



Data Glacier

Your Deep Learning Partner

G2M Case Study

Virtual Internship

09/20/2022

Agenda

- Executive Summary
- Problem Statement
- Datasets
- EDA
- Recommendations



Executive Summary

Problem statement : Understanding the market in order to make decision regarding investment in Cab industries, Yellow Cab versus Pink Cab

Approach :

- Cleaning and merge the data to create the Master data
- Conduct EDA and understand the features.
- Test multiple hypotheses

Recommendation : I recommend to invest in Yellow Cab company given the data and EDA result.

Problem Statement

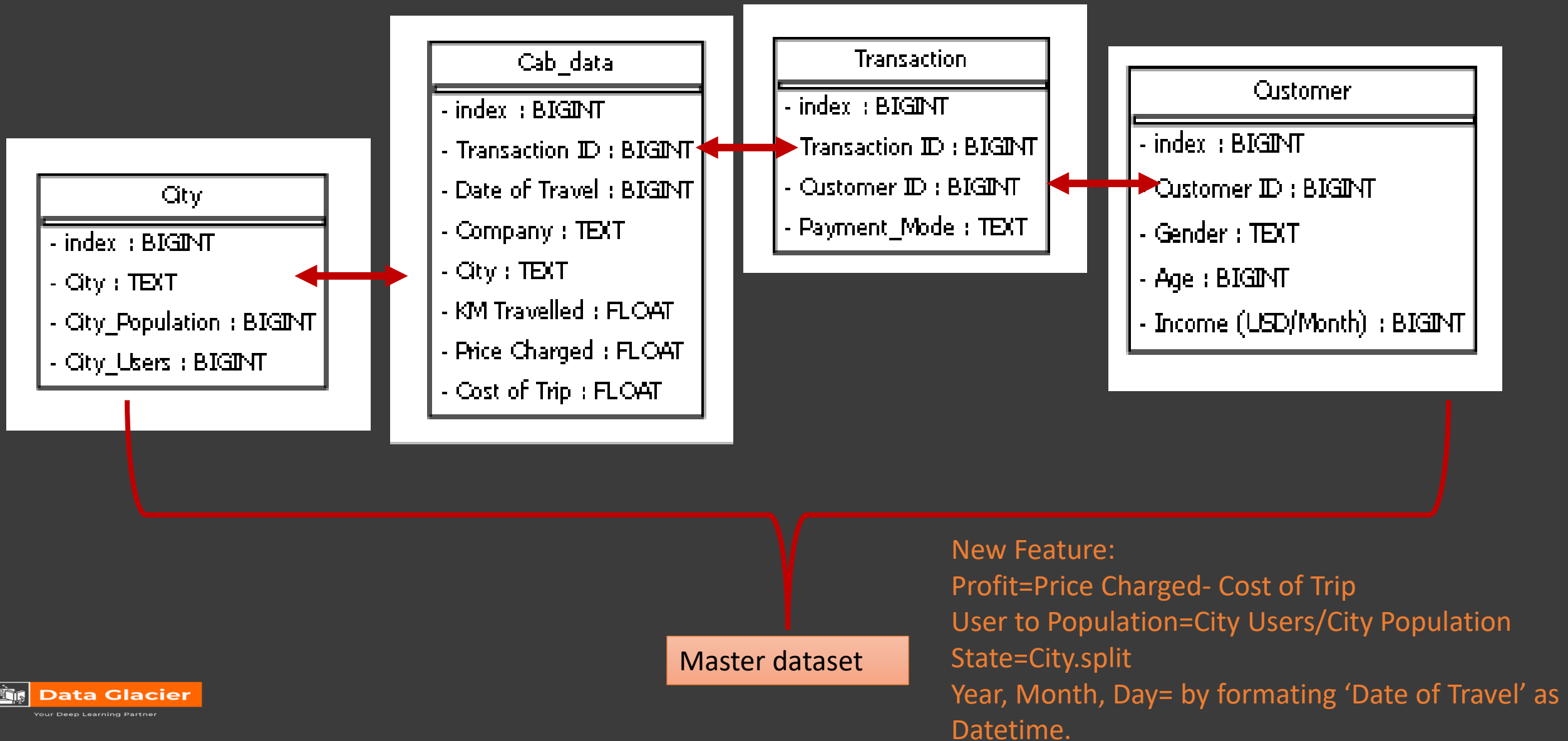
The remarkable growth in the Cab Industry in last few years make companies interested in investing in Cab industry. XYZ company is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Objective

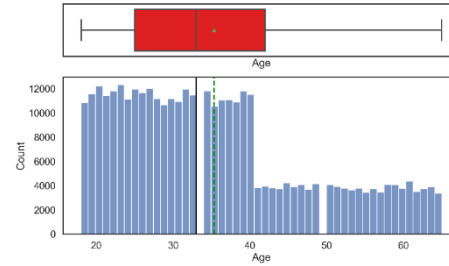
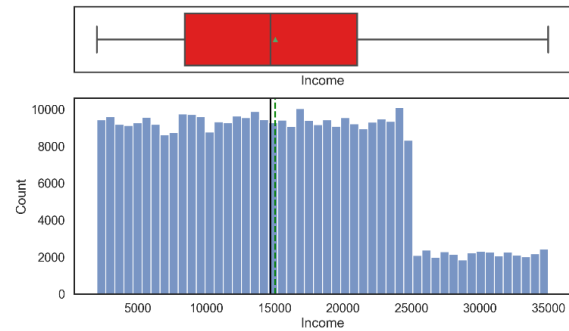
Understanding the market in order to make decision regarding investment in Cab industries.

Which company should XYZ invest in ?

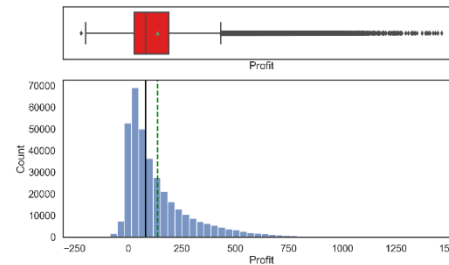
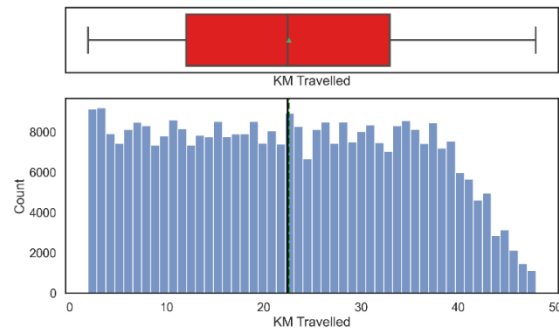
Data Sets



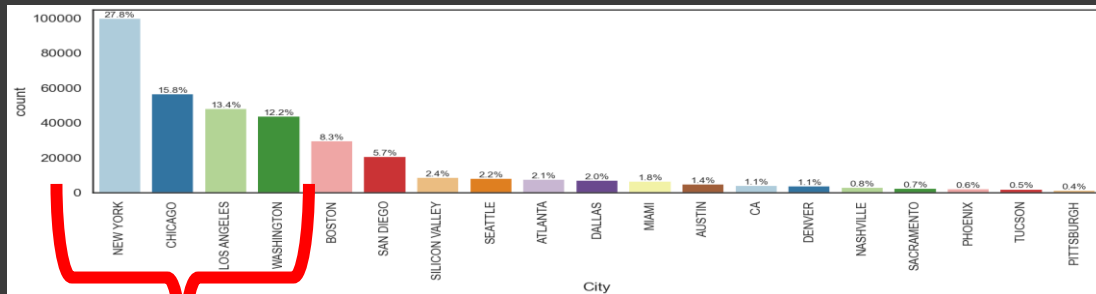
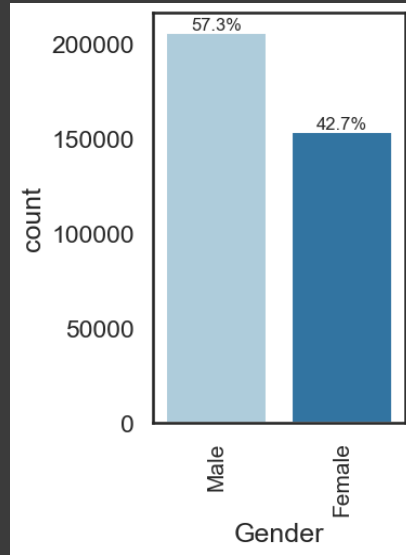
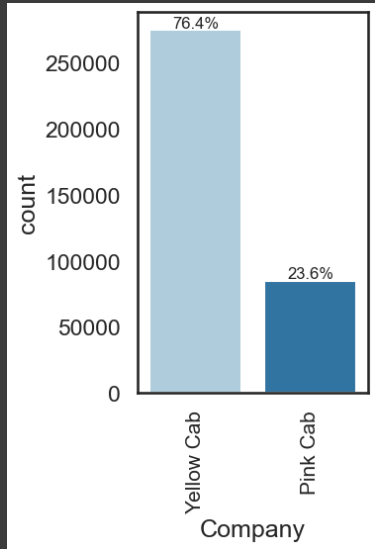
Univariate Analysis



- Customer with Income less 25k more frequently to use cab.
- Customer with Age less 40 more frequently to use cab.
- Average Travelled distance is 22km.
- Profit skewed toward the right with mean 125 and median 100.



Univariate Analysis

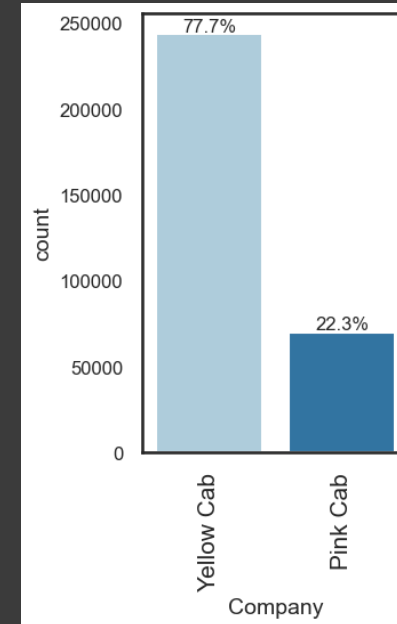


69%

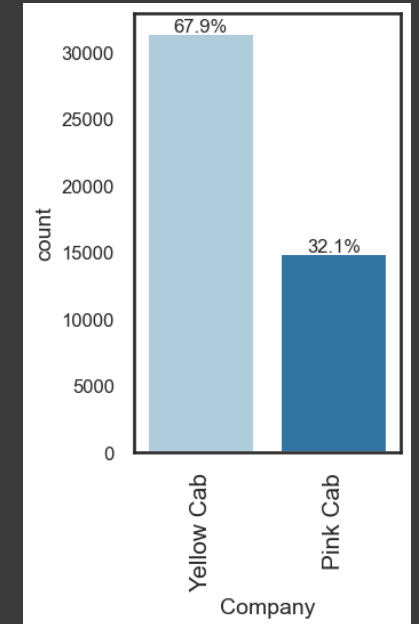
- 57% of the users are male while 42% are female.
- The largest three cities contain 55% of all the orders.

Univariate Analysis

Repeated customers

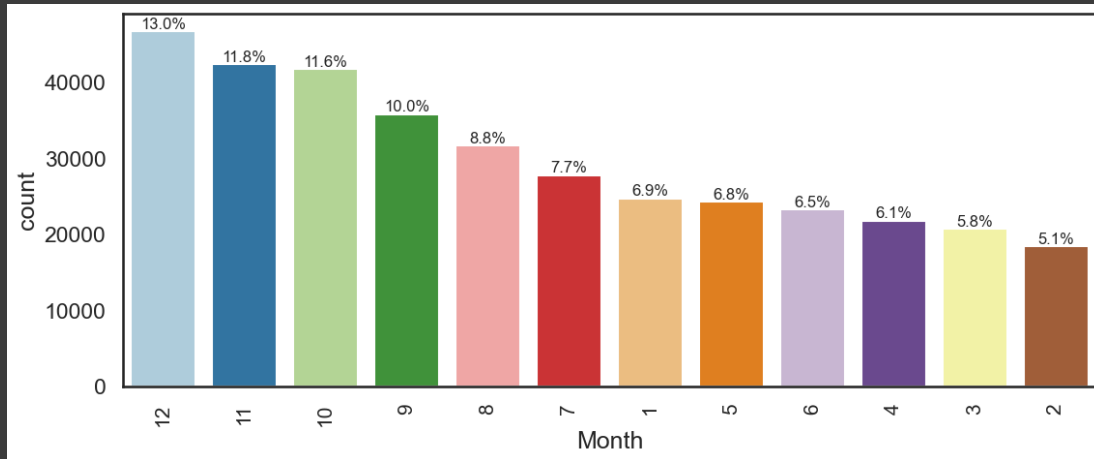
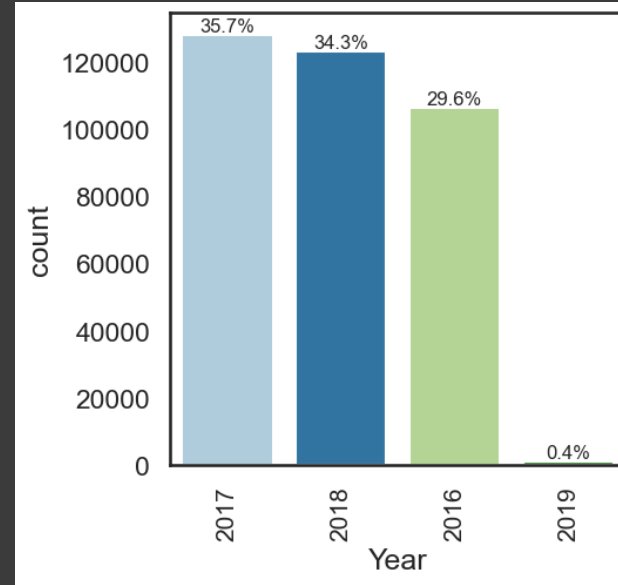
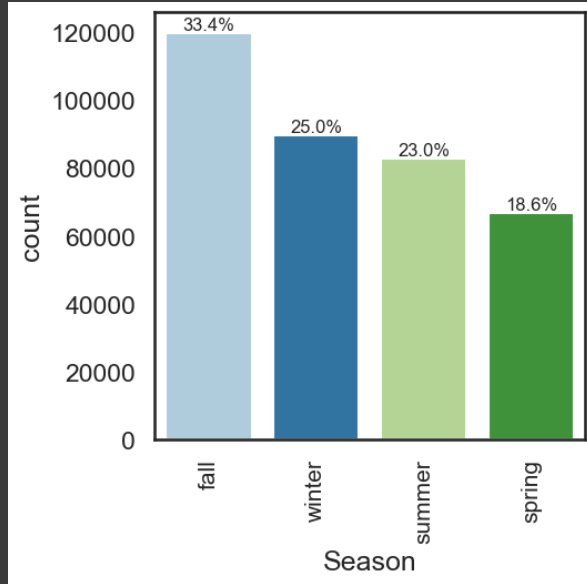


Unique customers



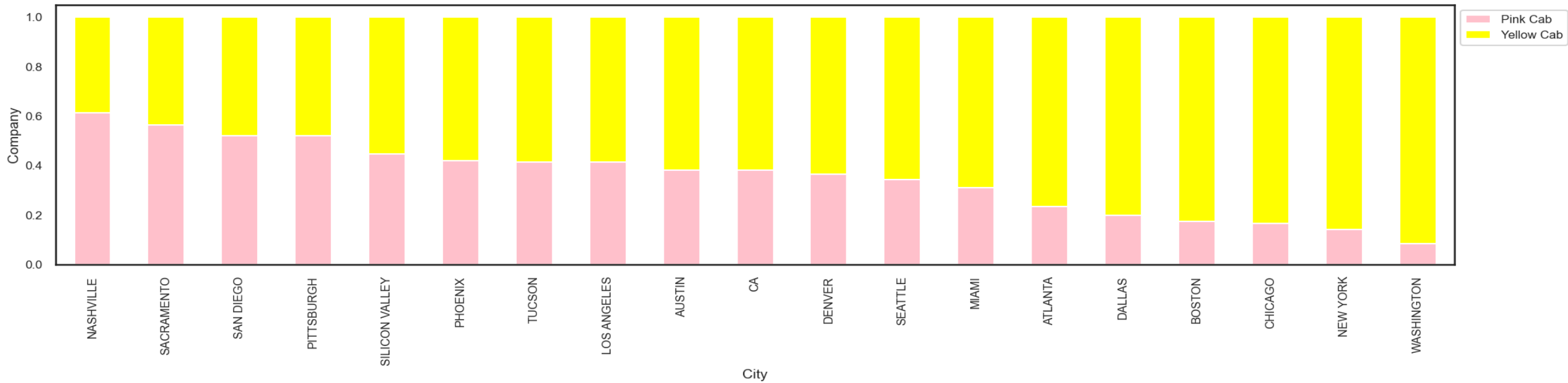
- 77% of the users are returned to the Yellow company while 67% did not return.
- 22% of the users are returned to the pink company while 32% did not return.

Univariate Analysis



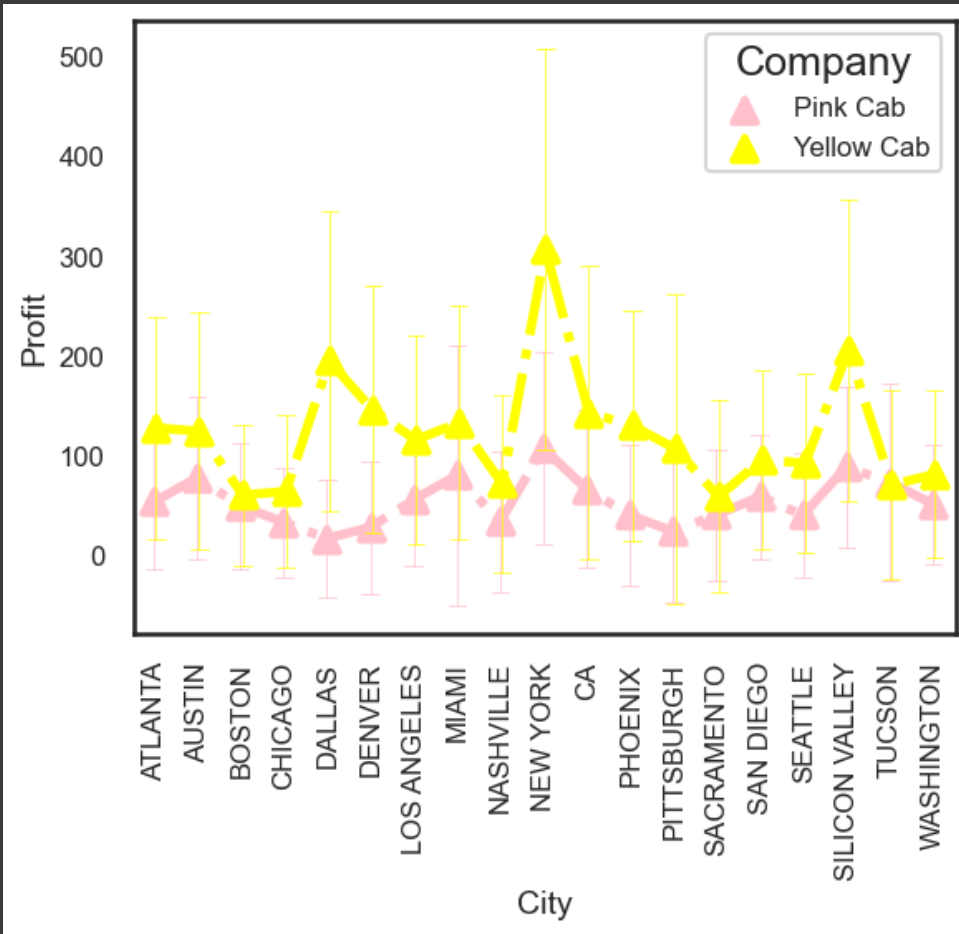
- **35% of order happed in 2017.**
- **33 % of order took place in Fall season.**
- **December orders 13% of all the orders.**

Company users per City



The number of order of yellow company is higher than 80% of the total user in Washington and New York while Nashville, Sacramento it is 40%, San Diego and Pittsburgh %45.

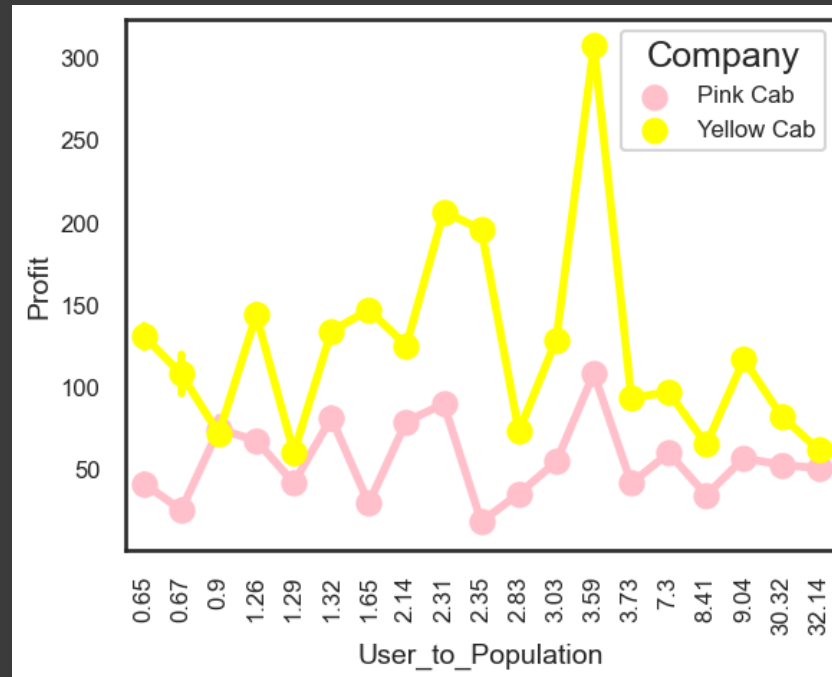
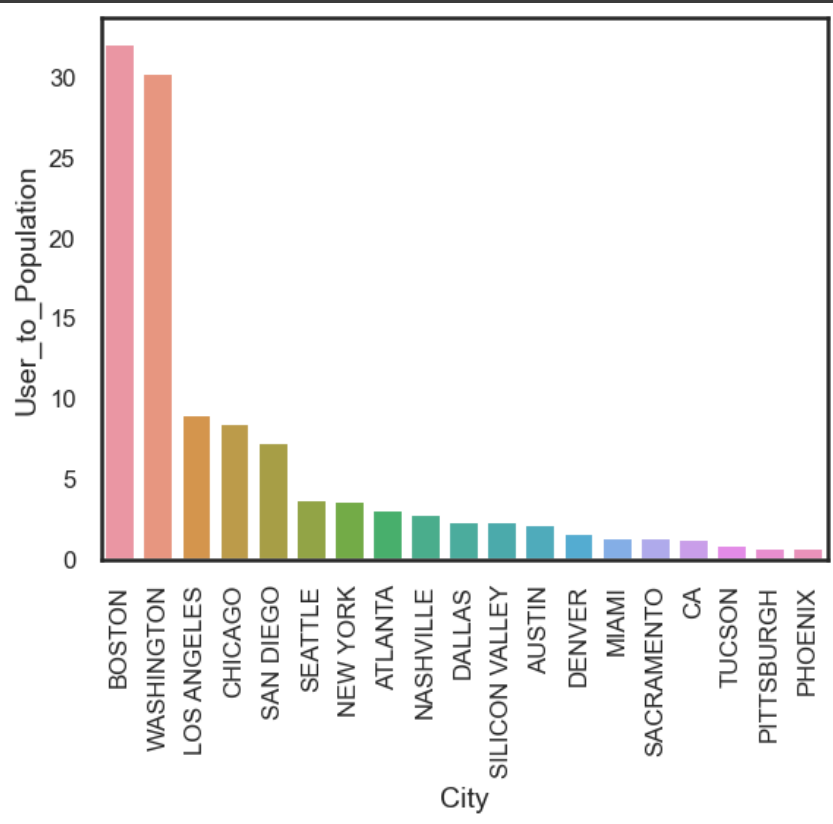
Profit with Cities



- New York city has the highest yellow cab profit ~1500 and Miami has the highest pink Cab profit~ 1000.
- Although the Yellow Cab has 80% of the user in Washington, It doesn't make a lot of profit.

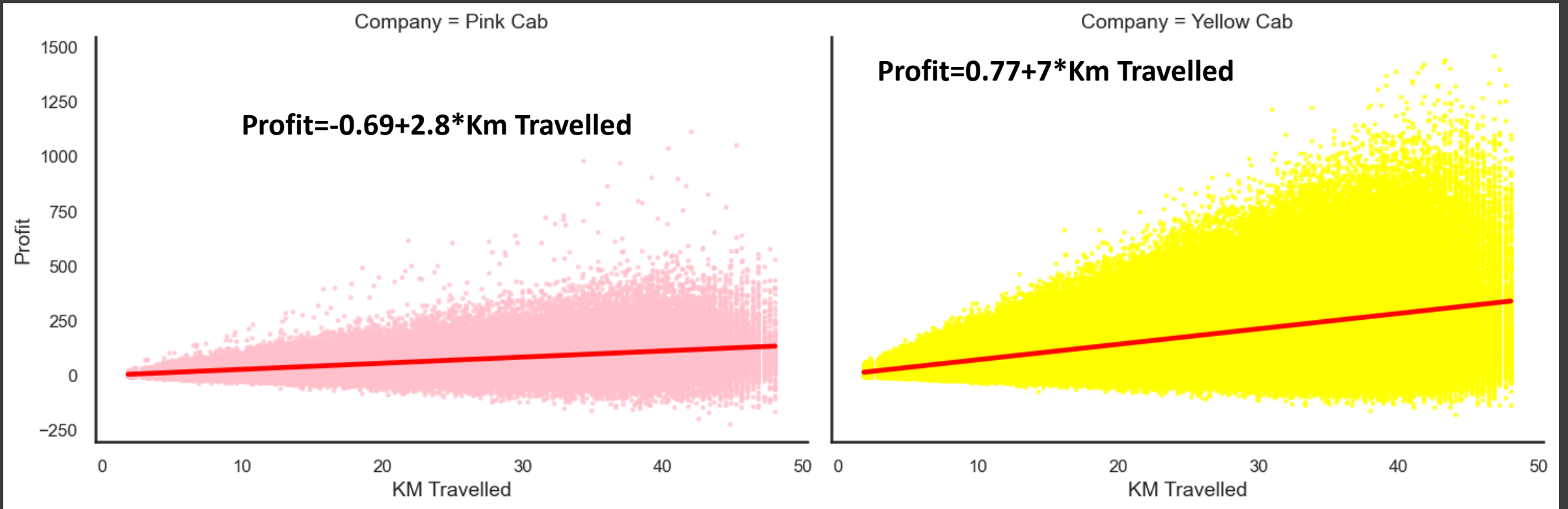
Multivariate Analysis

Users to Population ratio with City and Profit

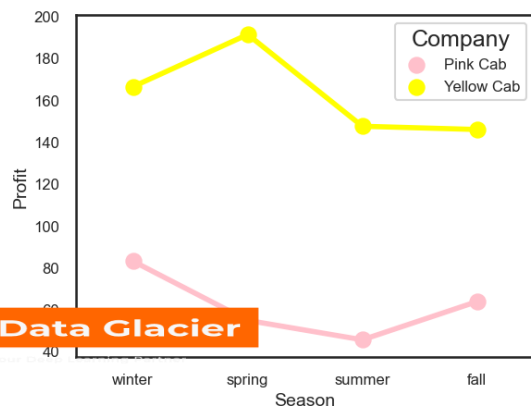
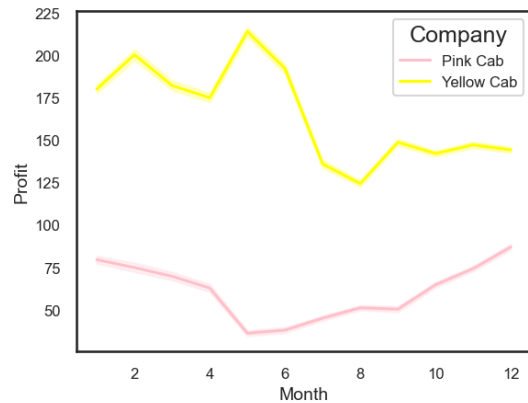
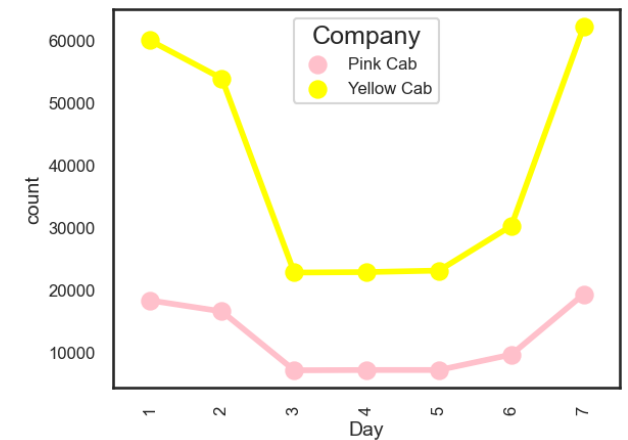
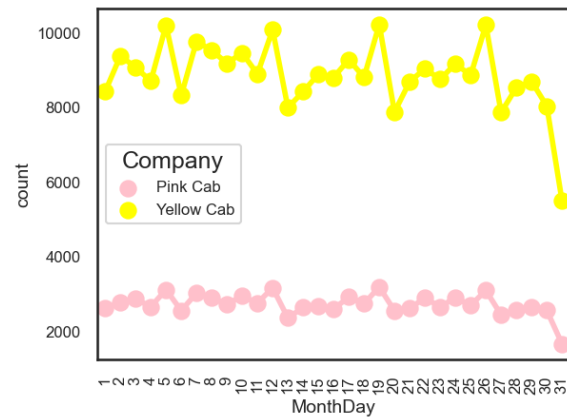
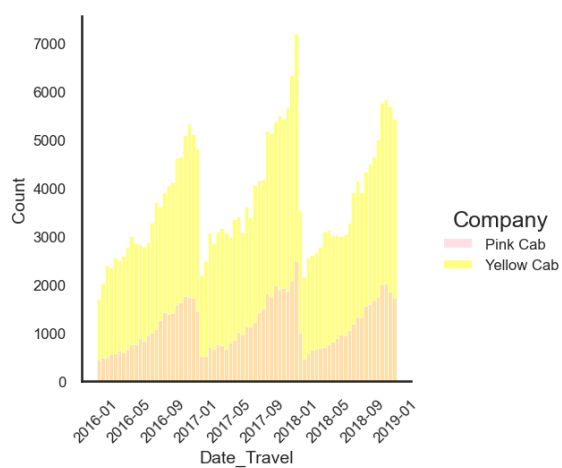


- Yellow Cab is existed in cities where the ratio of Cab users to city population is higher.
- The user to population ratio doesn't have affect on the profit.
- E.g. Boston high ratio of user to population but less profit for both cab company.

Profit with Distance



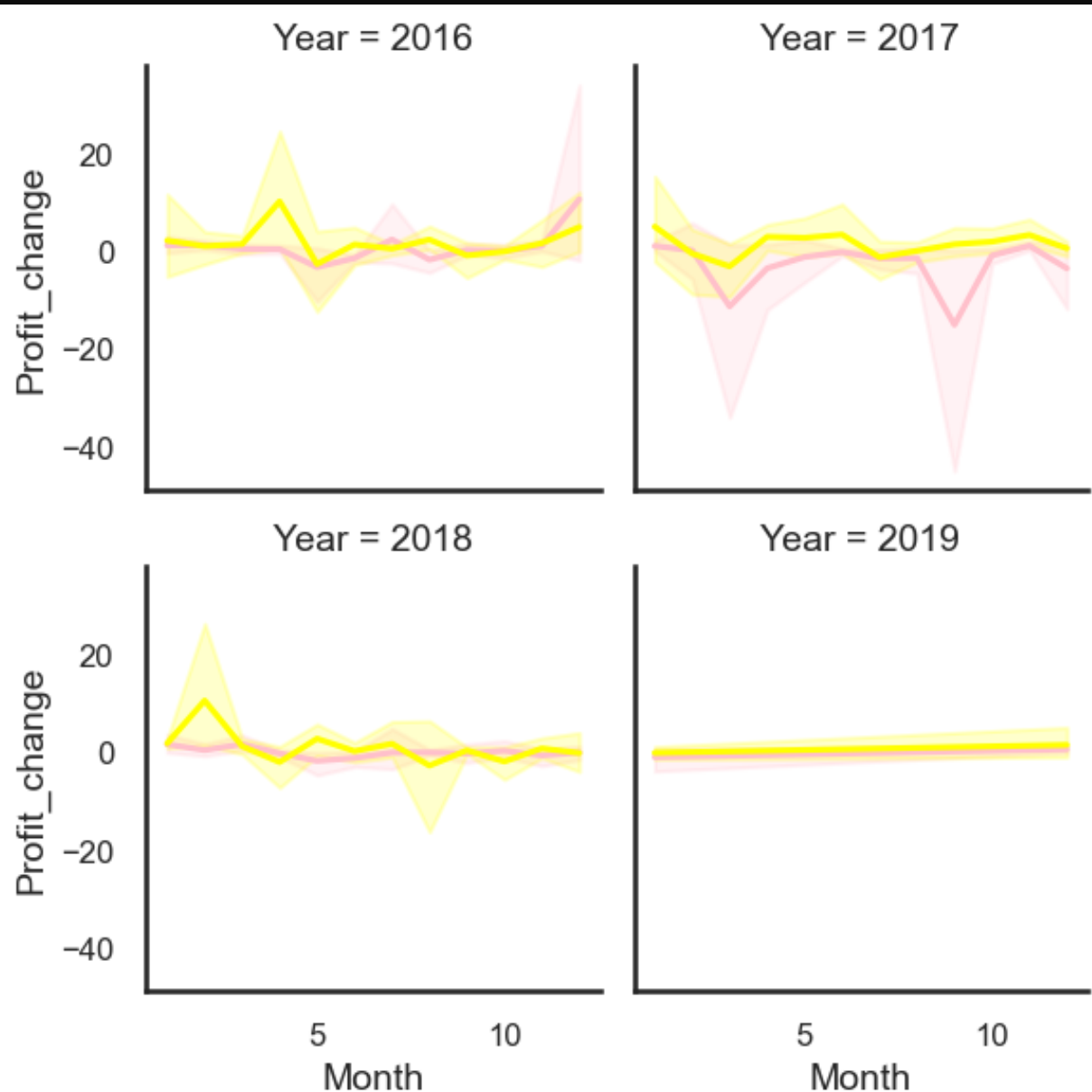
Yellow Cab makes more profit with longer distances than Pink company.



Periodicity

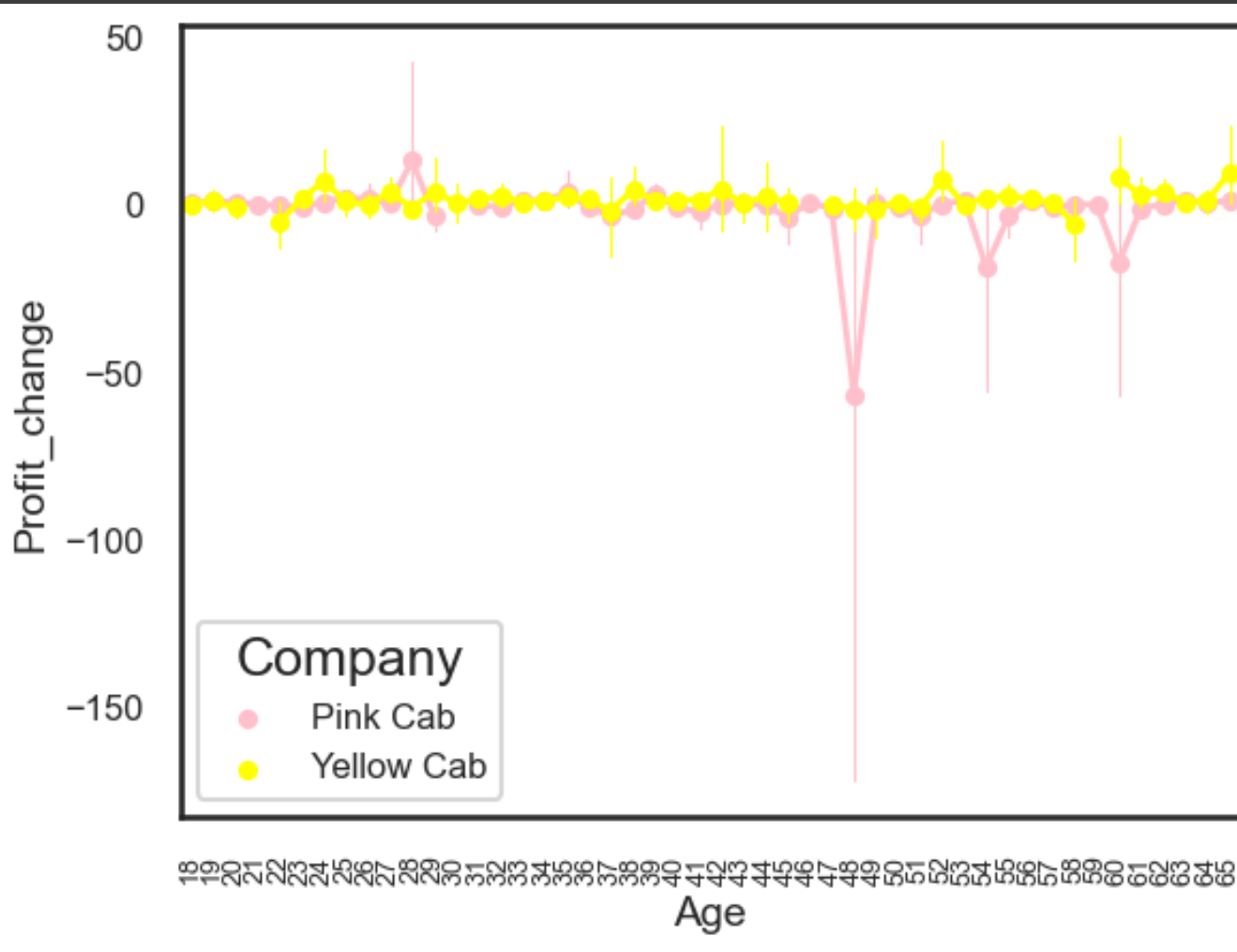
- The periodicity of the orders over time has the same pattern but it is much higher for Yellow Cab.
- Last months of 2017 has the higher demand.
- During the month the yellow cab shows more obvious pattern than pink cab and periodicity with increasing in the demand for the cab during the weekeands and Mondy.
- The profit that two companies make is vary differently around the year seasons. Yellow make the most on spring while pink make the most profit in winter.
- The two companies profit differently over year months and specifically around May

Changing the two Cabs profit over years



- The profit pct_change over time shows that the two Cab companies were doing good at 2016.
- In 2017 the pink Cab had much more decline in the profit than yellow Cab.
- Beginning 2018 the yellow cab recover and shows positive change in profit but did not last while the pink cab did not show positive change in profit.

Changing the two Cabs profit over user age



- The profit change (pct_change) over time shows that the two Cab companies were doing similar at user between 18-45.
- Users between 45-65, the yellow cab shows slightly positive changes in profit. while the pink cab show multiple negative change in profit.

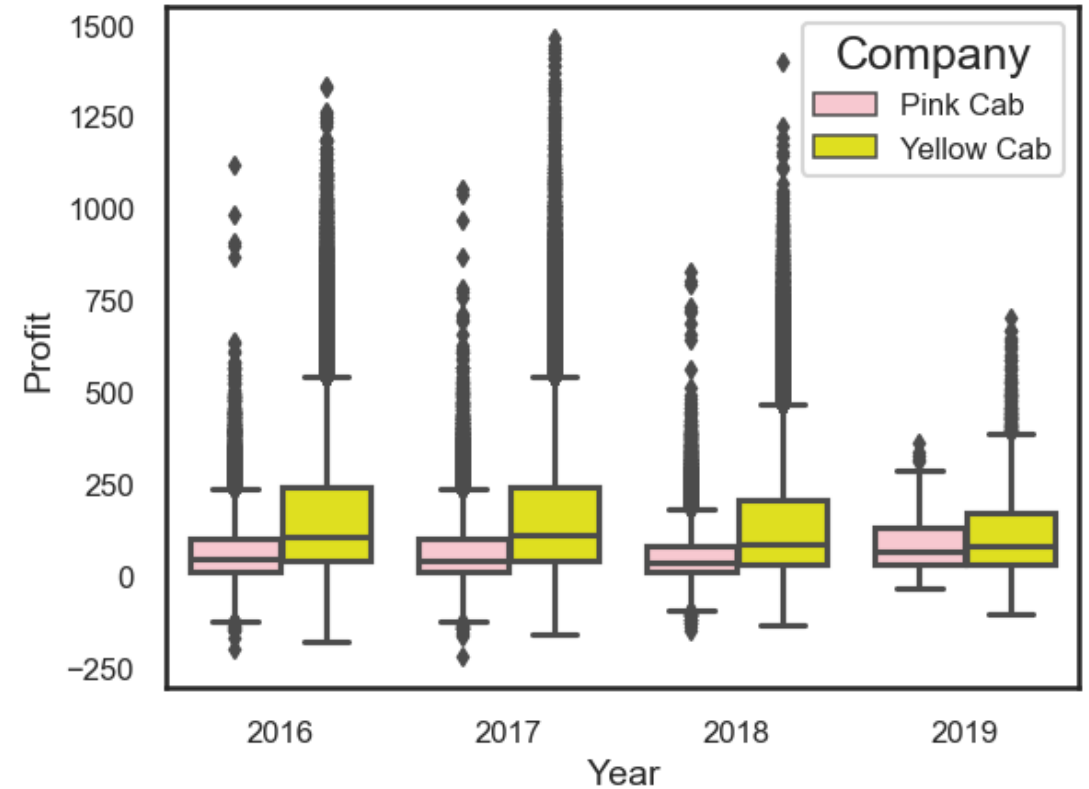
Hypothesis testing

Is there any difference in Profit regarding the year

- H_0 : There is no difference regarding the years in both cab companies
- * H_1 : There is difference regarding the years in both cab companies.

the p-value of Pink Cab $4.686e-140$ is less than the level of significance, we reject the null hypothesis

the p-value of Yellow cab $5.46e-295$ is less than the level of significance, we reject the null hypothesis.



The profit of the two companies varies per year

Hypothesis testing

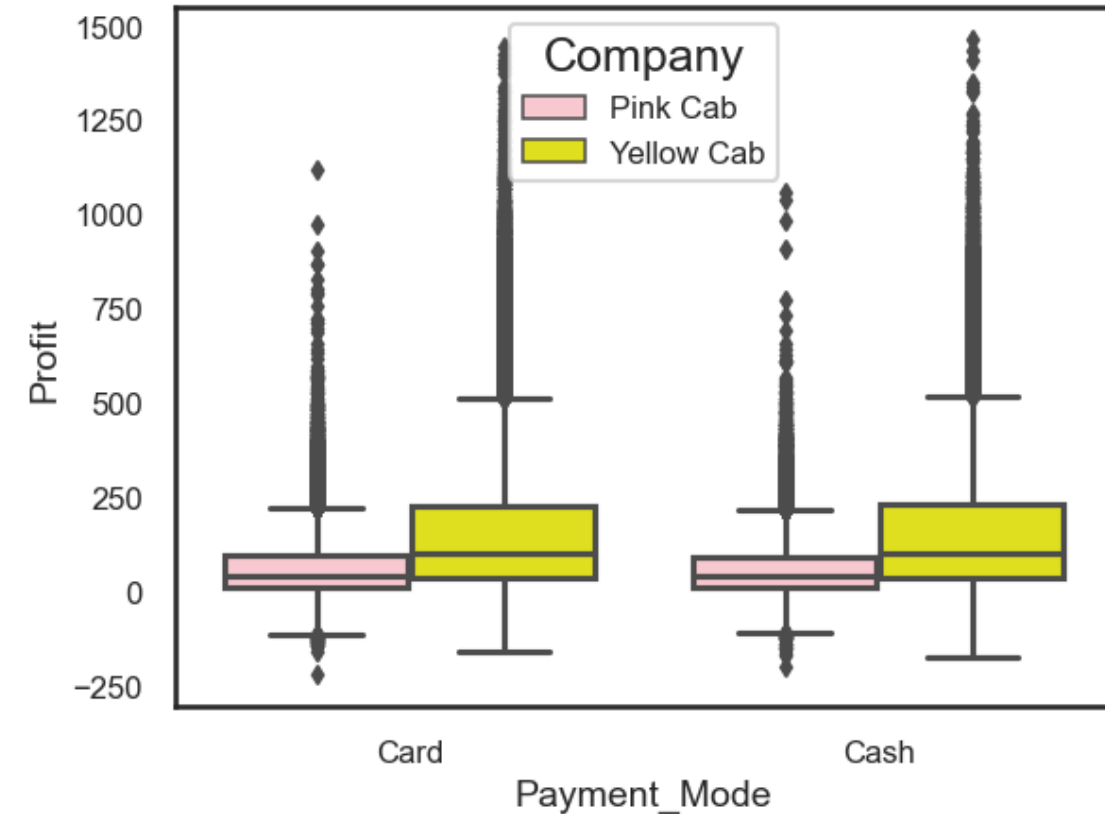
Is there any difference in Profit regarding Payment mode

* H_0 : There is no difference regarding Payment_Mode in both cab companies.

* H_1 : There is difference regarding Payment_Mode in both cab companies..

the p-value of Pink Cab 0.790 is greater than the level of significance, we fail to reject the null hypothesis

As the p-value of Yellow Cab 0.29 is greater than the level of significance, we fail to reject the null hypothesis.



The profit of the two companies similar for payment mode

Hypothesis testing

Is there any difference in Profit regarding Age

* H_0 : There is no difference regarding Age in both cab companies.

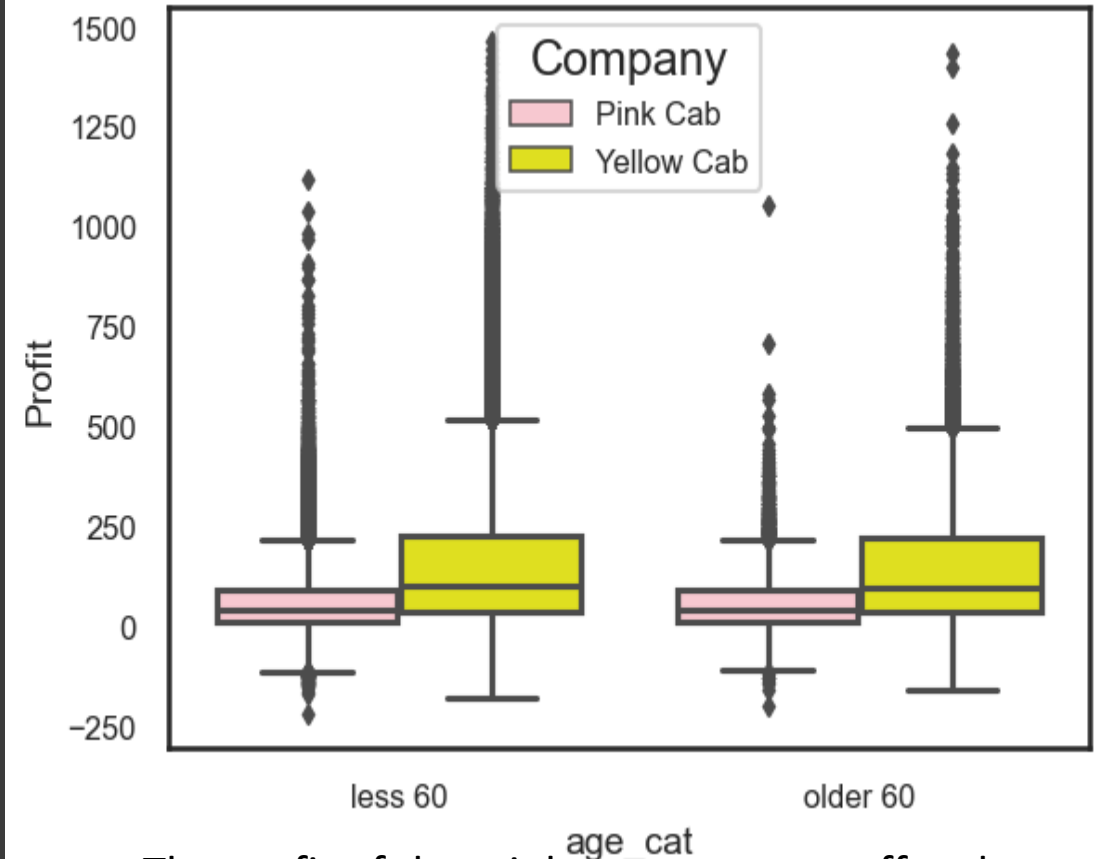
mean Yellow cab = mean Pink cab

* H_1 : There is difference regarding Age in both cab companies.

mean Yellow cab \neq mean Pink cab

The p-value of Pink Cab 0.48 is greater than the level of significance, we fail to reject the null hypothesis.

the p-value of Yellow Cab $6.32e-05$ is less than the level of significance, we reject the null hypothesis.



The profit of the pink company not affect by age.
The profit of the yellow company are higher for user >60 years.

Hypothesis testing

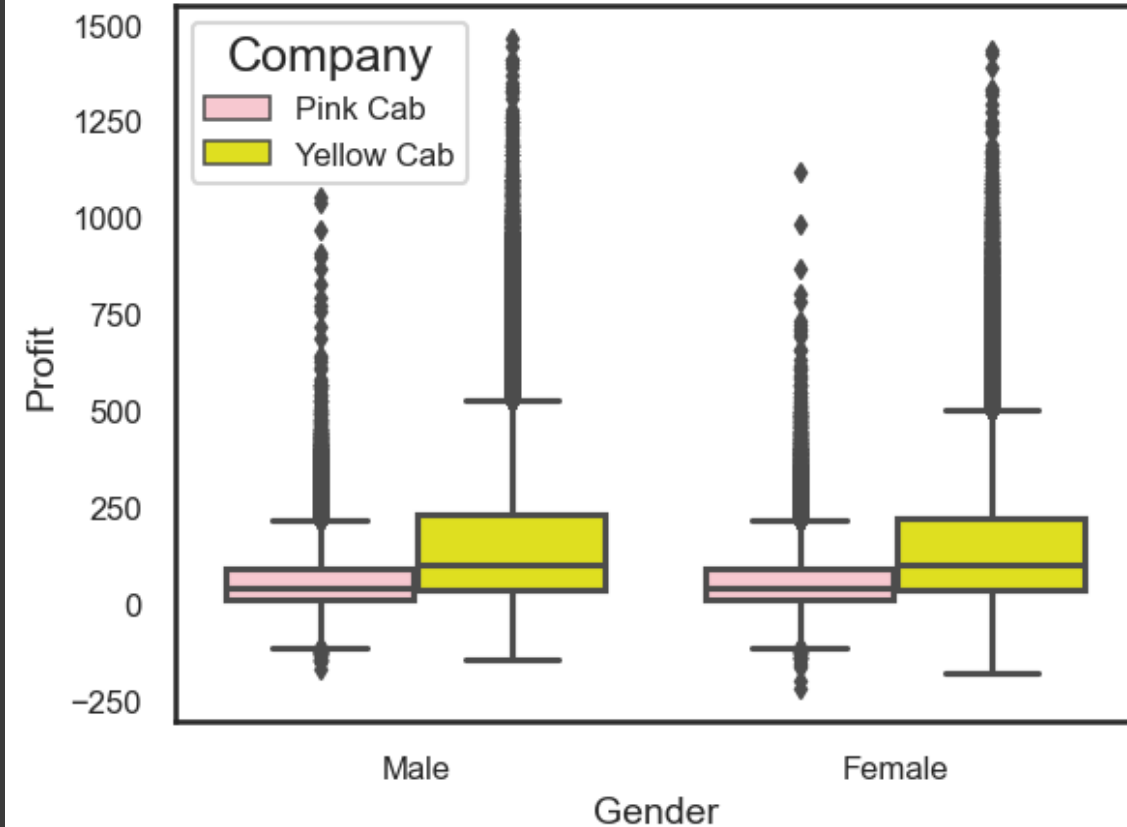
1. Is there any difference in profit regarding Gender

* H_0 : There is no difference regarding Gender in both cab companies.

* H_1 : There is difference regarding Gender in both cab companies.

the p-value of Pink Cab is 0.11 is greater than the level of significance, we fail to reject the null hypothesis.

The P values of the Yellow company is $6.060e-25$ is less than the level of significance, we reject the null hypothesis.



The profit of the pink company not affect by gender.
The profit of the yellow company are higher for male

Results and Recommendations

- I. Yellow cab has higher number of users around most of the Cities except for the Nashville, Sacramento it is 40%, San Diego and Pittsburgh 45%.
- II. Yellow cab is more popular in senior group while Pink cab has similar Yellow Cab popularity over <45 ages. For ages >45 the pink cab popularity is ups and down.
- III. The two companies profit differently over year and months, but Yellow Cab has higher profit.
- IV. Yellow Cab makes more profit with increasing the trip distance than Pink Cab.
- V. Pink Cab profits equally from both men and women, while Yellow Cab profit more from Men.
- VI. Both the companies show pattern of decline in the profit over 2017 . During 2018 Pink cab did not show much recovering as much as the yellow cab.

On the basis of above point , we will recommend Yellow cab for investment.



Thank You

