Exploratory Data Analysis

G2M Insight For Cab Investment Firm

John Nguyen

21 June 2024

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

Executive Summary

- Brief overview of the key points and findings

Problem Statement

XYZ is looking to invest in the Cab Industry due to its rapid growth in recent years.

Following their Go-to-Market(G2M) business strategy, they want to understand the market before taking the final decision.

The goal of the analysis is to provide company XYZ with actionable insight to identify the right company to make an investment.

Approach

Four datasets from January 31, 2016, to December 31, 2018, containing cab transaction details, customer demographics, transaction mappings, and city statistics were explored and analyzed.

The data was merged, processed, and compared, with insights visualized through various charts and graphs to aid in decision-making.

EDA (Exploratory Data Analysis)

Exploratory Data Analysis (EDA) refers to the process of analyzing datasets to summarize the main characteristics using visual methods.
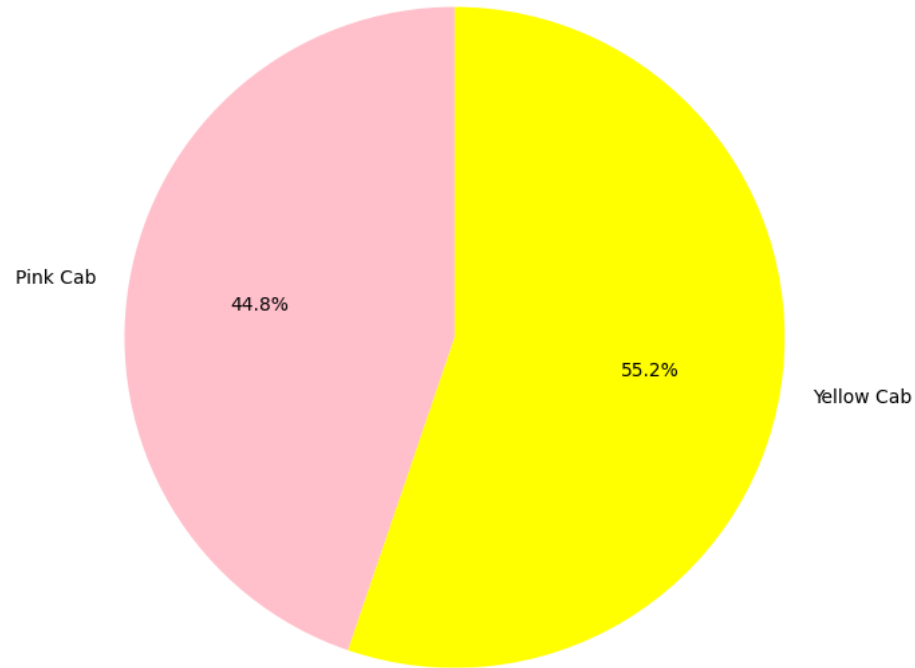
The following slides will look at an in-depth analysis of various aspects of cab service performance.
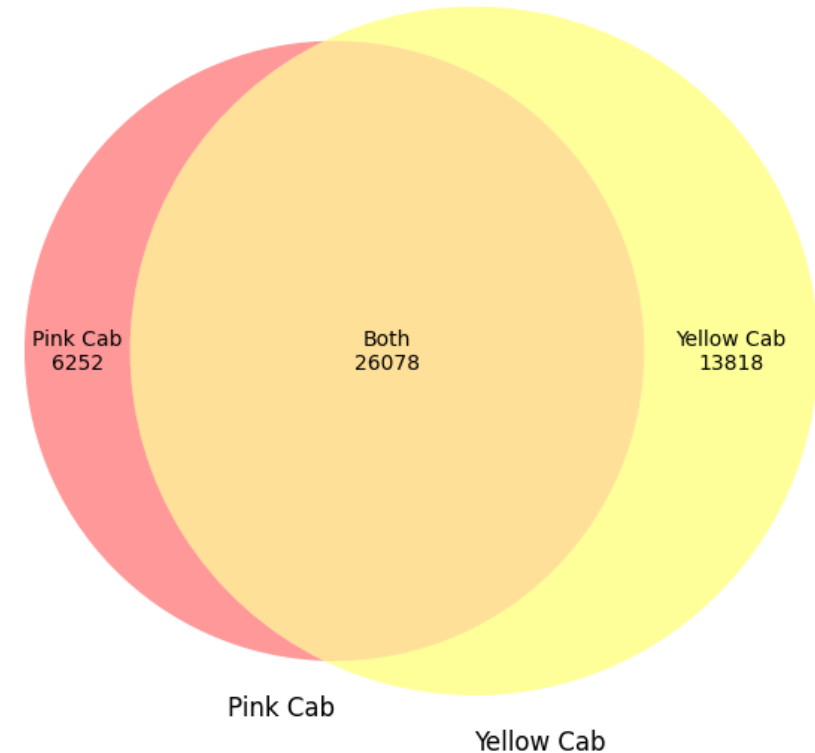
Total Users and Overlapping Users

- Yellow Cab has 55.2% of the market, while Pink Cab holds 44.8%.
- This indicates that Yellow Cab has a slightly larger customer base overall.

Total Users Comparison by Company

Pink Cab
44.8%

55.2%
Yellow Cab

Overlapping Customers Between Pink Cab and Yellow Cab

Pink Cab
6252

Both
26078

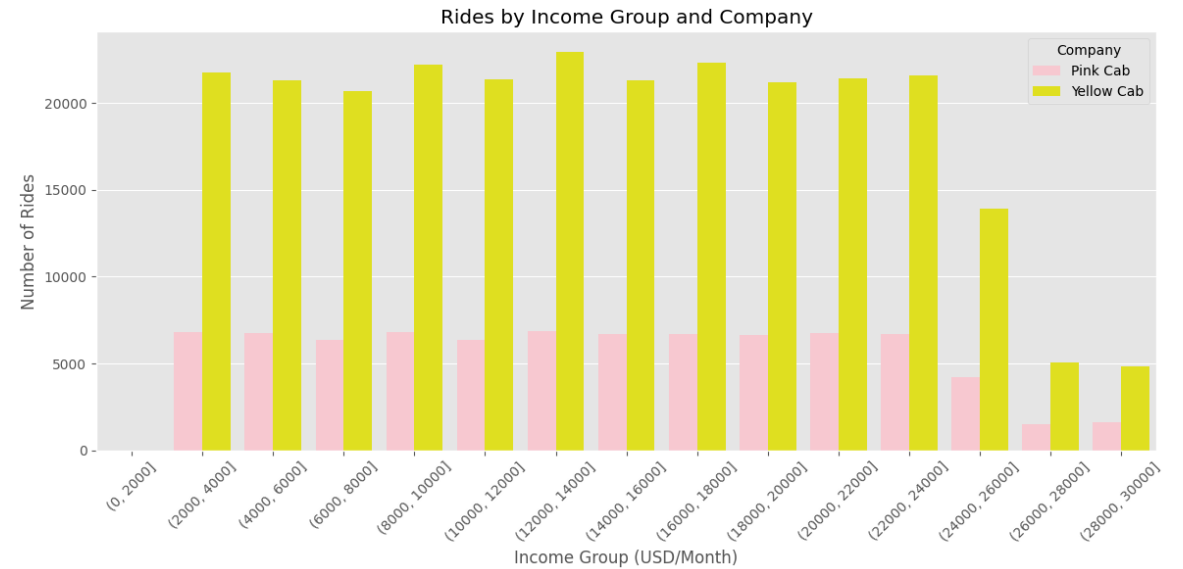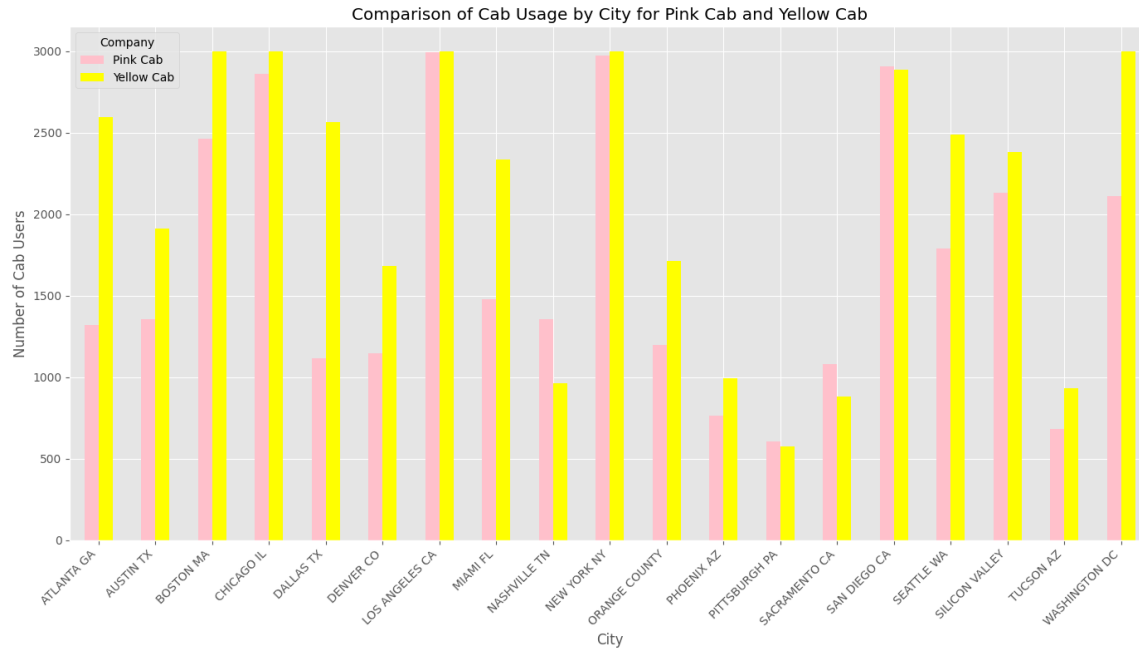Yellow Cab
13818

Pink Cab

Yellow Cab

Revenue Analysis

- Yellow Cab generates higher revenue per kilometer than Pink Cab, indicating more efficient pricing or higher value trips.

- Both companies show a wide range of revenue for longer distances, suggesting variability in pricing or trip types.
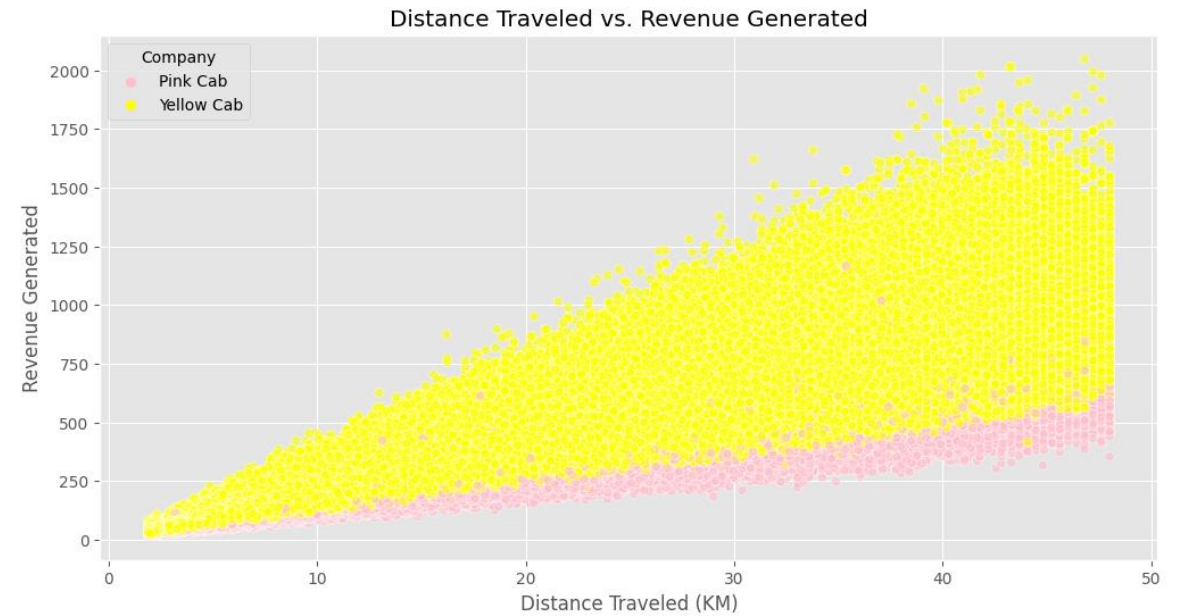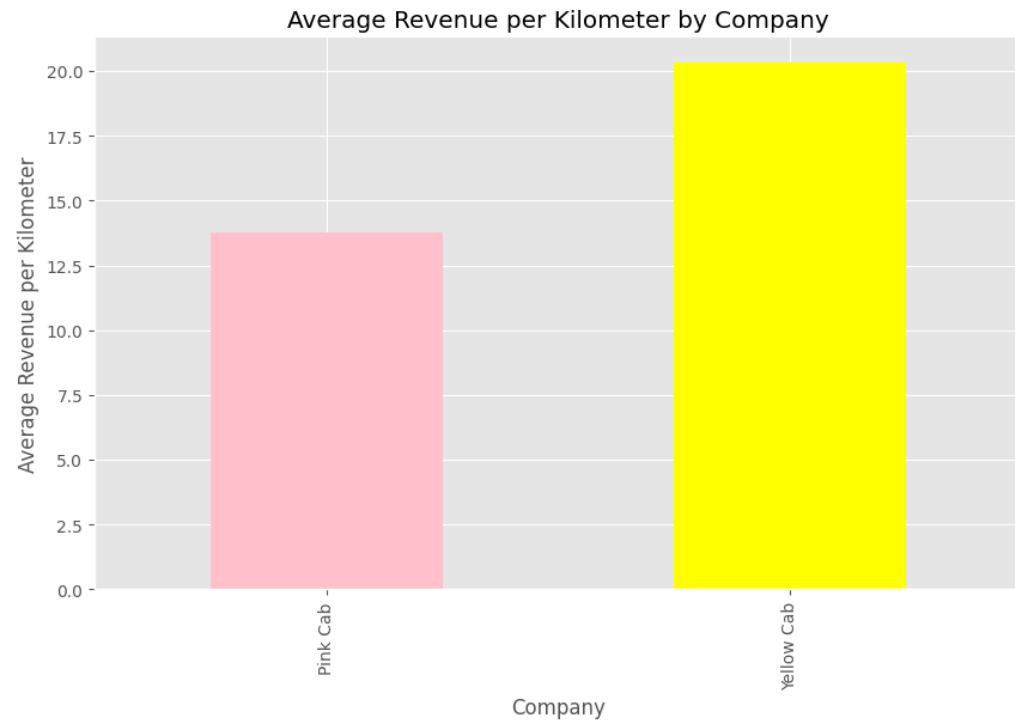
User Distribution and Demographics

- No significant difference in user distribution per city between Yellow Cab and Pink Cab.

- Similar income distribution for users of both Yellow Cab and Pink Cab.

- Both companies have similar market penetration in urban areas and appeal to users across various income levels.

Comparison of Cab Usage by City for Pink Cab and Yellow Cab

Rides by Income Group and Company

User Distribution and Demographics

- No significant preference for either cab company within specific age groups.

- Gender distribution is similar between Yellow Cab and Pink Cab, indicating comparable appeal across genders.

- Both companies effectively serve all age groups and have a higher usage among females.

Number of Customers by Age Group and Company


Rides by Gender and Company
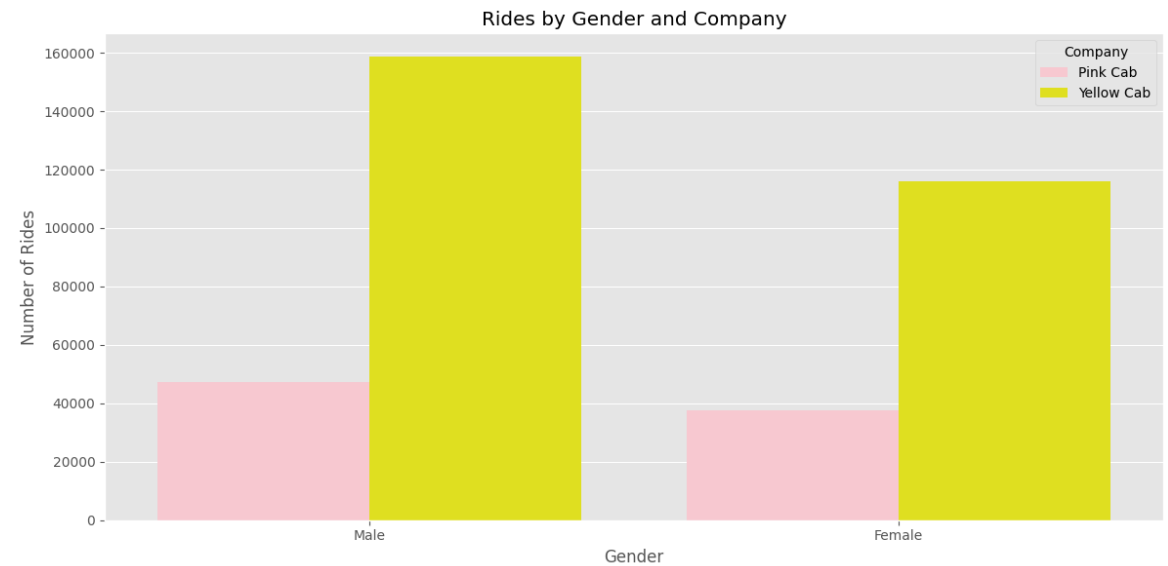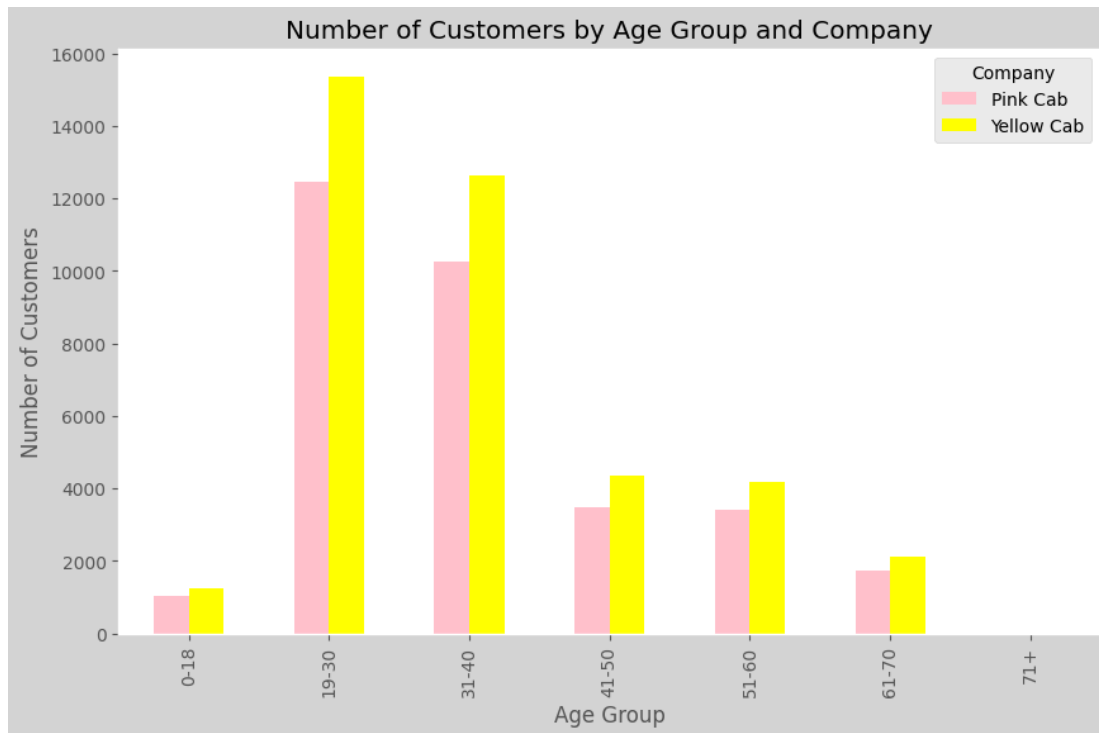
Repeated Customers to Total Revenue

- Yellow Cab has a higher proportion of repeat customers contributing to its total revenue.

Revenue Contribution: Repeat vs. One-Time Customers - Pink Cab

One-Time Customers
16.3%
83.7%
Repeat Customers
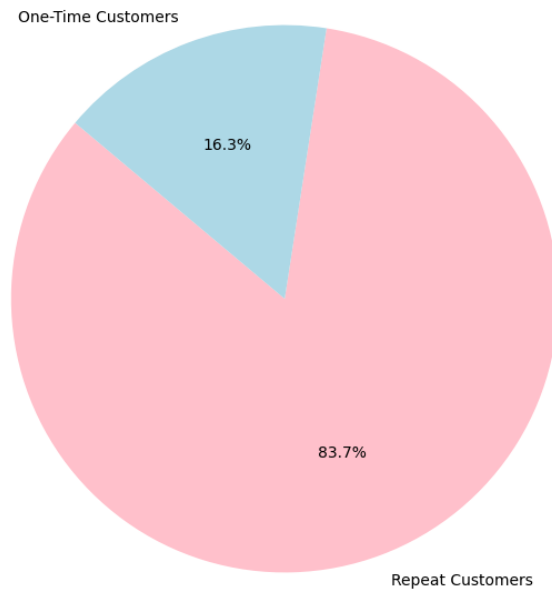
Revenue Contribution: Repeat vs. One-Time Customers - Yellow Cab

One-Time Customers
4.1%
95.9%
Repeat Customers
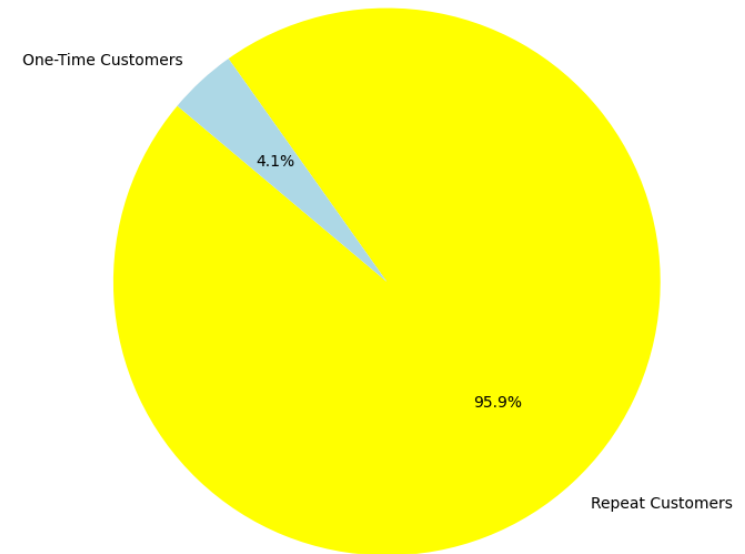
Time Series Analysis

- The trends in monthly cab usage for Yellow Cab and Pink Cab are similar.

- Peaks in one company's user base are mirrored by peaks in the other's.

- The similar trends suggest that external factors (e.g., seasonality, events) impact both companies equally.

Unique Users per Month by Company

Monthly Cab Usage by Company
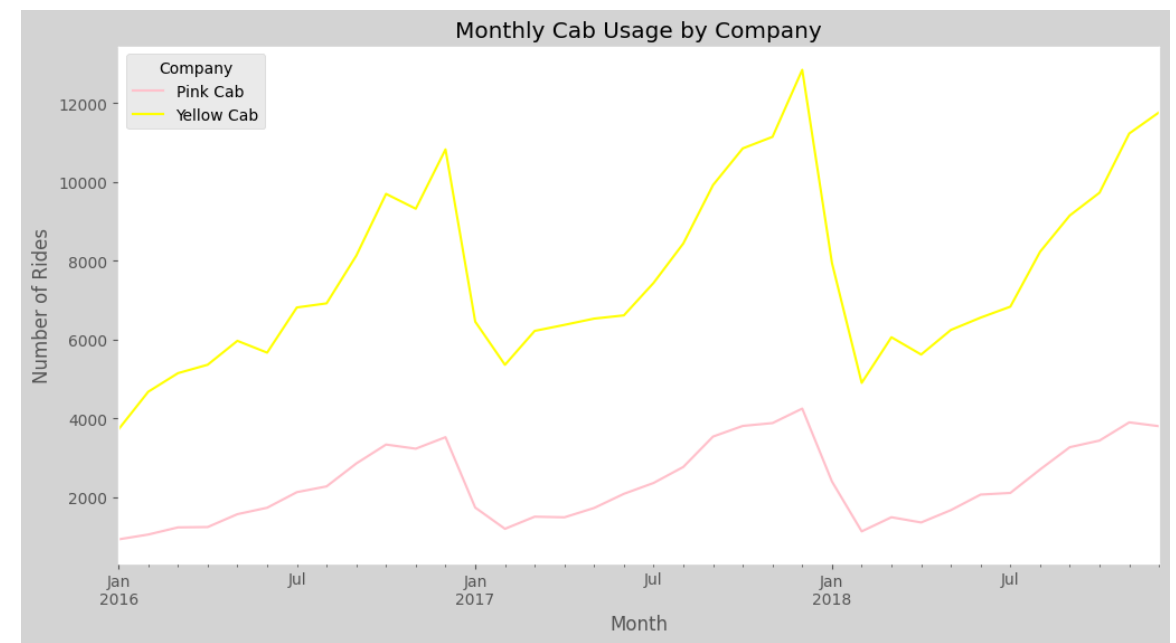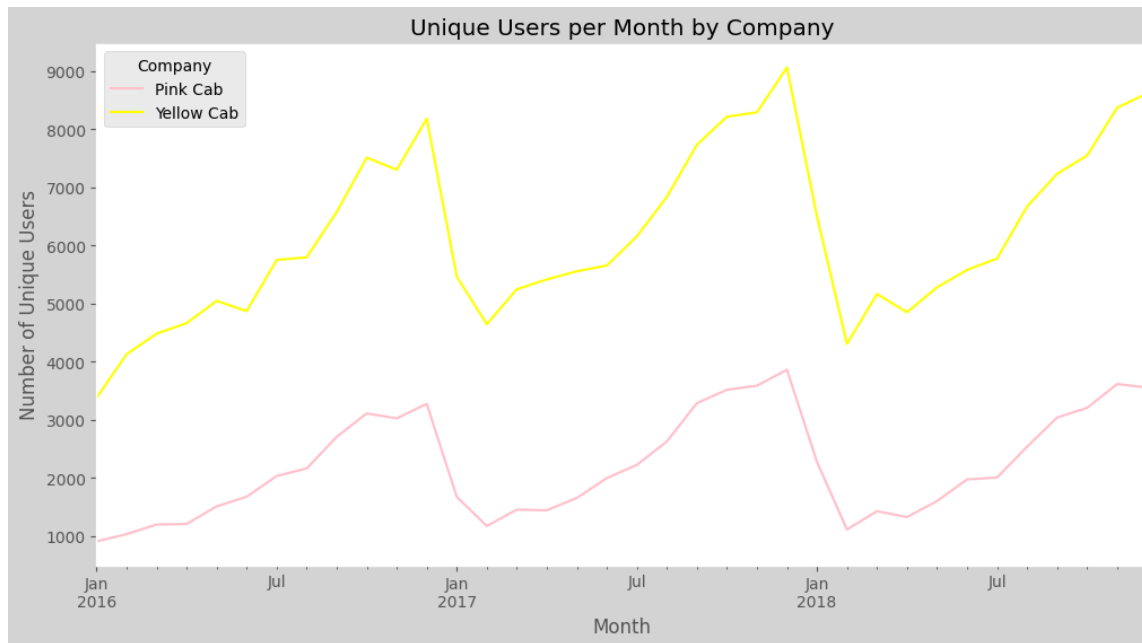
Total User vs. Total Revenue

- Yellow Cab has a higher number of users compared to Pink Cab, with a ratio of approximately 0.8:1.

Total Revenue Comparison by Company

Pink Cab 17.3%

Yellow Cab 82.7%

Total Users Comparison by Company

Pink Cab 44.8%
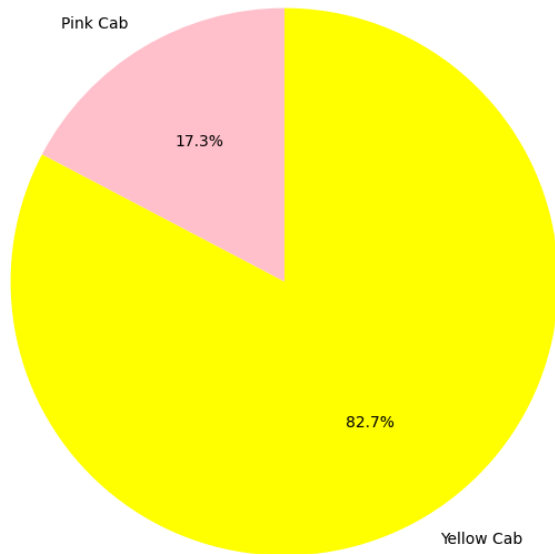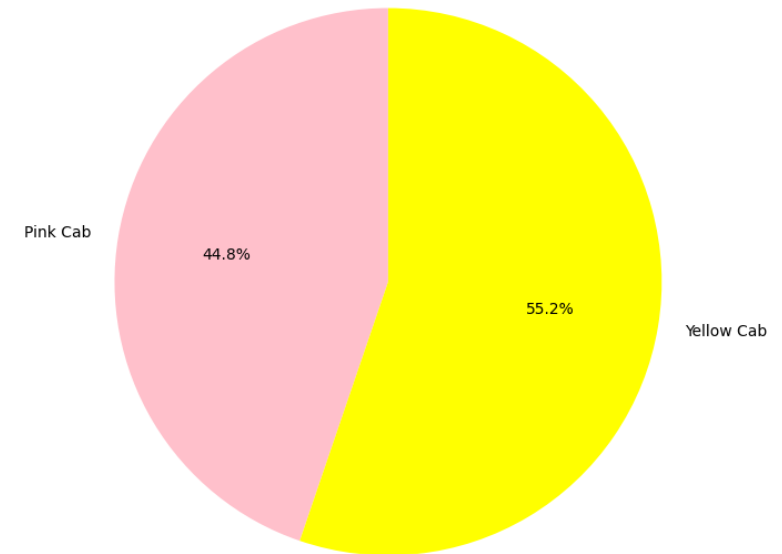
Yellow Cab 55.2%

Total Revenue Comparison by Company

Total Rides by Company
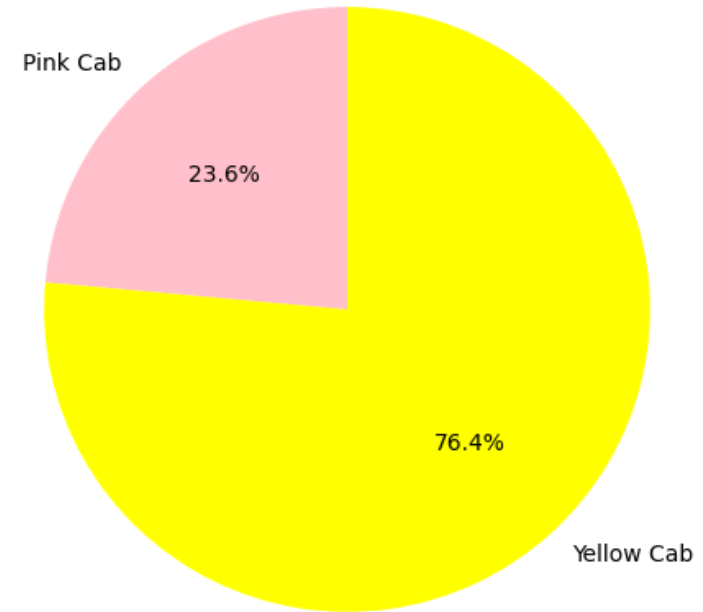
Total Revenue vs Total Rides

- Yellow Cab leads in both total revenue and total rides, indicating a strong market position in terms of both volume and revenue generation.

Hypothesis Testing

Impact of Distance on Revenue

Cab Usage and City Population

Certain demographic groups (age, gender, income) influence on cab services usage.

Contribution of Repeat Customers to Total Revenue

Age Group Preferences for Cab Companies

Impact of Distance on Revenue

H0 (Null Hypothesis): There is no difference in the average revenue per kilometer between Yellow Cab and Pink Cab.

H1 (Alternative Hypothesis): Yellow Cab has a higher average revenue per kilometer compared to Pink Cab.

Hypothesis 1: t-statistic = -320.9807762543478, p-value = 0.0

Summary of results here

Cab Usage and City Population

H0 (Null Hypothesis): There is no difference in the number of users per city between Yellow Cab and Pink Cab.

H1 (Alternative Hypothesis): Yellow Cab has more users per city compared to Pink Cab in the majority of cities, indicating stronger market penetration in larger urban areas.

Hypothesis 2: t-statistic = -1.4634295900210088, p-value = 0.1520272056721823

Summary of results here

Certain demographic groups (age, gender, income) influence on cab services usage.

H0 (Null Hypothesis): There is no difference in cab usage based on demographic groups (age, gender, income) between Yellow Cab and Pink Cab.

H1 (Alternative Hypothesis): There are significant differences in cab usage based on demographic groups, with more females using cab services compared to males, and similar income and age distributions for both companies.

Hypothesis 3 (Gender): chi2 = 107.22063897254299, p-value = 3.982674650131372e-25

Hypothesis 3 (Age): t-statistic = -0.3777700356771092, p-value = 0.7056016582376317

Hypothesis 3 (Income): t-statistic = 0.42711269788899975, p-value = 0.6692975005750657

Summary of results here

Contribution of Repeat Customers to Total Revenue

H0 (Null Hypothesis): There is no difference in the contribution of repeat customers to total revenue between Yellow Cab and Pink Cab.

H1 (Alternative Hypothesis): Yellow Cab has a higher proportion of repeat customers contributing to its total revenue compared to Pink Cab. Additionally, there is a notable segment of customers who use both cab services.

z-statistic = -39.3617148837474, p-value = 0.0

Summary of Results here

Age Group Preferences for Cab Companies

H0 (Null Hypothesis): There is no difference in age group preferences for either cab company.

H1 (Alternative Hypothesis): The age group analysis suggests no significant preference for either cab company within specific age groups, indicating that both Pink Cab and Yellow Cab serve customers across all age ranges fairly evenly.

chi2 = 18.95665910841074, p-value = 0.0019581936565151313

Summary of Results here

EDA Summary

- Summarize the findings from the EDA
- Link these findings to the business problems and objectives

Conclusion

Recommendations

- Provide actionable recommends
- Justify your recommendations
- Suggest next steps for XYZ's investment decision

Recommendations

- Provide actionable recommends
- Justify your recommendations
- Suggest next steps for XYZ's investment decision

Recommendations

- Provide actionable recommends
- Justify your recommendations
- Suggest next steps for XYZ's investment decision

Thank You