**MSIS 670 FINAL PROJECT REPORT**

**DATA VISUALIZATION OF UBER & LYFT PRICES USING TABLEAU**

For the final project of Business Intelligence, we have decided to base our work on the topic “Data Visualization of Uber and Lyft Prices using Tableau”. The purpose of this project is to visualize the pricing strategies of Uber and Lyft using data visualization techniques in Tableau. We will analyze the prices of rides across different cities in the United States and visualize the trends in prices over time. This report aims to provide insights into the pricing strategies of Uber and Lyft, as well as help consumers make informed decisions about which service to use based on their budget and travel needs. The visualizations will be created using the Tableau software package.

**Introduction to the Dataset:**

The dataset being used is publicly available.

**Cab Data**

<https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices?select=cab_rides.csv>

**Weather Data**

<https://www.kaggle.com/datasets/ravi72munde/uber-lyft-cab-prices?select=weather.csv>

The two links provided refer to datasets available on Kaggle, a platform for data science enthusiasts to share, discover and explore data. The first link contains a dataset called "Uber-Lyft Cab Prices'' that provides information about the cab rides taken by customers of Uber and Lyft in various cities in the United States. The dataset includes features such as distance, cab type, time stamp, destination, source, price, surge multiplier, and unique identifiers for the rides. The second link contains a dataset called "Weather" that provides information about weather conditions in various cities in the United States. The dataset includes features such as temperature, cloud cover percentage, air pressure, amount of rain, time stamp, relative humidity, wind speed, and location of the weather station. Both datasets offer valuable insights and can be used for various data analysis and modelling tasks.

**Data Description:**

**Strength of Datasets:**

"Uber-Lyft Cab Prices" dataset: The dataset contains a large amount of data on cab rides taken by customers of Uber and Lyft in various cities in the United States, which can be useful for analyzing trends and patterns in the ride-sharing industry. The dataset includes a range of features such as distance, cab type, time stamp, destination, source, price, and surge multiplier, which can be used to build predictive models for estimating the price of cab rides. The dataset is well-organized and structured, making it easy to analyze and work with.

"Weather" dataset: The dataset provides a comprehensive view of weather conditions in various cities in the United States, which can be useful for analyzing trends and patterns in weather data. The dataset includes a range of features such as temperature, cloud cover percentage, air pressure, amount of rain, relative humidity, wind speed, and location of the weather station, which can be used to build predictive models for estimating weather conditions. The dataset covers a wide range of dates and locations, providing a rich source of data for studying weather patterns over time.

**Limitations of Dataset:**

"Uber-Lyft Cab Prices" dataset: The dataset only includes data on cab rides taken by customers of Uber and Lyft, which may not be representative of the entire ride-sharing industry. The dataset does not include any demographic information about the customers, which may limit the ability to conduct deeper analyses on the factors influencing the price of cab rides. The dataset only includes data from a limited number of cities in the United States, which may limit the generalizability of any findings.

"Weather" dataset: The dataset only covers weather conditions in a limited number of cities in the United States, which may limit the generalizability of any findings. The dataset only includes data for a one week period, which may limit the ability to study long-term weather patterns and trends. The dataset does not include any information on extreme weather events such as hurricanes, tornadoes, or blizzards, which may limit the usefulness of the dataset for studying the impacts of such events.

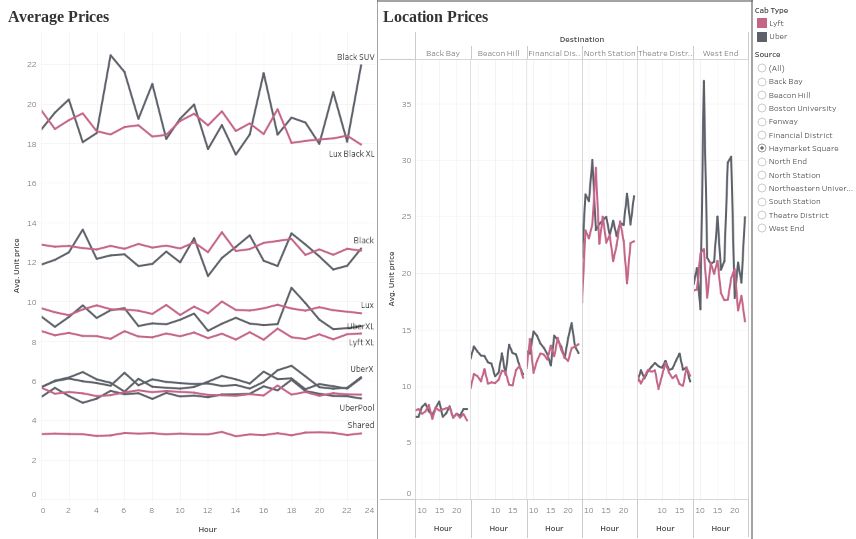
**Data Cleaning and Integration**

As we had a good quality of data, there was no need to make extensive changes for data cleaning. However, we did work on the Epoch timestamp to extract the hours and days from it. Integrating the data was a relatively simple process since we had a common timeframe and location in both datasets used for the visualizations, which allowed us to merge them directly in Tableau.

**Story telling by Data Visualization (using Tableau)**

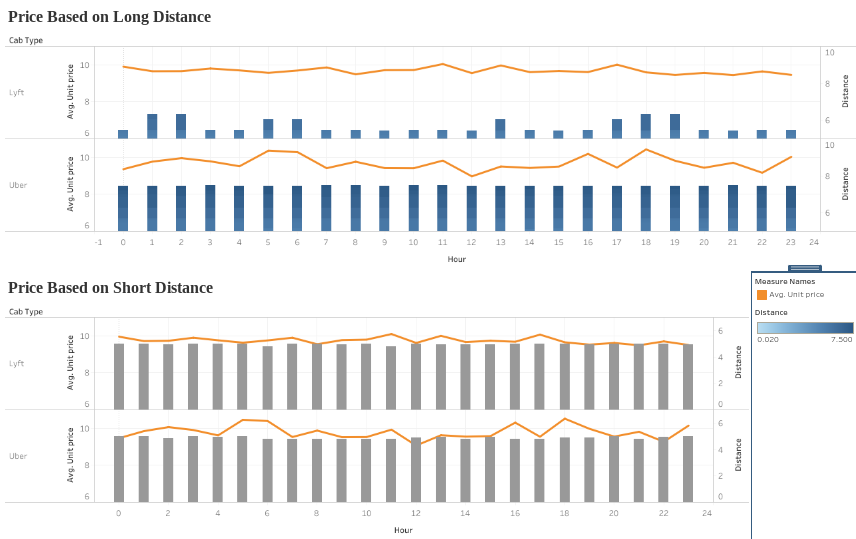
**Question 1: What are the factors that influence Uber and Lyft cab prices, and how do they vary by city and time of day?**

**Descriptive: What are the average cab prices for Uber and Lyft in each city?**

****

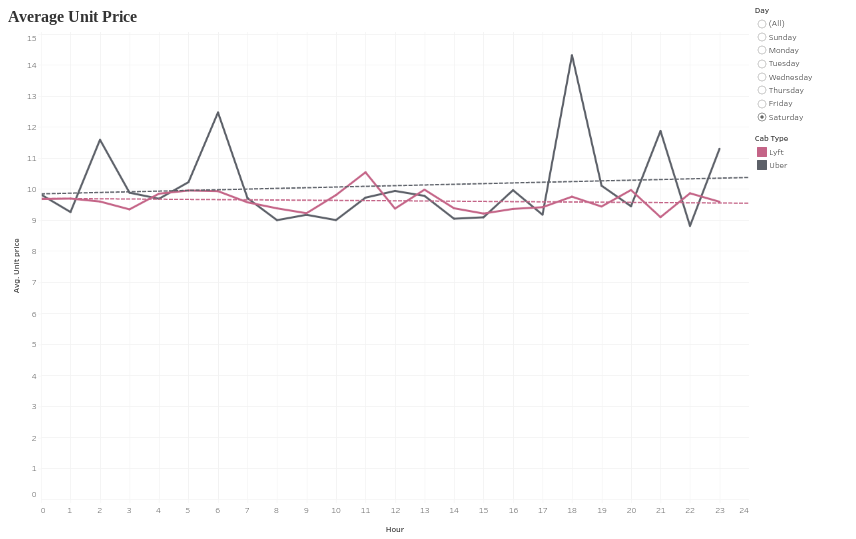
We created two sheets to visualize the average unit prices and hours for different cab types. The first sheet shows the overall average unit price and hour for each cab type. The second sheet shows the average unit price and hour for a particular source to destination. We used Tableau software to create the visualizations. The visualization of the average cab prices by hour and source-destination helps in understanding the variations in the cab fares.

**Predictive: Can we predict cab prices based on factors such as distance, time of day?**



For this dashboard, we created two sheets using Tableau software to visualize the average unit prices based on distance and time. The first sheet shows the average unit price based on the shortest distance for a specific time. The second sheet shows the average unit price based on the longest distance for a specific time. Our analysis revealed that the distance and time are significant factors that affect cab prices. The average unit price for the shortest distance is lower compared to the average unit price for the longest distance. This indicates that the distance of the cab ride has a direct impact on the cab fares.

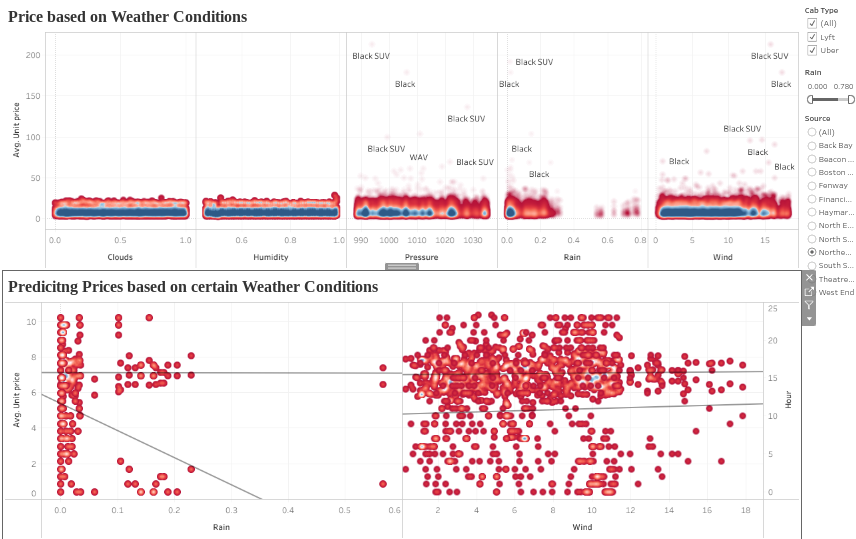
**Prescriptive: Can we recommend the best times to use Uber and Lyft to get the lowest prices?**

****

We created a sheet using Tableau software to visualize the average unit prices by time of day and cab type with trend lines. The sheet shows the average unit price for each cab type during different times of the day, and the trend line shows the estimate of the trend going on for the prices. We used the sheet to determine which cab type service will be cheaper during specific times of the day. Our analysis revealed that the time of day and cab type are significant factors that affect cab prices. The average unit price is higher during peak hours compared to non-peak hours for all cab types. However, the analysis shows that some cab types are cheaper during specific times of the day.

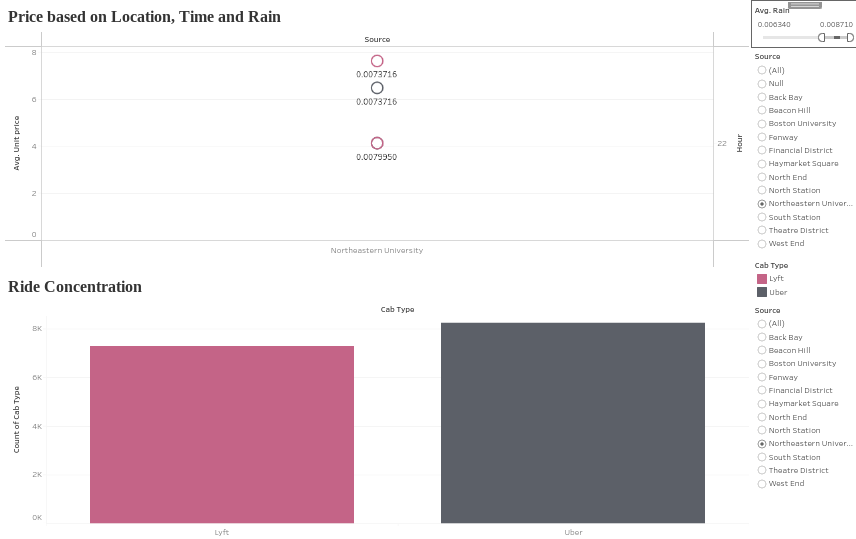
**Question 2: How do weather conditions affect Uber and Lyft cab prices?**

**Descriptive: How do typical weather conditions affect the prices for Uber and Lyft ?**

****

We created a dashboard using Tableau software to visualize the cab prices based on weather factors and time of day. The dashboard has two sheets. In the first sheet, we added clouds, humidity, pressure, rain, and wind to the "Columns" section and average unit price to the "Rows" section. Here we compare the prices of different cab types based on the weather factors. The second sheet contains the trend line, which shows the estimated trend line based on the average unit price for a specific time of day, also containing the source (location) from where the cab is booked. Our analysis revealed that weather factors have a significant impact on cab prices. The dashboard shows that during rainy weather, the cab prices are generally higher compared to non-rainy weather conditions. Cab prices are also higher during peak hours compared to non-peak hours for all cab types.

**Predictive: How to choose a ride service at certain times based on rain?**

****

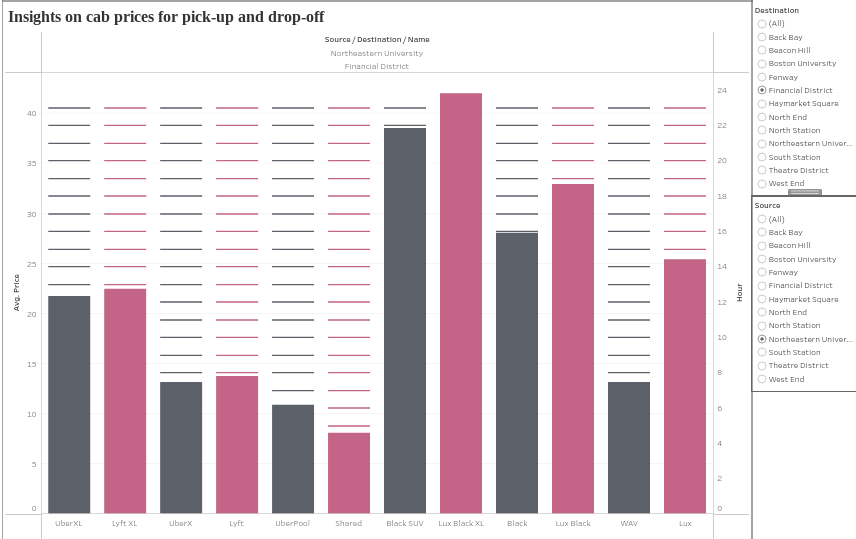
To create this dashboard, we developed a visualization that predicts the average unit price of a cab from a particular source based on the presence of rain during a specific time of day.

**Prescriptive: How should Uber and Lyft adjust prices based on weather conditions?**

Based on the visualization, it can be observed that users tend to book a cab service during the onset of rain or heavy rainfall, as there are no cab rides booked while it is already raining. This could be attributed to users waiting for the rain to stop or booking when they see it has started raining and might get worse. Additionally, it is evident that Uber is more responsive to rain as its pricing drastically changes with the onset of rain, whereas Lyft shows a relatively smaller change in pricing with changes in rain.

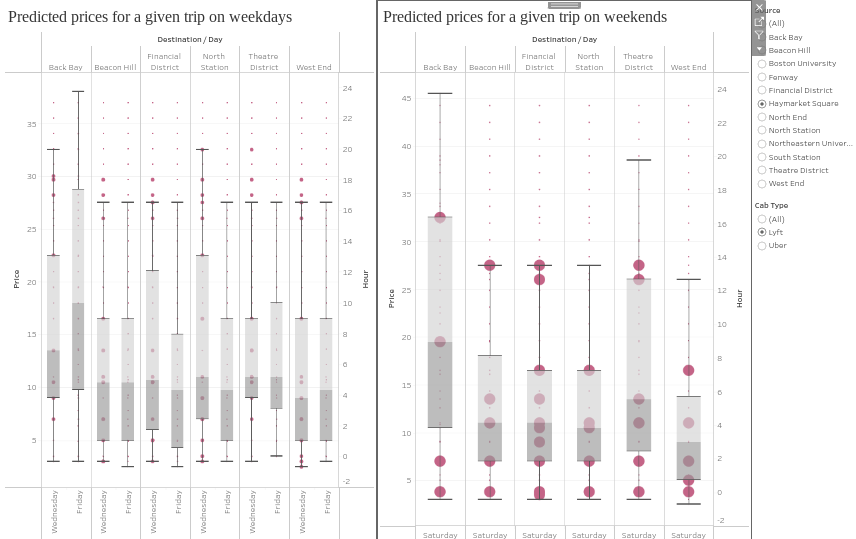
**Question 3: How do cab prices compare between Uber and Lyft, and how can users choose the most cost-effective option?**

**Descriptive: How do cab prices vary based on pick-up and drop-off locations?**

****

In order to create this dashboard, we developed visualizations for various cab types offered by Uber and Lyft. The comparison depicts the cab prices for each location based on the specific time of the day for specific pick-up and drop-off locations.

**Predictive: Can we predict which service (Uber or Lyft) will be cheaper for a given trip?**

****

To create this dashboard, we developed two sheets that provide insights on the factors affecting the cab prices. The first sheet shows the average unit price of cab types based on weather conditions at a specific time of the day. This sheet contains a scatter plot with the weather conditions as the independent variable and the average unit price as the dependent variable. The second sheet shows the average prices (low, mean, and high) for Uber and Lyft based on several factors, including the rain, source, and destination at a specific time of the day. The analysis can also help cab companies to develop pricing strategies that cater to different customer segments based on the time of day, cab type preferences and weather conditions.

**Prescriptive: Can we recommend which service (Uber or Lyft) to use based on the price and other factors such as distance and time of day?**

To create this visualisation, we have included graphs that depict the prices of various cab options based on location, time of day and rain conditions. These graphs enable a clear comparison of the prices offered by the different options.

**Lesson-Learned:**

Our group has learned the importance of understanding the data that we are working with and how it can be better represented through visualization. We have also learned how to use different software programs to create and interpret data visualization, and how to use them to analyse and interpret data for insights and trends. Additionally, we have gained an understanding of the various types of data visualization available, and how each type can be used to better understand a particular data set. Finally, we have learned how to effectively communicate our findings through data visualization, making the data more accessible and understandable to a wider audience.