

# Cut Video

Course 1: How Google Does ML

Module 5: Python notebooks in the cloud

Lesson Title: **Module Introduction**

Format: Talking head with slides

# Python notebooks in the cloud

## How Google does ML

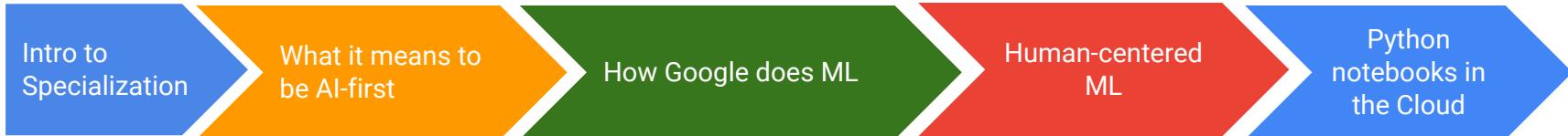
---

Machine Learning on Google Cloud Platform



# Module Learning Objectives

- Carry out data science tasks in notebooks
- Rehost notebooks on the cloud
- Execute ad-hoc queries at scale
- Invoke pre-trained ML models from Datalab



# Agenda

Cloud Datalab

---

Compute Engine and Cloud Storage

---

Data Analysis with BigQuery

---

Machine Learning APIs

---

# Cut Video

Course 1: How Google Does ML

Module 5: Python notebooks in the cloud

Lesson Title: **Cloud Datalab**

Format: Camtasia screenrecording + demo

# Increasingly, data analysis and machine learning are carried out in self-descriptive, shareable, executable notebooks

The screenshot shows a Google Cloud Databricks notebook titled "demandforecast" (autosaved). The interface is divided into four main sections:

- Share:** A green button.
- Code:** A blue button, currently selected. It contains the Python code: 

```
j = data[data['dayofweek'] == 7].plot(kind='scatter', x='maxtemp', y='numtrips')
```
- Output:** An orange box containing a scatter plot of "numtrips" vs "maxtemp". The x-axis ranges from 20 to 100, and the y-axis ranges from 50,000 to 85,000. The plot shows a positive correlation.
- Markup:** A red button. Below it is a red-bordered box containing explanatory text.

**Text in the Markup box:**

Removing the confounding factor does seem to reflect an underlying trend around temperature. But ... the data are a little sparse, don't you think? This is something that you have to keep in mind -- the more predictors you start to consider (here we are using two: day of week and maximum temperature), the more rows you will need so as to avoid overfitting the model.

A typical notebook contains code, charts, and explanations

Image Source:  
Git Logo from Wikipedia

# Demo of Cloud Datalab

# Cut Video

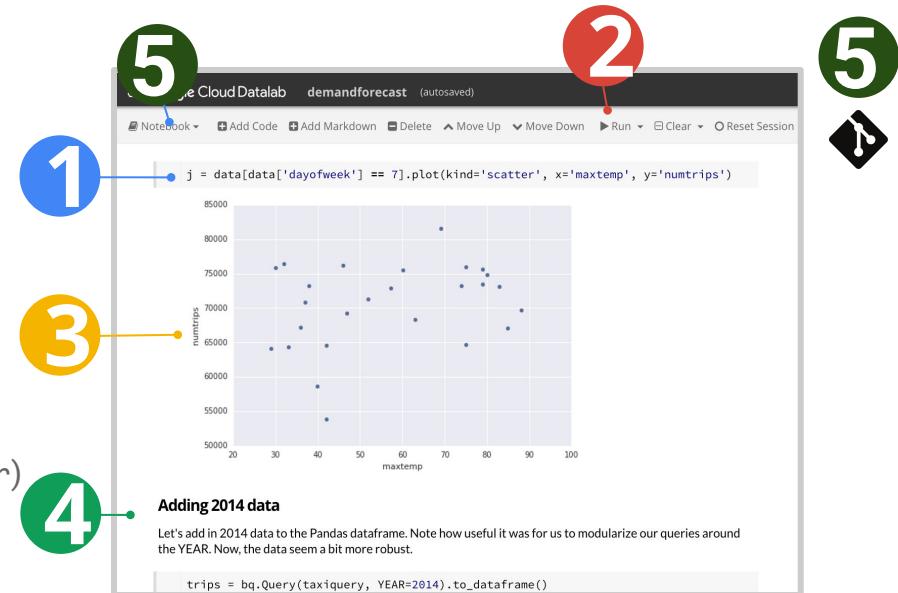
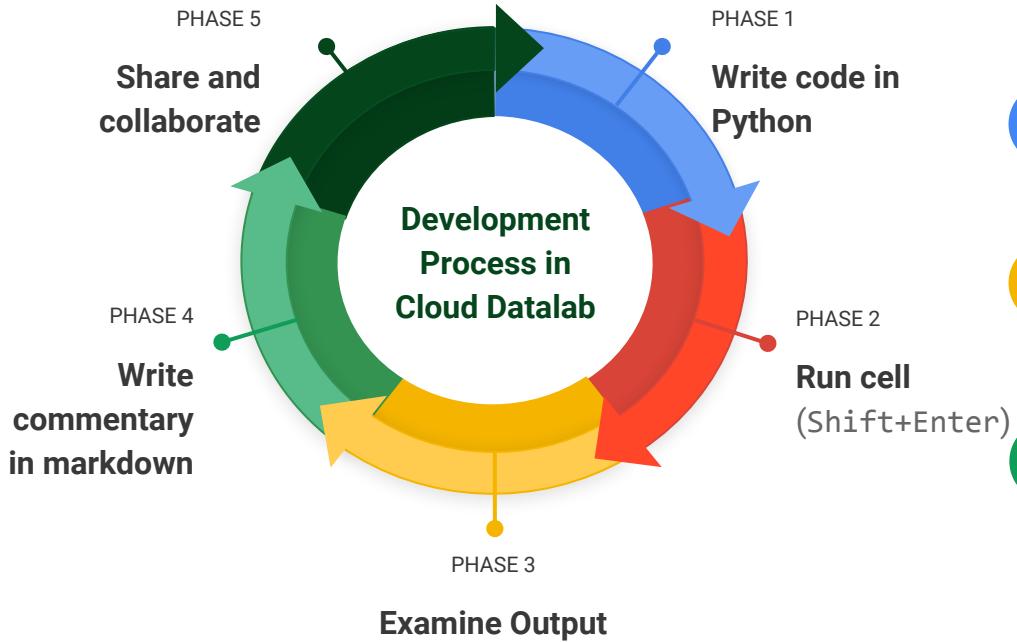
Course 1: How Google Does ML

Module 5: Python notebooks in the cloud

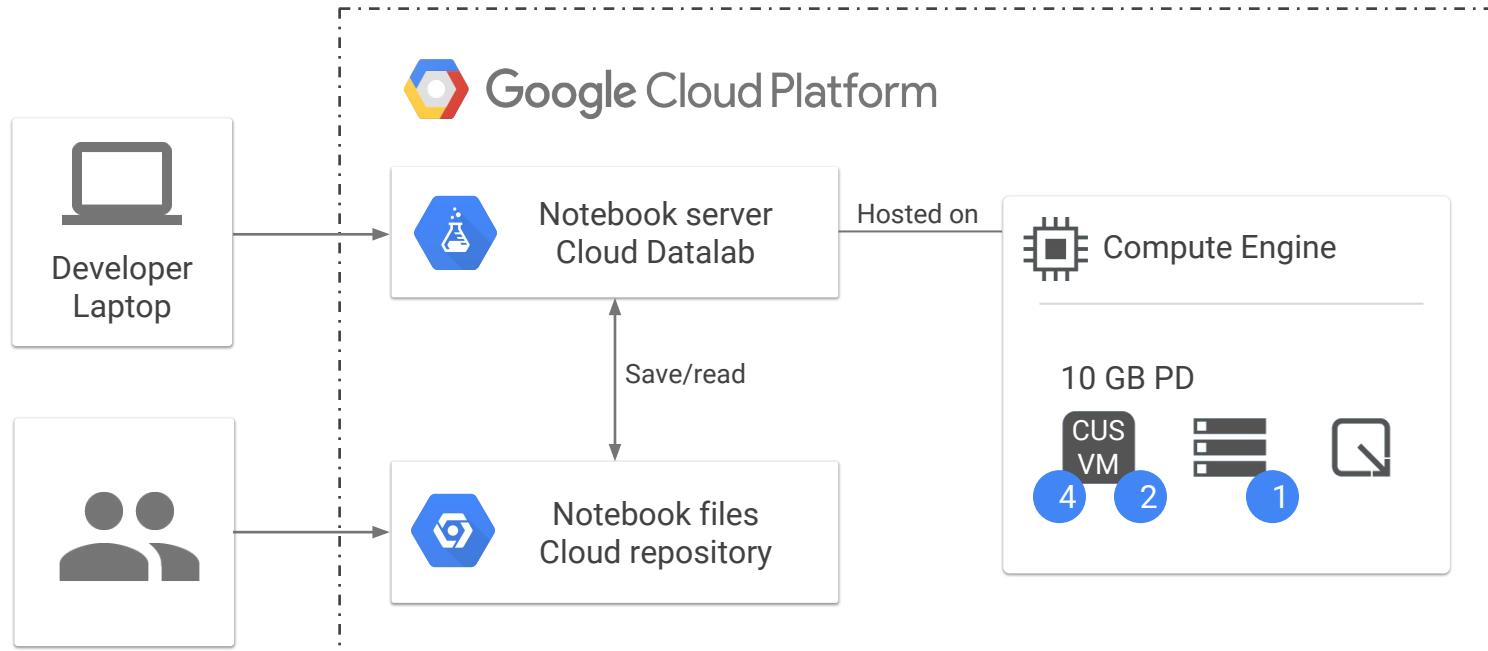
Lesson Title: **Development process**

Format: Camtasia screenrecording + demo

# Datalab notebooks are developed in an iterative, collaborative process



# Datalab notebooks let you change the underlying hardware



# Demo of rehosting Cloud Datalab

# Cut Video

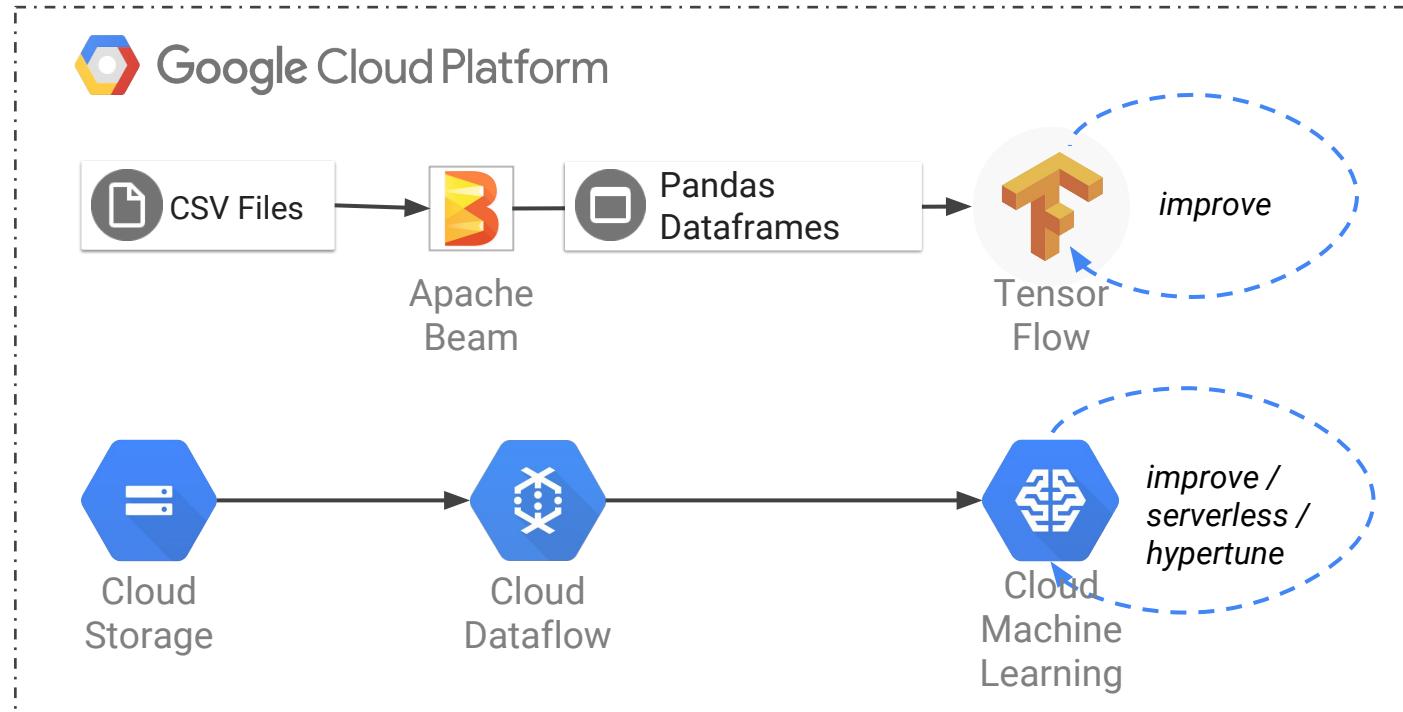
Course 1: How Google Does ML

Module 5: Python notebooks in the cloud

Lesson Title: **Working with managed services**

Format: Camtasia screenrecording

# You can develop locally with Datalab and then scale out data processing to the cloud



# Datalab integrates well with Google Cloud Platform products

Exploring and Analyzing

BigQuery, Google Cloud Storage

Machine Learning and Modeling

TensorFlow and GCML

Visualizing

Google Charts or Plotly or matplotlib

Seamless product combination

CMLE, Dataflow, CloudStorage

Integration

authentication and code source control

# Starting Cloud Datalab in Cloud Shell is quite simple ...

The screenshot shows the Google Cloud Platform interface with the following elements:

- Header:** Google Cloud Platform, Project dropdown, search icon, help icon, notifications icon, and more options icon.
- Left Sidebar:** Home icon, Activity tab, and a large green circle with the number **1**.
- Main Content Area:** A terminal window containing the command:

```
dataLab create my-datalab-vm \
--machine-type n1-highmem-8 \
--zone us-central1-a
```

A blue circle with the number **2** is positioned over the terminal window.
- Bottom Left:** A dropdown menu with options: Preview on port 8080, Change port, and About web preview. A yellow circle with the number **3** is positioned over the "Change port" option. The "Port 8081" option is highlighted with a red border.
- Bottom Right:** A callout box with a blue border containing the text: "what happens to your work when you are through? You do want to stop paying for the datalab machine ...".

# Cut Video

Course 1: How Google Does ML

Module 5: Python notebooks in the cloud

Lesson Title: **Compute and storage**

Format: Camtasia screenrecording

# Agenda

Cloud Datalab

---

Compute Engine and Cloud Storage

---

Data Analysis with BigQuery

---

Machine Learning APIs

---

# Google Cloud provides an earth-scale computer

Networking



FASTER (US, JP, TW) 2016

Data storage



Compute power



Indigo (SG, ID, AU) 2019

- Network
- Network sea cable investments
- Edge points of presence >100
- Edge node locations >1000

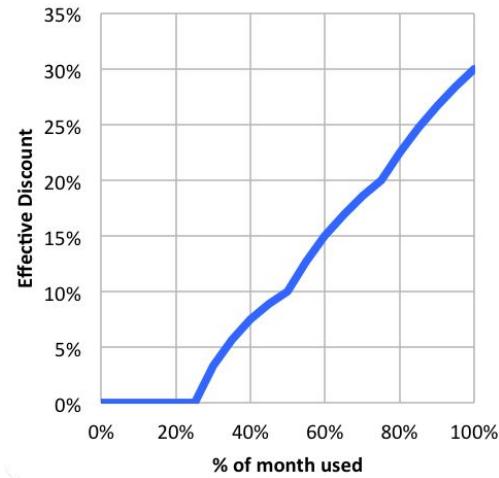
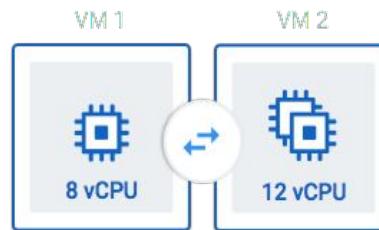
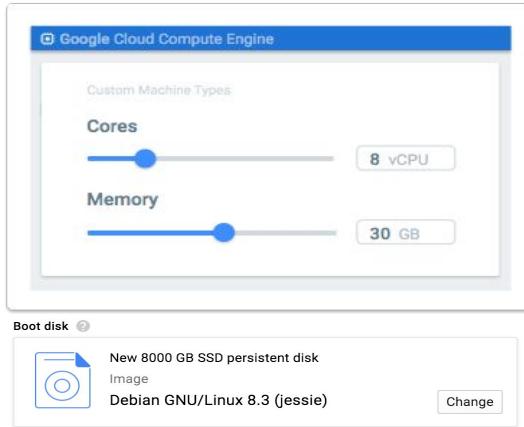
Monet (US, BR) 2017

Junior (Rio, Santos) 2017

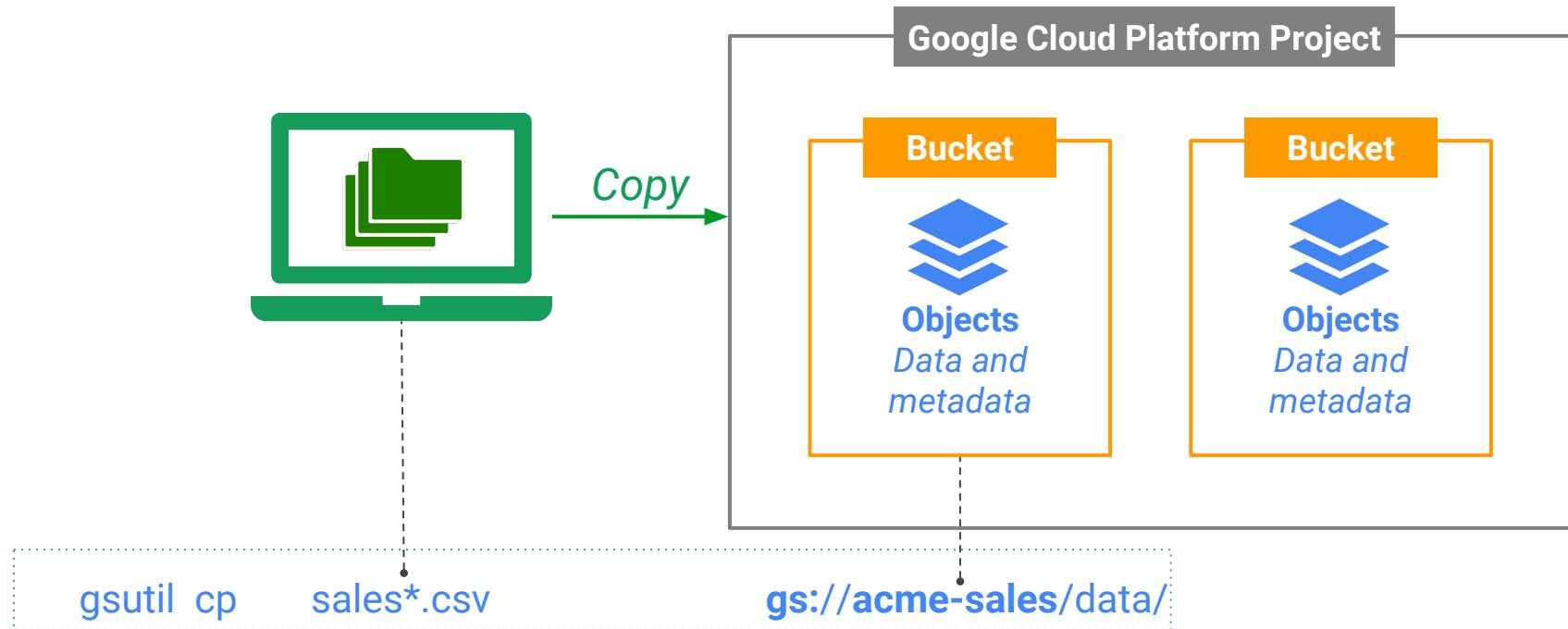
Tannat (BR, UY, AR) 2017

Google Cloud

# Compute Engine provides customizable machine types and flexible compute options

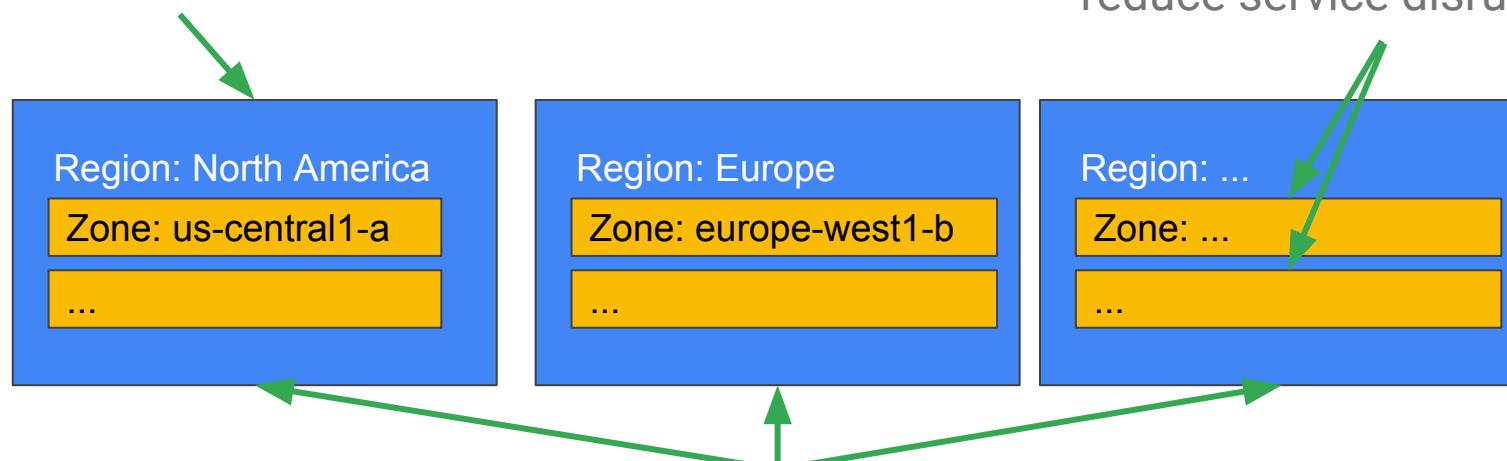


Cloud Storage is durable, persistent and organized in buckets



# Control latency and availability with zones and regions

Choose the closest zone/region so as to to reduce latency.



Distribute your apps and data across zones to reduce service disruptions.

Distribute your apps and data across regions for global availability.

# Cut Video

## Set up Qwiklabs

**NOTE TO VID EDITOR:**  
Lak intro lab

# Lab environment

# Use Qwiklabs to get a temporary GCP account

Welcome! You have signed up successfully.

In-Session Class: ML Immersion

Class Details

Lab 0 : Getting started in Qwiklabs

Lab 0 : Getting started in Qwiklabs

Select 1 Credit

In this lab you will use the Qwiklabs virtual classroom interface to access the Google Cloud Platform, launch Databus and clone the repository, and create a

Duration: 58000  
Access Time: 58000  
Setup Time: 0  
Level: introductory

Rate Lab: Lab 0 : Getting started in Qwiklabs

Start Lab

TIME REMAINING 40 days

© for the use of: ajayvshingrahi@gmail.com

## GETTING STARTED IN QWIKLABS

### Overview

In this lab, you

1. get a quick introduction to Qwiklabs,
2. learn how to get into GCP with your Qwiklabs generated account,
3. share your project info and user credentials with instructor

Primary Instruction

for the use of..

Getting started in Qwiklabs

Overview

An introduction to Qwiklabs

To use Qwiklabs, you need a standard internet browser.

Plan the timing of your lab.

Logging in to Google Cloud Platform

Step 1: Locate the Username, Password and Project Id

Step 2: Browse to Console

Step 3: Sign in to Console

Step 4: Accept the conditions

Step 5: Don't change the password

Step 6 Agree to the Terms of Service

Step 7: Console opens

Step 8: Switch project (if necessary)

# Come back to Qwiklabs to practice; do our Quests

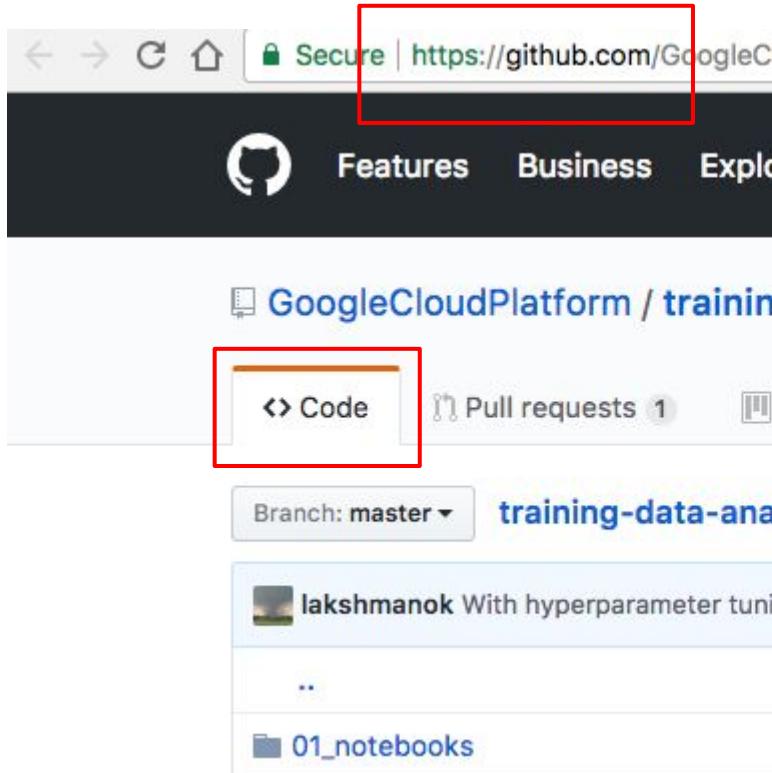
The screenshot shows the Qwiklabs catalog interface. At the top, there's a navigation bar with icons for back, forward, refresh, and home, followed by a secure connection indicator and the URL <https://google.qwiklabs.com/catalog>. Below the URL is the Qwiklabs logo and a search bar labeled "Search for labs". On the left, there's a sidebar with icons for dashboard, search, recent, calendar, users, and help. The main content area has a title "QUESTS (8)" which is highlighted with a red box. Next to it is "LABS (67)". Below this, a descriptive text reads: "Quests are a series of labs organized by technologies, specific services, or particular skills. Come back often as we continue to expand our catalog of". Underneath, there's a table-like structure with four rows, each representing a quest. The first row is "GCP Essentials" with a hexagonal icon containing a multi-layered cake. The second row is "Kubernetes in the Google Cloud" with a hexagonal icon containing a gear. The third row is "Scientific Data Processing" with a hexagonal icon containing a line graph. This row is also highlighted with a red box. The fourth row is "Deployment Manager" with a hexagonal icon containing a monitor. To the right of the quest titles, there are small descriptions and links.

QUEST TITLE
GCP Essentials
Kubernetes in the Google Cloud
Scientific Data Processing
Deployment Manager

# Source code for labs is on GitHub

[https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/machine\\_learning/deeplearning](https://github.com/GoogleCloudPlatform/training-data-analyst/tree/master/courses/machine_learning/deeplearning)

Later, practice taking the lab apart and trying to build it yourself on your own GCP account (strongly recommended)



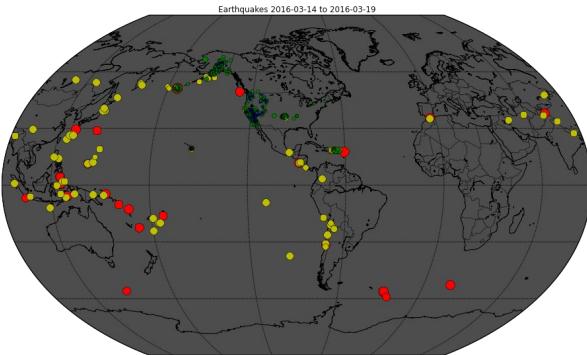
# Cut Video

## Lab: Rent-a-VM

# Lab: Rent-a-VM to process earthquakes data

# Lab: Rent-a-VM to process earthquakes data

In this lab you spin up a virtual machine, configure its security, access it remotely, and then carry out the steps of an ingest-transform-and-publish data pipeline manually.



1. Create a Compute Engine instance with the necessary Access and Security

3. Ingest data into a instance and then transform it

2. SSH into the instance  
Install the software package Git (for source code version control)

4. Store the transformed data on Cloud Storage;  
Publish data to the web

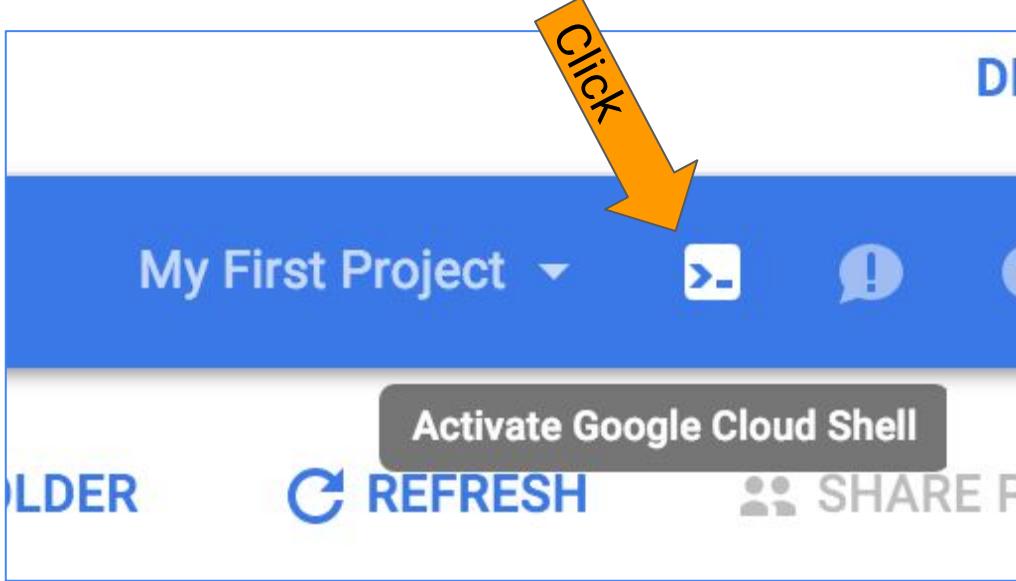
# Cut Video

## Lab debrief

# Cut Video

# CloudShell

# Follow-along: Cloud Shell gives you an easy command-line



- Pre-installed with the tools, libraries, and so on you need to interact with Google Cloud Platform (including a code editor)

# Cut Video

## 3rd wave of cloud

# Agenda

Cloud Datalab

---

Compute Engine and Cloud Storage

---

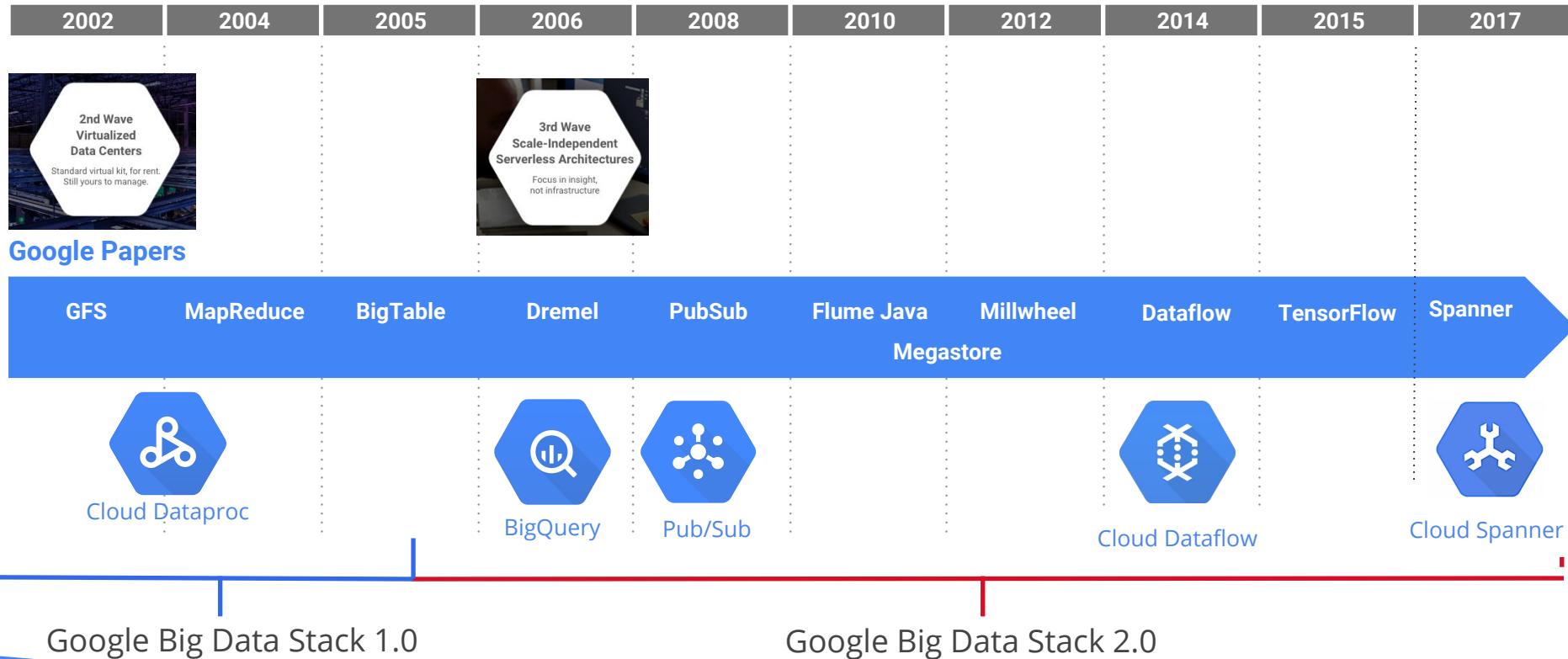
Data Analysis with BigQuery

---

Machine Learning APIs

---

# Spinning up VMs yourself doesn't scale ... what you want are managed services that autoscale for you



# Demo: Query large datasets within BigQuery

# Demo: Query large datasets in seconds

```
#standardsql

# medicare claims in 2014
SELECT
    nppes_provider_state AS state,
    ROUND(SUM(total_claim_count) / 1e6) AS total_claim_count_millions
FROM
    `bigquery-public-data.medicare.part_d_prescriber_2014`
GROUP BY
    state
ORDER BY
    total_claim_count_millions DESC
LIMIT 5;
```

Row	state	total_claim_count_millions
1	CA	116.0
2	FL	91.0
3	NY	80.0
4	TX	76.0
5	PA	63.0

# BigQuery offers...

1

Interactive analysis of petabyte scale databases

2

Familiar, SQL 2011 query language and functions

3

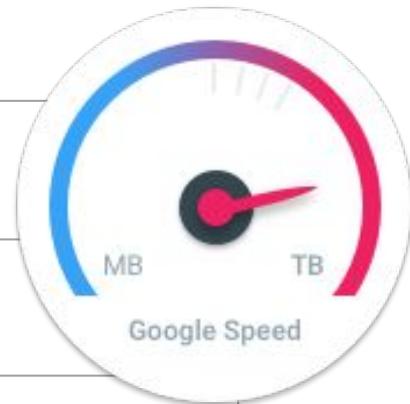
Many ways to ingest, transform, load, export data to/from BigQuery

4

Inexpensive data storage; queries charged on amount of data processed

5

Integration with Datalab for your data analysis needs



# Cut Video

## Lab: Datalab and BigQuery

# Lab: Analyzing data using BigQuery and Datalab

# Lab: Analyzing data Using Datalab and BigQuery

In this lab, you analyze a large dataset using Google BigQuery and Cloud Datalab.

**(70 million rows; 8 GB)**



1. Launch Cloud Datalab



2. Invoke a BigQuery query



3. Create graphs in Datalab

Notes to Video Editor:  
Cut Video: ML, not rules  
Insert: Lak talking head intro Sara

# Agenda

Cloud Datalab

---

Compute Engine and Cloud Storage

---

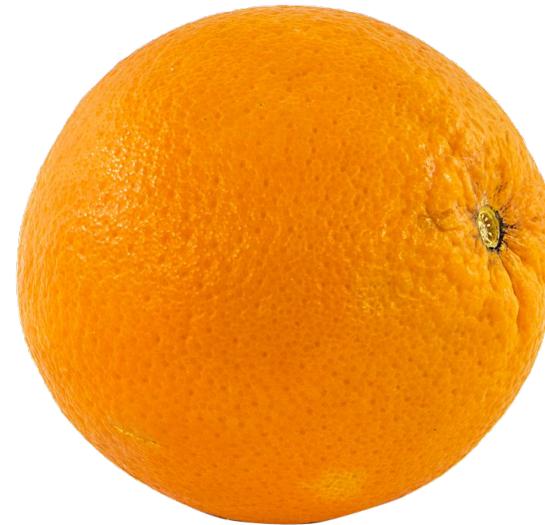
Data Analysis with BigQuery

---

Machine Learning APIs

---

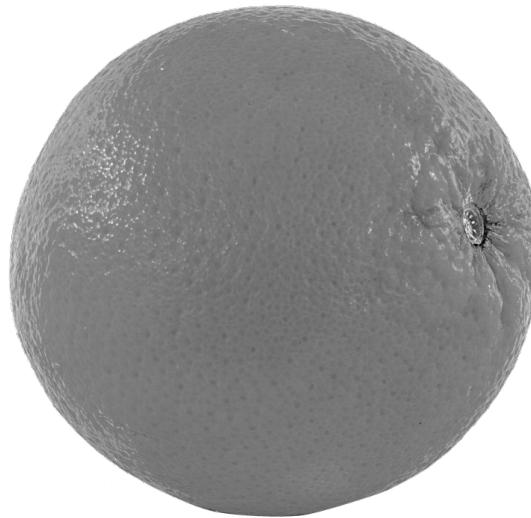
# How would we do this without ML?



CC-BY-SA 2.0 Wikimedia Commons

[https://commons.wikimedia.org/wiki/File:Apple\\_in\\_lightbox.png](https://commons.wikimedia.org/wiki/File:Apple_in_lightbox.png)

# How would we do this without ML?



CC-BY-SA 2.0 Wikimedia Commons

[https://commons.wikimedia.org/wiki/File:Apple\\_in\\_lightbox.png](https://commons.wikimedia.org/wiki/File:Apple_in_lightbox.png)

# How would we do this without ML?



CC-BY-SA 2.0 Wikimedia Commons  
[https://commons.wikimedia.org/wiki/File:Apple\\_in\\_lightbox.png](https://commons.wikimedia.org/wiki/File:Apple_in_lightbox.png)

# What about a dog... and a mop? Easy, right?



[CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Mop\\_and\\_bucket.jpg](https://commons.wikimedia.org/wiki/File:Mop_and_bucket.jpg)

# Not so fast...

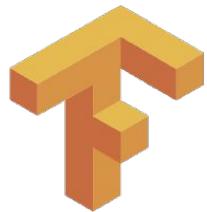


[CC-BY-SA-2.5](https://commons.wikimedia.org/wiki/File:Komondor_Westminster_Dog_Show_crop.jpg) Wikimedia Commons [https://commons.wikimedia.org/wiki/File:Komondor\\_Westminster\\_Dog\\_Show\\_crop.jpg](https://commons.wikimedia.org/wiki/File:Komondor_Westminster_Dog_Show_crop.jpg)  
[CC-BY-2.0](https://commons.wikimedia.org/wiki/File:2014_Westminster_Kennel_Club_Dog_Show_(12487315865).jpg) Wikimedia Commons [https://commons.wikimedia.org/wiki/File:2014\\_Westminster\\_Kennel\\_Club\\_Dog\\_Show\\_\(12487315865\).jpg](https://commons.wikimedia.org/wiki/File:2014_Westminster_Kennel_Club_Dog_Show_(12487315865).jpg)  
[CC-BY-2.0](https://www.flickr.com/photos/petsadviser-pix/16395099127) Petful <https://www.flickr.com/photos/petsadviser-pix/16395099127>  
[CC-BY-SA-2.0](https://www.flickr.com/photos/denverjeffrey/6903790333) Jeffrey Beall <https://www.flickr.com/photos/denverjeffrey/6903790333>

# Two ways to add ML to your apps

Friendly machine learning

## Custom ML models



TensorFlow



Machine Learning  
Engine



Vision API



Speech API



Translation  
API



Natural  
Language API



Video  
Intelligence API

# Cut Video

# Cloud Vision API

# Cloud Vision

Complex image detection with a  
simple REST request

# GIPHY: Cloud Vision in production



<http://engineering.giphy.com>



Label & web  
detection



OCR



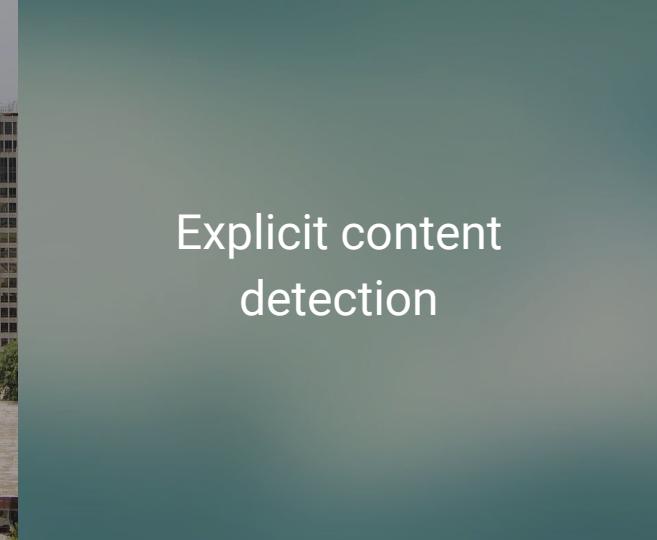
Logo  
detection



Landmark  
detection

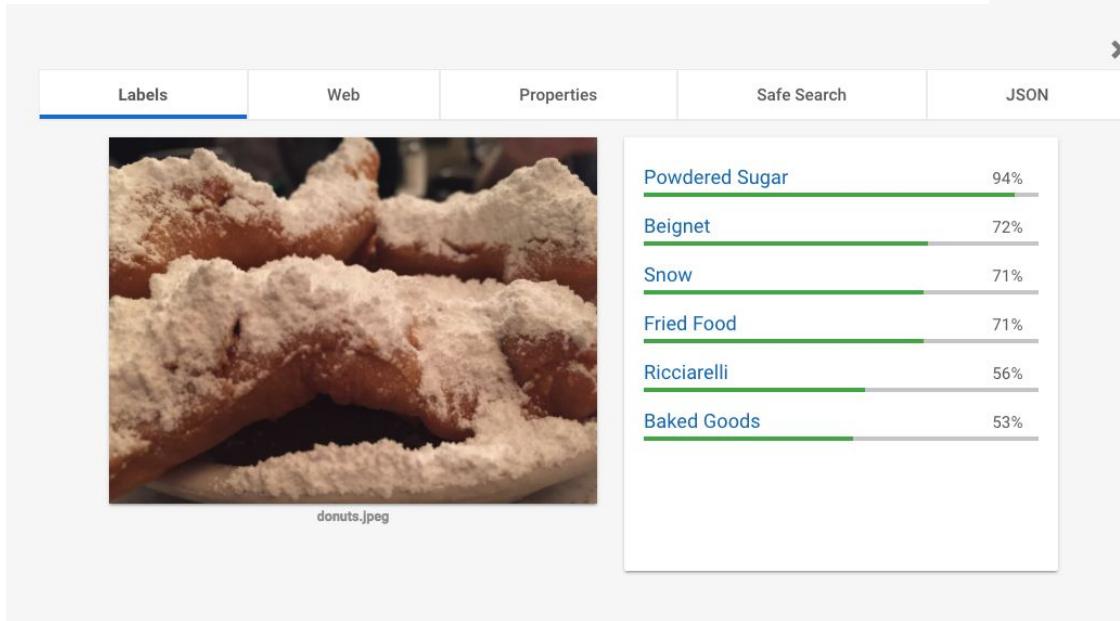


Crop  
hints



Explicit content  
detection

# You can try out the ML APIs in your browser



[cloud.google.com/vision](https://cloud.google.com/vision)

# Vision API demo

# Cut Video

# Video Intelligence API

# Cloud Video Intelligence

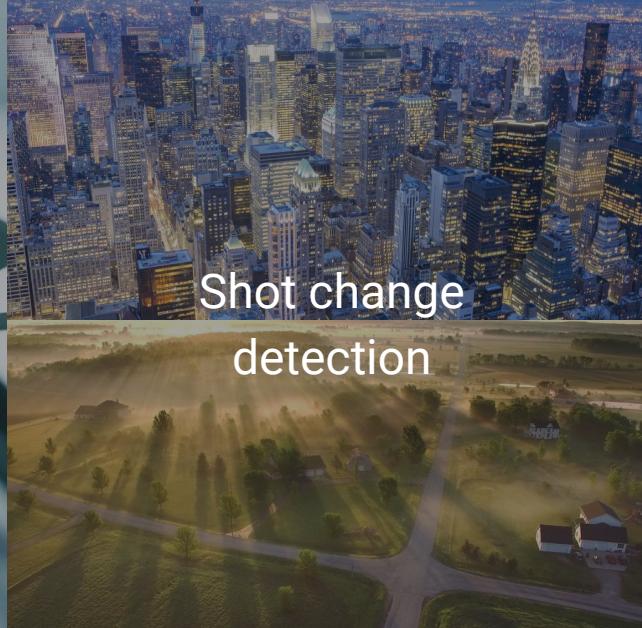
Understand your video's entities  
at shot, frame, or video level



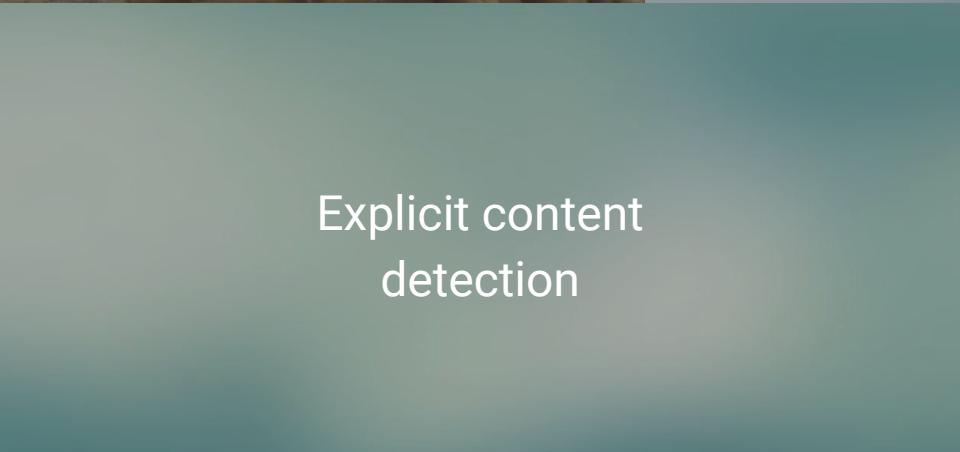
Label  
detection



Video &  
scene-level  
annotations



Shot change  
detection



Explicit content  
detection



Regionalization

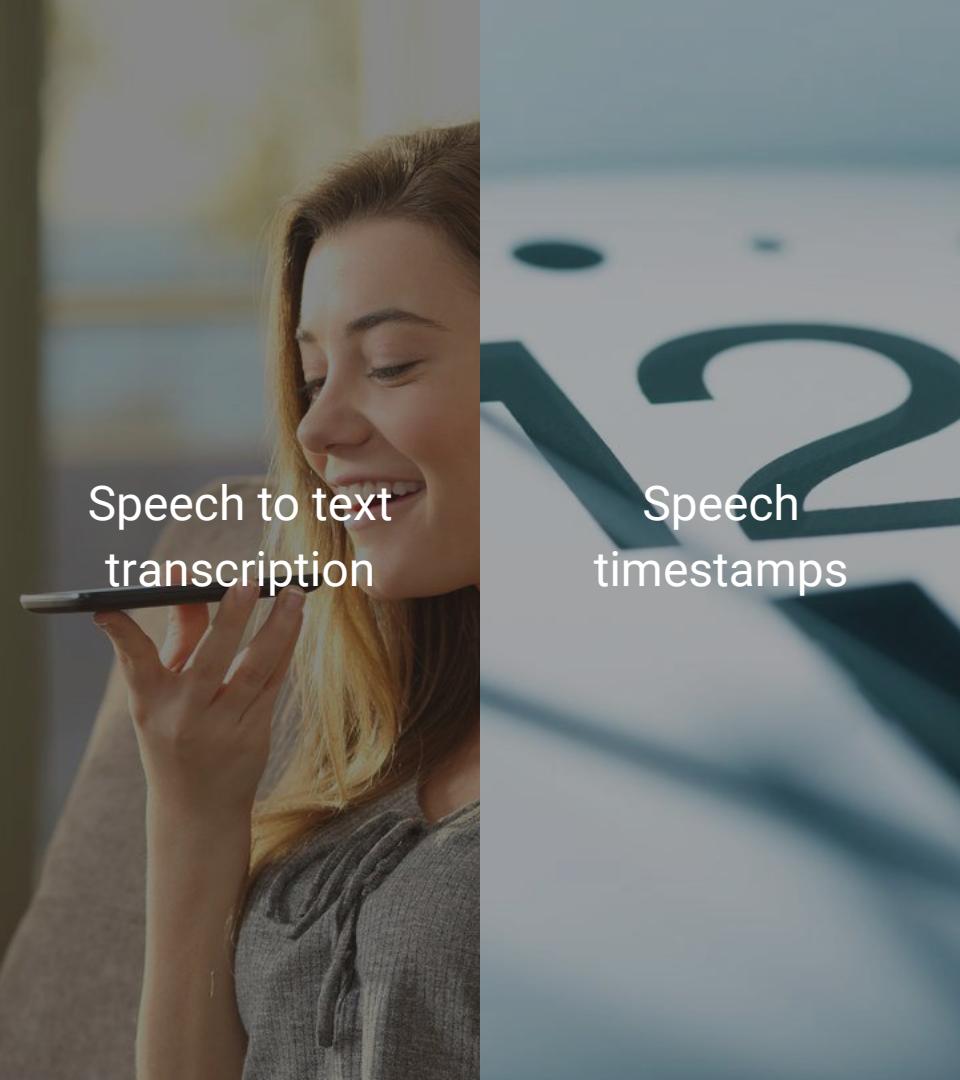
# Video API demo

# Cut Video

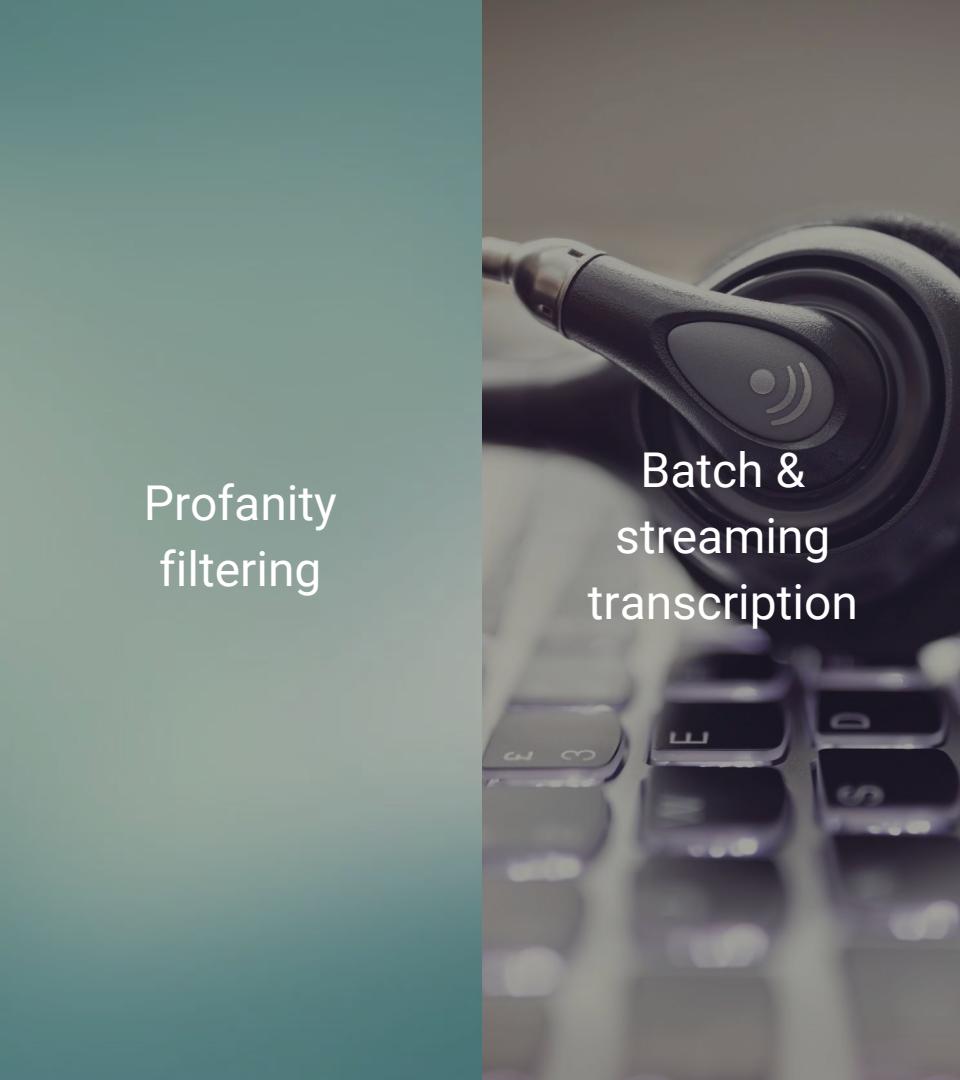
# Cloud Speech API

# Cloud Speech

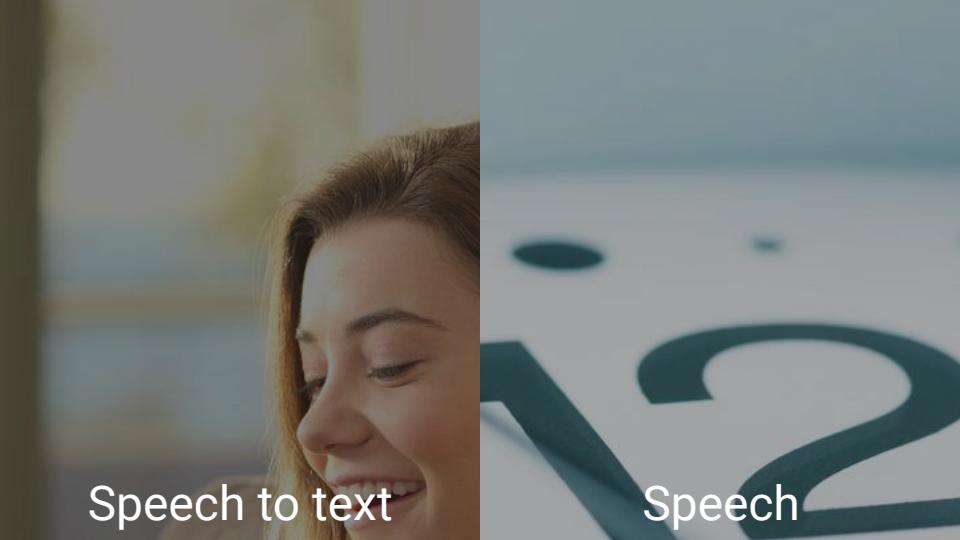
Speech to text transcription in  
100+ languages



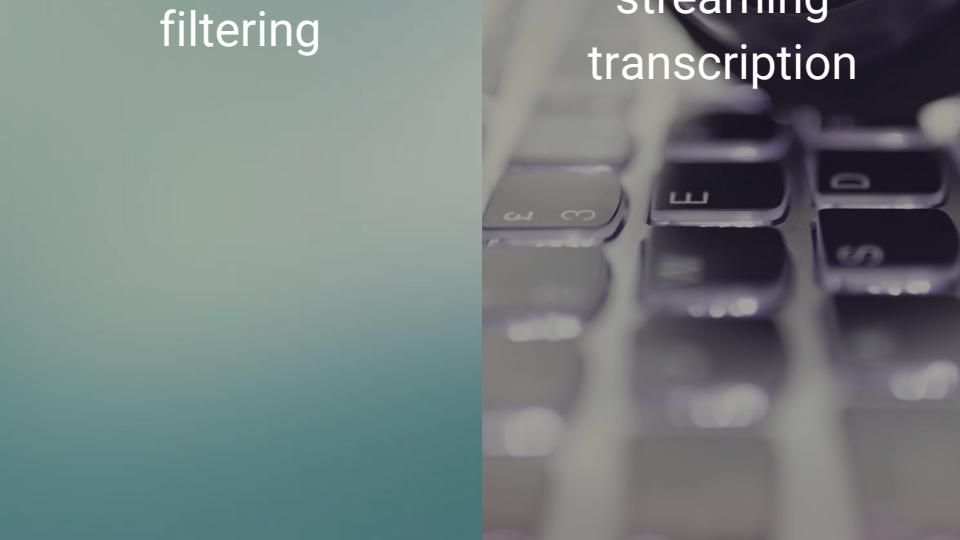
Speech to text  
transcription



Speech  
timestamps



Profanity  
filtering



Batch &  
streaming  
transcription

# Demo: speech timestamps

1

Extract audio from  
a video

2

Send audio to Cloud  
Speech for transcription  
& timestamps

3

Visualize & search  
videos in a UI

# Cut Video

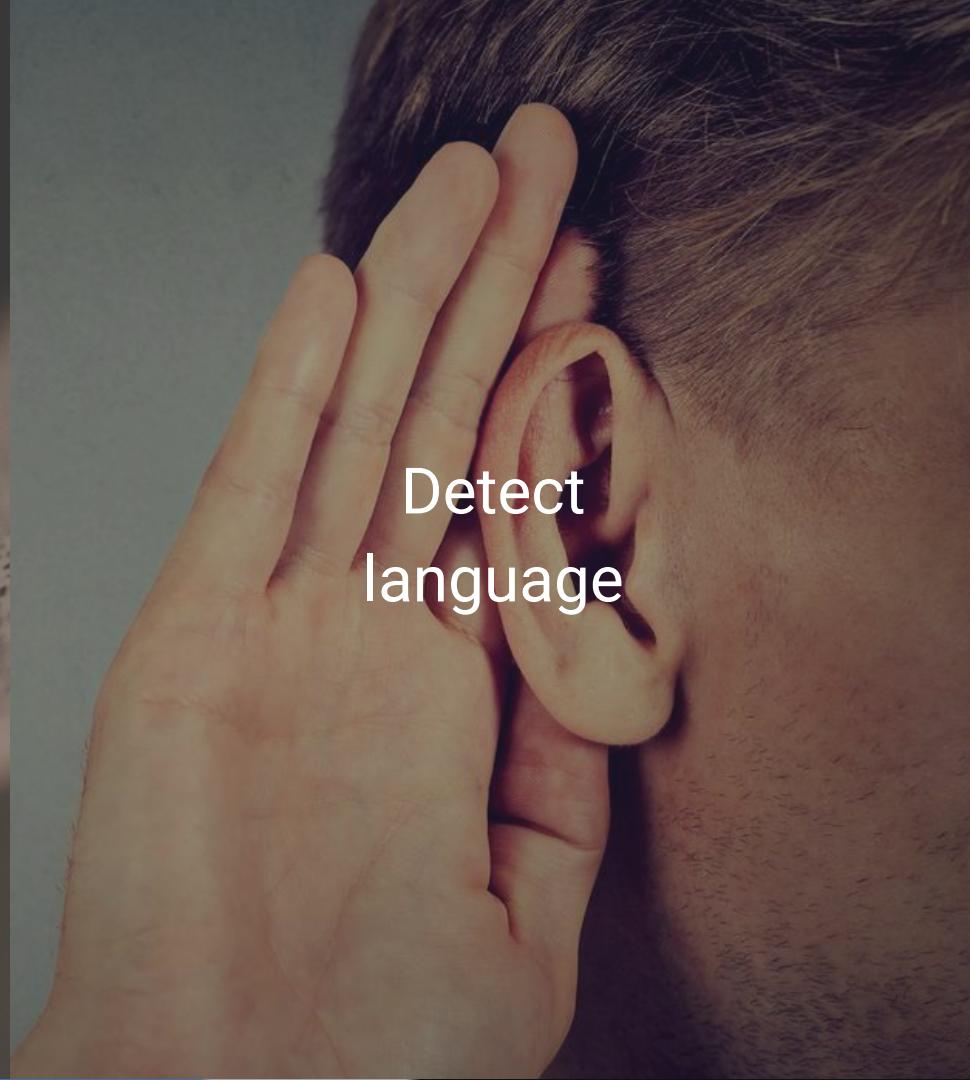
# Translation + NL

# Cloud Translation

Translate text into 100+ languages



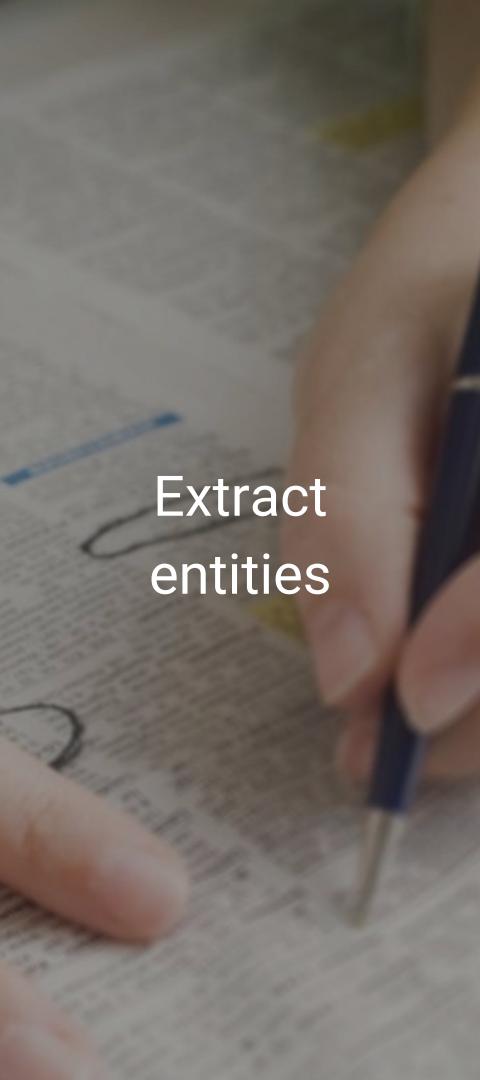
Translate  
text



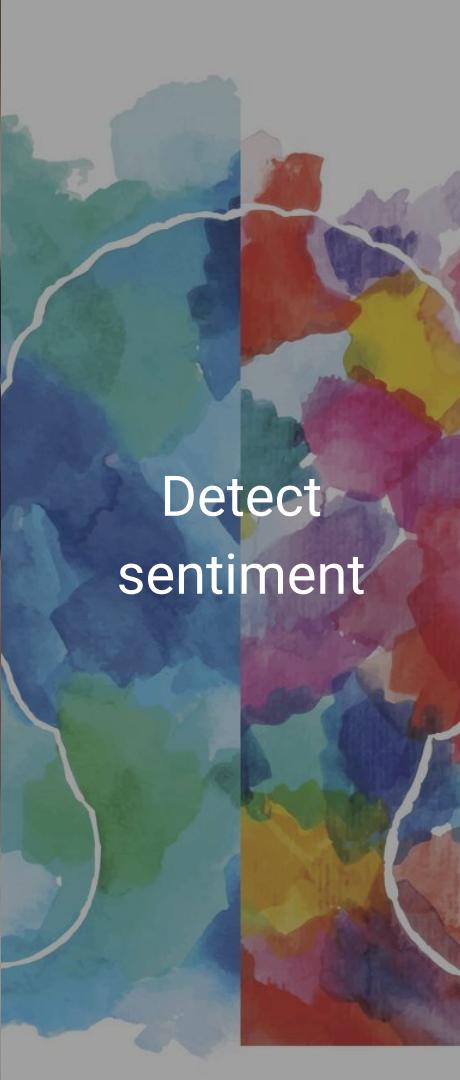
Detect  
language

# Cloud Natural Language

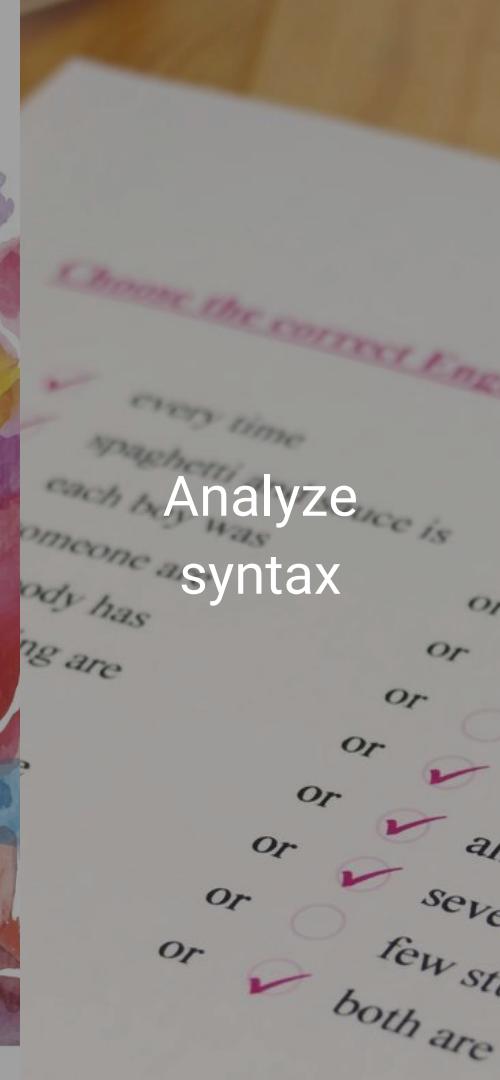
Understand text with a simple  
REST API request



Extract  
entities



Detect  
sentiment

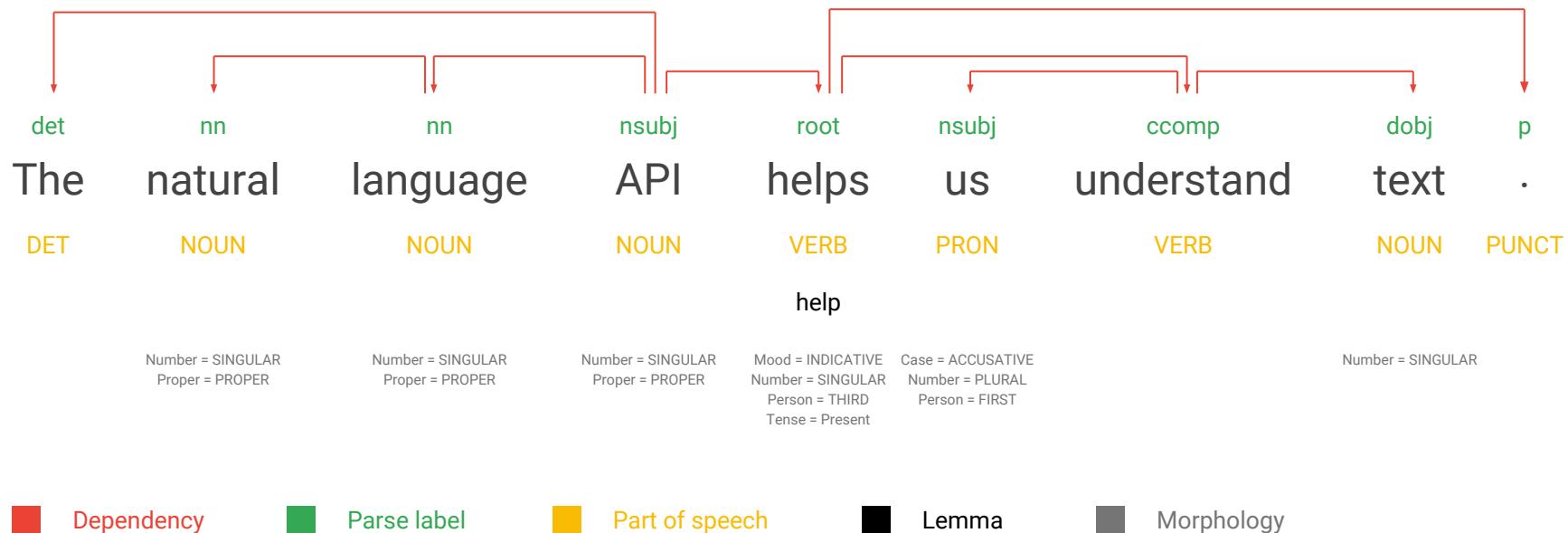


Analyze  
syntax



Classify  
content

# Analyze syntax



# Classify content using Cloud Natural Language

SPORTS | METS 2, REDS 0

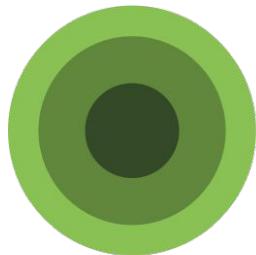
## ***Rafael Montero Shines in Mets' Victory Over the Reds***

By THE ASSOCIATED PRESS AUG. 30, 2017

Rafael Montero Shines in Mets' Victory Over the Reds. Montero, who was demoted at midseason, took a one-hitter into the ninth inning as the Mets continued to dominate Cincinnati with a win at Great American Ball Park.

```
{ categories:  
[  
 {  
 name: '/Sports/Team  
Sports/Baseball',  
 confidence: 0.99  
 }  
 ]  
}
```

# Wootric: Cloud Natural Language in production

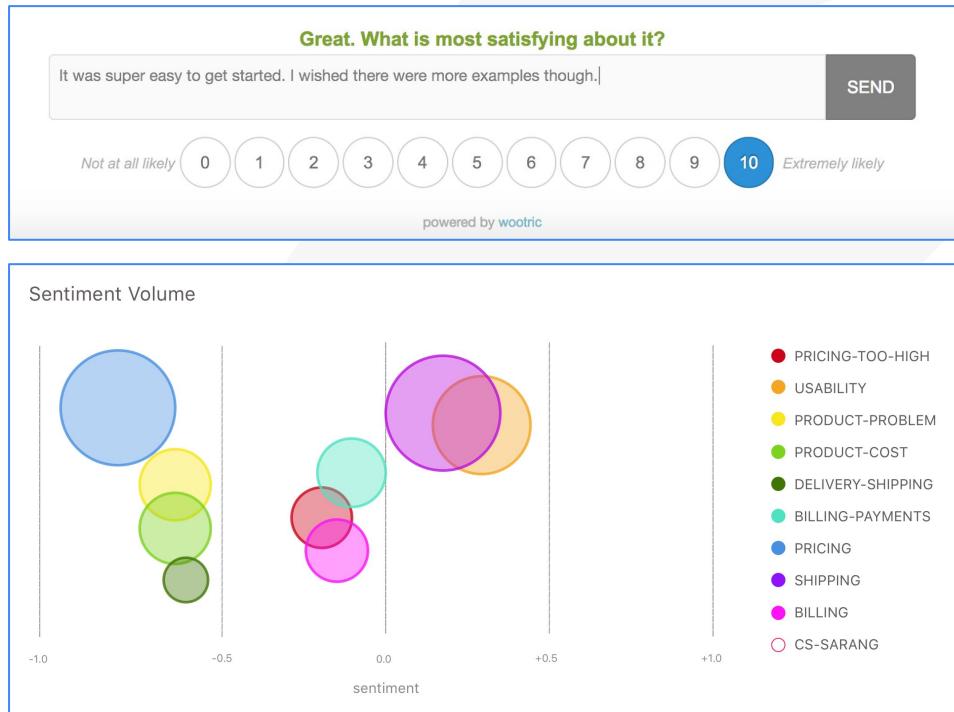


# WOOTRIC

WIN CUSTOMERS FOR LIFE

# Wootric: analyzing and routing feedback

- Make sense of millions of qualitative customer feedback each week using **entity** and **sentiment analysis**
- Route and respond to feedback in near realtime, compared to manually classifying each response



# Try Cloud NL in the browser

The Natural Language API helps us understand text.

ANALYZE

See supported languages

Entities Sentiment Syntax Categories

Dependency  Parse Label  Part of Speech  Lemma  Morphology

```
graph TD; The[The] -- det --> Natural[Natural]; Natural -- nn --> Language[Language]; Language -- nn --> API[API]; API -- nsubj --> helps[helps]; helps -- root --> us[us]; us -- nsubj --> understand[understand]; understand -- ccomp --> text[text]; text -- dobj --> p[p];
```

POS Features

DET	NOUN number=SINGULAR proper=PROPER	NOUN number=SINGULAR proper=PROPER	NOUN number=SINGULAR proper=PROPER	VERB mood=INDICATIVE number=SINGULAR person=THIRD tense=PRESENT	PRON case=ACCUSATIVE number=PLURAL person=FIRST	VERB	NOUN number=SINGULAR	PUNCT
-----	--	--	--	---	--	------	-------------------------	-------

[cloud.google.com/natural-language](https://cloud.google.com/natural-language)

# Cut Video

# Lab on ML APIs

# Lab: Invoking Machine Learning APIs

# Lab: Invoke Machine Learning APIs

In this lab, you invoke ready-to-use Machine Learning APIs in your Datalab environment

1. Clone the code repository within your Datalab
2. Save the code repo in your Cloud project's source repositories
3. Enable ML APIs in your Cloud project
4. Invoke ML APIs from Datalab

# 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:

- easy to use and
- stand-alone.

# cloud.google.com

