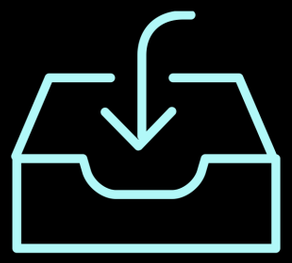

Download training and test data

Set-up data for pre-processing

STOP and think:
How should we use the 28 features?

Looked at correlations between variables and target

Gathered domain knowledge / other approaches on similar tasks

Insights!
Intuition!

Feature Engineering

Ensemble Approach using many models

YNC_Hacks

YaleNUSCollege

Important Insight 1: INTENTION

Not all customers who ask for a quote actually intend to convert.

- Out of 2.9 million quotes in the training set, there were only 206k unique customer keys.
- We realized that there was a median of 11 quotes per customer key, out of which there could be quotes for different car makes. There are 2 cases:
  (1) the customer has multiple cars; or
  (2) the customer is just “window-shopping”.
- We created 6 discrepancy flags to identify customers of the latter group: these customers are very likely to not convert. These identify differences in the following 4 types of information provided by the customer:
  (1) no. of unique car annual mileages;
  (2) no. of unique cars (car make ID);
  (3) no. of unique car ages; and
  (4) no. of unique car insured values.

Important Insight 2: NUMBER OF POLICIES

The greater the number of different policies offered, the more unlikely the conversion.

- We also realized that there was a threshold number of different policies offered beyond which customers were highly unlikely to convert. Thus, we included a feature on number of unique policies offered to a customer.

ENSEMBLE APPROACH

- We combined many classifiers in an ensemble approach, including:
  (1) XGBoost;
  (2) Gaussian Naive Bayes;
  (3) Logistic Regression; and
  (4) RandomForest.