Exploratory Data Analysis
G2M CAN FOR INVESTMENT FIRM

BACKGROUND – G2M CAB INDUSTRY

XYZ, a U.S.-based private equity firm, is exploring investment opportunities in the rapidly expanding cab industry, which has seen significant growth and the emergence of key players in recent years. The firm's objective is to gain insights that will aid in identifying the most suitable cab company for investment.

The analysis to achieve this objective is structured into four key segments:

1.**Data Understanding**: This involves a thorough analysis of the cab industry, including market trends, key players, and growth patterns. The aim is to develop a comprehensive understanding of the industry dynamics.

2.**Forecasting Profit and Number of Rides**: This section focuses on projecting the future profitability and ride volume for each cab company. Advanced forecasting techniques will be used to predict these key performance indicators.

3.**Identifying the Most Profitable Cab Company**: In this critical phase, the analysis will pinpoint which cab company currently stands as the most profitable. This involves comparing various metrics and financial performances across different companies.

4.**Investment Recommendations**: Based on the insights gathered from the previous sections, the final part of the analysis will offer strategic recommendations for XYZ's investment in the cab industry, ensuring the decision is data-driven and aligns with the firm's investment goals.

Data Exploration

•

We have a dataset with 24 features, which includes 9 derived features. The data spans from January 31, 2016, to December 31, 2018, with a total of 355,032 data points.

Here are some important assumptions:

1.Outliers exist in the "Price_Charged" feature, but we are not treating them as outliers due to the unavailability of trip duration details.
2.
3.Profit for each ride is calculated while keeping other factors constant. We calculate profit using only the "Price_Charged" and "Cost_of_Trip" features.
4.
5.The "Users" feature in the city dataset is considered to represent the number of cab users in the city. We assume that this includes users of other cab services as well, such as Yellow and Pink cabs.

Distribution of values across both Firms
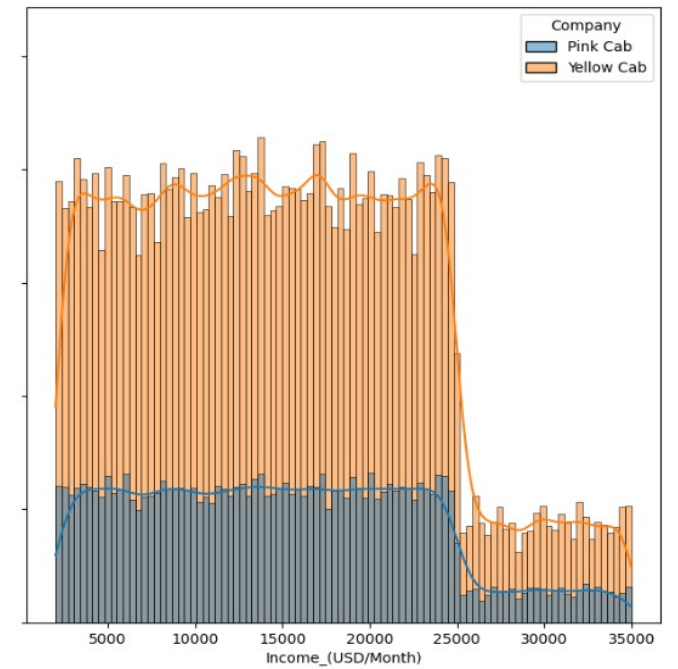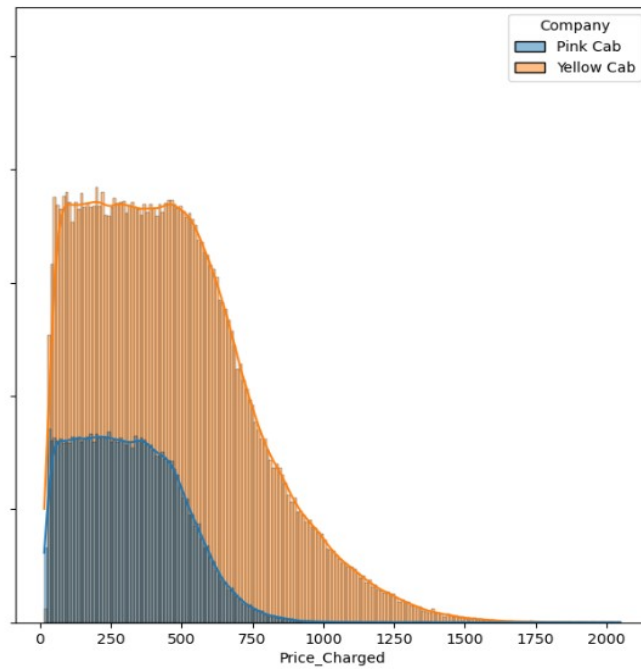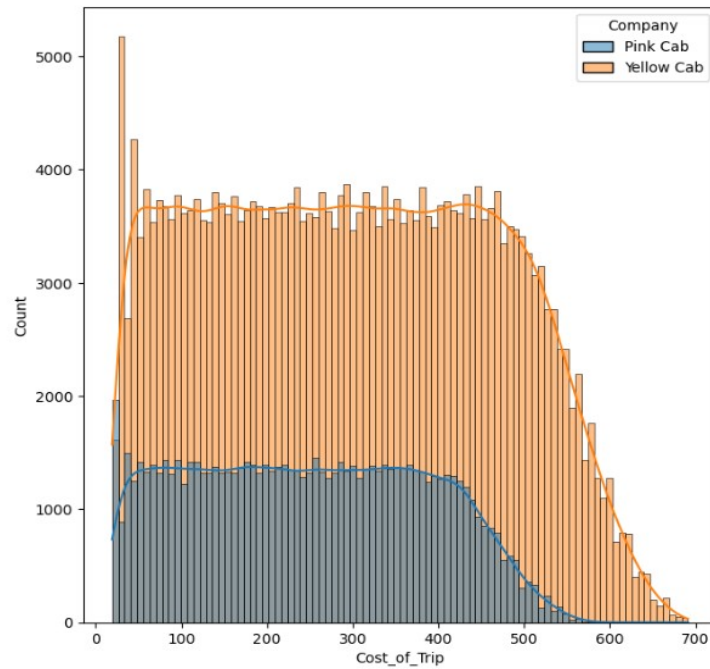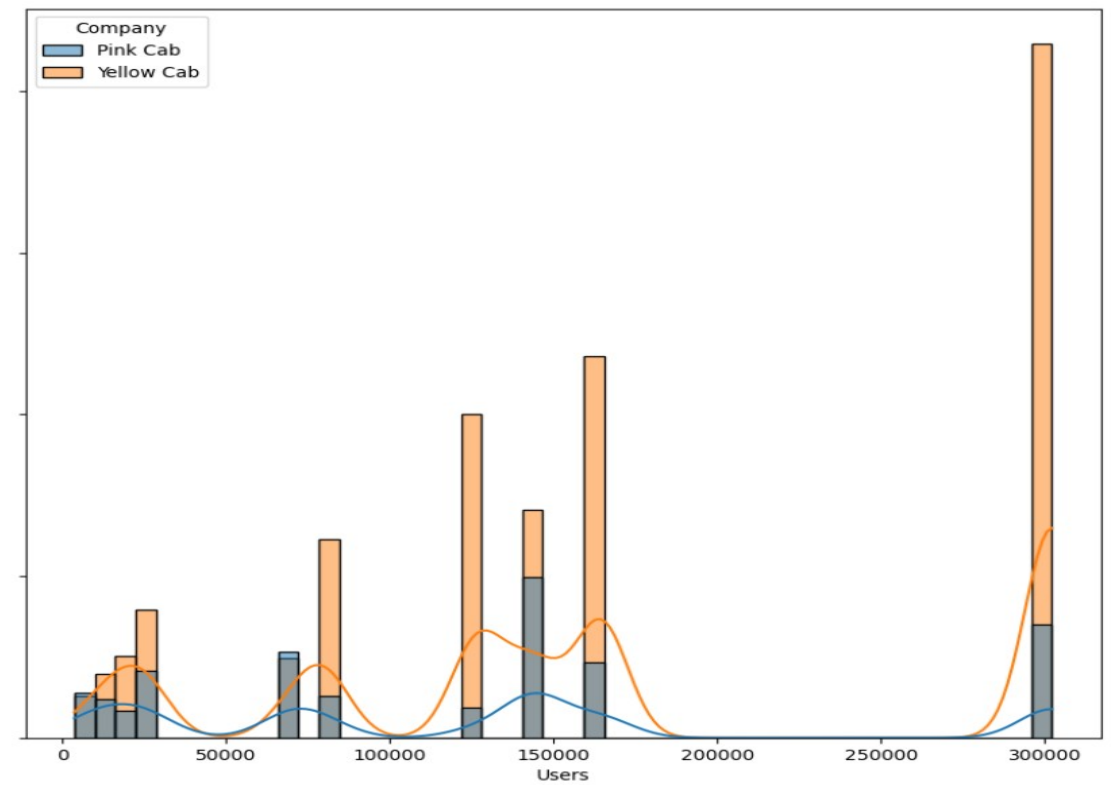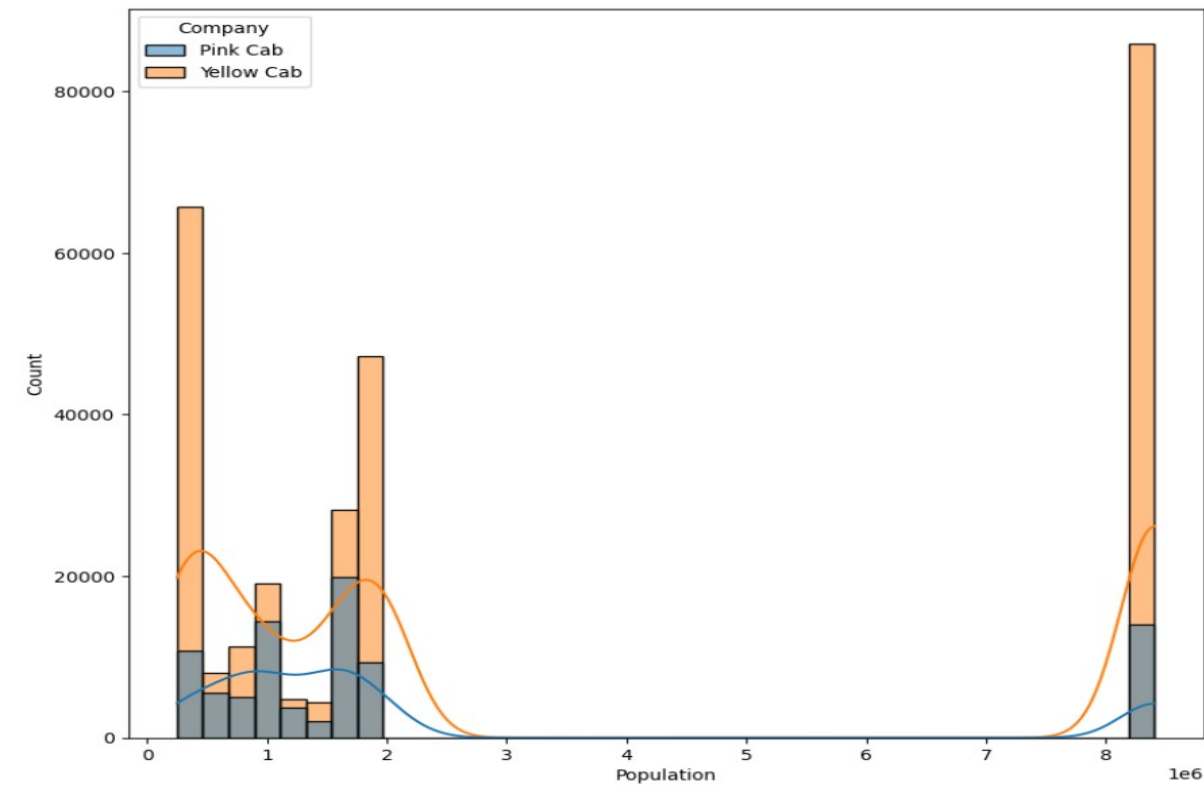
-

Distributions of Variables

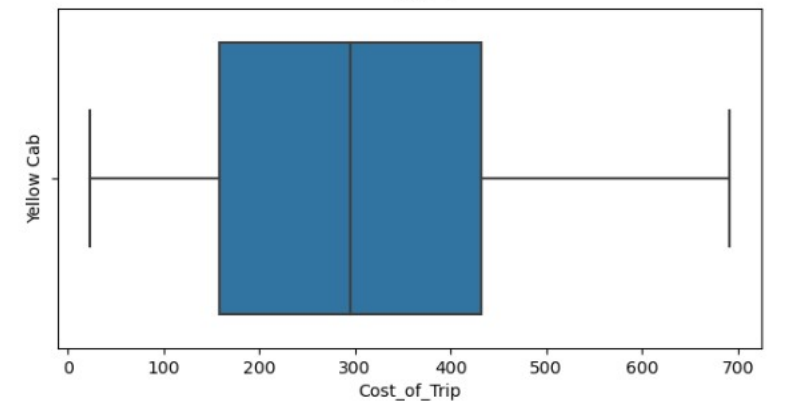# Distribution of Variables across both Firms
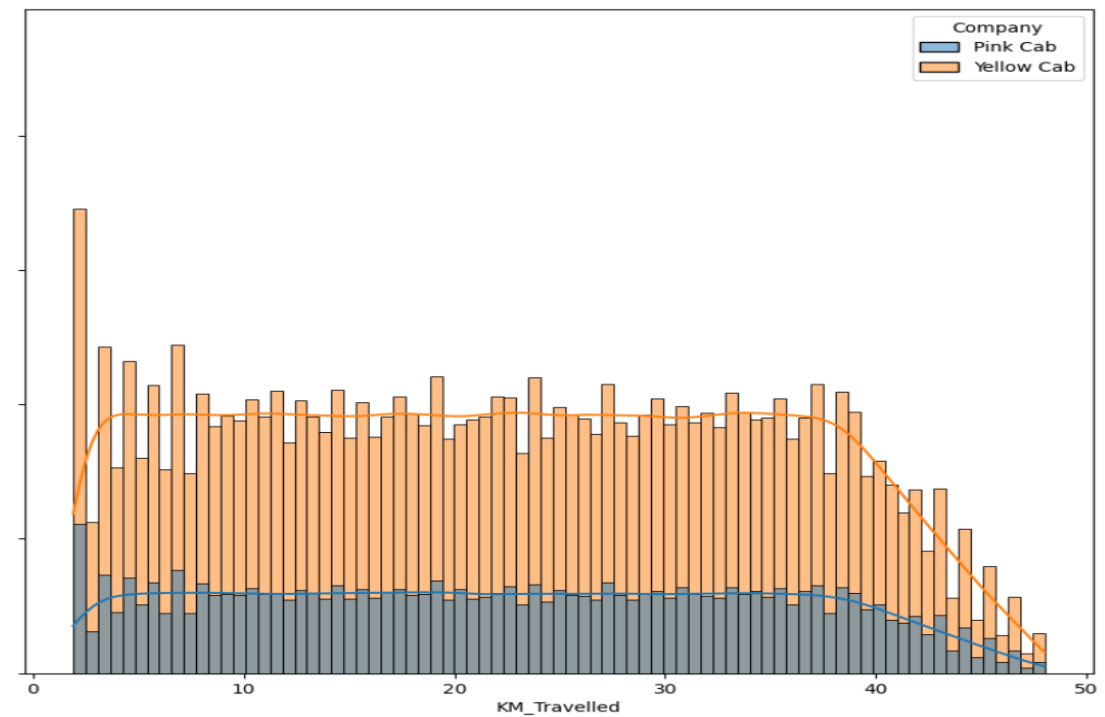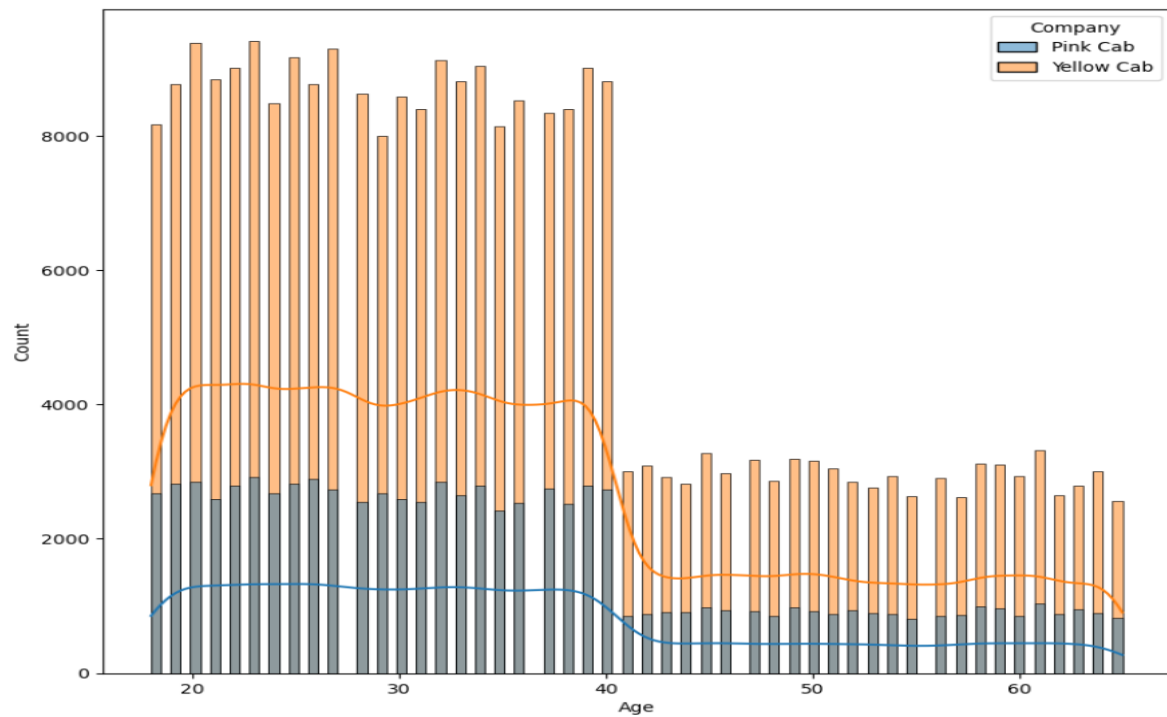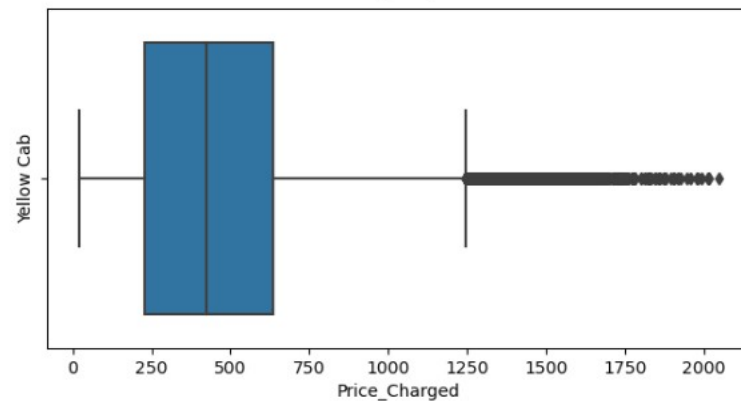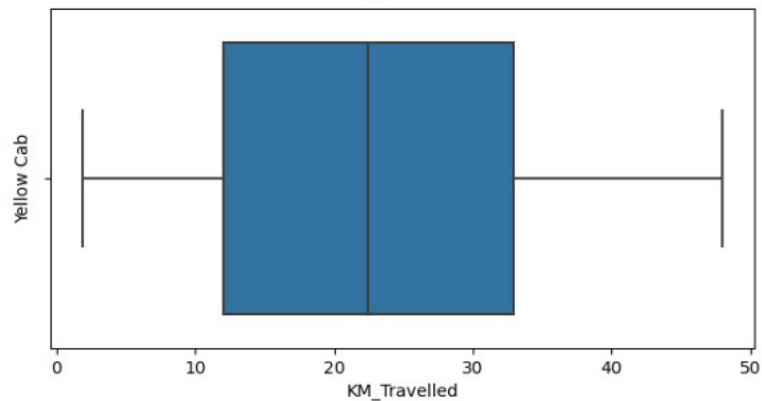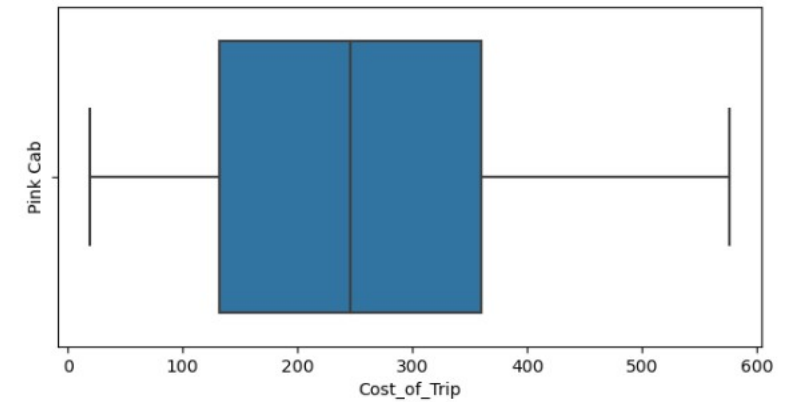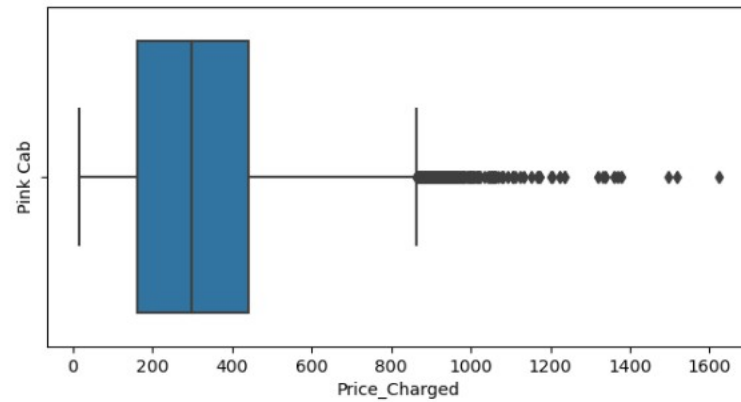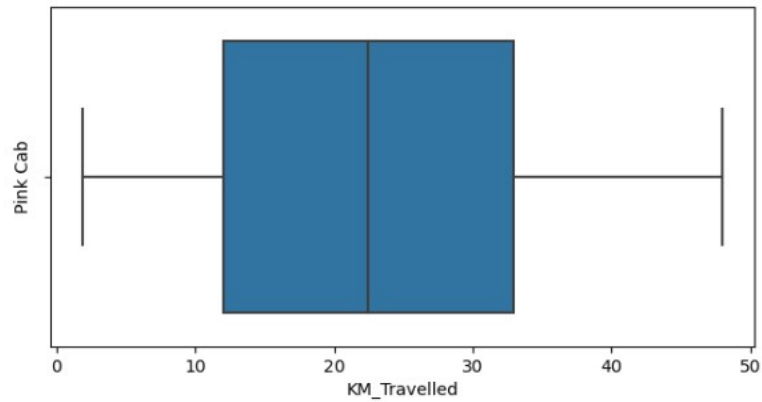

Distributions of Variables

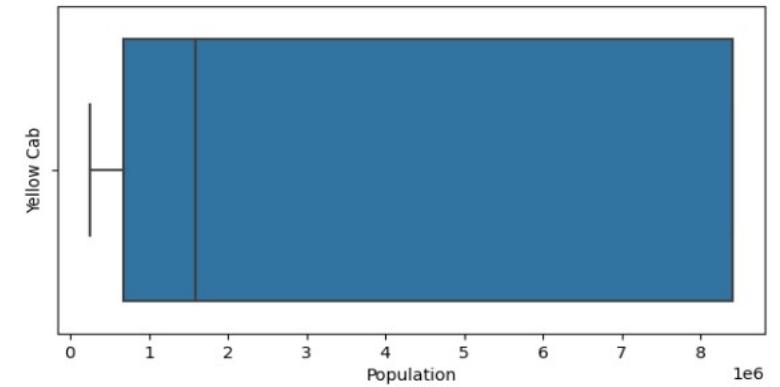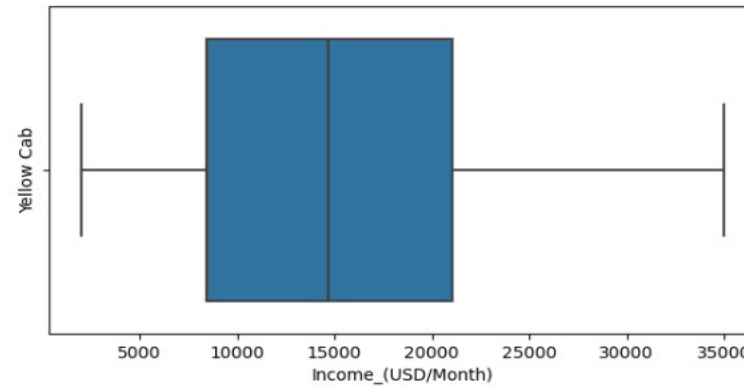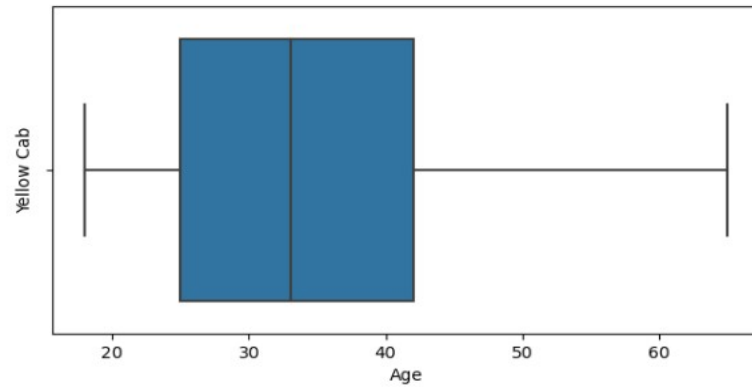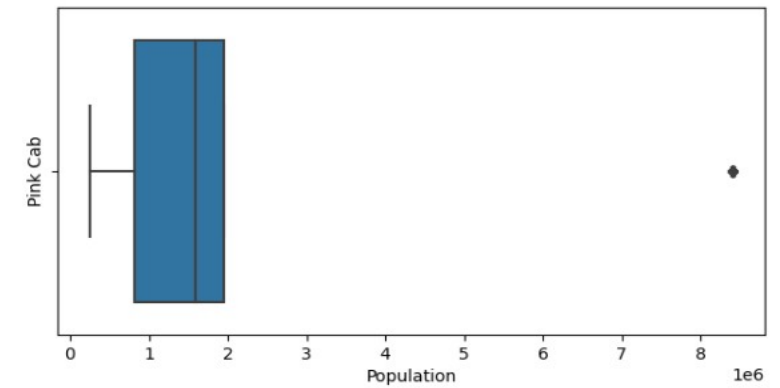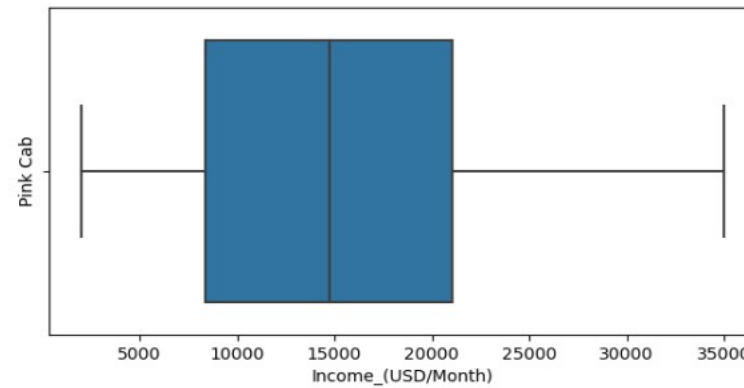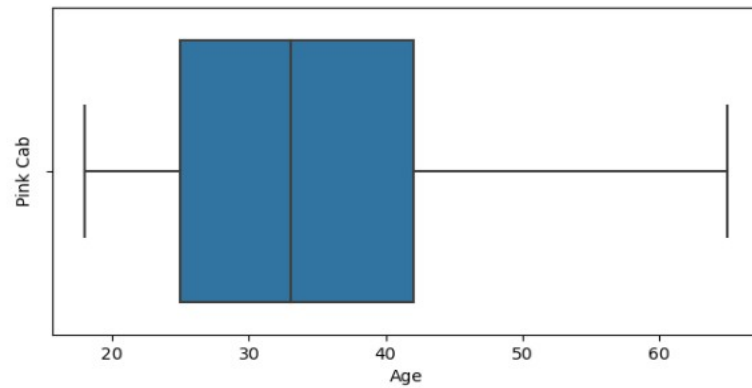# Distribution of Variables across both Firms

Boxplot Distribution of Variables

Boxplot Distributions of the Variables
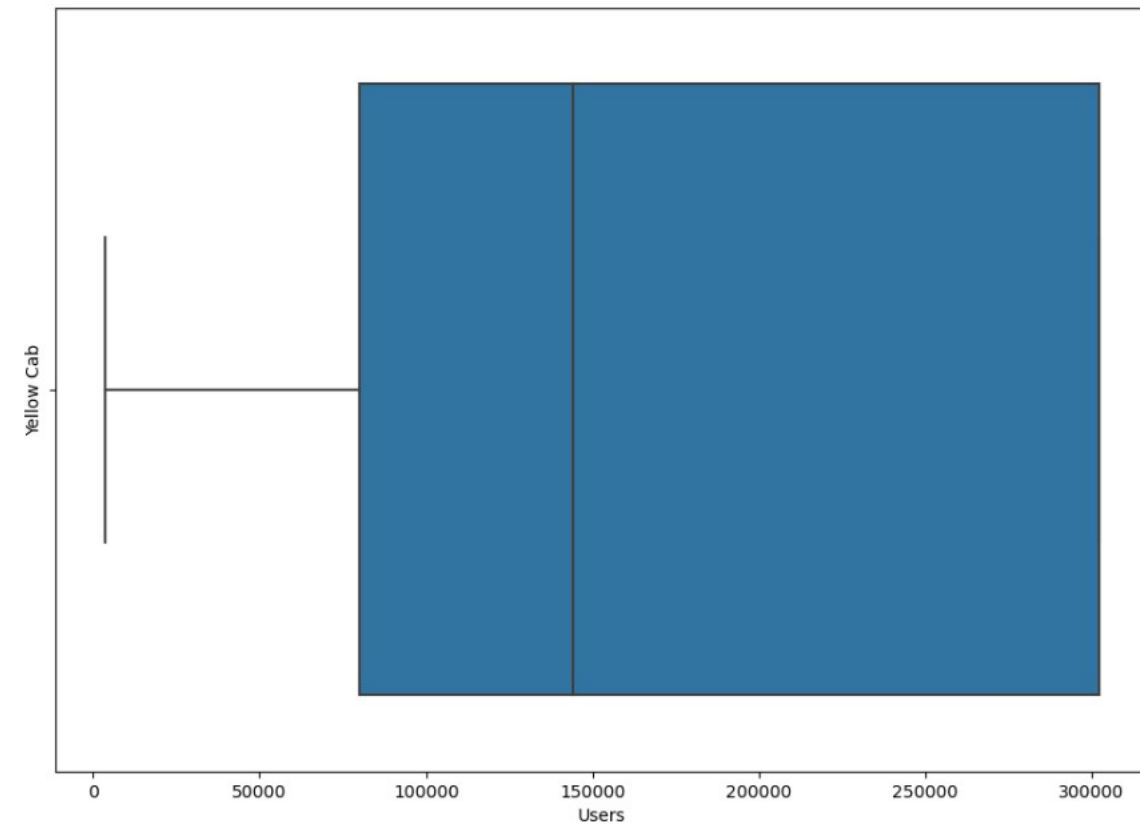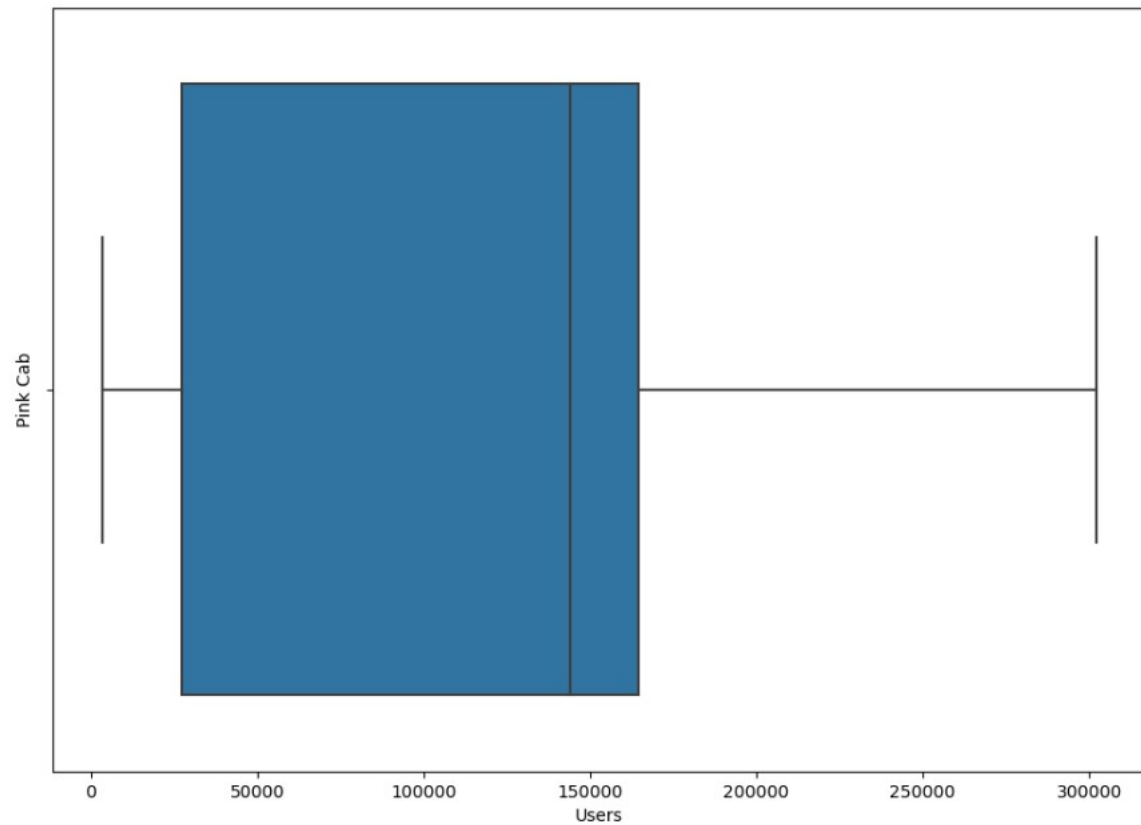
Boxplot Distribution of Variables

Boxplot Distributions of the Variables

Boxplot Distribution of User For Both Firms

Correlation of Master Data features by heatmap for Pink Cab Firm
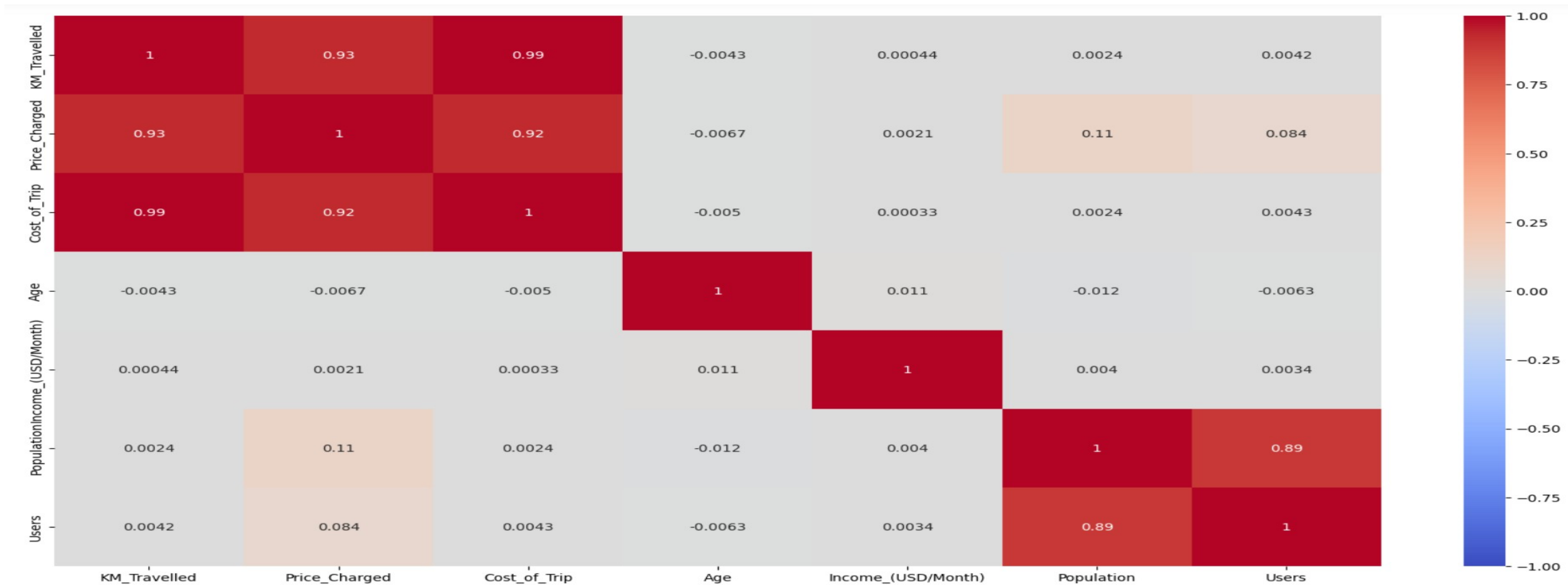
Correlation of Master Data features by heatmap for Yellow Cab Firm

Scatter Plots of Master Data Features For Both Firms
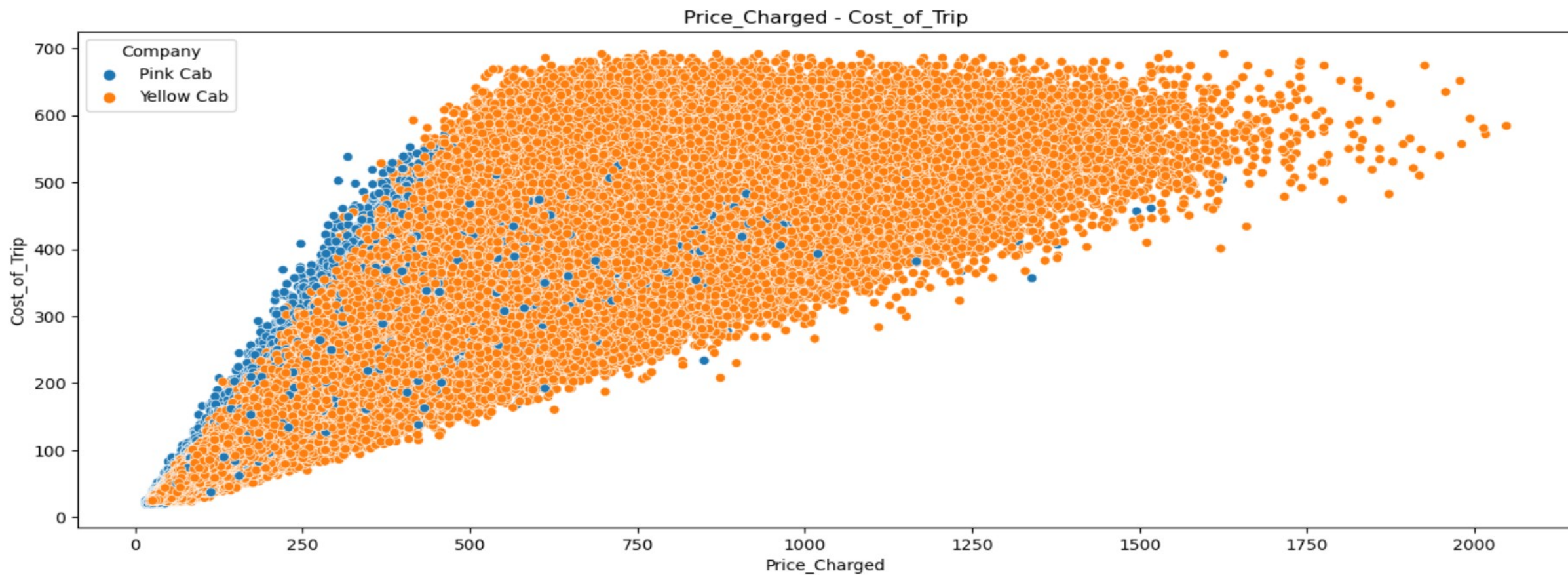
User Overview For Both Firms

Pink & Yellow Cab Firm Total Users Overview
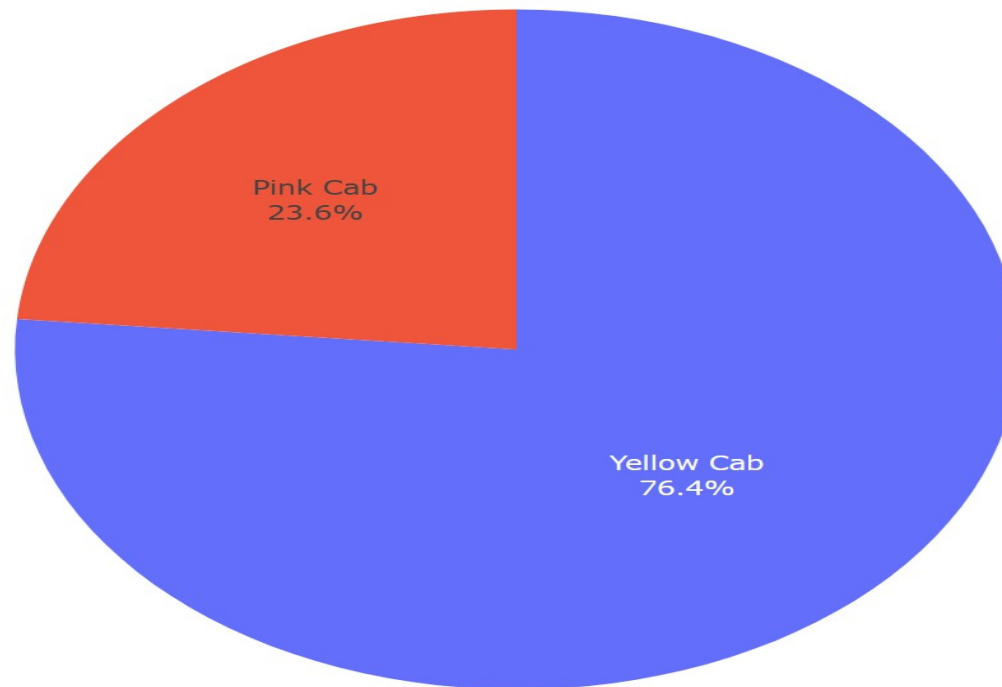
User Distribution By Cities For Both Firms

Pink & Yellow Cab Firm Users Distribution Over City
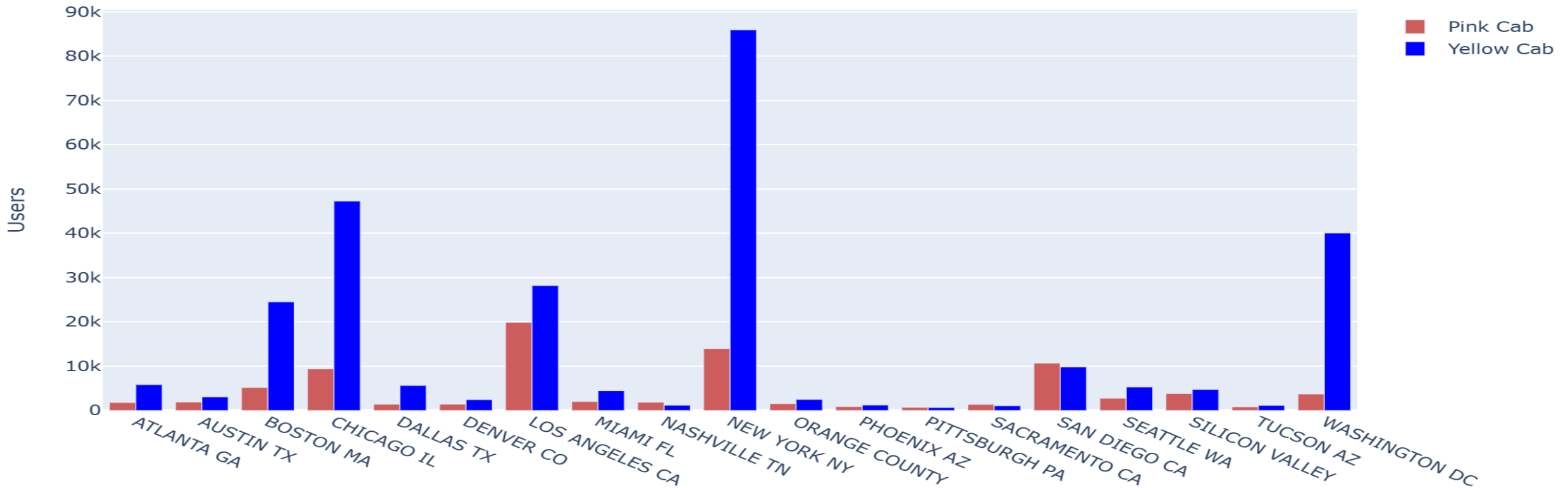
User Overview For Both Firms

Total Users Overview by Cities
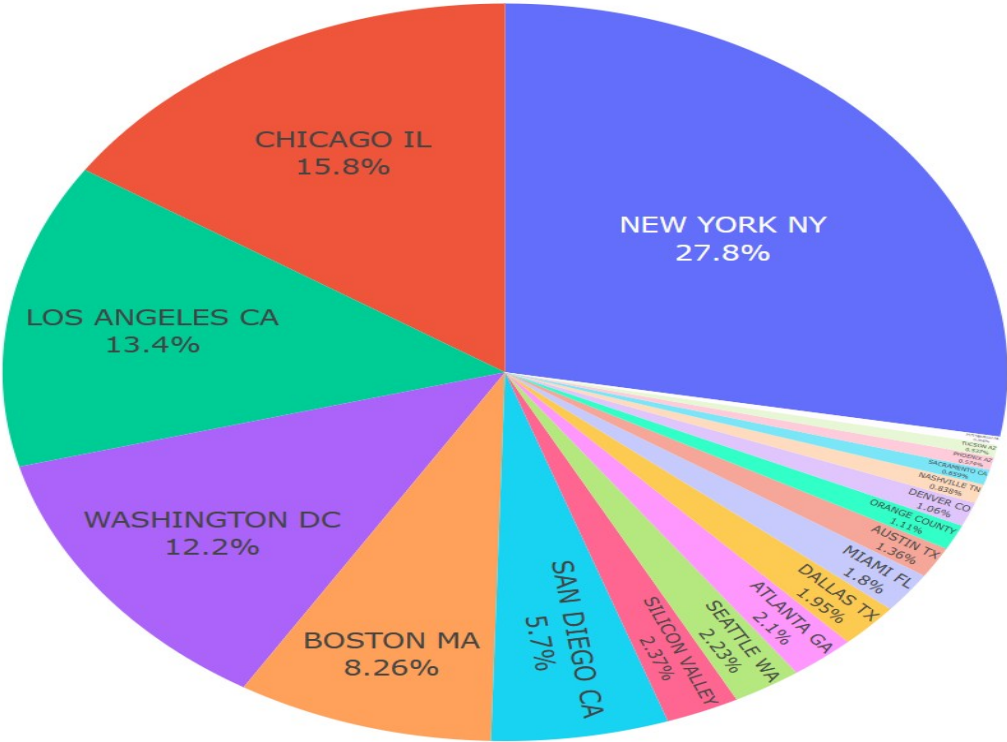
Profit Share By Cities For Both Firms

Total Market Profit Share by Cities
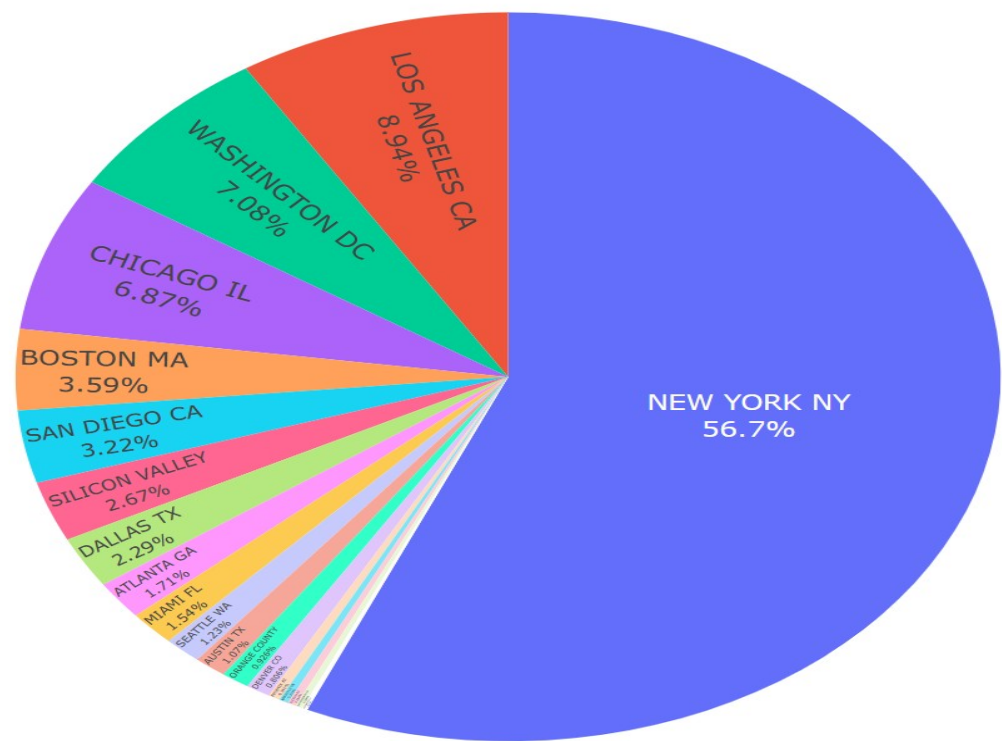
Profit Share For Both Firms

Total Market Profit Share by Cab Firms

Pink Cab
10.8%

Yellow Cab
89.2%

index=Yellow Cab
0=44,020,373.17
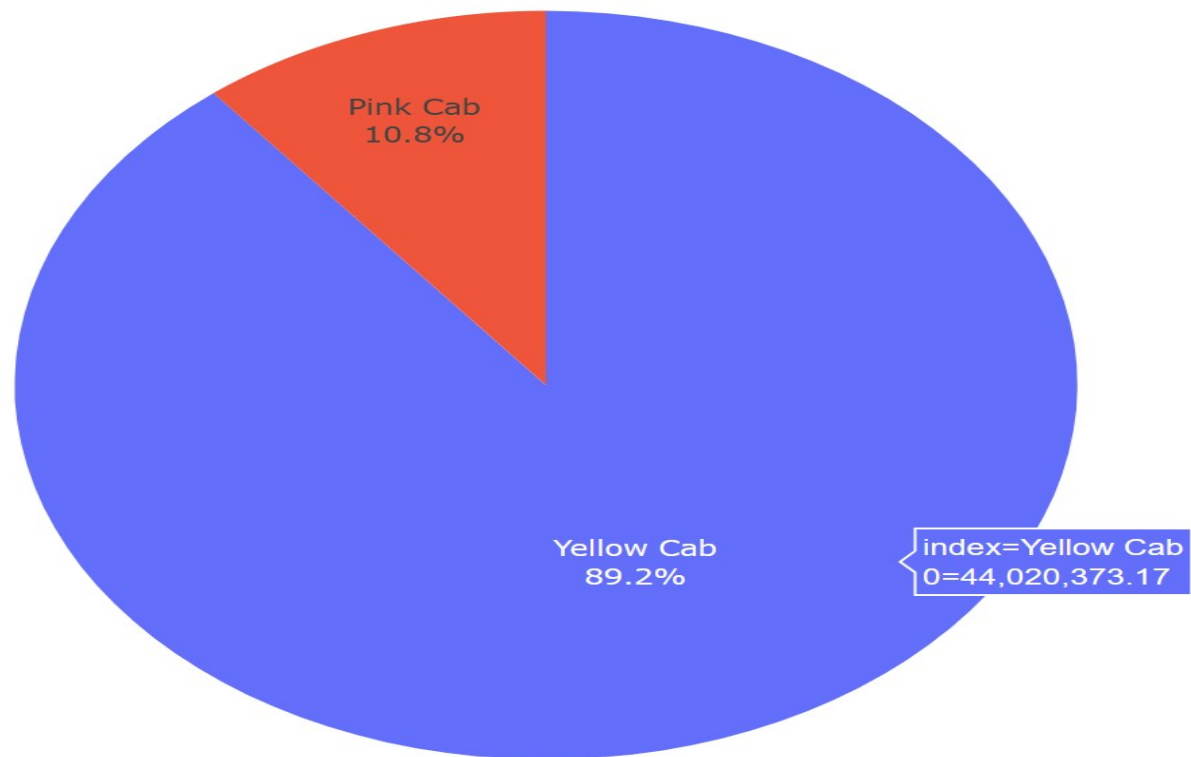
Gender Distribution For Both Firms

Pink & Yellow Cab Firm Users Distribution Over Gender
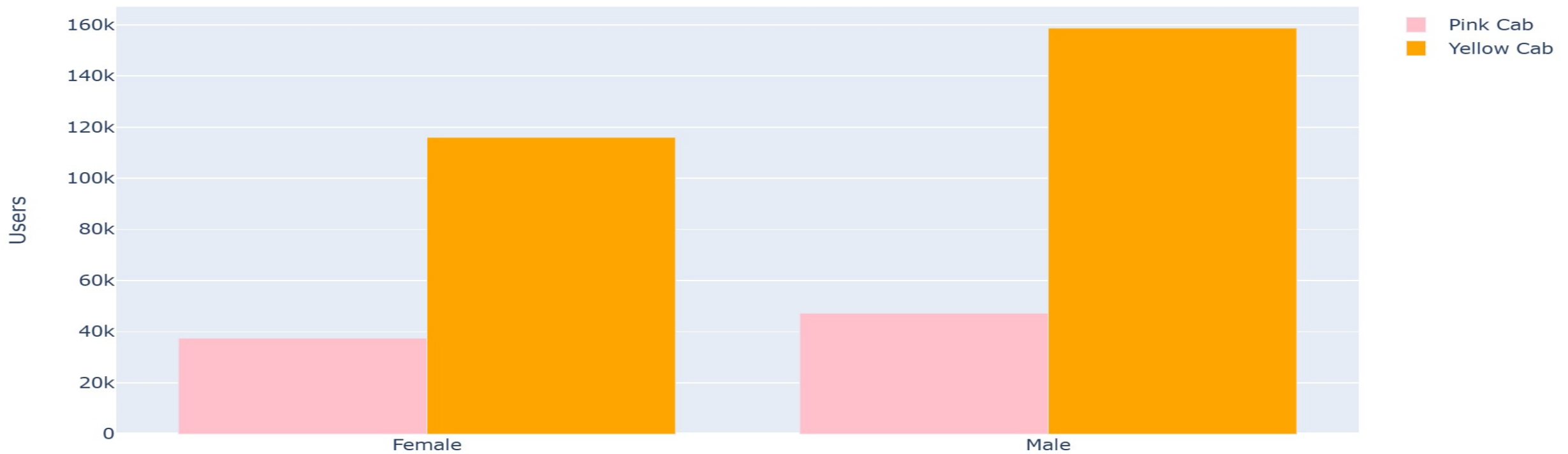
Gender Distribution For Both Firms

Profit forecasting for 2019

Recommendations

After a comprehensive evaluation of both cab companies, we have determined that Yellow Cab outperforms Pink Cab in several key aspects:

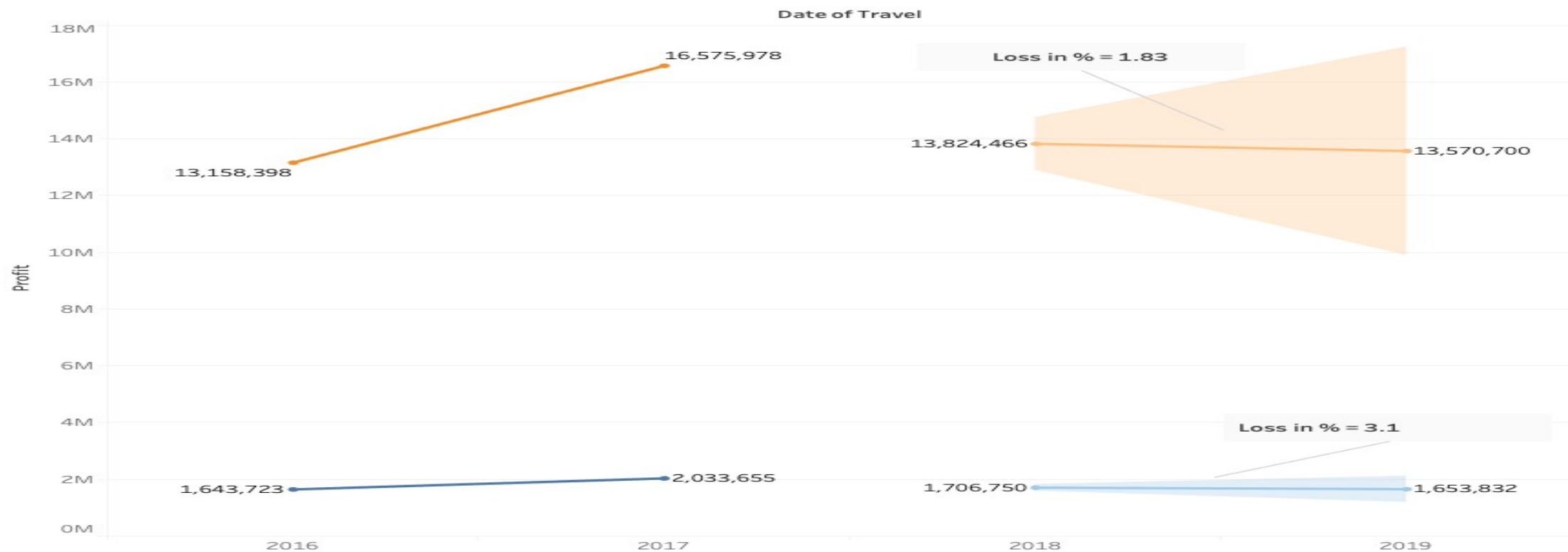1.**Customer Reach:** Yellow Cab has a higher customer reach in 25 cities, whereas Pink Cab has a higher reach in only 4 cities. Notably, Yellow Cab also excels in extending its services to other cab users beyond its own customer base.

2.**Customer Retention:** Our analysis considers two retention segments—customers with at least 5 rides and those with at least 10 rides with the same cab company. In both segments, Yellow Cab exhibits significantly better retention rates compared to Pink Cab.

3.**Age-wise Reach:** Yellow Cab caters to customers of all age groups, with a noteworthy presence in the 60+ age group, matching its popularity among the 18-25 age group.

4.**Average Profit per KM:** Yellow Cab boasts an average profit per kilometer nearly three times that of Pink Cab.

5.**Income-wise Reach:** Both cab companies are popular among high and medium-income classes. However, Yellow Cab surpasses Pink Cab in offering its services to all income class groups, including low, medium, and high-income individuals.

6.**Ride Count and Profit Forecasting:** Both companies are currently experiencing losses in both profit and the number of rides. However, Yellow Cab's forecasted profit loss is approximately 1.83%, while Pink Cab faces a higher loss at 3.1%. This is despite Pink Cab having a lower forecasted loss in the number of rides compared to Yellow Cab.

Based on these findings, we highly recommend considering an investment in Yellow Cab as the more promising option.

THANK YOU