



Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship, 21-Jan-2023

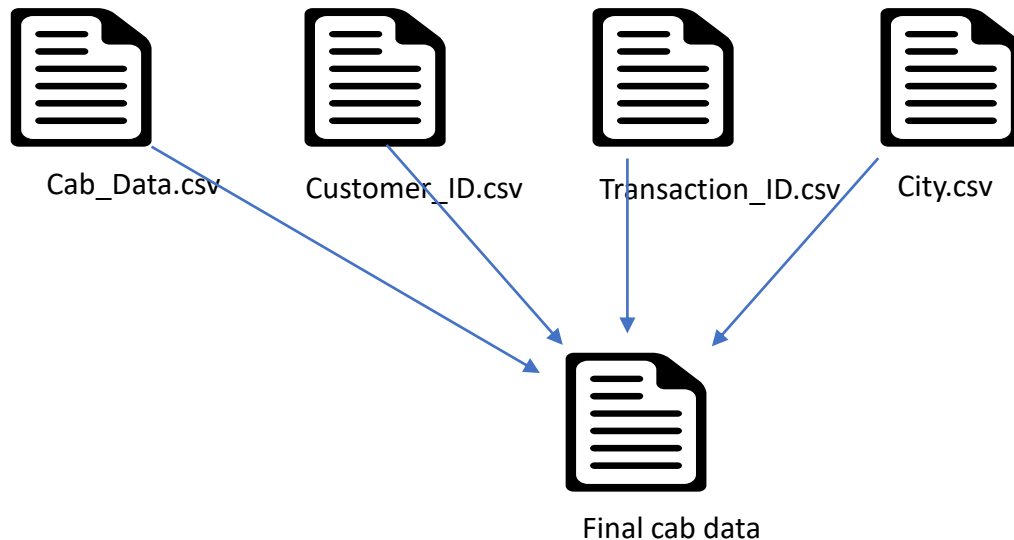
Vamsi Krishna Oruganti

Problem Statement

- XYZ is a private equity firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry.
- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

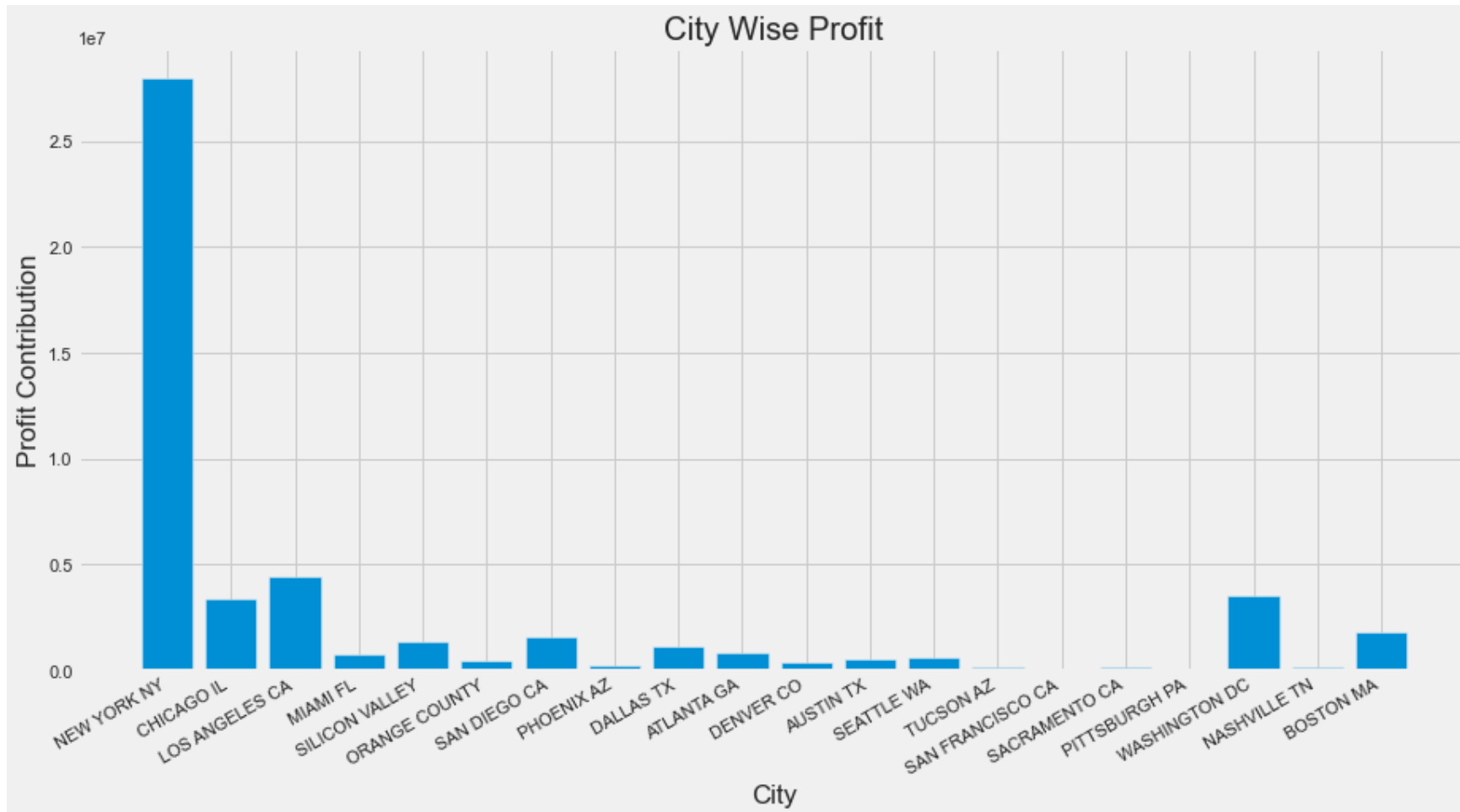
Data Exploration

Given data has 4 csv files and each of them are namely Cab, Customer ID, Transaction ID, City.



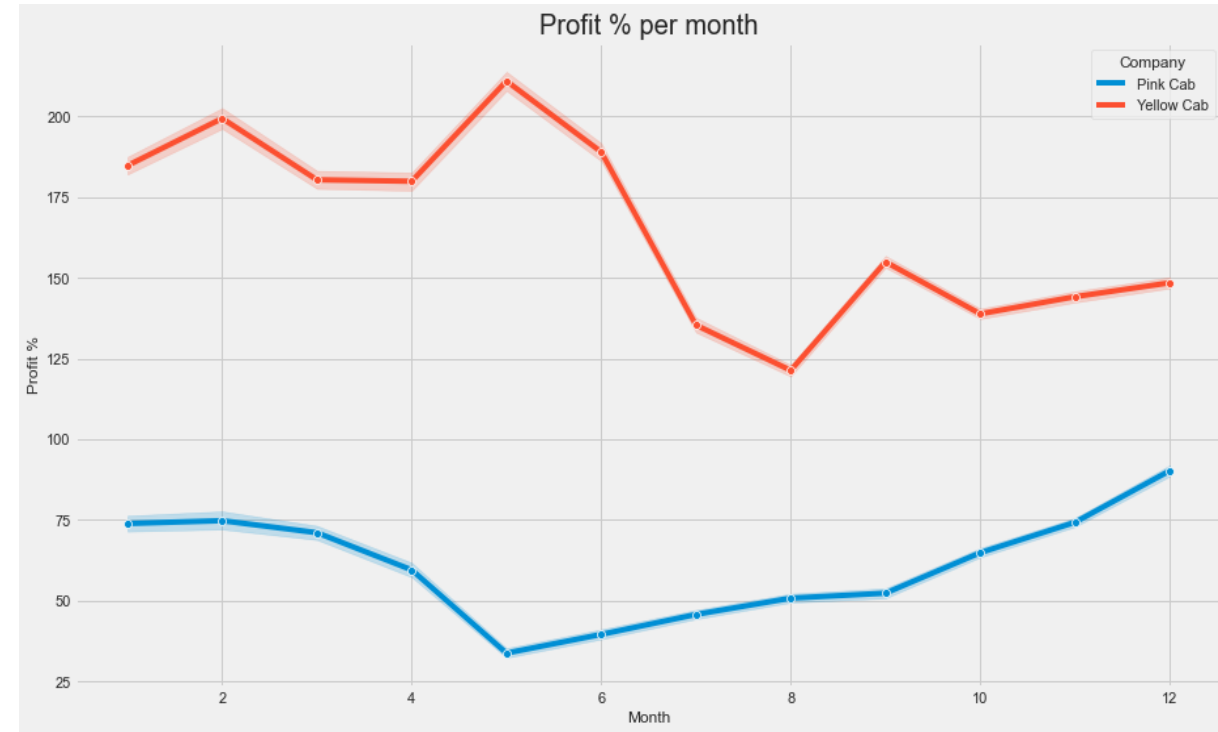
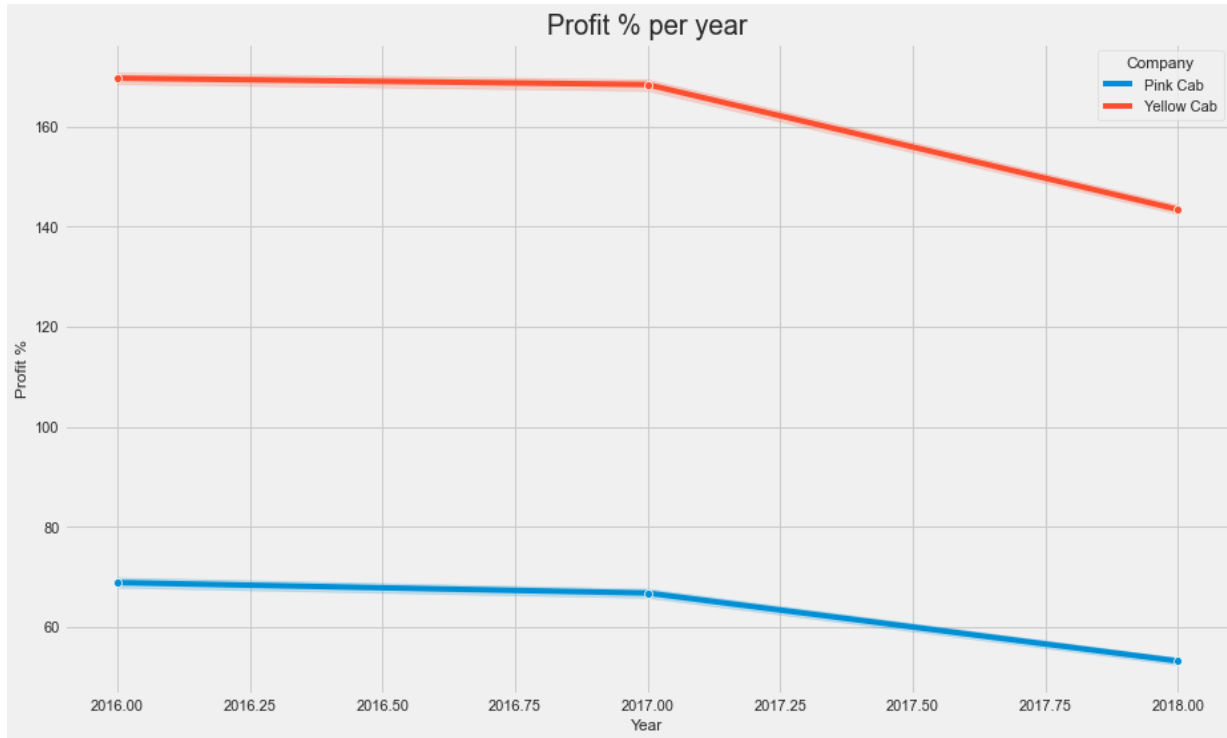
- **Cab_Data** – This file includes details of transactions for 2 cab companies and it has 7 columns and 359392 rows of data.
- **Customer_ID** – This file contains ID, Gender, Age and income of the customer and is comprised of 4 Columns and 49171 rows of data.
- **Transaction ID** – This is a mapping table that contains transaction ID, customer ID and payment mode with 3 columns and 440098 samples of data.
- **City** – This file contains a list of 20 US cities with the cities population, and the number of cab users in their respective cities.

Profit Analysis



- Here I have given the city wise profit information.
- In the x- axis is the city name and y- axis the profit contributed by the city.

Yearly and monthly Profit Analysis

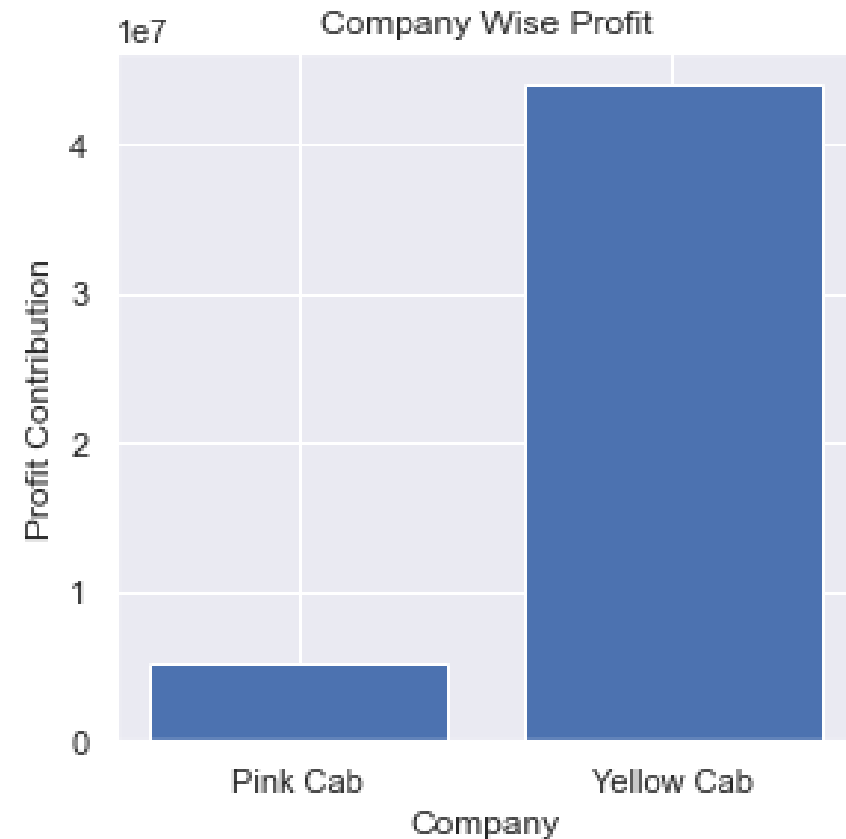


Yearly profit percentage and monthly profit percentage by both the pink cab and yellow cab is given.

Gender wise Profit and Company wise profit

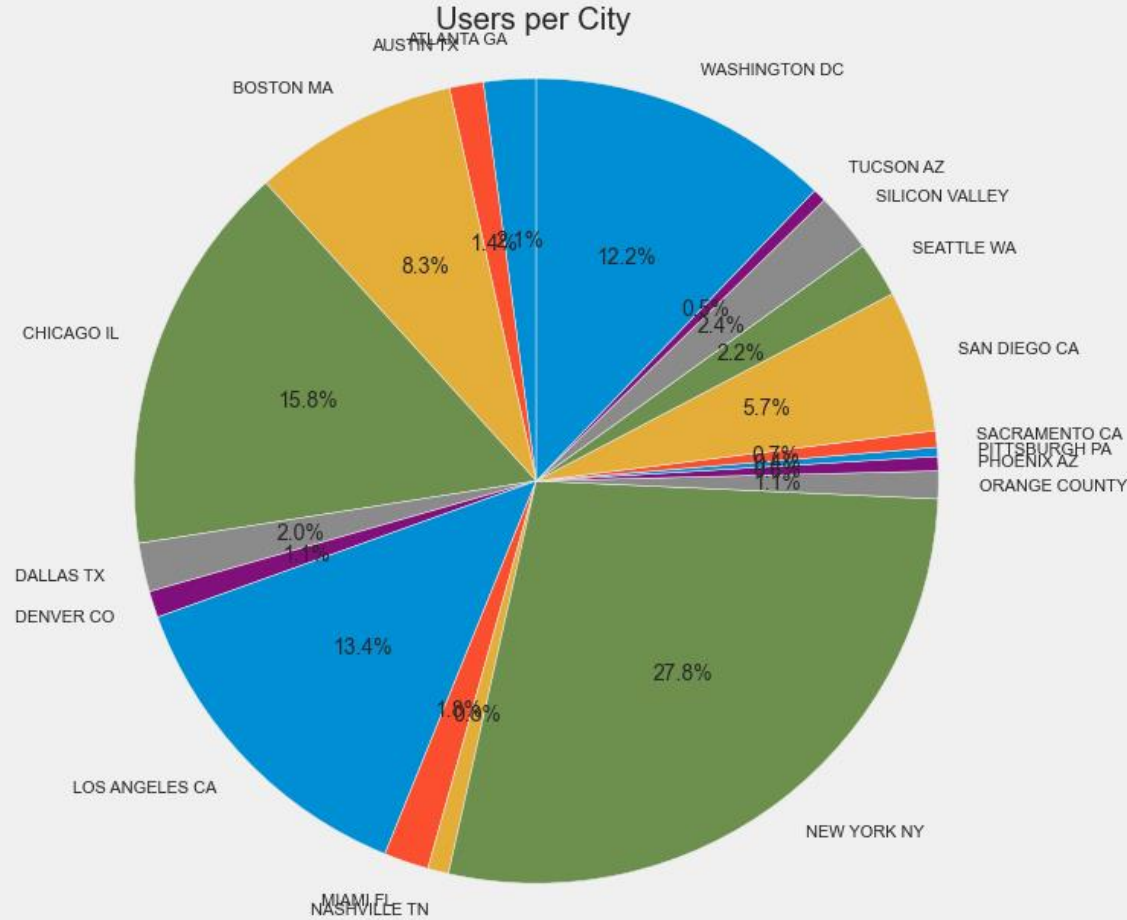


This is the gender wise profit analysis.



This is the profit earned by both the companies i.e Pink cab and Yellow cab.

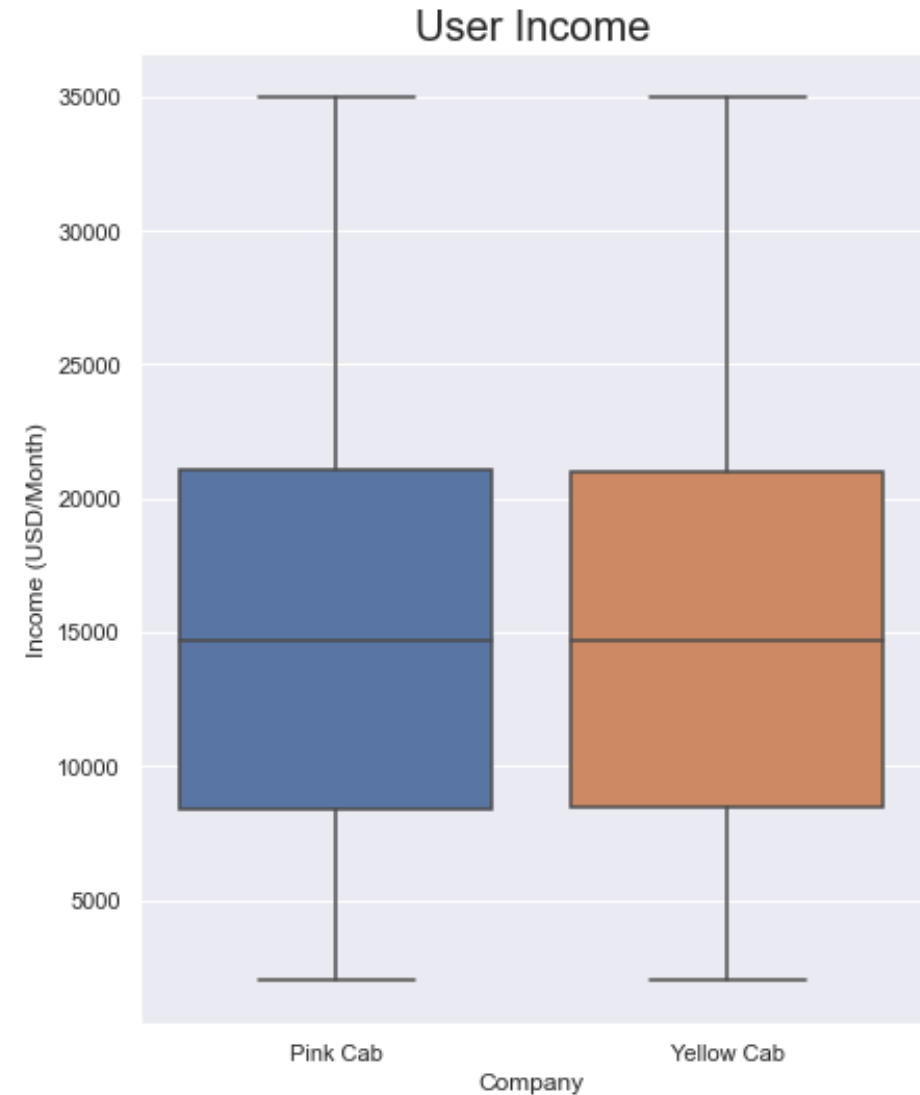
Users per city



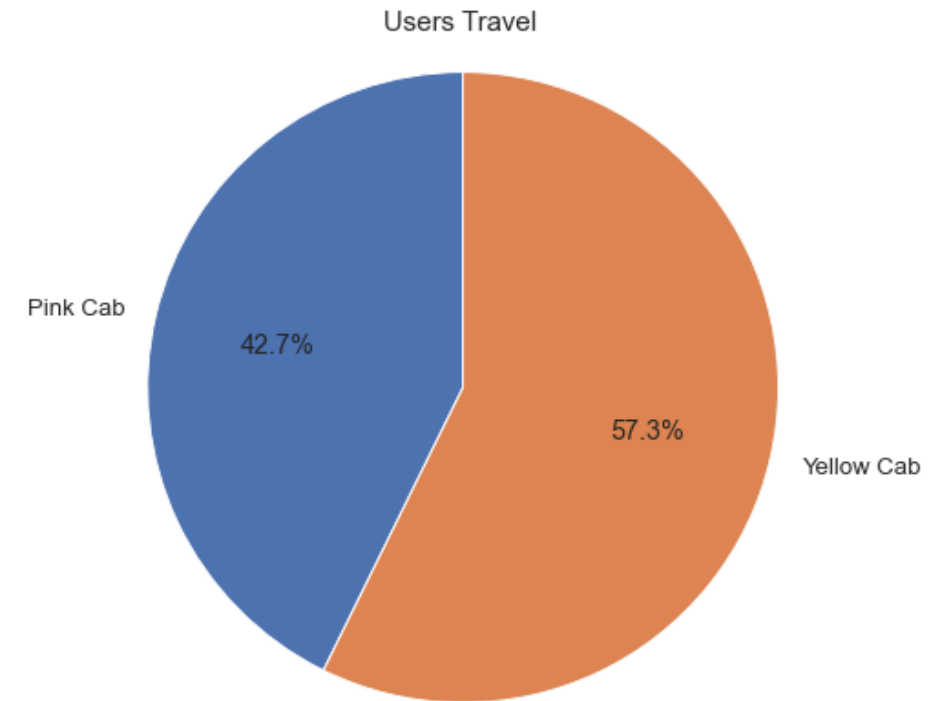
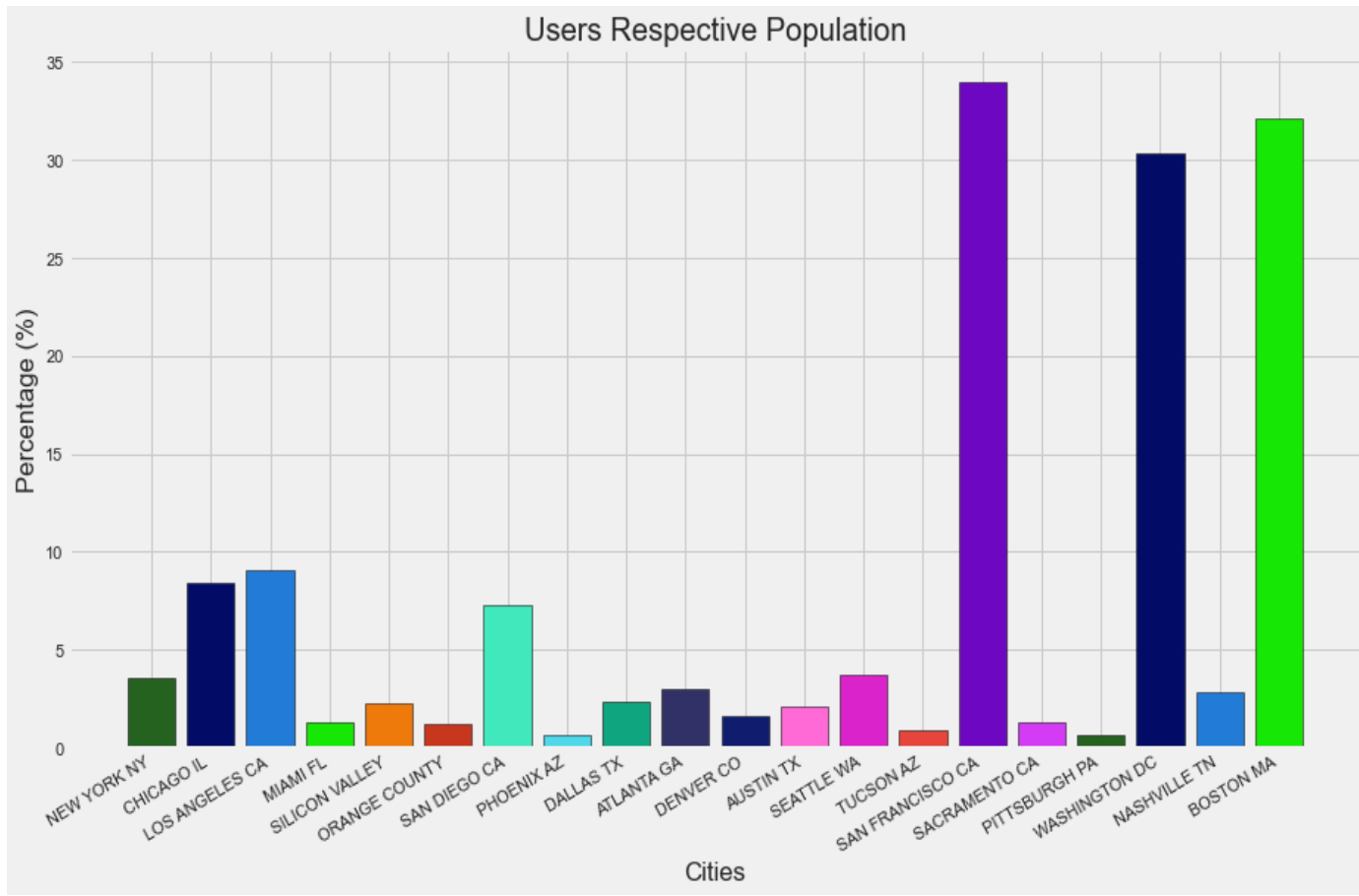
- This pie chart shows number of Cab users per city. And it consists of both Pink cab and Yellow cab Users per city.
- Altogether, the cab users for 20 cities are given in the chart.
- New York residents tops the pie with users of 27.8% , followed by Chicago with 15.8% in 2nd place and Los Angeles with 13.4% in the 3rd place.

User Income

- Clearly the average income of the cab users either Pink cab or Yellow cab is around **15000\$**.
- Here the boxplot is used for this analysis.

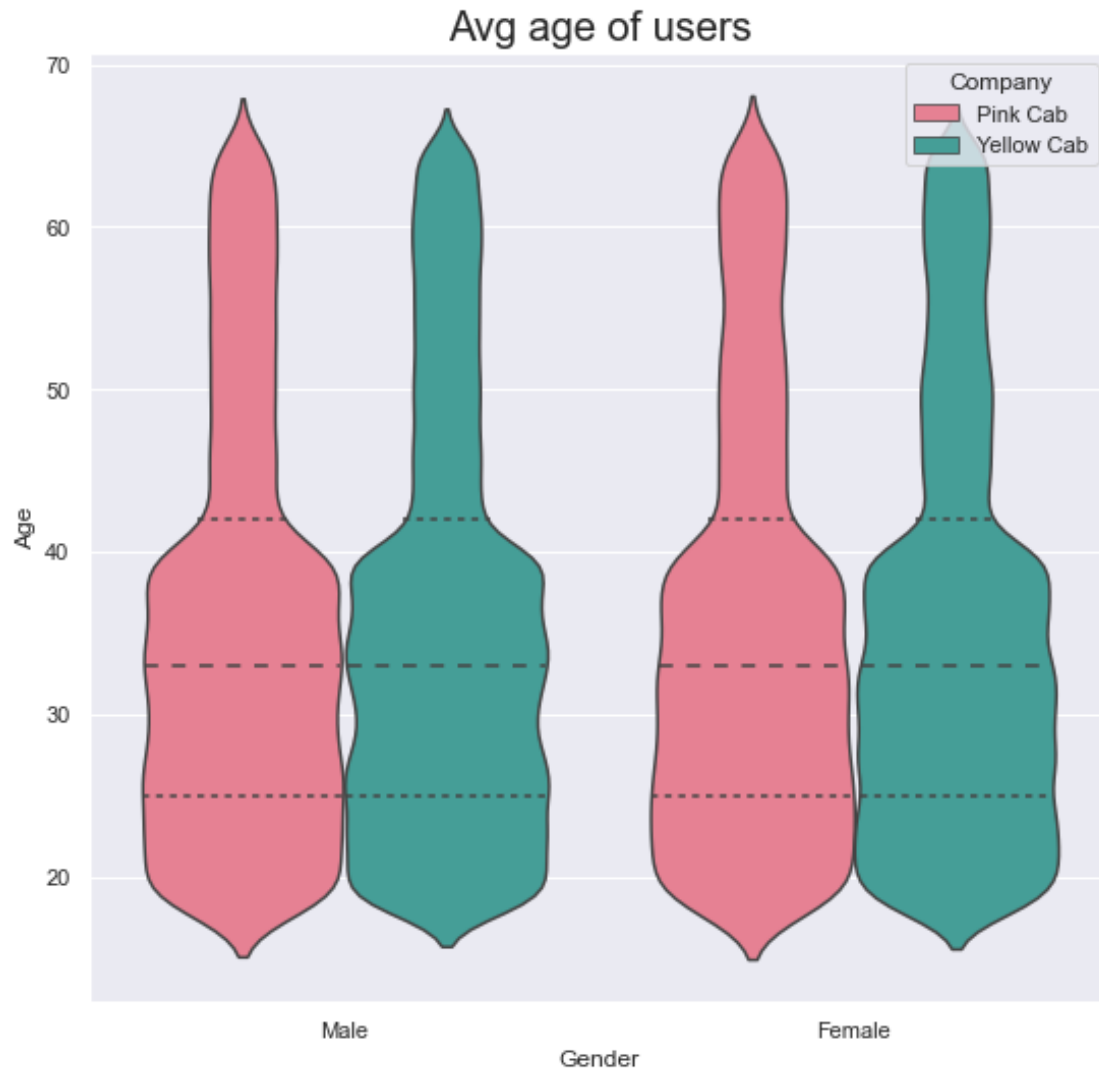


Users population and Users Travel



The Bar graph contains the users population for each city and the overall travel percentage of which cab they uses is given the pie chart.

Average age of the users

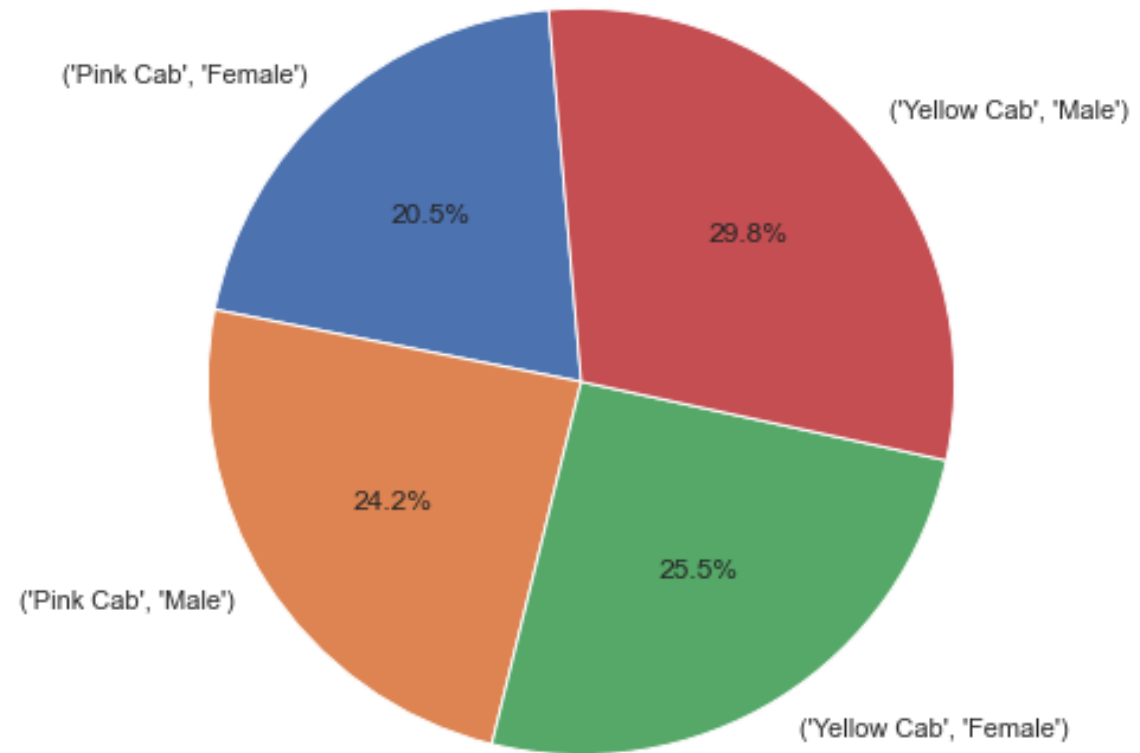


- Average age of the cab users with respect to the gender is given in the **Violin plot**.
- We can see that the average age of the both male and female users is between 30 and 40.

Customer share

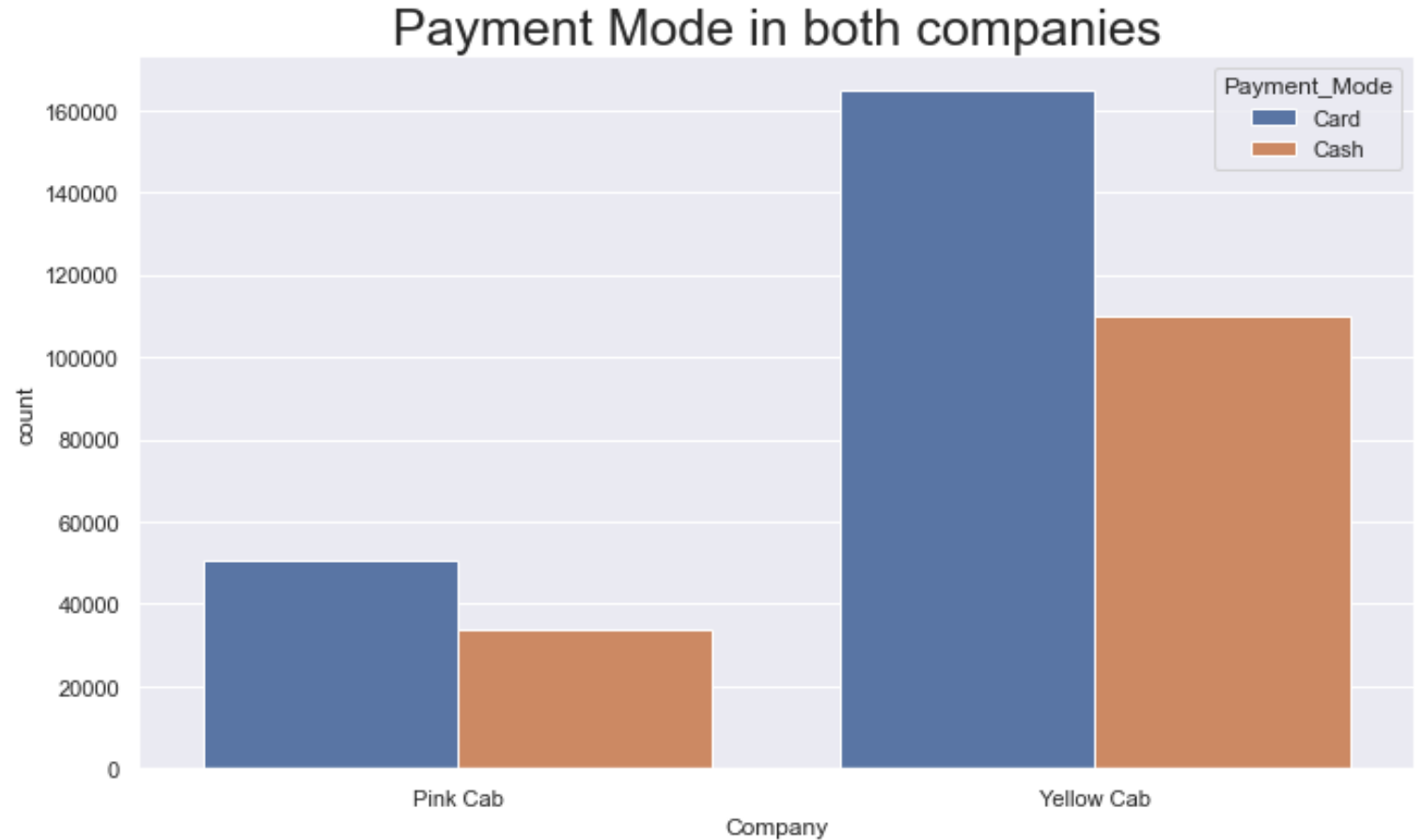
- This pie chart shows the users share as per gender.
- Overall males use cab service more than females.
- Both male and females yellow cab more than pink cab.

Customer share per gender per each cab

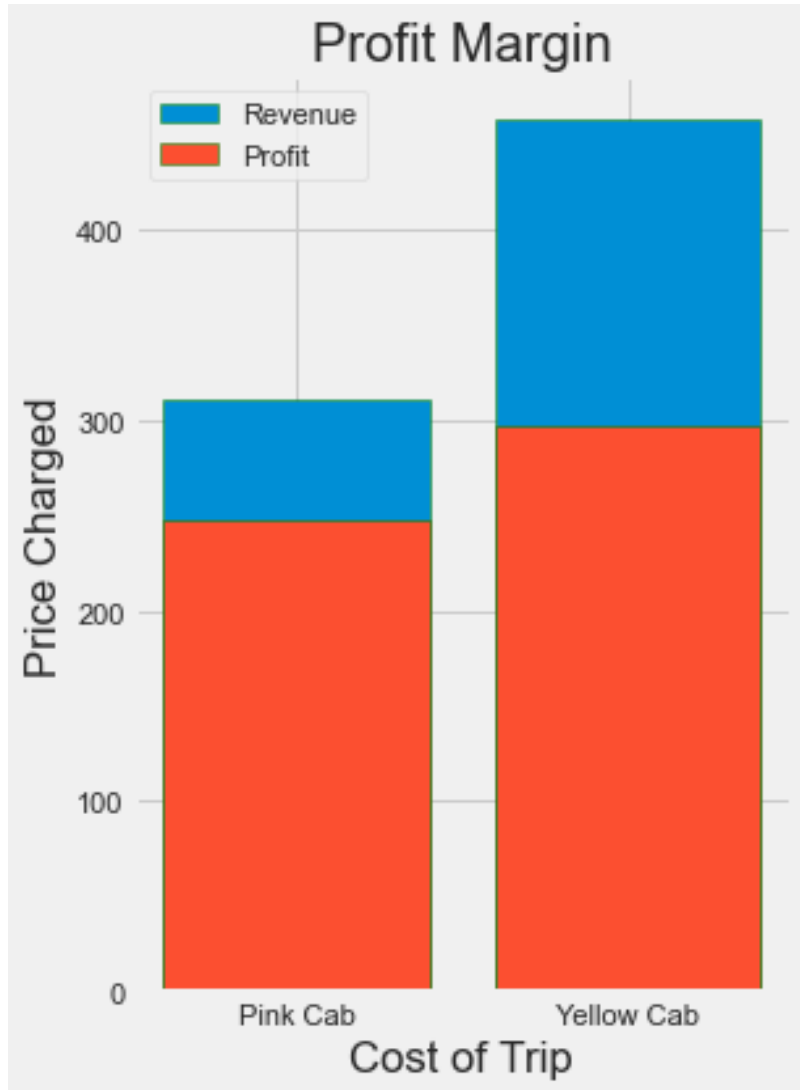


Payment methods

- Both yellow cab and pink cab companies receive more card payments to cash payments.
- Yellow cab card payments count is greater than 160000, cash payments count around 110000.
- Pink cab card payments count is around 50000 and cash payments count is around 30000.

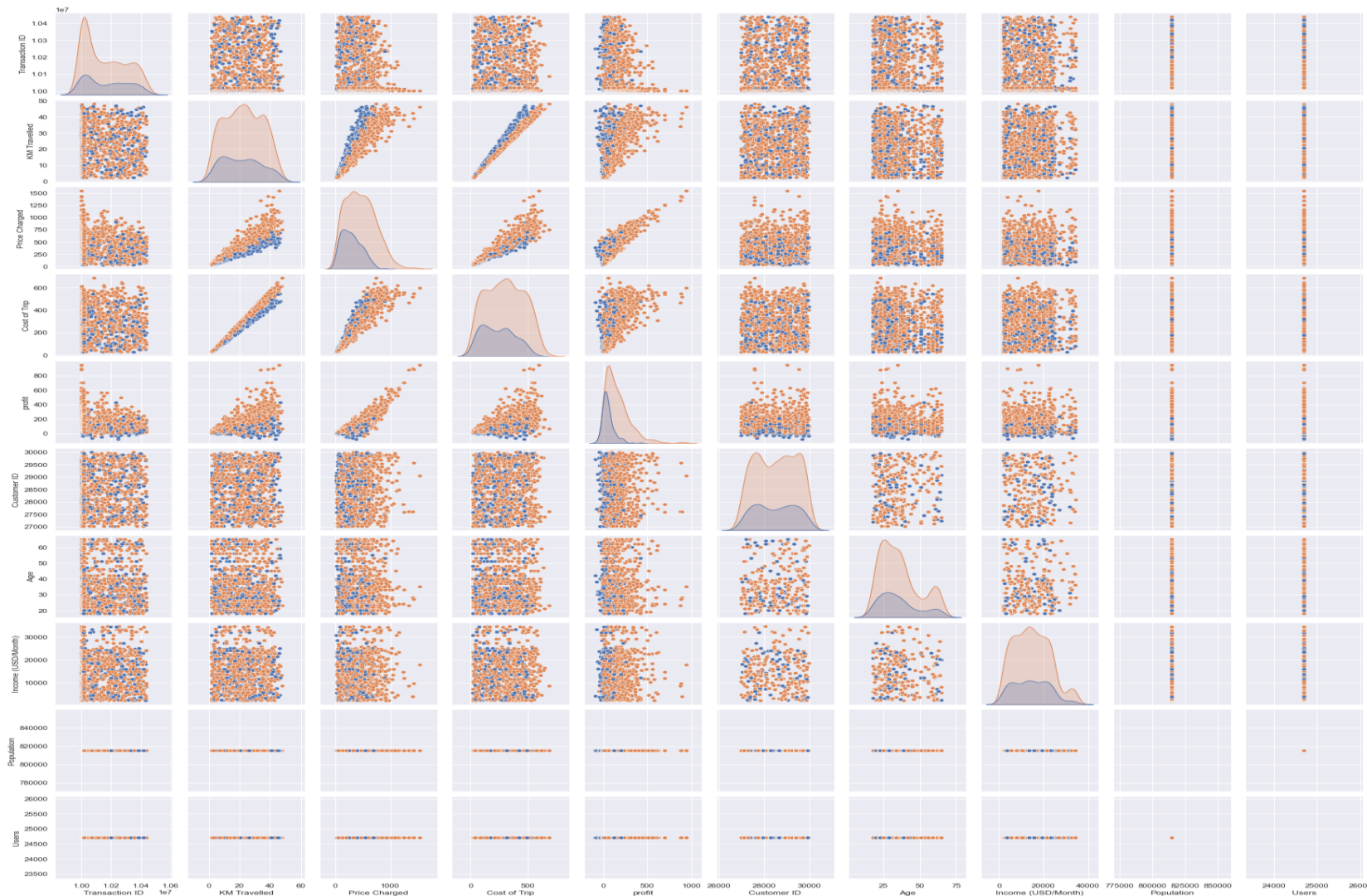


Profit margin



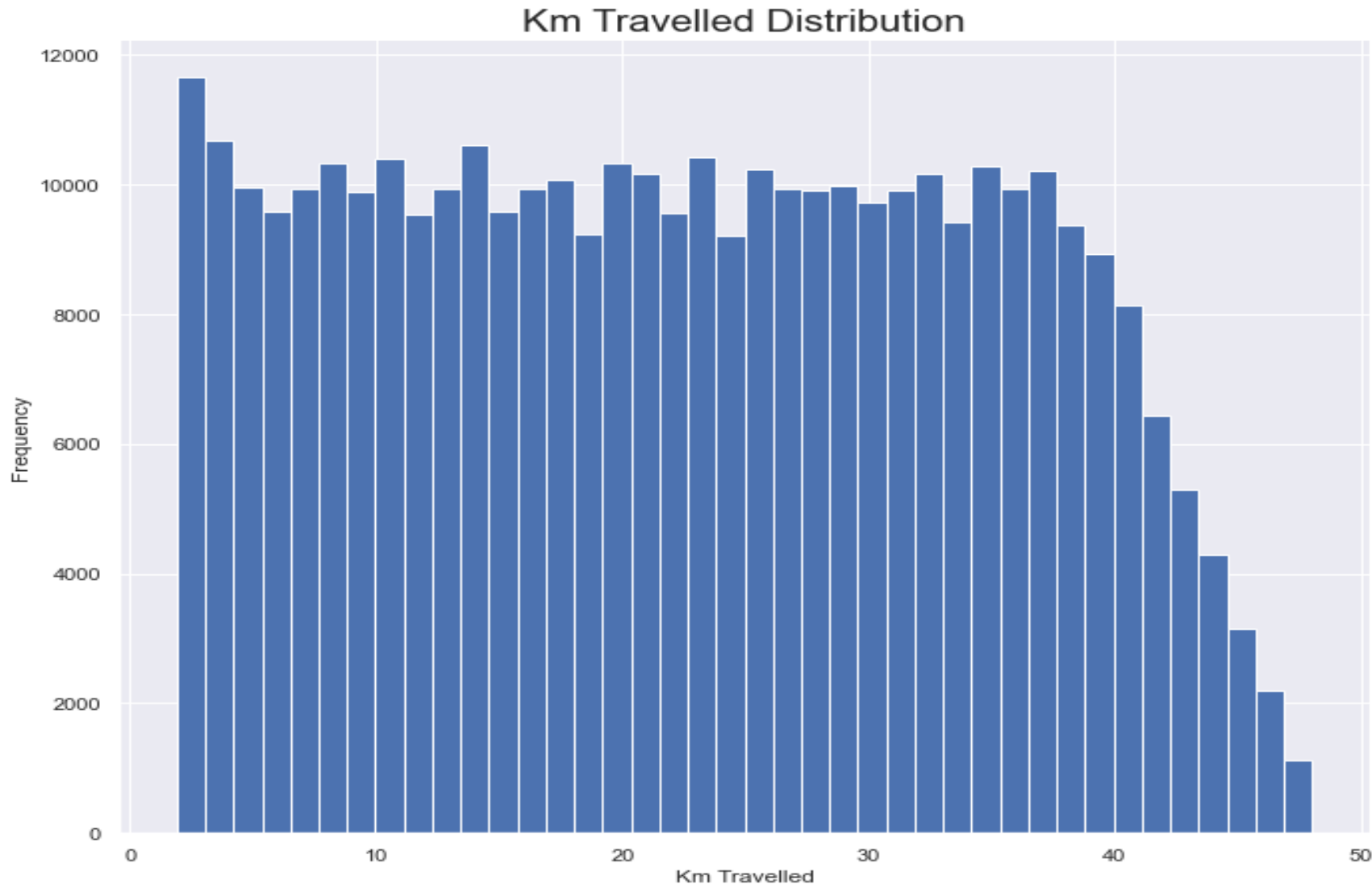
- As per the charts pink cab gets around 83.3% profit from the revenue it generates.
- Where as Yellow cab gets around 66.67% profits from the revenue it generates.
- Overall yellow cabs revenue and profits are more than the pink cab.

Relation between Variables



- This is the relationship between many variables in the data sets.
- The variables are **Users, Population, Income, Age, Customer ID, Profit, Cost of the trip, Price charged, KM Travelled, Transaction ID.**

KM Travelled



- This bar chart shows how long distances are frequently travelled.
- Most of the rides varies from around 2km and approx. 48 km.

Correlation

Correlation Heatmap of "G2M Insight for Cab Investment", fontsize = 20



This is a correlation between many variables.

Hypothesis

Hypothesis 1: Is there any difference in profit regarding Gender

- **Null:** There is no difference regarding Gender in both cab companies.
- **Alt:** There is a difference regarding Gender in both cab companies.
- **Conclusion:** There is no difference in profit basing on gender

Pink Cab :

P value is 0.11515305900425798

We accept null hypothesis (Null)

Yellow Cab:

P value is 6.060473042494144e-25

We accept alternative hypothesis (Alt)

Hypothesis

Hypothesis 2: Is there any difference in Profit regarding Age.

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

Pink Cab :

P value is 0.4816748536155635
We accept null hypothesis (Null)

Yellow Cab:

P value is 6.328485471267631e-05
We accept alternative hypothesis (Alt)

Hypothesis

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

Conclusion: There is no difference in profit basing on gender

Pink Cab :

P value is 0.7900465828793288
We accept null hypothesis (Null)

Yellow Cab:

P value is 0.2933060638298729
We accept null hypothesis (Null)

Conclusion

- From the analysis it is clear that Yellow cab has more number of users in each of the 20 states in USA.
- Yellow cab makes more profit compared to pink cab.

Thank You