

# Tableau Prep Help



# Contents

---

What's New in Tableau Prep Builder .....	1
Added in version 2020.2.1 .....	1
Connect to Data .....	1
Connect to your Salesforce.com data .....	1
Keep flow data fresh .....	1
Process only your newest data using incremental refresh .....	1
Added in version 2020.1.4 .....	4
Connect to Data .....	4
Connect to data with Impala .....	4
Clean and Shape Data .....	4
Convert date values in a single click .....	4
Save and Share your Work .....	5
Stay connected to your server .....	5
Added in version 2020.1.3 .....	6
Install and Deploy Tableau Prep Builder .....	6
Support for login-based license management .....	6
Connect to Data .....	6
Pass parameters to your data source using Initial SQL .....	6
Clean and Shape Data .....	7
Use Level of Detail (LOD), Rank and Row Number calculations to clean your data ..	7
Added in version 2020.1.1 .....	8
Connect to Data .....	8
Map Flow inputs back to their original connection .....	8
Clean and Shape Data .....	9

---

Pause data updates to boost performance .....	9
Version Compatibility with Tableau Prep .....	10
Detect and enable incompatible features .....	10
Keep flow data fresh .....	12
Add or embed credentials for cloud connectors when publishing your flow .....	12
Added in version 2019.4.2 .....	14
Connect to Data .....	14
Connect to more data with new connectors .....	14
Clean and Shape Data .....	15
Pivoting columns to rows just got easier .....	15
Additional Enhancements .....	16
See feedback for long-running operations .....	16
Support for login-based license management .....	17
Added in version 2019.4.1 .....	17
Connect to Data .....	17
Connect to more data in the cloud .....	17
Additional Enhancements .....	18
Share usage data with Tableau .....	18
Default database drivers no longer automatically installed for MacOS .....	18
Added in version 2019.3.2 .....	19
Build and Organize your Flow .....	19
Insert steps where you want them with one click .....	19
Easily review large flows with the flow navigator .....	20
Examine and Filter Your Data .....	20
Group similar values using custom data roles .....	20
Clean and Shape Data .....	22

Apply cleaning actions to fields using the new list view .....	22
Leverage your work and the work of others with reusable steps .....	22
Save and Share your Work .....	23
Disable the keychain dialog for Mac OS when running flows from the command line	24
Added in version 2019.3.1 .....	24
Connect to Data .....	24
Connect to data using Google BigQuery .....	24
Connect to published data sources for flow input .....	24
Examine and Filter Your Data .....	26
Create custom data roles .....	26
Clean and Shape Data .....	28
Use R and Python scripts in your flow .....	29
Save and Share your work .....	29
Spot incompatible features as you build your flow .....	29
Keep Flow Data Fresh .....	30
Tableau Data Management Add-on Includes Tableau Prep Conductor .....	30
Publish flows to Tableau Online .....	30
Additional Enhancements .....	31
Tableau Prep Builder now in Italian .....	31
Added in version 2019.2.3 .....	32
Examine and Filter Your Data .....	32
Select specific values to filter .....	32
Duplicate fields in one click .....	33
Clean and Shape Data .....	34
Fix invalid data role values with fuzzy matching and cleaning recommendations .....	35
Added in version 2019.2.2 .....	36

---

Connect to Data .....	36
Connect to more data with the Other Databases (ODBC) connector .....	36
Specify an Initial SQL command when connecting to your data .....	37
Examine and Filter Your Data .....	38
Keep only the fields you need while cleaning your data .....	38
Clean and Shape Data .....	38
Rearrange steps in your flow .....	38
Added in version 2019.2.1 .....	40
Connect to Data .....	40
Connect to more data with new database connectors .....	40
Keep Flow Data Fresh .....	40
View file safe list locations when publishing a flow .....	40
Other enhancements .....	41
Select from 10 different languages for your user interface .....	41
Quickly find your flow on Tableau server with flow thumbnail views .....	42
Added in version 2019.1.4 .....	43
Connect to Data .....	43
Connect to more data with new database connectors .....	43
Clean and Shape Data .....	43
Group similar values in a field by data role .....	43
Adjust your results when grouping field values .....	45
Added in version 2019.1.3 .....	46
Connect to Data .....	46
Quickly join your data based on table relationships .....	46
Clean and Shape Data .....	47
See all cleaning recommendations in the toolbar .....	47

Other enhancements .....	48
Find specific changes in your flow .....	48
Manage recently viewed flows from the Start page .....	49
Added in version 2019.1.2 .....	50
Keep Flow Data Fresh .....	50
Schedule flows to run automatically in Tableau Server .....	50
Help Design .....	50
Added in version 2019.1.1 .....	51
Clean and Shape Data .....	51
Pivot rows to columns .....	51
Use wildcard search to pivot fields .....	52
Use descriptions to add context to changes .....	53
Use cleaning recommendations to quickly split data in fixed-width text files .....	55
Added in version 2018.3.3 .....	56
Connect to Data .....	56
Automatically refresh input unions when new files are added .....	56
Clean and Shape Data .....	56
Reuse cleaning operations .....	57
Remove extra whitespace .....	57
Examine and Filter Your Data .....	57
Quickly find where fields are used in a flow .....	57
Other Enhancements .....	58
Open multiple windows of Tableau Prep Builder at once .....	58
Turn off usage reporting after install .....	58
Added in version 2018.3.2 .....	59
Clean and Shape Data .....	59

---

Improve data quality fast using cleaning recommendations .....	59
Other Enhancements .....	60
Added flexibility when running flows from the command line .....	60
Performance improvements .....	61
Added in previous versions .....	61
Install and Deploy Tableau Prep .....	62
Change your default repository location .....	62
Connect to Data .....	62
Clean data stored in PDF files .....	62
Wildcard union for database tables .....	63
Clean and Shape Data .....	64
Replace field values with Null .....	64
Apply cleaning operations in any step .....	65
Connect to Data .....	67
Connect to data stored in MongoDB Business Intelligence (BI) .....	67
Examine and Filter Your Data .....	67
Specify a data role for your field values .....	67
Join or Union Data .....	69
Fix mismatched fields directly in the join clause .....	69
Install and Deploy Tableau Prep .....	71
Set your display language .....	71
Connect to Data .....	72
Connect to data stored in Microsoft Access .....	72
Clean and Shape Data .....	72
Change the color scheme for your flow steps .....	72
Add descriptions to your steps .....	73



Use fuzzy match to find and fix spelling errors .....	74
Save and share your work .....	75
Run flows from the command line .....	75
Install and Deploy Tableau Prep .....	77
Deactivate Tableau Prep from the command line .....	77
Use virtual desktop support to optimize Tableau Prep Installations .....	77
Connect to Data .....	77
Connect to cloud data sources and Hadoop Hive .....	77
Explore Your Data .....	78
New filter options to keep only the data you want .....	78
Clean and Shape Data .....	78
Use the ISO-8601 date standard in calculated fields .....	79
Apply cleaning operations in the data grid .....	79
Use multi-select to group values in the Profile pane .....	80
Other enhancements .....	80
Join or Union Data .....	81
Use union recommendations to clean mismatched fields .....	81
Identify mismatched fields for all join types .....	83
Install and Deploy Tableau Prep .....	84
Activate and register Tableau Prep from the command line (Windows) .....	84
Connect to Data .....	84
Connect to data stored in statistical files or on Presto .....	85
Union sub-tables found by Data Interpreter in the Input step .....	85
Better feedback when loading tables .....	86
Explore Your Data .....	87
Reorder fields in the Profile pane and the Data grid .....	87

---

Clean and Shape Data .....	87
Use drag-select to remove multiple steps in your flow .....	87
Pivot multiple groups of fields in a single action .....	88
Improved field naming when merging fields .....	88
Other enhancements .....	89
Get Started with Tableau Prep Builder .....	91
Here's the story... .....	92
1. Connect to data .....	92
Check your work: Watch "Connect to data" in action. ....	95
2. Explore your data .....	96
3. Clean your data .....	98
Clean Orders_Central .....	99
Review your changes .....	105
Check your work: Watch "Clean Orders_Central" in action. ....	106
Clean Orders_East .....	107
Clean Orders_West .....	109
4. Combine your data .....	112
Union your data .....	113
Check your work: Watch "Union your data" in action. ....	116
Clean the product returns data .....	117
Join your data .....	122
Clean your join results .....	127
5. Run your flow and generate output .....	130
Wrap up and resources .....	132
About Tableau Prep Builder .....	133
Using Tableau Prep Builder .....	133

---

See Tableau Prep Builder in action .....	134
	135
A tour of the Tableau Prep Builder workspace .....	135
Connections pane .....	136
Flow pane .....	137
Profile pane .....	138
Data grid .....	139
How Tableau Prep Builder stores your data .....	140
Tableau Prep Visual Dictionary .....	141
Connect to Data .....	145
Start or open a flow .....	145
Start a new flow .....	147
Open an existing flow .....	149
Connect to published data sources (version 2019.3.1 and later) .....	149
Connect to cloud data sources .....	152
Connect to Salesforce data (version 2020.2.1 and later) .....	153
Connect to Google BigQuery (version 2019.3.1 and later) .....	153
Configure SSL to connect to Google BigQuery (MacOS only) .....	154
Set up and manage your Google BigQuery credentials .....	155
Connect to SAP HANA (version 2019.2.1 and later) .....	155
Connect using ODBC (version 2019.2.2 and later) .....	157
Before you connect .....	157
Connect using Other Databases (ODBC) .....	159
Use Initial SQL to query your connections (version 2019.2.2 and later) .....	161
Run Initial SQL .....	161
Include parameters in your Initial SQL statement (version 2020.1.3 and later) .....	162

---

Connect to Tableau data extracts .....	163
Use Data Interpreter to clean your files .....	163
Configure your Data Set .....	164
Connect to a custom SQL query .....	165
Apply cleaning operations in an input step .....	166
Select the fields to include in the flow .....	167
Apply filters to fields in the Input step .....	167
Change field names .....	168
Change data types .....	168
Configure field properties .....	168
Configure text settings in text files .....	169
Set your data sample size .....	169
Add More Data in the Input Step .....	170
Refresh data in the Input step .....	171
File input step types .....	171
File, database or Tableau extract input step types .....	171
Union files and database tables in the Input step .....	173
Union files .....	174
Union database tables (version 2018.3.1 and later) .....	177
Merge fields after a union .....	179
Join data in the Input step (version 2019.1.3 and later) .....	179
Build and Organize your Flow .....	182
Add or insert steps .....	182
Add steps .....	183
Insert steps .....	185
Change the flow color scheme .....	188

Add descriptions to flow steps and cleaning actions .....	189
Add a description to flow steps .....	189
Add a description to a change entry (version 2019.1.1 and later) .....	191
Reorganize the layout of your flow (version 2019.2.2 and later) .....	194
Use the flow navigator tool .....	195
Remove steps from the flow .....	196
Examine Your Data .....	198
Review the data types assigned to your data .....	198
See size details about your data .....	199
See the distribution of values or unique values .....	200
Search for fields and values .....	202
Sort values and fields .....	204
Highlight fields and values in a flow .....	205
Trace fields in a flow .....	205
See related values .....	206
Highlight identical values .....	206
Filter Your Data .....	207
Keep or remove fields .....	208
Filters available for each data type .....	209
Where are my filter options? .....	209
Calculation filter .....	209
Selected Values filter (version 2019.2.3 and later) .....	210
Range of Values filter .....	211
Range of Dates filter .....	211
Relative Date filter .....	212
Wildcard Match filter .....	212

---

Null Values filter .....	213
Use Data Roles to Validate your Data .....	213
Assign standard data roles to your data .....	214
Create custom data roles (version 2019.3.1 and later) .....	217
Create a custom data role .....	218
Apply a custom data role .....	222
View and manage custom data roles .....	224
Group similar values by data role .....	225
Clean and Shape Data .....	229
About cleaning operations .....	229
Available cleaning operations .....	229
Order of operations .....	231
Apply cleaning operations .....	233
Select your view .....	233
Pause data updates to boost performance .....	236
Apply cleaning operations .....	237
View your changes .....	239
Merge fields .....	241
Apply cleaning operations using recommendations (version 2018.3.2 and later) .....	243
Apply recommendations .....	244
Edit field values .....	246
Edit a single value .....	246
Edit multiple values .....	247
Edit multiple values using quick cleaning operations .....	247
Group and edit multiple values inline .....	248
Replace one or more values with Null .....	249

---

Manually map multiple values to a standard value .....	250
Map multiple values to a single selected field .....	250
Create a group by selecting multiple values .....	251
Add and identify values that aren't in the data set .....	252
Automatically map values to a standard value using fuzzy match .....	254
Group similar values using fuzzy match .....	256
Adjust your results when grouping field values .....	257
Copy steps, actions and fields .....	258
Copy and paste steps .....	259
Copy and paste cleaning operations .....	260
Copy fields (version 2019.2.3 and later) .....	263
Create reusable flow steps (version 2019.3.2 and later) .....	264
Create reusable steps .....	265
Insert reusable steps in a flow .....	266
Create level of detail and rank calculations .....	268
Calculate level of detail .....	268
Create Level of Detail (LOD) calculations .....	269
Calculation editor .....	269
Visual Calculation editor .....	271
Calculate rank or row number .....	272
Supported analytic functions .....	273
Create Rank or Row Number calculations .....	276
Calculation editor .....	276
Visual Calculation editor .....	279
Pivot Your Data .....	281
Pivot columns to rows .....	282

---

Watch "pivot on multiple field" in action. ....	285
Use wildcard search to pivot (version 2019.1.1 and later) ....	285
Pivot rows to columns (version 2019.1.1 and later) ....	287
Use R and Python scripts in your flow ....	290
Use R (Rserve) scripts in your flow ....	291
Prerequisites ....	292
Resources ....	292
Configure Rserve Server for Tableau Server ....	292
Additional Rserve configuration (optional) ....	293
Create your R script ....	294
Add a script to your flow ....	295
Use Python scripts in your flow ....	299
Prerequisites ....	300
Configure the Tableau Python (TabPy) server for Tableau Server ....	300
Create your python script ....	301
Add a script to your flow ....	303
Aggregate, Join, or Union Data ....	307
Aggregate and group values ....	307
Join your data ....	308
Inspect the results of the join ....	310
Common join issues ....	312
Fix mismatched fields and more ....	313
Union your data ....	314
Inspect the results of the union ....	315
Fix fields that don't match ....	317
Additional merge field options ....	320



---

Save and Share Your Work .....	321
Save a flow .....	321
View your data sample in Tableau .....	321
Create and publish data extracts and data sources .....	322
Create an extract file .....	323
Publish as a data source .....	325
Refresh Flow Data Using Incremental Refresh .....	326
Flow refresh options .....	327
Configure incremental refresh .....	327
Configure write options .....	330
Run your flow .....	333
Refresh flow output files from the command line .....	334
Before running the flow .....	335
Tips for creating your credentials file .....	338
Run the flow .....	339
Run the flow with incremental refresh enabled (version 2020.2.1 and later) .....	341
Command options .....	342
Syntax examples .....	344
The flow connects to and publishes to local files .....	344
The flow connects to and publishes to local files and uses the short form for incremental refresh .....	345
The flow connects to databases and publishes to a server .....	345
The flow publishes to a server and the credentials file is stored on a network share .....	346
Version Compatibility with Tableau Prep .....	347
Finding our version .....	347

---

Compatibility between different versions of Tableau Prep Builder .....	350
Fix compatibility issues with Tableau Prep Builder .....	351
Compatibility between different versions of Tableau Prep Builder and Tableau Server .....	351
Detect incompatible features .....	351
Tableau Prep Builder (version 2020.1.1 and later) .....	352
Tableau Prep Builder (version 2019.3.1 and later) .....	353
Tableau Prep Builder (all versions) .....	354
Fixing compatibility issues .....	356
Identify incompatible features .....	356
Remove incompatible features from the flow .....	358
Incompatible data sources .....	358
Incompatible features .....	359
Keep Flow Data Fresh .....	360
About Tableau Prep Conductor .....	362
Enabling Tableau Prep Conductor .....	363
Getting to know the flow workspace .....	363
Flow Overview page .....	363
Flow Connections page .....	365
Flow Scheduled Tasks page .....	365
Schedules page .....	366
Flow Run History .....	367
Flow Revision History .....	367
Who can do this .....	368
Licensing Tableau Prep Conductor .....	368
Enable and Configure Tableau Prep Conductor on Tableau Server .....	368

- Server Topology ..... 368
- Next step: ..... 370
  - Who can do this ..... 370
- Step 1 (New Install): Install Tableau Server with Tableau Prep Conductor ..... 370
  - Before you install ..... 370
  - Install Tableau Server and enable Tableau Prep Conductor ..... 371
    - Configure public gateway settings ..... 371
    - Enable Tableau Prep Conductor ..... 371
    - Verify Tableau Prep Conductor is enabled and running ..... 372
  - Dedicate a node for Tableau Prep Conductor ..... 374
    - Multi-node installations ..... 374
  - Next step ..... 374
    - Who can do this ..... 375
- Step 1 (Existing Install): Enable Tableau Prep Conductor ..... 375
  - Before you upgrade ..... 375
    - Prepare for upgrade: ..... 375
    - Configure public gateway settings ..... 375
- Tableau Server Installations using User-Based licenses ..... 376
  - Tableau Server single-node installations ..... 376
  - Tableau Server multi-node installations ..... 378
- Tableau Server Installations using Core-Based licenses ..... 381
  - Tableau Server single-node installations ..... 381
  - Tableau Server multi-node installations ..... 384
- Next step ..... 387
  - Who can do this ..... 387
- Step 2: Configure Flow Settings for your Tableau Server ..... 387

---

Publishing, Scheduling, and Credential Settings .....	387
Implication of disabling Tableau Prep Conductor .....	388
Configure notifications for flow failures .....	389
To enable the server-wide email notification .....	389
Set notification values .....	390
To configure email notification for a site: .....	391
Next step .....	391
Who can do this .....	391
Step 3: Create Schedules for Flow Tasks .....	392
Create a new schedule: .....	392
Next step .....	393
Who can do this .....	393
Step 4: Safe list Input and Output locations .....	393
How to safe list input and output locations .....	394
Next step .....	395
Who can do this .....	395
Step 5: Optional Server Configurations .....	396
Set the timeout period for flows .....	396
Set the threshold for suspended flow tasks .....	396
Who can do this .....	396
Enable Tableau Prep Conductor on your Tableau Online Site .....	397
Step 1: Install Tableau Prep Builder .....	397
Step 2: Log into Tableau Online .....	397
Publish a Flow to Tableau Server or Tableau Online .....	398
Before you publish .....	398
Publish a flow .....	402

---

Tableau Server .....	403
Files .....	404
Databases .....	406
Tableau Online .....	410
Files .....	411
Databases .....	412
Who can do this .....	417
Schedule a Flow Task .....	417
Schedule a flow task .....	418
Who can do this .....	421
Manage a Flow .....	421
Managing your flows .....	422
Who can do this .....	423
Monitor Flow Health and Performance .....	429
Detect issues as they occur and resolve them .....	430
Get email notifications when a flow fails: .....	430
View and resolve errors .....	430
Flow Overview page .....	431
Connections page .....	431
Scheduled Tasks page .....	432
Run History page .....	432
Alerts .....	433
Tableau Prep Conductor process status .....	434
Monitor flow activity and performance with Administrative Views .....	435
Who can do this? .....	435
Action by all users .....	436

---

Action by Specific User .....	436
Action by Recent Users .....	437
Backgrounder Task Delays .....	438
Background Tasks for Non Extracts .....	438
Performance of Flow Runs .....	439
Stats for Space Usage .....	440
Who can do this .....	441
Developer Resources - REST APIs .....	441
Day in the Life Scenarios .....	443
Hospital Bed Use with Tableau Prep .....	443
The Data .....	444
Preliminary Analysis .....	444
Desired Data Structure .....	445
Restructuring the Data .....	447
Bed Hour Matrix .....	447
Patient Bed Use .....	450
Analysis in Tableau Desktop .....	454
Recap and Resources .....	457
Finding the Second Date with Tableau Prep .....	458
The Data .....	459
Desired Data Structure .....	460
Restructuring the Data .....	461
Initial Aggregation for 1st Infraction Date .....	461
Second Aggregation for 2nd Infraction Date .....	465
Create full data sets for the 1st and 2nd infractions .....	466
Create the complete data set .....	469

Recap .....	470
Continue to Analysis with the Second Date in Tableau Desktop on page 1. ....	471
Analysis with the Second Date in Tableau Desktop .....	471
Analysis in Tableau Desktop .....	472
Go Further—Pivoted Data .....	479
The benefits of pivoted data .....	487
Go Further Still—Calculations Only .....	488
Reflection on Methods .....	494
Driver Infractions .....	494
Pivoted Driver Infractions .....	495
LOD Driver Infractions .....	495
Troubleshoot Tableau Prep Builder .....	499
Common errors when using the command line to run flows .....	499
Error: "These features were found that prevent this version of the application from using this file" .....	503
Error: "Failed to parse response from Tableau Server" when publishing Tableau Prep output .....	504
Error: "You are using Server version: null..." when signing in to an SSL-enabled Tableau Server using Tableau Prep .....	505
Maintain Licenses for Tableau Desktop and Tableau Prep .....	505
View data about your license .....	506
Track Tableau Desktop license usage and expiration data .....	507
Additional resources .....	508





# What's New in Tableau Prep Builder

Browse summaries of new features for currently supported versions.

**Note:** Starting in version 2019.1.2, Tableau Prep is now called Tableau Prep Builder.

## Added in version 2020.2.1

### Connect to Data

#### Connect to your Salesforce.com data

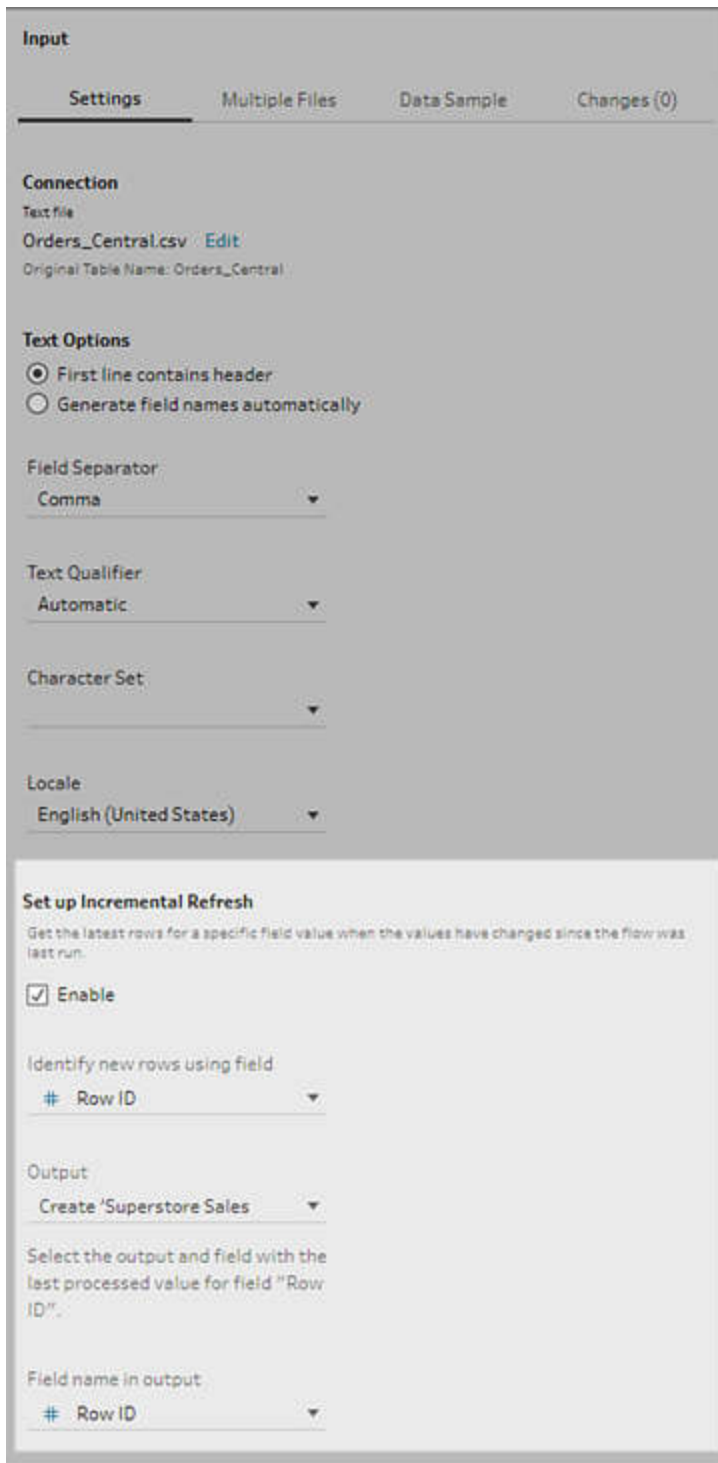
Tableau Prep Builder can now connect to your Salesforce data so you can clean and prepare it for analysis in Tableau. Using custom SOQL and creating a standard connection are not yet supported.

For more information see [Connect to Salesforce data \(version 2020.2.1 and later\)](#) on page 153.

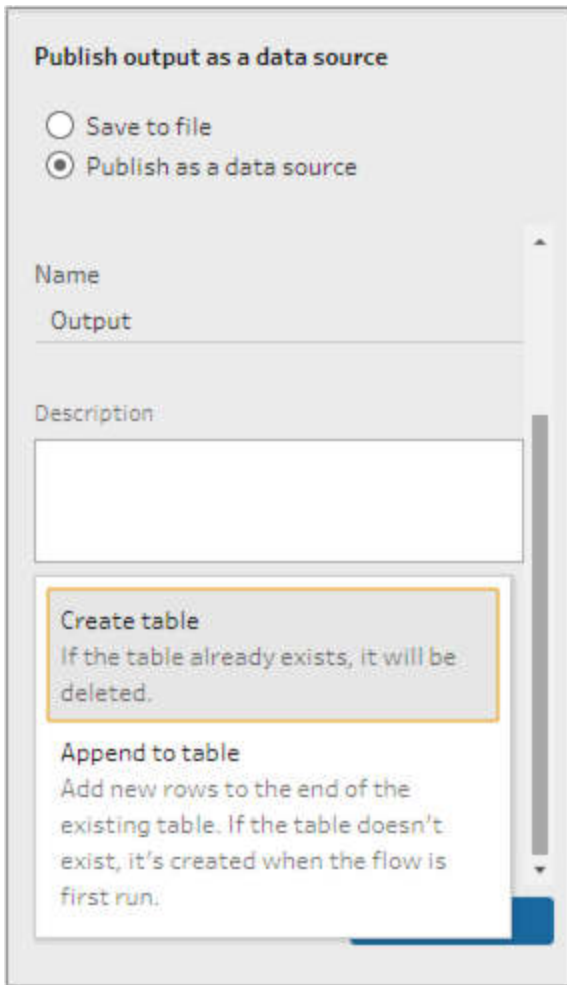
### Keep flow data fresh

#### Process only your newest data using incremental refresh

If your flow data is constantly changing, you can now save time and resources by using incremental refresh to retrieve and process only the newest rows of data since the flow was last run. Just configure incremental refresh parameters for your flow to tell Tableau Prep Builder how to detect and retrieve the new rows from your input data. You can also configure how you want to update your output tables when the flow is run in the Output step.

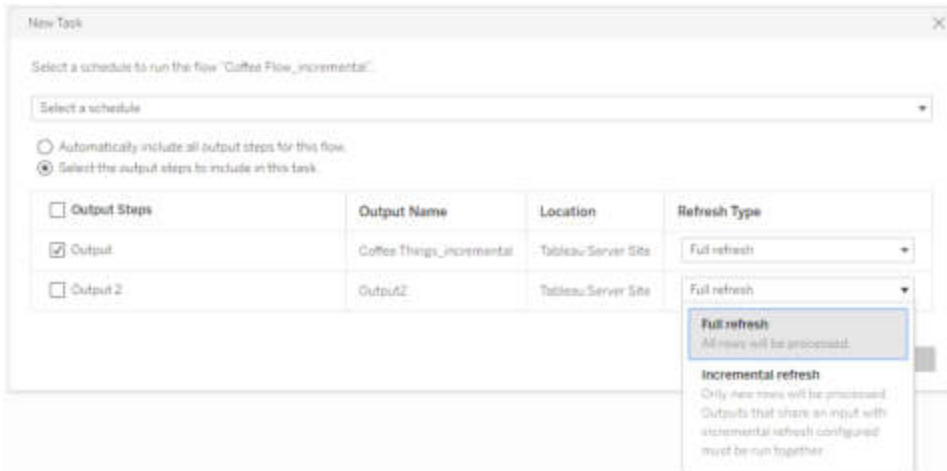


New options in the Output step enable you to choose to append new data to your existing tables or use Tableau Prep Builder's existing functionality and replace your table data with the new data. This option is available for both incremental and full refresh types.



Set up and run your flows on a schedule using Tableau Prep Conductor or run individual flows incrementally in Tableau Prep Builder or through the command line interface.

**Note:** Tableau Server version 2020.2 and later is required (with Tableau Prep Conductor enabled) to schedule flows using incremental refresh.



For more information about setting up and running flows using incremental refresh, see [Refresh Flow Data Using Incremental Refresh](#) on page 326.

## Added in version 2020.1.4

### Connect to Data

#### Connect to data with Impala

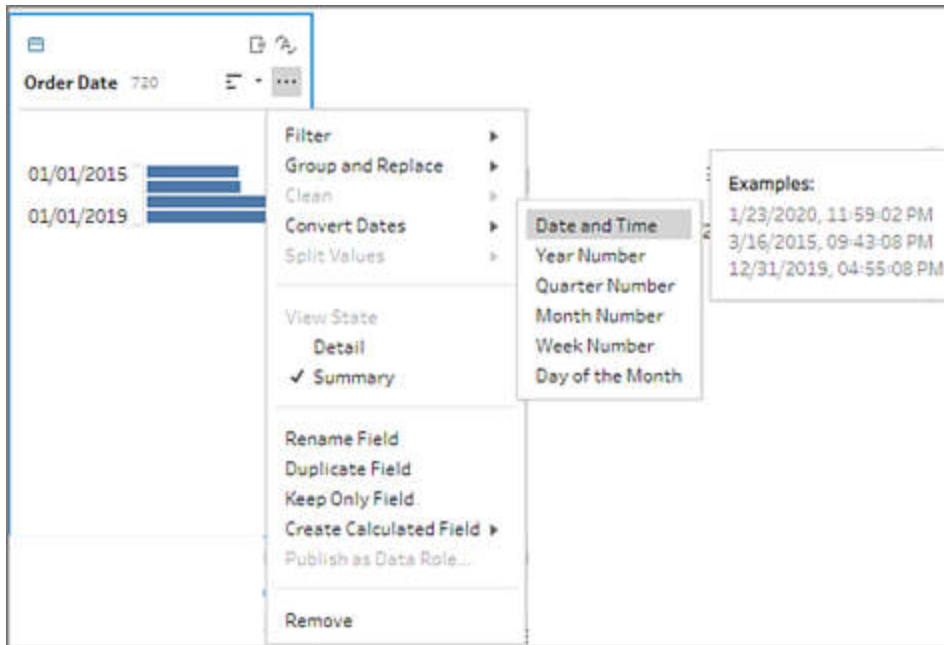
Just like Tableau Desktop, you can now easily connect to your larger data sets using Impala.

For more information about connecting to your data source using Impala, see [Impala](#) in the Tableau Desktop help.

### Clean and Shape Data

#### Convert date values in a single click

To help make data cleaning even easier, Tableau Prep Builder has added another quick cleaning menu for dates. If you need to convert your date values to show a DATEPART like year, month or day, you can select one of the following menu options from the new **Convert Dates** menu instead of writing a complex field calculation.

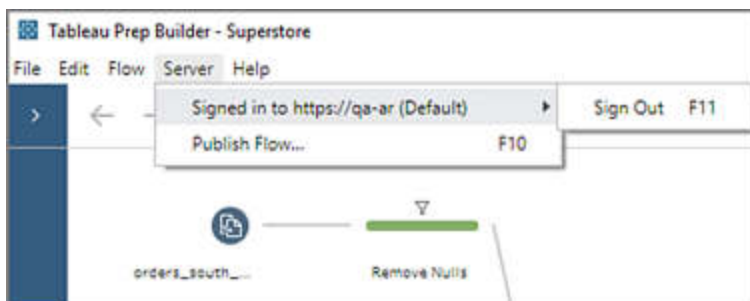


For more information about applying cleaning operations to your dates, see [Apply cleaning operations](#) on page 237.

## Save and Share your Work

### Stay connected to your server

You no longer need to sign into Tableau Server or Tableau Online every time you open Prep Builder. Now, Tableau Prep Builder remembers your server name and login information when you close the application so that the next time you open the application you are already logged into your server.



This time saving option will help you seamlessly connect to publish data sources or insert saved flow steps and if you need to know if a certain feature is compatible with your server version, compatibility warnings will show instantly without you having to sign into your server first.

For more information, see [Publish as a data source](#) on page 325.

## Added in version 2020.1.3

### Install and Deploy Tableau Prep Builder

#### Support for login-based license management

Login-based license management, helps you manage licensing for users with Creator roles on Tableau Server and Tableau Online. Users with Explorer or Viewer roles cannot use this feature. If you're using Role Based Subscriptions with Tableau Server or Tableau Online, you can simplify your license management using login-based license management to eliminate separate Tableau Desktop and Tableau Prep Builder product keys. You only need to manage one or more product keys for on-premises Tableau Server, or in the case of Tableau Online, you don't need to manage any product keys at all.

### Connect to Data

#### Pass parameters to your data source using Initial SQL

Now you can pass parameters to your data source using Initial SQL to provide more details about your Tableau Prep application, version and flow name. Tableau Prep Builder supports the following parameters:

- **TableauApp** - Returns the application being used to access your data source - Prep Builder or Prep Conductor.
- **TableauVersion** - Returns the application version number. For Tableau Prep Builder it will return the exact version number, for example 2020.1.1. For Tableau Prep Conductor, it will return the major server version where Tableau Prep Conductor is enabled, for example 2020.1.
- **FlowName** - Returns the name of the .tfl file.

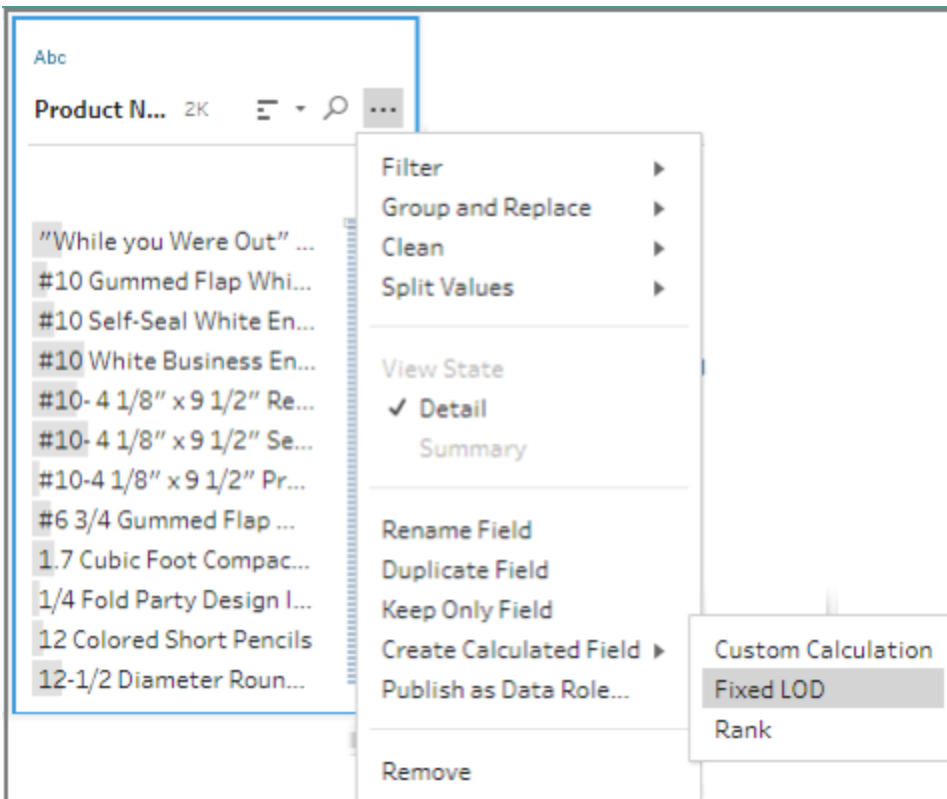
For more information about using parameters when connecting to your data source using Initial SQL, see [Include parameters in your Initial SQL statement \(version 2020.1.3 and later\)](#) on page 162.

## Clean and Shape Data

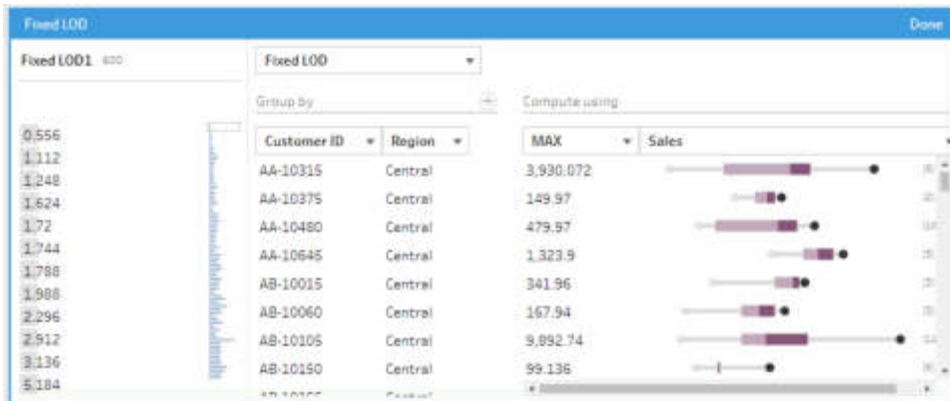
### Use Level of Detail (LOD), Rank and Row Number calculations to clean your data

If your data is aggregated at a different level than what you need or if you need to perform calculations across a set of rows you can now use FIXED Level of Detail (LOD) and RANK and ROW\_NUMBER analytic function calculations to create the data that you need in your data set.

For example, add a FIXED LOD calculation to change the granularity of fields in your table. Use the new ROW\_NUMBER () analytic function to quickly find duplicate rows, or use one of the new RANK () functions to find the top N or bottom N values for a selection of rows with similar data.



If you want a more guided experience when building these types of expressions, you can use the new visual calculation editor.



For more information about using these calculated fields to clean your data, see [Create level of detail and rank calculations](#) on page 268.

## Added in version 2020.1.1

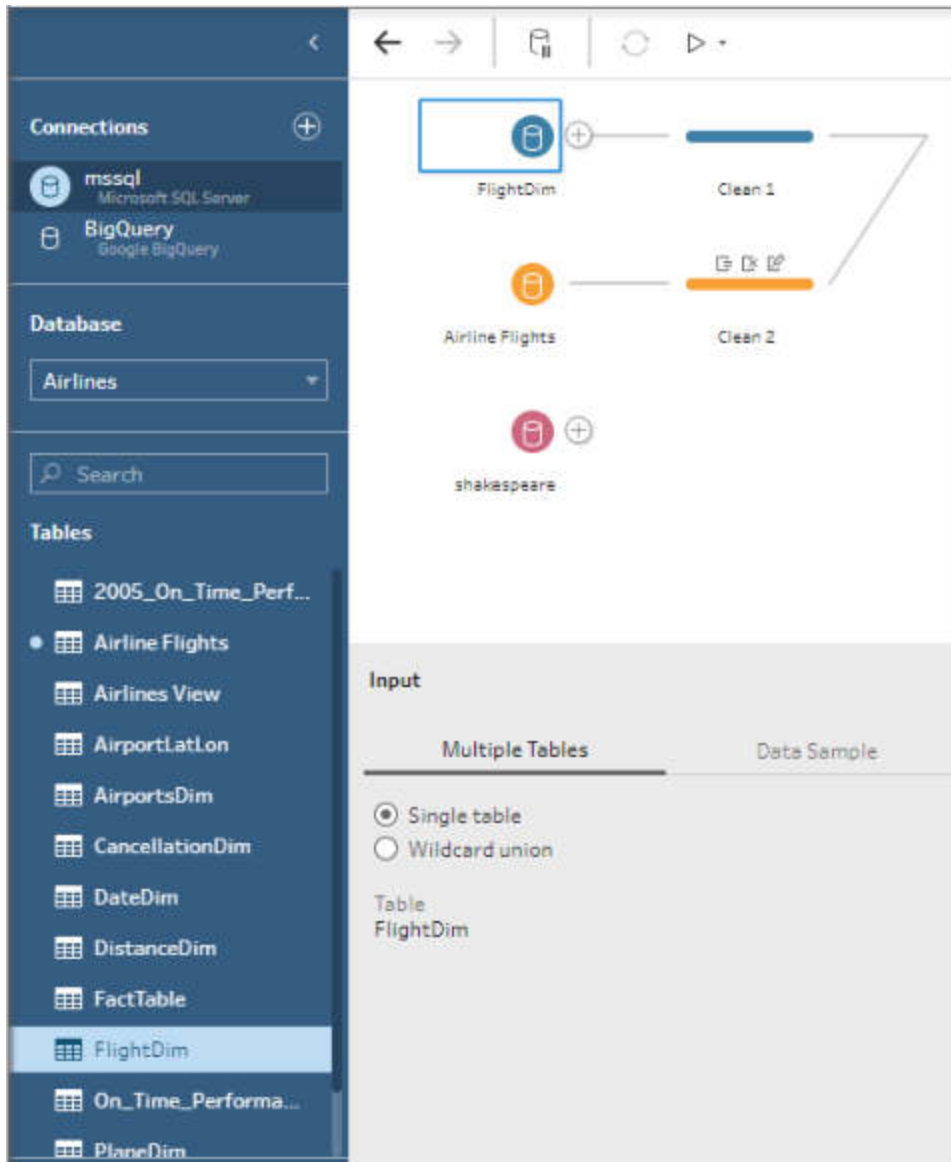
### Connect to Data

#### Map Flow inputs back to their original connection

If you use multiple inputs to build your flows, trying to find the individual tables or data sources you used for the different inputs when you need to make updates or swap out a connection isn't always obvious. To make finding your flow inputs and tables easier, we've added a new feature that highlights the connector and table for a selected input in the Connections pane.

Simply select one or more Input steps in the Flow pane and the related connector and tables are highlighted in the Connections pane.





For more information about how to connect to your data to build a flow see [Connect to Data](#) on page 145.

## Clean and Shape Data

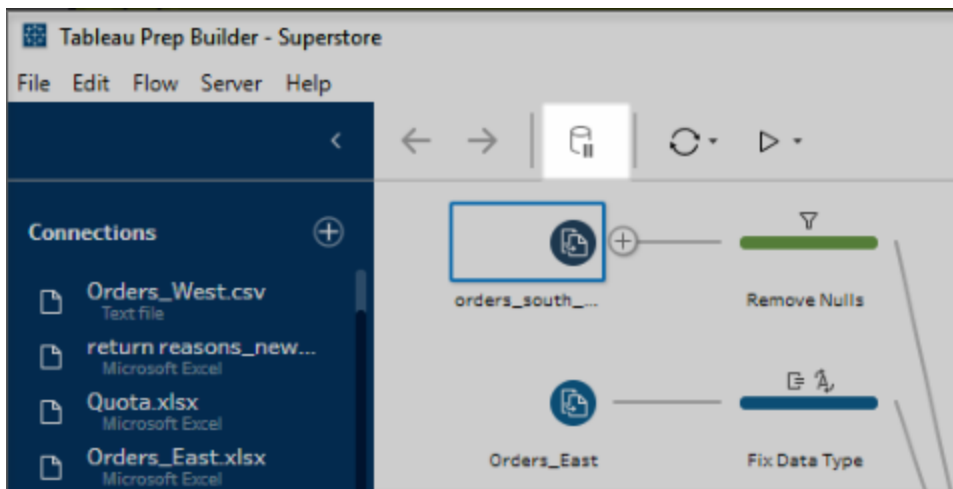
### Pause data updates to boost performance

Every time you apply cleaning operations to your data, Tableau Prep Builder refreshes your view so you can instantly see the results of your changes. While building new flows, this constant feedback helps you quickly know which steps to take next, but sometimes applying

these changes can take time. To boost performance and make applying changes to your data faster, you can now pause these constant data updates and apply your changes in bulk.

Simply select the **Pause data updates** button on the top menu and Tableau Prep Builder pauses your data updates and switches the Profile pane view into List view so you can quickly apply your changes.

Instead of refreshing your view with every change, only the field metadata is shown. When you finish your changes and want to see the results on your field values, click the **Resume data updates** button to switch back to the standard data update mode.



**Note:** Any cleaning operations that require data values to perform the task are disabled while data updates are paused.

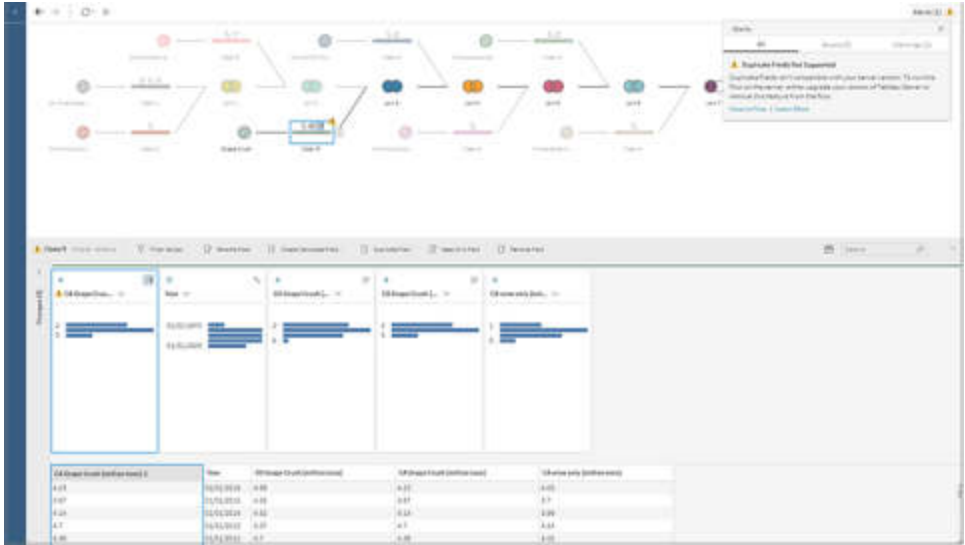
For more information about pausing data updates to apply cleaning operations in bulk, see [Pause data updates to boost performance on page 236](#).

## Version Compatibility with Tableau Prep

### Detect and enable incompatible features

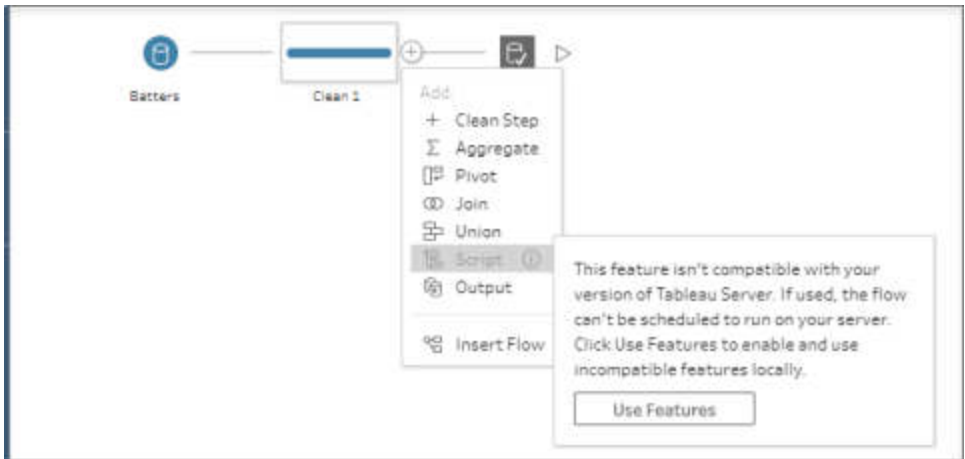
In Tableau Prep Builder version 2019.3.1, we added warning icons and tooltips so that you could more easily see which features were incompatible with your version of Tableau Server as you built your flow to help you make informed decisions about which features you included in your flow if you planned to publish the flow and run it on a schedule.

## Tableau Prep Help



In this release we have added even more transparency when working with new features that might not be compatible with your server version.

Now, after you sign into your server, Tableau Prep Builder can detect which features aren't compatible with your version of Tableau Server, and disable those features. This takes the guess work out of which features you can include if you plan to publish your flow and run it on a schedule.



If you still want to use the feature and run your flows manually using Tableau Prep Builder or from the command line, you can enable the feature right from the menu. Just click the **Use Features** button, and that feature is applied to your flow and all incompatible features are enabled and available to use.

The warning options introduced in Tableau Prep Builder version 2019.3.1 will still identify which features are incompatible with your server so that you can easily find and remove them if you want to run the flow using a schedule in Tableau Server.

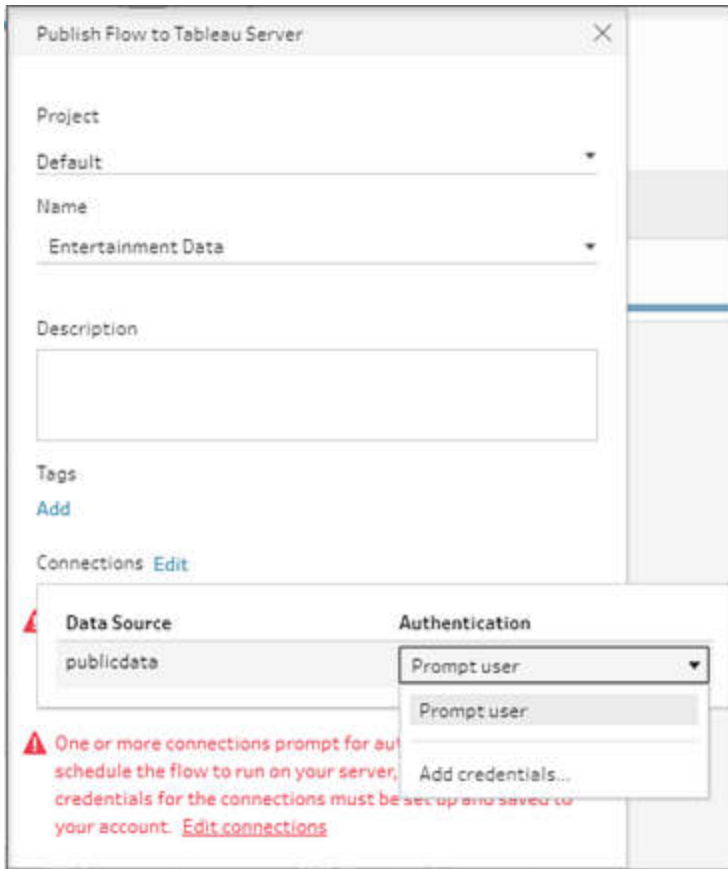
For more information about working with incompatible features, see [Version Compatibility with Tableau Prep](#) on page 347.

## Keep flow data fresh

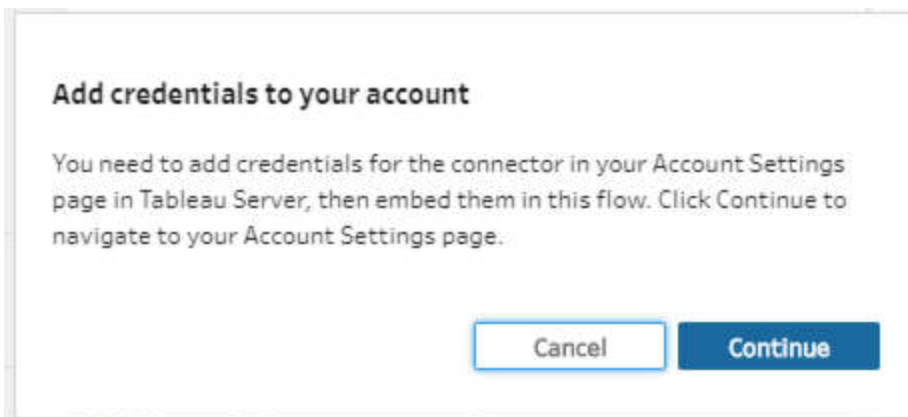
### Add or embed credentials for cloud connectors when publishing your flow

In previous versions, if you wanted to publish and schedule a flow that included cloud connectors, you would need to configure your credentials in Tableau Server or Tableau Online ahead of time. Otherwise, Tableau Prep Conductor couldn't detect your credentials and the flow would fail to run. Now you can add and embed your credentials for your cloud connectors directly from the Publish dialog in Tableau Prep Builder in a few simple steps.

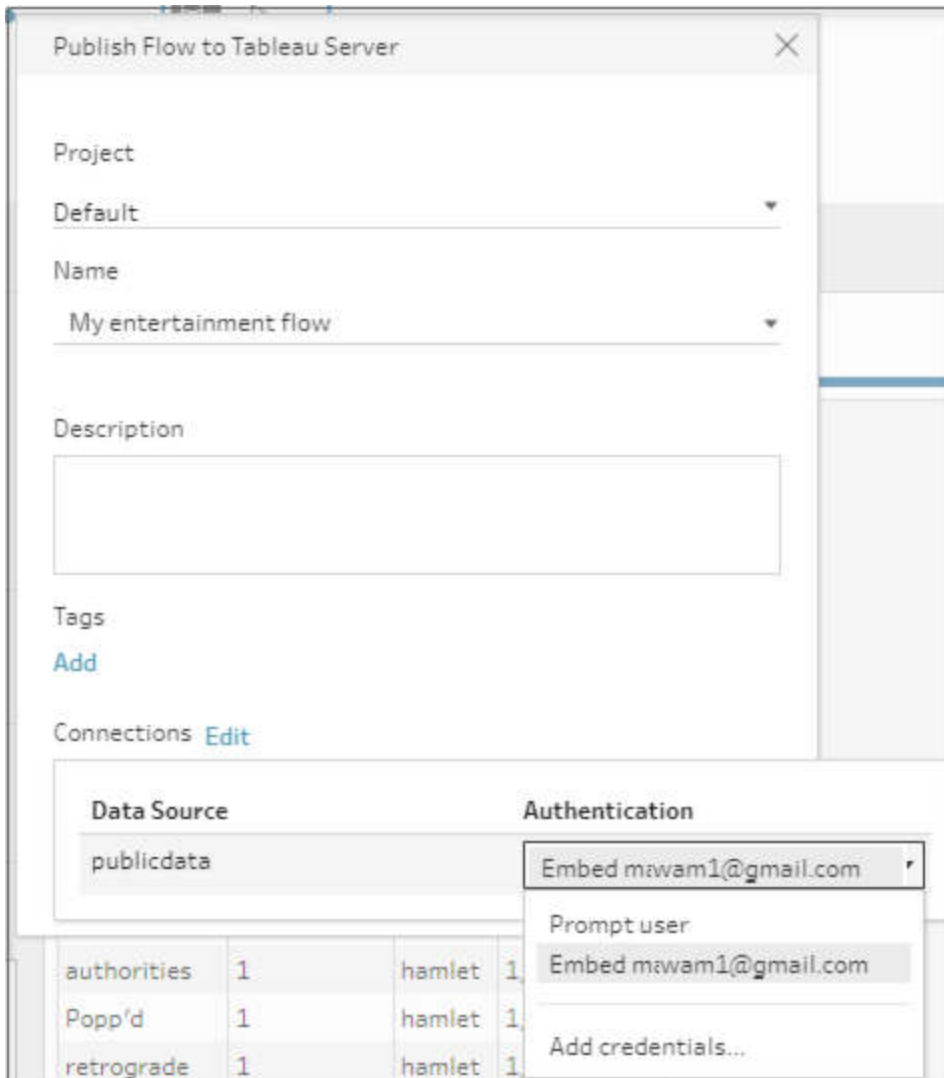
Simply sign into your server and select **Server > Publish Flow**. In the Publish Flow dialog, a new warning text shows, telling you that you have input connections that require authentication. Click the **Edit connections** link in the message or the Edit link in the Connections section, then in the Authentication drop-down, select **Add credentials**.



A secondary message confirms your selection. Click **Continue** to go to your **Account Settings** page in Tableau Server or Tableau Online, and add your credentials.



When you are finished, return to Tableau Prep Builder, click **Done** and your new credentials are automatically embedded in your flow.



For more information about adding credentials for cloud connectors during publish, see [Publish a flow](#) on page 402.

## Added in version 2019.4.2

### Connect to Data

#### Connect to more data with new connectors

In this release we've added the following new data connectors:

## Tableau Prep Help

- Alibaba AnalyticDB for MySQL
- Alibaba Data Lake Analytics
- Databricks
- Kyvos
- Qubole Presto

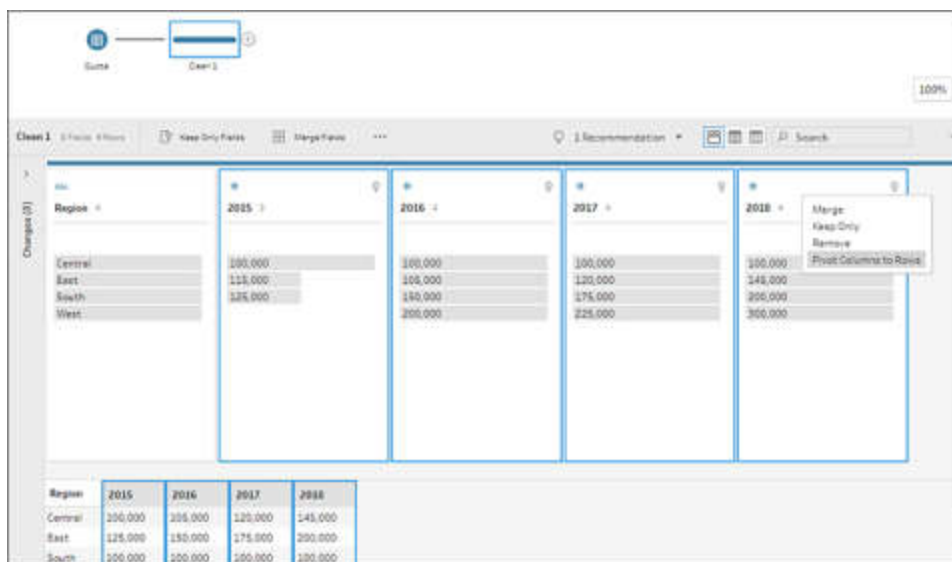
For more information about how to connect to your data using these connectors, see [Supported Connectors](#) in the Tableau Desktop and Web Authoring help. For general information about connecting to data to build a flow see [Connect to Data](#) on page 145.


## Clean and Shape Data

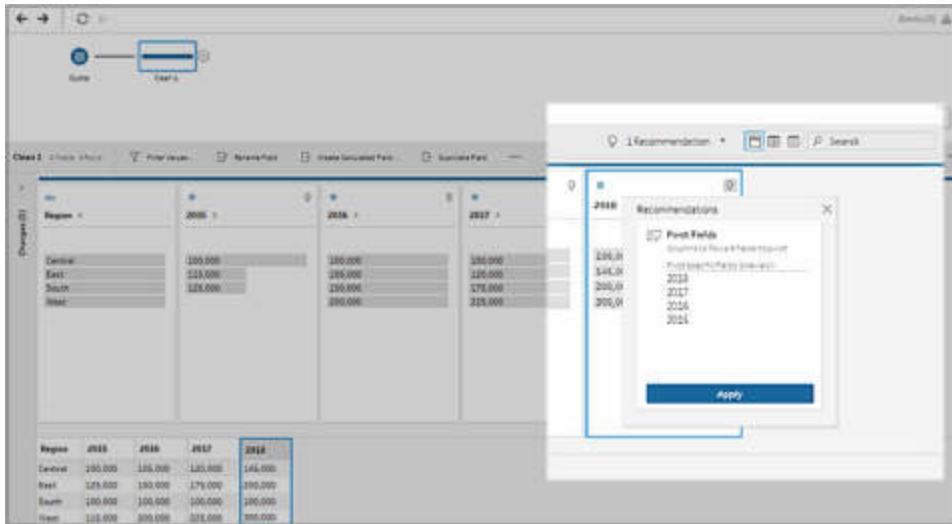
### Pivoting columns to rows just got easier

When you need to pivot your data from columns to rows, you can now manually select fields in the profile pane to pivot them or Tableau Prep Builder can recommend the fields to pivot and create the pivot for you. Both options automatically generate your pivot step so you can quickly perform any additional pivot actions like renaming the pivoted fields or pivoting on additional fields.

To manually pivot columns to rows from the Profile pane, simply select the fields you want to pivot, then right-click or Ctrl-click (MacOS) and select **Pivot Columns to Rows** from the menu to create the new pivot step and pivot your fields.



If Tableau Prep Builder detects fields that can be pivoted, a recommendation option will show in the Profile pane toolbar. Click the light bulb  icon in the toolbar and select the pivot recommendation. The recommendation shows which fields it will pivot. Simply click **Apply** to generate the Pivot step and pivot your fields.



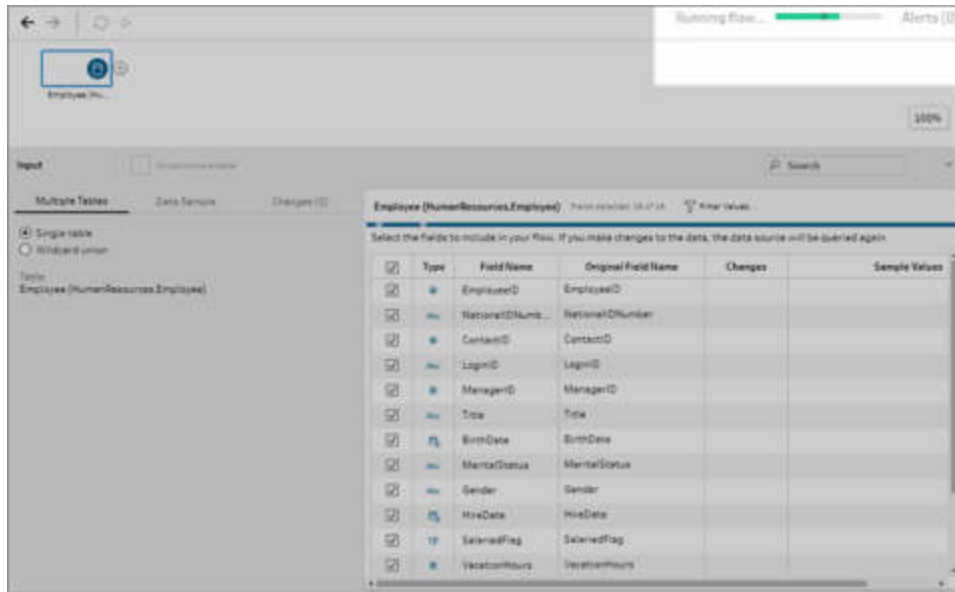
For more information about pivoting your data, see [Pivot Your Data](#) on page 281. For more information about how to use recommendations to clean your data, see [Apply cleaning operations using recommendations \(version 2018.3.2 and later\)](#) on page 243.

## Additional Enhancements

### See feedback for long-running operations

To give you more visibility when processing large data sets or flows, we've added a progress indicator at the top of the Flow pane to give you status updates as Tableau Prep Builder processes your data. Now you can have more visibility into what Tableau Prep Builder is doing behind the scenes so you no longer have to wonder whether your data is still being processed or if a problem has occurred.





## Support for login-based license management

Login-based license management, helps you manage licensing for users with Creator roles on Tableau Server and Tableau Online. Users with Explorer or Viewer roles cannot use this feature. If you're using Role Based Subscriptions with Tableau Server or Tableau Online, you can simplify your license management using login-based license management to eliminate separate Tableau Desktop and Tableau Prep Builder product keys. You only need to manage one or more product keys for on-premises Tableau Server, or in the case of Tableau Online, you don't need to manage any product keys at all.

## Added in version 2019.4.1

### Connect to Data

#### Connect to more data in the cloud

In this release we've added the following cloud connectors:

- Box
- DropBox
- Google Drive
- OneDrive

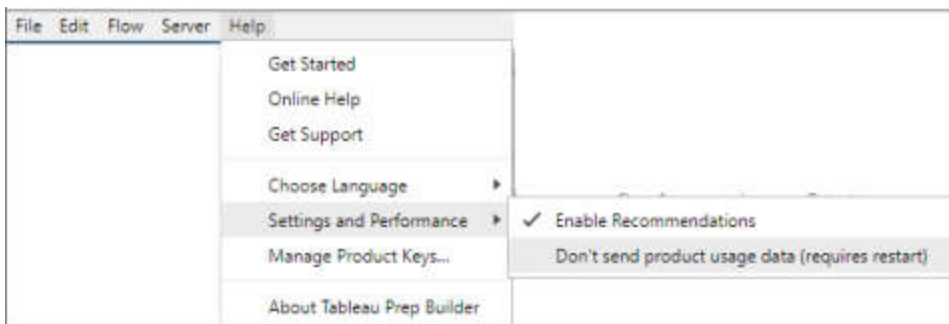
Additional configuration may be needed to save your credentials for these connectors if you plan to publish your flows to Tableau Server or Tableau Online. For more information about connecting to cloud data connectors, see [Connect to cloud data sources on page 152](#).

## Additional Enhancements

### Share usage data with Tableau

Tableau collects data that helps us learn how our products are being used so we can improve existing features and develop new ones. All usage data is collected and handled according to the [Tableau Privacy Policy](#). But you can opt out if you don't want to participate.

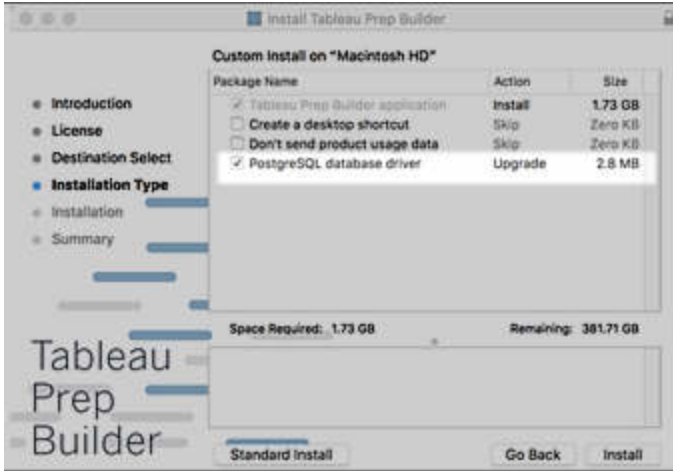
The opt out option has changed. Usage data collection is enabled by default. If you don't want to participate, select the **Don't send product usage data** check box from the installer or Customize installer (MacOS) dialog or from the Help menu after installation.



For more information, see [Turn off usage reporting](#) in the Tableau Desktop and Tableau Prep Deployment Guide.

### Default database drivers no longer automatically installed for MacOS

When installing Tableau Prep Builder on Mac, we only install the PostgreSQL database driver automatically. If you need other drivers to connect to your data, you can install them from the [Driver Download](#) page.



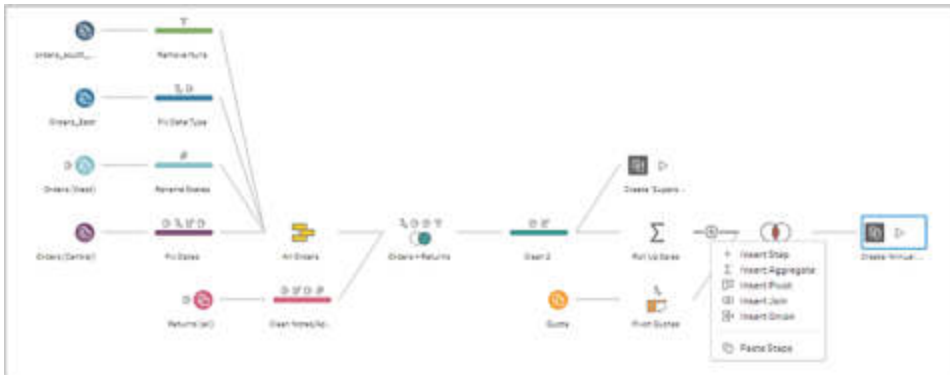
## Added in version 2019.3.2

### Build and Organize your Flow

Insert steps where you want them with one click

In prior versions you could insert a cleaning step after an existing step in a flow and the flow would extend to the right to make room for the new step. But adding any other step types after existing steps resulted in creating a branch to your flow that you would then have to reconnect by removing and adding flow lines.

In this release, we've added a new insert menu that lets you insert any step type (even copied steps) between existing steps in one click. Simply hover on the flow line where you want to insert a step and use the new Insert menu to select and insert the step type you want, right where you want it.

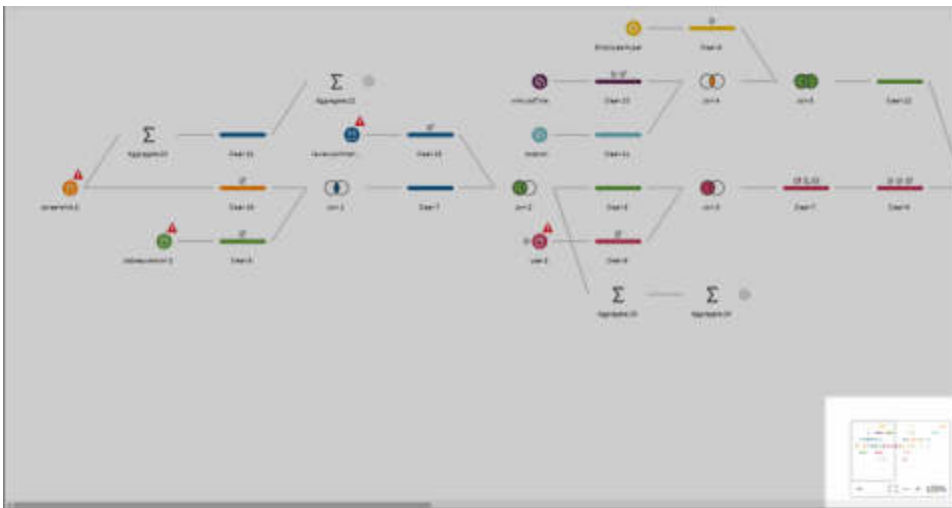


**Note:** The **Insert Step** menu item from the **Add** menu was moved to this new menu. This option is not available for Input and Output step types.

For more information about adding and inserting steps in a flow, see [Add or insert steps on page 182](#).

## Easily review large flows with the flow navigator

When working with large flows it can be difficult to scroll back and forth to find and fix a step or change. The new flow navigator tool shows in the lower right corner of the flow pane. This tool is a miniature image of your entire flow. You can use it to zoom in and out, click on parts of the image to navigate directly to that part of your flow, adjust your flow size and more. If you don't want to show the tool you can minimize it or restore it in a single click.



For more information about connecting to data and building a flow, see [Use the flow navigator tool on page 195](#).

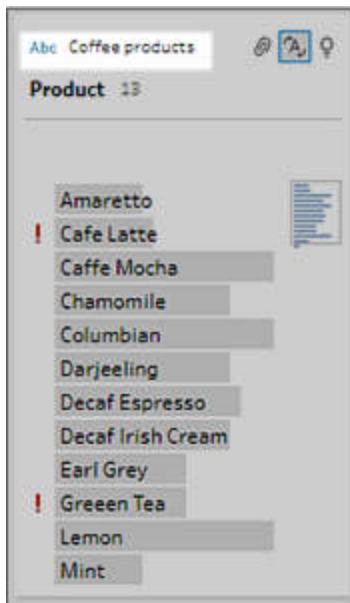
## Examine and Filter Your Data



### Group similar values using custom data roles

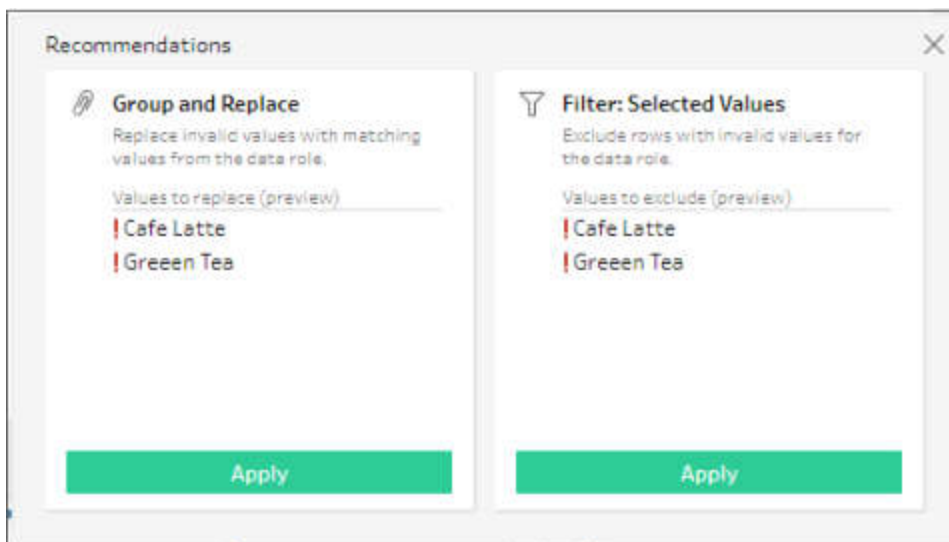
In Tableau Prep Builder version 2019.3.1 we introduced the ability to create your own custom data roles using the field values in your data sets. In this release you can now use those custom data roles to group values in a field based on spelling and pronunciation to automatically fix invalid values by matching and replacing them with the closest valid one.

## Tableau Prep Help

Simply apply the custom data role to your field and any values that don't match the standard values are marked with a red exclamation mark.



To fix the invalid values, you can click on the **Recommendations**  icon on the field to apply the recommendation to group and replace the invalid values with valid ones or select **Group and Replace > Spelling** or **Pronunciation + Spelling** from the **More options**  menu to select how you want to fix your invalid values.

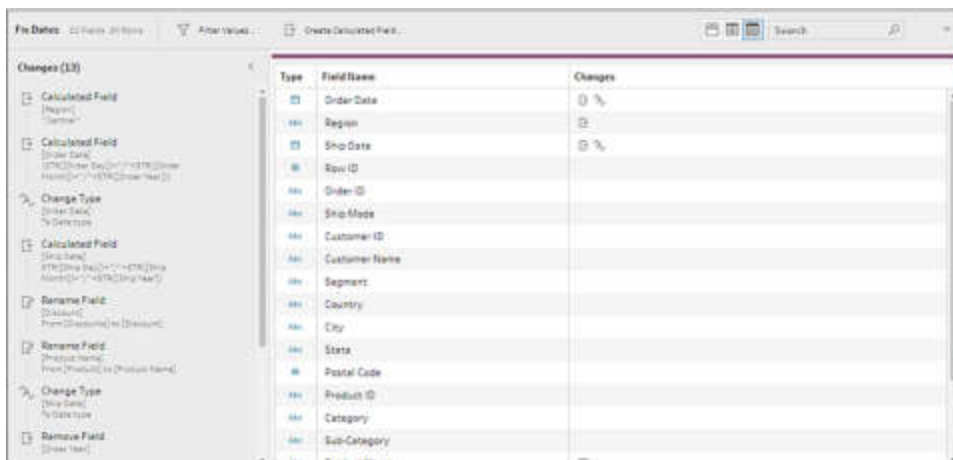


For more information about creating and using custom data roles see [Group similar values by data role](#) on page 225.

## Clean and Shape Data

### Apply cleaning actions to fields using the new list view

Currently you can apply cleaning operations in the profile pane, results pane, or data grid. But the number of fields you can see at one time is limited by the size of your screen. To make scanning all your fields and taking action easier, we've added a new list view that lets you convert the profile pane or results pane into a list. After you select this option, this view state persists across all steps in your flow but you can change it at any time.



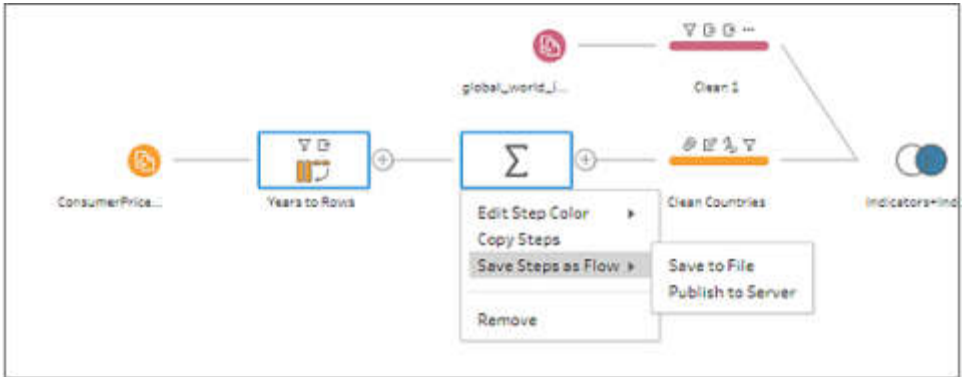
In this view you can apply operations to multiple fields like **Remove** or **Keep Only** or quickly apply cleaning operations like **Rename Field** or **Duplicate Field**. You can also apply other cleaning operations from the **More options**  $\dots$  menu, but operations like **Group and Replace**, **Filter**, or **Split Values** will return you to the Profile pane or Results pane view to complete your actions.

For more information about cleaning fields in the list view, see [Select your view on page 233](#).

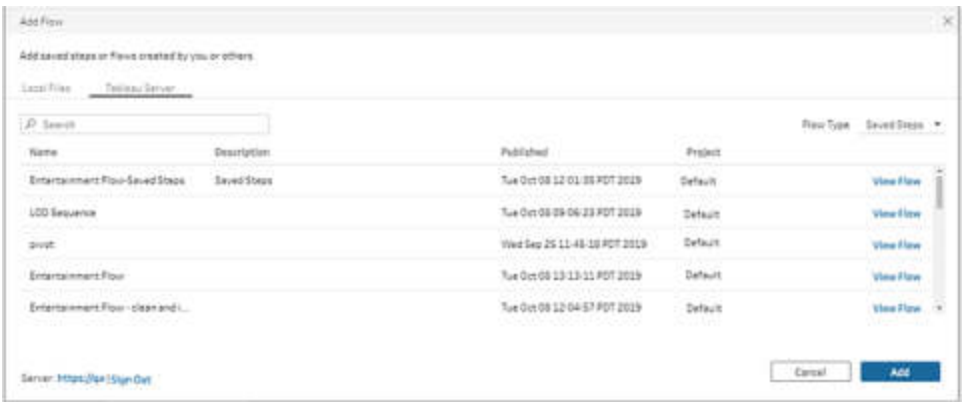
### Leverage your work and the work of others with reusable steps

If you commonly perform the same actions over and over again with your data you can select one or more flow steps and their associated actions, or the entire flow, and save it to reuse in other flows you create and manage. You can save your steps locally or publish them to Tableau Server or Tableau Online to share with others.

To create reusable steps, simply select the steps that you want to save, right-click and select **Save Steps as Flow** and select whether to save the steps to your local file or to a server. A **Saved Steps** tag is added automatically when publishing flow steps to your server to make it easy to find the flow steps later.



Once saved, you can open and insert any flow. Right-click or Ctrl-click (MacOS) anywhere in the flow pane canvas and select **Insert Flow**, or select **Edit > Insert Flow** from the top menu. In the **Add Flow** dialog, you can select and insert saved flow steps from a file or your server. The dialog automatically filters to show all flows with the **Saved Steps** tag, but you can change the filter and insert any flow that you have permissions to access.



The selected flow is added to the flow pane and you can then move and connect your steps where you want them in your flow. For more information about creating and using reusable flow steps, see [Clean and Shape Data on page 229](#).

## Save and Share your Work

## Disable the keychain dialog for Mac OS when running flows from the command line

When running flows from the command line on the MacOS, we've eliminated the need for a user to manually enter their MacOS keychain credentials every time the flow runs.

For more information, see the Disable SSL validation (MacOS) command option in [Refresh flow output files from the command line](#) on page 334

# Added in version 2019.3.1

## Connect to Data

### Connect to data using Google BigQuery

Connect to even more of your data in the cloud using Google BigQuery. Like Tableau Desktop, you can now connect to data stored in Google BigQuery and publish flows to Tableau Server and Tableau Online if Google OAuth connections are configured.

**Note:** Tableau Prep Builder doesn't currently support using Google BigQuery customization attributes. Also flows that include Google BigQuery connectors can't be run from the command line. Instead run the flow manually or run the flow on a schedule in Tableau Server or Tableau Online.

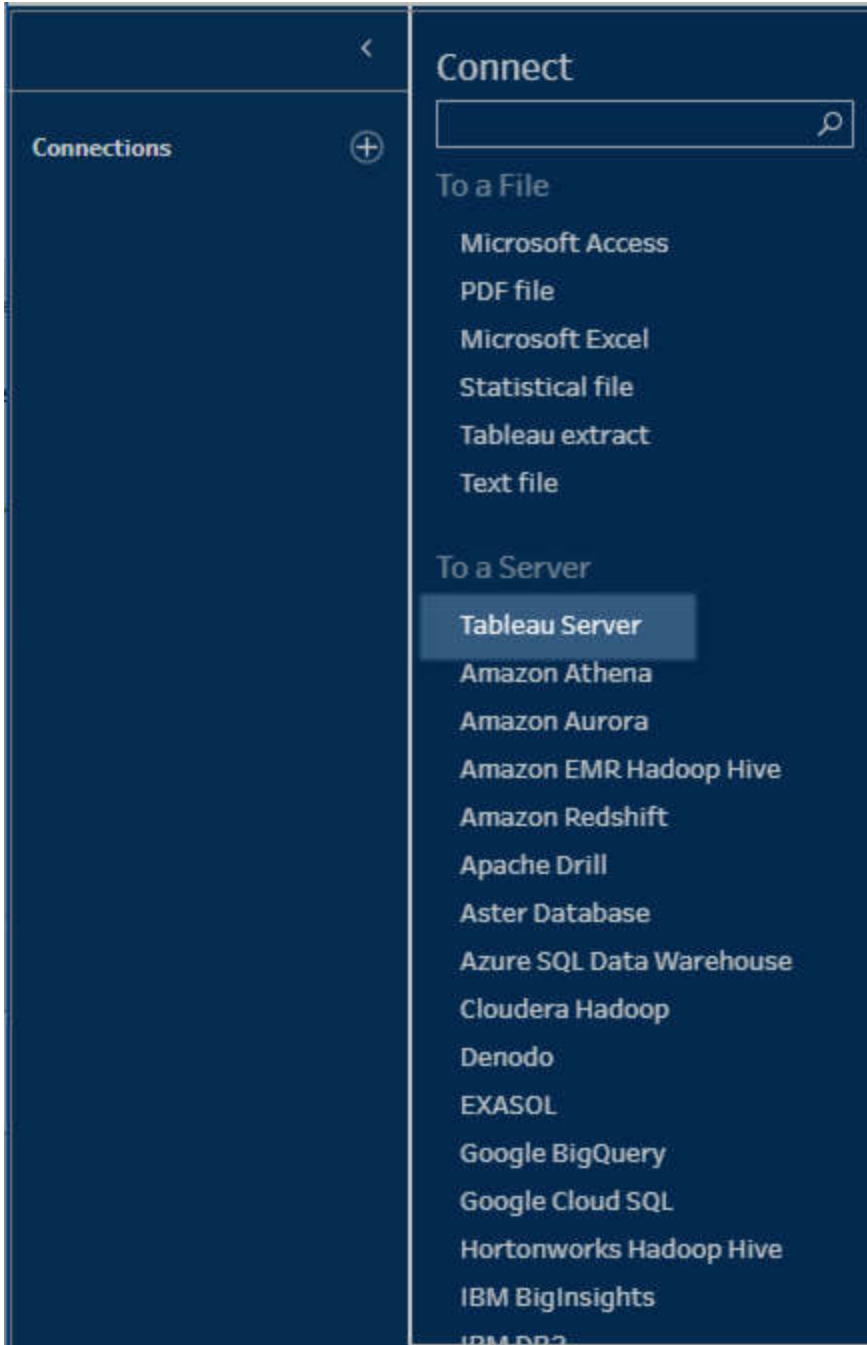
For more information about how to connect to Google BigQuery, see [Google BigQuery](#) in the Tableau Desktop help. For more information about connecting to data, see [Connect to Data](#) on page 145.

## Connect to published data sources for flow input

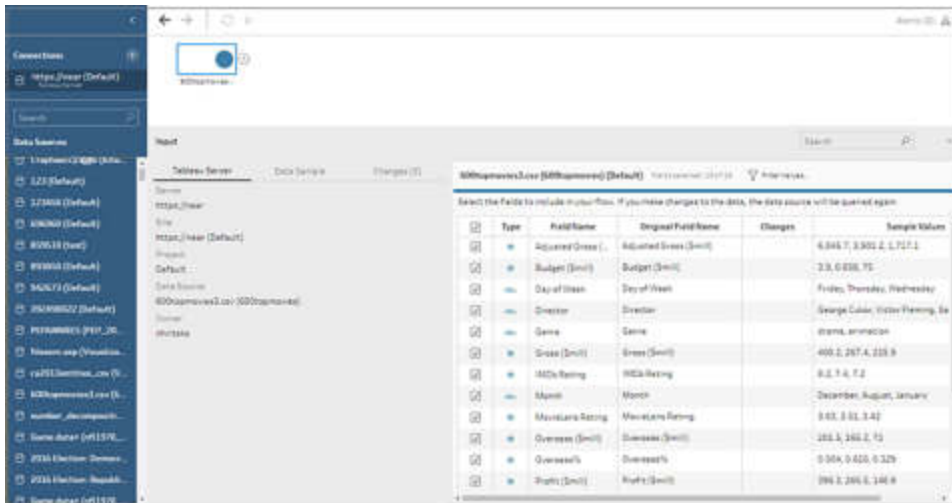
Whether you generate a published data source from Tableau Desktop or as output from Tableau Prep Builder and publish it to Tableau Server or Tableau Online, you can now connect to your published data sources and use them as an input data source for your flow.

Simply select **Tableau Server** from the **Connect** pane, then sign in to your server or site.





After you sign in, a list of available data sources are shown. Select your data source, or use the search option to find your data source and drag it to the flow pane to start your flow.



**Note:** Published data sources that include user filters aren't currently supported.

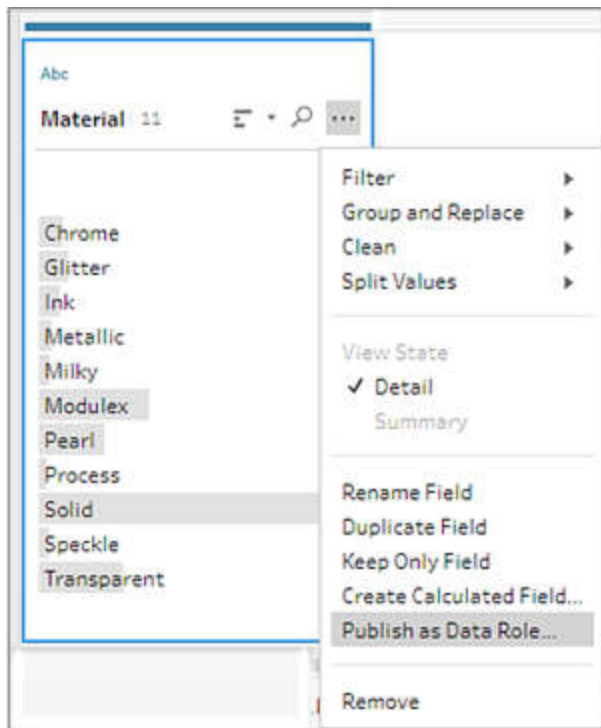
For more information about connecting to data and building a flow, see [Connect to Data on page 145](#).

## Examine and Filter Your Data

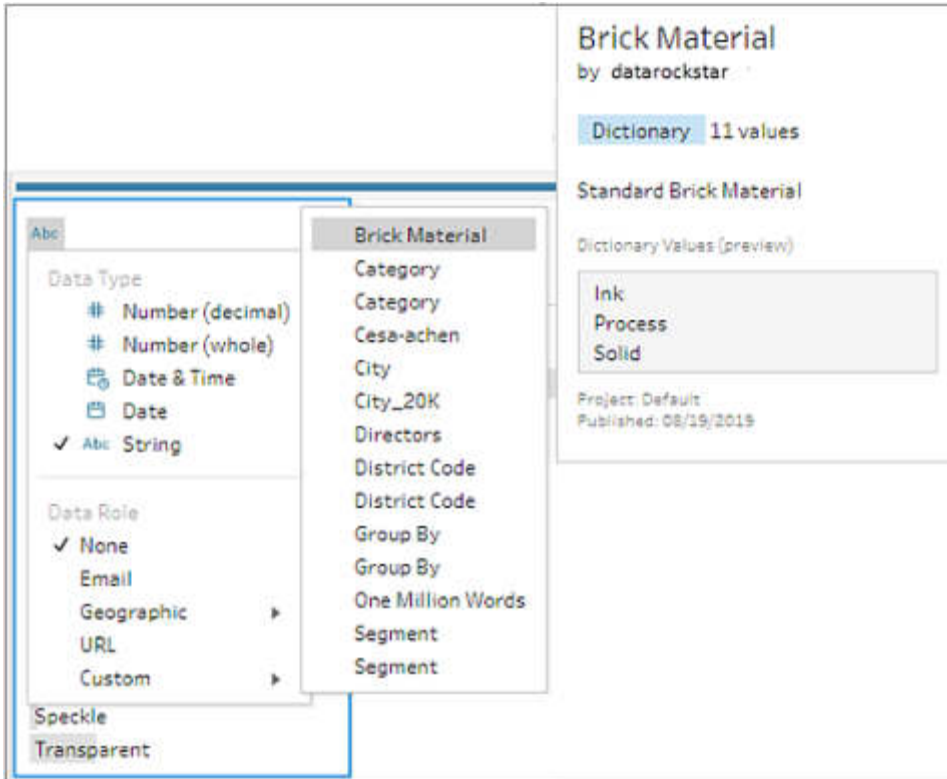
### Create custom data roles

You can use data roles to identify and validate field values. In prior releases, Tableau Prep Builder provided a set of standard data roles that you could apply to your fields to quickly identify invalid values. In this release, you can now create your own custom data roles using the field values in your data sets, then publish them to Tableau Server or Tableau Online to use them in your flow or share your data roles with others to make identifying and cleaning invalid values easier than ever.

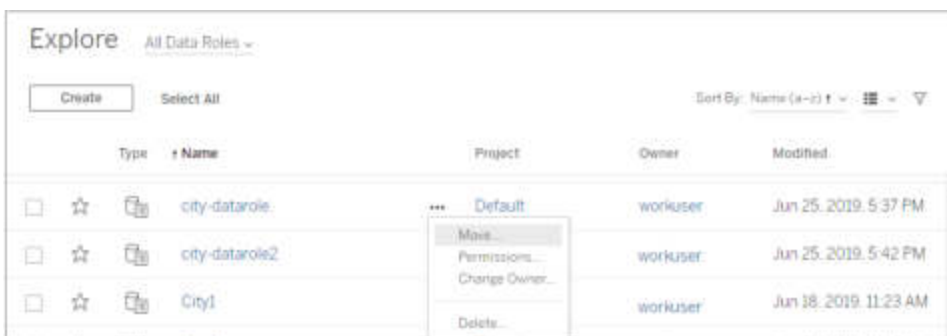
Simply select the field that you want to use and select **Publish as Data Role** from the **More options**  $\dots$  menu.



After the custom data role is published, once you sign into the site or server where the custom data role is published, you can select it from the new **Custom** menu option to apply it to your fields.



You can view and manage your published custom data roles on Tableau Server and Tableau Online. View all custom data roles published to your site or server, move them to a different project, change permissions or delete them.

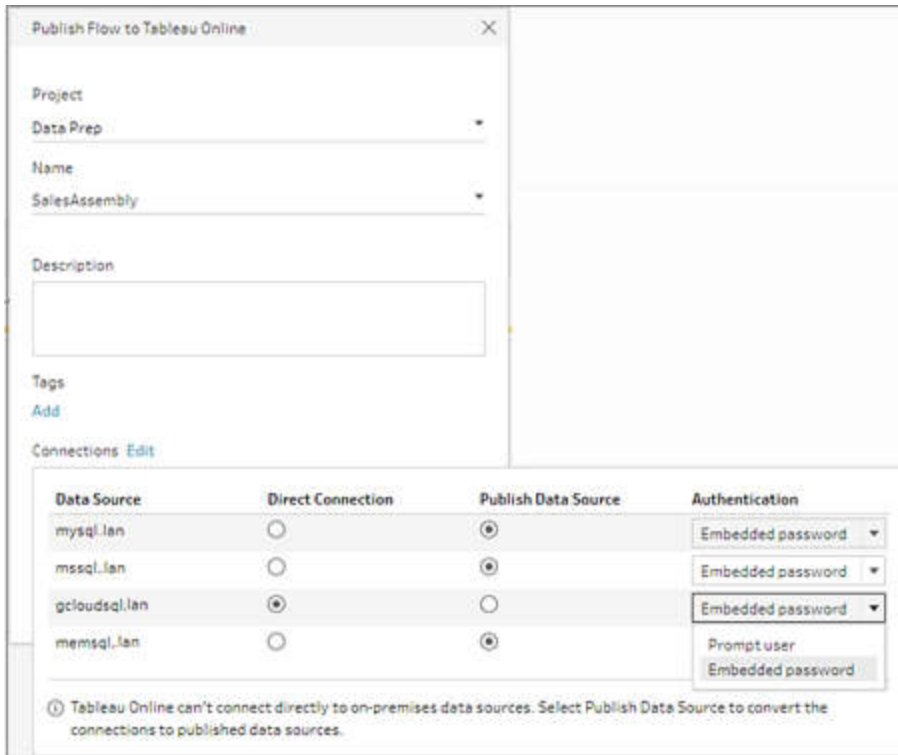


For more information about creating and using custom data roles see [Create custom data roles \(version 2019.3.1 and later\)](#) on page 217.

## Clean and Shape Data





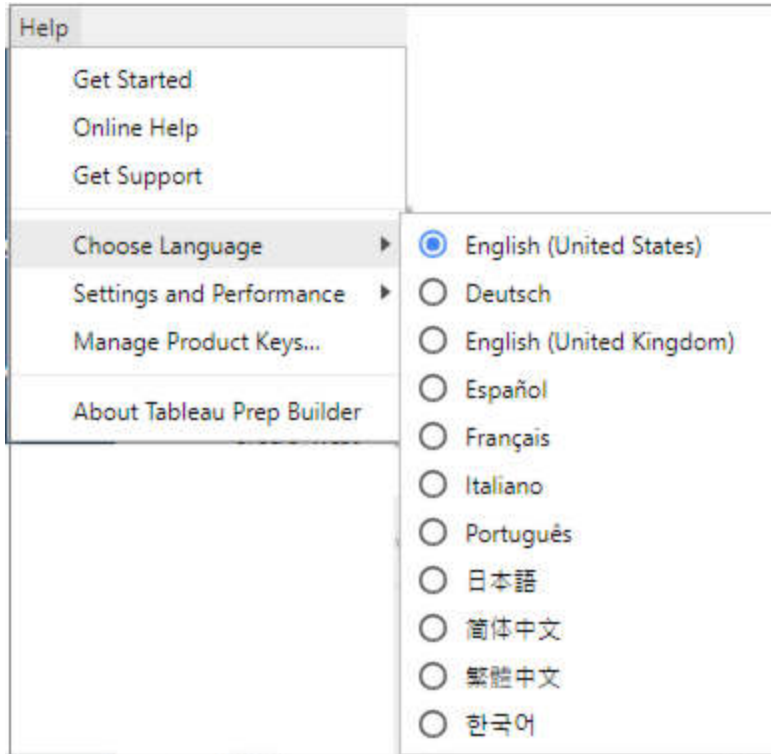


For more information about publishing flows to Tableau Online, see [Publish a Flow to Tableau Server or Tableau Online](#) on page 398.

## Additional Enhancements

### Tableau Prep Builder now in Italian

You can now configure your user interface (menus, messages, etc.) in one of 11 different languages. For more information about configuring your display language, see [Set your display language](#) in the Tableau Desktop and Tableau Prep Builder deployment guide



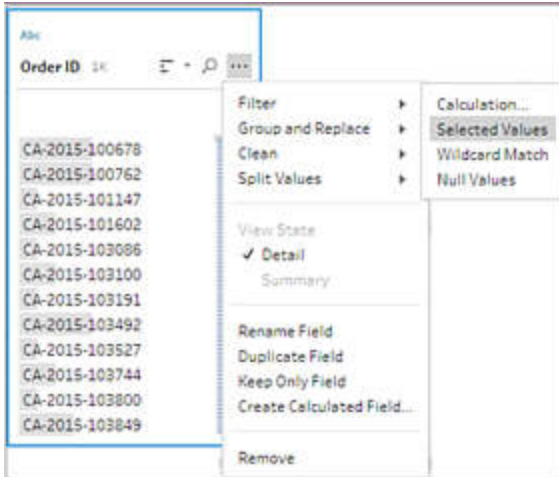
## Added in version 2019.2.3

### Examine and Filter Your Data

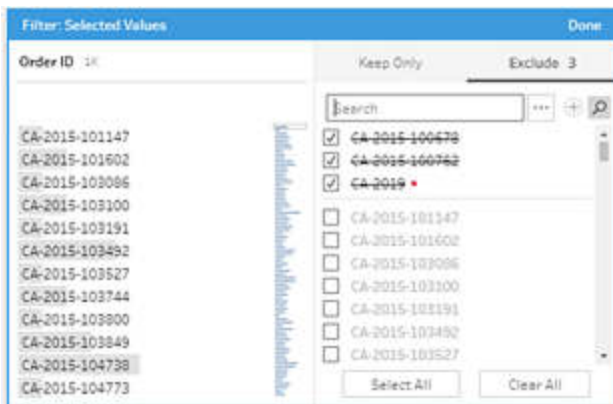
#### Select specific values to filter

With the new **Selected Values** filter, you can now pick and choose the values that you want to keep or exclude for a field, even values that aren't in your sample. In the Profile pane, Results pane or data grid click the **More options** ... button then select **Filter > Selected Values**.





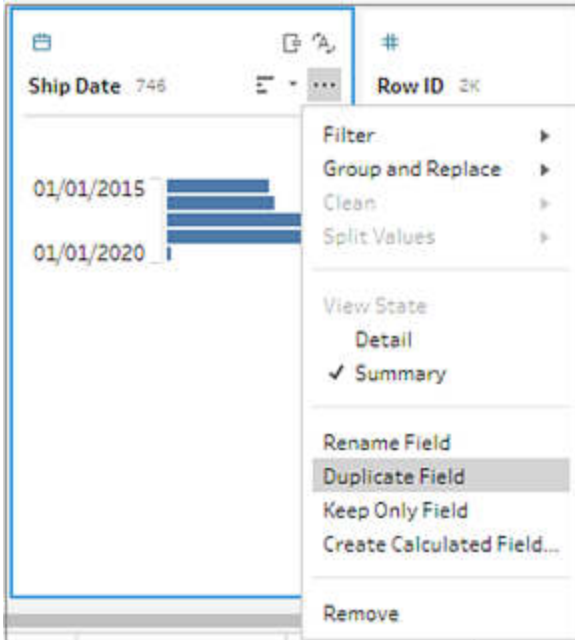
This new visual indicator lets you select the values that you want to keep or exclude and visually shows you what you've selected. If you are working with a sample of your data and want to include or exclude values that aren't in the sample you are working with, simply add the values you need and Tableau Prep Builder applies the keep or exclude option to these new values too.



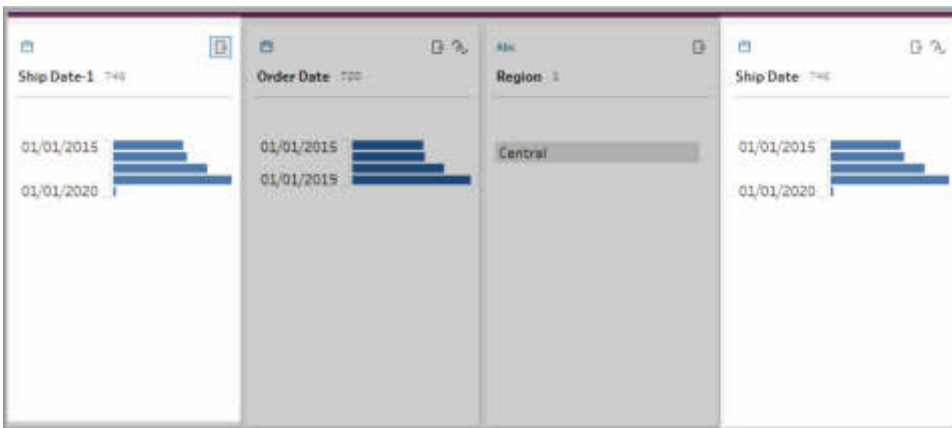
For more information about filtering field values, see [Examine Your Data](#) on page 198.

## Duplicate fields in one click

In prior releases if you wanted to experiment with your cleaning operations on a field but didn't want to change the original data, you would have to write a calculation to duplicate the field. Now you can easily duplicate your fields in one click. Select the field you want to duplicate, then from the **More options** menu, select **Duplicate Field**.



A new field is created and placed in the far left of the Profile pane, data grid, or Results pane with the same name and a modifier. For example, "Ship Date -1".

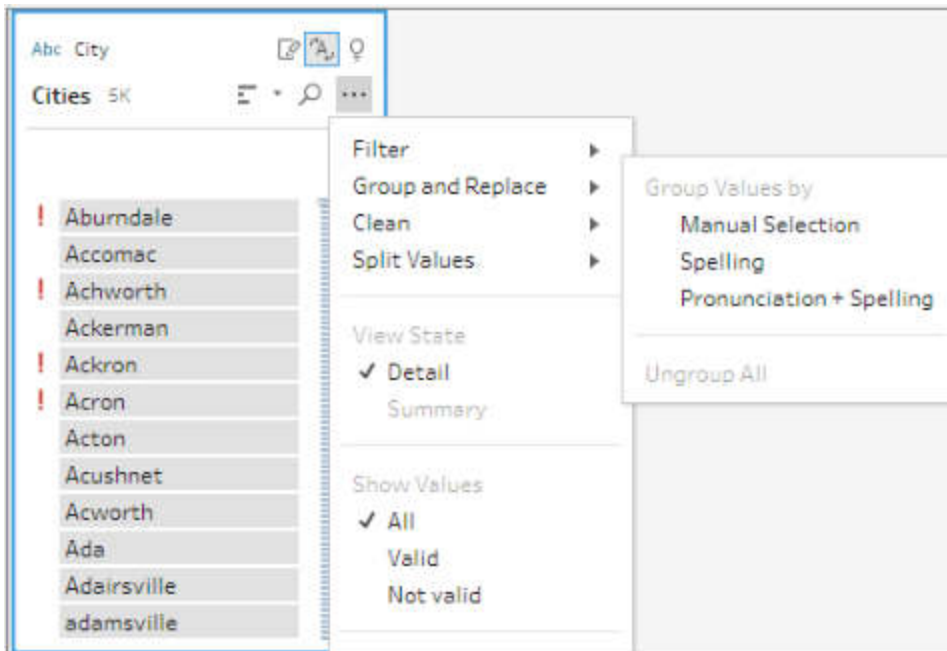


For more information about duplicating fields and other cleaning options, see [Clean and Shape Data](#) on page 229 [Apply cleaning operations](#) on page 233


## Clean and Shape Data

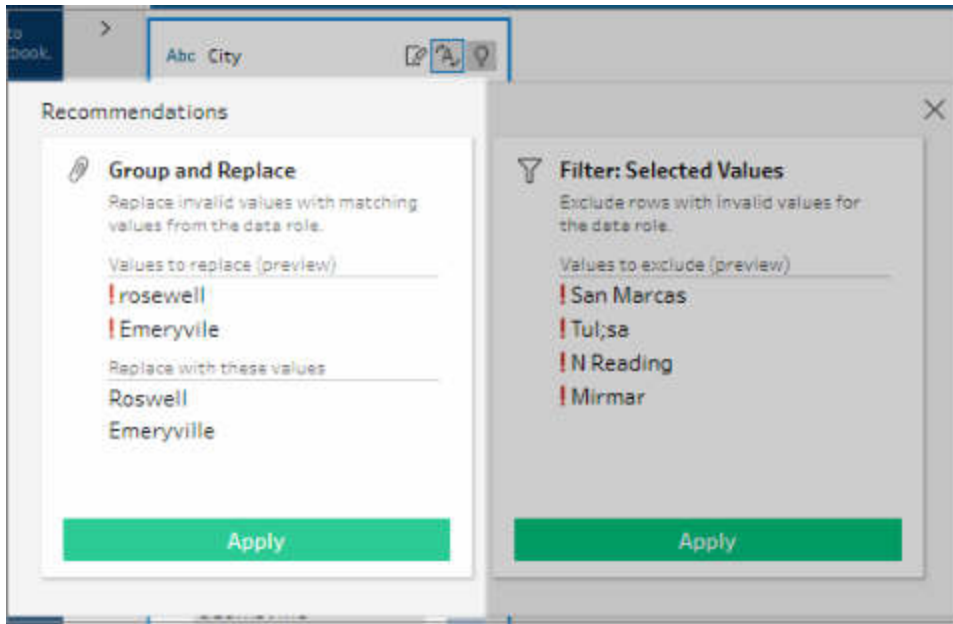
## Fix invalid data role values with fuzzy matching and cleaning recommendations

You can use data roles to quickly identify field values that aren't valid. After you assign the data role you can use one of two fuzzy matching algorithms; **Spelling** and **Spelling + Pronunciation** to quickly group and match invalid values for the data role to similar standardized values for that data role.



In addition to suggesting data roles for your fields, Tableau Prep Builder can now also suggest how to automatically fix invalid values for fields assigned to a data role using the **Spelling** and

**Spelling + Pronunciation** algorithms. If a suggested fix is detected, a light bulb  icon appears on the profile card. Click the icon and apply the suggested change to automatically group and replace invalid values with similar standardized values for that data role.



This feature is only available for the following data roles:

- Airport
- City
- Country/Region
- County
- State/Province

For more information about applying data roles to your fields and using group and replace to fix invalid values for data roles, see [Use Data Roles to Validate your Data](#) on page 213.

## Added in version 2019.2.2

### Connect to Data

#### Connect to more data with the Other Databases (ODBC) connector

If you need to connect to data sources that aren't listed in the **Connections** pane, you can connect to any data source using the **Other Databases (ODBC)** connector that supports the SQL standard and implements the ODBC API. Connecting to data using the **Other Databases (ODBC)** connector works similarly to how you might use it in Tableau Desktop, however there are a few differences:

- You can only connect using the DSN (data source name ) option.
- To publish and run your flow in Tableau Server, the server must be configured using a matching DSN.
- There is a single connection experience for both Windows and MacOS. Prompting for connection attributes for ODBC drivers (Windows) isn't supported.
- Only 64-bit drivers are supported by Tableau Prep Builder.

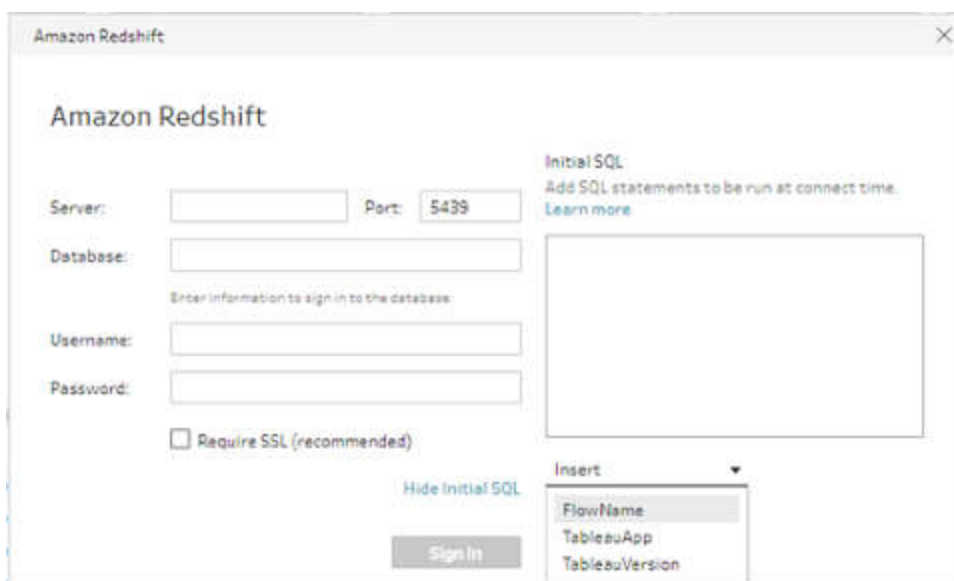
For more information about connecting to data using the Other Databases ODBC connector, see [Connect to Data on page 145](#).

## Specify an Initial SQL command when connecting to your data

You can now specify an Initial SQL command, when connecting to databases that support it, that will run when a connection is made to the database. For example when connecting to Amazon Redshift, you can enter a SQL statement to apply a filter when connecting to the database just like adding filters in the Input step. The SQL command will apply before data is sampled and loaded into Tableau Prep Builder.

If you change the Initial SQL command and refresh the Input step, the Initial SQL command is run again and the Input step is updated. When you run the flow, the initial SQL command is run before processing all of the data and if you schedule flows on Tableau Server, the initial SQL is run every time as part of the data loading experience.

For more information, see [Connect to Data on page 145](#).



**Note:** Parameters are not yet supported for Initial SQL in Tableau Prep Builder.

## Examine and Filter Your Data

### Keep only the fields you need while cleaning your data

In prior releases you could remove unwanted fields in the Input step by clearing the check boxes after connecting to your data. In a cleaning or action step you could select one or more fields and simply right-click to remove them. Now, you can select one or more fields in the Profile pane or data grid and right-click to keep only the selected fields and all other fields are removed.

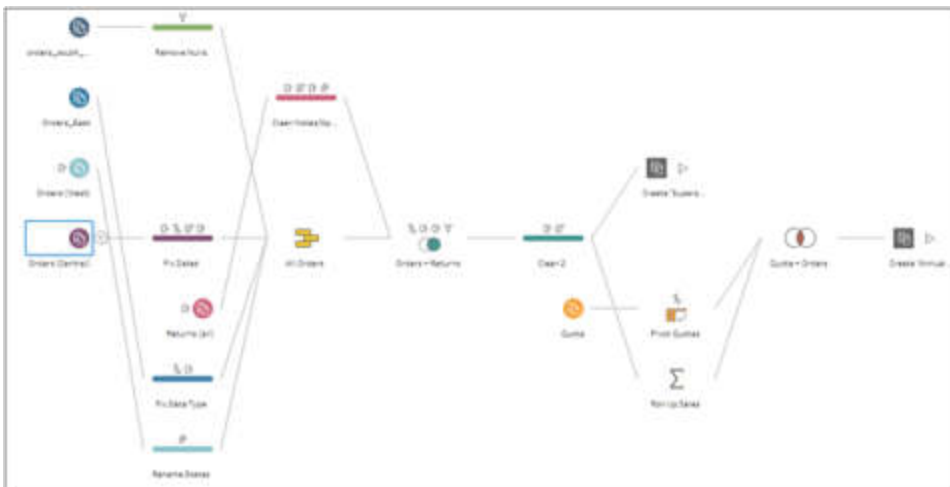
For more information, see [Keep or remove fields](#) on page 208.

## Clean and Shape Data

### Rearrange steps in your flow

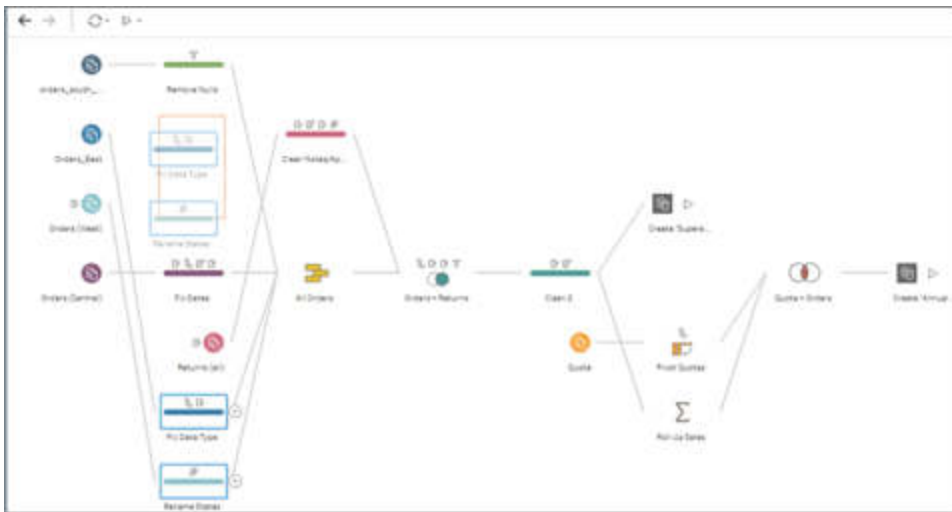
Have you ever built a large complicated flow, but the default layout that Tableau Prep Builder uses makes your flow hard to follow? Well now you can select and move your steps in the canvas to organize your flow in a way that makes sense to you. For example, you can fix crossed flow lines, move your flow steps to clean up extra white space, or rearrange your flow steps to show a clear sequence of events.

The following flow is confusing and hard to follow:



## Tableau Prep Help

Using this new feature, you can select one or more steps in the flow and drag them up, down, left or right and drop them right where you want them.



**Note:** Some restrictions do apply and flow steps can't be moved to a position that disrupts the left-to-right process flow. When dragging flow steps to an allowed location, an orange box displays. If the location isn't allowed, no orange box displays and the steps return to their original location when you try to drop them.

For more information , see [Build and Organize your Flow](#) on page 182.

# Added in version 2019.2.1

## Connect to Data

### Connect to more data with new database connectors

We've added more connectors to help you connect to and clean your data. Now you can connect to data stored in the following databases:

- Amazon Athena
- Other Databases (JDBC)

For more information about connecting to specific connectors, see [Supported Connectors](#) in the Tableau Desktop and Web Authoring help. For more information about connecting to data to build a flow see [Connect to Data on page 145](#).

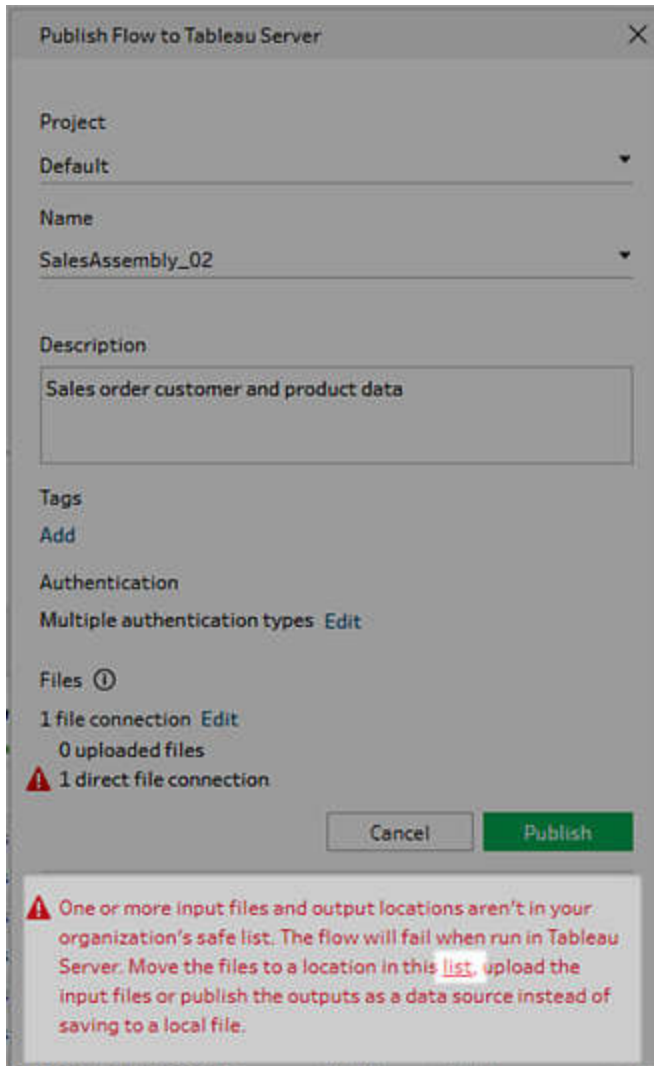
## Keep Flow Data Fresh

### View file safe list locations when publishing a flow

You can publish a flow to Tableau Server to automatically run your flows on a schedule and refresh the flow input or output using Tableau Prep Conductor. If your flow includes input or output steps with connections to a network share, your system administrator must include the file locations in your organization's safe list ahead of time, otherwise when you try to run the flow in Tableau Server, it will fail.

In Tableau Prep Builder version 2019.2.1 and later, if you publish a flow that includes input or output steps with connections to a network share, you are notified at publishing time if the files need to be included in your organization's safe list. To see safe listed locations, simply click the link in the warning message and move your files to one of the locations shown.





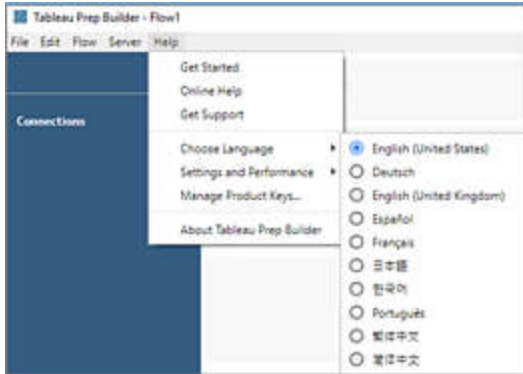
For more information about publishing flows, see [Publish a Flow to Tableau Server or Tableau Online on page 398](#). For more information about how to add file locations to your organization's safe list, see [Step 4: Safe list Input and Output locations on page 393](#).

## Other enhancements

### Select from 10 different languages for your user interface

Tableau has added two additional languages that you can choose from to configure your user interface (menus, messages, etc.):

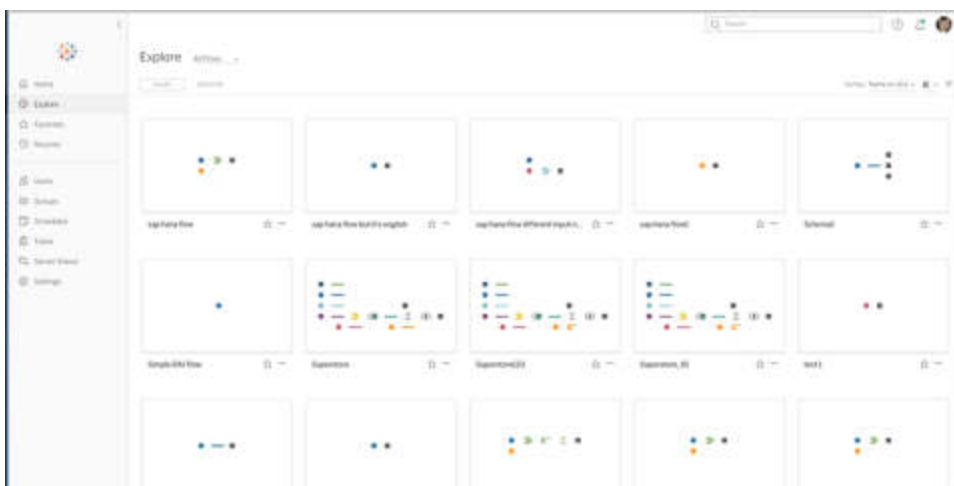
- en\_GB (British English)
- zh\_TW (Traditional Chinese)



For more information about configuring your display language, see [Set your display language](#) in the Tableau Desktop and Tableau Prep Builder deployment guide.

## Quickly find your flow on Tableau server with flow thumbnail views

When you publish your flows to Tableau Server, you can now switch to the thumbnail view and view a graphic representation of your flow layout to help you quickly spot your flow at a glance.



# Added in version 2019.1.4

## Connect to Data

### Connect to more data with new database connectors

We've added even more connectors to help you connect to and clean your data. Now you can connect to data stored in the following databases:

- SAP HANA
- Azure SQL Data Warehouse
- MariaDB

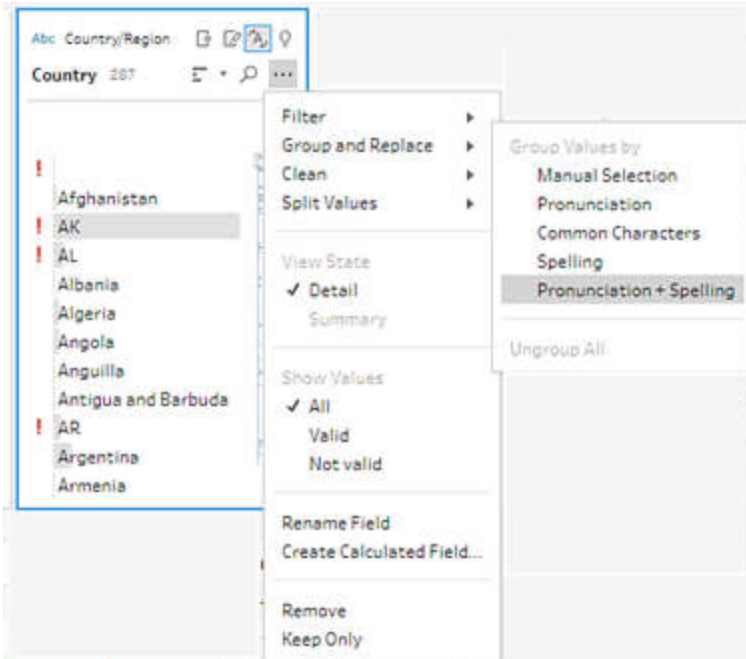
**Note:** When connecting to data using SAP HANA, prompting for variables and parameters when opening a flow and using Initial SQL to query the connection isn't supported.

For information about connecting to specific connectors, see connection details in the [Supported Connectors](#) section of the Tableau Desktop and Web Authoring help. For more information about connecting to data to build a flow see [Connect to Data on page 145](#).

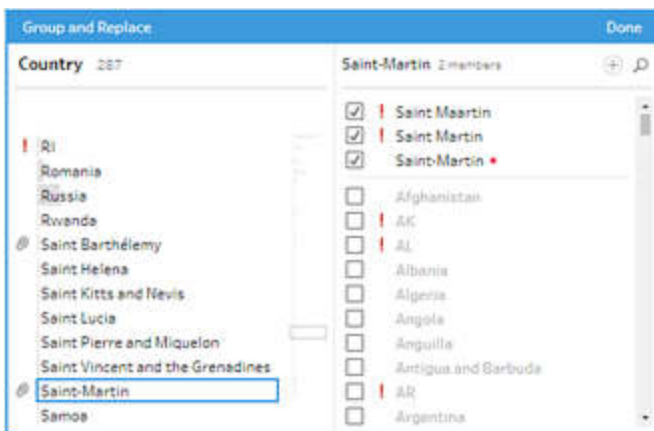
## Clean and Shape Data

### Group similar values in a field by data role

In Tableau Prep Builder version 2018.2.3 we introduced data roles to help you standardize field values and quickly find values that weren't valid. Now if you assign a data role to your fields, you can use that data role to match and group values with the standard value defined by your data role. Simply assign a data role to your field, then select **Data Role Matches** from the **Group and Replace** menu. This new option then matches invalid values to the most similar valid value based on spelling and pronunciation.



For example if you have a field that lists countries/regions such as Saint Martin and Saint Maarten, and you assign the **Country/Region** data role to the field, you can group fields using the new **Data Role Matches** option and map these values to the standardized data role value "Saint-Martin". If the standardized spelling of the value doesn't exist in your data set, Tableau Prep Builder automatically adds it to the group for you and groups the matching values using this standard value.



This feature applies to the following data roles and is currently only supported for U.S.English.

- Airport
- City

## Tableau Prep Help

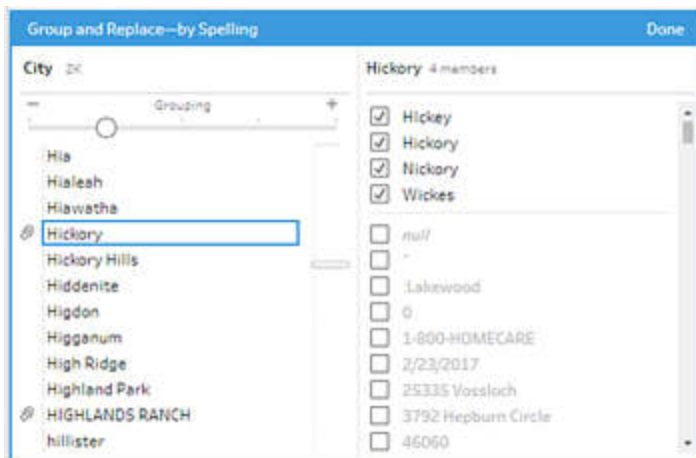
- Country/Region
- County
- State/Province

For more information about assigning data roles to a field or using data roles to group values, see [Use Data Roles to Validate your Data](#) on page 213.

### Adjust your results when grouping field values

When you use the **Group and Replace** cleaning option to group similar values by **Spelling** or **Pronunciation**, you can now use a simple slider on the field to adjust how strict the grouping parameters are, giving you more control over your results.

Select from 5 threshold settings to control the number of values included in a group and the number of groups that get created.



For more information about grouping fields using fuzzy matching, see [Edit field values](#) on page 246.

# Added in version 2019.1.3

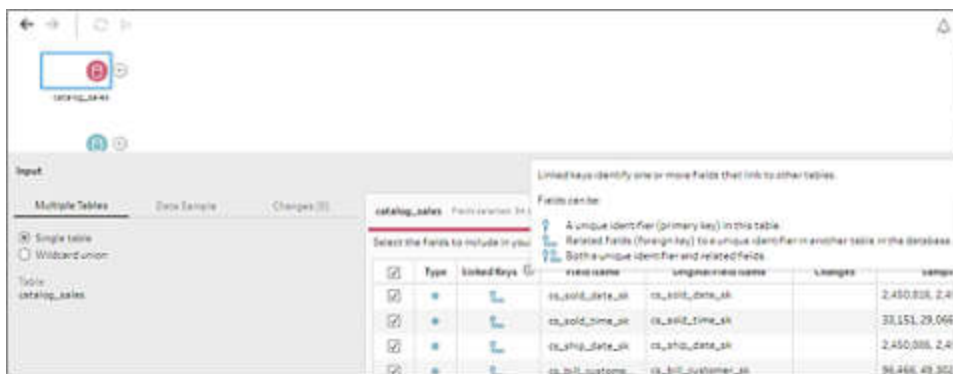
## Connect to Data

### Quickly join your data based on table relationships

When you connect to databases that include tables where table relationships are defined, such as Vertica or Microsoft SQL Server, Tableau Prep Builder can now detect and show this information.

For example, you can now see which fields in a table are unique identifiers and which fields are related fields (foreign key) as well any related tables for these fields.

A new column called **Linked Keys** shows in the Input pane with an icon next to the fields that are unique identifiers, related fields or a combination of both.

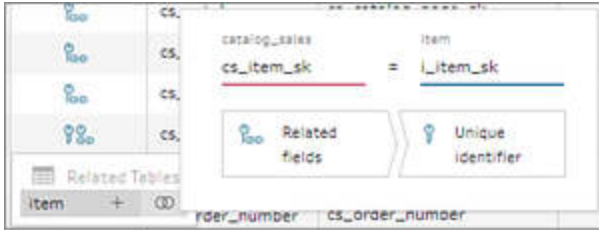


In the **Linked Keys** field, click on a related field or combination field to view related tables, then in one click, add the related table to your flow or create a join.

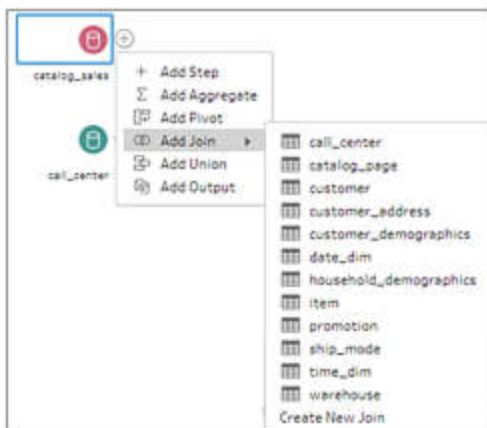


When creating a join, Tableau Prep Builder uses the defined field relationship to join your tables and shows a preview of the fields that define the join clause. From there you can click on the new Join step to clean and edit your join conditions as needed.

## Tableau Prep Help



For tables that contain related fields, you can also see a list of all related tables and create a join from the menu in the Flow pane. From the **Add Join** menu option, select the table that you want to join on from the list and Tableau Prep Builder creates the join based on the fields that make up the relationship between the two tables.




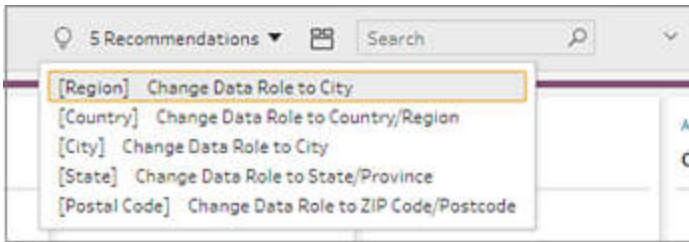
This feature is available for any supported database connector where table relationships are defined. For more information about using **Linked Keys** to add or join tables, see [Join data in the Input step \(version 2019.1.3 and later\)](#) on page 179.

## Clean and Shape Data

### See all cleaning recommendations in the toolbar

If Tableau Prep Builder detects and provides cleaning recommendations for your fields, a light

bulb  icon appears on the Profile card for the field. Now you can see all cleaning recommendations for any field from the toolbar in a selected step in one convenient list.



Simply click on an item in the list and Tableau Prep Builder takes you right to the field where you can apply the suggested change. For more information about cleaning recommendations, see [Apply cleaning operations using recommendations \(version 2018.3.2 and later\)](#) on page 243.

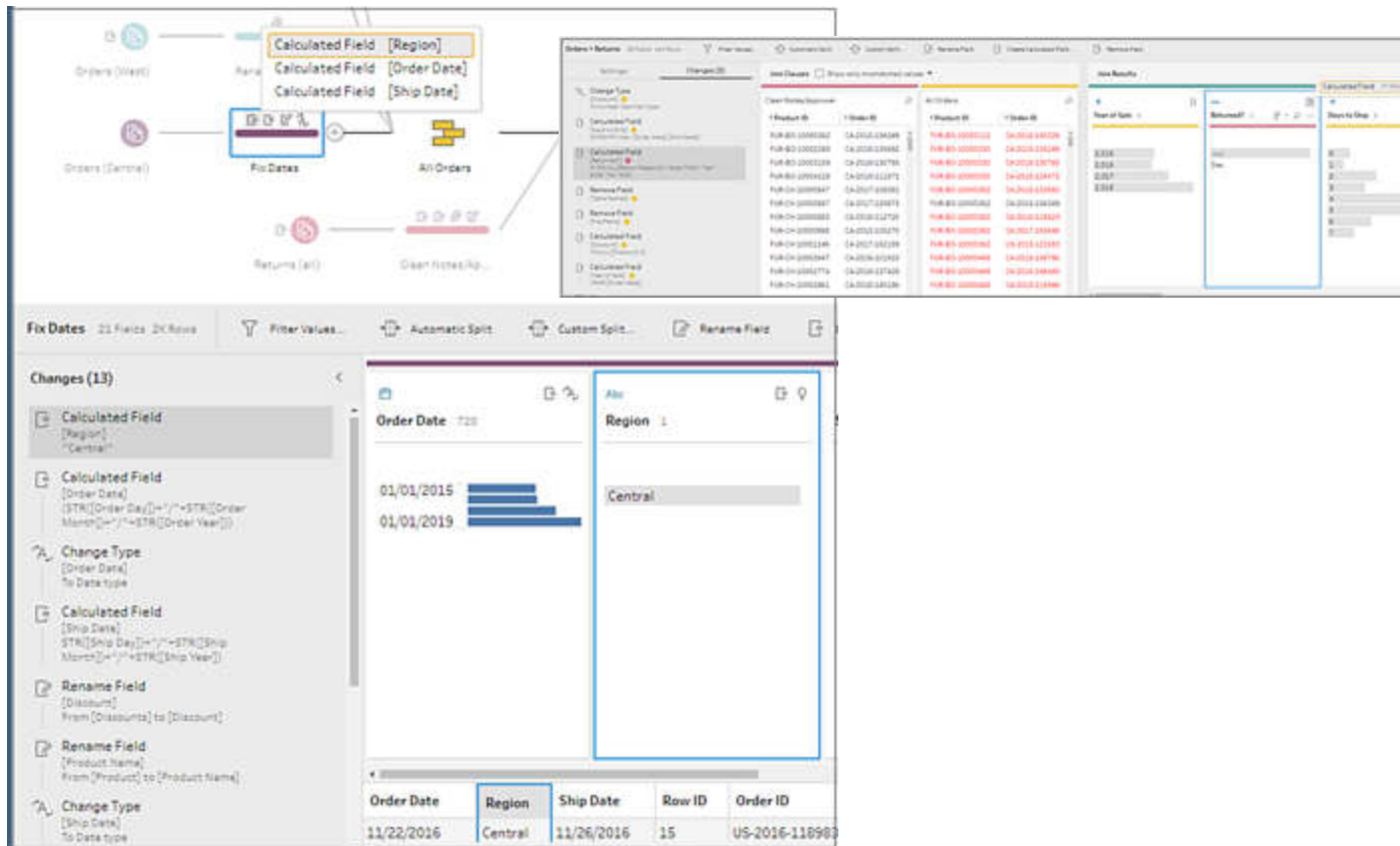
**Note:** This option isn't available for Join, Input and Output step types.

## Other enhancements

### Find specific changes in your flow

You can now quickly find specific changes that you made in a flow by clicking on the change icon on a step in the Flow pane or on a profile card in the Profile or Results pane. Simply click on the icon that represents the change you are looking for then select the annotation from the list of changes. The change is highlighted in the **Changes** pane, **Profile** or **Results** pane and data grid if applicable.



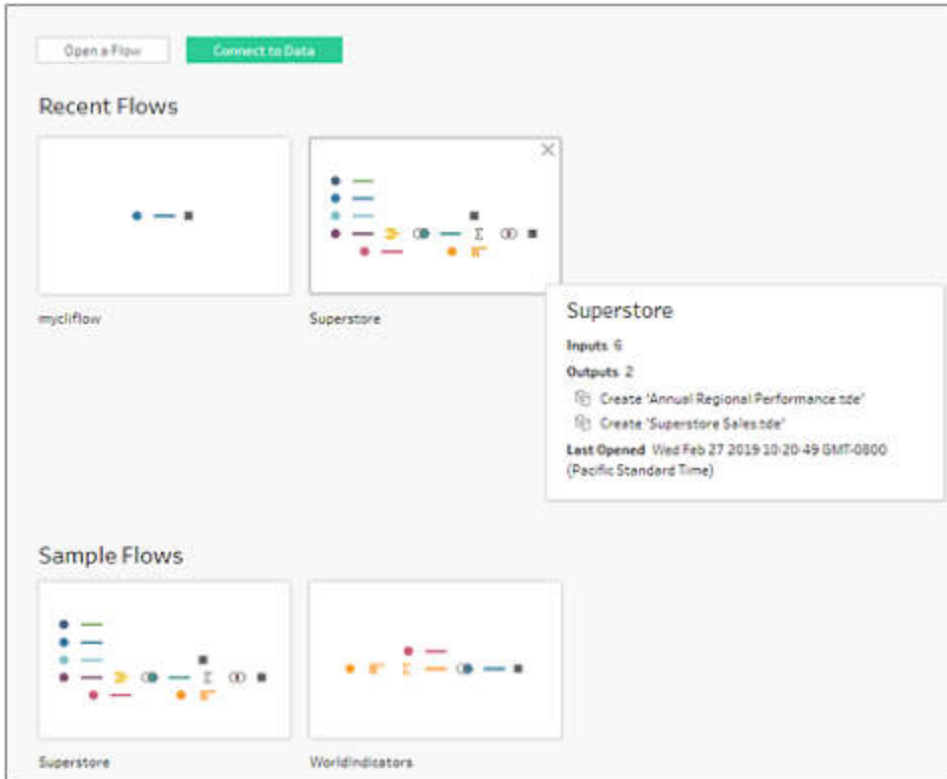


For more information about viewing your changes, see [View your changes](#) on page 239.

## Manage recently viewed flows from the Start page

When you save a flow and close Tableau Prep Builder, a thumbnail for the flow is saved to your Start page so you can quickly find your most used flows. But as a data prep rock star, your Start page can get crowded fast! To help you quickly find the flow you need and reduce the clutter when you're finished working with a flow, we've added a few new features to the Start page.

- Improved graphics on the flow thumbnail more accurately represents the flow layout to help you spot your flow at a glance.
- If the graphic doesn't give you enough information, hover over a flow thumbnail to instantly view the input and output details for the flow as well as when it was last opened.
- Clear the clutter by removing recently viewed flows from the start page that you no longer need. This option removes the flow from the **Recent Flows** view only and won't delete the flow.



## Added in version 2019.1.2

### Keep Flow Data Fresh

#### Schedule flows to run automatically in Tableau Server

Add the new Data Management Add-on to Tableau Server version 2019.1 and later and you can use Tableau Prep Conductor to schedule and manage your flow runs. Publish a flow to Tableau Server, then schedule tasks to run the flow and automatically update your flow inputs and outputs. Use the error and alert options and Administrative views to track and manage your flow processing health.

For more information, see [Keep Flow Data Fresh on page 360](#).

### Help Design

Updates have been made to help you navigate our content, including a new right-hand menu to quickly scan and select different topics on a page.

# Added in version 2019.1.1

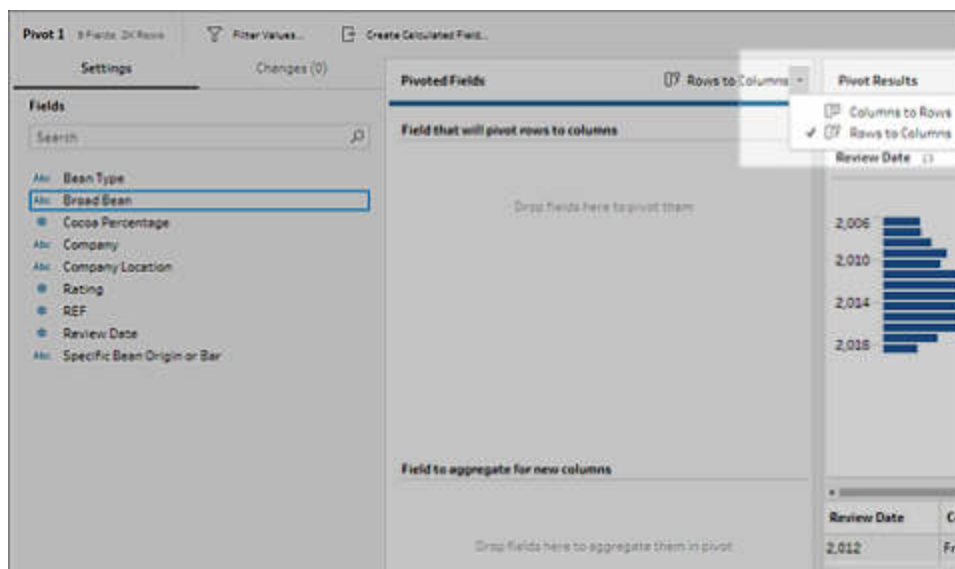
## Clean and Shape Data

### Pivot rows to columns

If you work with data in a cross-tab or spreadsheet it might not be in the format you need to easily analyze it in Tableau. Sometimes your data may have many columns of data that need to be converted into rows, or you may have fewer columns but the data values represent multiple things.

To make shaping your data easier than ever, you can now choose the pivot option that is right for you. In a pivot step, you can now select one of the following options:

- **Columns to Rows:** Drag one or more fields to the **Pivoted Fields** pane to pivot columns to rows.
- **Rows to Columns:** Drag a field to the **Pivoted Fields** pane to create new columns using the field values, then drag in the field to use to populate those columns and choose how you want to aggregate them.

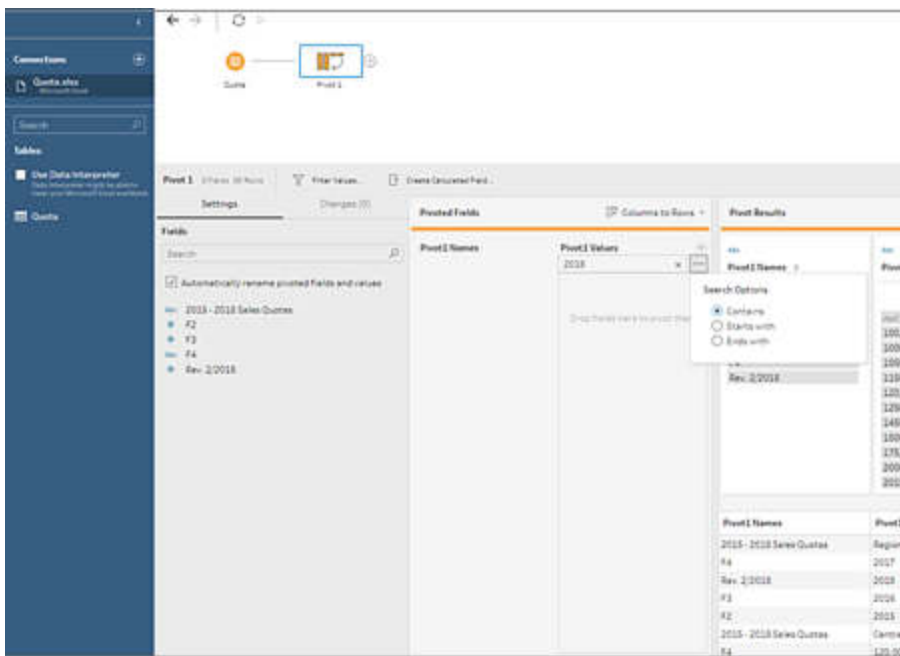


For more information, see [Pivot Your Data](#) on page 281

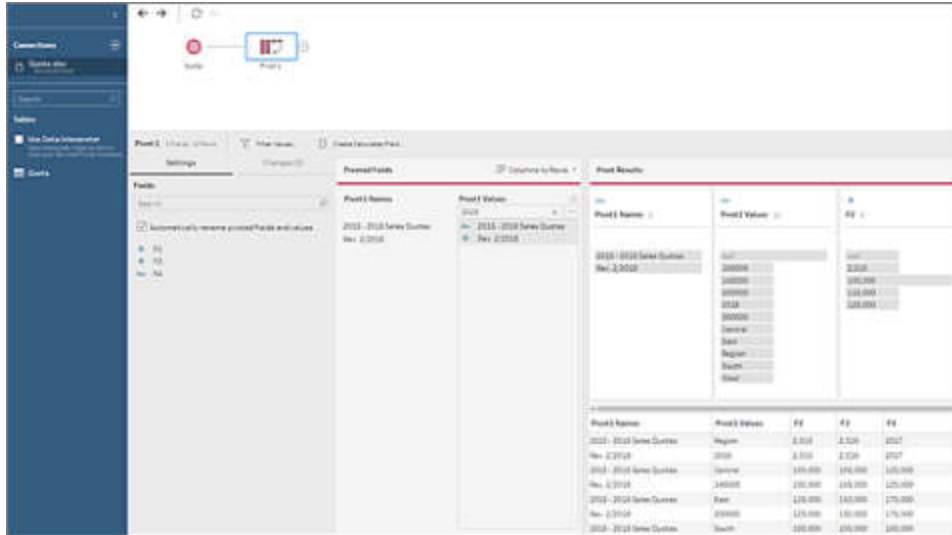
## Use wildcard search to pivot fields

If you have a data set with a large number of fields or if fields are frequently added or removed over time, you can use a wildcard search when pivoting columns to rows to instantly pivot your data based on a wildcard pattern match. If new fields are added or removed that match the pattern, Tableau Prep Builder detects the schema change when you run the flow and the pivot results are automatically updated.

To use wildcard search to pivot your data, add a pivot step then click the **Click here to create wildcard pivot** link to open the wildcard search field. Enter a wildcard search pattern to find fields with names that match the values you entered. Click the **Search Options...** button to select whether you want to include fields that contain, start with or end with the pattern you entered.



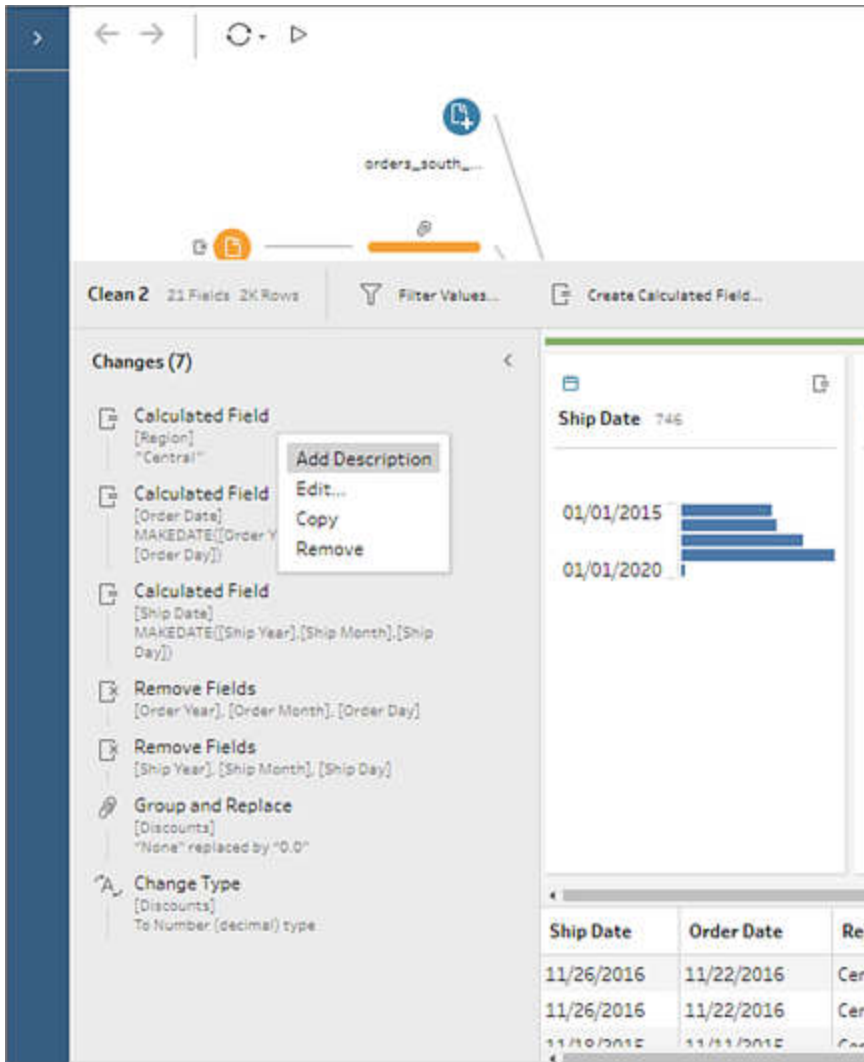
Click enter to apply the pattern and Tableau Prep Builder automatically pivots the fields.



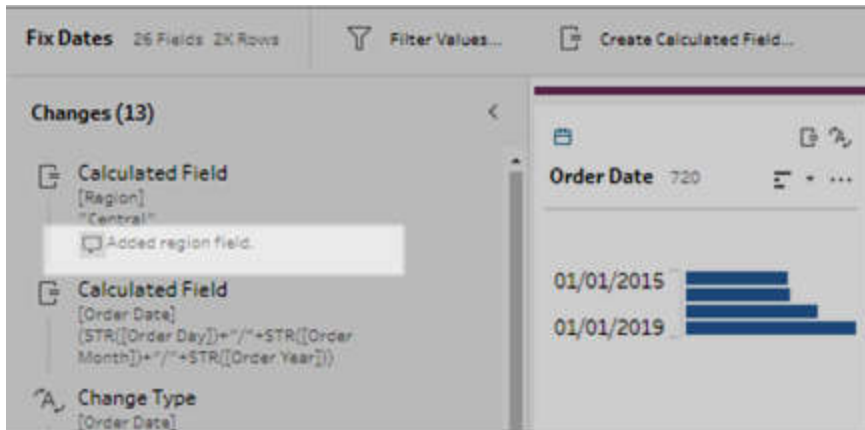
For more information, see [Pivot Your Data](#) on page 281

## Use descriptions to add context to changes

Add descriptions to any change in the **Changes** pane to add context to your cleaning operations. Simply open the **Changes** pane, right-click or Ctrl-click (Mac) on a change in the list and select **Add Description**. Then type a description in the text field that appears.



The description appears below the generated text for the change with a comment  icon.

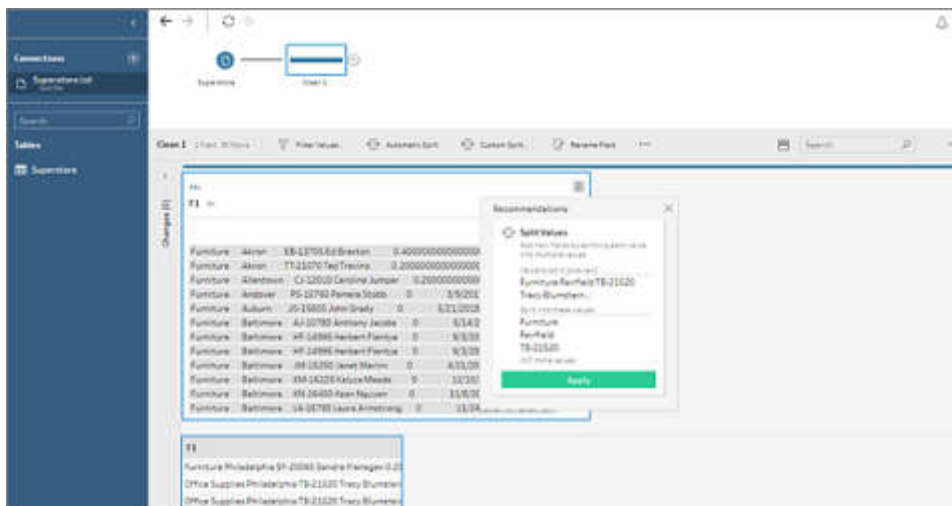


For more information see [Add descriptions to flow steps and cleaning actions on page 189](#).

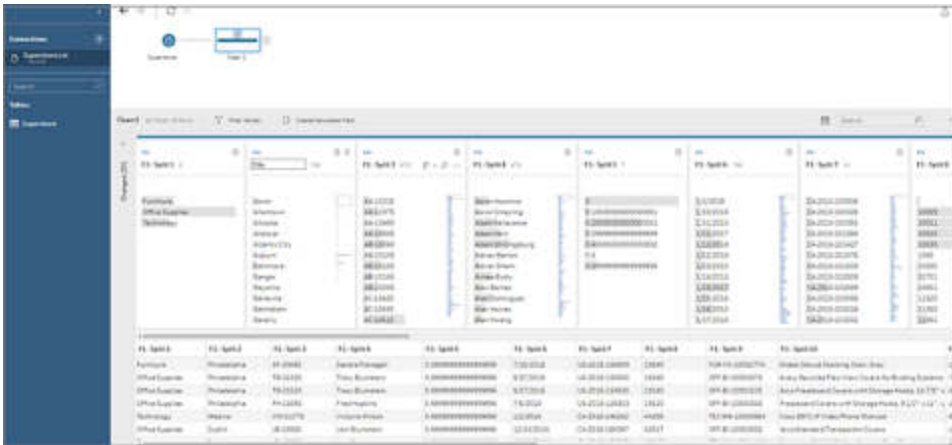
## Use cleaning recommendations to quickly split data in fixed-width text files

In prior releases if you work with data stored in fixed-width text files, files that contain strings of data values in a single column constrained by a fixed character limit, you might have had to perform some pre-cleaning on the file or use complex calculations to separate out the values into separate columns.

Now when you connect to this type of data source, the cleaning recommendations feature detects these types of columns and suggests split options for you. Simply connect to your data source then add a cleaning step. In the Profile pane, click on the recommendation icon on the profile card.



Click **Apply** to split the fields, then check the results and clean or rename each new field as needed.



For more information, see [Apply cleaning operations using recommendations \(version 2018.3.2 and later\)](#) on page 243.

## Added in version 2018.3.3

### Connect to Data

#### Automatically refresh input unions when new files are added

If you use a wildcard pattern to find and union Microsoft Excel or text (.csv) files in the input step, new files that are added to the same folder that match the pattern are automatically included in the union the next time you open the flow or run it from the command line.

**Note:** Packaged flow files (.tfx) won't automatically pick up new files because the files are already packaged with the flow. To include new files, open the flow file (.tfl) to pick up the new files then repackage the flow to include the new file data.

For more information about unioning files in the Input step, see [Union files and database tables in the Input step](#) on page 173.

### Clean and Shape Data



## Reuse cleaning operations

When cleaning your data you often perform the same cleaning operations or actions over and over throughout your flow. To help make cleaning and shaping your data more efficient, we've added the ability to copy and paste these operations or actions throughout your flow so you can perform a cleaning operation or action once, then reuse it where you need it.

- Copy one or more steps to use them in another area of your flow. This option is not available for Input steps that include a wildcard union.
- Copy an operation from the **Changes** pane in one step and paste it in the **Changes** pane for another step to apply that same operation in that step.
- Drag and drop an operation from the **Changes** pane to a field in the **Profile** pane to apply that operation to a different field. You can also copy an operation in the **Changes** pane and drag and drop it to other fields in the step to apply that operation to multiple fields. This option is not available for operations that impact multiple fields, such as calculated fields.

For more information, see [Clean and Shape Data on page 229](#).

## Remove extra whitespace

We've added two new quick cleaning operations to help you remove extra whitespace in your field values:

- **Remove Extra Spaces:** Use this option with text fields values to remove leading and trailing whitespace and replace extra whitespace in-between characters with a single space.
- **Remove All Spaces:** Use this option to remove all whitespace from text values, including leading and trailing whitespace and any whitespace in between characters.

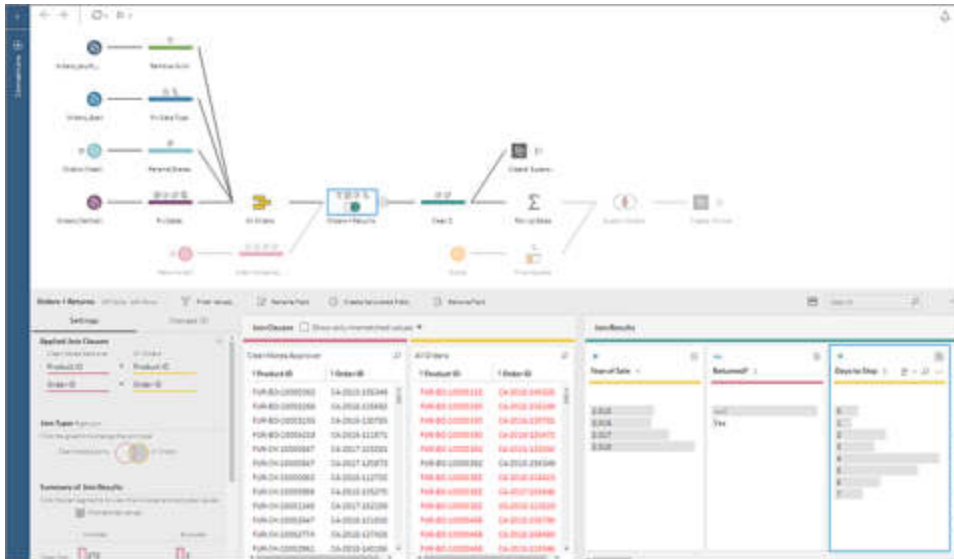
For more information, see [Edit field values on page 246](#).

## Examine and Filter Your Data

### Quickly find where fields are used in a flow

When building your flow, if you aren't seeing all the field values that you expect, you can now easily trace everywhere a field was used - even back to where a field originated. Just select a field in the **Profile** pane or **Results** pane in a step and the steps in the flow pane will highlight.

This shows you the path that field has taken in your flow, making troubleshooting your flow easier than ever.



For more information, see [Trace fields in a flow](#) on page 205.

## Other Enhancements

### Open multiple windows of Tableau Prep Builder at once

In Windows or on the Mac click **File>New** to open a new window for a flow instead of replacing the current window you are working in. When opening a flow file (.tfl or .tflx) this also opens a new window so you can easily work on or compare multiple flows at the same time.

### Turn off usage reporting after install

To help us improve Tableau Prep Builder, you can share usage pattern data with us. We never collect any specific user data, only anonymous usage data. For example data source connection types (such as .csv, sql or Amazon Redshift), large data sets versus small data sets, number of steps in a flow or types of operations used (but no data specifics such as step names or formulas used), and so on.

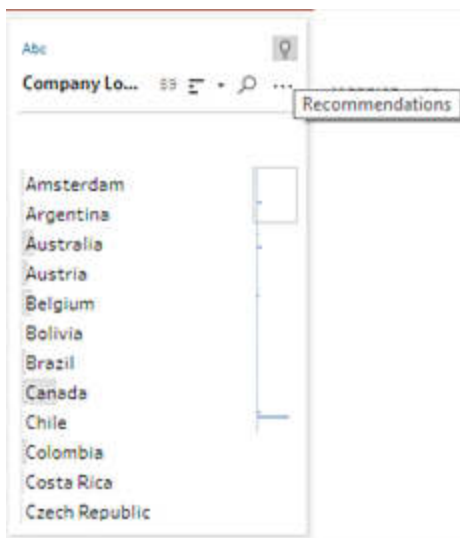
Your data is confidential and we will never share this information with anyone. But if you want to turn this option off after installation, you can. Go to **Help >Settings and Performance** and click on **Share Usage Data To Improve Tableau Prep Builder** to clear the check mark. For more information, see [Usage reporting](#) in the Tableau Desktop and Tableau Prep Builder Deployment guide.

## Added in version 2018.3.2

### Clean and Shape Data

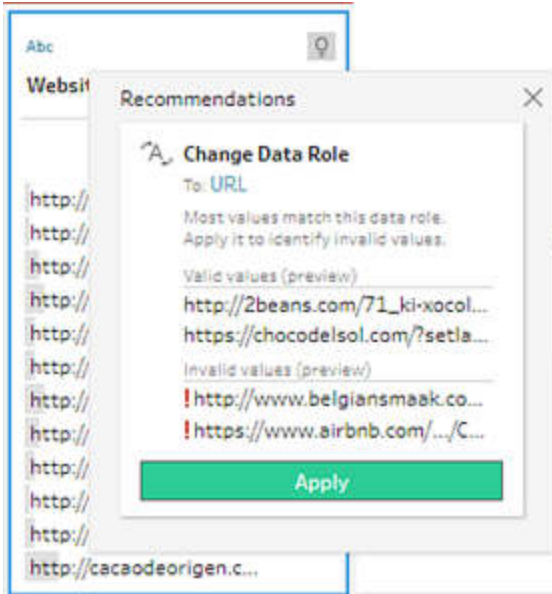
#### Improve data quality fast using cleaning recommendations

To help you find and fix problems with your data fast, Tableau Prep can now examine your data fields and suggest changes you can make to fix the problem. Tableau Prep examines your data fields and when a recommended change is identified, a light bulb icon appears on the Profile card for the field.



Click the light bulb icon to see the recommended changes, then hover over the recommendations card and click **Apply** to automatically apply them to the field. The change is recorded in the **Changes** pane and can be removed at any time.

In the example below Tableau Prep identified the field values as URL's and recommends applying a data role to the field to help identify values that aren't valid.



This feature is available in all step types except Input, Output and Join step types. In this release, recommendation types include:

- Data roles
- Filter
- Replace invalid values with Null values
- Remove fields
- Trim spaces

For more information, see [Apply cleaning operations using recommendations \(version 2018.3.2 and later\)](#) on page 243.

## Other Enhancements

### Added flexibility when running flows from the command line

When running flows from the command line, if you connect to databases or publish output files to a server, you need a credentials .json file that includes all required credentials. In past releases if you only connected to databases or published the flow output to a server, you would still need to include the empty array in the credentials .json file or the process would fail.

Now you no longer need to include the empty array in the credentials .json file. For example, if you only connect to databases you no longer need to include the outputConnection array in the .json file.

For more information about running flows from the command line, see [Refresh flow output files from the command line](#) on page 334.

## Performance improvements

To help make working with your flows faster and more efficient, we've made the following enhancements:

- Shutting down the product returns the memory space Tableau Prep was using right away.
- If you receive an unexpected error you can now retry your action or process to keep you in the flow.
- Improved performance when loading data or viewing data in a flow to reduce processing time.
- If you connect to the following connectors, you can now use a search field to quickly find your schema:
  - Amazon EMR Hadoop Hive
  - Cloudera Hadoop
  - Hortonworks Hadoop Hive
  - MapR Hadoop Hive
  - Presto
  - Spark SQL

## Added in previous versions

### Added in version 2018.3.1

#### Install and Deploy Tableau Prep

- [Change your default repository location](#) on the next page

#### Connect to Data

- [Clean data stored in PDF files](#) below
- [Wildcard union for database tables](#) on the facing page

### Clean and Shape Data

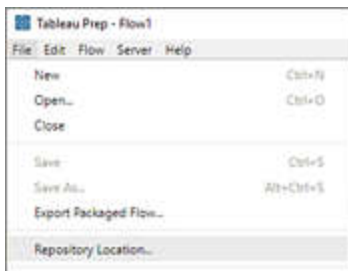
- [Replace field values with Null](#) on page 64
- [Apply cleaning operations in any step](#) on page 65

## Install and Deploy Tableau Prep

### Change your default repository location

By default, when you install Tableau Prep, the **My Tableau Prep Repository** is automatically installed in your Documents folder on your computer. But if you have restrictions about where you can store your flows and data sources, you can now move your My Tableau Repository to a different directory after Tableau Prep is installed.

To move your directory, open Tableau Prep, then from the top menu, select **File > Repository Location**. Select the folder where you want to move the repository.



For more information, see [Change your My Tableau Prep Repository location](#) in the Tableau Desktop and Tableau Prep Deployment Guide.

## Connect to Data

### Clean data stored in PDF files

Analyzing data stored in PDF files can be difficult to work with. Now you can connect directly to your PDF files and clean your data using Tableau Prep.

For more information about how to connect to your data in PDF files, see [PDF file](#) in the Tableau Desktop help.

**Note:** Data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

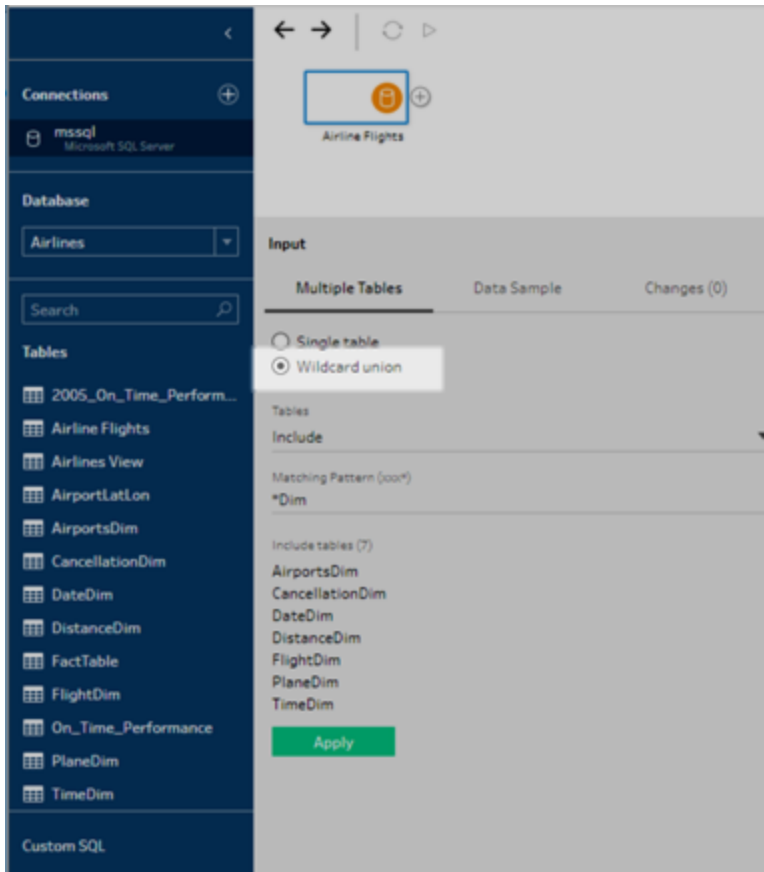
## Wildcard union for database tables

If you work with database tables stored in a single data source you can now union your tables in the input step using a wildcard search. You can then include all of the data from the selected tables in the Input step using the same process that you use to union multiple files in the Input step.

To use this type of union, the tables must be in the same data source and you must connect to a data source that supports wildcard union.

The following data sources support this type of union:

- Amazon Redshift
- Microsoft SQL Server
- MySQL
- Oracle
- PostgreSQL



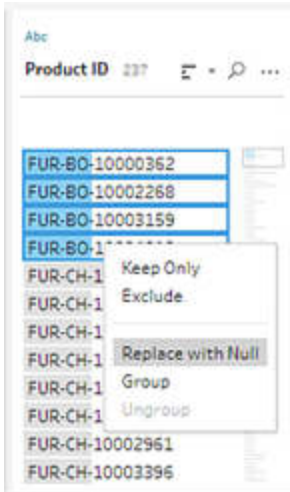
For more information, see [Union files and database tables in the Input step](#) on page 173.

## Clean and Shape Data

### Replace field values with Null

If you have data rows that you want to include in your analysis but you want to exclude certain field values you can change them to a Null value. Simply select the field values that you want to replace, then right-click (Ctrl+click on Mac) on the selected value and select **Replace with Null** from the menu.



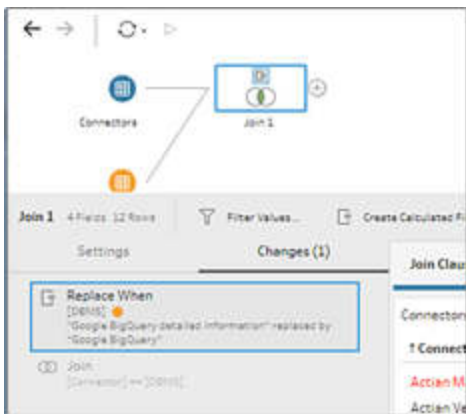


For more information about replacing field values, see [Edit field values](#) on page 246.

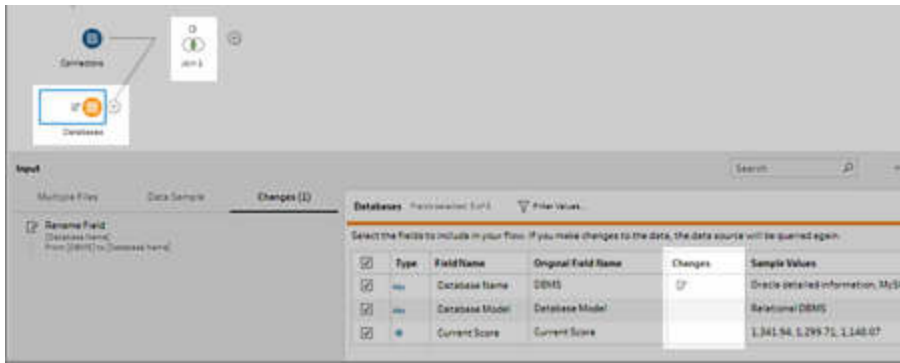
## Apply cleaning operations in any step

Fixing problems in your flow just got more efficient. Now if you spot problems in your data as you join, union, pivot or aggregate, you can fix these problems directly in that step without having to add an additional cleaning step.

Since the order of operations matters, the cleaning operation is placed either before or after the join, union, pivot or aggregate and this information is tracked in the **Changes** pane for that step.



Annotations for the change are also added to the step in the Flow pane. For Input steps, annotations show to the left of the step and are added to the Input field list so you can easily see which field had changes applied. For all other step types, the annotations show above the step in the Flow pane.



Some cleaning operations must still be performed in a cleaning step and only some cleaning operations are available for Input and Output steps. The table below lists the cleaning operations available for each step type:

	Input	Clean	Aggregate	Pivot	Join	Union	Output
Filter	X	X	X	X	X	X	
Group and Replace		X	X	X	X	X	
Clean		X	X	X	X	X	
Split Values		X	X	X	X	X	
Rename Field	X	X	X	X	X	X	
Edit Value		X	X	X	X	X	
Change Data Type	X	X	X	X	X	X	
Create Calculated Field		X	X	X	X	X	
Remove Field	X	X	X	X	X	X	

For more information, see [About cleaning operations](#) on page 229.

## Added in version 2018.2.3

### Connect to Data

- [Connect to data stored in MongoDB Business Intelligence \(BI\) below](#)

#### Examine and Filter Your Data

- [Specify a data role for your field values below](#)

#### Join or Union Data

- [Fix mismatched fields directly in the join clause on page 69](#)

## Connect to Data

### Connect to data stored in MongoDB Business Intelligence (BI)

If you store your data in Mongo DB Business Intelligence, you can now connect to your data and clean it with Tableau Prep.

For more information about how to connect to your data using Mongo DB Business Intelligence, see [MongoDB BI Connector](#) in the Tableau Desktop help.

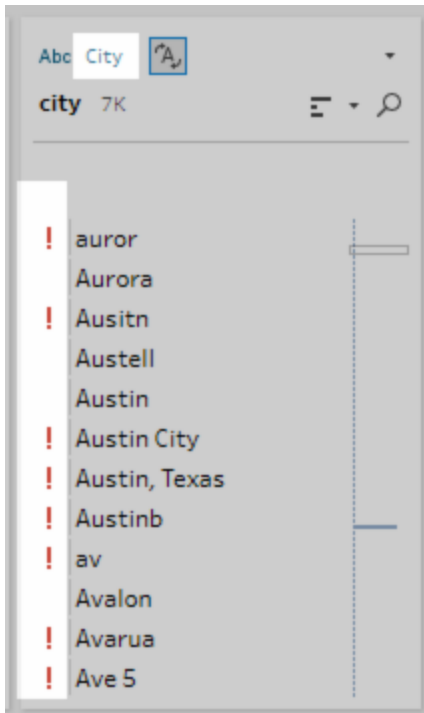
**Note:** Data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

## Examine and Filter Your Data

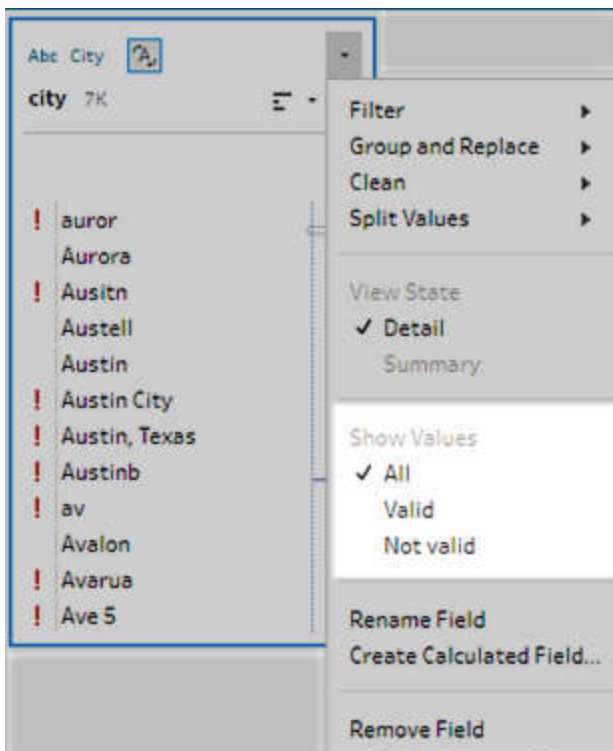
### Specify a data role for your field values

You can now assign a data role to your field values and use Tableau Prep to help you find inaccuracies or outliers in your data set.

Data roles tell Tableau Prep what the field values mean or represent, for example email addresses or a geographic role such as city or zip code. When a data role is assigned to a field, Tableau Prep examines the field values and flags the values that don't match so that you can take a closer look.



To view only the values that are valid or not valid, use the new filter option on the drop-down menu:



In this release we support the following data roles:

- Email
- URL
- Geographic roles (Based on current geographic data and is the same data used by Tableau Desktop)
  - Airport
  - Area code (U.S.)
  - CBSA/MSA
  - City
  - Congressional District (U.S.)
  - Country/Region
  - County
  - NUTS Europe
  - State/Province
  - Zip code/Postal code

For more information see [Use Data Roles to Validate your Data](#) on page 213.

## Join or Union Data

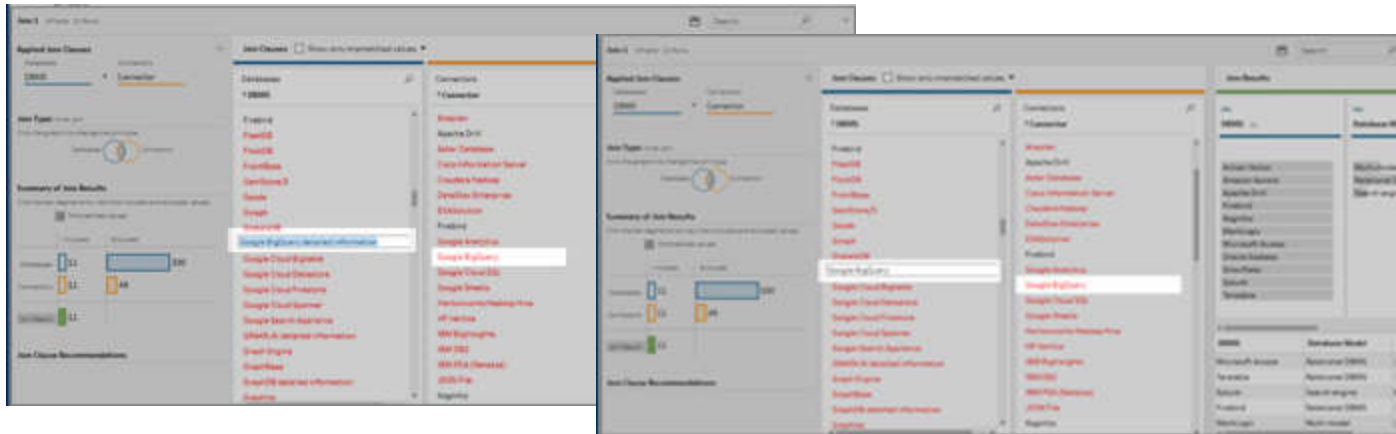
### Fix mismatched fields directly in the join clause

When you join two tables of data you will often have field values that are the same but are mismatched due to data entry errors. Tableau Prep helps identify mismatched fields in your join clauses by turning the mismatched field values red. But wouldn't it be great if you could just fix those field values right in your join clause? Well now you can.

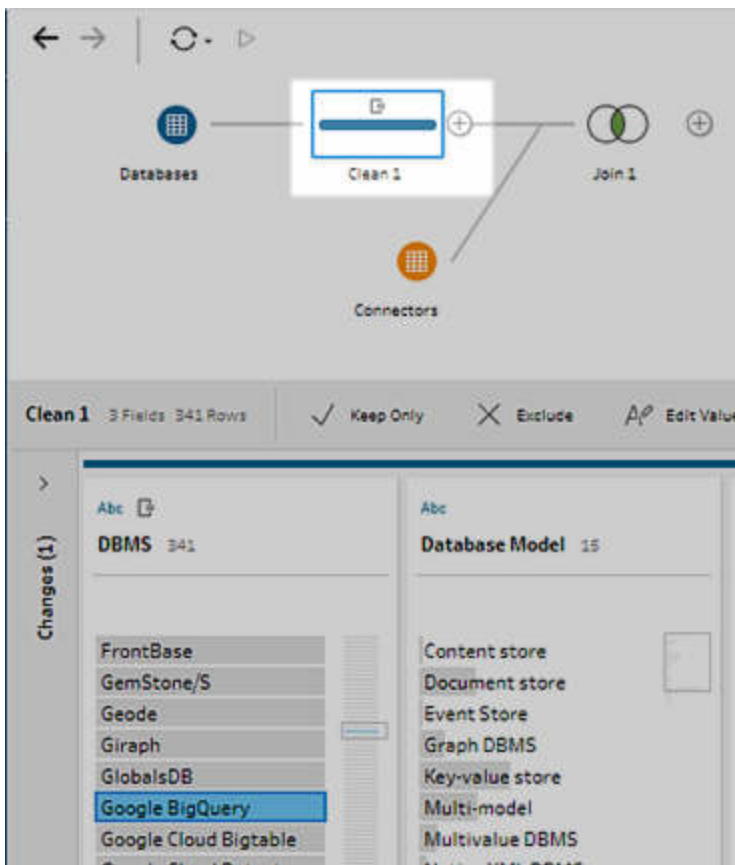
Now when you join two data sets you can edit the field values that you join on, right in the join clause to fix them. Simply double-click in a field value to edit it.

**Select field to change**

**Edit in-line**



The cleaning action is automatically pushed back to the previous cleaning step in the flow for the appropriate join data set.



No cleaning step before the join step? No problem. Tableau Prep automatically adds one for you to capture the cleaning operation from the join.

For information about how to clean field values directly in a join clause, see [Aggregate, Join, or Union Data](#) on page 307.

## Added in version 2018.2.2

### Install and Deploy Tableau Prep

- [Set your display language](#) below

### Connect to Data

- [Connect to data stored in Microsoft Access](#) on the next page

### Clean and Shape Data

- [Change the color scheme for your flow steps](#) on the next page
- [Add descriptions to your steps](#) on page 73
- [Use fuzzy match to find and fix spelling errors](#) on page 74

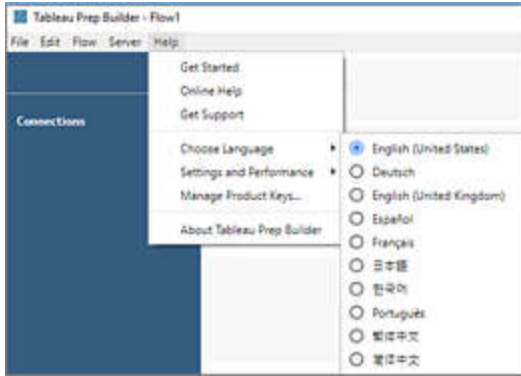
### Save and Share your work

- [Run flows from the command line](#) on page 75

## Install and Deploy Tableau Prep

### Set your display language

When you start up Tableau Prep, it now detects the locale set on your computer and displays the user interface, dates, and number formats in the appropriate language. If you want to change the display language for the user interface you can select from the supported languages from the top menu under **Help > Choose Language**.



For more information see [Set your display language](#) in the Tableau Desktop and Tableau Prep deployment guide.

## Connect to Data

### Connect to data stored in Microsoft Access

If you use Microsoft Access for data entry or to store your data tables you can now connect to your access files (from version 2007 or higher) through a file browser and clean your data with Tableau Prep.

This connector requires a 64-bit driver.

For more information about how to connect Tableau Prep to your data, see [Access](#) in the Tableau Desktop help.

**Note:** Data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

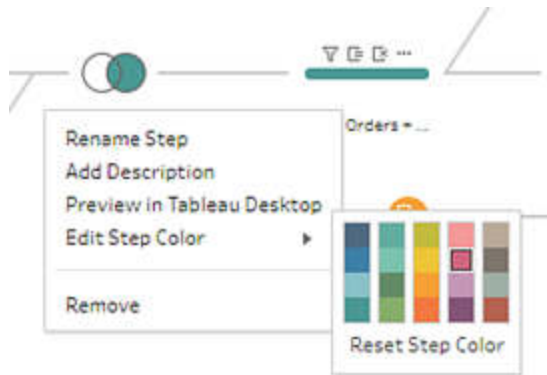
## Clean and Shape Data

### Change the color scheme for your flow steps

By default, Tableau Prep assigns each step in your flow a color to help you easily track the changes you make to your data as you build your flow. But you have choices when it comes to this color scheme.



You can now pick from a color palette to change the color scheme for one or more steps. Just select the steps in the Flow pane that you want to change, right-click the selected steps and select **Edit Step Color** from the context menu.

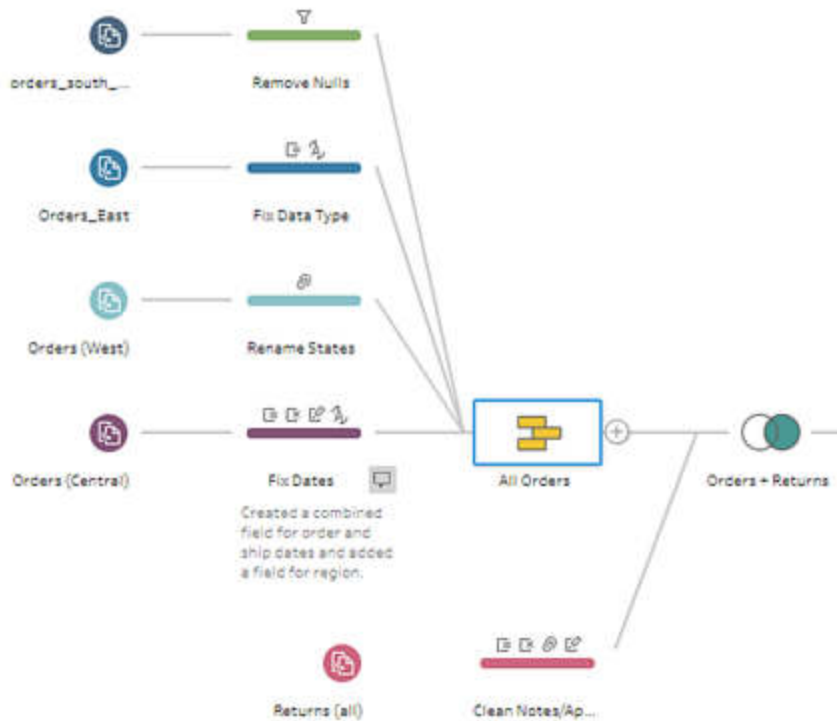


Don't like how it looks? Click **Undo** from the top menu or select **Reset Step Color** from the color palette menu.

For more information, see [Change the flow color scheme](#) on page 188.

## Add descriptions to your steps

If you share your flows with others, communicating the changes that you made and why can be cumbersome. To make it easier, you can now add a short description to any individual step in your flow and it displays right in the flow pane.



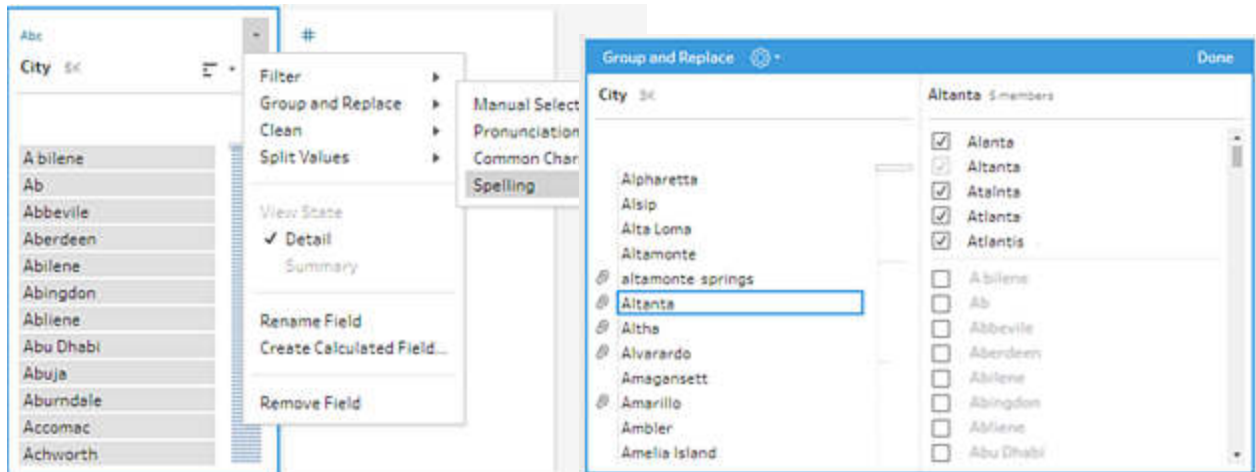
For more information about adding descriptions to flow steps, see [Add descriptions to flow steps and cleaning actions](#) on page 189.

## Use fuzzy match to find and fix spelling errors

To help you quickly identify and clean up multiple variations of the same value due to spelling errors, we've added another option to our fuzzy match cleaning feature. Use the new **Spelling** option to find and group text values that differ because of additional or missing letters. This option works in any supported language.

The **Spelling** option uses the Levenshtein distance algorithm to compute an edit distance between two text values and then groups them together when the edit distance is less than a default threshold value.

For more information, see [Edit field values](#) on page 246.



## Save and share your work

### Run flows from the command line

To keep data fresh in Tableau Prep you run your flow. However, opening Tableau Prep every time you want to run flows can take time. To help streamline this process, you can now run flows from the command line without having to open it in Tableau Prep.

To run a flow from the command line, you'll need:

- The path to the flow (.tfl) file that you want to run.
- A .json file that contains the database credentials for any databases that the flow connects to for its input steps and the credentials for the server where the output is published.

If your flow connects to or publishes to local files or files that are stored on a network share for inputs or outputs, then this file isn't needed.

**Note:** Connecting to or publishing files that are stored on a network share that are password protected isn't supported.

- Administrator permissions on the machine where you are running the flow.

This option is available on both Windows (Task Scheduler is supported) and Mac machines. To use this process, you need an activated version of Tableau Prep and the process must be run on the same machine where Prep is installed.

For information about how to run flows from the command line, see [Refresh flow output files from the command line](#) on page 334.

## Added in version 2018.2.1

### Install and Deploy Tableau Prep

- [Deactivate Tableau Prep from the command line](#) on the facing page
- [Use virtual desktop support to optimize Tableau Prep Installations](#) on the facing page

### Connect to Data

- [Connect to cloud data sources and Hadoop Hive](#) on the facing page

### Explore Your Data

- [New filter options to keep only the data you want](#) on page 78

### Clean and Shape Data

- [Use the ISO-8601 date standard in calculated fields](#) on page 79
- [Apply cleaning operations in the data grid](#) on page 79
- [Use multi-select to group values in the Profile pane](#) on page 80
- [Other enhancements](#) on page 80

### Join or Union Data

- [Use union recommendations to clean mismatched fields](#) on page 81
- [Identify mismatched fields for all join types](#) on page 83

# Install and Deploy Tableau Prep

## Deactivate Tableau Prep from the command line

Like Tableau Desktop, if you no longer need Tableau Prep on your computer you can now deactivate it from the command line using the `-return` option.

For more information see [Deactivate the product key](#). For more information about other installer properties that are available for Tableau Prep from the command line, see the Installer options and relevant sections in [Deploy Tableau Desktop](#).

## Use virtual desktop support to optimize Tableau Prep Installations

Just like Tableau Desktop you can now configure virtual desktop support to optimize your installations of Tableau Prep for non-persistent virtual desktops or for computers that are regularly reimaged. With virtual desktop support, Tableau Prep licenses are automatically deactivated after a predetermined period of time using a Tableau-hosted “Authorization to Run” (ATR) service, eliminating the need to manually deactivate the product key.

For more information about how to configure this option, see [Configure Virtual Desktop Support](#) in the Tableau Desktop and Tableau Prep deployment guide.

## Connect to Data

### Connect to cloud data sources and Hadoop Hive

We've added support for the following connectors so that you can connect to cloud data and data stored in Hadoop.

- Snowflake.
- Amazon EMR Hadoop Hive
- Cloudera Hadoop (Hive and Impala)
- Hortonworks Hadoop Hive
- MapR Hadoop Hive

- Apache Drill
- SparkSQL

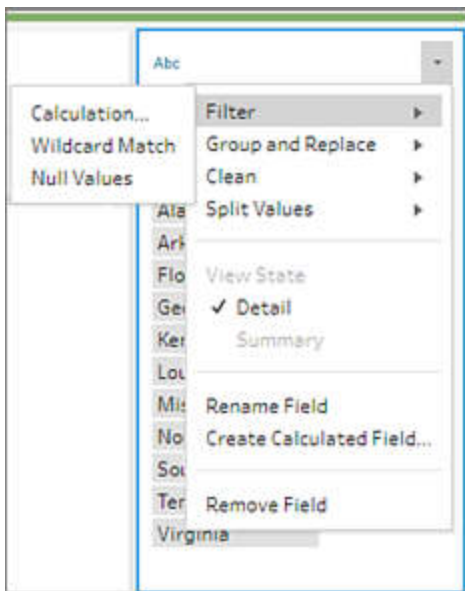
For more information about how to connect Tableau Prep to your data, see the topic for your connector under [Supported Connectors](#) in the Tableau Desktop help.

These data connectors are not backward compatible. Flows that include these connectors may open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

## Explore Your Data

### New filter options to keep only the data you want

No more writing complex calculations to keep or exclude Null values or to set up a wildcard match on text values. Instead select one of the new filter options on a field to see the impact of your changes right away without having to first write a calculation and then revert your change if it doesn't give you the results you want.



For more information, see [Examine Your Data](#) on page 198.

## Clean and Shape Data

## Use the ISO-8601 date standard in calculated fields

Creating calculated fields to support European calendars just got easier. Tableau Prep now supports the ISO-8601 international date standard for the following date parts:

- "iso-year"
- "iso-quarter"
- "iso-week"
- "iso-weekday"

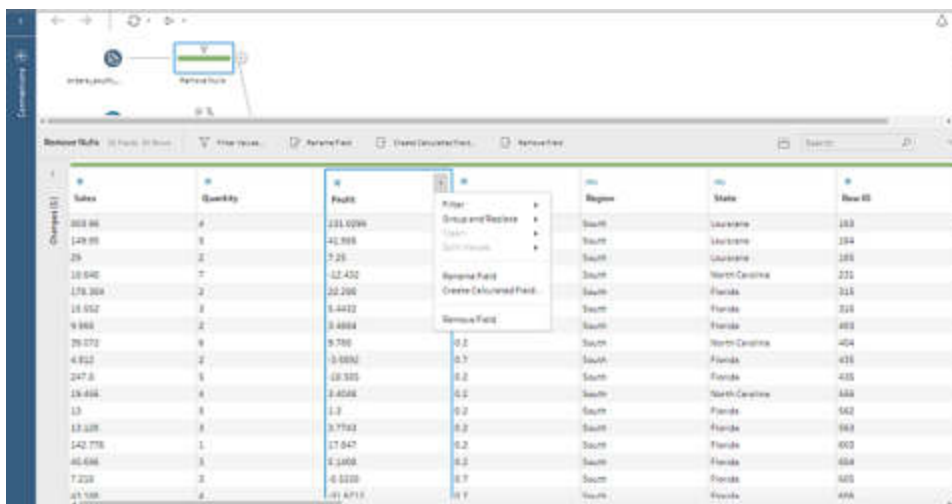
Use these date parts in functions DATEPART, DATETRUNC, DATENAME, DATEDIFF, and DATEADD.

For example `Week Number = STR(DATEPART('iso-year', [Week Date])) + "-" + STR(DATEPART('iso-week', [Week Date]))`

For more information about how to work with date functions in calculated fields, see [Date Functions](#) in the Tableau Desktop help.

## Apply cleaning operations in the data grid

In prior versions, the data grid showed you a preview of your data, but had few cleaning options available. Now you can act on your data anywhere. If you want to work with the detailed values in the data grid, collapse the Profile pane and perform the same cleaning operations that are available in the Profile pane in the data grid.

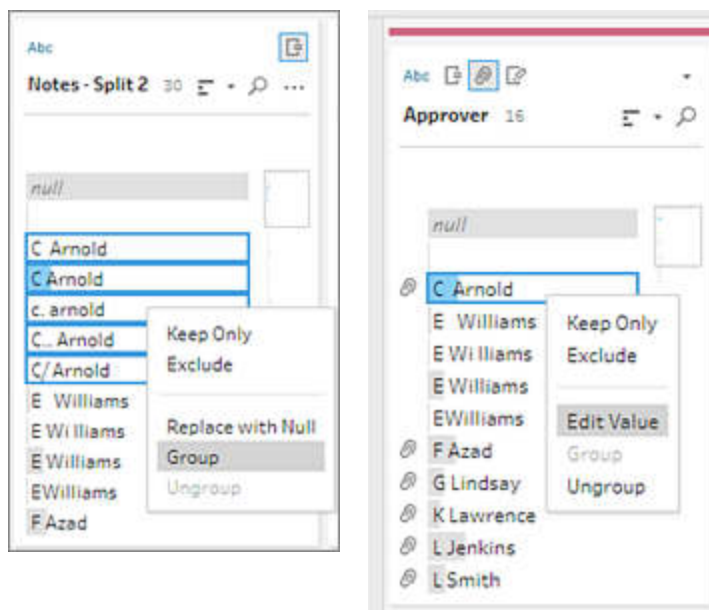


For more information, see [Clean and Shape Data](#) on page 229.

## Use multi-select to group values in the Profile pane

To quickly group a set of values for a field, you can now multi-select the values in the Profile card, then right-click to open the menu and select **Group**. The values are grouped under the field value that you select when you right-click to open the menu. A paperclip icon shows next to the grouped value.

Right-click the grouped values to open the menu again to ungroup or edit the values.



For more information about grouping values, see [Edit field values](#) on page 246

## Other enhancements

We've also made the following enhancements to improve usability:

- New icons show on the menu when adding steps to your flow to provide visual cues and help you learn the visual language of Tableau Prep.





- New animations in the Profile pane help catch your eye to better see the impact of your changes.

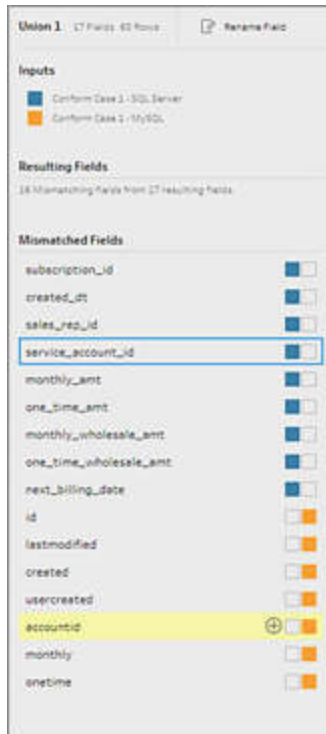
## Join or Union Data

### Use union recommendations to clean mismatched fields

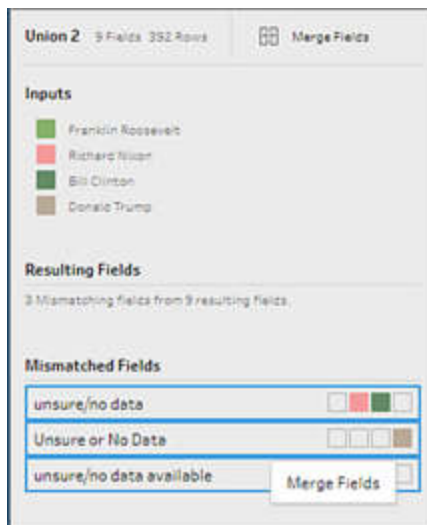
Fixing mismatched fields after unioning two or more files just got easier. A new **Mismatched Fields** section in the **Union summary** pane shows a list of fields that don't match and the data source where they originated. Merge your mismatched fields directly in this section using one of the following options:

- Select a field in the list. If Tableau Prep identifies a field with similar characteristics, based on similar data types and field names, it highlights the field in yellow, suggesting a match.

Click the plus (+) button that appears on hover on the suggested matched field to merge the two fields.



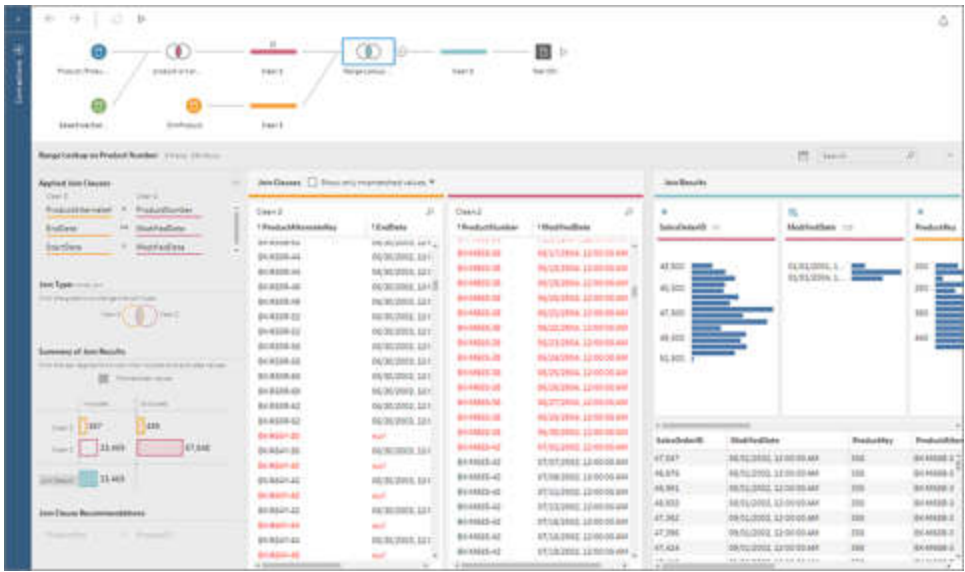
- Select two or more fields in the list, right-click on one of them and select **Merge Fields**.



For more information about unioning data and resolving mismatched fields, see [Union your data](#) on page 314.

## Identify mismatched fields for all join types

Easily identify fields that don't match no matter how you join tables. In prior versions the **Join Clauses** tables showed field values that were excluded only when one field equaled another. But now you can see this data using any operator when matching join clauses, such as "End Date >= Modified Date".



For more information, see [Join your data](#) on page 308.

## Added in version 2018.1.2

### Install and Deploy Tableau Prep

- **Activate and register Tableau Prep from the command line (Windows)** on the next page

### Connect to Data

- **Connect to data stored in statistical files or on Presto** on page 85

- **Union sub-tables found by Data Interpreter in the Input step** on the facing page
- **Better feedback when loading tables** on page 86

#### Explore Your Data

- **Reorder fields in the Profile pane and the Data grid** on page 87

#### Clean and Shape Data

- **Use drag-select to remove multiple steps in your flow** on page 87
- **Pivot multiple groups of fields in a single action** on page 88
- **Improved field naming when merging fields** on page 88
- **Other enhancements** on page 89

## Install and Deploy Tableau Prep

### Activate and register Tableau Prep from the command line (Windows)

Like Tableau Desktop, you can now activate and register Tableau Prep from the command line by including a command line with the following properties:

- `ACTIVATE_KEY="<key>"`. The installer runs `-activate` to apply the product key.
- `REGISTER="1"`. During the installation process, the installer will run the `-register` process and add the registration information.

For more information about these installer properties and how to activate and register Tableau Prep from the command line, see the Installer options and relevant sections in [Deploy Tableau Desktop](#).

## Connect to Data

## Connect to data stored in statistical files or on Presto

We've added two new connectors to help you connect to data from more locations.

- Statistical files. Connect to to SAS (\*.sas7bdat), SPSS (\*.sav), and R (\*.rdata) data files.
- Presto. For more information about how to configure your connection to Presto, see [Presto](#) in the Tableau Desktop help.

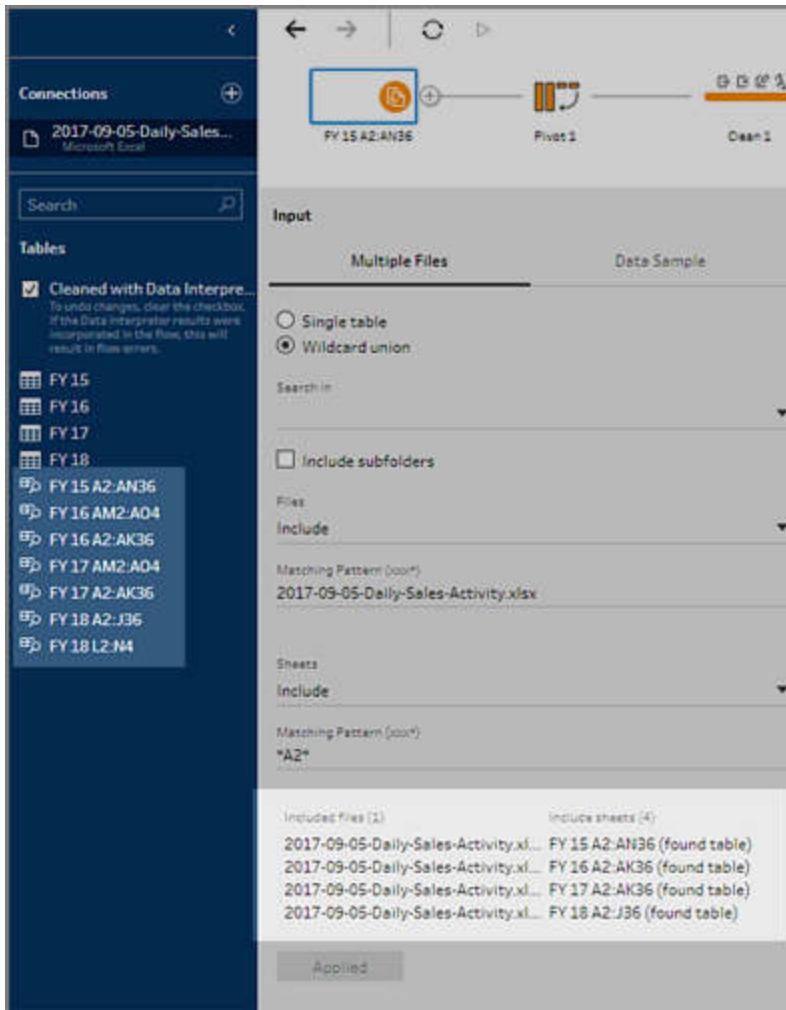
These data connectors are not backward compatible. Flows that include these connectors will open in a prior version of Tableau Prep, but will have errors or can't run unless the data connections are removed.

## Union sub-tables found by Data Interpreter in the Input step

Using Data Interpreter to clean your Microsoft Excel data and now you want to union the resulting sub-tables? You can now use wildcard union to union all of the found sub-tables in the Input step.

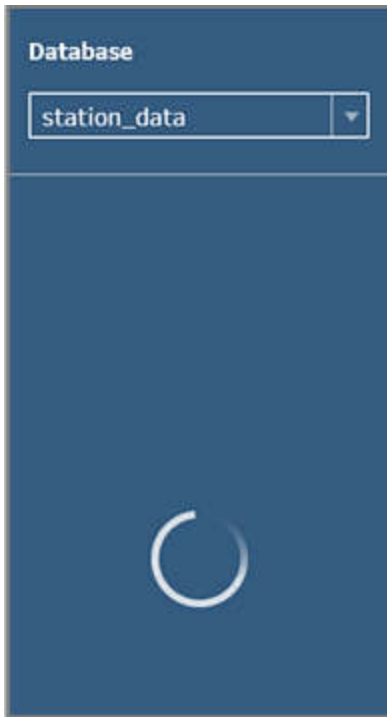
Simply drag one of your sub-tables to the **Flow** pane, and then use the wildcard search criteria to select the remaining sub-tables to union the data and include all the sub-table data in the Input step.

For more information see [Connect to Data](#) on page 145.



## Better feedback when loading tables

When you connect to a database for the first time, it can sometimes seem like nothing happened. Now a new indicator tells you that the data is still loading.



## Explore Your Data

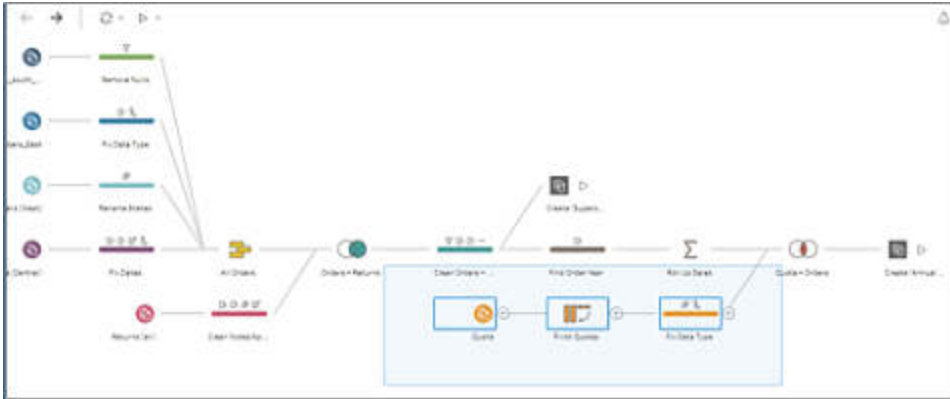
### Reorder fields in the Profile pane and the Data grid

You can now drag and drop to reorder your fields in both the **Profile** pane and the **Data** grid and the two panes will stay in sync. We also maintain the field order even when you rename a field. Fields are no longer reordered automatically.

## Clean and Shape Data

### Use drag-select to remove multiple steps in your flow

Removing whole sections of your flow just got easier. You can now click in the **Flow** pane and use your mouse to drag and select the section of the flow that you want to remove. Then right-click to remove all of the selected steps at once. For more information see [Clean and Shape Data](#) on page 229.



## Pivot multiple groups of fields in a single action

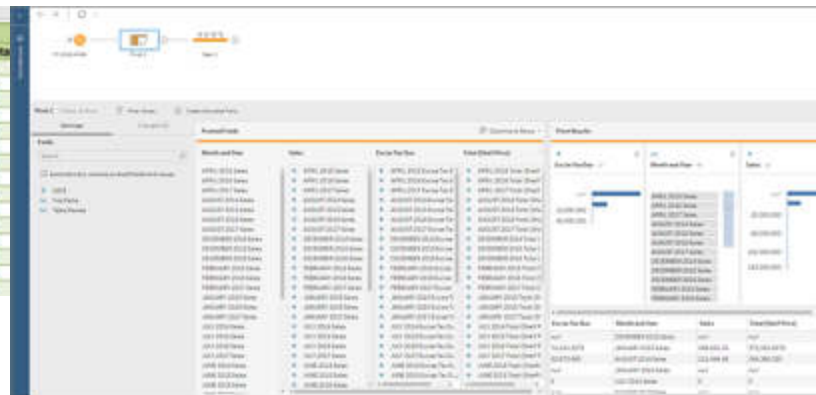
The pivot option in Tableau Prep has been expanded to better support more complex spreadsheets or text files. You can now perform either a single pivot or select groups of fields to pivot.

For example to pivot a spreadsheet to see sales, tax, and totals by month and year you can pivot each group of fields to get the results you want. For more information, see [Clean and Shape Data](#) on page 229.

### Original Spreadsheet

DATE	DECEMBER 2014			JANUARY 2015		
	Sales (Shelf Price)	Excise Tax Due	Total (Shelf Price)	Sales (Shelf Price)	Excise Tax Due	Total
1	\$ 448,111	\$ 112,628	\$ 560,739	\$ 296,843	\$ 74,270	\$ 371,113
2	\$ 425,472	\$ 106,368	\$ 531,840	\$ 754,061	\$ 188,515	\$ 942,576
3	\$ 435,525	\$ 108,881	\$ 544,406	\$ 482,497	\$ 120,624	\$ 603,121
4	\$ 634,765	\$ 158,691	\$ 793,456	\$ 332,228	\$ 83,057	\$ 415,285
5	\$ 695,425	\$ 173,856	\$ 869,282	\$ 981,525	\$ 250,382	\$ 1,231,907
6	\$ 456,720	\$ 114,180	\$ 570,900	\$ 527,374	\$ 131,843	\$ 659,217
7	\$ 236,481	\$ 59,120	\$ 295,601	\$ 360,162	\$ 90,026	\$ 450,188
8	\$ 421,422	\$ 105,356	\$ 526,778	\$ 579,974	\$ 144,993	\$ 724,967
9	\$ 543,816	\$ 135,954	\$ 679,770	\$ 683,408	\$ 170,852	\$ 854,260
10	\$ 616,271	\$ 154,067	\$ 770,338	\$ 442,352	\$ 110,588	\$ 552,940
11	\$ 796,542	\$ 199,135	\$ 995,677	\$ 288,605	\$ 72,155	\$ 360,760
12	\$ 726,270	\$ 181,567	\$ 907,837	\$ 674,127	\$ 168,530	\$ 842,657
13	\$ 477,288	\$ 119,302	\$ 596,590	\$ 526,451	\$ 131,813	\$ 658,264
14	\$ 245,898	\$ 61,475	\$ 307,373	\$ 573,842	\$ 143,811	\$ 717,653
15	\$ 456,254	\$ 114,064	\$ 570,318	\$ 698,952	\$ 174,738	\$ 873,690

### After pivoting on multiple groups of fields

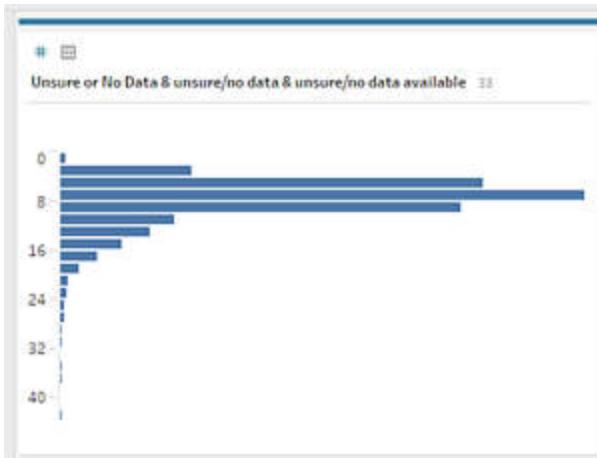


## Improved field naming when merging fields

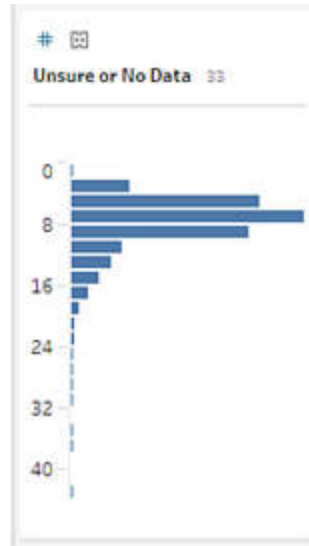
When you merge fields, the field names are no longer concatenated. Instead the field name of the target field persists. For more information, see [Merge fields](#) on page 241.



### Concatenated field naming



### Simplified field naming



## Other enhancements

We've also made the following enhancements to improve usability and performance:

- A new edit icon shows on hover on entries in the **Changes** pane to help you quickly see which items are editable.



- Renaming a step will no longer automatically run the flow so your authoring experience is seamless.
- The **Connections** pane automatically opens when you connect to a Microsoft Excel file with multiple sheets to help you quickly get to your data.



# Get Started with Tableau Prep Builder

**Note:** Tableau Prep version 2019.1.2 had changed its name to Tableau Prep Builder. If you are using an earlier version of the product you will still see "Tableau Prep" in the product.

This tutorial introduces you to the common operations that are available in Tableau Prep Builder. Using the sample data sets that come with Tableau Prep Builder, you will walk through creating a flow for Sample Superstore. This tutorial uses the most current version of Tableau Prep Builder. If you are using a previous version, your results may differ.

Watch for tips along the way to gain insights into how Tableau Prep Builder helps you clean and shape your data for analysis.

To install Tableau Prep Builder before continuing with this tutorial, see [Install Tableau Desktop or Tableau Prep Builder from the User Interface](#) in the Tableau Desktop and Tableau Prep Builder Deployment guide. Otherwise you can download the [free trial](#).

**Note:** To complete the tasks in this tutorial, you need to install Tableau Prep Builder, and you need to download the following sample Superstore data files to a single location:

- [Orders\\_South\\_2015](#)
- [Orders\\_South\\_2016](#)
- [Orders\\_South\\_2017](#)
- [Orders\\_South\\_2018](#)
- [Orders\\_Central](#)
- [Orders\\_East](#)
- [Orders\\_West](#)
- [returns\\_reasons\\_new](#)

If using version 2019.4.2 and earlier, you can also find the files in the following location on your computer after installing Tableau Prep Builder:

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en\_US\Superstore Files
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en\_US/Superstore Files

## Here's the story...

You work at the headquarters for a large retail chain. Your boss wants to analyze product sales and profits over the last four years for the company. You suggest that he use Tableau Desktop to do that. Your boss thinks that's a great idea and wants you to get right on that.

As you start gathering all the data you'll need, you notice that the data has been collected and tracked differently for each region. You also notice a lot of creative data entry in the different files, and that one region even has a separate file for each year!

Before you can start analyzing the data in Tableau, you'll have to do some serious data cleaning first, and it's going to be a long night.

As you rummage for restaurant menus to order some dinner, you remember that Tableau recently introduced a new product called Tableau Prep Builder that might help you with your Herculean data cleaning task.

You sign up for a [free trial](#) and decide to give it a try.

## 1. Connect to data

The first thing you see when you open Tableau Prep Builder is a Start page with a **Connections** pane, just like Tableau Desktop.

To get started, the first step is to connect to your data and create an Input step. From there you will start building a workflow or "flow", as it's called in Tableau Prep Builder, and add more steps to take action on your data as you go.

**Tip:** The Input step is the ingestion point for your data and the starting point for your flow. You can have multiple Input steps and some might include multiple data files. For more information about connecting to data, see [Connect to Data on page 145](#).

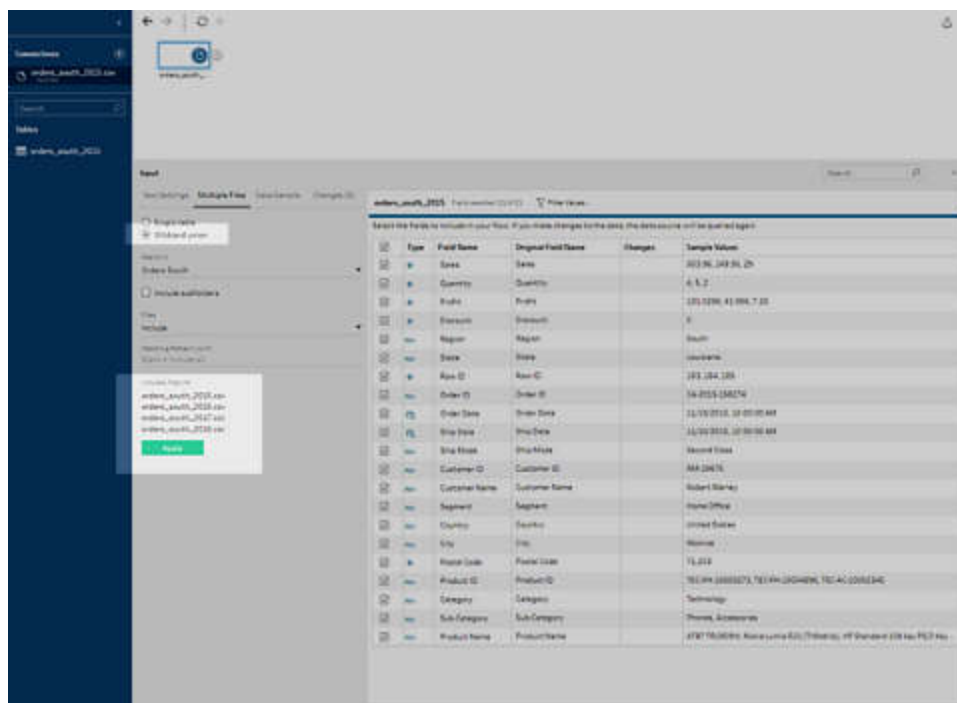


**Tip:** For single tables, Tableau Prep Builder automatically creates an Input step for you in the **Flow** pane when you add data to your flow. Otherwise you can use drag-and-drop to add tables to the **Flow** pane.

- You have three other files for your orders in the South. You could add each file individually, but you want to combine all the files together into one Input step, so you click the **Multiple Files** tab in the **Input** pane.
- You see an option for **Wildcard union**. Select it.

You notice that the directory where you selected your file is already populated and the other files you need are listed in the **Included files** section in the Input pane.

**Tip:** Using a wildcard union is a great way to connect to and combine multiple files from a single data source with a similar name and structure. To use this option, the files must be in the same parent or child directory. If you don't see the files you need right away, change your search criteria. For more information, see [Union files and database tables in the Input step](#) on page 173.

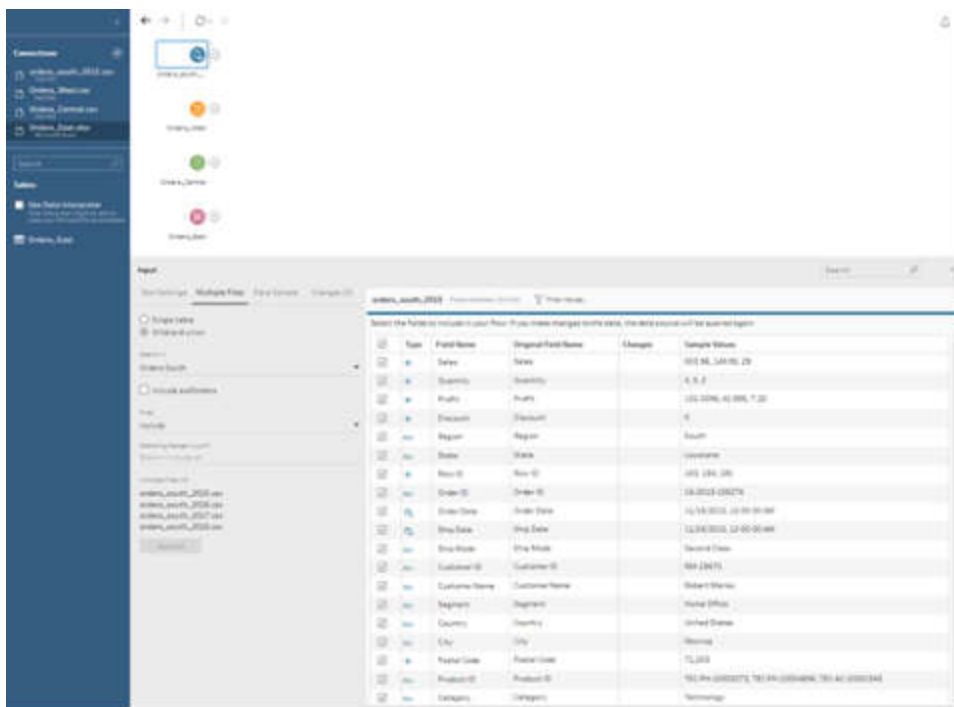


- Click **Apply** to add the data from these files to the **orders\_south\_2015** input step.

The files for the other regions are all single table files, so you can select all of the files at once and add them to your flow.

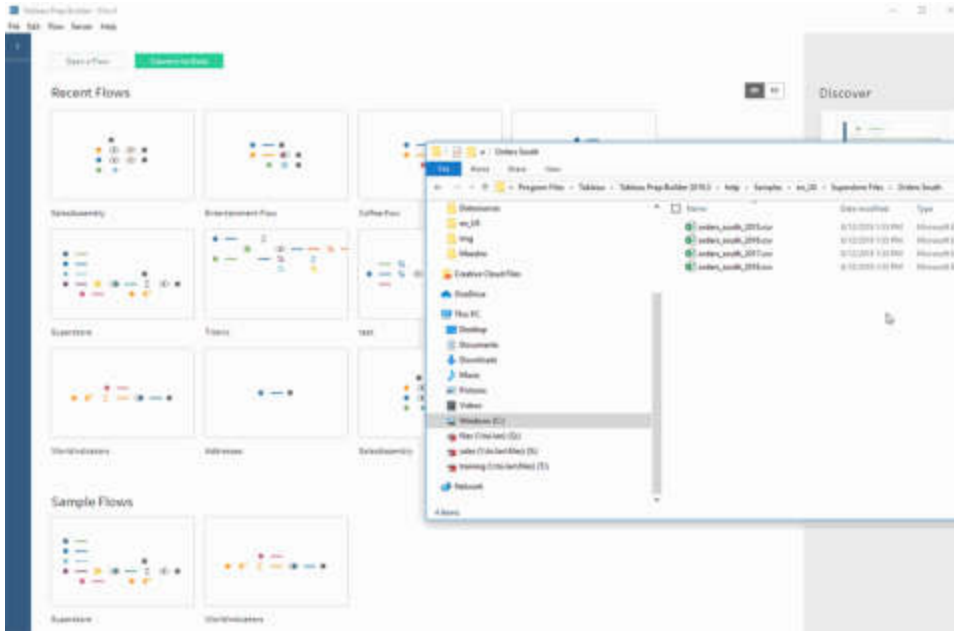
7. Open File Explorer or Finder and navigate to the directory for the files. Ctrl-click or Cmd-click (MacOS) to select the following files and drag-and-drop them onto the **Flow** pane to add them to your flow. (For file location, see [Wrap up and resources on page 132.](#))
  - Orders\_Central.csv
  - Orders\_East.xlsx
  - Orders\_West.csv

**Note:** These are different file types. If you don't see all of these files, make sure your file explorer or finder is set to view all file types.



Check your work: Watch "Connect to data" in action.

*Click the image to replay it*



## 2. Explore your data

Now that you have the data files loaded into Tableau Prep Builder, you're pretty sure that you want to combine all of the files together. But before you do that, it might be a good idea to take a look at them first and see if you can spot any issues.

When you select an Input step in the **Flow** pane, you can see the settings used to bring in the data, the fields that are included, and some sample values.

This is a good place to decide how much data you want to include in your flow and remove or filter fields that you don't want. You can also change any data types that were assigned incorrectly.

**Tip:** If you are working with large data sets, Tableau Prep Builder will automatically bring in a sample of the data to maximize performance. If you don't see the data you expect, you might need to adjust the sample. You can do this on the **Data Sample** tab. For more information about configuring your data options and sample size, see [Set your data sample size on page 169](#).

In the **Flow** pane, as you select each step and look over each data set, you notice a few things that you want to fix later and one thing that you can fix now in the **Input** step.



- In the **Flow** pane, click the **Orders\_Central** Input step to select it. In the **Input** pane, you notice the following issues:
  - The order dates and ship dates are separated out into fields for month, day, and year.
  - Some of the fields have different data types than the same fields in other files.
  - There is no field for **Region**.

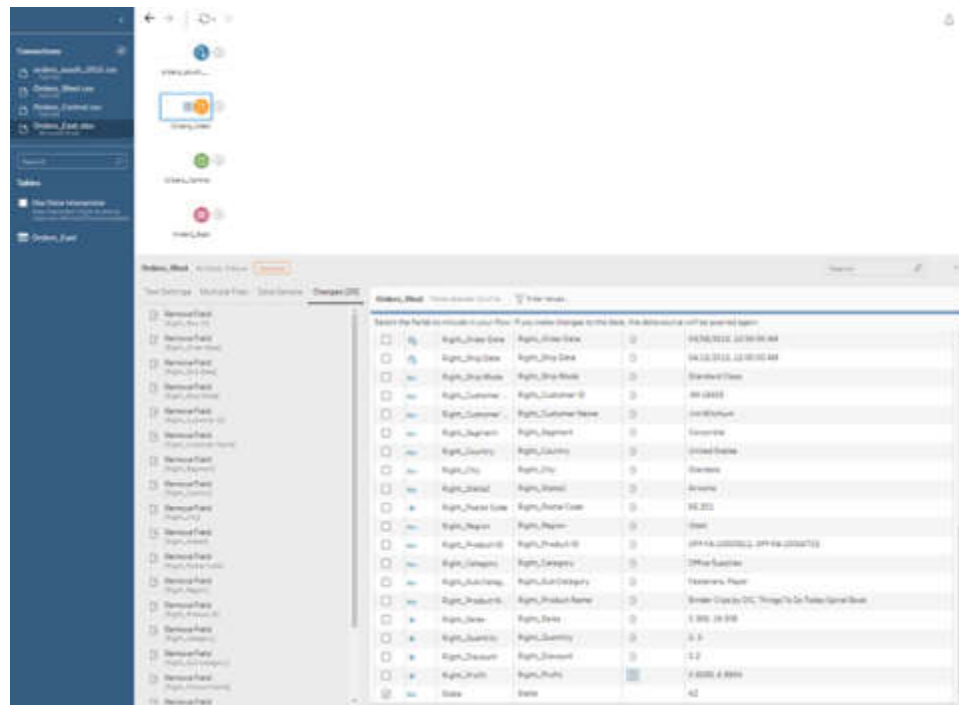
You'll need to do some cleaning on these fields before you can combine this file with the others files. But you can't fix that here in the **Input** step, so you make a note to do this later.

- Select the **Orders\_East** Input step.

The fields in this file look like they align pretty well with the other files. But the **Sales** values all seem to have the currency code included. You'll need to fix that later, too.

- Select the **Orders\_West** Input step. There are some issues in this file too.
  - The **State** field uses abbreviations for the state name. Other files spell this out, so you'll need to fix that later.
  - There are a lot of fields that start with **Right\_**. These fields appear to be duplicates of the other fields. You don't want to include these duplicate fields in your flow. This is something you can fix here in the **Input** step:

**To fix this now**, clear the check box for all fields that start with **Right\_**. This tells Tableau Prep Builder to ignore these fields and not to include them in the flow.



**Tip:** When you perform cleaning operations in a step, like removing fields, Tableau Prep Builder tracks your changes in the **Changes** pane and adds an annotation (in the form of a little icon) in the Flow pane to help you keep track of the actions you take on your data. For Input steps, an annotation is also added to each field.

In Tableau Prep Builder version 2018.3.1 and later, cleaning operations are available in most steps in your flow. In prior versions, most cleaning operations must be performed in a cleaning step.

Now that you've identified a few troublemakers in your data sets, the next step is to examine your data a bit more closely and clean up any issues that you find so that you can combine and shape your data and generate an output file that you can use for analysis.

### 3. Clean your data

In Tableau Prep Builder, examining and cleaning your data is an iterative process. After you decide on the data set that you want to work with, the next step is to examine and take action on that data by applying various cleaning, shaping, and combining operations to it. You apply these operations by adding steps to your flow. For more information about cleaning options in Tableau Prep Builder, see [Clean and Shape Data on page 229](#)


Steps come in many flavors, depending on what you are trying to do. For example, add a cleaning step (**Add Step** or **Add Clean Step** depending on your version) any time you want to apply cleaning operations to your fields like filter, merge, split, rename, and so on. Add an aggregation step (**Add Aggregate**) to group and aggregate fields and change the level of detail of your data. For more information about the different step types and their uses, see [Build and Organize your Flow on page 182](#).

**Tip:** As you add steps to your flow, a flow line is automatically added to connect the steps to one another. You can move these flow lines around and remove or add them as needed.

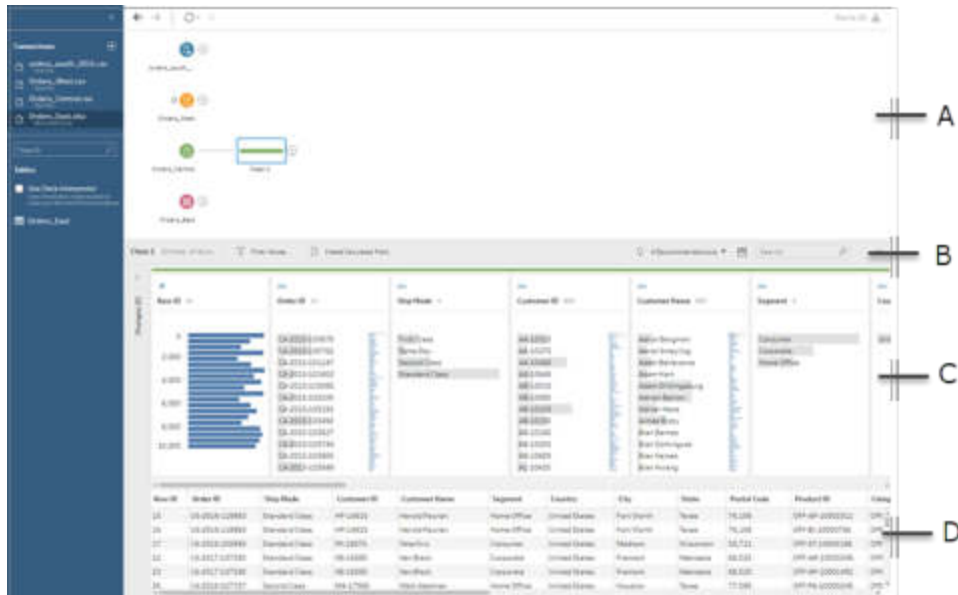
When you run your flow, these connection points are required so Tableau Prep Builder knows which steps are connected and in which order the steps apply in the flow. If a flow line is missing, the flow will be broken and you'll get an error.

## Clean Orders\_Central

To address the issues you noticed earlier and to see if there are any other issues, you start by adding a cleaning step to the **Orders\_Central** Input step.

1. In the **Flow** pane, select **Orders\_Central**, click the plus  icon and select **Add Step** or **Add Clean Step**, depending on your version.


When you add a cleaning step to your flow, the workspace changes and you see the details of your data.



A. Flow pane, B. Toolbar, C. Profile pane, D. Data grid

The workspace is now split into three parts: the **Flow** pane, the **Profile** pane with a toolbar, and the **Data** grid. The **Profile** pane shows you the structure of your data, summarizing the field values into bins so that you can quickly see related values and spot outliers and null values.

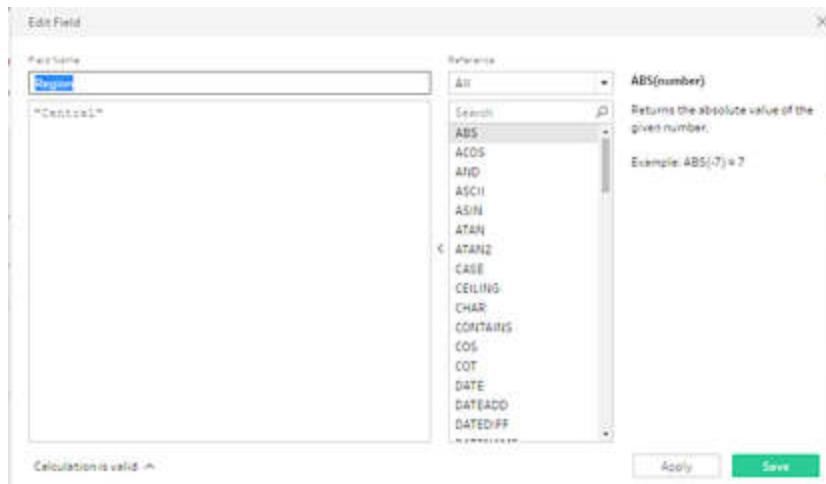
The **Data** grid shows you the row level detail for your fields.

**Tip:** Each field in the **Profile** pane is shown on a profile card. Use the **More options**  menu (drop-down arrow in prior versions) on each card to see and select the different cleaning options that are available for that field type. You can also sort the field values, change the data type, assign a data role to the field or drag and drop the profile cards and the columns in the **Data** grid to rearrange them.

### Clean data with calculated fields

This data set is missing a field for **Region**. Since the other data sets have this field you'll need to add it so that you can combine your data later. You'll need to use a calculated field to do this.

2. In the toolbar, click **Create Calculated Field**.
3. In the Calculation editor, name the calculated field **Region**. Then enter **"Central"** (including the quotes) and click **Save**.



You love the flexibility of being able to use calculated fields to shape your data. You are pleased to see that Tableau Prep Builder uses the same calculation editor language as Tableau Desktop.

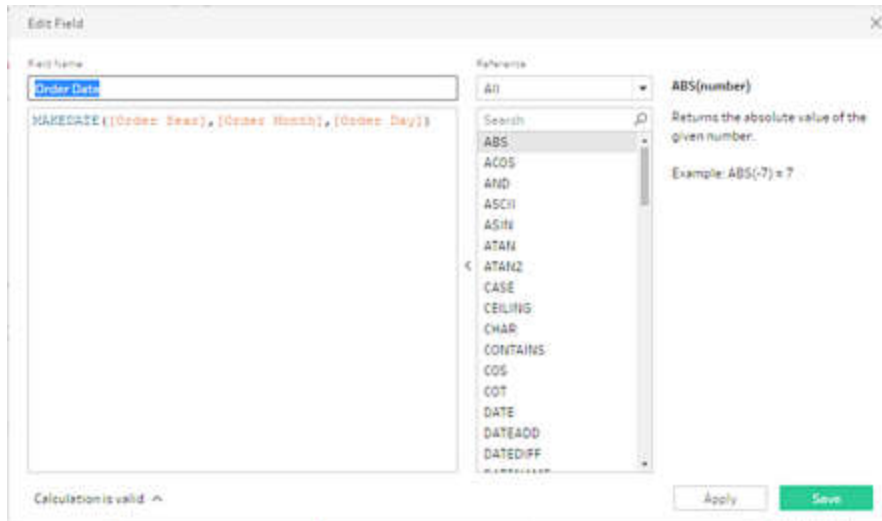
**Tip:** When you make changes to your fields and values, Tableau Prep Builder keeps track of them in the **Changes** pane. An icon (annotation) representing the change is also added to the cleaning step in the flow and to the field in the **Profile** pane. We'll look at the **Changes** pane after making more changes.

Next you want to address the separate order date and ship date fields. You want to combine them into two single fields, one for **Order Date** and one for **Ship Date** so they align with the same fields in the other data sets. Making sure your tables have the same fields will enable you to combine the tables using a union later.

You can use a calculated field again to do this in one easy step.

4. In the toolbar, click **Create Calculated Field** to combine the **Order Year**, **Order Month**, and **Order Day** fields into one field with the format "MM/DD/YYYY".
5. In the Calculation editor, name the calculated field **Order Date**. Then enter the following calculation and click **Save**:

```
MAKEDATE([Order Year],[Order Month],[Order Day])
```



Now that you have a new field for your order date, you want to remove the existing fields, as you no longer need them.

You have a lot of fields in the **Profile** pane. You notice a **Search** box in the top right corner on the toolbar. You wonder if you can use that to quickly find the fields that you want to remove. You decide to give it a try.

6. In the **Profile** pane, in the search box, type **Order**.

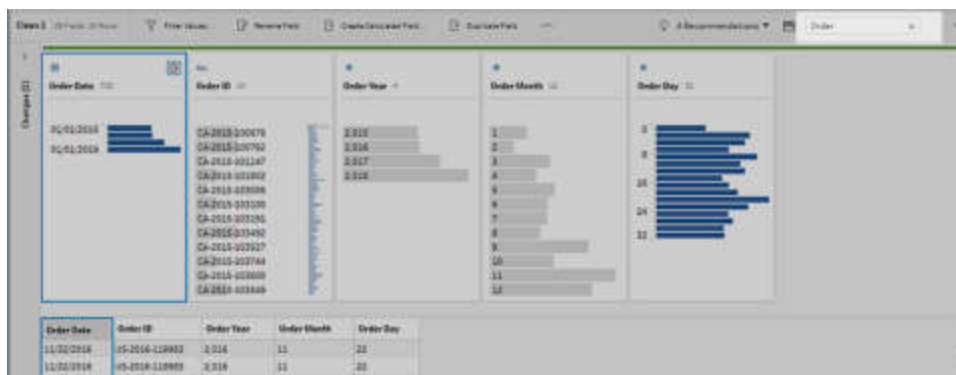


Tableau Prep Builder quickly scrolls all the fields with **Order** in the name into view. Cool!

7. Ctrl-click or Cmd-click (MacOS) to select the fields for **Order Year**, **Order Month**, and **Order Day**. Then right-click on the selected fields and select **Remove (Remove Field in prior versions)** from the menu to remove them.



8. Now repeat steps 4 through 7 above to create a single field for **Ship Date**. Try it on your own or use the steps below to help you.

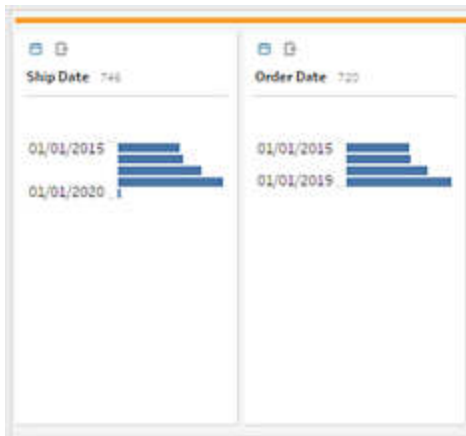
- In the toolbar, click **Create Calculated Field** to combine the **Ship Year**, **Ship Month**, and **Ship Day** fields into one field with the format "MM/DD/YYYY".
- Name the calculated field **Ship Date** and enter the calculation `MAKEDATE([Ship Year],[Ship Month],[Ship Day])`. Then click **Save**.
- Remove the **Ship Year**, **Ship Month**, and **Ship Day** fields. Search for the fields, select them, and select **Remove (Remove Field)** in prior versions) from the menu to remove the fields.

**Tip:** Tableau Prep Builder summarizes the data in the Profile pane into bins to help you quickly see the shape of your data, find outliers, spot relationships between fields, and so on.

In this scenario, the order and ship dates can now be summarized by year. Each bin represents a year from January of the starting year to January of the following year and is labeled accordingly. Because there are sales dates and ship dates that fall in the latter part of 2018 and 2019, we get a bin for that data that is labeled with the ending year 2019 and 2020 accordingly.

To change this view to the actual dates, click the **More options**  $\dots$  menu (drop-down

arrow in prior versions) in the Profile card and select **Detail**.



### Interact directly with fields to clean your data

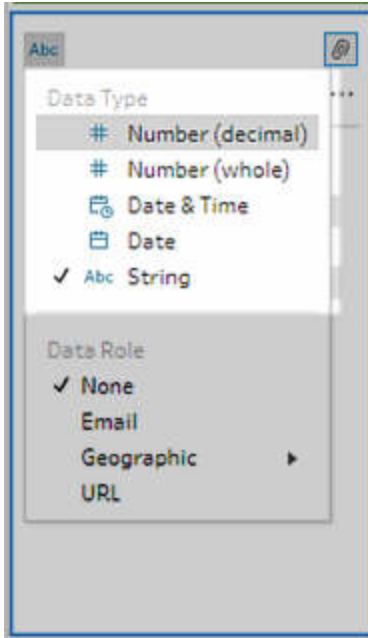
Your data is starting to look good. But, as you finish removing the extra fields for the order and ship dates, you notice that the **Discounts** field has a couple of issues.

- It's assigned to a **String** data type instead of a **Number (decimal)** data type.
- There's a field value **None** instead of a numeric value for no discount.

This will cause a problem when you combine the files, so you better fix that too.

9. Clear your search and enter **disc** in the search box to find the field.
10. Select the **Discounts** field, double-click the field value **None**, and change it to the numeric value **0**.
11. To change the data type for the **Discount** field from **String** to **Number (decimal)**, click **Abc** and select **Number (decimal)** from the drop-down menu.





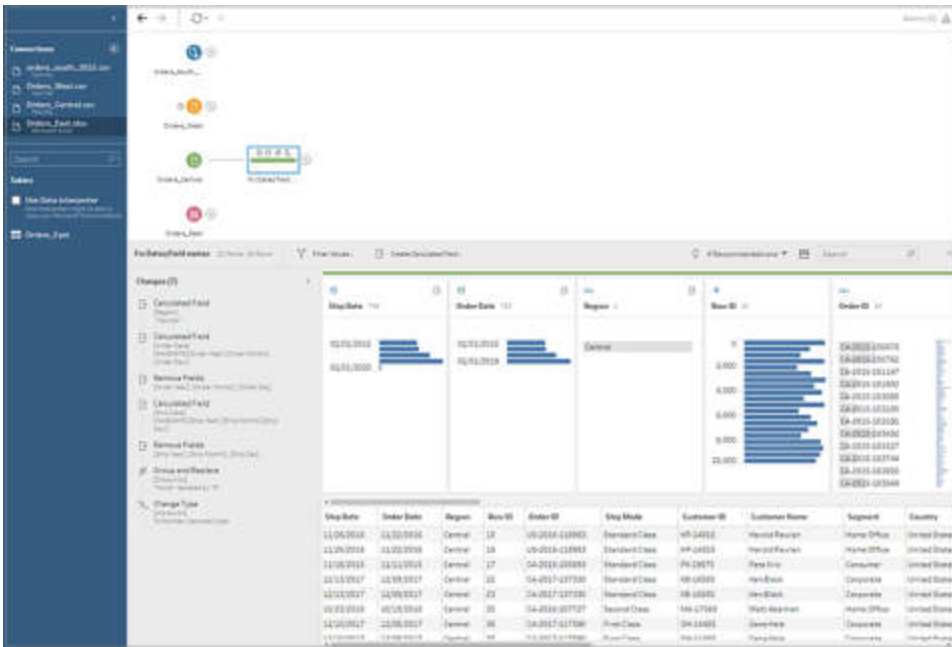
12. Finally name your step to help keep track of what you did in this step. In the **Flow** pane, double-click the step name **Clean 1** and type in **Fix dates/field names**.

## Review your changes

You made a lot of changes to this data set and you start to worry that you won't remember everything you did. As you look over your work, you see a column on the left of the **Profile** pane called **Changes**.

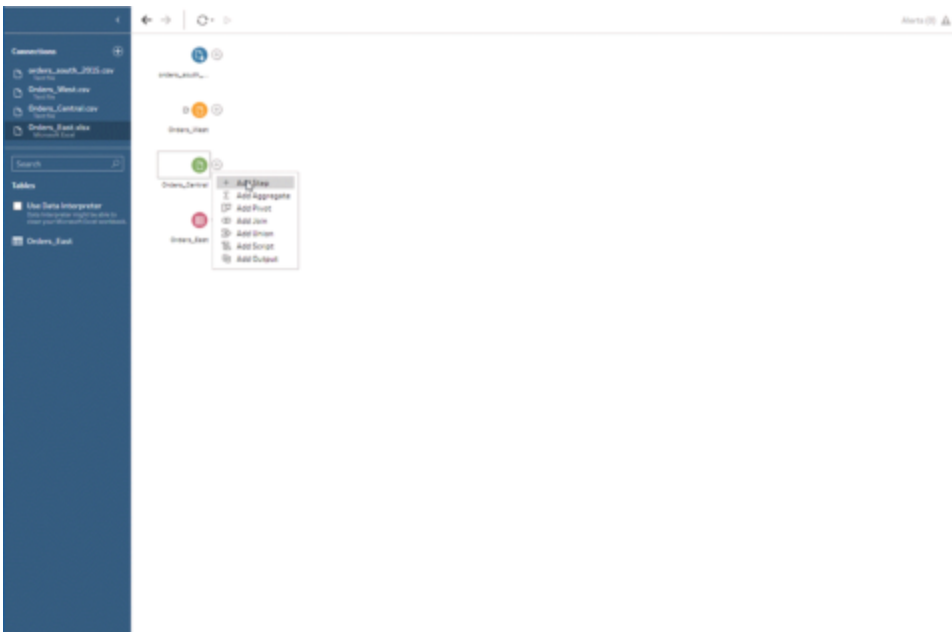
You click the arrow to open it and are delighted to see a list of every change you just made. As you scroll through the changes in the list, you notice that you can delete or edit your changes or even move them around to change the order that you did them in.

You love that you can easily find the changes you made in any step as you build your flow and experiment with the order of those changes to get the most out of your data.



Check your work: Watch "Clean Orders\_Central" in action.

*Click the image to replay it*




Now that you've cleaned one file, you take a look at the other files to see what other issues you need to fix.

You decide to look at the Excel file for **Orders\_East** next.

## Clean Orders\_East

As you look over the fields for the **Orders\_East** file, most of the fields look like they align with the other files, except for **Sales**. To take a closer look and see if there are any other issues to address, you add a cleaning step to the **Orders\_East** Input step.

1. In the **Flow** pane, select **Orders\_East**, click the plus  icon and select **Add Step** or **Add Clean Step**, depending on your version.


Looking at the **Sales** field you quickly see that the **USD** currency code has been included with the sales numbers, and Tableau Prep Builder interpreted these field values as a string.

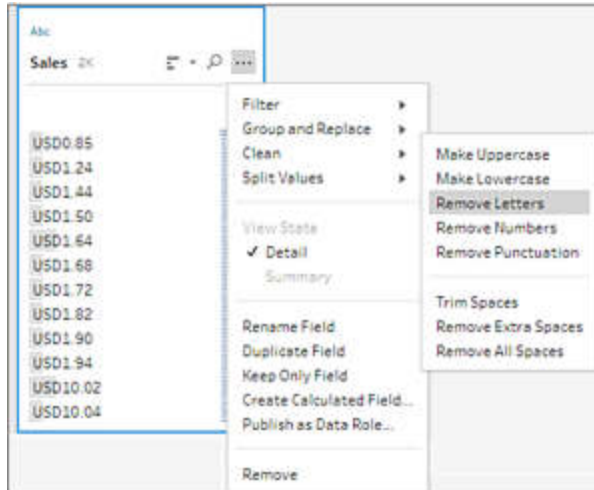
You'll need to remove the currency code from this field and change the data type if you want to get accurate sales data.

Fixing the data type is easy, you already know how to do that. But there are over 2000 unique rows of sales data and fixing every individual row to remove the currency code seems cumbersome.

But this is Tableau Prep Builder, and you decide to check out the drop-down menu to see if there is an option to fix this.

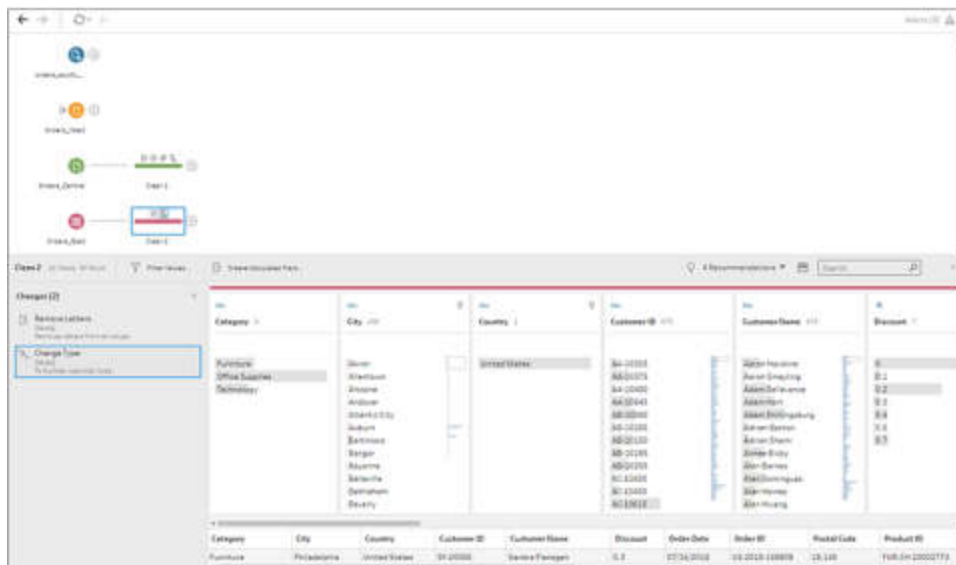
When you click the drop-down arrow for the **Sales** field, you see a menu option called **Clean** and an option under that to remove letters. You decide to give that a try and see what it does.

2. Select the **Sales** field. Click the **More options**  menu (drop-down arrow in prior versions) and select **Clean > Remove Letters**.



Wow! That cleaning option instantly removed the currency code from every field. Now you just need to change the data type from **String** to **Number (decimal)** and this file is looking good.

3. Click the data type for the **Sales** field and select **Number (decimal)** from the drop-down list to change the data type.




4. The rest of the file looks pretty good. Name your cleaning step to keep track of your work. For example, **Change data type**.


Next you look at your last file for **Orders\_West** to see if there are any issues there that you need to fix.

## Clean Orders\_West

As you look over the fields for the **Orders\_West** file, most of the fields look like they align with the other files, but you remember seeing that the **States** field used abbreviations for the values instead of spelling out the state name. To combine this file with the other files, you'll need to fix this. So you add a cleaning step to the **Orders\_West** Input step.

1. In the **Flow** pane, select **Orders\_West**, click the plus  icon and select **Add Step** or **Add Clean Step**, depending on your version.

Scroll or use Search to find the **State** field. You see that all the state name values use the short abbreviation. There are only 11 unique values for this field. You could manually change each one, but maybe Tableau Prep Builder has another way to do this?

You click the **More options**  menu (drop-down arrow in prior releases) for the field and see an option called **Group and Replace**. When you select it you see several options:

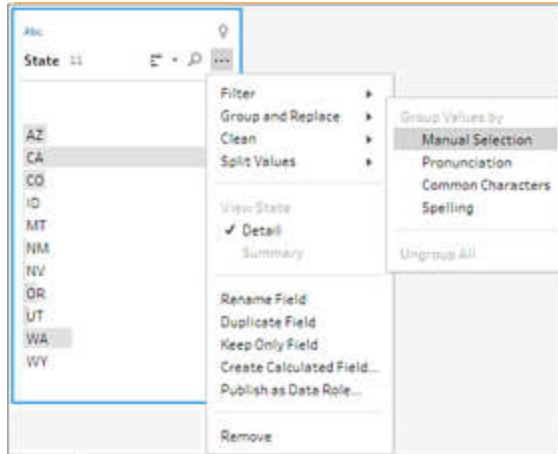
- Manual Selection
- Pronunciation
- Common Characters
- Spelling

The state names don't sound alike, they aren't spelled incorrectly, and they don't share the same characters, so you decide to try the **Manual Selection** option.

**Tip:** You can double-click a field name or field value to edit a single value. To edit multiple values you can select all the values and use the right-click menu option **Edit Values**. But when you want to map one or more values to specific values, use the **Group and Replace** option in the drop-down menu.

For more information about editing and grouping values, see [Edit field values on page 246](#).

2. Select the **State** field. Click the drop-down arrow and select **Group and Replace** > **Manual Selection**.



A two column card opens. This is the **Group and Replace editor**. The column on the left shows the current field values and the column on the right shows the fields that are available to map to the fields on the left.

You want to map your state abbreviations to the spelled out version of the state name, but you don't have those values in the **Orders\_West** data set. You wonder if you can just edit the name directly and maybe add it there, so you give that a try.

3. In the **Group and Replace editor** in the left pane, double-click **AZ** to highlight the value and type **Arizona**. Then press **Enter** to add your change.

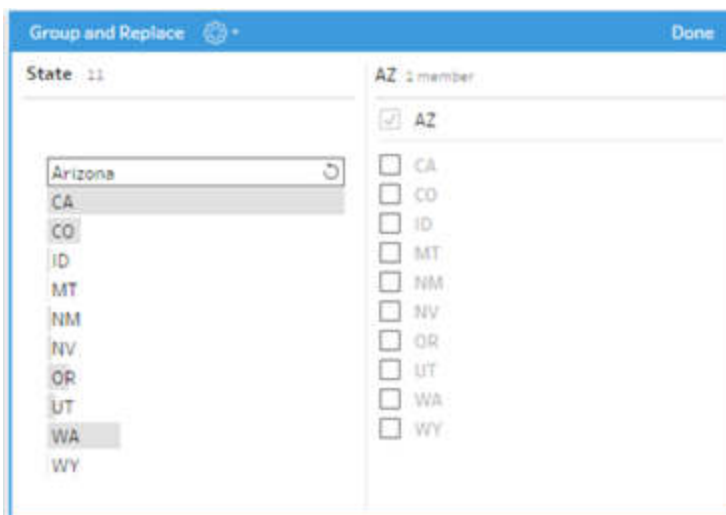


Tableau Prep Builder created a mapped value for your new value **Arizona** and automatically mapped the old value, **AZ** to it. Having a mapped relationship set up for these values will save you time if you get more data from this region entered like this.

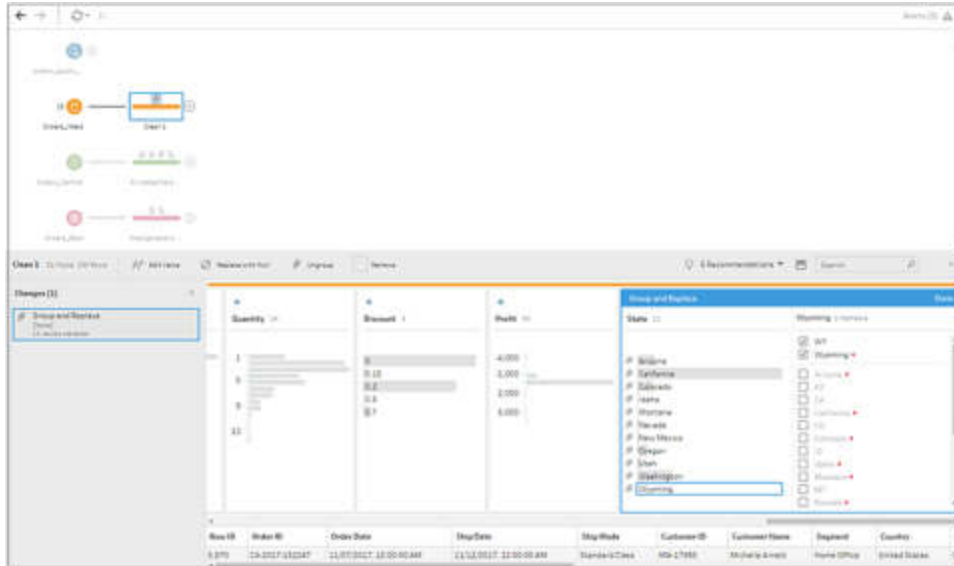
**Tip:** You can add field values that aren't in your data sample to set up mapping relationships to organize your data. If you refresh your data source and new data is added, you can add the new data to the mapping instead of manually fixing each value.

When you manually add a value that isn't in your data sample, the value is marked with a red dot to help you easily identify it.

4. Repeat these steps to map each state to the spelled out version of its name.

<b>Abbreviation</b>	<b>State Name</b>
AZ	Arizona
CA	California
CO	Colorado
ID	Idaho
MT	Montana
NM	New Mexico
NV	Nevada
OR	Oregon
UT	Utah
WA	Washington
WY	Wyoming

Then click **Done** to close the **Group and Replace** editor.



After all the states are mapped, you look at the **Changes** pane and see there is only one entry there instead of 11.

It looks like Tableau Prep Builder groups similar actions for a field together. You like that because it will make it easier to find changes you made to your data set later.

Fixing the **State** field values was the only change you needed to make here.

5. Name your cleaning step to keep track of your work. For example **Rename states**.

You've done a lot of clean up in your files, and you can't believe how quick and easy it was. You might make it home for dinner after all! To make sure that you don't lose all of your work so far, save your flow.

Click **File > Save** or **File > Save As**. Save your file as a flow file (.tfl) and give it a name. For example, **My Superstore**.

**Tip:** When you save your flow files, you can either save them as a flow file (.tfl) or you can save them as a packaged file (.tflx) and package your local data files with them to share the flow and files with someone else. For more information about saving and sharing your flows, see [Save and Share Your Work](#) on page 321.

## 4. Combine your data

Now that all the files are cleaned up, you are finally ready to combine them all.

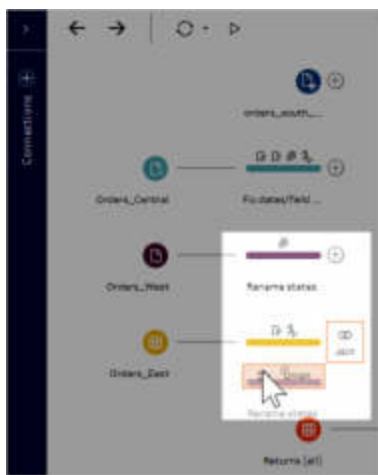


Because all the files have similar fields after your clean up efforts, to pull all the rows together into a single table, you need to union the tables.

You remember that there was a step option called **Add Union**, but you wonder if you can simply drag and drop the steps to union them. You decide to try it and see.

## Union your data

1. In the **Flow** pane, drag the second to last cleaning step on top of the last cleaning step in your flow and drop it on the **Union** option. For example drag the cleaning step **Rename states** on to the cleaning step **Changed data type** step and drop it on the **Union** option.



You see that Tableau Prep Builder added a new **Union** step to your flow. Great! Now you want to add the other files to this union too.

2. Drag the second cleaning step in the flow on to the new **Fix dates/field names** step on to the new **Union** step, then drop it on **Add** to add it to the existing union. For example, drag the **Fix dates/field names** cleaning step on to the new **Union** step then drop it on **Add** to add it to the existing union.






The only field that doesn't have a match now is the **File Paths** field. This field shows the file paths for the wildcard union that you did for your sales orders from the South. You decide to leave this field there as it has good information.

**Tip:** You have several options when fixing mismatched fields after a union.

Depending on the Tableau Prep Builder version you are using, you can select a field in the **Mismatched Fields** list (version 2018.2.1 and later) and if Tableau Prep Builder detects a possible match, it will highlight it in yellow. To merge the fields hover over the

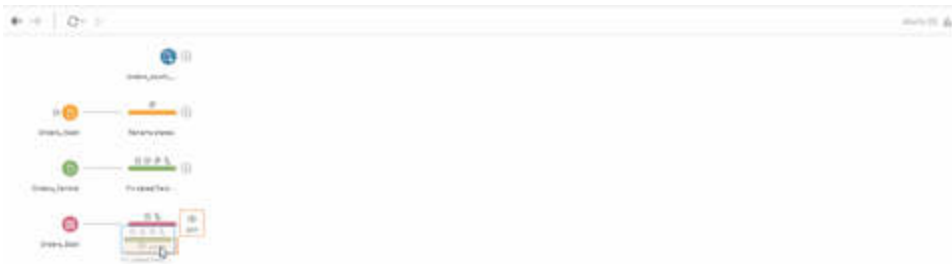
highlighted field and click the plus  button that appears.

For more ways to merge fields in a union, see [Fix fields that don't match on page 317](#).

7. Clear the **Show only mismatched fields** check box to show all the fields included in the union.
8. Name your Union step to represent what this union includes work. For example, **All orders**.

## Check your work: Watch "Union your data" in action.

*Click the image to replay it*



You are a cleaning genius! As you are admiring your results, your boss calls. He forgot to mention that he also wants you to include any product returns in your analysis. He hopes that won't be too much trouble. With Tableau Prep Builder in your toolkit, it's no problem at all!

## Clean the product returns data

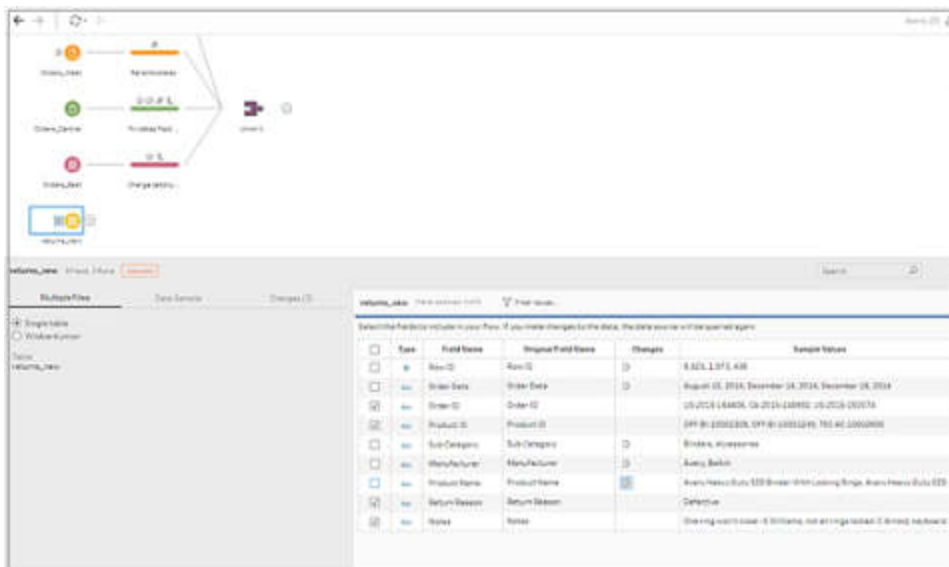
You look over the Excel file that your boss sent you for product returns and it looks a little messy. You add the new file **return\_reasons\_new** to your flow to take a closer look.

1. In the **Connections** pane, click **Add connection**. Select **Microsoft Excel** and navigate to the sample Superstore data files (see [Wrap up and resources on page 132](#) for the file location).

2. Select **return reasons\_new.xlsx**, and then click **Open** to add the file to the flow pane.

There are only 4 fields that you want to include from this file in your flow: **Order ID**, **Product ID**, **Return Reason** and **Notes**.


3. In the **Input** pane for **returns\_new** clear the check box at the top of the field grid to clear all the check boxes. Then select the check box for the **Order ID**, **Product ID**, **Return Reason** and **Notes** fields.



4. Rename the Input step to better reflect the data that is included in this input. In the **Flow** pane, double-click the Input step name **Returns\_new** and type in **Returns (all)**.

Looking at the sample field values, you notice that the **Notes** field seems to have a lot of different data combined together.

You have some cleaning to do in this file before you can do any further work with the data, so you add a cleaning step to check it out.

5. In the **Flow** pane, select the Input step **Returns (all)**, click the plus  icon, and then select **Add Step** or **Add Clean Step**, depending on your version.

In the **Profile** pane, click and drag the outer right edge of the field to the right to re-size the **Notes** field so you can see the entries better.

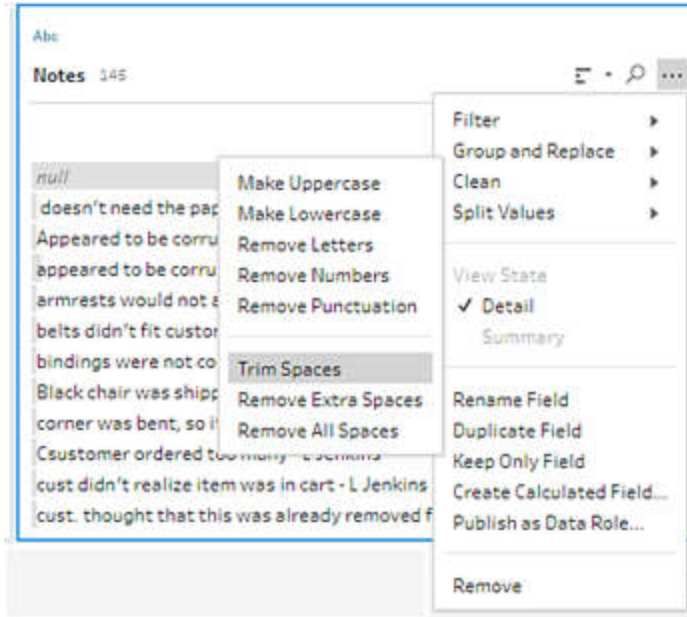
6. In the **Notes** field, use the visual scroll bar to the right of the field values to scan the values.

You notice a few things that are problematic:

- Some of the entries have an extra space in the entry. This can result in the field being read as a null value.
- It looks like the name of the approver is included in the return notes entry. To better work with this data you'll want that information in a separate field.

To tackle the extra spaces, you remember that there was a cleaning option to remove trailing spaces, so you decide to try that to see if it can fix that problem.

7. Select the **Notes** field. Click the **More options**  menu (drop-down arrow in prior releases) and select **Clean > Trim Spaces**.

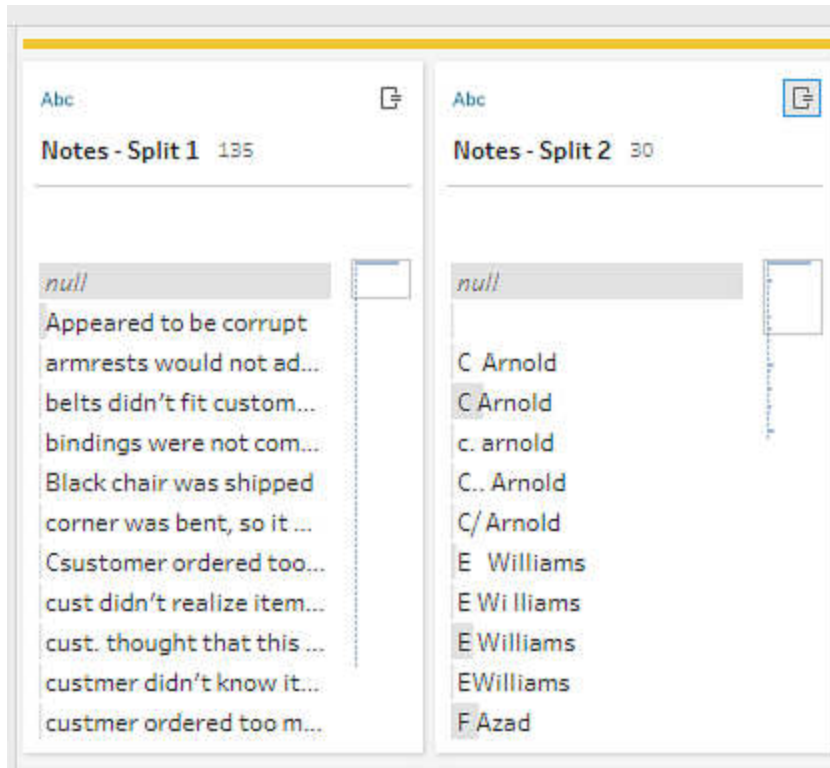


Yes! It did exactly what you wanted it to do. The extra spaces are gone.

Next you want to create a separate field for the approver name. You see a **Split Values** option in the menu, so you decide to try that.

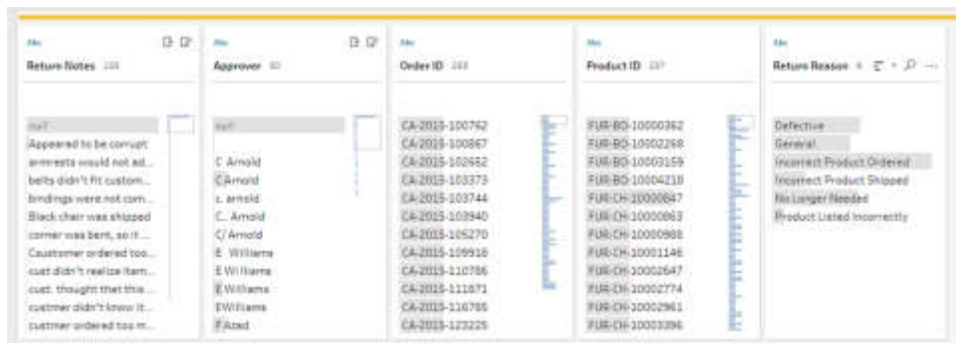
8. Select the **Notes** field. Click the **More options**  $\dots$  menu (drop-down arrow in prior releases) and select **Split Values > Automatic Split**.

This option did exactly what you were hoping it would do. It automatically split the return notes and the approver name into separate fields.



Just like Tableau Desktop, Tableau Prep Builder automatically assigned a name to those fields. So you'll need to rename the new fields to something meaningful.

9. Select the field **Notes-Split 1**. Double-click in the field name and type **Return Notes**.
10. Repeat this step for the second field and rename it to **Approver**.
11. Finally remove the original **Notes** field, as you no longer need it. Select the **Notes** field, click the **More options**  $\cdots$  menu (drop-down arrow in prior releases), and select **Remove** (**Remove Field** in prior versions) from the menu.



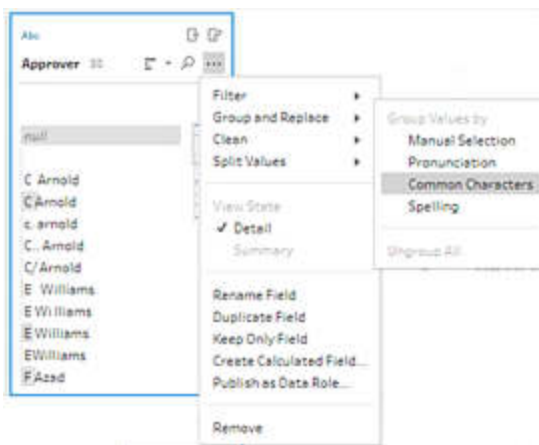


Looking at the new **Approver** field, you notice that the field values lists the same names but they are entered differently. You want to group them to eliminate multiple variations of the same value.

Maybe the **Group and Replace** option can help with that?

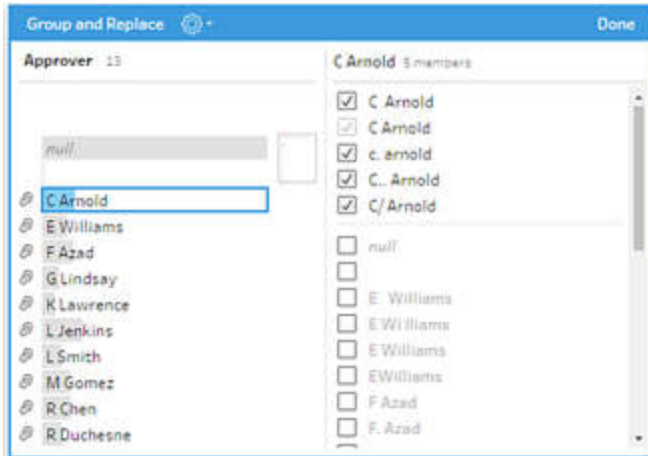
You remember there was an option for **Common Characters**. Since these values share the same letters, you decide to try that.

12. Select the **Approver** field. Click the **More options** ... menu (drop-down arrow in prior releases) and select **Group and Replace > Common Characters**.



This option grouped all of the variations of each name together for you. That's exactly what you wanted to do.

After checking the other names to make sure they are grouped properly, you click **Done** to close the **Group and Replace editor**.



This file is looking pretty good.

13. Name your cleaning step to keep track of your work. For example **Cleaned notes**.

Now that the product return data is all cleaned up, you want to add this data to the orders data from your unioned files. But many of these fields don't exist in the unioned files. To add these fields (columns of data) to your unioned data set, you need to use a join.

## Join your data

When you join data, the files must have at least one field in common. Your files share the **Order ID** and **Product ID** fields, so you can join on those fields to see all the rows that have those fields in common. You remember an option to create a join when you created your union using drag and drop, so you give that a try.

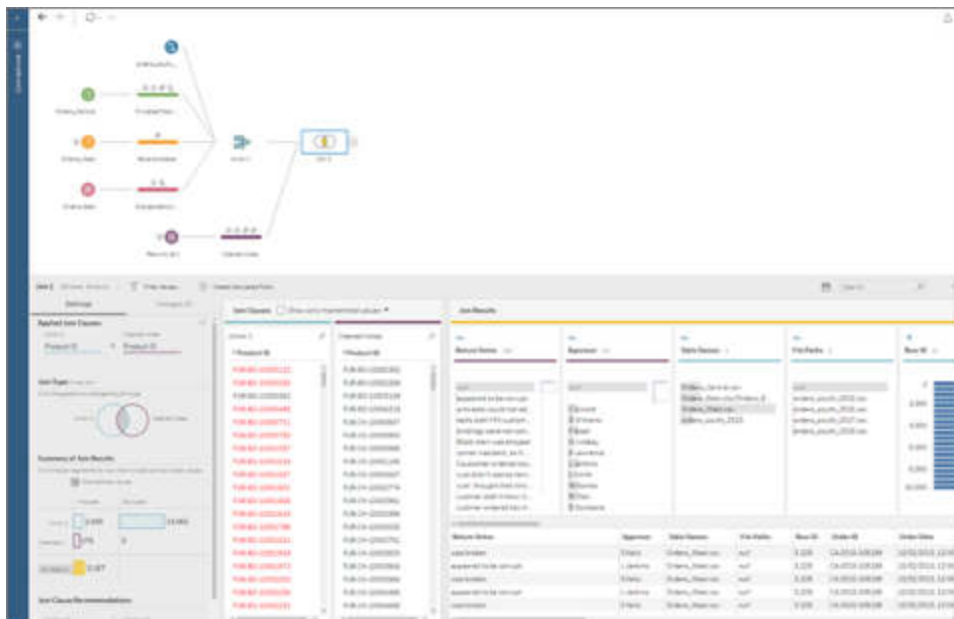
1. In the **Flow** pane, drag the **Cleaned notes** step on to the **All orders** Union step and drop it on **Join**.



When you join files, Tableau Prep Builder shows you the results of your join in the **Join Profile**.

Working with joins can be tricky. You often want to have a clear view of the factors that are included in the join, such as the fields used to join the files, the number of rows included in the results and any fields that aren't included or are null values.

As you review the results of the join in Tableau Prep Builder, you are delighted to see so much information and interactivity at your fingertips.



**Tip:** The far left pane of the join profile is where you can explore and interact with your join. You can also fix values directly in the **Join Clauses** panes and perform cleaning operations in the Join Results pane.


Choose the fields that you want to join on in the **Applied Join Clauses** section or add suggested join clauses from the **Join Clause Recommendations** section.

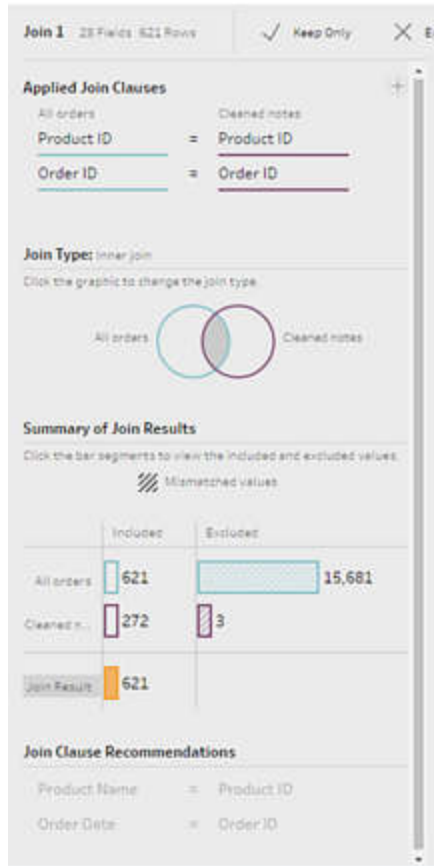
Click in the **Join Type** diagram to try different join configurations and see the number of rows included or excluded in your join for each table in the **Summary of Join Results** section.

For more information about working with joins, see [Aggregate, Join, or Union Data on page 307](#).

You see that you have over 13,000 rows excluded from your **All Orders** files. When you created your join, Tableau Prep Builder automatically joined on the **Product ID** field, but you wanted to also join on the **Order ID** field.

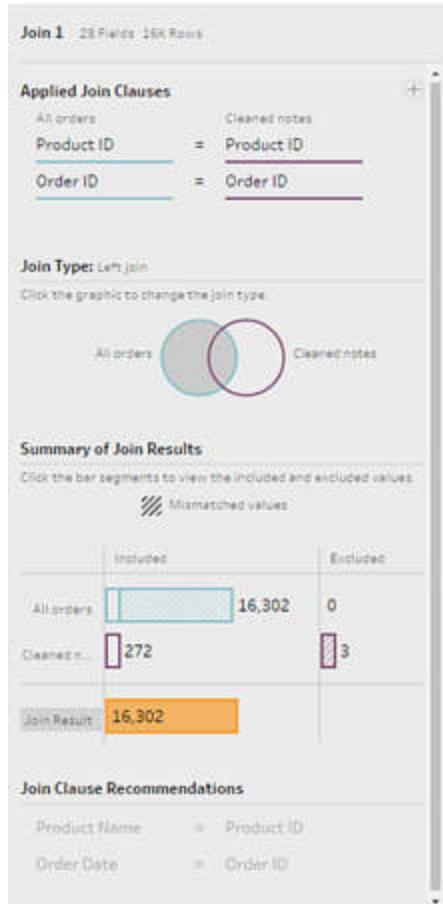
As you scan the left pane of the join profile, you see that **Order ID** is in the list of recommended join clauses, so you quickly add it from there.

2. In the left pane of the **Join** profile, in the **Join Clause Recommendations** section, select **Order ID = Order ID** and click the plus  button to add the join clause.



Because the **Join Type** is set to an inner join (the default setting for Tableau Prep Builder), the join is only including values that exist in both files. But you want all of the data from your **Orders** files as well as the return data for those files. So you'll need to change the join type.

3. In the **Join Type** section, click the side of the diagram to include all orders. In the example below, click the left side of the diagram to change the join type to a **Left** join and include all data from the **All orders** union step and any matching data from the **Cleaned notes** step.



Now you have all of the data from the sales order files and any return data that apply to those orders. You review the **Join Clauses** pane and see the distinct values that don't exist in the other file.

For example there are many order rows (shown in red) that have no corresponding return data. You love being able to explore this level of detail about your join.

You're anxious to start analyzing this data in Tableau Desktop, but you notice a few results from the join that you want to clean up before you do that. Good thing you know what to do!

**Tip:** Wonder if your data is clean enough? You can preview your data in Tableau Desktop from any step in your flow to check it out.

Just right-click on the step in the **Flow** pane and select **Preview in Tableau Desktop** from the menu.

You can experiment with your data and any changes that you make in Tableau Desktop

won't write back to your data source in Tableau Prep Builder. For more information see [View your data sample in Tableau on page 321](#).

4. Before you start cleaning your join results, name your **Join** step **Orders+Returns** and save your flow.


## Clean your join results

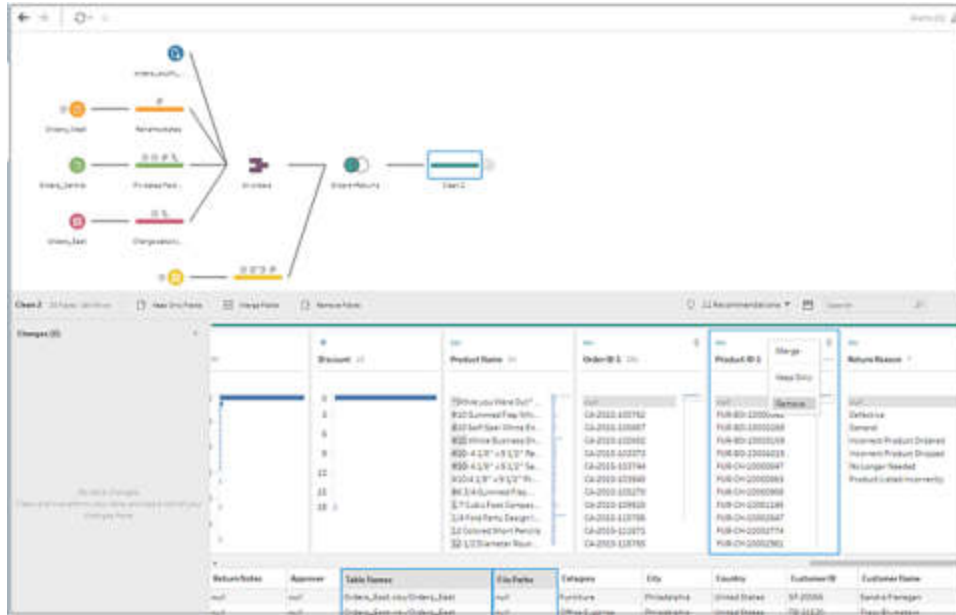
**Note:** To clean up the fields in your join, you can perform cleaning operations directly in the Join step. For the purposes of this tutorial we will add a cleaning step so you can clearly see your cleaning operations. If you want to try performing these steps directly in the join step skip steps 1 and 3 below.

When you joined the two steps, the common fields Order ID and Product ID were added for both tables.

You want to keep the Product ID field from all of your orders and the Order ID field from the returns file and remove the duplicate fields that came from those files. You also don't need the **File Paths** and **Table Names** fields in your output file, so you want to remove those fields as well.

**Tip:** When you join tables using fields that exist in both files, Tableau Prep Builder will bring in both fields and rename the duplicate field from the second file by adding a "-1" or a "-2" to the field name. For example Order ID and Order ID-1.

1. In the **Flow** pane, select **Orders+Returns**, click the plus  icon, and then select **Add Step** or **Add Clean Step**, depending on your version.
2. In the **Profile** pane, select and remove the following fields:
  - **Table Names**
  - **Order ID**
  - **File Paths**
  - **Product ID-1**



3. Rename the field **Order ID-1** to **Order ID**.

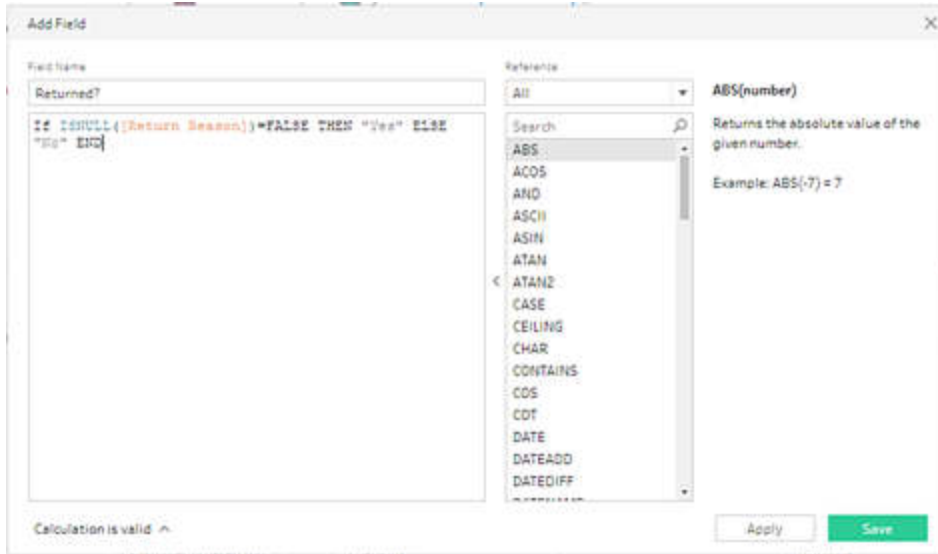
You have quite a few null values where the product was returned but there was no return note or approver indicated. To make this data easier to analyze, you want to add a field with a value of **Yes** and **No** to indicate whether the product was returned.

You don't have this field, so you can add it by creating a calculated field.

4. In the toolbar, click **Create Calculated Field**.
5. Name the field **Returned?** and then enter the following calculation and click **Save**.

```
If ISNULL([Return Reason])=FALSE THEN "Yes" ELSE "No" END
```



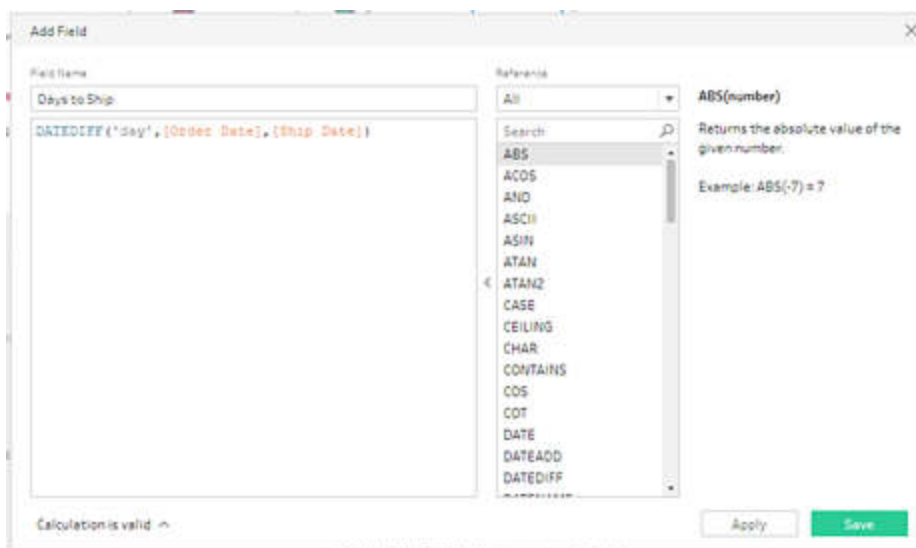


For your analysis you would also like to know the number of days it takes to ship an order, but you don't have that field either.

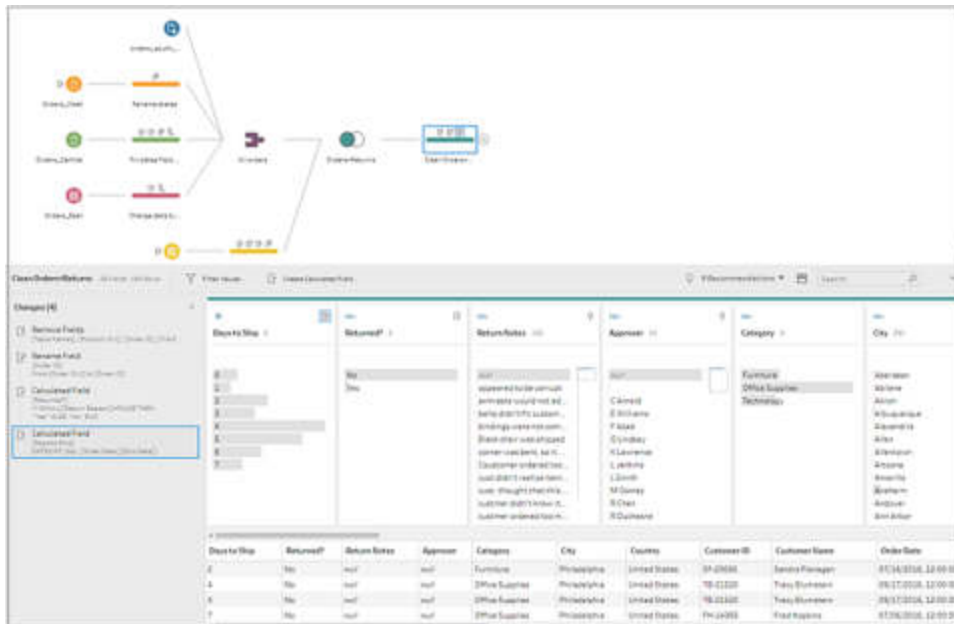
You have all the information that you need to create it though, so you add another calculated field to create it.

6. In the toolbar, click **Create Calculated Field**.
7. Name the field **Days to Ship** and then enter the following calculation and click **Save**.

`DATEDIFF('day', [Order Date], [Ship Date])`



8. Name your step **Clean Orders +Returns**.
9. Save your flow.



## 5. Run your flow and generate output

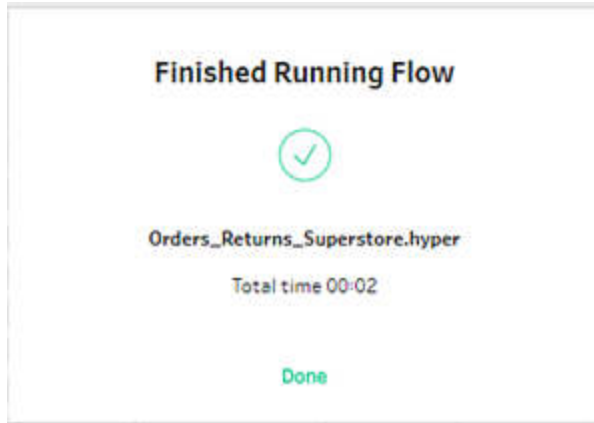
Your data is looking good and you're ready to generate your output file to start analyzing it in Tableau Desktop. All you need to do is run your flow and generate your extract file. To do this you need to add an **Output** step.

1. In the **Flow pane**, select **Clean Orders+Returns**, click the plus **+** icon and select **Add Output**,

When you add an Output step, the **Output** pane opens and shows you a snapshot of your data. Here you can select the type of output that you want to generate, and specify the name and where you want to save the file.

The default location is in the **My Tableau Prep Builder** repository in your data sources folder.





To keep your data fresh, you can run the flow manually here in Tableau Prep Builder or using the command line. You can also publish the flow to Tableau Server or Tableau Online to run your flow on a schedule. For more information about running flows using the command line, see [Refresh flow output files from the command line on page 334](#). For more information about running flows on a schedule, see [Publish a Flow to Tableau Server or Tableau Online on page 398](#).

## Wrap up and resources

You are a data prep rock star! You took dirty data and transformed it with ease! In no time, you cleaned and prepped your data from multiple data sets and turned it into a sleek, clean data set that you can now work with in Tableau Desktop to do your analysis.

Want more practice? Try replicating the rest of the sample flow for Superstore using the data files found here:

- [Orders\\_South\\_2015](#)
- [Orders\\_South\\_2016](#)
- [Orders\\_South\\_2017](#)
- [Orders\\_South\\_2018](#)
- [Orders\\_Central](#)
- [Orders\\_East](#)
- [Orders\\_West](#)
- [returns\\_reasons\\_new](#)
- [Quota](#)

You can also find the files in the following location on your computer after installing Tableau Prep Builder:

## Tableau Prep Help

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en\_US\Superstore Files
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en\_US/Superstore Files

Want more training? Check out the [training videos](#) for Tableau Prep Builder, check out these great [resources](#), or take an [in-person training](#) course.

Want more information about the topics we covered? Check out the other topics in the Tableau Prep Builder online help.

# About Tableau Prep Builder

Tableau Prep Builder is a tool in the Tableau product suite designed to make preparing your data easy and intuitive. Use Tableau Prep Builder to combine, shape, and clean your data for analysis in Tableau.

**Note:** Tableau Prep version 2019.1.2 had changed its name to Tableau Prep Builder. If you are using an earlier version of the product you will still see "Tableau Prep" in the product.

## Using Tableau Prep Builder

Start by connecting to your data from a variety of files, servers, or Tableau extracts. Connect to and combine data from multiple data sources. Drag and drop or double-click to bring your tables into the flow pane, and then add flow steps where you can then use familiar operations such as filter, split, rename, pivot, join, union and more to clean and shape your data.

Each step in the process is represented visually in a flow chart that you create and control. Tableau Prep Builder tracks each operation so that you can check your work and make changes at any point in the flow.

When you are finished with your flow, run it to apply the operations to the entire data set.

Tableau Prep Builder works seamlessly with other Tableau products. At any point in your flow, you can create an extract of your data, publish your data source to Tableau Server or Tableau Online, publish your flow to Tableau Server or Tableau Online to refresh your data using a schedule or open Tableau Desktop directly from within Tableau Prep Builder to preview your data.

For information about installing Tableau Prep Builder, see [Install Tableau Desktop or Tableau Prep Builder](#) in the Tableau Desktop and Tableau Prep Deployment Guide.

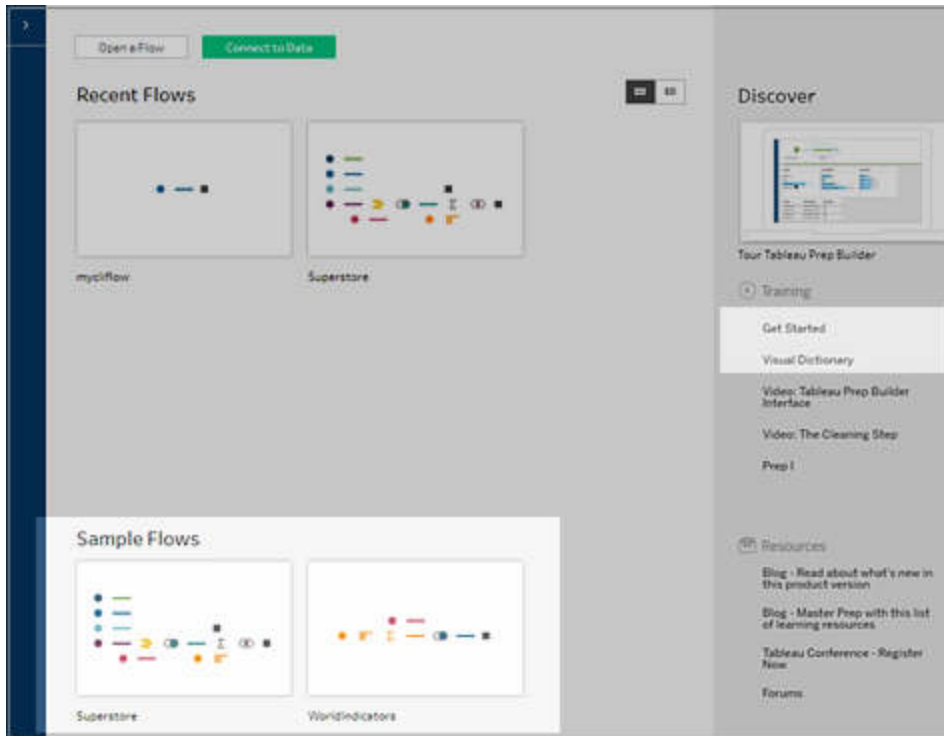
## See Tableau Prep Builder in action

Click the image to replay it.

The screenshot displays the Tableau Prep Builder interface. On the left, a sidebar shows a list of connections and tables. The main workspace features a flow diagram with three input tables (Orders\_2016, Orders\_2017, Orders\_2018) feeding into a 'Join' step, which then feeds into a 'Filter' step. Below the flow, the 'Output' pane shows a table of data with columns: Table Name, File Path, Sales, Quantity, Profit, Region, State, Row ID, Order ID, and Order Date. The table contains 18 rows of data representing orders from various states and years.

Table Name	File Path	Sales	Quantity	Profit	Region	State	Row ID	Order ID	Order Date
Orders_Central_2016	...	89.61	8	123.888	Central	Texas	15	05-0538-118993	11/02/2016, 12:00:00 AM
Orders_Central_2016	...	3.844	3	-1.838	Central	Texas	16	05-0538-118993	11/02/2016, 12:00:00 AM
Orders_Central_2016	...	863.80	6	23.2176	Central	Wisconsin	17	0A-2023-105893	11/11/2016, 12:00:00 AM
Orders_Central_2016	...	29.46	7	6.9896	Central	Nebraska	22	0A-2027-137330	12/04/2017, 12:00:00 AM
Orders_Central_2016	...	89.34	7	25.8834	Central	Nebraska	23	0A-2027-137330	12/05/2017, 12:00:00 AM
Orders_Central_2016	...	29.472	3	9.9488	Central	Texas	35	0A-2028-107037	10/19/2018, 12:00:00 AM
Orders_Central_2016	...	1,087.644	7	123.8737	Central	Texas	36	0A-2027-117940	12/02/2017, 12:00:00 AM
Orders_Central_2016	...	136.82	5	-147.953	Central	Texas	37	0A-2027-117940	12/02/2017, 12:00:00 AM
Orders_Central_2016	...	212.338	5	55.425	Central	Texas	38	0A-2028-117418	12/07/2018, 12:00:00 AM
Orders_Central_2016	...	310.288	2	-46.9764	Central	Texas	39	0A-2028-117418	12/07/2018, 12:00:00 AM
Orders_Central_2016	...	212.338	3	-15.247	Central	Texas	40	0A-2028-117418	12/07/2018, 12:00:00 AM
Orders_Central_2016	...	371.168	4	41.7884	Central	Texas	41	0A-2026-117412	12/07/2016, 12:00:00 AM
Orders_Central_2016	...	147.268	4	28.884	Central	Illinois	42	0A-2029-120949	09/10/2019, 12:00:00 AM
Orders_Central_2016	...	46.88	2	24.774	Central	Minnesota	43	0A-2027-118213	03/11/2017, 12:00:00 AM
Orders_Central_2016	...	17.46	3	6.2062	Central	Minnesota	44	0A-2027-118213	03/11/2017, 12:00:00 AM
Orders_Central_2016	...	211.90	6	8.4786	Central	Illinois	47	0A-2025-148763	10/30/2015, 12:00:00 AM

Ready to try it out? From the **Start** page, click on one of the sample flows to explore and experiment with the steps, try the [Get Started with Tableau Prep Builder](#) on page 91 hands-on tutorial to learn how to create a flow or try stepping through one of our [Day in the Life Scenarios](#) on page 443 using Tableau Prep Builder.



**Note:** You can find the sample data files used in the flows in these locations:

- (Windows) C:\Program Files\Tableau\Tableau Prep Builder <version>\help\Samples\en\_US
- (Mac) /Applications/Tableau Prep Builder <version>.app/Contents/help/Samples/en\_US

To learn more about Tableau Prep Builder and the different features and functions it offers, review the topics in this guide.

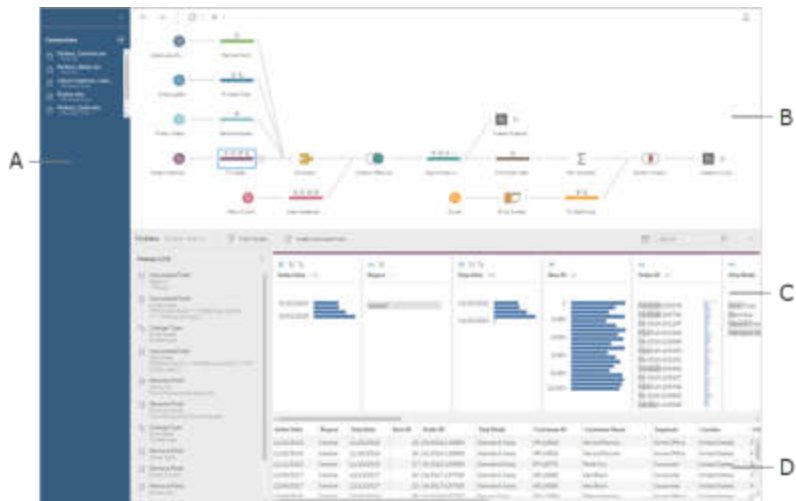
## A tour of the Tableau Prep Builder workspace

The Tableau Prep Builder workspace consists of the **Connections pane** (A) where you connect to your data sources, and three coordinated areas that help you interact with and explore your data:

- **Flow pane** (B): A visual representation of your operation steps as you prepare your data. This is where you add steps to build your flow.
- **Profile pane** (C): A summary of each field in your data sample. See the shape of your

data and quickly find outliers and null values.

- **Data grid (D):** The row level detail for your data.

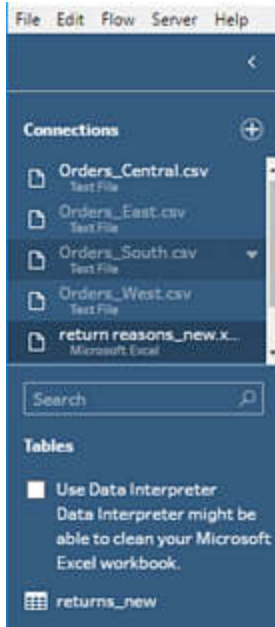


After you connect to your data and begin building your flow, you add steps in the **Flow** pane. These steps function as a lens into the structure of your data, as well as a summary of operations that is applied to your data. Each step represents a different category of operations that you define, all as part of your flow.

## Connections pane

On the left side of the workspace is the **Connections** pane, which shows the databases and files you are connected to. Add connections to one or more data sources and then drag the tables you want to work with into the **Flow** pane. For more information see [Connect to Data on page 145](#).



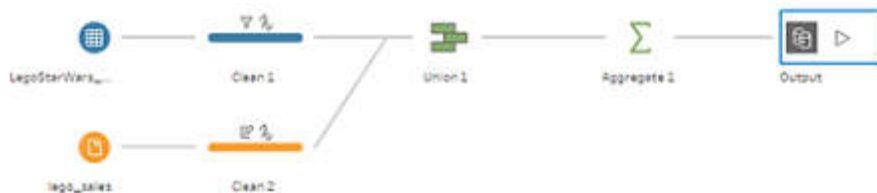


You can minimize the **Connections** pane if you need more room in your workspace.

## Flow pane

At the top of the workspace is the **Flow** pane. This is your canvas. As you connect to, clean, shape, and combine your data, steps appear in the **Flow** pane and align from left to right along the top. These steps tell you what kind of operation is being applied, in what order, and how your data is affected by it. For example, the Join step shows you which join type you've applied, the join clauses, recommended join clauses, and the fields of the tables that are included in the join.

You start your flow by dragging tables into the **Flow** pane. Here you can add additional data sets, pivot your data, union or join data, create aggregations, and generate output files in the form of .tde and .tds files or Hyper extract (.hyper) files that you can use in Tableau. For more information about generating output files, see [Save and Share Your Work](#) on page 321.



**Note:** If you make changes to the data while in Tableau Desktop, for example renaming fields, changing data types, and so on, these changes aren't written back to Tableau Prep Builder.

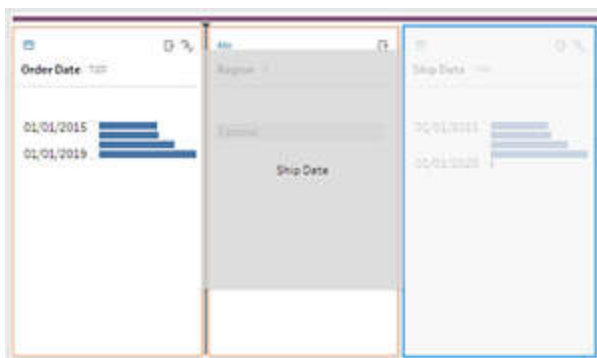
## Profile pane

In the center of the workspace is the **Profile** pane. The **Profile** pane shows you the structure of your data at any point in the flow. The structure of your data can be represented in different ways depending on the operation you want to perform on your data or the step that you select in the **Flow** pane.

At the top of the **Profile** pane is a toolbar that shows you the cleaning operations that you can perform for each step in your flow. An options menu also appears on each card in the **Profile** pane where you can select the different operations that you can perform on the data.

For example:

- Search, sort, and split fields
- Filter, include, or exclude values
- Find and fix null values
- Rename fields
- Clean up data entry errors using group and replace or quick cleaning operations
- Use automatic data parse to change data types
- Rearrange the order of your field columns by dragging and dropping them where you want them



Select one or more field values in a Profile card and right-click or Ctrl-click (MacOS) to see additional options to keep or exclude values, group selected values or replace values with Null.

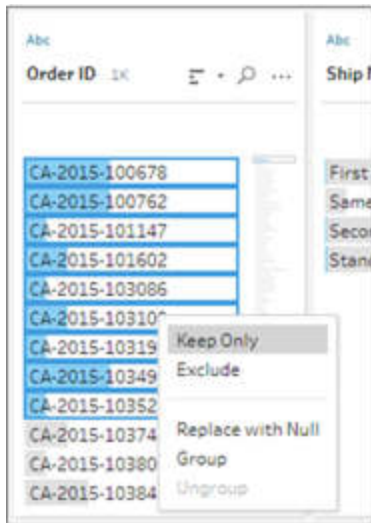
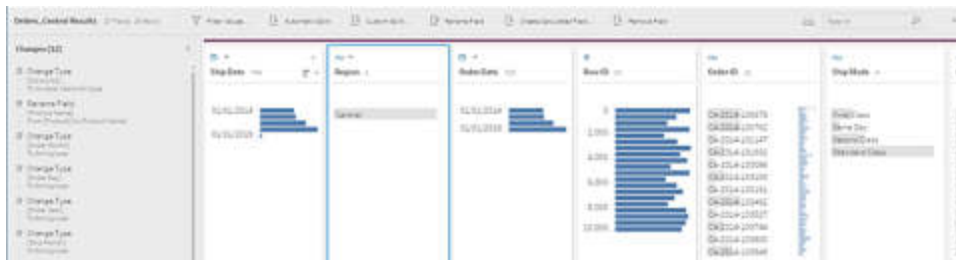


Tableau Prep Builder keeps track of any changes you make, in the order you make them, so you can always go back and review or edit those changes if needed. Use drag and drop to re-order those operations to experiment and apply changes in a different order.




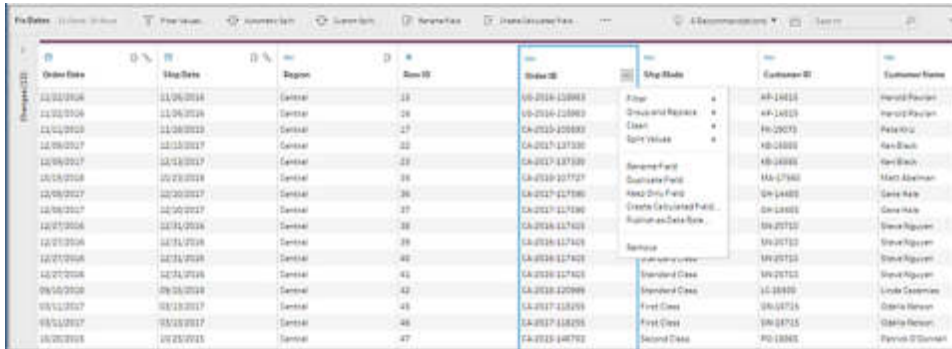
Click the arrow on the upper right of the pane to expand and collapse the **Changes** pane for more room to work with the data in the **Profile** pane.

For more information about applying cleaning operations to your data see [Clean and Shape Data](#) on page 229.

## Data grid

At the bottom of the workspace is the **Data** grid, which shows you the row level detail in your data. The values displayed in the **Data** grid reflect the operations defined in the **Profile** pane. You can perform the same cleaning operations here as you can in the Profile pane if you prefer to work at a more detailed level.

Click the **Collapse Profiles**  icon on the toolbar to collapse (and expand) the **Profile** pane to see your options.



Order Date	Ship Date	Region	Row ID	Order ID	Ship State	Customer ID	Customer Name
11/22/2016	11/26/2016	Central	25	US-2016-128863	Filter	AK-14515	Harold Pearson
11/22/2016	11/26/2016	Central	26	US-2016-128863	District of Columbia	AK-14515	Harold Pearson
11/22/2016	11/26/2016	Central	27	CA-2016-128863	Class	AK-14515	Peter King
12/09/2017	12/13/2017	Central	32	CA-2017-127330	San Vicente	KB-14600	Ron Black
12/09/2017	12/13/2017	Central	33	CA-2017-127330	Riverside	KB-14600	Ron Black
12/15/2016	12/19/2016	Central	35	CA-2016-127721	Dutchess	MA-17160	Mark Abraham
12/09/2017	12/20/2017	Central	36	CA-2017-127380	Hand On	SN-14400	Gina Hale
12/09/2017	12/20/2017	Central	37	CA-2017-127380	Green Calculated Field	SN-14400	Gina Hale
12/07/2016	12/11/2016	Central	38	CA-2016-127401	Support on Data Row	SN-20710	Shawn Nguyen
12/07/2016	12/11/2016	Central	39	CA-2016-127401	Berkshire	SN-20710	Shawn Nguyen
12/07/2016	12/11/2016	Central	40	CA-2016-127401	Standard Class	SN-20710	Shawn Nguyen
12/07/2016	12/11/2016	Central	41	CA-2016-127401	Standard Class	SN-20710	Shawn Nguyen
09/10/2016	09/16/2016	Central	42	CA-2016-120989	Standard Class	LC-18800	Linda Coleman
05/11/2017	05/17/2017	Central	45	LA-2017-118255	First Class	SN-14715	Clara Nelson
05/11/2017	05/17/2017	Central	46	LA-2017-118255	First Class	SN-14715	Clara Nelson
05/20/2016	05/23/2016	Central	47	CA-2016-148702	Second Class	PG-13800	Patrick O'Donnell

## How Tableau Prep Builder stores your data

When you connect Tableau Prep Builder to your data and create a flow, it stores the frequently used data in a .hyper file. For large data sets, this might be a sample of the data. Any stored data is saved under a secure temporary file directory in a file named Prep BuilderXXXXX, where XXXXX represents a universally unique identifier (UUID). After you save the flow, the file is deleted. For more information about how Tableau Prep Builder samples your data, see [Set your data sample size on page 169](#).

Tableau Prep Builder also saves data in the Tableau flow (.tfl) file to support the following operations (which can capture entered data values):

- Custom SQL used in Input steps
- Filtering (on data entry)
- Group and replace (on data entry)
- Calculations

# Tableau Prep Visual Dictionary



























































1







































**Input Step**

Start your flow by dragging data to the Flow pane to create an Input step. The icon shows you the type of data source.

- Data Source
- Data Source with Wildcard Union
- Excel
- Excel with Wildcard Union
- CSV
- CSV with Wildcard Union
- Tableau Extract

<p>2</p>	<p><b>Clean Step, Changes Pane, and Toolbar</b></p> <p>Easily keep track of the changes you make to your data. Find these icons on the Clean steps in the Flow pane, in the Changes pane, and on the toolbar.</p> <table border="0"> <tr> <td> Calculated Field</td> <td> Hide Profile Pane</td> </tr> <tr> <td> Change Data Type</td> <td> Show Profile Pane</td> </tr> <tr> <td> Edit Value</td> <td> Merge Fields</td> </tr> <tr> <td> Exclude Values</td> <td> Remove Field</td> </tr> <tr> <td> Filter Values</td> <td> Rename Field</td> </tr> <tr> <td> Group Values</td> <td> Search</td> </tr> <tr> <td> Keep Only</td> <td> Split Fields</td> </tr> </table>	 Calculated Field	 Hide Profile Pane	 Change Data Type	 Show Profile Pane	 Edit Value	 Merge Fields	 Exclude Values	 Remove Field	 Filter Values	 Rename Field	 Group Values	 Search	 Keep Only	 Split Fields
 Calculated Field	 Hide Profile Pane														
 Change Data Type	 Show Profile Pane														
 Edit Value	 Merge Fields														
 Exclude Values	 Remove Field														
 Filter Values	 Rename Field														
 Group Values	 Search														
 Keep Only	 Split Fields														
<p>3</p>	<p><b>Union Step</b></p> <p>Combine up to ten sources of data with similar fields in a single Union step.</p> <p> Union Data</p>														
<p>4</p>	<p><b>Join Step</b></p> <p>Connect two steps to join your data on one or more common fields. Select one of these options to choose the join type.</p> <ul style="list-style-type: none"> <li> Full Anti Join</li> <li> Inner Join</li> <li> Left Inner Join</li> <li> Left Outer Join</li> <li> Full Outer Join</li> <li> Right Inner Join</li> <li> Right Outer Join</li> </ul>														

<p>5</p>	<p><b>Output Step</b></p> <p>Add an Output step and run your flow to apply the changes to the complete data set and generate output files.</p> <ul style="list-style-type: none"><li> CSV</li><li> Published Data Source</li><li> Tableau Data Extract</li><li> Run Flow</li></ul>
<p>6</p>	<p><b>Pivot Step</b></p> <p>Add a Pivot step to change columns to rows.</p> <ul style="list-style-type: none"><li> Pivot Data</li></ul>
<p>7</p>	<p><b>Aggregate Step</b></p> <p>Add an Aggregate step to group and aggregate your data, which changes the level of detail of your data.</p> <ul style="list-style-type: none"><li> Aggregate Data</li></ul>

<p>8</p>	<p><b>Notification</b></p> <p>If there's a problem with your flow or something you need to know, check notifications. Errors include a <a href="#">Go to Error</a> link to help you quickly find the problem.</p> <ul style="list-style-type: none"> <li> No Notifications</li> <li> Notification Alert</li> <li> Error in the Step</li> </ul>																		
<p>9</p>	<p><b>Profile Pane</b></p> <p>See the exact row count of your data and know when your data is sampled.</p> <p> Only shows when data is sampled</p> <p> Hover over row summary to see exact row count</p>																		
<p>10</p>	<p><b>Profile Card</b></p> <p>Identify the data type and see the options available to apply to your data when you select a field in the Profile pane.</p> <table border="0"> <tr> <td> Calculated Field</td> <td> Rename Field</td> </tr> <tr> <td> Change Data Type</td> <td> Search</td> </tr> <tr> <td> Edit Value</td> <td> Split Fields</td> </tr> <tr> <td> Exclude Values</td> <td> Boolean Data Type</td> </tr> <tr> <td> Filter Values</td> <td> Date Data Type</td> </tr> <tr> <td> Group Values</td> <td> Date Time Data Type</td> </tr> <tr> <td> Keep Only</td> <td> Numeric Data Type</td> </tr> <tr> <td> Merge Fields</td> <td> Text Data Type</td> </tr> <tr> <td> Remove Field</td> <td></td> </tr> </table>	 Calculated Field	 Rename Field	 Change Data Type	 Search	 Edit Value	 Split Fields	 Exclude Values	 Boolean Data Type	 Filter Values	 Date Data Type	 Group Values	 Date Time Data Type	 Keep Only	 Numeric Data Type	 Merge Fields	 Text Data Type	 Remove Field	
 Calculated Field	 Rename Field																		
 Change Data Type	 Search																		
 Edit Value	 Split Fields																		
 Exclude Values	 Boolean Data Type																		
 Filter Values	 Date Data Type																		
 Group Values	 Date Time Data Type																		
 Keep Only	 Numeric Data Type																		
 Merge Fields	 Text Data Type																		
 Remove Field																			



# Connect to Data

To use Tableau Prep Builder to clean and prepare your data, start a new flow by connecting to your data, just like in Tableau Desktop. You can also open an existing flow and pick up where you left off.


You can see and access your most recent flows right on the Start page, so it's easy to find your work in progress. After you connect to your data, use the different options in the Input step to identify the data that you want to work with in your flow.

## Start or open a flow

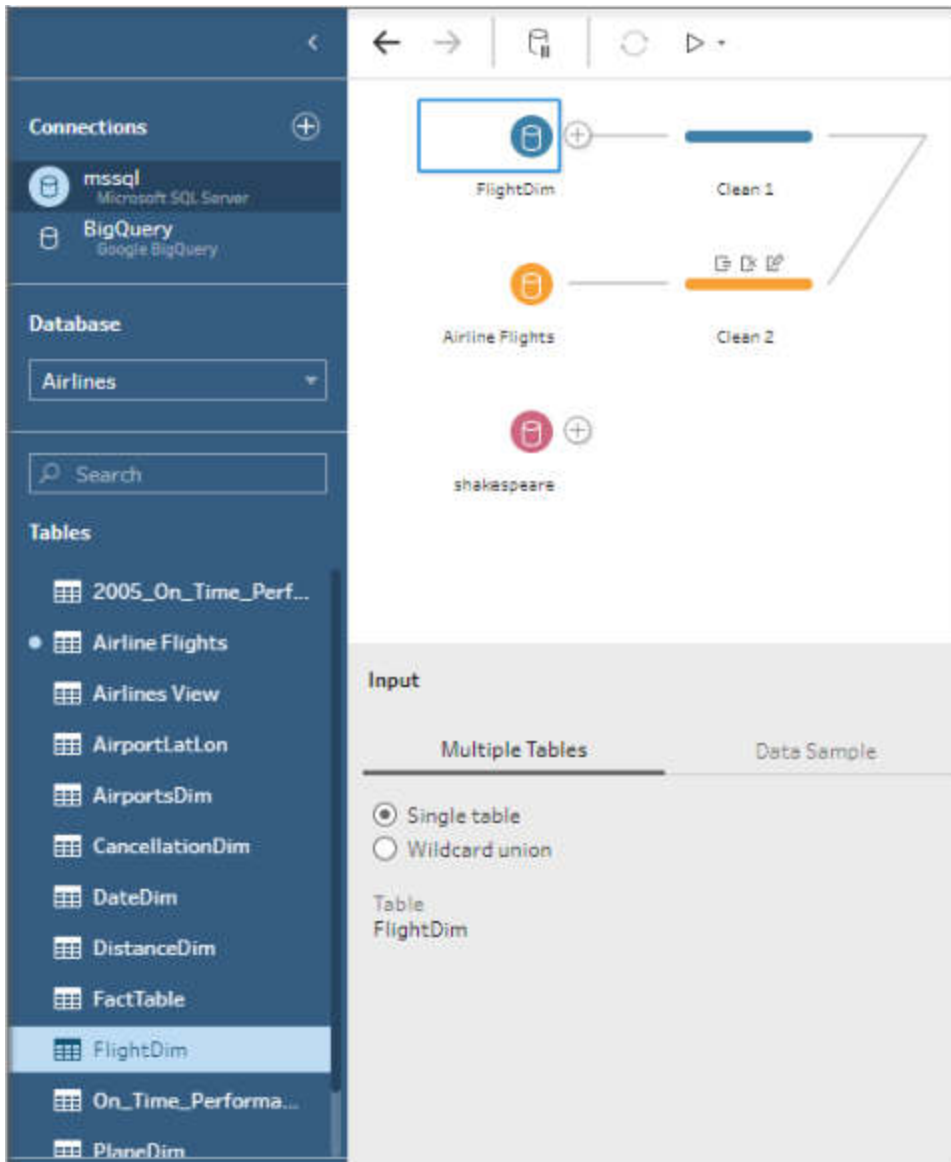
Tableau Prep Builder supports connections to popular types of data as well as Tableau data extracts (.tde or .hyper) and starting in version 2019.3.1, published data sources. Connectors that are supported in Tableau Desktop might not yet be supported in your version of Tableau Prep Builder. But new connectors are added with each new version, so check the **Connections** pane to see if your connector is available.

**Note:** If you open a flow in a version where the connector isn't supported, the flow may open but might have errors or won't run unless the data connections are removed.

You can also use custom SQL queries to connect to data just like you can in Tableau Desktop today. For more information, see [Connect to a Custom SQL Query](#) in the Tableau Desktop and Web Authoring Help.


To check whether you can connect to your data, open Tableau Prep Builder and click the **Add connection**  button to see if a connector for your data is listed in the left pane under **Connect**.

Starting in version 2020.1.1, you can also easily see which connectors and tables are associated with your Input steps. When you click on the Input step, the associated connector and data table is highlighted in the Connections pane.



**Note:** Some connectors might require you to download and install a driver before you can connect to your data. See the [Driver Download](#) page on the Tableau website to get driver download links and installation instructions.

## Start a new flow

1. Open Tableau Prep Builder and click the **Add connection**  button.
2. From the list of connectors, select the file type or server that hosts your data. If prompted,

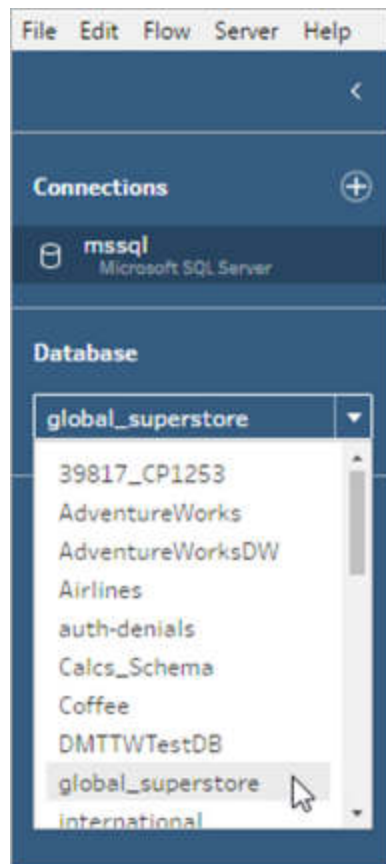
enter the information needed to sign in and access your data.

3. From the **Connections** pane, do one of the following:
  - If you connected to a file, double-click or drag a table to the **Flow** pane to start your flow.

For single tables, Tableau Prep Builder automatically creates an Input step for you in the **Flow** pane when you add data to your flow.

- If you connected to a database, select a database or schema, and then double-click or drag a table to the **Flow** pane to start your flow.

**Note:** If you connect to multiple files or database tables from a single data source, you can use a wildcard search and union the files or tables in the input step. For more information, see [Union files and database tables in the Input step](#) on page 173.



## Open an existing flow

To open an existing flow, on the **Start** page do one of the following:

- Under **Recent Flows**, select a flow.
- Click **Open a Flow** to navigate to your flow file and open it.

Click **File > New** or **File > Open** to open a new workspace. You can open multiple Tableau Prep Builder workspaces using these options if you want to work on multiple flows at the same time.

In Tableau Prep Builder version 2019.3.1 and earlier, if you select **File > Open**, Tableau Prep Builder replaces your current open flow with the new flow you select.

## Connect to published data sources (version 2019.3.1 and later)

If you publish data sources to Tableau Server or Tableau Online from Tableau Desktop or as output from your flow, you can connect to these published data sources and use them as an input data source for your flow.


**Note:** When you publish a flow that includes a published data source as an input, the publisher is assigned as the default flow owner. When the flow runs, it uses the flow owner for the **Run As** account. For more information about the Run As account, see [Run As Service Account](#). Only the Site or Server Administrator can change the flow owner in Tableau Server or Tableau Online.

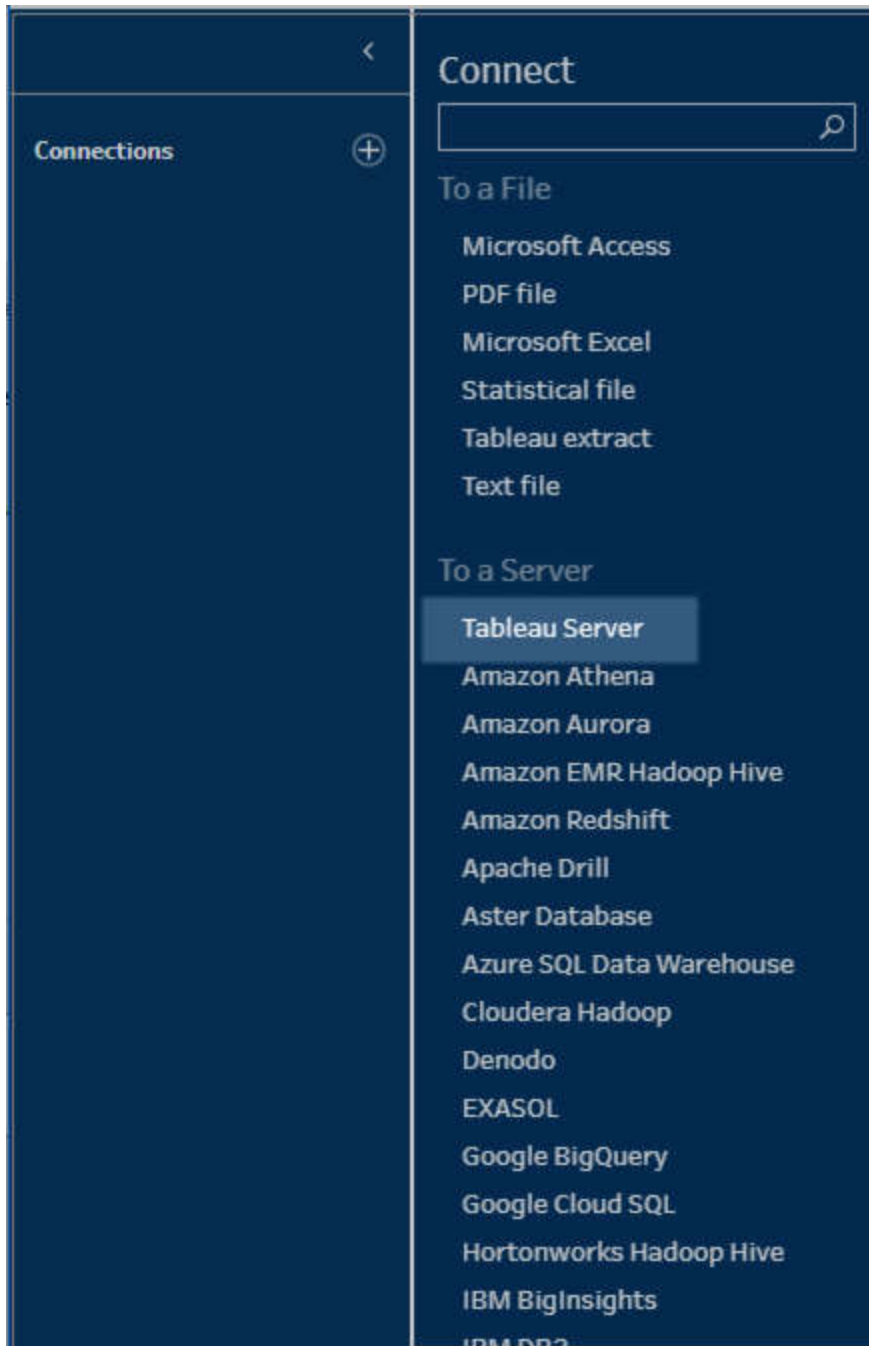
When using this feature the following requirements apply:

- Connecting to published data sources that are available on Tableau Server version 2019.3 are fully supported. All features may not be available if connecting to published data sources on earlier version of Tableau Server .
- You must be assigned a role of Explorer or higher in the Tableau Server site you are signed into to connect to published data sources. For more information about site roles, see [Set User's Site Roles](#) in the Tableau Server help.

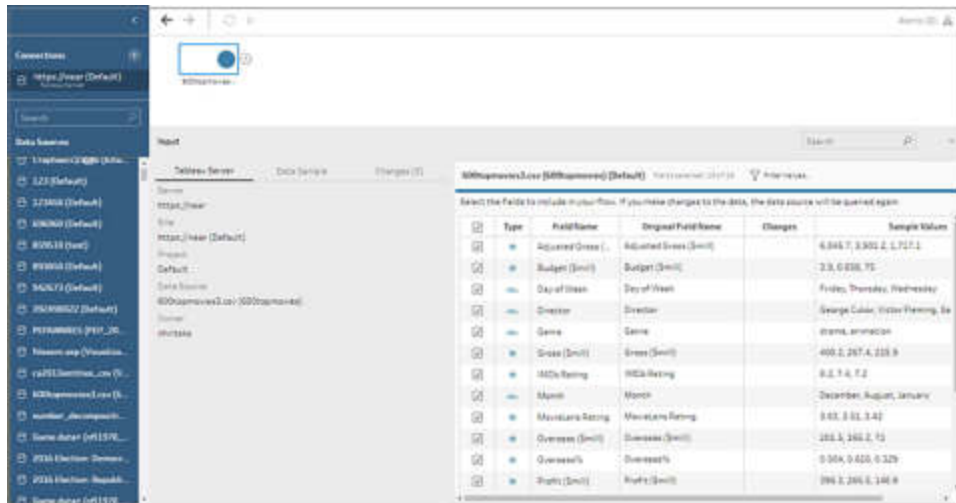
- Credentials must be embedded to connect to the published data source. If credentials aren't embedded for the data source, update the data source to include the embedded credentials.
- Published data sources that include multi-dimensional (cube) data or includes user filters are not currently supported.
- Multi-server connections aren't supported. You must use the same server or site connection to do the following:
  - Connect to the published data source.
  - Publish flow output to Tableau Server or Tableau Online.
  - Schedule the flow to run on Tableau Server or Tableau Online.

To connect to a published data source, do the following:

1. Open Tableau Prep Builder and click the **Add connection**  button.
2. From the list of connectors, select **Tableau Server**.



3. Sign in to connect to your server or site.
4. Select your data source or use the search option to find your data source and drag it to the flow pane to start your flow. The **Tableau Server** tab in the Input pane shows details about the published data source.



## Connect to cloud data sources

You can connect to cloud data sources in Tableau Prep Builder just like Tableau Desktop, but if you plan to publish flows that connect to cloud data sources and run them on a schedule in your server, you'll need to configure your credentials in Tableau Server or Tableau Online.

You set up your credentials in the **Settings** tab in the **My Account Settings** page and connect to your cloud connector input using these same credentials. When publishing the flow, on the **Publish** dialog, click **Edit** to edit the connection, then in the **Authentication** drop-down, select **Embed <your credentials>**.

Starting in version 2020.1.1, you can add credentials right from the publish dialog when publishing your flow and then automatically embed them in your flow when you publish. For more information, see [Publish a flow on page 402](#).

If you don't have saved credentials set up and select **Prompt user** in the **Authentication** drop-down, after you publish the flow you must edit the connection and enter your credentials in the **Connections** tab in Tableau Server or Tableau Online or the flow will fail when run.

In Tableau Prep Builder version 2019.4.1, the following cloud connectors were added:

- Box
- Dropbox
- Google Drive
- OneDrive



For more information about how to connect to your data using these connectors, see the [connector-specific](#) help topic in the Tableau Desktop help.

## Connect to Salesforce data (version 2020.2.1 and later)

Tableau Prep Builder supports connecting to data using the Salesforce connector, just like Tableau Desktop, but with a few differences.

- Tableau Prep Builder supports any join type you want to do.
- Custom SOQL isn't currently supported.
- Using a standard connection to create your own custom connection isn't currently supported.
- You can't change the default data source name to be something unique or custom.
- If you plan to publish flows to Tableau Server and want to use saved credentials, the server administrator will need to configure Tableau Server with an OAuth client ID and secret on the connector. For more information, see [Change Salesforce.com OAuth to Saved Credentials](#) in the Tableau Server help.

Tableau Prep Builder imports the data by creating an extract. Only extracts are currently supported for Salesforce. The initial extract may take some time depending on the amount of data that is included. You will see a timer in the Input step while the data loads.

For general information about using the Salesforce connector, see [Salesforce](#) in the Tableau Desktop and Web Authoring help.

## Connect to Google BigQuery (version 2019.3.1 and later)

Tableau Prep Builder supports connecting to data using Google BigQuery just like Tableau Desktop. You must configure OAuth credentials to enable Tableau Prep Builder to communicate with Google BigQuery. If you plan to publish flows to Tableau Server or Tableau Online, OAuth connections must also be configured for those applications.

**Note:** Tableau Prep Builder doesn't currently support using Google BigQuery customization attributes.

- [Set up OAuth for Google](#) - Configuring OAuth connections for Tableau Server.
- [OAuth Connections](#) - Configure OAuth connections for Tableau Online.

If you are using Tableau Prep Builder on Mac and you are using a proxy to connect to Big Query, you may need to modify the SSL configuration to connect to Google BigQuery

## Configure SSL to connect to Google BigQuery (MacOS only)

**Note:** No extra steps are required for Windows users.

To configure SSL for OAuth connections to Google BigQuery, complete the following steps:

1. Export the SSL certificate for your proxy to a file, for example proxy.cer. You can find your certificate in Applications > Utilities > Keychain Access > System > Certificates (under Category).
2. Locate the version of java that you are using to run Tableau Prep Builder. For example:  
/Applications/Tableau Prep Builder  
2020.1.app/Plugins/jre/lib/security/cacerts
3. Open the Terminal command prompt and run the following command for your Tableau Prep Builder version:

**Note:** The keytool command must be run from the directory that contains the version of java that you are using to run Tableau Prep Builder. You may have to change directories before running this command. For example `cd /Users/tableau_user/Desktop/SSL.cer -keystore Tableau Prep Builder 2020.1.1/Plugins/jre/bin`. Then run the keytool command.

```
keytool -import -trustcacerts -file /Users/tableau_
user/Desktop/SSL.cer -keystore Tableau Prep Builder
<version>/Plugins/jre/lib/security/cacerts -storepass
changeit
```

**Example:** `keytool -import -trustcacerts -file /Users/tableau_
user/Desktop/SSL.cer -keystore Tableau Prep Builder
2020.1.1/Plugins/jre/lib/security/cacerts -storepass
changeit`


If you get a **FileNotFoundException** (Access denied) when running the keytool command, try running the command with elevated permissions. For example: `sudo keytool -
import -trustcacerts -file /Users/tableau_user/Desktop/SSL.cer -
keystore Tableau Prep Builder
2020.1.1/Plugins/jre/lib/security/cacerts -storepass changeit`.

## Set up and manage your Google BigQuery credentials

The credentials that you use to connect to Google BigQuery in your Input step must match the credentials that are set up in the **Settings** tab in the **My Account Settings** page for Google BigQuery in Tableau Server or Tableau Online.

If you select different credentials or no credentials in your authentication setting when publishing your flow, the flow will fail with an authentication error until you edit the connection for the flow in Tableau Server or Tableau Online.

To edit your credentials do the following:

1. In Tableau Server or Tableau Online, on the **Connections** tab, on the Google BigQuery connection, click **More actions**  .
2. Select **Edit Connection**.
3. Select the saved credentials that are set up in the **Settings** tab in the **My Account Settings** page.

For more information about setting and managing your credentials, see the following topics:

[Manage Your Account Settings](#) in the Tableau Desktop and Web Authoring help.

[Publish a flow on page 402](#) for information about setting authentication options when publishing a flow.

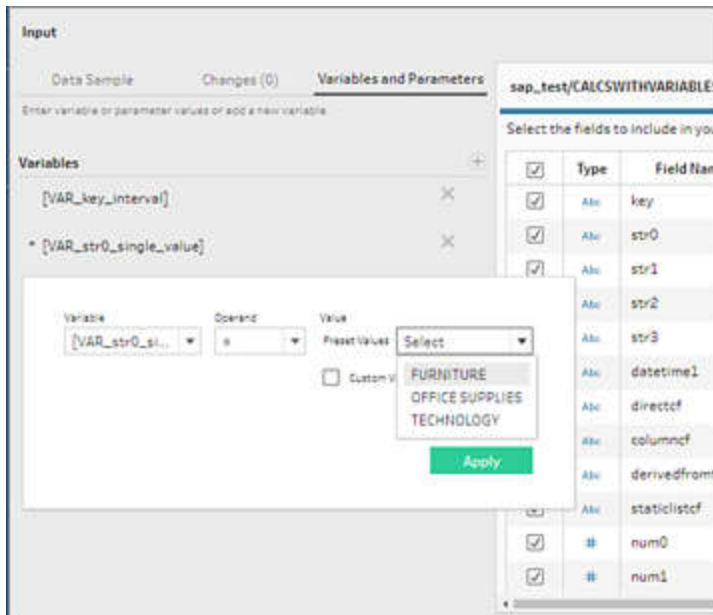
[View and resolve errors on page 430](#) for information about resolving connection errors in Tableau Server or Tableau Online.

## Connect to SAP HANA (version 2019.2.1 and later)

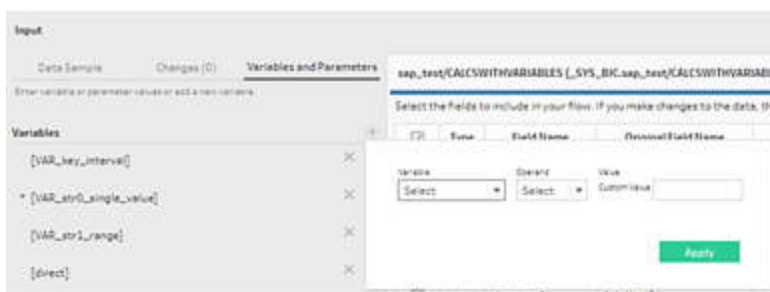
Tableau Prep Builder supports connecting to data using SAP HANA just like Tableau Desktop but with a few differences. Connect to the database using the same procedure you would use in Tableau Desktop. For more information see [SAP HANA](#). After you connect and search for your table, drag the table to the canvas to begin building your flow.

Prompting for variables and parameters when opening a flow isn't supported in Tableau Prep Builder. Instead, in the Input pane, click the **Variables and Parameters** tab and select the variables and operands you want to use, then select from a list of preset values or enter custom values to query your database and return the values you need.

**Note:** Starting in Tableau Prep Builder version 2019.2.2 you can use Initial SQL to query your connection and if you have multiple values for a variable, you can select the value you need from a drop-down list.



You can also add additional variables. Click the plus button **+** in the **Variables** section and select a variable and operand, then enter a custom value.



**Note:** This connector isn't supported in Tableau Prep Conductor version 2019.1. If you create a flow that uses this connector, you can publish the flow but you won't be able to run the flow using the scheduling features in Tableau Server. Instead you can run the flow using the command line interface to keep your data fresh. For more information

about running flows from the command line see [Refresh flow output files from the command line on page 334](#). For more information about version compatibility, see [Version Compatibility with Tableau Prep on page 347](#).

## Connect using ODBC (version 2019.2.2 and later)

If you need to connect to data sources that aren't listed in the **Connections** pane, you can connect to any data source using the **Other Databases (ODBC)** connector that supports the SQL standard and implements the ODBC API. Connecting to data using the **Other Databases (ODBC)** connector works similarly to how you might use it in Tableau Desktop, however there are a few differences:

- You can only connect using the DSN (data source name ) option.
- To publish and run your flow in Tableau Server, the server must be configured using a matching DSN.

**Note:** Running flows from the command line (Tableau Prep Builder) that include the Other Databases (ODBC) connector isn't currently supported.

- There is a single connection experience for both Windows and MacOS. Prompting for connection attributes for ODBC drivers (Windows) isn't supported.
- Only 64-bit drivers are supported by Tableau Prep Builder.

### Before you connect

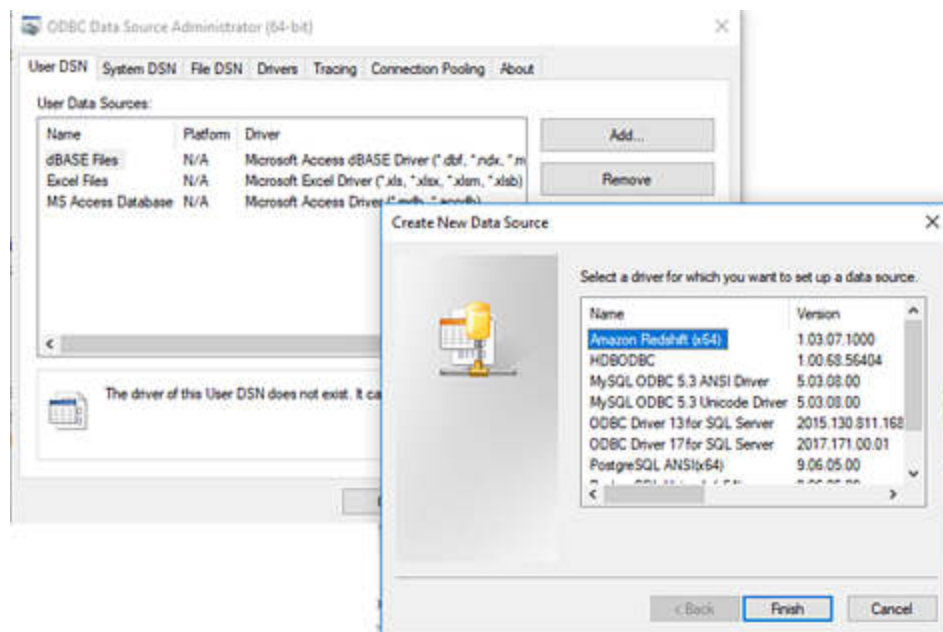
To connect to your data using the **Other Databases (ODBC)** connector, you must install the database driver and set up and configure your DSN (data source name). To publish and run flows to Tableau Server, the server must also be configured with a matching DSN.

**Important:** Tableau Prep Builder only supports 64-bit drivers. If you have a 32-bit driver already set up and configured, you may need to uninstall it and then install the 64-bit version if the driver doesn't allow both versions to be installed at the same time.

1. Create a DSN using either the **ODBC Data Source Administrator (64-bit)** (Windows) or the using an ODBC Manager utility (MacOS).

If you don't have the utility installed on your Mac, you can download one from ([www.odbcmanager.net](http://www.odbcmanager.net) for example) or you can manually edit the `odbc.ini` file.

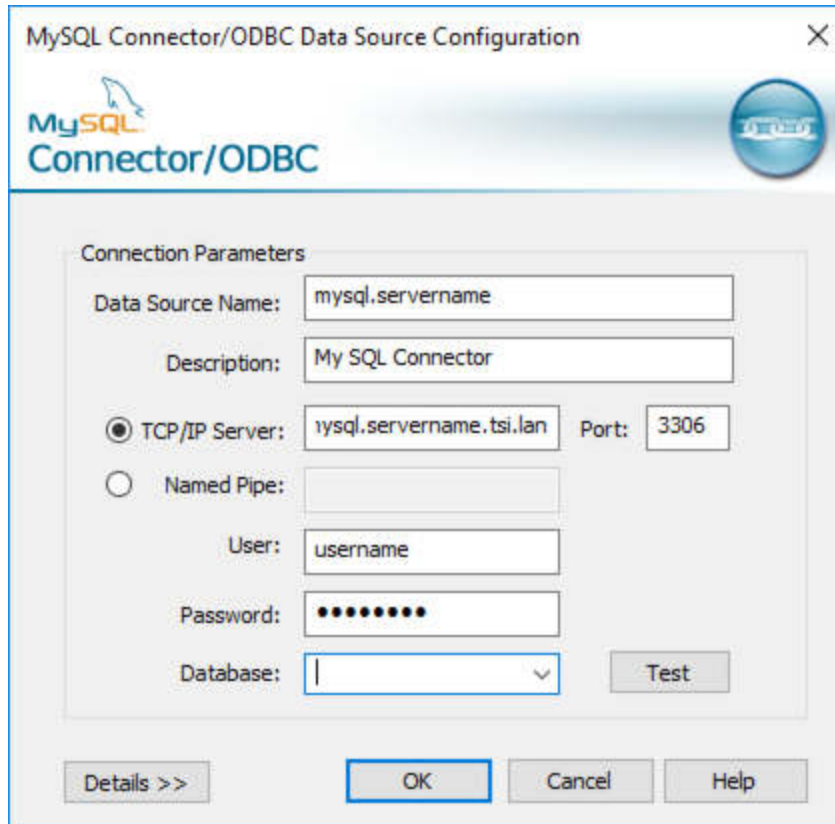
2. In the **ODBC Data Source Administrator (64-bit)** (Windows) or the ODBC Manager utility (MacOS), add a new data source then select the driver for the data source then click **Finish**.




3. In the **ODBC Driver Setup** dialog, enter the configuration information such as server name, port, user name and password. Click **Test** (if your dialog has that option) to verify that your connection is set up correctly, then save your configuration.

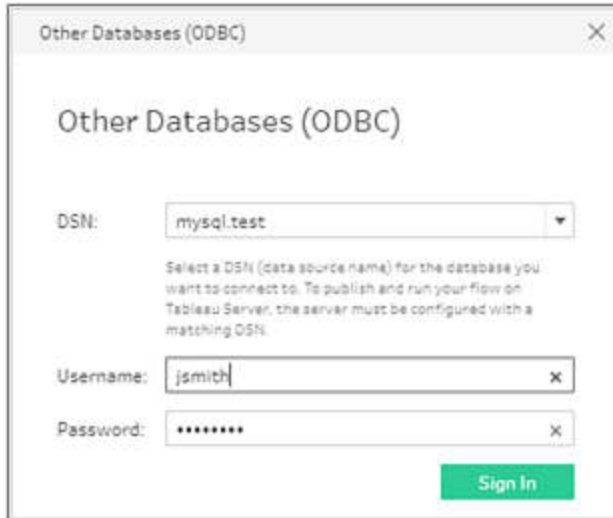
**Note:** Tableau Prep Builder doesn't support prompting for connection attributes so you must set this information when configuring the DNS.

This example shows the configuration dialog for a MySQL Connector.

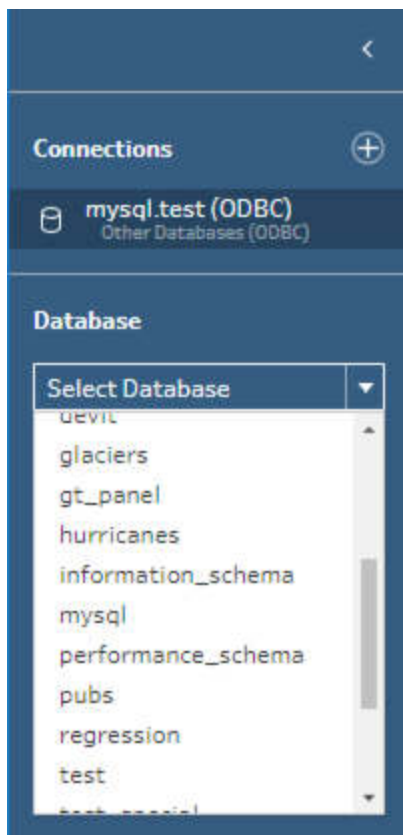


## Connect using Other Databases (ODBC)

1. Open Tableau Prep Builder and click the **Add connection**  button.
2. From the list of connectors, select **Other Databases (ODBC)**.
3. In the **Other Databases (ODBC)** dialog, select a DSN from the drop-down list and enter the user name and password. Then click **Sign In**.



4. From the **Connections** pane, select your database from the drop-down list.





## Use Initial SQL to query your connections (version 2019.2.2 and later)

You can specify an Initial SQL command that will run when a connection is made to the database that support it. For example when connecting to Amazon Redshift, you can enter a SQL statement to apply a filter when connecting to the database just like adding filters in the Input step. The SQL command will apply before data is sampled and loaded into Tableau Prep Builder.

Starting in version 2020.1.3, you can also include parameters to pass application name, version and flow name data to include tracking data when you query your data source.

### Run Initial SQL

To refresh your data and run the Initial SQL command do one of the following:

- Change the Initial SQL command and refresh the Input step by re-establishing the connection.
- Run the flow. The Initial SQL command is run before processing all of the data.
- Schedule the flow to run on Tableau Server or Tableau Online. The Initial SQL is run every time that the flow is run as part of the data loading experience.

**Note:** The [Data Management Add-on](#) is required to run your flow on Tableau Server or Tableau Online. For more information about the Data Management Add-on, see [Use the Data Management Add-on](#).

1. In the Connections pane, select a connector in the list that supports Initial SQL.
2. Click the **Show Initial SQL** link to expand the dialog and enter your SQL statements.

## Include parameters in your Initial SQL statement (version 2020.1.3 and later)

**Important:** Parameters are not supported for Initial SQL in Tableau Prep Builder version 2020.2.1 and earlier. If you try and enter the parameters like you would in Tableau Desktop, you will receive errors.

You can pass the following parameters to your data source to add additional detail about your Tableau Prep application, version and flow name. The **TableauServerUser** and **TableauServerUserFull** parameters are not currently supported.

Parameter	Description	Returned value
TableauApp	The application being used to access your data source.	Prep Builder Prep Conductor
TableauVersion	The application version number.	Tableau Prep Builder: Returns the exact version. For example 2020.1.3 Tableau Prep Conductor: Returns the major

		server version where Tableau Prep Conductor is enabled. For example 2020.1
FlowName	The name of the .tfl file in Tableau Prep Builder	Example: Entertainment Data_Cleaned

## Connect to Tableau data extracts

When you connect to a Tableau data extract, Tableau Prep Builder unpackages the extract and hyper expands, using a lot of temp space as it applies your flow operations to the resulting raw data.

This means you may need more RAM and disk space to accommodate a file that size. For example, an extract file with 18 columns and 1.2 million rows that is 360MB (8.5 GB uncompressed) may need up to 32GB RAM, 16-core, and 500GB of disk space available to support the file when it is unzipped.

## Use Data Interpreter to clean your files

When working with Microsoft Excel files, you can use Data Interpreter to detect sub-tables in your data as well as remove extraneous information to help prepare your data for analysis. When you turn on Data Interpreter, it detects these sub-tables and lists them as new tables in the **Tables** section of the **Connections** pane.

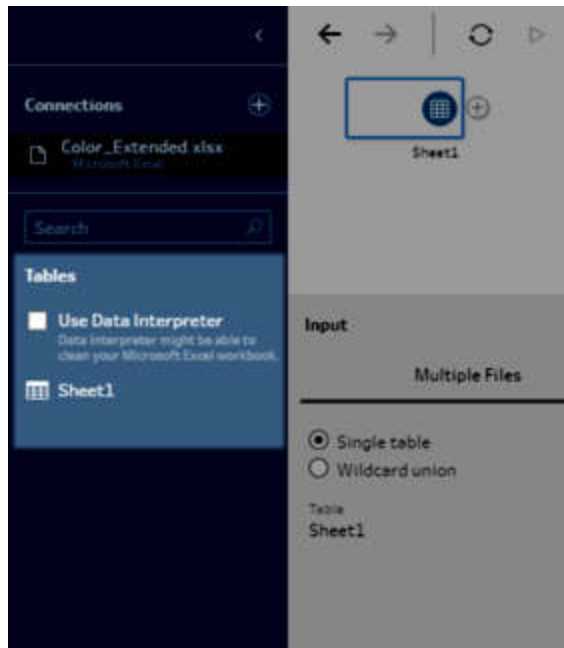
You can then drag them into the **Flow** pane. If you are using Tableau Prep Builder version 2018.1.2 or later, you can select the **Wildcard union** option in the **Multiple Files** tab to include all found sub-tables in your flow. For more information about using Wildcard union in the Input step see [Union files and database tables in the Input step on page 173](#).

If you turn Data Interpreter off, these tables are removed from the **Connections** pane. If these tables are already used in the flow, this will result in flow errors from the missing data.

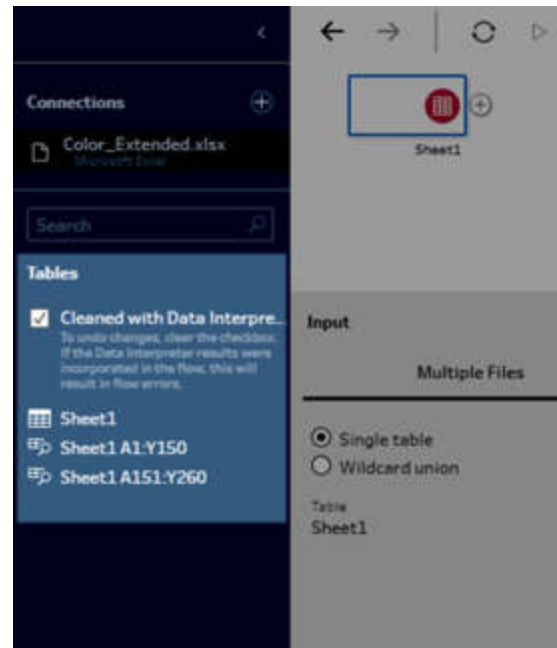
**Note:** Currently, Data Interpreter only detects sub-tables in your Excel spreadsheets and doesn't support specifying the starting row for text files and spreadsheets.

The example below shows the results of using Data Interpreter on an Excel spreadsheet in the **Connections** pane. Data Interpreter detected two additional sub-tables.

### Before Data Interpreter



### After Data Interpreter



To use Data Interpreter, complete the following steps:

1. Select **Connect to Data** then select **Microsoft Excel**.
2. Select your file and click **Open**.
3. Select the **Use Data Interpreter** check box.
4. Drag the new table to the **Flow** pane to include it in your flow. To remove the old table, right-click the Input step for the old table and select **Remove**.

## Configure your Data Set

To determine how much of your data set to include in the flow, you can configure your data set. When you connect to your data or drag tables into the **Flow** pane, an Input step is automatically added to the flow. This is always the first step in your flow. You can right-click the Input step to rename or remove it. If you're connected to an Excel or text file, you can also refresh the data from the Input step. For more information about how to refresh data from the input step, see [Add More Data in the Input Step](#) on page 170.

In the Input step, you can see the details about your data set. Here you can search for fields, see sample values, and perform actions to reduce the size of your data set, such as selecting

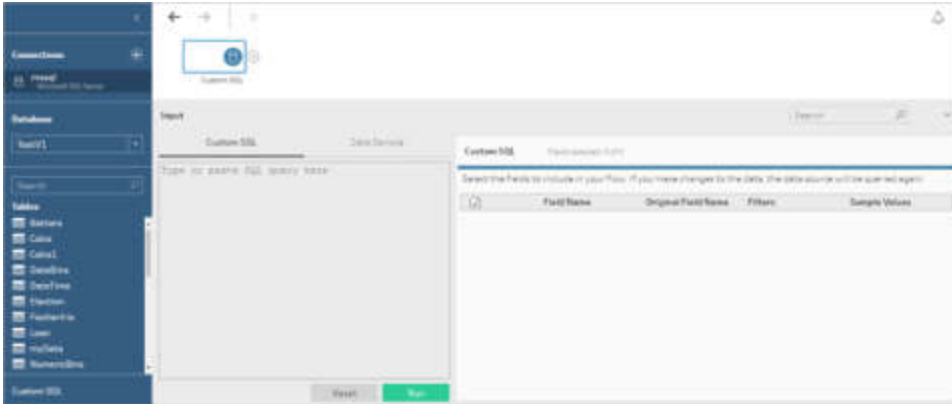
the fields to include, selecting the data sample to work with, or applying filters to selected fields or rows. You can also configure the field properties by changing the data type or field name for fields or configure the text settings for text files.

**Note:** Field values that include square brackets are automatically converted to parentheses.

<input type="checkbox"/>	Type	Field Name	Original Field Name	Changes	Sample Values
<input checked="" type="checkbox"/>		Row ID	Row ID		3,836, 3,837
<input checked="" type="checkbox"/>	Abc	Order ID	Order ID		CA-2015-100363
<input checked="" type="checkbox"/>		Order Date	Order Date		04/08/2015, 12:00:00 AM
<input checked="" type="checkbox"/>		Ship Date	Ship Date		04/15/2015, 12:00:00 AM
<input checked="" type="checkbox"/>	Data Type	Ship Mode	Ship Mode		Standard Class
<input checked="" type="checkbox"/>	Number (decimal)	Customer ID	Customer ID		JM-15655
<input checked="" type="checkbox"/>	Number (whole)	Customer Name	Customer Name		Jim Mitchell
<input checked="" type="checkbox"/>	Date & Time	Segment	Segment		Corporate
<input checked="" type="checkbox"/>	Date	Country	Country		United States
<input checked="" type="checkbox"/>	String	City	City		Glendale
<input checked="" type="checkbox"/>	Abc	Postal Code	Postal Code		85,301
<input checked="" type="checkbox"/>	Abc	Region	Region		West

## Connect to a custom SQL query

If your database supports using custom SQL, you will see **Custom SQL** displayed near the bottom of the **Connections** pane. Double-click **Custom SQL** to open the **Custom SQL** tab where you can enter queries to preselect data and use source-specific operations. After the query retrieves the data set, you can select the fields to include, apply filters, or change the data type before adding the data to your flow.



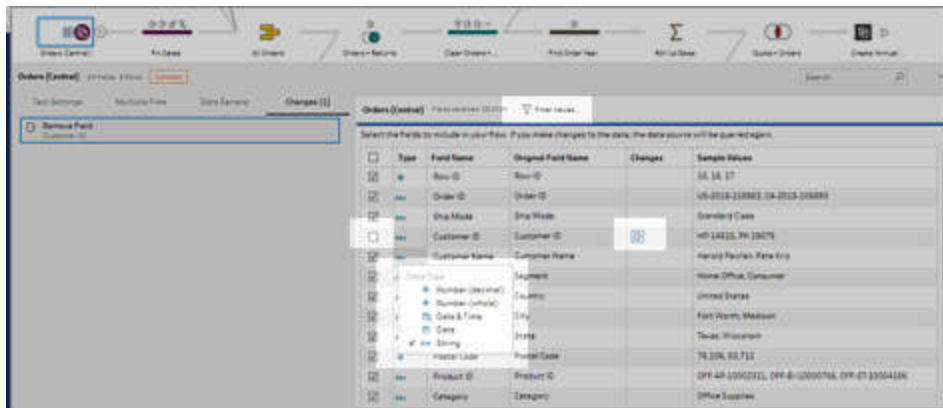
For more information about using custom SQL, see [Connect to a Custom SQL Query](#) in the Tableau Desktop and Web Authoring Help.

## Apply cleaning operations in an input step

Only some cleaning operations are available in an Input step. You can make any of the following changes in the Input field list. Your changes are tracked in the **Changes** pane and annotations are added to the left of the Input step in the **Flow** pane and in the Input field list.

- **Filter:** Click **Filter Values** in the toolbar then enter your filter criteria in the calculation editor.
- **Rename Field:** In the **Field Name** field, double-click or Ctrl-click (MacOS) on the field name and enter a new field name.
- **Change Data Type:** Click on the data type for the field and select a new data type from the menu.
- **Remove Field:** Clear the check box next to the fields that you don't want to include in

your flow.



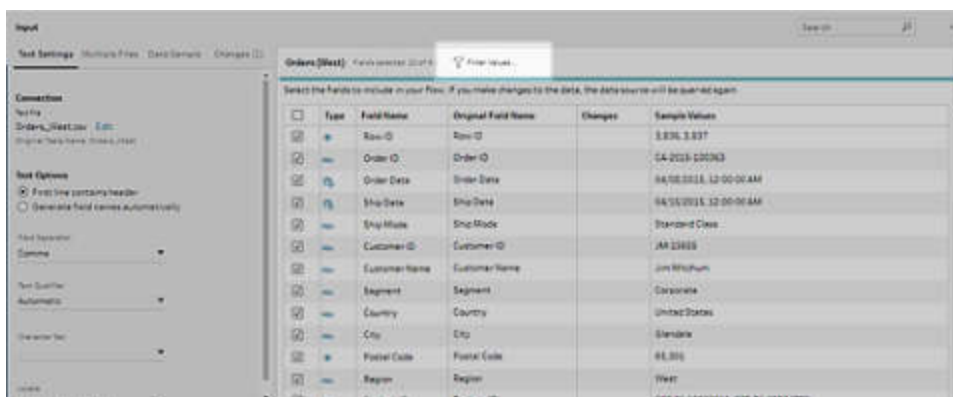
## Select the fields to include in the flow

The **Input** pane shows you a list of fields in your data set. You can use the **Search** field to find fields in the list, and then use the check boxes to select the fields to include or exclude. To include or exclude all fields from the flow, toggle the check box at the top left of the grid.

## Apply filters to fields in the Input step

To filter a field, do the following:

1. In the toolbar click **Filter Values**.



2. Enter your filter criteria in the calculation editor.

The calculation filter type is the only filter option available in the Input step. Other filter options are available in other step types. For more information, see [Filter Your Data on page 207](#)

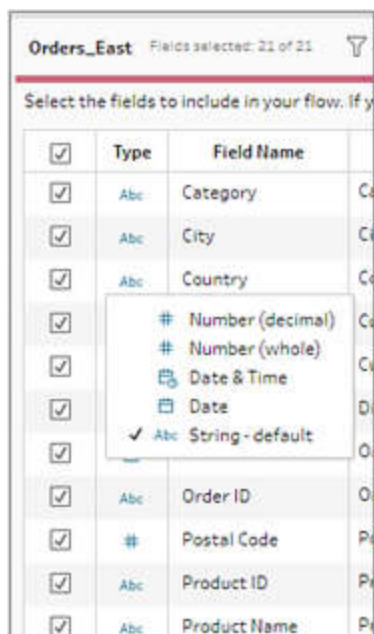
## Change field names

To change the name of a field, in the **Field Name** column, select the name, and then type the new name in the field. An annotation is added in the field grid and in the flow pane to the left of the Input step. Your changes are also tracked in the **Changes** pane.

## Change data types

To change the data type for a field, do the following:

1. Click the data type for the field.
2. Select the new data type from the menu.



You can also change the data type for fields in other step types in the flow or assign data roles to help validate your field values. For more information about changing your data type or using data roles, see [Review the data types assigned to your data on page 198](#) and [Use Data Roles to Validate your Data on page 213](#).

## Configure field properties

When you work with text files, you see a **Text Settings** tab where you can edit your connection and configure text properties, such as the field separator for text files. You can also edit the file connection in the Connections pane.



When you work with text or Excel files, you can correct data types that have been inferred incorrectly before you even start your flow. Data types can always be changed in subsequent steps in the **Profile** pane after you start your flow.

### Configure text settings in text files


To change the settings used to parse text files, select from the following options:

- **First line contains header** (default): Select this option to use the first row as the field labels.
- **Generate field names automatically**: Select this option if you want Tableau Prep Builder to auto-generate the field headers. The field naming convention follows the same model as Tableau Desktop. For example **F1**, **F2**, and so on.
- **Field Separator**: Select a character from the list to use to separate the columns. Select **Other** to enter a custom character.
- **Text Qualifier**: Select the character that encloses the values in the file.
- **Character Set**: Select the character set that describes the text file encoding.
- **Locale**: Select the locale to use to parse the file. This setting indicates which decimal and thousand separator to use.

### Set your data sample size

To maintain peak performance, Tableau Prep Builder limits the data included in the flow to a representative sample of your data set by default. The data sample is determined by calculating the optimal number of rows based on the total number of fields in the data set and the data types for those fields. Tableau Prep Builder then retrieves the top number of rows for the calculated amount as quickly as possible.

The resulting data sample may include all the rows you need, or it may not, depending on how the sample was calculated and returned. If you don't see the data that you expect, you can change the data sample settings to run the query again.

**Note:** If your data is sampled, a **Sampled**  indicator shows in the **Profile** pane and persists for every step you add. Any changes you make apply to the sample you are working with in the flow. All changes apply to your entire data set when you run the flow.

To change your data sample settings, select an Input step, then on the **Data Sample** tab select from the following options:

- **Default sample amount** (default): Tableau Prep Builder calculates the total number of rows to return.
- **Use all data:** Retrieve all rows in your data set regardless of size. This can impact performance or cause Tableau Prep Builder to time out.

**Note:** To maintain performance, even if you select this setting, a data sample limit of 1 million rows is applied to Aggregate and Union step types and a data sample limit of 3 million rows is applied to Join and Pivot step types.

- **Fixed number of rows:** Select the number of rows to return from the data set. The recommended number of rows is 1 million or less. Setting the number of rows to more than 1 million can impact performance.
- **Quick select** (default): The database returns the number of rows requested as quickly as possible. This might be the first N number of rows or the rows that the database had cached in memory from a previous query.
- **Random sample:** The database returns the number of rows requested but looks at every row in the data set and returns a representative sample from all of the rows. This option may impact performance when the data is first retrieved.

## Add More Data in the Input Step

After you connect to your data sources and begin to build your flow you may want to refresh your data connections as new data comes in. You can also join or union data sets in the input step to make working with larger data sources more efficient.

## Refresh data in the Input step

If data changes in your input files or tables after you begin working with your flow, you can refresh the Input step to bring in the new data.

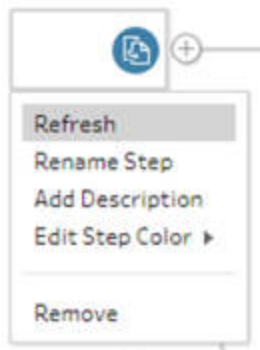
### File input step types

To refresh file inputs steps, do one of the following:

- In the flow pane on the top menu, click the **Refresh** button to refresh all Input steps. To refresh a single Input step, click the drop-down arrow next to the refresh button and select the Input step from the list.



- In the flow pane, right-click the Input step you want to refresh and select **Refresh** from the menu.



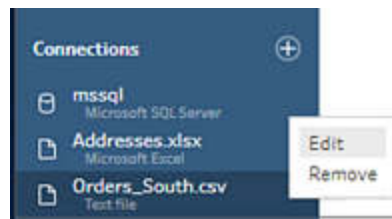
### File, database or Tableau extract input step types

To refresh database or tableau extract input steps, do one of the following:

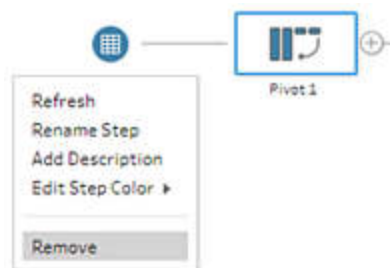
- Try editing the connection.

**Note:** To maintain performance, Tableau Prep Builder samples large data sets. If your data is sampled, you may or may not see your new data in the profile pane. You can change the settings for how your data is sampled in the Data Sample tab in the Input step, but it may impact performance. For more information about setting your data sample size, see [Set your data sample size on page 169](#).

1. In the **Connections** pane, right-click or Ctrl-click (MacOS) on the data source and select **Edit**.



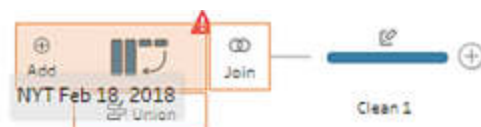
2. Re-establish your connection by signing into the database or re-selecting the file or Tableau extract.
- Remove and re-add the Input step to the flow.
    1. In the flow pane, right-click the Input step you want to refresh and select **Remove** from the menu.



This will temporarily put your flow in an error state.



2. Connect to the updated file again.
3. Drag the table to the flow pane on top of the second step in the flow where you want to add the Input step. Drop it on the **Add** option to reconnect it to the flow.



## Union files and database tables in the Input step

When working with multiple files or database tables from a single data source, you can search for files or tables using a wildcard search and then union the data to include all of the file or table data in the Input step. To union files, the files must be in the same parent or child directory.

New files that are added to the same folder that match the pattern are automatically included in the union the next time you open the flow or run it from the command line. Packaged flow files (.tflx) won't automatically pick up new files because the files are already packaged with the flow. To include new files for packaged flows, open the flow file (.tfl) to pick up the new files, then repackage the flow to include the new file data.

To union database tables, the tables must be in the same database and the database connection must support using a wildcard search to union. The following databases support this type of union:

- Amazon Redshift
- Microsoft SQL Server
- MySQL
- Oracle
- PostgreSQL

If you add or remove files or tables after you create the union you can refresh the Input step to update your flow with the new or changed data.

**Note:** Currently, this feature applies only to Excel and .csv (text) files and data tables stored in the specific databases listed above. This option is not available for Tableau data extracts.

Wildcard union for files is available in Tableau Prep Builder version 2018.1.2 and later. Wildcard union for database tables is available in Tableau Prep Builder version 2018.3.1 and later. Editing a flow connection with this type of union in a prior version can result in errors.

If you need to union data from different data sources, you can do that using a Union step. For more information about creating Union steps, see [Union your data on page 314](#).


## Union files

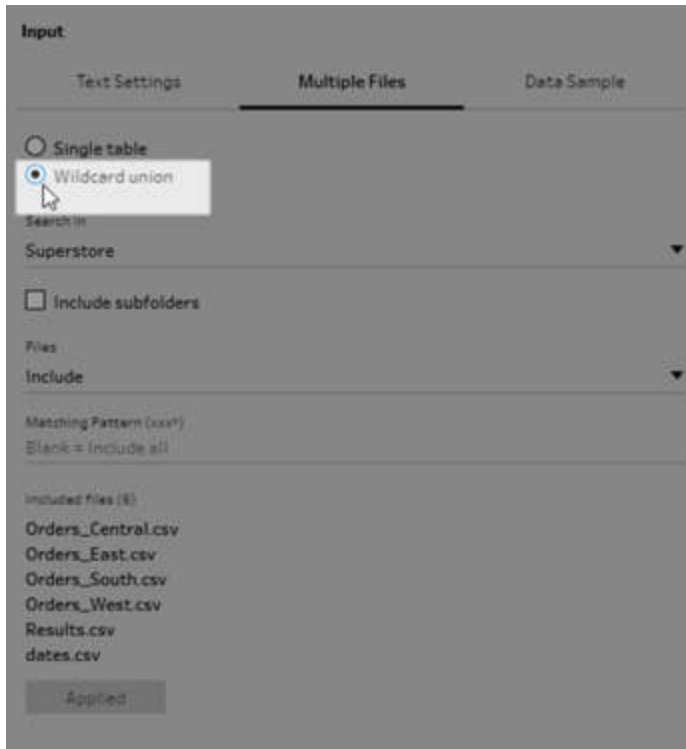
By default, Tableau Prep Builder unions all .csv files in the same directory as the .csv file you connected to or all the sheets in the Excel file you connected to. If you use Data Interpreter to clean Excel files and are using Tableau Prep Builder version 2018.1.2 or later, you can use the wildcard search to union and add any sub-tables that Data Interpreter found.

If you want to change the default union, use the following criteria to find the files or sheets you want to include in the union:

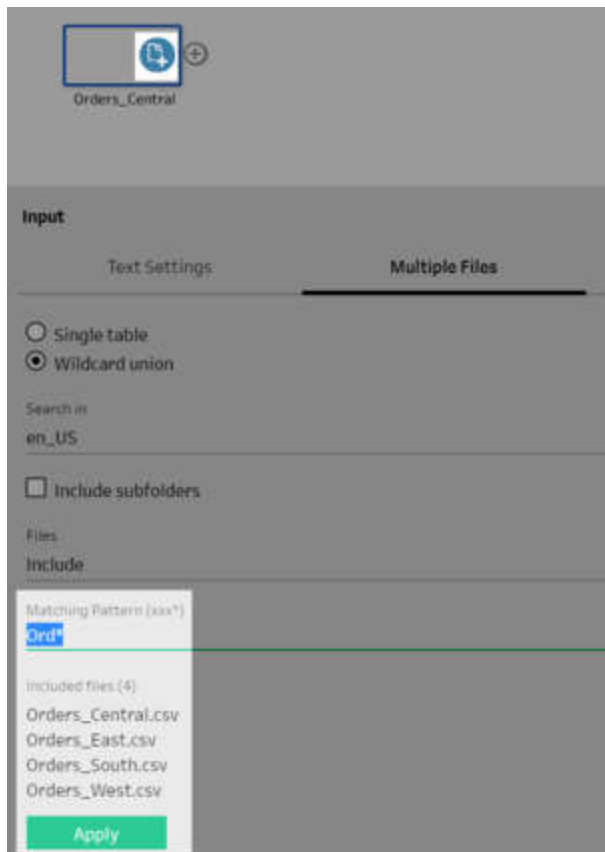
- **Search in:** Select the directory to use to search for files. Select the **Include subfolders** check box to include files in the sub-directory of the parent folder.
- **Files:** Select whether to include or exclude the files that match the wildcard search criteria.
- **Matching Pattern (xxx\*):** Enter a wildcard search pattern to find files that have those characters in the file name. For example, if you enter ord\* all files that include the file name are returned. Leave this field blank to include all of the files in the specified directory.

To union files in the input step, do the following:

1. Click the **Add connection**  button and under **Connect**, click **Text File** for .csv files or **Microsoft Excel** for Excel files, and then select a file to open.
2. In the **Input** pane, select the **Multiple Files** tab, and then select **Wildcard union**.



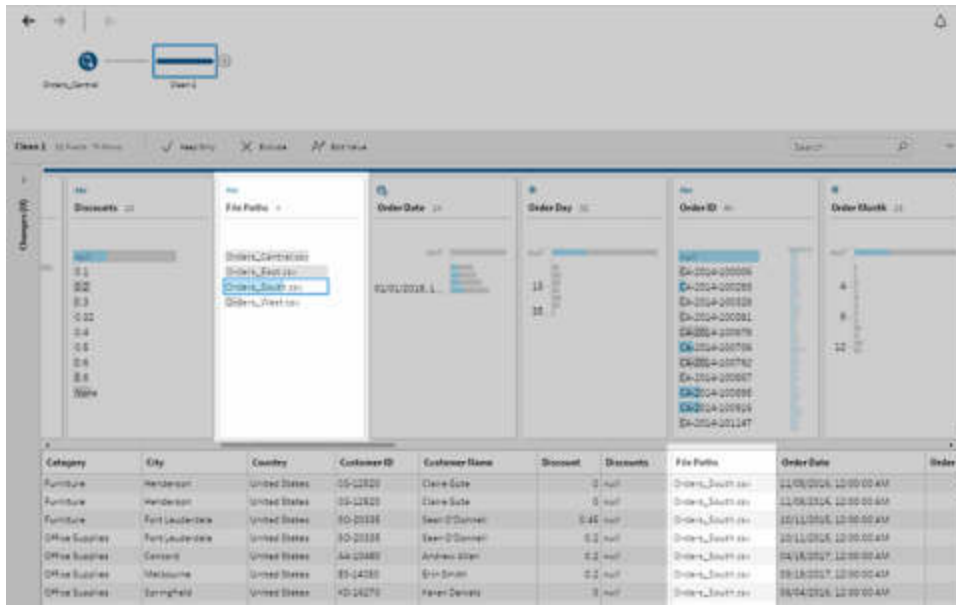
The example below shows a wildcard union using a matching pattern. The plus sign on the file icon on the Orders\_Central Input step in the **Flow** pane indicates that this step includes a wildcard union. The files in the union are listed under **Included files**.




3. Use the search, file and matching pattern options to find the files that you want to union.
4. Click **Apply** to union the files.

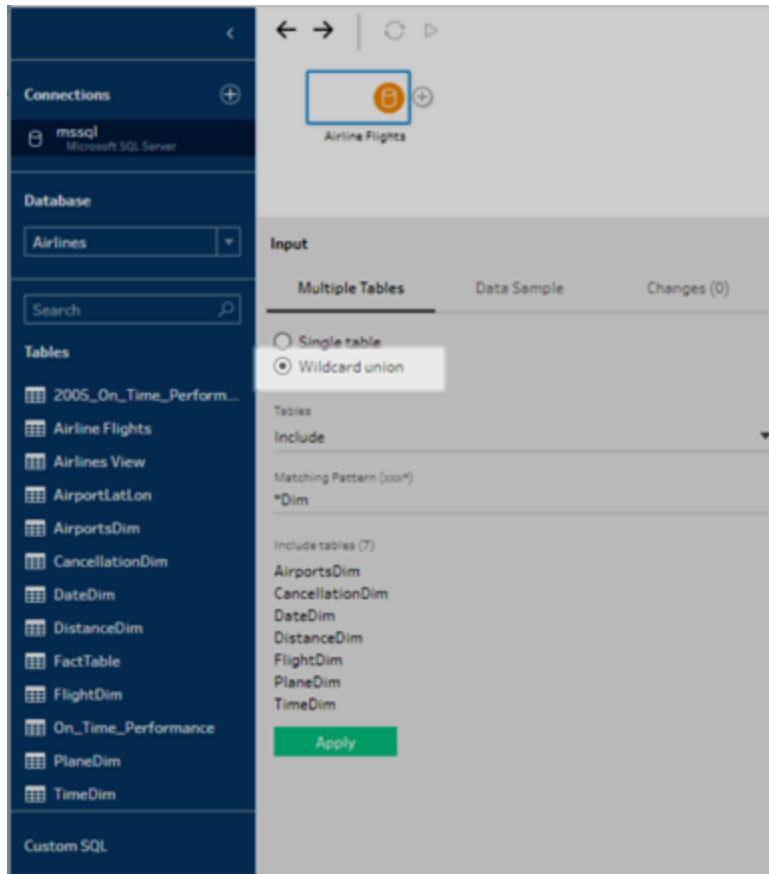
When you add a new step to the flow, you can see all the files added to the data set in the **File Paths** field in the Profile pane. This field is added automatically.





### Union database tables (version 2018.3.1 and later)

1. Click the **Add connection**  button and under **Connect**, connect to a database that supports wildcard union.
2. Drag a table to the flow pane.
3. In the **Input** pane, select the **Multiple Tables** tab, and then select **Wildcard union**.

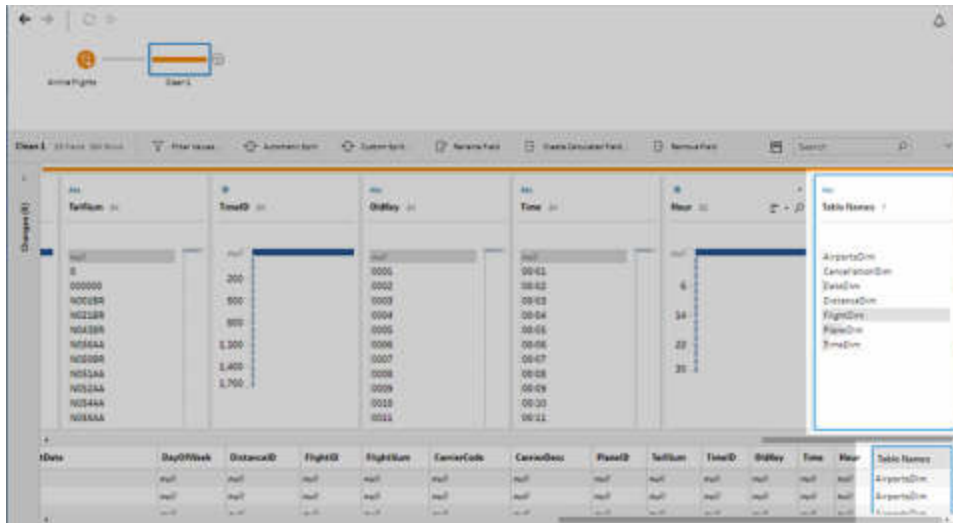


4. Use search, **Tables** and **Matching Pattern** options to find the tables that you want to union.

Only tables that display in the **Connections** pane in the **Tables** section can be included in the union. Wildcard search doesn't search across schemas or across the database connection to find tables.

5. Click **Apply** to union the table data.

When you add a new step to the flow, you can see all the tables added to the data set in the **Table Names** field in the Profile pane. This field is added automatically.





## Merge fields after a union


After you create a union in the input step, you might want to merge fields. You can do this in any subsequent step, except for the Input or Output steps. For more information, see [Additional merge field options](#) on page 320.

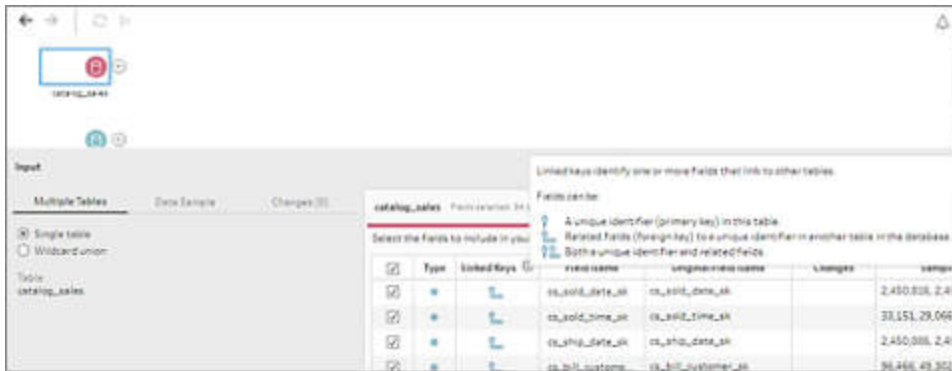
## Join data in the Input step (version 2019.1.3 and later)

When you connect to databases that include tables with relationship data, Tableau Prep Builder can detect and show which fields in a table are identified as the unique identifier and which fields are identified as a related field as well as show the related table names for these fields.



A new column called **Linked Keys** shows in the Input pane and shows the following relationships if they exist:

-  Unique identifier. This field uniquely identifies each row in the table. There can be multiple unique identifiers in a table. The values in the fields must be unique and cannot be blank or null.
-  Related field. This field relates the table to another table in the database. There can be multiple related fields in a table.

-  Both Unique Identifier and related field. The field is a unique identifier in this table and also relates the table to another table in the database.



You can leverage these relationships to quickly find and add the related tables to your flow or create joins from the Input step. This feature is available for any supported database connector where table relationships are defined.

1. Connect to a database (such as Microsoft SQL Server) that contains relationship data for fields, such as unique identifiers or related fields (foreign key).
2. In the Input pane, click on a field that is marked as a  related field or as  both a unique identifier and related field.

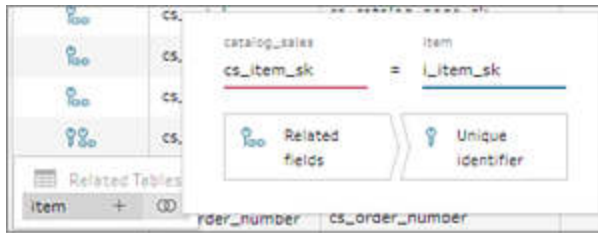
A dialog opens that shows a list of related tables.


3. Hover on the table that you want to add or join and click the plus button to add the table to your flow, or click the join button to create a join with the selected table.



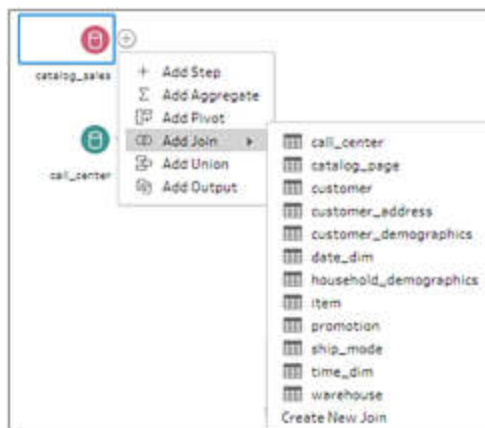
If you create a join, Tableau Prep Builder uses the defined field relationship to join the tables and shows you a preview of the join clauses that it will use to create the join.

## Tableau Prep Help



- Alternatively, you can join related tables from the menu in the Flow pane. Hover over a step until the plus  icon appears, then select **Add Join** to see a list of related tables. Tableau Prep Builder creates the join based on the fields that make up the relationship between the two tables.

**Note:** If your table doesn't have table relationships defined, this option is not available.



For more information about working with joins, see [Join your data](#) on page 308.

# Build and Organize your Flow

After you connect to the data that you want to include in your flow, you can begin cleaning and shaping your data by adding new steps to the flow or inserting steps in between existing steps.

To organize your flow, you can change the default step colors, add descriptions to provide context for your steps or cleaning actions, or reorganize your flow layout to make complex flows easier to follow.

## Add or insert steps

As you build out your flow, you can add different step types to perform the actions that you need. For example use the **Add Step** option to add a cleaning step to split fields, apply filters, or perform a variety of other operations to clean dirty data. Use a **Join** or **Union** step type to combine data tables or add a **Script** step type to incorporate your R or Python scripts into your flow.


As your flow begins to take shape, you may need to go back to earlier steps in your flow and insert different step types to perform various actions like adding an additional cleaning step or aggregating your data to use the same level of detail as a later step.

The menu options that you see will vary depending on your Tableau Prep Builder version and whether you are adding a step to build out the next step in the flow versus inserting a step in between existing steps. If you are using Tableau Prep Builder version 2019.3.1 or earlier, refer to that section to see your menu options.

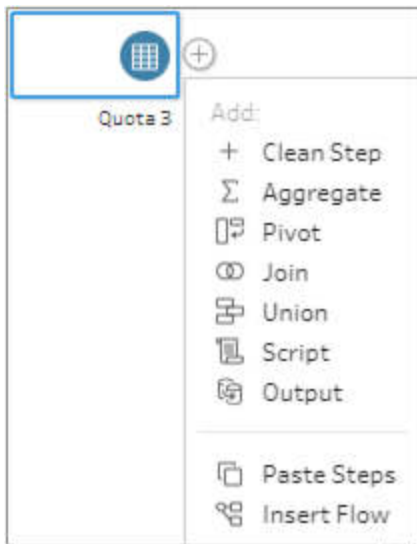
You can't add input steps using these menu's. Instead you'll need to drag tables from the **Connections** pane to the Flow pane. For more information, see [Connect to Data on page 145](#).

**Note:** Some features or actions that you include in your flow may not be compatible with the version of Tableau Server that you are signed into. Flows that include incompatible features can't be scheduled to run using Tableau Prep Conductor. For more information about managing version incompatibility, see [Version Compatibility with Tableau Prep on page 347](#).

## Add steps

Hover over a step until the plus  icon appears, then click the icon and select a step type. This menu appears when you are adding a step to build out the next step in the flow.

**Note:** Your options may look different depending on your product version. For example **Insert Flow** was added to this menu in Tableau Prep Builder version 2019.4.2.



Select from the following options:

- **Add Branch:** Split your flow into different branches. This option appears when opening the Add menu from between two existing steps.

**Note:** In Tableau Prep Builder 2019.4.2, this option was replaced with **Add Clean Step** when opening the Add menu from between two existing steps.

- **Add Step** or **Add Clean Step:** Add a cleaning step to perform a variety of cleaning actions. For more information about the various cleaning actions you can use, see [Clean and Shape Data](#) on page 229.
- **Add Aggregate:** Create an **Aggregation** step where you can select the fields that you want to aggregate or group. For more information, see [Aggregate and group values](#) on page 307.

- **Add Pivot:** Create a **Pivot** step where you can perform a variety of pivot options to convert column data to rows, or row data to columns. For more information, see [Pivot Your Data on page 281](#).
- **Add Join:** Create a **Join** step where you can manually add the other input to the join and add the join clauses. As an alternative, you can drag and drop a step to join files. The following example shows dragging the Orders\_Central Input step and dropping it on **Join**:



For more information about creating a join, see [Join your data on page 308](#).

In Tableau Prep Builder version 2019.1.3 and later, if you connect to databases that include tables with relationship data you can also create a join from the menu in the Flow pane. For more information about joining tables using this method, see [Join data in the Input step \(version 2019.1.3 and later\) on page 179](#).


- **Add Union:** Create a Union step. Add tables to the union by dragging them to the step and dropping them on the the **Add** option that displays. As an alternative, you can drag and drop a step onto another step to union files. For more information about creating a union, see [Union your data on page 314](#).
- **Add Script** (version 2019.3.1 and later): Create a Script step to include R and Python scripts in your flow. For more information, see [Use R and Python scripts in your flow on page 290](#).
- **Add Output:** Create an Output step to save the output to an extract file (.tde or .hyper), a .csv file, or publish the output as a data source to a server.
- **Paste Step** or **Paste Steps:** Add copied steps from the same flow. For more information about copying and pasting steps in the same flow, see [Clean and Shape Data on page 229](#).
- **Insert Flow** (version 2019.3.2 and later): Add flow steps that were saved from another flow into your current flow. You can add them to the end of a step or insert them between existing steps. This option was added to this menu in Tableau Prep Builder version



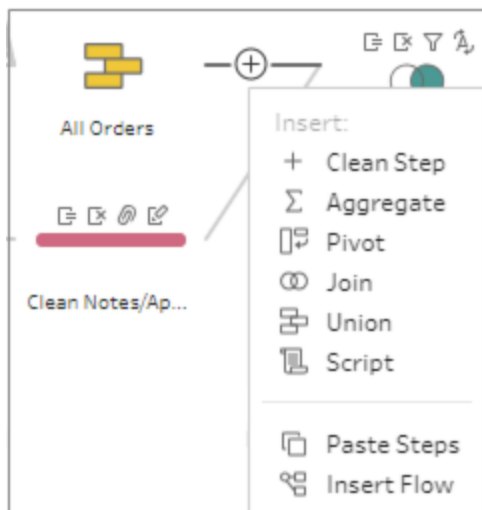
2019.4.2. In prior versions, you could add flow steps using right-click or Ctrl-click (MacOS) in the white space of the flow pane. For more information about using saved flow steps in your flow, see [Create reusable flow steps \(version 2019.3.2 and later\)](#) on page 264.

## Insert steps

Insert a step between existing steps. Input and Output step types aren't available from this menu. The options vary depending on your product version. If you are using an earlier version of Tableau Prep Builder, refer to the **Version 2019.3.1 and earlier** section below.

1. Hover in the middle of the flow line where you want to insert a step until the plus  icon appears. Then click the icon and select a step type.

**Note:** Your options may look different depending on your product version. For example **Insert Flow** was added to this menu in Tableau Prep Builder version 2019.4.2.



2. Select from the following options:
  - **Insert Step** or **Insert Clean Step**: Insert a cleaning step between existing steps to perform a variety of cleaning actions. For more information about the various cleaning actions you can use, see [Clean and Shape Data](#) on page 229.
  - **Insert Aggregate**: Insert an **Aggregation** step between existing steps. You can

select the fields that you want to aggregate or group. For more information, see [Aggregate and group values on page 307](#).

- **Insert Pivot:** Insert a **Pivot** step between existing steps. You can perform a variety of pivot options to convert column data to rows, or row data to columns. For more information, see [Pivot Your Data on page 281](#).
- **Insert Join:** Insert a **Join** step between existing steps. You can manually add the other input to the join and add the join clauses. As an alternative, you can drag and drop a step to join files. The following example shows dragging the Orders\_Central Input step and dropping it on **Join**:




For more information about creating a join, see [Join your data on page 308](#).

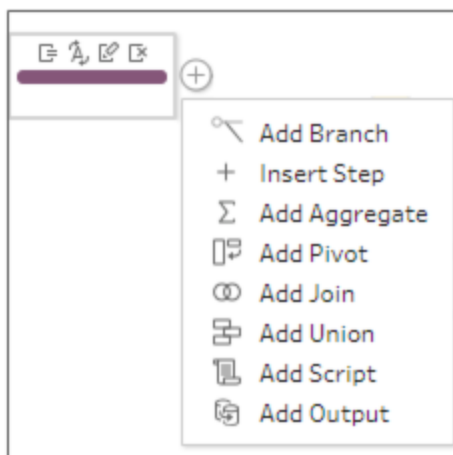
In Tableau Prep Builder version 2019.1.3 and later, if you connect to databases that include tables with relationship data you can also create a join from the menu in the Flow pane. For more information about joining tables using this method, see [Join data in the Input step \(version 2019.1.3 and later\) on page 179](#).

- **Insert Union:** Insert a Union step. Add tables to the union by dragging them to the step and dropping them on the **Add** option that displays. As an alternative, you can drag and drop a step onto another step to union files. For more information about creating a union, see [Union your data on page 314](#).
- **Paste Steps:** Insert copied steps from the same flow between existing steps. For more information about copying and pasting steps in the same flow, see [Clean and Shape Data on page 229](#).
- **Insert Flow** (version 2019.3.2 and later): insert flow steps that were saved from another flow into your current flow. You can add them to the end of a step or insert them between existing steps. This option was added to this menu in Tableau Prep Builder version 2019.4.2. In prior versions, you could add flow steps using right-click or Ctrl-click (MacOS) in the white space of the flow pane. For more

information about using saved flow steps in your flow, see [Create reusable flow steps \(version 2019.3.2 and later\)](#) on page 264.

## Version 2019.3.1 and earlier

1. Hover over a step until the plus  icon appears then click the icon and select a step type. **Insert Step** inserts a cleaning step between steps. All other options will create a branch from the flow.



2. Select from the following options:
  - **Add Branch:** Split your flow into different branches.
  - **Insert Step:** Insert a cleaning step between existing steps to perform a variety of cleaning actions. For more information about the various cleaning actions you can use, see [Clean and Shape Data](#) on page 229.
  - **Add Aggregate:** Create an **Aggregation** step where you can select the fields that you want to aggregate or group. For more information, see [Aggregate and group values](#) on page 307.
  - **Add Pivot:** Create a **Pivot** step where you can perform a variety of pivot options to convert column data to rows, or row data to columns. For more information, see [Pivot Your Data](#) on page 281.
  - **Add Join:** Create a **Join** step where you can manually add the other input to the

join and add the join clauses. As an alternative, you can drag and drop a step to join files. The following example shows dragging the Orders\_Central Input step and dropping it on **Join**:



For more information about creating a join, see [Join your data on page 308](#).

In Tableau Prep Builder version 2019.1.3 and later, if you connect to databases that include tables with relationship data you can also create a join from the menu in the Flow pane. For more information about joining tables using this method, see [Join data in the Input step \(version 2019.1.3 and later\) on page 179](#).

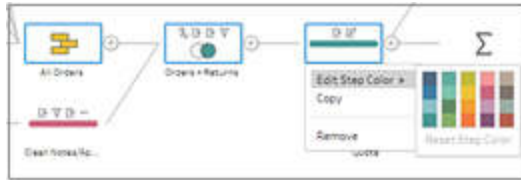
- **Add Union:** Create a Union step. Add tables to the union by dragging them to the step and dropping them on the **Add** option that displays. As an alternative, you can drag and drop a step onto another step to union files. For more information about creating a union, see [Union your data on page 314](#).
- **Add Script**(version 2019.3.1 and later): Create a Script step to include R and Python scripts in your flow. For more information, see [Use R and Python scripts in your flow on page 290](#).
- **Add Output:** Select this option to save the output to an extract file (.tde or .hyper), a .csv file, or publish the output as a data source to a server.

## Change the flow color scheme

Tableau Prep Builder assigns each step in your flow a color by default. This color scheme is applied throughout the flow to help you keep track of your data throughout the flow as you apply cleaning steps, join, union or aggregate the data so you know which files are impacted by your operations.

To select a different color scheme for your steps do the following:

1. Select one or more steps.
2. Right-click on a selected step and select **Edit Step Color**.



3. Click on a color in the color palette to apply it.

To reset the step color back to the default color, do one the following:


- Click **Undo** from the top menu.
- Cntrl+Z or Command-Shift-Z (Mac).
- Select the steps you changed, right-click on a selected step and select **Edit Step Color**, then select **Reset Step Color** from the bottom of the color palette.

## Add descriptions to flow steps and cleaning actions

As you build your flow and perform various cleaning operations, you might want to add a description to help others who might later look at or work with your flow to better understand your steps. You can add a description to any individual step in your flow directly on the Flow pane or to any cleaning action in the **Changes** pane to provide additional context for your changes. The description can be up to 200 characters long.

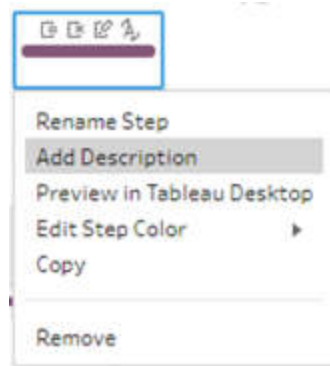
For more information about viewing changes in the Changes pane, see [View your changes on page 239](#).

### Add a description to flow steps

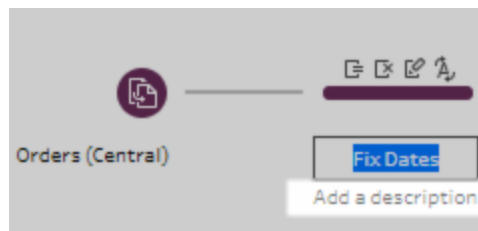
When you add a description, a message  icon is added underneath the step. Click the icon to show or hide the description text in the Flow pane.

1. In the Flow pane, select a step.
2. Do one of the following:

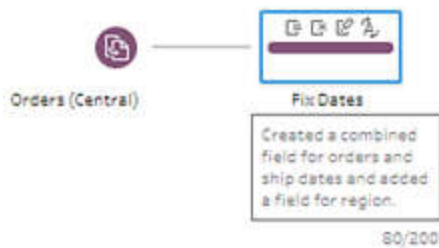
- Right-click the step and select **Add Description** from the menu.



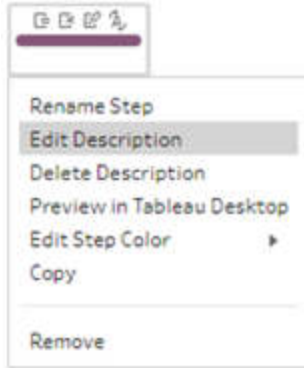
- Double-click in the name field for the step, then click on **Add a description**.



3. Type your description in the text box.

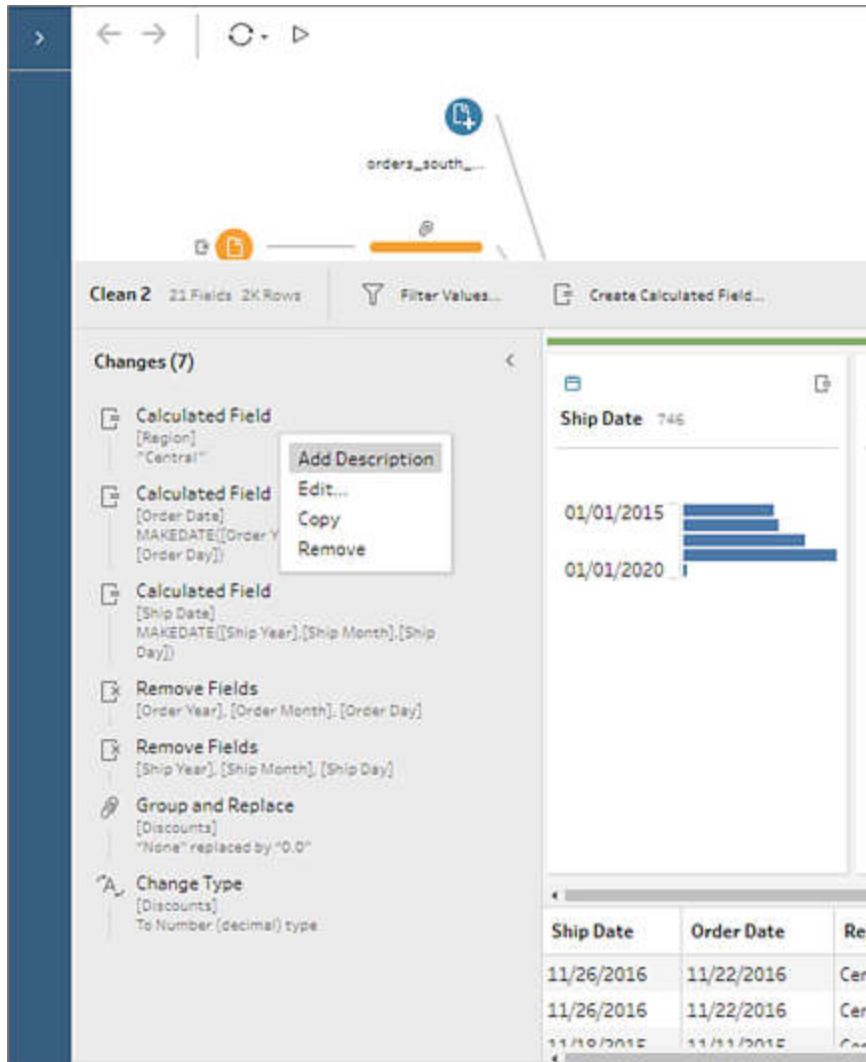


4. Click outside the text box or press Enter to apply your changes. By default, the description displays underneath the step. To hide the description click the message icon.
5. To edit or delete the description, right-click on the step or description then from the menu, select **Edit Description** or **Delete Description**.



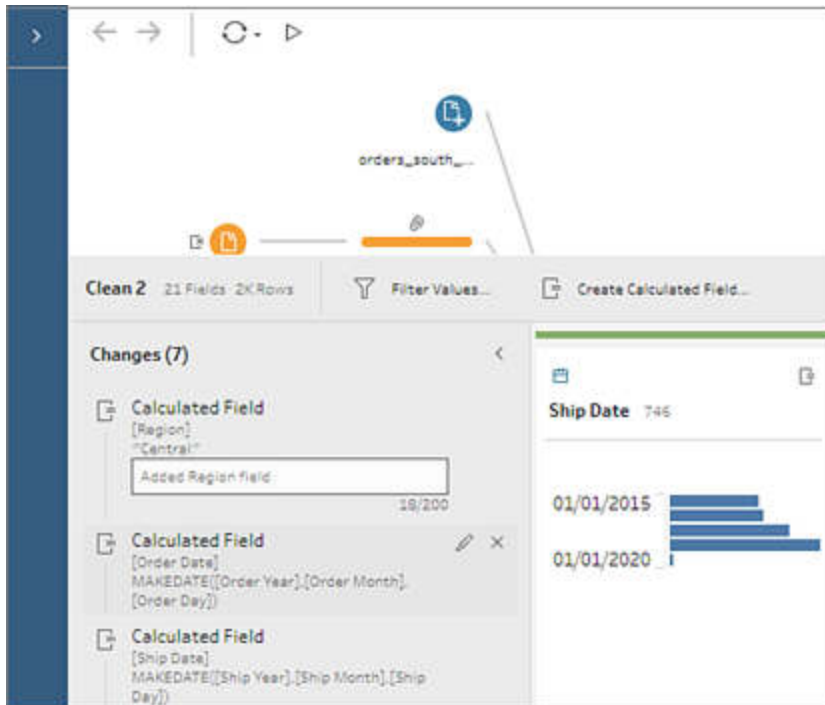
## Add a description to a change entry (version 2019.1.1 and later)

1. Select a step in the flow pane.
2. Open the **Changes** pane or **Changes** tab.
3. Right-click or Ctrl-click (MacOS) on an entry in the **Changes** pane and select **Add Description**.

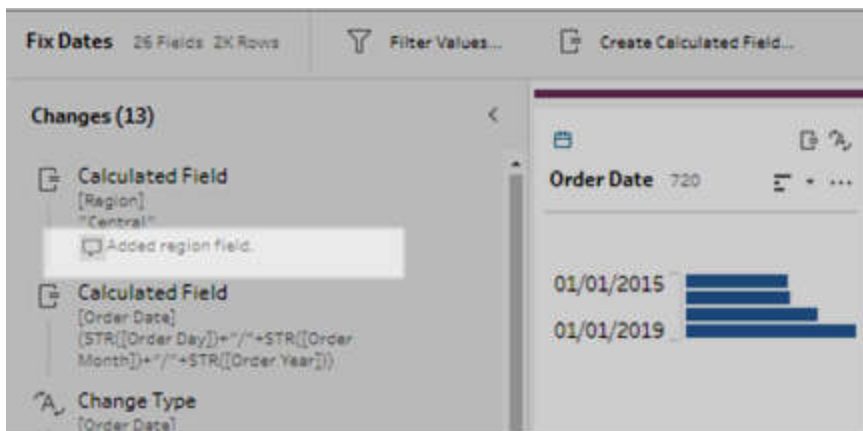


4. Enter a description for the change action.





The description appears below the generated text for the change with a comment icon.



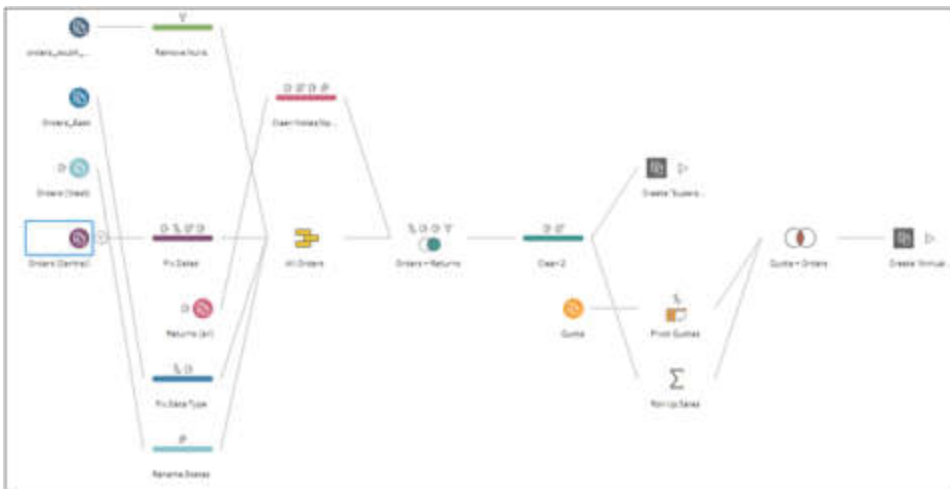
- To edit or delete the description, right-click or Ctrl-click (MacOS) on the change item, and select **Edit Description** or **Delete Description**.

## Reorganize the layout of your flow (version 2019.2.2 and later)

As you build a flow, Tableau Prep Builder uses a default layout. Each flow is laid out and processed from left to right, with Input steps beginning on the far left of the canvas and Output steps ending on the right side of the canvas. However if you build large, complex flows, they can quickly become hard to follow.

To clean up the layout of your flow, you can select and move steps so the flow layout is organized in a way that makes sense to you. For example, you can fix crossed flow lines, move your flow steps to clean up extra white space, or rearrange your flow steps to show a clear sequence of events.

For example, the following flow is confusing and hard to follow:



To clean up this flow you can select and drag steps up, down, left or right and drop them to a new location in the canvas. Flow steps can't be moved to a position that disrupts the left-to-right process flow. For example, you can't drag a union step that is positioned before a join step, to a position that is after that join step in the flow.

When dragging flow steps to an allowed location, an orange box displays. If the location isn't allowed, no orange box displays and the steps return to their original location when you try to drop them.

To move steps in your flow, do the following:

1. In the Flow pane, select the steps that you want to move. You can click on a specific step, drag to select multiple steps, or Ctrl-click or Cmd+Click (MacOS) to select steps that aren't next to each other.
2. Drag and drop the steps to the new location.

**Note:** If you don't like the reorganization moves that you make you can click **Undo** in the top menu to reverse them. However, if you perform cleaning actions in between moving steps, you may undo those actions as well. The **Undo** option reverses your actions in the order that you performed them.

### Watch "Reorganize flow steps" in action

The following example shows rearranging a flow using drag and drop.



## Use the flow navigator tool






When working with large flows, scrolling back and forth to search for a particular area of the flow that you need to go back to can be difficult. The flow navigator tool makes this easier. The flow navigator is a miniature version of your flow that appears in the lower right corner of the canvas.



Click in any area of the graphic to jump to that area of your flow or use the following toolbar



options to navigate:

Toolbar option	Description
	Collapse the flow navigator graphic. In the collapsed state, you may only see the percentage indicator. Simply hover on this to expand the toolbar and click the up arrow  to expand the graphic again.
	Expand the flow navigator graphic.
	Change the size of your flow to fit on your screen.
	Zoom in and out of your flow. You can click on the percentage indicator to restore the view to 100 percent.

## Remove steps from the flow

At any point in the flow, you can remove steps or the flow lines between steps.

## Tableau Prep Help

- To remove a step or flow line, select the step or line you want to remove, right-click the element, and then select **Remove**.
- To remove multiple steps or flow lines, do one of the following:
  - Use your mouse to drag and select a whole section of the flow. Then right-click on one of the selected steps and select **Remove**
  - Press Ctrl+A or Cmd+A (MacOS) to select all elements in the flow, or press Ctrl+click or Cmd+Click (MacOS) to select specific elements, and then press the **Delete** key.

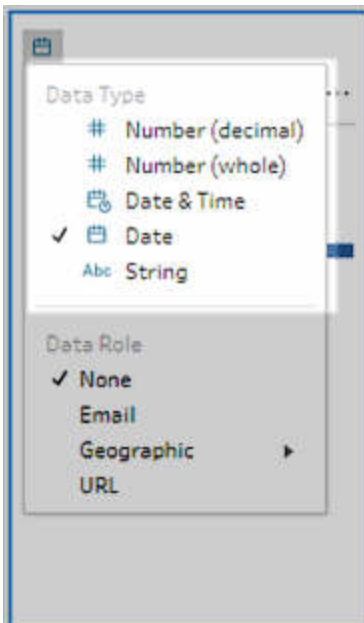
# Examine Your Data

Use the options in this topic to get a good understanding of the composition of your data to better understand changes you need to make and the effect of the operations you include in the flow.

## Review the data types assigned to your data

Like Tableau Desktop, Tableau Prep Builder interprets the data in your fields when you drag a connection to the **Flow** pane and automatically assigns a data type to it. Because different databases can handle data in different ways, Tableau Prep Builder's interpretation might not always be correct.

To change a data type, click the data type icon and select the correct data type from the context menu. You can change string or integer data types to Date or Date & Time, and Tableau Prep Builder will trigger Auto DateParse to change these data types. Like Tableau Desktop, if the change is not successful you will see Null values in the fields instead and you can create a calculation to make the change.



For more information about using DateParse, see [Convert a Field to a Date Field](#) in the Tableau Desktop and Web Authoring Help.

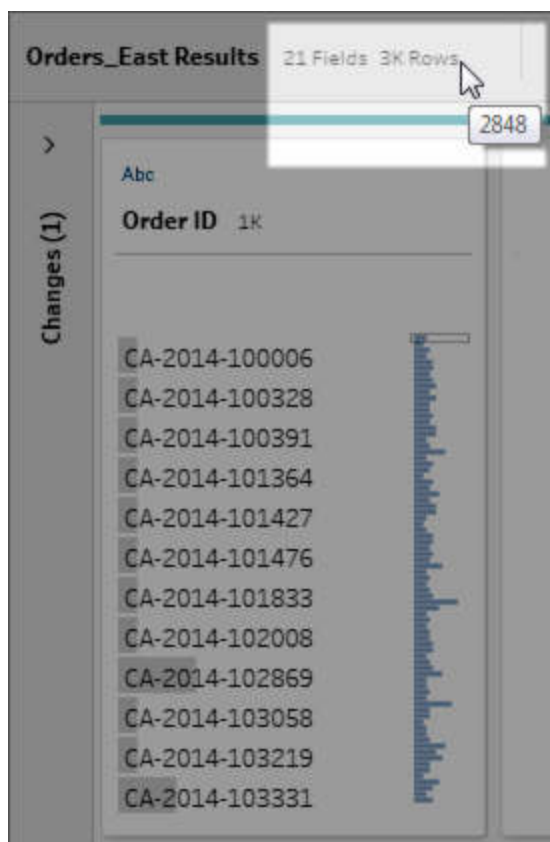
## See size details about your data

After you connect to your data, add a table to the flow, and then add a step. You can use the **Profile** pane to see the current state and structure of your data and spot nulls and outliers.

- **Number of fields and rows:** In the upper-left corner of the **Profile** pane you can find information that summarizes the number of fields and rows in the data at a particular point in the flow. Tableau Prep Builder rounds to the nearest thousand. In the example below, there are 21 fields and 3000 rows in the data set.



When you hover over the number of fields and rows, you can see the exact number of rows (in this example, 2848):



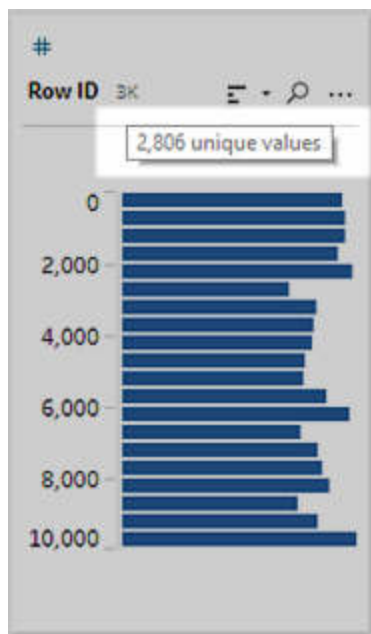
- **Data set size:** Work with a subset of your data by specifying the number of rows to

include in the **Data Sample** tab in the **Input** pane.

- **Sampled:** To enable you to interact directly with your data, Tableau Prep Builder works with a subset of your raw data. The number of rows is determined by the data types and number of fields that are being rendered. String fields take more storage space than integers, so if you have 10 fields of strings in your data set, you might get fewer rows

than if you had 10 fields of integers. A **Sampled** Sampled indicator displays next to the size details in the **Profile** pane to indicate that this is a subset of your data set. For more information about data samples, see [Configure your Data Set on page 164](#).

- **Number of unique values:** The number next to each field header represents the distinct values that are contained within that field. Tableau Prep Builder rounds to the nearest thousand. In the example below, there are 3,000 distinct values that are represented in the Description field, but if you hover over the number, you can see the exact number of unique values.



## See the distribution of values or unique values

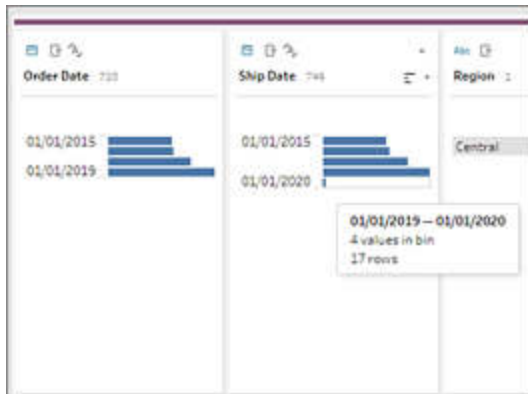
By default, Tableau Prep Builder groups numerical, date, and datetime values in a field into buckets. These buckets are also known as bins. The bins ensure that you can see the



## Tableau Prep Help

distribution of values as a whole and quickly identify outliers and null values. The bin size is calculated based on the minimum and maximum values in the field, and null values are always shown at the top of the distribution.

For example, order and ship dates are summarized or "binned" by year. Each bin represents a year from January of the beginning year to January of the following year and labeled accordingly. Because there are sales dates and ship dates that fall in the latter part of 2018 and 2019, a bin is created for the following year for those values.



If a discrete (or categorical) data field contains many rows or has a distribution that is large enough that it can't be displayed in the field without scrolling, you can see a summarized distribution to the right of the field. You can click and scroll through the distribution to target specific values.

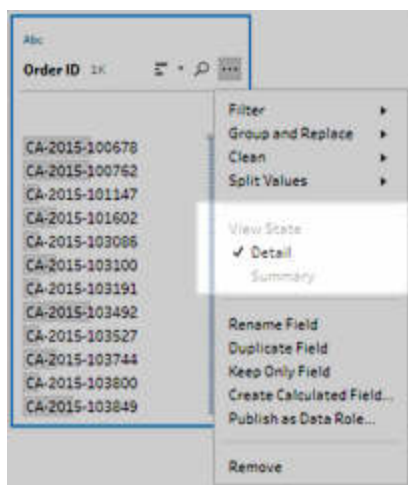


When your data contains numeric or date fields, you can toggle to display the detailed (discrete) version of the values or a summarized (continuous) version of the values. The summarized view shows you the range of values in a field and the frequency with which certain values appear.

This toggle can help you isolate unique values (like the number of “3” records in a field) or the distribution of values (like the sum of all “3” records in a field)

To toggle your view:

1. In the Profile pane, Results pane or data grid, click the **More options**  $\dots$  menu for a numeric or date field.



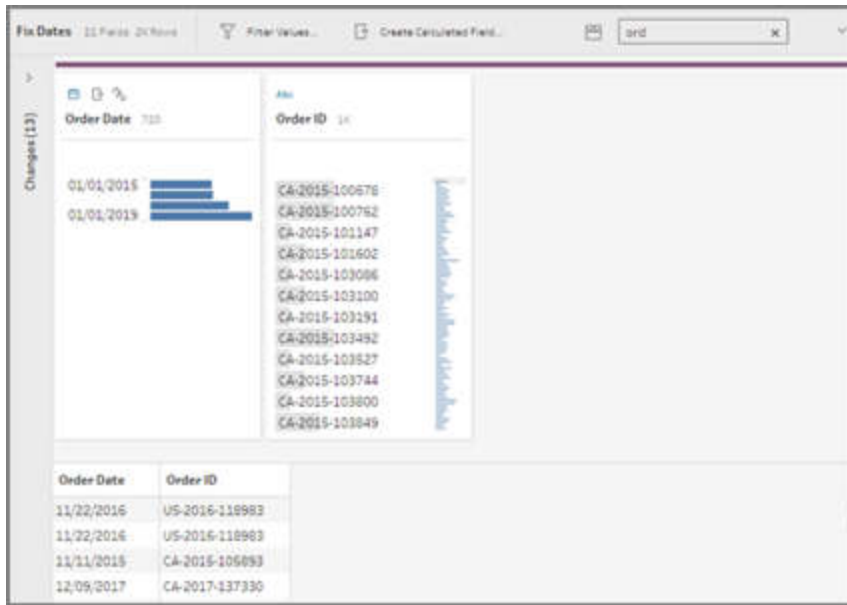
2. In the context menu, select **Detail** to see the detailed version of the values, or **Summary** to see the distributed version of the values.

## Search for fields and values


In the Profile pane or Results pane, you can search for fields or values of particular interest to you and use the search results to filter your data.

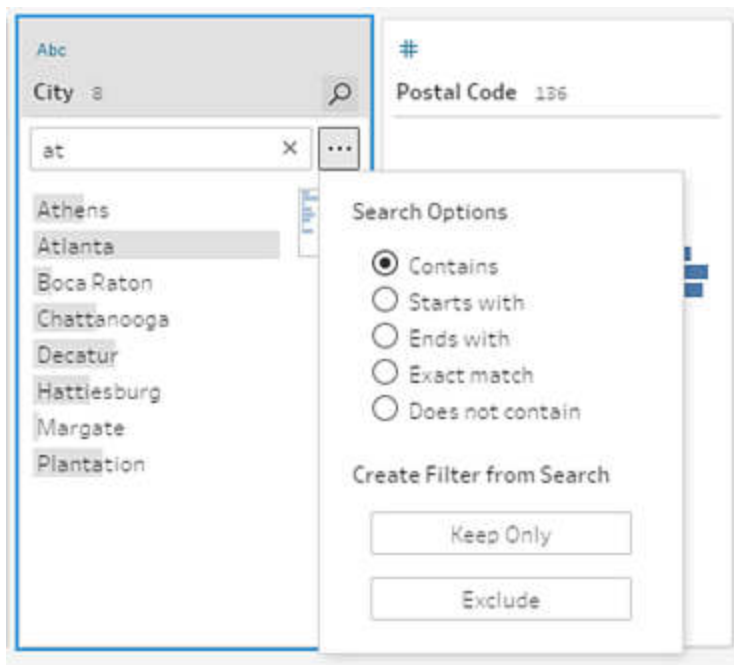
To search for fields, enter a full or partial search term in the search box on the toolbar.

## Tableau Prep Help

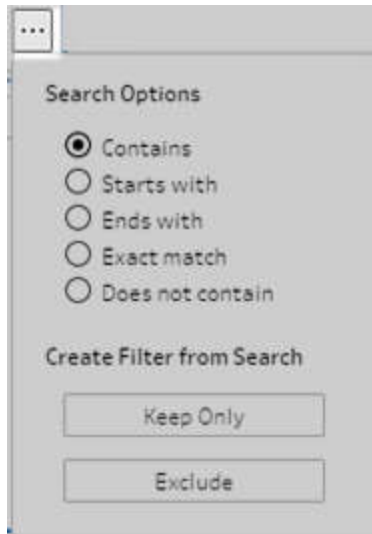


To search for a value in a field:

1. Click the Search icon  for a field, and enter a value.



2. To use advanced search options, click the **Search options...** button.

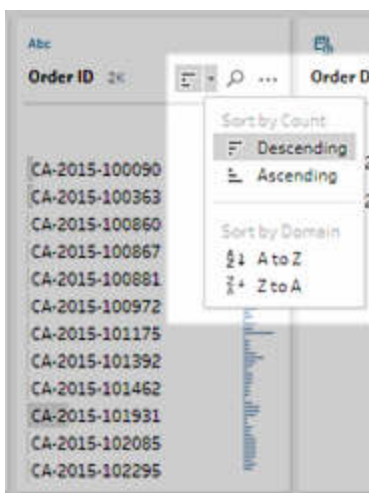


3. To use the search results to filter the data, select **Keep Only** or **Exclude**.

In the **Flow** pane, a filter icon appears above affected steps.

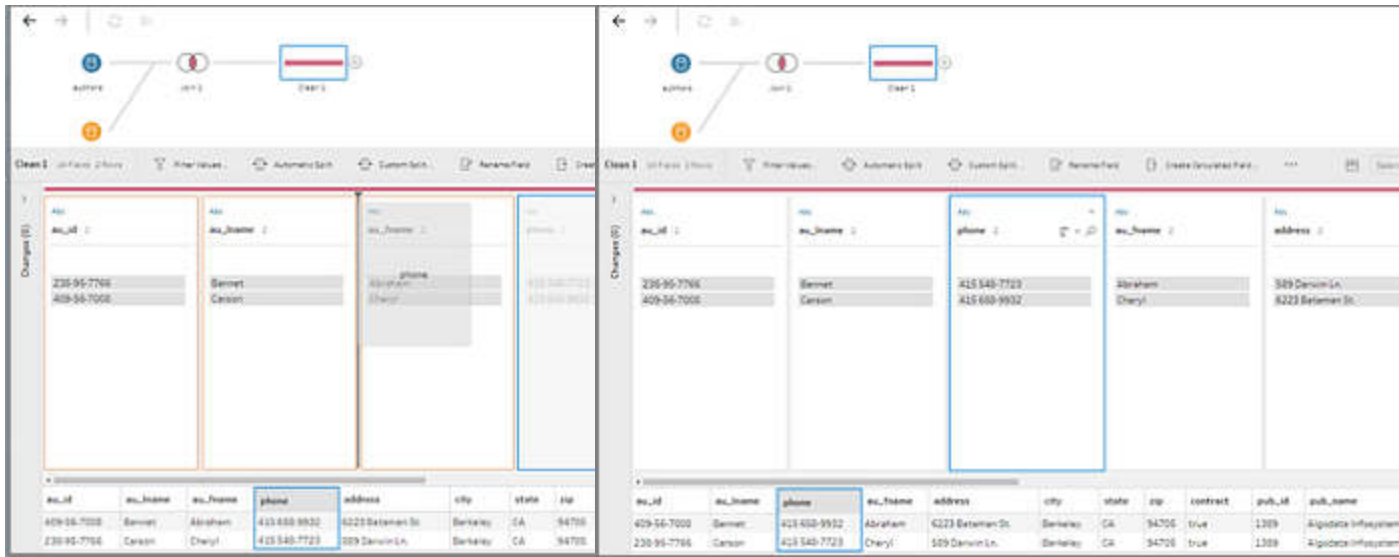
## Sort values and fields

Sort options on a profile card let you sort the bins (the count of values represented by the distribution bars) in ascending or descending order, or the individual field values in alphabetical order.



If you want to rearrange the order of your fields, in the **Profile** pane, **Results** pane or **Data** grid simply select a profile card or field in the data grid and drag it until you see the black target

line appear. Then drop it into place. The Profile pane and data grid are synced so the field will appear in the same order in both places.



## Highlight fields and values in a flow

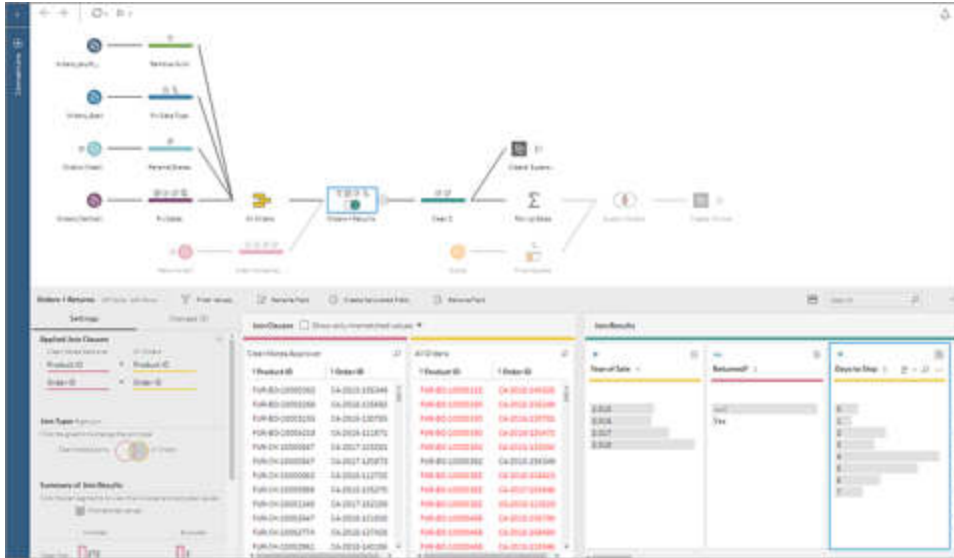
Tableau Prep Builder makes it easy to find fields and values in your flow data. Trace where a field originated and where it is used throughout the flow in the flow pane, or click individual values in a profile card or in the data grid to highlight related or identical values.

### Trace fields in a flow

In Tableau Prep Builder version 2018.3.3 and later, you can highlight everywhere a field is used in a flow, even where it originated to help you track down missing values or troubleshoot a flow when you aren't seeing the results you expect.

Click on a field in the **Profile** pane in a cleaning step or in the **Results** pane in any other step type and the flow pane will highlight the path where that field is used.

**Note:** This option is not available for Input or Output step types.



## See related values

You can use highlighting to find related values across fields. When you click a value in the **Profile** card in the Profile pane or Results pane, all the related values in the other fields are highlighted in blue. The blue color shows the relationship distribution between the value you selected and the values in the other fields.

For example, to highlight related values, in the **Profile** pane, click a value in a field. The related values in other fields turn blue and the proportion of the bar highlighted in blue represents the degree of association.



## Highlight identical values

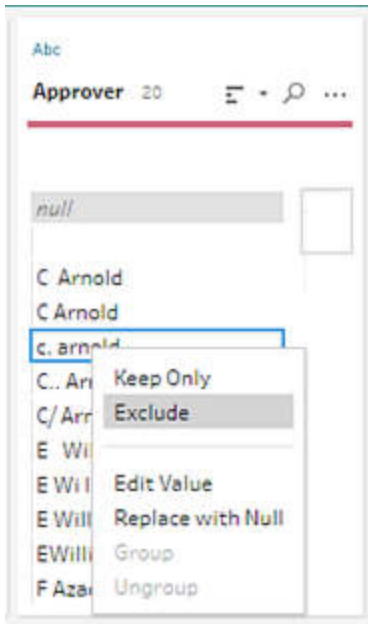
When you select a value in the data grid, all identical values are highlighted too. These highlights help you identify patterns or irregularities in your data.

Type	Customer	Purchases	Date
Cash	Wei	5	08/18/2016
Cash	Jim	7	07/15/2016
Credit	Arnold	5	06/29/2016
Credit	Lee	1	08/07/2016
Cash	Maria	2	08/30/2016
Cash	Wendy	1	07/21/2016
Credit	Max	2	07/02/2016
Credit	Juan	1	05/10/2016
Cash	Isaac	4	06/28/2016
Credit	Philip	1	08/09/2016
Credit	Lane	5	05/04/2016

## Filter Your Data

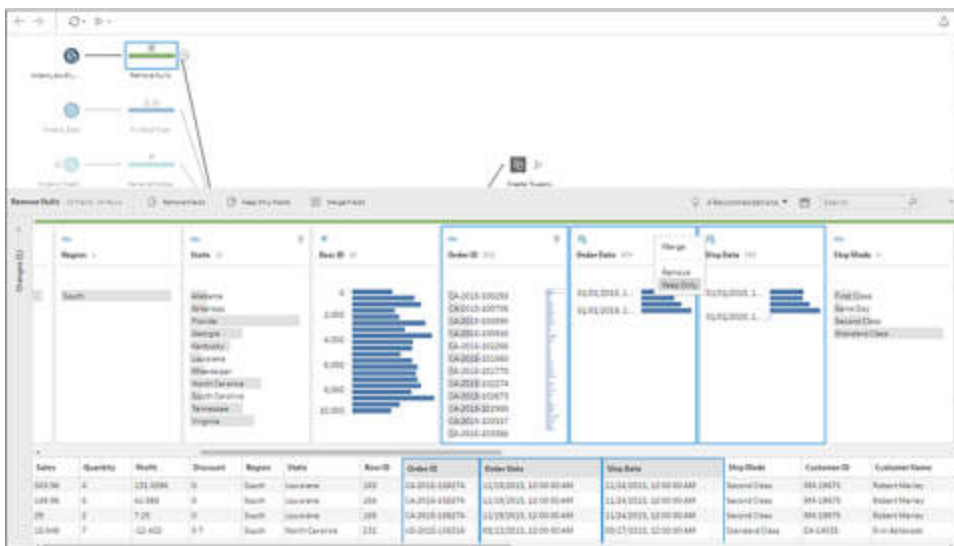
Tableau Prep Builder provides various options that you can use to filter your data. For example, use **Keep Only** or **Exclude** to do one-click filtering on a specific value for a field in a profile card, data grid or results card, or select from a variety of filter options for more complex filtering needs. You can also keep or remove entire fields.

Filter data at any step in the flow. If you want to simply change a specific value, you can select **Edit Value** to edit the value in-line or replace the value with Null. For more information about editing field values, see [Edit field values on page 246](#).



## Keep or remove fields

As you work with your data in your flow you might want to remove unwanted fields. In the Profile pane or the data grid in any cleaning or action step, select one or more fields and right-click or Ctrl-click (MacOS) and select **Remove** to remove the selected fields, or select **Keep Only** (version 2019.2.2 and later) to keep only the selected fields and remove all unselected fields.





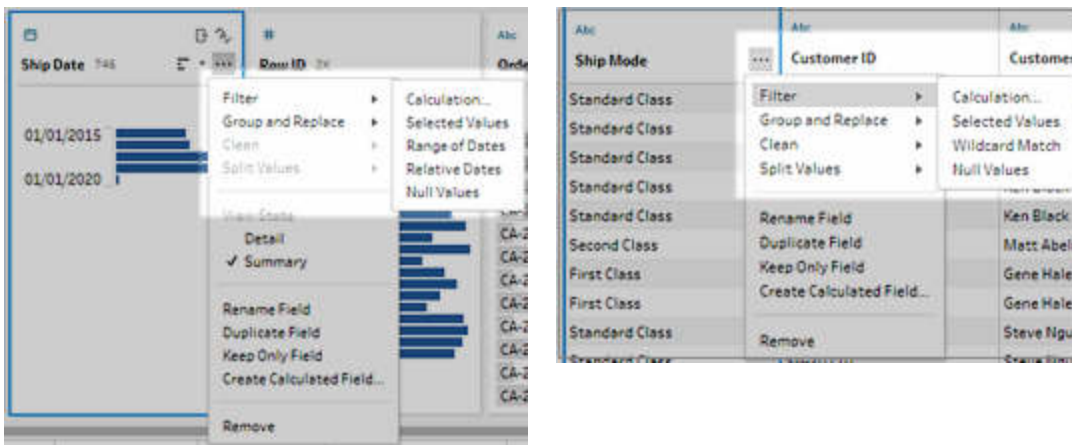
## Filters available for each data type

Data type	Available filters
String	Calculation, Wildcard Match, Null Values, Selected Values
Number	Calculation, Range of Values, Null Values, Selected Values
Date and Date & Time	Calculation, Range of Dates, Relative Date, Null Values, Selected Values

## Where are my filter options?

To see the different filter options available for your fields, on the profile card, in the data grid or in the results pane, click the **More options** ... menu. To see the menu on the data grid, you

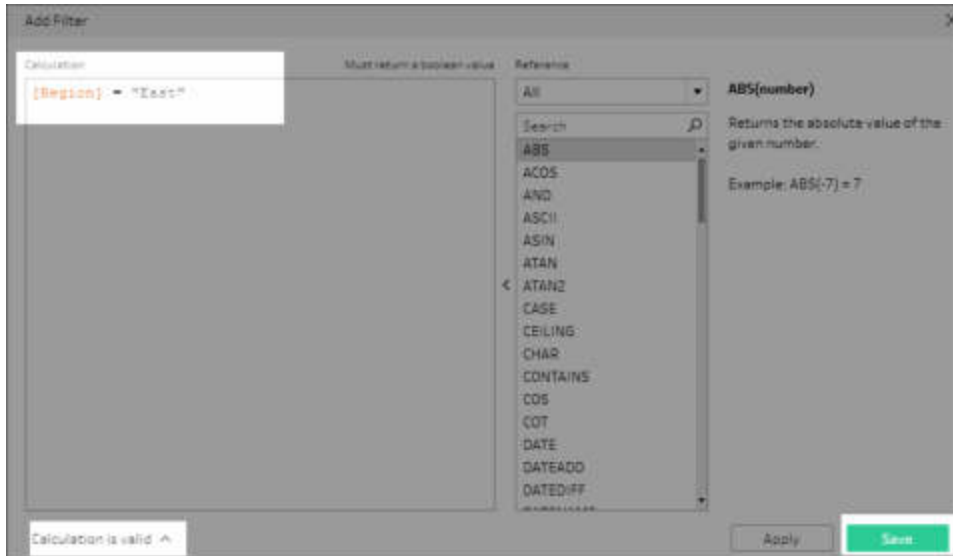
must click the **Hide profile pane**  button first, and then click **More options** ... .



## Calculation filter


When you select **Calculation**, the **Add Filter** dialog box opens. Enter the calculation, verify that it's valid, and click **Save**.

**Note:** In the Input step this is the only type of filter that is available. All other filter types are available in the profile cards, data grid or results pane.



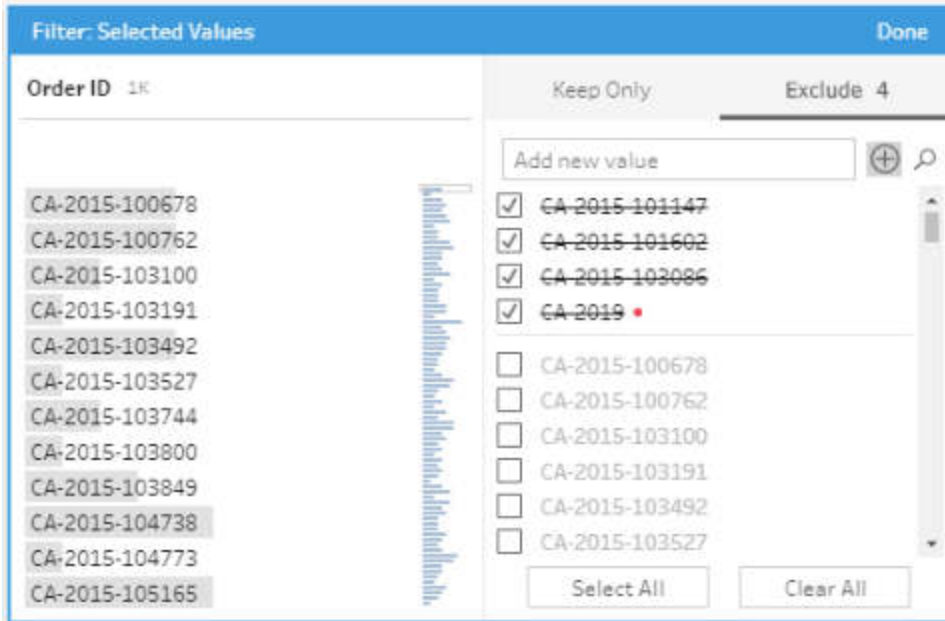
## Selected Values filter (version 2019.2.3 and later)

With the **Selected Values** filter you can pick and choose the values that you want to keep or exclude for a field, even values that aren't in your sample. In the right pane, click the **Keep Only** or **Exclude** tab to select your action, then enter search terms to search for values or click

**Add a value**  to add values that are in your data set but aren't included in your sample.

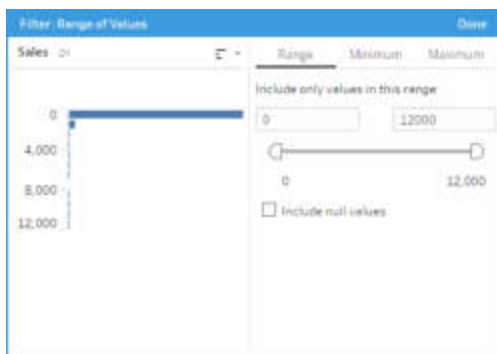
Click **Done** to apply your filter.

**Note:** this filter options isn't available for **Aggregation** or **Pivot** step types.



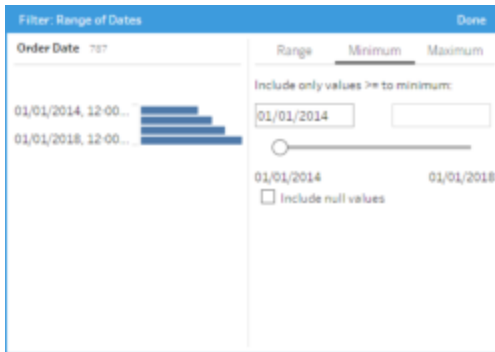
## Range of Values filter

Filter out values that fall within a specific range. When you select **Range of Values**, you can specify a range or set minimum or maximum values.



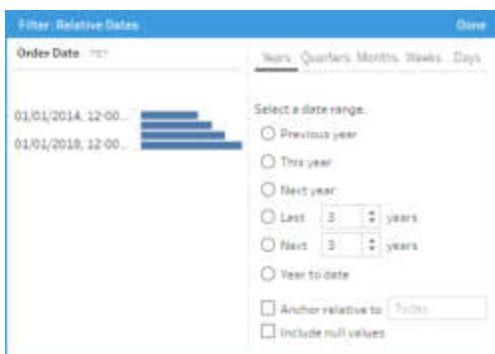
## Range of Dates filter

Filter out values that fall within a specific date range. When you select **Range of Dates**, you can specify a range of dates or set a minimum or maximum date.



## Relative Date filter

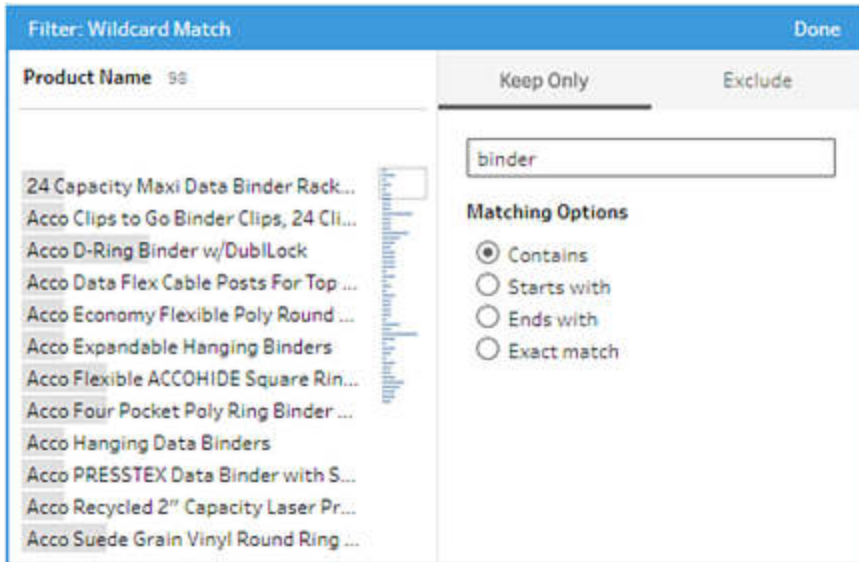
Use the **Relative Dates** filter to specify the exact range of years, quarters, months, weeks, or days that you want to see in your data. You can also configure an anchor relative to a specific date, and include null values.



## Wildcard Match filter

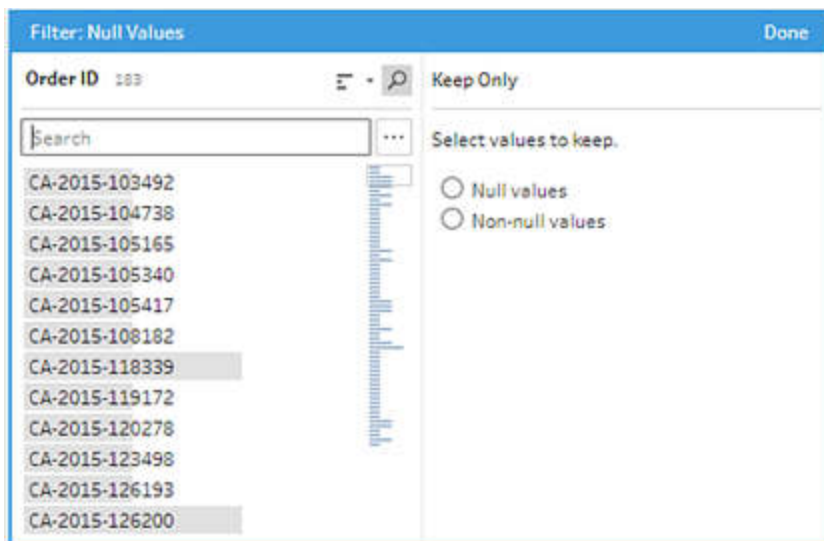
When you select **Wildcard Match**, you can filter the field values to keep or exclude values that match a pattern. In the filter editor, select the **Keep Only** or **Exclude** tab, enter a value to match and then set the **Matching Options** criteria to return the values you are looking for.

The filtered results display in the left pane of the filter editor so that you can review and experiment with your results. Once you have the results you want, click **Done** to apply your change.



## Null Values filter

When you select **Null Values** you can filter the values in the selected field to show only null values or exclude all null values.



## Use Data Roles to Validate your Data

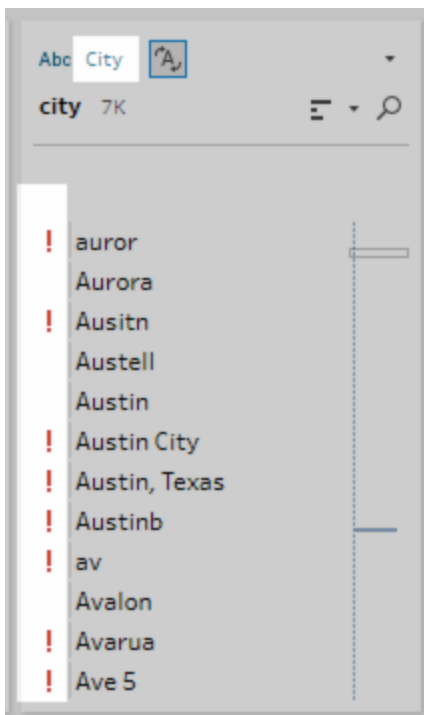
Use data roles to quickly identify whether the values in a field are valid or not. Tableau Prep Builder delivers a standard set of data roles that you can select from or you can create your own

using the unique field values in your data set.

When you assign a data role, Tableau Prep Builder compares the standard values defined for the data role with the values in your field. Any values that don't match are marked with a red exclamation mark. You can filter your field to view only the valid or invalid values and take the appropriate actions to fix them. Once you've assigned a data role to your fields, you can use the **Group and Replace** option to group and match invalid values to valid ones based on spelling and pronunciation.

## Assign standard data roles to your data

Assign data roles provided by Tableau Prep Builder to your field the same way you assign a data type. The data role identifies what your data values represent so Tableau Prep Builder can automatically validate values and highlight ones that aren't valid for that role.



For example if you have field values for geographical data, you can assign a data role of **City** and Tableau Prep Builder compares the values in the field to a set of known domain values to identify values that don't match.

**Note:** Each field is analyzed independently so a City value of "Portland" in State "Washington" in Country "USA" might not be a valid city and state combination, but it won't be identified that way because it is a valid city name.

Tableau Prep Builder provides the following data roles:

- Email
- URL
- Geographic roles (Based on current geographic data and is the same data used by Tableau Desktop)
  - Airport
  - Area code (U.S.)
  - CBSA/MSA
  - City
  - Congressional District (U.S.)
  - Country/Region
  - County
  - NUTS Europe
  - State/Province
  - Zip code/Postal code

**Tip:** In Tableau Prep Builder version 2019.1.4 and later, if you assign a geographic role to a field, you can also use that data role to match and group values with the standard value defined by your data role. For more information about grouping values using data roles, see [Clean and Shape Data on page 229](#).

To assign a data role to a field, do the following:

1. In the Profile pane, Results pane or data grid, click the data type for the field.
2. Select the data role for the field.

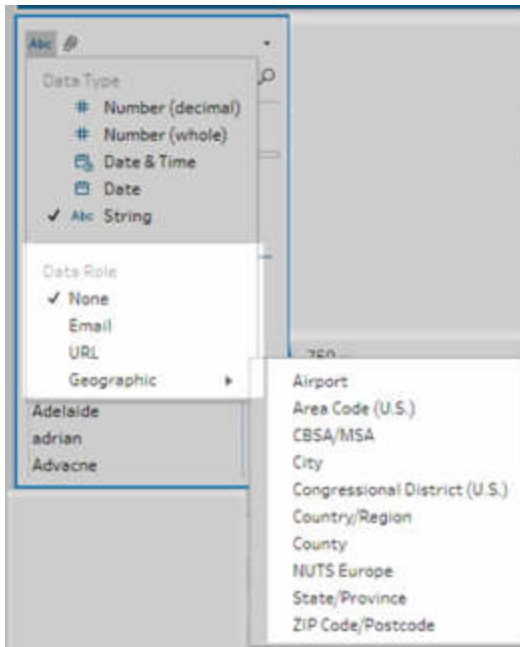
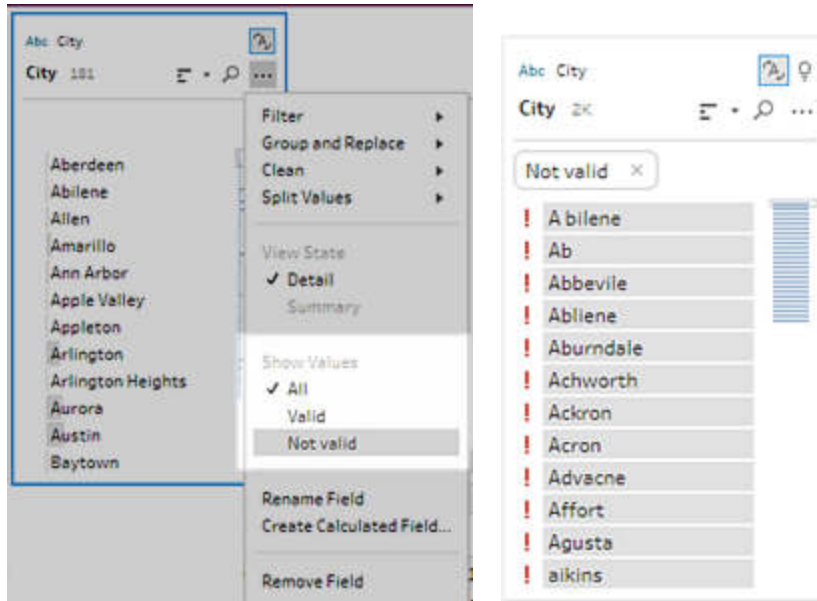


Tableau Prep Builder compares the field's data values to known domain values or patterns (for email or URL) for the data role you select and marks any values that don't match with a red exclamation point.

3. Click the drop-down arrow for the field and from the **Show Values** section select an option to show all values or only values that are valid or not valid for the data role.





4. Use the cleaning options on the **More options** ... menu for the field to correct any values that aren't valid. For more information about how to clean your field values see [About cleaning operations](#) on page 229.

## Create custom data roles (version 2019.3.1 and later)


You can create your own custom data roles using the field values in your data sets to create a standard set of values that you or others can use to validate your field values when cleaning data. Select the field that you want to use, apply any cleaning operations to it if needed, then, publish it to Tableau Server or Tableau Online to use it in your flow or share your data roles with others.

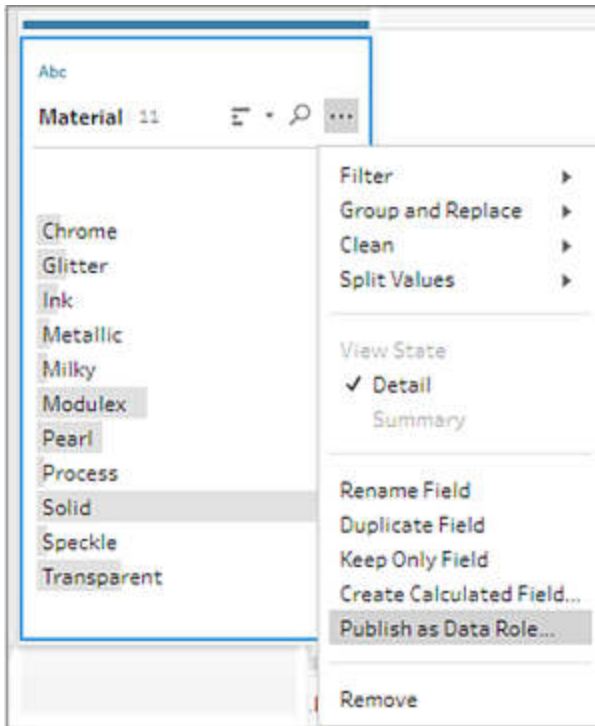
Before creating a custom data role, review the following:

- You can create custom data roles from single fields in your data set. Creating custom data roles from a combination of fields isn't supported.
- You can create custom data roles only for fields assigned to a data type of **String** and **Number (whole)**.
- When you create a custom data role Tableau Prep Builder creates an output step in your flow that is specific to publishing the data role.

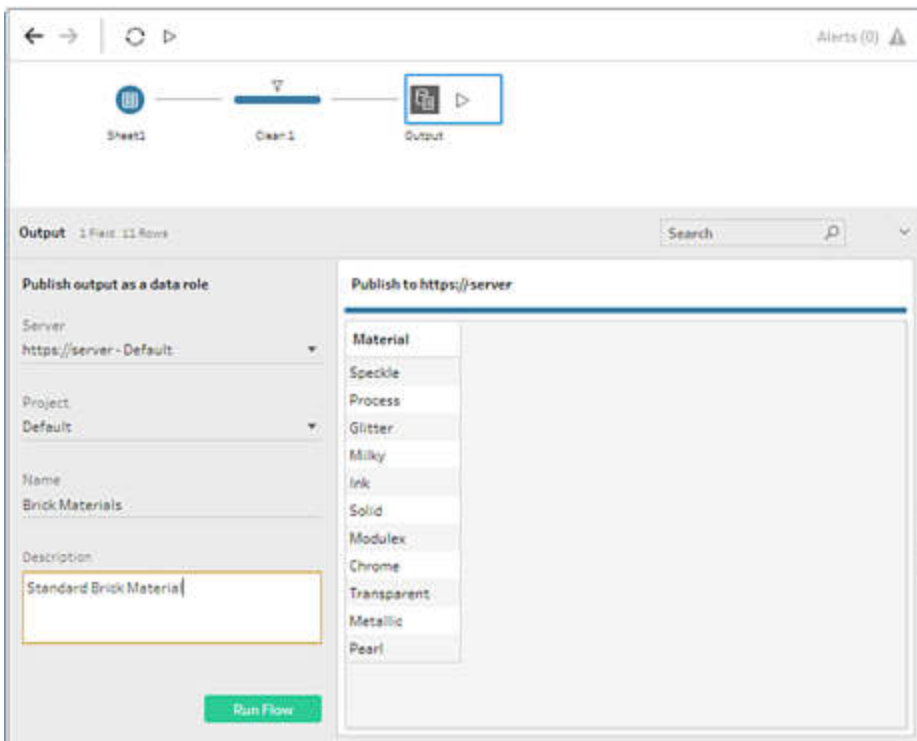
- Publishing custom data roles to multiple sites in the same flow isn't supported. If you publish the flow, you must publish the custom data role to the same site or server where the flow is published.
- Custom data roles are specific to the site, server and project where you publish them. All users with permissions to the location can use the custom data role, but must be signed in to the site or server to select it or apply it. Custom data roles are assigned the default permission for the **All Users** group to the **All Users** permissions for new projects instead of **None**.
- Custom data roles aren't version specific. When applying a custom data role, the most current version is applied.
- Once published to Tableau Server or Tableau Online user with access to the site, server and project can view all data roles in that location.
  - Users with appropriate permissions can move, delete or edit permissions for the data roles.
  - The permissions you can set and actions you can take on a custom data role are similar to what you can do with a flow. For more information, see [Manage a Flow on page 421](#). For more information on setting permissions, see [Permission capabilities](#) in the Tableau Server help.
- To edit a data role, you must make your changes in Tableau Prep Builder, then republish the data role using the same name to overwrite it, as you would to edit a published data source.

## Create a custom data role

1. In the **Profile** pane, data grid, or Results pane select the field you want to use to create a custom data role.
2. Click **More options**  for the field, and select **Publish as Data Role**.




3. Select the server and project where you want to publish the data role.



4. Click **Run Flow** to create the data role. After the publishing process completes successfully, you can view your data role in Tableau Server or Tableau Online. Processing the data role can take some time based on the load on your Tableau Server or Tableau Online site. If your data role isn't available right away, wait a few minutes, then try selecting it again.

[Explore](#) / [Default](#) / Brick Materials

 **Brick Materials** ☆ ...

Owner: [datarockstar](#) Modified: Aug 9, 2019, 3:39 PM

---

**Definition**

**About**

Role Type	Data Type
Dictionary	String
Description	Standard Brick Material

**Values (11)**

Name
Chrome
Glitter
Ink
Metallic
Milky
Modulex
Pearl
Process
Solid
Speckle
Transparent

## Apply a custom data role

1. In the Profile pane, Results pane or data grid, click the data type for the field where you want to apply the custom data role.
2. Select Custom then select the data role that you want to apply to the field.

**Important:** Make sure you are signed into the site or server where the data role was published or you won't see this option.

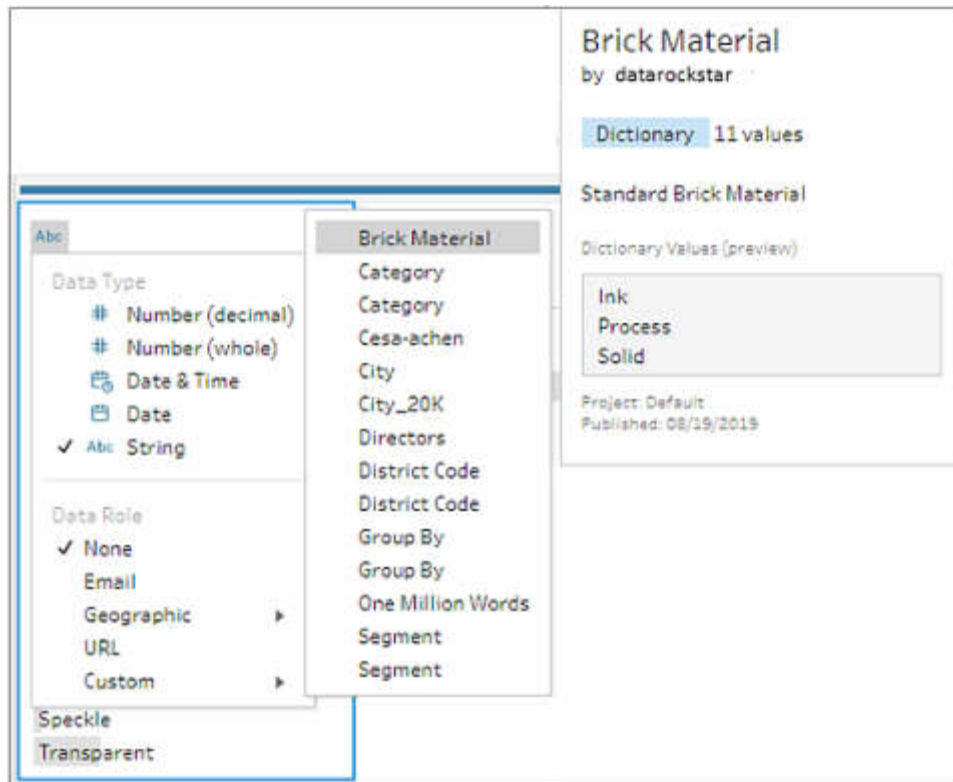
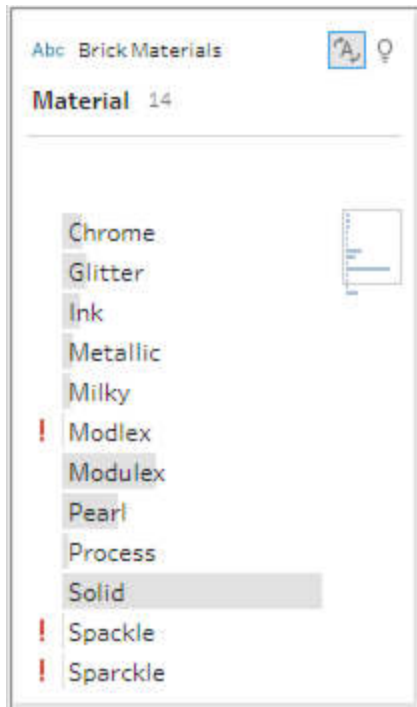
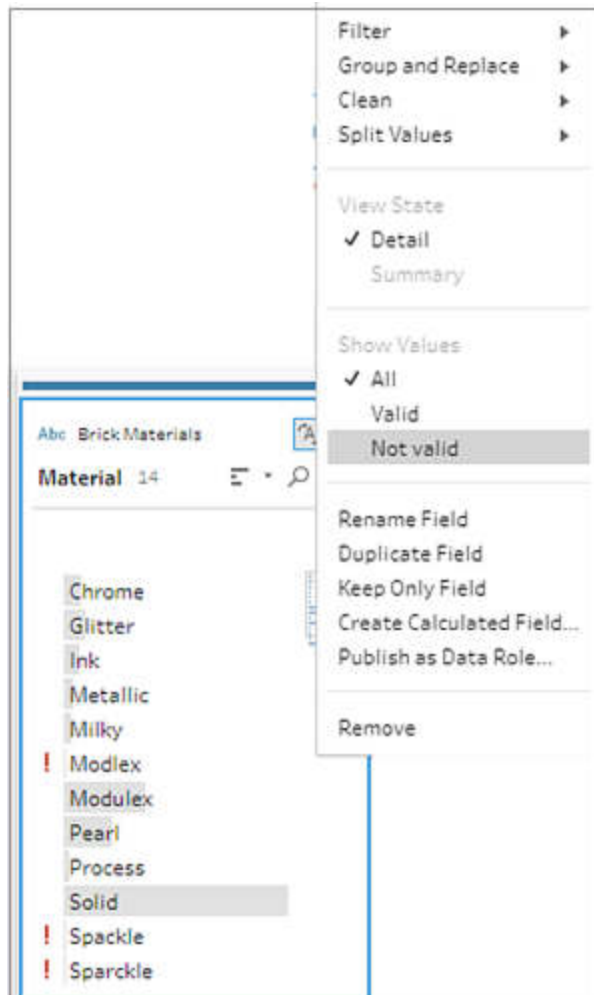


Tableau Prep Builder compares the field's data values to known domain values for the data role you select and marks any values that don't match with a red exclamation point.



3. Click the drop-down arrow for the field and from the **Show Values** section select an option to show all values or only values that are valid or not valid for the data role.



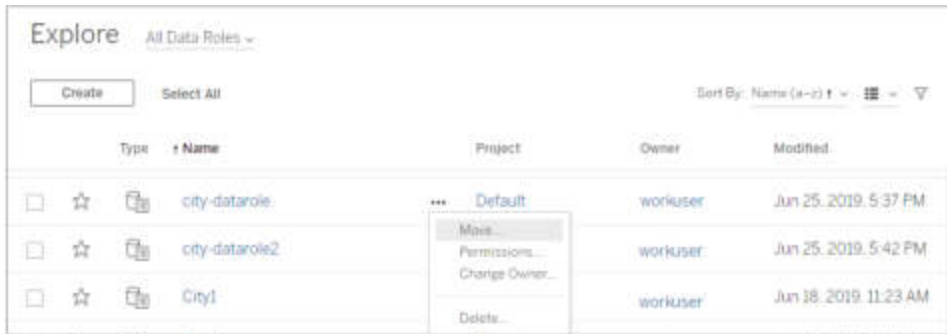
4. Use the cleaning options on the **More options**  $\cdots$  menu for the field to correct any values that aren't valid. For more information about how to clean your field values see [About cleaning operations on page 229](#).

## View and manage custom data roles

You can view and manage your published custom data roles on Tableau Server and Tableau Online. You can view all custom data roles published to your site or server. Click **More actions**

$\cdots$  for a selected data role to move it to a different project, change permissions or delete it.





## Group similar values by data role

**Note:** In Tableau Prep Builder version 2019.1.4 and 2019.2.1 this option was labeled **Data Role Matches**.

If you assign a geographic data role to a field you can use the values in the data role to group and match values in your data field based on spelling and pronunciation to standardize them. In Tableau Prep Builder version 2019.2.3, you can use either **Spelling** or **Spelling + Pronunciation** to group and match invalid values to valid ones.

These options use the standard value defined by the data role. If the standard value isn't in your data set sample, Tableau Prep Builder adds it automatically and marks the value as not in the original data set. For more information about assigning data roles to fields, see [Assign standard data roles to your data on page 214](#).

To use data roles to group values, complete the following steps.

1. In the Profile pane, Results pane or data grid, click the data type for the field.
2. Select one of the following data roles for the field:
  - Airport
  - City
  - Country/Region
  - County
  - State/Province

In Tableau Prep Builder version 2019.3.2 you can also select from your custom data roles

**Standard data roles (ver-**      **Custom data roles (version 2019.3.2 and later)**

### Tableau Prep Builder version 2019.1.4 and later)

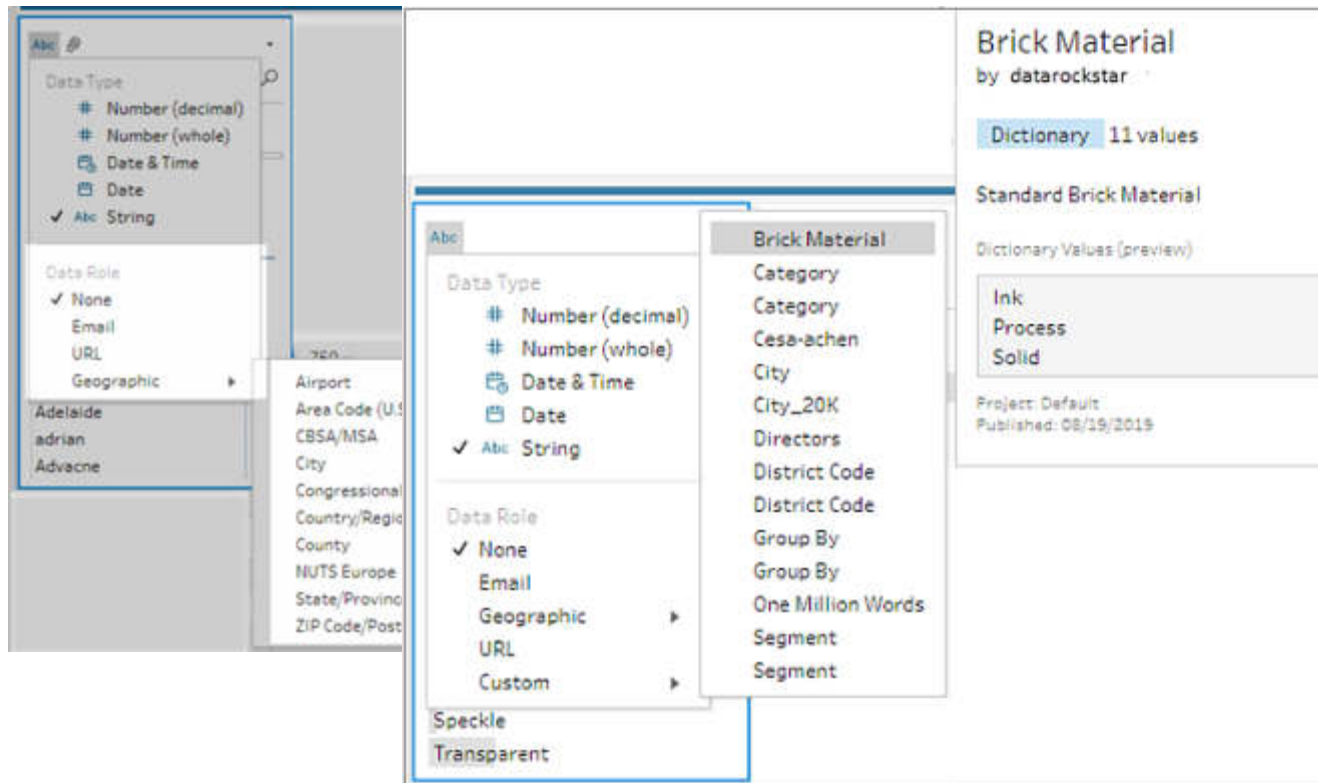

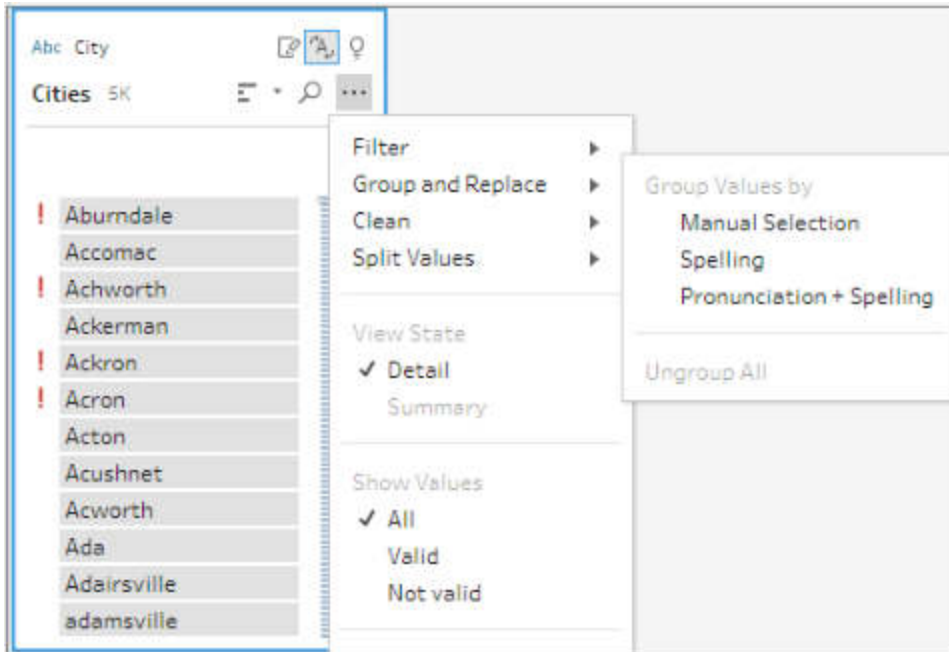


Tableau Prep Builder compares the field's data values to known domain values for the data role you select and marks any values that don't match with a red exclamation point.

- Click **More options** , select **Group and Replace**, then select one of the following options:
  - Spelling:** Matches invalid values to the closest valid values that differ by adding, removing, or substituting characters.
  - Pronunciation + Spelling:** Matches invalid values to the most similar valid value based on spelling and pronunciation.

**Note:** In Tableau Prep Builder version 2019.1.4 or 2019.2.1, this option was called **Data Role Matches**.




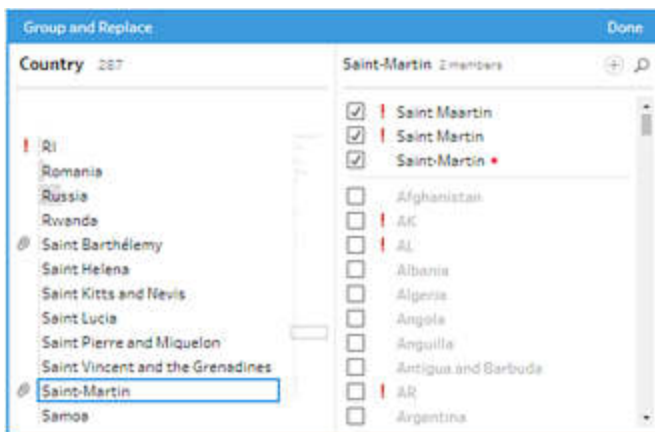
You can also click on the **Recommendations**  icon on the field to apply the recommendation to group and replace the invalid values with valid ones. This option uses the **Pronunciation + Spelling** Group and Replace option.

Tableau Prep Builder compares the values by spelling or spelling and pronunciation and then groups similar values under the standardized value for the data role. If the standardized value isn't in your data set, the value is added and marked with a red dot.





# Clean and Shape Data

Tableau Prep Builder provides various cleaning operations that you can use out of the box to clean and shape your data. Cleaning up dirty data makes it easier to combine and analyze your data or makes it easier for others to understand your data when sharing your data sets.

You can also clean your data using a pivot step or a script step to apply R or Python scripts to your flow. For more information, see [Pivot Your Data on page 281](#) or [Use R and Python scripts in your flow on page 290](#).

## About cleaning operations

You clean data by applying cleaning operations such as filtering, adding, renaming, splitting, grouping, or removing fields. In prior releases, cleaning operations could only be performed in the Clean step type. In Tableau Prep Builder version 2018.2.1 and later, you could also perform cleaning operations in the data grid in a cleaning step.

In Tableau Prep Builder version 2018.3.1 and later, you can perform cleaning operations in most step types in your flow. You can apply limited cleaning operations in the Input step and can't apply cleaning operations in the output step. For more information about applying cleaning operations in the Input step, see [Apply cleaning operations in an input step on page 166](#).

## Available cleaning operations

The following table shows which cleaning operations are available in each step type:

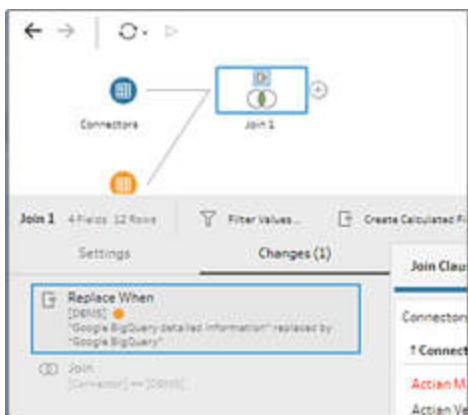
	Input	Clean	Aggregate	Pivot	Join	Union	Output
Filter	X	X	X	X	X	X	
Group and Replace		X		X		X	
Clean		X		X	X	X	
Convert Dates		X	X	X	X	X	
Split Values		X		X	X	X	

Rename Field	X	X	X	X	X	X	
Duplicate Field		X		X	X	X	
Keep Only Field	X	X	X	X	X	X	
Remove Field	X	X	X	X	X	X	
Create Calculated Field		X		X	X	X	
Edit Value		X		X	X	X	
Change Data Type	X	X	X	X	X	X	

As you make changes to your data, annotations are added to the corresponding step in the **Flow** pane and an entry is added in the **Changes** pane to track your actions. If you make changes in the Input step, the annotation shows to the left of the step in the **Flow** pane and is also shown in the **Input profile** in the field list.

The order that you apply your changes matters. Changes made in Aggregate, Pivot, Join and Union step types are performed either before or after those reshaping actions, depending on where the field is when you make the change. Where the change was made is shown in the **Changes** pane for the step.

The following example shows a field change from a single table in a Join step. The change is performed prior to the join action to give the corrected results.



## Order of operations

The following table shows where the cleaning action is performed in Aggregate, Pivot, Join and Union step types depending on where the field is in the step.

Action	Step Type:	Aggregate	Aggregate	Pivot	Pivot	Join	Join	Union	Union
	<i>Field Location:</i>	<i>Grouped fields</i>	<i>Aggregated fields</i>	<i>Not in pivot</i>	<i>Created from pivot</i>	<i>Included in one table*</i>	<i>Included in both tables*</i>	<i>Mismatched fields</i>	<i>Combined fields</i>
<b>Filter</b>		Before Aggregation	After Aggregation	Before Pivot	After Pivot	Before Join	After Join	Before Union	After Union
<b>Group and Replace</b>		NA	NA	Before Pivot	After Pivot	Before Join	After Join	Before Union	After Union
<b>Clean</b>		NA	NA	Before Pivot	After Pivot	Before Join	After Join	Before Union	After Union
<b>Convert Dates</b>		Before Aggregation	After Aggregation	Before Pivot	After Pivot	Before Join	After Join	Before Union	After Union
<b>Split Values</b>		NA	NA	Before Pivot	After Pivot	Before Join	After Join	Before Union	After Union

<b>Renam- e Field</b>		Before Aggreg- ation	After Aggreg- ation	Bef- ore Pivo- t	After Pivot	Befor- e Join	After Join	Before Union	After Union
<b>Duplic- ate Field</b>		NA	NA	Bef- ore Pivo- t	After Pivot	Befor- e Join	After Join	Before Union	After Union
<b>Keep Only Field</b>		After Aggreg- ation	After Aggreg- ation	Bef- ore Pivo- t	After Pivot	Befor- e Join	After Join	Before Union	After Union
<b>Remov- e Field</b>		Remove- s from Aggreg- ation	Remove- s from Aggreg- ation	Bef- ore Pivo- t	After Pivot	Befor- e Join	After Join	Before Union	After Union
<b>Create Cal- culated Field</b>		NA	NA	Bef- ore Pivo- t	After Pivot	After Join	After Join	Before Union	After Union
<b>Edit Value</b>		NA	NA	Bef- ore Pivo- t	After Pivot	Befor- e Join	After Join	Before Union	After Union
<b>Chang- e Data Type</b>		Before Aggreg- ation	After Aggreg- ation	Bef- ore Pivo- t	After Pivot	Befor- e Join	Befor- e Join	Before Union	After Union

**Note:** For joins, if the field is a calculated field that was created using a field from one table, the change is applied before the join. If the field is created with fields from both tables, the change is applied after the join.

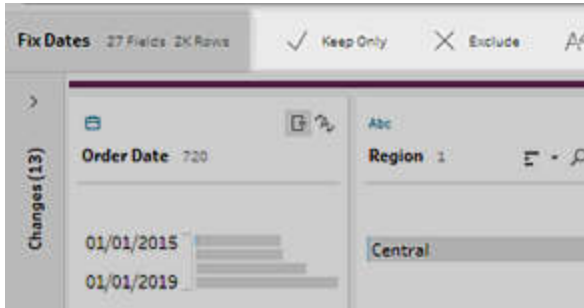


# Apply cleaning operations

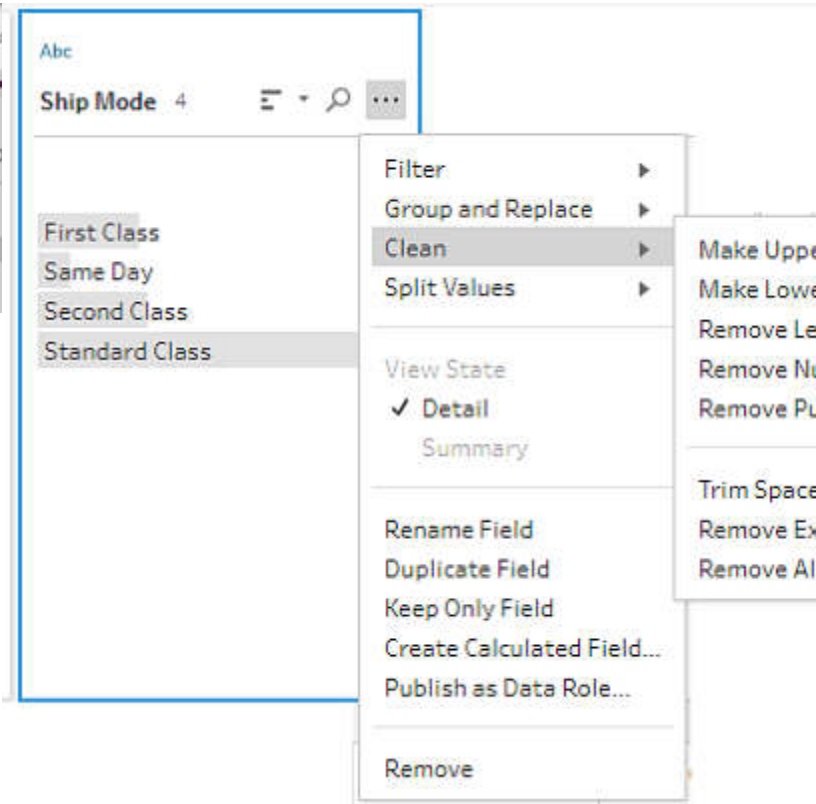
To apply cleaning operations to fields, use the toolbar options or click **More options** ... on the field profile card, data grid or Results pane to open the menu.

In Aggregate, Pivot, Join and Union step types, the **More options** ... menu is available on the profile cards in the Results pane and corresponding data grid. If you perform the same cleaning operations or actions over and over throughout your flow, you can copy and paste your steps, actions or even fields. For more information see [Copy steps, actions and fields on page 258](#).

## Profile pane toolbar



## Drop-down menu



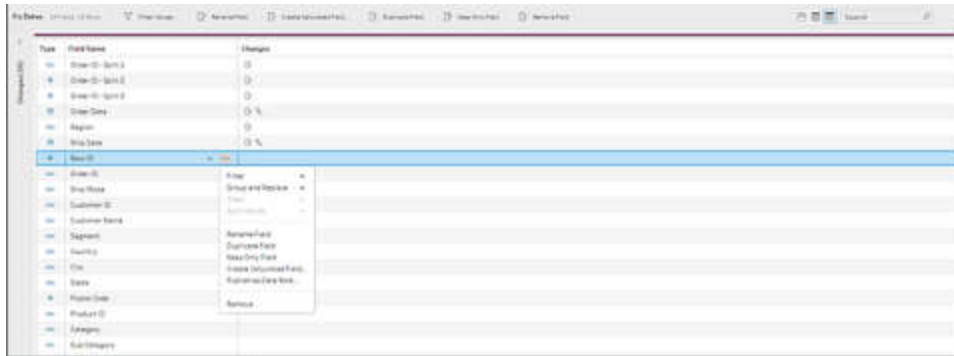
## Select your view

You can perform cleaning operations outside of the profile or results pane in the data grid or in the list view. Use the view toolbar



your flow but you can change it at any time.

You can select multiple rows and click the **X** to remove them or use the **More options** ... menu to apply operations to selected fields. If you assign a data role to the field, or select **Filter**, **Group and Replace**, **Clean**, or **Split Values**, you will be returned to the Profile or Results view to complete those actions. All other options can be performed in the list view.



## Version 2019.3.1 and earlier

Use the view toolbar    to hide the Profile pane and just show the

data grid. Then click **More options** ... on a field in the data grid to open the cleaning menu. This view provide a more detailed view of your data and can be useful when you need to work with specific field values. After you select this option, this view state persists across all steps in your flow but you can change it at any time.

**Note:** Not all cleaning operations are available in the data grid. For example if you want to edit a value in-line, you must do this in the Profile pane.

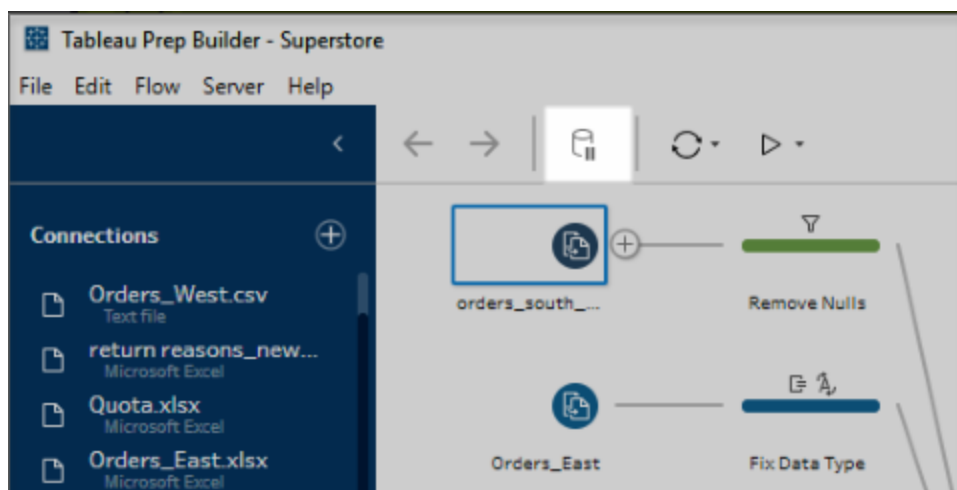
Order ID	Order ID	Order ID	Order Date	Region	Ship Date	Row ID	Row ID	Step Name
10	10	10	11/11/2016	Central	11/11/2016	10	10	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	11	11	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	12	12	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	13	13	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	14	14	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	15	15	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	16	16	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	17	17	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	18	18	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	19	19	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	20	20	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	21	21	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	22	22	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	23	23	Standard Case
10	10	10	11/11/2016	Central	11/11/2016	24	24	Standard Case

## Pause data updates to boost performance

As you perform cleaning operations on your data, Tableau Prep Builder applies your changes as you go to show you the results immediately. But if you know exactly what changes you want to make and don't need the immediate feedback as you make each change or just want a boost in performance, you can pause the data updates to save valuable processing time. When you pause data updates, you can make all your changes at once, then resume updates to see your results. You can resume data updates and enable all available operations at any time.

**Note:** When you pause data updates, any operations that require you to see your values will be disabled. For example if you want to apply a filter to selected values, you need to see the values you want to exclude.

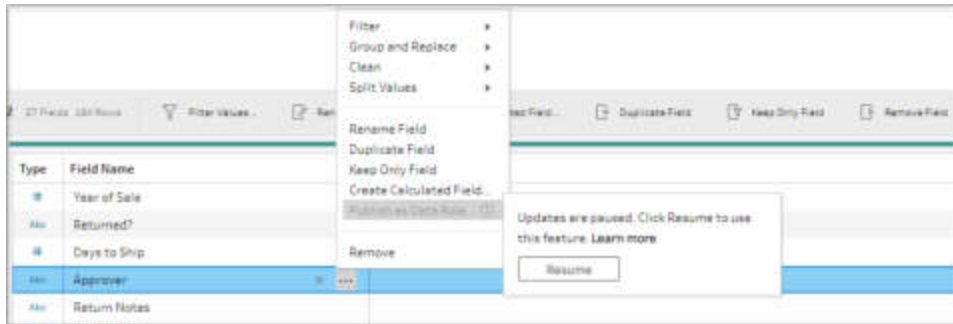
1. In the top menu, click **Pause data updates** to pause updates.



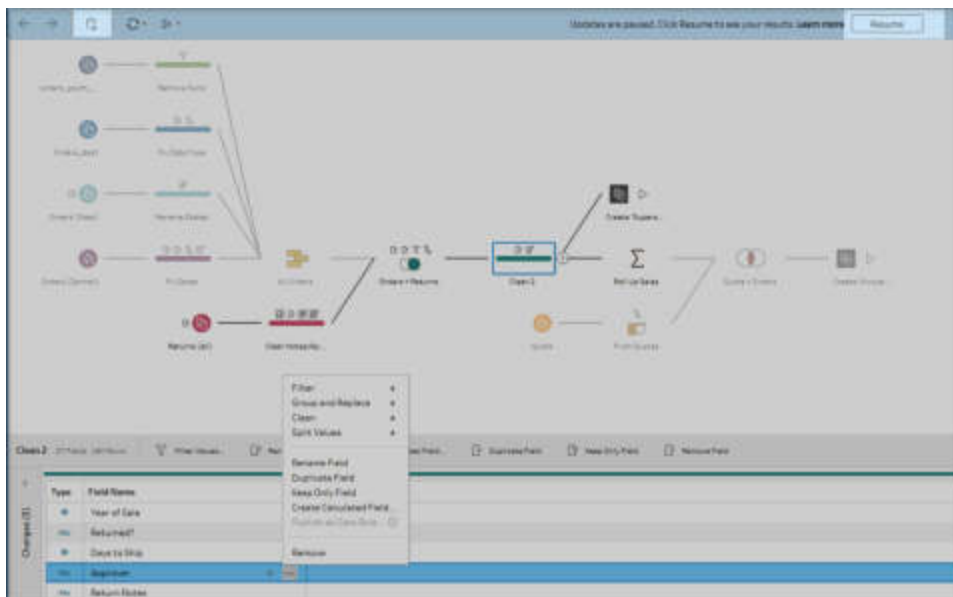
2. Tableau Prep Builder converts the Profile pane into the List view. In List view, use the

**More options** ... menu to apply operations to selected fields. If the operation requires you to see your values, it will be disabled. To enable the operation, you will need to resume data updates.

For more information about using List view mode, see [Select your view on page 233](#).




3. To see the results of your changes or enable a disabled feature, resume data updates. Click the **Resume data updates** button, click the **Resume** button in the menu dialog or in the message banner at the top of the Flow pane.



## Apply cleaning operations

To apply cleaning operations to a field, do the following:

**Note:** You can perform cleaning operations in a list view beginning in version 2019.3.2.

1. In the **Profile** pane, data grid, Results pane or list view, select the field you want to make changes to.
2. From either the toolbar or **More options**  menu for the field, select from the following options:
  - **Filter** or **Filter Values**: Select from one of the filter options, right-click or Ctrl-click (MacOS) a field value to keep or exclude values. You can also use the **Selected Values** filter to pick and choose the values to filter, included values not in your flow sample. For more information about filter options, see [Filter Your Data on page 207](#).
  - **Group and Replace**: Manually select values or use automatic grouping. You can also multi-select values in the Profile card and right-click Ctrl-click (MacOS) to group or ungroup values or edit the group value. For more information about using Group and Replace, see [Automatically map values to a standard value using fuzzy match on page 254](#).
  - **Clean**: Select from a list of quick cleaning operations to apply to all values in the field.
  - **Convert Dates** (version 2020.1.4 and later): For fields assigned to a Date or Date & Time data type, select from a list of DATEPART quick cleaning operations to convert your date field values to an integer value representing year, quarter, month, week, day, or a date and time value.
  - **Split Values**: Split values automatically based on a common separator or use custom split to specify how you want to split field values.  
  
Automatic split and custom split work the same as they do in Tableau Desktop. For more information, see [Split a Field into Multiple Fields](#) in the Tableau Desktop and Web Authoring Help.
  - **Rename Field**: Edit the field name.
  - **Duplicate Field** (version 2019.2.3 and later): Create a copy of your field and values.
  - **Keep Only Field**(version 2019.2.2 and later): Keep only the selected field and

exclude all other fields in the step.

- **Create Calculated Field:** Write a custom calculation in the Calculation editor or use the Visual Calculation editor (version 2020.1.1 and later) to create level of detail, rank or row number calculations. For more information, see [Create level of detail and rank calculations](#) on page 268.
- **Remove (Remove Field** in previous versions): Remove the field from the flow.

**Note:** Starting in version 2019.3.1, you can use the **Publish as Data Role** option to create custom data roles that you can then apply to your fields to validate the field values when cleaning data. For more information about this option, see [Create custom data roles \(version 2019.3.1 and later\)](#) on page 217.

3. To edit a value, right-click or Ctrl-click (MacOS) one or more values and select **Edit Value** then enter a new value. You can also select **Replace with Null** to replace the values with a Null value or double-click in a single field to edit it directly. For more information about editing field values see [Edit field values](#) on page 246.
4. Review the results of these operations in the **Profile** pane, Summary panes or data grid.

## View your changes

The different types of cleaning operations are represented by icons over the steps in your flow. If more than four types of operations are applied to a step, an ellipsis displays over the step. Hover over these icons to view annotations showing applied operations and the order in which they are performed.

Starting in Tableau Prep Builder version 2019.1.3 and later, you can click on an annotation on the change icon on a step in the Flow pane or on a profile card in the Profile or Results pane and the change and field it impacts will be highlighted in the **Changes** pane and the **Profile** or **Results** pane.

The screenshot displays the Tableau Prep Builder interface. At the top, a flow diagram shows data sources: 'Orders (Week)', 'Orders (Sample)', and 'Returns (All)'. These feed into a 'Fix Dates' step, which then leads to 'All Orders'. A 'Calculated Field' step is also shown, with fields for 'Region', 'Order Date', and 'Ship Date'. Below the flow, a data table is visible with columns for 'Product ID', 'Order ID', 'Ship Date', and 'Order ID'. The 'Changes (13)' pane is expanded, showing a list of changes applied to the data. The changes include:

- Calculated Field [Region] "Central"
- Calculated Field [Order Date] (STR[Order Day] + "/" + STR[Order Month] + "/" + STR[Order Year])
- Change Type [Order Date] To Date type
- Calculated Field [Ship Date] (STR[Ship Day] + "/" + STR[Ship Month] + "/" + STR[Ship Year])
- Rename Field [Discount] From [discounts] to [Discount]
- Rename Field [Product Name] From [Product] to [Product Name]
- Change Type [Ship Date] To Date type

The 'Changes' pane also shows a preview of the data with columns for 'Order Date', 'Region', and 'Ship Date'. The data rows are:

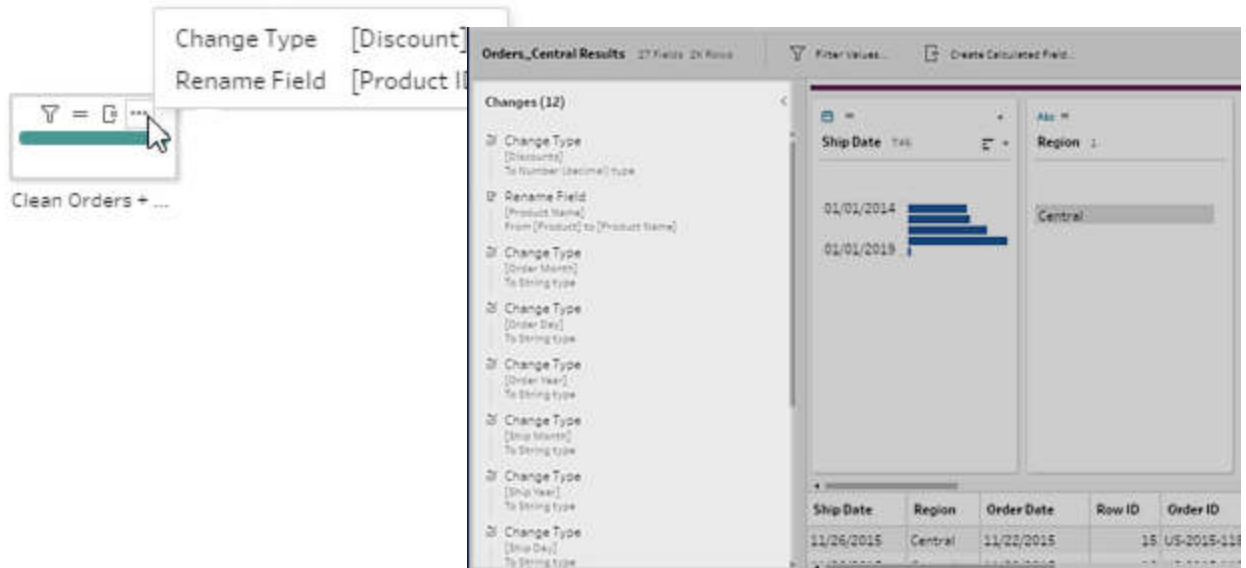
Order Date	Region	Ship Date	Row ID	Order ID
11/22/2016	Central	11/26/2016	15	US-2016-118983

You can also select a step and then expand the **Changes** pane to view the details for each change, edit or remove your changes, drag changes up or down to change the order in which they're applied and in Tableau Prep Builder version 2019.1.1, you can add a description to provide context to other users. For more information about adding descriptions to your changes, see [Add descriptions to flow steps and cleaning actions](#) on page 189

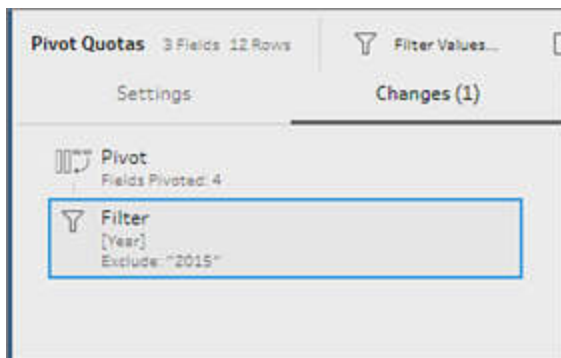
**Cleaning annotation**

**Changes pane**





When viewing changes in an Aggregation, Pivot, Join, or Union step, the order that the change is applied shows either before or after the reshaping action. The order of these changes is applied by the system and cannot be changed. You can edit and remove the change.



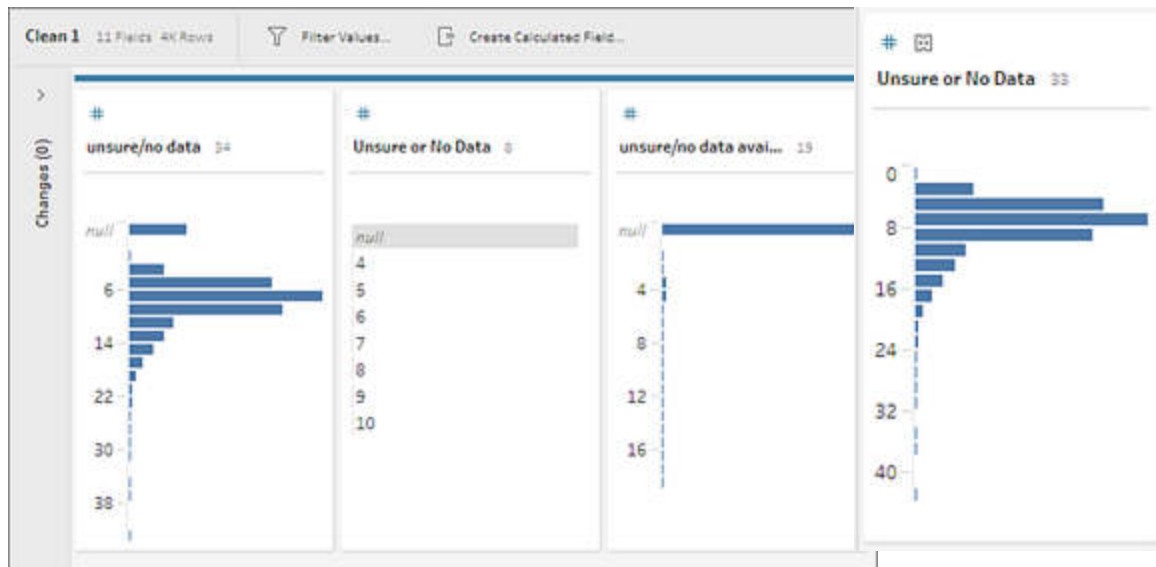
## Merge fields

If you have fields that contain the same values but are named differently, you can easily merge them into a single field to combine them by dragging one field on top of the other. When you merge the fields, the target field becomes the primary field and the field name of the target field persists. The field that you merge to the target field is removed.

### Example:

Wildcard union results in 3 fields with the same values

Merge 3 fields into 1



When you merge fields, Tableau Prep Builder keeps all of the fields from the target field and replaces any nulls in that field with values from the source fields that you merge with the target field. The source fields are removed.

### Example

Name	Contact_Phone	Business_Phone	Cell_Phone	Home_Phone
Bob	123-4567	123-4567	null	null
Sally	null	null	456-7890	789-0123
Fred	null	null	null	567-8901
Emma	null	234-5678	345-6789	null

If you merge the **Business\_Phone**, **Cell\_Phone** and **Home\_Phone** fields with the **Contact\_phone** field, the other fields are removed and results in the following:

Name	Contact_Phone
Bob	123-4567

Sally	456-7890
Fred	567-8901
Emma	234-5678

To merge fields, do one of the following:

- Drag and drop one field onto another. A **Drop to merge fields** indicator displays.
- Select multiple fields and right-click within the selection to open the context menu, and then click **Merge Fields**.
- Select multiple fields, and then click **Merge Fields** on the toolbar.

For information about how to fix mismatched fields as a result of a union, see [Fix fields that don't match on page 317](#).

## Apply cleaning operations using recommendations (version 2018.3.2 and later)

Sometimes it can be hard to identify which cleaning operation you need to use to fix problems in your data. Tableau Prep Builder can analyze your data and recommend cleaning operations that you can apply automatically to quickly fix problems in your data fields or help to identify problems so you can fix them. This feature is available in all step types except Input, Output and Join step types.

**Note:** If you don't want to use this feature, you can turn it off. From the top menu, go to **Help > Settings and Performance**. Then click on **Enable Recommendations** to clear the check mark next to the setting.

Recommendation types include:

- Data roles
- Filter
- Group and Replace (also applies to fields with data roles starting in version 2019.2.3)
- Pivot columns to rows (2019.4.2 and later)
- Replace values with Null values


- Remove fields
- Split (version 2019.1.1 and later)

**Note:** This option works specifically with data in fixed-width type text files. To use the split recommendation with this file type, after you connect to the data source, in the Input step, in the **Text Settings** tab, select a **Field Separator** character that is not used in the data so the data loads as a single field.

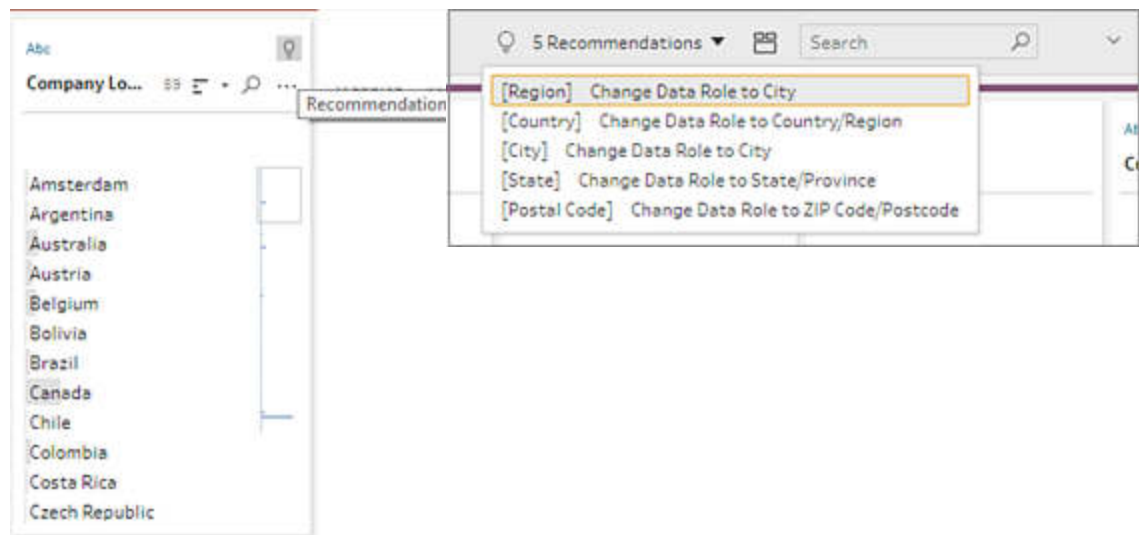
- Trim spaces

## Apply recommendations

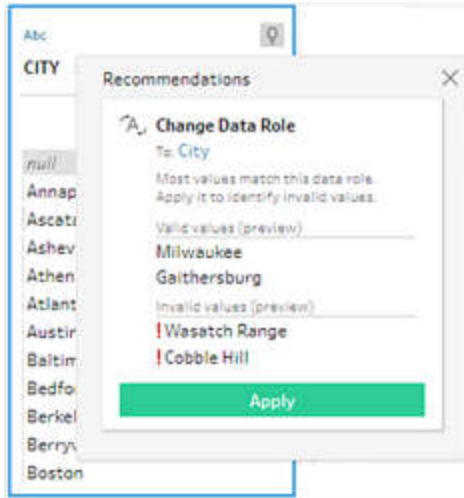
1. Do one of the following:

- Click the light bulb  icon in the top right corner of the profile card.
- From the toolbar, click the **Recommendations** drop-down arrow to view all recommendations for your data set and select a recommendation from the list.

This option only appears when recommended changes are identified by Tableau Prep Builder.



2. To apply the recommendation, hover on the Recommendations card and then click **Apply**.

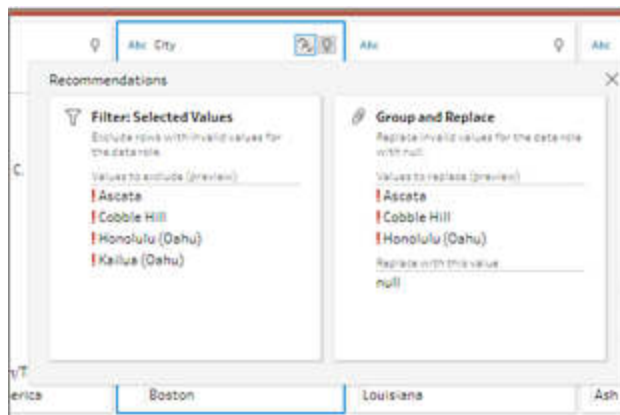


The change is automatically applied and an entry is added to the **Changes** pane. To remove the change, click **Undo** in the top menu or hover over the change in the **Changes** pane and click the **X** to remove it.

If you apply a recommendation to pivot fields (version 2019.4.2 and later), a Pivot step is automatically created where you can then perform any additional pivot actions like renaming the pivoted fields or pivoting on additional fields.

3. If Tableau Prep Builder identifies further recommendations as a result of the change, the light bulb icon remains on the Profile card until no further recommendations are found.

Repeat the steps above to apply any additional changes or ignore the suggested change and use the other cleaning tools in Tableau Prep Builder to address the data problems.




# Edit field values

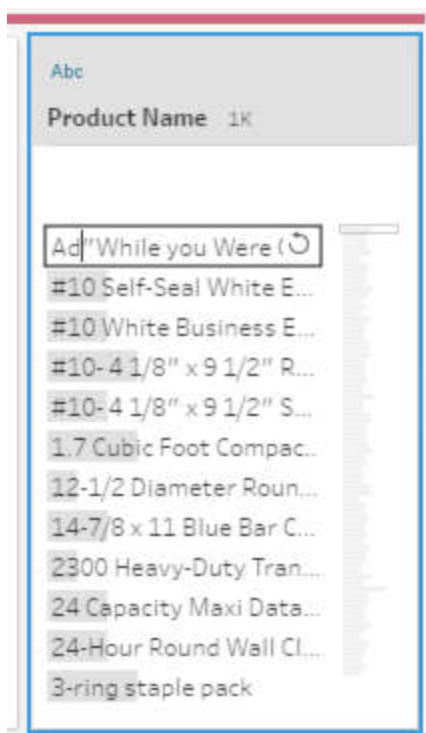
Multiple variations of the same value can prevent you from accurately summarizing your data. You can quickly and easily correct these variations using the following options in Tableau Prep Builder.

**Note:** Any edits that you make to the values must be compatible with the field data type.

## Edit a single value

1. In the **Profile** card, click the value you want to edit, and enter the new value. A group icon  shows next to the value.


Alternatively, right-click a value and click **Edit Value**. The change is recorded in the **Changes** pane on the left side of the screen.



2. View the results in the **Profile** pane, and data grid.


## Edit multiple values

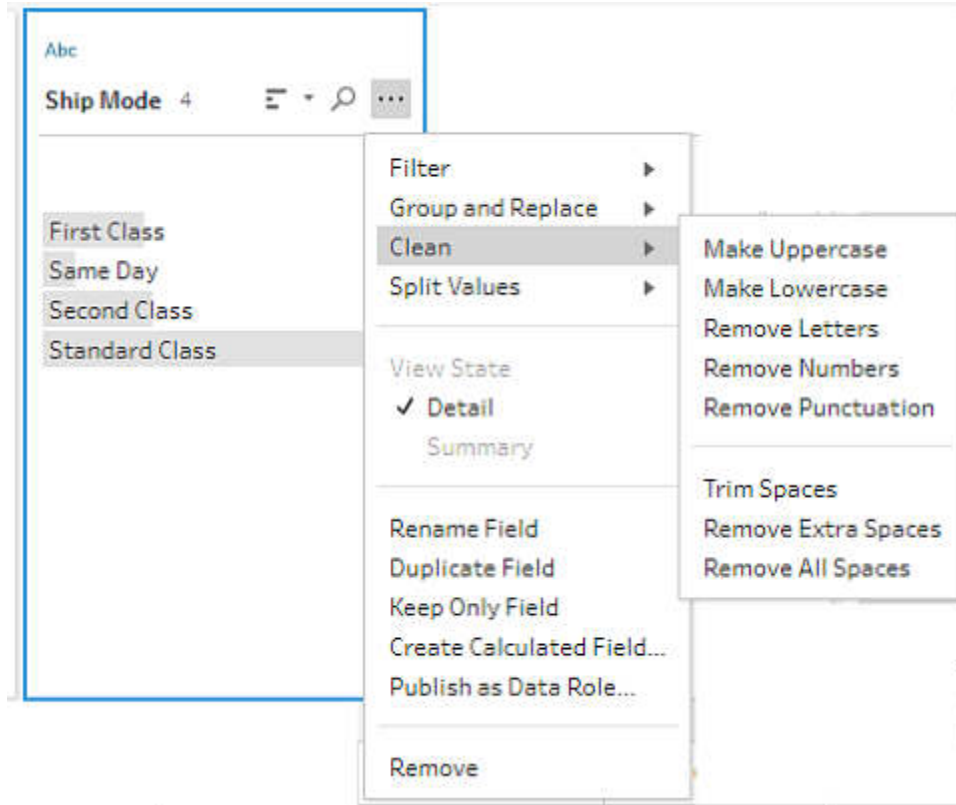
You have several options to edit multiple values at once. For example, use quick cleaning operations to remove punctuation for all values in a field, manually group values using multi-select, automatically group values together using fuzzy-match algorithms that find similar values or select multiple values and replace them with Null.

**Note:** When you map multiple values to a single value, the original field shows a group icon  next to the value, showing you which values are grouped together.

### Edit multiple values using quick cleaning operations

This option applies only to text fields.

1. In the **Profile** pane, Results pane or data grid, select the field you want to edit.
2. Click **More options**  , select **Clean**, and then select one of the following options:
  - **Make Uppercase:** Change all values to uppercase text.
  - **Make Lowercase:** Change all values to lowercase text.
  - **Remove Letters:** Remove all letters and leave only other characters.
  - **Remove Numbers:** Remove all numbers and leave letters and other characters.
  - **Remove Punctuation:** Remove all punctuation.
  - **Trim Spaces:** Remove leading and trailing spaces.
  - **Remove Extra Spaces** (version 2018.3.3 and later): Remove leading and trailing whitespace and replace extra whitespace in-between characters with a single space.
  - **Remove All Spaces** (version 2018.3.3 and later): Remove all whitespace, including leading and trailing whitespace and any whitespace in between characters.



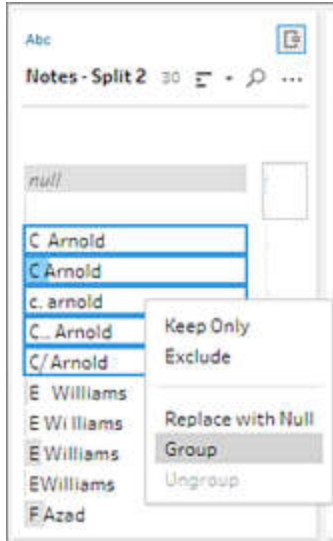
You can stack operations to apply multiple cleaning operations to the fields. For example first select **Clean > Remove Numbers** and then select **Clean > Remove Punctuation** to remove all numbers and punctuation from the field values.

3. To undo your changes, click the **Undo** arrow at the top of the **Flow** pane, or remove the change from the change list.

## Group and edit multiple values inline

1. In the **Profile** card, select the field you want to edit.
2. Press Ctrl or Shift+click or Command or Shift+click (MacOS), and select the values that you want to group.
3. Right-click, and select **Group** from the context menu. The value in the selection that you right-click becomes the default name for the new group but you can edit this in-line.





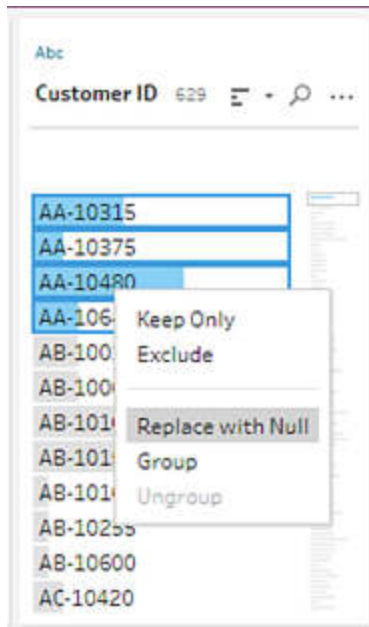
4. To edit the group name, select the grouped field and edit the value or right-click or Ctrl+click (Mac) on the grouped field and select **Edit Value** from the context menu.
5. To ungroup the grouped field values, right-click on the grouped field and select **Ungroup** from the context menu.

## Replace one or more values with Null

If you have data rows that you want to include in your analysis but you want to exclude certain field values you can change them to a Null value.

1. In the **Profile** card, press Ctrl or Shift+click or Command or Shift+click (on Mac), and select the values that you want to change
2. Right-click or Ctrl+click (Mac), and select **Repace with Null** from the menu. The values

are changed to Null and the group icon  shows next to the value.




## Manually map multiple values to a standard value

Use **Group and Replace** to map the value of a field from one value to another or manually select multiple values to group them. You can even add new values to set up mapping relationships to organize your data.

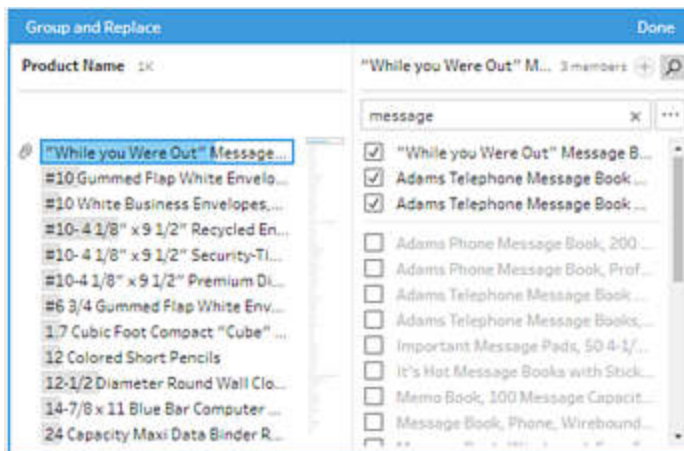
For example, let's say you have three values in a field: My Company, My Company Incorporated, and My Company Inc. All these values represent the same company, My Company. You can use **Group and Replace** to map the values My Company Incorporated and My Company Inc to My Company, so that all three values appear as My Company in the field.

### Map multiple values to a single selected field

1. In the **Profile** pane or Results pane, select the field you want to edit.
2. Click **More options**  and select **Group and Replace > Manual Selection** from the

menu.

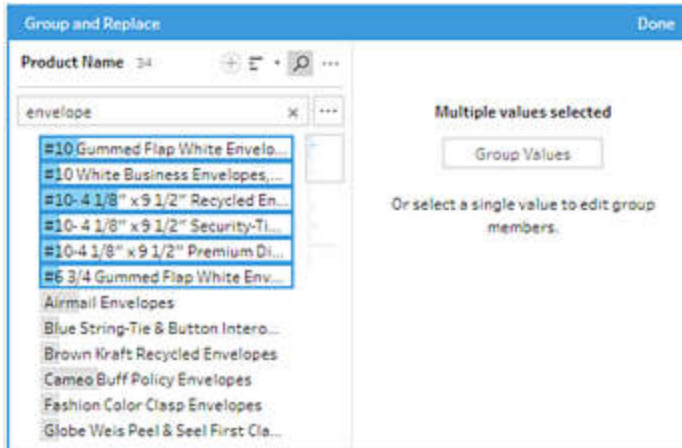
3. In the left pane of the **Group and Replace** editor, select the field value that you want to use as the grouping value. This value now shows at the top of the right pane.
4. In the lower section of the right pane in the **Group and Replace** editor, select the values you want to add to the group.



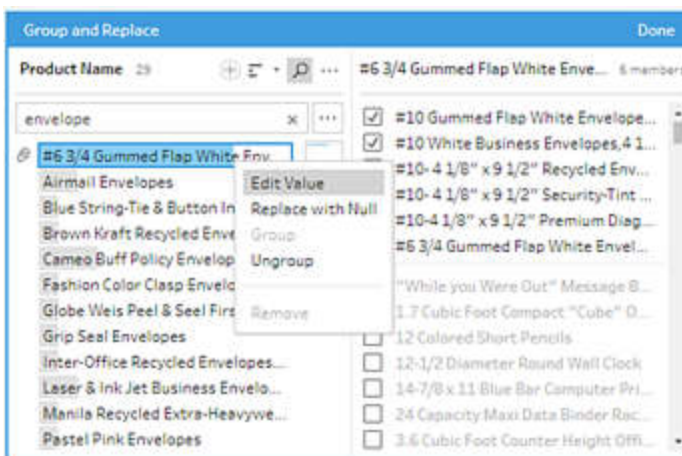
To remove values from the group, in the upper section of the right pane in the **Group and Replace** editor, clear the check box next to the values.

## Create a group by selecting multiple values

1. In the **Profile** pane or Results pane, select the field you want to edit.
2. Click **More options** ... and select **Group and Replace > Manual Selection** from the menu.
3. In the left pane of the **Group and Replace** editor, select multiple values that you want to group.
4. In the right pane of the **Group and Replace** editor, click **Group Values**.



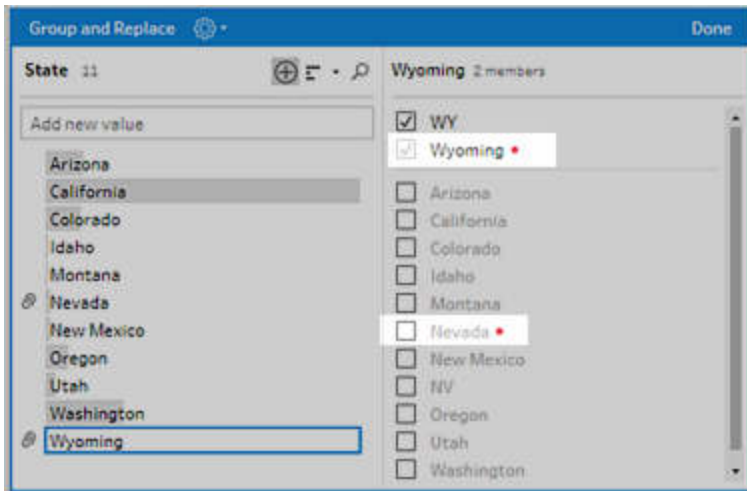
A new group is created using the last selected value as the group name. To edit the group name, select the grouped field and edit the value or right-click or Ctrl+click (Mac) on the grouped field and select **Edit Value** from the menu.



## Add and identify values that aren't in the data set

If you want to map values in your data set to a new value that doesn't exist, you can add it using Group and Replace. To easily identify any values that are not in the data set, these values are marked with a red dot next to the value name in the **Group and Replace** editor.



For example in the image below, Wyoming and Nevada aren't in the data set.

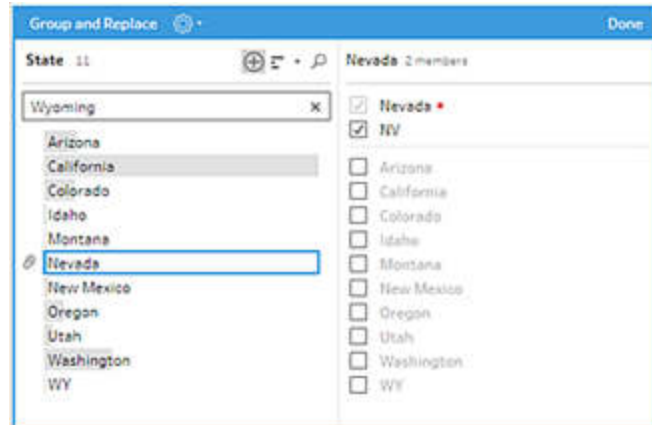


Some reasons why a value might not be in the data set include the following:

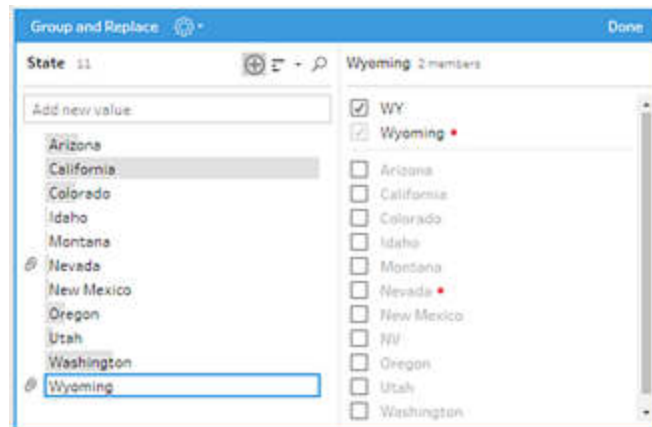
- You just added the new value manually.
- The value is no longer in the data.
- The value is in the data, but isn't in the sampled data set.

To add a new value:

1. In the **Profile** pane or Results pane, select the field you want to edit.
2. Click **More options**  and select **Group and Replace > Manual Selection** from the context menu.
3. In the left pane of the **Group and Replace** editor, click the plus  to add a new value.
4. Type a new value in the field and press Enter to add it.



- In the right pane, select the values that you want to map to the new value.



- (Optional) To add additional new values to your mapped value, click the plus (+) button in the right pane in the **Group and Replace** editor.

## Automatically map values to a standard value using fuzzy match

To search for and automatically group similar values, use one of the fuzzy match algorithms. Field values are grouped under the value that appears most frequently. Review the grouped values and add or remove values in the group as needed.

If you use data roles to validate your field values, you can use the Group and Replace option to match invalid values with valid ones. For more information, see [Group similar values by data role](#) on page 225

Choose one of the following options to group values:

- **Pronunciation:** Find and group values that sound alike. This option uses the Metaphone 3 algorithm that indexes words by their pronunciation and is most suitable for English words. This type of algorithm is used by many popular spell checkers. This option isn't available for data roles.
- **Common Characters:** Find and group values that have letters or numbers in common. This option uses the ngram fingerprint algorithm that indexes words by their unique characters after removing punctuation, duplicates, and whitespace. This algorithm works for any supported language. This option isn't available for data roles.

For example, this algorithm would match names that are represented as "John Smith" and "Smith, John" because they both generate the key "hijmnost". Since this algorithm doesn't consider pronunciation, the value "Tom Jhinois" would have the same key "hijmnost" and would also be included in the group.

- **Spelling:** Find and group text values that are spelled alike. This option uses the Levenshtein distance algorithm to compute an edit distance between two text values using a fixed default threshold. It then groups them together when the edit distance is less than the threshold value. This algorithm works for any supported language.

Starting in Tableau Prep Builder version 2019.2.3 this option is available to use after a data role is applied. In that case, it matches the invalid values to the closest valid value using the edit distance. If the standard value isn't in your data set sample, Tableau Prep Builder adds it automatically and marks the value as not in the original data set.

- **Pronunciation +Spelling:** (version 2019.1.4 and later) If you assign a data role to your fields, you can use that data role to match and group values with the standard value defined by your data role. This option then matches invalid values to the most similar valid value based on spelling and pronunciation. If the standard value isn't in your data set sample, Tableau Prep Builder adds it automatically and marks the value as not in the original data set. This option is most suitable for English words.

For more information see [Clean and Shape Data on page 229](#).

**Note:** In Tableau Prep Builder version 2019.1.4 and 2019.2.1 this option was labeled **Data Role Matches**.

## Group similar values using fuzzy match

1. In the **Profile** pane or Results pane, select the field you want to edit.
2. Click **More options** **⋮** and select **Group and Replace** then select one of these options:

- **Pronunciation**
- **Common Characters**
- **Spelling**

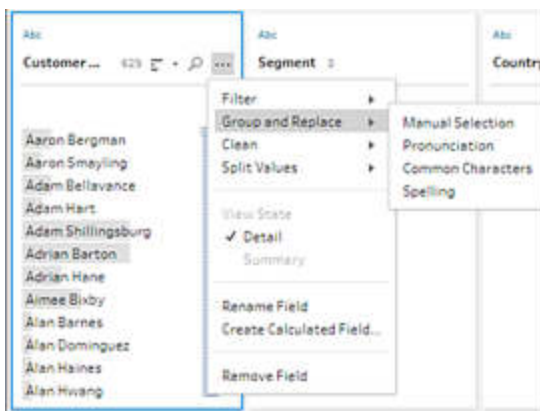
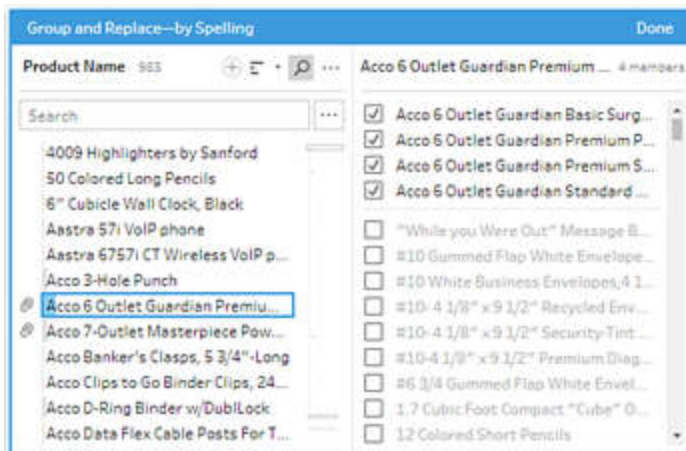


Tableau Prep Builder finds and groups values that match and replaces them with the value that occurs most frequently in the group.

3. Review the groupings and manually add or remove values or edit them as needed. Then



click **Done**.



## Adjust your results when grouping field values

If you group similar values by **Spelling** or **Pronunciation**, you can change your results by using the slider on the field to adjust how strict the grouping parameters are.

Depending on how you set the slider, you can have more control over the number of values included in a group and the number of groups that get created. By default, Tableau Prep Builder detects the optimal grouping setting and shows the slider in that position.

When you change the threshold, Tableau Prep Builder analyzes a sample of the values to determine the new grouping. The groups generated from the setting are saved and recorded in the **Changes** pane, but the threshold setting isn't saved. The next time the **Group and Replace** editor is opened, either from editing your existing change or making a new change, the threshold slider is shown in the default position, enabling you to make any adjustments based on your current data set.

1. In the **Profile** pane or **Results** pane, select the field you want to edit.
2. Click **More options** **...** and select **Group and Replace** then select one of these options:
  - **Pronunciation**
  - **Spelling**

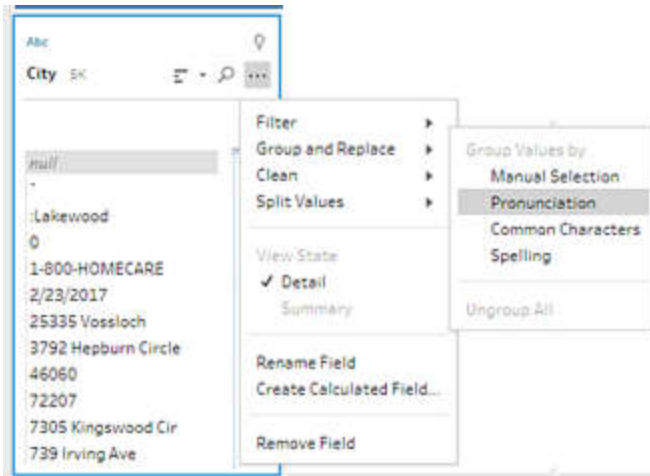
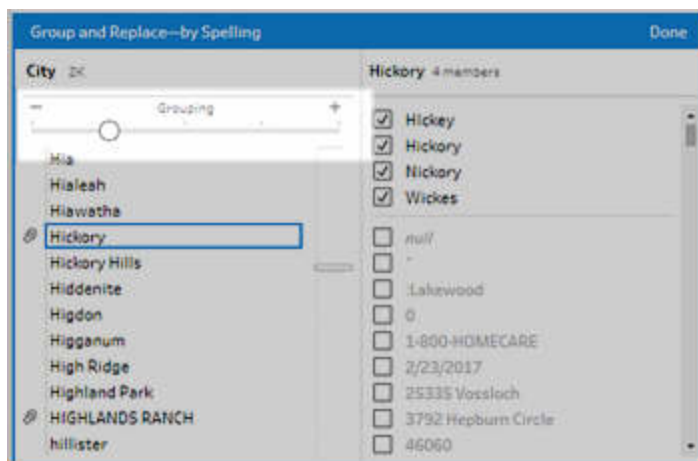


Tableau Prep Builder finds and groups values that match and replaces them with the value that occurs most frequently in the group.

3. In the left pane of the **Group and Replace** editor, drag the slider to one of the 5 threshold levels to change your results.



To set a stricter threshold, move the slider to the left. This results in fewer matches and creates less groups. To set a looser threshold, move the slider to the right. This results in more matches and creates more groups.

4. Click **Done** to save your changes.

## Copy steps, actions and fields

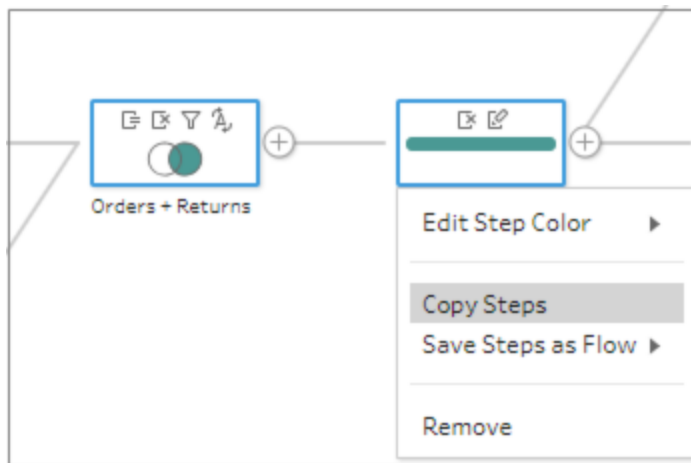
When cleaning your data you often perform the same cleaning operations or actions over and over throughout your flow. To help make cleaning and shaping your data more efficient, you


can copy and paste operations or actions throughout your flow, or even copy selected steps and save them so you can perform cleaning operations or actions once, then reuse it where you need it. You can even duplicate fields to experiment with different cleaning operations.

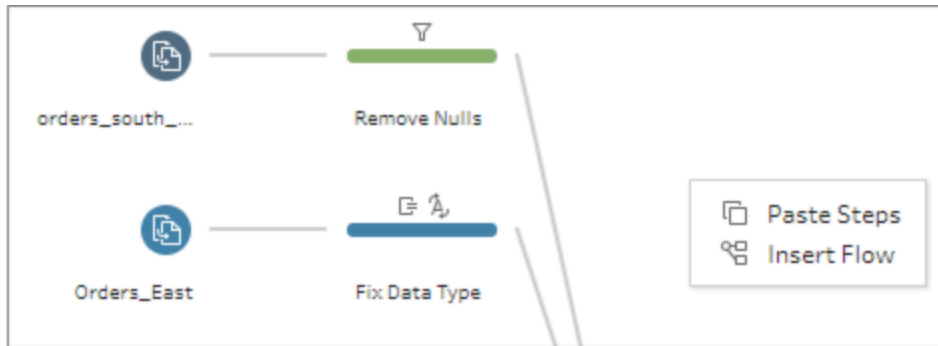
## Copy and paste steps

Copy one or more steps to use them in another area of the same flow. This option is not available for Input steps that include a wildcard union.

1. In the Flow pane, select one or more steps in the flow .
2. Right-click or Ctrl-click (MacOS) on the selected step, then select **Copy**.

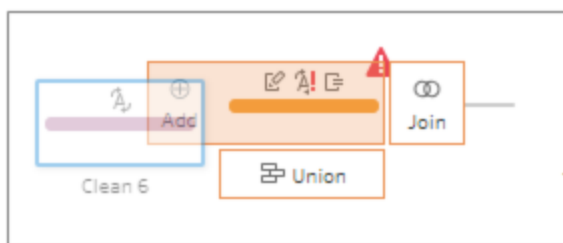


3. To paste the copied steps do one of the following:
  - **Version 2019.4.2 and later:** Hover over a step or flow line until the plus  icon appears, then click the icon and select **Insert Flow**
  - **All versions:** Right-click or Ctrl-click (MacOS) in any whitespace in the canvas and click **Paste** or **Paste Steps**.



4. Drag and drop the steps to where you want to place them in the flow. If adding steps to the end of a flow step, the steps are automatically added to the end of the step. If inserting steps between existing flow steps, move the steps where you want them in the flow and fix any errors.

You can remove flow lines or move steps around if needed. For example to connect a step to the copied steps, remove the existing flow line if there is one, then drag the existing step to the new step and drop on **Add**.



For more information about organizing your flow, see [Reorganize the layout of your flow \(version 2019.2.2 and later\)](#) on page 194.

## Copy and paste cleaning operations

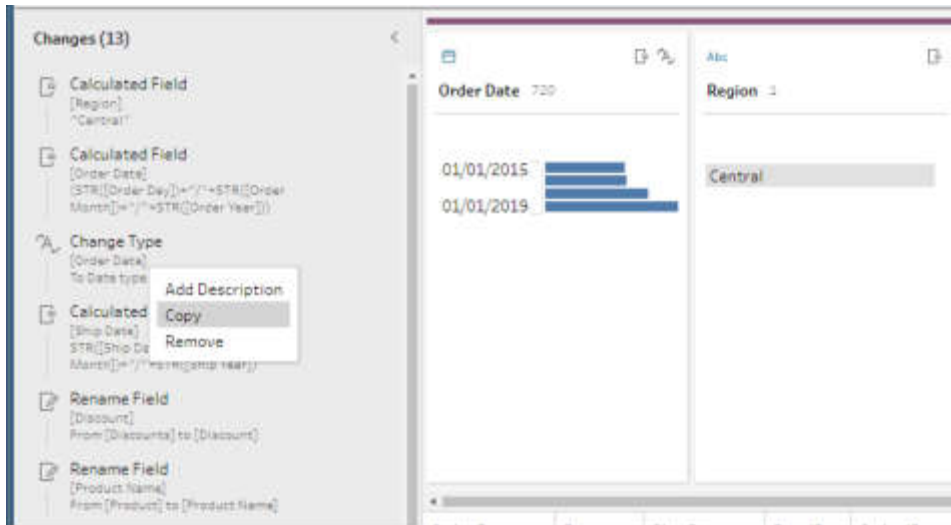
You can copy and paste cleaning operations in the same flow to reuse your actions using one of the following options:

- Copy an operation from the **Changes** pane in one step and paste it in the **Changes** pane for the same step or another step to apply that same operation in that step.
- Drag and drop an operation from the **Changes** pane and drop it to other fields in the Profile pane for that step to apply that operation to multiple fields. This option is not available for operations that impact multiple fields, such as calculated fields.

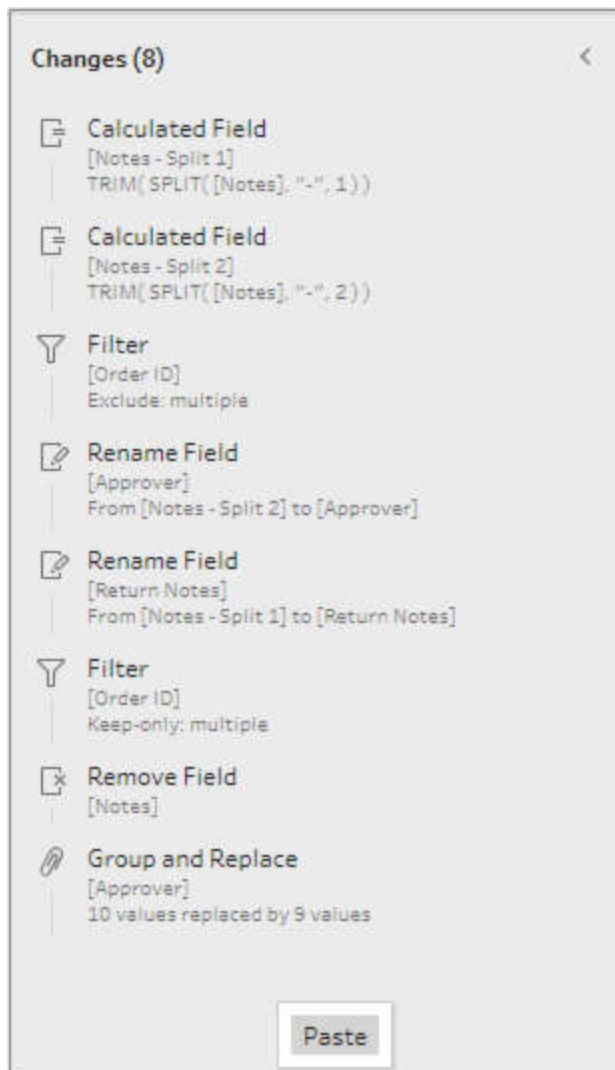
To copy and paste a change in a step to the same step or another step, do the following:

## Tableau Prep Help

1. In the **Changes** pane select the change you want to copy.
2. Right-click or Ctrl-click (MacOS) on the change item, then select **Copy** from the menu.



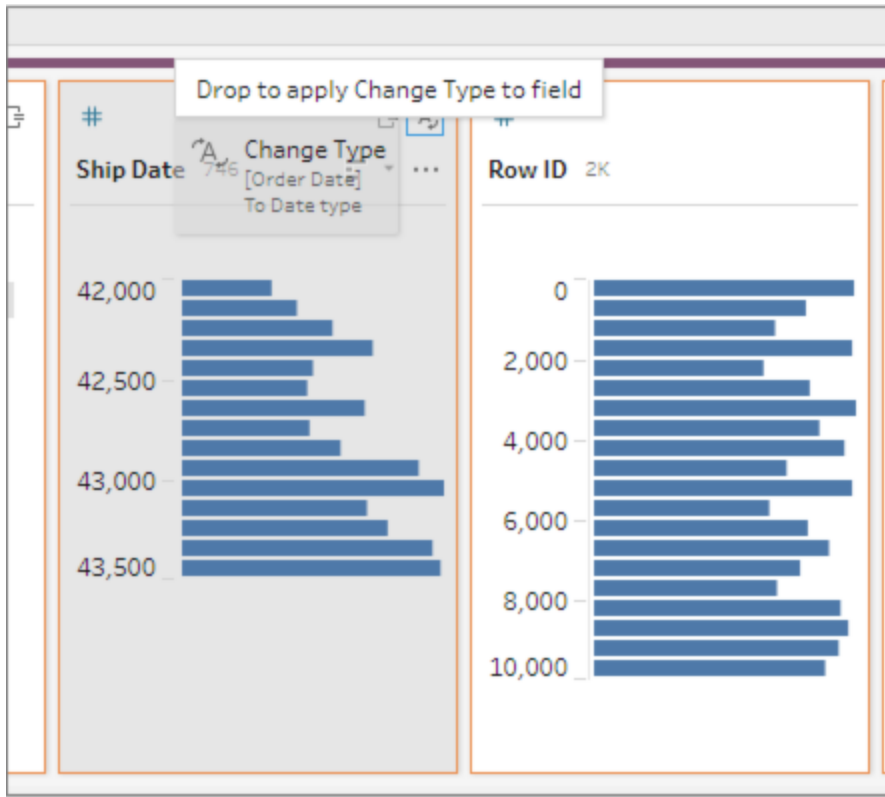
3. In the **Changes** pane where you want to paste the change right-click or Ctrl-click (MacOS) and select **Paste**. Select the change and click on **Edit** to make any adjustments as needed.



To drag and drop a change to other fields in the step do the following:

1. In the **Changes** pane select the change you want to copy.
2. Drag the change over the field where you want to apply it and drop it. Repeat this action

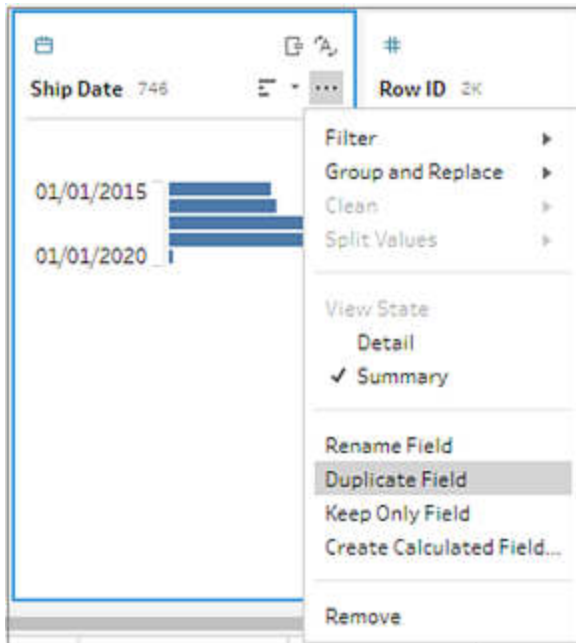
as needed.



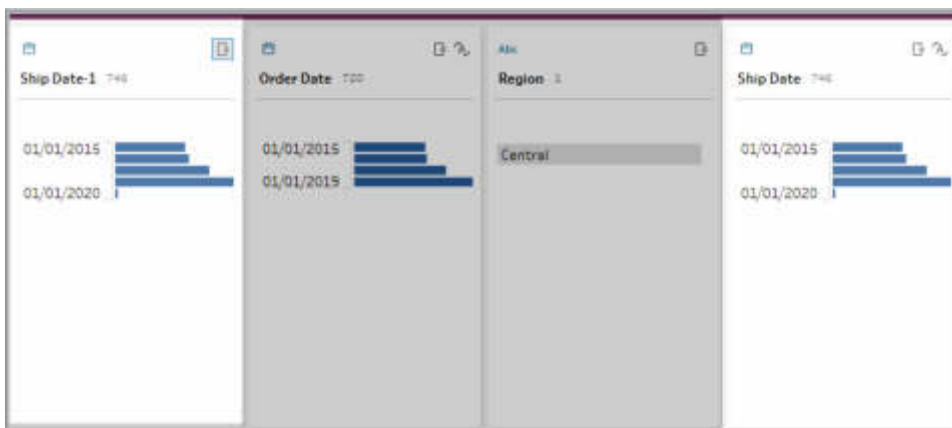
## Copy fields (version 2019.2.3 and later)

If you wanted to experiment with your cleaning operations on a field but don't want to change the original data, you can copy your fields .

1. In the profile pane, data grid, results pane, or list view, select the field you want to copy.
2. From the **More options** menu, select **Duplicate Field**.



A new field is created with the same name and a modifier. For example, "Ship Date -1".



## Create reusable flow steps (version 2019.3.2 and later)

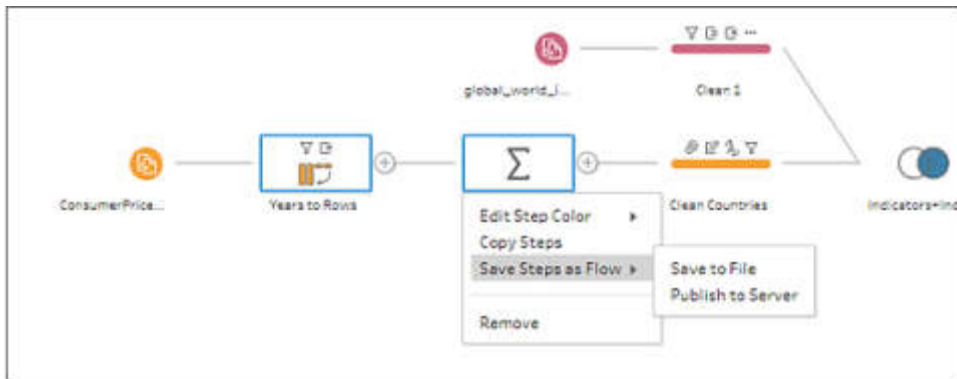
If you commonly perform the same actions over and over again with your data and you want to apply these same steps in other flows, you can select one or more flow steps and their associated actions or the entire flow and save it locally to a file on your computer or publish it to



Tableau Server or Tableau Online to share with others. When the flow steps are published to your server, a **Saved Steps** tag is automatically added so you can easily search and find them when adding them to your flows.

## Create reusable steps

1. Select one or more steps.
2. Right-click or Ctrl-click (MacOS) on a selected step and select **Save Steps as Flow**.



3. Select **Save to File** to save the flow locally or **Publish to Server** to publish the flow to Tableau Server or Tableau Online.
4. If you publish the flow to Tableau Server or Tableau Online, sign into your server if

needed, then complete the fields in the **Publish Flow** dialog then click **Publish**.

**Publish Flow to Tableau Server**

Project  
Default


Name  
Pivot + Aggregation

Description  
Pivot year columns to rows and then aggregate months to years.

Tags  
Saved Steps Add

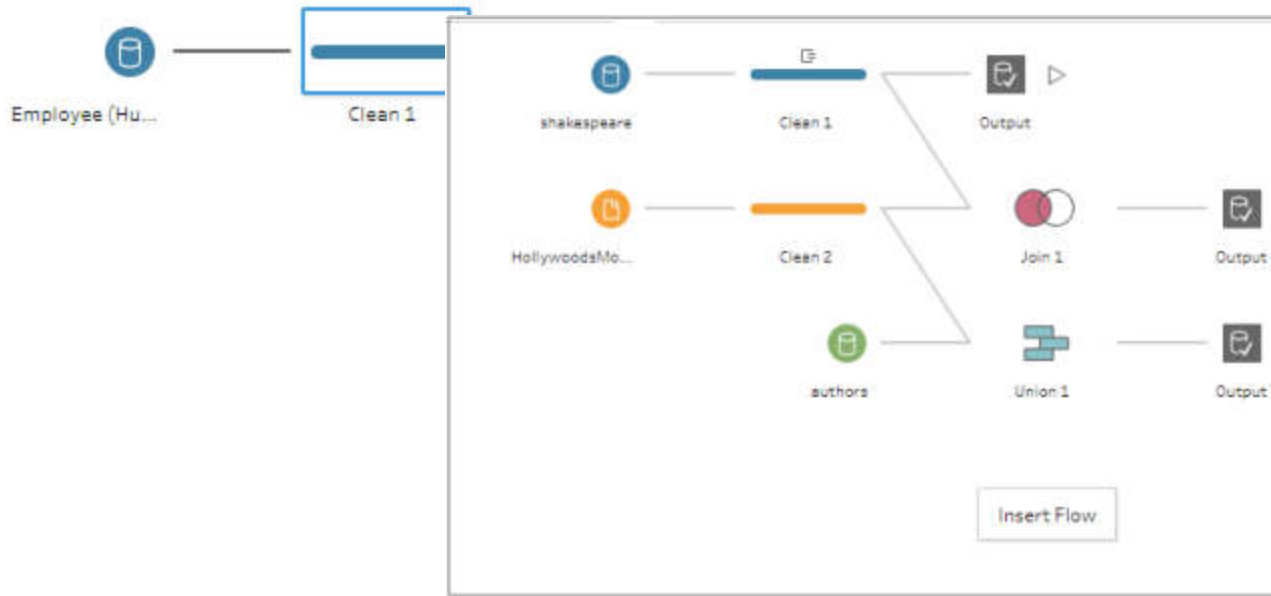
Cancel Publish

## Insert reusable steps in a flow

1. Open a flow.
2. In the flow pane, do one of the following:
  - **Version 2019.4.2 and later:** Hover over a step or flow line until the plus  icon appears, then click the icon and select **Insert Flow**
  - **Version 2019.3.2 and later:** In the white area of the canvas, right-click or Ctrl-click (MacOS) and click **Insert Flow** or click **Edit > Insert Flow** from the top menu.

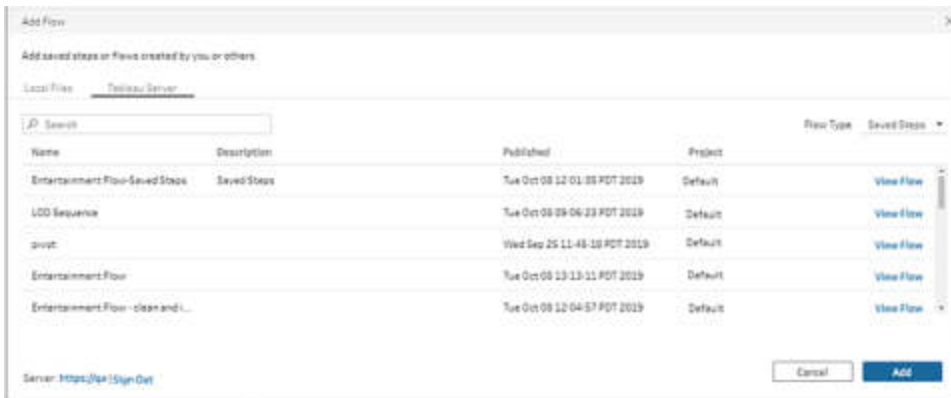
**Flow Step menu (version 2019.4.2 and later)**

**Canvas menu (version 2019.3.2 and later)**

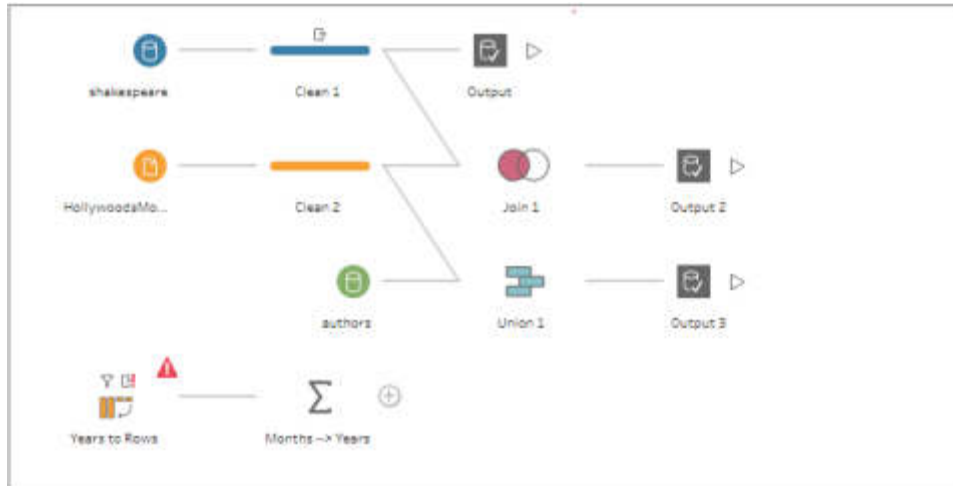


3. In the **Add Flow** dialog, select from flows saved to either your local file or your server, then click **Add**. The list of flows is automatically filtered to show flows tagged with **Saved Steps**. To insert other flows, change the **Flow Type** to **All Flows**.

In version 2019.4.2, you can click **View Flow** to open and view the published flow in the server you are signed into.



4. The flow is added to the flow pane. If adding a flow to the end of a flow step, the flow steps are automatically added to the end of the step. If inserting flow steps between existing flow steps, move the steps where you want them in the flow and fix any errors.



## Create level of detail and rank calculations

You can use calculated fields to create new data using data that already exists in your data source. Tableau Prep Builder supports many of the same calculation types as Tableau Desktop. For general information about creating calculations, see [Get Started with Calculations in Tableau](#).

**Note:** Some functions supported in Tableau Desktop may not yet be supported in Tableau Prep Builder. To view the available functions for Tableau Prep Builder, review the function list in the Calculation editor.

Starting in version 2020.1.3 Tableau Prep Builder you can use FIXED Level of Detail (LOD) and RANK and ROW\_NUMBER analytic functions to perform more complex calculations.

For example, add a FIXED LOD calculation to change the granularity of fields in your table, use the new ROW\_NUMBER () analytic function to quickly find duplicate rows, or use one of the new RANK () functions to find the top N or bottom N values for a selection of rows with similar data. If you want a more guided experience when building these types of expressions, you can use the new visual calculation editor.

### Calculate level of detail

When you need to calculate data at multiple levels of granularity in the same table, you can write a level of Detail (LOD) expression to do this. For example, if you wanted to find the total

sales for each region, you could write a calculation like `{FIXED [Region] : SUM ([Sales])}`.

Tableau Prep Builder supports the FIXED level of detail expression and uses the syntax `{FIXED [Field1],[Field2] : Aggregation([Field])}`.

LOD expressions have two parts to the equation that are separated by a colon.

- **FIXED [Field]** (required): This is the field or fields that you want to calculate the values for. For example if you wanted to find the total sales for customer and region, you would enter `FIXED [Customer ID], [Region] : .` If you don't select any fields, this is the equivalent to performing the aggregation defined on the right side of the colon and repeating that value for every row.
- **Aggregation ([Field])** (required): Select what you want to calculate and what level of aggregation you want. For example if you want to find the total sales, then enter `SUM ([Sales])`.


When using this feature in Tableau Prep Builder, the following requirements apply:

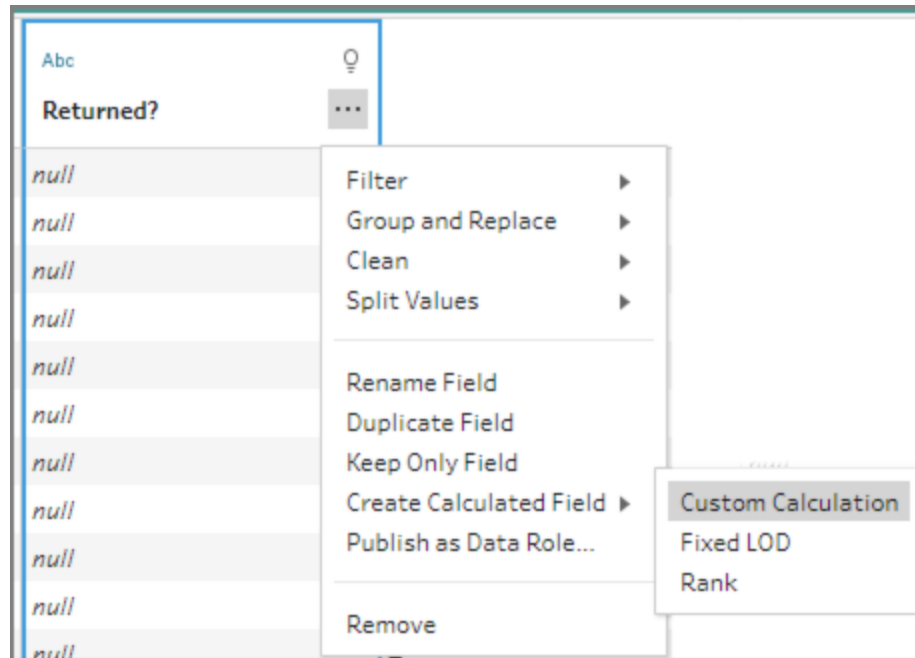
- INCLUDE and EXCLUDE LOD expressions aren't supported.
- Aggregation calculations are only supported inside an LOD expression. For example, `SUM([Sales])` would not be valid, but `{FIXED [Region] : SUM([Sales])}` is valid.
- Nesting expressions inside an LOD expression isn't supported. For example, `{ FIXED [Region] : AVG( [Sales] ) / SUM( [Profit] )}` isn't supported.
- Combining an LOD expression with another expression isn't supported. For example `[Sales]/{ FIXED [Country / Region]:SUM([Sales])}` isn't supported.

## Create Level of Detail (LOD) calculations

To create a level of detail calculation, you can use the Calculation editor to write the calculation yourself or if you want a more guided experience, you can use the Visual Calculation editor where you select your fields and Tableau Prep Builder writes the calculation for you.

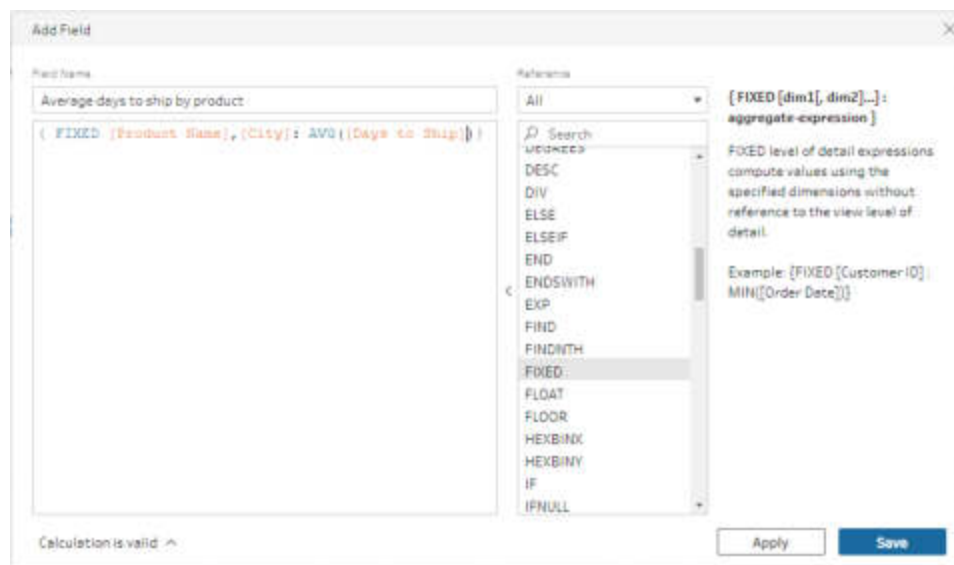
### Calculation editor

1. In the Profile pane toolbar click **Create Calculated Field**, or in a profile card or data grid, click the **More options**  menu and select **Create Calculated Field > Custom Calculation**.



2. In the Calculation editor, enter a name for your calculation and enter the expression.

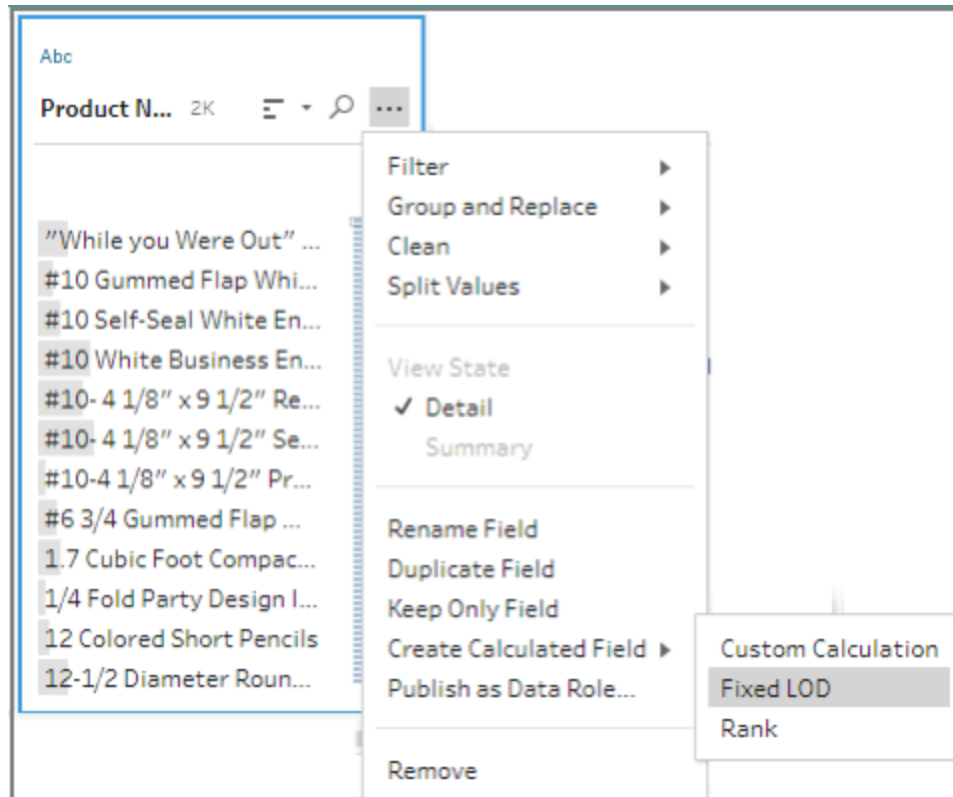
For example, to find the average days to ship products by city, create a calculation like the one shown below.



## Visual Calculation editor

Select fields from a list and Tableau Prep Builder builds the calculation for you as you make your selections. A preview of the results is shown in the left pane so you can see the results of your selections as you go.

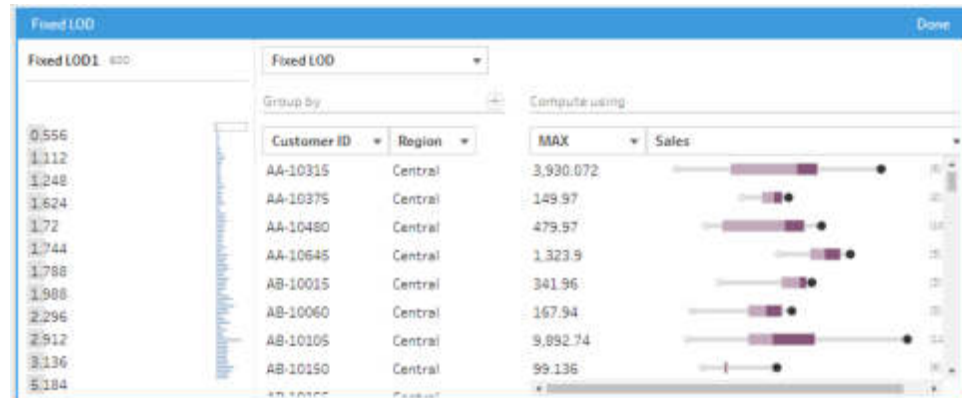
1. In a profile card or results pane, click the **More options** ... menu and select **Create Calculated Field >Fixed LOD**.



2. In the Visual Calculation editor, do the following:
  - In the **Group by** section, select the fields that you want to calculate the values for. The field where you selected the **Create Calculated Field >Fixed LOD** menu option is added by default. Click the plus  $\oplus$  icon to add any additional fields to your calculation. This populates the left side of the equation, `{FIXED [Field1],[Field2] :.`
  - In the **Compute using** section, select the field that you want to use to

calculate your new values. Then select your aggregation. This populates the right side of the equation, `Aggregation ([Field])`.

A graphic below the field shows the distribution of values and a total count for each value combination. Depending on the type of data, this can be a box plot, range of values, or the actual values.



**Note:** Available aggregation values vary by the data type assigned to the field.

- To remove a field, right-click or Cmd-click (MacOS) in the drop-down box for the fields in the **Group by** section and select **Remove Field**.
  - In the left pane, double-click in the field header and enter a name for your calculation.
3. Click **Done** to add your new calculated field. In the **Changes** pane, you can see the calculation that Tableau Prep Builder generated. Click **Edit** to open the visual calculation editor to make any changes.

```
Fixed LOD
[Fixed LOD1]
[[ FIXED [Product Name], [City] : AVG ([Days to Ship]) ]]
```

## Calculate rank or row number

Analytic functions, sometimes referred to as window calculations, enable you to perform calculations across the entire table, or a selection of rows (partition) in your data set. For



example, when applying a rank to a selection of rows, you would use the following calculation syntax:

```
{PARTITION [field]: {ORDERBY [field]: RANK() }
```

- **PARTITION** (optional): Designate the rows you want to perform the calculation on. You can specify more than one field, but if you want to use the entire table, omit this part of the function and Tableau Prep Builder treats all the rows as the partition. For example `{ORDERBY [Sales] : RANK() }`.
- **ORDERBY** (required): Specify one or more fields that you want to use to generate the sequence for the rank.
- **Rank ()** (required): Specify the rank type or `ROW_NUMBER ()` you want to calculate. Tableau Prep Builder supports `RANK()`, `RANK_DENSE()`, `RANK_MODIFIED()`, `RANK_PERCENTILE()`, and `ROW_NUMBER()` functions.
- **DESC** or **ASC** (optional): Represents descending (DESC) or ascending (ASC) order. By default, rank is sorted in descending order, so you don't need to specify this in the expression. If you want to change the sort order, add **ASC** to the expression. You can also include both options in the function. For example if you wanted to rank a selection of rows, but wanted to sort the rows in ascending order, then apply the rank in descending order, you would include these two options in the expression. For example: `{PARTITION [Country], [State]: {ORDERBY [Sales] ASC, [CUSTOMER] DESC: RANK() }}`

When using this feature in Tableau Prep Builder, the following requirements apply:

- Nesting expressions inside a `RANK ()` function isn't supported. For example, `[Sales] / {PARTITION [Country]: {ORDERBY [Sales]: RANK() }} / SUM([Profit] )}` isn't supported.
- Combining a `RANK ()` function with another expression isn't supported. For example `[Sales] / {PARTITION [Country]: {ORDERBY [Sales]: RANK() }}` isn't supported.

### Supported analytic functions

Function	Description	Example
----------	-------------	---------

RANK ()	<p>Assigns a whole number rank starting with 1, in ascending or descending order to each row. If rows have the same value, they share the rank that is assigned to the first instance of the value. The number of rows with the same rank is added when calculating the rank for the next row, so you may not get consecutive rank values.</p>	<table border="1"> <thead> <tr> <th>Sales Person</th> <th>Region</th> <th>Commission</th> <th>Rank</th> </tr> </thead> <tbody> <tr> <td>Jorge Ramire</td> <td>Central</td> <td>7000</td> <td>1</td> </tr> <tr> <td>Lisa Martin</td> <td>Central</td> <td>15000</td> <td>2</td> </tr> <tr> <td>Sarah Scott</td> <td>Central</td> <td>15000</td> <td>2</td> </tr> <tr> <td>Barbara Davis</td> <td>East</td> <td>13000</td> <td>4</td> </tr> <tr> <td>Charles Lee</td> <td>East</td> <td>9000</td> <td>5</td> </tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramire	Central	7000	1	Lisa Martin	Central	15000	2	Sarah Scott	Central	15000	2	Barbara Davis	East	13000	4	Charles Lee	East	9000	5
Sales Person	Region	Commission	Rank																							
Jorge Ramire	Central	7000	1																							
Lisa Martin	Central	15000	2																							
Sarah Scott	Central	15000	2																							
Barbara Davis	East	13000	4																							
Charles Lee	East	9000	5																							
RANK_DENSE()	<p>Assigns a whole number rank starting with 1 in ascending or descending order to each row. If rows have the same value, they share the rank that is assigned to the first instance of the value, but no rank values are skipped so you will see consecutive rank values.</p>	<table border="1"> <thead> <tr> <th>Sales Person</th> <th>Region</th> <th>Commission</th> <th>Rank</th> </tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td> <td>Central</td> <td>7000</td> <td>1</td> </tr> <tr> <td>Lisa Martin</td> <td>Central</td> <td>15000</td> <td>2</td> </tr> <tr> <td>Sarah Scott</td> <td>Central</td> <td>15000</td> <td>2</td> </tr> <tr> <td>Barbara Davis</td> <td>East</td> <td>13000</td> <td>3</td> </tr> <tr> <td>Charles Lee</td> <td>East</td> <td>9000</td> <td>4</td> </tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	2	Sarah Scott	Central	15000	2	Barbara Davis	East	13000	3	Charles Lee	East	9000	4
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	2																							
Sarah Scott	Central	15000	2																							
Barbara Davis	East	13000	3																							
Charles Lee	East	9000	4																							

<p><b>RANK_MODIFIED()</b></p>	<p>Assigns a whole number rank starting with 1, in ascending or descending order to each row. If rows have the same value, they share the rank that is assigned to the last instance of the value. No rank values are skipped. Rank_Modified is calculated as <math>\text{Rank} + (\text{Rank} + \text{Number of duplicate rows} - 1)</math>.</p>	<table border="1"> <thead> <tr> <th>Sales Person</th> <th>Region</th> <th>Commission</th> <th>Rank</th> </tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td> <td>Central</td> <td>7000</td> <td>1</td> </tr> <tr> <td>Lisa Martin</td> <td>Central</td> <td>15000</td> <td>3</td> </tr> <tr> <td>Sarah Scott</td> <td>Central</td> <td>15000</td> <td>3</td> </tr> <tr> <td>Barbara Davis</td> <td>East</td> <td>13000</td> <td>4</td> </tr> <tr> <td>Charles Lee</td> <td>East</td> <td>9000</td> <td>5</td> </tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	3	Sarah Scott	Central	15000	3	Barbara Davis	East	13000	4	Charles Lee	East	9000	5
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	3																							
Sarah Scott	Central	15000	3																							
Barbara Davis	East	13000	4																							
Charles Lee	East	9000	5																							
<p><b>RANK_PERCENTILE()</b></p>	<p>Assigns a percentile rank from 0 to 1 in ascending or descending order to each row. RANK_PERCENTILE is calculated as <math>(\text{Rank} - 1) / (\text{Total rows} - 1)</math>.</p>	<table border="1"> <thead> <tr> <th>Sales Person</th> <th>Region</th> <th>Commission</th> <th>Rank</th> </tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td> <td>Central</td> <td>7000</td> <td>0.0</td> </tr> <tr> <td>Charles Lee</td> <td>East</td> <td>9000</td> <td>0.25</td> </tr> <tr> <td>Barbara Davis</td> <td>East</td> <td>13000</td> <td>0.5</td> </tr> <tr> <td>Lisa Martin</td> <td>Central</td> <td>15000</td> <td>0.75</td> </tr> <tr> <td>Sarah Scott</td> <td>Central</td> <td>15000</td> <td>0.75</td> </tr> </tbody> </table>	Sales Person	Region	Commission	Rank	Jorge Ramirez	Central	7000	0.0	Charles Lee	East	9000	0.25	Barbara Davis	East	13000	0.5	Lisa Martin	Central	15000	0.75	Sarah Scott	Central	15000	0.75
Sales Person	Region	Commission	Rank																							
Jorge Ramirez	Central	7000	0.0																							
Charles Lee	East	9000	0.25																							
Barbara Davis	East	13000	0.5																							
Lisa Martin	Central	15000	0.75																							
Sarah Scott	Central	15000	0.75																							
<p><b>ROW_NUMBER()</b></p>	<p>Assigns a sequential row ID to each unique row. If you have duplicate rows and use this calculation, your results might change each time you run the flow if the order of rows changes.</p>	<table border="1"> <thead> <tr> <th>Sales Person</th> <th>Region</th> <th>Commission</th> <th>ROW ID</th> </tr> </thead> <tbody> <tr> <td>Jorge Ramirez</td> <td>Central</td> <td>7000</td> <td>1</td> </tr> <tr> <td>Lisa Martin</td> <td>Central</td> <td>15000</td> <td>2</td> </tr> <tr> <td>Sarah Scott</td> <td>Central</td> <td>15000</td> <td>3</td> </tr> <tr> <td>Barbara Davis</td> <td>East</td> <td>13000</td> <td>4</td> </tr> <tr> <td>Charles Lee</td> <td>East</td> <td>9000</td> <td>5</td> </tr> </tbody> </table>	Sales Person	Region	Commission	ROW ID	Jorge Ramirez	Central	7000	1	Lisa Martin	Central	15000	2	Sarah Scott	Central	15000	3	Barbara Davis	East	13000	4	Charles Lee	East	9000	5
Sales Person	Region	Commission	ROW ID																							
Jorge Ramirez	Central	7000	1																							
Lisa Martin	Central	15000	2																							
Sarah Scott	Central	15000	3																							
Barbara Davis	East	13000	4																							
Charles Lee	East	9000	5																							

## Create Rank or Row Number calculations

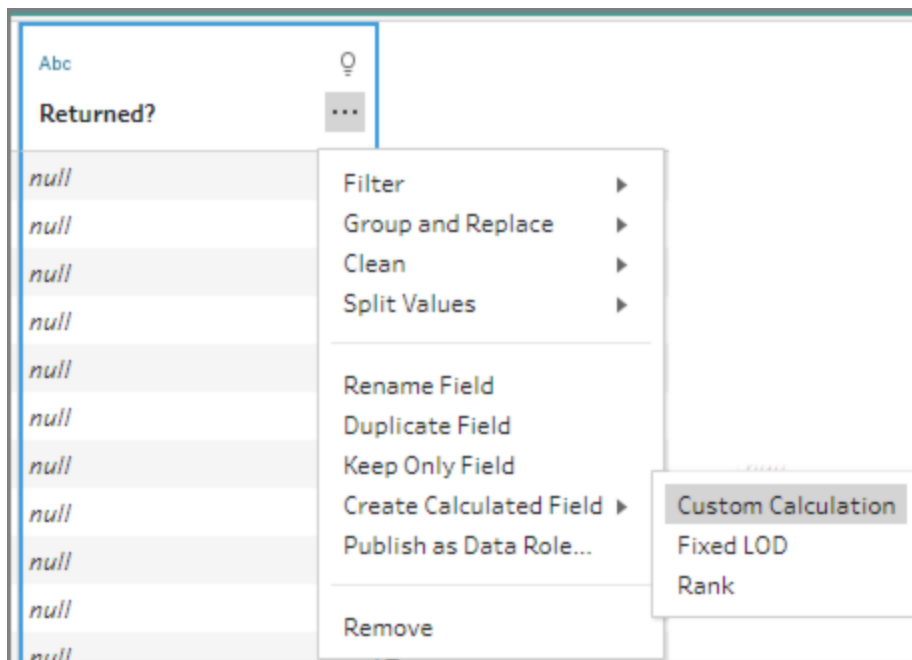
To create a Rank or Row\_Number calculations, you can use the Calculation editor to write the calculation yourself or if you want a more guided experience, you can use the Visual Calculation editor where you select your fields and Tableau Prep Builder writes the calculation for you.

**Note:** ROW\_NUMBER () calculations aren't available in the visual calculation editor.

### Calculation editor

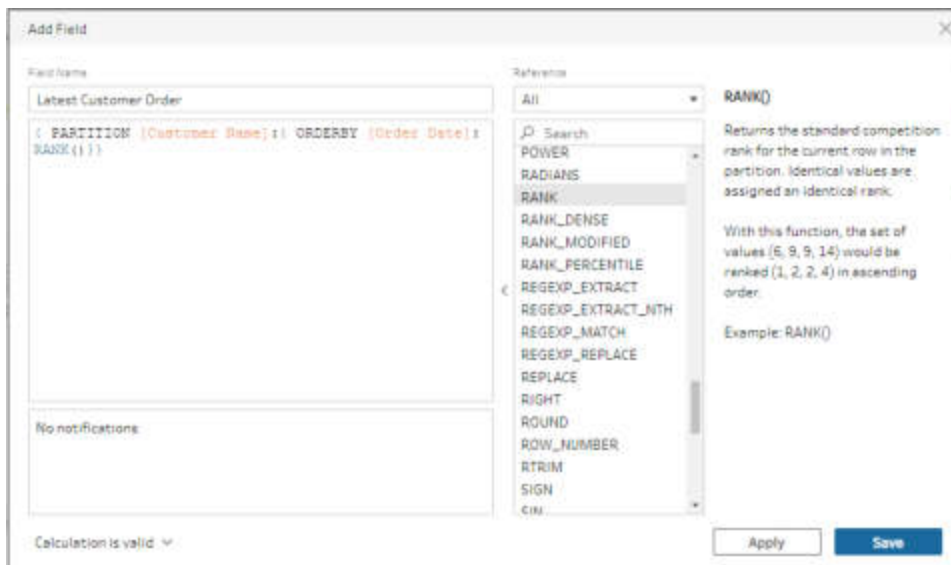
Use the Calculation editor to create any of the supported RANK () or ROW\_NUMBER() calculations. The list of supported analytic calculations is shown in the Calculation editor in the **Reference** drop-down under **Analytic**.

1. In the Profile pane toolbar click **Create Calculated Field**, or in a profile card or data grid, click the **More options** ... menu and select **Create Calculated Field > Custom Calculation**.



2. In the Calculation editor, enter a name for your calculation and enter the expression.

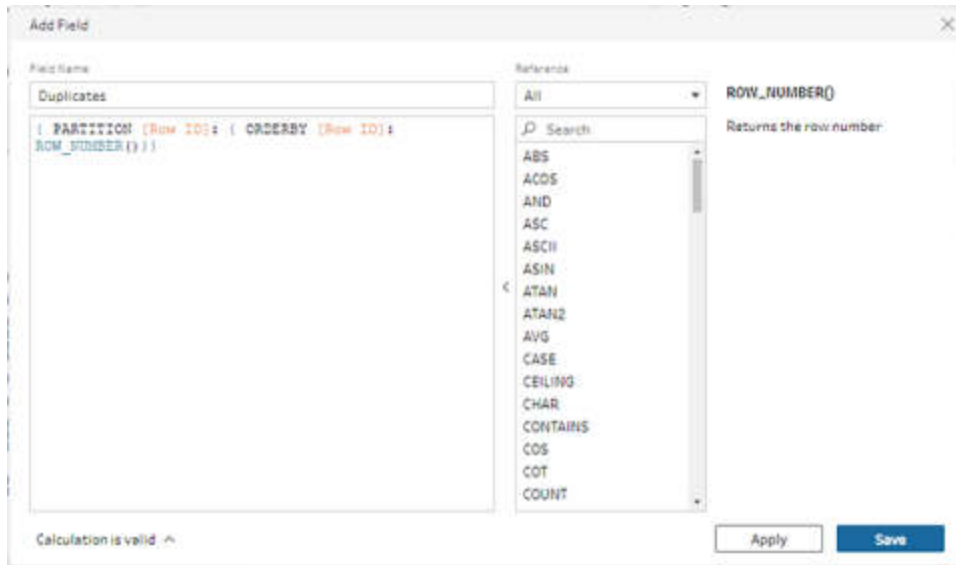
For example to find the latest customer order, create a calculation like the one shown below, then keep only the customer order rows that are ranked with the number **1**.



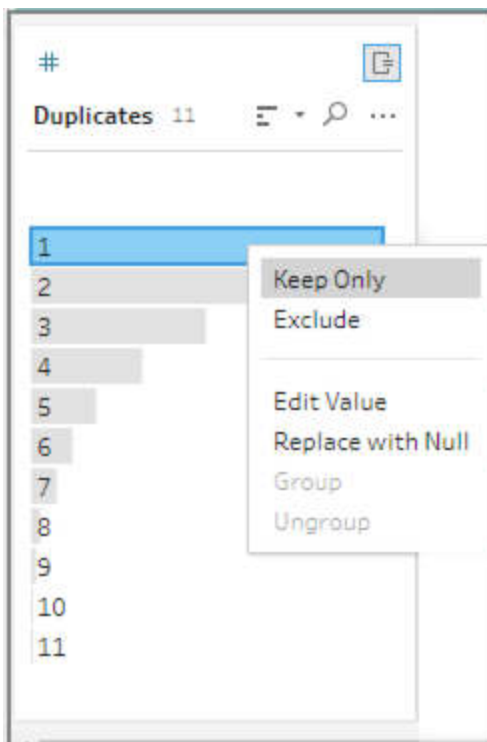
**Example:** Use ROW\_NUMBER to find and remove duplicate values

This example uses the Superstore sample data set to find and remove exact duplicate values for the field **Row ID** using the ROW\_NUMBER function.

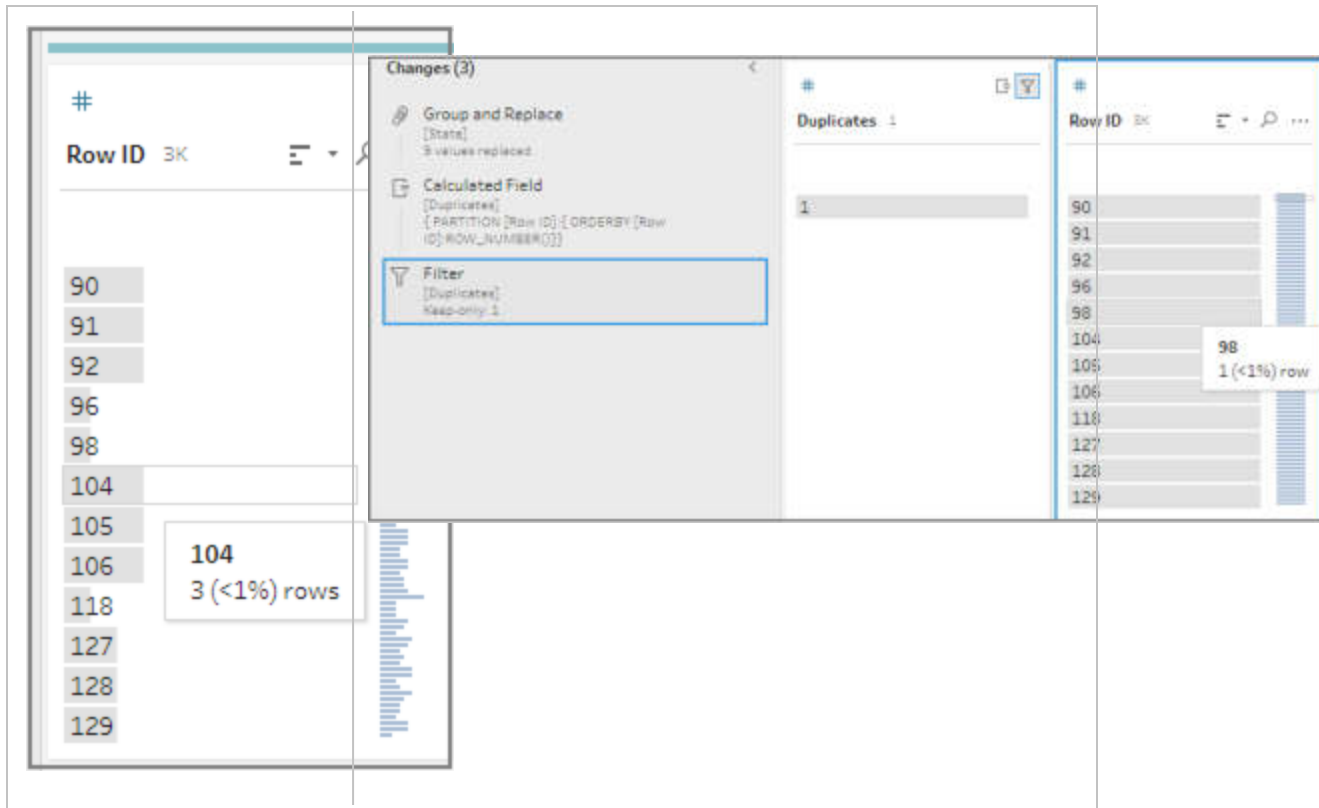
1. Open the **Sample Superstore** flow.
2. In the Flow pane, for the Input step **Orders West**, click on the Clean step **Rename States**.
3. In the toolbar, click **Create Calculated Field**.
4. In the Calculation editor, use the ROW\_NUMBER function to add a row number to the field **Row ID** using the expression `{ PARTITION [Row ID] : { ORDERBY [Row ID] : ROW_NUMBER() } }` and click **Save**.



5. In the new calculated field, right-click or Cmd-click (MacOS) on the field value **1**, then select **Keep Only** from the menu.



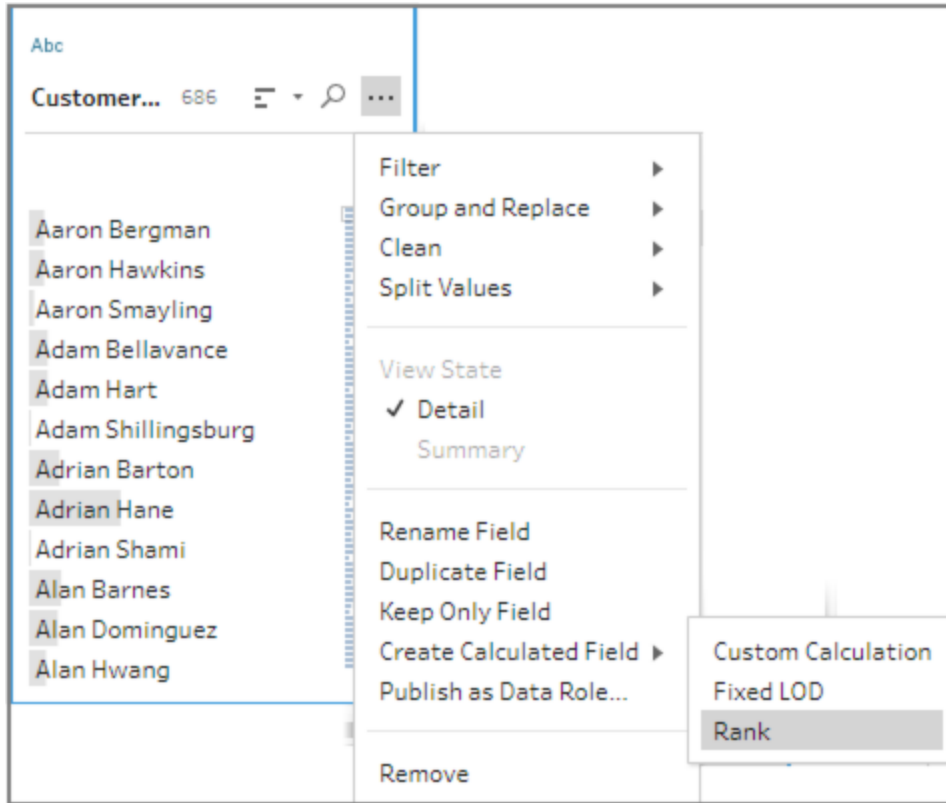
Before	After
--------	-------



### Visual Calculation editor

Just like when creating a level of detail calculation, you can use the visual calculation editor to build a rank calculation. Select the fields you want to include in the calculation, then select the fields you want to use to rank the rows and the type of rank you want to calculate. A preview of the results is shown in the left pane so you can see the results of your selections as you go.

1. In a profile card or results pane, click the **More options** ... menu and select **Create Calculated Field > Rank**.




2. In the Visual calculation editor, do the following:

- In the **Group by** section, select the fields whose rows you want to compute values for. This creates the Partition part of the calculation.

Click the plus (+) icon to add any additional fields to your calculation. If you want to include all rows or remove a selected field, right-click or Cmd-click (MacOS) in the drop-down box for the fields in the **Group by** section and select **Remove Field**.

- In the **Order by** section, select the fields that you want to use to rank your new values. The field where you selected the **Create Calculated Field > Rank** menu option is added by default.

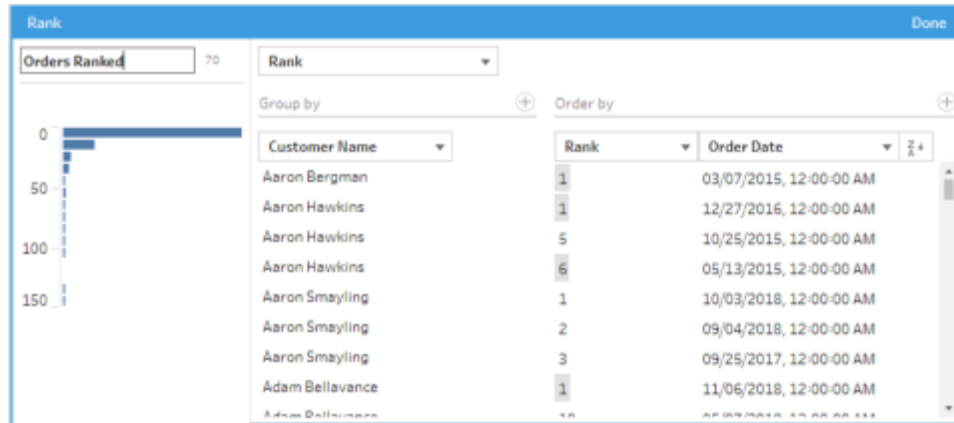
Click the plus (+) icon to add any additional fields to your calculation, then select

your **Rank** type. Click the sort icon  to change the rank order from descending (DESC) to ascending (ASC).



**Note:** Rank values vary by the data type assigned to the field.

- In the left pane, double-click in the field header and enter a name for your calculation.



3. Click **Done** to add your new calculated field. In the **Changes** pane, you can see the calculation that Tableau Prep Builder generated. Click **Edit** to open the visual calculation editor to make any changes.

```

Rank
[Rank1]
{{PARTITION [Customer Name]: {ORDERBY
[Order Date] DESC: RANK()}}}
    
```

## Pivot Your Data

Sometimes analyzing data from a spreadsheet or crosstab format can be difficult in Tableau. Tableau prefers data to be "tall" instead of "wide", which means that you often have to pivot your data from columns to rows so that Tableau can evaluate it properly.

However you may also have scenarios where your data tables are tall and narrow and are too normalized to properly analyze. For example a sales department that tracks advertising spend in two columns, one called **Advertising** that contains rows for **radio**, **television** and **print** and one column for total spent. In this type of scenario, to analyze this data as separate measures you would need to pivot that row data to columns.

But what about pivoting larger data sets or data that changes frequently over time? You can use a wildcard pattern match to search for fields that match the pattern and automatically pivot the data.

Use one of the following options when pivoting your data:


- Pivot columns to rows
- Use wildcard search to instantly pivot fields based on a pattern match (Tableau Prep Builder version 2019.1.1 and later).
- Pivot rows to columns (Tableau Prep Builder version 2019.1.1 and later).

No matter how you pivot your fields, you can interact directly with the results and perform any additional cleaning operations to get your data looking just the way you want it. You can also use Tableau Prep Builder's smart default naming feature to automatically rename your pivoted fields and values.

**Note:** Pivoting columns to rows on multiple groups of fields is not supported in Tableau Prep Builder 2018.1.1. Editing a pivot that includes pivoted columns on multiple field groups in Tableau Prep Builder version 2018.1.1 can result in errors or unexpected results

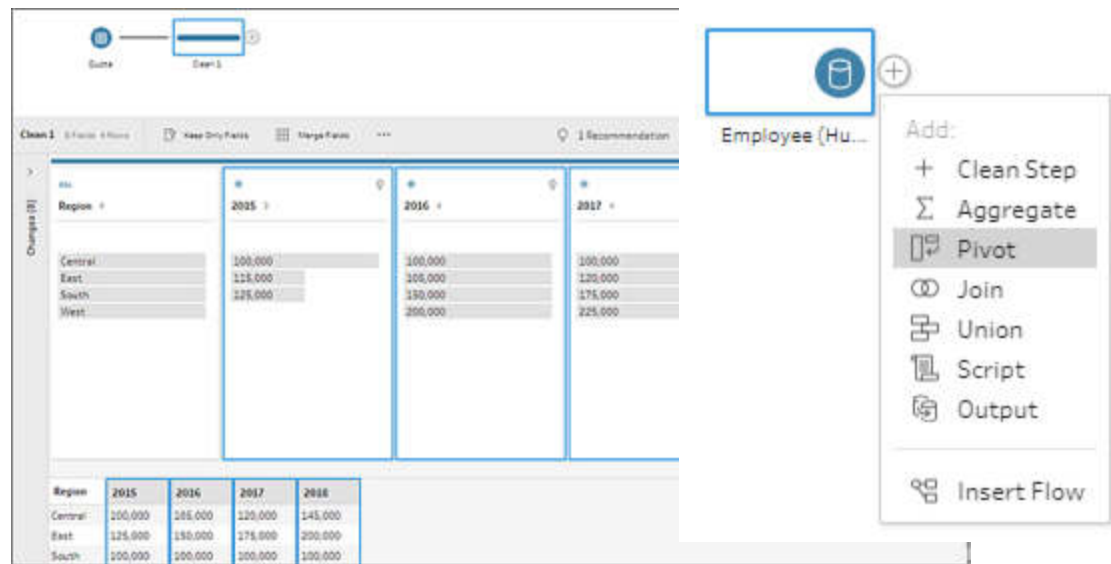
## Pivot columns to rows

Use this pivot option to go from wide data to tall data. Pivot from columns to rows on one or more groups of fields. Select the fields that you want to work with and pivot the data from columns to rows.

1. Connect to your data source.
2. Drag the table that you want to pivot to the **Flow** pane.
3. Do one of the following:
  - **Version 2019.4.2 and later:** In the Profile pane, select the fields that you want to pivot, then right-click or Ctrl-click (MacOS) and select **Pivot Columns to Rows** from the menu. If using this option, skip to step 7.
  - **All versions:** Click the plus  icon, and select **Add Pivot** from the context menu.

## Select Fields (version 2019.4.2 and later)

## Flow Step Menu (all versions)



- (Optional) In the **Fields** pane, enter a value in the **Search** field to search the field list for fields to pivot.
- (Optional) Select the **Automatically rename pivoted fields and values** check box to enable Tableau Prep to rename the new pivoted fields using common values in the data. If no common values are found, the default name is used.
- Select one or more fields from the left pane, and drag them to the **Pivot1 Values** column in the **Pivoted Fields** pane.
- (Optional) In the **Pivoted Fields** pane, click the plus  $\oplus$  icon to add more columns to pivot on, then repeat the previous step to select more fields to pivot. Your results appear immediately in both the **Pivot Results** pane and the data grid.

**Note:** You must select the same number of fields that you selected in Step 5. For example if you selected 3 fields to initially pivot on, then each subsequent column that you pivot on must also contain 3 fields.

- If you didn't enable the default naming option or if Tableau Prep couldn't automatically detect a name, edit the names of the fields. You can also edit the names of the original

fields in this pane to best describe the data.

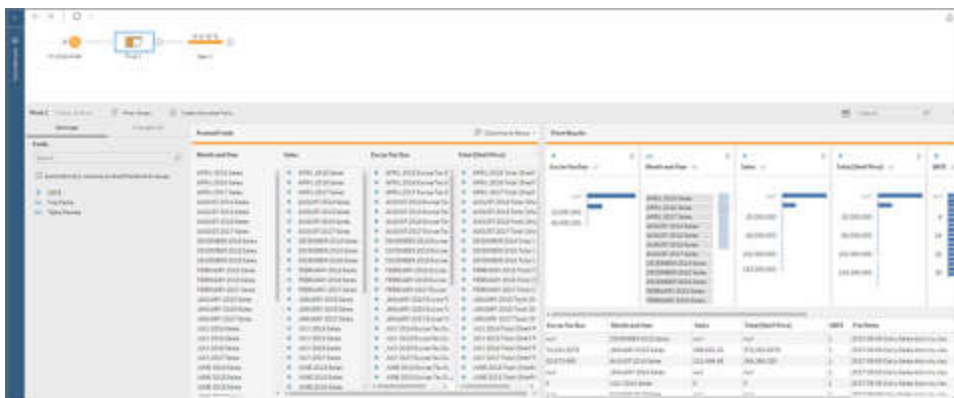
9. (Optional) Rename the new Pivot step to keep track of your changes. For example "Pivot months".
10. To refresh your pivot data when data changes, run your flow. If new fields are added to your data source that need to be added to the pivot, manually add them to the pivot.

**Example:** Pivoting on multiple fields

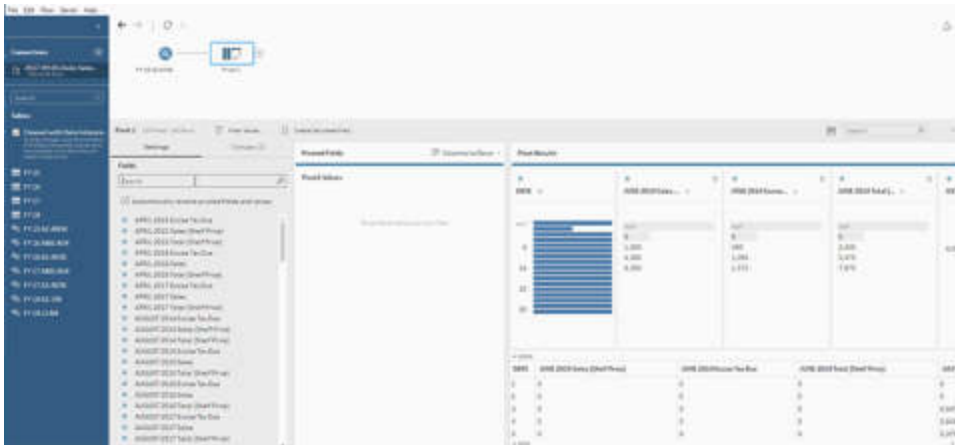
This example shows a spreadsheet for pharmaceutical sales, taxes and totals by month and year.

DATE	DECEMBER 2014			JANUARY 2015			FEBRUARY 2015		
	Sales (Shell Price)	Excise Tax Due	Total (Shell Price)	Sales (Shell Price)	Excise Tax Due	Total (Shell Price)	Sales (Shell Price)	Excise Tax Due	Total (Shell Price)
1	\$ 448,111	\$ 112,628	\$ 560,739	\$ 295,843	\$ 74,270	\$ 370,113	\$ 212,919	\$ 53,230	\$ 266,149
2	\$ 425,472	\$ 106,368	\$ 531,840	\$ 754,061	\$ 188,515	\$ 942,577	\$ 449,897	\$ 112,474	\$ 562,371
3	\$ 435,545	\$ 108,881	\$ 544,426	\$ 482,497	\$ 120,624	\$ 603,121	\$ 627,711	\$ 156,328	\$ 784,039
4	\$ 434,765	\$ 108,691	\$ 543,456	\$ 332,228	\$ 83,657	\$ 415,885	\$ 608,263	\$ 152,066	\$ 760,329
5	\$ 695,425	\$ 173,856	\$ 869,282	\$ 981,529	\$ 250,392	\$ 1,231,921	\$ 789,233	\$ 197,368	\$ 986,601
6	\$ 436,720	\$ 109,180	\$ 545,900	\$ 527,374	\$ 131,843	\$ 659,217	\$ 867,501	\$ 216,875	\$ 1,084,377
7	\$ 238,481	\$ 59,620	\$ 298,101	\$ 366,162	\$ 90,826	\$ 456,988	\$ 354,499	\$ 88,815	\$ 443,314
8	\$ 421,422	\$ 105,356	\$ 526,778	\$ 529,374	\$ 134,993	\$ 664,367	\$ 448,848	\$ 112,211	\$ 561,059
9	\$ 343,816	\$ 85,954	\$ 429,770	\$ 883,488	\$ 220,872	\$ 1,104,360	\$ 708,266	\$ 177,067	\$ 885,333
10	\$ 674,271	\$ 168,567	\$ 842,838	\$ 482,152	\$ 119,588	\$ 601,740	\$ 719,437	\$ 179,309	\$ 898,746
11	\$ 236,342	\$ 59,135	\$ 295,477	\$ 288,685	\$ 72,151	\$ 360,836	\$ 1,154,919	\$ 288,730	\$ 1,443,649
12	\$ 726,270	\$ 181,567	\$ 907,837	\$ 674,171	\$ 168,520	\$ 842,691	\$ 1,919,936	\$ 484,984	\$ 2,404,920
13	\$ 477,806	\$ 119,302	\$ 597,108	\$ 526,481	\$ 131,613	\$ 658,094	\$ 951,842	\$ 237,811	\$ 1,189,653
14	\$ 245,898	\$ 61,475	\$ 307,373	\$ 573,842	\$ 143,461	\$ 717,303	\$ 798,392	\$ 199,598	\$ 997,990
15	\$ 494,254	\$ 123,564	\$ 617,818	\$ 638,952	\$ 164,738	\$ 803,690	\$ 453,891	\$ 113,273	\$ 567,164

By pivoting the data you can create rows for each month and year and individual columns for sales, taxes and totals so that Tableau can more easily interpret this data for analysis.



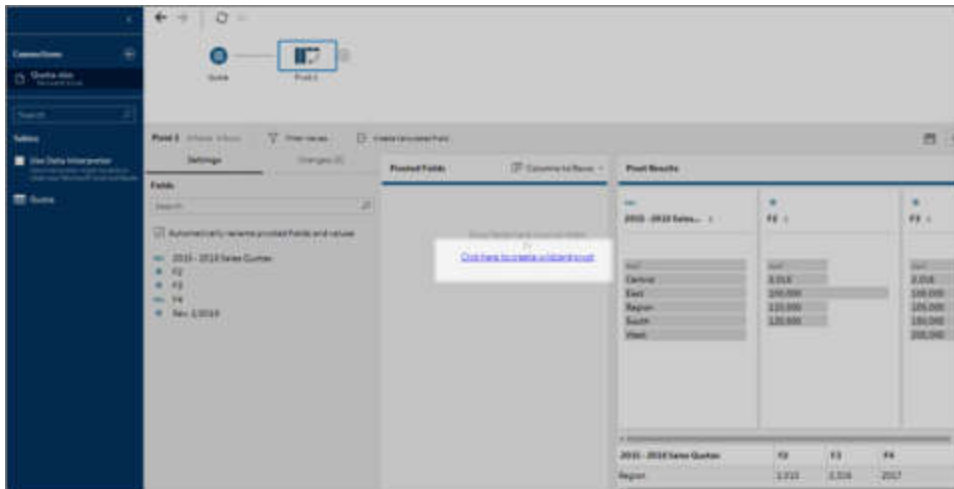
Watch "pivot on multiple field" in action.



## Use wildcard search to pivot (version 2019.1.1 and later)

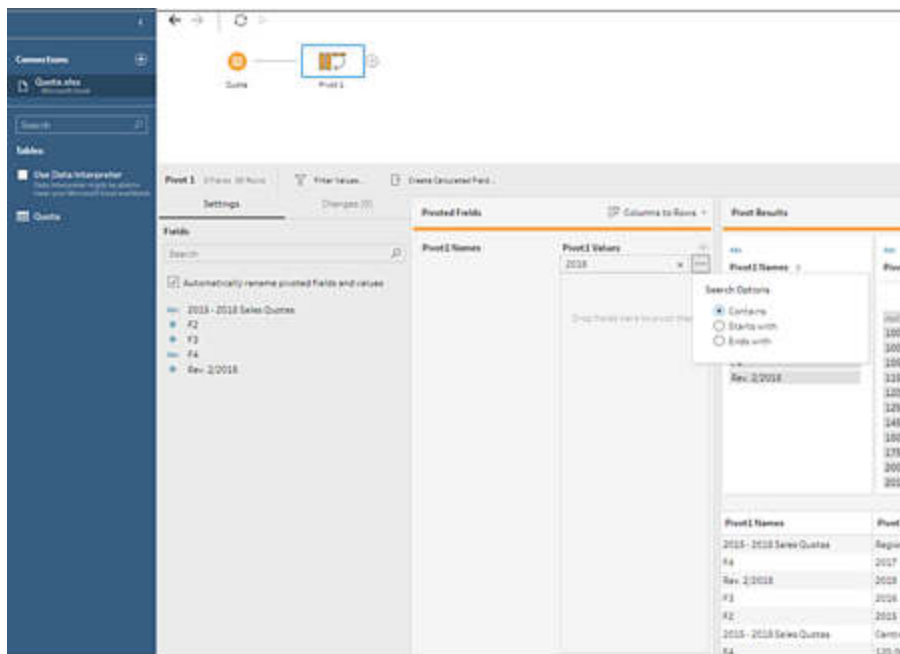
If you work with large data sets or if your data frequently changes over time, you can use a wildcard search when pivoting columns to rows to instantly pivot your data based on a wildcard pattern match. If new fields are added or removed that match the pattern, Tableau Prep Builder detects the schema change when the flow is run and the pivot results are automatically updated.

1. Connect to your data source.
2. Drag the table that you want to pivot to the **Flow** pane.
3. Click the plus **+** icon, and select **Add Pivot** from the context menu.
4. In the **Pivoted Fields** pane, click on the link **Click here to create wildcard pivot**.



5. Enter a value or partial value that you want to search for. For example, enter **Sales\_** to match fields that are labeled as **sales\_2017**, **sales\_2018** and **sales\_2019**.

Do not use asterisks to match the pattern unless they are part of the field value that you are searching for. Instead click the **Search Options** button to select how you want to match the value. Then press **Enter** to apply the search and pivot the matching values.



6. (Optional) In the **Pivoted Fields** pane, click the plus  $\oplus$  icon to add more columns to

pivot on, then repeat the previous step to select more fields to pivot.

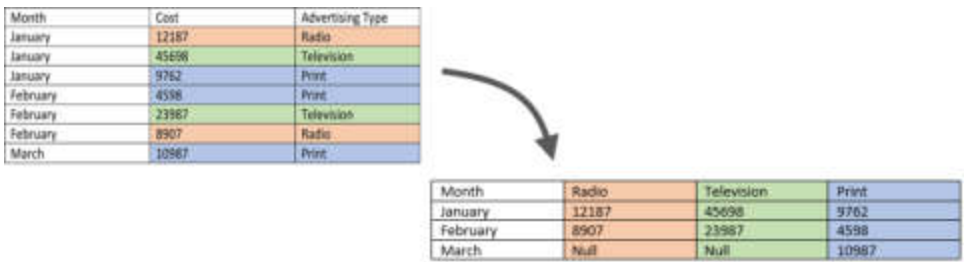
7. If you didn't enable the default naming option or if Tableau Prep Builder couldn't automatically detect a name, edit the names of the fields.
8. To refresh your pivot data when data changes, run your flow. Any new fields added to your data source that match the wildcard pattern are automatically detected and added to the pivot.
9. If the results aren't what you expect, try one of the following options:
  - Enter a different value pattern in the Search field and press enter. The pivot will automatically refresh and show the new results.
  - Manually drag additional fields to the **Pivot1 Values** column in the **Pivoted Fields** pane. You can also remove fields that were added manually by dragging them off the **Pivot1 Values** column and dropping them in the **Fields** pane.

**Note:** Fields that were added from the wildcard search results can't be removed by dragging them off the **Pivot1 Values** column. Instead try using a more specific pattern to match the search results you are looking for.

## Pivot rows to columns (version 2019.1.1 and later)

Use this option if your data is too normalized and you need to create new columns - going from tall data to wider data.

For example if you have advertising costs for each month that includes all advertising types in one column, if you pivot the data from rows to columns you can then have a separate column for each advertising type instead, making the data easier to analyze.



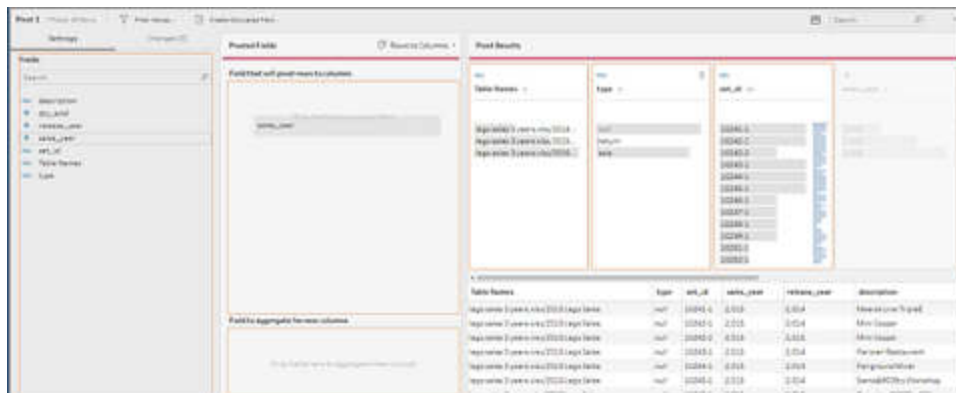
You can select one field to pivot on. The field values for that field are then used to create the new columns. Then, select a field to use to populate the new columns. These field values are aggregated and you can select the type of aggregation to apply.

Because aggregation is applied, pivoting columns back to rows won't reverse this pivot action. To reverse a row to column pivot type, you will need to undo the action. Either click the **Undo** button on the top menu, remove the fields from the **Pivoted Fields** pane or delete the pivot step.

1. Connect to your data source.
2. Drag the table that you want to pivot to the **Flow** pane.
3. Click the plus  $\oplus$  icon, and select **Add Pivot** from the context menu.
4. In the **Pivoted Fields** pane, select **Rows to Columns** from the drop-down list.
5. (Optional) In the **Fields** pane, enter a value in the **Search** field to search the field list for fields to pivot
6. Select a field from the left pane, and drag it to the **Field that will pivot rows to columns** section in the **Pivoted Fields** pane.

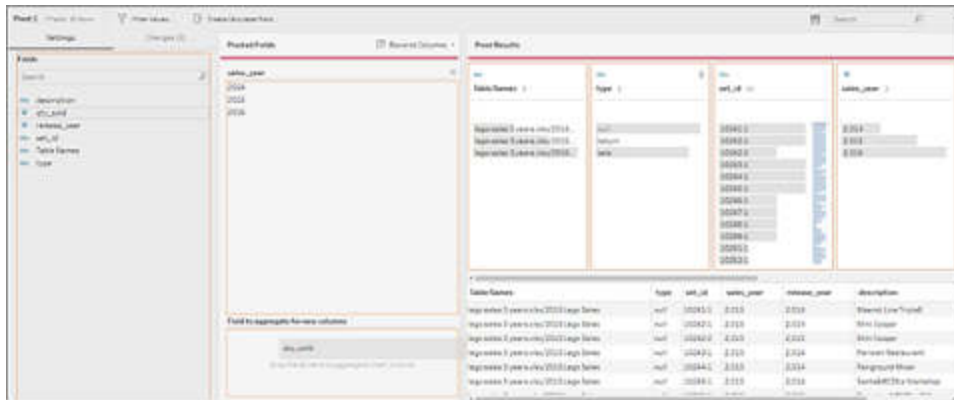
**Note:** If the field you want to pivot on has a data type of date or datetime, you will need to change the data type to string to pivot it.

The values in this field will be used to create and name the new columns. You can change the column names in the **Pivot Results** pane later.

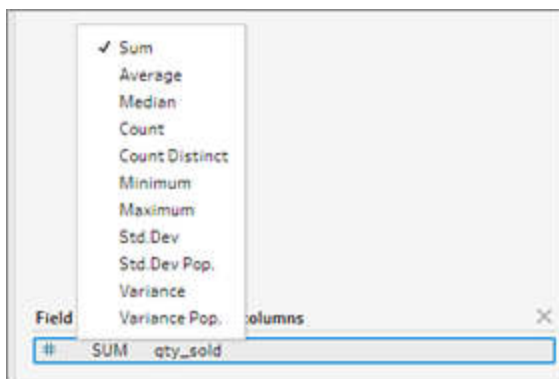


7. Select a field from the left pane and drag it to the **Field to aggregate for new columns** section in the **Pivoted Fields** pane. The values in this field are used to populate the new columns created from the previous step.

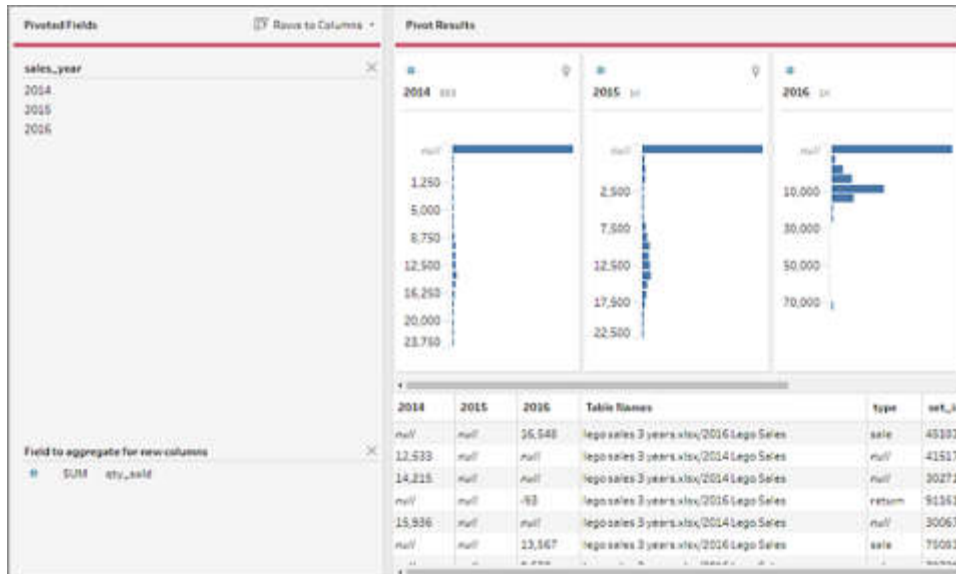




A default aggregation type is assigned to the field. Click the aggregation type to change it.



8. In the **Pivot Results** pane, review the results and apply any cleaning operations to the new columns that you created.

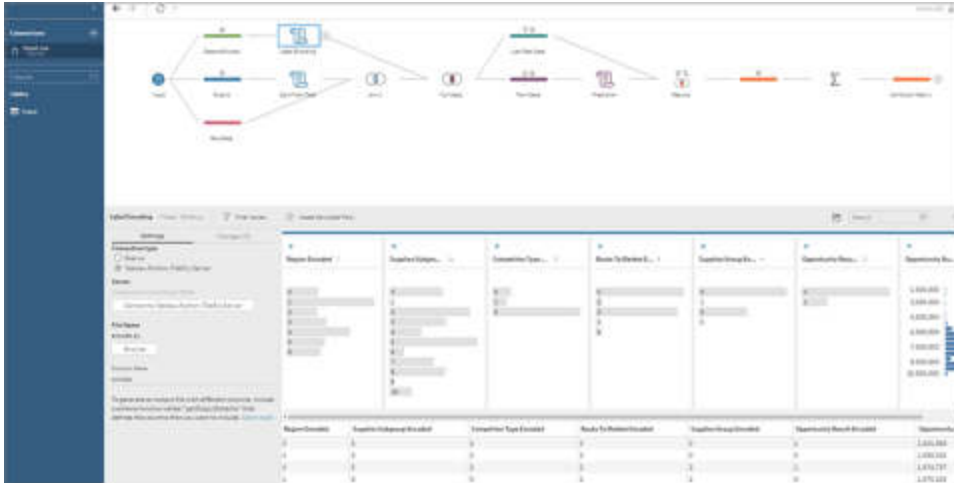


- If the field being pivoted has a change in its row data, right-click or Ctrl-click (Mac) on the Pivot step in the flow pane and select **Refresh**.

## Use R and Python scripts in your flow

Starting in version 2019.3.1 you can use R and Python scripts to perform more complex cleaning operations or incorporate predictive modeling data into your flow. Data is passed from the flow as input through the R or Python script step, then returned as output data that you can continue cleaning using the features and functions of Tableau Prep Builder.

**Note:** Connecting to scripts as an input step for your flow is not yet supported.



When you add a script step to your flow and enter in the configuration details for the selected service, Tableau Prep Builder passes the data to Rserve for R or Tableau Python server (TabPy) for Python and returns the resulting data back to the flow in the form of a table. You can continue to apply cleaning operations to the results and generate your output for analysis.

When you create your script, you will need to include a function that specifies a data frame as an argument of the function. If you want to return different fields than what you input, you'll need to include a `getOutputSchema` function in your script that defines the output and data types. Otherwise, the output will use the fields from the input data.

For information about how to configure R or Python to use in your flows and how to create your scripts, see [Use R \(Rserve\) scripts in your flow](#) below or [Use Python scripts in your flow](#) on page 299.

## Use R (Rserve) scripts in your flow

**Disclaimer:** This topic includes information about a third-party product. Please note that while we make every effort to keep references to third-party content accurate, the information we provide here might change without notice as R and Rserve changes. For the most up-to-date information, please consult the R and Rserve documentation and support.

R is an open source software programming language and a software environment for statistical computing and graphics. To extend the functionality of Tableau Prep Builder, you can create scripts in R to use in your flow that run through an Rserve server to produce output that you can further work with in your flow.

For example, you might want to add statistical modeling data or forecasting data to the data that you already have in your flow using a script in R, then use the power of Tableau Prep Builder to clean the resulting data set for analysis.

To include R scripts in your flow, you need to configure a connection between Tableau Prep Builder and an Rserve server. Then you can use R scripts to apply supported functions to data from your flow using R expressions. After you enter the configuration details and point Tableau Prep Builder to the file and function that you want to use, data is securely passed to the Rserve server, the expressions are applied, and the results are returned as a table (R data.frame) that you can clean or output as needed.

**Note:** To run flows that include script steps on Tableau Server (version 2019.3 and later), Tableau Server must also have a connection to an Rserve server.

## Prerequisites

To include R script steps in your flow, install R and configure a connection to an Rserve server.

## Resources

- [Download and Install R](#). Download and install the most current version of R for Linux, Mac, or Windows.
- [R Implementation notes](#) (community post). Install and configure a connection to R and Rserve for Windows.
- [Install and configure Rserve](#): Instructions for general installation and configuration for all platforms.
  - [Rserve for Windows \(release notes\)](#): This topic covers limitations when installing Rserve locally on Windows.

## Configure Rserve Server for Tableau Server

If you intend to publish flows that include script steps and run them on Tableau Server you will need to configure a connection between your Rserve server and Tableau Server version 2019.3 or later. Running flows with script steps in Tableau Online isn't currently supported.

1. Open the TSM command line.
2. Enter the following commands to set the host address, port values, and connect timeout:

## Tableau Prep Help

```
tsm security maestro-rserve-ssl enable --connection-type
{maestro-rserve-secure/maestro-rserve} --rserve-host <Rserve
IP address or host name> --rserve-port <Rserve port> --
rserve-username <Rserve username> --rserve-password <Rserve
password> --rserve-connect-timeout-ms <Rserve connect
timeout>
```

- **Select** {maestro-rserve-secure} to enable a secure connection or {maestro-rserve} to enable an unsecured connection.
- **If you select** {maestro-rserve-secure}, specify the certificate file `-cf<certificate file path>` in the command line.
- **Specify the** `--rserve-connect-timeout-ms <Rserve connect timeout>` in milliseconds. For example `--rserve-connect-timeout-ms 900000`.

### 3. To disable the Rserve connection enter the following command

```
tsm security maestro-rserve-ssl disable
```

## Additional Rserve configuration (optional)

You can create a file named `Rserv.cfg` to set default configuration values to customize Rserve and place it in the `/etc/Rserve.conf` installation location. To improve stability with the Rserve server and Tableau Prep Builder, you can add additional values to your Rserve configuration. When you launch Rserve you can refer to this file to apply your configuration options. For example:

- **Windows:** `Rserve (args="--RS-conf C:\\folder\\Rserv.cfg")`
- **MacOS and Linux:** `Rserve (args=" --no-save --RS-conf ~/Documents/Rserv.cfg")`

The following example shows some additional options you can include in your `Rserve.conf` configuration file:

```
# If your data includes characters other than ASCII, make it
explicit that data should be UTF8 encoded.
encoding utf8
# Disable interactive behavior for Rserve or Tableau Prep Builder
will stall when trying to run the script as it waits for an input
```

```
response.
interactive no
```

For information about setting up an Rserve.conf file, see the [Advanced Rserve configuration](#) section in the [R Implementation notes](#) (community post).

## Create your R script

When you create your script, include a function that specifies a data frame as an argument of the function. This will call your data from Tableau Prep Builder. You will also need to return the results in a data frame using supported data types.

For example:

```
postal_cluster <- function(df) {
  out <- kmeans(c(df$Latitude, df$Longitude), 3, iter.max=10)
  return(data.frame(Latitude=df$Latitude, Longitude=df$Longitude,
Cluster=out$cluster))
}
```

The following data types are supported:

Data type in Tableau Prep Builder	Data type in R
String	Standard UTF-8 string
Decimal	Double
Int	Integer
Bool	Logical
Date	String in ISO_DATE format “YYYY-MM-DD” with optional zone offset. For example, “2011-12-03+01:00” is a valid date.
DateTime	String in ISO_DATE_TIME format “YYYY-MM-DDT:HH:mm:ss” with optional zone offset. For example, “2011-12-03T10:15:30+01:00” is a valid date.

**Note:** Date and DateTime must always be returned as a valid string. Native Date (DateTime) types in R aren't supported as returned values but can be used in the script.

If you want to return different fields than what you input, you'll need to include a `getOutputSchema` function in your script that defines the output and data types. Otherwise, the output will use the fields from the input data, which are taken from the step that is prior to the script step in the flow.

Use the following syntax when specifying the data types for your fields in the `getOutputSchema`:


Function in R	Resulting data type
<code>prep_string ()</code>	String
<code>prep_decimal ()</code>	Decimal
<code>prep_int ()</code>	Integer
<code>prep_bool ()</code>	Boolean
<code>prep_date ()</code>	Date
<code>prep_datetime ()</code>	DateTime

The following example shows the `getOutputSchema` function for the `postal_cluster` script:

```
getOutputSchema <- function() {
  return (data.frame (
    Latitude = prep_decimal (),
    Longitude = prep_decimal (),
    Cluster = prep_int ());
}
```

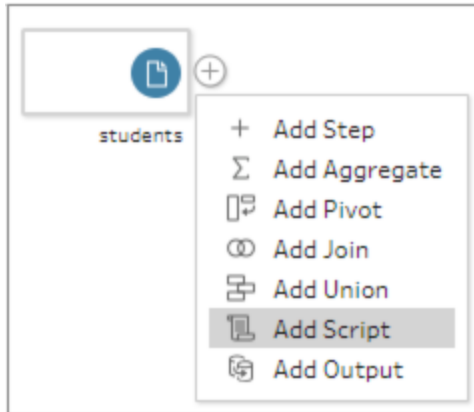
## Add a script to your flow

Start your Rserve server then complete the following steps:

1. Open Tableau Prep Builder and click the **Add connection**  button.
2. From the list of connectors, select the file type or server that hosts your data. If prompted,

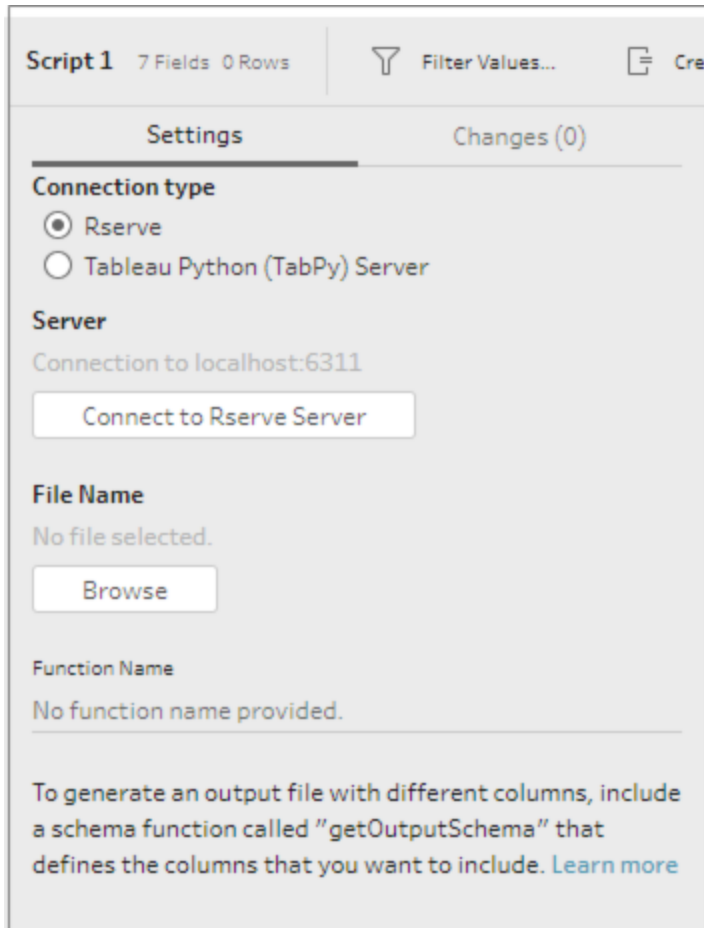
enter the information needed to sign in and access your data.

3. Click the plus  icon, and select **Add Script** from the context menu.

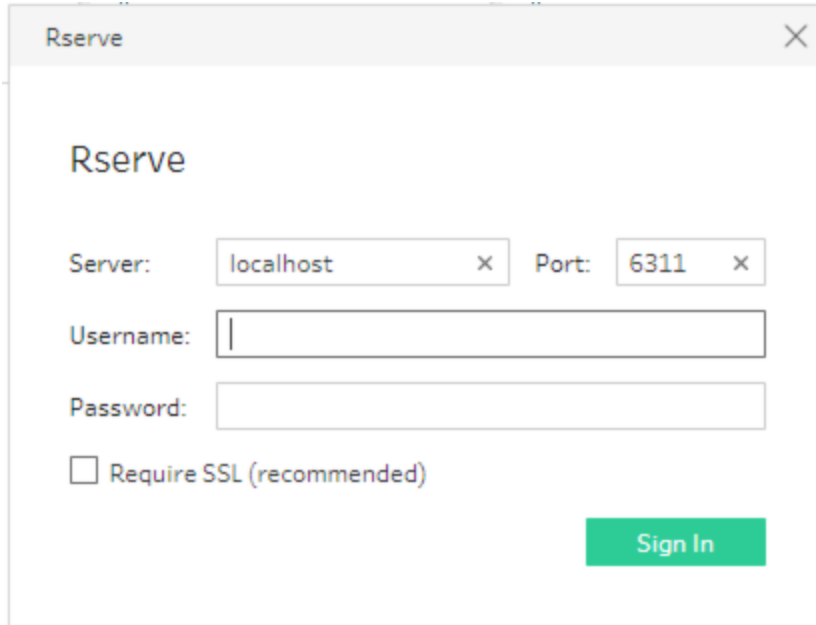


4. In the Script pane, under **Connection type** , select **Rserve**.





5. Click **Connect to Rserve Server** and enter your connection details.
  - Port 6311 is the default port for plaintext Rserve servers.
  - Port 4912 is the default port for SSL-encrypted Rserve servers.
  - If the server requires credentials, enter a Username and Password.
  - If the server uses SSL encryption, select the **Require SSL** check box, then click the **Custom configuration file** link to specify a certificate for the connection.

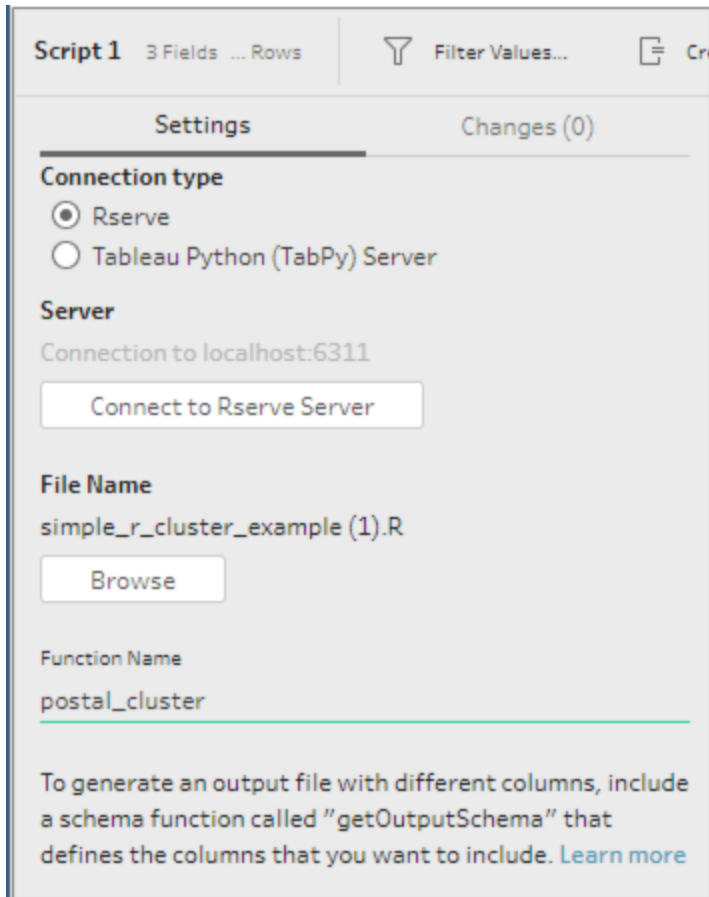


The image shows a dialog box titled "Rserve" with a close button (X) in the top right corner. The dialog contains the following fields and options:

- Server:** A text input field containing "localhost" with a small "x" icon to its right.
- Port:** A text input field containing "6311" with a small "x" icon to its right.
- Username:** An empty text input field.
- Password:** An empty text input field.
- Require SSL (recommended)**
- Sign In** button: A green button with white text located at the bottom right of the dialog.

**Note:** Tableau Prep Builder doesn't provide a way to test the connection. If there is a problem with the connection an error message shows when you try and run the flow.

6. In the **File Name** section, click **Browse** to select your script file.
7. Enter the **Function Name** then press **Enter** to run your script.



## Use Python scripts in your flow

**Disclaimer:** This topic includes information about a third-party product. Please note that while we make every effort to keep references to third-party content accurate, the information we provide here might change without notice as python changes. For the most up-to-date information, please consult the python documentation and support.

Python is a widely used high-level programming language for general-purpose programming. By sending Python commands to an external service through Tableau Prep Builder, you can easily extend your data preparation options by performing actions like adding row numbers, ranking fields, filling down fields and performing other cleaning operations that you might otherwise do using calculated fields.

To include Python scripts in your flow, you need to configure a connection between Tableau Prep Builder and a TabPy server. Then you can use Python scripts to apply supported functions

to data from your flow using a pandas dataframe. When you add a script step to your flow and specify the configuration details, file, and function that you want to use, data is securely passed to the TabPy server, the expressions in the script are applied, and the results are returned as a table that you can clean or output as needed.

**Note:** To run flows that include script steps on Tableau Server (version 2019.3 and later), Tableau Server must also have a connection to a TabPy server.

## Prerequisites

To include Python scripts in your flow, complete the following setup.

1. Download and install **Python**. Download and install the most current version of Python for Linux, Mac or Windows.
2. Download and install the Tableau Python server (**TabPy**). Follow the installation and configuration instructions for installing TabPy. Tableau Prep Builder uses TabPy to pass data from your flow through TabPy as the input, applies your script, then returns the results back to the flow.
3. Install Pandas. Run `pip3 install pandas`. You must use a pandas data frame in your scripts to integrate with Tableau Prep Builder.

## Configure the Tableau Python (TabPy) server for Tableau Server

If you intend to publish flows that include script steps and run them on Tableau Server you will need to configure a connection between your TabPy server and Tableau Server version 2019.3 or later. Running flows with script steps in Tableau Online isn't currently supported.

1. Open the TSM command line/shell .
2. Enter the following commands to set the host address, port values and connect timeout:

```
tsm security maestro-tabpy-ssl enable --connection-type
{maestro-tabpy-secure/maestro-tabpy} --tabpy-host <TabPy IP
address or host name> --tabpy-port <TabPy port> --tabpy-
username <TabPy username> --tabpy-password <TabPy password>
--tabpy-connect-timeout-ms <TabPy connect timeout>
```

- **Select** {maestro-tabpy-secure} **to enable a secure connection** or {maestro-tabpy} **to enable an unsecured connection.**
- **If you select** {maestro-tabpy-secure}, **specify the certificate file** `-cf<certificate file path>` **in the command line.**

- Specify the `--tabpy-connect-timeout-ms <TabPy connect timeout>` in milliseconds. For example `--tabpy-connect-timeout-ms 900000`.

3. To disable the TabPy connection enter the following command

```
tsm security maestro-tabpy-ssl disable
```

## Create your python script

When you create your script, include a function that specifies a pandas (`pd.DataFrame`) as an argument of the function. This will call your data from Tableau Prep Builder. You will also need to return the results in a pandas (`pd.DataFrame`) using supported data types.

For example to add encoding to a set of fields in a flow, you could write the following script:

```
def encode(input):
    le = preprocessing.LabelEncoder()
    Return pd.DataFrame({
        'Opportunity Number' : input['Opportunity Number'],
        'Supplies Subgroup Encoded' : le.fit_transform(input['Supplies Subgroup']),
        'Region Encoded' : le.fit_transform(input['Region']),
        'Route To Market Encoded' : le.fit_transform(input['Route To Market']),
        'Opportunity Result Encoded' : le.fit_transform(input['Opportunity Result']),
        'Competitor Type Encoded' : le.fit_transform(input['Competitor Type']),
        'Supplies Group Encoded' : le.fit_transform(input['Supplies Group']),
    })
```

The following data types are supported:

Data type in Tableau Prep Builder	Data type in Python

String	Standard UTF-8 string
Decimal	Double
Int	Integer
Bool	Boolean
Date	String in ISO_DATE format “YYYY-MM-DD” with optional zone offset. For example, “2011-12-03” is a valid date.
DateTime	String in ISO_DATE_TIME format “YYYY-MM-DDT:HH:mm:ss” with optional zone offset. For example, “2011-12-03T10:15:30+01:00” is a valid date.

**Note:** Date and DateTime must always be returned as a valid string.

If you want to return different fields than what you input, you'll need to include a `get_output_schema` function in your script that defines the output and data types. Otherwise, the output will use the fields from the input data, which are taken from the step that is prior to the script step in the flow.

Use the following syntax when specifying the data types for your fields in the `get_output_schema`:

Function in Python	Resulting data type
<code>prep_string ()</code>	String
<code>prep_decimal ()</code>	Decimal
<code>prep_int ()</code>	Integer
<code>prep_bool ()</code>	Boolean
<code>prep_date ()</code>	Date
<code>prep_datetime ()</code>	DateTime

The following example shows the `get_output_schema` function added to the field encoding python script:



## Tableau Prep Help

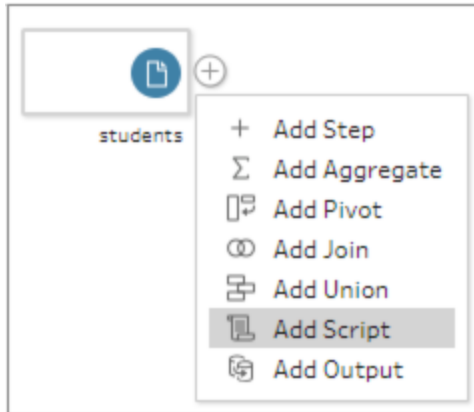
```
def get_output_schema():
    return pd.DataFrame({
        'Opportunity Number' : prep_int(),
        'Supplies Subgroup Encoded' : prep_int(),
        'Region Encoded' : prep_int(),
        'Route To Market Encoded' : prep_int (),
        'Opportunity Result Encoded' : prep_int (),
        'Competitor Type Encoded' : prep_int()
        'Supplies Group Encoded' : prep_int()
    })
```

### Add a script to your flow

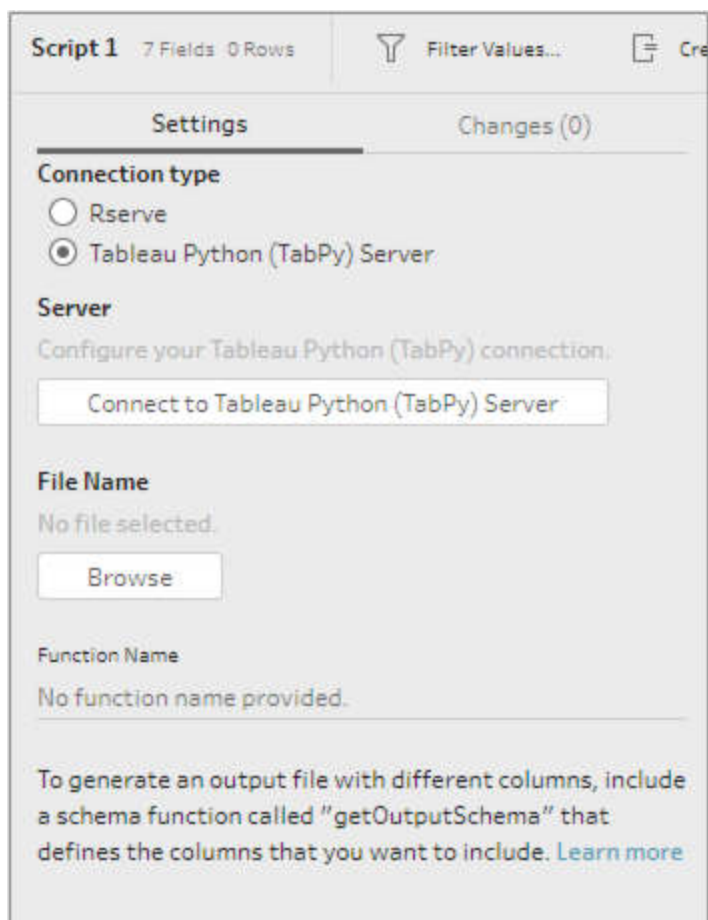
Start your TabPy server then complete the following steps:

**Note:** TabPy requires tornado package version 5.1.1 to run. If you receive the error 'tornado.web' has no attribute 'asynchronous' when trying to start TabPy, from the command line run `pip list` to check the version of tornado that was installed. If you have a different version installed, download the [tornado package version 5.1.1](#). Then run `pip uninstall tornado` to uninstall your current version, then run `pip install tornado==5.1.1` to install the required version.

1. Open Tableau Prep Builder and click the **Add connection**  button.
2. From the list of connectors, select the file type or server that hosts your data. If prompted, enter the information needed to sign in and access your data.
3. Click the plus  icon, and select **Add Script** from the context menu.



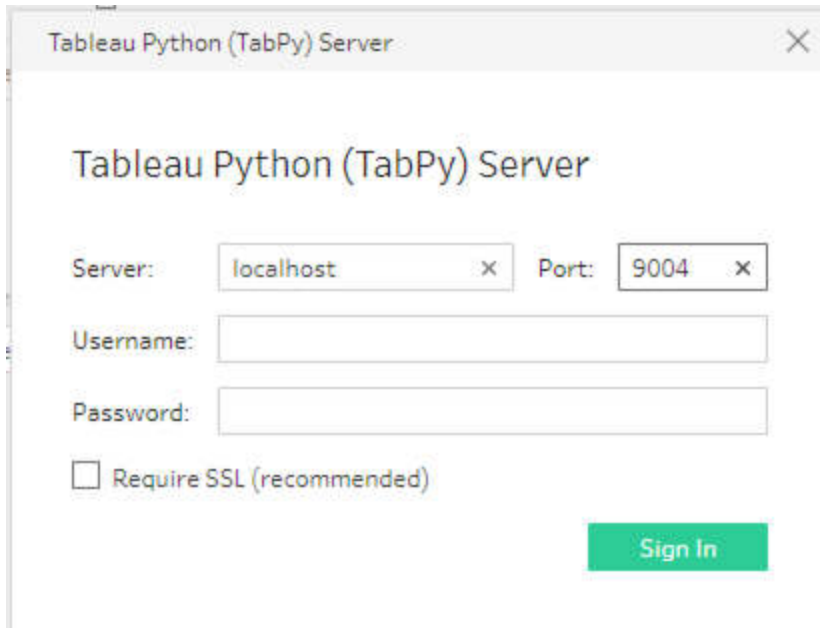
4. In the Script pane, in the **Connection type** section, select **Tableau Python (TabPy) Server**.



5. Click **Connect to Tableau Python (TabPy) Server** and enter your connection details.



- Port 9004 is the default port for TabPy.
- If the server requires credentials, enter a username and password.
- If the server uses SSL encryption, select the Require SSL check box, then click the **Custom configuration file...** link to specify a certificate for the connection.



The screenshot shows a dialog box titled "Tableau Python (TabPy) Server". It contains the following fields and options:

- Server:** A text box containing "localhost" with a close button (x).
- Port:** A text box containing "9004" with a close button (x).
- Username:** An empty text box.
- Password:** An empty text box.
- Require SSL (recommended)**
- Sign In** button (green)

**Note:** Tableau Prep Builder doesn't provide a way to test the connection. If there is a problem with the connection an error message shows.

6. In the **File Name** section, click **Browse** to select your script file.
7. Enter the **Function Name** then press **Enter** to run your script.

**Script 1** 6 Fields 65 Rows Filter Values... Cr

---

**Settings** Changes (0)

**Connection type**

Rserve

Tableau Python (TabPy) Server

**Server**

Connection to localhost:9004

**File Name**

FillDown.py

**Function Name**

FillDown

---

To generate an output file with different columns, include a schema function called "getOutputSchema" that defines the columns that you want to include. [Learn more](#)


# Aggregate, Join, or Union Data

Aggregate, join, or union your data to group or combine data for analysis.

## Aggregate and group values

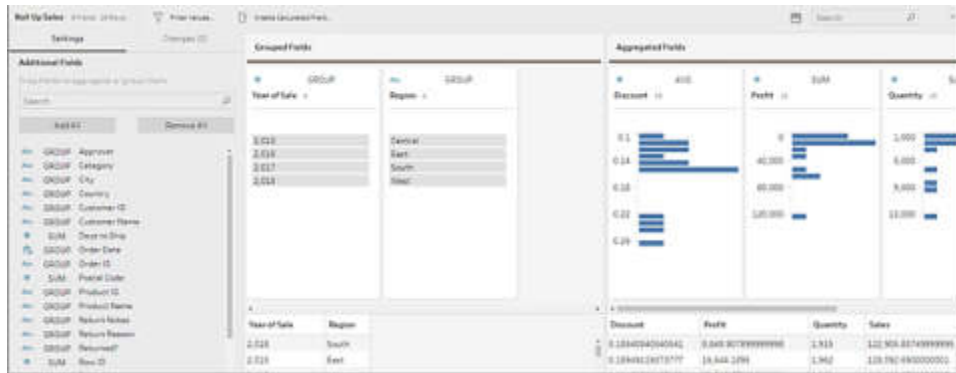
Sometimes you'll need to adjust the granularity of some data, either to reduce the amount of data produced from the flow, or to align data with other data you might want to join or union together. For example, you might want to aggregate sales data by customer before joining a sales table with a customer table.

If you need to adjust the granularity of your data, use the **Add Aggregate** option to create a step to aggregate or group data. Whether data is aggregated or grouped depends on the data type (string, number, or date).

1. In the **Flow** pane, click the plus  icon, and select **Add Aggregate**. A new aggregation step displays in the **Flow** pane and the **Profile** pane updates to show the aggregate and group profile.
2. To group or aggregate fields, drag them from the left pane to one of the columns in the right pane.

You can also:

- Drag and drop fields between the two panes.
- Search for fields in the list and select only the fields you want to include in your aggregation.
- Double-click a field to add it to the left or right pane.
- Change the function of the field to automatically add it to the appropriate pane.
- Click **Add All** or **Remove All** to bulk apply or remove fields.
- Apply certain cleaning operations to fields. For more information about which cleaning options are available, see [About cleaning operations on page 229](#).



Fields are distributed between the **Grouped Fields** and **Aggregated Fields** columns based on their data type. Click the group or aggregation type (for example, AVG or SUM) headings to change the group or aggregation type.

In the data grids below the aggregation and group profile, you can see a sample of the members of the group or aggregation.


Any cleaning operations that are made to the fields are tracked in the **Changes** pane.

## Join your data


The data that you want to analyze is often made up of a collection of tables that are related by specific fields. Joining is a method for combining the related data on those common fields. The result of combining data using a join is a table that's typically extended horizontally by adding fields of data.

Joining is an operation you can do anywhere in the flow. Joining early in a flow can help you understand your data sets and expose areas that need attention right away.

To create a join, do the following:

1. Join two tables using one of the following methods:
  - Add at least two tables to the flow pane, then select and drag the related table to the other table until the **Join** option displays or click the  icon and select **Add Join** from the menu.




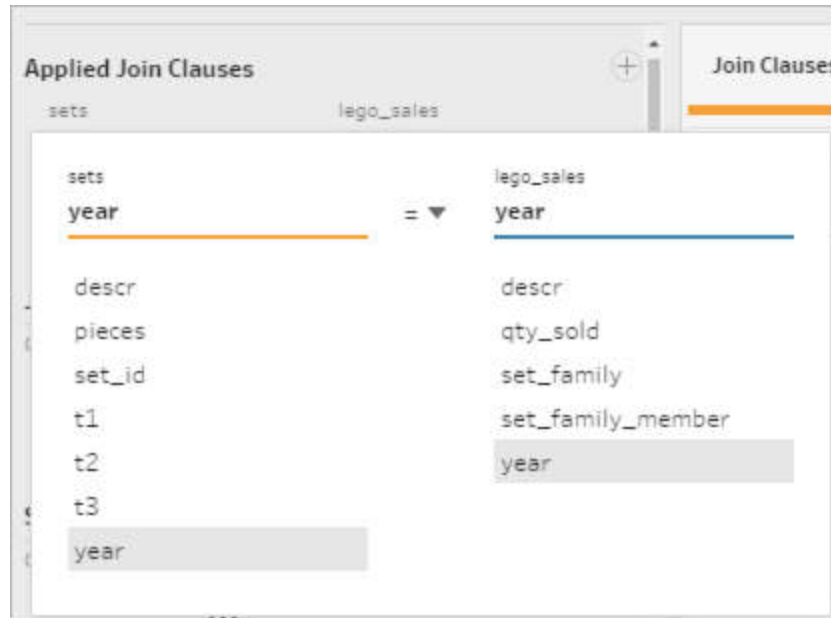
- Click the  icon and select **Add Join** from the menu, then manually add the other input to the join and add the join clauses.

**Note:** In Tableau Prep Builder version 2019.1.3 and later, if you connect to a table that has table relationships defined and includes related fields, you can select **Add Join** and select from a list of related tables. Tableau Prep Builder creates the join based on the fields that make up the relationship between the two tables.

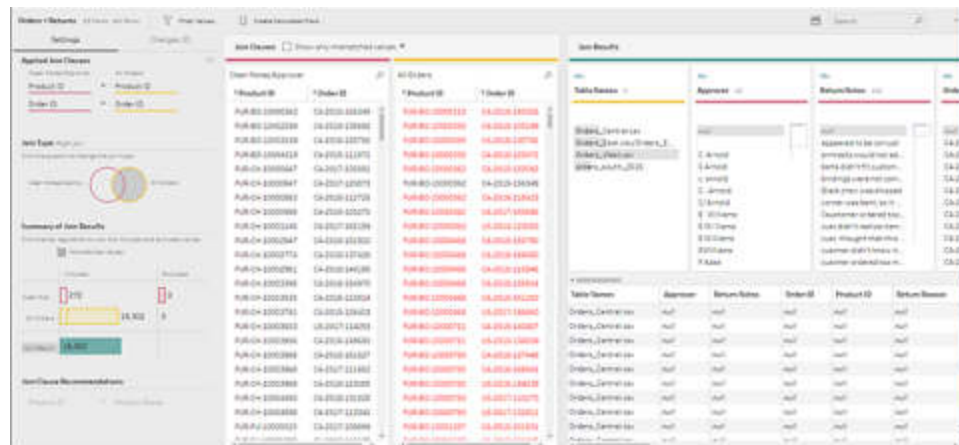
For more information about connectors with table relationships, see [Join data in the Input step \(version 2019.1.3 and later\)](#) on page 179.

A new join step is added to the flow and the profile pane updates to show the join profile.

2. To review and configure the join, do the following:
  - a. Review the **Summary of Join Results** to see the number of fields included and excluded as a result of the join type and join conditions.
  - b. Under **Join Type**, click in the Venn diagram to specify the type of join you want.
  - c. Under **Applied Join Clauses**, click the plus  icon or, on the field chosen for the default join condition, specify or edit the join clause. The fields you selected in the join condition are the common fields between the tables in the join.



- d. Alternatively, you can click the recommended join clauses shown under **Join Clause Recommendations** to add the clause to the list of applied join clauses.




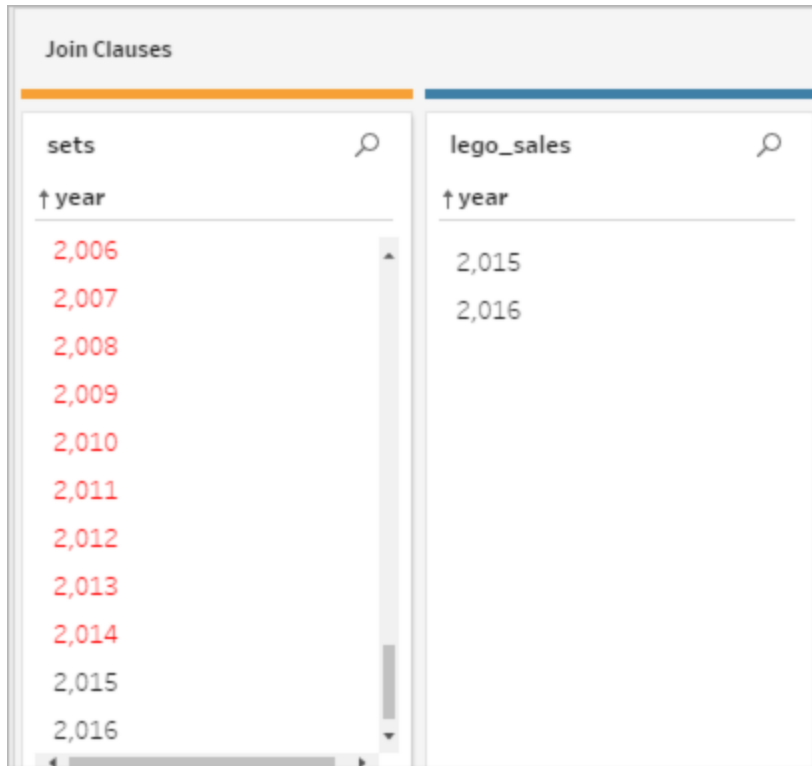
## Inspect the results of the join

The summary in the join profile shows the join metadata to help you validate that the join includes the data you expect.

- **Applied Join Clauses:** By default, Tableau Prep Builder defines the first join clause based on common field names in the tables being joined. Add or remove join clauses as

needed.

- **Join Type:** By default, when you create a join, Tableau Prep Builder uses an inner join between the tables. Depending on the data that you connect to, you might be able to use left, inner, right, or outer joins.
- **Summary of Join Results:** The Summary of Join Results shows you the distribution of values that are included and excluded from the tables in the join.
  - Click each **Included** bar to isolate and see the data in the join profile included in the join.
  - Click each **Excluded** bar to isolate and see the data in the join profile that are excluded from the join.
  - Click any combination of the **Included** and **Excluded** bars to see a cumulative perspective of the data.
- **Join Clause Recommendations:** Click the plus  icon next to the recommended join clause to add it to the **Applied Join Clauses** list.
- **Join Clauses** pane: In the **Join Clauses** pane, you can see the values in each field in the join clause. The values that don't meet the criteria for the join clause are displayed in red text.



- **Join Results** pane: If you see values in the **Join Results** pane that you want to change, you can edit the values in this pane.

## Common join issues

If you don't see the results you expect after joining your data, you may need to do some additional cleaning on your field values. The following issues will result in Tableau Prep Builder reading the values as not a match and exclude them from the join:

- Different capitalization: My Sales and my sales
- Different spelling: Hawaii and Hawai'i
- Misspelling or typo's: My Company Health and My Company Heath
- Name Changes: John Smith and John Smith Jr.
- Abbreviations: My Company Limited and My Company Ltd
- Extra separators: Honolulu and Honolulu (Hawaii)
- Extra spaces: This includes extra space between characters, tabbed spaces or extra



leading or trailing spaces

- Inconsistent use of periods: Returned, not needed and Returned, not needed.

The good news is that if your field values have any of these issues, you can fix the field values in-line directly in the **Join Clauses** or work with excluded values by clicking in the Excluded bars in the **Summary of Join Results** and use the cleaning operations in the profile card menu.

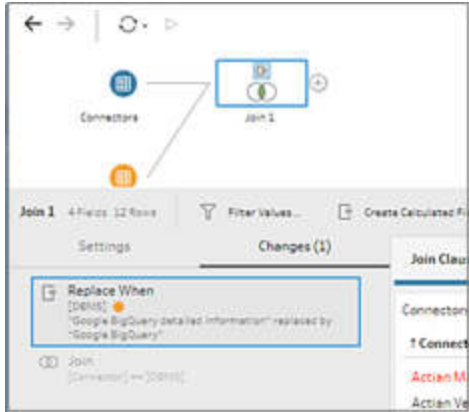
For more information about the different cleaning options available in the Join step, see [About cleaning operations on page 229](#).

## Fix mismatched fields and more

You can fix mismatched fields right in the join clause. Double-click or right-click the value and select **Edit Value** from the context menu on the field that you want to fix and enter a new value. Your data changes are tracked and added to the **Changes** pane right in the Join step.



You can also select multiple values to keep, exclude or filter in the Join Clauses panes, or apply other cleaning operations in the Join Results pane. Depending on which fields you change and where they are in the join process, your change is applied either before or after the join to give you the corrected results.



For more information about cleaning fields see [Apply cleaning operations](#) on page 233.


## Union your data

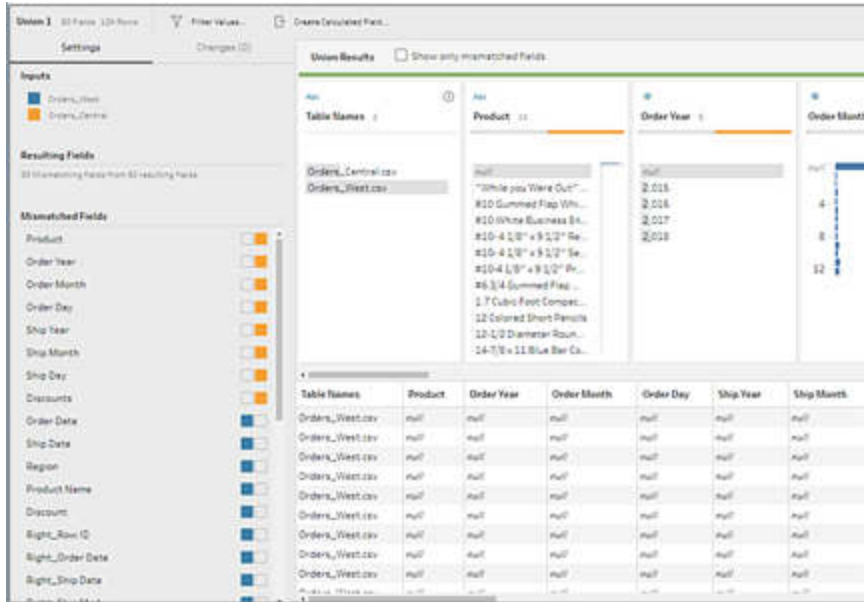
Union is a method for combining data by appending rows of one table onto another table. For example, you might want to add new transactions in one table to a list of past transactions in another table. Make sure the tables you union have the same number of fields, the same field names, and the fields are the same data type.

**Tip:** To maximize performance a single union can have a maximum of 10 inputs. If you need to union more than 10 files or tables, try unioning files in the Input step. For more information about this type of union, see [Union files and database tables in the Input step](#) on page 173.

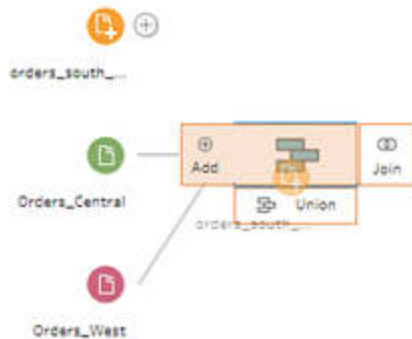
Similar to a join, you can use the union operation anywhere in the flow.

To create a union, do the following:

1. After you add at least two tables to the flow pane, select and drag a related table to the other table until you see the **Union** option. You can also click the  icon and select **Add Union** from the menu. A new union step is added in the **Flow** pane, and the **Profile** pane updates to show the union profile.



2. Add additional tables to the union by dragging tables toward the unioned tables until you see the **Add** option.

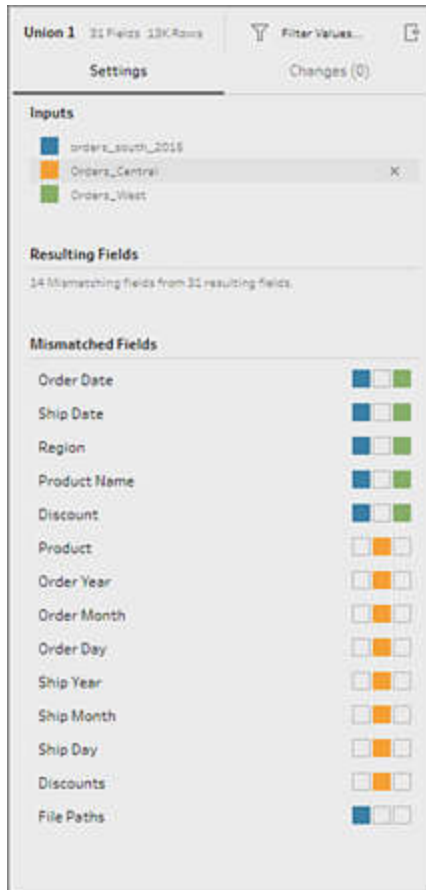


3. In the union profile, review the metadata about the union. You can remove tables from the union as well as see details about any mismatched fields.

## Inspect the results of the union

After you create a union, inspect the results of the union to validate that the data in the union is what you expect. There are a number of areas in the union profile that you can check to help you validate the data in the union.

- **Review the union metadata:** The union profile shows some metadata about the union. Here you can see the tables that make up the union, the resulting number of fields and any mismatched fields.



- **Review the colors for each field:** Next to each field listed in the Union summary and above each field in the union profile, is a set of colors. The colors correspond to each table in the union.

If all table colors show for that field, then the union performed correctly for that field. A missing table color indicates that you have mismatched fields.



Mismatched fields are fields that might have similar data but are different in some way. You can see the list of fields that don't match in the Union summary and the tables where they came from. If you want to take a closer look at the data in the fields, select the **Show only mismatched fields** check box to isolate the mismatched fields in the Union profile.



To fix these field, follow one of the suggestions in the “Fix fields that don’t match” section below.

## Fix fields that don’t match

When tables in a union don't match, the union produces extra fields. The extra fields are valid data being excluded from their appropriate context.

To resolve a field mismatch issue, you must merge the mismatched fields together.

There are a number of reasons why fields might not match.

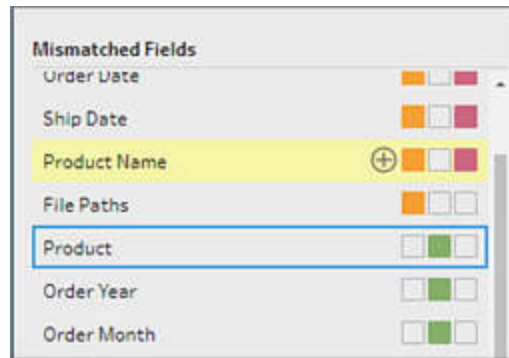
- **Corresponding fields have different names:** If corresponding fields between tables have different names, you can use union recommendations, manually merge fields in the **Mismatched Fields** list, or rename the field in the union profile to merge the mismatched fields together.

To use union recommendations, do the following:

1. in the **Mismatched Fields** list, click on a mismatched field. If a suggested match exists, the matching field is highlighted in yellow.

Suggested matches are based on fields with similar data types and field names.

2. Hover on the highlighted field and click the plus button to merge the fields.

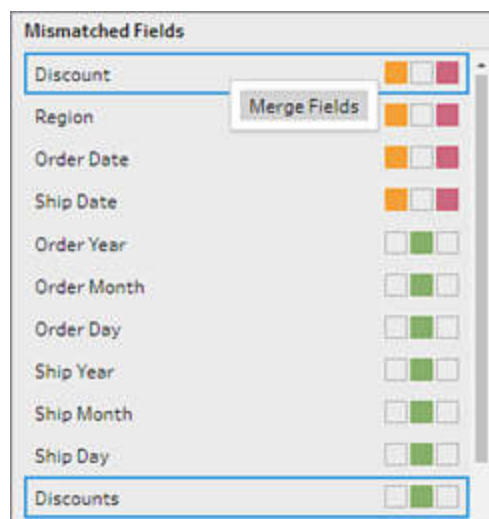


To manually merge fields in the **Mismatched Fields** list, do the following:

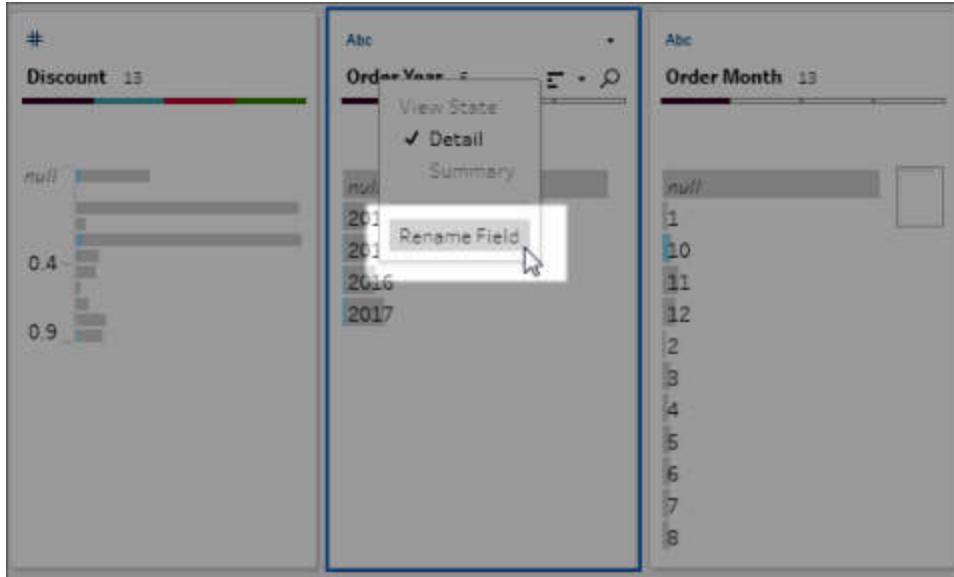
1. Select one or more fields in the list.
2. Right click a selected field and if the merge is valid, the **Merge Fields** menu option appears.

If you see **No options available** when you right-click the field, this is because the fields are not eligible to merge. For example trying to merge two fields from the same input.

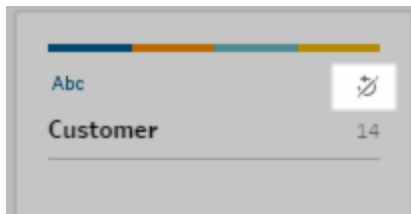
3. Click **Merge Fields** to merge the selected fields.



To rename the field in the union profile pane, right-click the field name and click **Rename Field**.



- **Corresponding fields have the same name but are a different type:** By default, when the name of corresponding fields match but the data type of the fields don't, Tableau Prep Builder will change the data type of one of the fields so they are compatible with each other. If Tableau Prep Builder makes this change, it's noted at the top of the merged field by the Change Data Type icon.



In some cases, Tableau Prep Builder might not pick the correct data type. If that happens and you want to undo the merge, right-click the Change Data Type icon and select **Separate Inputs with Different Types**.



You can then merge the fields again by first changing the data type of one of the fields and then using the suggestions in [Additional merge field options on the next page](#).

- **Corresponding tables have different number of fields:** To union tables, each table

in the union must contain the same number of fields. If a union results in extra fields, merge the field into an existing field.

## Additional merge field options

In addition to the methods described in the above section for merging fields you can also use one of the following methods to merge fields. You can merge fields in any step, except for the Output step.

For information about how to merge fields in the same file, see [Merge fields on page 241](#).

To merge fields, do one of the following:

- Drag and drop one field onto another. A **Drop to merge fields** indicator displays.
- Select multiple fields and right-click within the selection to open the context menu, and then click **Merge Fields**.
- Select multiple fields, and then click **Merge Fields** on the context-sensitive toolbar.



# Save and Share Your Work

At any point in the flow, you can save your flow, view a preview of the data in your flow in Tableau Desktop, or create an extract of your data that includes all the operations that you've applied to your flow. You can also package your data with your flow to share it with others or publish your data extract to Tableau Server or Tableau Online as a data source.

To keep data fresh you can run individual flows from within Tableau Prep Builder or from the command line. To run flows automatically on a schedule, you can publish flows to Tableau Server or Tableau Online. For more information about running flows, see [Refresh flow output files from the command line](#) on page 334 or [Publish a Flow to Tableau Server or Tableau Online](#) on page 398.

## Save a flow

Save your flow to back up your work before performing any additional operations. Your flow is saved in the Tableau Prep Builder flow (.tfl) file format.

You can also package your local files (Excel, Text Files, and Tableau extracts) with your flow to share with others, just like packaging a workbook for sharing in Tableau Desktop. Only local files can be packaged with a flow. Data from database connections, for example, aren't included.

When you save a packaged flow, the flow is saved as a Packaged Tableau Flow File (.tflx).

- To save your flow, from the top menu, select **File > Save**.
- To package your data files with your flow, from the top menu, do one of the following:
  - Select **File > Export Packaged Flow**
  - Select **File > Save As**. Then in the **Save As** dialog, select **Packaged Tableau Flow Files** from the **Save as type** drop down menu.

## View your data sample in Tableau

Sometimes when you're cleaning your data you might want to check your progress by looking at it in Tableau Desktop. When your flow opens in Tableau Desktop, Tableau Prep Builder creates a permanent Tableau extract (.tde or .hyper depending on your version of Tableau) and a

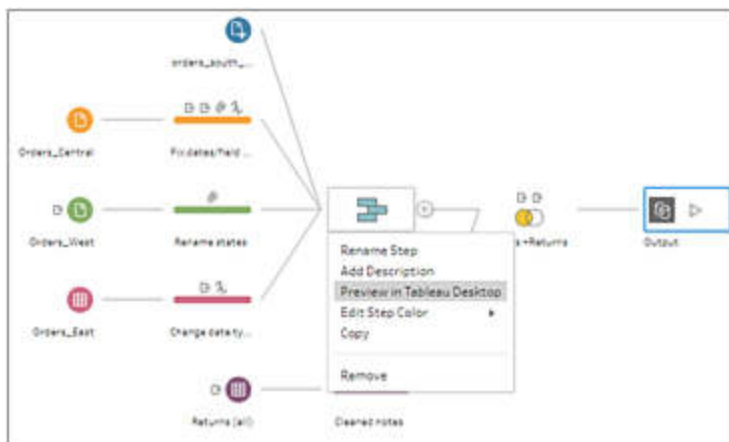
Tableau data source (.tds) file. The files are saved in your Tableau repository in the **Datasources** file so you can experiment with your data at any time.

**Note:** While you can experiment with your data, Tableau only shows you a sample of your data and you won't be able to save the workbook as a packaged workbook (.twbx). When you are ready to work with your data in Tableau, create an output step for your flow and save the output to a file or as a published data source, then connect to the full data source in Tableau.

When you open the flow in Tableau Desktop, you can see the data sample that you are working with in your flow with the operations applied to it, up to the step that you selected.

To view your data sample in Tableau Desktop do the following:

1. Right-click the step where you want to view your data, and select **Preview in Tableau Desktop** from the context menu.



2. Tableau Desktop opens on the **Sheet** tab.

## Create and publish data extracts and data sources

To create an extract, run your flow. When you run your flow your changes are applied to your entire data set. Running the flow results in a Tableau Data Source (.tds) and a Tableau Data Extract (.tde or .hyper) file. You can create an extract file from your flow output to use in Tableau Desktop or to share your data with third parties.

Starting in version 2020.2.1, you can run your flow using incremental refresh. This option saves time and resources and enables you to refresh only new data instead of your full data set. For information about how to configure and run your flow using incremental refresh, see [Refresh Flow Data Using Incremental Refresh](#) on page 326.

**Note:** You can publish data extracts or data sources to Tableau Server version 10.0 and later as well as to Tableau Online.


You can create an extract file in the following formats:


- **Tableau Data Extract (.tde):** The extract is saved as both a Tableau extract (.tde) and a Tableau data source (.tds) file. Use this file type if you use Tableau Desktop or Tableau Server version 10.0 through 10.4.
- **Hyper Extract (.hyper):** This is the new Tableau extract file type but can only be consumed by Tableau Desktop or Tableau Server version 10.5 and later.
- **Comma Separated Value (.csv):** Save the extract to a .csv file to share your data with third parties. The encoding of exported CSV file will be UTF-8 with BOM.

You can also publish your output as a data source to Tableau Server or Tableau Online to share your data and provide centralized access to the data you have cleaned, shaped, and combined.

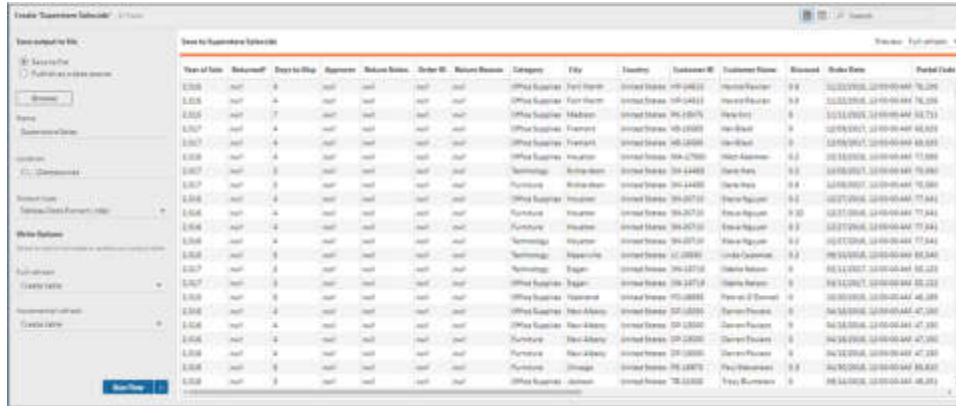
**Note:** To publish your output to Tableau Server, the Tableau Server REST API must be enabled. For more information see [Rest API Requirements](#) in the Tableau Rest API Help. To publish to a server that uses Secure Socket Layer (SSL) encryption certificates, additional configuration steps are needed on the machine running Tableau Prep Builder. For more information, see the [Before you Install](#) in the Tableau Desktop and Tableau Prep Builder Deployment Guide.

## Create an extract file

1. Click the plus icon  on a step and select **Add Output**.

If you have run the flow before, click the run flow  button on the Output step. This runs the flow and updates your output.

The **Output** pane opens and shows you a snapshot of your data.



2. In the left pane select **Save to file**.
3. Click the **Browse** button, then in the **Save Extract As** dialog, enter a name for the file and click **Accept**.
4. In the **Output type** field, select the output type. Depending on the version of Tableau Desktop you use you can choose from the following options:
  - Tableau Data Extract (.hyper) for Tableau Desktop version 10.5 and later.
  - Tableau Data Extract (.tde) for Tableau Desktop version 10.0 through 10.4.
  - Comma Separated Values (.csv) if you want to share the extract with a third party.


**Tip:** You have choices when generating output from your flow. You can generate an extract file, or you can publish your data as a data source to Tableau Server or Tableau Online. For more information about generating output files, see [Create and publish data extracts and data sources](#) on page 322.


5. (version 2020.2.1 and later) In the **Write Options** section, view the default write option to write the new data to your files and make any changes as needed. For more information, see [Configure write options](#) on page 330.
  - **Create table:** This option creates a new table or replaces the existing table with the new output.
  - **Append to table:** This option adds the new data to your existing table. If the table doesn't already exist, a new table is created and subsequent runs will add new rows to this table.

**Note:** **Append to table** isn't supported for .csv and .tde output types. For more information about supported refresh combinations, see [Flow refresh options on page 327](#).

6. Click **Run Flow** to run the flow and generate the extract file.

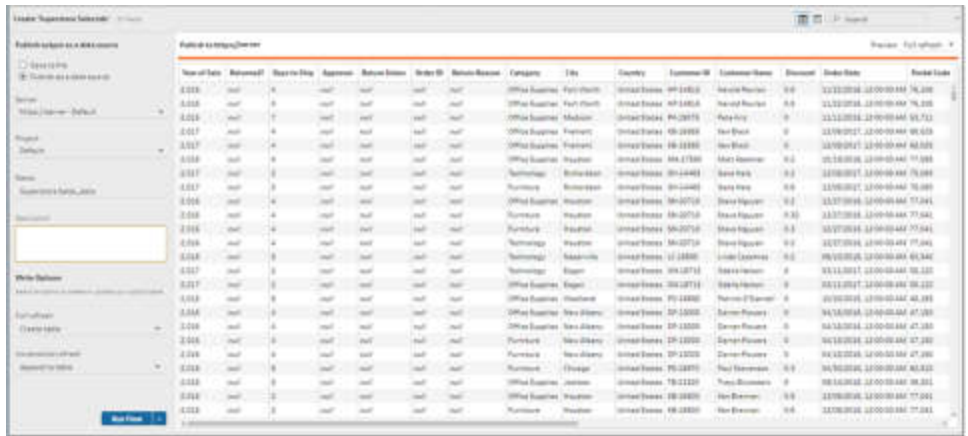
## Publish as a data source

1. Click the plus icon  on a step and select **Add Output**.

If you have run the flow before, click the run flow  button on the Output step. This runs the flow and updates your output.

**Note:** In version 2020.1.1, Tableau Prep Builder will refresh previously published data sources and maintain any data modeling (for example calculated fields, number formatting, and so on) that might be included in the data source. If the data source can't be refreshed, the data source will be replaced instead.

2. The output pane opens and shows you a snapshot of your data.



Row ID	Row ID	Report ID	Approval	Return Status	Order ID	Service Request	Category	Title	Country	Customer ID	Customer Name	Amount	Order Date	Order Date
0.001	1	1	1	1	1	1	Office Supplies	Form Sheets	United States	99-14812	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.002	2	2	1	1	2	2	Office Supplies	Post-it Notes	United States	99-14813	Acme Corp	8.0	11/21/2018	11/21/2018 09:00
0.003	3	3	1	1	3	3	Office Supplies	Markers	United States	99-14814	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.004	4	4	1	1	4	4	Office Supplies	Staplers	United States	99-14815	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.005	5	5	1	1	5	5	Office Supplies	Staplers	United States	99-14816	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.006	6	6	1	1	6	6	Office Supplies	Staplers	United States	99-14817	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.007	7	7	1	1	7	7	Office Supplies	Staplers	United States	99-14818	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.008	8	8	1	1	8	8	Office Supplies	Staplers	United States	99-14819	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.009	9	9	1	1	9	9	Office Supplies	Staplers	United States	99-14820	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.010	10	10	1	1	10	10	Office Supplies	Staplers	United States	99-14821	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.011	11	11	1	1	11	11	Office Supplies	Staplers	United States	99-14822	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.012	12	12	1	1	12	12	Office Supplies	Staplers	United States	99-14823	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.013	13	13	1	1	13	13	Office Supplies	Staplers	United States	99-14824	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.014	14	14	1	1	14	14	Office Supplies	Staplers	United States	99-14825	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.015	15	15	1	1	15	15	Office Supplies	Staplers	United States	99-14826	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.016	16	16	1	1	16	16	Office Supplies	Staplers	United States	99-14827	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.017	17	17	1	1	17	17	Office Supplies	Staplers	United States	99-14828	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.018	18	18	1	1	18	18	Office Supplies	Staplers	United States	99-14829	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.019	19	19	1	1	19	19	Office Supplies	Staplers	United States	99-14830	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.020	20	20	1	1	20	20	Office Supplies	Staplers	United States	99-14831	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.021	21	21	1	1	21	21	Office Supplies	Staplers	United States	99-14832	Acme Corp	9.0	11/21/2018	11/21/2018 09:00
0.022	22	22	1	1	22	22	Office Supplies	Staplers	United States	99-14833	Acme Corp	9.0	11/21/2018	11/21/2018 09:00

3. Select the **Publish as data source** radio button and complete the following fields:
  - **Server:** Select the Tableau server where you want to publish the data source and data extract. If you aren't signed in to a server you will be prompted to sign in.

**Note:** Starting in Tableau Prep Builder version 2020.1.4, once you sign into your server, Tableau Prep Builder remembers your server name and login information when you close the application so that the next time you open the application you are already logged into your server.

On the Mac you may be prompted to provide access to your Mac keychain so Tableau Prep Builder can securely use SSL certificates to connect to your Tableau Server or Tableau Online environment.

If you are outputting to Tableau Online include the pod your site is hosted on in the "serverUrl". For example, "https://eu-west-1a.online.tableau.com" not "https://online.tableau.com".

- **Project:** Select the project where you want to load the data source and extract.
  - **Name:** Enter file name
  - **Description:** Enter a description for the data source.
4. (version 2020.2.1 and later) In the **Write Options** section, view the default write option to write the new data to your files and make any changes as needed. For more information, see [Configure write options on page 330](#)
- **Create table:** This option creates a new table or replaces the existing table with the new output.
  - **Append to table:** This option adds the new data to your existing table. If the table doesn't already exist, a new table is created and subsequent runs will add new rows to this table.
5. Click **Run Flow** to run the flow and publish the data source.

## Refresh Flow Data Using Incremental Refresh

Starting in version 2020.2.1, you can configure selected inputs and outputs in your flow to refresh incrementally so that only the new rows are retrieved and processed when the flow runs, saving you time and resources. For example, if your flow includes transaction data that updates daily, you can set up incremental refresh to retrieve and process only the new transactions every day, then run a full refresh weekly or monthly to refresh all of your flow data.

To run your flow using incremental refresh, Tableau Prep Builder needs the following information:

- The field that detects new rows in the input table.
- The field to use to compare the last processed values in the flow output with the values in the input to determine which rows are new.
- How you want to write the new data to your tables. You can either add new data to your existing tables or overwrite your table data with the new data.

## Flow refresh options

Tableau Prep Builder enables you to select how your data is refreshed and how your tables are updated with the flow output. The following table describes the different options and their benefits.

Refresh Combination	Data Processed	Table Update	Benefits
<b>Full Refresh + Create Table</b>	All	Create or overwrite the existing table with the full data set.	Keep track of all the data on every flow run.
<b>Full Refresh + Append to Table</b>	All	Add new data to the existing table.	Keep track of both new and existing data on every flow run. <b>Append to table</b> isn't available for .csv or .tde output types.
<b>Incremental Refresh + Create Table</b>	New rows only	Create or overwrite the existing table with the new data.	Keep only the latest data as the complete data set.
<b>Incremental Refresh + Append to Table</b>	New rows only	Add the new data to the existing table.	Add the new data to the existing incremental data set for analysis. <b>Append to table</b> isn't available for .csv or .tde output types.

## Configure incremental refresh

To configure your flow to use incremental refresh, you need to specify settings on both the **Input** steps and the **Output** steps where you want to use this option. In the **Input** step, specify

how Tableau Prep Builder will find your new rows. In the **Output** step, specify how the new rows are written to your table. When you run the flow, you can select either a full or incremental refresh type.

1. In the flow pane, select the input step that you want to configure for incremental refresh.
2. In the **Input** pane on the **Settings** tab, under the **Set up Incremental Refresh section**, set the following options:

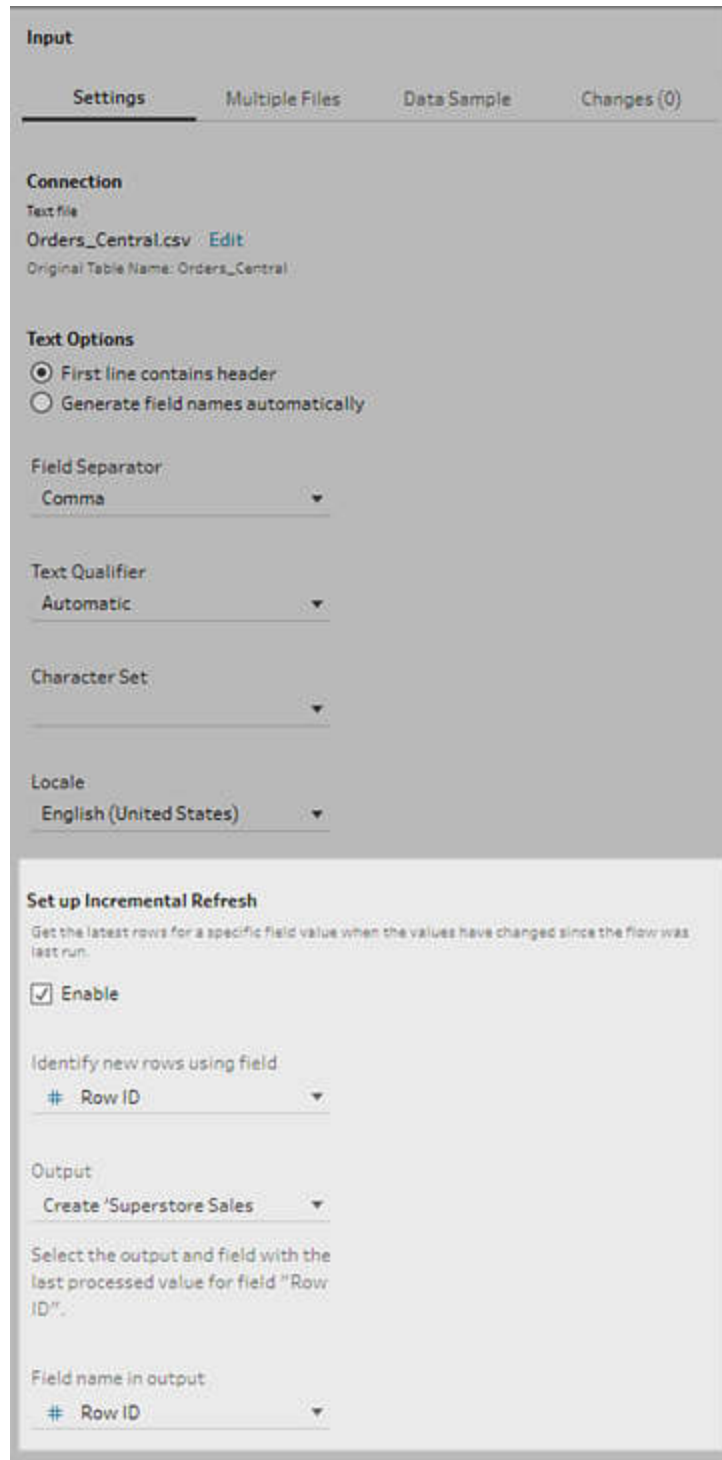
- Select **Enable**.
- **Identify new rows using field**: Select the field that you want to refresh in your input data. This field must be assigned a data type of **Number (whole)**, **Date**, or **Date & Time**. Currently, you can only select a single field.

**Note:** You can remove or rename this field later in the flow, as long as the field you specify in the **Field name in output** can be used to compare this field with the latest output to find new rows.

- **Output**: Select the output that is related to your input and that includes the field that will be used to compare rows.
- **Field name in output**: Select the field to use to compare the last processed values in the flow output with the values in the input to find new rows. This field must have the same data type as the field you specified in the field **Identify new**



**rows using field.**



## Configure write options

To finish setting up incremental refresh, set your output **Write Options** to specify how the new rows are written to your tables. All outputs that are related to the configured input step have a default write option selected, but you can change it to a supported option.

By default, outputs to local or published .hyper extracts are set to **Append to table**. Outputs to .csv or .tde file types are set to **Create table**.

1. In the flow pane, select the output step that you want to configure for incremental refresh.
2. In the Output pane, in the **Write Options** section, view the default write option and make any changes as needed.
  - **Create table:** This option creates a new table or replaces the existing table with the new output.
  - **Append to table:** This option adds the new data to your existing table. If the table doesn't already exist, a new table is created when the flow is first run and subsequent runs will add new rows to this table.

**Note:** **Append to table** isn't supported for .csv and .tde output types. For more information about supported refresh combinations, see [Flow refresh options](#)

on page 327

**Save output to file**

Save to file  
 Publish as a data source

**Browse**

Name  
Superstore Sales

Location  
C:\...\Datasources

Output type  
Tableau Data Extract (.hyper) ▼

**Write Options**  
Select an option to create or update your output table.

Full refresh  
Create table ▼

Incremental refresh  
Append to table ▼

**Create table**  
If the table already exists, it will be deleted.

**Append to table**  
Add new rows to the end of the existing table. If the table doesn't exist, it's created when the flow is first run.

**Run Flow** ▼

## Run your flow

You can run individual flows using incremental refresh in Tableau Prep Builder or from the command line. For information about running your flow from the command line, see [Run the flow with incremental refresh enabled \(version 2020.2.1 and later\)](#) on page 341.

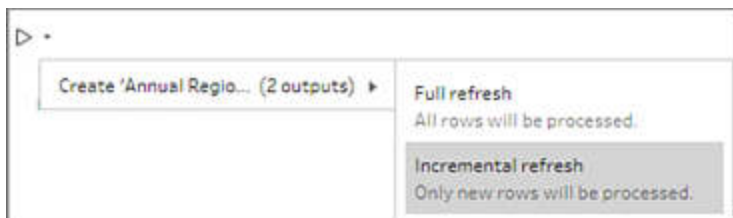
If you have Tableau Prep Conductor (part of the Data Management Add-on) enabled on your server, you can run your flow using incremental refresh by either setting up a schedule in Tableau Server or Tableau Online. For information about running your flow on a schedule, see [Schedule a Flow Task](#) on page 417.

**Note:** Write options are set in Tableau Prep Builder and can't be changed when running your flow in Tableau Server or Tableau Online.

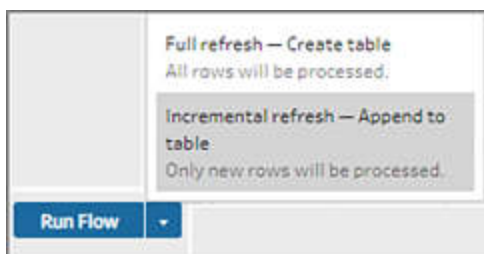
Tableau Prep Builder runs a full refresh for all outputs regardless of the run option you select if no existing output is found. Subsequent flow runs use the incremental refresh process and retrieve and process only the new rows unless incremental refresh configuration data is missing or the existing output is removed.

To run the flow in Tableau Prep Builder using incremental refresh, select **Incremental refresh** from one of the following locations:

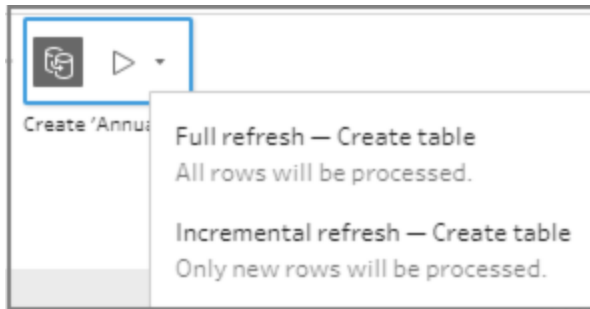
- From the top menu, click the drop-down option on the Run button.



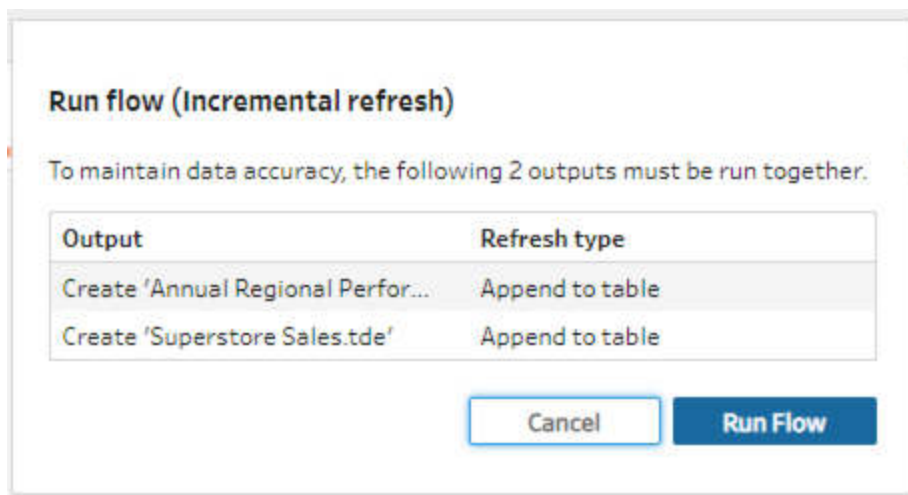
- From the Output pane, click the drop-down option on the **Run Flow** button.



- From the Flow pane, click the drop-down on the Run button next to the Output step.



If one input with incremental refresh enabled is associated with multiple outputs, those outputs must be run together and must use the same refresh type. When you run your refresh in Tableau Prep Builder, a dialog shows letting you know that you must run both outputs together.



## Refresh flow output files from the command line

**Note:** Running flows from the command line that include connections using either the JDBC or ODBC connectors is supported starting in version 2019.2.3.

Also, flows that include cloud connectors, such as Google BigQuery, can't be run from the command line. Instead run the flow manually or run the flow on a schedule in Tableau Server or Tableau Online.

If you want to refresh output files for your flow you can run the flow from the command line instead of opening it in Tableau Prep Builder and running the flow from there. You can run one

flow at a time using this method. This option is available on both Windows and Mac machines where Tableau Prep Builder is installed.

For Windows machines, you can also schedule this process using Windows Task Scheduler.

For more information about using Windows Task Scheduler, see [Task Scheduler](#) in the Microsoft online help.

When you run flows from the command line, Tableau Prep Builder refreshes all outputs for the flow.

You can also schedule your flow to run automatically in Tableau Server or Tableau Online using Tableau Prep Conductor. For more information about using Tableau Prep Conductor to run flows, see [Keep Flow Data Fresh on page 360](#).

**Note:** The output location and write options (version 2020.2.1 and later) for the files is specified in the output step for the flow in Tableau Prep Builder. When you refresh the files from the command line, this process uses the settings configured in the output step to process the data and append or create the data tables.

For information about how to specify an output location for your flow files, see [Refresh flow output files from the command line on the previous page](#). For information about setting your write options, see [Configure write options on page 330](#).

To run the flow from the command line, you'll need the following:

- Administrator privileges on the machine where you are running the flow.
- The path where Tableau Prep Builder is installed.
- If connecting to databases and publishing output files to a server, a credentials .json file that includes all required credentials.
- The path where the Tableau Flow (.tfl) file is located.

## Before running the flow

If you are running a flow that connects to database files, published data sources, or publishes output files to a server, then you'll need to create a .json file that includes the credentials that are required to connect to these locations.

When you run the process, the hostname, port and username provided in the credentials .json file is used to find the matching connection in the Tableau flow file (tfl) and updated before running the process.

If connecting to a published data source, include hostname and contentUrl and port (80 for http and 443 for https) in the input connections. The hostname is required to find the matching connection in the Tableau flow file (tfl), and the contentUrl and port is used to establish the connection to the server.

**Note:** Skip this step if the flow connects to and outputs to local files, files stored on a network share or input files that use Windows Authentication (SSPI). For more information about Windows Authentication, see [SSPI Model](#) in the Microsoft online help.

If you plan to reuse the file, place it in a folder where it won't be overwritten by the Tableau Prep Builder install process.

The following table lists the credentials that you need to include in the .json file. Port ID and Site ID are optional if your connections don't require this information.

Input connections	Output location
<ul style="list-style-type: none"> <li>• hostname (Server name)</li> <li>• contentUrl (published data sources) Site ID. This appears after /site/ in the URL when you sign into Tableau Server or Tableau Online</li> <li>• port (Port ID)</li> <li>• username</li> <li>• Password</li> </ul>	<ul style="list-style-type: none"> <li>• serverUrl</li> <li>• contentUrl (Site ID. This appears after /site/ in the URL when you sign into Tableau Server or Tableau Online)</li> <li>• username</li> <li>• password</li> </ul>

The following example shows the syntax to use for the credentials .json file when connecting to a published data source.



**Note:** If the `inputConnection` or `outputConnection` uses the Default site, set `ContentUrl` to blank. For example `"contentUrl": ""`

```
{
  "inputConnections": [
    {
      "username": "jsmith",
      "contentUrl": "my-testing-site",
      "hostname": "https://my-server",
      "port": 443,
      "password": "passw0rd$"
    }
  ],
  "outputConnections": [
    {
      "serverUrl": "https://my-server",
      "contentUrl": "my-testing-site",
      "username": "jsmith",
      "password": "passw0rd$"
    }
  ]
}
```

The following example shows the syntax to use for the credentials `.json` file when connecting to two databases (MySQL and Oracle) and outputs files to a server that includes a Site ID.

```
{
  "inputConnections": [
    {
      "username": "jsmith",
      "hostname": "mysql.example.lan",
      "port": 1234,
      "password": "passw0rd"
    },
    {
```

```

    "username": "jsmith",
    "hostname": "Oracle.example.lan",
    "port": 5678,
    "password": "passw0rd"
  }
],
"outputConnections": [
  {
    "serverUrl": "http://MyServer",
    "contentUrl": "FinanceTeam",
    "username": "jsmith",
    "password": "passw0rd$"
  }
]
}

```

## Tips for creating your credentials file

Tableau Prep Builder uses information from the flow file and from the credentials .json file to run the flow when you have remote connections. For example, the database name for your remote connections and the project name for your output files come from the flow, and the server name and the log in credentials come from the .json file.

To avoid errors when running the flow, make sure your credentials file follows these guidelines:

- If using Tableau Prep Builder version 2018.2.2 through 2018.3.1, always include the "inputConnections" and "outputConnections" arrays even if the flow doesn't have remote connections for inputs or outputs. Just leave those arrays blank. If you are using Tableau Prep Builder version 2018.3.2 and later you don't need to include the blank array.
  - No remote input connection? Include this syntax at the top of the .json file

```

{
  "inputConnections": [
  ],

```

- No remote output connection? Include this syntax at the bottom of the .json file

```
"outputConnections": [
  ]
}
```

- No port ID for your input connection or the port is specified as part of the server name? If there is no port ID for your connection, don't include the "port":xxxx, reference in the .json file, not even "port": ". If the port ID is included in the server name, include the port ID in the host name. For example "hostname":  
"mssql.example.lan,1234"
- When referencing the "serverUrl": don't include a "/" at the end of the address. For example, use this "serverUrl": "http://server" not this "serverUrl":  
"http://server/".
- If you have multiple input or output connections include the credentials for each one in the file.
- If connecting to published data sources, make sure to include the hostname and contentUrl in the input connections.

## Run the flow

**Important:** The examples below include the name change for "Tableau Prep" version 2019.1.2 to "Tableau Prep Builder". If you are using an earlier version of the product use "Tableau Prep" instead.

1. Open the command prompt or terminal command prompt (Mac) as an Administrator.
2. Run one of the following commands.
  - The flow connects to local files or files stored on a network share and publishes to local files, files stored on a network share or uses Windows authentication:

**Note:** If connecting to or outputting to files stored on a network share, use the UNC format for the path: \\server\path\file name. It can't be password protected.

### Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep
Builder <version>\scripts"\tableau-prep-cli.bat -t
"path\to\[your flow file name].tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep
Builder version].app/Contents/scripts/./tableau-prep-cli
-t path/to/[your flow file name].tfl
```

- The flow connects to databases or publishes to a server:

### Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep
Builder <version>\scripts"\tableau-prep-cli.bat -c
"path\to\[your credential file name].json" -t "path\to\
[your flow file name].tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep
Builder version].app/Contents/scripts/./tableau-prep-cli
-c path/to/[your credential file name].json -t path/to/
[your flow file name].tfl
```

- The flow file or credentials file is stored on a network share (use the UNC format for the path: \\server\path\file name):

### Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep
Builder <version>\scripts"\tableau-prep-cli.bat -c
"\server\path\[your credential file name].json" -t
"\server\path\[your flow file name].tfl"
```

**Mac:** Map the network share to /Volumes in Finder so that it is persistent, then use /Volumes/.../[your file] to specify the path:

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep  
Builder version].app/Contents/scripts/./tableau-prep-cli  
-c /Volumes/.../[your credential file name].json -t  
path/to/[your flow file name].tfl
```

For examples of sample commands see [Syntax examples on page 344](#). For common errors and resolutions see [Common errors when using the command line to run flows on page 499](#).

## Run the flow with incremental refresh enabled (version 2020.2.1 and later)

If you don't have Tableau Prep Conductor enabled on your server to schedule your flow runs, you can run your flow using incremental refresh from the command line. Simply include the parameter `--incrementalRefresh` in your command line as shown in the example below.

For example:

### Windows

```
"\[Tableau Prep Builder install location]\Tableau Prep Builder  
<version>\scripts"tableau-prep-cli.bat --incrementalRefresh -t  
"path\to\[your flow file name].tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder ver-  
sion].app/Contents/scripts/./tableau-prep-cli --incre-  
mentalRefresh -t path/to/[your flow file name].tfl
```

If the input steps in your flow have incremental refresh enabled and the incremental refresh parameters are properly configured, Tableau Prep Builder will do the following:

- All inputs in the flow that have incremental refresh enabled will run all corresponding outputs using incremental refresh.
- If no input in the flow has incremental refresh enabled, all outputs will be run using full refresh. A message will show the refresh method details.
- If some inputs in the flow have incremental refresh enabled, the corresponding outputs will run using incremental refresh. The other outputs will be run using full refresh and a message will show the refresh method details.

For more information about configuring flows to use incremental refresh, see [Refresh Flow Data Using Incremental Refresh](#) on page 326

## Command options

If you want to view the help options, include `-h` in the command line.

Command options	Description	Notes
<code>-c, --connections &lt;arg&gt;</code>	The connection path to the credentials file.	Requires the path to where the credentials file is located.
<code>-d, --debug</code>	Debug the flow process.	Include this option to view more information to help debug a problem with refreshing the flow. Log files are stored in: My Tableau Prep Builder Repository\Command Line Repository\Logs
<code>-dsv, --disableSSLvalidation</code>	Disable SSL validation (Mac-	When running flows using the command line on the MacOS, a dialog may show asking for the keychain user and password. Starting with Tableau Prep Builder version 2019.3.2, you can pass in this additional parameter to disable this keychain dialog. For example: /Applications/Tableau\ Prep\ Builder\ [Tableau Prep Builder version].app/Contents/scripts/./tableau-prep-cli -

Tableau Prep Help

	cO-S)	dsv -c path/to/[your credential file name].json -t path/to/[your flow file name].tfl
-h, --help	View the help for syntax options.	<p>The help option or a syntax error shows the following information:</p> <pre>usage: tableau-prep-cli [-c &lt;arg&gt;] [-d] [-h] [-t &lt;arg&gt;]</pre> <p>-c, --connections &lt;arg&gt; Path to a file with all connection information</p> <p>-d, --debug This option is for debugging</p> <p>-dsv, --disableSslValidation Disable SSL validation</p> <p>-h, --help Print usage message</p> <p>-inc, --incrementalRefresh Run incremental refresh for all outputs that are configured to support it</p> <p>-t, --tflFile &lt;arg&gt; The Tableau Prep Builder flow file</p>
-inc, --incrementalRefresh	Run incremental refresh for all outputs	<p>Include this option to run incremental refresh for all inputs that are configured to use it. Incremental refresh enables Tableau Prep Builder to retrieve and process only new rows instead of all rows in a flow.</p> <p>The incremental refresh configuration settings on the input steps determine which flow outputs can be run incrementally. All other outputs will be run using a full refresh and a message will show the refresh method details.</p> <p>For more information about running flows using incremental refresh, see <a href="#">Refresh Flow Data Using Incremental Refresh</a> on page 326.</p>

	that are configured to use it.	
-t, -- tflFile <arg>	The .tfl flow file	Requires the path to where the .tfl flow file is located.

## Syntax examples

The command lines below show four different examples for running a flow using the following criteria:

- **Tableau Prep Builder version:** 2020.2.1  
**Important:** The examples below include the name change for Tableau Prep version 2019.1.2 to Tableau Prep Builder. If you are using an earlier version of the product use "Tableau Prep" instead.
- **Flow name:** Flow1.tfl
- **Flow location:** C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows
- **Credentials file name:** Flow 1.json
- **Credentials file location:** C:\Users\jsmith\Desktop\Flow credentials
- **Credentials file location stored on a network share:** \tsi.lan\files\Flow credentials

The flow connects to and publishes to local files

### Windows



## Tableau Prep Help

```
"\Program Files\Tableau\Tableau Prep Builder 2020.2.1\scripts"\tableau-prep-cli.bat -t "\C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows\Flow1.tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ 2020.2.1.app/Contents/scripts/./tableau-prep-cli -t /Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\ Repository/Flows.Flow1.tfl
```

The flow connects to and publishes to local files and uses the short form for incremental refresh

### Windows

```
"\Program Files\Tableau\Tableau Prep Builder 2020.2.1\scripts"\tableau-prep-cli.bat -inc -t "\C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows\Flow1.tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ 2020.2.1.app/Contents/scripts/./tableau-prep-cli -inc -t /Users/jsmith/Documents/My\ Tableau\ Prep\ Builder\ Repository/Flows.Flow1.tfl
```

The flow connects to databases and publishes to a server

### Windows

```
"\Program Files\Tableau\Tableau Prep Builder 2020.2.1\scripts"\tableau-prep-cli.bat -c "\C:\Users\jsmith\Desktop\Flow credentials\Flow1.json" -t "\C:\Users\jsmith\Documents\My Tableau Prep Builder Repository\Flows\Flow1.tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ 2020.2.1.app/Contents/scripts/./tableau-prep-cli -c
```

```
/Users/jsmith/Desktop/Flow\ credentials/Flow1.json -t /User-  
s/jsmith/Documents/My\ Tableau\ Prep\ Builder\ Repos-  
itory/Flows.Flow1.tfl
```

The flow publishes to a server and the credentials file is stored on a network share

### Windows

```
"\Program Files\Tableau\Tableau Prep Builder 2020.2.1\scripts"\t-  
ableau-prep-cli.bat -c "\\tsi.lan\files\Flow cre-  
dentials\Flow1.json" -t "\C:\Users\jsmith\Documents\My Tableau  
Prep Builder Repository\Flows\Flow1.tfl"
```

### Mac

```
/Applications/Tableau\ Prep\ Builder\ 2020.2.1.ap-  
p/Contents/scripts/./tableau-prep-cli -c /Volumes/files/Flow\ cre-  
dentials/Flow1.json -t /Users/jsmith/Documents/My\ Tableau\ Prep\  
Builder\ Repository/Flows.Flow1.tfl
```

# Version Compatibility with Tableau Prep

Tableau Prep Builder frequently releases new versions, generally every month. If new features or connectors are introduced in a new version and you are working in an older version of Tableau Prep Builder, compatibility may be an issue if you try to open a flow.

Similarly, if you publish flows to Tableau Server or Tableau Online to schedule them to run using Tableau Prep Conductor and your flows include new features or connectors that aren't supported in your version of Tableau Server or Tableau Online, you can run into compatibility errors that might prevent you from scheduling and running your flows.

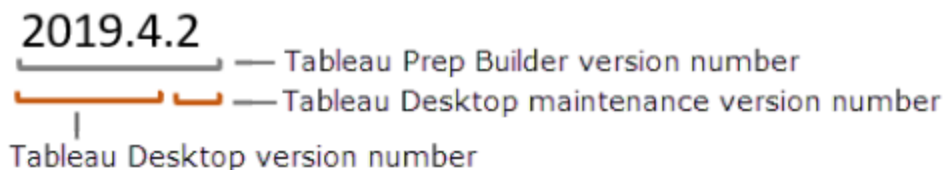
**Note:** To download a specific version of Tableau Prep Builder, open the [Product Downloads and Release Notes](#) page and select **Tableau Prep Builder** from the list on the left side of the page.

## Finding our version

Follow the instructions below to find your product version number.

### Tableau Prep Builder

Tableau Prep Builder release versions are numbered like Tableau Desktop and Tableau Server maintenance versions for example 2019.4.2, except Tableau Prep Builder doesn't release maintenance versions.

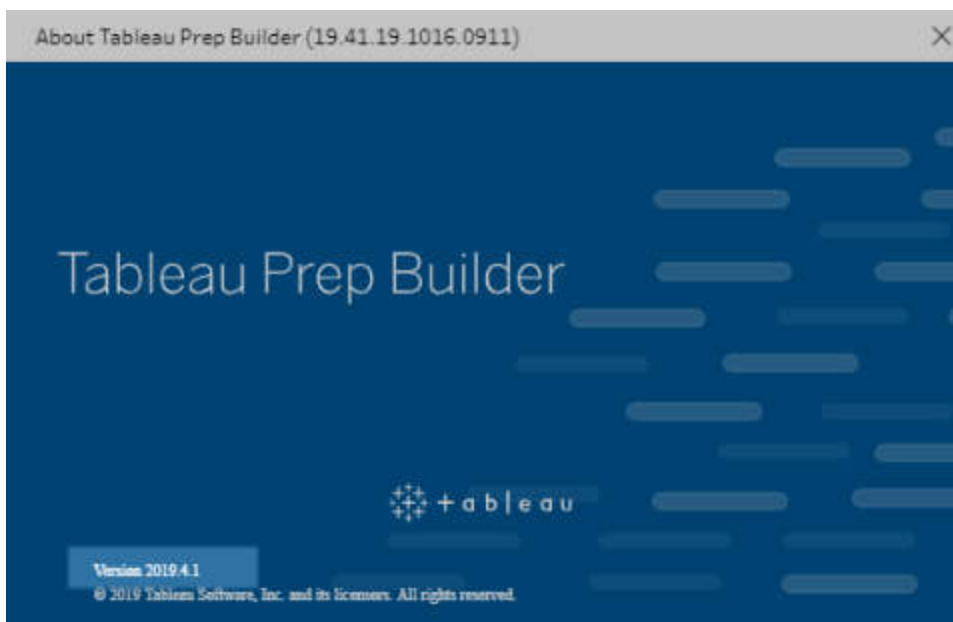


Each release represents the latest version of the product including any new features and updates. So Tableau Prep Builder version 2019.4.2 might not be fully compatible with Tableau Server version 2019.4.2, but would be compatible with the next major release version of Tableau Server, for example 2020.1.

To find the release version for your product, open Tableau Prep Builder, then in the top menu do one of the following:

- **Windows:** In the top menu, click **Help > About Tableau Prep Builder** or **About Tableau Prep**, depending on your version.
- **Mac:** In the top menu, click **Tableau Prep Builder > About Tableau Prep Builder** or **Tableau Prep > About Tableau Prep**, depending on your version.


The release number displays in the lower left corner of the dialog.



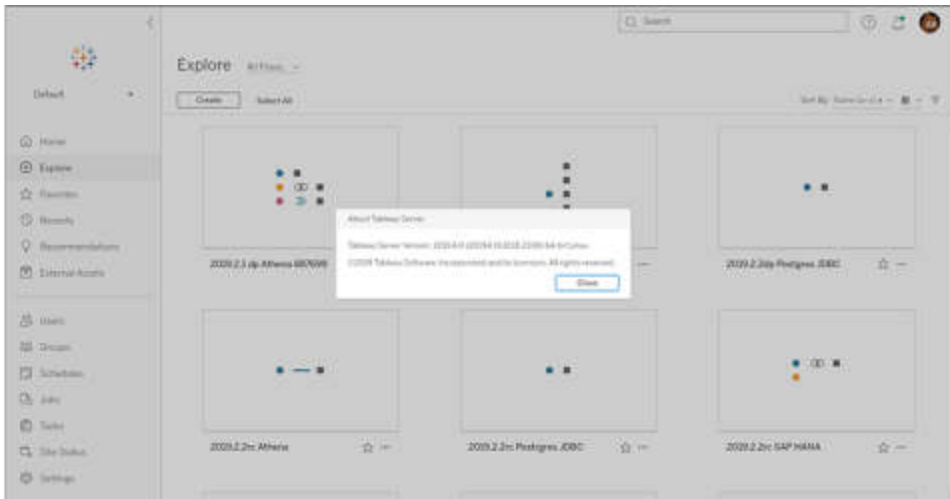
## Tableau Server

Tableau Prep Conductor was introduced as part of the Data Management Add-on in Tableau Server version 2019.1. To publish and schedule flows to run on Tableau Server, you must be using Tableau Server version 2019.1 or later and Tableau Prep Conductor must be enabled.

To find your version of Tableau Server, open Tableau Server in your web browser. In the top

menu bar click the information icon  in the top right corner and select **About Tableau Server**. A dialog opens that tells you which version of Tableau Server you are using. For information about how to enable Tableau Prep Conductor, see [Step 2: Configure Flow Settings for your Tableau Server](#) on page 387.

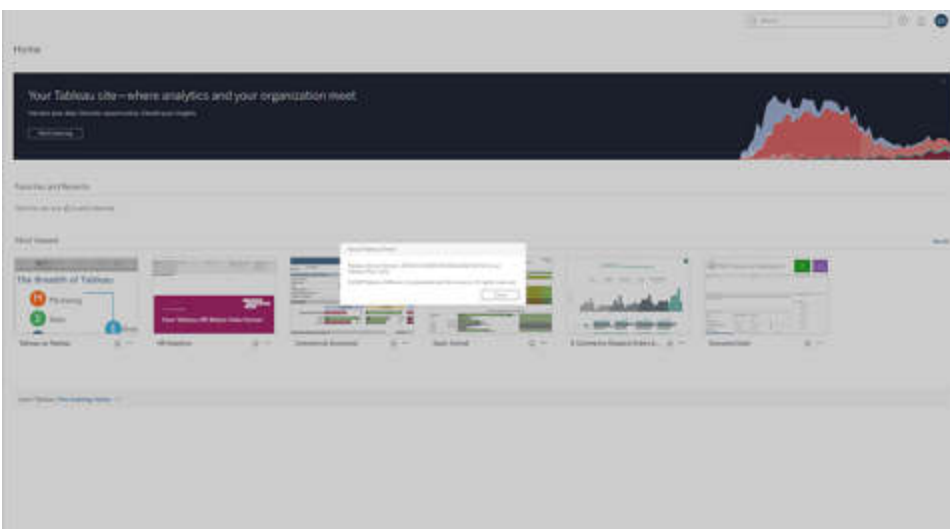
## Tableau Prep Help



## Tableau Online

Tableau Prep Conductor was introduced as part of the Data Management Add-on in Tableau Online version 2019.3. To publish and schedule flows to run on Tableau Online, you must be using Tableau Online version 2019.3 or later and Tableau Prep Conductor must be enabled .

To find your version of Tableau Online, open Tableau Server in your web browser. In the top menu bar click the information icon in the top right corner and select **About Tableau Online**. A dialog opens that tells you which version of Tableau Online you are using. For information about enabling Tableau Prep Conductor, see [Enable Tableau Prep Conductor on your Tableau Online Site](#) on page 397.



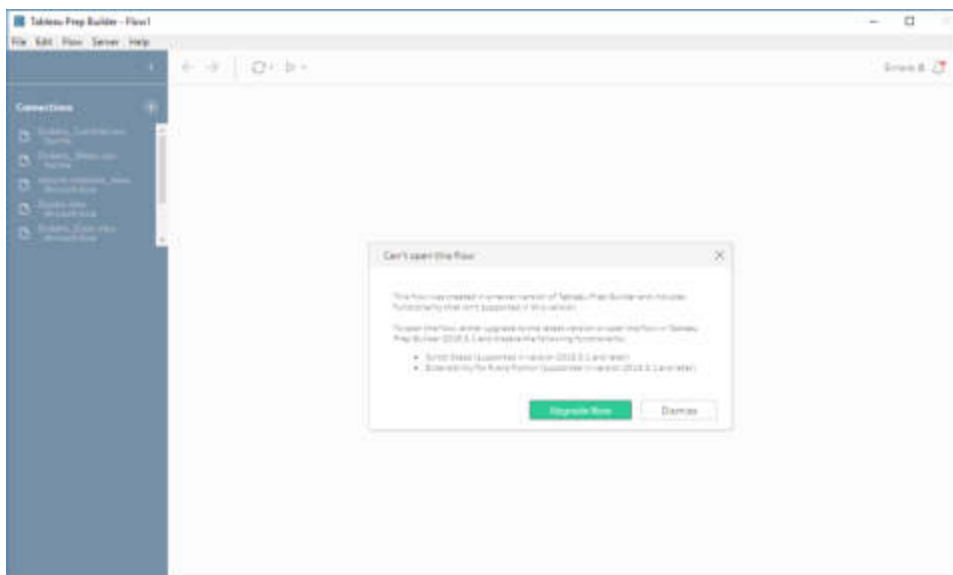
# Compatibility between different versions of Tableau Prep Builder

Generally, a new version of Tableau Prep Builder can open flows created in an older version. However, compatibility issues can occur when you try to open a flow between newer and older versions of Tableau Prep Builder or even when opening flows in the same version of Tableau Prep Builder using different computers.

For example:

- The flow includes input connectors or features that aren't supported in the version where the flow is opened.
- The machine that you use to open the flow doesn't have the required input connectors installed or has a driver version for the connector that isn't compatible. Tableau Prep Builder requires 64-bit drivers to be installed to work with flow input connectors.

If compatibility is an issue, when you try to open the flow, the flow may open but contains errors or the flow won't open at all and you receive an error message. In the example below, the flow won't open and an error message displays and lists the incompatible features and options for resolving the issue.



## Fix compatibility issues with Tableau Prep Builder

To fix compatibility issues, try one of the following:

- Upgrade to the latest version of Tableau Prep Builder.

Click the update button on the bottom of the Discover pane to download the latest version of the product and follow the instructions to [Install Tableau Prep Builder](#) in the Tableau Desktop and Tableau Prep Builder Deployment Guide. If you don't have access to the update button on the Discover pane, instructions about how to download the latest version of the product are included in the **Install Tableau Prep Builder** topic.

- Make sure your computer is compatible with Tableau Prep Builder. For example, make sure that you have the 64-bit drivers installed for the connectors used by the flow. To install drivers, see the [Driver Download](#) page.
- Open a copy of the flow that has the incompatible features removed.

## Compatibility between different versions of Tableau Prep Builder and Tableau Server

Publishing from a newer version of Tableau Prep Builder to an older version of Tableau Server can result in compatibility issues. For example, new features added in Tableau Prep Builder version 2019.4.2 may not be compatible with Tableau Server version 2019.4.2 but would be compatible with Tableau Server version 2020.1 and any later major versions of Tableau Server, such as version 2020.2.

In Tableau Server, Tableau Prep Conductor detects the features that are included in a flow when it has been published. If it finds features that it doesn't support, the flow can still be published to Tableau Server, but the flow can't be run, scheduled, or added to a task. Tableau Online is updated automatically on a regular basis, so is generally compatible with all versions of Tableau Prep Builder.

If you have an older version of Tableau Server, you can still run incompatible flows manually in Tableau Prep Builder or using the command line. For more information about using this process, see [Refresh flow output files from the command line on page 334](#).

### Detect incompatible features

Depending on the version of Tableau Prep Builder you are using, you can spot incompatible features in different ways.

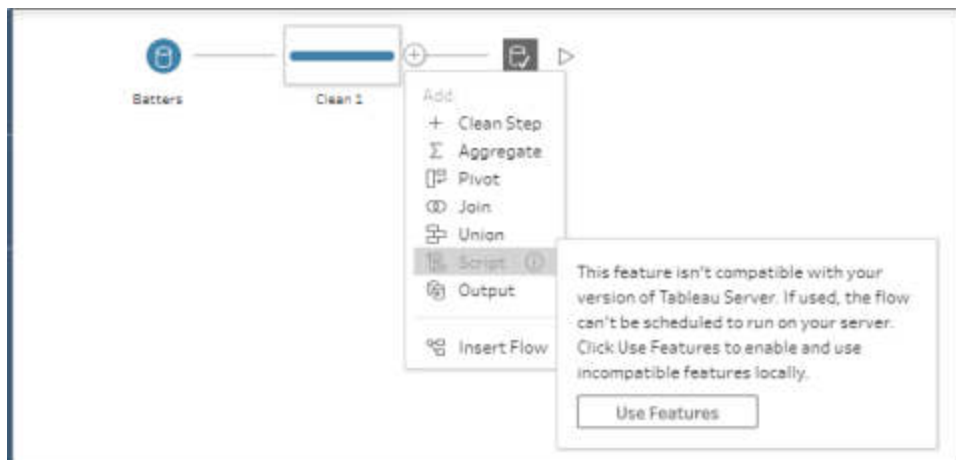
## Tableau Prep Builder (version 2020.1.1 and later)

Sign into Tableau Server, and Tableau Prep Builder will detect and disable incompatible features for you. Any features that aren't compatible will show as grayed out. If you want to still use the feature and run the flow manually or from the command line, you can enable it from the menu.

**Note:** Starting in Tableau Prep Builder version 2020.1.4, once you sign into your server, Tableau Prep Builder remembers your server name and login information when you close the application so that the next time you open the application you are already logged into your server.

1. Hover over the disabled feature to see if it's disabled because it isn't compatible with your server version, then click the **Use Features** button. This option is available in the Flow pane and from the menus in the Profile pane, Results pane and data grid.

**Note:** Features can be disabled for other reasons, such as data updates being paused or if the option isn't available for a particular step or data type.



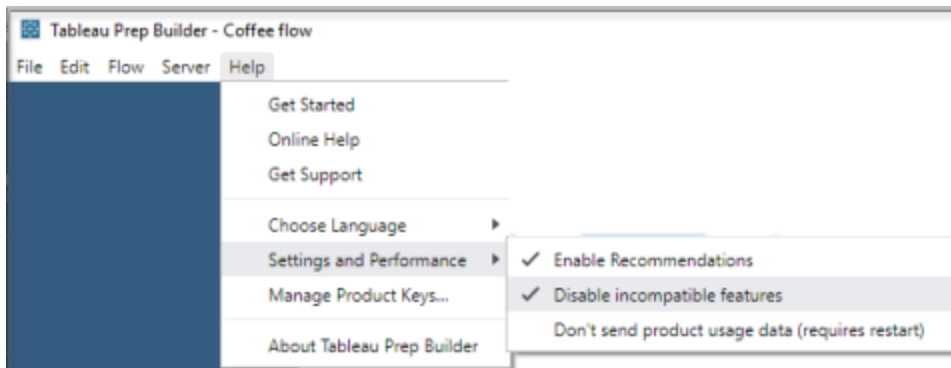
2. The selected feature is applied and all incompatible features are enabled and available to use. Incompatible features are flagged with a warning so that you can easily find and remove them if you want to run the flow using a schedule in your version of Tableau Server.

To disable this feature entirely and enable all incompatible features, do the following:



## Tableau Prep Help

1. From the top menu, select **Help > Settings and Performance > Disable Incompatible Features**.

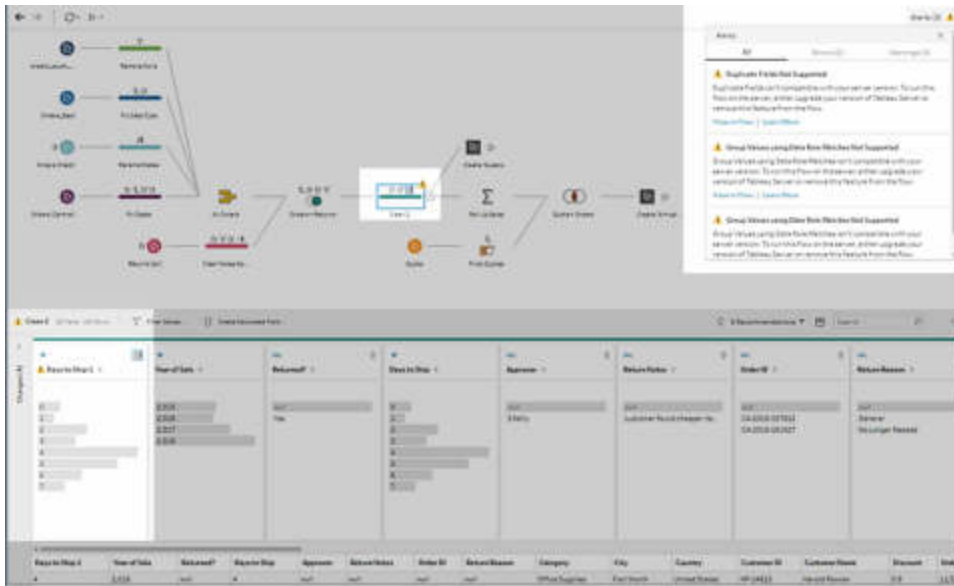


2. Select **Disable Incompatible Features** to clear the check mark next to this option. To enable the feature again, select **Disable Incompatible Features**. This option should be enabled by default.

## Tableau Prep Builder (version 2019.3.1 and later)

As you build your flow, Tableau Prep Builder can detect incompatible features as you add them and flags these features with an alert icon. You must be signed into your server to see these alerts. This new alert system helps you quickly identify incompatible features in your flow so you can decide whether to keep the feature in your flow or remove it.

Hover over alerts in the Flow pane to view information about the incompatible feature, or use the new alert center to see more details. In the alert center, click the **View in Flow** link to navigate directly to the step, annotation, field or change that triggered the warning.

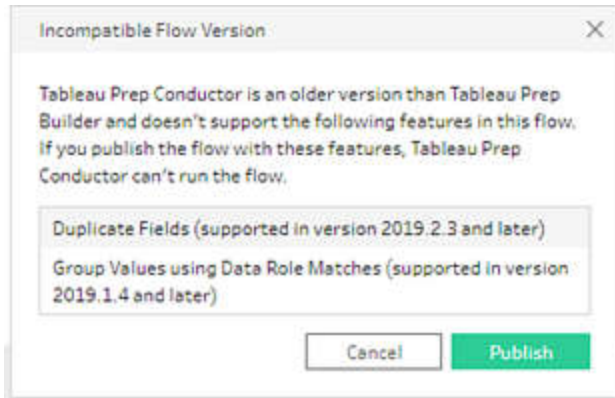


## Tableau Prep Builder (all versions)

If you publish a flow with incompatible features, the following message is displayed and lists the features that aren't supported in the version of Tableau Server that you are signed into. In Tableau Prep Builder version 2019.2.3 and earlier, this is the only way to see which features are incompatible in your flow.

**Note:** The error message lists the Tableau Prep Builder version when the feature was introduced. Tableau Prep Builder doesn't release maintenance versions, so for the feature to be compatible, Tableau Server must be running the next major release version. In the example below, the Duplicate Fields feature was introduced in Tableau Prep Builder version 2019.2.3 so it won't be compatible with the 2019.2.3 Tableau Server maintenance release version. Instead it would be compatible with the next major release for Tableau Server, version 2019.3.

## Tableau Prep Help



If you continue to publish the flow, publishing will complete successfully. However, when you open the flow in Tableau Server or Tableau Online, you will see the following message:



To schedule and run the flow in Tableau Server, you can do one of the following:

- Look for the latest major release of **Tableau Server** that is compatible with the version of Tableau Prep Builder that you are using. For example, Tableau Prep Builder releases a new build each month. Features released in Tableau Prep Builder version 2019.2.3 may not be supported in Tableau Server version 2019.2.2 or 2019.2.3, but would be supported in the next major release for Tableau Server, version 2019.3.

Tableau Online is updated automatically on a regular basis, usually every quarter. Test your flow first, to make sure it is compatible with your current version of Tableau Online before publishing.

- Before publishing the flow, remove the incompatible features from the flow, then publish the flow.

- If it is already published your flow to Tableau Server, download the flow and remove the features or create the flow in an older version of Tableau Prep Builder using only the features available in that version.

**Note:** To download a specific version of Tableau Prep Builder, open the [Product Downloads and Release Notes](#) page and select **Tableau Prep Builder** from the list on the left side of the page.

## Fixing compatibility issues

If the flow is already published to Tableau Server, try the steps below to remove the incompatible features using your current version of Tableau Prep Builder. After you remove the features and no longer see the version incompatibility message or warnings, republish your flow to Tableau Server or Tableau Online and schedule it using the Tableau Prep Conductor.

**Note:** Tableau Prep Conductor is part of the Data Management Add-on. It must be enabled in Tableau Server or Tableau Online to run flows using the scheduling functionality. For more information about the Data Management Add-on, see [Tableau Data Management Add-On](#). For more information about enabling Tableau Prep Conductor in Tableau Server or Tableau Online, see see [Step 2: Configure Flow Settings for your Tableau Server](#) on page 387 or [Enable Tableau Prep Conductor on your Tableau Online Site](#) on page 397.

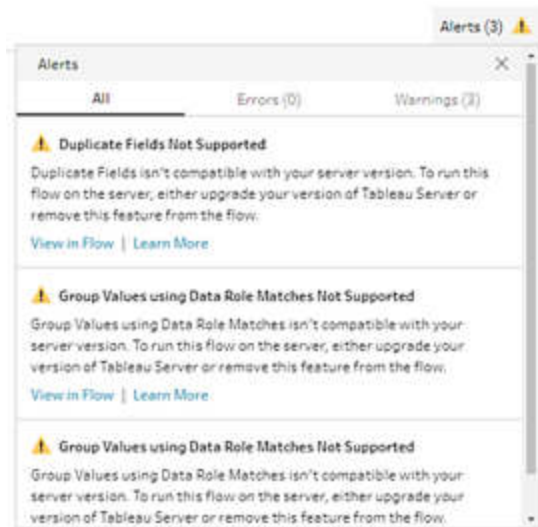
## Identify incompatible features

If you are working in Tableau Server, it doesn't currently list the incompatible features in your flow. To identify the list of features to remove from the flow, you need to open the flow in Tableau Prep Builder then find and remove them in your flow.

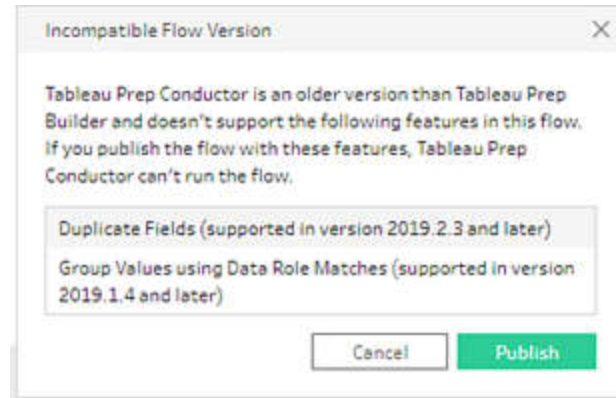
1. Open your flow. If you are in Tableau Prep Conductor, from the **More actions** ... menu, click **Download** to download and open the flow in Tableau Prep Builder or simply open the flow in Tableau Prep Builder.
2. If you downloaded the flow, click on the downloaded flow to open it.
3. Depending on your version, do one of the following:

- **Version 2019.3.1 and later:** From the top menu select **Server >Sign In**. Make sure you select the same server that is incompatible with the flow. Any incompatible steps, annotations, fields, or changes should be marked with an alert icon.

In the top right corner of the flow pane, click **Alert** to view the details for each incompatible feature. Click **View in Flow** to navigate to the incompatible feature to take action.



- **Version 2019.2.3 and earlier:** From the top menu select **Server >Publish Flow**. If you need to sign into the server again, make sure you select the same server that is incompatible with the flow. A warning dialog opens that lists the features that are not compatible with your server version. Note the features so you can identify and remove them from the flow. Then click **Cancel** to close the dialog.



4. From the top menu, click **File > Save As** to save a copy of your flow. Use the options in the following sections to remove incompatible features from your flow.

## Remove incompatible features from the flow

You can use various methods to find and remove features from your flow. This section shows some options to help you resolve incompatibility errors.

## Incompatible data sources

If the data source isn't compatible, for example a new connector was added that isn't yet supported in Tableau Prep Conductor, you'll need to connect to a data source that is supported.

To change your data connection do the following:

1. In the **Connections** pane, click the drop-down arrow on the connector and select **Remove** to remove the data source connection.  
This will temporarily put your flow in an error state.
2. Connect to a compatible data source and drag it to the Flow pane. To connect the new Input step with your flow, drag the step to the first step in the flow where you want to connect it and drop it on the connecting step on the **Add** option.



3. Repeat these steps to replace any other connectors. Then save your flow and republish it.

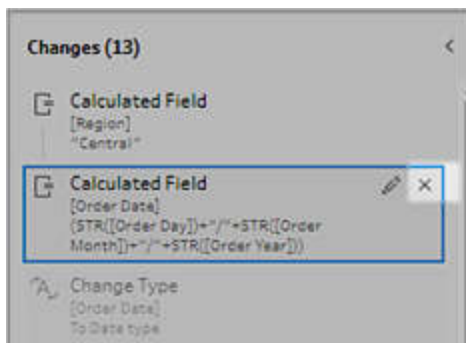
## Incompatible features

To remove incompatible features you'll need to find the steps where the features were used and remove them. You can follow the instructions in [Identify incompatible features on page 356](#) to locate the incompatible features.

1. If the feature is a step type, in the Flow pane click on the step where the feature is used. Right-click or Ctrl-click (MacOS) on the step and select **Remove**.
2. If the feature is a cleaning operation, in the Flow pane click on the step where the feature is used. You can hover over the annotations in the **Flow** pane or in the **Profile** or **Results** panes to see a list of changes.

**Note:** In Tableau Prep Builder version 2019.1.3 and later you can hover on the icon that represents the change you are looking for over a step in the Flow pane or in the profile card then select the annotation from the list of changes. The change is highlighted in the **Changes** pane, **Profile** or **Results** pane and data grid.

3. Open the **Changes** pane if needed, and select the change that matches the feature you need to remove. Click on the change to select it and click **Remove** to delete it from the flow.



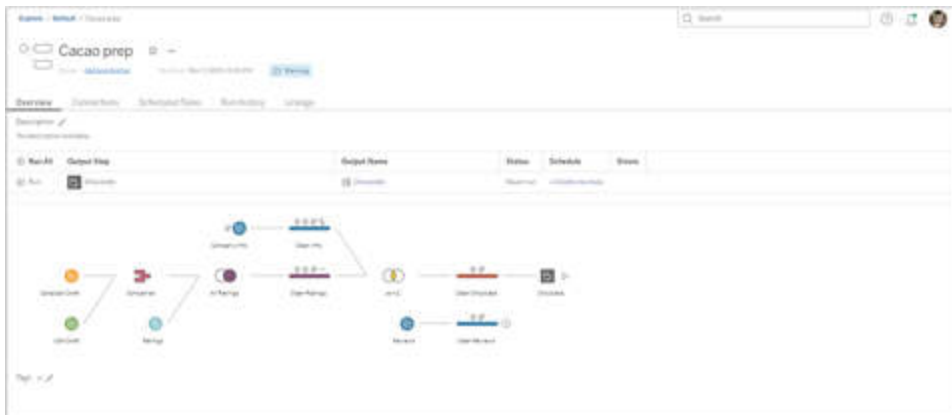
4. Repeat these steps to replace any other features. Then save your flow and republish it.

# Keep Flow Data Fresh

You've built your flow and cleaned your data, but now you want to share your data set with others and you want to keep that data fresh. You can always run your flow in Tableau Prep Builder and publish an extract to Tableau Server, but now there's a better way.

Meet Tableau Prep Conductor, part of the new Data Management Add-on to Tableau Server version 2019.1 and later. If you add this option to your Tableau Server or Tableau Online installation, you can use Tableau Prep Conductor to run your flows automatically and keep your flow data fresh.

**Note:** If Tableau Catalog is installed, you can also see data quality warnings about your flow input data and view the upstream and downstream impact of fields in your flow on the new **Lineage** tab. For more information about Tableau Catalog, see [About Tableau Catalog](#) in the Tableau Server online help.



With Tableau Prep Conductor you can do the following:

- Configure your Server or Site to use Tableau Prep Conductor
  - Enable or disable Tableau Prep Conductor for individual sites
  - Set up email notifications for flow failures for flows that are run either on-demand or using a schedule
  - Configure flow timeout settings
- Publish a flow from Tableau Prep Builder to Tableau Server or Tableau Online



- Upload data files or connect directly to your files or databases. If connecting to databases, you can either embed the database credentials or require a user prompt.

**Note:** If you connect to data files through a direct connection or publish your flow output to a file share, the files need to be in a location that Tableau Server can access. For more information see [Step 4: Safe list Input and Output locations](#) on page 393.

- Select from a project hierarchy when publishing your flows
- Enter tags and a description to help others find your flow
- Manage the flow
  - Set permissions
  - Move the flow to a different project
  - Change the flow owner
  - Add or edit tags
  - View the version history and select from the list to restore the flow to a previous version
  - Mark a flow as a favorite and add it to your favorites list
  - Edit an input connection and update credentials
  - View data sources created from a flow and link back to the flow that created it
- Create schedules to run your flows
- Add scheduled tasks to run the flow and select which flow outputs to update
- Run the flow on demand without a schedule
- Monitor the flow
  - Set up email alert notifications
  - View errors
  - Monitor and restart flows that have been suspended
  - View run history
  - Use Admin views

# About Tableau Prep Conductor

Tableau Prep Conductor enables you to seamlessly publish flows from Tableau Prep Builder to Tableau Server or Tableau Online, then leverage the scheduling and tracking functionality available in those applications to run your flows automatically to update the flow output. Tableau Prep Conductor is part of the Tableau Data Management Add-on introduced in Tableau Server version 2019.1 and must be enabled for the features to be available.

Publishing flows is similar to publishing data sources and workbooks with Tableau Desktop. You can package files with the flow or specify a direct connection to data sources to update the flow input as data changes. If your flow connects to databases, specify the authentication type and set credentials to access the data.

You can also publish a flow to share it with others. For example, publish an incomplete flow to Tableau Server or Tableau Online and then download it to another computer to continue working on it. Create a flow with only Input steps (that are properly configured) and share it with co-workers who can then download the flow to their computers and create and publish their own flows.

Flows can only be run in Tableau Server or Tableau Online if the flow includes output steps and the flow has no errors or incompatible features. For more information about publishing a flow, see [Publish a Flow to Tableau Server or Tableau Online on page 398](#). For more information about incompatibility, see [Version Compatibility with Tableau Prep on page 347](#).

Keeping track of the health of your flows is easy. With Tableau Prep Conductor, you can:

- View and monitor the details about your flow, including recent activity in the **Content** pages.
- View the results of the flow runs and any errors in the **Run History** tab.
- Use **Administrative Views** to monitor server and site activity including a new view that tracks flow performance history.
- View detailed alerts for failed flow runs.
- Set up email notification alerts to send emails to flow owners notifying them when the flow failed to run and why.

For more information about setting up alerts, see [Monitor Flow Health and Performance on page 429](#).

If a flow fails to run due to connectivity errors, you can fix the error right in Tableau Server or Tableau Online by editing the connection. If a flow fails to run due to errors in one or more of the flow steps, such as a calculation that isn't valid, simply download the flow to Tableau Prep Builder, fix the problems and then republish the flow to Tableau Server or Tableau Online to pick up where you left off.

## Enabling Tableau Prep Conductor

Tableau Prep Conductor is available with the Data Management Add-on and can be installed with Tableau Server or Tableau Online or added to an existing installation of Tableau Server or Tableau Online version 2019.1 or later. For more information about how to install and enable Tableau Prep Conductor, see [Enable and Configure Tableau Prep Conductor on Tableau Server on page 368](#) or [Enable Tableau Prep Conductor on your Tableau Online Site on page 397](#).

Tableau Prep Conductor is designed to work with Tableau Prep Builder version 2019.1.2 and later. You must install or upgrade Tableau Prep Builder to use version 2019.1.2 or later to be able to publish flows to Tableau Server or Tableau Online. For information about how to install Tableau Prep Builder, see [Install Tableau Desktop or Tableau Prep Builder from the User Interface](#).

## Getting to know the flow workspace

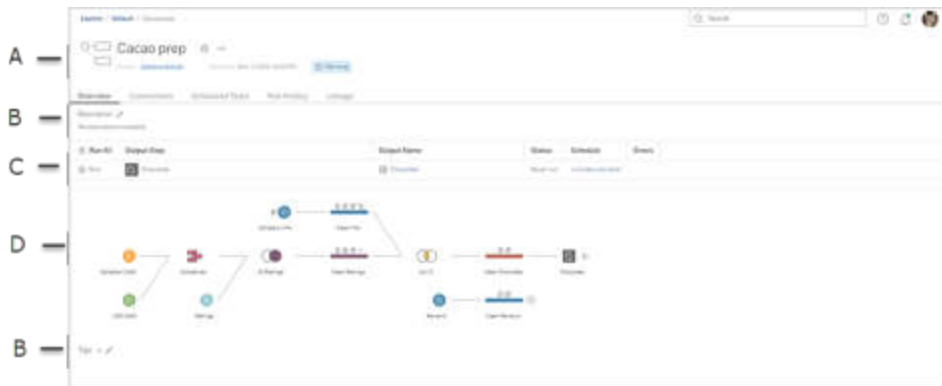
After you publish your flow you can schedule tasks in Tableau Server or Tableau Online to automatically run your flow on a regular basis to keep your output data fresh.

Tableau Prep Conductor leverages much of the same functionality for managing flows that you might see when managing workbooks or data sources from Tableau Desktop in Tableau Server or Tableau Online. For example, just like extract refreshes, scheduled flow tasks and on-demand flow runs are queued as background tasks. But when it comes to working with flows, there are a few differences.

### Flow Overview page


The flow **Overview** page is the main landing page where you can view data about your flow and schedule, monitor and maintain running the flow.

Open the flow **Overview** page by clicking on a flow in your list. You can navigate there from **Content > Explore > All Flows** or by opening the project that contains your flows.



- A. The header lists the name of the flow, the flow owner and the date that the flow was last modified.


Add a flow to your favorites, or from the **More actions** ... menu you can also download the flow, set permissions, change the flow owner, restore previous flow versions, and more.

- B. View and edit the flow description and set tags to help others find the flows they are looking for.
- C. View the output steps for a flow along with the status of the last update, any schedule the output is assigned to, and any errors from the last flow run. You can also click the **Run**  button to run all output steps or individual output steps on-demand.

If the flow has errors, the flow run will fail. Connectivity errors can be resolved directly in Tableau Server or Tableau Online by navigating to the **Connections** tab for the flow and editing the input connections. To resolve any other flow errors, from the **More actions** ... menu select **Download** to download and open the flow in Tableau Prep Builder. Edit the flow then republish it and try running the flow again.

After a flow has run successfully, outputs that are data sources become links that you can click to open the **Data Source** page to view more information about the data source or edit the flow input connection.


In the **Schedule** field, view the scheduled tasks that the output step is assigned to. A flow output can be assigned to one or more tasks.

If no schedule has been assigned yet, click **Create new task** to add the output step to a schedule. To immediately run the flow to update a specific output step, click the **Run**  button on the left-hand side of the row.

D. View an image of the flow.

## Flow Connections page

View both the input and output locations for a flow, connection types, authentication settings, input and output steps and any connectivity errors. You can set authentication settings when publishing a flow. For more information, see [Publish a flow on page 402](#).

For database input types, click the **More actions**  menu for an input connection to edit the connection and change the server name, port, user name and password.



## Flow Scheduled Tasks page

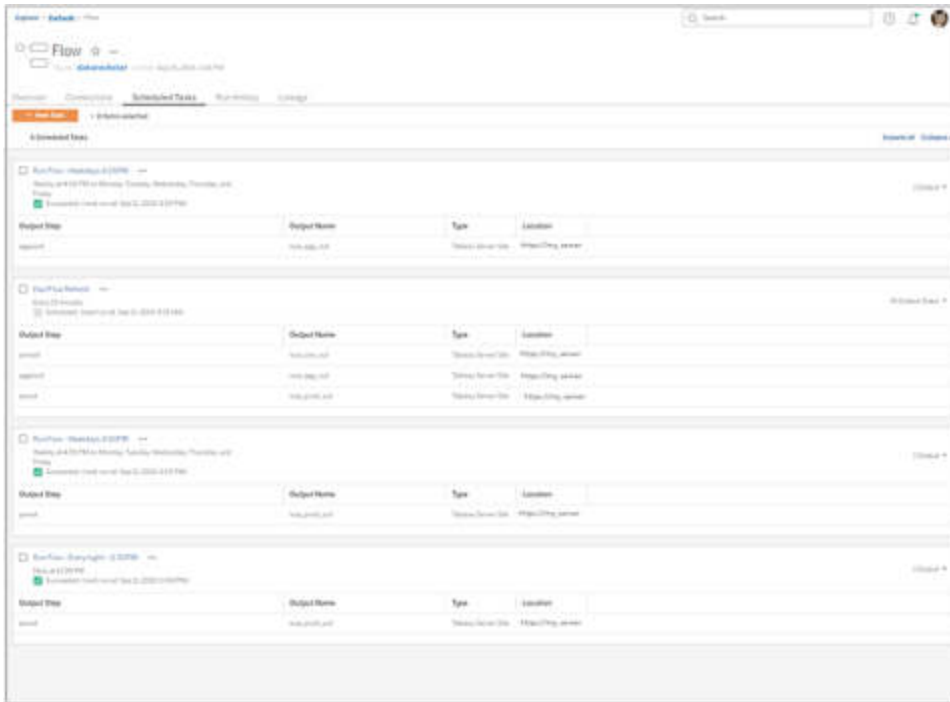
View any schedules that the flow is assigned to and the outputs that are included in those schedules. Click the schedule link at the top of the task card to open the **Schedules** page and see a list of flows that are assigned to that schedule. For more information about assigning flows to a schedule, see [Schedule a Flow Task on page 417](#).

You can also add new tasks or manage existing ones from this page. To take action on an existing task, select the check box on a task card then click the **Actions** drop-down menu to run, edit, or delete the task.

In Tableau Server, if the flow fails to run after a configured number of consecutive attempts, the flow is automatically suspended. In Tableau Online, if a flow task fails to run after 5 consecutive attempts, the flow is automatically suspended.

You can see that status on the **Overview** tab as well as this tab. You can resume suspended tasks from this menu.

For information about how to set the threshold for suspended flow tasks, see [Step 5: Optional Server Configurations on page 396](#). For more information about suspended flow tasks, see [View and resolve errors on page 430](#).



## Schedules page

On the **Schedules** page, you can view the flows assigned to a schedule and the details about the flow runs. You can run the schedule on-demand and run all flows assigned to it. You can also select one or more flows, then use the **Actions** menu to change the flow schedule or priority, delete selected flows from the schedule or resume suspended flows.

For information about how to set up a schedule, see [Step 3: Create Schedules for Flow Tasks on page 392](#).

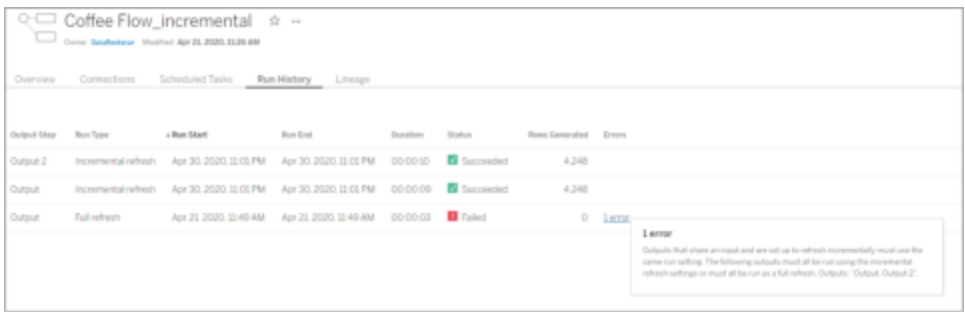


## Flow Run History

See, search, and sort through a list of historical runs for a flow. This page also includes details about the flow run such as run type, duration and number of rows that were generated.

If the flow output has an error, hover over the error to view the messages. If applicable, click the **Go to Connections** link in the error message to navigate to the **Connections** page to fix connectivity errors. If shown, click the **Download the Flow** link to download the flow to fix flow errors in Tableau Prep Builder, then republish the flow to continue to manage it using Tableau Prep Conductor.

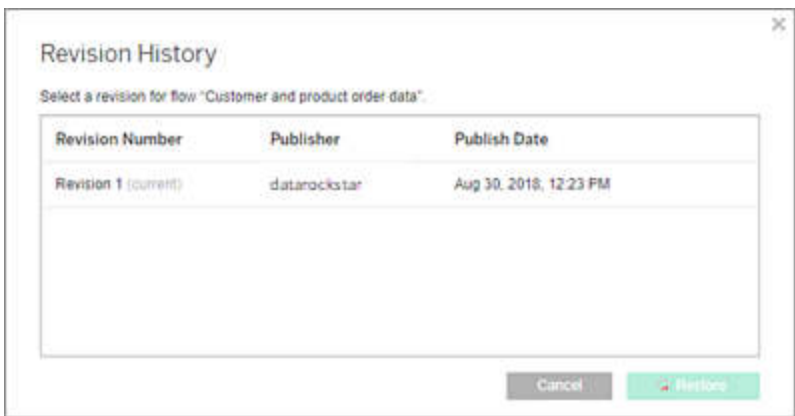
**Note:** The run history for a flow will persist unless the flow is deleted.



Output Step	Run Type	Run Start	Run End	Duration	Status	Rows Generated	Errors
Output 2	Incremental refresh	Apr 30, 2020 11:02 PM	Apr 30, 2020 11:02 PM	00:00:30	Successful	4,248	
Output	Incremental refresh	Apr 30, 2020 11:02 PM	Apr 30, 2020 11:02 PM	00:00:09	Successful	4,248	
Output	Full refresh	Apr 21, 2020 12:49 AM	Apr 21, 2020 12:49 AM	00:00:03	Failed	0	<a href="#">View Error</a>

## Flow Revision History

If you need to revert a flow to a previous version, from the **More actions** ... menu for the flow, select **Revision History**. On the **Revision History** dialog, select the flow version from the list that you want to revert to.



Revision Number	Publisher	Publish Date
Revision 1 (current)	datarockstar	Aug 30, 2019, 12:23 PM

## Who can do this

Server Administrators can activate Data Management Add-on license keys.

Server administrators can enable Tableau Prep Conductor.

# Licensing Tableau Prep Conductor

Tableau Prep Conductor is licensed through the Data Management Add-on.

For more information, see:

- [License the Data Management Add-on for Tableau Server](#)
- [License the Data Management Add-on for Tableau Online](#)

# Enable and Configure Tableau Prep Conductor on Tableau Server

Tableau Prep Conductor is supported only on Tableau Server versions 2019.1 or later.

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis, which is User-Based or Core-Based. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the [EULA Documentation](#).

For more information on how Tableau Prep Conductor licensing works, see [Licensing Tableau Prep Conductor for Tableau Server](#).

## Server Topology

When you install Tableau Server and enable Tableau Prep Conductor, using the Data Management product key, Tableau Prep Conductor is automatically enabled by default by the setup program.

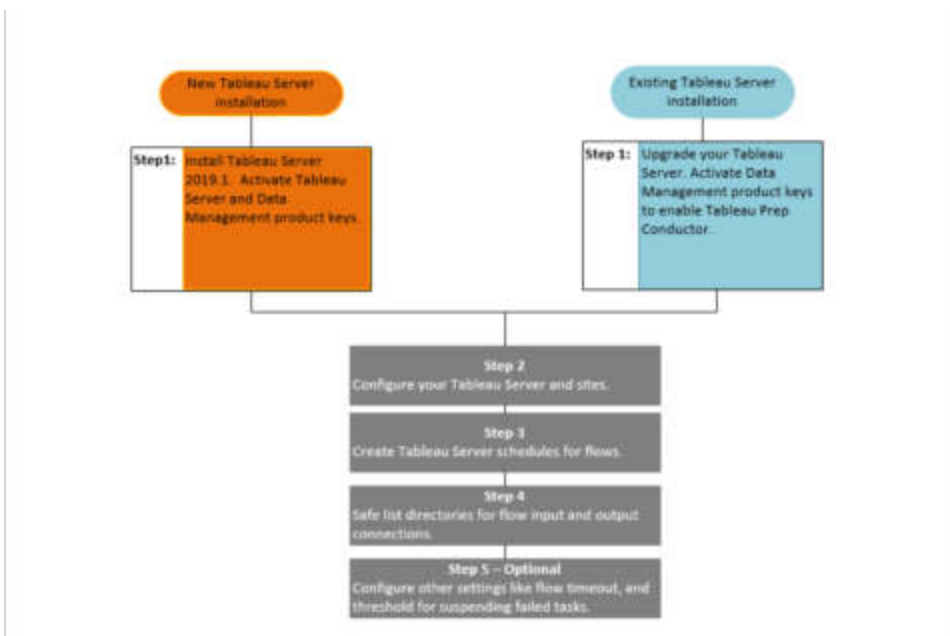
For multi-node installations, by default, one instance of Tableau Prep Conductor is enabled on any node that has backgrounder installed. In the example below, Tableau Prep Conductor is enabled on node 2 and 3 where the backgrounders are also enabled, but not on node 1,4, and 5.



# Tableau Prep Help

+ a b l e a u					
STATUS MAINTENANCE CONFIGURATION					
Process	node1	node2	node3	node4	node5
Gateway	✓				
Application Server	✓				
Interactive Microservice Container	✓				
VizQL Server	✓	✓	✓	✓	
Cache Server	✓	✓			
Cluster Controller	✓	✓	✓	✓	✓
Search & Browse	✓				
Backgrounder		✓	✓	✓	✓
Background Microservice Container		✓	✓	✓	✓
Data Server	✓	✓			
Data Engine	✓	✓	✓		
File Store	✓				
Repository	✓				
Tableau Prep Conductor		✓	✓		
Ask Data	✓				
Elastic Server	✓				

Below is a visual representation of that work-flow:



## Next step:

New Tableau Server Installations: [Step 1 \(New Install\): Install Tableau Server with Tableau Prep Conductor](#) below

Existing Tableau Server Installations: [Step 1 \(Existing Install\): Enable Tableau Prep Conductor](#) on page 375

## Who can do this

Server administrators can install Tableau Server and enable Tableau Prep Conductor.

Server-level settings can be configured by Tableau Server administrators, and site-level settings can be configured by Tableau Server and Site administrators.

## Step 1 (New Install): Install Tableau Server with Tableau Prep Conductor

This topic describes how to Tableau Prep conductor on a new installation of Tableau Server.

Tableau Prep Conductor is supported only on Tableau Server versions 2019.1 or later.

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the [EULA Documentation](#).

## Before you install

The recommended topology for a production Tableau Server installation is a dedicated node for running flows. If you are currently planning to have a single node Tableau Server installation it is recommend that you add a second node and dedicate it to run flows.

- Review the hardware recommendations for Tableau Server and Tableau Prep conductor.
  - [Minimum Hardware Requirements and Recommendations for Tableau Server installation on Windows](#).
  - [Minimum Hardware Requirements and Recommendations for Tableau Server installation on Linux](#).

## Install Tableau Server and enable Tableau Prep Conductor

Use the instructions provided in the following topics to install Tableau Server.

[Windows: Install Tableau Server](#)

[Linux: Install Tableau Server topic](#)

When you get to the **Activate** step, use the Tableau Server product keys to activate Tableau Server.

All product keys are available through the [Customer Portal](#).

### Configure public gateway settings

If your Tableau Server is set up with one of the following:

- Load balancer to distribute requests across gateways.
- Reverse proxy to authenticate external (internet) client requests and offloading SSL-based encryption.

You must configure the following public gateway settings:

```
tsm configuration set -k gateway.public.host -v <name> (This should be the URL that your users are using to access Tableau Server)
```

```
tsm configuration set -k gateway.public.port -v 443
```

For more information on configuring gateway settings, see [Configuring Proxies for Tableau Server](#).

### Enable Tableau Prep Conductor

**Use the following steps to add the Data Management product key to your Tableau Server:**

**Note:** This process requires a restart of the Tableau Server.

1. If the computer where you are running Tableau Server has been configured to connect to the internet through a forward proxy, follow the procedure in the topic, [Configure Product Key Operations with Forward Proxy](#), before continuing.
2. Open TSM in a browser:

```
https://<tsm-computer-name>:8850
```

3. Click **Licensing** on the **Configuration** tab and click **Activate License**:
4. Enter or paste your **Data Management product key** and click **Activate**:
5. On the Register page, enter your information into the fields and click **Register**.
6. Follow the prompts and restart Tableau after registration is complete.

### Verify Tableau Prep Conductor is enabled and running

When you activate the Data Management product key, a single instance of Tableau Prep Conductor is automatically enabled on any node that has Backgrounder enabled.

#### Use the following steps to verify that it is enabled and running:

1. Open a browser and enter the Tableau Server URL, and append the dedicated TSM web UI port. Here are some examples of what the URL might look like:

`https://localhost:8850/` (if you're working directly on the server computer)

`https://MarketingServer:8850/` (if you know the server's name)

`https://10.0.0.2:8850/` (if you know the server's IP address)

In the sign-in page that appears, enter your administrator user name and password.

**Note:** Tableau Server creates and configures a self-signed certificate during the installation process. This certificate is used to encrypt traffic to the TSM Web UI. Because it's a self-signed certificate, your browser will not trust it by default. Therefore, your browser will display a warning about the trustworthiness of the certificate before allowing you to connect.

2. In the Tableau Services Manager web interface, click the **Status** tab to see the status.
  - If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active** on at least on one node. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information for any of the nodes.

#### Tableau Prep Conductor not enabled:

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VuQL Server	✓✓✓✓		
Cache Server	✓✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓✓		✓✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor	✓		
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

**Tableau Prep Conductor enabled and running. In the image below Tableau Prep Conductor is enabled on node1 and node3:**

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VuQL Server	✓✓✓✓		
Cache Server	✓✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓✓		✓✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor	✓		✓
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

## Dedicate a node for Tableau Prep Conductor

On the node you are planning to dedicate to running flows, enable Backgrounder process if it is not already enabled. It is recommended that you do not run other processes like VizQL server on this node.

Because you are dedicating this node to running flows, you must configure Backgrounder to run only flow tasks. By default, the Backgrounder process runs tasks of all types, including flows, extract refreshes, and subscriptions. For more information, see [Node Roles in Tableau Server](#).

Run the following tsm commands on that dedicated node to run only flow tasks:

1. Run the following command to allow Backgrounders on this node to run only flow tasks.

```
tsm topology set-node-role -n node1 -r flows
```

2. Set the node role on the initial node to no flows. The backgrounder on this node will run all jobs except flows:

```
tsm topology set-node-role -n node1 -r no-flows
```

3. Apply the changes and restart Tableau Server:

```
tsm pending-changes apply
```

### Multi-node installations

If you have more than 2 nodes in your Tableau Server installation, you can choose to configure other nodes to run all tasks other than flows:

1. Restrict a node to not allow flows. This command removes Tableau Prep Conductor from this node and Backgrounders on this node will not run flow tasks.

```
tsm topology set-node-role -n node1 -r no-flows
```

2. Apply the changes and restart Tableau Server:

```
tsm pending-changes apply
```

### Next step

**Step 2: Configure Flow Settings for your Tableau Server** on page 387

## Who can do this

Server administrators can install Tableau Server and enable Tableau Prep Conductor.

## Step 1 (Existing Install): Enable Tableau Prep Conductor

This topic describes how to enable Tableau Prep conductor on your existing installation of Tableau Server.

Tableau Prep Conductor is supported only on Tableau Server versions 2019.1 or later. If you are using Tableau Server 2018.3 or earlier, you must first upgrade your Tableau Server to 2019.1 before enabling Tableau Prep Conductor on your Tableau Server installation.

Tableau Prep Conductor is licensed through the Data Management Add-on, on a per Deployment basis, which is User-Based or Core-Based. A Deployment includes a licensed production Tableau Server installation and licensed non-production Tableau Server installations that support the production installation. For more information on Deployment, see the [EULA Documentation](#).

This topic describes how to enable Tableau Prep conductor on your existing installation of Tableau Server.

## Before you upgrade

Prepare for upgrade:

- [Know before you upgrade](#)
- [Licensing Tableau Prep Conductor](#)
- [Tableau Server Hardware Requirements and Recommendations](#)

## Configure public gateway settings

If your Tableau Server is set up with one of the following:

- Load balancer to distribute requests across gateways.
- Reverse proxy to authenticate external (internet) client requests and offloading SSL-based encryption.

You must configure the following public gateway settings:

```
tsm configuration set -k gateway.public.host -v <name> (This should be the URL that your users are using to access Tableau Server)
```

```
tsm configuration set -k gateway.public.port -v 443
```

For more information on configuring gateway settings, see [Configuring Proxies for Tableau Server](#).

## Tableau Server Installations using User-Based licenses

The recommended topology for a production Tableau Server installation is a dedicated node for running flows. For more information, see [Minimum Hardware Requirements and Recommendations for Tableau Server](#).

### Tableau Server single-node installations

If you currently have a single node Tableau Server installation, it is recommended that you add a second node and dedicate it to running flows.

1. Run upgrade on your current Tableau Server installation using the information in the topics below:

- [Windows](#)
- [Linux](#)

When you get to the **Activate** step, use the Tableau Server product keys to activate Tableau Server.

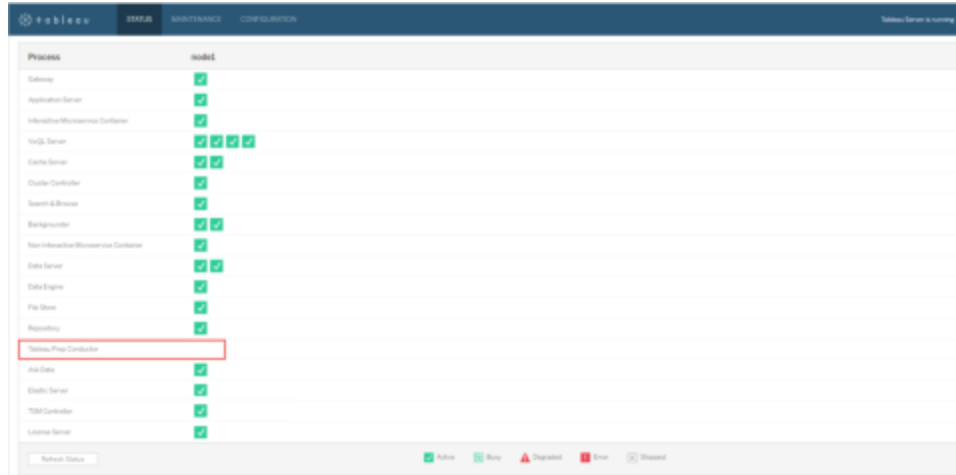
All product keys are available through the [Customer Portal](#).

2. After completing the installation, add the Data Management product key to enable Tableau Prep Conductor on your node. The Data Management product key, like your other server keys, are available through the [Customer Portal](#).
  - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
  - Enter or paste your new product key and click **Activate**.
  - On the **Register** page, enter your information into the fields and click **Register**.
3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and running.
  - In the Tableau Services Manager web interface, click the **Status** tab to see the status. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep

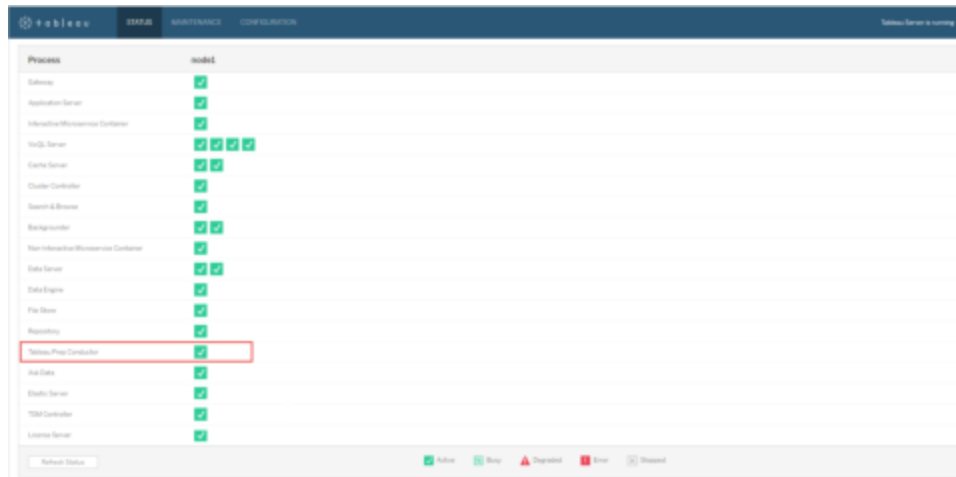


Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

**Tableau Prep Conductor not enabled:**



**Tableau Prep Conductor enabled and running:**



4. Add a second node to your Tableau Server installation. The installer will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine and Tableau Prep Conductor on the node. Do not add any other processes on this node.
5. Run the following commands to dedicate this node to do only flow tasks. For more

information on node roles, see [Node Roles in Tableau Server](#).

- Get the nodeID for your dedicated node to see the list of services on each node:

```
tsm topology list-nodes -v.
```

- Set the node role for the dedicated node using the nodeID that you got from running the command described above:

```
tsm topology set-node-role -n <nodeID> -r flows.
```

- Apply the changes, and restart the server:

```
tsm pending-changes apply.
```

- Review the status to ensure that all the processes are up and running and configured correctly:

```
tsm status -v.
```

You have successfully added Tableau Prep Conductor to your Tableau Server installation.

## Tableau Server multi-node installations

1. Run upgrade on your current Tableau Server Installation using the information in the topics below:
  - [Windows](#)
  - [Linux](#)

When you get to the **Activate** step, use the Tableau Server product keys to activate Tableau Server.

All product keys are available through the [Customer Portal](#).

2. After completing the installation, add the Data Management product key to enable Tableau Prep Conductor. Tableau Prep Conductor is automatically enabled on the nodes where you already have the Backgrounder process enabled. The Data Management product key, like your other server keys, are available through the [Customer Portal](#).
  - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
  - Enter or paste your new product key and click **Activate**.
  - On the **Register** page, enter your information into the fields and click **Register**.

3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and is running.
  - In the Tableau Services Manager web interface, click the **Status** tab to see the status of all the processes. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

**Tableau Prep Conductor not enabled:**

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VoQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor			
Ask Data	✓		
Elastic Server	✓		
TDM Controller	✓		
License Server	✓		

Refresh Status

Active Busy Degraded Error Stopped

## Tableau Prep Conductor enabled and running:

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VueQL Server	✓ ✓ ✓ ✓		
Cache Server	✓ ✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓ ✓		✓ ✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓ ✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
Tableau Prep Conductor	✓		✓
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

4. Add a new node to your Tableau Server installation. The installer will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine and Tableau Prep Conductor on the node. Do not add any other processes on this node.

**Note:** The dedicated node counts towards to total count of the Coordination Service ensemble. You may need to deploy a Coordination Service on the new node depending on the total number of nodes you have in your cluster including the new dedicated node. For more information, see [Deploy a Coordination Service Ensemble](#).

5. Run the following command to dedicate this node to only doing flow related operations. For more information on node roles, see [Node Roles in Tableau Server](#).
  - Get the nodeID for your dedicated node to see the list of services on each node:
  - `tsm topology list-nodes -v`.
  - Set the node role for the dedicated node using the nodeID that you got from running the command described above:

## Tableau Prep Help

- `tsm topology set-node-role -n <nodeID> -r flows.`
  - Apply the changes and restart the server:
  - `tsm pending-changes apply.`
  - Review the status to ensure that all the processes are up and running and configured correctly:
  - `tsm status -v.`
6. At this stage, you may have Tableau Prep Conductor enabled on other nodes. By default, the Backgrounder process on a node performs all tasks of all types including flow tasks. To isolate Tableau Prep Conductor and flow tasks to only certain nodes, you can configure the Backgrounders to do one of the following:
- To run only flow tasks: `tsm topology set-node-role -n <nodeID> -r flows.`
  - To run all other tasks except flows: `tsm topology set-node-role -n <nodeID> -r no-flows.`

You have successfully added Tableau Prep Conductor to your Tableau Server installation.

## Tableau Server Installations using Core-Based licenses

The recommended topology for a production Tableau Server installation is a dedicated node for running flows. For more information, see [Minimum Hardware Requirements and Recommendations for Tableau Server](#).

The Data Management Add-on for Core-Based licenses includes product keys that enable Tableau Prep Conductor for your Tableau Server, and Tableau Prep Conductor cores that comes in units of four. The Tableau Prep Conductor cores should be applied to the node dedicated to running the flows. These product keys, like your other server keys, are available through the [Customer Portal](#).

To learn more about Tableau Prep Conductor licensing, see [Licensing Tableau Prep Conductor for Tableau Server](#).

### Tableau Server single-node installations

If you currently have a single node Tableau Server installation, it is recommended that you add a second node and dedicate it to running flows.

1. Run upgrade on your current Tableau Server Installation using the information in the topics below:

- **Windows**
  - **Linux**
2. Activate the product keys. This will enable Tableau Prep Conductor on the nodes where you already have the Backgrounder process enabled.
    - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
    - Enter or paste your new product key and click **Activate**.
    - On the **Register** page, enter your information into the fields and click **Register**.
  3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and is running.
    - In the Tableau Services Manager web interface, click the **Status** tab to see the status. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

#### Tableau Prep Conductor not enabled:

Process	Status
Colony	Active
Application Server	Active
Infrastructure Management Container	Active
Web Server	Active
Cache Server	Active
Cluster Controller	Active
Search & Browse	Active
Backgrounder	Active
Non-Interactive Management Container	Active
Data Server	Active
Data Engine	Active
File Store	Active
Repository	Active
<b>Tableau Prep Conductor</b>	
Web Data	Active
Entity Server	Active
TSM Controller	Active
Licensing Server	Active

**Tableau Prep Conductor enabled and running:**

Process	node1
Authentic	✓
Application Server	✓
Interactive Workspaces Container	✓
WebUI Server	✓ ✓ ✓ ✓
Cache Server	✓ ✓
Cluster Controller	✓
Server & Services	✓
Backgrounder	✓ ✓
Non-Interactive Workspaces Container	✓
Data Server	✓ ✓
Data Engine	✓
File Store	✓
Repository	✓
<b>Tableau Prep Conductor</b>	<b>✓</b>
Job Data	✓
Health Server	✓
TSM Controller	✓
License Server	✓

4. Add a second node to your Tableau Server installation. The installer will enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine and Tableau Prep Conductor on the node. Do not add any other processes on this node.

**Important:** The number of physical cores on this machine must be equal to, or less than the Tableau Prep Conductor cores you purchased. For example, if you purchased four Tableau Prep Conductor cores, your node can only have up to four physical cores. To understand about how Tableau Prep Conductor licensing works, see [Licensing Tableau Prep Conductor for Tableau Server](#).

5. Run the following commands to dedicate this node to only doing flow tasks. For more information on node roles, see [Node Roles in Tableau Server](#).
  - Get the nodeID for your dedicated node to see the list of services on each node:
 

```
tsm topology list-nodes -v.
```
  - Set the node role for the dedicated node using the nodeID that you got from running the command described above:
 

```
tsm topology set-node-role -n <nodeID> -r flows.
```
  - Apply the changes and restart the server: `tsm pending-changes apply.`
  - Review the status to ensure that all the processes are up and running and

configured correctly:

```
tsm status -v.
```

You have successfully added Tableau Prep Conductor to your Tableau Server installation.

## Tableau Server multi-node installations

1. Run upgrade on your current Tableau Server installation using the information in the topics below:
  - [Windows](#)
  - [Linux](#)
2. Activate the product keys. This will enable Tableau Prep Conductor on the nodes where you already have the Backgrounder process enabled.
  - In the Tableau Services Manager web interface, click **Licensing** on the **Configuration** tab and click **Activate License**.
  - Enter or paste your new product key and click **Activate**.
  - On the **Register** page, enter your information into the fields and click **Register**.
3. You will be prompted to restart the server. Restart the server and verify that Tableau Prep Conductor is enabled and is running.
  - In the Tableau Services Manager web interface, click the **Status** tab to see the status. If Tableau Prep Conductor is enabled and running, you should see Tableau Prep Conductor in the list of processes as **Active**. If Tableau Prep Conductor is not enabled, you will see Tableau Prep Conductor in the list of processes, but with no status information.

### **Tableau Prep Conductor not enabled:**



Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VueQL Server	✓✓✓✓		
Cache Server	✓✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓✓		✓✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
<b>Tableau Prep Conductor</b>	✓		
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

**Tableau Prep Conductor enabled and running:**

Process	node1	node2	node3
Gateway	✓		
Application Server	✓		
Interactive Microservice Container	✓		
VueQL Server	✓✓✓✓		
Cache Server	✓✓		
Cluster Controller	✓	✓	✓
Search & Browse	✓		
Backgrounder	✓✓		✓✓
Non-Interactive Microservice Container	✓		✓
Data Server	✓✓		
Data Engine	✓	✓	✓
File Store	✓	✓	
Repository	✓		
<b>Tableau Prep Conductor</b>	✓		✓
Ask Data	✓		
Elastic Server	✓		
TSM Controller	✓		
License Server	✓		

4. Add a new node to your Tableau Server installation. A dedicated node to run flow related operations is recommended for production Tableau Server installations. The installer will

enable certain required processes like the Cluster Controller. Enable Backgrounder process on it as it is required to run scheduled flow tasks. When you enable the Backgrounder process, the installer automatically enables a single instance of Data Engine on the node. Do not add any other processes on this node.

**Note:** The dedicated node counts towards to total count of the Coordination Service ensemble. You may need to deploy a Coordination Service on the new node depending on the total number of nodes you have in your cluster including the new dedicated node. For more information, see [Deploy a Coordination Service Ensemble](#).

**Important:**

The number of physical cores on this machine must be equal to, or less than the Tableau Prep Conductor cores you purchased. For example, if you purchased four Tableau Prep Conductor cores, your node can only have up to four physical cores. To understand about how Tableau Prep Conductor licensing works, see [Licensing Tableau Prep Conductor for Tableau Server](#).

5. Run the following commands to dedicate this node to only doing flow tasks. This will enable Tableau Prep Conductor on your new node. For more information, see [Node Roles in Tableau Server](#).
  - Get the nodeID for your dedicated node to see the list of services on each node:
 

```
tsm topology list-nodes -v.
```
  - Set the node role for the dedicated node using the nodeID that you got from running the command described above:
 

```
tsm topology set-node-role -n nodeID -r flows.
```
  - Apply the changes and restart the server:
 

```
tsm pending-changes apply.
```
  - Review the status to ensure that all the processes are up and running and configured correctly:
 

```
tsm status -v.
```
6. At this stage, you may have Tableau Prep Conductor enabled on other nodes that have the Backgrounder process. By default, the Backgrounder process on a node performs

all tasks of all types including flow tasks. To isolate Tableau Prep Conductor and flow operations to only certain nodes, you can configure the backgrounders to do one of the following:

- To run only flow tasks:

```
tsm topology set-node-role -n <nodeID> -r flows.
```

- To run all other tasks except flows:

```
tsm topology set-node-role -n <nodeID> -r no-flows.
```

## Next step

**Step 2: Configure Flow Settings for your Tableau Server** below.

### Who can do this

Tableau Server Administrators can install or upgrade Tableau Server, and enable Tableau Prep Conductor on Tableau Server.

## Step 2: Configure Flow Settings for your Tableau Server

This topic describes the various flow settings that you can configure for your Tableau Server.

### Publishing, Scheduling, and Credential Settings

When you activate Tableau Prep Conductor using the Data Management product key, Tableau Prep Conductor is enabled for the entire Tableau Server installation. You can further modify and customize the setting for sites.

Use the following instructions to configure settings related to flows for all your sites or for individual sites:

Use the following instructions to sign in to Tableau Server Admin pages:

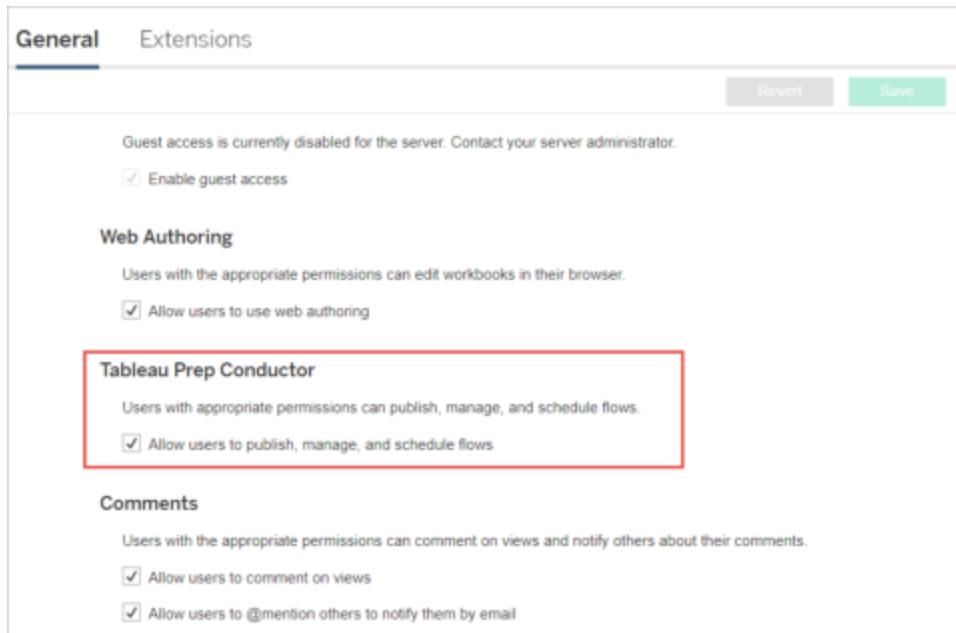
- Windows: [Tableau Server Admin Pages](#).
- Linux: [Tableau Server Admin Pages](#).

### Configure whether publishing and scheduling flow should be allowed for a site:

1. **Enable users to publish and schedule flows:** This setting is enabled by default when you enable Tableau Prep Conductor. If you have multiple sites, you can selectively turn

off Tableau Prep for Server for individual sites. If you disable this setting for a site that once allowed flows, see [Implication of disabling Tableau Prep Conductor](#) below to learn more.

On the **General** page under **Settings**, scroll to the **Flows** setting and clear the **Allow user to publish and schedule flows** check box to turn off Tableau Server for the site.



## 2. Embed Credentials

### **-Allow publishers to embed credentials in a data source, flow or workbook:**

This setting allows publishers to attach passwords to published flows that will automatically authenticate web users.

**-Allow publishers to schedule flow runs and data extract refreshes:** This option is only available if setting above is enabled. When this setting is enabled, publishers will see scheduling options in the Publish dialog box.

## Implication of disabling Tableau Prep Conductor

If you disable Tableau Prep Conductor after using it for a while, you will not be able to see the flows, schedules, tasks, and other things related to flows. The following table gives you more information on what you can and cannot see when you disable Tableau Prep for Server completely or only for specific sites:

	<b>Prep not enabled at Server level</b>	<b>Prep enabled at Server level, but disabled for a site</b>	<b>Prep enabled for both Server and site</b>
<b>Show flows</b>	No	No	Yes
<b>Show tasks/schedules in Server view</b>	No	Yes	Yes
<b>Show tasks/schedules in Site view</b>	No	No	Yes
<b>Show Site setting (only for Server Admins)</b>	Yes (disabled)	Yes	Yes
<b>Show TSM status</b>	Yes (Tableau Prep Conductor is not shown)	Yes	Yes
<b>Show TSM settings</b>	Yes (disabled)	Yes	Yes

**Important:** Scheduled tasks will continue to run even when Tableau Prep Conductor is disabled for that site, but will fail.

## Configure notifications for flow failures

You can configure Tableau Server to send email notifications for flow run failures. The notifications are sent for failures that occur when running the flows through either a scheduled task or a manual run using the **Run now** menu option. You must first enable the server-wide setting, and then configure at the site level.

### To enable the server-wide email notification

You can either use the Tableau Services Manager (TSM) web interface or TSM CLI as described below:

## Use the TSM web interface

1. Open TSM in a browser:  
`https://<tsm-computer-name>:8850.`
2. Click **Notifications** on the **Configuration** tab and click **Email Server**.
3. Enter the email server information.
4. Click the **Events** tab.
5. Under **Content Updates**, select **Send emails for extract refresh failures and flow run failures** if not already turned on by default.
6. Click **Save Pending Changes** after you've entered your configuration information.
7. Click **Apply Changes and Restart**.

## Use the TSM CLI

The notification values can be set individually with the tsm configuration set command:

Windows: [tsm configuration](#).

Linux: [tsm configuration](#).

### Set notification values

Use the tsm configuration set command with the following syntax to

to enable flow failure notifications, run the following command:

```
tsm configuration set -k backgrounder.notifications_enabled -v true
```

**Note:** This will enable email notification for both extract refresh failures and flow failures.

After you are done setting values, you must run the following command:

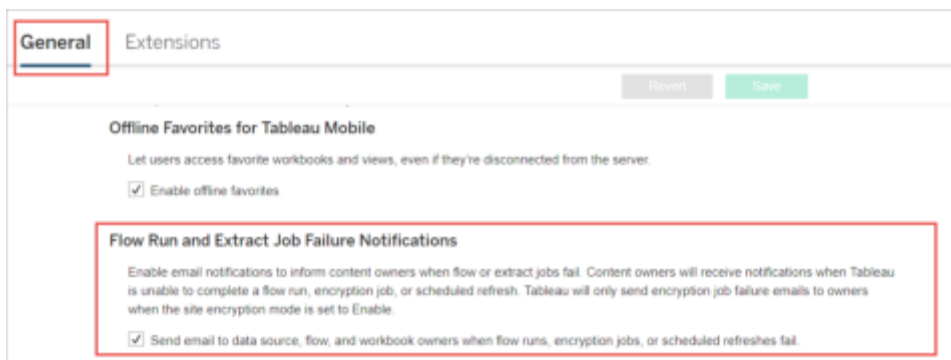
```
tsm pending-changes apply
```

The pending-changes apply command displays a prompt to let you know this will restart Tableau Server if the server is running. The prompt displays even if the server is stopped, but

in the case there is no restart. You can suppress the prompt using the --ignore-prompt option, but this does not change the restart behavior.

To configure email notification for a site:

On the **General** page under **Settings**, scroll to the **Flow Run and Refresh Failure Notifications** setting and clear the **Send email to data source, flow, and workbook, when flow runs and scheduled refreshes fail** check box to turn off refresh failure and flow run notifications.



## Next step

**Step 3: Create Schedules for Flow Tasks** on the next page

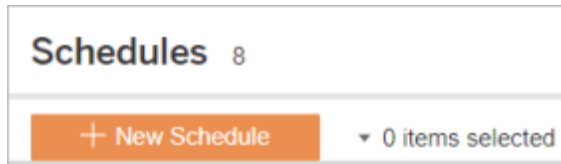
## Who can do this

Tableau Server Administrators can configure server and site level settings. Tableau Site Administrators can configure site level settings.

## Step 3: Create Schedules for Flow Tasks

Create a new schedule:

1. On the **Schedules** tab, click **New Schedule**.



2. Enter the following information in the **New Schedule** dialog box and click **Create**.
  - **Name:** Enter a descriptive name for the schedule. Typically, this includes the description of the schedule frequency.
  - **Task Type:** Select Flow as the task type.
  - **Default Priority:** You can define a default priority from 1 to 100, where 1 is the highest priority. This value will be assigned to the tasks by default. If two tasks are pending in the queue, the one with the higher priority runs first.
  - **Execution:** Choose whether a schedule will run in parallel or serially. Schedules that run in parallel run on all available background processes so that they can complete faster.



- **Frequency:** You can define an hourly, daily, weekly, or monthly schedule.

**New Schedule**

Create a schedule users can choose for running extract refreshes, flows, or subscriptions.

Name:

Task type:

Default priority:   
Tasks are executed in priority order from 1 to 100

Execution:  Parallel: Use all available background processes for this schedule  
 Serial: Limit this schedule to one background process

Frequency:  Hourly  
 Daily  
 Weekly  
 Monthly

every:   
from:  :    
to:  :

## Next step

### Step 4: Safe list Input and Output locations below

#### Who can do this

Tableau Server Administrators can create and modify schedules. Schedules are created at the server level and apply across all the sites on a server.

## Step 4: Safe list Input and Output locations

This topic describes the rules that apply to this feature and how to safe list the directories on your network.

Flow input and output connections may need to connect to databases or files in the directories on your network. You must safe list the directories you want to allow access to. Input and Output connections will only be allowed to connect to data in the safe listed locations. By default, no connections are allowed.

**Note:** You can still publish the flows and any data that is embedded in the flow file (tflx) to Tableau Server, but the flow will fail to run if the directories aren't included in your organization's safe list.

## How to safe list input and output locations

The following rules apply and must be considered when configuring this setting:

- The directory paths should be accessible by Tableau Server. These paths are verified during server startup and at flow run time and are **not** verified at the time of publishing the flow to Tableau Server.
- Network directory paths have to be absolute and cannot contain wildcards or other path traversing symbols. For example, `\\myhost\myShare\*` **or** `\\myhost\myShare*` are invalid paths and would result in all the paths as disallowed. The correct way to safelist any folder under *myShare* would be `\\myhost\myShare` **or** `\\myhost\myShare\`.

**Note:** The `\\myhost\myShare` configuration will not allow `\\myhost\myShare1`. In order to safe list both of these folders safe list them as `\\myhost\myShare; \\myhost\myShare1`.

- **Windows:**
  - The value can be either `*`, (for example, `tsm configuration set -k maestro.input.allowed_paths -v "*"`) to allow any network directory, or a specified list of network directory paths, delimited by a semicolon (;).
  - If the path contains spaces or special characters you will have to either use single or double quotes. Whether you use single or double quotes depends on the shell that you are using.
  - No local directory paths are allowed even when the value is set to `*`.
- **Linux:**
  - The value can be either `*`, (for example, `tsm configuration set -k maestro.input.allowed_paths -v "*"`) meaning that any path, including local (with the exception of some system paths configured using "native\_api.internal\_disallowed\_paths"), or a list of paths, delimited by a semicolon (;).

- You must be using a kernel version of equal to or later than 4.7. Safe listing is not supported on kernel version earlier than 4.7. To check the kernel version, in the Linux terminal, type the command `uname -r`. This will display the full version of the kernel you are running on the Linux machine.

**Note:** If a path is both on the flows allowed list and `internal_disallowed` list, `internal_disallowed` takes precedence.

Use the following commands to create a list of allowed network directory paths:

### For input connections:

```
tsm configuration set -k maestro.input.allowed_paths -v your_
networkdirectory_path_1;your_networkdirectory_path_2

tsm pending-changes apply
```

### For output connections:

```
tsm configuration set -k maestro.output.allowed_paths -v your_net-
workdirectory_path_1;your_networkdirectory_path_2

tsm pending-changes apply
```

### Important:

These commands overwrite existing information and replace it with the new information you provided. If you want to add a new location to an existing list, you must provide a list of all the locations, existing, and the new one you want to add. Use the following commands to see the current list of input and output locations:

```
tsm configuration get -k maestro.input.allowed_paths
tsm configuration get -k maestro.output.allowed_paths
```

## Next step

[Step 5: Optional Server Configurations on the next page](#)

### Who can do this

Tableau Server Administrators who also are a member of the **tsmadmin** group can configure settings using *tsm* commands.

## Step 5: Optional Server Configurations

The options described in this topic are not required to enable flow publishing and scheduling flows on Tableau Server. They can be used to customizing your environment according to your requirements.

### Set the timeout period for flows

You can set time limits for how long a flow can run to make sure that subsequent tasks are not held up due to stalled tasks. The following two *tsm* command options determine how long a flow task can run before the flow background task is canceled. These two commands together determine the total timeout value for flow tasks.

Use the following *tsm* commands to change the default timeout period of 4 and half hours:

```
backgrounder.extra_timeout_in_seconds
```

(Default value: 14400 seconds or 4 hours)

```
backgrounder.default_timeout.run_flow
```

(Default value: 1800 seconds or 30 minutes)

### Set the threshold for suspended flow tasks

By default, a flow task is suspended after 5 consecutive flow task failures. To change the threshold number of flow task failures that can occur before they are suspended, use the following *tsm* configuration set command:

```
tsm configuration set -k backgrounder.flow_failure_threshold_for_run_prevention -v <number>
```

This sets the threshold for the number of consecutive failed flow tasks necessary before suspending the tasks. This is a server-wide setting.

### Who can do this

Tableau Server administrators can make changes to server configurations.

# Enable Tableau Prep Conductor on your Tableau Online Site

**Note:** Tableau Bridge is not currently supported. You can only connect directly to cloud-hosted data sources. If you connect to on-premises data sources, you can convert those data sources to a published data source and connect using a Tableau Bridge client if Tableau Bridge is configured for the data source. For more information, see [Publish a flow on page 402](#) for Tableau Online.

Prep Conductor is automatically enabled on your Tableau Online site after you purchase the Data Management Add-on. For information on how to purchase the Data Management Add-on, contact your account manager.

## Step 1: Install Tableau Prep Builder

Install Tableau Prep Builder on your local machine.

## Step 2: Log into Tableau Online

1. Sign into the server through the **Server->Sign In** menu. For more information about how to sign in to Tableau Online, see [Sign In to Tableau Online](#).
2. On the General page under Settings, if you see a section that says **Notifications for Extract Jobs and Flow Runs** then Tableau Online is licensed with the Data Management Add-on and Prep Conductor is enabled.

**Note:** If you see a section that only says **Notifications for Extract Jobs** then Tableau Online is not licensed with the Data Management Add-on and Prep Conductor is not available. Contact your account manager to purchase the Data Management Add-on.

3. Verify that the flow run failure notification emails are enabled for your site. This should be turned on by default.

# Publish a Flow to Tableau Server or Tableau Online

Publish your flows to Tableau Server or Tableau Online to automatically run them on a schedule and refresh the flow output using Tableau Prep Conductor.

**Note:** The content in this topic applies to both Tableau Server and Tableau Online, exceptions are called out specifically.

## Before you publish

To make sure that you can run your flow in Tableau Server or Tableau Online, check the following:

1. Verify that there are no errors in the flow.

Flows that contain errors will fail when you try to run them in Tableau Server or Tableau Online. Errors in the flow are identified by a red exclamation mark and a red dot with an **Errors** indicator in the upper right corner of the canvas.

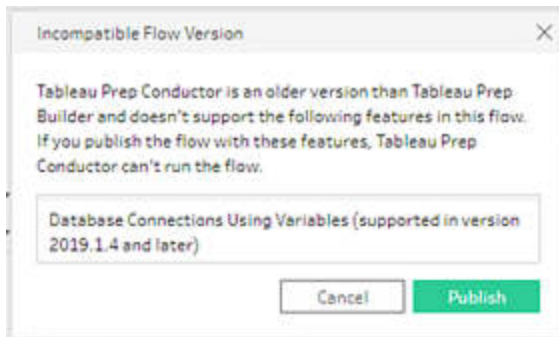


2. Verify that your flow doesn't include input connectors or features that aren't compatible with your version of Tableau Server. Tableau Online should always be running the most current version.

Flows that include connectors or features that aren't yet supported in your version of Tableau Server can be published, but can't be scheduled to run in Tableau Server. To

publish and schedule flows to run on Tableau Server, you must be using Tableau Server version 2019.1 or later and Tableau Prep Conductor must be enabled.

For example, the SAP HANA connector was introduced in Tableau Prep Builder version 2019.1.4 but this connector isn't supported until Tableau Server version 2019.2 for Tableau Prep Conductor. When you publish the flow, you would see a message like the example below.



To run your flow in Tableau Server, you need to take the appropriate actions to make the flow compatible. For more information about working with incompatible flows, see [Version Compatibility with Tableau Prep on page 347](#).

3. Flows that include input or output steps with connections to a network share require safe listing. Tableau Online doesn't support this option and files must be packaged with the flow on publish.

Flow input and output steps that point to files stored in a network share (UNC path) aren't permitted unless the file and path is accessible by the server and are included in your organization's safe list. If you publish the flow without adding the file location to your safe list, the flow will publish, but you will get an error when you try and schedule or run the flow in Tableau Server.

If the files aren't stored in a safe listed location, you will see a warning message when you publish the flow.

Publish Flow to Tableau Server

Project  
Default

Name  
My Superstore data

Description

Tags  
Add

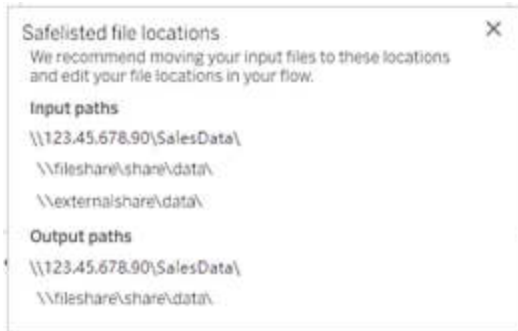
Connections Edit  
6 uploaded files ⓘ  
0 direct file connections ⓘ

Cancel Publish

**⚠** One or more output locations aren't in your organization's safe list. The flow will fail when run in Tableau Server. Save the outputs to a location in this [list](#) or publish the outputs as a data source.

Click the "list" link in the message to see a list of allowed locations. Move your files to one of the locations in the list, and make sure that your flow points to these new locations.





In Tableau Server, to configure the allowed network paths, use the tsm command options described in [Step 4: Safe list Input and Output locations on page 393](#).

If you don't want to move your files to a safe listed location, you will need to package the input files with the flow and publish the flow output to Tableau Server as a published data source. For more information about setting these options, see [Publish a flow on the next page](#) in this topic.

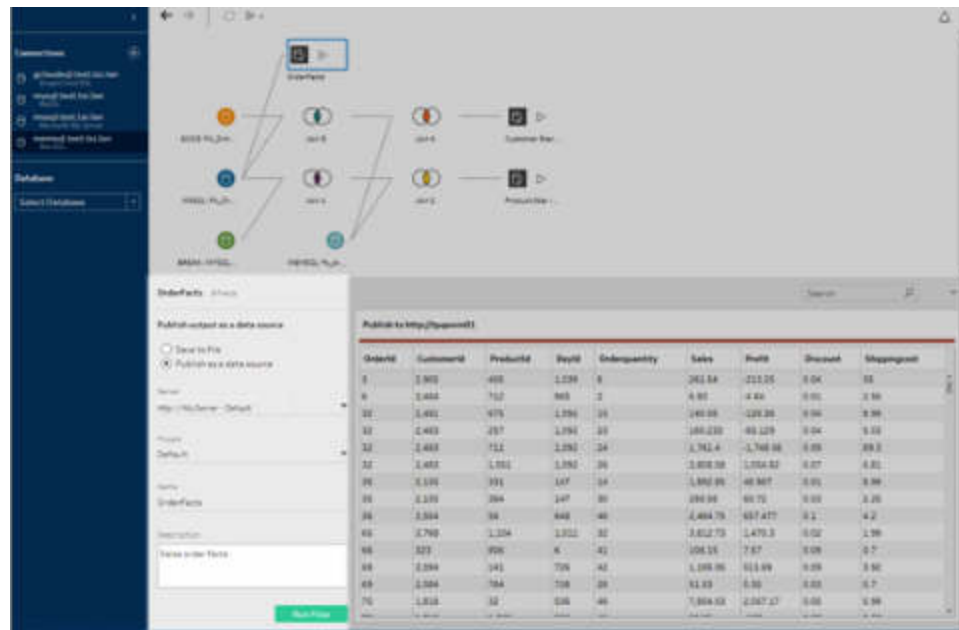
4. Make sure each flow output step is set to **Publish as a data source**. All flow output steps must point to the same server or site where the flow is published but can point to different projects on that server or site. Only one server or site can be selected.

To set the publishing location for all output steps, do the following:

- a. In the flow pane, select the output step.
- b. In the publishing pane, select **Publish as a data source**.
- c. Select the server or site and the project where you want to publish the flow. Sign in to the server or site if needed.
- d. Enter a name and description for each output.

The output file name should be distinctive enough so that the person running the flow can easily identify which output files to refresh. The file name shows on the **Overview** and **Connections** page for the flow in Tableau Server or Tableau Online.

- e. Save your flow.

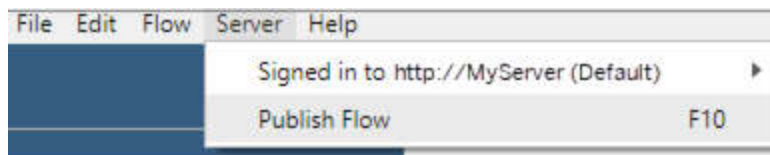


For more information about how to configure output steps for publishing, see [Create and publish data extracts and data sources](#) on page 322.

## Publish a flow

**Note:** When you publish a flow that includes a published data source as an input, the publisher is assigned as the default flow owner. When the flow runs, it uses the flow owner for the **Run As** account. For more information about the Run As account, see [Run As Service Account](#). Only the Site or Server Administrator can change the flow owner in Tableau Server or Tableau Online.

1. Open your flow in Tableau Prep Builder.
2. From the top menu select **Server > Publish Flow**.



3. Complete the fields for your platform. Then click **Publish**. Tableau Server or Tableau Online opens automatically in your default browser on the flow **Overview** page.

## Tableau Server

1. In the **Publish to Tableau Server** dialog, complete the following fields:
  - **Project:** Click the drop-down option to select your project from the project hierarchy. This should be the same project that the output files are published to.
  - **Name:** Enter a name for your flow. This name shows on the server on the Flow pages. If you want to overwrite an existing flow, click the drop-down option to select a name from the list.
  - **Description (optional):** Enter a description for the flow.
  - **Tags (optional):** Click **Add** to type in one or more tags to identify your flow so users can easily find it. Tags can also be added after publishing in the Flow pages in Tableau Server.
2. Click **Edit** in the Connections section to edit connections settings or change authentication.

**Publish Flow to Tableau Server** [X]

Project  
Default [v]

Name  
Superstore [v]

Description  
[Text Input Field]

Tags  
Add

Connections [Edit](#)

6 uploaded files ⓘ  
0 direct file connections ⓘ

Cancel Publish

## Files

By default, file input connections are packaged with the flow. Packed files aren't refreshed when the flow is run in Tableau Server. All files must have the same setting, either **Upload** or **Direct Connection**.

- To retrieve the most current data when refreshing the output files, select **Direct Connection** if Tableau Server can connect to the file location and the location is included in your organization's safe list.

Publish Flow to Tableau Server

Project  
Default

Name  
My Superstore data

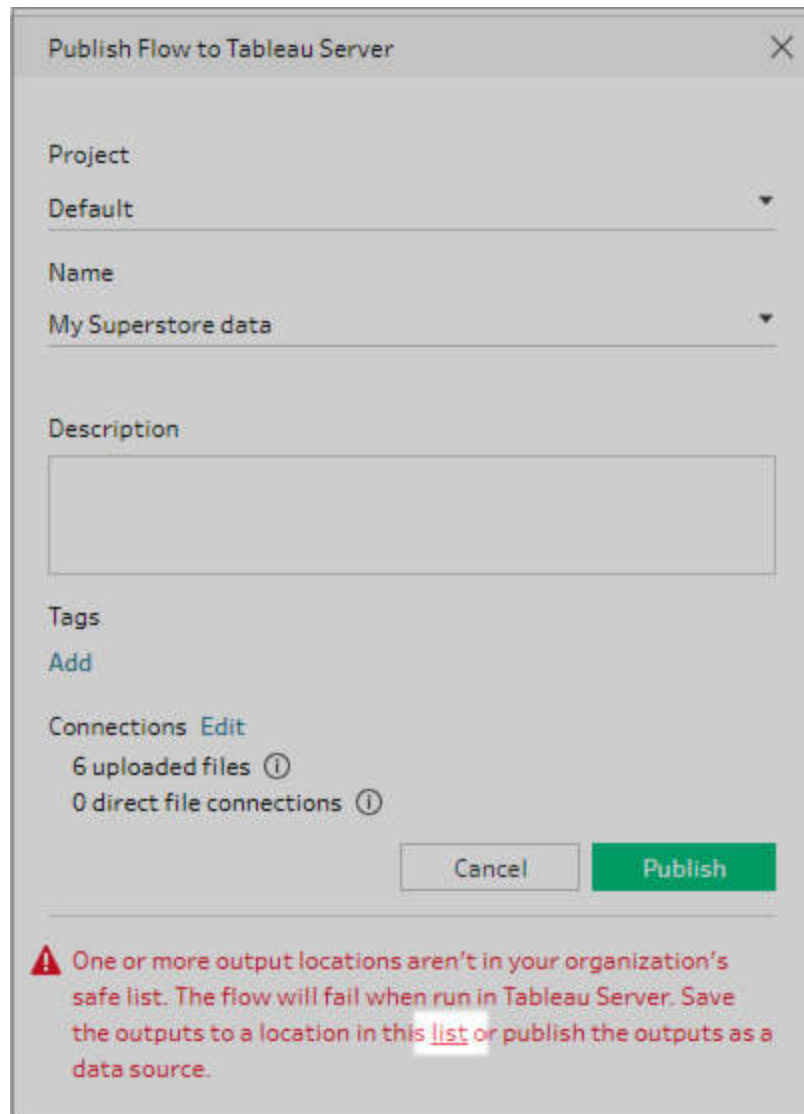
Description

Tags  
Add

Connections Edit

File	Upload	Direct Connection
Select all	<input type="radio"/>	<input checked="" type="radio"/>
Orders_Central.csv	<input type="radio"/>	<input checked="" type="radio"/>
orders_south_2015.csv	<input type="radio"/>	<input checked="" type="radio"/>
return reasons_new.xlsx	<input type="radio"/>	<input checked="" type="radio"/>
Orders_West.csv	<input type="radio"/>	<input checked="" type="radio"/>
Orders_East.xlsx	<input type="radio"/>	<input checked="" type="radio"/>
Quota.xlsx	<input type="radio"/>	<input checked="" type="radio"/>

- If your input or output steps point to files stored in a network share (UNC path) and the location isn't included in your organization's safe list, you will see a warning message. Click the link in the message to see a list of safe listed locations, move your files and point your input and output steps to the new file location. For more information, see Step 3 in [Before you publish on page 398](#)



For information about how to add locations to your organization's safe list, see [Step 4: Safe list Input and Output locations on page 393](#).

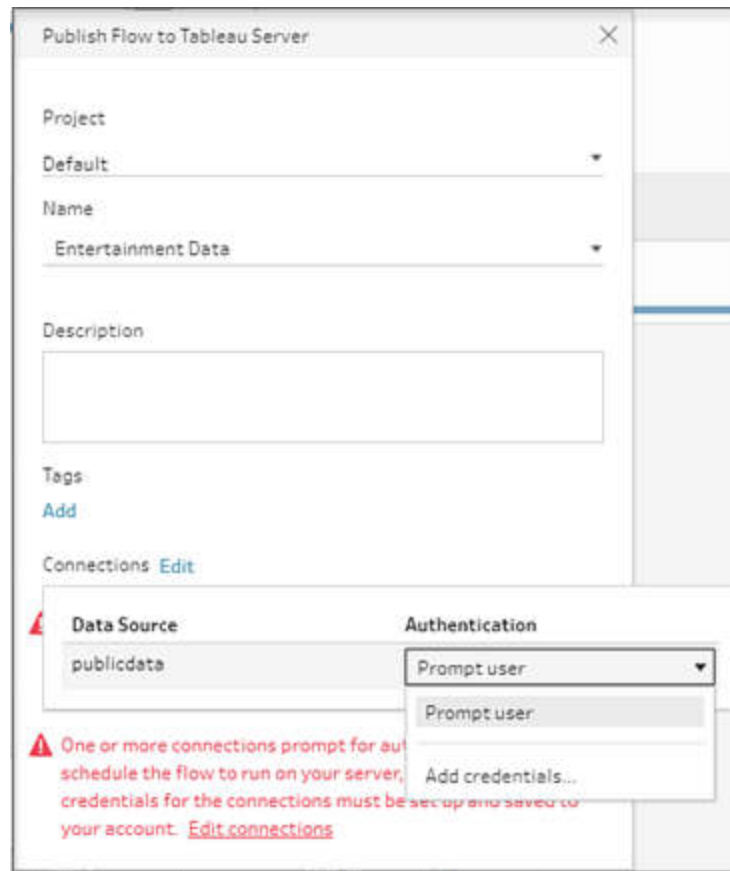
## Databases

If your flow connects to one or more databases, select one of the following authentication types to use to connect to the flow input data sources.

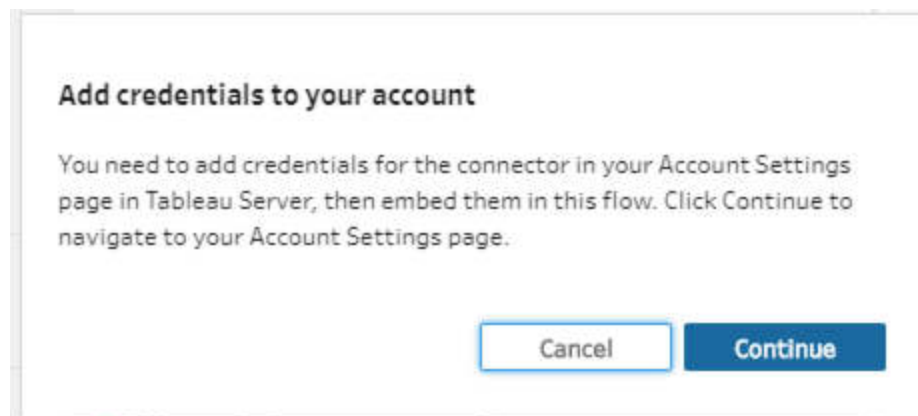
- **Server Run As Account:** The server's **Run As User** account will authenticate all users.
- **Prompt User:** You must edit the connection in Tableau Server and enter the database credentials before running the flow.
- **Embedded Password:** The credentials you used to connect to the data will be saved with the connection and used by everyone who accesses the flow you publish.

Data Source	Authentication
mysql.lan	Embedded password
mssql.lan	Embedded password
gcloudsql.lan	Prompt user
memsql.lan	Embedded password

- **Add credentials:** (version 2020.1.1 and later) If you connect to cloud connectors, you can add your credentials directly from the **Publish Flow** dialog to embed them in the flow. To add credentials, do the following:
  - Click **Edit** in the Connections section, or click **Edit credentials** from the warning message. Then click **Add credentials** from the **Authentication** drop-down menu.

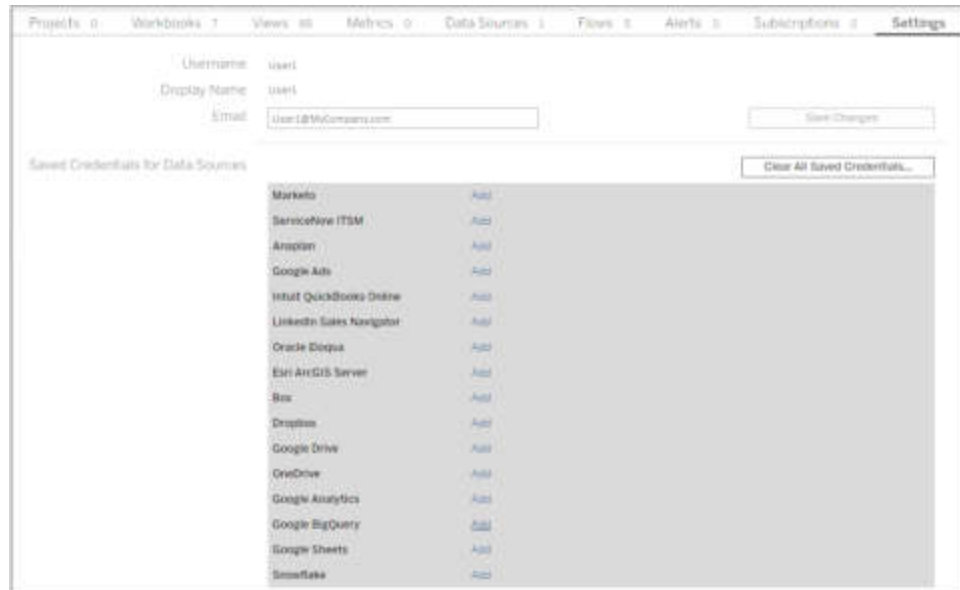


- In the confirmation dialog, click **Continue**. Tableau Prep Builder automatically opens the Account Settings page for the server you are signed into.

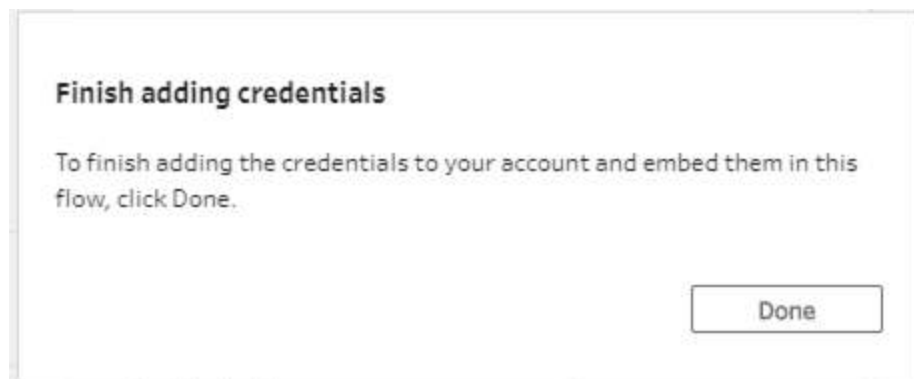


- Add your credentials, then navigate back to Tableau Prep Builder.

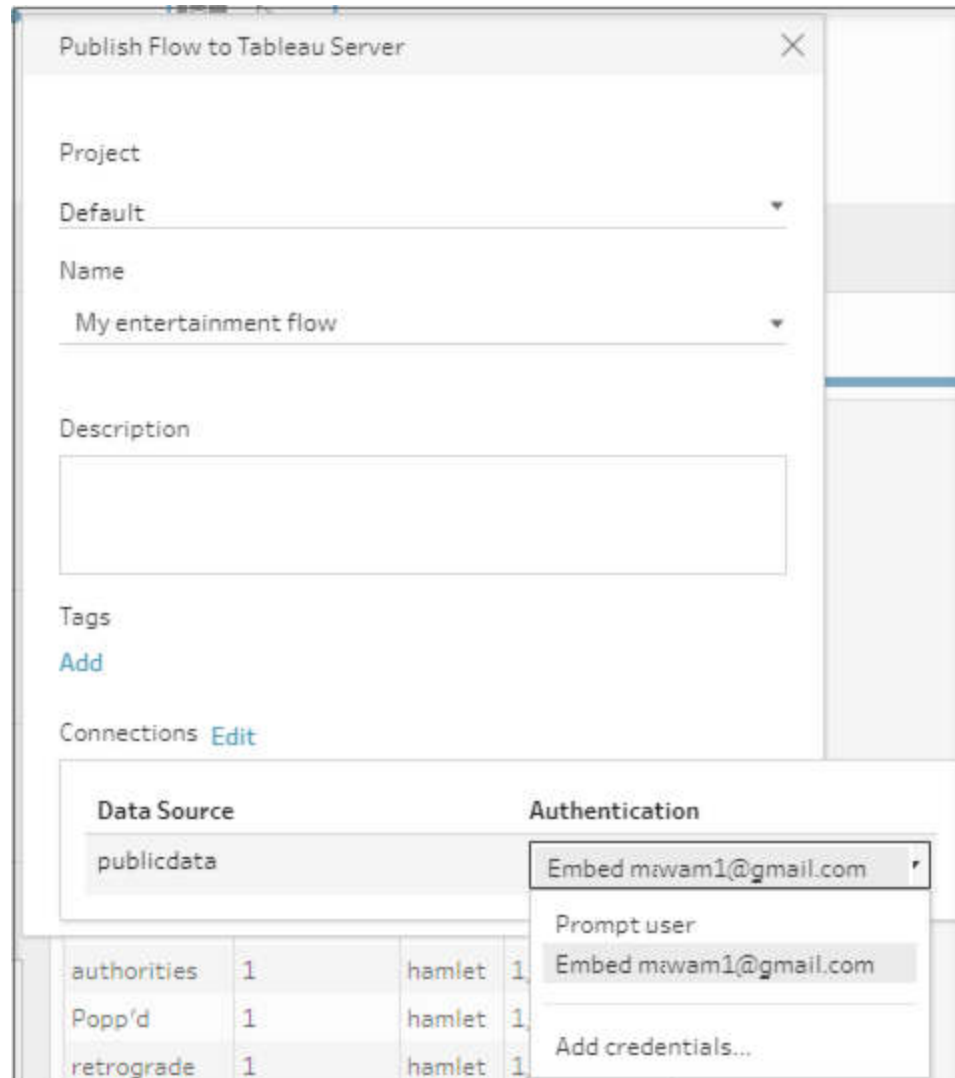




- In the **Finish adding credentials** dialog, click **Done**.



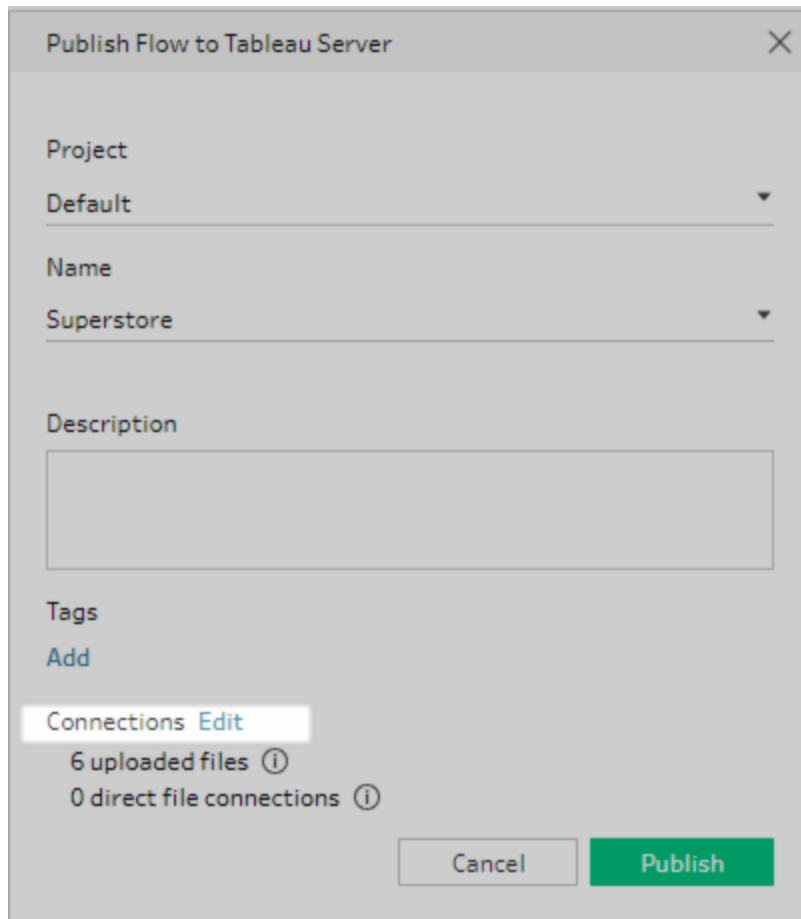
- Click **Edit** in the **Connections** section and verify that your credentials were added and embedded in your flow.



## Tableau Online

1. In the **Publish to Tableau Online** dialog, complete the following fields:
  - **Project:** Click the drop-down option to select your project from the project hierarchy. This should be the same project that the output files are published to.
  - **Name:** Enter a name for your flow. This name shows on the server on the Flow pages. If you want to overwrite an existing flow, click the drop-down option to select a name from the list.
  - **Description (optional):** Enter a description for the flow.

- **Tags (optional):** Click **Add** to type in one or more tags to identify your flow so users can easily find it. Tags can also be added after publishing in the Flow pages in Tableau Server.
2. Click **Edit** in the Connections section to edit connections settings or change authentication.



## Files

Tableau Online doesn't support direct file connections for input step data and you must package your files with the flow. Packed files aren't refreshed when the flow is run in Tableau Online.

## Databases

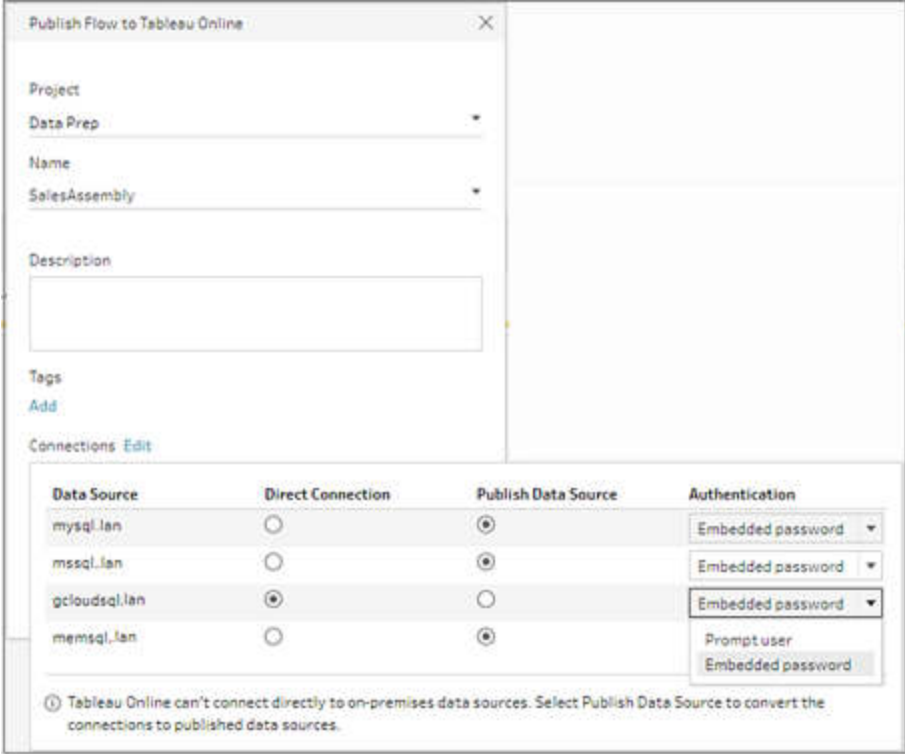
To keep data fresh when publishing flows to Tableau Online, you can only connect directly to cloud-hosted data sources. But when connecting to on-premises data sources, you must convert the data sources to a published data source and Tableau Online can use a Tableau Bridge client to connect to your data if Tableau Bridge is configured for the data source.

For more information about direct connections supported by Tableau Online, see [Allow Direct Connections to Data Hosted on a Cloud Platform](#).

For more information about using a Tableau Bridge, see [Allow your Publishers to Maintain Live Connections to On Premises Data](#).

If your flow connects to a cloud-based data source that supports a direct connection, select one of the following authentication types to use to connect to the flow input data sources.

- **Prompt User:** You must edit the connection in Tableau Online and enter the database credentials before running the flow.
- **Embedded Password:** The credentials you used to connect to the data will be saved with the connection and used by everyone who accesses the flow you publish.



Data Source	Direct Connection	Publish Data Source	Authentication
mysql.lan	<input type="radio"/>	<input checked="" type="radio"/>	Embedded password
mssql.lan	<input type="radio"/>	<input checked="" type="radio"/>	Embedded password
gcloudsql.lan	<input checked="" type="radio"/>	<input type="radio"/>	Embedded password Prompt user Embedded password
memsql.lan	<input type="radio"/>	<input checked="" type="radio"/>	Embedded password

Tableau Online can't connect directly to on-premises data sources. Select Publish Data Source to convert the connections to published data sources.

- Select the **Publish Data Source** radio button for on-premises data sources. Tableau Online can't connect directly to these data sources to refresh your data. Selecting this option converts the data source input connection to a published data source when you publish the flow to Tableau Online.

If Tableau Bridge is configured for the data source and the data source is supported by Tableau Online, the data can be refreshed when the flow is run. See [Allow Direct Connections to Data Hosted on a Cloud Platform](#) for more information.

- To replace the on-premises data source connections for the flow in Tableau Prep Builder with the published data source, select **Update flow inputs to use published data sources** in the **More options** section before publishing your flow.

If you don't select the check box, the flow in Tableau Prep Builder remains connected to the local on-premises data source and the flow in Tableau Prep Builder can become out of sync with the published version of the flow. To continue working with your flow, you would need to download the flow from Tableau Online to edit it, then republish it.

Publish Flow to Tableau Online

Project

Name  
SalesAssembly

Description

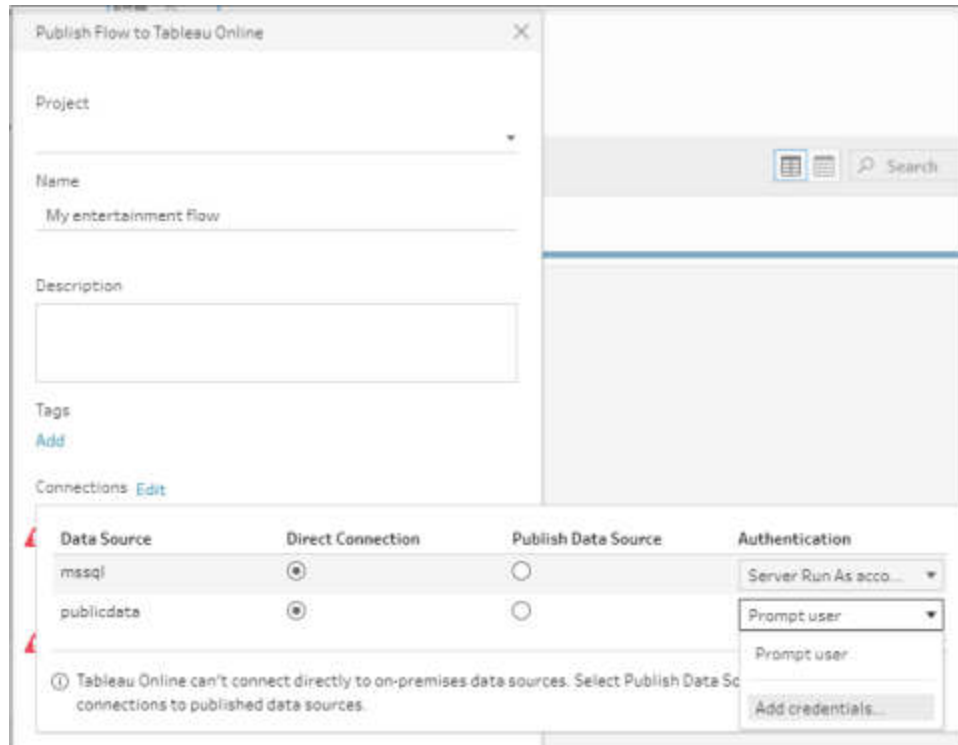
Tags  
Add

Connections Edit  
1 direct database connection ⓘ  
3 published data source connections ⓘ

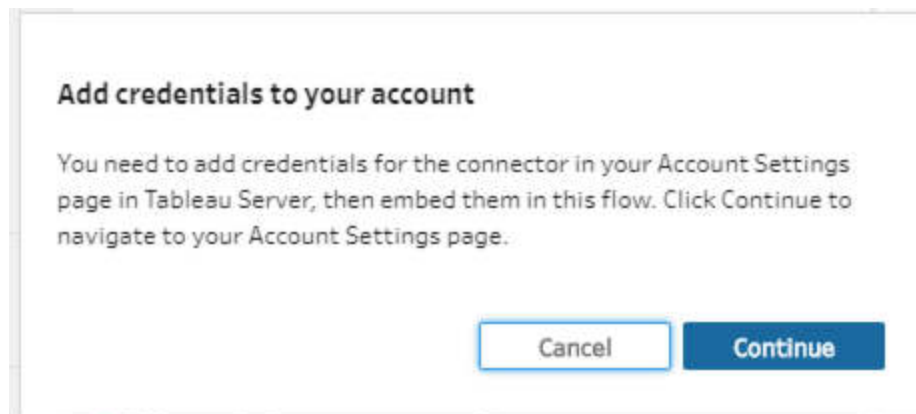
More options:  
 Update flow inputs to use published data sources

Cancel Publish

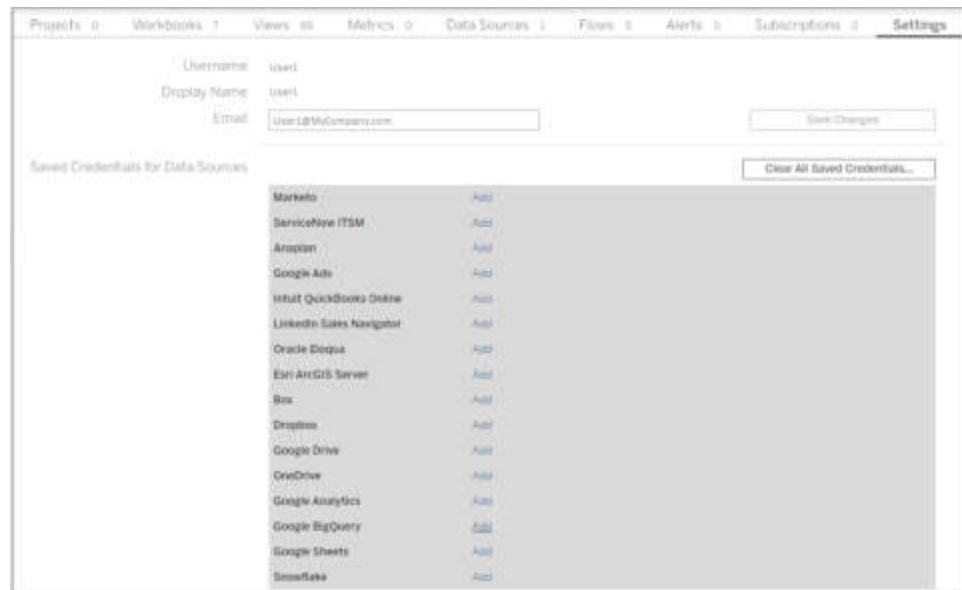
- **Add credentials:** (version 2020.1.1 and later) If you connect to cloud connectors, you can add your credentials directly from the **Publish Flow** dialog to embed them in the flow. To add credentials, do the following:
  - Click **Edit** in the Connections section, or click **Edit credentials** from the warning message. Then click **Add credentials** from the **Authentication** drop-down menu.



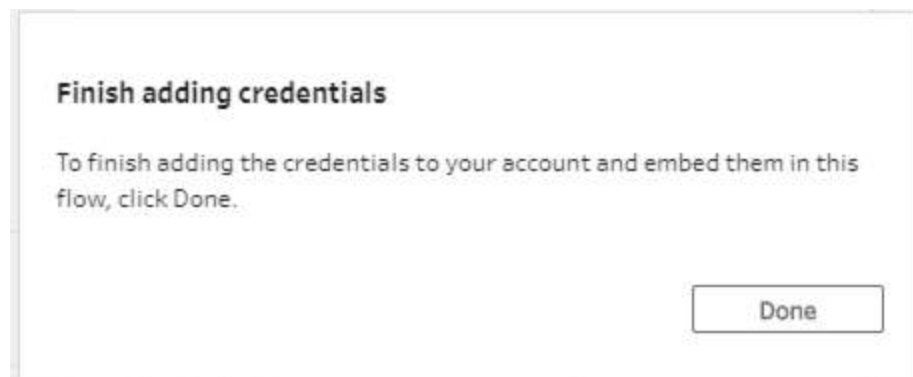
- In the confirmation dialog, click **Continue**. Tableau Prep Builder automatically opens the Account Settings page for the server you are signed into.



- Add your credentials, then navigate back to Tableau Prep Builder.

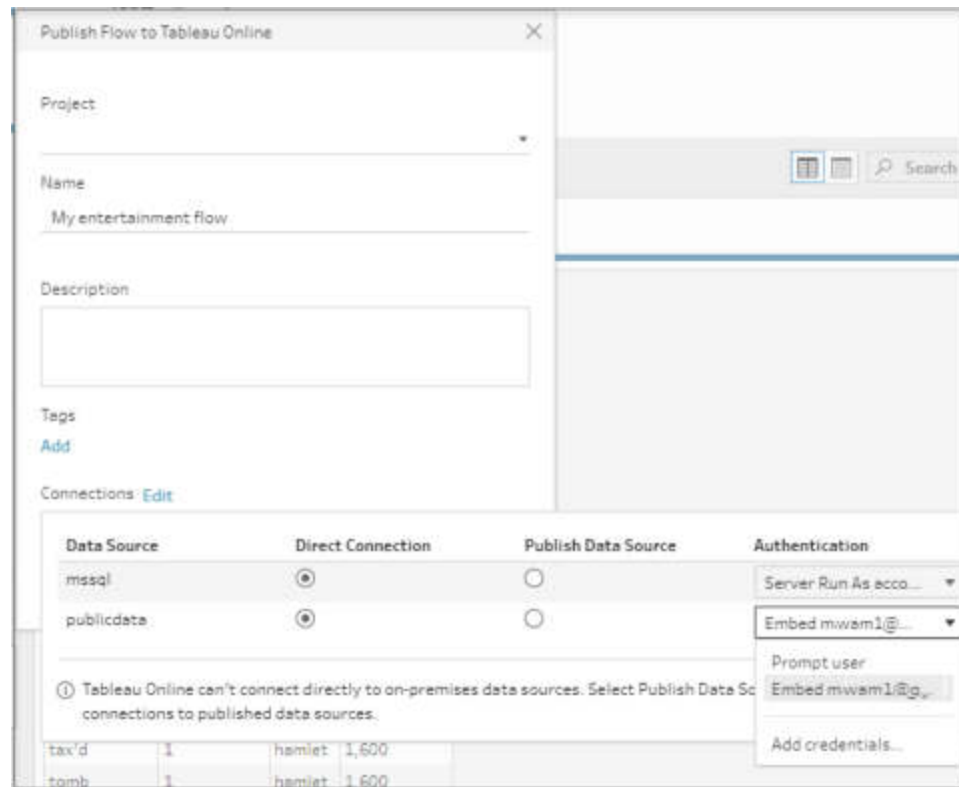


- In the **Finish adding credentials** dialog, click **Done**.



- Click **Edit** in the **Connections** section and verify that your credentials were added and embedded in your flow.





## Who can do this

Server Administrator, Site Administrator Creator, and Creator allow full connecting and publishing access.

Explorer (can publish) and Site Administrator Explorer have limited publishing capabilities, as described in the following topics:

### Tableau Server:

- Windows: [General capabilities allowed with each site role.](#)
- Linux: [General capabilities allowed with each site role.](#)

**Tableau Online:** [General capabilities allowed with each site role.](#)

## Schedule a Flow Task

To schedule flows to run at a specific time or on a recurring basis you can create scheduled tasks. Scheduled tasks rely on pre-configured schedules. In Tableau Server. Schedules are

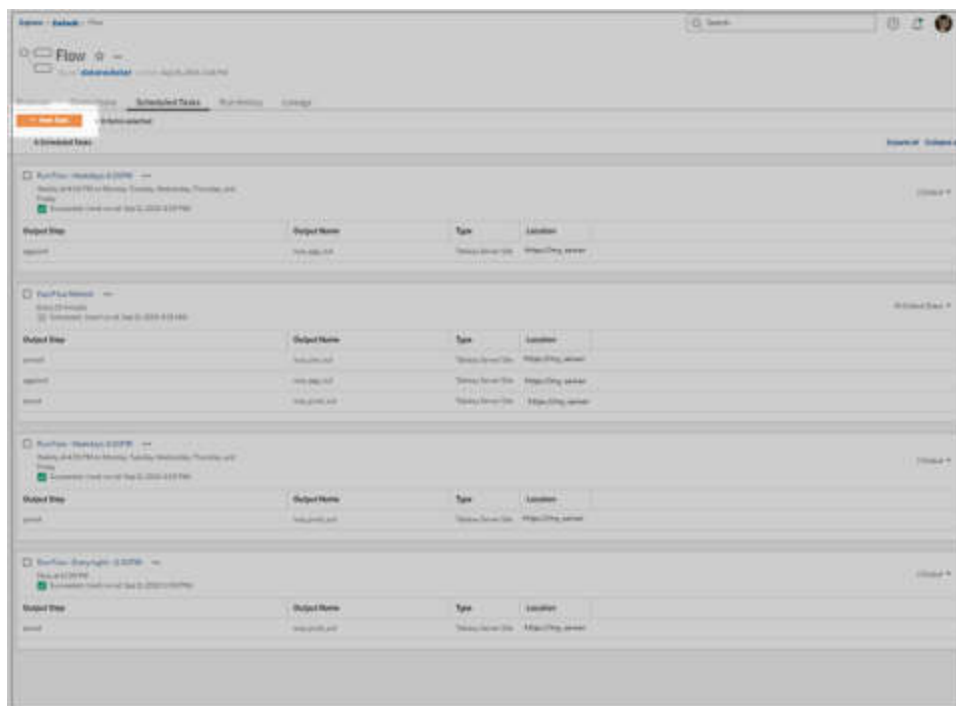
created by the System Administrator. For information about how to create schedules on Tableau Server, see [Step 3: Create Schedules for Flow Tasks on page 392](#).

In Tableau Online, to assign a flow to a schedule, select one of the pre-defined schedules provided with Tableau Prep Conductor. New schedules cannot be created in Tableau Online.

## Schedule a flow task

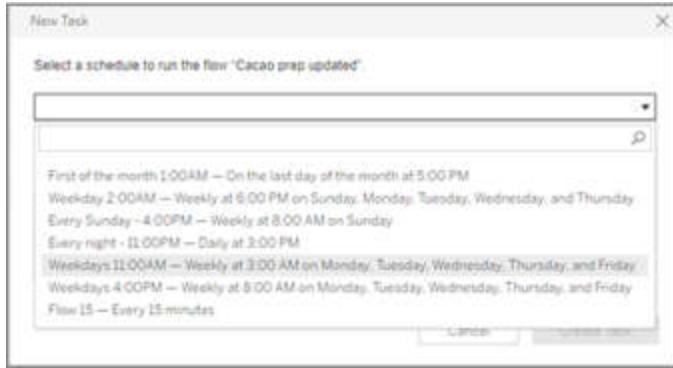
1. From the **Scheduled Tasks** tab, click **New Task**.

If the output step isn't assigned to a task, you can also create a new task from the **Overview** page. On that page, in the **Schedules** field, click **Create new task**.

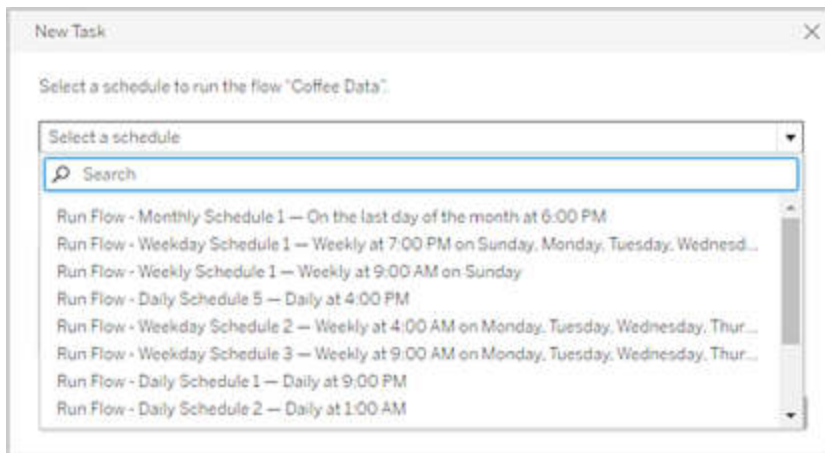


2. In the **New Task** dialog, select a schedule from the drop-down list.

### Tableau Server



### Tableau Online



3. Select one of the following options:

**Note:** Starting in version 2020.2.1 and later, if one input is configured to use incremental refresh and it is associated with multiple outputs, those outputs must be run together and must use the same refresh type. Otherwise the flow will fail.

- **Automatically include all output steps for this flow:**(default) Select this option to include all current and future output steps for this flow in the scheduled task. As new output steps are added to the flow over time, they are automatically included in the schedule when it runs.
- **Select the output steps to include in this task:** Select this option to manually select the output steps to include in this scheduled task. Then select the output steps to include in the flow task.

To include all output steps in the flow task, select the check box next to **Output Steps**. This area can't be edited if the **Automatically include all output steps for this flow** radio button is selected. Select the other radio button to enable this section.

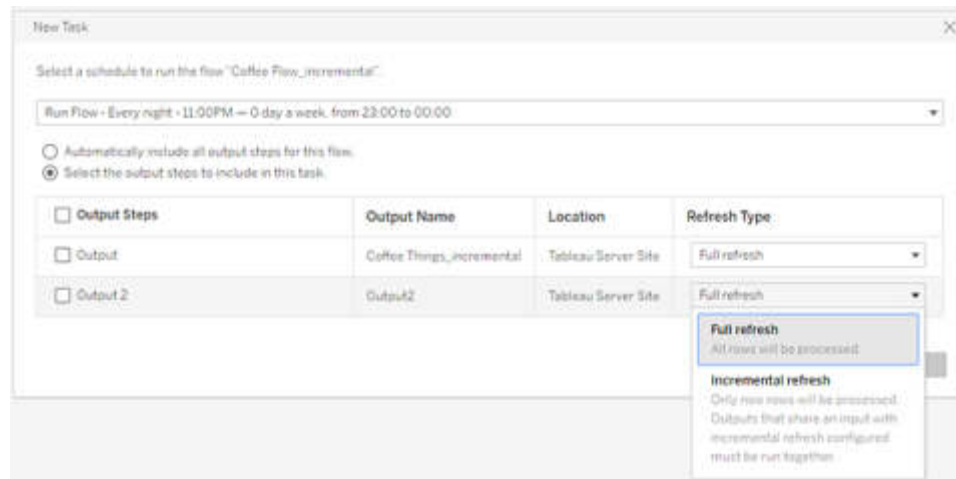
<input type="checkbox"/> Output Steps	Output Name	Location	Refresh Type
<input checked="" type="checkbox"/> Create 'Annual Regional Performance.tde'	Annual Regional Performance	Tableau Server Site	Full refresh
<input type="checkbox"/> Create 'Superstore Sales.tde'	Superstore Sales	Tableau Server Site	Full refresh

4. (version 2020.2.1 and later) Select a **Refresh Type**. For more information about these settings, see [Refresh Flow Data Using Incremental Refresh](#) on page 326.

- **Full refresh** (default): Refresh all data and create or append data to your table based on the output setting in Tableau Prep Builder.
- **Incremental refresh**: Refresh only the new rows and create or append data to your table based on the output setting in Tableau Prep Builder. The incremental refresh option is only available when the flow is configured in Tableau Prep Builder to use this refresh type.

**Note:** Tableau Prep Conductor runs a full refresh for all outputs regardless of the run option you select if no existing output is found. Subsequent flow runs will use the incremental refresh process and retrieve and process only your new rows unless incremental refresh configuration data is missing or

the existing output is removed.



5. Click **Create Task** to create the scheduled task.

## Who can do this

- The Server administrator can do this on all sites on the server. The Site administrator can do this on sites they have access to if the site settings to allow users to publish and schedule tasks is enabled.
- Flow owners and project leaders can create flow tasks for flows or projects that they own respectively.

For more information, review the following articles:

- Windows: [Set Users' Site Roles](#) and [Content Permissions and Ownership](#)
- Linux: [Set Users' Site Roles](#) and [Content Permissions and Ownership](#)

Site administrators, flow owners, project leaders, can create flow tasks for the flows and projects that they own respectively.

For more information, see [Set Users' Site Roles](#) and [Content Ownership and Permissions](#)

## Manage a Flow

Once you publish a flow to Tableau Server or Tableau Online, you can manage your flows and make changes to them as necessary. This topic describes the various actions you can take to manage your flows.

The content in this topic applies to both Tableau Server and Tableau Online, exceptions are called out specifically.

## Managing your flows

### Following is a list of actions you can take to manage your flows:

- **Run flows:** You can manually run a flow in addition to creating scheduled flows tasks that run at a specific time.

Navigate to the list of flows, select one or more flows you want to run, select Actions and click Run now.

- **Tag:** Tags are keywords you can create for flows to help you find, filter, and categorize content. Authors can add tags to flows when they publish it. But you can also add tags to any workbook, view, or data source that you are allowed to access and you can delete any tags you have added. You can add a tag to a list of flows.

Navigate to the list of flows, select one or more items you want to tag, select Actions and click Tag. To add a tag to a specific flow you can do it from the list of flows as described above. Open the Flow, from the Overview tab, select Actions, and click Tag.

- **Change Owner:** Administrators, flow owners, and project leaders can change owners.
- **Permissions:** You can set permissions for users and specify if they can perform edit actions like save, download, move to a different project and delete. In addition, you can specify who can view and run the flow.
- **Download:** You can download a flow to view or modify it using your Tableau Prep. To download a flow, you need download permissions. You'll have that by default if you are the owner, but you might need to add it for other users.
- **Revision History:** When you make a change to the flow, and republish it to the same project with the same name, a new version of the flow is created. You can view the revision history by selecting revision history from the actions menu. Flow owners have permissions to restore a previous version of a flow.
- **Move:** You can move flows between projects. To move a flow, users need Move permission. You'll have that by default for flows you own, but you might need to add it for other users.
- **Rename:** You can rename a flow. To rename a flow, users need the Save permission.

You'll have that by default for flows you own, but you might need to add it for other users.

- **Delete:** You can delete a flow. To delete a flow, users need the Delete permission. You'll have that by default for flows you own, but you might need to add it for other users.

## Who can do this

# Tableau Server Administrator

### Can do the following tasks across all the sites:

- Run flows
- Delete flows
- Download flows
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
- Create a Flow Task\*
- Version Management\*

\*Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:

- Revision history must be enabled on the site.
- User role is allowed to publish to the project.

## Tableau Site Administrator

**Can do the following tasks on flows published to the sites that they are site administrator for:**

- Run flows
- Delete flows
- Download flows
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
- Create a Flow Task\*
- Version Management\*

Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:



- Revision history must be enabled on the site.
- User role is allowed to publish to the project.

## Project Leader

**Can do the following tasks on flows published to the projects where they have project leader permissions:**

- Run Flows
- Delete
- Download
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
- Create a Flow Task\*
- Version Management\*

\*Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:

- Revision history must be enabled on the site.
- User role is allowed to publish to the project.

## Project Owner

**Can do the following tasks on flows published to the projects that they own:**

- Run Flows
- Delete
- Download
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
- Create a Flow Task\*
- Version Management\*

\* Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

## Flow Owner

### Can do the following tasks on flows that they own:

- Run flows
- Delete flows
- Download flows
- Change Owner
- Change Permissions
- Change Project
- Add/Remove Tags
- Change Description
- Change Name
- Update Flow Task
- Delete Flow Task
- Create a Flow Task\*
- Version Management\*

\*Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

## User with Creator License

### Can do the following tasks:

- Run flows (with Run flow permissions)
- Delete (with delete permissions)
- Download (with download or Save as, and read permissions)
- Change Permissions (with Change Permissions)
- Change Project (with move permissions, and write permissions on the destination project)
- Add/Remove Tags (with read permissions)
- Change Description (with Save permissions)
- Change Name (with Save permissions)
- Update Flow task (with Execute permissions)
- Delete Flow Task (with Execute permissions)
- Create a Flow Task\* (with Execute permissions)
- Version Management\* (with view, read, save as, download permissions)

\* Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

## User with Explorer License

### Can do the following tasks:

- Run flows (with Run flow permissions)
- Delete (with delete permissions)
- Download (with download or Save as, and read permissions)

- Change Permissions (with Change Permissions)
- Change Project (with move permissions and write permissions on the destination project)
- Add/Remove Tags (with read permissions)
- Change Description (with Save permissions)
- Change Name (with Save permissions)
- Update Flow task (with Execute permissions)
- Delete Flow Task (with Execute permissions)
- Create a Flow Task\* (with Execute permissions)
- Version Management\* (with view, read, save as, download permissions)

\*Some additional conditions apply to these actions:

- To create a flow task:
  - A flow schedule must be available. Only Server Administrators can create a schedule.
  - The flow must have at least one output step.
  - The flow version must be compatible with the Tableau Server version.
- Version management:
  - Revision history must be enabled on the site.
  - User role is allowed to publish to the project.

## User with Viewer License

Viewers cannot manage flows, they can however view the flow and the different versions of the flow.

For more information about the full capabilities you can set on flows, see [Permission capabilities](#).

## Monitor Flow Health and Performance

After you publish flows and schedule them to run periodically, you want to know that they are running as expected and resolve any issues as they occur. You will also want to monitor and understand the performance of your flows.

This topic describes the various methods that Tableau Server and Tableau Online provide to help you monitor your flows.

## Detect issues as they occur and resolve them

You can set up Tableau Server to send email notifications when flows fail, or find and review errors on Tableau Server using the Alerts menu or by reviewing the flow pages for the flows that you are interested in. This type of monitoring allows you to detect problems as they occur.

### Get email notifications when a flow fails:

You can configure Tableau Server to send email notifications to flow owners when a flow fails.

To set up email notifications for your Tableau Server and sites, follow the instructions in [Step 2: Configure Flow Settings for your Tableau Server on page 387](#).

Email notifications is turned on by default for your site in Tableau Online. You can verify this by going to **Settings > General** page.

### View and resolve errors

The following errors can happen when running a flow:

- **Connection errors:** Connection errors generally happen when Tableau Server or Tableau Online is unable to connect to one or more data inputs or is unable to make a connection in one or more output steps.
  - For Input connection errors, use the **Edit connections** option on the **Connections** tab to make changes to connection details, then run the flow again.
  - For output connection errors, download the flow and check the output location for the flow output steps. If the flow output is going to a network share, make sure the output steps are pointing to a safe listed location. After you make any changes republish the flow and try running it again.
- **Errors in the flow:** If there are errors in one or more steps in the flow, you will see an error message. In such cases, download the flow to Tableau Prep Builder, resolve the errors, republish the flow to the server and run the flow again.
- **Suspended flow tasks:** When a scheduled flow task fails to run after a configured number of attempts, the flow task is suspended. A flow can have multiple scheduled tasks assigned to it, but only the failed tasks are suspended. All other flow tasks will continue to run unless they have errors. To resolve a suspended task, review and

resolve the errors then run the flow on-demand or let the flow run automatically based on the assigned schedule.

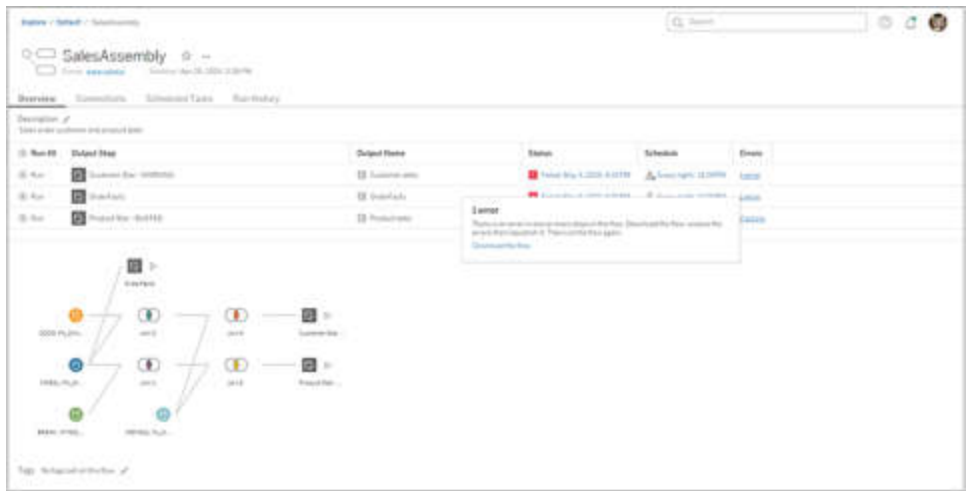
**Note:** By default, a flow task is suspended after 5 consecutive flow tasks failures. Server administrators can change this default by using the tsm configuration set option. For more details, see [Step 5: Optional Server Configurations on page 396](#).

You can view errors on the following pages:

### Flow Overview page

On this page you can see the status of the most recent flow run and any errors. Hover on the error text to review the error details. If a scheduled task is suspended, a warning icon shows next to the schedule. Hover on the icon to view the status.

After you resolve the error that caused the flow to fail or the task to be suspended, you can run the flow manually or let the flow run based on the assigned schedule. For suspended flow tasks, click the **Go to Scheduled Task** link on the tooltip for the suspended task to navigate to the **Scheduled Tasks** page and click the **Resume Scheduled Tasks** button to resume the suspended tasks.



### Connections page

The **Connections** page shows the most recent status and any related connectivity errors. To correct input errors, click the **More actions** \*\*\* menu for an input connection to edit the

connection and change the server name, port, user name and password.

To fix output connection errors, download the flow in Tableau Prep Builder, correct the file path, then republish the flow to continue running it.

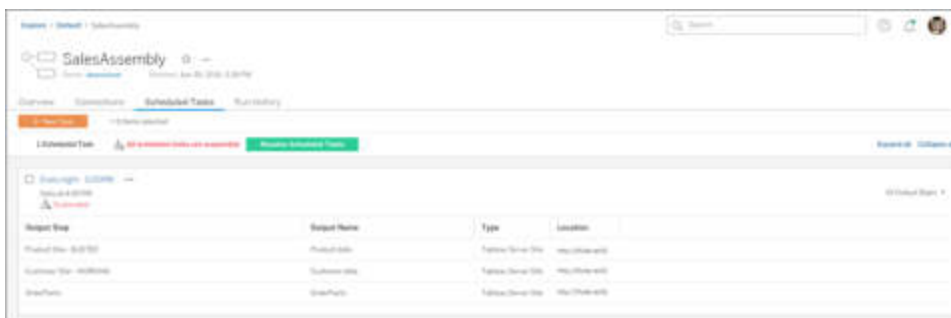


## Scheduled Tasks page

View the scheduled tasks assigned to a flow. If a scheduled task is suspended, you can see the status of that tasks here and you can manually resume the flow tasks from this page. Before resuming a suspended task, resolve any errors in the flow. Errors are not shown on this page, but you can review them on the **Overview** or **Run History** pages.

A suspended scheduled task will automatically resume when the flow is republished, if you edit a connection for the flow or manually run the flow tasks. To manually resume a suspended tasks, on the **Scheduled Tasks** page, click **Resume Scheduled Tasks**. This resumes all suspended tasks for the flow.

To resume individual tasks, click the **More actions** ... menu for a scheduled task and select **Resume**. You can also click **Run Now** to run all tasks for the flow immediately.



## Run History page

The **Run History** page shows the details of all the flow runs that have either completed or are in progress for each output. View any error details by hovering over the errors in the **Errors** column. The duration column shows you the run time of the flow.



**Note:** Starting in version 2020.2.1, the **Run Type** field shows the refresh type for the output. In prior releases this field showed whether the output was run on a schedule or on-demand. For more information about setting up output refresh types, see [Refresh Flow Data Using Incremental Refresh](#) on page 326.

The screenshot shows the 'Run History' tab for a flow named 'Coffee Flow\_incremental'. The table below lists the run history:

Output Step	Run Type	Run Start	Run End	Duration	Status	Rows Generated	Errors
Output 2	Incremental refresh	Apr 30, 2020 11:00 PM	Apr 30, 2020 11:00 PM	00:00:00	Succeeded	4,240	
Output	Incremental refresh	Apr 30, 2020 11:00 PM	Apr 30, 2020 11:00 PM	00:00:09	Succeeded	4,240	
Output	Full refresh	Apr 21, 2020 11:40 AM	Apr 21, 2020 11:40 AM	00:00:03	Failed	0	<a href="#">Error</a>

An error message is displayed for the failed run:

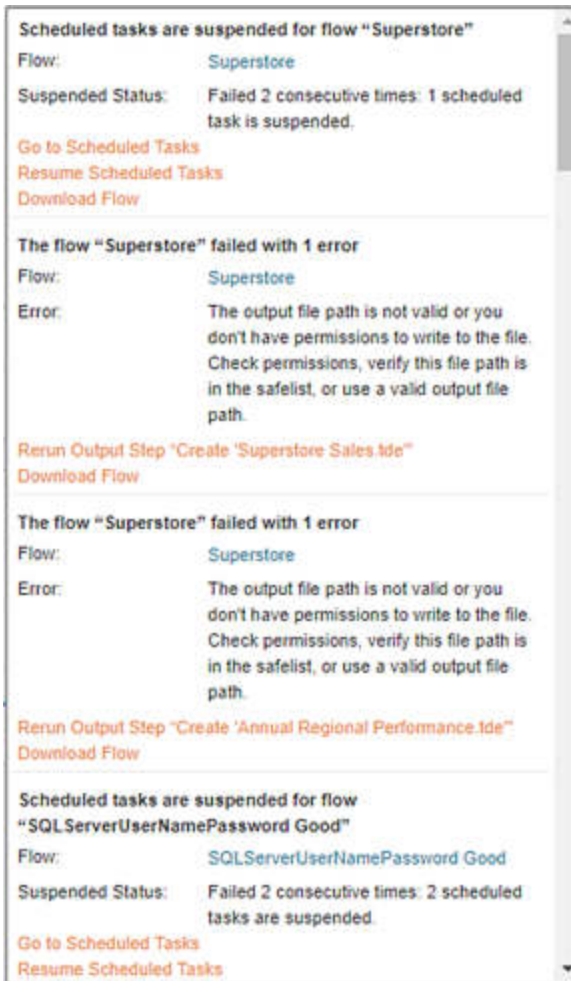
```

Error
Outputs that share an input and are set up to refresh incrementally must use the same run setting. The following outputs must all be run using the incremental refresh settings or must all be run as a full refresh: Outputs: "Output, Output 2"
    
```

## Alerts

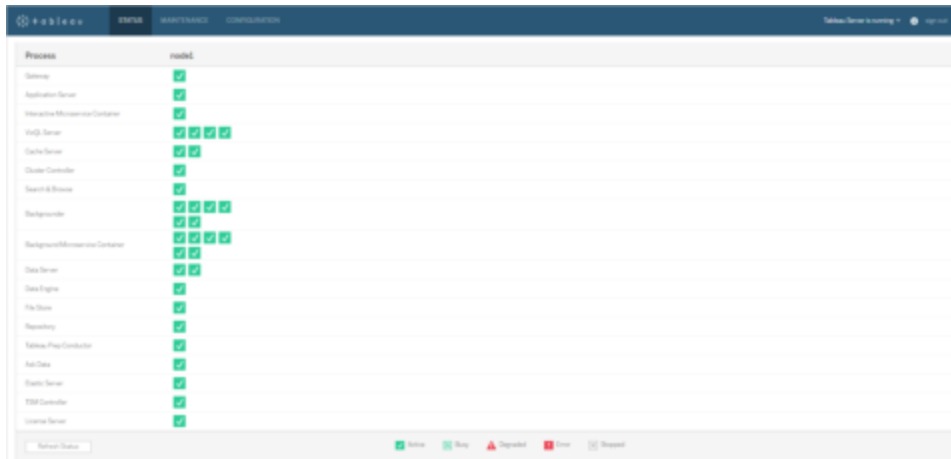
When a flow fails, the alerts menu is populated with the error details with the option to re-run the flow, or download the flow to troubleshoot.

**Note:** In Tableau Server, Flow owners, Server or Site Administrators can see this menu. In Tableau Online, Flow owners and Site Administrators can see this menu.



## Tableau Prep Conductor process status

- The Tableau Services Manager (TSM) status page is accessible in TSM and can be viewed by TSM administrators. You must be able to log into TSM to see this page.
- The Tableau Server status page appears in the Tableau Server web UI and is accessible by Tableau Server administrators. If you hover your mouse pointer over the status indicator for a process, a tooltip shows the node name and the port the process is running on. The Tableau Server status page does not show TSM processes.



When Tableau Server is functioning properly, Tableau Prep Conductor will show as Active or Busy:

- **Active**—The process is functioning as intended.
- **Busy**—The process is completing some task.
- **Down**—The process is down. The implications of this differ depending on the process.
- **Status unavailable**—Tableau Server is unable to determine the status of the process.

## Monitor flow activity and performance with Administrative Views

Administrative views can be used to monitor the activities related to flows, performance history, and the disk space used. The **Status** page contains an embedded Tableau workbook with various administrative views that can be used to monitor different types of server or site activity.

For **Tableau Server** installations, you must install PostgreSQL drivers before you can see Administrative views. For more information, see [Database Drivers \(Linux\)](#), [Database Drivers \(Windows\)](#). Server administrators can use these views to see activity both at the server level (aggregated for all sites) or for a specific site. Only server administrators can filter by site.

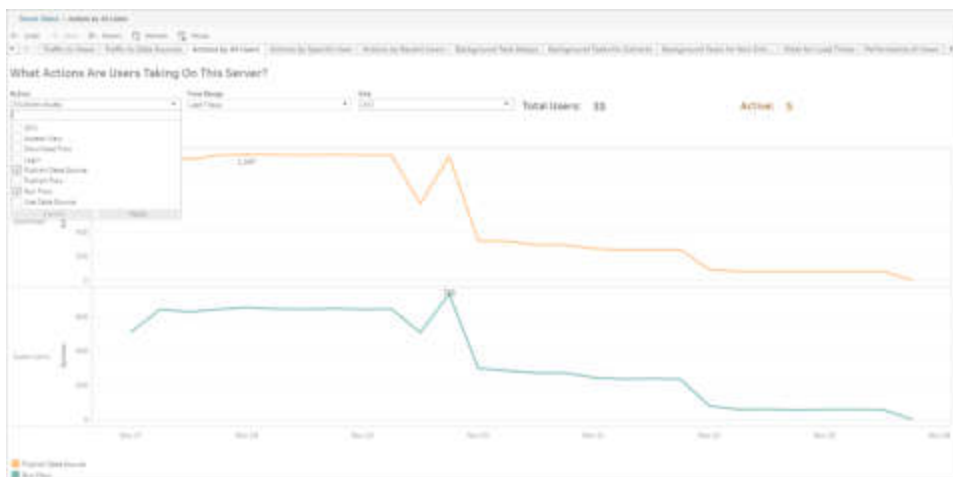
### Who can do this?

- **Tableau Server:** Tableau Server administrators and Tableau Site administrators can both view and work with Administrative Views. Only Server administrators can filter by site.

- **Tableau Online:** Tableau Site administrators can view and work with Administrative Views.

## Action by all users

Use this view to gather insight into how flows are being used. This includes actions like publish, download, and flow runs. You can filter the view by actions, by site, and by time range. The Total Users count shows the number of users who have performed an action. This value is not affected by any filtering. The Active user count shows the number of users who have been active during the selected time period and performed one of the selected actions.



## Action by Specific User

Use this view to gather insights about how an individual user is working with flows.

In Tableau Server, you can filter the view by user name, the type of action, time range, and by site.

In Tableau Online, you can filter the view by user name, the type of action, and by time range.



### Action by Recent Users

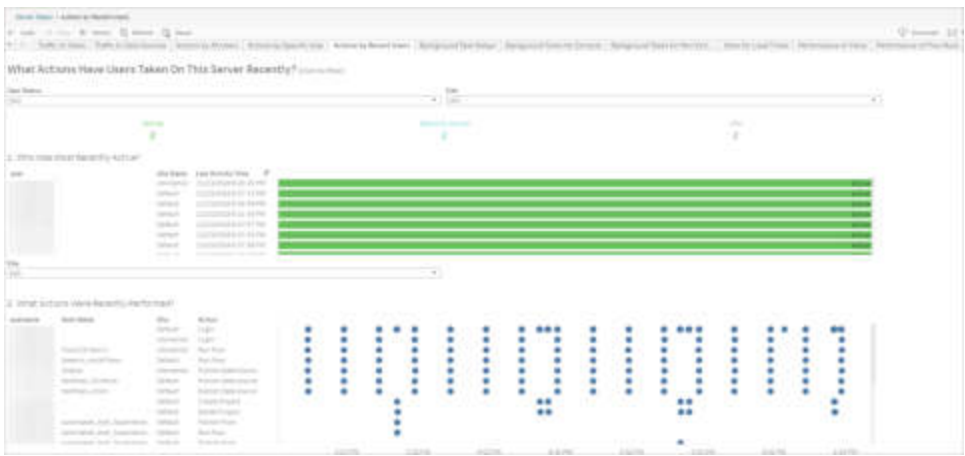
This view shows you which users have been active on Tableau Server or Tableau Online over the past 24 hours.

This can be useful if you need to do some maintenance activity on the server and want to know how many and which users this will affect, and what they are doing.

The view **Active**, **Recently Active**, and **Idle** users that are currently signed in to Tableau Server or Tableau Online.

For this view, an active user is one who took an action in the last 5 minutes, a recently active user is one who last took an action within 30 minutes, and an idle user is one who last took an action more than 30 minutes ago.

Select a user to see only the actions that user performed recently. Hover over an action to see details of the action.



## Backgrounder Task Delays

The Background Task Delay view displays the delay for extract refresh tasks, subscription, and flow tasks—that is, the amount of time between when they are scheduled to run and when they actually run. You can use the view to help you identify places you can improve server performance by distributing your task schedules and by optimizing tasks.



Possible reasons for the delays and ways that you might reduce the delays include the following:

- Many tasks are scheduled for the same time. In the example view, tasks that show long delays are clustered at the same time every day, which creates spikes in the wait time. Note that you can set the Timeline filter to a single day to view task delays by hour and identify the hours of the day which have many tasks scheduled at the same time. A solution to this issue can be to distribute the tasks to off-peak hours to reduce load on the server.
- In Tableau Server, other server processes are running at the same time and are consuming server resources and slowing down performance. Monitor the CPU and memory usage of server processes to see which processes are consuming the most resources and then adjust the configuration of processes on your server. For more information on monitoring processes, see [Collect Data with Windows Performance Monitor](#).

## Background Tasks for Non Extracts

Background Tasks are created to run flows (scheduled and ad hoc). You can use this view to see how many flow tasks succeeded or failed on this site. For details on a task, hover over its icon.



### Performance of Flow Runs

Use this view to see the performance history for all the flows on a site. You can filter by Flow Name, the Output Step Name, Flow Owner, Run Type (Scheduled or Ad Hoc), and the time the flow runs were started.



Here are some questions you can answer using this view:

- **What flow tasks are currently scheduled?** – To do this, use the Start Time filter and select the time frame you want to look at. For example, to see flow tasks that are scheduled in the next 3 hours, select **Hours** -> **Next** -> and enter **3**.
- **What is the duration of flow tasks?** - To answer this, click on a mark in the view and you should see details including the task duration.

**How many flows were run ad hoc, and how many were scheduled runs?** - To answer this, use the **Run Type** filter and select **Ad hoc** or **Scheduled**.

**Note:** This is not functional in this release and will not actually filter the data.

In addition to the questions described above, here are some examples of insights you might be able to gather:

- The flows that are running most frequently will have the most marks.
- To see how many flows are running at the same time currently, hover over a mark that shows **“In Progress”** or **“Pending and select “Keep Only”** to filter all flow runs that are currently running.
- To see how many flows are running at the same time during a specific time range, select a range for the **Start Time** filter. For example, you can choose **“Next three hours”** to see which flows will be running in the next three hours.

## Stats for Space Usage

The Stats for Space Usage view can help you identify which flow outputs are taking up the most disk space on the server. Disk space usage is displayed by user, project, and by the size of flow output and is rounded down to the nearest number.



Use the Min Size filter to control which flow outputs are displayed, based on the amount of space they take up. Use the object type filter for flows.

- **What Users Use the Most Space** – This section shows the users who own flows (when filtered for flows) that are taking up the most space. Click a user name to filter the next two graphs for that user.
- **What Projects Use the Most Space** – This section shows the projects with flows (when filtered for flows) that are using the most space.
- **What Workbooks, Data Source and Flows Use the Most Space** – This section shows the flows (when filtered for flows) that take up the most space.



## Who can do this

- **Tableau Server Administrators:**
  - Set up email notifications at the server level
  - Set up email notifications for a site
  - View errors
  - Resume suspended tasks
  - View alerts
  - view process status
- **Tableau Site Administrators:**
  - Set up email notifications at the site level
  - View errors
  - Resume suspended tasks
  - View alerts
- **Flow owners, project leaders and any user who is granted permissions to view the flow:**
  - View errors
  - Resume suspended tasks
  - View alerts (Flow owners)

## Developer Resources - REST APIs

Use Tableau Server REST APIs to automate and seamlessly integrate Tableau into your existing workflows. Tableau Server REST API gives you programmatic access to work with your content, users, sites, and now flows. Manage provisioning, permissions, and publishing on Tableau Server or Tableau Online via HTTP. The REST API gives you access to the functionality behind the data sources, projects, workbooks, site users, sites, and flows. You can use this access to create custom applications or to script interactions with server resources.

Tableau REST API supports the following flow functionality:

New flow endpoints have been added to support publishing flows, scheduling flows, running flows on demand, managing permissions, downloading flows and such. For a full list of all the new REST API endpoints for flows, see [Flow Methods](#).

In addition, [existing endpoints](#) have been updated to support flow functionality such as Creating New Schedules for Flows, Creating new sites, updating existing sites, and managing default permissions.



# Day in the Life Scenarios

What does it mean to shape data? How does that impact what visualizations can be built and what analysis can be performed? In the tutorials below, we explore scenarios for analysis and visualization, identify the data limitations holding us back, then see how Tableau Prep can help us shape the data to reach our intended outcome.

Download the data sets and follow along with these day in the life scenarios using Tableau Prep and Tableau Desktop. Learn how to apply the features and functions in Tableau Prep to get your data ready for analysis in Tableau Desktop.

**Give us your feedback.** We are just starting to build this section of the online help. If there are specific scenarios you'd love to see here, please let us know. Use the feedback bar at the top of the page to tell us more.

To complete the tasks in these tutorials, you need Tableau Prep and Tableau Desktop installed, and you'll need to download and save the data to your computer.

For information about how to install Tableau Prep and Tableau Desktop, see [Install Tableau Desktop or Tableau Prep Builder from the User Interface](#) in the Tableau Desktop and Tableau Prep Deployment guide. Otherwise you can download the [Tableau Prep](#) and [Tableau Desktop](#) free trials.

## Hospital Bed Use with Tableau Prep

Reaching capacity in a hospital is problematic but so is an overabundance of resources. It's important to understand hospital beds from the perspective of the bed as a resource. However, the data is often stored from the perspective of a patient. How can we take data that captures when patients are in beds and determine the bed usage?

**Note:** To complete the tasks in these tutorials, you need Tableau Prep and optionally Tableau Desktop installed:

To install Tableau Prep and Tableau Desktop see the [Tableau Desktop and Tableau Prep Deployment guide](#). Otherwise you can download the [Tableau Prep](#) and [Tableau Desktop](#) free trials.

You will also need to download three data files. It is recommended to save them in your My Tableau Prep Repository > Datasources folder.

- [Beds.xlsx](#)
- [Hours.xlsx](#)
- [Patient Beds.xlsx](#)

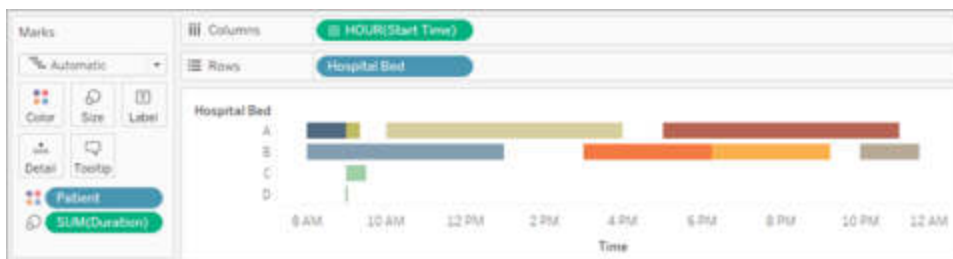
## The Data

For our four beds, A, B, C, and D, we track which patient was in the bed and their start and end time there. The data looks like this:

	A	B	C	D
1	<b>Hospital Bed</b>	<b>Patient</b>	<b>Start Time</b>	<b>End Time</b>
2	A	Person 1	1/1/2018 8:34	1/1/2018 9:34
3	A	Person 5	1/1/2018 9:55	1/1/2018 10:15
4	A	Person 9	1/1/2018 10:34	1/1/2018 16:34
5	A	Person 8	1/1/2018 17:00	1/1/2018 23:00
6	B	Person 2	1/1/2018 8:45	1/1/2018 13:45
7	B	Person 6	1/1/2018 15:13	1/1/2018 18:27
8	B	Person 7	1/1/2018 18:41	1/1/2018 21:56
9	B	Person 10	1/1/2018 22:13	1/1/2018 23:43
10	C	Person 3	1/1/2018 9:05	1/1/2018 9:35
11	D	Person 4	1/1/2018 9:30	

## Preliminary Analysis

If we bring this data into Tableau Desktop, we can create a Gantt chart to show when patients are in beds.



This is a useful visual. We can see that there are only small gaps in use for beds A and B, but bed C is very under-used. Bed D's patient has no end time, but we could address that with some calculations. This gives us a visual overview of how the beds are used.

However, what if we wanted to count the hours when a bed was empty? Or compare open bed time before and after a new policy is put in place? There's no easy way to do that with the data as it's currently structured.

## Desired Data Structure

By creating some very basic data sets and combining them in Tableau Prep, we can modify this data set into a form that will allow us to perform deeper analysis and create even more useful visualizations.

Before we jump into Tableau Prep, let's step back and think about what we need to create to answer the question, "How many hours was each bed empty?"

We need to be able to look at each bed for each hour, and know whether or not there was a patient in the bed. Right now, the data is solely when a patient was in the bed; we haven't given Tableau information about the *empty* hours.

To create that full matrix of all beds and all hours, we'll create two new data sets. One is simply a list of beds (A, B, C, D) and the other is a list of hours (1, 2, 3, ..., 23, 24). By performing a cross join (joining every row in one data set with every row in the other data set) we'll wind up with every possible combination of beds and hours.

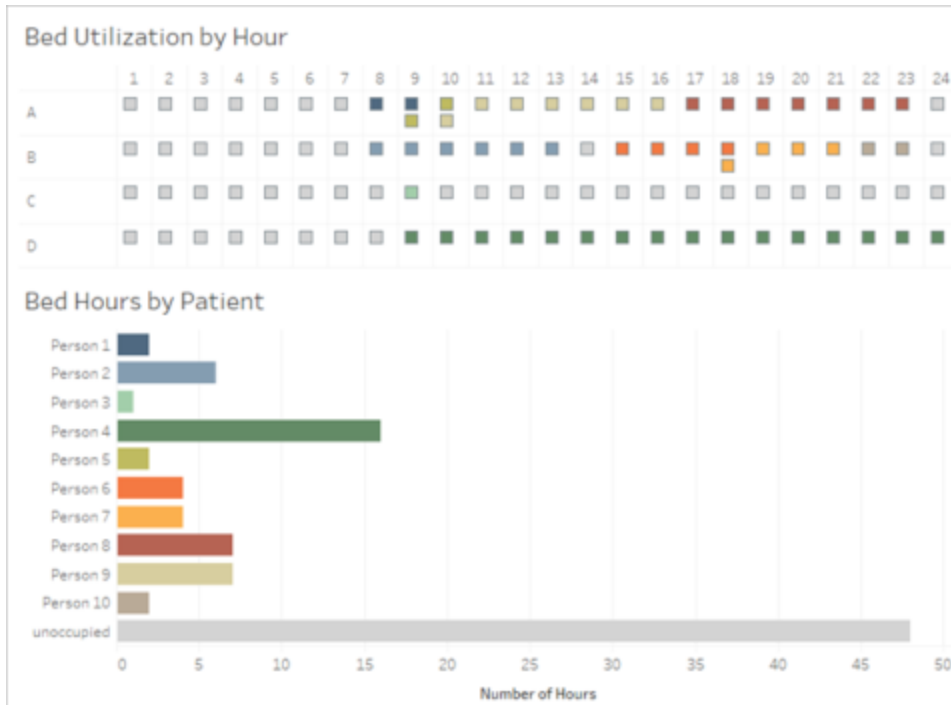
The <b>Beds.xlsx</b> data set looks like this:	The <b>Hours.xlsx</b> data set looks like this:	And the cross joined results look like this:																																																							
<table border="1"> <thead> <tr> <th></th> <th>A</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Bed</td> </tr> <tr> <td>2</td> <td>A</td> </tr> <tr> <td>3</td> <td>B</td> </tr> <tr> <td>4</td> <td>C</td> </tr> <tr> <td>5</td> <td>D</td> </tr> </tbody> </table>		A	1	Bed	2	A	3	B	4	C	5	D	<table border="1"> <thead> <tr> <th></th> <th>A</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Hour</td> </tr> <tr> <td>2</td> <td>1</td> </tr> <tr> <td>3</td> <td>2</td> </tr> <tr> <td>4</td> <td>3</td> </tr> <tr> <td>5</td> <td>4</td> </tr> <tr> <td>6</td> <td>5</td> </tr> <tr> <td>7</td> <td>6</td> </tr> </tbody> </table>		A	1	Hour	2	1	3	2	4	3	5	4	6	5	7	6	<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>Bed</td> <td>Hour</td> </tr> <tr> <td>2</td> <td>A</td> <td>1</td> </tr> <tr> <td>3</td> <td>B</td> <td>1</td> </tr> <tr> <td>4</td> <td>C</td> <td>1</td> </tr> <tr> <td>5</td> <td>D</td> <td>1</td> </tr> <tr> <td>6</td> <td>A</td> <td>2</td> </tr> <tr> <td>7</td> <td>B</td> <td>2</td> </tr> <tr> <td>8</td> <td>C</td> <td>2</td> </tr> </tbody> </table>		A	B	1	Bed	Hour	2	A	1	3	B	1	4	C	1	5	D	1	6	A	2	7	B	2	8	C	2
	A																																																								
1	Bed																																																								
2	A																																																								
3	B																																																								
4	C																																																								
5	D																																																								
	A																																																								
1	Hour																																																								
2	1																																																								
3	2																																																								
4	3																																																								
5	4																																																								
6	5																																																								
7	6																																																								
	A	B																																																							
1	Bed	Hour																																																							
2	A	1																																																							
3	B	1																																																							
4	C	1																																																							
5	D	1																																																							
6	A	2																																																							
7	B	2																																																							
8	C	2																																																							

Next, we'll bring in the **Patient Beds** information, labeling each bed-hour combination as having a specific patient or not. We wind up with a data set that has a row for each bed-hour,

and if a patient was in the bed, their number and start and end times. Null values indicate the bed was unoccupied.

	A	B	C	D	E
1	Bed	Hour	Patient	Start Time	End Time
29	D	7			
30	A	8	Person 1	1/1/2018 8:34	1/1/2018 9:34
31	B	8	Person 2	1/1/2018 8:45	1/1/2018 13:45
32	C	8			
33	D	8			
34	A	9	Person 5	1/1/2018 9:55	1/1/2018 10:15
35	A	9	Person 1	1/1/2018 8:34	1/1/2018 9:34
36	B	9	Person 2	1/1/2018 8:45	1/1/2018 13:45
37	C	9	Person 3	1/1/2018 9:05	1/1/2018 9:35
38	D	9	Person 4	1/1/2018 9:30	
39	A	10	Person 9	1/1/2018 10:34	1/1/2018 16:34
40	A	10	Person 5	1/1/2018 9:55	1/1/2018 10:15
41	B	10	Person 2	1/1/2018 8:45	1/1/2018 13:45
42	C	10			
43	D	10	Person 4	1/1/2018 9:30	
44	A	11	Person 9	1/1/2018 10:34	1/1/2018 16:34

With the data in this structure, we can perform analyses like this, which enables us to investigate unoccupied beds as easily as patient beds.



## Restructuring the Data

So how do we get there with Tableau Prep? We'll build out the flow in two parts, first building the Bed Hours matrix, then combining it with the Patient Beds data. Make sure you've downloaded all three Excel files (**Beds.xlsx**, **Hours.xlsx**, and **Patient Beds.xlsx**) to follow along.


### Bed Hour Matrix

First, we'll connect to the **Beds.xlsx** file.

1. Open Tableau Prep.
2. From the start screen, click **Connect to Data**.
3. On the **Connections** pane, click **Microsoft Excel**. Navigate to where you saved **Beds.xlsx** and click **Open**.
4. The **Beds** sheet should automatically be brought out to the **Flow** pane.

**Tip:** For more information about connecting to data, see [Connect to Data on page 145](#).



Next, we need to create a field we can use to do the cross join with the **Hours** data set. We'll add a calculation that is simply the value **1**.

5. In the **Flow** pane, select **Beds**, click the plus  icon, and select **Add Step**.
6. With the **Clean** step we just added, the **Profile** pane will come up. Click **Create Calculated Field** in the toolbar.
7. Name the field **Cross Join** and enter the value **1**.
8. The **Data** grid should update show the current state of the data.

Cross Join	Bed
1	A
1	B
1	C
1	D

Now we'll repeat the process with the Hours data set.

## Click for directions

9. On the **Connections** pane, click the **Add connection**  button to add another data connection.
10. Choose **Microsoft Excel** and then select the **Hours.xlsx** file and click **Open**.
11. In the **Flow** pane, select **Hours**, click the plus  icon, and select **Add Step**.
12. From the toolbar in the **Profile** pane, create a calculated field named **Cross Join** and enter the value **1**.

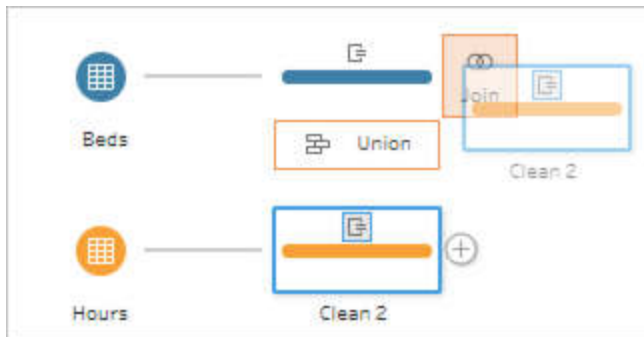


## Tableau Prep Help

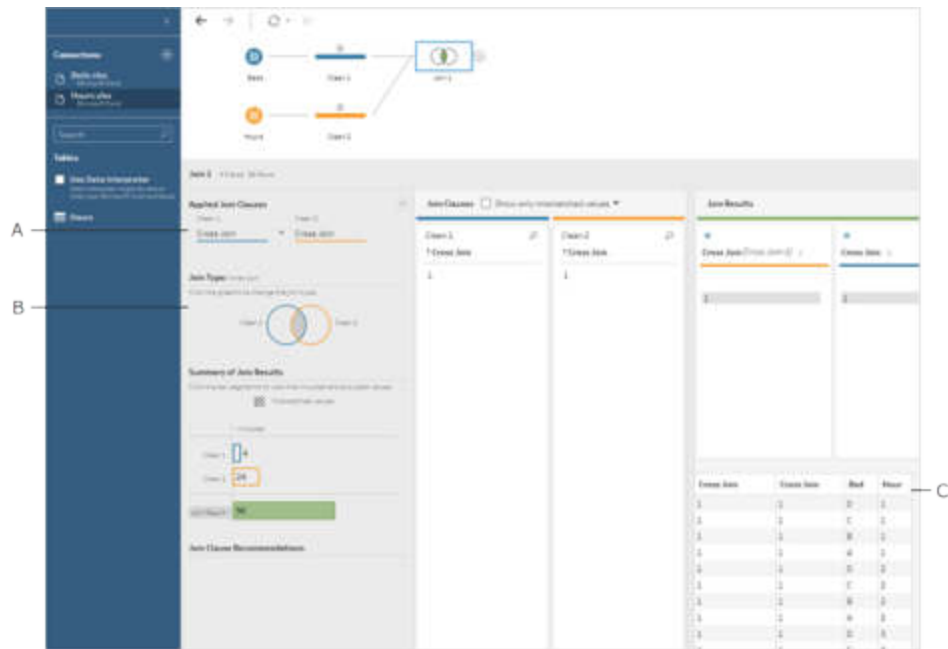
Cross Join	Hour
1	1
1	2
1	3
1	4
1	5

Both data sets now have a shared field, **Cross Join**, and can be joined.

13. Join the two cleaning steps by dragging **Clean 2** onto **Clean 1** and dropping it on the **Join** option.



14. In the **Join Profile** below, the join configurations should populate automatically.
  - Because we named both fields **Cross Join**, Tableau Prep automatically identifies them as the shared field and creates the appropriate **Applied Join Clauses**.
  - The default **Join Type** is inner, which is what we want.
  - This join will match all rows from **Beds** with all rows from **Hours**, as seen in the **Data** grid.



A. Join clause, B. Join type, C. Data grid results

**Tip:** For more information about joins, see [Join your data on page 308](#).

We no longer need the **Cross Join** fields, so we can remove them.


15. In the **Flow** pane, select **Join 1**, click the plus  $\oplus$  icon, and select **Add Step**.
16. Select the fields **Cross Join-1** and **Cross Join**, and click **Remove Fields**.
17. Double click on the **Clean 3** label and rename that step **Bed Hour Matrix**.

We now have the Bed Hour Matrix data set that contains all beds and all hours and have finished the first part of building our data set.

## Patient Bed Use

Part two is bringing in the patient bed usage. To start, we'll connect to the data.

1. On the **Connections** pane, click the **Add connection**  $\oplus$  button to add another data connection.

2. Choose **Microsoft Excel** and then select the **Patient Beds.xlsx** file, and click **Open**.
3. In the **Flow** pane, select **Patient Beds**, click the plus  icon and select **Add Step**.

Because the Bed Hour Matrix file is based on *hour* but Patient Beds is based on *actual time*, we need to pull the hour out of the Patient Beds start and end times. Additionally, for the end time, we want to ensure that if a patient is still in the bed at the end of the day (midnight, hour 24) we indicate that the bed is occupied even though there's no end time in the data set. We'll add a calculated field in this new step.

4. In the toolbar, click **Create Calculated Field**.
5. Name the field **Start Hour**. For the calculation, enter `DATEPART('hour', [Start Time])`.

This takes the hour of the start time and pulls it out. Therefore, "1/1/18 9:35 AM" becomes simply "9".

6. Create another calculated field named **End Hour**. For the calculation, enter `IFNULL(DATEPART('hour', [End Time]), 24)`.

The `DATEPART` portion takes the hour of the end time. The `IFNULL` portion will assign an end time of 24 (midnight) to any missing end time.

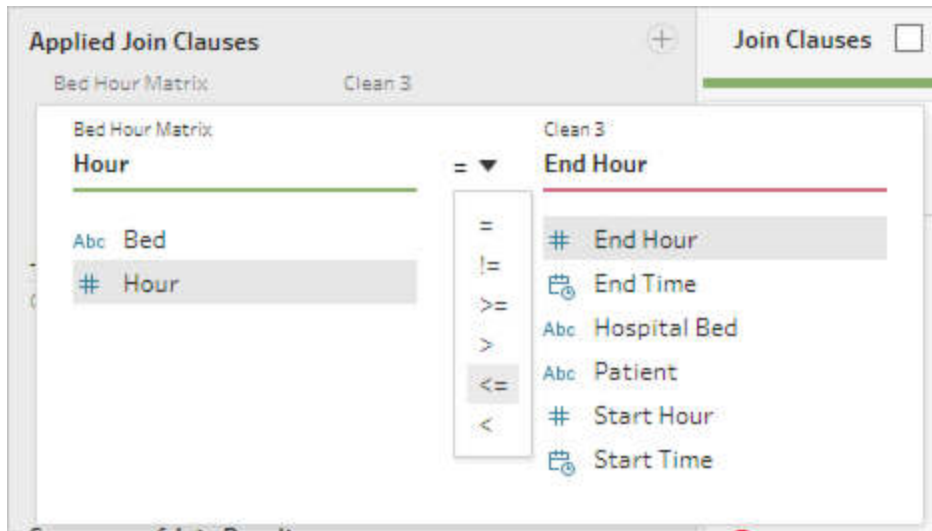
Now we're ready to join patient bed usage to the **Bed Hour Matrix**. This is a bit more complex join than we did previously. An inner join would only return values present in both data sets. Because we want to make sure we keep all the bed-hour slots, regardless of whether or not a patient was in the bed, we'll need to do a left join. This will result in a lot of nulls, but that's appropriate.


We also need to match when a bed-hour slot is taken by a patient (or patients). In addition to matching the bed the patient is in, we need to consider the time. The Bed Hour Matrix data set just has an **Hour** field, and the **Patient Beds** data set has **Start Hour** and **End Hour**. We'll use some basic logic to determine if a patient should be assigned to a given bed-hour slot: *A patient is considered in a bed if their start hour is less than or equal to ( $\leq$ ) the bed-hour slot AND their end hour is greater than or equal to ( $\geq$ ) the bed-hour slot.*

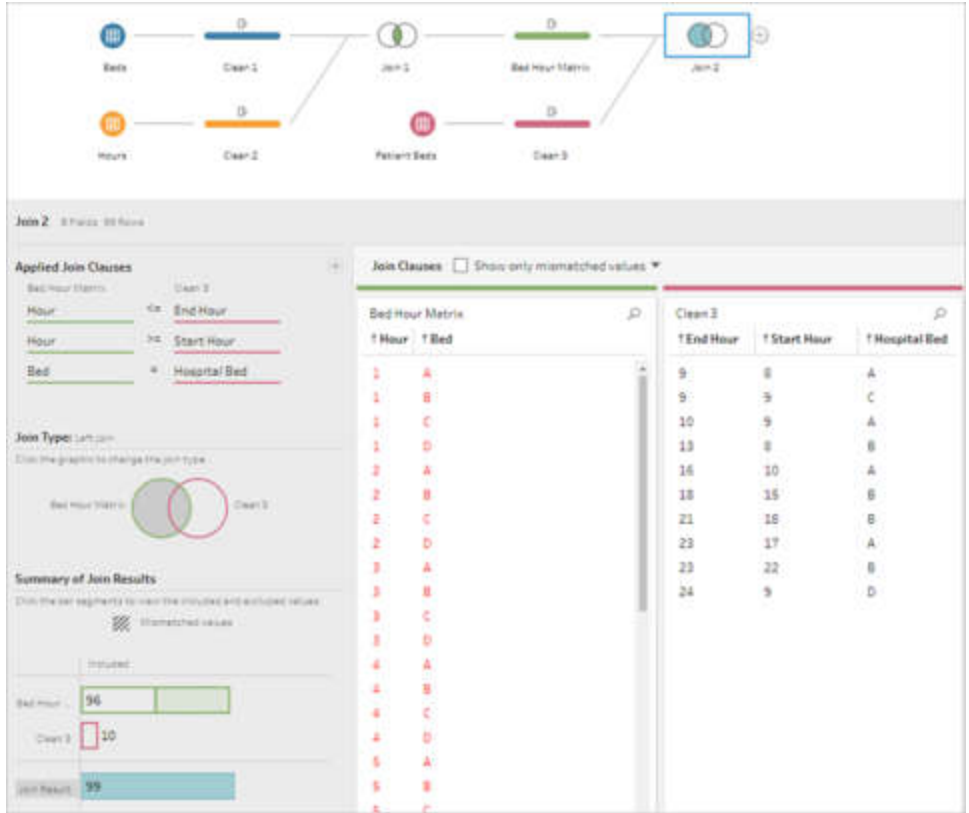
Therefore, three join clauses are needed to appropriately match these two data sets together.

9. Join the **Clean 3** step with the **Bed Hour Matrix** step.
10. In the **Applied Join Clauses** area, the default should be **Hour = End Hour**. Click the

join clause to change the operator from "=" to "<=".



11. Click the plus  button in the upper right corner of the **Applied Join Clauses** area to add another join clause. Set it to be **Hour >= Start Hour**
12. Add a third join clause for **Bed = Hospital Bed**.
13. In the **Join Type** section, click the unshaded area of the graphic next to **Bed Hour Matrix** to change the join type to a **Left** join.





**Note:** If you drag the **Bed Hour Matrix** to **Clean 3** instead of the other way around, the desired results can be obtained by using a right join instead of a left join. The order of dragging the steps matters for the orientation of the join. The join clauses will also be in reverse order—be sure to preserve the correct logic of comparing the hours.

Our data is now joined, but we should clean up some artifacts from the join and make sure the fields are tidy. We no longer need **Start Hour** and **End Hour**. **Hospital Bed** and **Bed** are also redundant. Finally, a value of null in the **Patient** field really means the bed is unoccupied.

14. In the **Flow** pane, add a cleaning step so we can tidy up the joined data.
15. Ctrl+click (Command+click on Mac) to multi select the fields **End Hour**, **Start Hour**, and **Hospital Bed**, then click **Remove Fields** in the toolbar.
16. On the **Patient** field profile card, double click the **null** value and type **Unoccupied**.

We now have a data structure with a row for every bed-hour; if there was a patient in bed during that hour, we have the patient information as well. All that remains to do is add an output step and generate the data set itself.

17. In the **Flow** pane, select **Clean 4**, click the plus  icon, and select **Add Output**.
18. In the **Output** pane, change the **Output type** to .csv then click **Browse**.
19. Enter **Bed Hour Patient Matrix** for the name and choose the desired location before clicking **Accept** to save.
20. Click the **Run Flow**  button at the bottom of the pane to generate your output. Click **Done** in the status dialog to close the dialog.

**Tip:** For more information about outputs and running a flow, see [Save and Share Your Work](#) on page 321.

The final flow should look like this:



## Analysis in Tableau Desktop

To install Tableau Desktop before continuing with this tutorial, you can download the [free trial](#).

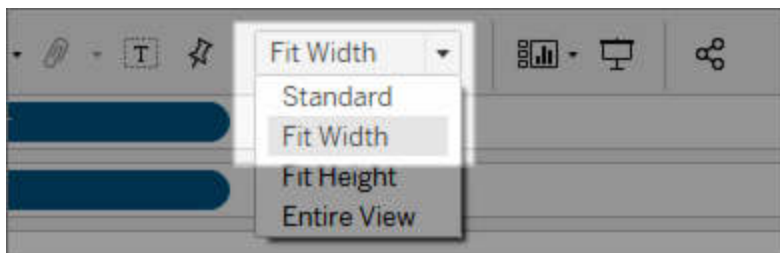
Now that we have the data set in the desired structure, we can perform deeper analysis than with the original data.

1. Open Tableau Desktop. In the **Connect** pane, select **Text file**, navigate to the **Bed Hour Patient Matrix.csv** file, and click **Open**.
2. On the **Data source** tab, the data should appear in the canvas by default. Click to **Sheet 1**.
3. In the **Data** pane, drag **Hour** from **Measures** to **Dimensions** to make it a discrete dimension.
4. Drag **Bed** to the **Rows** shelf and **Hour** to the **Columns** shelf.
5. Drag **Patient** to the **Color** shelf.

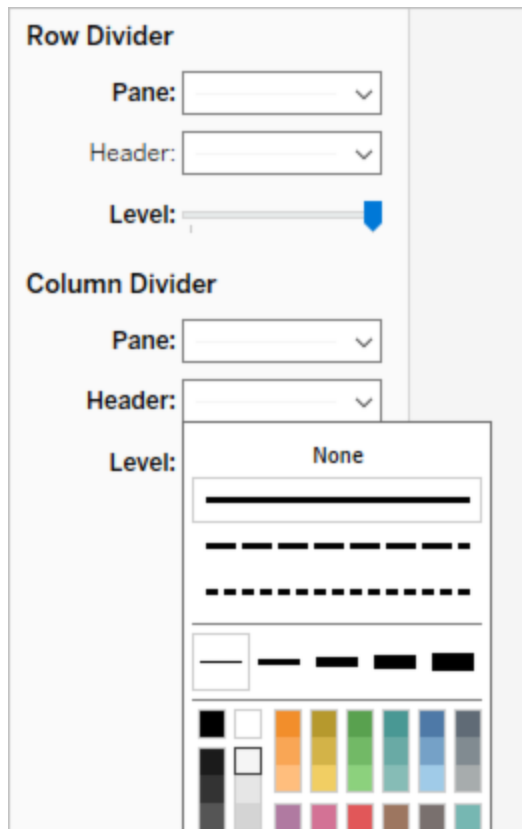


Formatting is optional, but may help make the visual more readable.

6. Click on the **Color** shelf and select **Edit Colors**.
7. In the area to the left, select **Unoccupied**. From the drop down on the right, choose the **Seattle Grays** color palette.
8. Select the fourth, lightest gray, and click **OK**.
9. Click the **Color** shelf again, then click the **Border** dropdown. Choose the second gray option at the far right.
10. In the toolbar, from the Size dropdown, change from **Standard** to **Fit Width**.

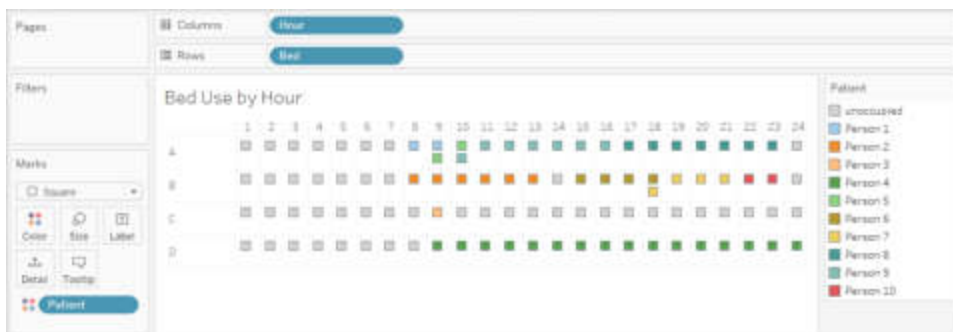


11. Click the **Format** menu and then **Borders**.
12. For **Row Divider**, click the Pane dropdown and choose a very light gray.
13. Adjust the **Level** slider to the second tick mark.
14. Repeat with the **Column Divider**. Set the **Pane** color to be light gray and the **Level** to the second tick mark.




15. Double click the sheet tab at the bottom and rename it **Bed Use by Hour**.

This view lets us quickly see when a given bed was occupied or open.



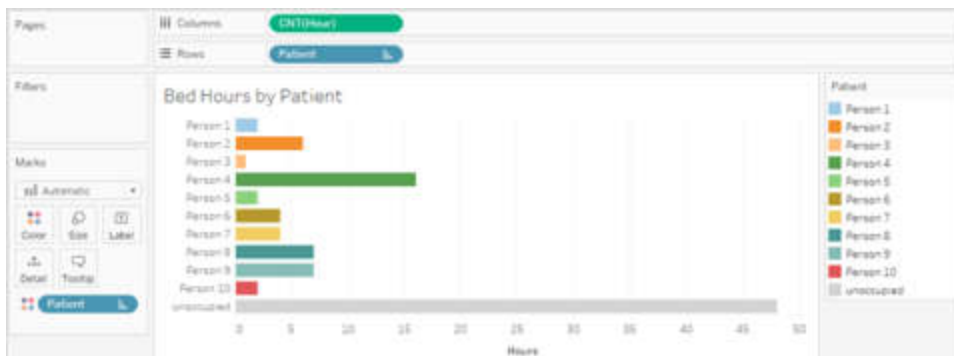
But we can go further and count the number of hours each bed was unoccupied.

16. Click the new sheet tab  icon at the bottom to open a clean sheet.

17. Drag **Patient** to **Rows**.



18. Drag **Hour** to **Columns**. Right click to open the menu. Choose **Measure > Count**.
19. Drag another copy of the **Patient** field from the **Data** pane to the **Color** shelf.
20. Right click on the axis and select **Edit Axis**. Change the title to **Hours** and close the dialog.
21. Rename the sheet tab **Bed Hours by Patient**.



This view lets us identify how many unoccupied bed hours we had, something we couldn't do with the original data set. What other charts or dashboards can you create? Give it a try now that your data is in the right structure.

## Recap and Resources

To build this data structure using Tableau Prep, we needed to perform the following actions:

1. Build a data set for each aspect we want to analyze, in this case, **Beds** and **Hours**.
2. Cross join those data sets to create a **Bed Hour Matrix** data set with every possible combination of beds and hours.
3. Join the **Bed Hour Matrix** with the **Patient Bed** data, making sure the join keeps all bed-slot hours and the join clauses appropriately match patient bed data with the bed-hour slots.

We used the following calculations to create fields we could join on. The second and third pull out the hour information from the original datetime fields.

- **Cross Join = 1**
  - This simply assigns the value 1 to every row
- **Start Hour = DATEPART('hour', [Start Time])**

- This takes the hour of the start time and pulls it out. Therefore, "1/1/18 9:35 AM" becomes simply "9".
- **End Hour** = `IFNULL(DATEPART('hour', [End Time]), 24)`
  - We could use `DATEPART('hour', [End Time])`, as we did for **Start Time**. This takes the hour of the end time and pulls it out. Therefore, "1/1/18 4:34 PM" becomes simply "4".
  - But we want to indicate that the patient bed that is still occupied (no end time) is in use, not empty. To do so, we'll assign an end time of 24 (midnight) to any missing end time using the `IFNULL` function. If the first argument `DATEPART('hour', [End Time])` is null, the calculation will return "24" instead.

**Note:** Want to check your work? Download the Tableau Prep packaged flow file ([Hospital Beds.tflx](#)) and the Tableau Desktop packaged workbook file ([Hospital Beds.twbx](#)).

**Resources:** Need more training? Check out the new [training videos](#) for Tableau Prep or take an [in-person training](#) course. Curious about the features we covered? Check out the other topics in the Tableau Prep online help. Looking for additional resources? The [Master Tableau Prep with this list of learning resources](#) blog post is for you.

## Finding the Second Date with Tableau Prep

A common need in analytics is to determine the date a *second* event happens, such as when a customer made a second purchase—thereby becoming a repeat customer—or when a driver gets a second traffic violation. Finding the date of a first event is easy, it's simply the minimum date. Finding the second date is trickier.

In this two-part tutorial, we'll shape traffic infraction data and answer the following questions:

1. What was the length of time in days between the first and second infraction for each driver?
2. Compare the fine amounts for the first and second infractions. Are they correlated?
3. Which driver paid the most overall? Who paid the least?

4. How many drivers had multiple infraction types?
5. What was the average fine amount for drivers who never attended traffic school?

In the first stage, we'll use Tableau Prep to restructure the data for our analysis. In the second stage, [Analysis with the Second Date in Tableau Desktop on page 471](#), we'll move on to analysis in Tableau Desktop.

The goal of this tutorial is to present various concepts in the context of a real-life scenario and work through options—not prescriptively establishing which is best. At the end, you should have a better sense of how data structure impacts calculations and analysis, as well as greater familiarity with various aspects of Tableau Prep and calculations in Tableau Desktop.

**Note:** To complete the tasks in this tutorial, you need Tableau Prep and optionally Tableau Desktop installed and the data downloaded.

To install Tableau Prep and Tableau Desktop before continuing with this tutorial, see the [Tableau Desktop and Tableau Prep Deployment guide](#). Otherwise you can download the [Tableau Prep](#) and [Tableau Desktop](#) free trials.

The data set is [Traffic Violations.xlsx](#). It is recommended to save it in your My Tableau Prep Repository > Datasources folder.

## The Data

For this example, we're looking at traffic infraction data. Each infraction is a row. The driver, date, type of infraction, if the driver was required to attend traffic school, and fine amount are recorded.

	A	B	C	D	E
1	Driver ID	Infraction Date	Infraction Type	Traffic School	Fine Amount
2	JO-151451402	1/8/2017	Speeding	Yes	115
3	CM-127151402	3/1/2017	Running a red light	No	55
4	AP-109151404	3/2/2017	Non-moving violation	No	95
5	SH-199751404	3/4/2017	Speeding	Yes	130
6	BT-114401404	3/20/2017	Non-moving violation	No	130
7	MO-175001406	5/30/2017	Speeding	Yes	118
8	RA-1988558	6/2/2017	Speeding	Yes	144
9	BT-1168027	6/5/2017	Speeding	Yes	128
10	MO-175001406	6/18/2017	Speeding	Yes	115
11	MP-174701406	6/19/2017	Speeding	No	125
12	AA-106451404	7/5/2017	Running a red light	No	60
13	RA-199151402	7/20/2017	Speeding	Yes	146
14	SC-202601404	8/31/2017	Running a red light	No	150
15	MO-175001406	9/7/2017	Non-moving violation	No	320
16	AS-100451404	9/26/2017	Running a red light	No	50

## Desired Data Structure

The data is currently structured such that each *infraction* is a row. A driver with multiple infractions appears on multiple rows, and there's no easy way to tell which was their first or second infraction.

To investigate our repeat offenders, we want a data set that separates out the first and second infraction dates, and the information associated with each of those infractions, and each row is a *driver*.

	A	B	C	D	E	F	G	H	I
1	Driver ID	1st Infraction Date	1st Infraction Type	1st Traffic School	1st Fine Amount	2nd Infraction Date	2nd Infraction Type	2nd Traffic School	2nd Fine Amount
2	BD-117701406	12/25/2017	Speeding	Yes	140	3/7/2018	Speeding	Yes	125
3	JO-151451402	1/8/2017	Speeding	Yes	115	11/21/2018	Rackless driving	Yes	350
4	DN-207101403	12/7/2017	Speeding	Yes	380	4/26/2018	Speeding	Yes	130
5	CJ-120101402	11/26/2017	Speeding	Yes	122	3/28/2018	Speeding	Yes	118
6	JR-156701404	12/24/2017	Speeding	No	148	7/28/2018	Speeding	Yes	210
7	AP-109151404	3/2/2017	Non-moving violation	No	95	9/24/2018	Speeding	No	101
8	PC-187431406	11/11/2017	Speeding	Yes	220	12/30/2018	Non-moving violation	No	600
9	TS-214301406	9/13/2018	Speeding	Yes	115	11/10/2018	Non-moving violation	No	95
10	NR-187001404	12/11/2018	Non-moving violation	No	80	12/20/2018	Speeding	No	120
11	DB-129701402	3/13/2018	Running a red light	No	118	11/11/2018	Speeding	Yes	80
12	AJ-107951404	10/15/2017	Speeding	Yes	130	12/31/2017	Running a red light	No	85
13	BT-114401404	3/20/2017	Non-moving violation	No	130	11/11/2018	Speeding	Yes	96
14	AF-108851406	3/9/2018	Non-moving violation	No	200	5/2/2018	Speeding	No	130
15	SC-302601404	8/31/2017	Running a red light	No	150	11/16/2018	Speeding	Yes	50
16	KL-166401406	10/4/2017	Speeding	No	115	11/11/2017	Speeding	Yes	104
17	MO-175001406	5/30/2017	Speeding	Yes	118	6/18/2017	Speeding	Yes	115
18	CM-127151402	3/1/2017	Running a red light	No	55	8/1/2018	Running a red light	No	180
19	KT-164801402	3/11/2018	Non-moving violation	No	190	11/10/2018	Speeding	No	74
20	JR-169001402	11/18/2018	Speeding	Yes	220	12/5/2018	Non-moving violation	No	195
21	LR-170201404	3/9/2018	Running a red light	No	110	9/17/2018	Speeding	Yes	230
22	BG-1103055	12/25/2017	Speeding	Yes	195	12/8/2018	Speeding	Yes	313
23	MP-174701406	6/19/2017	Speeding	No	125	10/12/2017	Running a red light	No	175
24	AS-100451404	9/26/2017	Running a red light	No	50	12/28/2017	Speeding	Yes	114

## Restructuring the Data

So how do we get there with Tableau Prep? We'll build out the flow in stages, beginning with pulling out the first infraction date, then the second, then shaping the final data set as desired. Make sure you've downloaded the Excel file ([Traffic Violations.xlsx](#)) to follow along.

### Initial Aggregation for 1st Infraction Date


First, we'll connect to the **Traffic Violations.xlsx** file.

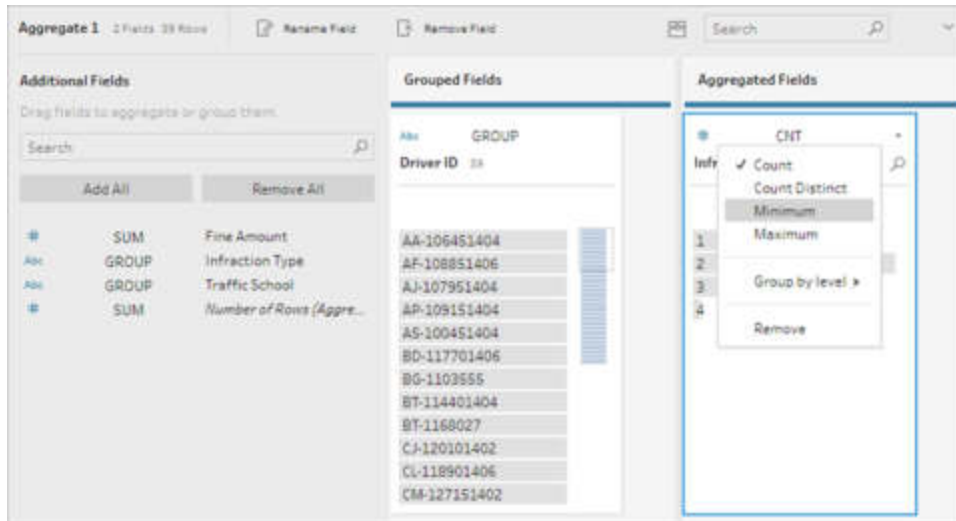
1. Open Tableau Prep.
2. From the start screen, click **Connect to Data**.
3. In the **Connections** pane, click **Microsoft Excel**. Navigate to where you saved **Traffic Violations.xlsx** and click **Open**.
4. The **Infractions** sheet should automatically be brought out to the **Flow** Pane.

**Tip:** For more information about connecting to data, see [Connect to Data on page 145](#).

Next, we need to identify the first infraction date per driver. We'll use an **Aggregate** step to do this, creating a mini data set of **Driver ID** and **Minimum Infraction Date**.

**Note:** When using an **Aggregate** step in Tableau Prep, any field that should define what makes a row is a **Grouped Field**. (For us, that's **Driver ID**.) Any field that will be aggregated and presented at the level of the grouped fields is an **Aggregated Field**. (For us, that's **Infraction Date**).

5. In the **Flow** pane, select **Infractions**, click the plus  icon, and select **Add Aggregate**.
6. Drag **Driver ID** to the **Grouped Fields** drop area.
7. Drag **Infraction Date** to the **Aggregated Fields** area. The default aggregation is **CNT** (count). Click **CNT** and change the aggregation to **Minimum**.

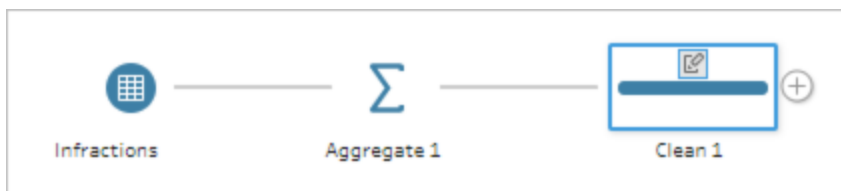


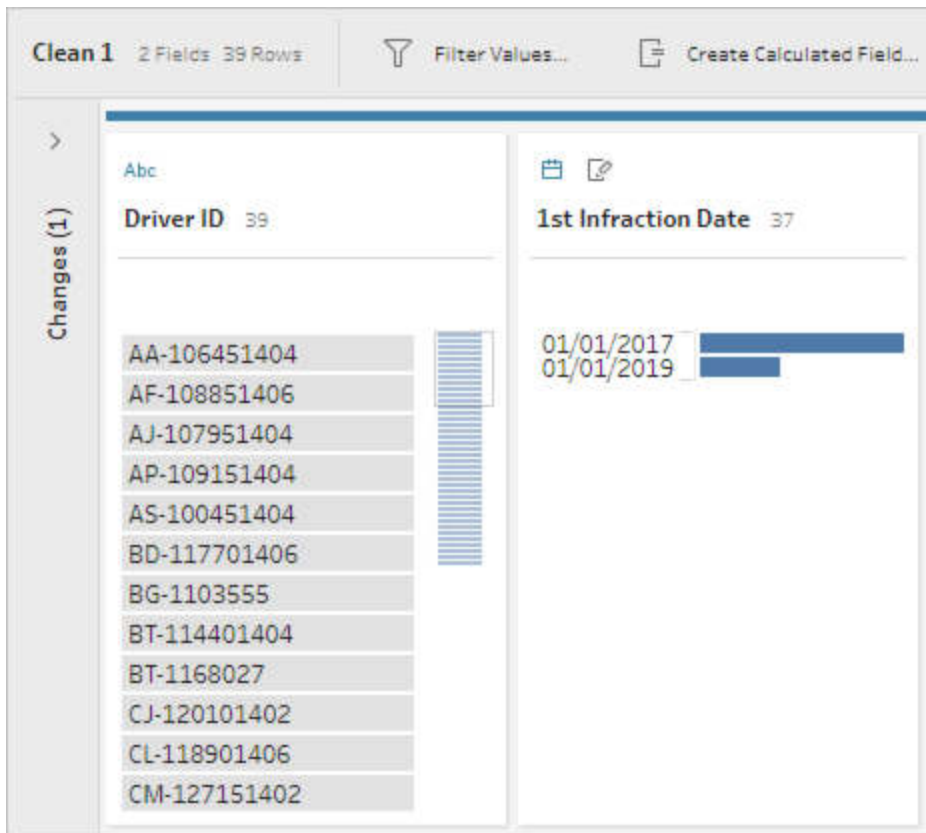
This identifies the smallest (earliest) date, which is the first infraction date per driver.

**Tip:** For more information about aggregations, see [Clean and Shape Data on page 229](#).

8. In the **Flow** pane, select **Aggregate 1**, click the plus (+) icon, and select **Add Step** so we can clean up the output of the aggregation.
9. In the **Profile** pane, double-click on the field name **Infraction Date** and change it to **1st Infraction Date**.

*At this stage, the flow and profile should look like this:*





From the **Profile** pane in this **Clean** step, we can see that our data now consists of 39 rows and only 2 fields. Any field not used for grouping or aggregation is lost. But we want to be able to keep some of the original information. We could either add those fields to the grouping or aggregation (but doing so would change the level of detail or require the fields to be aggregated), or join this mini data set back to the original (essentially adding a new column to the original data for **1st Infraction Date**). Let's do the join.

10. In the **Flow** pane, select **Infractions**, click the plus  $\oplus$  icon, and select **Add Branch**.

This branch has all the original data. We'll join the results of the aggregation to this copy of the full data. By joining on **Driver ID**, we're adding the minimum date from our aggregation into the original data.

11. Select step **Clean 2** and drag it on top of step **Clean 1**, and drop it on **Join**.
12. The default join configuration should be correct: an inner join on **Driver ID = Driver ID**.

The screenshot displays the Tableau Prep interface for a join operation. On the left, the 'Applied Join Clauses' section shows 'Clean 1 Driver ID = Clean 2 Driver ID'. Below this, the 'Join Type' is set to 'Inner join', illustrated with a Venn diagram. The 'Summary of Join Results' section shows a bar chart with three segments: 'Clean 1' (39), 'Clean 2' (81), and 'Join Result' (81). On the right, the 'Join Clauses' section shows two columns, 'Clean 1' and 'Clean 2', both containing '1 Driver ID' and a list of driver IDs.

**Tip:** For more information about joins, see [Join your data](#) on page 308.

Because some fields may be duplicated during a join, such as the fields in the join clause, it's often a good idea to clean up extraneous fields after performing a join.

13. In the **Flow** pane, select **Join 1**, click the plus  $\oplus$  icon, and choose **Add a Step**.
14. In the **Profile** pane, click the card for **Driver ID-1**, then click **Remove Field** in the toolbar.
15. To change the field order, drag the **1st Infraction Date** card between **Driver ID** and **Infraction Date** where you see the black line appear.

*At this stage, the flow should look like this:*





Looking at the data grid below, we can see our new, combined data set. We have the minimum—that is, first—infraction date for each driver added to each row in the data set.

Driver ID	1st Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	01/08/2017	01/08/2017	Speeding	Yes	115
CM-127151402	03/01/2017	03/01/2017	Running a red light	No	55
AP-109151404	03/02/2017	03/02/2017	Non-moving violation	No	95
SH-199751404	03/04/2017	03/04/2017	Speeding	Yes	130
BT-114401404	03/20/2017	03/20/2017	Non-moving violation	No	130
MO-175001406	05/30/2017	05/30/2017	Speeding	Yes	118
RA-1988558	06/02/2017	06/02/2017	Speeding	Yes	144
BT-1168027	06/05/2017	06/05/2017	Speeding	Yes	128
MO-175001406	06/30/2017	06/18/2017	Speeding	Yes	115
MP-174701406	06/19/2017	06/19/2017	Speeding	No	125
AA-106451404	07/05/2017	07/05/2017	Running a red light	No	60
RA-199151402	07/20/2017	07/20/2017	Speeding	Yes	146
SC-202601404	08/31/2017	08/31/2017	Running a red light	No	150
MO-175001406	05/30/2017	09/07/2017	Non-moving violation	No	320
AS-100451404	09/26/2017	09/26/2017	Running a red light	No	50
SH-199751404	03/04/2017	09/27/2017	Speeding	Yes	225
AA-106451404	07/05/2017	09/28/2017	Running a red light	No	195

## Second Aggregation for 2nd Infraction Date

We need to also determine the second infraction date. To do this, we want to filter out any row where the infraction date is equal to the minimum—thus removing the first date. We can then take the minimum of the remaining dates using another aggregate step, leaving us with the second infraction date, which we'll rename for clarity.

**Note:** Because we'll want to use the data as it currently is in **Clean 3** later on in the flow, we'll add another **Clean** step to break out the process of getting the second infraction date.

16. In the **Flow** pane, select **Clean 3**, click the plus  $\oplus$  icon, and select **Add Step**.
17. On the toolbar in the **Profile** pane, choose **Filter Values**. Create a filter [Infraction Date]  $\neq$  [1st Infraction Date].

18. Remove the field **1st Infraction Date**.
19. In the **Flow** pane, select **Clean 4**, click the plus  $\oplus$  icon, and select **Add Aggregate**.
20. Drag **Driver ID** to the **Grouped Fields** drop area. Drag **Infraction Date** to the **Aggregated Fields** area and change the aggregation to **Minimum**.
21. In the **Flow** pane, select **Aggregate 2**, click the plus  $\oplus$  icon, and select **Add Step**.  
Rename **Infraction Date** to **2nd Infraction Date**.

*At this stage, the flow should look like this:*



We now have our second infraction date identified for each driver. To get all the other information associated with each infraction (type, fine, traffic school) we again need to join this back to the entire data set.

22. Select **Clean 5** and drag it on top of **Clean 3**, dropping it on **Join**.
23. Again, the default join configuration should be correct: an inner join on **Driver ID = Driver ID**.
24. In the **Flow** pane, select **Join 2**, click the plus  $\oplus$  icon, and select **Add Step**. Delete the fields **Driver ID-1** and **1st Infraction Date** as they are no longer needed.

*At this stage, the flow should look like this:*



## Create full data sets for the 1st and 2nd infractions

Before we go any further, let's step back and think about everything we have and how we want to bring it all together. Our desired end state is a data set that looks like this, with a column for

**Driver ID**, then columns for date, type, traffic school, and fine amount for the 1st and 2nd infractions.

	A	B	C	D	E	F	G	H	I
1	Driver ID	1st Infraction Date	1st Infraction Type	1st Traffic School	1st Fine Amount	2nd Infraction Date	2nd Infraction Type	2nd Traffic School	2nd Fine Amount
2	BO-11701406	12/25/2017	Speeding	Yes	140	3/7/2018	Speeding	Yes	125
3	JO-151451402	1/8/2017	Speeding	Yes	115	11/21/2018	Rackless driving	Yes	350
4	SN-20701403	12/27/2017	Speeding	Yes	380	4/26/2018	Speeding	Yes	130
5	CA-12010402	11/26/2017	Speeding	Yes	122	3/28/2018	Speeding	Yes	118
6	JR-154701404	12/24/2017	Speeding	No	148	7/28/2018	Speeding	Yes	310
7	AP-108151404	3/2/2017	Non-moving violation	No	95	9/24/2018	Speeding	No	103
8	PC-187431406	11/11/2017	Speeding	Yes	220	12/30/2018	Non-moving violation	No	600
9	TS-214301400	8/13/2018	Speeding	Yes	115	11/30/2018	Non-moving violation	No	95
10	NP-187001404	12/11/2018	Non-moving violation	No	80	12/20/2018	Speeding	No	120
11	DB-129701402	3/13/2018	Running a red light	No	131	11/11/2018	Speeding	Yes	80
12	AJ-107951404	10/15/2017	Speeding	Yes	130	12/31/2017	Running a red light	No	83
13	BT-114401404	3/20/2017	Non-moving violation	No	130	11/13/2018	Speeding	Yes	96
14	AF-108851406	5/9/2018	Non-moving violation	No	200	9/2/2018	Speeding	No	130
15	SC-302601404	8/31/2017	Running a red light	No	150	11/18/2018	Speeding	Yes	30
16	KL-166431406	10/4/2017	Speeding	No	115	11/13/2017	Speeding	Yes	104
17	MO-175001406	5/30/2017	Speeding	Yes	118	6/18/2017	Speeding	Yes	115
18	CM-127151402	3/1/2017	Running a red light	No	55	8/1/2018	Running a red light	No	180
19	KT-164801402	3/31/2018	Non-moving violation	No	190	11/30/2018	Speeding	No	74
20	JB-360901402	11/18/2018	Speeding	Yes	220	12/5/2018	Non-moving violation	No	195
21	LN-170201404	5/6/2018	Running a red light	No	110	9/13/2018	Speeding	Yes	230
22	BG-1103555	12/25/2017	Speeding	Yes	195	12/8/2018	Speeding	Yes	313
23	MP-174701406	6/19/2017	Speeding	No	125	10/12/2017	Running a red light	No	175
24	...	...	...	...	...	...	...	...	...

How do we get there from here?

In the step **Clean 3**, we have our complete data set with a column that repeats the first infraction date for each driver.

Driver ID	1st Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	01/08/2017	01/08/2017	Speeding	Yes	115
CM-127151402	03/01/2017	03/01/2017	Running a red light	No	55
AP-109151404	03/02/2017	03/02/2017	Non-moving violation	No	95
SH-199751404	03/04/2017	03/04/2017	Speeding	Yes	130
BT-114401404	03/20/2017	03/20/2017	Non-moving violation	No	130
MO-175001406	05/30/2017	05/30/2017	Speeding	Yes	118
RA-1988558	06/02/2017	06/02/2017	Speeding	Yes	144
BT-1168027	06/05/2017	06/05/2017	Speeding	Yes	128
MO-175001406	05/30/2017	06/18/2017	Speeding	Yes	115
MP-174701406	06/19/2017	06/19/2017	Speeding	No	125


We want to eliminate all the rows for a driver that aren't the first infraction, building a data set of only first infractions. That is, we only want to keep the information for a given driver when **1st Infraction Date = Infraction Date**. Once we've filtered to keep only the row of the first infraction, we can remove the **Infraction Date** field and tidy up field names.

Similarly, after the second aggregation and join, we have our complete data set with a column for the second infraction date.

Driver ID	2nd Infraction Date	Infraction Date	Infraction Type	Traffic School	Fine Amount
JO-151451402	11/21/2018	01/08/2017	Speeding	Yes	115
CM-127181402	08/01/2018	03/01/2017	Running a red light	No	55
AP-109151404	09/24/2018	03/02/2017	Non-moving violation	No	95
SH-199751404	09/27/2017	03/04/2017	Speeding	Yes	130
BT-114401404	11/13/2018	03/20/2017	Non-moving violation	No	130
MO-175001406	06/18/2017	05/30/2017	Speeding	Yes	118
MO-175001406	06/18/2017	06/18/2017	Speeding	Yes	115
MP-174701406	10/12/2017	06/19/2017	Speeding	No	125
AA-106451404	09/28/2017	07/05/2017	Running a red light	No	60
RA-199151402	12/31/2017	07/20/2017	Speeding	Yes	146
BT-305601404	11/01/2018	08/21/2017	Running a red light	No	150

We can perform a similar filter of **2nd Infraction Date = Infraction Date** to keep only the row of information for each driver's 2nd infraction. Again, we can also remove the now-redundant **Infraction Date** and tidy up field names.

We'll start with the first infraction data set.

25. In the **Flow** pane, select **Clean 3**, click the plus  icon, and select **Add Branch**.
26. With this new **Clean** step selected, in the **Profile** pane, click **Filter Values** in the toolbar. Create a filter `[1st Infraction Date] = [Infraction Date]`.
27. Remove the field **Infraction Date**.
28. Rename the **Infraction Type**, **Traffic School**, and **Fine Amount** fields to start with "1st".
29. Double-click on the name **Clean 7** under the step in the **Flow** pane and rename it **Robust 1st**.

Now for the second infraction data set.

30. In the **Flow** pane, select **Clean 6**, after the last join.
31. Click **Filter Values** in the toolbar. Create a filter `[2nd Infraction Date] = [Infraction Date]`.
32. Remove the field **Infraction Date**.
33. Rename the **Infraction Type**, **Traffic School**, and **Fine Amount** fields to start with "2nd".
34. Double-click on the name **Clean 6** under the step in the **Flow** pane and rename it **Robust 2nd**.

*At this stage, the flow should look like this:*



## Create the complete data set

Now that we have these two tidy data sets with complete information for the first and second infractions per driver, we can join them back together on **Driver ID** and wind up with our desired data structure.

35. Select **Robust 2nd** and drag it on top of **Robust 1st**, dropping it on **Join**.
36. The default join clause should be correct as **Driver ID = Driver ID**.
37. Because we don't want to drop drivers who didn't have a second infraction, we need to make this a left join. In the **Join Type** area, click the unshaded area of the diagram next to **Robust 1st**, turning it into a **Left** join.
38. In the **Flow** pane, select **Join 3**, click the plus  $\oplus$  icon, and select **Add Step**. Remove the field duplicate **Driver ID-1**.

The data is in the desired state, so we can create an output and proceed to analysis.

39. In the **Flow** pane, select the newly added **Clean 6**, click the plus  $\oplus$  icon, and select **Add Output**.
40. In the **Output** pane, change the **Output type** to .csv then click **Browse**. Enter **Driver Infractions** for the name and choose the desired location before clicking **Accept** to save.
41. Click the **Run Flow**  $\triangleright$  button at the bottom of the pane to generate your output. Click **Done** in the status dialog to close the dialog.

**Tip:** For more information about outputs and running a flow, see [Save and Share Your Work](#) on page 321.

*The final flow should look like this:*



**Note:** You can download the completed flow file to check your work: [Driver Infractions.tflx](#)

## Recap

For the first stage of this tutorial, our goal was to take our original data set and prepare it for analysis involving the first and second infraction dates. The process consists of three phases:

Identify the first and second infraction dates:

1. Create an aggregation that keeps **Driver ID** and **MIN Infraction Date**. Join this with the original data set to create an "intermediate data set" that has the first (minimum) infraction date repeated for every row.
2. On a new step, filter out all rows where the **1st Infraction Date** is the same as the **Infraction Date**. From that filtered data set, create an aggregation that keeps **Driver ID** and **MIN Infraction date**. Join this with the intermediate data set from the first step. This identifies the second infraction date.

Build out clean data sets for the first and second infractions:

3. Go back and create a branch from the intermediate data set and filter to keep only rows where the **1st Infraction Date** is the same as the **Infraction Date**. This builds a data set for just the first infraction. Tidy it up by removing any unnecessary fields and rename all the desired fields (except **Driver ID**) to indicate they're for the first infraction. This is the **Robust 1st** data set.
4. Tidy the data set for the second infraction date. Clean the join results from step 2 by filtering to keep only rows where the **2nd Infraction Date** is the same as the **Infraction Date**. Remove any unnecessary fields and rename all the desired fields (except **Driver ID**) to indicate they're for the second infraction. This is the **Robust 2nd** data set.

Combine the first and second infraction data into one data set:

5. Join the **Robust 1st** and **Robust 2nd** data sets, making sure to keep all records from **Robust 1st** to prevent losing any drivers without a second infraction.

Next, we want to explore how this data can be analyzed in Tableau Desktop.

Continue to [Analysis with the Second Date in Tableau Desktop below](#).

**Note:** Special Thanks to Ann Jackson's Workout Wednesday topic [Do Customers Spend More on Their First or Second Purchase?](#) and Andy Kriebel's Tableau Prep Tip [Returning the First and Second Purchase Dates](#) that provided the initial inspiration for this tutorial. Clicking these links will take you away from the Tableau website. Tableau cannot take responsibility for the accuracy or freshness of pages maintained by external providers. Contact the owners if you have questions regarding their content.

## Analysis with the Second Date in Tableau Desktop

This is the second stage of the tutorial and assumes the first stage, [Finding the Second Date with Tableau Prep on page 458](#), has been completed.

In the first stage, we took our original data set and shaped it to answer the following questions:

1. What was the length of time in days between the first and second infraction for each driver?
2. Compare the fine amounts for the first and second infractions. Are they correlated?
3. Which driver paid the most overall? Who paid the least?
4. How many drivers had multiple infraction types?
5. What was the average fine amount for drivers who never attended traffic school?

As we now explore these questions, it becomes clear that there are some pros and cons to the first data structure we created. We'll go back into Tableau Prep and do some additional reshaping, then see how that impacts the same analysis in Tableau Desktop. Finally, we'll look at a Tableau Desktop-only approach to the analysis using Level of Detail (LOD) expressions with the original data.

The goal of this tutorial is to present various concepts in the context of a real-life scenario and work through options—not prescriptively establishing which is best. At the end, you should have a better sense of how data structure impacts calculations and analysis, as well as greater familiarity with various aspects of Tableau Prep and calculations in Tableau Desktop.

**Note:** To complete the tasks in this tutorial, you need Tableau Prep and optionally Tableau Desktop installed and the data downloaded.

To install Tableau Prep and Tableau Desktop before continuing with this tutorial, see the [Tableau Desktop and Tableau Prep Deployment guide](#). Otherwise you can download the [Tableau Prep](#) and [Tableau Desktop](#) free trials.

The data set is the output from [Driver Infractions.tflx](#), as built in the first stage.

## Analysis in Tableau Desktop

Now that we have our data configured, we'll bring it into Tableau Desktop. We can easily answer some questions, but others involve a few (or a lot of) calculations. Try your hand at the questions below; you can expand each one for basic information about how to proceed if you get stuck.

**Note:** You can download the workbook [Driver Infractions.twbx](#) to look at the solutions in context. Remember that there may be alternative ways to interpret the analysis or pursue answers.

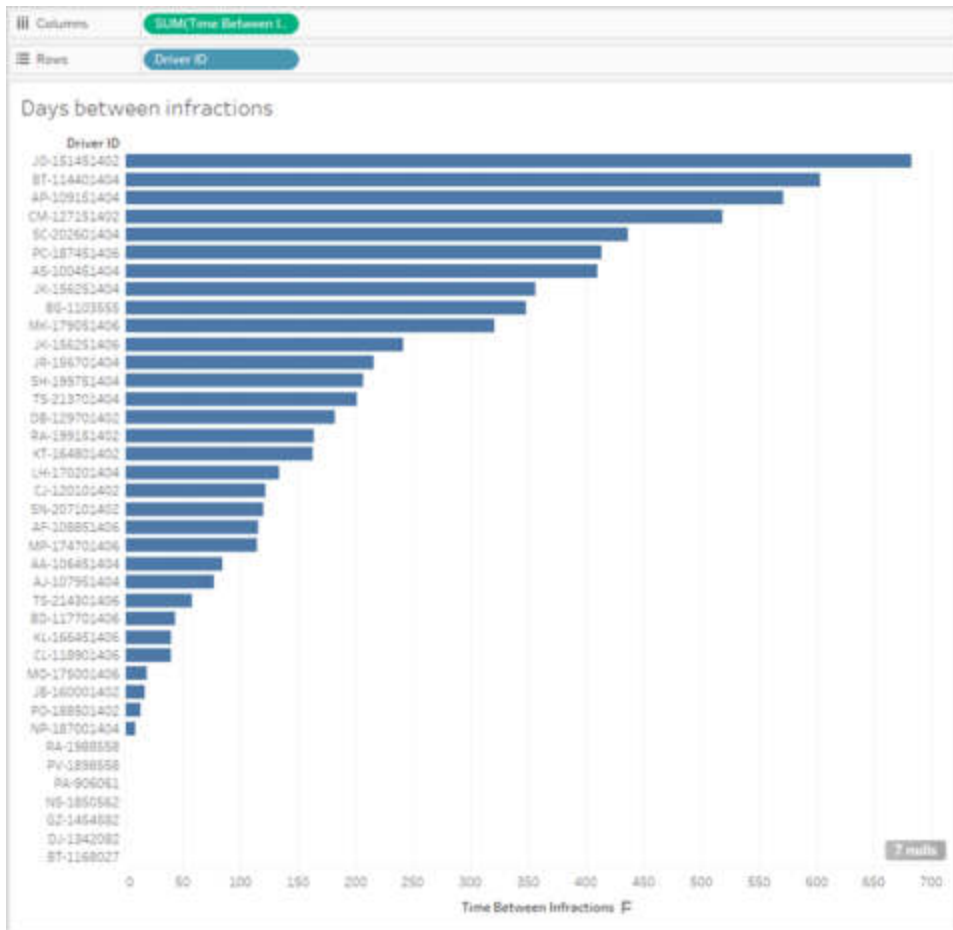
### 1. What was the length of time in days between the first and second infraction for each driver?

- A. To answer this question in Tableau Desktop, we'll use the DATEDIFF function. This function takes three arguments—the date part, the start date, and the end date. Since we want to know the days between these events, we'll use the date part 'day'. Our start and end dates are in the data set as **1st Infraction Date** and **2nd Infraction Date**.
- B. The calculation is:



**Time Between Infractions** = DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])

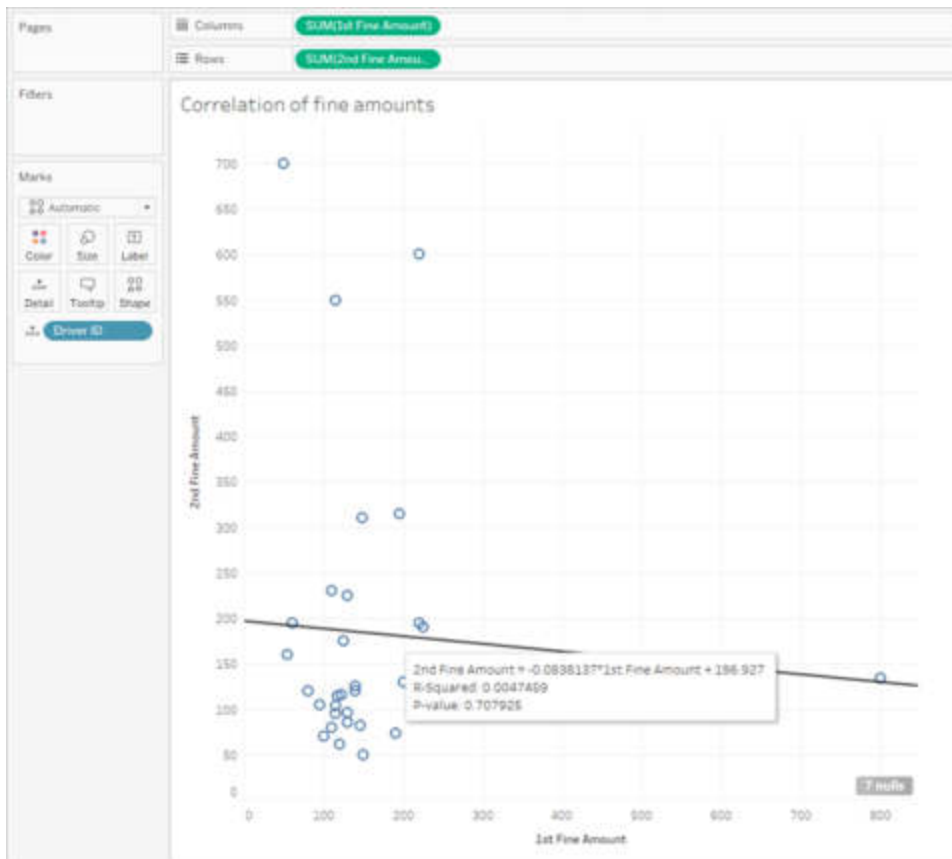
- C. We can plot that against **Driver ID** as a bar chart. Note that seven drivers had no second infraction, so there are seven nulls.



## 2. Compare the fine amounts for the first and second infractions. Are they correlated?

- A. To answer this question in Tableau Desktop, we'll create a scatter plot of **1st Fine Amount** and **2nd Fine Amount**. By bringing **Driver ID** to the **Detail** shelf on the **Marks** card, we can create a mark for each driver.

- B. To add a trend line, use the **Analytics** tab in the left-hand pane and bring out a linear trend line. Hovering over the trend line, we can see the R-squared value is practically zero, and the p-value is far above any cutoff for significance. We can determine that there is no correlation between first and second fine amount.



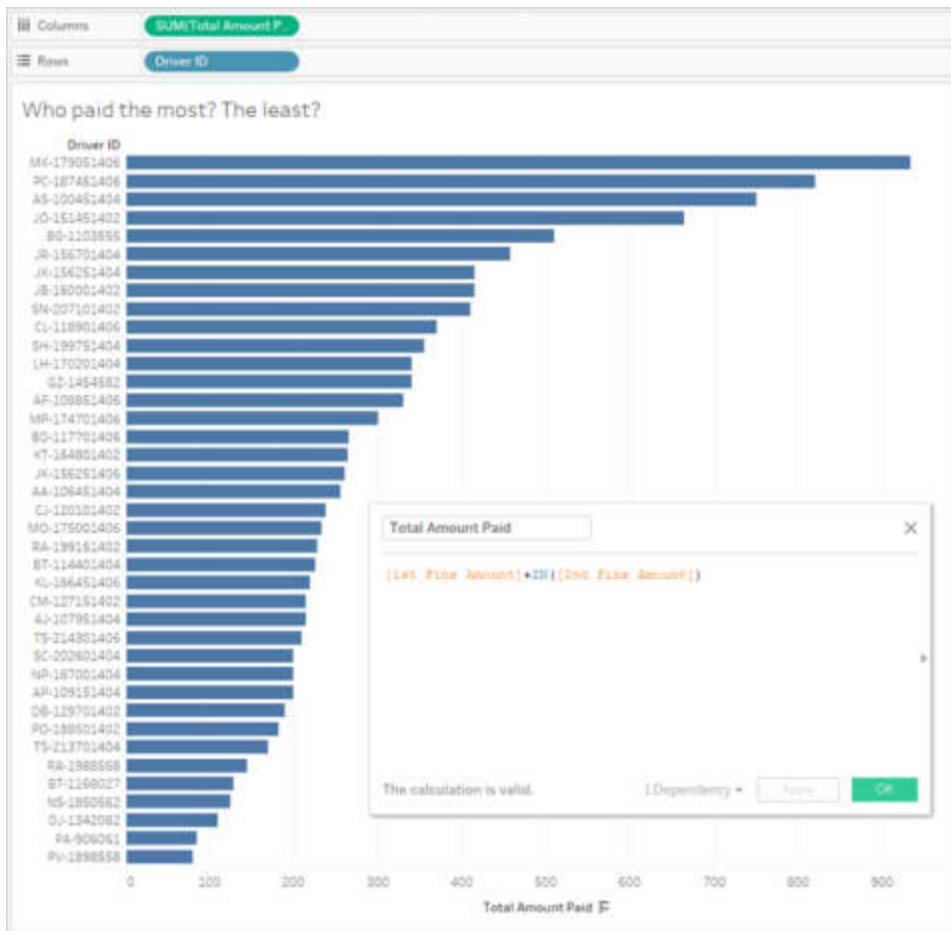
### 3. Which driver paid the most overall? Who paid the least?

When we want to go deeper in our analysis, we may need to create some calculations.

- A. To answer this in Tableau Desktop, we need to add the fines for both infractions into a single field. Because some drivers may not have had a second infraction, we need to use the zero null `ZN` function to turn any nulls for **2nd Fine Amount** into zeros.
- B. The calculation is:

**Total Amount Paid** = [1st Fine Amount] + ZN([2nd Fine Amount])

C. We can plot **Total Amount Paid** against **Driver ID** and sort the bar chart.



#### 4. How many drivers had multiple infraction types?

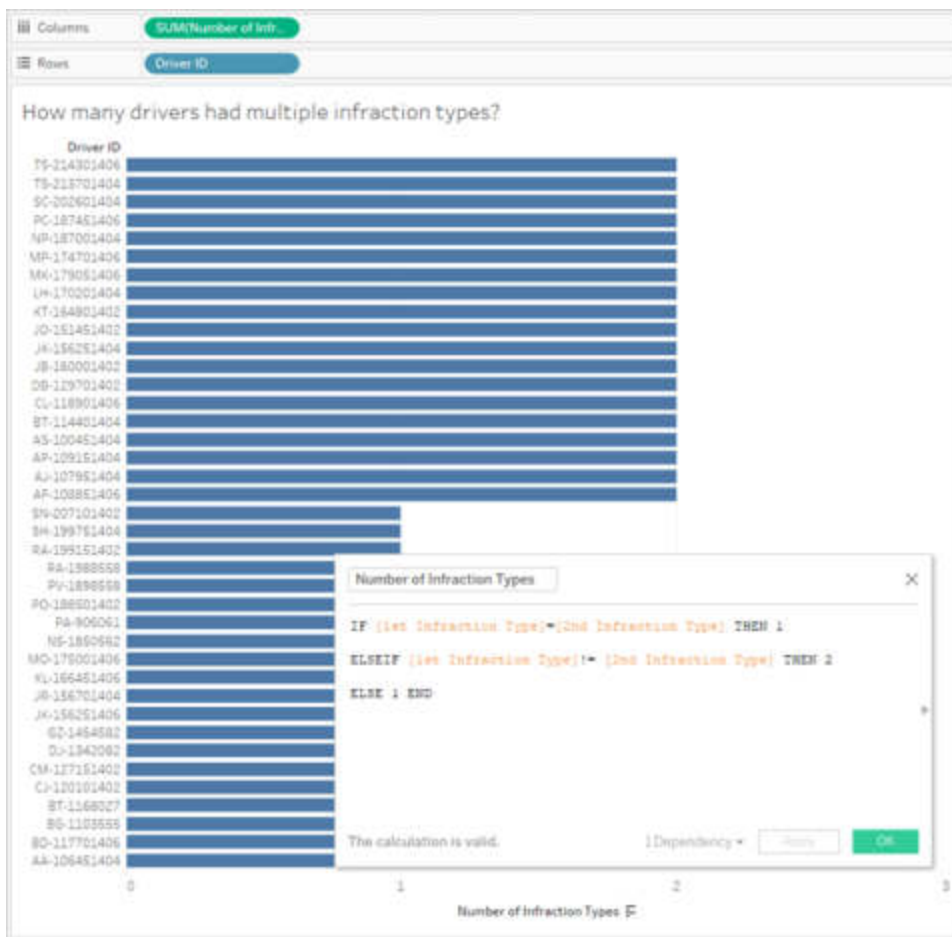
A. To answer this in Tableau Desktop, we need to do a fancier `IF` calculation, comparing if the first and second infraction types are the same. If they are, we want to assign the value "1". If they are not the same, we'll assign "2". Any other result, such as a null second infraction type, will be assigned "1".

B. The calculation is:

**Number of Infraction Types** =

```
IF [1st Infraction Type]=[2nd Infraction Type] THEN 1
ELSEIF [1st Infraction Type]!=[2nd Infraction Type] THEN 2
ELSE 1 END
```

- C. We can then plot **Number of Infraction Types** against **Driver ID** and sort the bar chart.



## 5. What was the average fine amount for drivers who never attended traffic school?

- A. To answer this in Tableau Desktop, we cannot simply divide the total fine amount by two, since some drivers only had one infraction. We also can't calculate the average fine per

driver and take the average of those values, because averaging averages can lead to inconsistencies. Instead, we need to calculate the total amount paid by drivers who never attended traffic school, then divide by the total number of infractions associated with those fines.

1. First, we need to determine if each driver had a second infraction. We can leverage the fact the information in all the "2nd" fields will be null if there was no second infraction and start building the calculation:

```
IFNULL([2nd Infraction Type], 'no')
```

This will return an infraction type if it exists, or "no" if there was no second infraction.

2. Next, we need to turn this information into the number of infractions, 1 or 2. If the result of our `IFNULL` calculation is "no", then the driver should be marked as having one fine. Any other result should be marked as having two fines. The calculation is:

**Number of Infractions =**

```
IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1
ELSE 2
END
```

3. Now we need to consider the total fine amount. Similarly to question 3 above, we'll add the first and second fine amounts, with a `ZN` function around the second. However, because we want this to be computed at the level of the entire data set, it's a best practice to specify the aggregations, **SUM**, in the calculation itself. The calculation is:

```
SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) )
```

4. To bring it all together, we'll take this total fine amount and divide it by our new **Number of Infractions** calculated field to determine the average fine amount:

```
Average Fine = ( SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) ) ) / SUM([Number of Infractions])
```

- B. We also need to filter out drivers who ever attended traffic school—but that information is also stored across two fields.

1. Tableau is very efficient at numerical calculations. We'll phrase this with numbers to help performance as much as we can. To combine these two fields, we'll create a calculation for each one that says "Yes = 1" and "No = 0" (null should also = 0, for drivers with no second infraction). By summing the outcome of these calculations, any driver with an overall value of 0 never went to traffic school (and a value of 1 or 2 represents how many times they went). We can then filter to keep only drivers with a value of 0.
2. This time, we'll use a `CASE` statement instead of `IF`. They function very similarly but have different syntax. The start of the calculation should look like this:

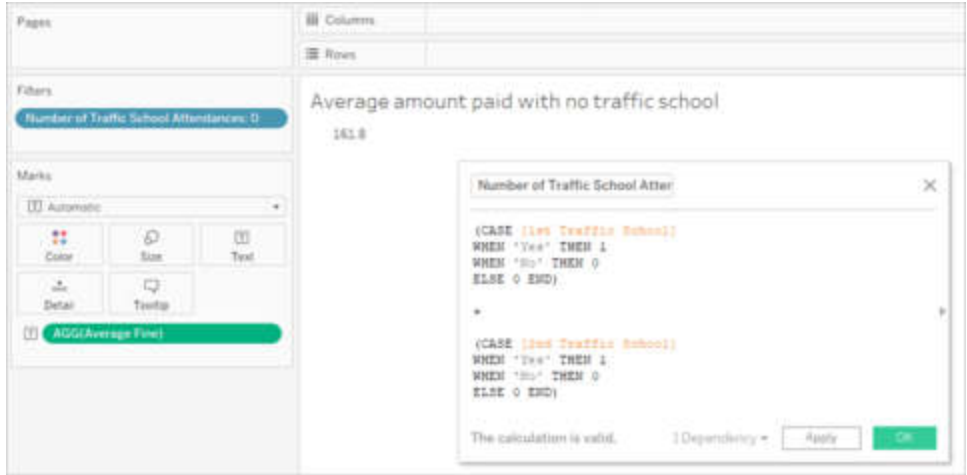
```
CASE [1st Traffic School]
WHEN 'Yes' THEN 1
WHEN 'No' THEN
ELSE 0
END
```

3. And then we'll do the same thing for 2nd Traffic School. We can add both pieces in the same calculation by wrapping each case statement in parentheses and adding a plus between them. Removing some of the line breaks, it looks like this:

**Number of Traffic School Attendances =**

```
(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
THEN 0 ELSE 0 END)
+
(CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
THEN 0 ELSE 0 END)
```

4. If we drag **Number of Traffic School Attendances** to the **Dimensions** area of the **Data** pane, the values 0–2 will become discrete.
  5. Now if we filter on **Number of Traffic School Attendances**, we can select just the 0 and know we're getting drivers who have never attended traffic school.
- C. To answer the original question, we'll simply bring **Average Fine** to the **Textshelf** on the **Marks** card. Because we built the aggregations into the calculation, the aggregation on the field will be **AGG** and we cannot change it. This is as expected.



## Go Further—Pivoted Data

While the data we've been working with is well structured to address questions specifically around first and second infractions, it isn't the standard structure recommended for use with Tableau Desktop. The more our analysis diverges from basic questions around the infraction dates, the more complicated our calculations become to combine the relevant information into useable form.

Usually, when data is stored with multiple columns for the same type of data (such as two columns for date, two columns for fine amount, etc.) and unique information is stored in the field name (such as whether it's the first or second infraction), this is an indication the data should be pivoted.

Performing a multiple pivot can handle this nicely. Back in Tableau Prep, and working from the end of the **Driver Infraction** Tableau Prep flow created in the step-by-step tutorial above:

1. From the final clean step, add a **Pivot** step that pivots by every duplicated field.



**Tip:** Use the plus  $\oplus$  icon in the upper right corner of the **Pivoted Fields** area to add more **Pivot Values**. Each set of fields (such as 1st and 2nd Fine

Amounts) should be pivoted together.

For more information about pivoting, see [Clean and Shape Data](#) on page 229.

The results can be tidied by removing null dates as well as renaming and reordering fields.

2. Add a cleaning step after the pivot. In the **Infraction Date** column, right-click on the null bar and choose **Exclude**.
3. Double-click the field name **Pivot1 Names** and rename it **Infraction Number**.
4. Drag fields as appropriate to reorder them as below:

Driver ID	Infraction Number	Infraction Date	Infraction Type	Traffic School	Fine Amount
MO-175001406	1st	05/30/2017	Speeding	Yes	118
SH-199751404	1st	03/04/2017	Speeding	Yes	130
AA-106451404	1st	07/05/2017	Running a red light	No	60
MP-174701406	1st	06/19/2017	Speeding	No	125
PO-188501402	1st	10/30/2017	Speeding	Yes	120
KL-166451406	1st	10/04/2017	Speeding	No	115
RA-199151402	1st	07/20/2017	Speeding	Yes	146
AJ-107951404	1st	10/15/2017	Speeding	Yes	130
BD-117701406	1st	12/25/2017	Speeding	Yes	140
CJ-120101402	1st	11/26/2017	Speeding	Yes	122
SN-207101402	1st	12/27/2017	Speeding	Yes	280
TS-213701404	1st	10/23/2017	Speeding	Yes	100
JR-156701404	1st	12/24/2017	Speeding	No	148
CM-127151402	1st	03/01/2017	Running a red light	No	55
JK-156251406	1st	12/25/2017	Speeding	Yes	140

5. From the new, pivoted data, create an output named **Pivoted Driver Infractions** and bring it into Tableau Desktop. (Don't forget to run the flow after adding the **Output** step.)

Now we can look at our five questions again with this pivoted data structure; you can expand each one for basic information about how to proceed if you get stuck.

**Note:** You can download the completed flow file [Pivoted Driver Infractions.tflx](#) to check your work, or download the workbook [Pivoted Driver Infractions.twbx](#) to look at the solutions in context. Remember that there may be alternative ways to interpret the analysis or pursue answers.

## 1. What was the length of time in days between the first and



## second infraction for each driver?

A. To answer this in Tableau Desktop, as we did with the first data set, we'll use the `DATEDIFF` function. This function requires a start date and an end date. This information is present in our data, but all in one field. We need to pull it out into two fields.

1. Create two preliminary calculated fields:

```
1st Infraction = IF [Infraction Number] = "1st" THEN
[Infraction Date] END
```

```
2nd Infraction = IF [Infraction Number] = "2nd" THEN
[Infraction Date] END
```

2. Because we want to make sure both of these values are available to be compared for each driver, we need to fix them to the level of **Driver ID**.

**Note:** Don't believe me? Try to do a `DATEDIFF` calculation with these two fields as they are: **Time Between Infractions** = `DATEDIFF('day', [1st Infraction], [2nd Infraction])`  
You'll get null results everywhere, because Tableau is trying to compare across a data structure that looks like this:

Driver ID	1st Infraction Date	2nd Infraction Date	Time between infractions
AA-106451404	Null	9/28/2017	Null
	7/5/2017	Null	Null
AF-108851406	Null	9/2/2018	Null
	5/9/2018	Null	Null
AJ-107951404	Null	12/31/2017	Null
	10/15/2017	Null	Null

Here, the row that knows what the first date is doesn't know what the second date is, and vice versa. To get around this, we'll use a `FIXED` Level of Detail expression to force these first and second dates to be related by **Driver ID**.

Edit each calculation as follows:

```
1st Infraction = { FIXED [Driver ID] : MIN ( IF [Infraction
Number] = "1st" THEN [Infraction Date] END ) }
```

```
2nd Infraction = { FIXED [Driver ID] : MIN ( IF [Infraction
Number] = "2nd" THEN [Infraction Date] END ) }
```

**Note:** The original calculation must be aggregated when embedded in an LOD expression. We can use any basic aggregation that will preserve the date value (so aggregations like SUM, AVG, or MIN work, but not CNT or CNTD).

- Now we can create the `DATEDIFF` calculation as follows:

```
Time Between Infractions = DATEDIFF('day', [1st Infraction],
[2nd Infraction])
```

- If we want to look at weeks or months, simply modify the date part (currently 'day').

The results will be identical to the outcome with the unpivoted data structure.

## 2. Compare the fine amounts for the first and second infractions. Are they correlated?

- To answer this in Tableau Desktop, we'll use very similar logic to the previous question. We'll use **Infraction Number** to identify if a given row is the first or second infraction, then pull out the fine amount accordingly.

- If all we want to do is make a scatter plot, we can skip the LOD portion and just use the `IF` calculation:

```
1st Fine Amount = IF [Infraction Number] = "1st" THEN [Fine
Amount] END
```

```
2nd Fine Amount = IF [Infraction Number] = "2nd" THEN [Fine
Amount] END
```

- However, if we wanted to compare and see the difference in amount between the first and second fines for a single driver, we'd run into the same null issue as with the dates. It can't hurt to wrap these calculations in a `FIXED` LOD, so it might be good just to do so from the start:

```
1st Fine Amount = { FIXED [Driver ID] : MIN ( IF [Infraction  
Number] = "1st" THEN [Fine Amount] END ) }
```

```
2nd Fine Amount = { FIXED [Driver ID] : MIN ( IF [Infraction  
Number] = "2nd" THEN [Fine Amount] END ) }
```

3. Create a scatterplot and bring out a linear trend line as before.

The results will be identical to the outcome with the unpivoted data structure.

### 3. Which driver paid the most overall? Who paid the least?

- A. To answer this question in Tableau Desktop, the pivoted data structure is ideal. All we need to do is bring out **Driver ID** and **Fine Amount** into a bar chart. The default aggregation is already **SUM**, so the total amount paid by the driver will automatically be plotted.

The results will be identical to the outcome with the unpivoted data structure.

### 4. How many drivers had multiple infraction types?

- A. To answer this question in Tableau Desktop, the pivoted data structure is ideal. All we need to do is bring out **Driver ID** and a **Count Distinct** of **Infraction Type** as a bar chart, and we'll have our answer.

The results will be identical to the outcome with the unpivoted data structure.

### 5. What was the average fine amount for drivers who never attended traffic school?

- A. To answer this in Tableau Desktop, we cannot simply divide the total fine amount by two, since some drivers only had one infraction. We also can't calculate the average fine per driver and take the average of those values, because averaging averages can lead to inconsistencies. Instead, we need to calculate the total amount paid by drivers who never attended traffic school, then divide by the total number of infractions associated with those fines.

1. First, we need to determine if each driver had a second infraction. We can leverage the fact **2nd Infraction Date** will be null if there was no second infraction and start building the calculation:

```
IFNULL(STR([2nd Infraction]), 'no')
```

This will return the date of the second infraction if it exists, or "no" if there was no second infraction.

**Note:** The `STR` portion of this calculation is necessary because `IFNULL` needs consistency of data type in its arguments. Because we want to return the string "no" for null values, we need to convert the date to a string as well.

2. Next, we need to turn this information into the number of infractions, 1 or 2. If the result of our `IFNULL` calculation is "no", then the driver should be marked as having one fine. Any other result should be marked as having two fines. The calculation is:

**Number of Infractions =**

```
IF IFNULL(STR([2nd Infraction]), 'no')= 'no' THEN 1
ELSE 2
END
```

3. Now we need to consider the average fine amount. We already have a single field for **Fine Amount**. All we need to do is divide that by our new **Number of Infractions** field, wrapping both in **SUM**:

**Average Fine =** (SUM([Fine Amount]) / SUM([Number of Infractions]))

- B. We also need to filter out drivers who attended traffic school. It looks like we could use the **Traffic School** field and filter on **Traffic School = no**. However, this would filter on *infractions* not associated with traffic school, not *drivers* who never went to traffic school. If a driver went to traffic school for one infraction but not a second, we don't want the second infraction to be considered here.

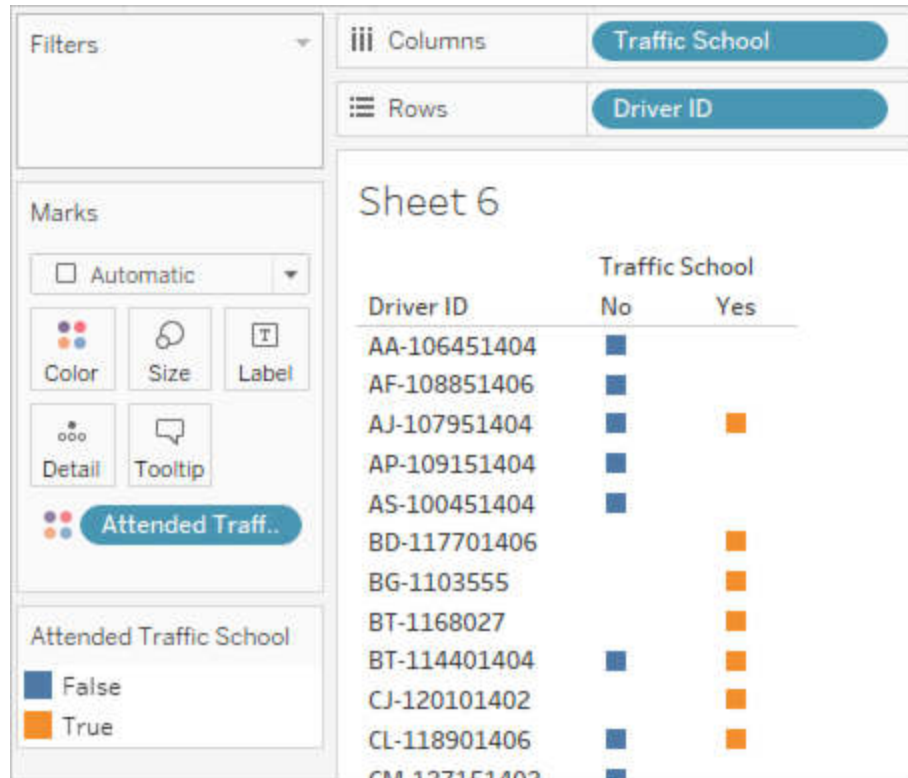
What we want to do is filter out any driver who's attended traffic school. In terms of the data, we want to filter out any driver who has a "Yes" for **Traffic School** on any row.

Let's build our calculation in stages, using a simple view to help keep track of what's happening:

1. First, we want to know if a driver has a "Yes" for Traffic School. Drag **Driver ID** to **Rows** and **Traffic School** to **Columns**. We'll get a text table with placeholder "Abc" text indicating the relevant values for each driver.
2. Next, we want to build a calculation that will identify if the value of **Traffic School** is "Yes". The first stage of the calculation is:

**Attended Traffic School** = CONTAINS ([Traffic School], 'Yes')

If we bring **Attended Traffic School** to the **Color** shelf on the **Marks** card, we see it accurately labels "False" for every mark in the "No" column, and "True" for every mark in the "Yes" column.



3. However, what we really want is this information at the level of the *driver*, not the *infraction*. An LOD expression is a natural fit when trying to compute a result at a different level of detail than the basic structure of the data. We'll make this a `FIXED` LOD expression. But, as we know, the aggregate expression portion of an

LOD must be aggregated. Previously, we've used **MIN**. Will that work here? We'll modify the calculation to be:

```
Attended Traffic School = { FIXED [Driver ID] : MIN( CONTAINS ([Traffic School], 'Yes')) }
```

The screenshot shows the Tableau interface with the following configuration:

- Columns:** Traffic School
- Rows:** Driver ID
- Marks:** Attended Traff.. (Color)
- Legend:** Attended Traffic School (False: Blue, True: Orange)

Driver ID	No	Yes
AA-106451404	■	
AF-108851406	■	
AJ-107951404	■	■
AP-109151404	■	
AS-100451404	■	
BD-117701406		■
BG-1103555		■
BT-1168027		■
BT-114401404	■	■
CJ-120101402		■
CL-118901406	■	■

With that change applied in the view, we see the opposite of what we want. Any driver that has a "No" is marked as "False" across the board. Instead, we want to carry the "Yes" as a "True" for every record for that driver. What is MIN doing here? It's picking the first response alphabetically, that is, "No".

4. What if we changed it to **MAX**? Would that take the last response alphabetically? We'll modify the calculation to be:

```
Attended Traffic School = { FIXED [Driver ID] : MAX ( CONTAINS ([Traffic School], 'Yes') ) }
```

The screenshot shows the Tableau interface with the following configuration:

- Columns:** Traffic School
- Rows:** Driver ID
- Marks:** Automatic (set to 'Attended Traff..')
- Legend:** Attended Traffic School
  - False (Blue square)
  - True (Orange square)

The resulting pivot table is as follows:

Driver ID	Traffic School	
	No	Yes
AA-106451404	■	
AF-108851406	■	
AJ-107951404	■	■
AP-109151404	■	
AS-100451404	■	
BD-117701406		■
BG-1103555		■
BT-1168027		■
BT-114401404	■	■
CJ-120101402		■
CL-118901406	■	■
CM-127151402	■	

And here we have it: if a driver has "Yes" anywhere in the data, they are marked as "True" for having attended traffic school, even on the infraction that didn't involve traffic school.

5. If we bring **Attended Traffic School** to the **Filter** shelf and select only "False", we'll be left with only drivers who never attended traffic school.
- C. To answer the original question, with our filter in place we'll simply bring **Average Fine** to the **Text** shelf on the **Marks** card. Because we built the aggregations into the calculation, the aggregation on the field will be **AGG** and we cannot change it. This is as expected.

The results will be identical to the outcome with the unpivoted data structure.

## The benefits of pivoted data

We could stick with the original data structure from the tutorial if we know we'd only need to answer questions that are easy to answer with that structure. However, the pivoted data format is more flexible. Even though it requires some calculations, once they're in place the resulting data set is well suited to answer broader questions.

## Go Further Still—Calculations Only

What if you don't have access to Tableau Prep? Are you out of luck entirely if you're stuck with the original data? Not at all!

Tableau Desktop and LOD expressions can answer all of our analytical questions. If we connect to the original **Traffic Violations.xlsx**, it looks very similar to the pivoted data set—just without the crucial **Infraction Number** field. We'll need to mimic the outcome of the aggregation steps via LOD expressions.

**Note:** You can download the workbook **LOD Driver Infractions.twbx** to look at the solutions in context. Remember that there may be alternative ways to interpret the analysis or pursue answers.

### 1. What was the length of time in days between the first and second infraction for each driver?

- A. To answer this in Tableau Desktop, we'll again use the `DATEDIFF` function. This function requires a start date and an end date. This information is present in our data, but all in one field. We need to pull it out into two fields. Because we want to make sure both of these values are available to be compared for each driver, we need to fix them to the level of **Driver ID**.

1. To find the first infraction date, we use the calculation:

```
1st Infraction = { FIXED [Driver ID] : MIN ( [Infraction Date] ) }
```

2. We'll do the second infraction date in stages.

- a. To start, we need to look at just the dates that are larger than the first date:

```
IF [Infraction Date] > [1st Infraction] THEN
  [Infraction Date] END
```

- b. But this will give us *every* infraction after the first, and we only want the second. So we want the smallest of these dates. Wrap the whole thing in **MIN**:



```
MIN( IF [Infraction] : [1st Infraction] THEN
[Infraction Date] END )
```

- c. We also want to recalculate the second infraction date for each driver. That's where LOD expressions come in. We'll fix this to the level of **Driver ID**:

```
2nd Infraction = { FIXED [Driver ID] : MIN ( IF
[Infraction Date] > [1st Infraction] THEN
[Infraction Date] END ) }
```

3. And we can now create the `DATEDIFF` calculation:

```
Time Between Infractions = DATEDIFF('day', [1st Infraction],
[2nd Infraction])
```

The results will be identical to the outcomes with the other two data structures.

## 2. Compare the fine amounts for the first and second infractions. Are they correlated?

- A. To answer this in Tableau Desktop, we'll use similar logic to the pivoted data version of this question. We'll use the **1st Infraction** and **2nd Infraction** fields we created for question 1 to identify if a given row is the first or second infraction, then pull out the fine amount accordingly.

1. If all we want to do is make a scatter plot, we can skip the LOD portion and just use an `IF` calculation:

```
1st Fine Amount = IF [1st Infraction] = [Infraction Date]
THEN [Fine Amount] END
```

```
2nd Fine Amount = IF [2nd Infraction] = [Infraction Date]
THEN [Fine Amount] END
```

2. However, if we want to compare and see the difference in amount between the first and second fines for a single driver, we'd run into issues with nulls, as in the first data structure. It can't hurt to wrap these calculations in a `FIXED` LOD, so it might be good just to do so from the start:

```
1st Fine Amount = { FIXED [Driver ID] : MIN ( IF [1st
Infraction] = [Infraction Date] THEN [Fine Amount] END )
}
```

```
2nd Fine Amount = {FIXED [Driver ID] : MIN( IF
[2ndInfraction] = [Infraction Date] THEN [Fine Amount]
END ) }
```

The results will be identical to the outcomes with the other two data structures.

### 3. Which driver paid the most overall? Who paid the least?

- A. To answer this in Tableau Desktop, we need to first realize something about the LOD-only method. Both methods using Tableau Prep filter out records that are not the first or second infraction for a driver. The LOD method in Tableau Desktop keeps all records. This means that if we were to create a viz of **SUM(Amount Paid)** by **Driver ID**, the Tableau Desktop-only version will show higher amounts for drivers with more than two infractions. To get a **Total Amount Paid** value from the complete data that matches the other methods, instead of using the original **Fine Amount** field, we instead need to sum the first and second fines like we did with the first data structure.
- B. Using the fields we created for question 2, we'll add the two fine amounts. **ZN** is necessary to prevent a null result for any drivers who only had one infraction. The calculation is:

```
Total Amount Paid = [1st Fine Amount] + ZN([2nd Fine Amount])
```

The results will be identical to the outcomes with the other two data structures.

### 4. How many drivers had multiple infraction types?

- A. To answer this question in Tableau Desktop, we can't simply bring out **Driver ID** and a **Count Distinct** of **Infraction Type**. Because this data set has infractions beyond the second, some drivers may have more than two infraction types. To match the results with the other methods, we need to limit the scope to just the first two infractions.

- B. We can to pull out the 1st and 2nd infraction types, wrap them in LOD expressions to make them `FIXED` to the driver, then use an `IF` calculation to count the types:

1. **1st Infraction Type** = { `FIXED [Driver ID] : MIN ( IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END )` }

2. **2nd Infraction Type** = { `FIXED [Driver ID] : MIN ( IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END )` }

3. **Number of Infraction Types** =

```
IF [1st Infraction Type] = [2nd Infraction Type] THEN 1
ELSEIF [1st Infraction Type] != [2nd Infraction Type]
THEN 2
ELSE 1 END
```

**Note:** It's also possible to create many of these calculations as a single field by nesting the initial calculations directly in the larger calculation. Here, the combined calculation would look like this:

```
IF
  {FIXED [Driver ID] : MIN(IF [1st Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
=
  {FIXED [Driver ID] : MIN(IF [2nd Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
THEN 1

ELSEIF
  {FIXED [Driver ID] : MIN(IF [1st Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
!=
  {FIXED [Driver ID] : MIN(IF [2nd Infraction]=
[Infraction Date] THEN [Infraction Type] END)}
THEN 2
```

```
ELSE 1
END
```

Which is a bit harder to make sense of, but works if preferred. (Note that line breaks and some spaces do not impact how a calculation is interpreted by Tableau.)

- C. We can then plot **Number of Infraction Types** against **Driver ID** and sort the bar chart.

The results will be identical to the outcomes with the other two data structures.

## 5. What was the average fine amount for drivers who never attended traffic school?

- A. To answer this in Tableau Desktop, we cannot simply divide the total fine amount by two, since some drivers only had one infraction. We also can't calculate the average fine per driver and take the average of those values, because averaging averages can lead to inconsistencies. Instead, we need to calculate the total amount paid by drivers who never attended traffic school, then divide by the total number of infractions associated with those fines.

1. First, we need to determine if each driver had a second infraction. We can leverage the fact the information in all the "2nd" fields will be null if there was no second infraction and start building the calculation:

```
IFNULL([2nd Infraction Type], 'no')
```

This will return an infraction type if it exists, or "no" if there was no second infraction.

2. Next, we need to turn this information into the number of infractions, 1 or 2. If the result of our `IFNULL` calculation is "no", then the driver should be marked as having one fine. Any other result should be marked as having two fines. The calculation is:

**Number of Infractions =**

```
IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1
ELSE 2
END
```

3. For the Total Amount Paid, we can use the calculation from question 3. To bring it all together, we'll take this total fine amount and divide it by our new **Number of Infractions** calculated field to determine the average fine amount:

```
Average Fine = SUM([Total Amount Paid]) / SUM([Number of
Infractions])
```

- B. We also need to filter out drivers who attended traffic school. Because this data set contains some drivers with a third or fourth infraction, we can't use the same method as the pivoted data structure. Instead, we'll follow the same method as the unpivoted data, summarized here:

1. First, we need to built two calculations identifying if the first and second infractions involved traffic school or not:

```
1st Traffic School = { FIXED [Driver ID] : MIN (IF [1st
Infraction] = [Infraction Date] THEN [Traffic School]
END ) }
```

```
2nd Traffic School = { FIXED [Driver ID] : MIN (IF [2nd
Infraction] = [Infraction Date] THEN [Traffic School]
END ) }
```

2. Then we'll add those values to get the overall number of traffic school attendances:

**Number of Traffic School Attendances =**

```
(CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
THEN 0 ELSE 0 END)
+
(CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No'
THEN 0 ELSE 0 END)
```

3. If we drag **Number of Traffic School Attendances** to the **Dimensions** area of the **Data** pane, the values 0–2 become discrete.
4. Now if we filter on **Number of Traffic School Attendances**, we can select just the 0 and know we're getting drivers who have never attended traffic school.

- C. To answer the original question, we'll simply bring **Average Fine** to the **Textshelf** on the **Marks** card. Because we built the aggregations into the calculation, the aggregation on the field will be **AGG** and we cannot change it. This is as expected.

The results will be identical to the outcomes with the other two data structures.

It's important to remember that this solution has a lot of nested calculations and LOD expressions. Depending on the size of the data set and the complexity of the data, performance could be an issue.

## Reflection on Methods

So which route should you go? That's entirely up to you and the tools at your disposal.

- If you want to steer clear of LODs, there's a data-shaping solution, though calculations might be necessary for some analysis ([Analysis in Tableau Desktop on page 472](#)).
- If you can shape the data and are comfortable with calculations—including LODs—the middle-of-the-road option provides the best flexibility ([Go Further—Pivoted Data on page 479](#)).
- If you're comfortable with LODs, there's minimal impact on performance, and/or you don't have access to Tableau Prep, solving this with LODs alone is a viable option ([Go Further Still—Calculations Only on page 488](#)).

At the very least, it's valuable to understand how aggregation in Tableau Prep and Level of Detail expressions in Tableau Desktop are interrelated and impact data analysis. As with most things in Tableau, there's more than one way to do anything. Exploring all the various options can help bring concepts together and let you pick the best solution for you.

## Calculations used:

### Driver Infractions

- **Time Between Infractions** = `DATEDIFF('day', [1st Infraction Date], [2nd Infraction Date])`
- **Total Amount Paid** = `[1st Fine Amount] + ZN([2nd Fine Amount])`
- **Number of Infraction Types** = `IF [1st Infraction Type]=[2nd`

```
Infraction Type] THEN 1 ELSEIF [1st Infraction Type]!=[2nd
Infraction Type] THEN 2 ELSE 1 END
```

- **Number of Infractions** = IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END
- **Average Fine** = ( SUM([1st Fine Amount]) + SUM( ZN([2nd Fine Amount]) ) ) / SUM([Number of Infractions])
- **Number of Traffic School Attendances** = (CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)

## Pivoted Driver Infractions

- **1st Infraction** = {FIXED [Driver ID] : MIN(IF [Infraction Number] = "1st" THEN [Infraction Date] END)}
- **2nd Infraction** = {FIXED [Driver ID] : MIN(IF [Infraction Number] = "2nd" THEN [Infraction Date] END)}
- **Time Between Infractions** = DATEDIFF('day', [1st Infraction], [2nd Infraction])
- **1st Fine Amount** = {FIXED [Driver ID] : MIN( IF [Infraction Number] = "1st" THEN [Fine Amount] END ) }
- **Number of Infractions** = IF IFNULL(STR([2nd Infraction]), 'no')= 'no' THEN 1 ELSE 2 END
- **Average Fine** = SUM([Fine Amount])/SUM([Number of Infractions])
- **Attended Traffic School** = { FIXED [Driver ID] : MAX( CONTAINS ([Traffic School], 'Yes')) }

## LOD Driver Infractions

- **1st Infraction** = {FIXED [Driver ID] : MIN([Infraction Date])}
- **2nd Infraction** = { FIXED [Driver ID] : MIN( IF [Infraction Date] > [1st Infraction] THEN [Infraction Date] END ) }
- **Time Between Infractions** = DATEDIFF('day', [1st Infraction], [2nd

```
Infraction])
```

- **1st Fine Amount** = {FIXED [Driver ID] : MIN( IF [1st Infraction] = [Infraction Date] THEN [Fine Amount] END ) }
- **2nd Fine Amount** = {FIXED [Driver ID] : MIN( IF [2nd Infraction] = [Infraction Date] THEN [Fine Amount] END ) }
- **Total Amount Paid** = [1st Fine Amount] + ZN([2nd Fine Amount])
- **1st Infraction Type** = {FIXED [Driver ID] : MIN( IF [1st Infraction] = [Infraction Date] THEN [Infraction Type] END ) }
- **2nd Infraction Type** = {FIXED [Driver ID] : MIN( IF [2nd Infraction] = [Infraction Date] THEN [Infraction Type] END ) }
- **Number of Infraction Types** = IF [1st Infraction Type]=[2nd Infraction Type] THEN 1 ELSEIF [1st Infraction Type]!=[2nd Infraction Type] THEN 2 ELSE 1 END
- **Number of Infractions** = IF IFNULL([2nd Infraction Type], 'no') = 'no' THEN 1 ELSE 2 END
- **Average Fine** = SUM ([Total Amount Paid]) / SUM([Number of Infractions])
- **1st Traffic School** = {FIXED [Driver ID] : MIN (IF [1st Infraction] = [Infraction Date] THEN [Traffic School] END ) }
- **2nd Traffic School** = {FIXED [Driver ID] : MIN (IF [2nd Infraction] = [Infraction Date] THEN [Traffic School] END ) }
- **Number of Traffic School Attendances** = (CASE [1st Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END) + (CASE [2nd Traffic School] WHEN 'Yes' THEN 1 WHEN 'No' THEN 0 ELSE 0 END)

**Note:** Special Thanks to Ann Jackson's Workout Wednesday topic [Do Customers Spend More on Their First or Second Purchase?](#) and Andy Kriebel's Tableau Prep Tip



[Returning the First and Second Purchase Dates](#) that provided the initial inspiration for this tutorial. Clicking these links will take you away from the Tableau website. Tableau cannot take responsibility for the accuracy or freshness of pages maintained by external providers. Contact the owners if you have questions regarding their content.



# Troubleshoot Tableau Prep Builder

This article lists problems you might encounter when using Tableau Prep Builder and suggestions for how to resolve them.

## Common errors when using the command line to run flows

You can run flows from the command line to refresh your output files programmatically instead of opening Tableau Prep Builder to run each flow manually. While this process helps build efficiency in your flow process, if your syntax is incorrect, or you are missing credentials for your connections or output locations, you will receive errors when running this process.

The following table describes common errors and how to resolve them. For information about how to run flows from the command line, see [Refresh flow output files from the command line on page 334](#).

Error	Cause	How to fix it
"Missing arguments"	One of the required command line arguments is missing.	Use "tableau-prep-cli -help" to see a list of the arguments for the command line.
"Unable to read the connections file."	There are errors in the syntax or format in the credentials.json file for the input connections.	Check the syntax for the input connections in the .json file. For more information and examples, see <a href="#">Refresh flow output files from the command line on page 334</a> .

<p>"There are errors in the flow. Unable to run the flow.</p> <p>Check that the credentials .json file includes all required credentials. Open the flow in Tableau Prep Builder to view error details."</p>	<p>There are missing credentials in the credentials.json file for the input connections or the flow has errors.</p>	<p>Check that the .json file has the credentials for all connections, and open the flow file in Tableau Prep Builder to see if there are any errors in the flow.</p> <p>If the flow has errors, you must fix them and republish the flow to Tableau Server, then try running the process again.</p>
<p>"Could not find match for &lt;hostname of inputConnections &gt;"</p>	<p>The credentials.json file is missing an entry for the hostname (server name).</p>	<p>Make sure the credentials.json file includes the correct credentials for the hostname (server name).</p> <p>For more information and examples, see <a href="#">Refresh flow output files from the command line on page 334</a></p>
<p>"We don't have credentials of all connections in tfl/t-flx file. The following connection(s) were not found: &lt;hostname of inputConnections&gt;"</p>	<p>The credentials.json file is missing or has incorrect</p>	<p>Make sure credentials.json file includes the correct credentials for the hostname (server</p>

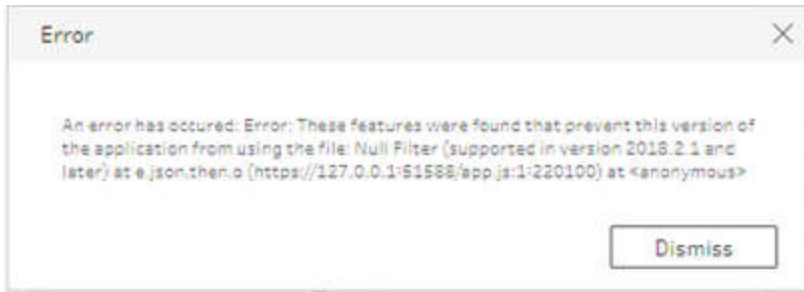
	<p>credentials for the hostname (server name) shown in the error message.</p>	<p>name) listed in the error message.</p> <p>For more information and examples, see <a href="#">Refresh flow output files from the command line on page 334</a>.</p>
<p>"Error signing in server &lt;serverUrl&gt; as a user &lt;userName&gt;. Please check the credentials."</p>	<p>The credentials.json file has the incorrect credentials for Tableau Server.</p>	<p>Make sure the credentials.json file includes all the correct credentials and elements for the output connection.</p> <p>For more information and examples, see <a href="#">Refresh flow output files from the command line on page 334</a></p>
<p>"Could not sign in successfully as &lt;userName&gt; to server &lt;serverUrl&gt;(&lt;contentUrl&gt;)"</p>	<p>The credentials.json file has the incorrect credentials for Tableau Server.</p>	<p>Make sure the credentials.json file includes all the correct credentials and elements for the output connection.</p> <p>For more information and examples, see</p>

		<a href="#">Refresh flow output files from the command line on page 334</a>
"We don't have credentials for Tableau Server to publish extract for one or more output nodes in tfl/t-flx file."	The credentials.json file was not passed in as a command line argument or it is missing the credentials for the output connection.	<p>Make sure the the path to the credentials.json file is included in the command line and verify that the credentials.json file includes all the correct credentials and elements for the output connection.</p> <p>For more information and examples, see <a href="#">Refresh flow output files from the command line on page 334</a></p>
"Loom rest api server not started"	The installation or environment setup is incorrect.	<p>Make sure that Tableau Prep Builder is installed correctly and that you are running the command as an Administrator.</p> <p>For information about how to install Tableau Prep</p>

		Builder, see <a href="#">Install Tableau Desktop or Tableau Prep Builder from the User Interface.</a>
"Error. Flow file does not exist."	The path to the flow file is incorrect.	Make sure that the correct path to the flow file is included in the command line.
"Error. Connections file does not exist."	The path to the credentials.json file is incorrect.	Make sure that the correct path to the credentials.json file is included in the command line.
"Could not find match for <mapr01:5181>,<mapr02:5181>,<mapr03:5181>"	You must specify a specific Port ID when connecting to Apache Drill using ZooKeeper .	Include a credentials.json file in the command line that specifies "port": 31010 for the input credentials.

## Error: "These features were found that prevent this version of the application from using this file"

If you open a flow that was created in version 2018.2.1 or later in an earlier version of Tableau Prep Builder, you may see the following error:



Flows that include features that are not supported in earlier releases will result in this incompatibility error. To resolve the error, open the flow in the later version, and save a copy of the flow without the indicated features. In the above example, remove the null filter from the field where it is applied.

Then open the copy that has the feature removed in the earlier version of Tableau Prep Builder.

## Error: "Failed to parse response from Tableau Server" when publishing Tableau Prep output

To successfully publish output from Tableau Prep Builder to Tableau Server, the REST API must be enabled on Tableau Server. If the REST API is not enabled, you will see the following error:

```
Failed to parse response from Tableau server due
to:javax.xml.bind.UnmarshalException - with linked
exception: [org.xml.sax.SAXParseException; lineNumber: 1;
columnNumber: 10; DOCTYPE is disallowed when the feature
"http://apache.org/xml/features/disallow-doctype-decl" set
to true.]
```

For information about enabling the REST API on Tableau Server, see [REST API Requirements](#) in the REST API Help.

For information about publishing output from Tableau Prep Builder, see [Create and publish data extracts and data sources](#) on page 322.



# Error: "You are using Server version: null..." when signing in to an SSL-enabled Tableau Server using Tableau Prep

When you sign in to an SSL-enabled Tableau Server from Tableau Prep Builder, you must have a root certificate installed on the computer where Tableau Prep Builder is installed. If the certificate is not installed, you might see the following error:

**You are using Server version: null but the minimum compatible version is: 10.0. Please upgrade to a compatible version**

If you see this error, work with your IT department or system administrator to install the required root certificate on the computer where Tableau Prep Builder is installed. For more information, see [System requirements](#) in the Tableau Desktop and Tableau Prep Builder Deployment Guide.

## Maintain Licenses for Tableau Desktop and Tableau Prep

**Important:** The Tableau licensing service was moved to a new data center on October 6, 2018. This means that any environments that required special configuration (static IP safe listing for example) to access [licensing.tableau.com](#) or [licensing.tableau.com](#) will need to be updated before you can activate, refresh, or deactivate a Tableau product key. For more information, see [Tableau Community](#).

Tableau Desktop and Tableau Prep can be licensed under a term license model. Term licenses, also known as subscription licenses, allow you to use and update Tableau Desktop and Tableau Prep for a specified period of time.

Term licenses must be renewed and the product key refreshed to continue providing uninterrupted service. You can continuously renew the term license as each specified period expires. If you don't renew your term license and the term expires, Tableau will stop working and you will no longer have access to the software. For more information about renewing your license, see [How to Renew your Tableau Licenses](#).

**Note:** Trial licenses for Tableau Desktop or Tableau Prep expire after a set period of time, usually 14 days. After the trial period expires, you'll need to [purchase a license](#) to continue using the product.

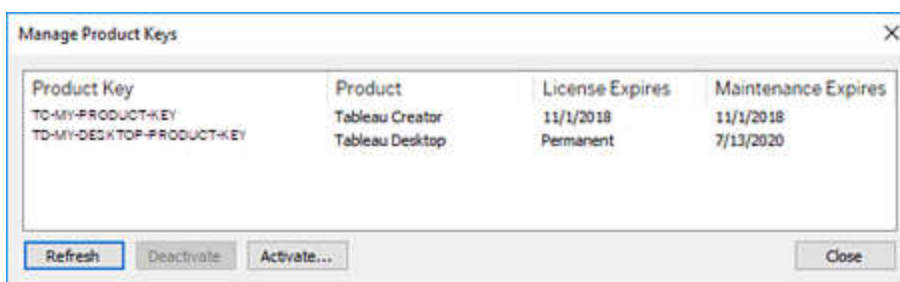
## View data about your license

After you install Tableau Desktop or Tableau Prep open the application and then navigate to **Help > Manage Product Keys** from the top menu to see information about the type of license you have and when it expires.

You can also activate or deactivate a product key or refresh a maintenance product key from this dialog.

**Note:** Tableau offers term licenses that provide a range of capabilities. The type of license that you have is displayed in the **Product** field. For more information about the different type of user-based licenses that are available, see [User-based licenses](#) in the Tableau Server help.

Existing Tableau Desktop users may have a perpetual (permanent) license. Perpetual licenses don't expire. However, to get access to product updates and technical support you must purchase Support and Maintenance services. These services must be renewed to continue receiving the service. Perpetual (permanent) licenses are no longer available for Tableau Desktop.



Use the following buttons to take action on your product key:

- **Refresh** (Tableau Desktop only): Click the **Refresh** button to refresh a maintenance license that is expiring, then close and restart Tableau Desktop. If the **Maintenance**

**Expires** date doesn't update, check with your license administrator as the key or maintenance agreement may have changed.

To refresh a maintenance key from the command line see [Refresh the product key](#) in the Tableau Desktop and Tableau Prep Deployment guide.

**Note:** You cannot refresh the product key if Tableau Desktop is offline. If you are activating Tableau Desktop in offline mode, you must obtain and activate a new key from the Tableau Customer Portal.

- **Deactivate:** Select a product key in the list then click **Deactivate** to deactivate the product key. Deactivate a product key if you need to move the product key to another computer or when you no longer need the product key on this computer.

For more information about deactivating a product key, see [Move or Deactivate Product Keys](#) in the Tableau Desktop and Tableau Prep Deployment guide.

- **Activate:** After Tableau Desktop or Tableau Prep is installed, click **Activate** to open the activation dialog and enter your product key. If you get an error and can't activate Tableau Desktop or Tableau Prep using your product key, contact [Tableau Support](#).

For more information about activating a product key, see [Activate and Register your Product](#) in the Tableau Desktop and Tableau Prep Deployment guide.

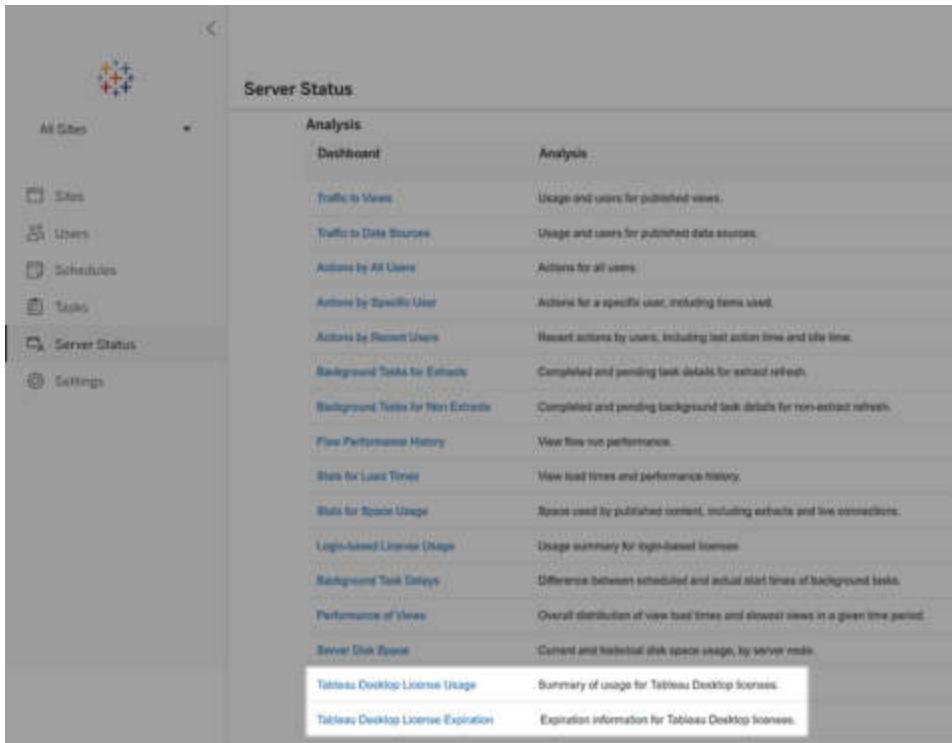
## Track Tableau Desktop license usage and expiration data

If you want to track and view license usage and expiration data for Tableau Desktop in Tableau Server you must configure Tableau Desktop to send license data to Tableau Server on a set interval, and then enable reporting on Tableau Server.

This enables server administrators to access two reports:

- **Desktop License Usage:** This report lets server administrators see usage data for Tableau Desktop licenses in your organization.
- **Desktop License Expiration:** This report gives server administrators information about which Tableau Desktop licenses in your organization have expired or need maintenance renewal.

If Tableau Desktop and Tableau Server are configured for license reporting, when signed in to Tableau Server as an Administrator, you will see these two reports listed on the **Server Status** page in the **Analysis** section.



If you don't see these reports listed, then Tableau Desktop and Tableau Server may not be configured for Tableau Desktop usage reporting.

For information about how to configure Tableau Desktop and Tableau Server for usage reporting, see [Manage Tableau Desktop License Usage](#) in the Tableau Desktop and Tableau Prep Deployment guide.

## Additional resources

For more information about managing your license refer to the following topics:

- To find your product key and activate Tableau Desktop or Tableau Prep Builder, see [Where's my product key.](#)
- To deactivate a product key or move it to another computer, see [Move or Deactivate Tableau Desktop.](#)
- To learn more about product keys for non-persistent virtual desktops or for computers that are regularly re-imaged, see [Configure Virtual Desktop Support.](#)
- To learn more about product key management for Tableau Server or Tableau Online, see Licensing Overview ([Linux](#) | [Windows](#))

## Tableau Prep Help

- To learn more about the license renewal process or to renew a license, see [How to Renew your Tableau Licenses](#).