Project: G2M Insight For Cab Investment Firm

Outline

Problem Statement

Datasets Information

Exploratory Data Analysis

Multiple Hypothesis Test

Summary

Problem Statement

Problem Statement

XYZ is a private equity firm in the US. Due to remarkable growth in the Cab Industry in the last few years and multiple key players in the market, it is planning for an investment in the Cab industry.

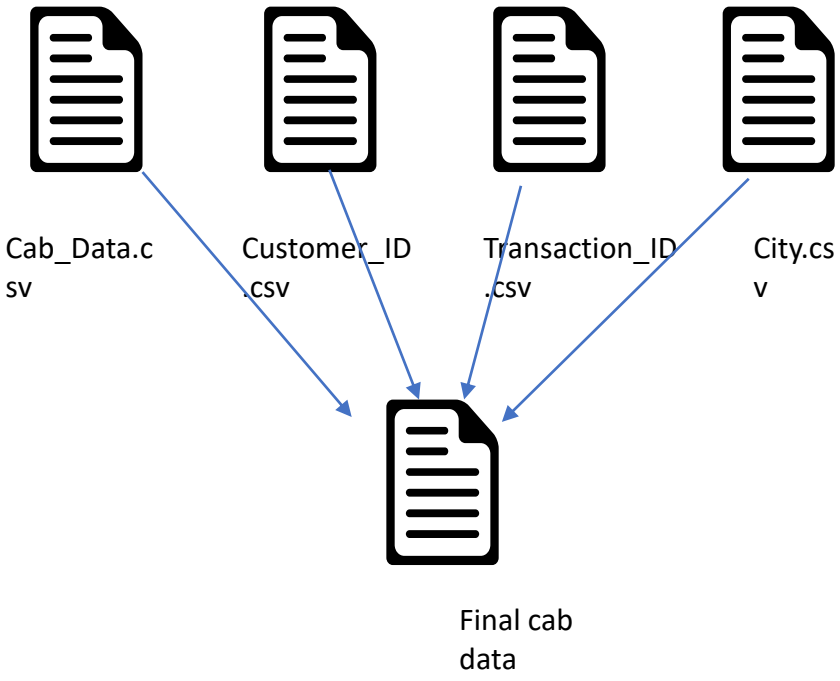
Objective: Provide actionable insights to help XYZ firm in identifying the right company for making an investment.

The analysis has been divided into four parts:

1. Data Understanding and Visualization
2. Finding the most users Cab company
3. Finding the cheapest Cab company for users
4. Finding the most profitable Cab company

Data Information

Data Information



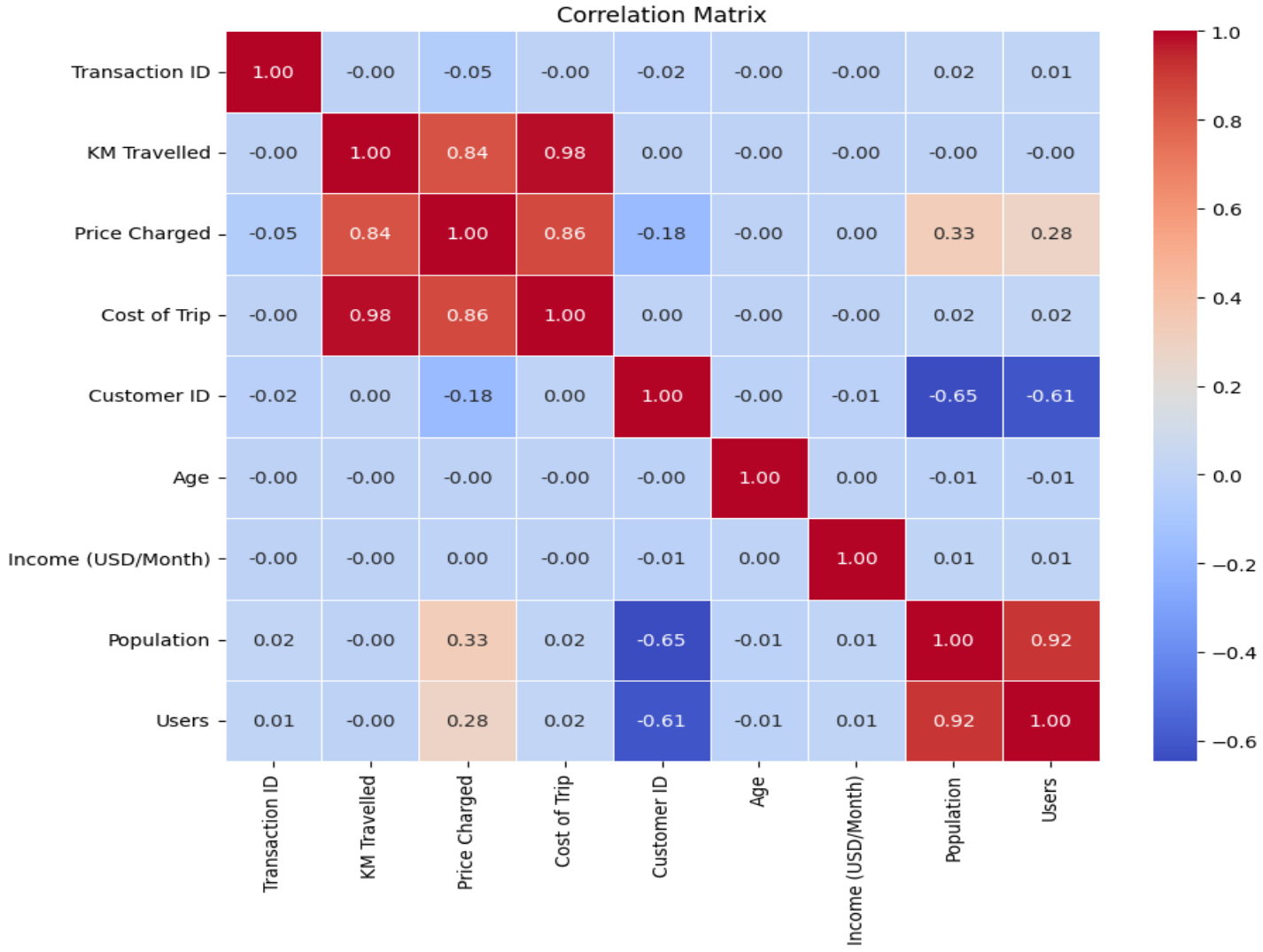
- **Dataset Information:**
  - Cab Data: This dataset contains information about cab rides and includes details like the transaction ID, date of travel, company, city, kilometers traveled, price charged, cost of the trip, and state.
  - City Data: This dataset contains information about various cities, including their population, number of users, city name, and state.
  - Customer Data: This dataset contains information about customers, including customer ID, gender, age, and income (USD/month).
  - Transaction Data: This dataset includes information about transactions, including the transaction ID, customer ID, and payment mode.

- **Timeframe:**
  - Time period of data is from 31/01/2016 to 31/12/2018.

- **Total Data Points:**
  - Total number of records in the dataset: 359392

Correlation Between Variables

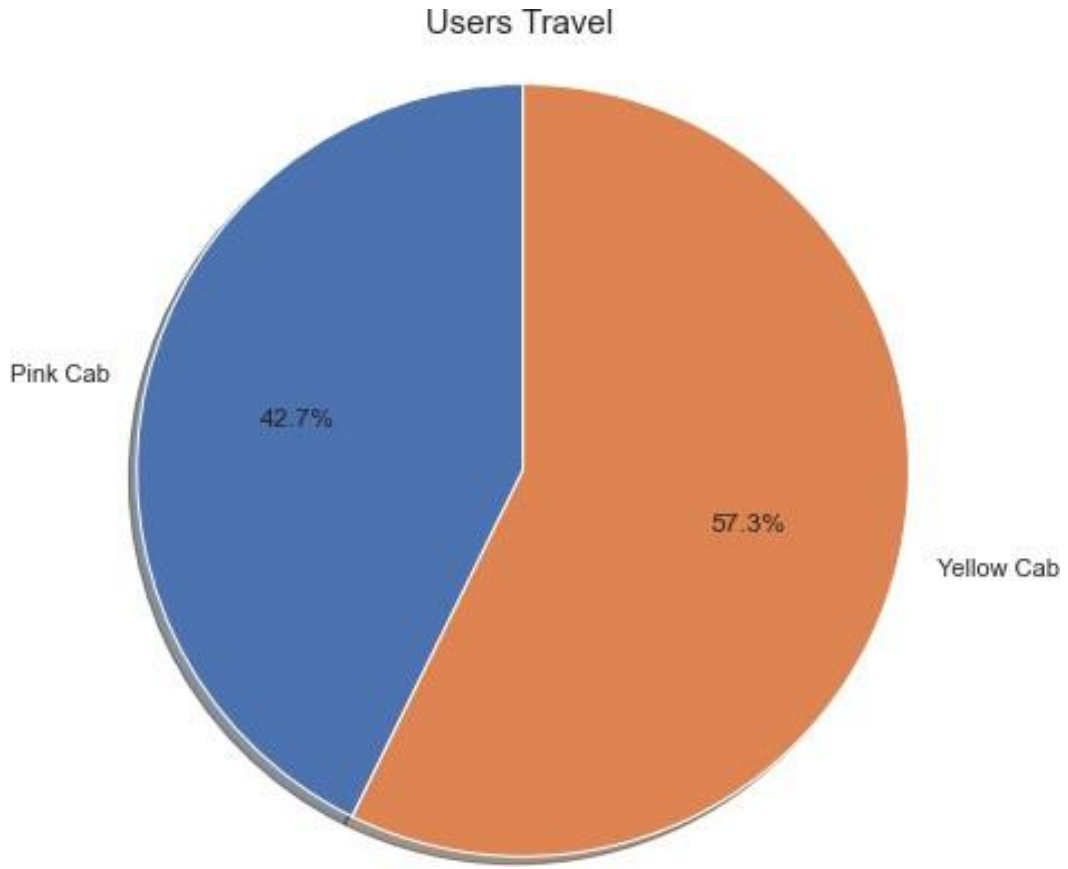
- As we can see there is a strong correlation between
  - Population vs Users
  - Price Charged vs Cost of Trip vs KM Travelled



EDA

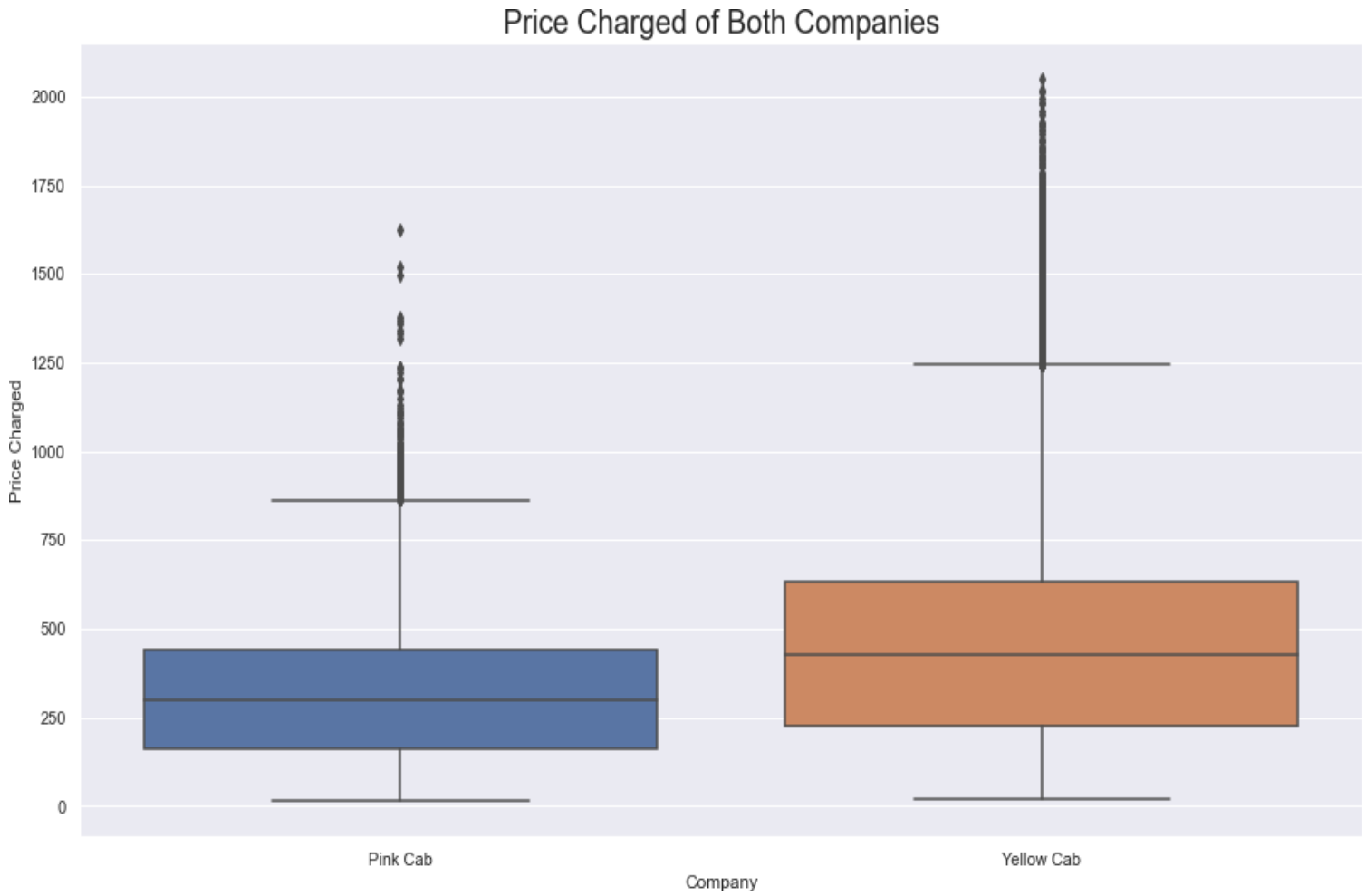
Which Company has more Users?

- As we can see users like to ride on Yellow cab more as compared to Pink Cab



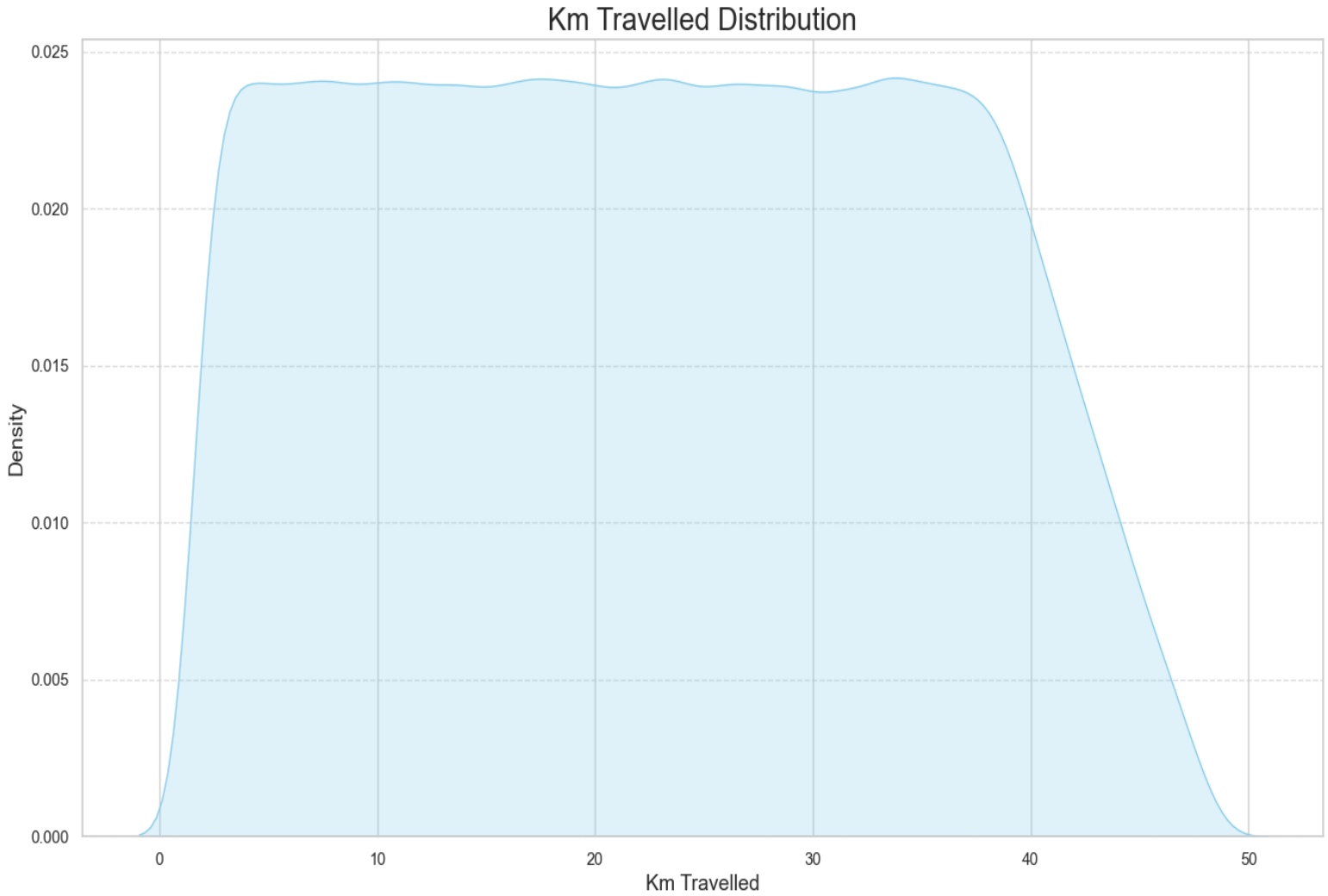
Which Company has a high price charged?

- As we can see Price Charged for Yellow Cab is highest as compared to Pink Cab



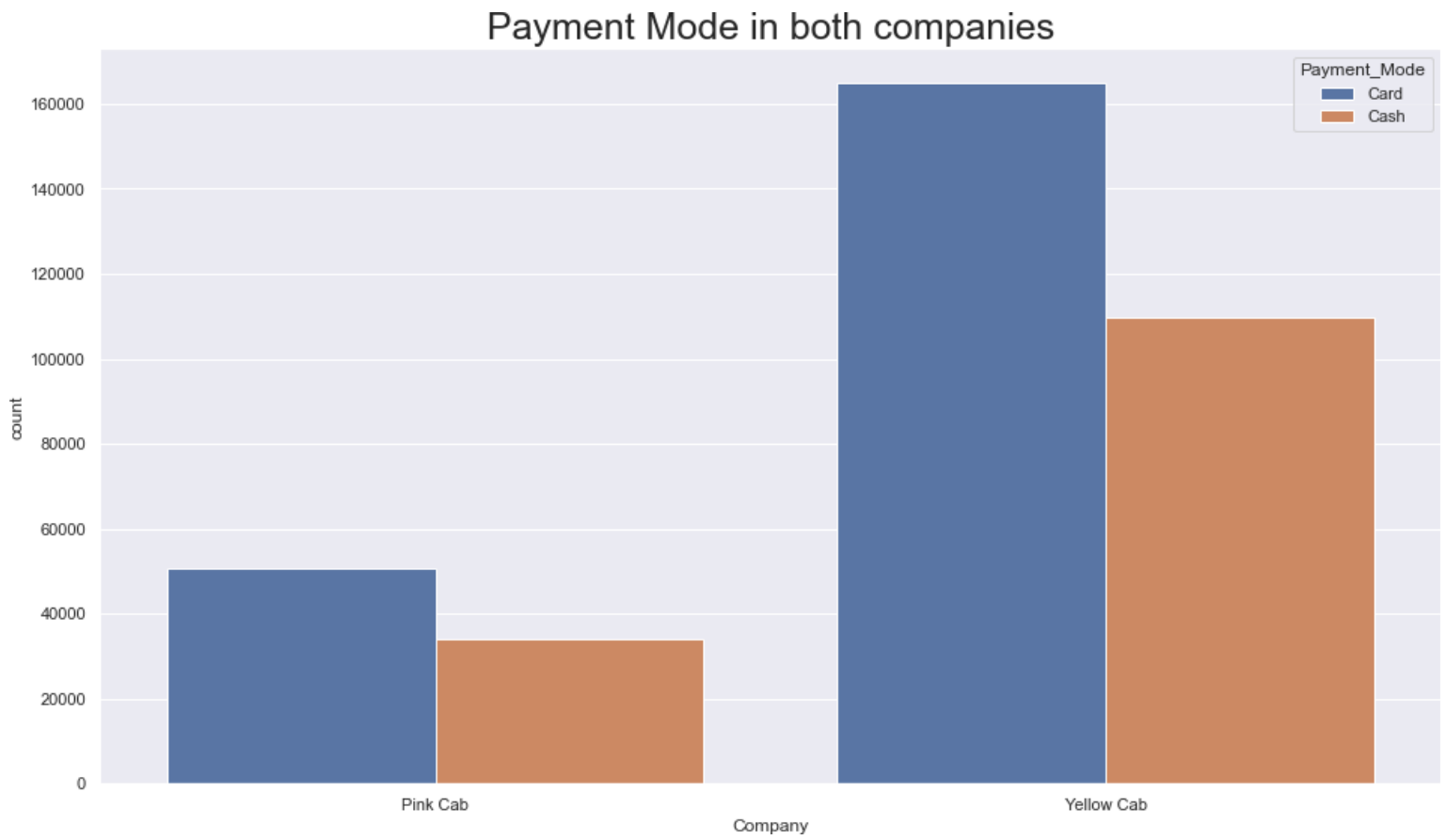
KM Travelled Distribution

- Most of the rides varies from 2 to 48 KM.



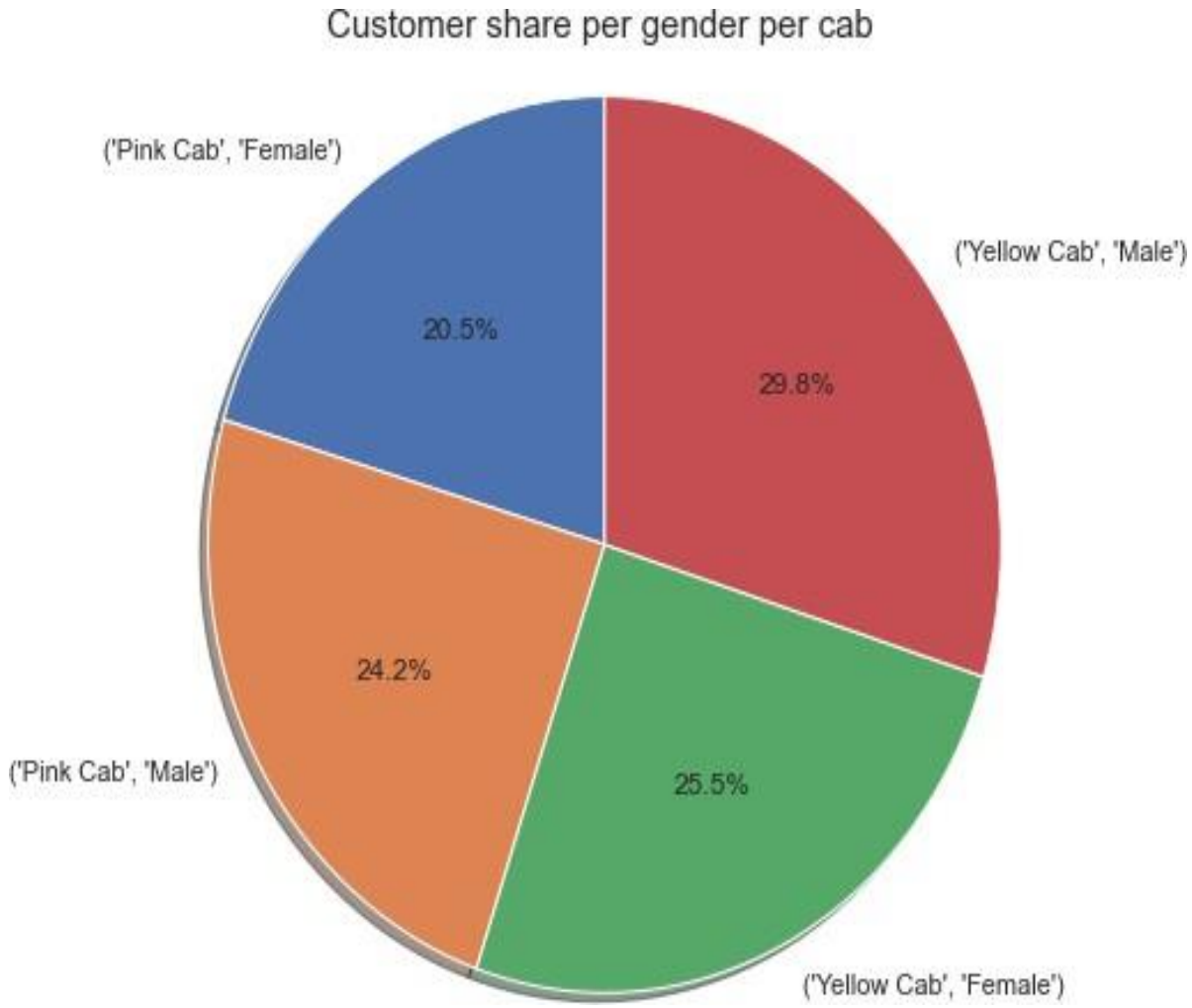
Payment Mode

- As we can see that users prefer to pay with a card more compared to cash



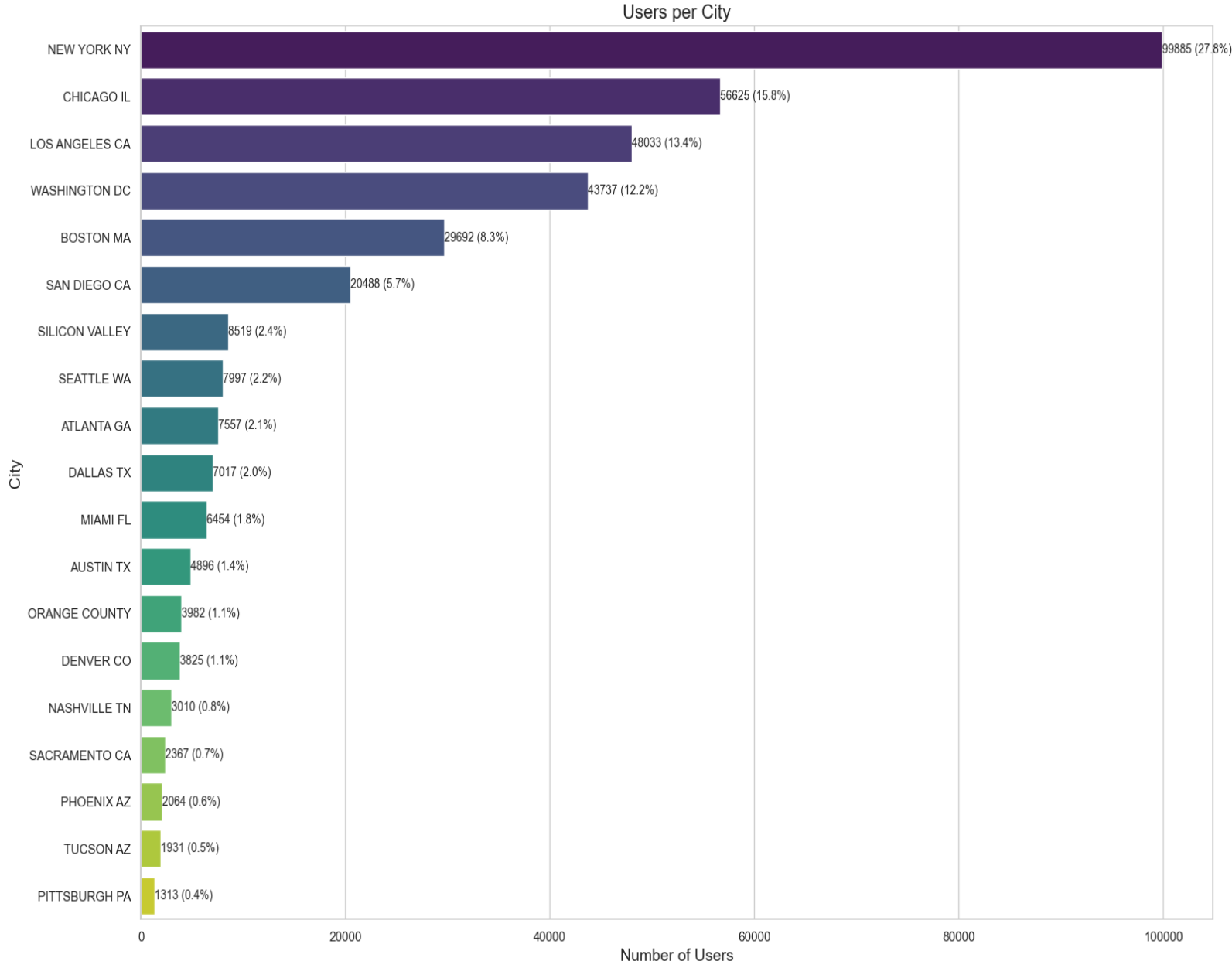
Users w.r.t Gender

- Male users prefer more to travel in Cab
- Also Users prefer to travel in Yellow Cab



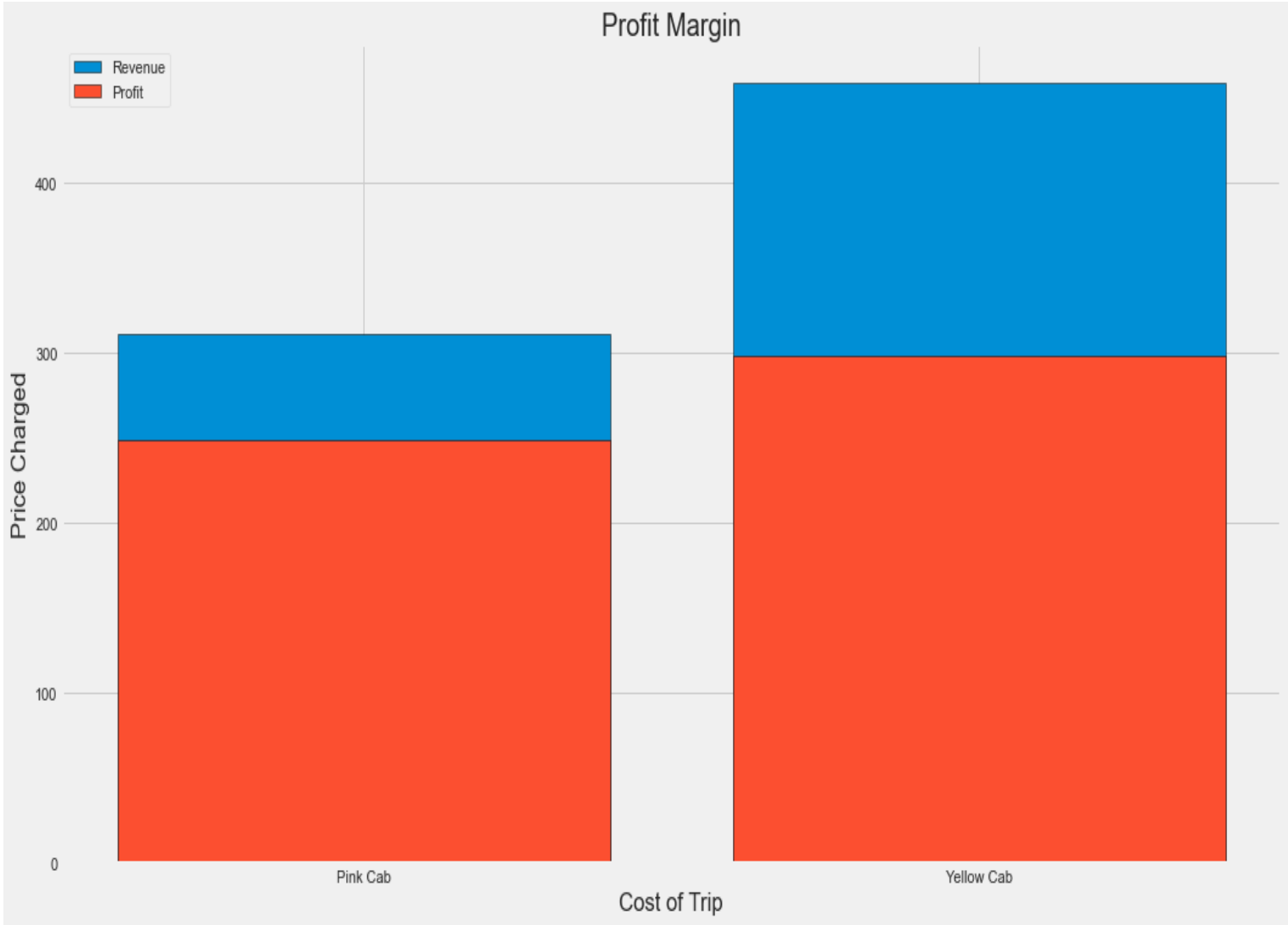
Users w.r.t Cities

- New York City has the highest Cab users with 28% followed by Chicago with 16% and Los Angeles with 13%



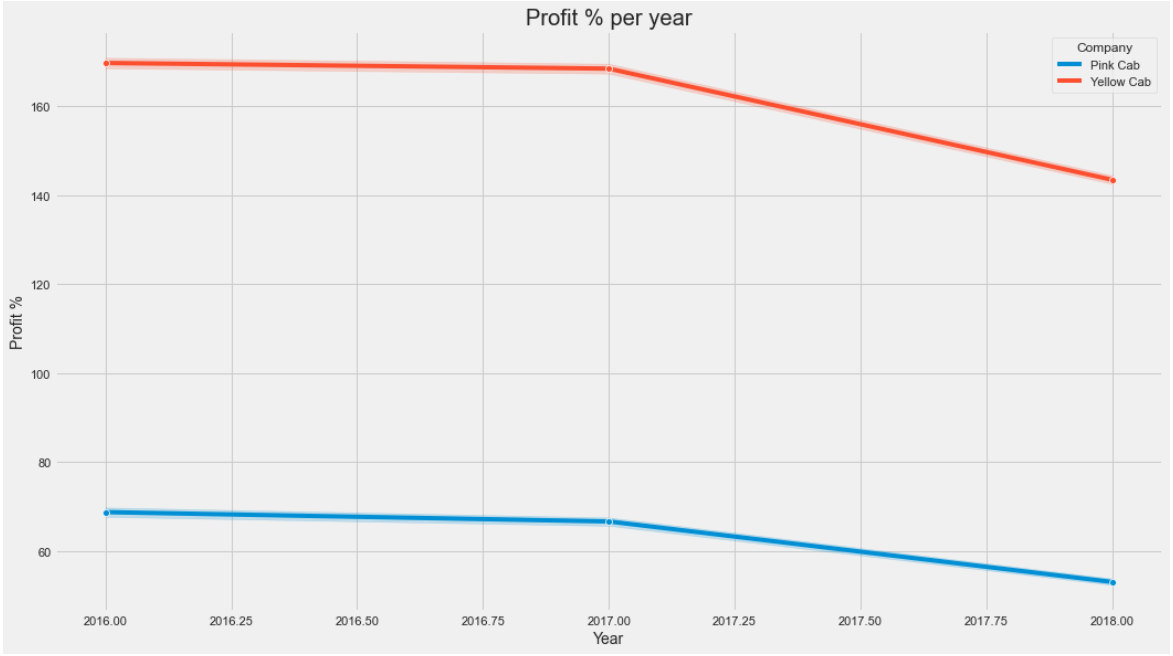
Profit Margin

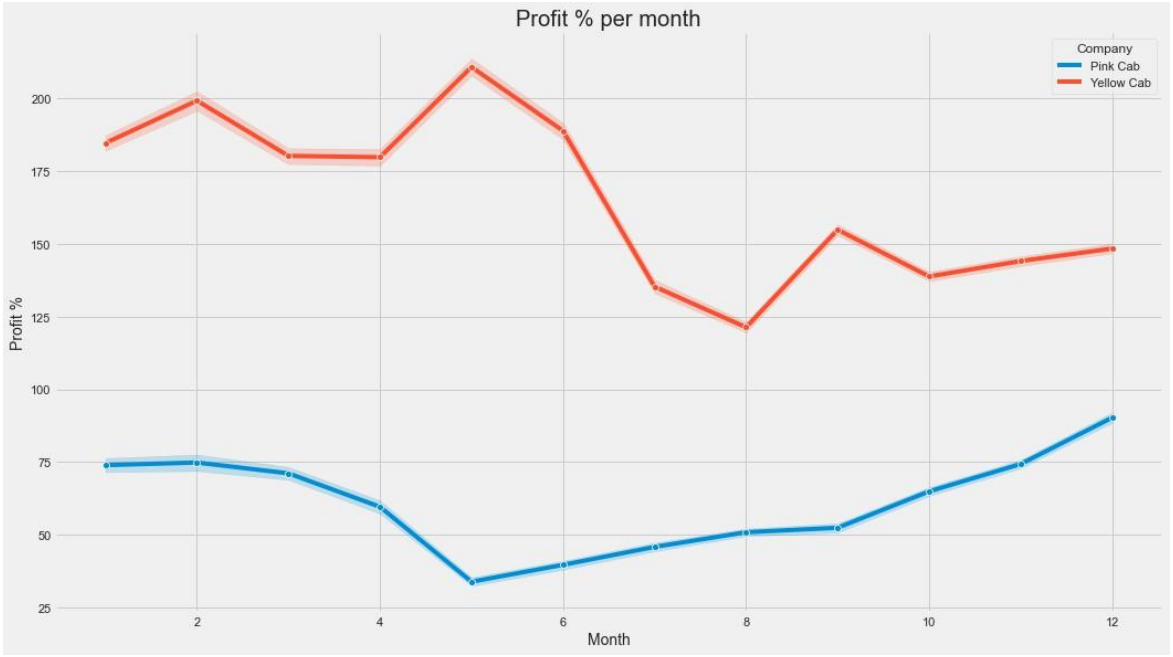
- The Yellow cab has a higher Profit Margin (Price Charged - Cost of Trip) compared to Pink cab



Profit Margin w.r.t Time

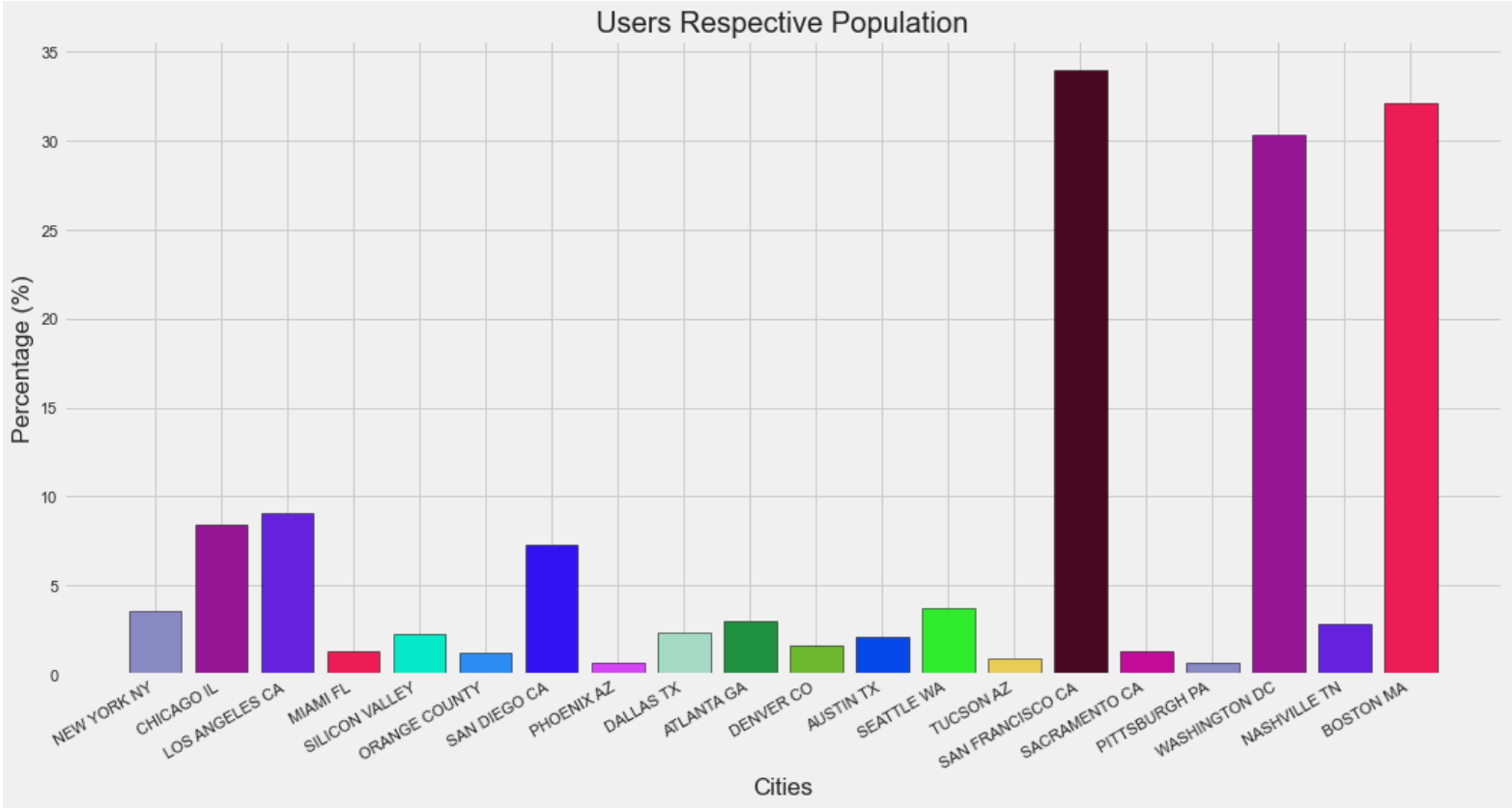
- The profit margin decreased w.r.t year
- The profit margin varies w.r.t month





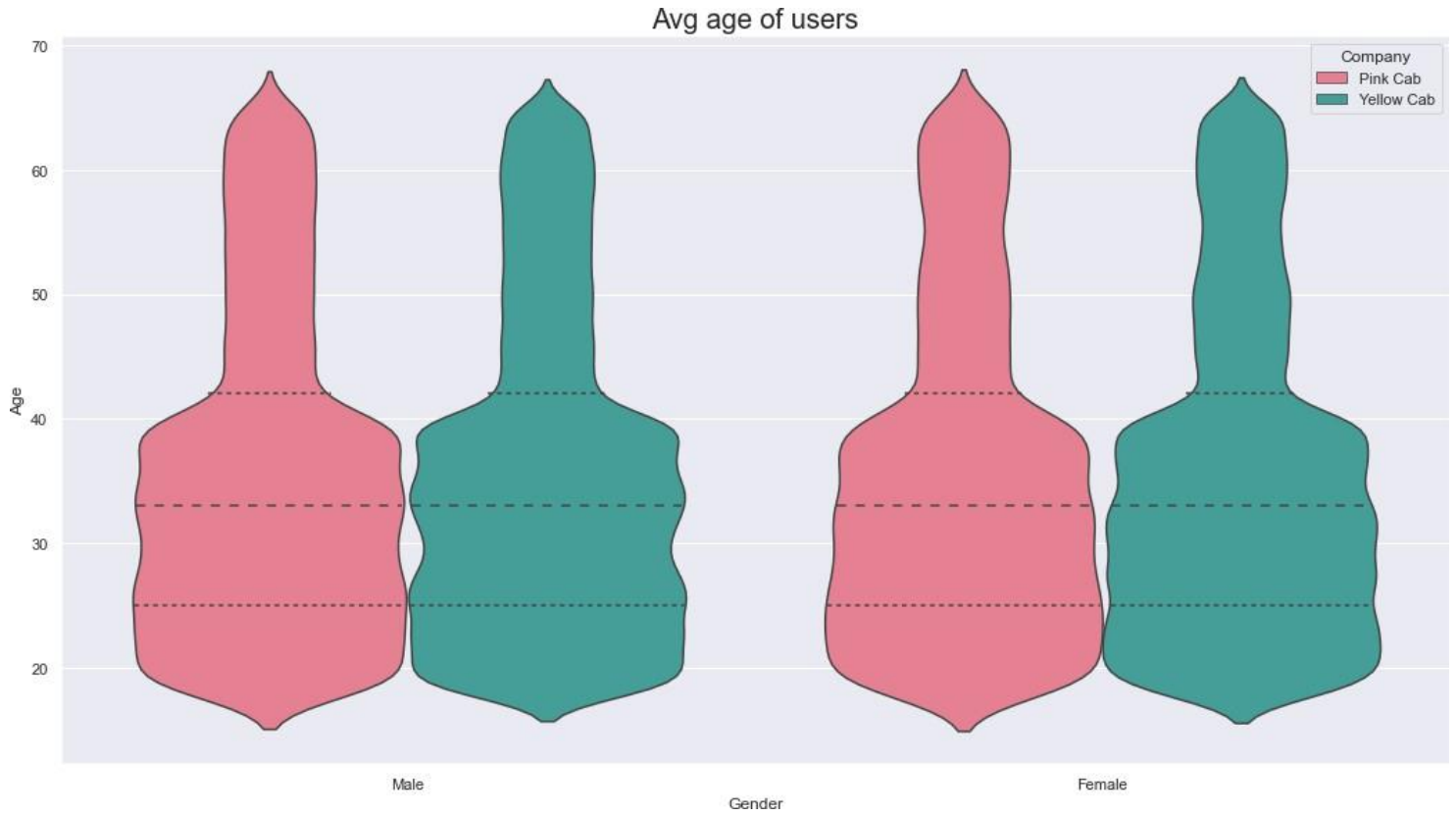
Users w.r.t Population

- As we can see in the cities of San Francisco, Washington, and Boston more than 30% of the population use cab service



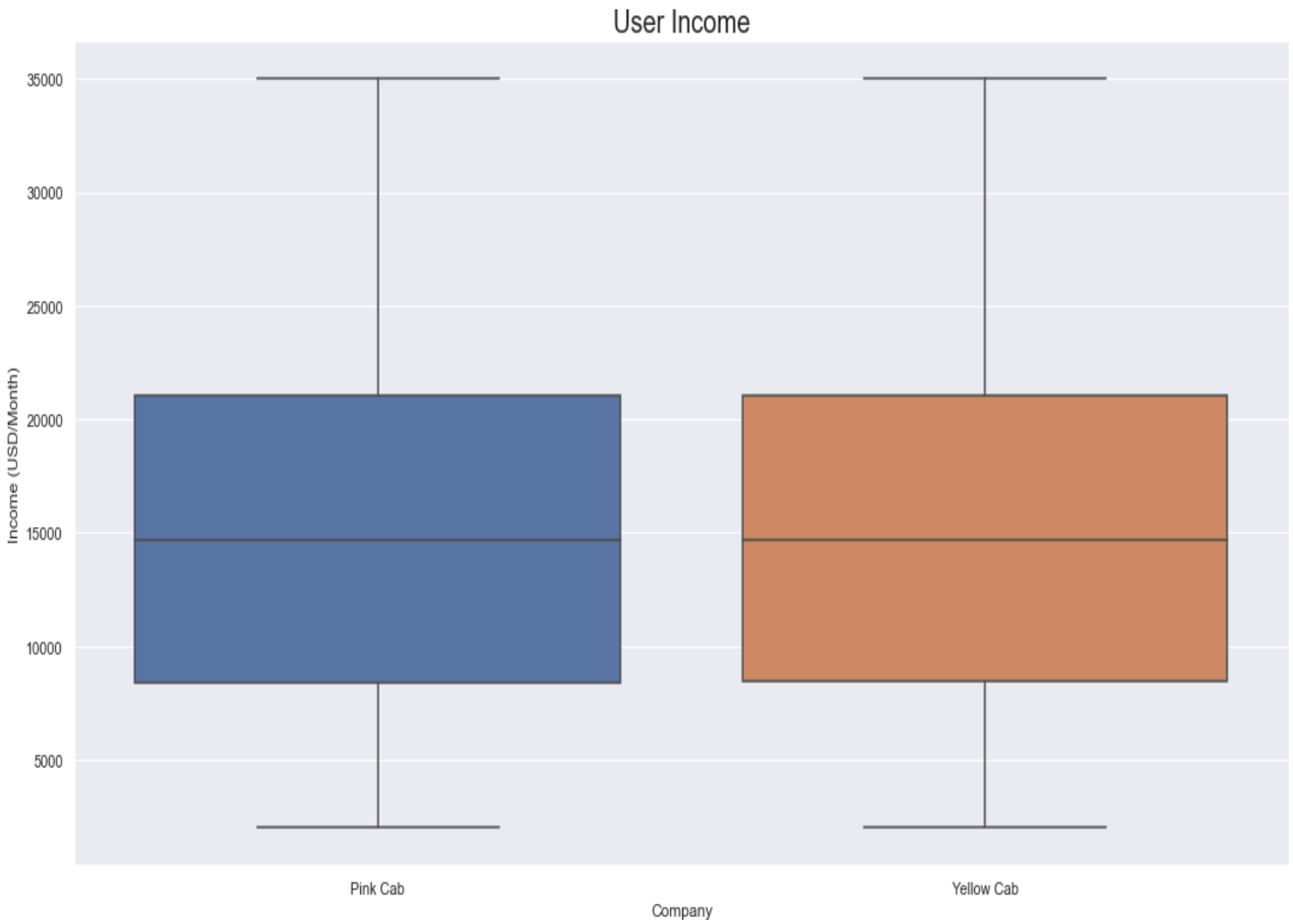
Average Age of Users

- As we can see 35 Avg age of females and Male who use Cab service



Average Income of Users

- As we can see Avg income is around 15k$ who use cab service



Price Charged w.r.t Distance

- As we can see there is a linear relationship between KM traveled and Price Charged as we expected. However, Yellow Cab has high charges compared to Pink.

Hypothesis 1: Is there any difference in profit regarding Gender

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

Pink Cab Company

```
P value is  0.11515305900425798
We accept null hypothesis (H0) that there is no difference regarding gender for Pink Cab
```

Yellow Cab Company

```
P value is  6.060473042494144e-25
We accept alternative hypothesis (H1) that there is a difference regarding gender for Yellow Cab
```

Conclusion

There is no difference regarding Gender in both cab companies

Hypothesis 2: Is there any difference in Profit regarding Age

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

Pink Cab Company

```
P value is  0.4816748536155635
We accept null hypothesis (H0) that there is no difference regarding age for Pink Cab
```

Yellow Cab Company

```
P value is  6.328485471267631e-05
We accept alternative hypothesis (H1) that there is a difference regarding age for Yellow Cab
```

Conclusion

Looks like Yellow Cab company offers discounts for their customers who are older than 60 years old.

Hypothesis 3: Is there any difference in Profit regarding Payment mode

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

Pink Cab Company

```
P value is  0.7900465828793288
We accept null hypothesis (H0) that there is no difference in payment mode for Pink Cab
```

Yellow Cab Company

```
P value is  0.2933060638298729
We accept null hypothesis (H0) that there is no difference in payment mode for Yellow Cab
```

Conclusion

There is no difference in payment mode for both cab companies

Conclusion

- Yellow Cab Company is better than Pink Cab Company Because:
  - Profit Margin
  - More Users
  - More transactions per Year

Thank You