
# Exploratory Data Analysis

G2M insight for Cab Investment firm

**8th Aug 2019**

# Agenda

Executive Summary

Problem Statement

Data Exploration

Approach

EDA

EDA Summary

Recommendations

# Background –G2M(cab industry) case study

XYZ is a private 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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

The analysis is divided into following parts-

- Data Exploration
- Trends and Patterns in Various Profit forms ( Average Profit, Profit Percentage and Profit per KM)
- City wise EDA
- Payment Method Analysis
- Customer Analysis

# Problem Statement

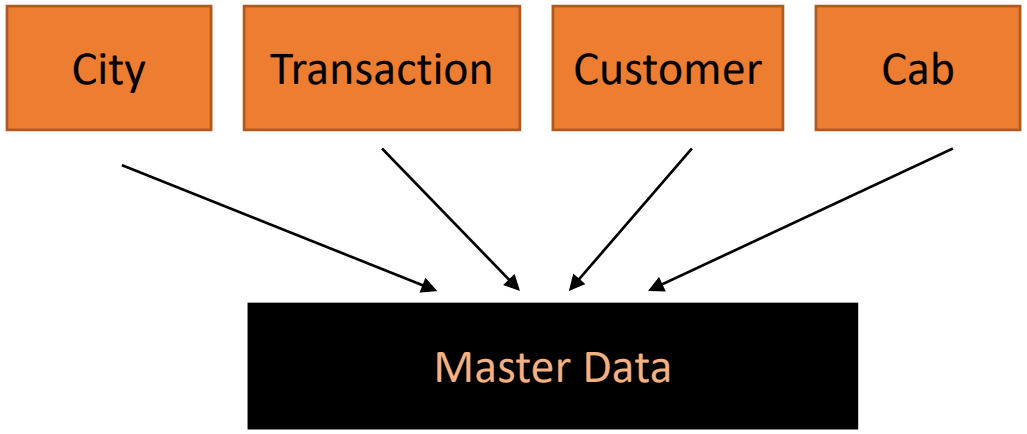
Perform EDA on given Datasets to get-

- Profit Insights
- City wise investment opportunities
- General trends in customers

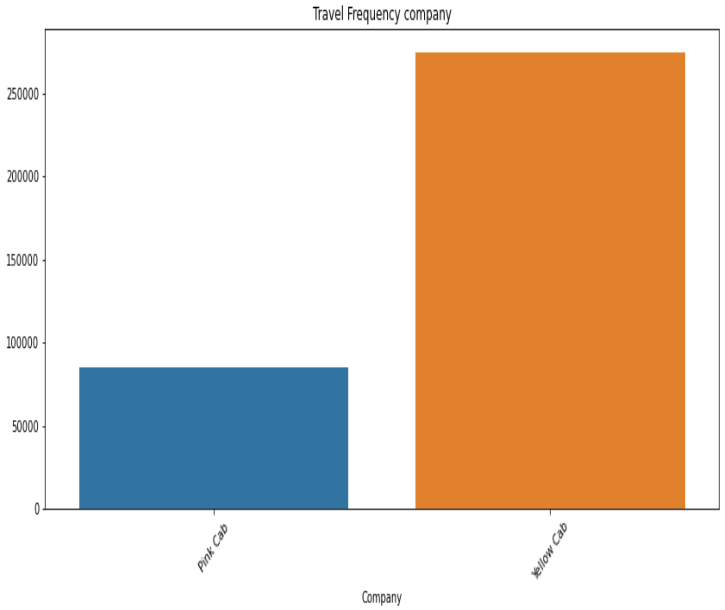
# Data Exploration

Data Glacier
Your Deep Learning Partner

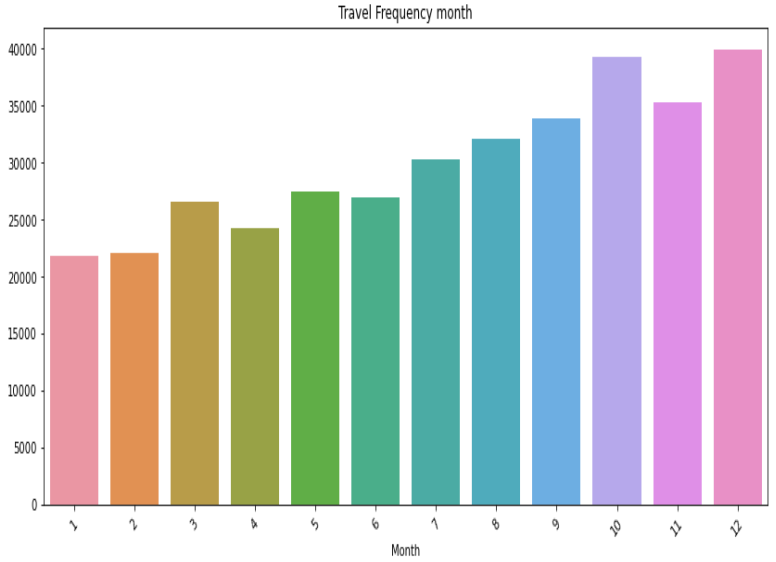
| | Number of Observations | Null Values | Derived Features | Total Features |
|---|---|---|---|---|
| **Cab_Data** | 359391 | None | 4 | 10 |
| **Customer_ID.** | 49170 | None | 0 | 4 |
| **Transaction_ID** | 440097 | None | 0 | 3 |
| **City** | 20 | None | 1 | 4 |
| Master Data | 359392 | None | 5 | 15 |



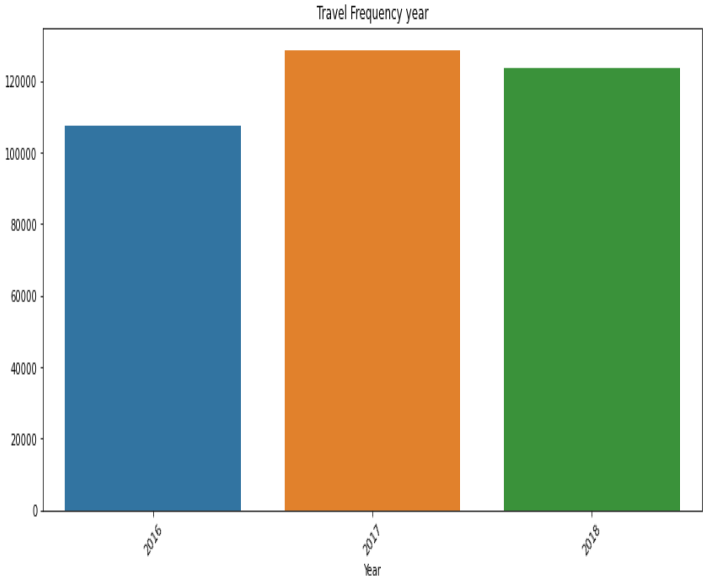
# General Trends




- Yellow Cab is used more than Pink Cab.
- 76.4% of the rides are by Yellow Cab

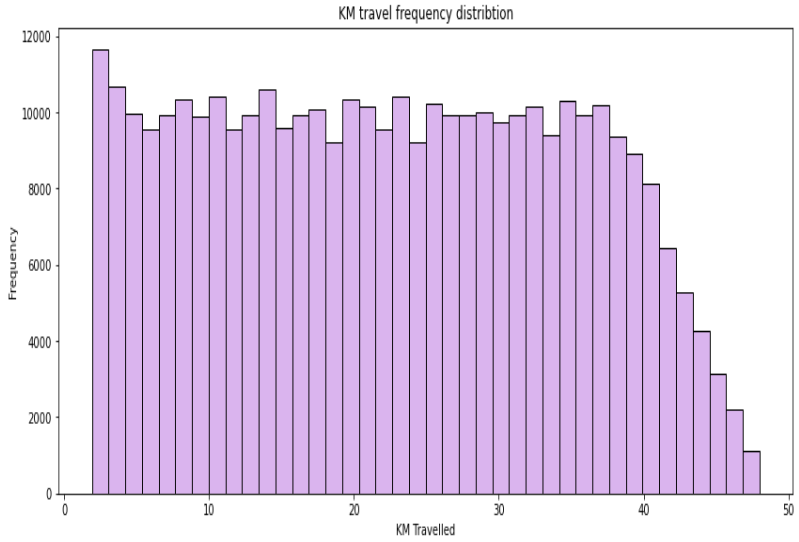


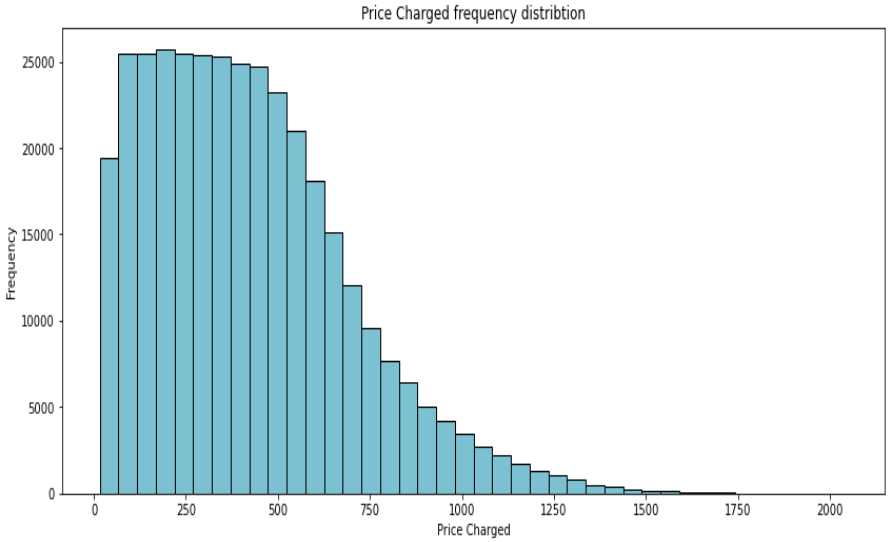

- October and December have slightly higher frequency of cab rides then other months, with January and February being the lowest

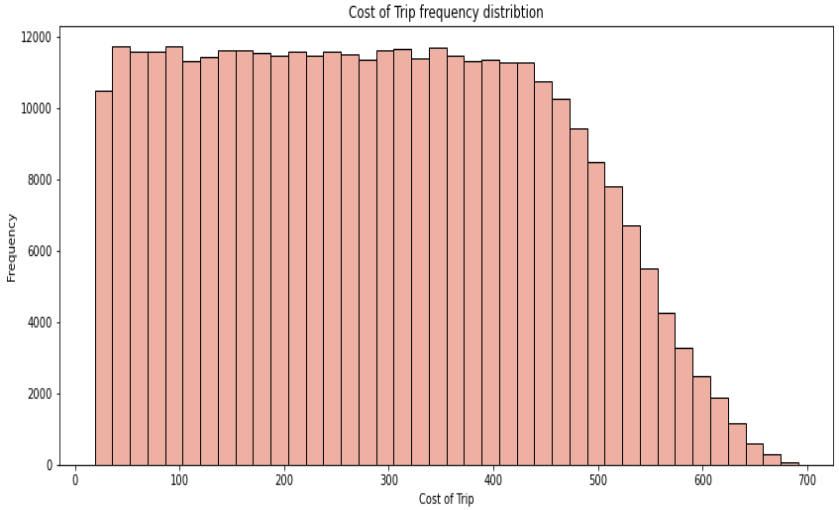



- 2017 has the highest cab rides and the 'Year'' column approximately has fairly distributed data

# General Trends

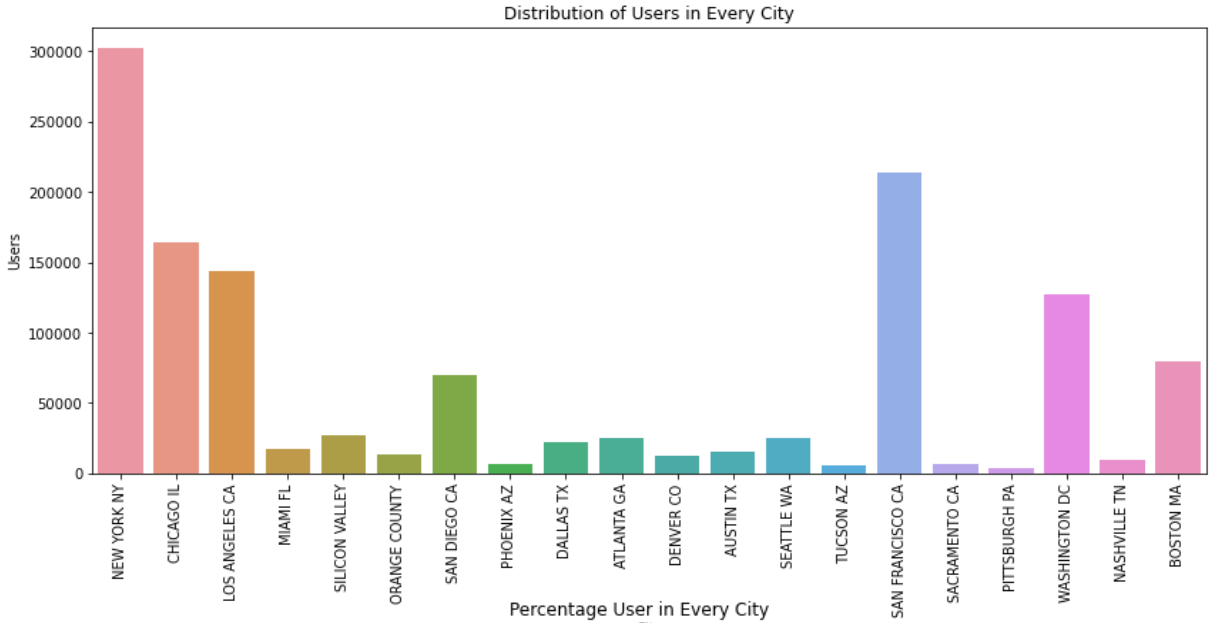


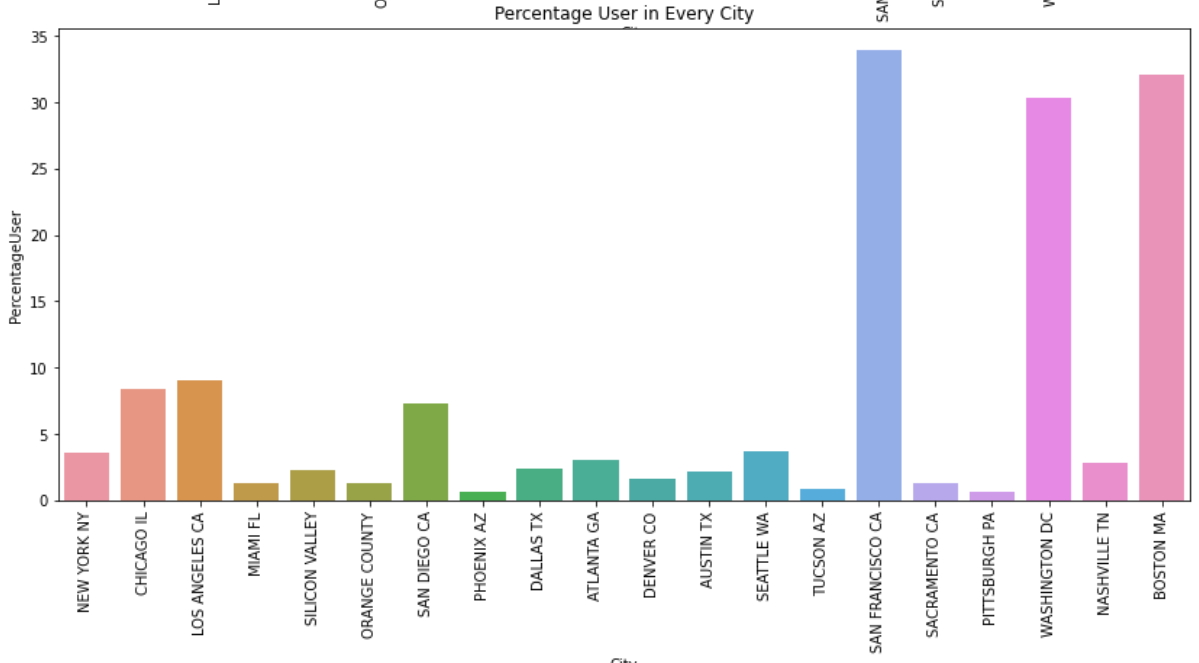




- After 40KM there is gradual linear decrease in number of rides
- After 400, there is an exponential decrease in number of rides
- After around 420 there is linear decrease in number of rides till it reaches 0 around 700

# City Analysis



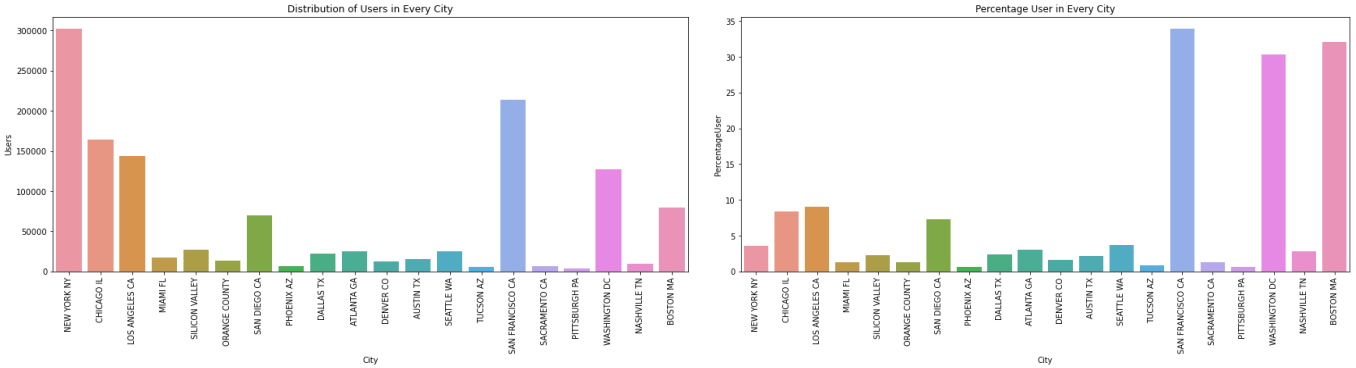




*(Percentage users = user/population)*

- Even if NY has maximum users, San Francisco has the highest percentage of users.
- Cities with high number of users but low percentage users can be a potential market for investment opportunities (E.g. NY, Chicago and LA).
- Cities with both low users and percentage users, would require more focus since people there in general don't travel via cab.

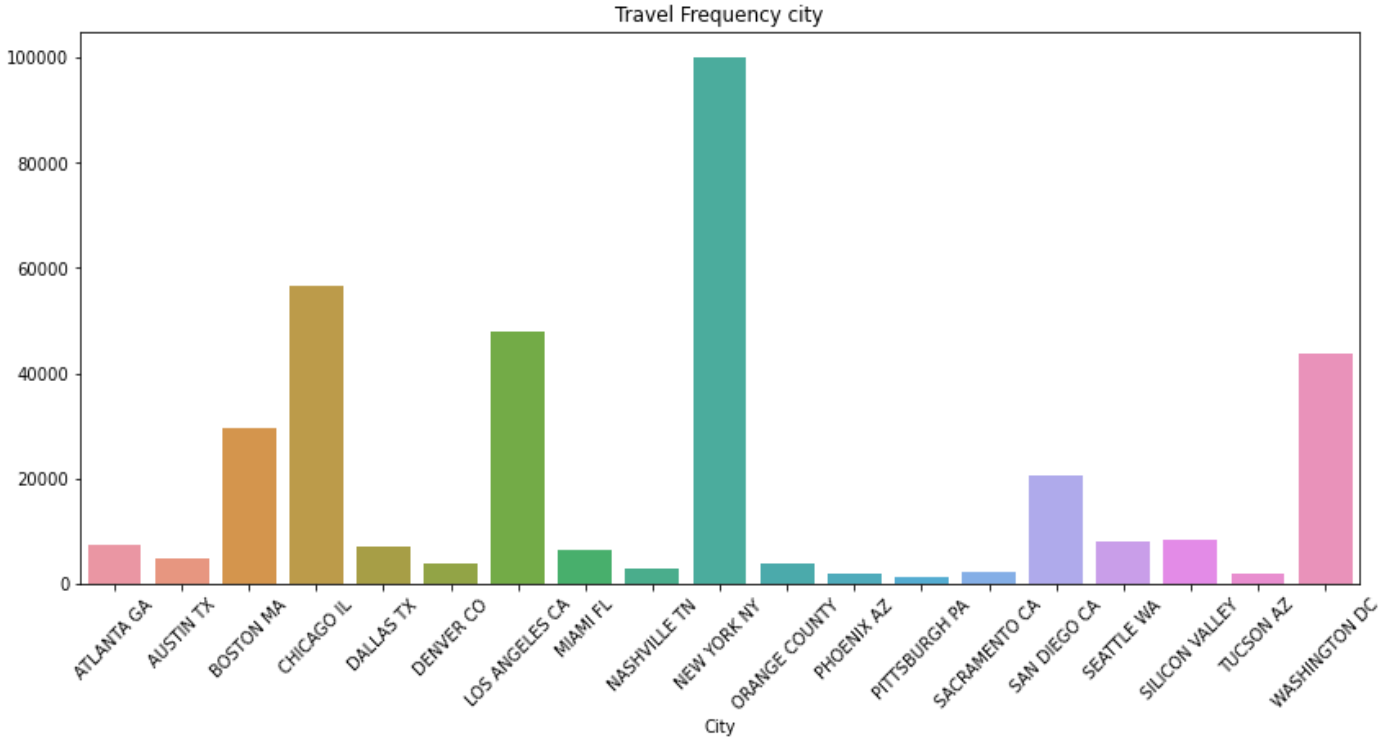
*Graph1: No of users vs City*
*Graph2: Percentage user vs City*

# City









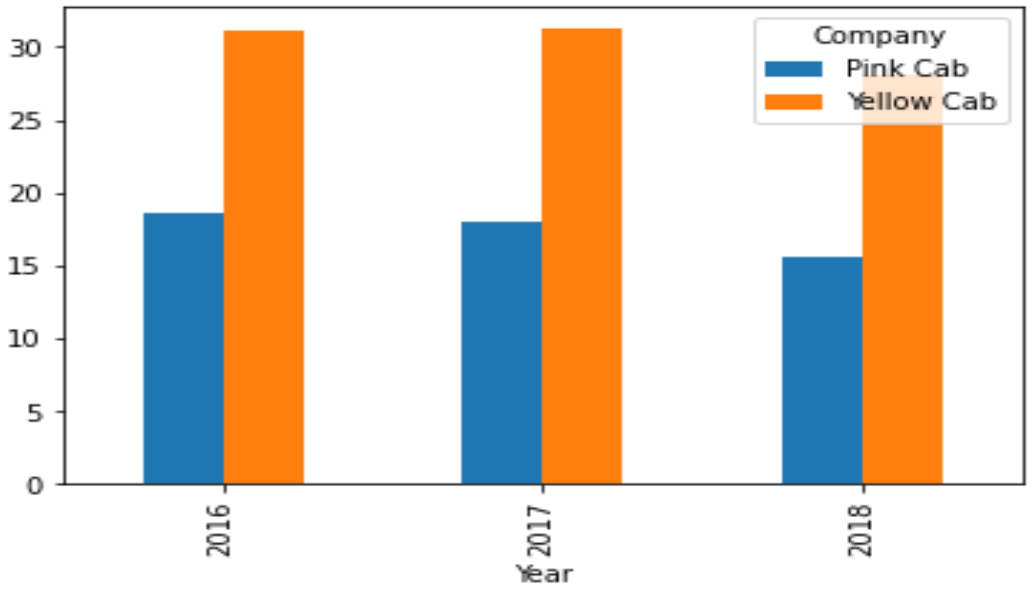

*(above two graphs are from the previous slide)*

- New York Contributes most in number of cab rides.
- Cities like Boston, Chicago and Washington are potential enough to generate greater travel frequency
- Even if San Francisco contributes towards users we don't have any travel records in San Francisco.

*Graph: Total Number of Rides vs City*

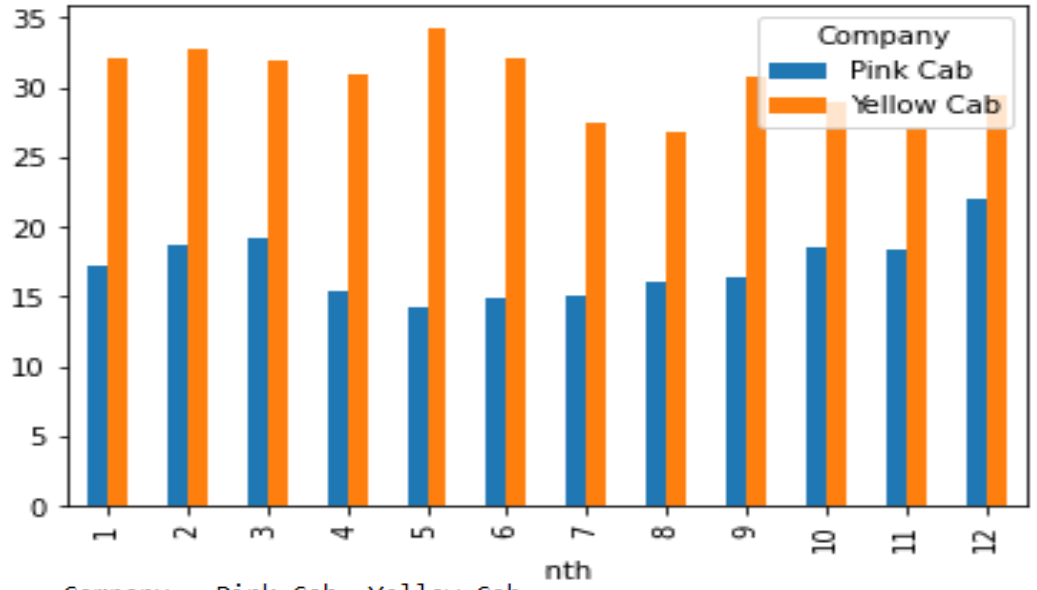
# Profit Analysis

- Year vs Profit Percentage



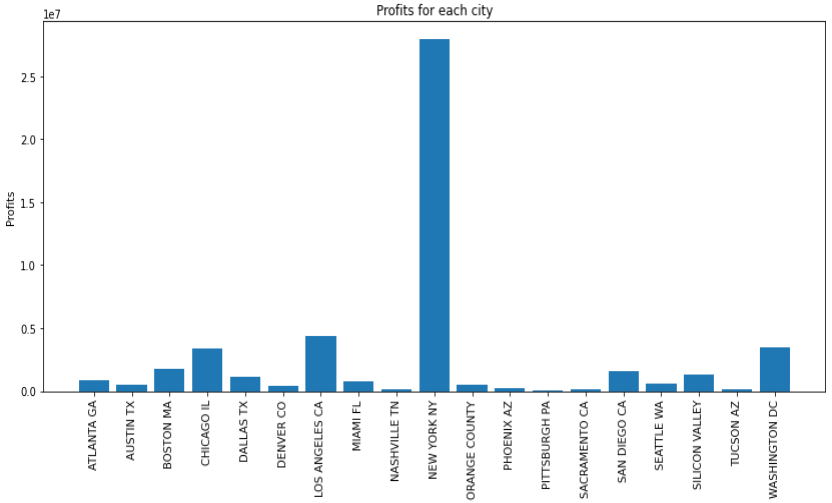

| Company | Pink Cab | Yellow Cab |
|---|---|---|
| Year | | |
| 2016 | 18.583079 | 31.201838 |
| 2017 | 18.067983 | 31.246804 |
| 2018 | 15.641241 | 28.044586 |

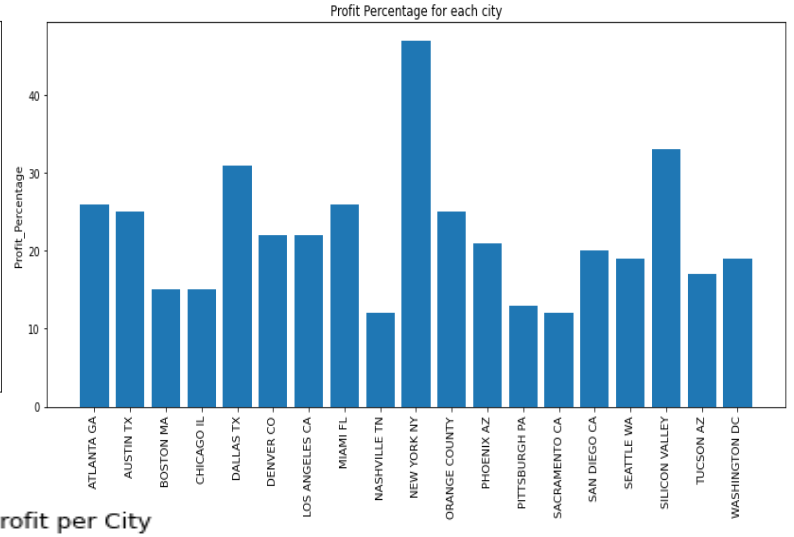
- Months vs Profit Percentage

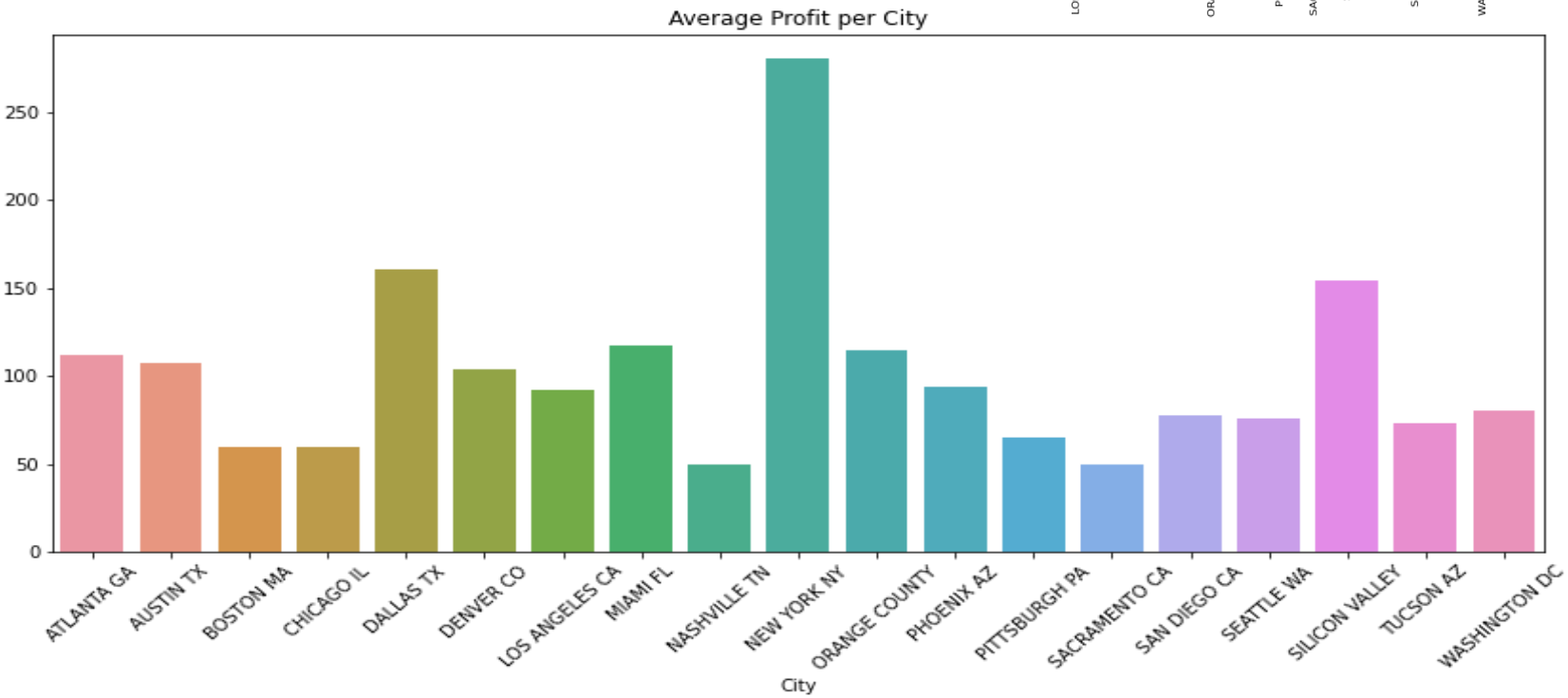



| Company | Pink Cab | Yellow Cab |
|---|---|---|
| Month | | |
| 1 | 17.213460 | 32.165921 |
| 2 | 18.671782 | 32.681342 |
| 3 | 19.214736 | 31.846191 |
| 4 | 15.310962 | 30.932702 |
| 5 | 14.179417 | 34.248129 |
| 6 | 14.852102 | 32.174926 |
| 7 | 14.985581 | 27.461039 |
| 8 | 16.129198 | 26.778640 |
| 9 | 16.418108 | 30.826651 |
| 10 | 18.595578 | 28.906862 |
| 11 | 18.358509 | 27.118507 |
| 12 | 22.025052 | 29.528205 |

# City



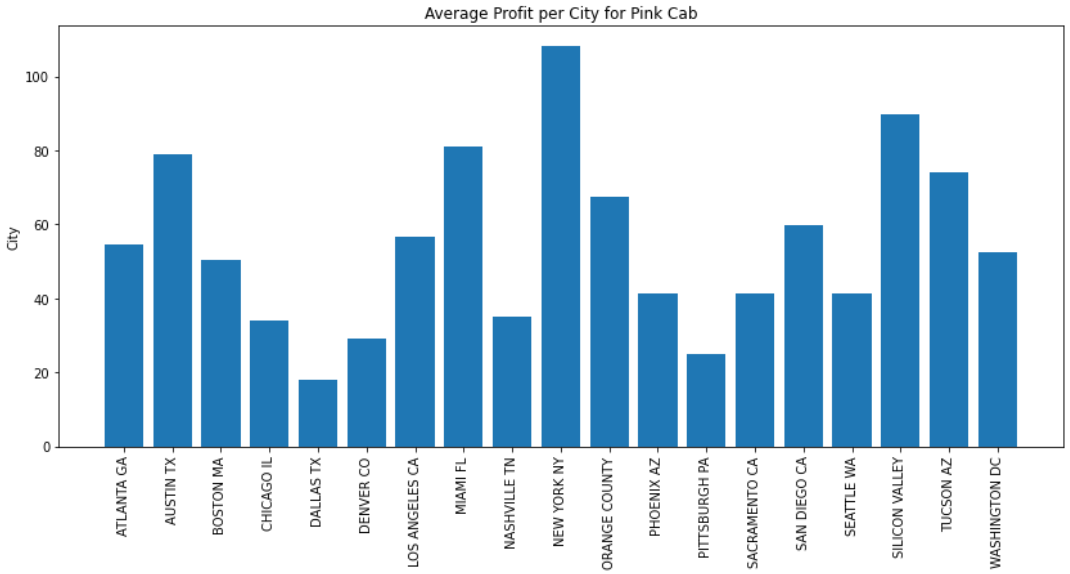



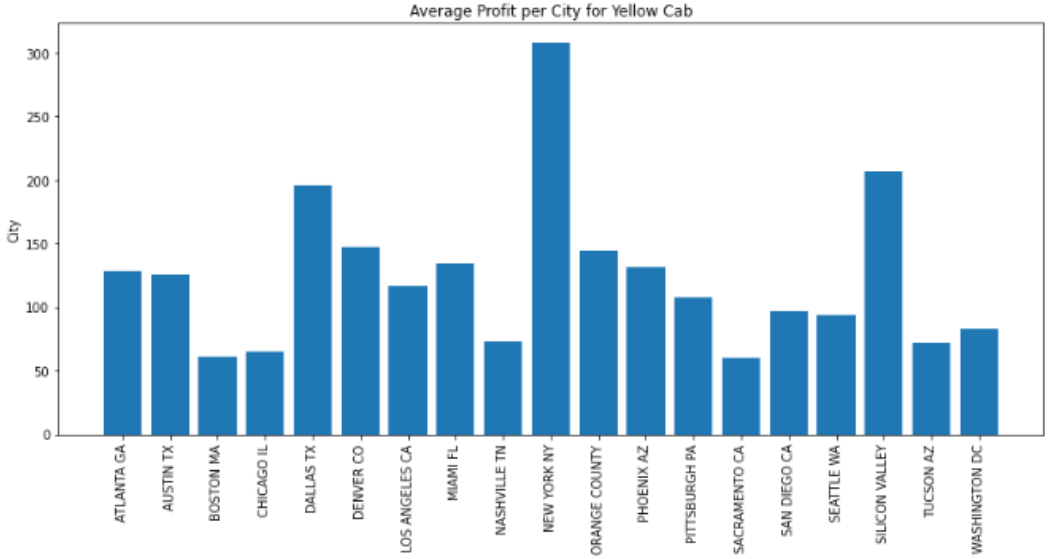

*(Also refer graph from slide 8 – Travel Frequency)*

- *Dallas and Silicon Valley has comparatively lower travel frequency but higher average profit*
- *Profit and average profit have relatively different graphs*
- *Travel Frequency Profit ,Average Profit and Profit Percentage have different impacts on different cities*
- *Trends in Profit Percentage and Average Profit graphs are very similar*

*Graph1: (from left) Total Profit vs City*
*Graph2: Sum of Profit Percentage vs City*
*Graph 3: (below)Average Profit Vs City*

# Profit Analysis

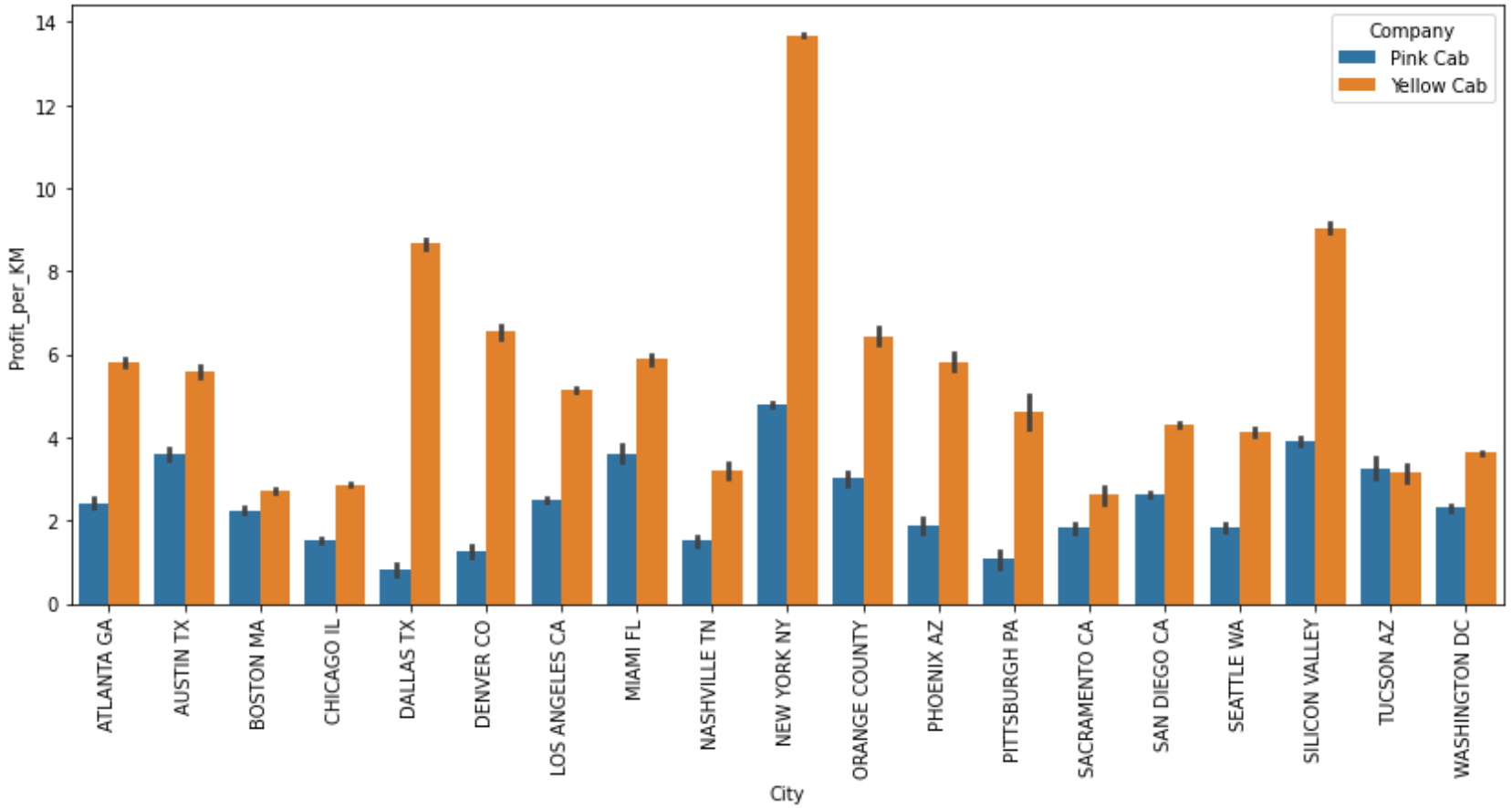



- Highest Average Profit for Pink City is in New York - 100 , and for Yellow Cab it is 300
- For Pink Cab lowest average profit is in the city of Dallas. Yellow Cab has lowest average profit is in the city of Boston
- Overall Average profit for Pink Cab – 54.6
- Overall Average profit for Yellow Cab- 123.58

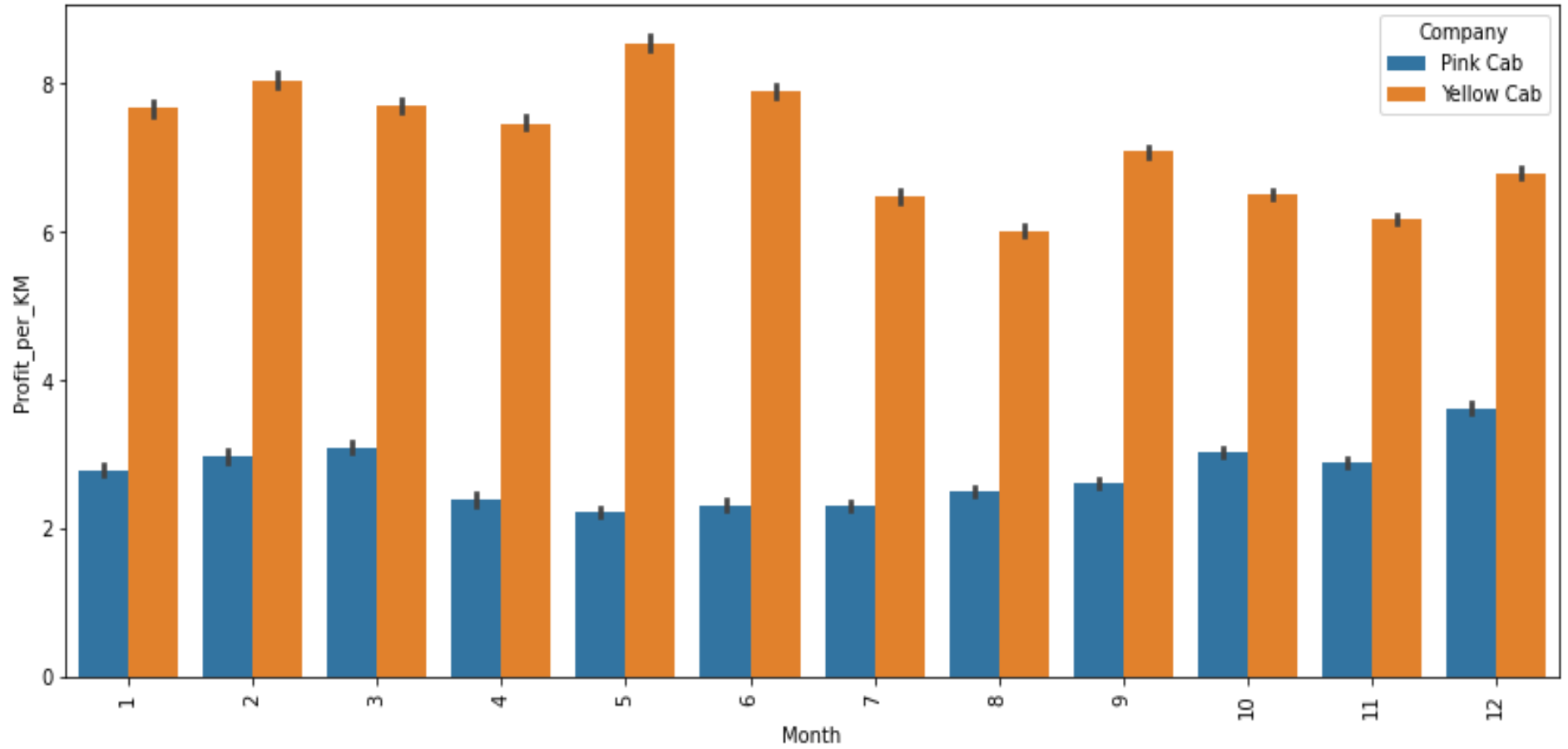
*Graph1: (from top) Average Profit Vs City for Pink Cab*
*Graph2: Average Profit Vs City for Yellow Cab*

# Profit Analysis



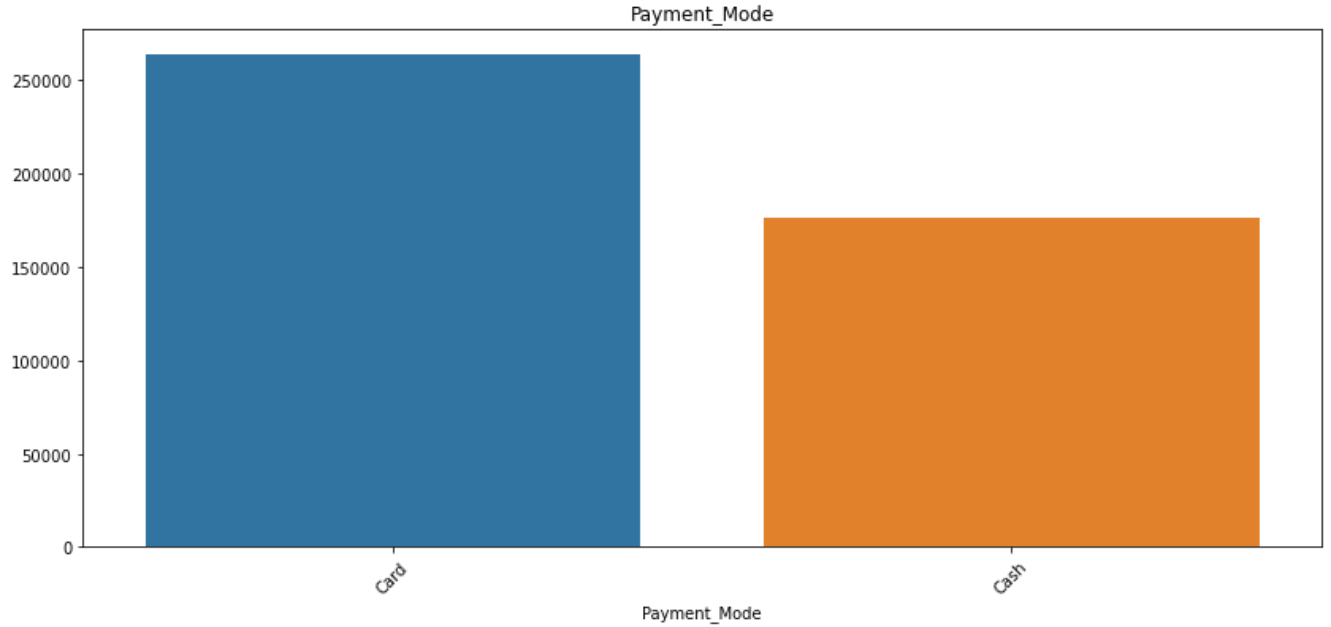

- Following graph represents Profit per KM vs City for both the company
- Dallas, Denver and New York have higher gap for profit per km between the two companies
- Tucson is the only city with Profit per KM is higher for Pink Cab

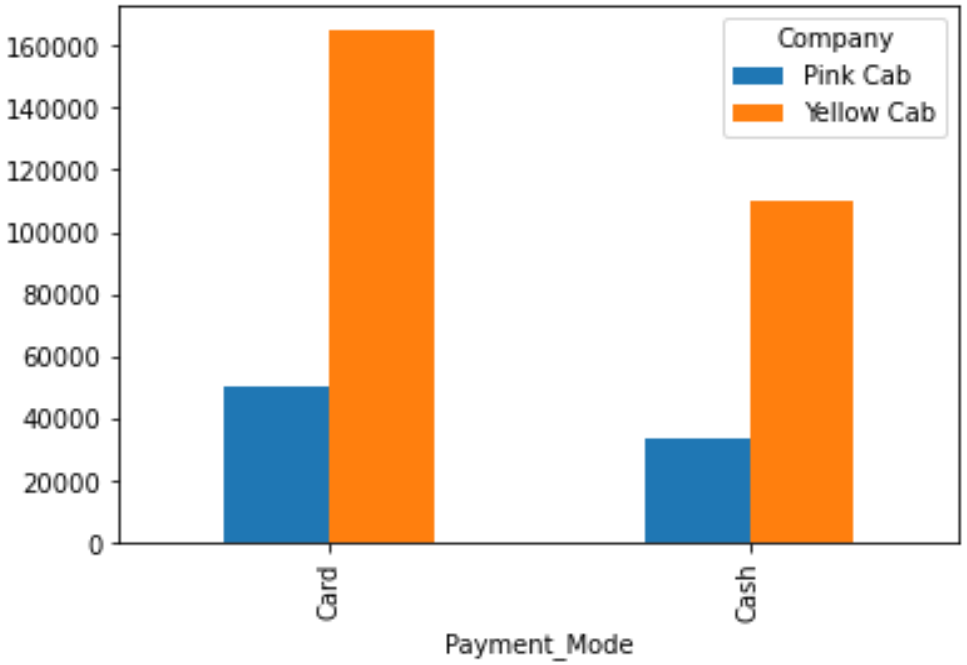
# Profit Analysis




The month of May has slightly higher profit per KM while August and November have comparatively lower profit per km

# Payment Mode



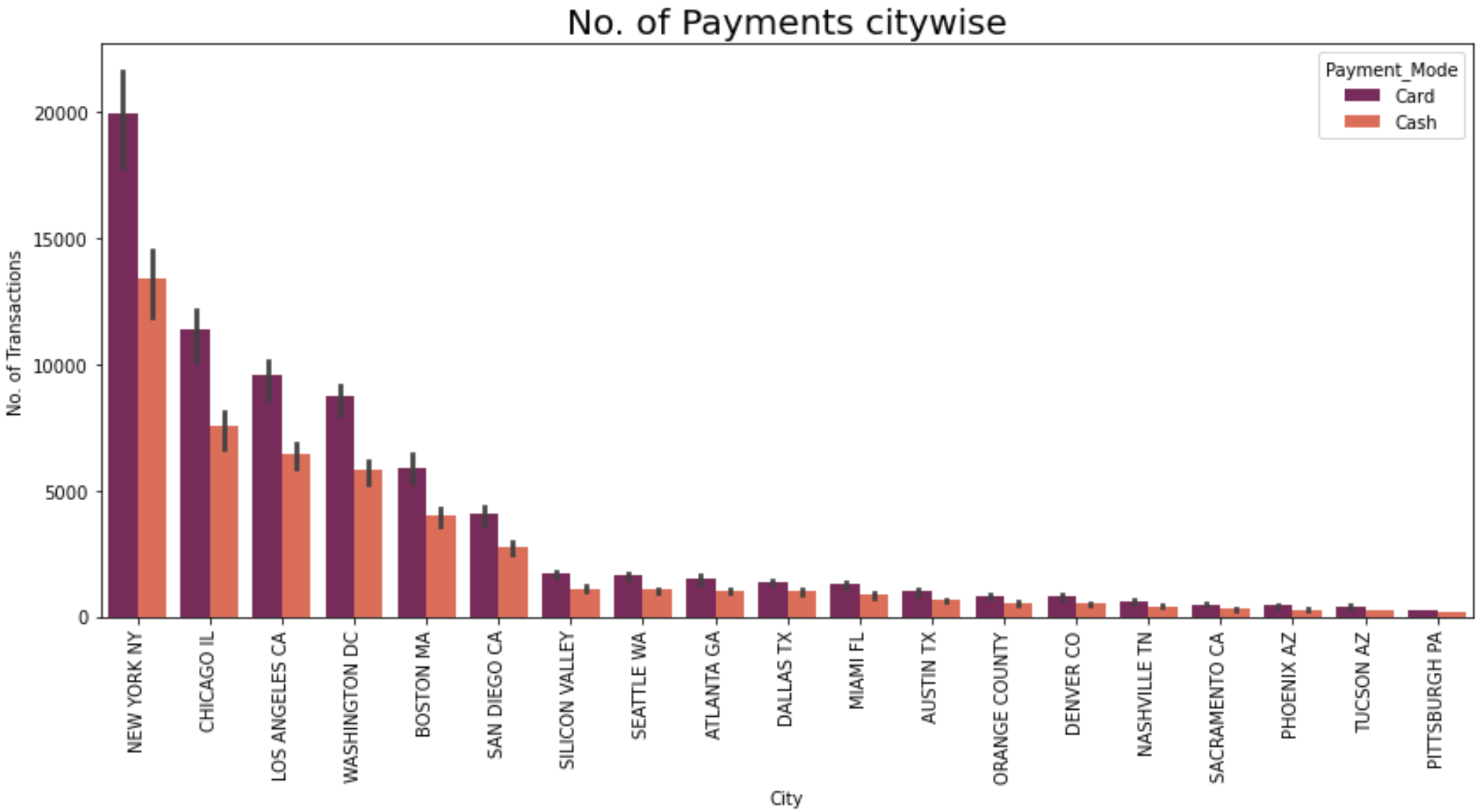




| Payment_Mode | |
|---|---|
| Card | 263991 |
| Cash | 176107 |

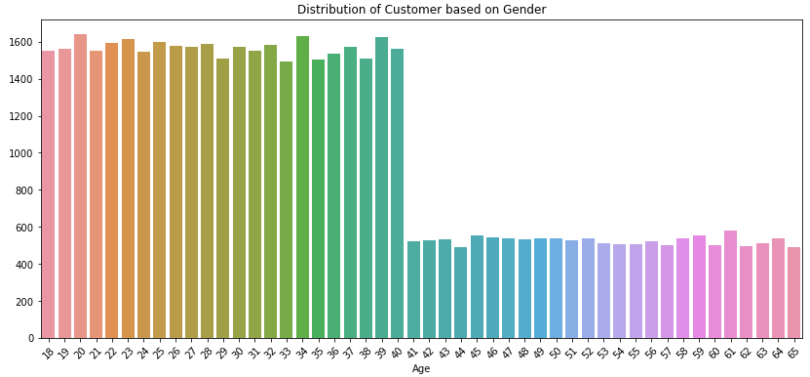
| Company | Pink Cab | Yellow Cab |
|---|---|---|
| Payment_Mode | | |
| Card | 50719 | 164785 |
| Cash | 33992 | 109896 |

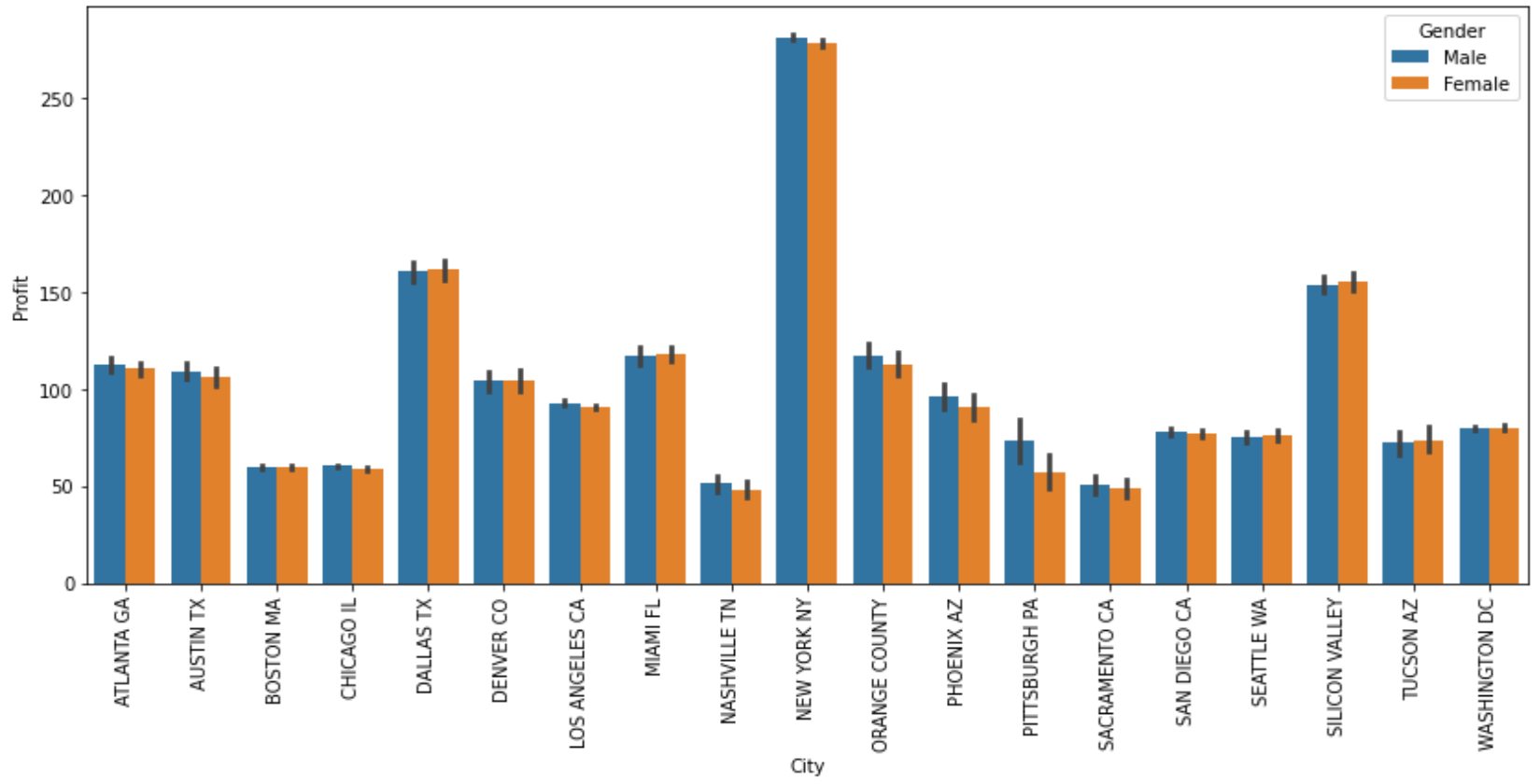
# Payment Analysis

Data Glacier
Your Deep Learning Partner



# Client Analysis

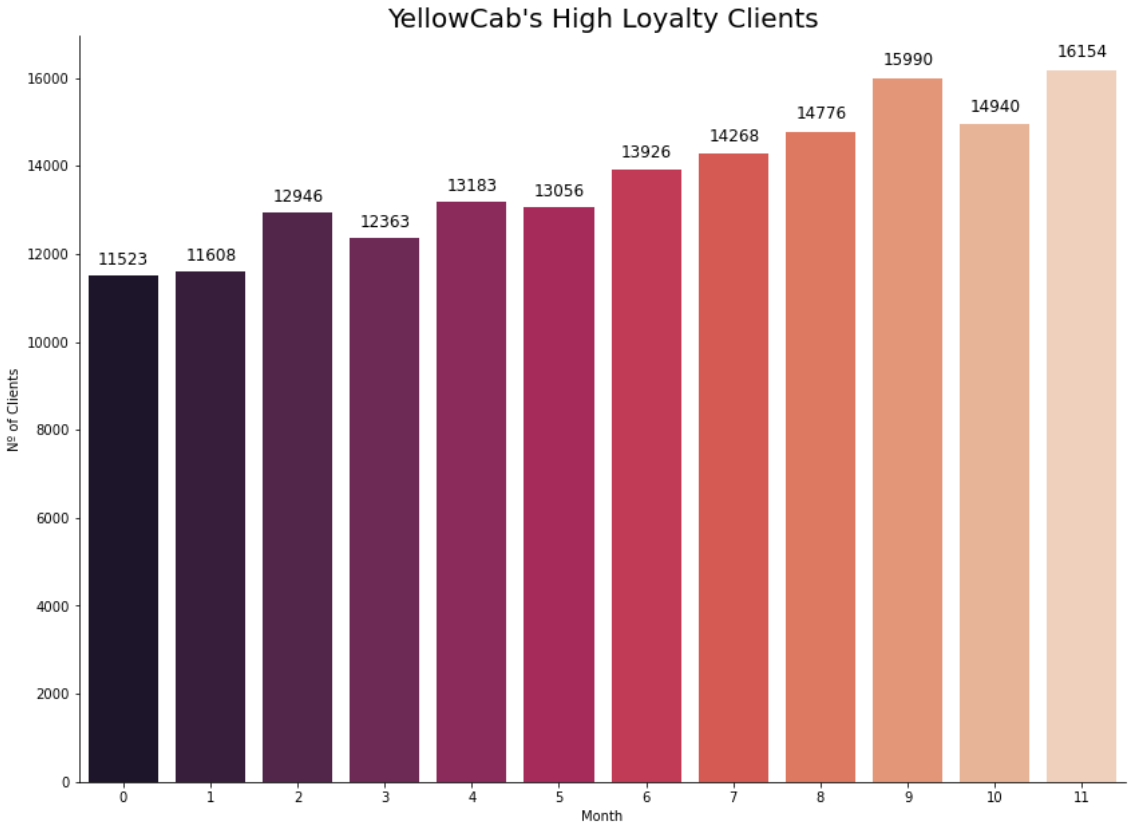



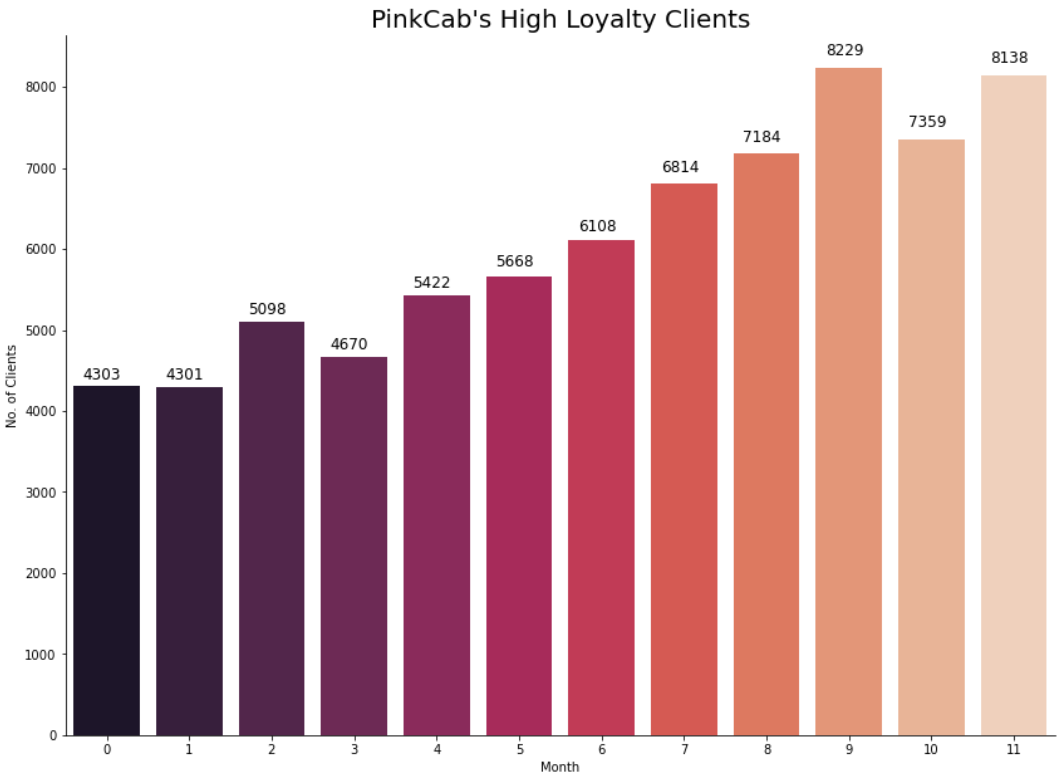




- There are slightly more number of male customers vs female customers
- After the age of 40 there is sudden decrease in customers
- The trends in the third graph are similar to that of Avg Profit vs City graph in slide 11

*Graph1: (from left)Total number of Customers Vs Gender*
*Graph2: Total number of Customers Vs Age*
*Graph 3: (below)Average Profit Vs City (for Genders)*

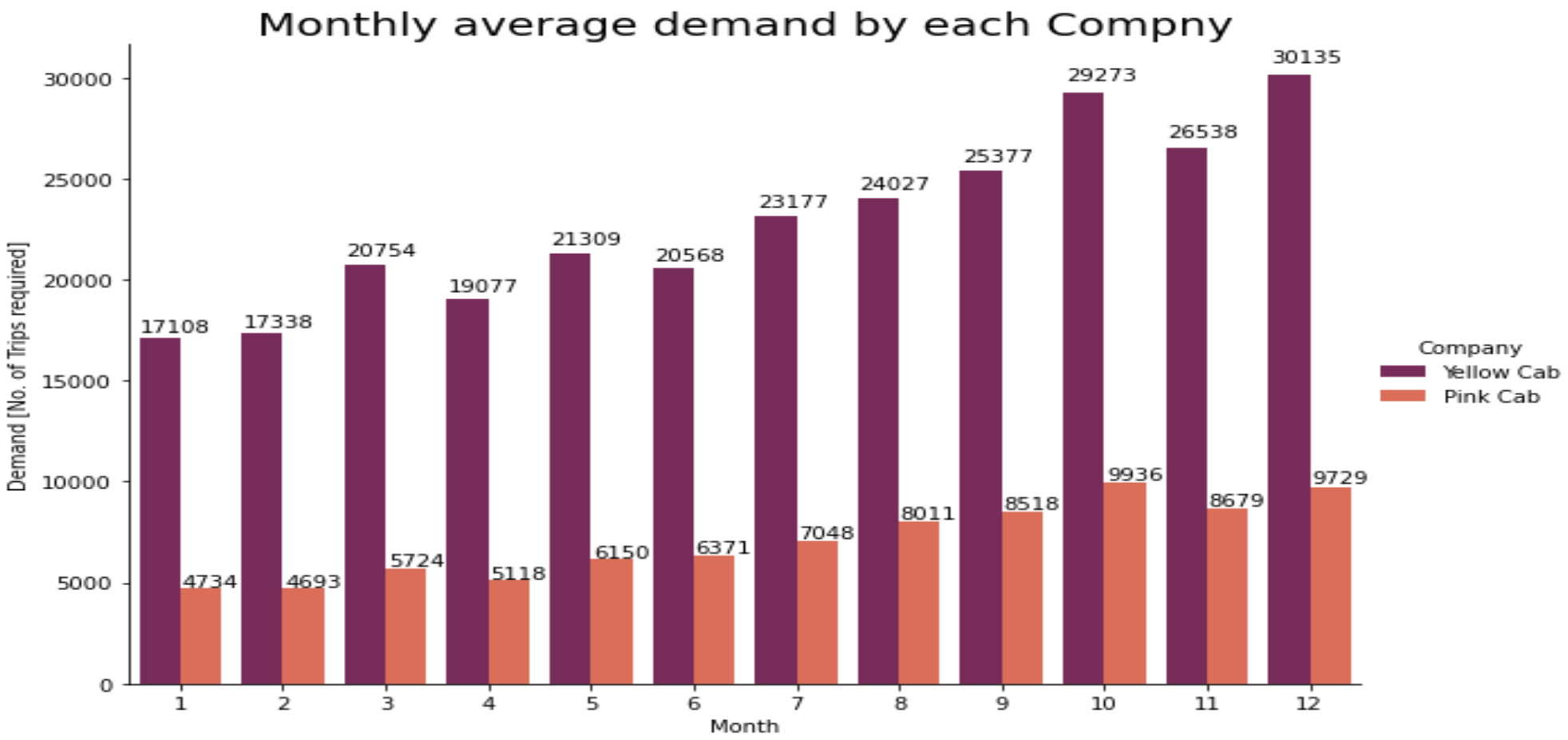
# Client Analysis




Both the Cab companies have similar trends in Number of Rides by Loyalty Clients per Months

# Client Analysis




- October and December receives highest customer demands while January and February receives the lowest

# Summary

Data Glacier
Your Deep Learning Partner

On Average Yellow Cab yields more Profit than Pink Cab

Months of October and December have highest rides and profit yields, while we can increase engagement and number of rides in Jan and Feb by giving discounts.

Profit per KM could be increased for Oct and Dec.

New York is a safe city to invest whereas cities like Boston Chicago and  Washington are potential markets

Special scheme can be introduced for people above the age of 40 to encourage them to use cabs more often

Even if number of Males Customers are greater than Females, both the groups generate approximately same average profit trends across cities

We should focus more on cities with comparatively large number of users and lesser user percentage.

# Thank You