Exploratory Data Analysis

G2M insight for Cab Investment firm

22 June 2021

Agenda

Data Glacier
Your Deep Learning Partner

EXECUTIVE SUMMARY

❑ 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.

❑ As per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

PROBLEM STATEMENT

Identify which cab company is performing better and is a better investment for XYZ. The two cab companies to choose from are:

➢ *Pink Cab*

➢ *Yellow Cab*

APPROACH

The analysis include:

- Data Understanding

- Data Visualization

- Recommendation for Investment

Understanding the Data

There are 4 data sets to be used for analysis:

- **Cab_Data.csv –** this file includes details of transaction for 2 cab companies
- **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
- **Transaction_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode
- **City.csv –** this file contains list of US cities, their population and number of cab users

❑The time period for which we perform EDA is 31-01-2016 to 31-12-2018 so the data is filtered accordingly.

❑Profit was calculated as Price Charged – Cost of Trip

❑It was assumed that some customers received discounts on trips hence negative values of profit were kept.

❑The data was joined before visualizing.

❑There were no NULL or duplicate values.

EXPLORATORY DATA ANALYSIS
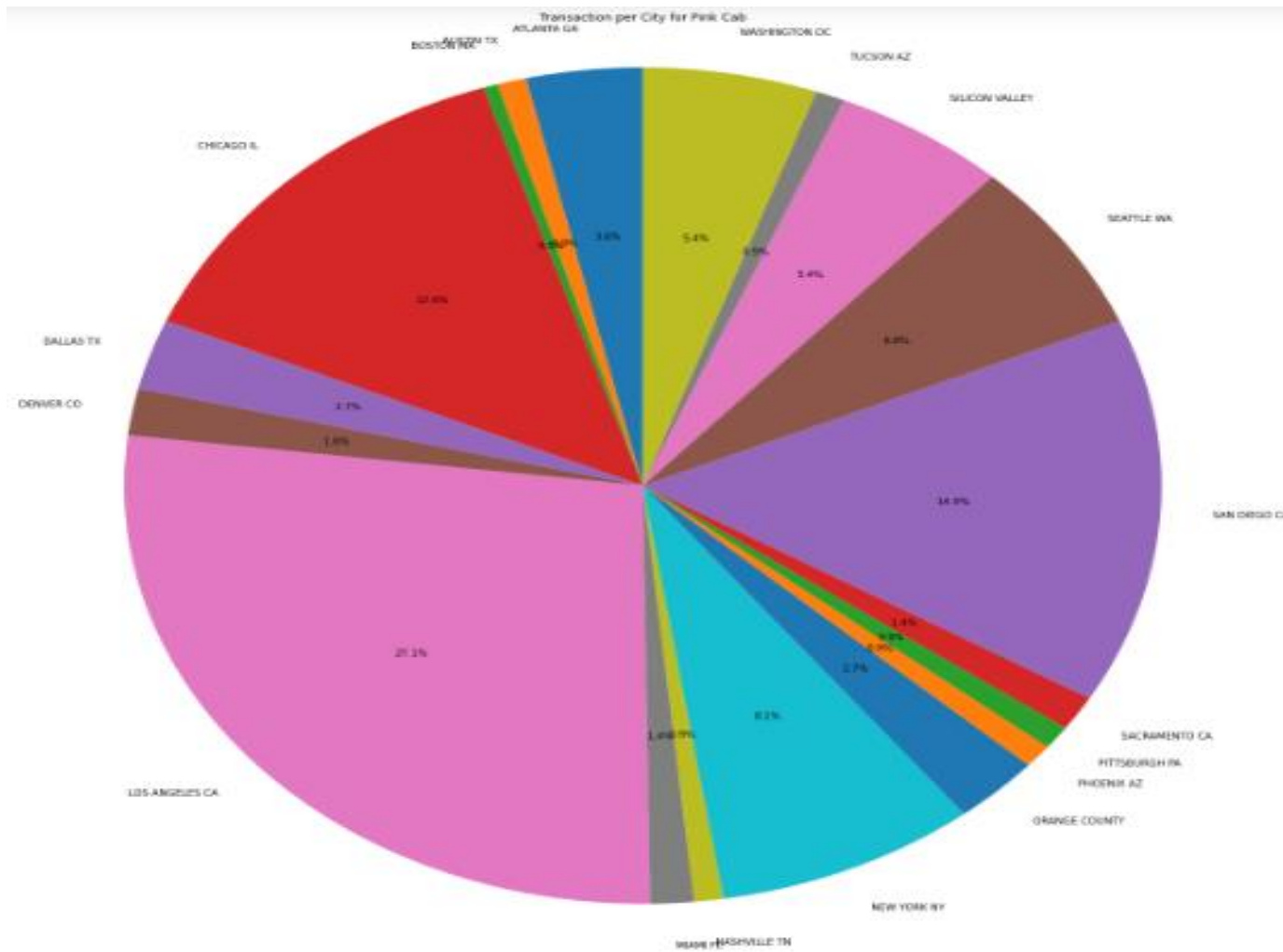
Data Visualization

Cab Users per City
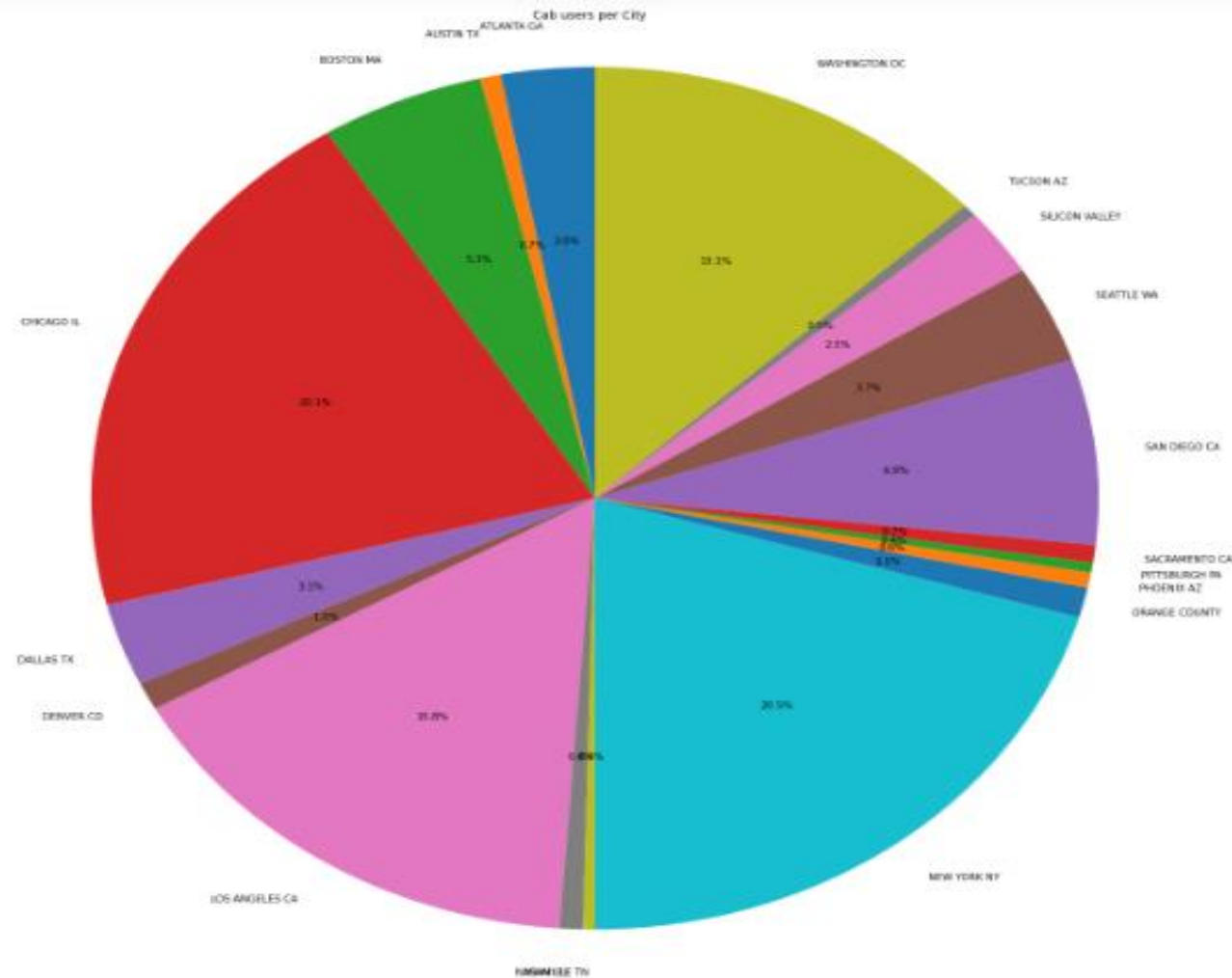
It is assumed that cab users include other cab companies and not just Pink Cab and Yellow Cab

The highest number of cab users are in:
❖ New York(20.5%)
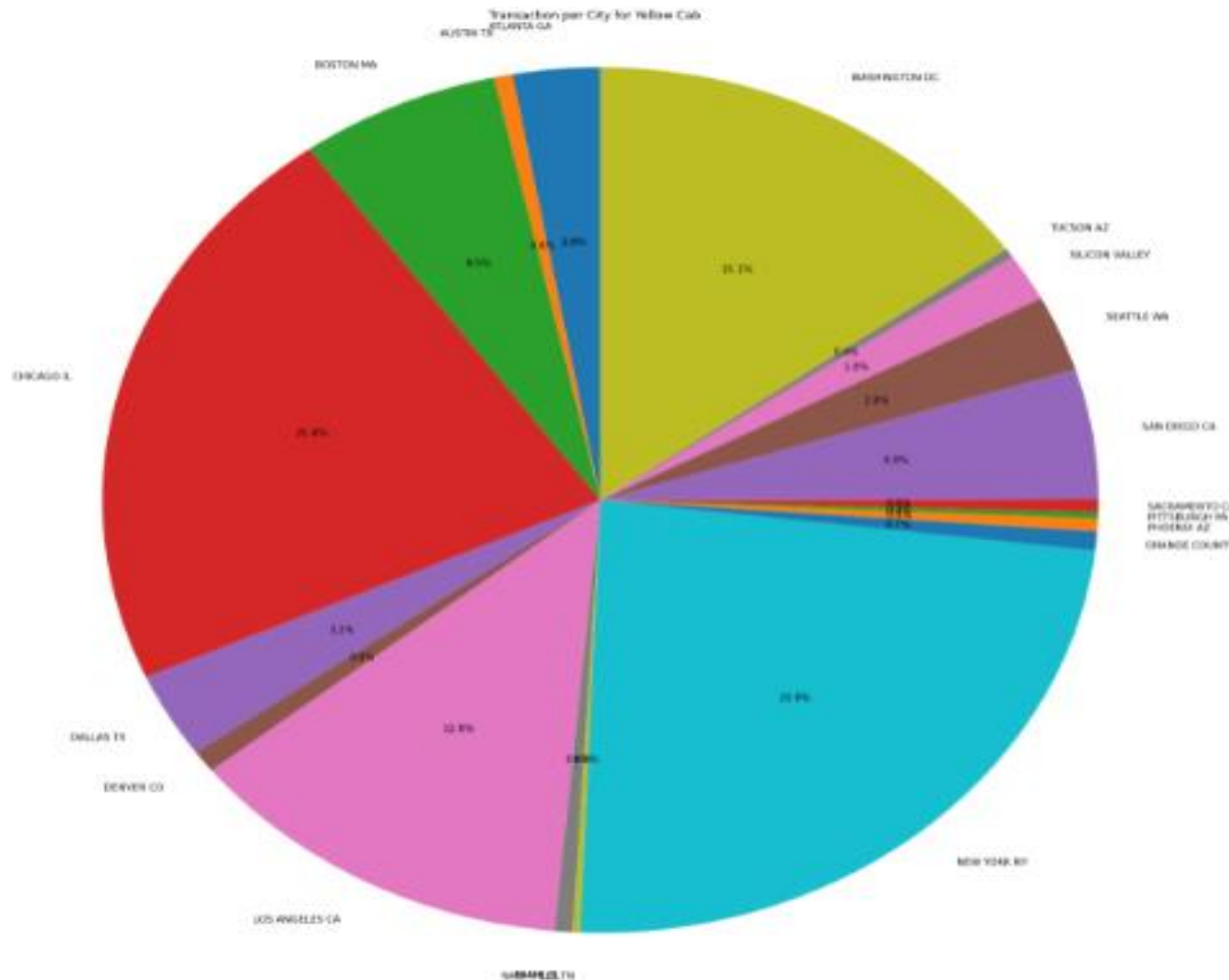❖ Chicago(20.1%)
❖ Los Angeles(15.8%)

Transaction per City for Pink Cab

Customers for Pink Cab are highest in:
- ❖ Los Angeles(27.1%)
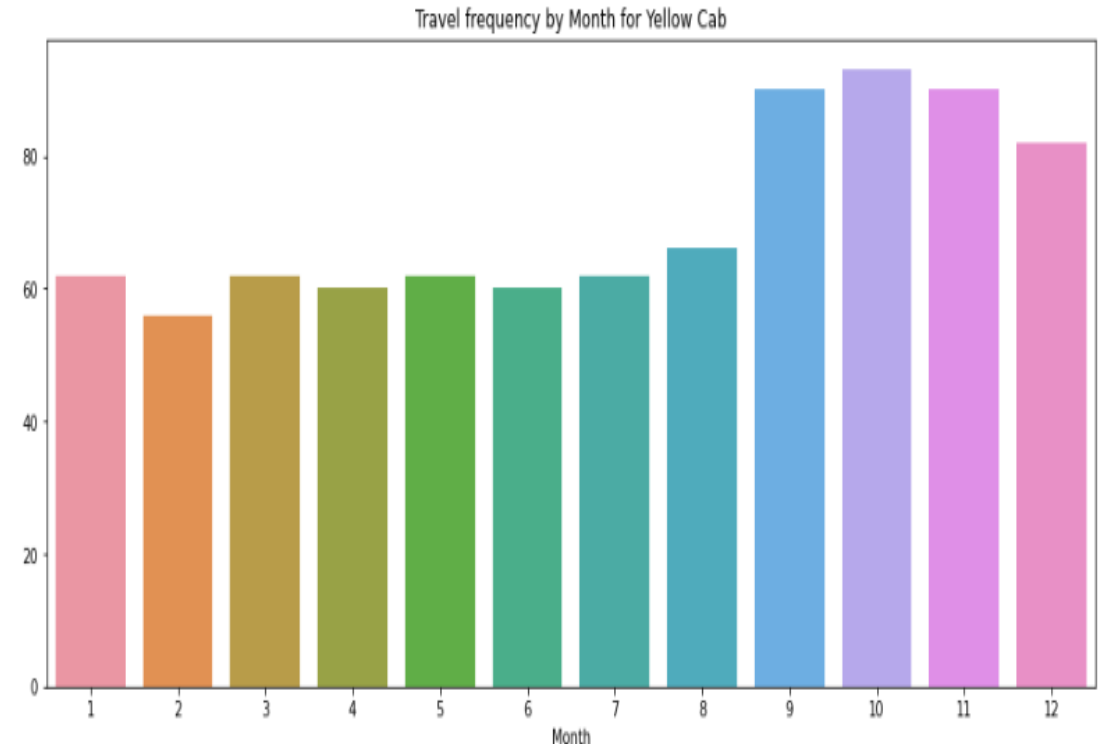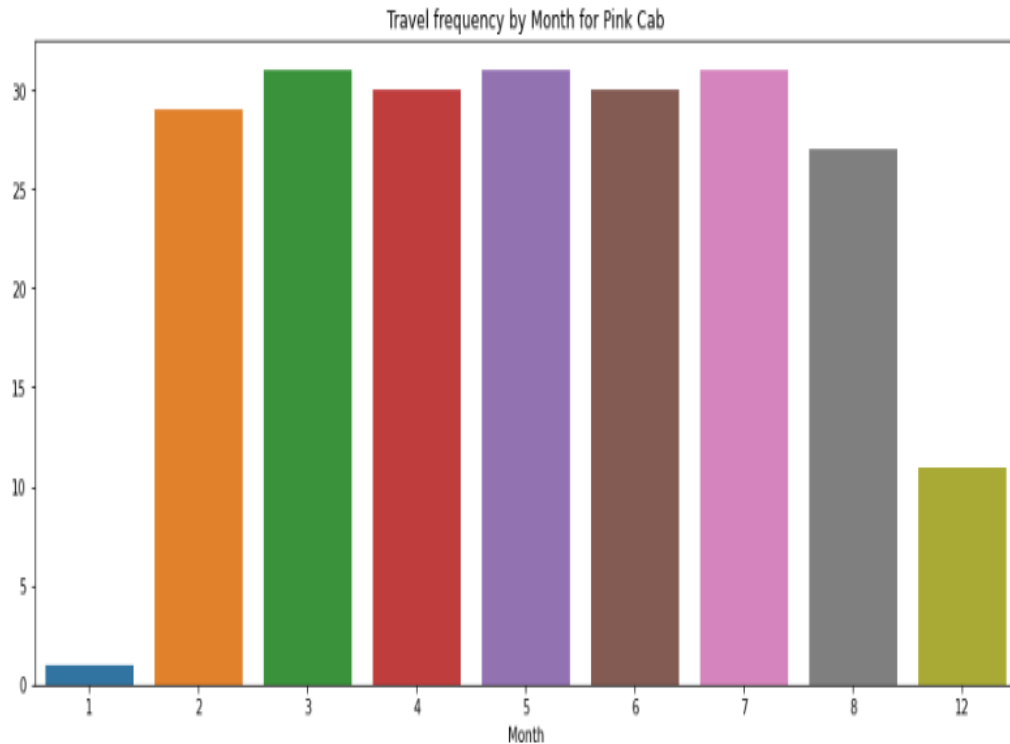- ❖ San Diego(14.9%)
- ❖ Chicago(13.6%)

Transaction per City for Yellow Cab


Customers for Yellow Cab are highest is:
- ❖ New York(23.8%)
- ❖ Chicago(21.8%)
- ❖ Washington(15.1%)

Monthly Travel Frequency for each Cab

Travel frequency by Month for Pink Cab

Travel frequency by Month for Yellow Cab

The lower frequency during the month of December and January could be due to people being on holiday and therefore staying at home or travelling outside the city(Where either cabs are not operating). Pink Cab experiences a period of no travel in the months of September, October and November and very low travels in December and January. This could be due to a shutdown in those months and time taken to recover from it. Yellow Cab has a consistent travel frequency with a rise towards the end of the year

KM Travelled vs Cost of Trip for each cab
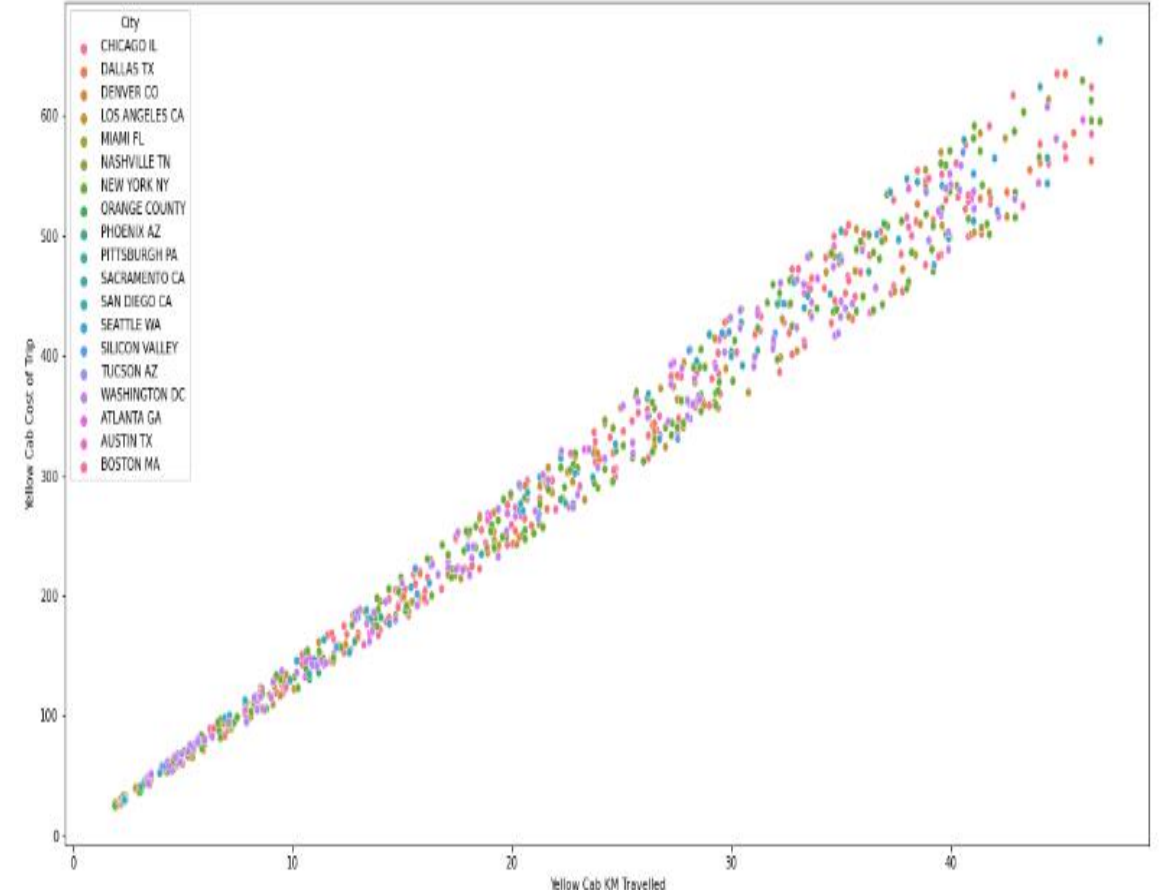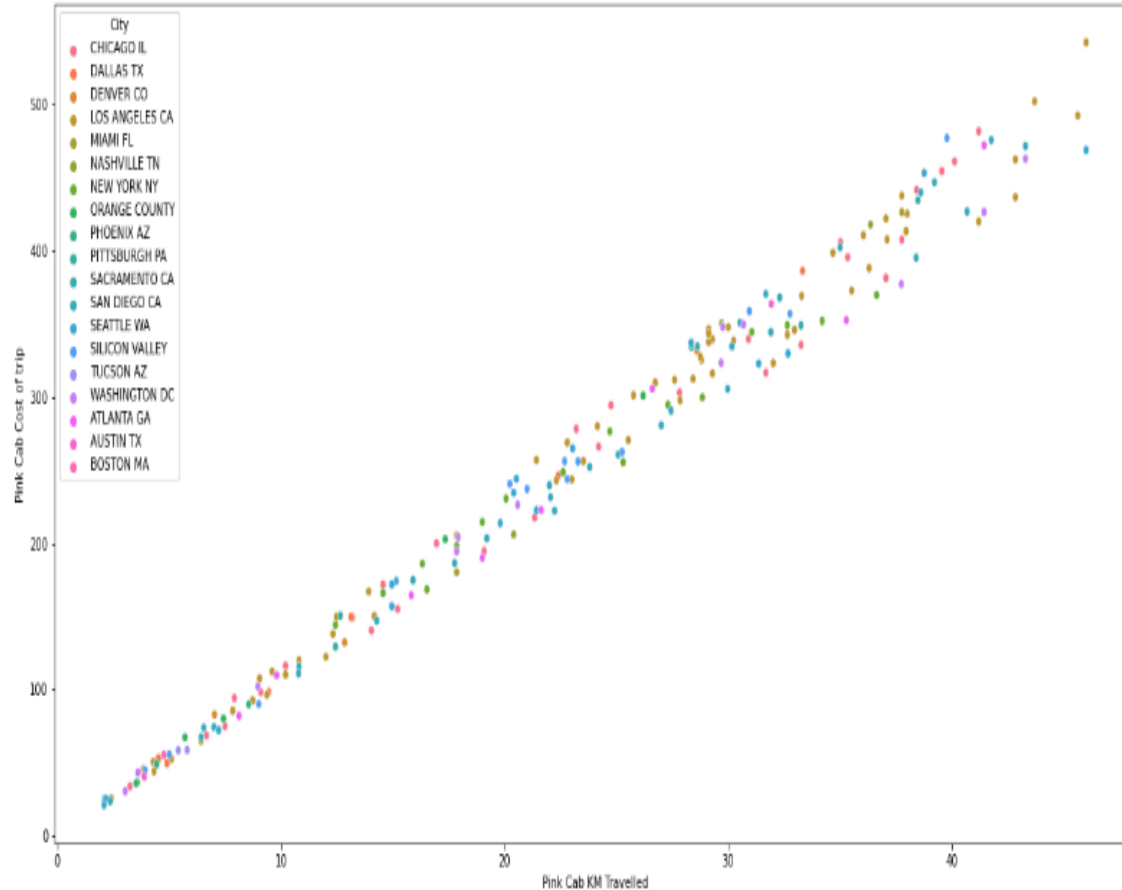
Both cabs approximately travel between 2km to 48km in all cities and the cost of trip of Yellow Cab are slightly higher. Due to the cluster in the graph for Yellow Cab we can clearly see that Yellow Cab has more users overall

# KM Travelled vs Price Charged for each cab
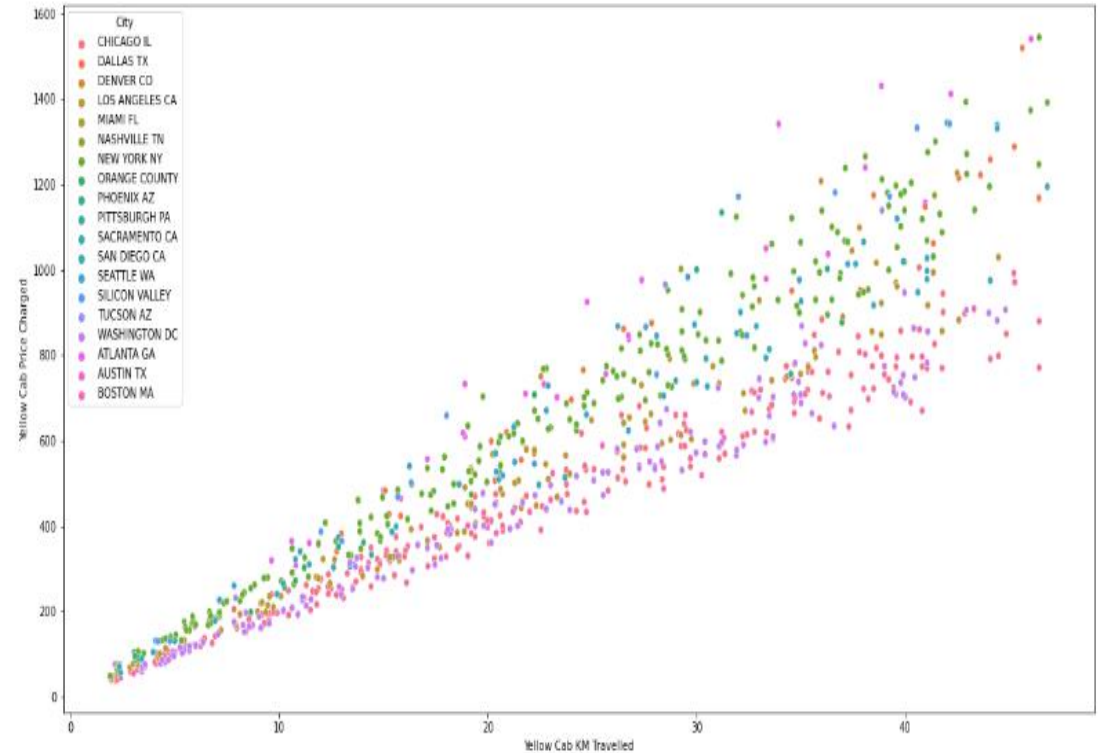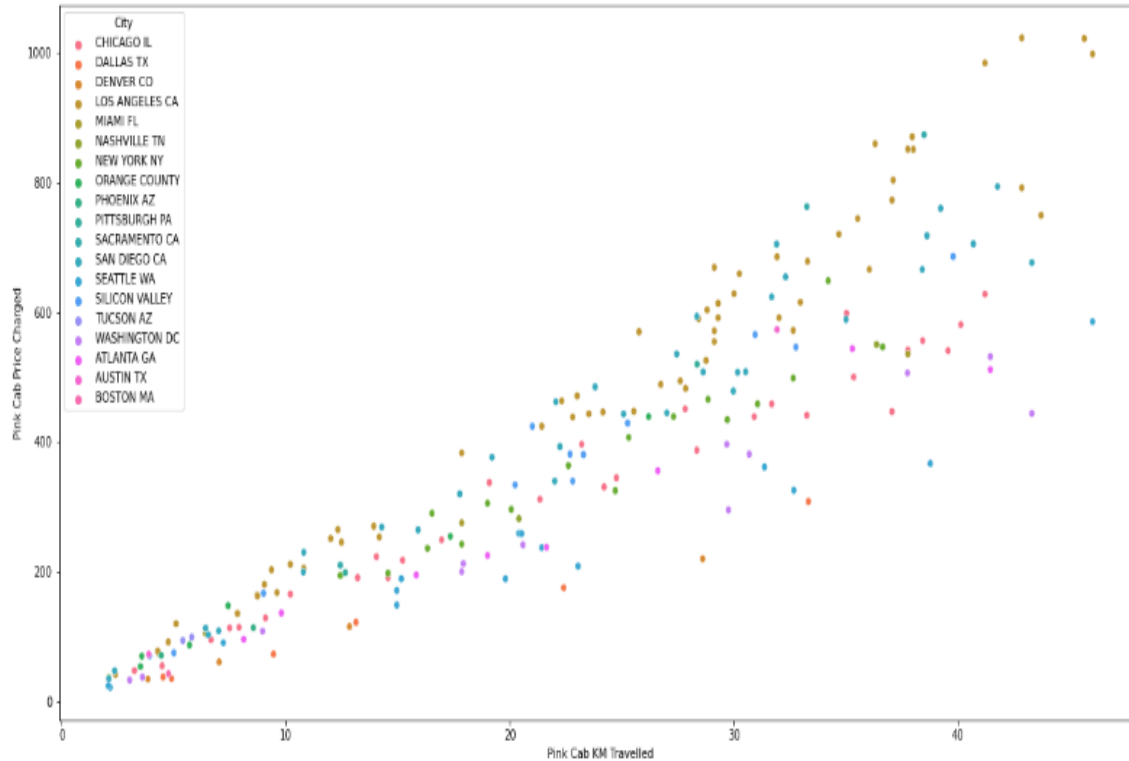
There is a positive correlation between price charged and kilometers travelled for both Pink and Yellow Cab. However, from the above two plots we see that Yellow cab charges more per kilometers travelled. The maximum price charged for Pink Cab is 1000 USD while for Yellow Cab it is 1600 USD

Average profit per month for each cab
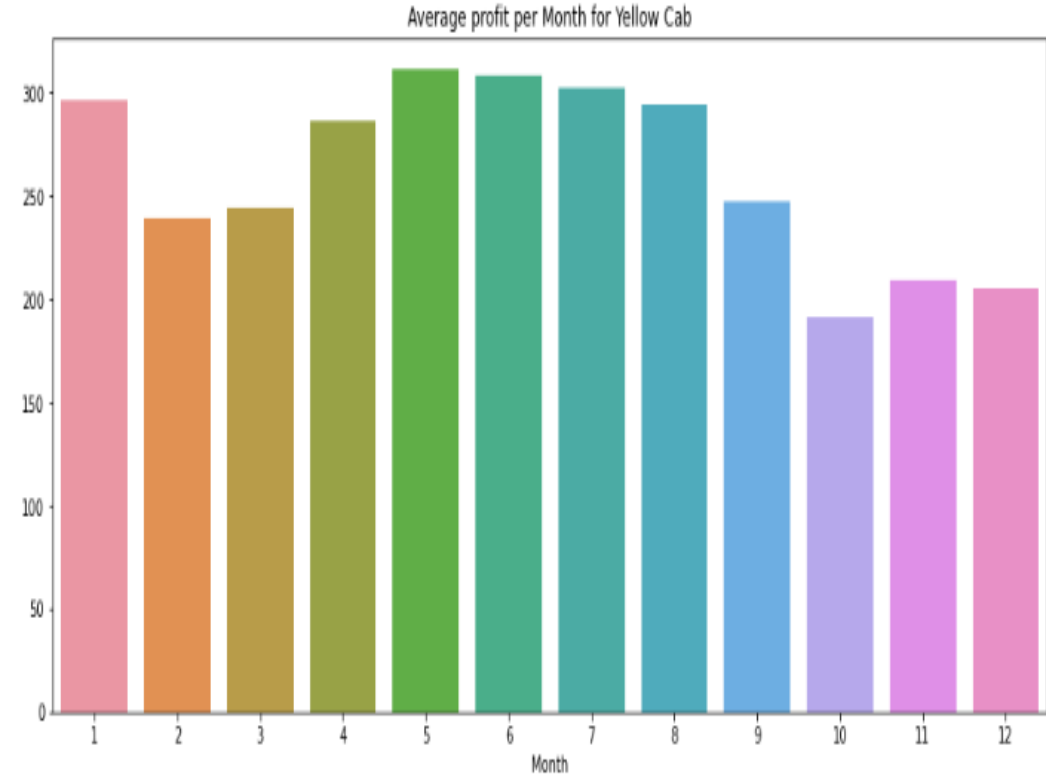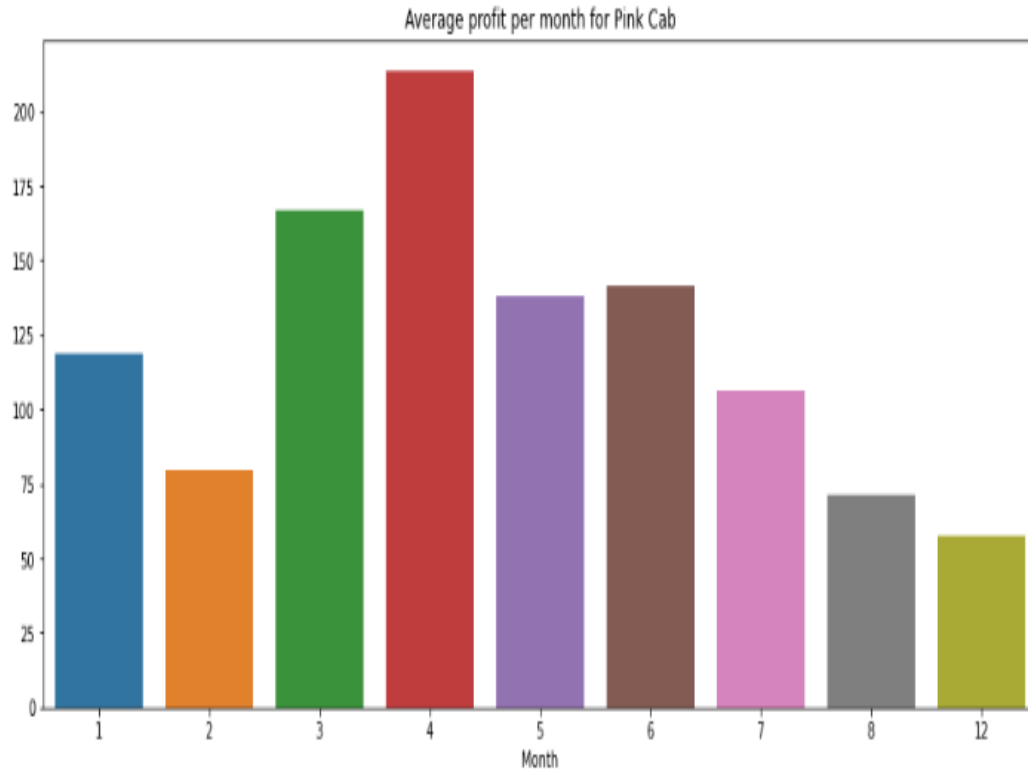
We see that Yellow Cab makes a greater average profit than Pink Cab every month. While Pink Cab makes a maximum average profit of approximately 200 USD in April, Yellow Cab makes more than 200 USD almost every month and reaches a maximum of 300 USD. Pink Cab makes no profit in the months of September, October and November possibly due to a shutdown in those months(as there were also no transactions made). A low profit could also be due to customers earning discounts on their trips.

Correlation of various factors
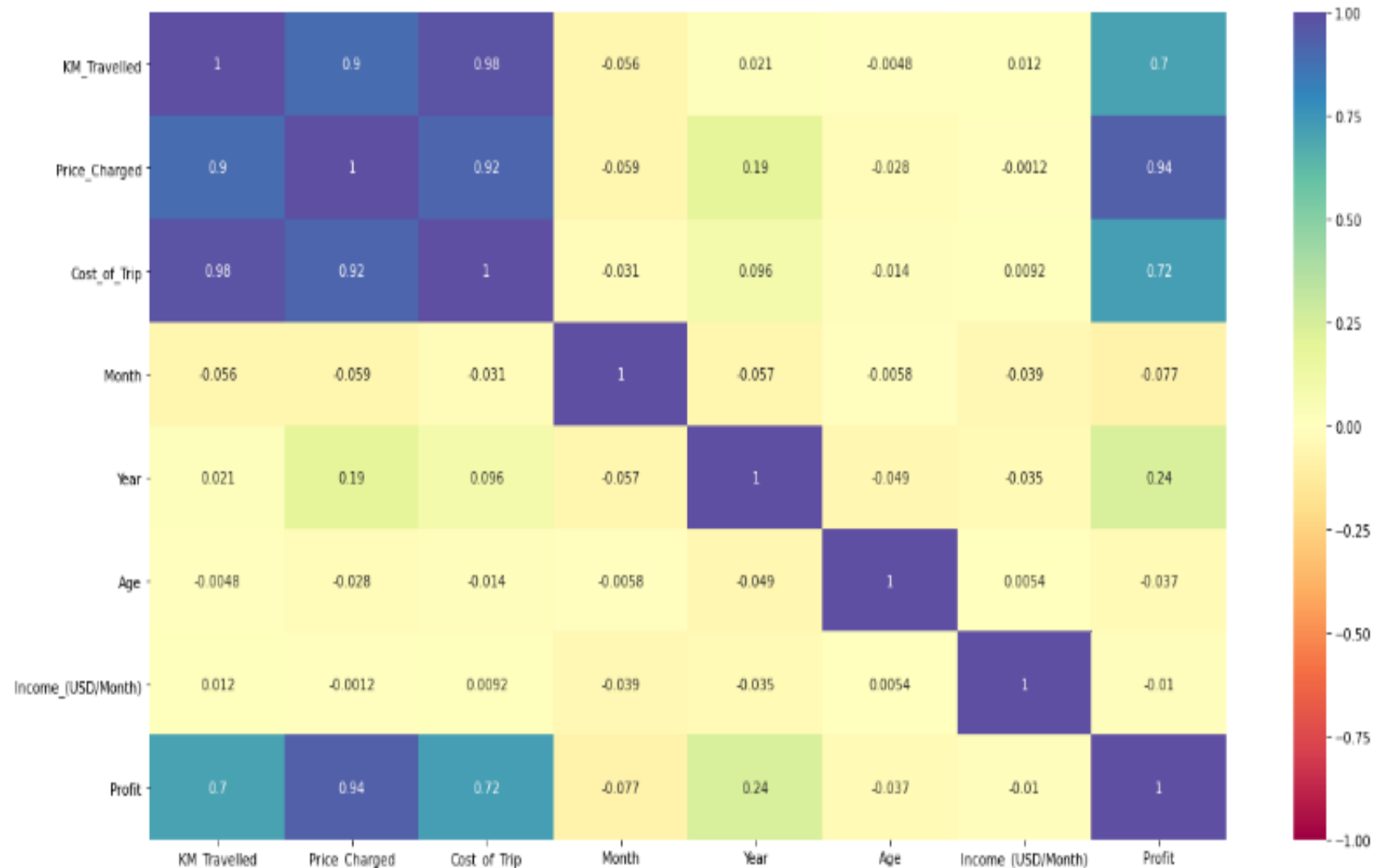
Price Charged is correlated to KM Traveled, Cost of Trip and Profit.

EDA SUMMARY

➤ The highest number of cab users are in New York which also has the highest number of Yellow Cab users.

➤ Both cab rides are in the range of 2km to 48km.

➤ The cost of trip per km travelled is slightly higher for Yellow Cab.

➤ There are overall more users of Yellow Cab than Pink Cab.

➤ Yellow Cab has a higher price charge per km than Pink Cab.

➤ Yellow Cab makes a higher average profit than Pink Cab every month.

➤ Pink Cab has faced a few months of no profit.

RECOMMENDATION

From the visualizations and analysis, it is evident that **Yellow Cab** would be a better option for investment and is therefore recommended to XYZ.

Thank You