



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

## G2M Insight for Cab Investment Firm

**April 21<sup>st</sup>, 2025**

# Agenda

Executive Summary

EDA

EDA Summary

Recommendations

# Executive Summary

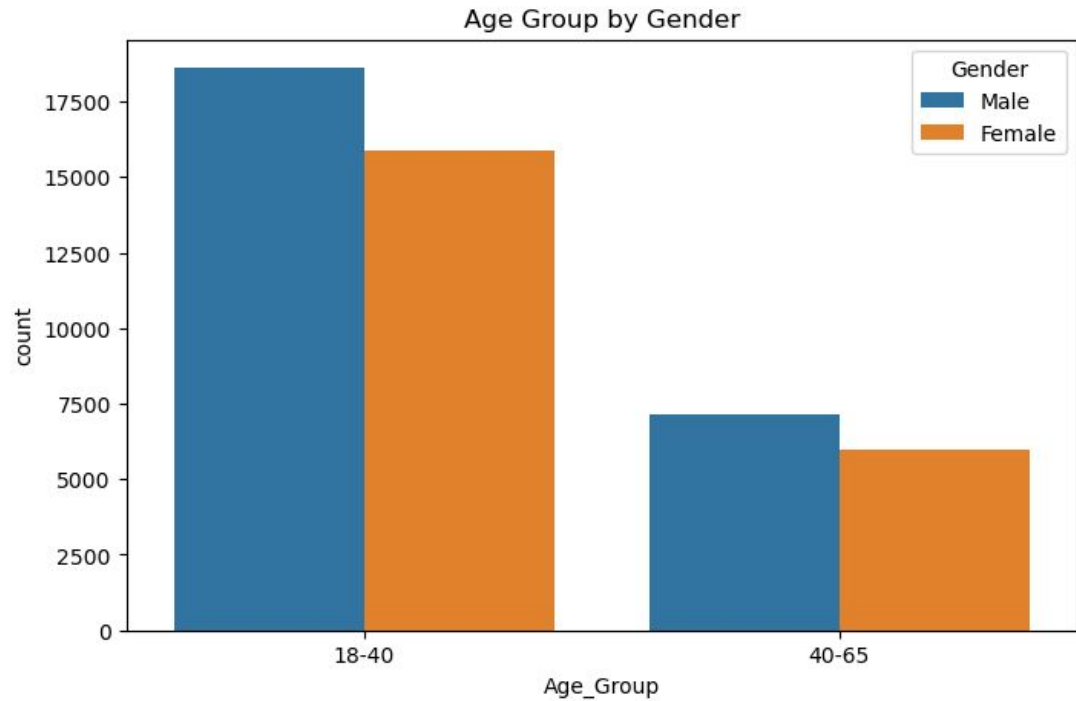
- **Objective:** Assist XYZ, a private US investment firm, in identifying the most promising cab company for investment based on market, customer, and performance insights.
- **Scope:** Conduct exploratory data analysis (EDA) on datasets representing transactions, customers, city demographics, and payment methods across two cab companies from 2016 to 2018.
- **Key Focus Areas:**
  - Customer demographics, behavior, and geographic distribution.
  - Revenue trends and growth patterns for each cab company.
  - Market penetration and cab user density by city.
- **Methodology:** Data wrangling, descriptive statistics, and visualization techniques were used to uncover trends, patterns, and outliers.
- **Outcome:** Actionable insights and strategic recommendations are provided to guide XYZ in making a data-driven investment decision.

# EDA

- 23 Features (including 6 derived features)
- Timeframe: 2016-1-02 to 2018-12-31
- Total trips: 359392
- Number of cities: 20
- Number of unique customers: 49171
- Datasets:
  - Cab\_Data.csv
  - City.csv
  - Customer\_ID.csv
  - Transaction\_ID.csv

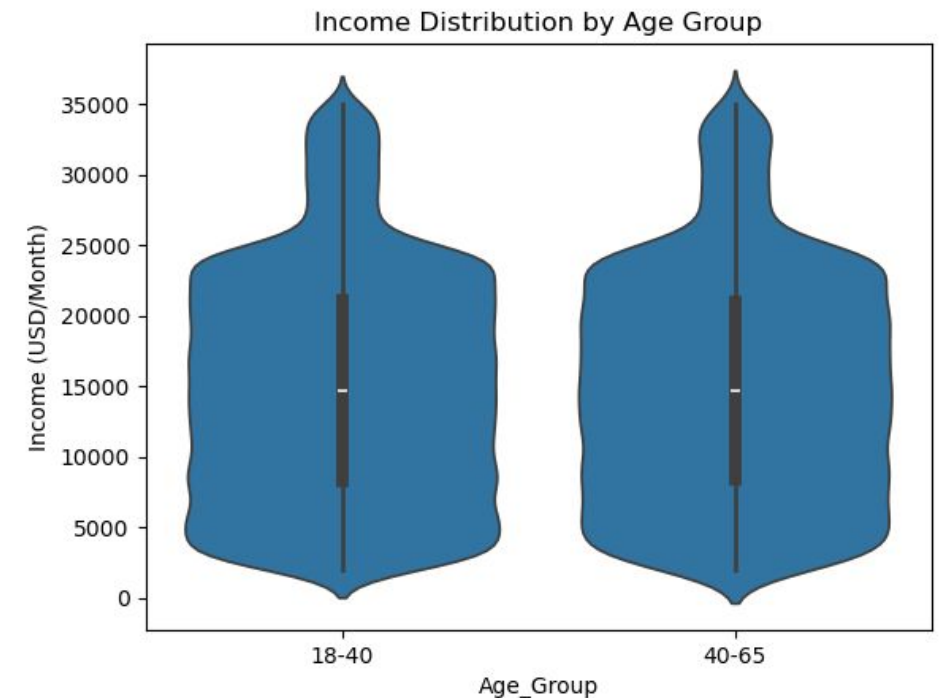
Check	Status	Notes
Missing values	✓ None	No missing value found
Duplicate entries	✓ None	No duplicates in primary keys
Date format consistency	✓ Clean	All dates parsed successfully
Mismatched keys (joins)	✓ None	All transaction have matching customer ID

# Customers Profiles

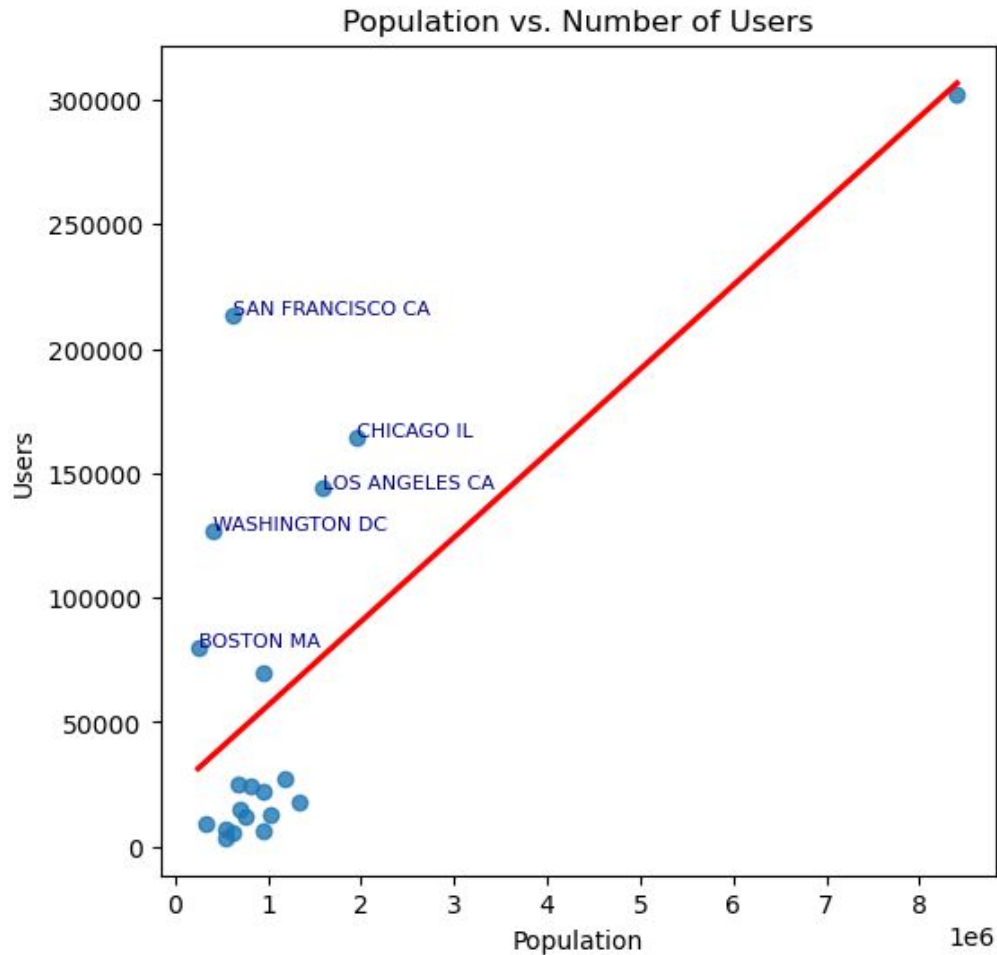


- More male users than female users for both age groups.

- No apparent income distribution difference between two age groups.

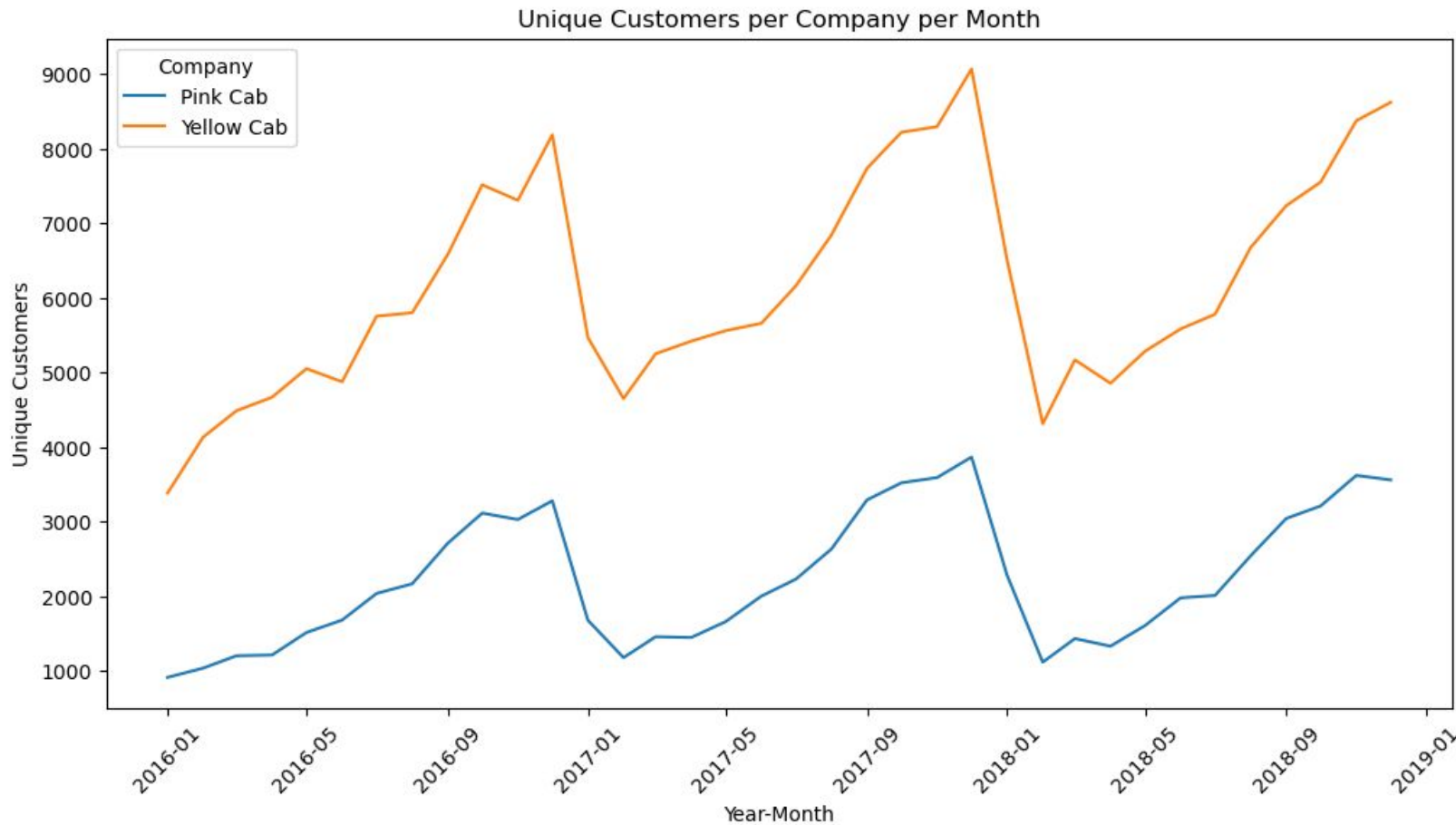


# Customers Distribution



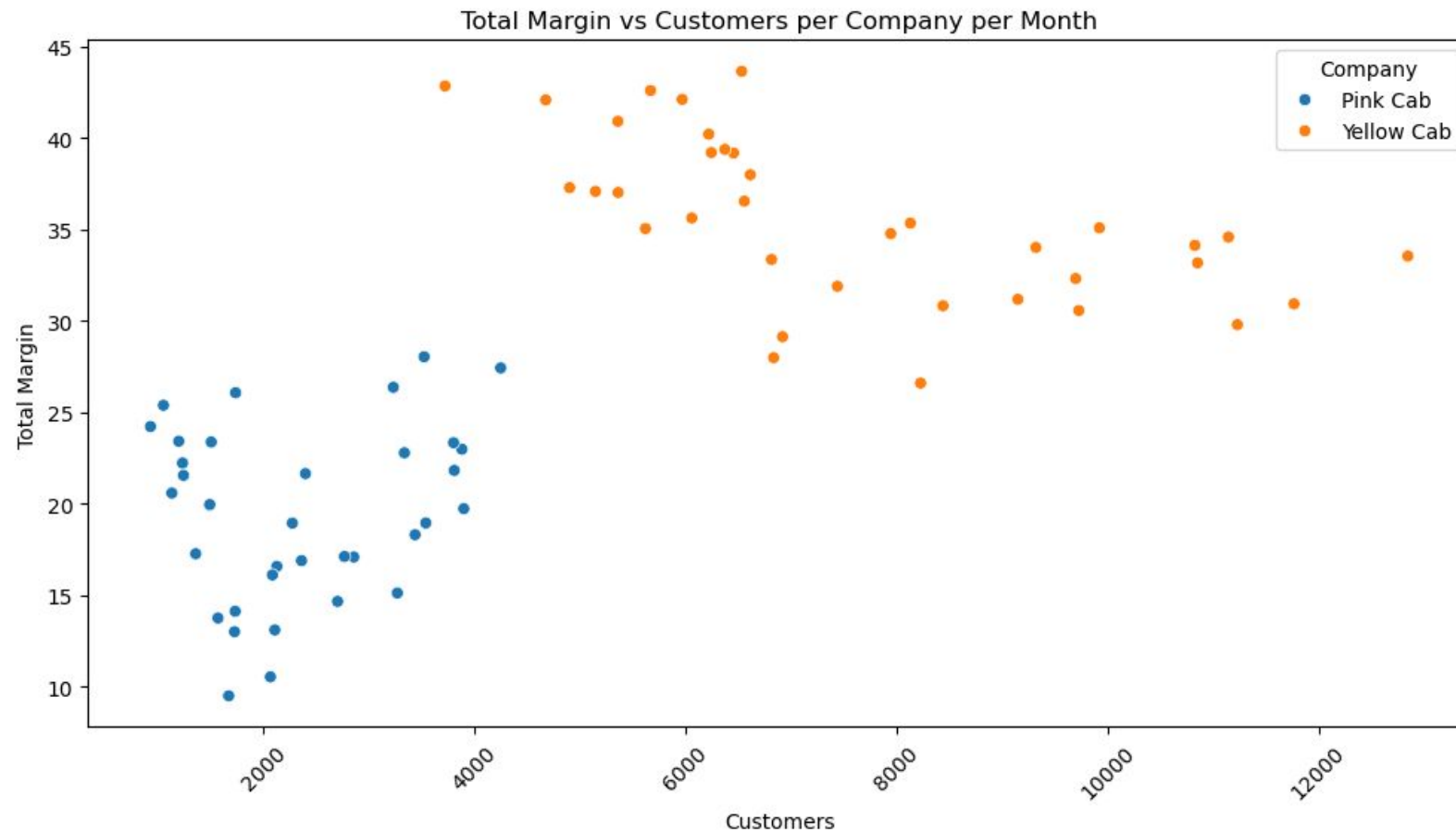
- **Larger cities** with higher populations tend to have **more users**.
- However, several cities stand out for having a particularly **high user-to-population ratio**. These include San Francisco, Chicago, Los Angeles, Washington D.C., and Boston — indicating a relatively **higher adoption rate** of the **cab services** in these areas compared to their population size.

# Seasonality in the demand



- Both companies exhibit a **clear seasonal pattern** in the number of unique monthly users.
- Each year, the **number of users steadily increases** from January to December, **peaking in December**, and then drops back to the starting level at the beginning of the following year.
- Notably, **Yellow Cab** consistently has 3,000 to 4,000 **more unique users** than Pink Cab each month.

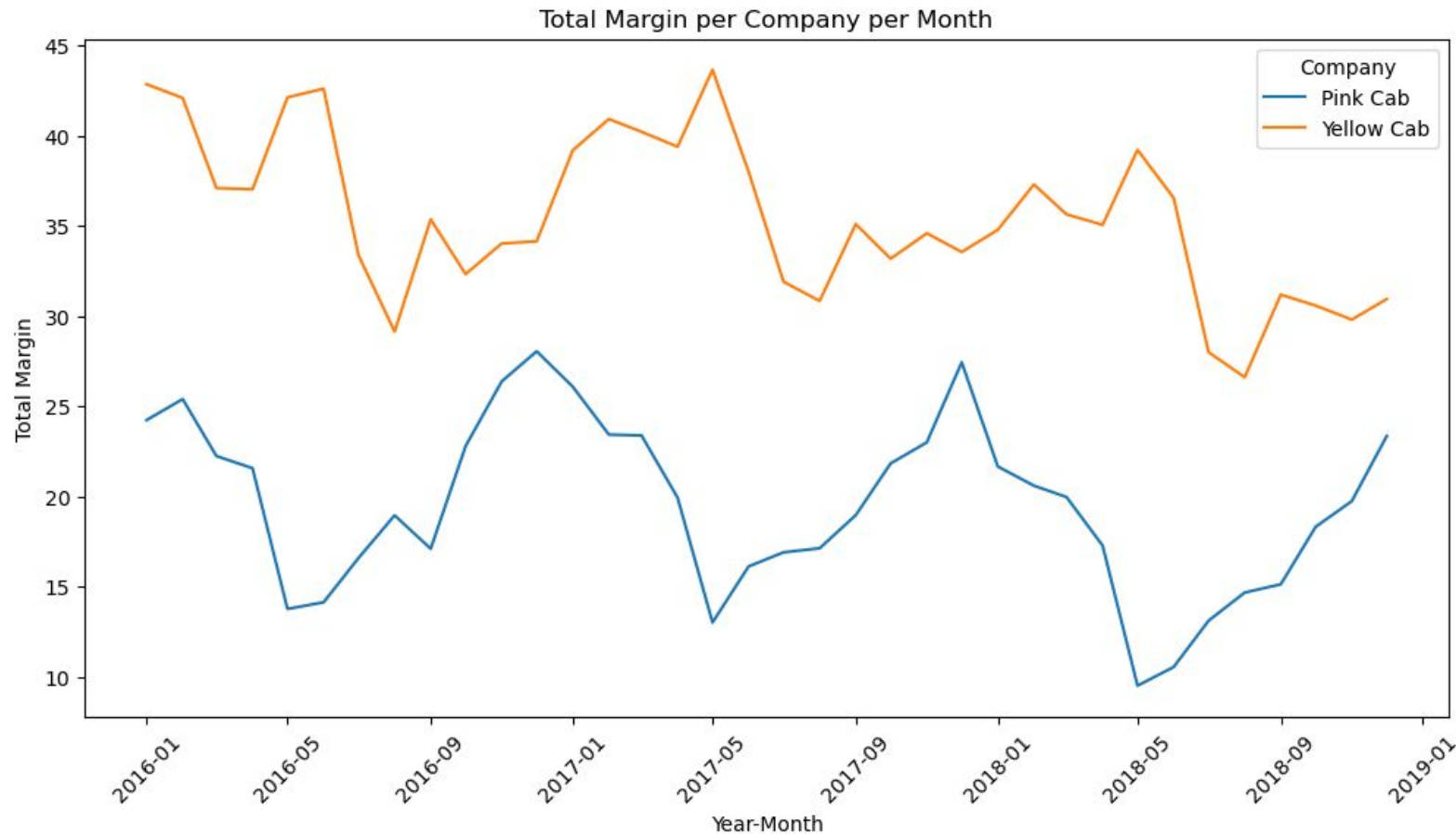
# Monthly Margin vs. Users



- Yellow Cab has negative relationship between users and monthly margin, while Pink Cab exhibits positive relationship.
- Margin reaches plateau once monthly customers reaching 8k.
- Overall high margin per customers in **Yellow Cab**



# Margin Trend



- **Yellow Cab** maintains a profit margin **twice as large** as that of Pink Cab, peaking between January and May, followed by a rebound in winter after a mid-year dip. It also exhibits a decreasing trend of margin from 2016 to 2018.
- **Pink Cab** displays a repeating annual pattern, with its margin peak from September to May, suggesting stronger performance during the academic or colder months.

# Recommendations

Based on evaluation of both cab companies, **Yellow** cab is **better** than Pink cab:

- **Customer Reach:** Yellow cab has higher monthly customer in general than pink cab, consistently covering more cab users than Pink cab.
- **Margin per customer:** Yellow cab retains a higher profit for each customer served than pink cab for almost all months than Pink cab.
- **Margin trend:** Yellow cab maintains a profit margin twice as large as that of Pink Cab.

On the basis of the above factors, I recommend Yellow cab for investment.

# Thank You