Exploratory Data Analysis

G2M for cab investment

15/06/2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

Executive summary

During the past few years, the cab industry has experienced tremondous growth. It is being widely used by people from various age groups, income classes and across various states across the United States.

Lets consider a few facts about the industry:

- **1.2 million+ customers** | Across the united states
- **30 percent +** | of the population in some states using cab services

These remarkable figures are attracting a lot of investment from companies worlwide. XYZ is a private firm in the United States that would like to follow suit.

Problem statement

XYZ is a private firm in the united States that would like to invest in the cab industry. Using data from 2 local cab services, Yellow cab and Pink cab over the period of 3 years, I have been tasked to give the firm insight on which of the two cab services they should invest in.

Approach

We will use exploratary data analysis to discover trends in the cab industry, for example seasonality and customer base.

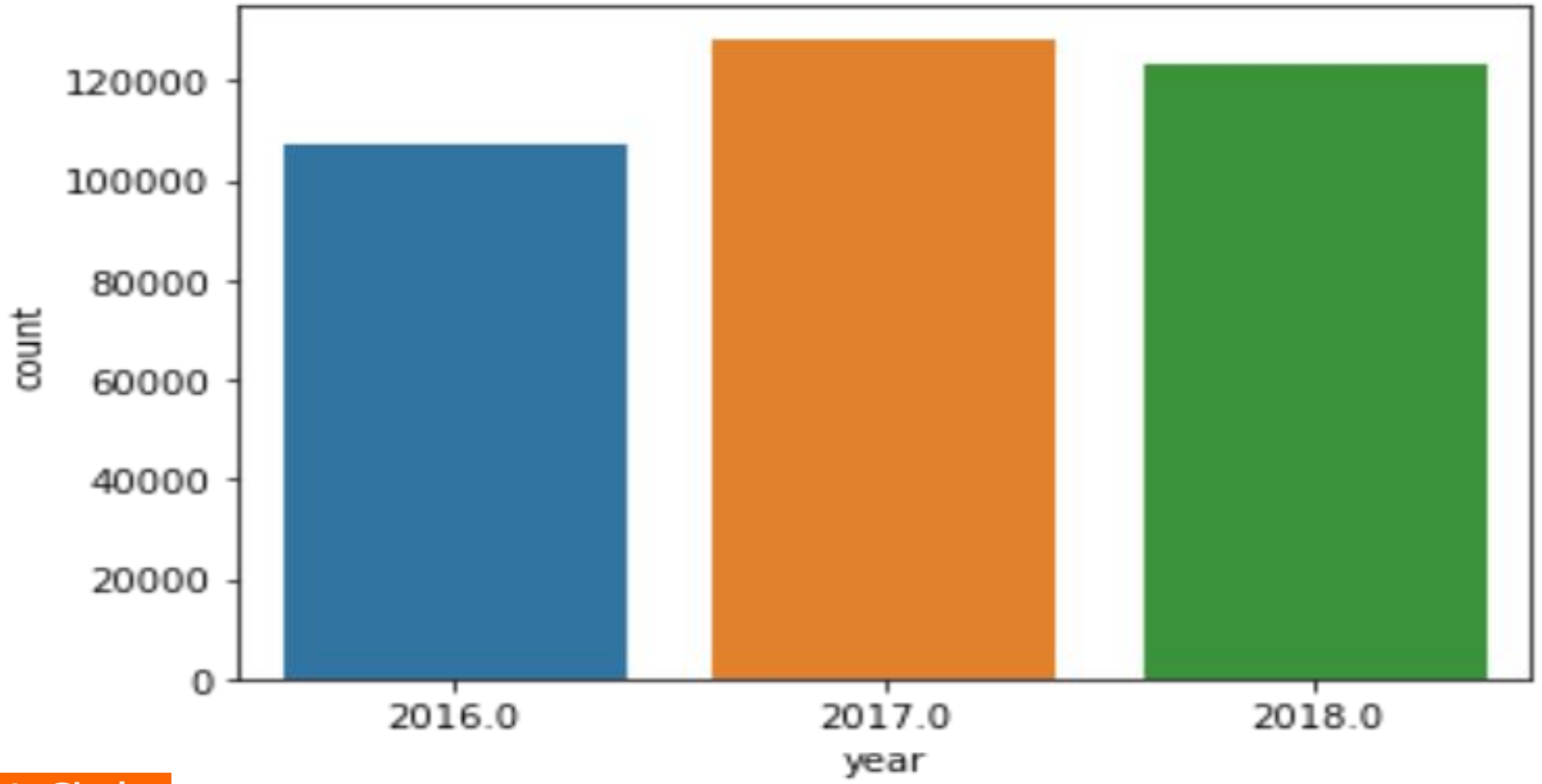
The same approach will be used to determine which of the two cab service would be a better choice for investment.

We have formulated the following hypothesis that may be tested using this approach:

- Pink cab is more affordable compared to Yellow cab
- There is some gender preference when it comes to selecting a cab service
- There is a preffered mode of payment by customers
- There is seasonality in the number of people using the cab service
- A certain income group prefers a certain cab service
- A cerain age group prefers a certain cab service
- Customers using a certain cab sevice tend to use it again

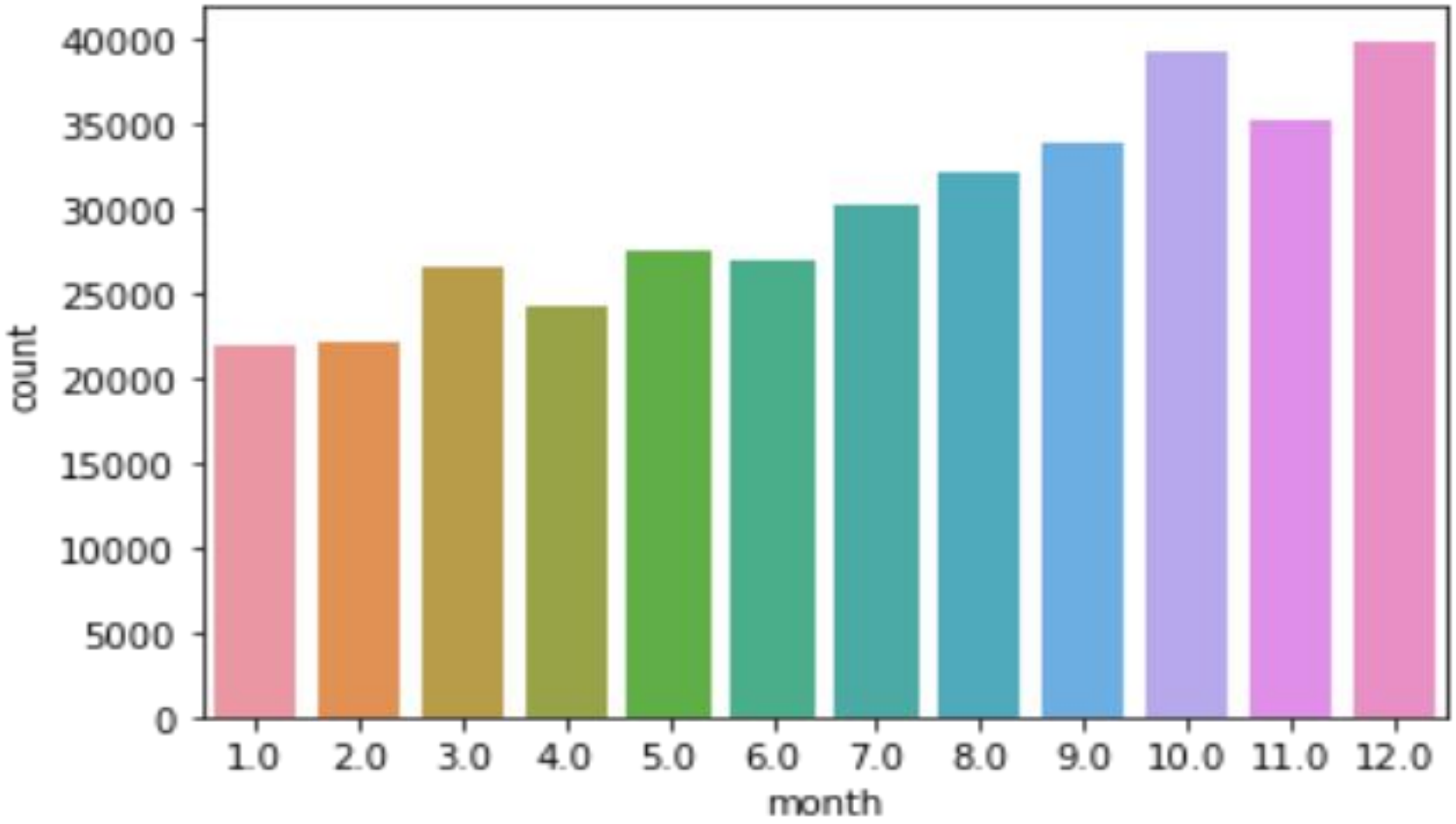
Exploratory data analysis

General travel frequency using cab services



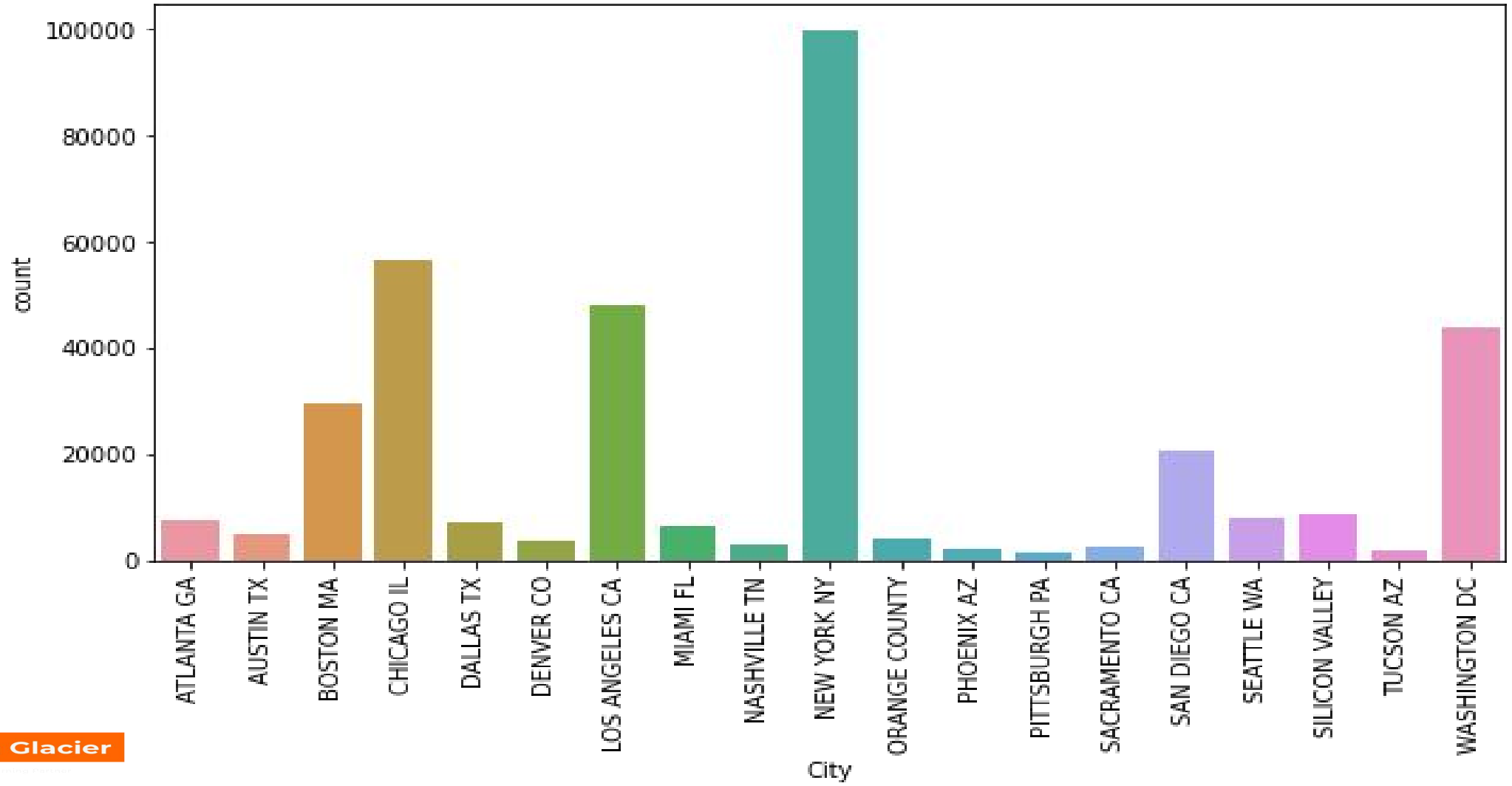
Exploratory data analysis

General travel frequency using cab services by month



Exploratory data analysis

General Travel frequency using cab services in various cities



Exploratory data analysis

Summary

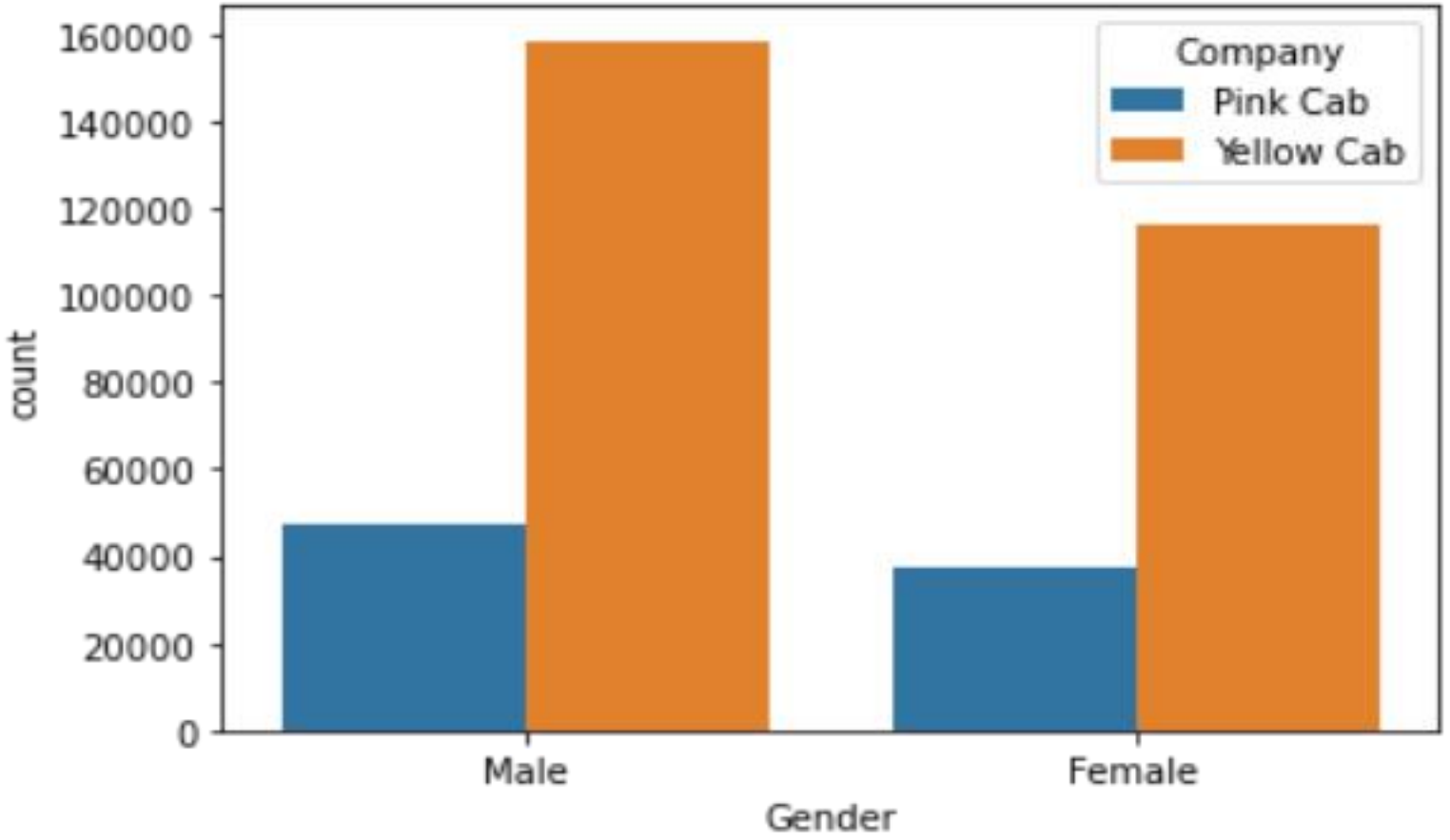
We can draw the following conclusions from the above visualizations:
- pink cab generally provides more rides compared to yellow cab
- 2017 was the year cab services was used the most, although the data is more-less fairly distributed
- Cab services are most popular in New York and least popular in Pittsburgh
- Most frequented months for cab services are December and October

Data Glacier

Exploratory data analysis

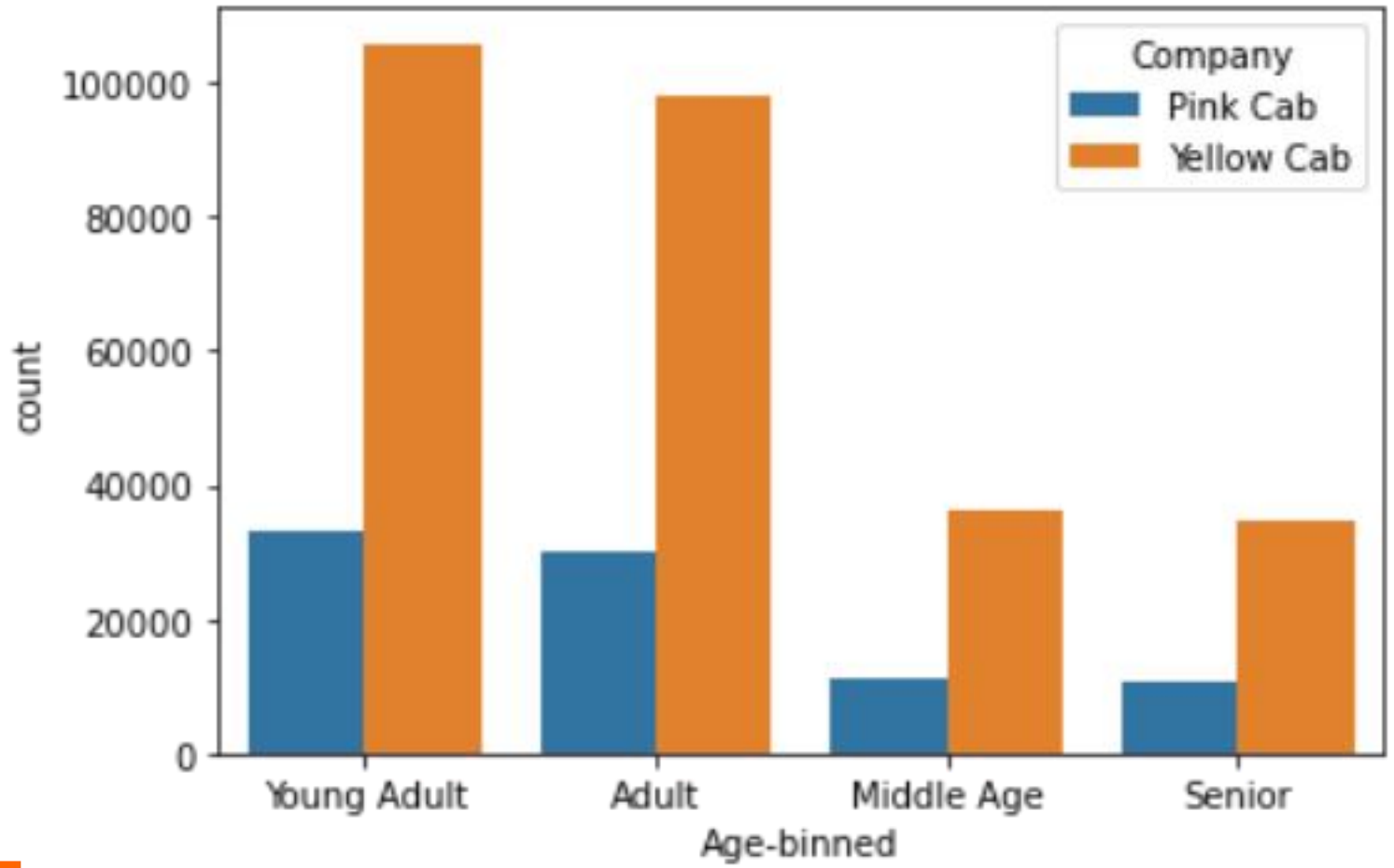
Customer base analysis

Customer analysis by Gender



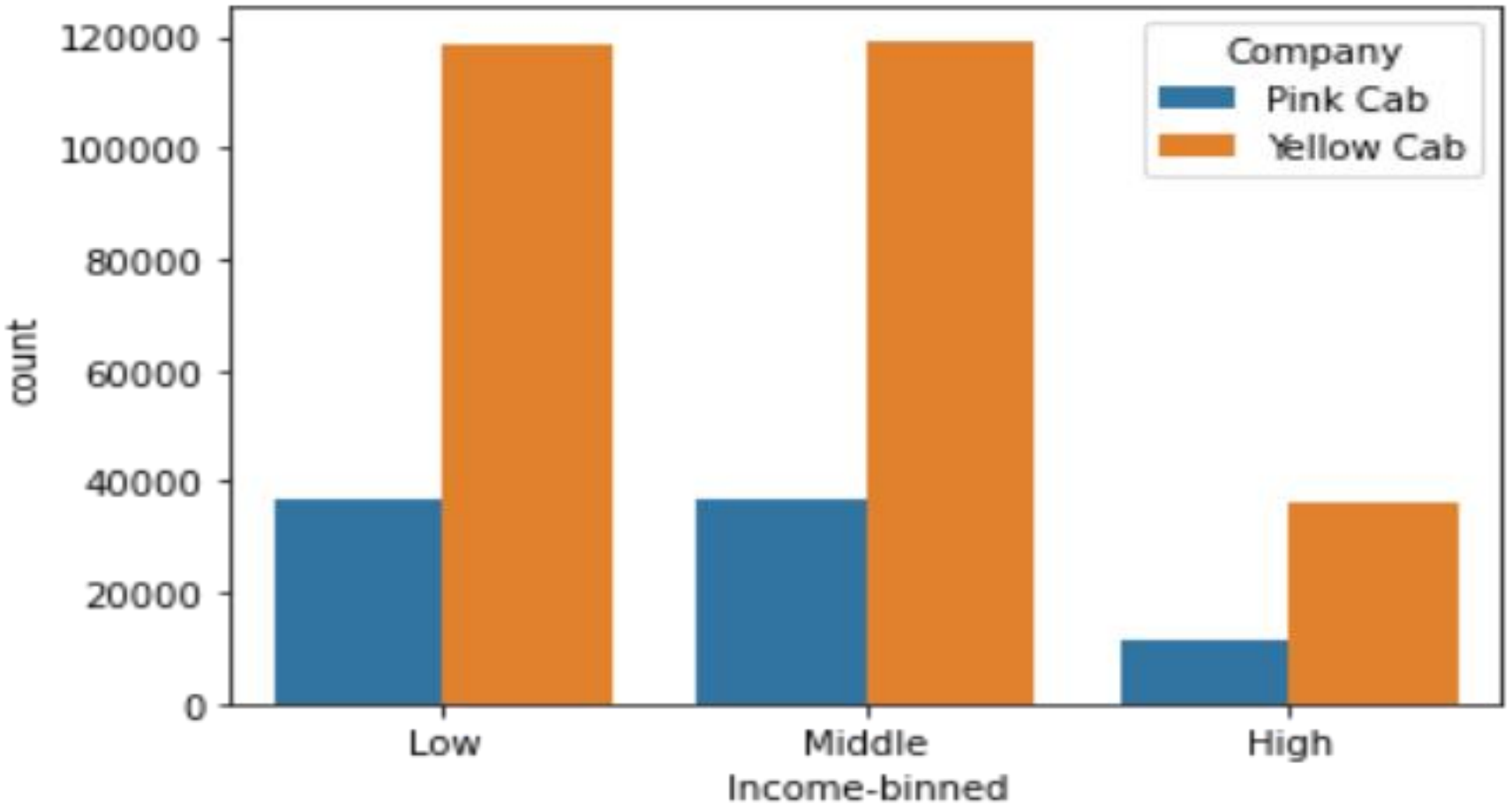
Exploratory data analysis

Customer Analysis by Age



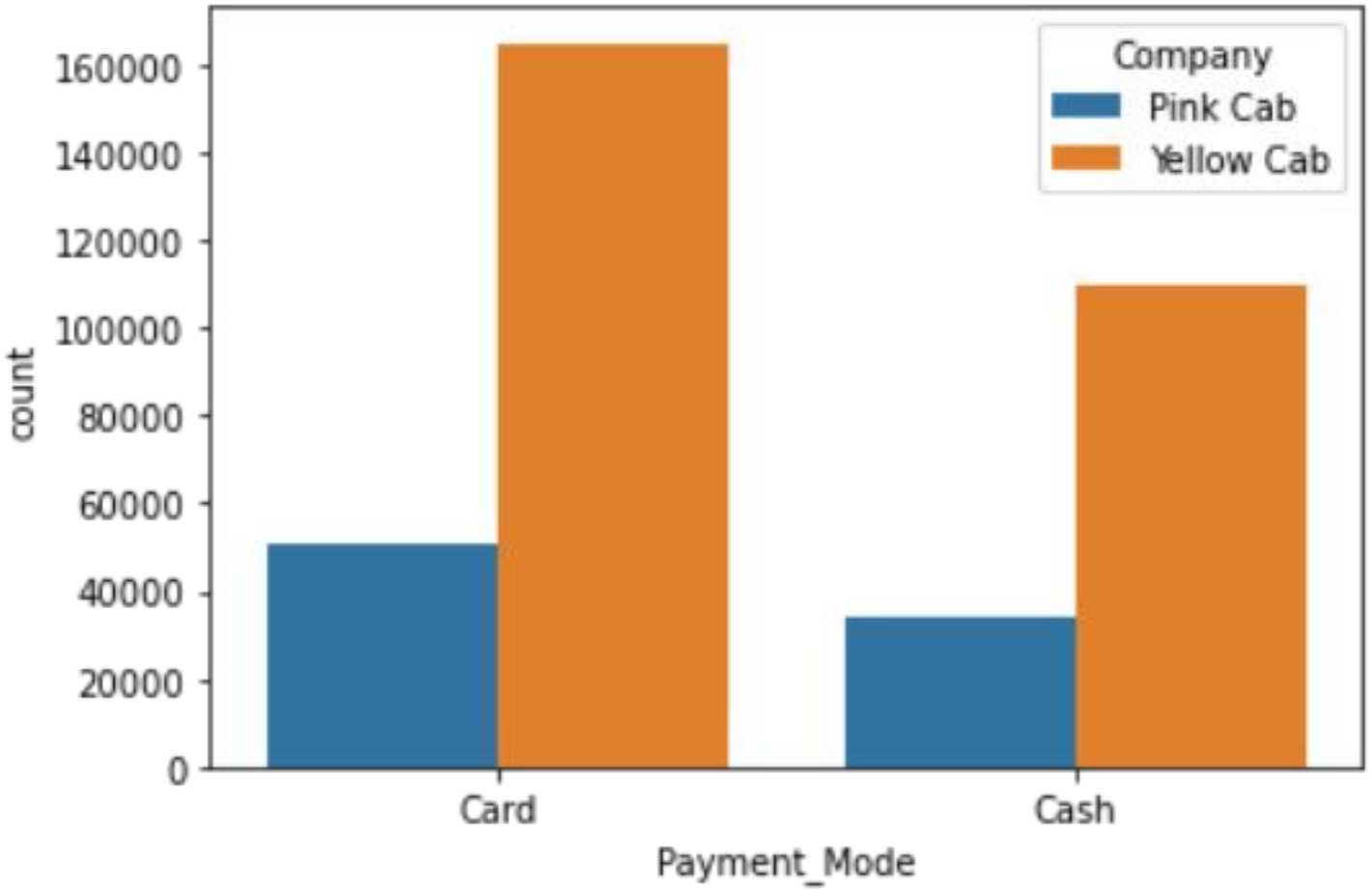
Exploratory data analysis

Customer Analysis by Income groups



Exploratory data analysis

Customer analysis on preffered mode of payment



Exploratory data analysis

Summary

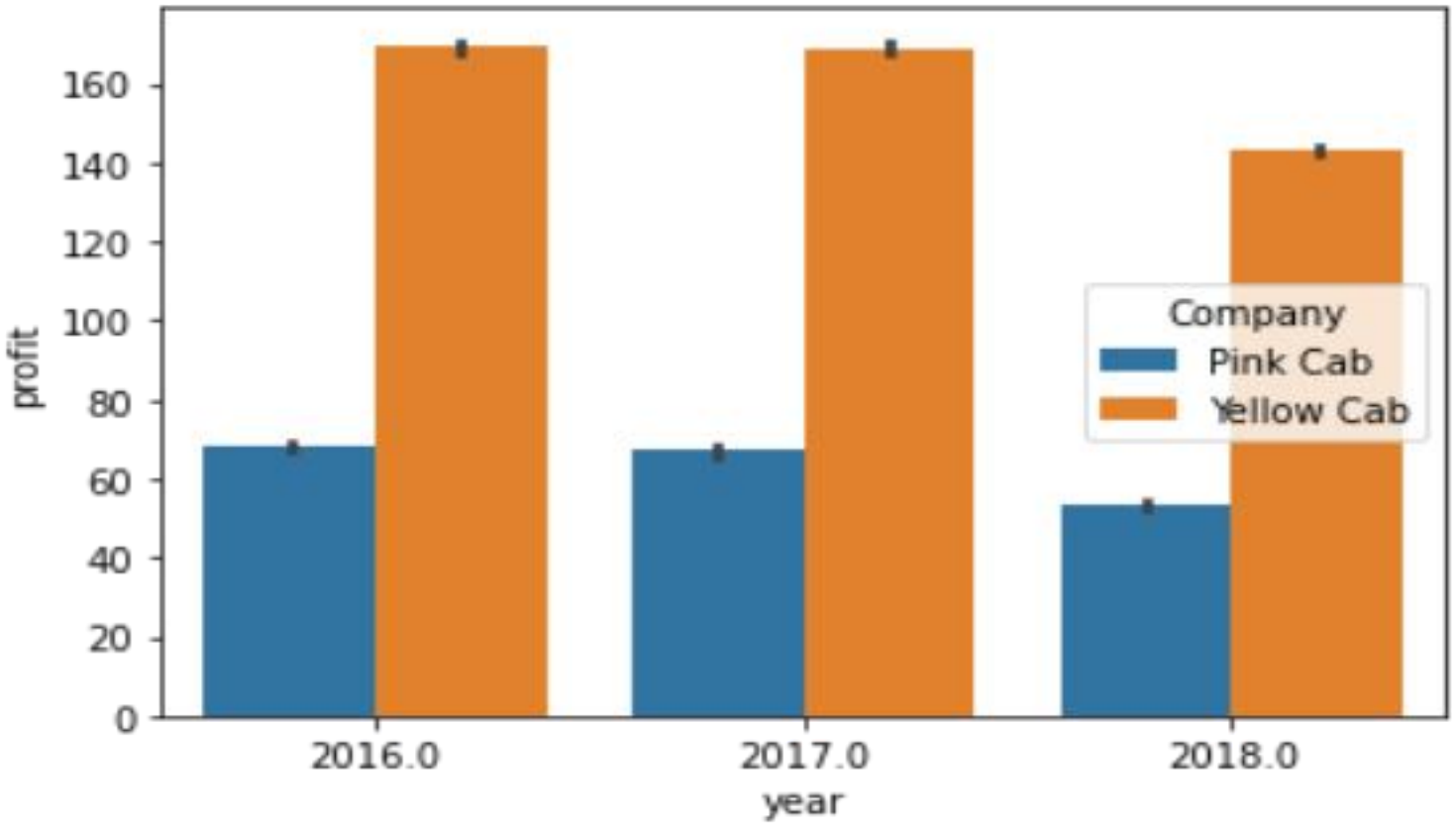
We can draw the following conclusions on our customer base:
- Men generally use cab services more than women do
- Young adults and adults are the most frequent uses of cab services
- low and middle income induvuals use cab services more than those of high income
- The preffered mode of payment by customers is card

Data Glacier

Exploratory data analysis

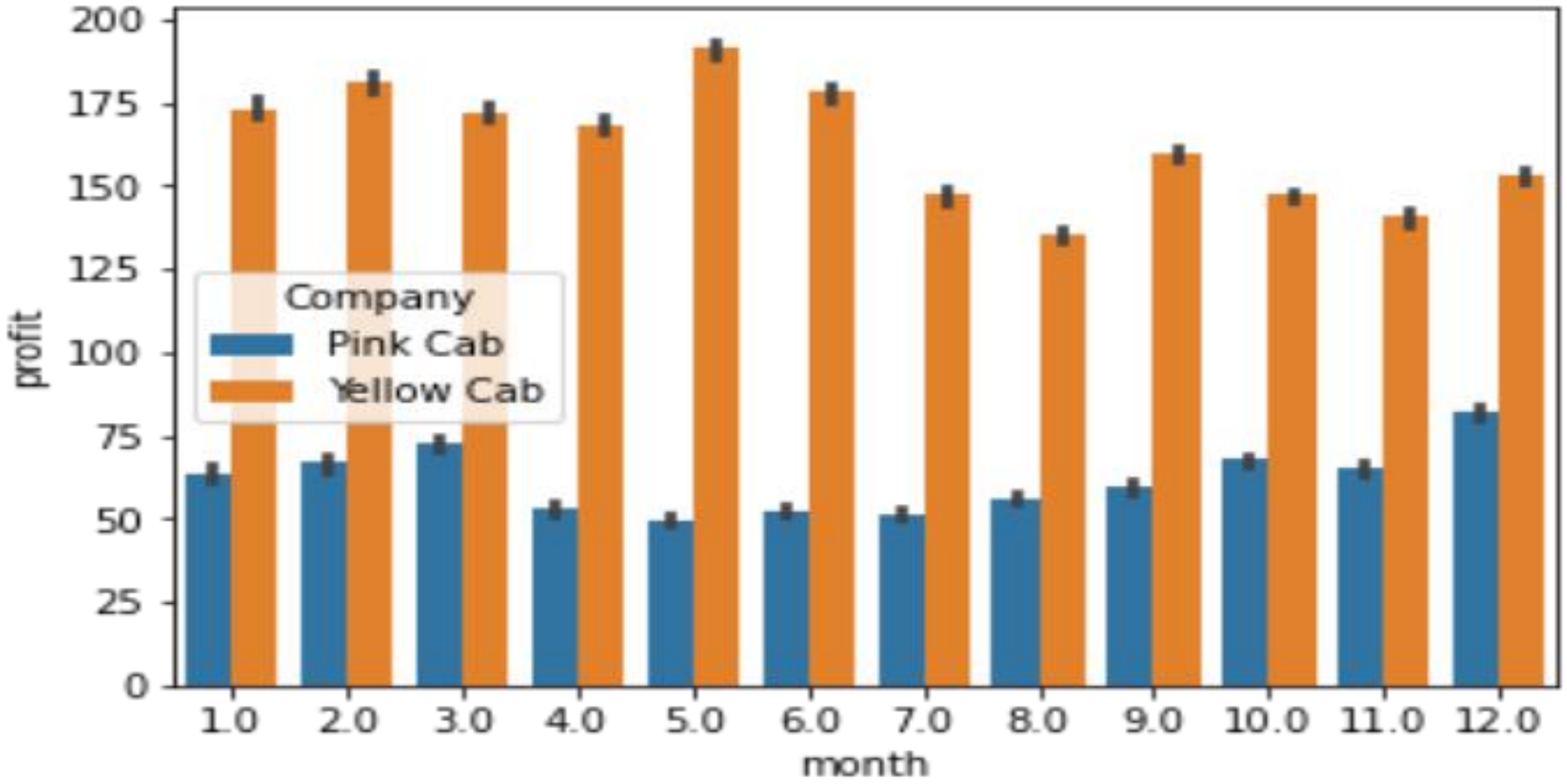
Analysis for each cab service

Annual Profit analysis



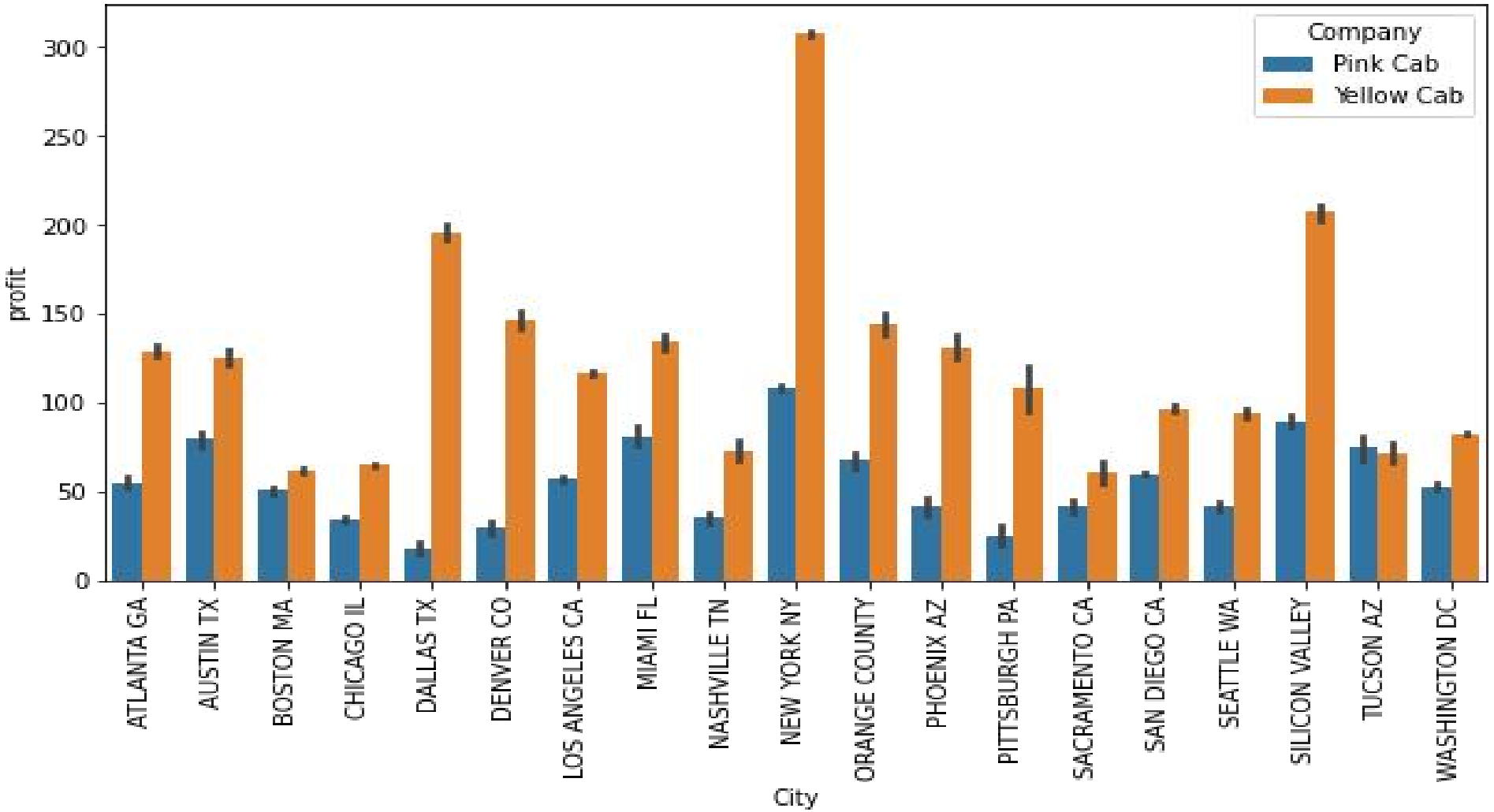
Exploratory data analysis

Monthly profits analyis for each cab service



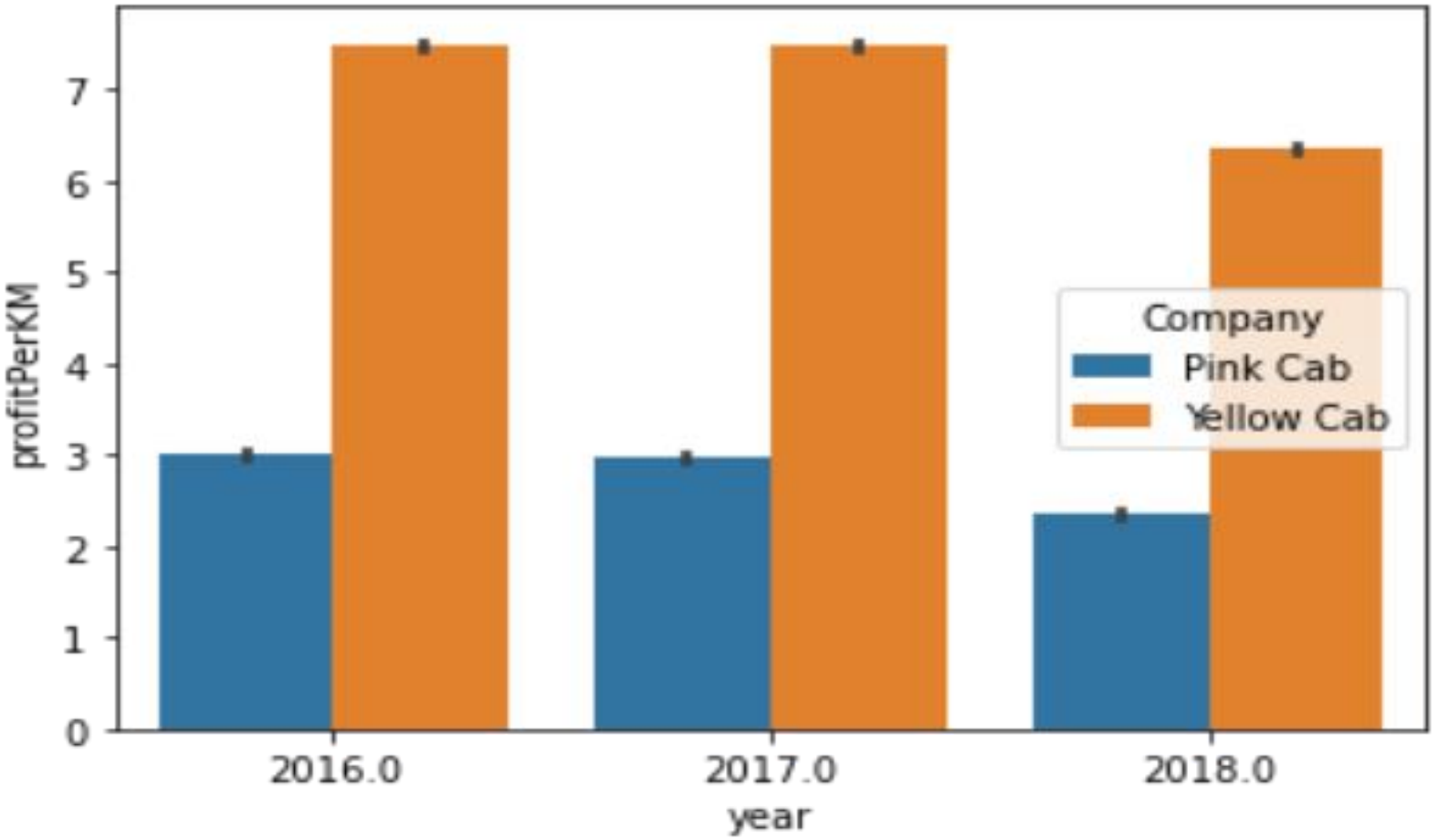
Exploratory data analysis

Profits by City for each cab service



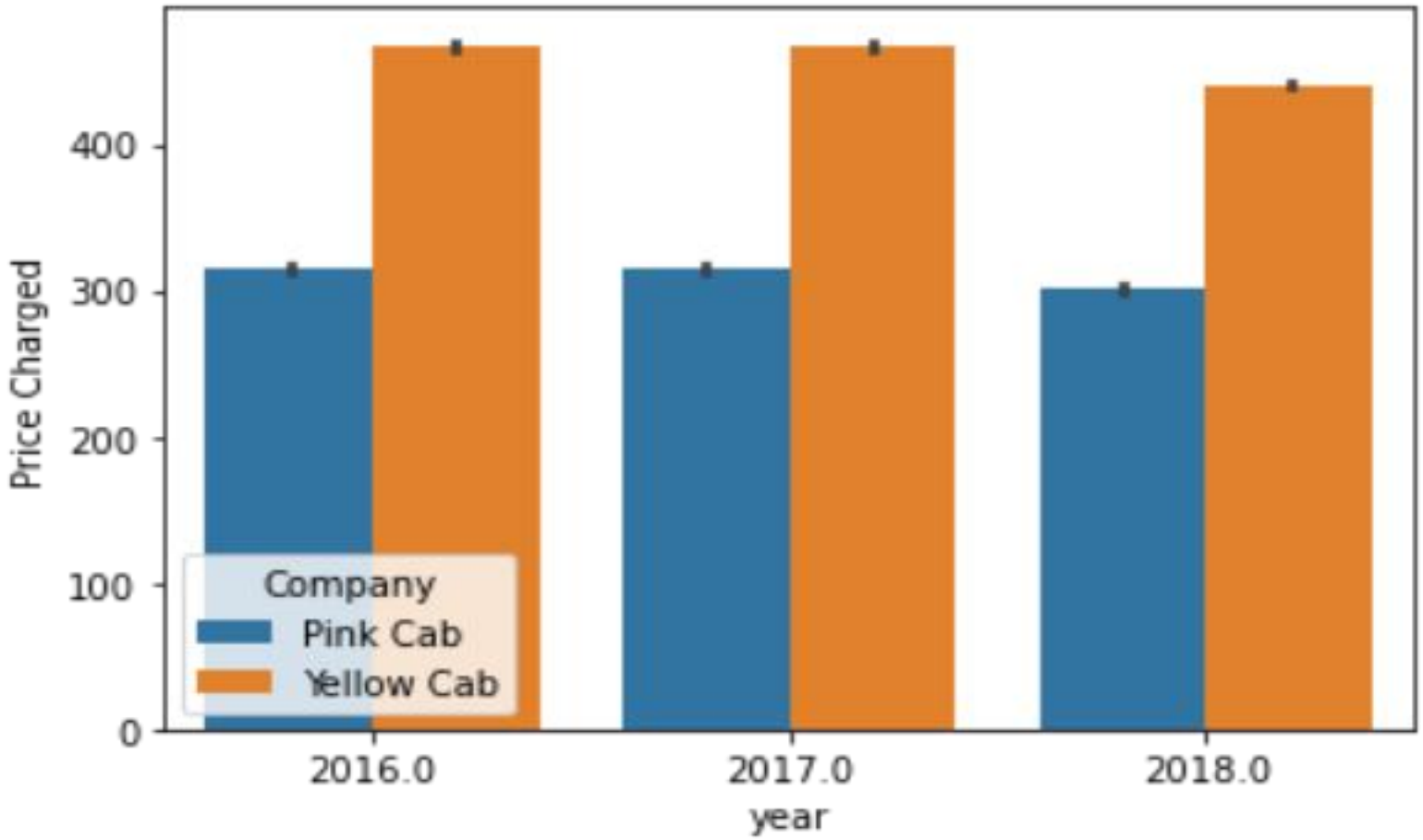
Exploratory data analysis

Average profit per KM of the cab services



Exploratory data analysis

Prices being charged by each cab service

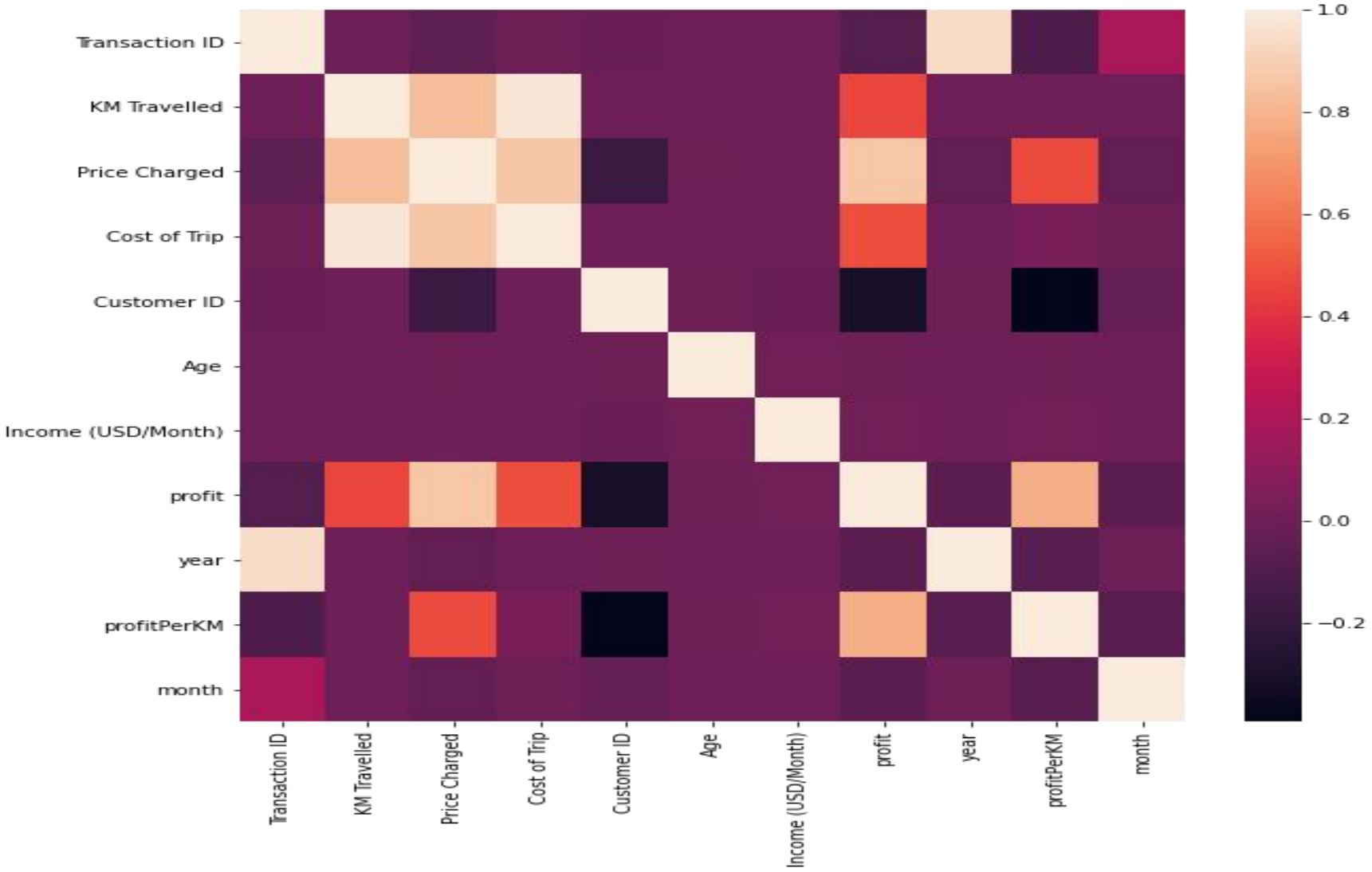


Exploratory data analysis

- From the above data, it is evident that Yellow cab is making more profit than Pink cab across all observed cities and through out the entire time period, despite Pink Cab being the more frequented service
- This may have been brought about by Yellow cab charging more for their services

Exploratory data analysis

Heatmap to visualize correlation between variables



Data Glacier

Exploratory data analysis

using the above heatmap, we can curate a list of highly correlated columns:

- KM travelled and Price charged
- cost of trip and price charged
- profit and profitPerKM
- profit and Price charged
- cost of trip and Km travelled

Exploratory data analysis

Final conclusions

Since we have analysed our data, we can test some of the hypothesis we had formulated:

- Cab services are the most popular in New York and the least popular in Pittsburgh
- Yellow cab charges more per KM
- Males are generally the most prevalent Cab services users but there is generally no prefernce of cab service
- Young adults and adults are the most frequent cab service users but there is generally no prefernce of cab service
- low and middle income induvuals use cab services more than those of high income but there is generally no prefernce of cab service
- The preffered mode of payment by customers is card
- There is some seasonality in the cab service industry as the most frequented months for cab services are December and October

Exploratory data analysis

Recommendations

I would recommend the firm to invest in Yellow Cab due to the following reasons:

- Yellow cab has proven to have a better city outreach compared to Pink cab
- Yellow Cab has been consistently making more profit throughout the given time period
- Yellow Cab is preffered by the younger generation, which in most cases is the most populous

Data Glacier

Thank You