Exploratory Data Analysis

G2M Case Study

19-Mar-2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

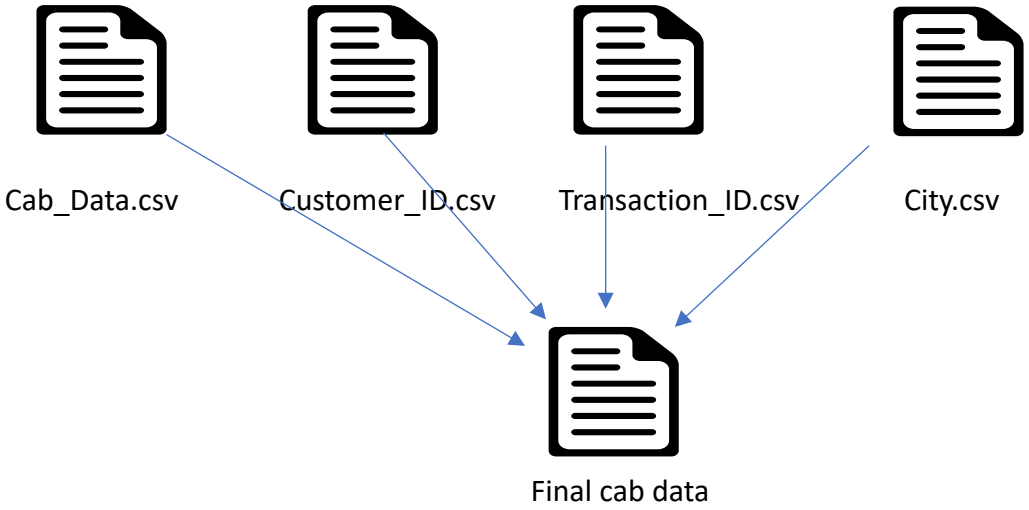
The Exploratory Data Analysis was conducted to compare the performance of two taxi companies, Yellow Cab and Pink Cab, across various criteria such as **customer reach**, **pricing strategy**, **market share**, **customer loyalty**, **age range target**, and **profit margin**. The The Exploratory Data Analysis showed that Yellow Cab appears to be a better company to invest in, based on all the characteristics above. However, investment decisions should not solely be based on the results of The Exploratory Data Analysis , and other factors such as financial stability and market trends should be taken into account.

Problem Statement

1. What is the distribution of the data, and are there any outliers or anomalies?
2. How can we visualize and summarize the data to gain insights and make decisions?
3. Who has better pricing strategy?
4. Who has greater profit and profit margin?
5. Who is performing better in each city?
6. Who is performing better over time?

Approach

- 16 Features(including 4 derived features)
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points: 440,098
- Total Nulls: 80,706



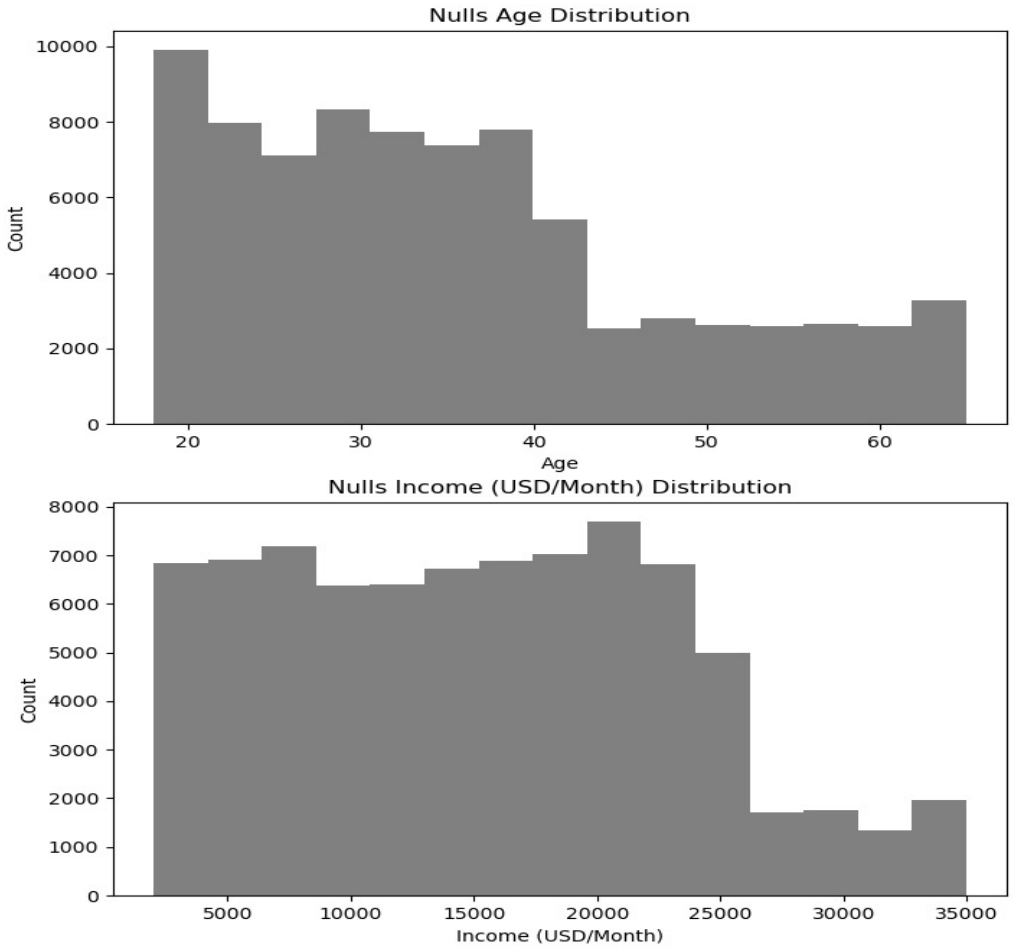
Assumptions:

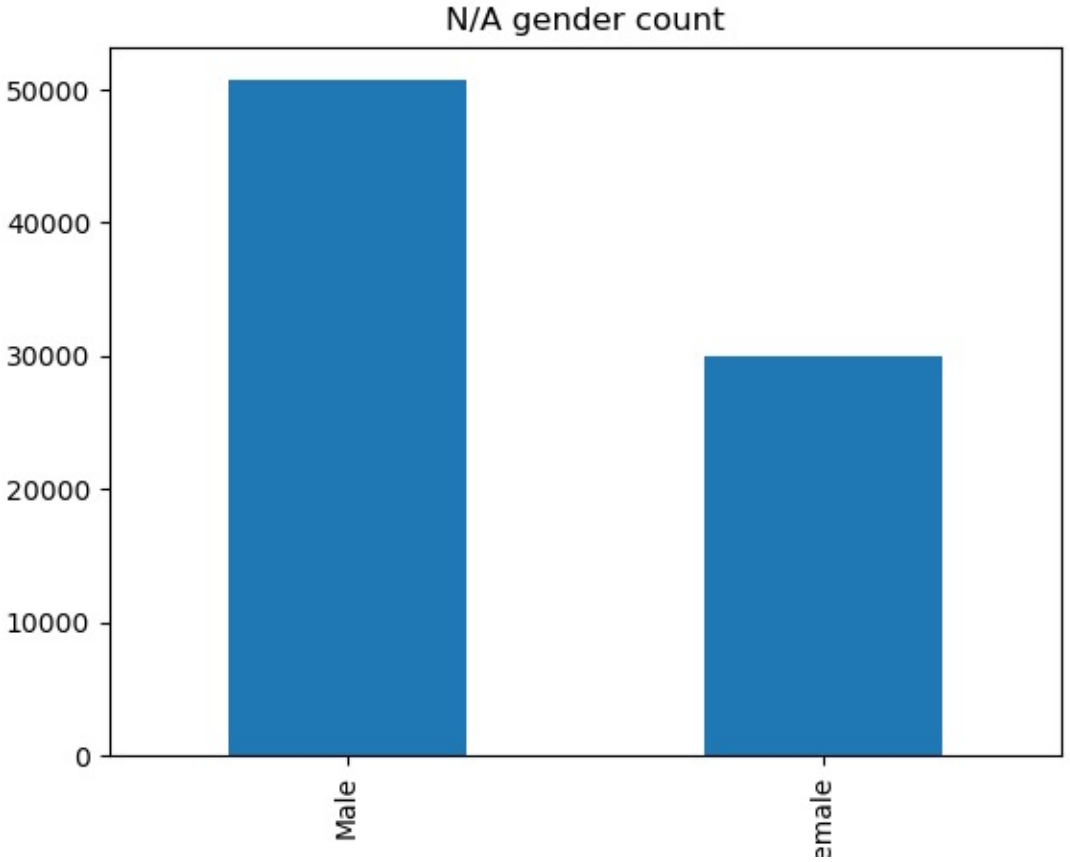
- Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.

- Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.

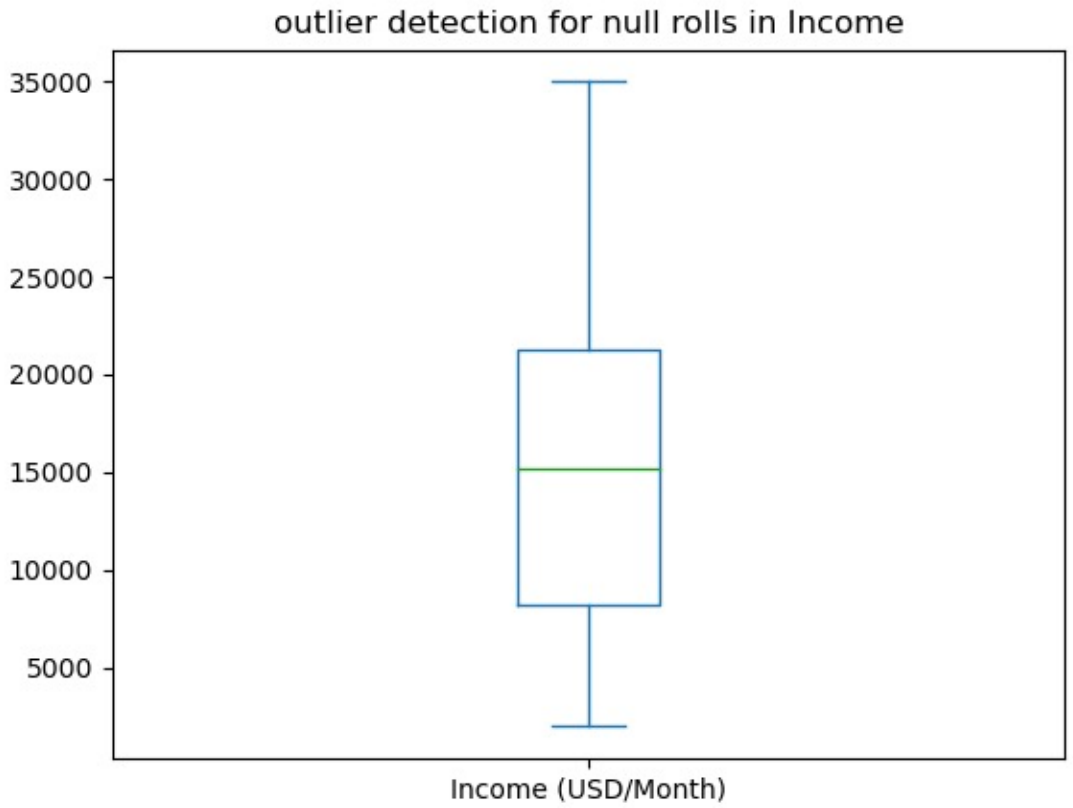
- Users feature of city dataset is treated as number of cab users in the city. we have assumed that this can be other cab users as well(including Yellow and Pink cab)

EDA – NULLs Analysis

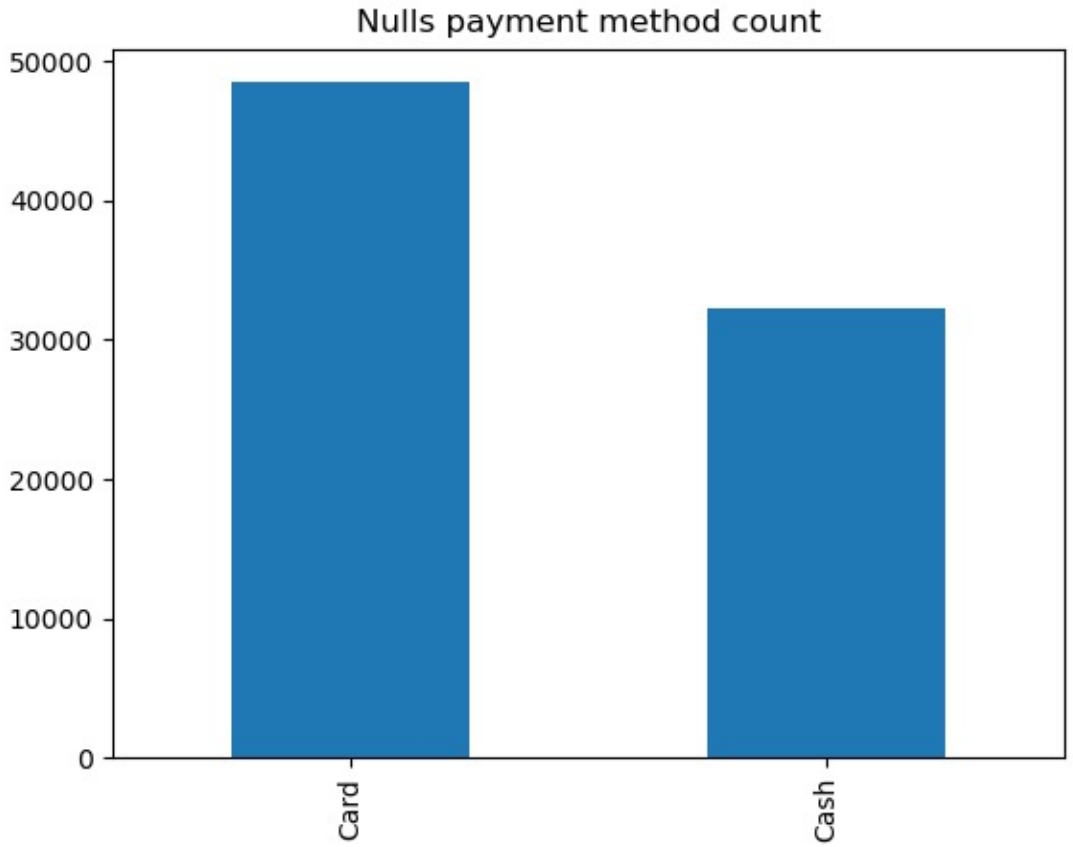




EDA – NULLs Analysis



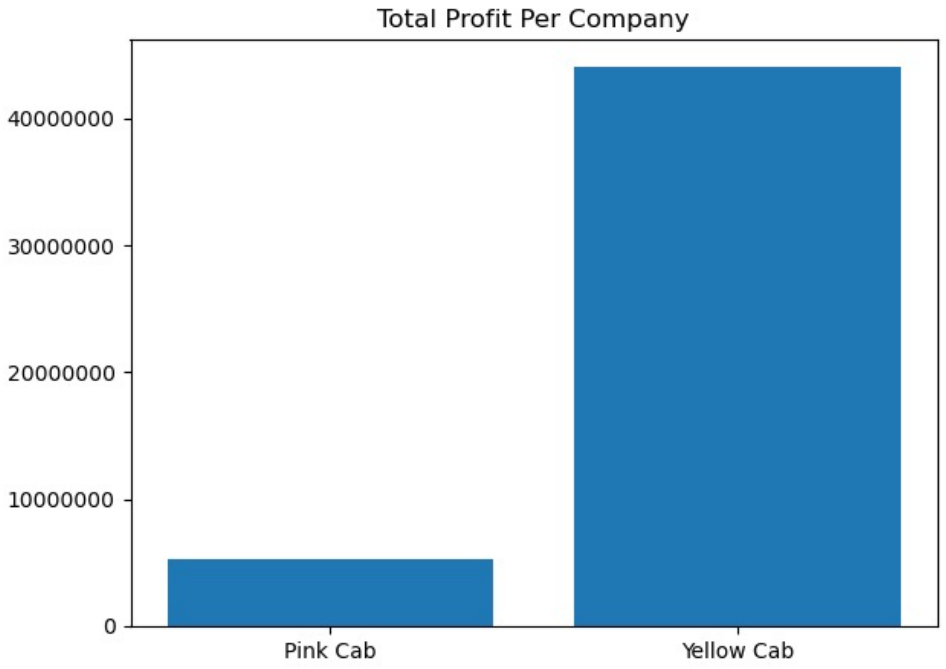
EDA – NULLs Analysis

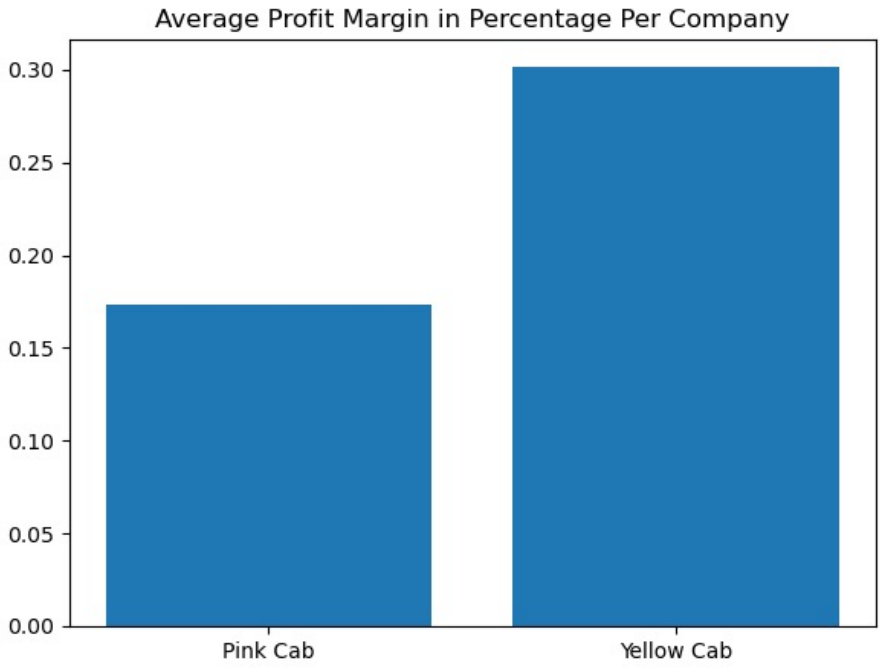



```
full.isnull().sum()
```

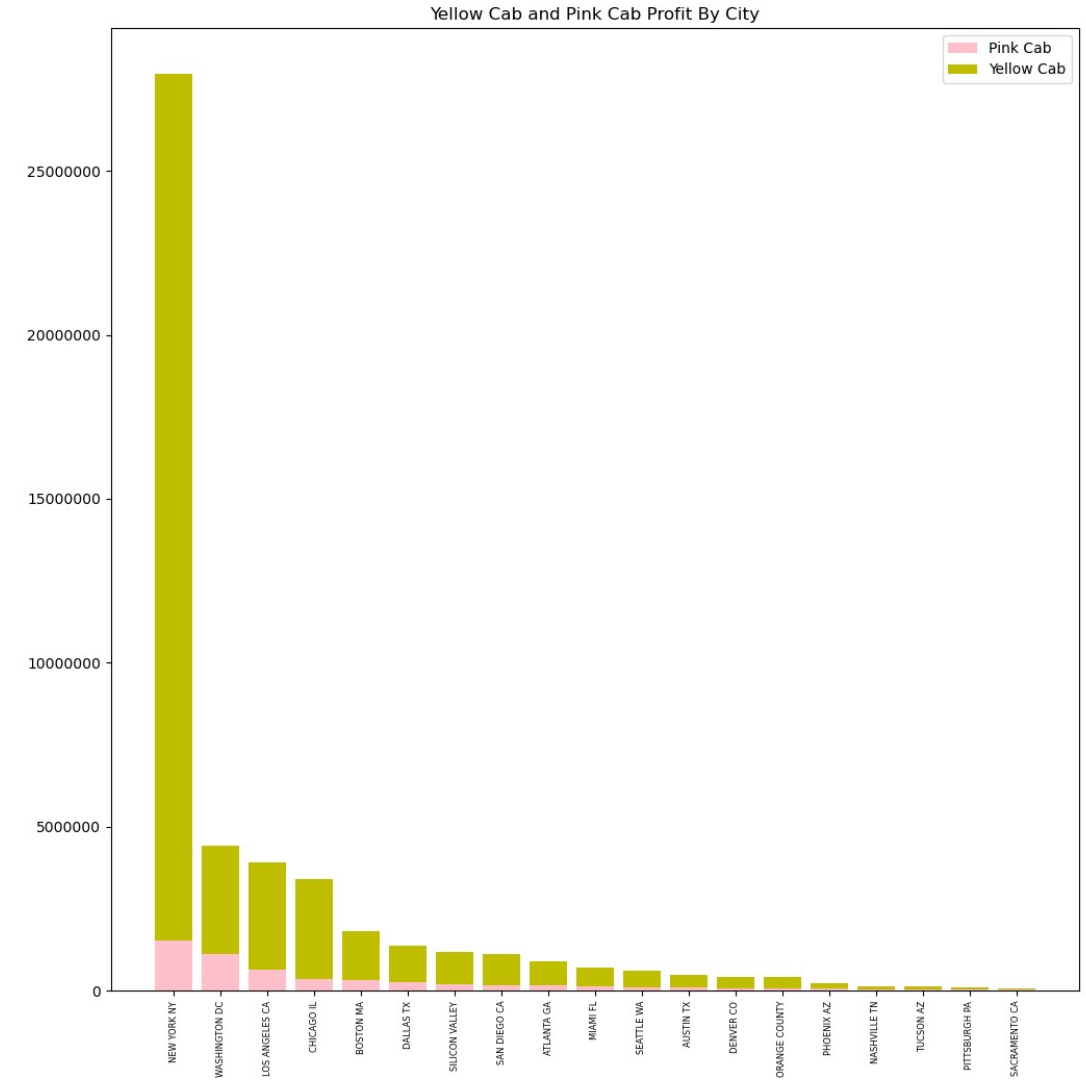
```
Transaction ID           0
Customer ID              0
Payment_Mode             0
Date of Travel       80706
Company              80706
City                 80706
KM Travelled         80706
Price Charged        80706
Cost of Trip         80706
Gender                   0
Age                      0
Income (USD/Month)       0
profit               80706
profit_margin_in_%   80706
price_per_KM         80706
profit_per_KM        80706
dtype: int64
```

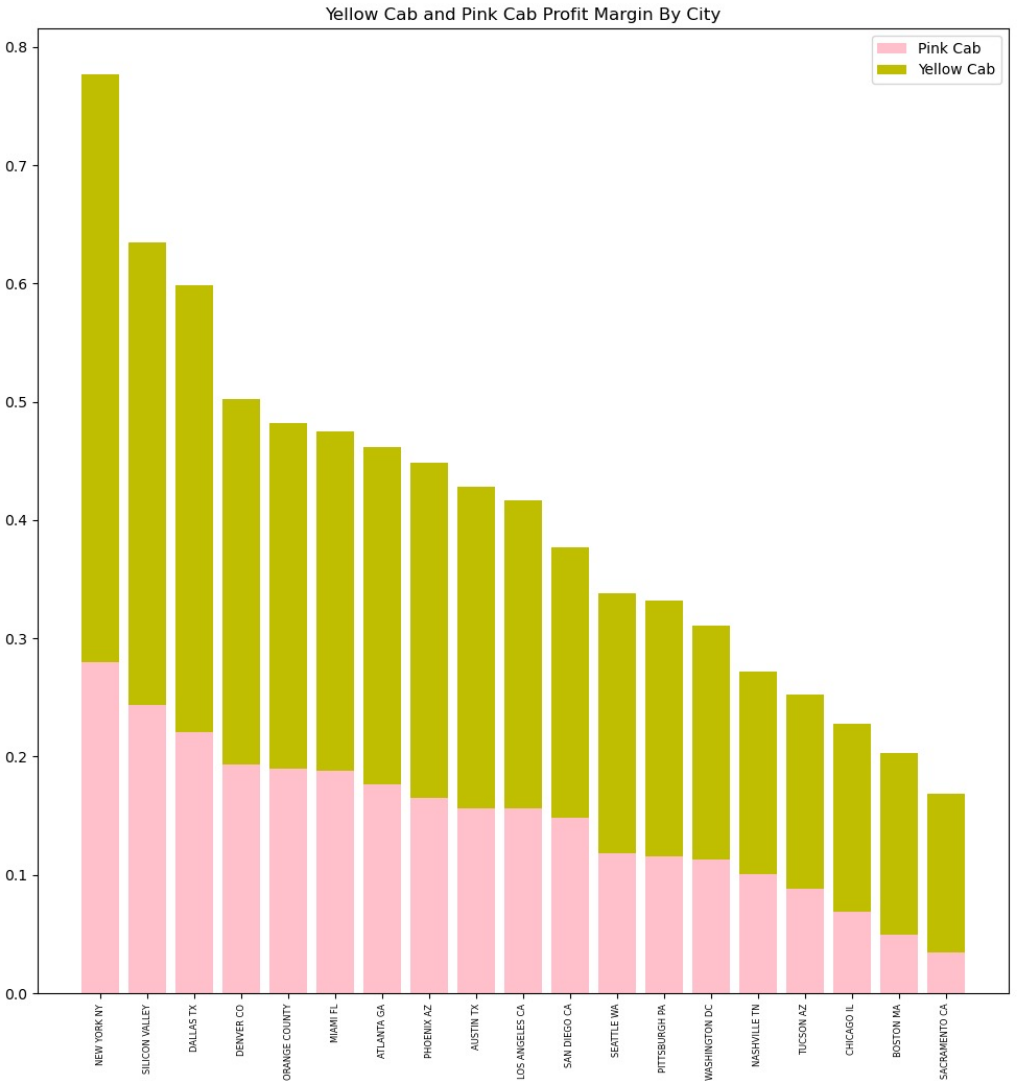
EDA - Profit Analysis



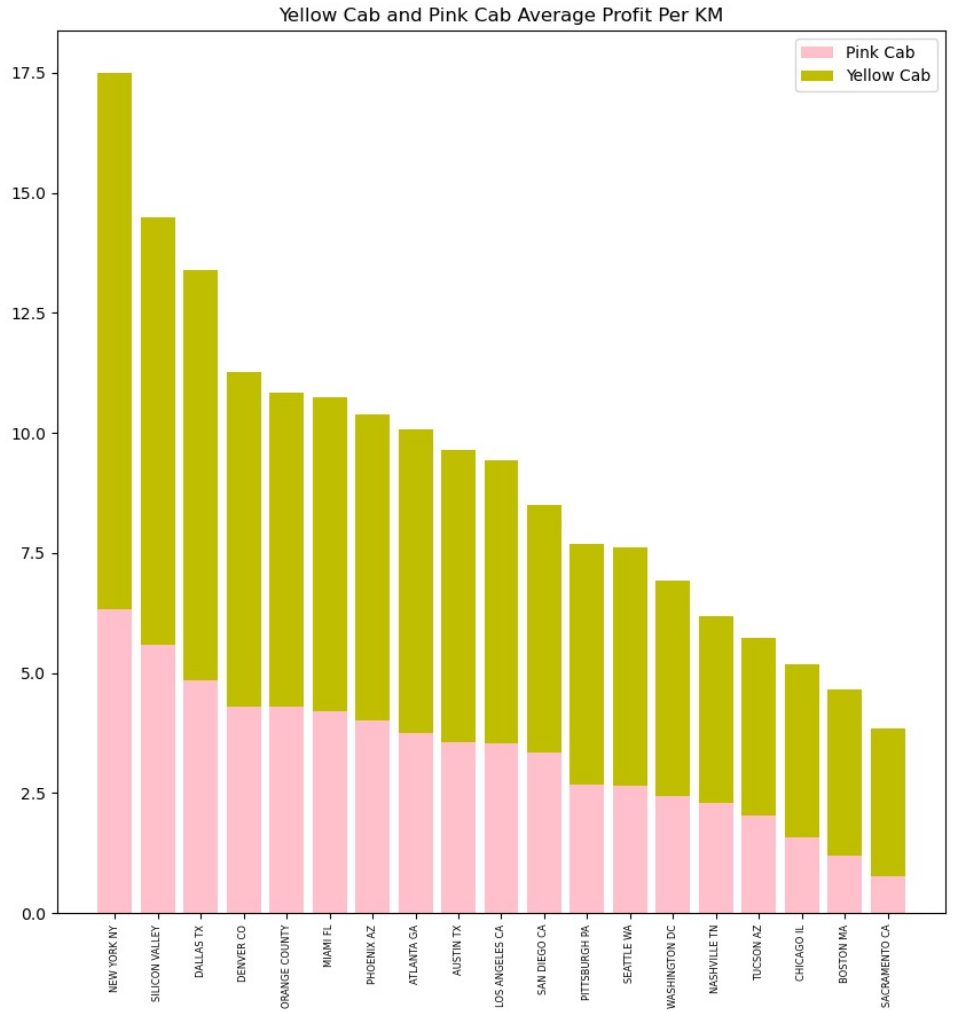


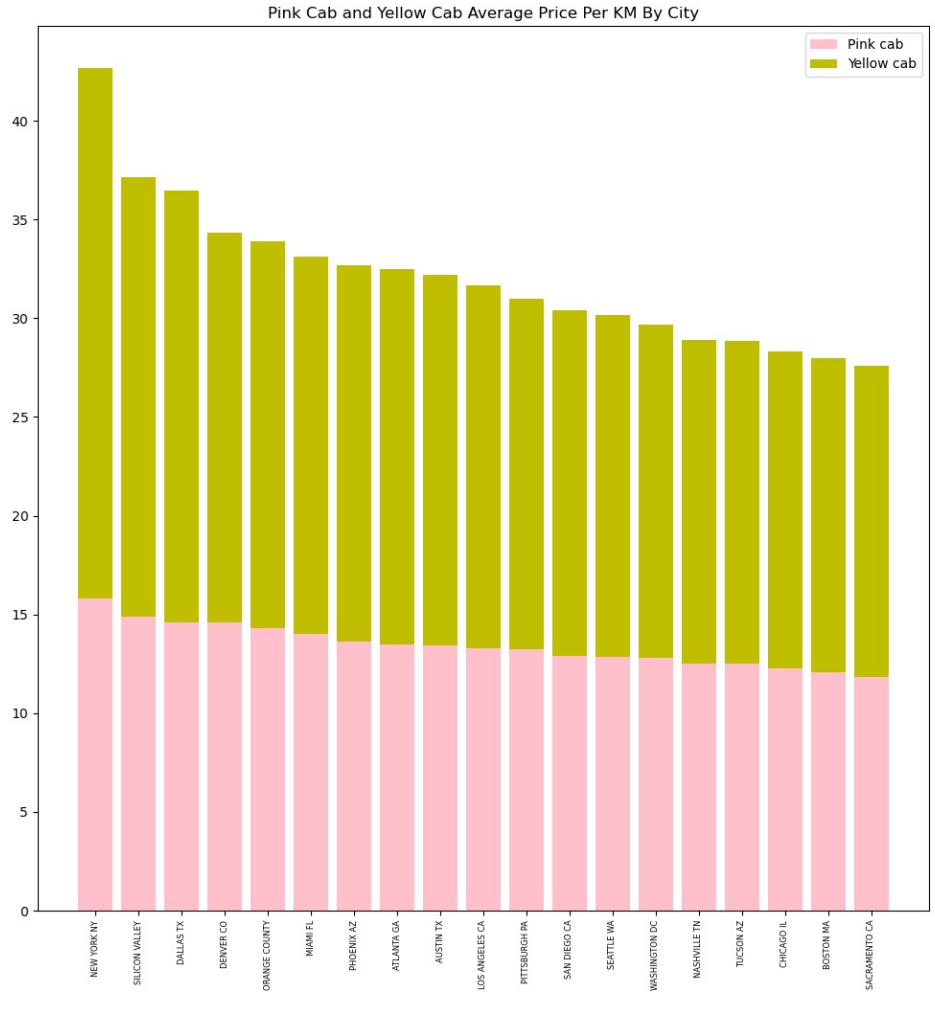
EDA - Profit Analysis



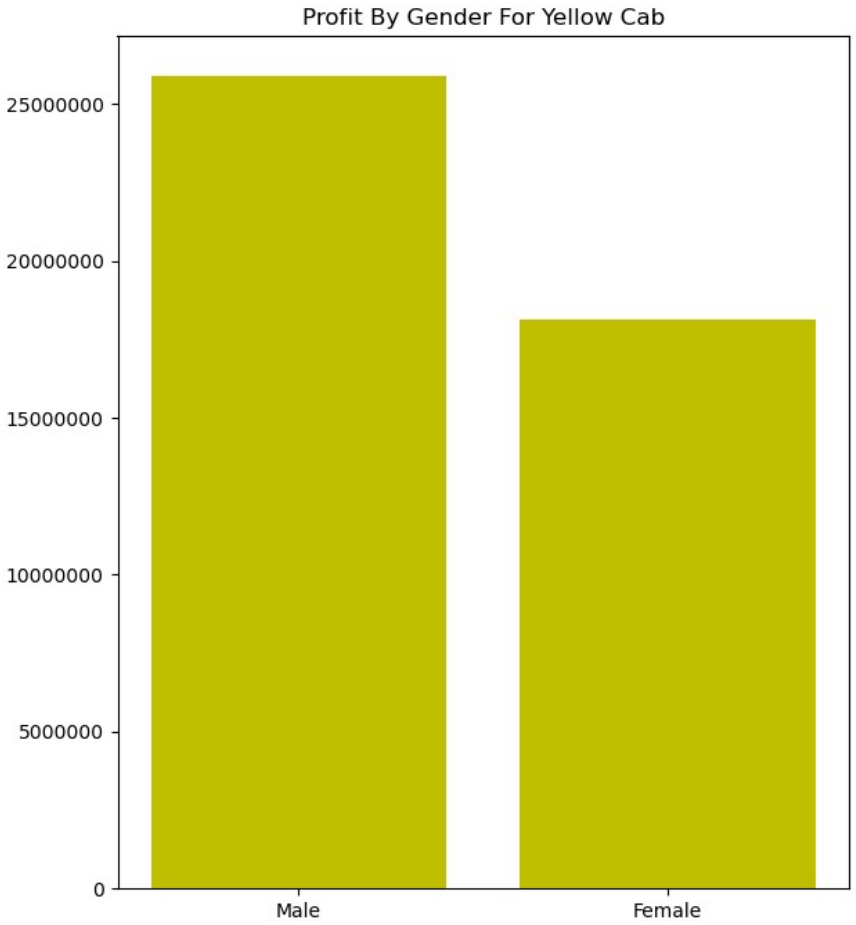


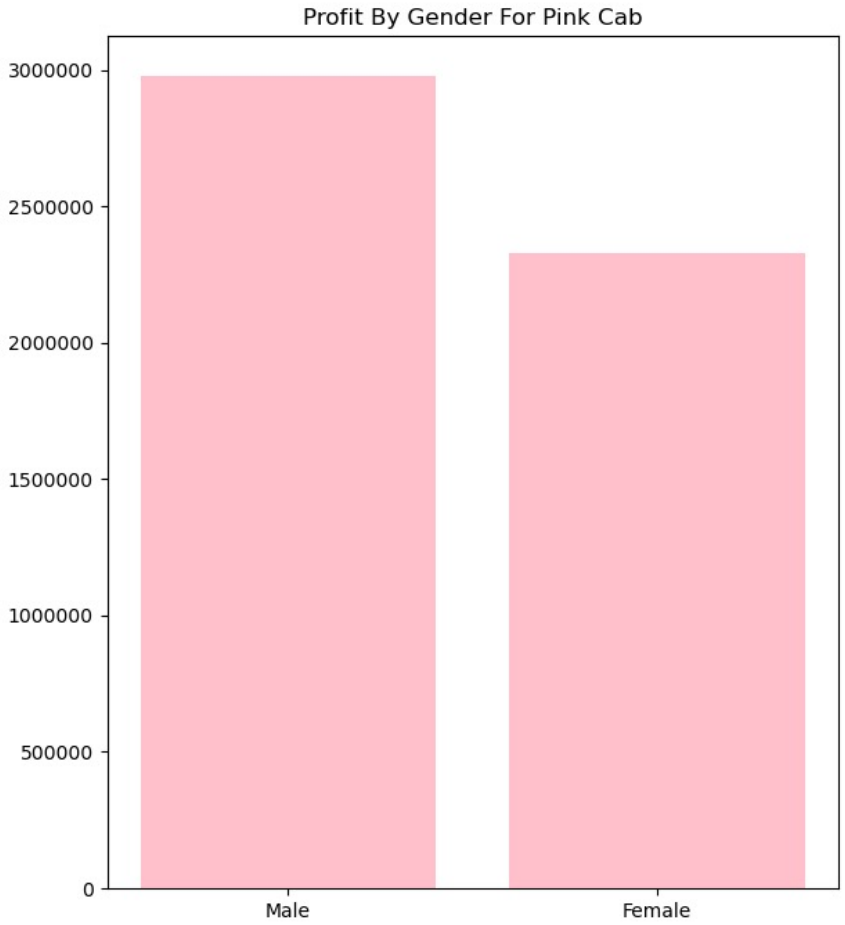
EDA - Profit Analysis



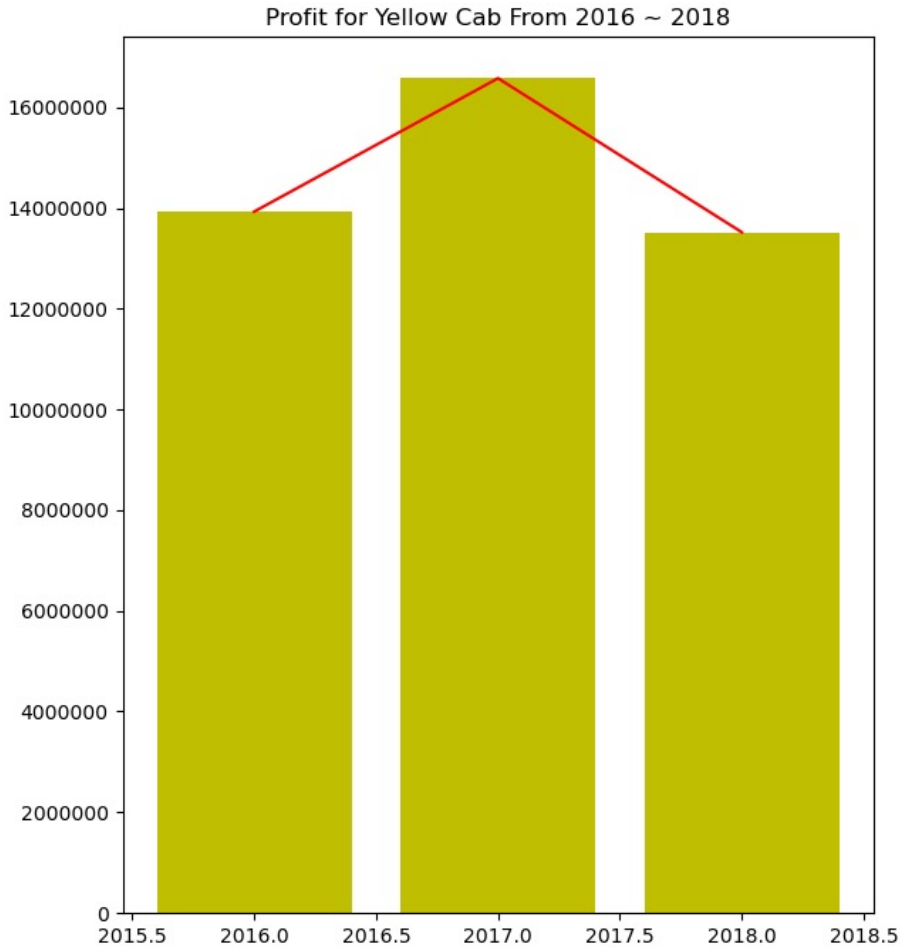


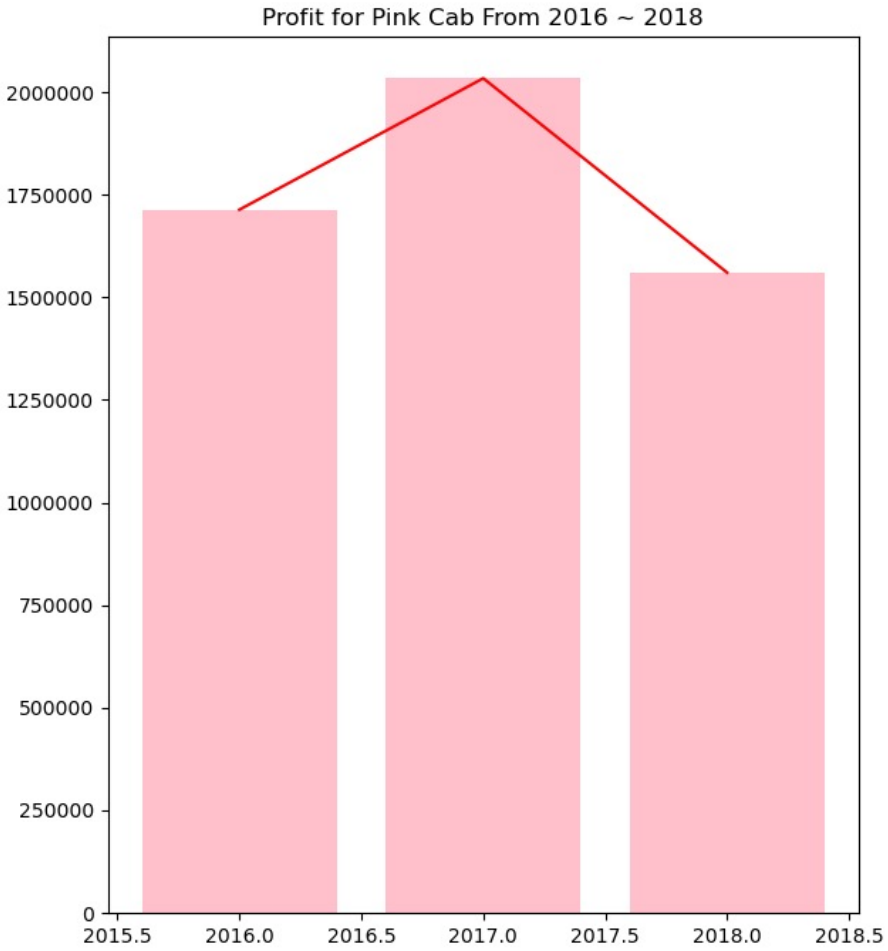
EDA - Profit Analysis



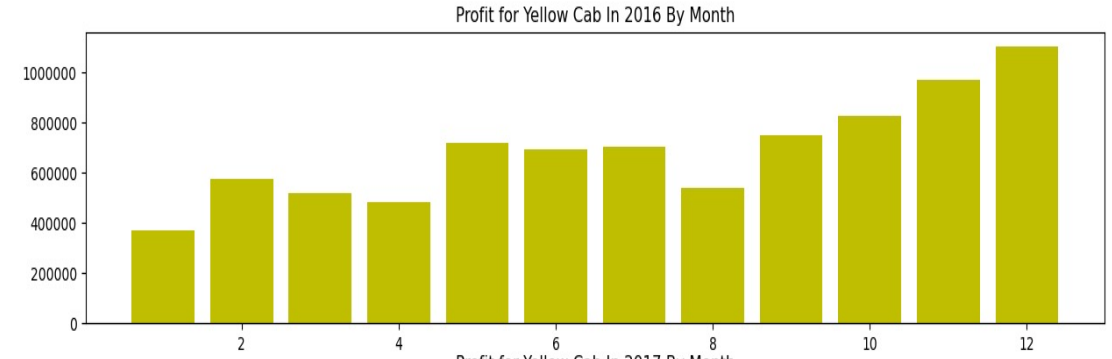


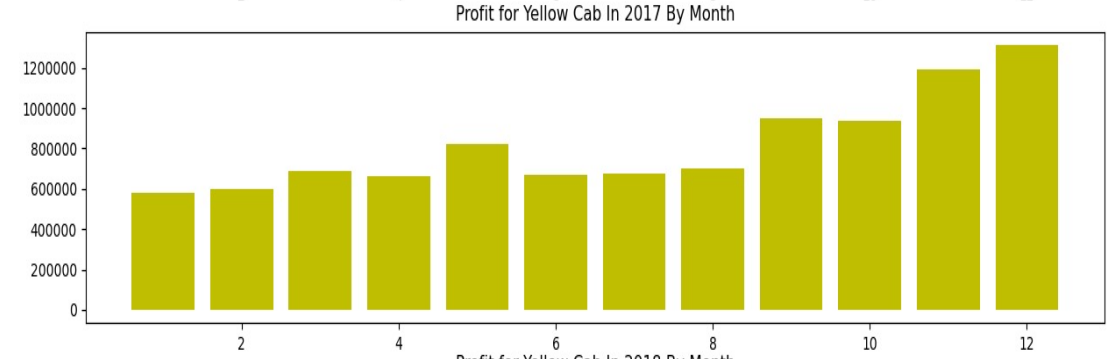
EDA - Profit - Time Series

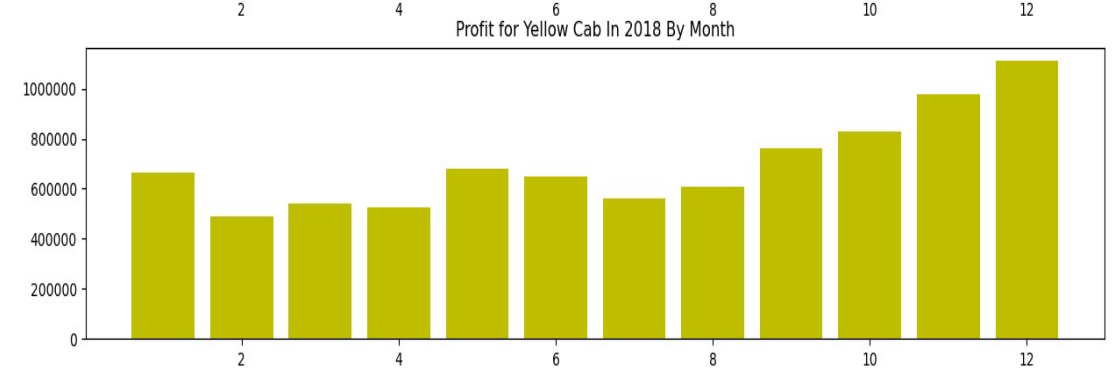


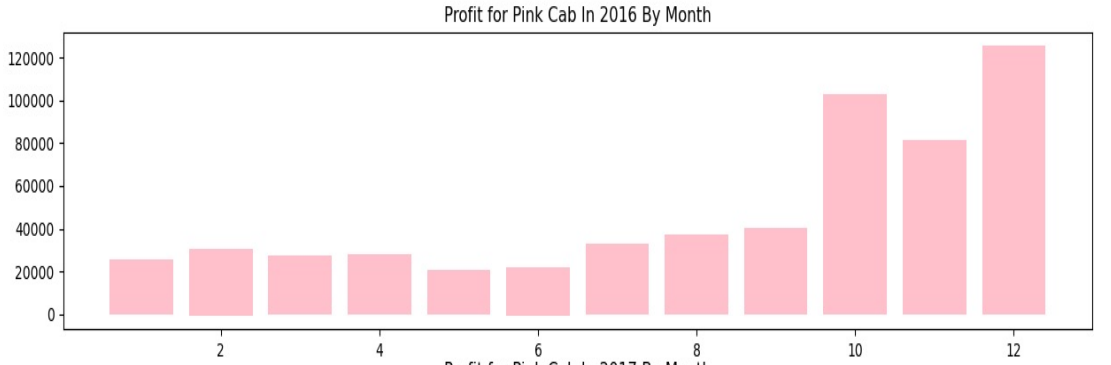


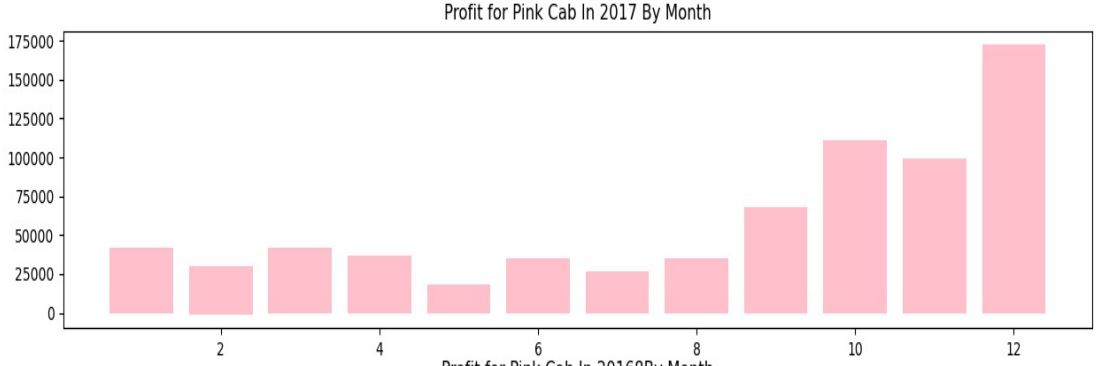
EDA - Profit - Time Series

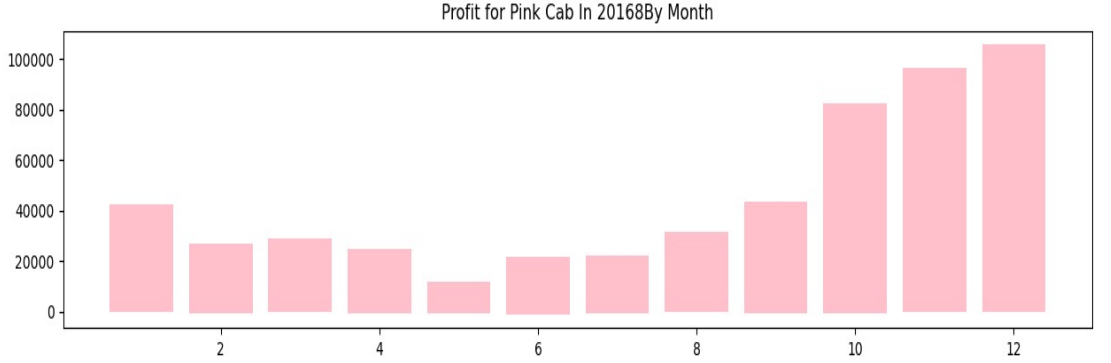




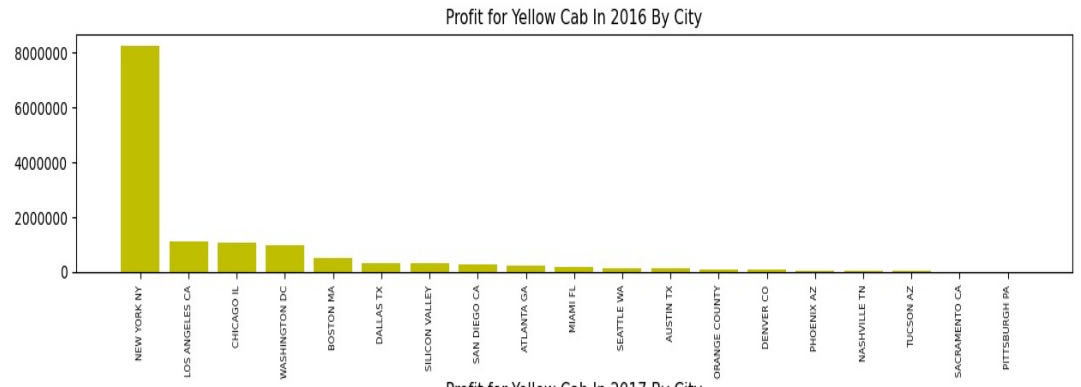


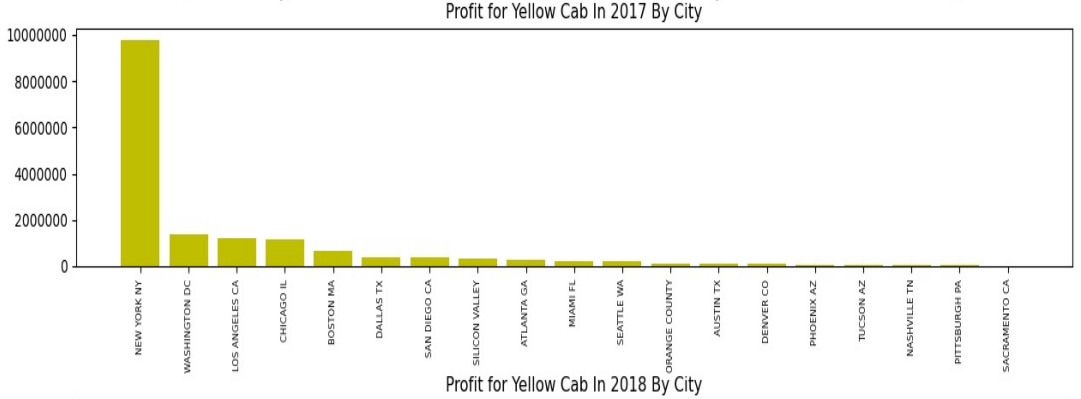


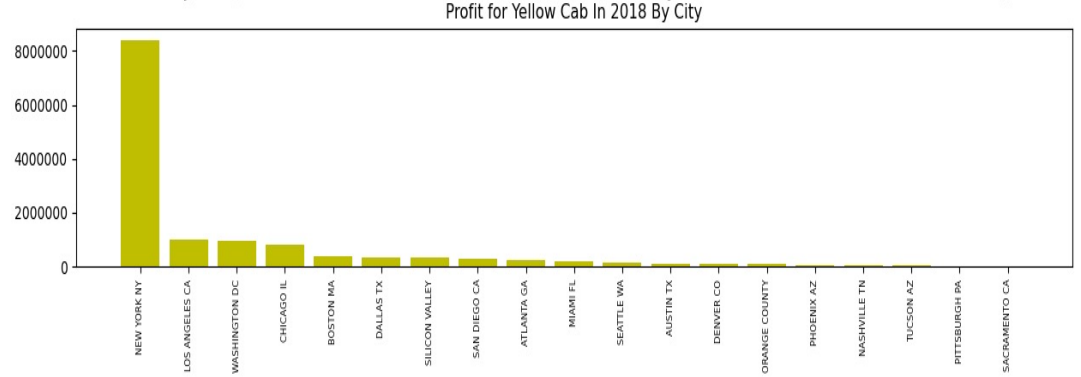


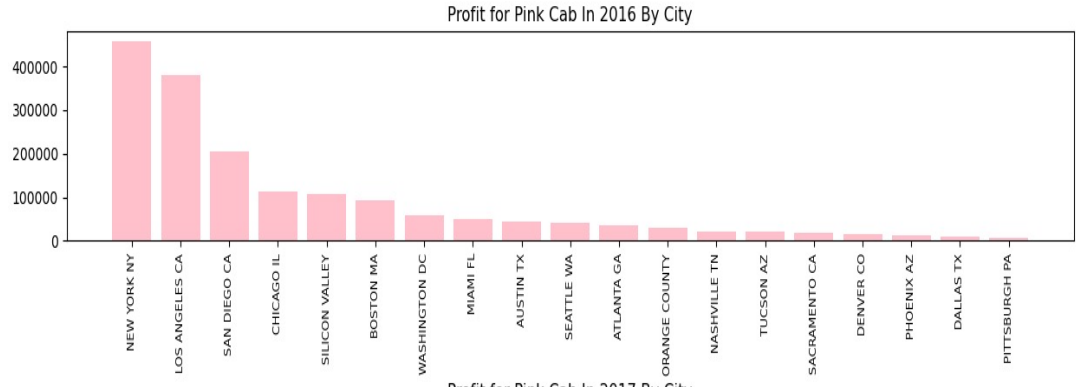


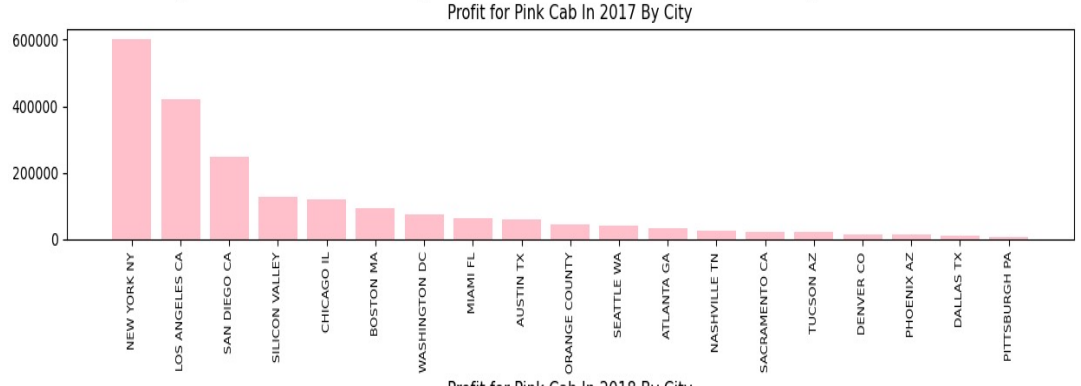
EDA - Profit - Time Series

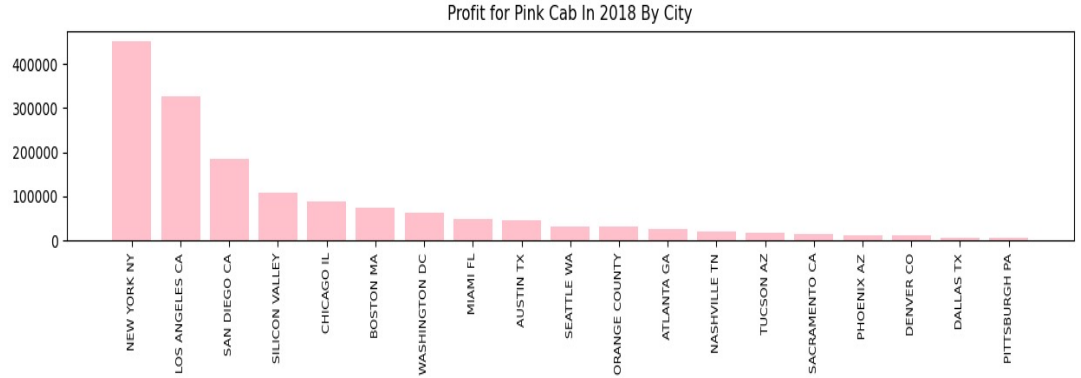




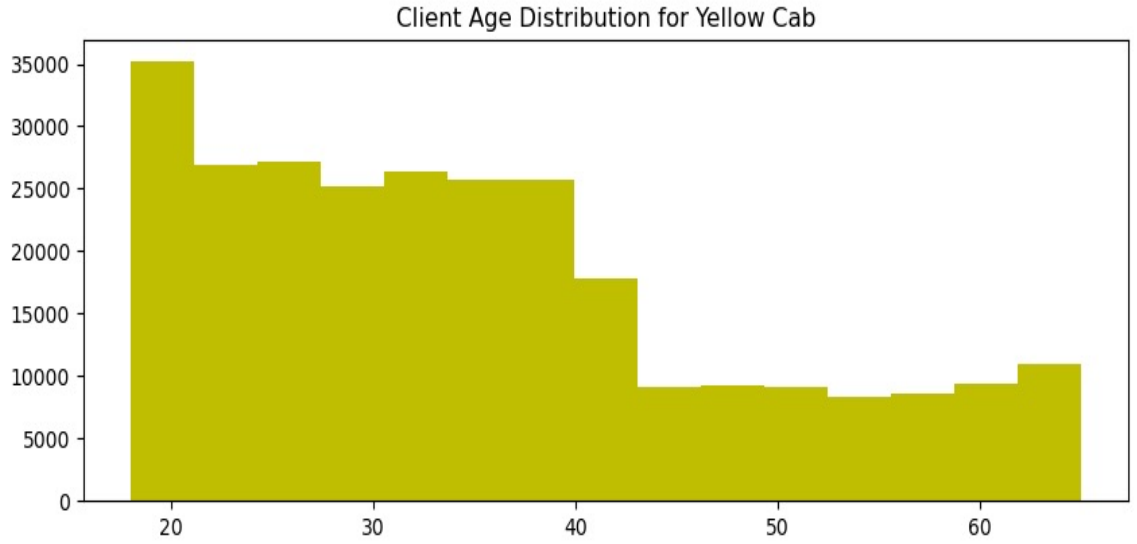


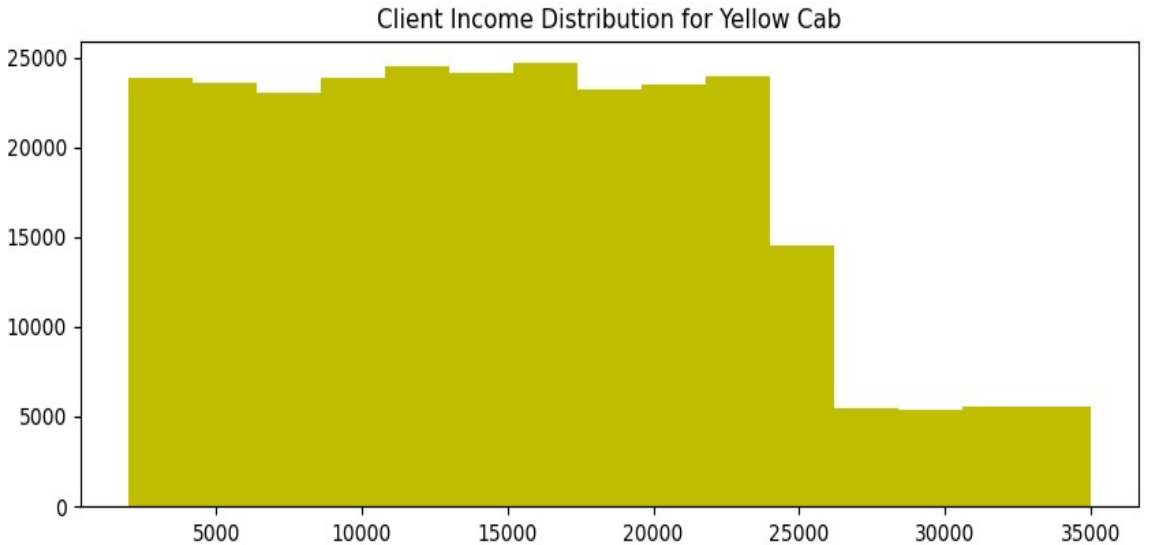


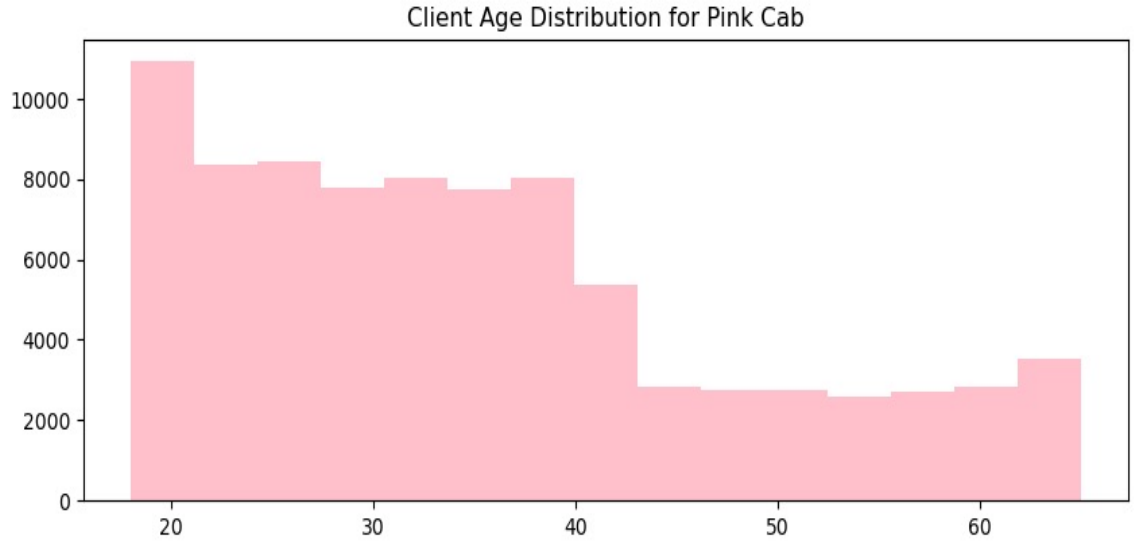




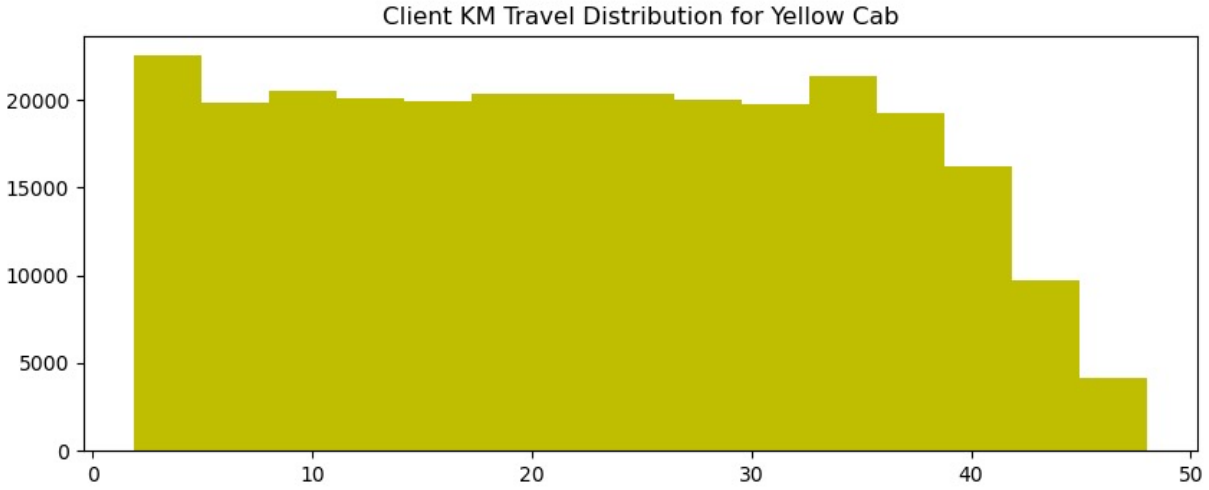
EDA – Customer Analysis

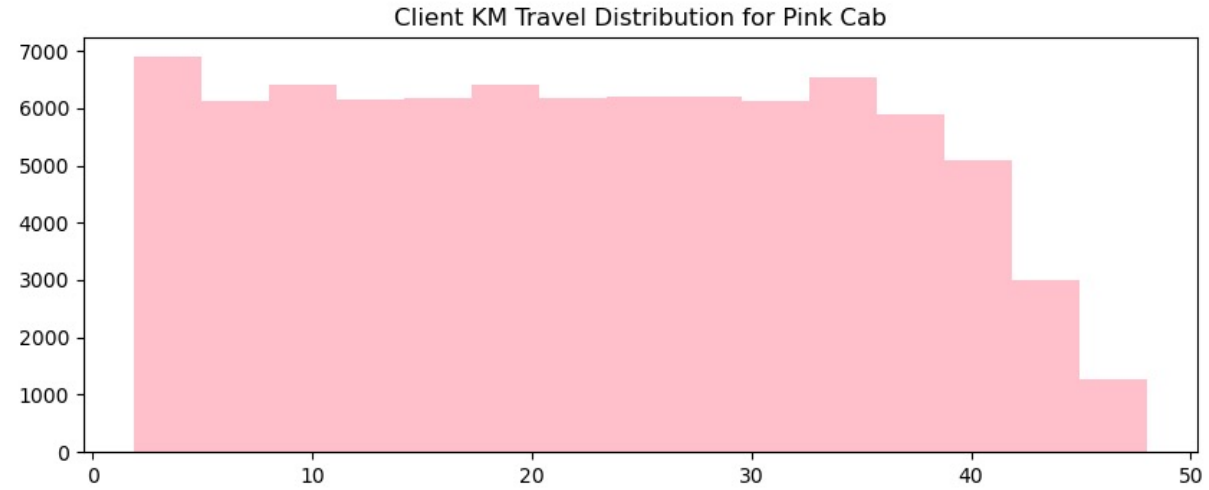




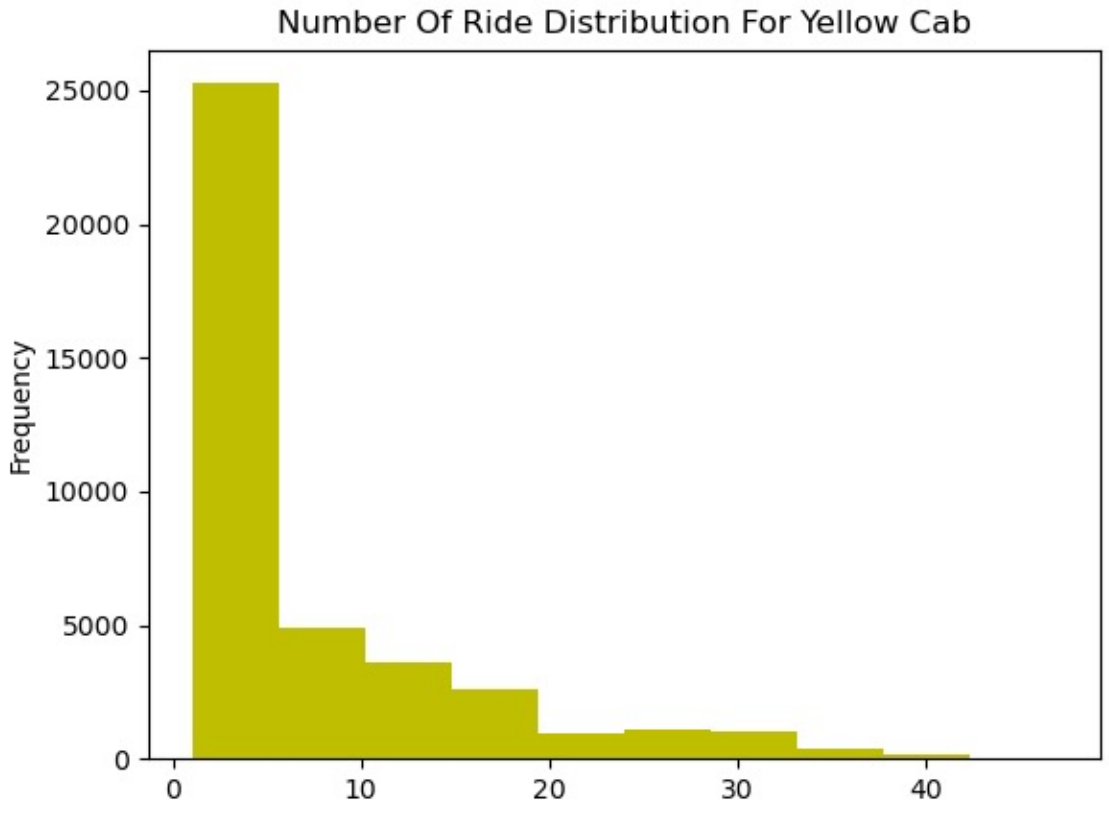


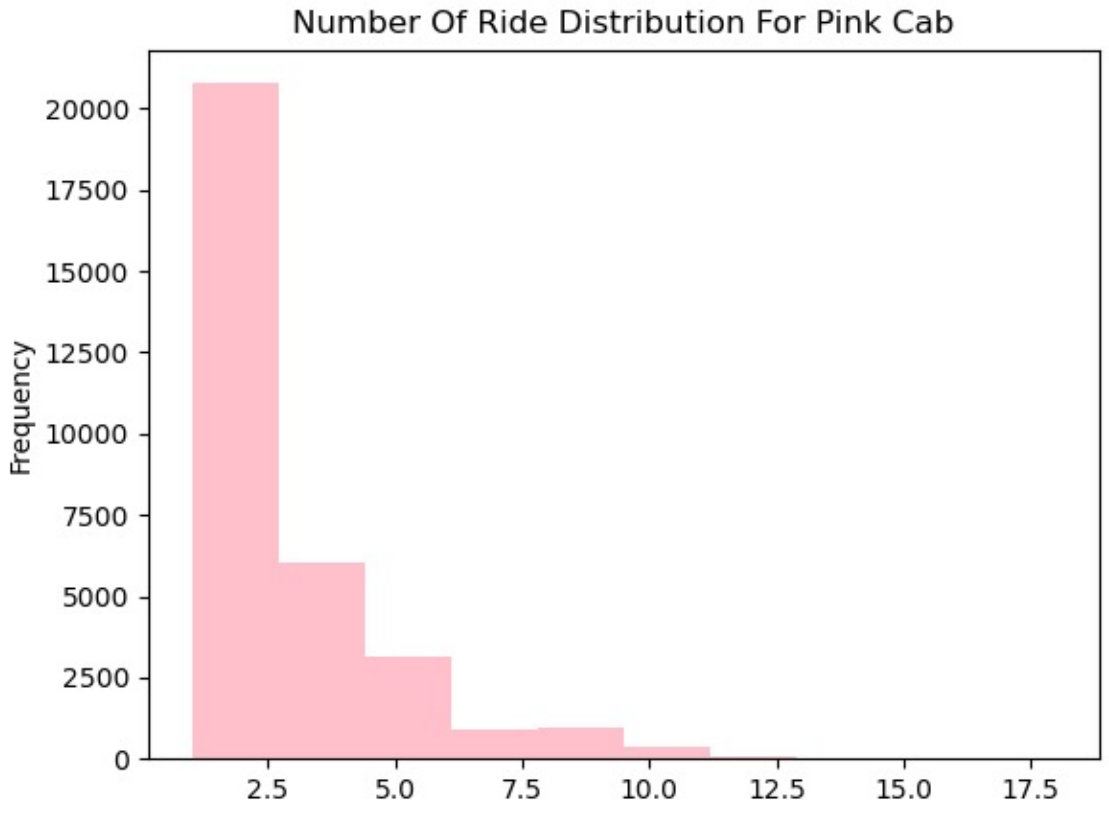
EDA – Customer Analysis



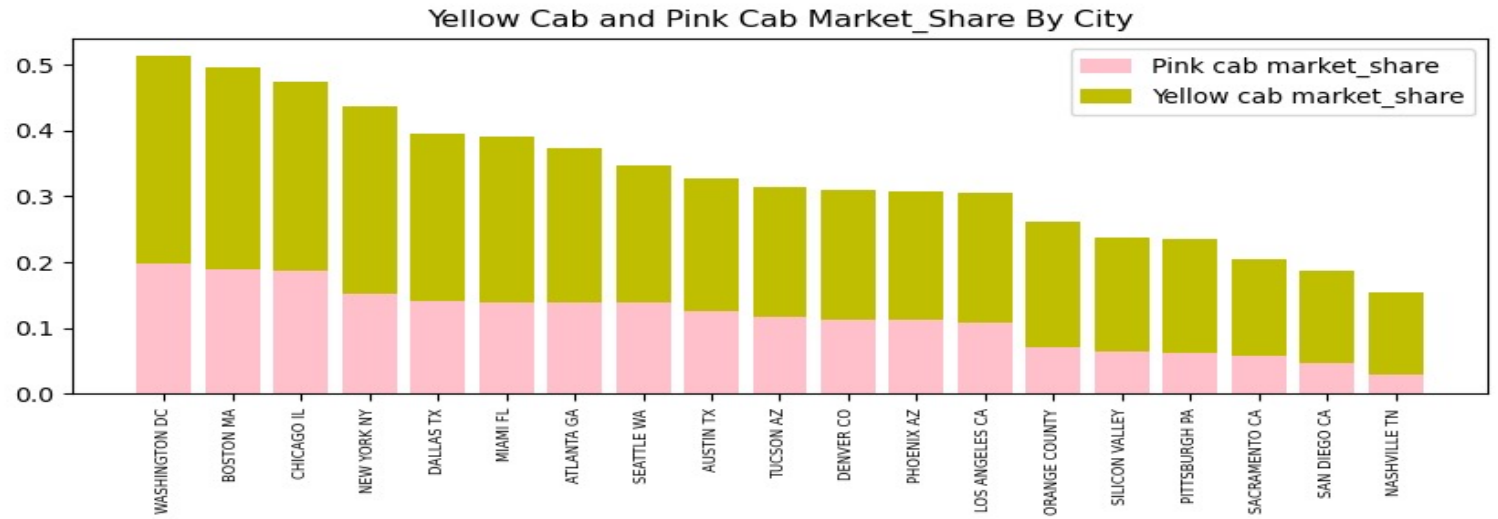


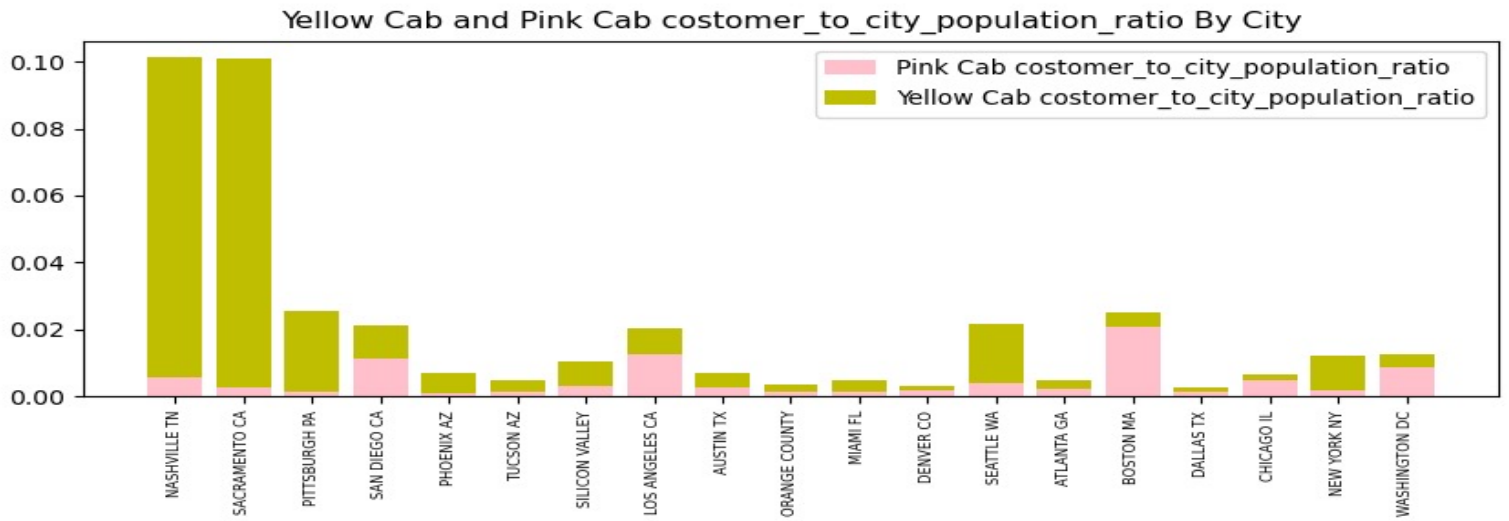
EDA – Customer Retention





EDA – Market Share





Company	Pink Cab	Yellow Cab
Customer ID_count	84711.0000	274681.0000
profit_sum	5307328.3210	44020373.1708
profit_mean	62.6522	160.2600
Age_mean	35.3224	35.3411
Income (USD/Month)_mean	15059.0471	15045.6698
Price Charged_mean	310.8009	458.1820
Cost of Trip_mean	248.1487	297.9220

EDA Summary

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

- **Customer Reach:** Yellow cab has higher customer reach all 19 cities. We have also observed that Yellow cab is doing good in covering other cab users as compared to Pink cab.
- **Pricing**: Yellow Cab has a better pricing strategy, we observed price distribution for Yellow Cab in different city and leads to better profit
- **Market Share**: Yellow cab has a higher market share in all 19 cities.
- **Customer Retention:** we found that Yellow cab is doing far better than Pink cab in terms of customer retention.
- **Age wise Reach :** Yellow cab has customer in all age group, and it's been observed that it's even popular in 60+ age group as equally as its in 18-25 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.
- **Income wise Reach :**Both the cabs are very popular in high and medium income class but here also Yellow cab is performing better than Pink cab in offering their services to all the three income class group (low, medium and high)

Recommendations

Based on the EDA summary previously, we will recommend Yellow cab for investment.

- Better Customer Reach
- More Effective Pricing Strategy
- Greater Market Share
- Stable Customer Loyalty
- Wider Age Range Target
- Greater Profit Margin

Thank You