Data Glacier

Your Deep Learning Partner

Project:
G2M Insight For Cab Investment Firm
Data Science Virtual Internship

Submitted by : Siddharth Mahindru
Date   : 19-July-2022

Approach

Understanding Bussiness Problem

Understanding Datasets Provided

Bivariate And Multivariate Analysis

Profit Analysis

EDA Results

Recommendation

Hypothesis Testing

Understanding Bussiness Problem

Data Glacier
Your Deep Learning Partner

Bussiness Problem

XYZ is a private firm in US. XYZ wants to invest in US Cab Company. US has 2 major cab companies - Pink Cab and Yellow Cab. We must find out which Cab company would be good choice to invest in for XYZ company. For this purpose, we will mostly focus on profitable cab company. Whichever cab company seems suitable in terms on market value, we would invest in that.

Understanding Datasets Provided

Information About the Datasets Provided

1. Cab_Data.csv – this data file includes the details of transactions for 2 cab companies
2. Customer_ID.csv – this file contains the information about customer's demographics.
3. Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and mode of payment.
4. City.csv – this data file contains an information about US cities, their population, and the number of cab users

All These datasets were merged, to form the final dataset for performing EDA.
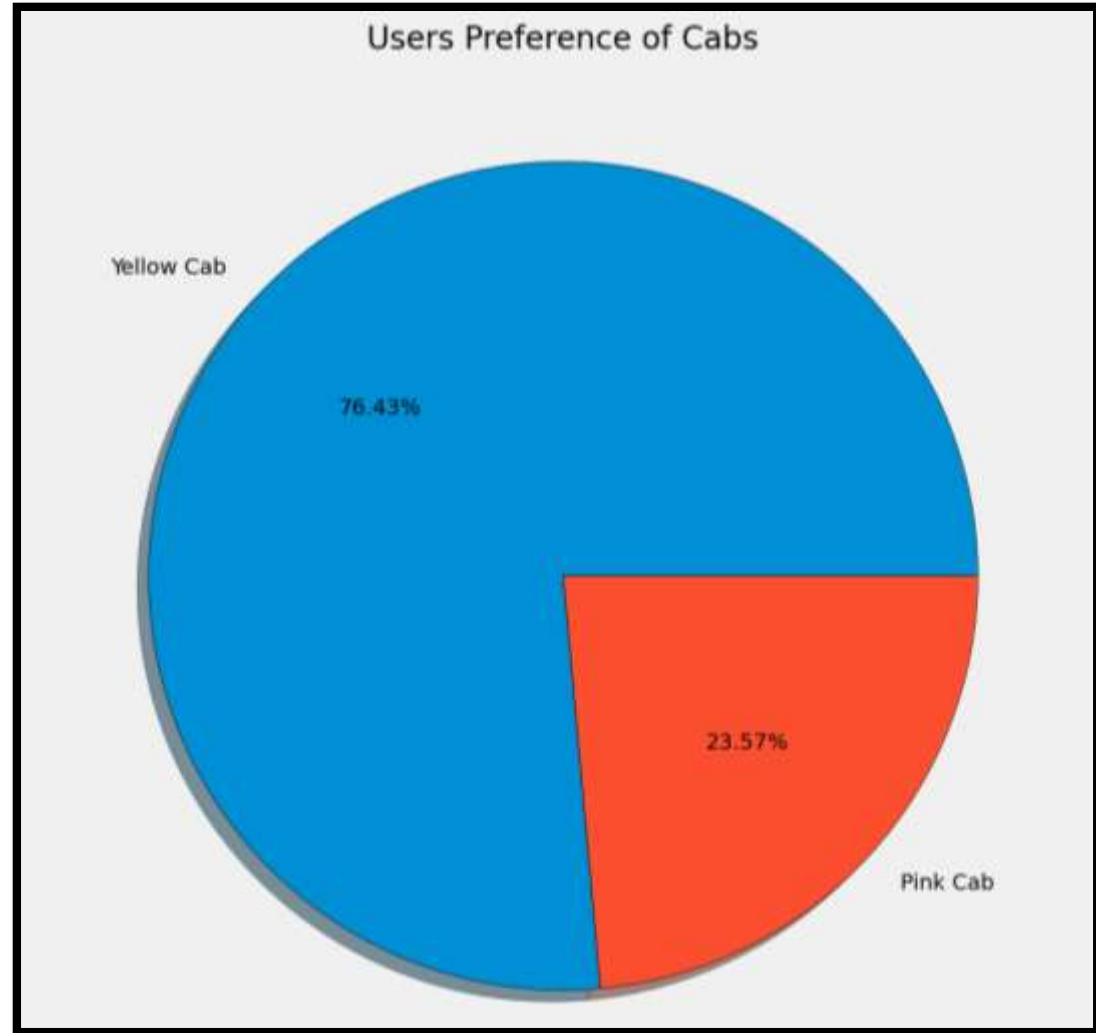
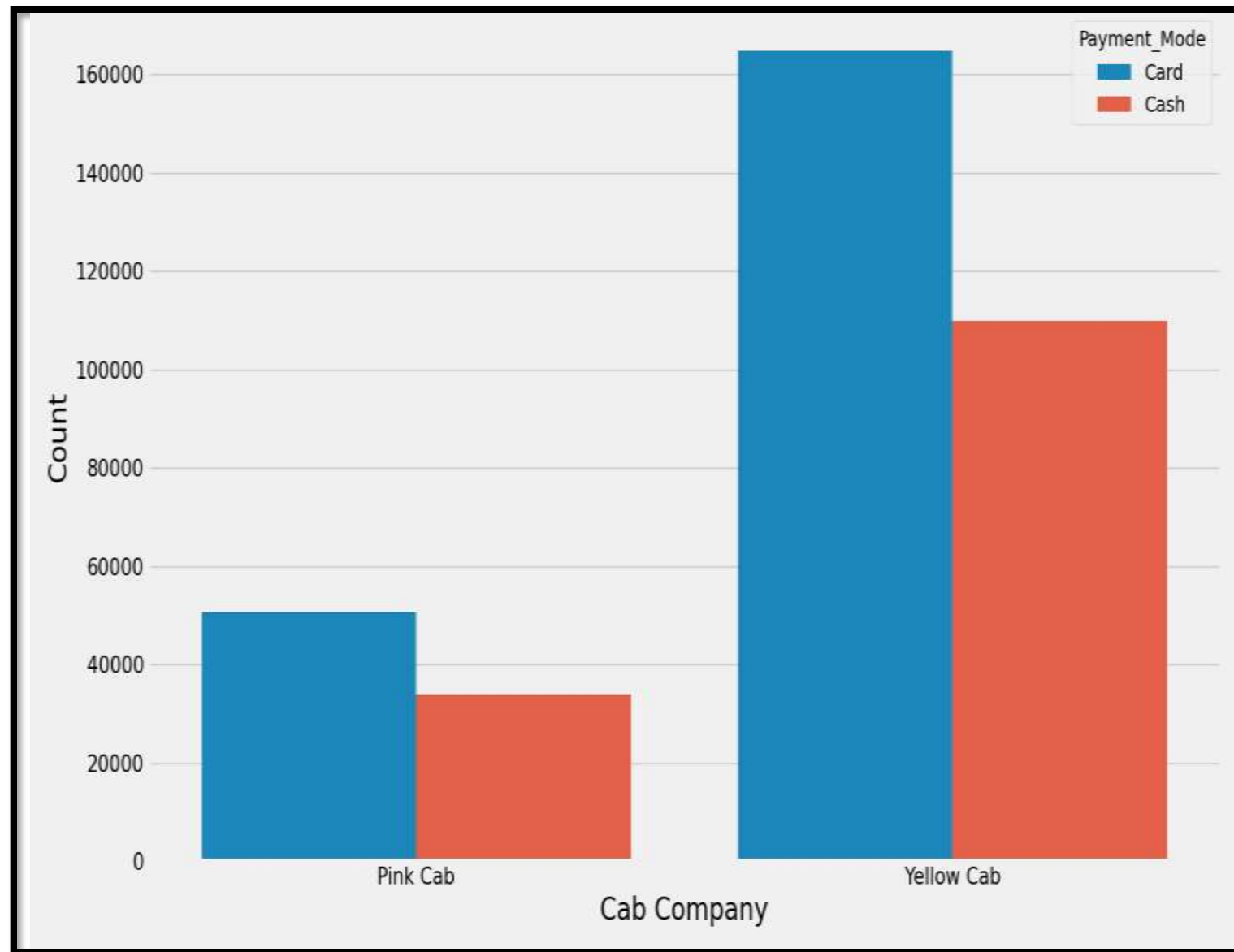Performing Bivariate And Multivariate Analysis to find out Meaning-full Insights

Which Company has more Users?

Users Preference of Cabs

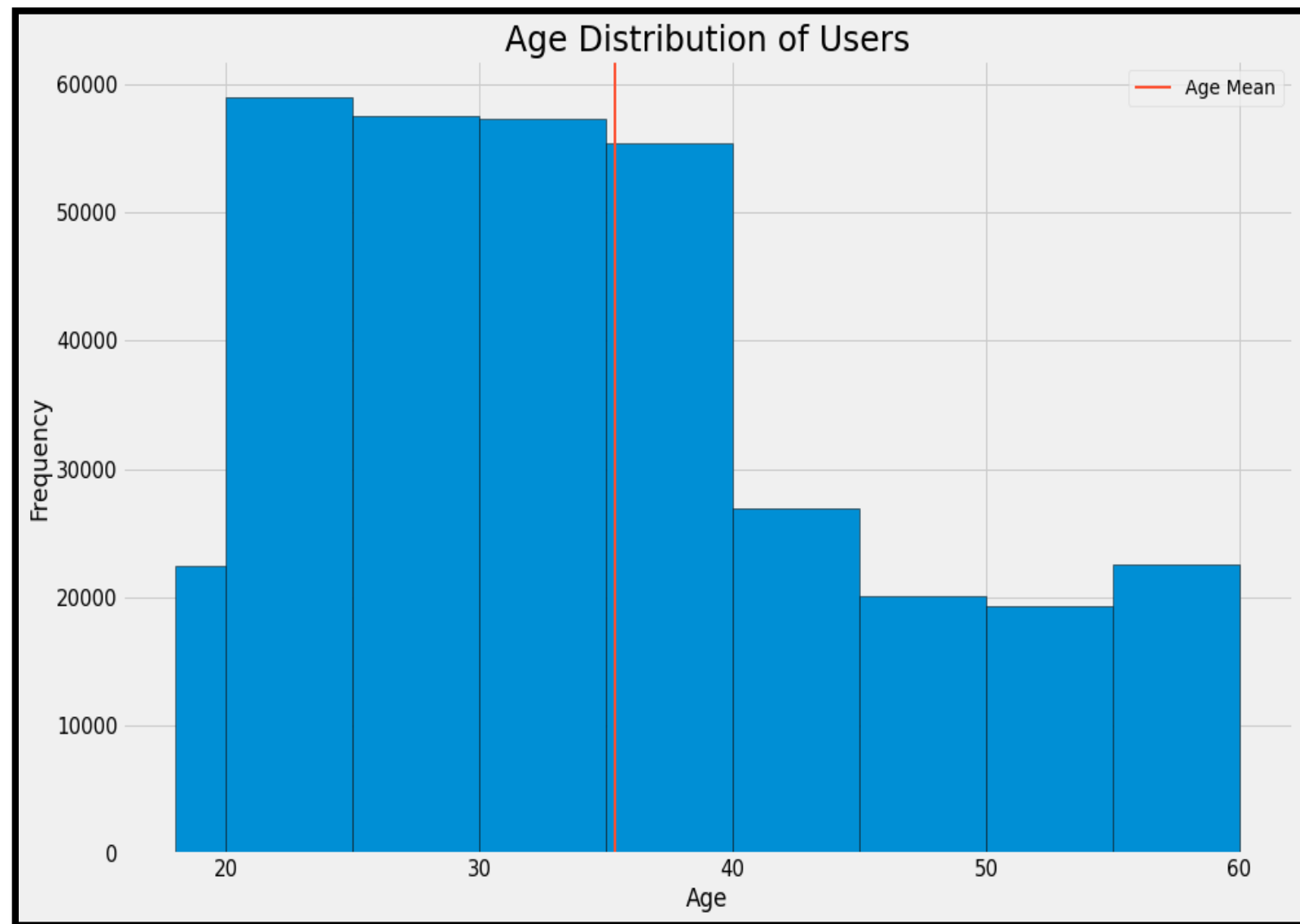Yellow Cab — 76.43%

Pink Cab — 23.57%

Most Users prefer to use Yellow Cab Than Pink Cab

Payment Method of Each Company Users

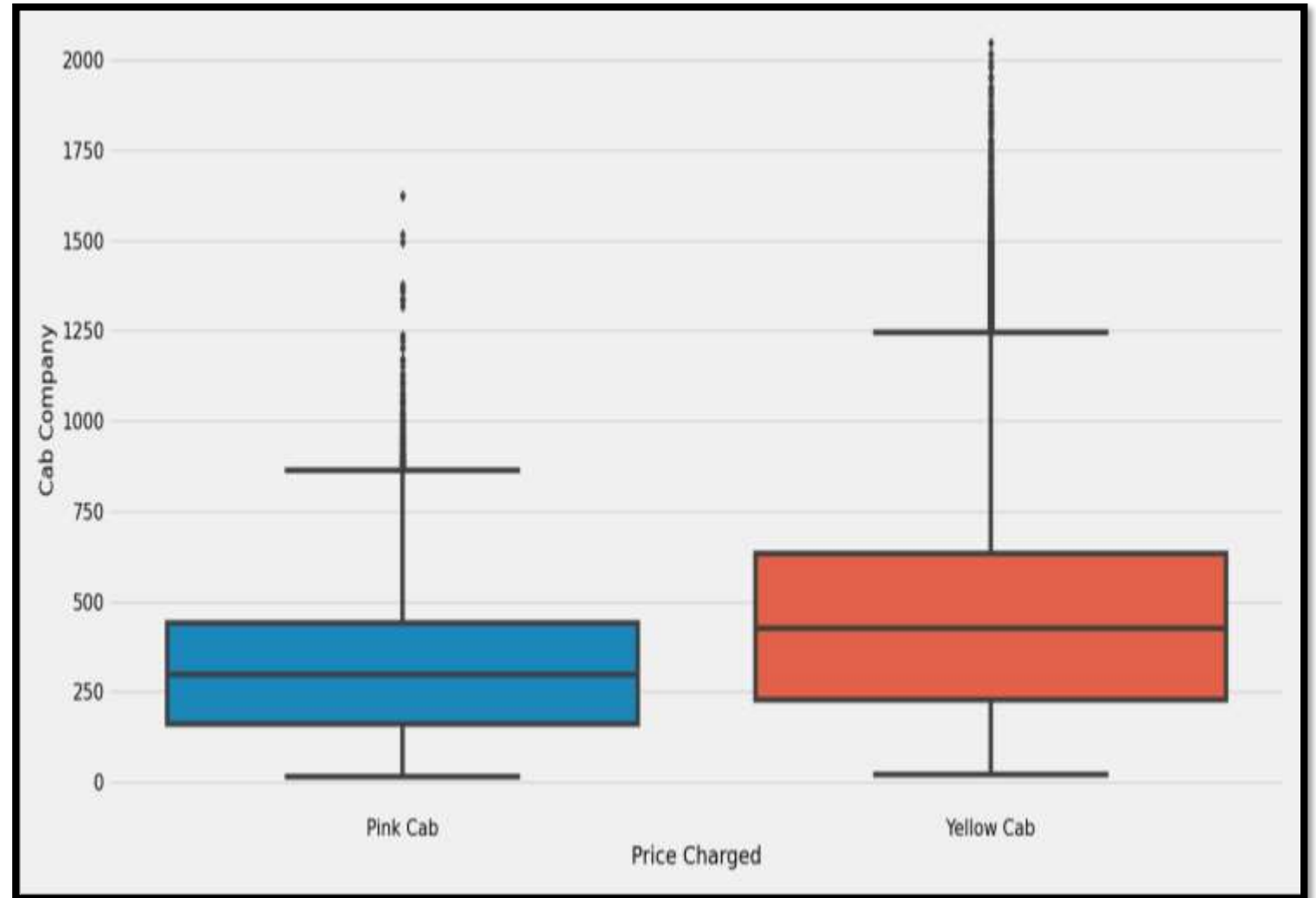Most users prefer to pay by Card than Cash

Age Distribution of Users

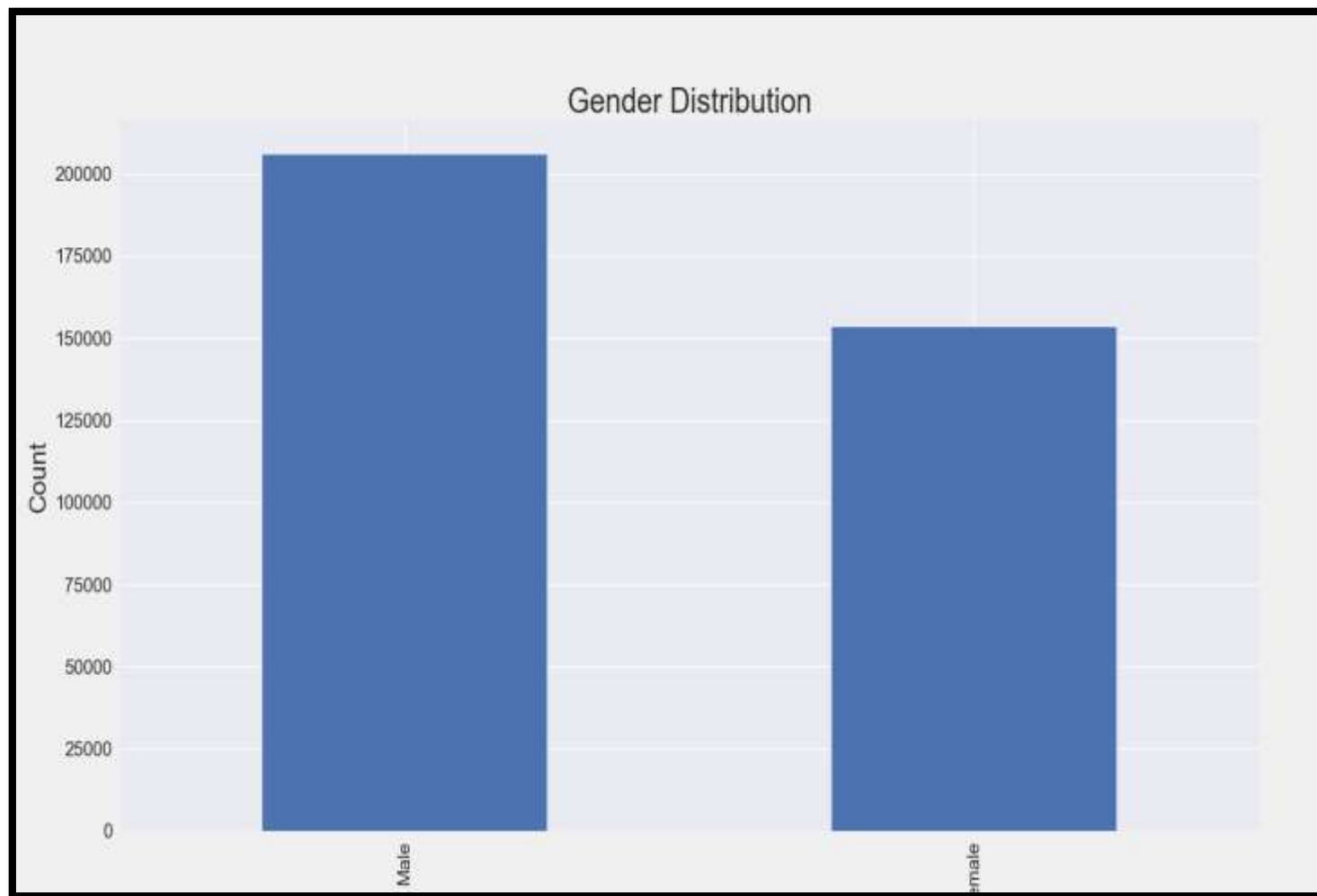This graph shows that Most of the Users were in between age 20 to 40.

Price Charged for Both Cab Companies

This graph shows that the price charged by Yellow cab is comparatively more than Pink Cab.
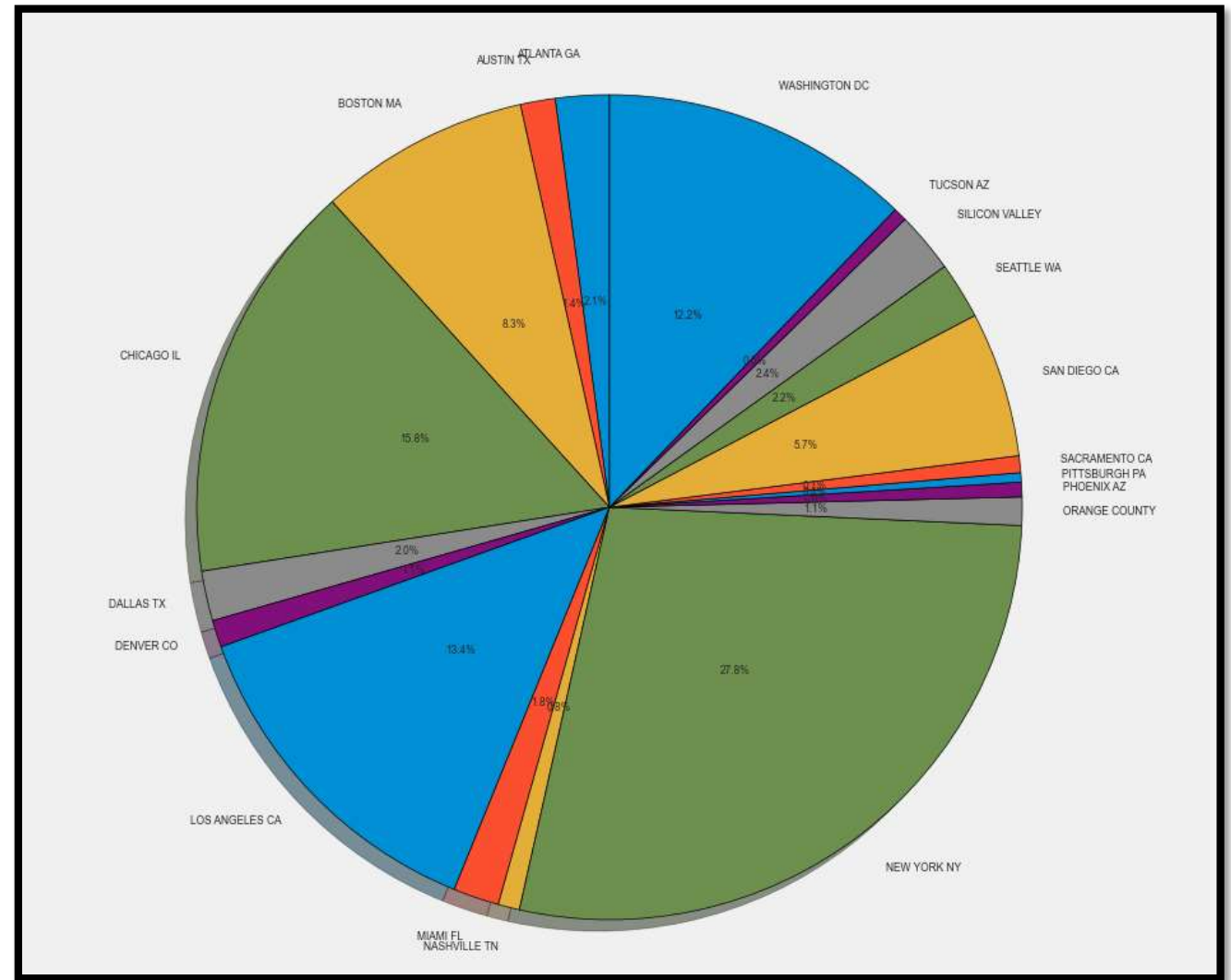
Gender Distribution

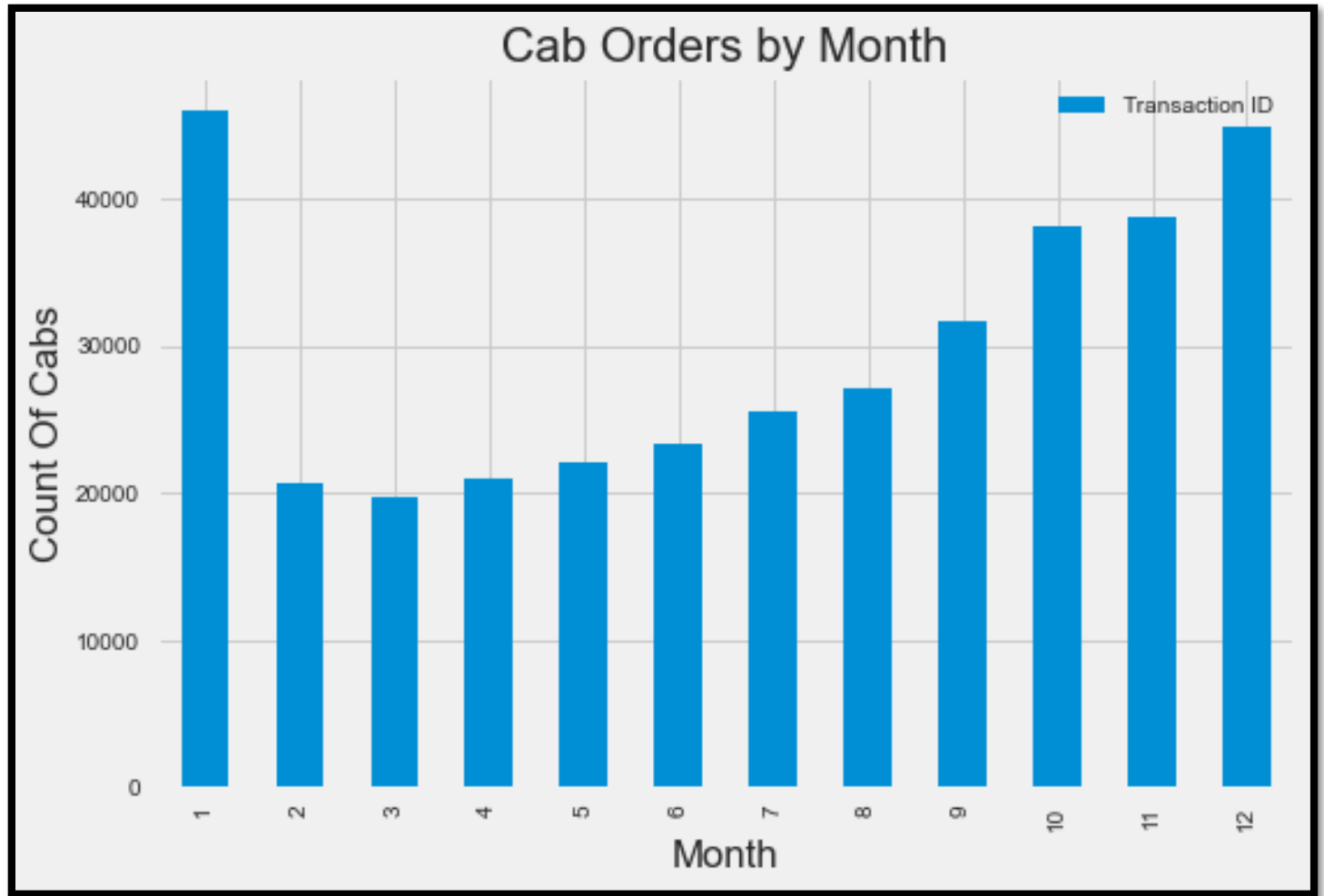This graph shows that the Male users prefer to travel by cab more.

Users Per City

This pie chart shows that New York City has the highest Cab users with 27% followed by Chicago with 16% and Los Angeles with 13%.
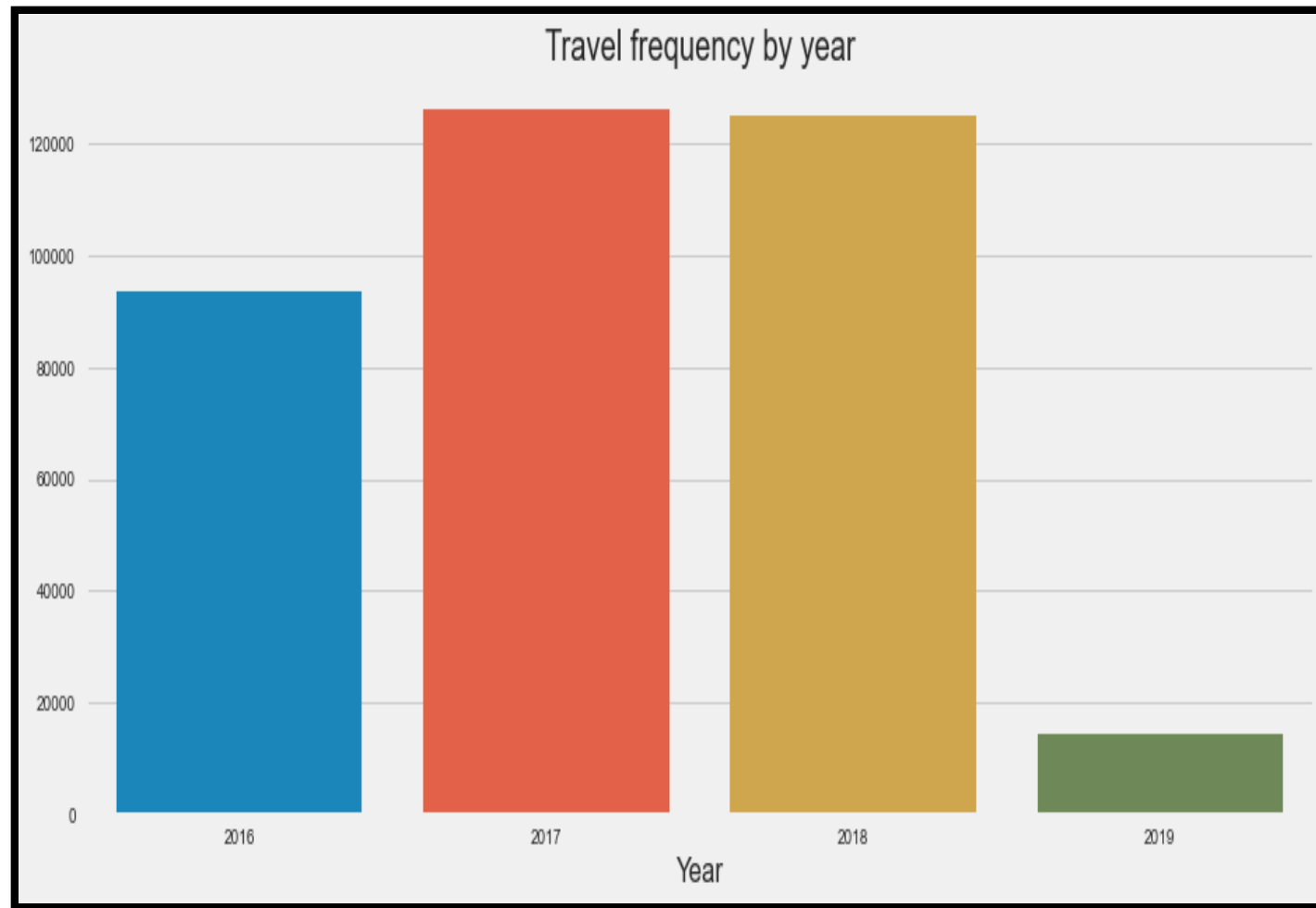
Cab Orders per Month

This graph shows that the Cab was ordered the most in the month of January and December

Travel frequency by year

This graph shows that the travel by cab was more in 2017 and 2018.
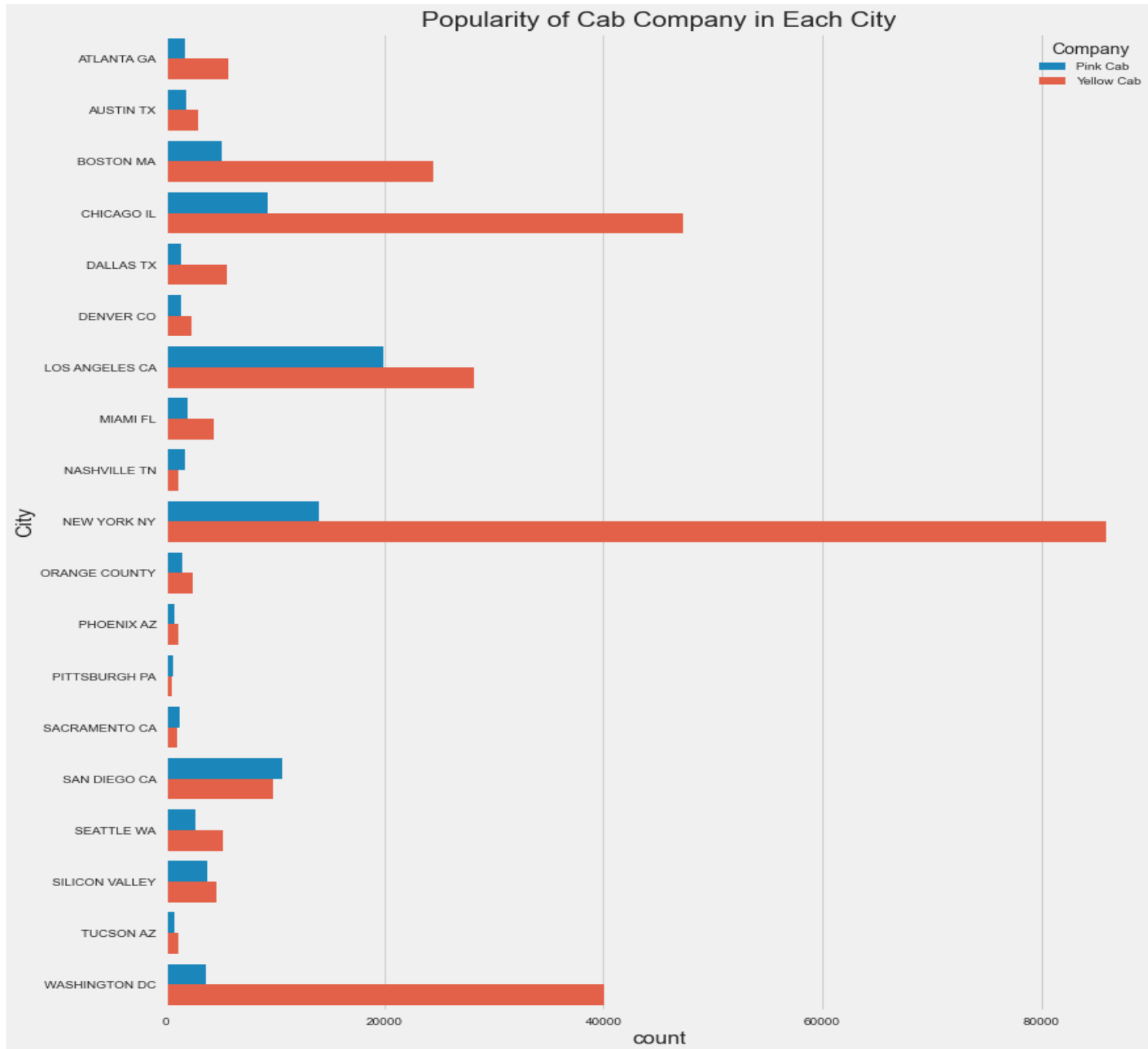
Popularity of Cab Company in Each City

This Graph shows that Yellow Cab is more popular 15 countries including Washington DC, New York NY, Chicago IL, Boston MA. However, Pink Cab is more popular in San Diego CA, Sacramento CA and Nashville TN.
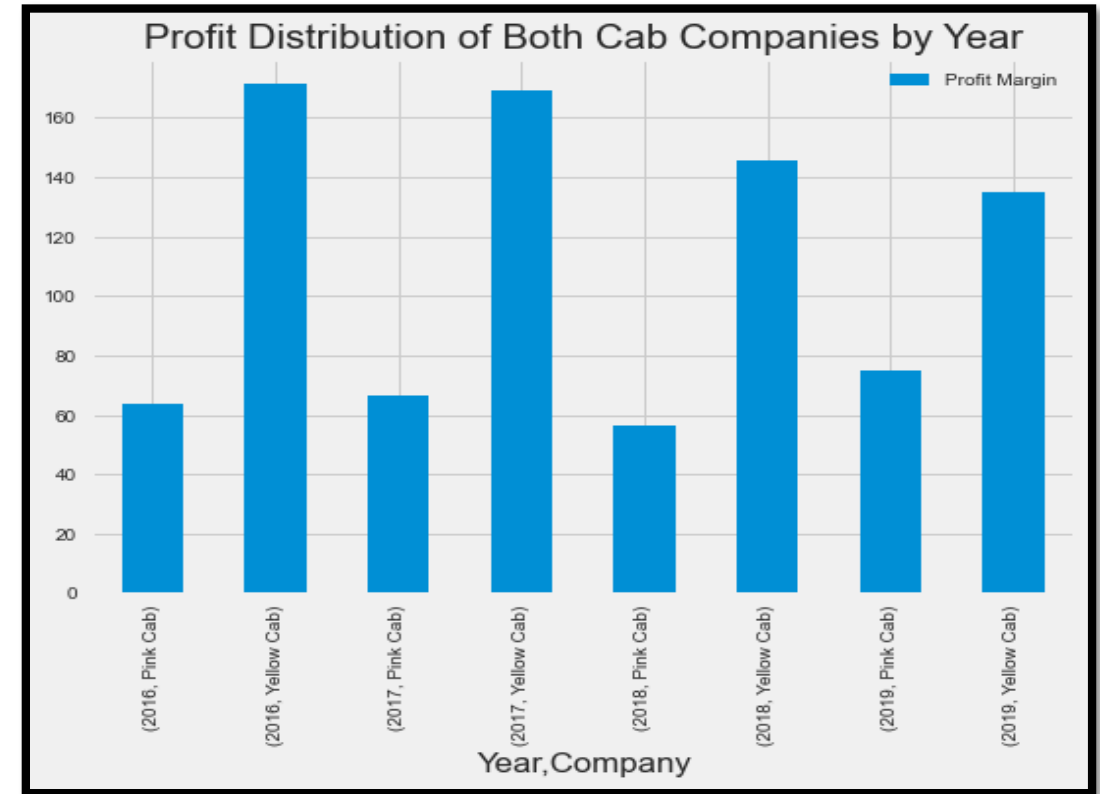
Gender Distribution of Cab Orders in Each City
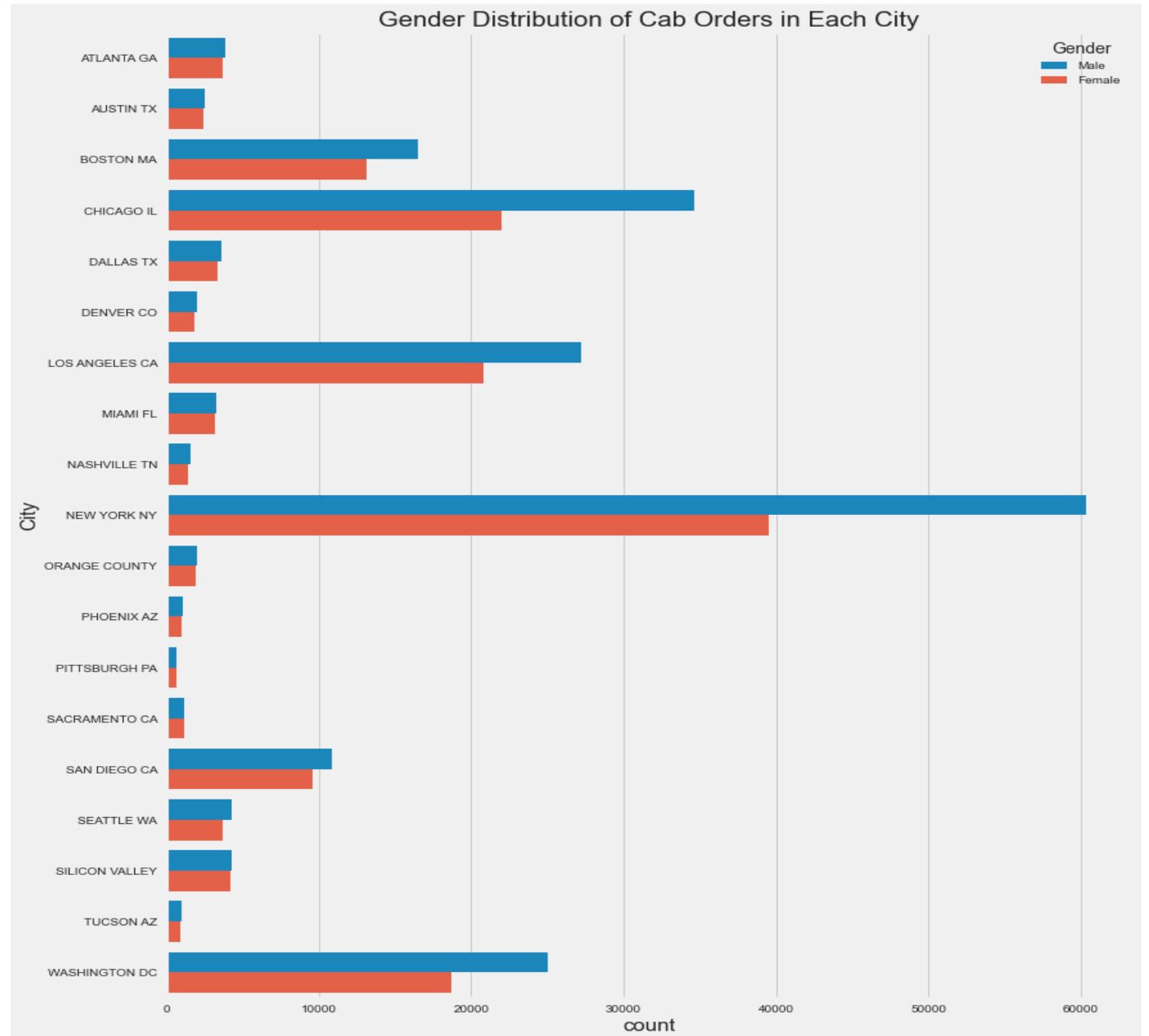
Profit Analysis

Profit Analysis
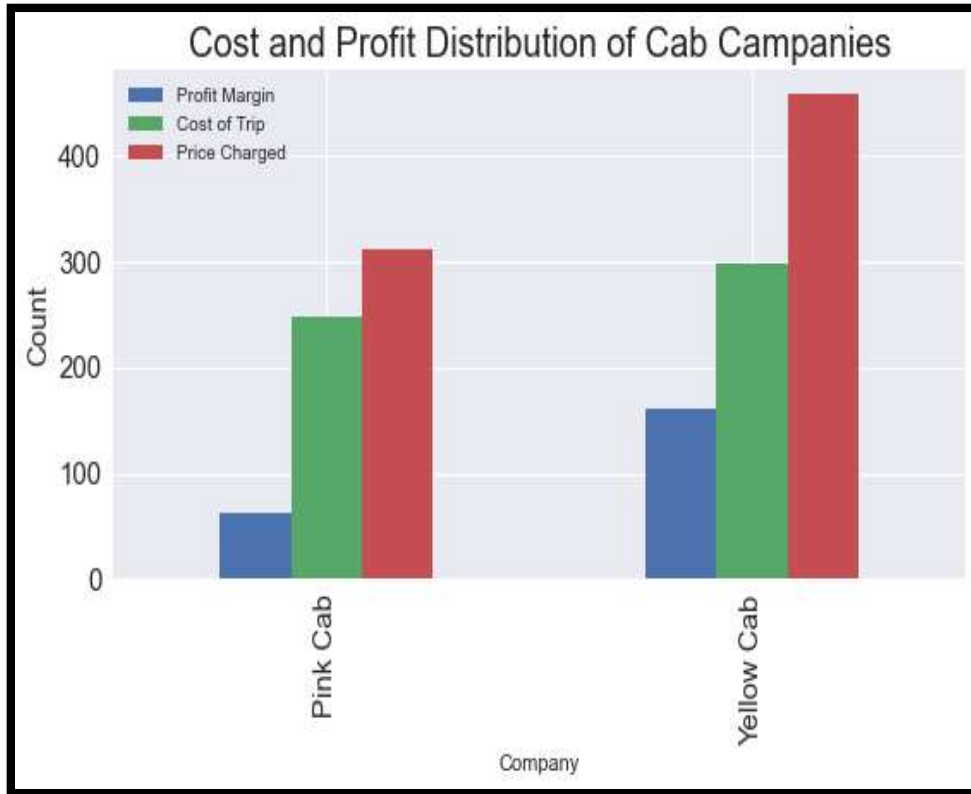
This graph shows the profit margin of Each Cab company per year. We can see that the profit margin of Yellow Cab is more than Pink Cab. However, Yellow Cab company profit was more in year 2016 and 2017 and it has a decreasing trend. Also, Note that Pink Cab company has an increasing profit growth trend.

Profit Analysis
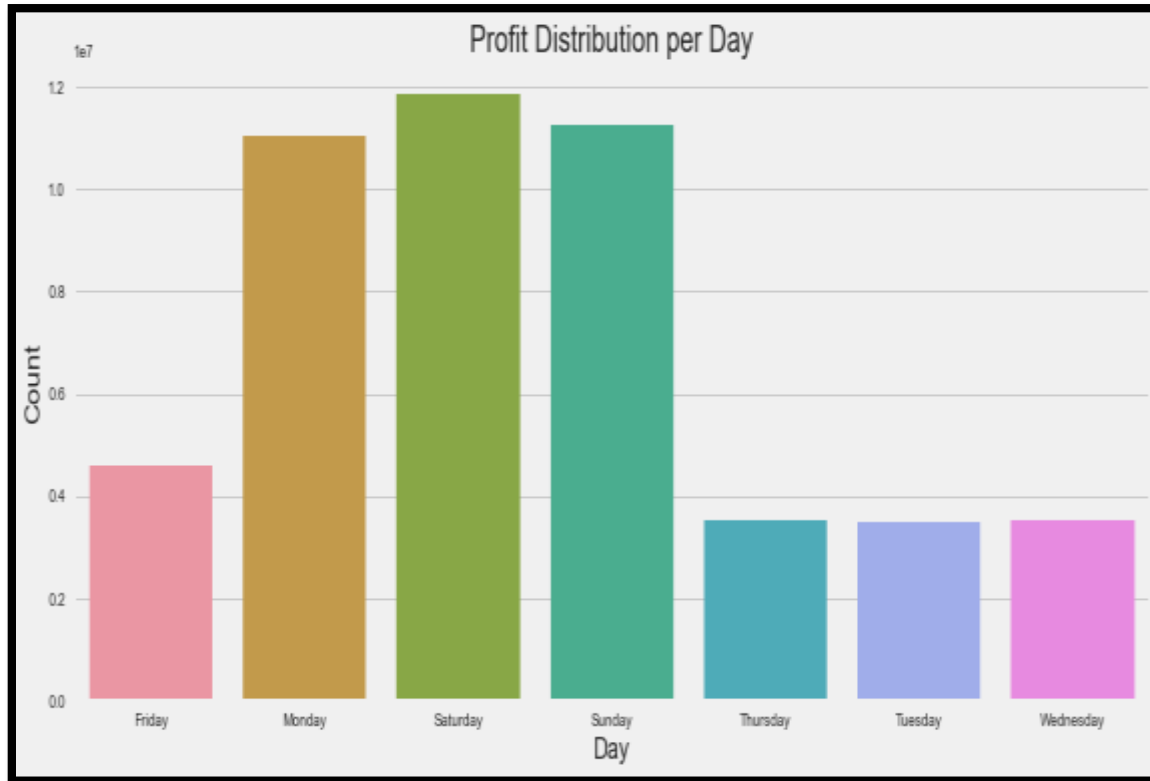
This graph shows that on Mondays, Saturdays and Sundays Cab Companies are making more profit.
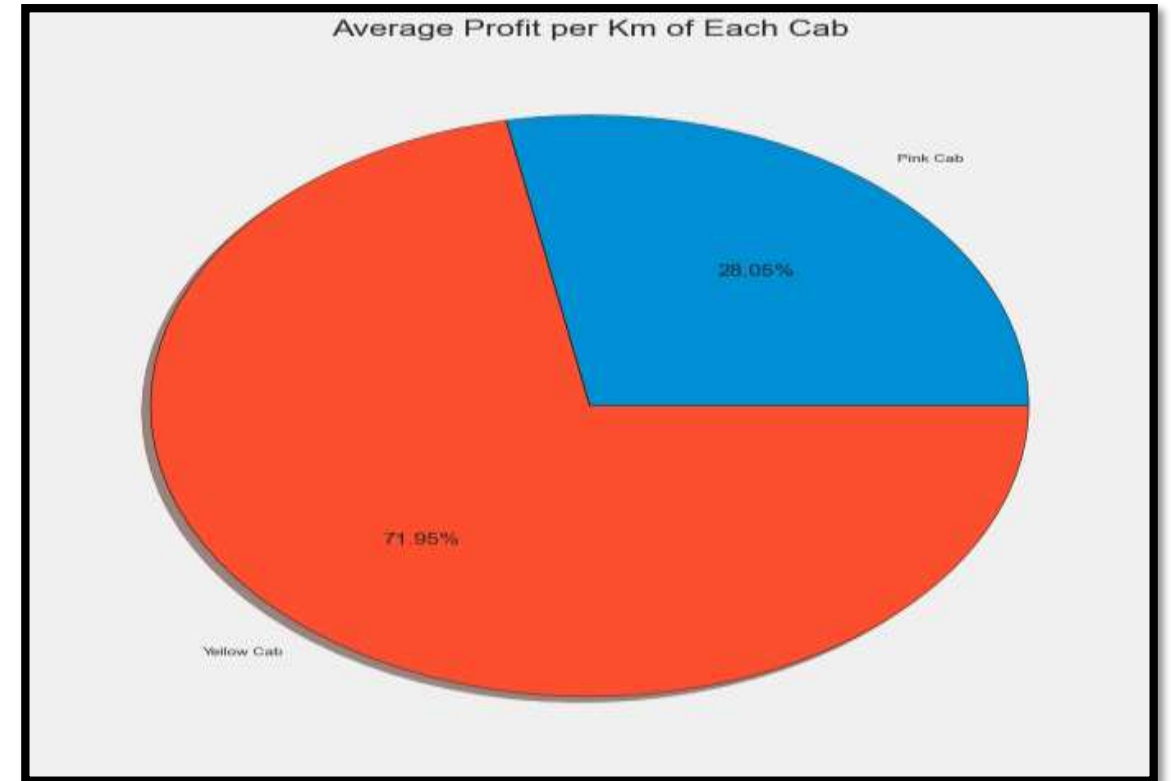
Yellow Cab's average profit per KM is almost three times the average profit per KM of the Pink cab.

The Results of the Above EDA are :

- Most Users prefer to use Yellow Cab Than Pink Cab
- Most Users prefer to pay by Card than Cash.
- Most of the Users travel in between 5 to 40 km.
- Most of the Users where in between age 20 to 40.
- Price charged by Yellow cab is comparatively more than Pink Cab.
- Male users prefer to travel by cab more.
- New York City has the highest Cab users with 27% followed by Chicago with 16% and Los Angeles with 13%.
- Cab was ordered the most in the month of January and December.
- Travel by Cab was more in 2017 and 2018.
- Yellow Cab is more popular 15 countries including Washington DC, New York NY, Chicago IL, Boston MA. However, Pink Cab is more popular in San Diego CA, Sacramento CA and Nashville TN.
- Yellow cab has customer in all age group, and it's been observed that it's even popular in 50+ age group as equally as 18-29 age group

Recommendation

We have evaluated both the cab companies on following points and found Yellow cab better than Pink cab:

Profit Analysis : The profit margin of Yellow Cab is more than Pink Cab.

Users Reach : Yellow Cab has more number of users.

Customer Preference : Most Users prefer to use Yellow Cab Than Pink Cab.

Age wise Reach : Yellow cab has customer in all age group, and it's been observed that it's even popular in 50+ age group as equally as 18-29 age group.

Average Profit per KM: Yellow Cab's average profit per KM is almost three times the average profit per KM of the Pink cab.

City wise Reach : Yellow Cab is more popular 15 countries including Washington DC, New York NY, Chicago IL, Boston MA.

Based on above point , we will recommend Yellow cab for investment.

Hypothesis Testing

Hypothesis 1:

Is there any difference in Profit with mode of Payment?

Defining Null Hypothesis

H0 : There is no significant difference regarding mode of payment in both cab companies.

Defining Alternate Hypothesis

H1 : There is significant difference regarding mode of payment in both cab companies.

Pink Cab

P value is 0.7900465828793288
We accept null hypothesis (H0) that there is no difference regarding mode of payment for Pink Cab

Yellow Cab

P value is 0.2933060638298729
We accept null hypothesis (H0) that there is no difference regarding mode of payment for Yellow Cab

We can now say that there is no difference in profit with mode of payment

Hypothesis 2:

Is there any difference in Profit with Weekend?

Defining Null Hypothesis

H0 : There is no significant difference regarding weekend on Profit, in both cab companies.

Defining Alternate Hypothesis

H1 : There is significant difference regarding weekend on Profit, in both cab companies.

Pink Cab

```
P value is  3.6685059417138927e-85
We reject null hypothesis or accept alternative hypothesis (H1) that there is a difference regarding weekend on Profit for Pink
Cab
```

Yellow Cab

```
P value is  6.841211801770341e-115
We reject null hypothesis or accept alternative hypothesis (H1) that there is a difference regarding weekend on Profit for Yell
ow Cab
```

We can now say that there is significant difference in profit with Weekend

Thank You