PROJECT REPORT

UBER LYFT PRICING ANALYSIS

Project Contributors	Gaurav Chatterjee
	Mayur Sinha
	Megha Shirodkar

Introduction

Uber and Lyft's ride prices are not constant like public transport. They are greatly affected by the demand and supply of rides at a given time. So, what exactly drives this demand? The first guess would be the time of the day; times around 9 am and 5 pm should see the highest surges on account of people commuting to work/home. Another guess would be the weather; rain/snow should cause more people to take rides. The trip costs of Uber and Lyft are not static like those of public transport. They are also hugely impacted at a given time by the demand and supply of trips.

Data Description

There are two datasets. The cab ride data covers the different types of Uber & Lyft cabs taken and their price for the trip along with any surge. The weather data includes attributes such as temperature, rain, cloud, etc. for the listed locations.

Goal:

We tried to collect real-time information about Lyft and Uber rides and corresponding weather conditions. We chose a few locations in Boston from the map.
The Cab ride data covers various types of cabs for Uber & Lyft and their price for the given location. We can also find if there was any surge in the price during that time. Weather data contains weather attributes like temperature, rain, cloud, etc. for all the locations taken into consideration. The analysis is proposed to help start-ups to enter the transportation industry with existing competitors by strategizing the pricing structure, cut costs, and drive profits.

Execution:

- We started with the loading of data into the database using ingestion queries and segregated data from the whole dataset into cab rides and weather tables.
- SQLite studio is used to load the .csv files into database.
- We used matplotlib and seaborn to plot graphs and visualize the data.
- Pandas and NumPy is used for data cleansing and preprocessing.
- Did feature engineering on date and time columns for modeling purposes.
- We used Linear Regression model and Random Forest Regression model to predict the prices.

Analysis Methods:

Analyses techniques employed in the project are:

1. Statistical analysis of the numerical data to organize and summarize data and draw conclusions about cab fares.
2. Predictive analysis to make forecasts about future outcomes of cab fares based on current data.

Analysis:

- Our Analysis of cab and weather data has been done using the Random Forest Regression model to predict the prices based on the hour of the day, day of the week and weather conditions, and its effect on cab rides.
- We compared number of rides for Uber with Lyft and could observe that Uber is preferred by users when compared to Lyft.
- On the basis of price range, Lyft was found to be more expensive than Uber. Comparatively, people preferred Uber for longer distances and Lyft for shorter distances.
- Shared rides in Lyft are found to be cheaper than Uber Pool.
- Weather influences cab rides: During unfavorable weather, people prefer to travel by cabs and we found Monday to have maximum number of rides booked due to high precipitation of rainfall.
- Peak times affect cab rides: Maximum rides were booked during 6 PM – 8 PM, generally during the end of office hours.

Conclusion and Future Research Directions:

We were successfully able predict the prices of Lyft with an accuracy of **81.76%**. We plan to perform some feature engineering and improve our prediction model in future.