Exploratory Data Analysis

G2M Cab Investment Case Study

15th March 2023

Agenda

Data Glacier
Your Deep Learning Partner

Executive Summary

- XYZ is a private firm in the US, looking to invest in a key player in the Cab Industry, due to remarkable growth, as part of their Go-to-Market (G2M) strategy.

- **Objective of Report:** Based on data from 2 major players in the industry, we must provide XYZ recommendations and actionable insights, backed by the data, to find the right company to make the investment.

Data Sets

- The following datasets were used:
    1. Cab_Data.csv – Transactions for Pink and Yellow companies (359392 Rows, 7 Columns)
    2. Customer_ID.csv – Customer's demographic details (49171 Rows, 4 Columns)
    3. Transaction_ID.csv – Mapping from transactions to customers (440098 Rows, 3 Columns)
    4. City.csv – Details of Cab Users in different cities in the US (20 Rows, 3 Columns)

- Time Period of Data: 01/01/2016 to 31/12/2018

Data Cleaning & Combining

Approach

- From the 4 datasets, extracted insights of dataset.

- Calculated the profit at the 'per ride' level, where profit = **price – cost**.

- Next, we mapped the cities and and populations to each record.

- Similarly, the customer's data was mapped based on the transaction ID.

- For the purpose of this evaluation, only the transactions that were done by the Pink or Yellow companies were considered.

Exploratory Data Analysis
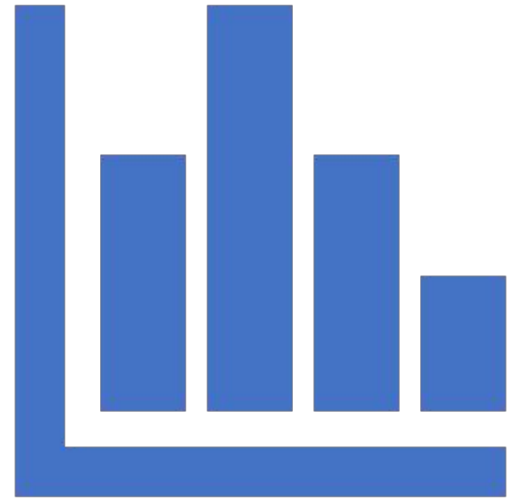
Average Profit Per Trip

- In the distribution created a Profit and Profit% column .
- Calculated the Profit per trip as Profit= Price Charged - Cost of Trip
- Then using above calculated the Profit% = Profit / KM Travelled
- The Box plot shows that the MEdian of yellow cab is higher than Pink Cab.
- Whereas the bar chart Showed the average profit and average profit per km travelled.
- Yellow cab have 2.5 times overall avg profit over pink cab.
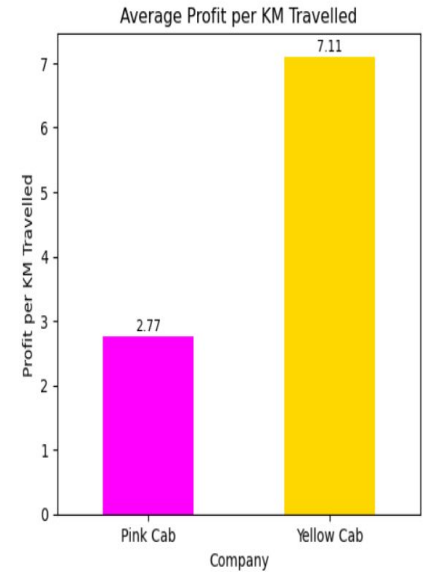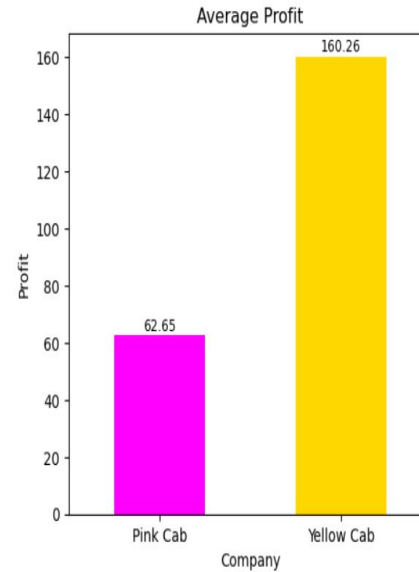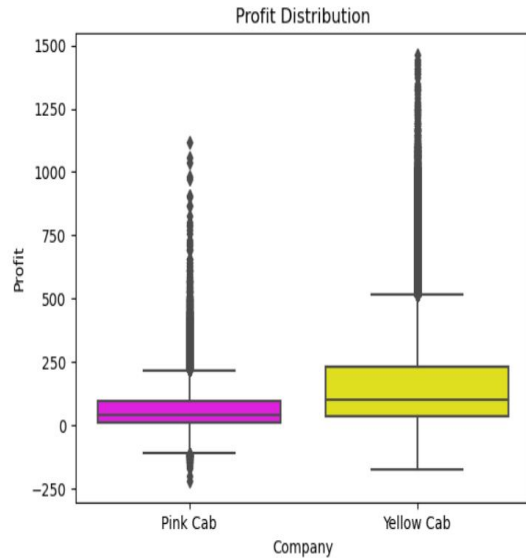
# Is there any Seasonality?

- Yes, there is a seasonal component in both companies,they are opposed to each other.
- While profit for Pink Cab improves between January and March, it decreases for Yellow cab during those months.
- However, when profits for Pink Cab decreases between May and June, it improves in Yellow cab.
- This seasonal pattern repeats every year.
-

Seasonal Distribution
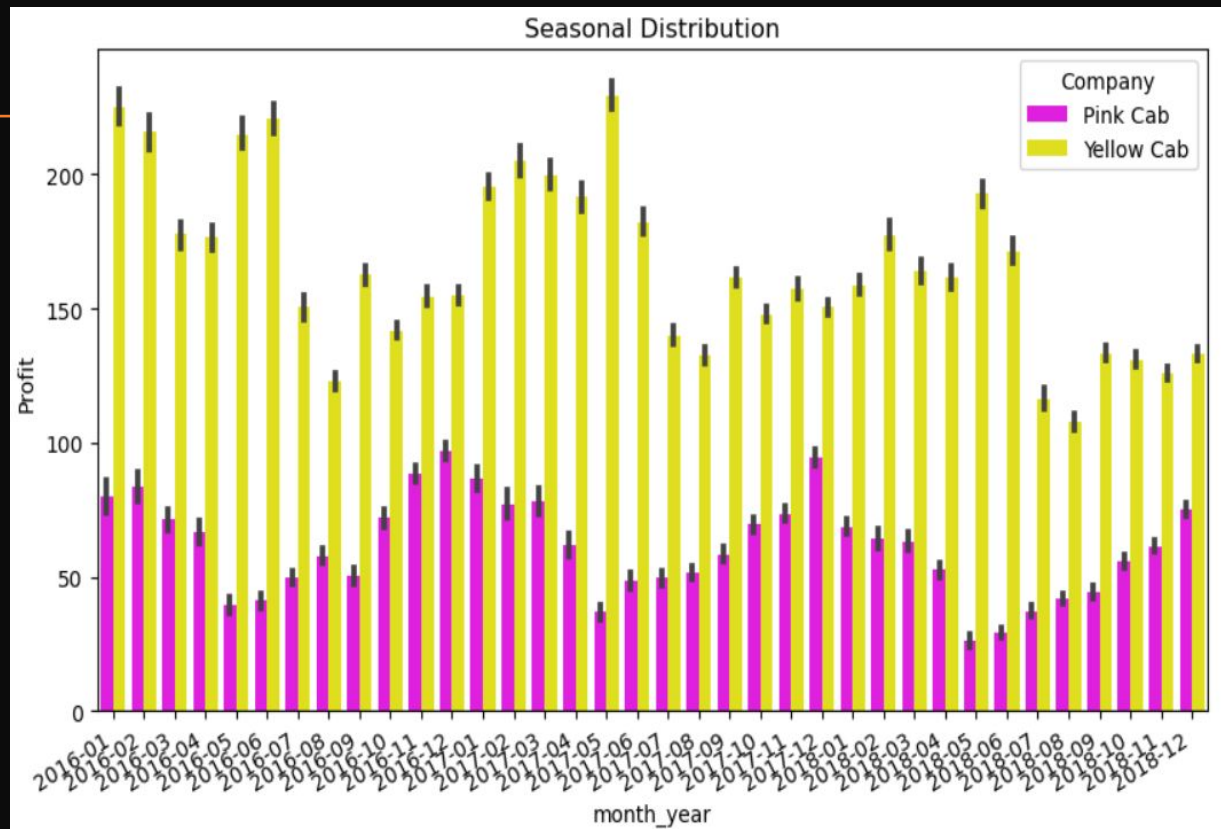
Users in each City

- As per the outcomes there are more customers from Yellow Cab of the major cities such as New York, Chicago, Washington.
- But the pink cab have highest only in LA.
- Overall,Yellow company has more than 50%of the customers in most cities.


Customer in each City
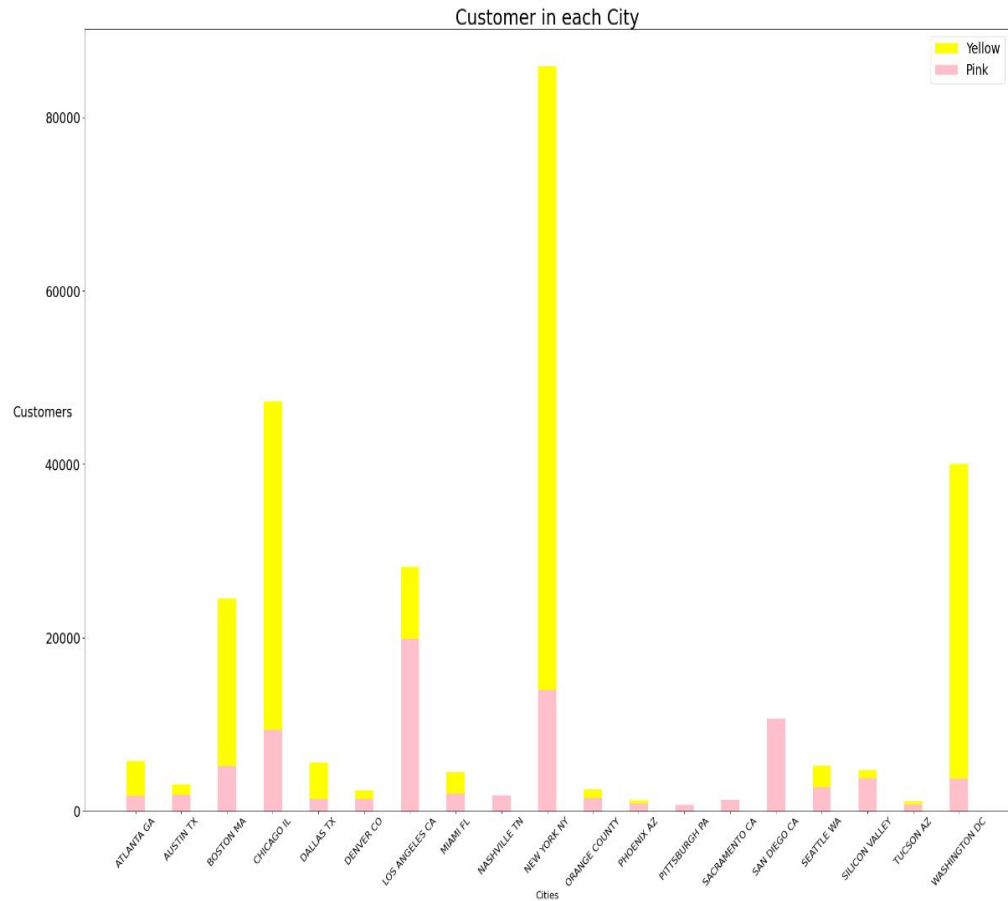
Distribution based on Gender

- Here, we see the customers based on the gender.

- It is observed that the Yellow company has more than 50%of the customers.

- Both men and women are the customers of both cab companies.

- Overall, Yelow cab have more customers of both genders.

Customer based on Gender
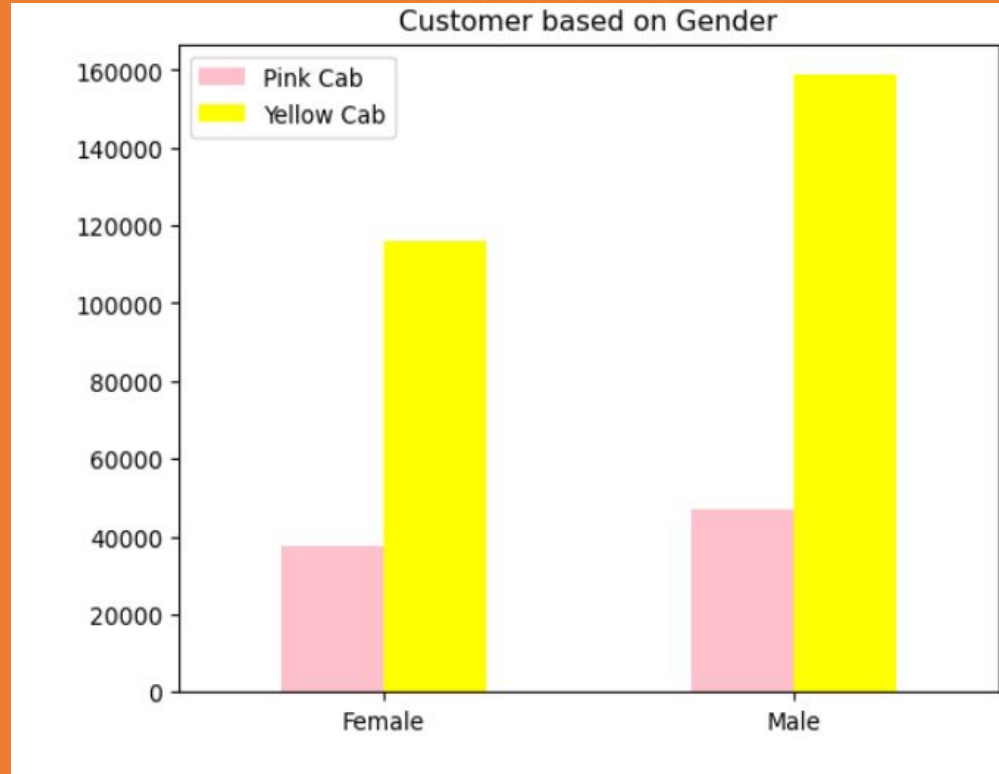
Customer Base by Age Group

- Finally, we see the customers based on age group.

- We see that the Yellow company has a very high average users compared to the Pink company, exceeding all age group.

- This is interesting as Adults are major customers of cab companies with yellow cab exceeding.

Customer Base by Age Group
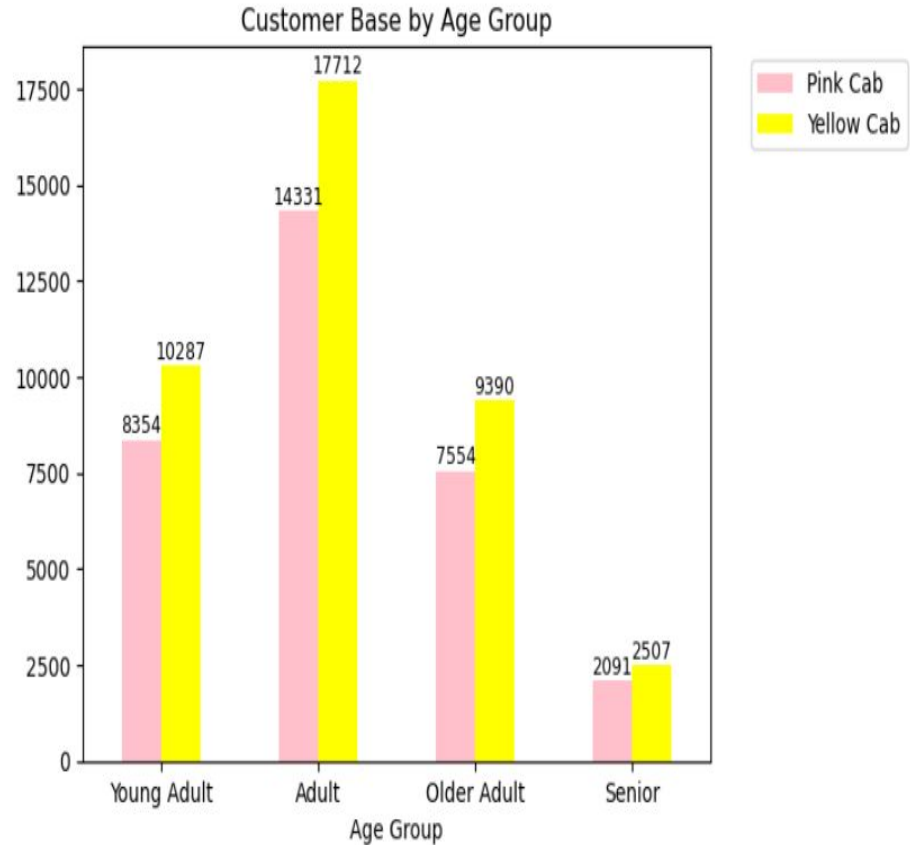
Exploratory Data Analysis (EDA) Summary

- After going through the plots, we see that the Yellow company has a higher cost per ride, higher price per ride and higher profit per ride. A higher price might lead us to the conclusion that it is the unpopular choice.

- This, however, isn't the case at all. In , majority of rides were done by the Yellow company.

- At the beginning of the analysis, the significantly higher average cost per ride of the Yellow company was a bit concerning. But now, this isn't a concern anymore. The higher cost per ride could be attributed to many factors such as Premium Cars used, better maintenance of the vehicles and higher wages for the drivers.

Recommendations

- **NOTE:** This recommendation is only based on profit and cost analysis. A final decision cannot be taken before looking at each of their Financial Statements to ensure that XYZ is investing in the right company.

We have evaluated both the cab companies on following points and found Yellow cab better than Pink cab:

- **Customer Reach** : Yellow cab has higher customer reach in 18 cities while Pink cab has higher customer reach in 4 cities. We have also observed that Yellow cab is doing good in covering other cab users as compared to Pink cab.

- **Age wise Reach :** Yellow cab has customer in all age group and it's been observed that it's very popular in 18-60 age group

- **Average Profit per KM:** Yellow cab's average profit per KM is almost three times the average profit per KM of the Pink cab.

- **Income wise Reach :**Both the cabs are very popular in high and medium income class but here also Yellow cab is performing better than Pink cab in offering their services to all the three income class group (low, medium and high)

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

Thank You