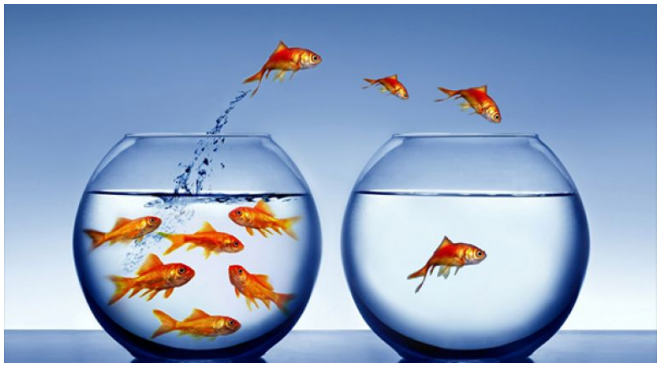
Churn Prediction

Case Study Galvanize

Overview

- A ride-sharing company in San-Francisco is interested in predicting their rider retention and make data based business decision

- This case study helps the company to achieve the following goals:
  - Analyze their user data
  - Build a churn prediction model
  - Analyze the business impact
  - Make recommendations

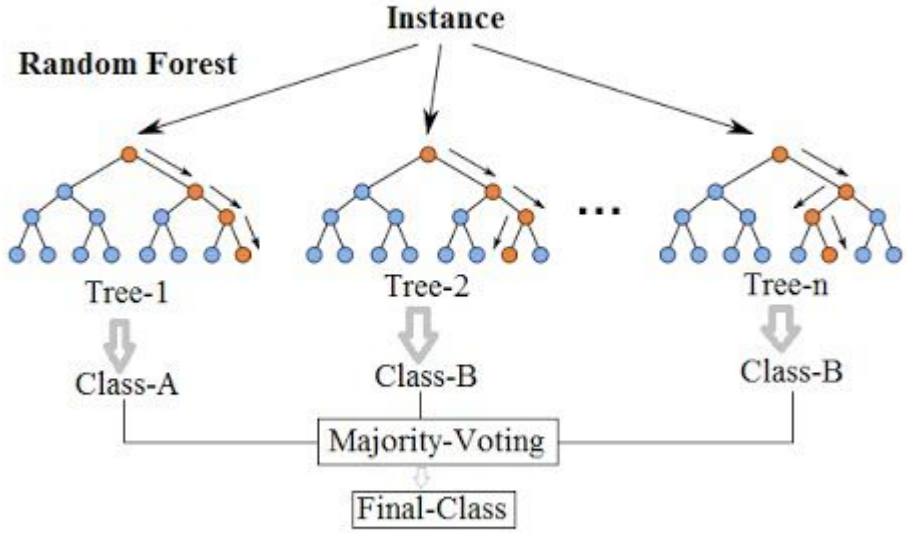


Feature Engineering

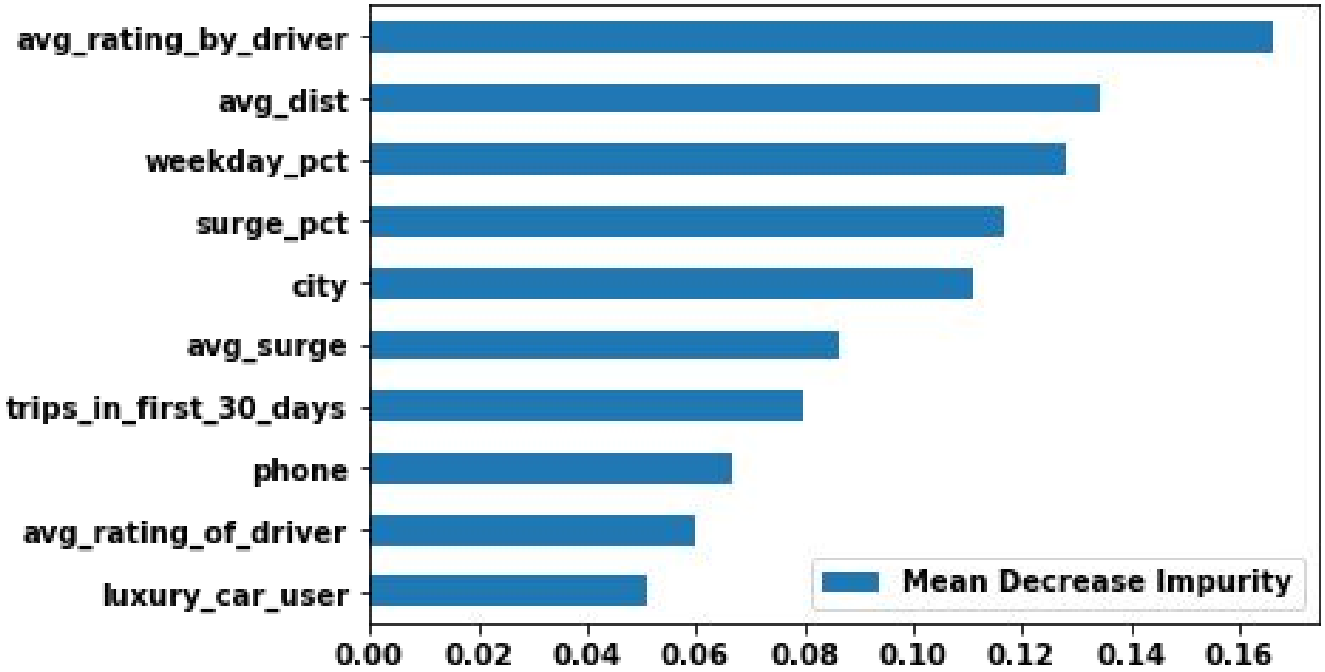
- There are 40,000 user activity items, and each item has 14 features. Based on the exploratory data analysis, we did the following feature engineering.
  - Create a “Churn” Feature
  - Conversion the signup_date to datetime object, create a churn rate features
  - Dealing with missing values.
  - Convert categorical column into numeric labels
  - Fill in the nan values for numeric column with the median value of each column
- Before model building, we split the raw data sets into train/test by 80:20. The test data was only used to evaluate the model in the last step.

Model Selection: Random Forest

- Random Forest has several advantages for this project
  - Flexibility to handle both numeric and categorical features
  - Easy interpretation and fast training
  - Lower risk of overfitting and less variance
  - Feature importance output
- Cross-validation and grid search was used to optimize the model parameters

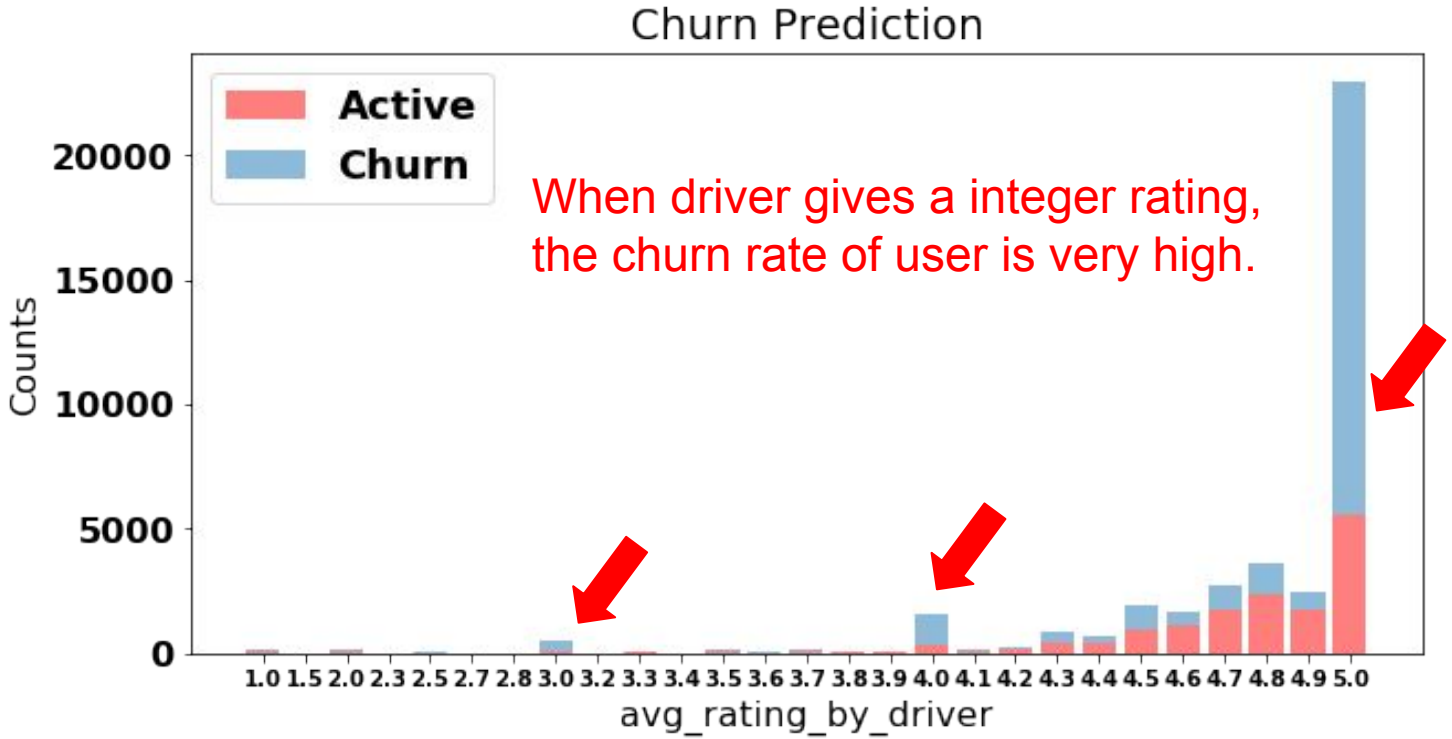


Feature Importance Analysis



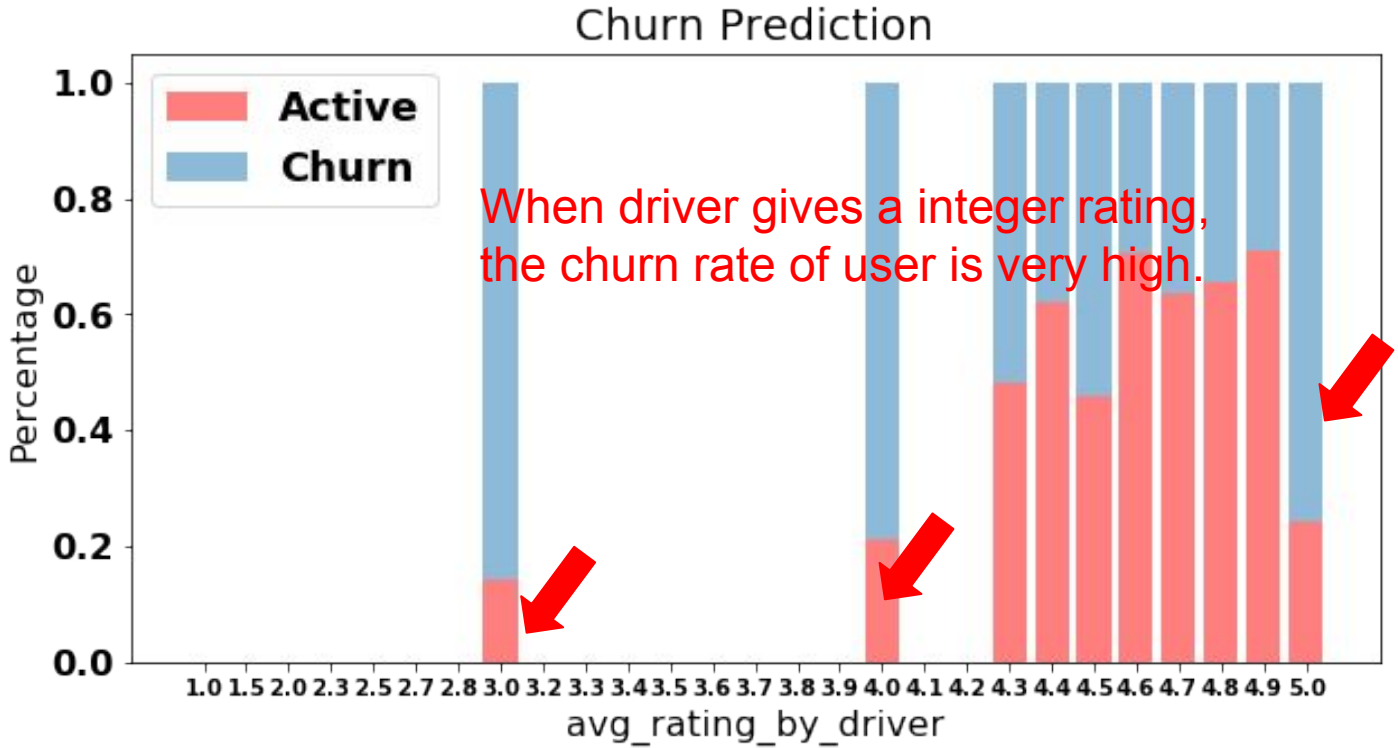
Feature Importance Analysis

avg_rating_by_driver

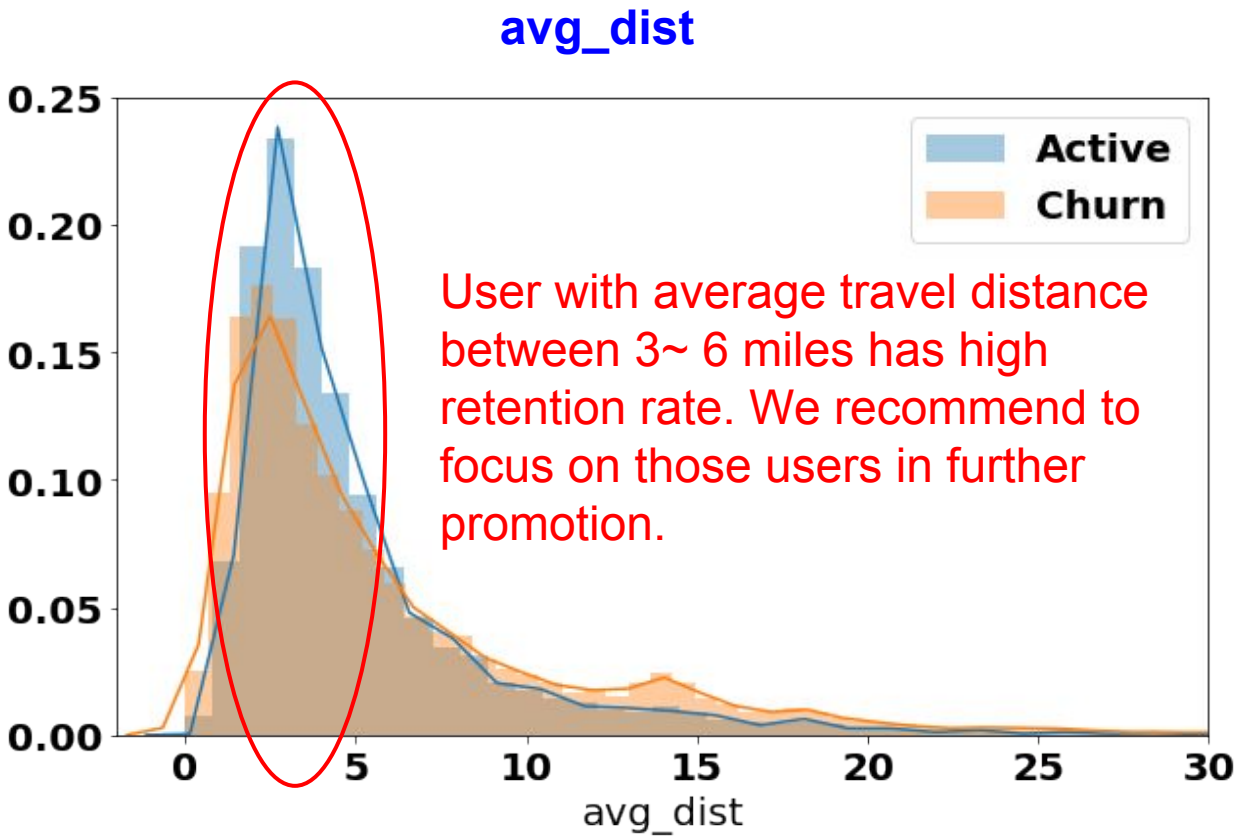


Feature Importance Analysis

avg_rating_by_driver



Feature Importance Analysis



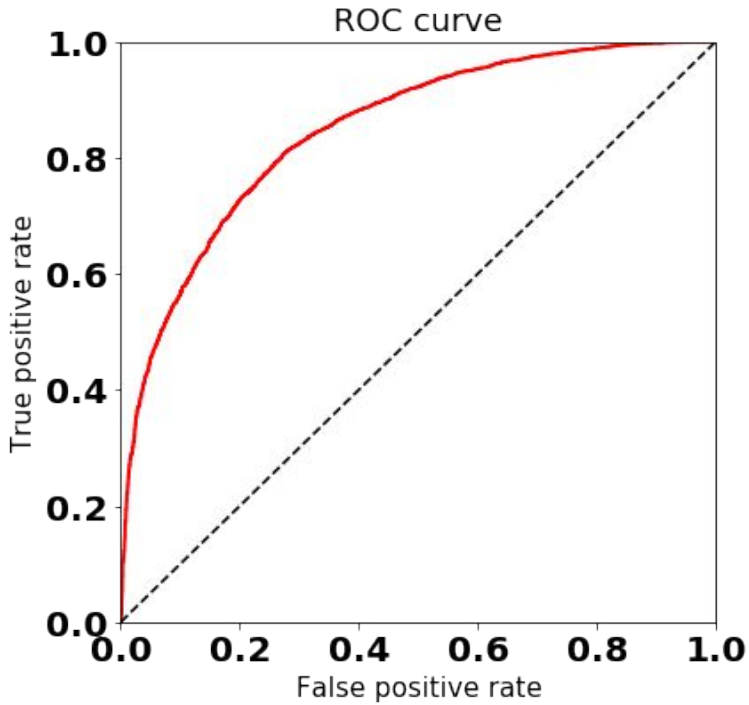
Model Prediction Performance

Accuracy Score: 0.779 Precision: 0.85

Recall : 0.80

Confusion Matrix on Test Data

	Condition Positive	Condition Negative
Predicted Positive	5318	910
Predicted Negative	1322	2450



Business Impact

Cost/Benefit Matrix

	Condition Positive	Condition Negative
Predicted Positive	-20 $	0
Predicted Negative	-100$	0

Confusion Matrix on Test Data

	Condition Positive	Condition Negative
Predicted Positive	5318	910
Predicted Negative	1322	2450

- Business Information: Assume the company has 10,000 customer. If one customer churns, the company lose $100 in CLTV revenue. When we predict a customer churn, we recommend to send out a 10$ coupon, and assume 90% of customer stays. The cost/benefit matrix was shown as above.
- Based on the prediction performance of our model, the data-driven action based on our random forest model is going to save the company **394,000$ CLTV revenue.**

Recommendation

- Based on feature analysis, we recommend to promote the driver engagement. The more detail the driver rate the rider, the lower the churn rate of the user.
- The customer service should be more focused on users who has average distance between 3 ~ 6 miles
- When our model predict a user has a high chance of churning, send out a promotion coupon to keep them.