G2M Insight for Cab Investment Firm

Virtual Internship

Marina Tsvetkova

13-May-2024

Agenda

- **Problem Statement**

- **Executive Summary**

- **Approach**

- **Exploratory Data Analysis (EDA)**

- **EDA Summary**

- **Hypothesis tests**

- **Recommendations**

Data Glacier
Your Deep Learning Partner

Problem Statement

XYZ is a private equity 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.

• Objective: Provide actionable insights to help XYZ firm in identifying the right company for making investment.

Data Information

The dataset contains four different csv files. These datasets were combined in the 'merged_df'.

Cab_Data.csv – this file includes details of transaction for 2 cab companies

Customer_ID.csv – this is a mapping table that contains a unique identifier which links the customer's demographic details

Transaction_ID.csv – this is a mapping table that contains transaction to customer mapping and payment mode

City.csv – this file contains list of US cities, their population and number of cab users

merged_df: This dataset is a combination of 4 files. The dataset is combined based on the "Customer ID" and "Transaction ID" features. Also, new features such as profit, state have been added to the dataset (16 features).

Executive Summary

Key Findings:

Seasonality: Cab usage shows notable seasonal trends, which can inform resource allocation and marketing timing.

Company Preference: Yellow Cab and Pink Cab have different peak usage periods, suggesting varying customer preferences or market strategies.

Margin Dynamics: Increasing customer numbers have a slight positive effect on profit margins.

Age and Income: Both companies attract customers with similar age and income profiles.

Gender: A distinct difference in gender preferences presents opportunities for targeted marketing (Men use the service more)

Approach

- Data Preparation: Merged datasets, cleaned data, and handled duplicates/missing values to ensure data integrity.
- Analysis Steps: EDA to explore patterns, followed by hypothesis tests to confirm findings.

Exploratory Data Analysis

Here, the total transactions by cab companies and gender is shown. More transactions for the company Yellow Cab.Men use it more often than women.
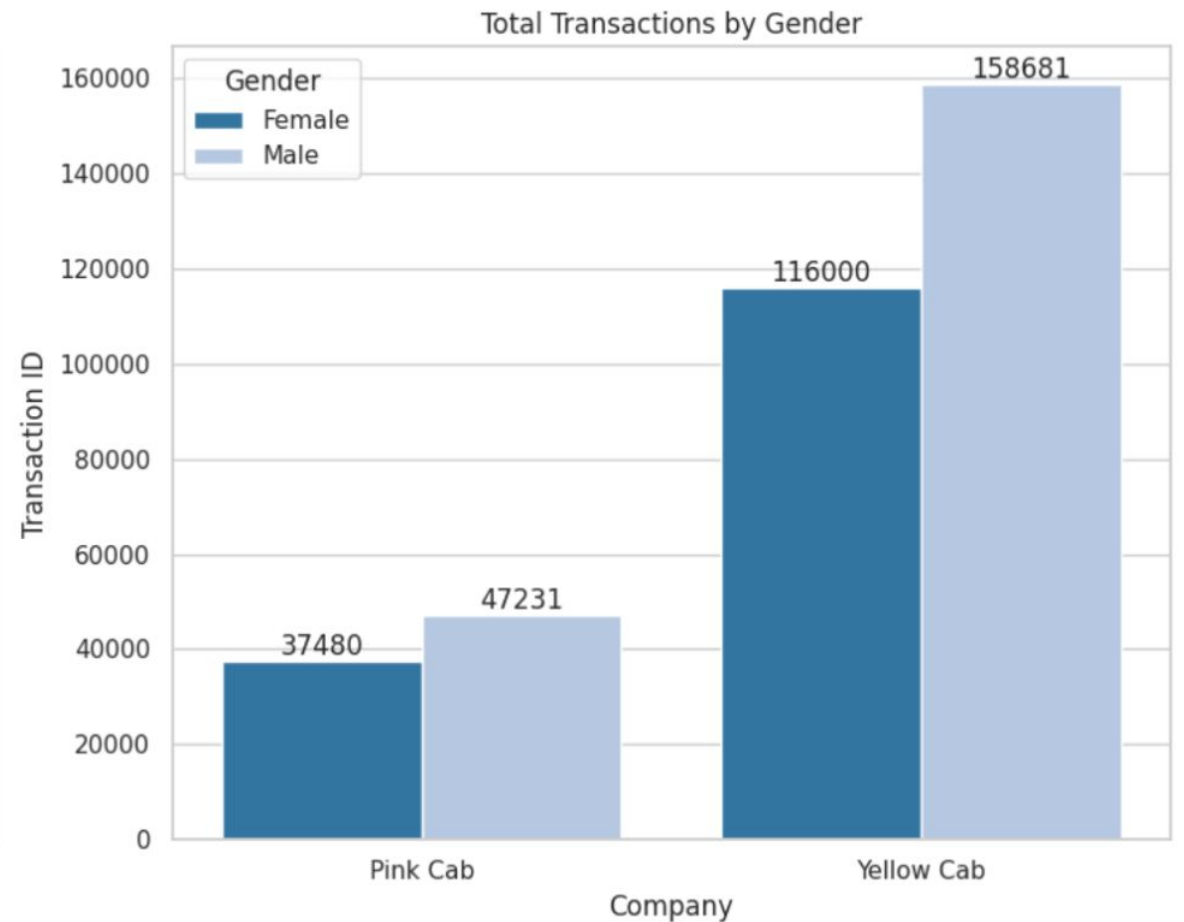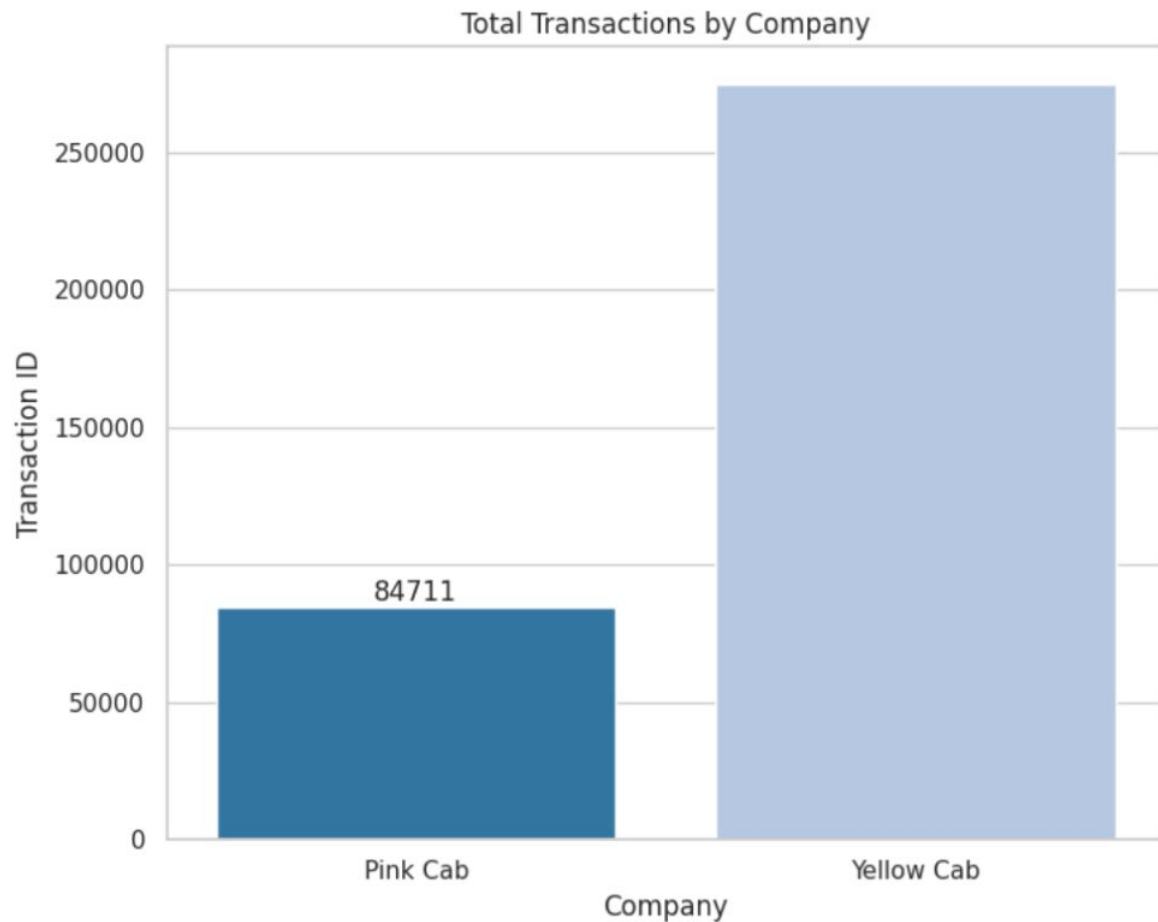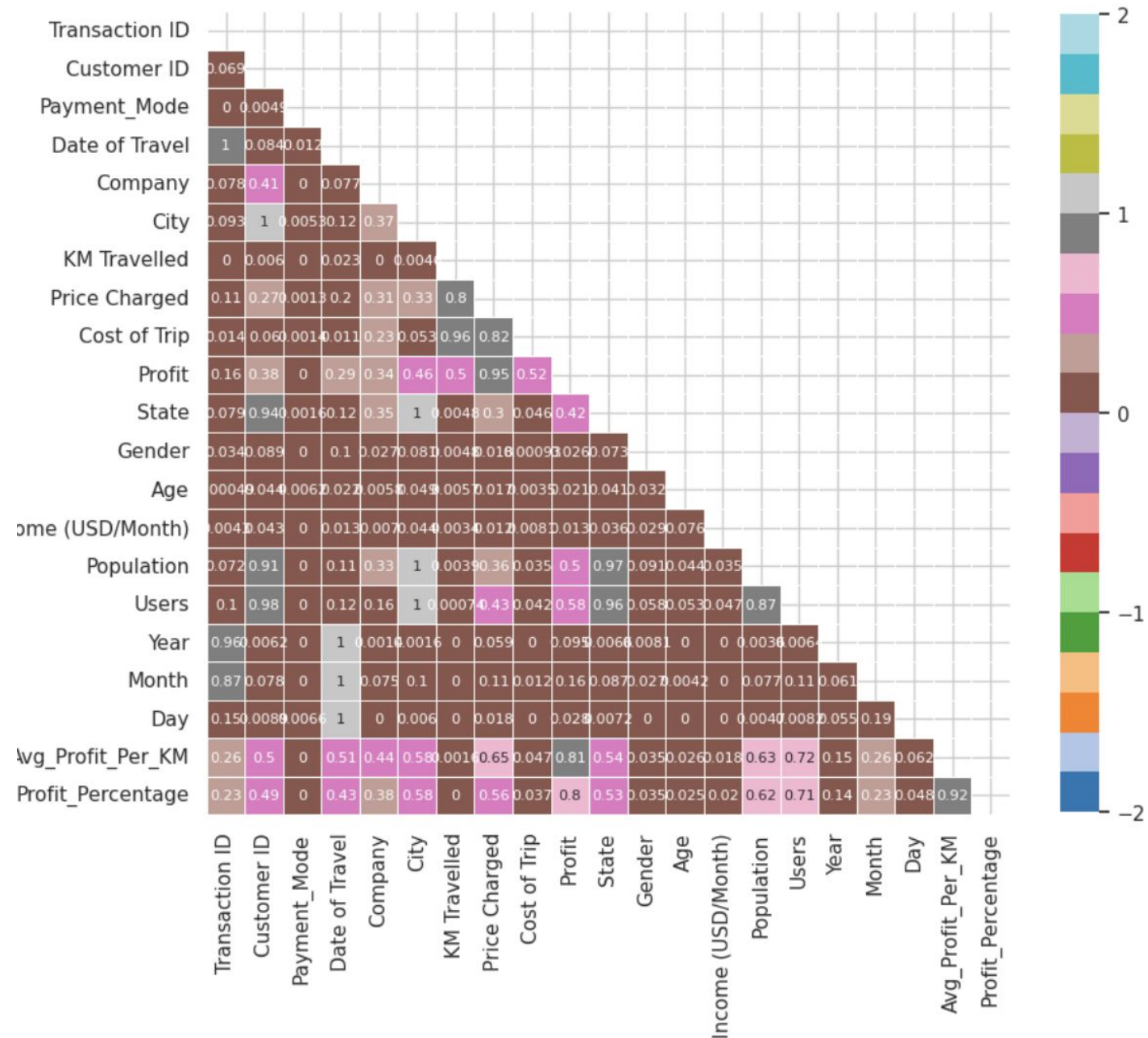
Exploratory Data Analysis

Here, the total transaction by payment mode and total profit by company is shown. Yellow Cab brings in more income than Pink
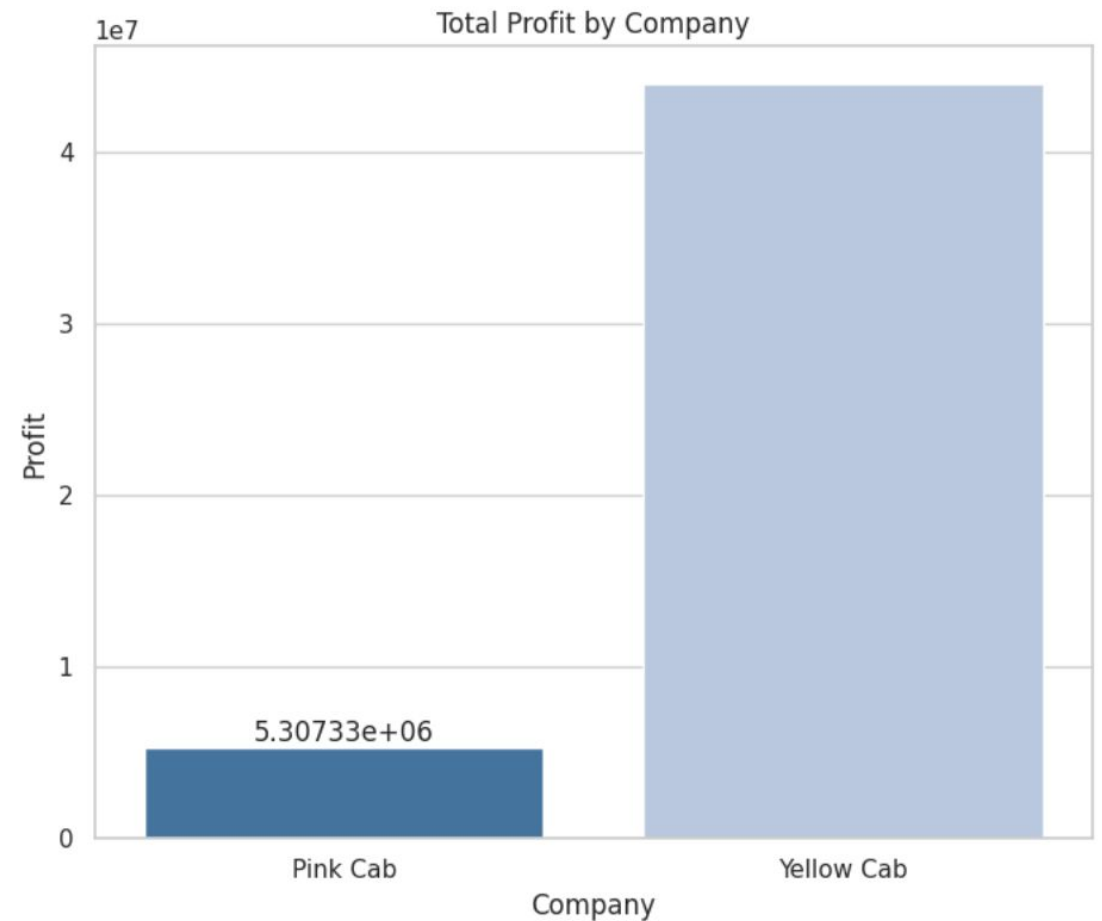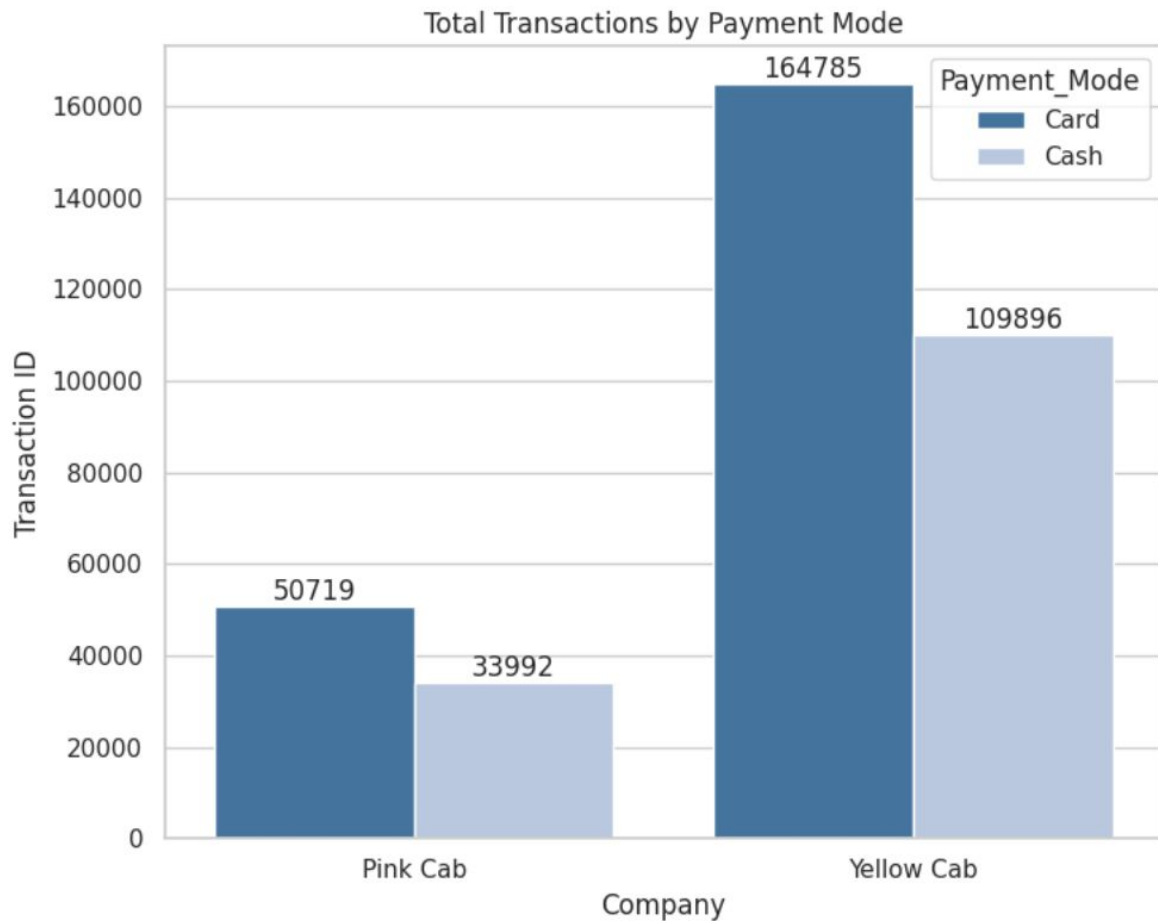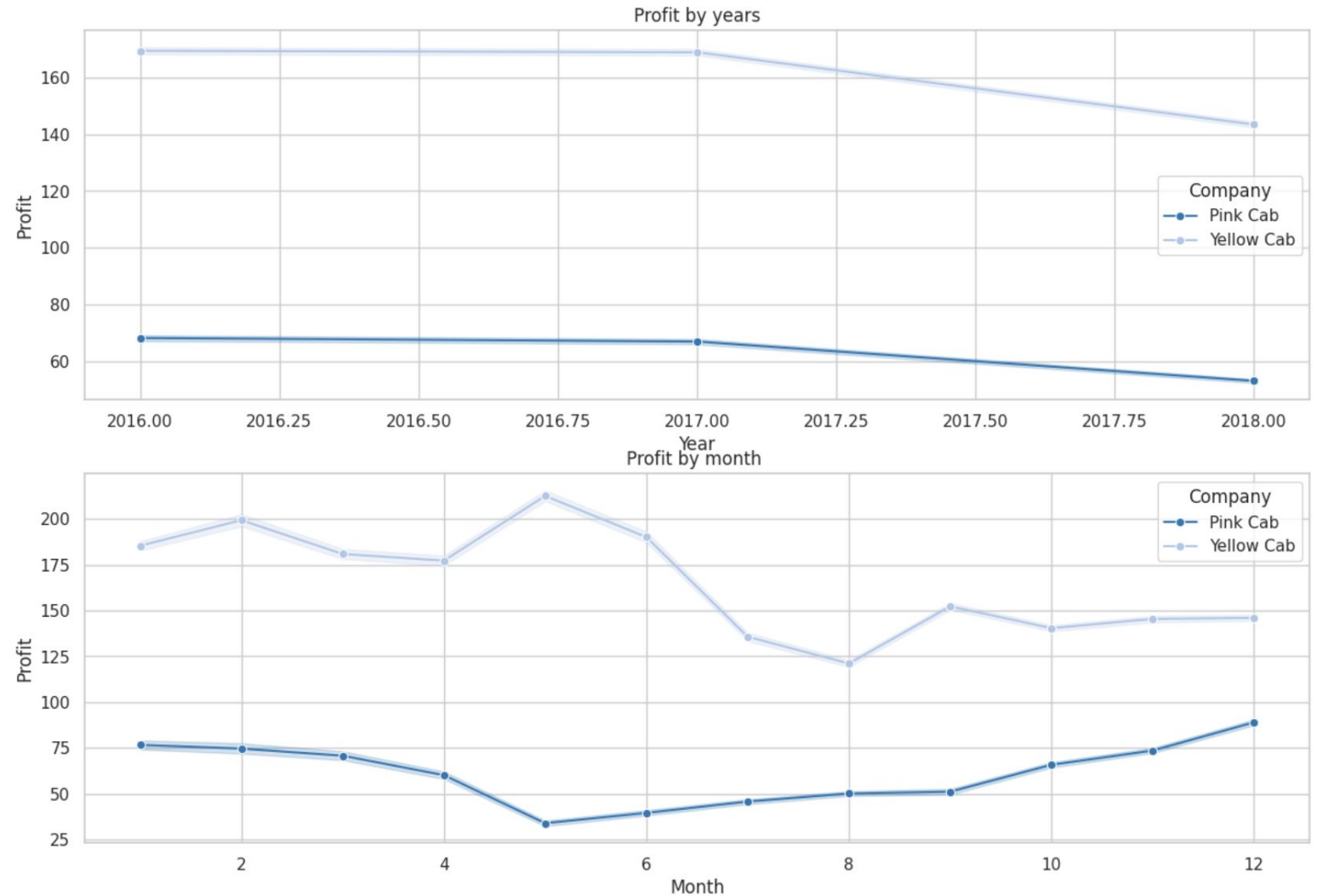
Exploratory Data Analysis (EDA)

•Price Charged (0.951922 with Profit) This indicates that increasing the price charged significantly boosts profitability.

•Average Profit Per KM (0.813103 with Profit). The correlation with Profit suggests that improving the average profit per kilometer is crucial for enhancing overall profitability.

•Profit Percentage (0.798361 with Profit). The correlation with Profit implies that higher profit margins directly contribute to increased total profit.
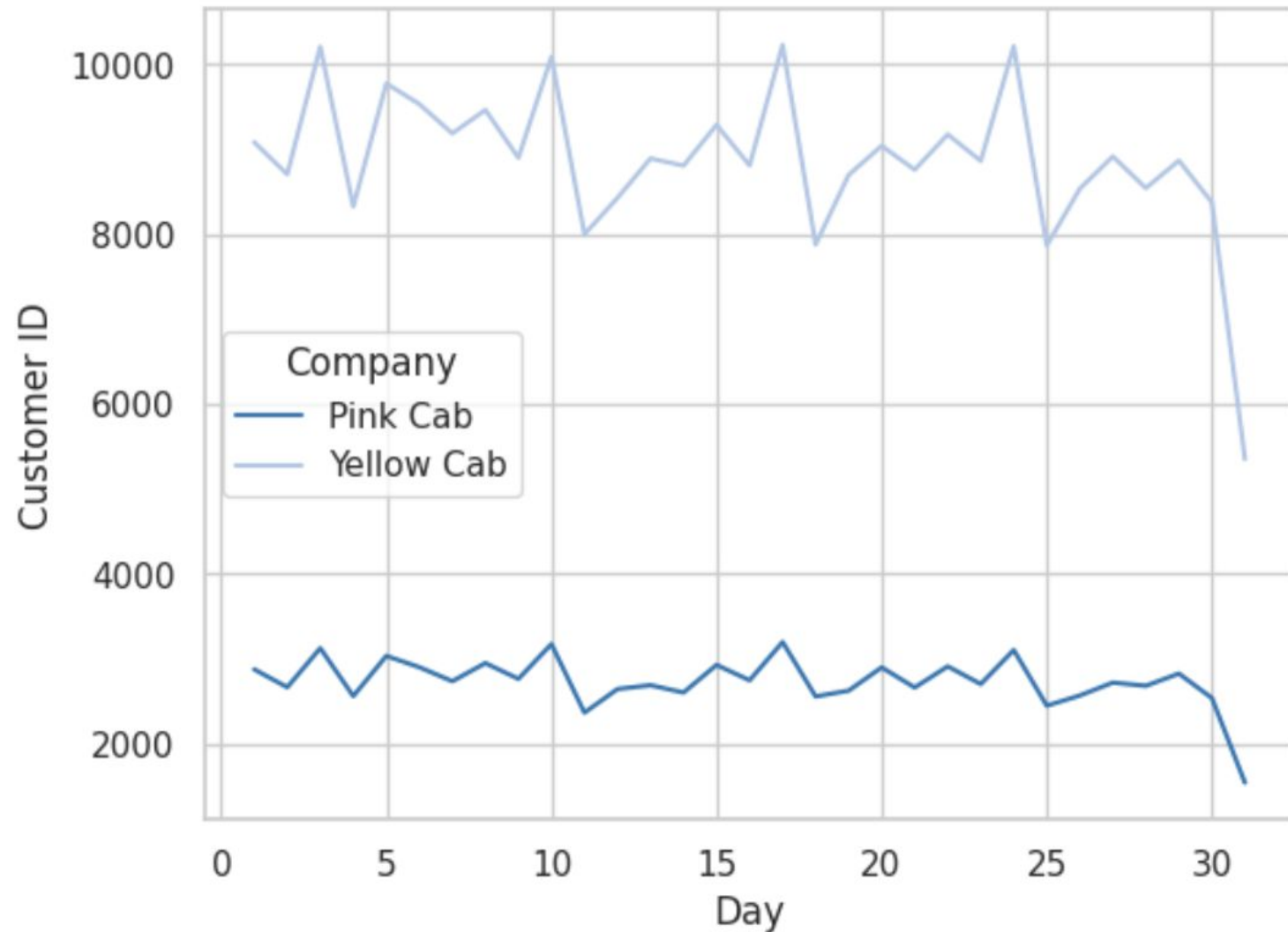
Exploratory Data Analysis (EDA)

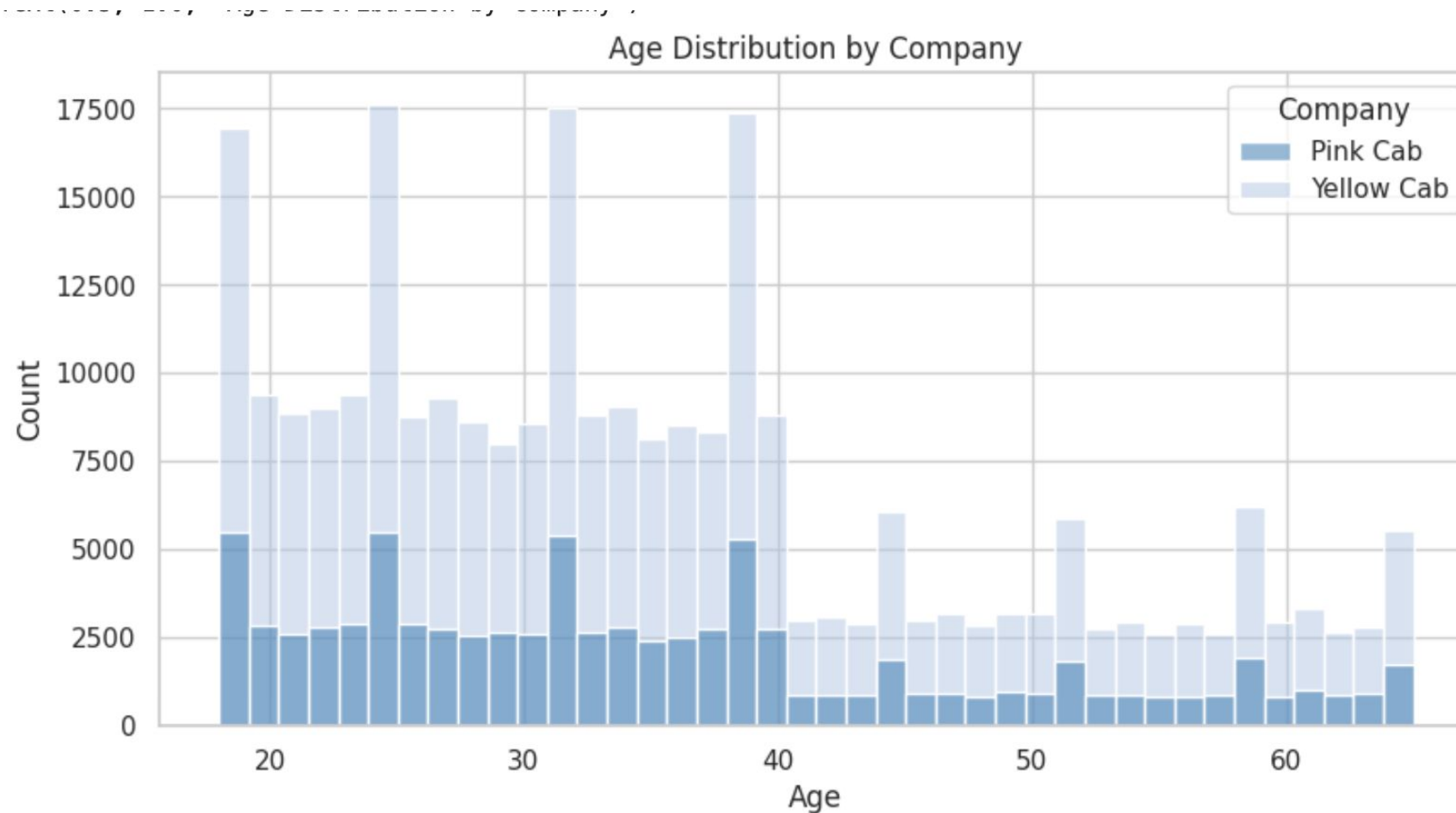- Profits of cab companies on a monthly and annual basis

Exploratory Data Analysis

It is seen that the demand is higher on the third day of the week for both companies.
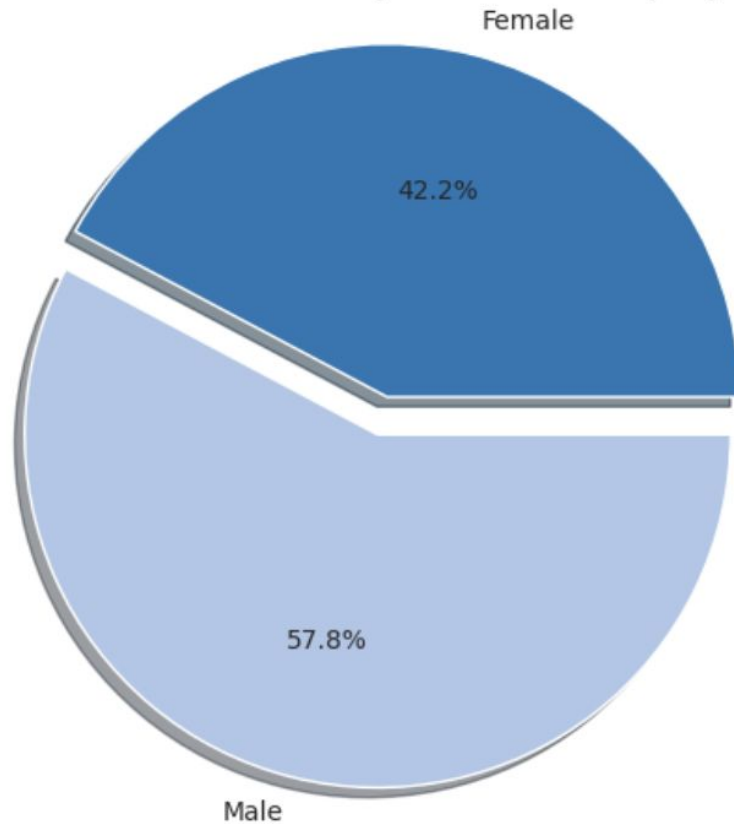
Exploratory Data Analysis

It is seen that the customer age distribution in both companies is between 20 and 40.

Exploratory Data Analysis

Gender distribution by cab company.

Gender Distribution by Yellow Cab Company
Female
42.2%
57.8%
Male

Gender Distribution by Pink Cab Company
Female
44.2%
55.8%
Male
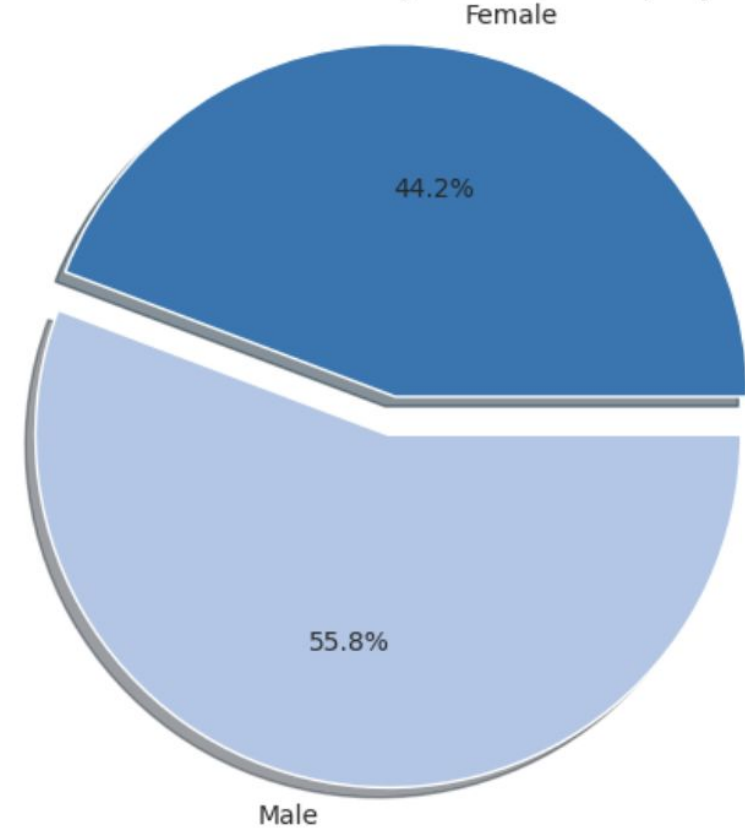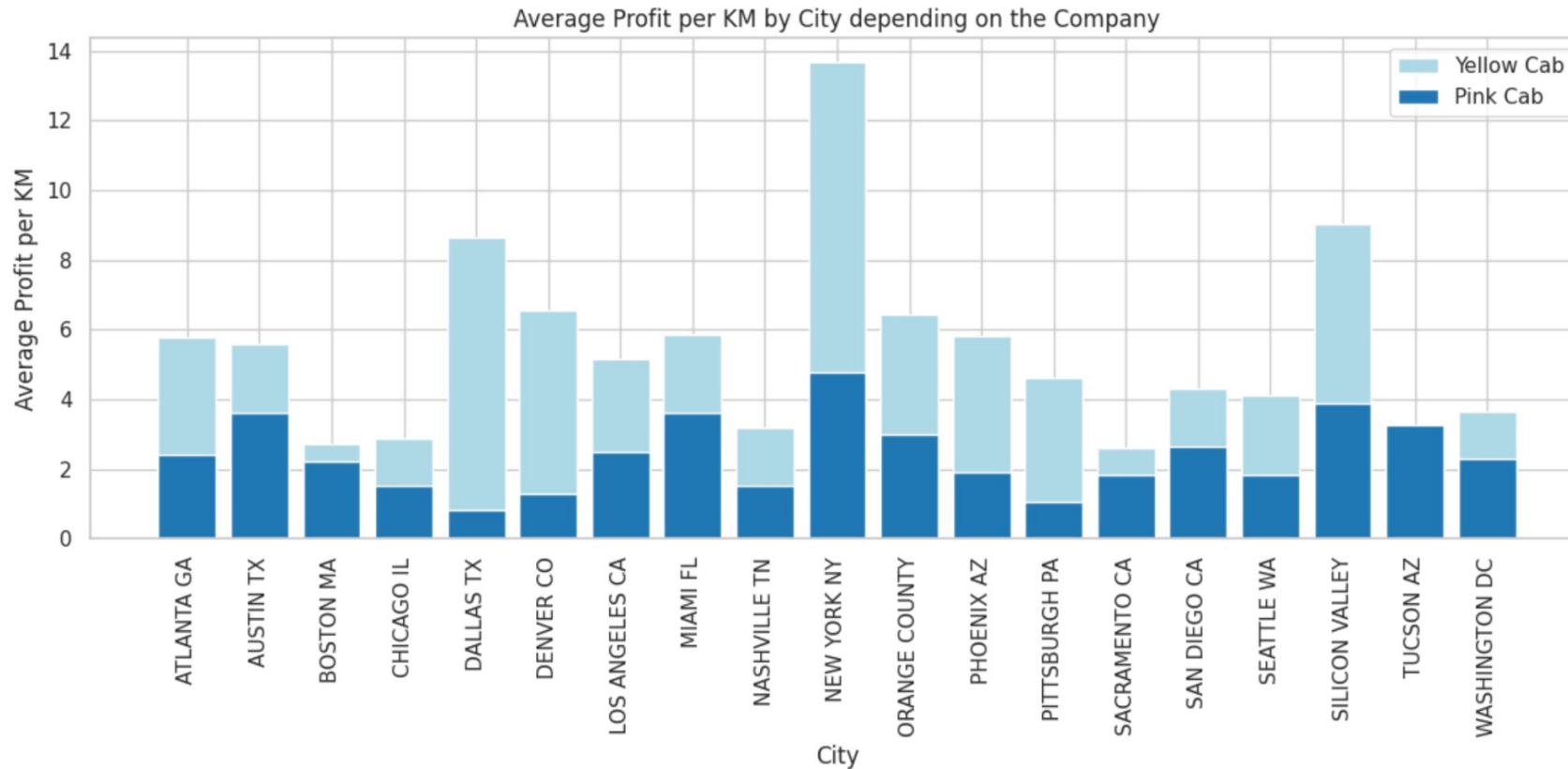
Exploratory Data Analysis

Profit in Cities by Company. New York makes the highest profit for both companies
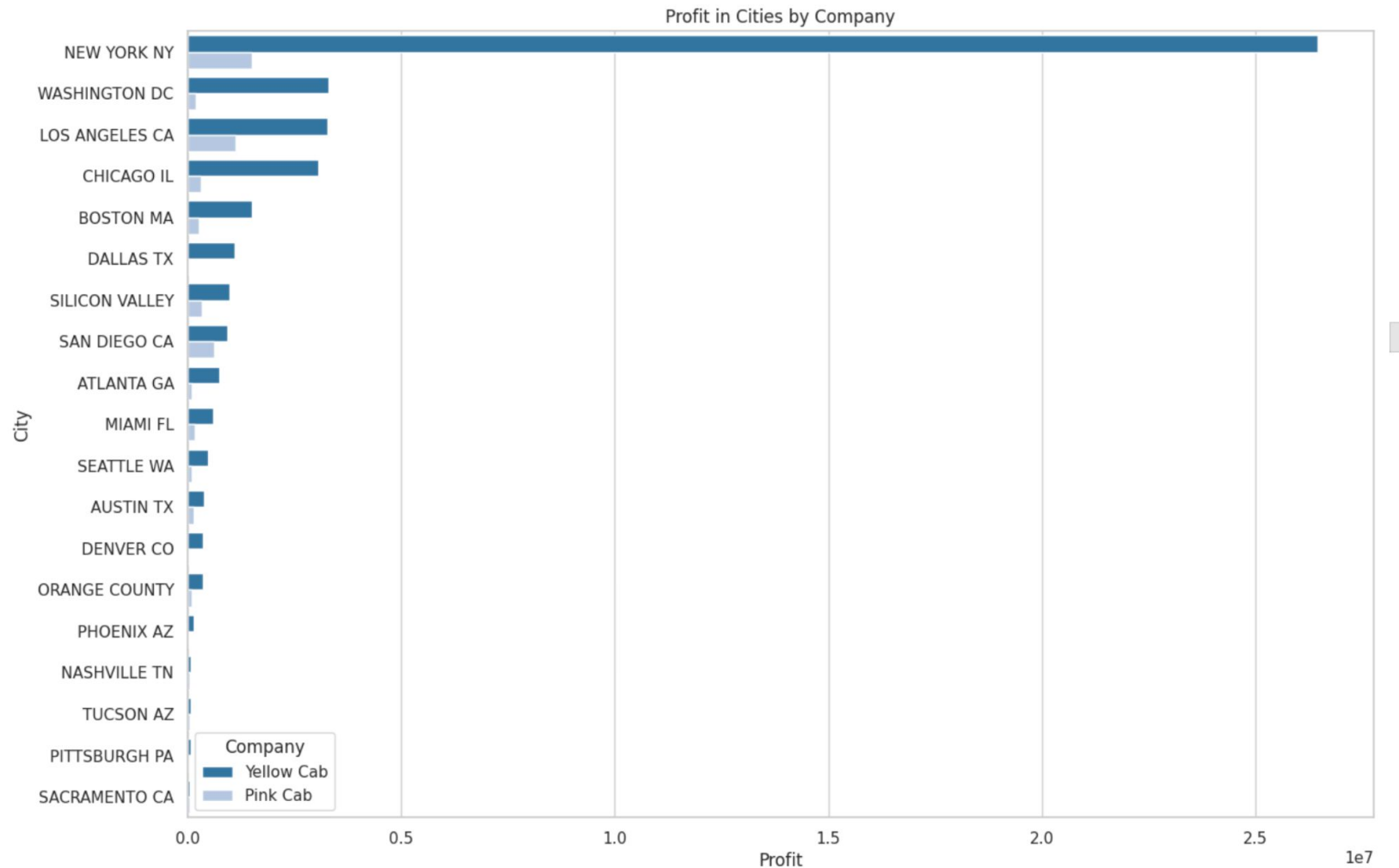
Exploratory Data Analysis

Average Profit by City depending on the Company.

Average Profit per KM by City depending on the Company

Exploratory Data Analysis (EDA)

- Income distribution by state. New York and California - the highest total income.

Sum Income by State

CA 23.2%
NY 27.8%
AZ 1.1%
WA 2.2%
TX 3.3%
TN 0.8%
PA 0.4%
MA 8.3%
IL 15.8%
GA 2.1%
FL 1.8%
DC 12.2%
CO 1.1%

Data Glacier
Your Deep Learning Partner
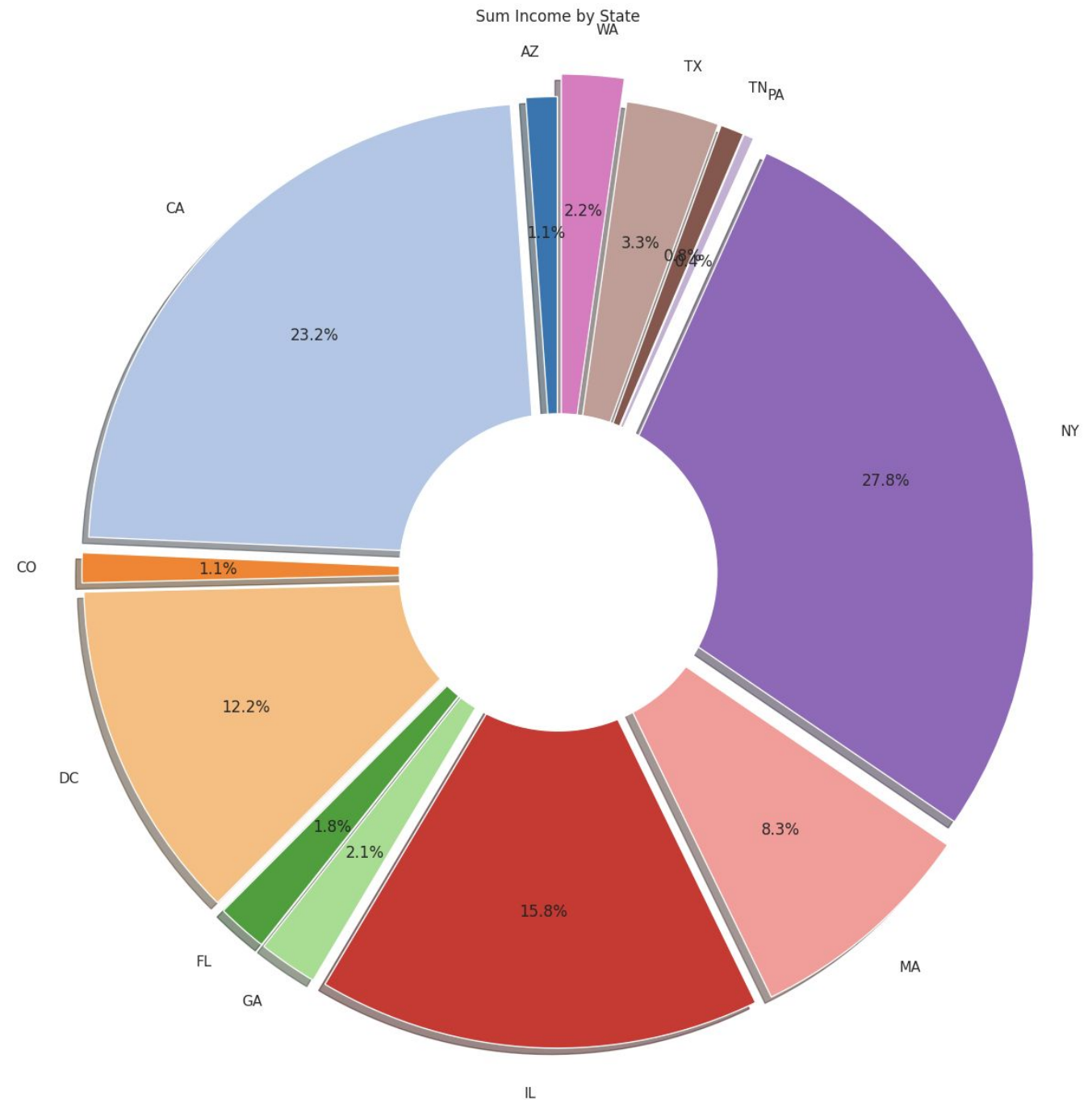
Exploratory Data Analysis (EDA)

Seasonal pattern 4 quarter cycle by Profit. The highest profit in last quarter for 2016, 2017,2018
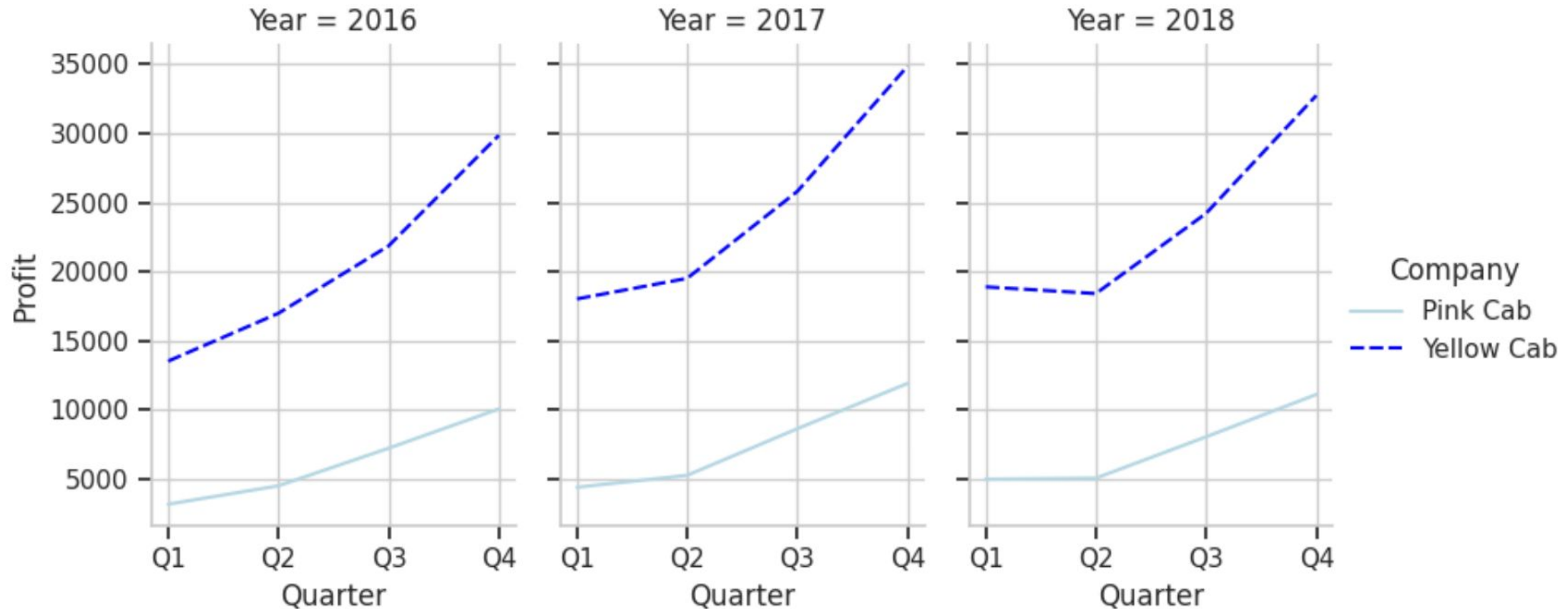
Hypothesis tests

1. **Seasonality in Cab Usage**:
   - Is there seasonality in the number of customers using the cab service?
   - **Result**: significant seasonality in cab usage, indicating that customer numbers vary notably across different months.
2. **Company Preference by Time Period**:
   - Which company has maximum cab users at a particular time period?
   - **Result**: Analysis revealed a significant difference in the number of daily cab users between Yellow Cab and Pink Cab.
3. **Margin and Number of Customers Relationship**:
   - Does the margin proportionally increase with the increase in the number of customers?
   - **Result**: There is a significant but modest positive relationship between the number of customers and the margin, suggesting that as customer numbers increase, margins slightly improve.
4. **Customer Attributes (Age)**:
   - Do customer ages significantly differ between Yellow Cab and Pink Cab users?
   - **Result**: No significant difference was found in the age distribution of customers between the two companies.
5. **Customer Attributes (Income)**:
   - Do customer incomes significantly differ between Yellow Cab and Pink Cab users?
   - **Result**: No significant difference was found in the income levels of customers between the two companies.
6. **Customer Attributes (Gender)**:
   - Does the gender distribution significantly differ between Yellow Cab and Pink Cab users?
   - **Result**: A significant difference was identified in gender distribution, indicating distinct gender preferences for each company.

p_value <0.05

Recommendations

- **Seasonality Strategy**: Adjust marketing campaigns and operational resources to align with seasonal demand patterns.
- **Time-Based Promotions**: Tailor promotions to peak usage times specific to each company.
- **Leverage Margins**: Optimize operational efficiencies to further enhance margins as customer numbers grow.
- **Targeted Marketing**: Develop gender-specific marketing strategies to better appeal to the dominant customer segments of each company.
- **Service Customization**: Customize services to meet the needs and preferences of the identified customer segments.

Thank You