LA-Metro Bike Sharing

Analysis and Modelling

Project 3

For this, Project 3, I have performed analysis and Machine Learning modelling on the LA-Metro Bike Sharing Network
This dataset was obtained from Kaggle.com

The Network

762 Bikes

66 Stations

3 Membership Types

9 months

>130,000 Rides!

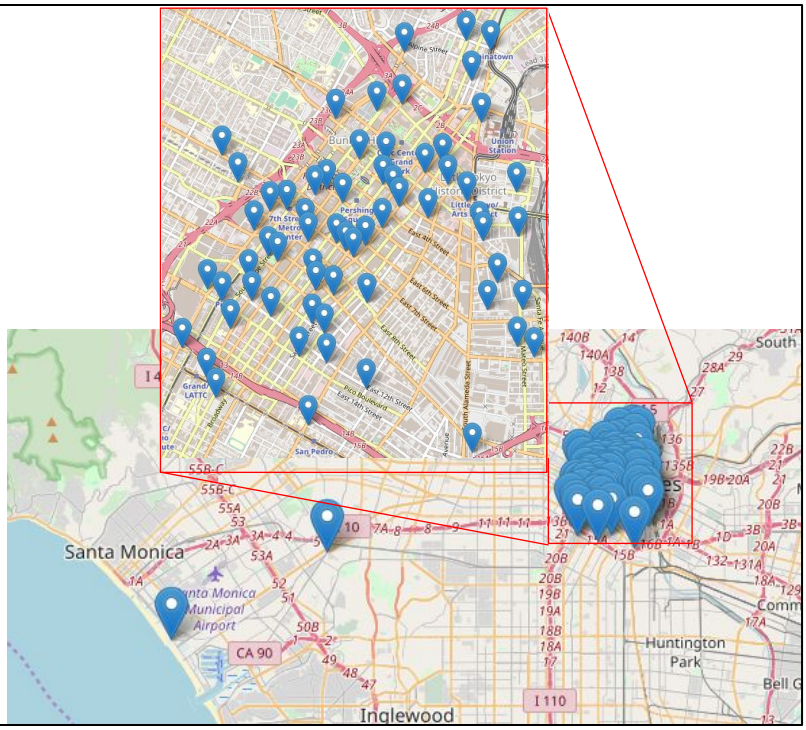
The network consists of 762 bikes which can be ridden between 66 stations by riders with 3 different membership levels. The Dataset is a complete log of rides taken over a 9 month period from from July 2016 to March 2017 encompassing over 130,000 rides

The Network

Downtown LA

Venice Beach

Culver City



The network of stations is clustered tightly in Downtown LA but there are also 2 “outpost: locations at Culver City and at Venice Beach

Machine Learning

- Understanding Customer Behaviour

Who is doing **What?**

This machine learning exercise is ultimately about modelling customer behaviour. Our riders are our customers and the better insights we have into their usage of the network ultimately the better the network can be

Modelling ride Duration from starting conditions

- Drop end location and time and trip type features
- Grid Search 2 Regression Models
- Decision Tree - R-squared: 0.266
- Random Forest- R-squared: 0.30

A rider picks up a bike from a station, How long will they have that bike for?
Ride duration was modelled using 2 Regressor methods. With the best result, a tuned Random Forest Regressor, able to account for 30% of variance from observed mean.
On our way to understanding rider behaviour

Modelling Membership Type

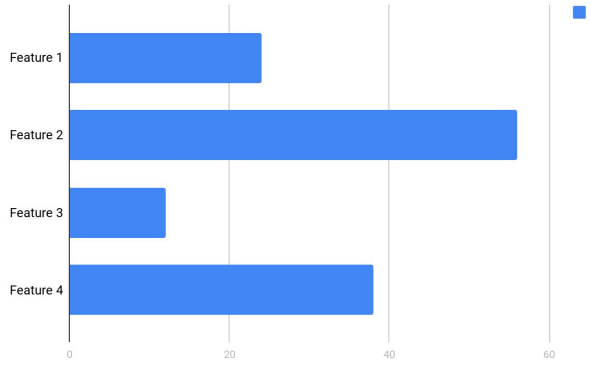
Grid Search 3 ensemble Classifiers

- Random Forest: 72% Accuracy
- ADABoost: 73% Accuracy
- XGBoost: 74% Accuracy

Who are our customers? How distinct are the riding patterns of different membership types?
3 ensemble classifier were built, tuned XGBoost model had highest overall accuracy at 74% of predictions.
Understanding the different needs of different membership types

Future Work

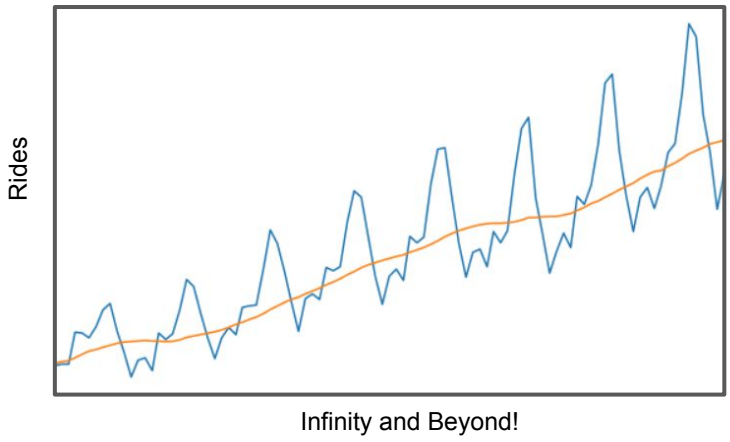
Extract most influential features from Models



This is our ultimate goal. By extracting the most influential features from our models of ride duration and membership classification we can understand our riders needs and guide the growth of the Network to encourage usage!

Future Work

Yearly Ridership Patterns



This dataset spans almost 9 months. But even right now someone out there in LA is probably riding a bicycle.
If we can expand scope of time series to we can incorporate year on year growth trends

Future Work

Weather data correlation

Weather is very likely to influence ridership and future models could correlate to weather data

Thank You!