Exploratory Data Analysis

G2M Insight for Cab Investment

May 21, 2025

Agenda

Executive Summary

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

Executive Summary

Client:

XYZ is a private 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 and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Objective:

To determine which cab company is worth investing into for the best possible profits.

Analysis Pieces:

Cab Company popularity, relationships between Unique customers & Margin, Age, Gender & Income Ranges.

Data Exploration:

There were total 359,352 rows in the Cab dataset and a total of 15 features in all the 4 datasets combined and 3 features were derived. The data ranges from 2016 to 2018.

Assumptions:

1. Data analysis was performed for both the companies under the assumption of external noise beyond the data that was provided.
2. Data is assumed to be from 2016 to 2018.
3. The datasets were drawn from random selection.
4. Payment by cash or card were the only payment methods considered.
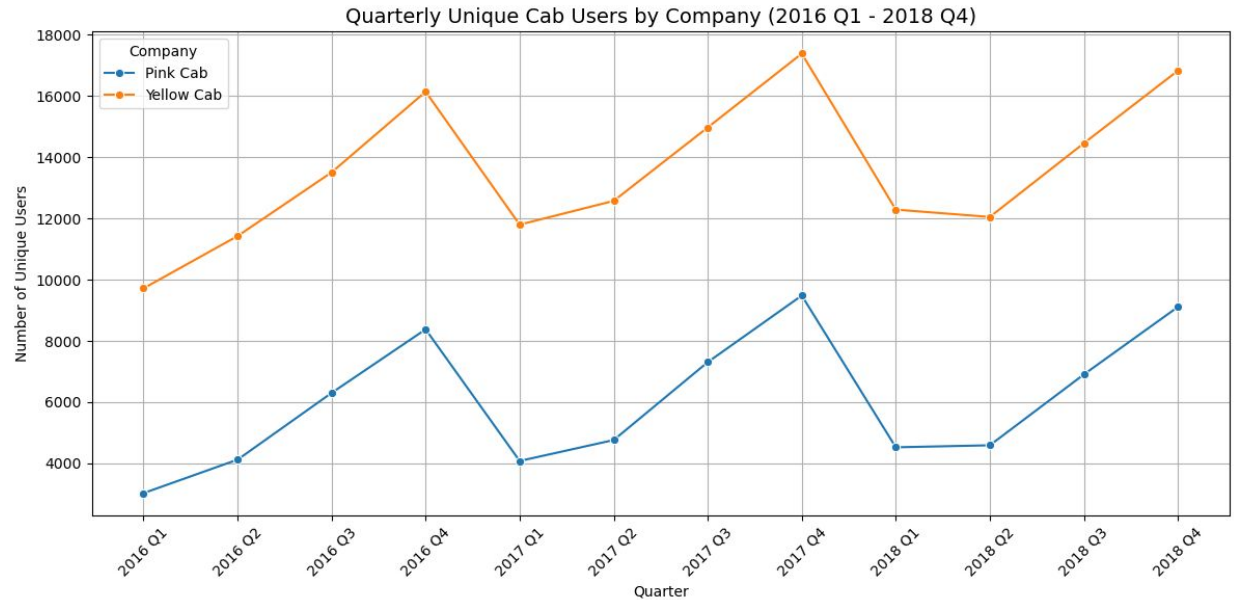
Cab Company Popularity Analysis

Motivation:

First, we needed to figure out which cab company (Yellow or Pink) was more popular among users in terms of numbers during each quarter beginning from from January of 2016 to December of 2018.
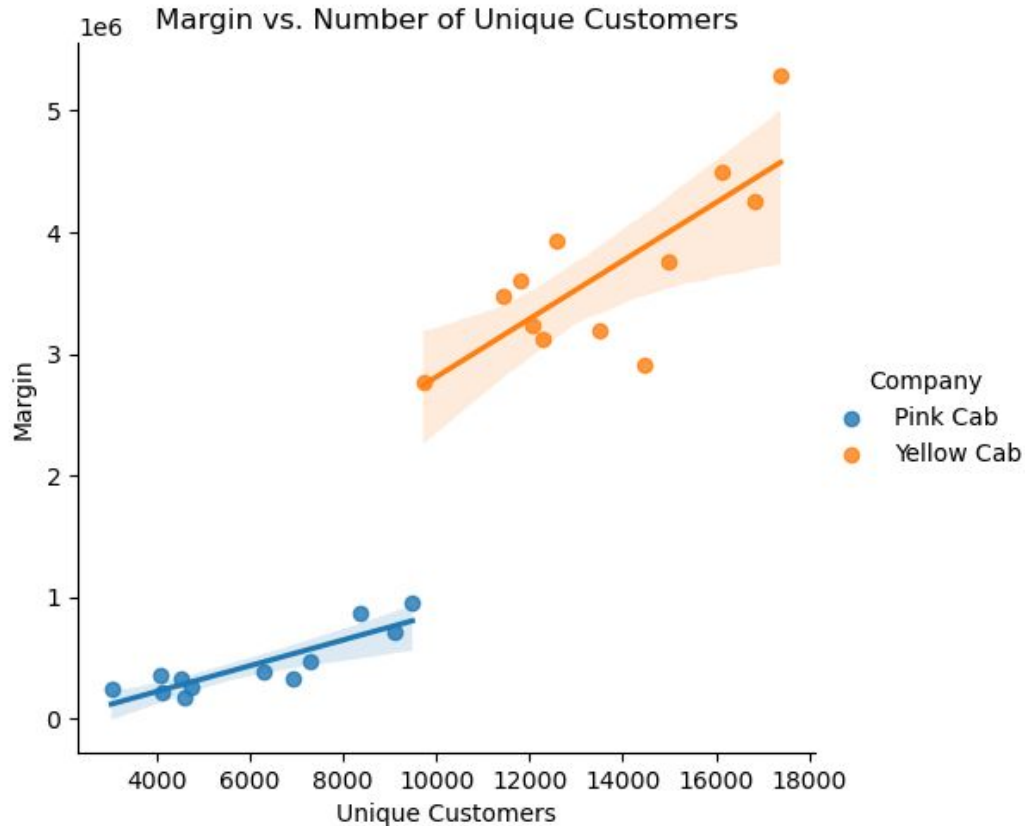
Test:

The first test showed the maximum unique number of users for each company. The 2nd and 3rd tests were to show by how much frequency one cab company is more profitable than the other.

Cab Company Popularity Analysis Results

	Company	Month	Unique Users
12	Yellow Cab	2016Q1	9718
13	Yellow Cab	2016Q2	11426
14	Yellow Cab	2016Q3	13508
15	Yellow Cab	2016Q4	16144
16	Yellow Cab	2017Q1	11800
17	Yellow Cab	2017Q2	12576
18	Yellow Cab	2017Q3	14975
19	Yellow Cab	2017Q4	17388
20	Yellow Cab	2018Q1	12291
21	Yellow Cab	2018Q2	12051
22	Yellow Cab	2018Q3	14459
23	Yellow Cab	2018Q4	16830

Quarterly Unique Cab Users by Company (2016 Q1 - 2018 Q4)

Cab Company Popularity Analysis Results

Margin vs. Number of Unique Customers

Conclusion:

We can see the the *Yellow Cab* company is outperforming in **Margin** (profits calculated) and **no. of unique users** by very large amounts.

Insight: Relationship Between Unique Customers and Margin

Furthermore, if we check the correlation between the no. of unique customers and margin, we see that the correlation coefficient between Unique Customers and Margin is 0.938, which indicates a very strong positive correlation. This suggests that as the number of customers grows, the profit margin also increases consistently.

So this is the Implication:

- Customer growth directly contributes to revenue growth, reinforcing the importance of expanding the user base
- Yellow Cab, with more customers, generates significantly higher margins, indicating economies of scale or better pricing efficiency.

```
margin_customers[['Unique Customers', 'Margin']].corr()
```

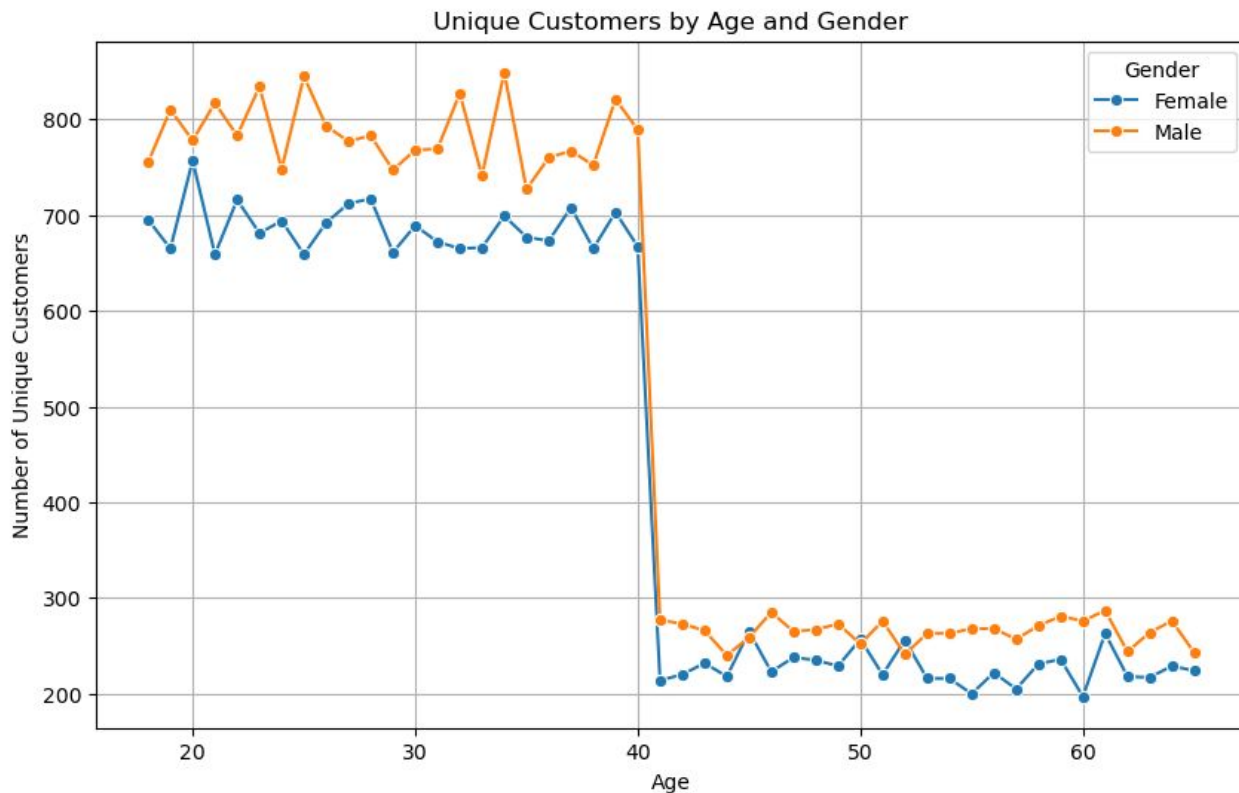|	Unique Customers	Margin
Unique Customers	1.000000	0.937945
Margin	0.937945	1.000000

Motivation:

In order to create the best marketing strategy for the future, one needs to figure what kind of demographic is being catered to to take the appropriate business decisions to increase revenue.

Test:

1. The 1st graph shows unique customers on the basis of age and gender.
2. For the 2nd graph, the incomes were binned into bin ranges of about 5k USD each to determine the income range of the customers and their preferred cab company.
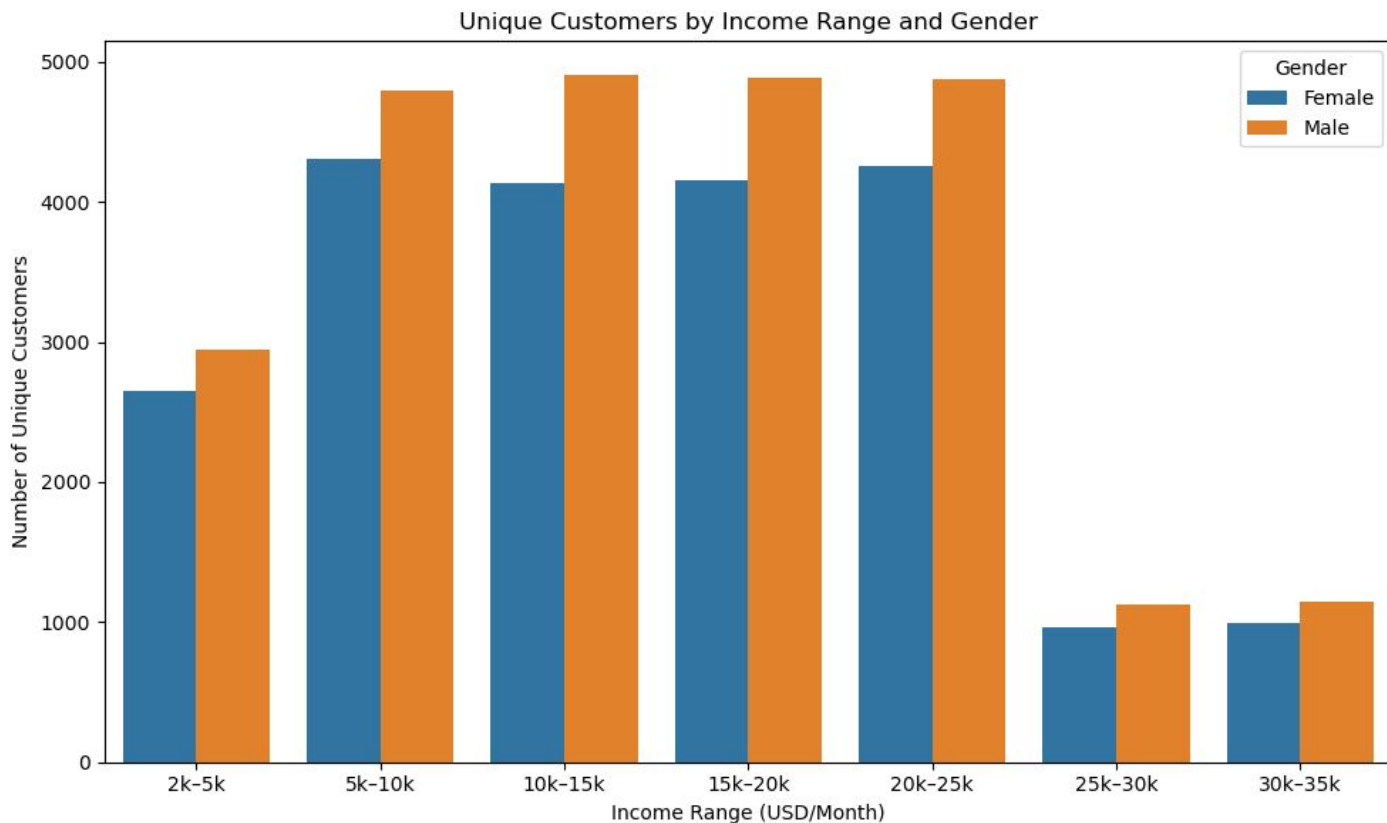3. The 3rd graph showed income ranges based on the 2 genders.

Unique users based on Age & Gender & Income range

Unique users based on Income ranges & their cab preferences
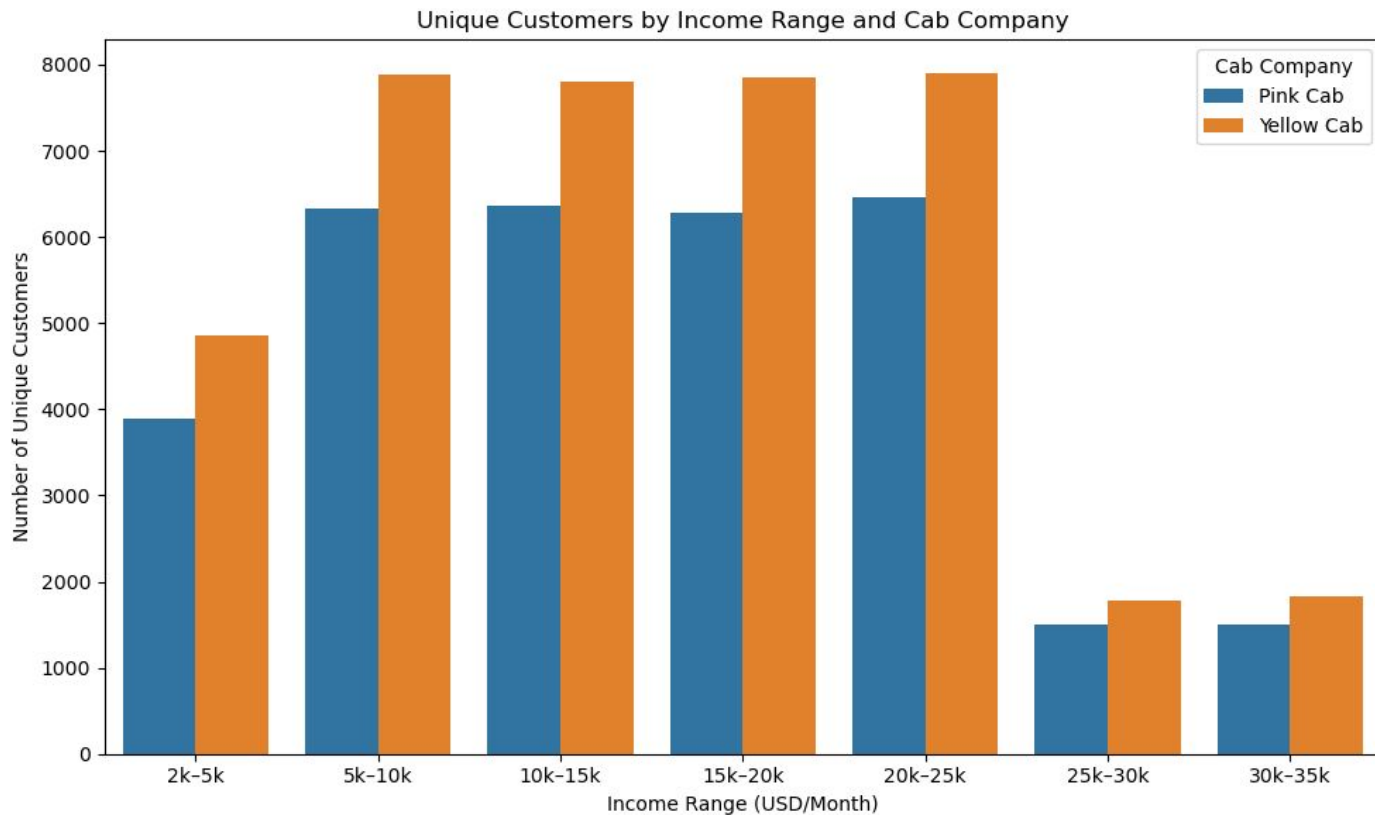
Unique users based on Income ranges & gender

Unique Customers by Income Range and Gender

Unique Customers' Analysis Results

One can derive the following **_key insights_** from the aforementioned graphs:

1. **Yellow Cab consistently has more unique users** per quarter compared to Pink Cab, showing stronger market presence over time.
2. There is a strong positive correlation (0.938) between the number of unique customers and margin for both companies, indicating that **higher customer volume leads to higher profits.**
3. Most cab users fall in the **15–40 age group**, with usage declining sharply after 40.
4. Men use cab services more frequently than women across all income ranges and ages.
5. The **dominant customer base earns between 2k- 25k USD/month**, with fewer high-income users (less than 25k USD/month).
6. **Yellow Cab is the preferred option** across all income brackets.

My Recommendations based on my detailed Analysis

1. Investment in Yellow Cab should be prioritized due to its larger and more consistent user base and higher profit margins.
2. The marketing strategies should be focused on the 15–40 age segment and mid-income groups (5k–25k USD/month), which make up the bulk of users.
3. It needs to be explored how Yellow Cab maintains broader appeal across demographics and apply similar tactics to improve Pink Cab's positioning.
4. The strong customer-margin relationship should be leveraged by investing in customer acquisition to drive profitability.

Thank You