Tableau Prep Tutorial

Use the tutorials below to help you gain a deeper understanding on what Tableau Prep can do for your data, business and overall visuals. Follow each step in the tutorial of jump around to the tutorial that fits your needs.

Getting Started with Tableau Prep

Go to the Tutorial

Tableau Prep is a great tool. Get used to the new interface and build your first Flow. What you learn

- Connect Data
- Build a Flow
- Add Steps
- Clean and Change Data

Tableau Prep Interface and Steps

Go to the Tutorial

Master the Tableau Prep Interface and learn about the different steps in a Flow

- Input Step
- Cleaning Step
- Pivoting
- Joining and Union
- Output Step

Profile Pane and Pivoting

Go to the Tutorial

Explore data on a deeper level with the profile pane and pivoting.

- Make Data Tidy
- Data Distribution
- Pivot and Unpivot

The Data Cleaning Step

View Tutorial

Learn how to replace, filter, and clean data with this essential step. Follow these steps to make your data ultra tidy.

- Replace
- Filter
- Clean
- Split
- Annotate

in **TABLEAU** | August 21, 2019

Related Posts

- Tableau Prep – The Cleaning Step, Group and Replace
- Tableau Prep – The Profile Pane and Pivot Step
- Getting Started with Tableau Prep
- The Tableau Prep Interface and the Input Step
- Waffle Chart in Tableau
- Pareto Chart in Tableau
- EXCLUDE – LOD Deep Dive in Tableau
- INCLUDE – LOD Deep Dive in Tableau

Gaelim Holland

Leave a Reply

Start the discussion...

Subscribe

Previous post Next post

Help us improve this website
Question 1 of 4 or fewer:
Overall, how satisfied are you with this website?
Very dissatisfied
Somewhat dissatisfied
Neither satisfied nor dissatisfied
Somewhat satisfied

Copyright text 2019 by AbsentData.

Getting Started with Tableau Prep

Many data scientists have heard the line – "People spend 80% of their time prepping data and only 20% of their time analyzing it." And now, as an overload of data becomes available, dealing with unlikely data becomes excessively difficult. Cleaning data is known as one of the most tedious jobs in a data scientist's career.

What does Tableau Prep do?

But what does Tableau have to do with this? Tableau happens to be one of the leading data visualization tools in the industry, and now, with Tableau Prep, it has incorporated a missing feature into its offerings. Tableau Prep makes doing routine tasks, such as joins, unions and pivots a lot easier to do. And, of course, you don't need any programming skills to do this, which is a major reason for Tableau's success.

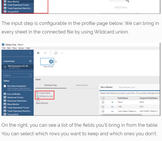
Let's get started!

Building a Flow

- **Connecting to Data**

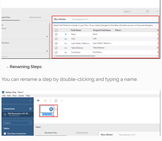
First, we'll connect to our data set.

Once you've opened the file, you will see this screen.

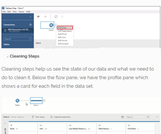


We'll drag out a sheet and here we have our first step in the file.

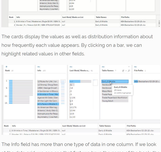
- **The Input Step**



The input step is configurable in the profile page below. We can bring in every sheet in the connected file by using Wildcard union.

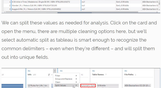


On the right, you can see a list of the fields you'll bring in from the table. You can select which ones you want to keep and which ones you don't.



- **Renaming Steps**

You can rename a step by double-clicking and typing a name.



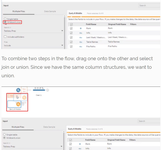
Adding Steps

To add another step in the flow, we'll click the plus button. Since we want just a basic cleaning step to start, we'll select 'Add Step'.

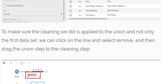
- **Cleaning Steps**

Cleaning steps help us see the state of our data and what we need to do to clean it. Below the flow pane, we have the profile pane which shows a card for each field in the data set.

The cards display the values as well as distribution information about how frequently each value appears. By clicking on a bar, we can highlight related values in other fields.

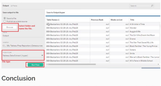


The info field has more than one type of data in one column. If we look at the data grid, we can see that first, we have the name of the book, the author name, the dollar price, and the ISBN number, each separated by a pipe.



We can split these values as needed for analysis. Click on the card and open the menu, there are multiple cleaning options here, but we'll select automatic split as tableau is smart enough to recognize the common delimiters – even when they're different – and will split them out into unique fields.

To rename the fields, just double click on the name, and type out another name.



As we no longer need the original, un-split field, we can delete it.

- **Changing Data Types**

Price is currently a string data type, but it should be a decimal. We can click the data type icon and select number (decimal).

We have some more weeks' worth of data in the data set, so we can add that. Add a new connection, and bring in a table, and select wildcard union.

To combine two steps in the flow, drag one onto the other and select join or union. Since we have the same column structures, we want to union.

To make sure the cleaning we did is applied to the union and not only the first data set, we can click on the line and select remove, and then drag the union step to the cleaning step.

- **Output Step**

Our data is prepped and ready for use, so we'll add an output step. We'll click the plus and select 'Add Output'. We can choose what file type to save the file as, where to save it, and what to name it. Now when we run the flow, we create a new file. This file contains all our data, cleaned and combined.

Conclusion

Prepping data in Tableau Prep can be simple or complex, and each step has robust options. Thank you for reading this article on Tableau Prep, and I hope you will continue to read the next articles in this series.

in TABLEAU / August 19, 2019

Related Posts

- Tableau Prep – The Cleaning Step, Group and Replace
- Tableau Prep Tutorial
- Tableau Prep – The Profile Pane and Pivot Step
- The Tableau Prep Interface and the Input Step
- Waffle Chart in Tableau
- Pareto Chart in Tableau
- EXCLUDE – LOD Deep Dive in Tableau
- INCLUDE – LOD Deep Dive in Tableau

Authors

Leave a Reply

Start the discussion...

Subscribe

Previous post Next post

Copyright text 2019 by AbsentData

Search Blog

Recent Posts

- Rolling Averages in Power BI
- Python Moving Average
- Sales Prediction Model in Power BI
- Easy Forecasting in Excel
- Data Analyst Toolkit

Tableau Prep – The Cleaning Step, Group and Replace

Introduction

Data is like blood flowing through an organization and just like blood quality is directly linked with the health of the body, data quality impacts function effectively when the quality of their data meets the minimum quality standards.

In real life, data that is used for visualization is typically sourced from a variety of different sources. Because data is coming from different sources it could be in different formats, may have different meanings for the same data values, may have different purpose for the same data field, and may have different levels of data quality. To ensure that this data can be aggregated in a meaningful and consistent manner for visualization and analytical purposes, data must be cleaned as part of the data preparation step.

This article is part of a series of 5 articles on Tableau Prep and focused on Cleaning, Grouping and Replace. It assumes that the reader has fundamental knowledge about data visualization and has some basic level of exposure tino one or more tools for data visualization and analytics. Having read the previous articles, is recommended but not a pre-requisite.

Having read the previous articles is recommended but not a pre-requisite. As a case study, I will use an Excel file that contains foreign exchange rates against the Canadian dollar. this file was downloaded from the Bank of Canada website. Additionally, I will also use a data file that was downloaded from the Statistics Canada Website.

Cleaning Step

In this section, we will use the cleaning step available as part of the Tableau Prep data flow. But before we get into the details of Tableau Prep, first let's have a quick look at the sample data that we are going to use. As I mentioned, this is an Excel file that contains foreign exchange data.

Using Data Interpreter to Remove Extra Rows in Source Data

Look at the screenshot below showing the Excel file. You will notice that the first 39 date rows in the Excel file do not contain any data. These rows contain information about the data itself. While it is useful to be aware of this information it is unlikely that this information will be used directly as part of the visualization process, by Tableau or any other visualization tool. Hence, these rows have to be removed from the data.

In order to do that, let's open Tableau Prep and connect the Excel file. Once you have it connected the file, select **cleaned with data interpreter.**

You will notice that the Tableau Prep has suggested 2 tables to you after an automated analysis of the data in the Excel file. Select the second table on the bottom.

Remove and Rename

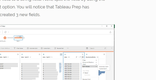
In order to do that, let's open Tableau Prep and connect the Excel file. Once you have it connected the file, select **cleaned with data interpreter.**

You will notice that the Tableau Prep has suggested 2 tables to you after an automated analysis of the data in the Excel file. Select the second table on the bottom.

Remove and Rename

Split and Change Data Type

To demonstrate the split function, let's assume we want to split the date fields into the year, month and date. In order to do this, let's first convert the type of the field to a string field. Next, split the field by using the automatic split option. You will notice that Tableau Prep has automatically created 3 new fields.

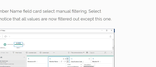


Clean

For this step, we will use another data file from https://www.ncdc.noaa.gov/cdo-web/datasets#GHCND. Download the Excel file and use the data interpreter in the input step to select the correct data table. Once loaded, add a Clean step after input. From the clean options, convert all the member name to Upper Case. Also, remove all the punctuations from the data by selecting the remove punctuations options.

Filtering

Using the Member Name field card select manual filtering. Select Extortion and notice that all values are now filtered out except this one.

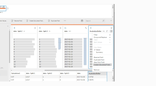


Group and Replace

Select the Member Name field, group and replace by Pronunciation. Notice that identifies Fraud and Identity Theft are grouped together now as an example. Please also refer to the section "Grouping" in this article for additional information.

Calculated Values

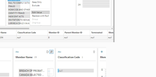
Now let's go back to our original data set of foreign exchange data from Bank of Canada. Go to the Australian dollar card and select the calculated field option.



You will notice a library of available functions to use. Select the Round() function to round off the exchange rate to decimal places.

Edit Values

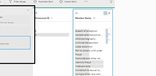
Go to the Member Name Card. Select the Extortion value, select the edit values and change it to Cyber Extortion.



Annotations and Changes

When you perform changes to the field cards, annotations are automatically added to them to show something has changed. This is shown in the screenshot.

Changes, on the other hand, is a concept which is similar to the system audit log in the sense that they are generated automatically by Tableau to capture what changes you applied manually to the data as part of the flow. It also allows you to maintain a log of all changes in a meaningful format, modify each step subsequently, change the sequence of the steps or remove steps altogether. Please refer to the screenshot.



Group and Replace

Significance of group and replace

A lot of times, when we use data, we find variations and inconsistencies within the same field. Especially, when data is being aggregated from various sources. For e.g. Muhammad may be spelled as Mohammed or Mohammed and USA may be spelled as America, United States etc. For effective and accurate analytics, it is important to be able to group them together.

Tableau Prep provides different options to achieve grouping. Some of the key approaches are outlined below:

Manual Selection

When manually selecting the values, the user can select which values to group together using the user interface. You can do this by going to the field card options, then selecting manual selection under group and replace the menu.

Out of domain values

In addition to selecting and grouping from available data values, you can also manually insert values that may not exist in the current data set but may appear in the future. These are called out of domain values and are indicated through a coloured dot against the respective value. You can add an out of domain value by pressing the + icon in the group and replace menu.

Fuzzy matching algorithms – Pronunciation and Common Characters

Tableau Prep has built-in algorithms to perform automated matching and grouping. Two such frequently used algorithms are matching by pronunciation and characters. Pronunciation groups together values for which the pronunciation is similar for e.g. John and Jon. Common Characters on the other works by comparing values for common characters in between them and grouping the ones where there is a match so for e.g. Wasge, Gaelim and Gaelim, Wasge would match since they have the same set of characters.

Conclusion

Thank you for reading this article. We hope that you have developed a foundational understanding about the Cleaning and Grouping features in Tableau Prep.

in TABLEAU / August 28, 2019

Related Posts

- Tableau Prep Tutorial
- Tableau Prep – The Profile Pane and Pivot Step
- Getting Started with Tableau Prep
- The Tableau Prep Interface and the Input Step
- Waffle Chart in Tableau
- Pareto Chart in Tableau
- EXCLUDE – LOD Deep Dive in Tableau
- INCLUDE – LOD Deep Dive in Tableau

Authors

Leave a Reply

Start the discussion…

Subscribe

Previous post | Next post

MeserData
I created this blog as a launch pad for my ideas and to inspire you to evaluate data that matters.

Blog
Google Tag Manager
Tableau
Analytics Tutorials
Adwords
Excel
Power BI
PANDAS

Join Our Mailing List
Your Email
Subscribe

Copyright text 2019 by MeserData

Tableau Prep – The Profile Pane and Pivot Step

Introduction and Significance

Exploring data is probably one of the first things a data analyst would do when dealing with a new data set. It is important to understand the profile or the characteristics of the data before you start analyzing and visualizing it.

Case Study

For this article, let's use a hypothetical case study of a large bank in Canada. Consider aggregating IT risk data for this bank with each of the various geographies reporting their respective data on IT risk incidents, loss events and issues. Now each geography, in turn, has different operational processes and systems of how they are managing this risk. So, for e.g. Canadian Business units are using RSA GRC System. US business units use excel file stored on SharePoint. Business Units in India use their home-grown system developed in Visual Studio .Net. Your job is to collect data from all of these different sources and aggregate them for reporting purposes to the board of directors.

As you would have figured out by now, the real challenge here it to make the data consistent for aggregation.

Profiling

Understanding the relationship within the data set is a key step. You can identify patterns and outliers with the data using this step. Consider the following fields in the source data

- Finding ID
- Residual Risk Rating
- Issue Title
- Issue Summary
- Line of Business
- Issue Owner
- Created Date
- Health Status

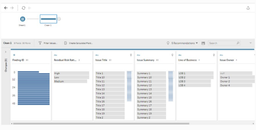
Considering the above, it will be important to see the profile for each of the fields and how do they relate to each other.

So, for example, you could have the following situations (only for illustration purposes, not an exhaustive list):

Distribution of Data:

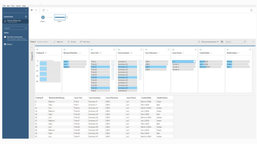
You can review the distribution of data – visually. Refer to the screenshot below and note for e.g.

- Data contains issue for the Month of Jan, Feb, and March of 2019
- There are 4 lines of business in this data set (LOB1, 2, 3 and 4)
- There are only 3 owners for all the 50 issues



Identification of Errors and Relationship:

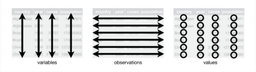
Null or Missing Values in Issue Owner Field: For legitimate business reasons it might be possible to have issues without an owner defined yet, although this is expected the situation to have as an exception, it is not expected to have a large portion of issues without owners. You can use Tableau Prep to identify patterns for such values. For e.g. Refer to the screenshot below. You can click on Null Values in the Owner Field to quickly reveal that there is one line of business (i.e. LOB 3) which is causing the problem.



To identify additional relationships

1. Click on any data value, Tableau Prep will highlight the related data values in other field cards.

Tidy Data and Columns to Rows



It is important to put your data in the correct format. The tidy data philosophy will help you to understand when to pivot and unpivot your data. The above image highlights how tidy data should be.

1. Variables should be a single column
2. Observations should be a single row
3. Each value should be in a cell.

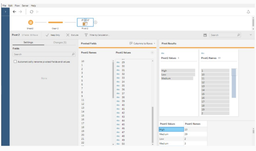
Short and Wide Data Bad , Long Skinny Data Good

Pivot is a frequently used method in data analytics. In Tableau Prep, the key use of Pivot Step is to convert data sets that have many columns into a smaller number of columns which are easier to then analyze in Tableau. Consider the following data format.

Notice that it has 1 finding in each column (versus a row). This format is hard for Tableau to analyze and must be pivoted (or unpivoted) to make it analyzable.

Pivot to Unpivot

Using the Pivot Step in Tableau we can achieve this. See the screenshot below. Notice how we have created a column for Finding ID and another column for risk rating for those finding ID.



Conclusion

Data Profiling is a fundamental activity for a data analyst before they use the consume the data in Tableau for visualizations. Typically pivoting step is used to unpivot Excel data with a large number of columns which are easier to read by humans but difficult to be used by Tableau for visualization purposes, and hence, it needs to be transformed as part of preparation.

in **TABLEAU** | August 20, 2019

Related Posts

- Tableau Prep – The Cleaning Step, Group and Replace
- Tableau Prep Tutorial
- Getting Started with Tableau Prep
- The Tableau Prep Interface and the Input Step
- Waffle Chart in Tableau
- Pareto Chart in Tableau
- EXCLUDE – LOD Deep Dive in Tableau
- INCLUDE – LOD Deep Dive in Tableau

Author1

Leave a Reply

Start the discussion...

Subscribe

Previous post | Next post

Copyright text 2019 by AbsentData

The Tableau Prep Interface and the Input Step

Welcome back to the articles on Tableau Prep. Today we'll be discussing the interface of Tableau, how to use it, and the Input step in Tableau.

The Start Screen and Connecting to Data

When we open Tableau, we begin on the start screen. At the far right, we have the discover pane, which shows relevant resources and jobs to relevant content. In the middle, we are shown previously opened files, and to the left, we have the connections pane. We will open a file in the way described in the previous article.

The Input Step

We will need a data file from the connections manager and drag out a table. This creates the first step in the flow and input step. There can be multiple input steps, and whenever you click on one, the related input pane will be brought up.

In the input pane, you can select which rows you want in your flow, and how you want to sample your data from the Data Sampler tab.

From the 'Multiple Files' tab, you can add files to the rows by clicking the 'filter values' option. Upon clicking, you will be shown a calculations screen, where you can add the filters.

If you want to remove the filter, right-click the filter option in the table and select 'remove'.

To use a new connection, simply click the plus sign. We have a multiple of options to select from: a cleaning step, an aggregate, pivot, join, union, and an output.

In case of multiple sets of data, select 'Wildcard union'. This will have to select 'Wildcard Union for every sheet'.

Cleaning Step

To add another step, we can click the plus sign. We have a multitude of options to select from: a cleaning step, an aggregate, pivot, join, union, and an output.

A cleaning step shows up as a simple tab on the flow pane. Below, we have the profile pane, which shows different cards for each column in the data. Each card displays the value in that column and the number of rows for each value. For more information, check out the article on the Profile Pane and Pivot Step.

The cleaning operations are shown in the menu on the individual cards or the contextual toolbar.

We can also perform cleaning operations by direct interaction, such as changing the data type or renaming the field.

Each type of cleaning step we perform will be shown above the cleaning step icon and will be tracked in the changes to the left.

For more information on the cleaning step, check out the article on The Cleaning Step and Group and Replace.

Types of Steps

Here, we have opened a more complex flow. This flow can be broken down into algorithmic parts. There are a series of pivots that restructure January's transaction data, bringing it all together with joins.

There's a union that combines several weeks' worth of data for February.

March is also pivoted, and all three months are combined with a union.

Another data source is brought in, which brings with it sales information and aggregates it from transactional data to bons level data, then joins it with bonsheets.

Finally, income escalation data is also brought in via another join, and an output is generated to fully cleaned and combined data.

Pivoting

We can also join as deep to bring up the pane related to that step. First, let's start with a pivot step, shown by the rulers/ruler icon.

Fields are brought in from the left rulers list of field to the Pivot Values Drop area. The result, that is, the pivoted fields, are shown on the right. Pivot values are shown in the Pivot Values Drop area. For more information, the Pivot step, read the article on the Profile Pane and Pivot Step.

Next, let's check out an aggregation step, which is indicated by the sigma symbol. Once again, fields are shown on the left and brought either to the grouped fields' area or the aggregated fields area. For more information on the aggregation step, check out the article on the Aggregate Step and the Join Step.

Joining and Unioning

Combining data can be done by either joining data or unioning it. Let's look at a union step.

A union shows the inputs going in on the left, and on the right, we are shown the resulting fields and the source of the data. Unions are a field-level display a colored bar, showing which inputs each field has fed. There's also a Table Names column, generated by Tableau, which shows which data source each row came from.

Another way to combine data is joining.

The join step pane provides a visual interface for configuring the join. To the left, we can build a join clause or clauses, choosing which field to join and how.

We can select the join type by clicking on the diagram, and under the join, we can see the summary of results, showing what has been included in the join and what has been excluded.

Immediately to the right of the join configuration pane, we can see a summary of the resulting fields, including which inputs the resulting fields are sourced from. Further to the right, we can the profile pane, and the data grid pane. For more information on joining, take a look at the article on the Aggregate step and the Join step.

Output Step

Finally, once the data has been cleaned and prepared, we must create the output step. Clicking on the output step, we can see the output pane, an output of the data, if packaged flow, the includes the flow and extracts of the flat files, through not the files from the connections. To use the file, an output step must be created and the flow must be run.

Clicking Run Flow will execute the flow and generate an output file or publish the data source. Note that if a flow has multiple outputs, they can either be run separately or simultaneously. For more information on the Output step, read the article on the Union step and the Output step.

Conclusion

Thank you for reading this article, and we hope you learned some more basics of Tableau.

- TABLEAU - August 25, 2019

Related Posts

- **Tableau Prep – The Cleaning Step, Group and Replace**
- **Tableau Prep Tutorial**
- **Tableau Prep – The Profile Pane and Pivot Step**
- **Getting Started with Tableau Prep**
- **Waffle Chart in Tableau**
- **Pareto Chart in Tableau**
- **EXCLUDE – LOD Deep Dive in Tableau**
- **INCLUDE – LOD Deep Dive in Tableau**

Authors

Leave a Reply

SAY YO DISCUSSION...

Subscribe ▾

Search Blog

Recent Posts

Nesting Averages in Power BI
Python Missing Average
Semi-Freedom Analytics Excel Tutorial
Easy Forecasting in Excel
Daily Analytic Teaser

Blog
Google Sheets Manager
Tableau
Analytics Tutorials
Salesforce
Excel
Power BI
Website

Join Our Mailing List

Your Email

Subscribe

Copyright © 2020 Analytics Demystified.