



**Data Glacier**

Your Deep Learning Partner

# Exploratory Data Analysis

**G2M insight for Cab Investment firm**

**Nimanthi Yaseema**

**July 20, 2025**

# Agenda

Background

Data Exploration

EDA

EDA Summary

Recommendations

# Background

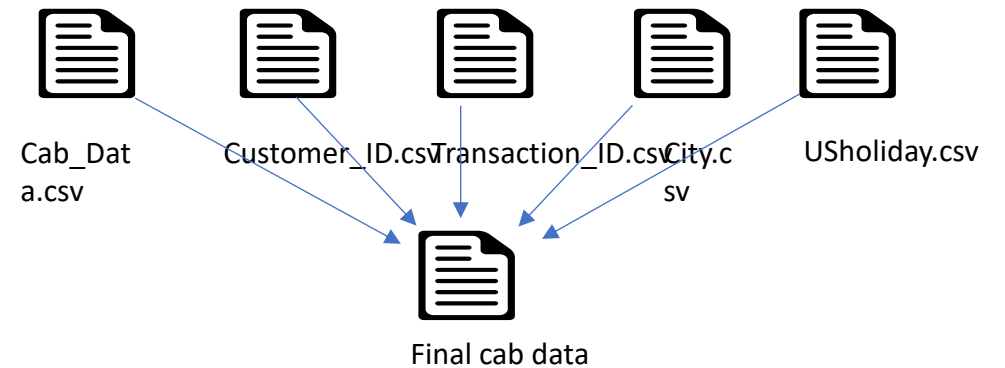
- **XYZ** is a private equity firm in the US. Due to remarkable growth in the **Cab Industry** in the last few years and the presence of multiple key players in the market, XYZ is planning an investment in the cab industry.
- **Objective** : Provide **actionable insights** to help XYZ identify the right company for investment.
- **Analysis Overview**:
  1. Data Understanding
  2. Forecasting profit and number of rides for each cab type
  3. Finding the most profitable Cab company
  4. Recommendations for investment

# Data Exploration

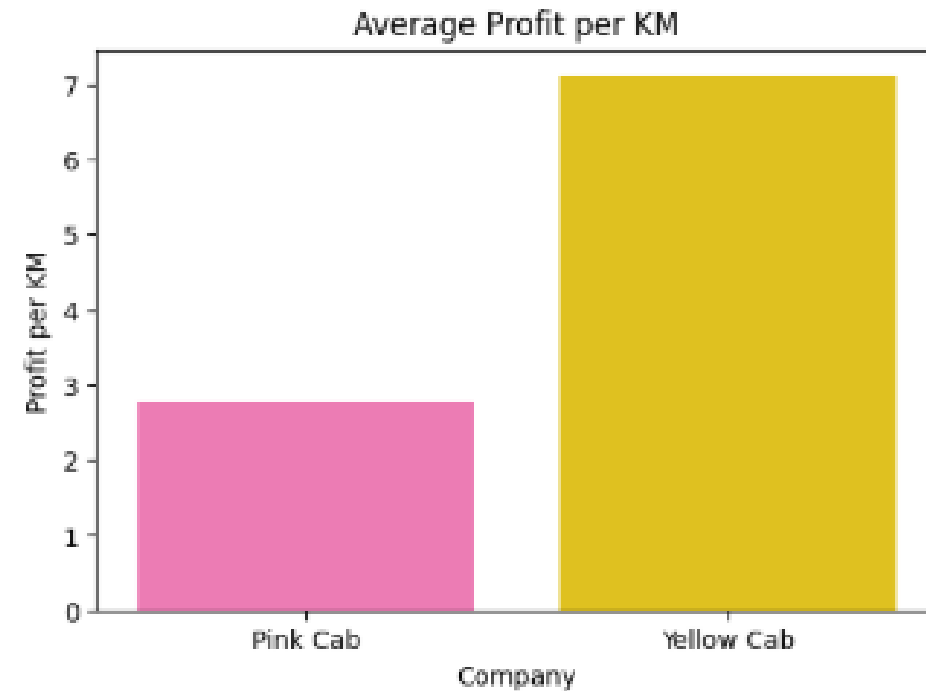
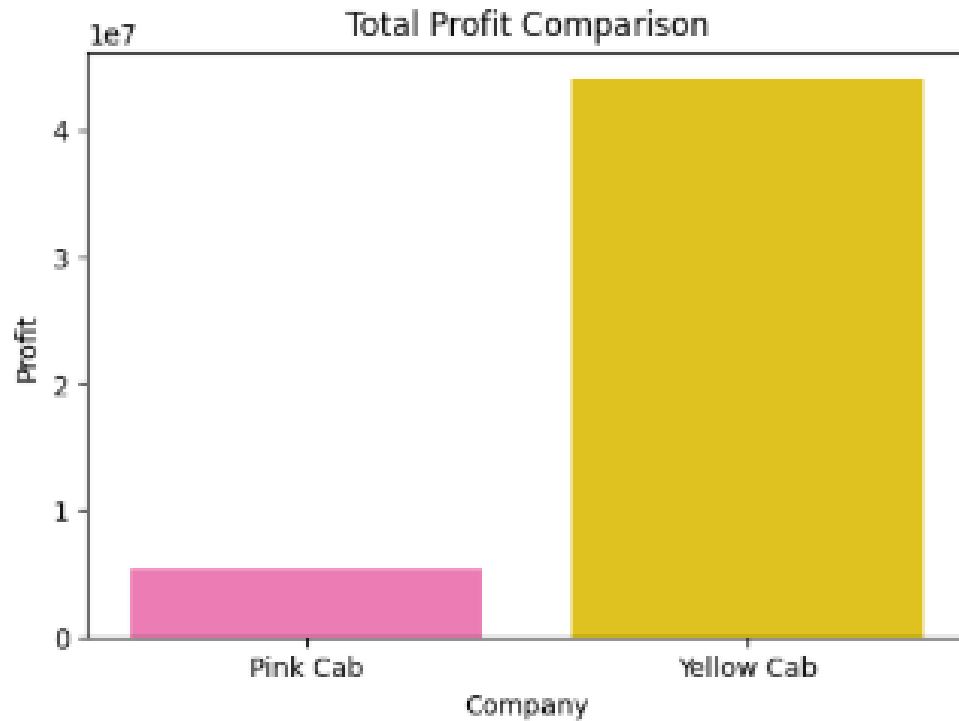
- **Features:** 24 total features, including **9 derived features**
- **Timeframe:** January 31, 2016 – December 31, 2018
- **Total data points :**355,032

## Assumptions:

- Outliers are present in the **Price\_Charged** feature, but due to the **lack of trip duration data**, they are not treated as outliers.
- **Profit** is calculated using only **Price\_Charged** and **Cost\_of\_Trip**, assuming other factors remain constant.
- The **Users** feature in the city dataset is treated as the total number of cab users in the city — assumed to include **Yellow Cab, Pink Cab, and other cab users**.

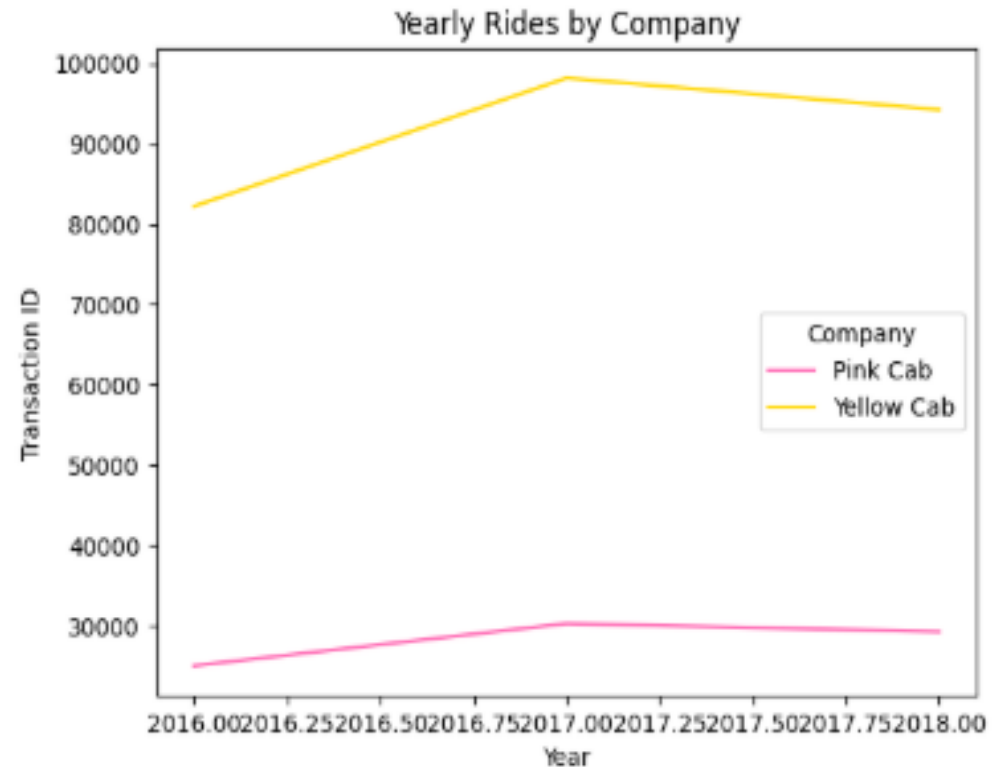
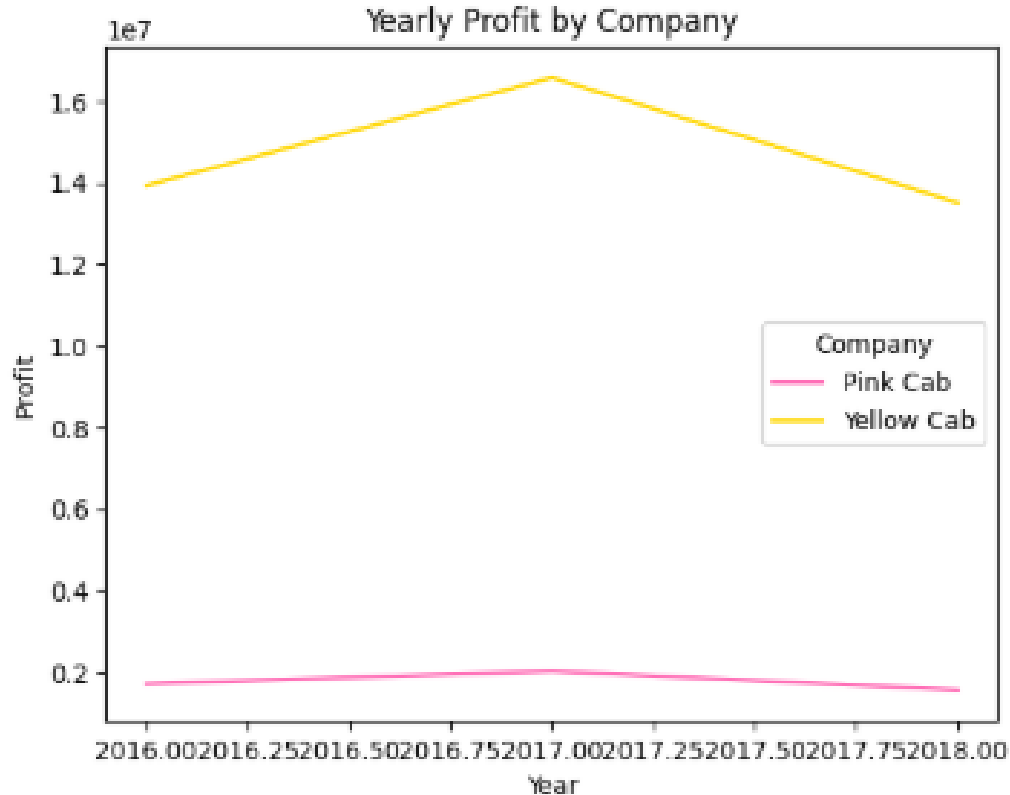


# Profit Analysis



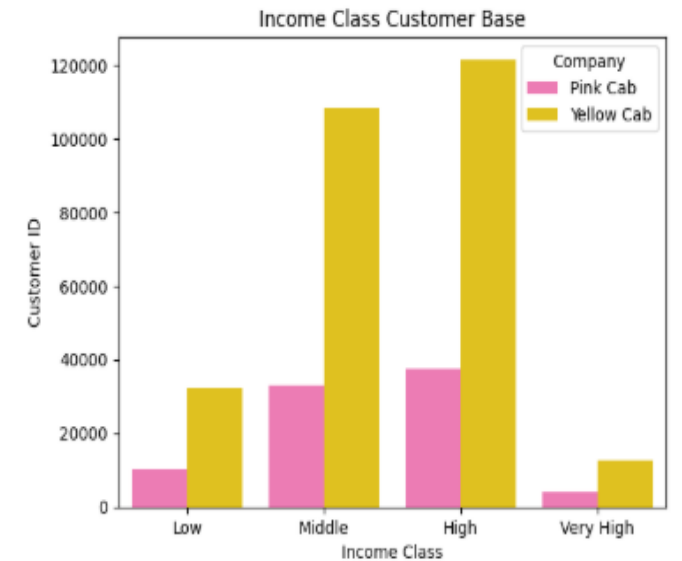
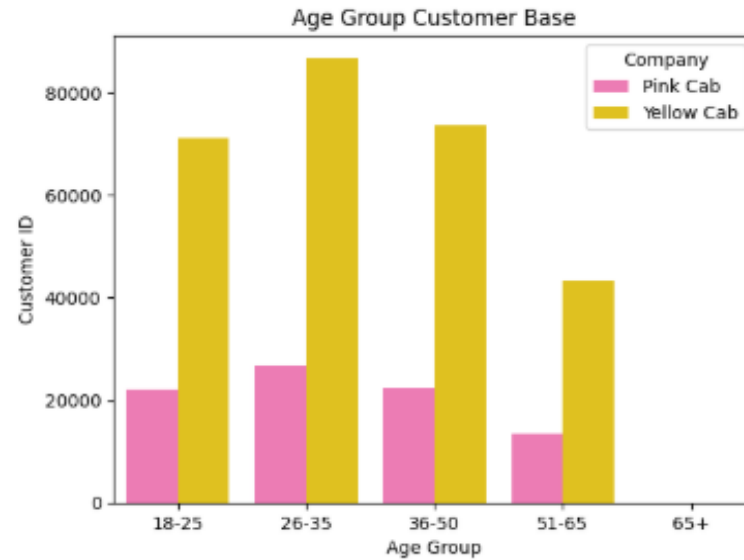
The Yellow Cab company shows higher total profit and higher average profit per kilometer compared to the Pink Cab company.

# Yearly Profit & Ride Analysis

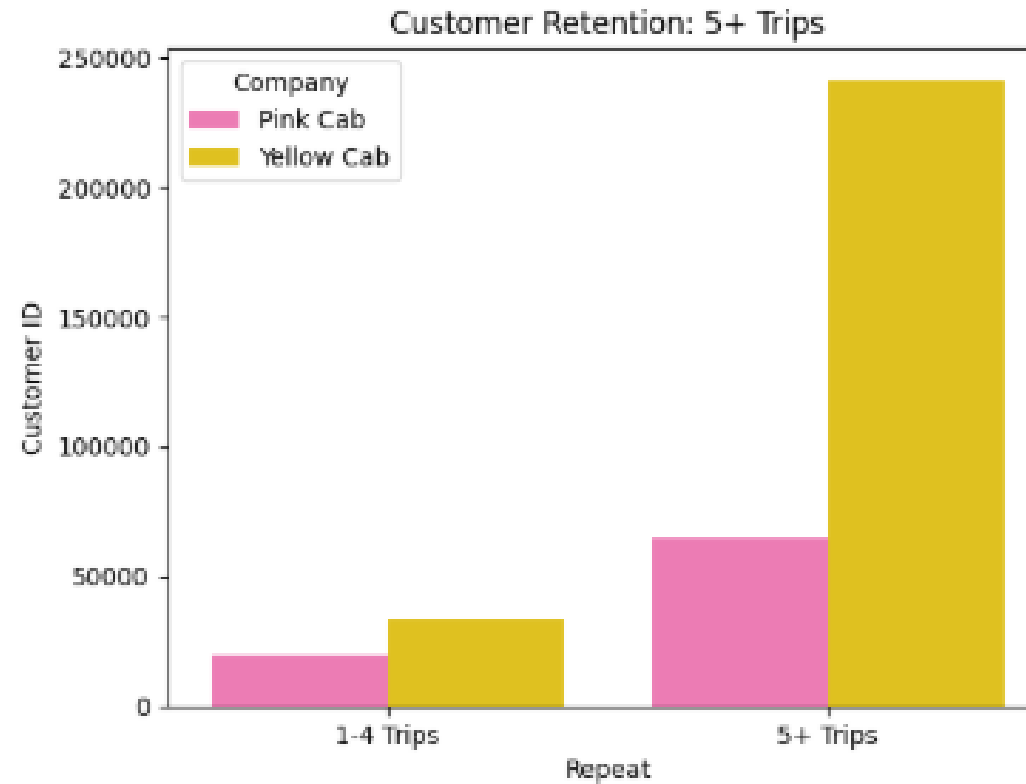


The Yellow Cab company had higher annual profit and a greater number of rides in 2016, 2017, and 2018 compared to the Pink Cab company.

# Customer Segmentation: Gender, Age, Income



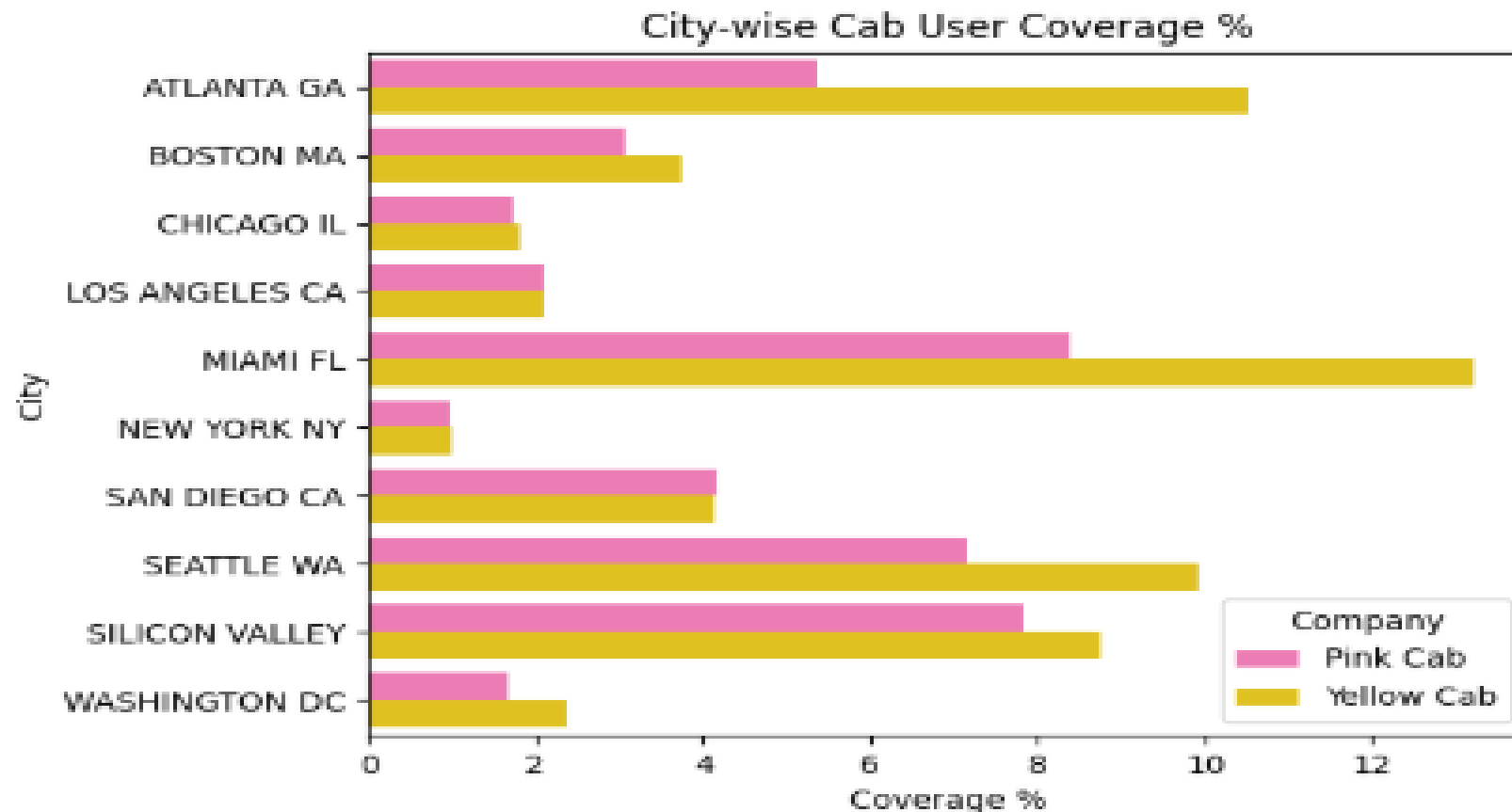
## Customer Retention



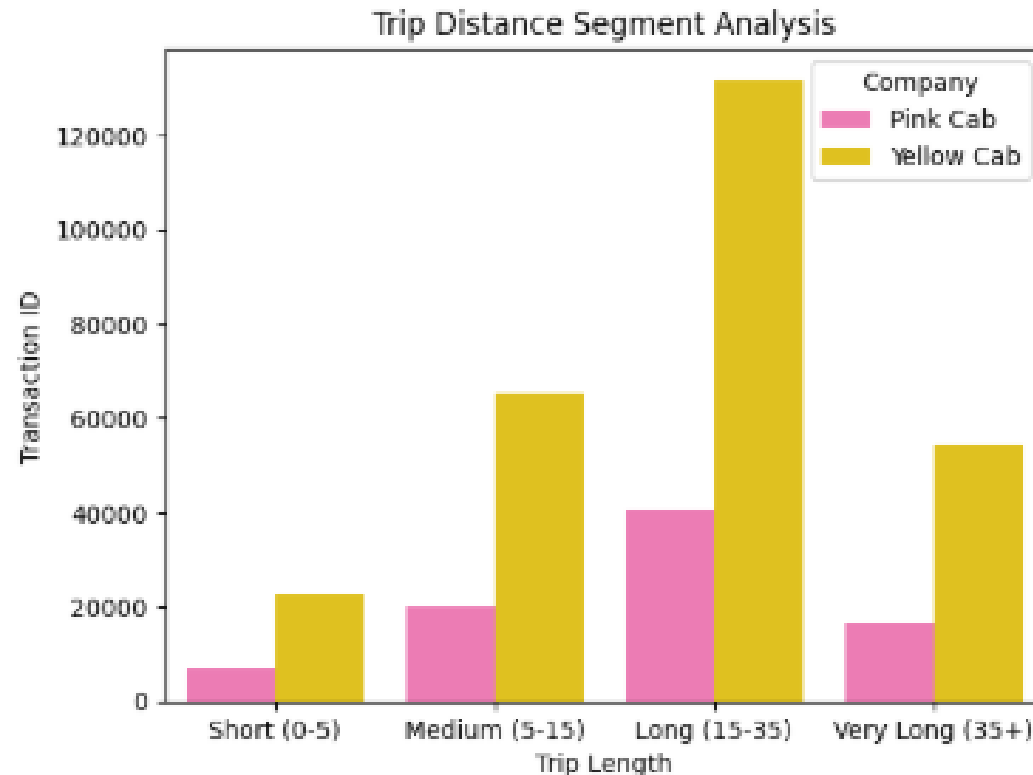
More customers take 5+ trips than 1–4 trips, and Yellow Cab outperforms Pink Cab in both retention segments.



# City-Wise Reach

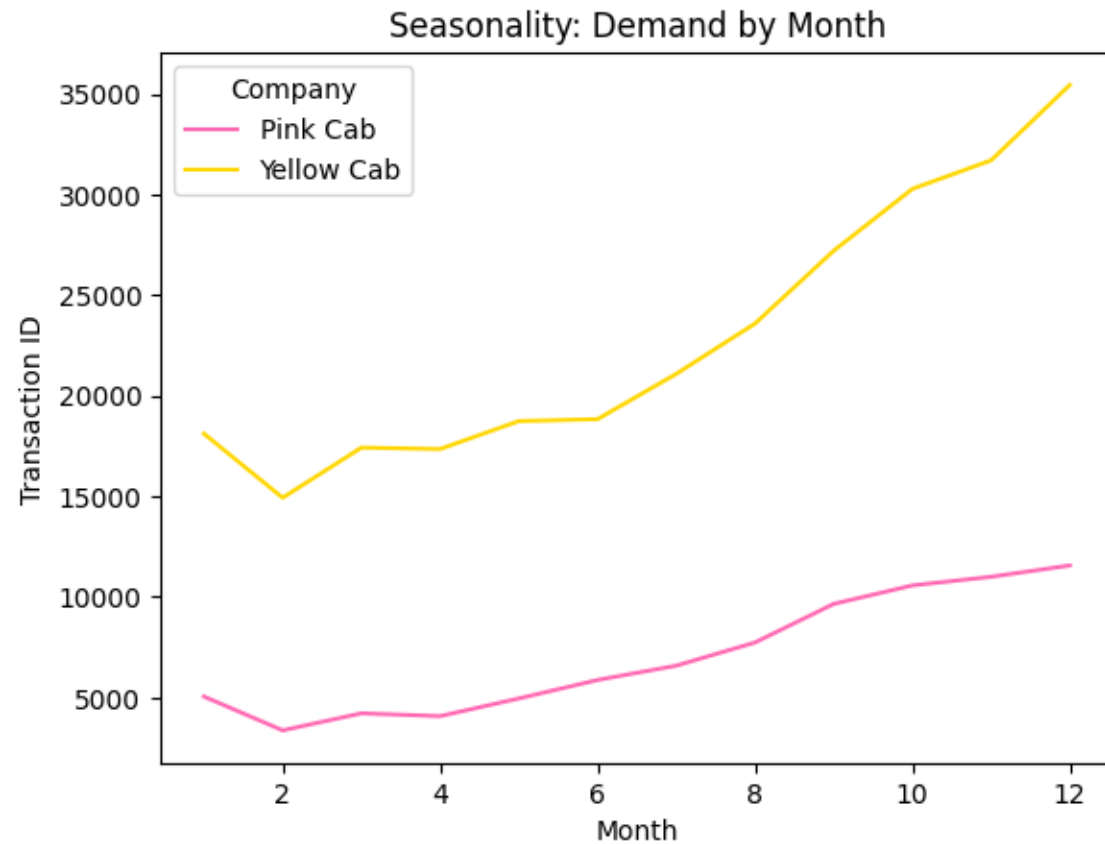


# Ride Distance Preference



In the trip distance segment analysis, long trips (15–35 km) account for the highest share, while short trips (0–5 km) are the smallest segment. Across all trip distance segments, Yellow Cab serves more trips than Pink Cab.

# Seasonality



In the seasonality demand analysis by month, the number of cab users increases steadily from month 0 to 12, with Yellow Cab consistently having higher demand than Pink Cab throughout the year.

# Recommendations

Overall, Yellow Cab scores higher than Pink Cab on all key factors evaluated.

- **Profit Analysis** : The Yellow Cab company shows higher total profit and higher average profit per kilometer compared to the Pink Cab company.
- **Yearly Profit and Rides Analysis:** The Yellow Cab company had higher annual profit and a greater number of rides in 2016, 2017, and 2018 compared to the Pink Cab company.
- **Customer Segmentation** : Customer segmentation by gender, age, and income shows that Yellow Cab serves a higher proportion of customers in all these segments compared to Pink Cab.
- **Customer Retention:** Customer retention analysis (5+ trips) shows that there are more customers taking over 5 trips compared to those taking only 1–4 trips. In both cases, Yellow Cab consistently retains more customers than Pink Cab.
- **City-wise reach analysis:** In the city-wise reach analysis, more people use Yellow Cabs than Pink Cabs overall. However, in Los Angeles, New York, San Diego, and Chicago, both companies have almost the same market share.
- **Ride distance preference:** Long trips (15–35 km) are the largest segment, short trips (0–5 km) are the smallest, and Yellow Cab leads Pink Cab in every distance segment.
- **Seasonality:** Monthly demand rises steadily, and Yellow Cab usage is higher than Pink Cab every month.

Based on the above points, we recommend investing in Yellow Cab.

# Thank You