G2M Insight for Cab Investment Firm

Company Name: XYZ

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Problem Statement

XYZ is a private equity firm in the USA. 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.

Provide meaningful insights to help XYZ firm identify the right company to invest with.

There are two cab companies to consider:

- Yellow cab
- Pink cab

The analysis incudes:

- Data Understanding
- Data visualization
- Hypothesis Testing

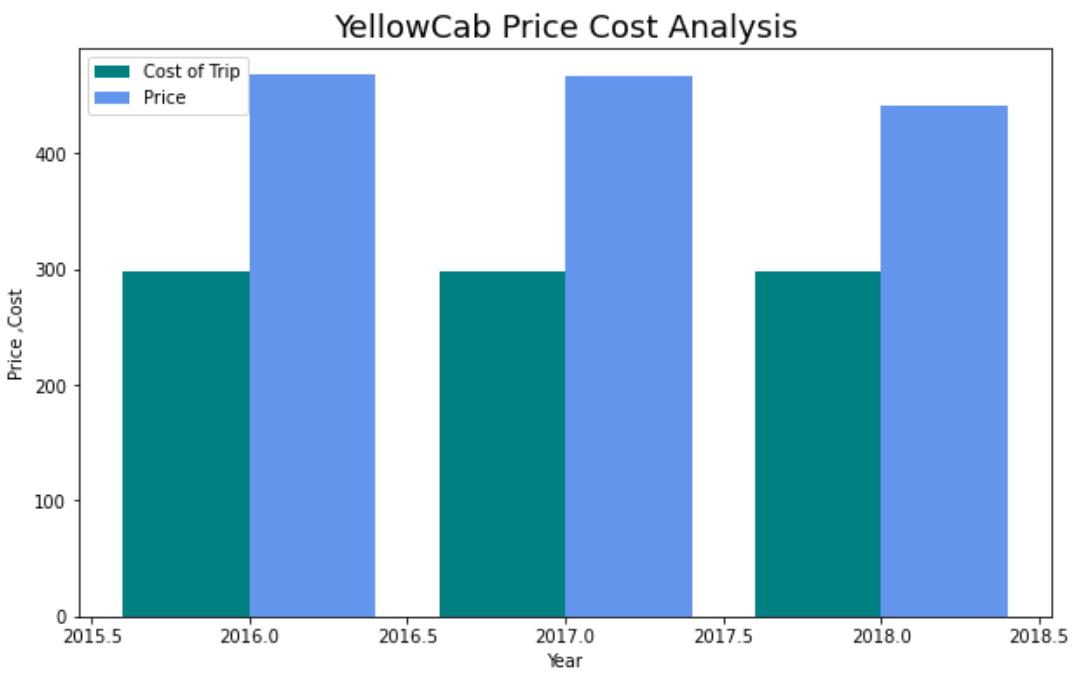
Data Preparation

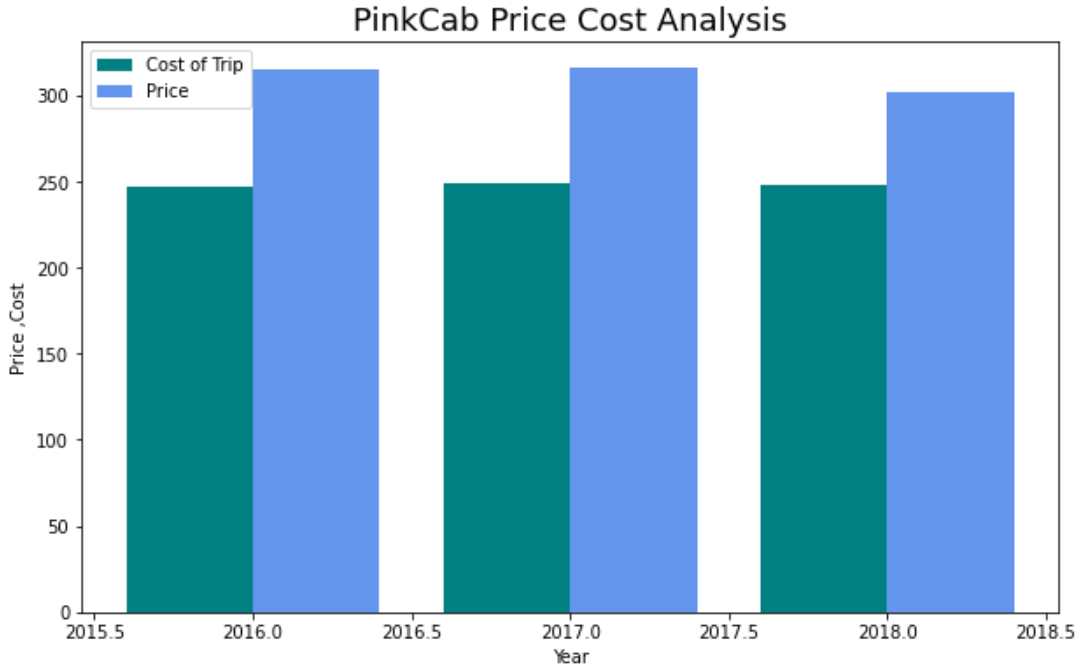
There are four datasets :

- City.csv –This file contains a list of the US cities ,their population and the number of cab users
- Customer_ID.csv –This is a mapping table that contains a unique identifier which links the customers' demographic details.
- Cab_Data.csv –This file contains the details of transactions for the yellow cab company and the pink cab company.
- Transaction_ID.csv-This is a mapping table that contains the transactions details of a customer and the payment mode.

EXPLORATORY DATA ANALYSIS

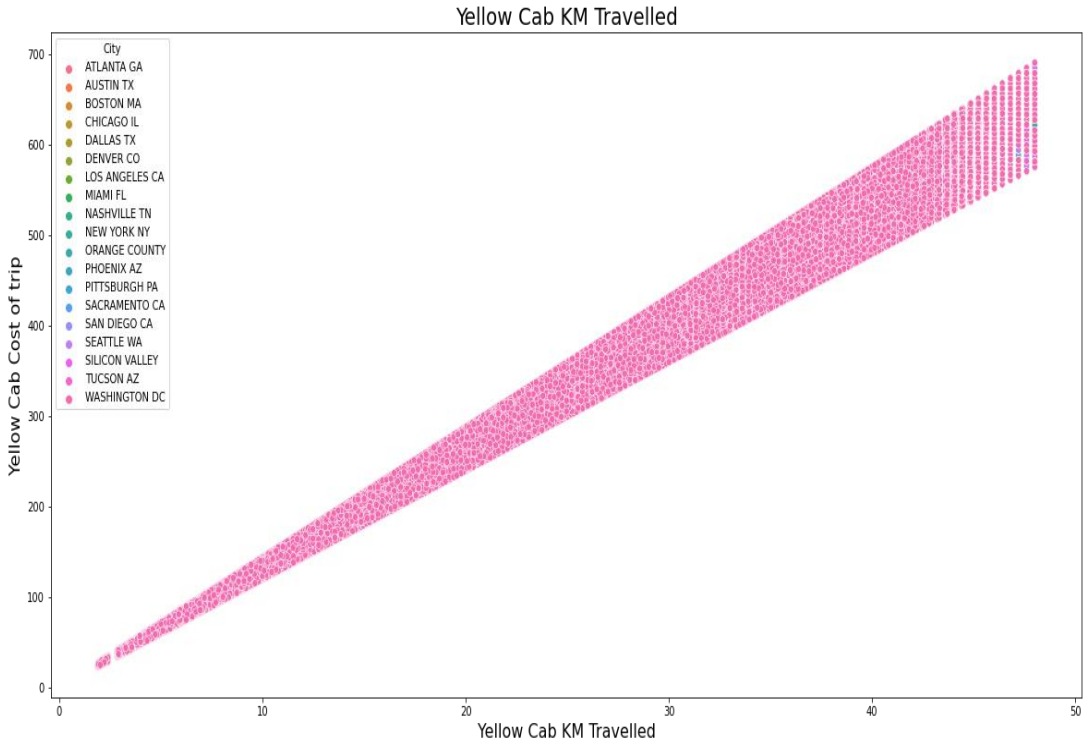
Price Cost Analysis

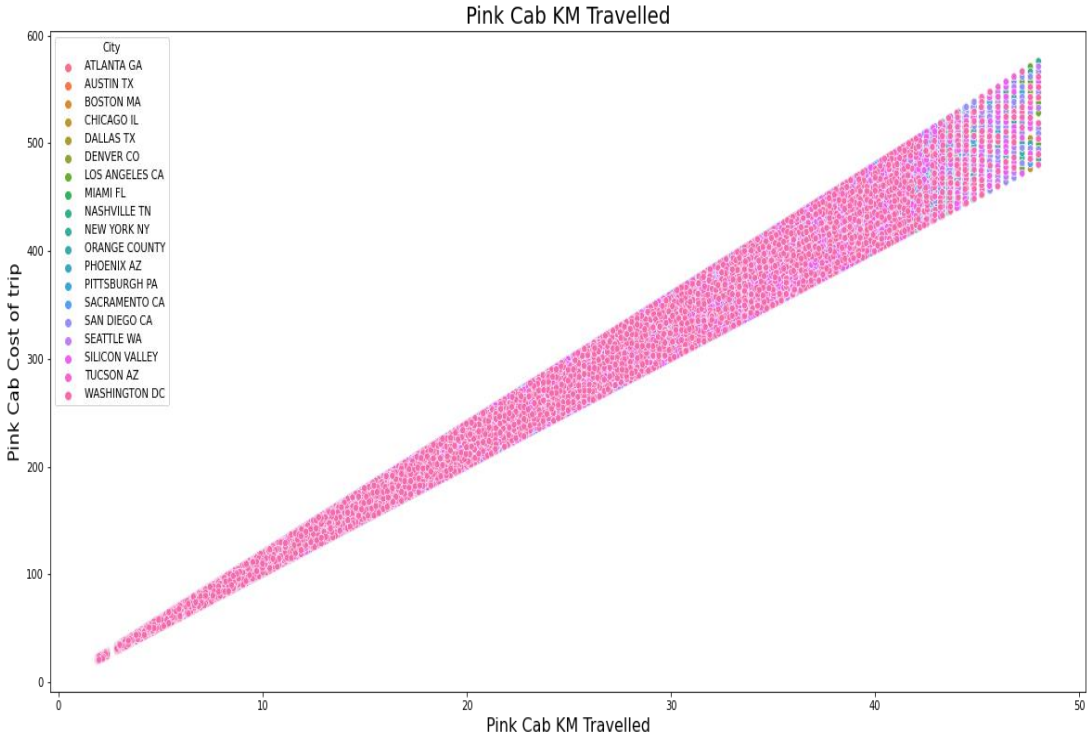




- **From the graphs we can see that yellow cab has higher margins(price-cost) throughout the years 2016-2018 compared to pink cab**

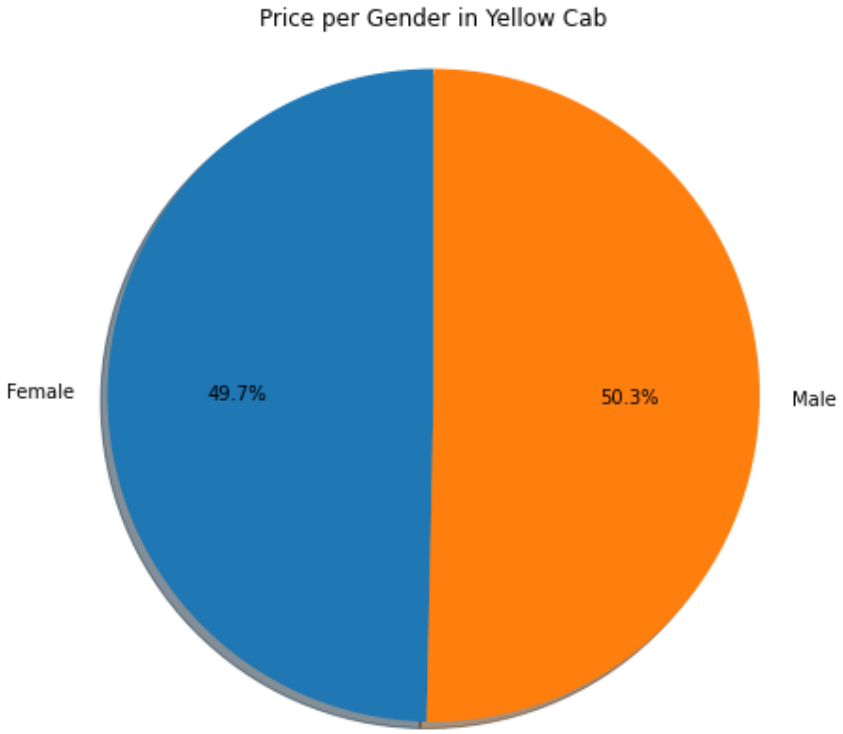
KM Travelled vs Cost

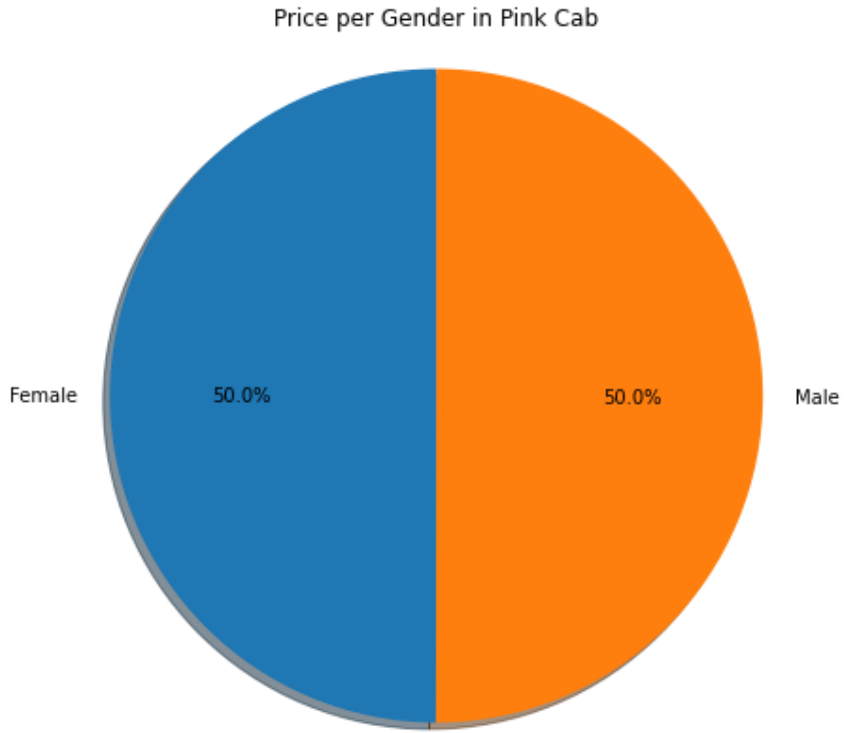




- **We see that New York charges the highest prices in comparison to other cities**

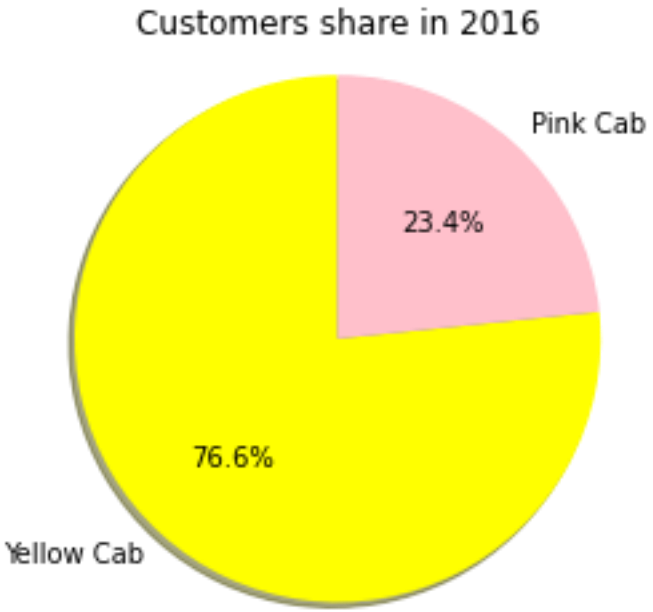
Price Per Gender Analysis

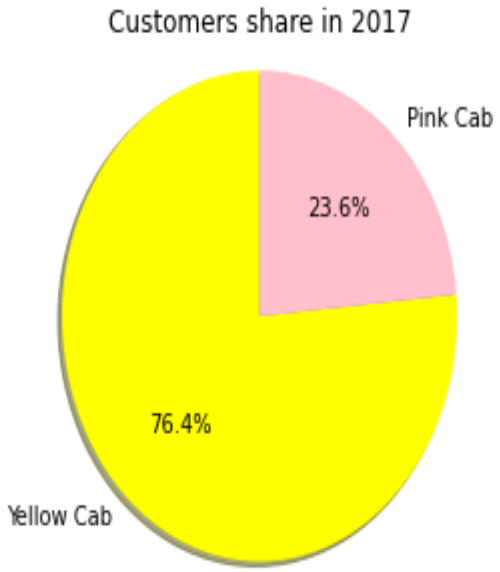


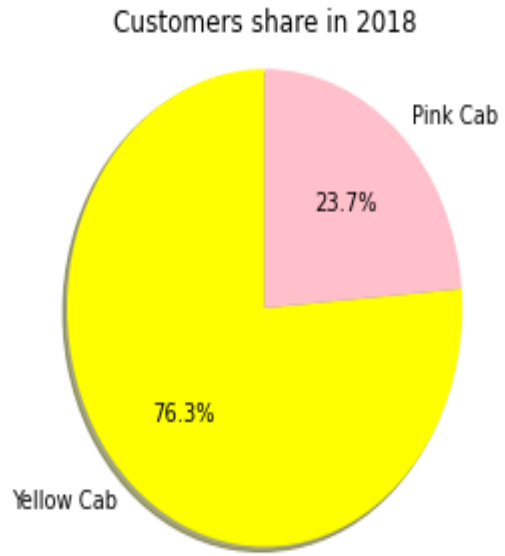


- **Pink cab charges the same for all customers regardless of gender whereas yellow cab charges male customers more**

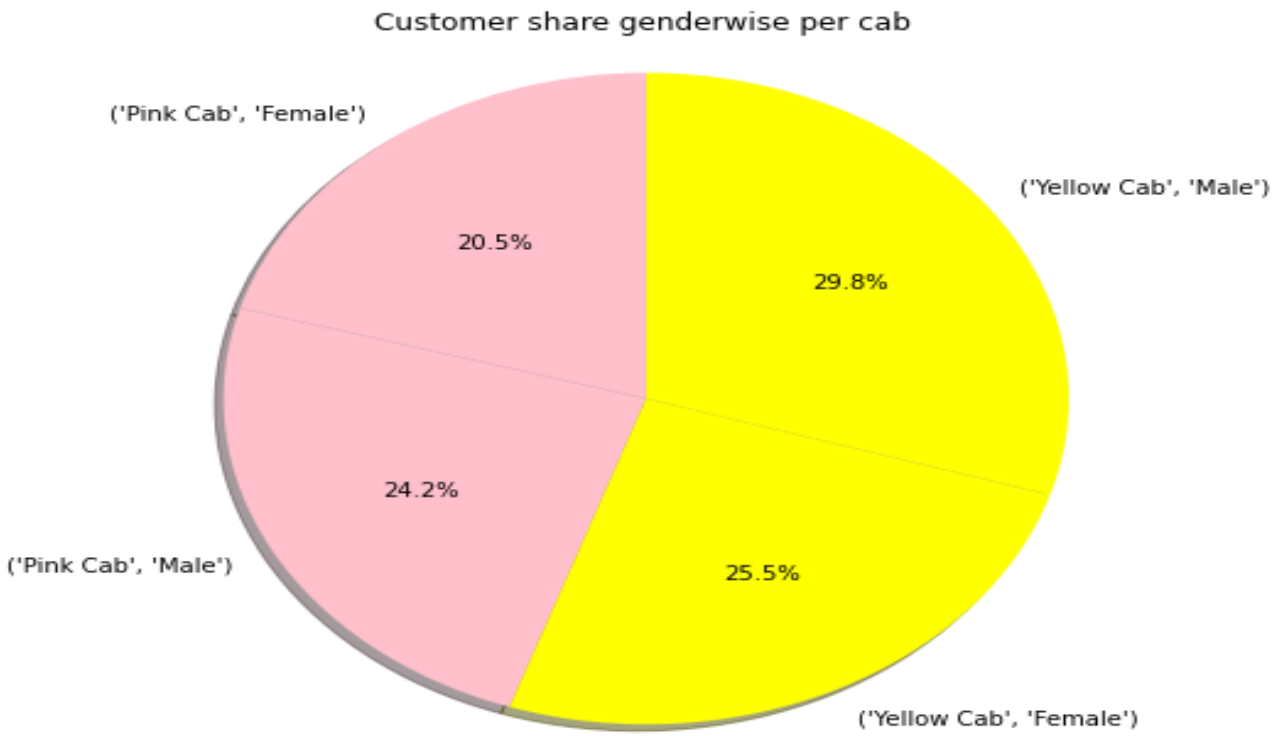
Customer Share Analysis





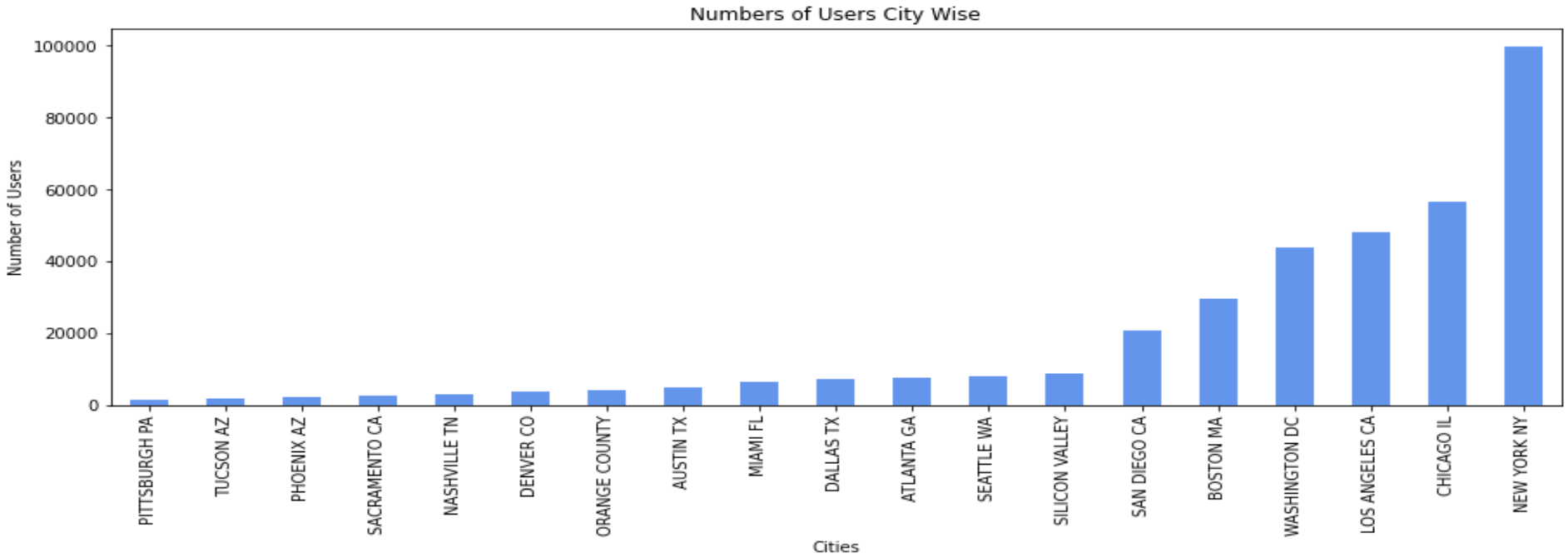


Customer Share Gender wise Per Cab



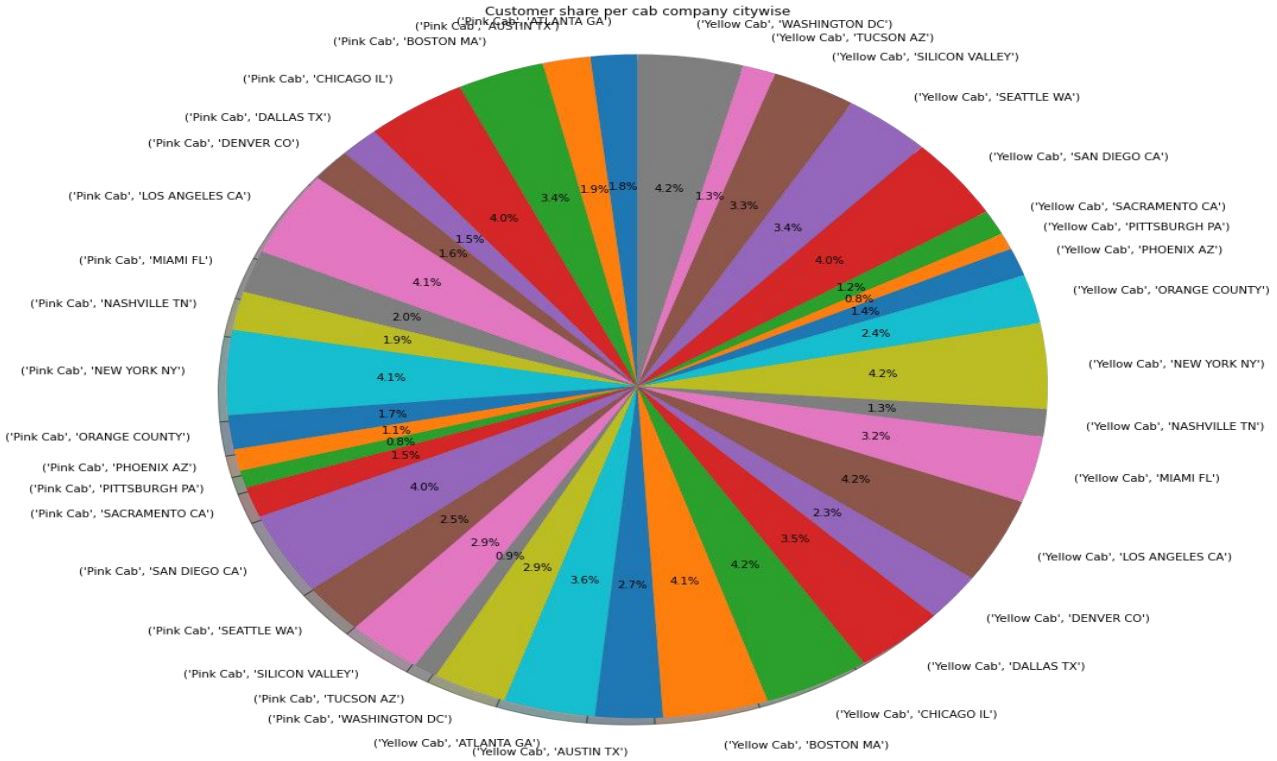
- **From the graph we see that yellow cab has more customers of each gender compared to pink cab**

Number of Cab Users Per City



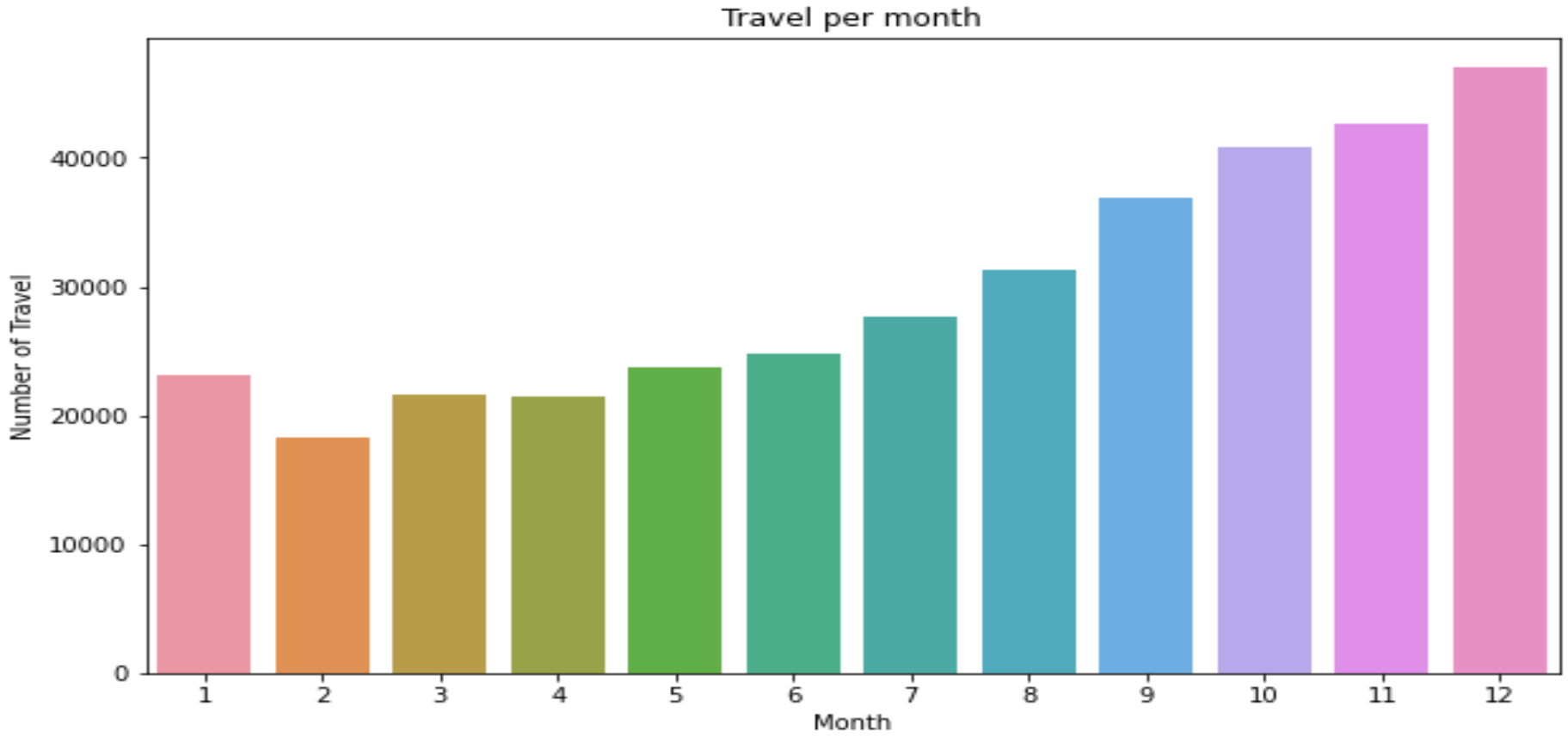
- **New York City has the highest number of cab users**

Customer Share Per Cab Company Per City



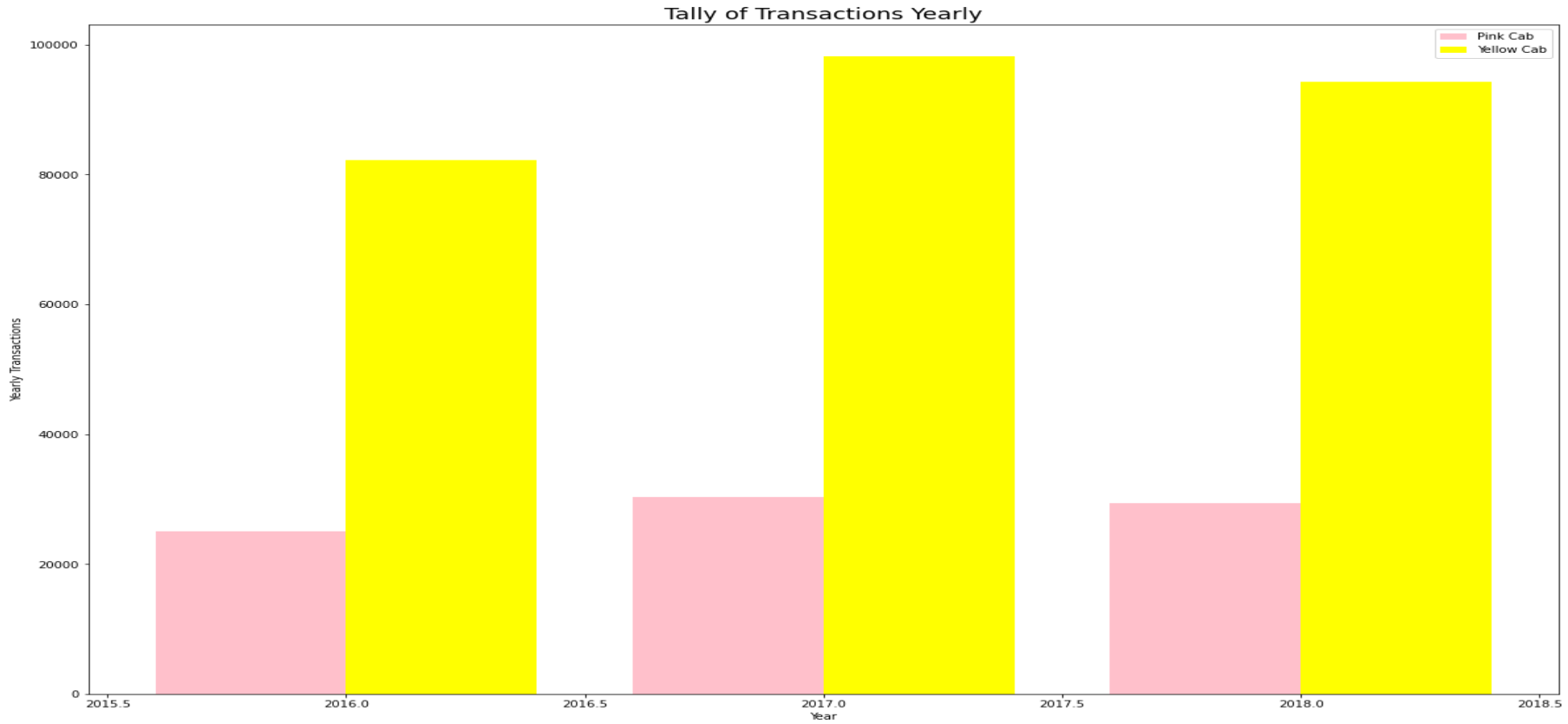
- **Yellow cab has a higher customer share compared to pink cab in majority of the cities**

Travel Per Month Analysis



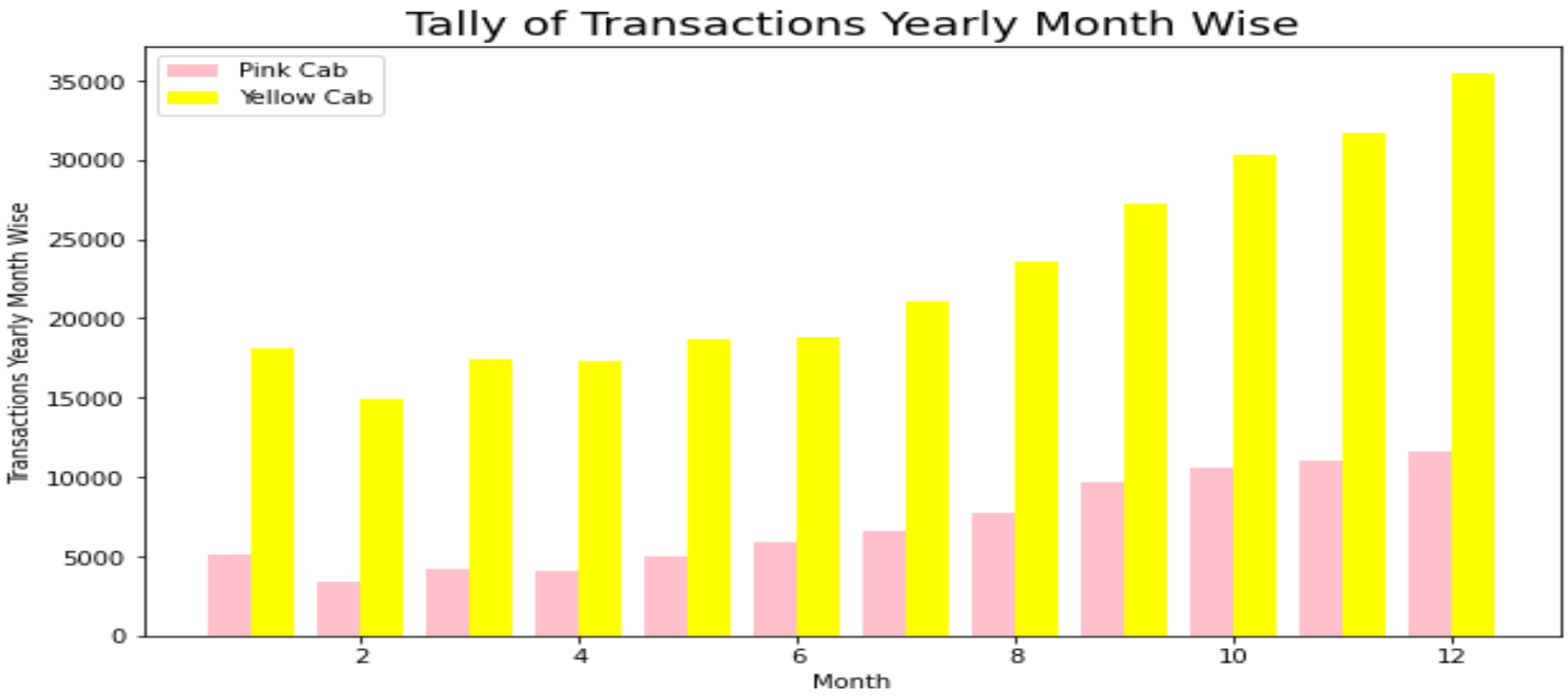
- **From the above graph we see that cabs are mostly used in the month of December most probably due to the festivities.**

Tally of Transactions Yearly



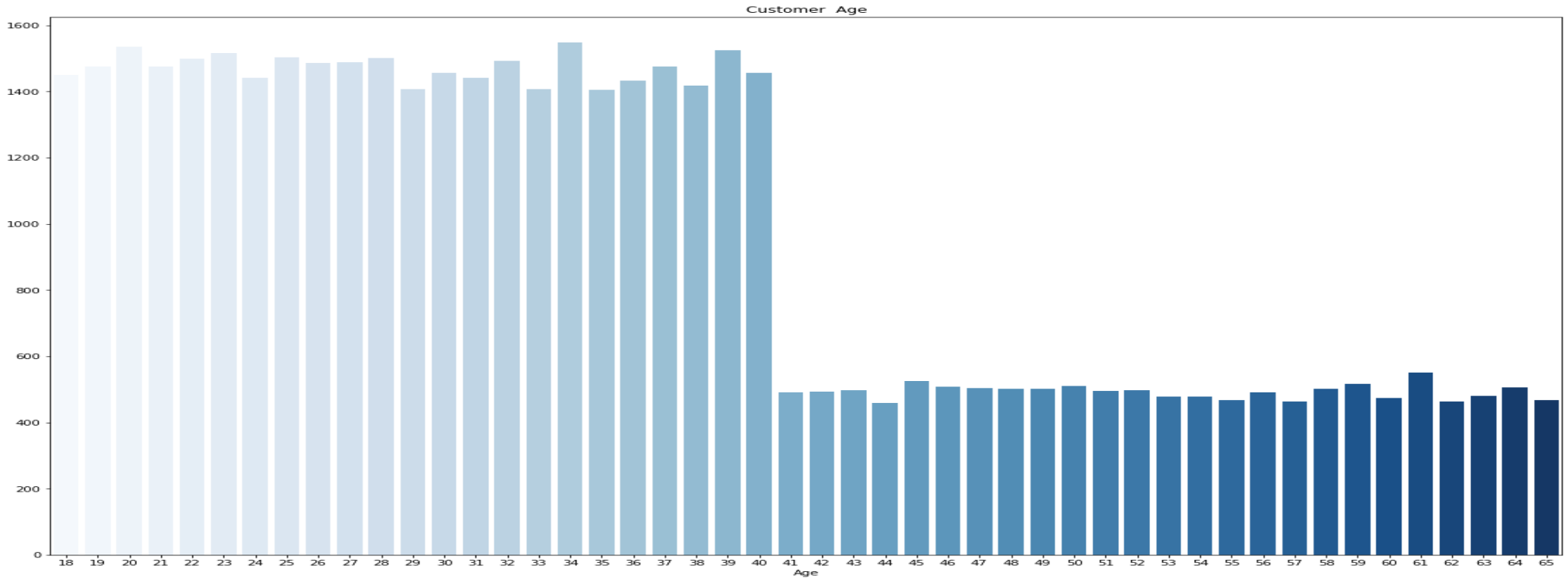
- **We can see that Yellow cab had more transactions compared to pink cab throughout the years**

Tally of Transactions month wise



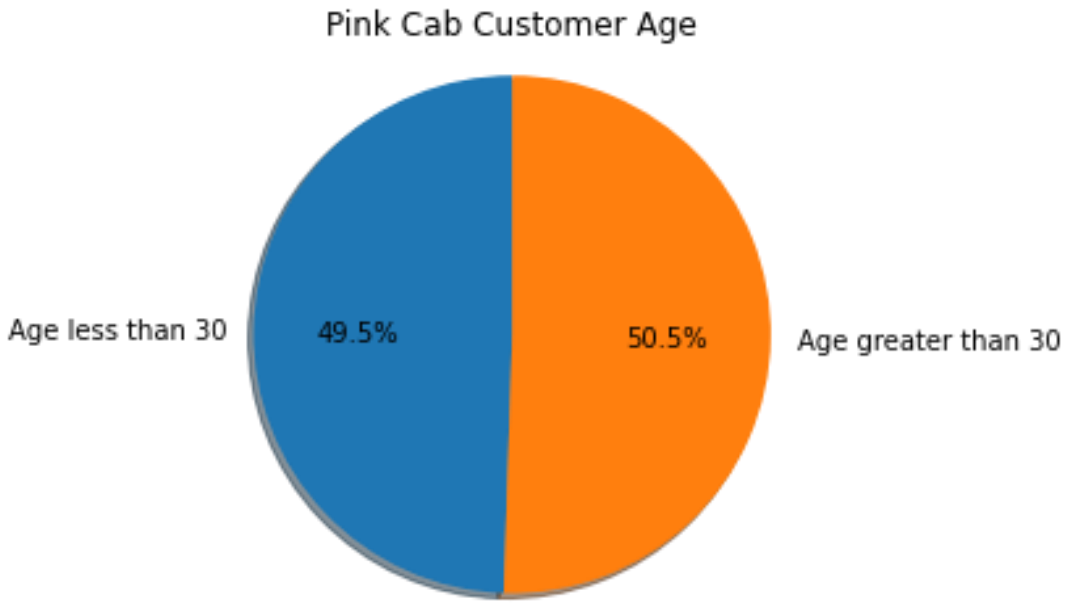
- **Pink cab had a lower number of transactions compared to Yellow cab in every month**

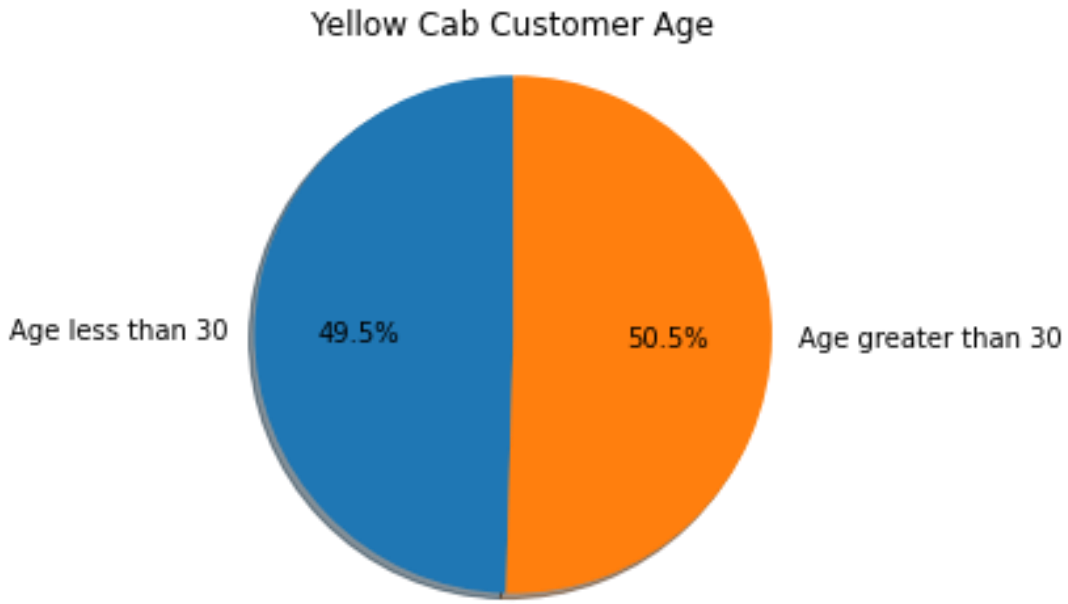
Customer Age Analysis



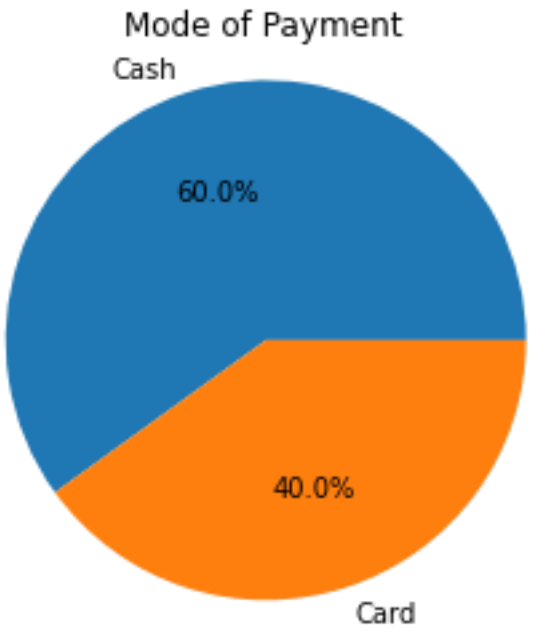
- **Majority of cab users are 18-40 years**

Customer Age Analysis per Cab





Mode of Payment



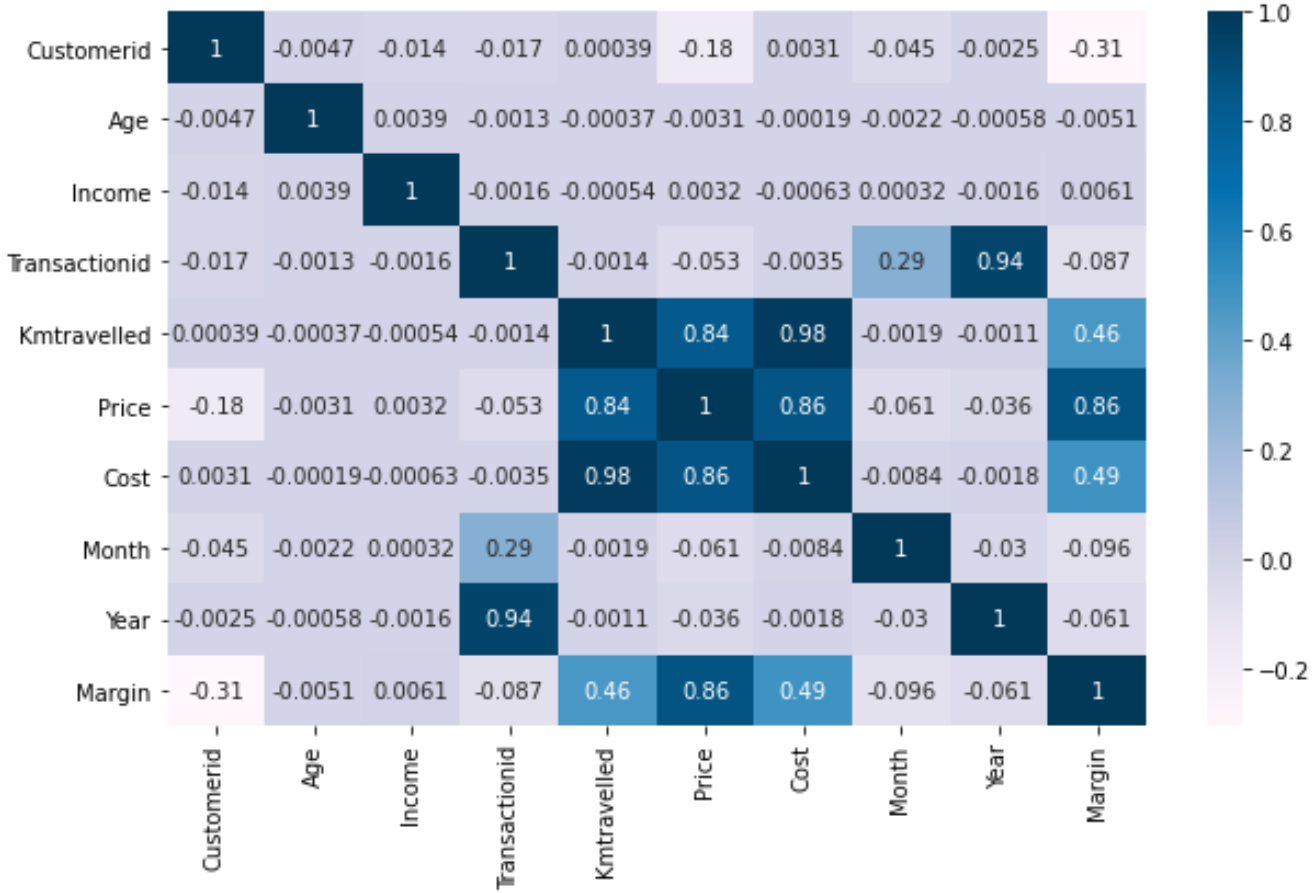
- **Majority of the cab users prefer to pay using cash**

EDA SUMMARY

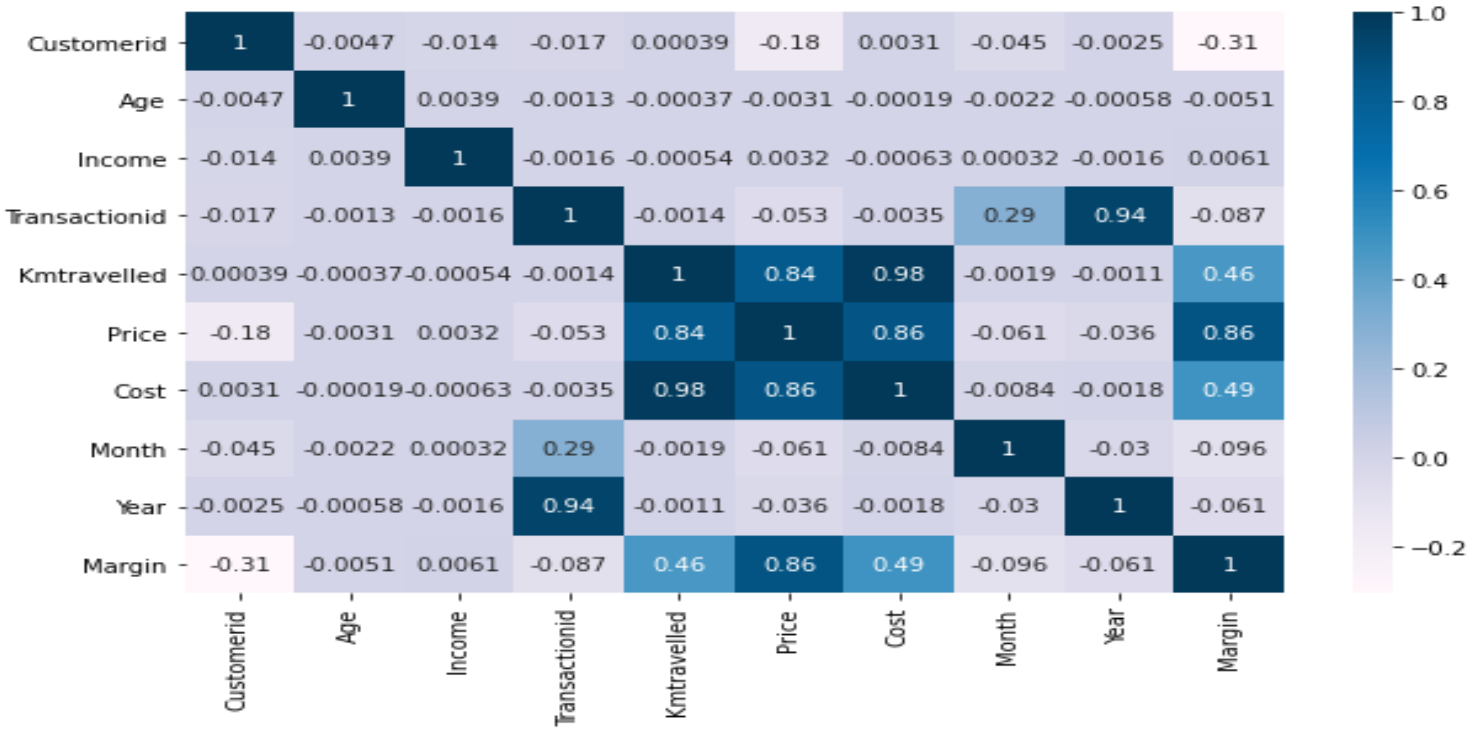
- Yellow cab charges the male customers more compared to the female customers whereas the pink cab charges the same for all their customers regardless of gender.
- Yellow cab has a high margins of profit from 2016 to 2018 while the profit margin for the pink cab is low from 2016 to 2018.
- Yellow cab has higher customer share (76%) in every year from 2016 t0 2018 while pink cab has a low customer share (23%).
- Yellow cab has way more transactions(300,000) compared to pink cab

Which has (120,000) transactions from 2016 to 2018

Correlation Analysis



Correlation Analysis



- **We See that there is a positive correlation between price and the cost charged**

HYPOTHESIS TESTING

HYPOTHESIS : COMPANY MARGIN REMAINS THE SAME REGARDLESS OF THE GENDER

Yellow cab : Gender does affect the company margin

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

```
We accept alternate hypothesis that theres a difference
P value is  6.060473042494144e-25
```

Pink Cab : Gender does not affect the company margin

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

```
We accept null hypothesis that theres no effect
P value is  0.115153059004258
```

HYPOTHESIS: Gender has no effect on the distance travelled

Yellow Cab: There is no difference in the distance travelled regardless of the gender

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

```
We accept null hypothesis that theres no effect
P value is  0.5146654429411317
```

Pink Cab: There is no difference in the distance travelled regardless of the gender.

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

```
We accept null hypothesis that theres no effect
P value is  0.6164626165258722
```

HYPOTHESIS : Customer Age has no effect on their cab company preference

Yellow Cab : The customers age does not influence their cab preference for yellow cab

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

```
We accept null hypothesis that theres no effect
P value is  0.08738489267415496
```

Pink Cab : The customers age does not influence their cab preference for pink cab

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

```
We accept null hypothesis that theres no effect
P value is  0.2872878897358956
```

HYPOTHESIS : Customers income affects their choice of cab

Yellow Cab : The customer's income does not affect their choice of the cab

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

```
We accept null hypothesis that theres no effect
P value is  0.699146952631954 3
```

Pink Cab: The customer's income does not affect their choice of cab

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

```
We accept null hypothesis that theres no effect
P value is  0.609741651636732 2
```

RECOMMENDATION

Profit Margin : Yellow cab has higher profit margins compared to pink cab form the year 2016 to 2018

Customer Share : Yellow cab has a higher customer share compared to pink cab

Transactions : Yellow cab has more transactions that pink cab

Thank You