Exploratory Data Analysis
G2M insight for Cab Investment firm

Name:Madoka Fujii

July 21, 2023

Agenda

Data Glacier

Your Deep Learning Partner

Executive Summary

Our Client XYZ is planning to invest a cab service company because recently the market is growing.

-Objective: Using the two cab company as examples in dataset, provide the findings and insights to them to know and understand the industry and the market for the preparation of investment.

-Out Client XYZ needs to know the important key points to identify which cab company is right and profitable to invest.

-We provide our analysis and insights so that Our Client can invest a right company.

-Yellow Cab is larger business size than Pink Cab in terms of the numbers of customers, the profit, locations, and transactions.

Problem Statement

Our Client does not know what kind of cab service is worth to invest.

We provide recommendations, analysis, and insights to them to choose right cab service.

Approach

-Merge and join 5 data sets to analyze.

    -Cab_Data.csv

    -City.csv

    -Customer_ID.csv

    -Transaction_ID.csv

    -US_Holiday_Dates.csv

-Data cleaning.

-Data type check to analyze.

-Explorer Data Analysis.

-Creation Profit variable.

-Profit Analysis.

-Seasonality Analysis.

Merge and Join 5 datas

-There were no missing values so no needed to impute or drop for that.

-Transaction IDs that had duplications. I deleted.

-The duplication of Customer IDs were same as Transaction IDs.

-Changed data types of Date of Travel, Population, and Users.

-Used left join method that can check any missing values to understand datas.

Data Type after Cleaning

```
Transaction_ID                     object
Customer_ID                        object
Payment_Mode                       object
Date_of_Travel             datetime64[ns]
Company                            object
City                               object
KM_Travelled                      float64
Price_Charged                     float64
Cost_of_Trip                      float64
Gender                             object
Age                                 int64
Income_(USD/Month)                  int64
Population                         float64
Users                             float64
Holiday                            object
WeekDay                            object
Profit                            float64
Profit/Cost_of_Trip               float64
KM_Travelled/Cost_of_Trip         float64
Year                                int64
Month                               int64
Day                                 int64
dtype: object
```

Comparison of Profitability
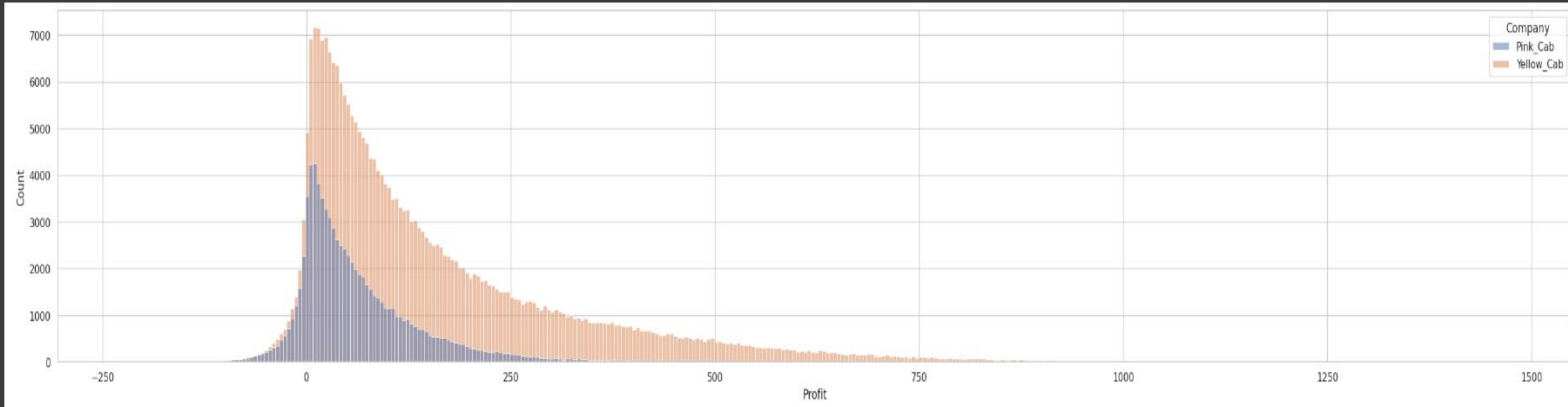
-Yellow Cab (Orange) makes profits significantly higher than Pink Cab (Blue).

-Pink Cab has some minus profit while Yellow Cab has almost no minus profit.

-The count transactions of Yellow Cab is about double of Pink Cab's. That contributes more profit.

-Pink Cab's profit is limited according to the graph. But Yellow Cab extends the profit that we can see the long tail in left side.
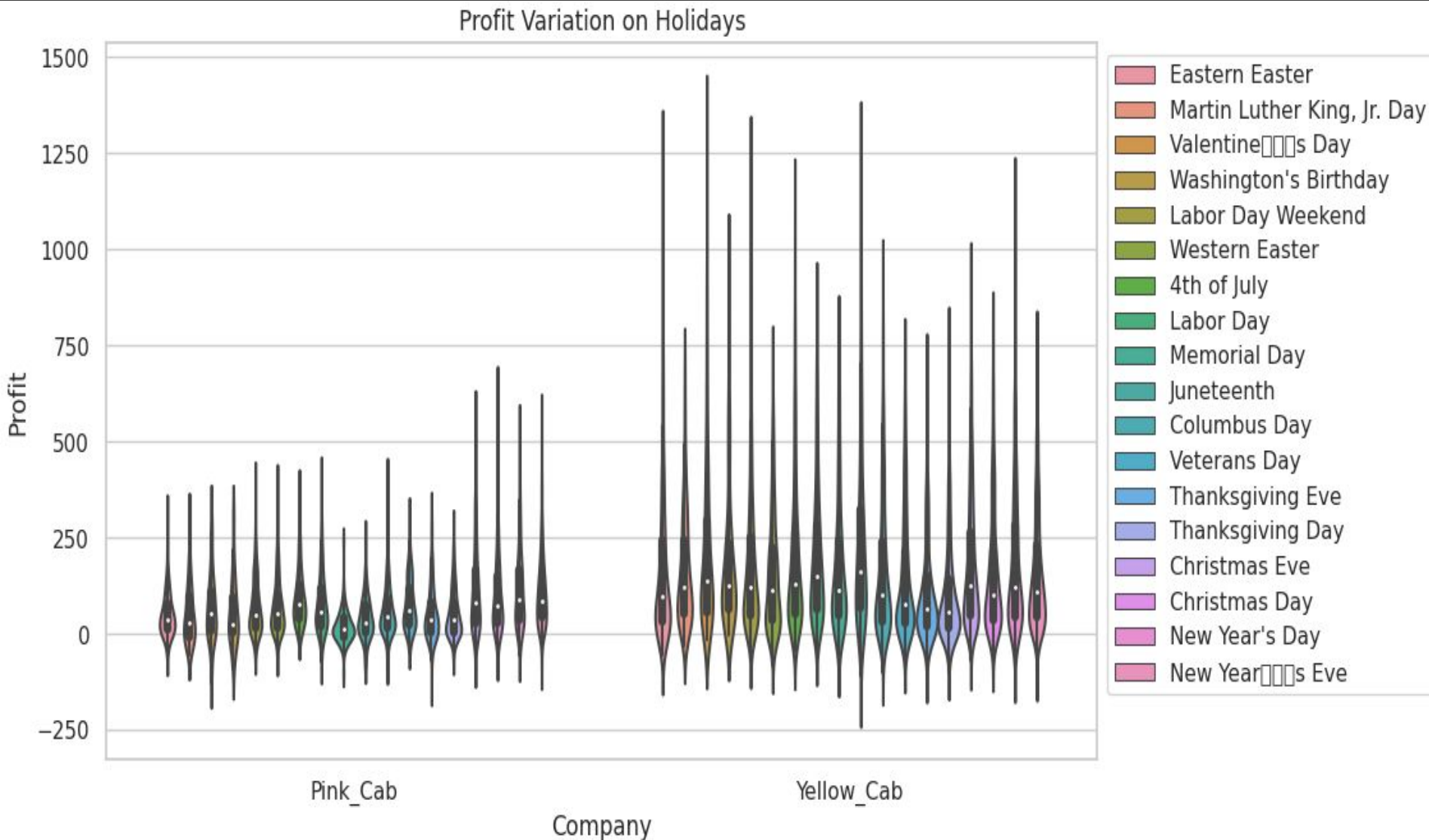
Average Profit Distribution by Citi

Profit Distribution by Company and City
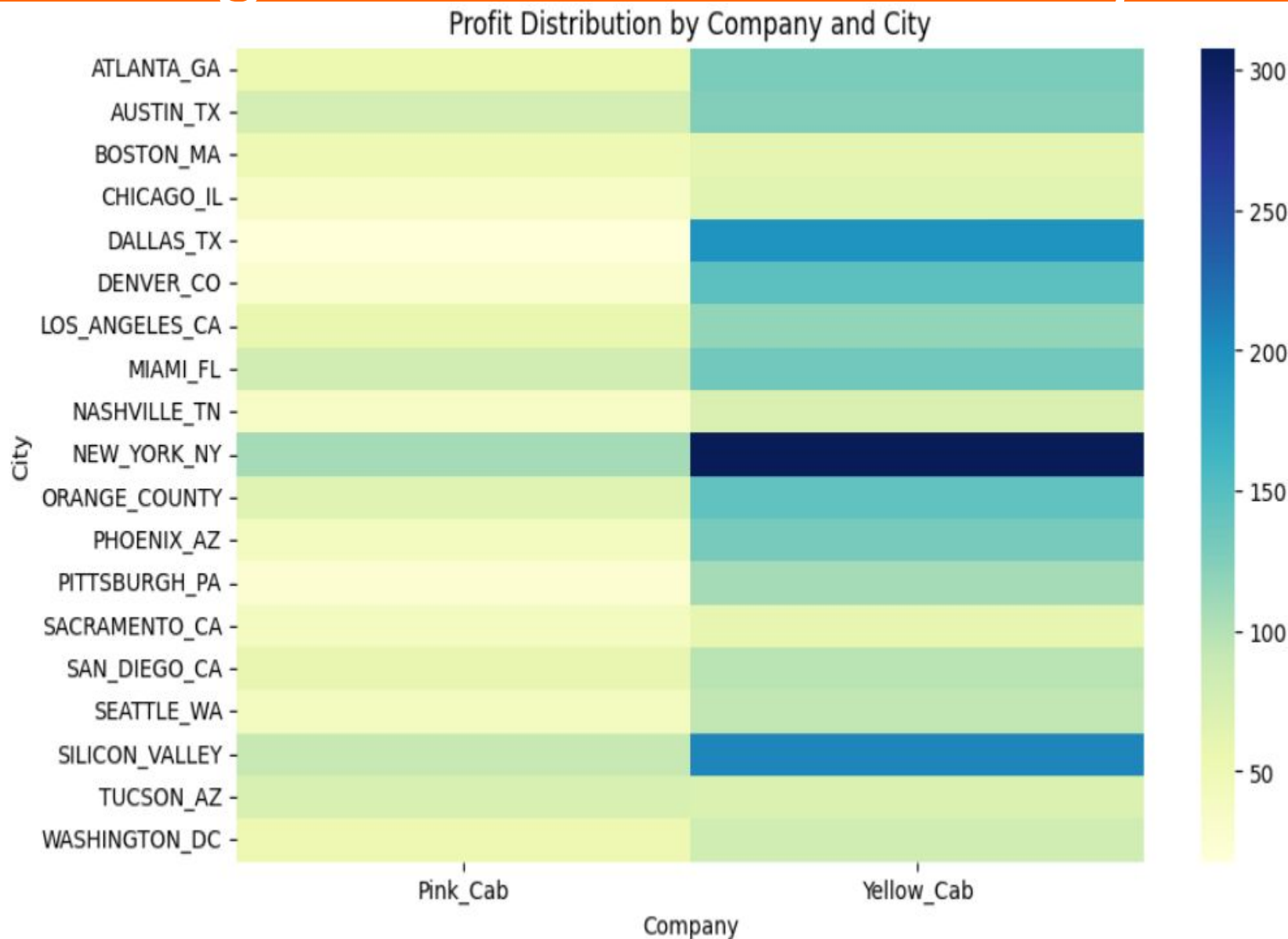
-The highest profitable city is New York with Yellow Cab as about average profit $300 per transaction.

-The second profitable city is Silicon Valley with Yellow Cab as about average profit $250 per transaction.

-The highest profitable city with Pink Cab is also New York as about about profit $150 per transaction but that is not really profitable compared to profit of Yellow Cab in many cities.
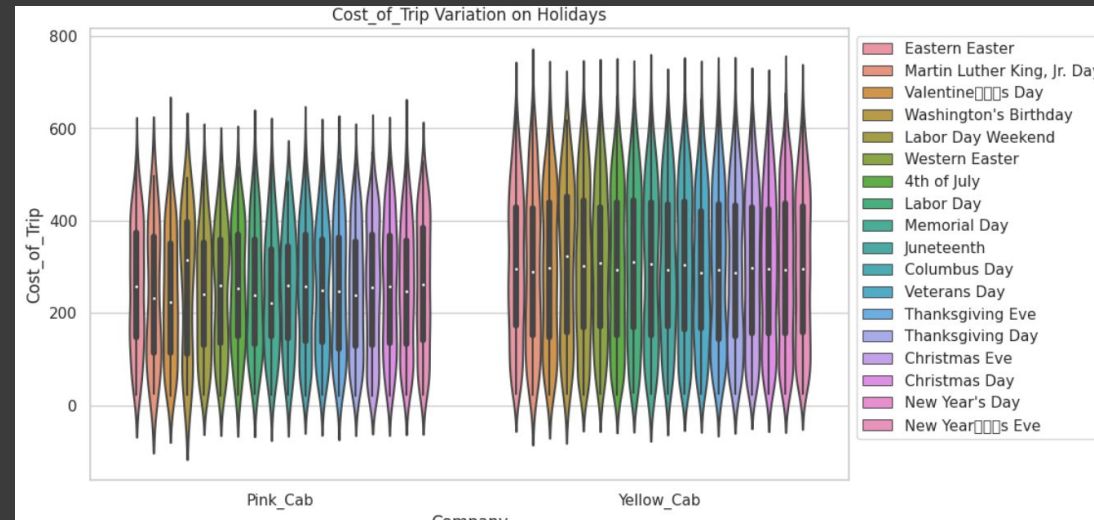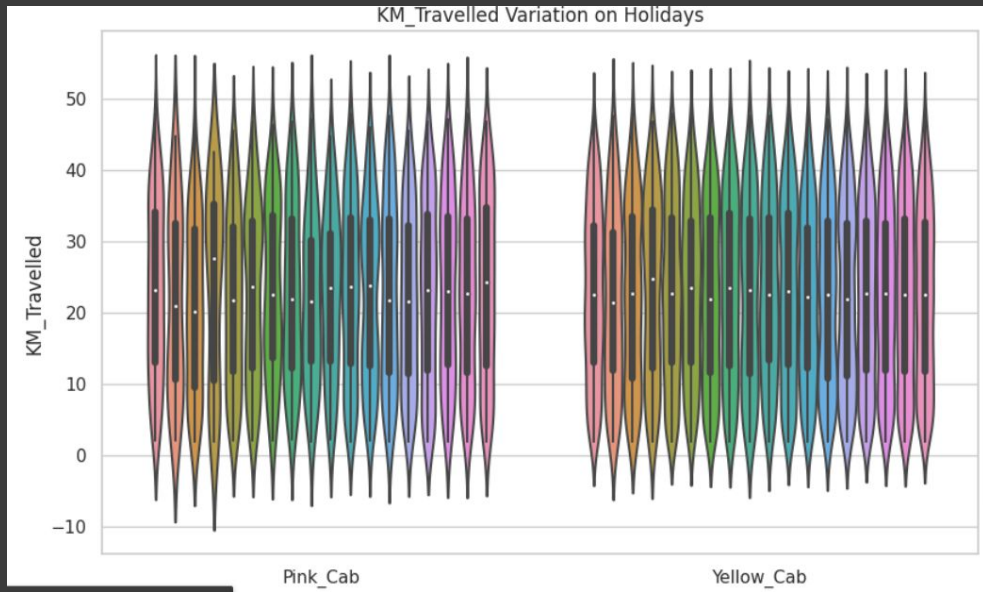
Profit on Holidays

Profit Variation on Holidays

Legend:
- Eastern Easter
- Martin Luther King, Jr. Day
- Valentine□□□s Day
- Washington's Birthday
- Labor Day Weekend
- Western Easter
- 4th of July
- Labor Day
- Memorial Day
- Juneteenth
- Columbus Day
- Veterans Day
- Thanksgiving Eve
- Thanksgiving Day
- Christmas Eve
- Christmas Day
- New Year's Day
- New Year□□□s Eve

-In any holidays, Yellow Cab is more profitable than Pink Cab. Yellow Cab has large amount of customers compared to Pink Cab. That is proportional to profit in holidays as well.

-Pink Cab's large profit are from Christmas, the Eve, New Year Eve, and New Year. On the other hand, Yellow Cab's profits are not specified with specific holidays.

Holidays Margin Analysis

-In Holidays, both companies' KM_Travelled are almost same. However, Pink Cab charges not much while Yellow Cab charges a lot. As a result of this difference, the profit of each company becomes significantly different.

-The cost of trip of Yellow Cab is more than Pink Cab. It seems Yellow Cab provides better service to customers compared to Pink Cab.
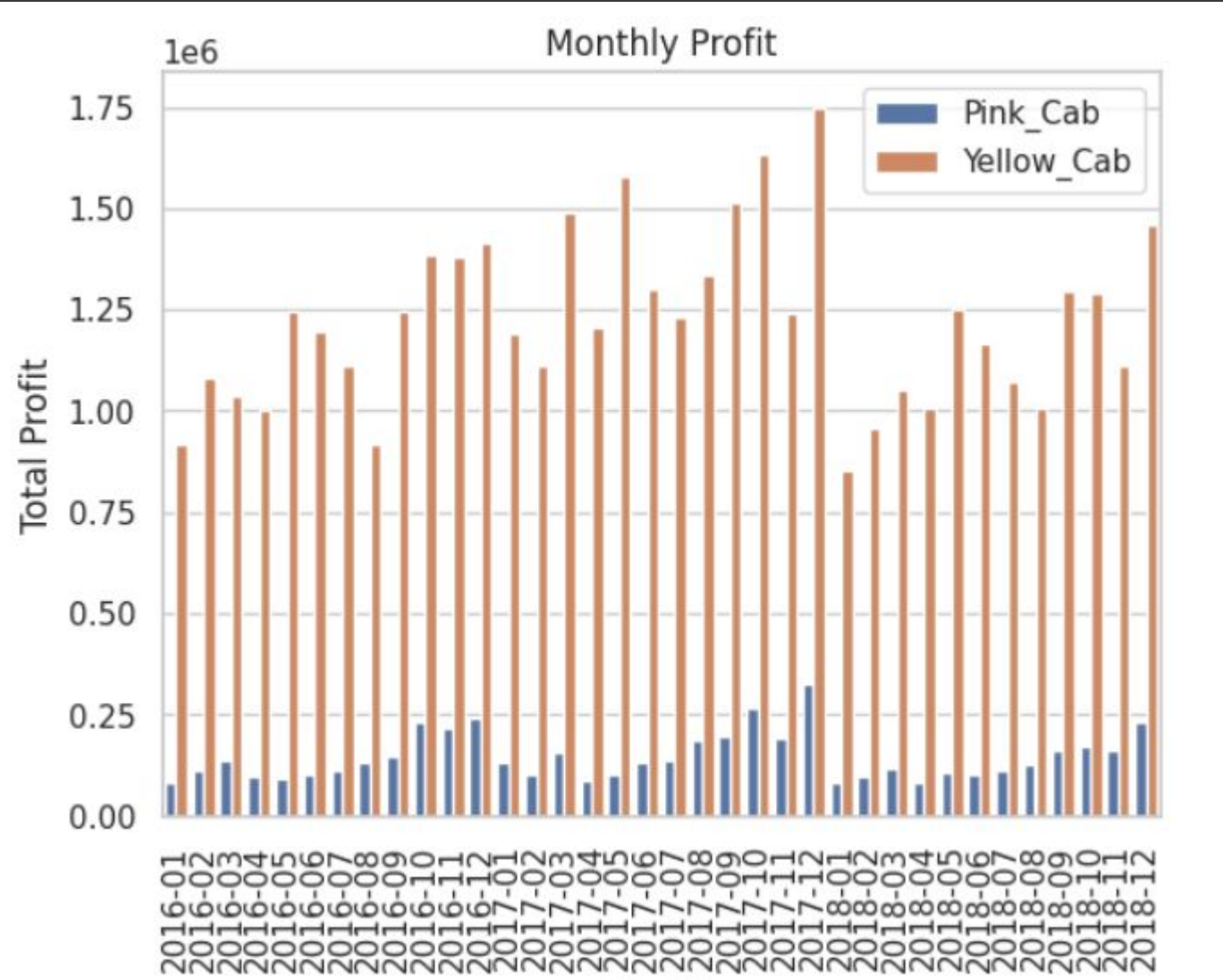
Yearly Profit
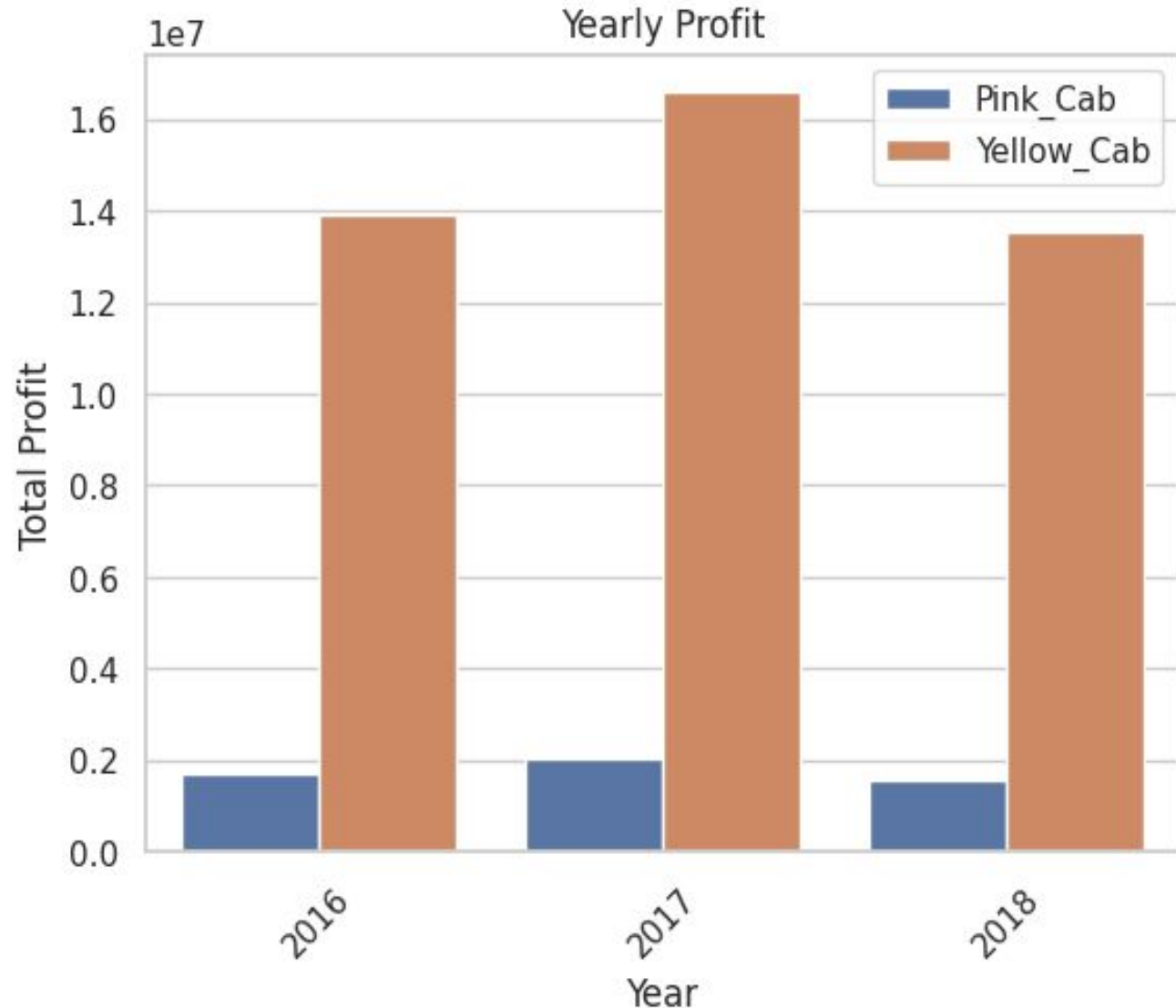
The yearly profit of Yellow Cab is significantly higher than Pink Cab's in every year.

-However, the highest profit of both company's profit is in 2017.

-In 2018, their profit becomes decreasing even lower than the profit in 2016.
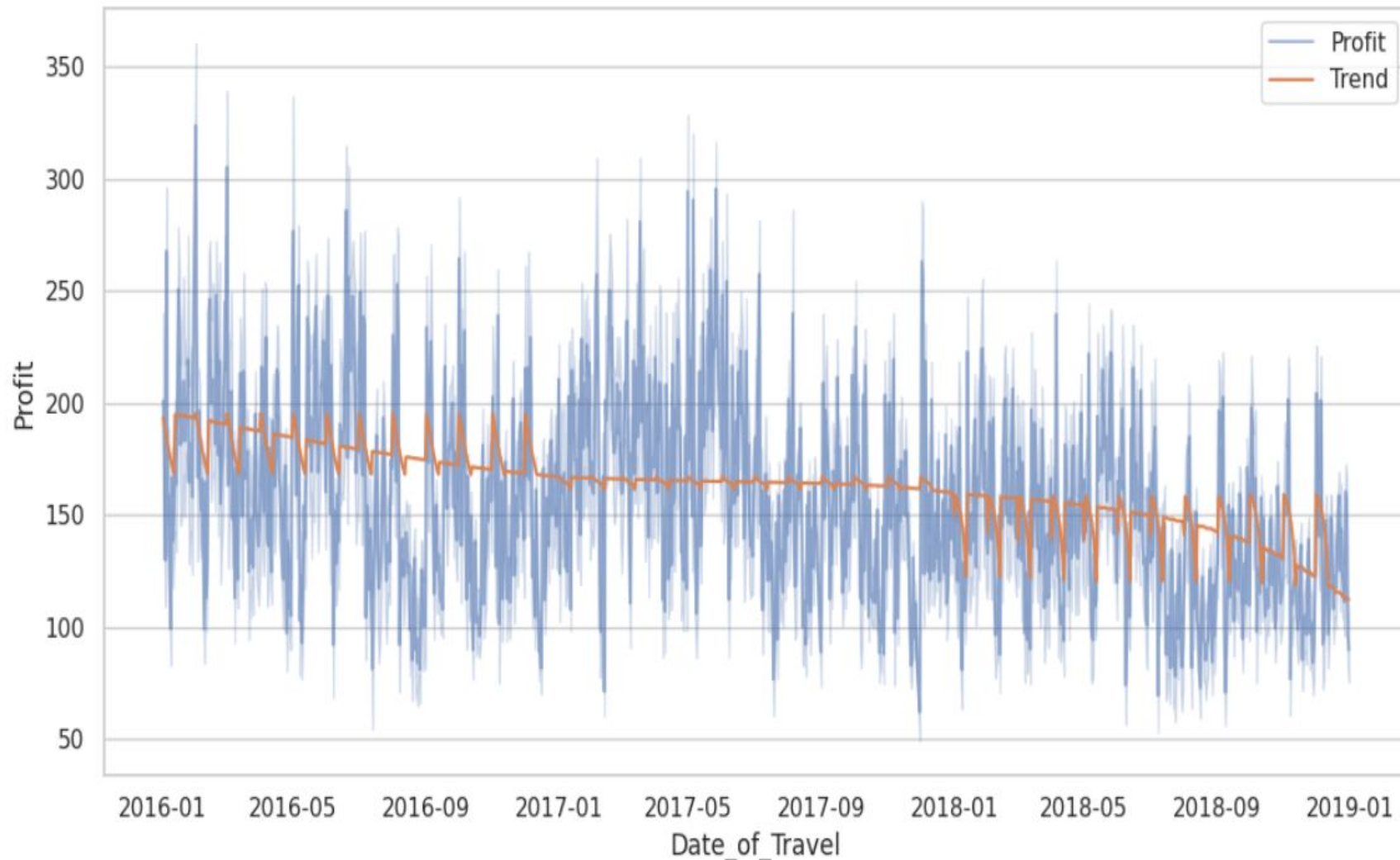
Monthly Profit

-Yellow Cab's monthly profit is larger than Pink Cab's. It seems the amount of profit is about more than 7 times different.

-There are trend both company. When the season is getting the end of the year, the profits are increasing.

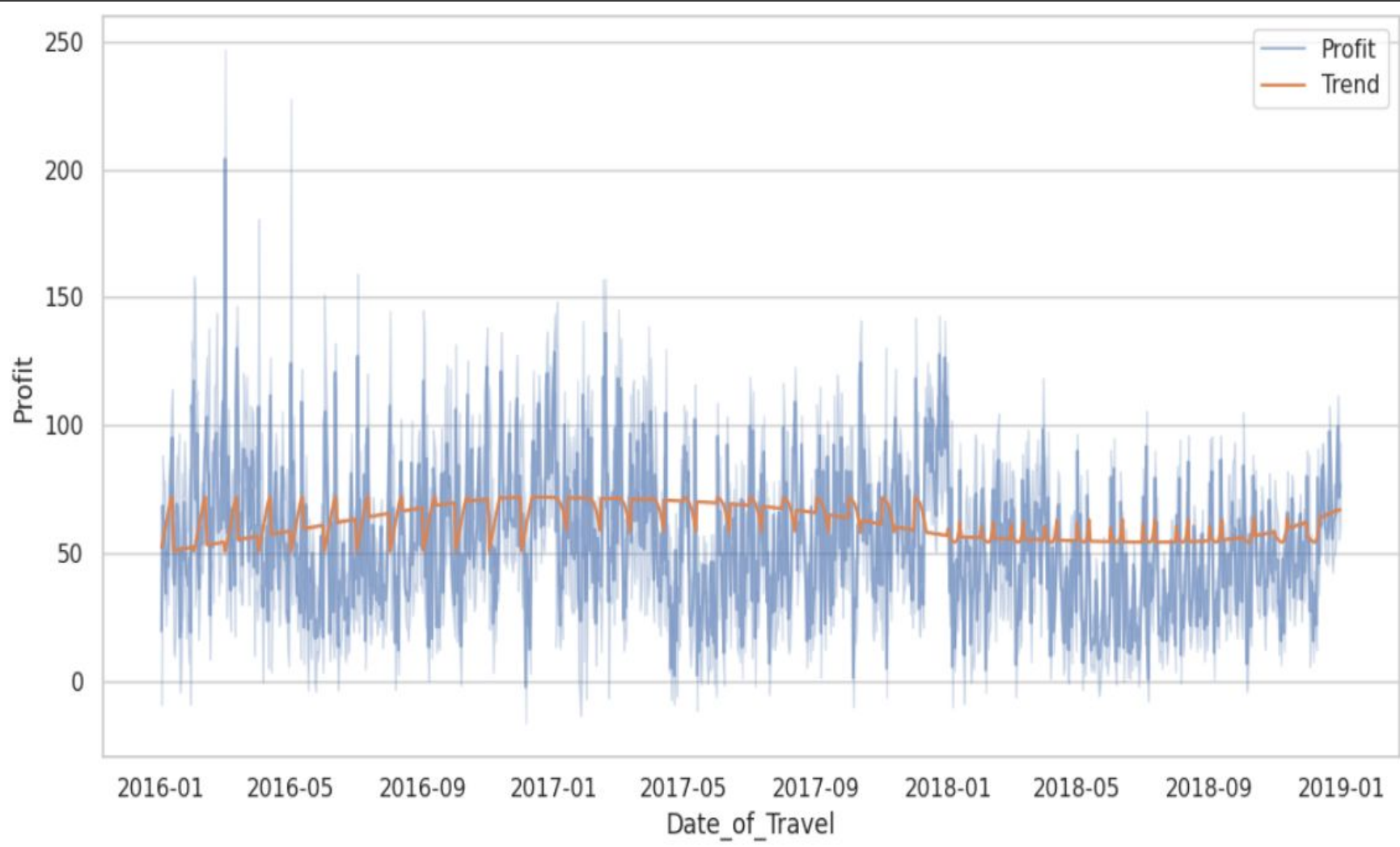-We can see the month profit in 2018 of Yellow Cab was decreasing.

Profit with Seasonality of Yellow Cab

-Used time series with seasonality method.

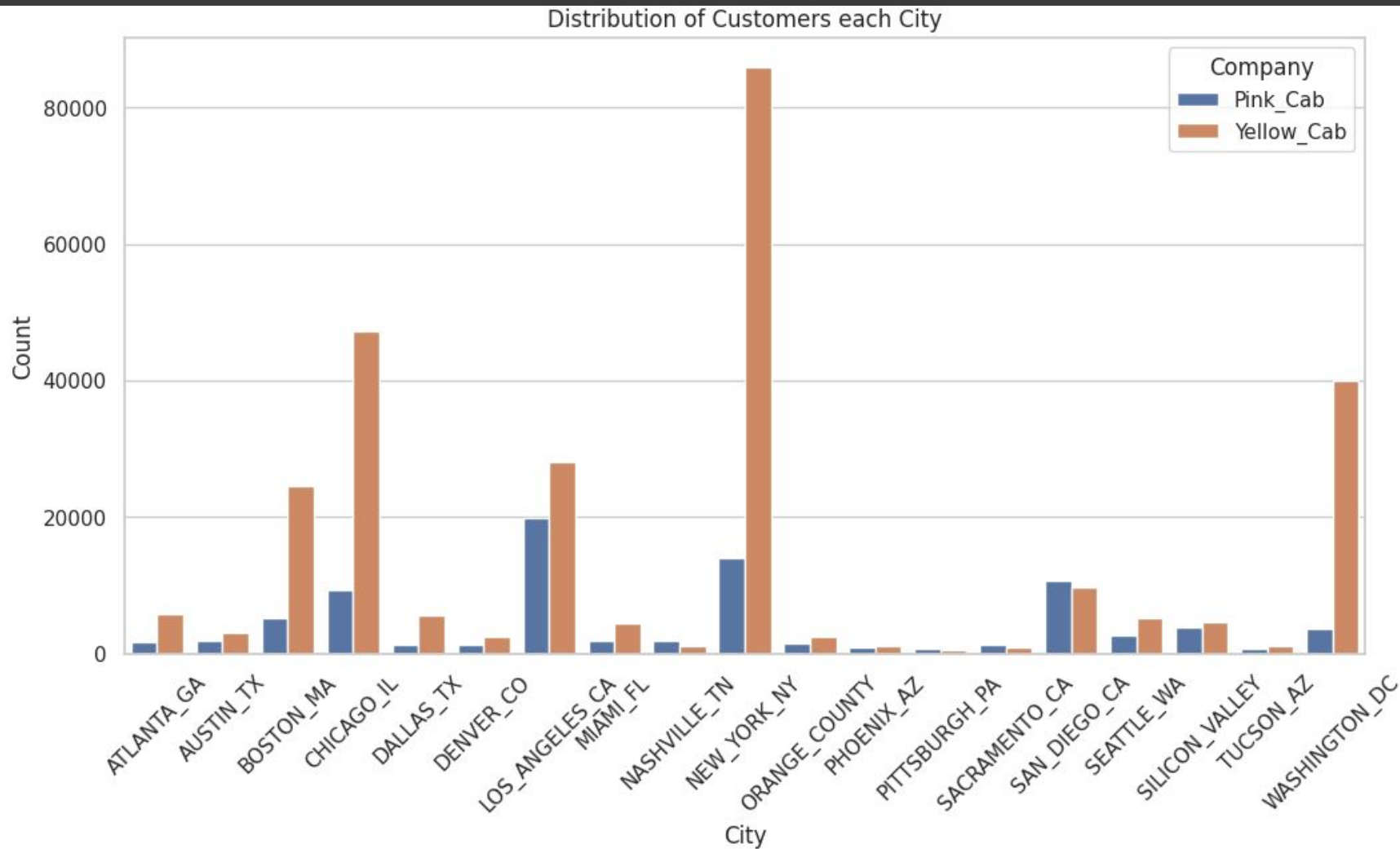-The profit is slowly decreasing from 2016 to 2019.

Profit with Seasonality of Pink Cab

-Compared to Yellow Cab, Pink Cab's profit is a little better.

-This is because Pink Cab's profit is not fractuated so much as Yellow Cab. So, impact from volatility is not big as Yellow Cab.

Distribution of Usage each City

Distribution of Customers each City

-In Yellow Cab, the location with most highest usage is New York. On the other hand, in Pink Cab, it is Los Angeles.

-But Pink Cab's usage in Los Angeles is lower than Yellow Cab's.

-In some cities, Pink Cab has more usage than Yellow Cab as below.
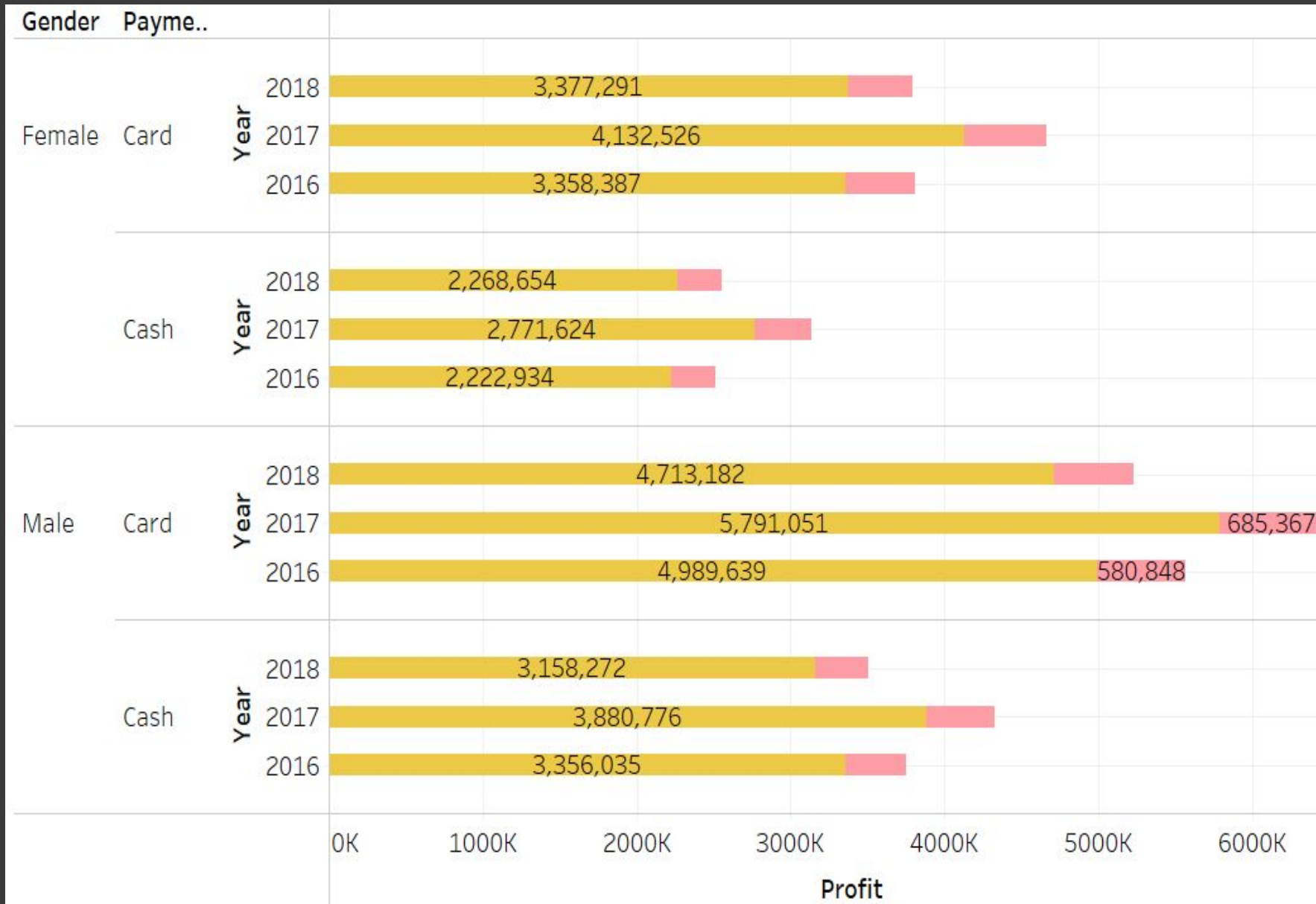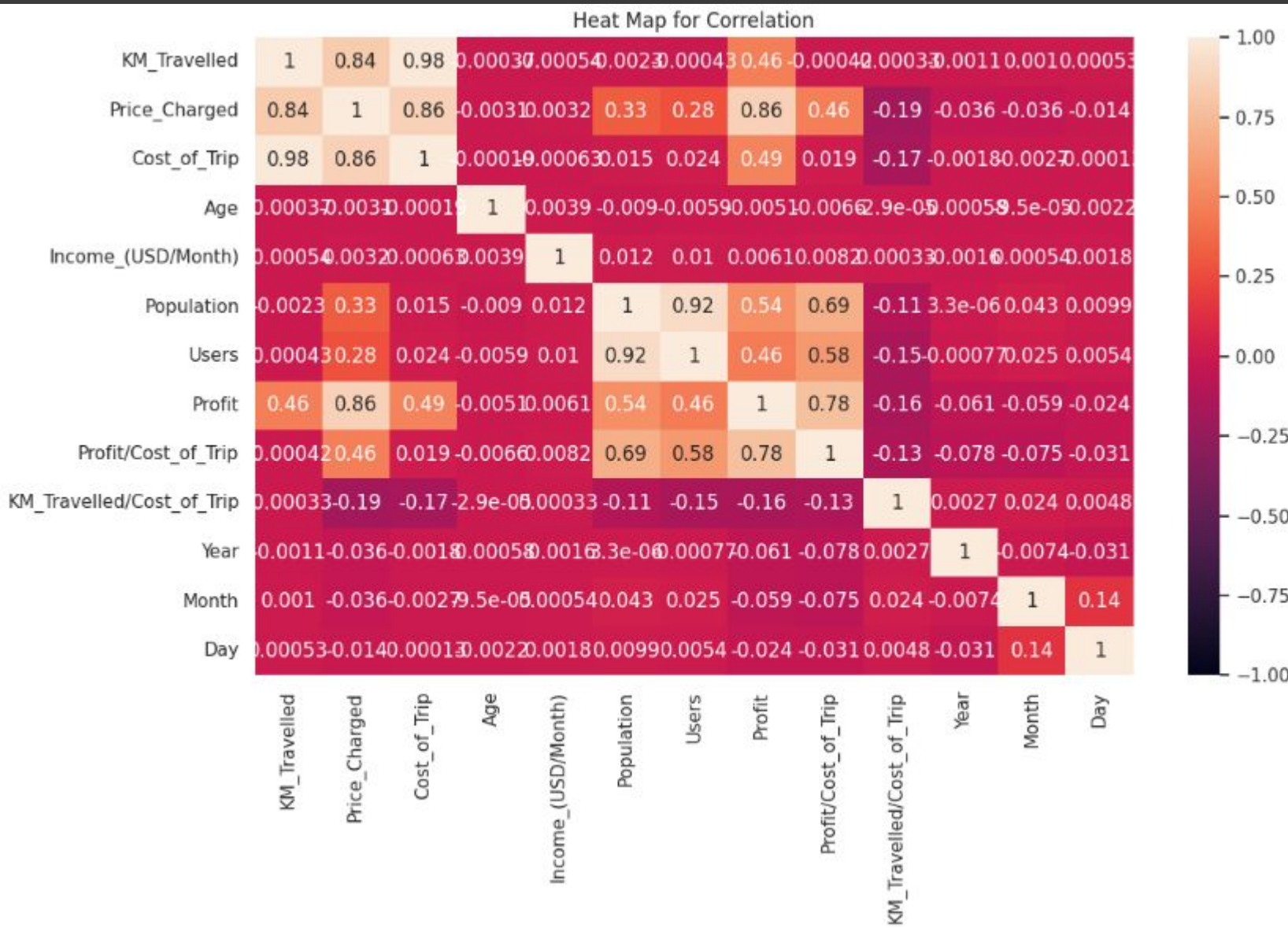
-San Diego CA

-Pittsburgh PA

-Sacramento CA

-Nashville TN

Payment Mode, Gender, Profit

-Most profitable group is male with using card who is using Yellow Cab.

# Correlation of Each Items

Heat Map for Correlation

-There is correlation between Profit and Price_Charged as 86% that is high. That make sense more charges is more profit in general.

-When KM_Travelled is getting long, then Price_Charged is also increasing. The correlation rate is 84%.

- When Users increase, Population is also increasing as 92%. It is very strong relationship.

Correlation with numeric items
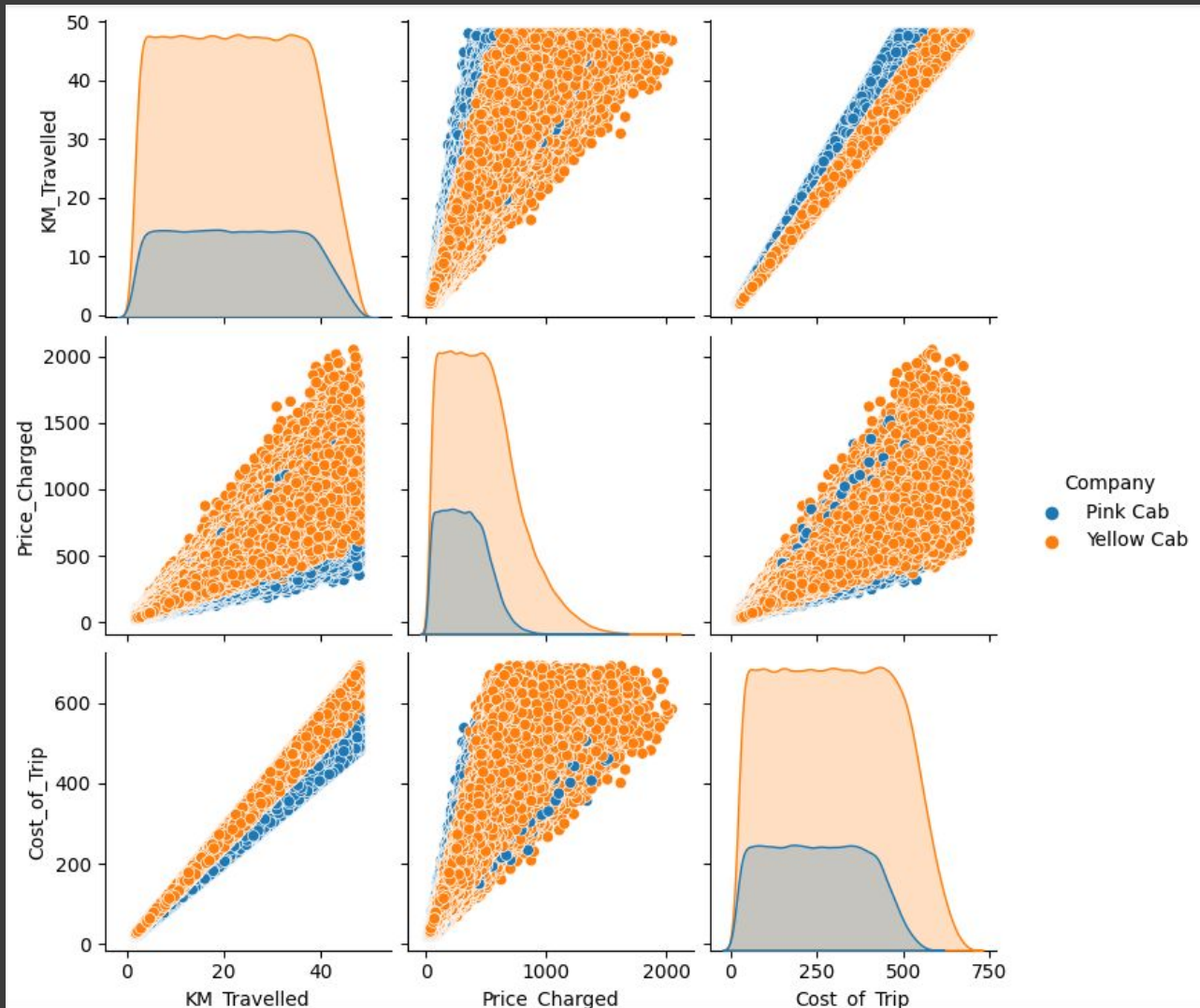
-Comparing of Yellow Cab and Pink Cab, many of dots are Yellow Cab's data.

-As we can see, the scatter plots of Cost_of_Trip and KM_Travelled shows Yellow Cab is more costly than Pink Cab.

-We can see that trend in Cost_of_Trip and Price_Charged as well. Some of Pink Cab's data is higher than Yellow Cab's but in general Yellow Cab has more dominated and profitable.

Profit/Cost_of_Trip
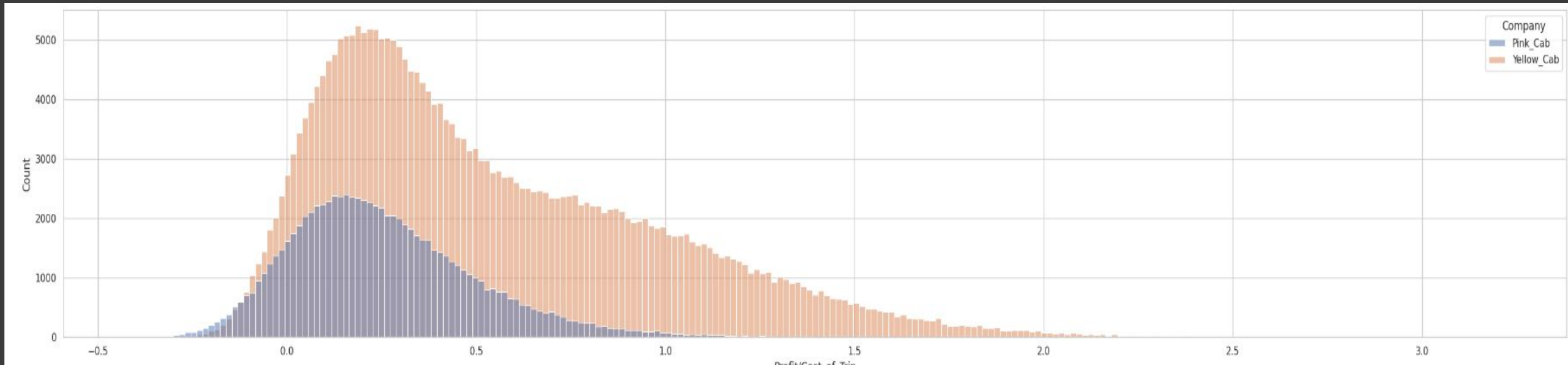
-In Profit/Cost_of_Trip, it shows how much they make Profit against Cost of Trip.

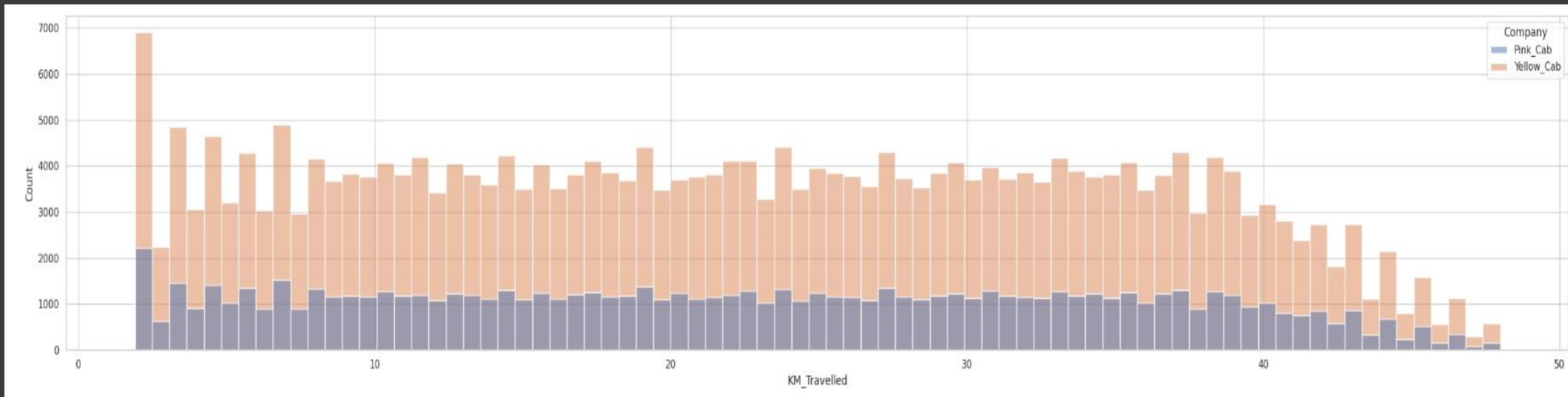Yellow Cab make max 2.5 times profit of Pink Cab.

Distance per $1
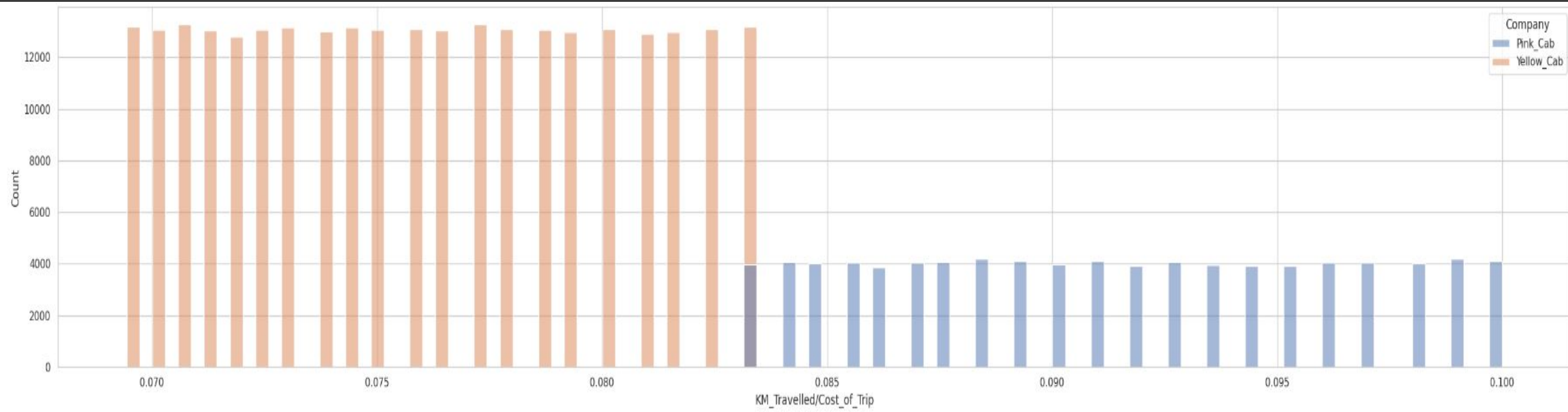
This shows how many KM Travelled when each Cab service charges $1.

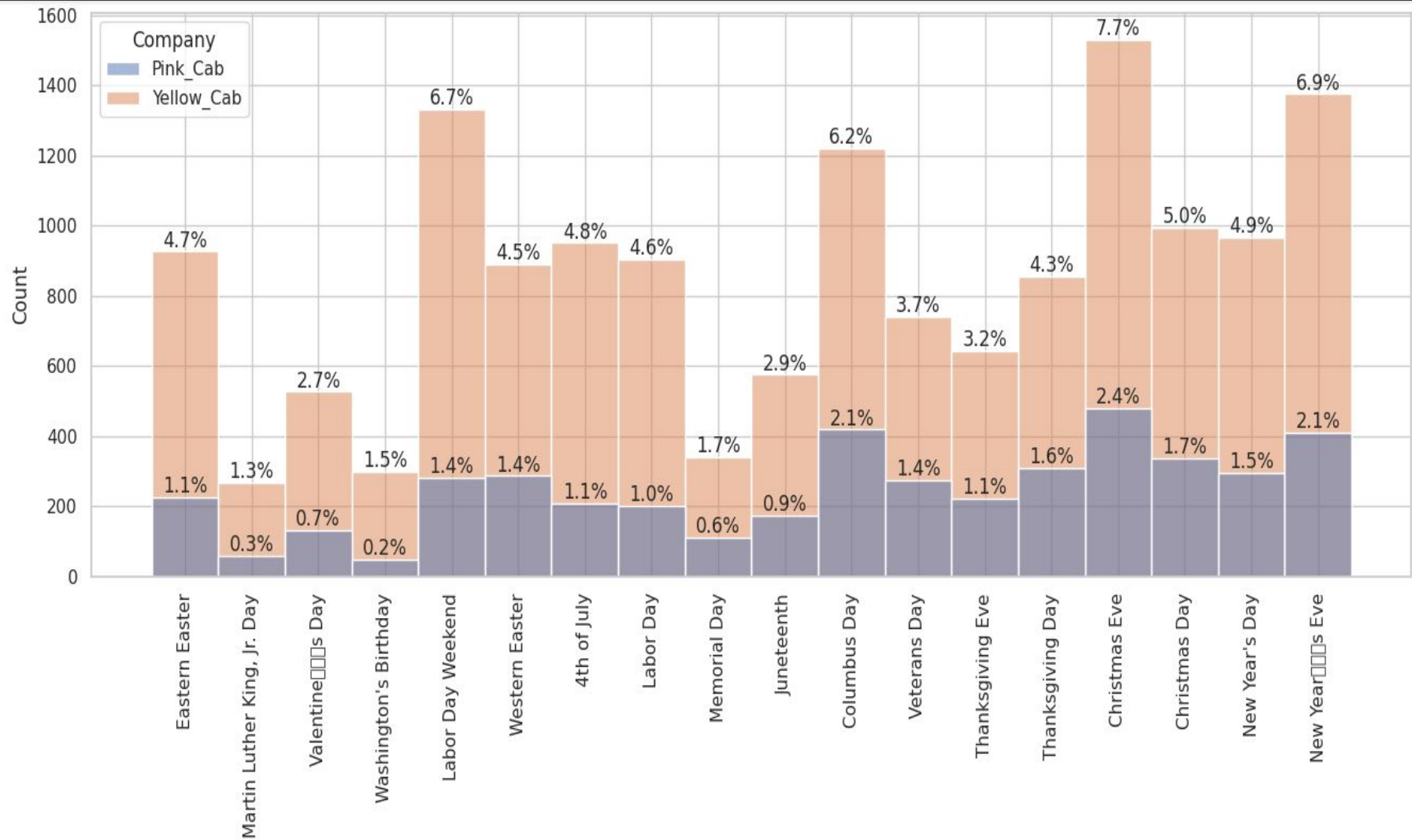This analysis absolutely shows Yellow Cab charges more than Pink Cab.

Yellow Cab never go over than 0.084km per $1. On the other hand, Pink Cab starts charging from 0.085km per $1.

Distance comparison

Both cab companies covers all of ranges of distance. But Yellow Cab's transactions are approximately 2.5 to 3 times of Pink Cab's.
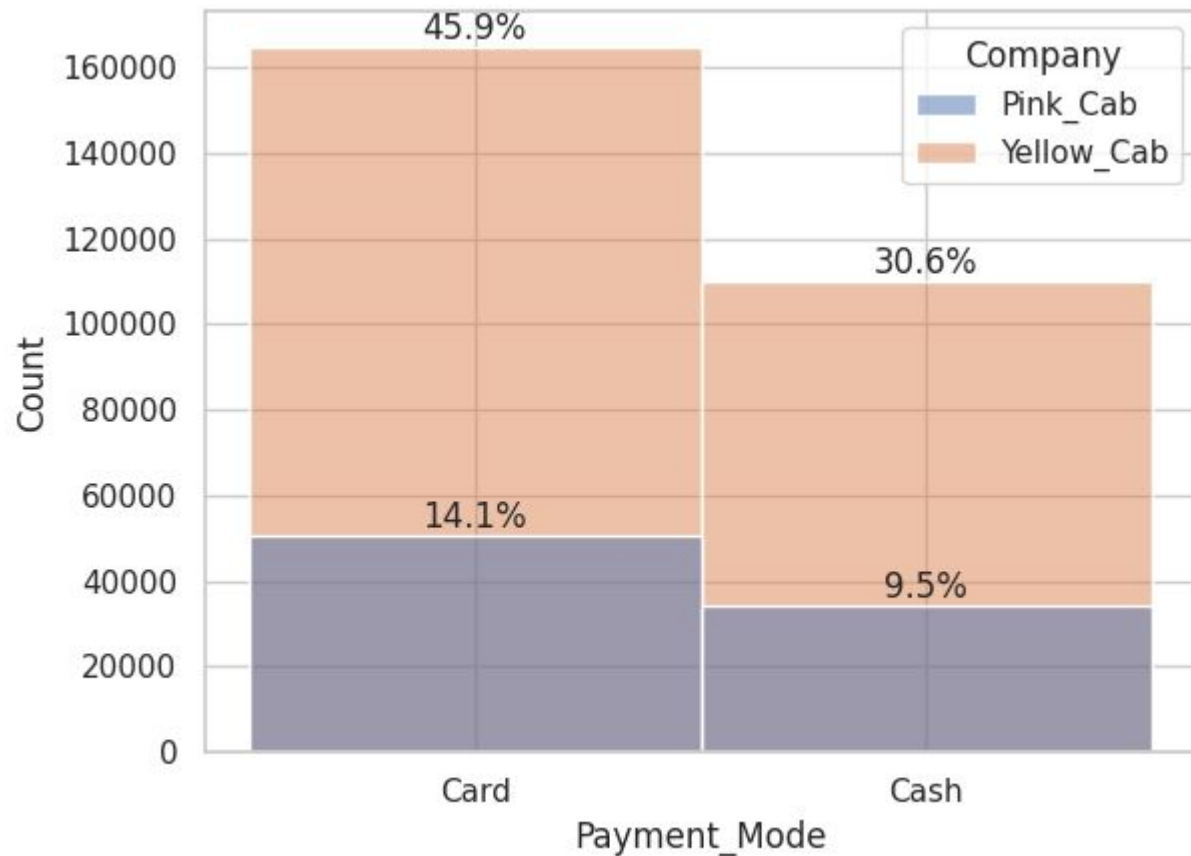
Usage in Holidays

-In any holidays, Yellow Cab are used more compared to Pink Cab.

-Christmas Eve is the holiday that most customers use the service.

-New Year's Eve is the 2nd largest usage holiday.

-Labor Day Weekend is the 3rd largest usage holiday.
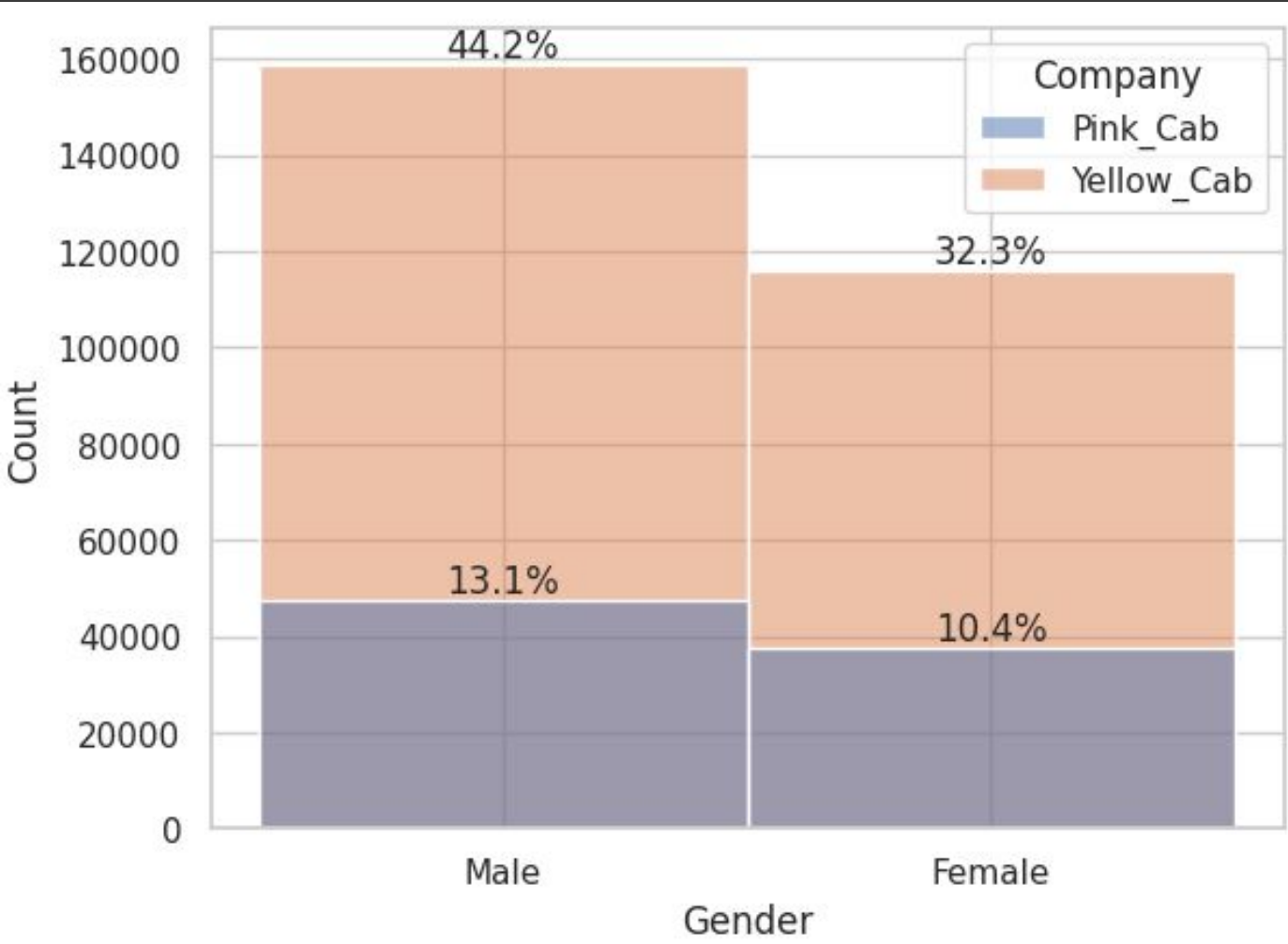
Payment Mode Comparison

-In both company, the customer with using Card is higher than using Cash.

-The percentage of Payment Mode is same in both companies as Card 60% and Cash 40%.

|	Yellow	Pink
Card	46%	14%
Cash	31%	10%
Total %	77%	24%
Card/Total	60%	60%
Cash/Total	40%	40%

Gender Comparison

-Both companies has more Male customers compared to Female.

-The percentage of Gender in both companies is about same as Male 57% and Female 43%.

	Yellow	Pink
Male	44%	13%
Female	32%	10%
Total	77%	24%
Male/Total	58%	56%
Female/Total	42%	44%

Transactions by Month

Transaction Count by Company and Month
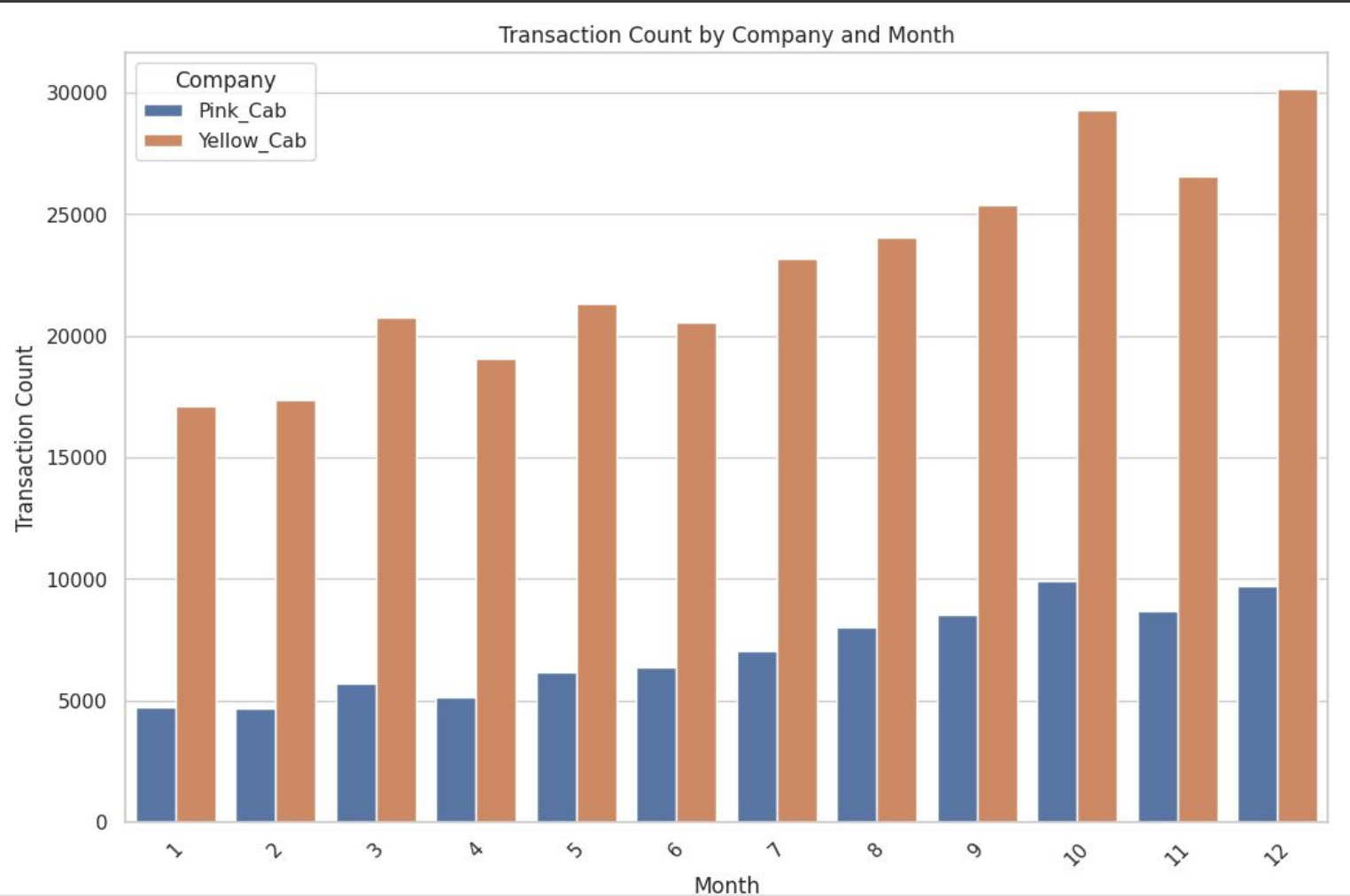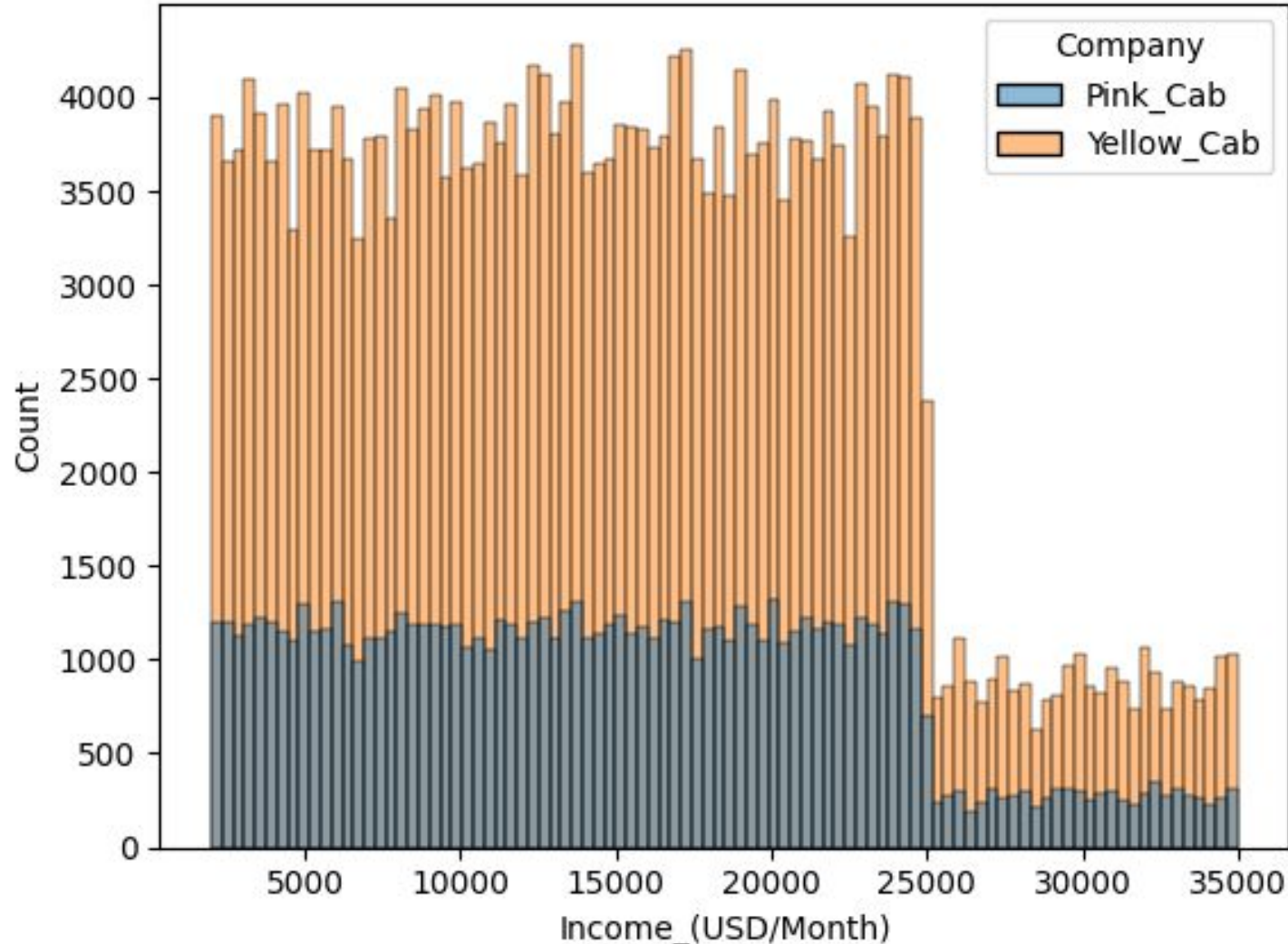
-Every month, Yellow Cab's transitions are approximately 3 - 4 times more than Pink Cab's.

-Both company's transactions increase toward the end of the year.

Income Comparison

Both companies have wide range of income. However, customers with Yellow Cab share larger amounts than customers of Pink Cab in any income ranges.

The amount of over 25000 customers significantly decrease both Pink Cab and Yellow Cab.

However, Yellow Cab is popular among the over 25k income groups compared to Pink Cab.

It seems the difference over 25k and under 25k are some rule or restriction making this gap because of the gap is significantly large.

Age Comparison
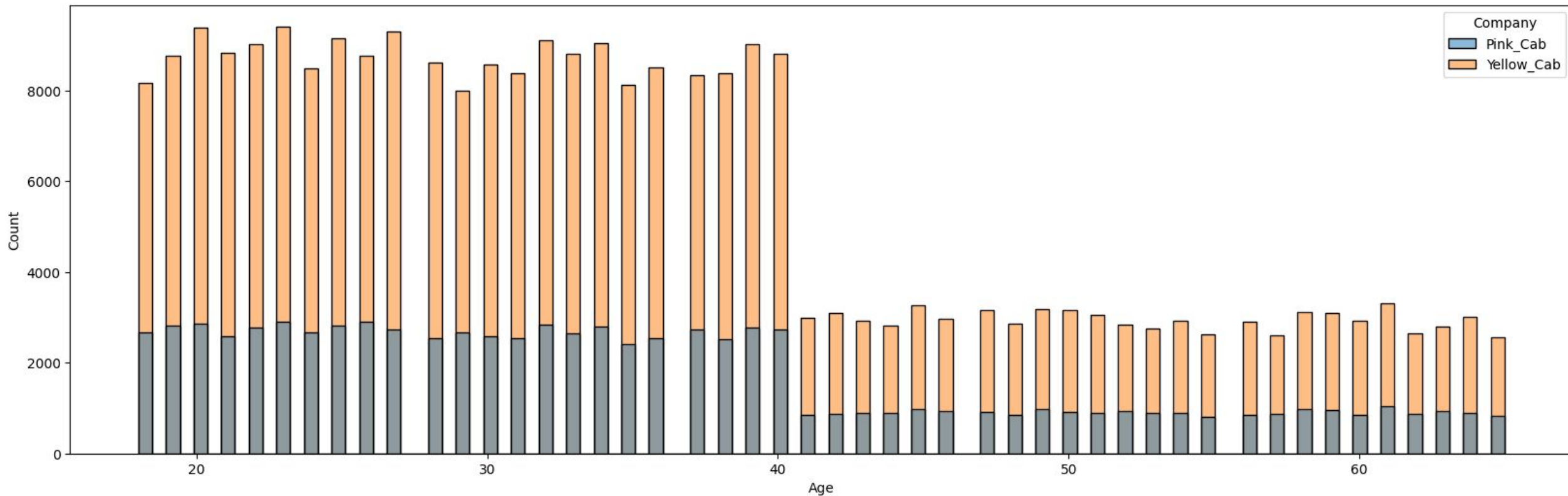
Generally, customers use more Yellow Cab compared to Pink Cab.

Under 40 years old customers actively use Cab service compared to over 40 years old customers.

Over 40 years old customers especially more use Yellow Can than Pink Cab.

It seems the difference over 40 and under 40 are some rule or restriction making this gap because of the gap is significantly large.

Repetitive Customers
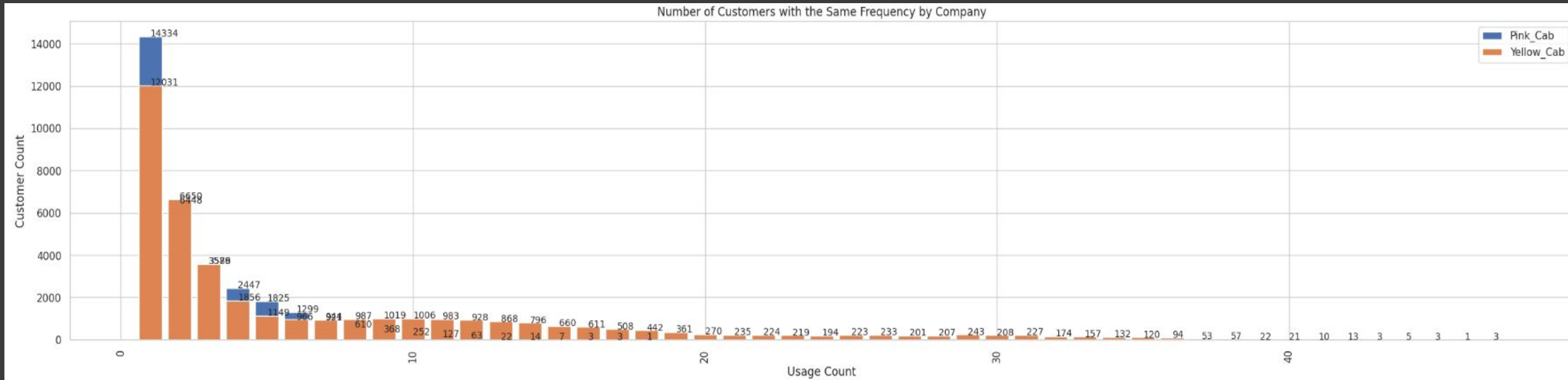
Using the service just 1 time is the highest number of counts. Both Yellow Cab and Pink Cab have the largest amount of customers in this. 84% of 1 time usage is shared by Yellow Cab. 6% of 1 time usage is shared by Pink Cab.

The count of using 2nd time customers is approximately half of the amount of using 1 time customers.
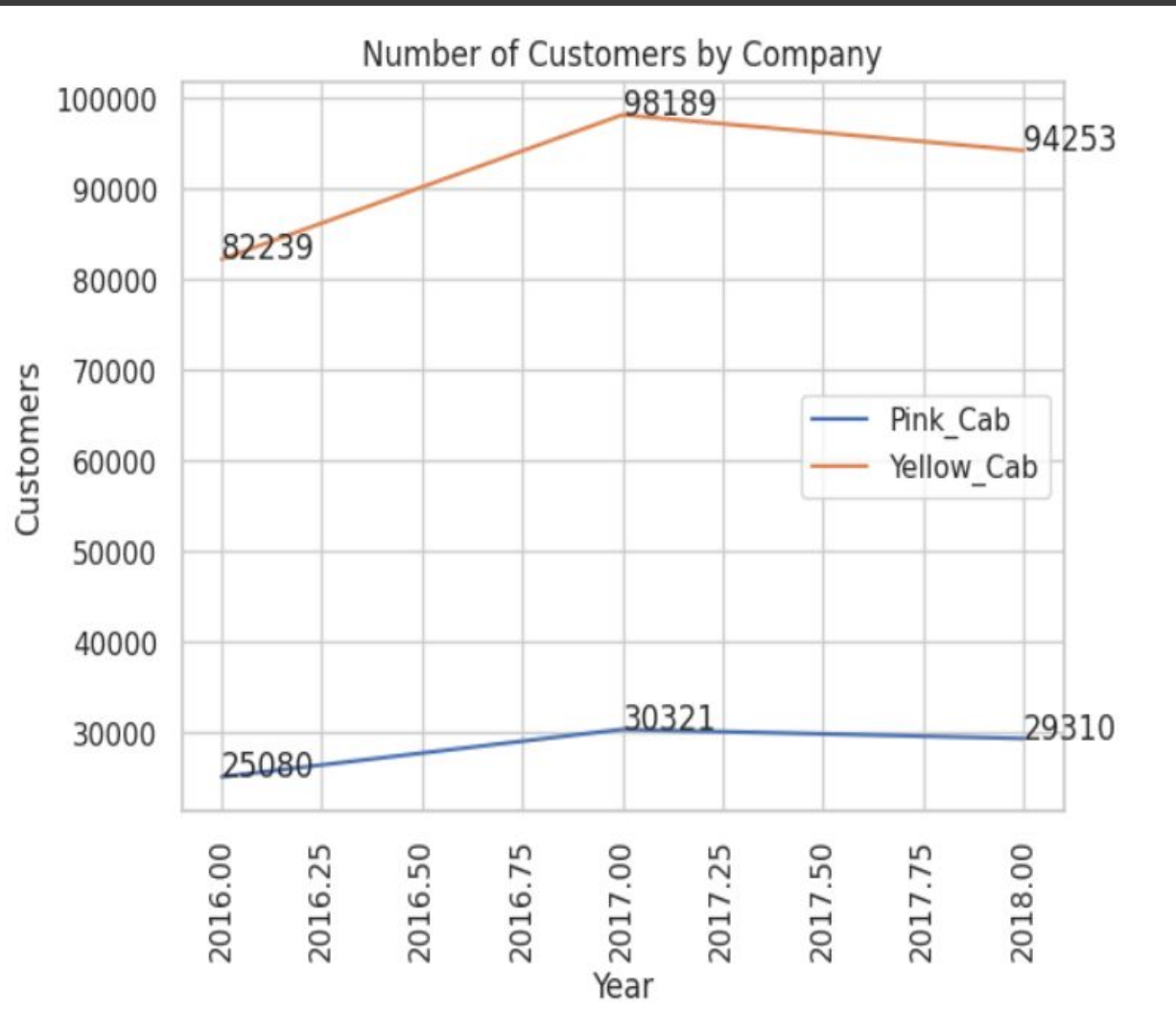
After using 3rd time and so on is decreasing the amount. But the most of the customers are belonging with Yellow Cab.

4th times usage: Yellow: 76%   Pink: 24%

5th times usage: Yellow: 63%   Pink: 37%

6th times usage: Yellow: 74%   Pink: 26%

Number of Customers

Yellow Cab has 3-4 times numbers of customer of Pink Cab every year.

Their holding of the largest numbers of customer is in 2017. However, in 2018, the number of customers of both of companies is decreasing.

In 2017, new customers are increasing from both companies.
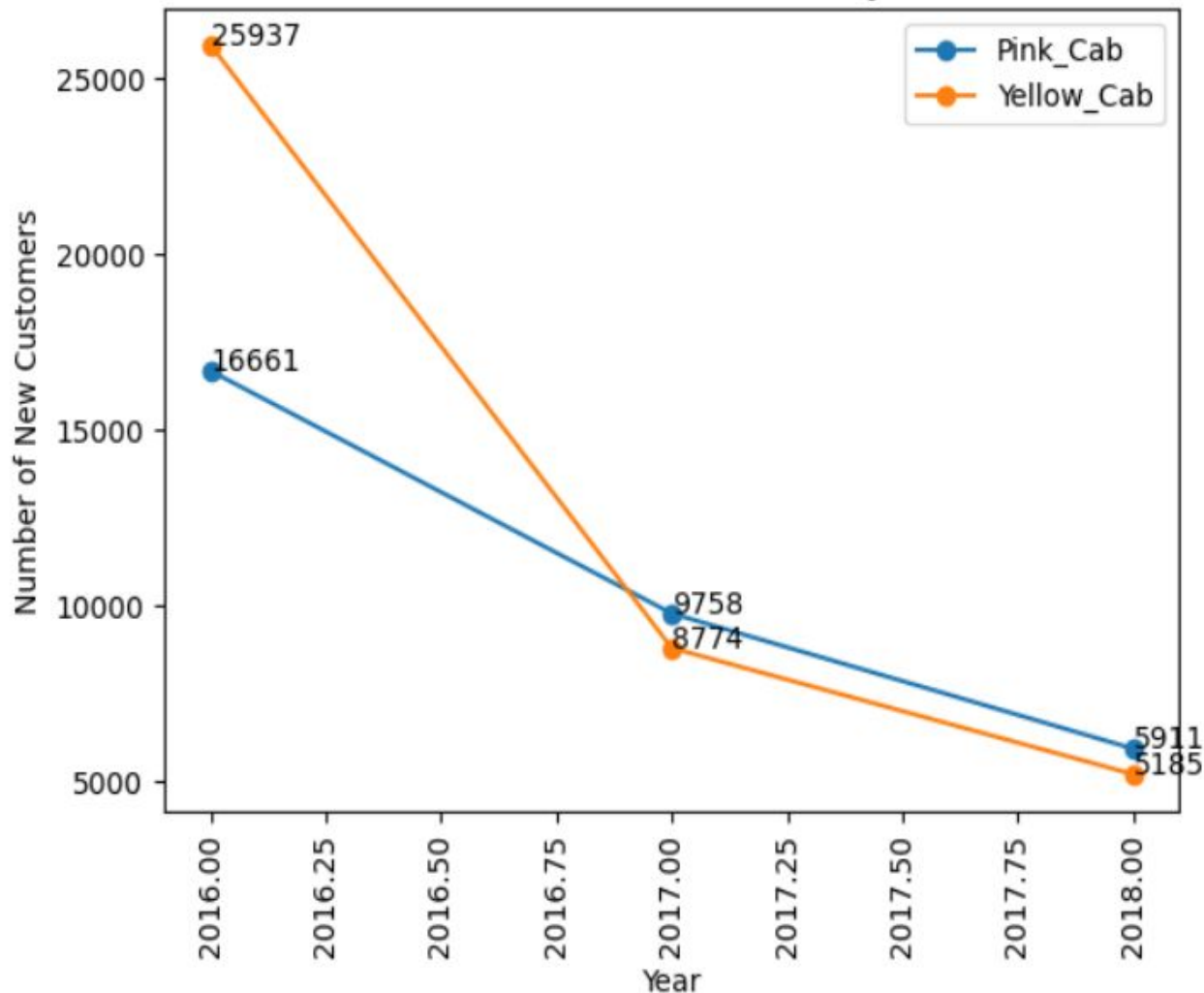
Yellow Cab: 119% (YOY)

Pink Cab: 120% (YOY)

In 2018, new customers are decreasing from both companies.

Yellow Cab: 95% (YOY)

Pink Cab: 97% (YOY)

New Customers

Number of New Customers by Year

In 2016, we don't have a data with 2015 so we can't compare if there are new customers in 2016 from 2015.

In 2017, new customers are decreasing from both companies.
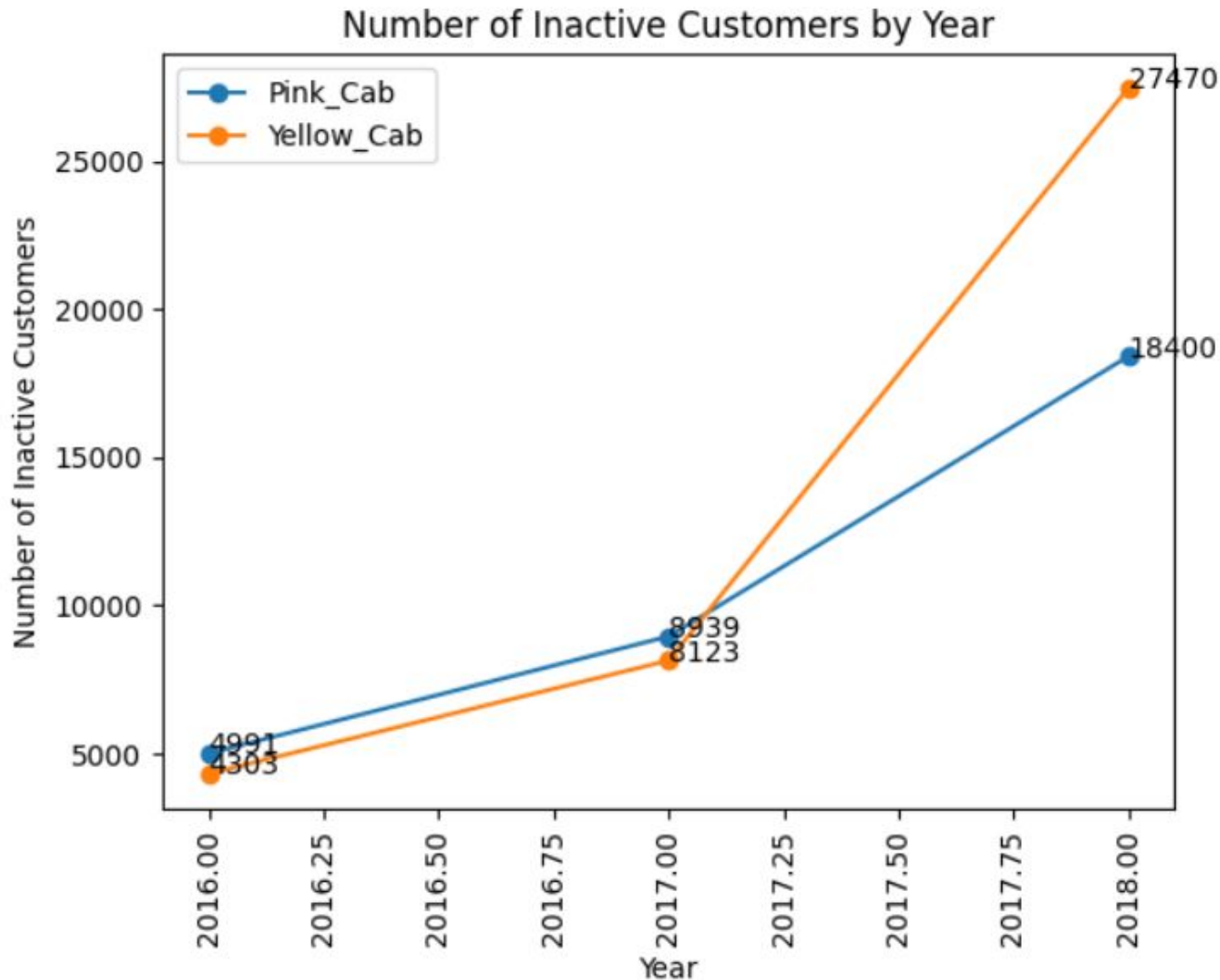
Yellow Cab: 34% (YOY)

Pink Cab: 59% (YOY)

In 2018, new customers are decreasing from both companies.

Yellow Cab: 60% (YOY)

Pink Cab: 60% (YOY)

Inactive Customer Comparison

Number of Inactive Customers by Year

In 2016, we don't have a data with 2015 so we can't compare if there are inactive customers in 2016 from 2015.

In 2017, inactive customers are increasing from both companies.

Yellow Cab: 188% (YOY)

Pink Cab: 180% (YOY)

In 2018, inactive customers are increasing from both companies.

Yellow Cab: 338% (YOY)

Pink Cab: 205% (YOY)

Recommendation: Yellow Cab is good for investments

-**Customer Reach:** Yellow Cab has 15 cities that the larger amount of usage compared to Pink Cab. Pink Cab has 4 cities that the larger amount of usage compared to Yellow Cab.

-**Repeat Customer**: Yellow Cab has large amount of repetitive customers while Pink Cab does not have them a lot. In 4-6 times usage customers, 60-70% of them use Yellow Cab while 40-30% of them use Pink Cab. In other usage times, most of them are dominated by Yellow Cab.

-**Profit from short distance**: Yellow Cab provides the service to customer with less 0.084 KM per $1 while Pink Cab provides the service to customer more than 0.085 KM per $1.

-**Profit from Holidays:** Yellow Cab is used by many customers any holidays not only Christmas Eve to New Year's day. Those usages provide the company more profitable compared to Pink Cab. Pink Cab's most profitable holidays are Christmas Eve to New Year's day.

-**Age Distribution**: Over 40 years old customers prefer Yellow Cab to Pink Cab. Even though the number of customers of over 40 years old is less than under 40 years old, that is noticeable thing.

-**Income wise Reach**: Yellow Cab is popular at over 25k income range customers compared to Pink Cab. This is great opportunity to increase profit.

-**Gender wise Reach**: Both of cab companies has more Male customers compared to Female. Male who use card will provide more profit with using the cab service.

Thank You