Data Glacier

Your Deep Learning Partner

# G2M Case Study

## Virtual Internship

14-Sep-2024

# Background –G2M(cab industry) case study

- 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.

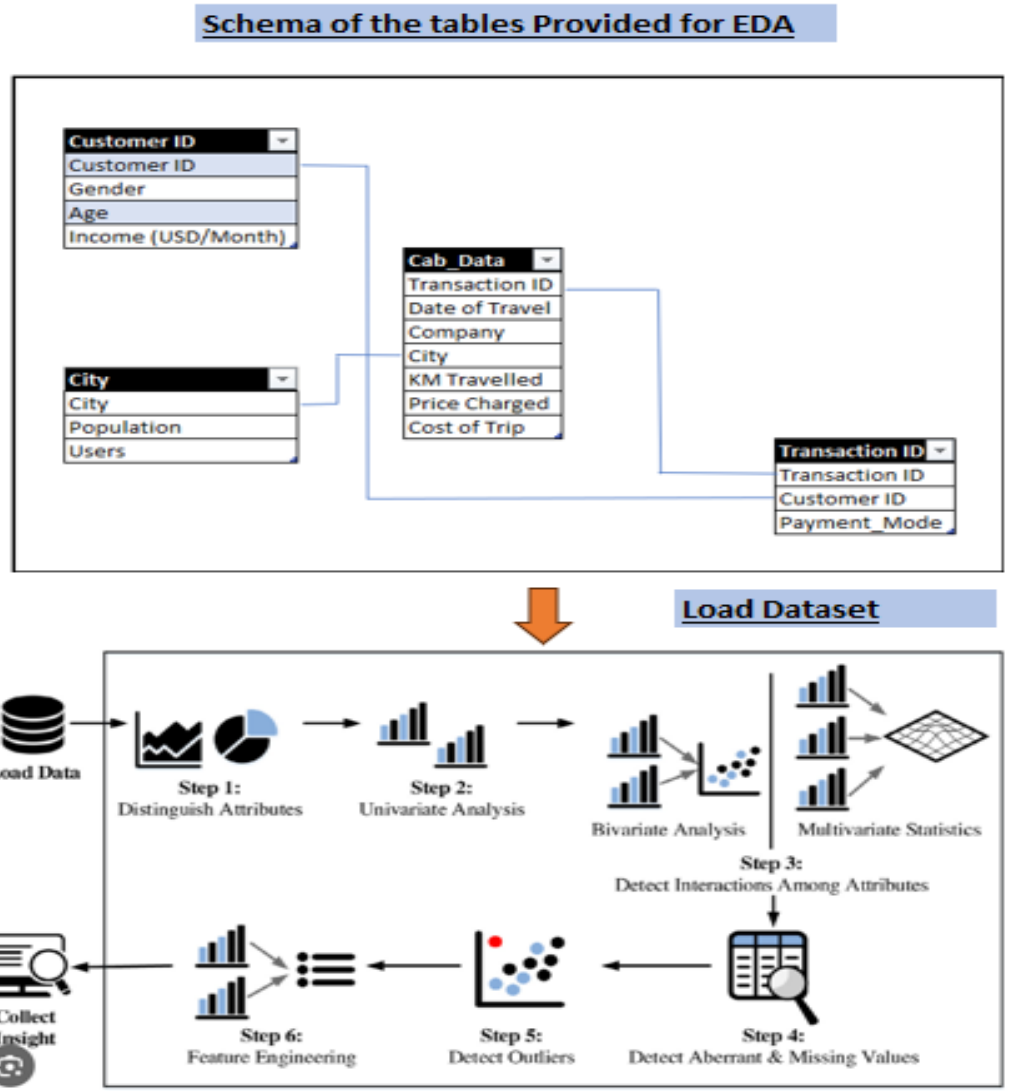
The analysis has been divided into four parts:

- Data Understanding
- Forecasting profit and number of rides for each cab type
- Finding the most profitable Cab company
- Recommendations for investment

# Data Exploration

**Initial Observations:**

- minimum year is in the year 2016 and Date of Travel is in the year 2019
- **"*Price charged*"** column has the highest variance amongst all the columns. Hence may be the people with different income groups travelling
- ***"Km* Travelled"** people boarding the cab have travelled a maximum of 48 miles indicating that the cabs are not used for outstation travel
- Total data points :355,032
- There are no null values in the dataset.
- There are 24 variables in the dataset



# Data Exploration: *Outlier Detection*

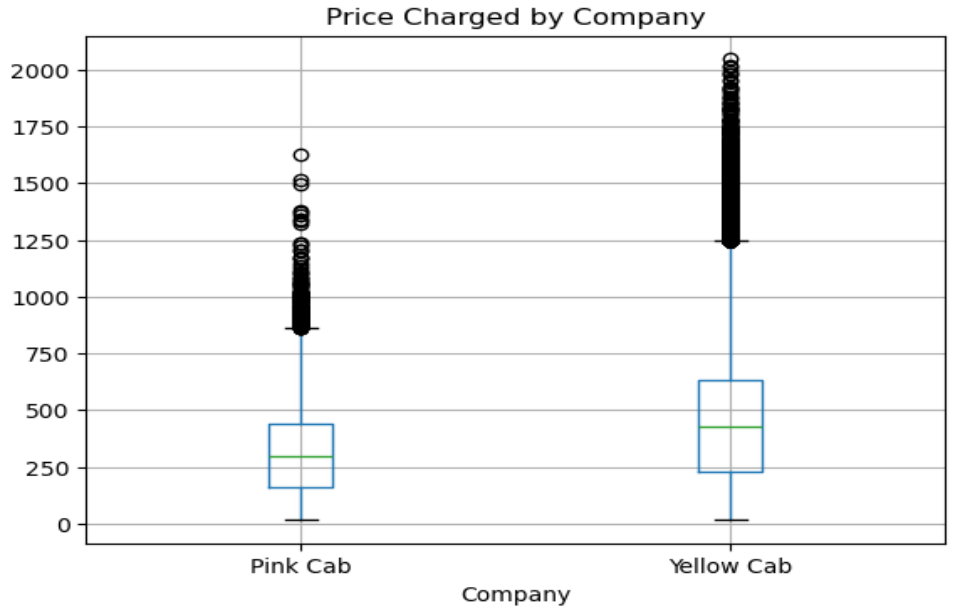
**Initial Observations:**

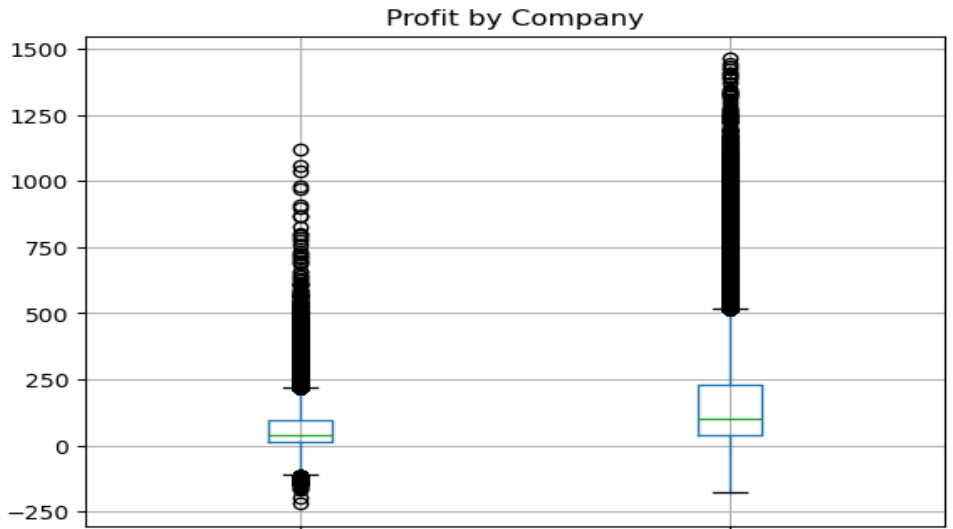
***Price charged variable-***

- *Yellow cabs* has a very price rates in comparison to *Pink cab*.
- The data for yellow cab are more spread in comparison to pink cabs and hence the business will cover more customers then pink cab does.

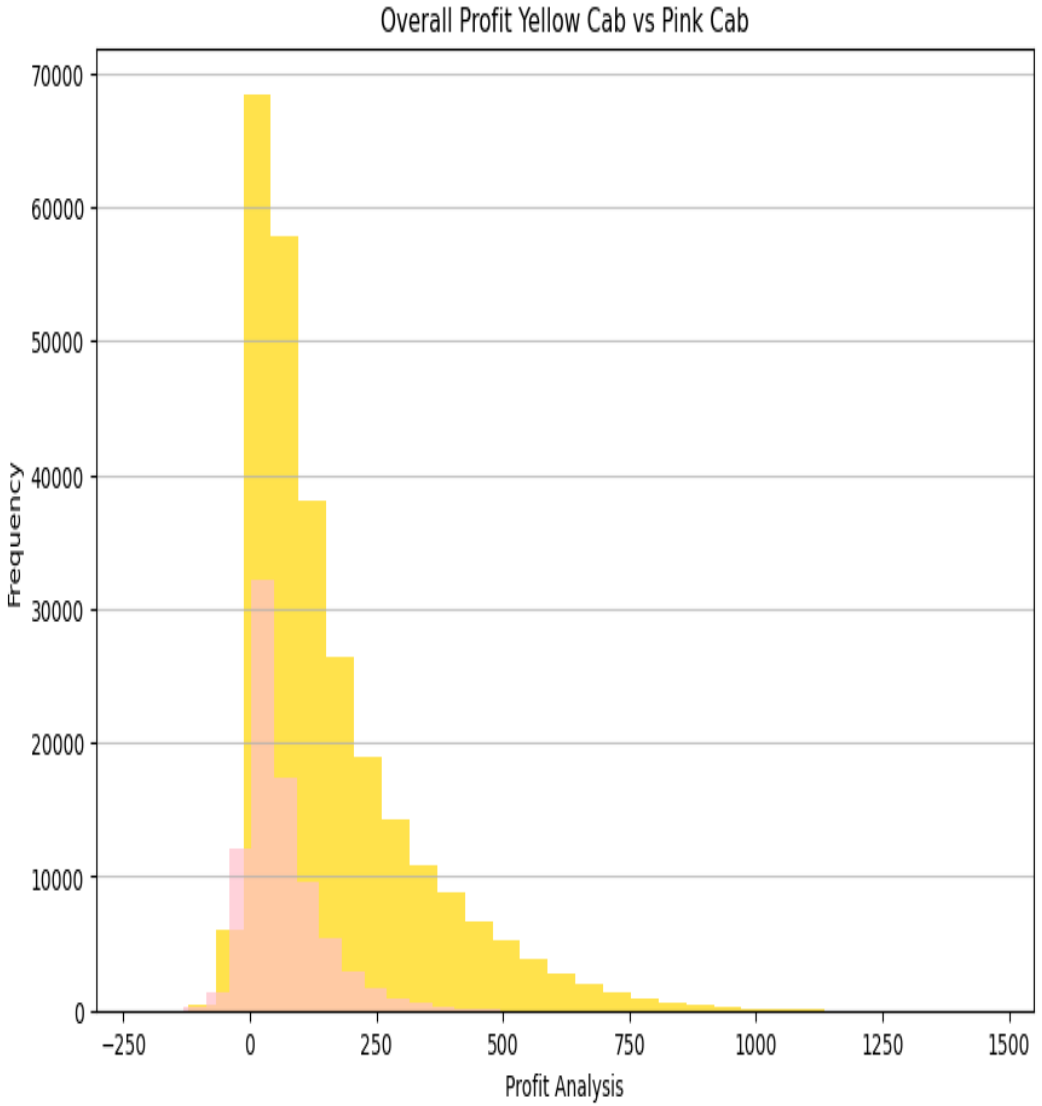
***Profit variable-***

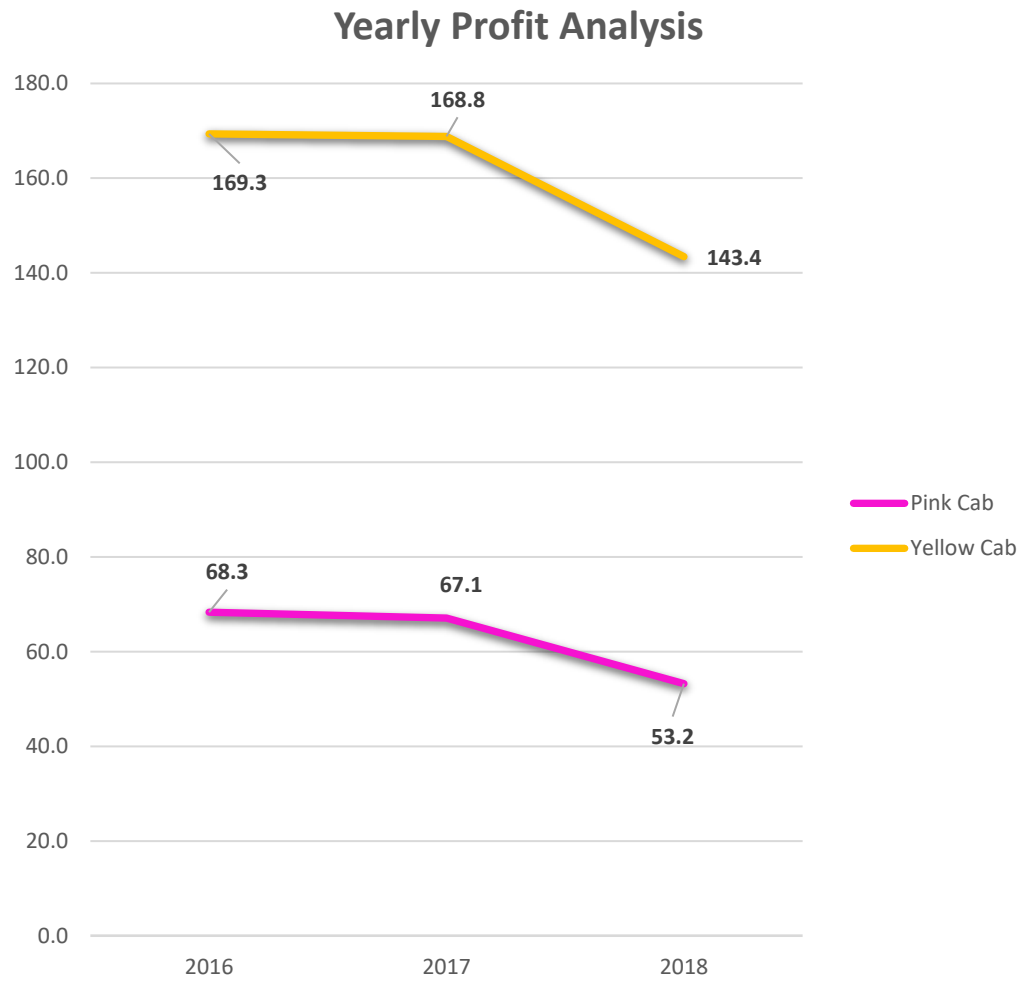
- *Yellow cabs* has a very high Profit margin in comparison to *Pink cab*
- The outliers in yellow cab lies more towards the upper limit then lower limit indicating that far more affluent customers travel in yellow cab then Pink cab
- To understand the Company's profitability, Customers Age and Socio-economic analysis needs to be performed



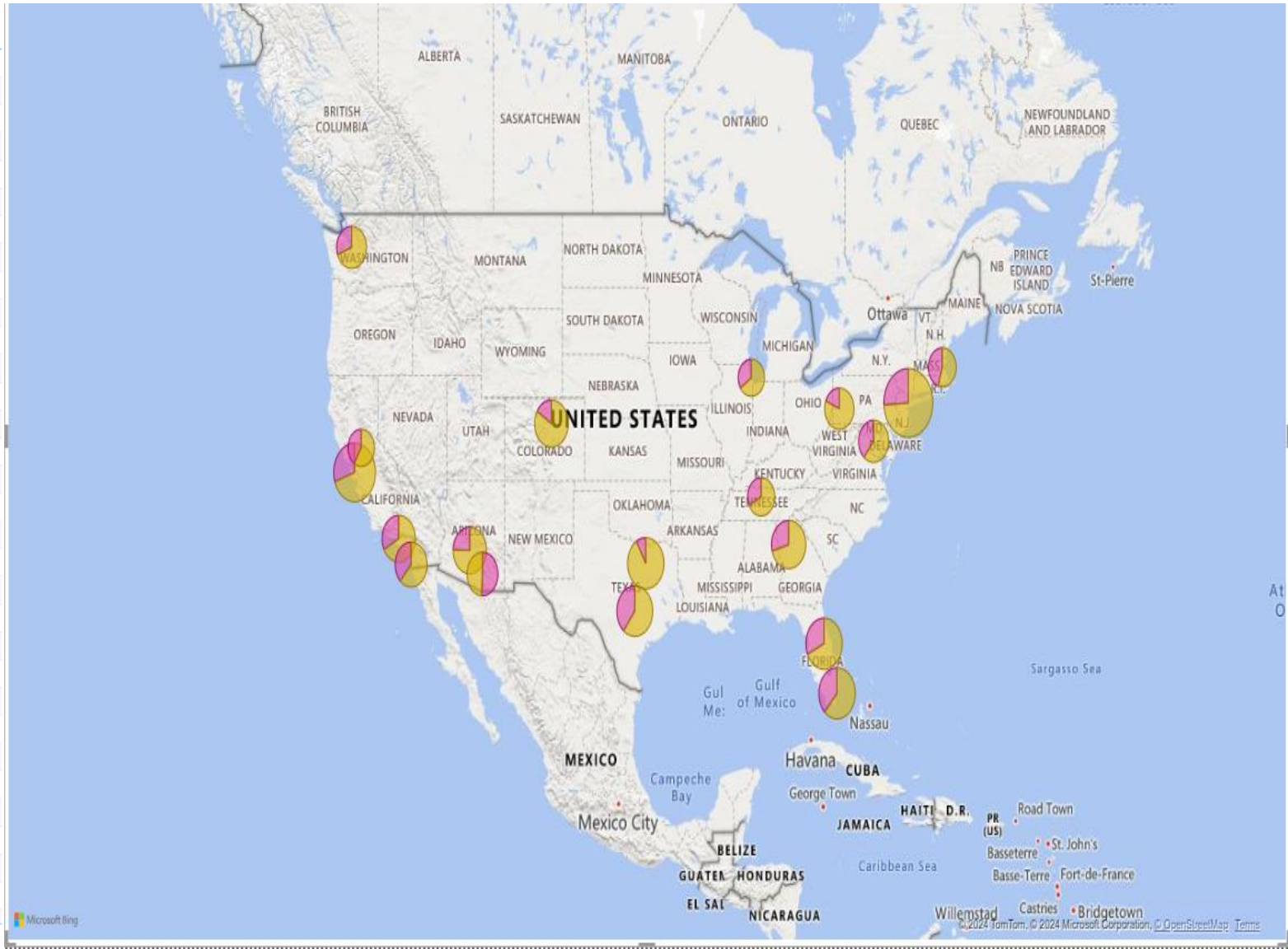
# Profit Analysis: *Overall & Yearly wise*



# Profit Analysis : *Geography*

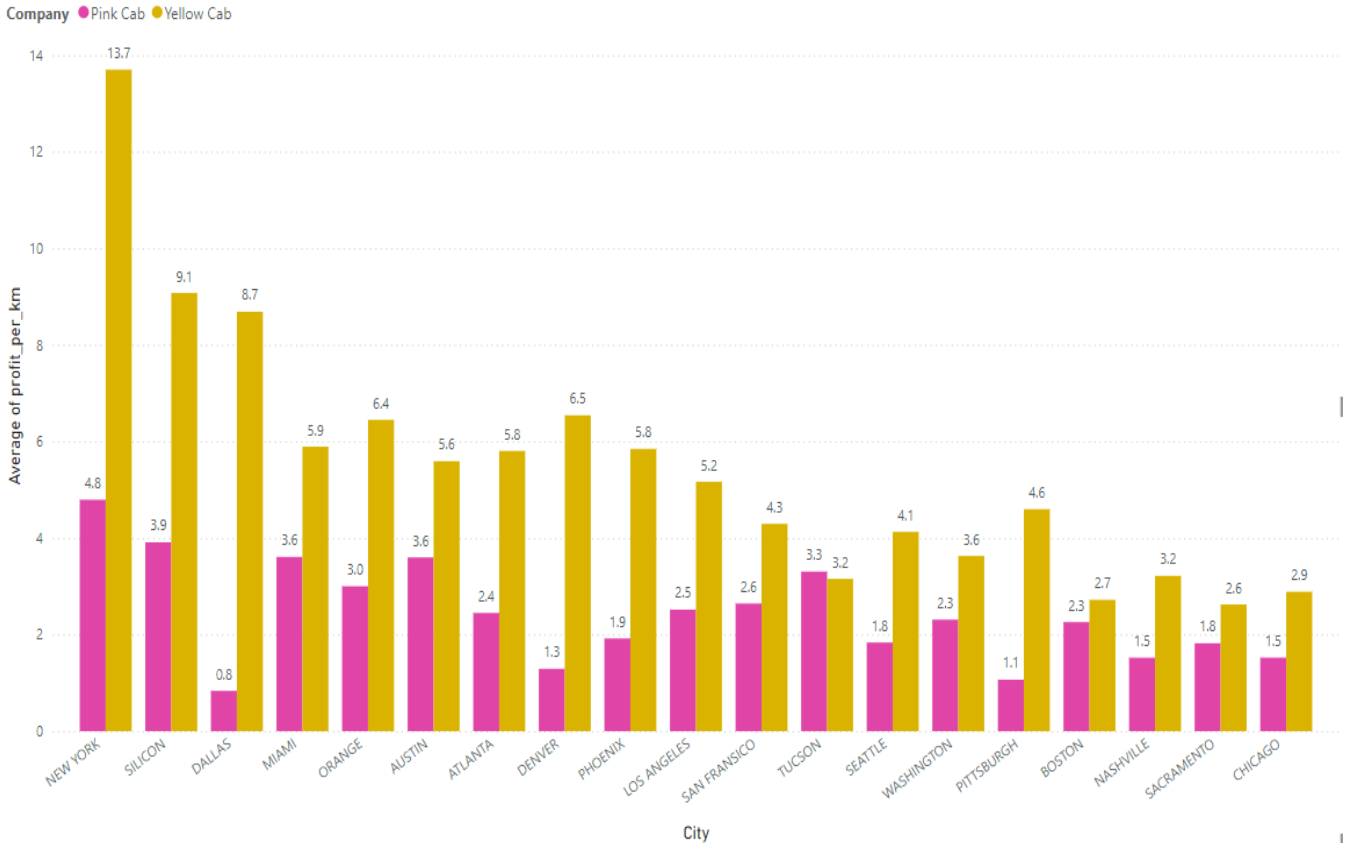
| City | Company | Sum of profit_per_km |
|---|---|---|
| NEW YORK | Yellow Cab | 13.68 |
| SILICON VALLEY | Yellow Cab | 9.05 |
| DALLAS | Yellow Cab | 8.66 |
| DENVER | Yellow Cab | 6.54 |
| ORANGE COUNTY | Yellow Cab | 6.44 |
| MIAMI | Yellow Cab | 5.88 |
| PHOENIX | Yellow Cab | 5.83 |
| ATLANTA | Yellow Cab | 5.79 |
| AUSTIN | Yellow Cab | 5.57 |
| LOS ANGELES | Yellow Cab | 5.14 |
| NEW YORK | Pink Cab | 4.79 |
| PITTSBURGH | Yellow Cab | 4.61 |
| SAN DIEGO | Yellow Cab | 4.31 |
| SEATTLE | Yellow Cab | 4.13 |
| SILICON VALLEY | Pink Cab | 3.90 |
| WASHINGTON | Yellow Cab | 3.63 |
| AUSTIN | Pink Cab | 3.60 |
| MIAMI | Pink Cab | 3.60 |
| TUCSON | Pink Cab | 3.27 |
| NASHVILLE | Yellow Cab | 3.20 |
| TUCSON | Yellow Cab | 3.16 |
| ORANGE COUNTY | Pink Cab | 3.01 |
| CHICAGO | Yellow Cab | 2.86 |
| BOSTON | Yellow Cab | 2.72 |
| SAN DIEGO | Pink Cab | 2.64 |
| SACRAMENTO | Yellow Cab | 2.61 |
| LOS ANGELES | Pink Cab | 2.50 |
| ATLANTA | Pink Cab | 2.42 |
| WASHINGTON | Pink Cab | 2.31 |
| BOSTON | Pink Cab | 2.23 |
| PHOENIX | Pink Cab | 1.89 |
| SEATTLE | Pink Cab | 1.82 |
| **Total** | | **149.82** |



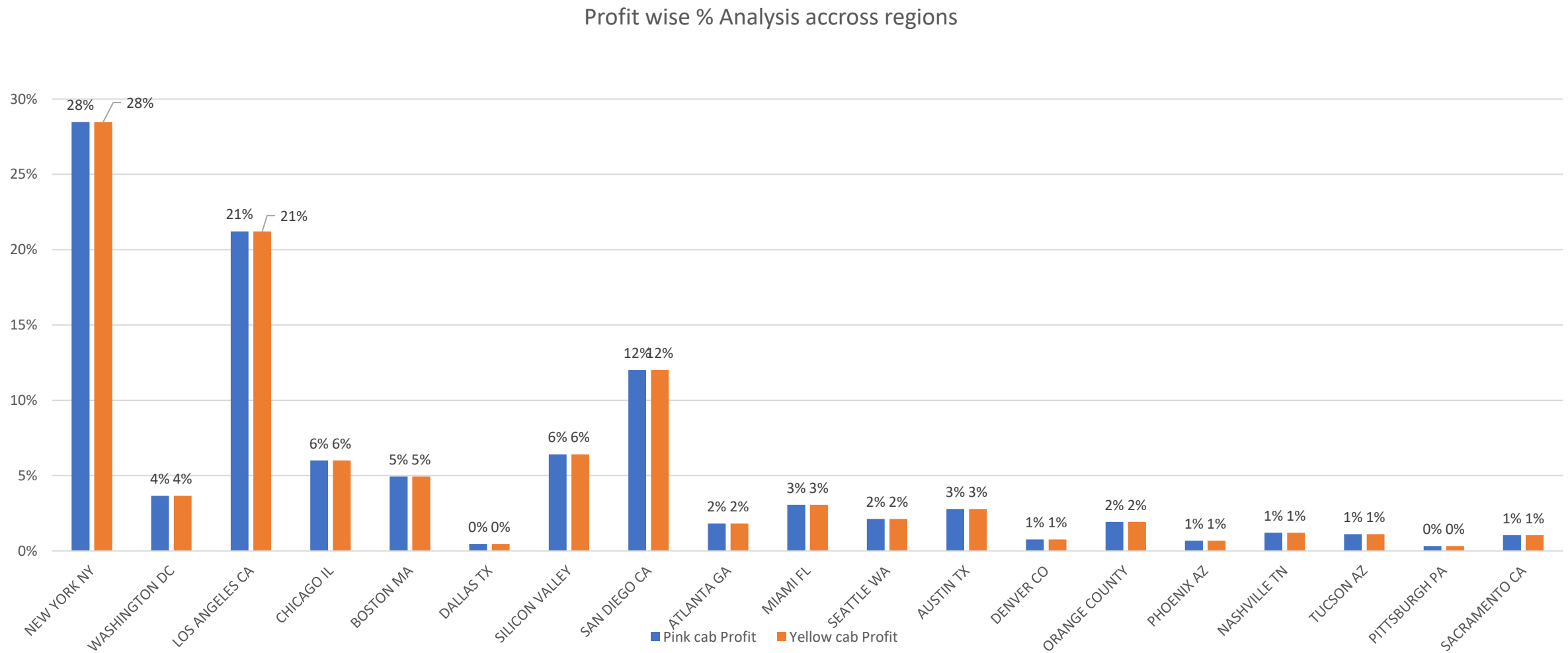
# Profit Analysis: *Geography wise*

## Insights

- *New York* had the highest customer base as opposed to *Chicago* which was lowest customer base
- In *New York*, the Yellow cab company has more than double the profit margin as compared to other metropolitan cities
- The difference in average Profit per km for pink and yellow cabs differed in almost all metropolitan cities except *Boston*
- The average profit across all the cities for Pink cab and yellow cab are 2.42 and 5.46



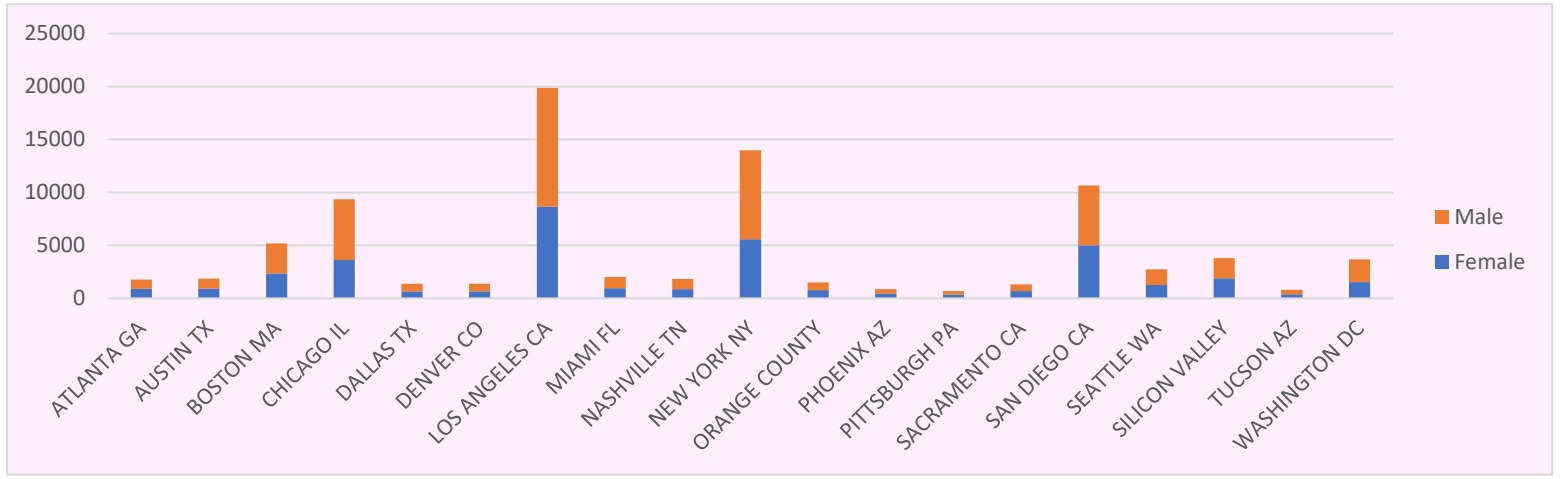
# Profit Analysis: *Geography wise*

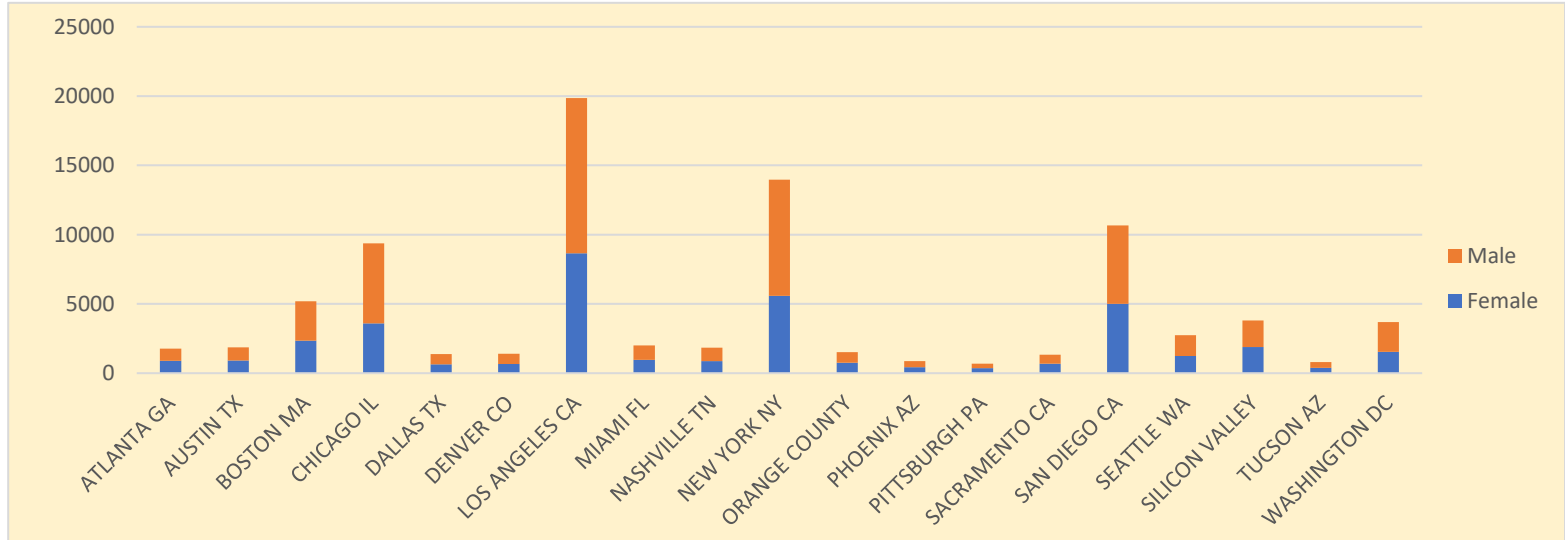


# Profit Analysis: *Gender wise*

## Insights

- There are more number of females travelling through Pink Cabs then Yellow cabs
- The average *Female* to *Male* ratio in case of yellow cab is 0.73 and pink cab is 0.77
- *New York* has the highest number of male and female customers travelling in the city
- *Pittsburgh* has the lowest number of customers in case of yellow cab
- Potential to target more female population in *yellow cab* exists as the business base for Yellow customers are high in almost all metropolitan cities

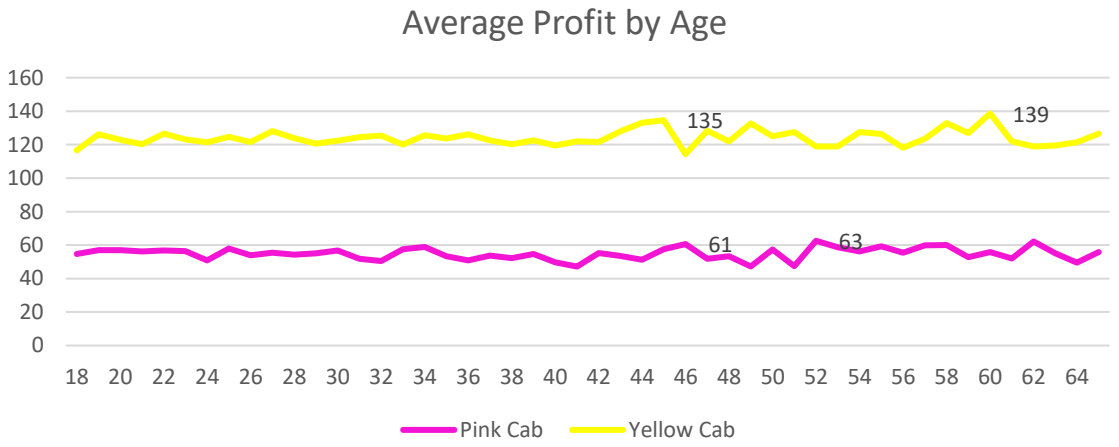


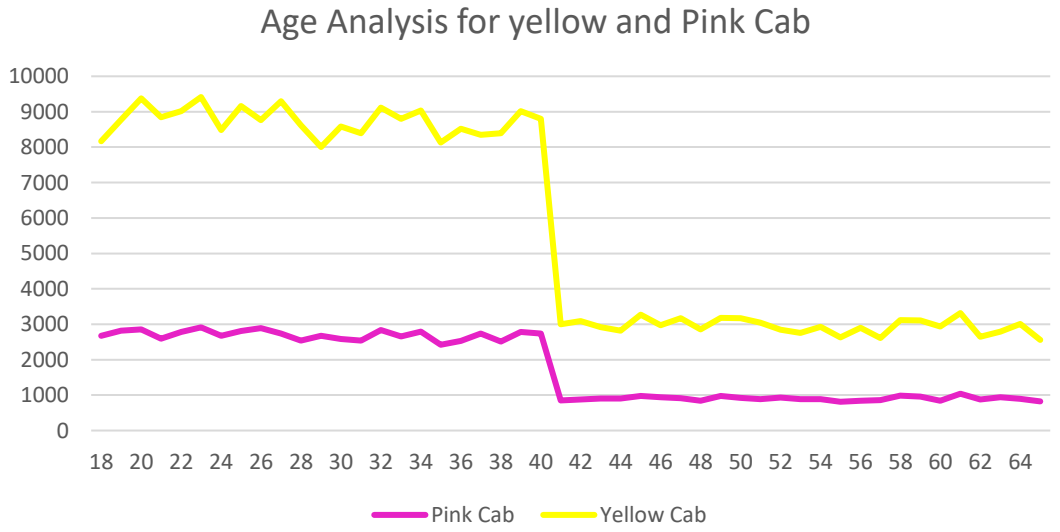



# Profit Analysis: *Age wise*

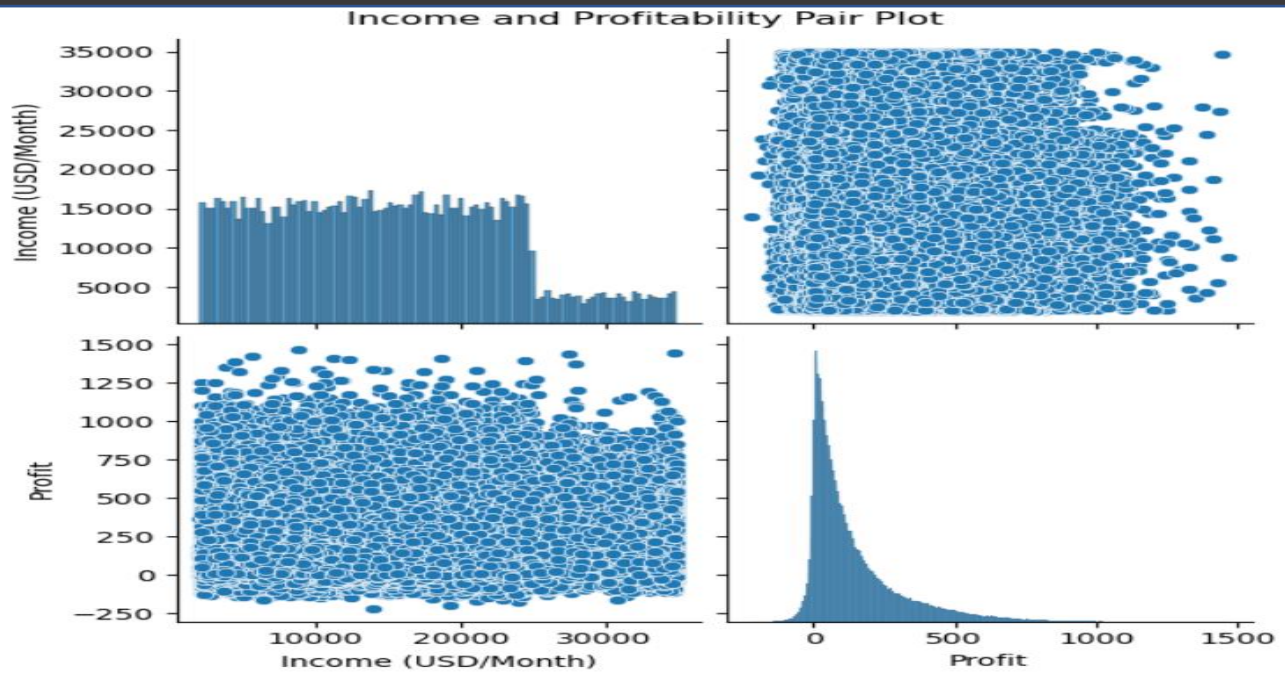
## Insights

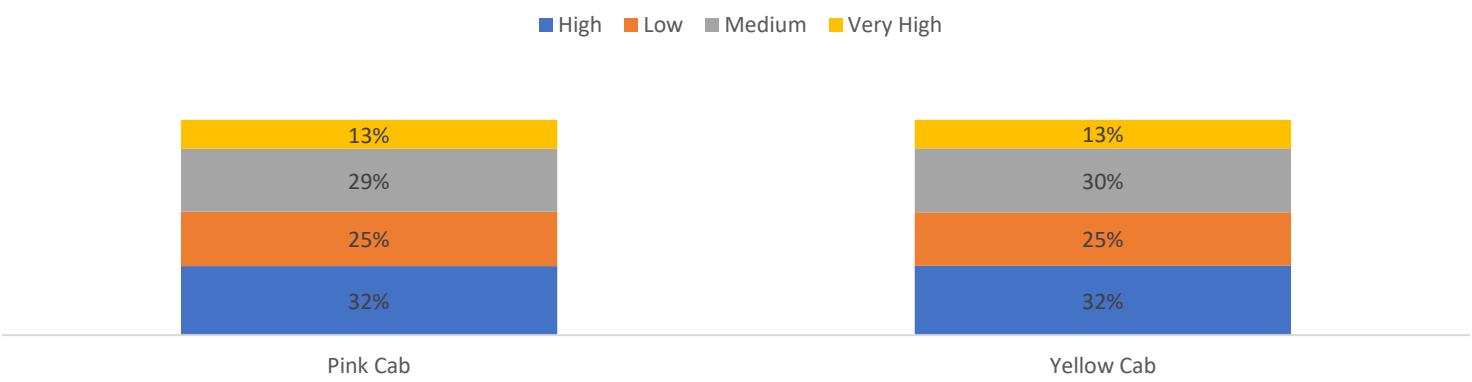
- The average Profit earned by the customers for yellow cab are 135$ and 139 $ for 46 and 60 years of age
- The average Profit earned by the customers for pink cab are 61$ and 63 $ for 48 and 54 years of age
- Majority of the cab user both yellow and Pink are youngsters and account to 73% of the approximately total volume of customers.
- Older Age customers do not frequently use the cab services.
- Clustering algorithms can be used further understand the behavioral patterns in the dataset



# Profit Analysis: *Income*



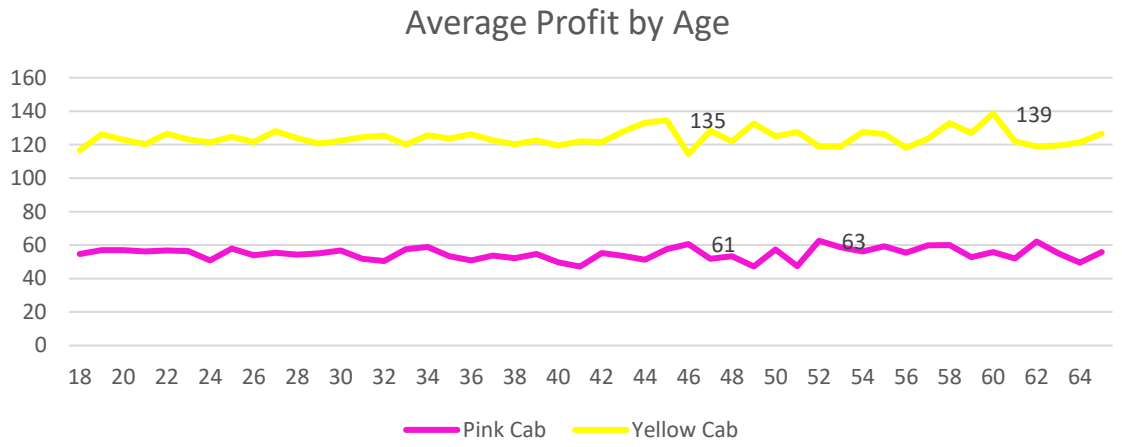



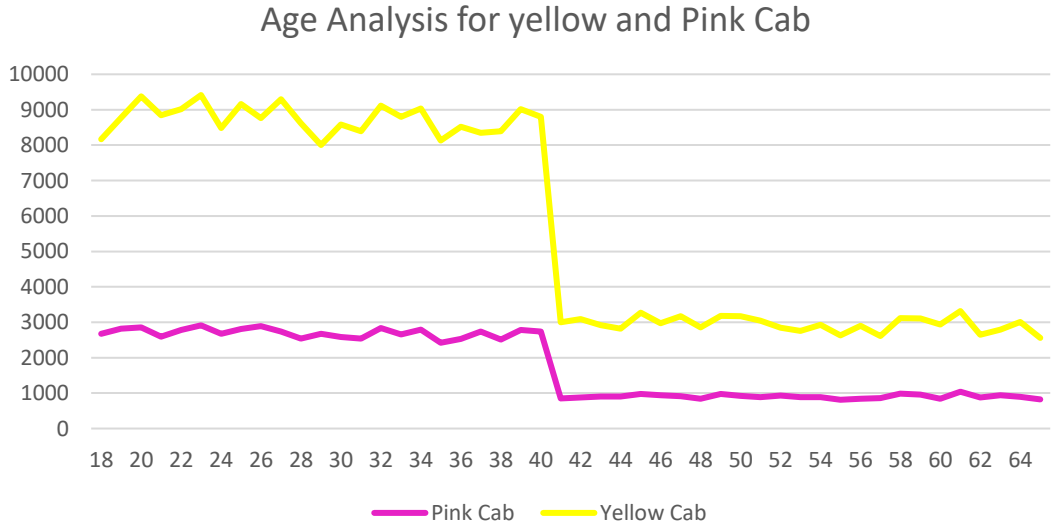

- There is no correlation between Income earned and Profit.
- Middle class and high class contributes more in the profit as well as in the customer base of both the cabs
- There is equal proportion of all classes travelling in both Yellow and Pink Cab

# Profit Analysis: *Gender wise*

## Insights

- The average Profit earned by the customers for yellow cab are 135$ and 139 $ for 46 and 60 years of age
- The average Profit earned by the customers for pink cab are 61$ and 63 $ for 48 and 54 years of age
- Majority of the cab user both yellow and Pink are youngsters and account to 73% of the approximately total volume of customers.
- Older Age customers do not frequently use the cab services.
- Clustering algorithms can be used further understand the behavioural patterns in the dataset

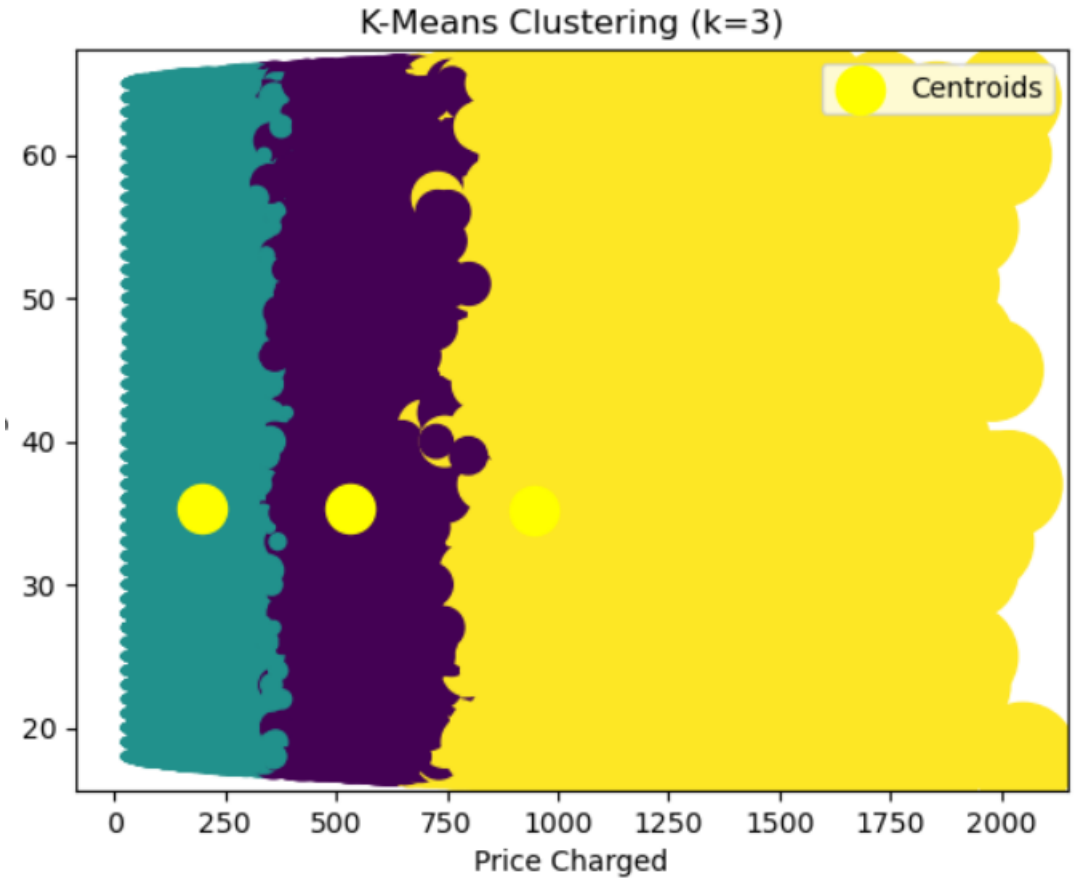


# Profit Analysis: *Clustering Analysis*

## Insights

- Customers are segmented based on *'Price Charged, Age and profit* continuous variables.
- The number of clusters are determined as per K-elbow method (k=3)
- The average Age of Cluster 0 is 35.The profit gained by the customer is 149$ and Price Charged is 532$
- The average Age of Cluster 1 is 30. The profit gained by the customer is 40$ and Price Charged is 199$
- The average Age of Cluster 2 is 40. The profit gained by the customer is 478$ and Price Charged is 946$



# Recommendations

Based on the detailed analysis of both Pink and Yellow Cabs following are the recommendations

- Yellow cabs operate in more cities compared to pink cabs. As a result of this broader coverage, the profits associated with yellow cabs exhibit significantly more outliers than those of pink cabs.

- Customers with medium incomes can be encouraged to use cab services by offering them additional discounts and vouchers. This strategy aims to boost sales by employing the K-Means approach to create distinct clusters for more effective targeting..

- The customer attrition rate for Yellow Cab is significantly lower than that of Pink Cab, indicating better customer retention. This advantage makes Yellow Cab more suitable for profitability, as it can capitalize on repeat business and enhance revenue streams compared to its competitor.

- Yellow Cab's broader geographic reach, higher utilization rates, lower customer churn, economies of scale, and more effective marketing all contribute to its superior profitability per kilometer compared to Pink Cab across major cities

- More females travel in pink cabs then yellow cabs. It would be very important to analyze the factors contributing to business growth for Pink cab so that we can take the learnings from Pink cab

**On the basis of above point , we will recommend Yellow cab for investment.**

# Thank You