



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Case Study

July 20, 2023

Agenda

Background

Data Overview

EDA

Hypothesis

EDA Summary

Recommendations

Background

Background

- XYZ is a private firm in the US. and is planning for an investment in Cab industry.
- They want to understand the market before taking final decision.
- I have been provided with multiple data sets that contains information on 2 cab companies.
- The goal is to provide XYZ with insights to help them identify the right company to make their investment.

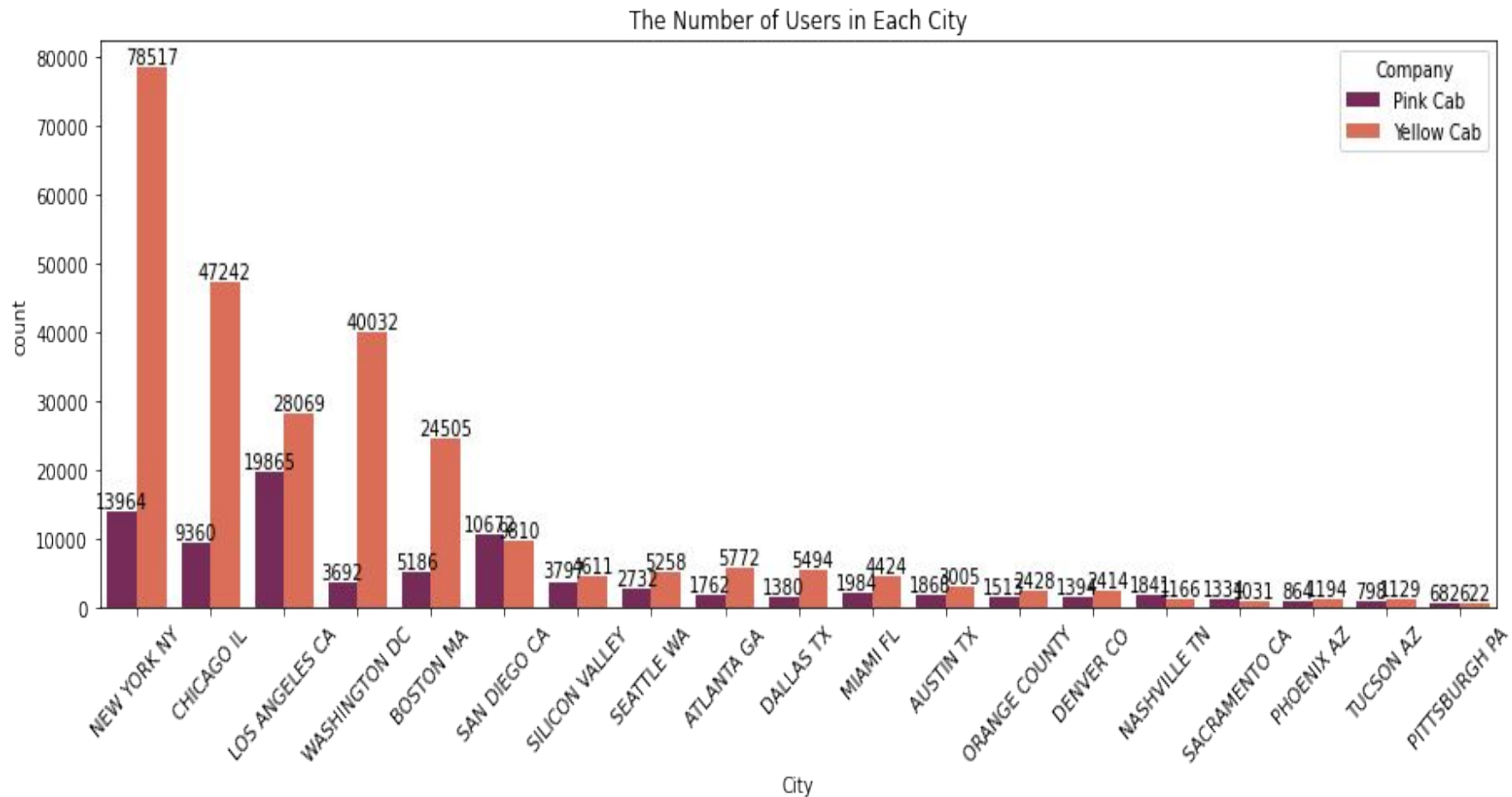
Data Overview

Data Overview

- There are four datasets. One for cab data, customer data, transaction data, and city data.
- There were no null, missing, or duplicate values in all four datasets.
- I was able to combine all four datasets based on Transaction ID, Customer ID, and city.
- There were some outliers in the “Price Charged” column that I removed.
- The master dataset has 359,392 entries and 14 columns.

EDA

Users

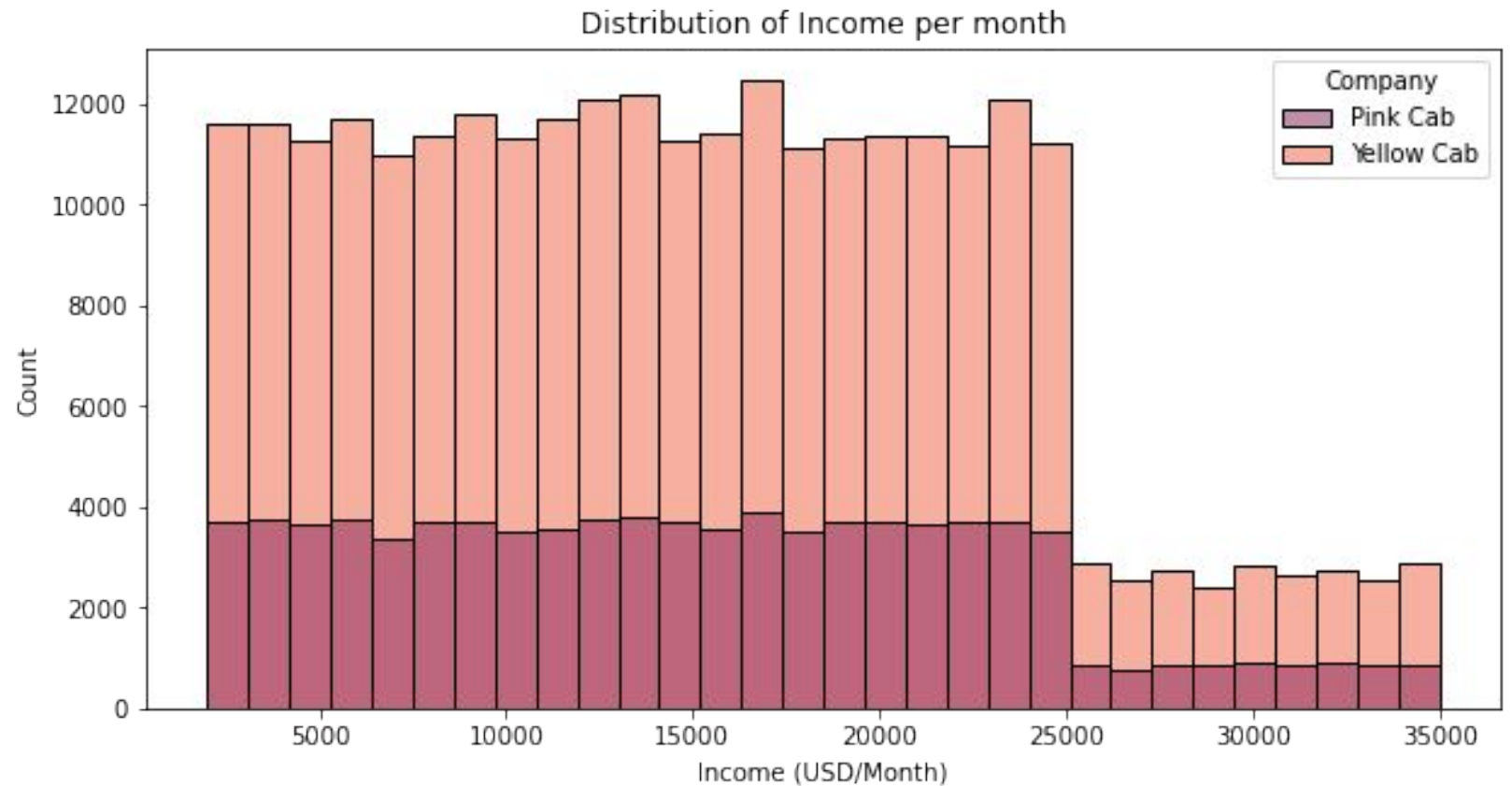


The Yellow Cab company has more users compared to the Pink Cab company.

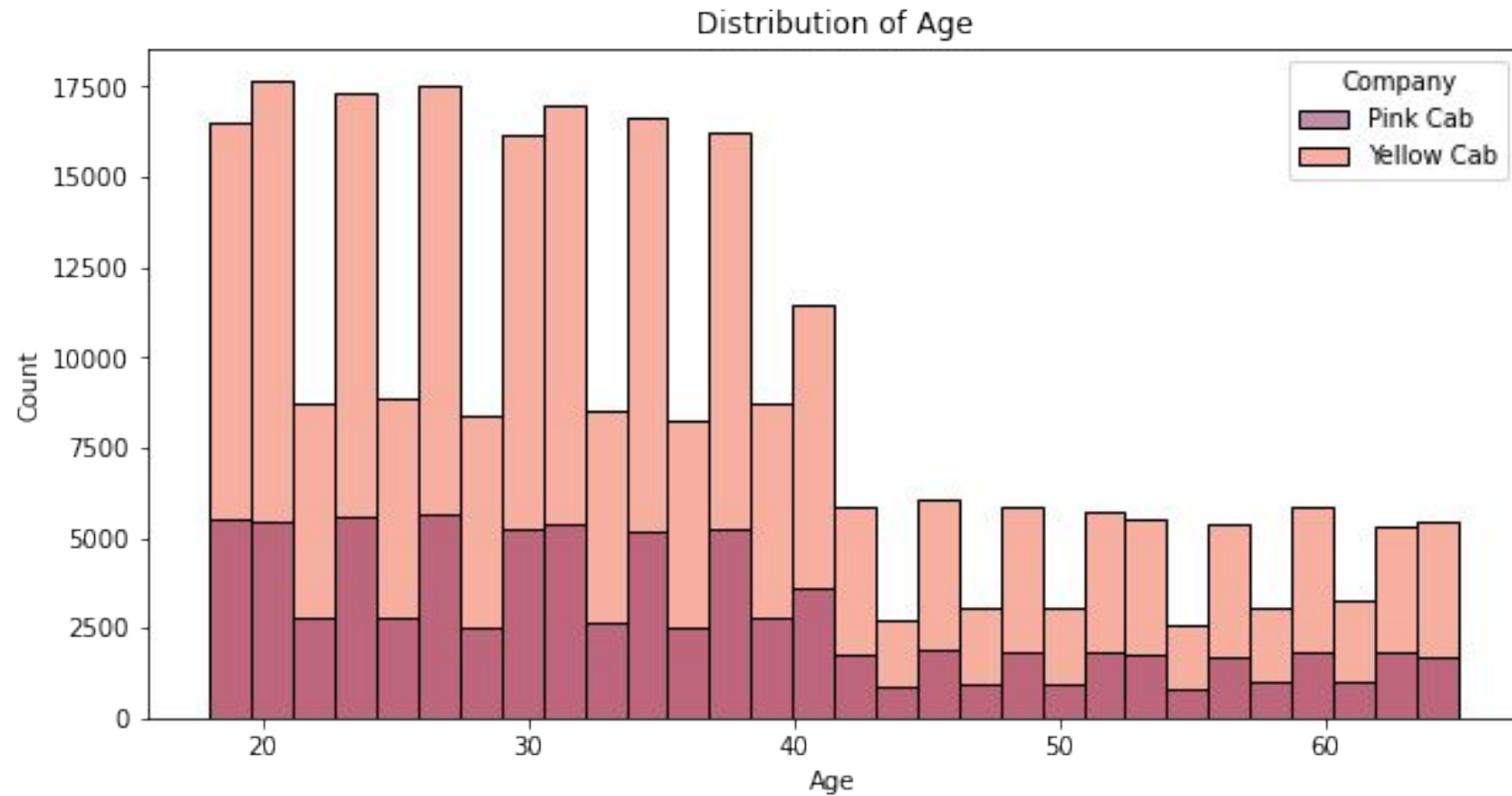
Income

The monthly income of users for the Pink Cab and the Yellow Cab company is about the same.

Both set of customers seem to be mainly middle and upper class.



Age



Most of the customer for both companies are between the ages of around 18 to 40.

Gender

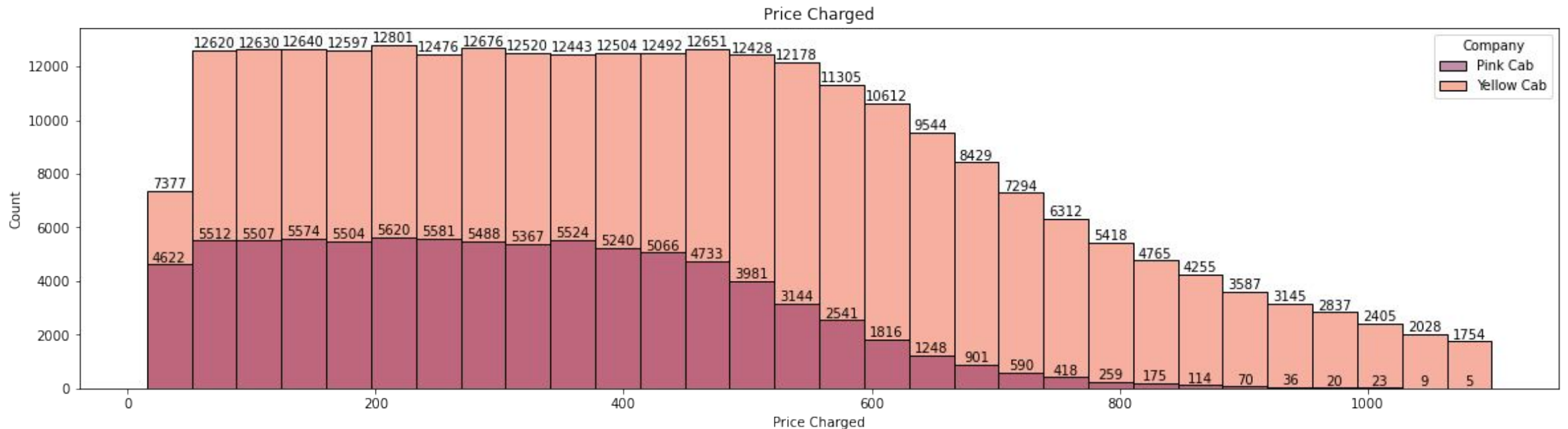
Majority of the customers are male even though male customers are not significantly higher.

About 57% of all customers are male and 43% are female.



Price Charged

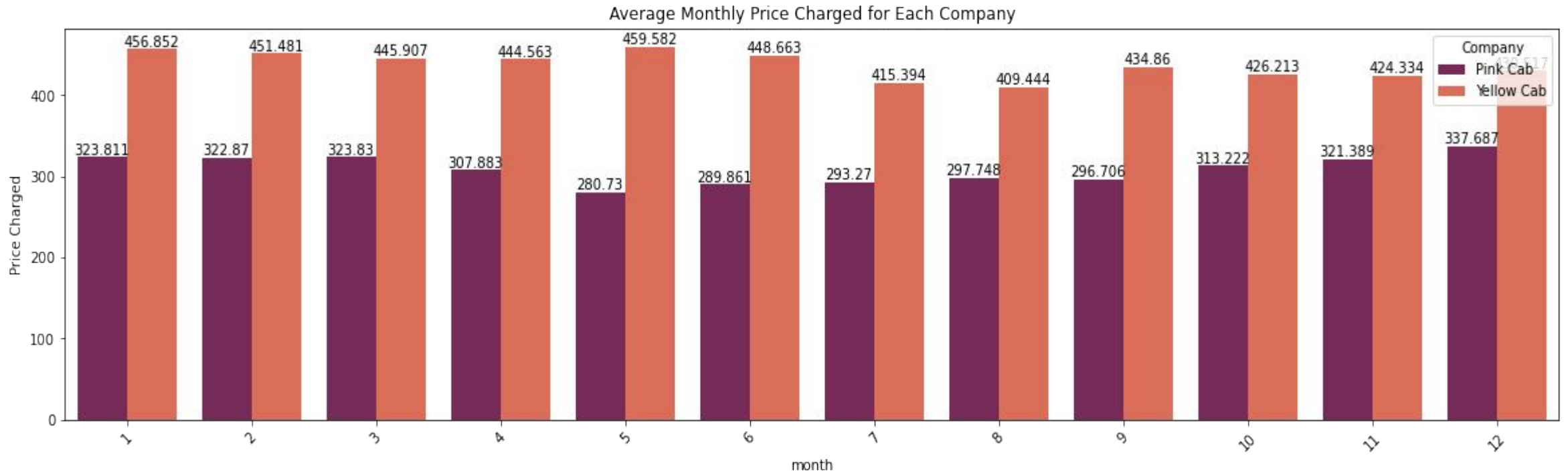
The Yellow Company has a higher distribution of price charged. Rides that cost \$800+ were mainly paid by the Yellow Cab company customers.



Average Price Charged per Month

The Yellow Company has a higher monthly price charged over the three years.

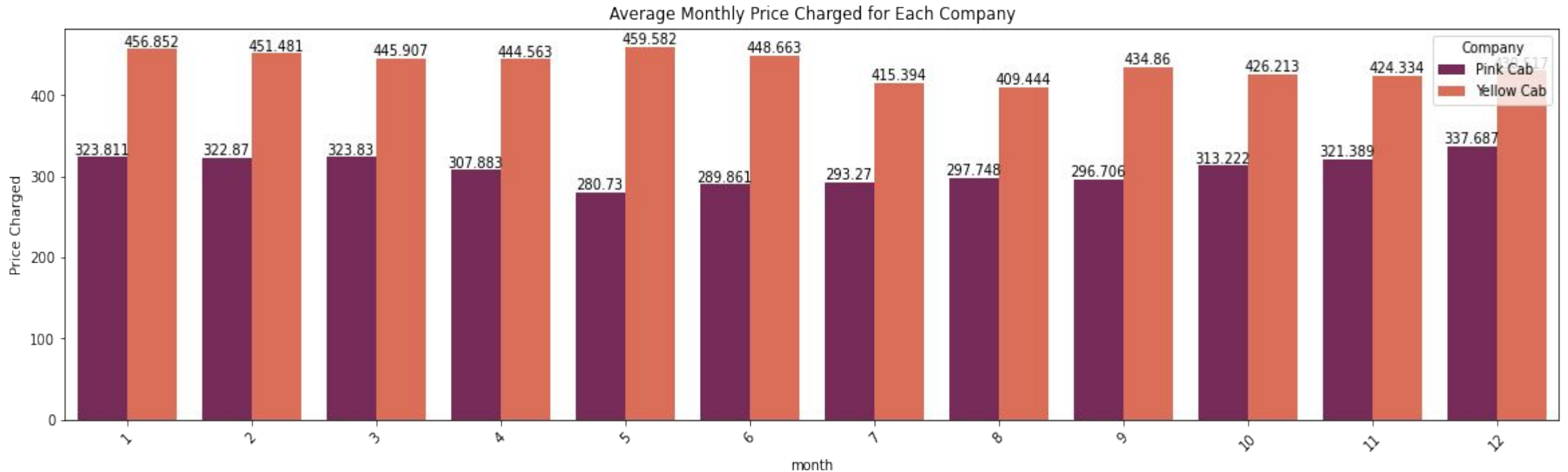
Most of the Yellow Cab price charge is around \$450, and July and August seems to be their cheaper months.



Average Price Charged Cont.

The Pink Cab average price charge is around \$320, and May and June is their cheaper months.

May is the Yellow Cab company peak price charge month. December is the Pink Can peak price charge month.



Seasonal Trends

It's a steady increase from the beginning of the year until the end of the year. November and December are the months with the most customers and significantly drops off at in January of the next year.

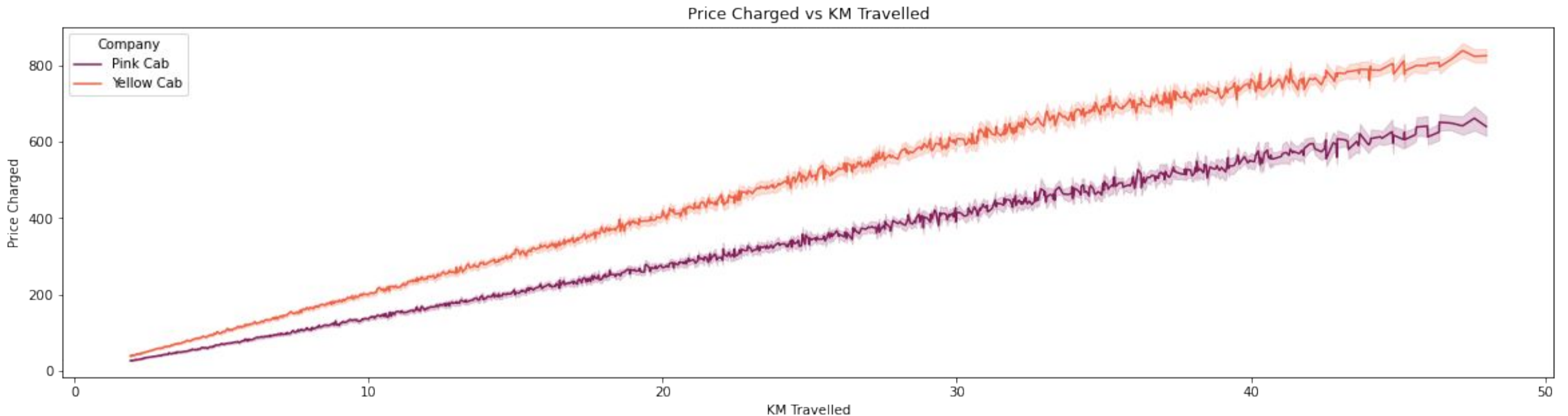


Hypothesis

Hypothesis: KM Travelled

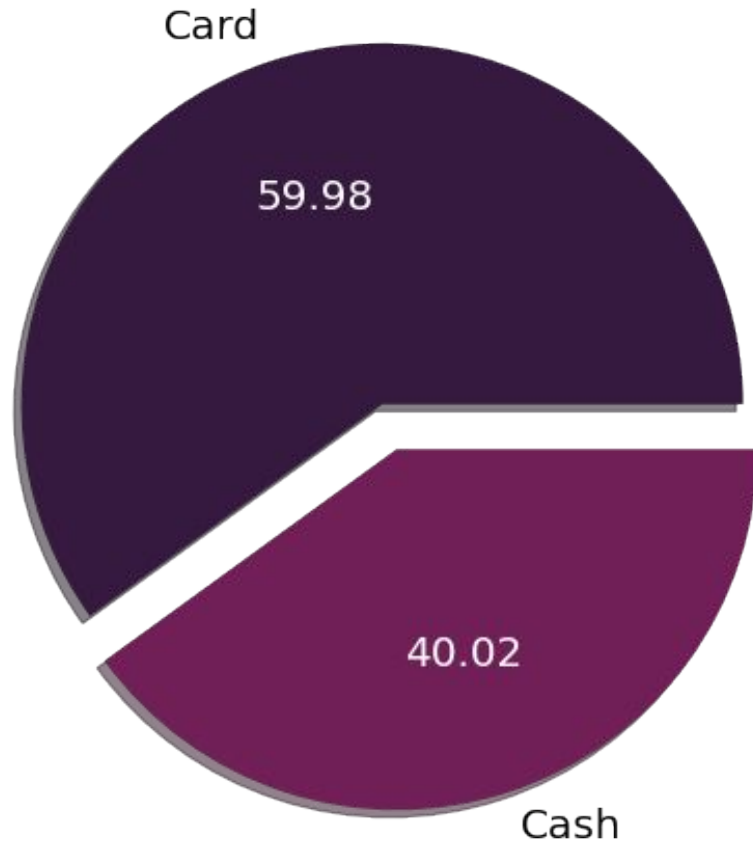
The more KM travelled, the higher the price charged is.

The Yellow Cab charges more than the Pink Cab company.



Hypothesis: Payment Mode

Proportion Based on Payment_Mode



Most customers pay with a card.

There are still a good portion of customers that have paid in cash.

EDA Summary

EDA Summary

Users: The Yellow Cab company is performing better than the Pink Cab company. It has significantly more customers with about 76% of customers of The Yellow Cab company.

Income: Both groups have similar distributions of monthly income. The Yellow Cab just have more customers.

EDA Summary

Price Charged: The Yellow Company has a higher distribution of price charged. Rides that cost \$800+ were mainly paid by the Yellow Cab company customers.

Monthly Average Price Charged: The Yellow Company has a higher monthly price charged.

Recommendations

Recommendation

Based on the users, income, price charged, etc. I would recommend to invest in the Yellow Cab company.

The Yellow Cab company is outperforming the Pink Cab company.

Thank You