



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

G2M Case Study  
Virtual Internship

21-Jan-2024

# Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

# Background –G2M(cab industry) case study

## Background

XYZ is a private equity firm based in the United States. Recognizing the rapid growth and competitive landscape of the cab industry in recent years, the firm is planning to invest in this sector to capitalize on its market potential.

## Objective

The primary objective is to provide actionable insights that will guide XYZ in selecting the most profitable cab company for investment, ensuring long-term growth and returns.

## Analysis Scope

The analysis is divided into four major components:

**Data Exploration and Preparation:** Understanding the datasets, cleaning the data, and deriving useful features to support the analysis.

**Profitability and Demand Analysis:** Evaluating each cab company's profitability and customer demand to identify key drivers of success.

**Comparison of Companies:** Assessing financial and operational metrics to determine the stronger investment opportunity.

**Investment Recommendations:** Providing strategic advice on which company offers the highest potential for growth and returns.

## Data Exploration

For this project, the data exploration phase involved a detailed review and preparation of multiple datasets to ensure they were ready for analysis. Here's an outline of the exploration process:

### Key Features of the Data

**Number of Features:** A total of 24 features were analyzed, including 9 derived fields created during the data preparation stage.

**Timeframe:** The data spans from January 31, 2016, to December 31, 2018.

**Total Records:** The combined dataset contains 355,032 individual transaction records.

### Datasets Used

**Cab\_Data.csv:** Contains trip details such as kilometers traveled, price charged, and cost of the trip.

**Transaction\_ID.csv:** Links transactions to customers and identifies payment methods.

**Customer\_ID.csv:** Provides demographic details, including age, gender, and income.

**City.csv:** Includes city-level data such as population and number of cab users.

## Data Cleaning and Integration

**Data Merging:** The datasets were merged to form a unified master dataset, linking transaction data to customer and city attributes. Key relationships between Transaction ID, Customer ID, and City were used to ensure a seamless integration.

**Handling Missing Values:** No missing values were found in the data, which streamlined the preparation process.

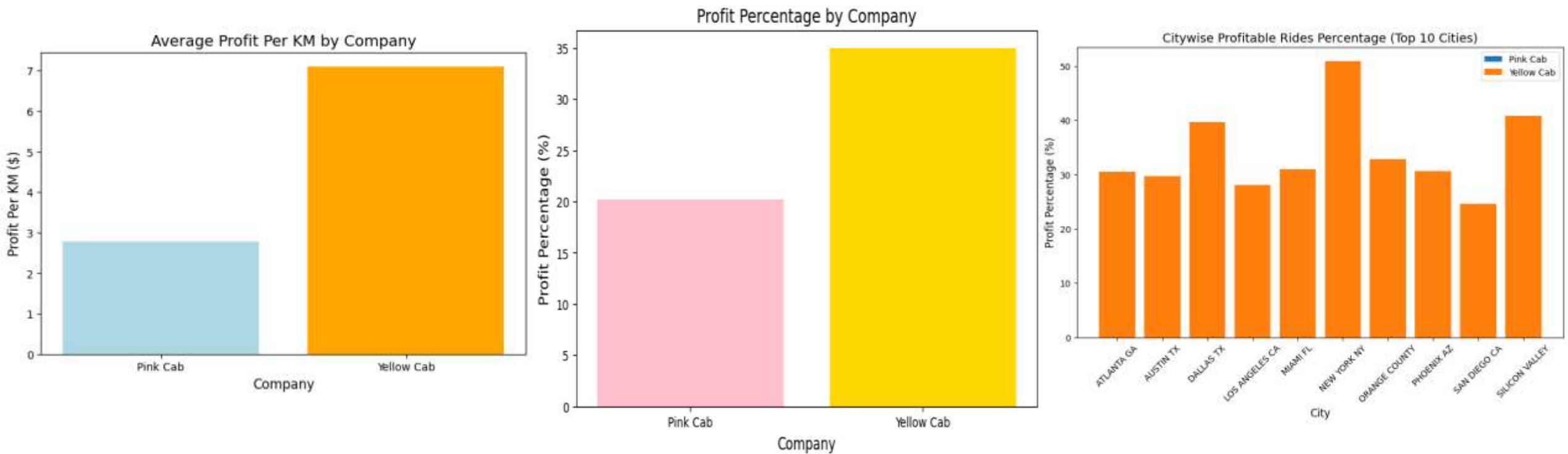
**Derived Features:** Additional fields, such as Year and Month, were created from the Date of Travel field to facilitate time-based trend analysis.

## Assumptions and Approach

**Outliers:** While outliers were identified in the Price Charged field, they were retained, as the lack of trip duration details made it impractical to accurately determine anomalies.

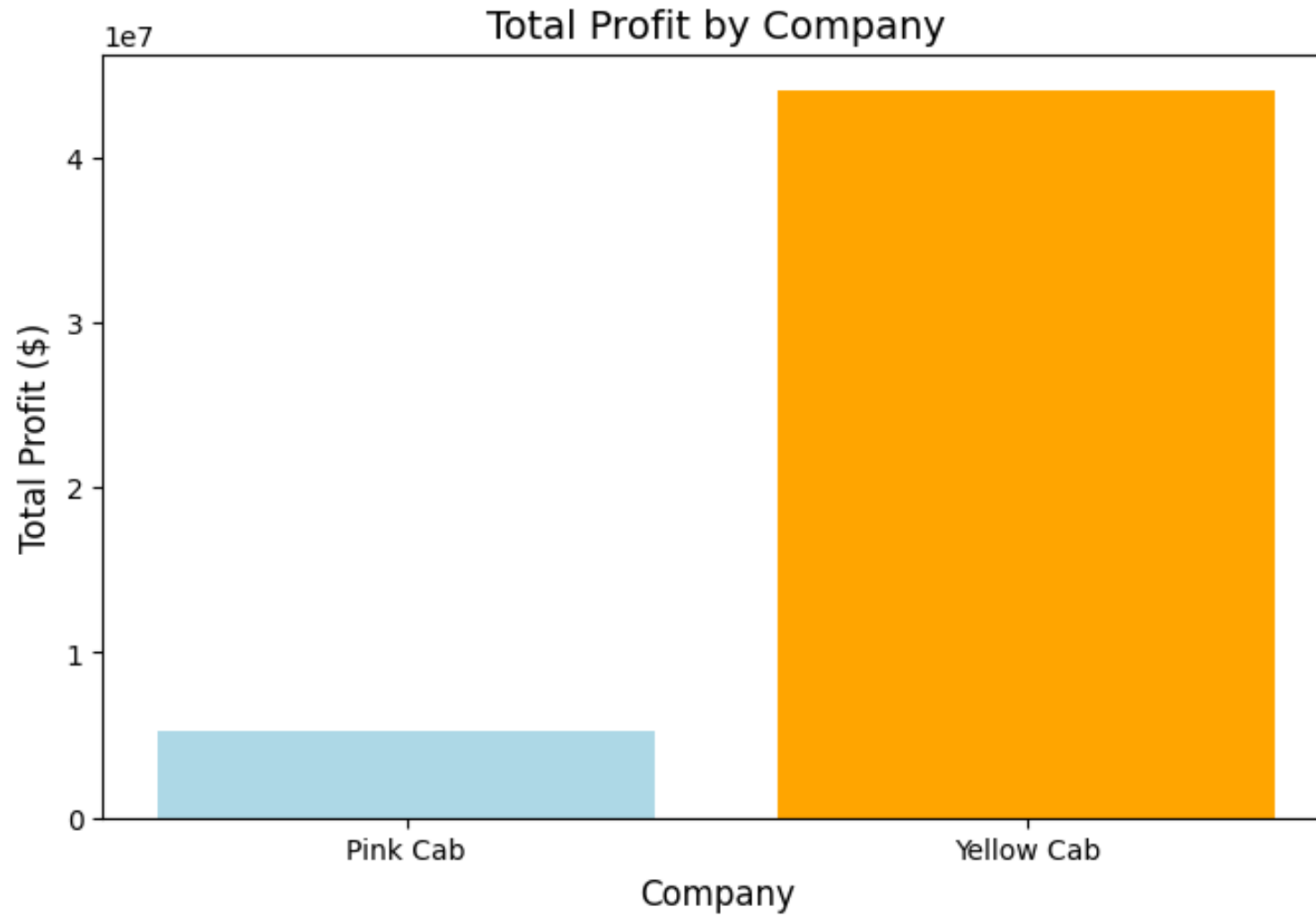
**Profit Calculation:** Profits were calculated based on the difference between Price Charged and Cost of Trip, assuming other factors remained constant.

**City Users:** The Users field in the city dataset was treated as the total number of cab users, including both Yellow Cab and Pink Cab customers.

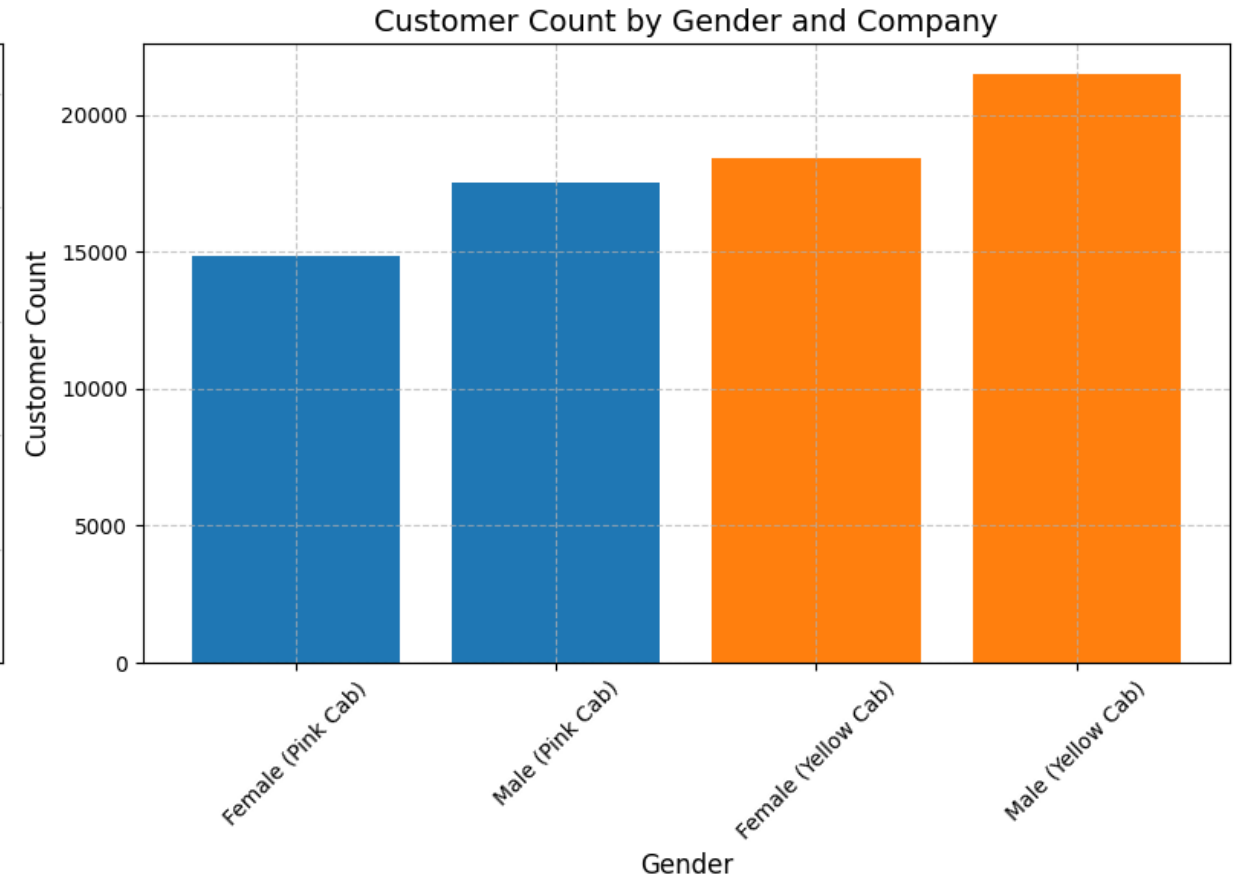
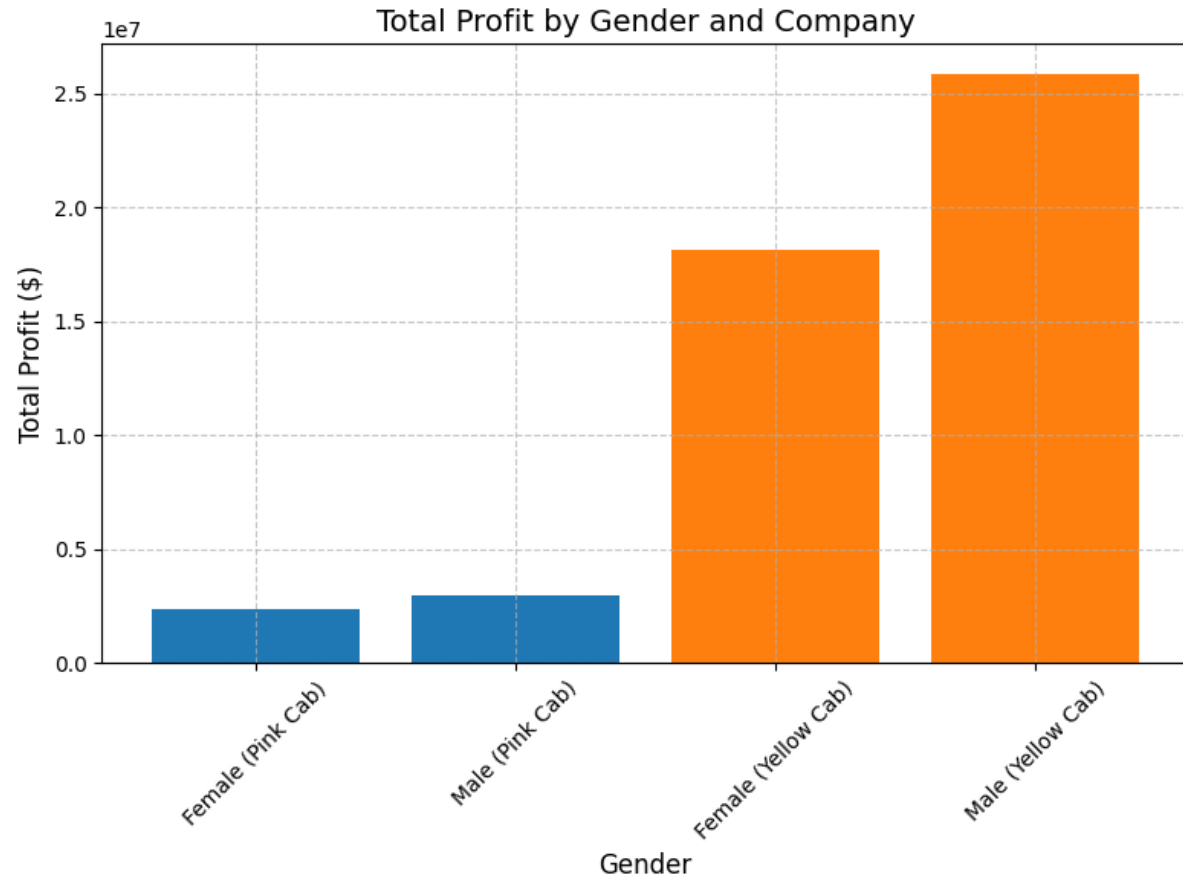


Company	Total Profit	Total Rides	Profit Per Ride
Pink Cab	\$5,237,540.00	83,825	\$62.48
Yellow Cab	\$43,251,775.00	271,207	\$159.48

# Total Profit Comparison Between Companies



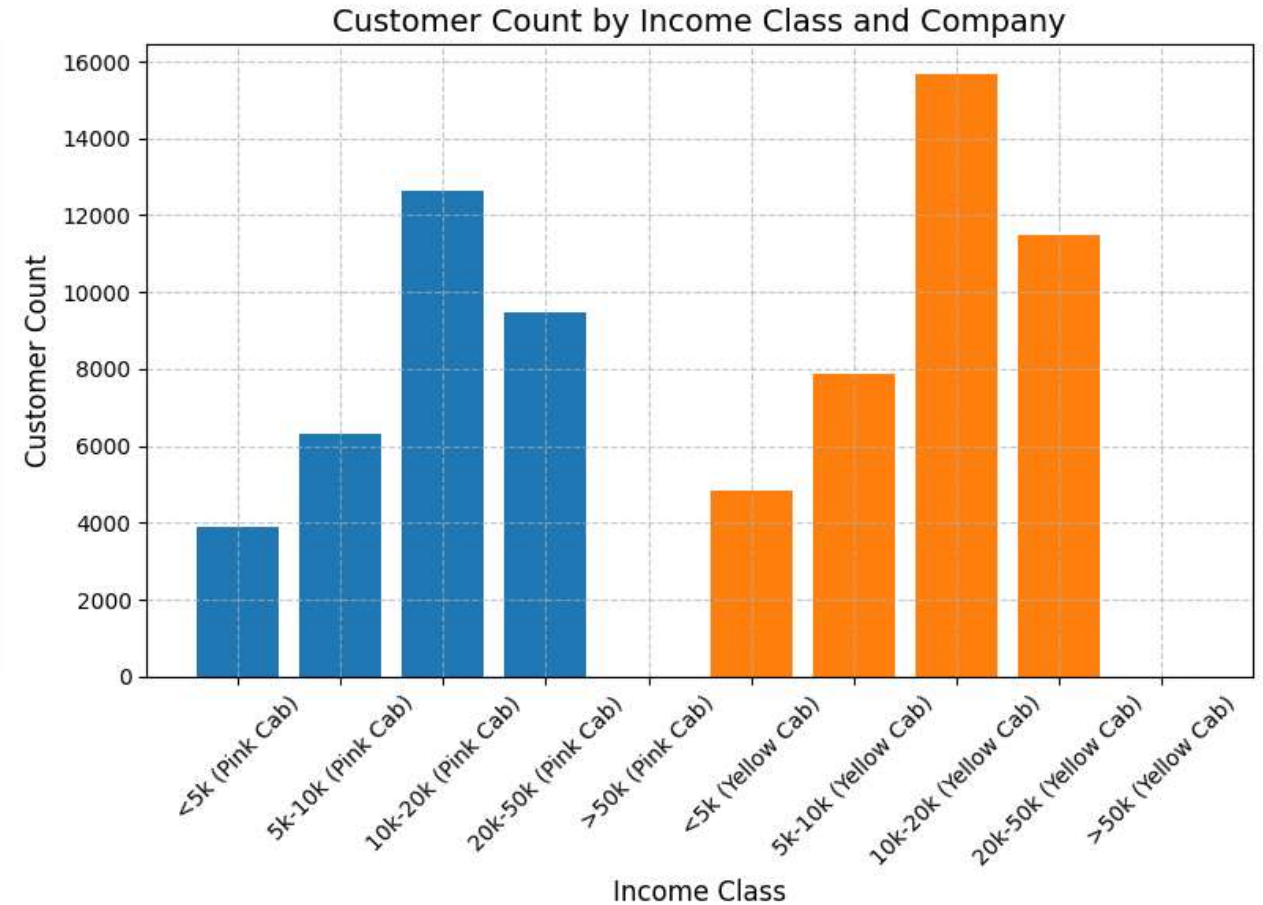
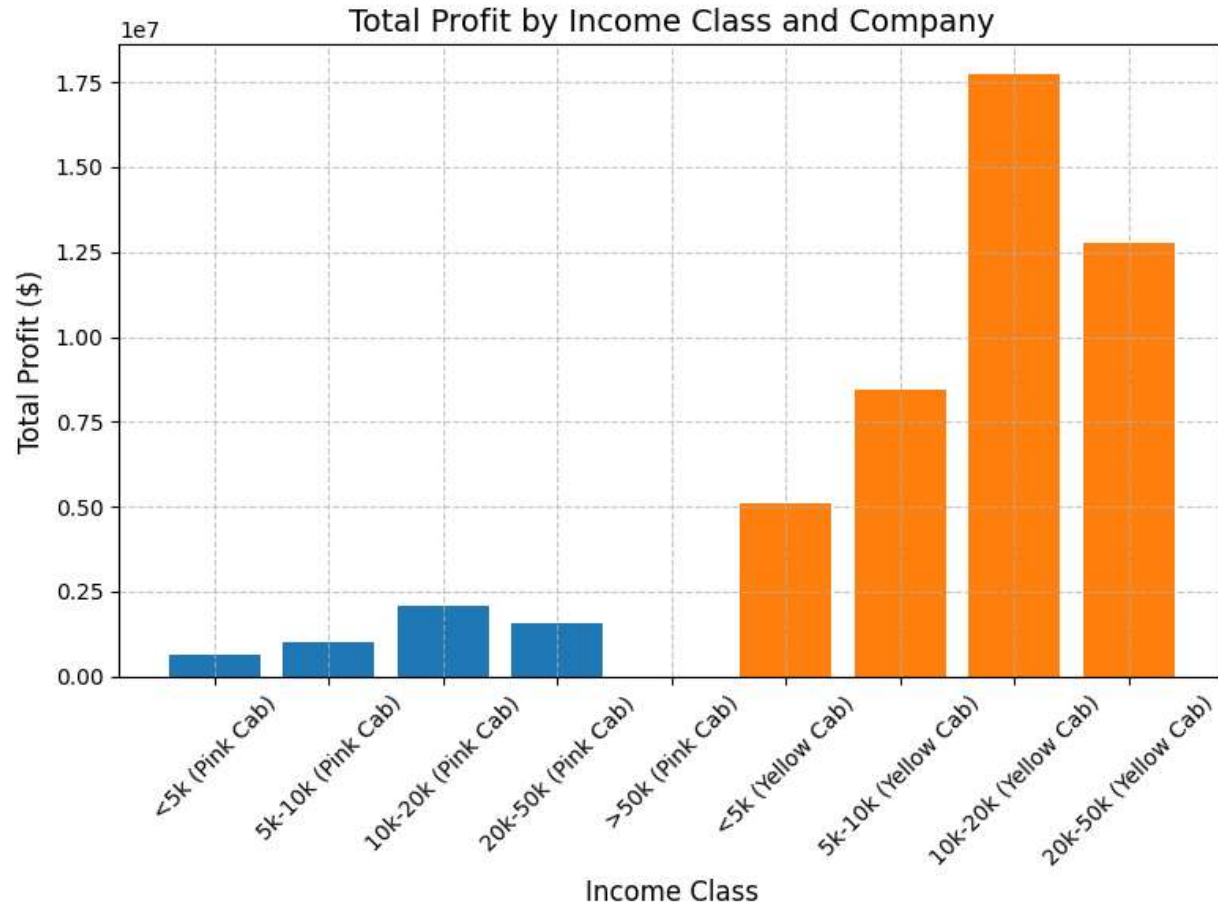
## Gender-Wise Profit and Customer Base Analysis



**Yellow Cab** outperforms **Pink Cab** in profit and customer base across both genders, reflecting broader appeal and stronger profitability.

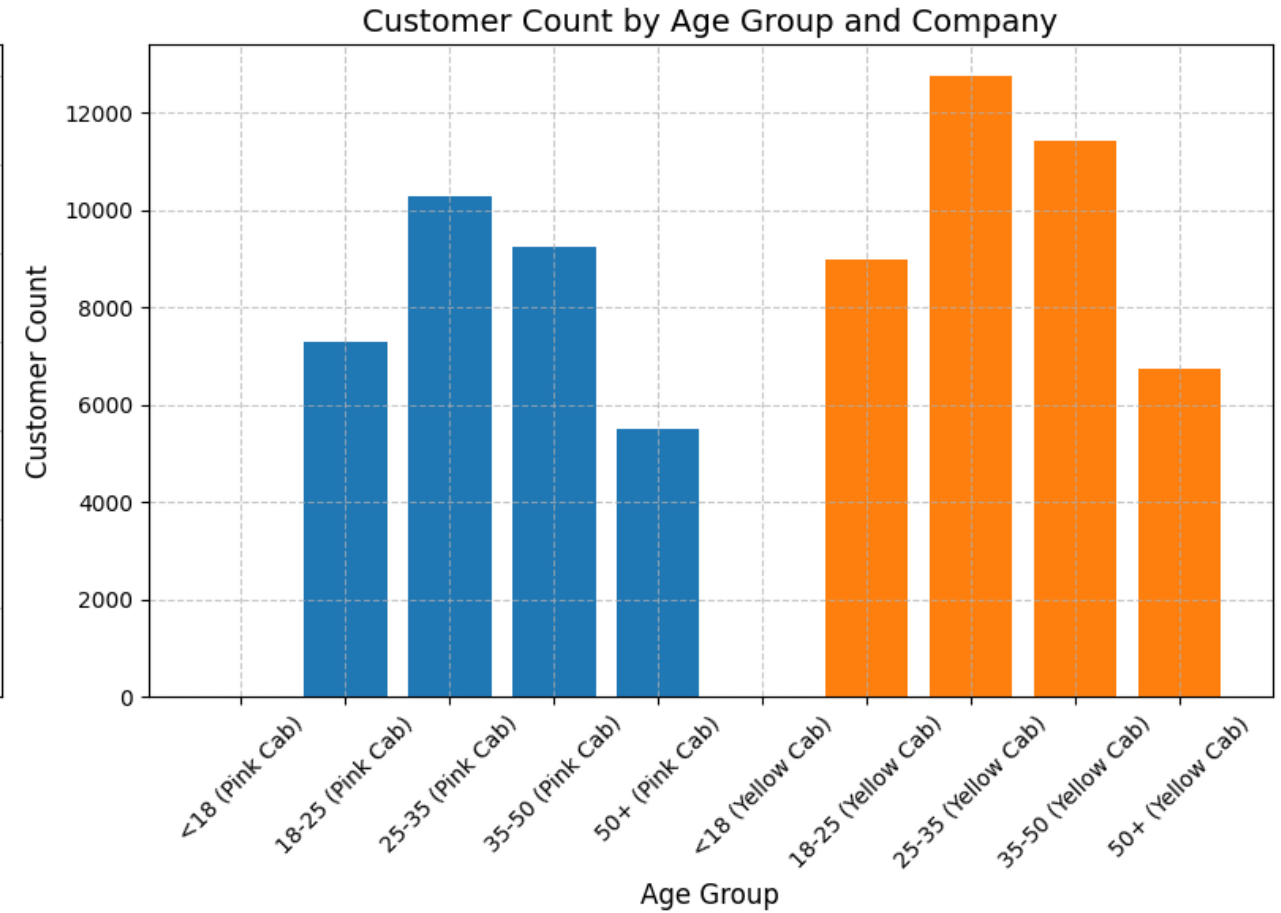
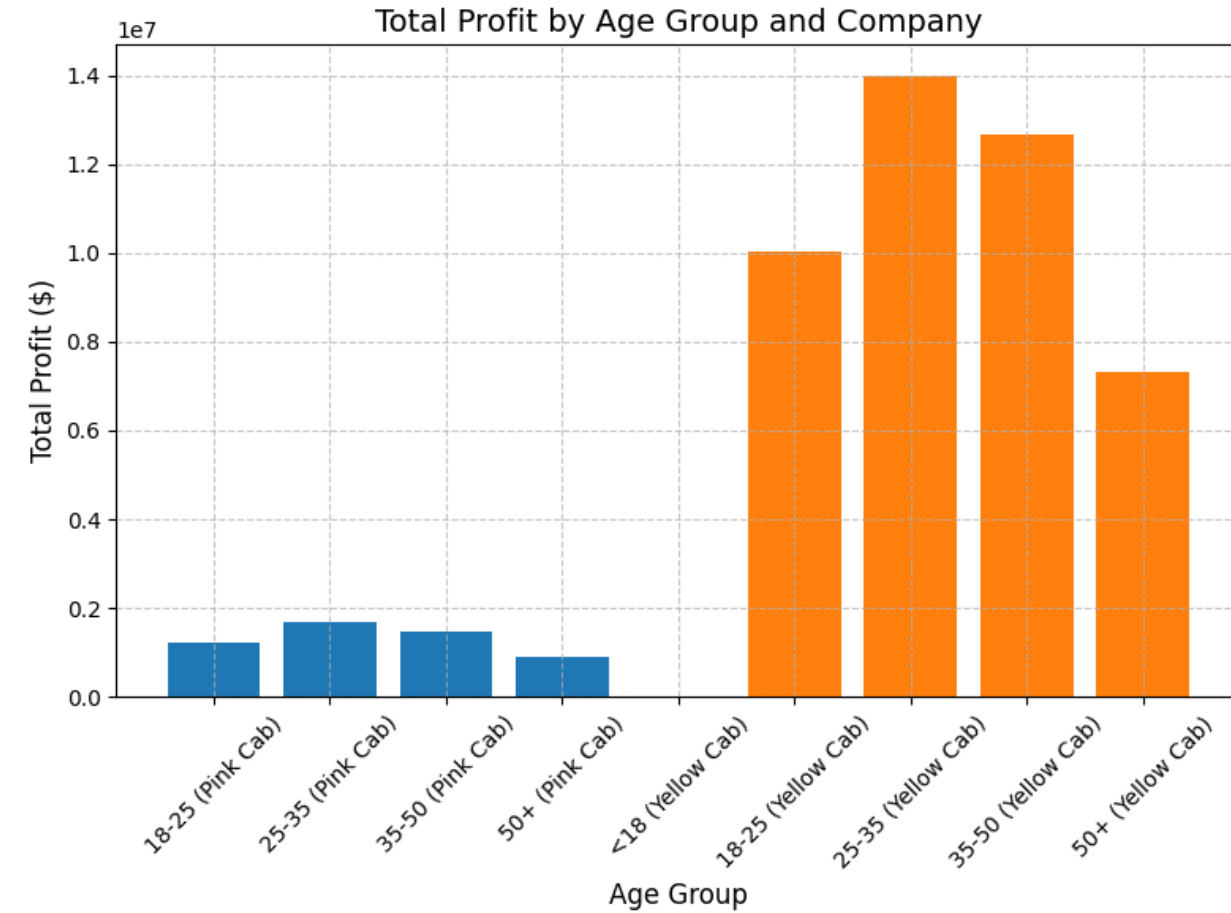


## Income Class-Wise Profit and Customer Base Analysis

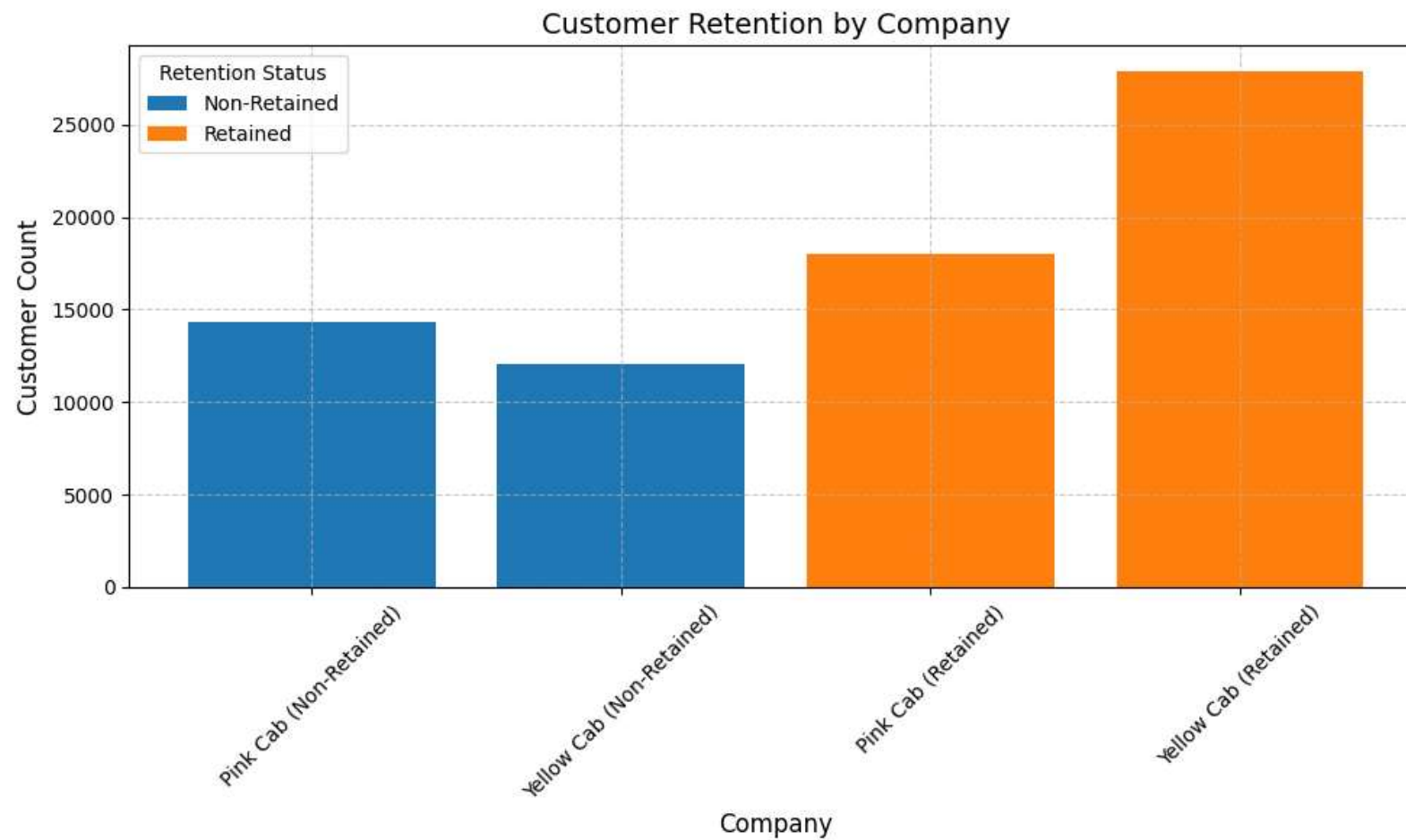


**Yellow Cab** dominates in both profit and customer base across all income segments, solidifying its position as the more lucrative and widely adopted service provider.

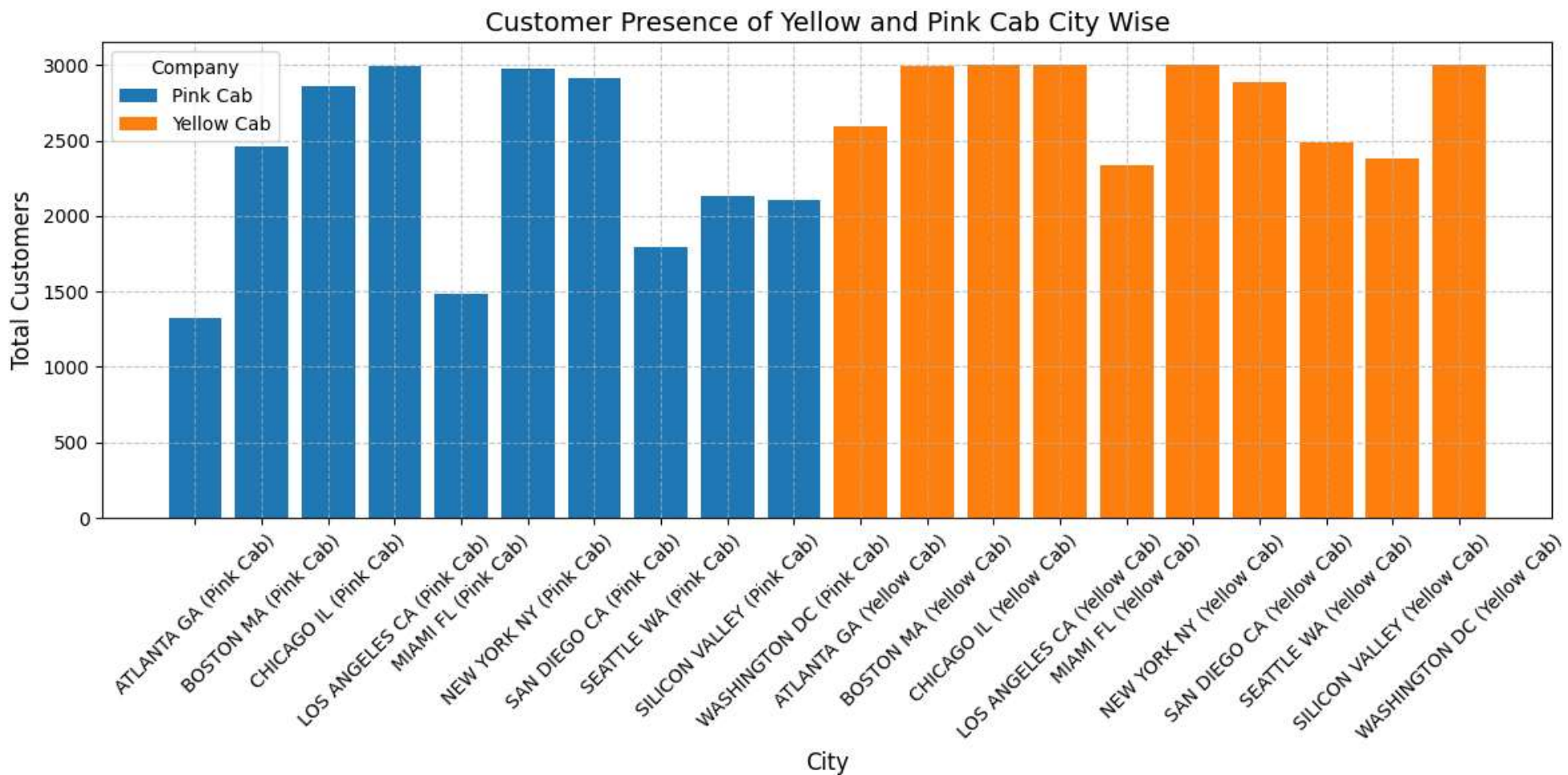
## Age Group-wise Profit and Customer Base Analysis



The **25-35 age group** contributes the highest to both **profit** and **customer base** for both companies

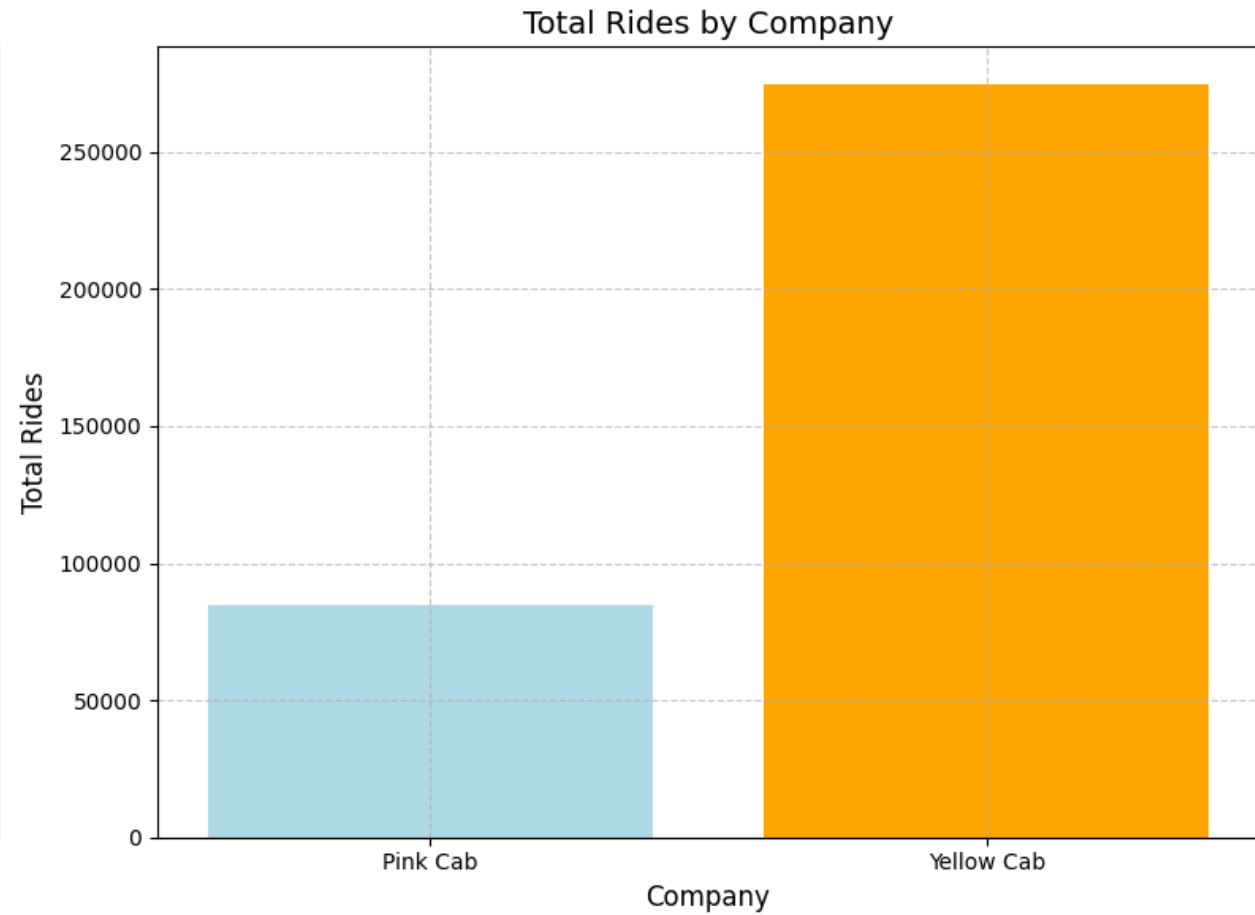
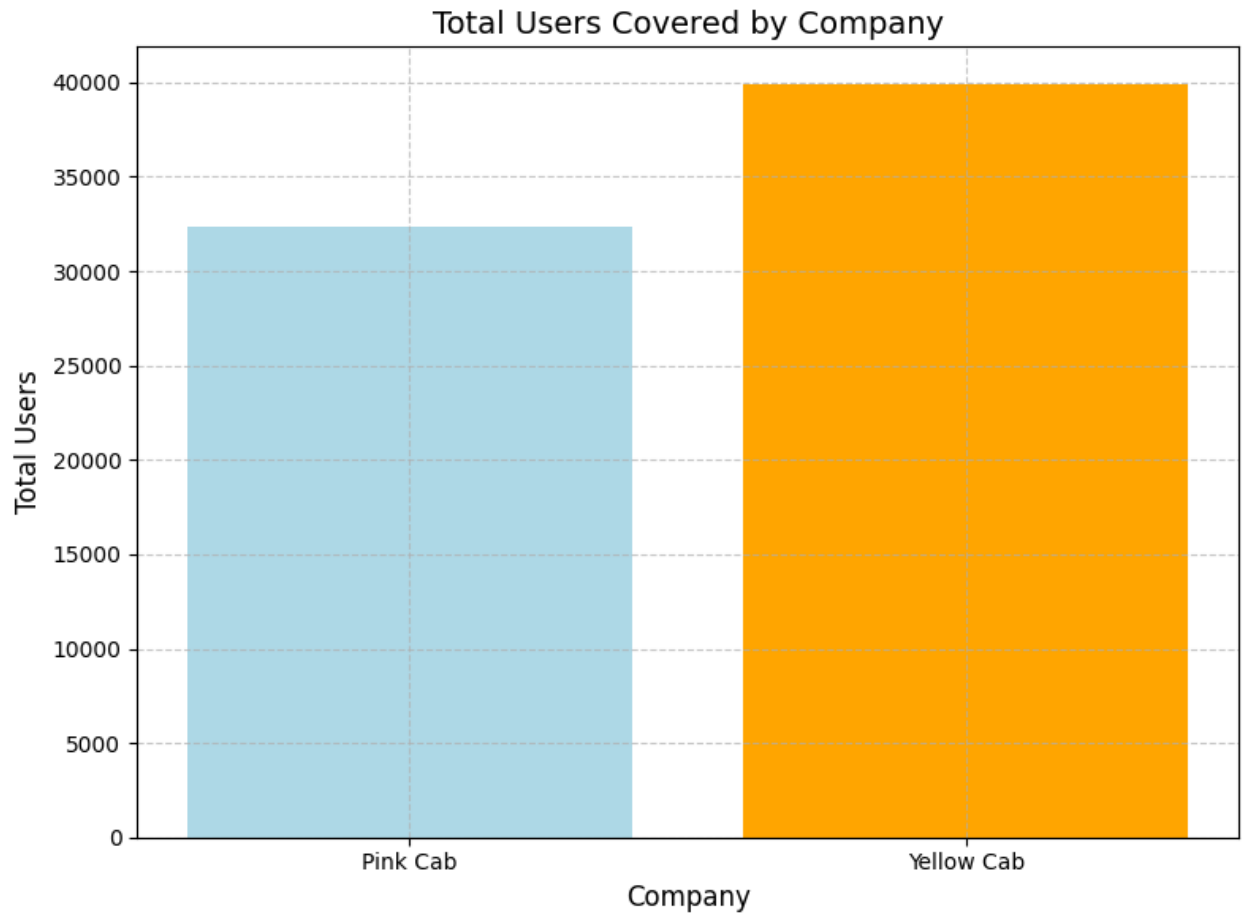


**Yellow Cab** retains a significantly higher number of customers compared to **Pink Cab**



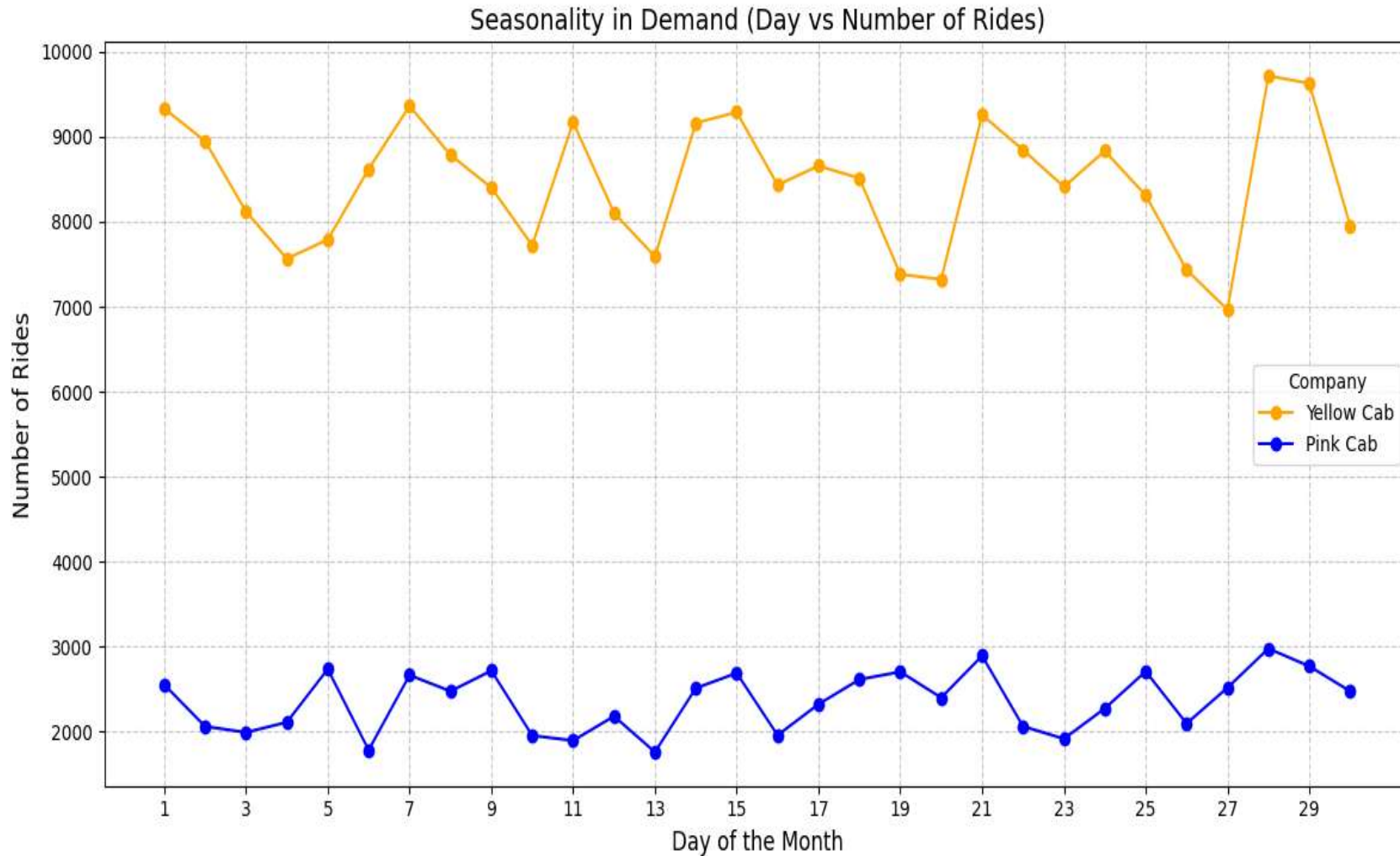
**Yellow Cab** consistently maintains a stronger presence across the top cities

# Users Covered by Company and Customer Base



Yellow Cab outperforms Pink Cab in both total users and rides, showcasing stronger market reach, operational efficiency, and a dominant position in the cab market.

# Seasonality in the demand



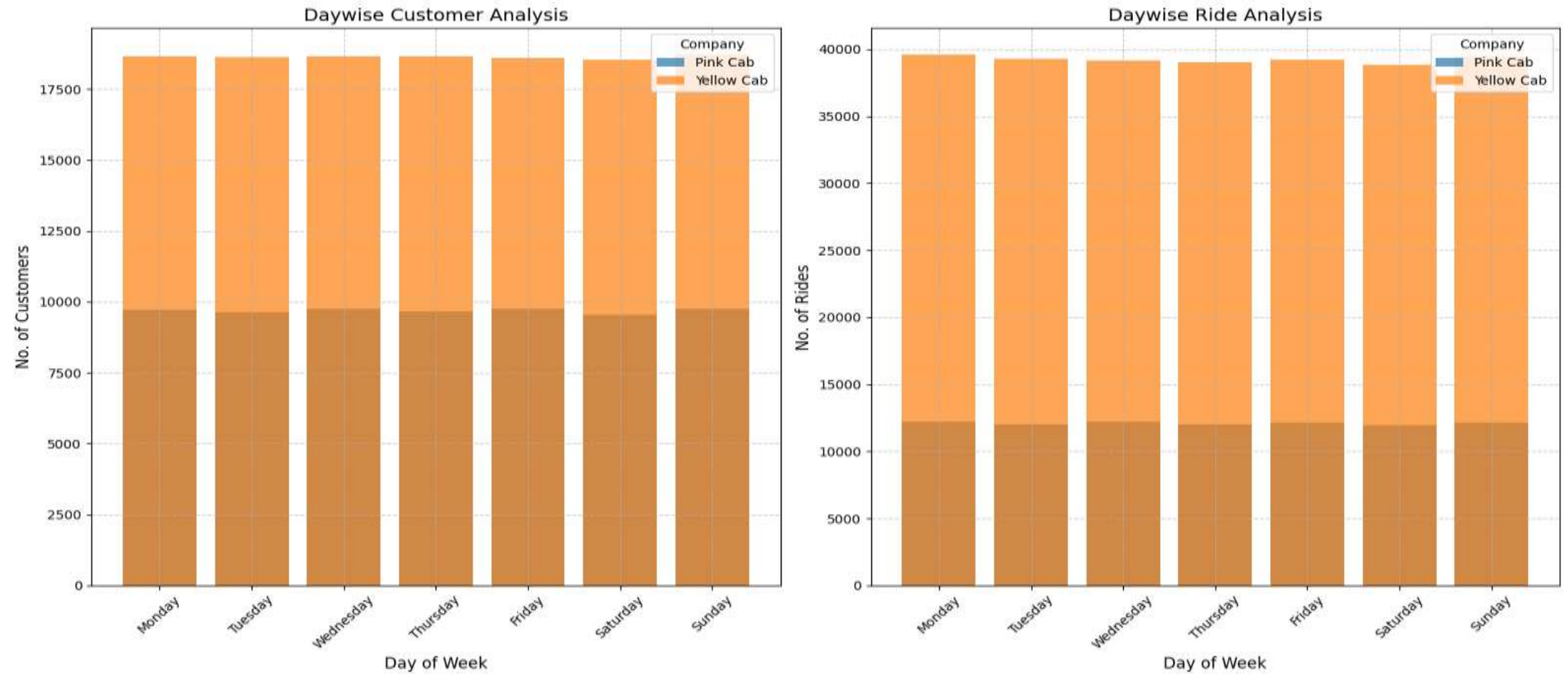
- Seasonal Pattern:** Both Yellow Cab and Pink Cab show a recurring weekly demand pattern.

- Demand Peaks:** Demand is consistently higher on the 3rd day of the week and follows a similar 7-day cycle.

- Yellow Cab Dominance:** Yellow Cab experiences significantly higher demand compared to Pink Cab across all days

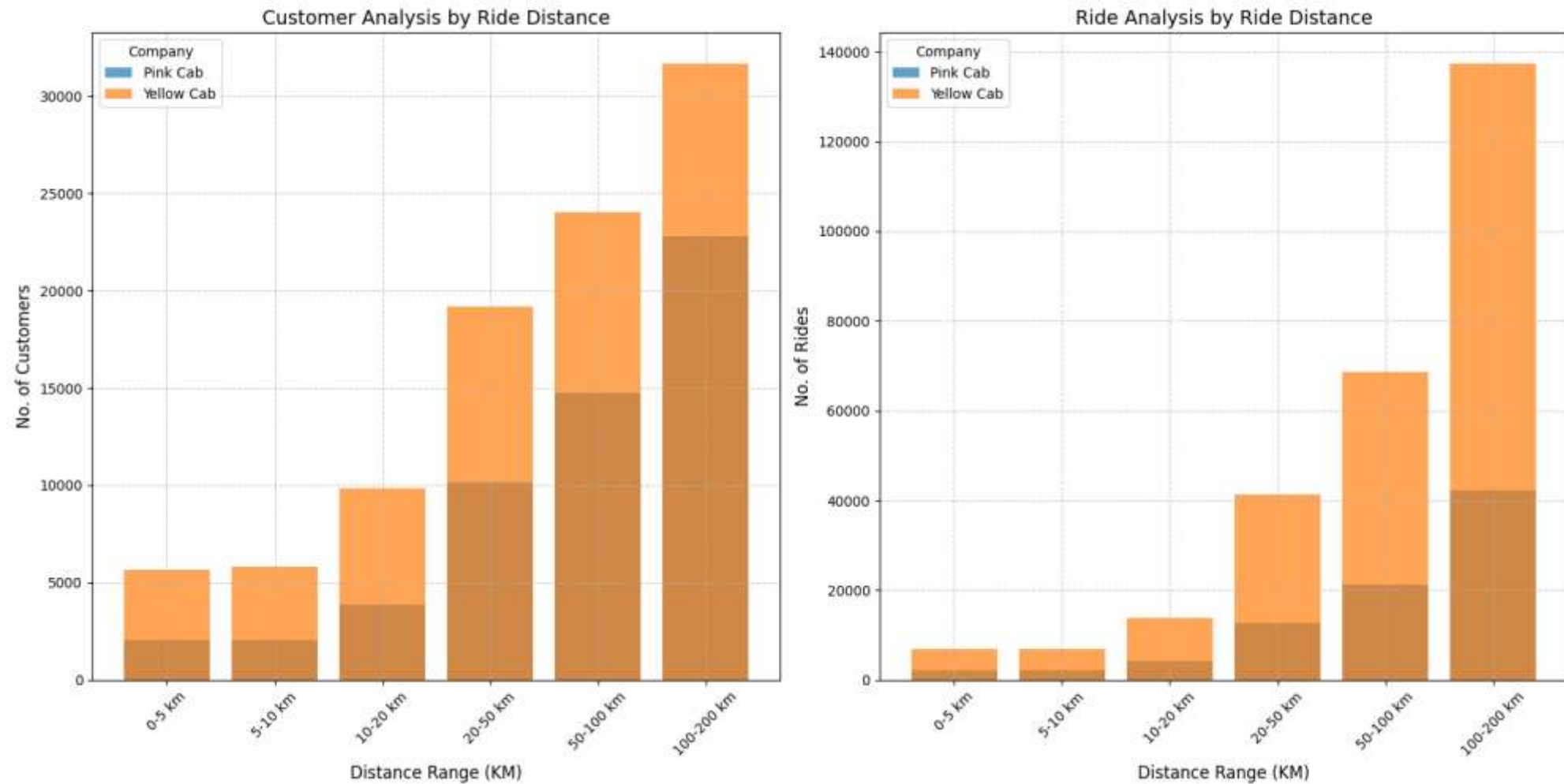


# Customer And Ride Analysis Day Wise



- Uniform Customer Distribution
- Higher Demand for Yellow Cab

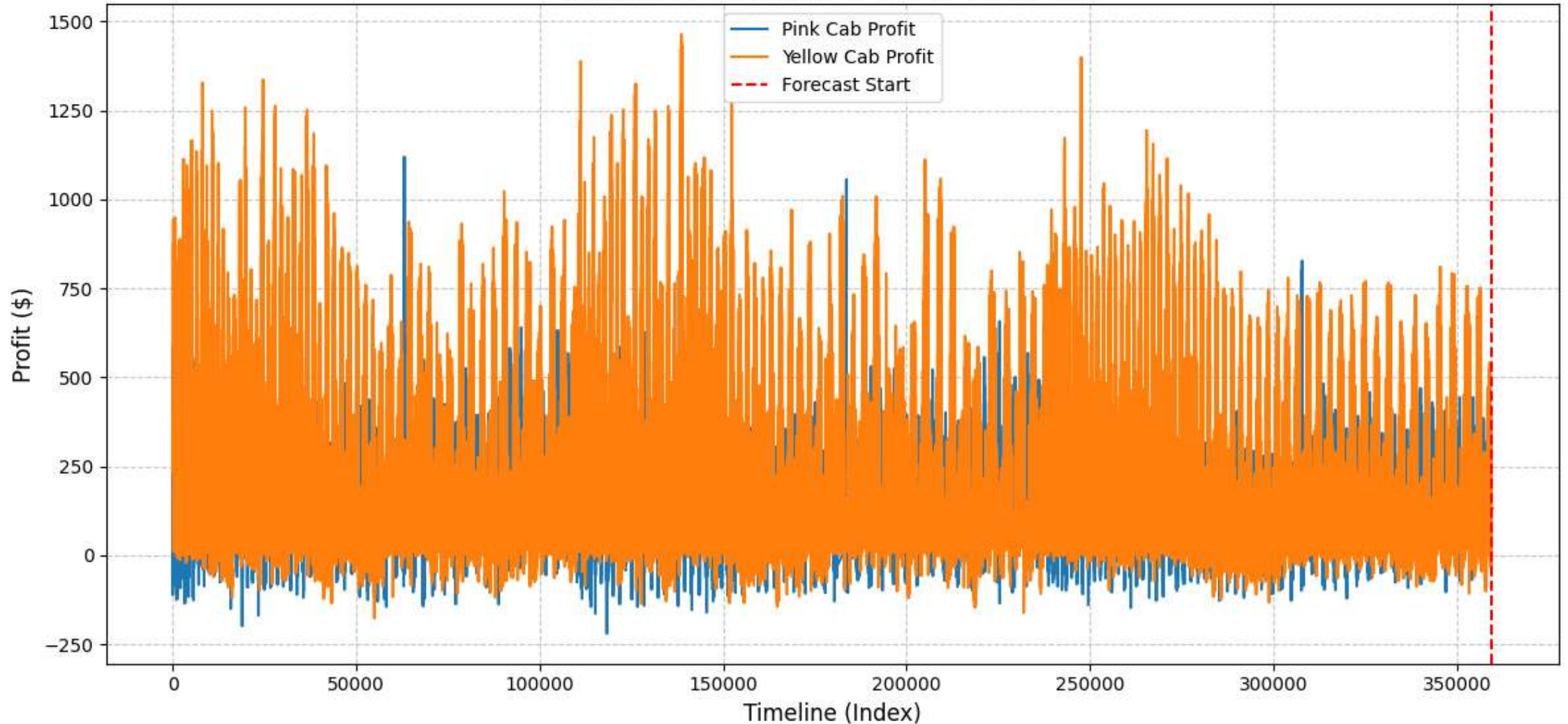
# Customer analysis based on ride distance



- Short Rides Lead:** Most customers and rides are in the **0-10 km range**.
- Yellow Cab Dominance:** Yellow Cab outperforms Pink Cab across all distance ranges.
- Low Long-Distance Demand:** Fewer rides occur in the **50+ km range** for both companies.
- Retention Opportunity:** Short-distance customers offer potential for loyalty programs.



Profit Forecasting for Each Company



- **Yellow Cab** shows higher historical profits and a steeper growth trajectory.
- **Both companies** have positive profit trends, but Yellow Cab's performance is more robust.
- **Yellow Cab** is the better investment due to its consistent growth and market dominance.

## Recommendations

Based on the analysis conducted, we have evaluated both Yellow Cab and Pink Cab on multiple parameters to identify the better investment opportunity. Below are the key points and recommendations:

### Customer Reach:

**Yellow Cab** has a broader customer reach across multiple cities, consistently outperforming Pink Cab in city-wise coverage. Pink Cab shows a stronger presence in a few cities but lacks Yellow Cab's extensive reach.

### Customer Retention:

Yellow Cab demonstrates higher retention rates, with more customers using their services repeatedly (5+ and 10+ rides). Pink Cab struggles in retaining its customer base compared to Yellow Cab.

### Age-Wise Reach:

Yellow Cab serves a more diverse age group, performing well across all age categories, including 18-25 and 60+ age groups. Pink Cab shows limited appeal across broader demographics.

### Average Profit Per KM:

Yellow Cab's average profit per KM is **substantially higher** (almost three times) compared to Pink Cab, showcasing better profitability per ride.

### Income-Wise Reach:

Both companies perform well in high and medium income segments, but Yellow Cab leads in reaching all income groups, including the low-income class.

### Ride Count and Profit Forecasting:

Yellow Cab's forecasted profit trends indicate **lower losses** compared to Pink Cab, which faces steeper declines in both ride count and profitability. Yellow Cab's resilience in financial forecasting suggests better long-term scalability.

### Final Recommendation:

Based on customer reach, retention, age and income demographics, profitability per ride, and future forecasting:

**Yellow Cab** is the recommended investment option due to its consistent performance, broader market coverage, higher retention, and stronger financial outlook.

# Nitish Jena

# Thank You