



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

22 June 2021

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

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:

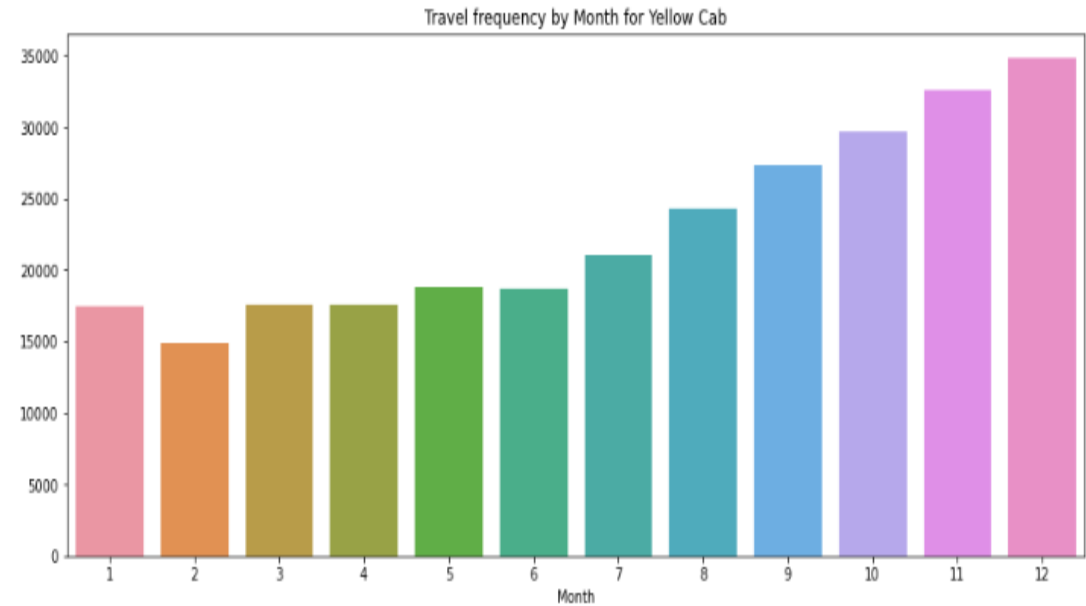
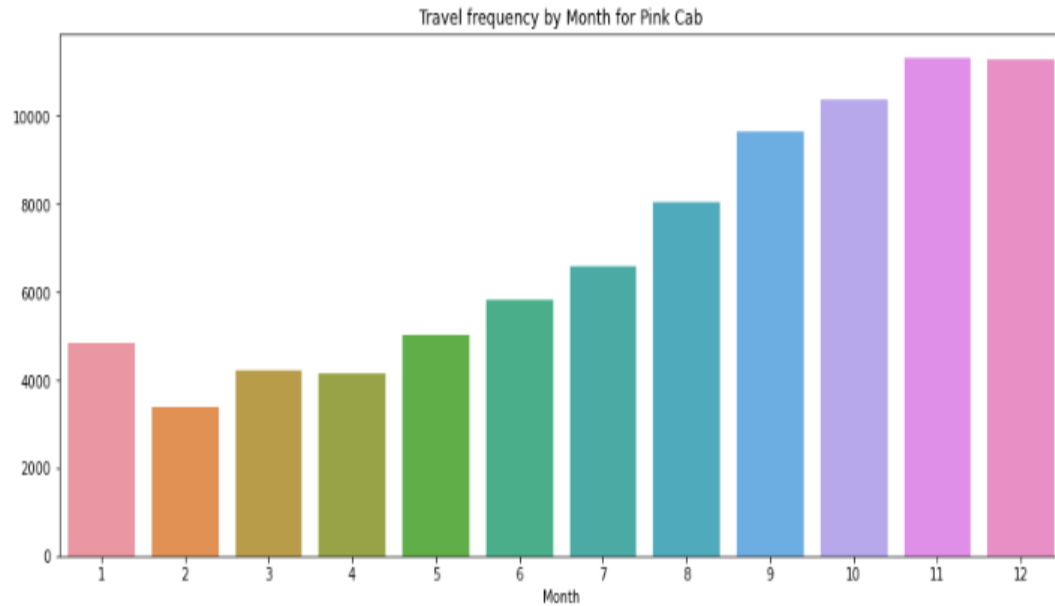
- **Cab Data.csv** – this file includes details of transaction for 2 cab companies
- **Customer ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details
- **Transaction ID.csv** – this is a mapping table that contains transaction to customer 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.
- ❑ Profit was calculated as $\text{Price Charged} - \text{Cost of Trip}$
- ❑ 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

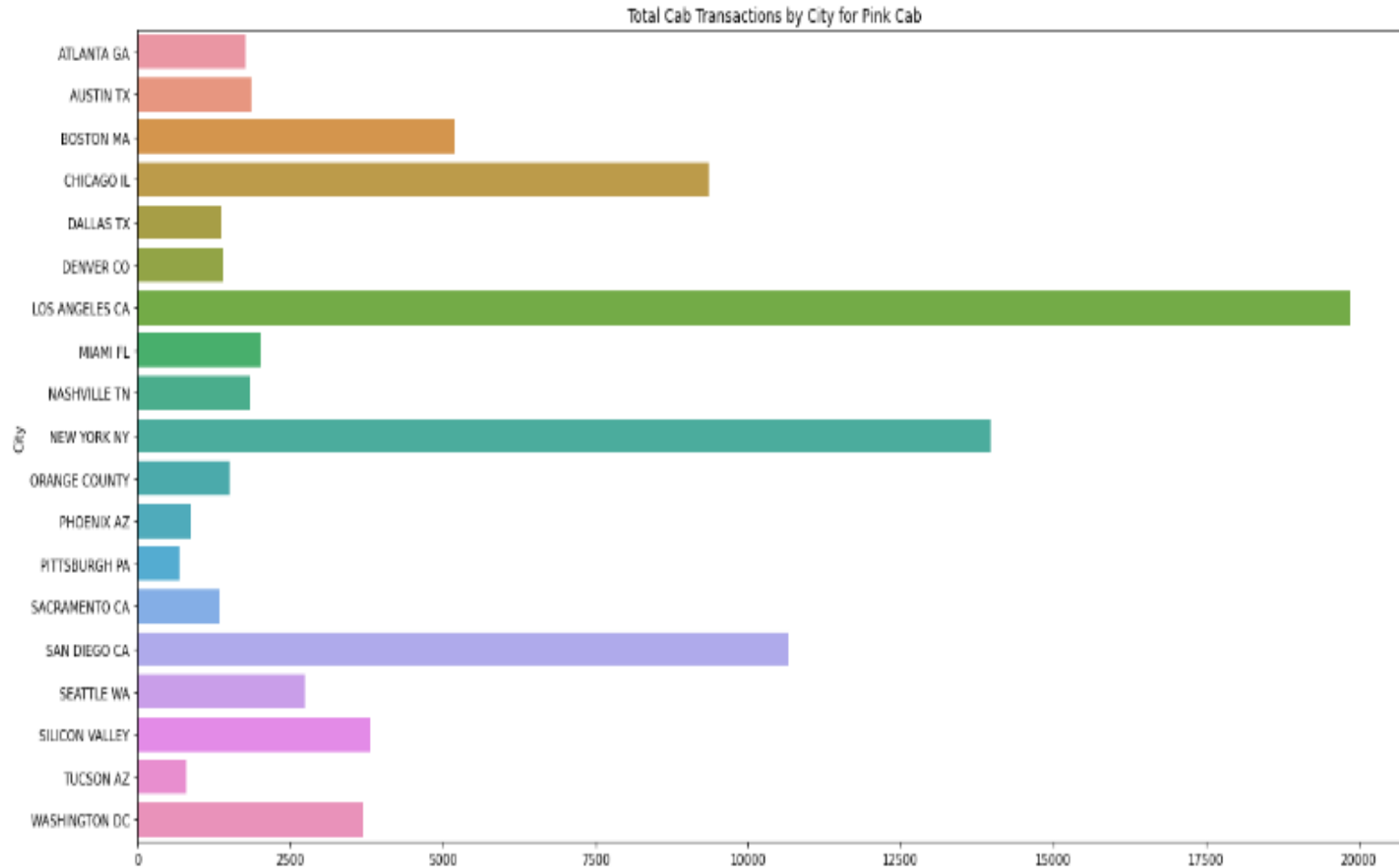
Data Visualization

Monthly Travel Frequency



We see an increase in travel frequency towards the end of the year for both cabs. This could be due to the holiday season.

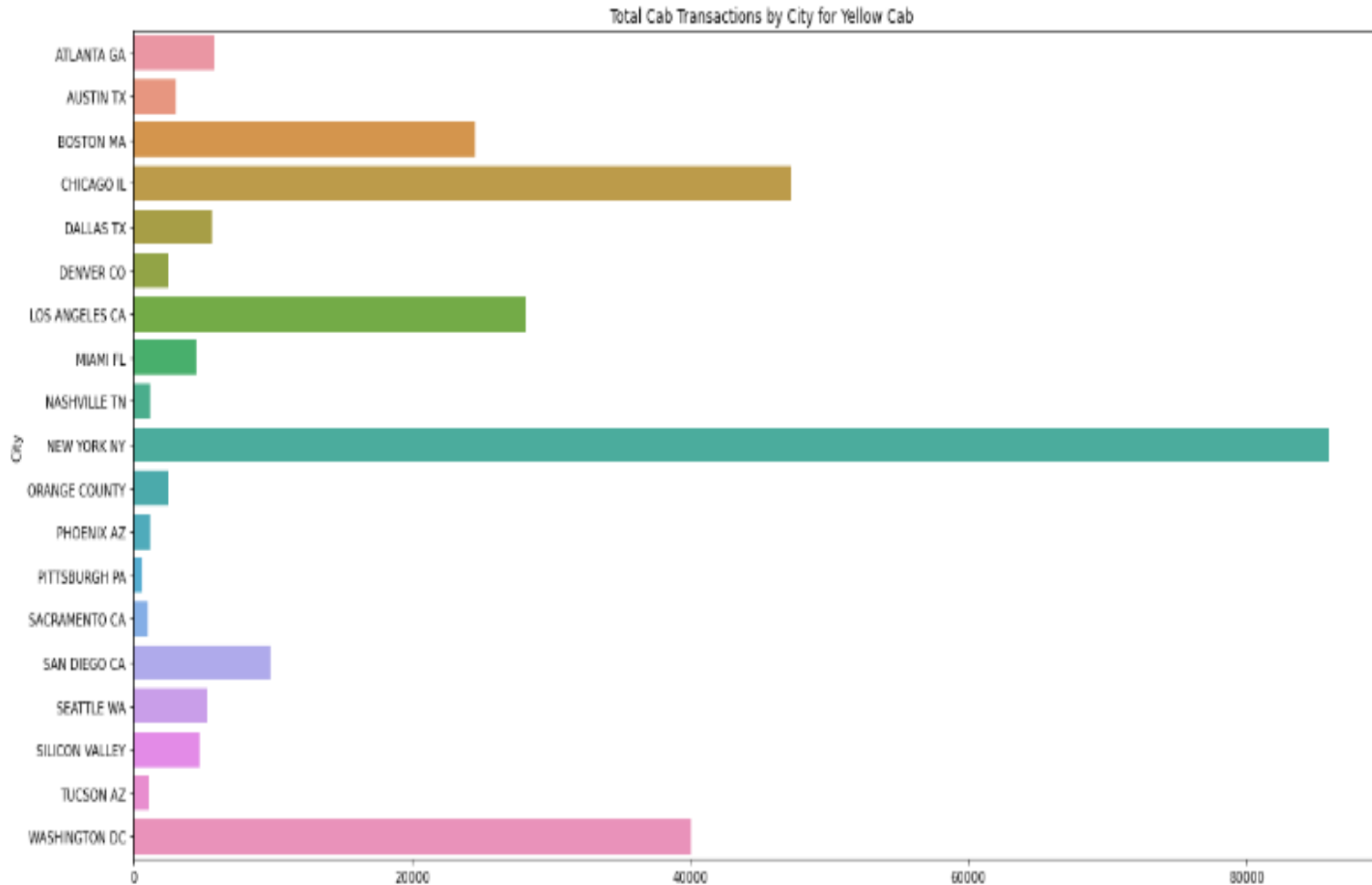
Pink Cab Users



The highest number of Pink Cab Users are in:

- Los Angeles
- New York
- San Diego

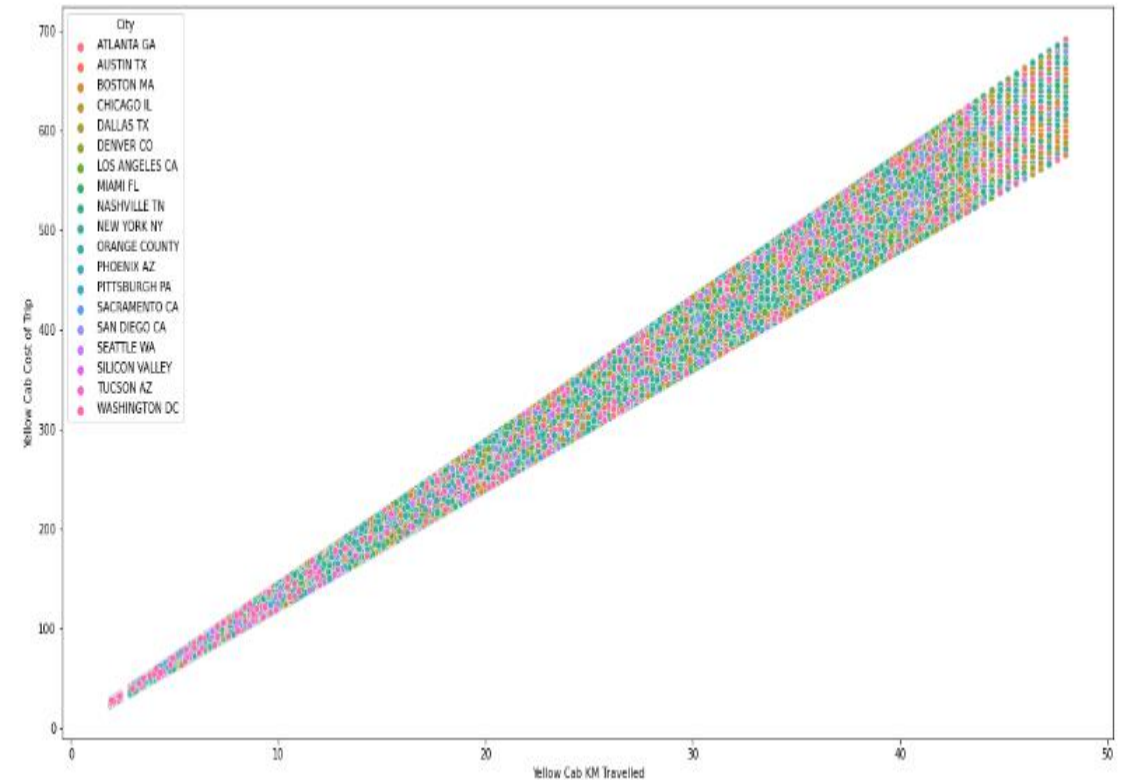
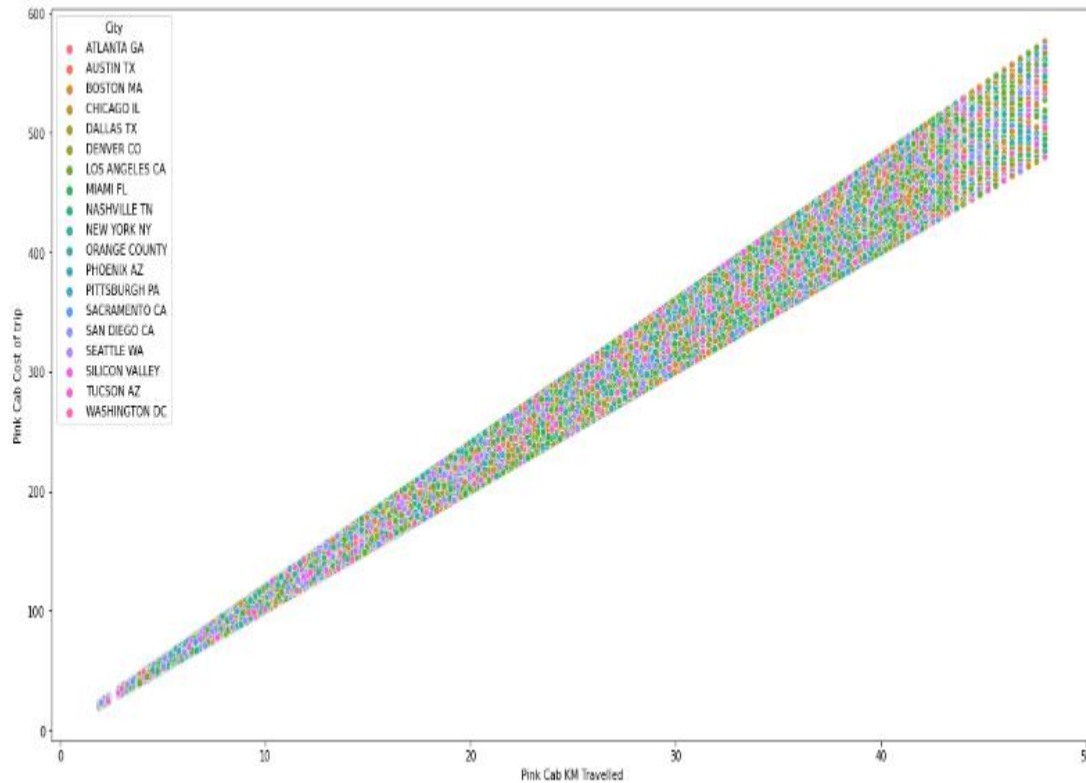
Yellow Cab Users



The highest number of Yellow Cab Users are in:

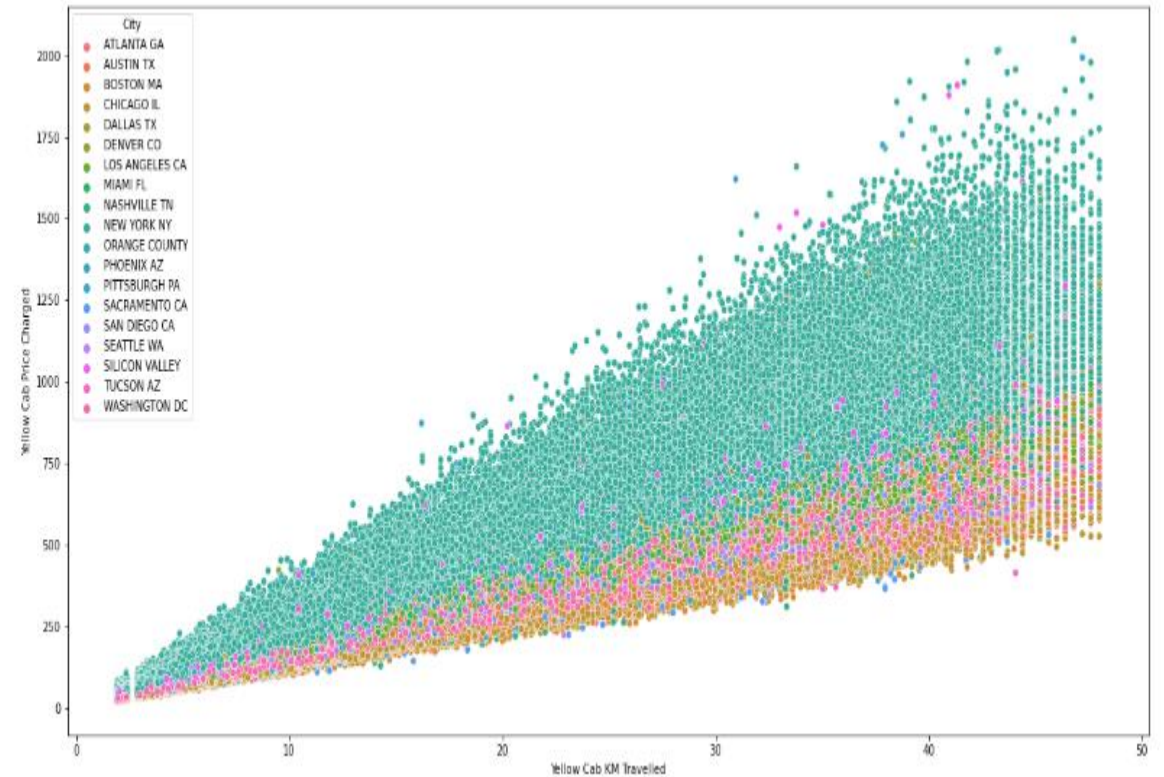
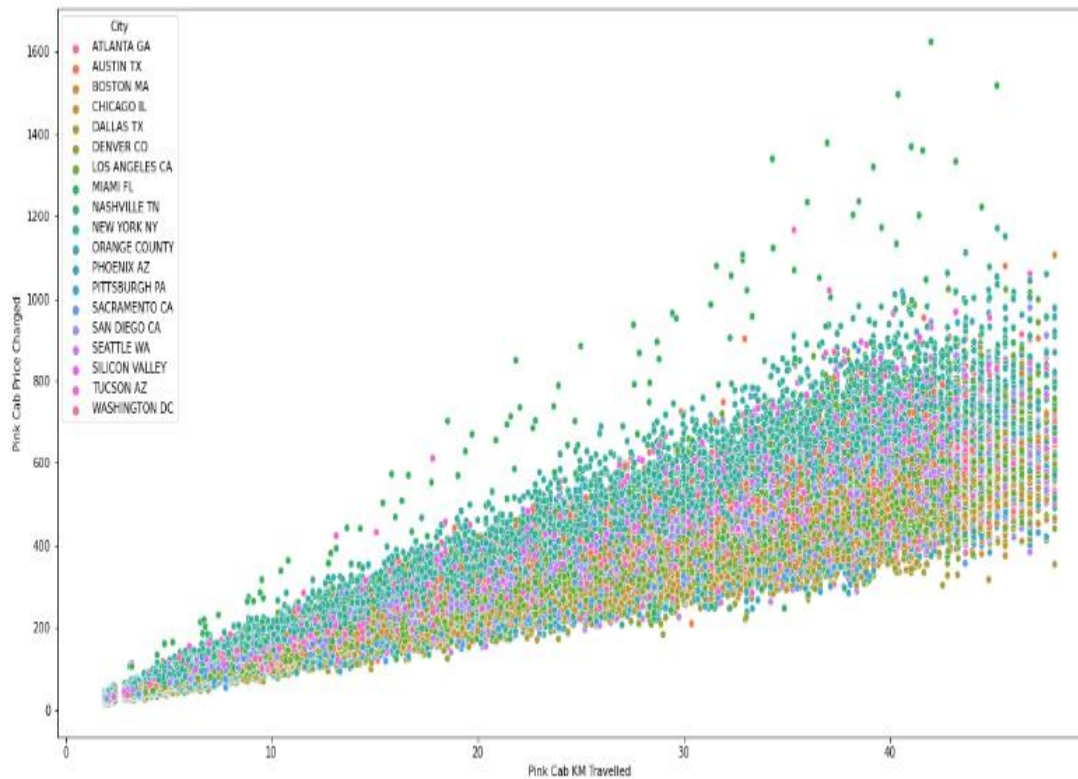
- New York
- Chicago
- Washington

KM Travelled vs Cost of Trip



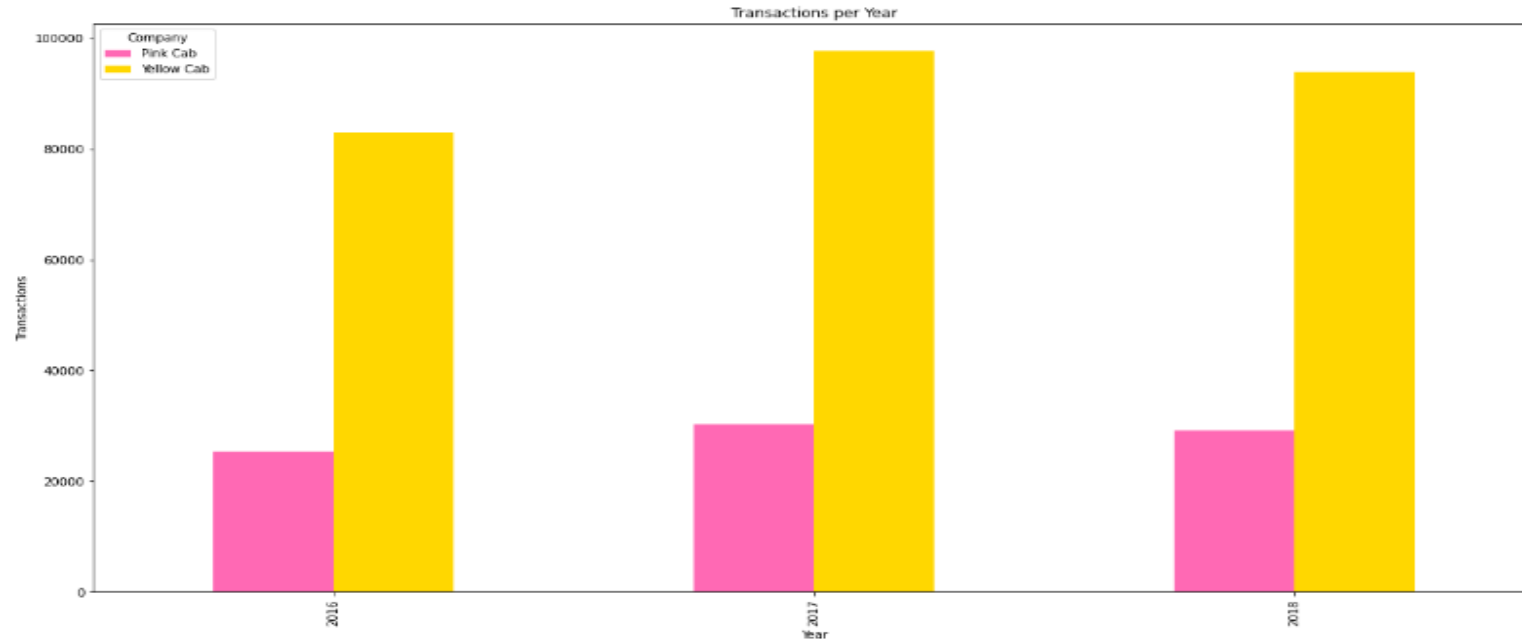
Both cabs travel between 2km to 48 km, however the cost of trip for Yellow Cab is slightly higher

KM Travelled vs Price Charged



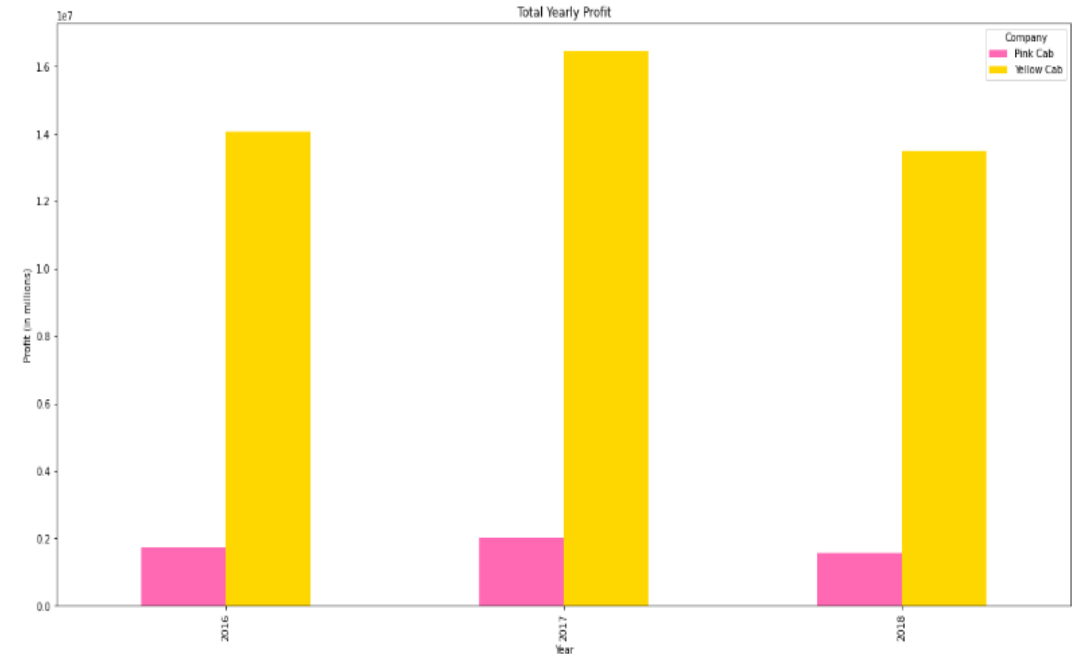
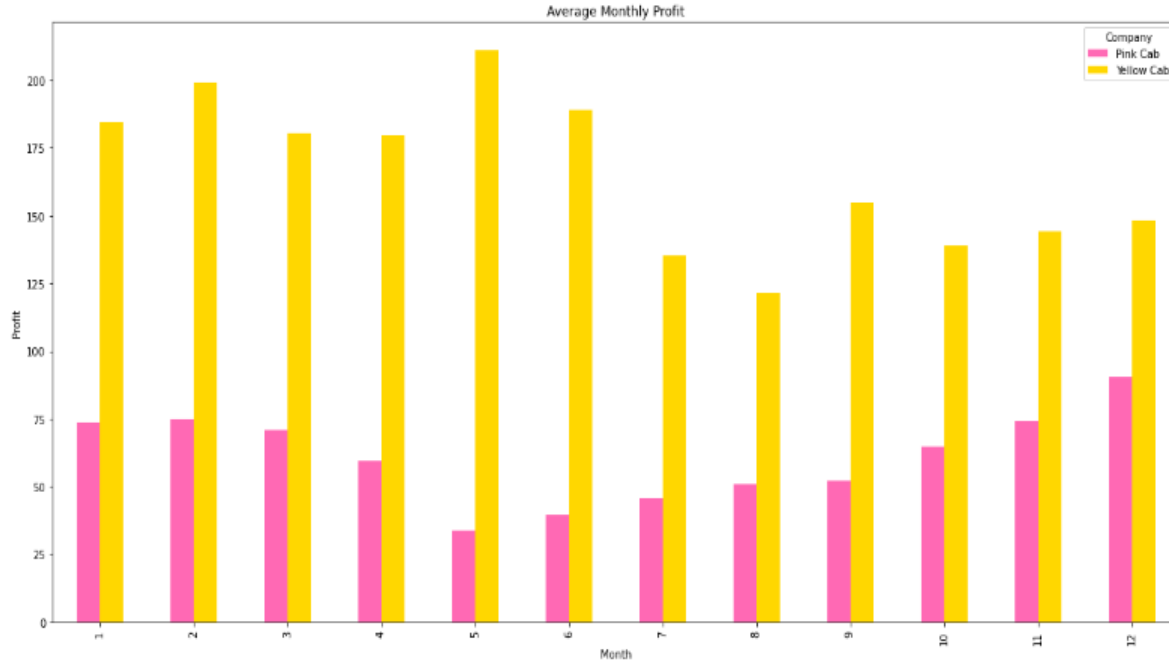
It is quite clear that Yellow Cab prices are much higher than Pink Cab. Yellow Cab also charges a high price in New York compared to other cities.

Yearly Transactions



Over the three years Yellow Cab has more than double the amount of transactions as Pink Cab

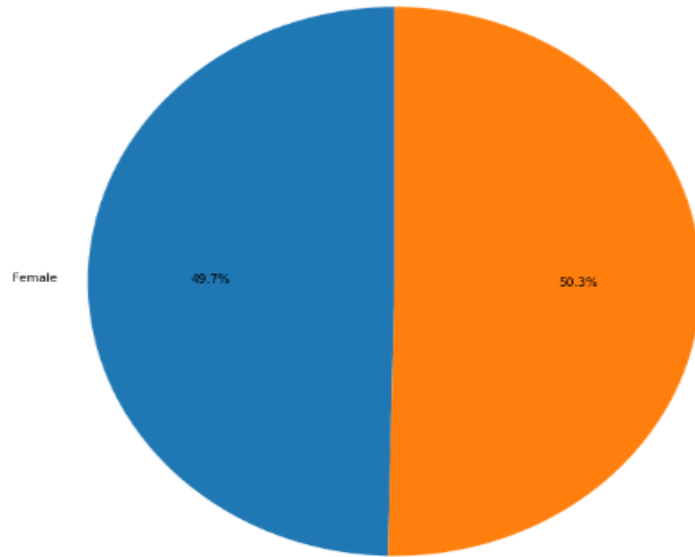
Monthly and Yearly Profit



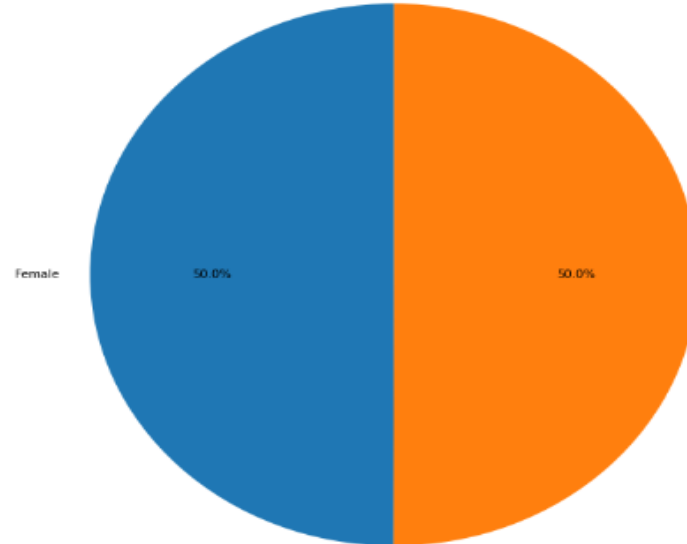
Yellow Cab makes its highest profit in the months of May, February and June. Pink Cab makes its highest profit in the months of December, February and November. However, Yellow Cab makes a much higher profit every month as well as every year. Pink Cab annual profit is almost the same every year. Although, there is a decrease in profit for Yellow Cab in 2018, it is still much higher than Pink Cab.

Gender Analysis

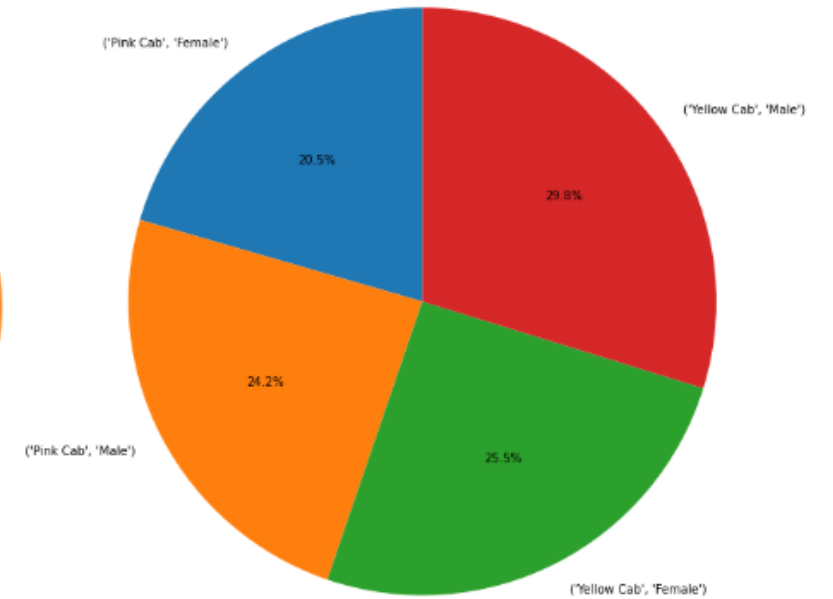
Price Charged per Gender for Yellow Cab



Price Charged per Gender for Pink Cab

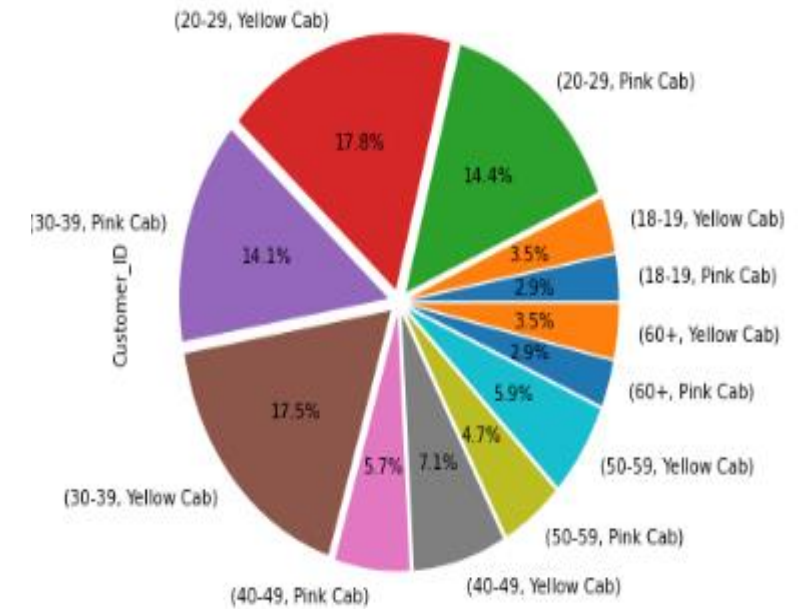
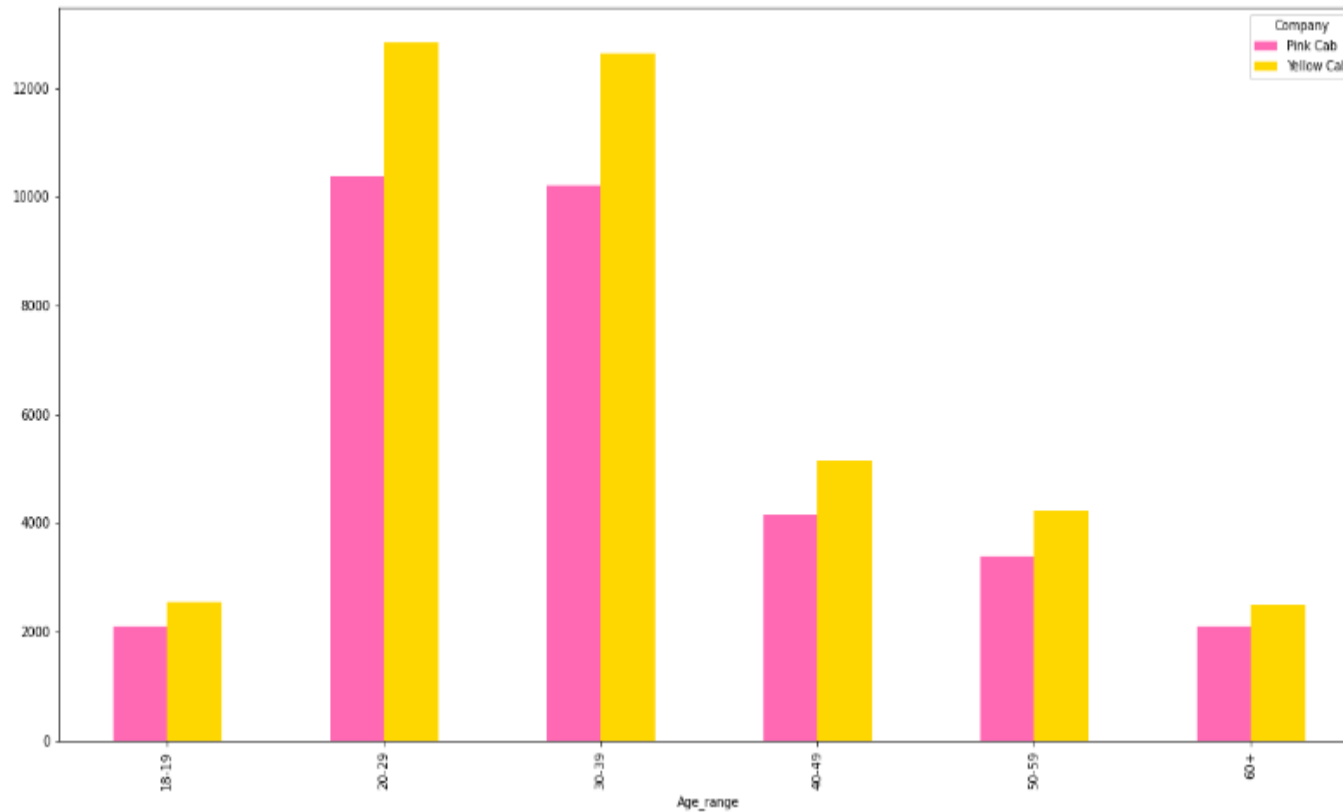


Customer share per gender per cab



Pink Cab charges equally for both genders while Yellow Cab charges slightly less for females. Both Cabs have more male customers

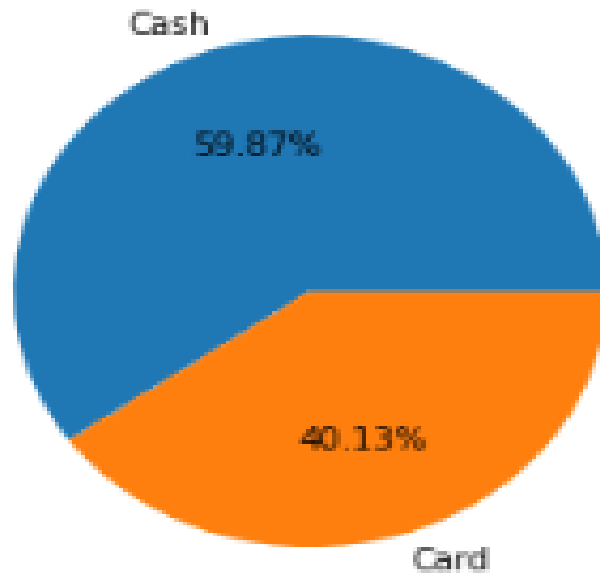
Age Analysis



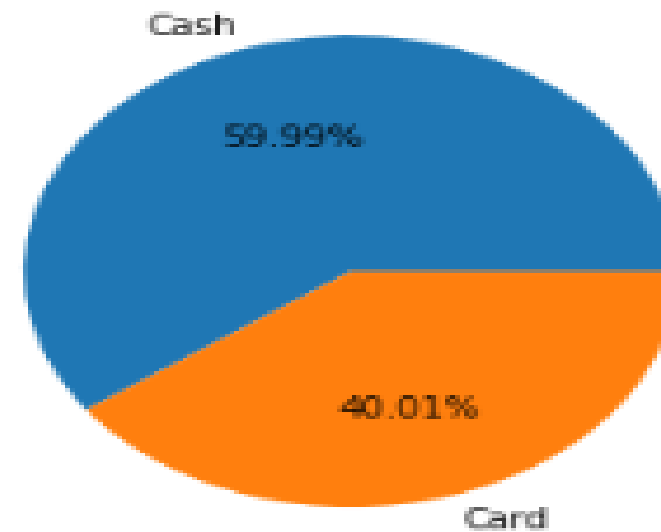
For both cabs, most customers are aged between 20 – 40 years with only a few aged 18-19 and 60+ years

Payment Mode Analysis

Payment Method for Pink Cab

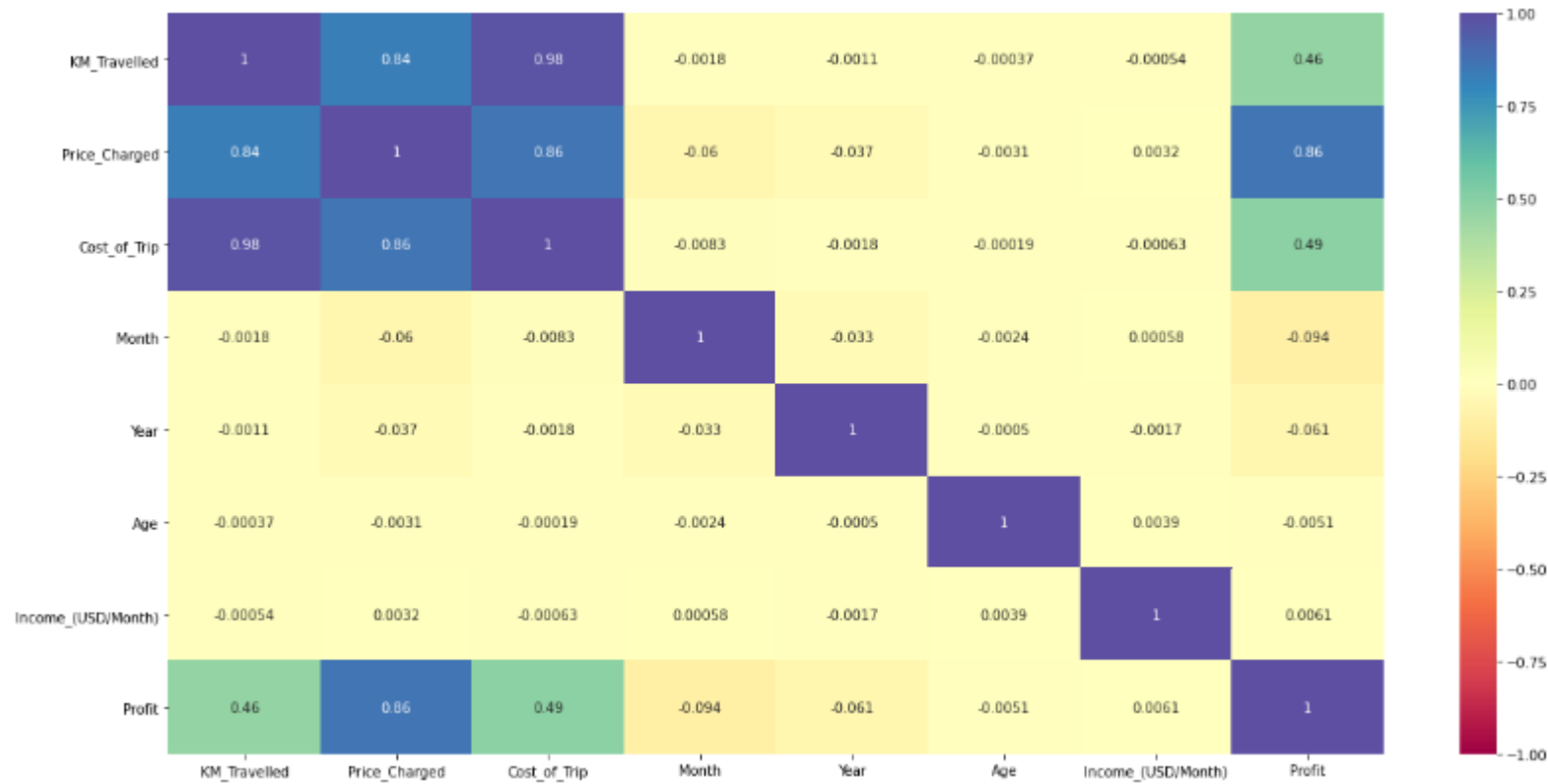


Payment Method for Yellow Cab



Both cabs have fairly the same % of transactions made in cash. Cash seems to be the more convenient payment mode for customers in both cabs

Correlation chart



Factors such as KM Travelled, Price Charged and Cost of Trip are highly correlated to each other. Other factors are uncorrelated.

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 and year.

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