Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis
G2M insight for Cab Investment firm

19-10-2023

Agenda

Data Glacier

Your Deep Learning Partner

Executive Summary

- **Project Objective:** XYZ, a private investment firm in the US, seeks to invest in the Cab Industry. To inform their decision. The client is pursuing a Go-to-Market (G2M) strategy that requires understanding the market dynamics. The goal is to identify the right company for their investment based on actionable insights.

- **Data Sets Provided:** Four data sets spanning from 31/01/2016 to 31/12/2018 have been made available for analysis:

- **Hypothesis Generation:** Multiple hypotheses are to be formulated and investigated, with a suggested range of 5-7

Problem Statement

- XYZ, a private investment firm in the US, is facing the challenge of identifying the most promising investment opportunity within the dynamic Cab Industry.

- With the sector experiencing remarkable growth and hosting multiple key players, XYZ needs to navigate this complex landscape strategically.

Approach

Their Go-to-Market (G2M) approach demands actionable insights to ascertain the most suitable cab company for investment.

To address this challenge effectively, we must:

- Analyze multiple data sets

- Explore potential seasonality

- Evaluate customer segments, and discern relationships within the data.

- Our goal is to provide XYZ with a comprehensive understanding of the market, enabling them to make a well-informed and advantageous investment decision.
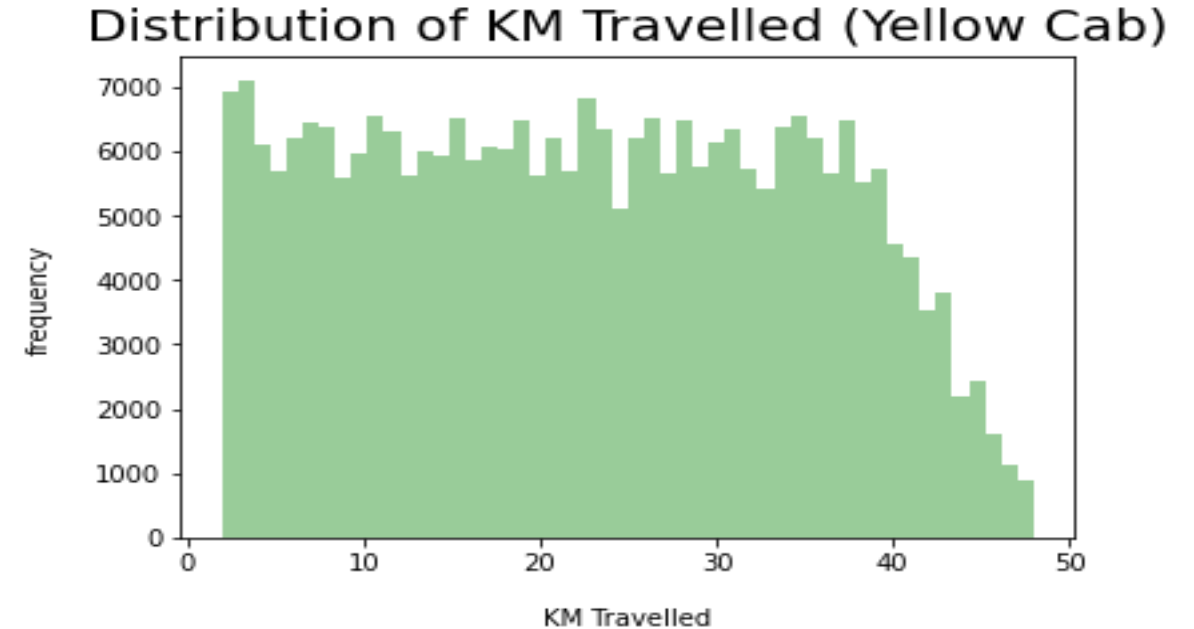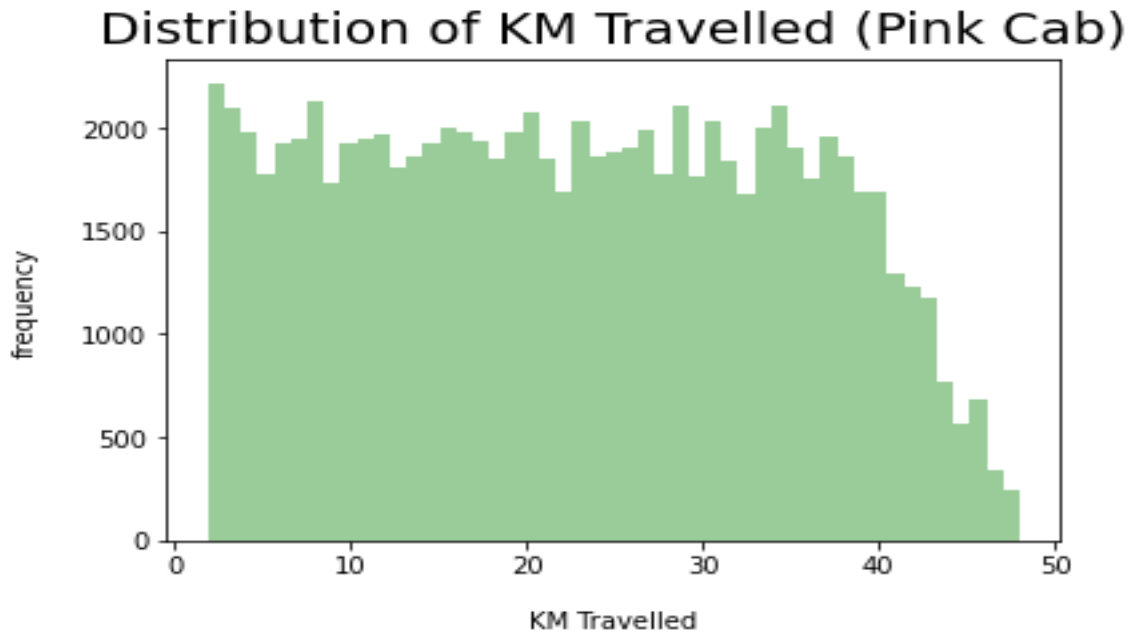
EDA: Exploratory Data Analysis

- **Feature Understanding (Travel Frequency)**
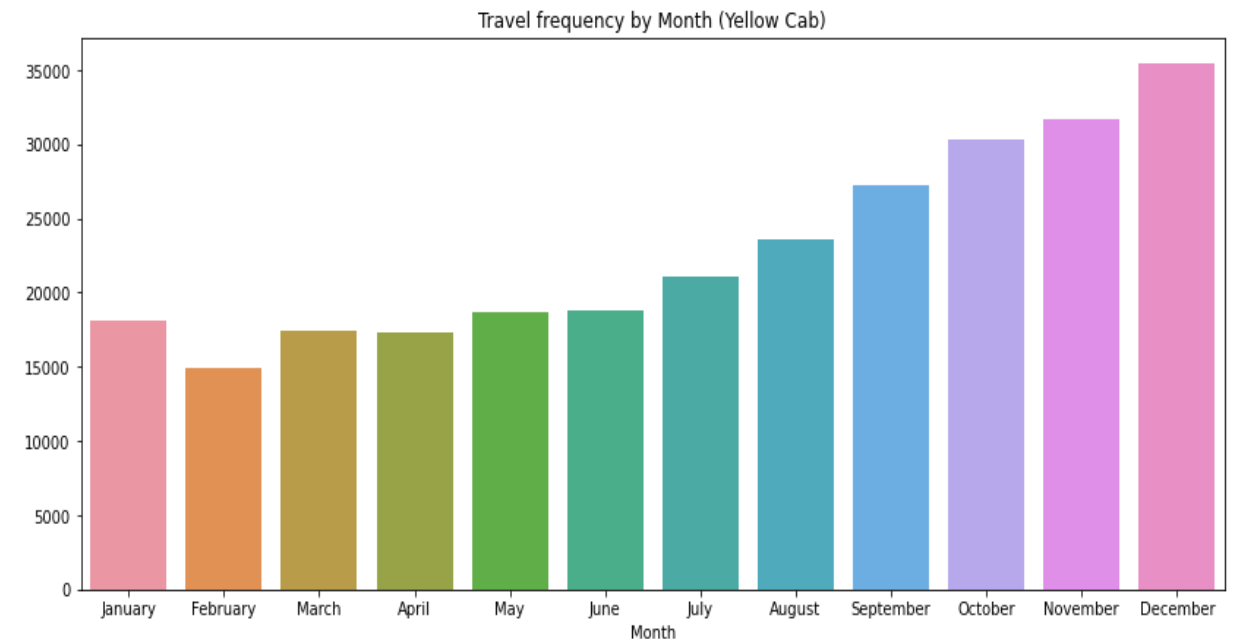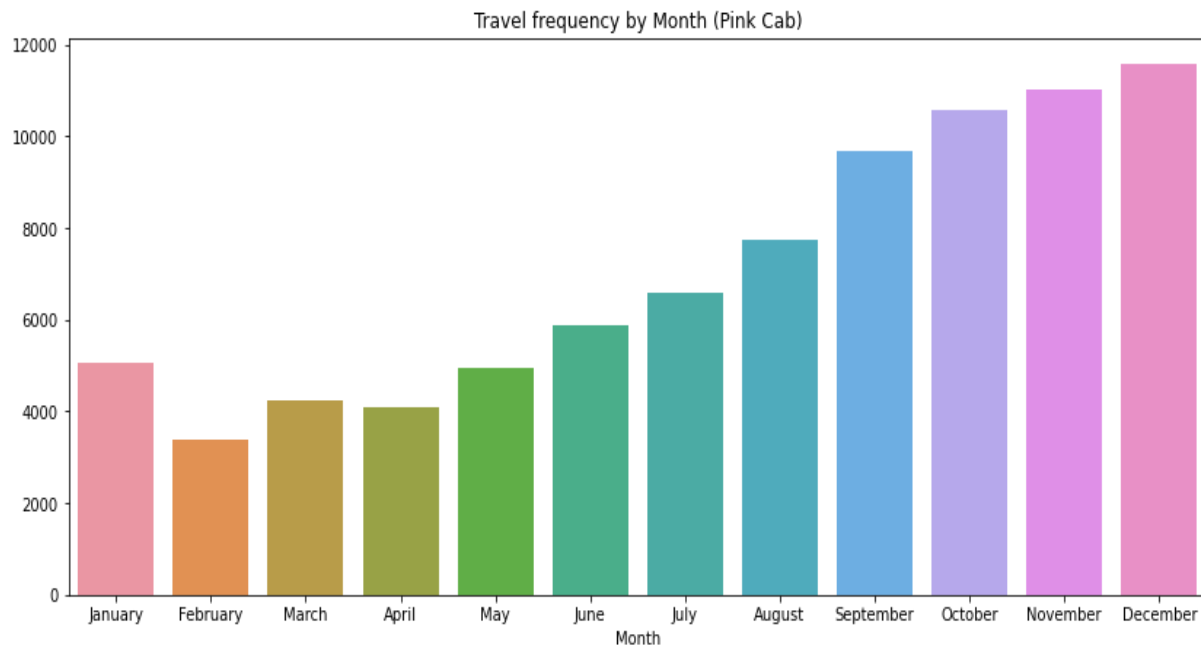
From the above charts, we can see clearly that Yellow Cab (14,932 KM - 35,439 KM) generally has higher travel distances than Pink Cab (3,375 KM - 11,568 KM) with a very wide margin. A notable comparison can be seen in the holiday season (December) where Yellow Cab has a higher travel distance (35,439 KM) than Pink Cab (11,568 KM).
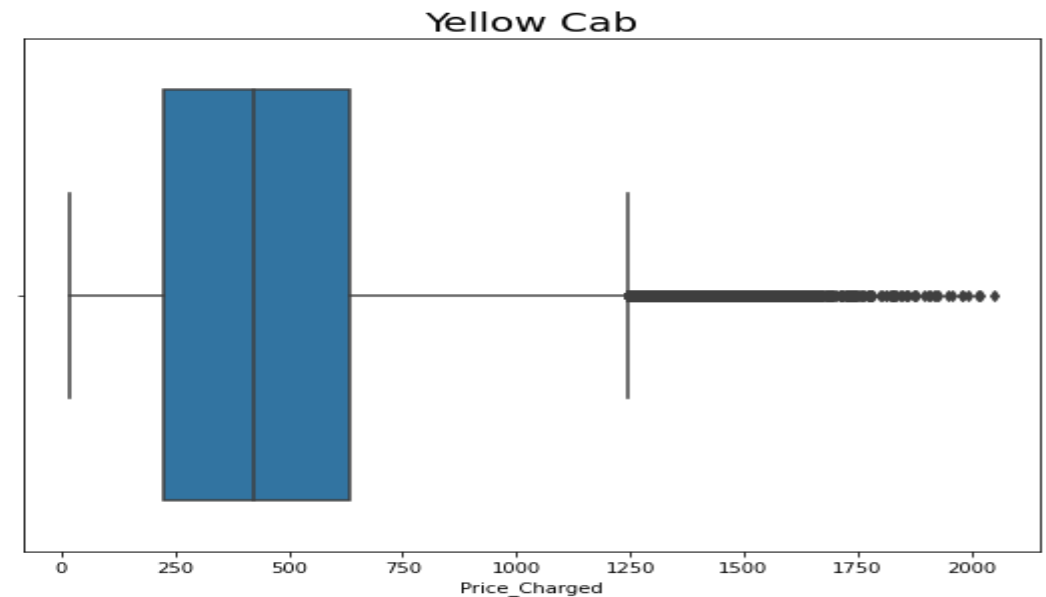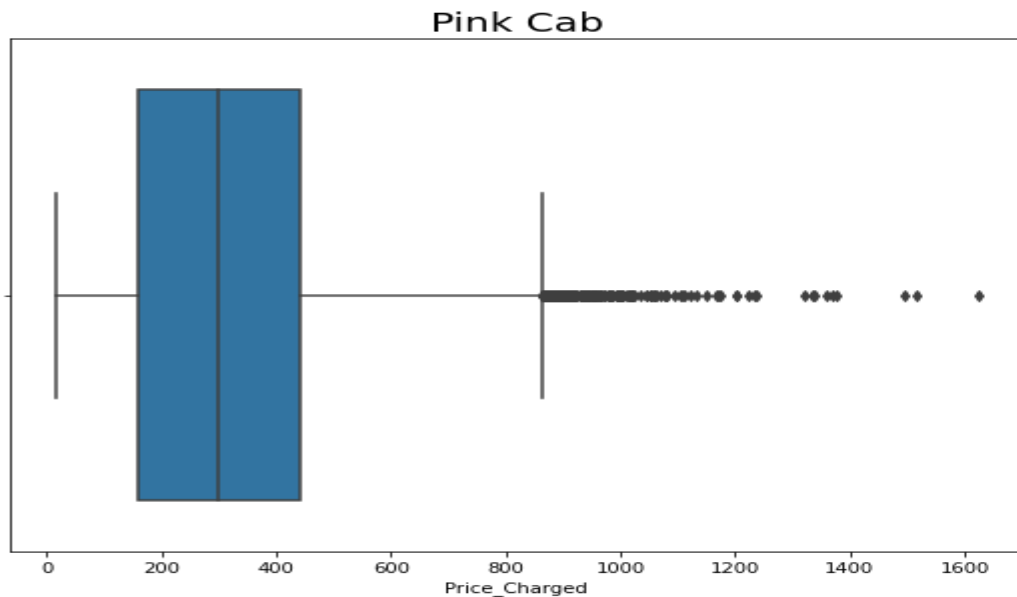
EDA: Exploratory Data Analysis

- **Feature Understanding (KM Travelled)**

From the analysis of the above charts, we can see that the distance traveled in the two cabs are about the same which is approximately between the range of 2 to 48 KMs.

# EDA: Exploratory Data Analysis

- **Feature Understanding (Outliers Detection)**

- From the above Boxplots, we could see that there are some outliers in price charged. This could be some cases where the prices charged are high compared to the regular charges.
- We can also see that the price charge range for Yellow cab is generally more than that of the Pink cab.

EDA: Exploratory Data Analysis

- **Feature Understanding (Price VS Distance Travelled)**
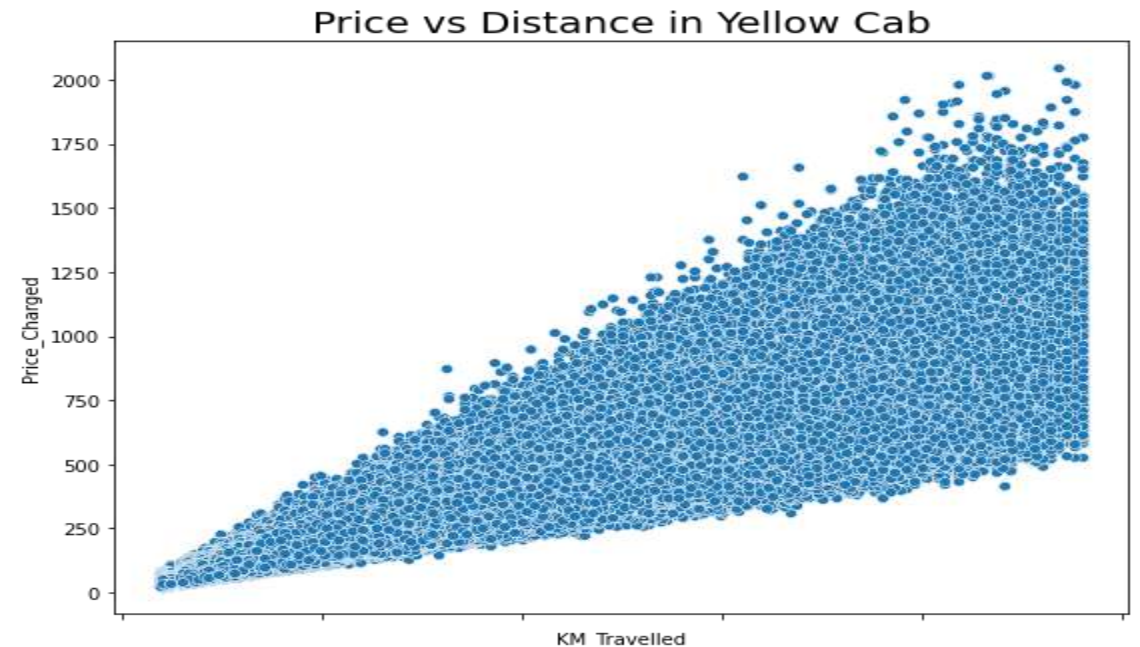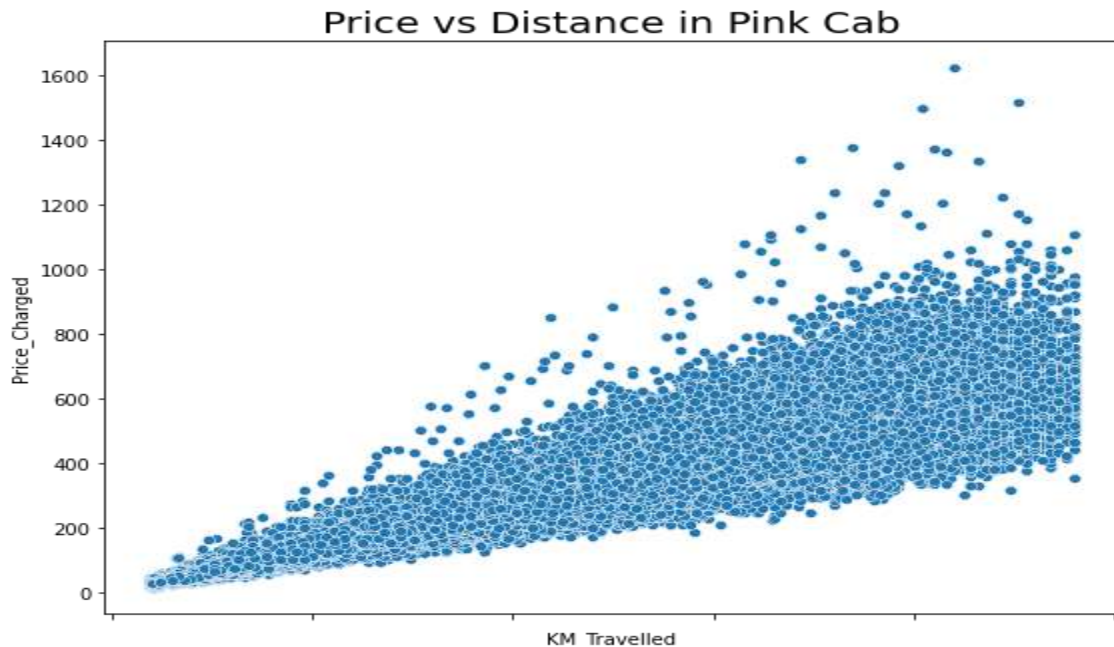
- Generally, Distance travelled is one of the factors that determines the price charged.¶
- We could see that there is a positive correlation in the graph for both Pink and Yellow cab.
- The prices charged for the Yellow cab is higher than that of the Pink cab.

EDA: Exploratory Data Analysis
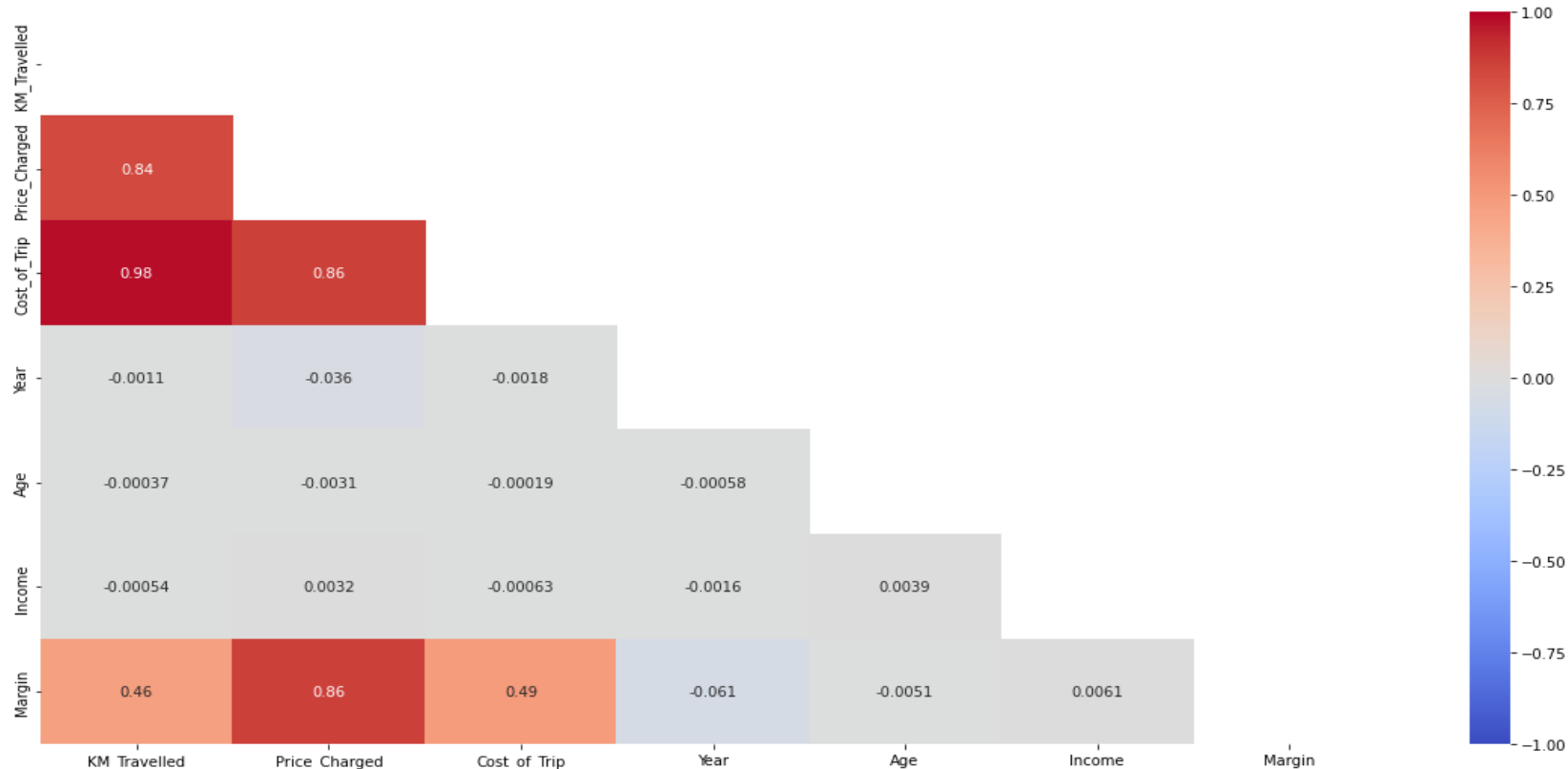
- From the above heatmap, we can see that there is a correlation between Margin and KM_Travelled, Price_Charged and Cost_of_Trip. However, it is more correlated with Price_Charged.
- KM_Travelled is correlated with Cost_of_Trip and Price_Charged. And finally, Price Charged is correlated with Cost_of_Trip.

# EDA: Exploratory Data Analysis

EDA recommendation and hypothesis of results

Final Investigation Areas:

- Which company has maximum cab users at a particular time period?

- Does margin proportionally increase with increase in number of customers?

- What are the attributes of these customer segments?
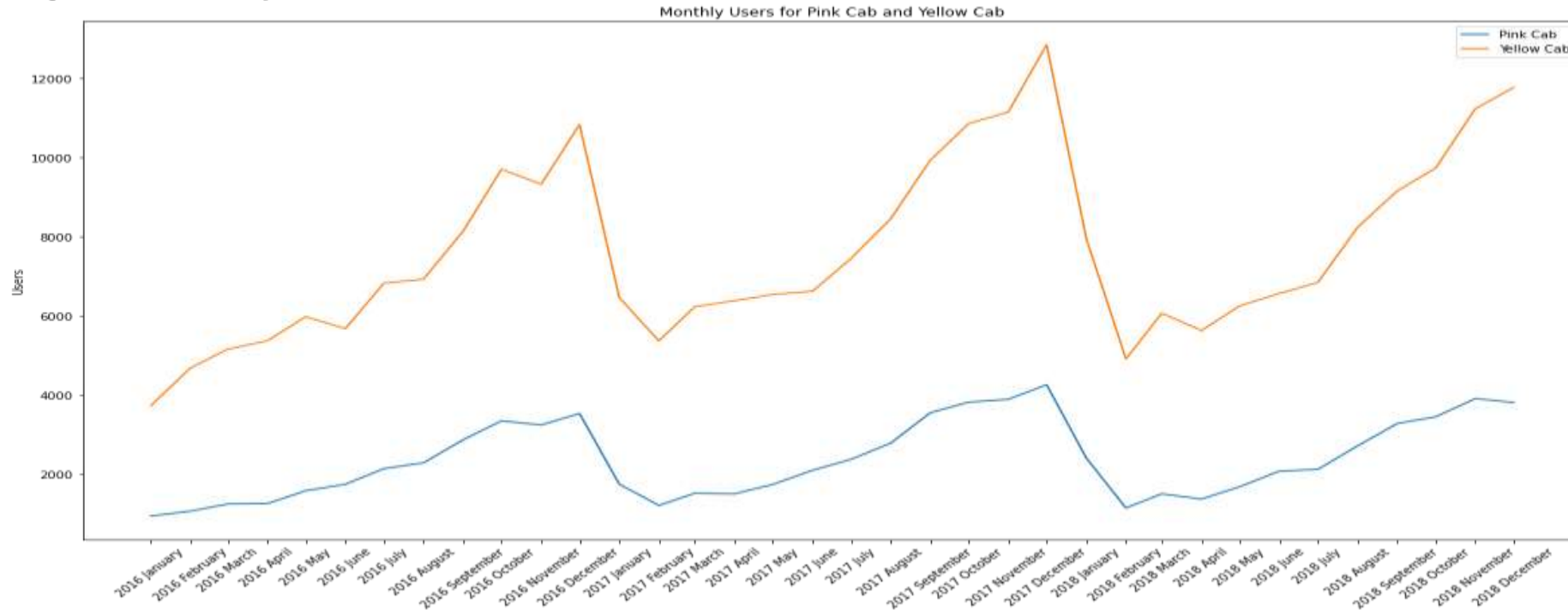
EDA: Exploratory Data Analysis

- **Which company has maximum cab users at a particular time period?**

Monthly Users for Pink Cab and Yellow Cab
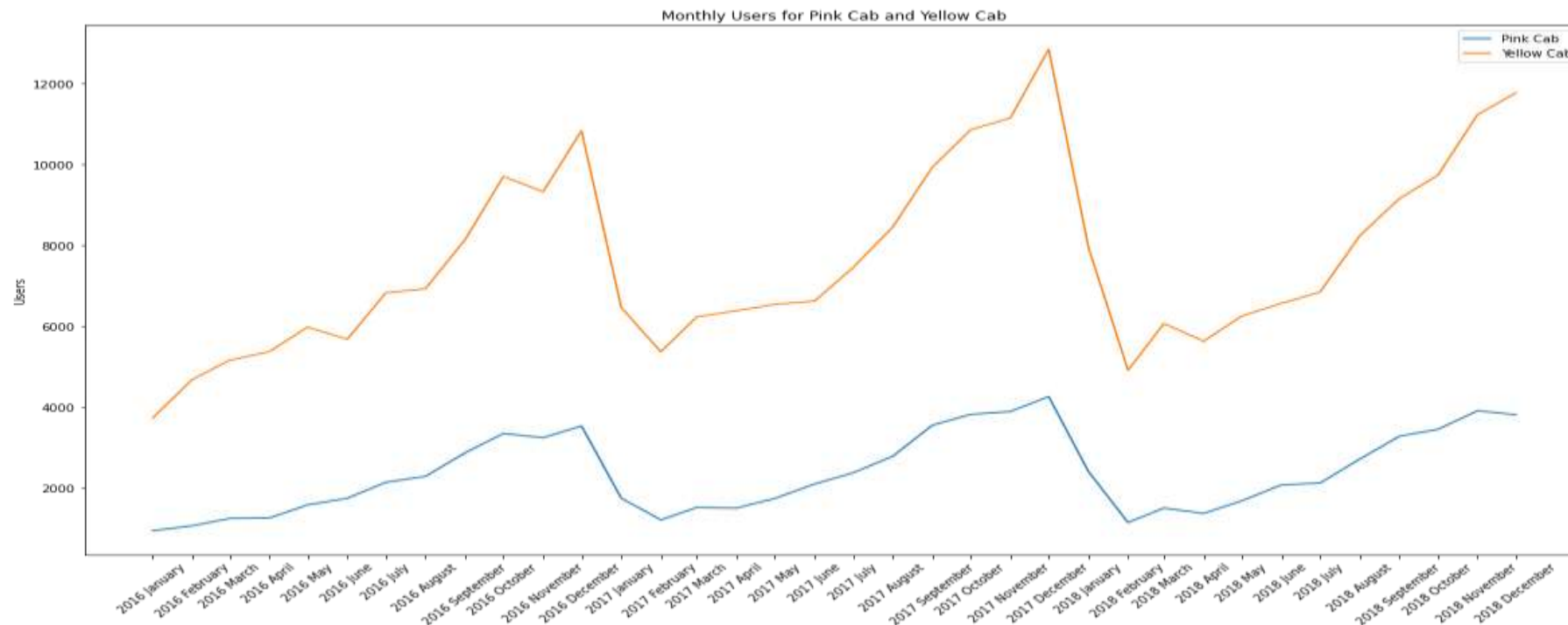
- From the chart and the analysis carried out to ascertain the maximum users at a particular time period, we can see that the peak period with maximum users for Pink Cab of 4247 is 2017 December and the peak period for the Yellow Cab users of 12846 is also 2017 December.
- It is clear that the Yellow Cab users are much more than the users of the Pink Cab.
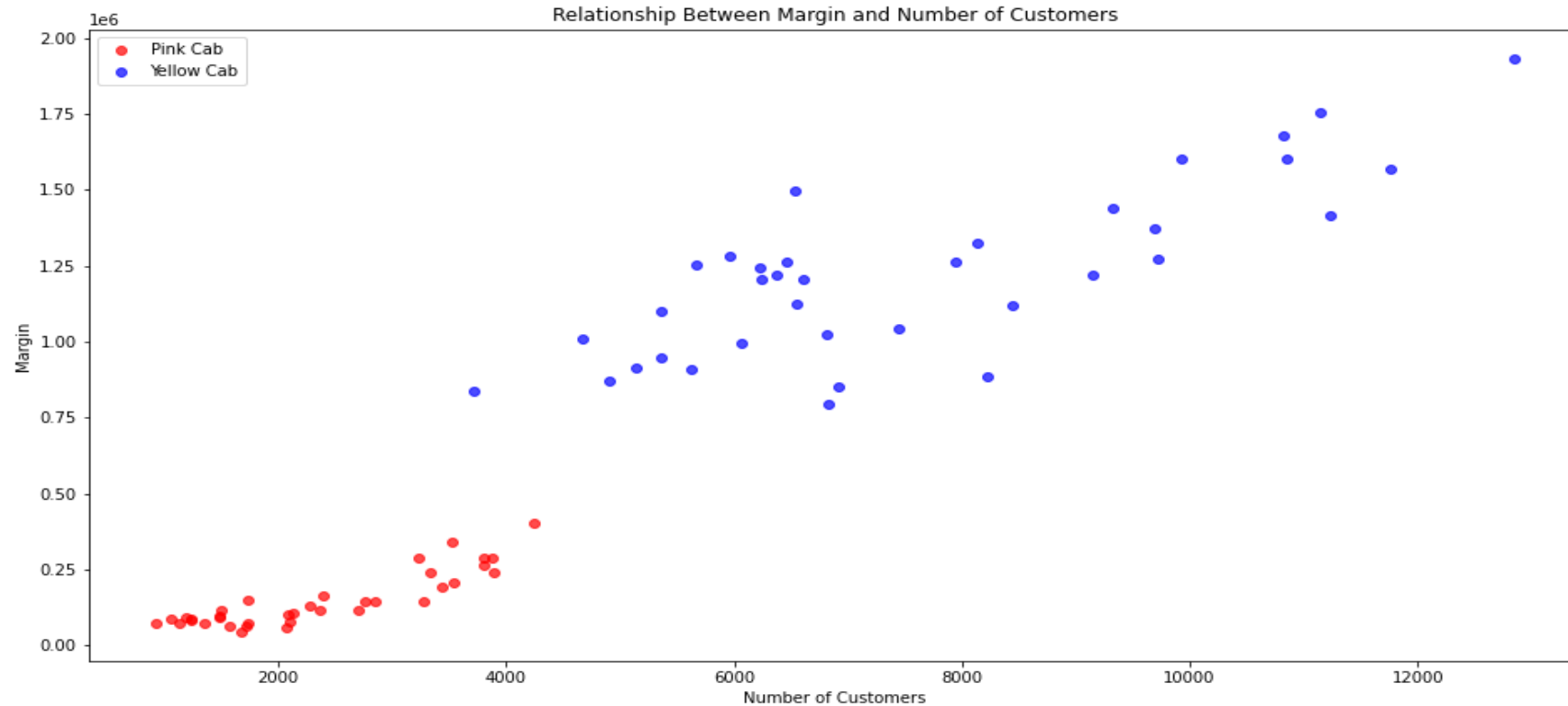
EDA: Exploratory Data Analysis

- **Margin Analysis for Pink Cab and Yellow Cab**

Monthly Users for Pink Cab and Yellow Cab

• From the chart and the analysis carried out to ascertain the maximum monthly margin at a particular time period, we can see that the peak margin for Pink Cab of 402094 is 2017 December and the peak margin for the Yellow Cab of 1932446.43 is 2017 December also.

• The Yellow Cab margin is much more than the margin of the Pink Cab.
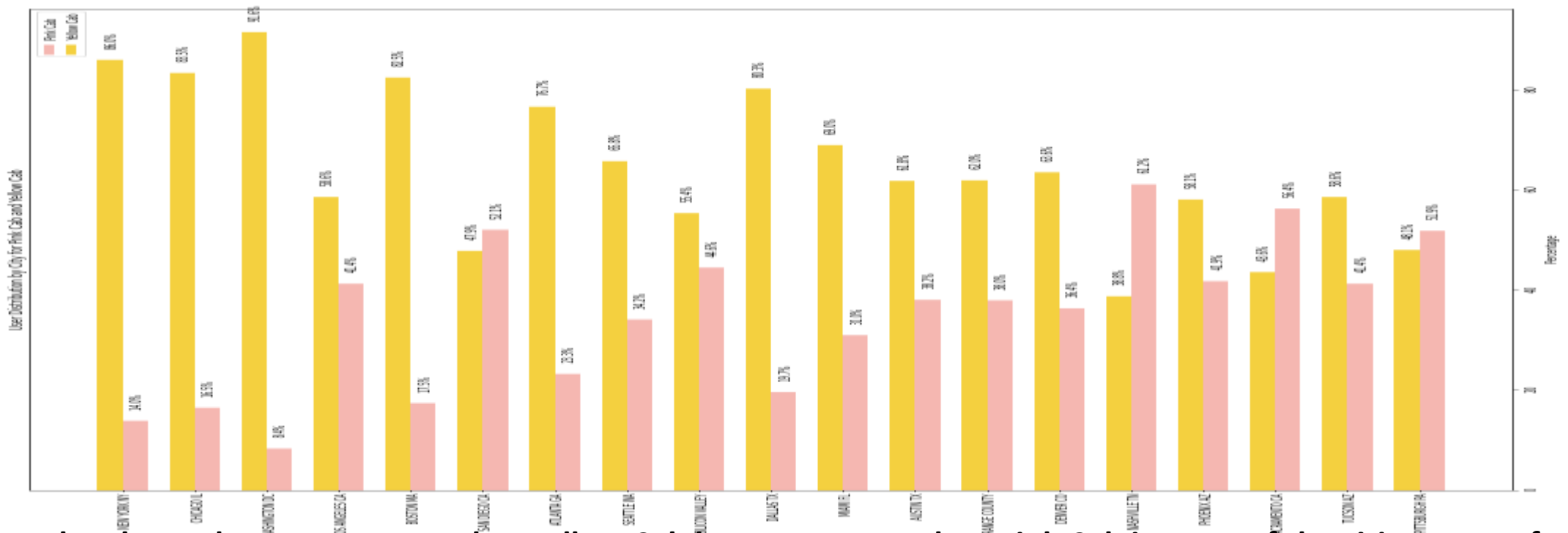
EDA: Exploratory Data Analysis

- **Relationship Between Margin and Number of Customers**

- **From the above chart, we can see that as number of customers increases, so is the margin.**
- **The Yellow Cab customers has higher margin compared to Pink Cab customers.**

EDA: Exploratory Data Analysis

- **Attributes of customer segments (Group by City)**

- From the above chart, we can see that yellow Cab has more users than Pink Cab in most of the cities except for few cities such as San Diego CA, Nashville TN, Sacramento CA and Pittsburgh PA. For Cities like New York NA, Chicago IL, Washington DC, Boston MA and Dallas TX, the Yellow Cab users are more than three times higher than Pink Cab users which makes a whole lot of difference between them.

EDA: Exploratory Data Analysis

- **Attributes of customer segments (Group by Gender)**
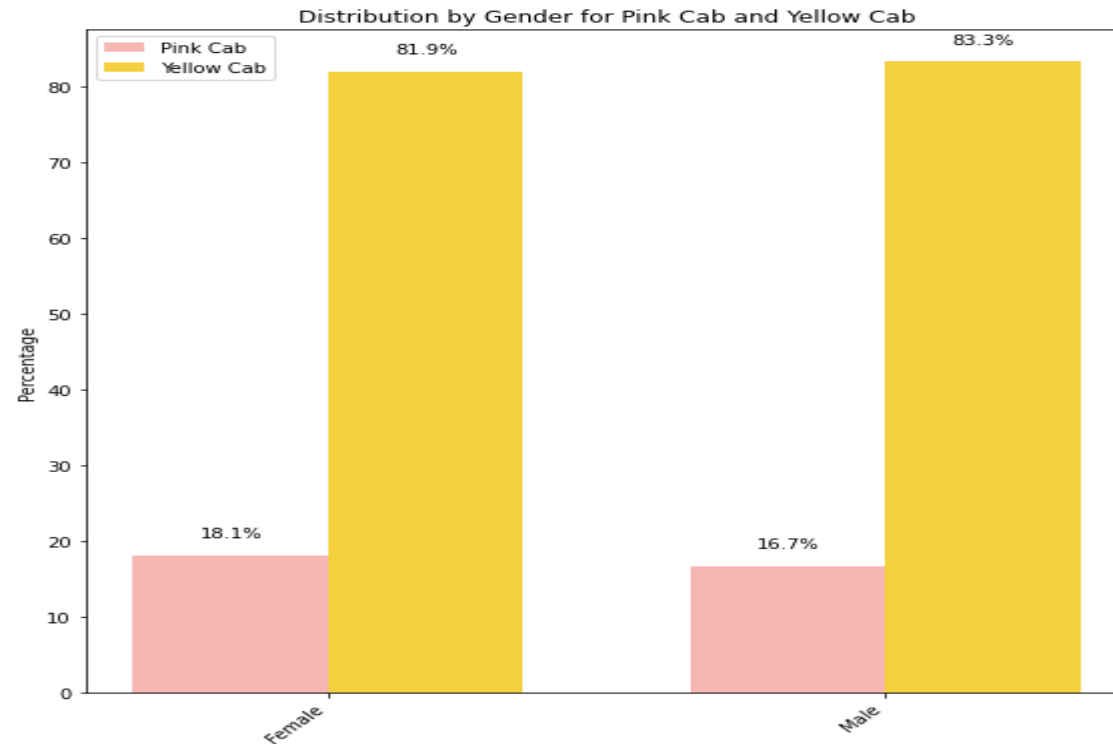
Distribution by Gender for Pink Cab and Yellow Cab

• From the above bar chart, the Yellow Cab are about 4 times higher than Pink Cab in terms of Priced_Charged for both male and female customers. This shows that the Yellow Cab are more patronized than Pink Cab.

EDA: Exploratory Data Analysis

- **Attributes of customer segments (Group by Age)**
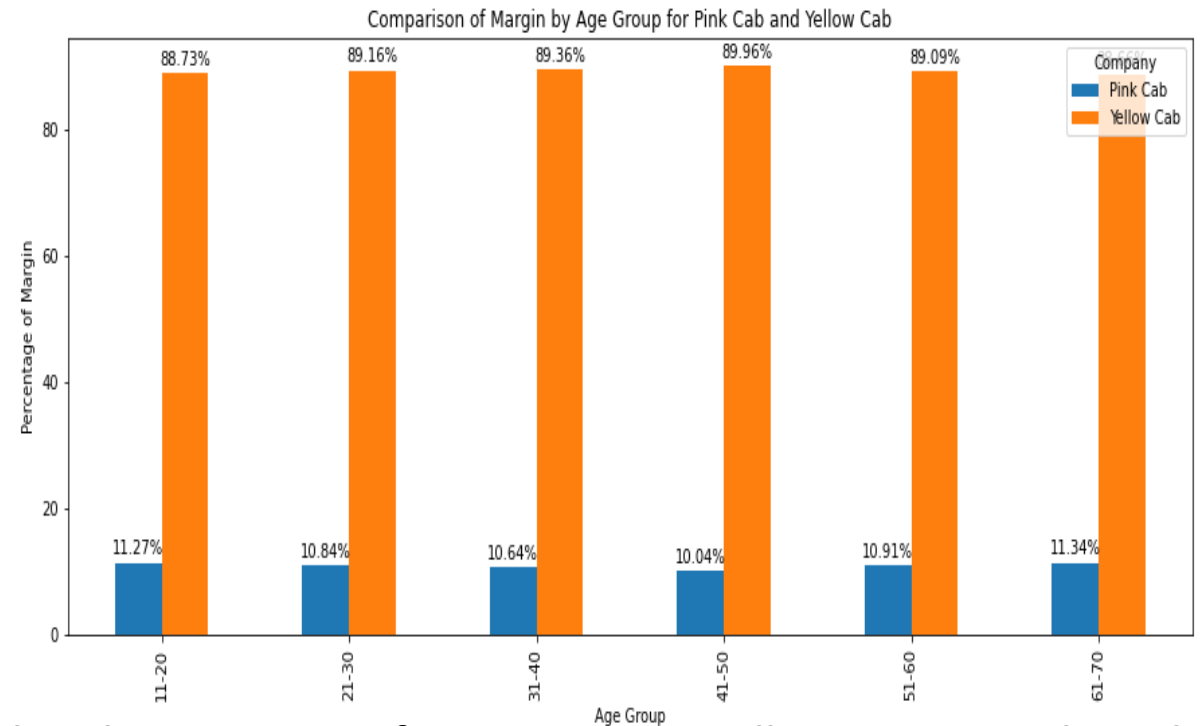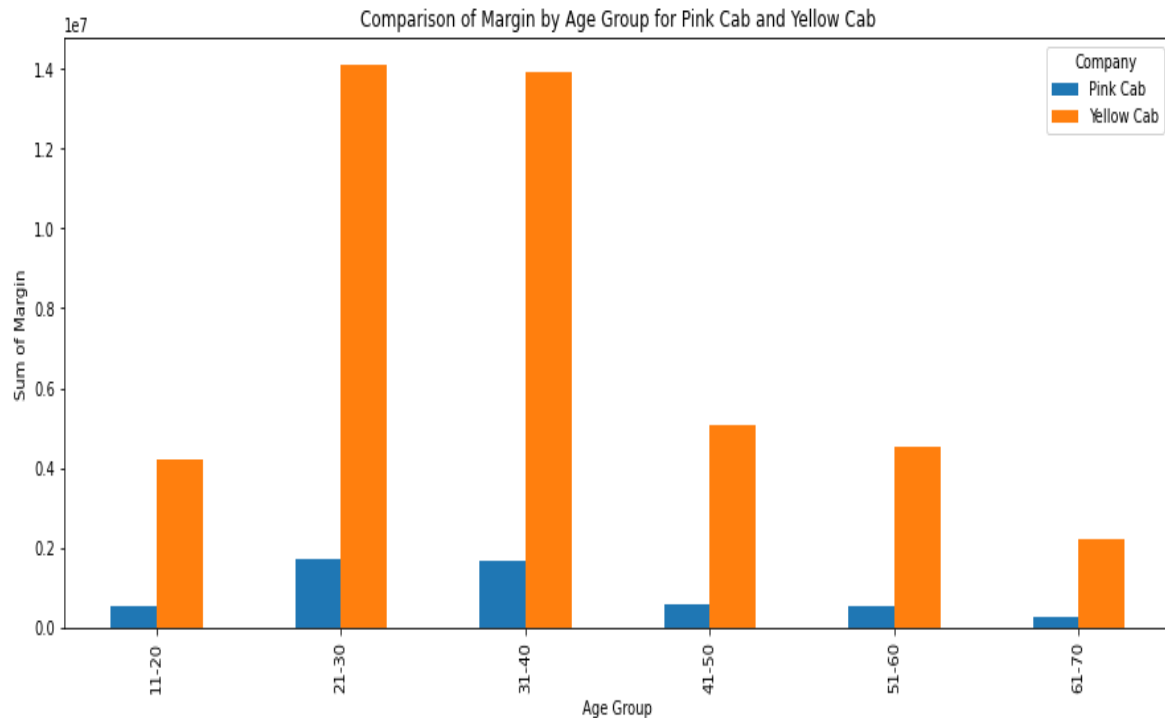
- From the above charts, we can see that customers within the age range of 21-40 are generally patronizing the Cab business more than customers from other age groups for both Pink Cab and Yellow Cab.
- When comparing both companies, the Yellow Cab customers for different age groups patronize more than 7 times the Pink Cab customers making Yellow Cab a more suitable business for investment.

EDA: Exploratory Data Analysis

- **Attributes of customer segments (Group by Income)**
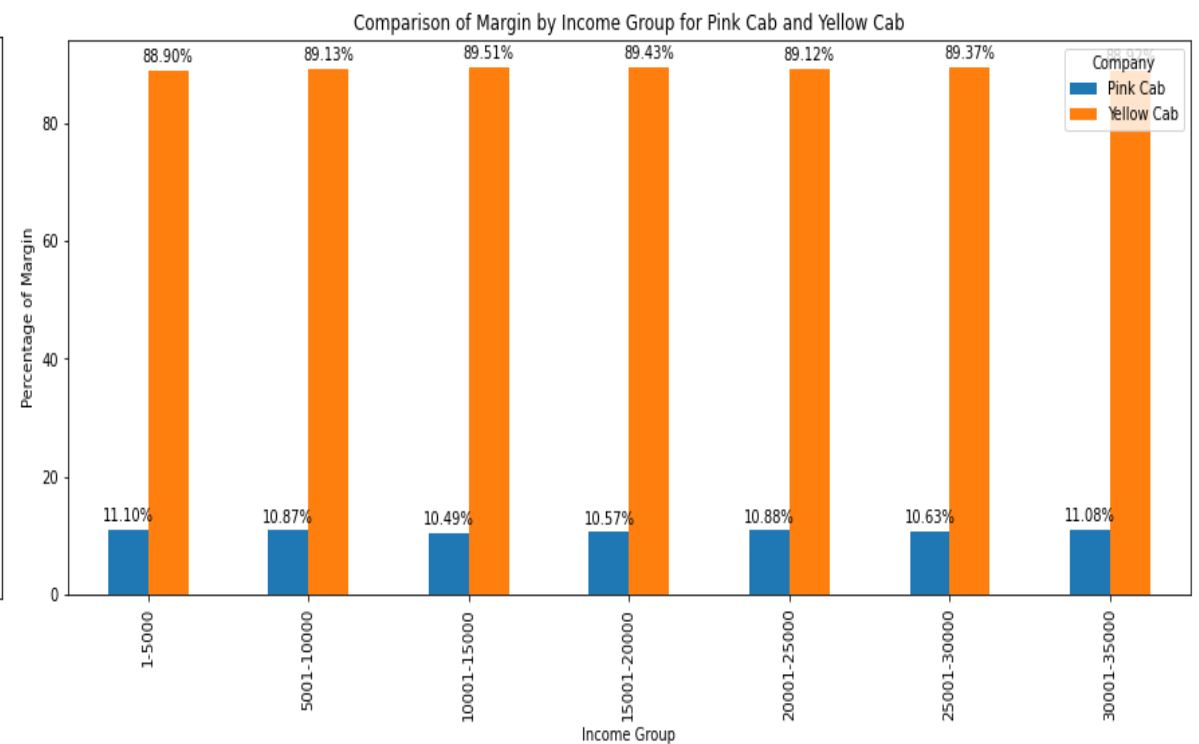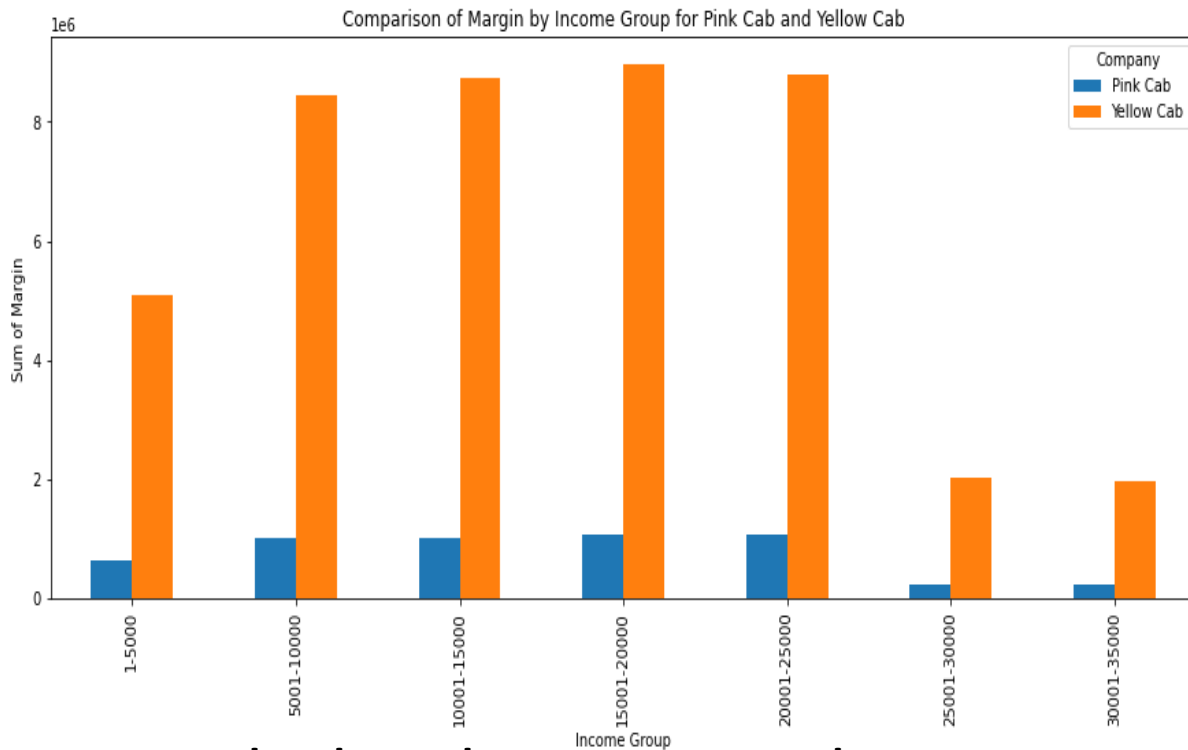
- From the above charts, we can see that customers earning within 5001 and 25000 are generally patronizing the Cab business more than customers from other income groups for both Pink Cab and Yellow Cab.
- When comparing both companies, the Yellow Cab customers for different income groups patronize more than 7 times the Pink Cab customers making Yellow Cab a more suitable business for investment.

EDA: Exploratory Data Analysis

- **Attributes of customer segments (Group by Payment Mode)**

Distribution by Payment Mode for Pink Cab and Yellow Cab

• From the above bar chart, the Yellow Cab are about 4 times higher than Pink Cab in terms of Priced_Charged for both cash and card customers. This shows that the Yellow Cab are more petronized than Pink Cab

EDA: Exploratory Data Analysis
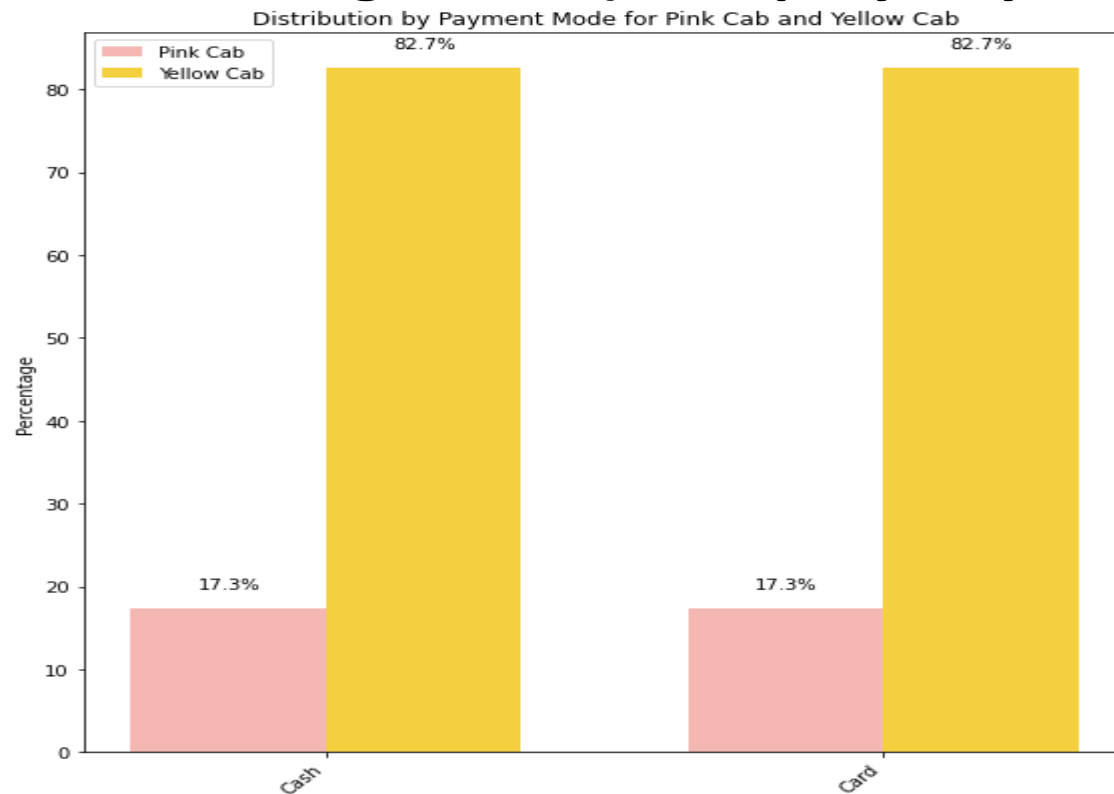
- ## **Summary**

- **Feature Understanding:**
The EDA revealed that Yellow Cab consistently covers longer travel distances than Pink Cab, with notable disparities during peak seasons. Both companies tend to operate within a similar range of distances, roughly 2 to 48 kilometers. Additionally, price charged for Yellow Cab is generally higher than for Pink Cab, with a positive correlation observed between distance traveled and price charged. This feature understanding can be pivotal in assessing market dynamics and pricing strategies for the investment decision.

- **Attributes of Customer Segments:**
The analysis of customer segments based on gender, age, income, and payment mode highlights Yellow Cab's dominance in most categories. Regardless of the segment, Yellow Cab consistently outperforms Pink Cab in terms of price charged, often by a substantial margin. These insights offer valuable guidance for identifying target demographics and tailoring marketing and investment strategies effectively within the cab industry.

Recommendations (Part 1)

- **XYZ Company Investment Recommendations**

1. Investment Focus:
Given the consistently superior performance of Yellow Cab across various customer segments and in terms of price charged, we recommend that XYZ Company consider Yellow Cab as the primary investment focus. The data suggests that Yellow Cab has a larger customer base, provides higher-priced services, and exhibits more favorable margins, making it a more promising choice for investment within the cab industry.

2. Market Positioning:
To maximize the investment's potential, XYZ Company should strategize market positioning, leveraging Yellow Cab's strengths. This may involve targeted marketing campaigns, pricing strategies, and service enhancements to further solidify Yellow Cab's position as the preferred choice for customers in various demographics. Additionally, it's important to capitalize on the insights from the Feature Understanding section to optimize pricing strategies and operational efficiency.

Recommendations (Part 2)

- **XYZ Company Investment Recommendations**

3. Continuous Monitoring:

After making the investment, continuous monitoring of market trends, customer segments, and performance metrics is essential. Regularly reassess the market landscape and make data-driven adjustments to stay competitive and sustain growth within the cab industry.

These recommendations aim to guide XYZ Company toward a well-informed and strategic investment decision in the dynamic cab industry, with a focus on maximizing returns and market dominance.

Exploratory Data Analysis
G2M insight for Cab Investment firm
(Analysis done by: Okeoma Ihunwo)

Thank You