Uber vs Lyft prices

Capstone project by Samuel Ma
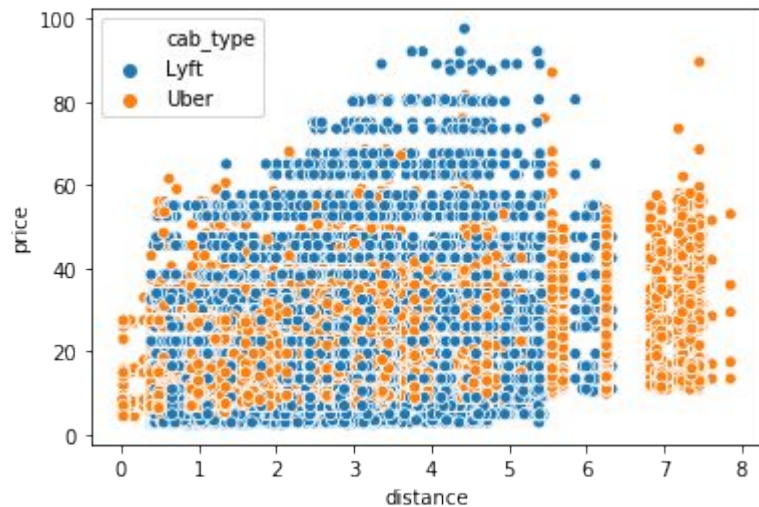
Data taken from Kaggle, Uber vs Lyft prices for a week in Nov-Dec '18
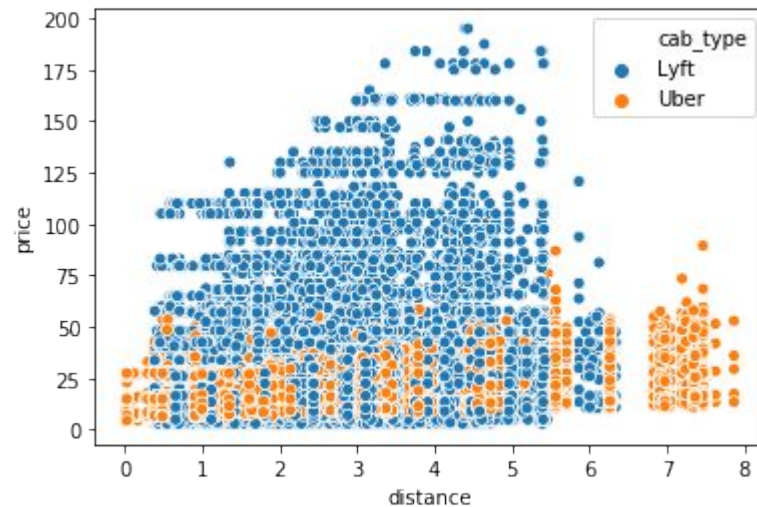
https://www.kaggle.com/ravi72munde/uber-lyft-cab-prices
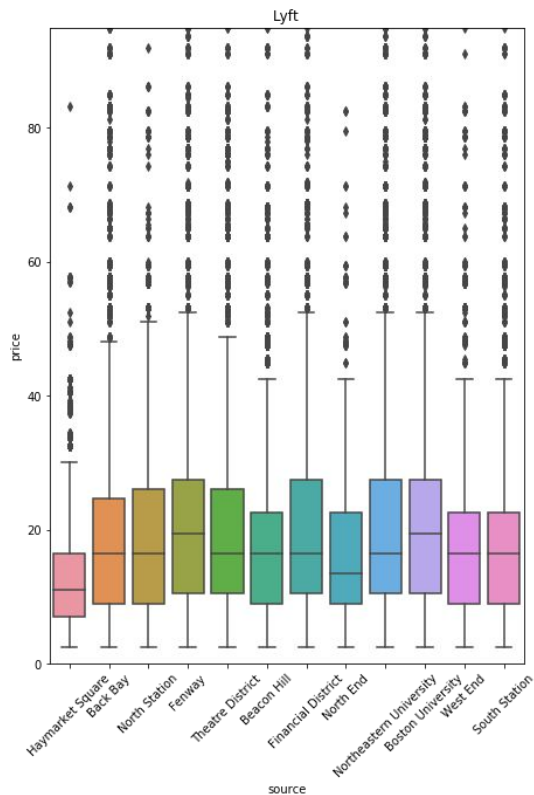
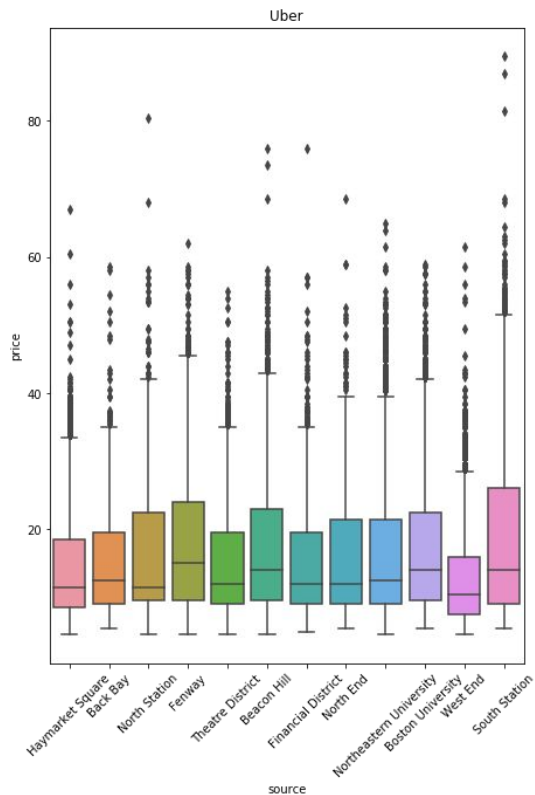***Clean all given data, but primarily work with cab.csv data
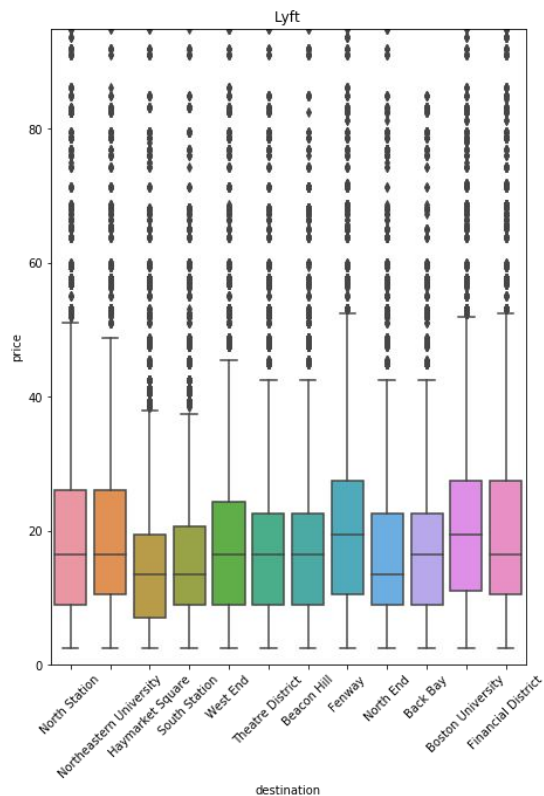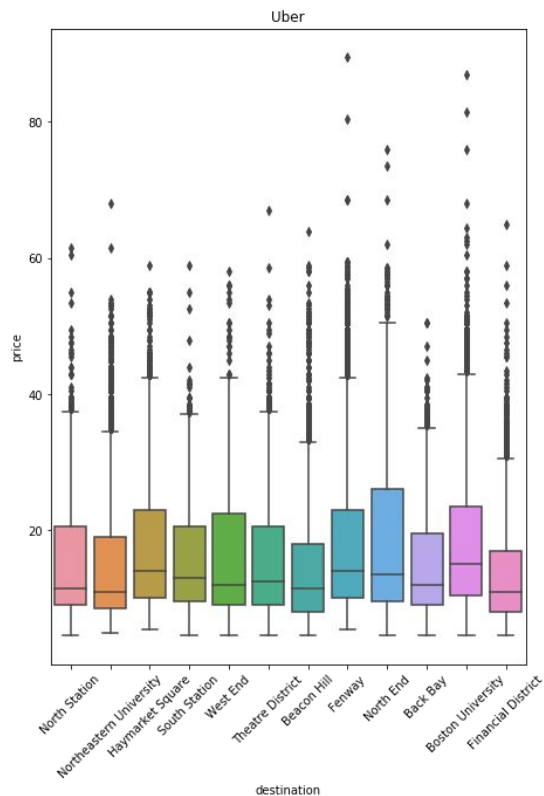
Initial Analysis

Raw data
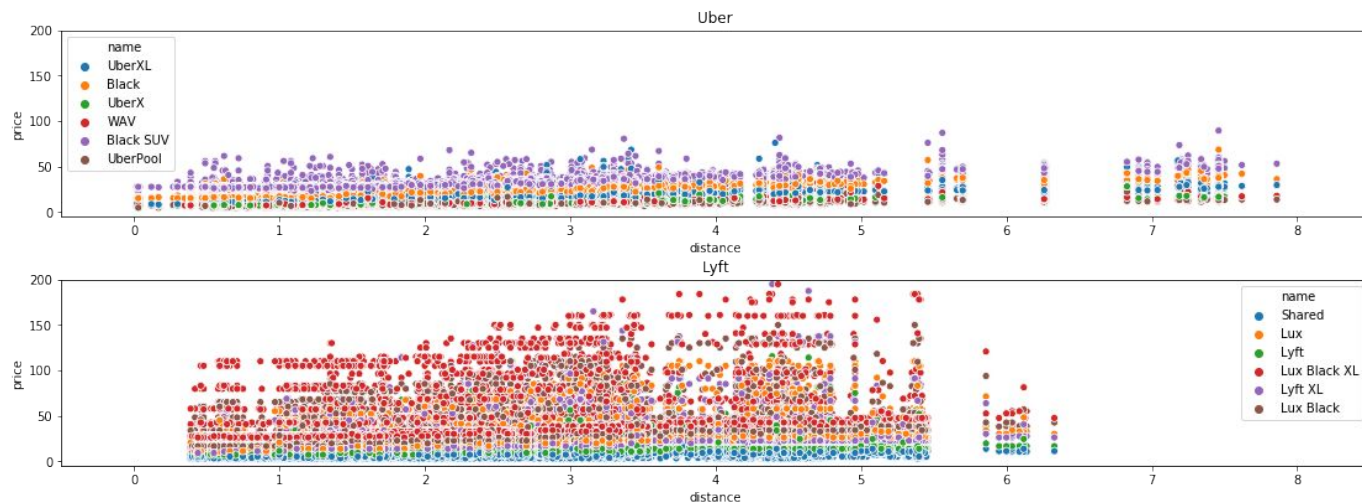
Cleaned data w/ Surge Factor

Location Comparisons (Sources)
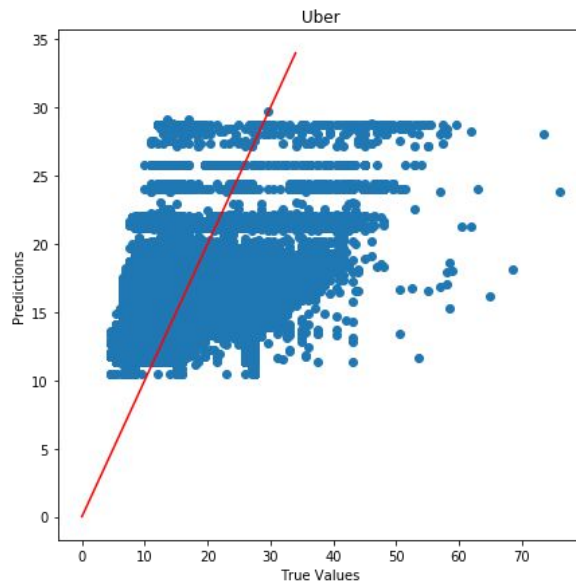
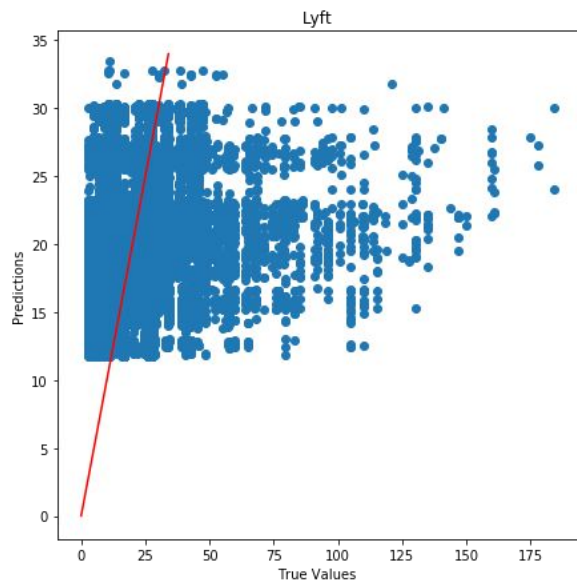# Location Comparisons (Destinations)

Different Services from each company
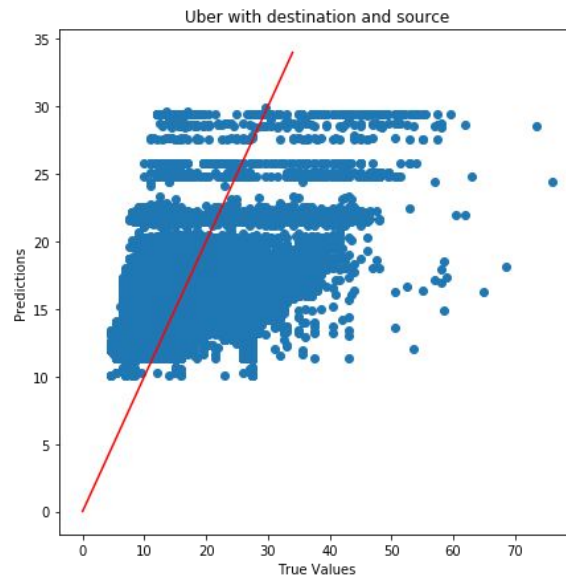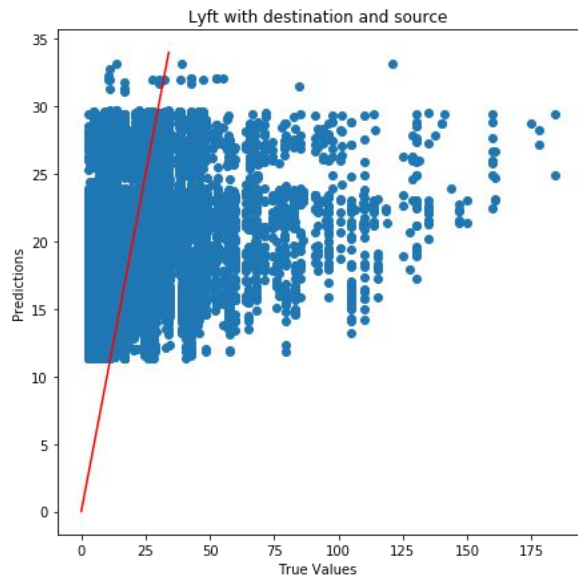
Machine Learning models

Linear Regression without source/destination

Machine Learning models

Linear Regression with source/destination

Not much difference than previous models

Machine Learning models

Random Forests

Without Dummies, we are only using distance as a factor.

With Dummies, we can see how important distance as well as both source and destination are in determining pricing.

Lyft

Distance importance: 0.87862

Uber

Distance importance 0.95012

Conclusions

- Distance is biggest factor

- Location has very small influence in pricing

- Very difficult to predict pricing, even if just given distance

- Potential hidden factors or variables we have yet to explore

Moving Forward

- Rain

- UberXL vs Lyft XL, UberPool vs Shared, etc…

- Time of day