Google Cloud

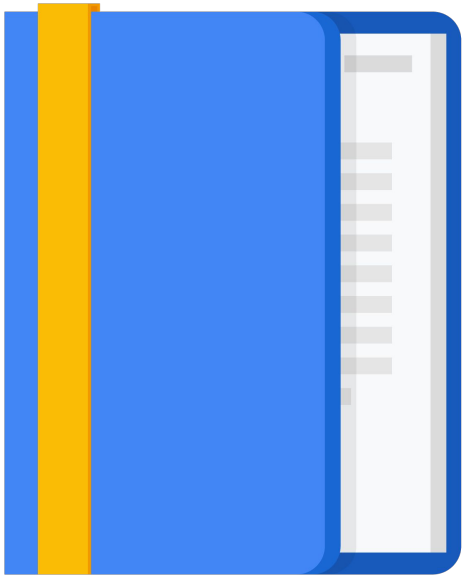
Analytics and AI: Introduction

Agenda

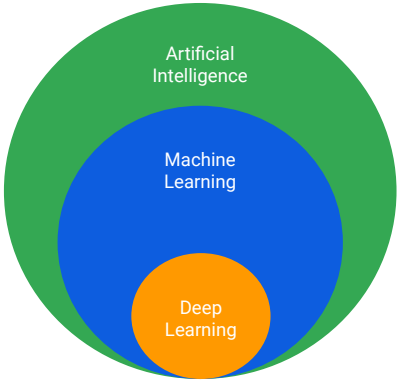
What is AI?

From Ad-hoc Data Analysis to Data Driven Decisions

Options for ML models on GCP



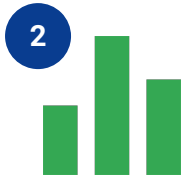
Machine Learning is a type of AI, and deep learning is a type of machine learning



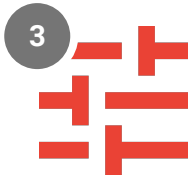
Class of problems we can solve when **computers think/act like humans**

ML is a way to use standard **algorithms** to derive **predictive insights** from **data** and make **repeated decisions**

Algorithm



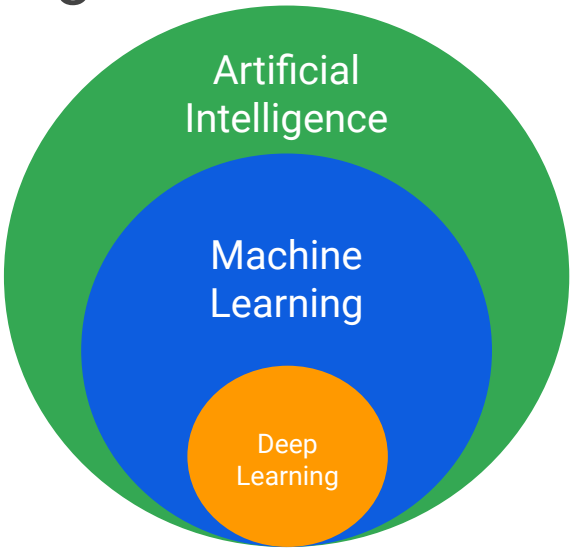
Data



Predictive insight

Decision

Why are Machine Learning and Deep Learning so exciting?



Class of problems we can solve when **computers think/act like humans**

Scalably solve those problems using **data** examples (**not custom code**)

Even when that data consists of **unstructured data** like images, speech, video, natural language text, etc.

Keller Williams uses AutoML Vision to automatically recognize common elements of house furnishings and architecture

Cloud AutoML Vision

"Modern" style

Granite countertops

Kewpie uses ML to sort out the bad potatoes in baby food

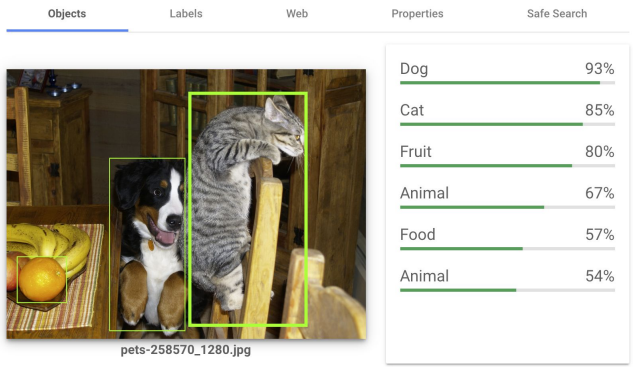
Original process required humans to identify low-quality ingredients, which was expensive and stressful.

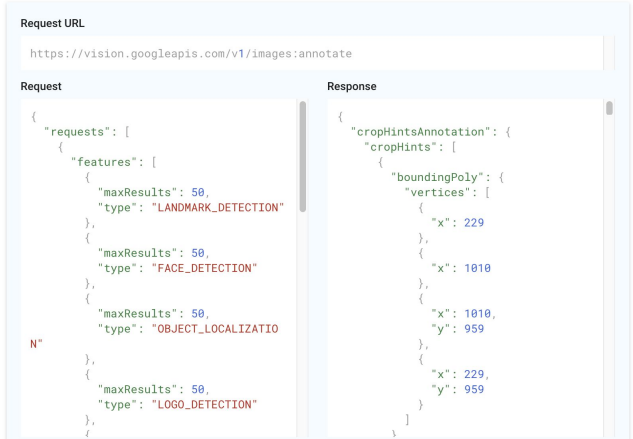
Machine learning was used to replicate the quality control process.

kewpie

Google Cloud

Play around with the power of AI yourself...



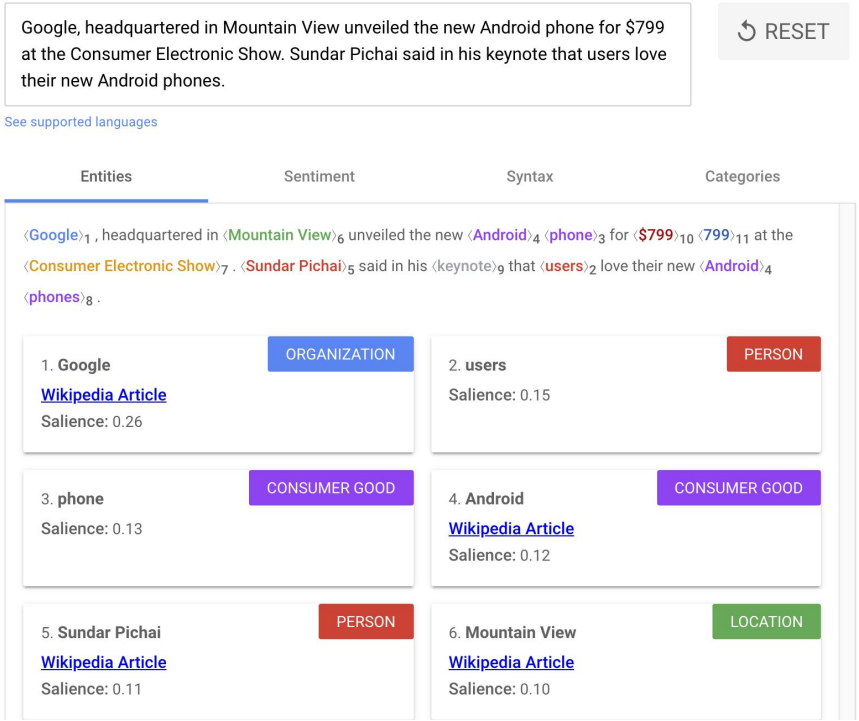


cloud.google.com/vision/

- Object detection
- Labeling and confidence
- Web lookup
- Pre-trained (call the API)

Goo

Try Google's natural language API



cloud.google.com/natural-language/

- Entity extraction
- Sentiment analysis
- Sentence structure
- Pre-trained (call the API)

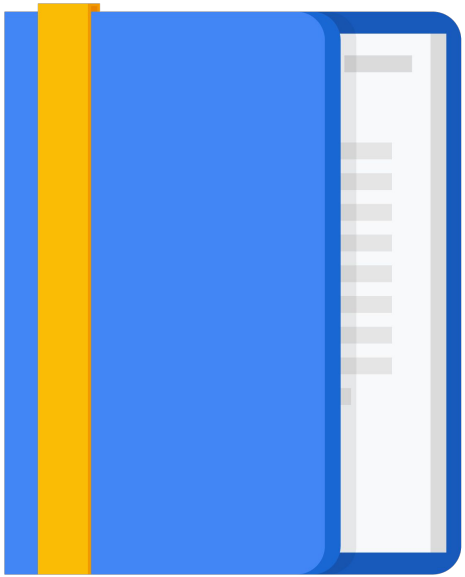
Google Cloud

Agenda

AI

From Ad-hoc Data Analysis to Data Driven Decisions

Options for ML models on GCP



Imagine you're the owner of a bicycle rental business (in London). How do you stock enough bicycles?

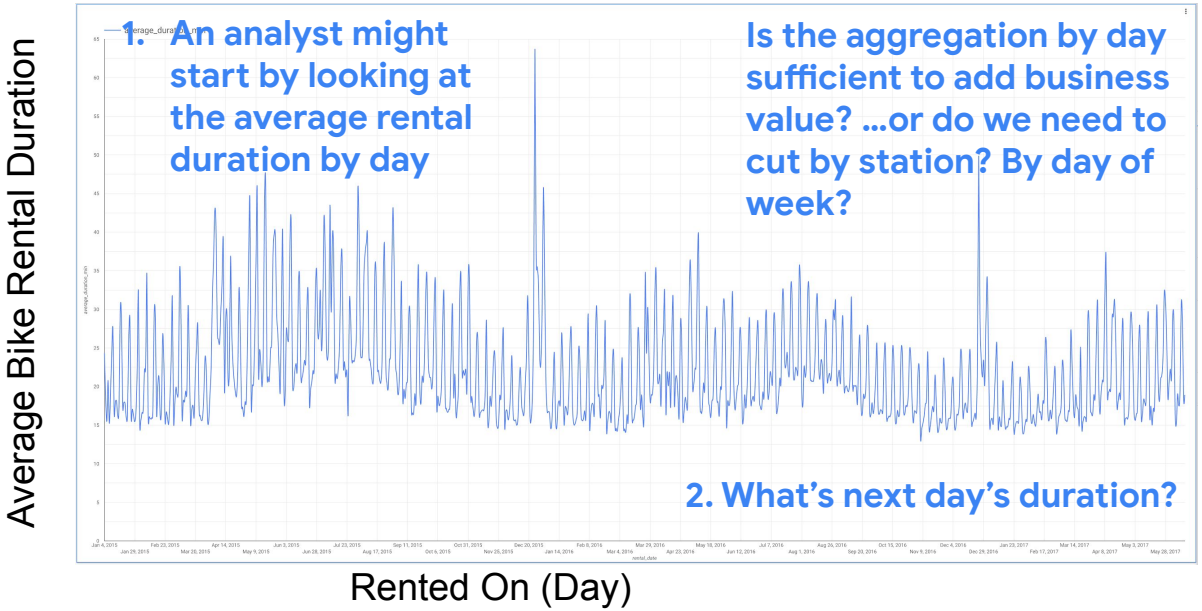
Commuter Bikes

If rental is likely to be for a **short duration**, we need to have commuter bikes in stock

Road Bikes

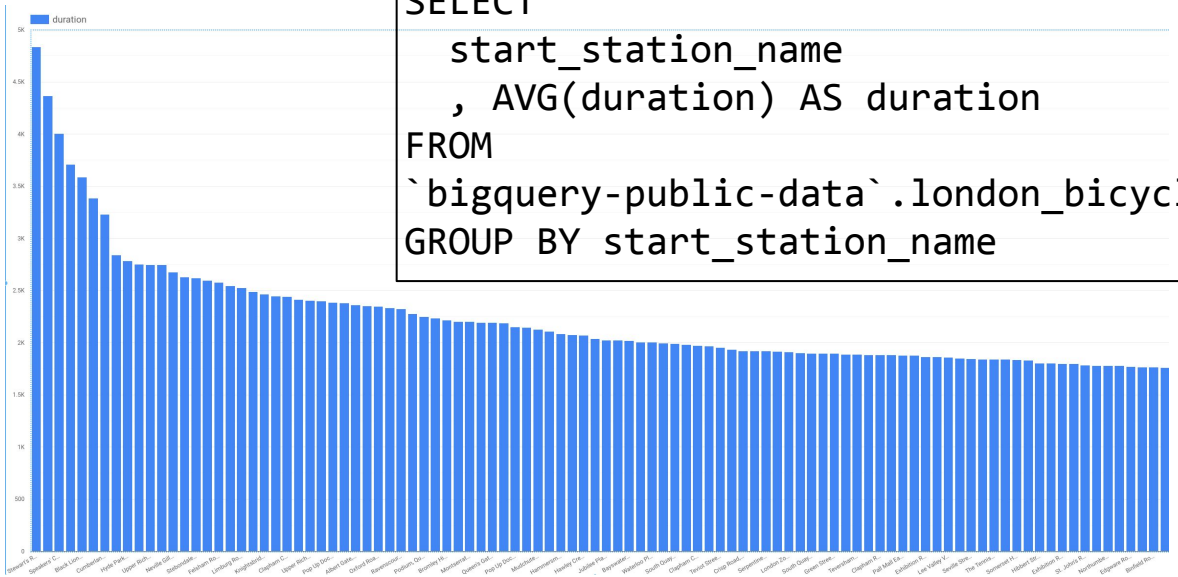
If rental is likely to be for a **long duration**, we need to have road bikes in stock

You hire a data analyst to help get you insights on how to keep the right bicycles in stock



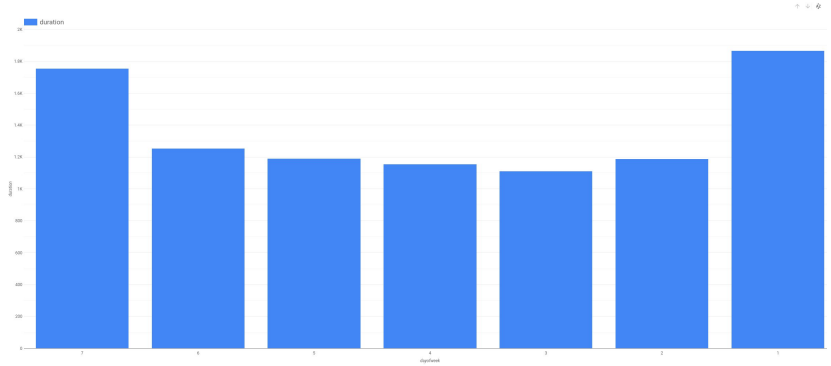
Google Cloud

Does the duration of a rental vary by station?

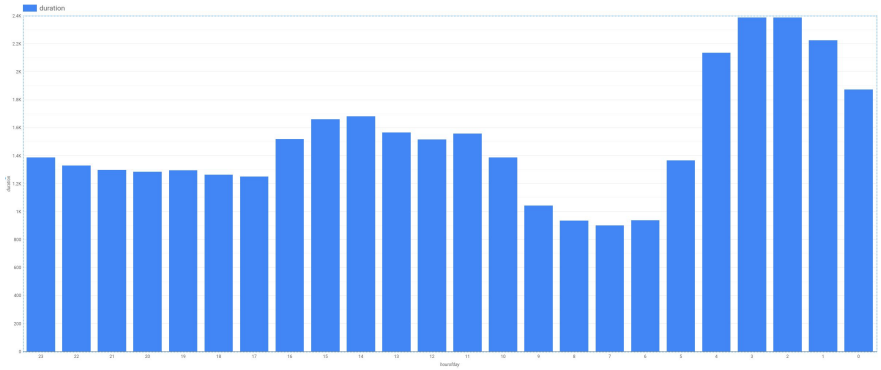



```
SELECT
  start_station_name
  , AVG(duration) AS duration
FROM
`bigquery-public-data`.london_bicycles.cycle_hire
GROUP BY start_station_name
```

How about the day of the week? Hour of day?




```
SELECT
  EXTRACT(dayofweek
  FROM
    start_date) AS dayofweek,
  AVG(duration) AS duration
FROM
  `bigquery-public-data`.london_bicycles.cycle_hire
GROUP BY
  dayofweek
```




```
SELECT
  EXTRACT(hour
  FROM
    start_date) AS hourofday,
  AVG(duration) AS duration
FROM
  `bigquery-public-data`.london_bicycles.cycle_hire
GROUP BY
  hourofday
```

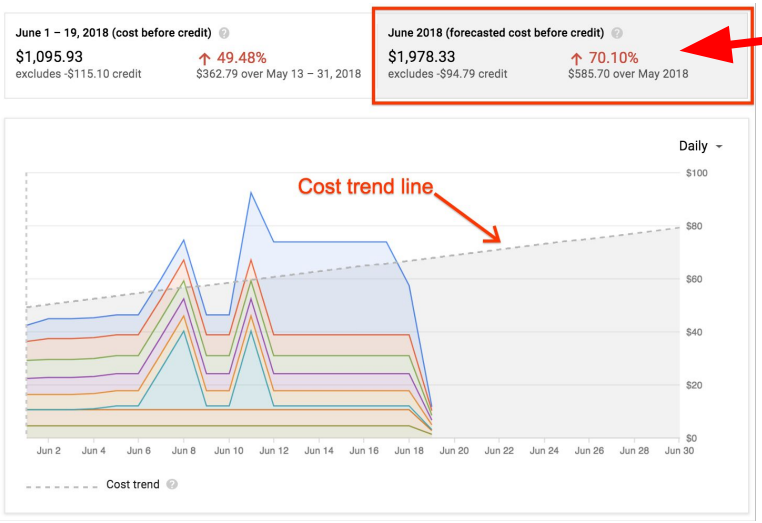
Google Cloud

This ad-hoc analysis is great but...

- A lot of manual, repetitive work involved for the data analyst
- Any decisions made will be based on hunches on how all these factors interact
- Wouldn’t it be better if we could automate this analysis?

What we need is a ML model to be able to make predictions

- Goal: Augment our dashboards with predicted values
  e.g. prediction for the duration of a rental



As an example, Google augments GCP cost dashboards (descriptive) with forecasted (predictive) usage costs

Use the ML model to anticipate what type of bike/how many to stock at your locations

- The ML model takes some of the drudgery out of ad-hoc analysis to help you make truer data-driven decisions
- Can build a ML model in BigQuery or AI Platform or AutoML

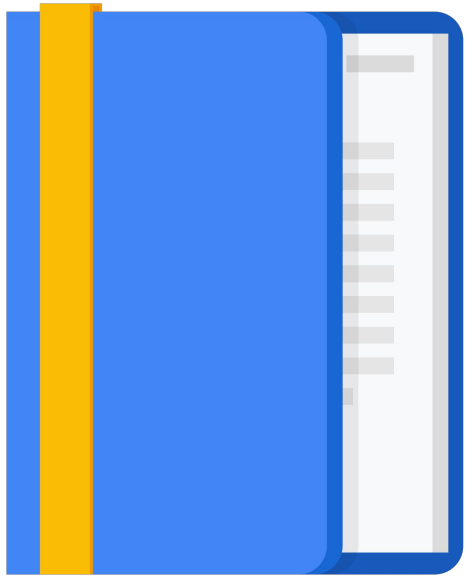
```
CREATE OR REPLACE MODEL
  bike_model.model_bucketized TRANSFORM(* EXCEPT(start_date),
  IF
    (EXTRACT(dayofweek
      FROM
        start_date) BETWEEN 2 AND 6,
      'weekday',
      'weekend') AS dayofweek,
    ML.BUCKETIZE(EXTRACT(HOUR
      FROM
        start_date),
      [5, 10, 17]) AS hourofday )
OPTIONS
  (input_label_cols=['duration'],
    model_type='linear_reg') AS
SELECT
  duration,
  start_station_name,
  start_date
FROM
  `bigquery-public-data`.london_bicycles.cycle_hire
```

Agenda

AI

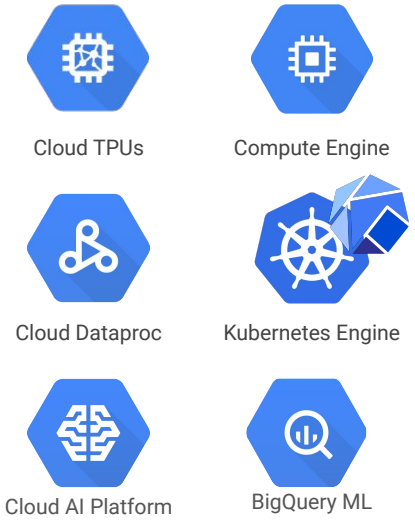
From Ad-hoc Data Analysis to Data Driven Decisions

Options for ML models on GCP

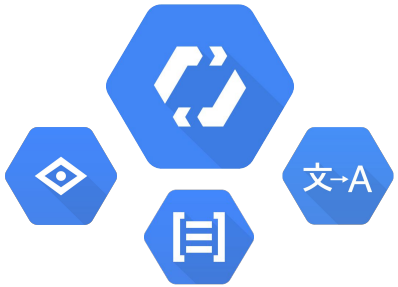


Leverage pretrained models or build your own

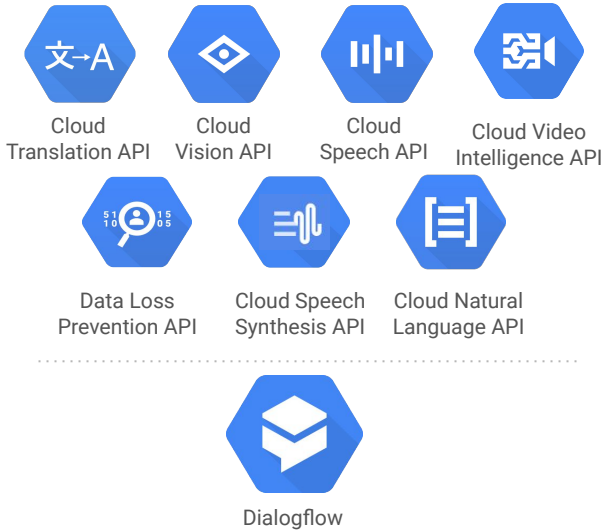
Build a Custom Model



Build Custom Model (codeless)



Call a Pretrained Model



Module Summary

- AI's impact on industry is huge
- Predictive modeling takes data-driven decision making to a new level
- The typical data science workflow