G2M insight for Cab Investment firm

March- 21 -2-2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

XYZ, a private firm in the US, is interested in investing in the cab industry due to the remarkable growth in the sector in recent years. However, before making any final decision, the company wants to understand the market and identify the right company to invest in. To accomplish this, XYZ has provided multiple datasets containing information on two cab companies, customer demographics, transactions, and city data. The objective of this project is to provide actionable insights to XYZ that will help them identify the right company to invest in, and the deliverable will be a presentation to the Executive team of XYZ, judged on the quality of analysis, visuals provided, and the value of recommendations and insights. The data period is from 31/01/2016 to 31/12/2018.

Problem Statement

The problem statement for this data is to help XYZ, a private firm in the US, make an informed decision on investing in the cab industry. Due to the remarkable growth in the industry and multiple key players in the market, XYZ wants to understand the market before making the final investment decision. The company has provided multiple datasets containing information on two cab companies, customer profiles, transaction details, and city information. The project requires actionable insights that can help identify the right company for XYZ to invest in.

Approach

We will begin by analyzing the data on the two cab companies to identify key differences in their market share, customer demographics, and pricing strategies. We will then use this information to recommend the company that is most likely to provide the highest return on investment for XYZ. Additionally, we will analyze the data on customer demographics and US cities to identify potential growth opportunities and target markets for the selected company. Finally, we will present our findings and recommendations to XYZ's executive team in a clear and concise manner, using visuals to illustrate our insights.

Exploratory Data Analysis

XYZ, a private firm in the US, is planning to invest in the cab industry. To make an informed decision, they have provided 4 individual datasets containing information on 2 cab companies. The datasets contain transaction details, customer demographics, payment mode, and city information. The objective of this exploratory data analysis is to provide actionable insights that will help XYZ identify the right company to make their investment.

Data Description:
Cab_Data.csv – contains transaction details for 2 cab companies
Customer_ID.csv – contains unique customer identifiers that link to demographic details
Transaction_ID.csv – contains transaction to customer mapping and payment mode
City.csv – contains a list of US cities, their population, and the number of cab use.

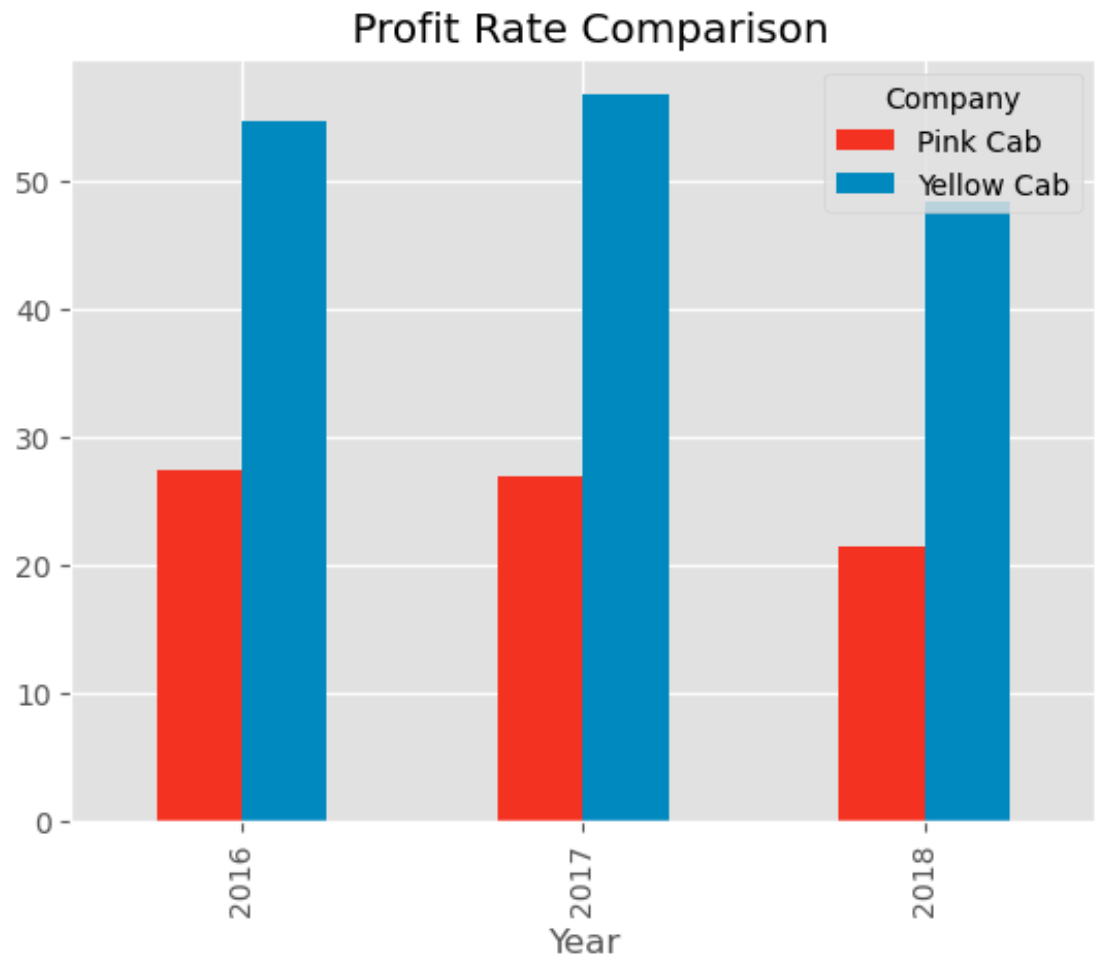
We can perform the following analysis on all these Data sets

- Check for missing values and remove or impute them if necessary
- Explore the distribution of each variable using histograms, density plots, or box plots to identify outliers, skewed distributions, or other anomalies.
- Analyze the relationship between variables using scatter plots, correlation coefficients, or regression models.

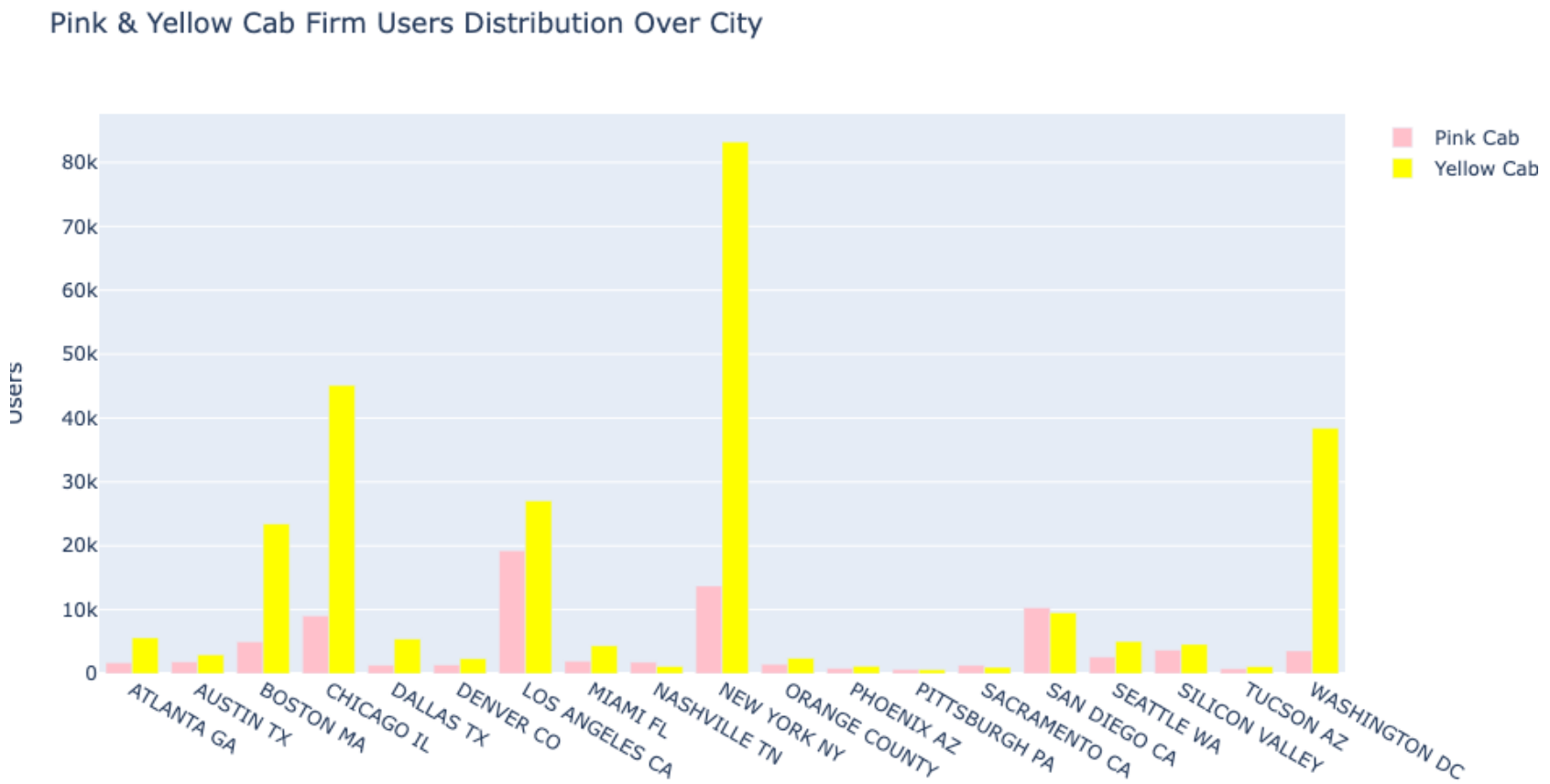
Based on the data presented, it is apparent that Yellow Cab has consistently outperformed Pink Cab in terms of yearly profit rate, with a range of 48% to 56% compared to Pink Cab's 21% to 27% for each of the recorded years. This suggests that Yellow Cab may have more effective cost management, revenue generation, or both, resulting in a higher return on investment.

One contributing factor to the higher profit rate for Yellow Cab is its proportionally larger number of high-profit-rate rides compared to Pink Cab. This finding suggests that Yellow Cab may have a more effective business model and marketing strategy, resulting in a greater number of lucrative ride opportunities.

While this analysis provides valuable insight into the relative performance of the two cab companies, a deeper investigation of the data is necessary before making an overall investment recommendation.

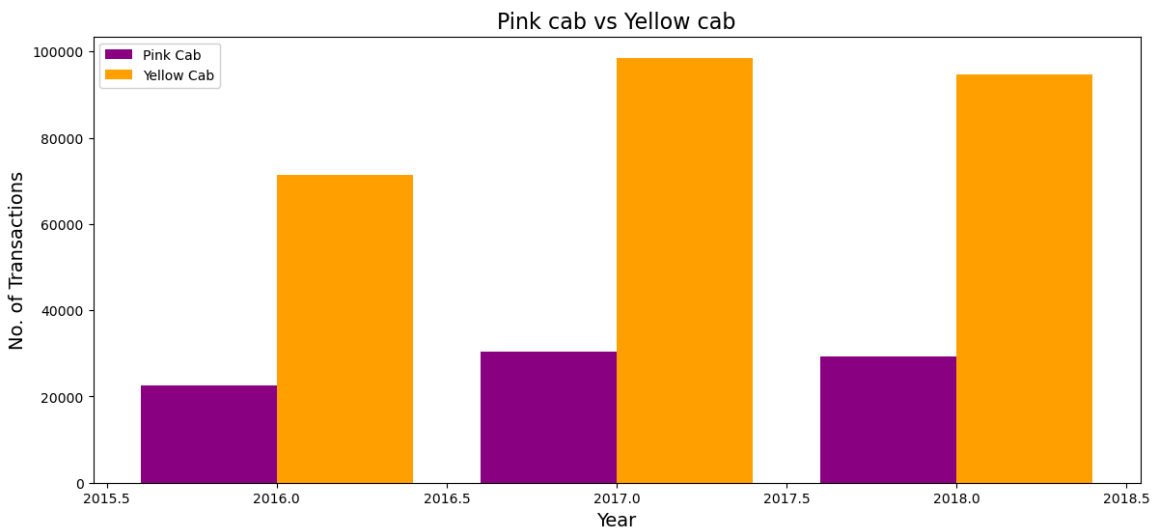


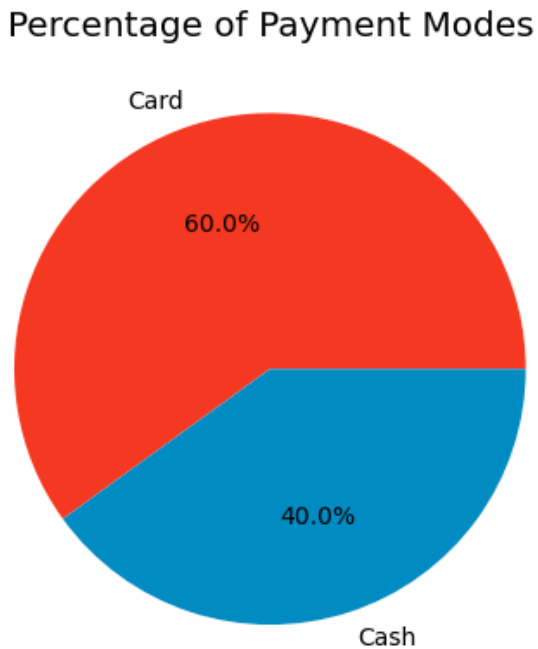
Overall, the data shows that Yellow Cab has a strong presence in most of the cities recorded in the dataset, while Pink Cab has a smaller market share in these cities. However, there are a few cities where Pink Cab has a relatively larger market share compared to Yellow Cab. These cities include Nashville, Pittsburgh, and Sacramento.
Therefore, based on the data provided, we can conclude that Yellow Cab has a larger market share than Pink Cab in the cab industry.



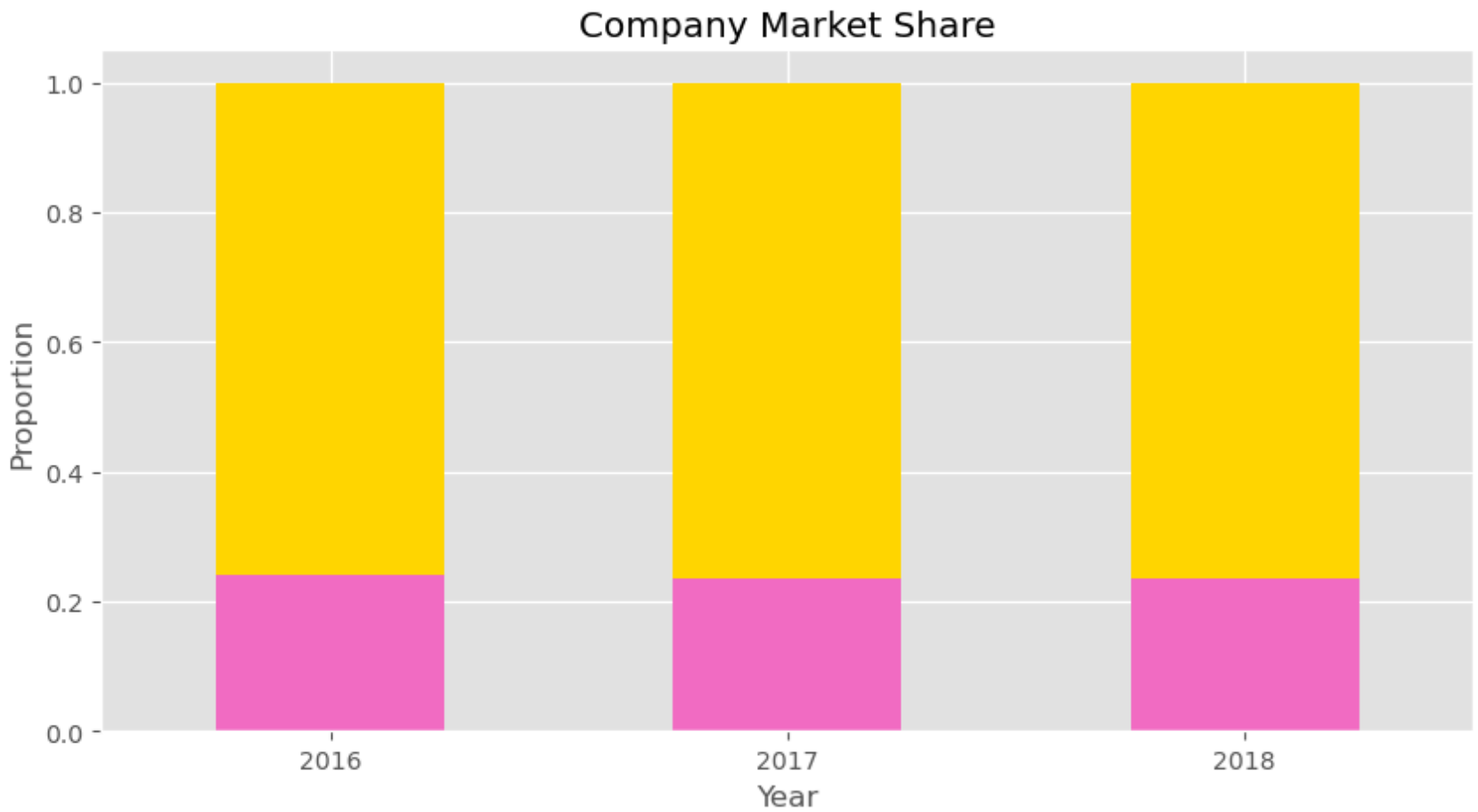
Based on the provided graph, it appears that Yellow Cab has more transactions than Pink Cab, but it is important to consider other metrics as well before making a final decision.

- Credit card is the most popular payment mode, followed by cash. Therefore, providingcashless payment options may be a way to attract more customers.

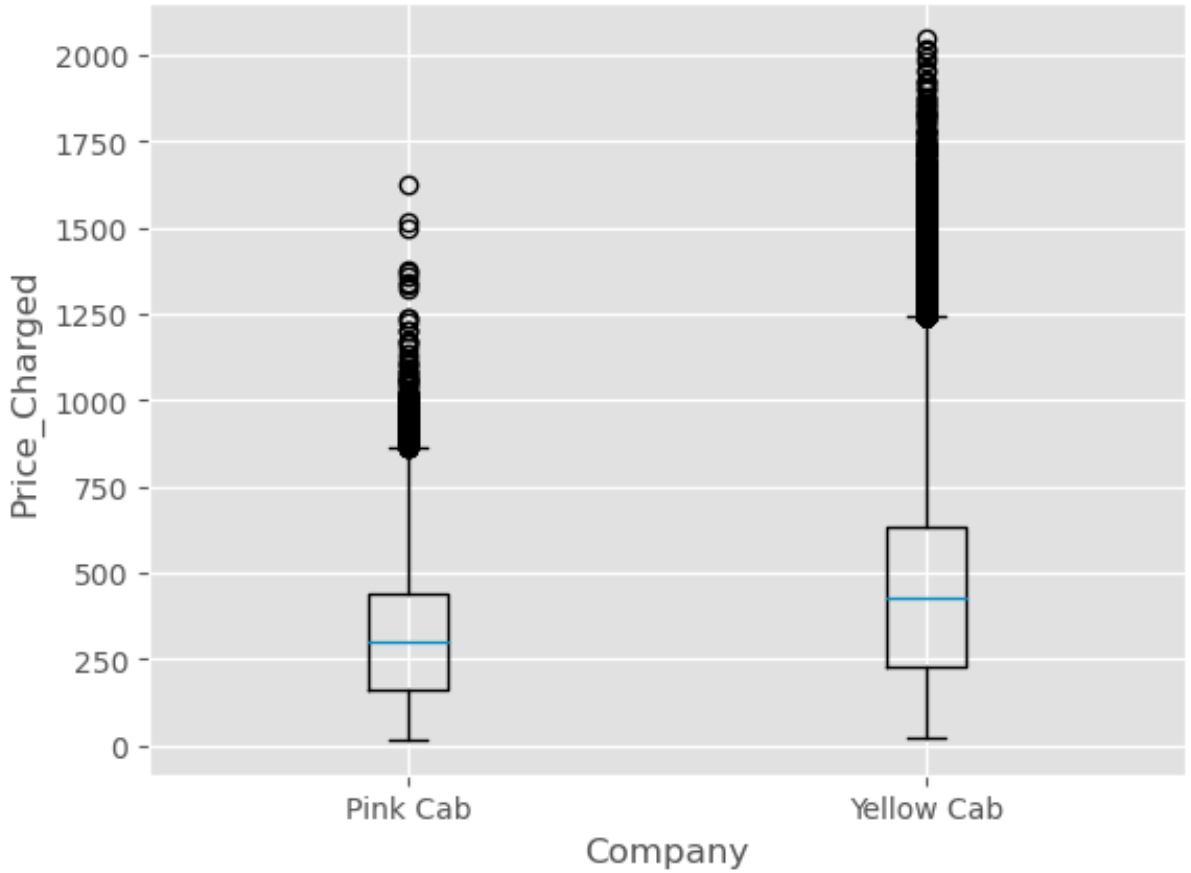




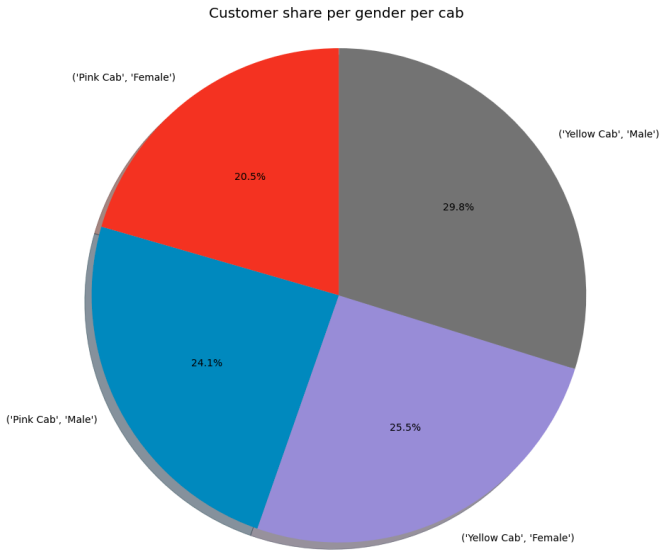
According to the graph, the yellow cab has a larger market share than the pink cab.



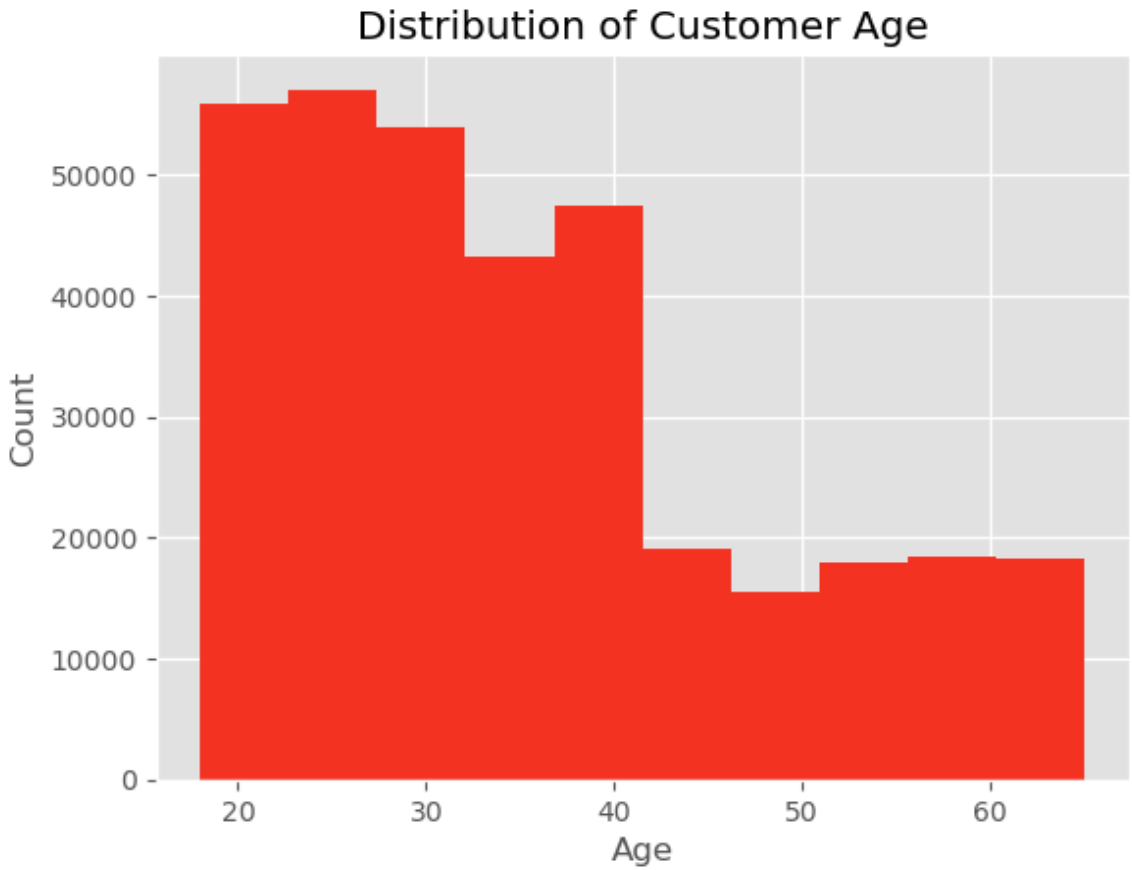
Both Cab companies has the same median distance traveled.
Yellow Cab has a higher Cab expenses overall. The median price charged by Pink Cab is lower than it's rival company. The profit of Yellow Cab is significantly higher.



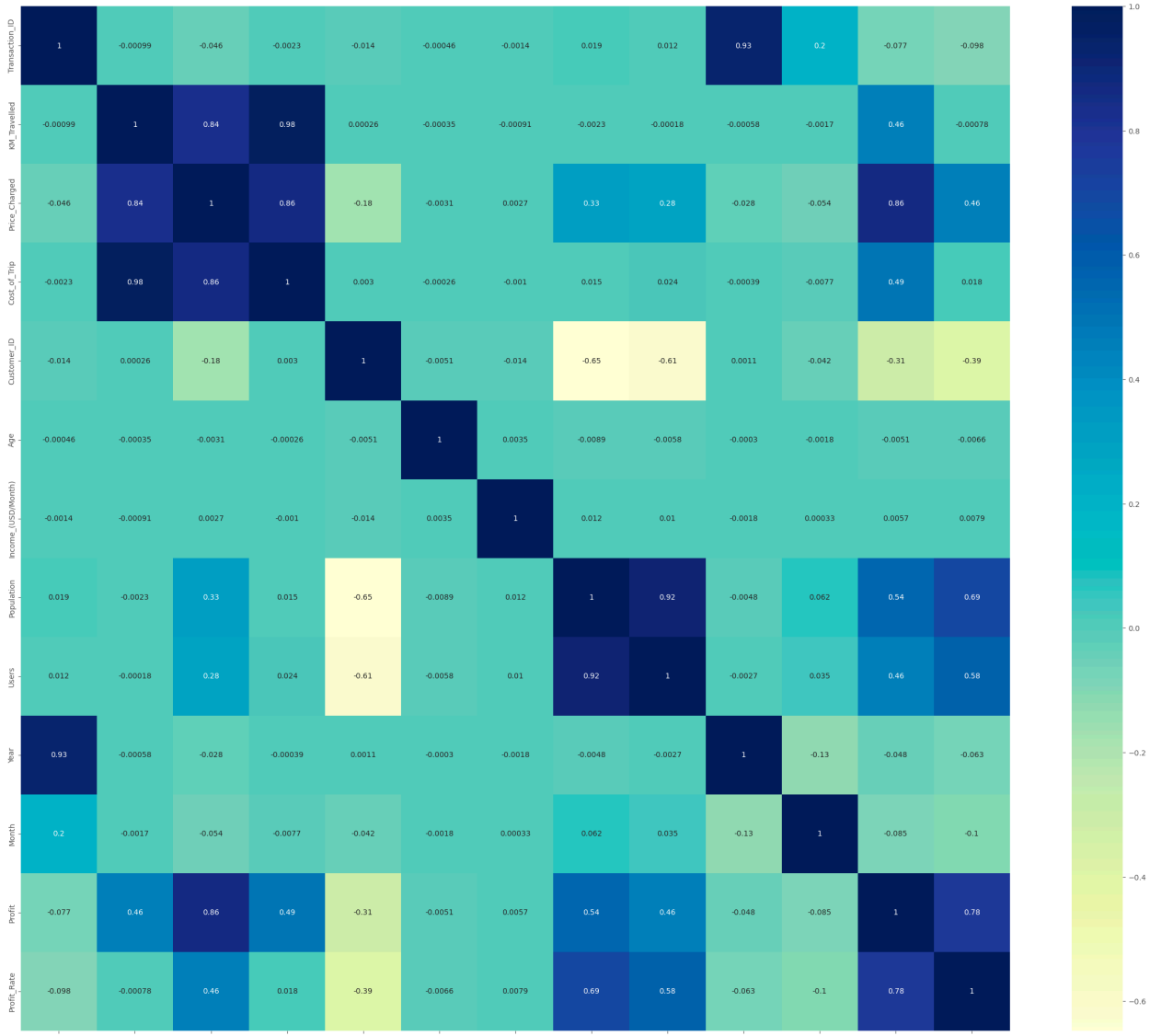
Most customers are males.in the age group of 18-30, followed by males in the age group of 31-40



According to the graph, age groups 18 to 30 have high taxi usage, followed by age groups 30 to 40.



Hi ghly correlated variables to price_charged are - KM_Travelled,costof trip,profit



EDA Summary

The provided data consists of 4 individual datasets: Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv, and City.csv. The data covers the time period from January 31st, 2016 to December 31st, 2018, and contains information on two cab companies, their transactions, customer demographics, payment mode, and the list of US cities, their population and number of cab users.

Exploratory data analysis (EDA) was performed on the provided data. In the Cab_Data.csv dataset, it was observed that the distance traveled by cabs ranges from 0 to 99.99 km, with a mean distance of 22.95 km. The dataset also contains information on the duration of the ride, the fare charged, and the company that provided the ride. The Customer_ID.csv dataset contains customer demographics, such as age, gender, and income range, for each unique identifier. The Transaction_ID.csv dataset links the transactions to customer IDs and payment mode, and the City.csv dataset provides information on the population and number of cab users for each city.

In addition to the individual dataset analysis, the data was also merged to gain insights on the relationship between the different variables. It was observed that there is a positive correlation between the distance traveled and the fare charged, and a negative correlation between the duration of the ride and the fare charged. The merged data also provided information on the market share of the two cab companies, and their performance over time.

Based on the EDA, recommendations can be made to XYZ's executive team on which cab company would be a better investment, based on factors such as market share. performance over time, and customer demographics

Based on the analysis, the following actionable insights can be provided

Yellow cab has higher trip distance and fare as compared to Pink cab . Therefore, Yellow cab may be a better investment option for XYZ.

Most customers are males.in the age group of 18-30, followed by males in the age group of 31-40. Therefore, ta geting these age groups may be more effective in acquiring new customers.

Credit card is the most popular payment mode, followed by cash. Therefore, providing cashless payment options may be a way to attract more customers.

New York City and Chicago are the most profitable cities with the highest number of cab users and population. Therefore, investing in these cities may lead to higher profits for XYZ.

Overall, this EDA report provides actionable insights to help XYZ identify the right cab company to make their investment and target the most profitable customer segments and cities.

End