Exploratory Data Analysis

Project: Go-to Market Analysis for an Investment Firm

Date:18 Jan-2023

Presented by: Shruthi Madgi

Agenda

The Client:

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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Objective:

Provide insights to firm via Exploratory Data Analysis of the data provided and help XYZ to make a decision on where to invest.

Outline

1.Provide the information on the dataset, i.e. data used and any outliers.

2. Analysis of the data, i.e. which company has more users, more profit per ride etc.

3.Hypothesis tests, i.e. any difference in profit in regards with gender, etc.

4. Summary and result of the analysis.

Dataset Information

Total of 5 datasets are used:

1.Cab_Data.csv: Details the transactions of two companies: Yellow and Pink.

2.City.csv: Contains list of cities, population of each city and total number of cab users.

3.Customer_ID.csv: A mapping table which contains unique identifier that links a customer's demographic details.

4.Transaction_ID.csv: A mapping table which details payment and mode of payment a customer used.

5.US Bank holidays.csv: A file detailing number of US Holidays.

ASSUMPTIONS ABOUT DATA:

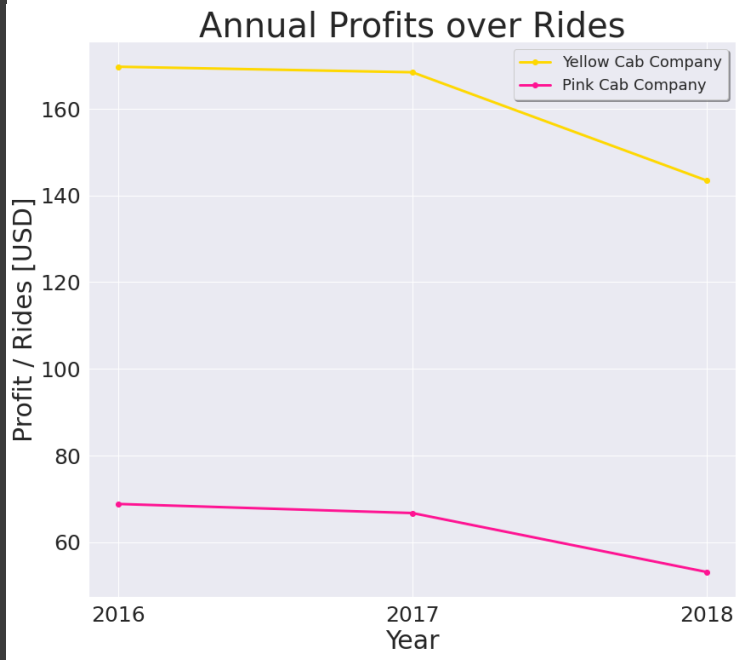
a)Outliers are present in the data such as Price charged, but due to lack of enough information, it is not treated as an outlier.

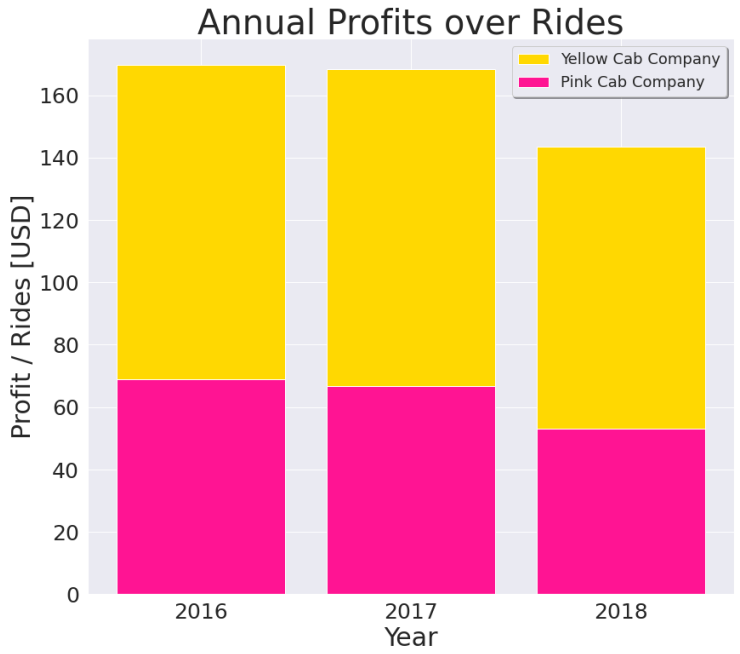
b)After Merging if any record has at least 1 NaN value, the whole record is dropped. it means there is not enough information of that trip.

c) Profit is calculated by : Profit per Ride = Total Profits over a certain Period of Time / Number of Rides over that period of Time.

Profits per company

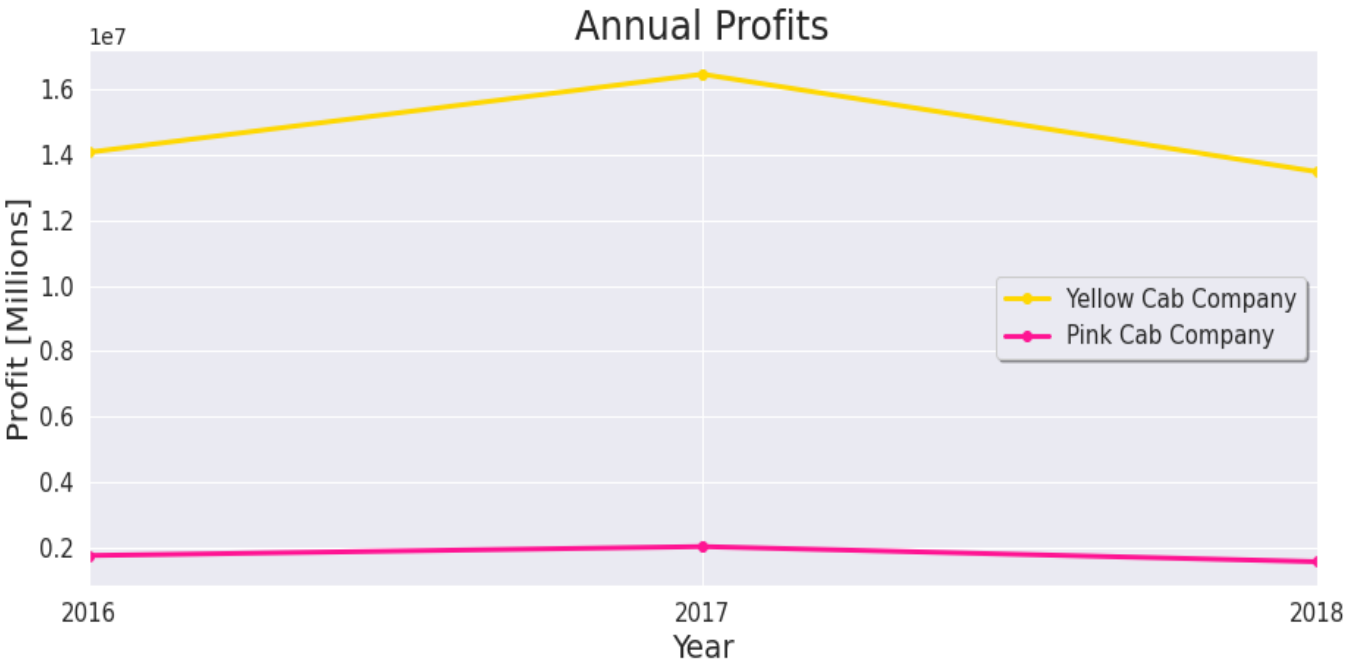
Ride per KM, City wise profitable ride





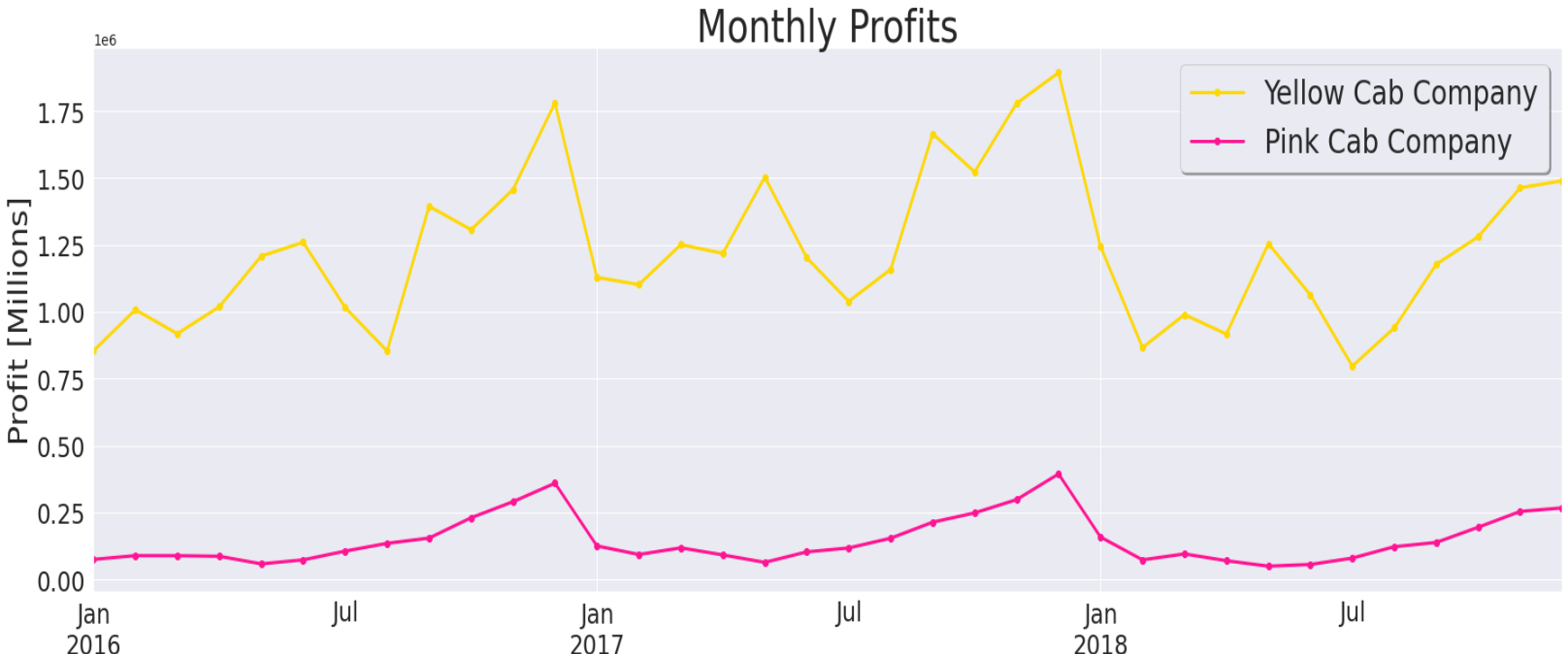
Profit per ride decreases over time in both companies.

Data Glacier
Your Deep Learning Partner



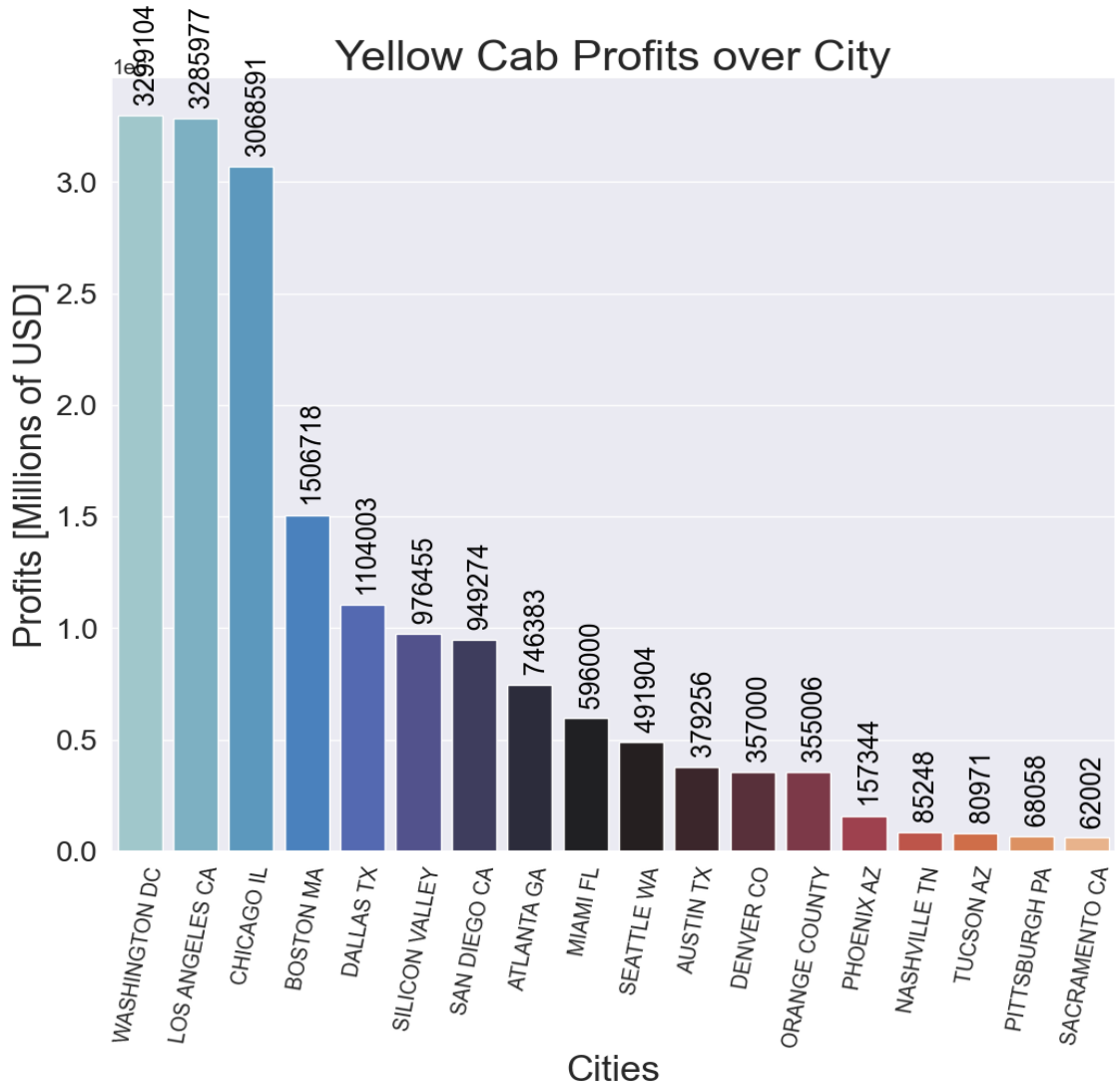
Yellow Cab's profits are 8 times higher than Pink Cab's Company over the last 3 years.

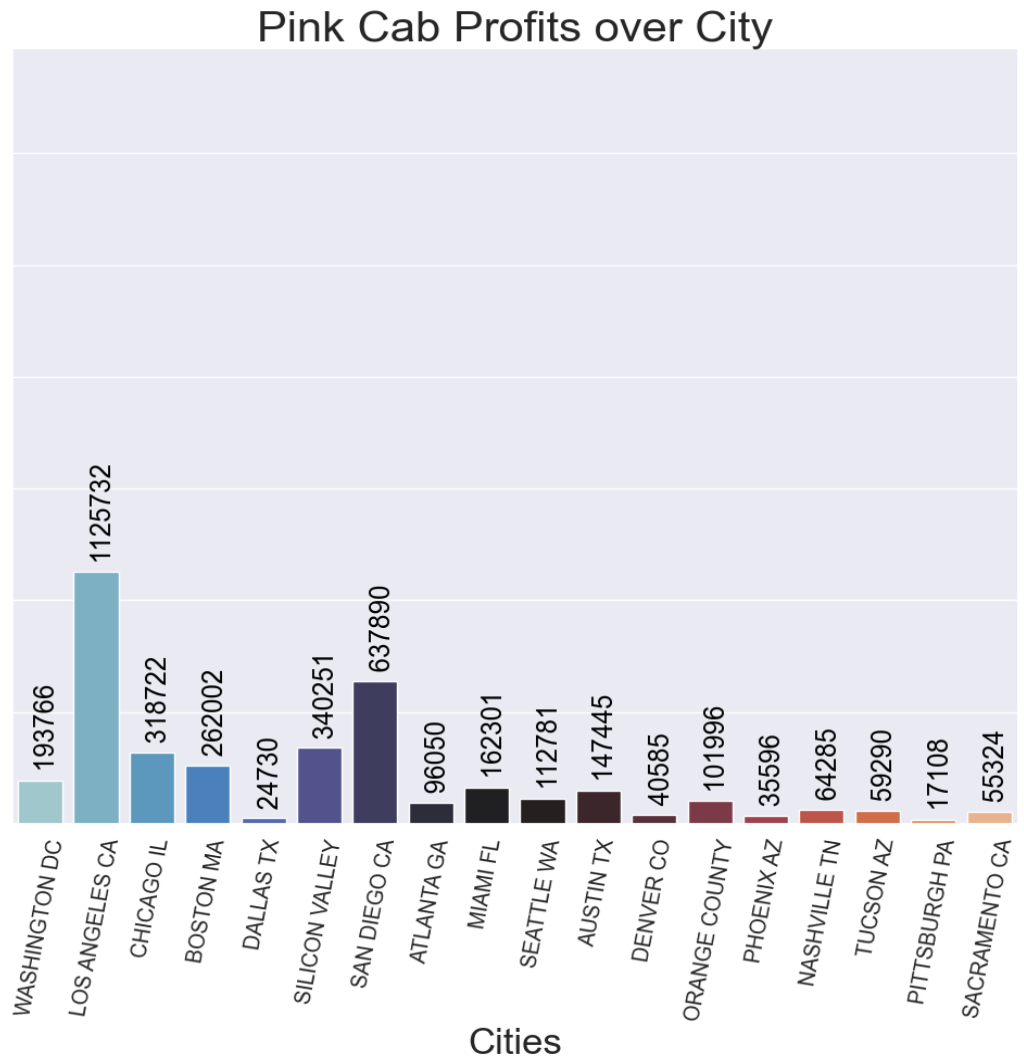
Yellow Cab Company's earnings are more stable with fluctuations of 23.08% while Pink Cab Company's earnings vary with fluctuations of 61.22%

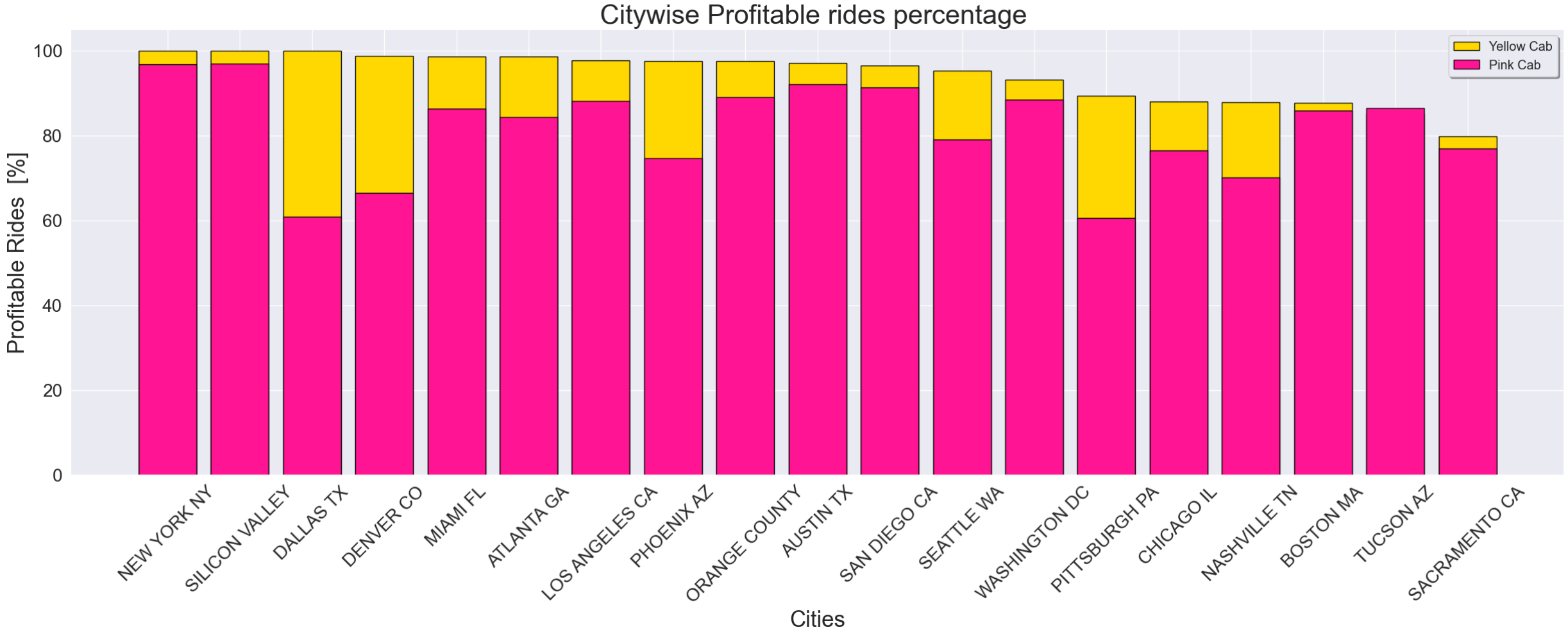


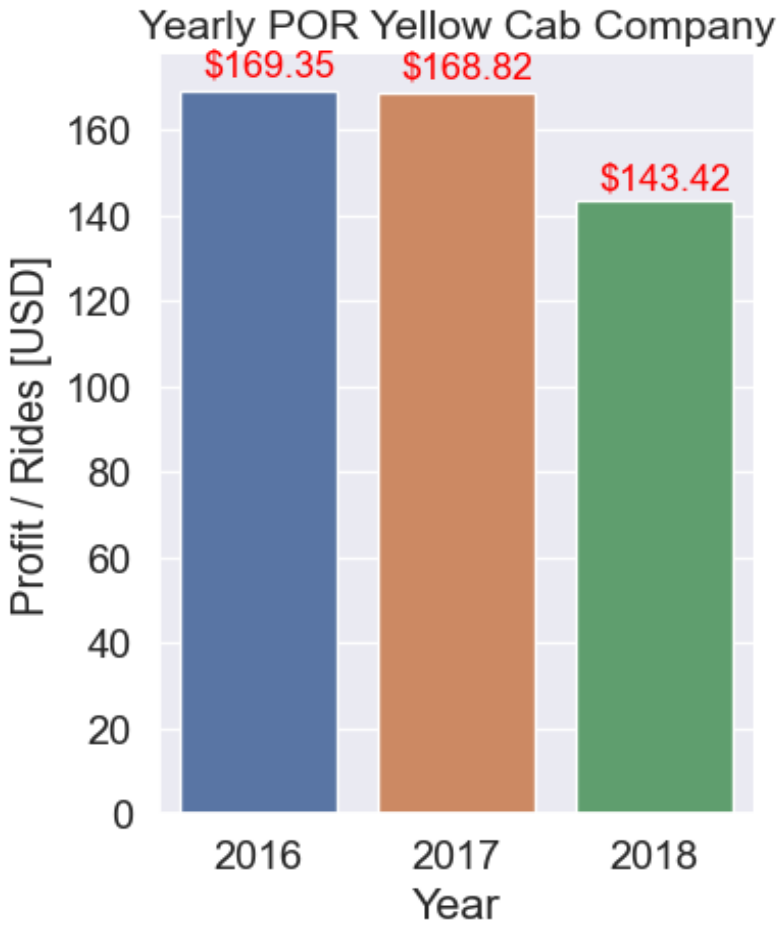
Data Glacier

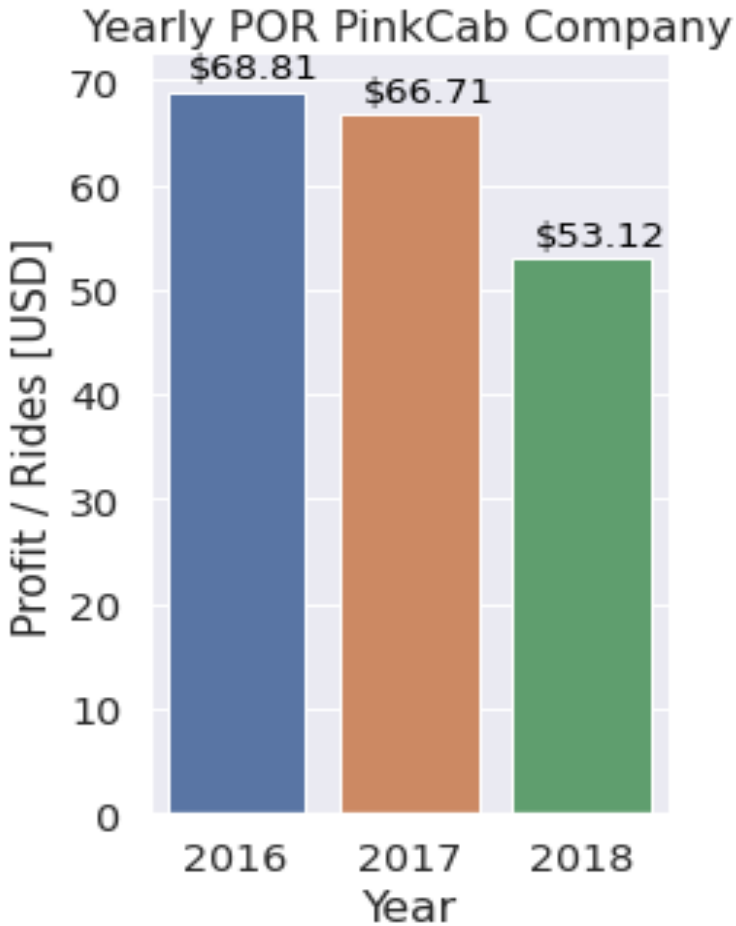
In this analysis, New York City has been removed from both cab companies to get a better sight of the profits over the other cities.
Conclusion is that Yellow Cab has greater market share in every City.



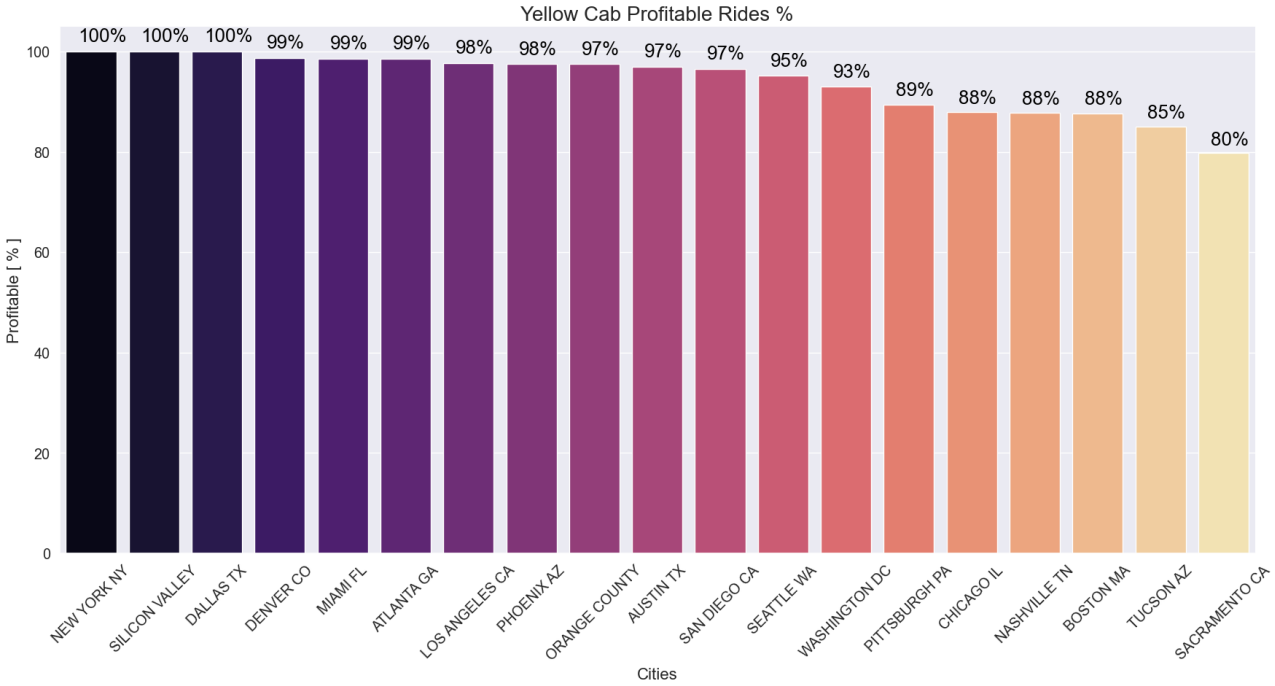




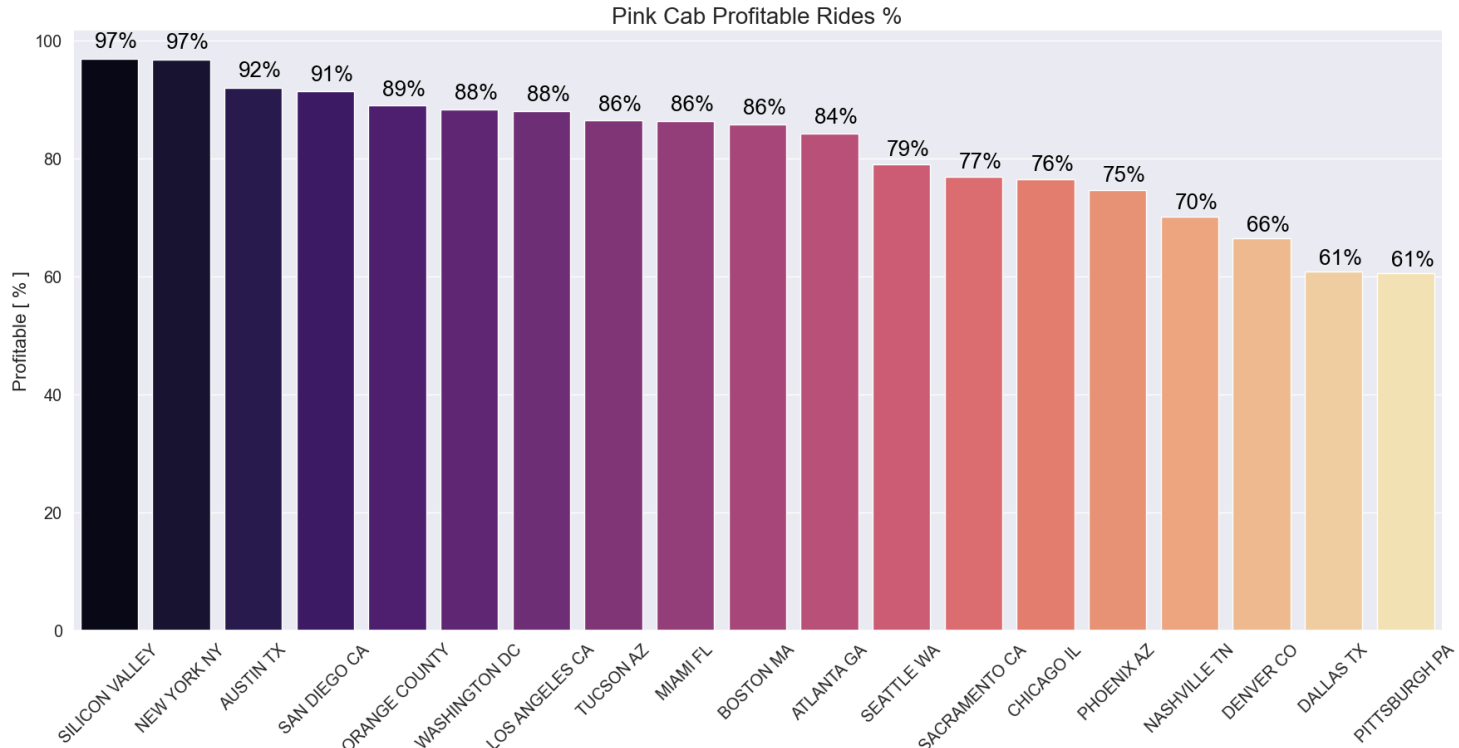




Profit per ride of Yellow cab company is higher than Pink cab company over three years.

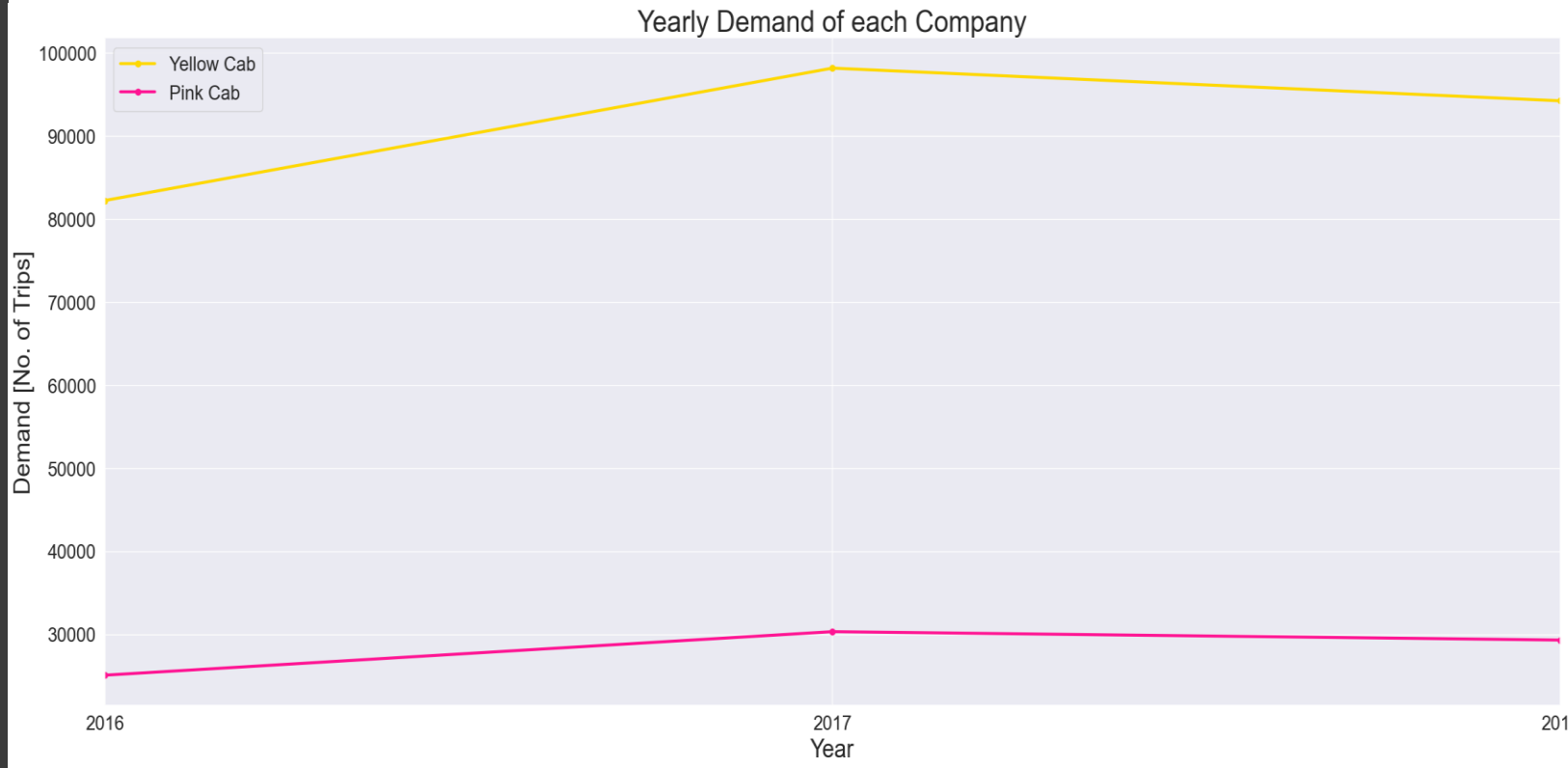


Assumption is if the profitability percentage of rides per city is higher than 80%, it will perform well. Profitability percentage of rides change by cities and Yellow Cab has a high performance by maintaining a high percentage of profitable rides in every city.



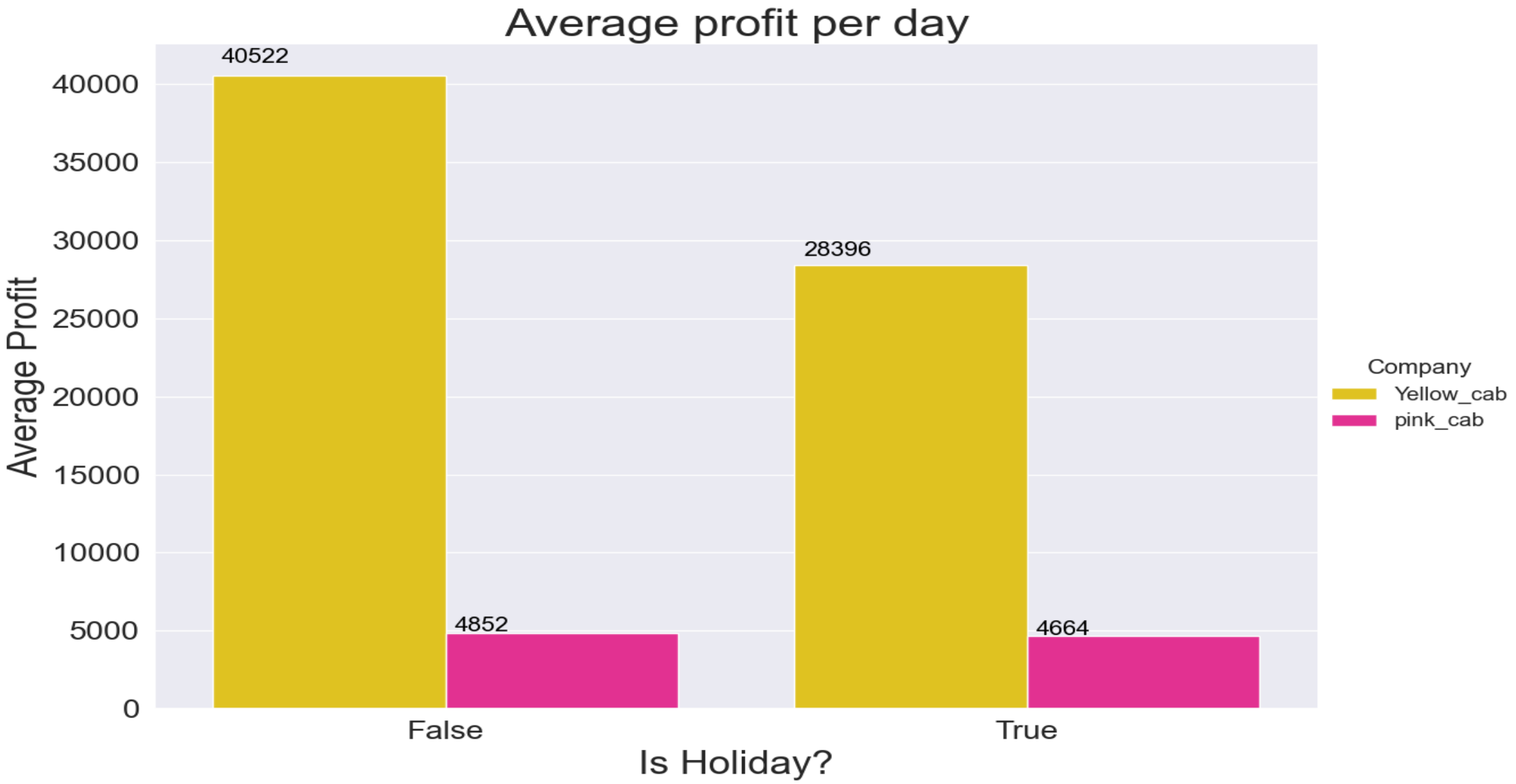
Data Glacier

Yearly Demand, Holiday Profits,& and Demand by month and Age

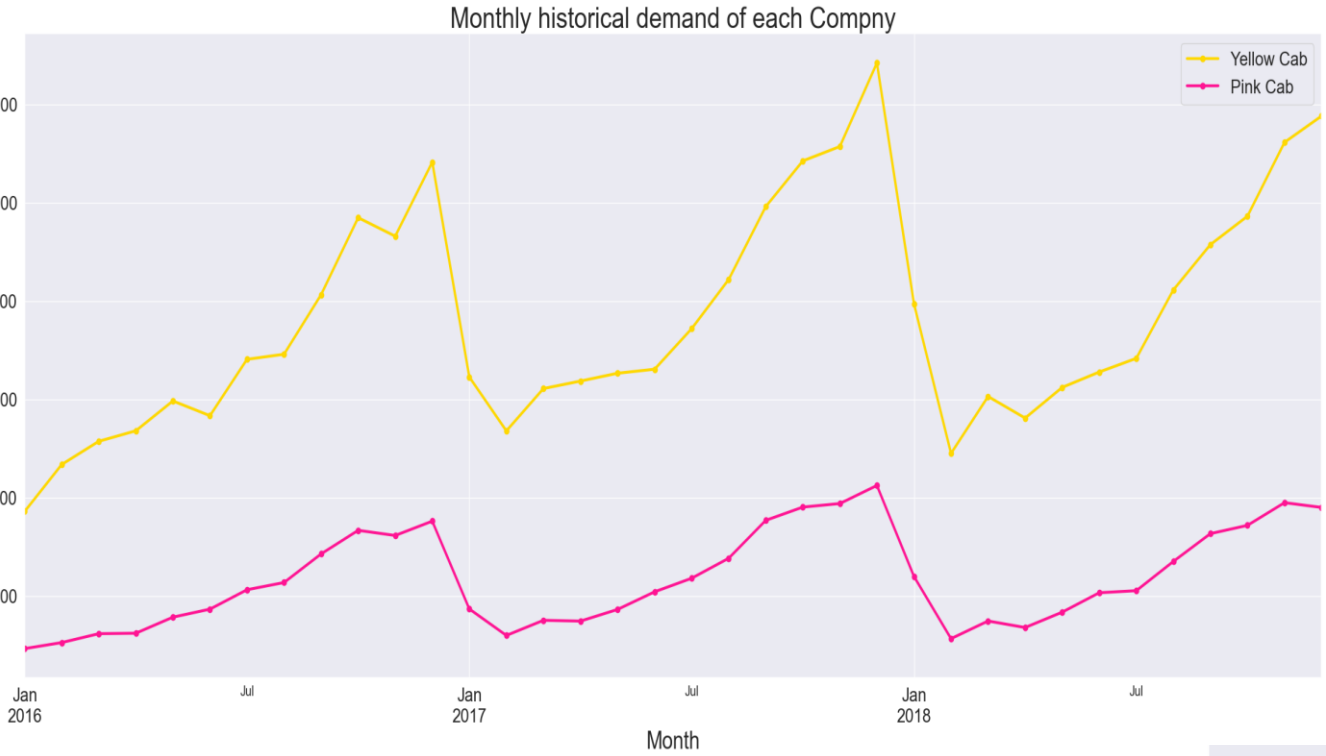


Yearly demand of Yellow cab company is nearly 4 times greater than that the of Pink cab company.

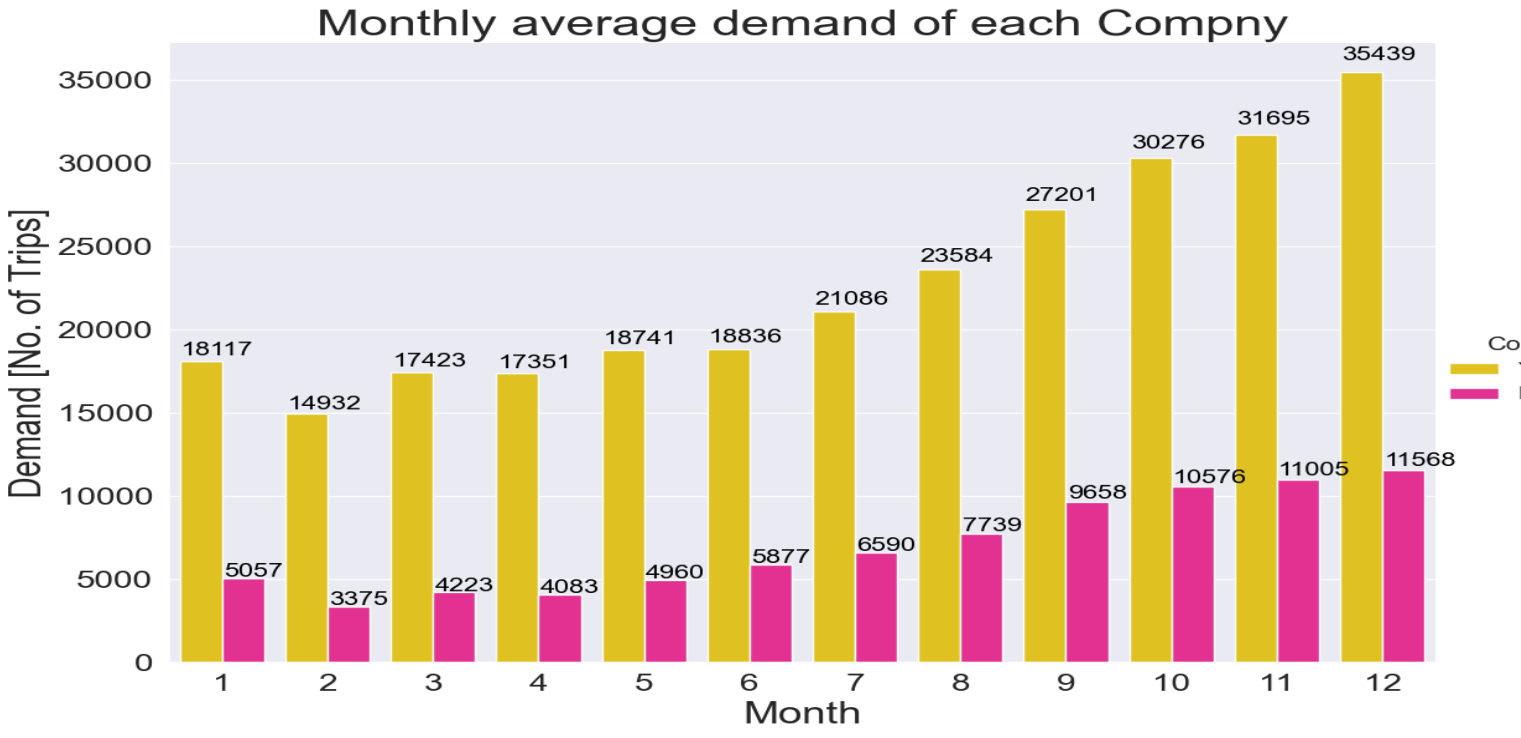
Data Glacier
Your Deep Learning Partner



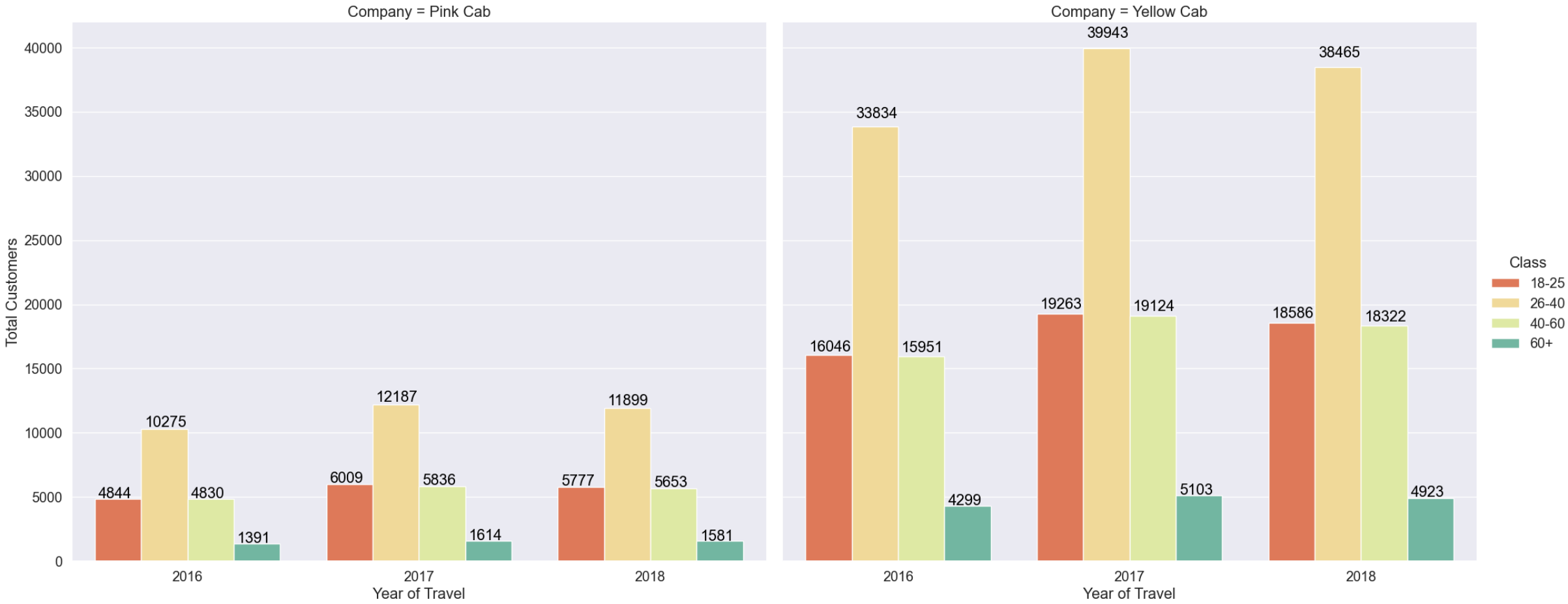
In Yellow cabs, the average profit per holiday is less than the average profit per normal day. But when considering the average profit per day in pink cab company, they have gained a bit of higher profit on holidays than normal days.



The highest demand is in December.



Data Glacier



Each year, both the companies have more customers in the age group of 20-40 years.

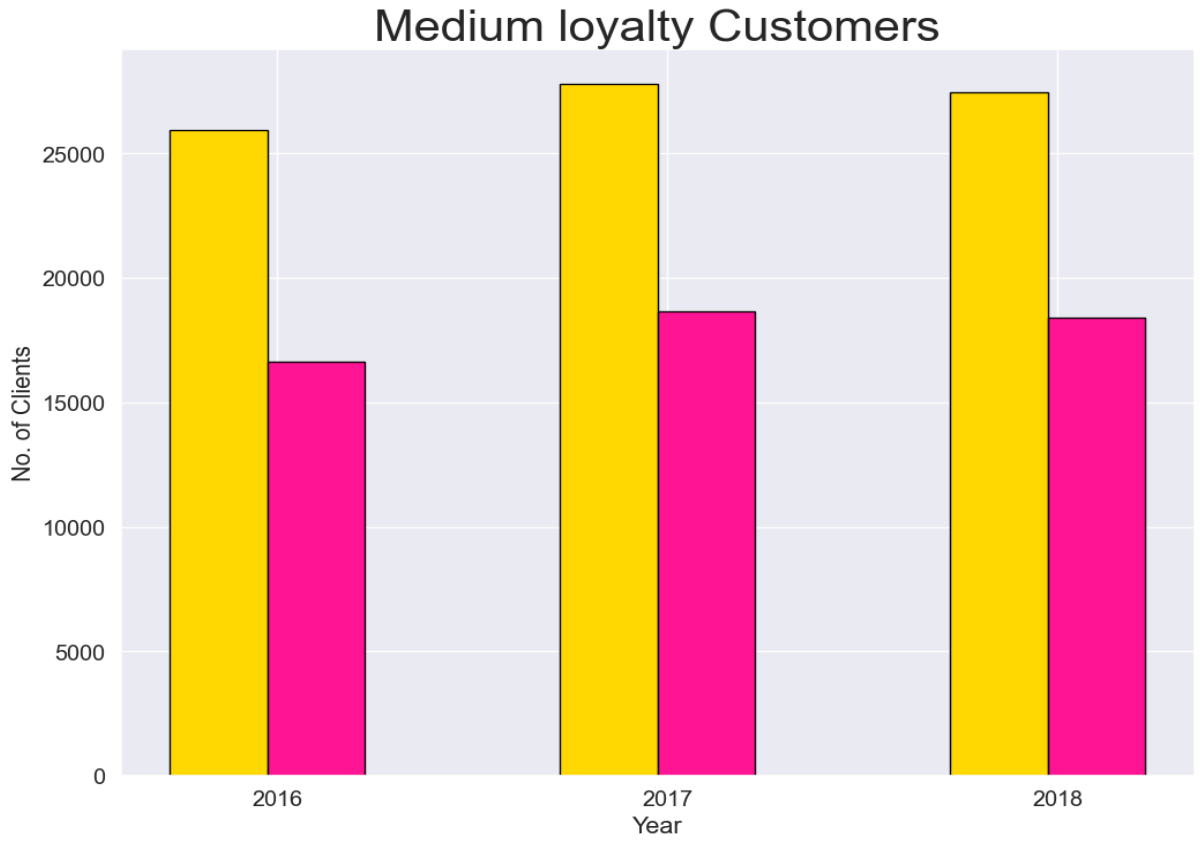
Client usage analysis: Retention rate & Payment modes

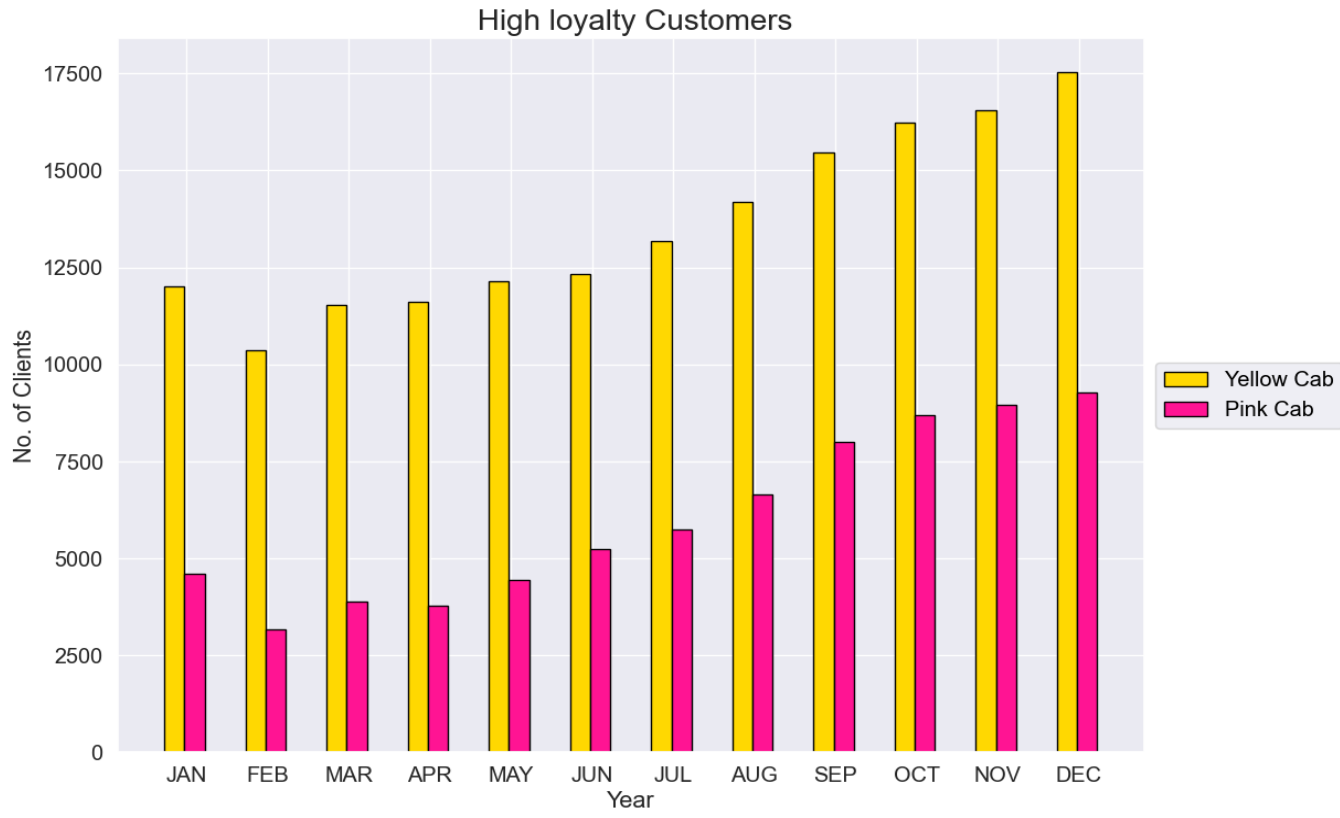
To analyze the Loyalty rates, two classes are defined in the code:

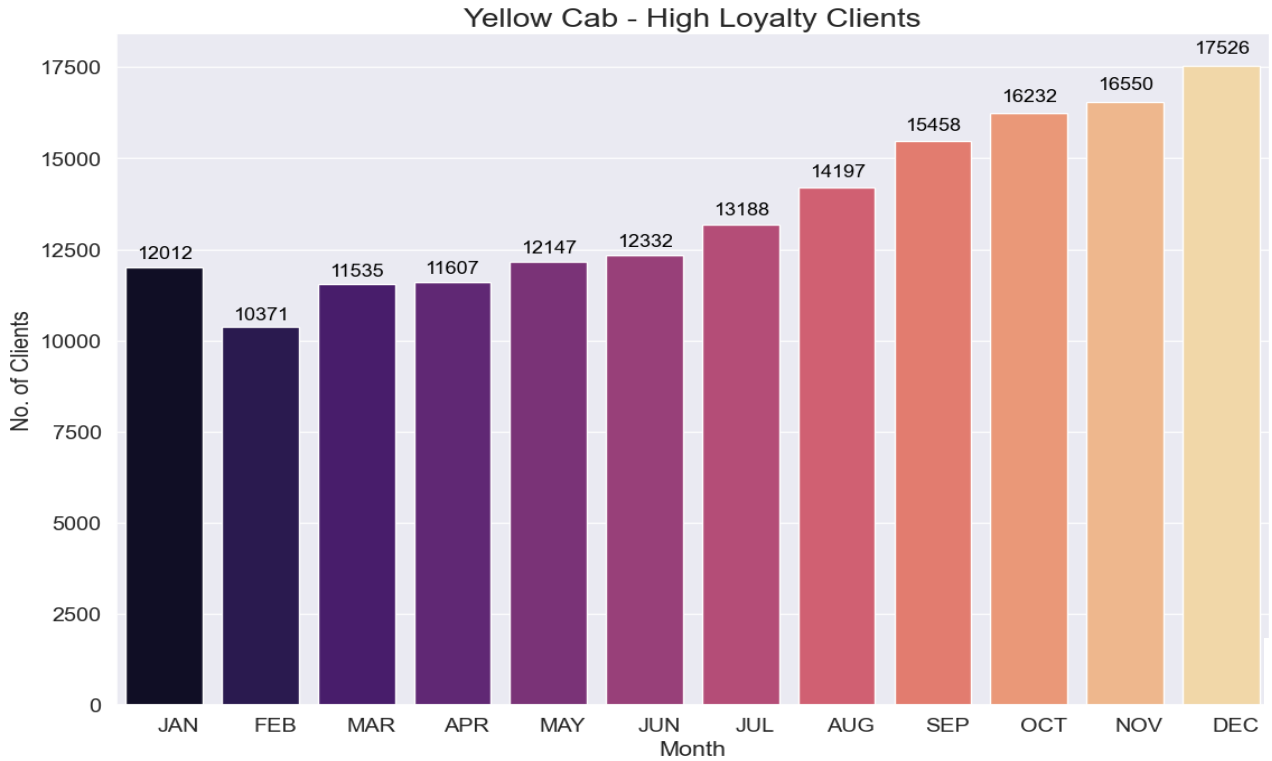
1. Medium loyalty Customers: Customers who took more than 10 rides yearly.

2. High loyalty Customers: Customers who took more than 10 rides monthly.

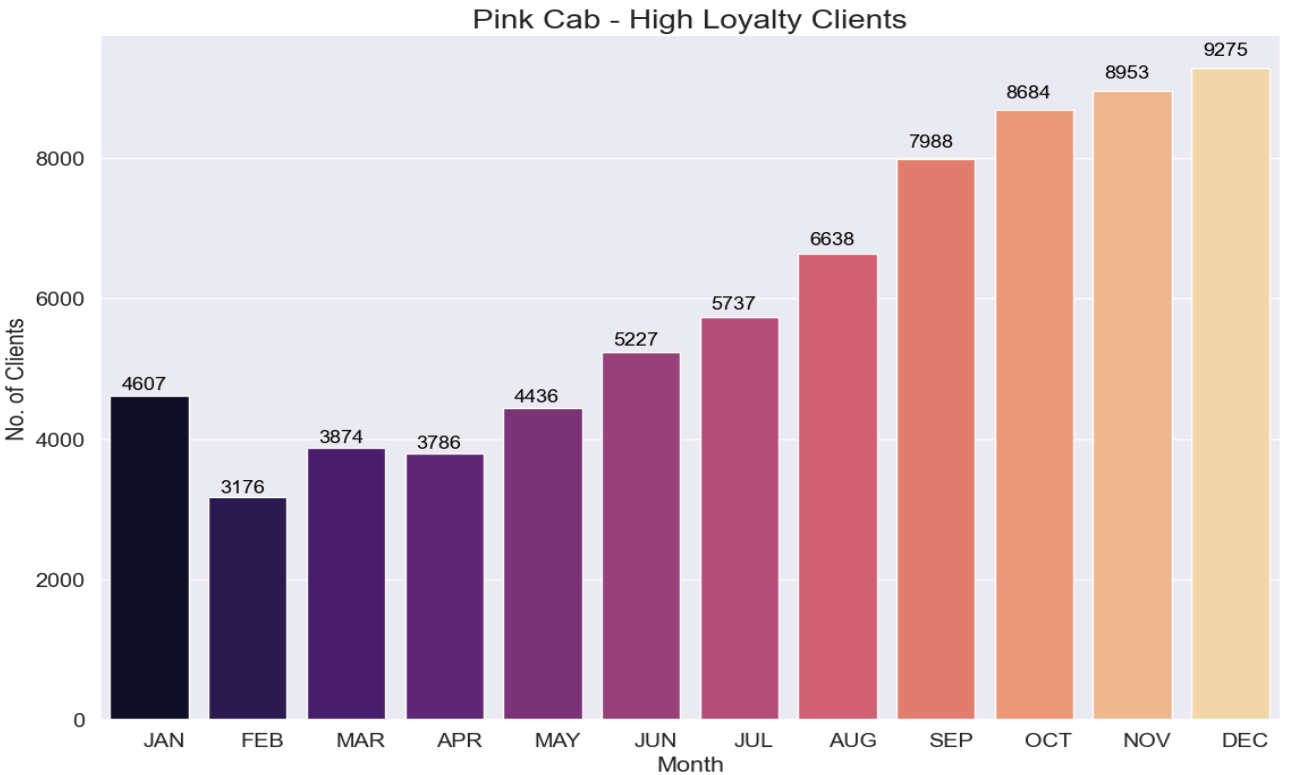
Minimal deviations in payment modes are observed over the time.



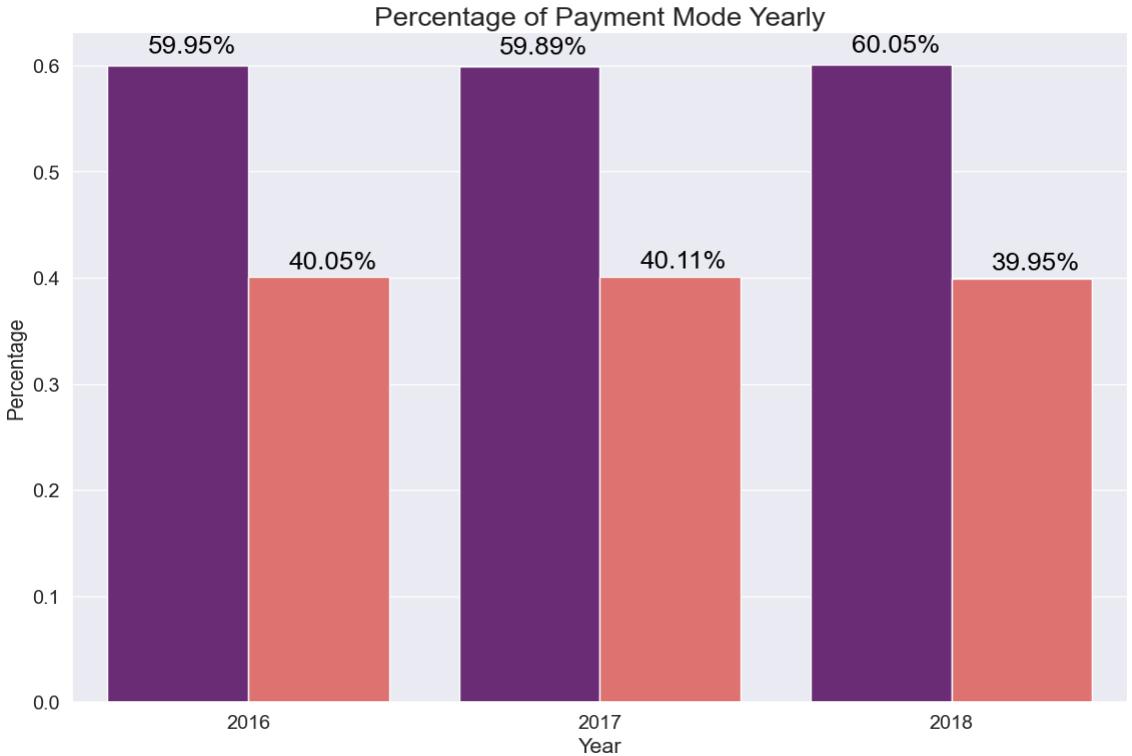




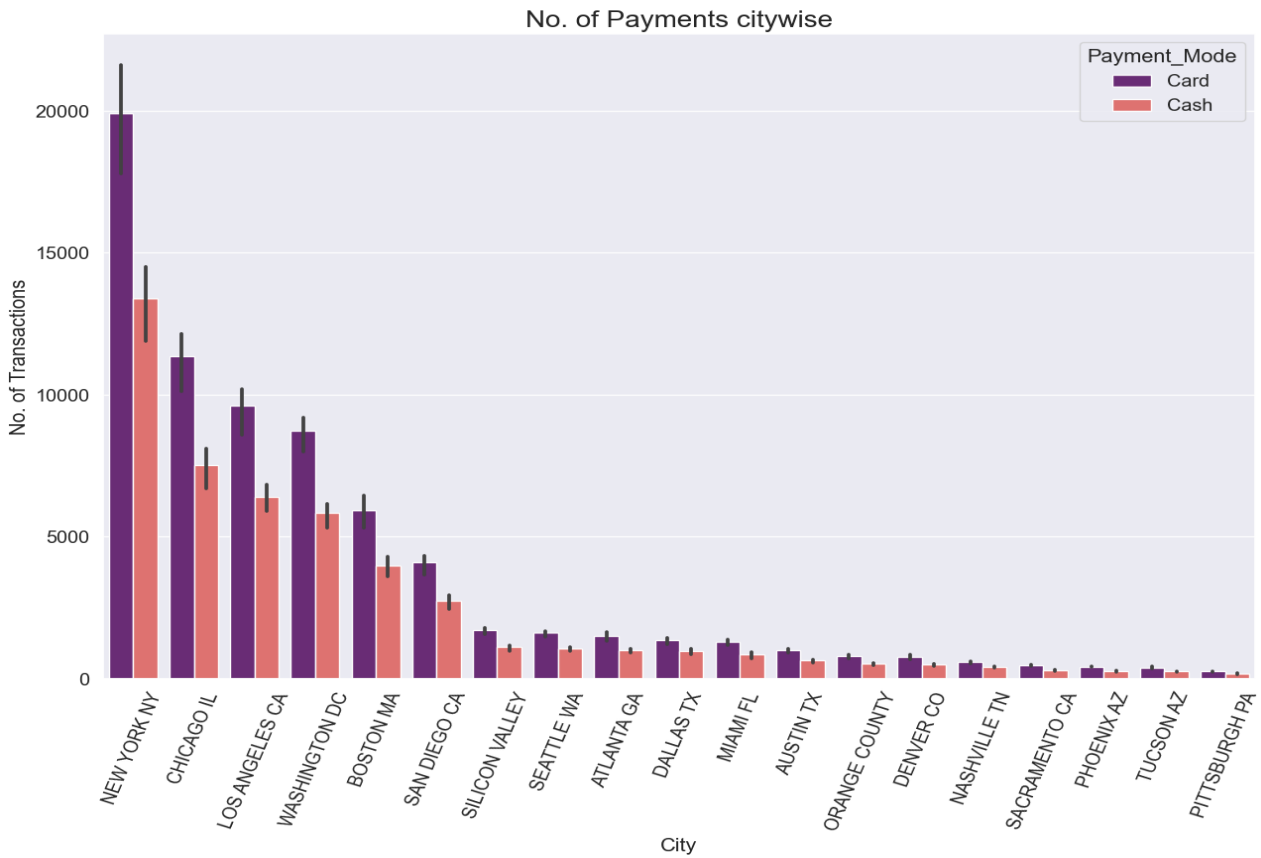
Yellow cab company is doing better in retaining customers than the Pink Company.



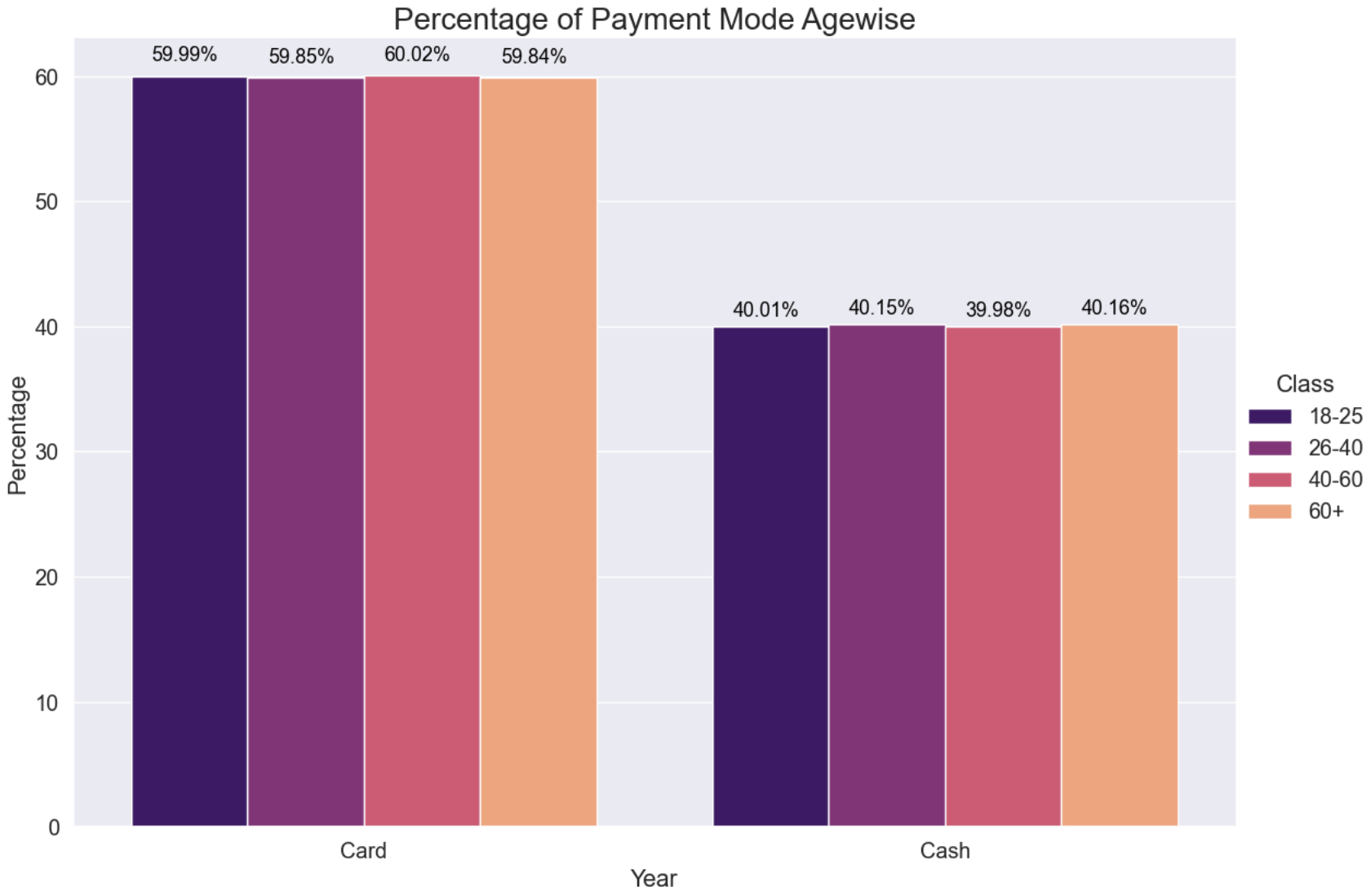
Data Glacier



Minimal deviations in payment modes, both yearly and city wise are observed.

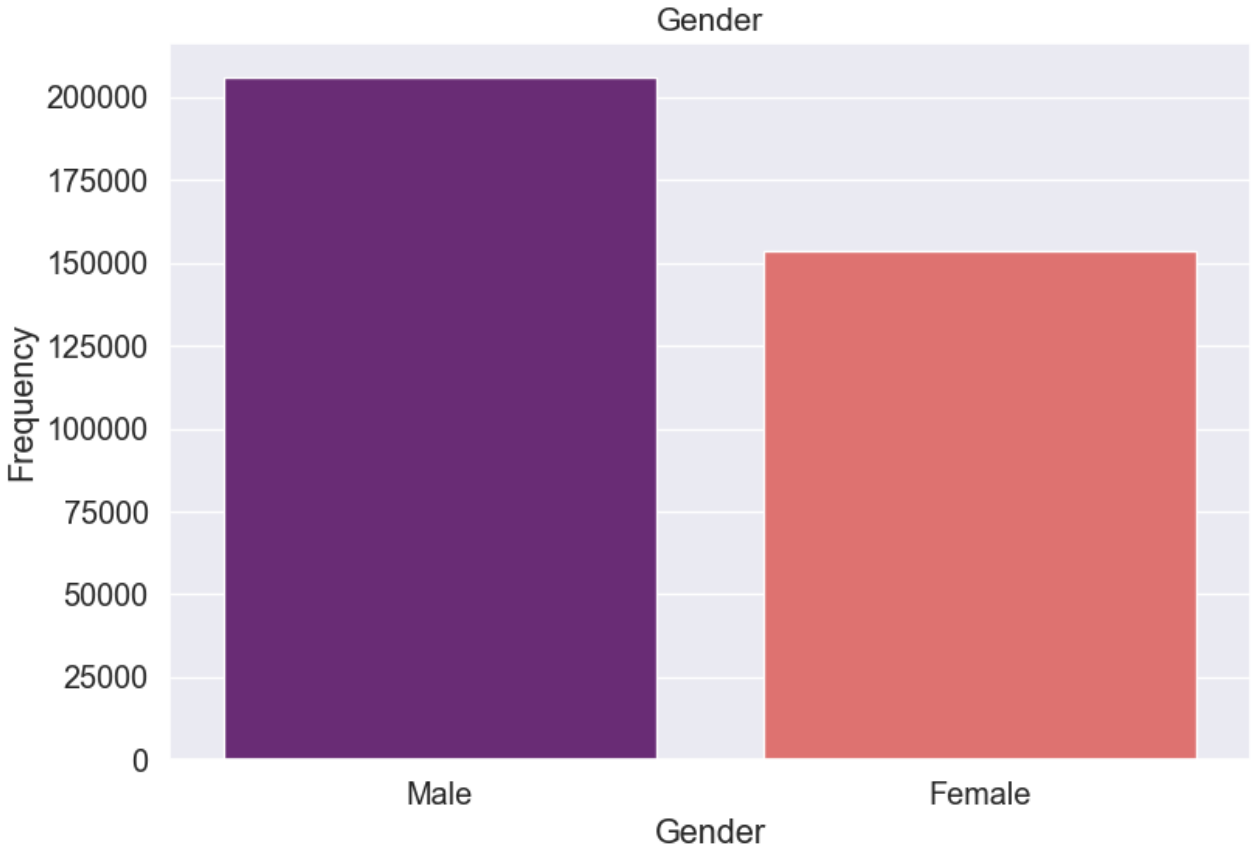


Data Glacier



Minimal deviations in payment modes are observed across the ages.

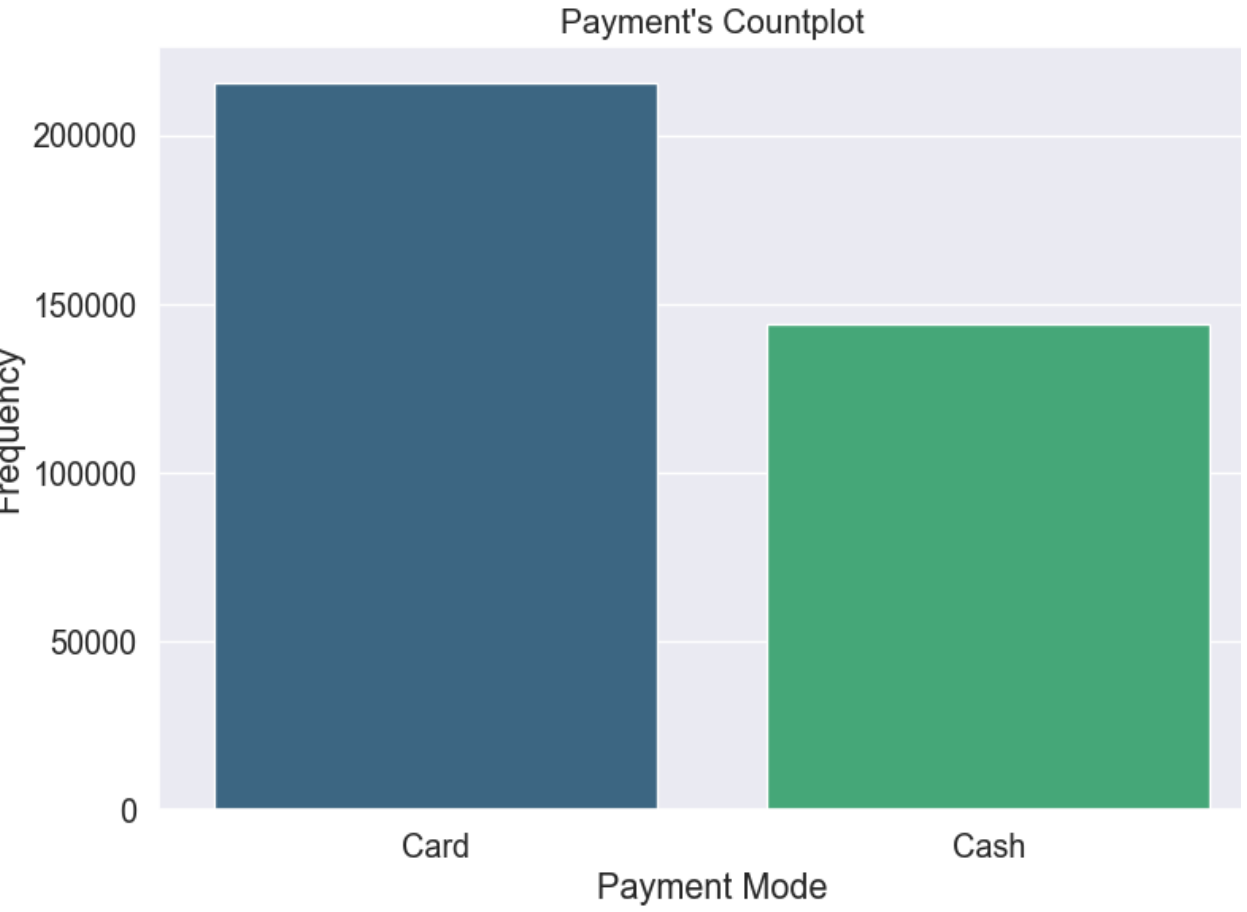
Hypothesis: Difference in profit with respect to gender, age and payment

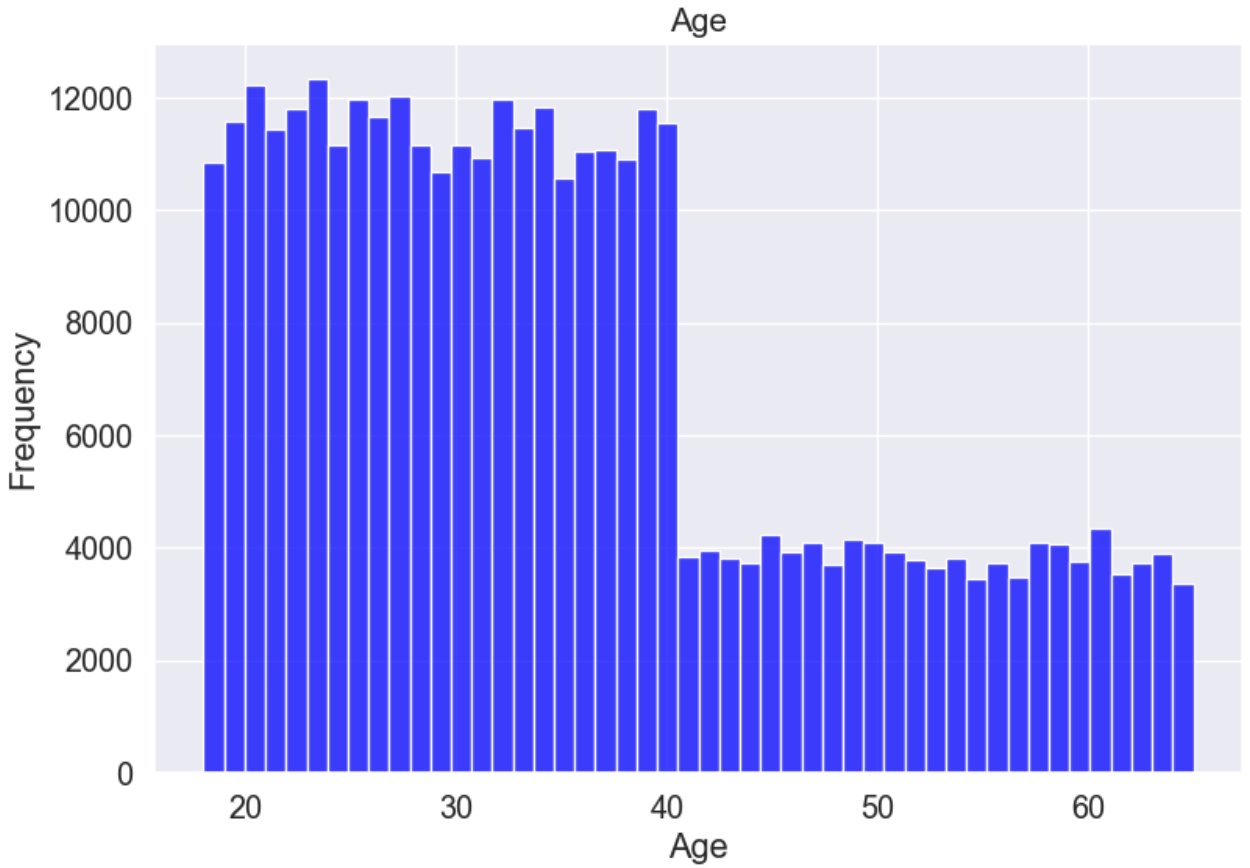


1.Men are majority users by slim margin in both companies.
2.Men seem to be more profitable for both the cab companies.
3.Yellow Cab has more customers in both genders.

Gender has no significant effect on profits in both companies.

Data Glacier
Your Deep Learning Partner





Payment preference is strongly biased towards Card to cash; ratio is 3:2

Looking at the age distribution of all users, it is seen that approximately half of them are between the ages of 18-25.

Data Glacier

Conclusion

- Yellow Cabs experience more sharpening effects than Pink Cab, both in increasing and decreasing profit and other factors.
- Investing in Yellow Cab can be a better solution, since the data indicates that risk factors associated with Yellow Company is lesser than that of the Pink.
- Yellow Cab company is more profitable in every aspect.

Yellow Cab Company seems to be the perfect choice for the investment and return on profits.