baywheels

Bike Share Subscriber or Casual Rider?

Matt Ranalletta

Two Types of Users

Casual Rider

$2 for first 30 minutes
$3 for additional 15 min.
Extra fee for e-bikes

01

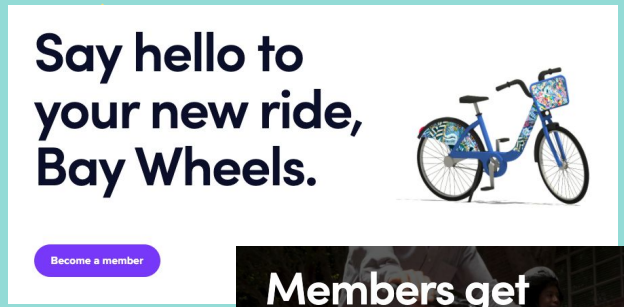
Subscriber

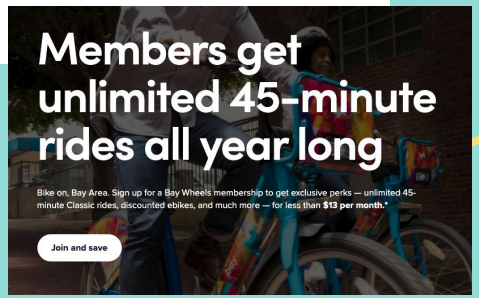
Unlimited 45-minute rides
$2 for additional 15 min.
Discounted e-bikes
Can link transit card

02

Main Goal

Identify casual riders and turn them into subscribers.





Classification Process

Research & EDA

01

SQL & Pandas

Preprocessing & regularization

02

One Hot Encoding, oversampling, scaling, training and validating

Optimizing & evaluating

03

Feature engineering, parameter tuning, testing on holdout data

Selecting Model

04

Feature importance, metrics

Classification Algorithms

KNN

Logistic Regression

Decision Tree

Random Forest

Naive Bayes

XGBoost

Classification Algorithms

KNN

Logistic Regression

Decision Tree

Random Forest

Naive Bayes

XGBoost

Classification Algorithms

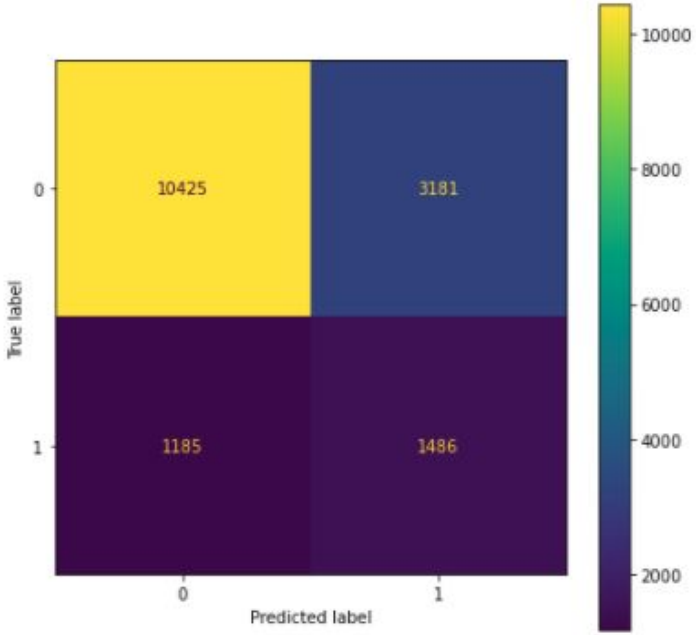
KNN	Decision Tree	Naive Bayes
Logistic Regression	Random Forest	XGBoost

Classification Metrics

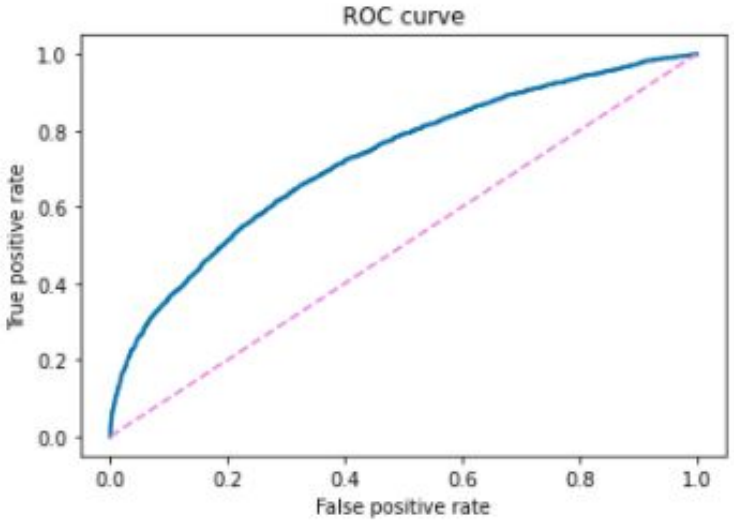
	Precision	Recall	F1-score	Accuracy
Casual	0.90	0.77	0.83	0.73
Subscriber	0.32	0.56	0.41	

Metrics

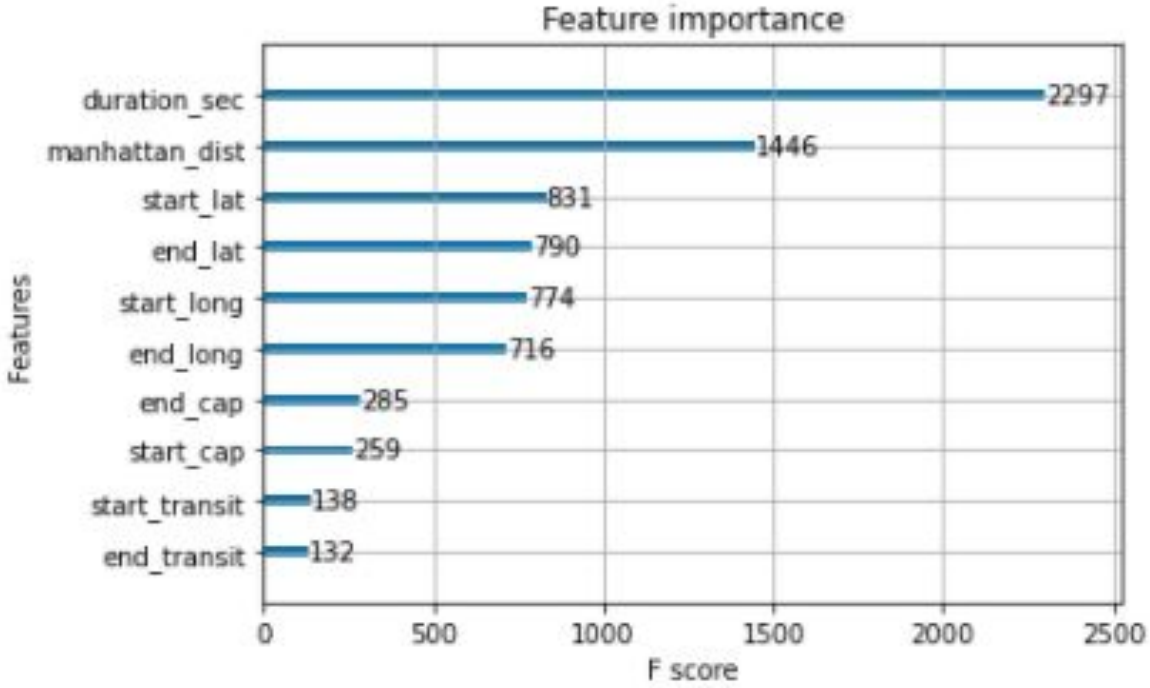
Predicting on Test Set



ROC AUC score = 0.7244331059925608



Metrics


Feature importance
duration_sec
manhattan_dist
start_lat
end_lat
start_long
end_long
end_cap
start_cap
start_transit
end_transit
2297
1446
831
790
774
716
285
259
138
132
Features
0
500
1000
1500
2000
2500
F score

Streamlit App Demo

Improvements

Time Data

Consistent start and stop data with hours and minutes

Web App

More sophisticated to less technical users

Extra Time

Could try out more modeling techniques and additional parameters

Thank you!