

Exploratory Data Analysis

Cab G2M Report

Report date: 30-03-2021

Internship Batch:LISP01

Version: 2.0

Data intake by: BRIAN ORINA BUNDI

Data intake reviewer: Data Glacier

Data storage location: <https://github.com/Brianorinah/Cab-EDA>

Background – G2M(cab industry) Case Study

- 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.
- Objective :- Provide actionable insights to help XYZ firm in identifying the right company for making investment.
- The analysis has been divided into three parts:
 - Data Understanding
 - Finding the most profitable Cab company
 - Recommendations for investment

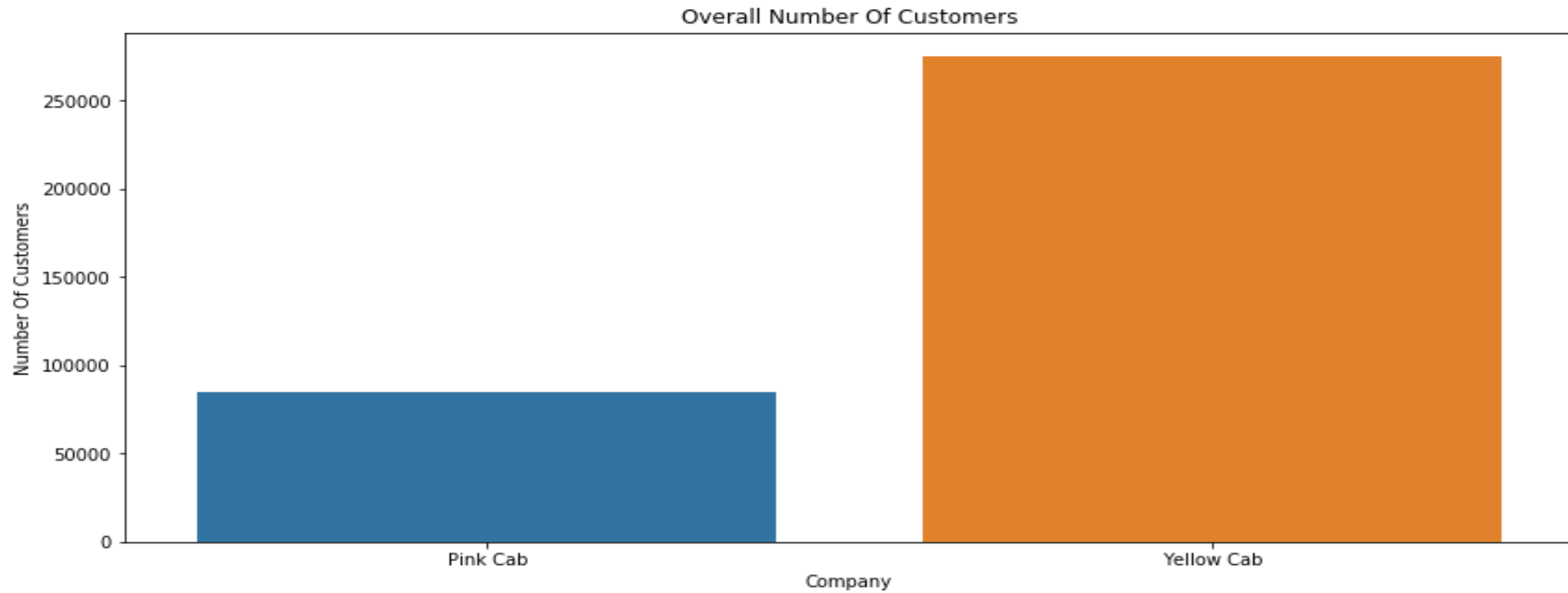
Data Exploration

Total number of observations	355,032
Total number of files	4
Total number of features	15
Base format of the file	.csv
Size of the data	31,904 KB

- **Assumptions:-**

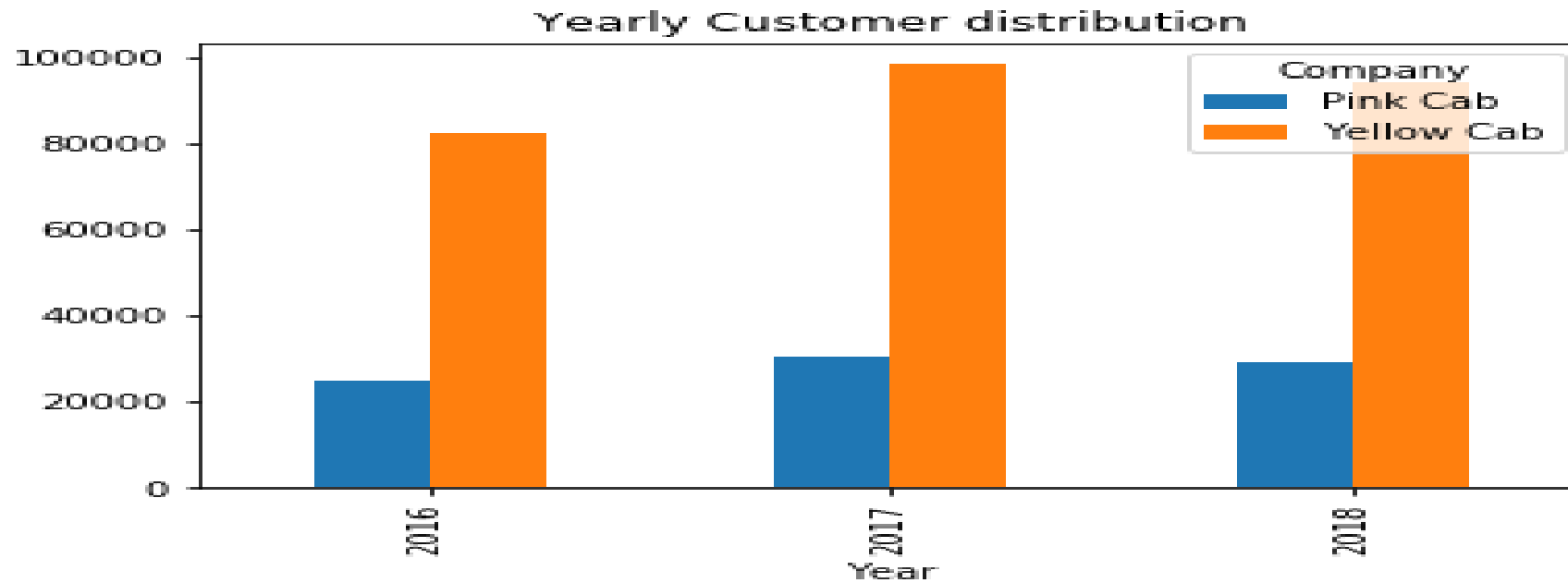
- Profit of rides are calculated keeping other factors constant and only Price_Charged and Cost_of_Trip features used to calculate profit.

Data Understanding



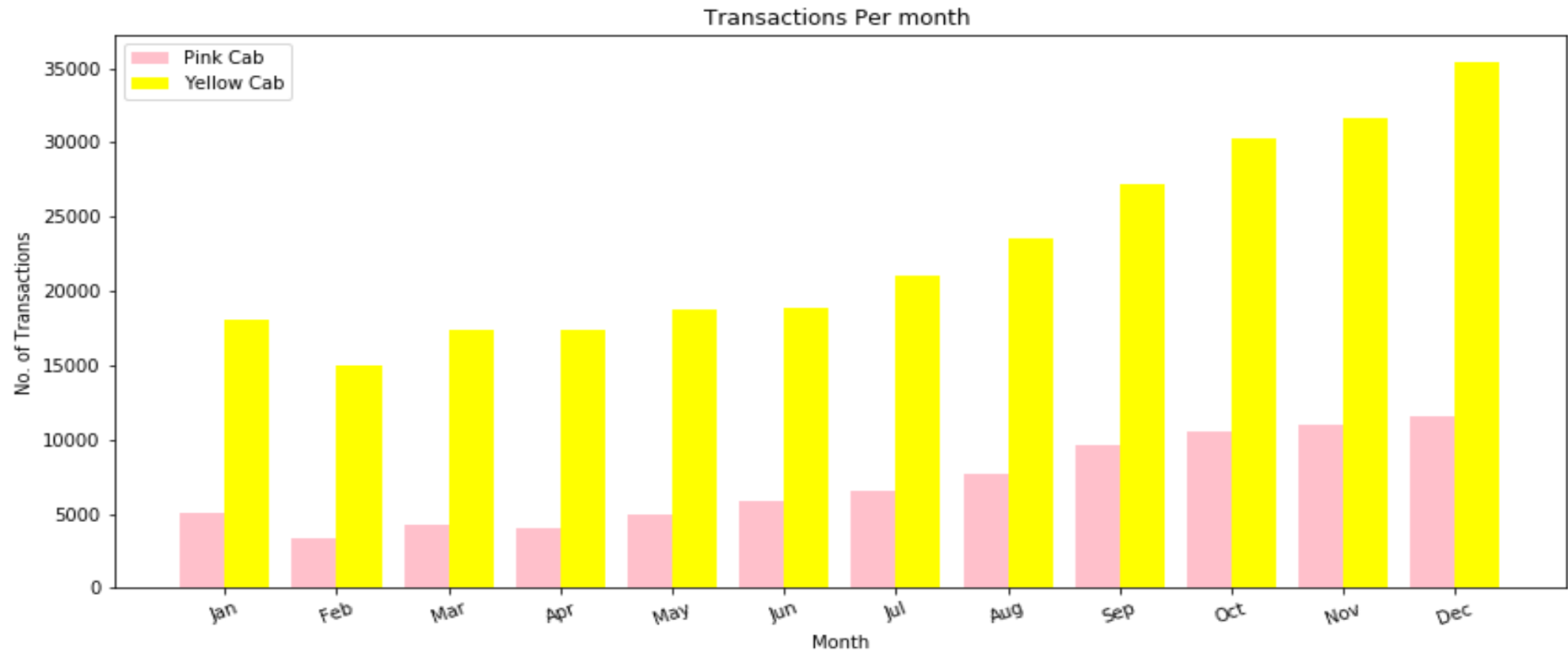
✓ Yellow cab has the majority of transactions between the two cabs.

Yearly Customer Distribution



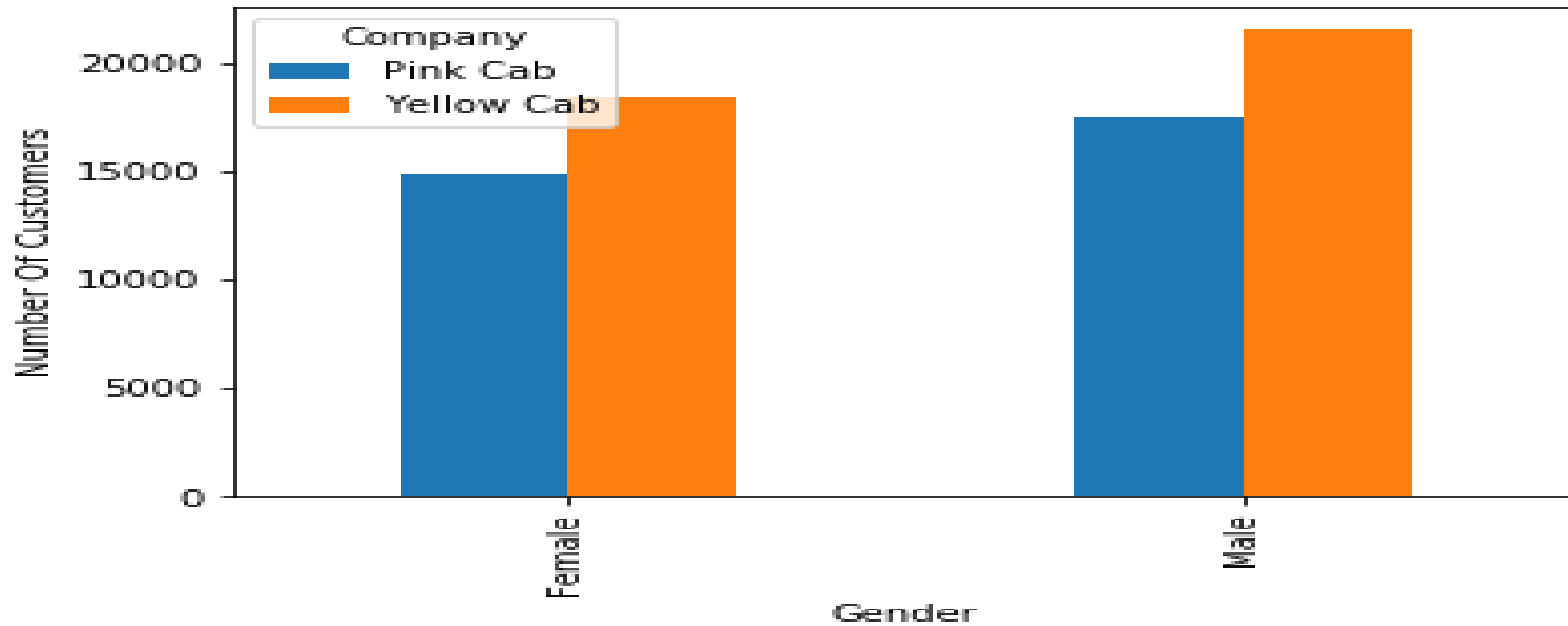
- ✓ Yellow Cab has had the majority of customers accross the three years.

Monthly Transactions Analysis



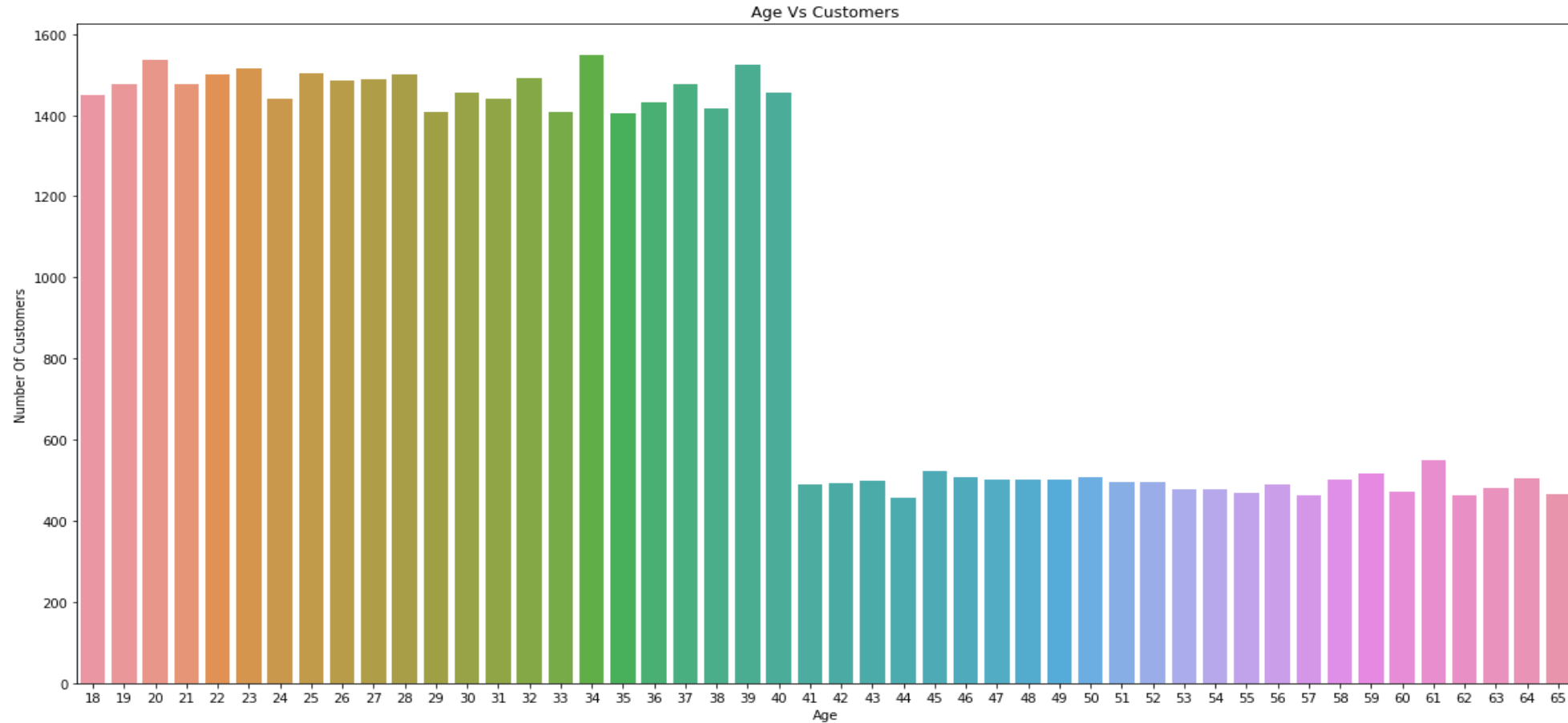
- ✓ The busiest Months are the holiday months heading towards January for both Cabs with Yellow cab taking the lions share

Gender Distribution



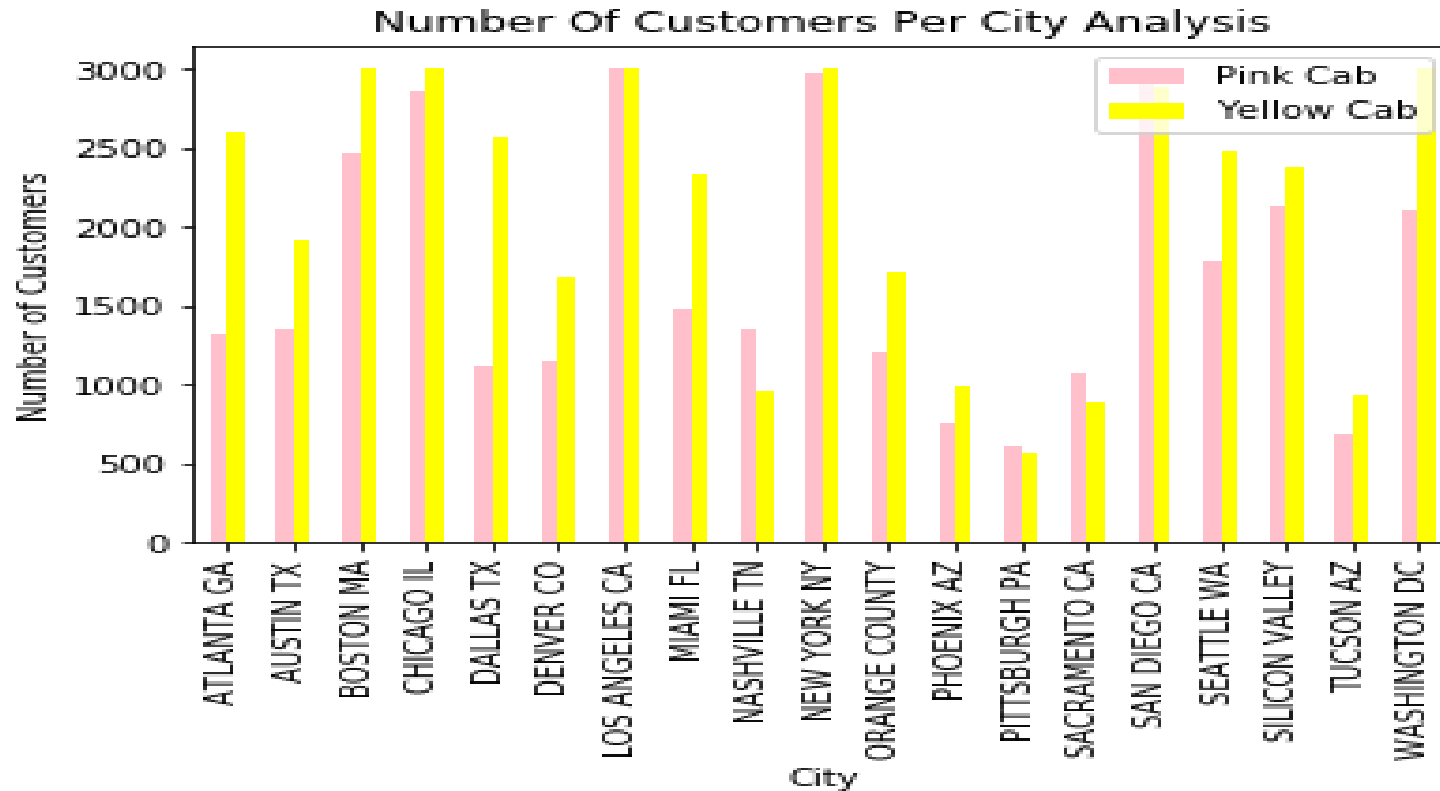
- ✓ Most of the customers for both Cabs are Male with the Yellow Cab taking the lions share in both

General Age Distribution Analysis



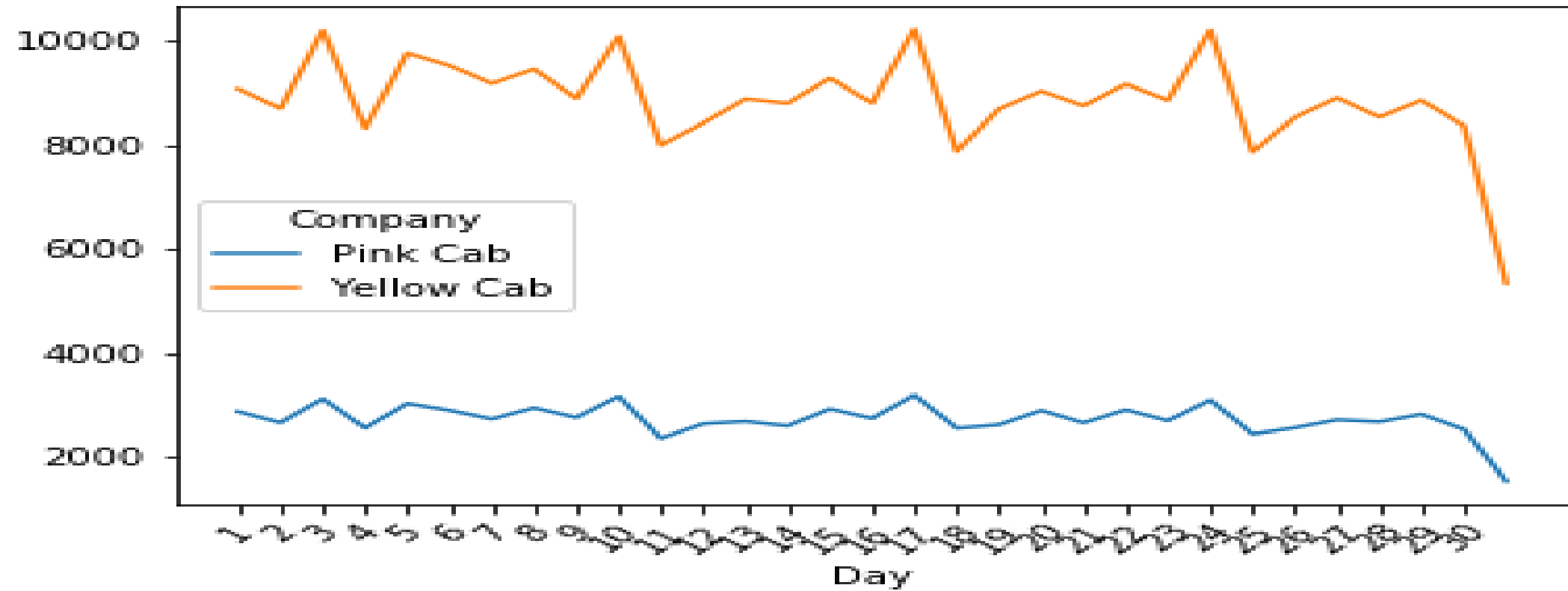
✓ Most of the Customers are Young i.e below 40 Years of age

City Customer Analysis

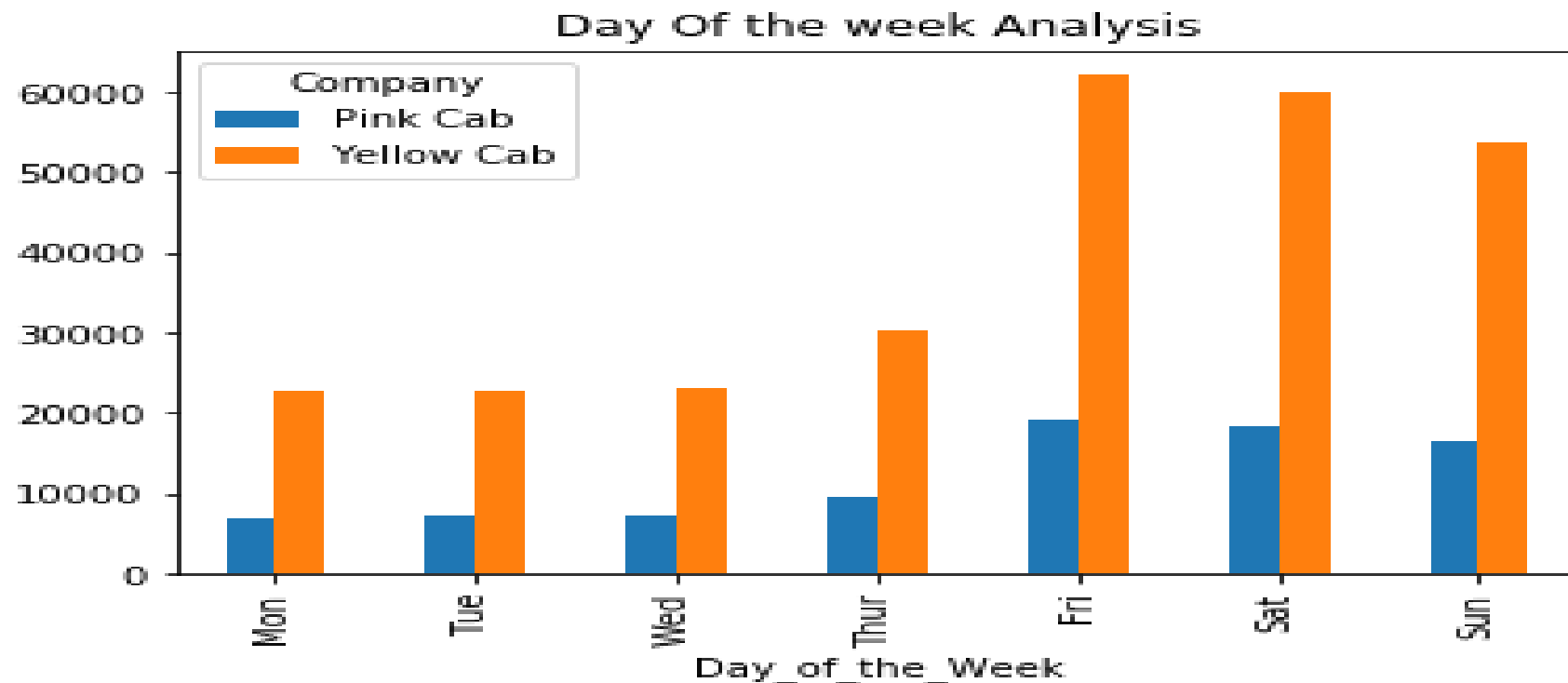


- ✓ Yellow cab leads on most of the cities with Pink Cab only leading in 4 cities:-
- ✓ San Diego
 - ✓ Nashville
 - ✓ Sacramento
 - ✓ Pittsburgh

Day Of the Month Analysis



Day Of the week Analysis



- ✓ On average most of the customers use the cabs as from Friday heading into the weekend.

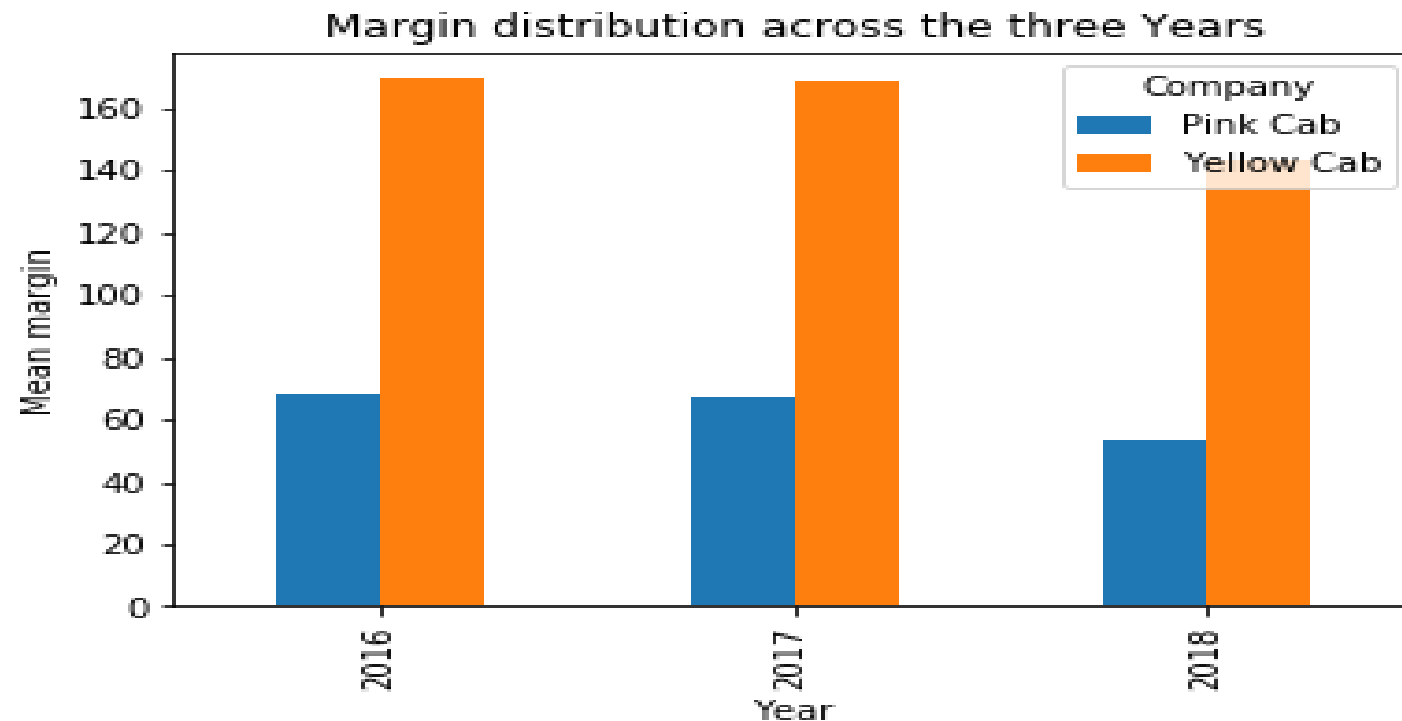
Most frequent Customers for both Cabs(Customer Retention)

Pink Cab(Customer_ID)	Frequency
8120	18
6159	17
7927	17
8595	17
8915	16

Yellow Cab(Customer_ID)	Frequency
1360	47
1803	47
494	47
636	46
126	45

- ✓ Yellow cab customers frequent the cabs more often showing that they have a high customer retainability. Also the most frequent Pink Customers are also Yellow Cab Customers.

Margin Analysis



- ✓ Yellow cab has had the most Profit compared to Pink Cab across the three years this very little fluctuations(Decrease in profit)

Recommendations and Conclusion

- I believe that Yellow cab gonna maintain its monopoly in the market for a longer period of time because:-
 1. Yellow cab has a very high customer retainability rate compared to Pink Cab.
 2. Through the three Years yellow cab has had the most customers and is not showing any signs of letting go off this trend.
 3. Yellow cab a high customer following in high population cities like New York.
 4. Most of the young aged customers prefer Yellow Cab. It is to be noted that the young aged customers are the biggest consumers in this cab business.
 5. Yellow cab seems to be making a higher profit margin compared to pink cabb

Brian Orina Bundi.

Thank You