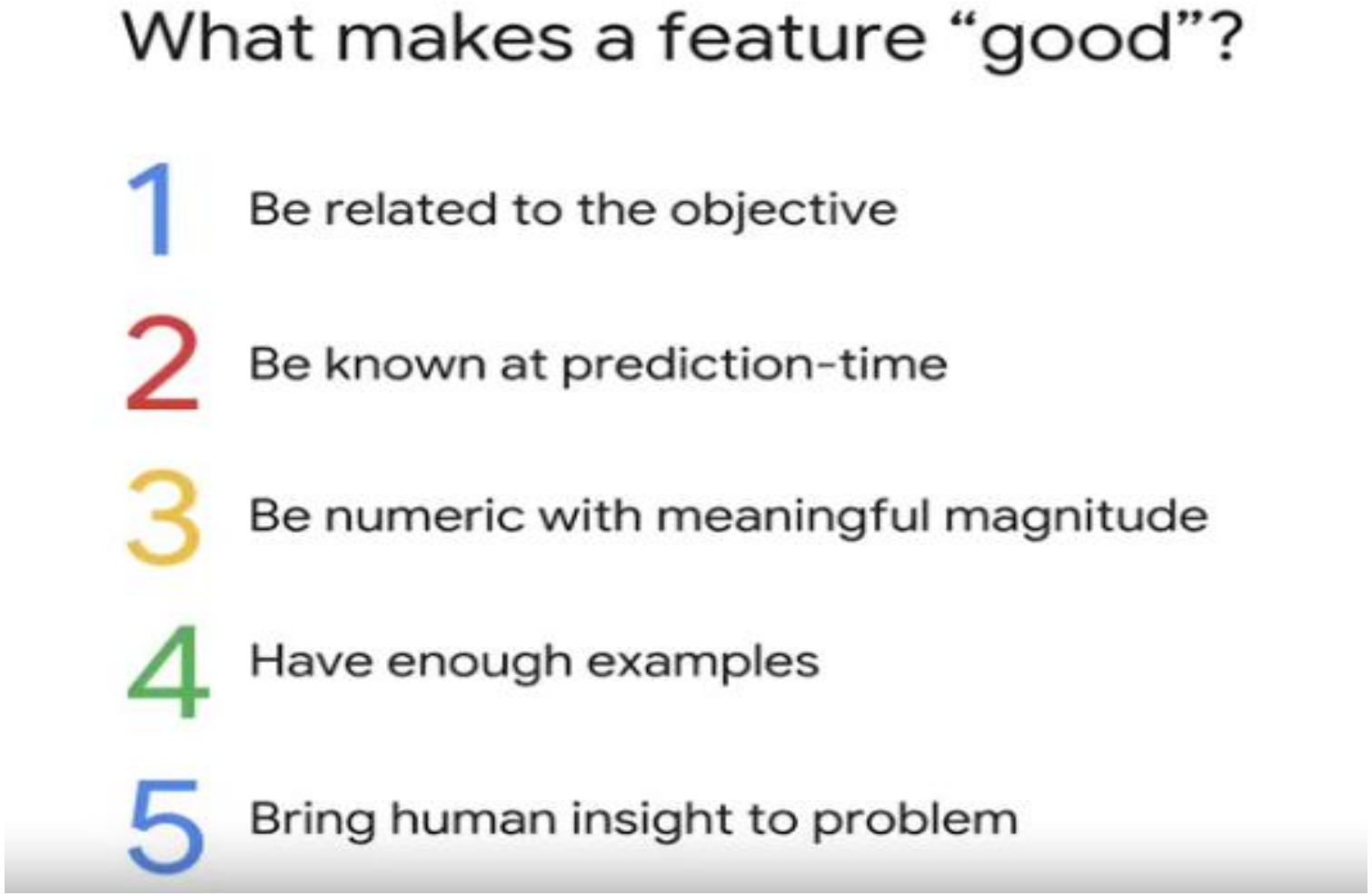
Criar Dataset e Construir Modelo

Criar Dataset



Criar Dataset

Will we know all these things at prediction time?

With ultrasound

Sex: Male/Female
Plurality: 1, 2, 3, 4, or 5

Without ultrasound

Sex: Unknown
Plurality: Single/Multiple

Criar Dataset

Split a dataset into training/validation using hashing and modulo operators

```
#standardSQL
SELECT
  date,
  airline,
  departure_airport,
  departure_schedule,
  arrival_airport,
  arrival_delay
FROM
 `bigquery-samples.airline_ontime_data.flights`

WHERE
  MOD(ABS(FARM_FINGERPRINT(date)),10) < 8
```

Note: Even though we select date, our model wouldn't actually use it during training.

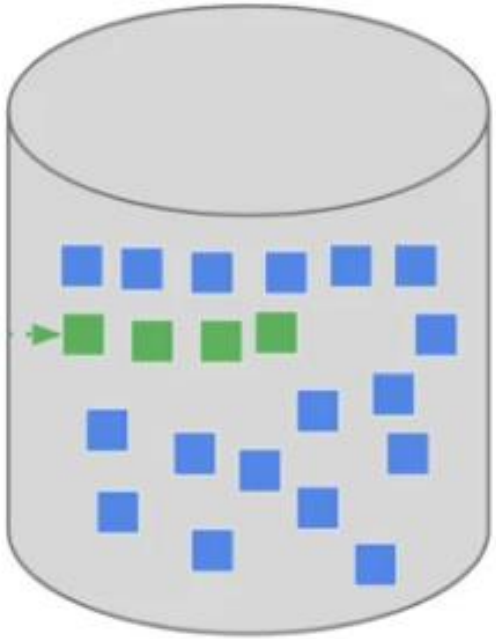
Hash value on the Date will always return the same value.

Then we can use a modulo operator to only pull 80% of that data based on the last few hash digits.

Criar Dataset

Developing the ML model software on the entire dataset can be expensive; you want to develop on a smaller sample

Develop your TensorFlow code on a small subset of data, then scale it out to the cloud



Full Dataset

Criar Dataset

Sampling the split so that we have a small dataset to develop our code on

```
#standardSQL
SELECT
  date,
  airline,
  departure_airport,
  departure_schedule,
  arrival_airport,
  arrival_delay
FROM
 `bigquery-samples.airline_ontime_data.flights`

WHERE
  MOD(ABS(FARM_FINGERPRINT(date)),10) < 8  AND   RAND() < 0.01
```

Criar Dataset

Lab ou vídeo “04-Criar dataset”

Construir modelo

Vídeo “05-Construir modelo”

Glossário
DNN: deep neural network

Construir modelo

Vídeo “06-Construir modelo - lab”