Data Glacier

Your Deep Learning Partner

G2M insight for Cab Investment firm

Company Name : XYZ
Location: Denmark
Date: 27-06-22
By: Hiten Chadha

Agenda

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

Description:

XYZ is a private 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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

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

Cab Companies:
- Yellow Cab
- Pink Cab

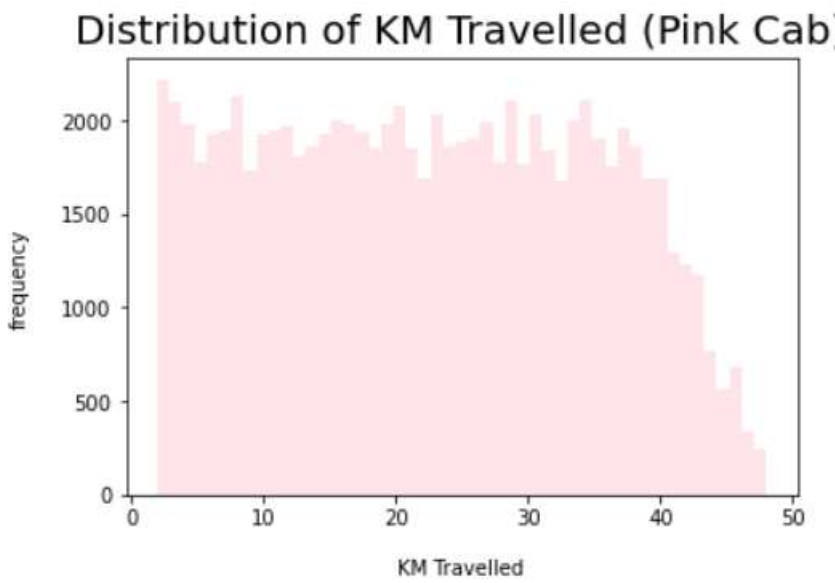
The Analysis include :
- Data Understanding,
- Data Visualization,
- Creating multiple hypothesis,
- Building models and finding the best fit model based on Accuracy

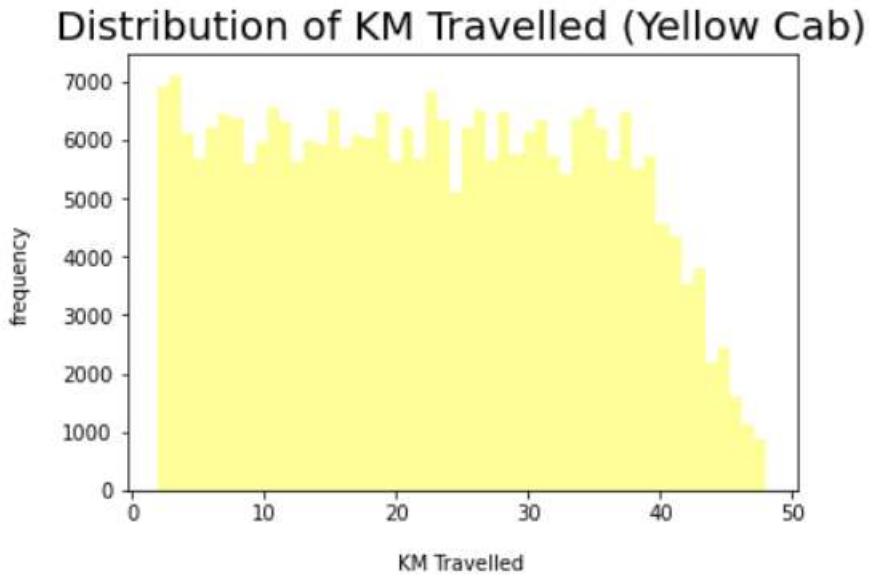
Data Preparation:

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.

EDA

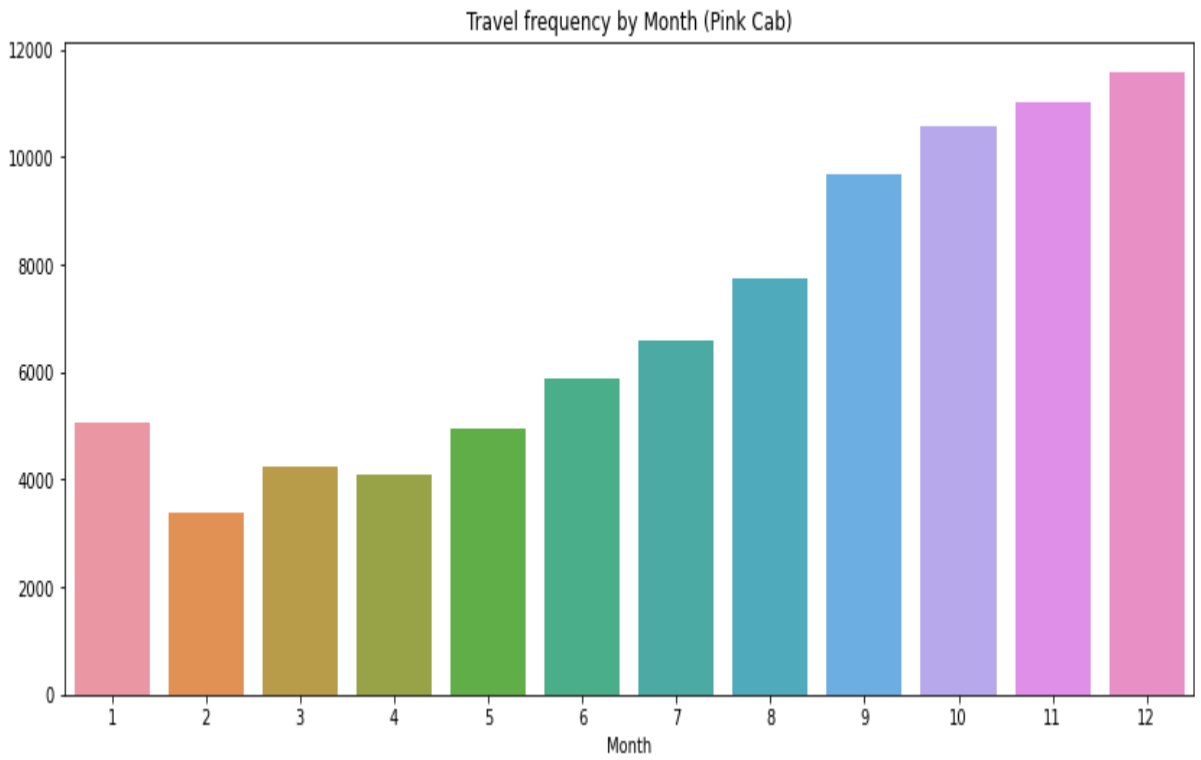
Distribution of KM Travelled for both Cabs:

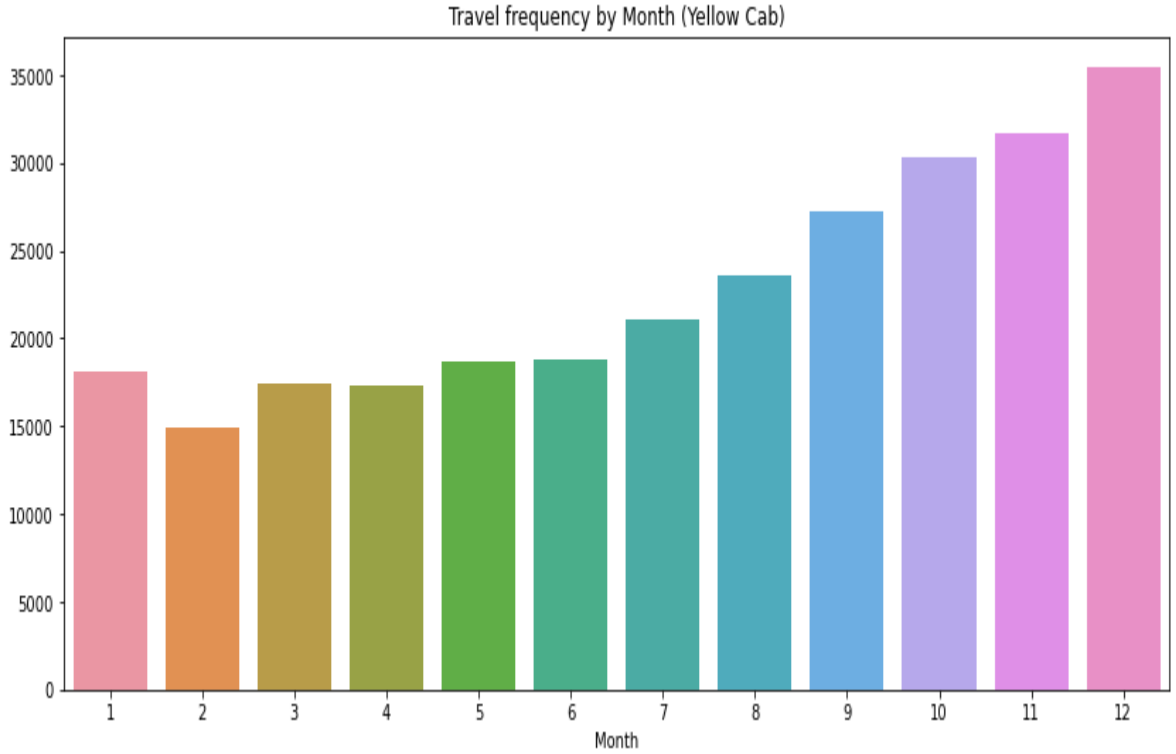




- From the above graphs, we can see that for both Pink and Yellow Cab most of the rides are in the range of approximately 2 to 48 KM.

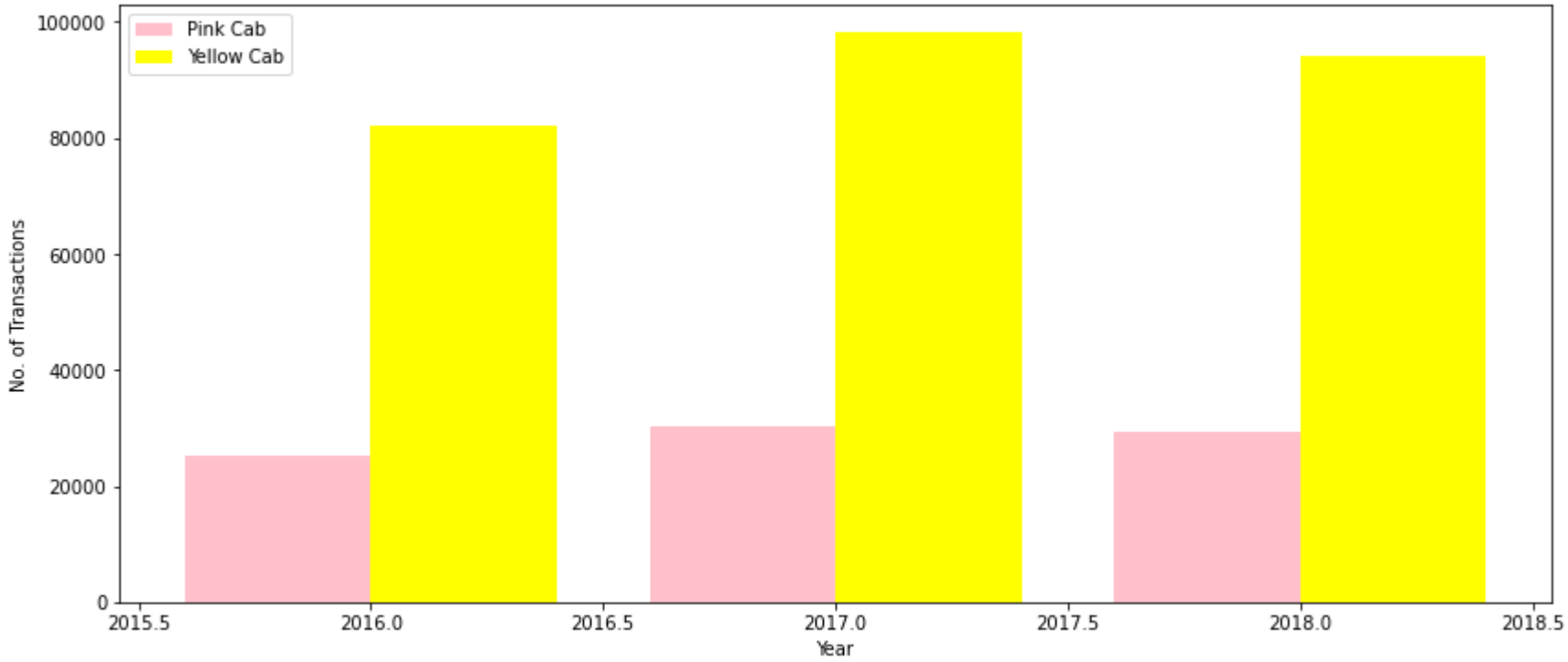
Travel Frequency per Month:





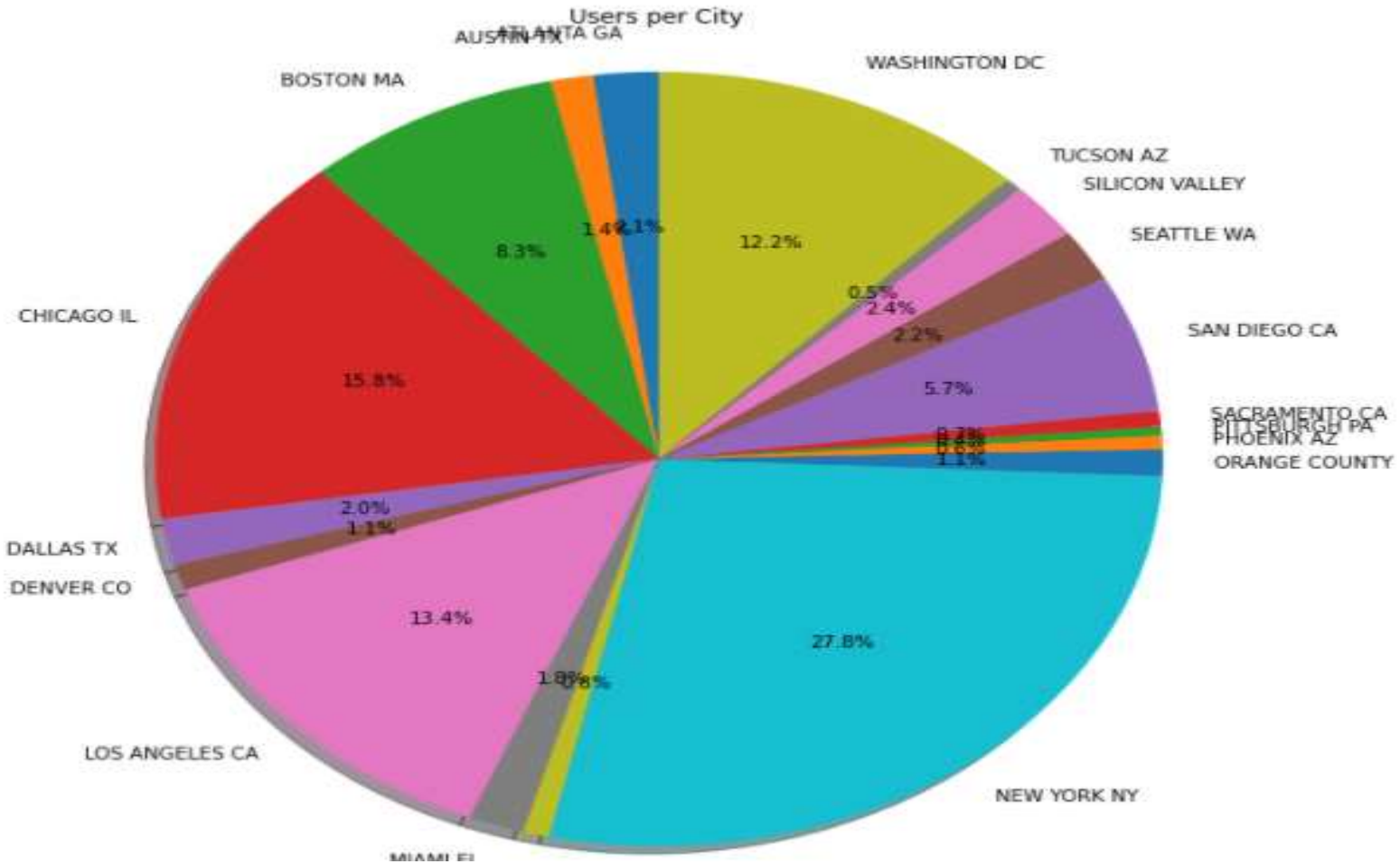
Yellow Cab has higher travels (35000) in the month of December which is the holiday season compared to Pink Cab (11000).

Transaction per Year for both Cabs:



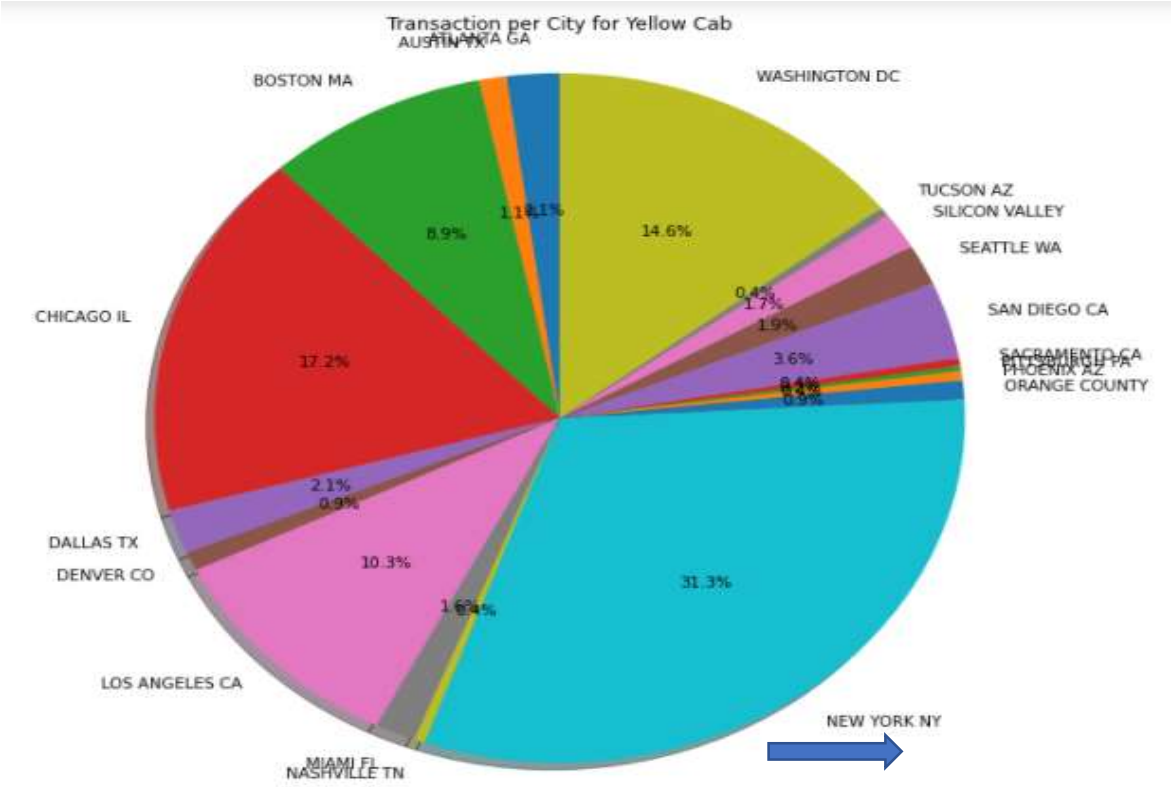
From the graph it shows that on yearly basis no. of transactions for Yellow cab is higher than Pink cab.

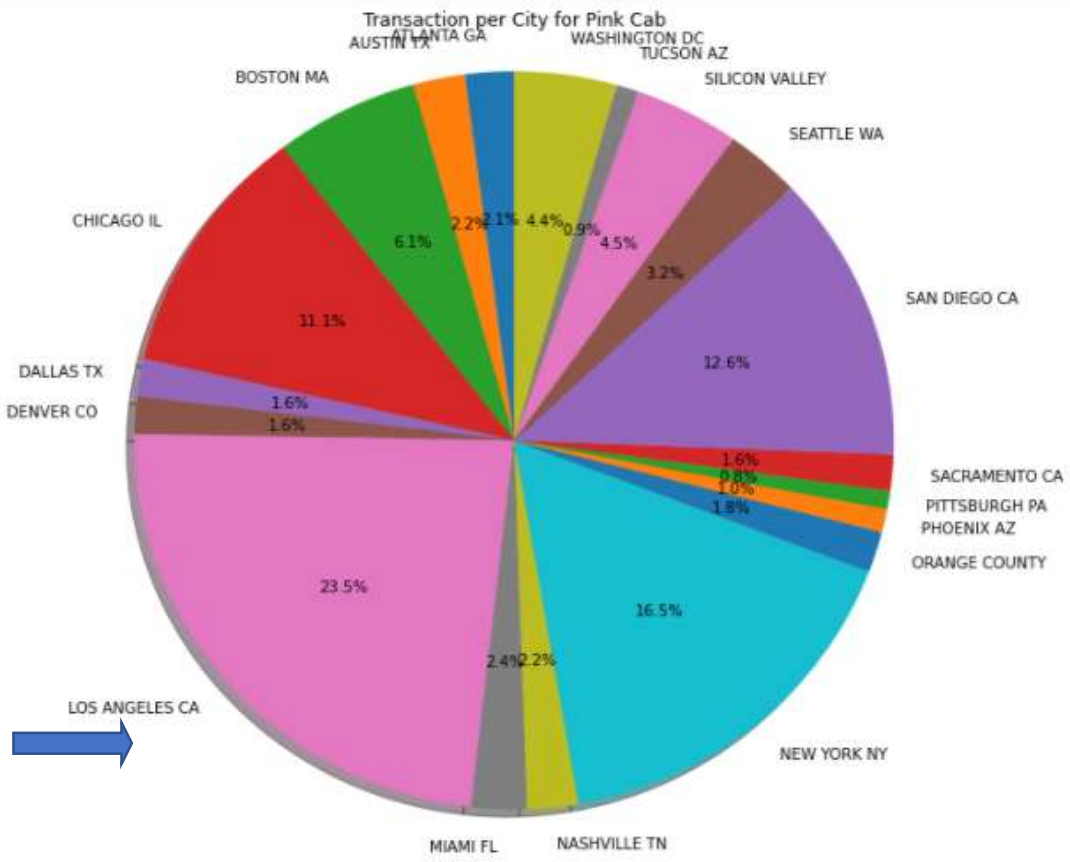
Cab Users per City:



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

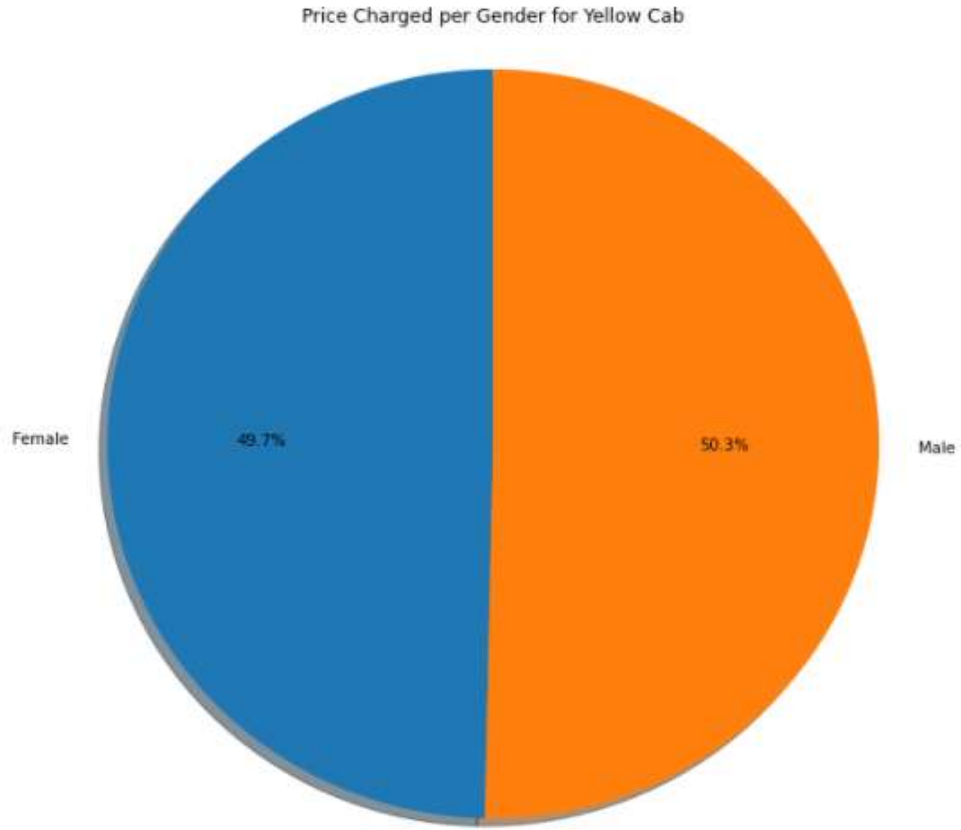
Transaction per City for both Cabs:

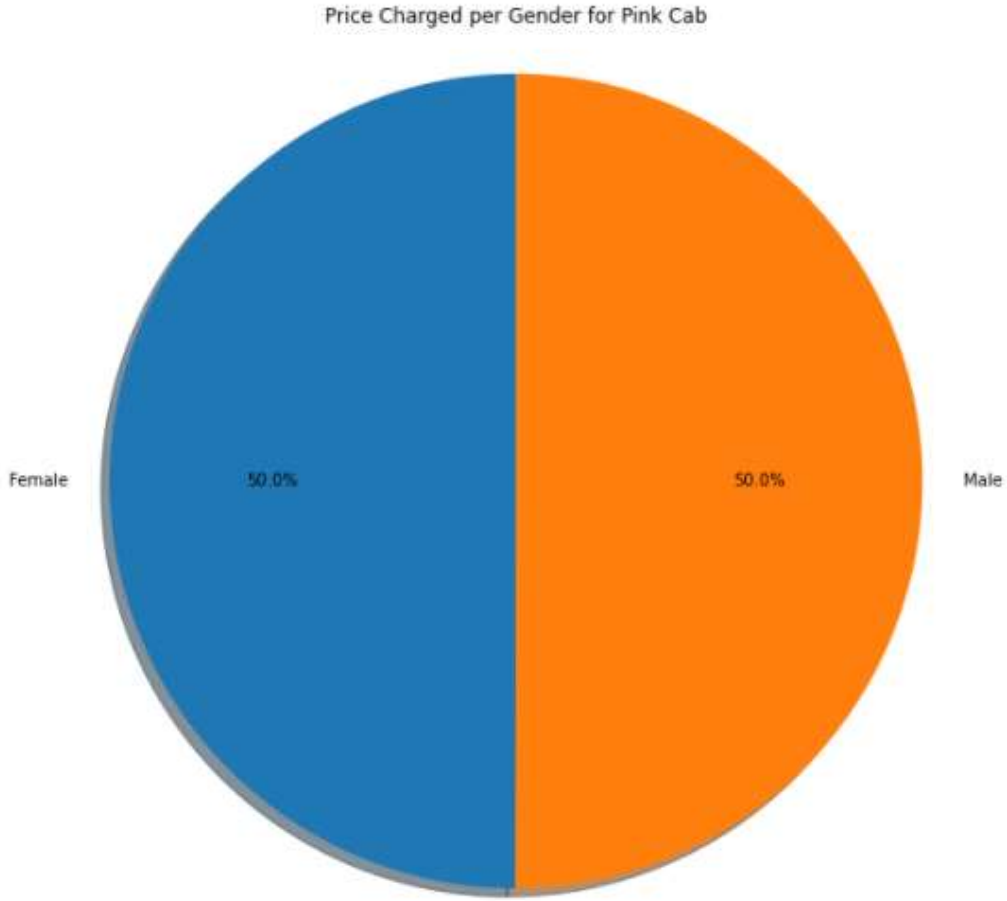




Transaction for Yellow Cab is highest in New York City(31%) and New York City has the highest Cab Users of 28% as per the previous slide.
Transaction for Pink Cab is highest in Los Angeles City.

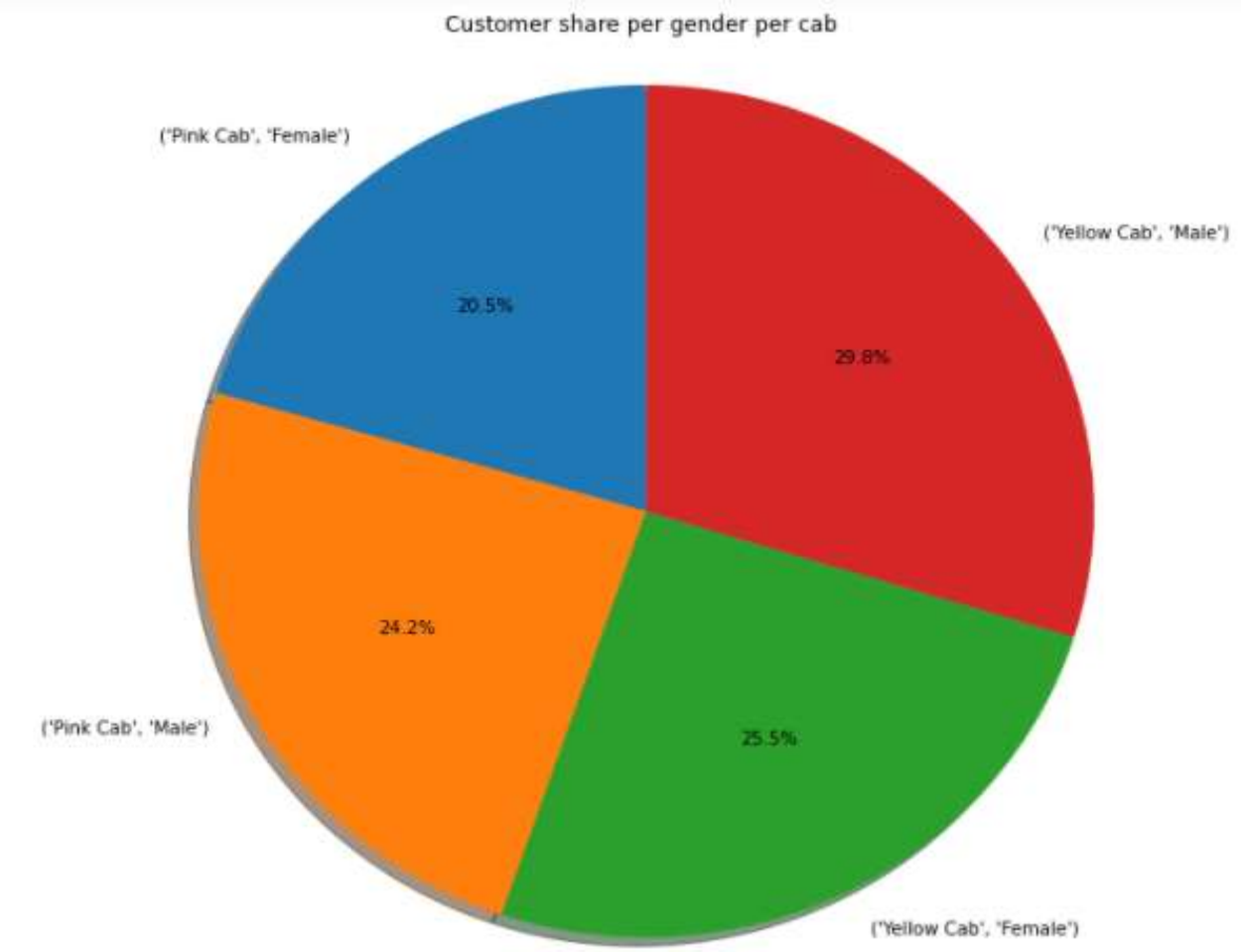
Price Charged per Gender for both Cabs:





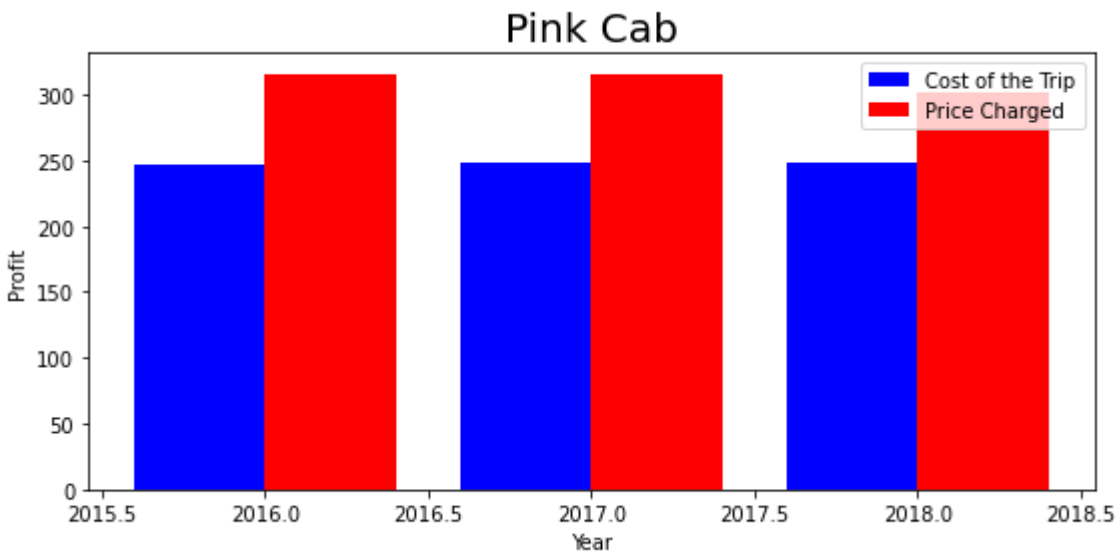
Yellow Cab charge less from Female Customers whereas Pink Cab charges same for both Male and Female Customers.

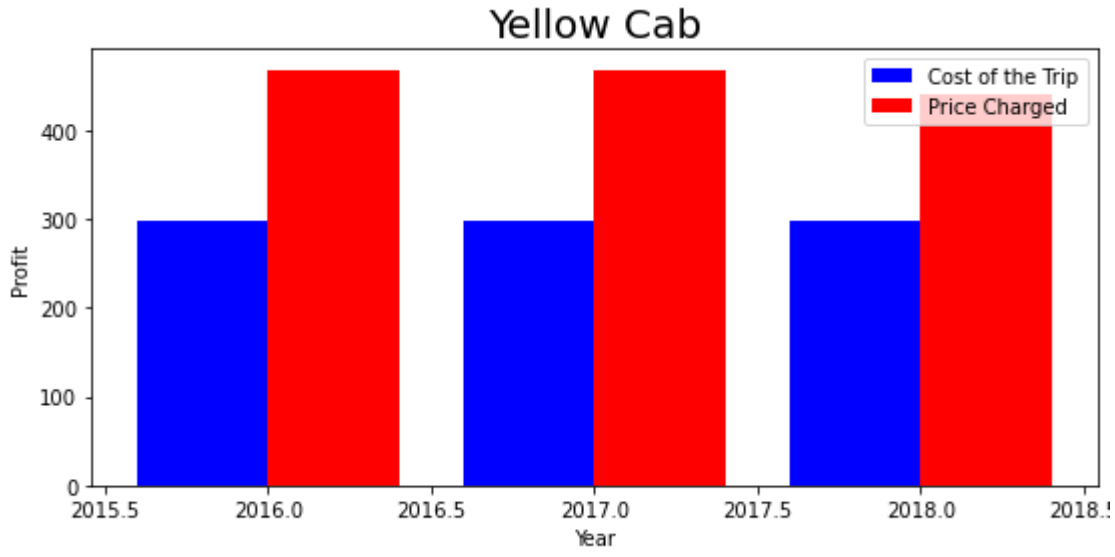
Customer Share per Gender for both Cabs:



Female Customers in Yellow Cab(25.5%) is higher compared to Pink Cab (20.5%)

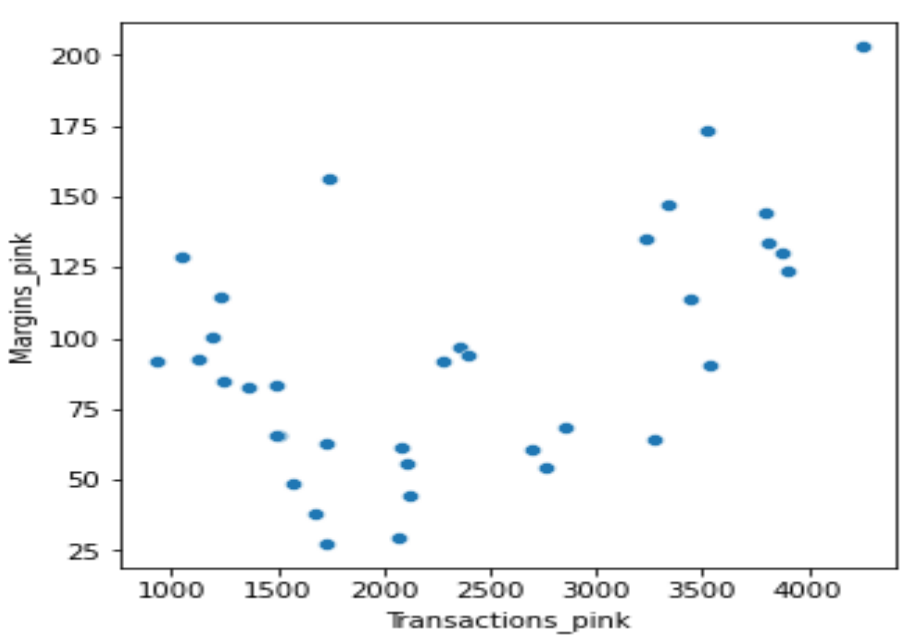
Profit Margin per year for both Cabs:

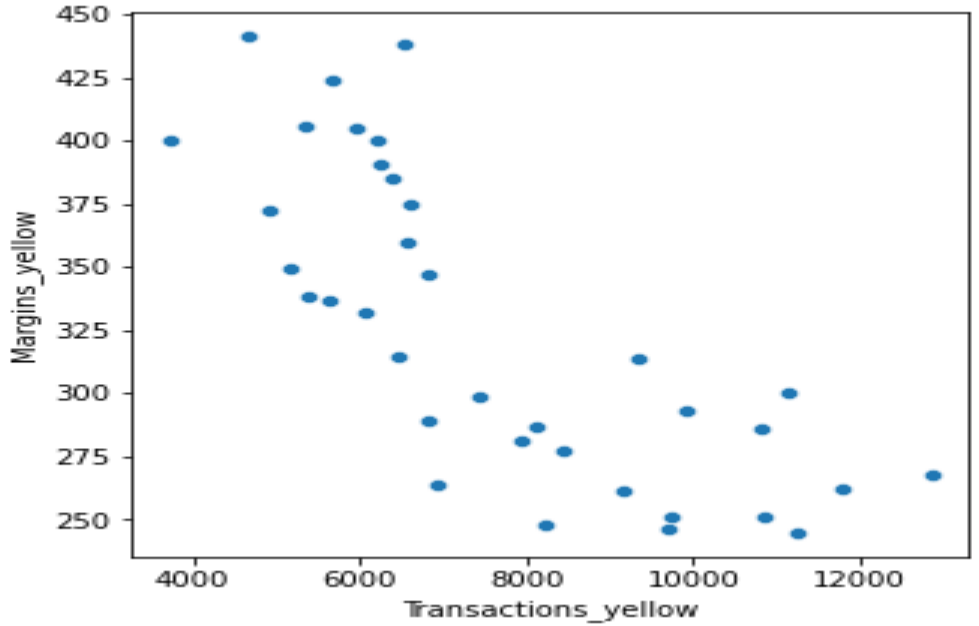




From the Graphs, it shows that the Yellow cab has a higher Profit Margin (Price Charged - Cost of Trip) compared to Pink cab.

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.

Hypothesis Testing

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

  - 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 difference
P value is  0.11515305900425798
```

  - 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  6.060473042494144e-25
```

- **Hypothesis** : **Margin remain the same for all Age group for both Yellow Cab & Pink Cab.**

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

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

```
71228 13483
We accept null hypothesis that theres no difference
P value is  0.3281748754798163
```

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

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

```
231480 43201
We accept alternate hypothesis that theres a difference
P value is  6.4942568177993685e-09
```

- **Hypothesis**: **There is difference in margins for Card payer and Cash payers.**

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

Pink Cab:

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

```
We accept null hypothesis that theres no difference
P value is  0.7900465828793288
```

Yellow Cab:

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

```
We accept null hypothesis that there is no statistical difference
P value is  0.29330606382985325
```

Recommendation

- **Transaction per year**: For Yellow Cab Transaction per year from 2016 to 2018 is almost double than Pink Cab.
- **Margin per Gender**: For Yellow Cab there is difference in Margin between Male and Female Customers due to which Female Customer percentage is higher in Yellow Cab in comparison to Pink Cab.
- **Profit Margin**: For Yellow Cab the Profit Margin is higher per year from 2016 to 2018 in comparison to Pink Cab.
- **Margin per Age**: In Yellow Cab there is difference in Margin for people older than 50 yrs, whereas in Pink Cab there is no difference in Margin of all age group.
- 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%.

On the basis of the above points, Yellow Cab is recommended for investment.

Thank You