Case Study

Exploratory Data Analysis

Introduction:

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.

The analysis has been divided into Two parts:

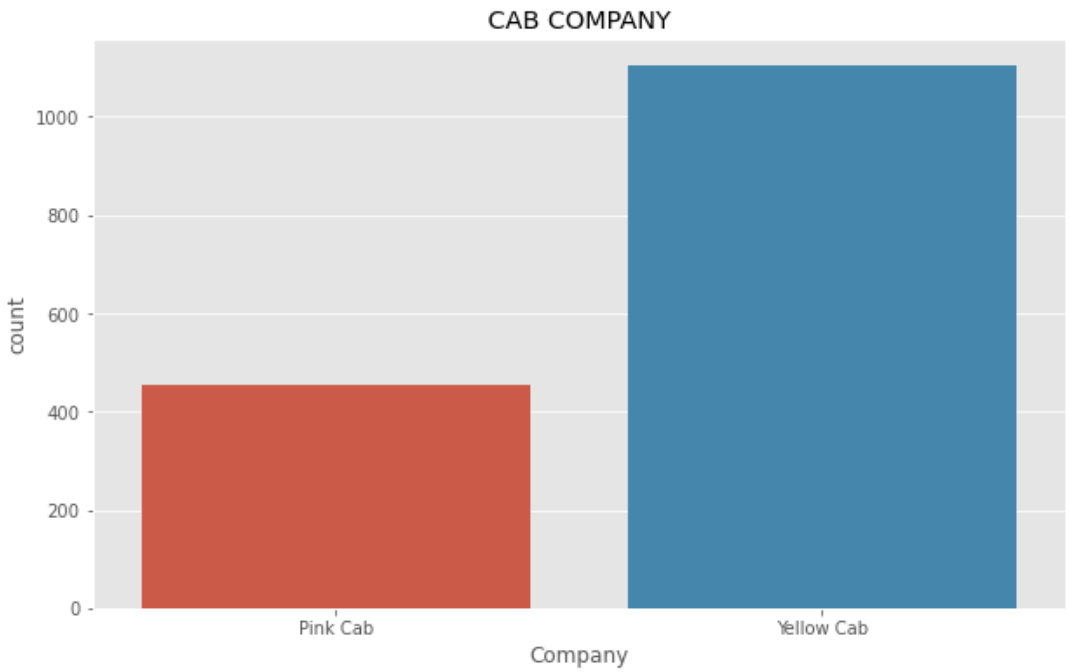
- Data Understanding & Most Profitable Cab Company.
- Hypothesis Testing
- Recommendations for investment

Data Understanding & Most Profitable Cab Company.

Data Exploration:

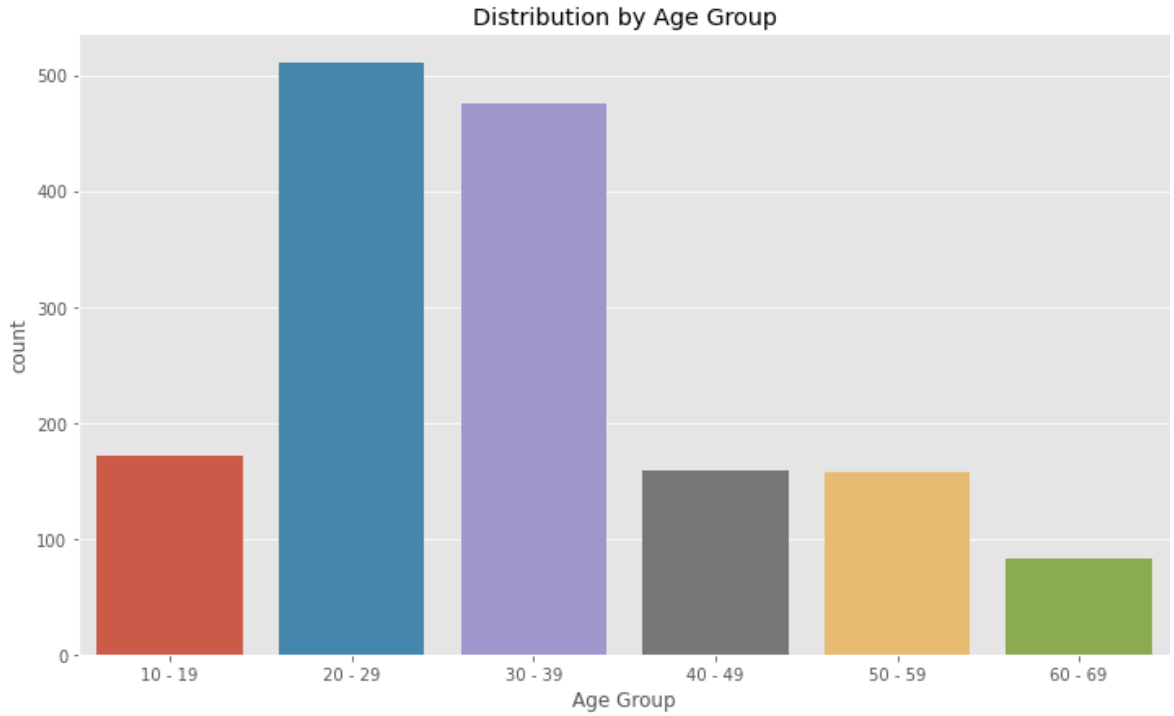
- **Cab_Data.csv** – this file includes details of transaction for 2 cab companies, shape (1998,7).
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details, shape (49171,4).
- **Transaction_ID.csv** – this is a mapping table that contains transaction to customer mapping and payment mode, shape (1999,3).
- **City.csv** – this file contains list of US cities, their population and number of cab users, shape (20,3)
- **Global_data** – This is Combination of all files. Shape (1558,12)

Cab Company Users:



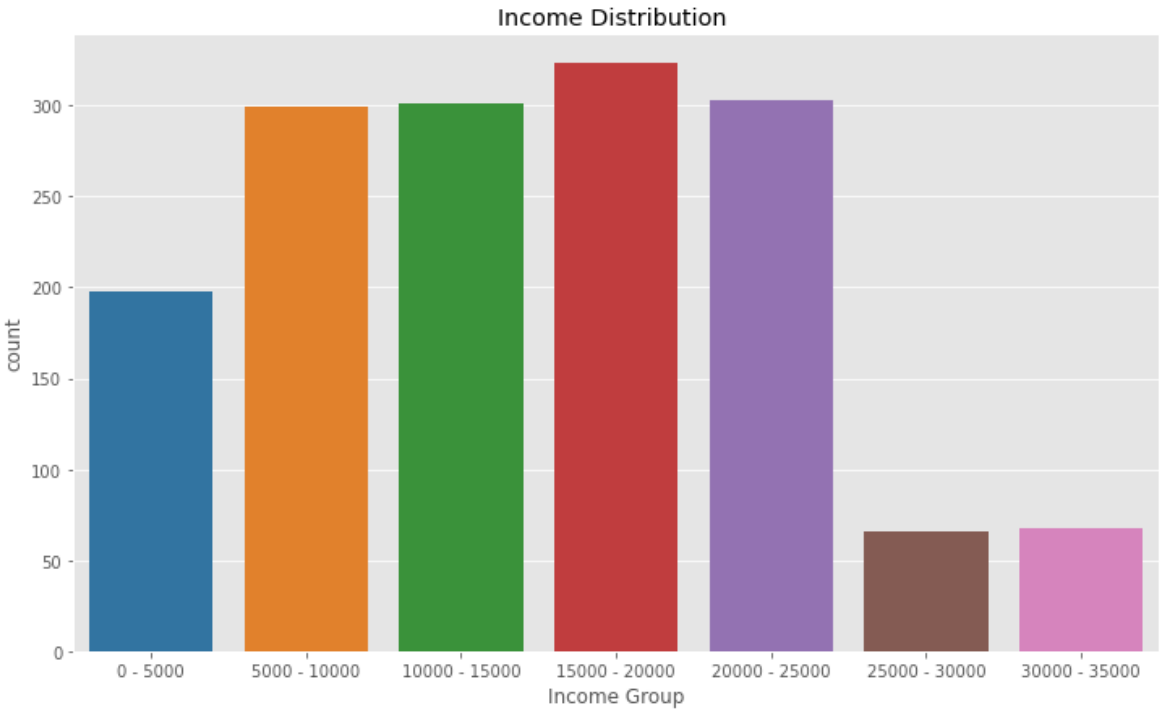
- Yellow Cab has had the majority of customers.

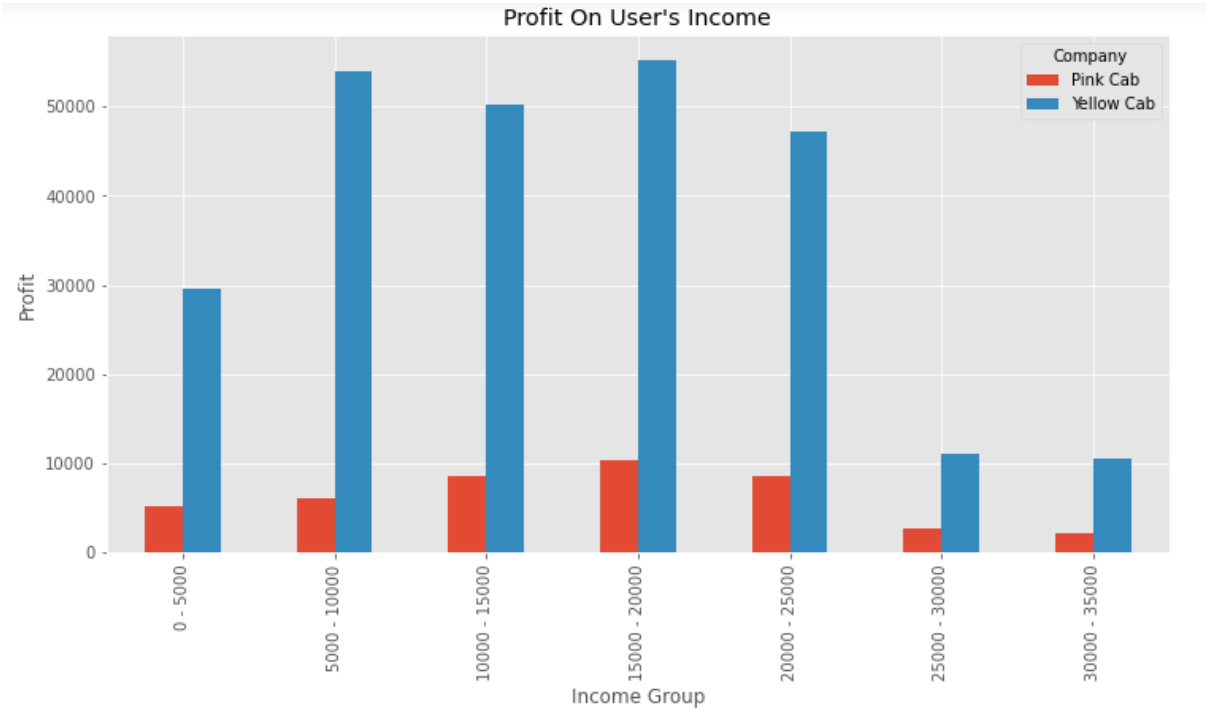
Age Group:



- 20-39 age group has the is most profit and most rides.
- Yellow Cab Company is making More Profit In each Age Group.

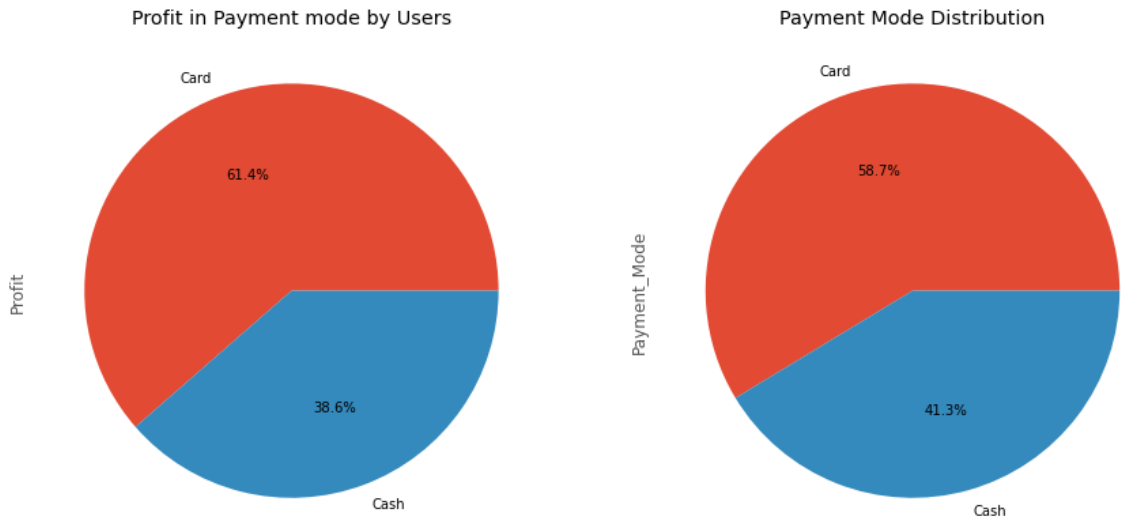
Income Group:





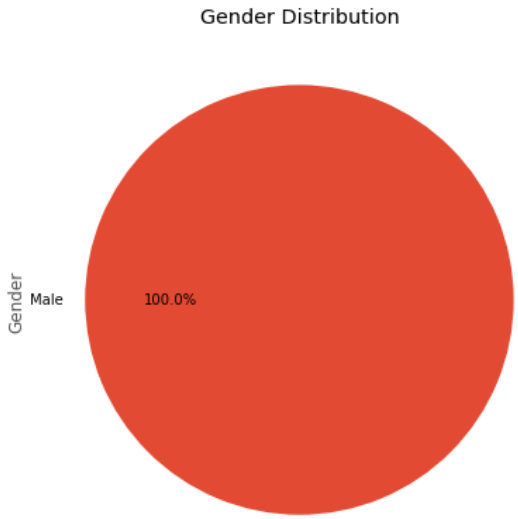
- User With Payment range 5000-25000 gives more profit & more rides.
- Yellow Cab Company is making More Profit In each Income Group

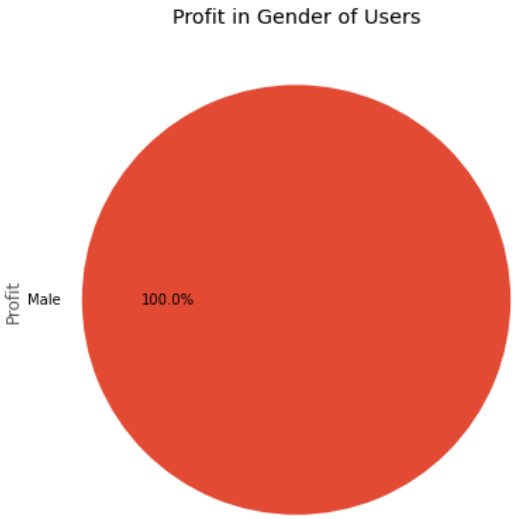
Payment Mode:



- 58.7% users prefers paying with Card which is 61.4% of Total Profit.
- 41.3% users prefers paying with Card which is 38.6% of Total Profit.

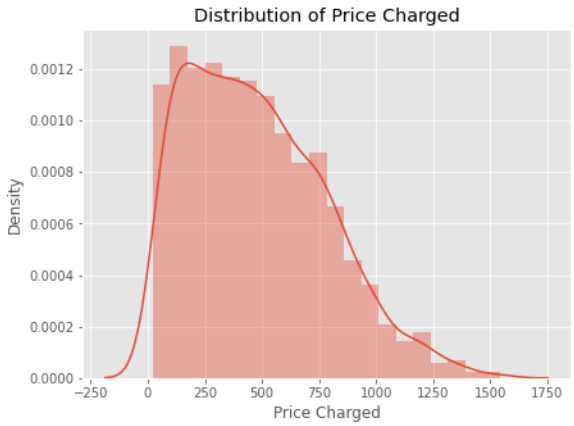
Gender:

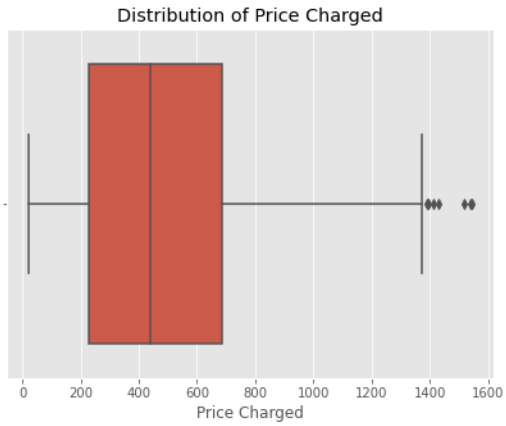


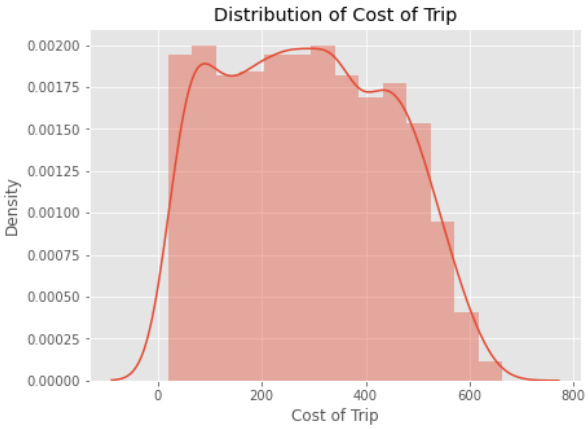


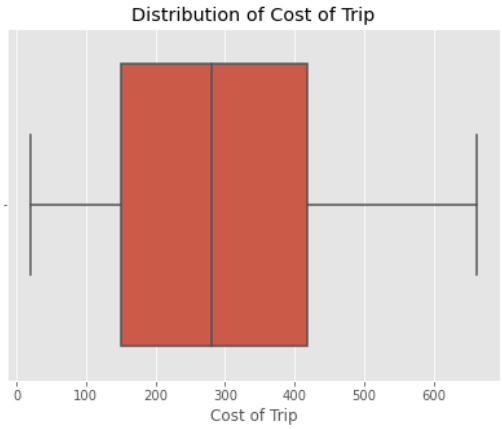
- All the users are Male.

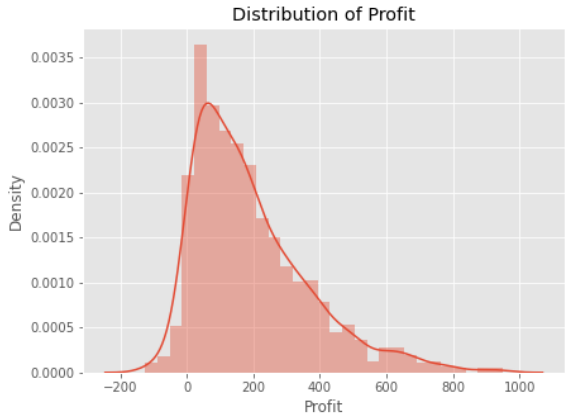
Prices In Global Data:

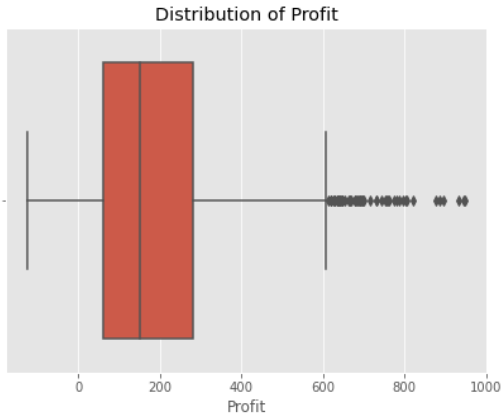






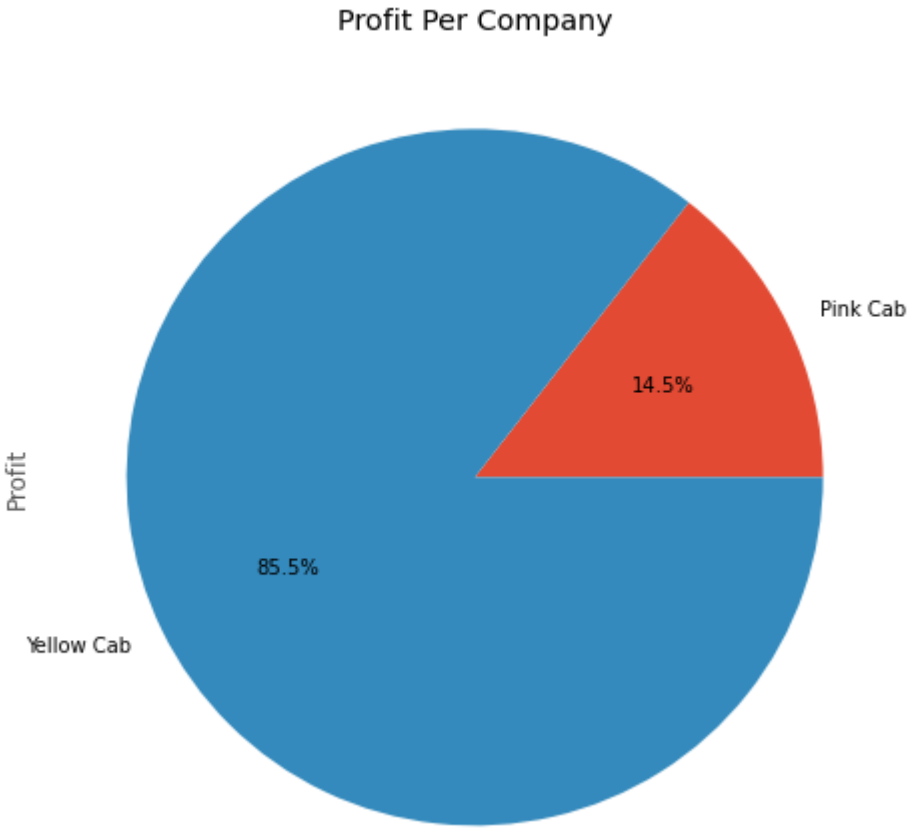






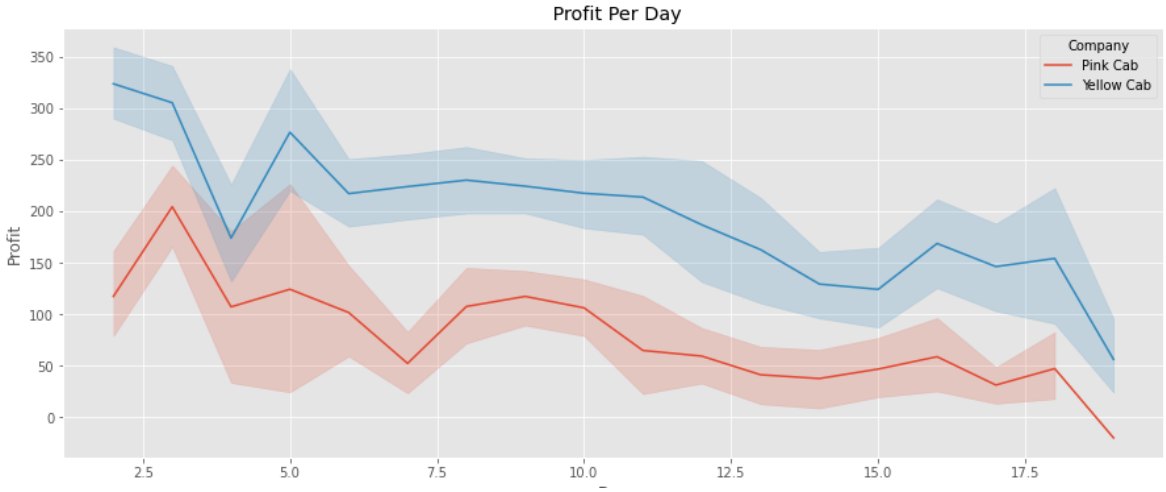
- This distributions show that there is a clear difference between the actual cost of trip and the priced charged by both companies.
- There is presence of outliers in the Profit and Price Charged.
- There is no Outliers Presence in Cost of Trip.

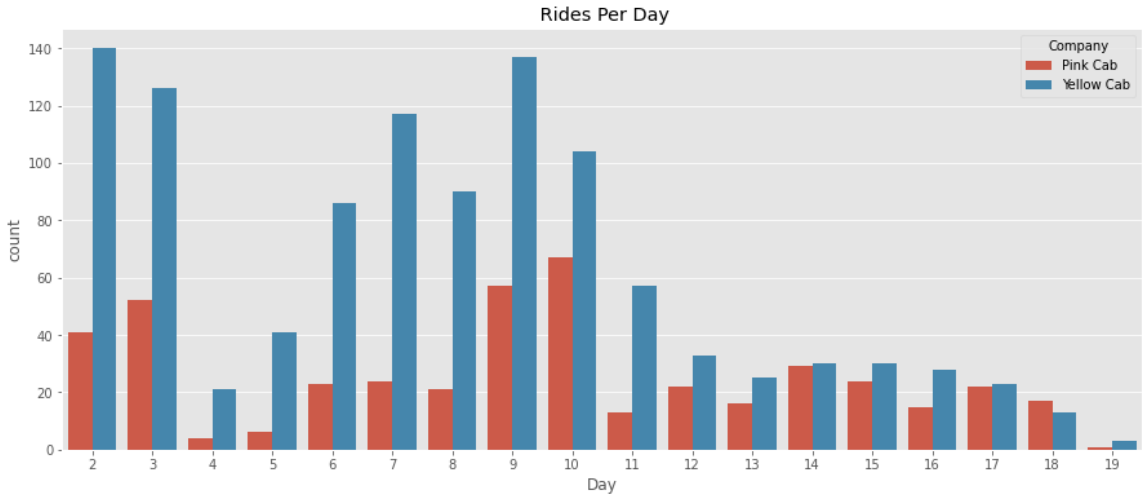
Profit per Company:



- Yellow cab Company made 85.5% of the total profit

Profit & Rides Per Day:



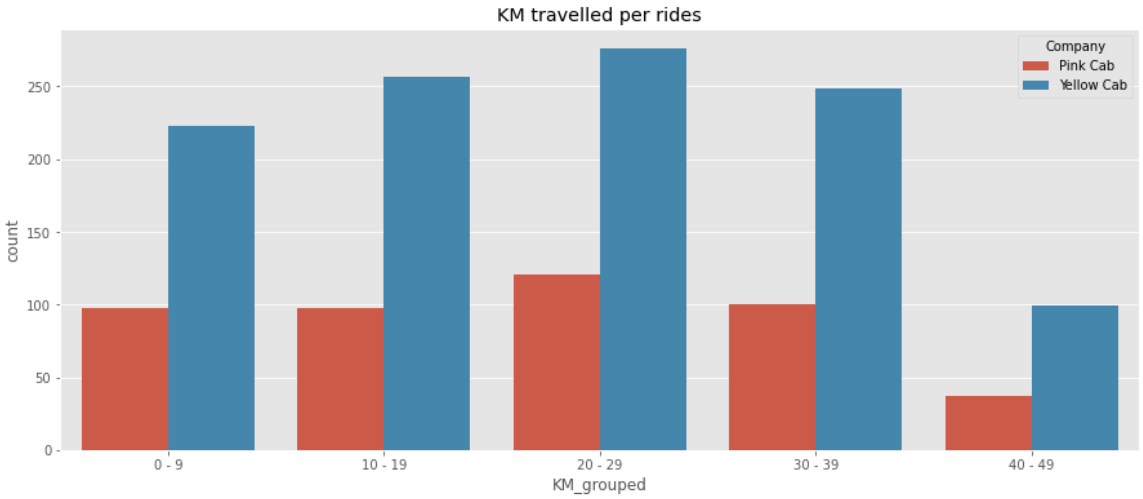


- Yellow Company is taking more rides and making more profit.

Profit Per KM Travelled:

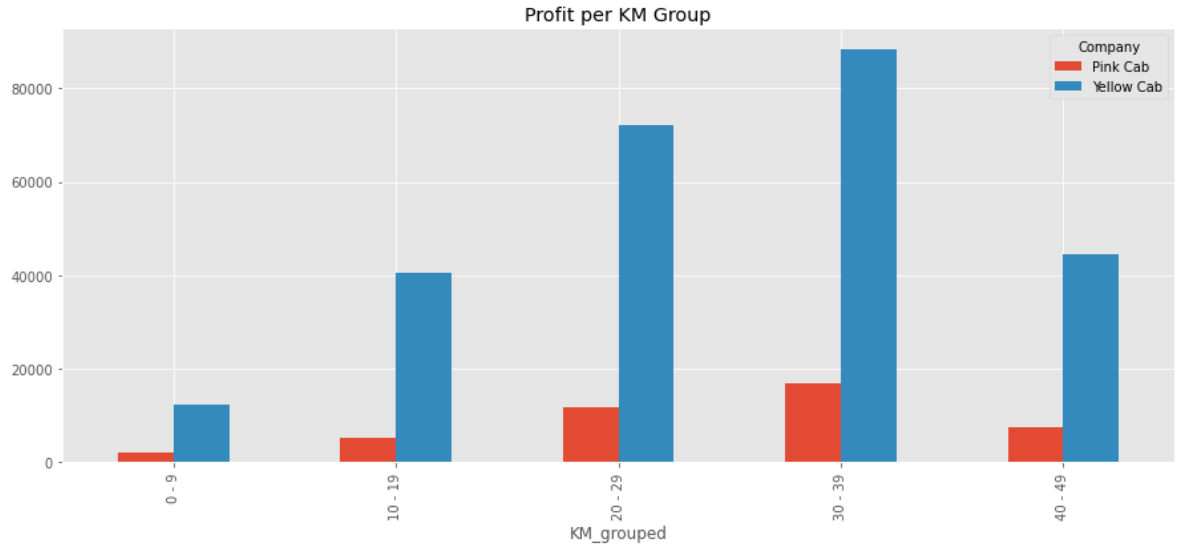
- The longer the KM travelled the higher the price charges from both companies however the Yellow cab company charged higher than the Pink cab company.

KM Travelled per Rides:



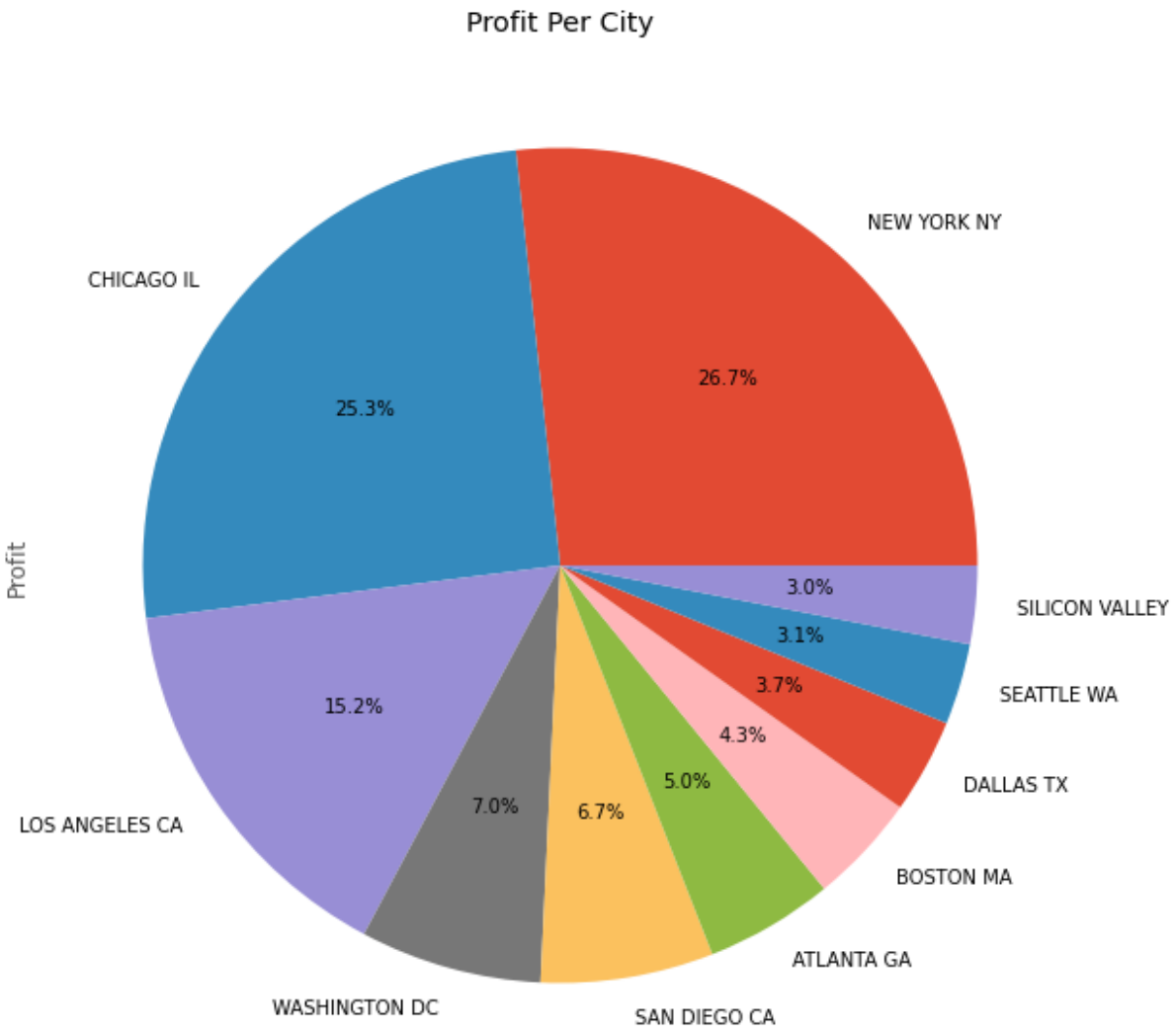
- Most users take a trip of 10 to 39 km

Profit per KM Group:



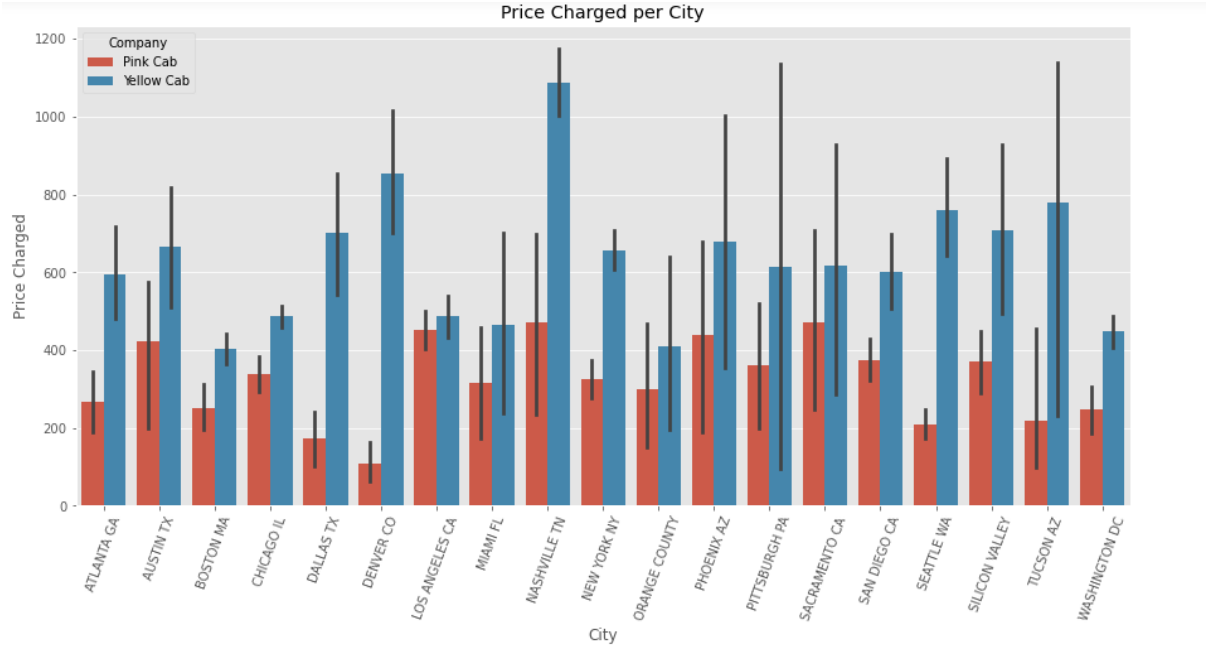
- This chart represents the number of Km travelled registered from Yellow cab and Pink cab company. Most users took a ride of 10 to 30 KM and Most users choose Yellow cab, from where the yellow cab made more profit compared to the other.

Profit Per City:



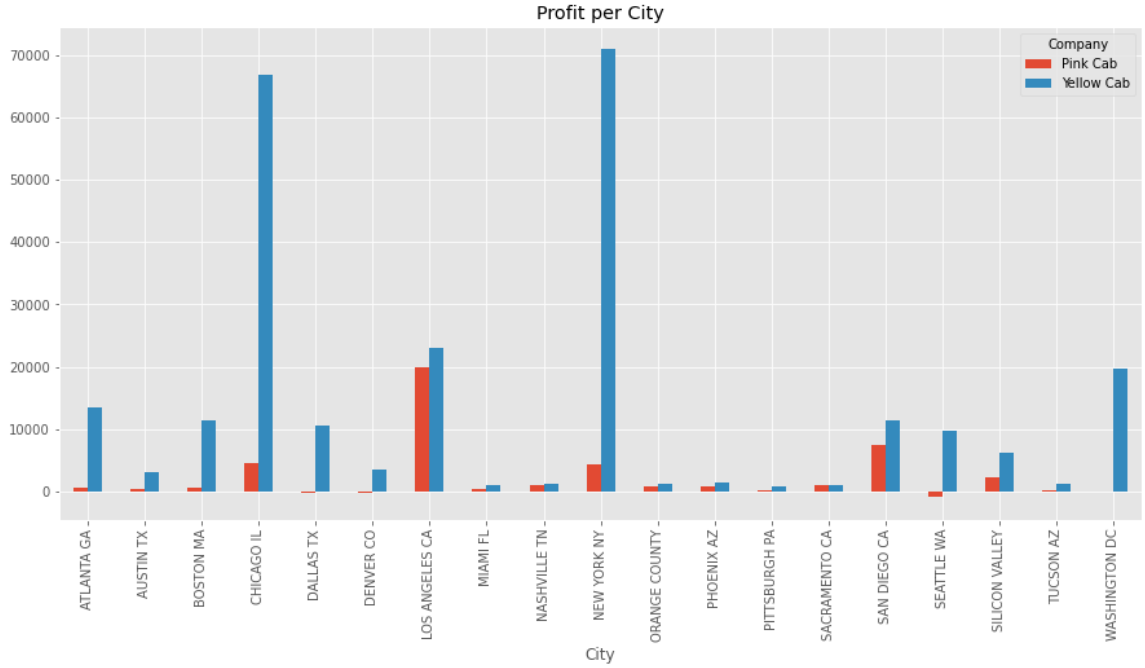
- New York City has 26.7% profit, followed by Chicago 25.3%, Los Angeles 15.2%.
- These 3 cities contribute more than 65% of profit.

Price Charged per City:



- Yellow Cab company charges higher than the Pink Cab company in all cities, pink Cab profit looks constant and ranges from 110 to 500 while The yellow cab charges higher in populated city in terms of users and has an instant profit that ranges wider.

Profit Per City:



- Cities with more Population have more profit.
- Yellow cab company made more profit in every city.

THE PINK CAB COMPANY DATA:

- The Revenue of Pink Cab Company is: 43626.5
- The Gross Profit Margin of Pink Cab Company is: 28.06
- The Profit per Ride of Pink Cab Company is: 96.09

THE YELLOW CAB COMPANY DATA:

- The Revenue of Yellow Cab Company is: 257873.42
- The Gross Profit Margin of Pink Cab Company is: 43.75
- The Profit per Ride of Pink Cab Company is: 233.58

Global Data:

- The Revenue of Yellow Cab Company is: 301499.92
- The Gross Profit Margin of Pink Cab Company is: 40.47
- The Profit per Ride of Pink Cab Company is: 193.52

	Company	Revenue	Profit Margin [%]	Rides	Profit per Ride [AVG]
0	Pink Cab Company	43626.50	28.06	454	96.09
1	Yellow Cab Company	257873.42	43.75	1104	233.58
2	Global Data	301499.92	40.47	1558	193.52

Hypothesis Testing

Areas to investigate:

• Which company has maximum cab users at a particular time period?

• Does margin proportionally increase with increase in number of customers?

• What are the attributes of these customer segments?

1.Maximum Cab Users

Which company has maximum cab users at a particular time period?

We have the data between 2 Jan to 19 Jan, so we'll check for the max Cab users in this time/date period

we'll check whether Yellow cab users are the maximun cab users in our global data

We'll make our hypothesis as bolow:

Null Hypothesis [H0] - "Yellow Cab Users Are Maximum in Global Data in Date Period"

Alternate Hypothesis [H1] - "Yellow Cab Users Are Not Maximum in Global Data in Date Period"

The P value with Pink Cab Users is: 0.0002712033246883 31

We accept Alternate Hypothesis that there is a statistical difference
Pink Cab Users Are Not Maximum in Global Data in Date Period

The P value with Yellow Cab Users is: 0.17425882747731186

We accept Null Hypothesis that there is no statistical difference
Yellow Cab Users Are Maximum in Global Data in Date/Time Period
---

So, We Have Statistically Prove that Yellow Cab Company Have the Maximum Users in the Date/Time Period.

---

2. Margin Proportionality with Increasing Customers

Does margin proportionally increase with increase in number of customers?

We've unique Transaction Id so we'll check profit with respect to them.

we'll check whether Profit Margin Increase with Increase in No of Customers in Yellow and Pink Cab Company

We'll make our hypothesis as below:

Null Hypothesis [H0] - "Profit Margin Increases with Increase in No of Customers in Yellow and Pink Cab Company"

Alternate Hypothesis [H1] - "Profit Margin Does not Increases with Increase in No of Customers in Yellow and Pink Cab Company"

Checking With All Data:

The P value with All: 1.2865793052940668e-45

We accept Alternate Hypothesis that there is a statistical difference
Profit Margin Does Not Increases with Increase in No of Customers.

Checking with Global & Pink Company Data:

The P value with Global & Pink Company Data: 2.395879929877656e-28

We accept Alternate Hypothesis that there is a statistical difference
Profit Margin Does Not Increases with Increase in No of Customers in Pink Cab Com
pany.

Checking with Global & Yellow Company Data:

The P value with Pink & Yellow Company Data: 1.8908630848986554e-48

We accept Alternate Hypothesis that there is a statistical difference Profit Margin Does Not Increases with Increase in No of Customers in Pink/Yellow Cab Company.

	Company	Revenue	Profit Margin [%]	Rides	Profit per Ride [AVG]
0	Pink Cab Company	43626.50	28.06	454	96.09
1	Yellow Cab Company	257873.42	43.75	1104	233.58
2	Global Data	301499.92	40.47	1558	193.52

3.Attributes of Customer Segments

What are the attributes of these customer segments?

Attributes of Custoners Segments:

1. Gender: We Only Have Transactions of Male Users so We can neglect this feature.

2. Age Group: This is one of the Important Feature of customers. We have created it from Age column.

3. Income Group: This is also on of the Important Feature of Customers. We have created it from Income Column

We'll Check whether this features are attributes of Customer Segments are not.

We'll make our hypothesis as bolow:

Null Hypothesis [H0] - "Attribute is Part of Customer Segment"

Alternate Hypothesis [H1] - "Attribute is Not Part of Customer Segment"

Age Group:

The P value with Profit Margin by Age Group: 5.72245598308611 36e-08

We accept Null Hypothesis that there is no statistical difference
Attribute Age Group is Part of Customer Segment

Income Group:

The P value with Profit Margin by Income Group: 1.063098937942807e-12

We accept Null Hypothesis that there is no statistical difference
Attribute Income Group is Part of Customer Segment

So, The Customer Segment Attributes are Age Group or Age and Income Group or Income

Recommendations and Conclusion:

I believe that Yellow cab going to maintain its monopoly in the market for a longer period of time because:

1. **Customer Retention:** Yellow cab has a very high customer retainability rate compared to Pink Cab.

2. **Customer Reach:** Yellow cab has had the most customers and is not showing any signs of letting go off this trend.

3. **Company Availability:** Yellow cab a high customer following in high population cities like New York, Chicago, Los Angeles.

4. **Age wise Reach:** Most of the young aged customers prefer Yellow Cab. It is to be noted that the young aged customers are the biggest consumers in this cab business.

5. **Company Profitability:** Yellow cab seems to be making a higher profit margin compared to Pink cab Company.

6. **Income wise Reach:** Both the cabs are very popular in all Income Groups, but here also Yellow cab is performing better than Pink cab in offering their services to any class.

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

We will advise the XYZ company to invest in Yellow Cab company for its glorious benefit.