Exploratory Data Analysis

<G2M insight for Cab Investment firm>

<21-May-2024>

Agenda

- Executive Summary
- Problem Statement
- Data Source
- EDA
- EDA Summary
- Recommendations
- Conclusion

Executive Summary

- 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.

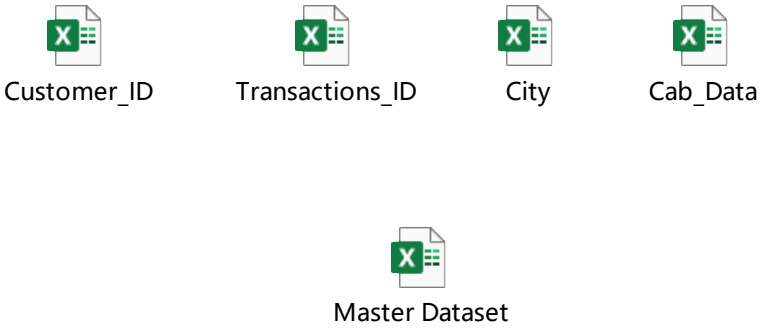
- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

Problem Statement

- Areas to investigate:
- Which company has maximum cab users at a particular time period?
- Does margin proportionally increase with increase in number of customers?
- What are the attributes of these customer segments?

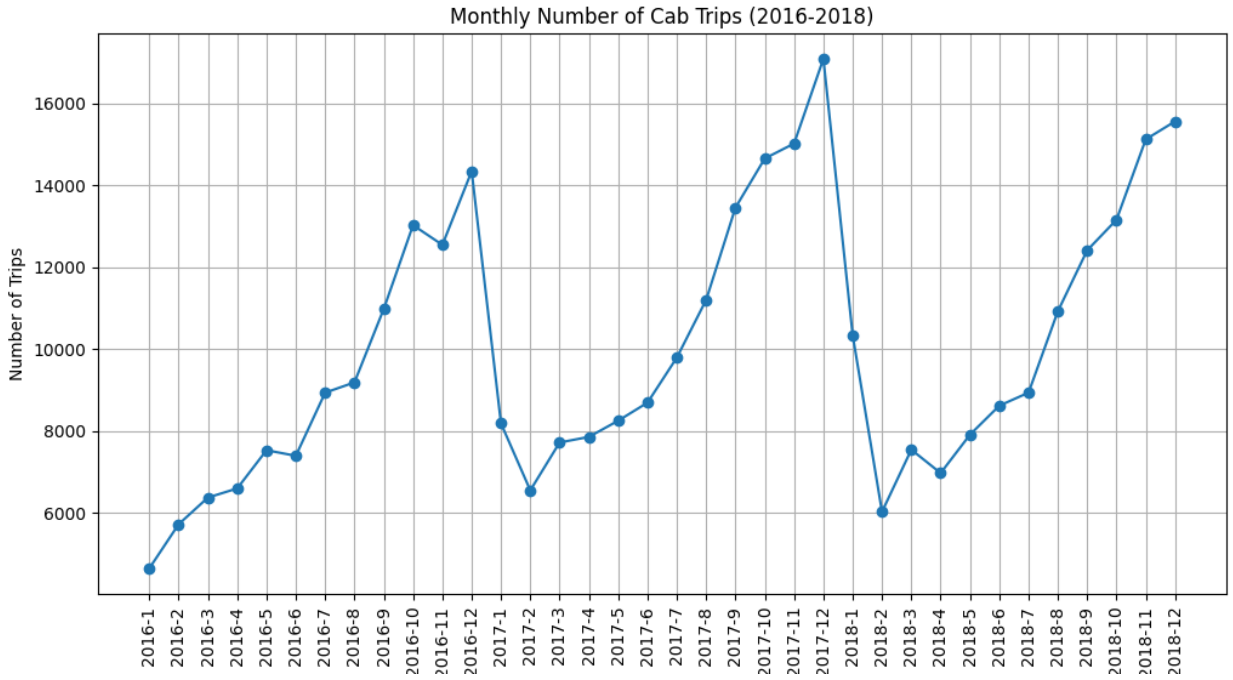
Data Source

- 4 individual data sets. Time period of data is from 31/01/2016 to 31/12/2018.
- Hypothesis: The number of Cab users are seen as users feature of city dataset.



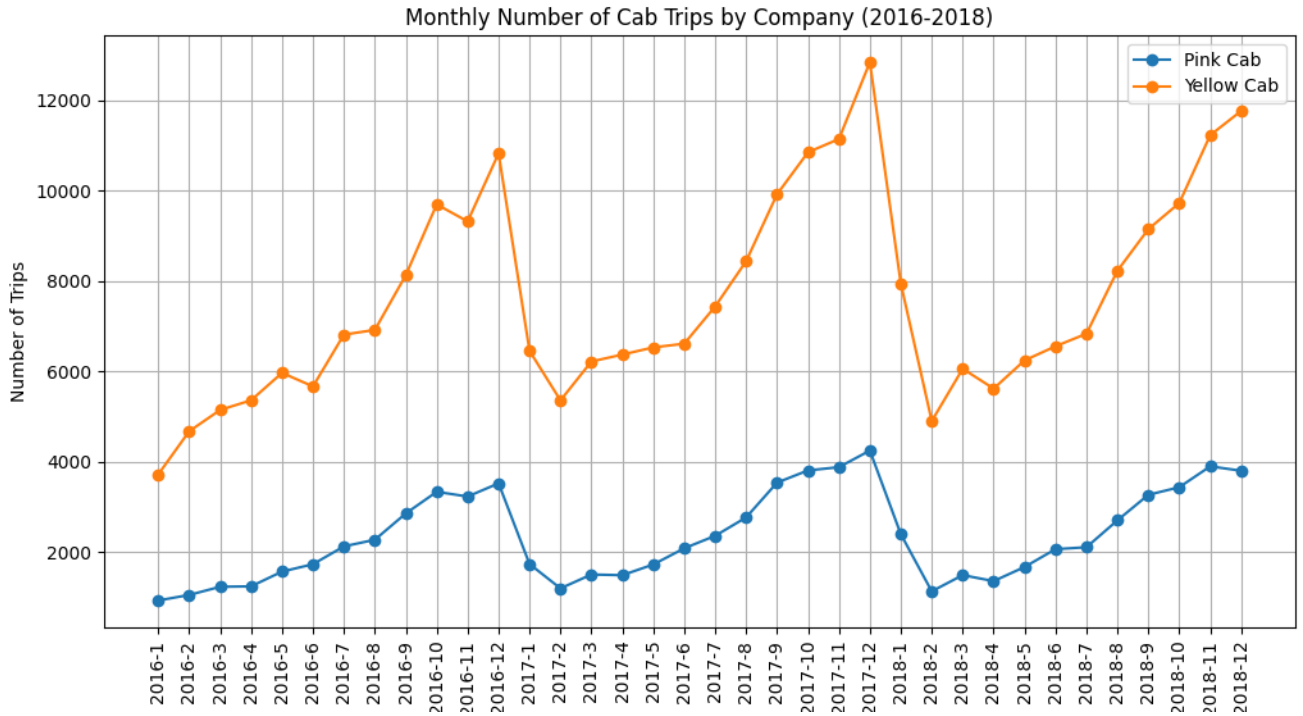
- Data Preprocessing:
- - Merging datasets to create a master dataset.
- - Converting data types for relevant fields.

EDA: Seasonality in Cab Usage



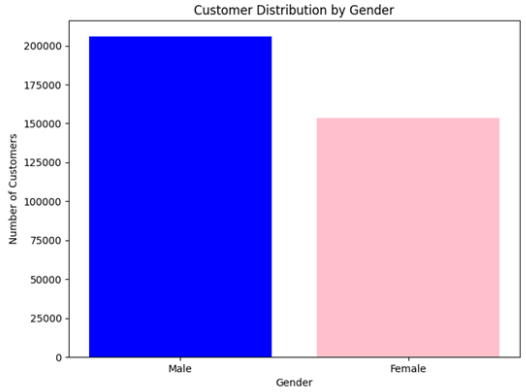
- Monthly number of cab trips from January 2016 to December 2018.
- The analysis shows fluctuations in cab usage over time.

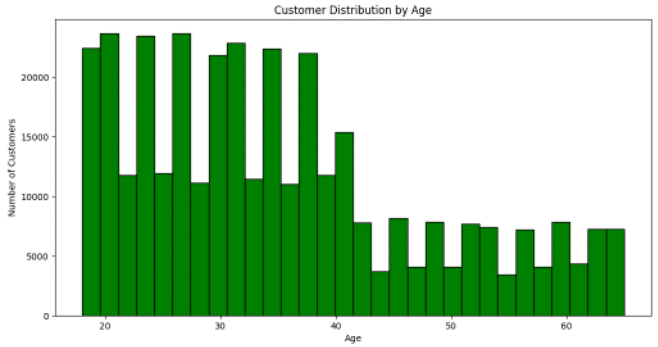
Company Performance Comparison

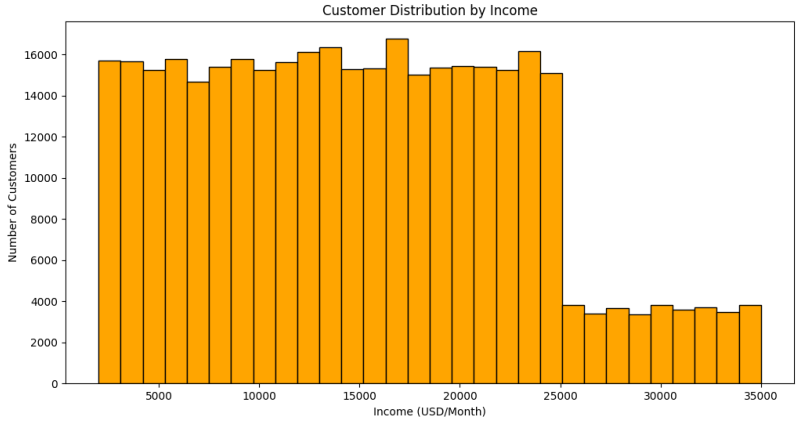


- Comparison of monthly number of trips for Pink Cab and Yellow Cab.
- Yellow Cab consistently had more users each month.

Customer Segmentation







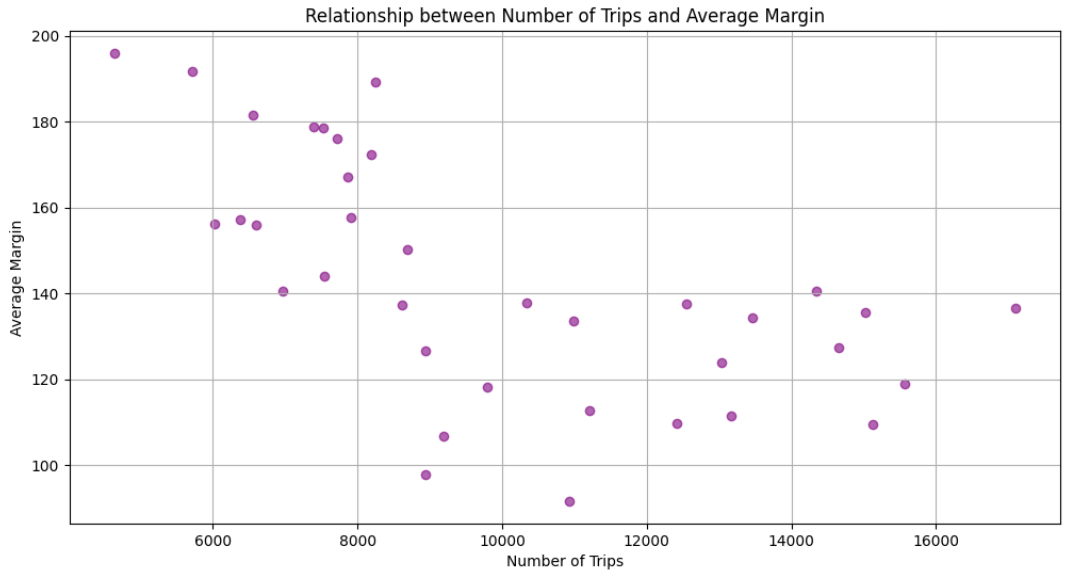
Customer Segmentation

- Gender Analysis:
- - Majority of customers are male.

- Age Analysis:
- - Most customers are between 25-45 years old.

- Income Analysis:
- - Wide income range, with most earning between $5,000 and $20,000 per month.

Revenue and Cost Analysis

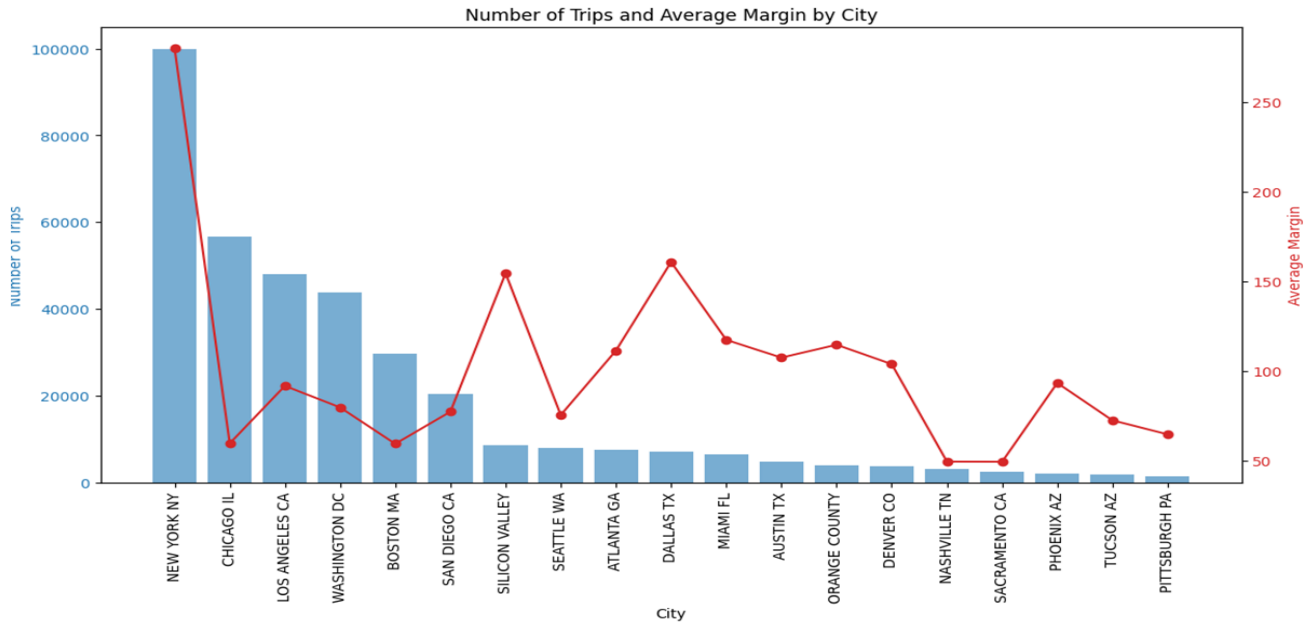


Analysis:

- Calculation of margin for each trip.
- Relationship between number of trips and average margin.
- Negative correlation between the number of trips and average margin, indicating higher volume does not necessarily result in higher margins.

Geographical Analysis

Geographical Analysis



Analysis:

- Number of trips and average margin by city.
- Identification of cities with highest cab usage and margins.

EDA Summary

- 1. Yellow Cab Performance:
- - Consistently had the maximum number of users each month from January 2016 to December 2018.
- - Higher volume of trips but with a negative correlation to margins.

- 2. Customer Insights:
- - Predominantly male customer base.
- - Concentration of customers aged 25-45.
- - Majority of customers have a monthly income between $5,000 and $20,000.

- 3. Geographical Insights:
- - Certain cities have higher cab usage and better margins.
- - Focus on cities with higher average margins for better profitability.

Recommendations for XYZ

1. Investment in Yellow Cab:
- Given its consistent performance in terms of user base and trip volume, Yellow Cab appears to be a better investment opportunity.
- Consider strategies to improve margins despite the high volume of trips.

- 2. Targeted Marketing:
- - Focus on male customers aged 25-45 with a monthly income of $5,000-$20,000.
- - Tailor marketing campaigns to attract this demographic.

- 3. Geographical Expansion:
- - Prioritize expansion in cities with higher cab usage and better margins.
- - Invest in improving services in high-margin cities to increase profitability.

Conclusion

- Summary:
- - Yellow Cab is the better investment opportunity based on user base and trip volume.
- - Targeted marketing and geographical expansion can further enhance profitability.

Thank You