



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

<G2M insight for Cab Investment firm>

<21 November 2024>

Done by: Ethan Dy during LISUM39

# Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

# Executive Summary

## Investment Analysis of the US Cab Industry for XYZ

**Objective:** Evaluate two cab companies to determine the best investment opportunity for XYZ.

### Key Findings:

- Yellow Cab demonstrates stronger performance across ride volume, profitability, and customer loyalty.
- Middle-aged customers (30-50 years) drive the highest ride frequency.
- City-level demand varies and does not align strictly with population density.

# Problem Statement

## Industry Context:

- The US cab industry has seen remarkable growth in recent years with increasing competition among key players.
- XYZ, a private firm, aims to enter the market and needs data-driven insights to decide between two cab companies for investment.

## Data Overview:

- Cab\_Data.csv**: Transaction details for two companies from 2016 to 2018.
- Customer\_ID.csv**: Customer demographics and unique identifiers.
- Transaction\_ID.csv**: Links transactions to customers and payment modes.
- City.csv**: City population, cab usage, and other geographical details.

## Challenges:

- Split and diverse datasets require management and cleaning.
- Understanding the profitability trends, customer wants/needs, and demographics.
- Identifying the extreme insights for a strategic investment decision.

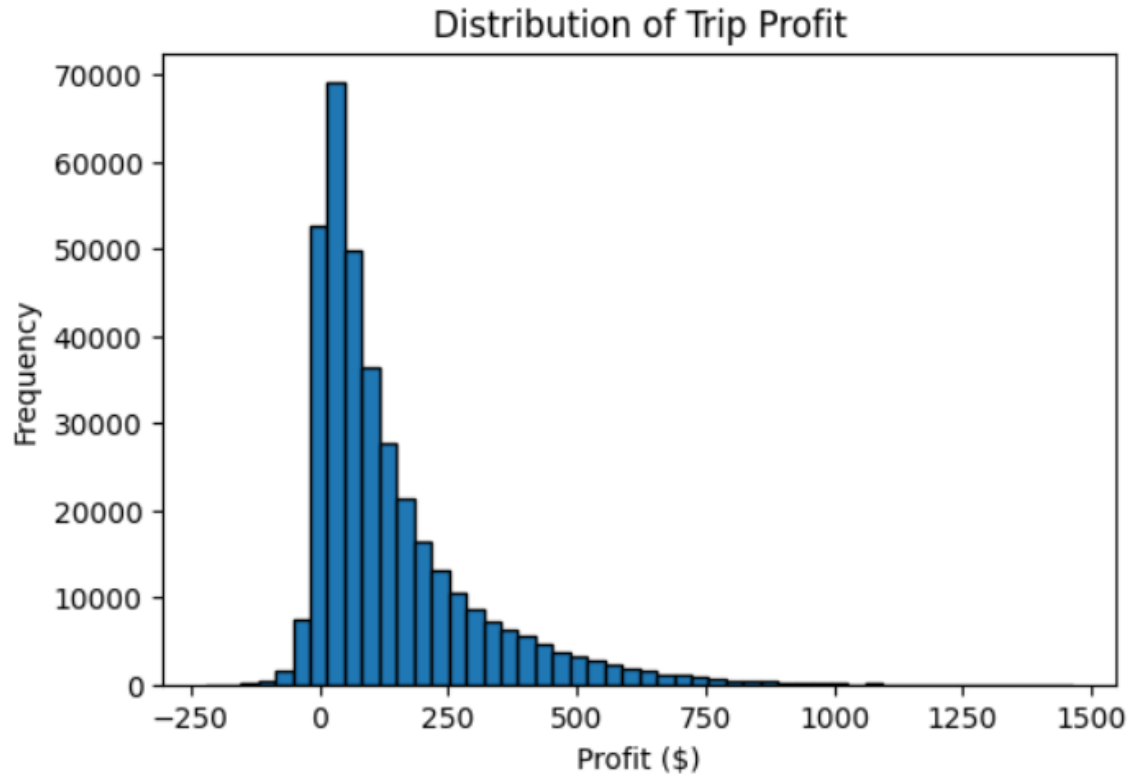
# Approach

## Content:

- 1.Data Preparation:** Cleaned, transformed, and merged datasets (Cab\_Data, Customer\_ID, Transaction\_ID, City.csv).
- 2.Exploratory Data Analysis (EDA):** Investigated trends in profitability, customer demographics, and city-level usage.
- 3.Hypothesis Testing:** Validated assumptions about ride volume, city size, and customer loyalty.
- 4.Comparison:** Evaluated Yellow Cab vs. Pink Cab on key performance metrics.

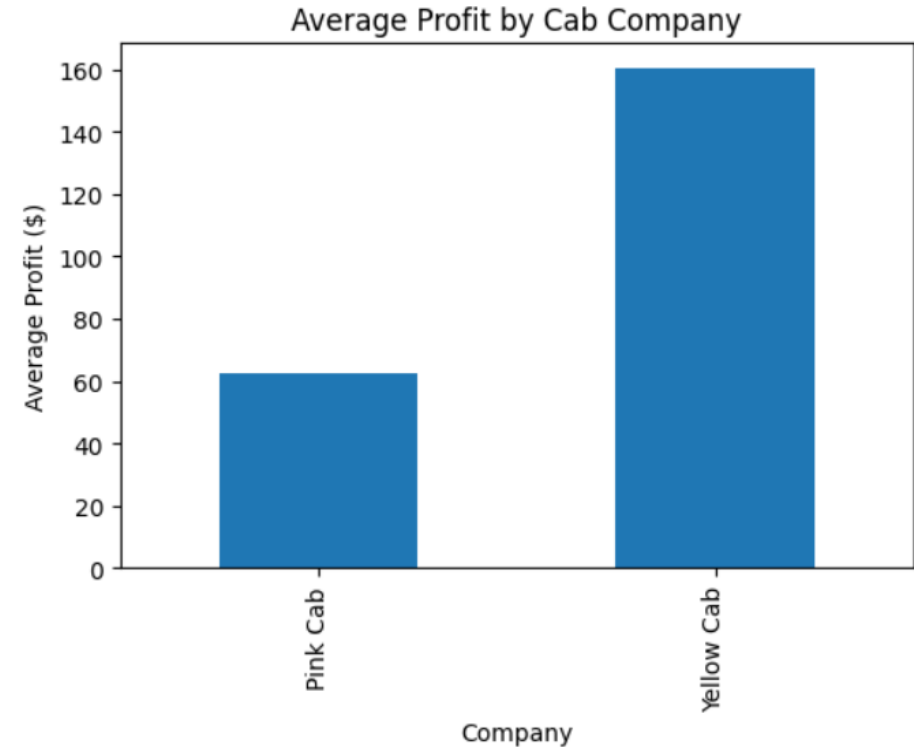
# Exploratory Data Analysis (EDA)

## Distribution of Trip Profit



- Most trips skew towards lower profit margins
- Majority of profit is below \$250

## Average Profit by Cab Company

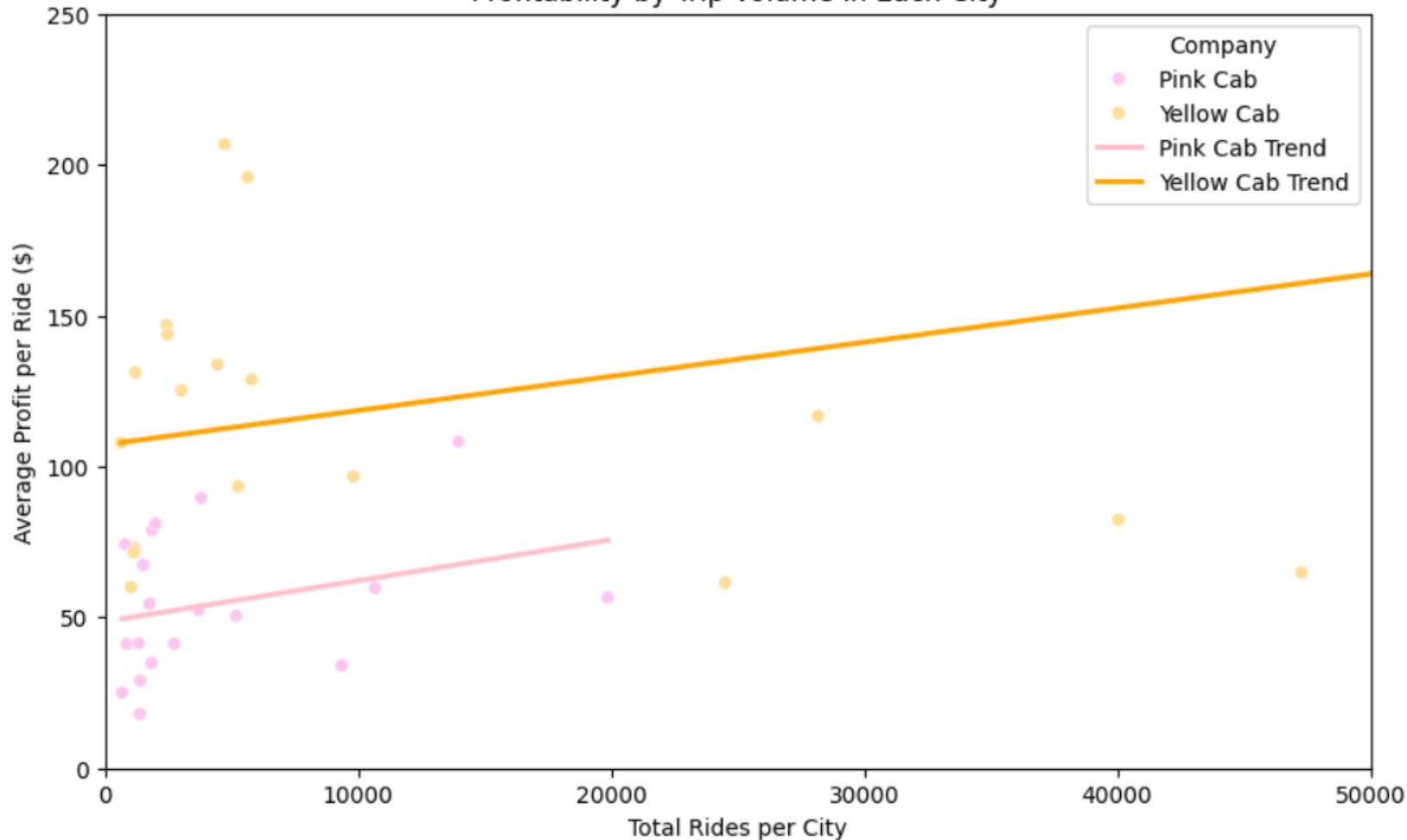


- Yellow cab shows higher profits
- Profit gap favors yellow cab

# Exploratory Data Analysis (EDA)

## Cities Profitability by Trip Volume

Profitability by Trip Volume in Each City

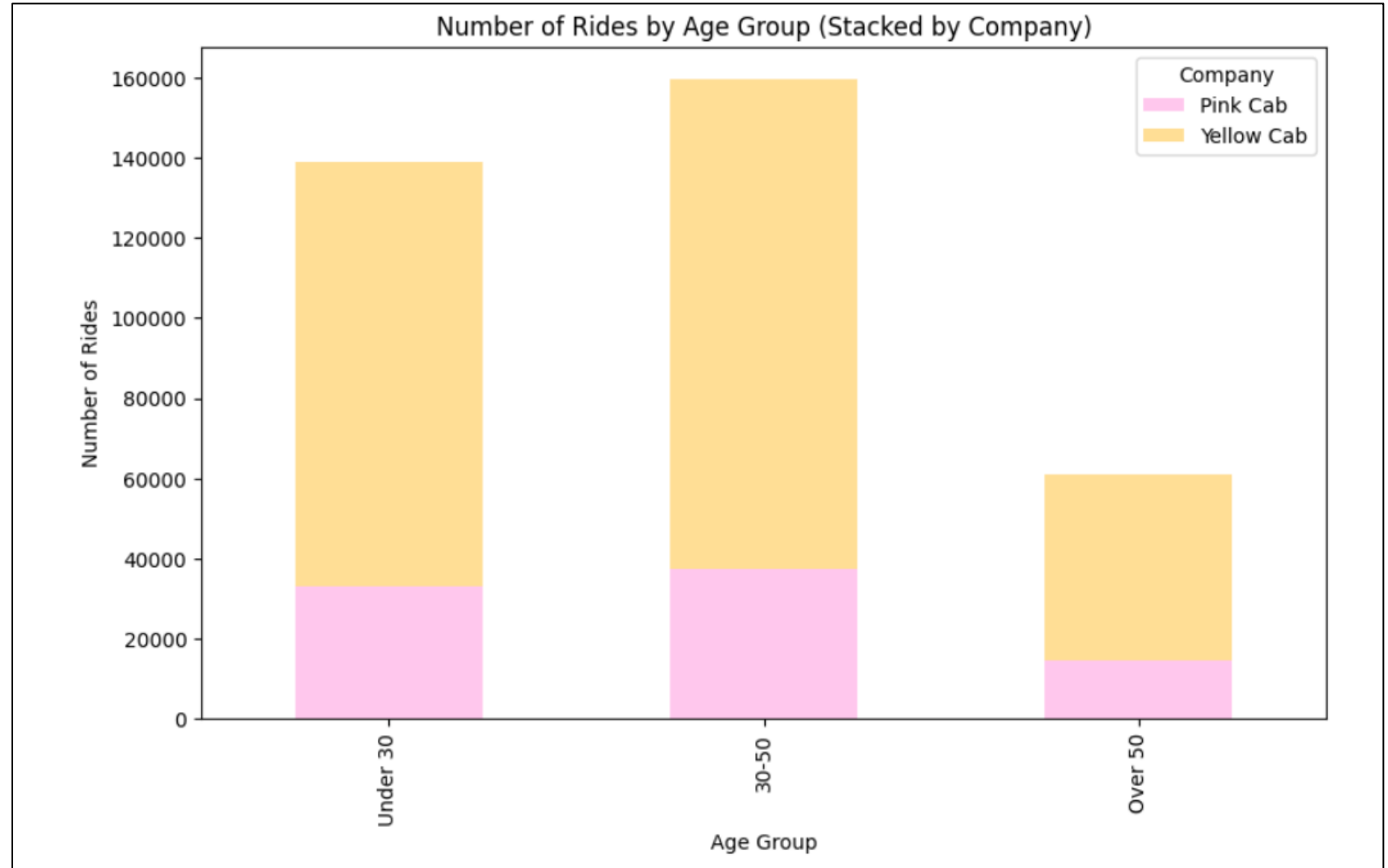


- **Null Hypothesis (H0):** There is no relationship between the number of rides in a city and the average profit per ride for either company.
- **Alternative Hypothesis (H1):** Cities with a higher number of rides have a lower average profit per ride for both companies.

# Exploratory Data Analysis (EDA)

## Customer Age Impact on # of Rides

- **Null Hypothesis (H0):** There is no difference in the average number of rides taken by older customers compared to younger customers, regardless of the cab company.
- **Alternative Hypothesis (H1):** Older customers take more rides on average than younger customers, regardless of the cab company.



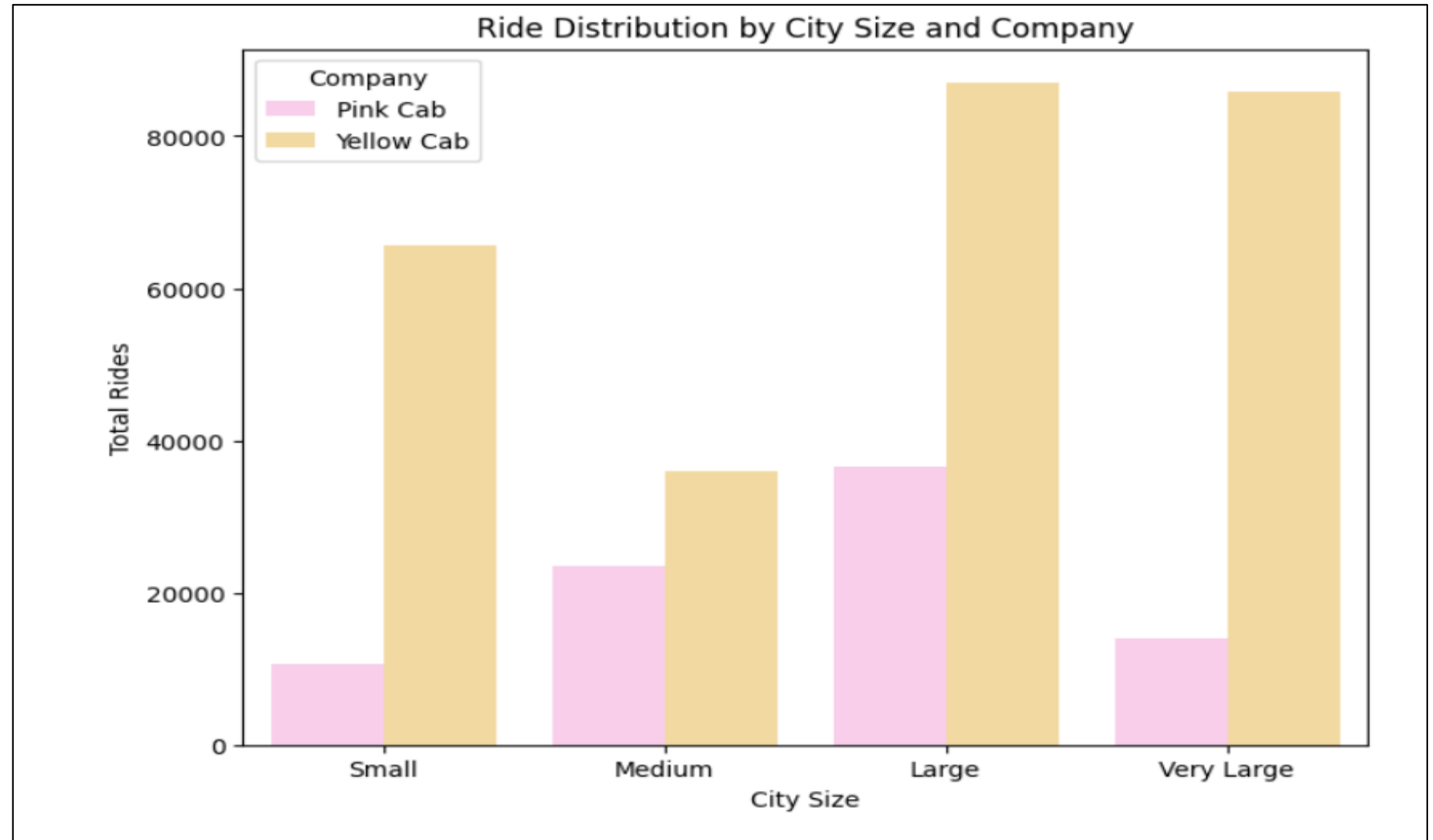


# Exploratory Data Analysis (EDA)

## Ride Distribution from City Size

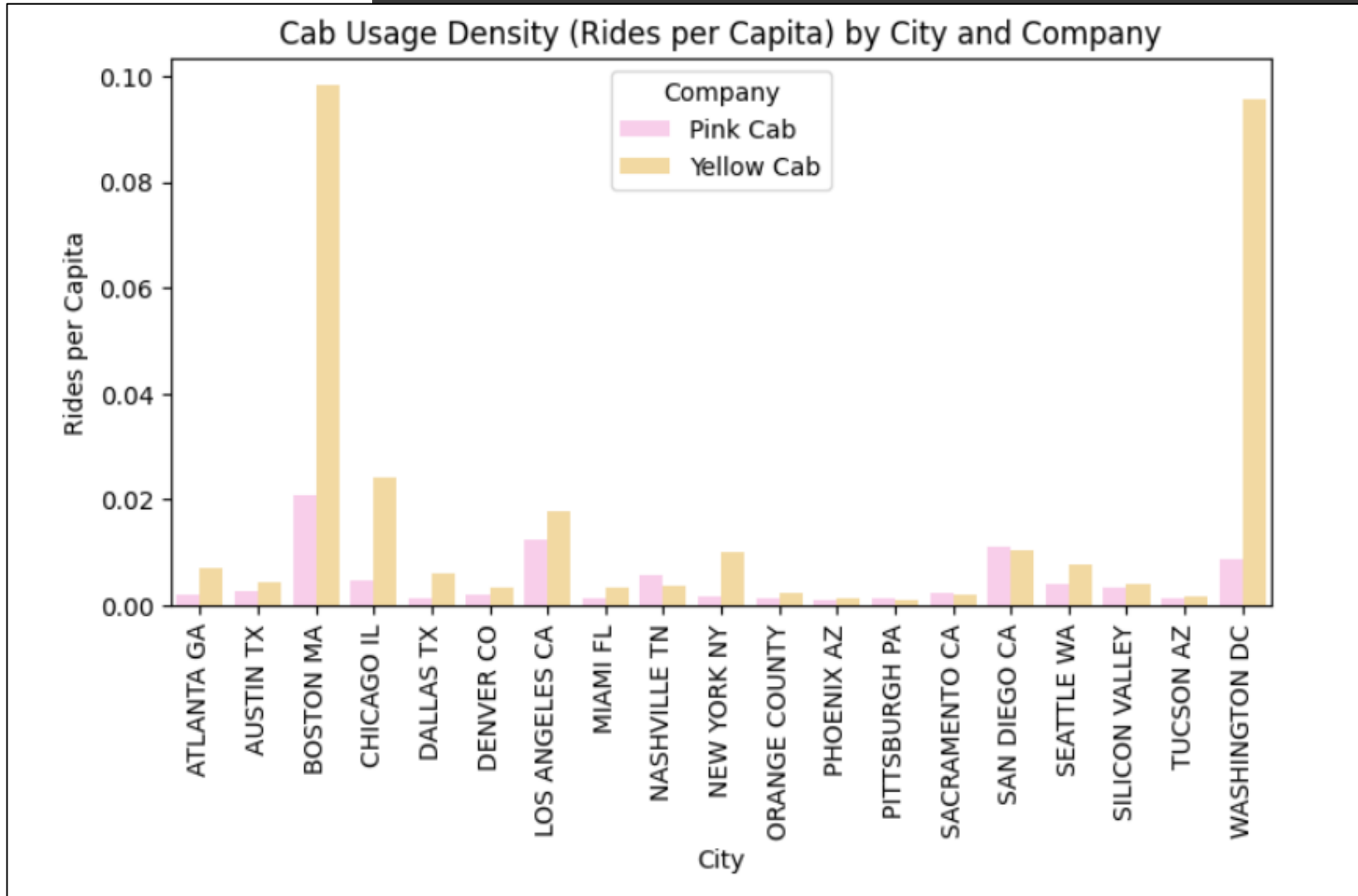
- **Null Hypothesis (H0):** There is no difference in the presence of Pink Cab and Yellow Cab across cities of different sizes.

- **Alternative Hypothesis (H1):** Pink Cab has a stronger presence in smaller cities, while Yellow Cab is more dominant in larger cities.



# Exploratory Data Analysis (EDA)

## City Pop. Density vs Cab Usage



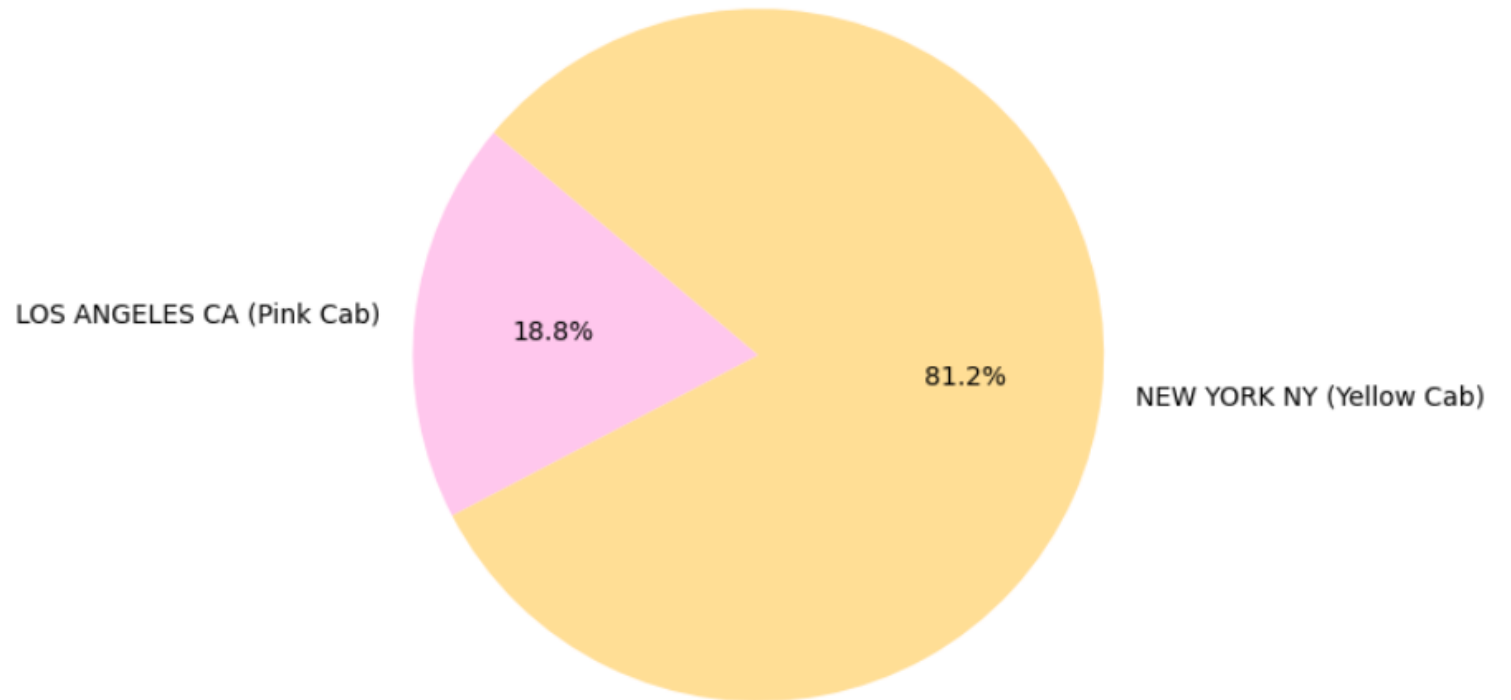
- **Null Hypothesis (H0):** Cab usage density (rides per capita) does not vary with city population size for either company.
- **Alternative Hypothesis (H1):** Cab usage density is higher in more densely populated cities, indicating a stronger customer base in these areas for both companies.

# Exploratory Data Analysis (EDA)

## Top City for Repeat Customers

- **Null Hypothesis (H0):** There is no difference in the concentration of repeat customers between Yellow Cab and Pink Cab in any specific city.
- **Alternative Hypothesis (H1):** Yellow Cab has a higher concentration of repeat customers in certain cities compared to Pink Cab.

Top City for Repeat Customers by Company



# EDA Summary

## Hypothesis Results

### 1. Profitability by Trip Volume:

Result: **Not supported**. The data did not show an inverse relationship between ride volume and profit per ride; in fact, both companies showed stable or slightly increased profitability with higher ride counts.

### 2. Customer Age Impact on Ride Frequency:

Result: **Not supported**. Middle-aged customers (age 30-50) took the most rides, while the "Over 50" group took the fewest, contradicting the initial hypothesis.

### 3. Ride Distribution Based on City Size:

Result: **Not supported**. Yellow Cab was dominant across all city sizes, with no indication that Pink Cab had a stronger presence in smaller cities.

### 4. City Population Density vs. Cab Usage:

Result: **Partially supported**. Some densely populated cities showed high usage density, but this was inconsistent across all cities, suggesting other factors beyond population size influence cab demand.

### 5. Top City for Repeat Customers by Company:

Result: **Supported**. The data showed a significantly higher concentration of repeat customers for Yellow Cab in New York City compared to Pink Cab's top city (Los Angeles).

# Recommendations

After analyzing the data and testing my hypotheses, here are some of my recommendations:

## 1. **Target Middle-Aged Customers to Boost Ride Numbers**

- Finding: Customers aged 30-50 are taking the most rides, while those over 50 are taking the least

## 2. **Explore Different Opportunities in High-Demand Cities**

- Finding: Certain cities like Boston, MA and Washington, DC have higher cab usage density, but this doesn't always match up with population size

## 3. **Differentiate Services to Stand Out**

- Finding: Yellow Cab dominates in ride numbers across all city sizes, and Pink Cab doesn't have a stronger presence in smaller cities as initially thought

## 4. **Increase Customer Loyalty and Repeat Business**

- Finding: Yellow Cab has more repeat customers, especially in cities like New York

# Thank You