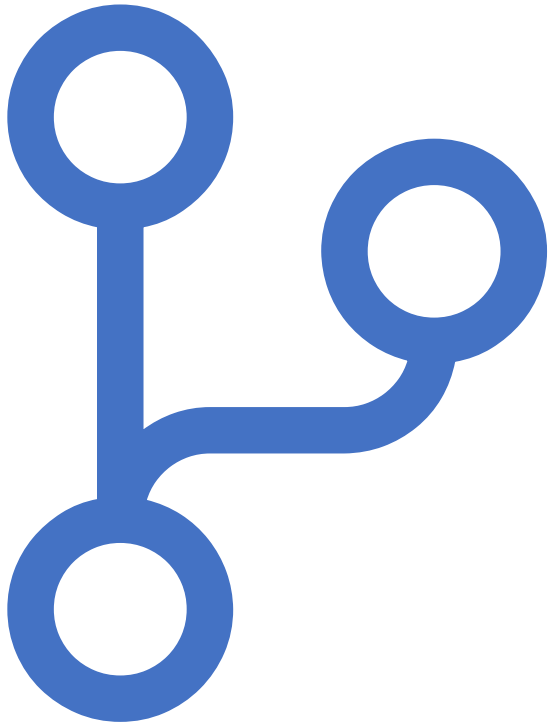
G2M insight for Cab Investment firm

Company Name : XYZ
Location: USA
Team: Data and Analytics
Date: 12-FEB-2024

Content



- Dataset information
- Problem Statement
- Exploratory data analysis
- EDA Summary
- Hypothesis Testing
- Conclusion.

Data Glacier

Description:

- **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.

- Cab Companies:
  - Yellow Cab.
  - Pink Cab.

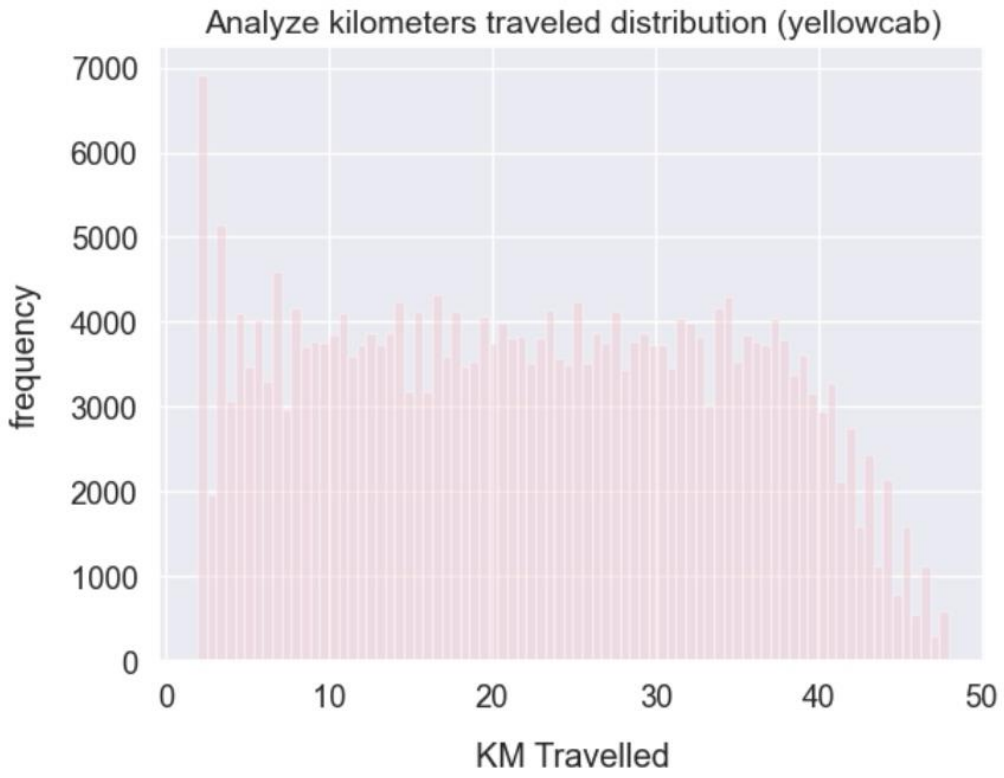
- The Analysis include :
  - Data Understanding.
  - Data Visualization .
  - Creating multiple hypothesis.

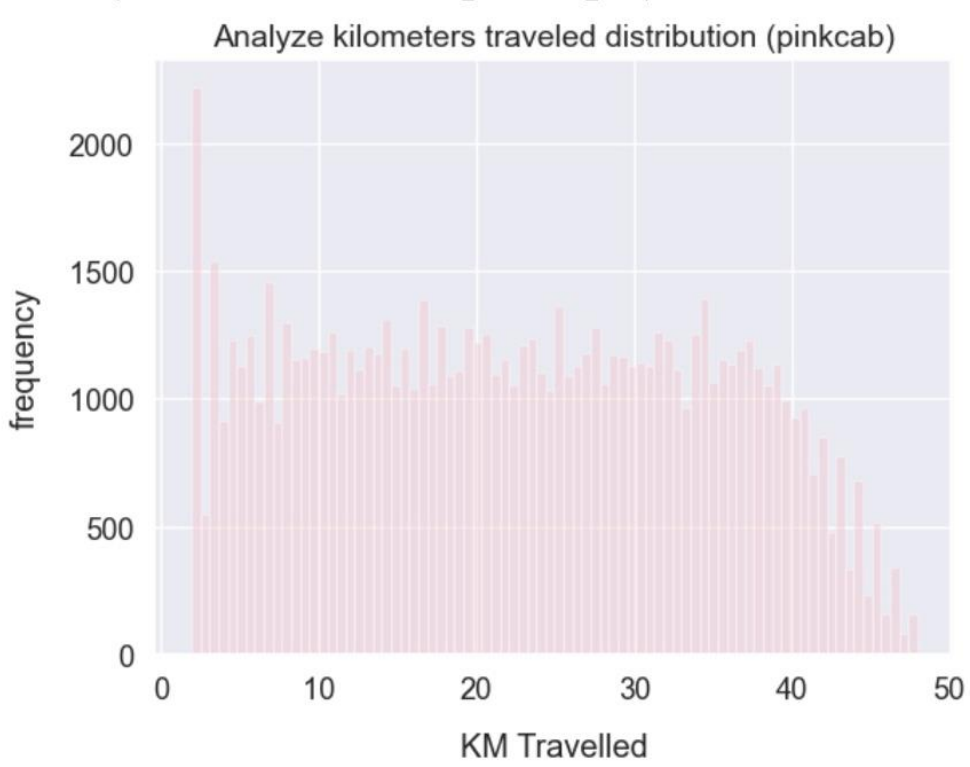
Data Preparation:

There are 4 datasets:

- Cab_Data.csv – this file includes details of transaction for 2 cab companies.
- Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details.
- Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode.
- City.csv – this file contains list of US cities, their population and number of cab users.

EXPLORATORY DATA ANALYSIS



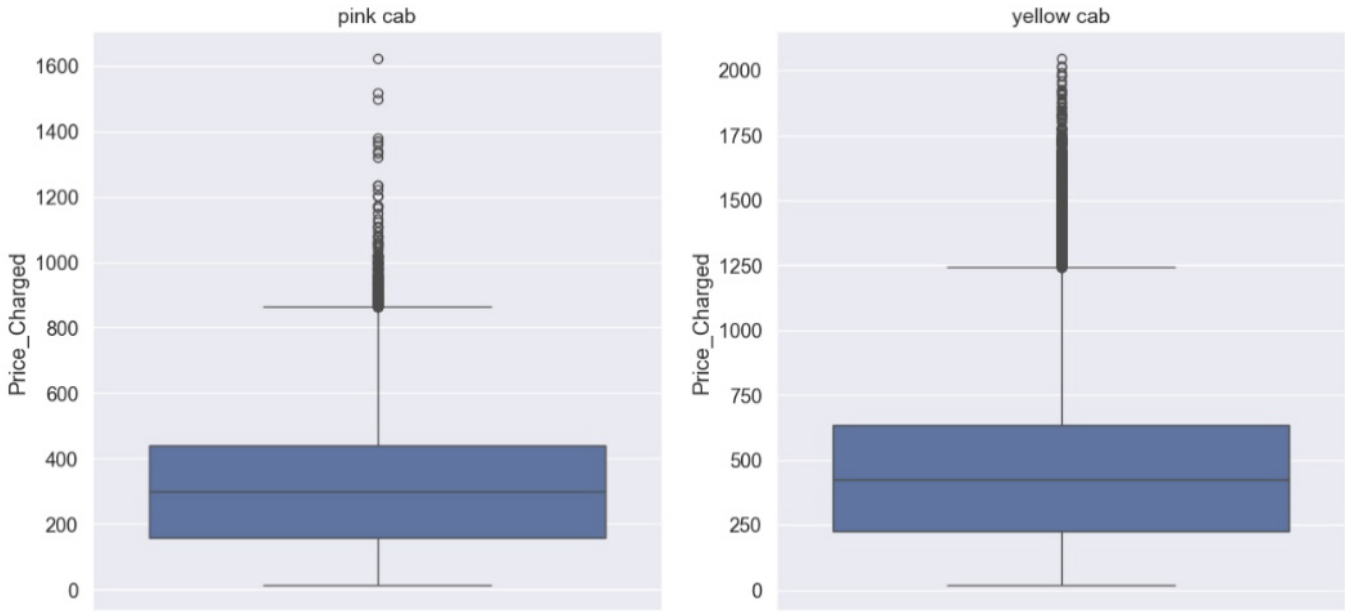


<u>The distribution of kilometers travelled for both cabs is as follows:</u>

- The figures above show that the majority of trips for both Pink and Yellow Cab fall between around 2 and 48 kilometers.

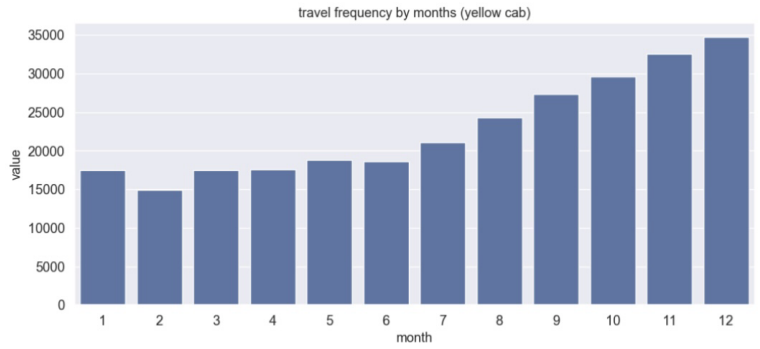
Distribution of Price Charged for both Cabs:

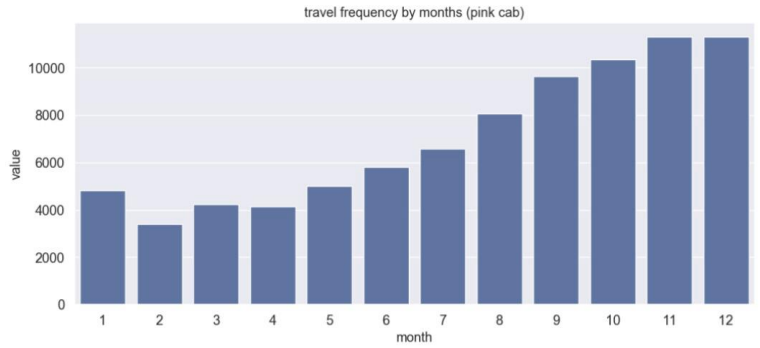
- **The Yellow Cab has a higher price range than the Pink Cab.**
- **The use of expensive cars is the cause of the outliers.**



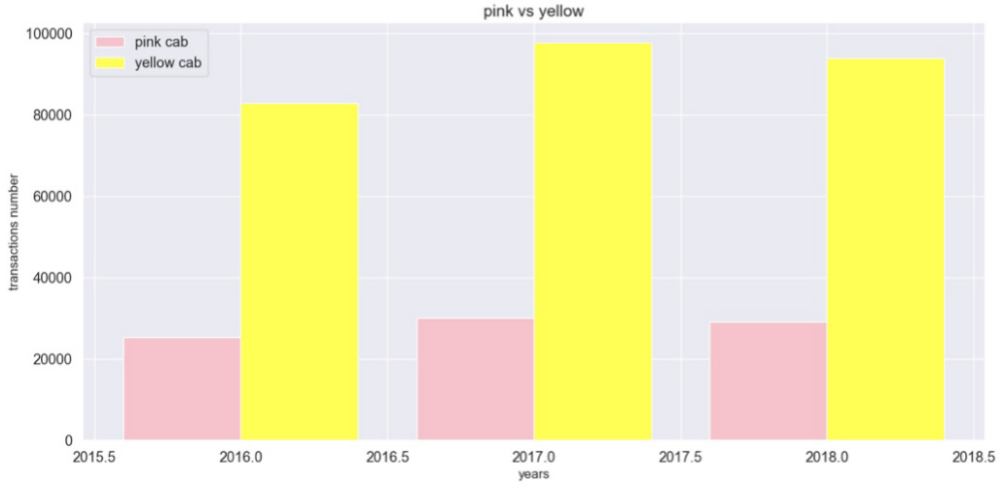
Travel Frequency per Month:

- **In December, when everyone is celebrating the holidays, Yellow Cab travels more (35,000) than Pink Cab (11000).**





Transaction per Year for both Cabs:



- The graph indicates that Yellow Cab has more transactions than Pink Cab on an annual basis.

Pink Cab:The cost per kilometer within a specific city is assessed at a fixed rate per kilometer.

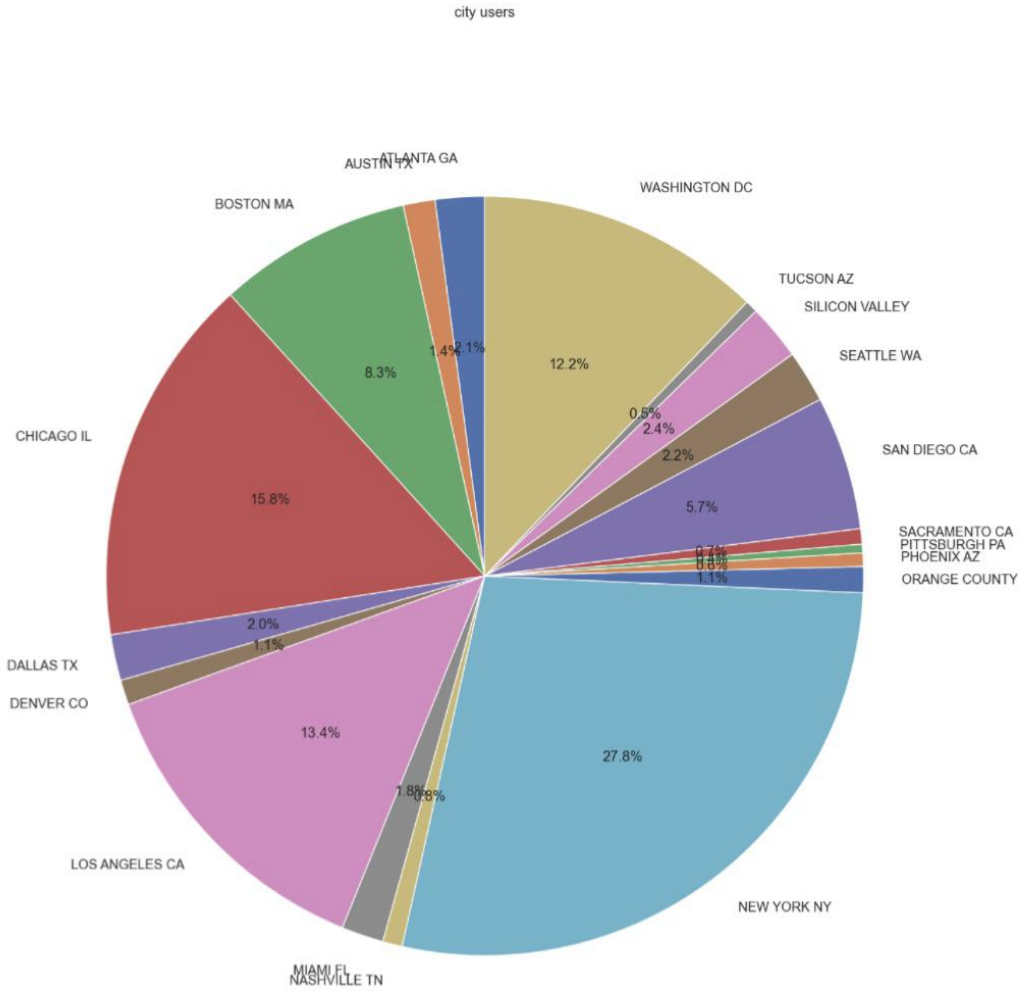
- Every city has the same pricing rise for Pink Cab as the distance traveled increases.

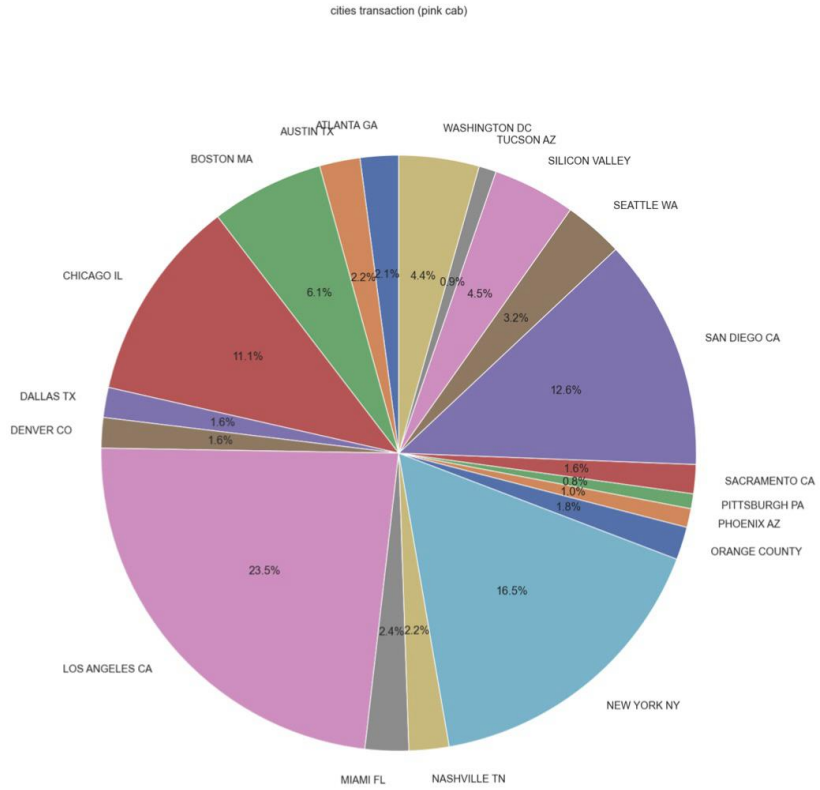
Yellow Cab: The cost per kilometer within a specific city is assessed at a fixed rate per kilometer.

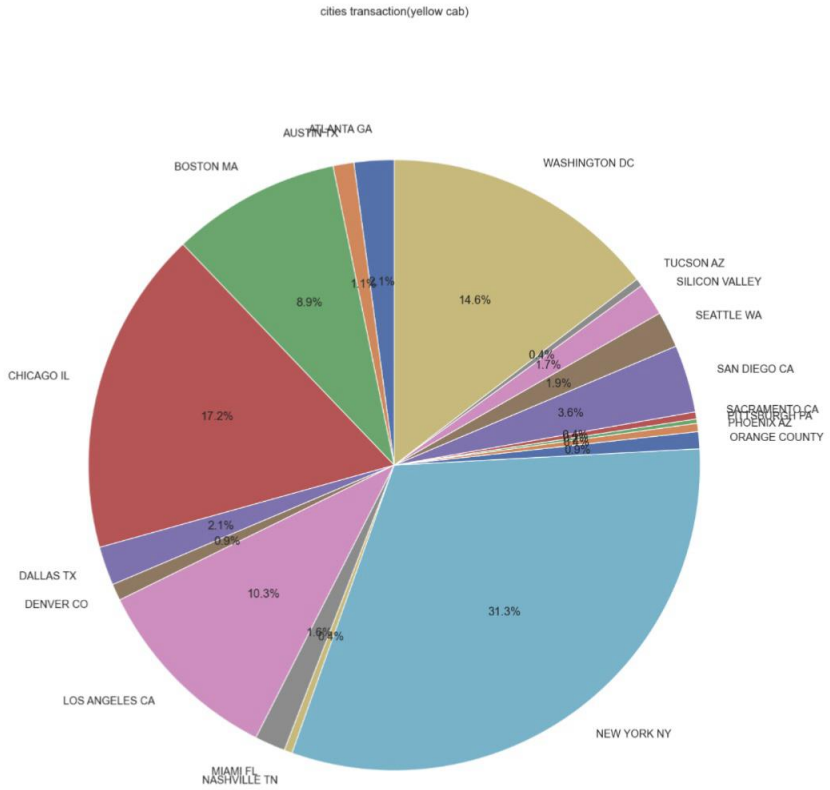
- In New York City, the cost of using a yellow cab is higher than in other cities.

Cab users in every city:

- According to recent data, New York City boasts the largest proportion of ride-hailing users, comprising approximately 28% of the total market, followed closely by Chicago with 16% and Los Angeles with 13%.

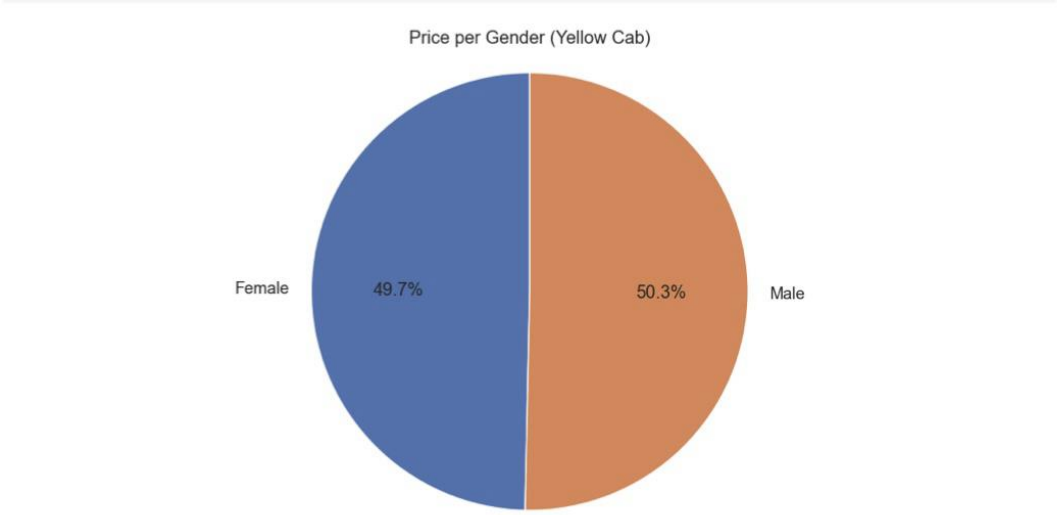


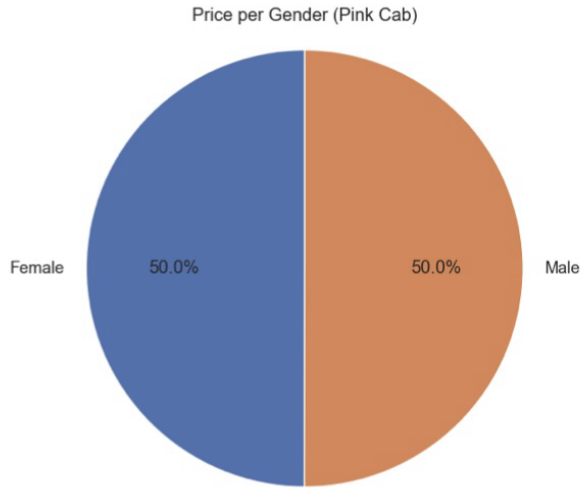




Transaction per City for both Cabs:

- According to the provided information, the transaction volume for Yellow Cab is highest in New York City, with a percentage of 31%. Additionally, New York City has the highest percentage of cab users, accounting for 28% of the total. On the other hand, the transaction volume for Pink Cab is highest in Los Angeles City.

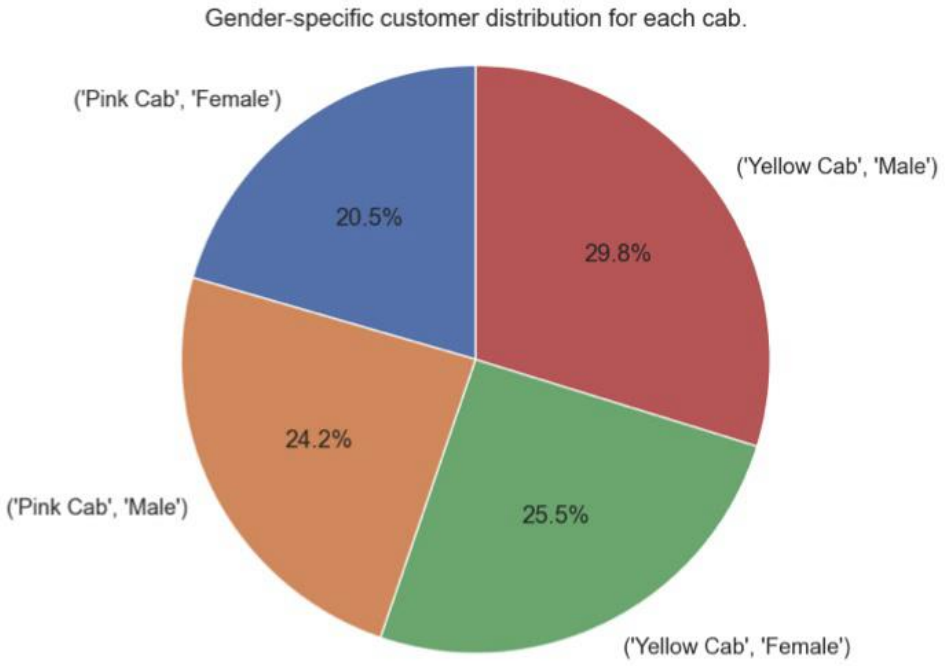




Price Charged per Gender for both Cabs:

- Yellow Cab has been observed to charge lower fares for female customers compared to Pink Cab, which charges the same fare for both male and female customers.

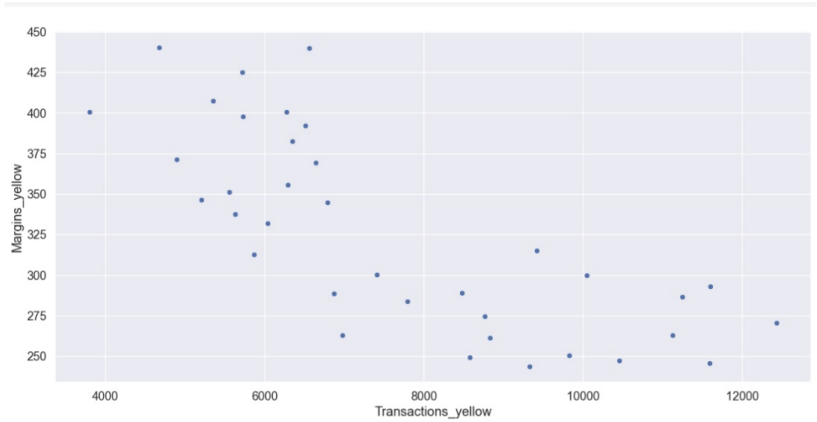
Customer Share per Gender for both Cabs:

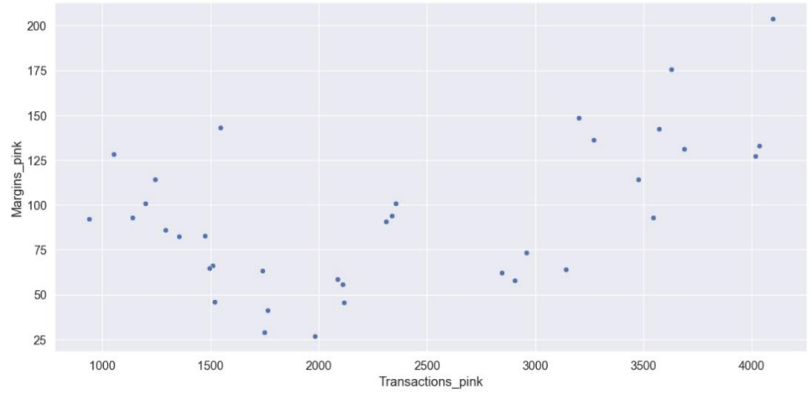


- According to the data, there is a slightly higher percentage of female customers among Yellow Cab users (25.5%) compared to Pink Cab users (20.5%).

Margins per Transactions:

- As the number of transactions increases, Pink Cabs' margins rise due to the additional price charged for their services, while Yellow Cab's margins decrease with the same increase in transactions.





EDA SUMMARY

Pink Cab

- According to the provided information, the range of ride distances is approximately between 2 and 48 kilometers, with price charges varying between $150 and $450. During the holiday season of December, the number of travels recorded was around 11,000. Notably, the price charge for all cities exhibited a similar increase in correlation with the distance traveled.

- Transaction per year:
  2016: 20000 – 40000
  2017: 20000 – 40000
  2018: 20000 – 40000

Yellow Cab

- According to the text, the distance of rides can vary from approximately 2 to 48 kilometers, with prices ranging from $250 to $600. During the holiday season in December, the number of trips made was around 35,000. Additionally, it is noted that the price charged for yellow cabs in New York City is higher compared to other cities.

- Transaction per year:
  2016: 80000 – 100000
  2017: 80000 – 100000
  2018: 80000 – 100000

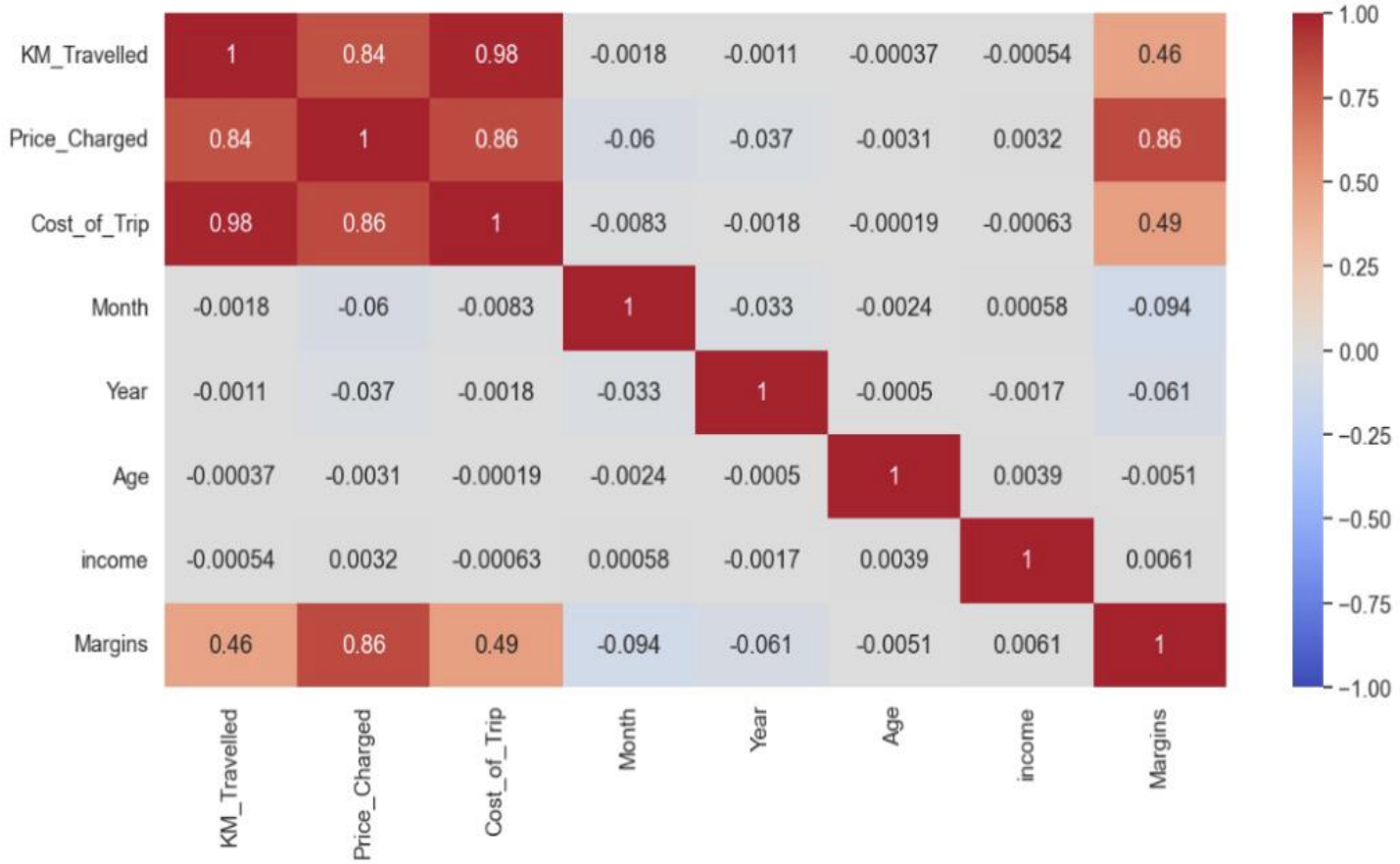
Pink Cab

- Pink Cab's pricing policy is consistent for both male and female customers, with no gender-based discrimination in pricing. According to the data, female customers make up around 20.5% of the total customer base. Unfortunately, Pink Cab's profit margins have been relatively low in recent years (2016-2018) compared to Yellow Cab. However, as the number of transactions increases, Pink Cab's profit margins tend to improve.

Yellow Cab

- According to the text, Yellow Cab appears to have a pricing strategy that charges female customers at a lower rate compared to male customers. Female customers make up approximately 25.5% of the total customer base. Additionally, Yellow Cab has reported a higher profit margin in each of the years 2016-2018 compared to its competitor, Pink Cab. However, as the number of transactions increases for Yellow Cab, its profit margin decreases.

Correlation:



- According to the graph, there exists a positive relationship between margin and the price charged.

Hypothesis Testing

Hypothesis :The margin for both taxi services remains unchanged between genders.

- Pink Cab: There is no difference in Margin between Male and Female customers.

```
print('P value is ', p_value)
```

```
37480 47231
We accept null hypothesis that there
is no statistical difference
P value is  0.11515305900425 8
```

- Yellow Cab: There is difference in Margin between Male and Female customers.

```
print('P value is ', p_value)
```

```
116000 158681
We accept alternate hypothesis that
there is a statistical difference
P value is  6.060473042494056e-25
```

Hypothesis : No matter what age group you belong to, the margins remain the same in both cabs.

- Pink Cab: There is no difference in Margin for all Age group.

```
print('P value is ', p_value)
```

```
71228 14405
We accept null hypothesis that there
is no difference
P value is  0.3905344473325604
```

- Yellow Cab: There is difference in Margin for people older than 50 years.

```
print('P value is ', p_value)
```

```
231480 46368
We accept alternate hypothesis that
there is a difference
P value is  3.893841736668538e-05
```

Hypothesis :The payment methods used by customers, whether cash or card, result in differing margins for businesses.

- There is no difference in Margin regarding mode of Payment for both Yellow & Pink Cab.

Pink Cab:

```
print('P value is ', p_value)
```

```
33992 50719
We accept null hypothesis that there
is no difference
P value is  0.7900465828793286
```

Yellow Cab:

```
print('P value is ', p_value)
```

```
109896 164785
We accept null hypothesis that there
is no difference
P value is  0.29330606382987284
```

Recommendation

- From 2016 to 2018, the average number of transactions per year for Yellow Cab is nearly twice that of Pink Cab. Additionally, Yellow Cab exhibits a noticeable disparity in margin between male and female customers, resulting in a higher percentage of female customers. Furthermore, Yellow Cab's profit margin is higher than Pink Cab's over the three-year period. Interestingly, Yellow Cab's margin varies by age group, with people over the age of 50 experiencing a decrease in margin. Notably, Yellow Cab's travel frequency during the holiday season in December is significantly higher than Pink Cab's. Finally, Yellow Cab serves the most customers in New York City, which accounts for 28% of all cab users.

Based on the previously discussed factors, it is advisable to invest in Yellow Cab.

Thank You