Exploratory Data Analysis
G2M Cab Investment case study

26/06/2021

By William Okomba
Virtual-internship

Background

Cab service is a booming industry in US, there are several companies offering the service where they carry passengers from point A to B, there are rules and regulations that plays a big role in their operations.

Some cabs are restricted to pick and drop passenger enroute, and offer airport services while others are allowed to do so.

The most common cab companies are yellow cab, and pink cab with operates in various cities across the US.

The industry is affected by both holidays as well as weather patterns.

Problem Statement

XYZ is a private equity 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. There is a need to identify the most profitable company where the XYZ can invest an recoup profit, and as per their Go-to-Market(G2M) strategy they what to understand the market before taking final decision

Objectives

1. To identify the right company to invest in.

   Other objectives:

2. To identify the most profitable company.

3. To identify the market segment where it is more viable to invest in.

Data Description

We were provided with 4 datasets, and added one external dataset.

1. **Cab_Data.csv –** this file includes details of transaction for 2 cab companies

2. **Customer_ID.csv** – this is a mapping table that contains a unique identifier which links the customer's demographic details

3. **Transaction_ID.csv –** this is a mapping table that contains transaction to customer mapping and payment mode

4. **City.csv –** this file contains list of US cities, their population and number of cab users

5. **Us_holidays.csv** - contains the list of US holidays, non-holidays days, and the days of the week.

Experimental Design

- upload and read our csv files

- clean our dataset

- perform EDA

- Provide observations, conclusion, and recommendations.

Procedure

Clean 5 datasets.

1. Cab dataset timeframe is 01/01/2016 to 12//31/2018.

2. Outlier in price charged variable was removed.

Merging Dataset

1.The 5 datasets were merged to form one final merge dataset. The final dataset has 23 variables and 353434 records.

2. Data Exploratory Data Analysis
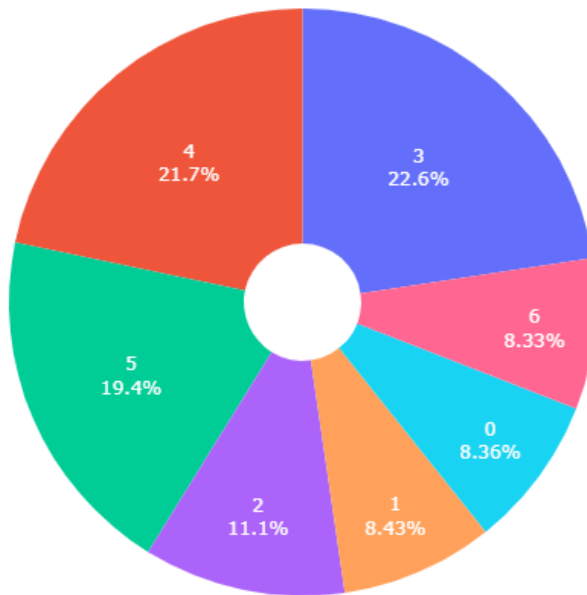
3. Conclusion

4. Recommendation.

Univariate Analysis

Univariate analysis- company variable
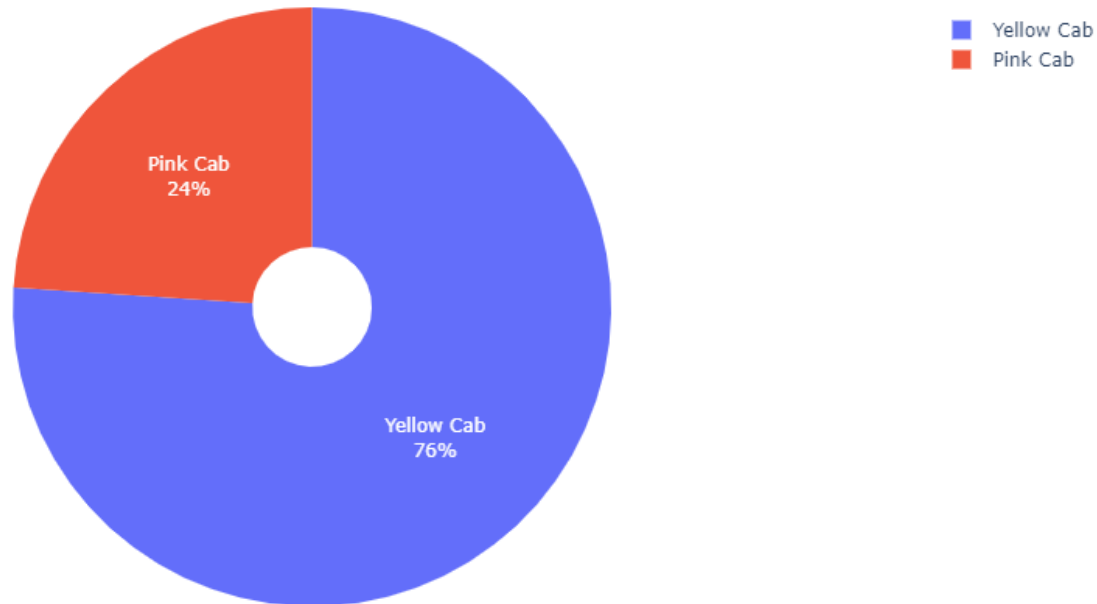
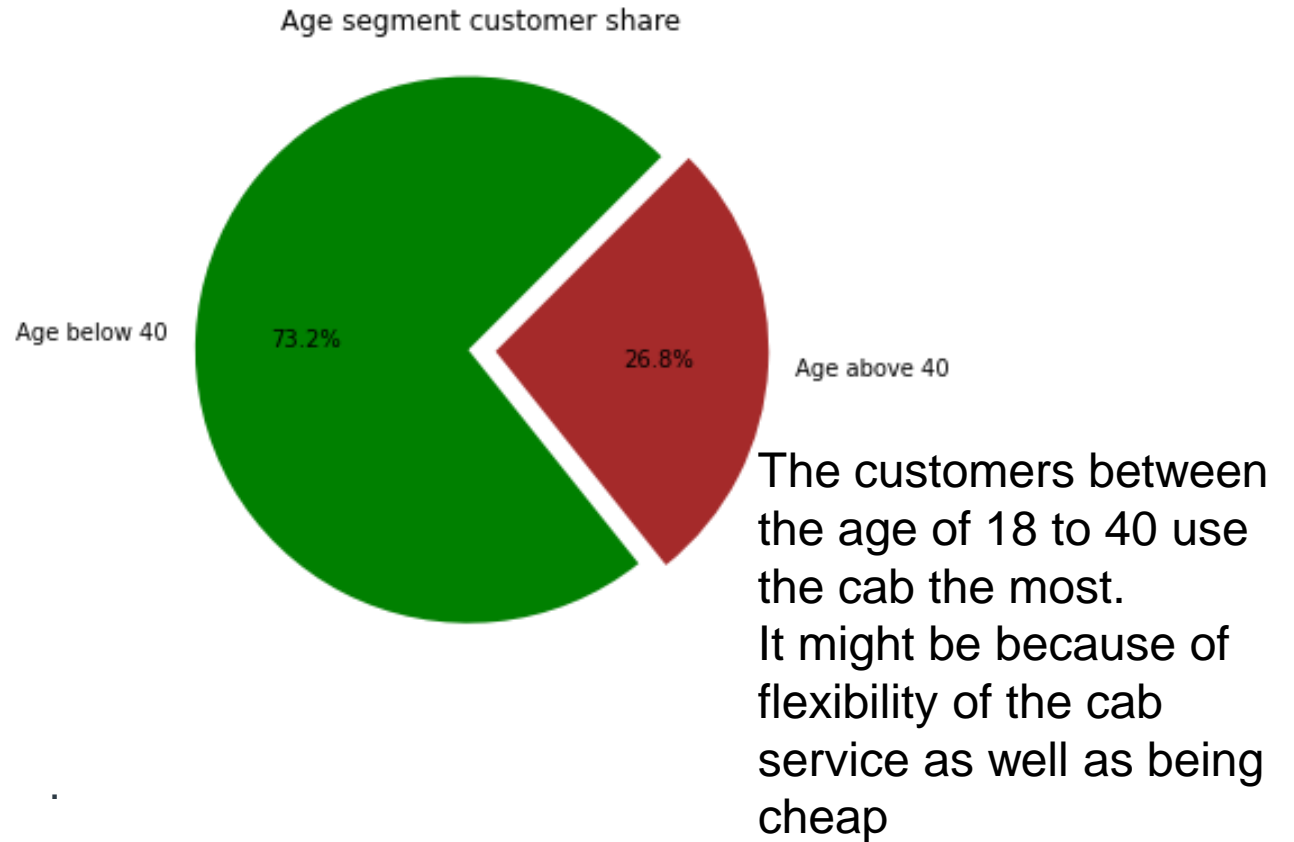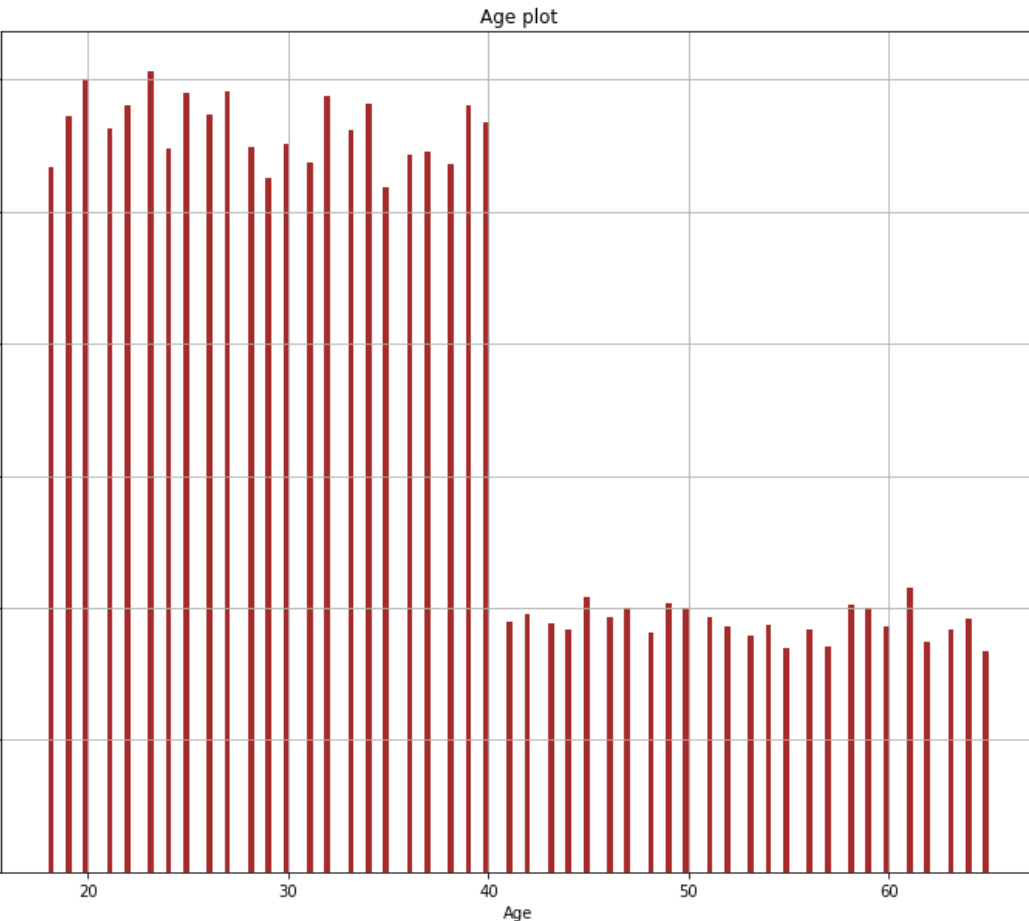Yellow cab dominate the customer usage by 76 % compared to it's rival.

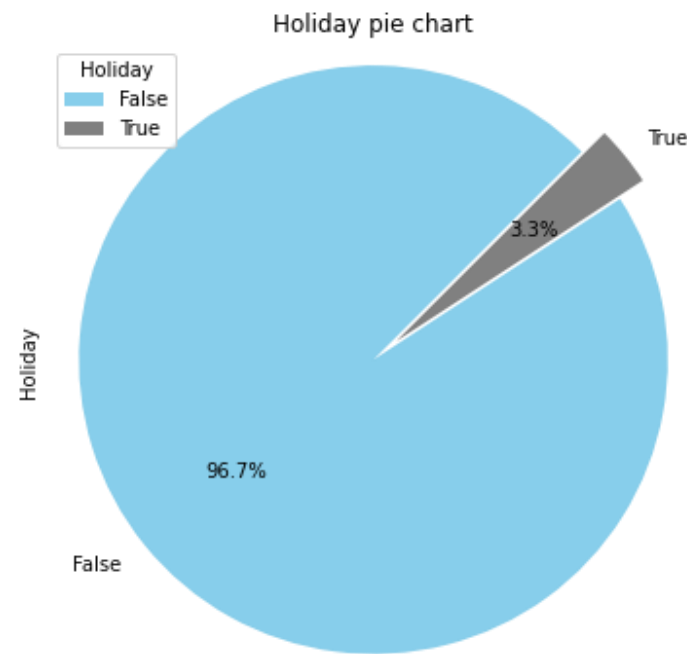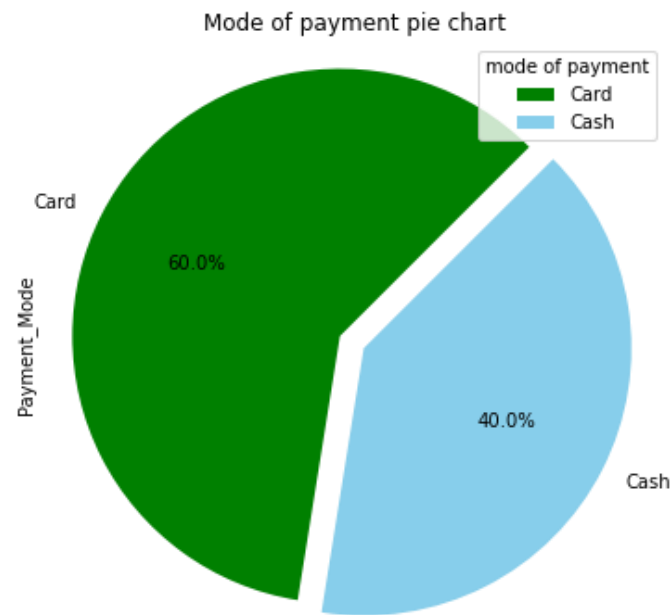# Univariate analysis: weekday variable

Monday = 0, **Tuesday** = 1, **Wednesday** = 2, **Thursday** = 3, **Friday** = 4, **Saturday** = 5, and **Sunday** = 6.

•Most of the customers prefer to use taxi on Thursday(3) and Friday(4), this is because this days near the weekend.
•Day 6(Sunday) is the lowest, this is because it is a weekend and most of the customers prefer spending their time at home.

Univariate Analysis: Age variable

Age plot

Age segment customer share

Age below 40 — 73.2%

Age above 40 — 26.8%

The customers between the age of 18 to 40 use the cab the most.
It might be because of flexibility of the cab service as well as being cheap

Data Glacier

Univariate analysis: mode of payment and holiday variable

Mode of payment pie chart

mode of payment
- Card
- Cash

Card 60.0%
Cash 40.0%

Payment_Mode

Holiday pie chart

Holiday
- False
- True

True 3.3%
False 96.7%

Holiday

1. Customers prefer card mode of payment, this can be attributed to safety.
2. Customers use cab service more in non holidays days.

# Bivariate Analysis: Company vs Gender, and Company vs Year variable.
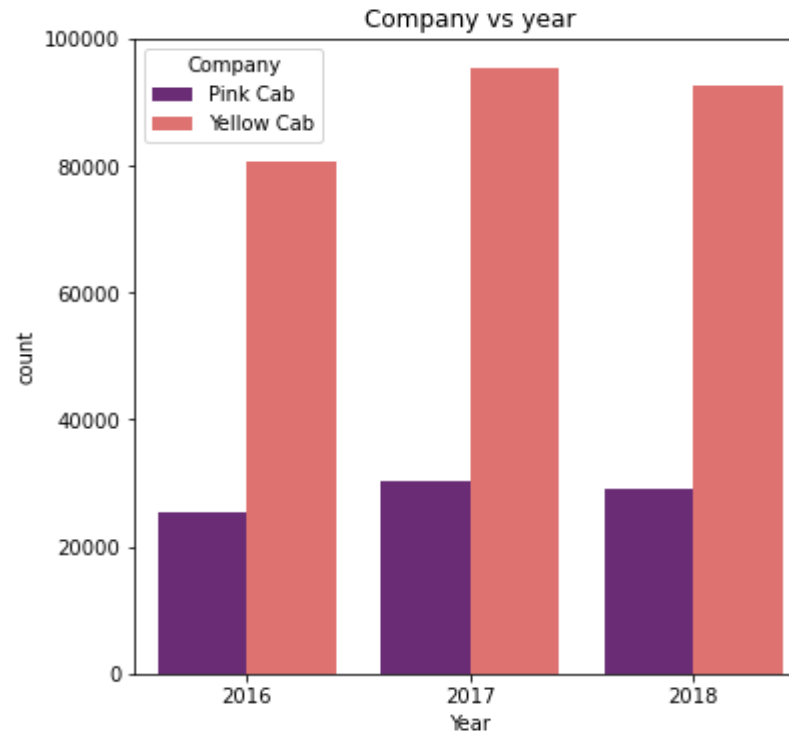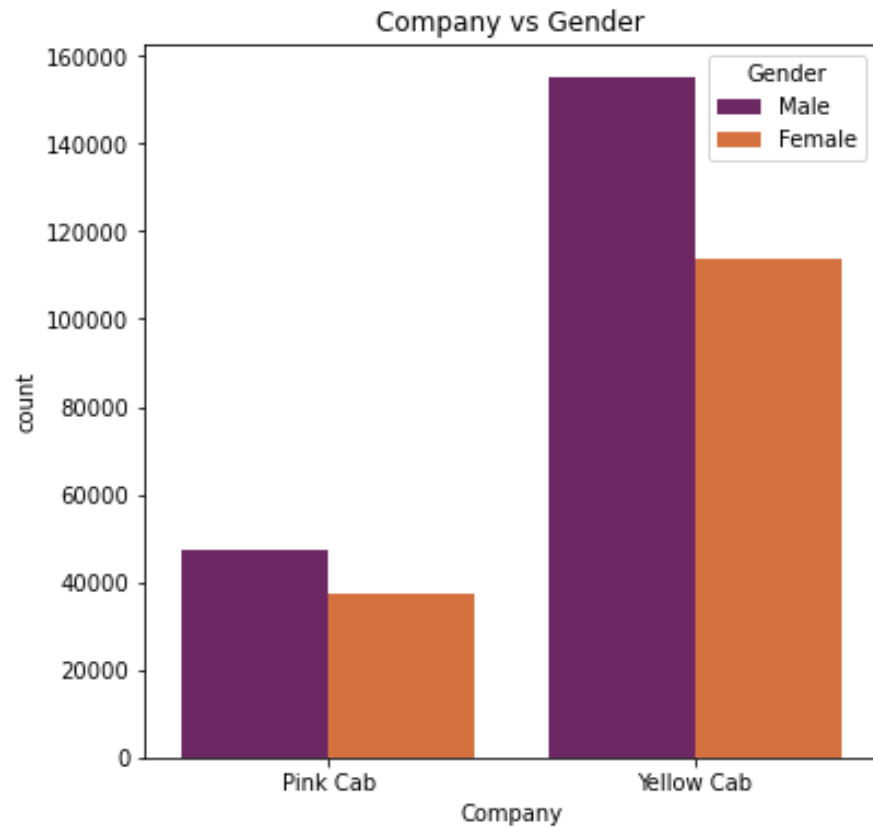
1. yellow cab is preferred with both male and female customers compared to pink cab.
2. Male uses both cabs more compared to female customers.
3. Customer used cabs in 2017 more compared to other years
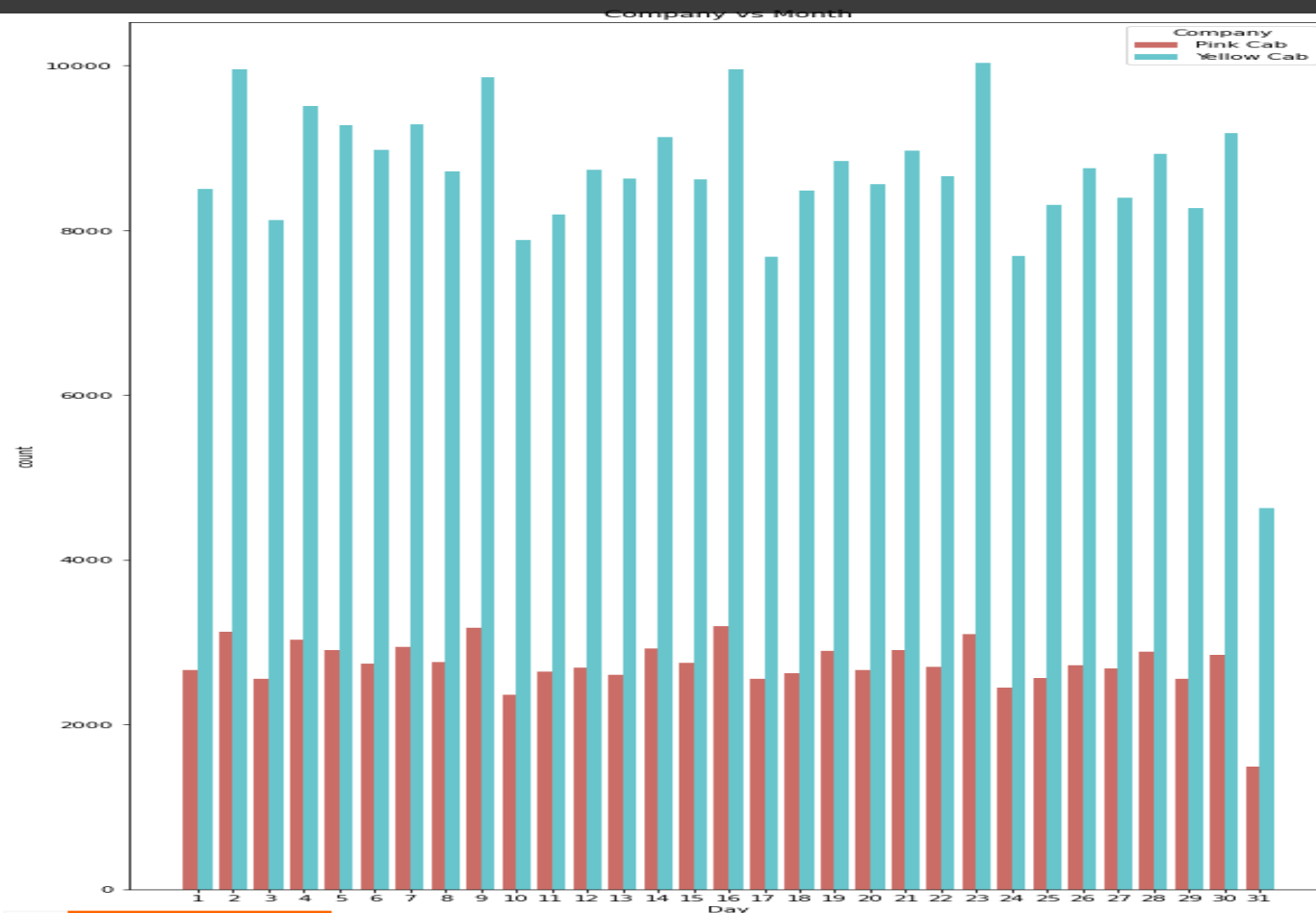
Yellow cab is used more than pink cab all month round. on the second day of the month, mid month 16th almost the end of the month customers use the cab more, this can attributed to pay check days thus they have money to spend.

# Bivariate analysis: Company Vs Month, and Company vs Weekend.
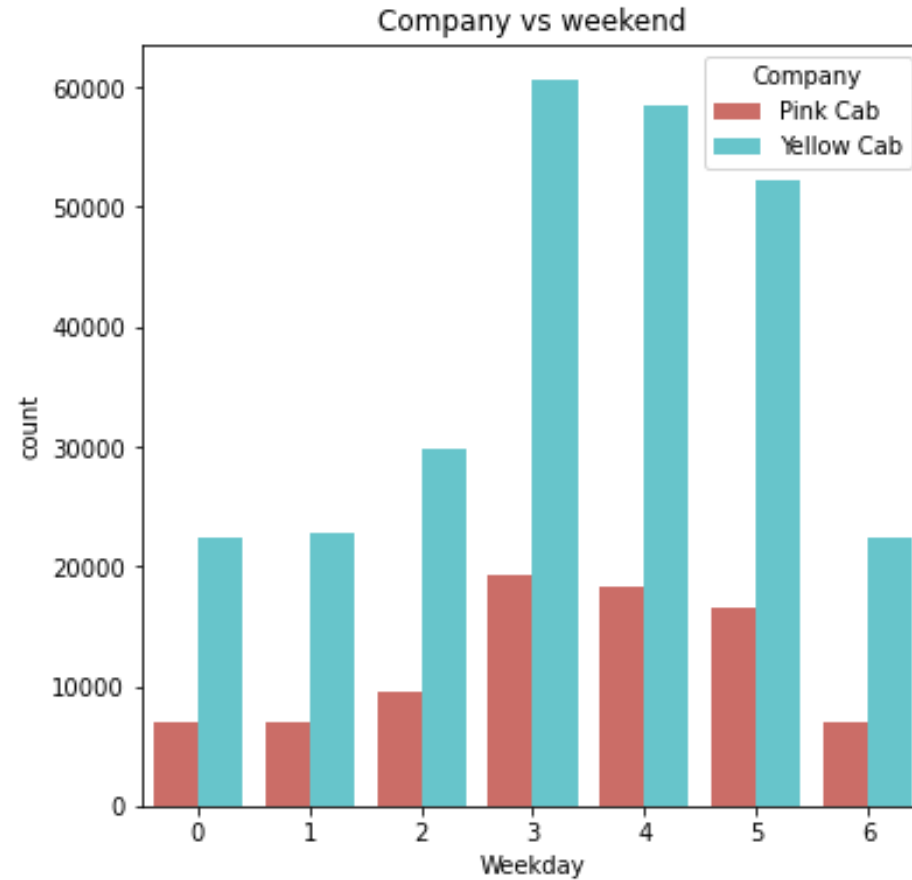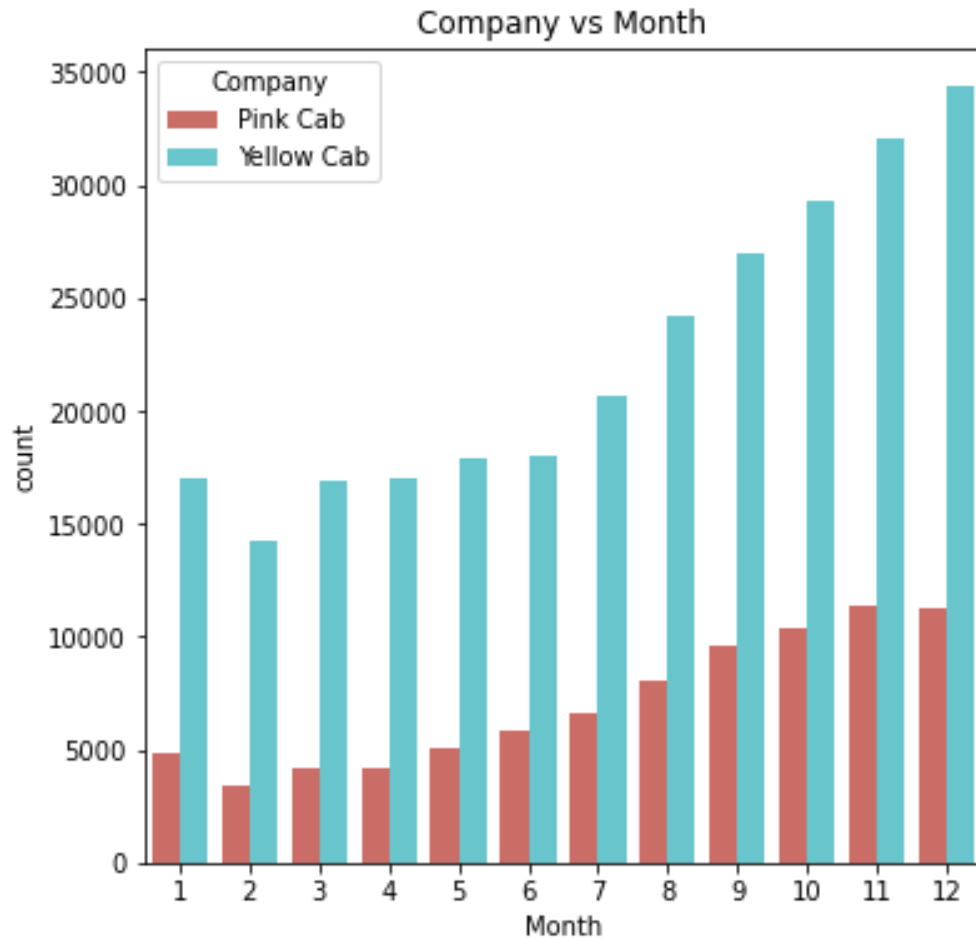
1. customers use the cab more at the of the year, this can be attributed to people havening money they have been saving all year round
2. Both yellow and pink cabs are on high demand near the weekend on Thursday, Friday, and Saturday.

Bivariate Analysis: profit per year.

Profit per trip

2017 registered higher profit both cab companies.
Yellow cab realized bigger profit for all the 3 years compared to the pink cab.

Data Glacier

Bivariate analysis: Relationship between variables.

correlation betwen features
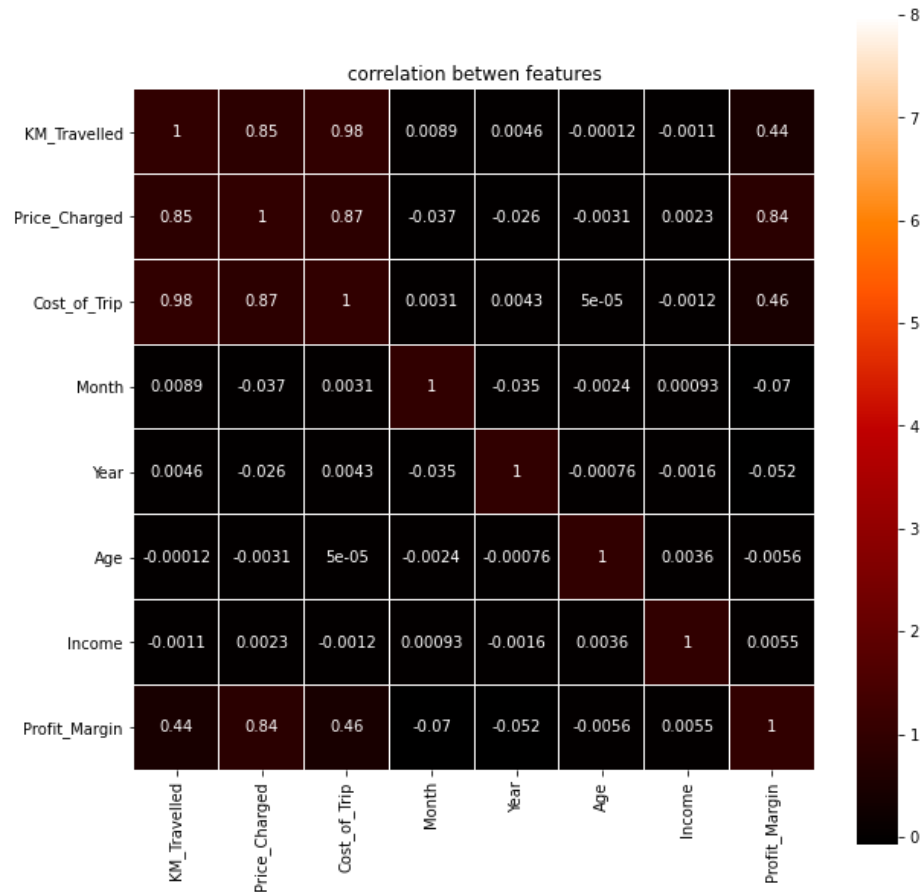
- There is a strong positive correlation between price charges, kilometer travelled, cost of trip, and profit margin.
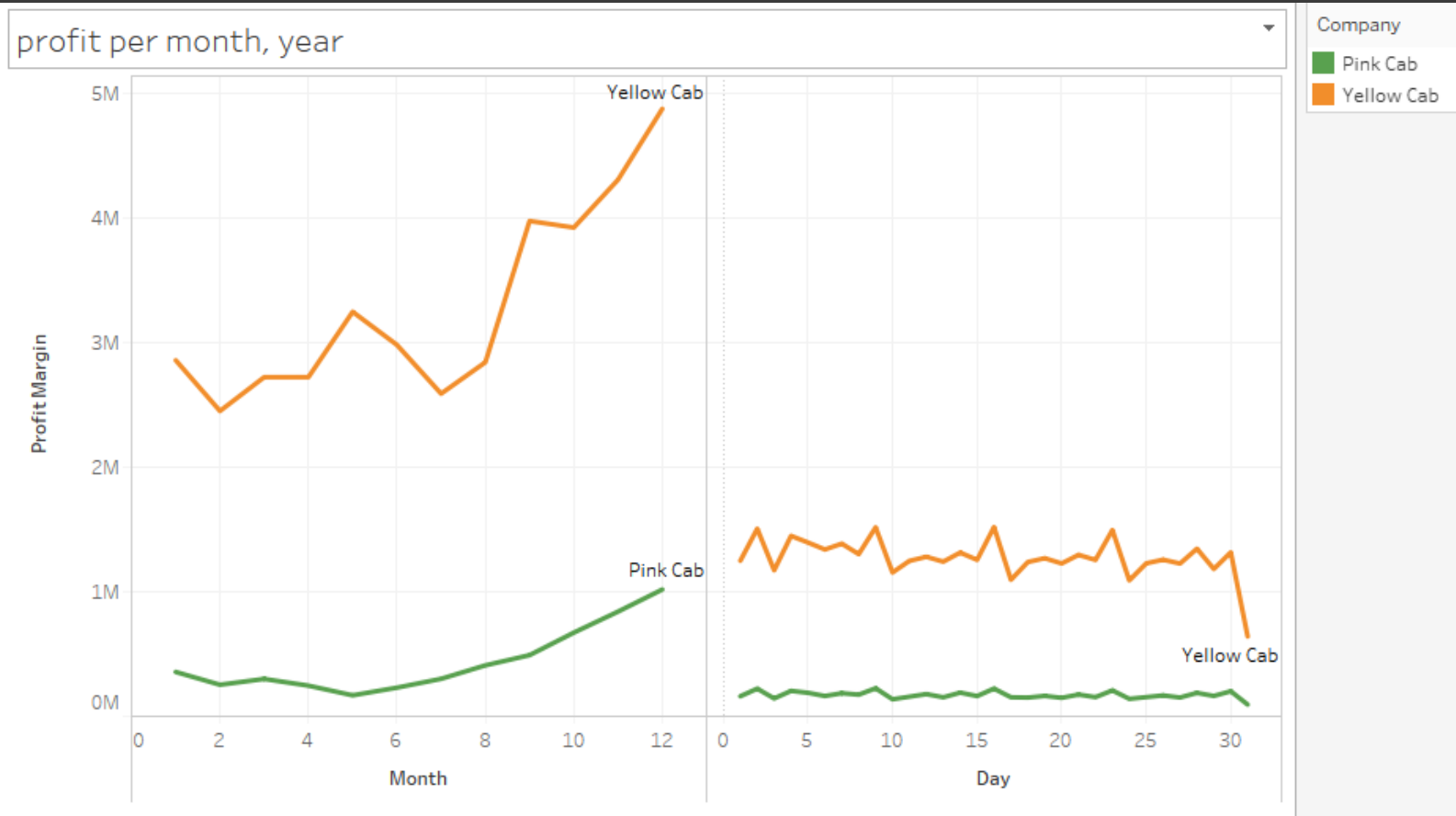- There is a negative correlation between km travelled, age, income, and month etc.

Seasonality
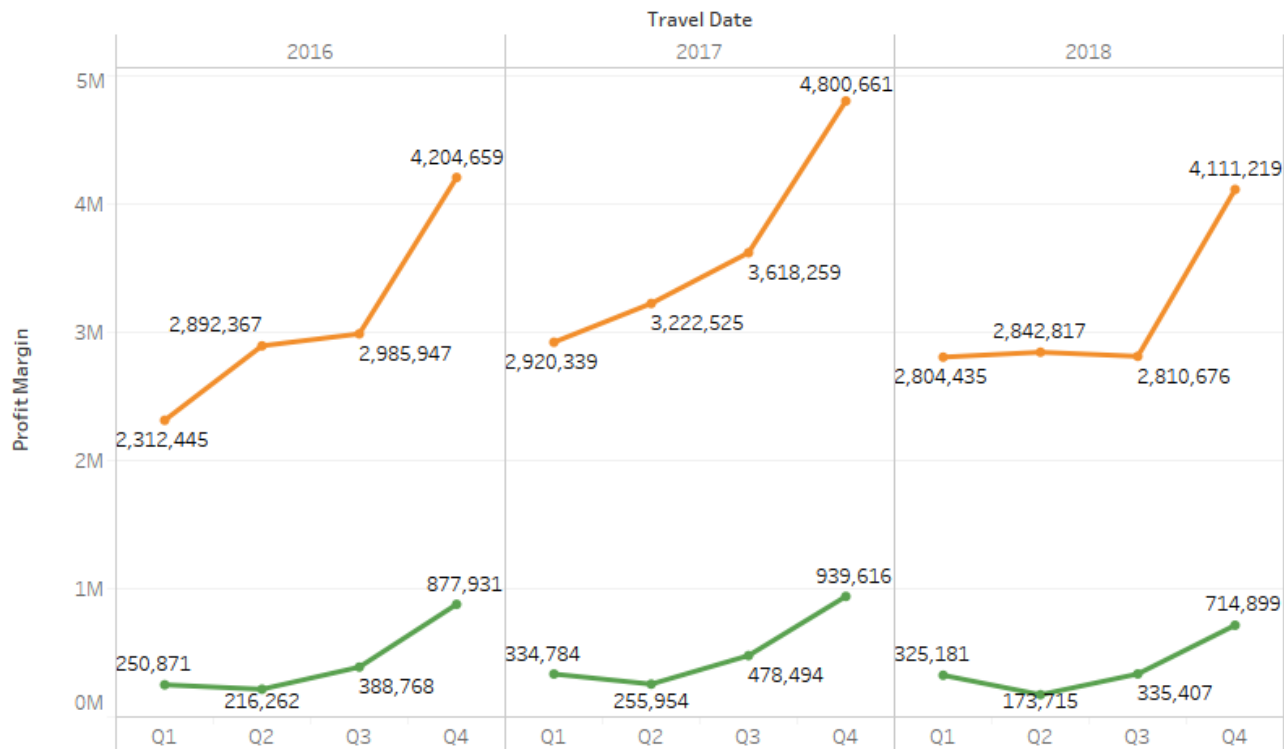
There is a sharp increase in the 4th quarter profit for all the years. But there is a higher increase in profit for yellow cars compares compare to pink cars.
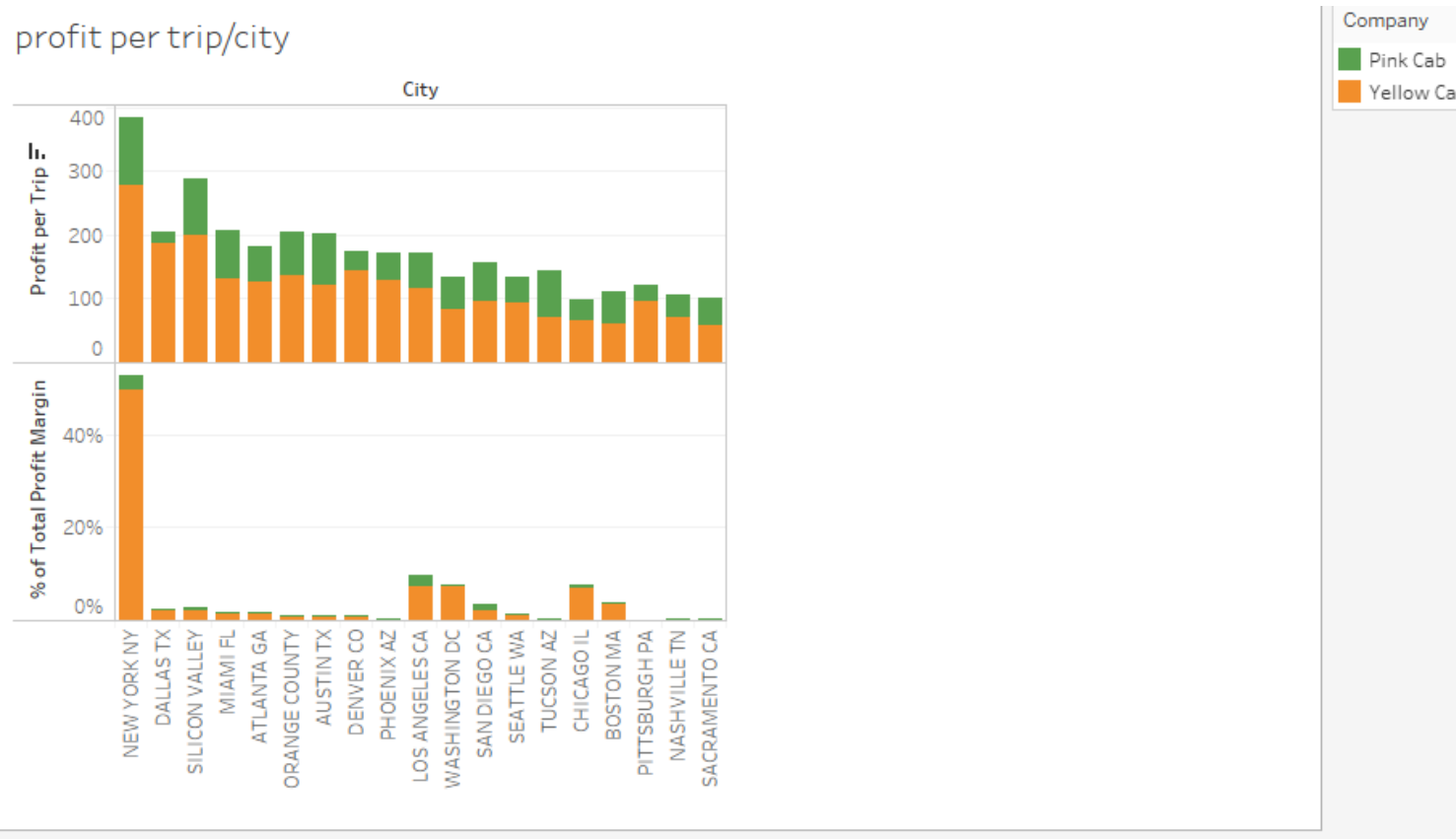
Data Glacier

Profit trend.

1. Profit tend to increase towards the month of December, This can be attribute to people have enough money to spend.

2. Profit fluctuates all month long where the usage is slightly higher on a certain dates.

Multivariate analysis: Profit per trip per company, and profit per company

profit per trip/city

Yellow cab leads in profit per trip, this makes the yellow cab to lead in total profit margin all years round.

Data Glacier

Conclusion

- pink cab has small market share across the cities but it is present in all major cities.

- yellow cab has a giant share of the market especially in New York and it enjoys the preference by customers compared to pink cab.

- Yellow cab makes a big profit when the travel short distance as compared to long haul. it seems this is a perfect strategy since customer can board and stop at any place, besides they have access to all the major airports.

- customers don't use cab service much on holidays, but there is a slight increase on use on long holiday like charismas break.

- customers prefer paying with the card as compared o the cash, this can be attributed to credit card payment preferences among customers as a mode of payment.

Recomenation

- Investor should go ahead and invest in yellow cab since it has a large-solid market, good strategy to attract customers as well versed on how to enjoy the dominance all year round.

- cab can expand their market by targeting other big cities where yellow cab has less dominance.
- yellow cab should continue with short distance strategy if they need to make more profits.

- Government policies should favors pink cab if they need to expand since most cities are saturated by the yellow cab.

Thank You