Exploratory Data Analysis

G2M Insight Cab Investment Firm

21/03/24

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

EXECUTIVE SUMMARY

- XYZ, a private firm in the US, is considering investment in the cab industry. They aim to understand the market dynamics and identify the right company for investment based on thorough analysis.
- The Dataset Below are the list of datasets which are provided for the analysis:
  - Cab_Data.csv : this file includes details of transaction for
  - Customer_ID.csv : this is a mapping table that contains a unique identifier which links the customer's demographic details
  - Transaction_ID.csv : this is a mapping table that contains transaction to customer mapping and payment mode
  - City.csv : this file contains list of US cities, their population and number of cab users

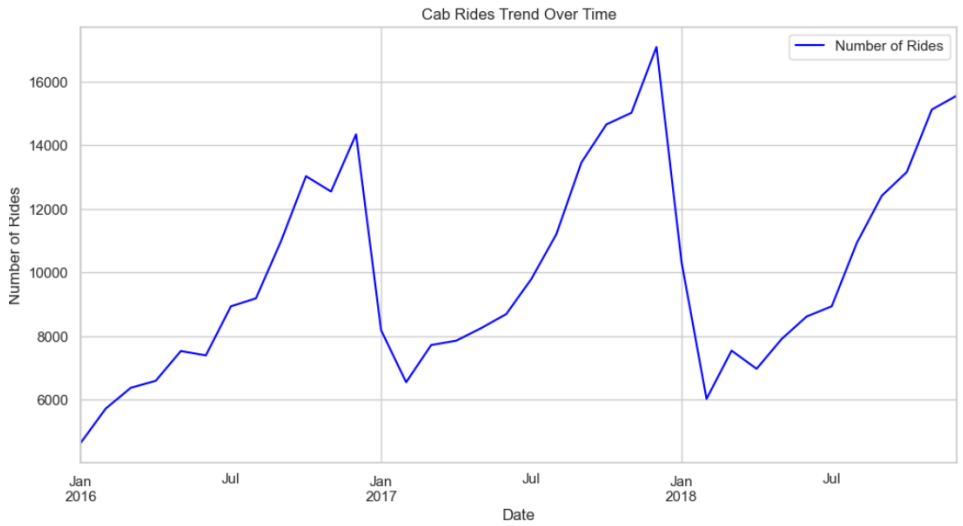
PROBLEM STATEMENT

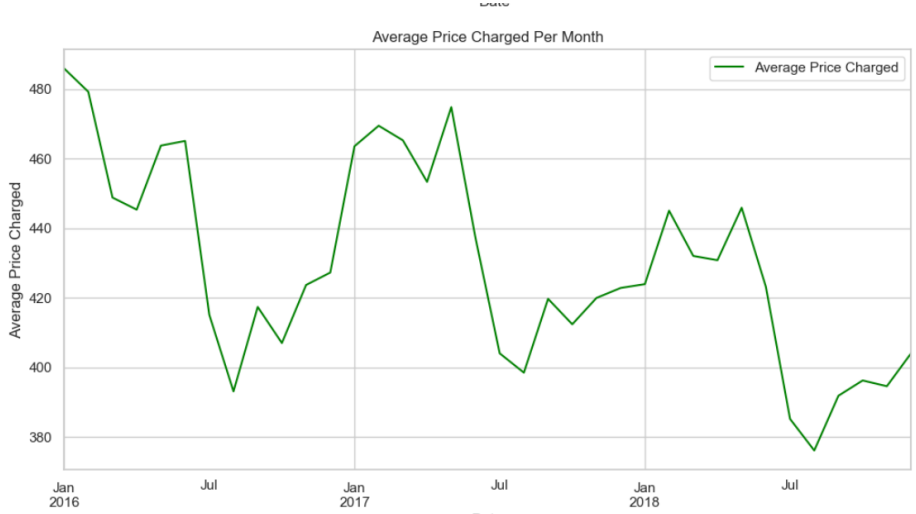
- XYZ, a private firm in the US, is facing a crucial decision regarding investment in the cab industry. With remarkable growth in the market and multiple key players, XYZ needs to carefully evaluate the available data and market trends to identify the most promising investment opportunity. The primary challenge is to navigate through complex datasets, understand market dynamics, and derive actionable insights to inform XYZ's investment decision-making process. XYZ seeks assistance in analyzing various aspects of the cab industry, including customer behavior, company performance, and market trends, to determine which company presents the best investment opportunity. The goal is to provide XYZ with comprehensive analysis and recommendations that enable them to make an informed investment decision aligned with their Go-to-Market strategy.

APPROACH

- We conducted Exploratory Data Analysis (EDA) on multiple datasets provided by XYZ, along with incorporating external data sources related to US holidays and weather. Our analysis focused on generating hypotheses, investigating trends, and deriving actionable insights to aid XYZ in their investment decision-making process.

EDA





	Company	Price Charged	Cost of Trip	KM Travelled	Number of Rides	\
0	Pink Cab	2.632825e+07	2.102092e+07	1911073.11	84711	
1	Yellow Cab	1.258539e+08	8.183351e+07	6199417.47	274681	

	Profit	Average Fare per KM
0	5.307328e+06	13.776685
1	4.402037e+07	20.300921

HYPOTHESIS 1

- **Objective:** To analyze whether there is a gender preference in cab usage and to determine if there are differences in gender preferences between Pink Cab and Yellow Cab.
- **Conclusion :** Yellow Cab appears to have a larger customer base overall, with both male and female customers preferring it over Pink Cab.
- Male customers seem to utilize both Pink Cab and Yellow Cab more than female customers.
- Further analysis may be required to understand the underlying factors influencing gender preferences in cab usage, such as pricing, availability, and service quality.

```
Number of Rides by Female Customers of pink cab: 37480
Number of Rides by Male Customers pink cab: 47231
Number of Rides by Female Customers of yellow cab: 116000
Number of Rides by Male Customers yellow cab: 158681
```

HYPOTHESIS 2

- **Objective:** To investigate whether there is a difference in the average fare charged per kilometer between Pink Cab and Yellow Cab, and to identify which company has a higher average fare per kilometer.
- **Conclusion:** Yellow Cab has a significantly higher average fare per kilometer compared to Pink Cab.This suggests that Yellow Cab may have different pricing strategies or cater to a different customer segment that is willing to pay higher fares for their services.

```
Pink Cab Average Fare per KM: 13.78
Yellow Cab Average Fare per KM: 20.30
```

HYPOTHESIS 3

- **Objective:** To assess the average income of cab users segmented by age group and determine if there are differences in income levels between age groups for Pink Cab and Yellow Cab.
- **Conclusion:** Surprisingly, despite Yellow Cab charging more per kilometer, both young and older customers with lower average income tend to prefer Yellow Cab over Pink Cab. Additionally, the analysis reveals that only individuals with higher average income in the older age group tend to prefer using cabs, whereas even those with lower incomes in the younger age group opt for cab services, regardless of the company.

```
Young Customers Average Income for pink cab: 15030.15
Older Customers Average Income for pink cab: 15098.31
Young Customers Average Income for yellow cab: 15026.32
Older Customers Average Income for yellow cab: 15071.85
```

HYPOTHESIS 4

- **Objective:** To examine the profitability of long and short rides for Pink Cab and Yellow Cab, and determine which type of ride generates more profit for each company.
- **Conclusion:** Both Pink Cab and Yellow Cab tend to generate higher profits from long rides compared to short rides. Particularly for Yellow Cab, long rides are significantly more profitable compared to Pink Cab, indicating potential advantages in pricing or operational efficiencies for longer distance trips.This suggests that both cab companies may benefit from strategies that encourage longer rides, such as offering discounts for longer distances or promoting services tailored for longer trips.

```
Long Rides Average Profit for pink cab: 92.32
Short Rides Average Profit for pink cab: 33.07
Long Rides Average Profit for yellow cab: 235.09
Short Rides Average Profit for yellow cab: 85.46
```

HYPOTHESIS 5

- **Objective:** To explore whether there is a difference in the average price charged by cab companies based on the income group of customers, and to determine if income level affects the price charged for cab services.
- **Conclusion:** The analysis suggests that there is minimal difference in the average price charged by both Pink Cab and Yellow Cab based on the income group of customers. Despite initial assumptions that locations of affluent areas might incur higher costs, the difference in price charged is not statistically significant. This indicates that income level may not directly influence the pricing strategy of cab companies, and other factors such as distance traveled, demand, and operational costs may play a more significant role in determining prices.

```
High Income Average Price Charged for yellow cab: 459.82
Low Income Average Price Charged for yellow cab: 456.54
High Income Average Price Charged for Pink cab: 311.34
Low Income Average Price Charged for Pink cab: 310.26
```

RECOMMENDATION

- Investment Opportunity: Invest in Yellow Cab.
- Market Presence: Yellow Cab has a larger customer base and higher profitability compared to Pink Cab.
- Profitability: Yellow Cab's higher average fare per kilometer and profitability from both long and short rides indicate potential revenue advantages.
- Customer Segmentation: Despite higher fares, lower-income groups show a preference for Yellow Cab, indicating strong brand loyalty or perceived value.
- Next Steps: Conduct further due diligence on Yellow Cab's financial stability, operational efficiency, and growth prospects. Explore market dynamics, competition, and regulatory factors affecting the cab industry. Consider opportunities for strategic partnerships or initiatives to enhance market positioning. Implement mechanisms for continuous monitoring and evaluation of investment performance and market trends.

Thank You