



Data Glacier

Your Deep Learning Partner

Project: G2M Insight For Cab Investment Firm

Data Science Virtual Internship

Submitted by: Aparna Virupakshi

Date: 13-June-2024

Outline

- ✓ Problem Statement
- ✓ Understanding the Data
- ✓ Exploratory Data Analysis
- ✓ Hypothesis Tests
- ✓ Summary

Problem Statement

XYZ is a private equity firm based in the US. Due to the remarkable growth in the cab industry over the last few years and the presence of multiple key players in the market, XYZ is considering an investment in this sector.

- Objective: The goal is to provide actionable insights to help XYZ identify the most suitable cab company for investment.

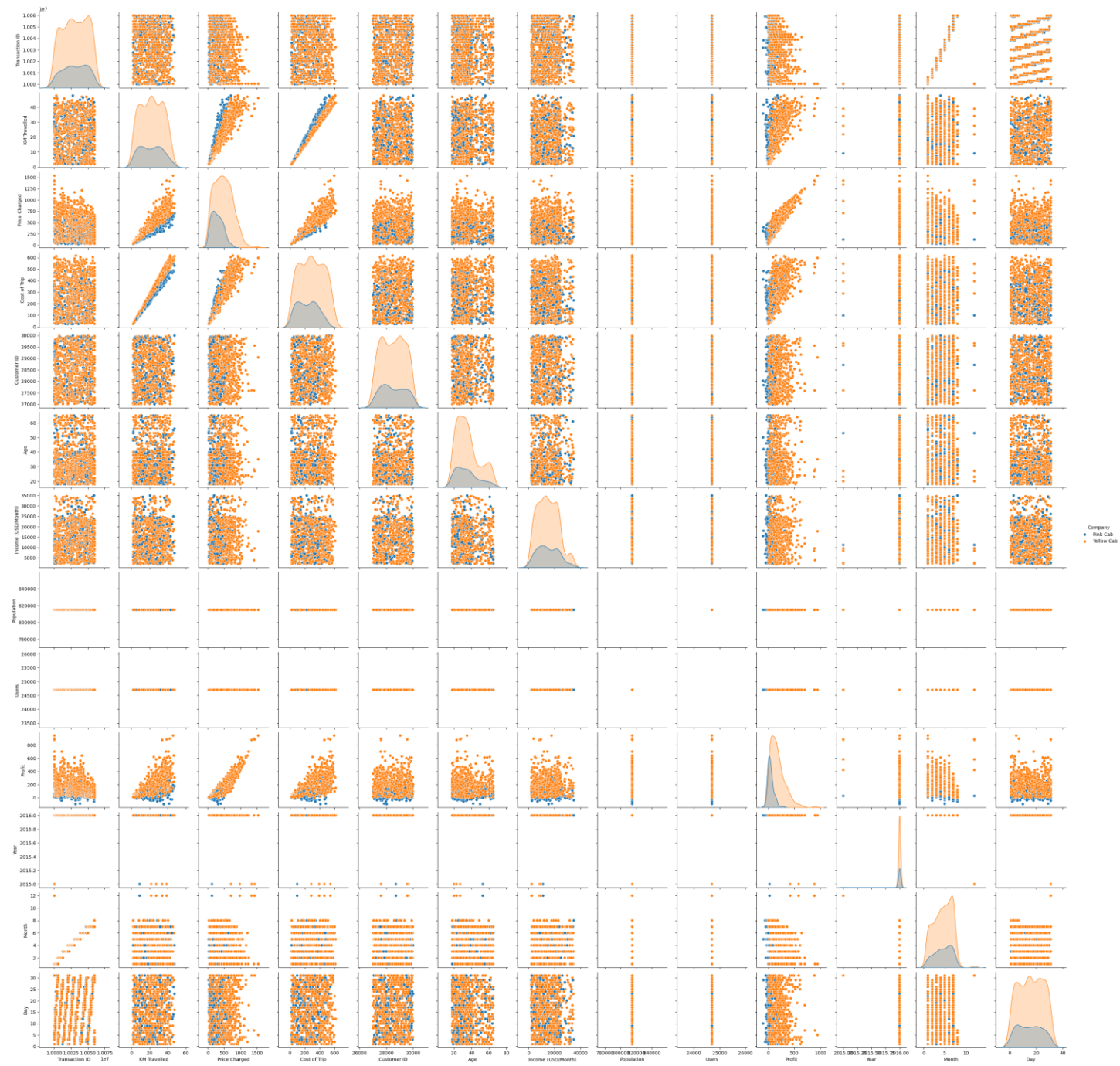
Analysis:

- Data Understanding and Visualization
- Finding the Most Users Cab Company
- Finding the cheapest Cab Company for users
- Finding the Most Profitable Cab Company
- Multi Hypothesis Testing and Investigation

Understanding the Data

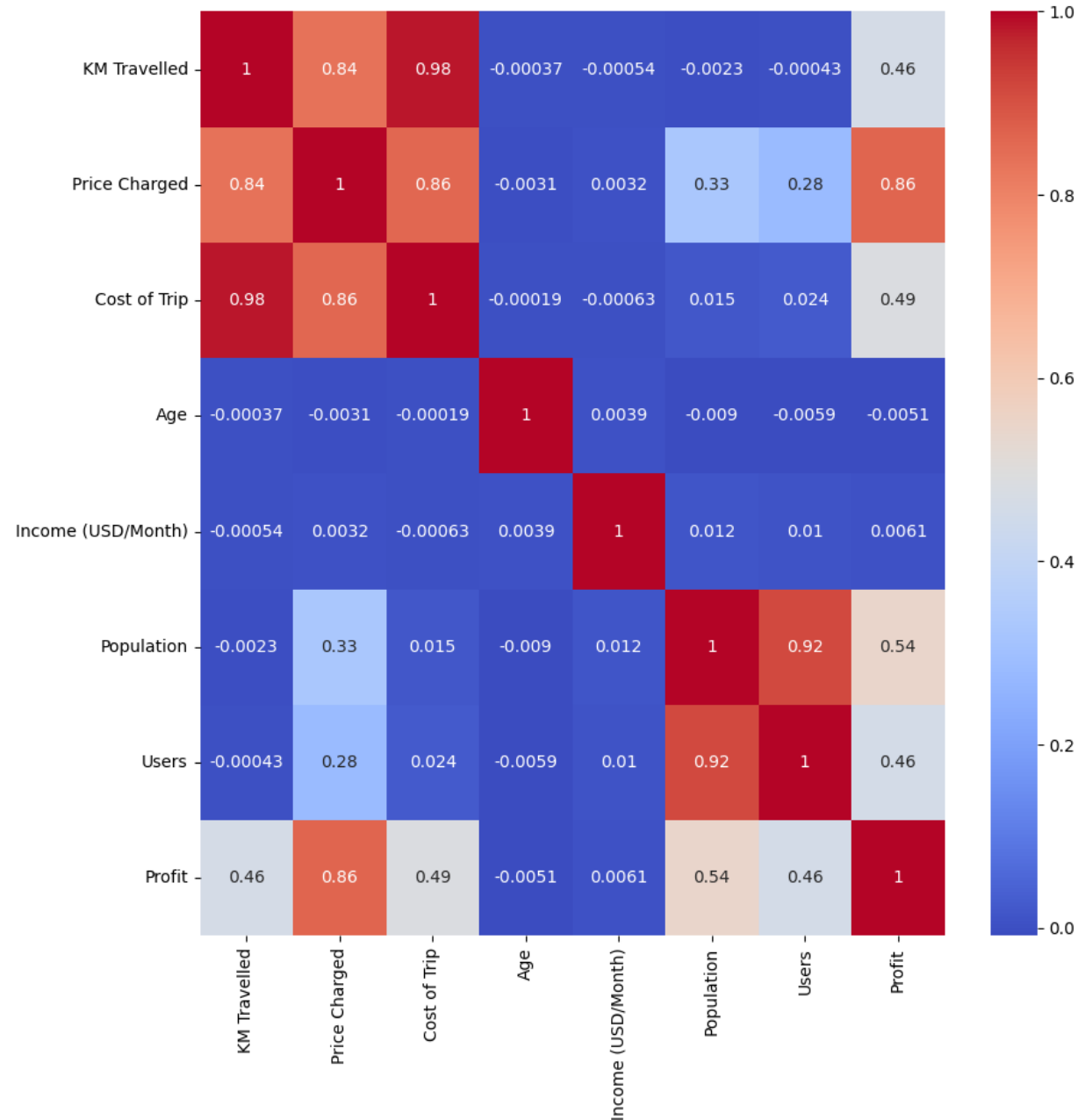
- ✓ **Cab_Data.csv**: Contains detailed records of transactions from two cab companies.
- ✓ **Customer_ID.csv**: A mapping table that provides a unique identifier to link customers with their demographic information.
- ✓ **Transaction_ID.csv**: A mapping table that links transactions to customers and specifies the payment mode.
- ✓ **City.csv**: Lists various US cities along with their population and the number of cab users.

Relationships Between Variables



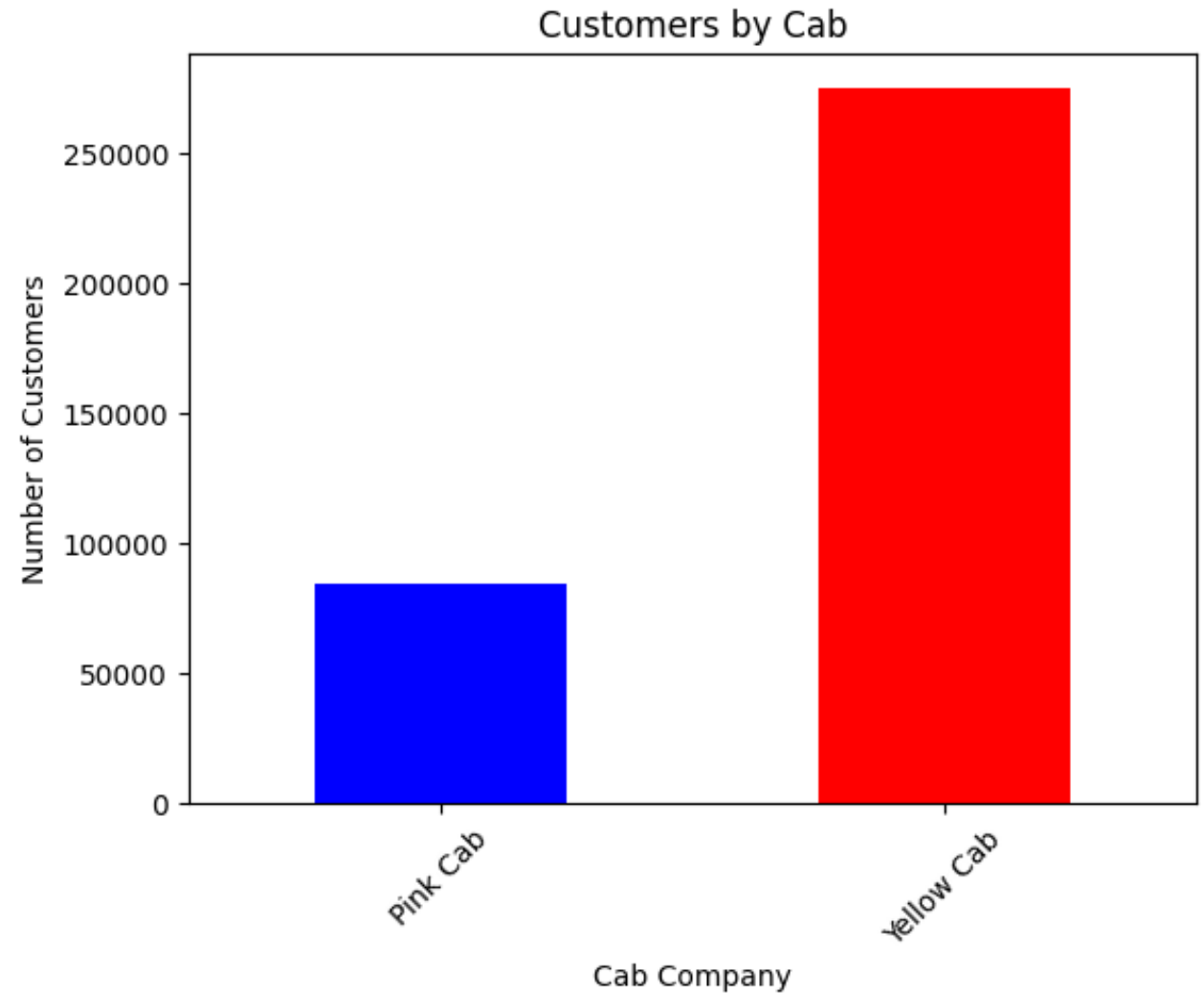
Correlation Between Variables

- 1. The heatmap shows that KM Travelled is very much related to the Cost of Trip and Price Charged
- 2. Strong correlation between Population and Users as well which is very obvious.



Which Company has more Customers?

- Most of the customers ride on Yellow Cab compared to Pink Cab



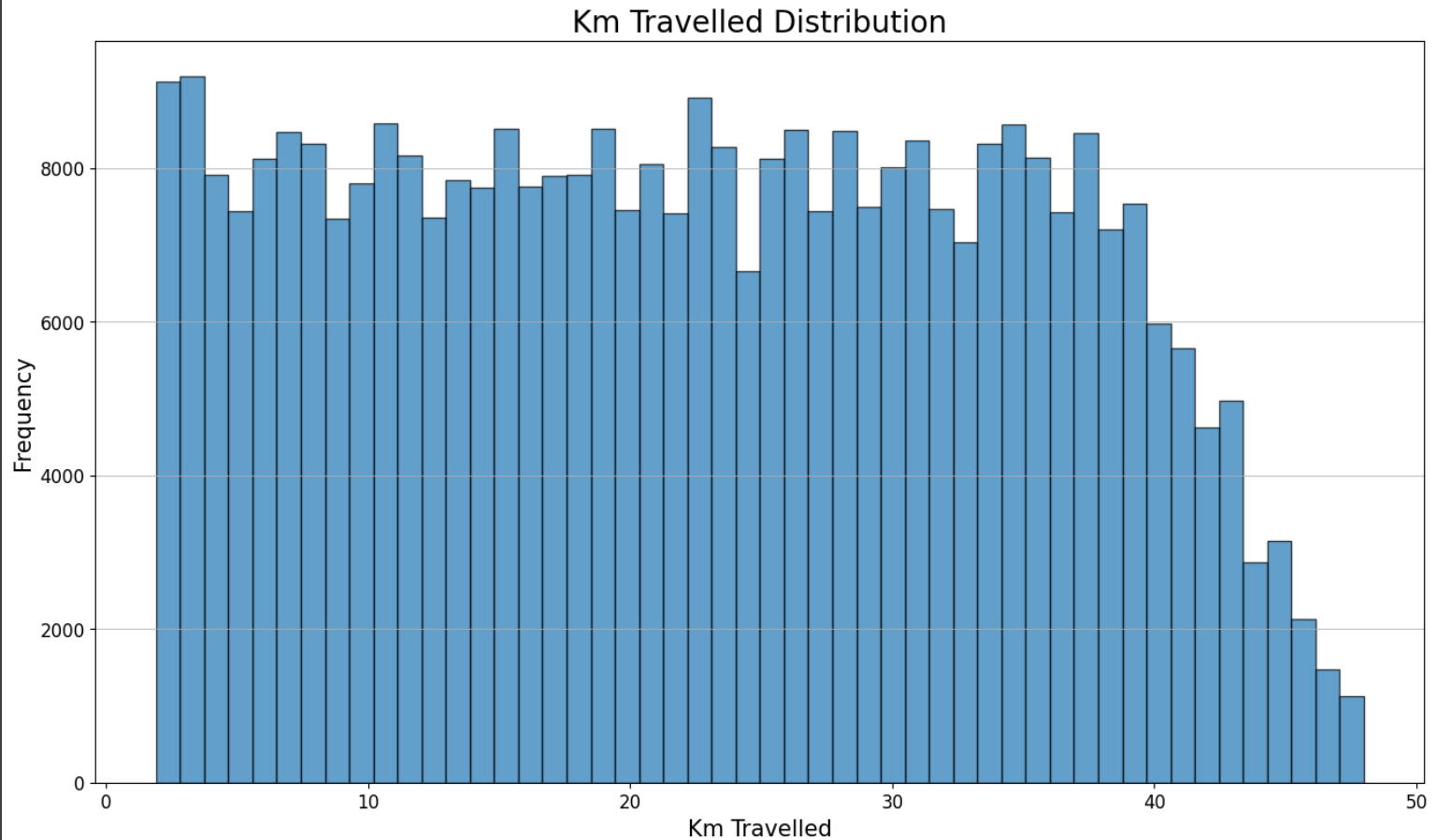
Which Company has a high price charged?

- The price charged for Yellow Cab is higher than Pink Cab.



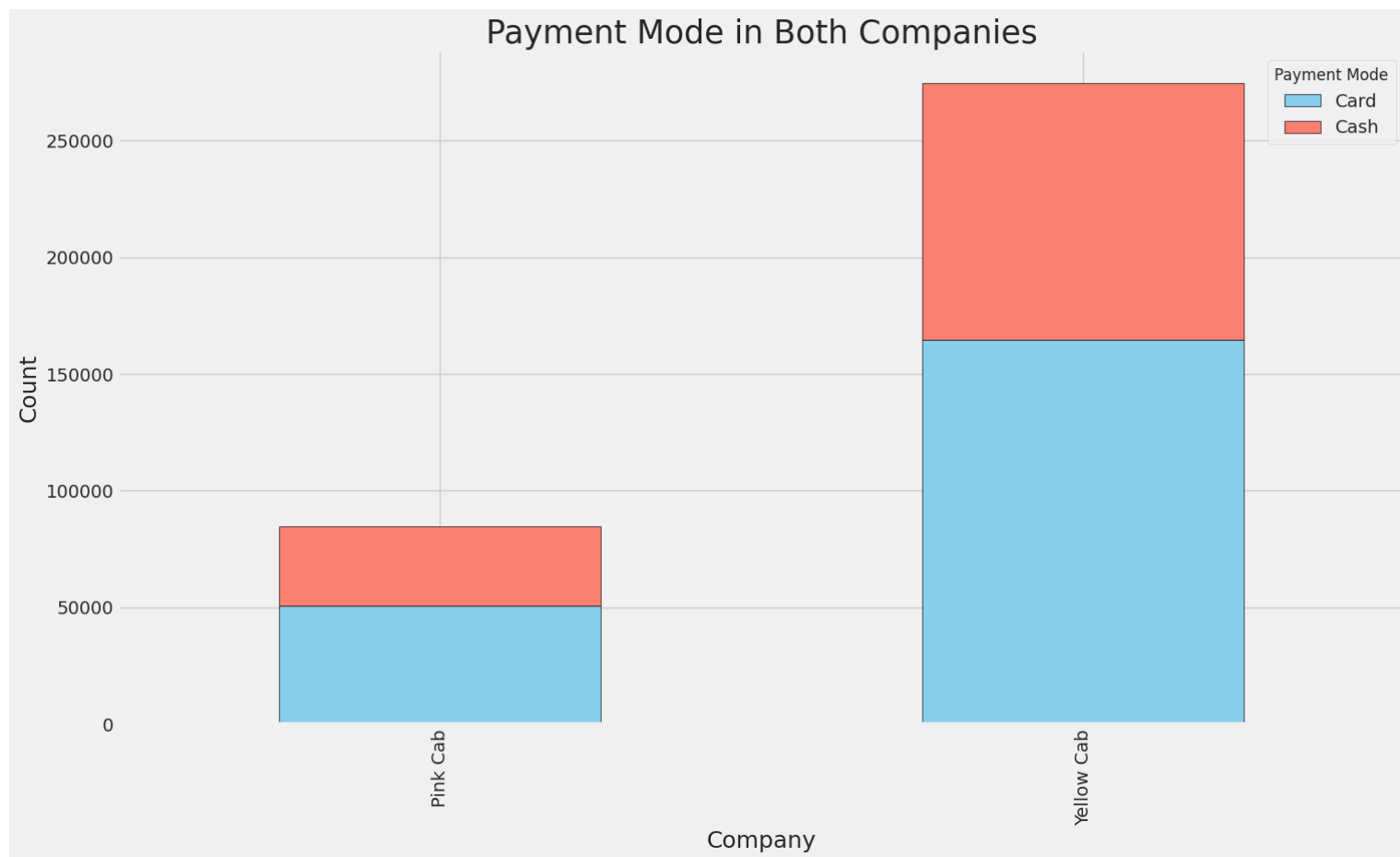
KM Travelled Distribution

- The highest frequencies are observed in the range of 0 to 10 kilometers, suggesting that short-distance trips are the most common.
- The histogram shows a relatively uniform distribution of trips up to around 30 kilometers, indicating a high frequency of short to mid-range trips.



Payment Mode In Both Companies

- For Yellow Cab, the number of card payments is almost double the number of cash payments. This indicates a strong preference for cashless transactions among Yellow Cab users.
- For Pink Cab, the distribution is more even, suggesting that both payment modes are similarly popular among its users.



Users concerning Gender

➤ Male Customers:

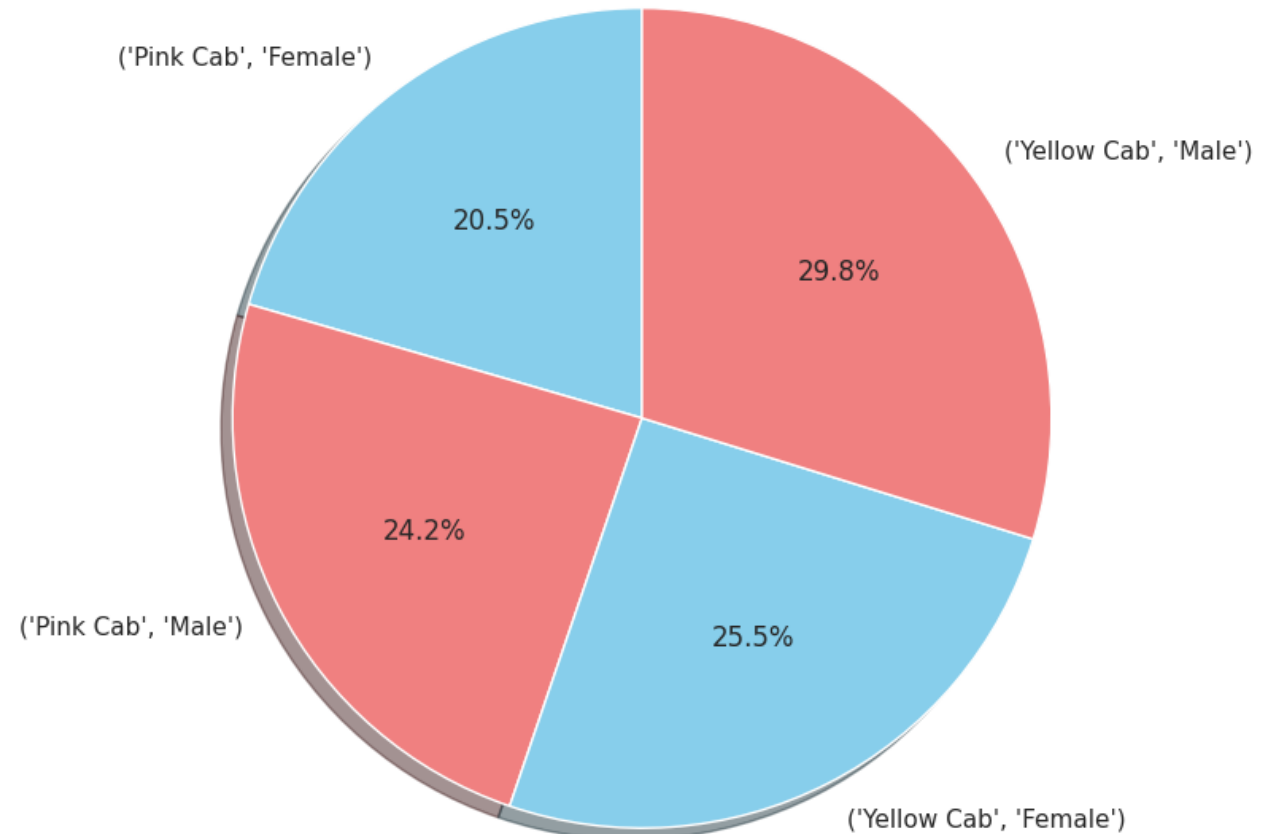
Yellow Cab has a higher percentage of male customers (29.8%) compared to Pink Cab (24.2%).

➤ Female Customers:

Yellow Cab also has a slightly higher percentage of female customers (25.5%) compared to Pink Cab (20.5%).

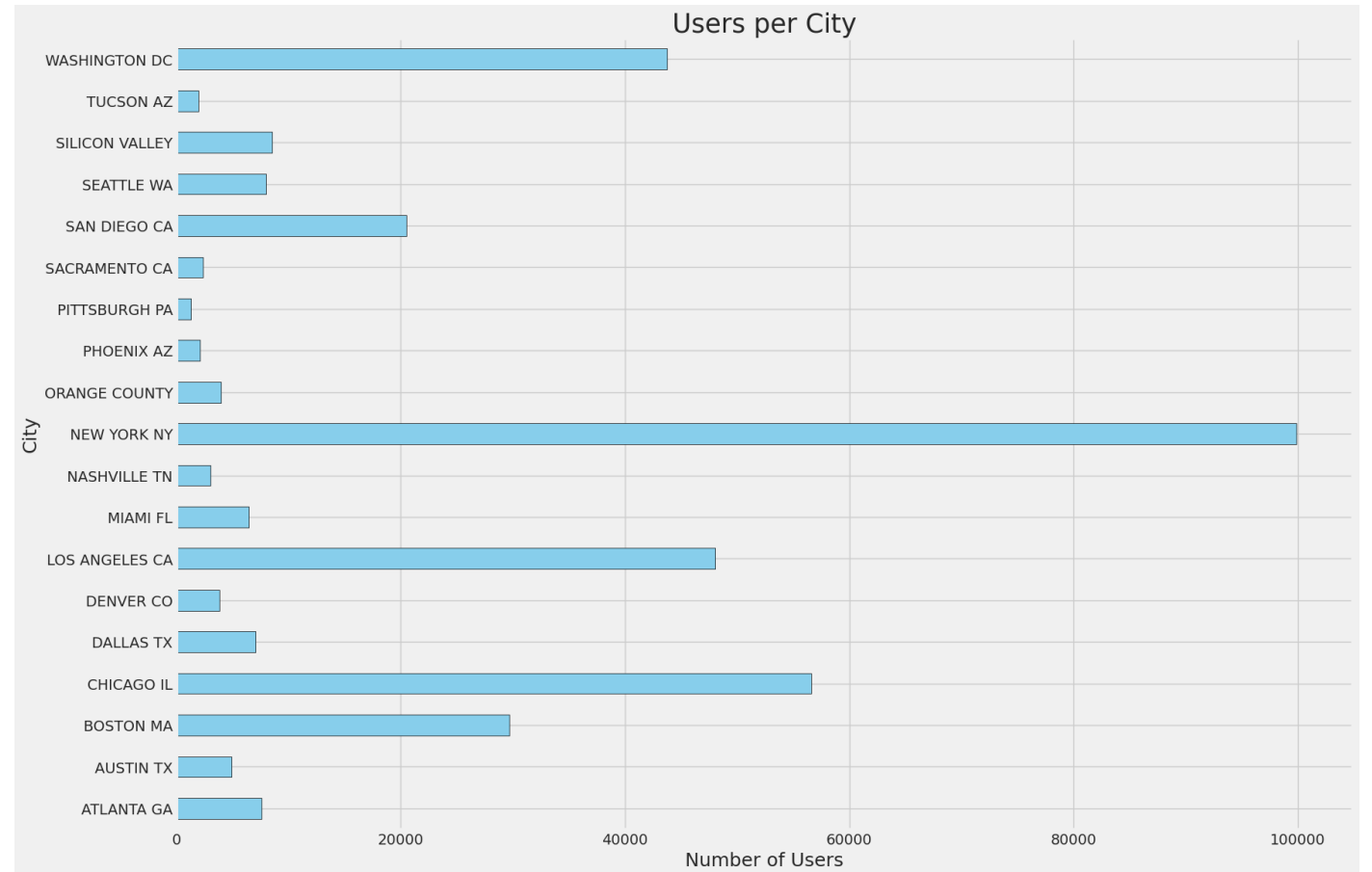
Overall, Yellow Cab has a larger share of both male and female customers compared to Pink Cab.

Customer share per gender per cab



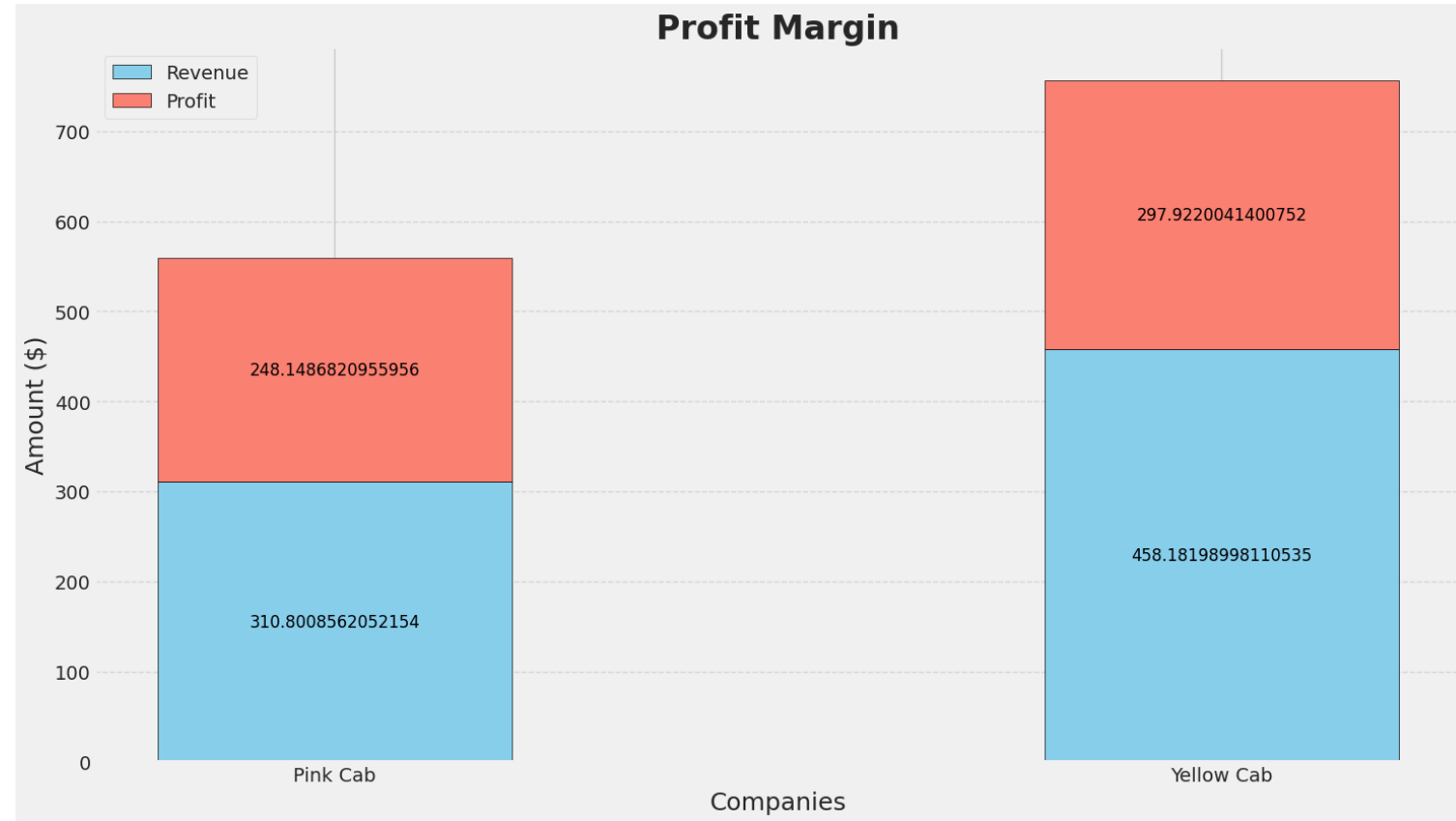
Users concerning Cities

- New York City has the highest Cab Users followed by
- Chicago,
- Los Angeles



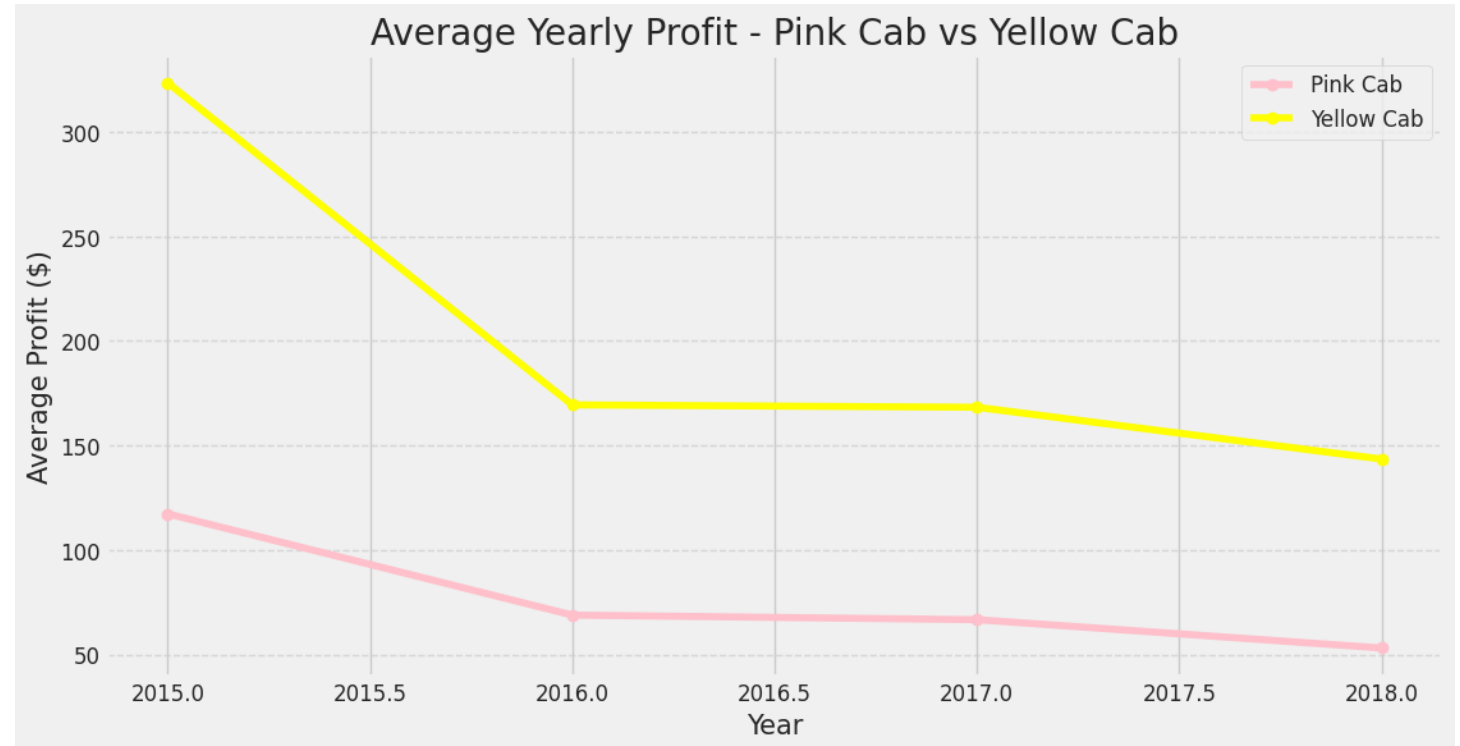
Profit Margin

- Yellow cabs has a higher profit margin (297.922) Compared to Pink Cab



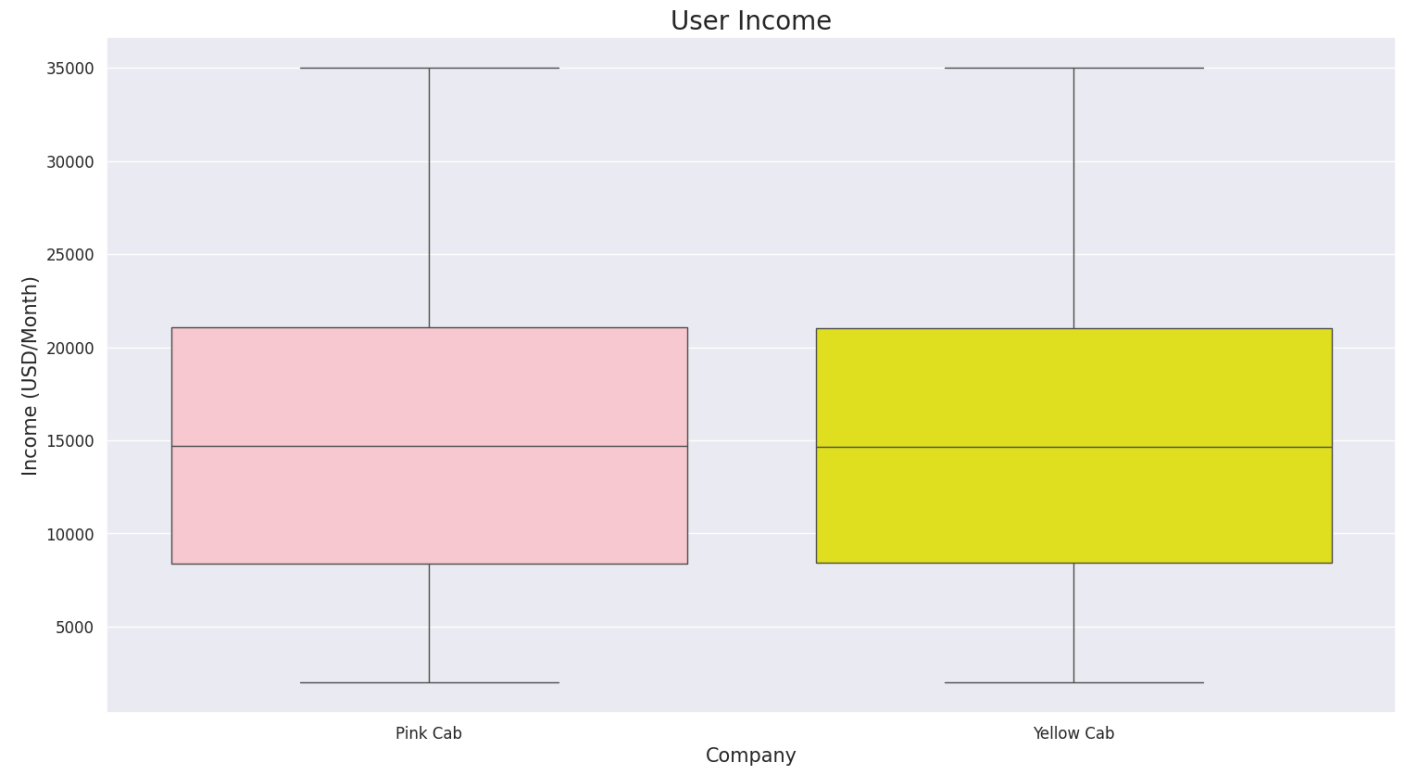
Average Profit Margin With Respect to Year

- The Average Profit Margin with respect to Year is decreased
- Yellow cabs has a higher profit margin compare to Pink Cab



Average Income of Users

- The median income for users of both Pink Cab and Yellow Cab is around \$15,000 USD/month. This is indicated by the line inside each box, representing the median.



Price Charged With Respect to KM Travelled

➤ Yellow Cab:

The data points for Yellow Cab are consistently higher compared to Pink Cab for the same distances, indicating that Yellow Cab generally charges more for similar distances compared to Pink Cab.

➤ Pink Cab:

The data points for Pink Cab are lower, showing that Pink Cab charges less for the same distance.




Hypothesis 1:


Does anything make a difference in Profit regarding Payment Mode

- H_0 : There is no difference regarding payment mode in both cab companies.
- H_1 : There is a difference regarding payment mode in both cab companies.

Pink Cab Company

 P value is 0.7900465828793286
Accept the null hypothesis (H_0) that there is no difference in payment mode for Pink Cab

Yellow Cab Company

 P value is 0.2933060638298729
Accept the null hypothesis (H_0) that there is no difference in payment mode for Yellow Cab

Conclusion

There is no difference regarding Payment Mode in both cab companies

Hypothesis 2:

Does anything
make a difference
in gender
distribution
between
companies

- H0: There is no difference regarding Gender in both cab companies.
- H1: There is a difference regarding Gender in both cab companies.

Chi-Square Test for Proportions:

↔ P value is 3.982674650131372e-25
Reject the null hypothesis (H1) that there is a difference in gender distribution between companies

Conclusion

There is a difference in gender distribution between companies

Hypothesis 3: Correlation between KM Traveled & Price Charged

- H0: There is no significant correlation between KM traveled and Price charged.
- H1: There is a significant correlation between KM traveled and Price charged.

Correlation Test:

```
P value is 0.0  
Correlation coefficient is 0.8357531580209414  
Reject the null hypothesis (H1) that there is a significant correlation between KM Travelled and Price Charged
```

Conclusion

There is a significant correlation between KM travelled and Price charged.

Conclusion

- Yellow Cab Company is better than Pink Cab Company

Because:

- ✓ Higher Profit Margin
- ✓ Larger User Base
- ✓ Higher Transaction per Year
- ✓ Better Geographical Coverages

Thank You



Data Glacier

Your Deep Learning Partner