



Data Glacier

Your Deep Learning Partner

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



Mission Statement: Company XYZ aims to invest in the cab industry by identifying the most promising company based on market presence, revenue generation, customer loyalty, and overall usage.



Data Analysis Approach: Analyzed four datasets from January 2016 to December 2018, covering cab transaction details, customer demographics, transaction mappings, and city statistics.



Key Findings: Yellow Cab holds 55.2% of the market, generates higher revenue, has a higher proportion of repeating customers, but share the same market and follows similar usage trends with Pink Cab.



Overall Conclusion: Invest in Yellow Cab but leave room for Pink Cab's potential

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.

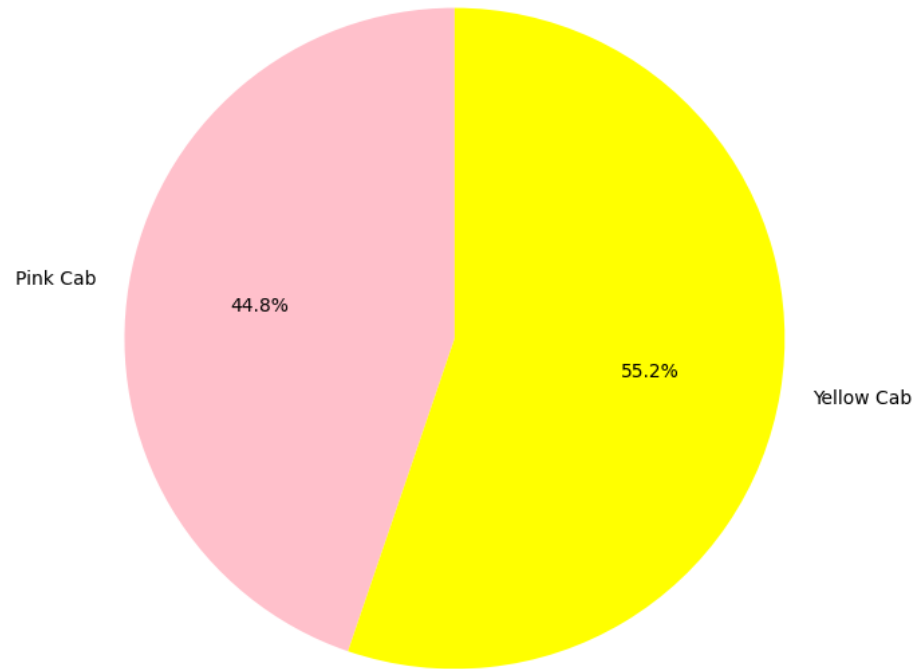


The following slides will look at an in-depth analysis of various aspects of cab service performance.

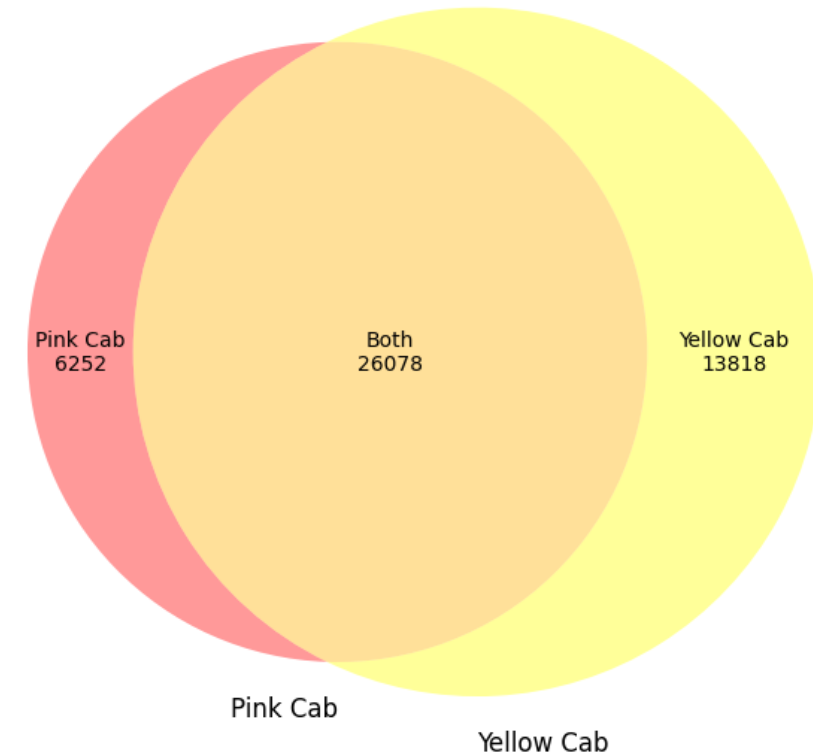
Total Users and Overlapping Users

- 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

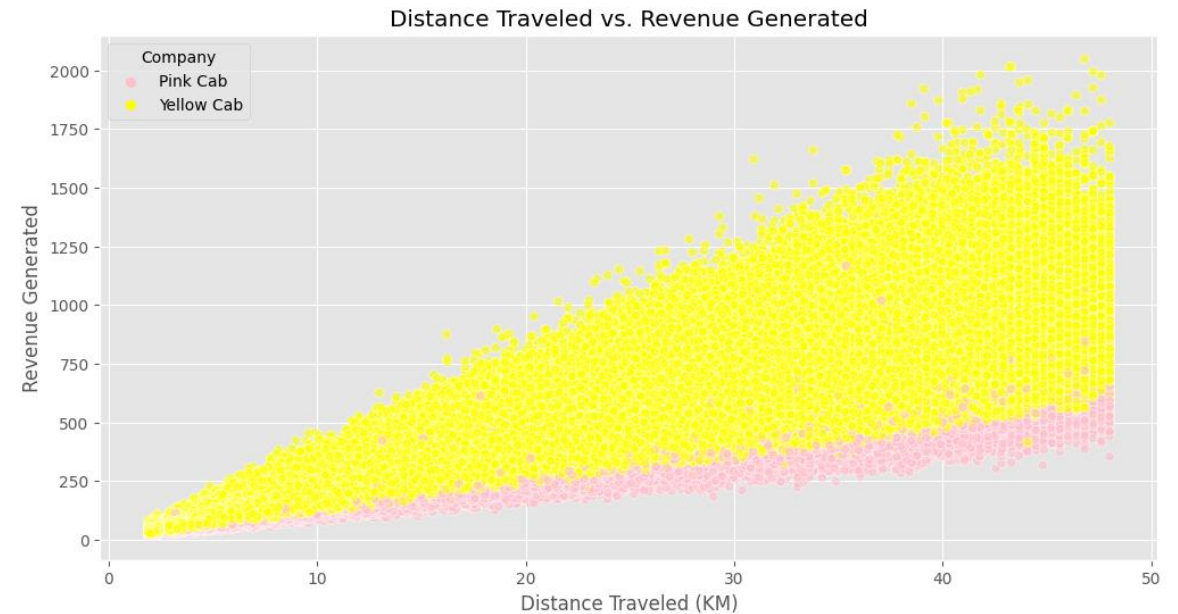
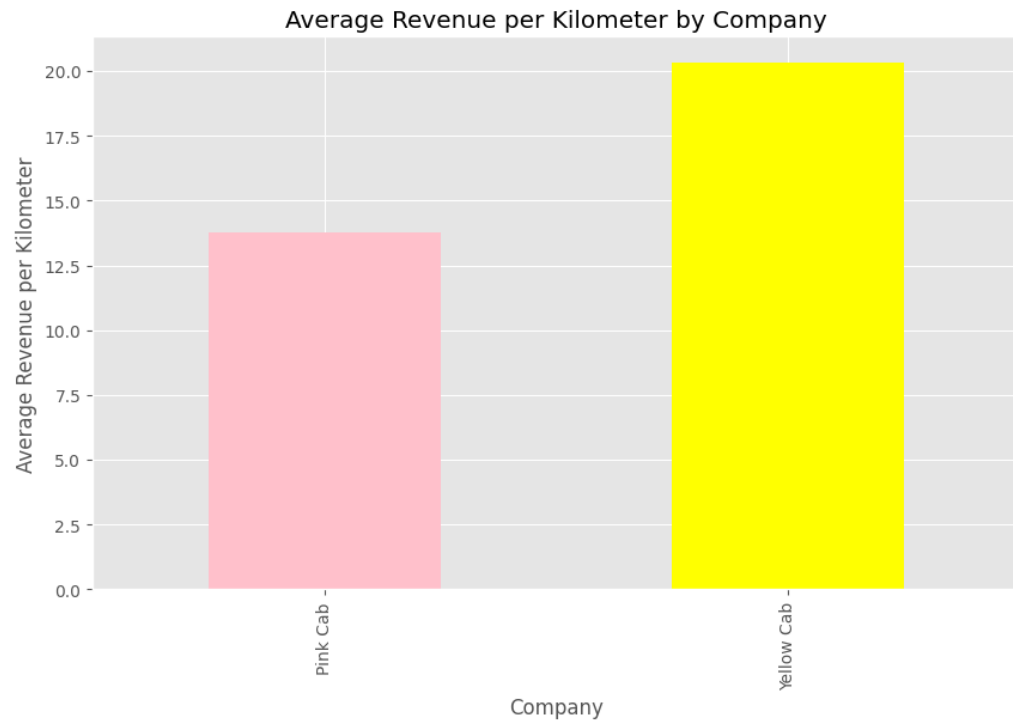


Overlapping Customers Between Pink Cab and Yellow Cab



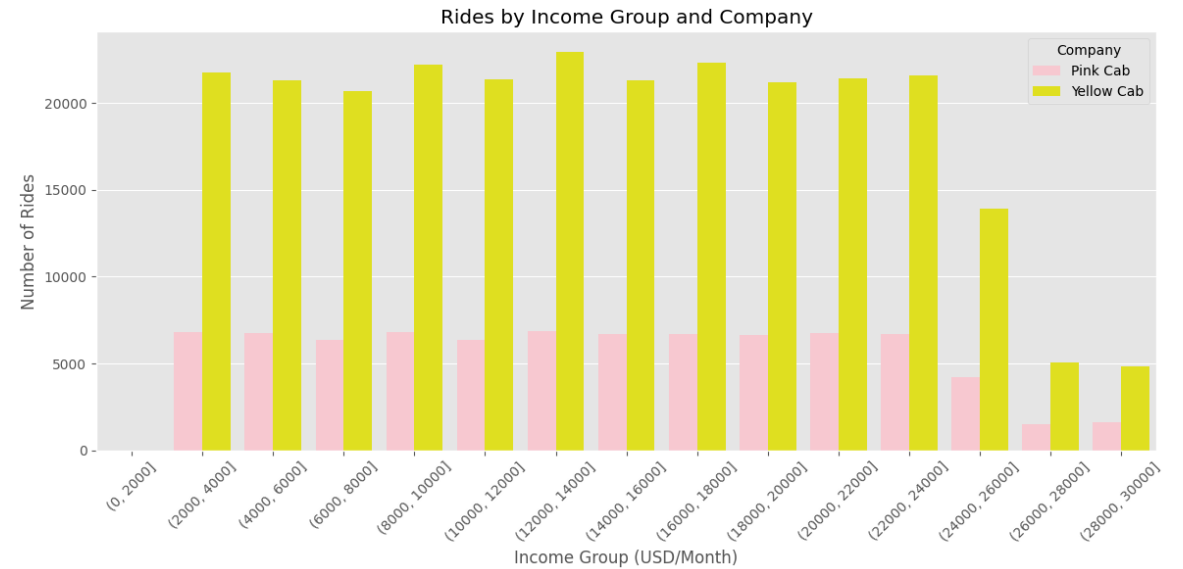
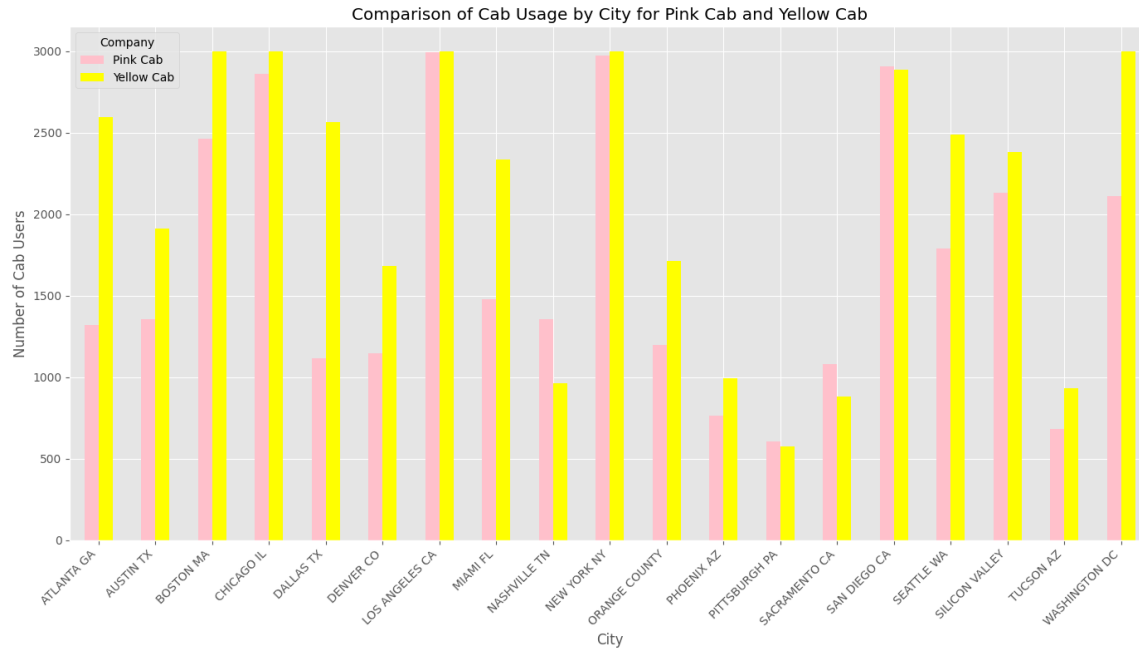
Revenue Analysis

- 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.



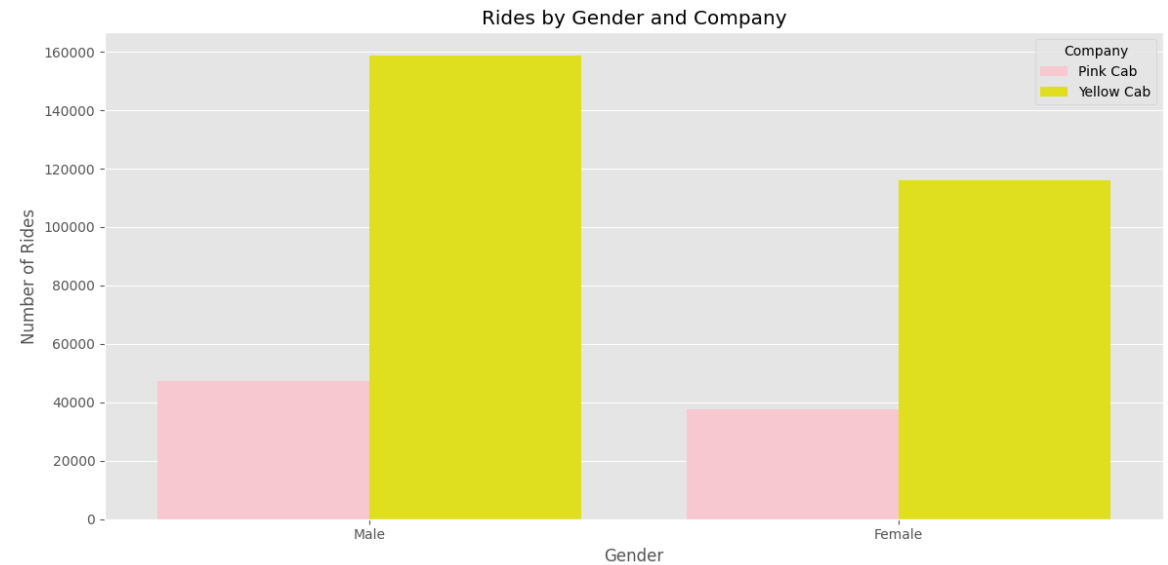
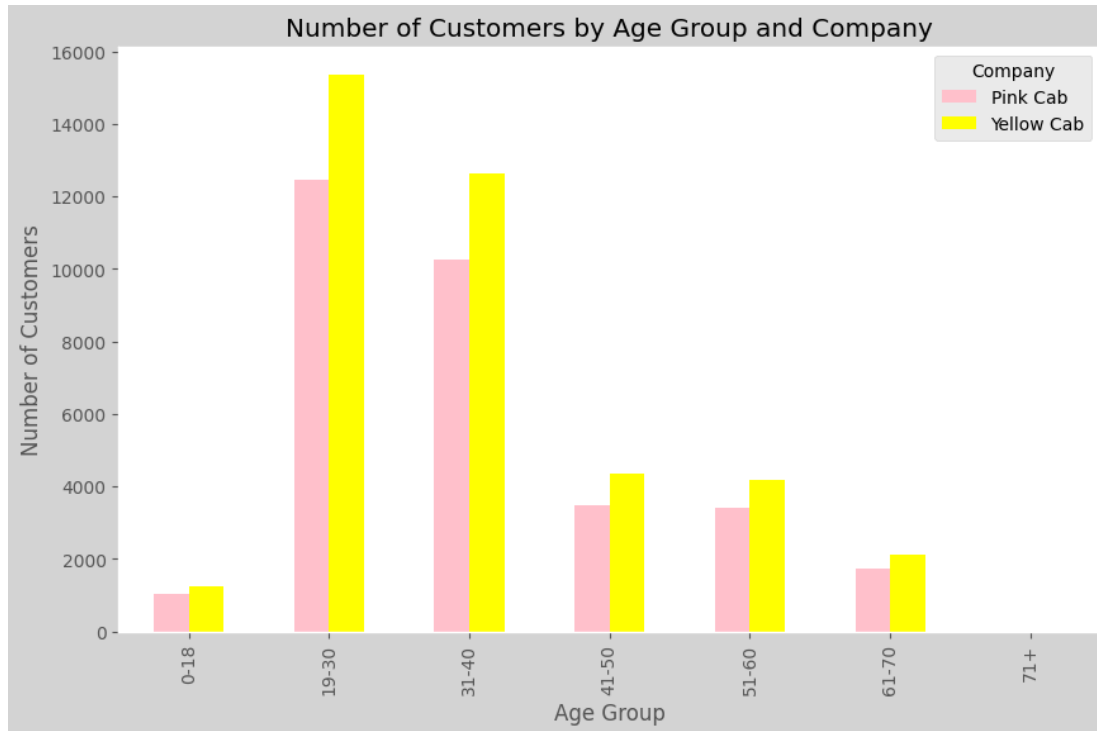
User Distribution and Demographics

- 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.



User Distribution and Demographics

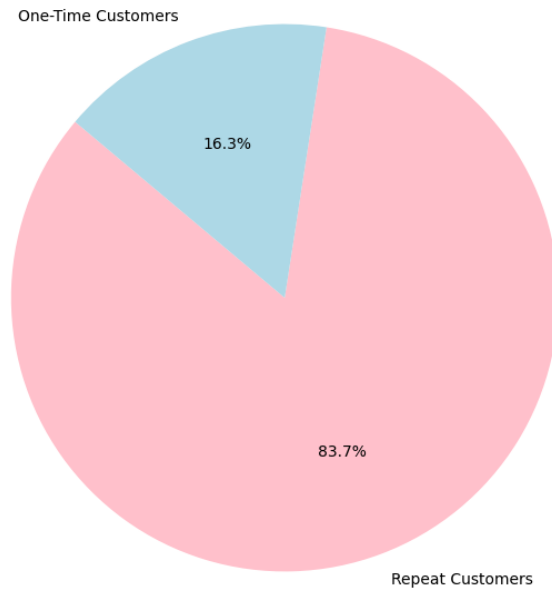
- No significant preference for either cab company within specific age groups.
- Gender distribution is similar between Yellow Cab and Pink Cab, indicating comparable appeal across genders.
- Both companies effectively serve all age groups and have a higher usage among females.



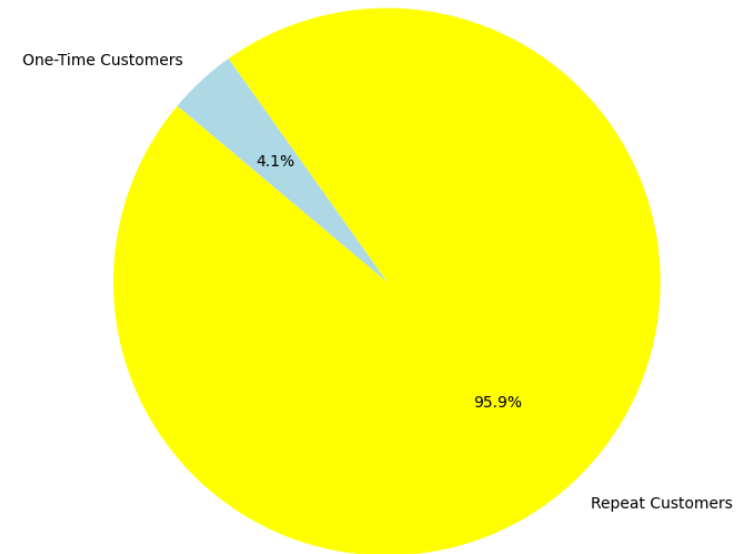
Repeated Customers to Total Revenue

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

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

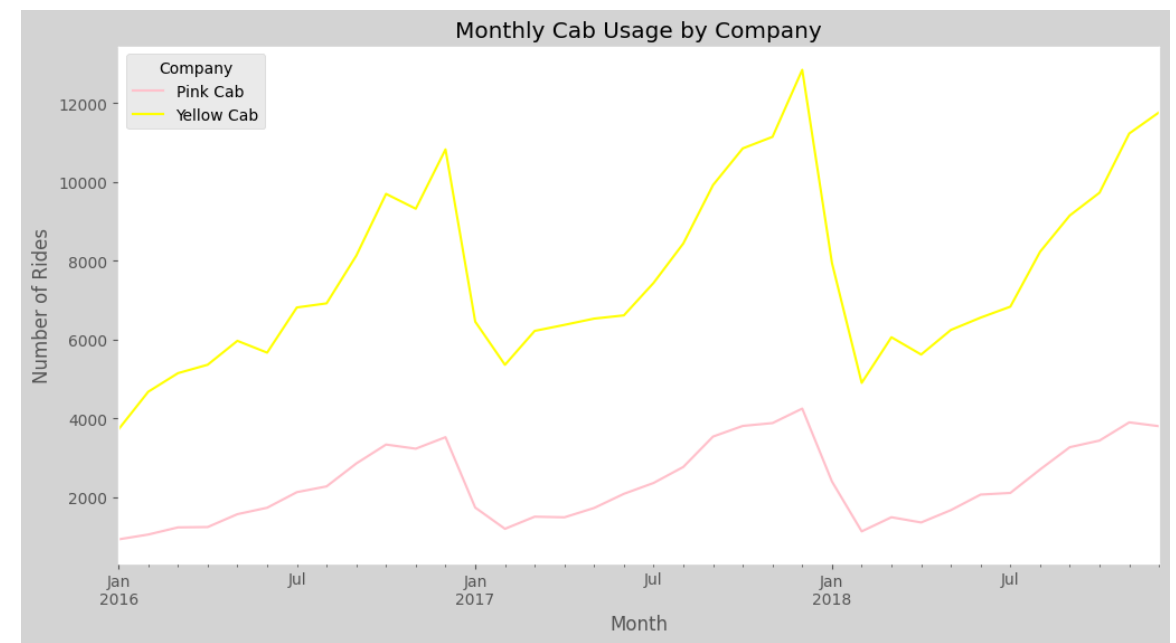
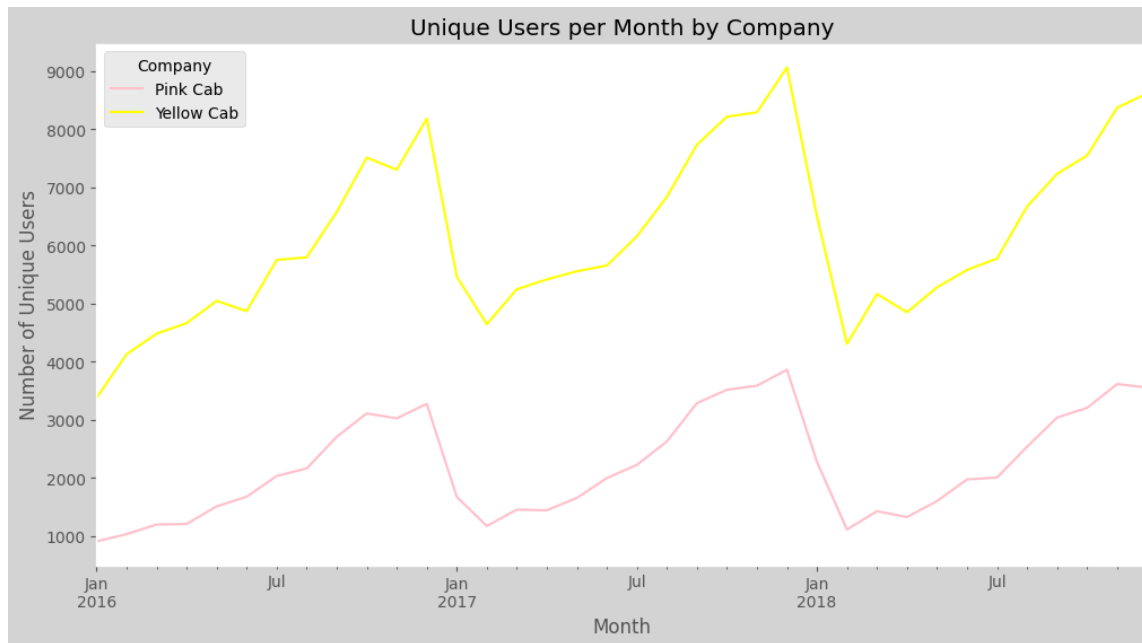


Revenue Contribution: Repeat vs. One-Time Customers - Yellow Cab



Time Series Analysis

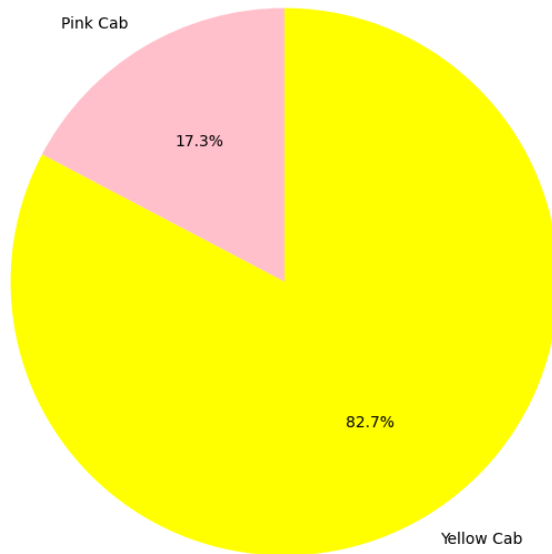
- The trends in monthly cab usage for Yellow Cab and Pink Cab are similar.
- Peaks in one company's user base are mirrored by peaks in the other's.
- The similar trends suggest that external factors (e.g., seasonality, events) impact both companies equally.



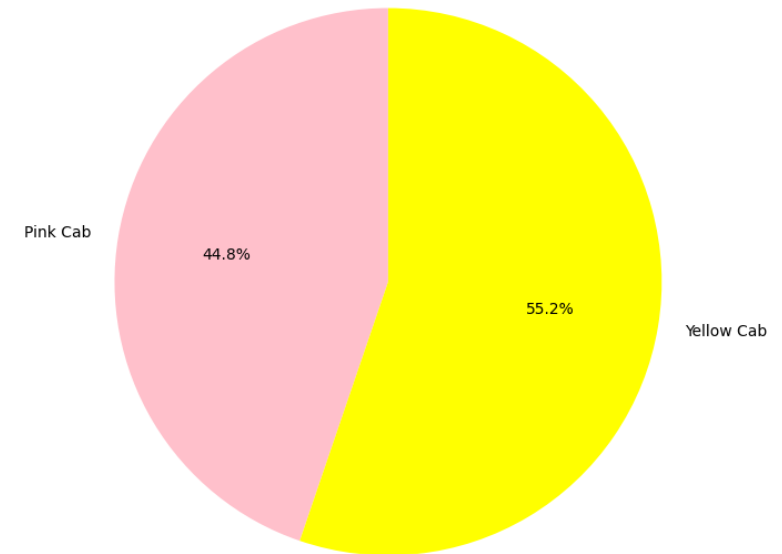
Total User vs. Total Revenue

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

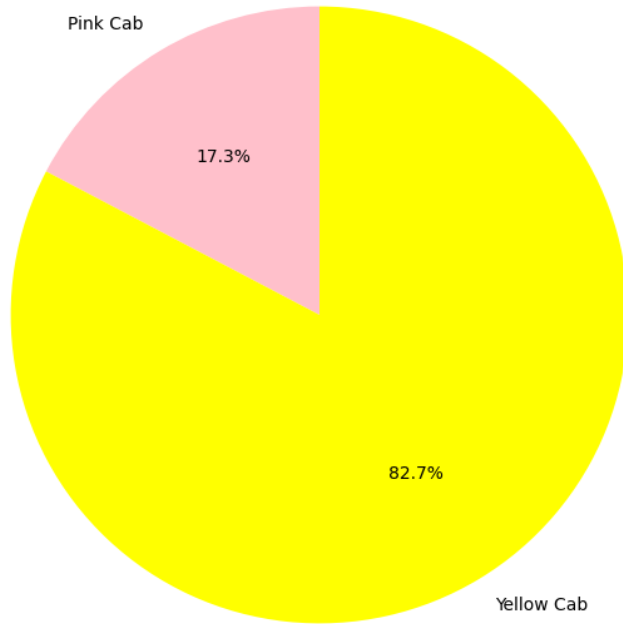
Total Revenue Comparison by Company



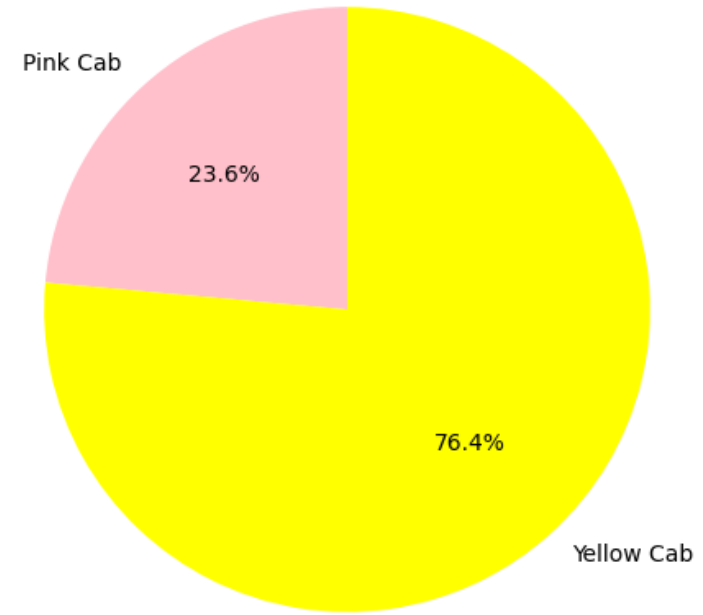
Total Users Comparison by Company



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.



t-statistic = -320.98
p-value = 0.0



Reject the null hypothesis. Yellow Cab generates significantly higher revenue per kilometer than Pink Cab, indicating more efficient pricing or higher value trips.

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.



t-statistic = -1.463
p-value = 0.152



Fail to reject the null hypothesis.
There is no significant difference in the number of users per city between Yellow Cab and Pink Cab.

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.



Gender: $\chi^2 = 107.2$, p-value = $3.98e-25$

Age: t-statistic = -0.378
p-value = 0.706

Income: t-statistic = 0.427, p-value = 0.6692975005750657



Gender: **Reject the null hypothesis.** There are significant differences in cab usage based on gender, with more females using cab services.

Age: **Fail to reject the null hypothesis.** There is no significant difference in cab usage based on age.

Income: **Fail to reject the null hypothesis.** There is no significant difference in cab usage based on income.

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 = -39.362
p-value = 0.0



Reject the null hypothesis. Yellow Cab has a significantly higher proportion of repeat customers contributing to its total revenue.

Age Group Preferences for Cab Companies



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



$\chi^2 = 18.957$
 $p\text{-value} = 0.002$



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.



Reject the null hypothesis. There is no significant preference for either cab company within specific age groups, indicating that both companies serve customers across all age ranges fairly evenly

EDA Summary

Yellow Cab generates significantly higher revenue per kilometer.

No significant difference in the number of users per city between Yellow Cab and Pink Cab.

Significant gender-based differences in cab usage, with more females using cab services.

No significant differences in cab usage based on age or income.

Higher contribution of repeat customers to Yellow Cab's total revenue.



Recommendations

- Invest in Yellow Cab
 - Market Share: 55.2% compared to 44.8%.
 - Revenue Generation: Higher revenue per kilometer.
 - Repeated Customer: Higher repeating customers.
 - Ride Count & Usage: Higher total revenue & total rides.
- Pink Cab has potential...



Recommendations

- Invest in Yellow Cab but leave room for Pink Cab's potential
 - Significant User Base: Pink Cab still holds a significant market share with 44.8% , although less than Yellow Cab, still a significant number due to its low revenue.
 - Similar Demographics: Share the same market as Yellow Cab.
 - Repeating Customers: Although slightly behind Yellow Cab, Pink Cab has a notable proportion of repeat customers.
 - Overall Usage: The monthly usage trends for Pink Cab mirror those of Yellow Cab.

Thank You