Exploratory Data Analysis
G2M Cab Industry

27 Dec 2023

Agenda

Executive Summary

XYZ, a private firm in the US, is considering an investment in the cab industry and seeks insights to aid their decision-making process. The analysis encompassed multiple datasets spanning the period from 31/01/2016 to 31/12/2018, including Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv, City.csv and US_Holiday.csv.

Problem Statement

XYZ aims to grasp US cab market dynamics for informed investment. The challenge involves comprehensively analyzing diverse datasets to evaluate market dominance, pricing strategies, and seasonal demand patterns for investment decisions.

Approach

The analysis commences with reviewing, cleaning, transforming datasets, and conducting Exploratory Data Analysis (EDA). This encompasses hypothesis generation, outlier detection, and visualizations to comprehend the data. I have identified outliers in the 'Price_Charged' column, and these outliers are also affecting the 'Margin' profit column

EDA (Exploratory Data Analysis)

EDA dives into transactional behaviors, city-wise usage, pricing dynamics, and seasonal trends. Visualizations illuminate customer behaviors, market dominance, profit margins, and geographical preferences. Correlations reveal insights into population, infrastructure, and cab usage.
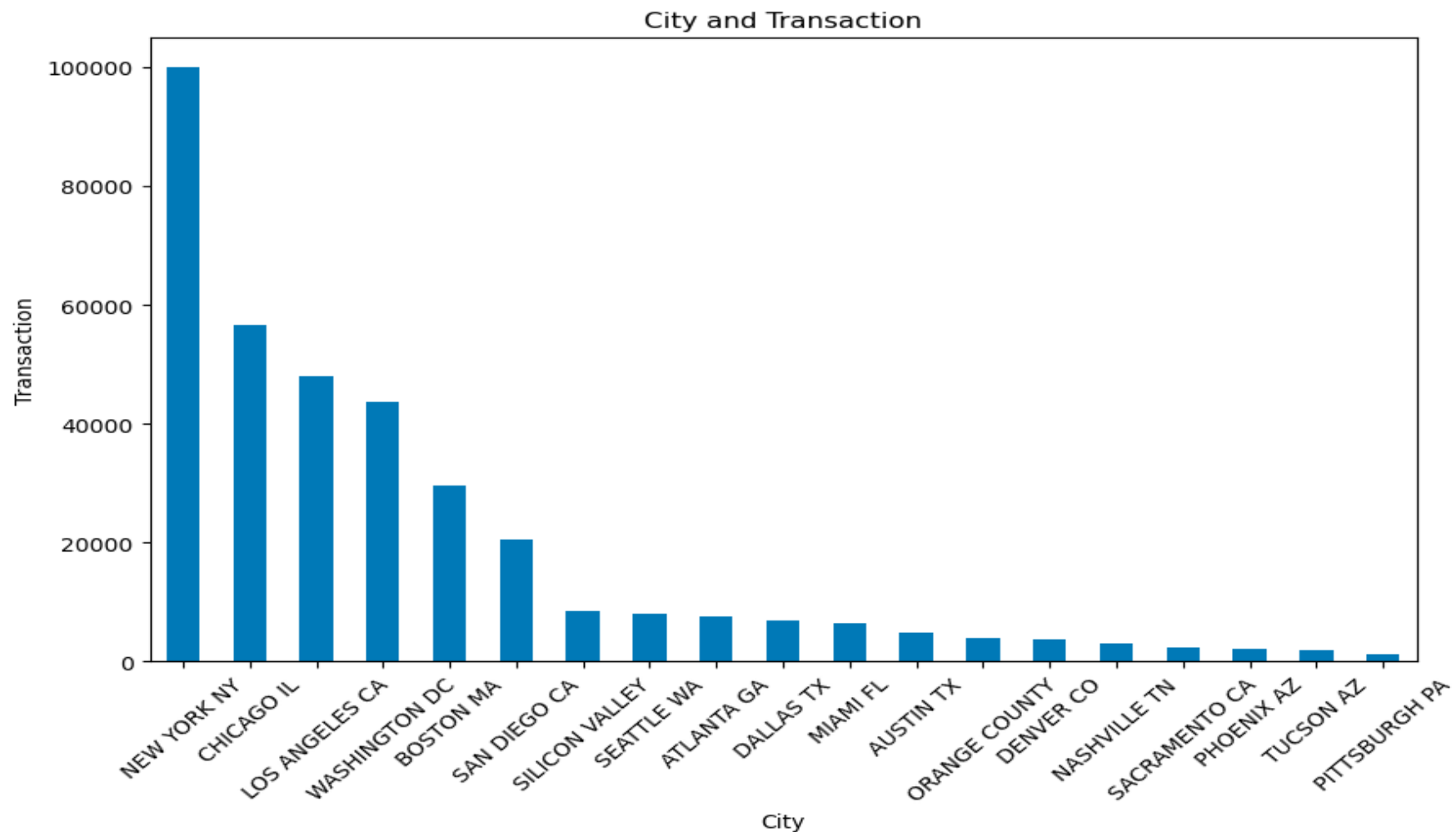
EDA
(Exploratory Data Analysis)
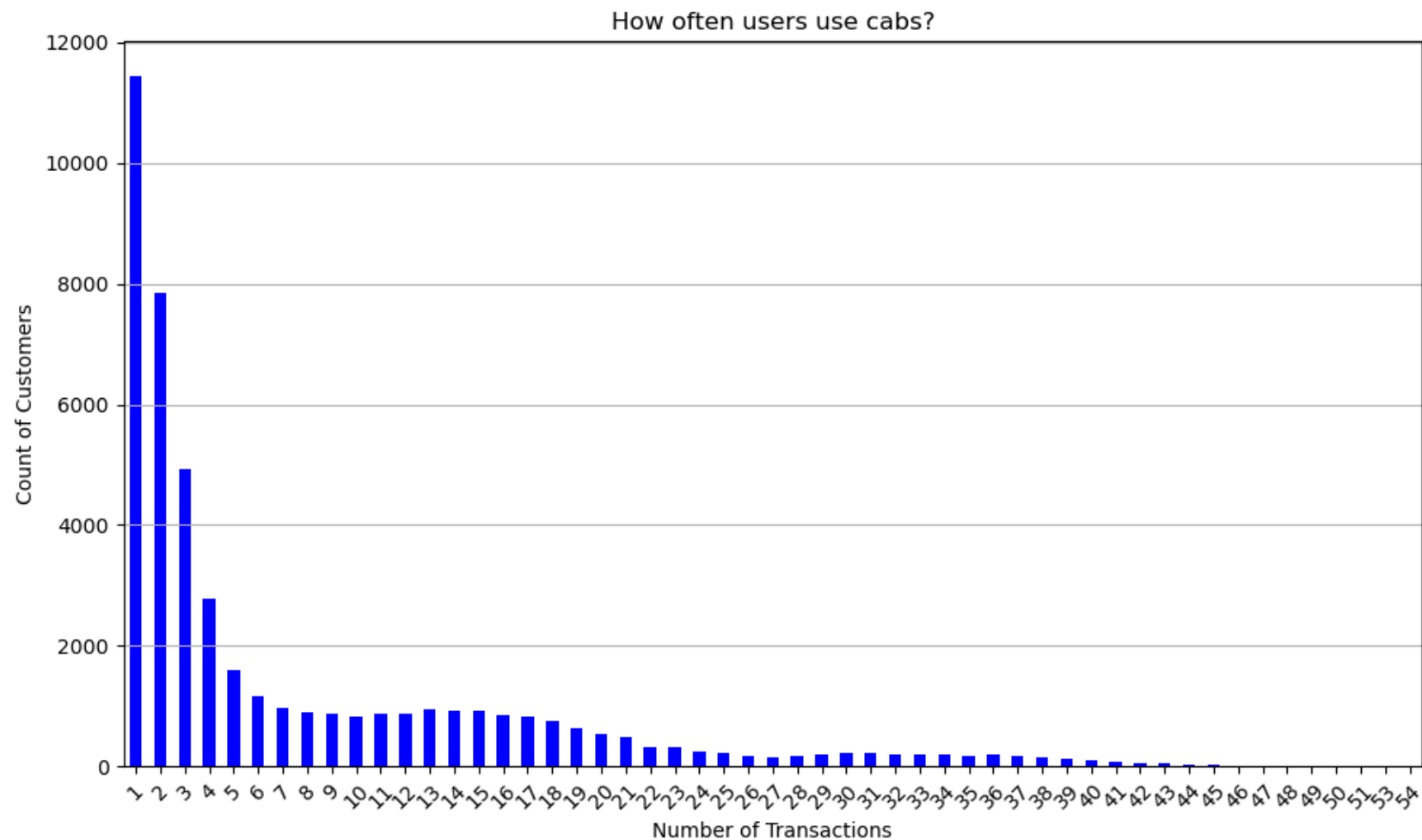
Most customers order the cab from 1 to 5 times

Transaction Counts by City and Customer
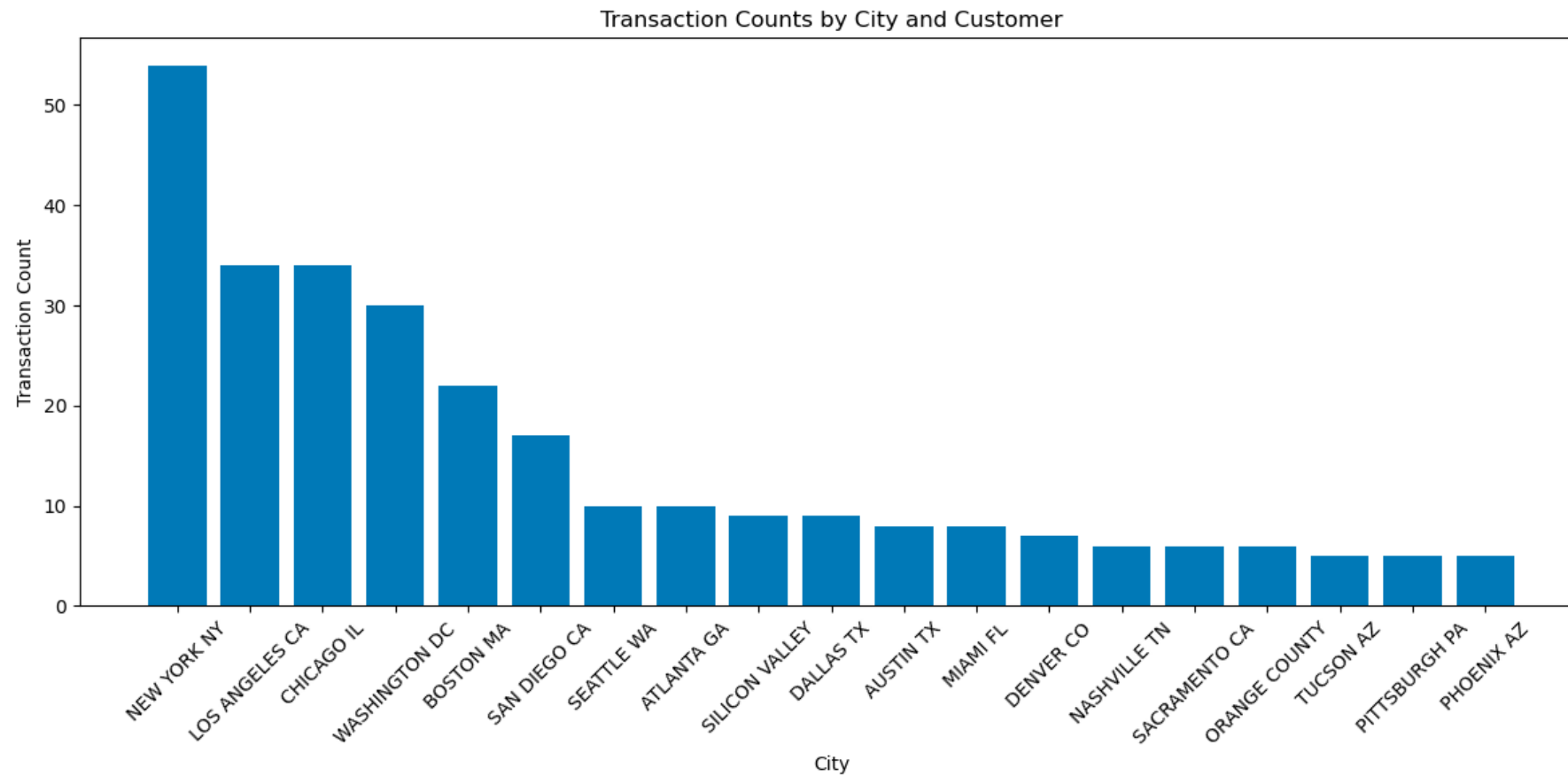
Many customers who live in New York use the cab service, which makes sense as New York is a big city.
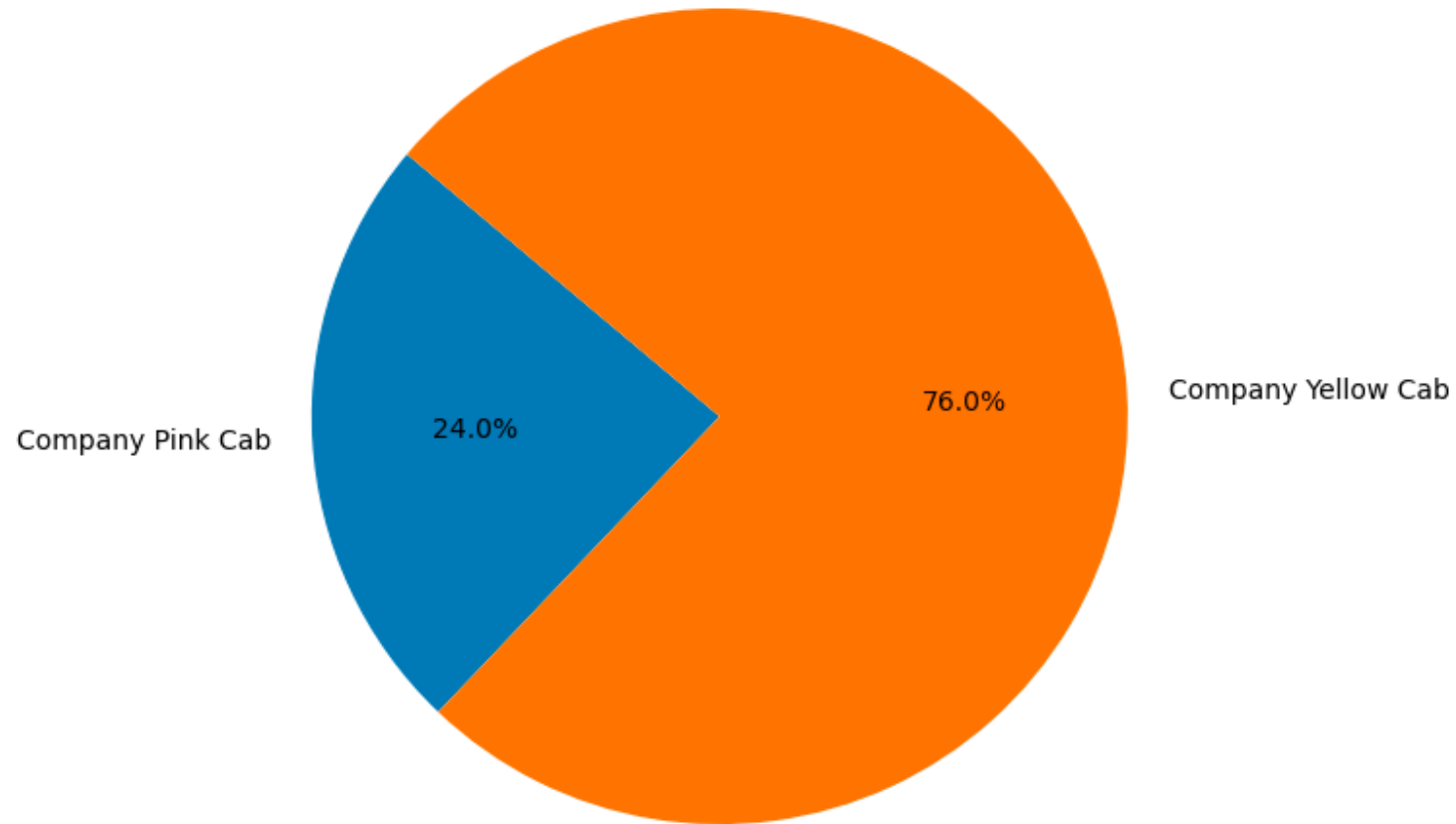
City and Transaction

The highest number of transactions occurs in New York. This analysis closely aligns with the customer distribution, as most transactions take place in New York

Friday is the busiest day, followed by Saturday and Sunday. During this period, people use cabs the most

Peak Hours Based on Day of the Week and Hour of the Day

Day of the Week

Hour of the Day
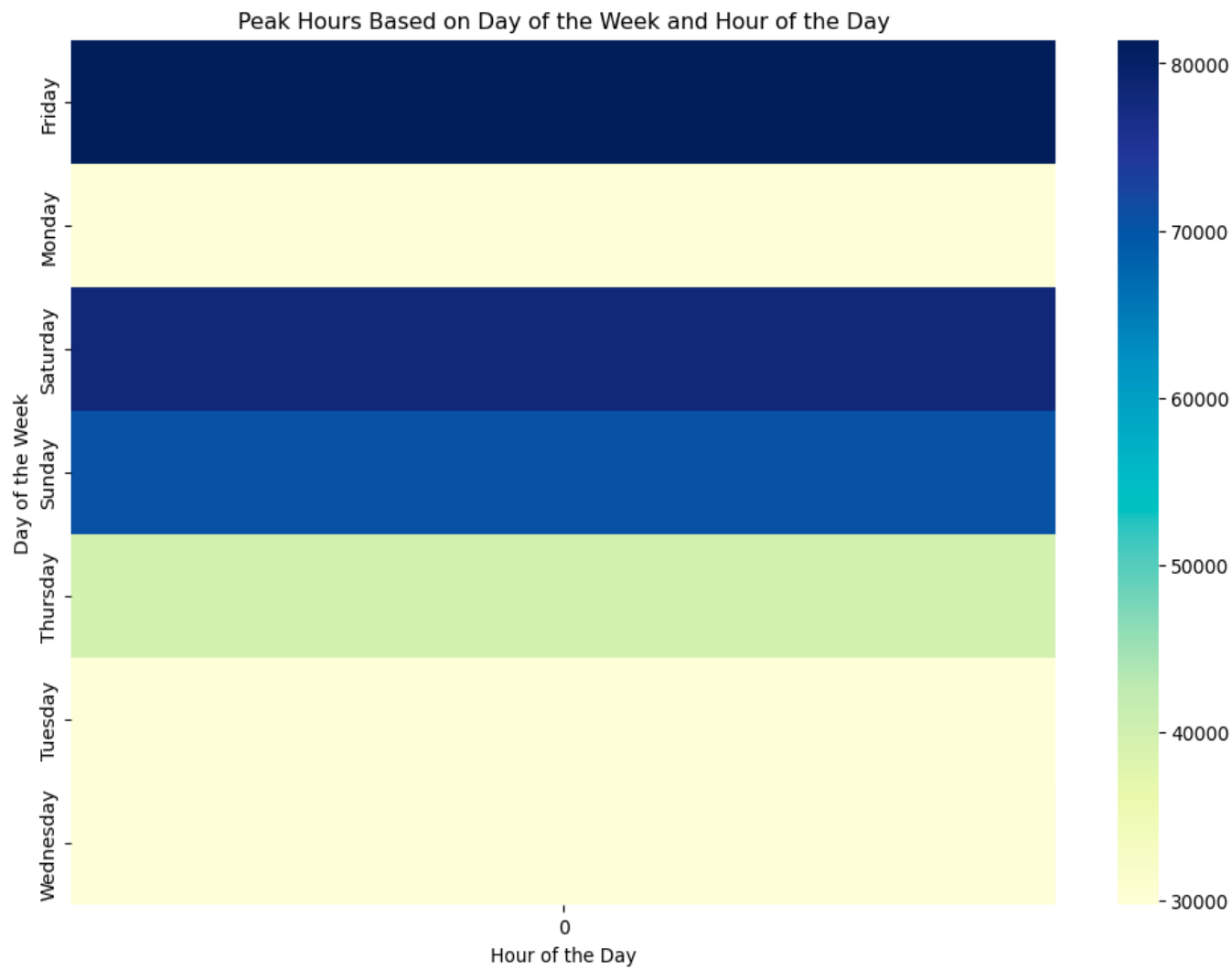
Market Share of Cab Companies

Company Pink Cab 24.0%

Company Yellow Cab 76.0%

Based on the transactions and customer, the yellow cab has a market share of 76%, whereas the pink cab only holds 26%. This demonstrates the dominance of the yellow cab in the cab industry.

As depicted in the bar chart, it's evident that customers tend to return more frequently using the yellow cab rather than the pink cab
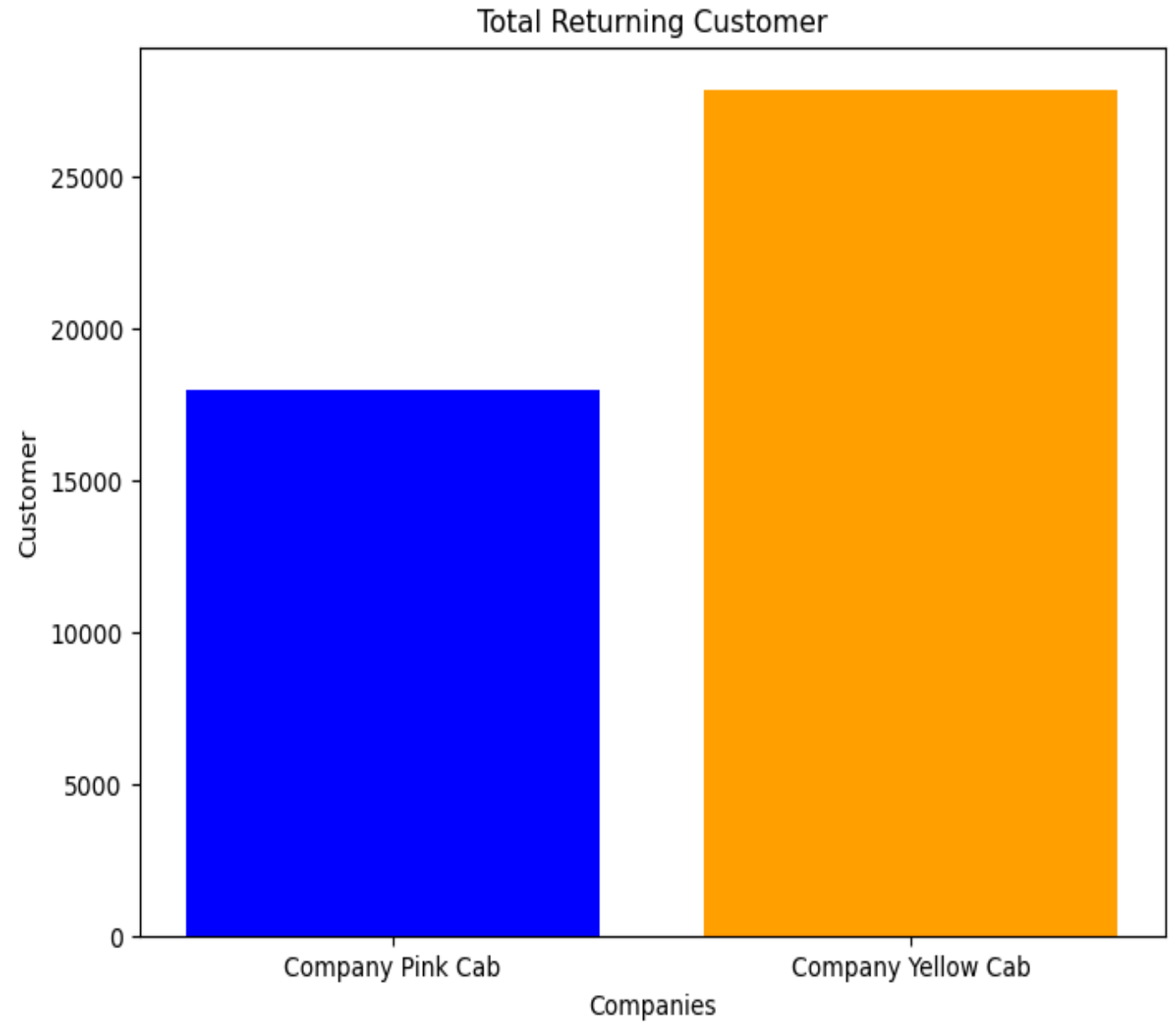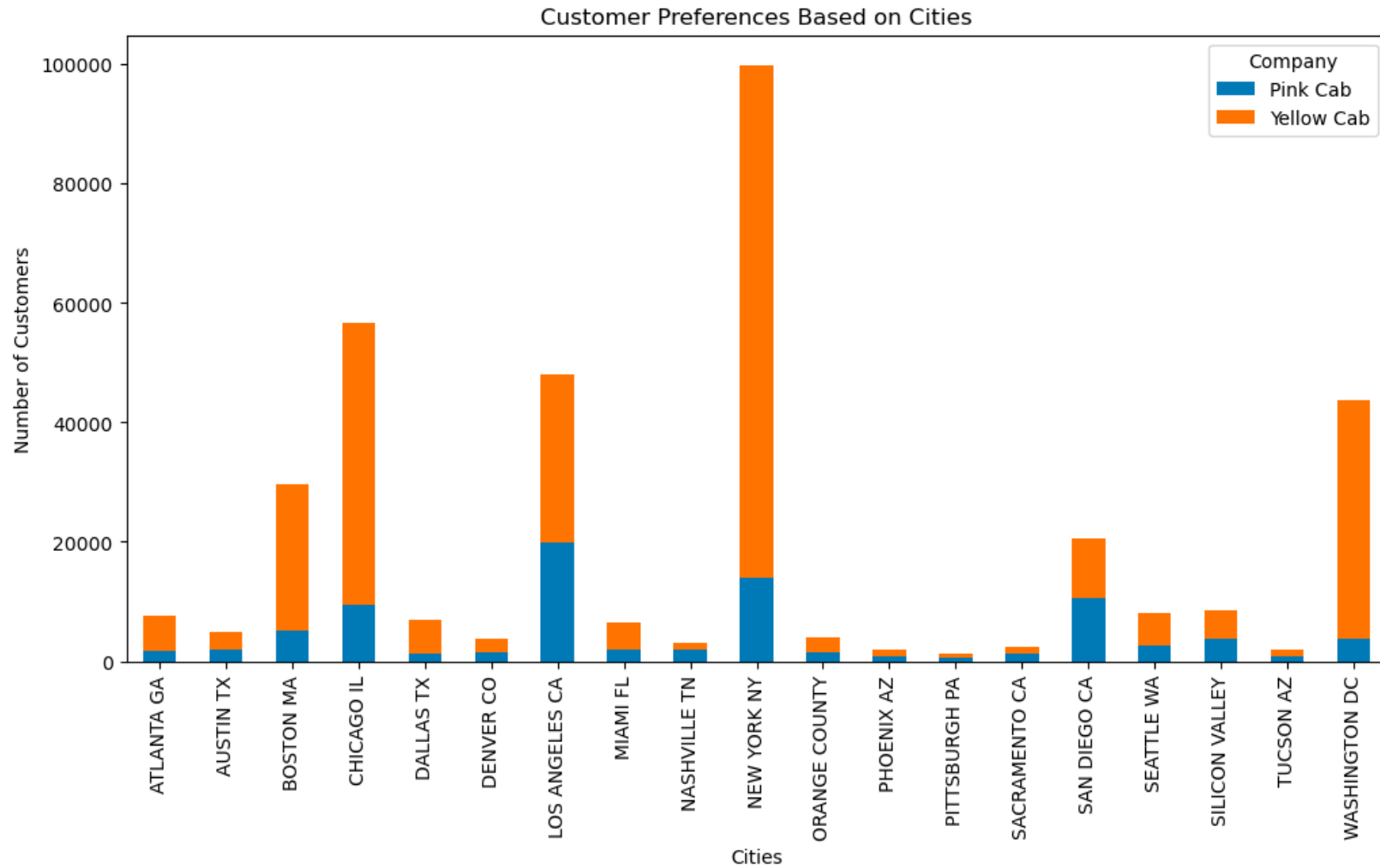
While there's not a significant difference, yellow cabs have more customers than pink cabs
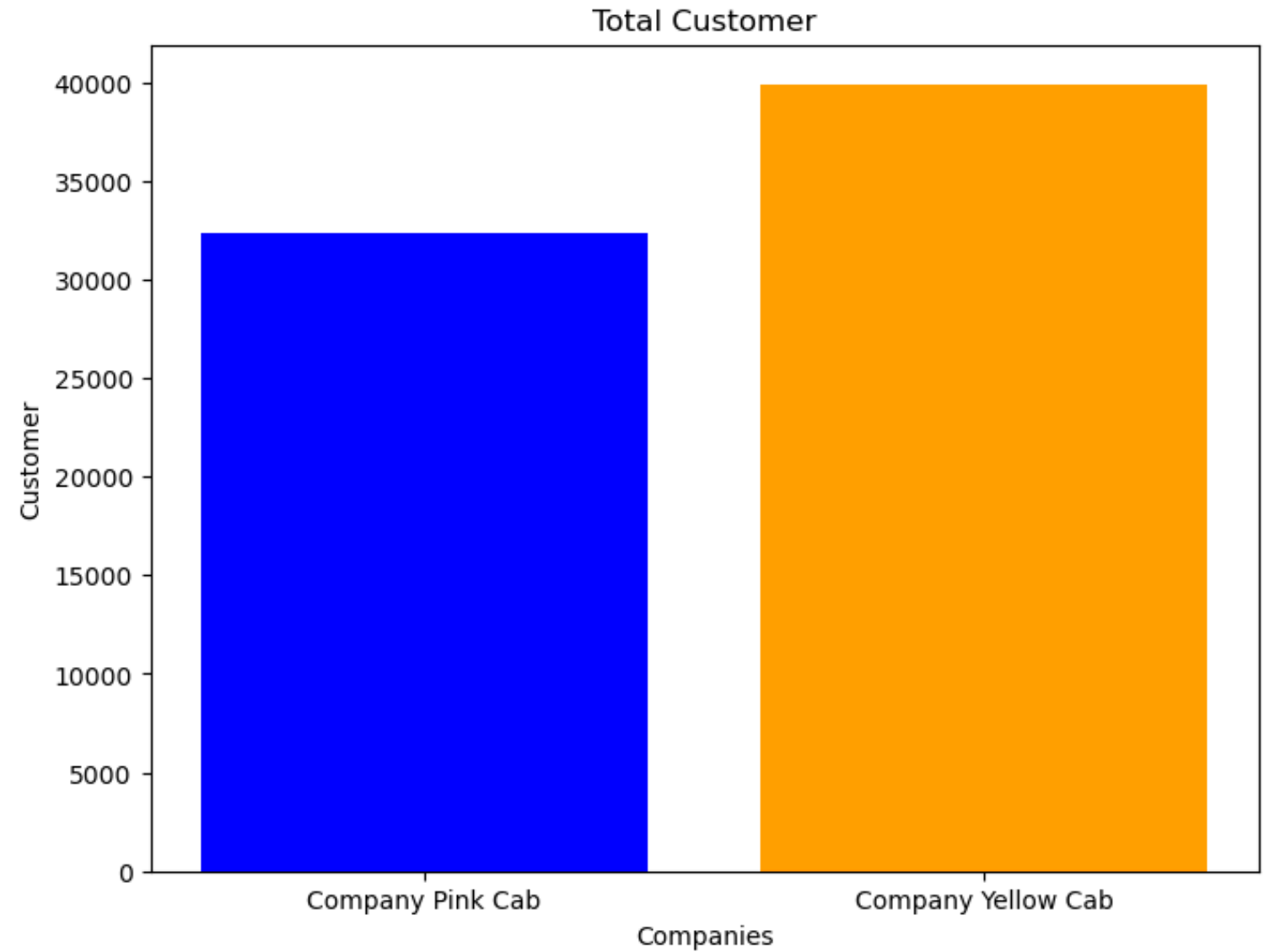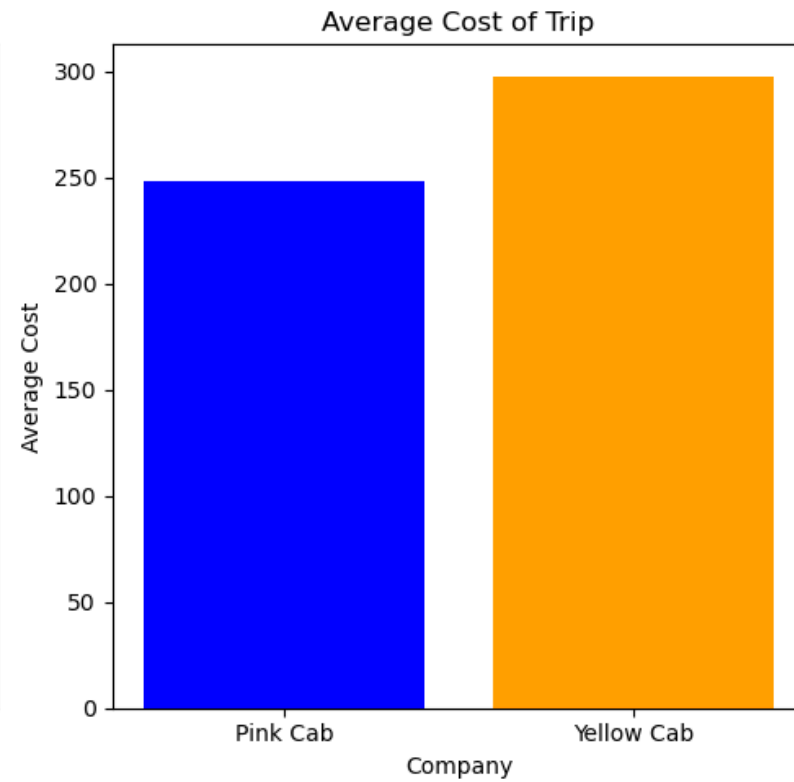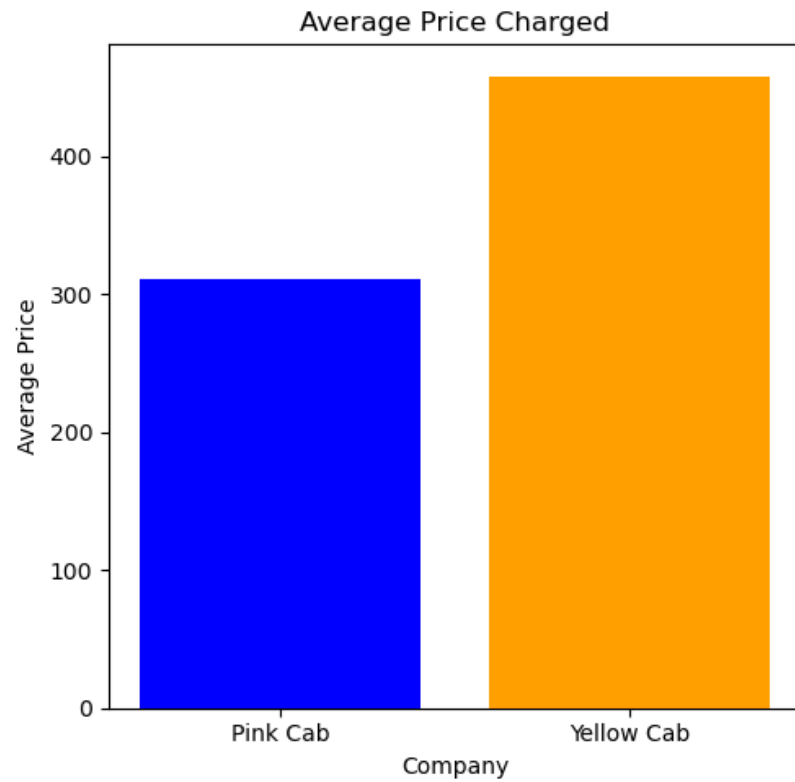
Yellow Cab has a higher rate of retaining customers compared to Pink Cab.

Customer Retention Rate of Cab Companies

Customer Preferences Based on Cities
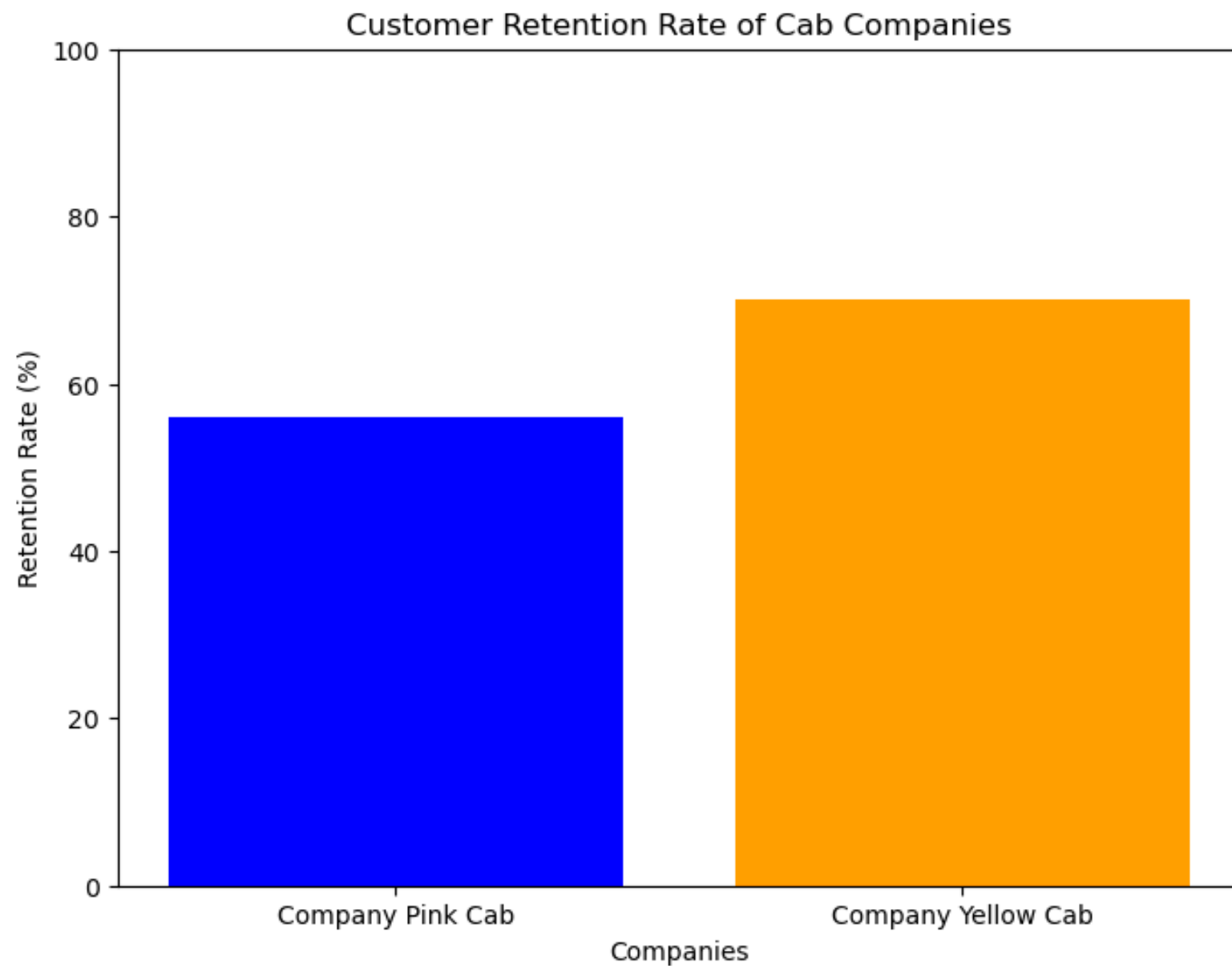
New York remains the city with the highest cab usage, followed by Chicago and Los Angeles. While more people use the pink cab in Nashville, in other cities, the majority tends to favor the yellow cab
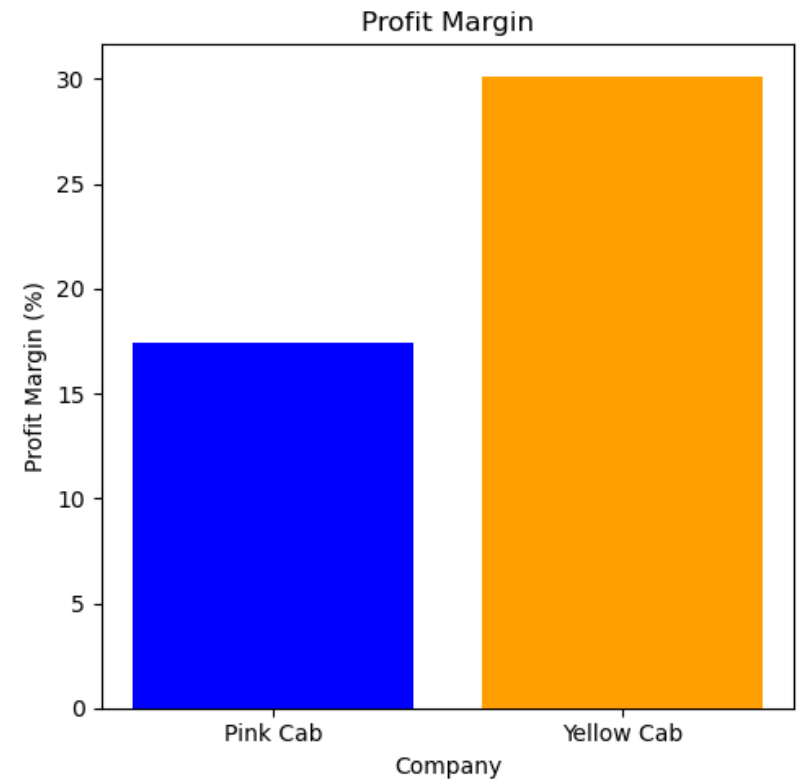
Yellow cabs have higher prices and trip costs compared to pink cabs. Despite the increased expenses, yellow cabs still maintain a higher profit margin than pink cabs

As we can see, the total profit margin for yellow cabs is four times higher than that of pink cabs.

Total Margin/profit by Company

Monthly Profit Comparison Between Companies
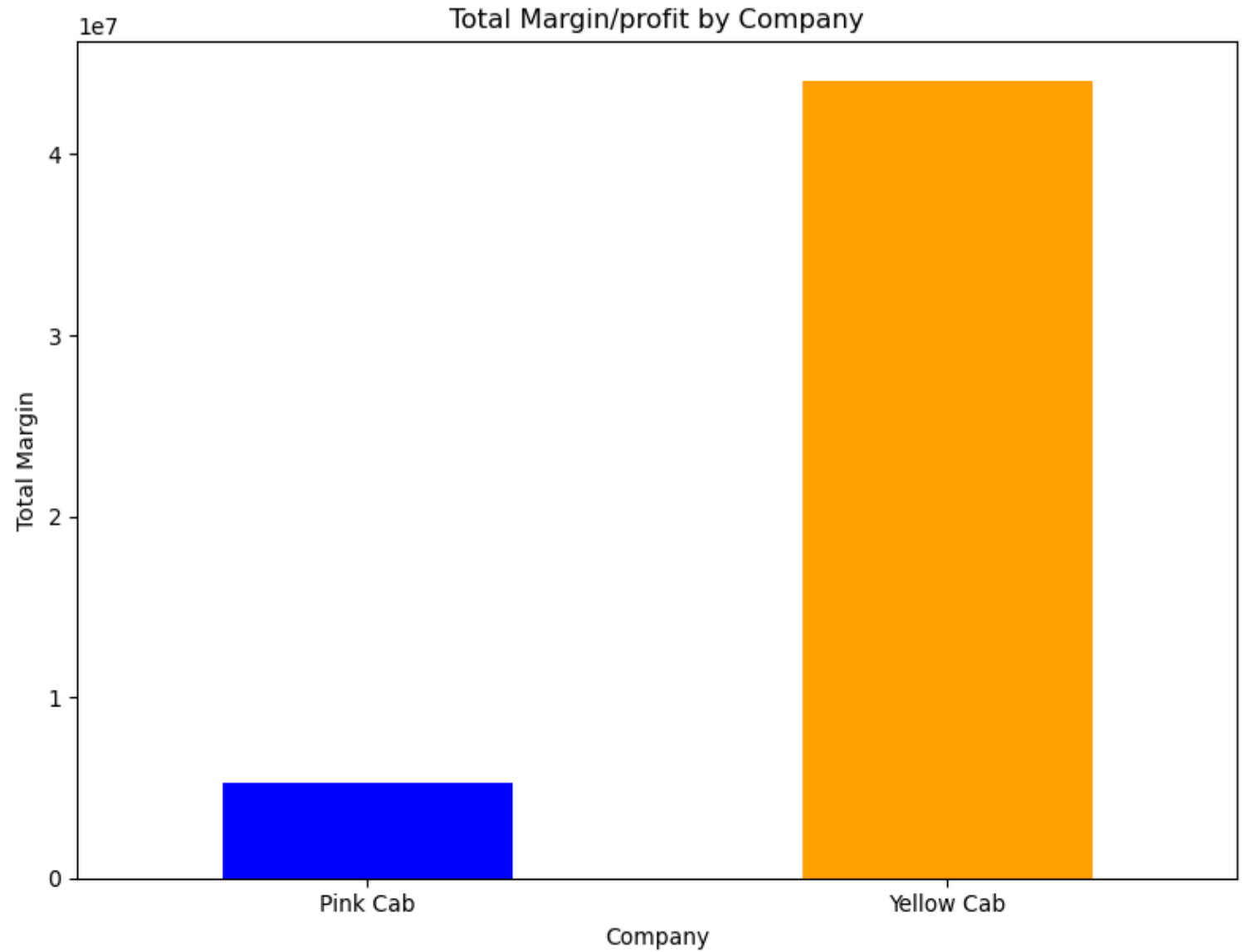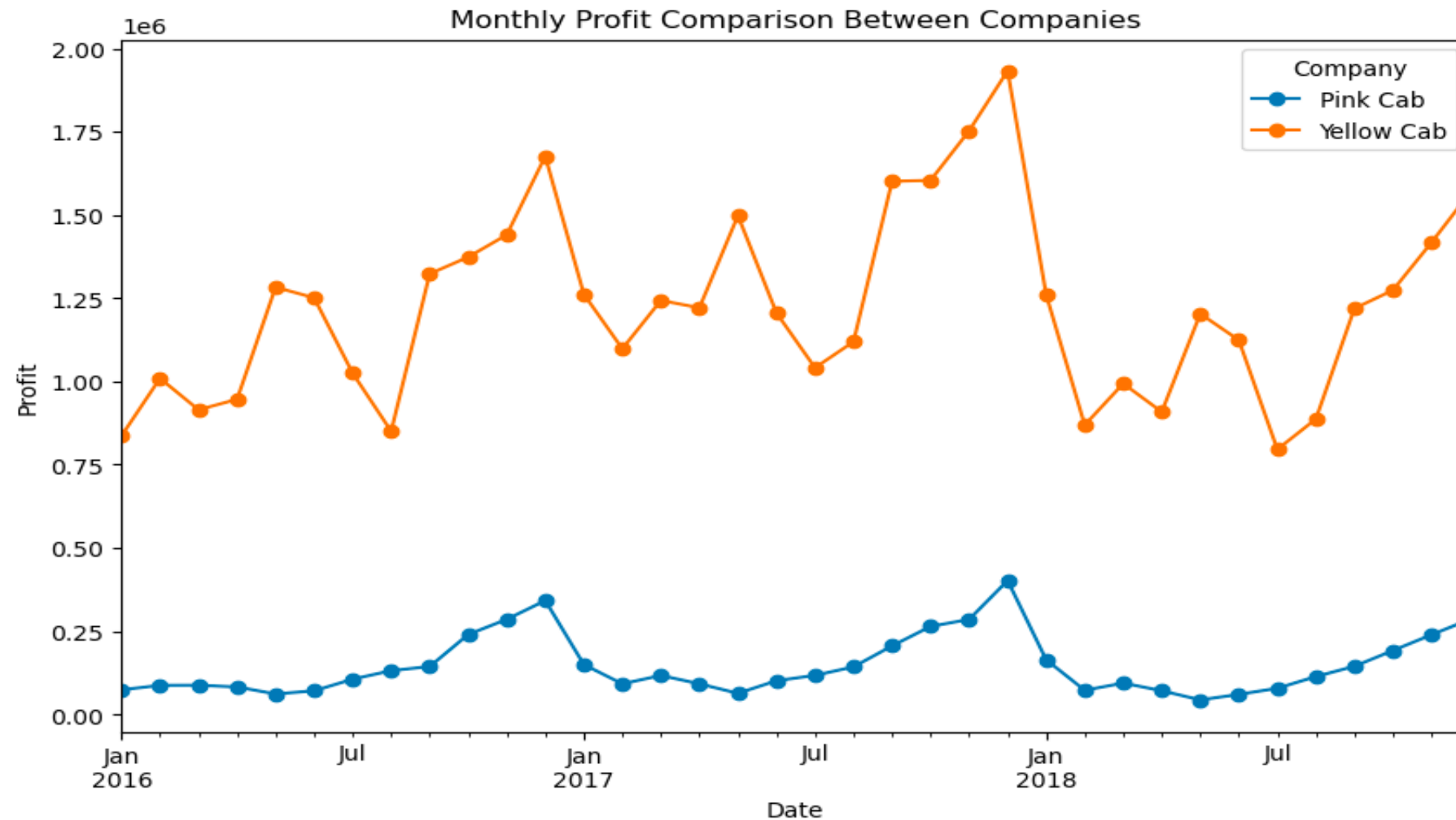
In terms of monthly profit, yellow cabs consistently earn a higher profit margin than pink cabs. Both companies experience a surge in profits in December, likely attributed to the Christmas season, marking the peak earnings for both

Yellow cabs make more money each year because they earn more every month. The difference in yearly profits shows that yellow cabs make a lot more money than pink cabs.


Yearly Profit Comparison Between Companies
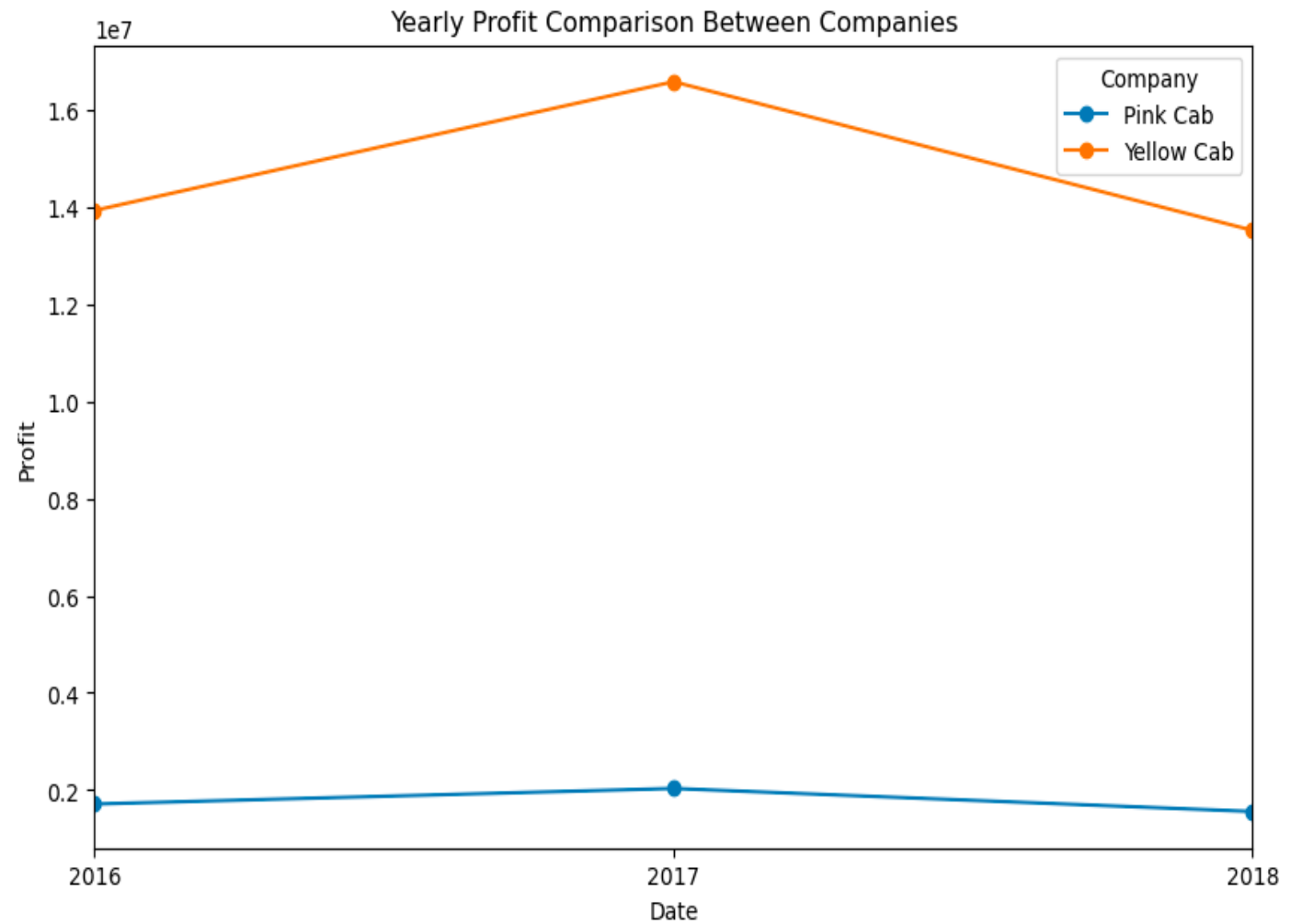
Yellow cabs charge higher prices, which is why they make a lot more money in revenue compared to pink cabs
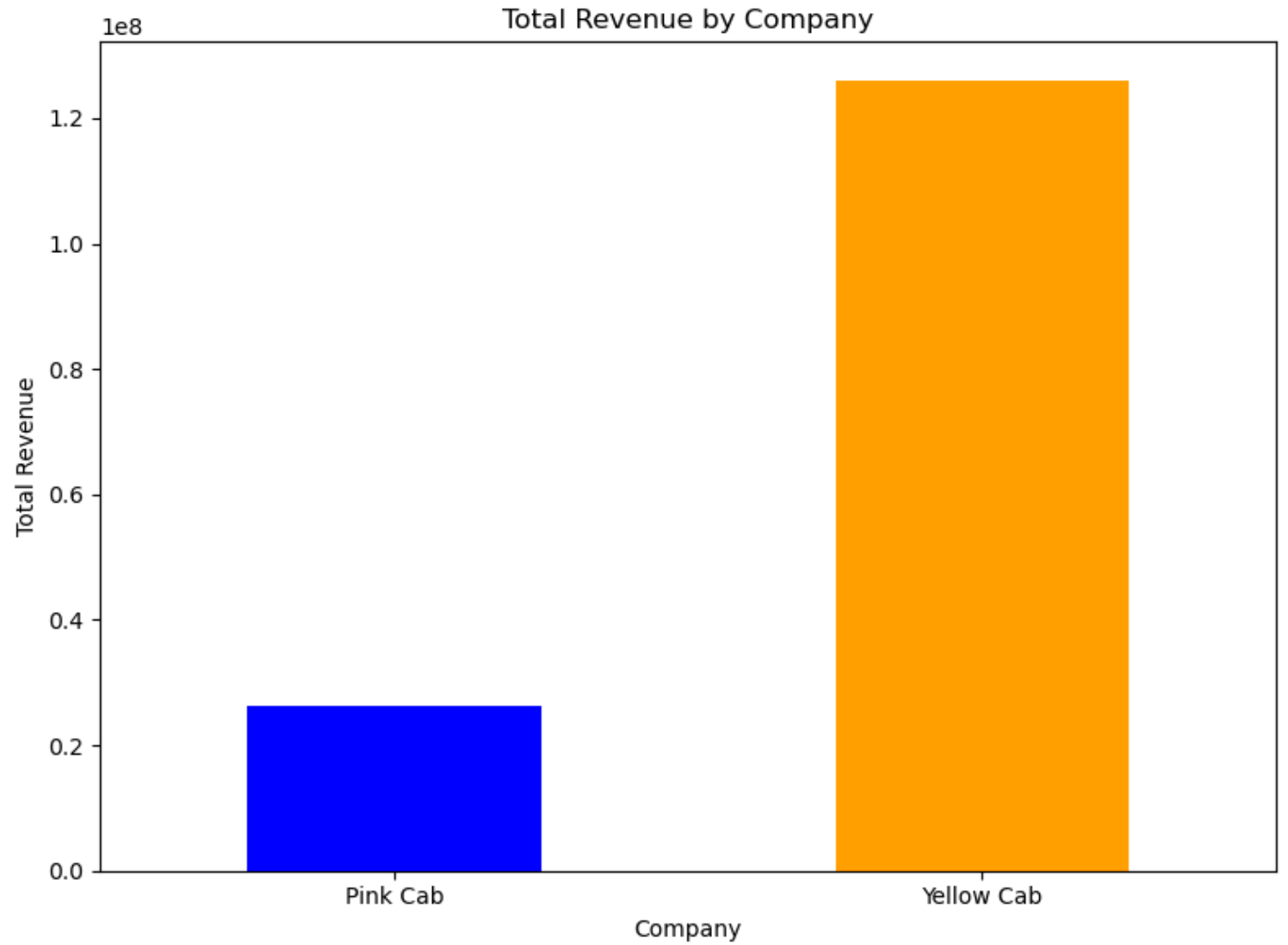
More customers prefer using cards, but there are still many who use cash, especially with pink cabs.

Payment Modes by Company
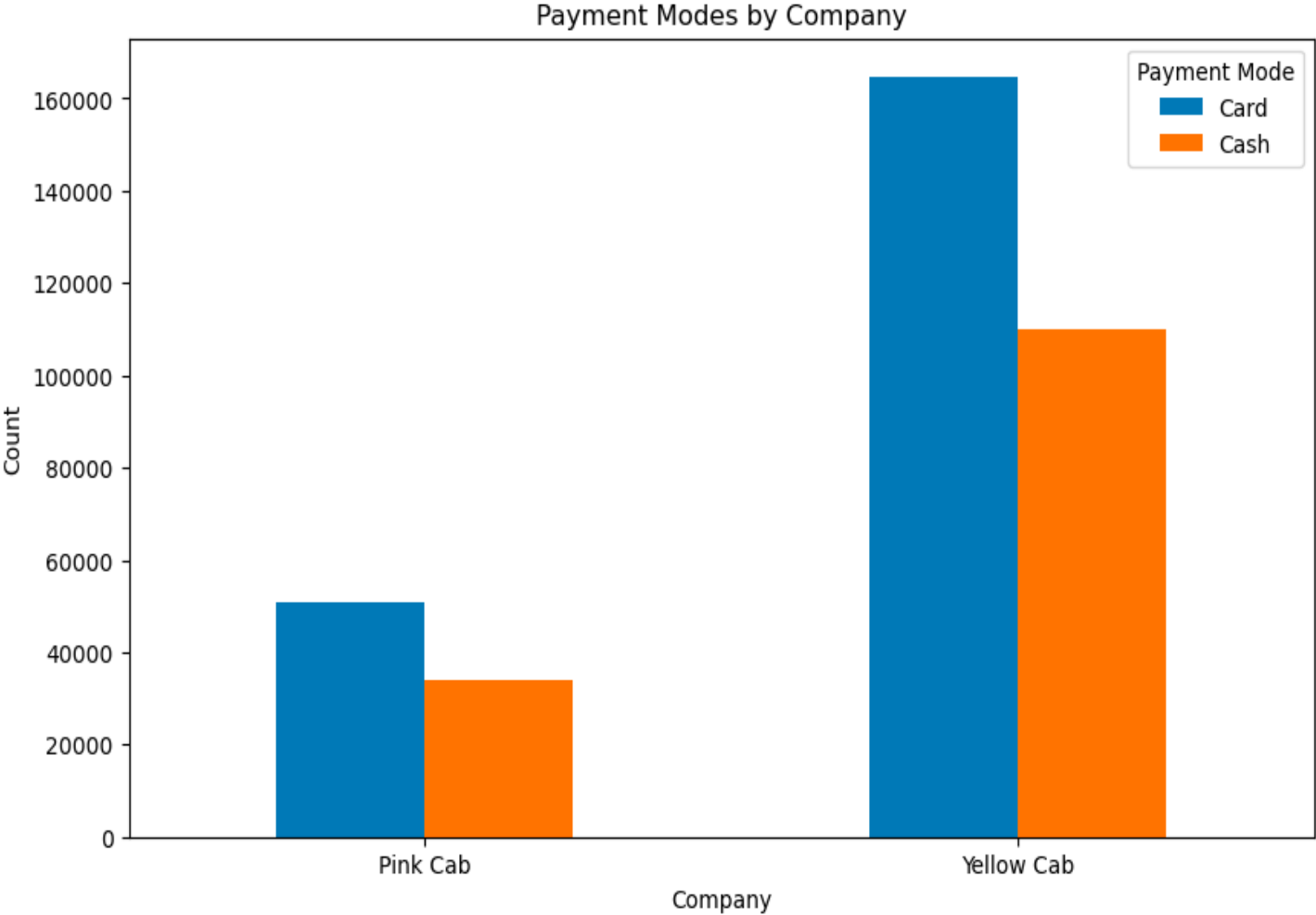
The chart indicates a correlation between pricing and factors like profit margin, trip cost, and distance traveled, suggesting pricing's influence on these elements. Population size impacts cab users, yet infrastructure, transport options, and lifestyle choices collectively shape usage patterns within a city.
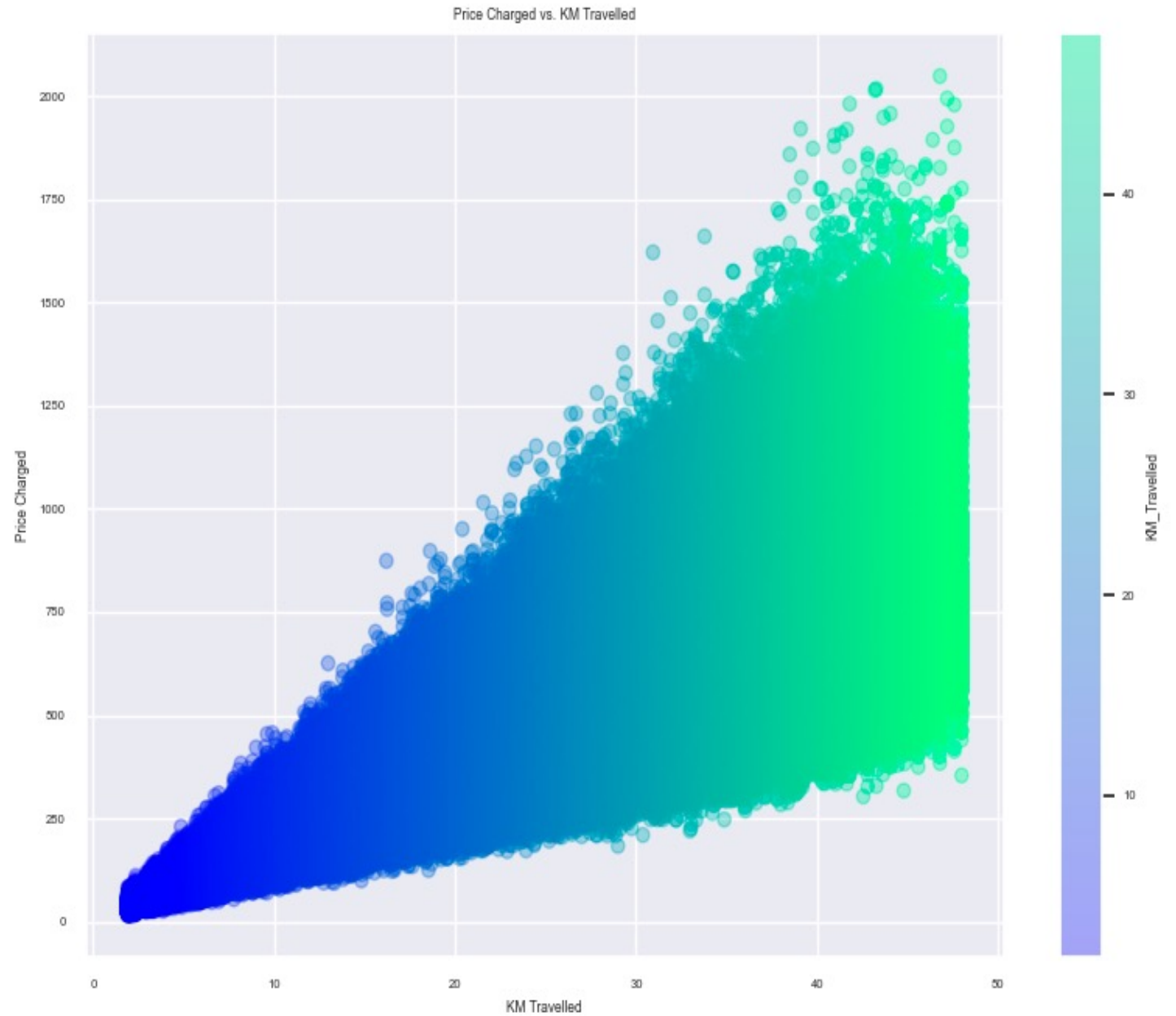

Correlation Heatmap
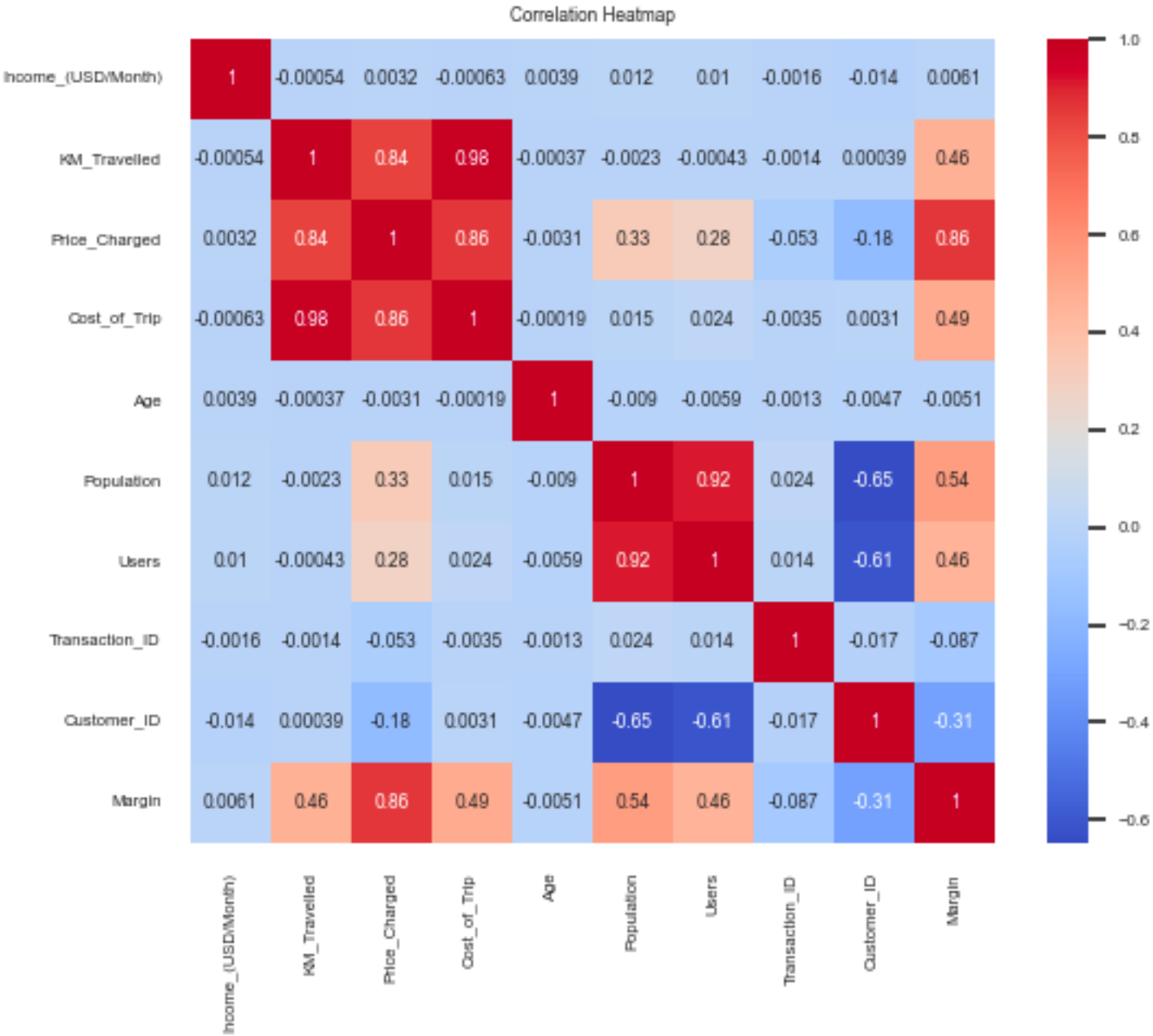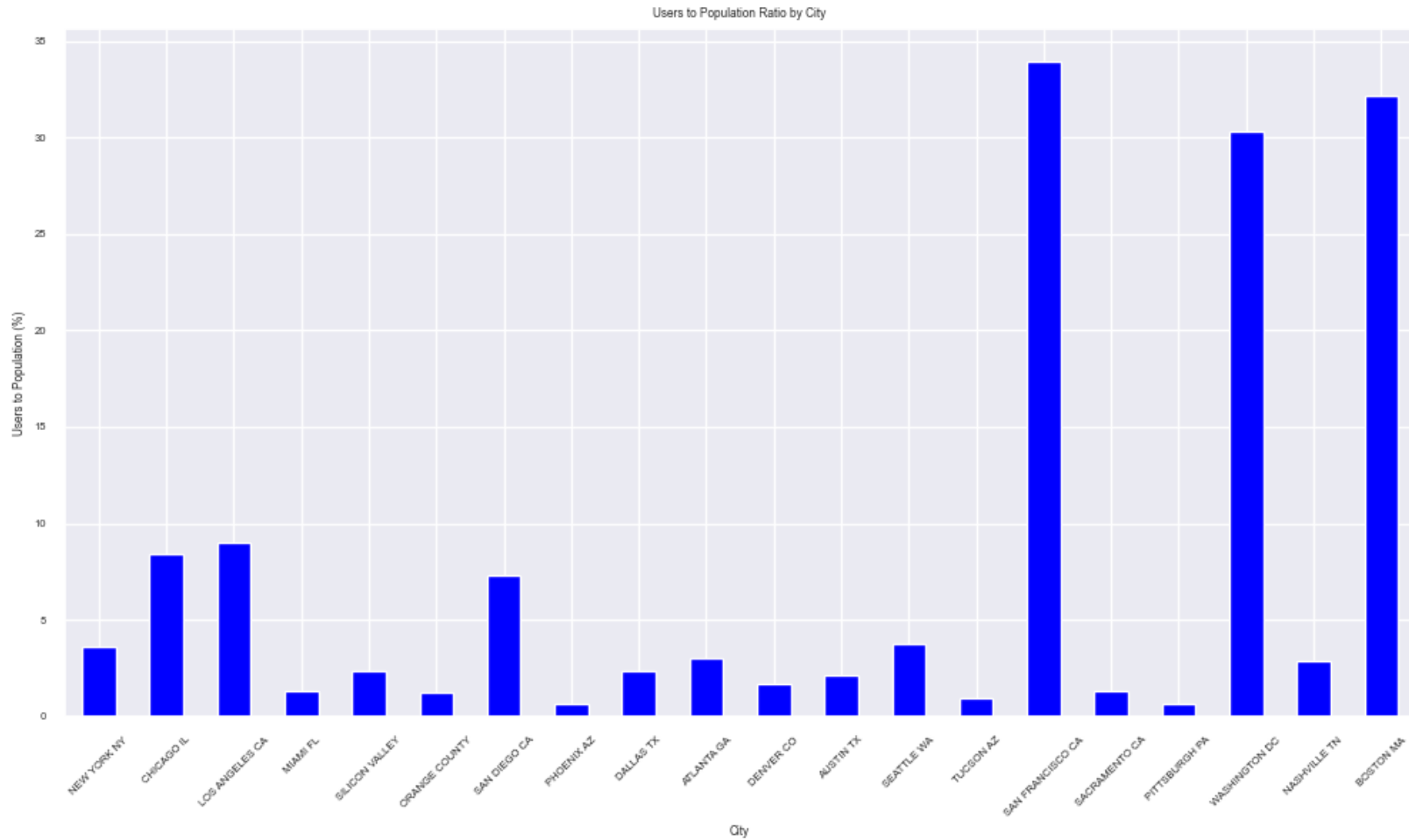
Males tend to use more cabs than females
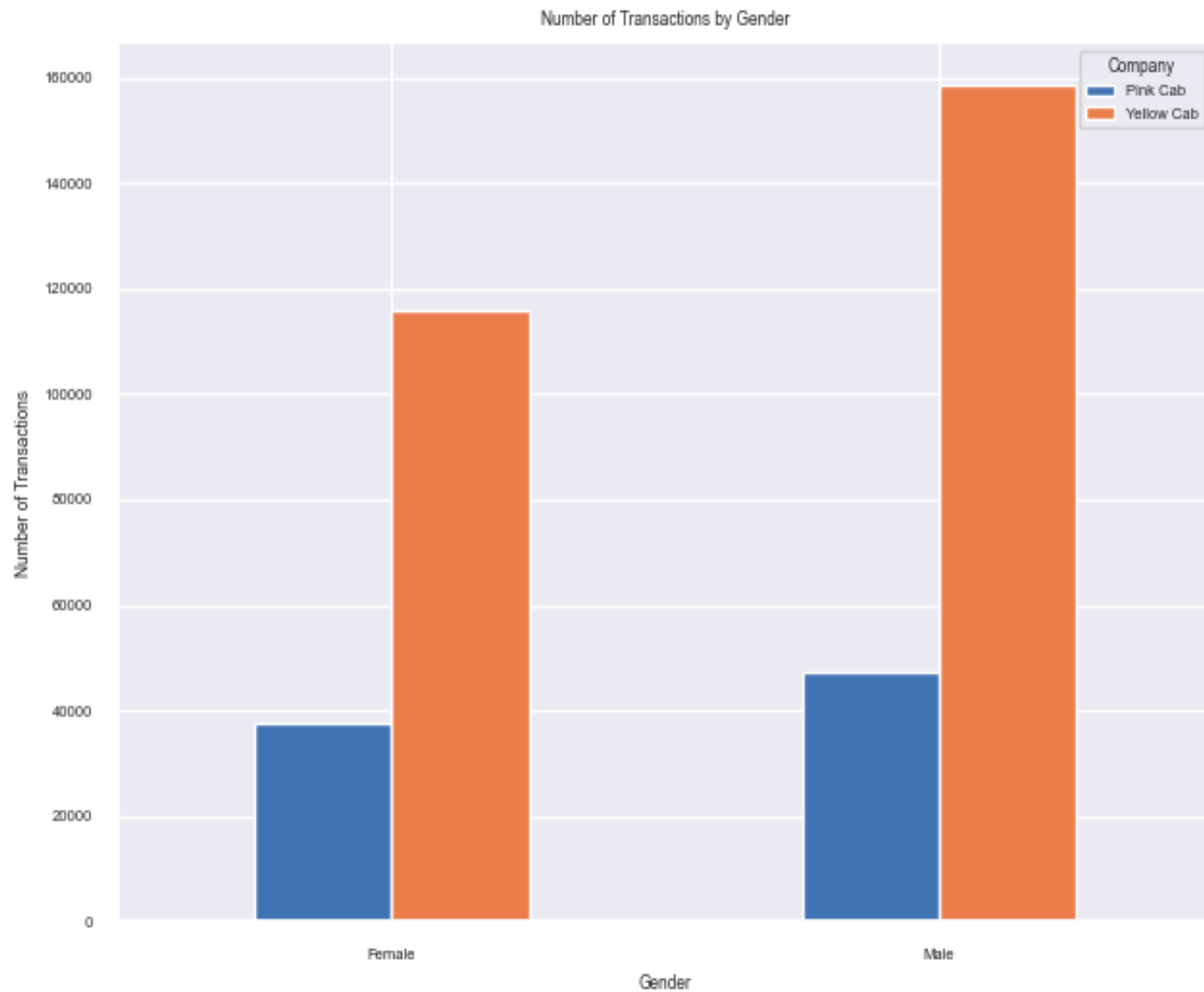
Total transaction by Gender
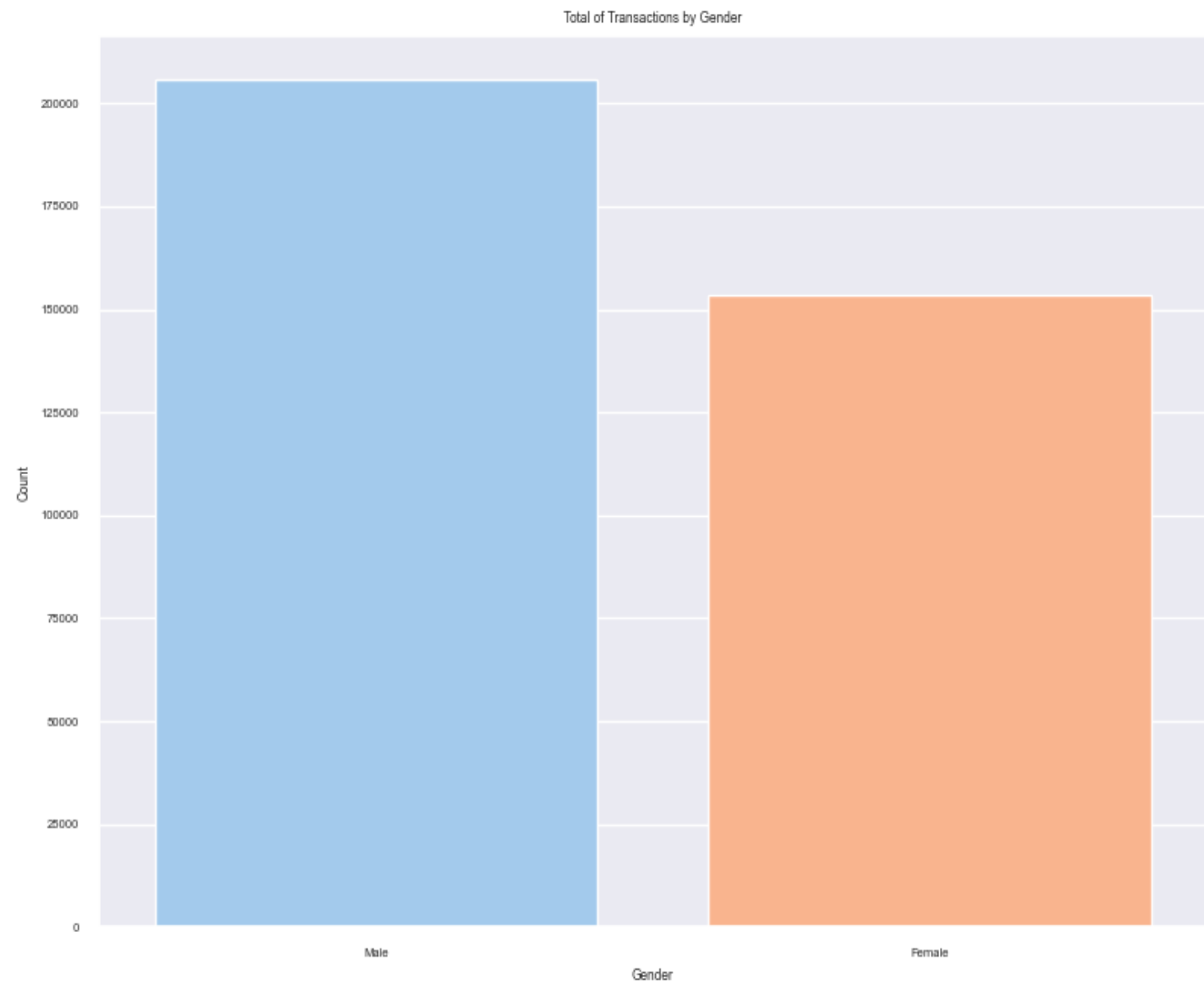
The farther you travel in kilometers, the higher the price charged will be

Users to Population Ratio by City

San Francisco, Boston, and Washington show higher cab user ratios relative to their population compared to New York, Chicago, and Los Angeles. Despite a lower user ratio, New York still holds the highest total customers due to its larger population, indicating higher absolute numbers compared to other cities

Average income by City
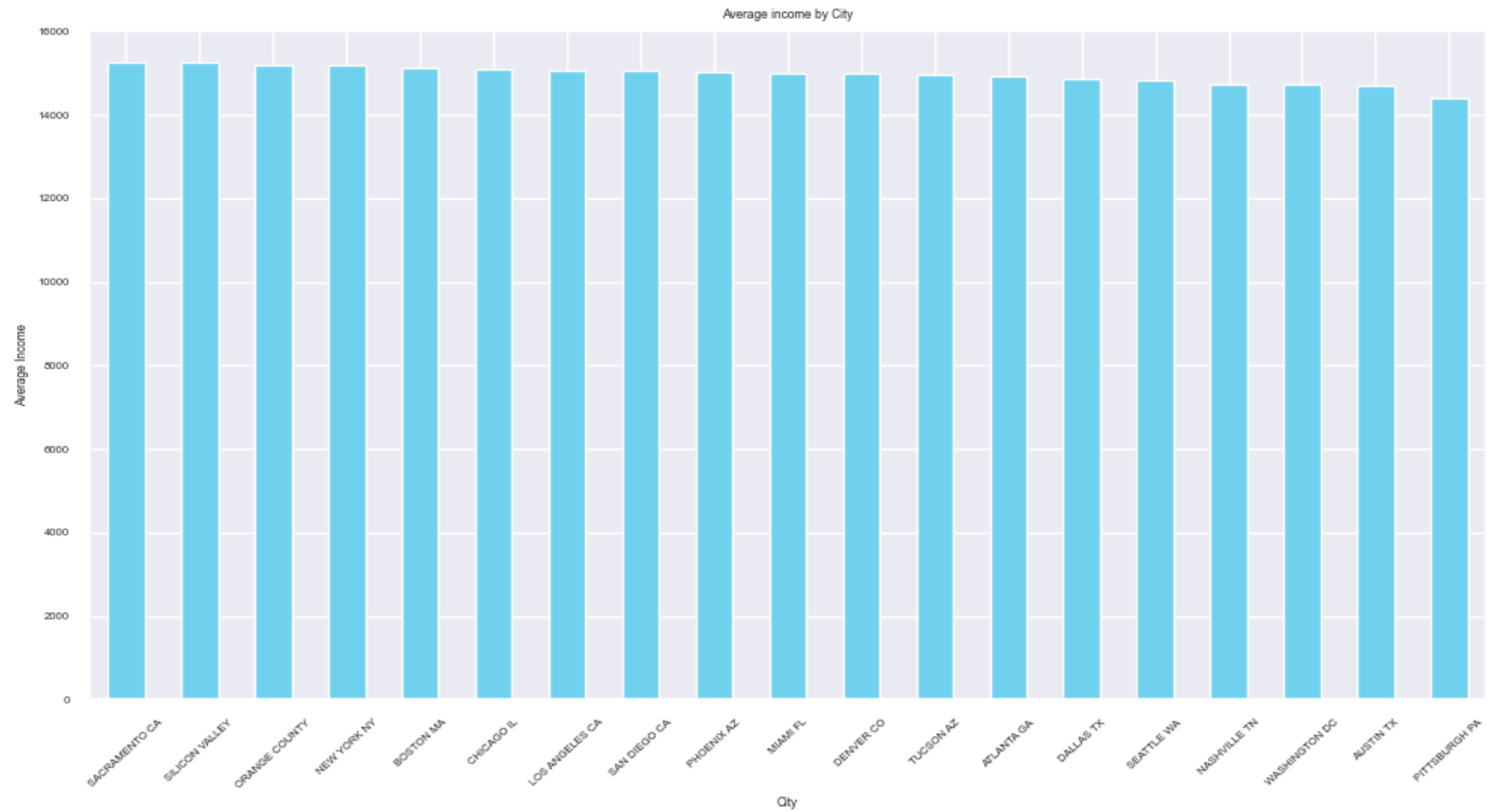
The average income is quite similar in every city

Count of Users by Age
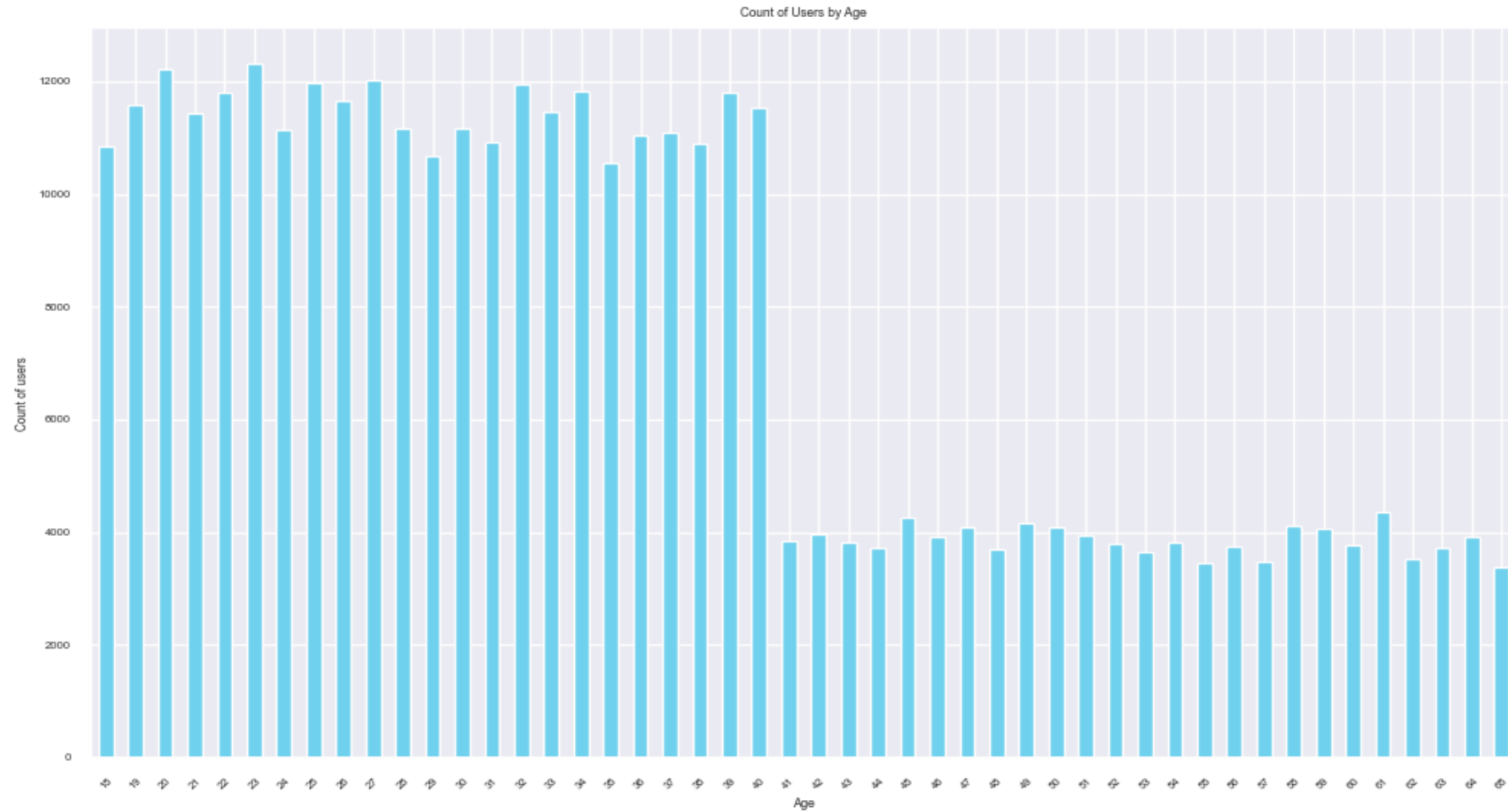
Younger individuals under 40 exhibit higher cab usage, possibly due to preferences for faster transport, unique work or lifestyle commutes, and convenience. Factors like tech-savviness and familiarity with app-based services also influence their higher usage
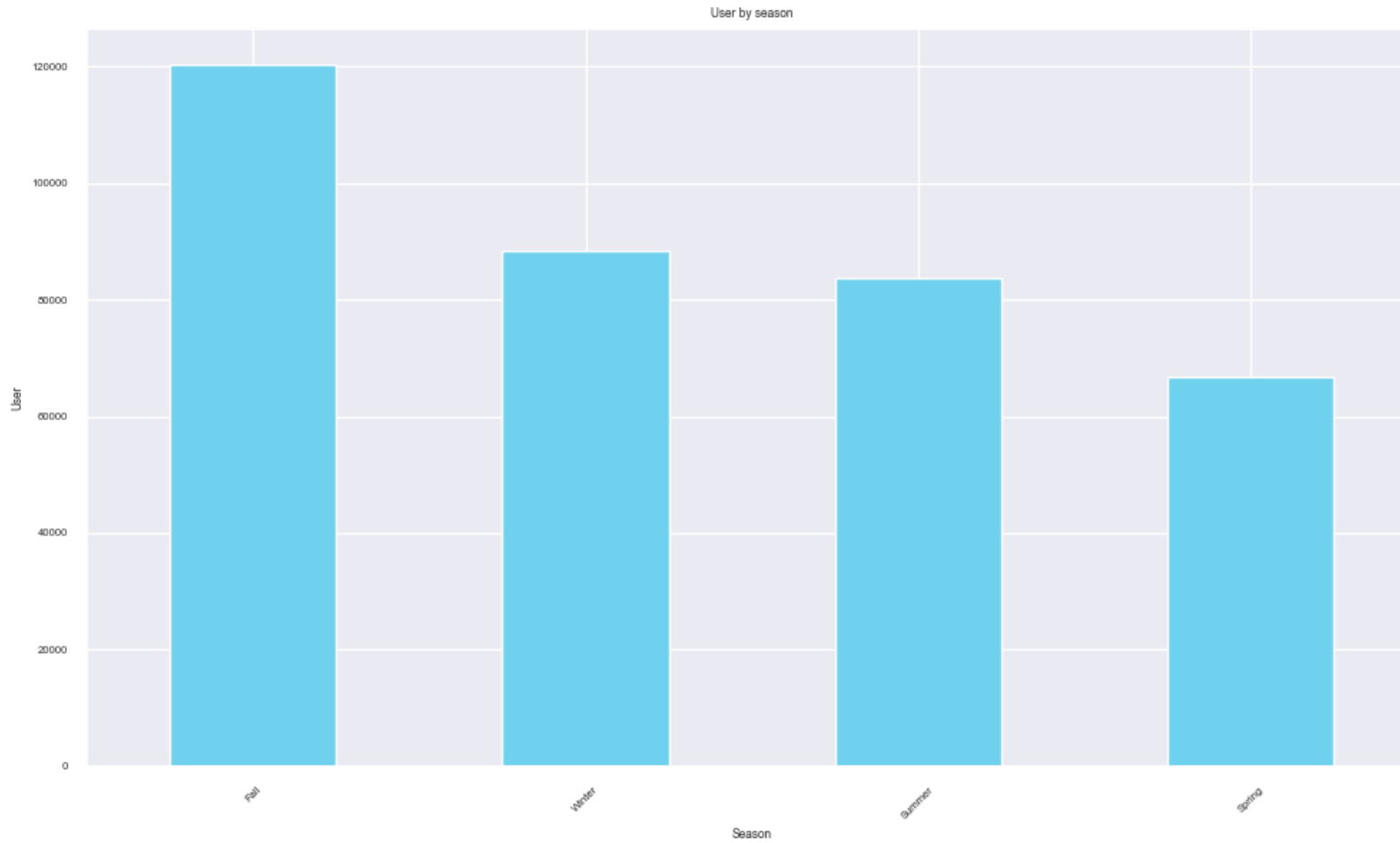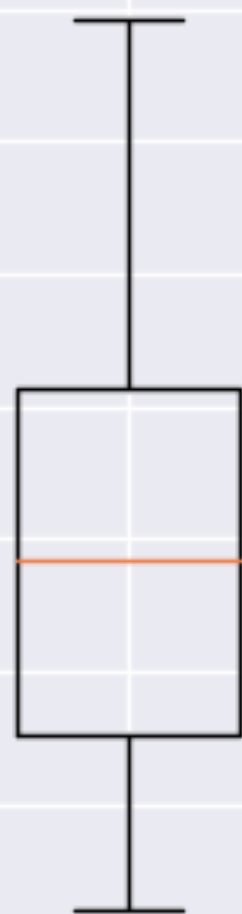
It appears that fall is a more popular season for people using cabs. Possibly during the fall, there might be more events, festivities, or cultural activities happening, prompting increased travel and hence, a higher demand for cab services.
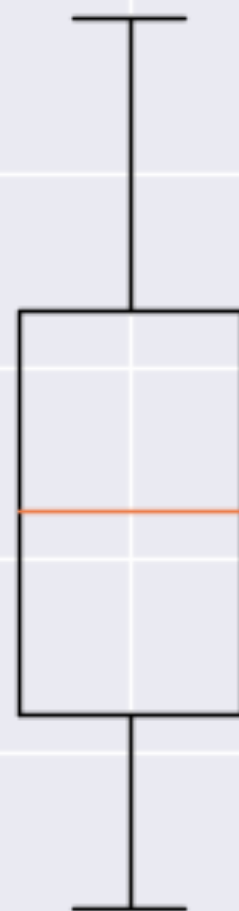
User by holiday season
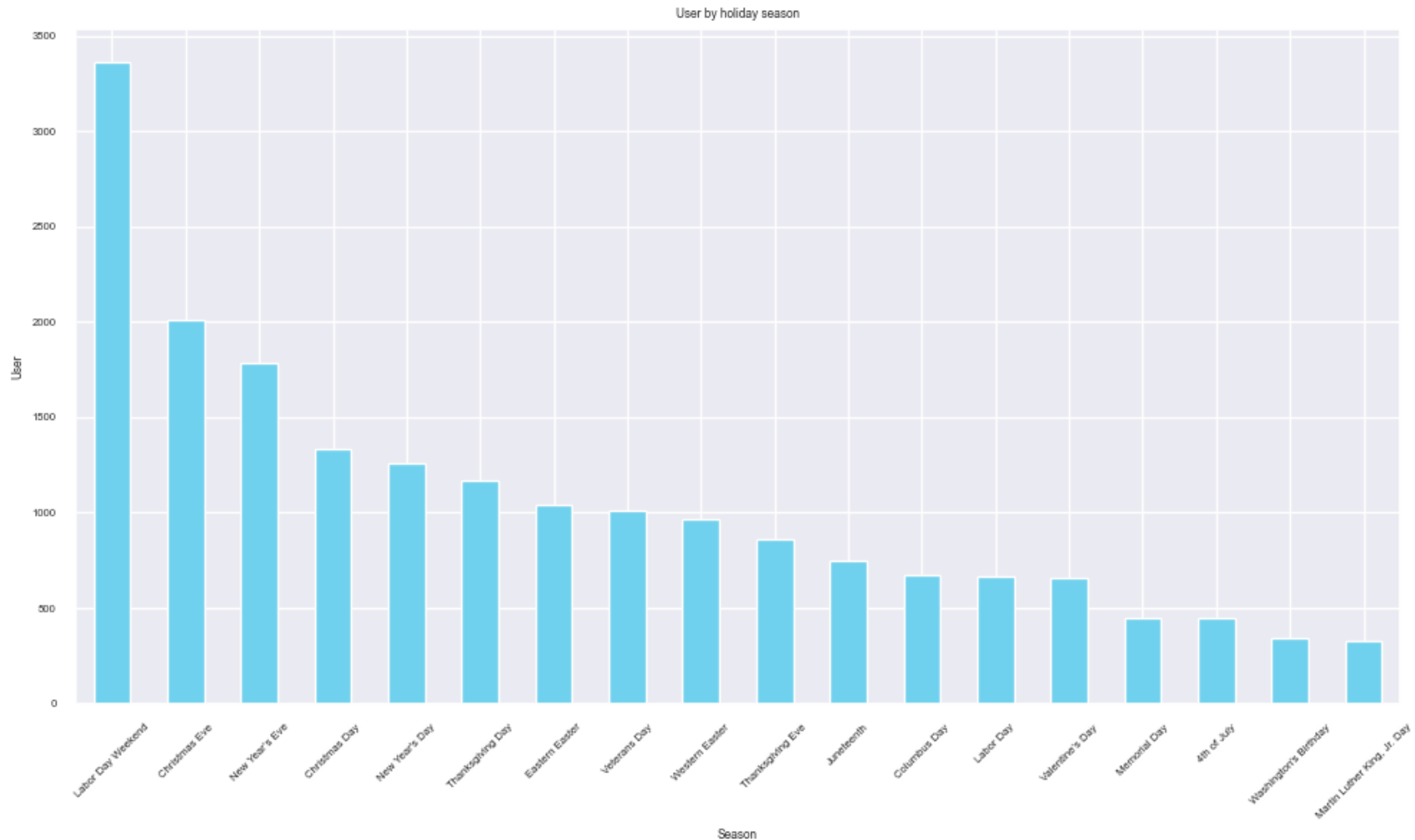
Labor Day marks the busiest holiday season for cab service users, followed by Christmas and New Year's. The monthly profit analysis also indicates a surge in profits starting from September, continuing to increase steadily until December

Boxplot of Cost_of_Trip

Boxplot of KM_Travelled

KM_Travelled
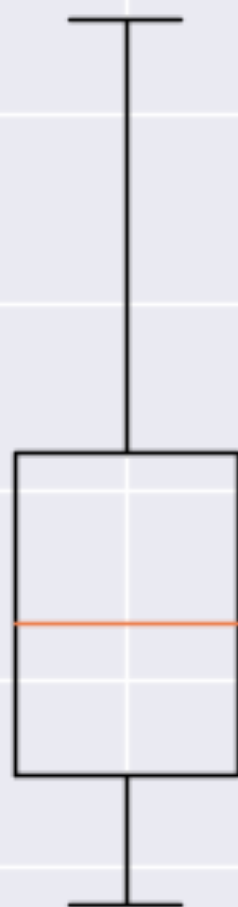
Boxplot of Price_Charged
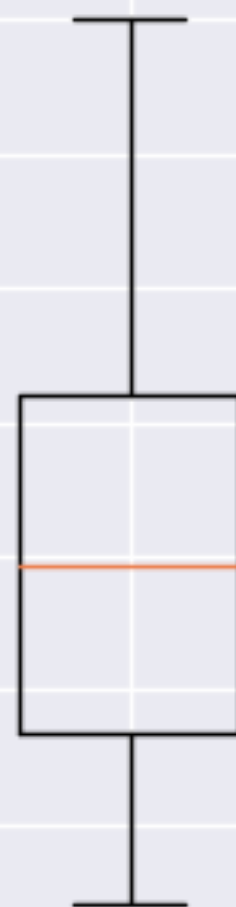
Cost_of_Trip

Price_Charged

Boxplot of Age

Boxplot of Margin

Boxplot of Income_(USD/Month)

Age

Margin

Income_(USD/Month)

Recommendations

1.**Investment Focus:** Yellow Cab holds 76% market share, suggesting a more profitable investment over Pink Cab (26%). Directing investment toward Yellow Cab could yield higher returns.

2.**Customer Retention:** Prioritize loyalty programs, personalized offers, and service improvements based on Yellow Cab's better customer retention to reinforce loyalty industry-wide.

3.**Geographical Strategy:** Target markets like Chicago and Los Angeles, focusing on cities with potential Yellow Cab growth, despite New York's high usage.

4.**Pricing Strategy:** Despite higher costs, Yellow Cabs maintain better profit margins. Consider maintaining slightly higher prices while optimizing costs without service compromise.

5.**Seasonal Demand:** Optimize marketing during peak holidays like Labor Day, Christmas, and New Year's to capitalize on increased demand and maximize profits.

6.**Demographic Targeting:** Customize marketing for the under-40 demographic, recognizing their higher cab service usage, emphasizing app convenience.

7.**Data Integration:** Utilize weather and event datasets to uncover correlations and enhance insights into seasonal usage patterns.

8.**Continuous Analysis:** Regularly monitor trends, conduct customer surveys, and adapt strategies based on evolving customer preferences and market dynamics for ongoing success.

Thank You