UC San Diego

Analysis of Uber & Lyft Prices

Group 13

Jiayuan Shen **John Xie**
Robert Young **Jialu Zhao**

Overview

UC San Diego

- Context
- Data Overview
- Data Visualization
- Data Prediction

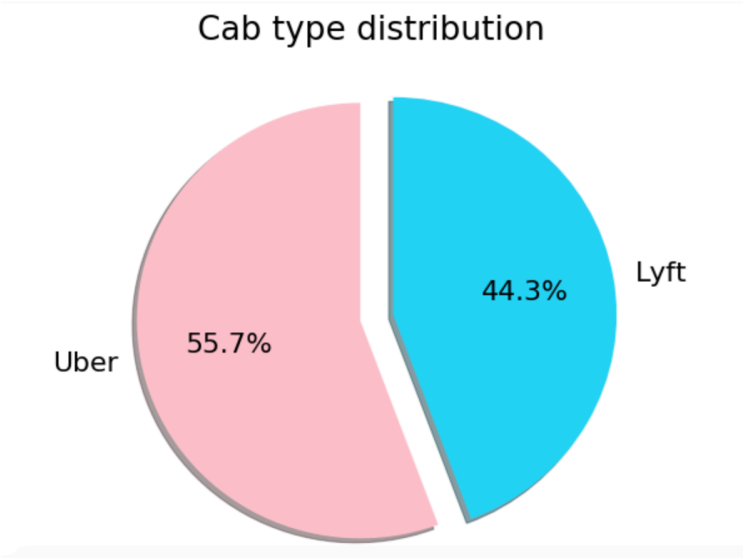
Context

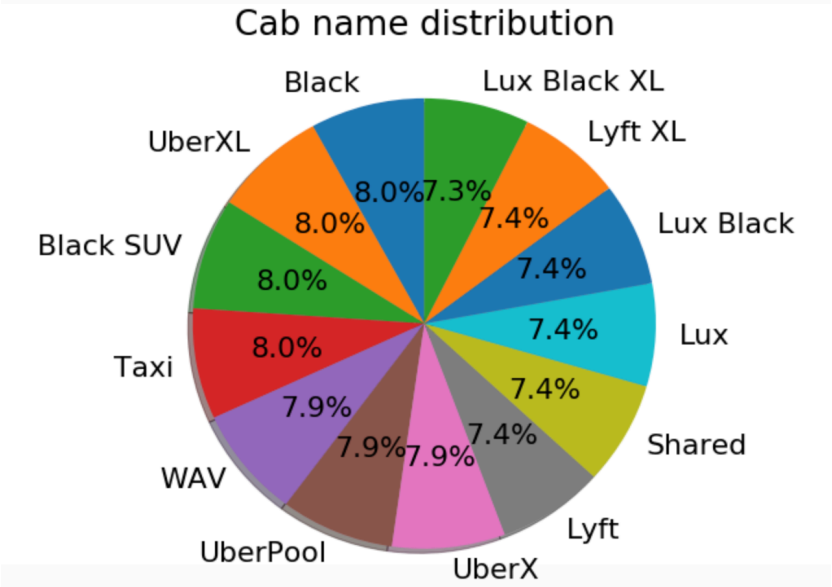
?

- **Uber and Lyft are the two dominant business rideshare apps in the states.**
- **Prices for those two rideshare apps depends on numerous factors such as weather, time, destination and etc.**

Data Overview

Dataset of cab rides collected for a week in Nov - Dec 2018 in Boston.
Collected at a regular interval of 5 mins.

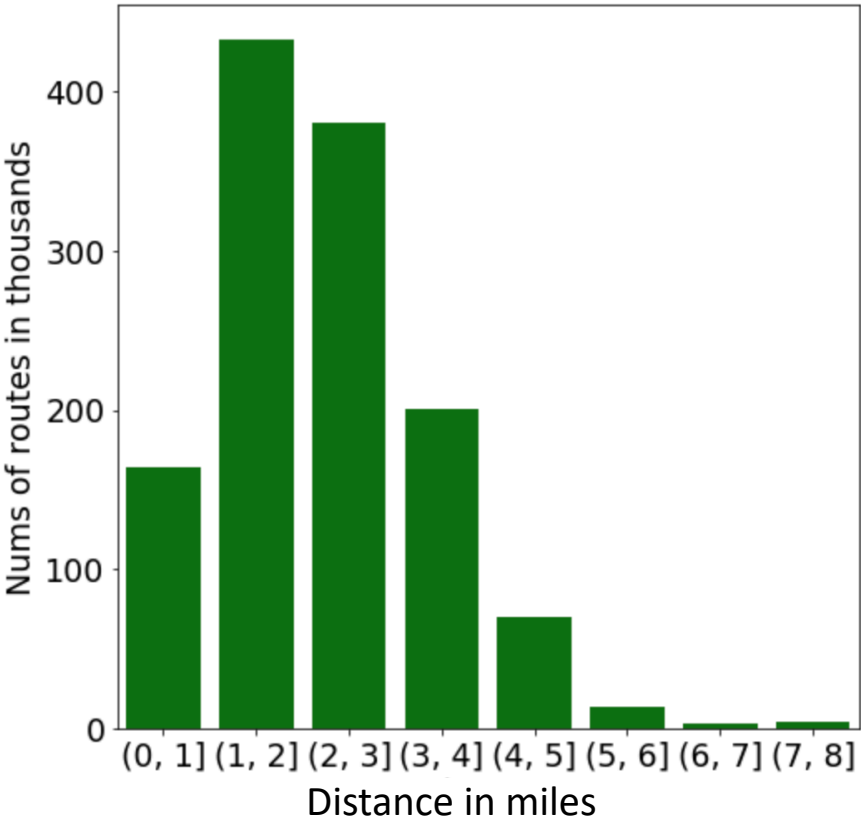




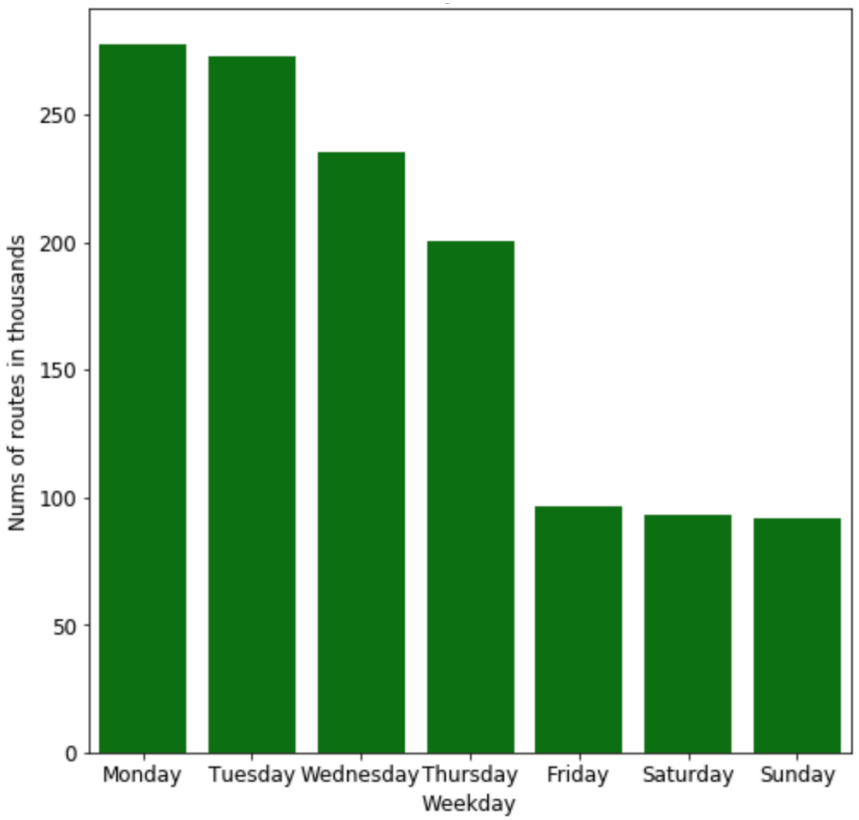
Data Overview

UC San Diego

Distance Distribution



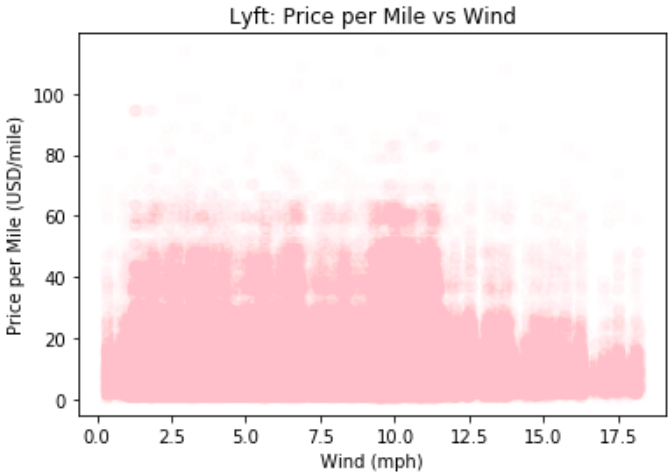
Weekday Distribution

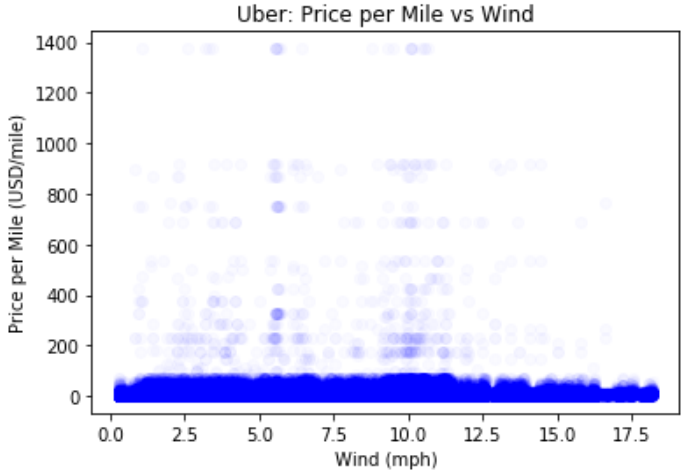


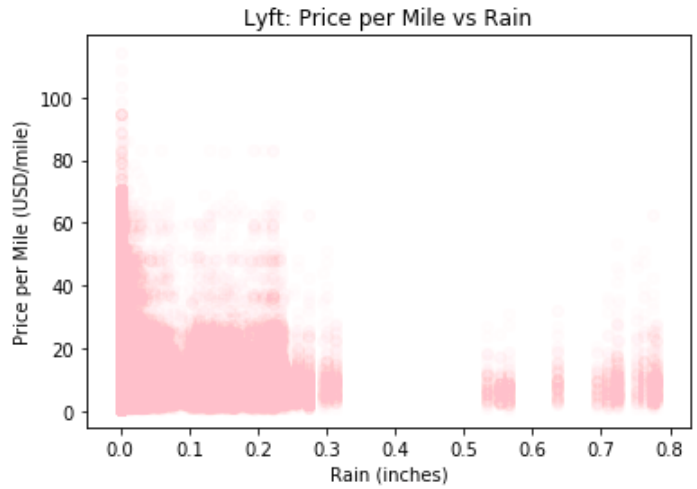
Data Visualization

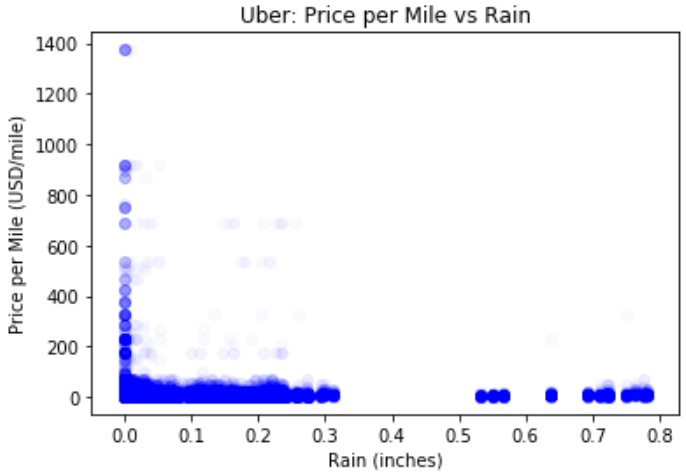
Uber vs Lyft: Effect of Weather on Price per Mile

Weather does not affect the prices of rides







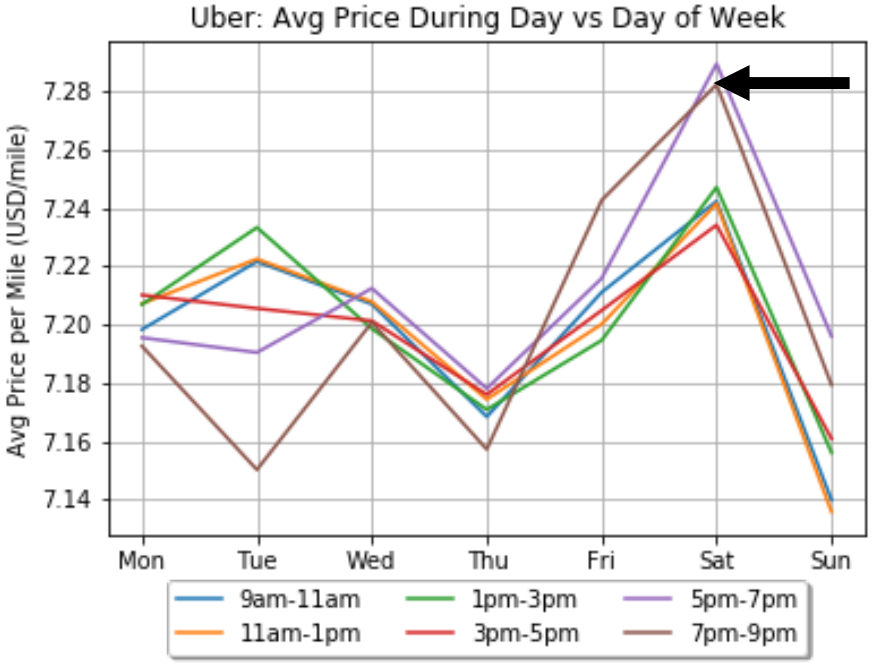


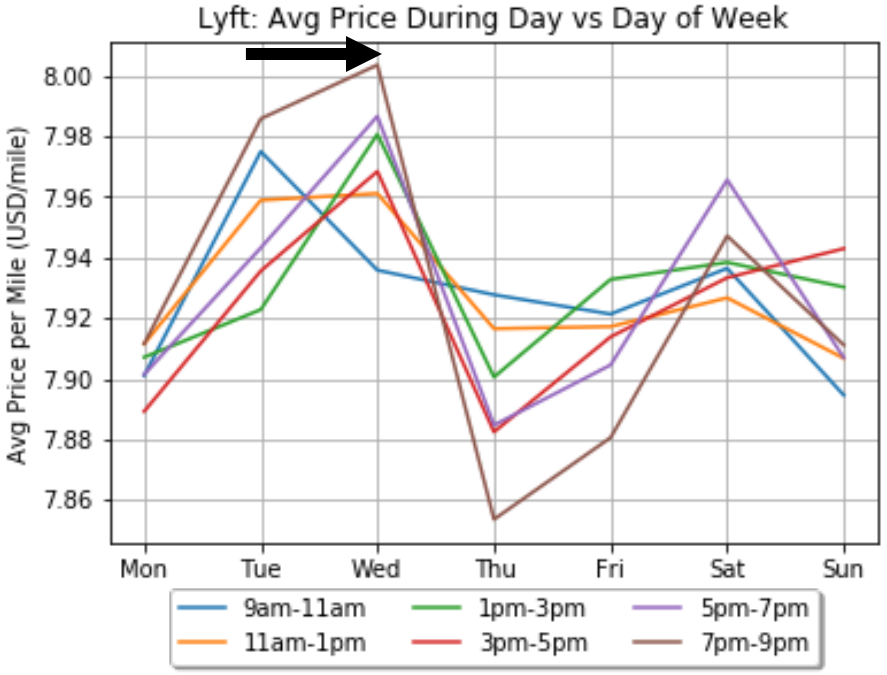
Data Visualization

UC San Diego

Uber vs Lyft: Price per Mile vs Day of Week

- Uber is generally, slightly less expensive than Lyft
- Uber's prices fluctuate less than Lyft's
- Prices spike from 5pm-9pm



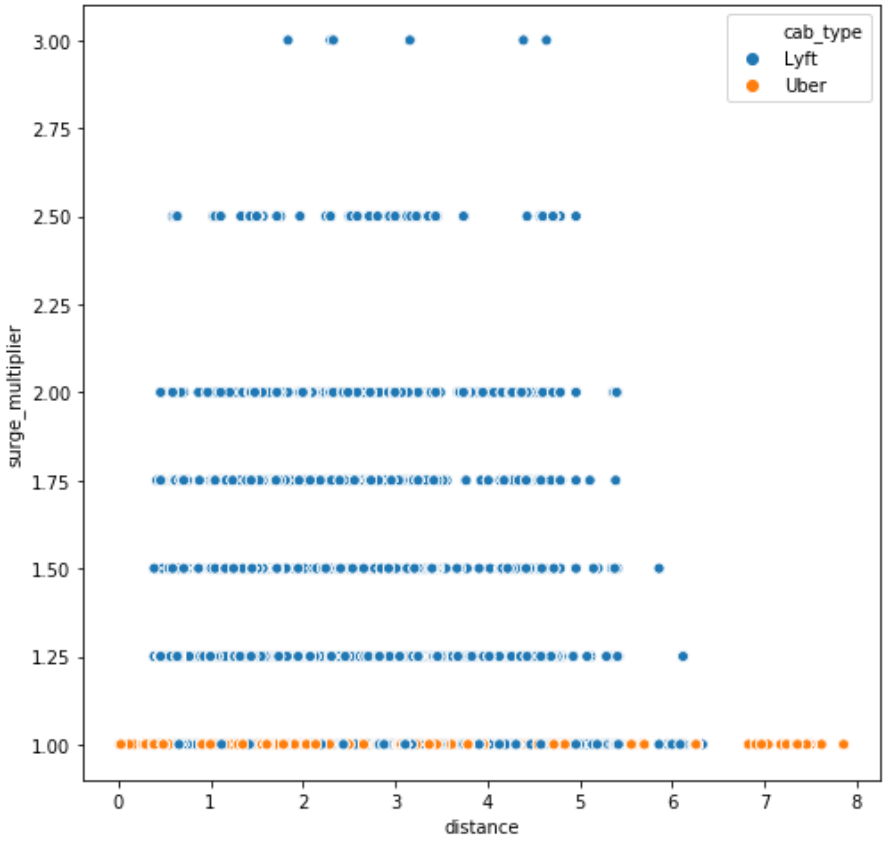


Data Visualization

UC San Diego

Uber vs Lyft: Surge Multiplier vs Distance

Uber Surge Multiplier is always 1

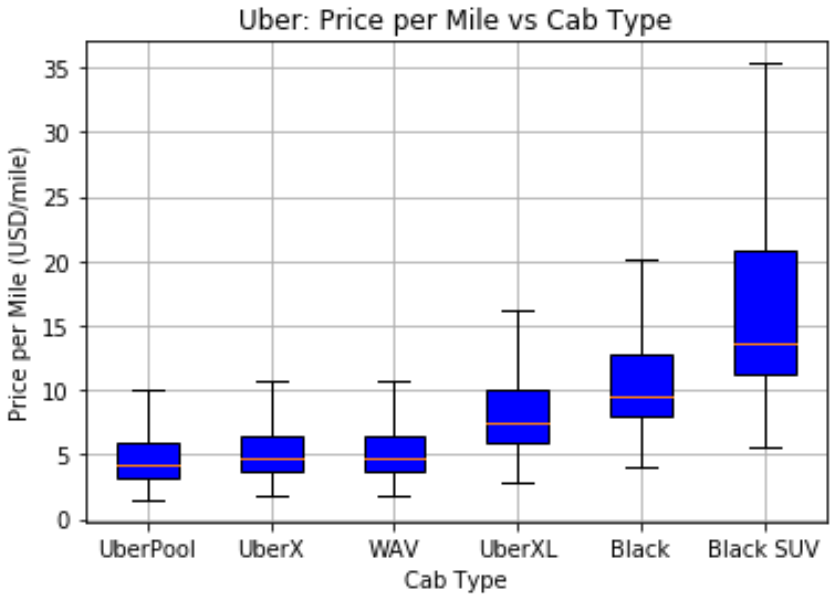


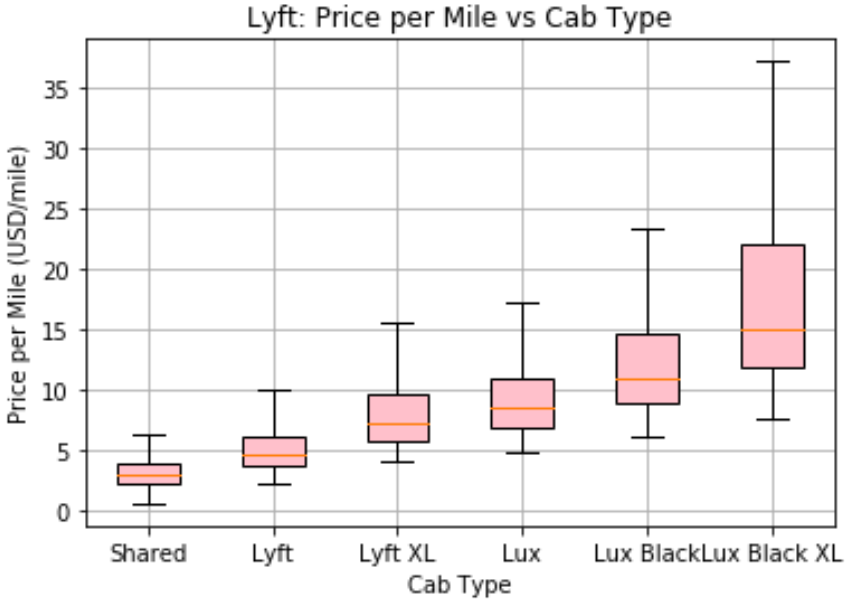
Data Visualization

UC San Diego

Uber vs Lyft: Price per Mile vs Cab Type

- Uber is generally cheaper than Lyft
- Lyft has much more price variation
- Better cab type leads higher price





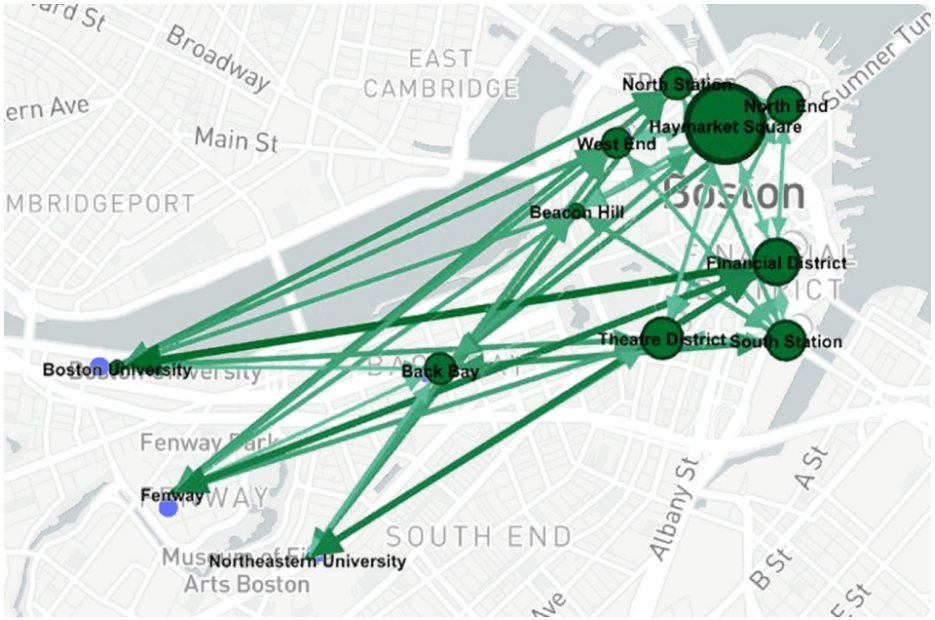
Data Visualization

UC San Diego

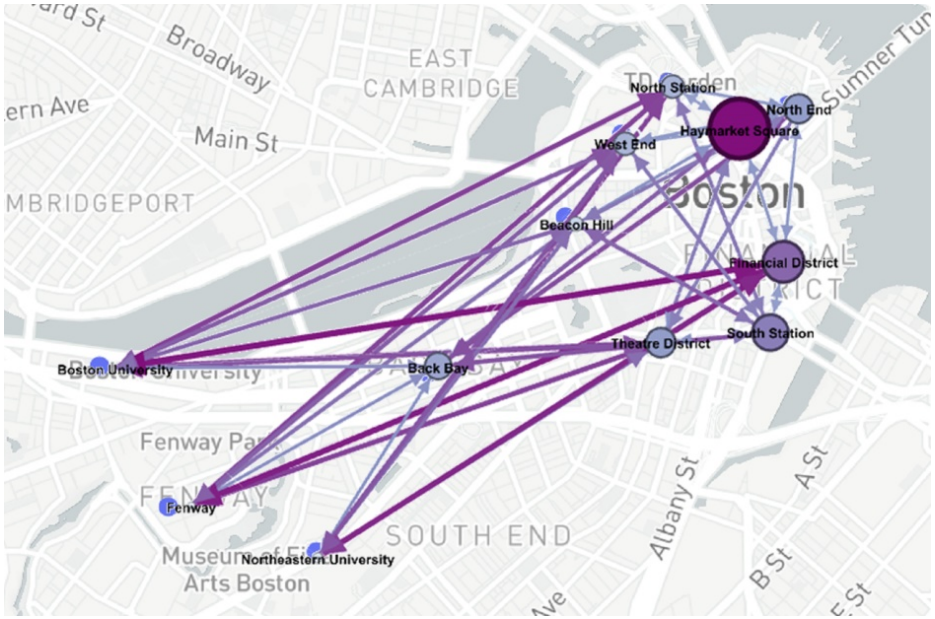
Price Networks for Uber & Lyft

- Edge weight = mean price
- Node weight = price/distance
- Suburb area is cheaper
- Downtown area node is quite large
- Generally: Lyft < Uber

Uber Price Network



Lyft Price Network



Data Prediction

UC San Diego

- Goal
  - Predict future Uber & Lyft rides price
- Training set
  - Splitting original dataset into 80% training set and 20% validation set.
- Test set
  - Manually collected 3000 entries from Uber & Lyft App.
- Model
  - Random Forest
  - SVM
  - Decision Tree

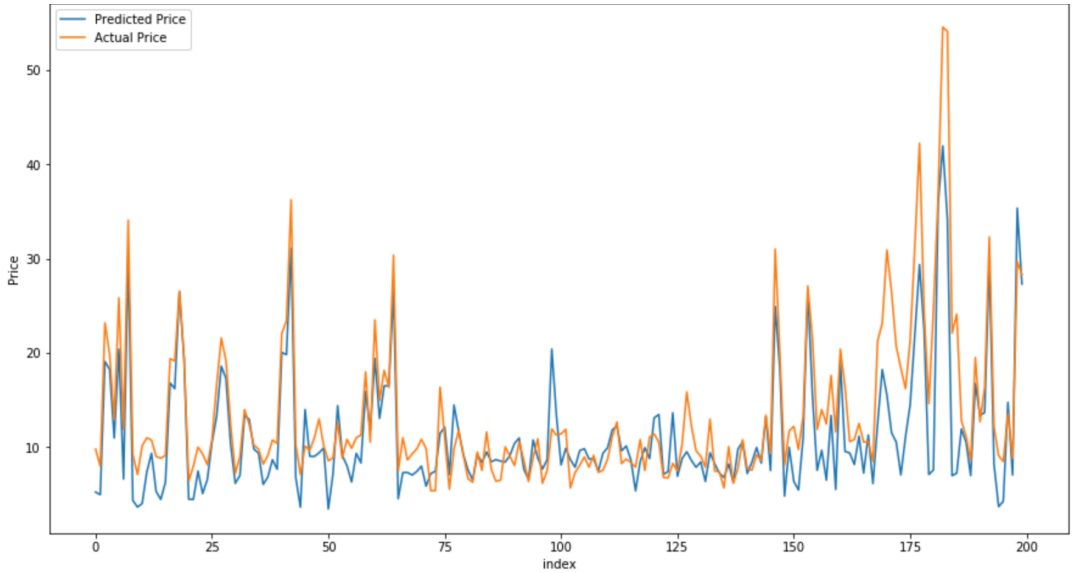
Data Prediction

UC San Diego

Accuracy	Training set	Validation set	Test set
Random Forest	95.95 %	89.4 %	81.99 %
Decision Tree	98.67 %	93.01 %	79.58 %
SVM	84.56 %	59.77 %	61.56 %

Random Forest:

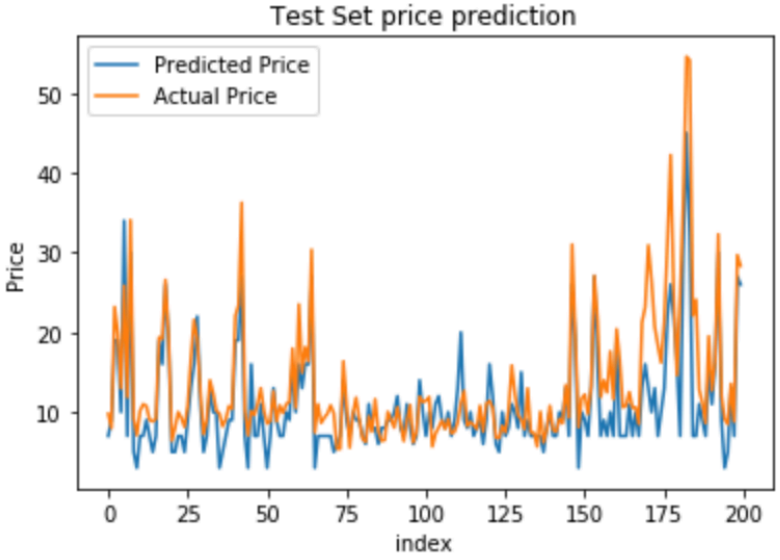
Test set price prediction



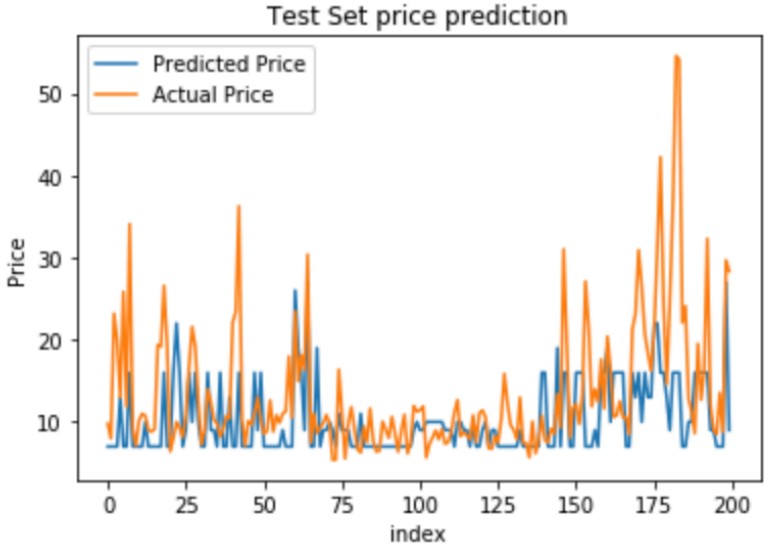
Data Prediction

UC San Diego

Decision Tree:



SVM:



Conclusion:

Random Forest achieves the best results.

Thank You