Course 4: Applying Machine Learning to your Datasets

Module 0: Course Intro

Lesson Title: **Introduction**

Format: Talking Head

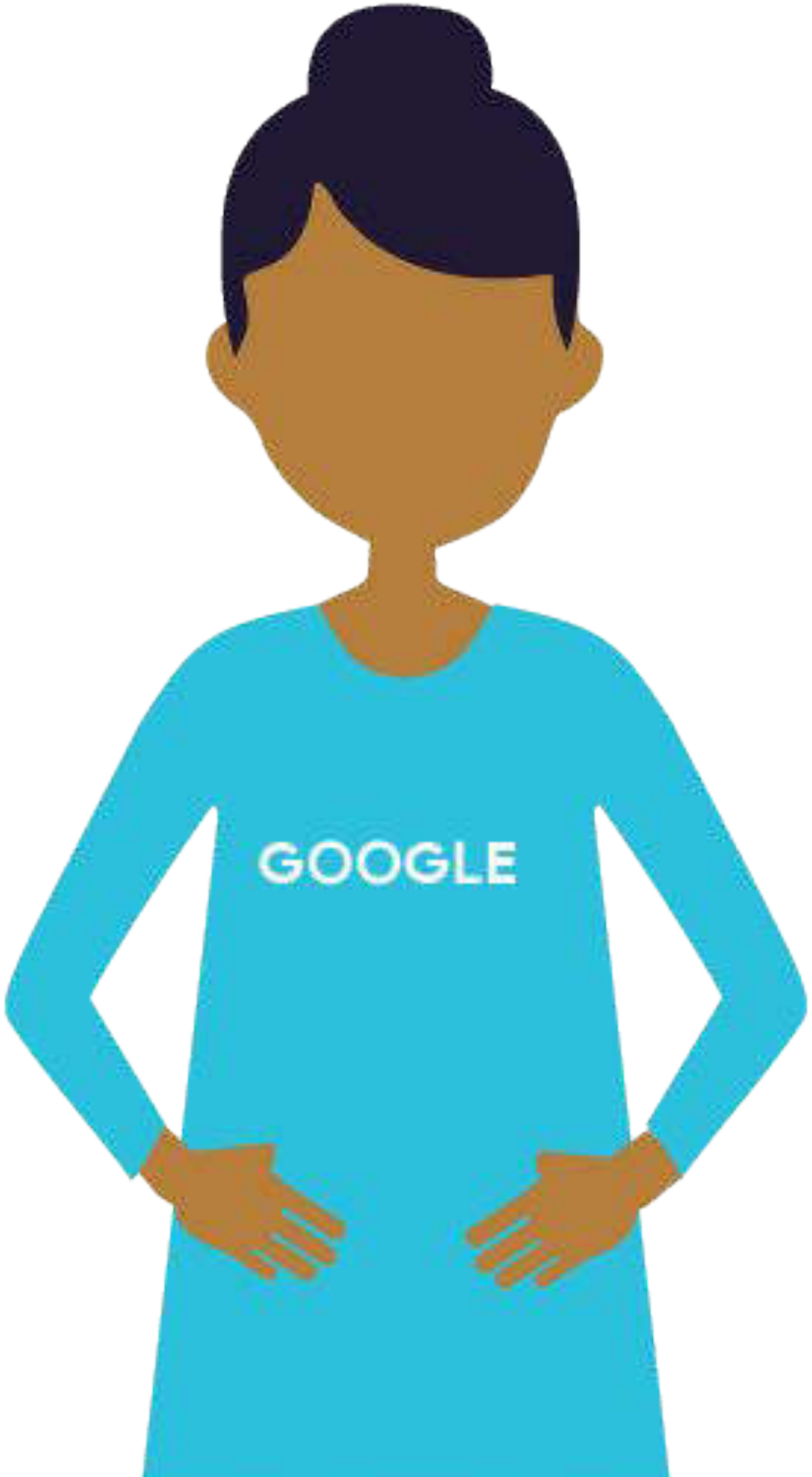
Video Name: T-BQML-O_0_I1_course_introduction

Google Cloud

From Data to Insights

On Google Cloud Platform

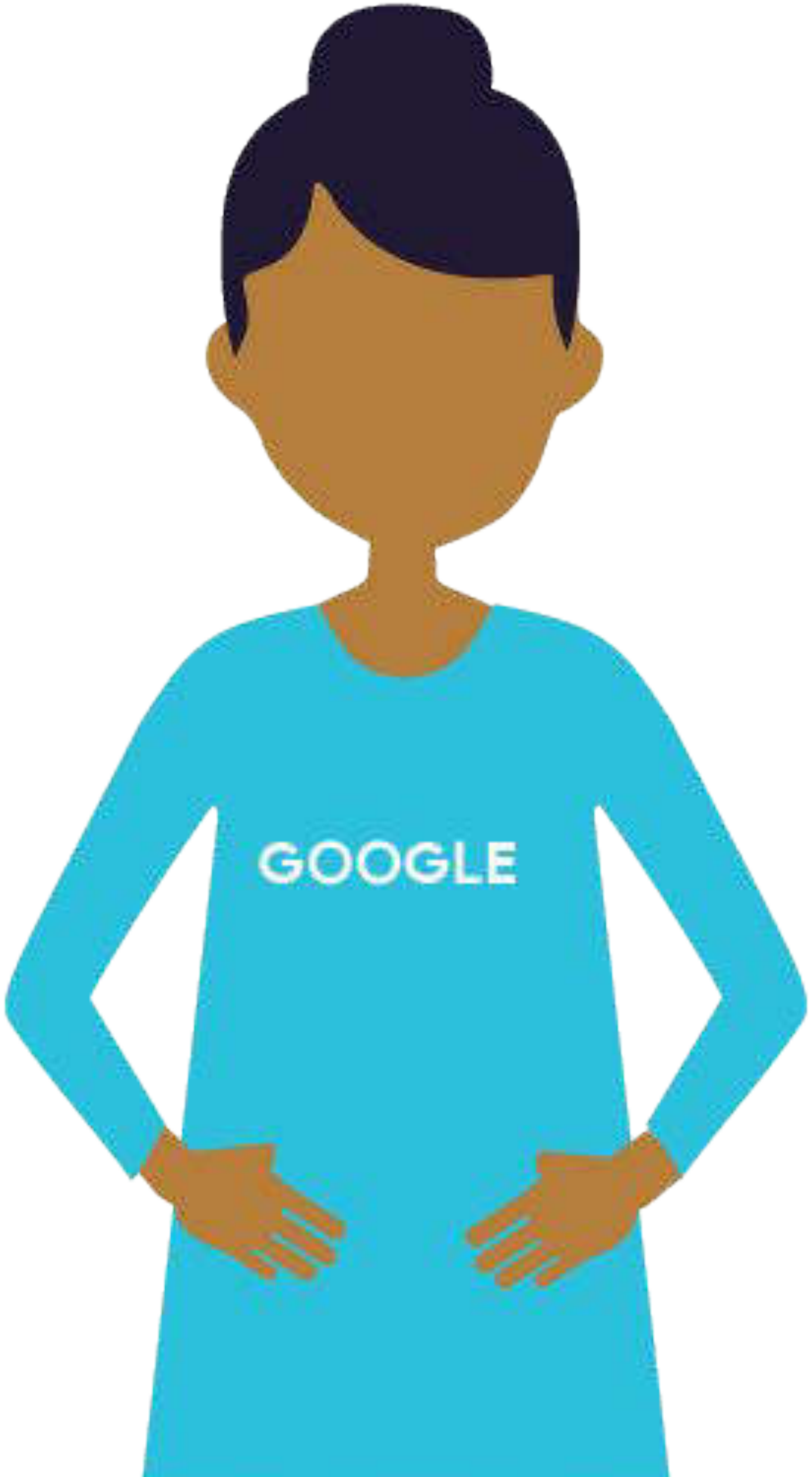
Evan Jones



From Data to Insights

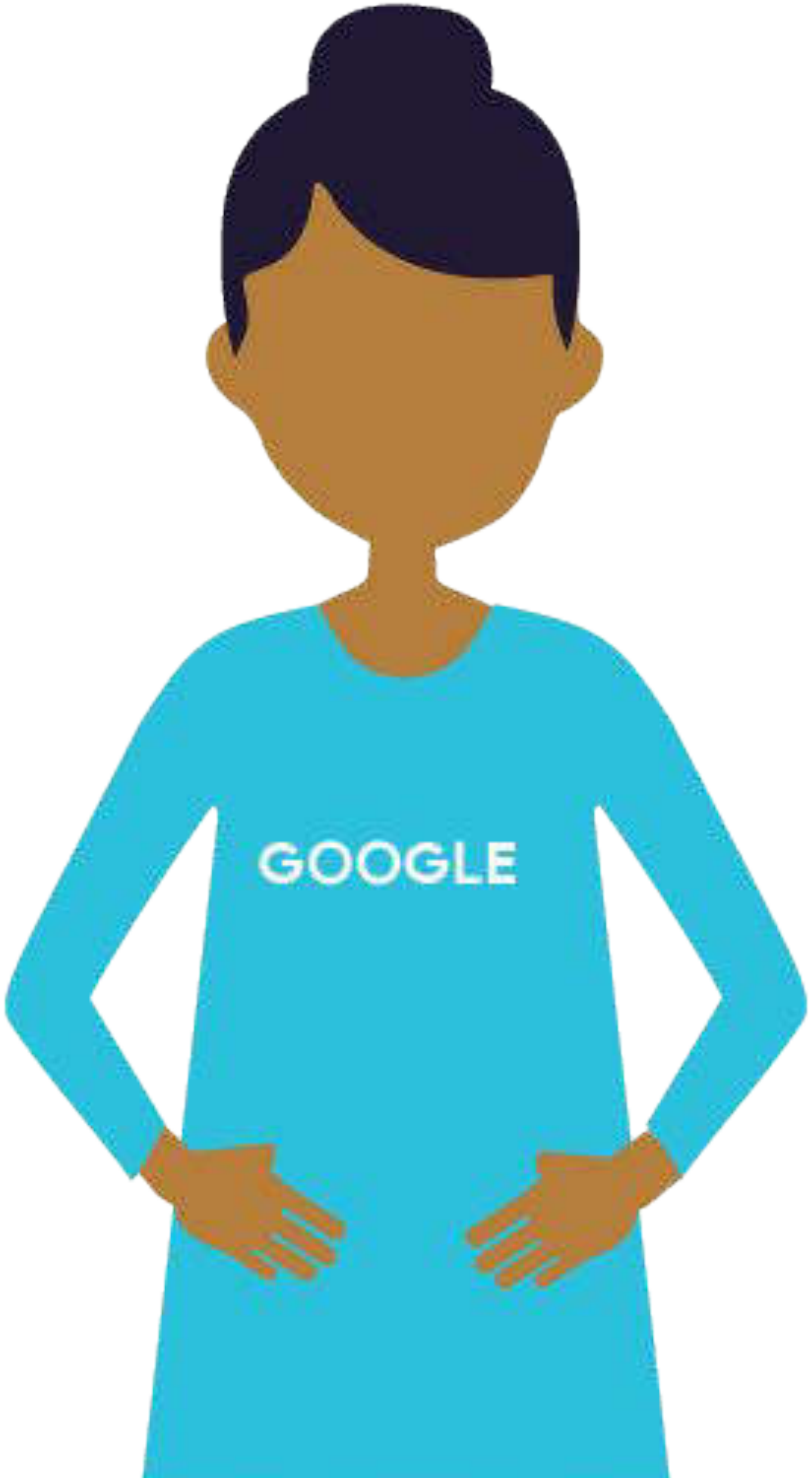
on Google Cloud Platform

Evan Jones

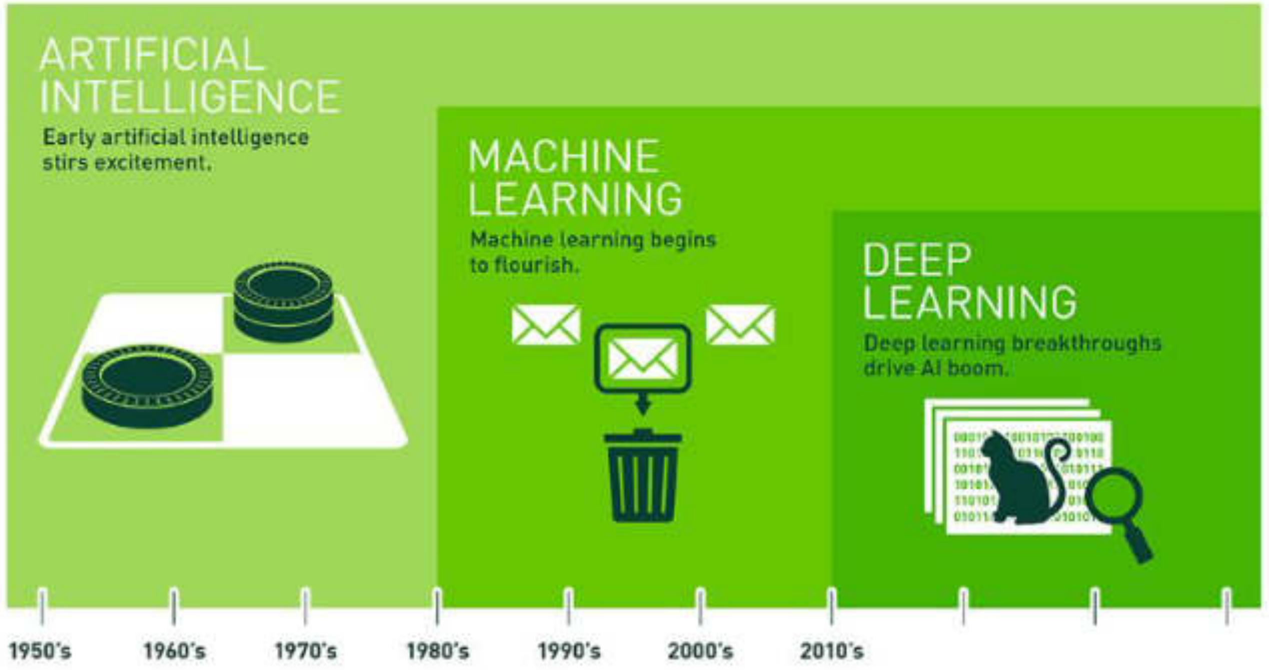


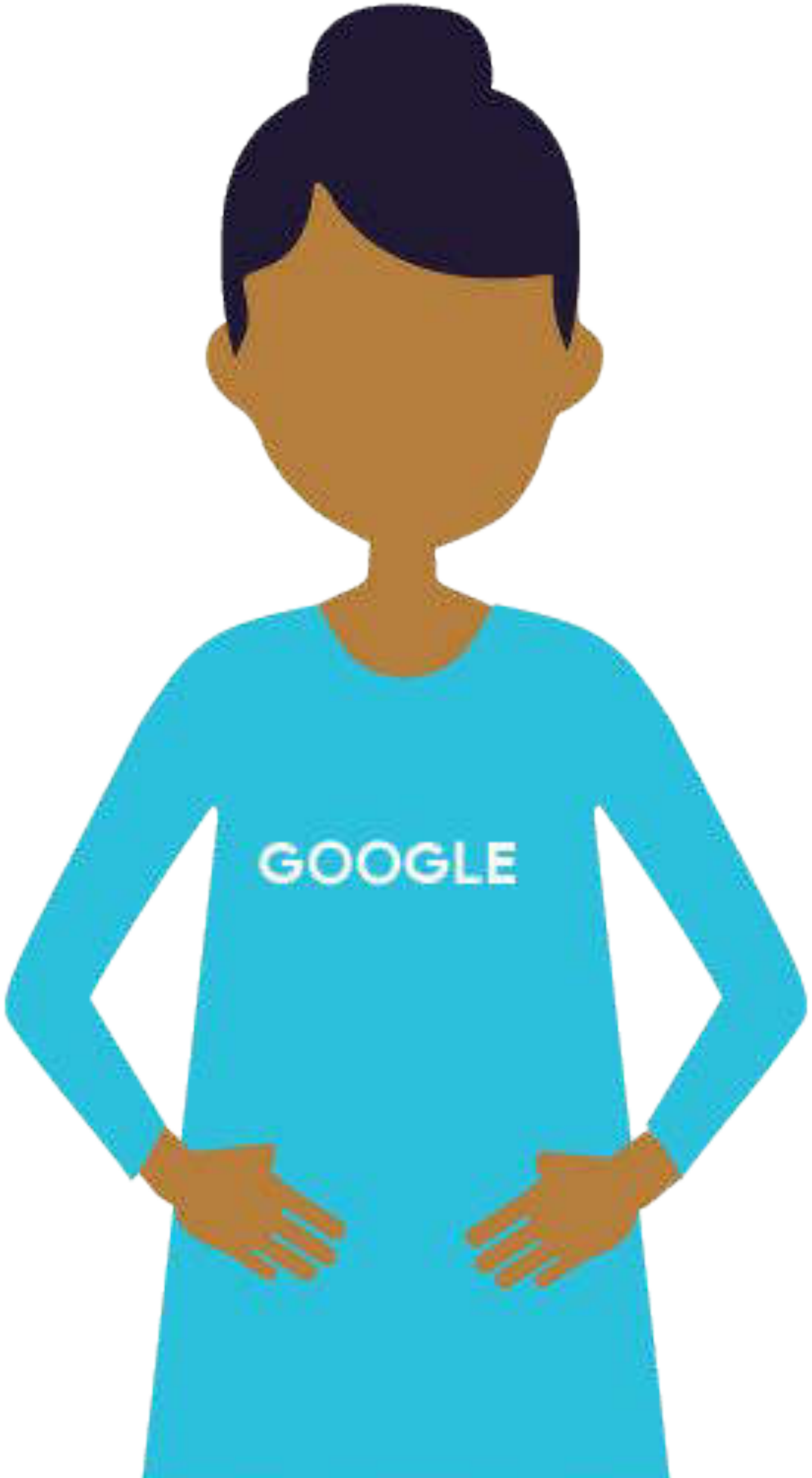
4 Courses in the Data to Insights Specialization

- ✓ 1 - Exploring and Preparing your Data with BigQuery
- ✓ 2 - Creating New BigQuery Datasets and Visualizing Insights
- ✓ 3 - Achieving Advanced Insights with BigQuery
- ✓ 4 - Applying Machine Learning to your Data with GCP

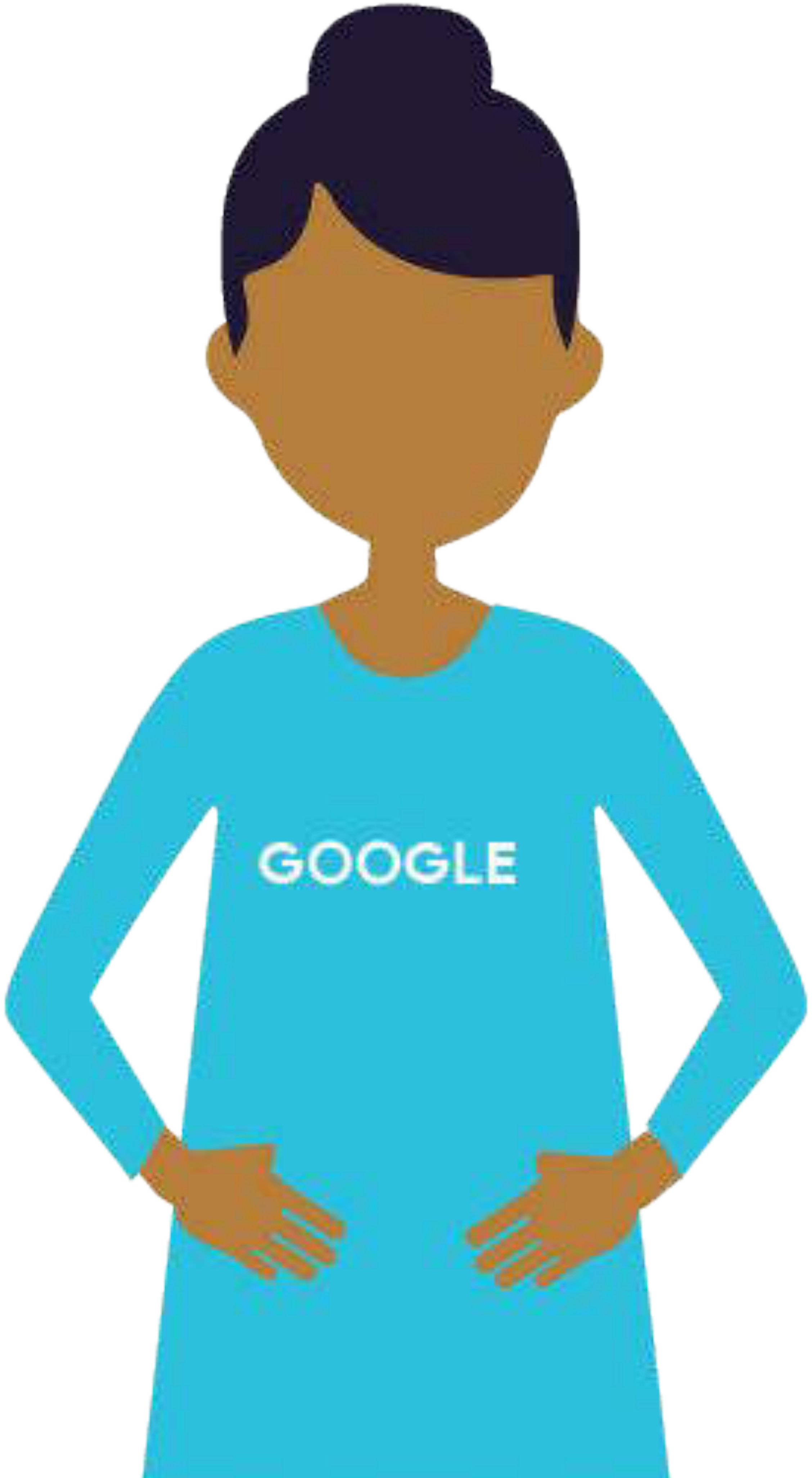

GOOGLE

ML, AI, and DL


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

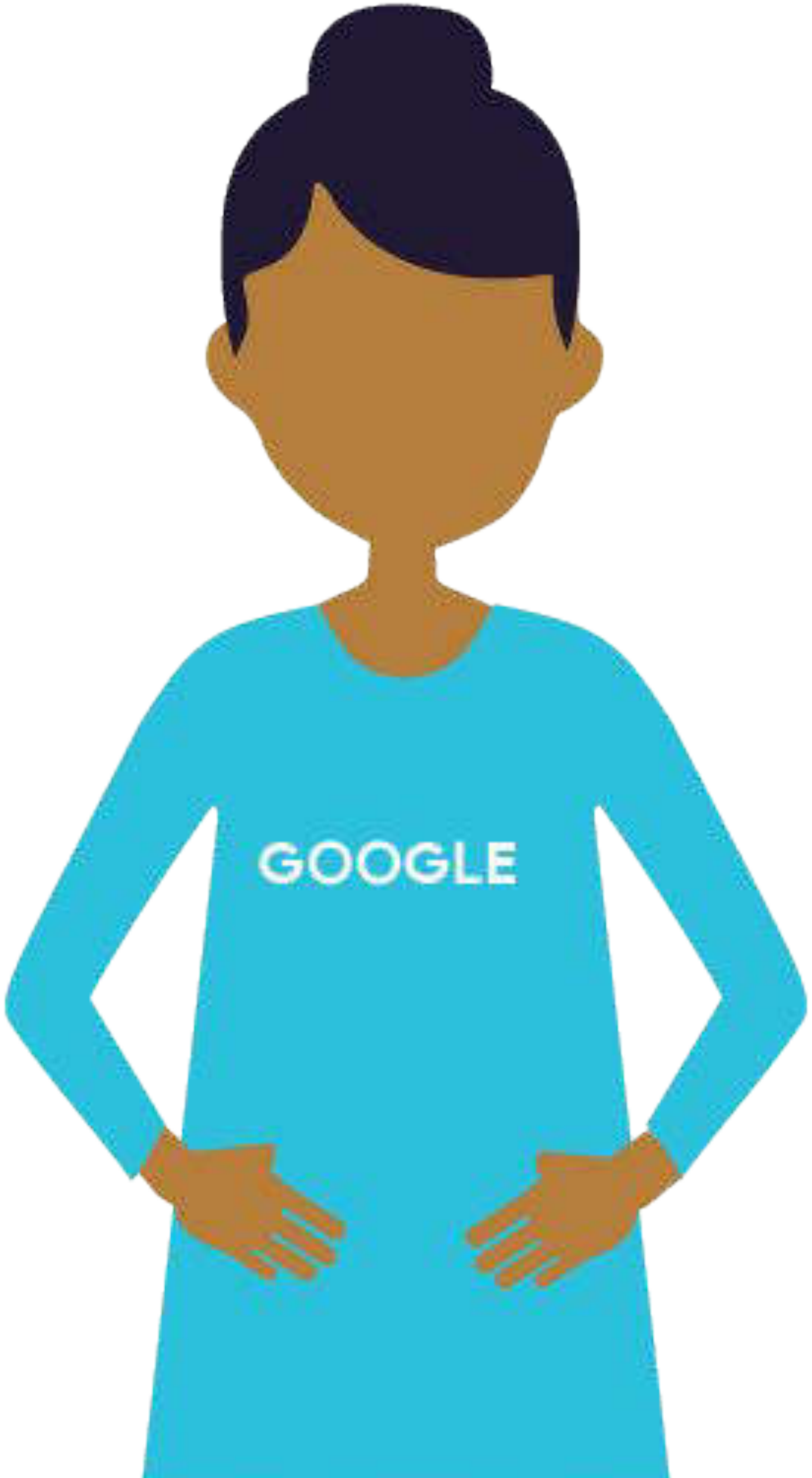


ML Applications for Businesses

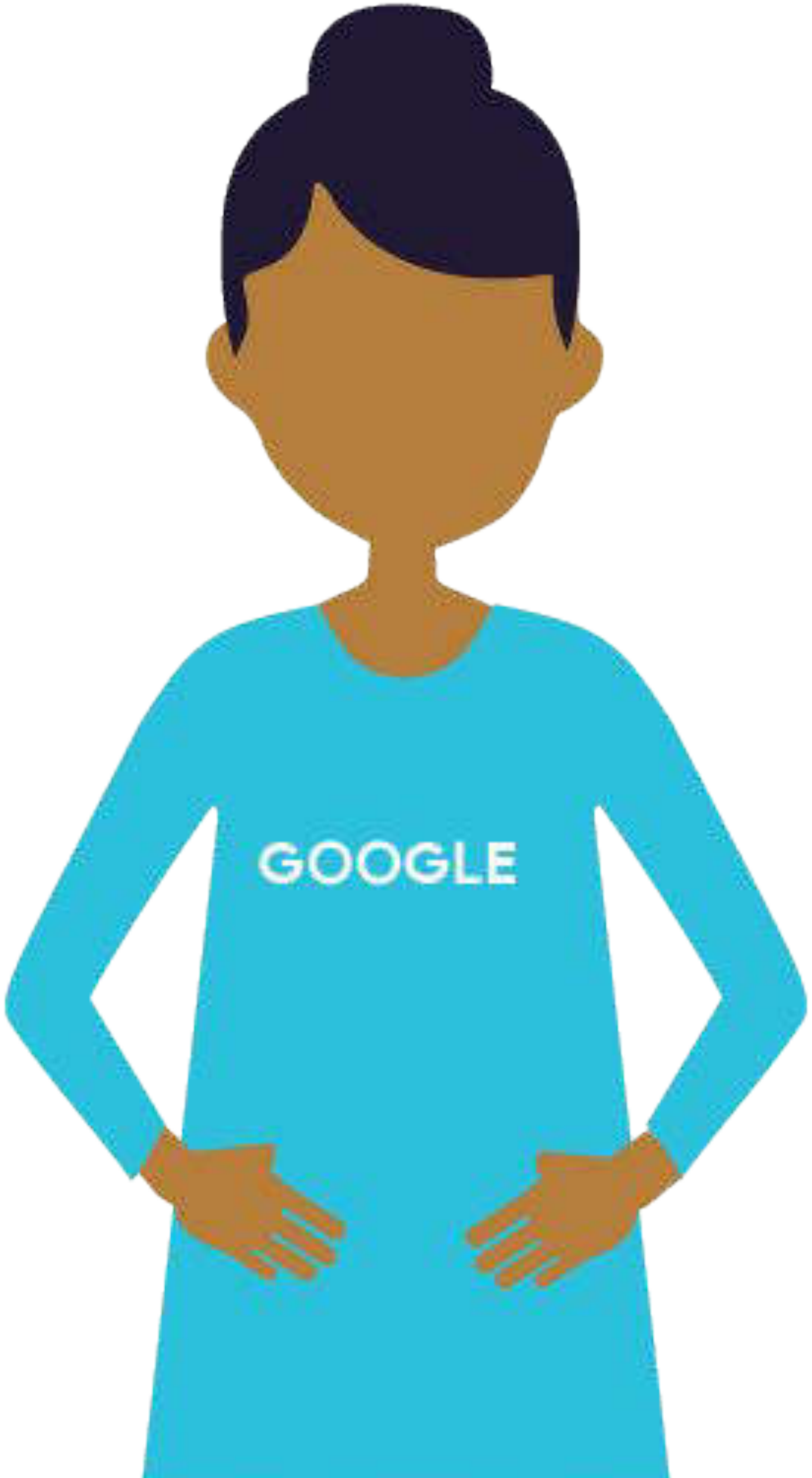


ML Terminology

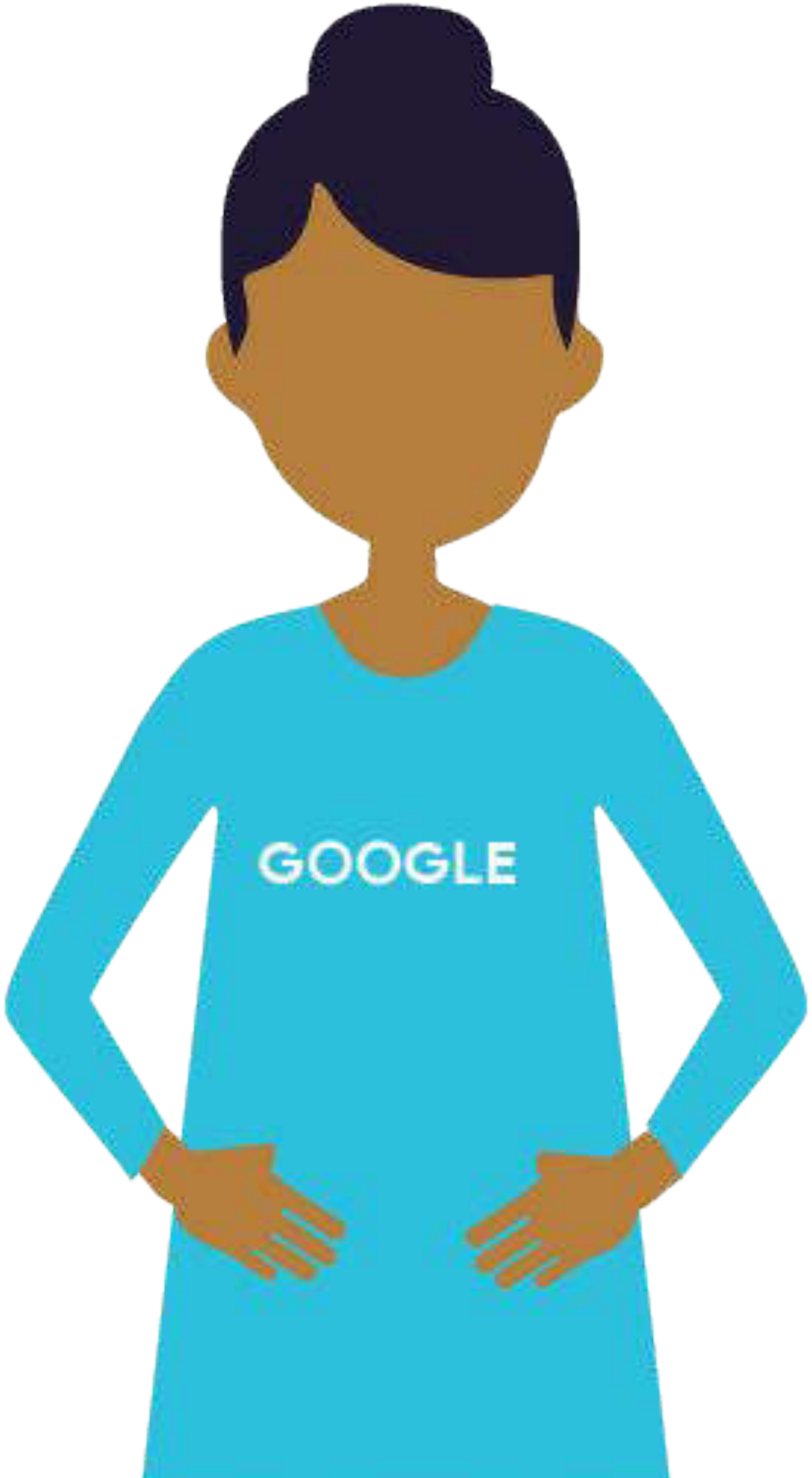
Instances, Labels, Features, and Models



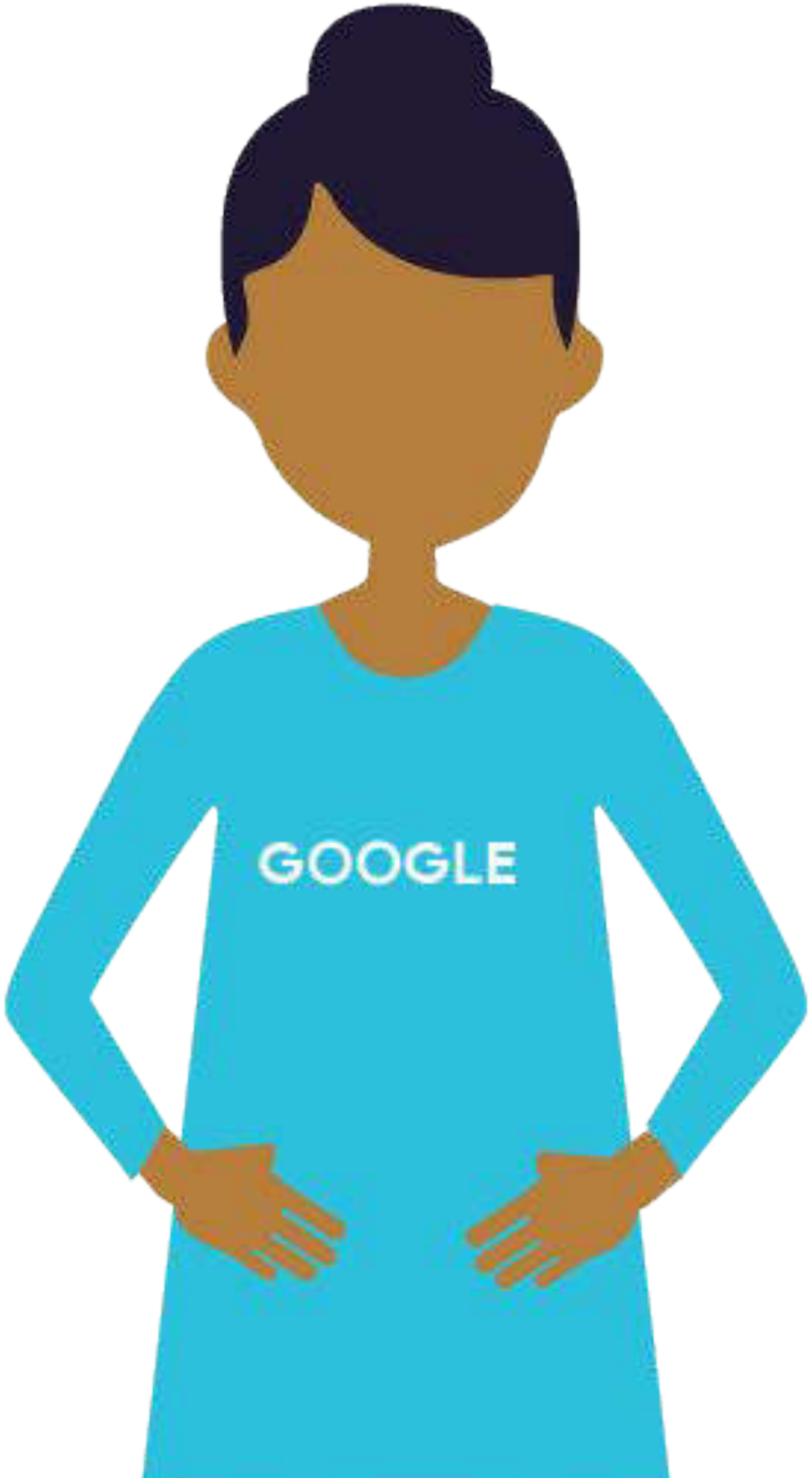
The 3 Secrets of ML



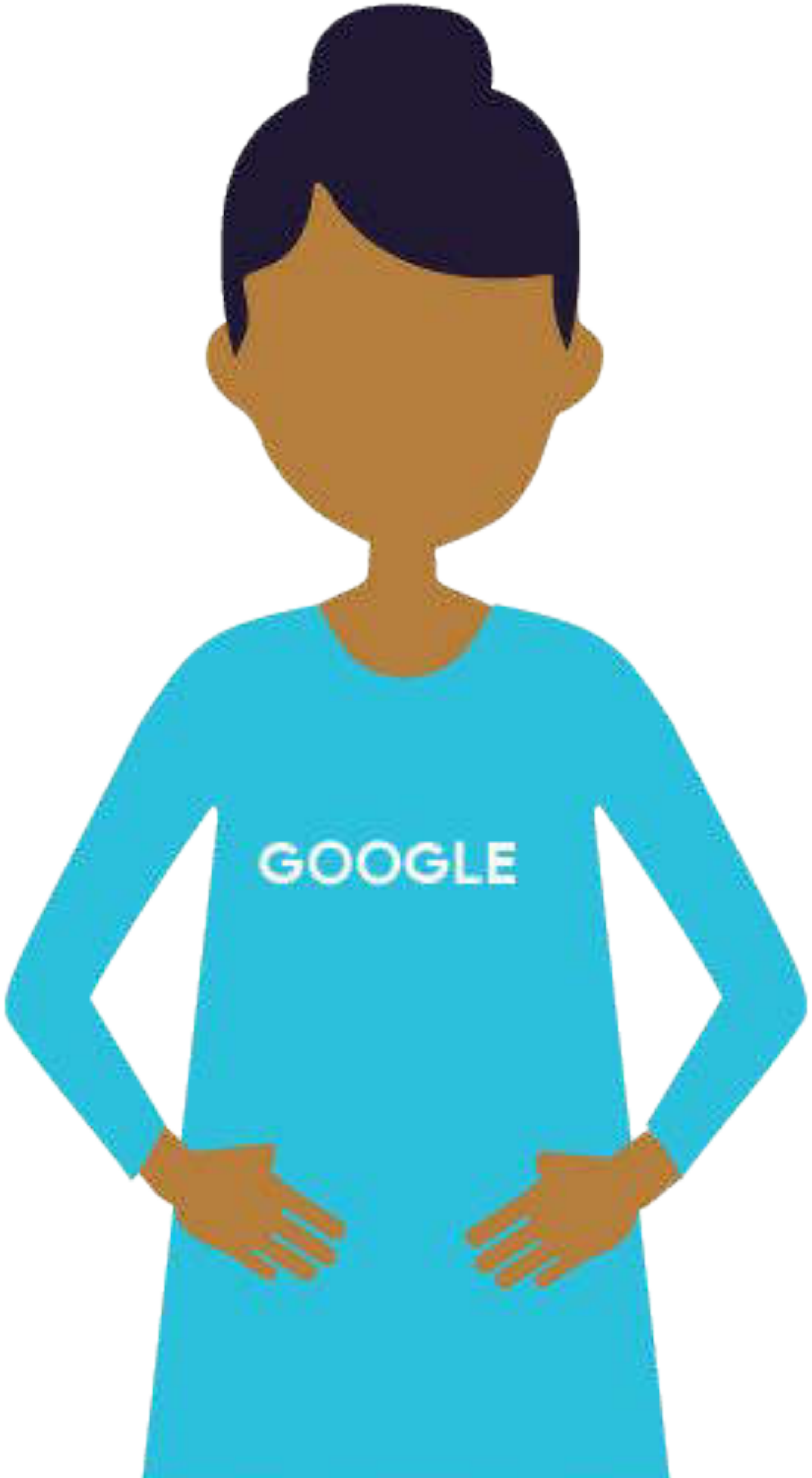
The ML Tool Spectrum for Data Analysts



Pre-trained ML APIs



Creating ML Datasets for in BigQuery



Creating ML Models inside of BigQuery

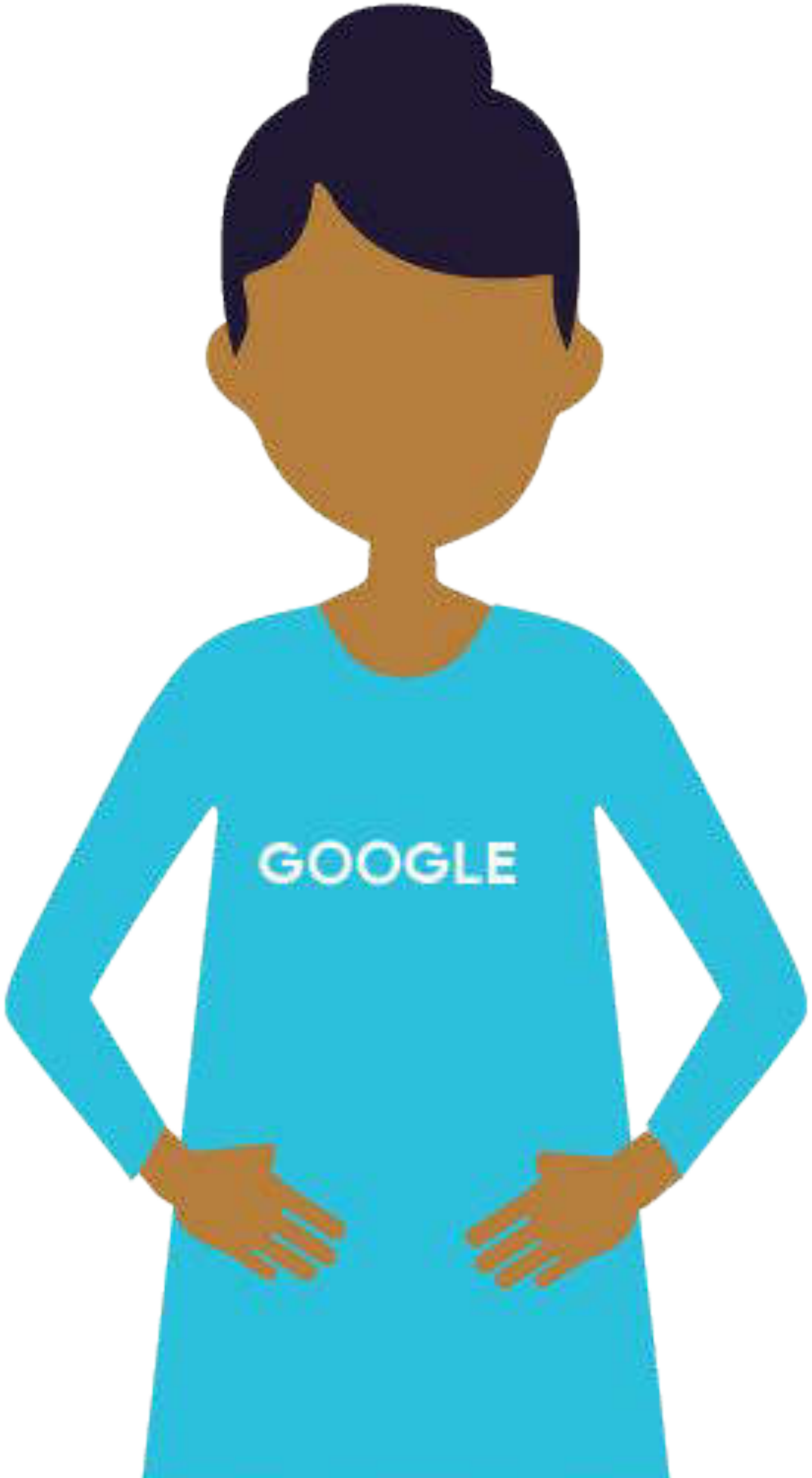
Course 4: Applying Machine Learning to your Datasets

Module 1: Introduction to Machine Learning

Lesson Title: **Introduction to Machine Learning**

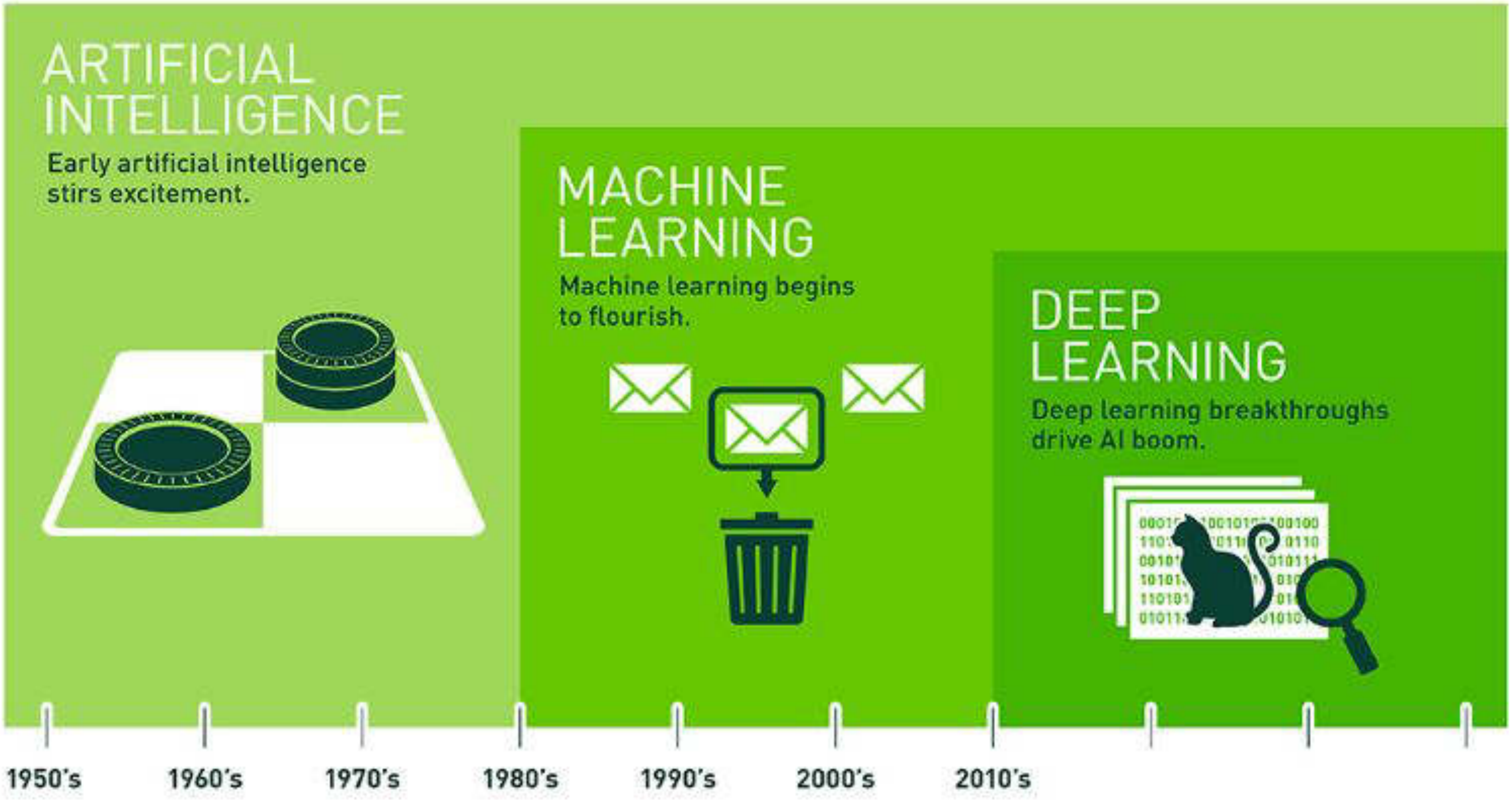
Format: Talking Head

Video Name: T-BQML-O_1_l1_introduction_to_machine_learning

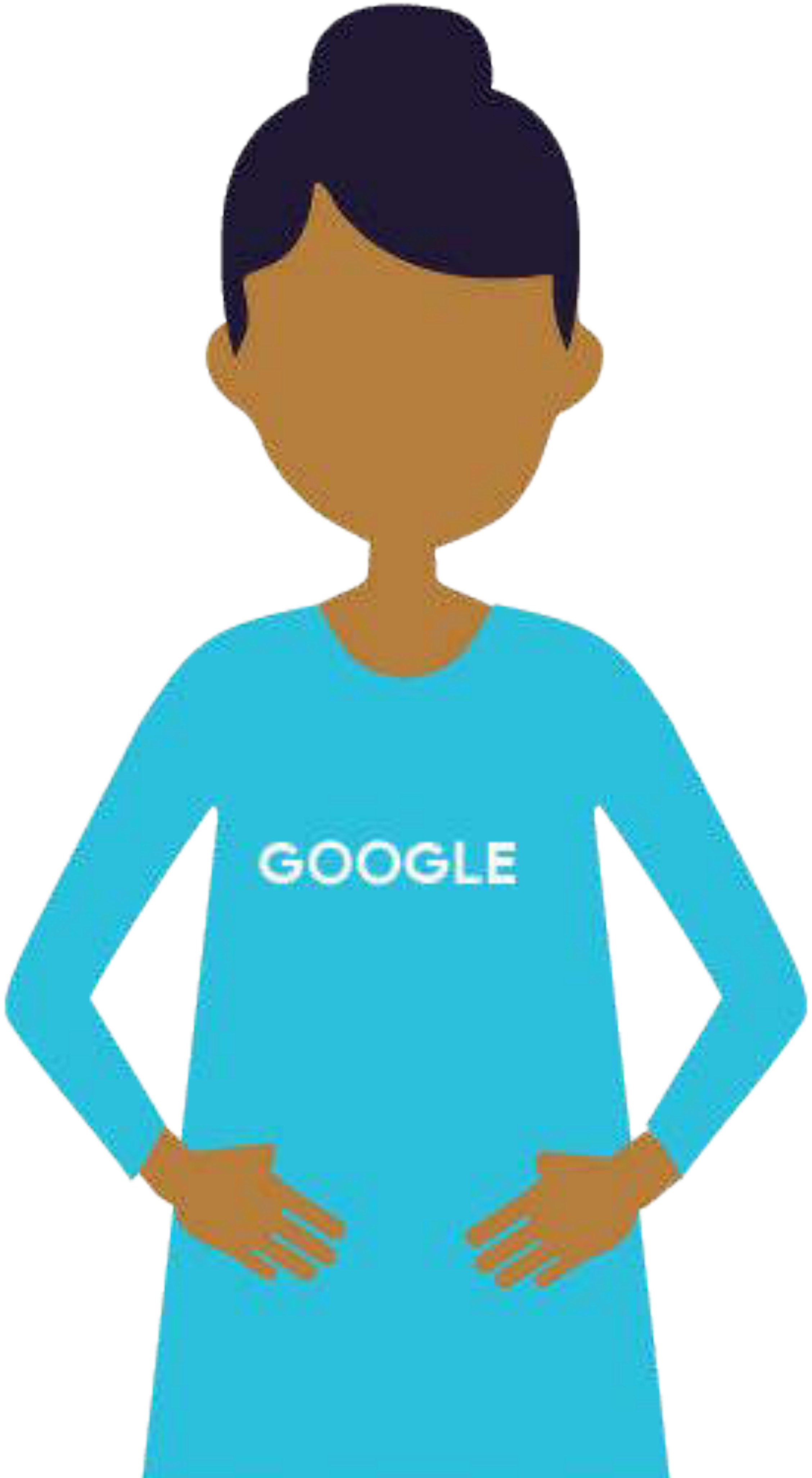


Machine Learning is a discipline inside of AI

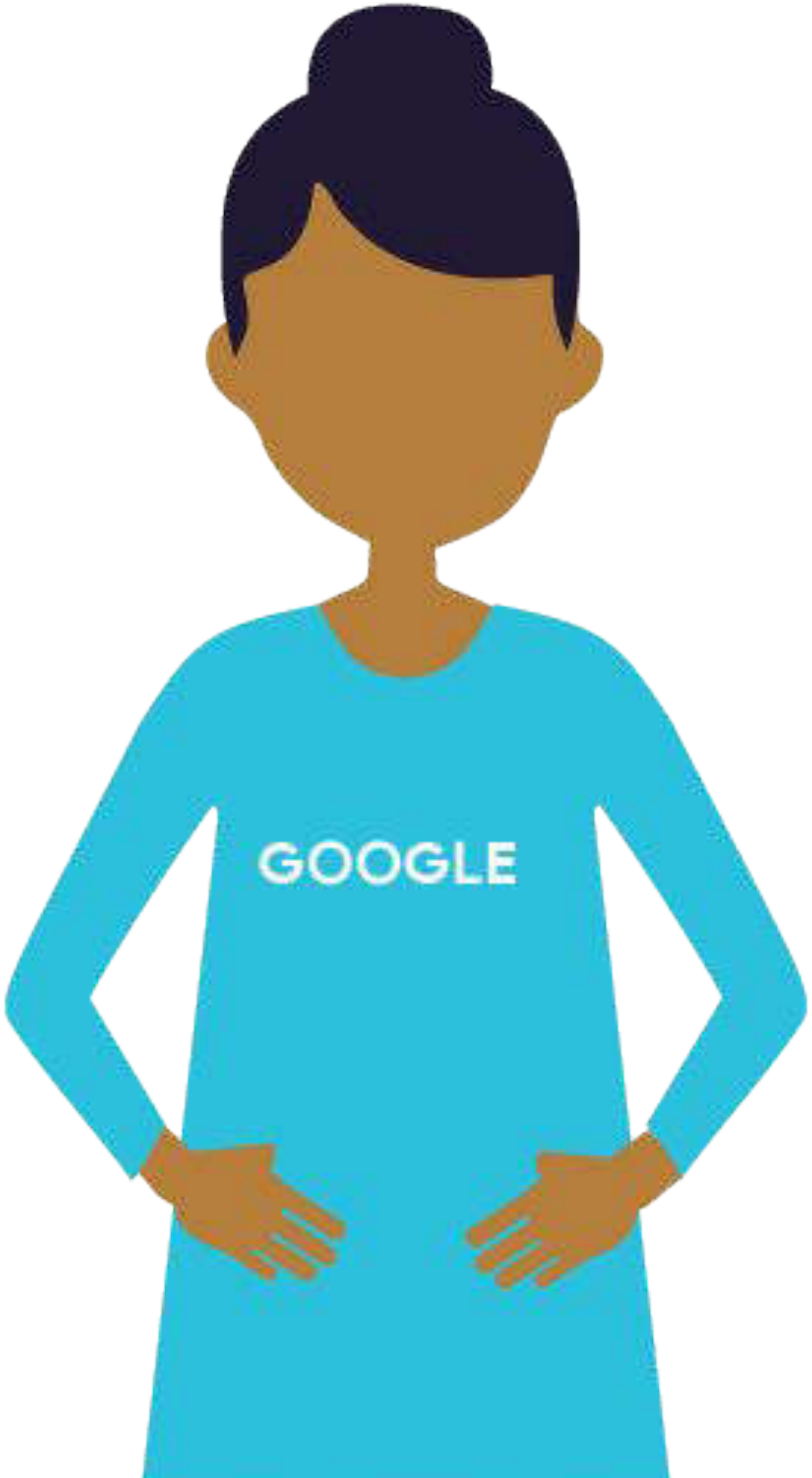
Machine Learning is a discipline inside of AI


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

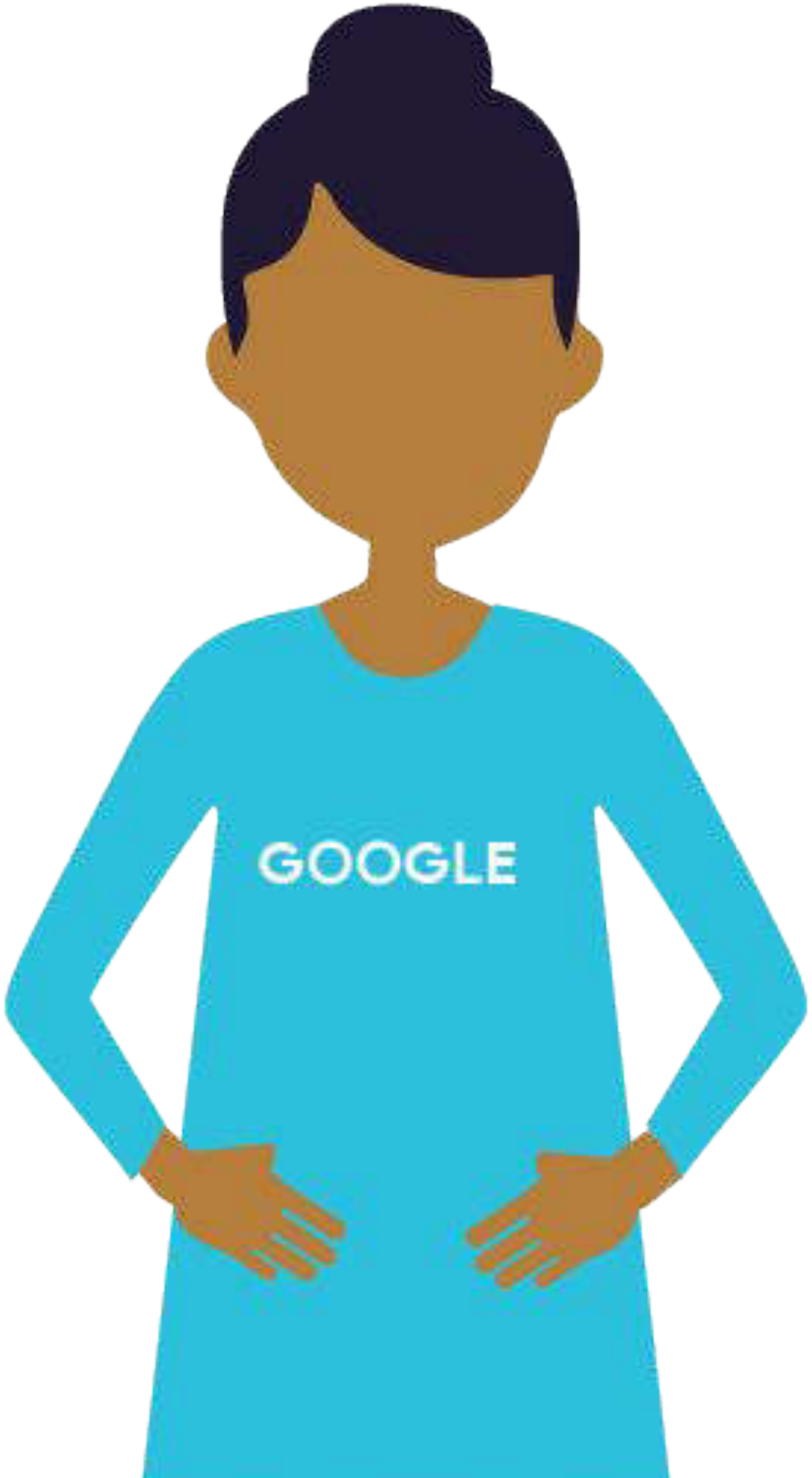
Source: Cassie Kozyrkov
https://becominghuman.ai/are-you-using-the-term-ai-incorrectly-911ac23ab4f5



Machine Learning labels things for you



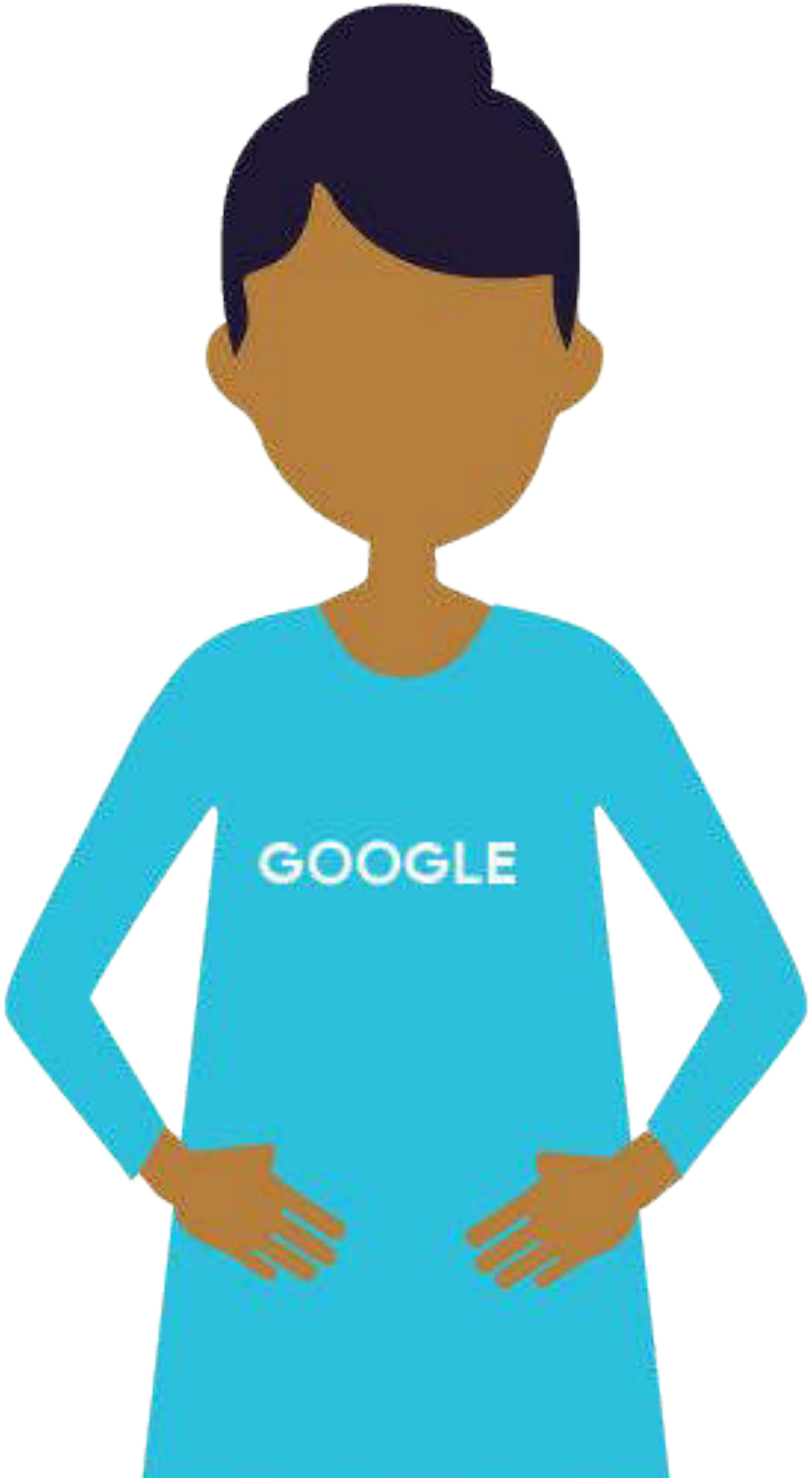
Science Fiction Movies



Science Fiction Movies

Movies I like:

- Set in space
- In the near future
- Shorter than 2 hours
- Not horror

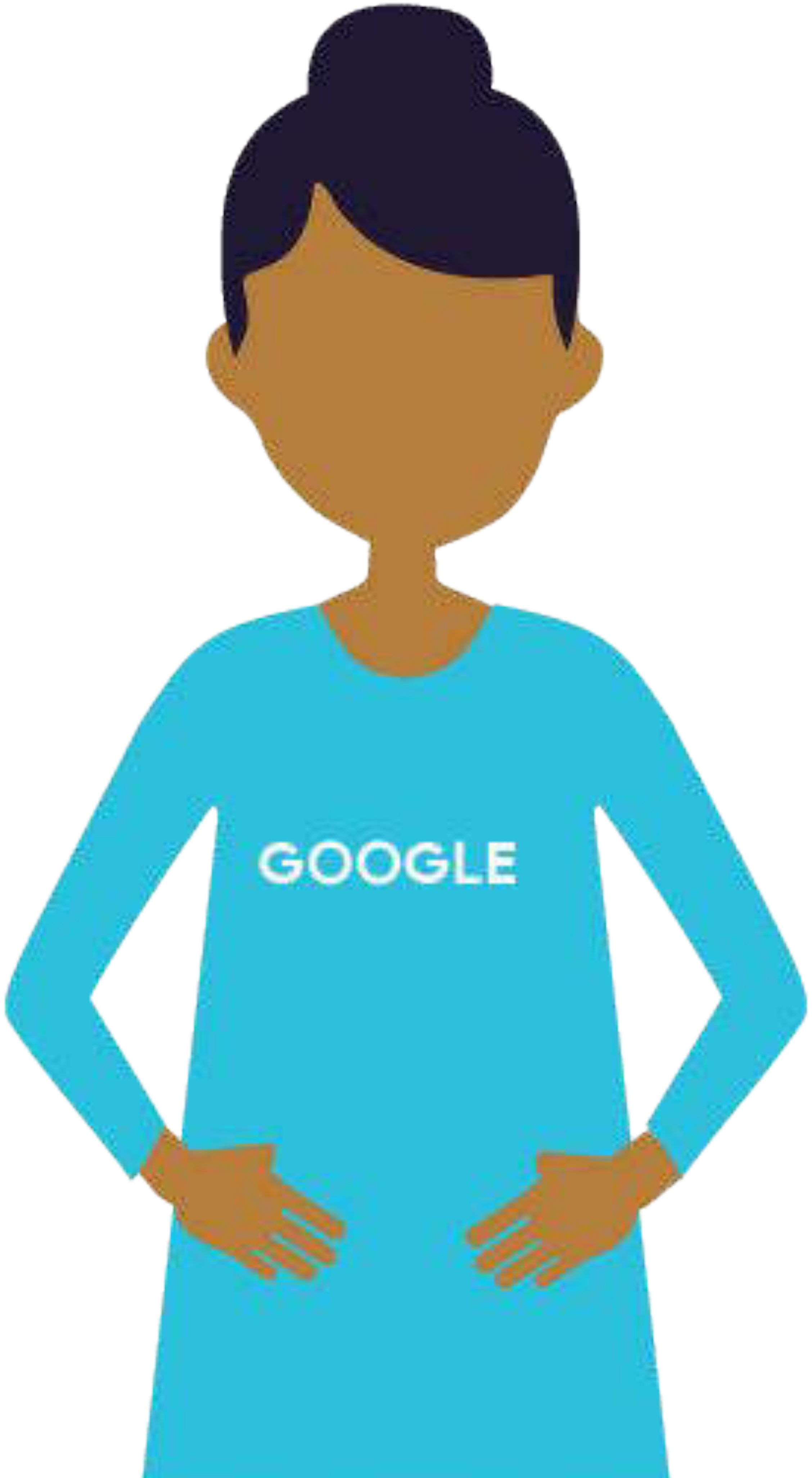


Science Fiction Movies

Movies I like:

- Set in space
- In the near future
- Shorter than 2 hours
- Not horror

List of Past Sci-Fi Movies I Like

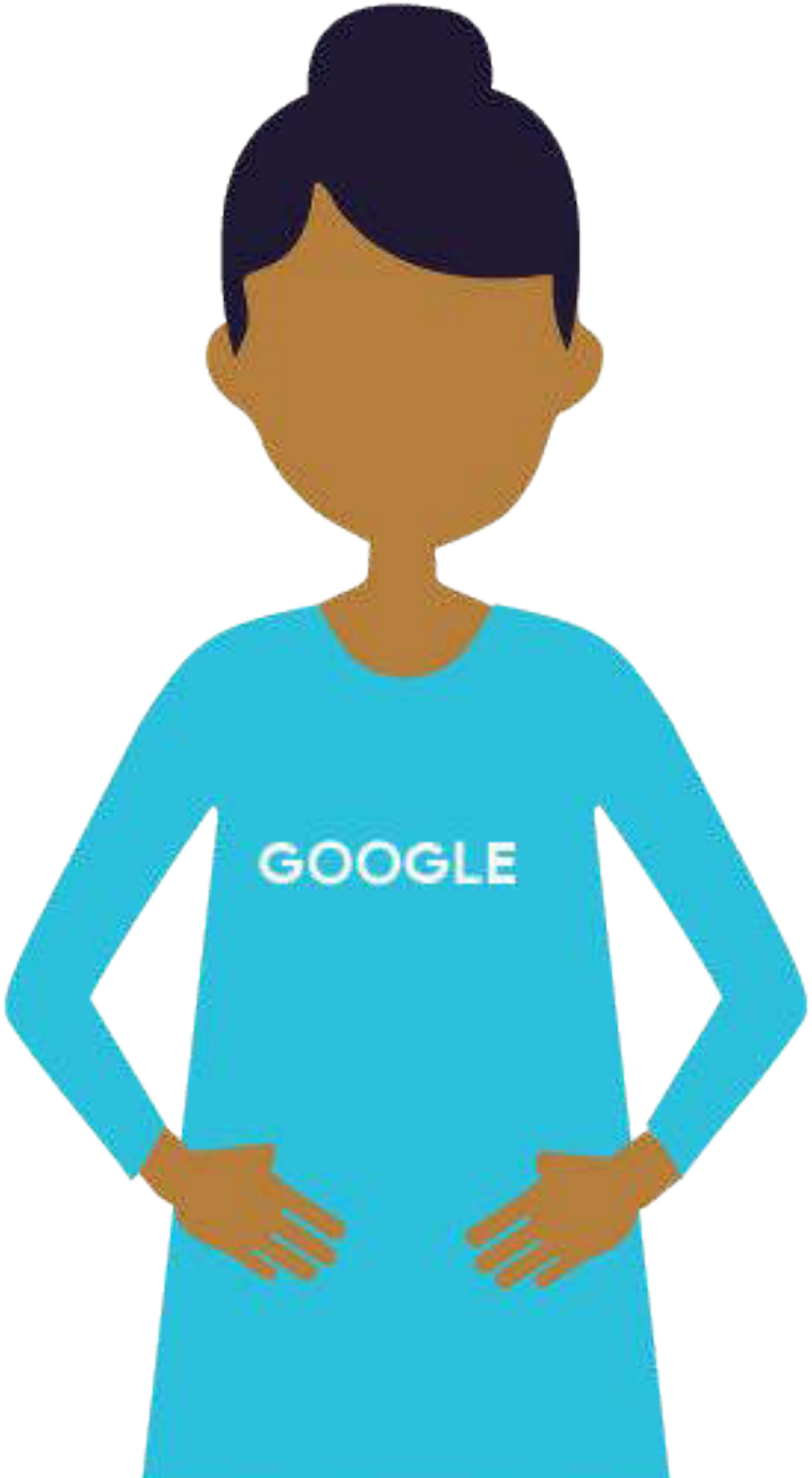


Train a movie recommendation model

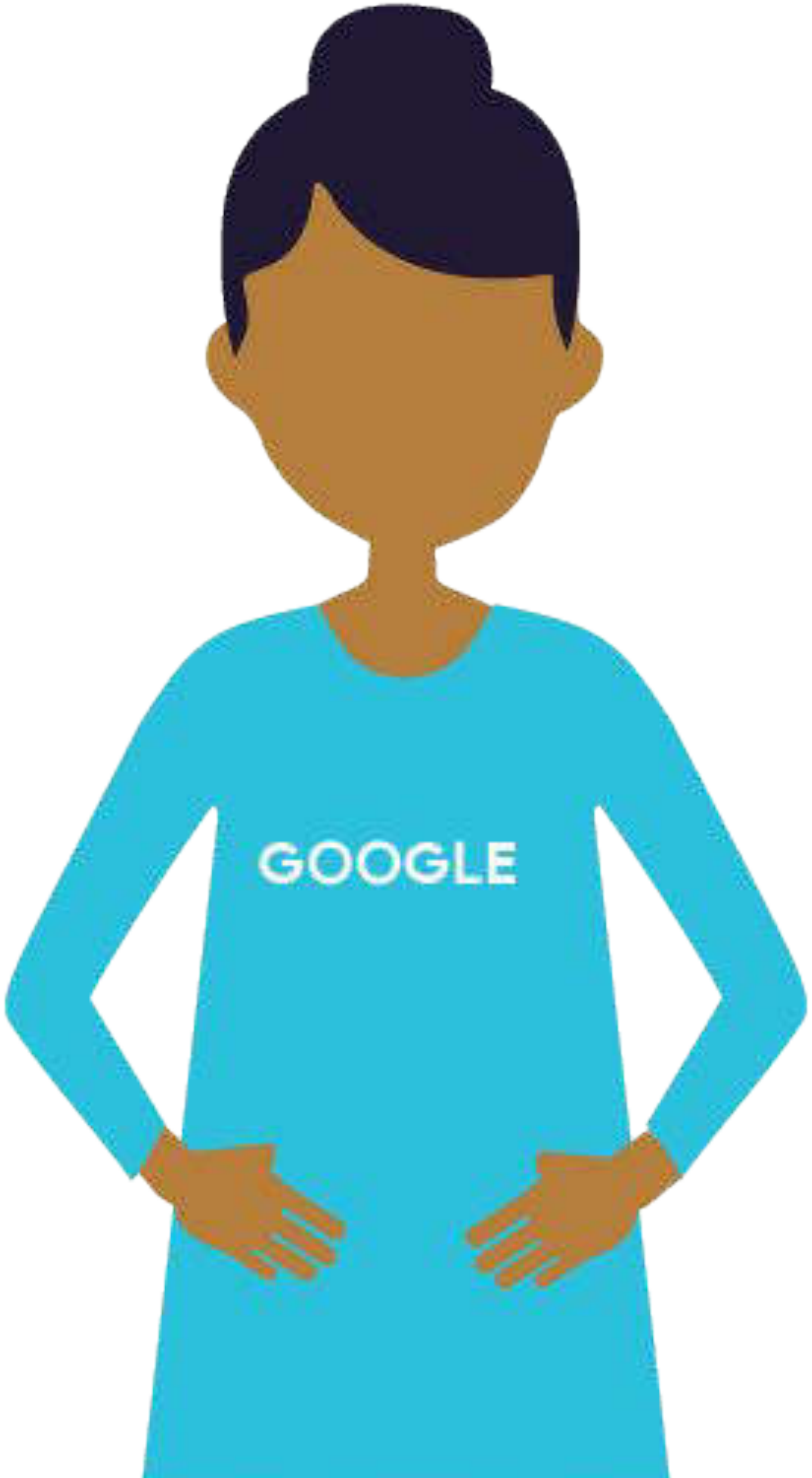
Provide model with rules? No.

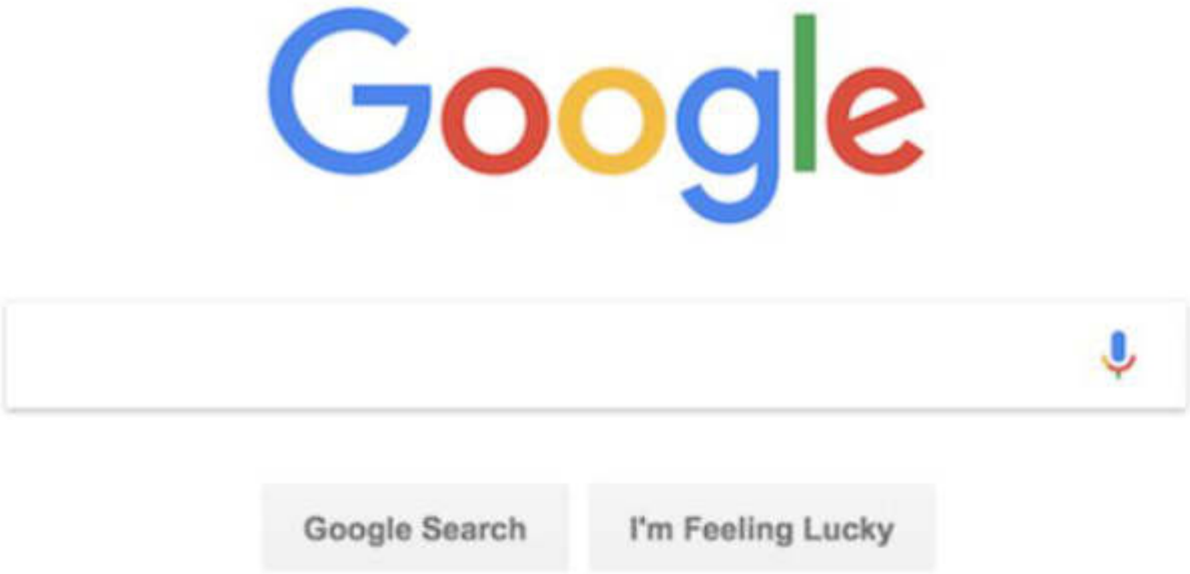
✗
- IF Set in space THEN
- IF In the near future THEN
- IF Shorter than 2 hours THEN
- IF Not horror THEN

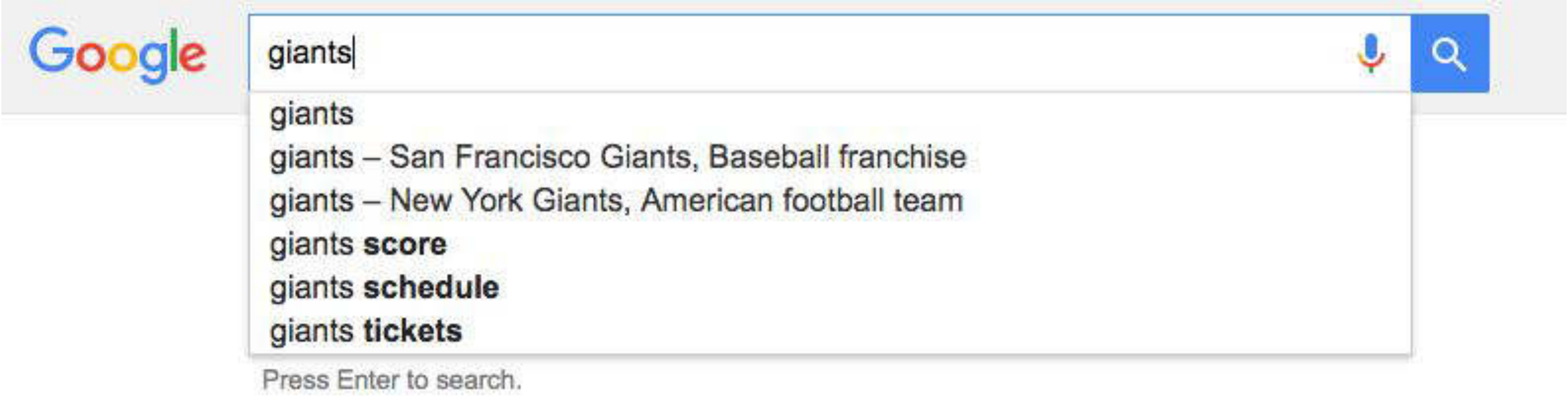
✓ List of Past Sci-Fi Movies I Like



ML enables scale


GOOGLE

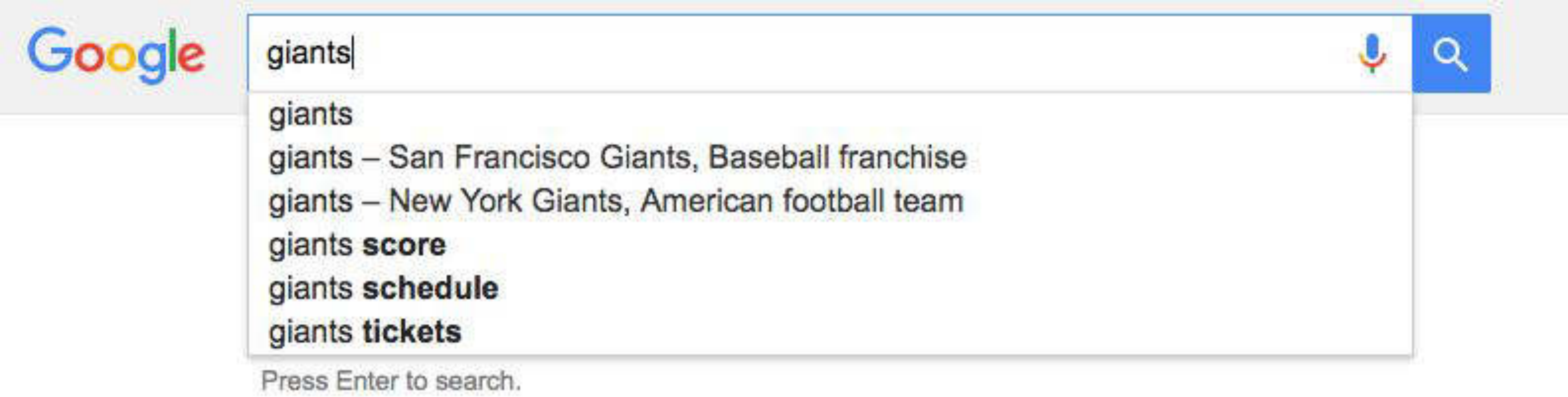

Google
Google Search
I'm Feeling Lucky



giants

giants
giants – San Francisco Giants, Baseball franchise
giants – New York Giants, American football team
giants **score**
giants **schedule**
giants **tickets**

Press Enter to search.

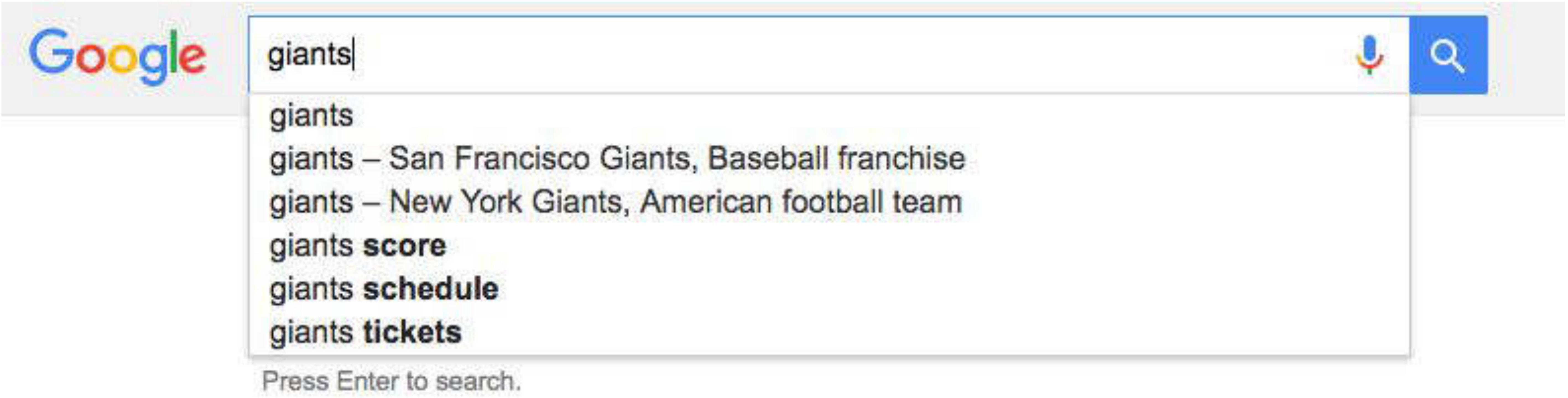


giants

giants
giants – San Francisco Giants, Baseball franchise
giants – New York Giants, American football team
giants **score**
giants **schedule**
giants **tickets**

Press Enter to search.

CALIFORNIA

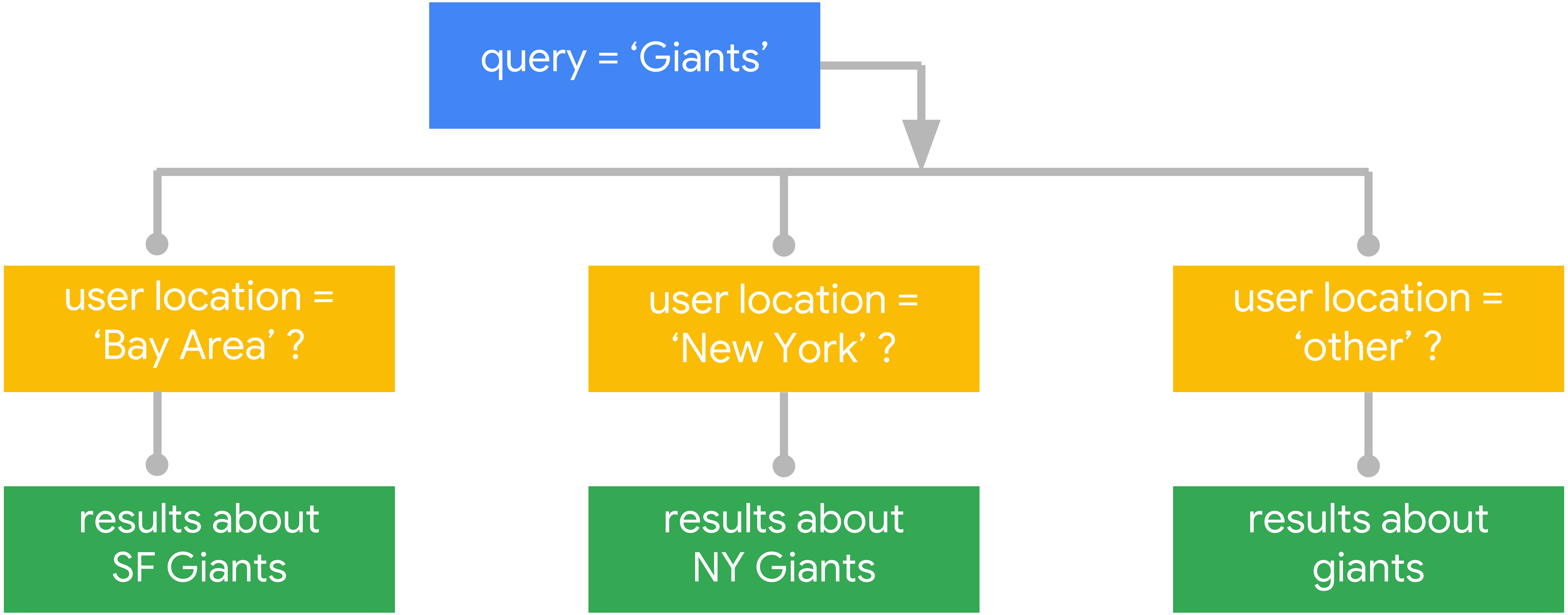


giants

giants
giants – San Francisco Giants, Baseball franchise
giants – New York Giants, American football team
giants **score**
giants **schedule**
giants **tickets**

Press Enter to search.

CALIFORNIA



query = ‘Giants’

user location = ‘Bay Area’ ?

user location = ‘New York’ ?

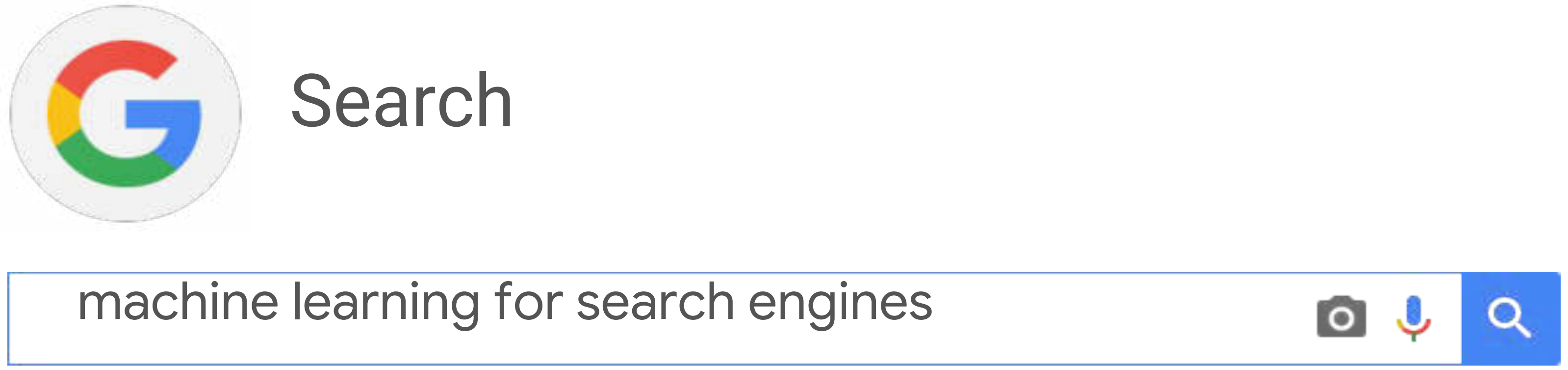
user location = ‘other’ ?

results about SF Giants

results about NY Giants

results about giants

RankBrain (ML for search ranking) improved performance significantly



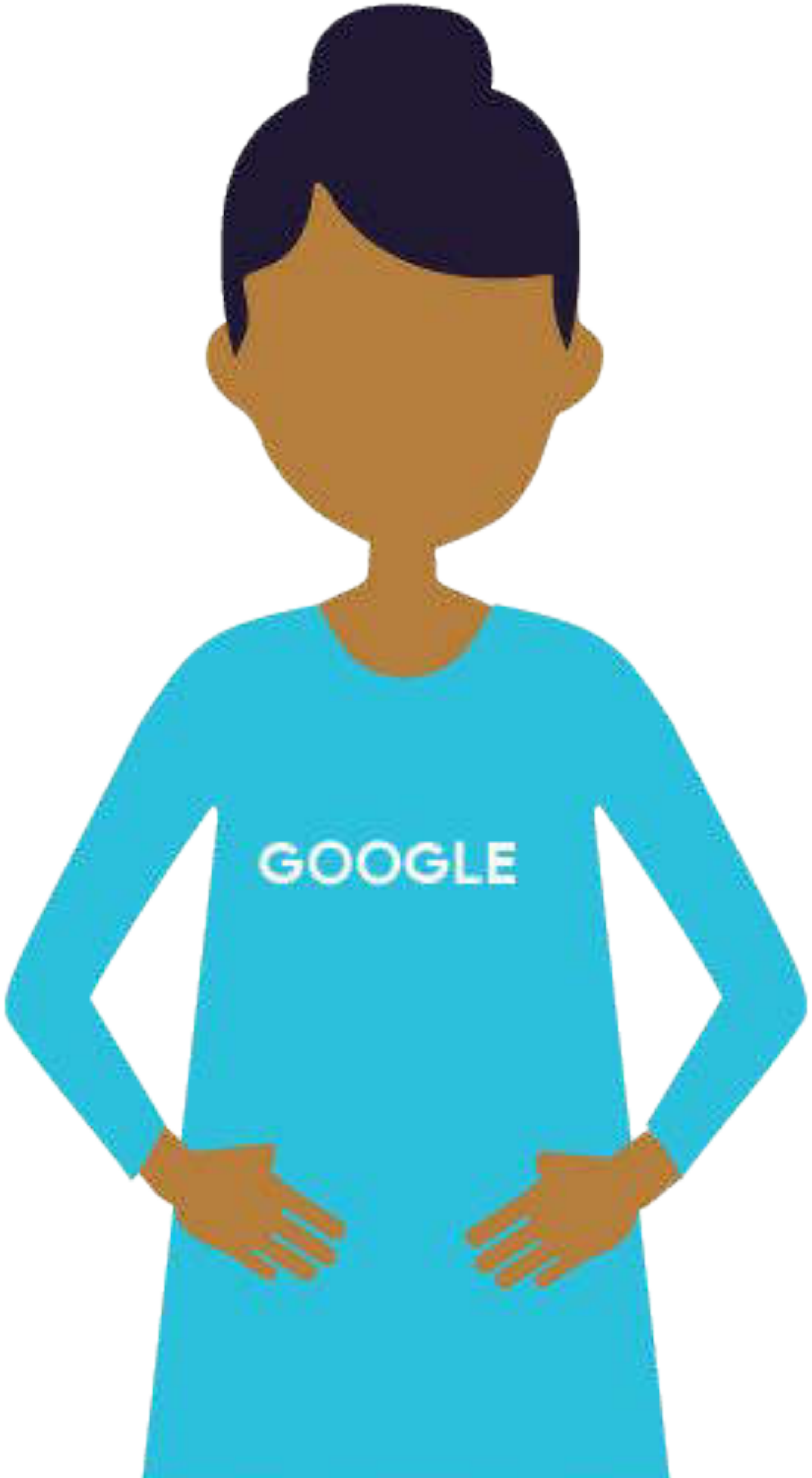
machine learning for search engines

#3

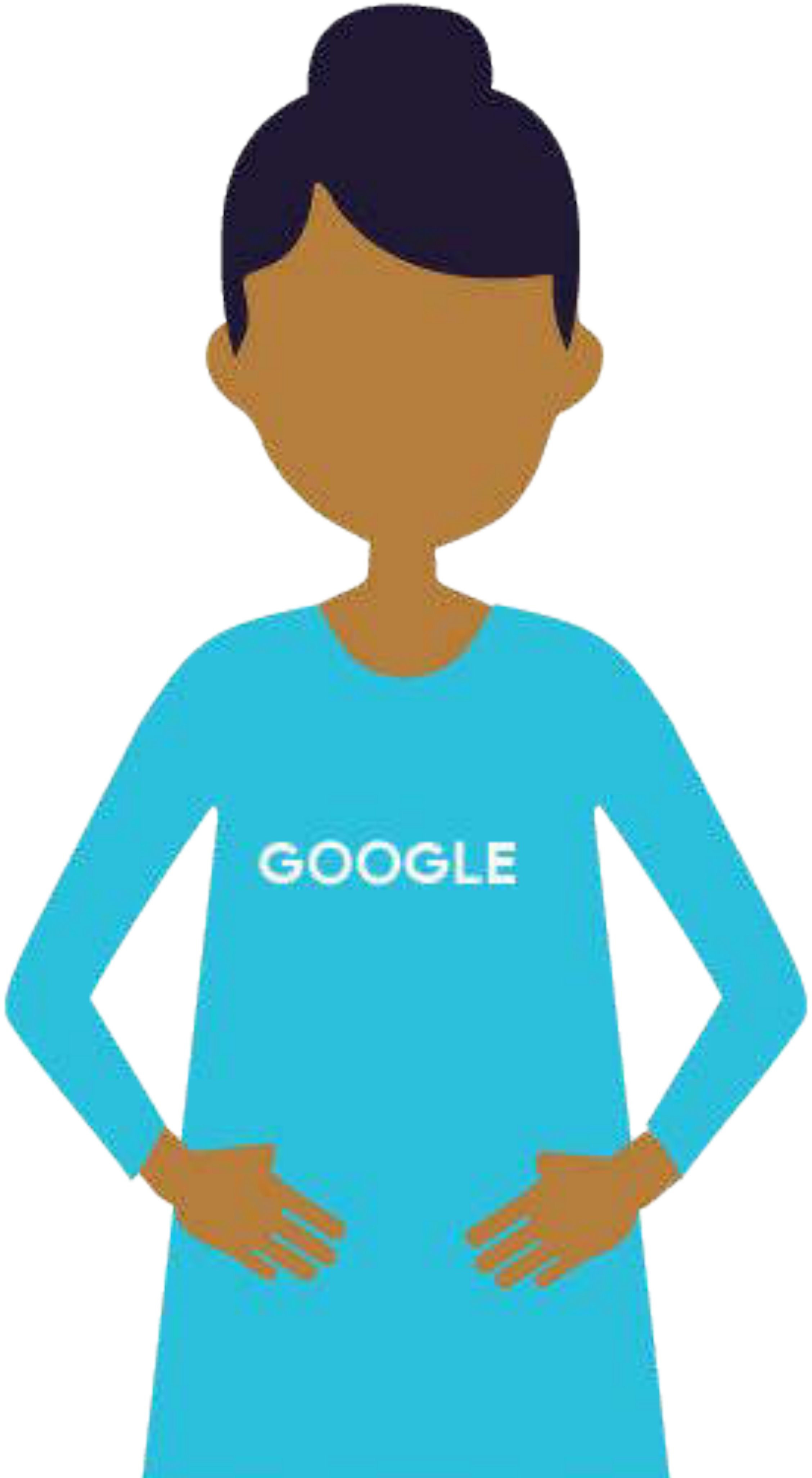
signal
for Search ranking, out of hundreds

#1

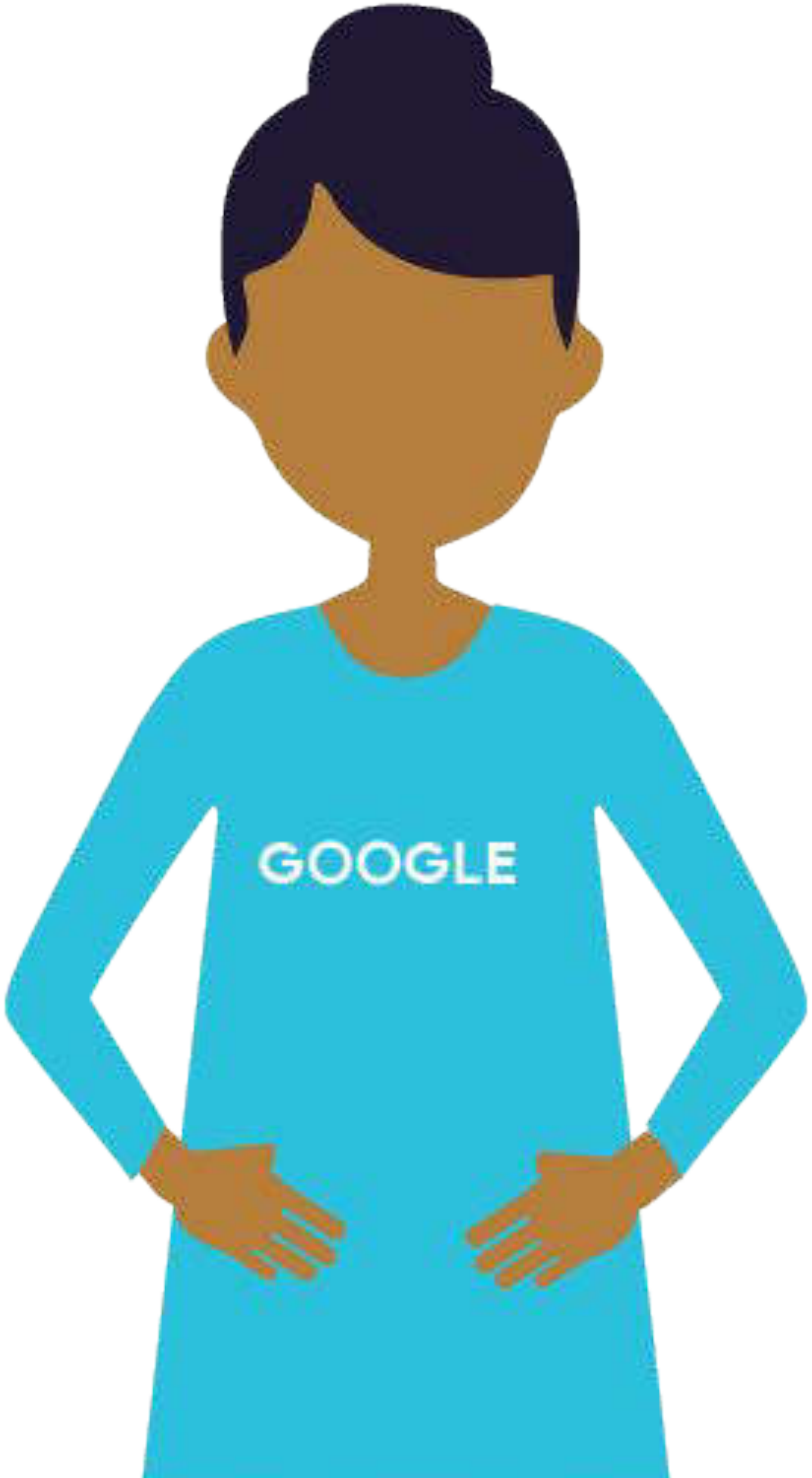
improvement
to ranking quality in 2+ years

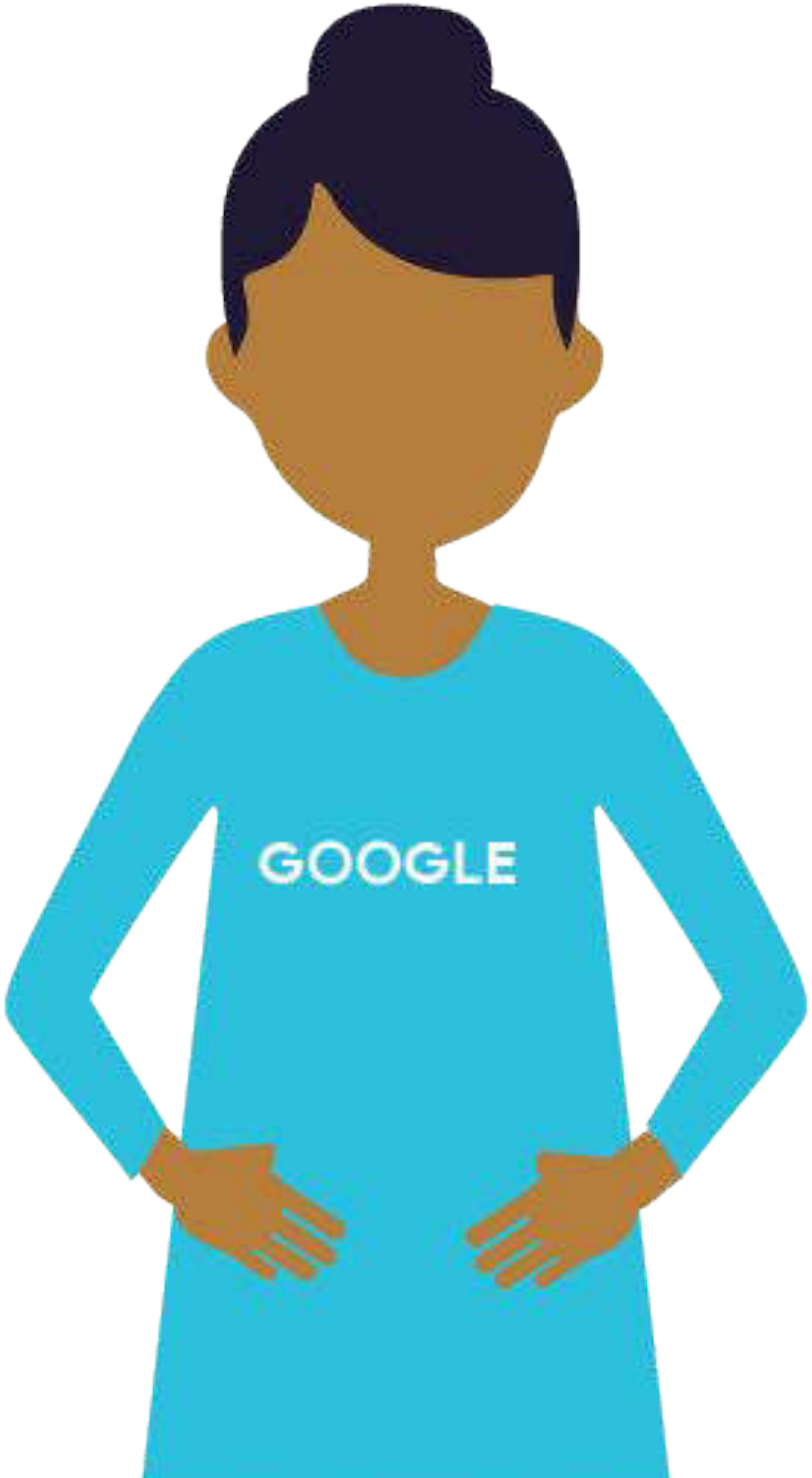


Machine Learning =
Lead with examples,
not instructions

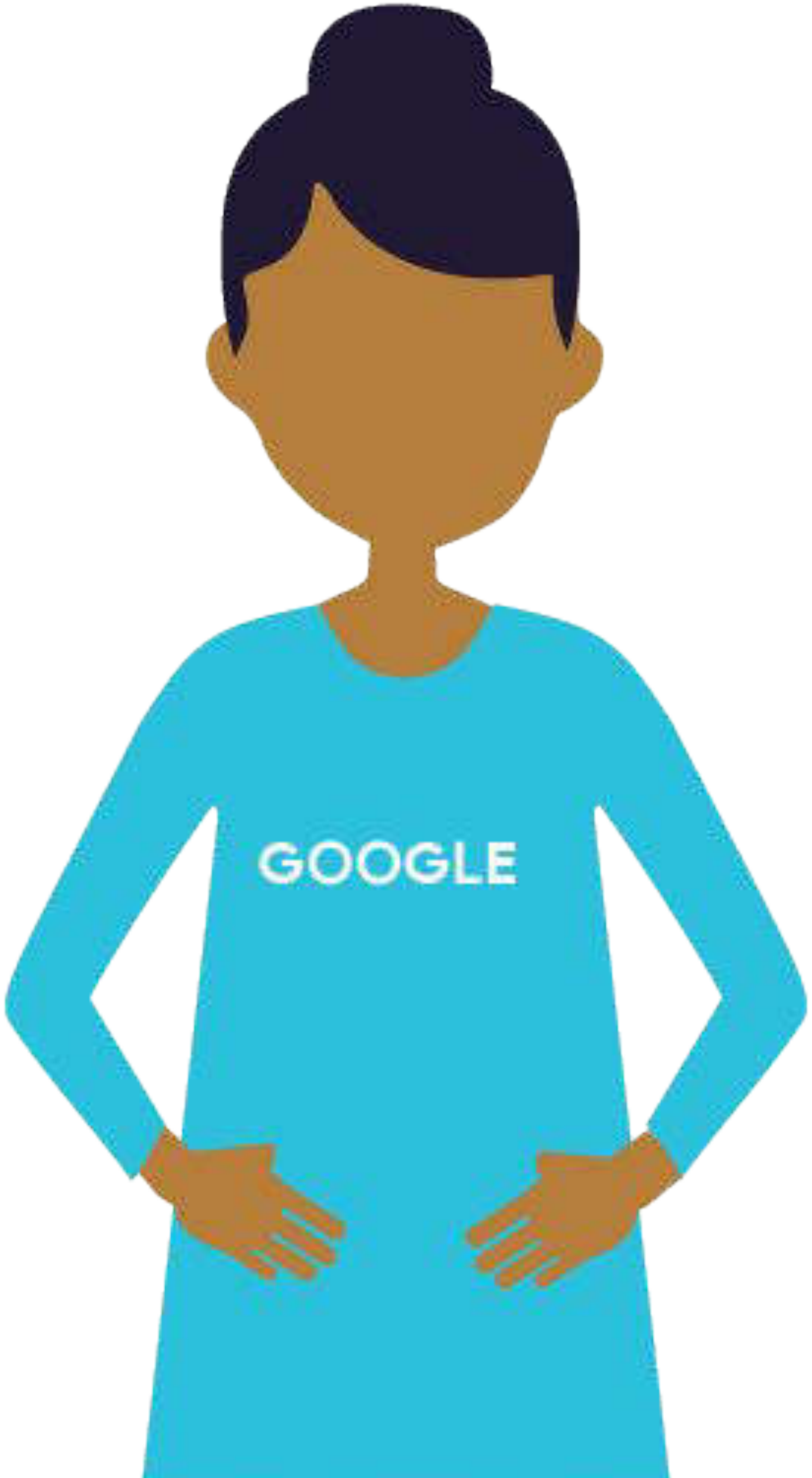


Use Deep Learning when you can't explain the labeling rules

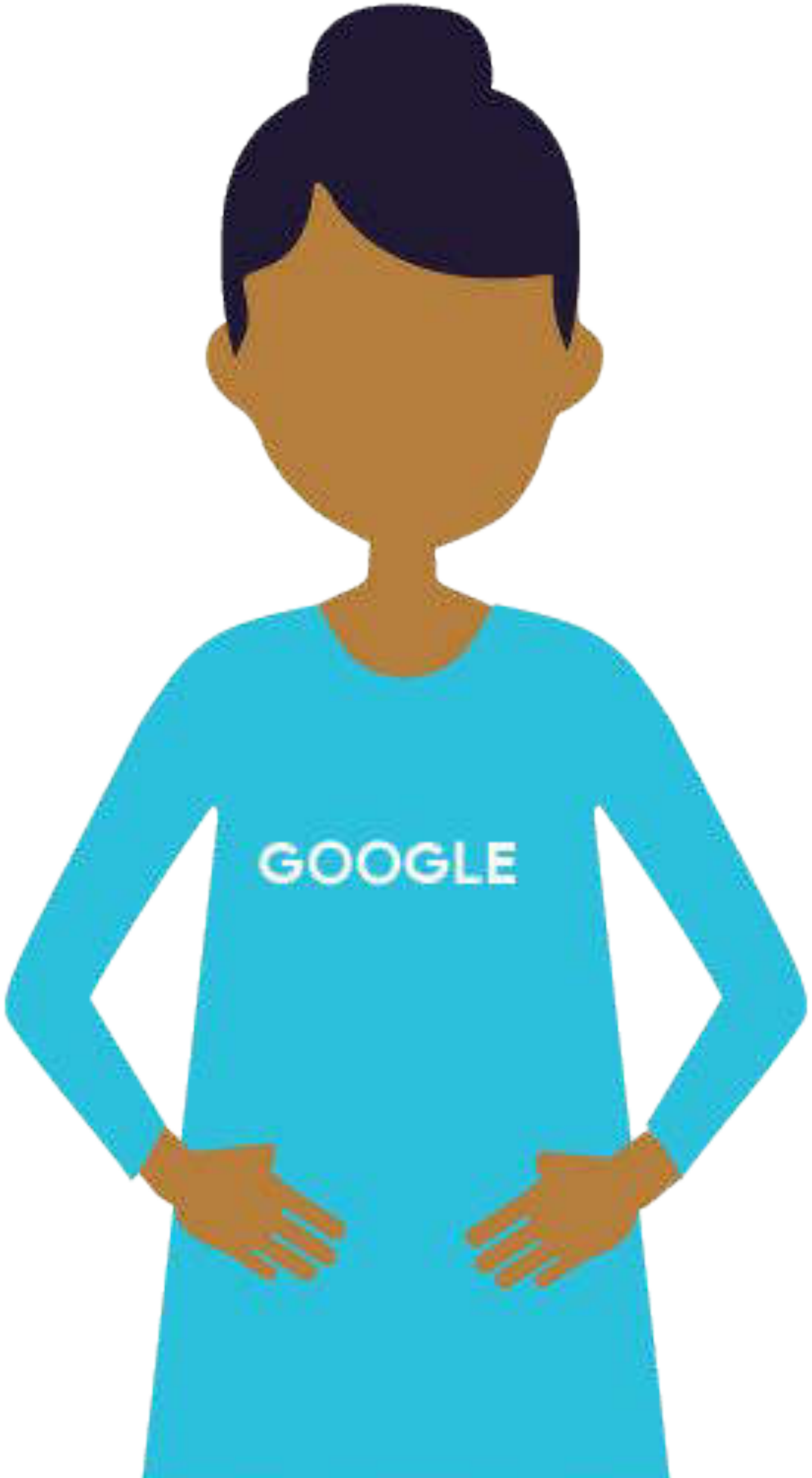

GOOGLE



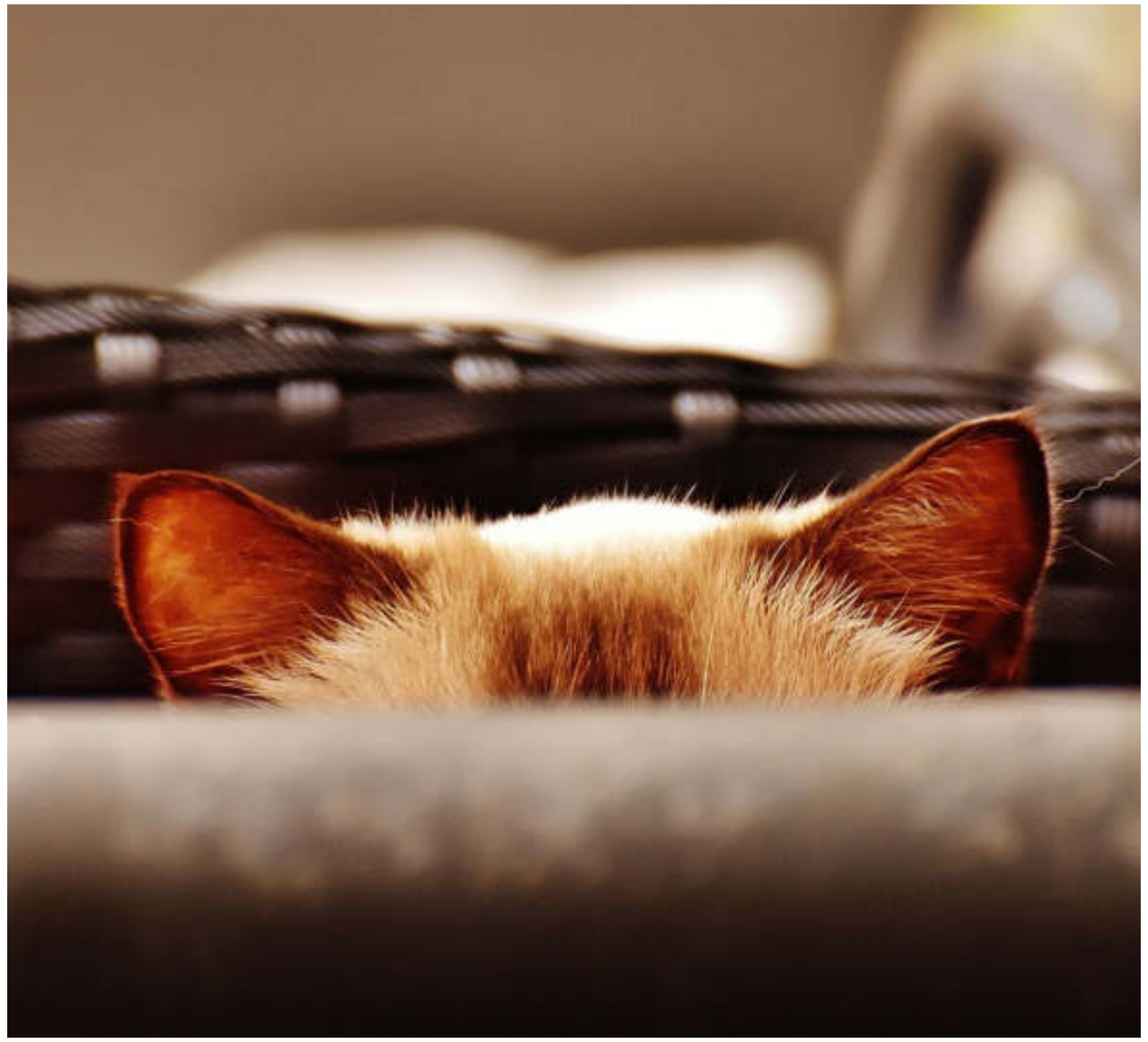
We know this is a cat, but how would you teach a machine?

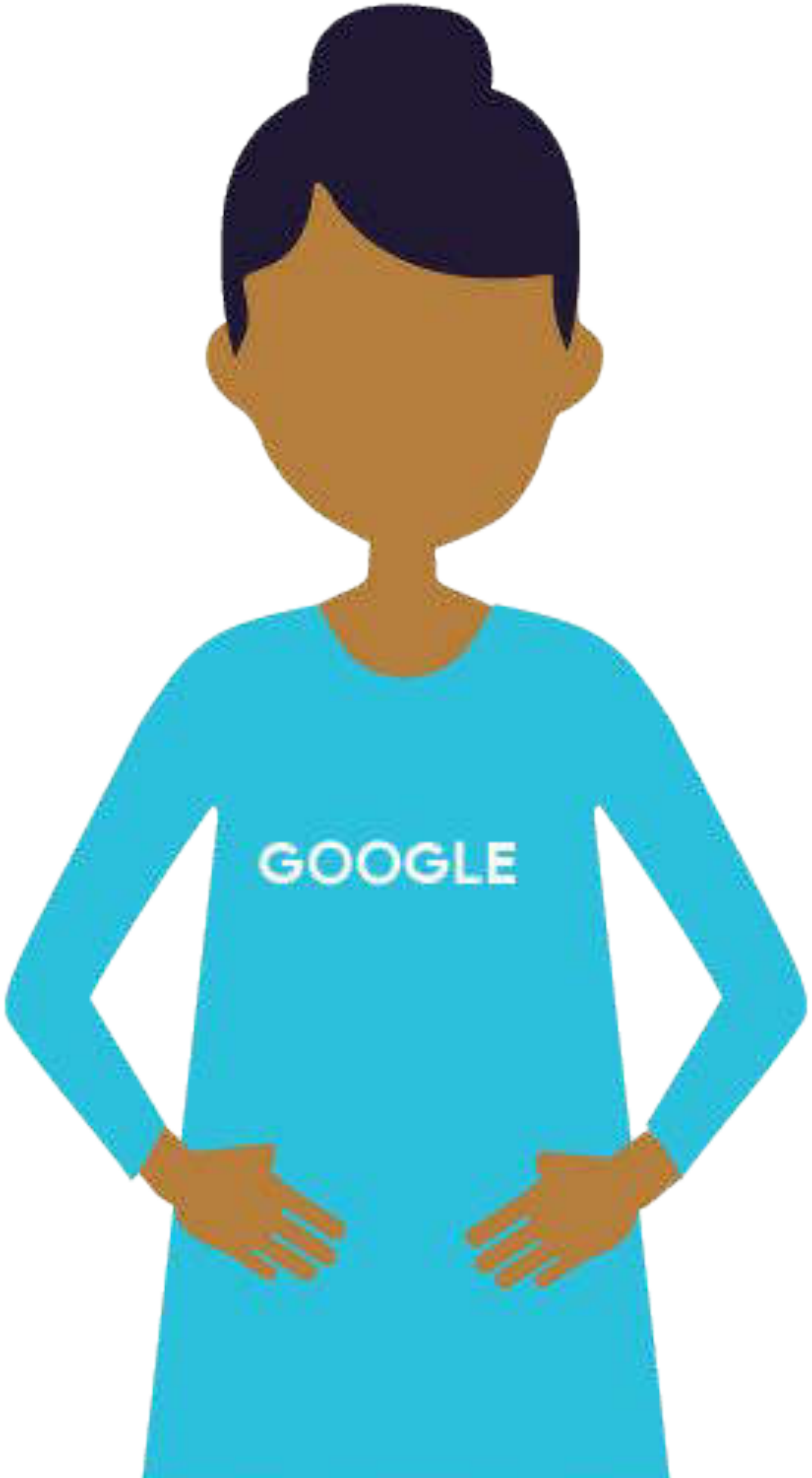


We know this is a cat, but how would you teach a machine?

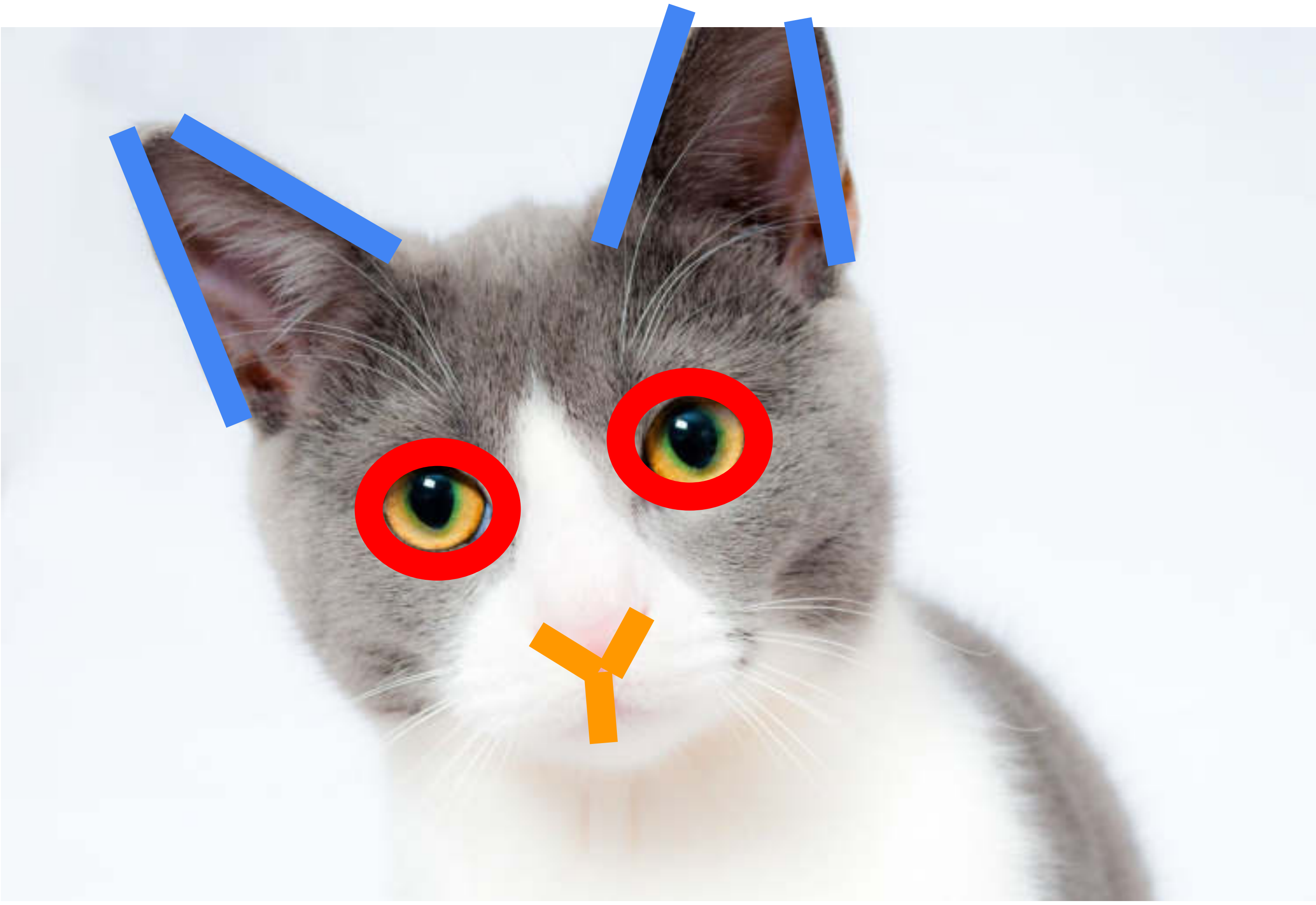


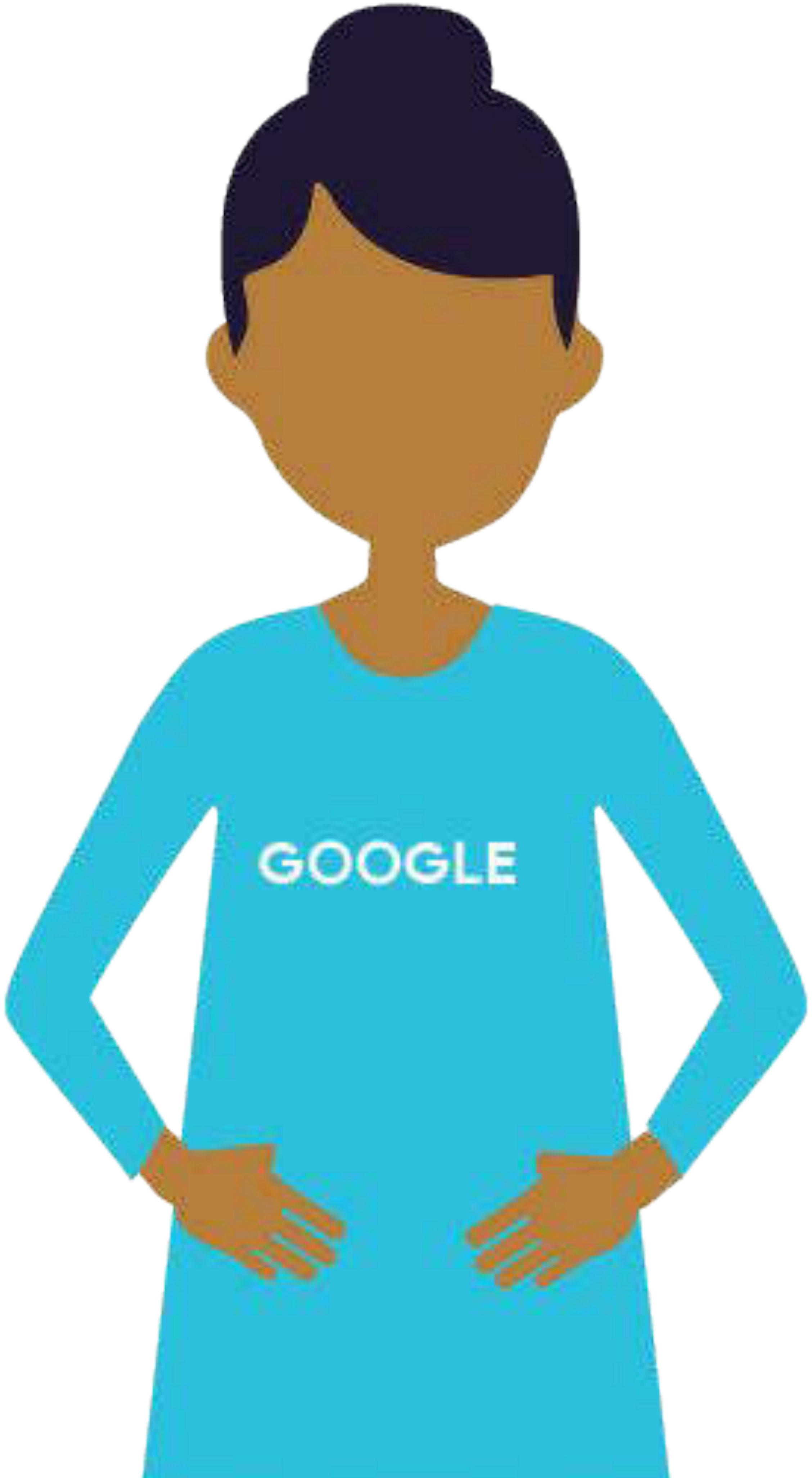
What about this?



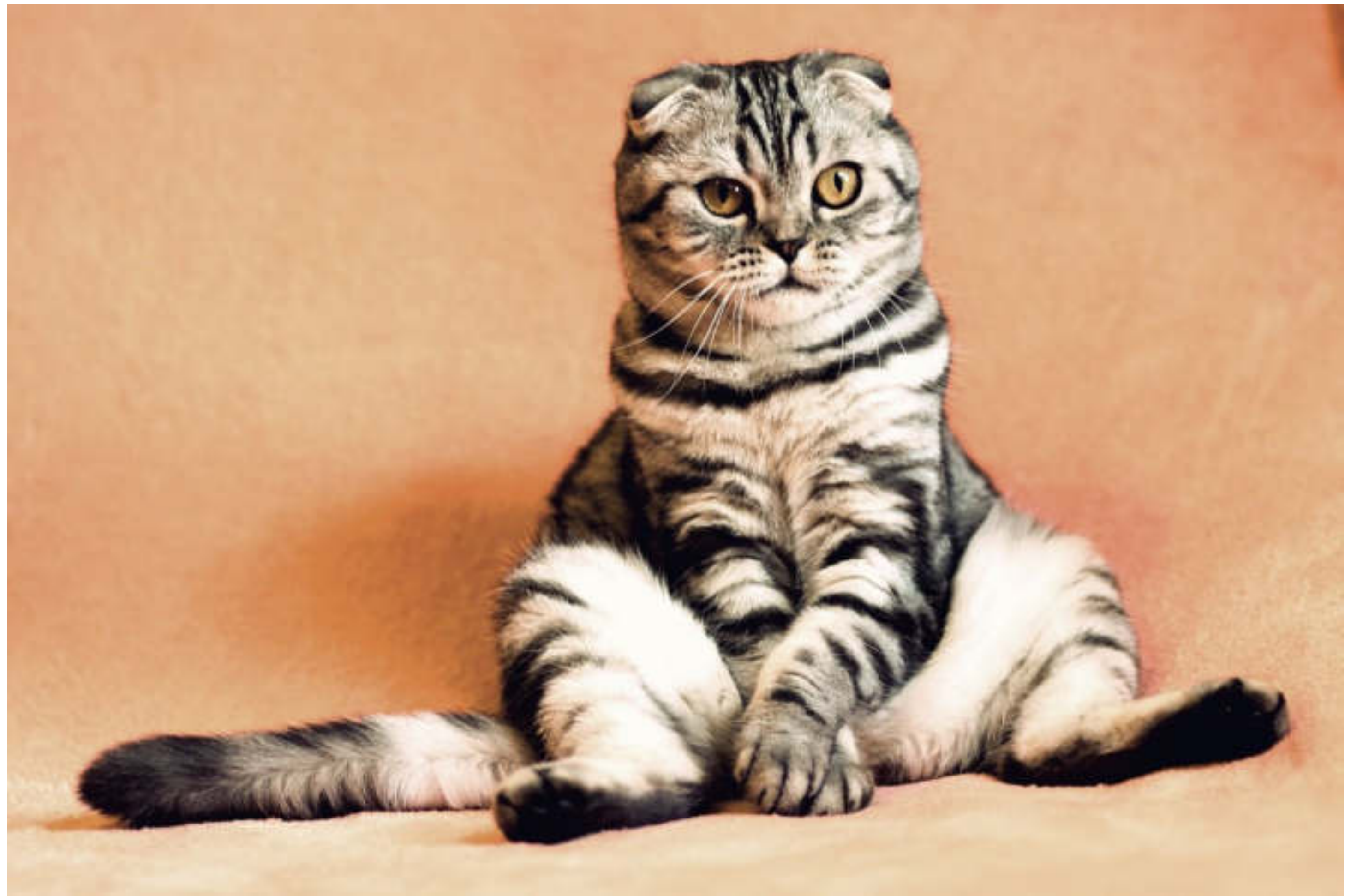


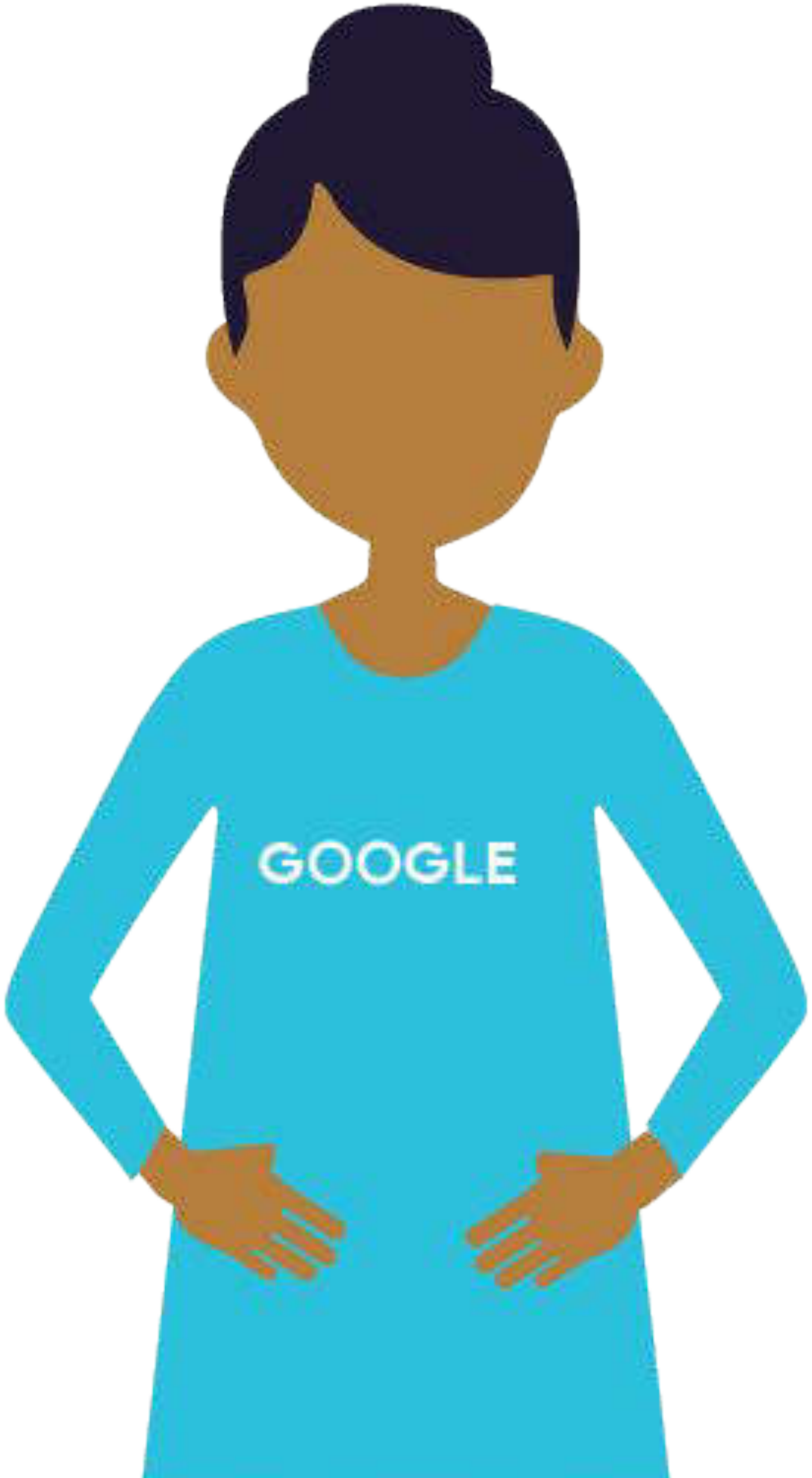
We know this is a cat, but how would you teach a machine?



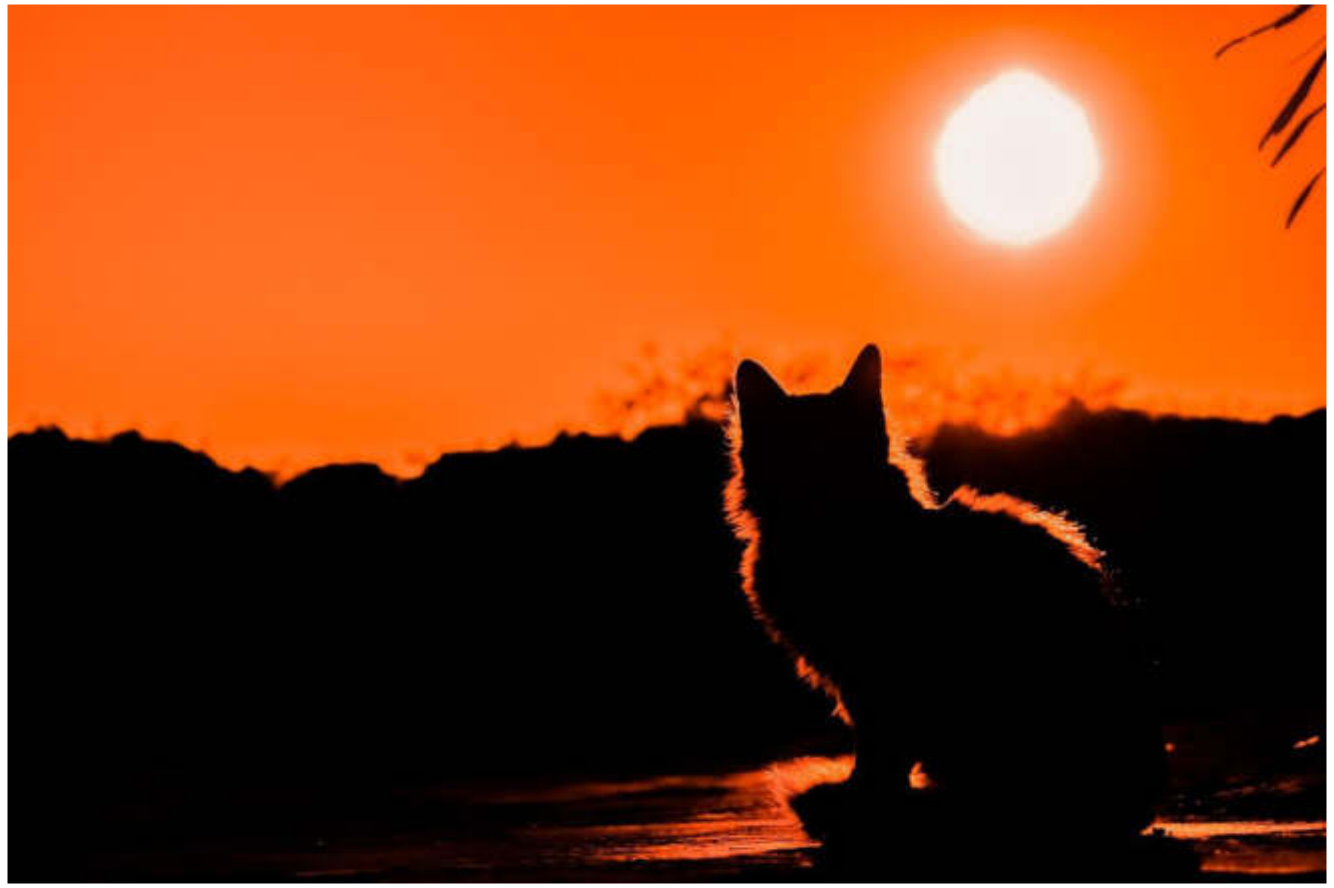


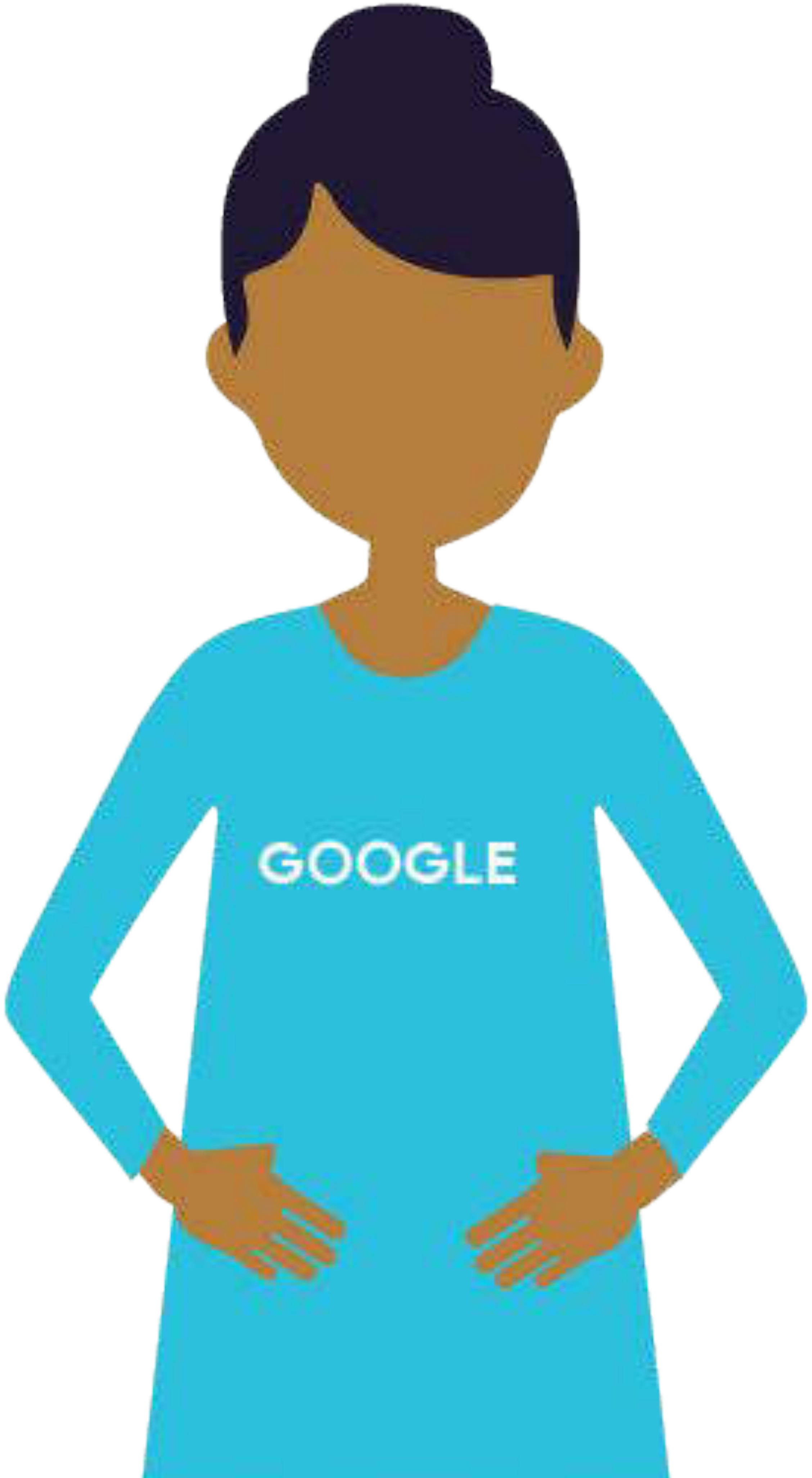
What about this?





Or even this?

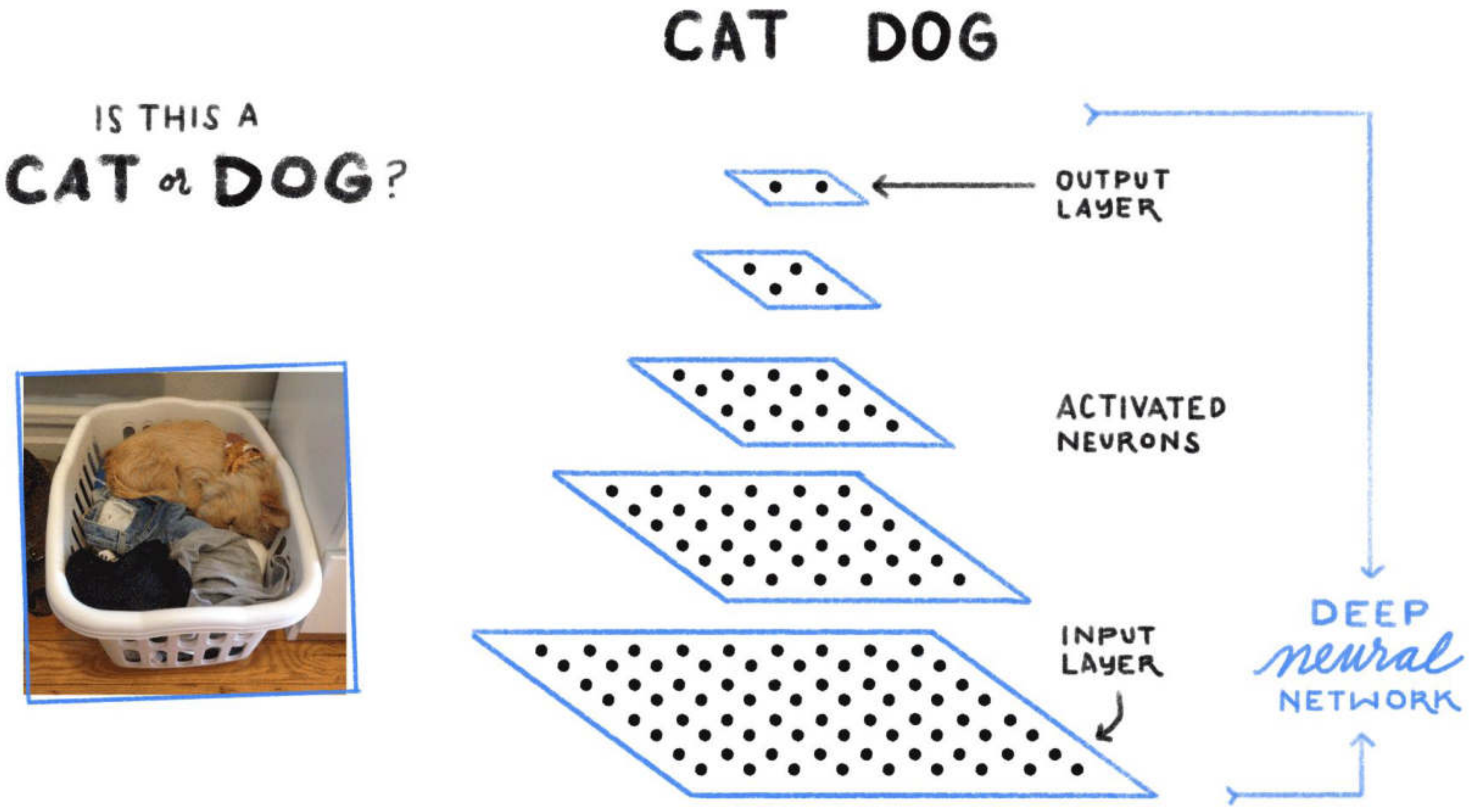


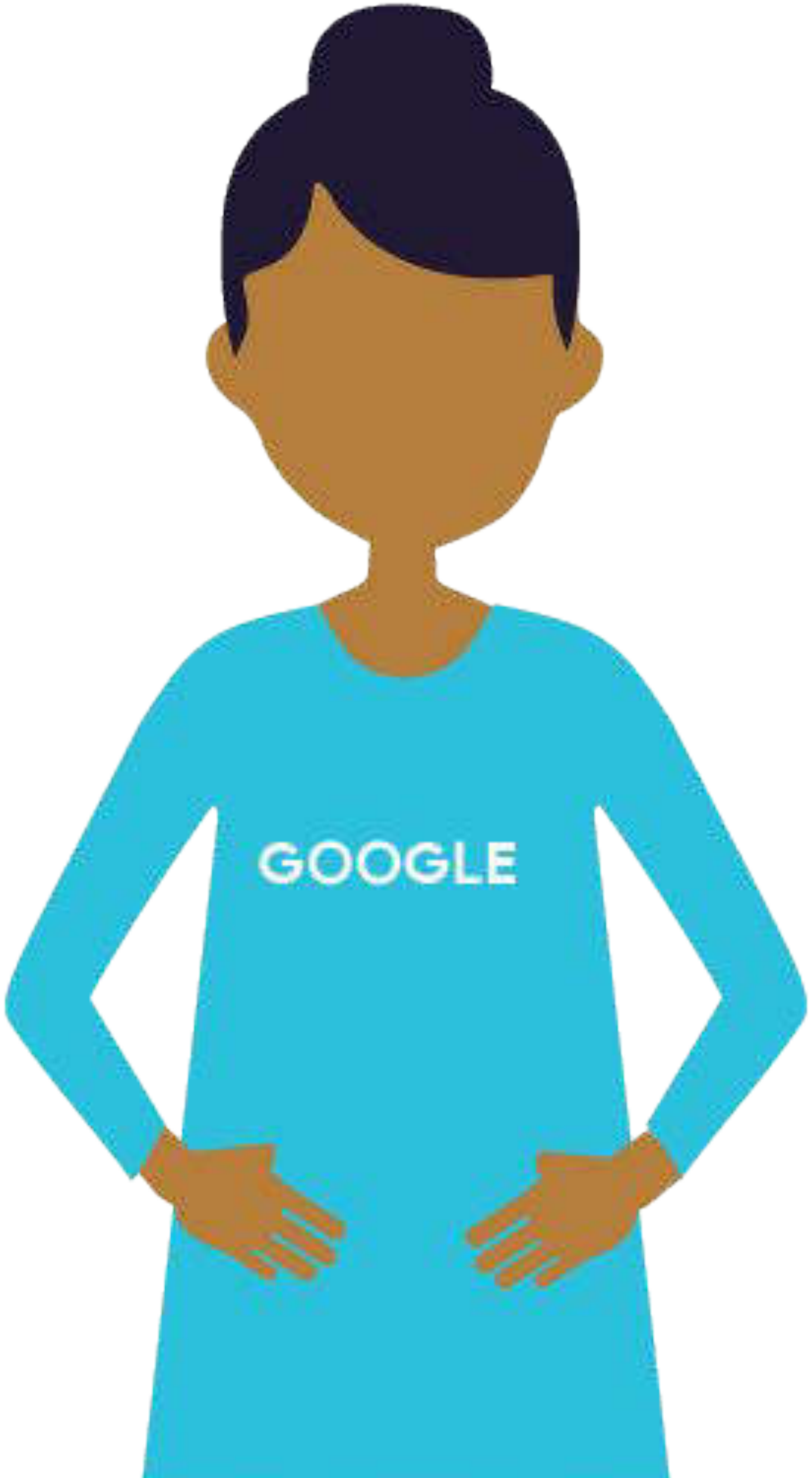


Google in 2012:

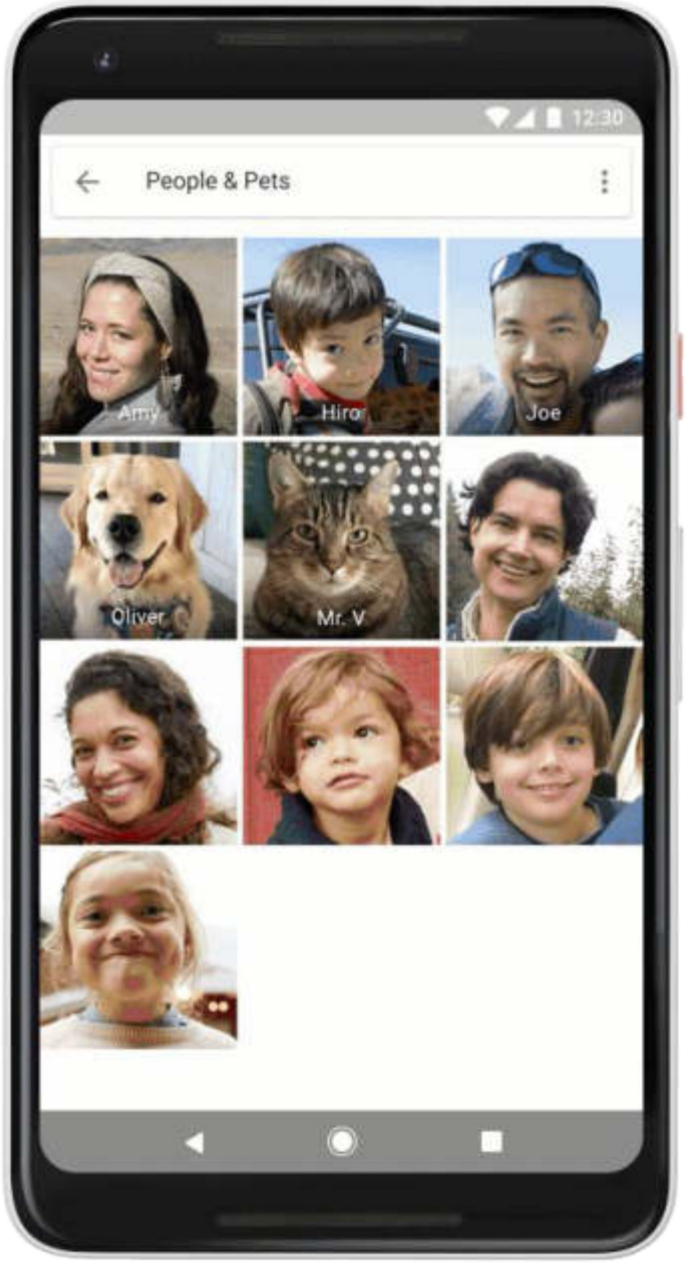
Show the computer 10 million images, have it find cats

Modern AI Applications use Deep Learning





Google Photos



Course 4: Applying Machine Learning to your Datasets

Module 1: Introduction to Machine Learning

Lesson Title: **Demo: ML in Google Photos**

Format: Talking Head + Lab Screencast

Video Name: T-BQML-O_1_l2_demo:_google_photos

Demo:
Google Photos Rex

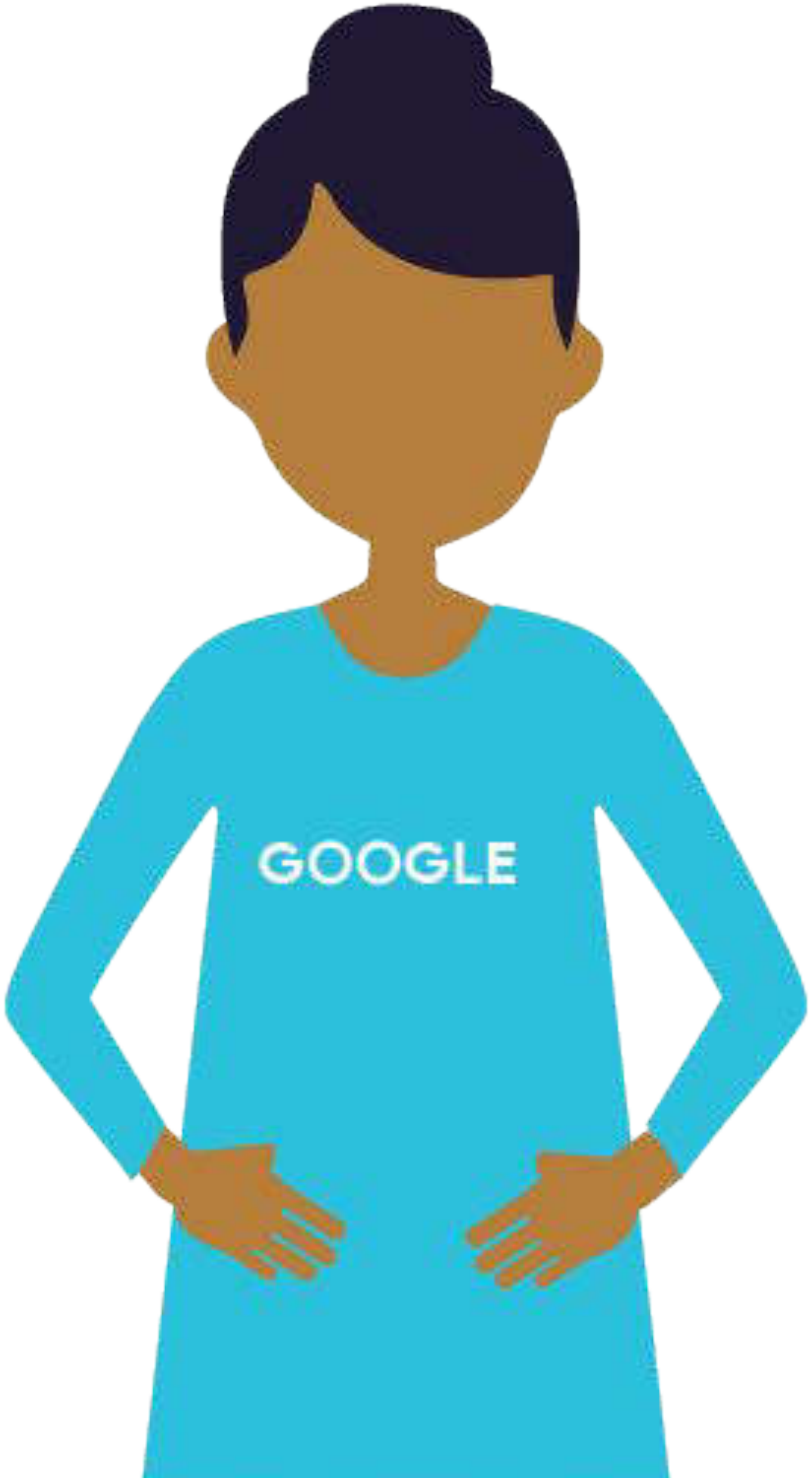
Course 4: Applying Machine Learning to your Datasets

Module 1: Introduction to Machine Learning

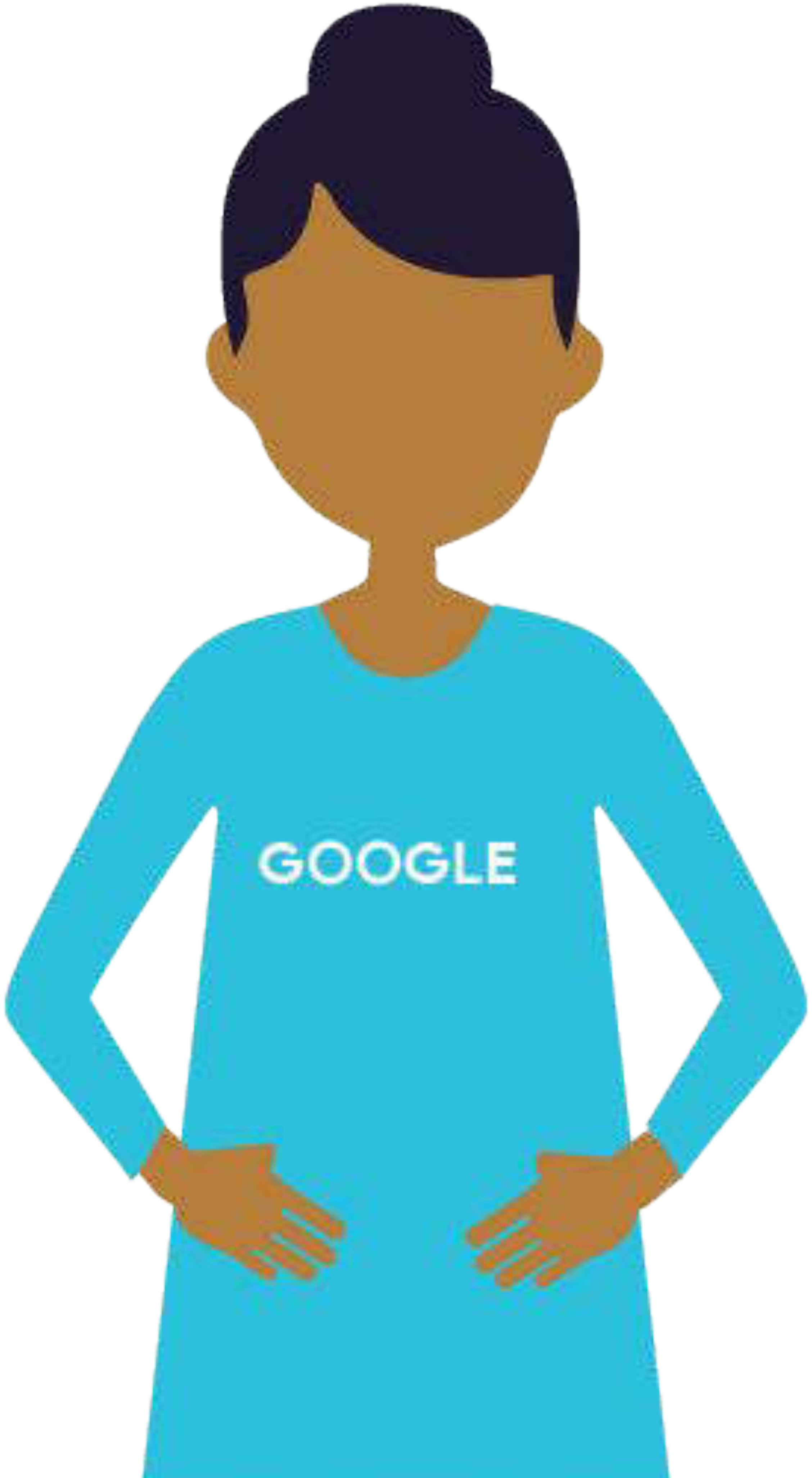
Lesson Title: **Deep Learning**

Format: Talking Head

Video Name: T-BQML-O_1_l3_deep_learning



Modern AI Applications
use Deep Learning



Waymo Self-Driving Cars

Image Recognition and Translation

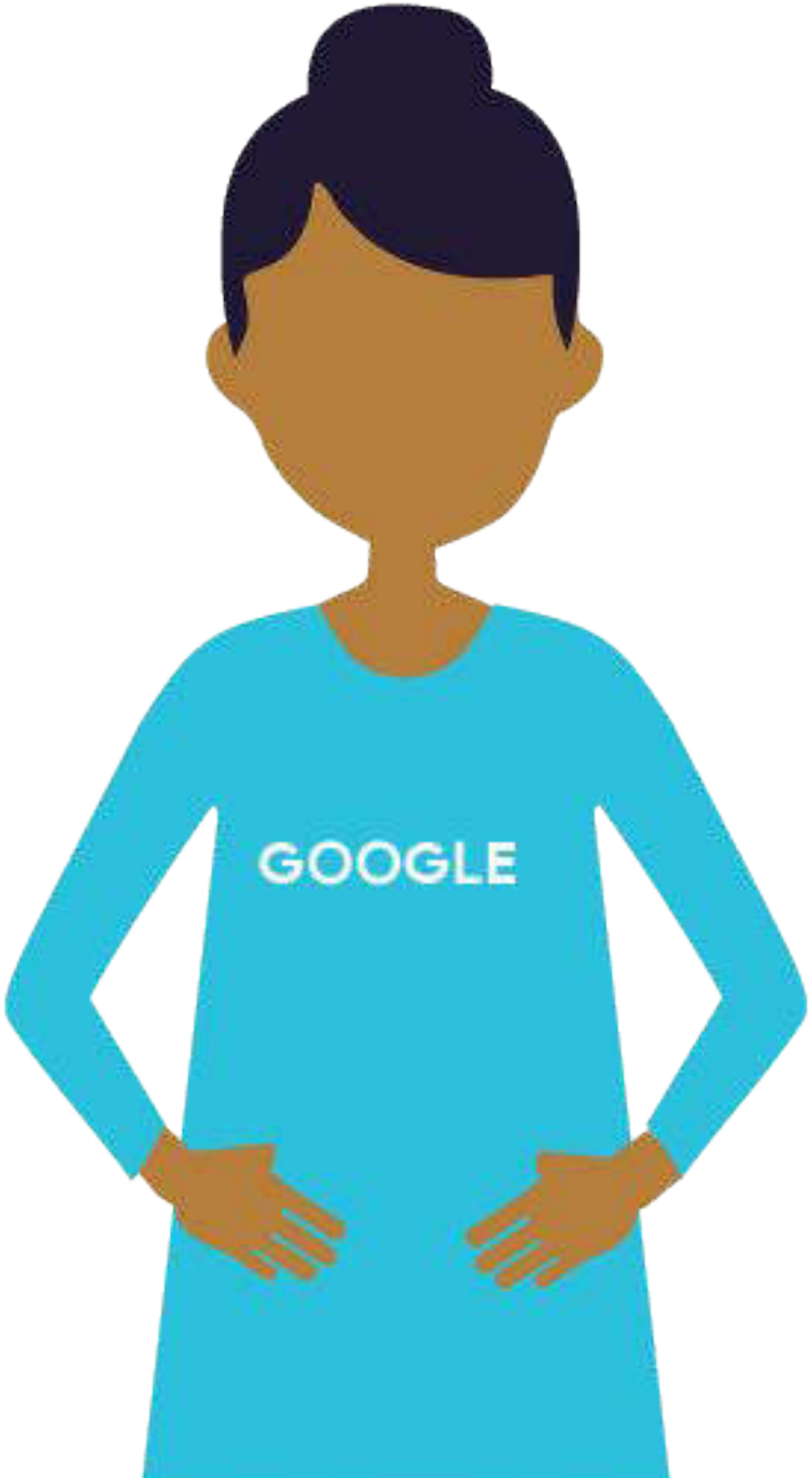
Course 4: Applying Machine Learning to your Datasets

Module 1: Introduction to Machine Learning

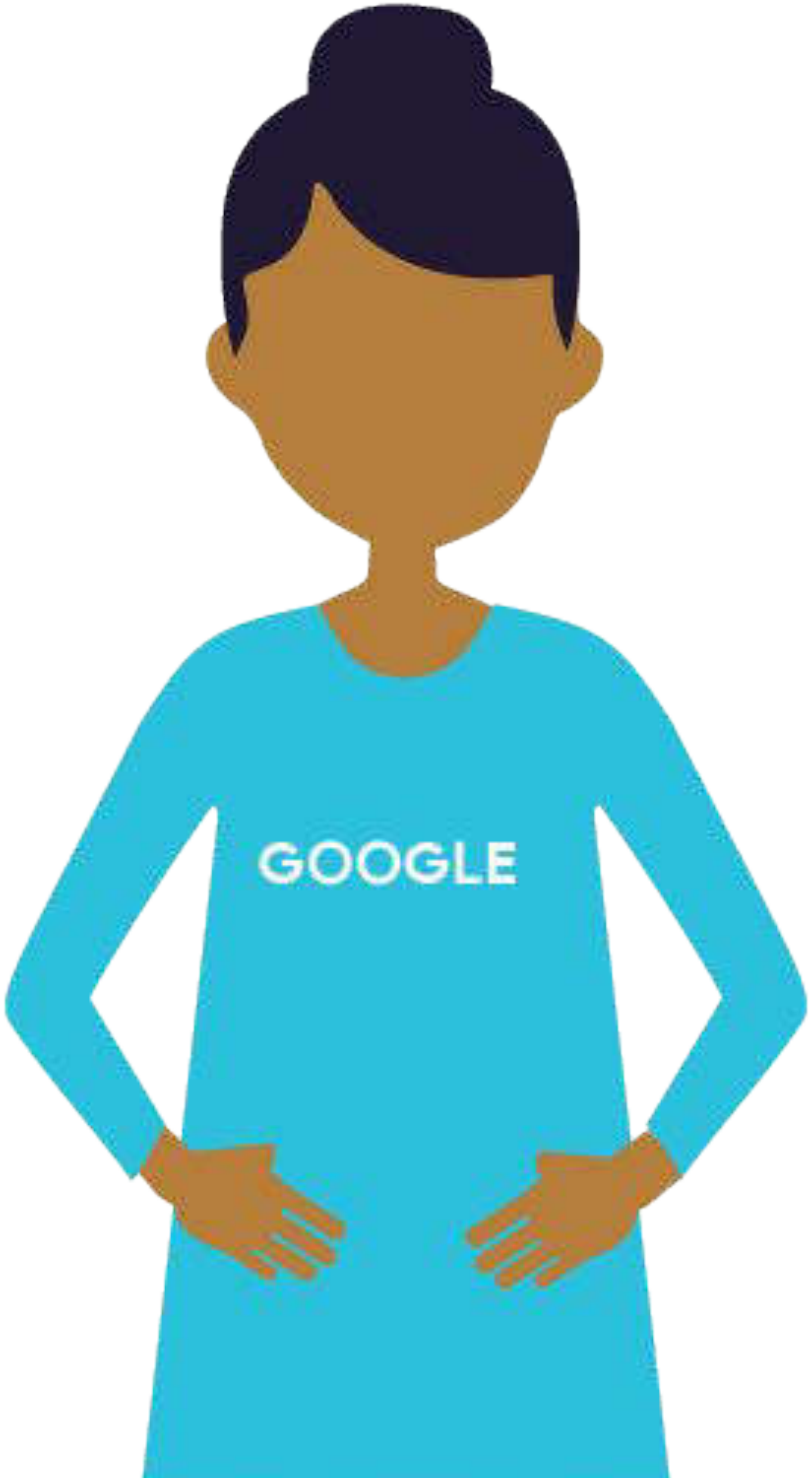
Lesson Title: **ML Applications for Business**

Format: Talking Head

Video Name: T-BQML-O_1_l4_ml_applications_for_business

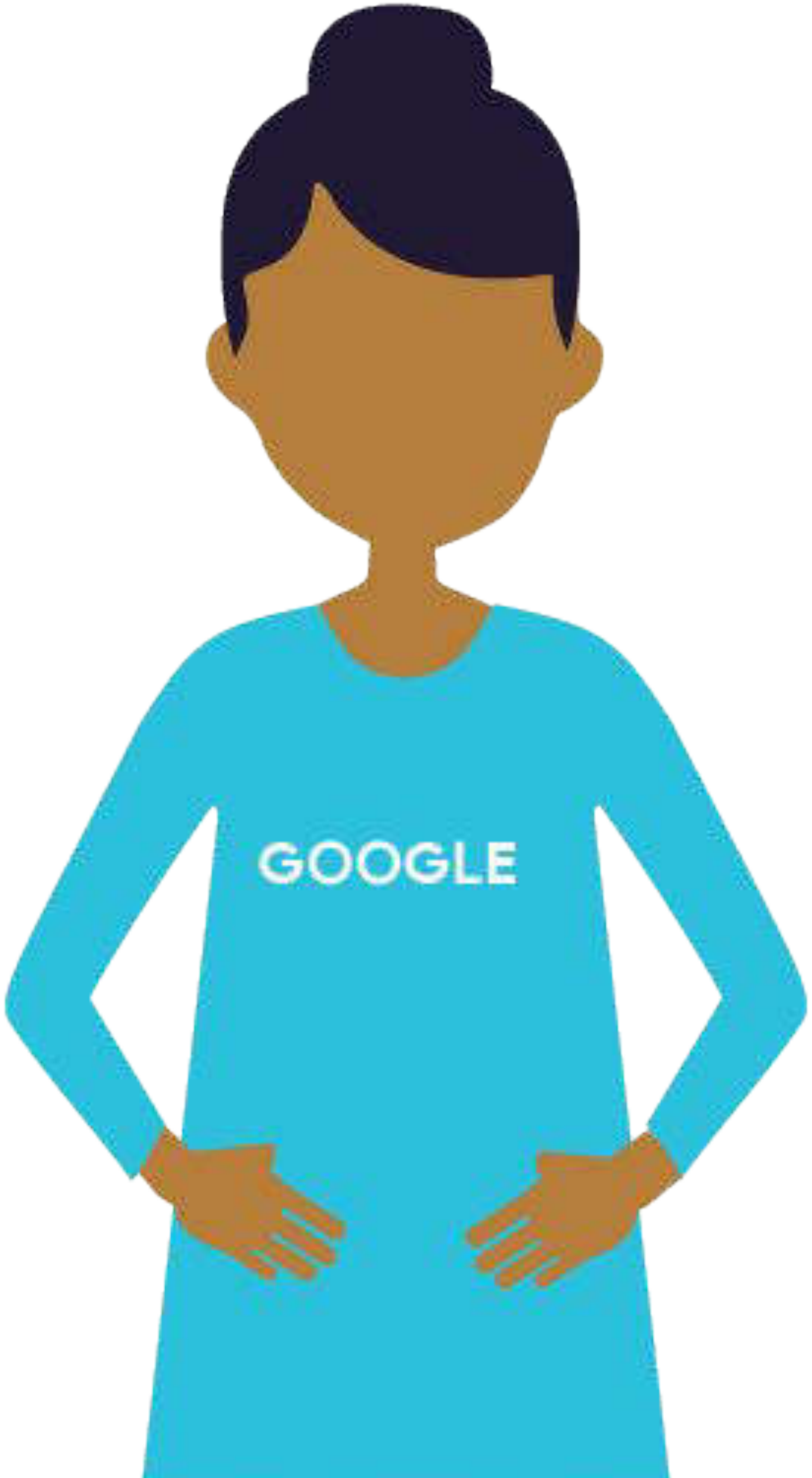


Where else can ML be applied?



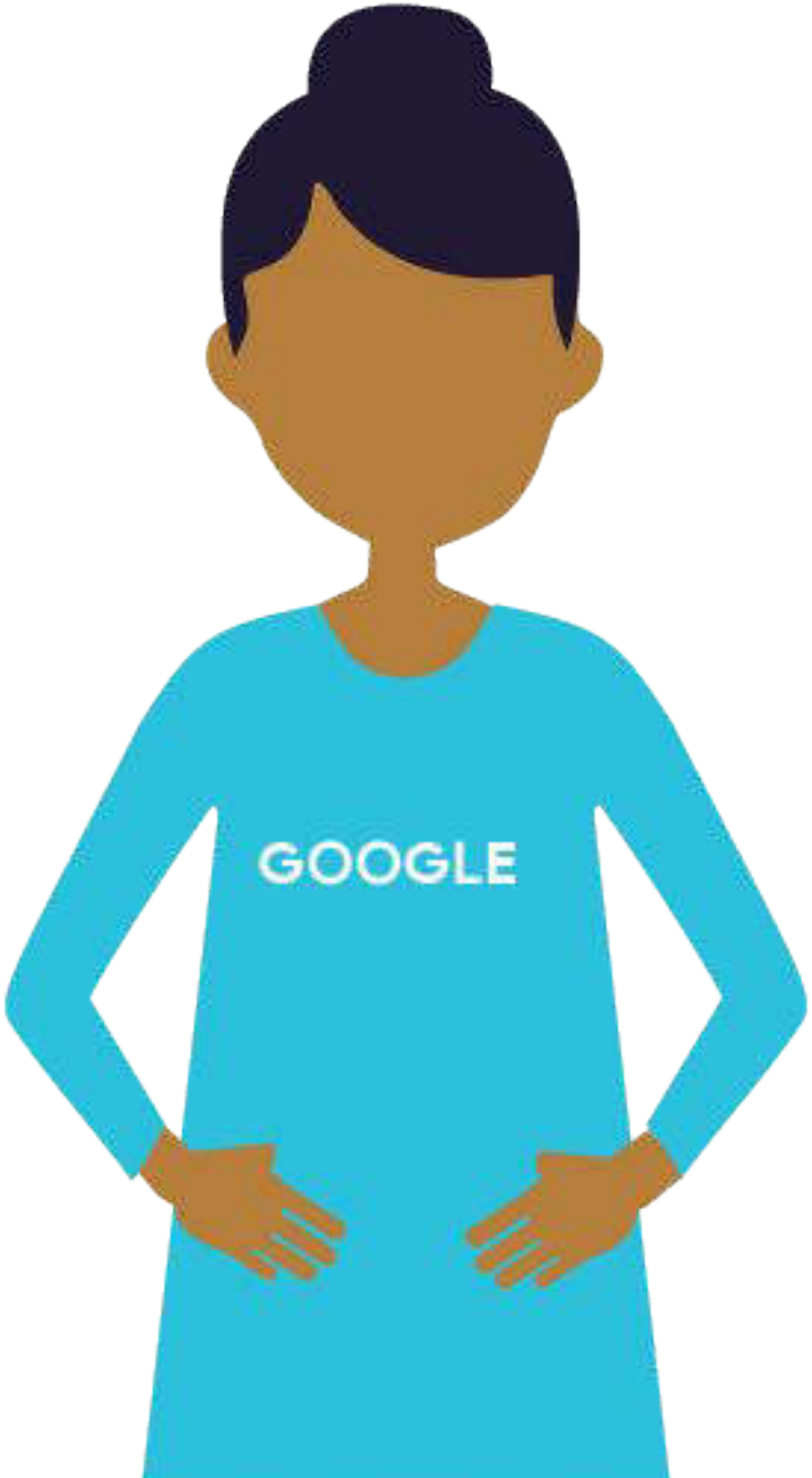
Operations

- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Process optimization
- Customer complaint resolution
- Support automation



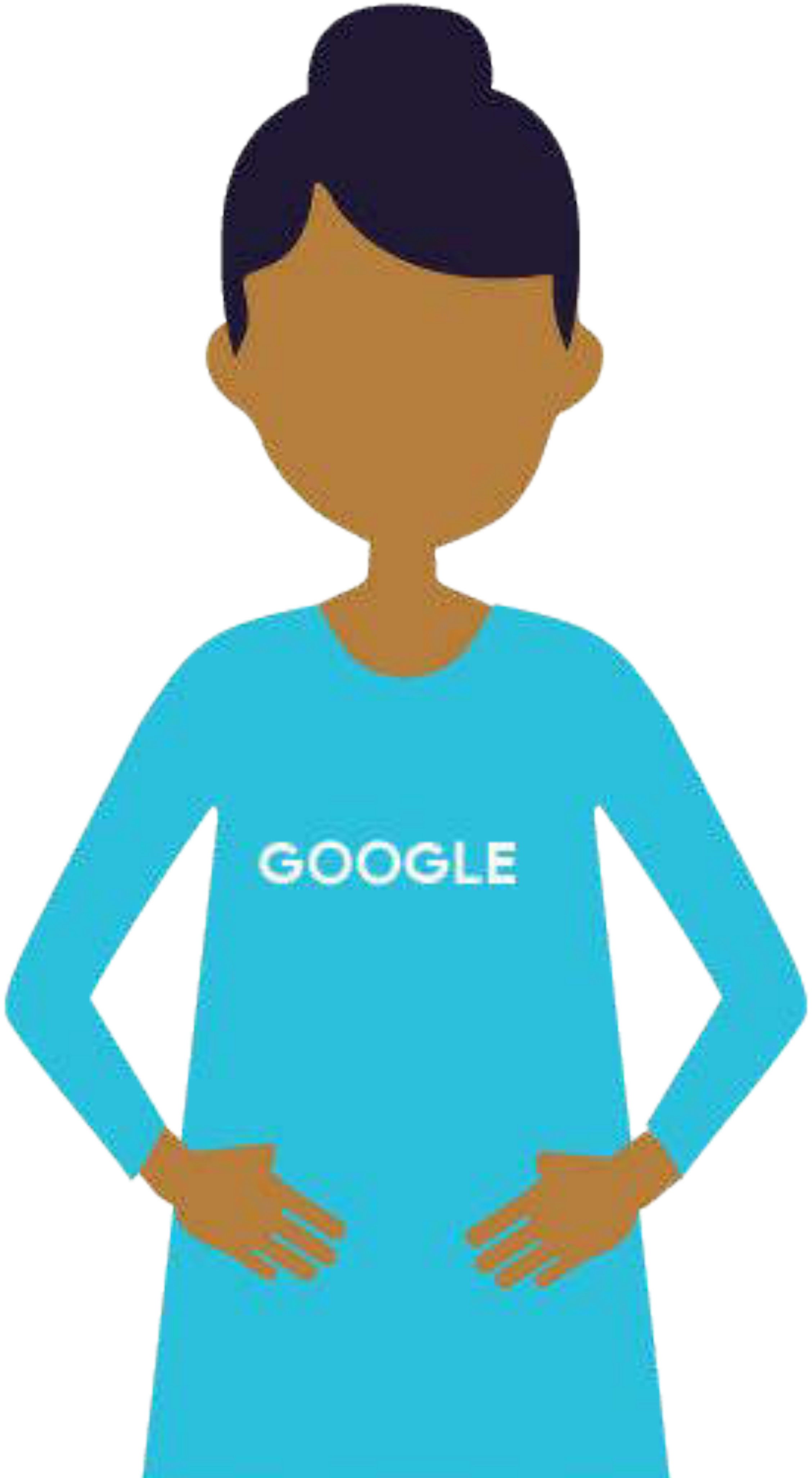
Sales

- Product usage analytics
- Recommendation engine
- Cross-selling and upselling
- Sales campaign management
- Propensity to buy



Marketing

- Social media feedback analysis
- Upsell + cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value
- Customer segmentation
- Marketing campaign management



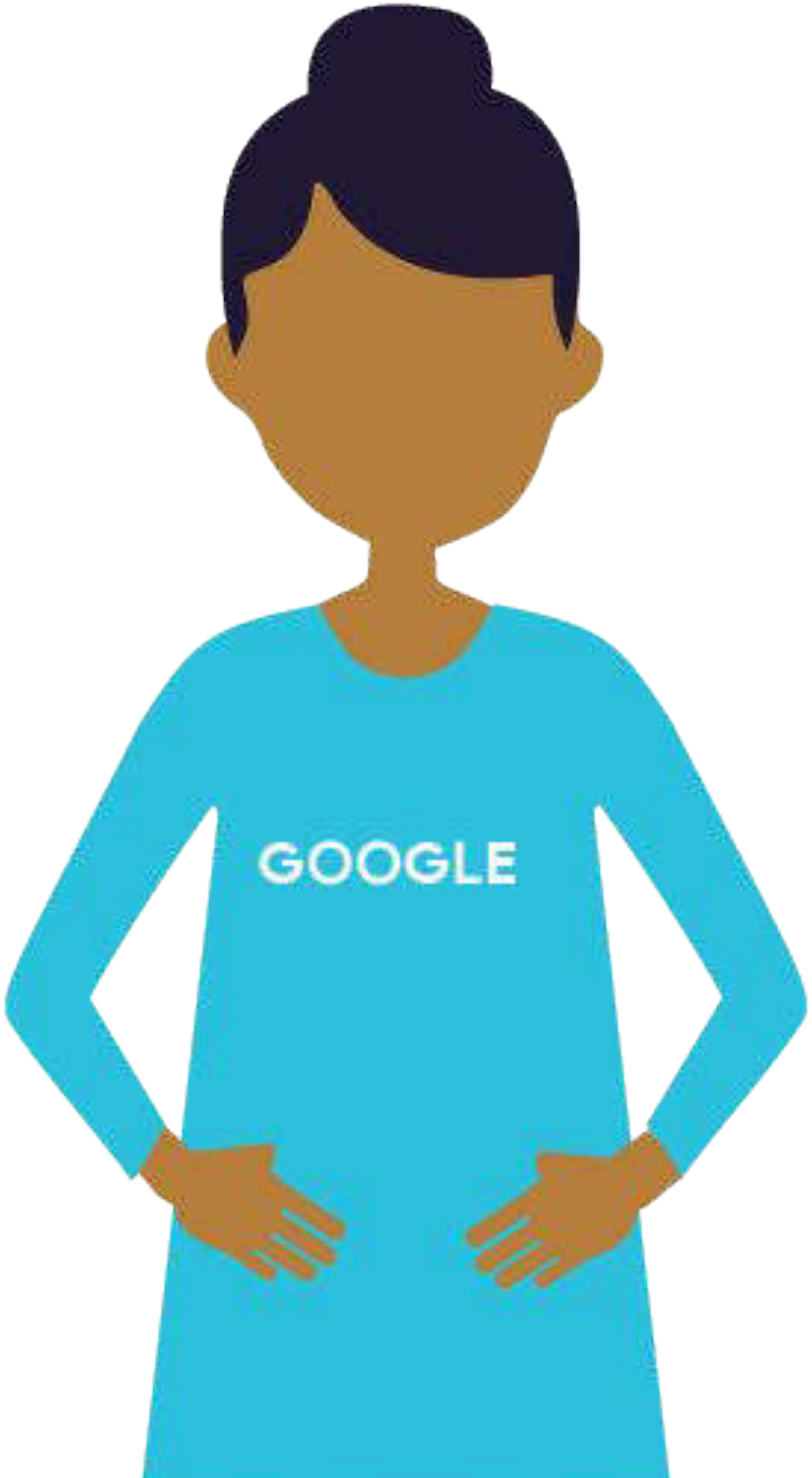
Finance

- Demand forecasting
- Risk analytics and regulation
- Creditworthiness evaluation

Discussion Question:

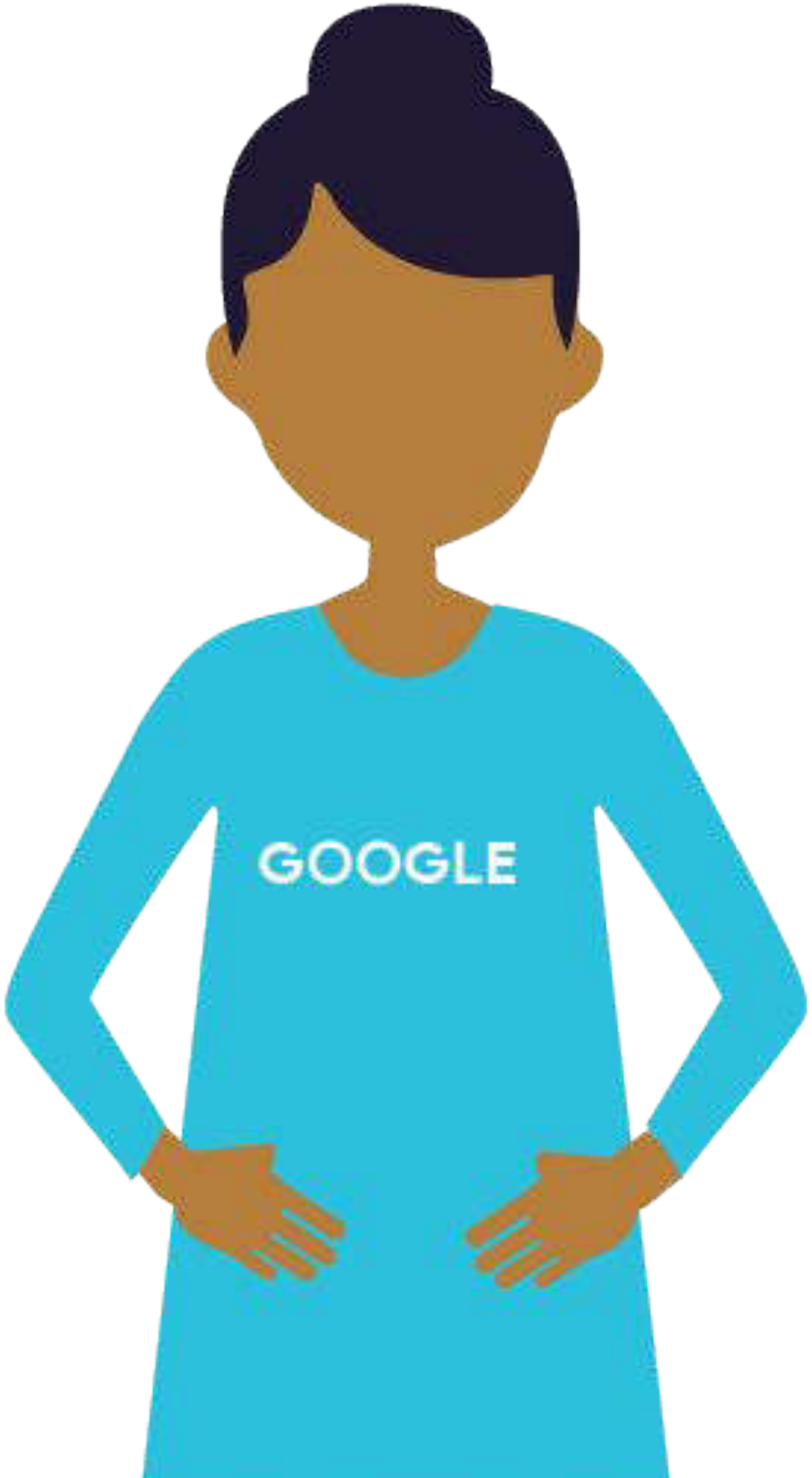
What other areas could you lead with examples and train a model?

Come up with your own or expand on one from the list

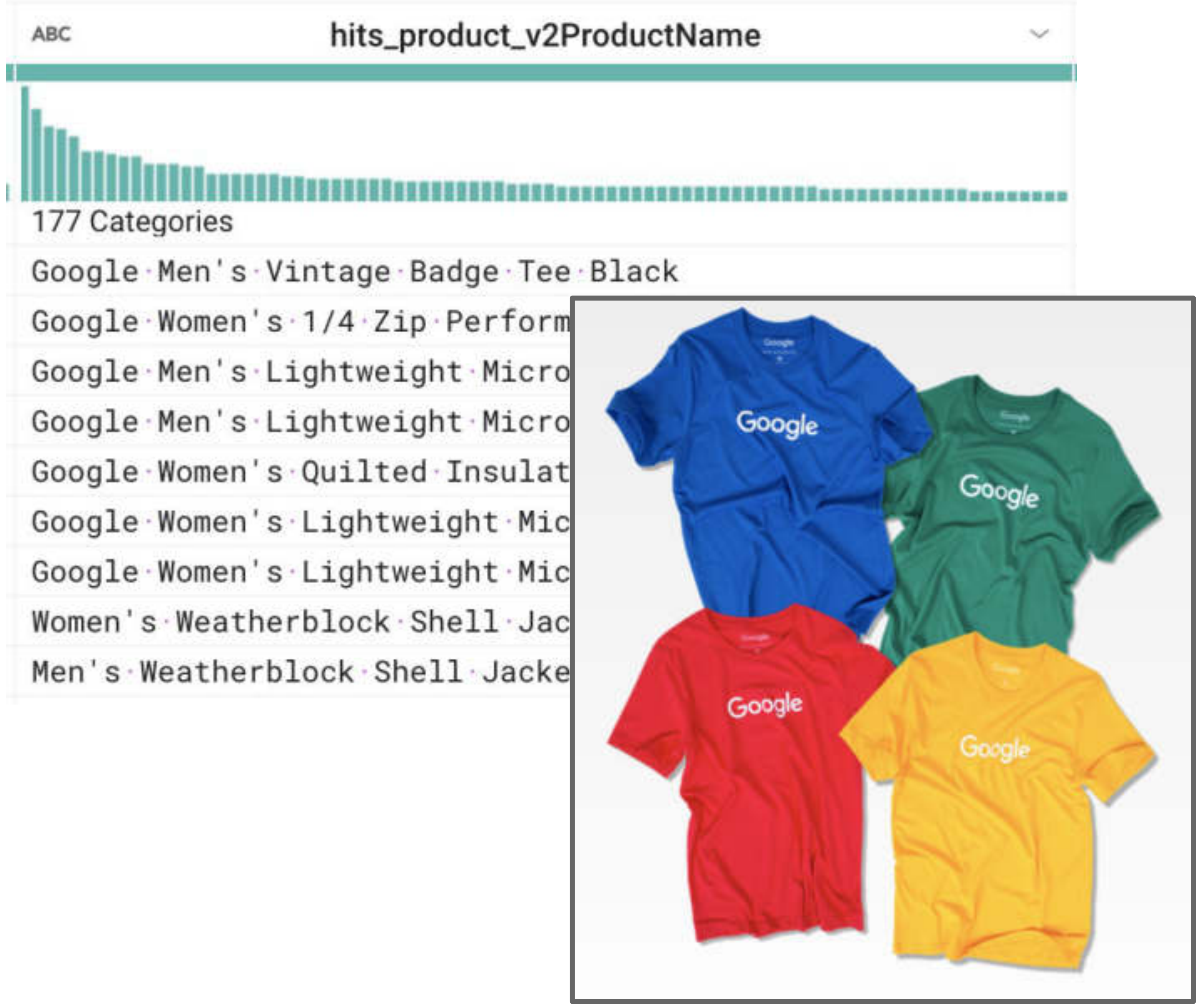


Marketing

- Social media feedback analysis
- Upsell + cross-channel marketing
- Market segmentation and targeting
- **Customer ROI and lifetime value**
- Customer segmentation
- Marketing campaign management



ML for Customer LTV

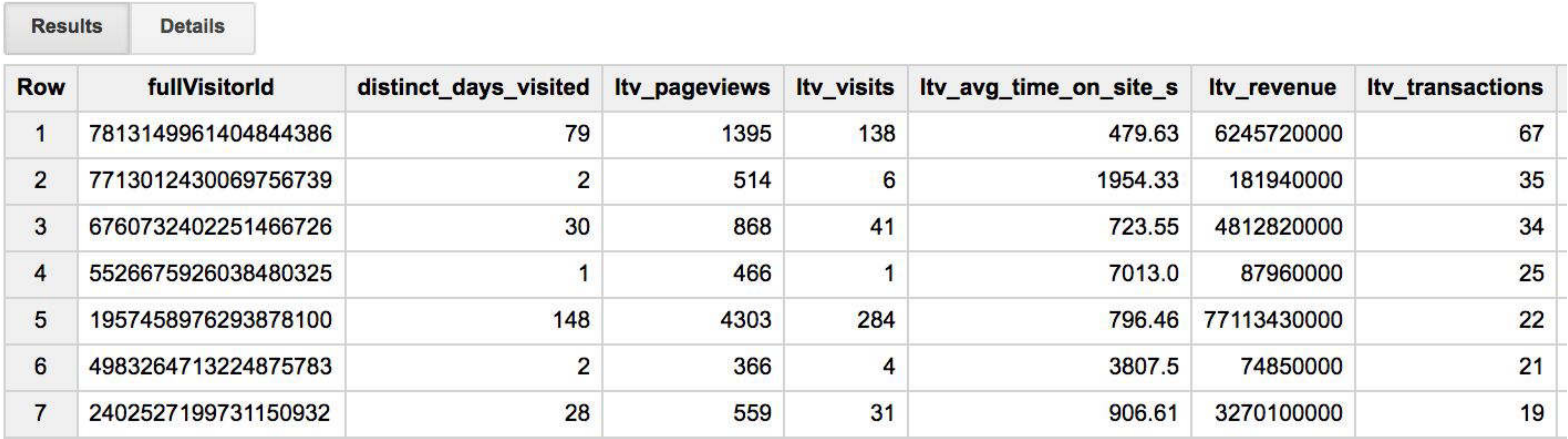


ABC hits_product_v2ProductName

177 Categories

Google Men's Vintage Badge Tee Black
Google Women's 1/4 Zip Perform
Google Men's Lightweight Micro
Google Men's Lightweight Micro
Google Women's Quilted Insulat
Google Women's Lightweight Mic
Google Women's Lightweight Mic
Women's Weatherblock Shell Jac
Men's Weatherblock Shell Jacke

Predict Lifetime Value (LTV) of a Customer



Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions
1	7813149961404844386	79	1395	138	479.63	6245720000	67
2	7713012430069756739	2	514	6	1954.33	181940000	35
3	6760732402251466726	30	868	41	723.55	4812820000	34
4	5526675926038480325	1	466	1	7013.0	87960000	25
5	1957458976293878100	148	4303	284	796.46	77113430000	22
6	4983264713224875783	2	366	4	3807.5	74850000	21
7	2402527199731150932	28	559	31	906.61	3270100000	19

Course 4: Applying Machine Learning to your Datasets

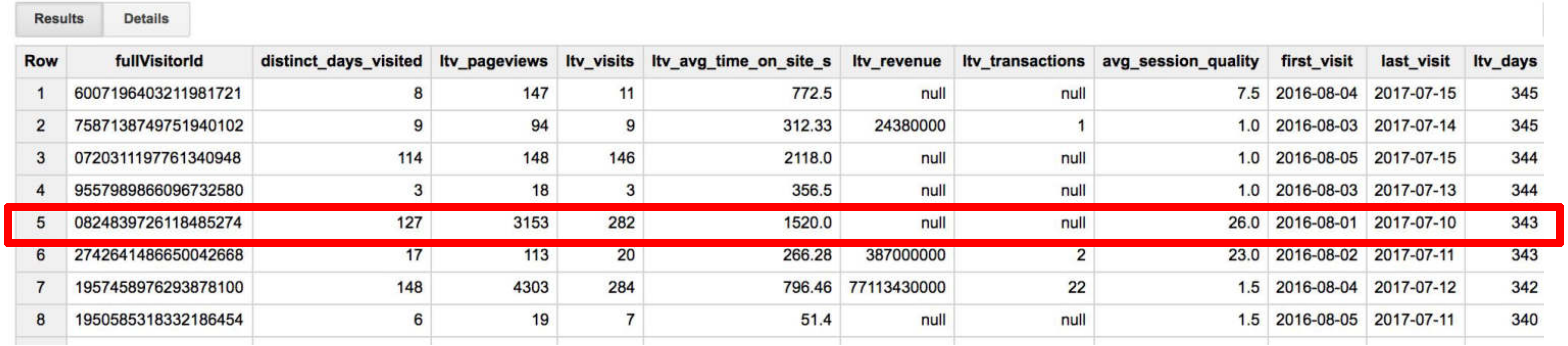
Module 1: Introduction to Machine Learning

Lesson Title: **Instances, Labels, Features, and Models**

Format: Talking Head

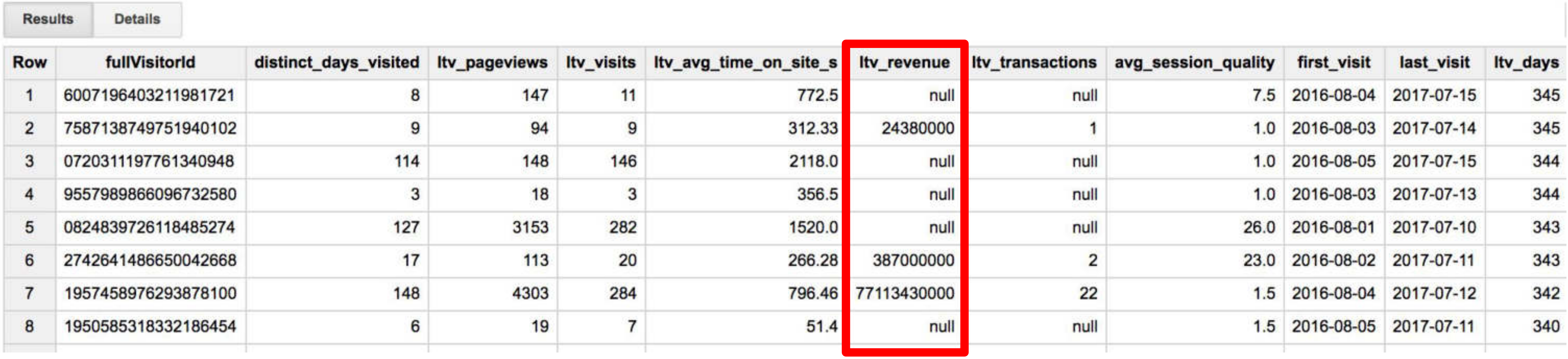
Video Name:
T-BQML-O_1_l5_instances,_labels,_features,_and_models

An instance (or observation) is a row of data



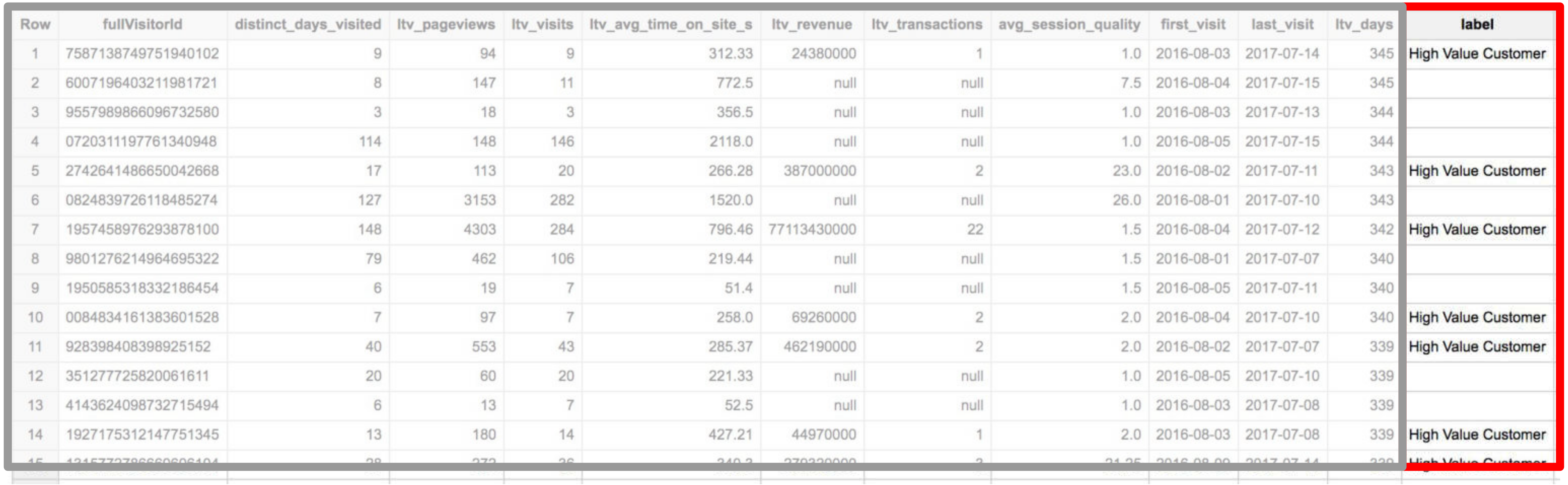
Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days
1	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345
2	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345
3	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344
4	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344
5	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343
6	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342
8	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340

A label is the correct answer



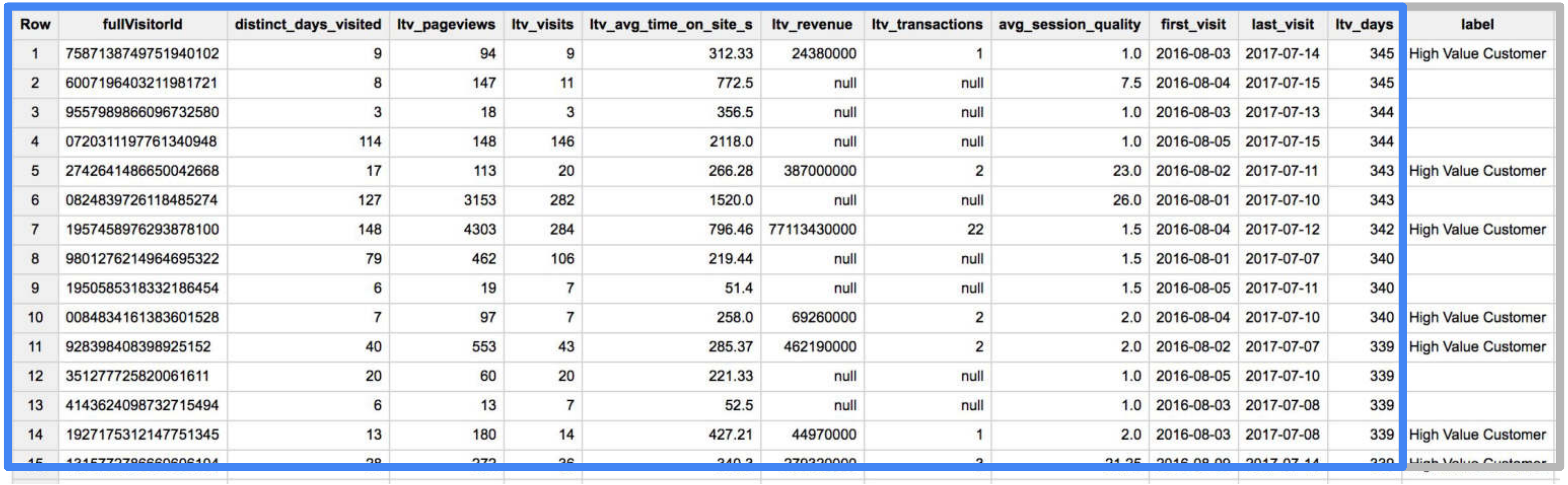
Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days
1	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345
2	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345
3	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344
4	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344
5	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343
6	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342
8	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340

A label is the correct answer



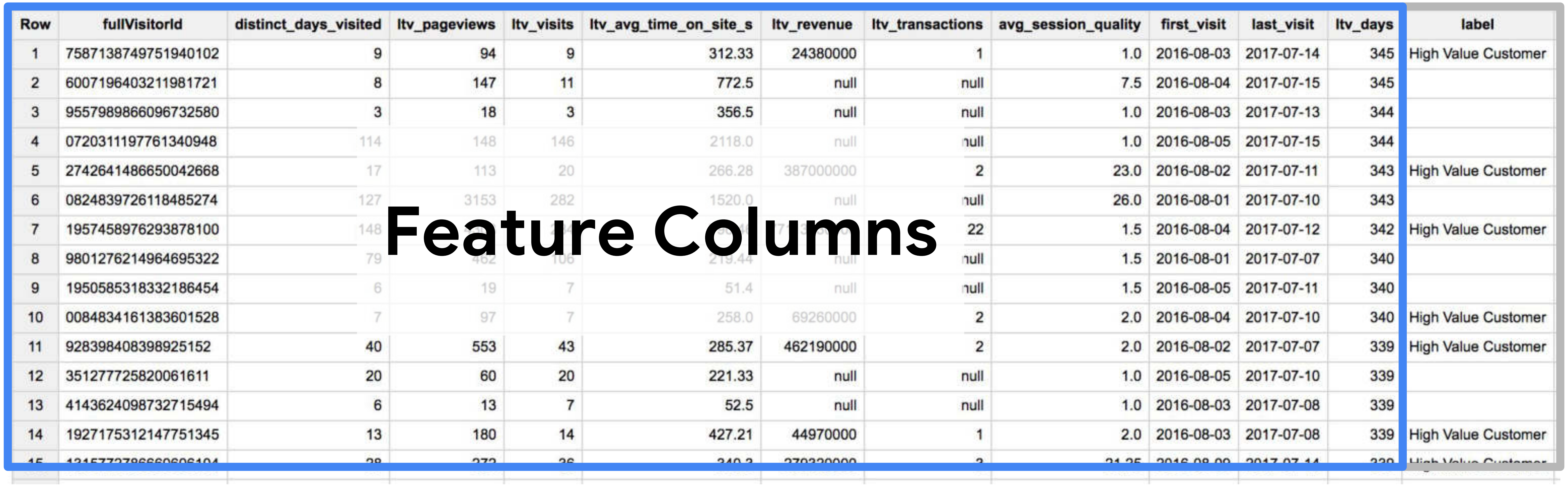
Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days	label
1	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345	High Value Customer
2	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345	
3	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344	
4	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344	
5	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343	High Value Customer
6	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343	
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342	High Value Customer
8	9801276214964695322	79	462	106	219.44	null	null	1.5	2016-08-01	2017-07-07	340	
9	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340	
10	0084834161383601528	7	97	7	258.0	69260000	2	2.0	2016-08-04	2017-07-10	340	High Value Customer
11	9283984083989251 52	40	553	43	285.37	462190000	2	2.0	2016-08-02	2017-07-07	339	High Value Customer
12	3512777258200616 11	20	60	20	221.33	null	null	1.0	2016-08-05	2017-07-10	339	
13	4143624098732715494	6	13	7	52.5	null	null	1.0	2016-08-03	2017-07-08	339	
14	1927175312147751345	13	180	14	427.21	44970000	1	2.0	2016-08-03	2017-07-08	339	High Value Customer

What about the other columns?



Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days	label
1	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345	High Value Customer
2	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345	
3	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344	
4	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344	
5	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343	High Value Customer
6	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343	
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342	High Value Customer
8	9801276214964695322	79	462	106	219.44	null	null	1.5	2016-08-01	2017-07-07	340	
9	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340	
10	0084834161383601528	7	97	7	258.0	69260000	2	2.0	2016-08-04	2017-07-10	340	High Value Customer
11	928398408398925152	40	553	43	285.37	462190000	2	2.0	2016-08-02	2017-07-07	339	High Value Customer
12	351277725820061611	20	60	20	221.33	null	null	1.0	2016-08-05	2017-07-10	339	
13	4143624098732715494	6	13	7	52.5	null	null	1.0	2016-08-03	2017-07-08	339	
14	1927175312147751345	13	180	14	427.21	44970000	1	2.0	2016-08-03	2017-07-08	339	High Value Customer
15	1315773786660606104	28	272	36	340.3	279320000	3	21.25	2016-08-09	2017-07-14	339	High Value Customer

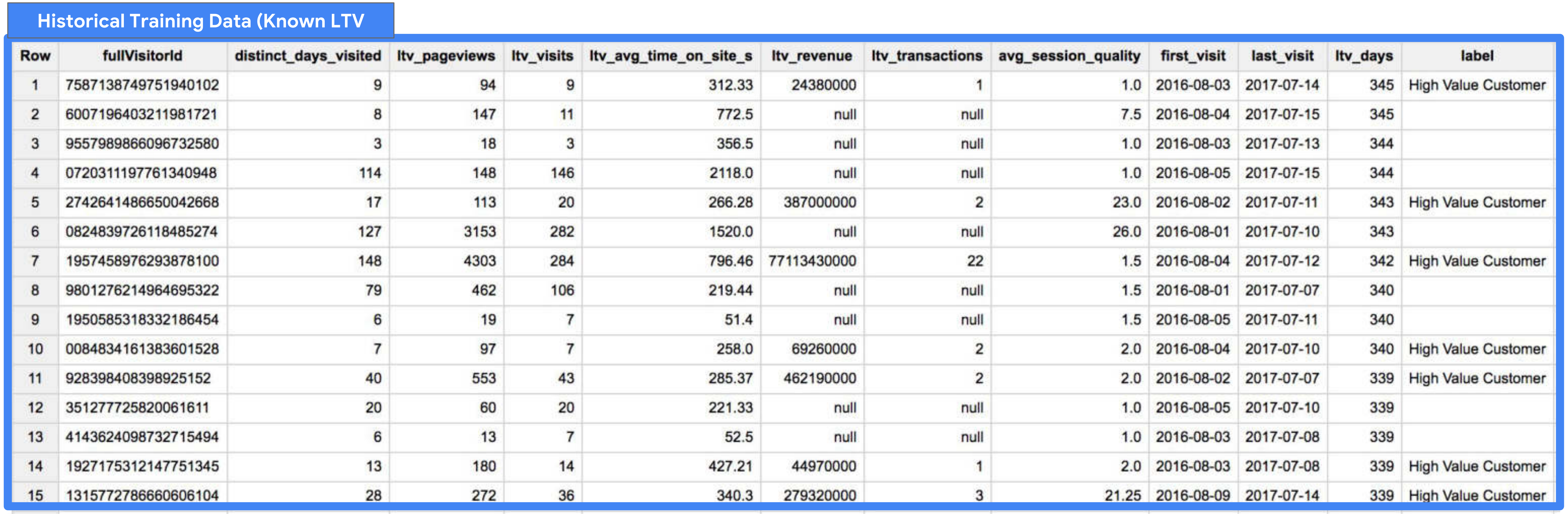
What about the other columns?



Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days	label
1	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345	High Value Customer
2	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345	
3	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344	
4	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344	
5	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343	High Value Customer
6	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343	
7	1957458976293878100	148	[illegible]	[illegible]	[illegible]	[illegible]	22	1.5	2016-08-04	2017-07-12	342	High Value Customer
8	9801276214964695322	79	462	106	219.44	null	null	1.5	2016-08-01	2017-07-07	340	
9	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340	
10	0084834161383601528	7	97	7	258.0	69260000	2	2.0	2016-08-04	2017-07-10	340	High Value Customer
11	928398408398925152	40	553	43	285.37	462190000	2	2.0	2016-08-02	2017-07-07	339	High Value Customer
12	351277725820061611	20	60	20	221.33	null	null	1.0	2016-08-05	2017-07-10	339	
13	4143624098732715494	6	13	7	52.5	null	null	1.0	2016-08-03	2017-07-08	339	
14	1927175312147751345	13	180	14	427.21	44970000	1	2.0	2016-08-03	2017-07-08	339	High Value Customer
15	[illegible]	28	272	26	[illegible]	[illegible]	[illegible]	21.25	[illegible]	2017-07-14	[illegible]	High Value Customer

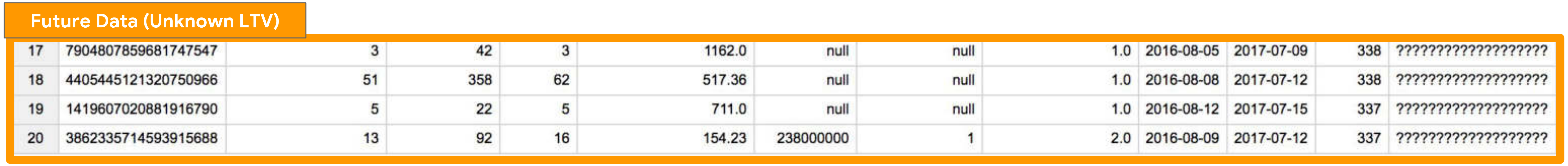
Feature Columns

What if I don’t know where a new customer will fit?



Historical Training Data (Known LTV

Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days	label
1	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345	High Value Customer
2	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345	
3	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344	
4	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344	
5	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343	High Value Customer
6	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343	
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342	High Value Customer
8	9801276214964695322	79	462	106	219.44	null	null	1.5	2016-08-01	2017-07-07	340	
9	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340	
10	0084834161383601528	7	97	7	258.0	69260000	2	2.0	2016-08-04	2017-07-10	340	High Value Customer
11	928398408398925152	40	553	43	285.37	462190000	2	2.0	2016-08-02	2017-07-07	339	High Value Customer
12	351277725820061611	20	60	20	221.33	null	null	1.0	2016-08-05	2017-07-10	339	
13	4143624098732715494	6	13	7	52.5	null	null	1.0	2016-08-03	2017-07-08	339	
14	1927175312147751345	13	180	14	427.21	44970000	1	2.0	2016-08-03	2017-07-08	339	High Value Customer
15	1315772786660606104	28	272	36	340.3	279320000	3	21.25	2016-08-09	2017-07-14	339	High Value Customer

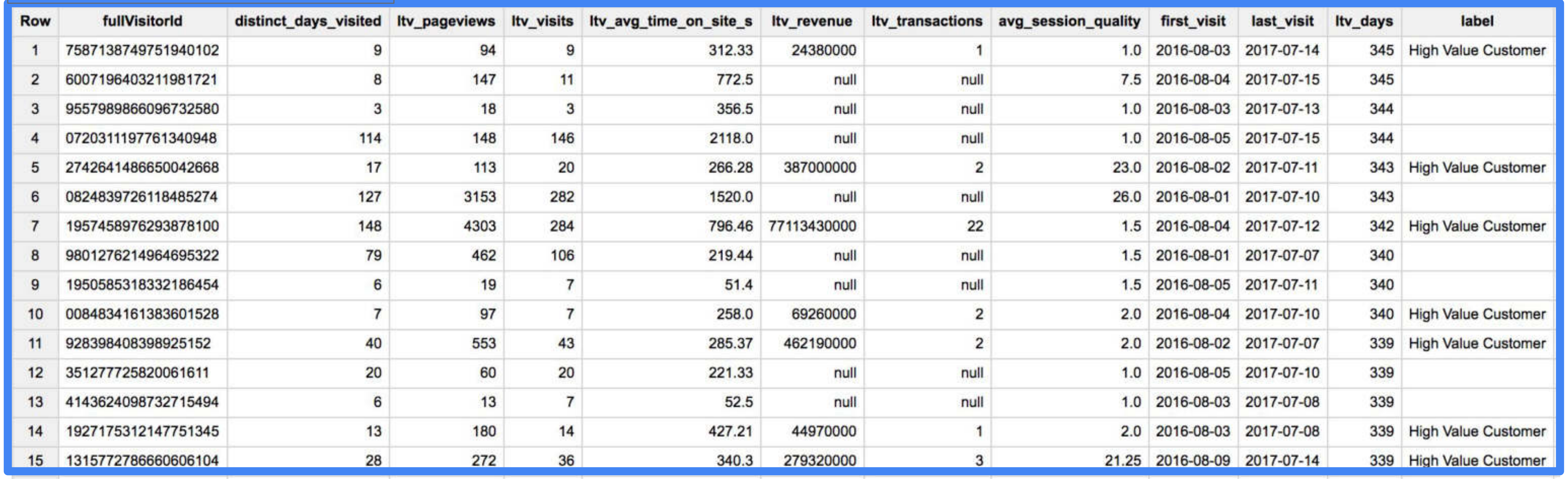


Future Data (Unknown LTV)

17	7904807859681747547	3	42	3	1162.0	null	null	1.0	2016-08-05	2017-07-09	338	?????????????????
18	4405445121320750966	51	358	62	517.36	null	null	1.0	2016-08-08	2017-07-12	338	?????????????????
19	1419607020881916790	5	22	5	711.0	null	null	1.0	2016-08-12	2017-07-15	337	?????????????????
20	3862335714593915688	13	92	16	154.23	238000000	1	2.0	2016-08-09	2017-07-12	337	?????????????????

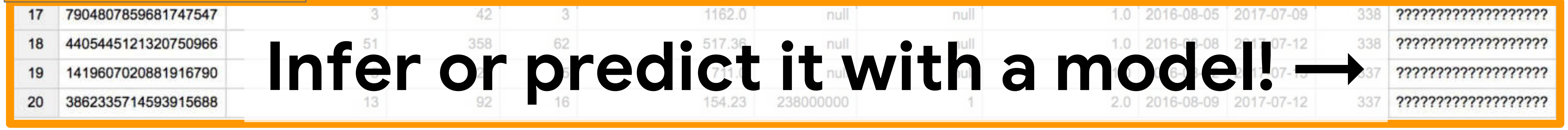
What if I don’t know where a new customer will fit?

Historical Training Data (Known LTV



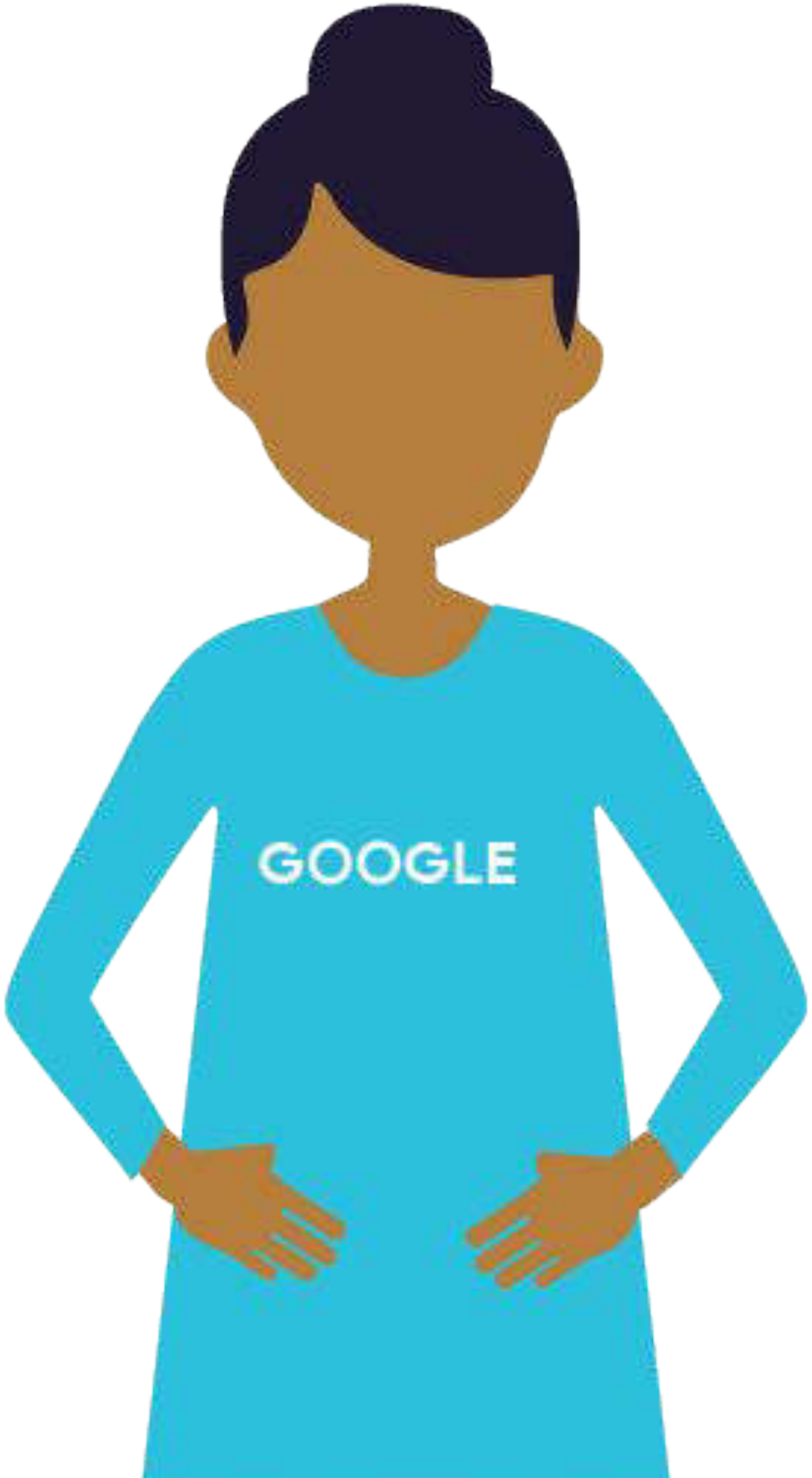
Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days	label
1	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345	High Value Customer
2	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345	
3	9557989866096732580	3	18	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344	
4	0720311197761340948	114	148	146	2118.0	null	null	1.0	2016-08-05	2017-07-15	344	
5	2742641486650042668	17	113	20	266.28	387000000	2	23.0	2016-08-02	2017-07-11	343	High Value Customer
6	0824839726118485274	127	3153	282	1520.0	null	null	26.0	2016-08-01	2017-07-10	343	
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342	High Value Customer
8	9801276214964695322	79	462	106	219.44	null	null	1.5	2016-08-01	2017-07-07	340	
9	1950585318332186454	6	19	7	51.4	null	null	1.5	2016-08-05	2017-07-11	340	
10	0084834161383601528	7	97	7	258.0	69260000	2	2.0	2016-08-04	2017-07-10	340	High Value Customer
11	928398408398925152	40	553	43	285.37	462190000	2	2.0	2016-08-02	2017-07-07	339	High Value Customer
12	351277725820061611	20	60	20	221.33	null	null	1.0	2016-08-05	2017-07-10	339	
13	4143624098732715494	6	13	7	52.5	null	null	1.0	2016-08-03	2017-07-08	339	
14	1927175312147751345	13	180	14	427.21	44970000	1	2.0	2016-08-03	2017-07-08	339	High Value Customer
15	1315772786660606104	28	272	36	340.3	279320000	3	21.25	2016-08-09	2017-07-14	339	High Value Customer

Future Data (Unknown LTV)

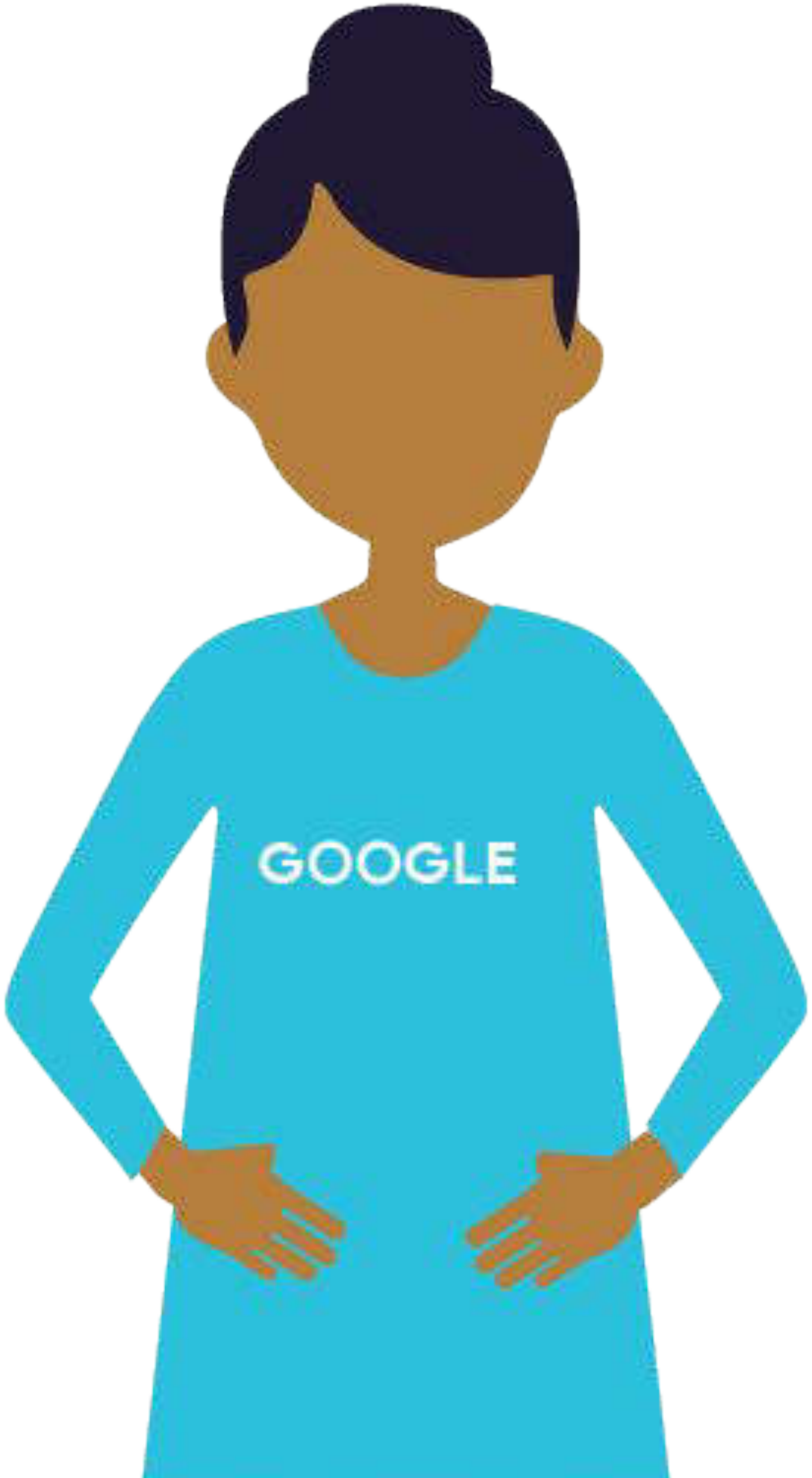


17	7904807859681747547	3	42	3	1162.0	null	null	1.0	2016-08-05	2017-07-09	338	??????????????????
18	4405445121320750966	51	358	62	517.36	null	null	1.0	2016-08-08	2017-07-12	338	??????????????????
19	1419607020881916790	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	337	??????????????????
20	3862335714593915688	13	92	16	154.23	238000000	1	2.0	2016-08-09	2017-07-12	337	??????????????????

Infer or predict it with a model! ⟶

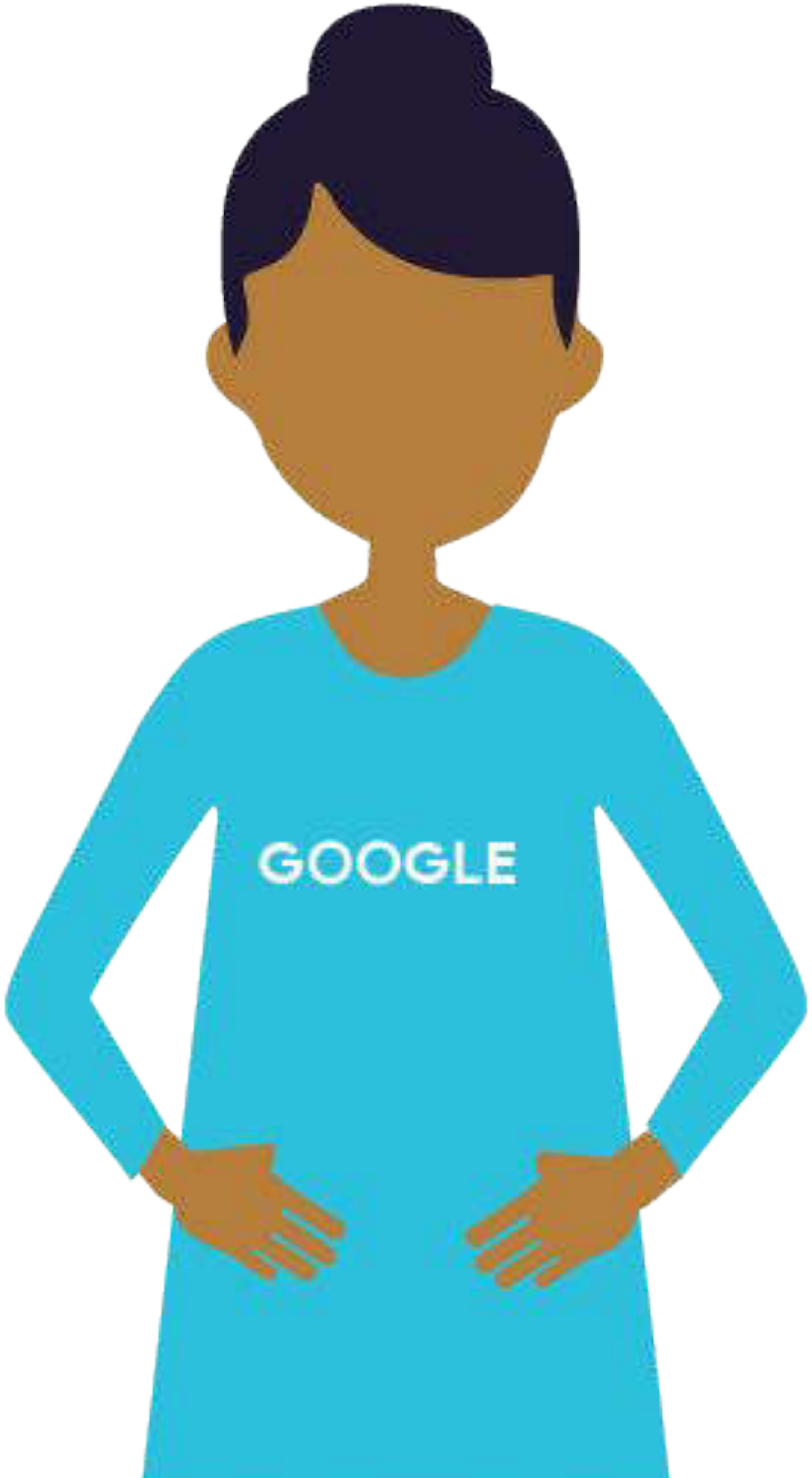


Choose the right model
for your use case



Choose the right model for your use case

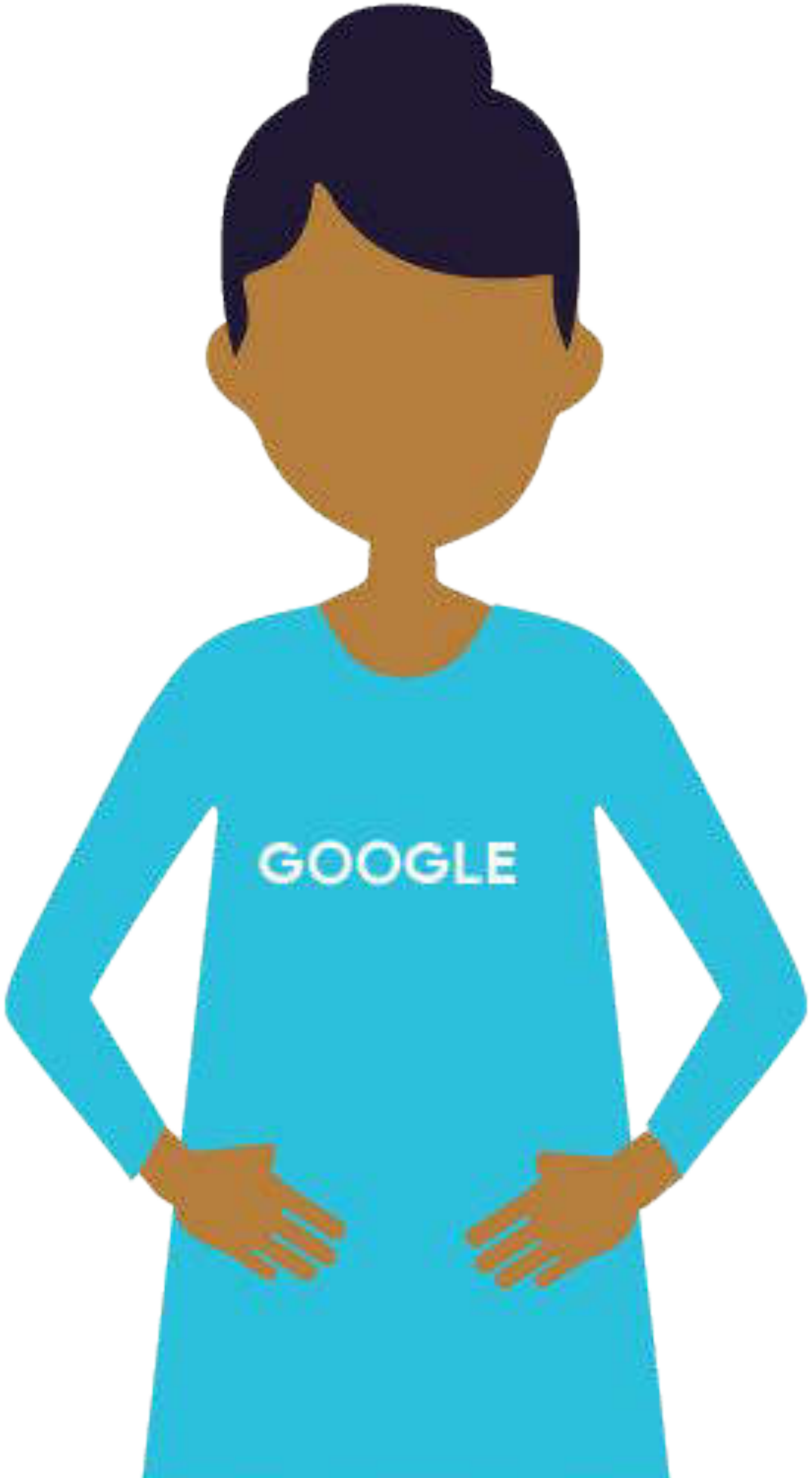
Forecasting a number?
Try linear regression



Choose the right model for your use case

Classifying a label?
Try logistic regression

(among many more)



Last Note:

Supervised vs
Unsupervised Learning

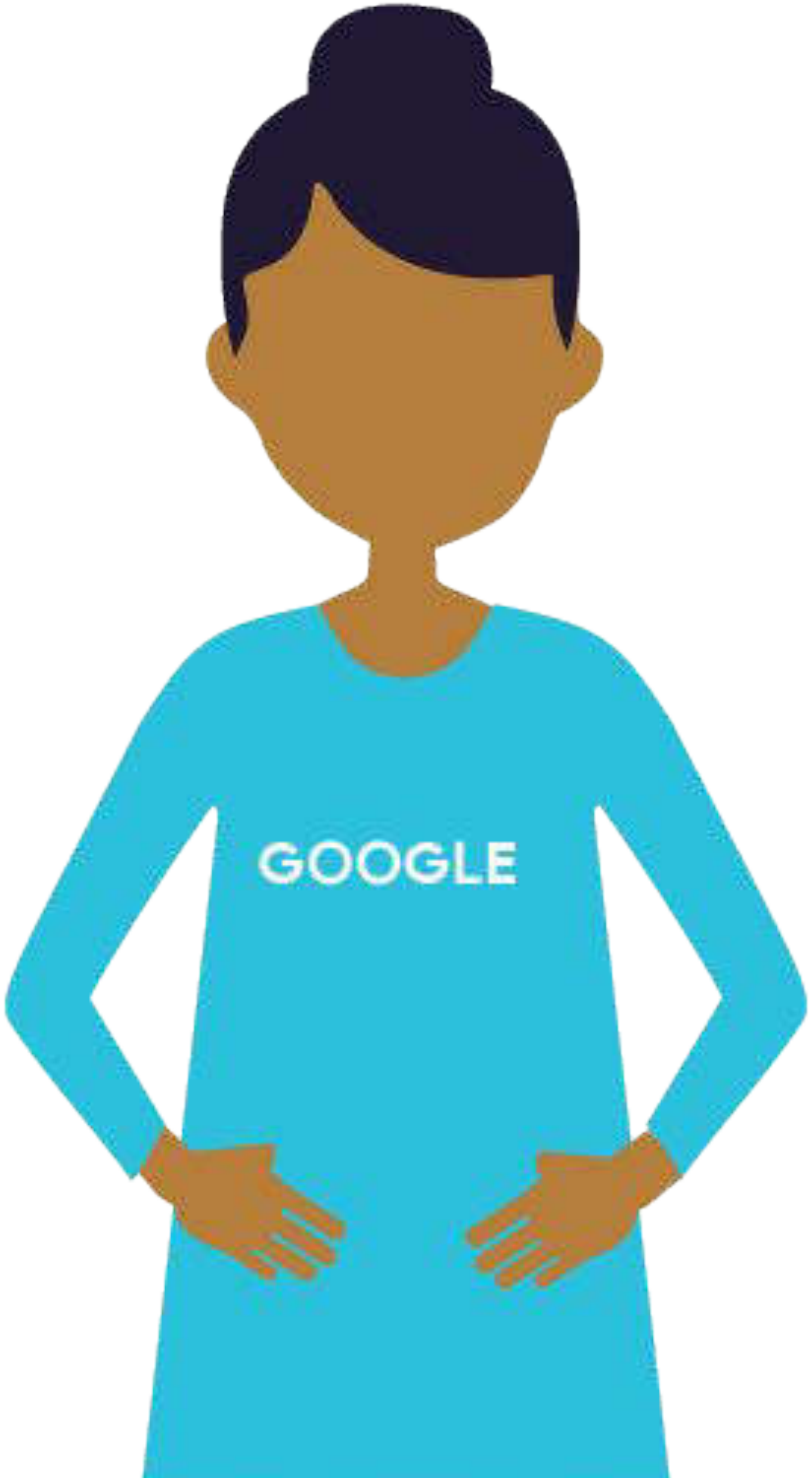
Course 4: Applying Machine Learning to your Datasets

Module 1: Introduction to Machine Learning

Lesson Title: **The 3 Secrets of ML**

Format: Talking Head

Video Name: T-BQML-O_1_l6_the_3_secrets_of_ml

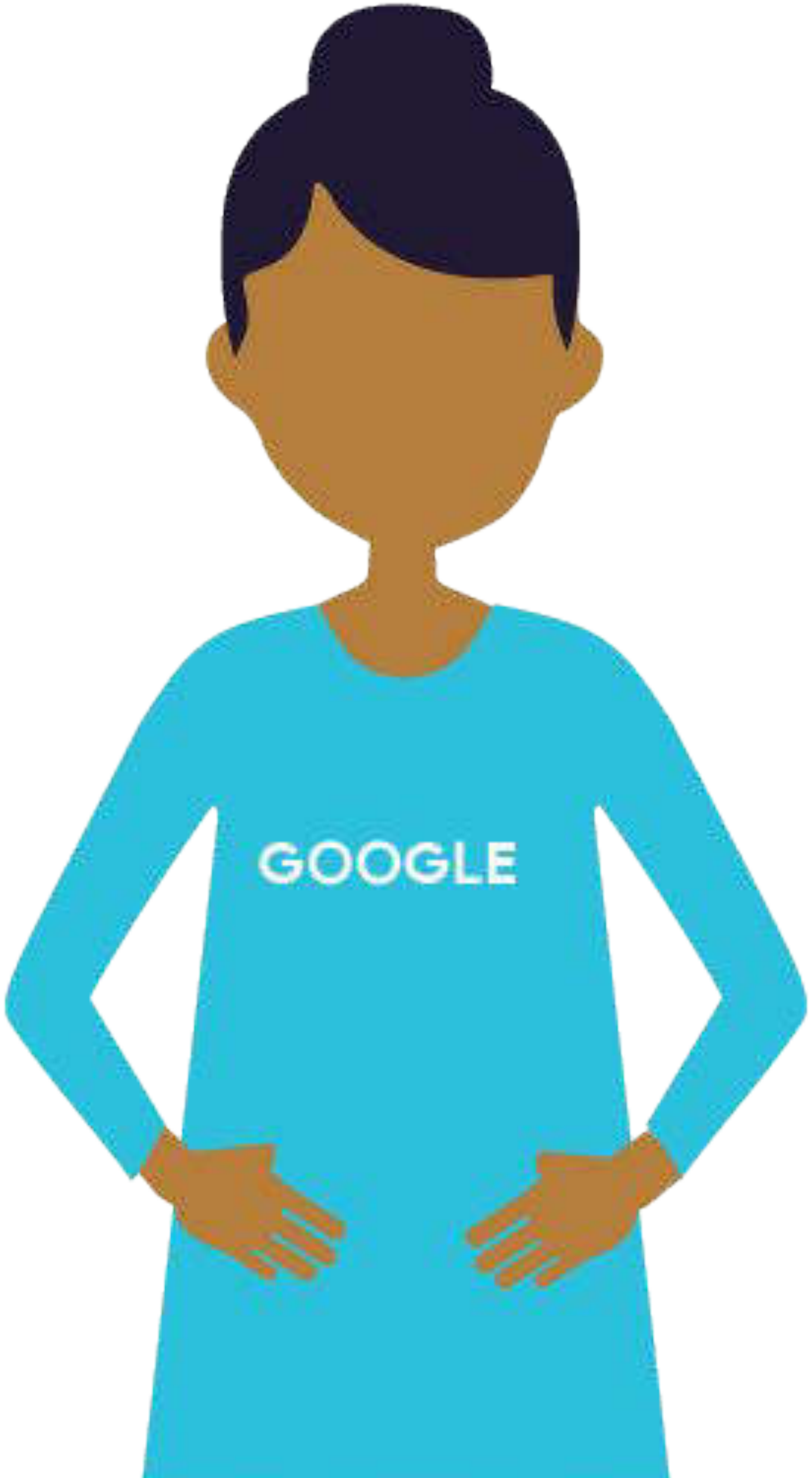


The 3 Secrets of ML

1. You don't have to set out to do an ML project

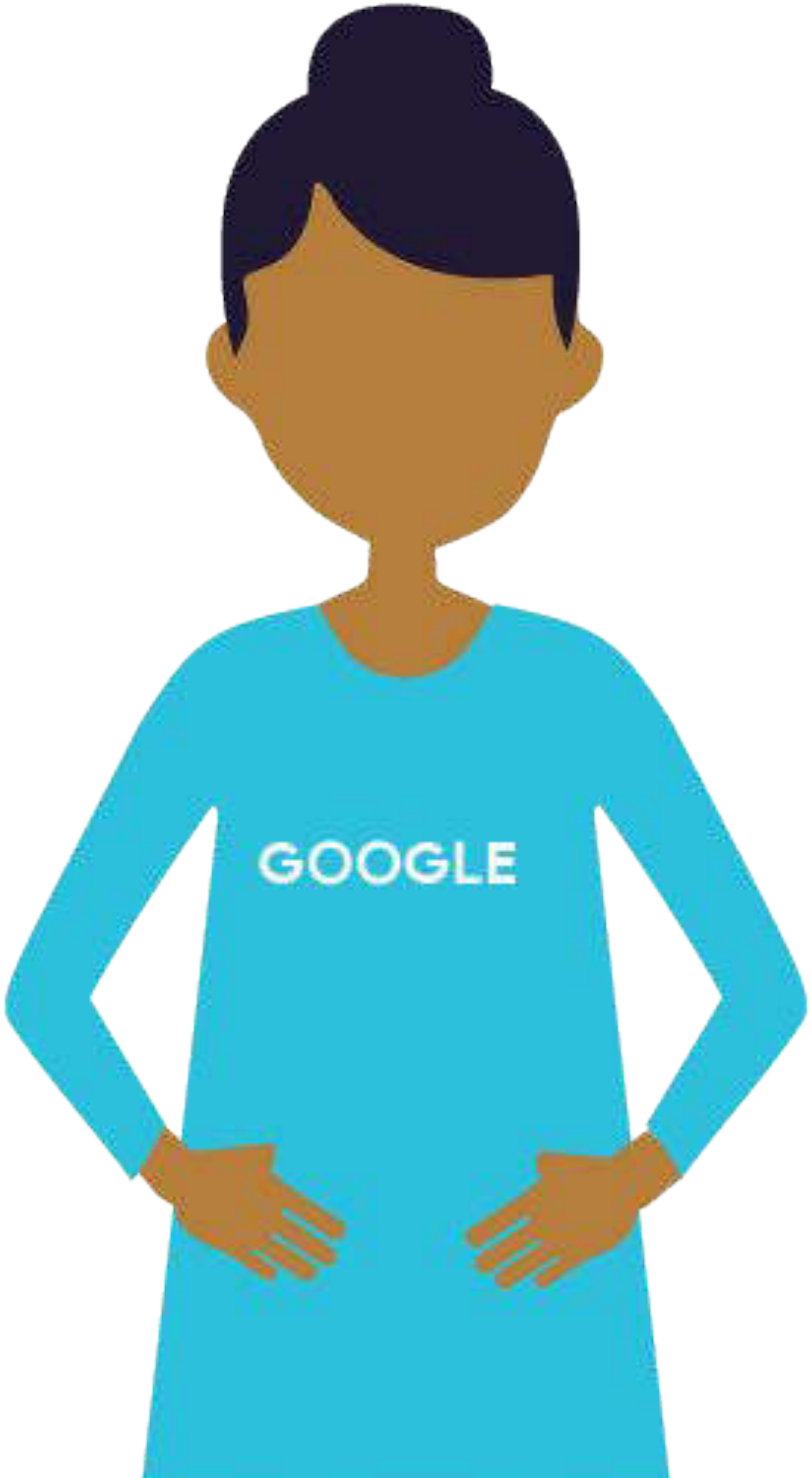
2. It's not just about training models

3. You need lots of good examples to train from*



The 3 Secrets of ML

1. **You don’t have to set out to do an ML project**
2. It’s not just about training models
3. You need lots of good examples to train from*



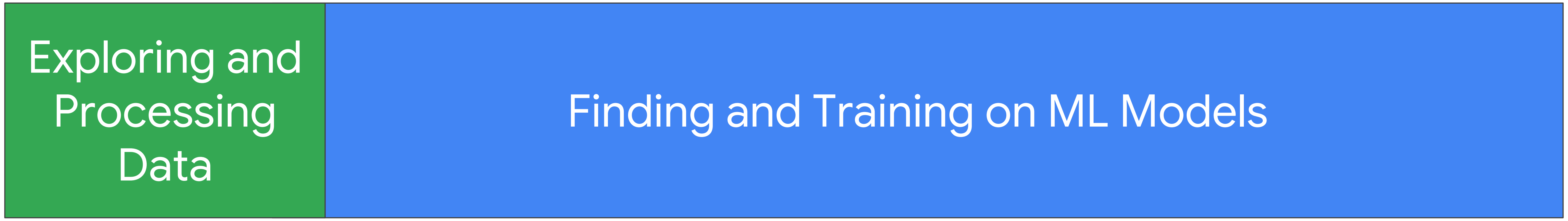
The 3 Secrets of ML

1. You don’t have to set out to do an ML project

2. **It’s not just about training models**

3. You need lots of good examples to train from*

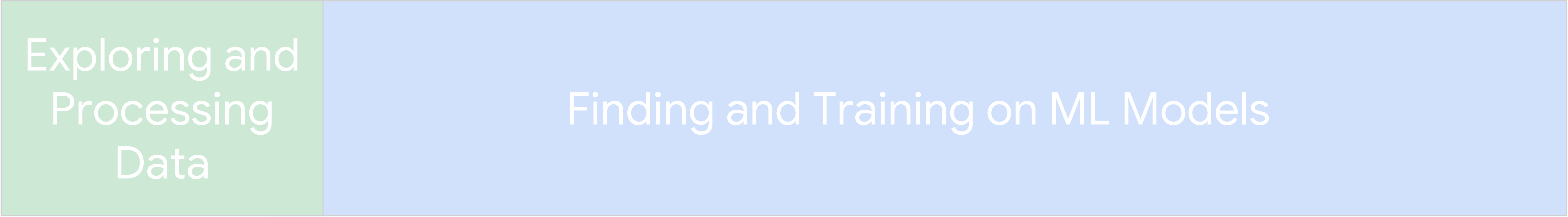
Expectation (your time spent)



Exploring and Processing Data

Finding and Training on ML Models

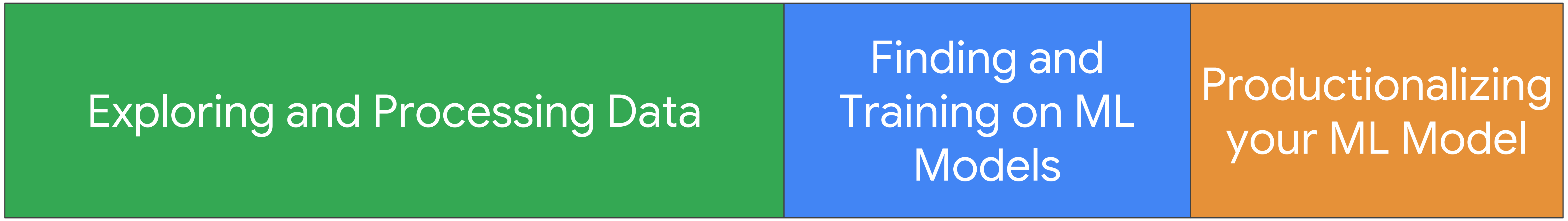
Expectation (your time spent)



Exploring and Processing Data

Finding and Training on ML Models

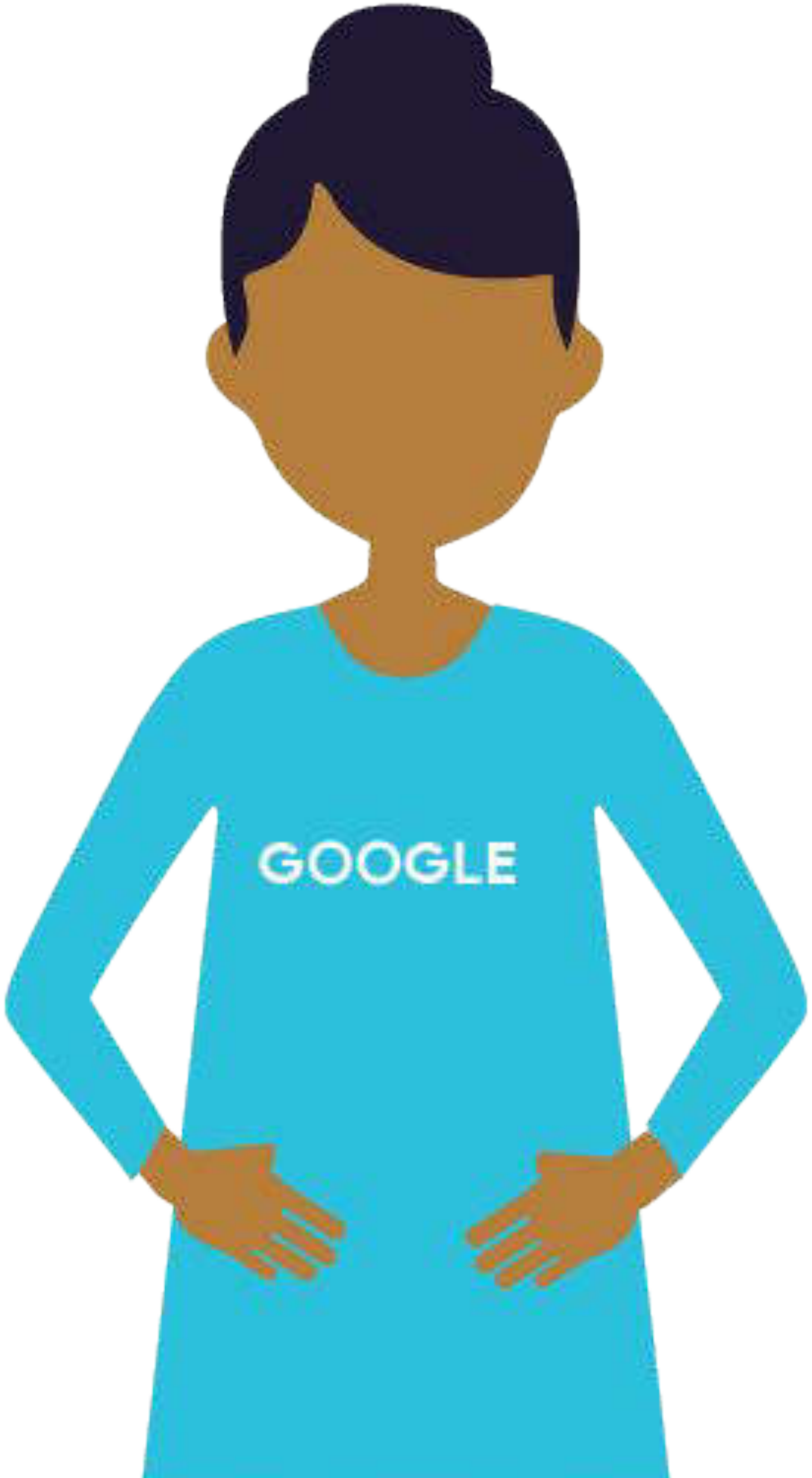
Reality



Exploring and Processing Data

Finding and Training on ML Models

Productionalizing your ML Model



The 3 Secrets of ML

1. You don’t have to set out to do an ML project
2. It’s not just about training models
3. **You need lots of good examples to train from***

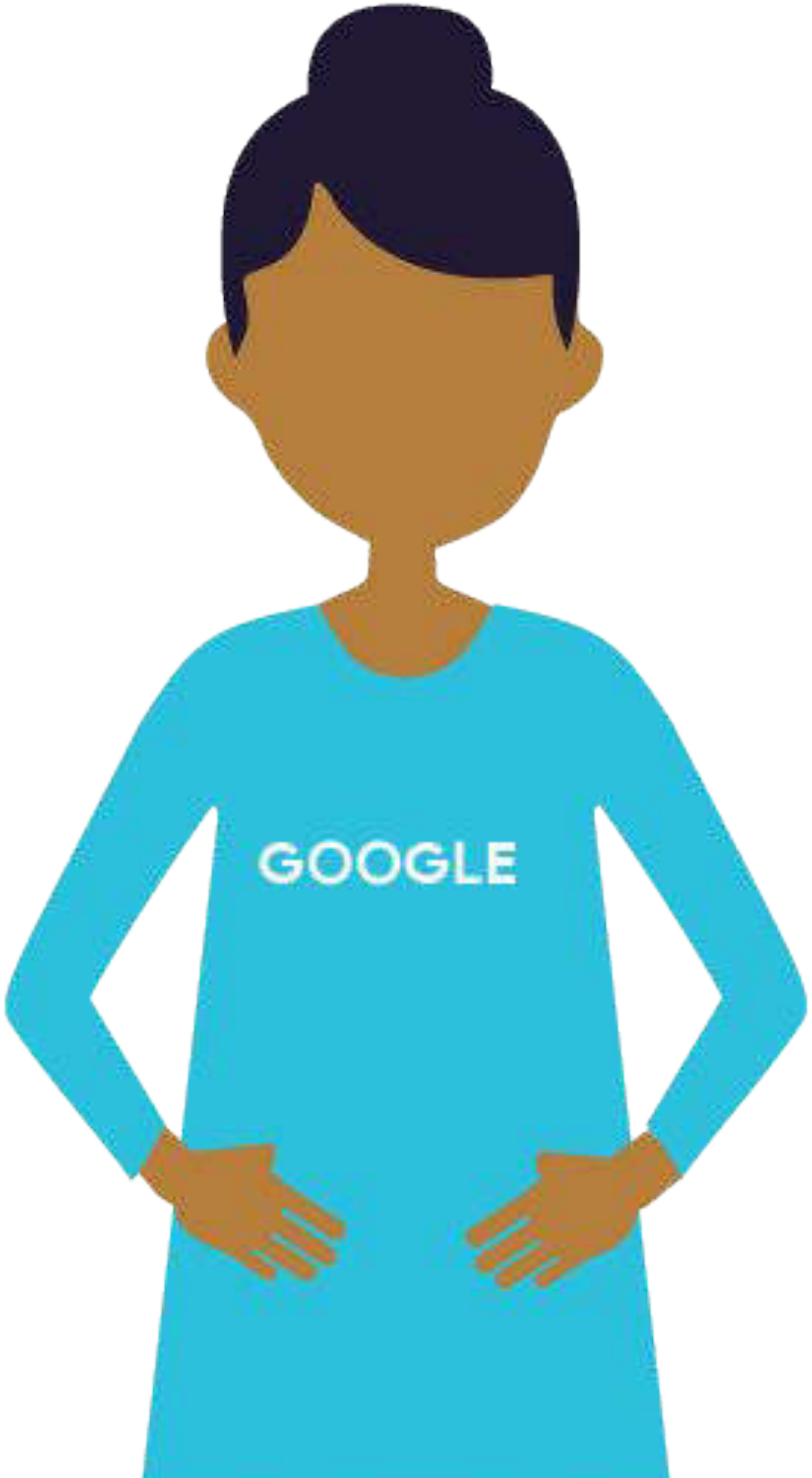
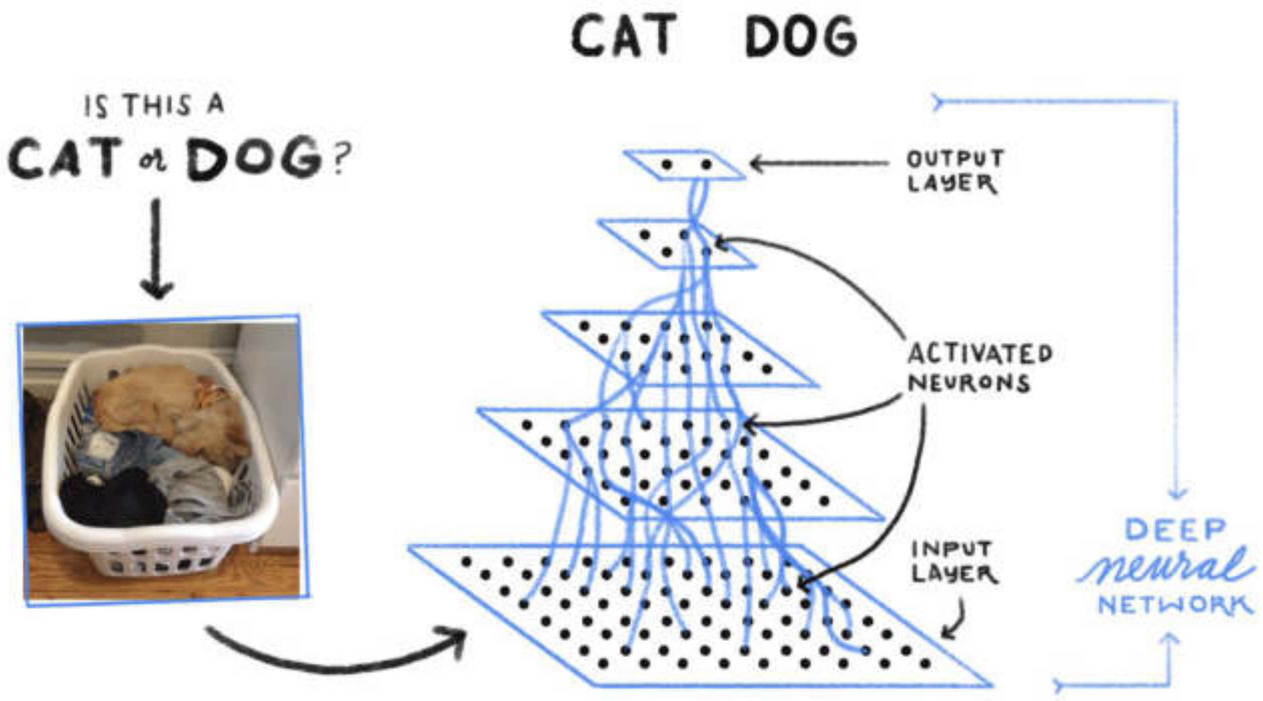
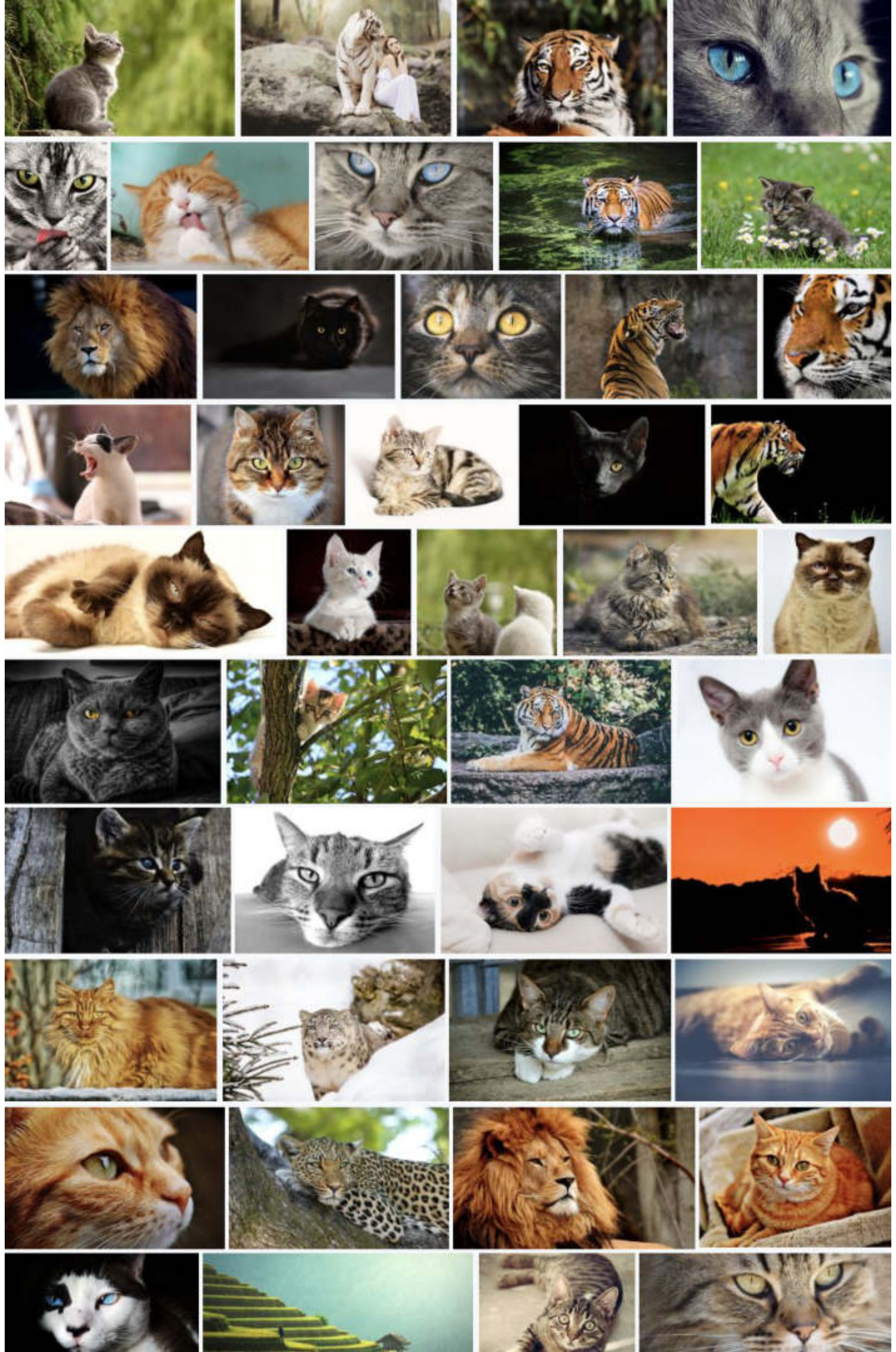

GOOGLE

Image Classification Model (Neural Network)





United States

Course 4: Applying Machine Learning to your Datasets

Module 2: Machine Learning Tool Options on GCP

Lesson Title: **The ML Tool Spectrum**

Format: Talking Head

Video Name: T-BQML-O_2_l1_the_ml_tool_spectrum

Machine Learning is a continually evolving field

The GCP Machine Learning Tool Spectrum

Advanced Models	Modeling for Analysts	Pretrained Models	Minimal Effort
TensorFlow	**ML on BigQuery (beta)**	**Pretrained ML APIs**	**AutoML (soon)**
● Data Scientists ● Data Engineers	● Data Analysts	● Data Analysts ● Data Scientists ● Data Engineers	● Everyone

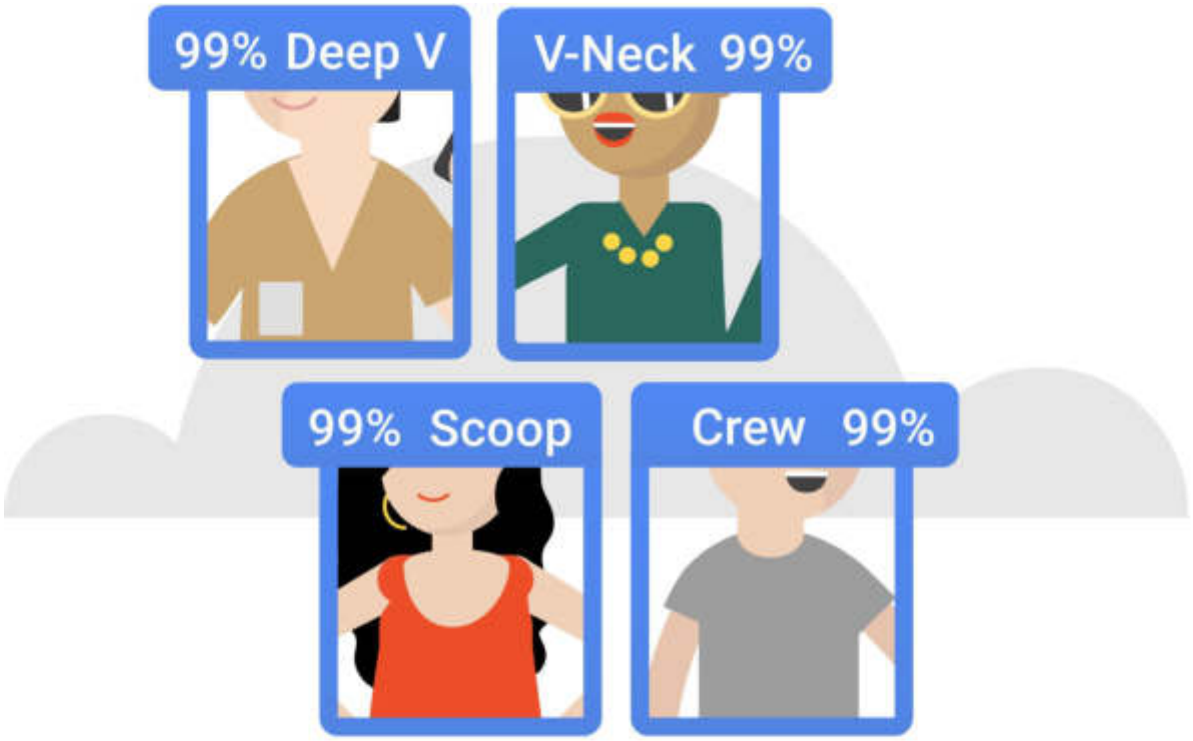
Create custom ML models with TensorFlow

Train and run ML in the familiar BigQuery UI

BigQuery

Access Pretrained ML APIs for common applications

Train and run ML with minimal effort



Examples of real-world ML tool use

Custom image model to price cars

AUCNET

Build off NLP API to route customer emails

ocado

Use Vision API as-is to find text in memes

GIPHY

Use Dialogflow to create a new shopping experience

UNIQLO

Course 4: Applying Machine Learning to your Datasets

Module 3: Pre-trained ML APIs

Lesson Title: **Overview**

Format: Talking Head

Video Name: T-BQML-O_3_l1_overview

Don’t Reinvent the ML/Distributed Computing Wheel

Cloud Vision API

Cloud Speech API

Cloud Jobs API

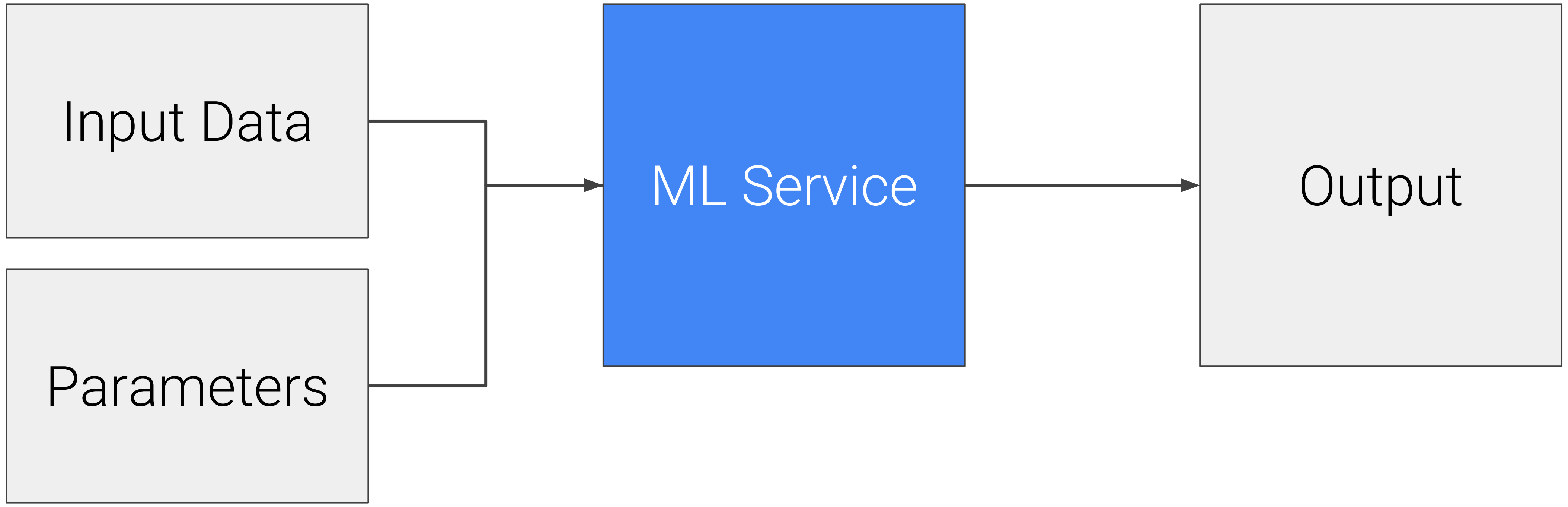
Cloud Translation API

Cloud Natural Language API

Cloud Video Intelligence

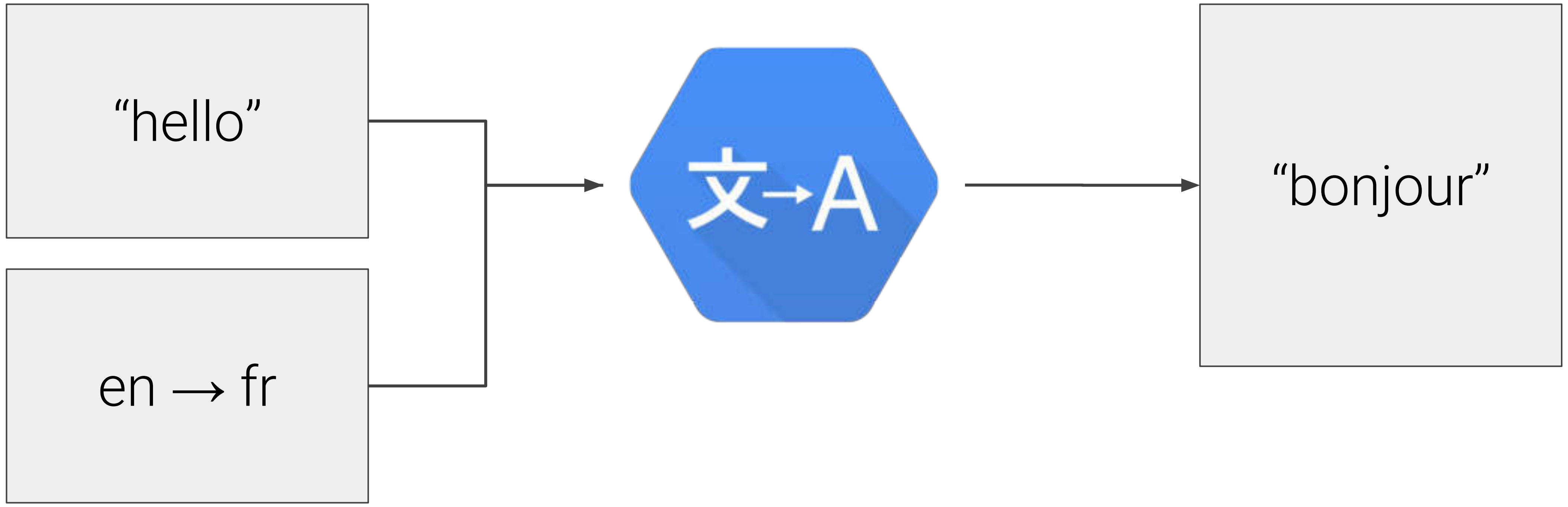
The ML APIs are microservices that provide a high level of abstraction

When we build ML models ourselves, it should be our goal to make them as easy to use and stand-alone.



Pass data values and parameters into the API

The Google Translate API expects certain values and will output the result



“hello”

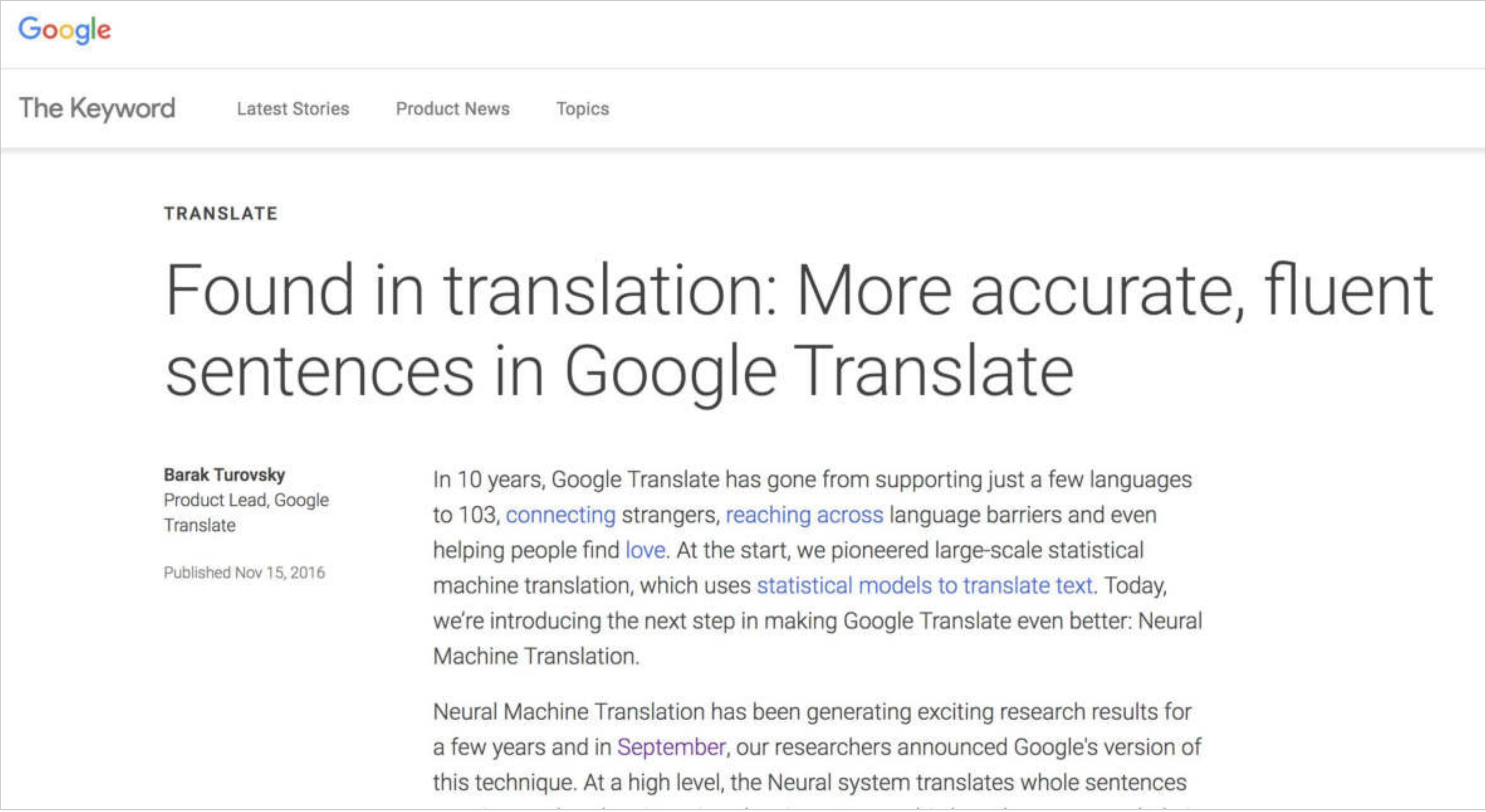
en → fr

“bonjour”

Demo
Machine Language Translation

Language Translation leaps forward with ML

In 2016, Google Translate adopts more deep neural networks which allows for more natural-sounding translations



TRANSLATE

Found in translation: More accurate, fluent sentences in Google Translate

Barak Turovsky
Product Lead, Google Translate

Published Nov 15, 2016

In 10 years, Google Translate has gone from supporting just a few languages to 103, connecting strangers, reaching across language barriers and even helping people find love. At the start, we pioneered large-scale statistical machine translation, which uses statistical models to translate text. Today, we're introducing the next step in making Google Translate even better: Neural Machine Translation.

Neural Machine Translation has been generating exciting research results for a few years and in September, our researchers announced Google's version of this technique. At a high level, the Neural system translates whole sentences

blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/

Google Cloud

Course 4: Applying Machine Learning to your Datasets

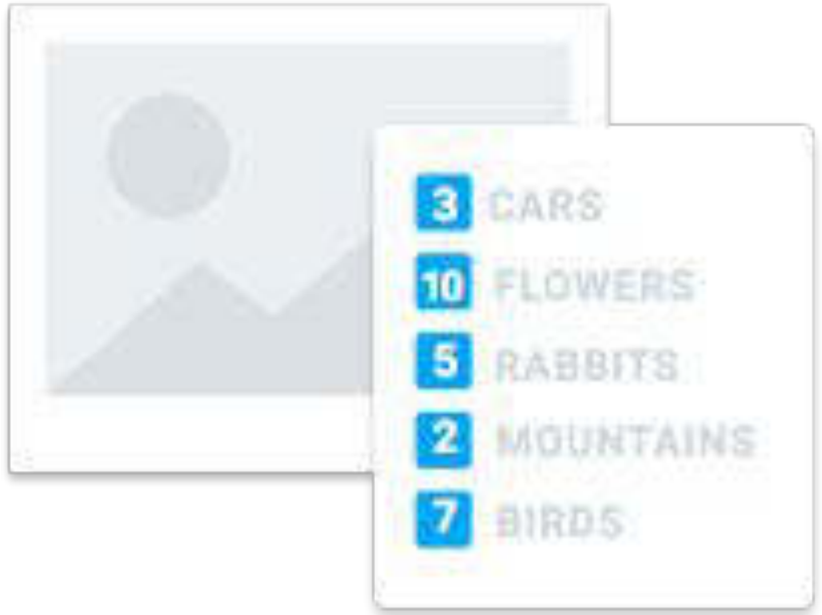
Module 3: Pre-trained ML APIs

Lesson Title: **Cloud Vision API**

Format: Talking Head

Video Name: T-BQML-O_3_l2_cloud_vision_api

Use the Cloud Vision API to understand image content



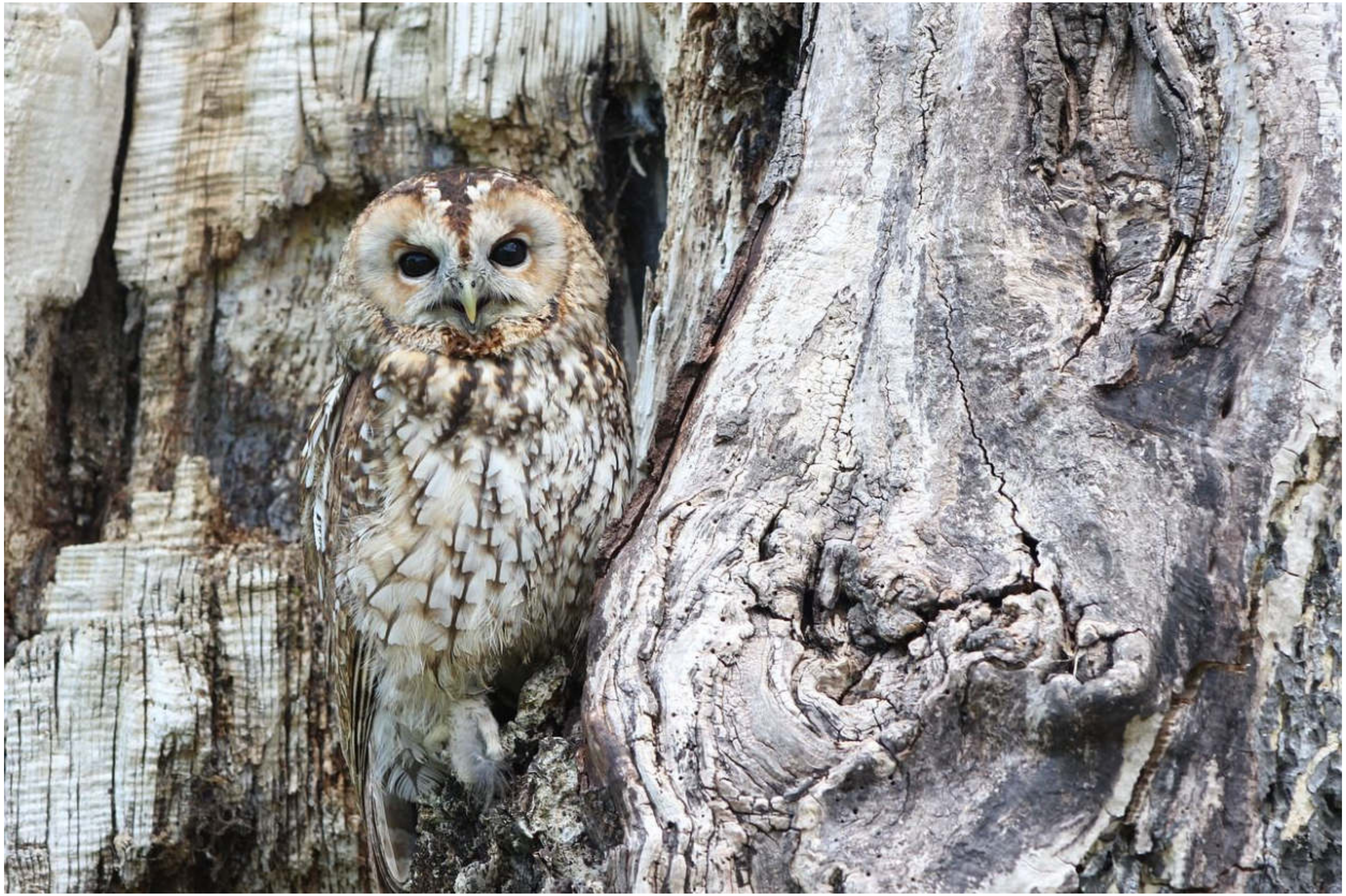
Detect and Label

Extract Text

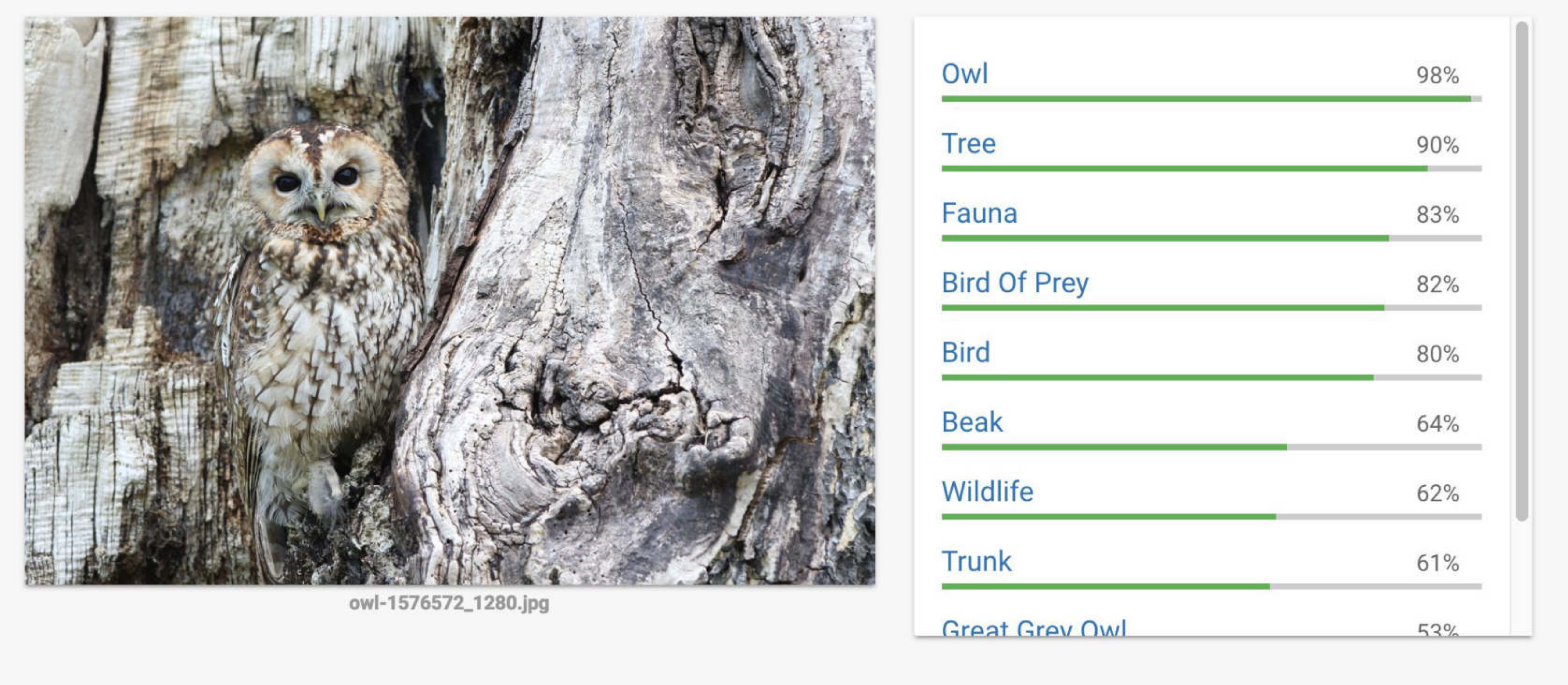
Identify Entities

Demo: Cloud Vision API

Let's see how well the ML API recognizes this owl



Demo: Cloud Vision API

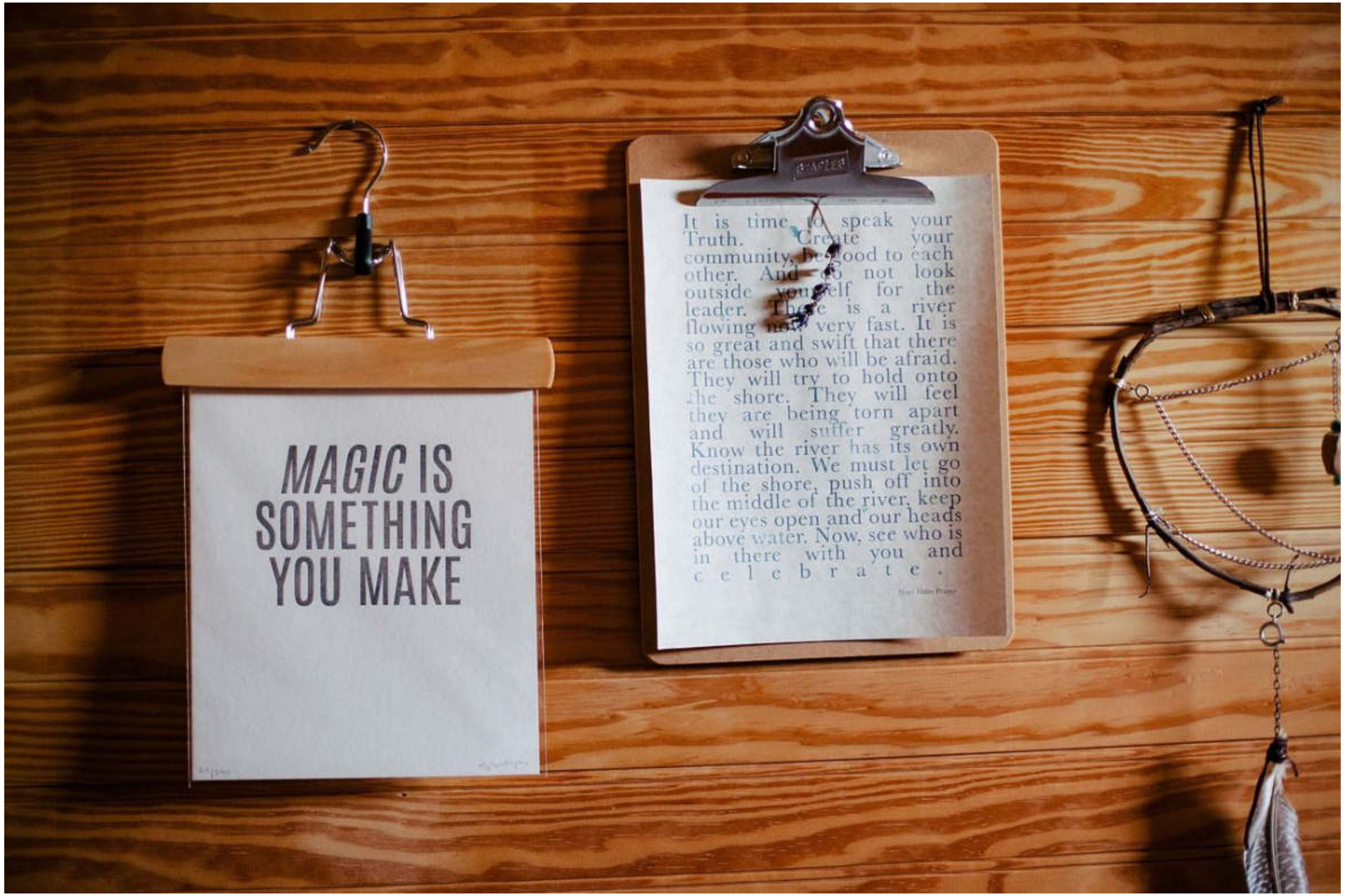


owl-1576572_1280.jpg

Owl	98%
Tree	90%
Fauna	83%
Bird Of Prey	82%
Bird	80%
Beak	64%
Wildlife	62%
Trunk	61%
Great Grey Owl	53%

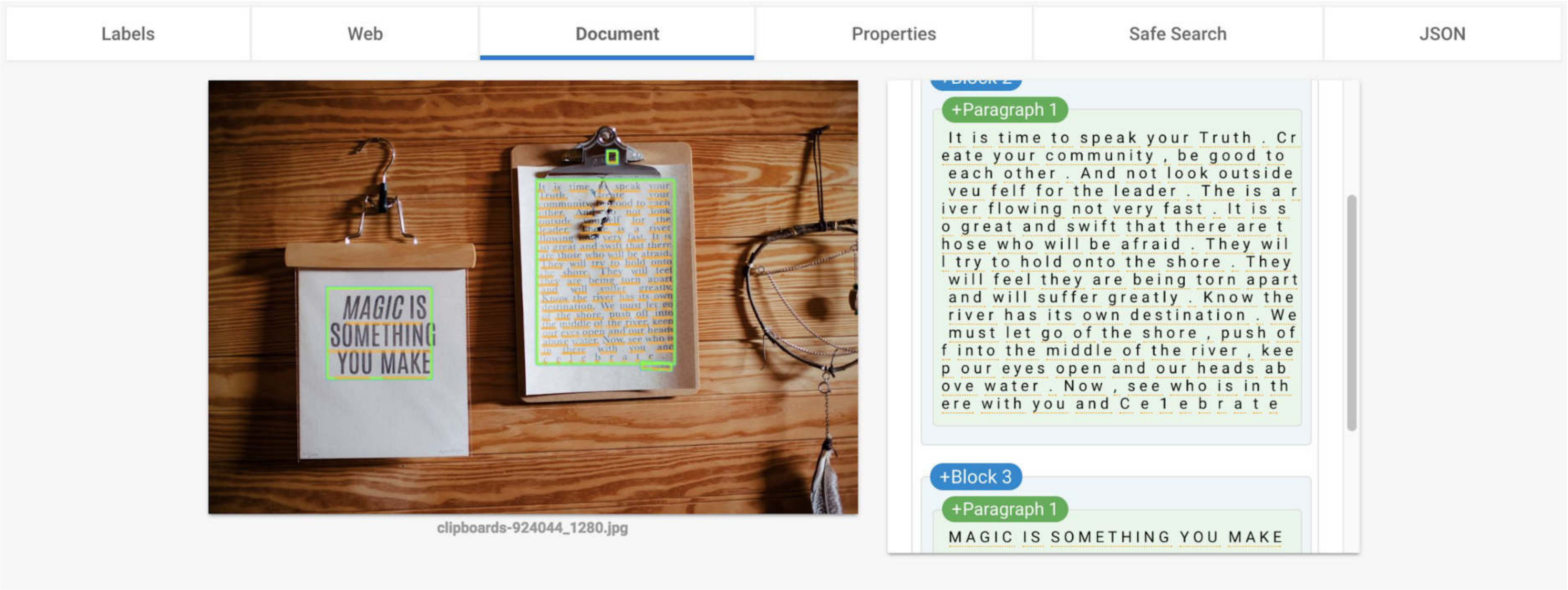
Demo: Cloud Vision API

what about embedded text?



It is time to speak your Truth. Create your community, be good to each other. And do not look outside yourself for the leader. There is a river flowing now very fast. It is so great and swift that there are those who will be afraid. They will try to hold onto the shore. They will feel they are being torn apart and will suffer greatly. Know the river has its own destination. We must let go of the shore, push off into the middle of the river, keep our eyes open and our heads above water. Now, see who is in there with you and celebrate.

Demo: Cloud Vision API

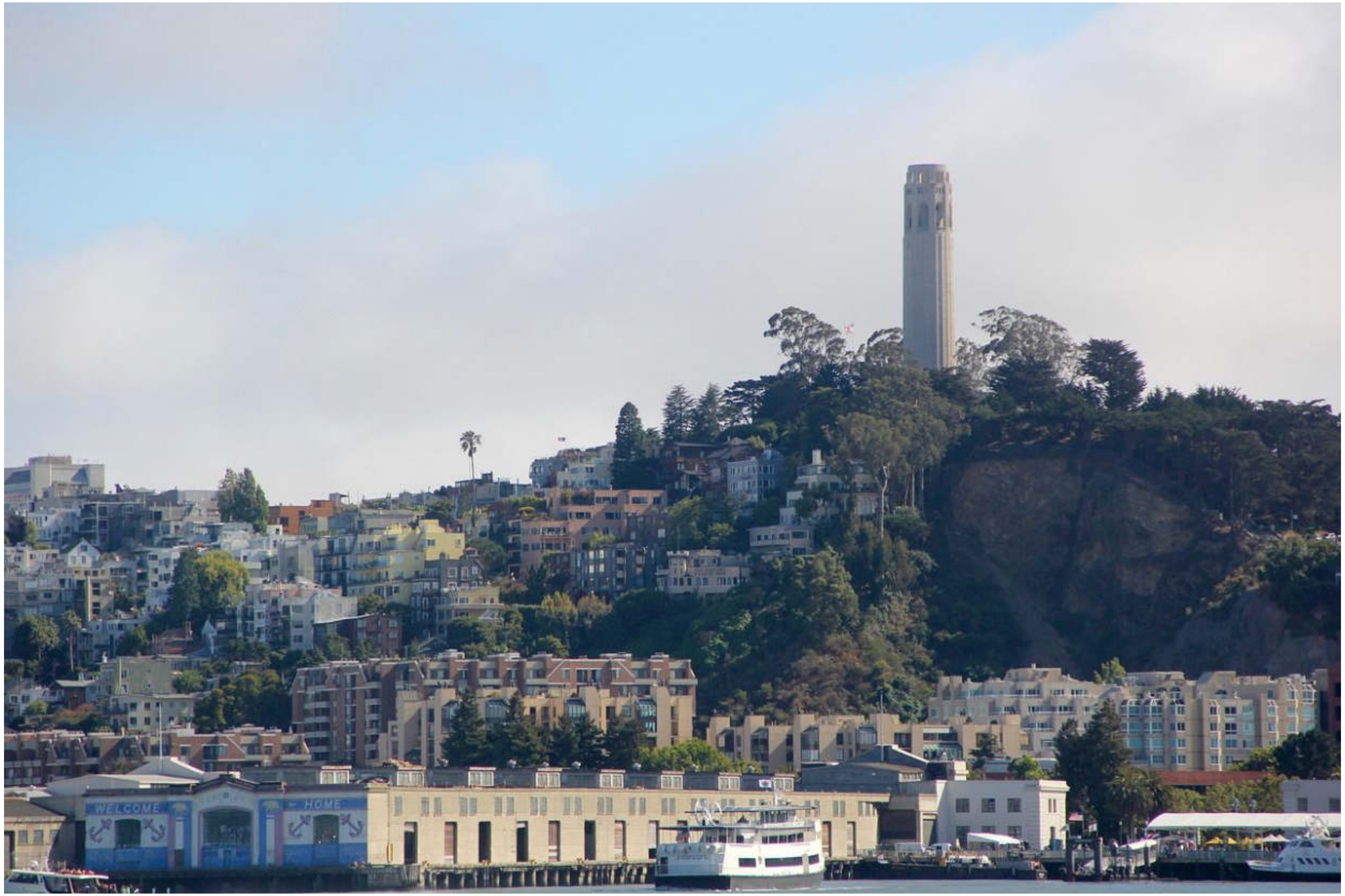


```
It is time to speak your Truth . Cr
eate your community , be good to
each other . And not look outside
veu felf for the leader . The is a r
iver flowing not very fast . It is s
o great and swift that there are t
hose who will be afraid . They wil
l try to hold onto the shore . They
will feel they are being torn apart
and will suffer greatly . Know the
river has its own destination . We
must let go of the shore , push of
f into the middle of the river , kee
p our eyes open and our heads ab
ove water . Now , see who is in th
ere with you and C e 1 e b r a t e
```

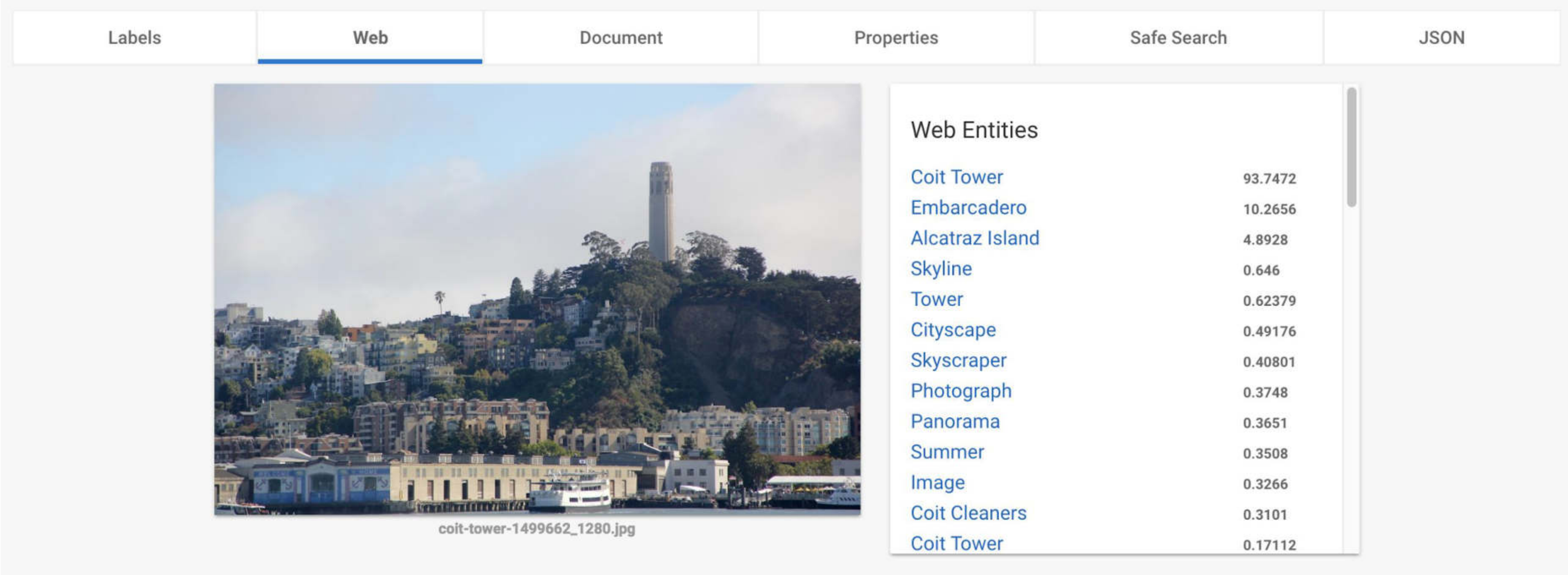
```
MAGIC IS SOMETHING YOU MAKE
```

Demo: Cloud Vision API

What about known entities like Coit Tower in San Francisco?



Demo: Cloud Vision API

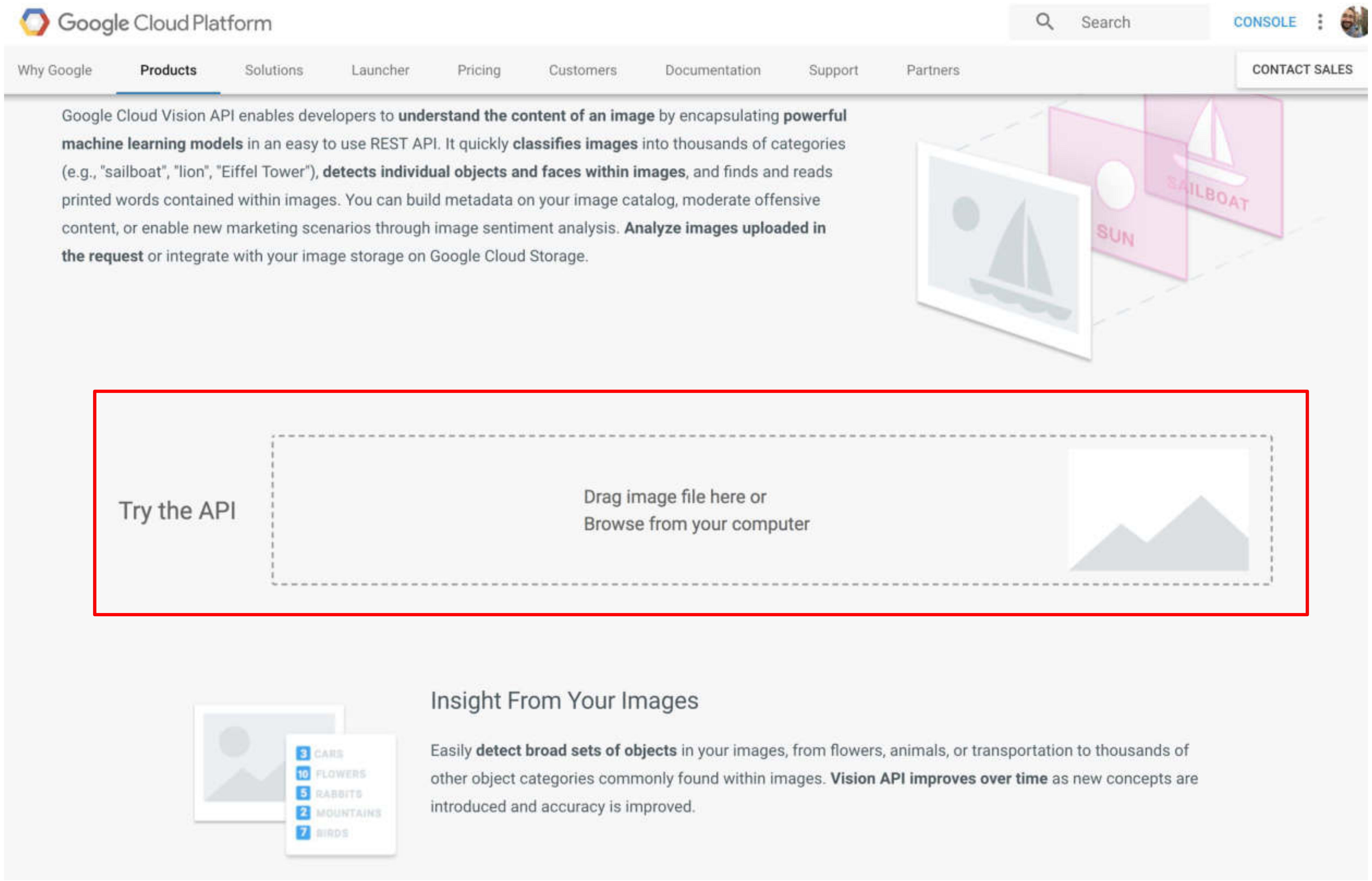


coit-tower-1499662_1280.jpg

Web Entities

Coit Tower	93.7472
Embarcadero	10.2656
Alcatraz Island	4.8928
Skyline	0.646
Tower	0.62379
Cityscape	0.49176
Skyscraper	0.40801
Photograph	0.3748
Panorama	0.3651
Summer	0.3508
Image	0.3266
Coit Cleaners	0.3101
Coit Tower	0.17112

Your turn: https://cloud.google.com/vision/



Google Cloud Vision API enables developers to **understand the content of an image** by encapsulating **powerful machine learning models** in an easy to use REST API. It quickly **classifies images** into thousands of categories (e.g., "sailboat", "lion", "Eiffel Tower"), **detects individual objects and faces within images**, and finds and reads printed words contained within images. You can build metadata on your image catalog, moderate offensive content, or enable new marketing scenarios through image sentiment analysis. **Analyze images uploaded in the request** or integrate with your image storage on Google Cloud Storage.

Try the API

Drag image file here or
Browse from your computer

Insight From Your Images

Easily **detect broad sets of objects** in your images, from flowers, animals, or transportation to thousands of other object categories commonly found within images. **Vision API improves over time** as new concepts are introduced and accuracy is improved.

Google Cloud

Course 4: Applying Machine Learning to your Datasets

Module 3: Pre-trained ML APIs

Lesson Title: **Natural Language API**

Format: Talking Head

Video Name: T-BQML-O_3_l3_natural_language_api

Use the to Cloud NLP API understand and parse language

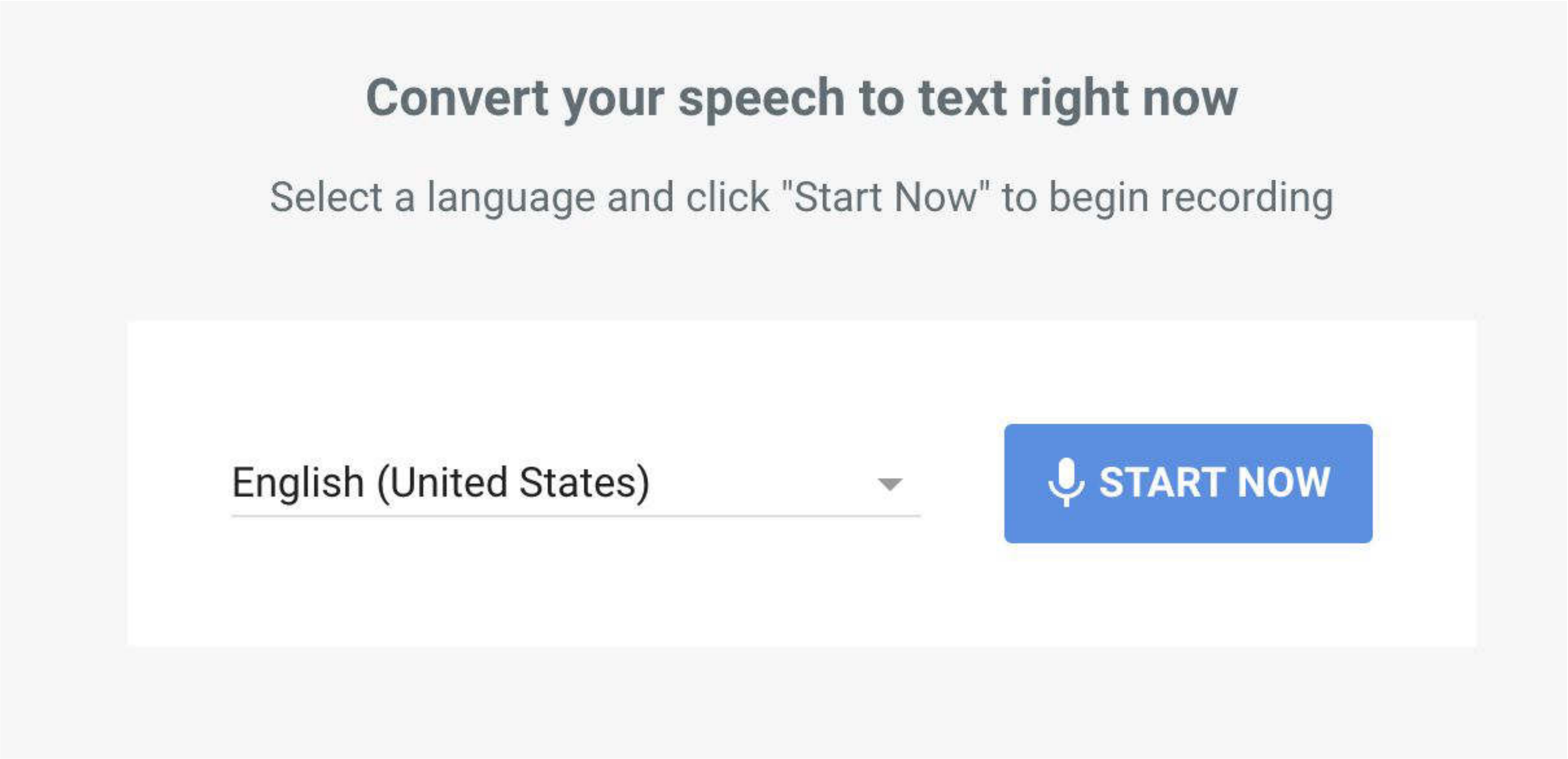
Speech
Recognition

Neural Machine
Translation

Identify Sentiment
and Entities

Demo: Cloud Speech API

Let's see how well the ML API understands us



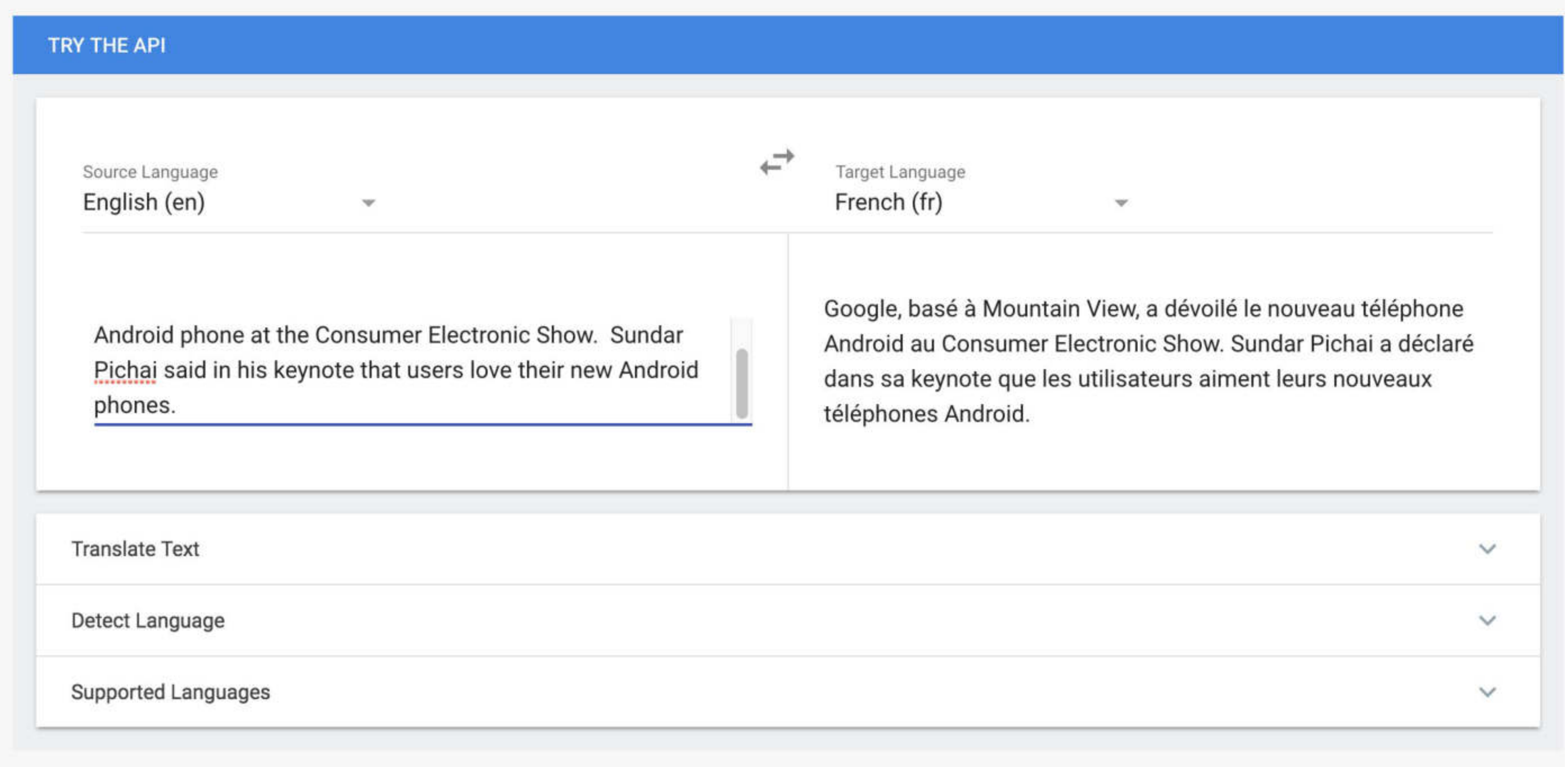
Convert your speech to text right now

Select a language and click "Start Now" to begin recording

English (United States)

Demo: Google Translate API

Time to translate



Source Language: English (en)

Target Language: French (fr)

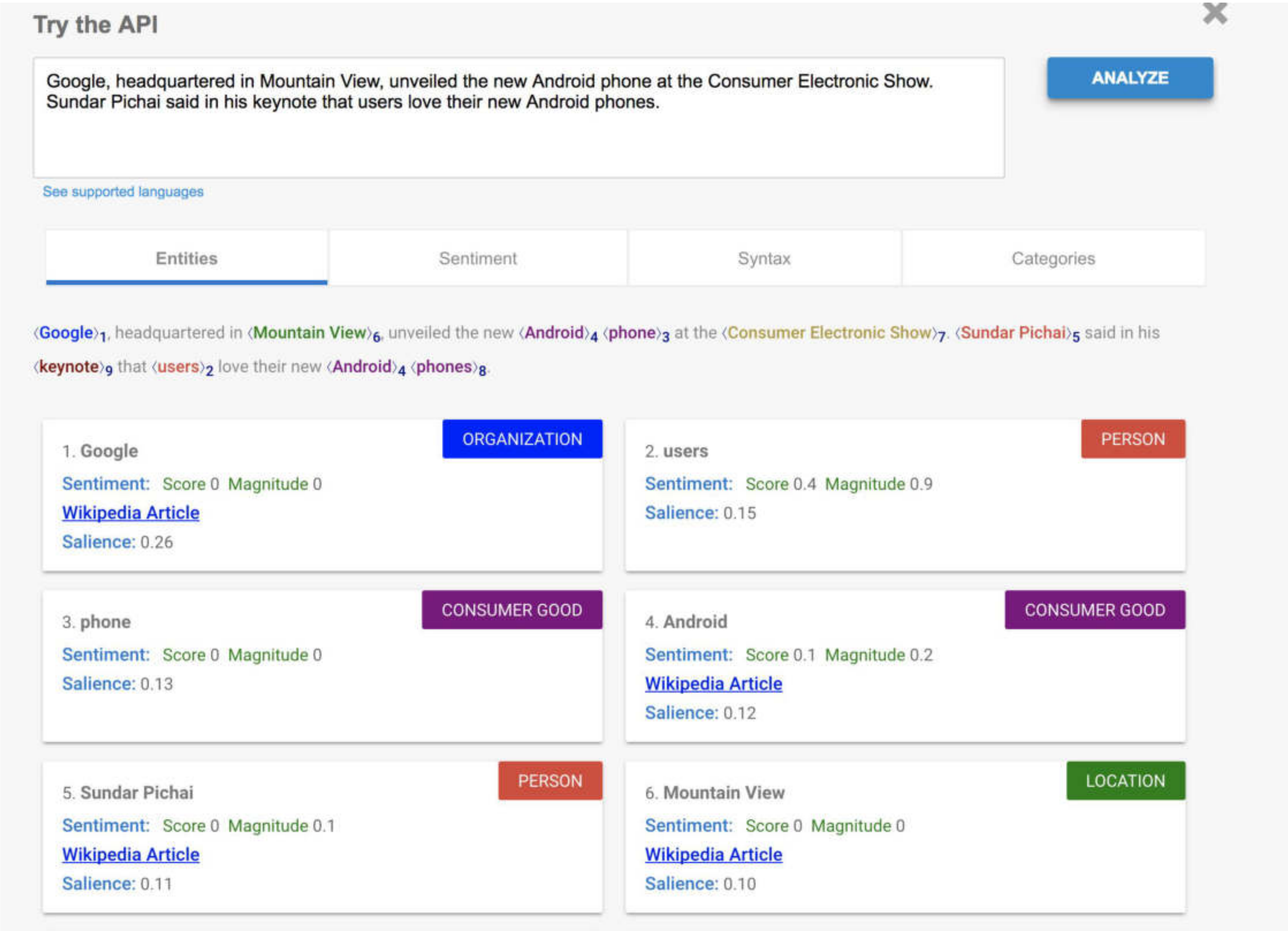
Android phone at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

Google, basé à Mountain View, a dévoilé le nouveau téléphone Android au Consumer Electronic Show. Sundar Pichai a déclaré dans sa keynote que les utilisateurs aiment leurs nouveaux téléphones Android.

Google Cloud

Demo: Cloud Natural Language Processing API

what entities are recognized in our text?



Google, headquartered in Mountain View, unveiled the new Android phone at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

⟨Google⟩$_1$, headquartered in ⟨Mountain View⟩$_6$, unveiled the new ⟨Android⟩$_4$ ⟨phone⟩$_3$ at the ⟨Consumer Electronic Show⟩$_7$. ⟨Sundar Pichai⟩$_5$ said in his ⟨keynote⟩$_9$ that ⟨users⟩$_2$ love their new ⟨Android⟩$_4$ ⟨phones⟩$_8$.

1. Google — ORGANIZATION
Sentiment: Score 0 Magnitude 0
Wikipedia Article
Salience: 0.26

2. users — PERSON
Sentiment: Score 0.4 Magnitude 0.9
Salience: 0.15

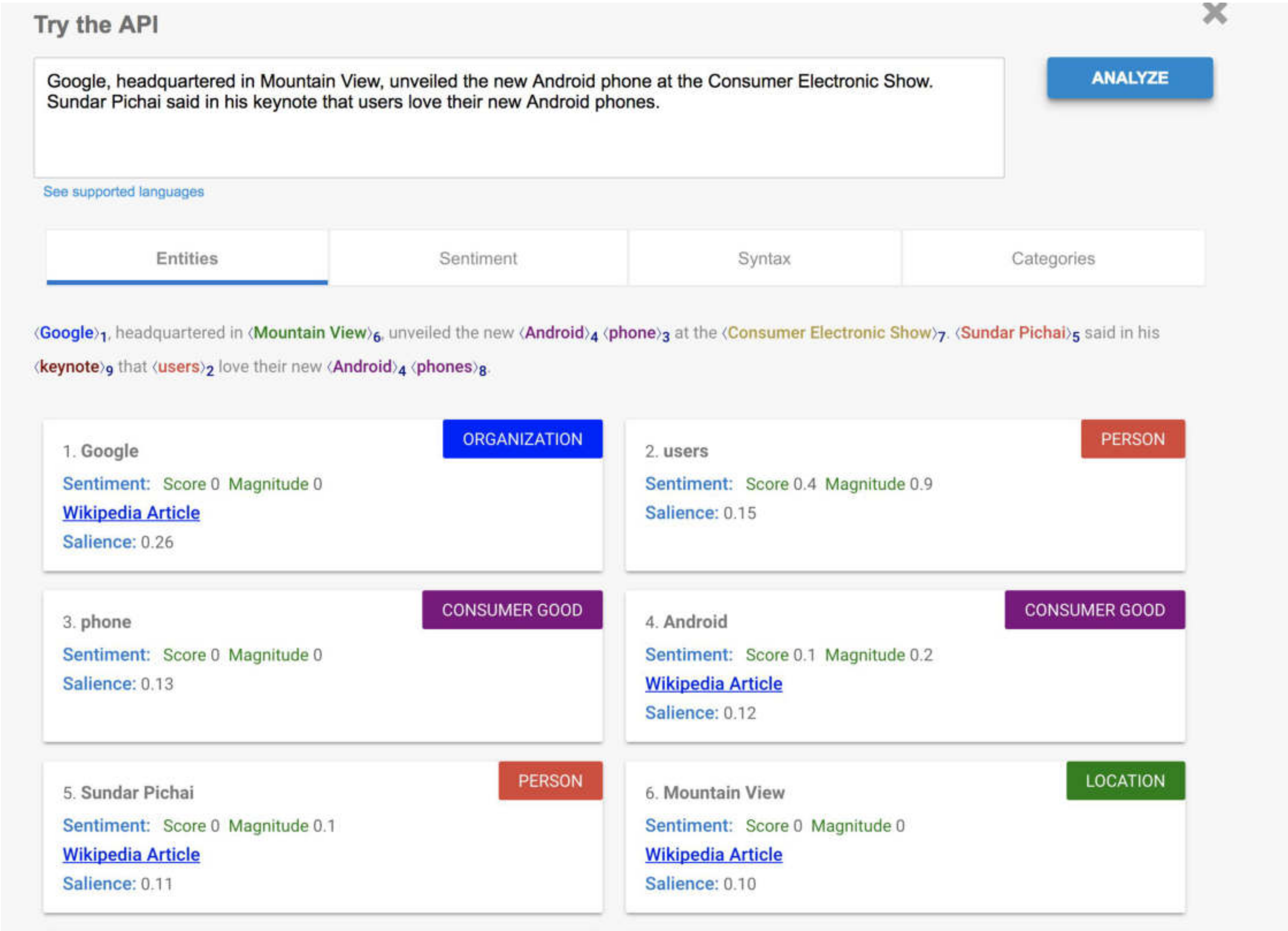
3. phone — CONSUMER GOOD
Sentiment: Score 0 Magnitude 0
Salience: 0.13

4. Android — CONSUMER GOOD
Sentiment: Score 0.1 Magnitude 0.2
Wikipedia Article
Salience: 0.12

5. Sundar Pichai — PERSON
Sentiment: Score 0 Magnitude 0.1
Wikipedia Article
Salience: 0.11

6. Mountain View — LOCATION
Sentiment: Score 0 Magnitude 0
Wikipedia Article
Salience: 0.10

Your Turn: https://cloud.google.com/natural-language/



Try the API

Google, headquartered in Mountain View, unveiled the new Android phone at the Consumer Electronic Show. Sundar Pichai said in his keynote that users love their new Android phones.

See supported languages

⟨Google⟩$_1$, headquartered in ⟨Mountain View⟩$_6$, unveiled the new ⟨Android⟩$_4$ ⟨phone⟩$_3$ at the ⟨Consumer Electronic Show⟩$_7$. ⟨Sundar Pichai⟩$_5$ said in his ⟨keynote⟩$_9$ that ⟨users⟩$_2$ love their new ⟨Android⟩$_4$ ⟨phones⟩$_8$.

1. Google — ORGANIZATION
Sentiment: Score 0 Magnitude 0
Wikipedia Article
Salience: 0.26

2. users — PERSON
Sentiment: Score 0.4 Magnitude 0.9
Salience: 0.15

3. phone — CONSUMER GOOD
Sentiment: Score 0 Magnitude 0
Salience: 0.13

4. Android — CONSUMER GOOD
Sentiment: Score 0.1 Magnitude 0.2
Wikipedia Article
Salience: 0.12

5. Sundar Pichai — PERSON
Sentiment: Score 0 Magnitude 0.1
Wikipedia Article
Salience: 0.11

6. Mountain View — LOCATION
Sentiment: Score 0 Magnitude 0
Wikipedia Article
Salience: 0.10

Course 4: Applying Machine Learning to your Datasets

Module 3: Pre-trained ML APIs

Lesson Title: **Lab: Pretrained ML APIs**

Format: Talking Head

Video Name: T-BQML-O_3_l4_lab_intro:_pretrained_ml_apis

LAB:

Pretrained ML APIs

hello
Hello
你好
3 CARS
10 FLOWERS
5 RABBITS
2 MOUNTAINS
7 BIRDS
Aa

Course 4: Applying Machine Learning to your Datasets

Module 3: Pre-trained ML APIs

Lesson Title: **Lab Solution: Pretrained ML APIs**

Format: Talking Head + Lab Screencast

Video Name: T-BQML-O_3_l6_lab_solution:_pretrained_ml_apis

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **What makes a dataset good for ML?**

Format: Talking Head

Video Name: T-BQML-O_4_l1_what_makes_a_dataset_good_for_ml

Building a ML Model involves:

Create the dataset

Build the model

Operationalize the model

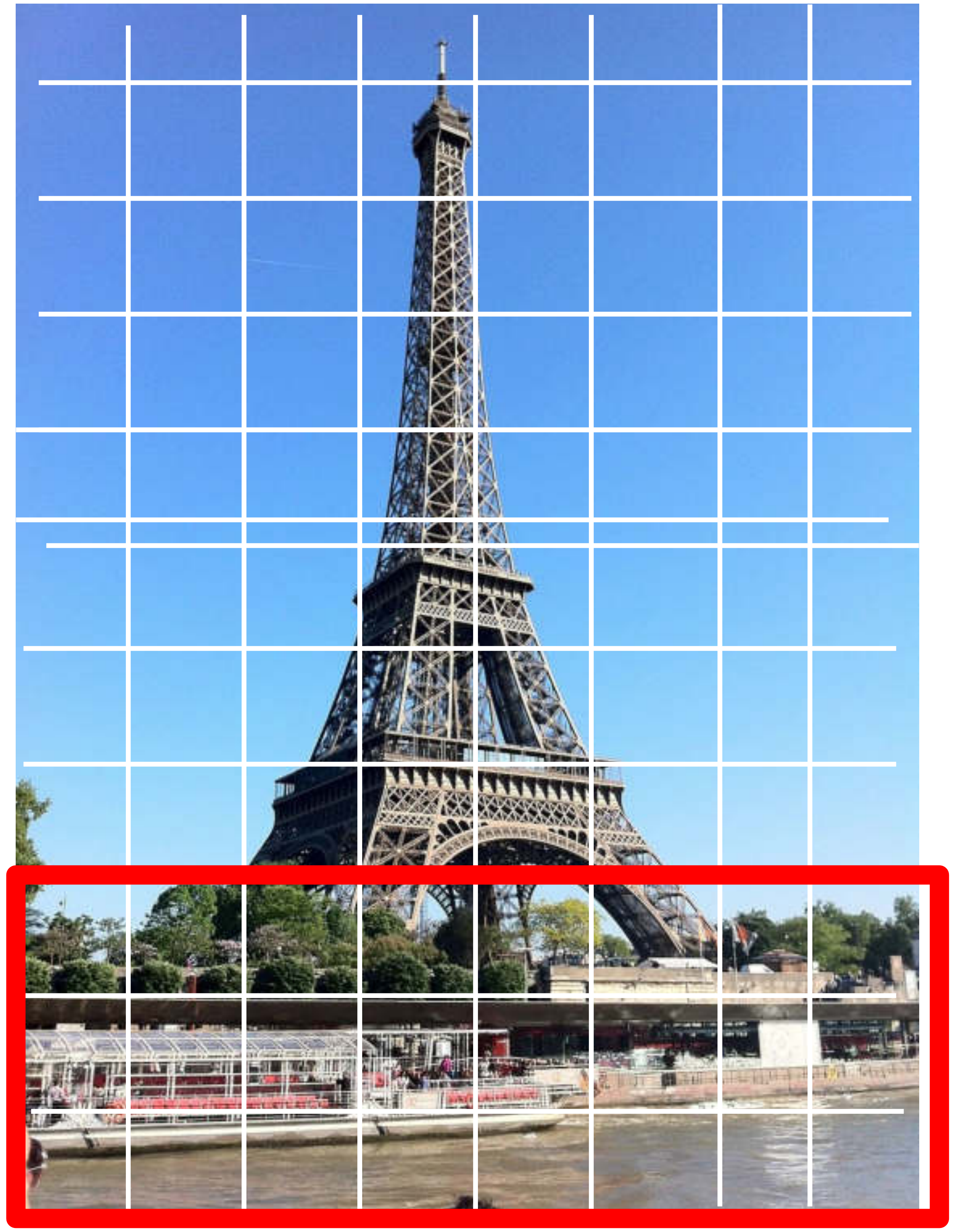
Building a ML Model involves:

Create the dataset

Build the model

Operationalize the model

Don’t assume datasets have high quality or complete data



Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Choosing Good Features**

Format: Talking Head

Video Name: T-BQML-O_4_l2_choosing_good_features

Good dataset feature columns must be:

✓ 1. Related to the objective

✓ 2. Known at prediction-time

✓ 3. Numeric with meaningful magnitude

✓ 4. Have enough examples

✓ 5. Bring human insight to problem

Good dataset feature columns must be:

✓ 1. Related to the objective

✓ 2. Known at prediction-time

✓ 3. Numeric with meaningful magnitude

✓ 4. Have enough examples

✓ 5. Bring human insight to problem

Choose the good features

A) Breed

B) Age

C) Eye Color

Objective: Good racehorse

✓ **A) Breed**

✓ **B) Age**

C) Eye Color

Objective: Eye disease

- ✓ A) Breed
- ✓ B) Age
- ✓ C) Eye Color

Good features are:

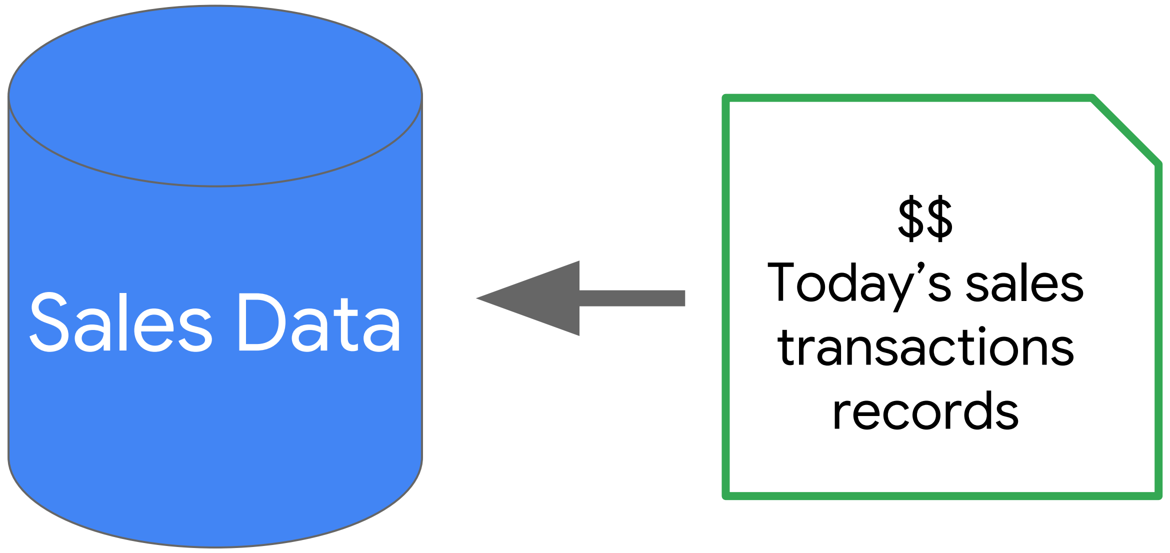
✓ 1. Related to the objective

✓ 2. Known at prediction-time

✓ 3. Numeric with meaningful magnitude

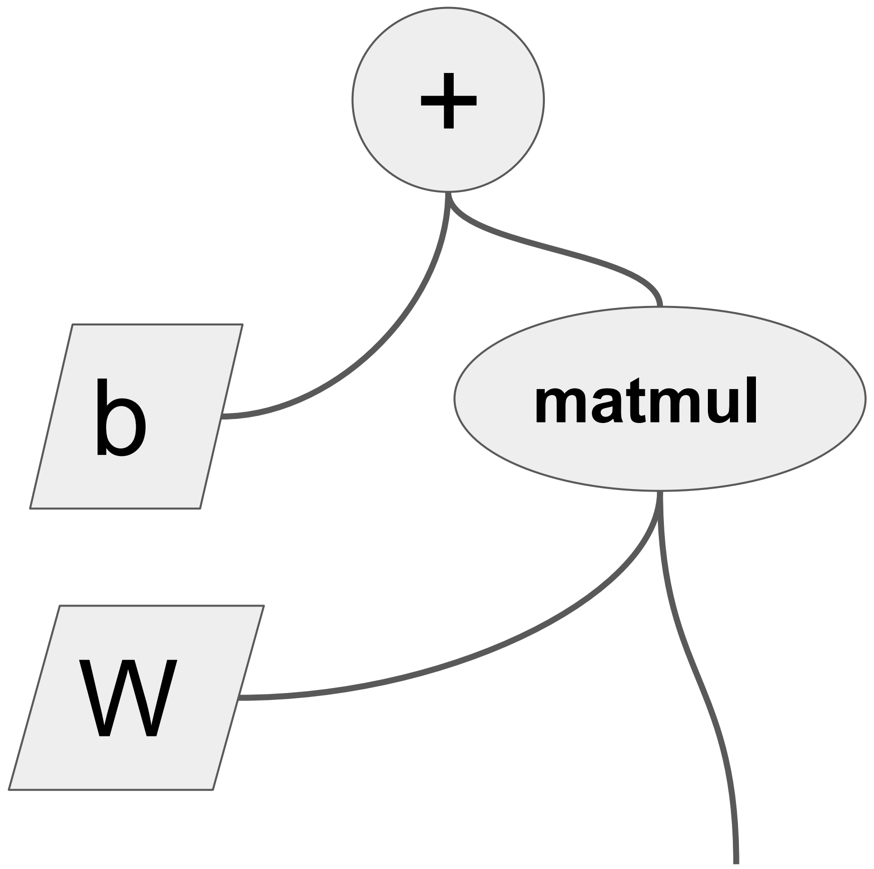
✓ 4. Have enough examples

✓ 5. Bring human insight to problem


Sales Data
$$
Today's sales
transactions
records

Good features are:

- ✓ 1. Related to the objective
- ✓ 2. Known at prediction-time
- ✓ 3. Numeric with meaningful magnitude
- ✓ 4. Have enough examples
- ✓ 5. Bring human insight to problem


+
b
matmul
W

Predict total number of customers who will use a certain discount coupon

PROMOCODE1234

Features must be numeric with meaningful magnitude

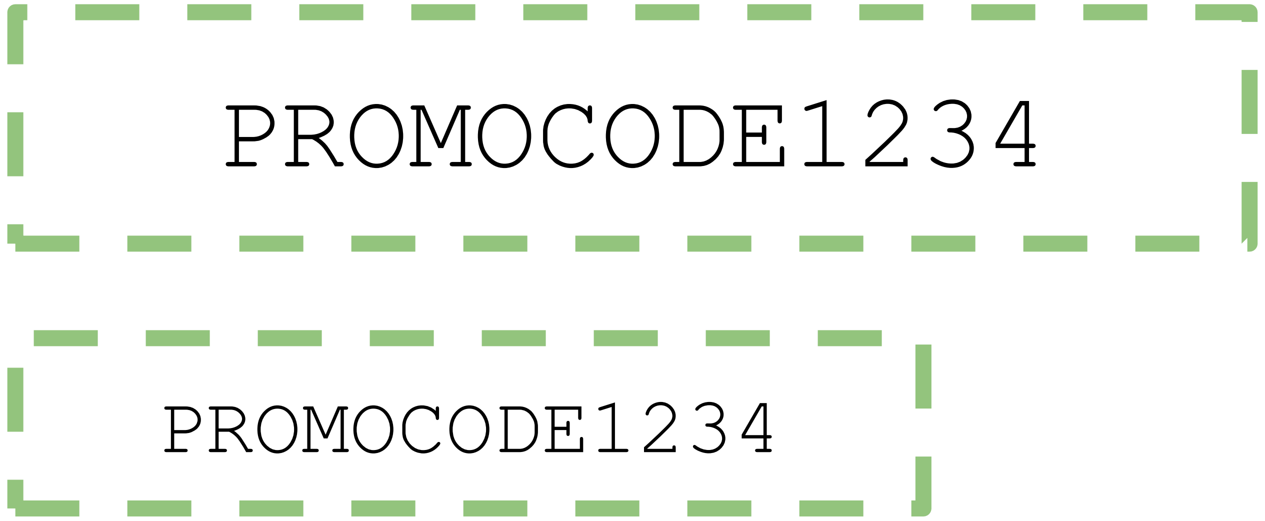
1 Percent value of the discount (e.g. 10% off, 20% off, etc.)

```
PROMOCODE1234 10%
```

```
PROMOCODE1234 20%
```

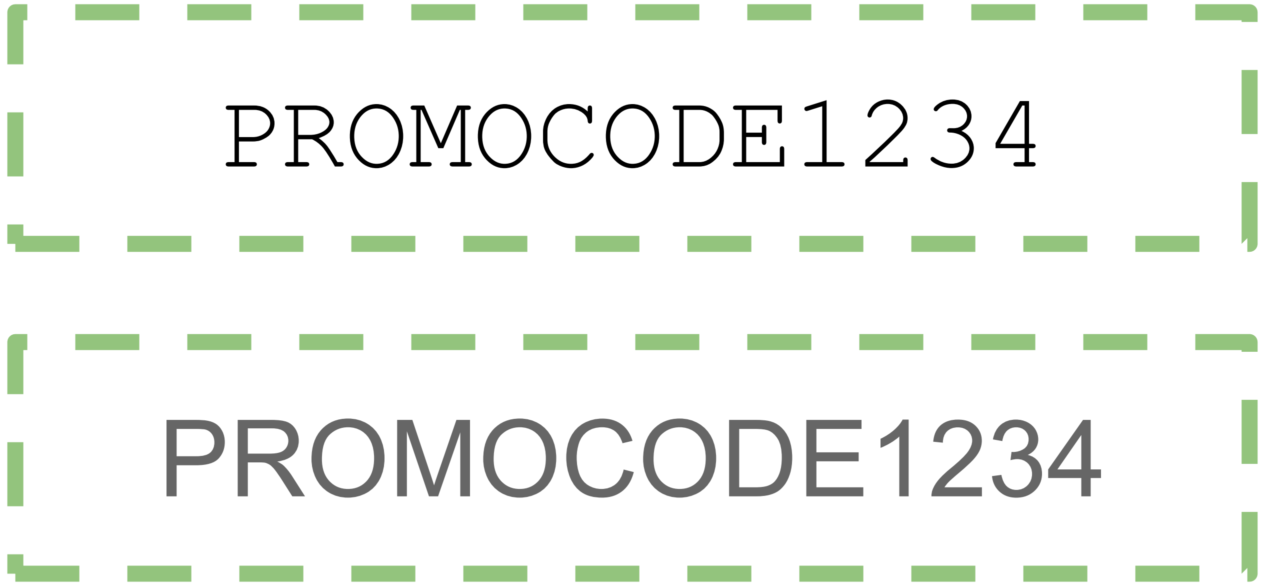
Features must be numeric with meaningful magnitude

2 Size of the coupon



Features must be numeric with meaningful magnitude

3 Font an advertisement is in (Arial, Times New Roman, etc.)

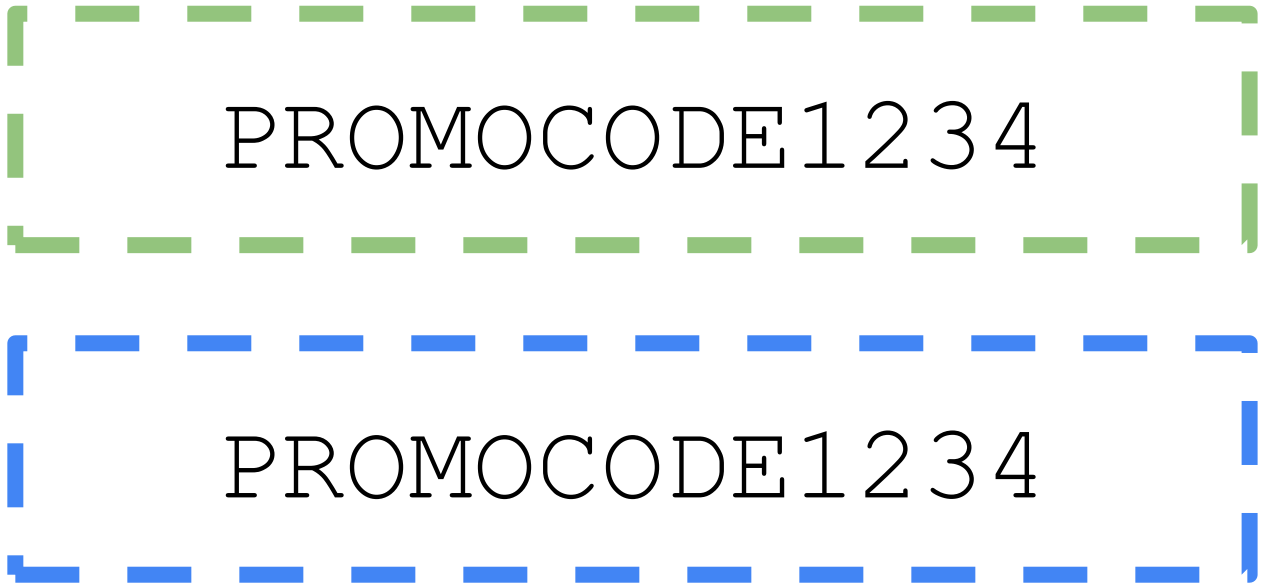


PROMOCODE1234

PROMOCODE1234

Features must be numeric with meaningful magnitude

4 Color of coupon (red, black, blue, etc.)



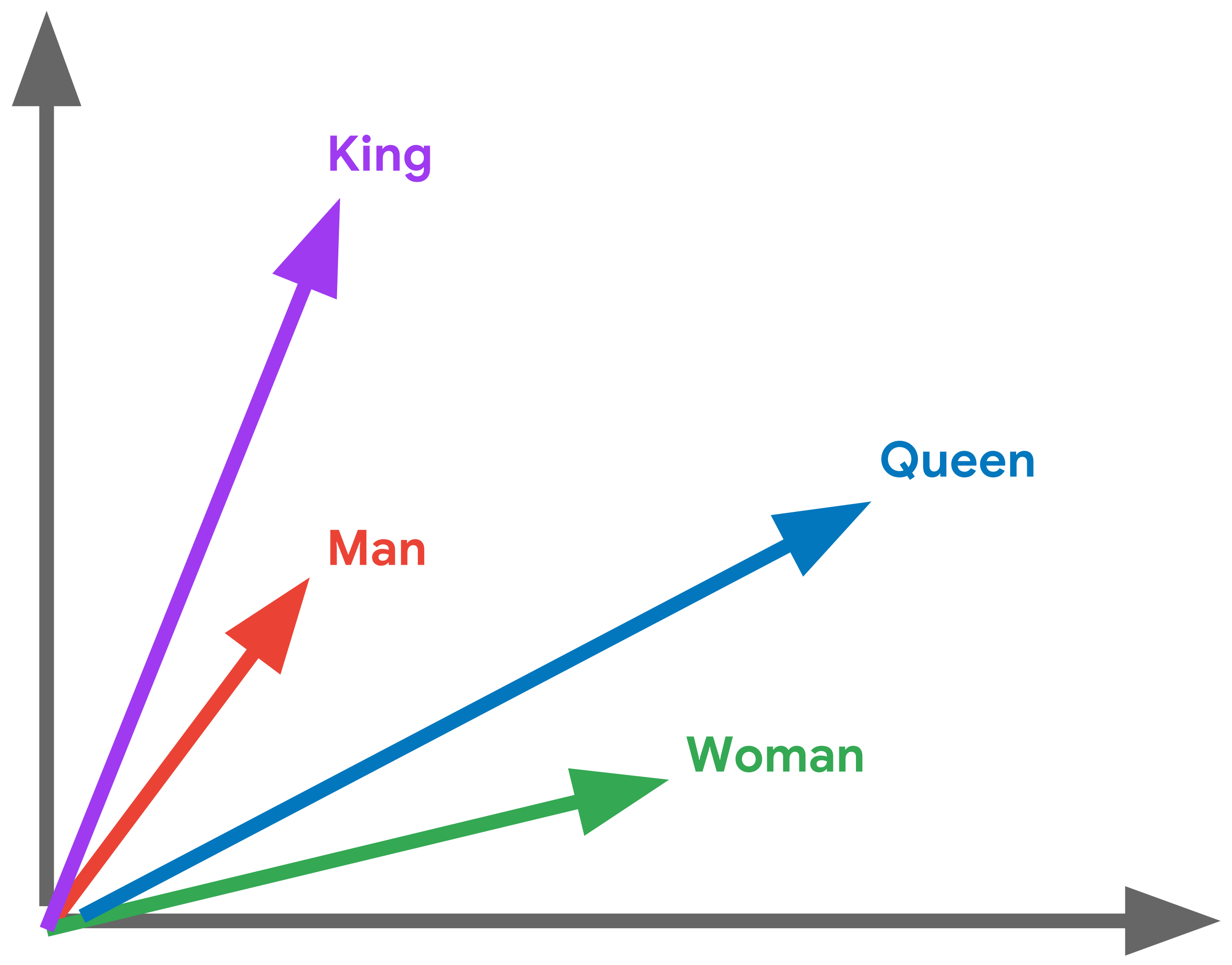
Features must be numeric with meaningful magnitude

5 Item category (1 for dairy, 2 for deli, 3 for canned goods, etc.)

`PROMOCODE1234 - Deli`

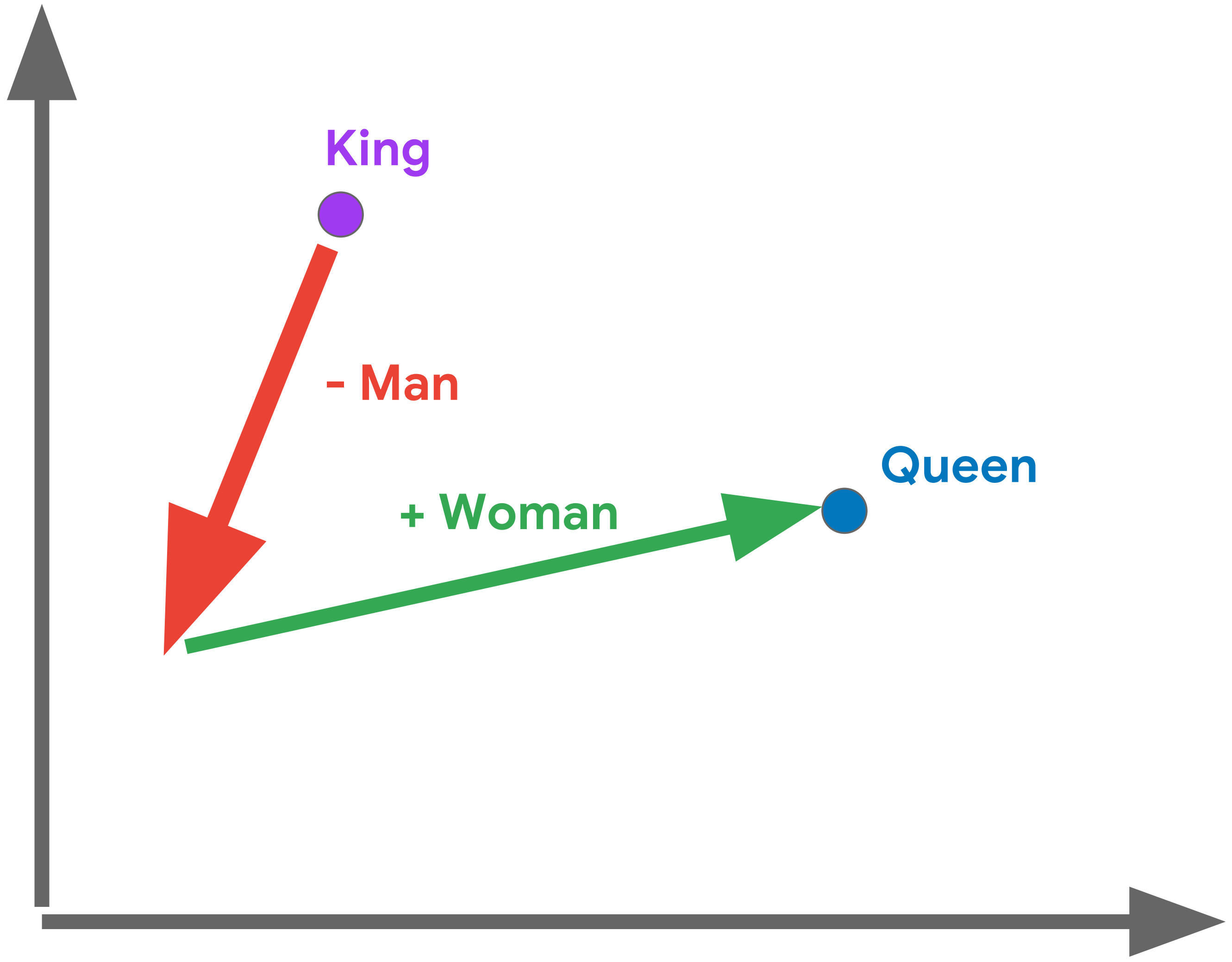
`PROMOCODE1234 - Canned Goods`

Word2Vec



Word Vectors

Word2Vec



Vector Composition

Good features are:

✓ 1. Related to the objective

✓ 2. Known at prediction-time

✓ 3. Numeric with meaningful magnitude

✓ 4. Have enough examples

✓ 5. Bring human insight to problem

Features must be numeric with meaningful magnitude

Percent value of the discount
(e.g. 10% off, 20% off, etc.)

`PROMOCODE1234 10%`

`PROMOCODE1234` **`87%`**

Avoid having values of which you don't have enough examples

Good features are:

- ✓ 1. Related to the objective
- ✓ 2. Known at prediction-time
- ✓ 3. Numeric with meaningful magnitude
- ✓ 4. Have enough examples
- ✓ 5. Bring human insight to problem

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Exploring and Preprocessing Data**

Format: Talking Head

Video Name: T-BQML-O_4_l3_exploring_and_preprocessing_data

Building a ML Model involves:

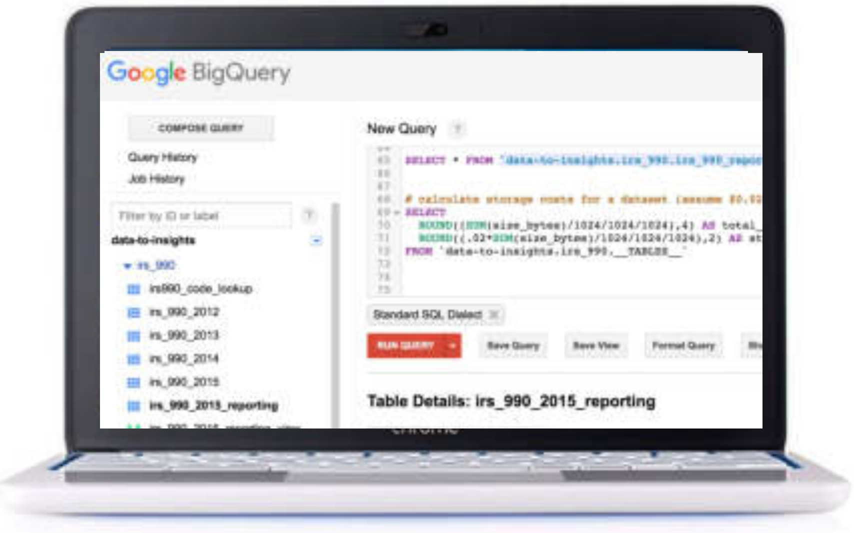
Create the dataset

Build the model

Operationalize the model

Recall: Options for Exploring and Preparing Datasets

SQL + Web UI

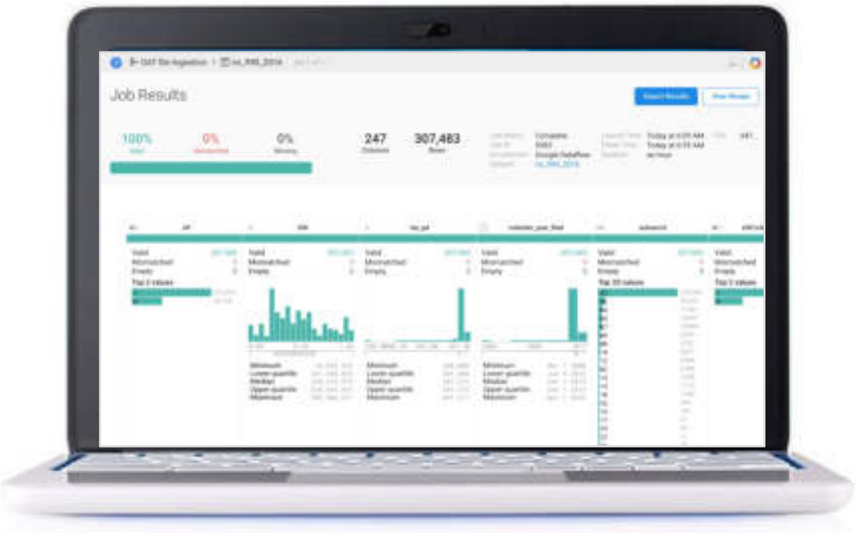


```
SELECT * FROM `data-to-insights.irs_990.irs_990_repor


# calculate storage costs for a dataset (assume $0.02
SELECT
  ROUND((SUM(size_bytes)/1024/1024/1024),4) AS total_
  ROUND((.02*SUM(size_bytes)/1024/1024/1024),2) AS st
FROM `data-to-insights.irs_990.__TABLES__`
```

Table Details: irs_990_2015_reporting

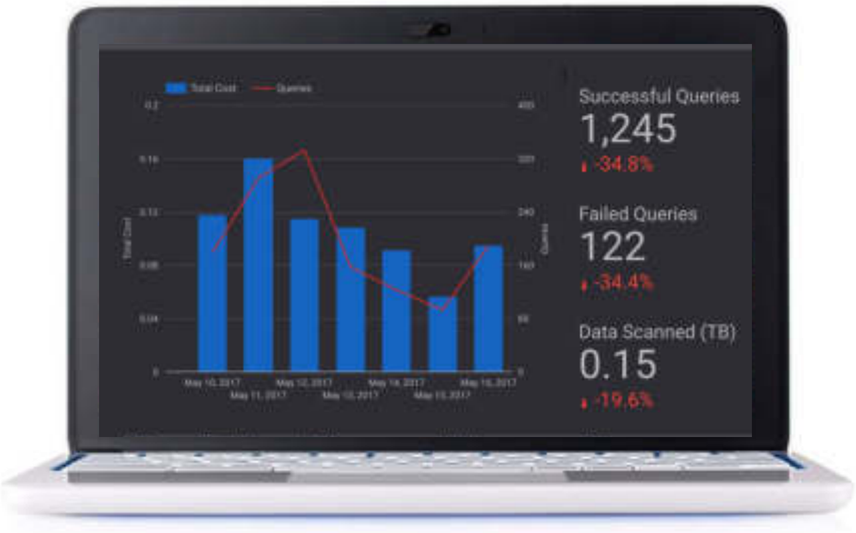
- Flexible, Fast, and Familiar
- Requires SQL knowledge

Data Preparation Tools



- GUI for Exploring Columns and Rows
- Fast Summary Statistics

Visualization Tools



- Visually Shape and Re-Shape Quickly
- See Data a Different Way

Course 4: Applying Machine Learning to your Datasets

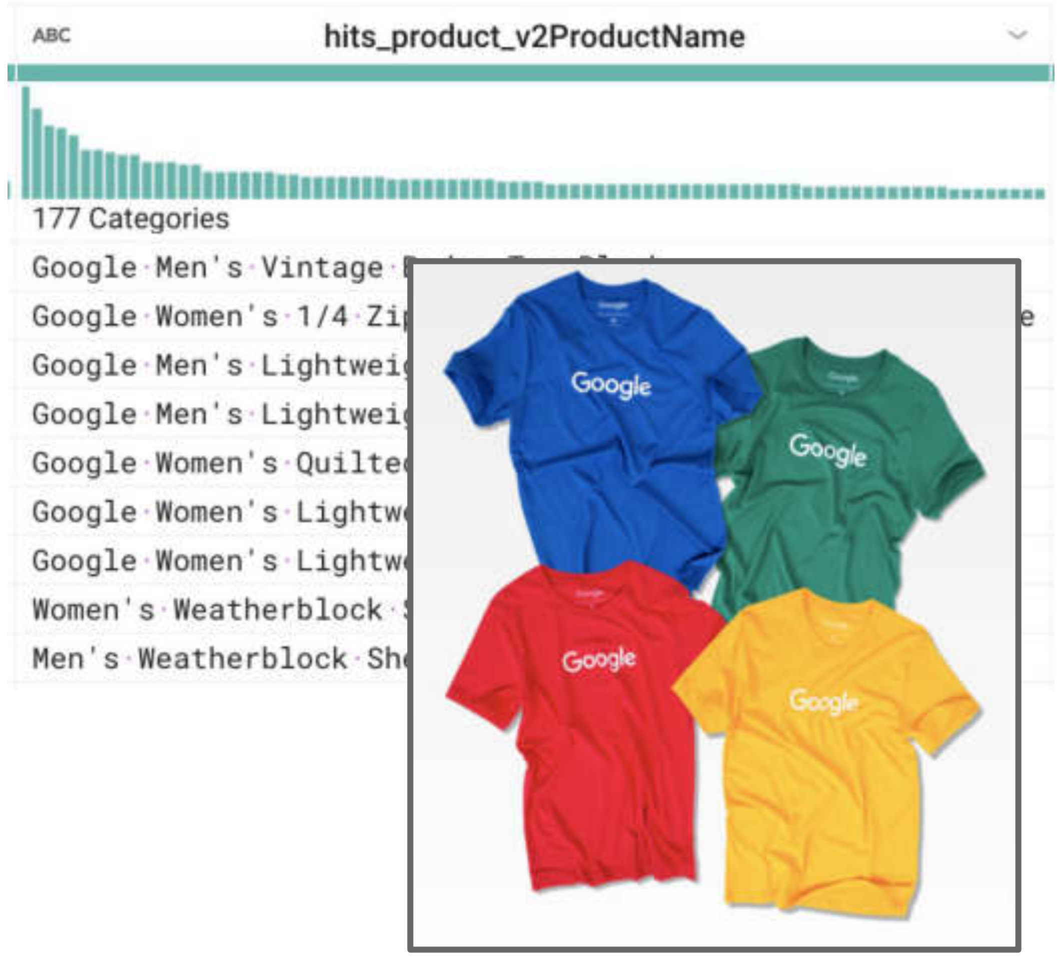
Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Demo: Exploring and Preprocessing Data**

Format: Talking Head

Video Name:
T-BQML-O_4_l4_demo:_exploring_and_preprocessing_data

Dataset Exploration



hits_product_v2ProductName

177 Categories

```
Google·Men's·Vintage
Google·Women's·1/4·Zi
Google·Men's·Lightwei
Google·Men's·Lightwei
Google·Women's·Quilte
Google·Women's·Lightw
Google·Women's·Lightw
Women's·Weatherblock
Men's·Weatherblock·Sh
```

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Lab Intro: Exploring and Preprocessing Data**

Format: Talking Head

Video Name:
T-BQML-O_4_l5_lab_intro:_exploring_and_preprocessing_data

Lab

Exploring and Preprocessing Data

Evan Jones

LAB:

Exploring and Preprocessing Data

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Lab Solution: Exploring and Preprocessing Data**

Format: Talking Head + Lab Screencast

Video Name:
T-BQML-O_4_l7_lab_solution:_exploring_and_preprocessing_data

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

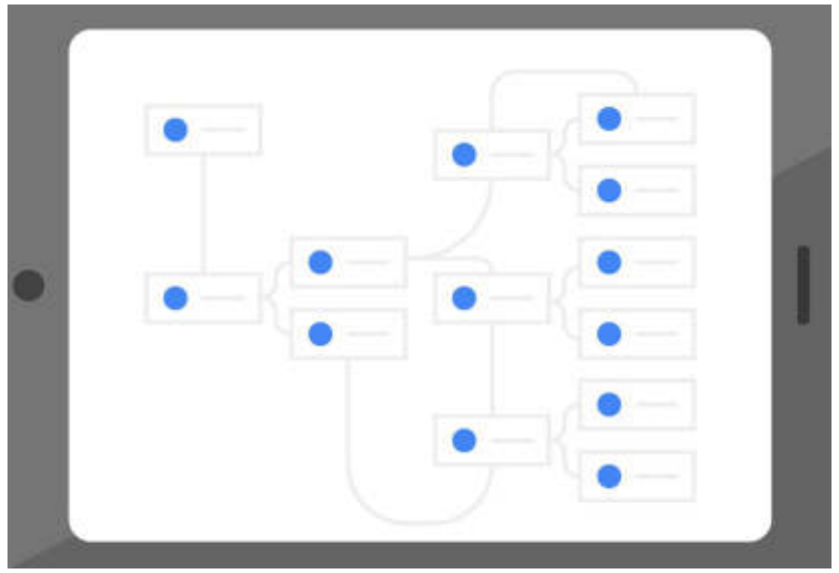
Lesson Title: **Pipeline Creation**

Format: Talking Head

Video Name: T-BQML-O_4_l8_pipeline_creation

Other options for creating data pipelines

- Dataprep (batch)
- Dataflow (batch/stream)
- Cloud Composer



Course 4: Applying Machine Learning to your Datasets

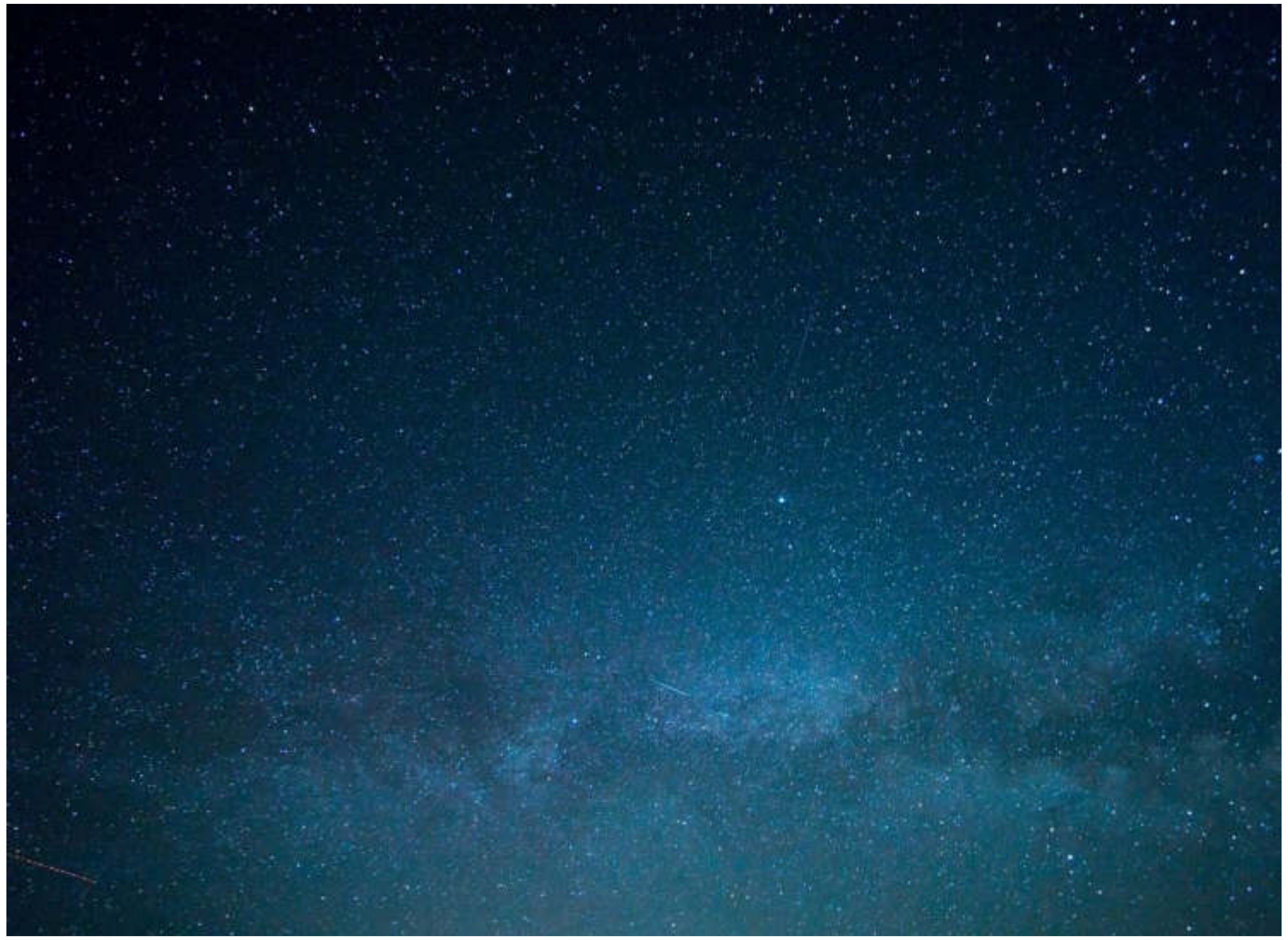
Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Knowing the Unknowable**

Format: Talking Head

Video Name: T-BQML-O_4_l9_knowing_the_unknowable

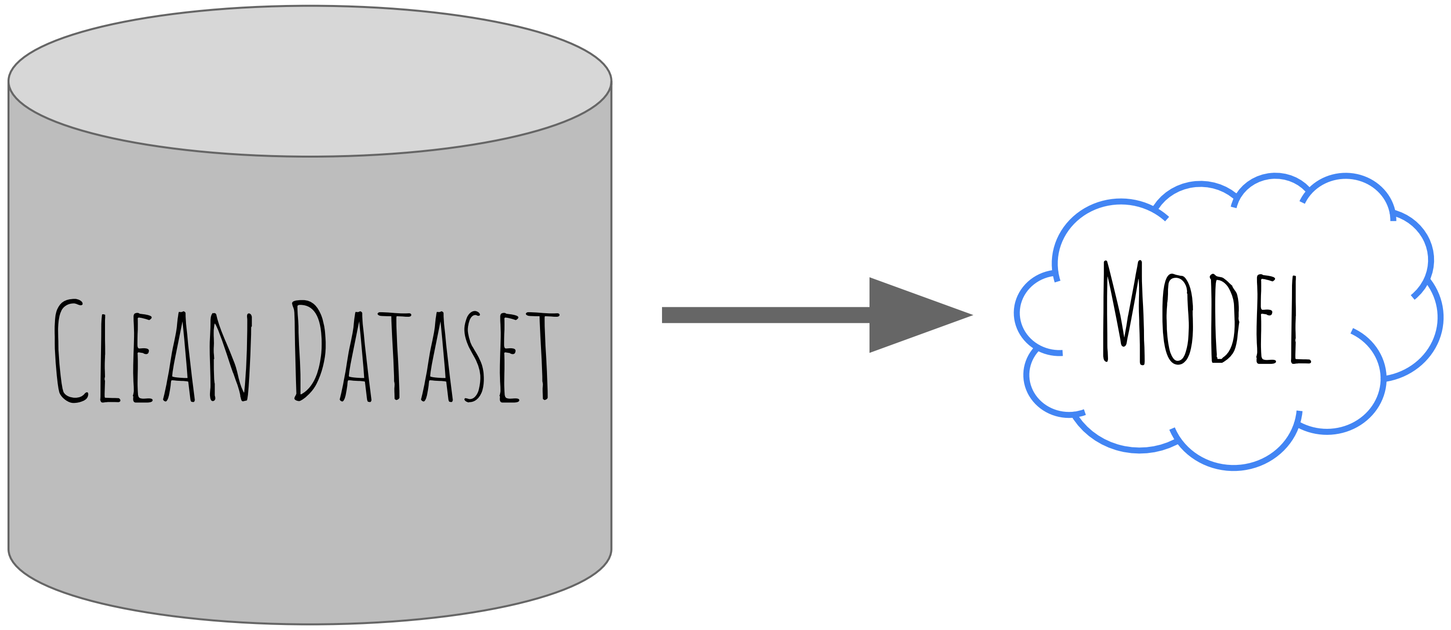
Knowing the Unknowable



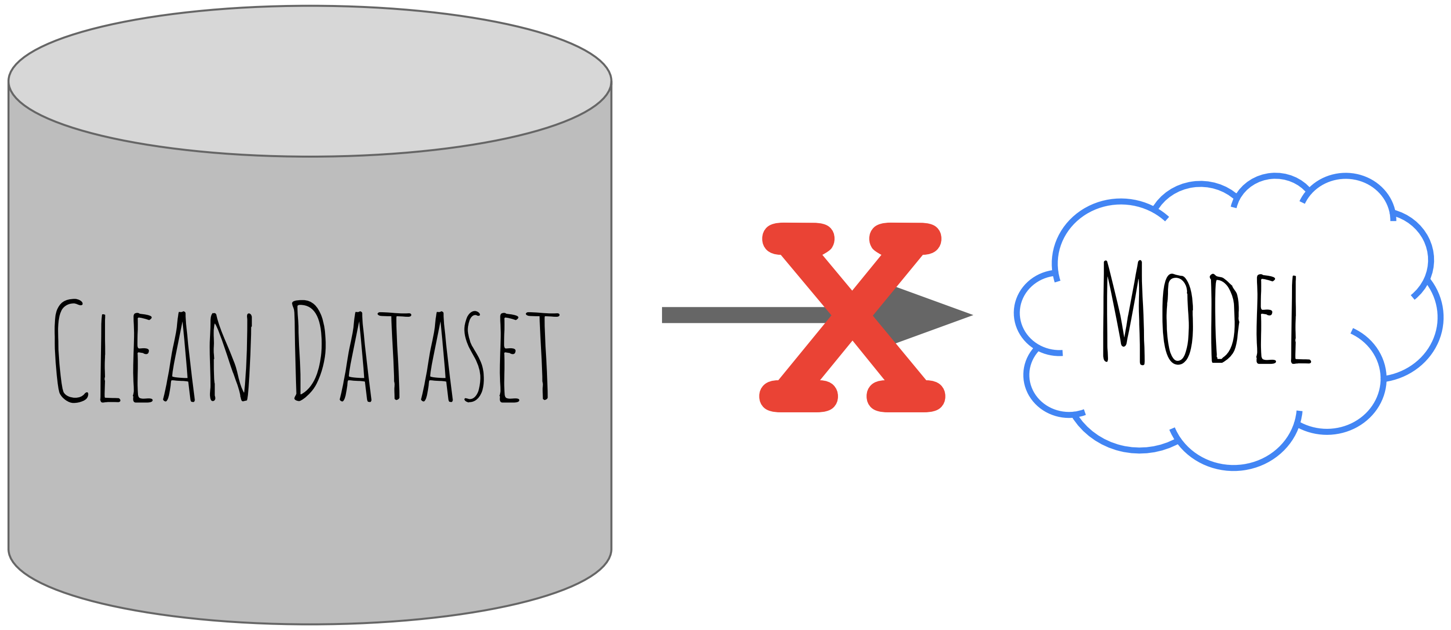
Knowing the Unknowable

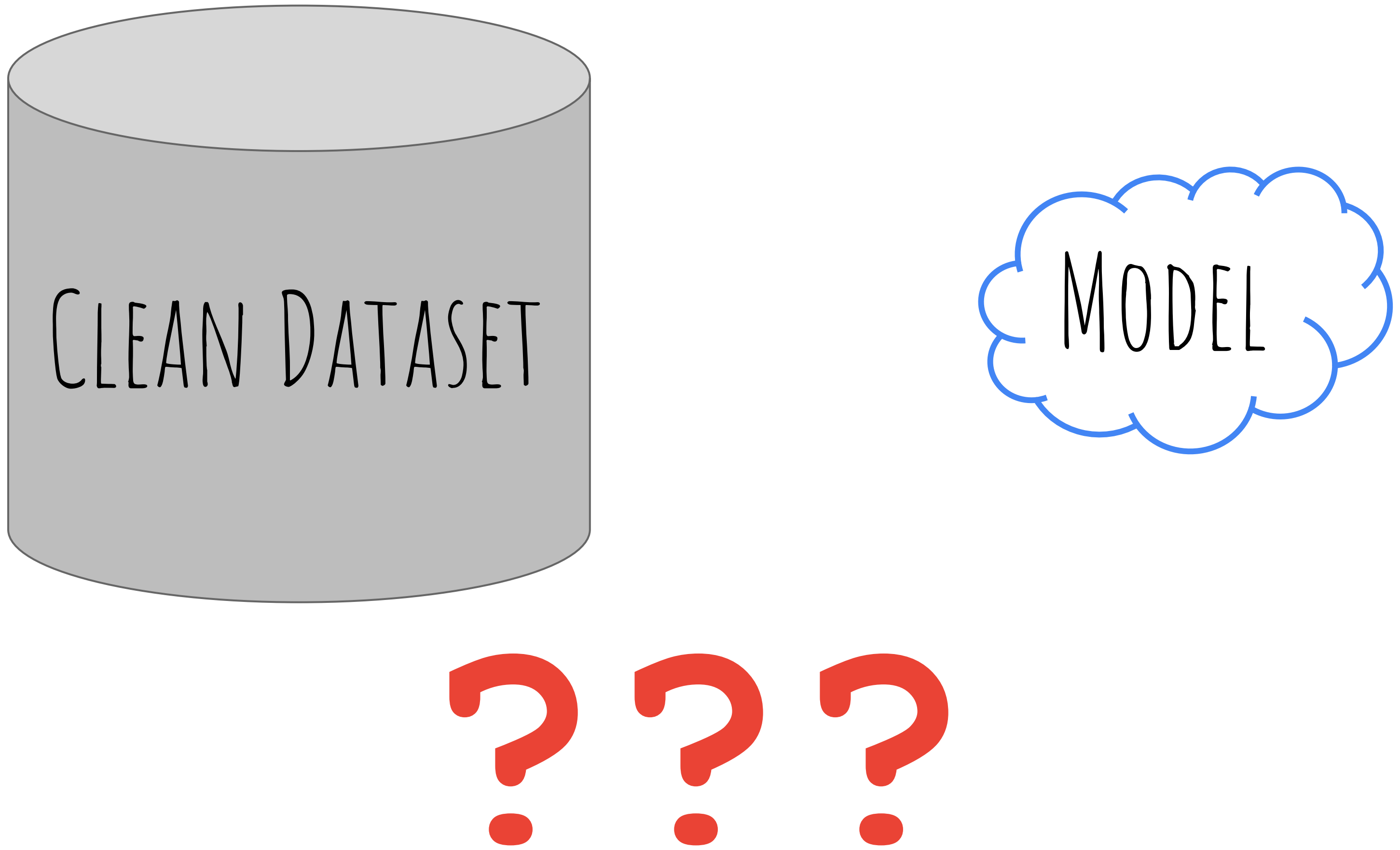
What we have to work with:

Can we feed it all to the model?



Your dataset has the answers already



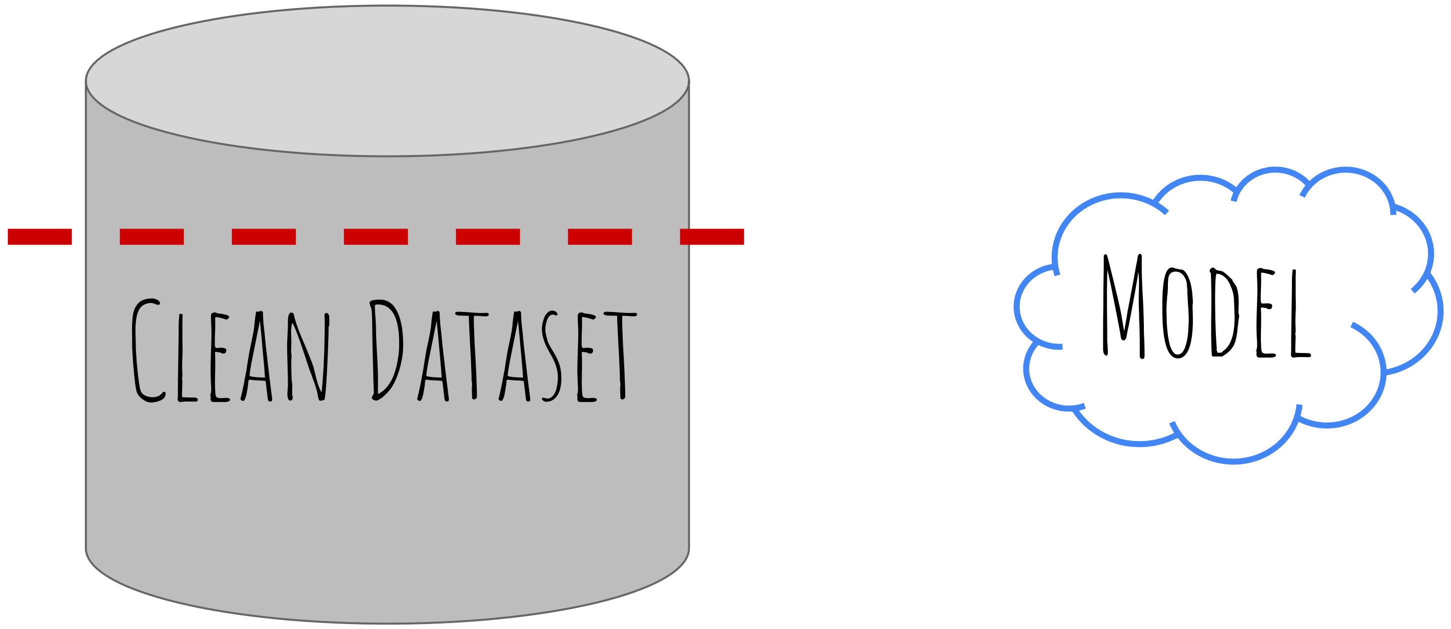


Clean Dataset

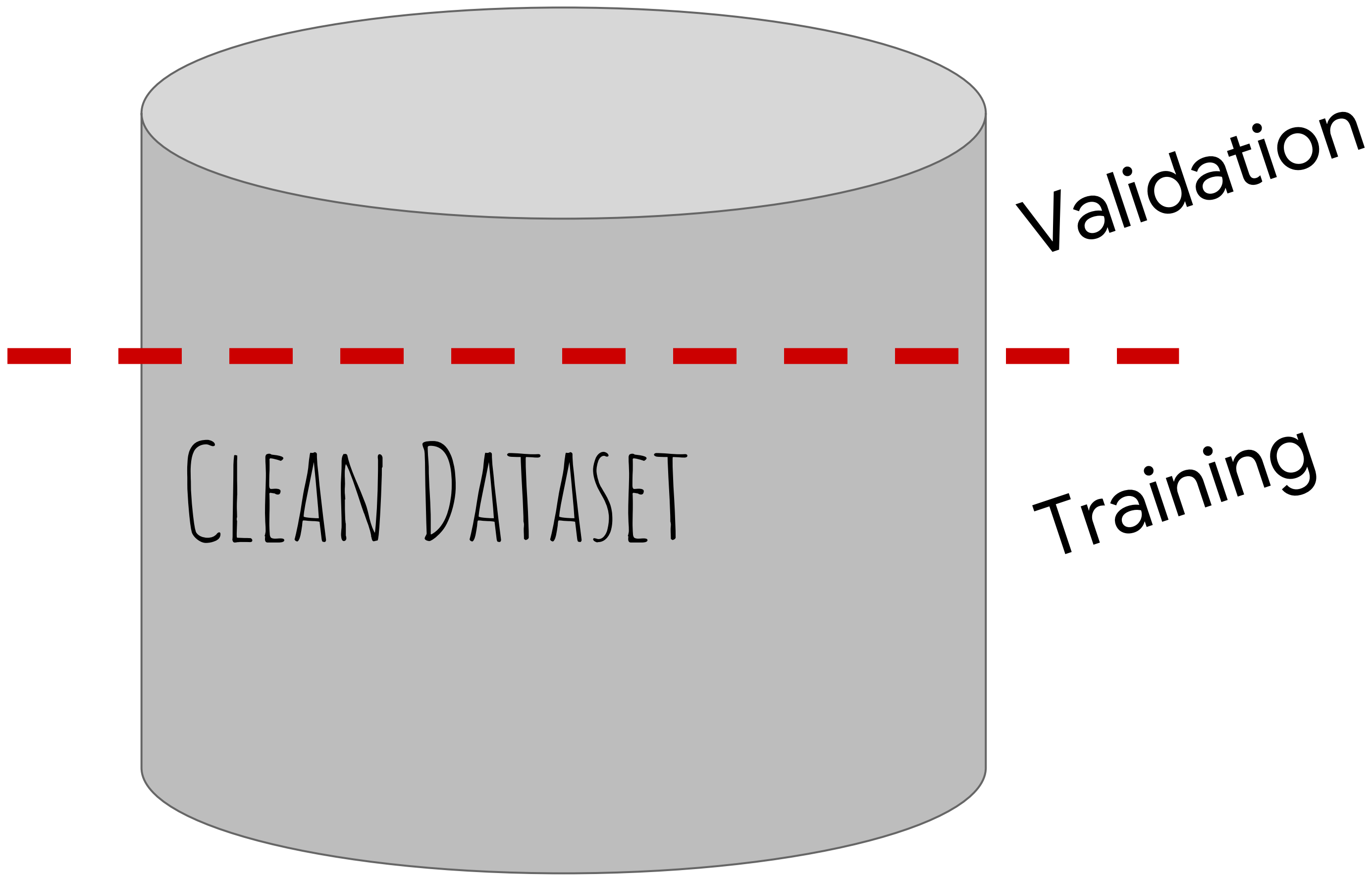
???

Model

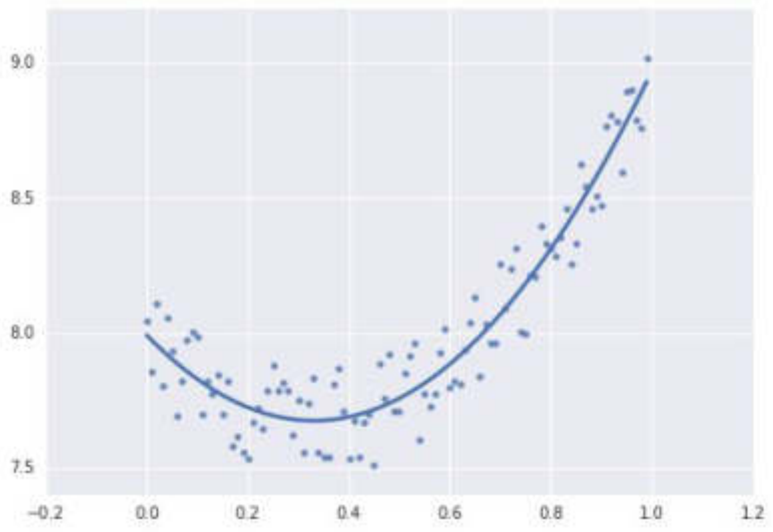
Split your Dataset



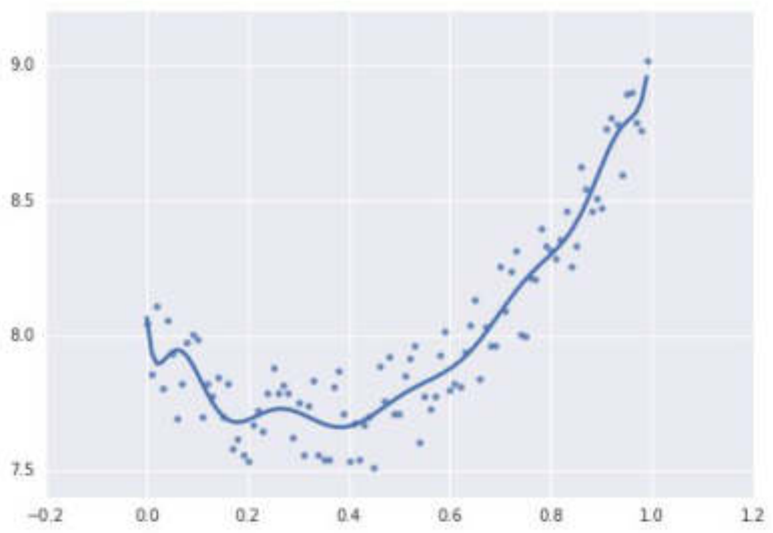
Split your Dataset



Validation helps prevent overfitting



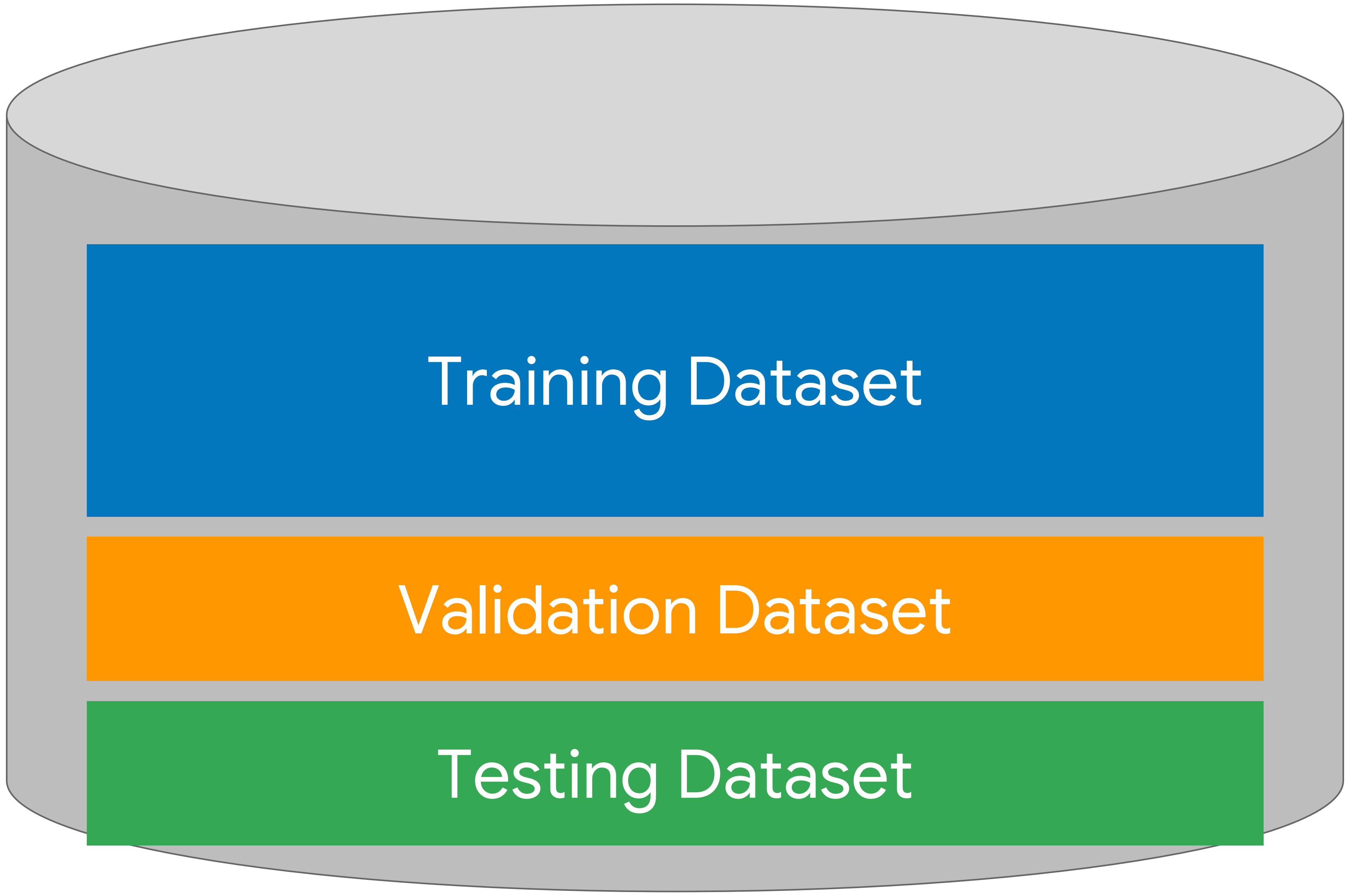
Fit



Overfit

What about retraining the model? It’s already seen the validation data

Split your data to train and simulate the real-world unknown



Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

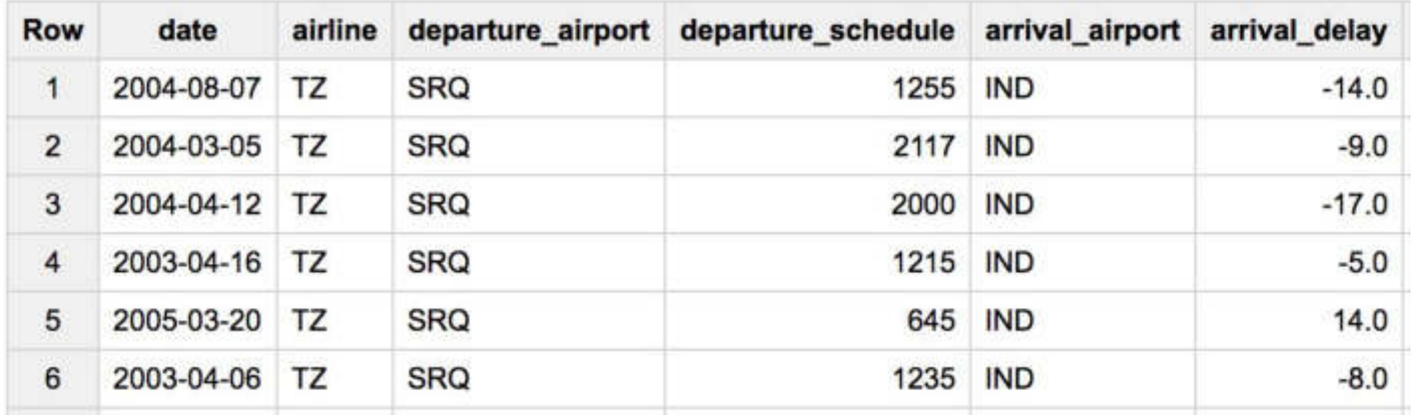
Lesson Title: **Creating Repeatable Dataset Splits**

Format: Talking Head

Video Name: T-BQML-O_4_l10_creating_repeatable_dataset_splits

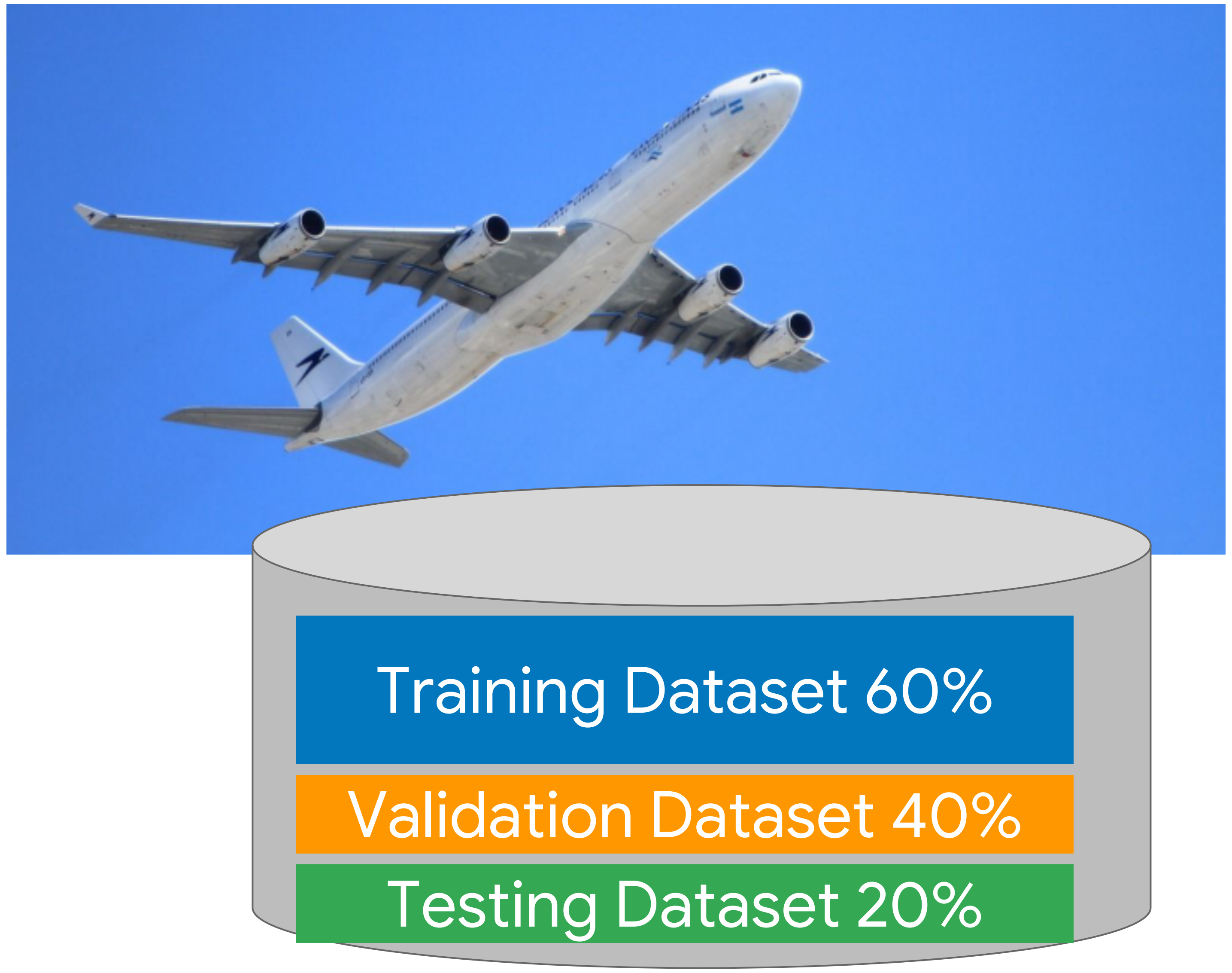
How do I actually split my dataset?

Example Dataset: Millions of Flights



Row	date	airline	departure_airport	departure_schedule	arrival_airport	arrival_delay
1	2004-08-07	TZ	SRQ	1255	IND	-14.0
2	2004-03-05	TZ	SRQ	2117	IND	-9.0
3	2004-04-12	TZ	SRQ	2000	IND	-17.0
4	2003-04-16	TZ	SRQ	1215	IND	-5.0
5	2005-03-20	TZ	SRQ	645	IND	14.0
6	2003-04-06	TZ	SRQ	1235	IND	-8.0

Our Goal: Sample and Split the Data



Training Dataset 60%

Validation Dataset 40%

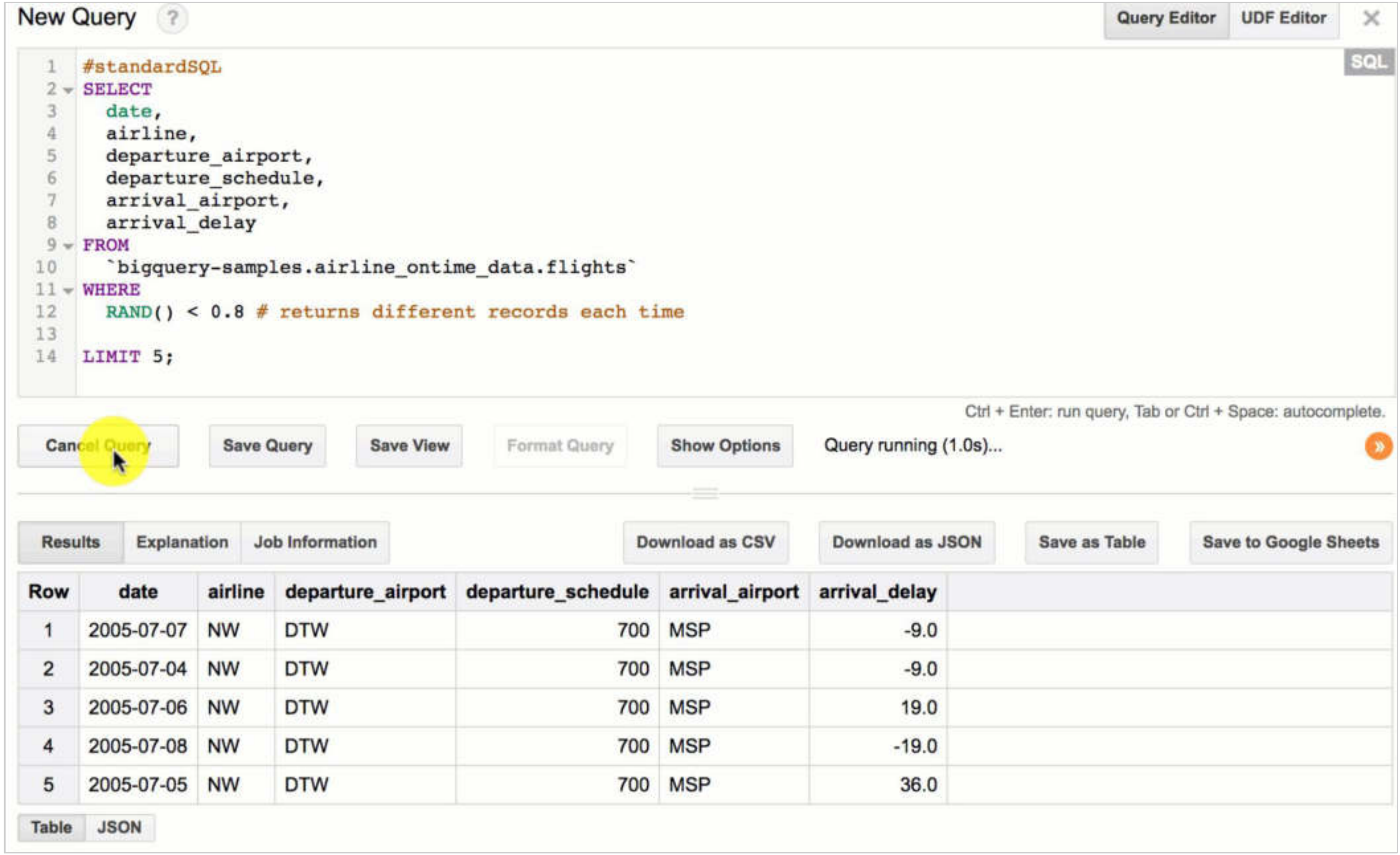
Testing Dataset 20%

Can’t we just use a WHERE clause and pull 80% of the rows?

Hard to identify and split the remaining 20% of data for validation and testing if the data in each slice is changing each time

RAND( )

will return different results each time →



```
#standardSQL
SELECT
  date,
  airline,
  departure_airport,
  departure_schedule,
  arrival_airport,
  arrival_delay
FROM
  `bigquery-samples.airline_ontime_data.flights`
WHERE
  RAND() < 0.8 # returns different records each time

LIMIT 5;
```

Row	date	airline	departure_airport	departure_schedule	arrival_airport	arrival_delay
1	2005-07-07	NW	DTW	700	MSP	-9.0
2	2005-07-04	NW	DTW	700	MSP	-9.0
3	2005-07-06	NW	DTW	700	MSP	19.0
4	2005-07-08	NW	DTW	700	MSP	-19.0
5	2005-07-05	NW	DTW	700	MSP	36.0

Splitting the data must
be a repeatable process

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Lab Intro: Creating Repeatable Dataset Splits**

Format: Talking Head

Video Name:
T-BQML-O_4_l11_lab_intro:_creating_repeatable_dataset_splits

Lab

Creating Repeatable Dataset Splits

Evan Jones

LAB:

Creating Repeatable Dataset Splits

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Demo: Creating Repeatable Dataset Splits**

Format: Talking Head

Video Name:

Course 4: Applying Machine Learning to your Datasets

Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Lab Solution: Creating Repeatable Dataset Splits**

Format: Talking Head + Lab Screencast

Video Name:
T-BQML-O_4_l13_lab_solution:_creating_repeatable_dataset_splits

Course 4: Applying Machine Learning to your Datasets

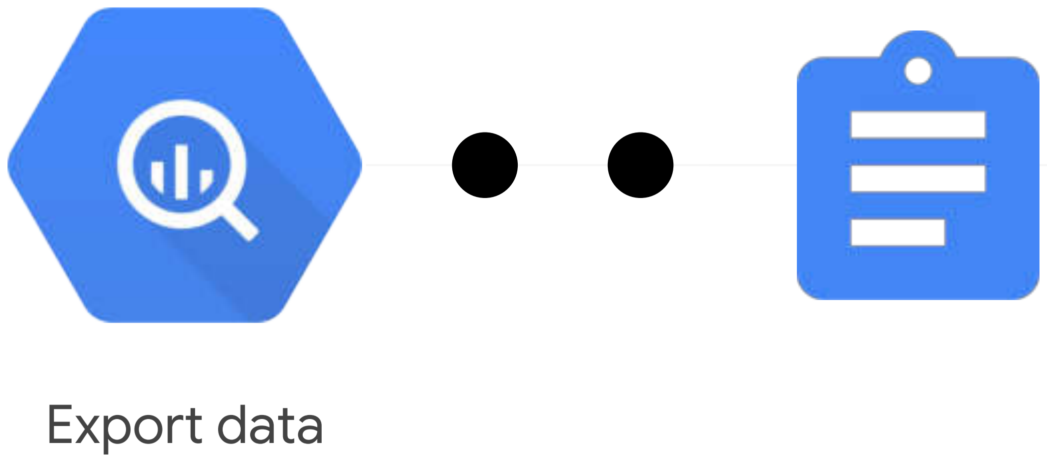
Module 4: Creating ML Datasets in BigQuery

Lesson Title: **Introducing BigQuery Machine Learning (BQML)**

Format: Talking Screencast

Video Name: T-BQML-O_4_l13b_bqml_intro

Days to months to create an ML model



1 Regression in Excel/Sheets:

Export small amounts of data from BQ
Run linear regression
Get a model with **low accuracy** due to small data for training
Go back and get more data to create new features, and improve performance
Repeat. It's hard, so you **stop after a few iterations**

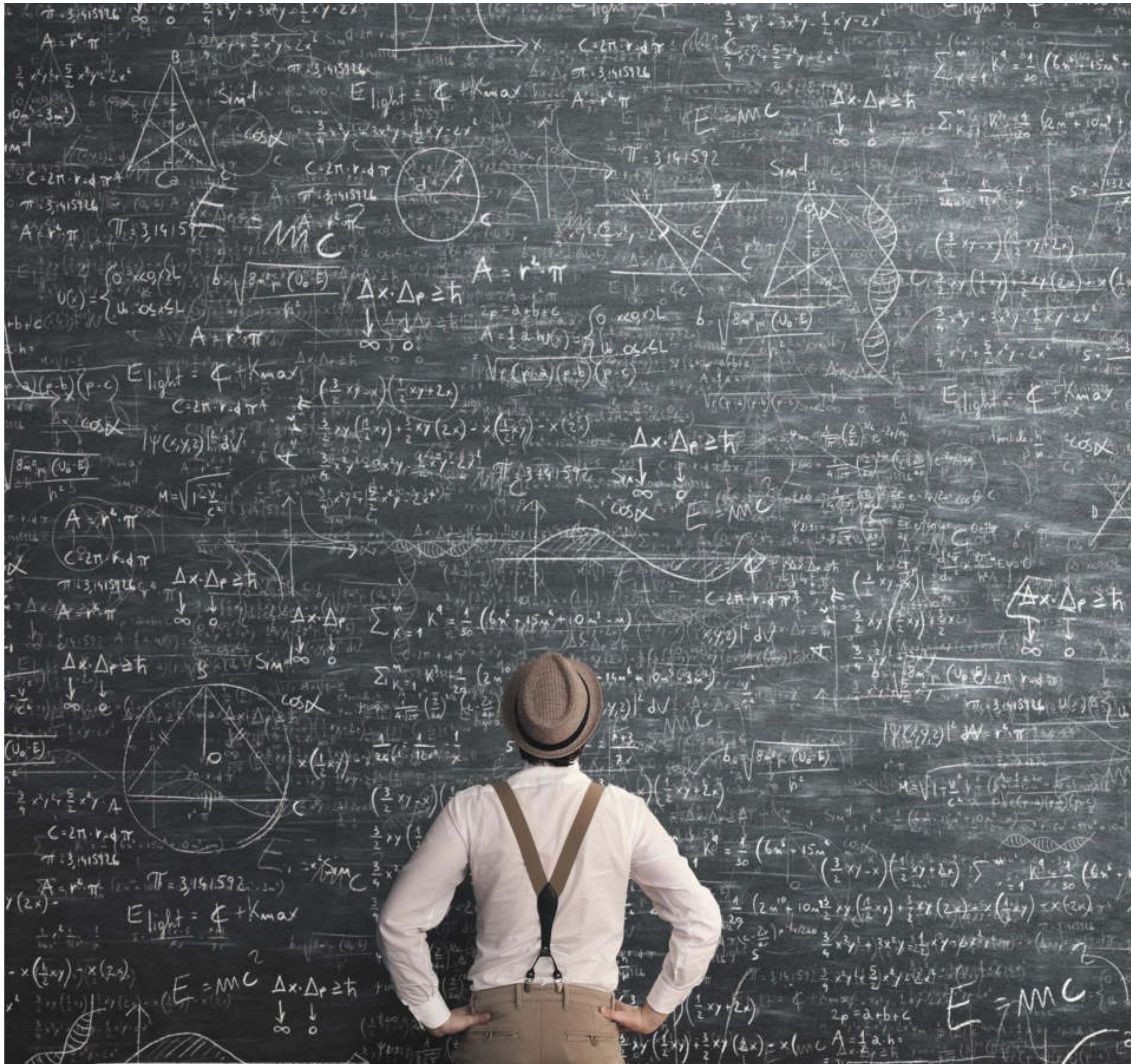
2 TensorFlow or scikit-learn:

Only an expert data scientist can do this
Export small amounts of data from BQ
Create frames of data for use with TensorFlow
Build model
Go back and get more data to create new features, and improve performance
Repeat. It's hard, so you **stop after a few iterations**

Key challenges affecting ML

Expensive for companies to hire enough data scientists

Complex and time consuming to move data out of BigQuery



Introducing

BigQuery ML

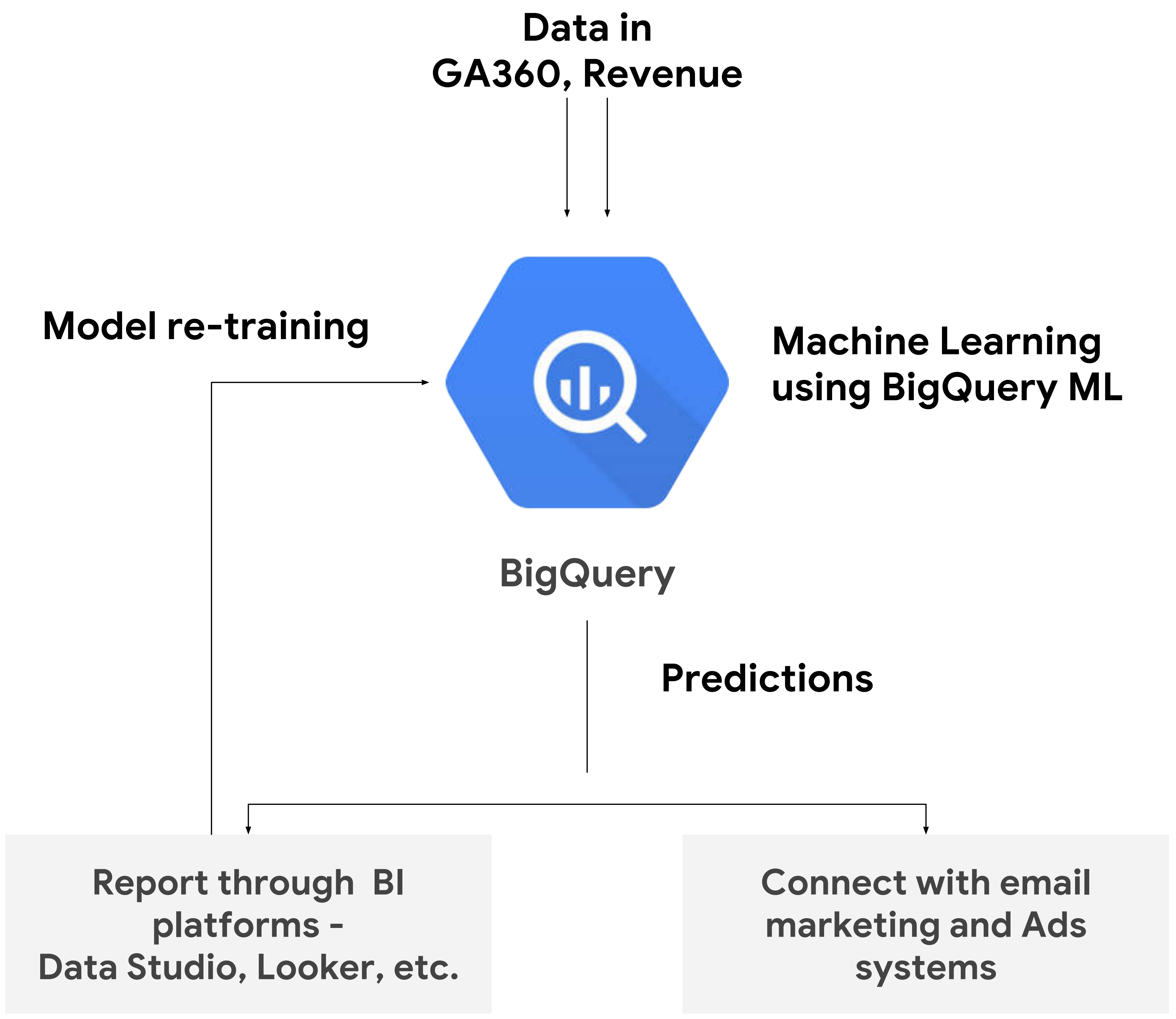
Machine learning using SQL in BigQuery

Bring ML to your data with **BigQuery ML**

Data analysts and data scientists can

1. Use familiar SQL for machine learning
2. Train models over all their data in BigQuery
3. Not worry about hypertuning or feature transformations

Example



Data in GA360, Revenue

Model re-training

Machine Learning using BigQuery ML

BigQuery

Predictions

Report through BI platforms - Data Studio, Looker, etc.

Connect with email marketing and Ads systems

Behind the scenes

With 2 lines of code:

- Leverages BigQuery’s processing power to build a model
- Auto-tunes learning rate
- Auto-splits data into training and test

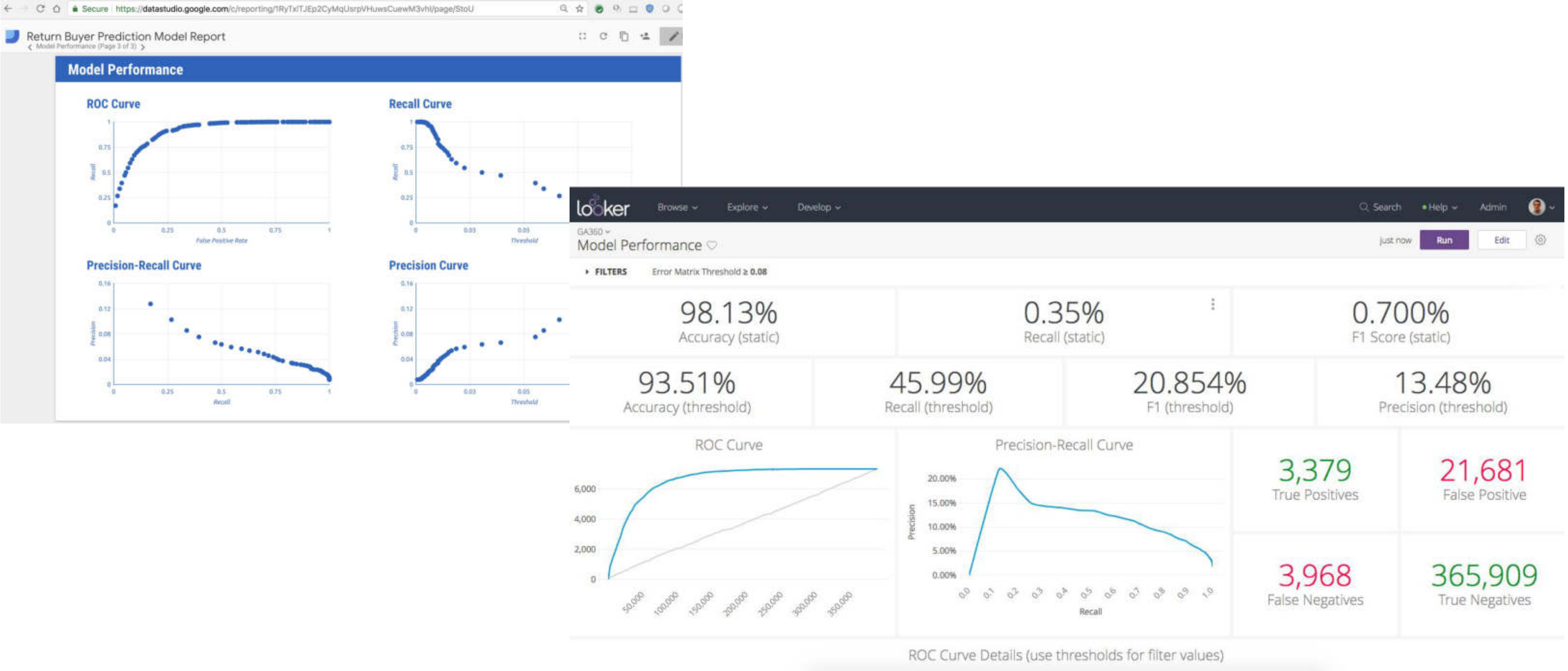
For the advanced user:

- L1/L2 regularization
- 3 strategies for training/test split: Random, Sequential, Custom
- Set learning rate

Supported features

- StandardSQL and UDFs within the ML queries
- Linear Regression (Forecasting)
- Binary Logistic Regression (Classification)
- Model evaluation functions for standard metrics, including the ROC curve
- Model weight inspection
- Feature distribution analysis through standard functions

Available through your favorite BI Platform



Secure https://datastudio.google.com/c/reporting/1RyTxlTJEp2CyMqUsrpVHuwsCuewM3vhl/page/StoU

Return Buyer Prediction Model Report

Model Performance (Page 3 of 3)

Model Performance

Model Performance

FILTERS Error Matrix Threshold ≥ 0.08

98.13% Accuracy (static)

0.35% Recall (static)

0.700% F1 Score (static)

93.51% Accuracy (threshold)

45.99% Recall (threshold)

20.854% F1 (threshold)

13.48% Precision (threshold)

3,379 True Positives

21,681 False Positive

3,968 False Negatives

365,909 True Negatives

ROC Curve Details (use thresholds for filter values)

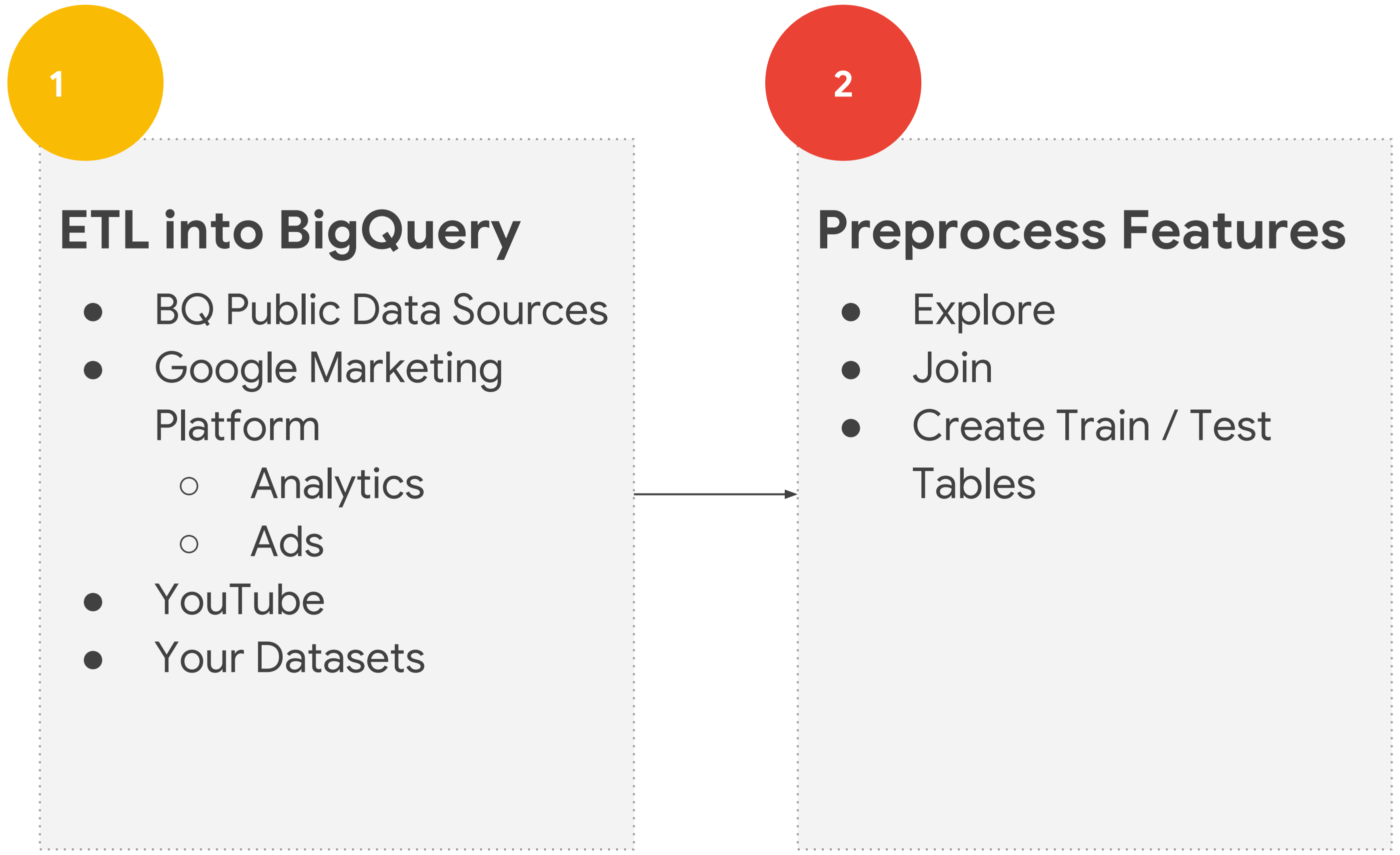
The End-to-End BQML Process

1

ETL into BigQuery

- BQ Public Data Sources
- Google Marketing Platform
  - Analytics
  - Ads
- YouTube
- Your Datasets

The End-to-End BQML Process



1

ETL into BigQuery

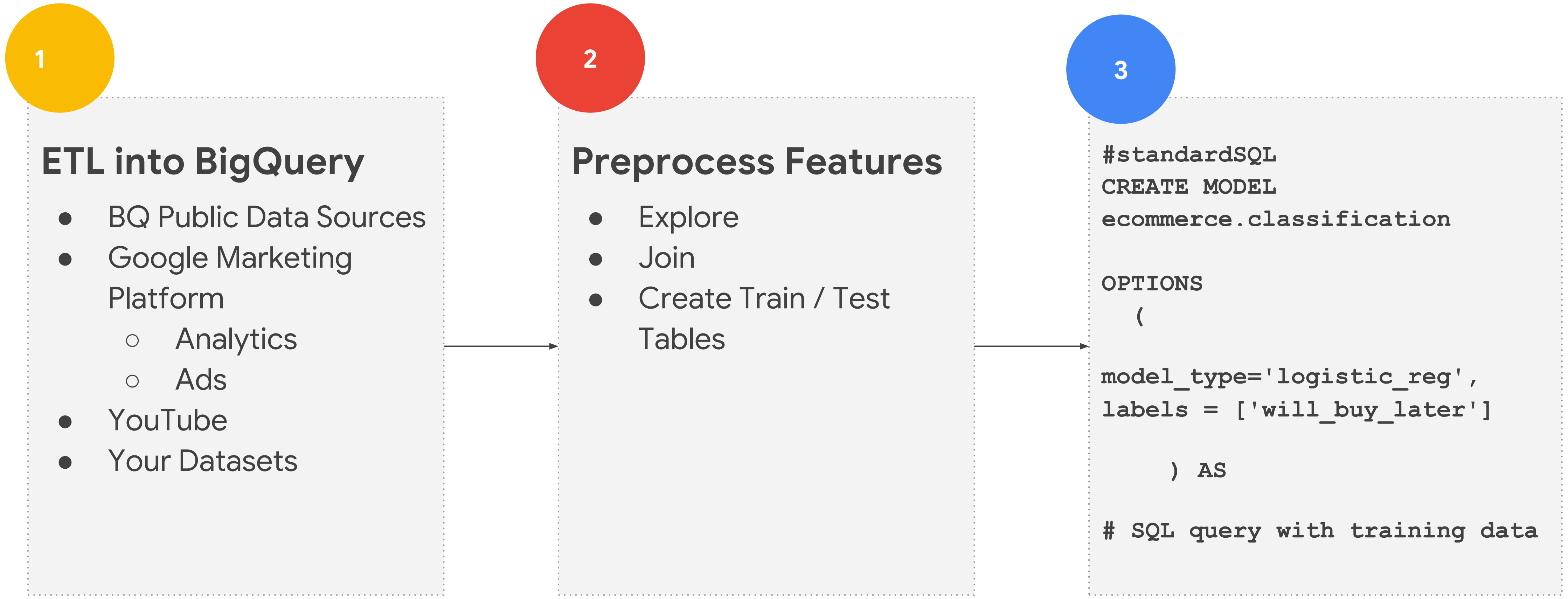
- BQ Public Data Sources
- Google Marketing Platform
  - Analytics
  - Ads
- YouTube
- Your Datasets

2

Preprocess Features

- Explore
- Join
- Create Train / Test Tables

The End-to-End BQML Process



1

ETL into BigQuery

- BQ Public Data Sources
- Google Marketing Platform
  - Analytics
  - Ads
- YouTube
- Your Datasets

2

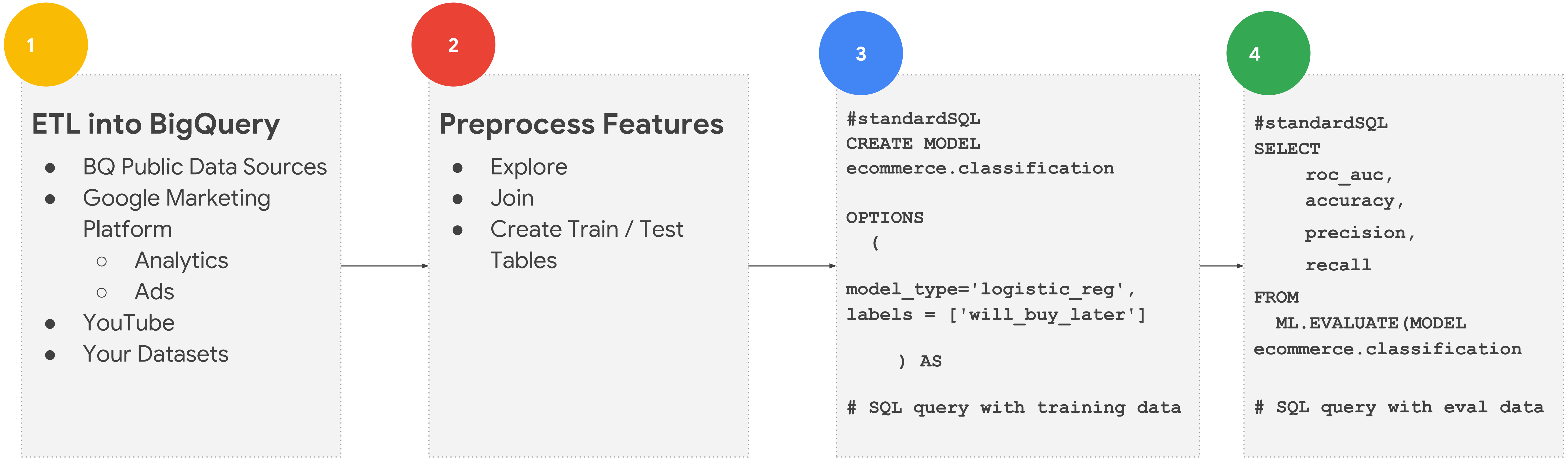
Preprocess Features

- Explore
- Join
- Create Train / Test Tables

3

```
#standardSQL
CREATE MODEL
ecommerce.classification

OPTIONS
  (

model_type='logistic_reg',
labels = ['will_buy_later']

     ) AS

# SQL query with training data
```

The End-to-End BQML Process



1

ETL into BigQuery

- BQ Public Data Sources
- Google Marketing Platform
  - Analytics
  - Ads
- YouTube
- Your Datasets

2

Preprocess Features

- Explore
- Join
- Create Train / Test Tables

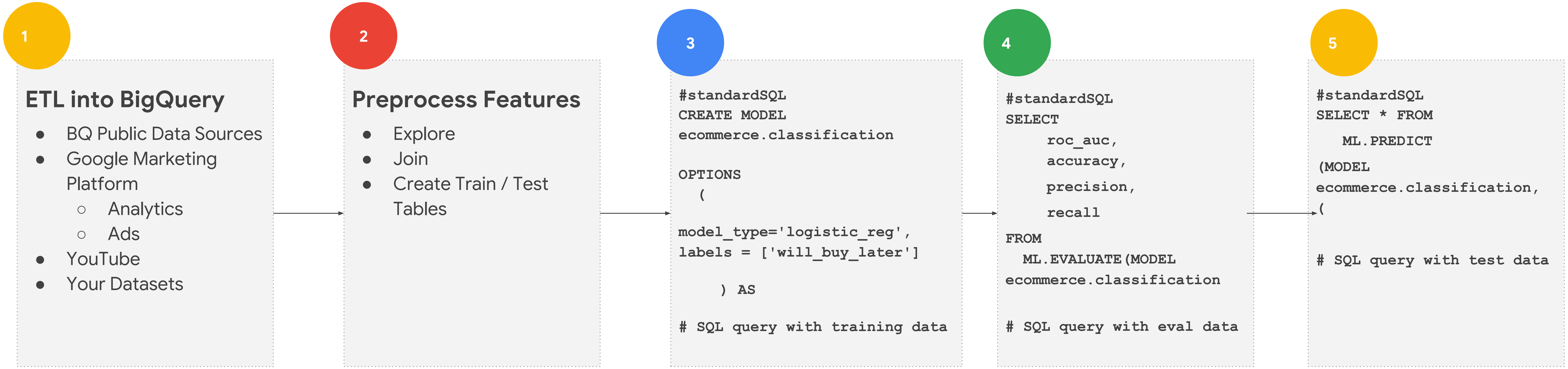
3

```
#standardSQL
CREATE MODEL
ecommerce.classification

OPTIONS
  (

model_type='logistic_reg',
labels = ['will_buy_later']

    ) AS

# SQL query with training data
```

4

```
#standardSQL
SELECT
     roc_auc,
     accuracy,
     precision,
     recall
FROM
  ML.EVALUATE(MODEL
ecommerce.classification

# SQL query with eval data
```

The End-to-End BQML Process



1

ETL into BigQuery

- BQ Public Data Sources
- Google Marketing Platform
    - Analytics
    - Ads
- YouTube
- Your Datasets

2

Preprocess Features

- Explore
- Join
- Create Train / Test Tables

3

```
#standardSQL
CREATE MODEL
ecommerce.classification

OPTIONS
  (

model_type='logistic_reg',
labels = ['will_buy_later']

     ) AS

# SQL query with training data
```

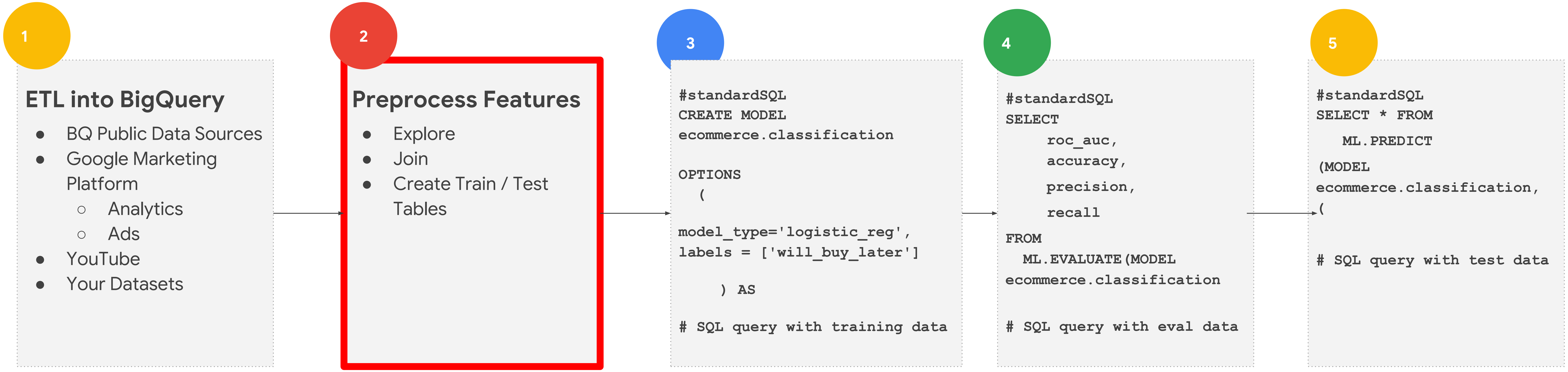
4

```
#standardSQL
SELECT
     roc_auc,
     accuracy,
     precision,
     recall
FROM
  ML.EVALUATE(MODEL
ecommerce.classification

# SQL query with eval data
```

5

```
#standardSQL
SELECT * FROM
   ML.PREDICT
(MODEL
ecommerce.classification,
(

# SQL query with test data
```

Feature Engineering is often the hardest part of ML



1

ETL into BigQuery

- BQ Public Data Sources
- Google Marketing Platform
  - Analytics
  - Ads
- YouTube
- Your Datasets

2

Preprocess Features

- Explore
- Join
- Create Train / Test Tables

3

```
#standardSQL
CREATE MODEL
ecommerce.classification

OPTIONS
  (

model_type='logistic_reg',
labels = ['will_buy_later']

    ) AS

# SQL query with training data
```

4

```
#standardSQL
SELECT
     roc_auc,
     accuracy,
     precision,
     recall
FROM
  ML.EVALUATE(MODEL
ecommerce.classification

# SQL query with eval data
```

5

```
#standardSQL
SELECT * FROM
   ML.PREDICT
(MODEL
ecommerce.classification,
(

# SQL query with test data
```

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Demo: Using BQML to Predict Taxi Fare**

Format: Talking Head

Video Name:
T-BQML-O_5_l1_demo:_using_bqml_to_predict_taxi_fare

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Phases of Building the Model**

Format: Talking Head

Video Name: T-BQML-O_5_l2_phases_of_building_the_model

Building a ML Model involves:

Create the dataset

Build the model

Operationalize the model

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. Select a model
4. Review loss metrics
5. Improve and re-train

Steps in Model Building

1. **Review our goal**
2. Establish benchmark
3. Select a model
4. Review loss metrics
5. Improve and re-train

Our Ecommerce Goal #1

Forecast Monthly Site Visits

Steps in Model Building

1. Review our goal
2. **Establish benchmark**
3. Select a model
4. Review loss metrics
5. Improve and re-train

Benchmark

+- XXXX Visits

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. **Select a model**
4. Review loss metrics
5. Improve and re-train

Model Selection

Linear Regression

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. Select a model
4. **Review loss metrics**
5. Improve and re-train

Review Loss Metrics

Linear Regression uses MSE or RMSE

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. Select a model
4. Review loss metrics
5. **Improve and re-train**

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Demo: Creating a Forecasting Model**

Format: Talking Head

Video Name: T-BQML-O_5_l3_demo:_creating_a_forecasting_model

BQML

BQML has three main features: training, prediction and evaluation

- What can we forecast on our ecommerce dataset? (think numeric)
- What model do we use? (linear regression)
- What is our measure of success? (MSE or RMSE)
- Demo: Linear Regression w BQML
- Intro to BQML

  Demo? BQML example query for taxis
  https://medium.com/@lakshmanok/10ab44a37fbe
  - Use the WITH clause train = 1, eval = 2 for explaining BQML pieces

- Lab: forecast visits by device type, etc. (regression)
  - <LINK TO R STUDIO LAB>

    Predict Bounce Rate Based on Page Load Time (and time on site?)
    https://www.r-bloggers.com/predict-bounce-rate-based-on-page-load-time-in-google-analytics/
    - Try this in BQML
    - https://support.google.com/analytics/answer/3437719?hl=en hits.page.
    - x_id – Id of the page
    - ismobile – page visited is by mobile or not
    - Country
    - pagePath
    - pageTitle
    - avgServerResponseTime
    - avgServerConnectionTime
    - avgRedirectionTime
    - avgPageDownloadTime
    - avgDomainLookupTime
    - avgPageLoadTime
    - Entrances
    - Pageviews
    - Exits

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Lab intro: Forecast Ecommerce Visits in BigQuery ML**

Format: Talking Head

Video Name:
T-BQML-O_5_l4_lab_intro:_forecast_ecommerce_visits_in_bigquery_ml

Lab

Forecast Ecommerce Visits with BigQuery ML

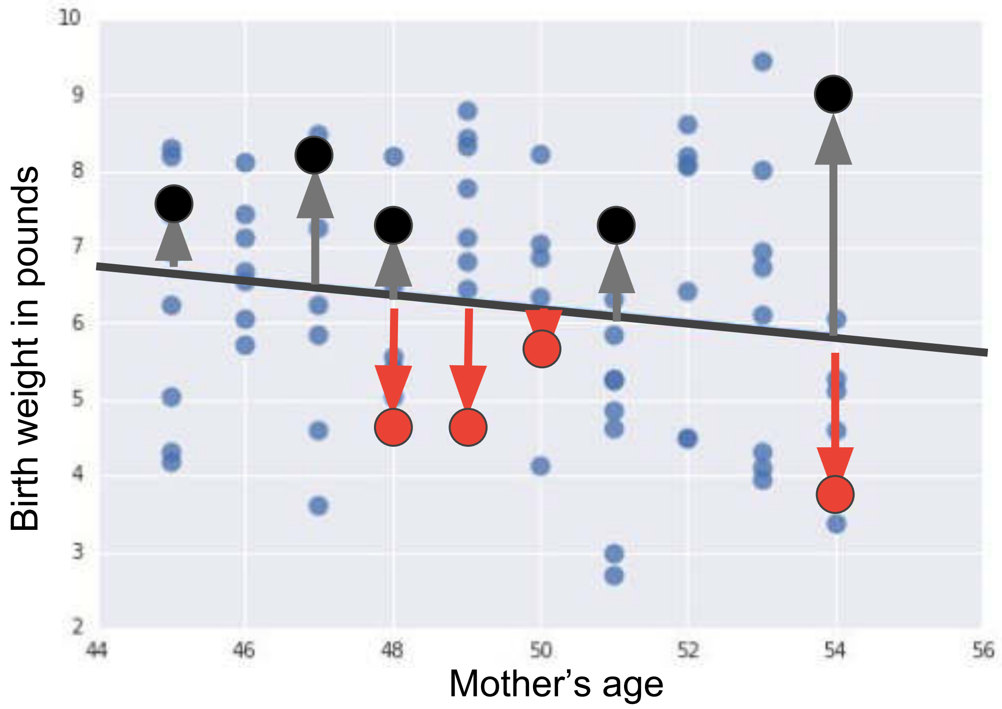
Evan Jones

Forecasting Model:
Calculating Model Error

Error = actual (true) - predicted value

Computed errors:

+0.70
+1.10
+0.65
-1.20
-1.15
+1.10
+3.09
-2.10



Forecasting Model:
Lowest **Root Mean Squared Error**

1. Get the errors for the training examples	2. Compute the squares of the error values
+0.70	0.49
+1.10	1.21
+0.65	0.42
-1.20	1.44
-1.15	1.32
+1.10	1.21
+3.09	9.55
-2.10	4.41

3. Compute the **mean** of the squared error values

2.51

Forecasting Model:
Lowest **Root Mean Squared Error**

1. Get the errors for the training examples	2. Compute the squares of the error values
+0.70	0.49
+1.10	1.21
+0.65	0.42
-1.20	1.44
-1.15	1.32
+1.10	1.21
+3.09	9.55
-2.10	4.41

3. Compute the **mean** of the squared error values

2.51

4. Take a **square root of the mean**

1.58

$$\sqrt{\frac{1}{n} \times \sum_{i=1}^{n} (\hat{Y}_i - Y_i)^2}$$

$\hat{Y}_i$ predicted value

Y_i labeled value

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Lab Solution: Forecast Ecommerce Visits in BigQuery ML**

Format: Talking Head + Lab Screencast

Video Name: T-BQML-O_5_l6_lab_solution:_forecast_ecommerce_visits_in_bigquery_ml

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Creating a Classification Model**

Format: Talking Head

Video Name: T-BQML-O_5_l7_creating_a_classification_model

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. Select a model
4. Review loss metrics
5. Improve and re-train

Steps in Model Building

1. **Review our goal**
2. Establish benchmark
3. Select a model
4. Review loss metrics
5. Improve and re-train

Our Ecommerce Goal #2

Predict whether a user will return within a day

Steps in Model Building

1. Review our goal
2. **Establish benchmark**
3. Select a model
4. Review loss metrics
5. Improve and re-train

Benchmark

70%+ Accurate

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. **Select a model**
4. Review loss metrics
5. Improve and re-train

Model Selection

Logistic Regression

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. Select a model
4. **Review loss metrics**
5. Improve and re-train

Review Loss Metrics

Cross Entropy

Steps in Model Building

1. Review our goal
2. Establish benchmark
3. Select a model
4. Review loss metrics
5. **Improve and re-train**

- What can we classify?
- What model do we use? (logistic regression)
- What is our measure of success? Model performance xentropy vs Criteria performance: (accuracy, precision, recall)
- Lab: Model to predict whether a user will return to the site in 24 hours (logistic)

  Predict If User Will Return within 24 hours
  https://www.tatvic.com/blog/predict-users-return-visit-within-a-day-part-1/
  - Try this in BQML
  - visitor_ID
  - visitCount
  - daysSinceLastVisit
  - Medium
  - landingPagePath
  - exitPagePath
  - pageDepth

  Will they return:
  http://pingax.com/predictive-analysis-ecommerce-part-3/
  https://www.google.com/amp/s/www.tatvic.com/blog/predict-users-return-visit-within-a-day-part-1/amp/

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Demo: Creating a Classification Model**

Format: Talking Head

Video Name:
T-BQML-O_5_I99_demo:_creating_a_classification_model

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Lab intro: Predict User Return Visits in BigQuery ML**

Format: Talking Head

Video Name: T-BQML-O_5_l8_lab_intro:_predict_user_return_visits_in_bigquery_ml

Lab

Predict User Return Visits with BigQuery ML

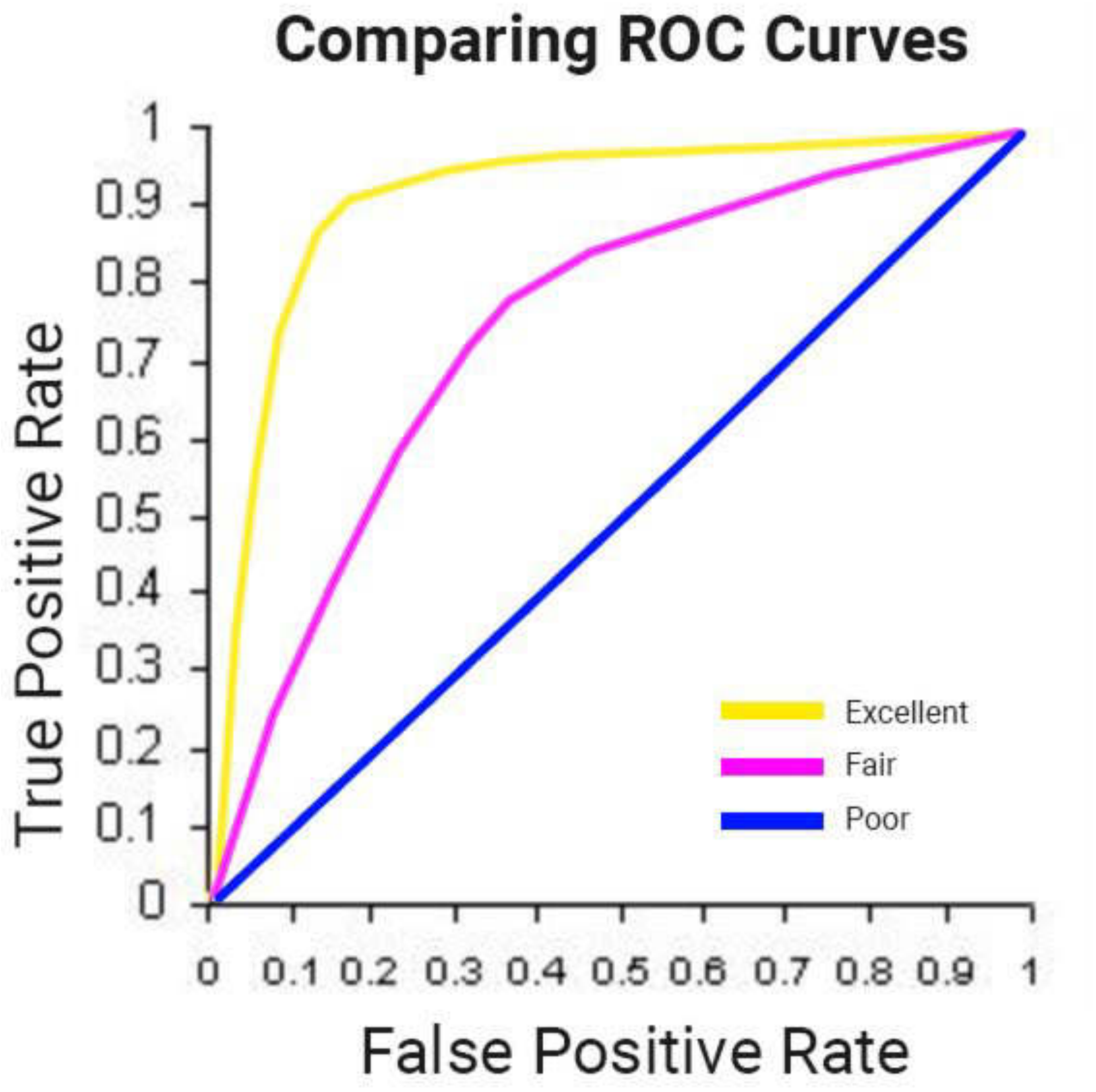
Evan Jones

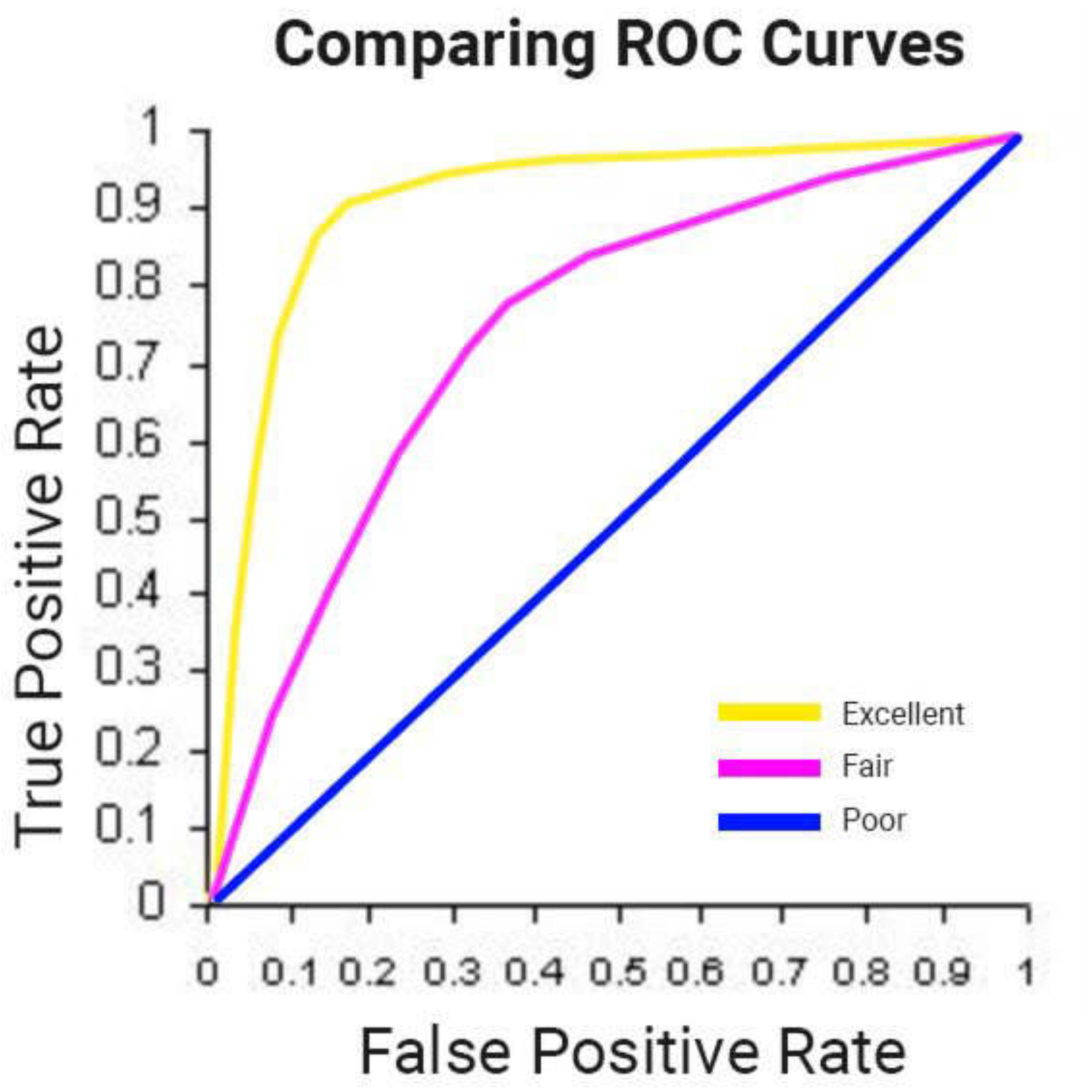
Lab Steps:

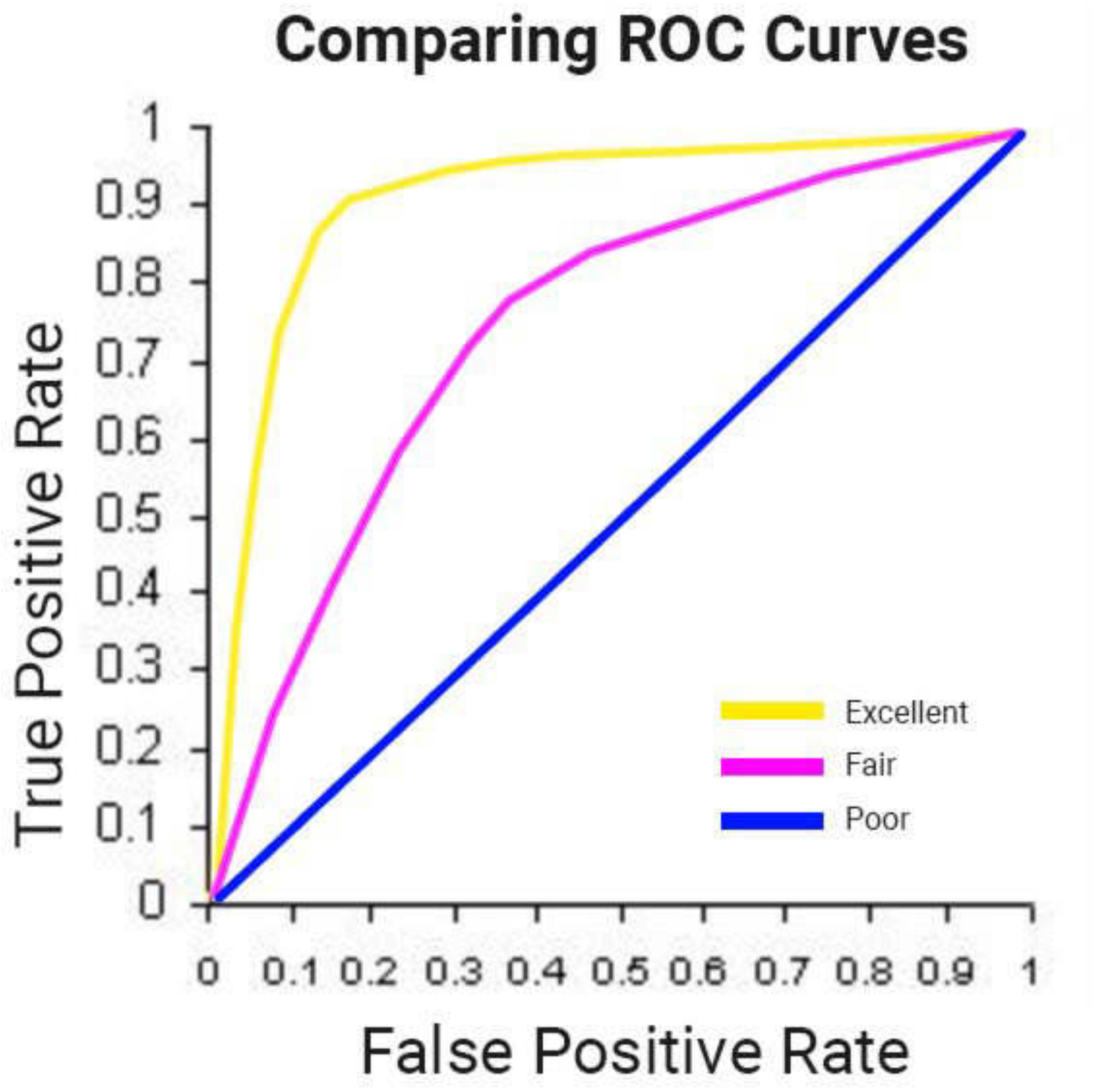
- Explore the dataset features
- Split the data
- Build a Classification Model
- Evaluate it against criteria

True Positive Rate
(where we predicted
the user ***will* return**
and they **actually *did***)

False Positive Rate
(where we predicted
the user ***will* return**
and **they *didn't***)


Comparing ROC Curves
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
True Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
False Positive Rate
Excellent
Fair
Poor


Comparing ROC Curves
True Positive Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
False Positive Rate
Excellent
Fair
Poor


Comparing ROC Curves
True Positive Rate
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
False Positive Rate
Excellent
Fair
Poor

Assess classification model performance with ROC AUC

- .90-1 = excellent (A)
- .80-.90 = good (B)
- .70-.80 = fair (C)
- .60-.70 = poor (D)
- .50-.60 = fail (F)

Course 4: Applying Machine Learning to your Datasets

Module 5: Creating Forecasting and Classification Models in BigQuery

Lesson Title: **Lab Solution: Predict User Return Visits in BigQuery ML**

Format: Talking Head + Lab Screencast

Video Name: T-BQML-O_5_l10_lab_solution:_predict_user_return_visits_in_bigquery_ml

Course 4: Applying Machine Learning to your Datasets

Module 6: End of Course Recap

Lesson Title: **End of Course Recap**

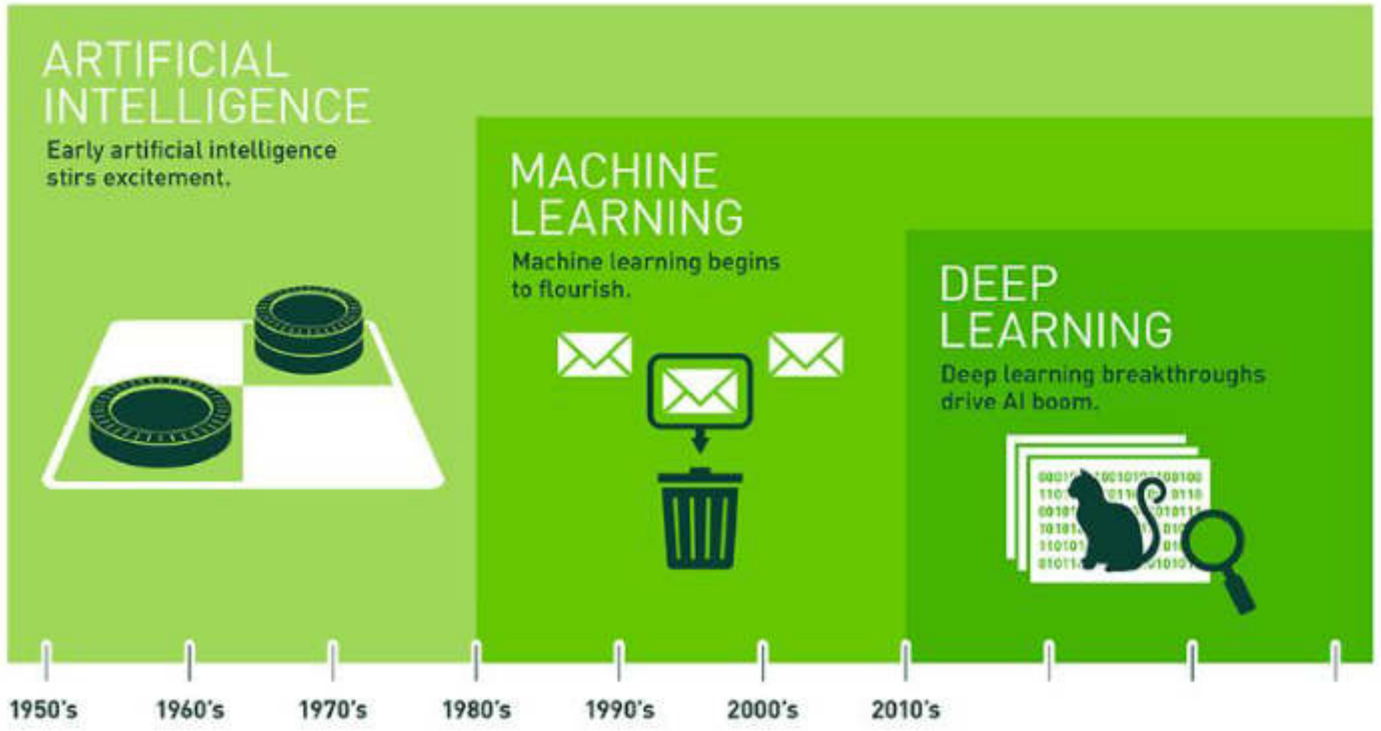
Format: Talking Head

Video Name: T-BQML-O_6_l1_end_of_course_recap

4 Courses in the Data to Insights Specialization

- ✓ 1 - Exploring and Preparing your Data with BigQuery
- ✓ 2 - Creating New BigQuery Datasets and Visualizing Insights
- ✓ 3 - Achieving Advanced Insights with BigQuery
- ✓ 4 - Applying Machine Learning to your Data with GCP

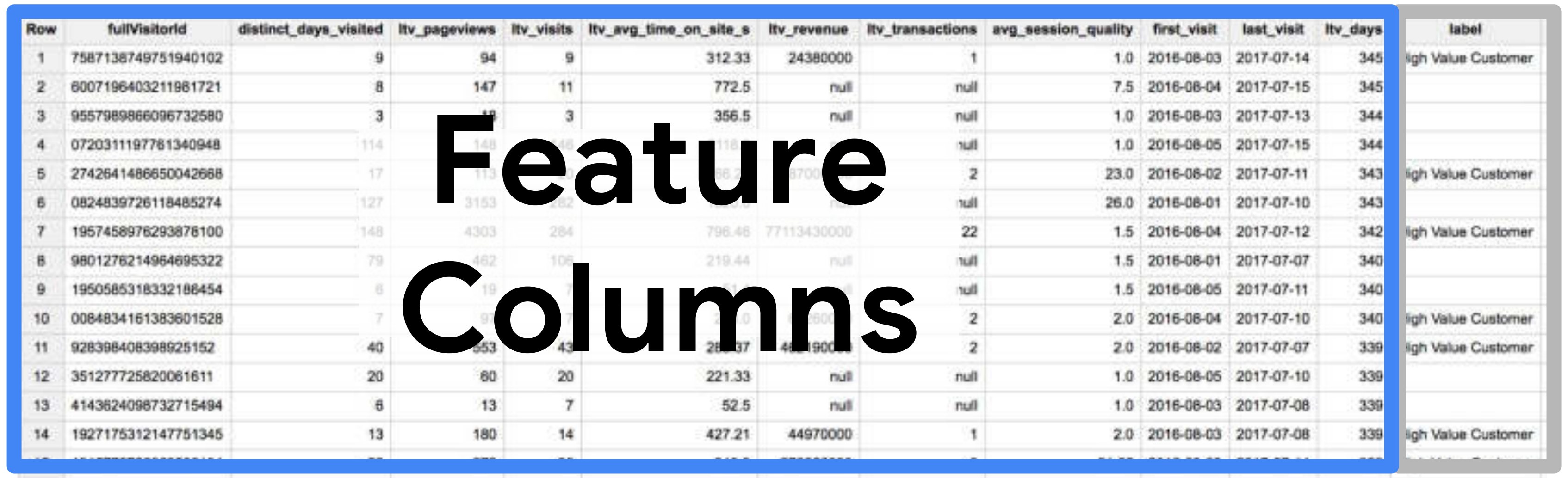
Machine Learning is a discipline inside of AI


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

Source: Cassie Kozyrkov
https://becominghuman.ai/are-you-using-the-term-ai-incorrectly-911ac23ab4f5

ML can transform business operations

Instances, Labels, Feature Columns



Feature Columns

Row	fullVisitorId	distinct_days_visited	ltv_pageviews	ltv_visits	ltv_avg_time_on_site_s	ltv_revenue	ltv_transactions	avg_session_quality	first_visit	last_visit	ltv_days	label
1	7587138749751940102	9	94	9	312.33	24380000	1	1.0	2016-08-03	2017-07-14	345	igh Value Customer
2	6007196403211981721	8	147	11	772.5	null	null	7.5	2016-08-04	2017-07-15	345	
3	9557989866096732580	3	[illegible]	3	356.5	null	null	1.0	2016-08-03	2017-07-13	344	
4	0720311197761340948	114	148	146	[illegible]	[illegible]	null	1.0	2016-08-05	2017-07-15	344	
5	2742641486650042668	17	113	[illegible]	[illegible]	[illegible]	2	23.0	2016-08-02	2017-07-11	343	igh Value Customer
6	0824839726118485274	127	3153	282	[illegible]	[illegible]	null	26.0	2016-08-01	2017-07-10	343	
7	1957458976293878100	148	4303	284	796.46	77113430000	22	1.5	2016-08-04	2017-07-12	342	igh Value Customer
8	9801276214964695322	79	462	106	219.44	null	null	1.5	2016-08-01	2017-07-07	340	
9	1950585318332186454	6	19	7	[illegible]	[illegible]	null	1.5	2016-08-05	2017-07-11	340	
10	0084834161383601528	7	[illegible]	7	[illegible]	[illegible]	2	2.0	2016-08-04	2017-07-10	340	igh Value Customer
11	9283984083989225152	40	[illegible]	43	[illegible]	[illegible]	2	2.0	2016-08-02	2017-07-07	339	igh Value Customer
12	3512777258200616111	20	60	20	221.33	null	null	1.0	2016-08-05	2017-07-10	339	
13	4143624098732715494	6	13	7	52.5	null	null	1.0	2016-08-03	2017-07-08	339	
14	1927175312147751345	13	180	14	427.21	44970000	1	2.0	2016-08-03	2017-07-08	339	igh Value Customer

The 3 Secrets of ML

1. You don’t have to set out to do an ML project

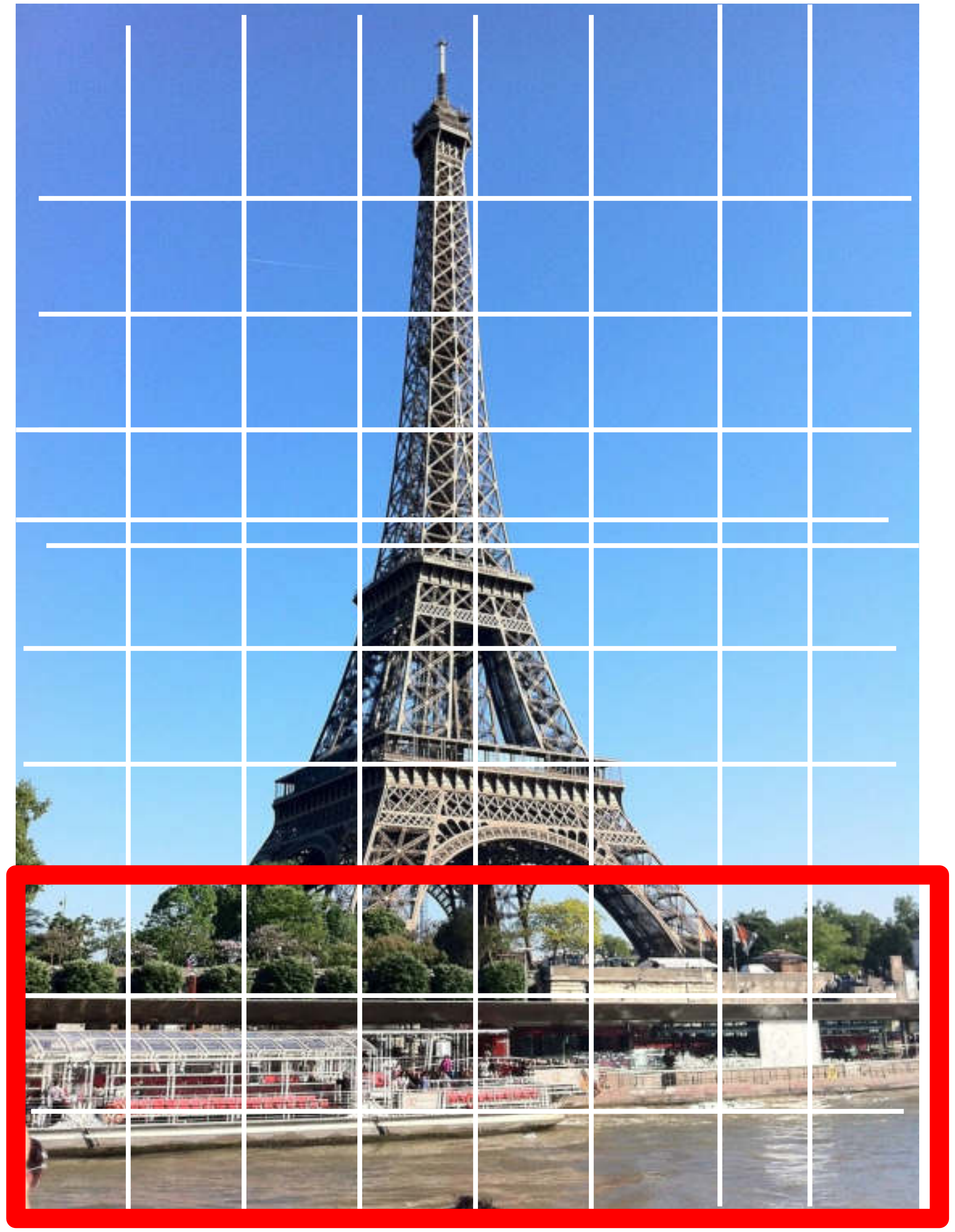
2. It’s not just about training models

3. You need lots of good examples to train from*

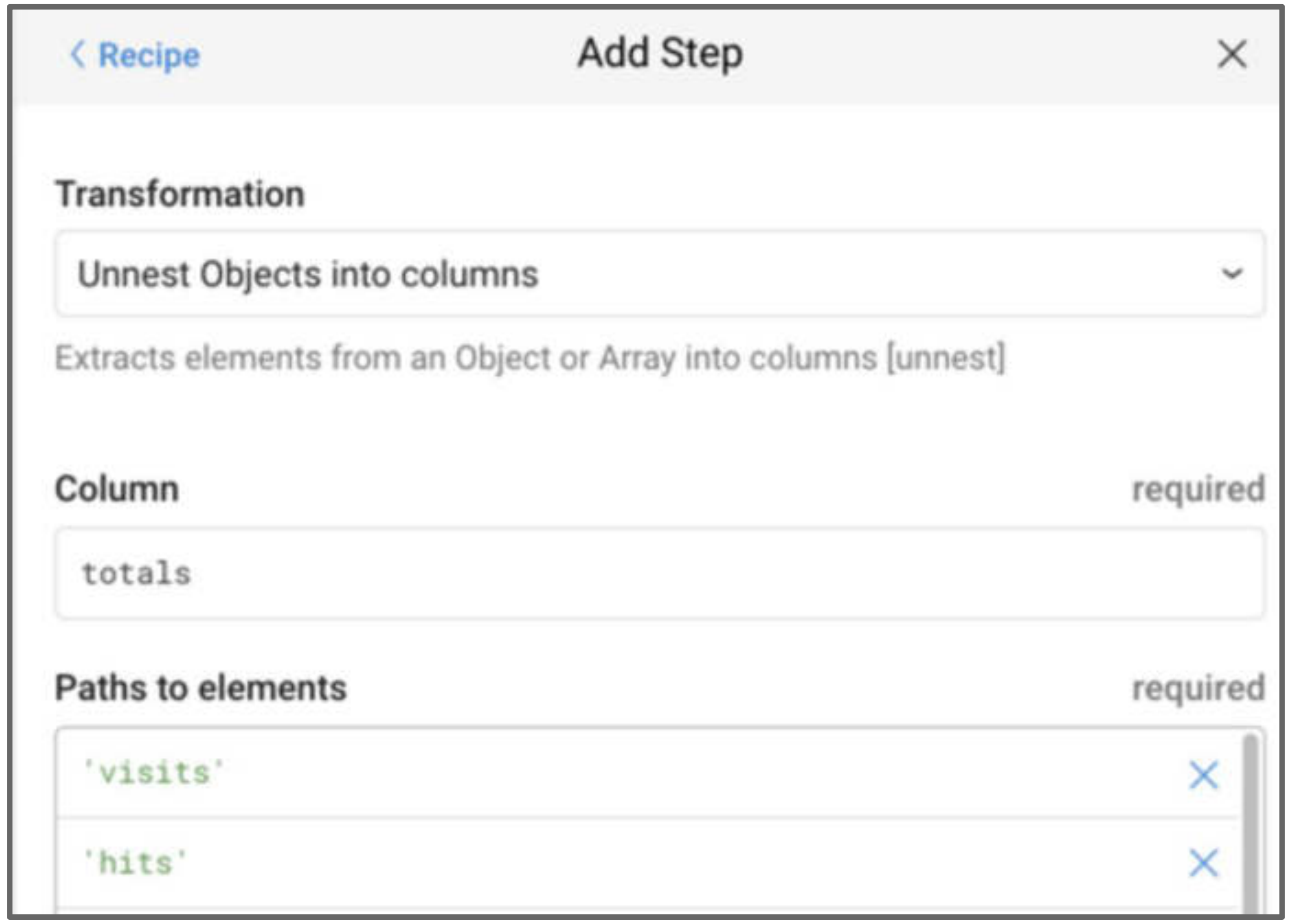
The GCP Machine Learning Tool Spectrum

Advanced Models	Modeling for Analysts	Pretrained Models	Minimal Effort
TensorFlow	**ML on BigQuery (beta)**	**Pretrained ML APIs**	**AutoML (soon)**
● Data Scientists ● Data Engineers	● Data Analysts	● Data Analysts ● Data Scientists ● Data Engineers	● Everyone

Access Pretrained ML APIs for common applications



Advanced Dataprep Transformations



Transformation

Unnest Objects into columns

Extracts elements from an Object or Array into columns [unnest]

Column required

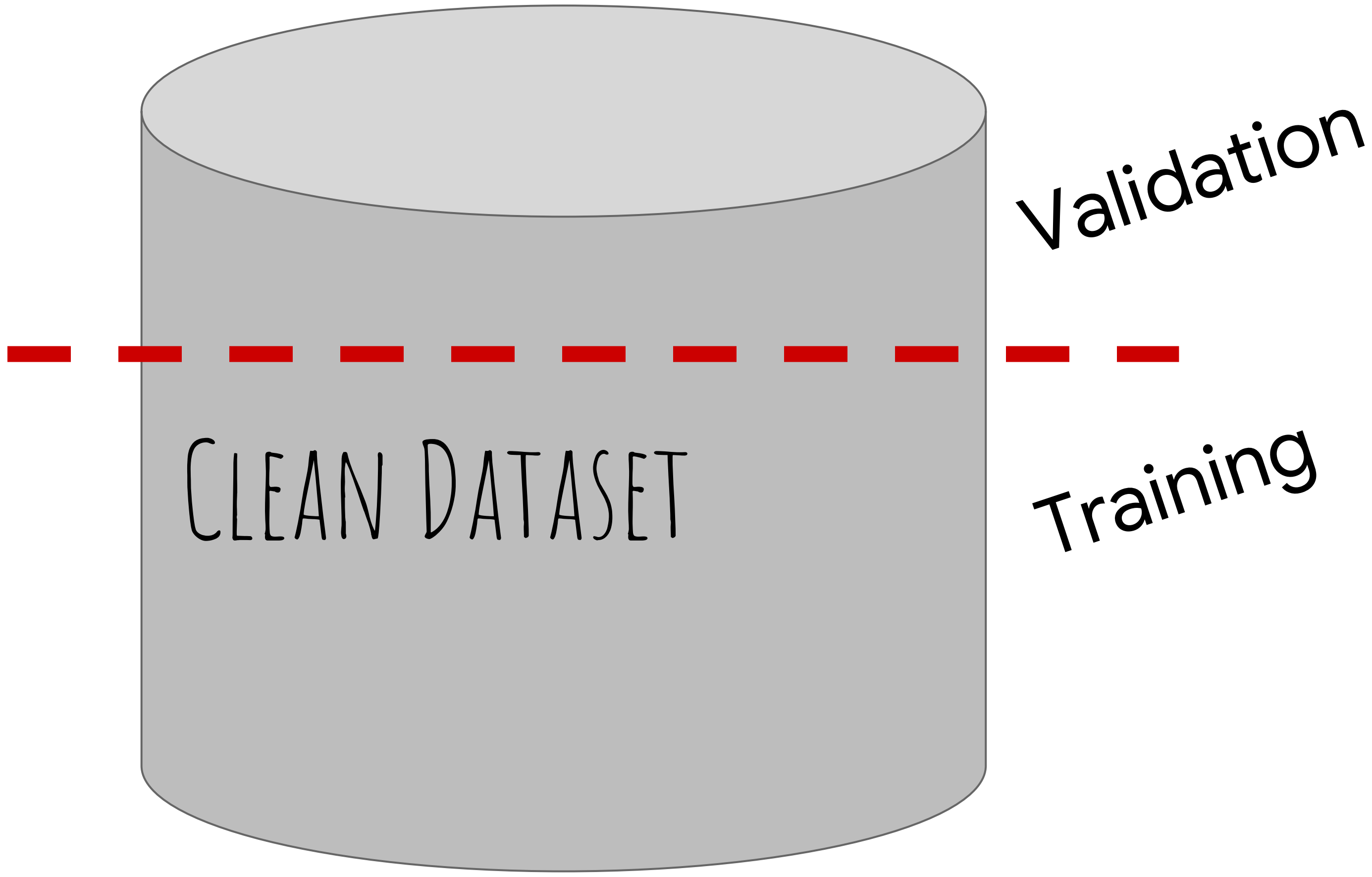
totals

Paths to elements required

'visits'

'hits'

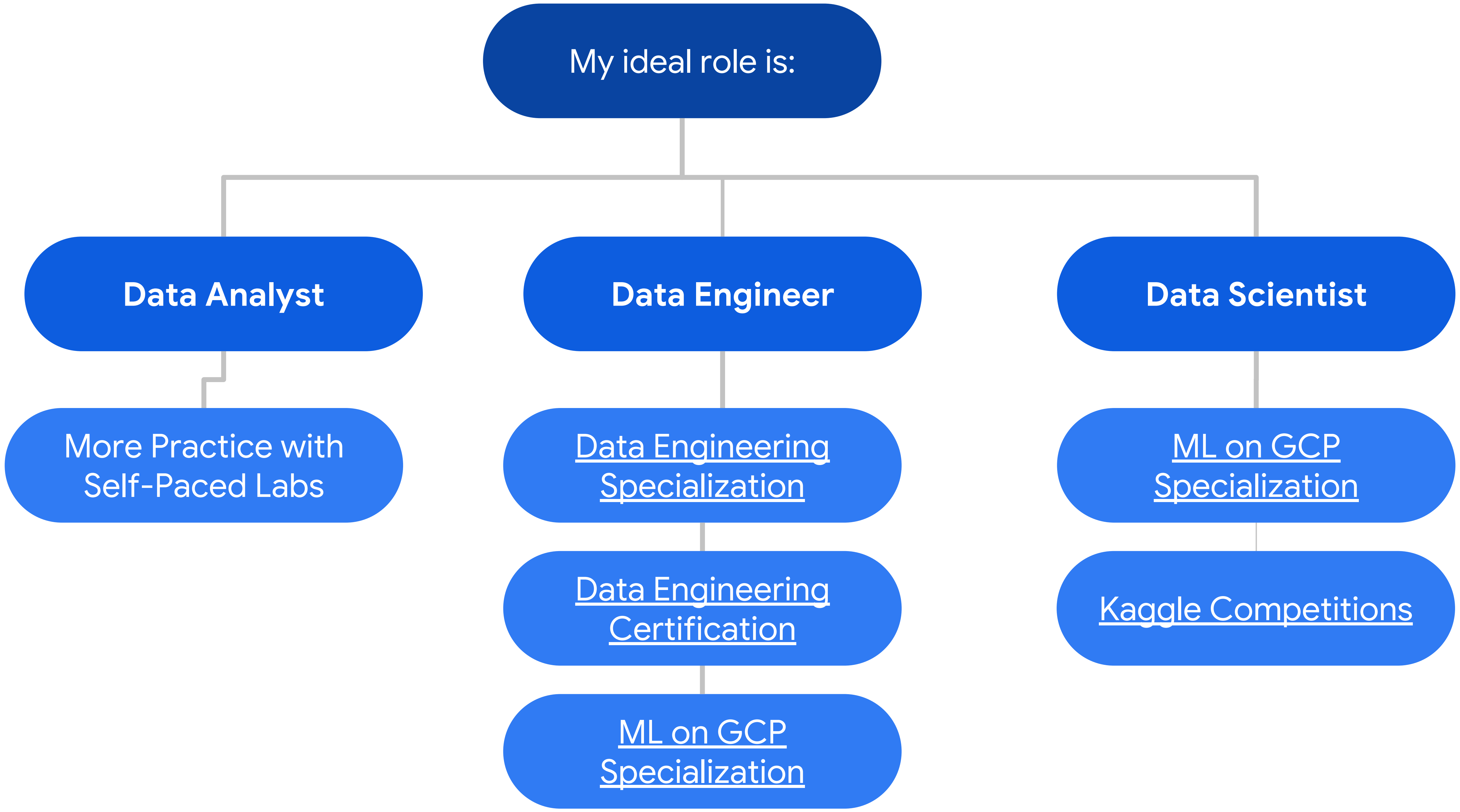
Split your Dataset



Create ML Models inside of BigQuery

BigQuery

Recommended Learning Paths



My ideal role is:

Data Analyst

More Practice with Self-Paced Labs

Data Engineer

Data Engineering Specialization

Data Engineering Certification

ML on GCP Specialization

Data Scientist

ML on GCP Specialization

Kaggle Competitions