Bike Sharing

AMNA JAMAL

Objective

Objective of this exercise was to accurately predict the bikes at each bike station so as to adjust the bike supply across the stations.

Data Overview

Three data sets were made available

- Station data

Id	Name	Lat	Long	Dock Count	City

- Trip data

Trip ID	Start Date	Start Hour	Start Station	End Date	End Hour	End Station	Duration	Subscriber Type

Date	Temperature	Dew	Humidity	Precipitation	Events	Zip

Feature Selection

Net Rate was evaluated for each station for each hour

- Number of cycles leaving stations
- Number of cycles ending at stations

Weather data for each station was obtained by joining station data with daily weather data

Lag variables were used to account for past three hours of activity at each station

Model Preparation & Results

Time series data for each station was used to train RandomForest model

Following libraries were used:

- Randomforest
- Catools
- Google's geocode API

Data for each set was divided into training and testing sets

- Model was trained on training set and validated on test set

RMSE for two stations are as follows

Station ID	RMSE
2	0.22
27	0.28

Results

Error distribution ranged between 0.1 and 0.5

Majority of the stations had RMSE error between 0.1 and 0.3