G2M insight for Cab Investment firm

Company : Name : XYZ
Location : Lebanon
Team : Data and Analytics
Date : 05-08-2021

Agenda

- Executive Summary
- Problem Statement
- Approach
- EDA
- EDA Summary
- Hypothesis Testing
- Building Models
- Recommendations

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.

- 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,
  - Building models and finding the best fit model based on Accuracy.

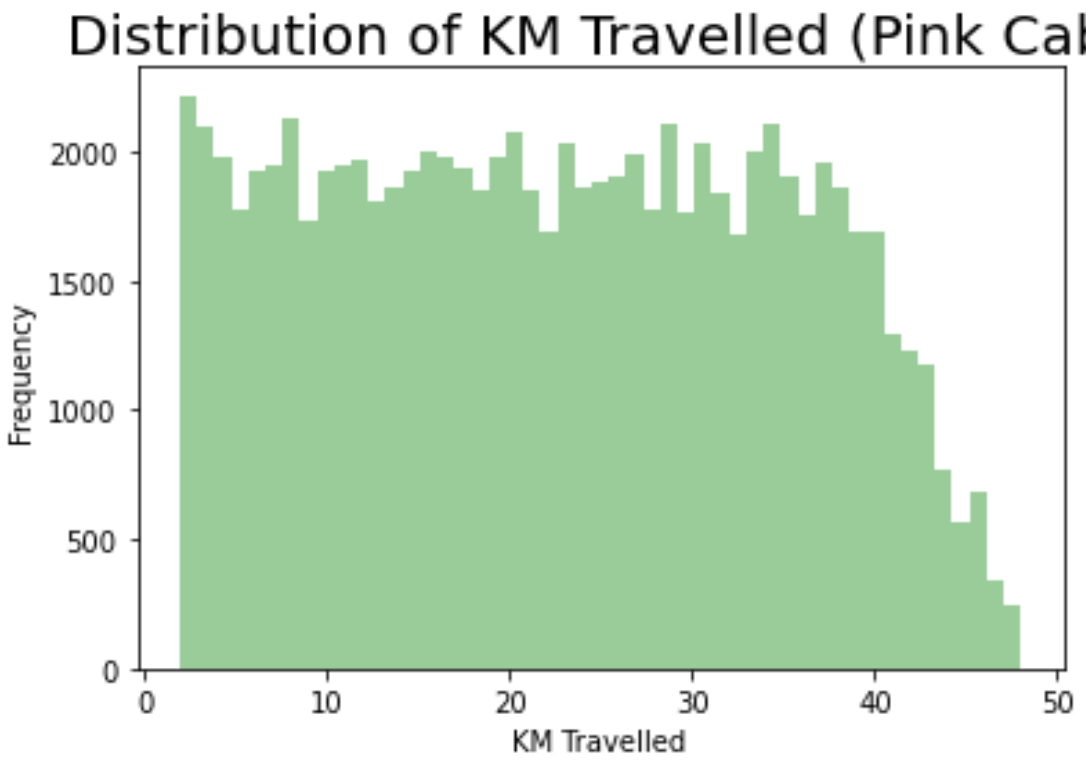
Data Preparation:

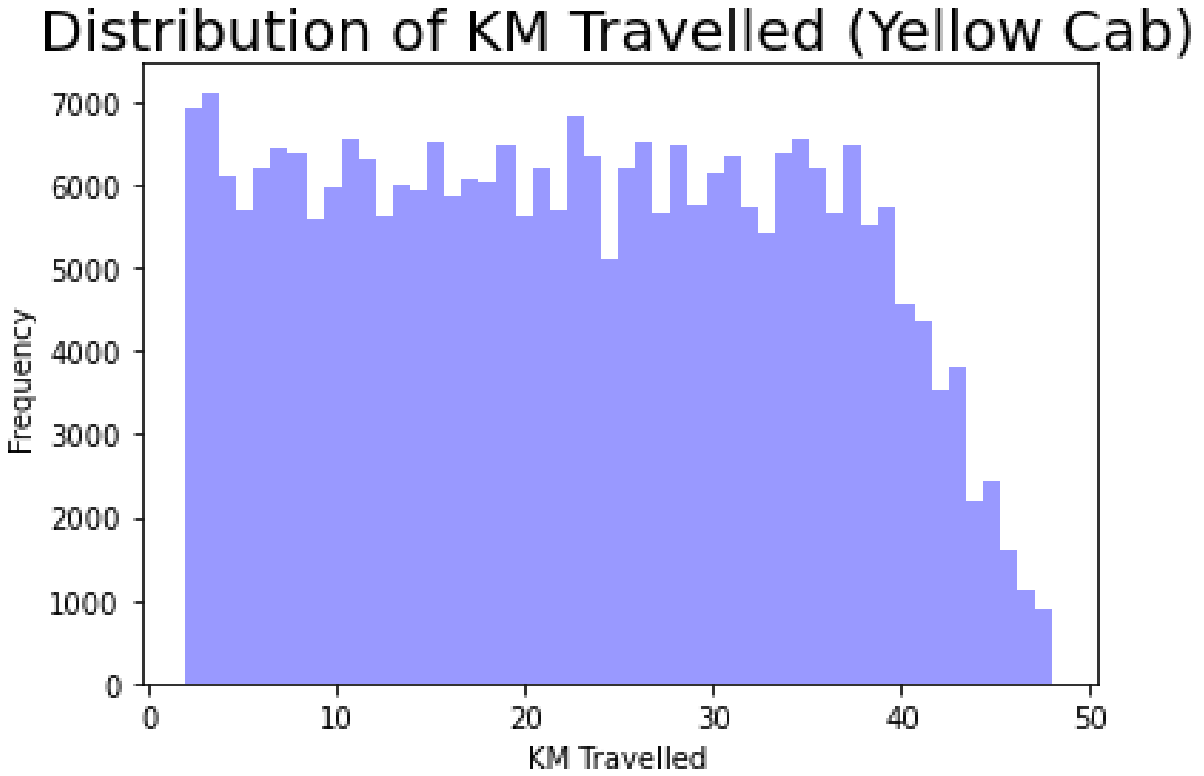
There are 4 datasets:

- **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

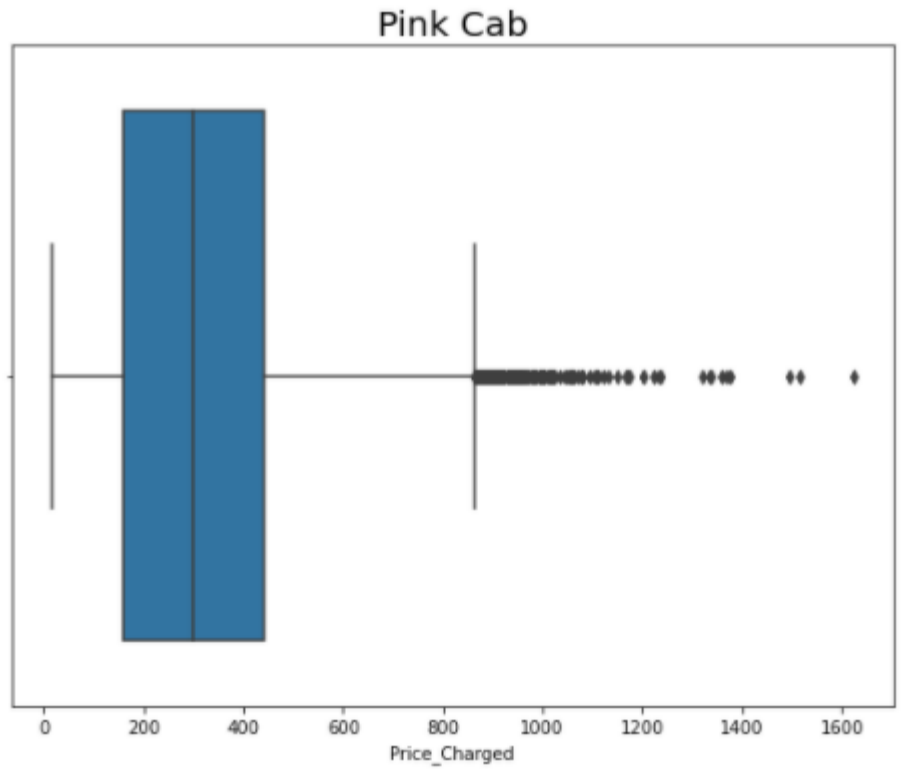
Distribution of KM Travelled for both Cabs:

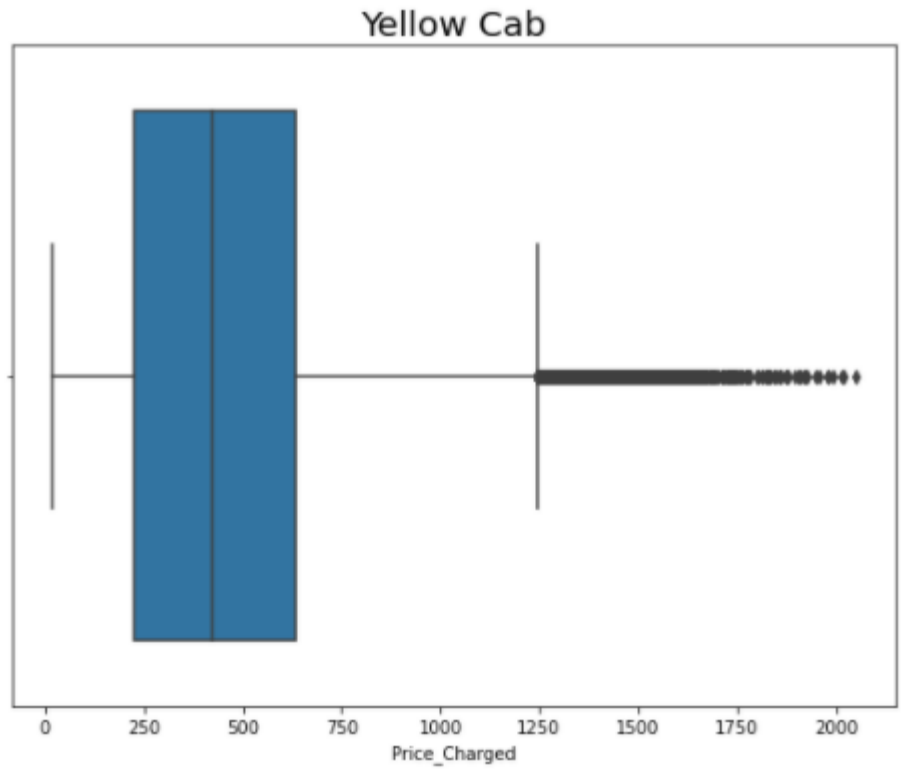




- From the above graphs, we can see that for both Pink and Yellow Cab most of the rides are in the range of approximately 2 to 48 KM.

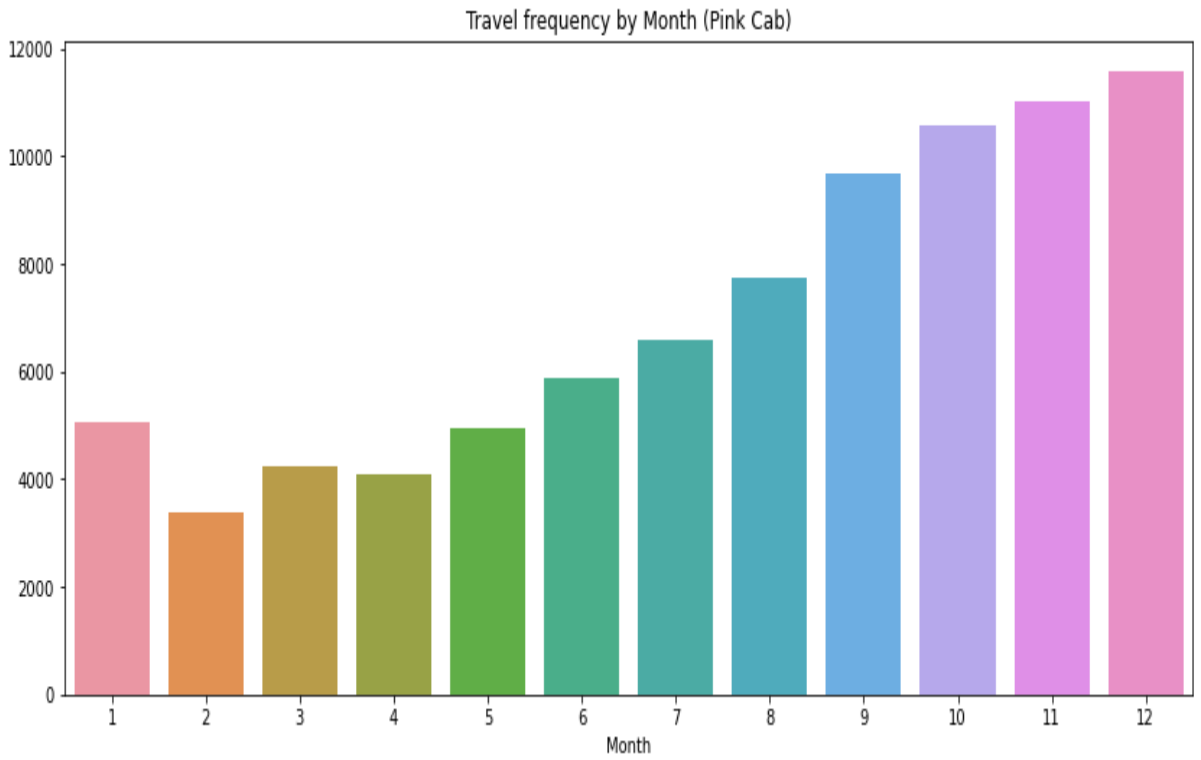
Distribution of Price Charged for both Cabs:

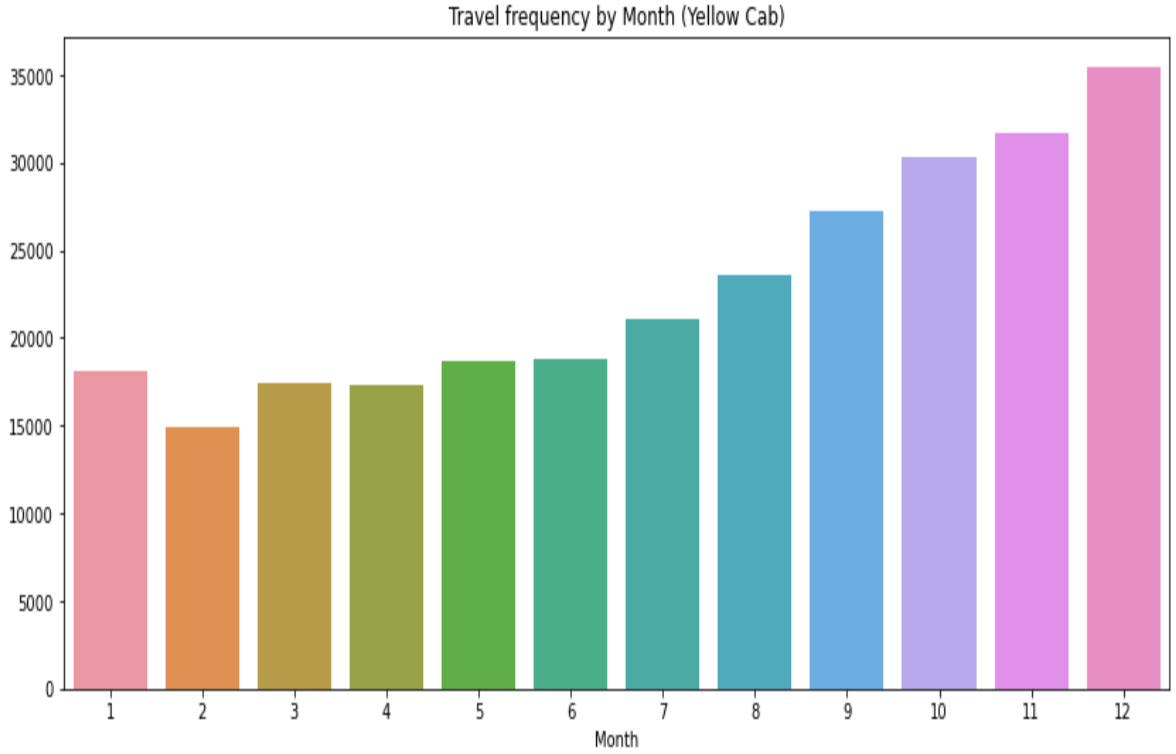




- The Price Charge range for Yellow cab is more than the Pink cab.
- The Price Charged is an outlier but it is important for the study so it should be kept.

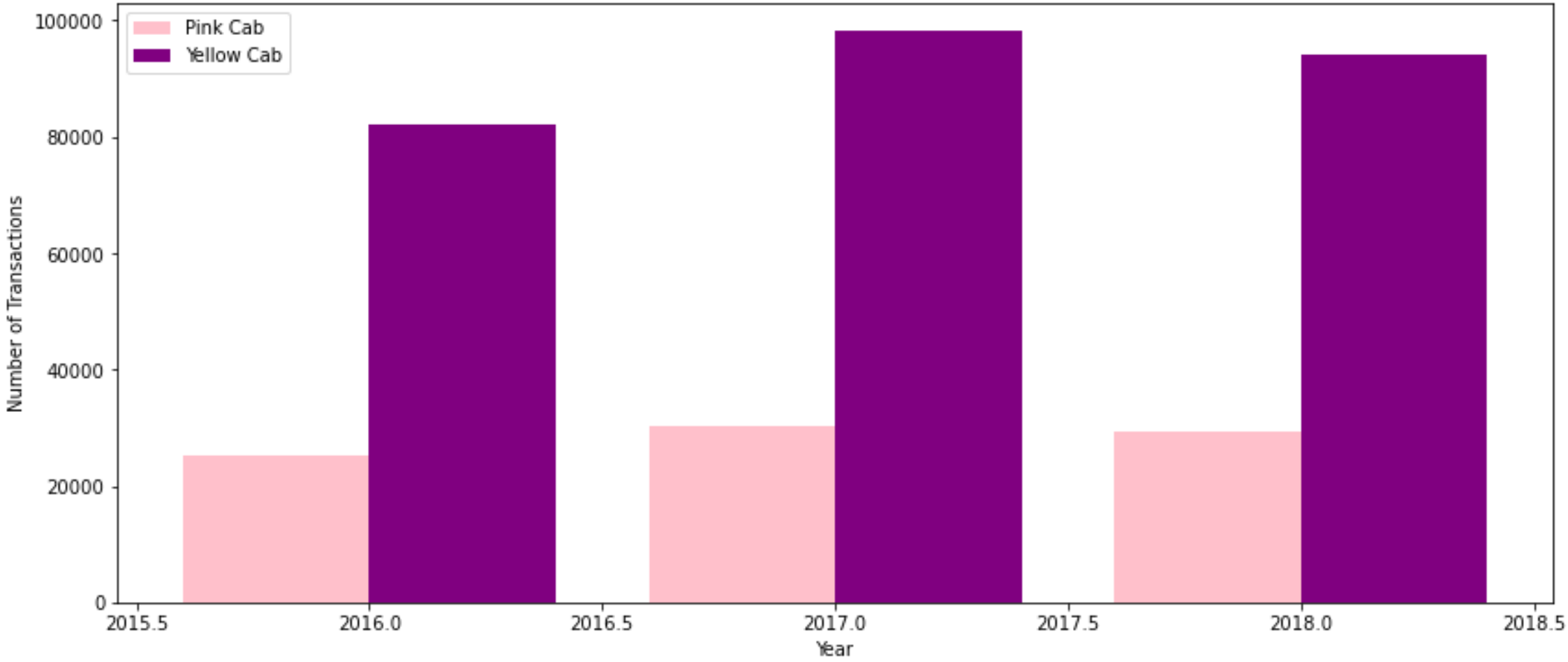
Travel Frequency per Month:





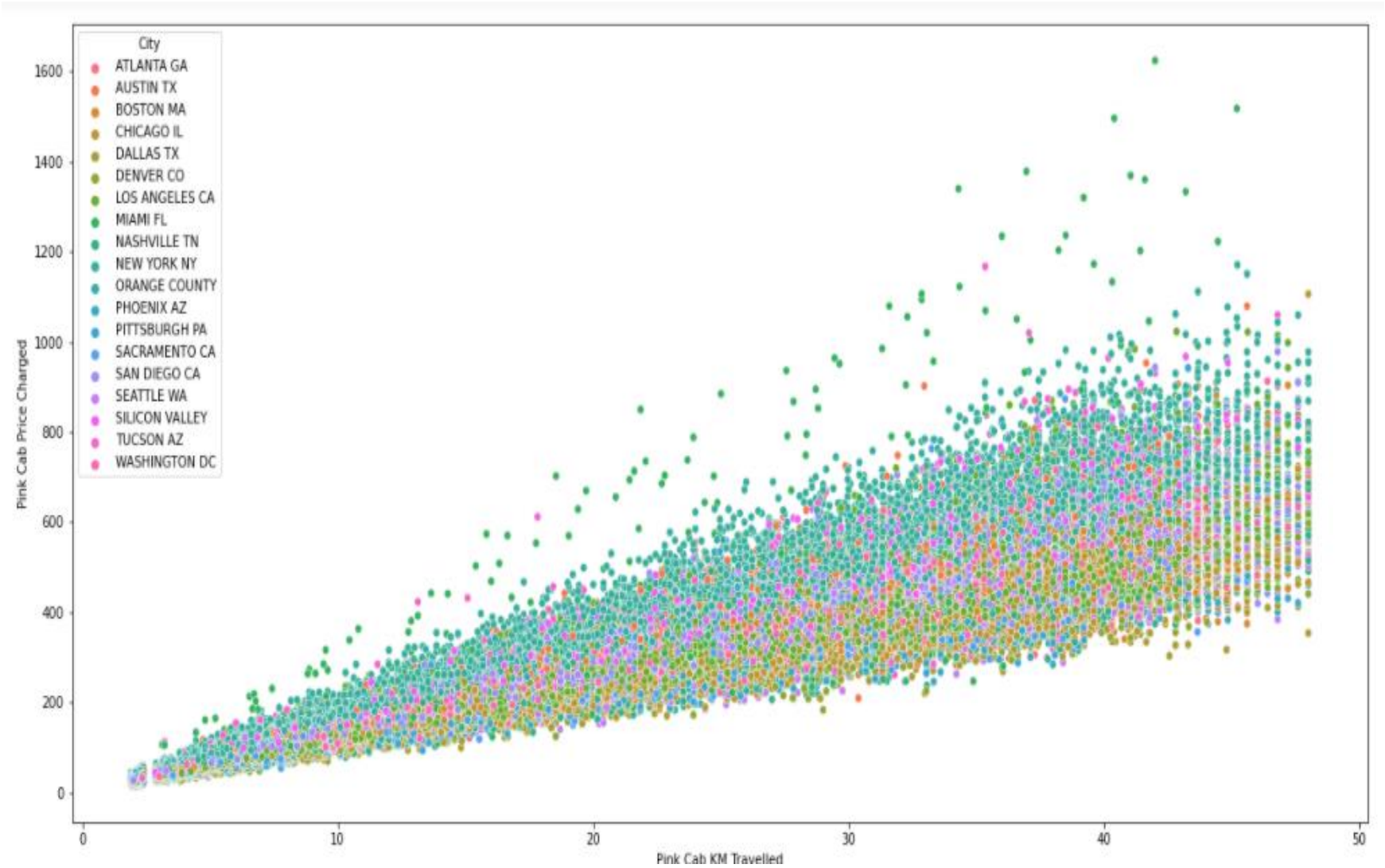
- Yellow Cab has higher travels (30135) in the month of December which is the holiday season compared to Pink Cab (9729).

Transaction per Year for both Cabs:



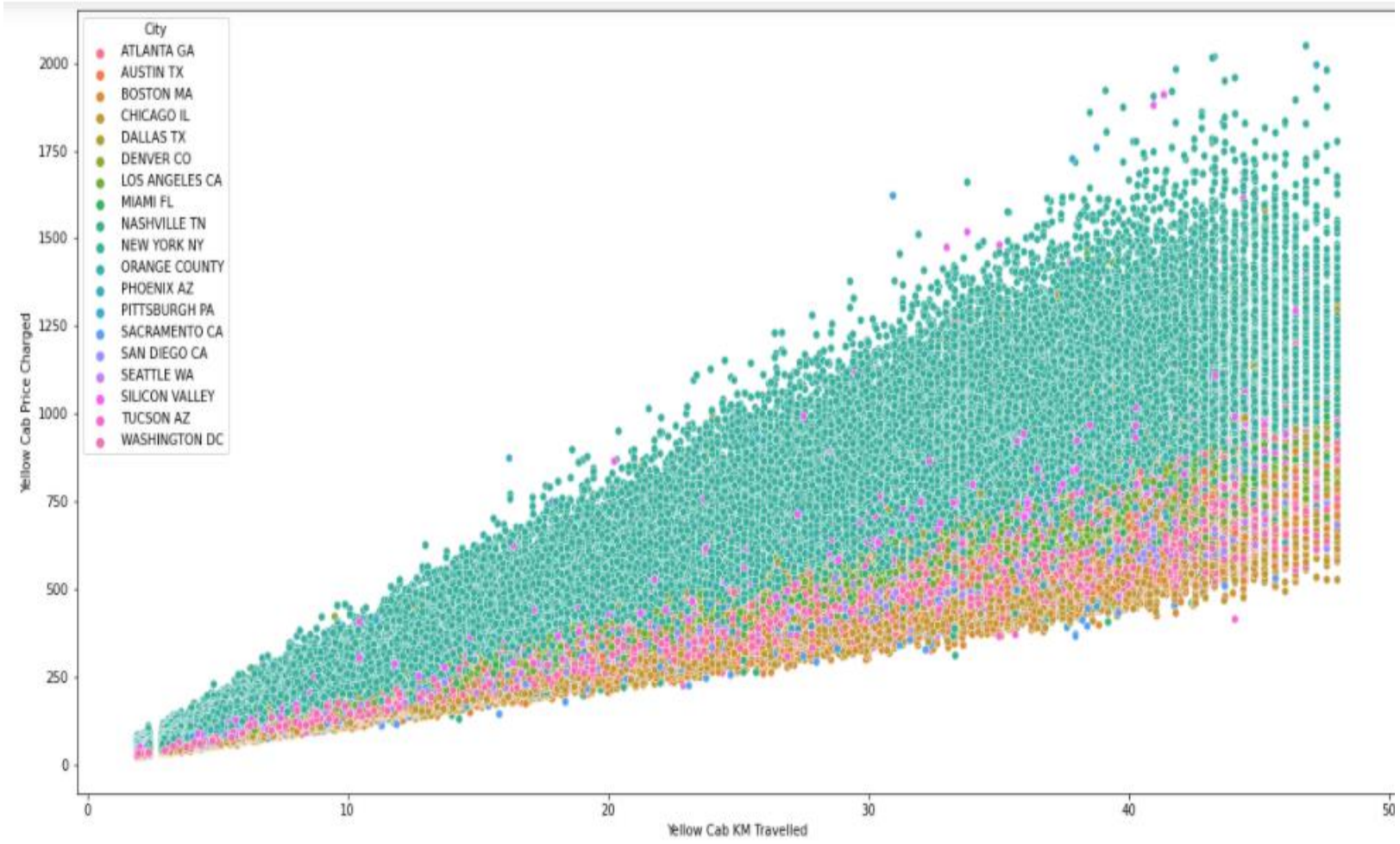
- From the graph it shows that on yearly basis no. of transactions for Yellow cab is higher than Pink cab.

Pink Cab: Price Charged per KM per City



- In Pink cab we can say that all cities have approximately the same price of a ride with the increase of distance travelled.

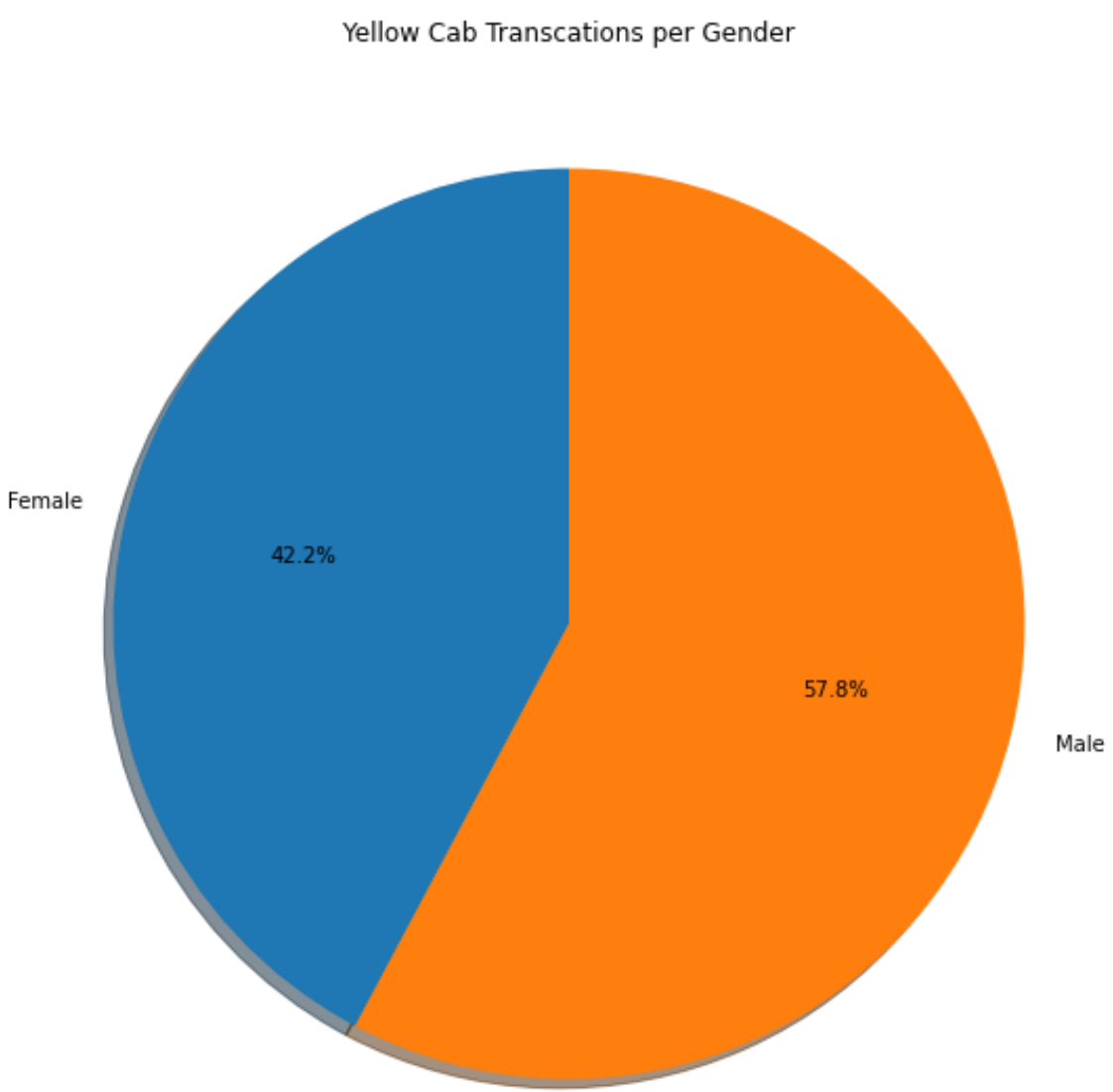
Yellow Cab: Price Charged per KM per City

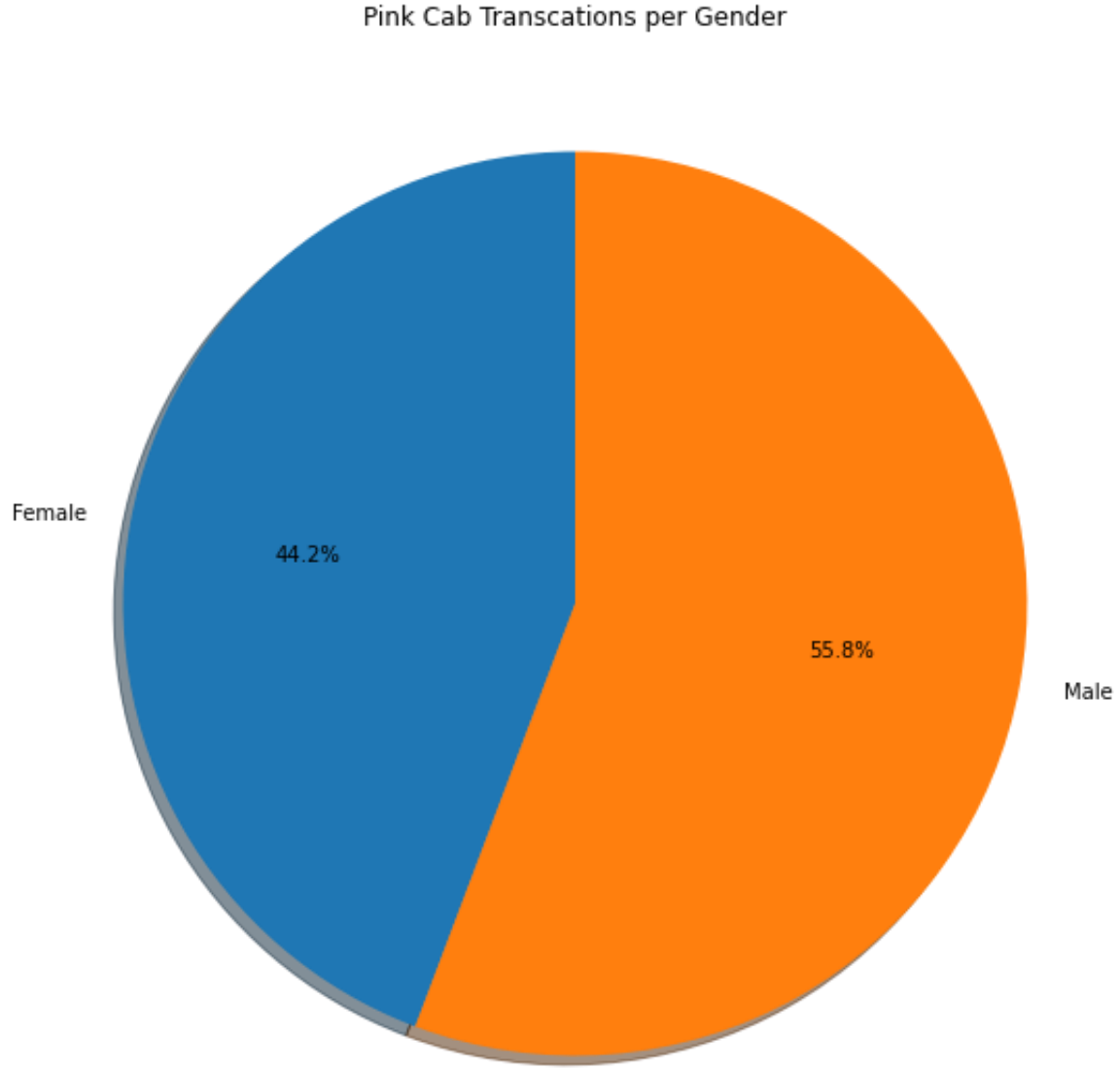


- From these graphs we can see that for Yellow cab the New York city have the highest price of a ride compared to other cities iththe increase in distance travelled

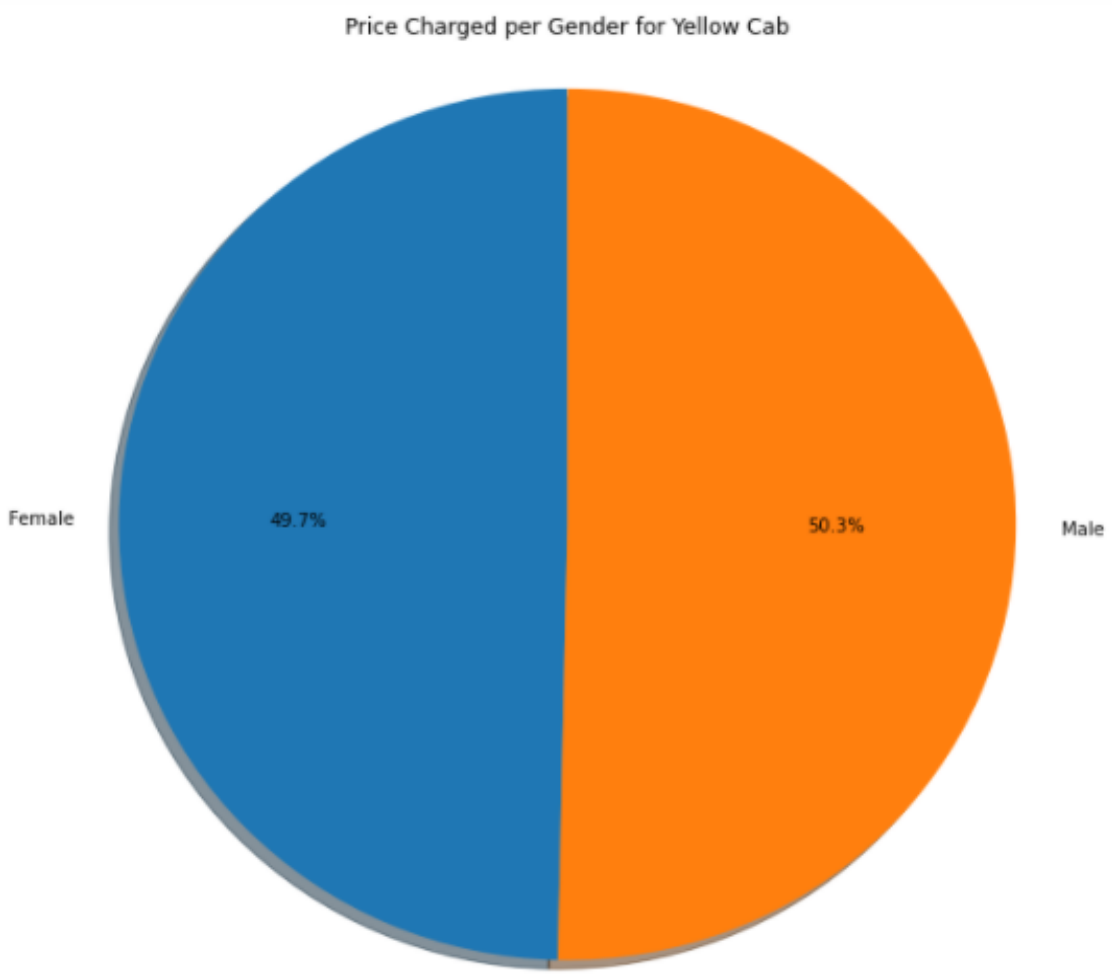
Transaction Per Gender :

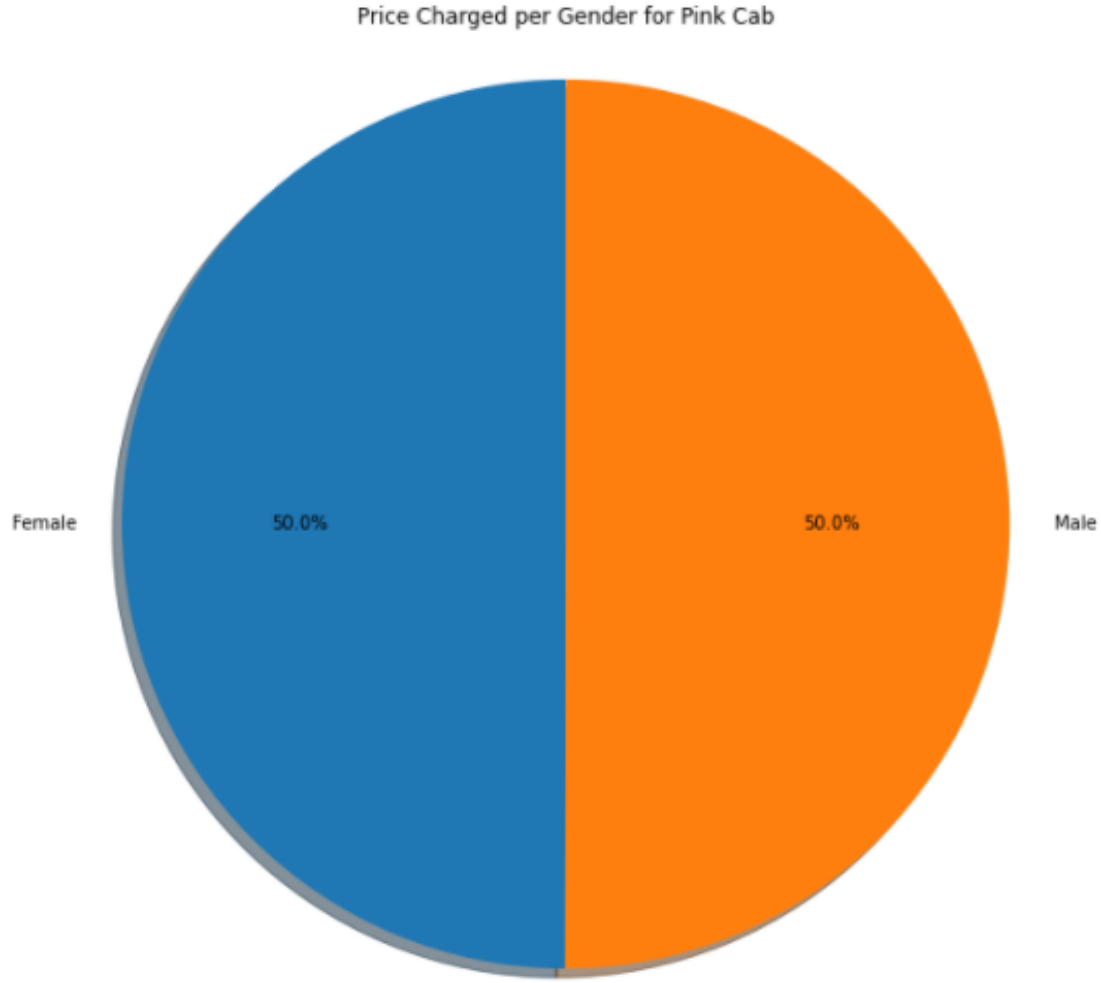
- **Both Cabs have the same number of Females and Males.**





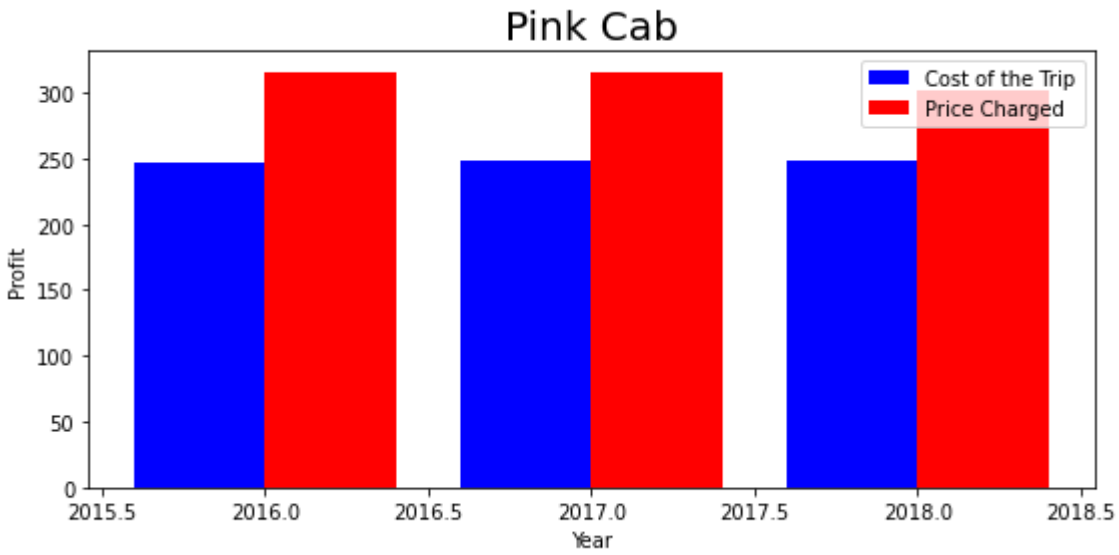
Price Charged per Gender for both Cabs:

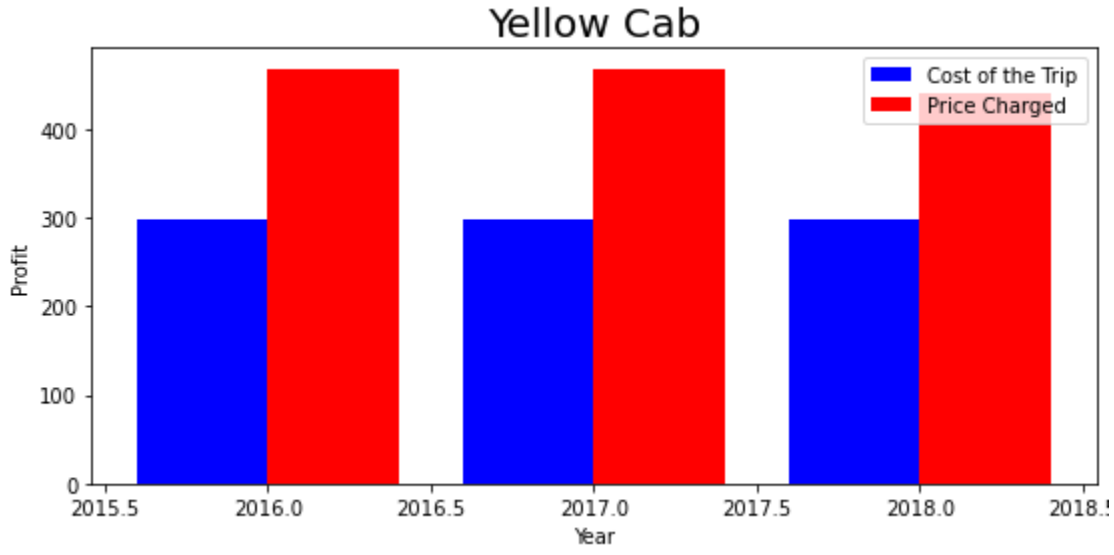




- We can notice that the price of a ride is equal for males and females in Pink cab users while it is less for female users in Yellow cab. Which mean that females might have a discount in Yellow cabs

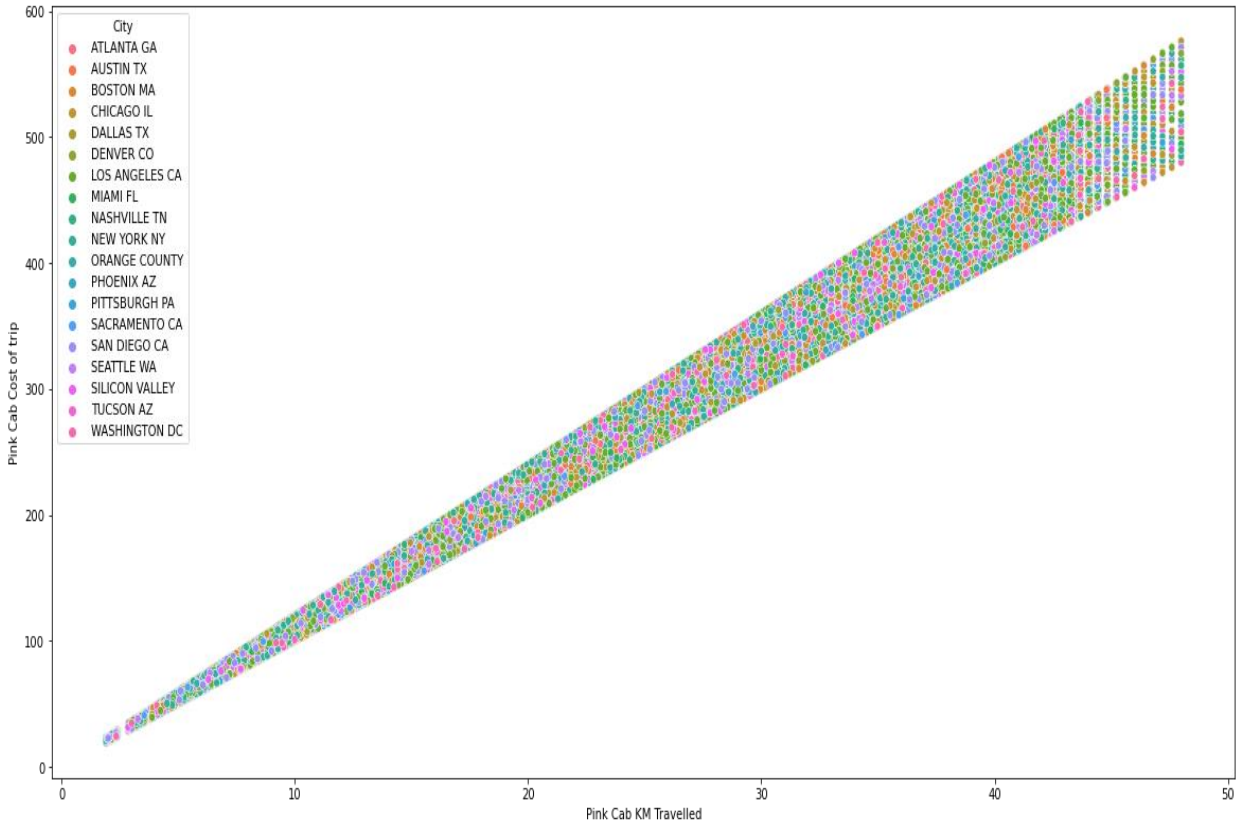
Profit Margin per year for both Cabs:

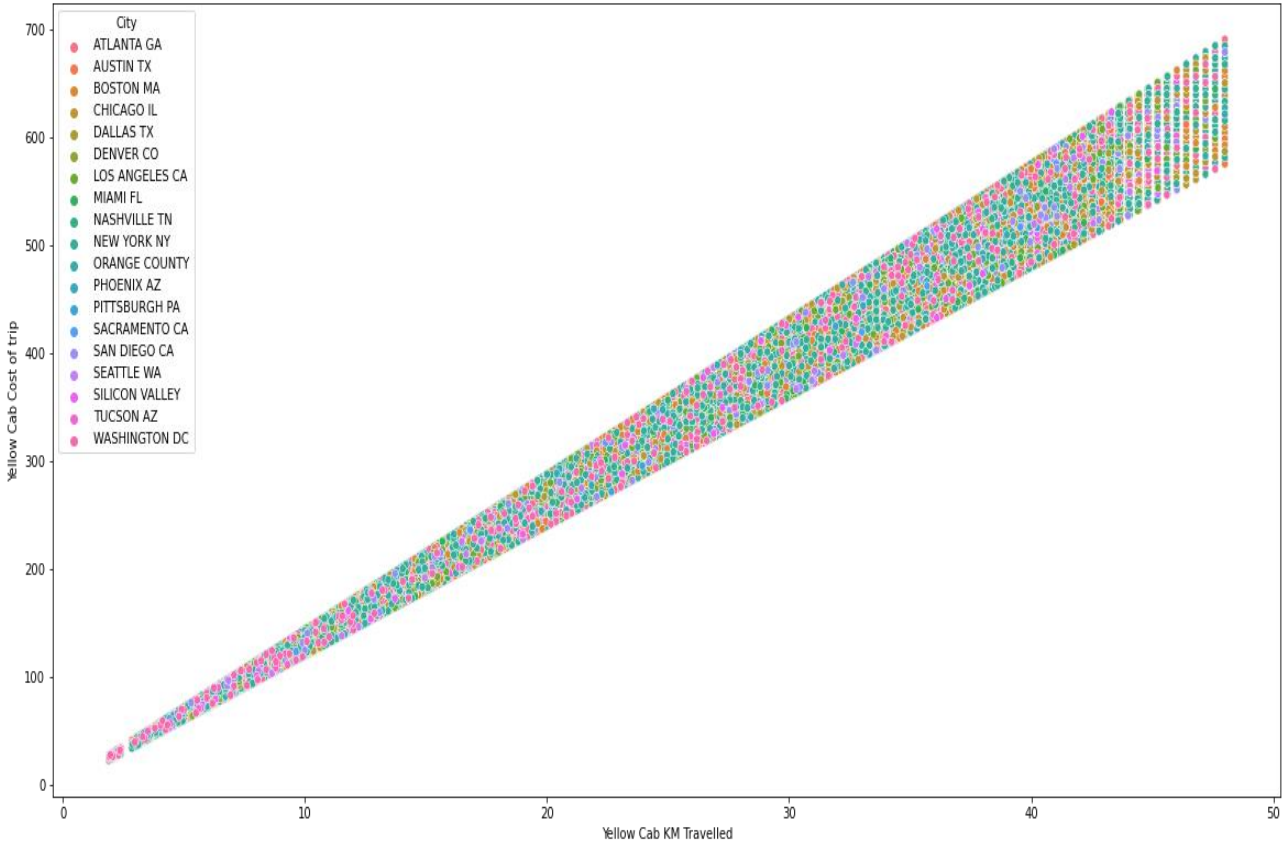




- From the graphs above we have that the Profit (Price Charged - Cost of Trip) of a Yellow Cab yearly is higher compared to Pink Cab
- Yellow cab is more expensive than Pink cab
- For each graph of Pink and Yellow cabs we notice that always the Price Charged is greater than the Cost Of Trip which is normal.

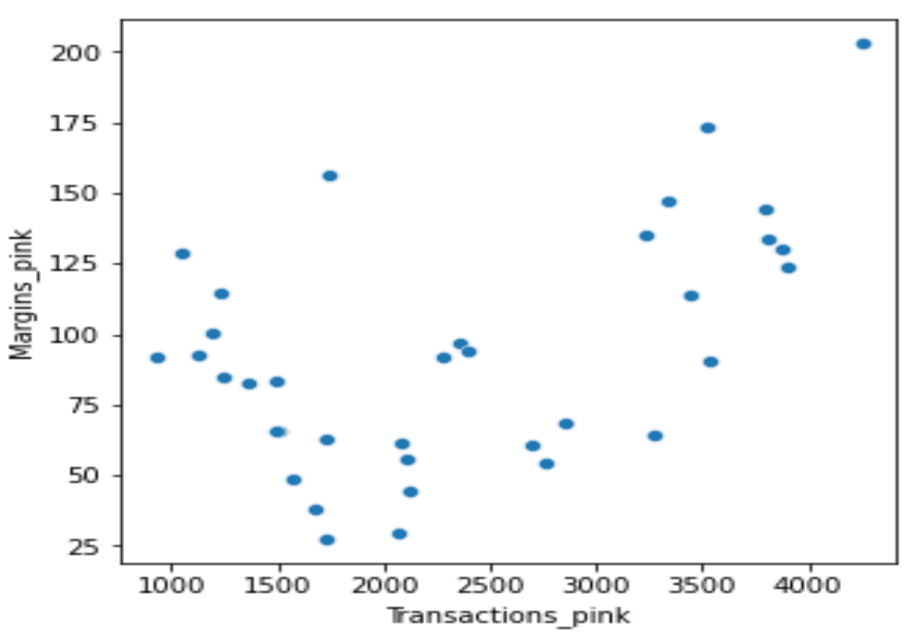
Cost Of Trip For Both Cabs :

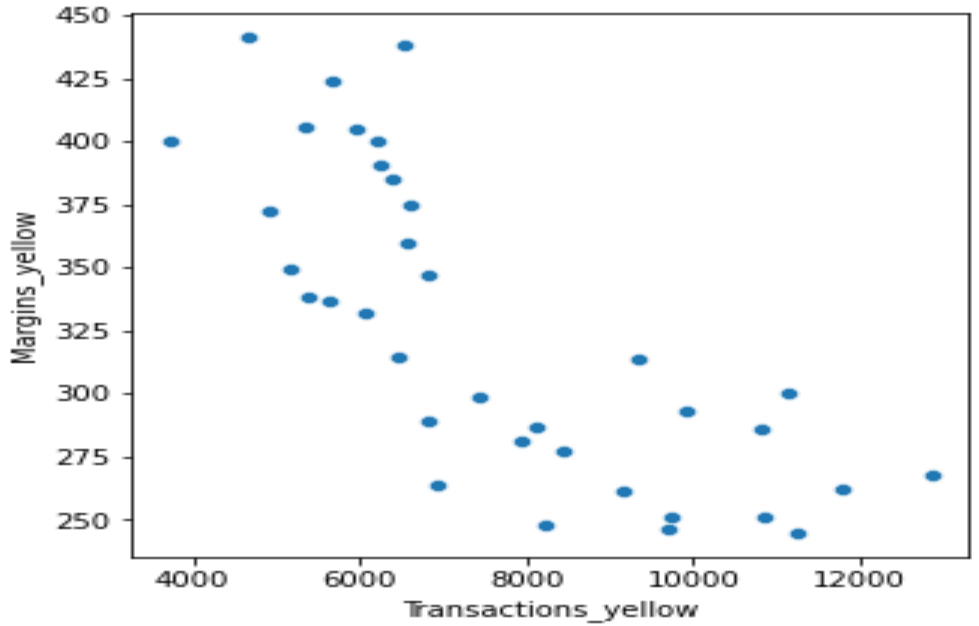




- The Cost of Trip in both Cabs Pink and Yellow is the same .

Margins per Transactions:

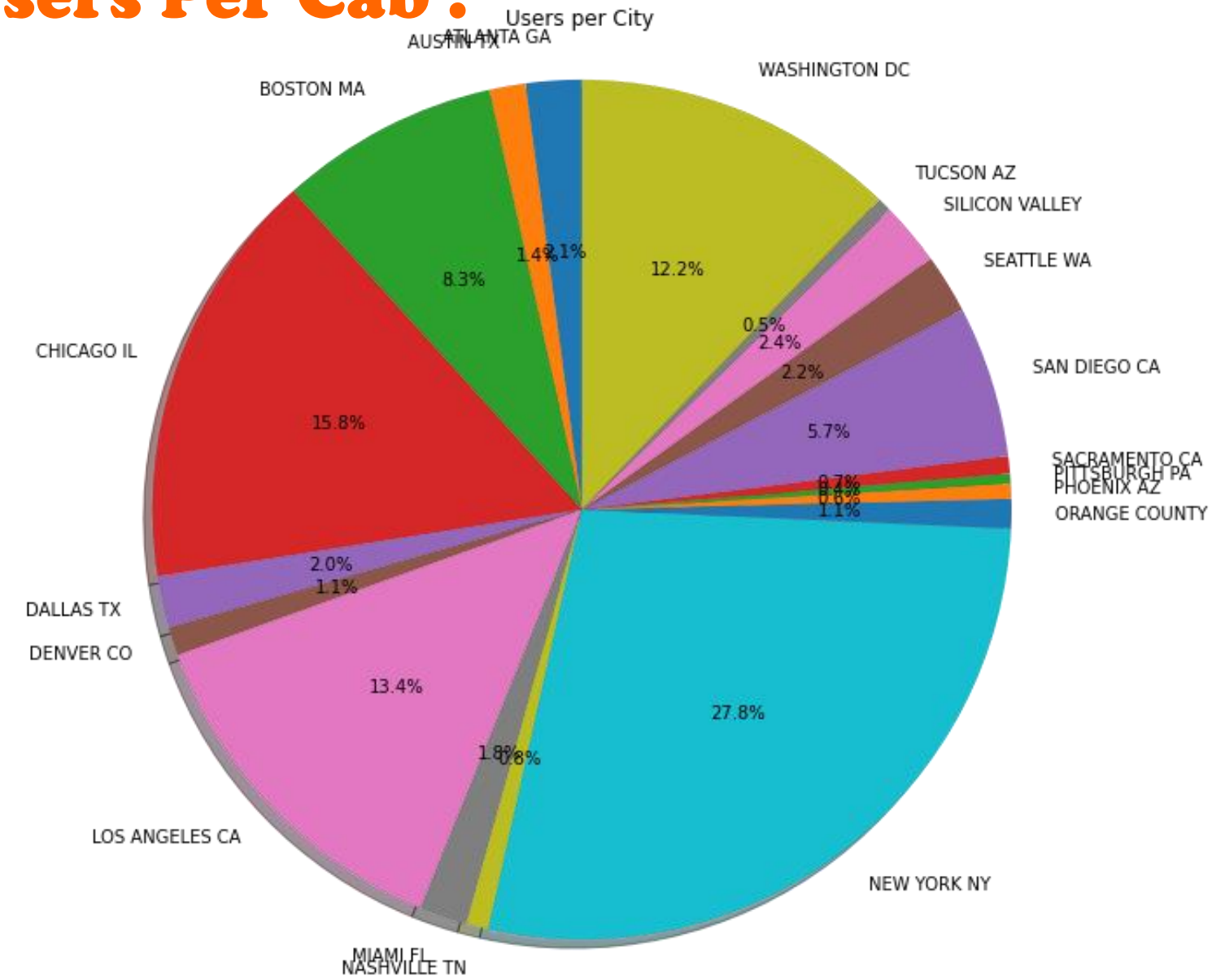




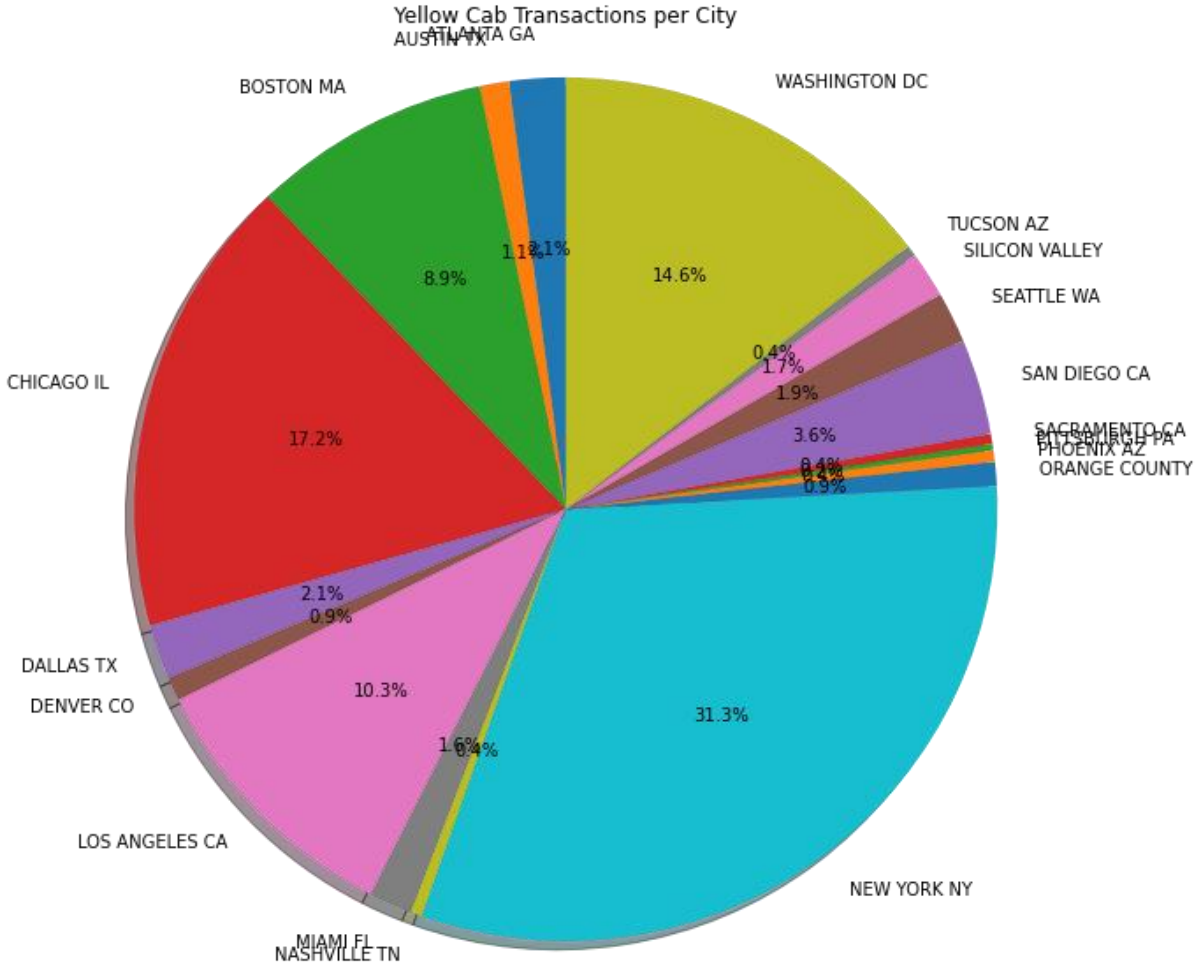
- Margins: Price Charged – Cost of Trip
- Pink Cabs increase margins with increase in number of Transactions.
- Yellow Cab decrease Margins with the increase in Transaction.

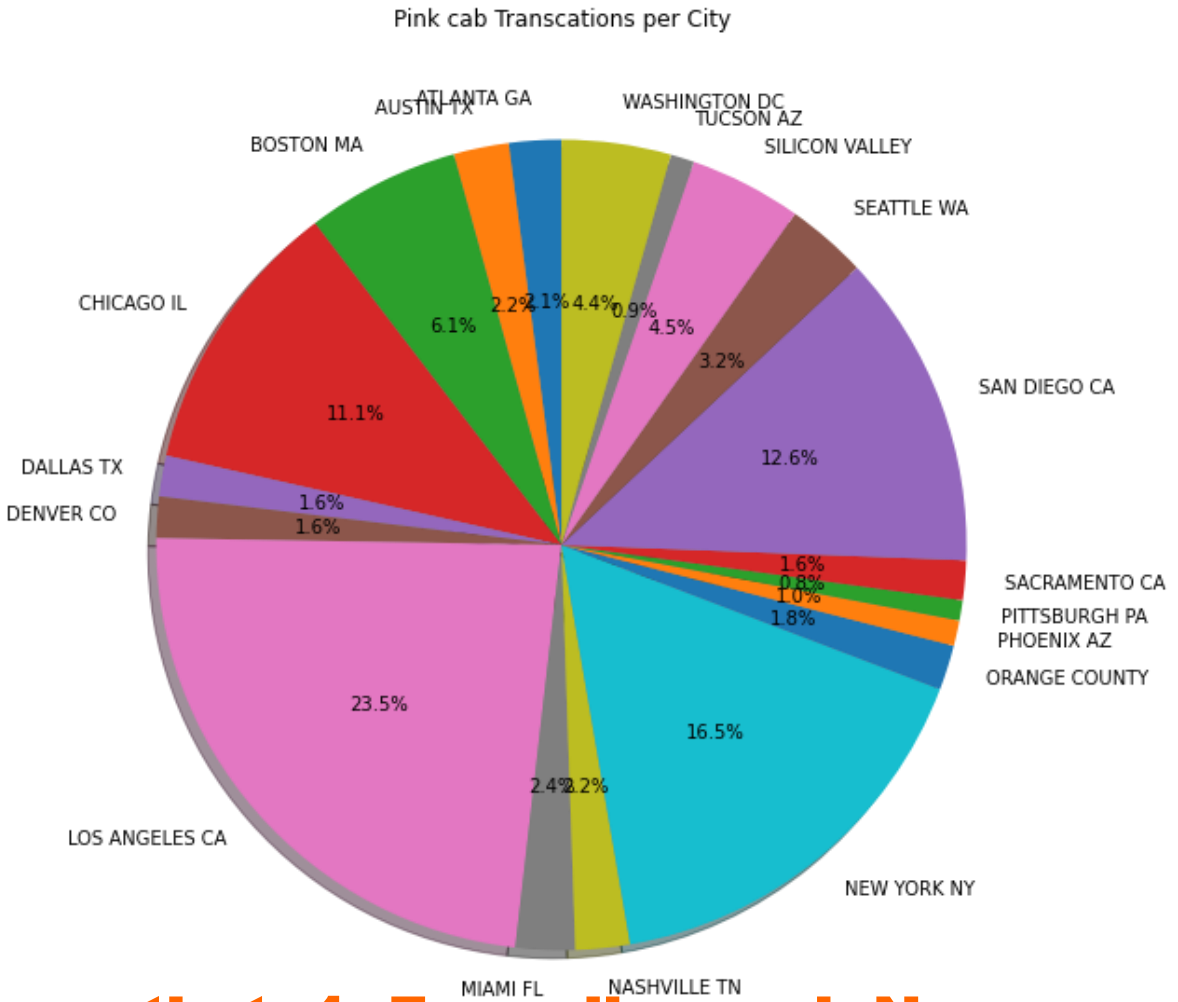
Users Per Cab :

- We can notice that New York city have the most cab users about (28%) and the city that follows new York is the Chicago IL city with approximately 16% then Los Angeles CA with approximately 13.4% .



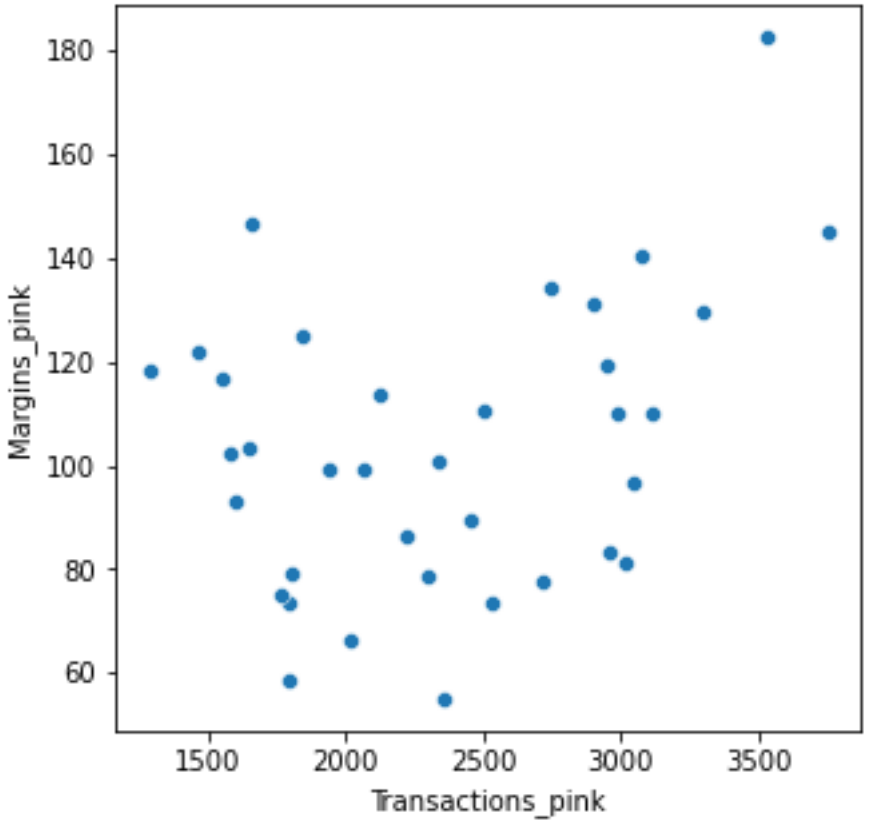
Transaction Per City For Each Cab :

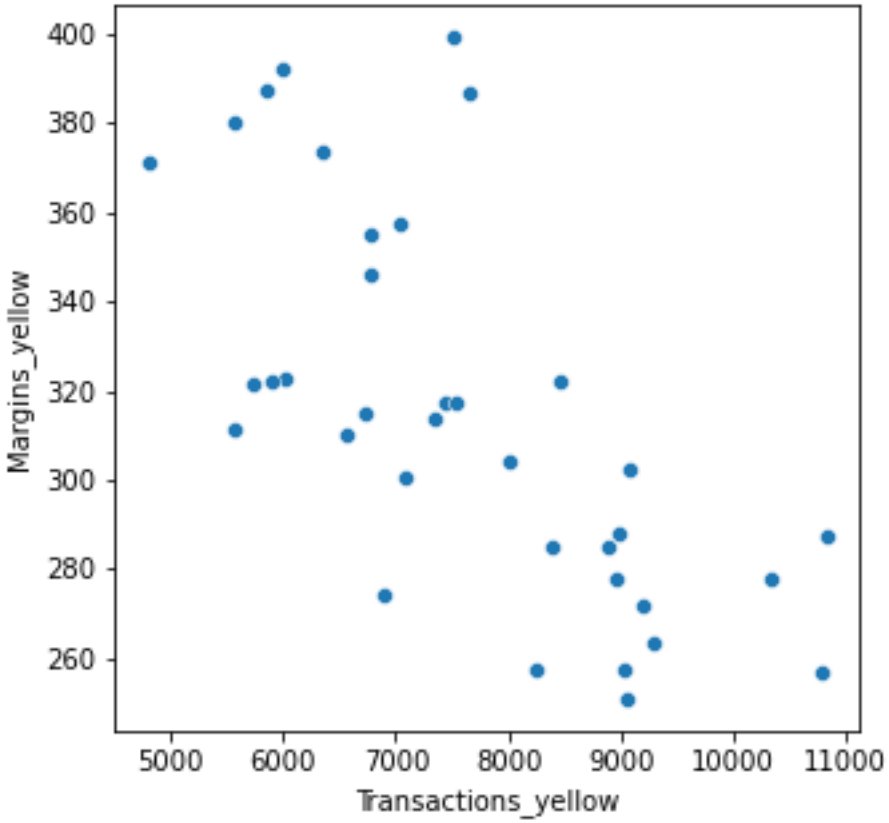




- From the above pie charts we can see that: 1. For yellow cab New York city has the highest number of transaction which contains the highest users about 28%. 2. For pink cab Los Angeles city has the highest number of transaction

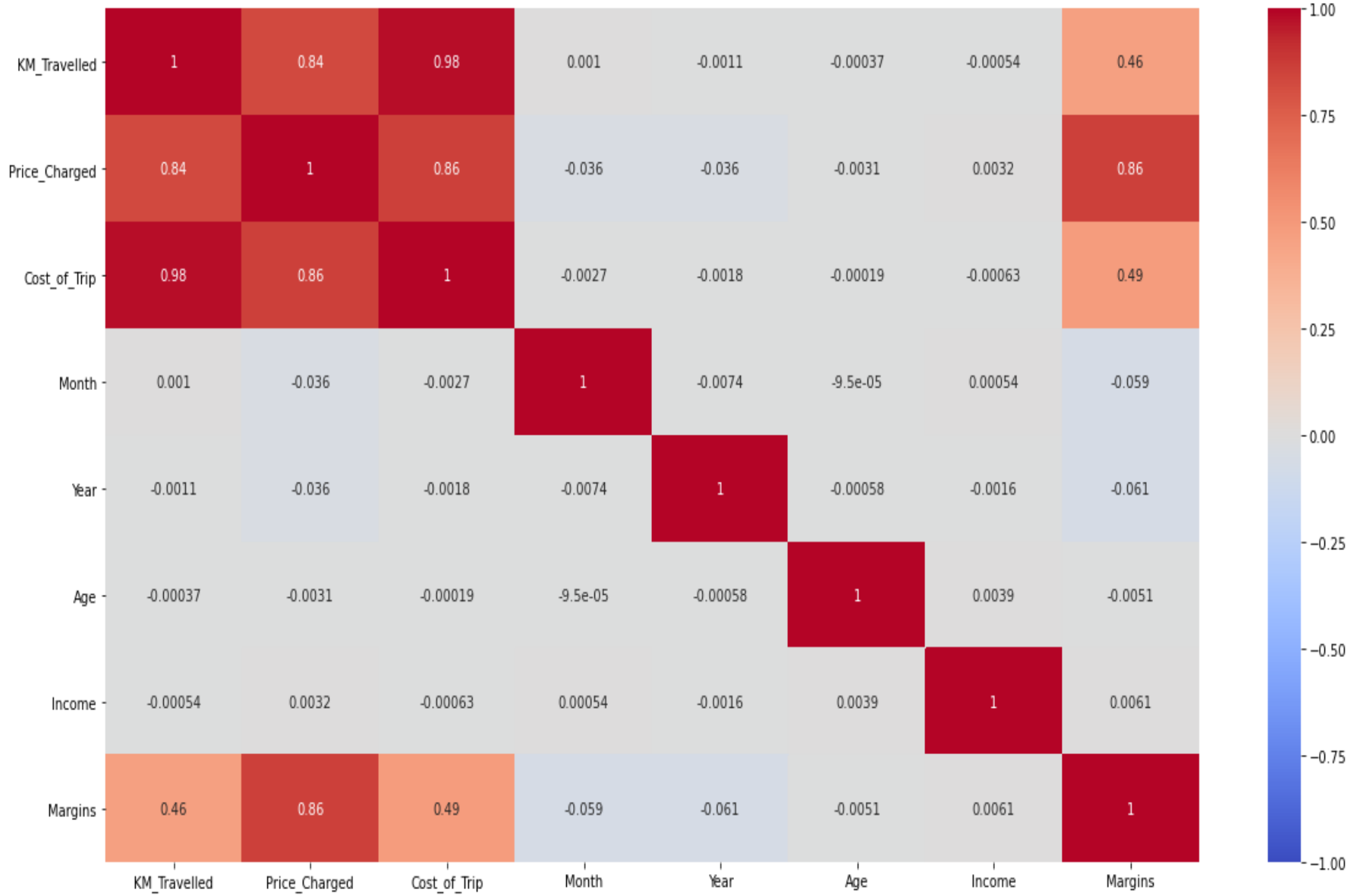
Margins Per Transaction :





- Yellow cabs decrease their margins with the increase in the number of transactions.
- Pink Cabs increase their margin with the increase in number of transactions

Correlation :



- There is a correlation between the Margins and Price Charged

EDA

SUMMARY

- Rides are in the range between approximately 2 to 48 KM.
- All the cities have the same price charged as the distance travelled increases .
- The number of travels in the month of December which is a holiday month is about 9729 .
- The Yearly Transactions is:

  2016 → 25080

  2017 → 30321

  2018 → 29310

- Rides are in the range between approximately 2 to 48 KM.
- The New York city has the highest price charged compared to other cities as the distance travelled increases .
- The number of travels in the month of December which is a holiday month is about 30135.
- The Yearly Transaction is :

  2016 → 82239

  2017 → 98189

  2018 → 29310

- Pink Cab have same charge for both females and males.
- Profit Margin is low each year (2016-2018) compared to Yellow Cab.
- Pink Cabs increase margins with increase in number of Transactions.

- Yellow Cab charge less from Female Customers.
- Profit Margin is high each year (2016-2018) compared to Pink Cab.
- Yellow Cab decrease Margins with the increase in Transaction.

HYPOTHESIS

TESTING

Null Hypothesis : Margin remain the same regarding Gender for both Yellow Cab & Pink Cab.

```
116000 158681
We accept alternate hypothesis that there is a statistical
difference
The p-value = 6.060473042494144e-25
```

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

```
37480 47231
We accept null hypothesis that there is no difference
P value is 0.11515305900425798
```

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

Null Hypothesis : Margin remain the same for all Age groups for both Male and Female

```
71228 13483
We accept null hypothesis that there is no difference
P value is 0.32817487547980695
```

Pink Cab: There is no difference between Margin per Ages

```
231480 43201
We accept alternate hypothesis that there is a difference
P value is 6.494256817799368e-09
```

Yellow Cab: There is discount for older than 50

Null Hypothesis : the Margin remain the same for Card payer and Cash payer in both Yellow and Pink Cabs.

```
Pink Cab:
We accept null hypothesis that there is no difference
P value is 0.7900465828793288
```

```
Yellow Cab:
We accept null hypothesis that there is no statistical difference
P value is 0.293306063829872 9
```

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

BUILDING A MODEL

LINEAR REGRESSION MODEL :

- Linear Regression is a method for predicting target value and attempts to model the linear relationship between target and one or more predictors.
- In our dataset, Price Charge is the target value and all the other variables are predictors.
- Splitting the data into a training set (75%), and test set (25%).
- Splitting the dataset into X_train , Y_train, X_test, Y_test .

RECOMMENDATION:

- **Transaction per year :** For Yellow Cab the number of transaction yearly from 2016 to 2018 is higher compared to Pink Cab.
- **Margin per Gender :** For Pink Cab the margin remain the same for both genders but in Yellow Cab there is difference between margin of Female and Male due to % of Females in Yellow Cab is higher than in Pink Cab.
- **Margin per Age :** For Yellow Cab there is significant difference in Margin for older than 50 people, whereas there is no difference in Margin for all Age groups in Pink Cab.
- **Profit Margin :** For Yellow Cab the profit margin is higher yearly compared to the Pink Cab.
- Yellow Cab **decreases Margins with the increase in Transaction**, hence for Yellow Cab the travel frequency during the Month of December which is the holiday season is 3 times more than Pink Cab.
- Customers for Yellow Cab is highest in New York City which has the highest Cab Users of 28%.

Yellow Cab is recommended for investment

Thank You