Exploratory Data Analysis

G2M Insight For Cab Investment

16/7/2024

Munirah Alfehaid

Agenda

Executive Summary

Approach

EDA Analysis

Hypothesis Test

Recommendation

Executive Summary

Client:

XYZ is a private firm in the US. Given the significant growth in the Cab Industry over the past few years and the presence of numerous key players in the market, the company is considering an investment in the cab industry. As part of their Go-to-Market (G2M) strategy, they want to thoroughly understand the market before making a final decision.

Objective:

XYZ seeks your actionable insights to help them determine the best company for their investment.

Approach

Data Exploration:

The dataset contains a total of 359,352 rows. There are 20 provided features. The data covers the period from 2016 to 2018.

Assumptions:

-Data analysis was conducted for both companies under the assumption that there may be external noise beyond the provided data.

-Profit calculations were based solely on the difference between the Price Charged and the Cost of Trip.

-Data analysis was conducted for both companies assuming the dataset is time-constrained to the years 2016 to 2018.

-Only cash and card payments were considered.

Data Glacier
Your Deep Learning Partner

EDA Analysis

Analysis of Seasonality in Number of Customers by Company (Monthly):
_____

Consistent Seasonal Pattern: Both Pink Cab and Yellow Cab exhibit a clear seasonal pattern in the number of customers throughout the year.

Peak in Q4: The number of customers peaks around the fourth quarter (October-December) for both companies. This suggests increased demand during the holiday season or year-end festivities.

Trough in Q2: Both companies experience a dip in customer numbers during the second quarter (April-June). This could be attributed to factors like summer vacations or lower overall activity levels.

Relative Performance: While both companies follow a similar seasonal trend, Yellow Cab consistently has a higher number of customers compared to Pink Cab throughout the year.
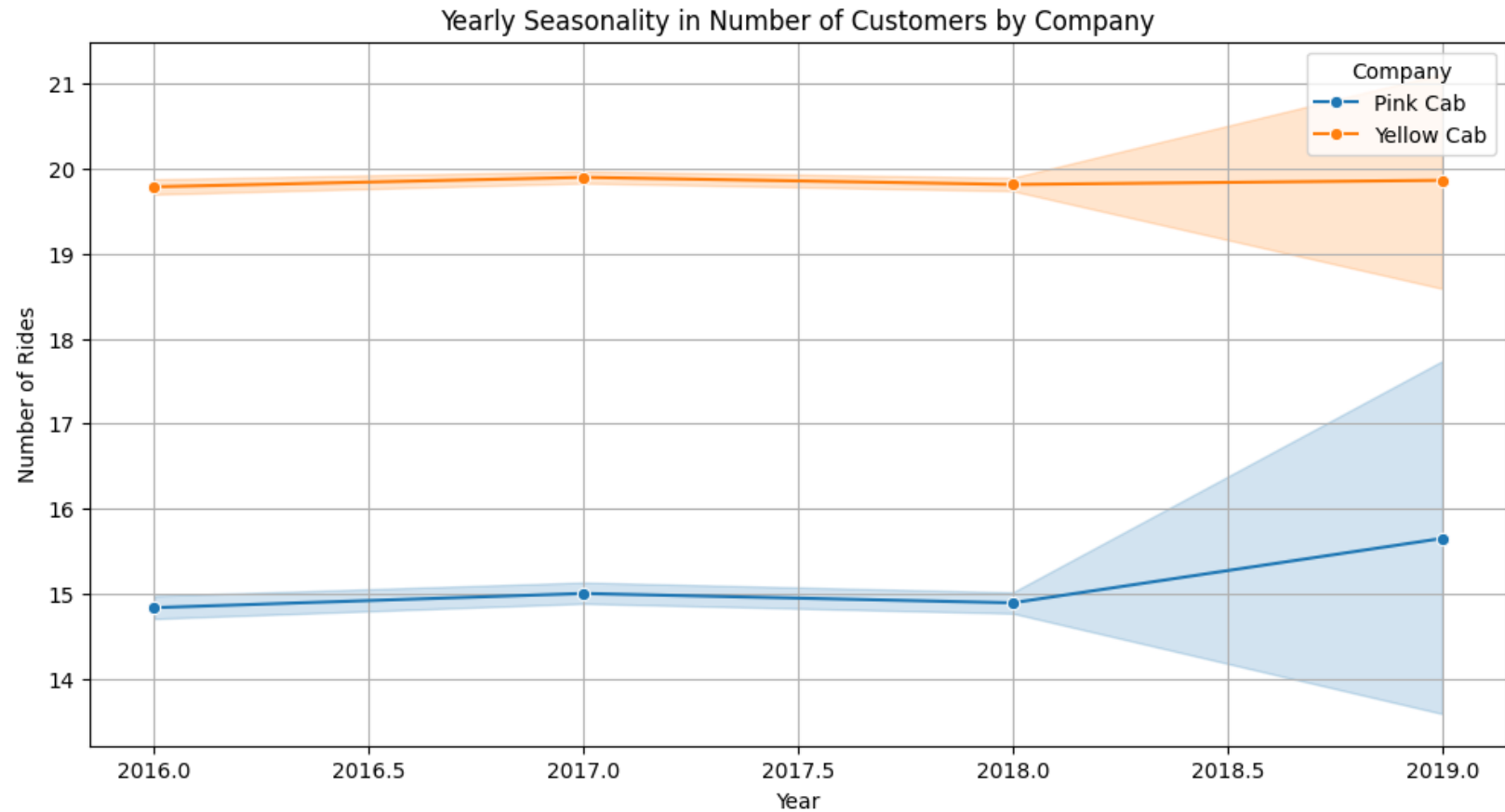

Monthly Seasonality in Number of Customers by Company

Analysis of Seasonality in Number of Customers by Company (Yearly):
_____

Consistent Seasonality: Both Pink Cab and Yellow Cab exhibit a clear yearly seasonal trend in the number of customers.

Yellow Cab Dominance: Throughout the observed years, Yellow Cab consistently maintains a higher number of customers compared to Pink Cab.

Slight Decline: There seems to be a slight downward trend in the number of customers for both companies over the years, though it's not very pronounced.

Similar Fluctuations: Both companies experience similar fluctuations in customer numbers from year to year, suggesting they might be influenced by the same external factors.


Yearly Seasonality in Number of Customers by Company

Total Number of Rides by Company and Year
_____

Consistent Seasonality: Both Pink Cab and Yellow Cab exhibit a clear yearly seasonal trend in the number of customers.

Yellow Cab Dominance: Throughout the observed years, Yellow Cab consistently maintains a higher number of customers compared to Pink Cab

Slight Decline: There seems to be a slight downward trend in the number of customers for both companies over the years, though it's not very pronounced.

Similar Fluctuations: Both companies experience similar fluctuations in customer numbers from year to year, suggesting they might be influenced by the same external factors.
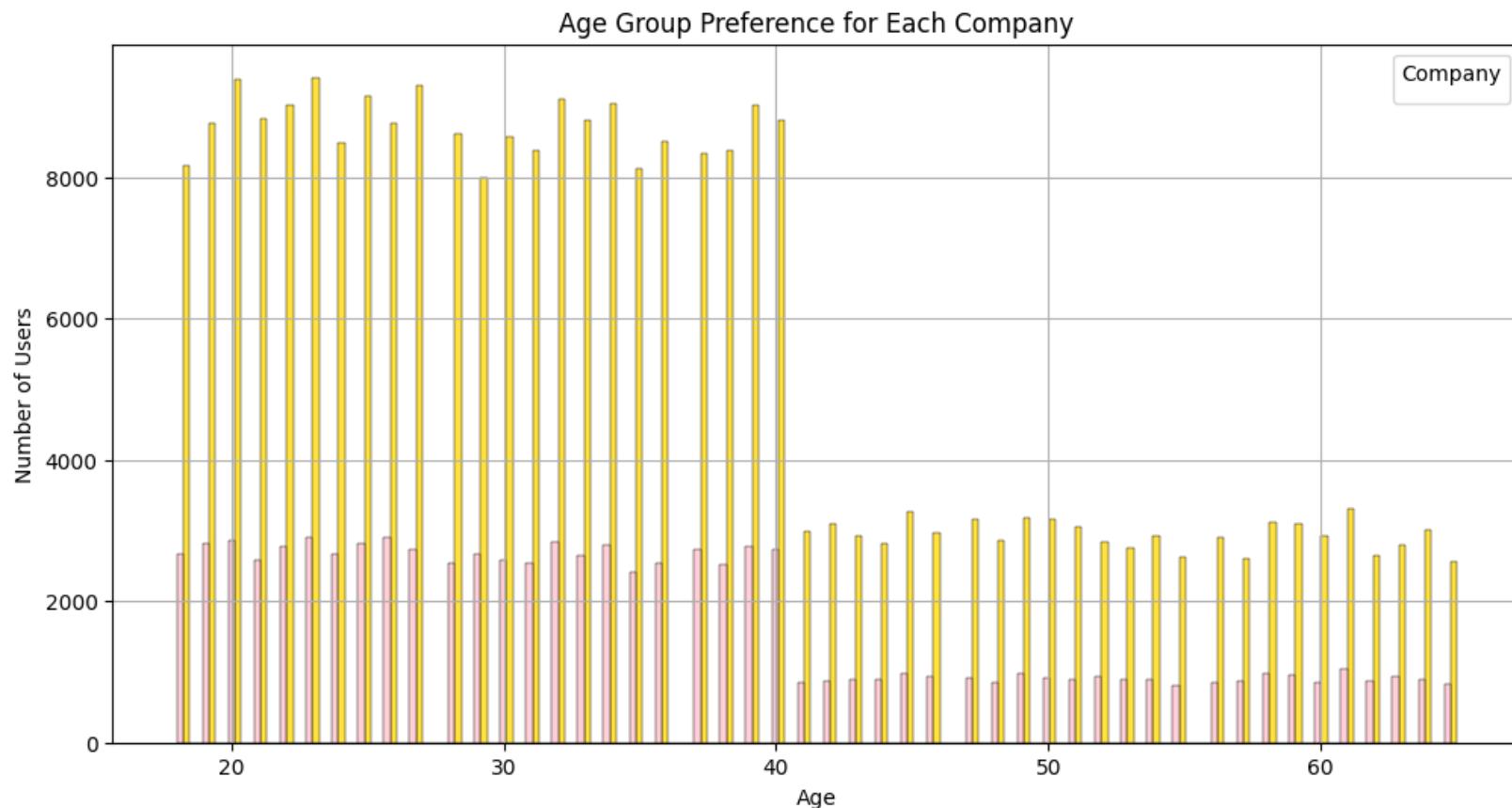

Data Glacier
Your Deep Learning Partner
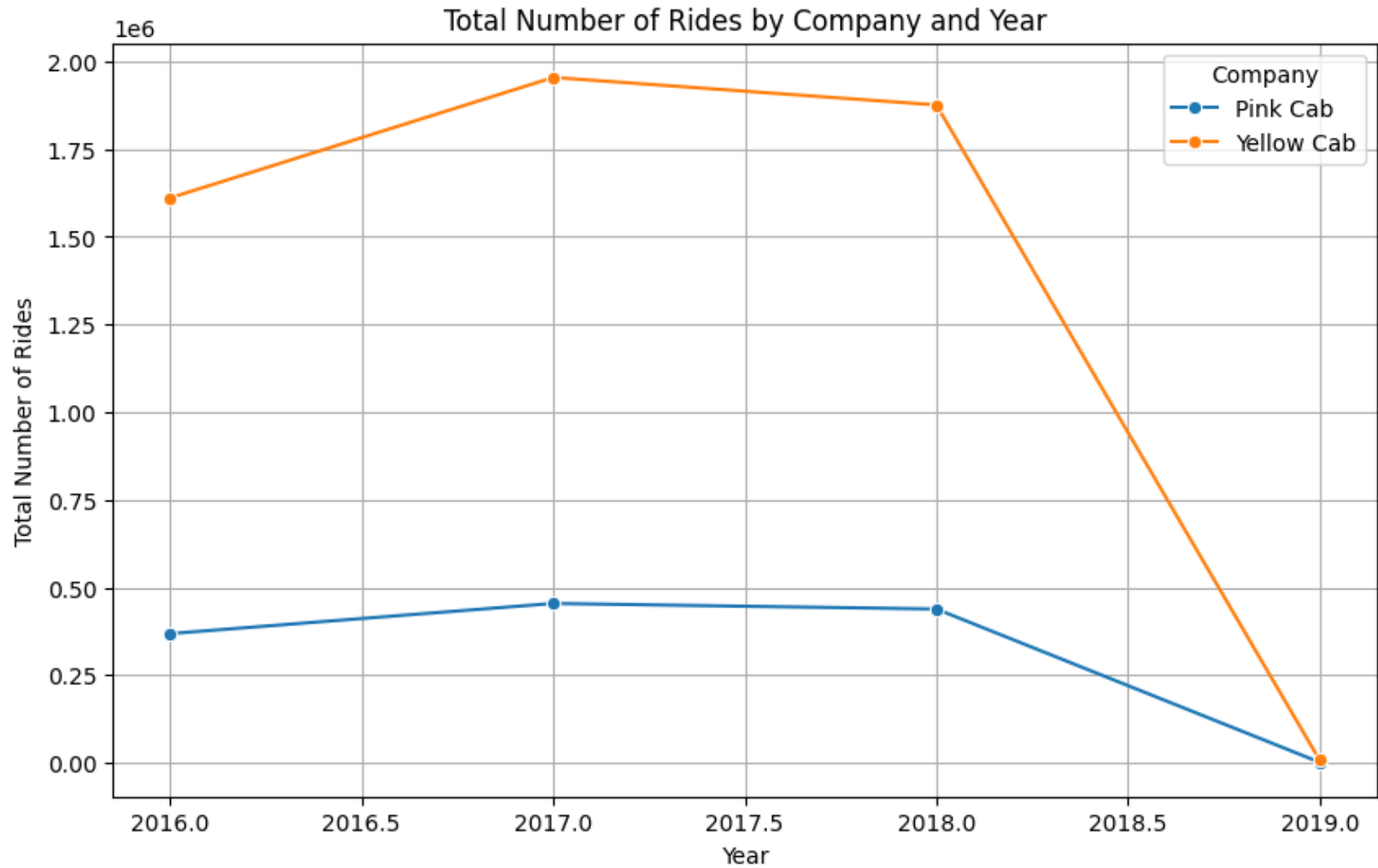
Age Group Preference for Each Company
_____

Consistent Preference: Both companies have a relatively consistent user base across all age groups.

Slight Decline: There is a slight decline in the number of users for both companies as age increases.

Yellow Cab Dominance: Yellow Cab maintains a larger user base compared to Pink Cab across all age groups.


Data Glacier
Your Deep Learning Partner

Income Level and Cab Usage by Company
_____

Consistent Preference: Both companies have a relatively consistent user base across all income levels.

Slight Decline: There is a slight decline in the number of users for both companies as income increases.

Yellow Cab Dominance: Yellow Cab maintains a larger user base compared to Pink Cab across all income levels.

Data Glacier
Your Deep Learning Partner
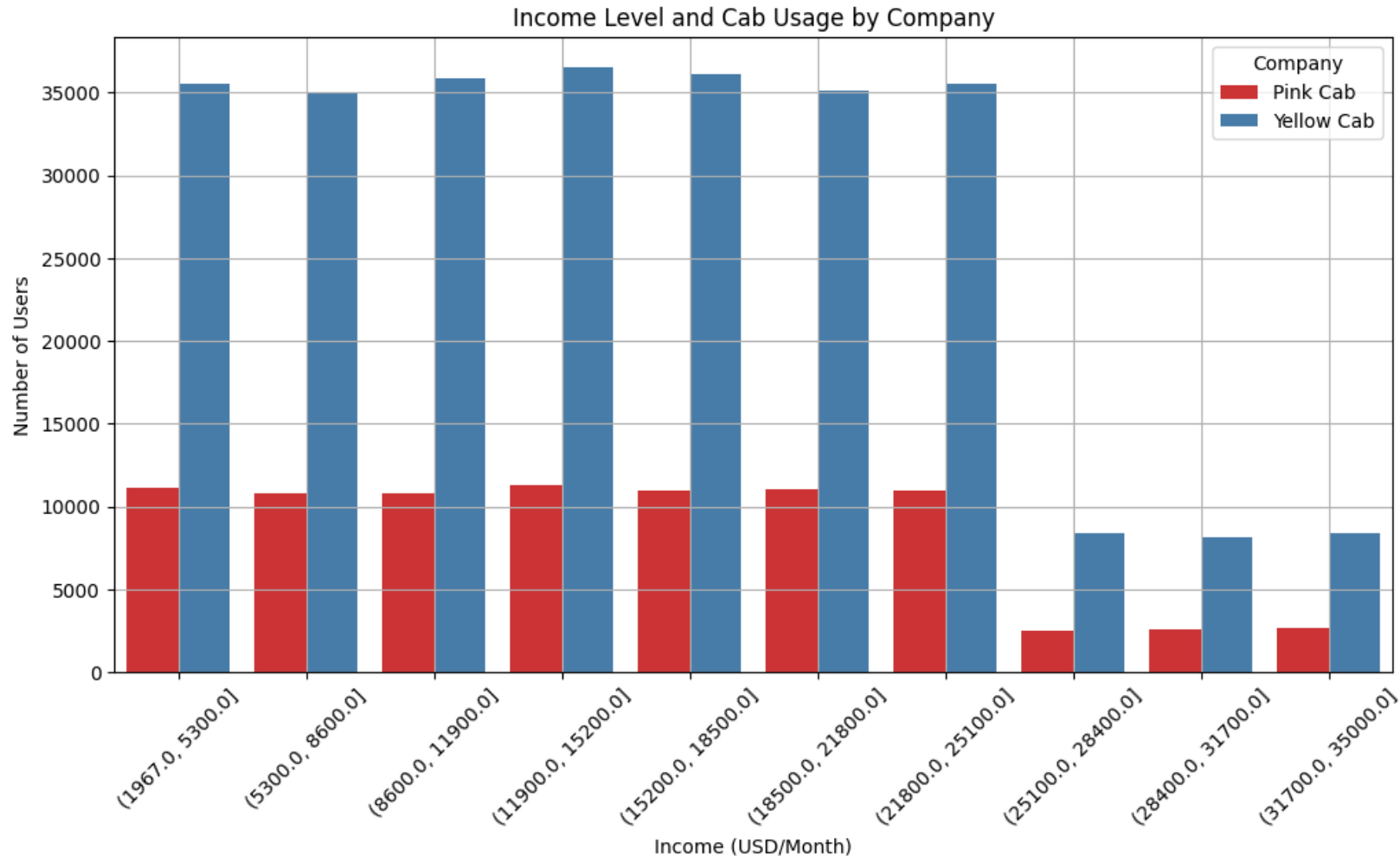
Gender Preference for Each Company
_____

Male Dominance: The number of male users is significantly higher than the number of female users for both Pink Cab and Yellow Cab.

Yellow Cab Dominance: Yellow Cab has a larger user base than Pink Cab for both genders.

Data Glacier
Your Deep Learning Partner
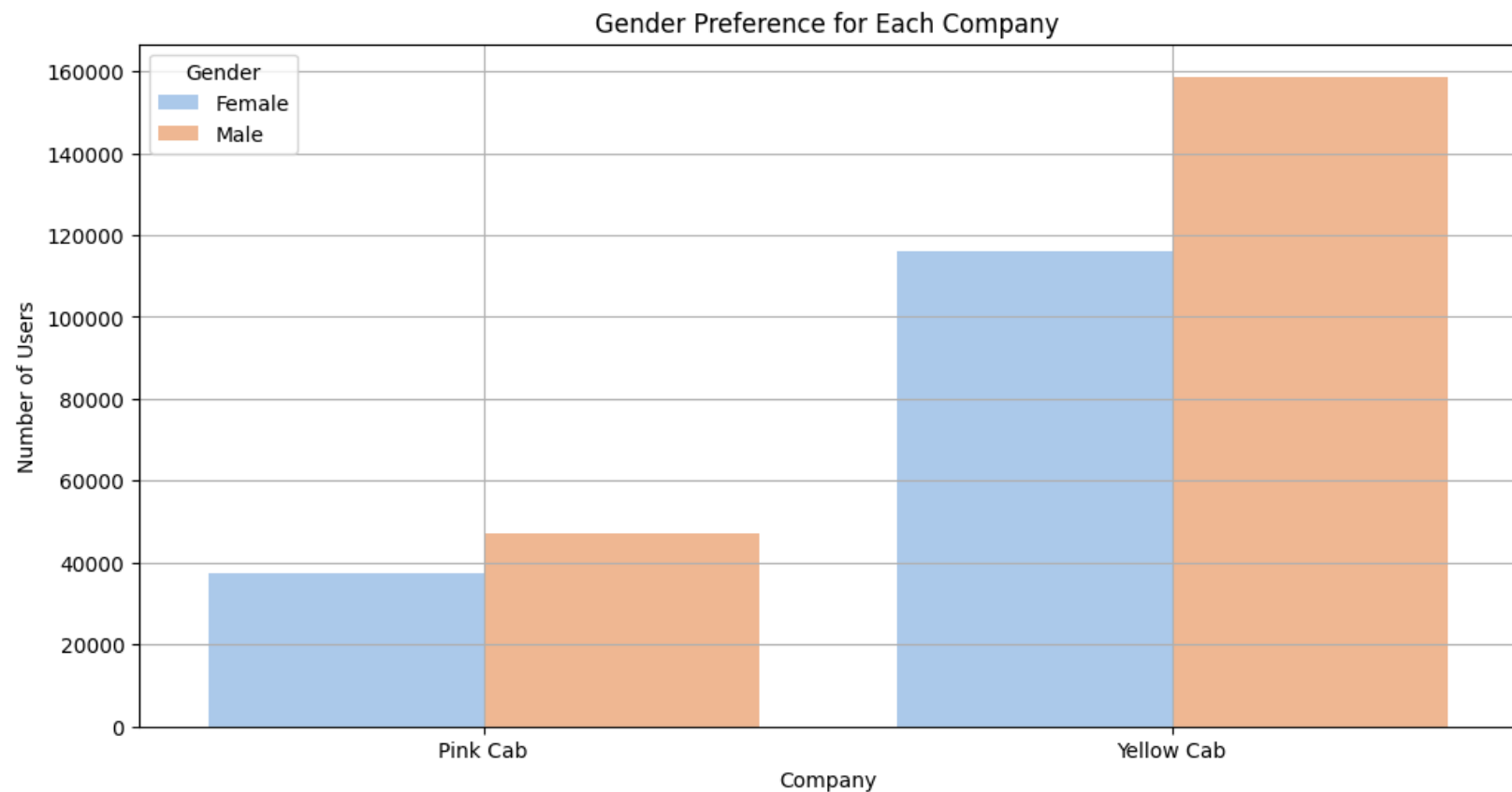
Impact of Distance on Profit by Company
_____

Positive Correlation: Both Pink Cab and Yellow Cab generally show a positive correlation between distance traveled and profit. This suggests that, on average, longer trips tend to generate higher profits for both companies.

Profit Variability: Yellow Cab exhibits a wider range of profits for a given distance compared to Pink Cab. This indicates that Yellow Cab's profits are more influenced by factors other than distance.

Impact of Distance on Profit by Company

Total Profit by Company
_____

Yellow Cab has a significantly higher total profit than Pink Cab. The total profit of Yellow Cab is approximately 40 million USD, while the total profit of Pink Cab is approximately 6 million USD. This suggests that Yellow Cab is a more profitable company than Pink Cab.


Total Profit by Company

Data Glacier
Your Deep Learning Partner
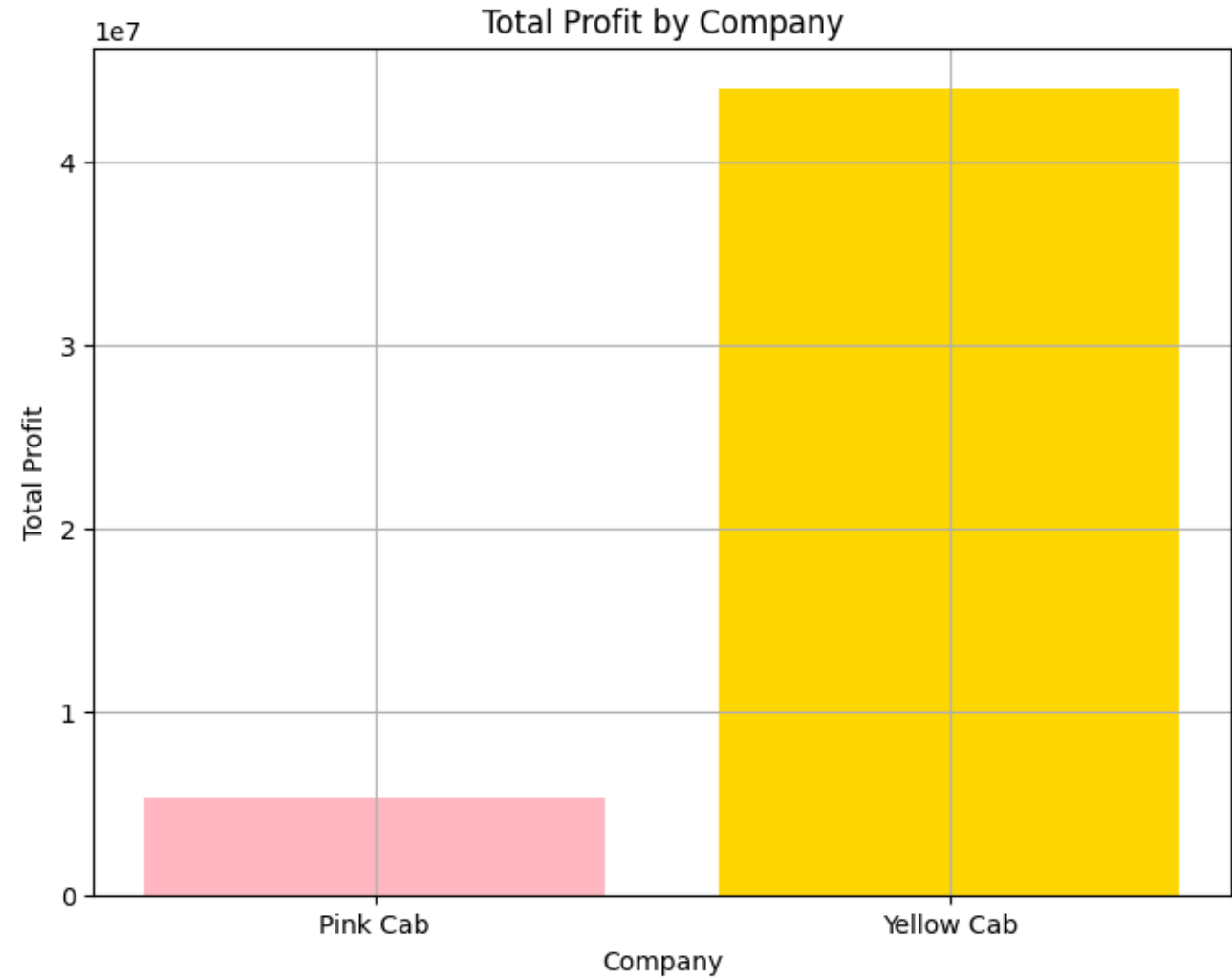
Total Transactions by Payment Mode and Company
_____

Yellow Cab Dominance: Yellow Cab has a significantly higher number of total transactions compared to Pink Cab for both card and cash payment modes.

Card Preference: For both companies, card payments account for a significantly higher number of transactions compared to cash payments.

Payment Mode Gap: The gap in transaction numbers between the two companies is more pronounced for card payments than for cash payments.

Total Transactions by Payment Mode and Company
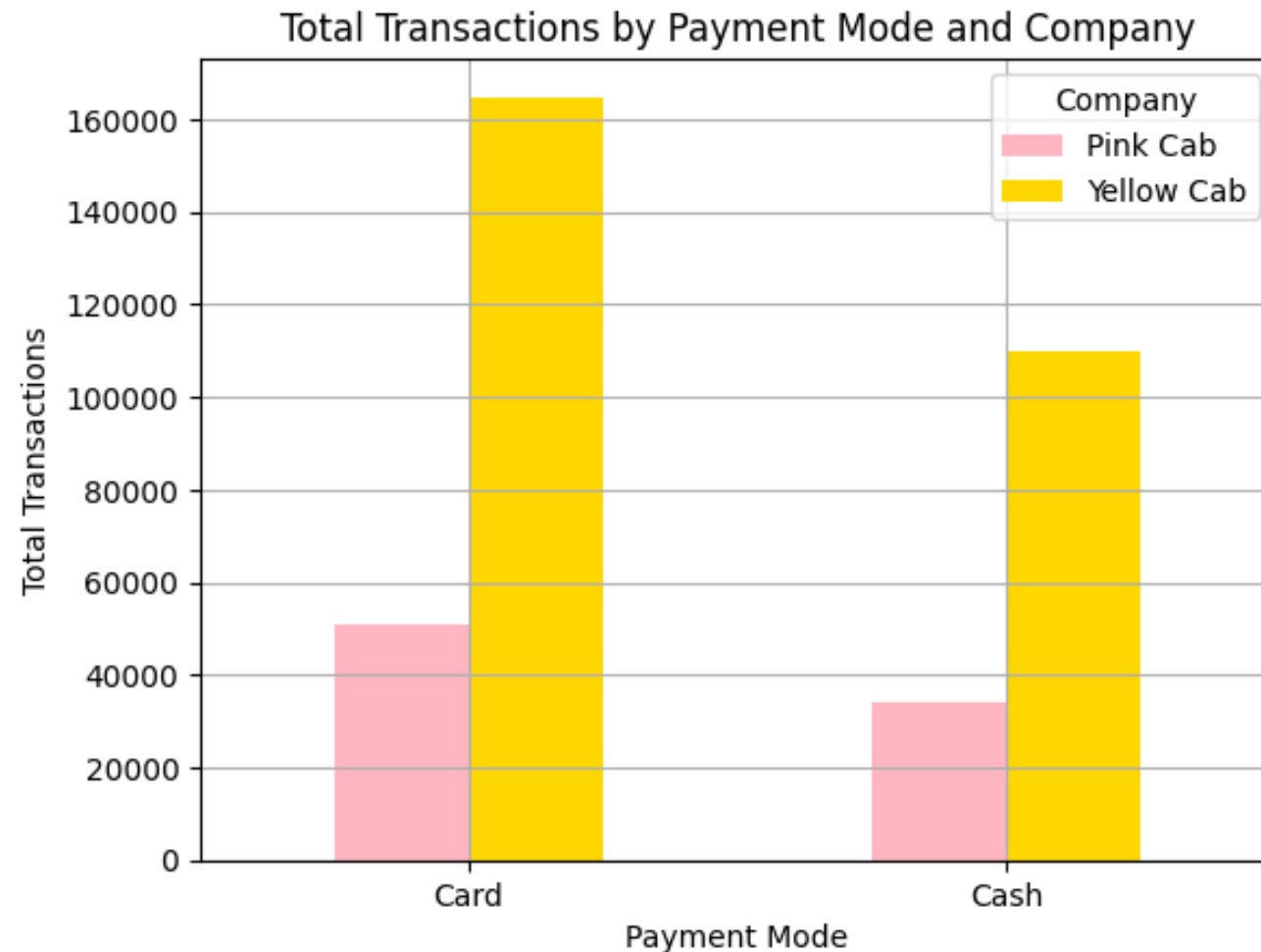
Number of Customers in Cities by Company

Yellow Cab Dominance: Yellow Cab consistently has a higher number of customers in all cities compared to Pink Cab.

Market Concentration: The customer base for both companies is concentrated in a few major cities, with New York and Chicago having the highest number of customers.

Number of Customers in Cities by Company
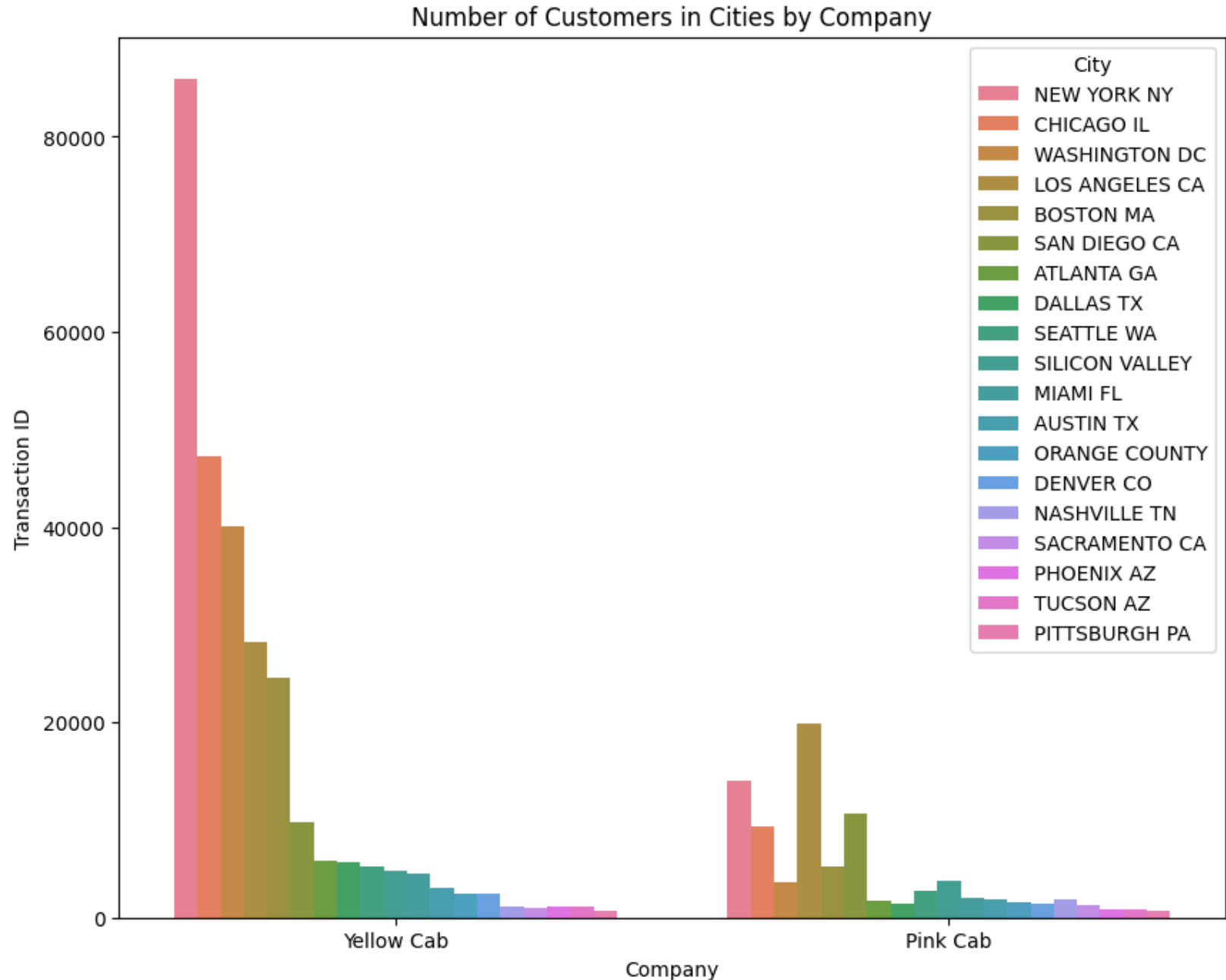
Profit Margin by Company
_____

Yellow Cab exhibits a significantly higher profit margin compared to Pink Cab.

Profit Margin by Company

Data Glacier
Your Deep Learning Partner
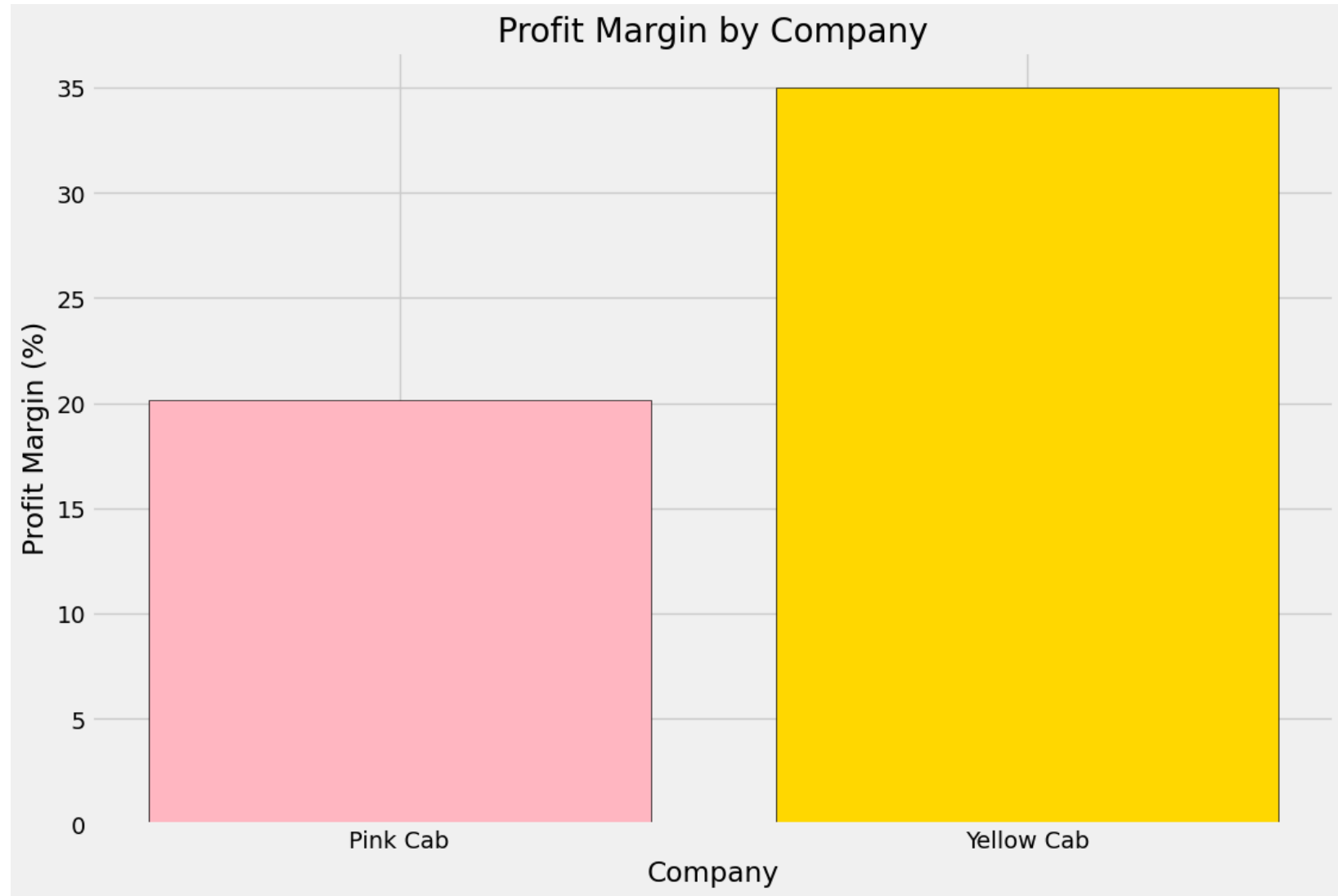
Hypothesis Test

Hypothesis 1: Average Profit Margin Comparison

Hypothesis: The average profit margin of Yellow Cab is higher than that of Pink Cab.

P value is 0.0

We reject the null hypothesis: There is a significant difference in average profit margin between Pink Cab and Yellow Cab.

Hypothesis 2: Weekdays vs. Weekends Usage

Hypothesis: There is a significant difference in the number of rides between weekdays and weekends for both Pink Cab and Yellow Cab.

P value (Pink Cab) is 0.201763491284747l

P value (Yellow Cab) is 0.14969354943135718

We accept the null hypothesis for Pink Cab: There is no significant difference in rides between weekdays and weekends.

We accept the null hypothesis for Yellow Cab: There is no significant difference in rides between weekdays and weekends.

Hypothesis 3: Income Level and Cab Usage

Hypothesis: Customers with higher incomes tend to use Yellow Cab more frequently than Pink Cab.

Number of high income customers - Pink Cab: 21261

Number of high income customers - Yellow Cab: 68551

P value is 0.405007544879073387

We accept the null hypothesis: There is no significant difference in high income customer usage between Pink Cab and Yellow Cab.

Hypothesis 4: Age Group Preferences

Hypothesis: There are age group preferences that differ significantly between Pink Cab and Yellow Cab.

P value is 4.24756254342697272e-07

We reject the null hypothesis: There is a significant difference in age group preferences between Pink Cab and Yellow Cab.

Hypothesis 5: Effect of Population Density on Usage

Hypothesis: There is a correlation between population density in cities and the number of rides for Yellow Cab.

Correlation between Population and Number of Rides for Yellow Cab: 0.8496096914969222

P value is 0.0

We reject the null hypothesis: There is a significant correlation between population density and number of rides for Yellow Cab.

Recommendation

Yellow Cab Company is better than Pink Cab Company because:

- **Profit Margin:** Yellow Cab Company has a higher profit margin.
- **More Users:** Yellow Cab Company serves a larger number of users.
- **More Transactions per Year and per Month :** Yellow Cab Company records more transactions both annually and monthly.

Thank You