Prediction of Cab Prices

Uber and Lyft Data

Group members:

1. Elena Mwangi
2. Gladwel Wanjau
3. Mercy Gichuhi
4. Elizabeth Josephine
5. Wanjiku Githu

Problem Statement:

Uber and Lyft ride prices are not constant like public transport. They are greatly affected by the demand and supply of rides at a given time.
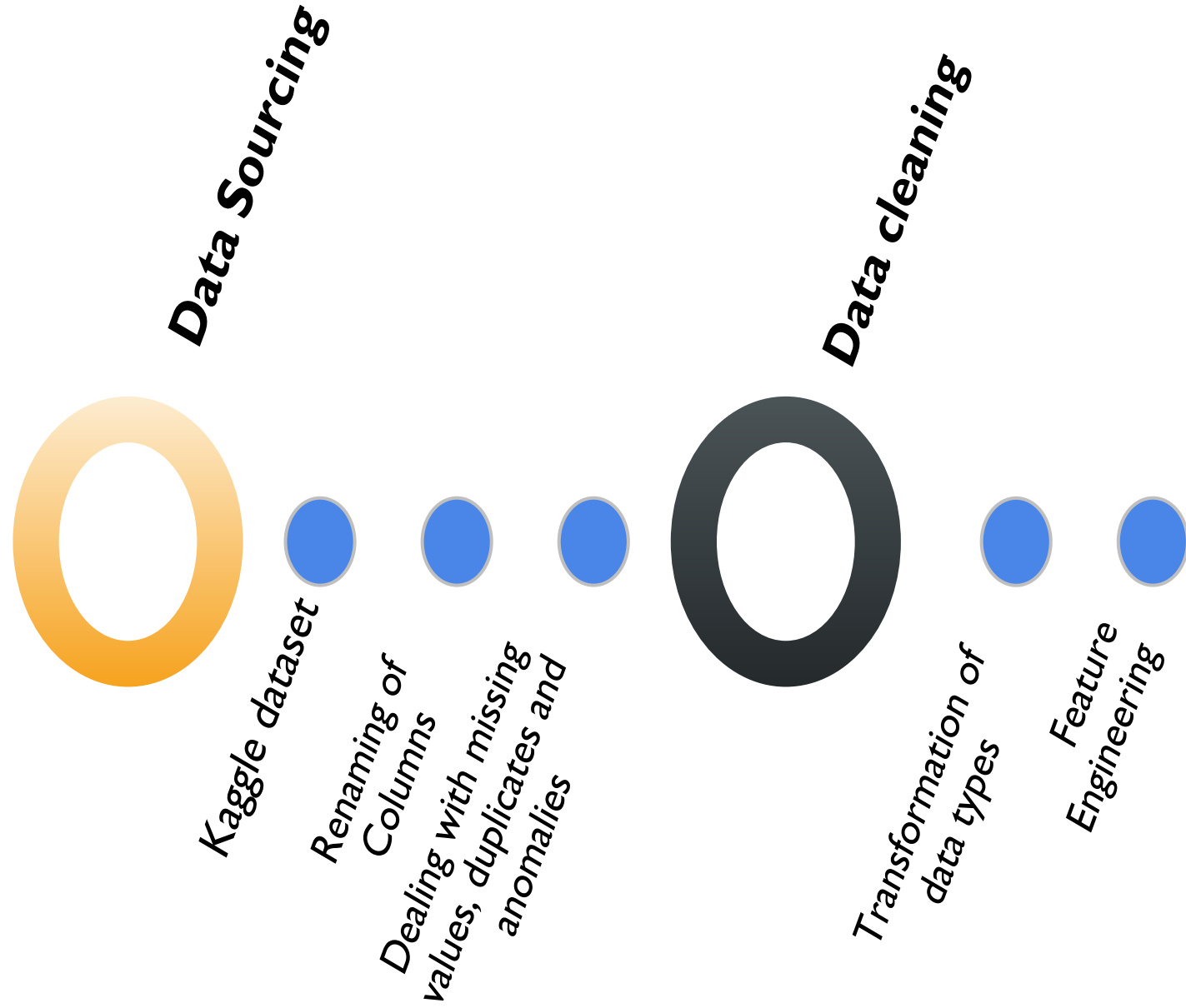We would like to understand what affects the prices of cabs.

Objectives:

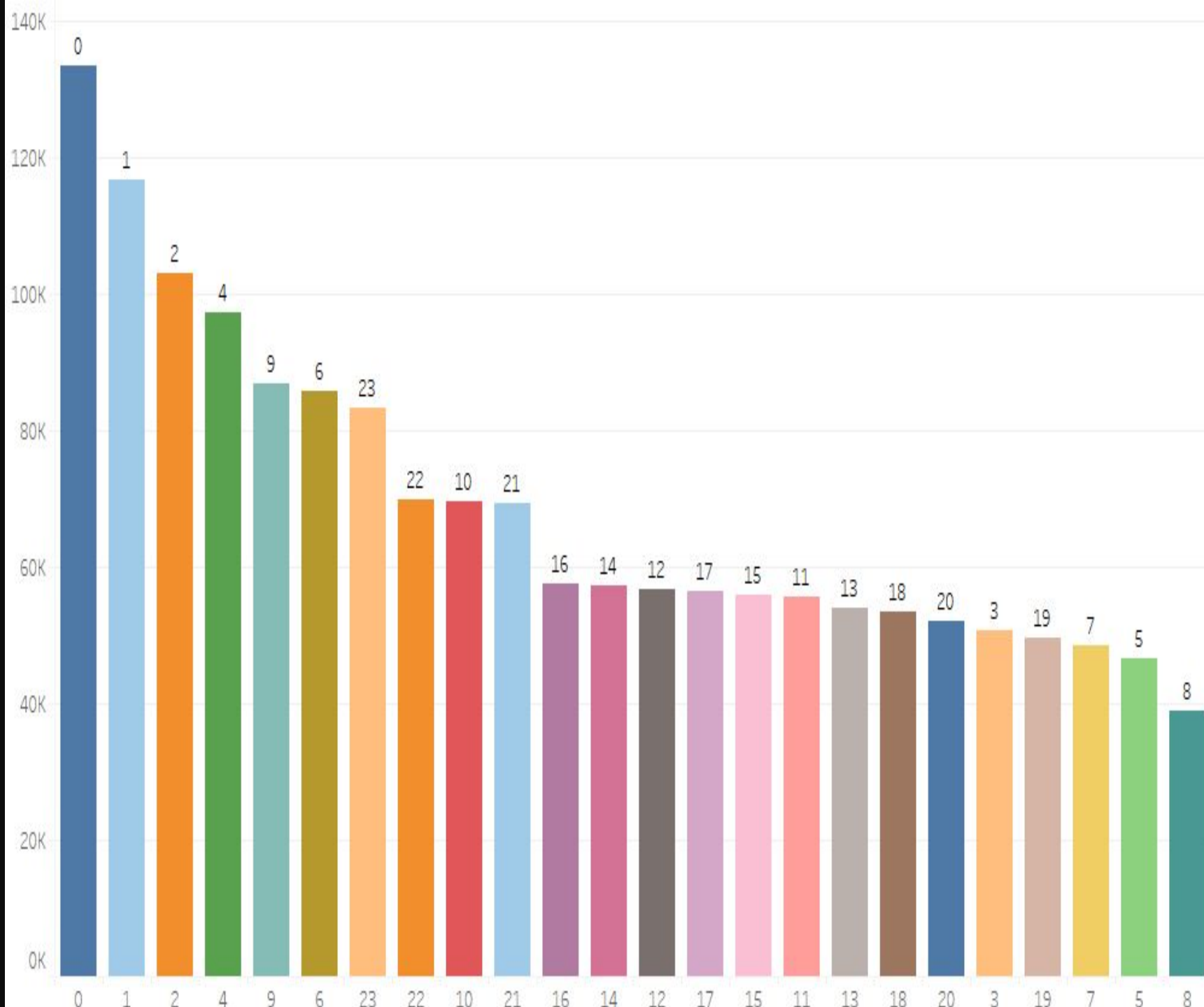1. To understand the factors that affect the price of a cab.

2. To predict trip prices based on this factors.
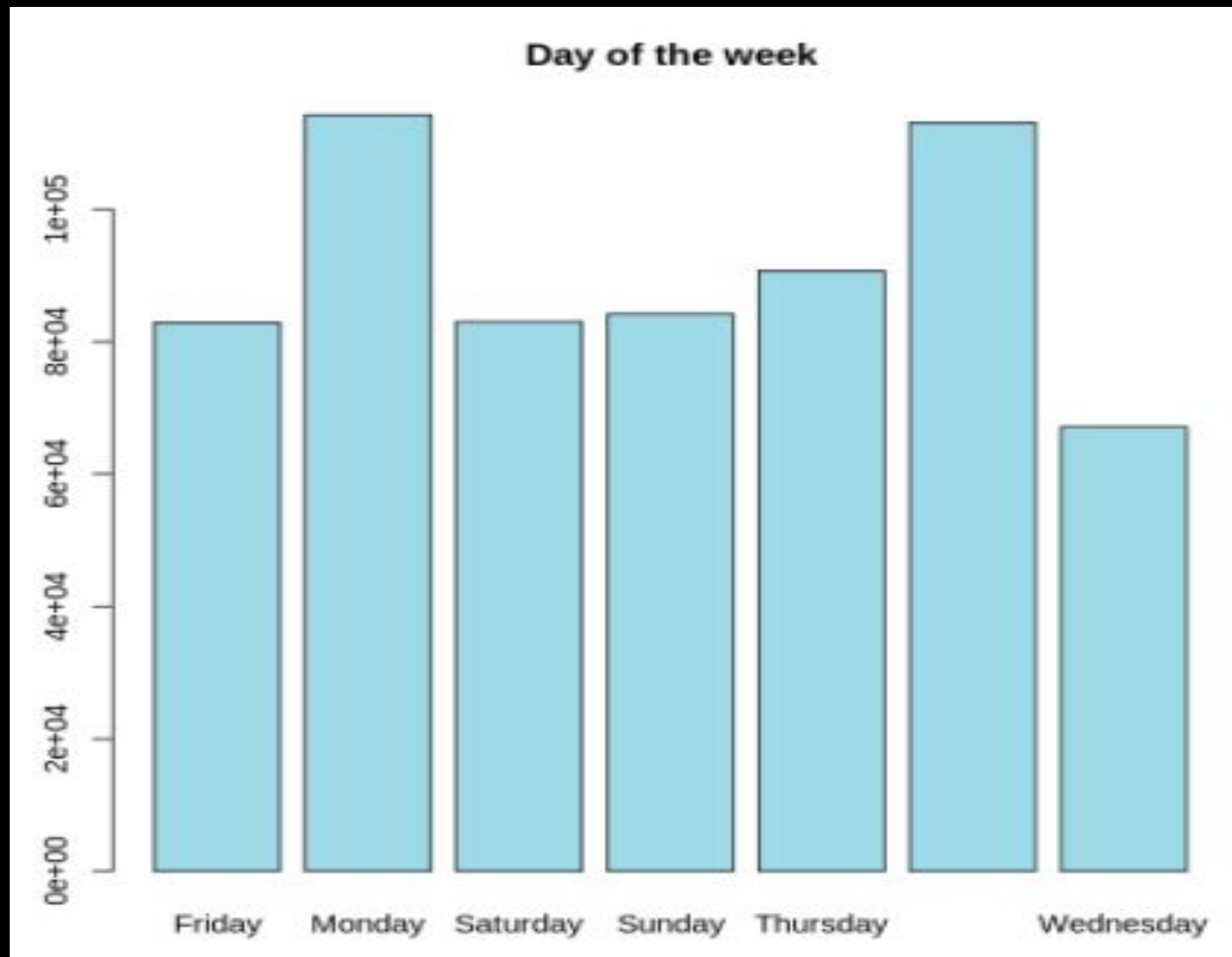
Data Sourcing and cleaning

Data Sourcing

Data cleaning

Kaggle dataset

Renaming of Columns

Dealing with missing values, duplicates and anomalies

Transformation of data types

Feature Engineering

Data Analysis
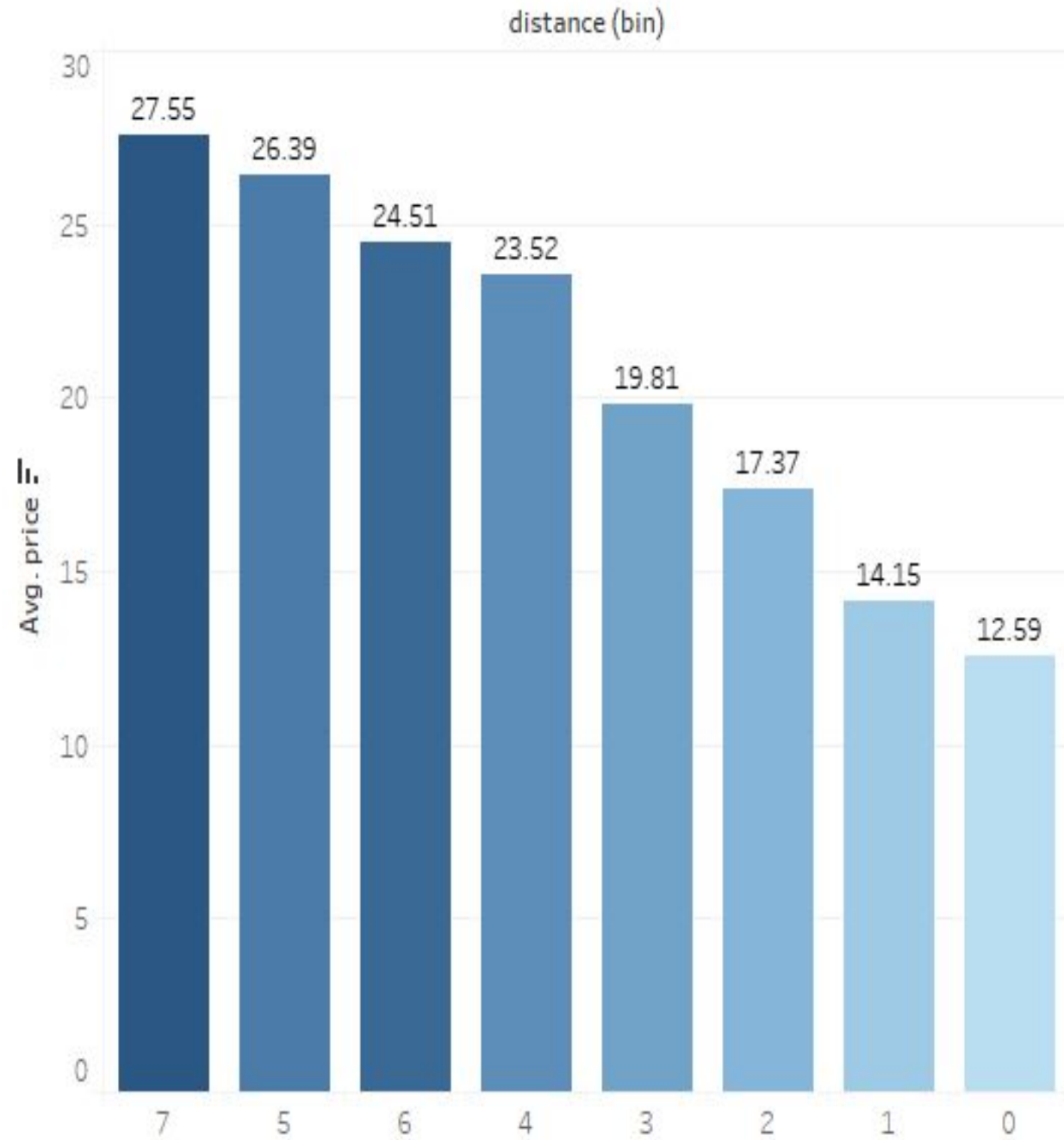
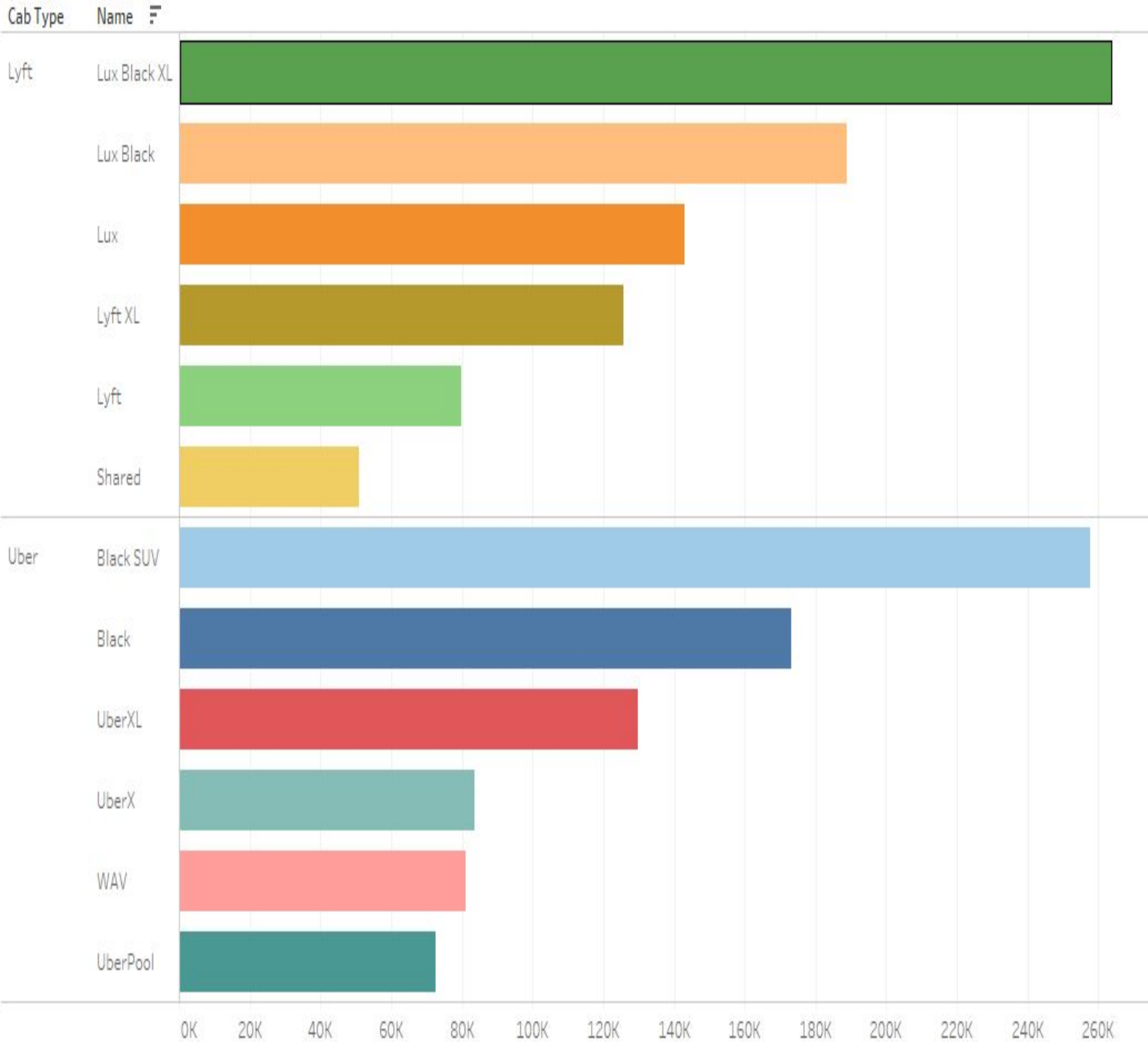Prices across the hours in a day

Rides per day of the week:

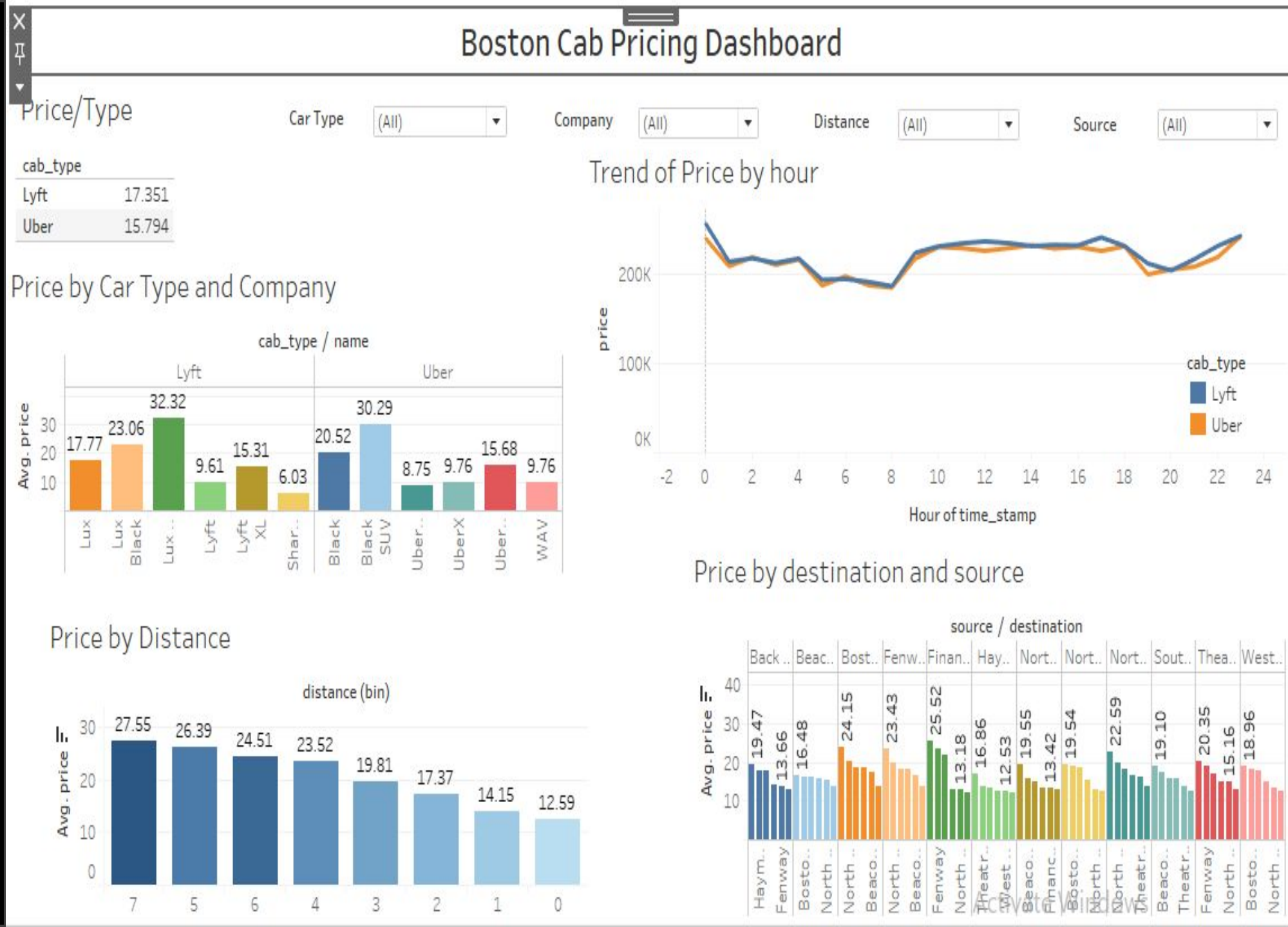**Effect of distance on Price:**

Price by Distance

distance (bin)

There are different factors that affect pricing of a ride therefore the customer should look at these:

❖ Timing
❖ Distance
❖ Company Type (Lyft has generally higher prices than uber)
❖ Type of Car (The most expensive for Lyft is Black Lux Xl and for Uber is Black SUV and the cheapest are Uber Pool and Shared)

Tableau dashboard

Boston Cab Pricing Dashboard

Hypothesis and Modelling:

H0: No difference between the prices of the cab companies

H1: Difference in the pricing of the cab companies

Rejected the null hypothesis & accepted the alternative

H0: There's no difference in pricing of car types

H1: Difference in the pricing of car types

Rejected the null hypothesis & accepted the alternative

Linear, Ridge and Lasso Regression: RMSE of 0.2696 and R Squared of 0.9273

Random Forest Regressor: RMSE: 0.3883 & R Squared: 0.849

Conclusions:

❖ The most popular destination is the Financial District.

❖ Luxurious cabs have higher average prices compared to the shared cabs.

❖ The most popular hour for ride requests is between 11:00 pm and 12:00 midnight.

❖ The most expensive hour to request for uber rides is at 9:00 am and 8:00 pm while for lyft is between 11:00 pm and 12:00 midnight.

❖ The regression model has a higher R Squared score compared to the Random Forest Regressor.

Recommendations:

❖ For an increase in price, the both companies can ensure there are sufficient luxurious cars in the different locations since their average price is higher.

❖ Increase cabs during the hours 11pm and 12.midnight, since they are the most popular hrs.

THANK YOU