Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

Michael Udonna Egbuzobi

20-August-2024

Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

Michael Udonna Egbuzobi

20-August-2024

Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- Hypothesis Testing
- Recommendations

Executive Summary

XYZ, a private firm in the US, is considering an investment in the competitive cab industry. To support this decision, a detailed analysis of two cab companies' data from 2016 to 2018 was conducted. The analysis focuses on ride volumes, profitability, customer demographics, and market trends. Key insights will guide XYZ in identifying the optimal investment opportunity, ensuring alignment with market conditions and customer preferences.

Problem Statement

Given the presence of multiple key players in the cab industry, XYZ seeks to make an informed investment decision. The goal is to understand the market dynamics, customer behaviour, and financial performance of two cab companies. This analysis provides actionable insights to determine the better-performing company and the most promising investment opportunity for XYZ.

The Datasets

The analysis utilizes four datasets:

- **Cab_Data.csv:** Contains transaction details for two cab companies.
- **Customer_ID.csv:** A mapping table linking customer IDs to their demographic details.
- **Transaction_ID.csv:** A mapping table linking transactions to customers and payment modes.
- **City.csv:** Lists US cities along with their populations and number of cab users.

Data pre-processing approach

- **Data Integration:** The analysis started with merging the provided datasets in Excel using VLOOKUP functions to create a unified dataset.

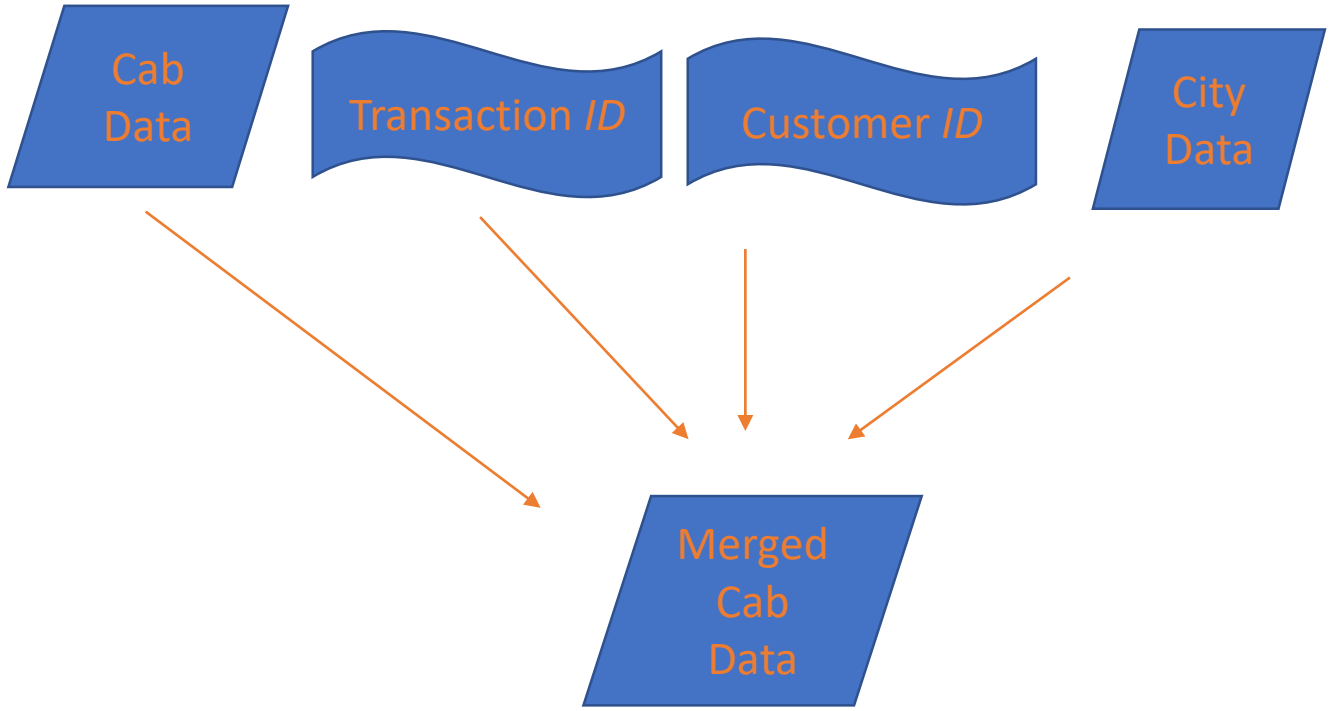
- **Profit Calculation:** A new profit column was calculated and added to the dataset to facilitate financial analysis.

- **Data Integrity Check:** A thorough examination was conducted using the filter function to ensure no blanks or missing values, confirming the dataset's completeness and reliability for further analysis.

Data pre-processing approach

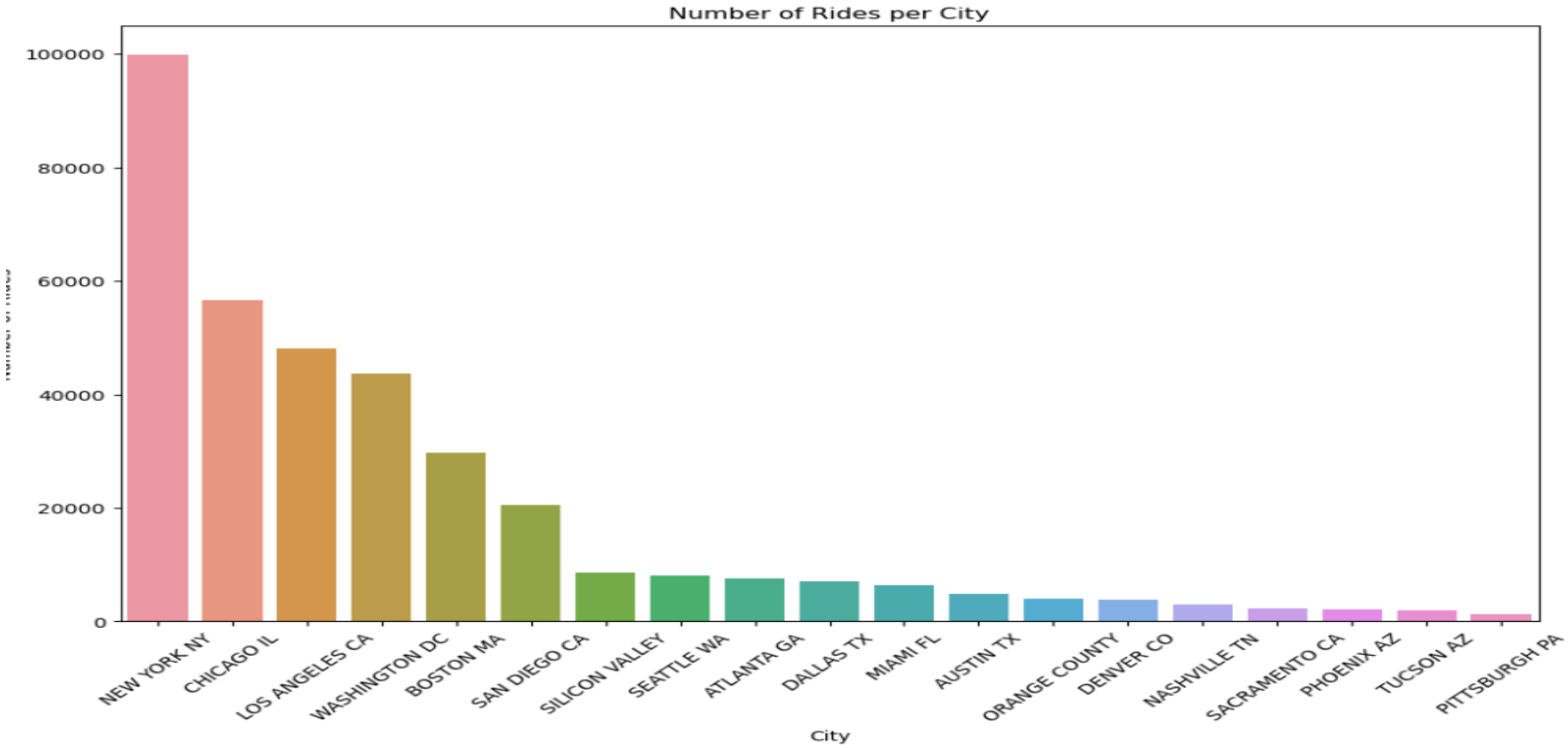
- The final dataset contained 14 features, including five calculated features.

- Dataset contained 359, 394 records.



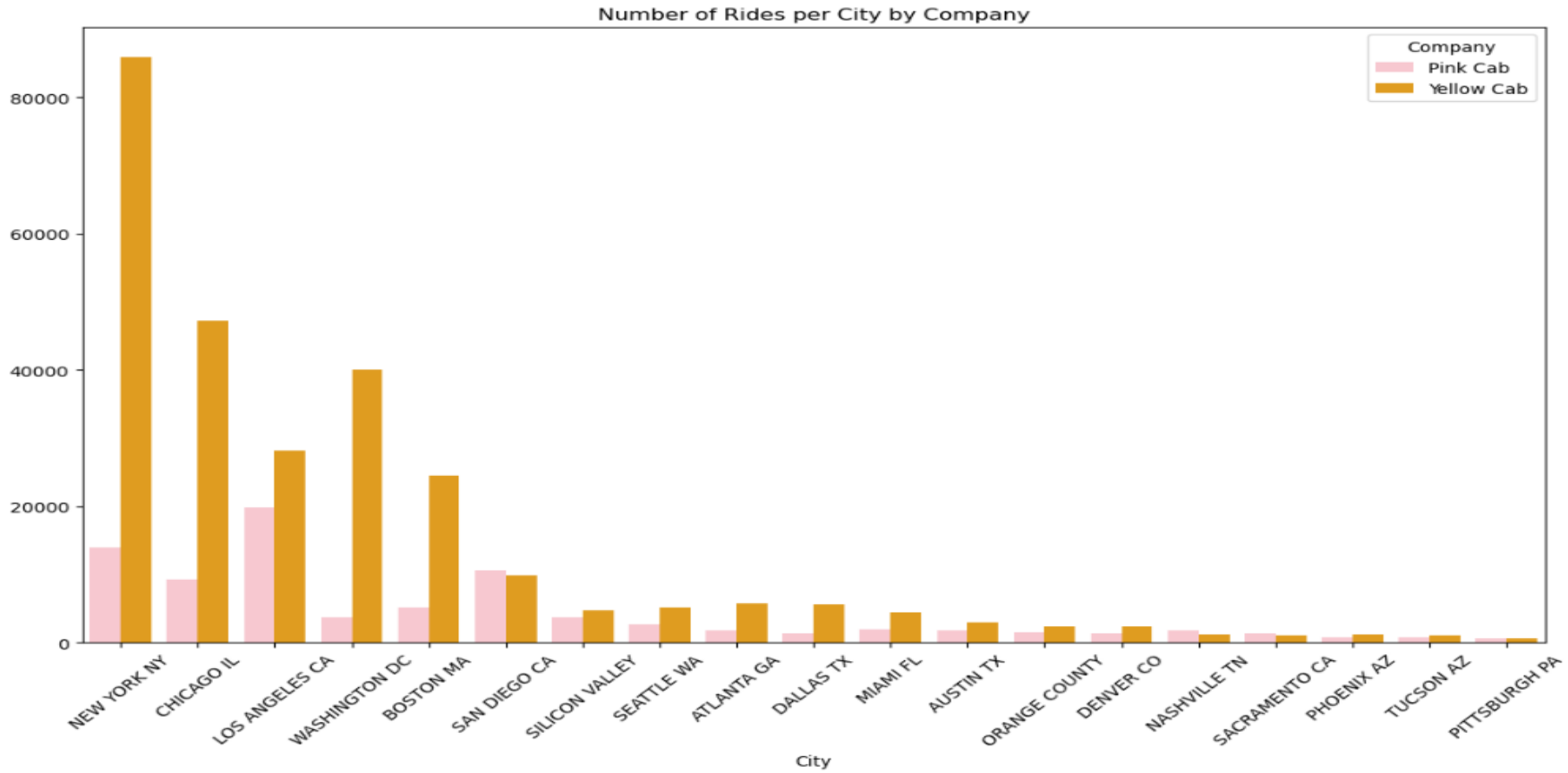
Exploratory Data Analysis

Number of Rides per City



The chart indicates that New York has the highest number of rides, followed by Chicago, Los Angeles, and Washington D.C. This suggests that larger metropolitan areas tend to have a higher volume of ride bookings compared to smaller cities.

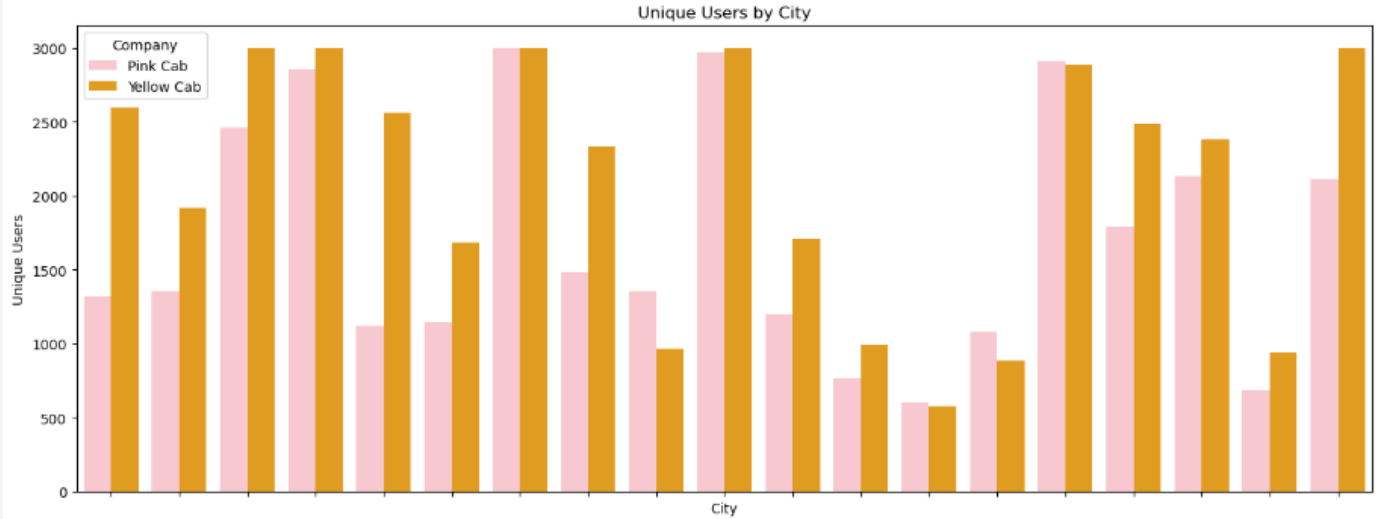
Number of Rides per City by Company

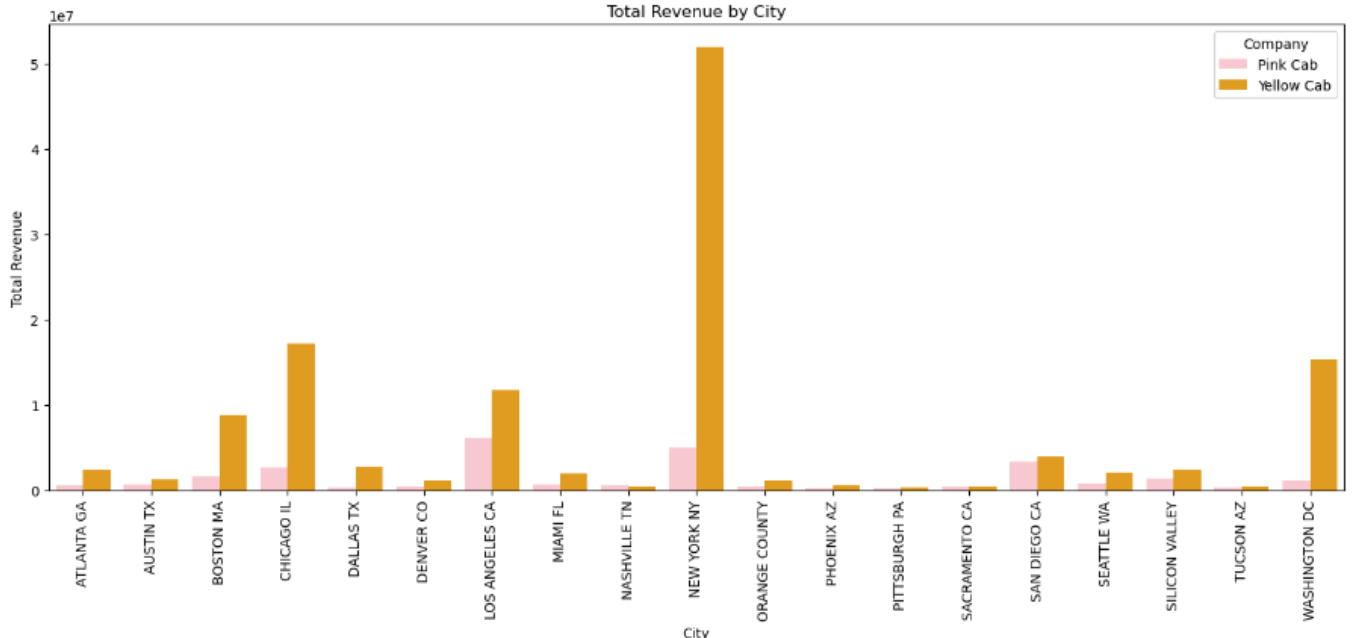


After adding "company" as a variable, it was observed that the majority of the customer base for both companies is concentrated in large metropolitan areas, as described by the number of rides per city, with Yellow Cab having the highest number of rides.

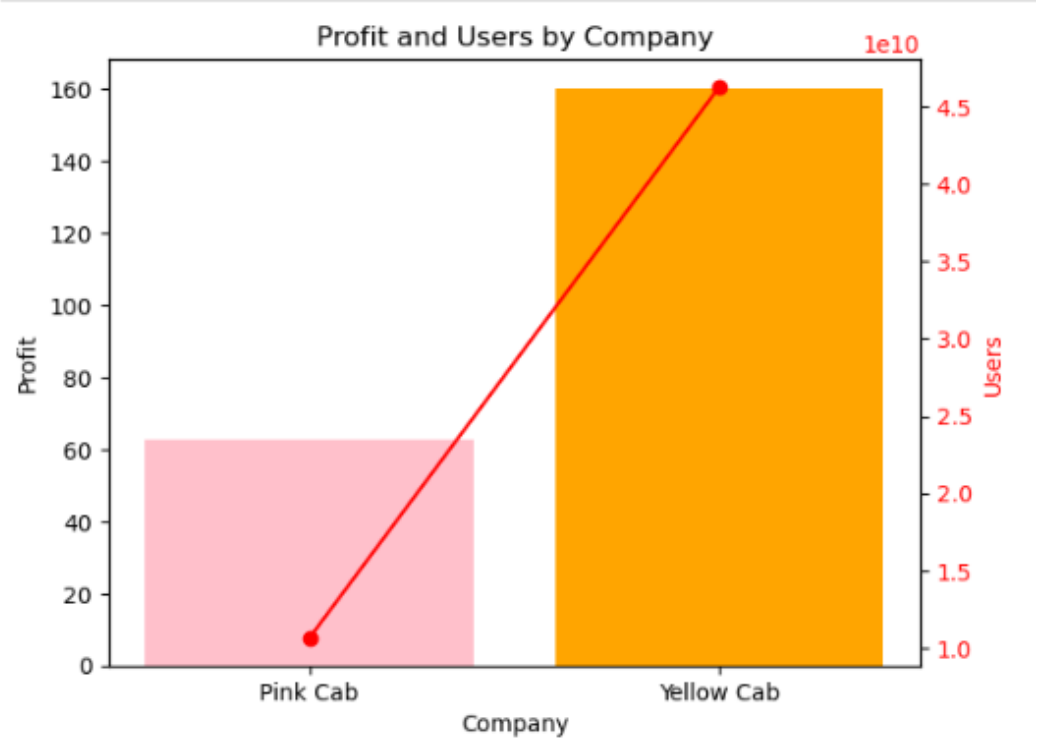
Number of Unique Users and Total Revenue per City by Company

The analysis of the number of unique users—defined as distinct individuals who utilized the service, including first-time users—and total revenue by city aligns with previous findings. Large metropolitan cities generate more revenue and attract a higher number of unique users, with Yellow Cab slightly outperforming Pink Cab in both metrics.



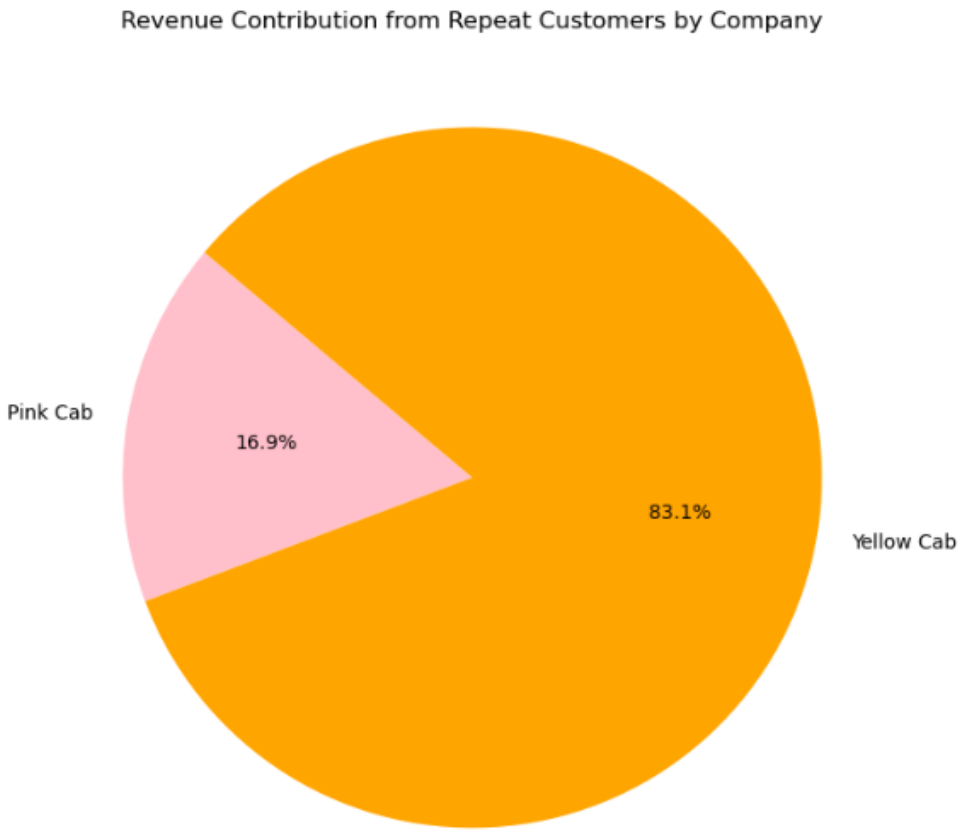


Profit and Users by Company



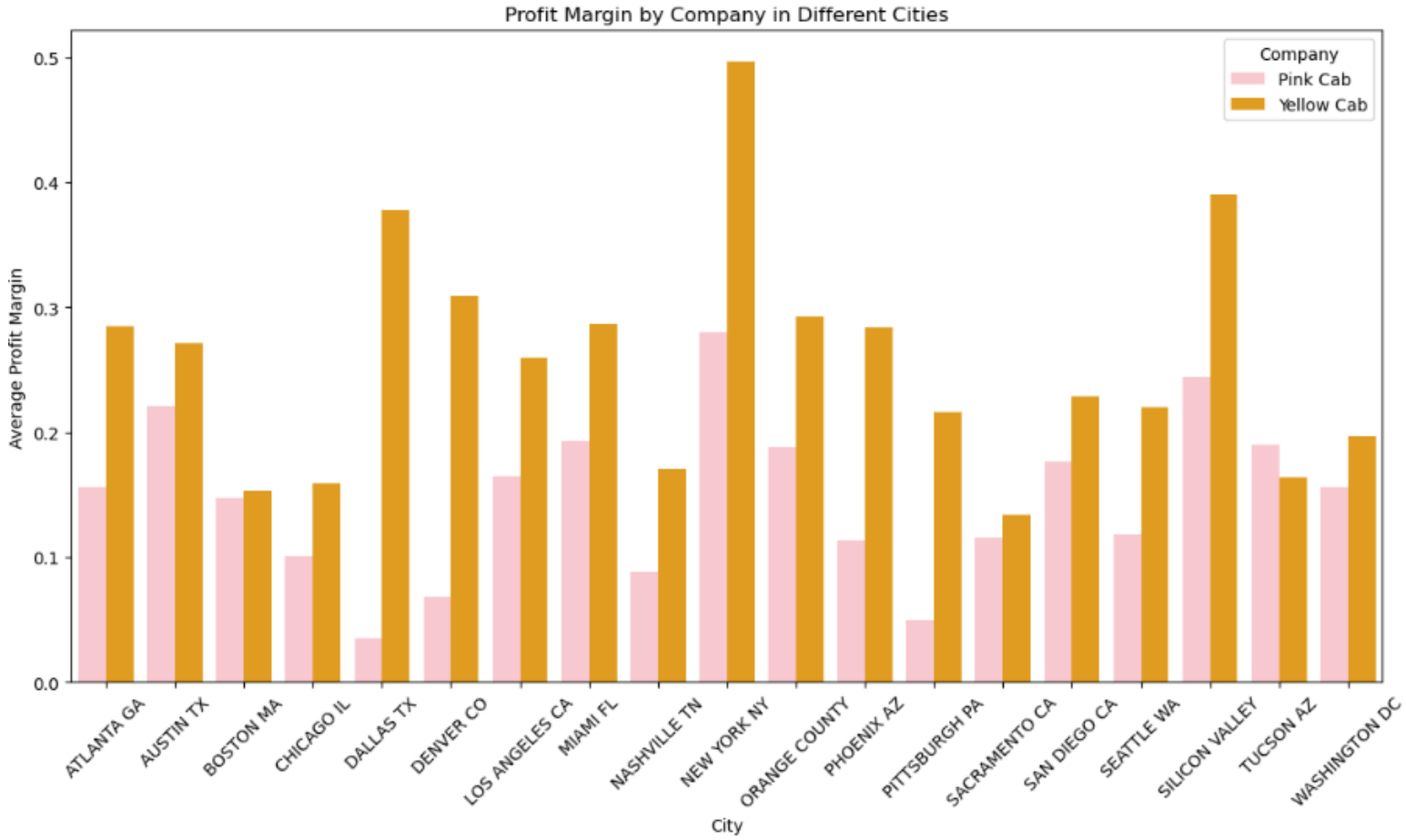
Once again, Yellow Cab outperforms Pink Cab in both total profit and number of users per company.

Revenue Contribution from Repeat



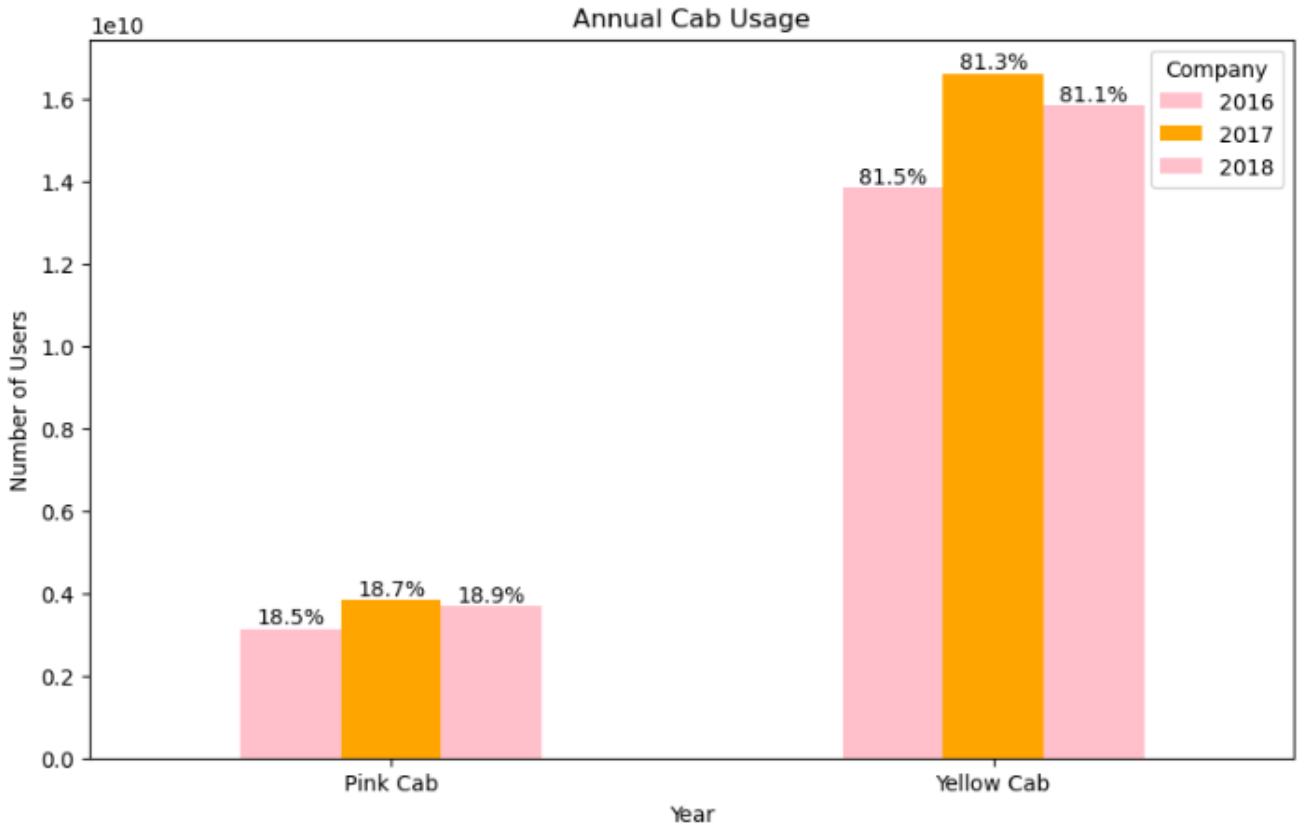
Yellow Cab has a significantly higher percentage of repeat customers, accounting for 83.1%, while Pink Cab has only 16.9%, indicating that Yellow Cab retains a larger proportion of its customer base.

Profit Margin by Company in Different Cities



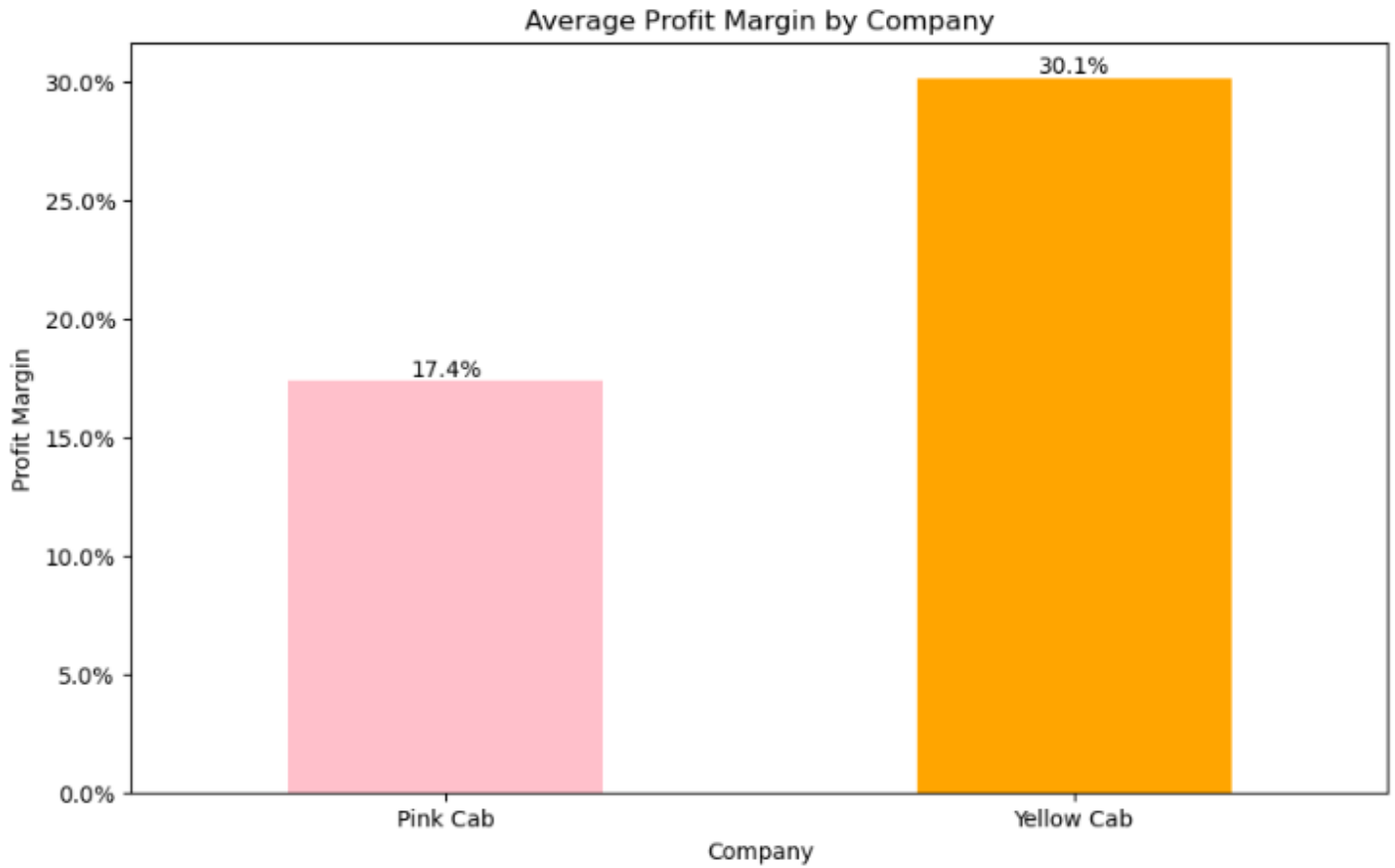
Same as revenue, profit is higher in large metropolitan cities, with yellow cab outperforming pink cab.

Annual Cab Usage



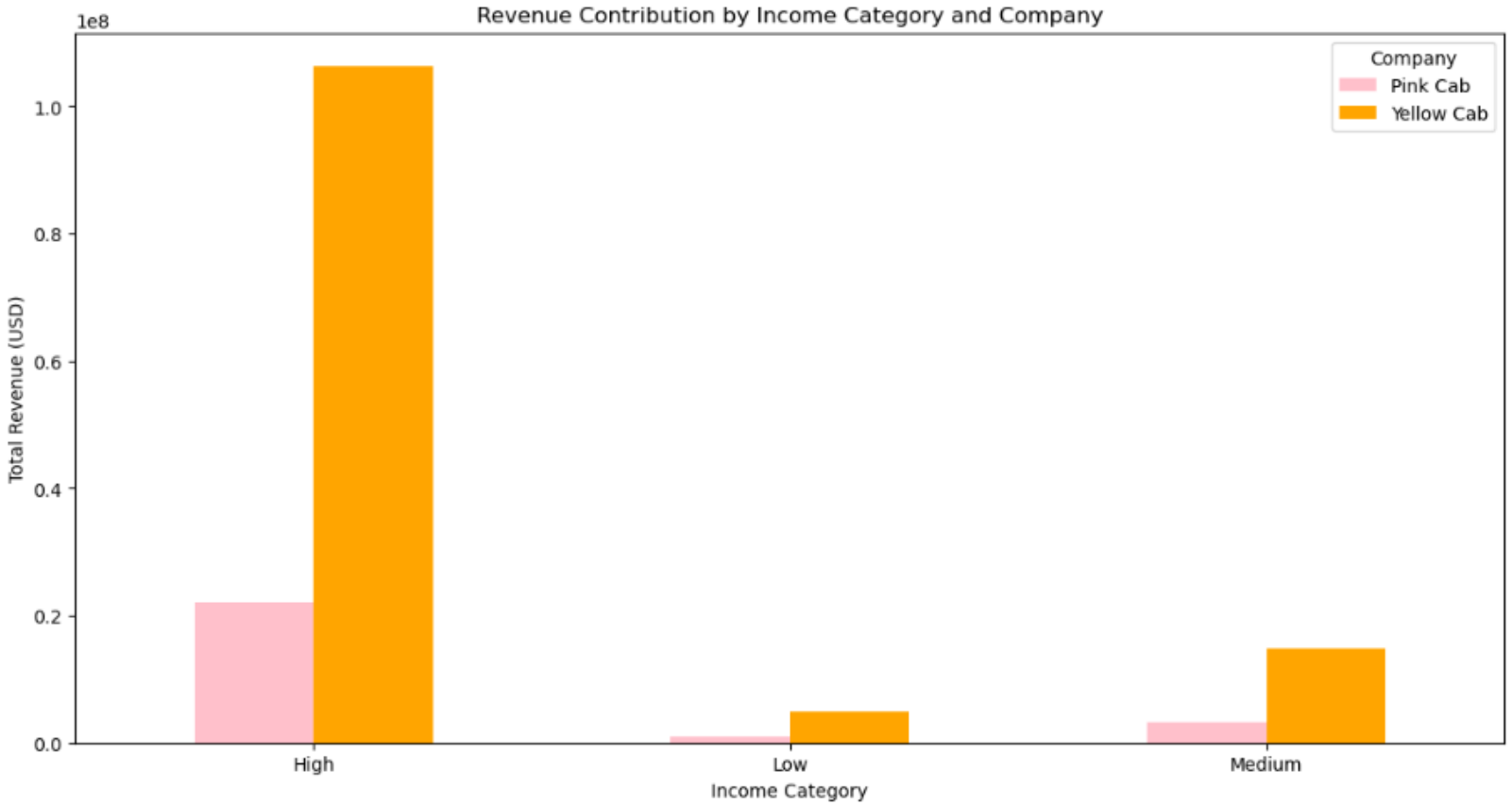
Annual cab usage data shows that Yellow Cab consistently dominates the market, with usage rates of 81.5%, 81.3%, and 81.1% for 2016, 2017, and 2018, respectively, compared to Pink Cab's 18.5%, 18.7%, and 18.9%. Additionally, the data indicates that 2017 was a peak period for both companies.

Average Profit Margin by Company



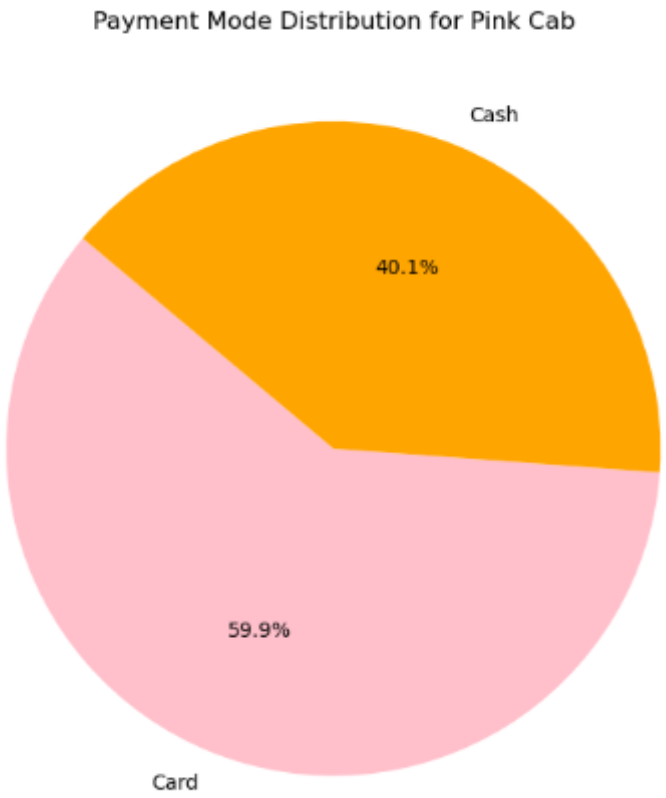
The average profit margins are 17.4% for Pink Cab and 30.1% for Yellow Cab, indicating that Yellow Cab enjoys a substantially higher profit margin compared to Pink Cab.

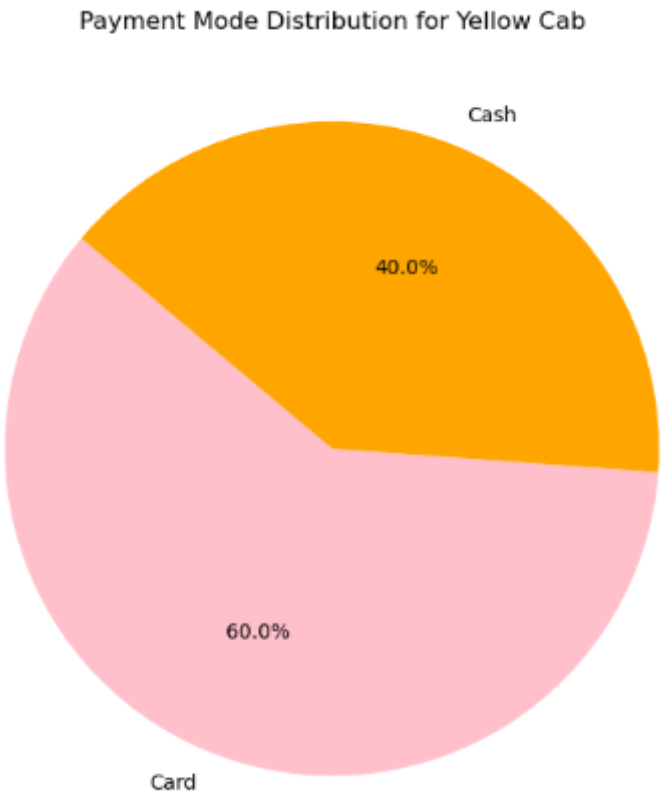
Revenue Contribution by Income Category and Company



The analysis of revenue contribution by income category and company indicates that high-income earners utilize cab services more frequently than other income groups. Furthermore, Yellow Cab derives greater benefits from this demographic compared to its competitor.

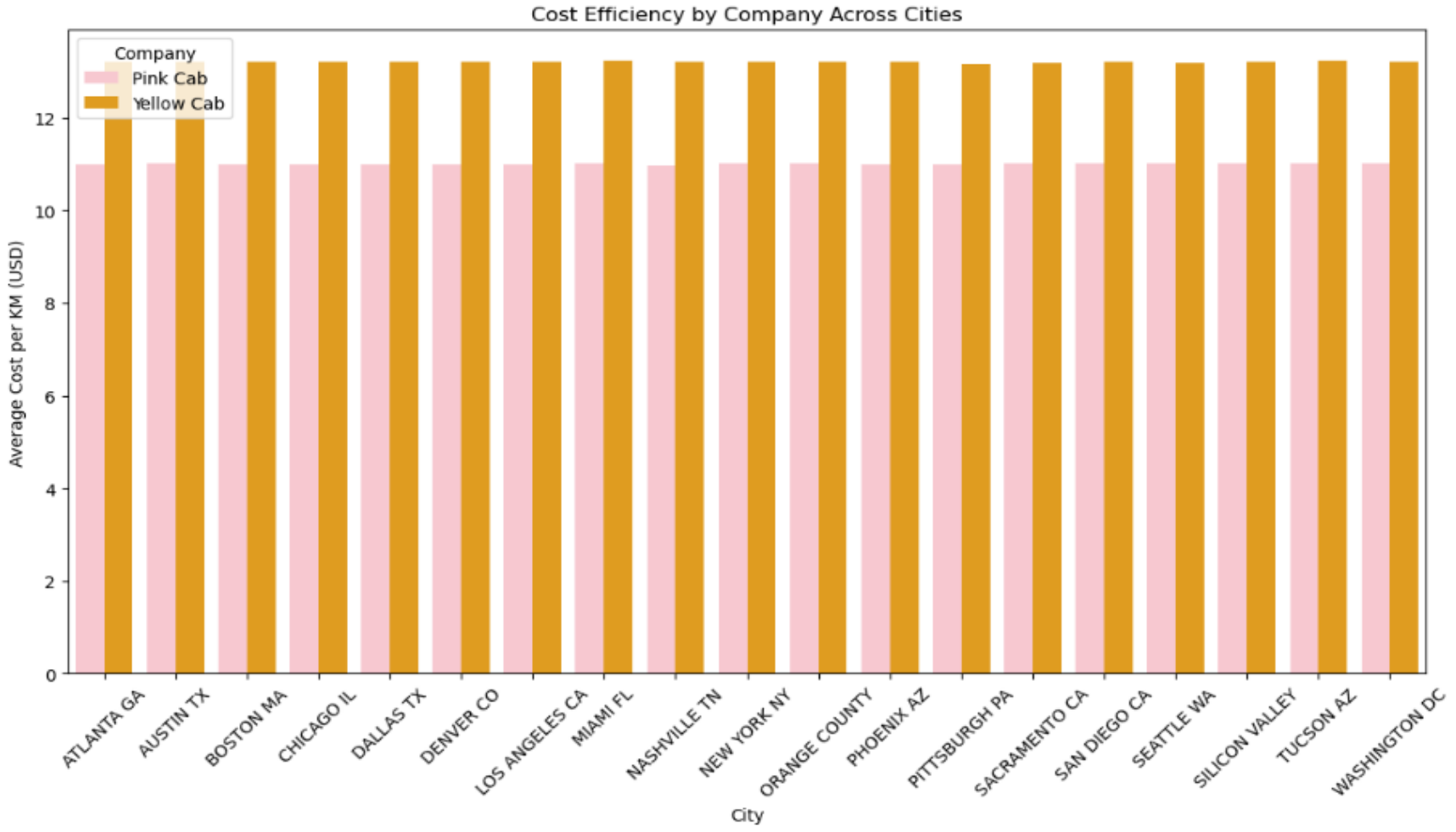
Payment Mode Distribution by Company





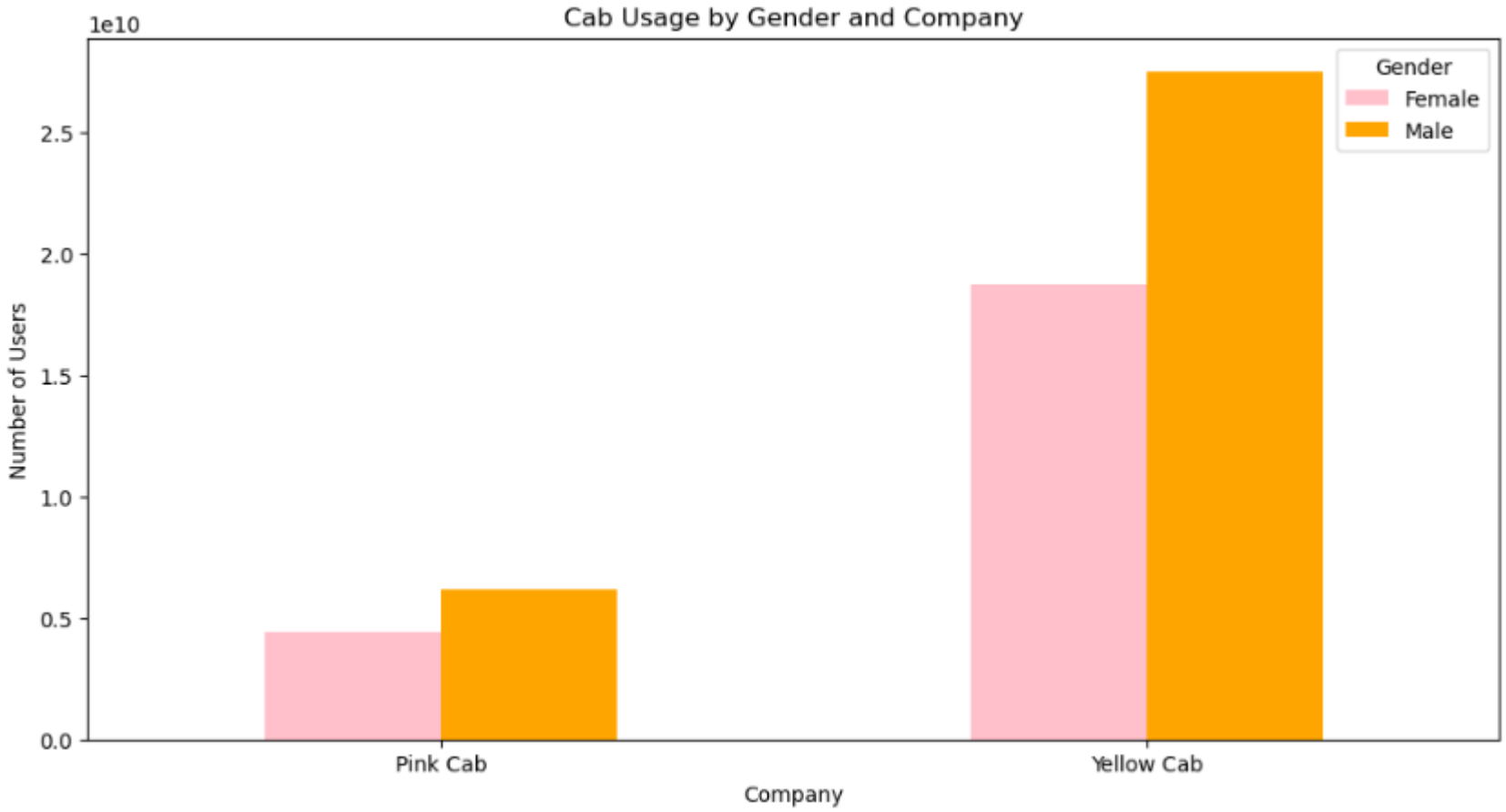
The data reveals that customers predominantly use card payments for both companies, suggesting a strong preference for card transactions over other payment methods.

Cost Efficiency by Company Across Cities



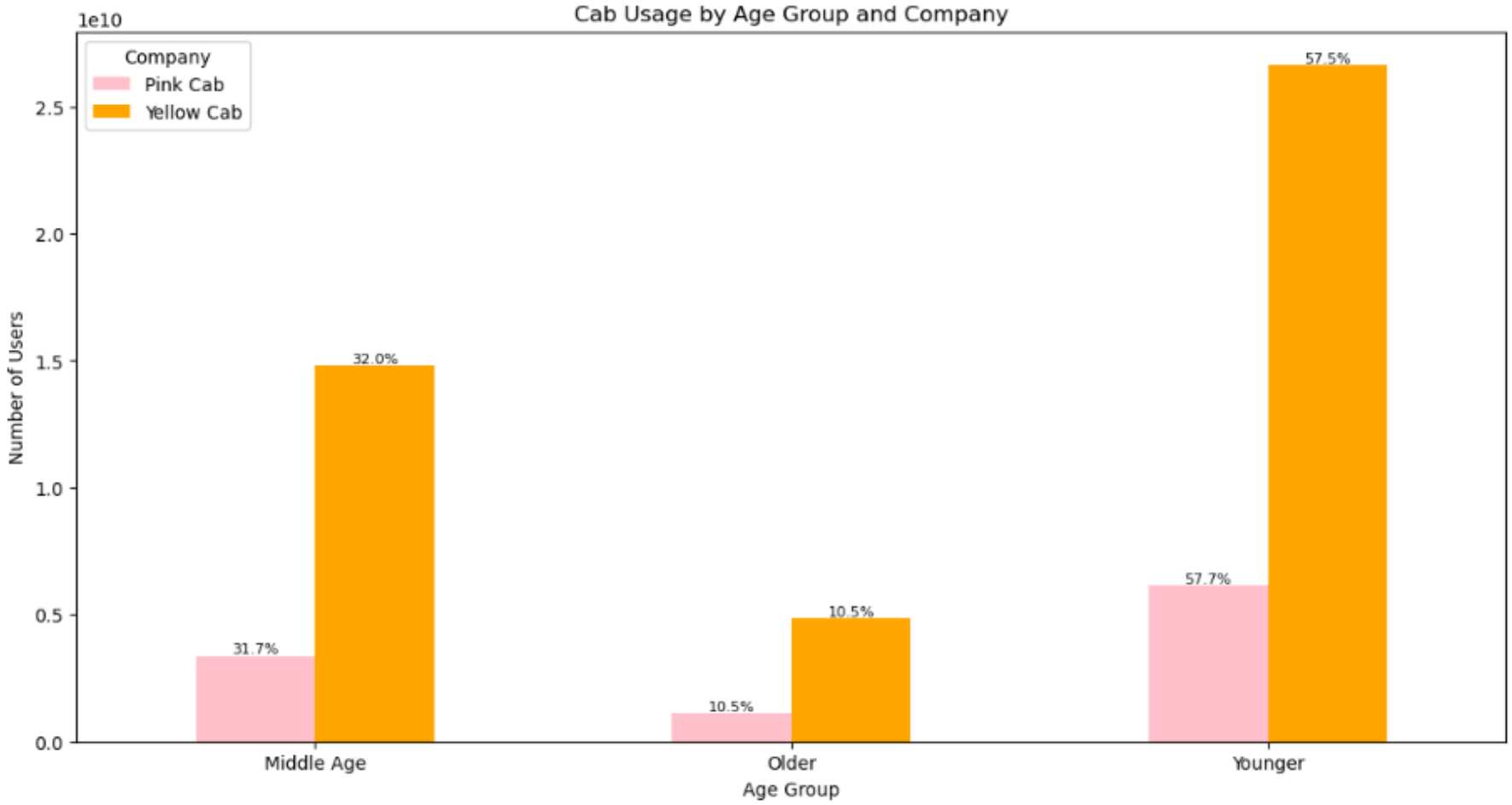
Yellow Cabs demonstrate greater cost-efficiency compared to Pink Cabs.

Cab Usage by Gender and Company



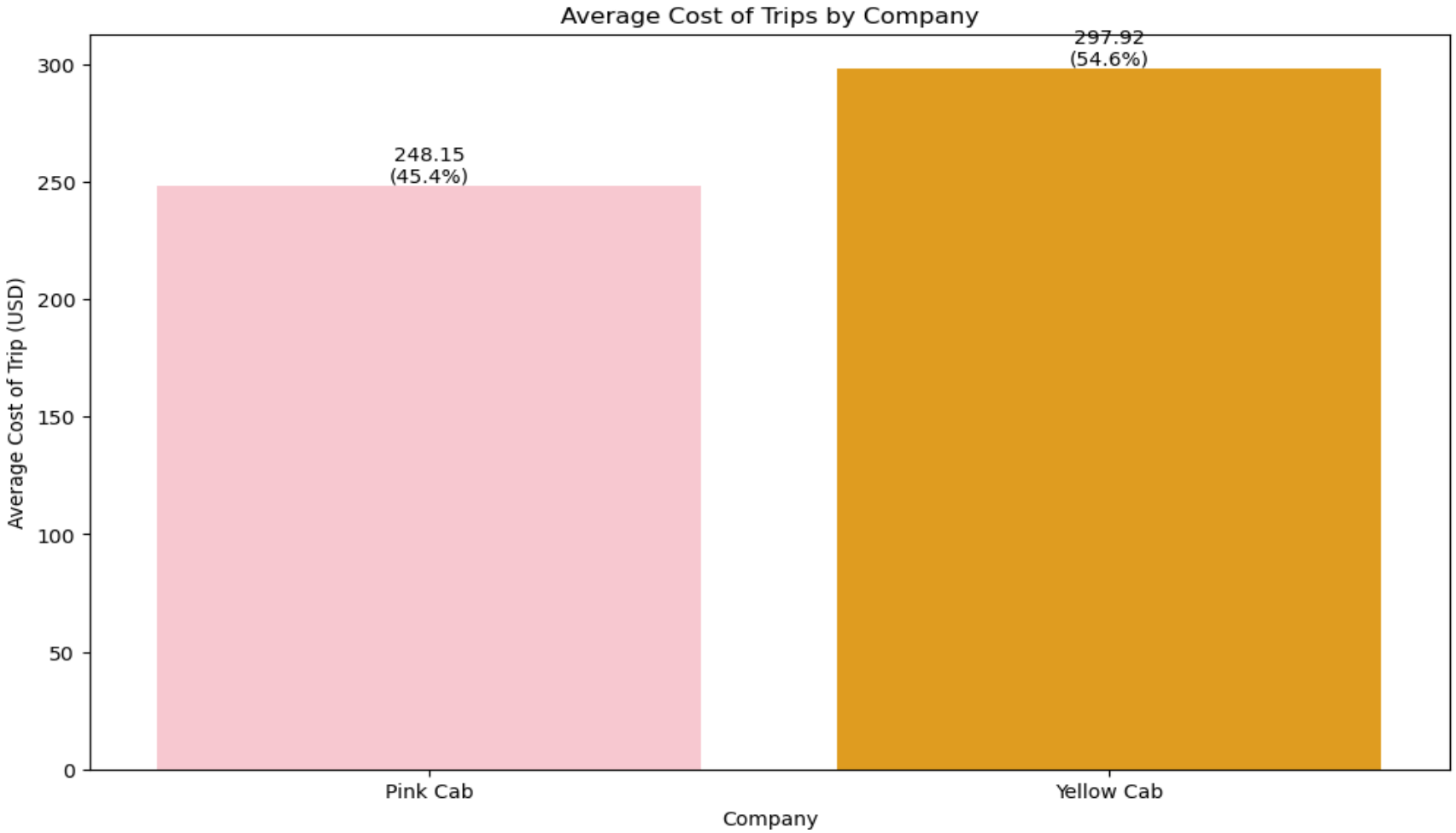
Based on the data, males utilize cab services more frequently than females. Additionally, Yellow Cabs once again outperform Pink Cabs in terms of usage.

Cab Usage by Age Group and Company



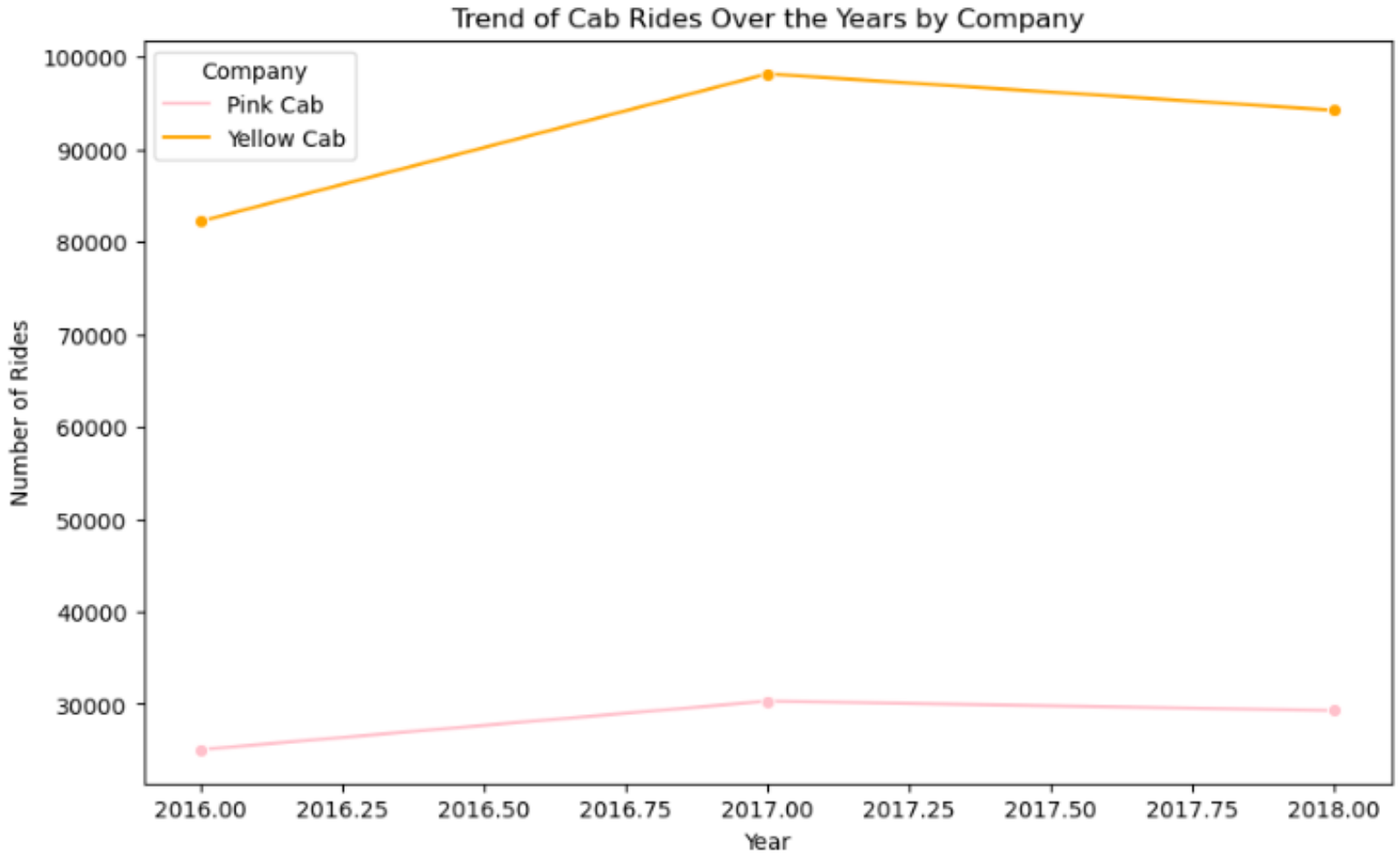
Younger age groups utilize cab services more frequently than older age groups and tend to favour Yellow Cabs over Pink Cabs.

Average Cost of Trips by Company



The average cost of trips indicates that Yellow Cabs are more expensive on average compared to Pink Cabs.

Trend of Cab Rides Over the Years by Company



As indicated by the annual cab usage data, the trend in cab rides over the years shows that 2017 was the peak year for both companies.

Hypothesis Testing

Hypothesis

H1: Is there a significant difference in cab usage between males and females?

```
T-statistic: 36.522236572546824
P-value: 1.8983851995043967e-291
There is a significant difference in cab usage between males and females.
```

H2: Does Yellow Cab have a significantly higher average profit margin compared to Pink Cab?

```
T-statistic: 230.99551452746311
P-value: 0.0
Yellow Cab has a significantly higher average profit margin compared to Pink Cab.
```

H3: Does revenue generation differ between high-income and low-income earners for Yellow Cab?

```
H-statistic: 14917.81734623754
P-value: 0.0
There is a significant difference in revenue generation between high-income and low-income earners across both companies.
```

Hypothesis

H4: How does the average trip cost compare between Yellow Cab and Pink Cab?

```
T-statistic: 89.01971247300362
P-value: 0.0
The average trip cost is significantly higher for Yellow Cab compared to Pink Cab.
```

H5: Is cab usage uniformly distributed across different age groups?

```
Chi-Square Statistic: 18950752960.84848
P-value: 0.0
The distribution of cab usage across different age groups is not uniform.
```

H6: Does the choice of payment mode affect cab usage frequency?

```
T-statistic: -1.0242536163472624
P-value: 0.30571630387777504
There is no significant difference in cab usage between different payment modes.
```

Hypothesis

H7: Is there a significant difference in cab usage between males and females?

```
T-statistic: 126.47873145242187
P-value: 0.0
The average revenue generated per ride is significantly higher in large metropolitan areas compared to smaller cities.
```

Recommendations

Based on the comprehensive exploratory data analysis (EDA), the following insights and recommendations favour investing in Yellow Cab:

Market Share and Usage: Yellow Cab consistently demonstrates a dominant market share with higher annual usage compared to Pink Cab. This is evident from the data showing Yellow Cab's significant share in overall rides and revenue across large metropolitan areas.

Profitability: Yellow Cab exhibits a notably higher average profit margin of 30.1% compared to Pink Cab's 17.4%. This superior profit margin indicates better financial performance and efficiency in generating profits.

Customer Retention: Yellow Cab has a higher proportion of repeat customers (83.1%) compared to Pink Cab (16.9%). This suggests stronger customer loyalty and satisfaction with Yellow Cab’s services.

Recommendations

Demographic and Usage Trends: Younger age groups and high-income earners show a clear preference for Yellow Cab, indicating a robust customer base with potential for future growth.

Performance Trends: Historical data indicates that Yellow Cab has maintained its leading position consistently over the years, with 2017 being a peak year for both companies but Yellow Cab showing stronger performance.

Revenue Contribution: Analysis of revenue contribution by income category shows that high-income earners contribute significantly more to Yellow Cab’s revenue. This reflects Yellow Cab's strong appeal to affluent customer segments who may also be more likely to use cab services regularly

Recommendations

Payment Preferences: Yellow Cab benefits from a higher proportion of card payments, suggesting a more modern and potentially higher-value customer base compared to Pink Cab, which may influence overall financial stability and transaction efficiency.

Cost of Trips: The average cost of trips for Yellow Cab is higher, indicating that it caters to a premium segment of the market. This premium positioning can lead to higher revenue per ride and suggests Yellow Cab's ability to command better pricing.

Operational Consistency: Yellow Cab shows more consistent performance across different metrics and time periods. Its ability to maintain high usage rates and profitability even during peak periods underscores its operational strength and market resilience.

Recommendations

Cost Efficiency: Yellow Cab is more cost-efficient, managing operational costs effectively while maintaining higher revenue and profit margins.

Thank You