G2M Case Study

(cab industry)

Nicolette Peterkin
Virtual Internship
22-June-2021

Background

The cab industry in the United States has grown significantly, and as a result, XYZ firm is interested in investing in the cab sector. They handed over four datasets to the senior Data Scientist to analyse and give actionable insights. The goal of this project is to clean, append, join, and analyse the datasets provided in order to provide sufficient advice to XYZ's Executive team on the most lucrative cab company.

Business Problem

Buying into a new industry requires a lot of research and data analysing to ensure that the right Go-to-Market(G2M) strategy is used to enter the market. The issue XYZ is having currently is that they do not understand the cab industry market in which they are thinking to invest. Solution.

SOLUTION

As a Data Scientist my job is to analyse the given data with much scrutiny in order to provide XYZ with vital insights into the cab industry. Once the exploratory data analyse report is completed XYZ will have actionable insights which can then be used to inform their G2M strategy.

Layout of Case Study

The analysis has been divided into four parts:

- Data Understanding

- Find the company that has maximum cab users

- Finding the most profitable Cab company

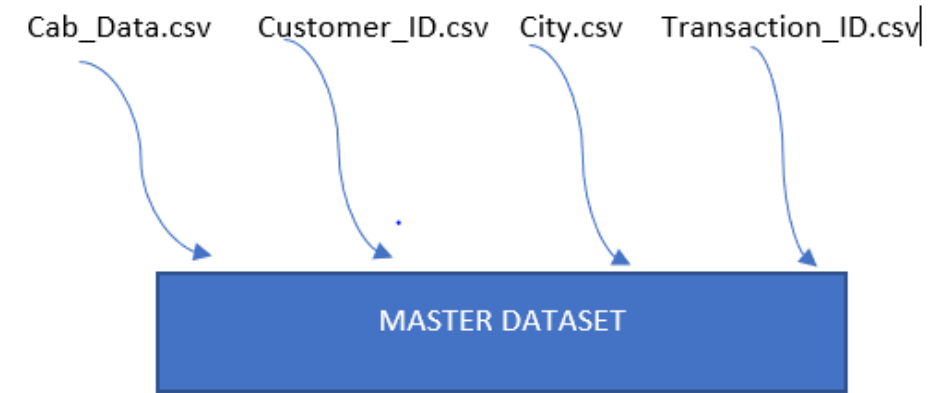- Recommendations for investment

Hypothesis to Test

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

2. Does margin proportionally increase with increase in number of customers?

3. What are the attributes of these customer segments?

4. What age range of customer use cabs the most?

5. Is there a correlation between price charged and cost of trip?

Data Understanding

➤ Timeframe of the data: January 31, 2016 to December 31,2018
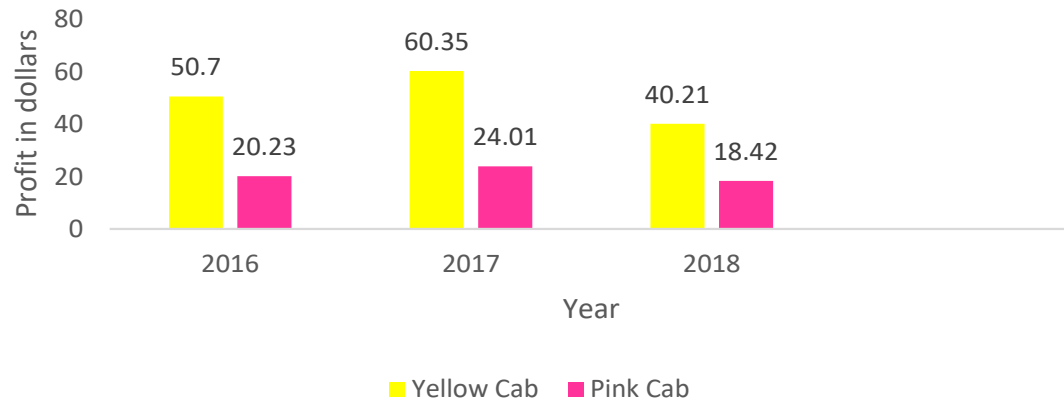
➤ Total data points : 359,392

Assumptions:

➤ Outliers are present in Price_Charged feature but due to unavailability of trip duration details ,we are not treating this as outlier.

➤ Profit of rides are calculated keeping other factors constant and only.

➤ Price_Charged and Cost_of_Trip features used to calculate profit.
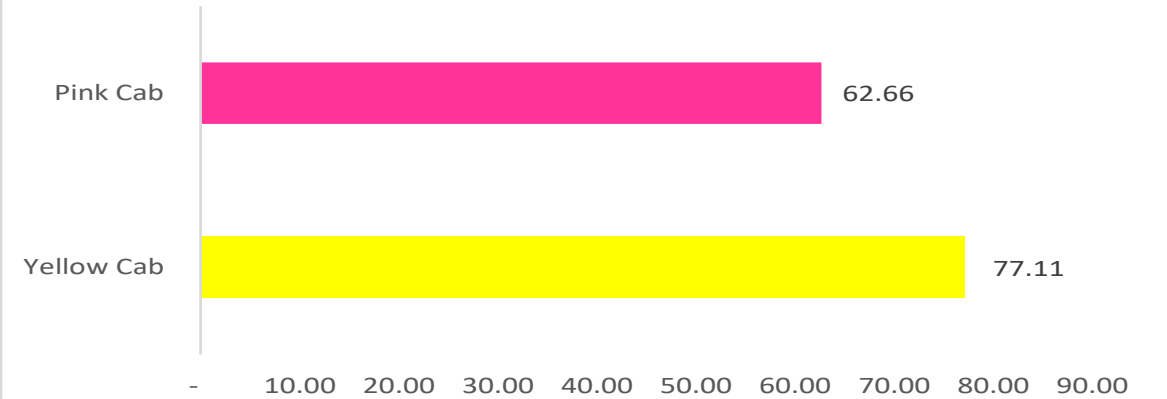
➤ Yellow Cab has the most customers.

Cab_Data.csv   Customer_ID.csv   City.csv   Transaction_ID.csv

MASTER DATASET

Profit Analysis

Average Yearly Profit

Profit in dollars

Year	Yellow Cab	Pink Cab
2016	50.7	20.23
2017	60.35	24.01
2018	40.21	18.42

■ Yellow Cab ■ Pink Cab

Total Average Profit

Pink Cab — 62.66

Yellow Cab — 77.11

- 10.00 20.00 30.00 40.00 50.00 60.00 70.00 80.00 90.00

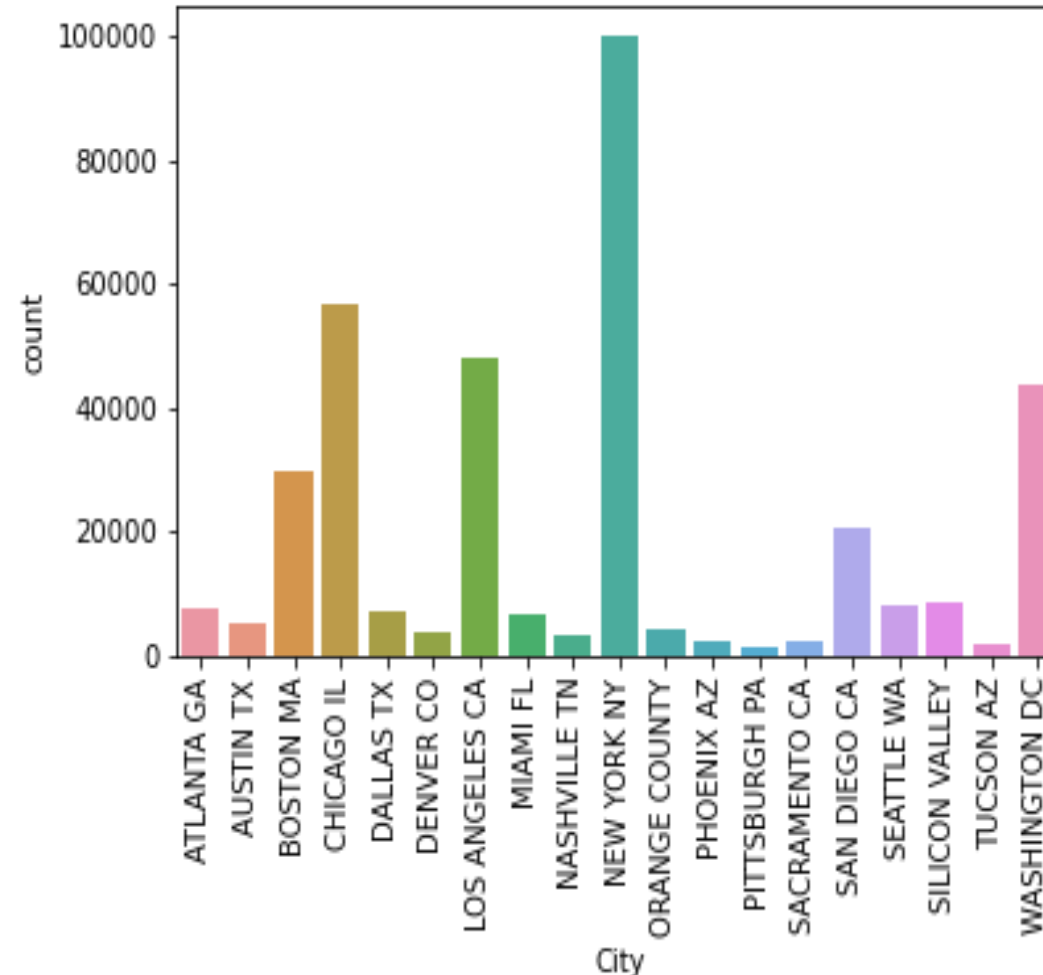|	Yellow Cab	Pink Cab
Total Profit	6,531,387.1	5,307,328.32

Univariate Plots Analysis
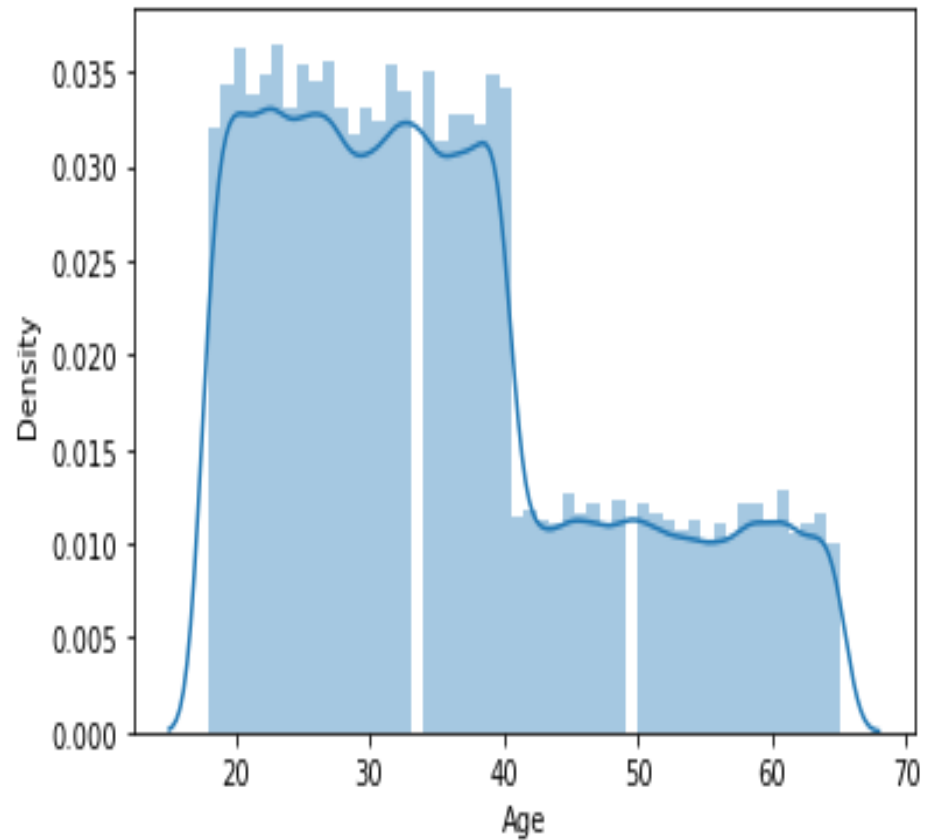
Customer Analysis based on Cab Company
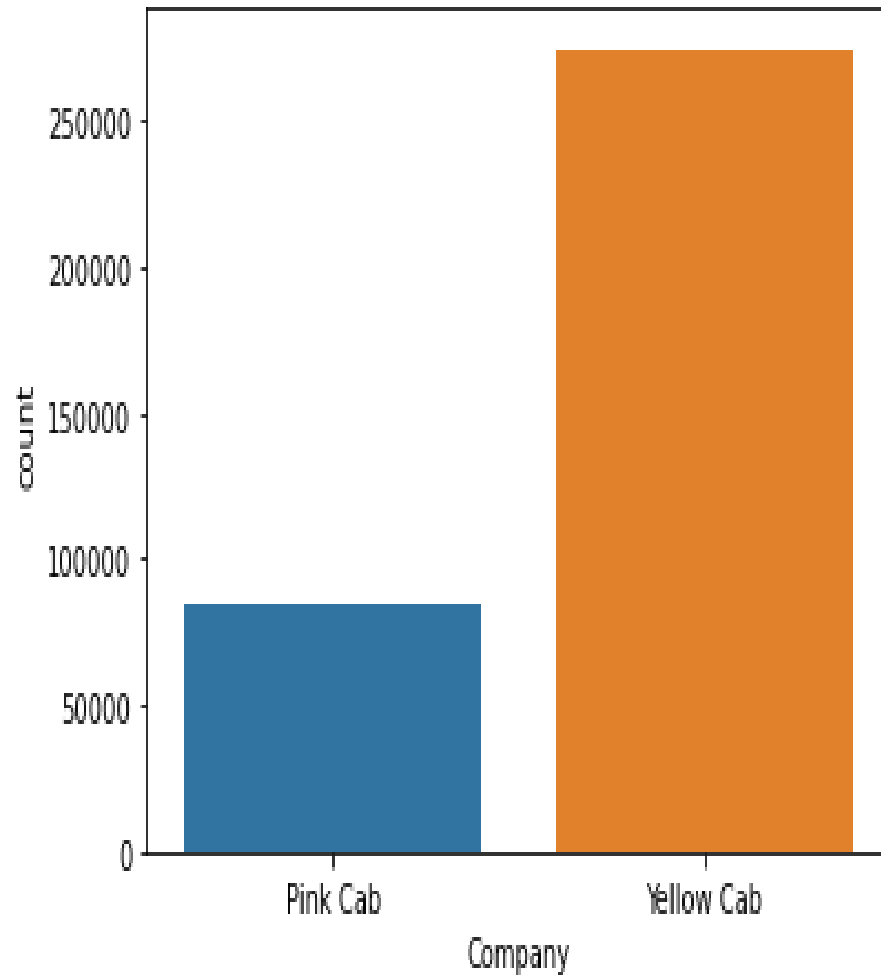
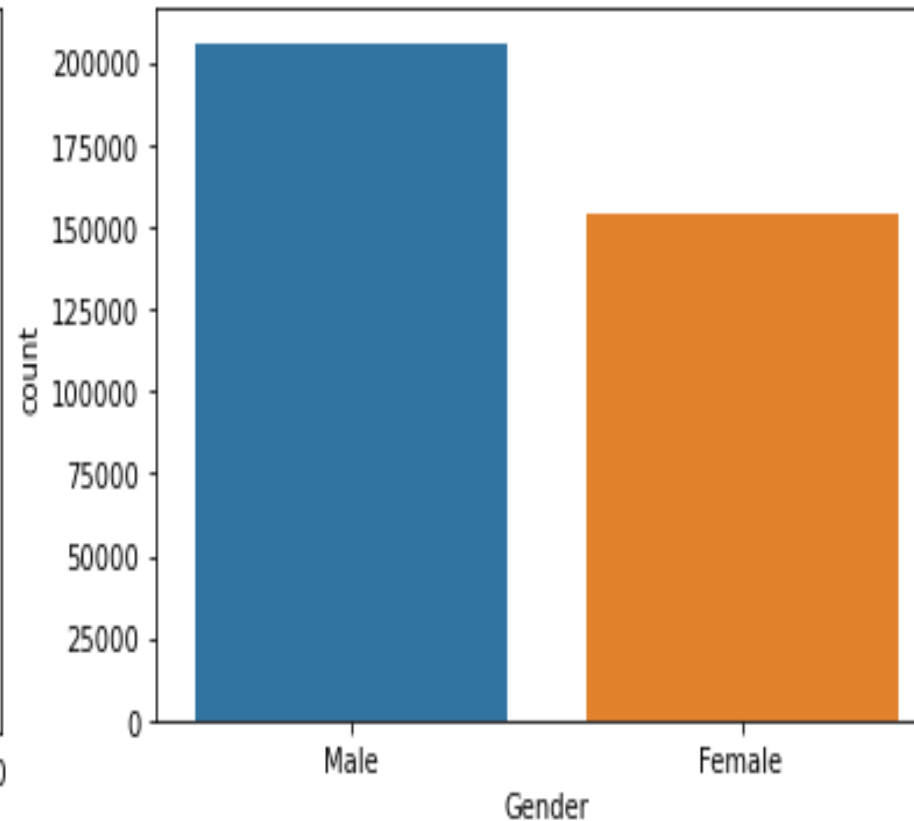Customer Analysis based on City

➢ Yellow cab has a larger customer base.

➢ New York City has the highest amount of cab users.

Univariate Plots Analysis

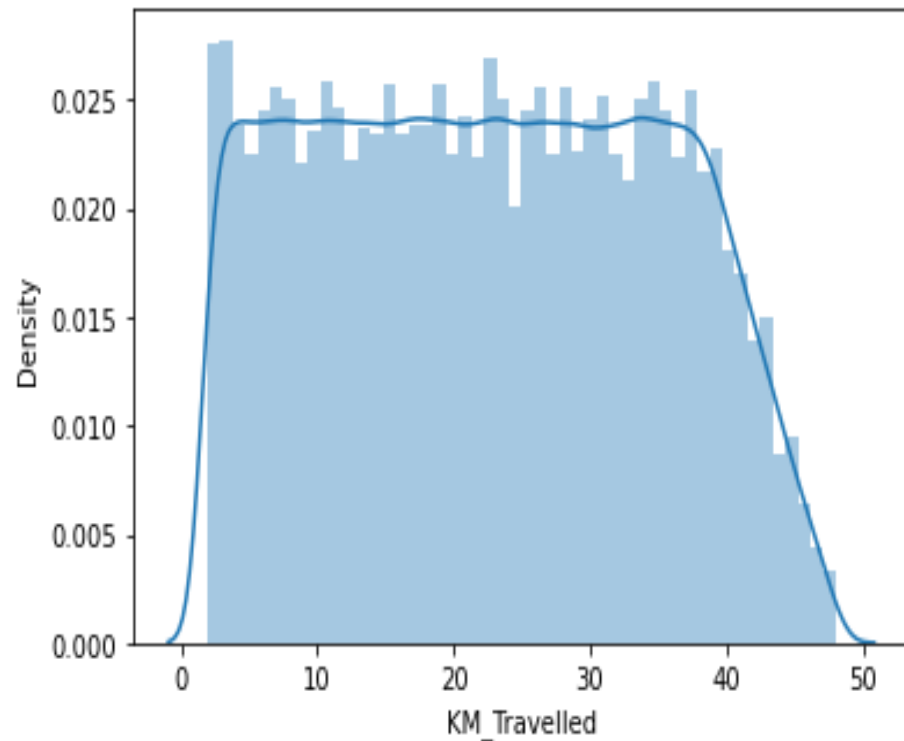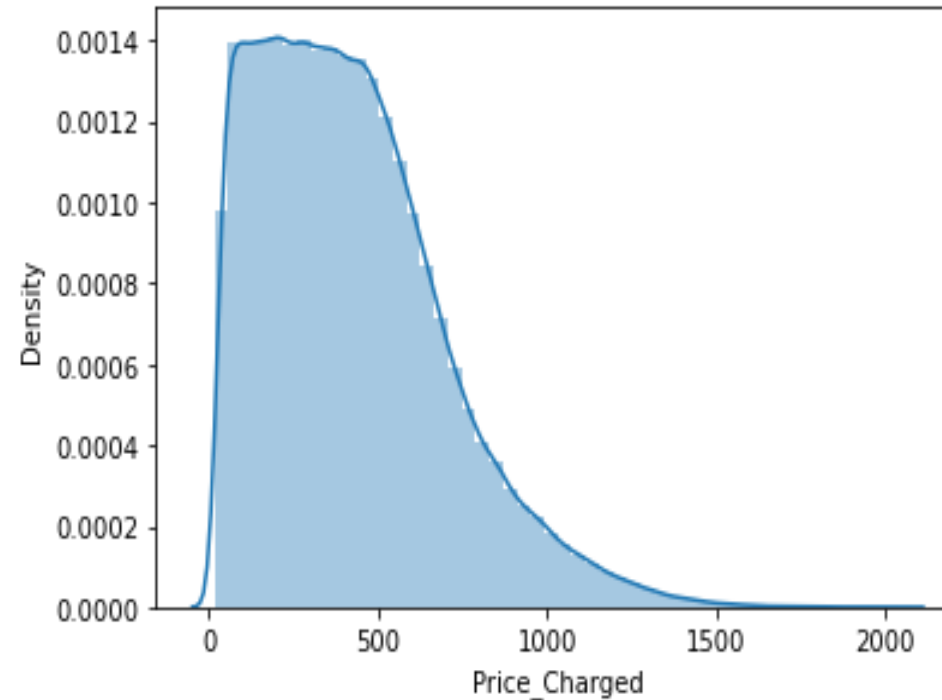Customer Analysis based on Age

Customer Analysis based on Gender

➤ It is evident that users are of a wide variety (All Ages). However the majority of users are between the ages of 20-40.

➤ Majority of customers are male.

Univariate Plots Analysis

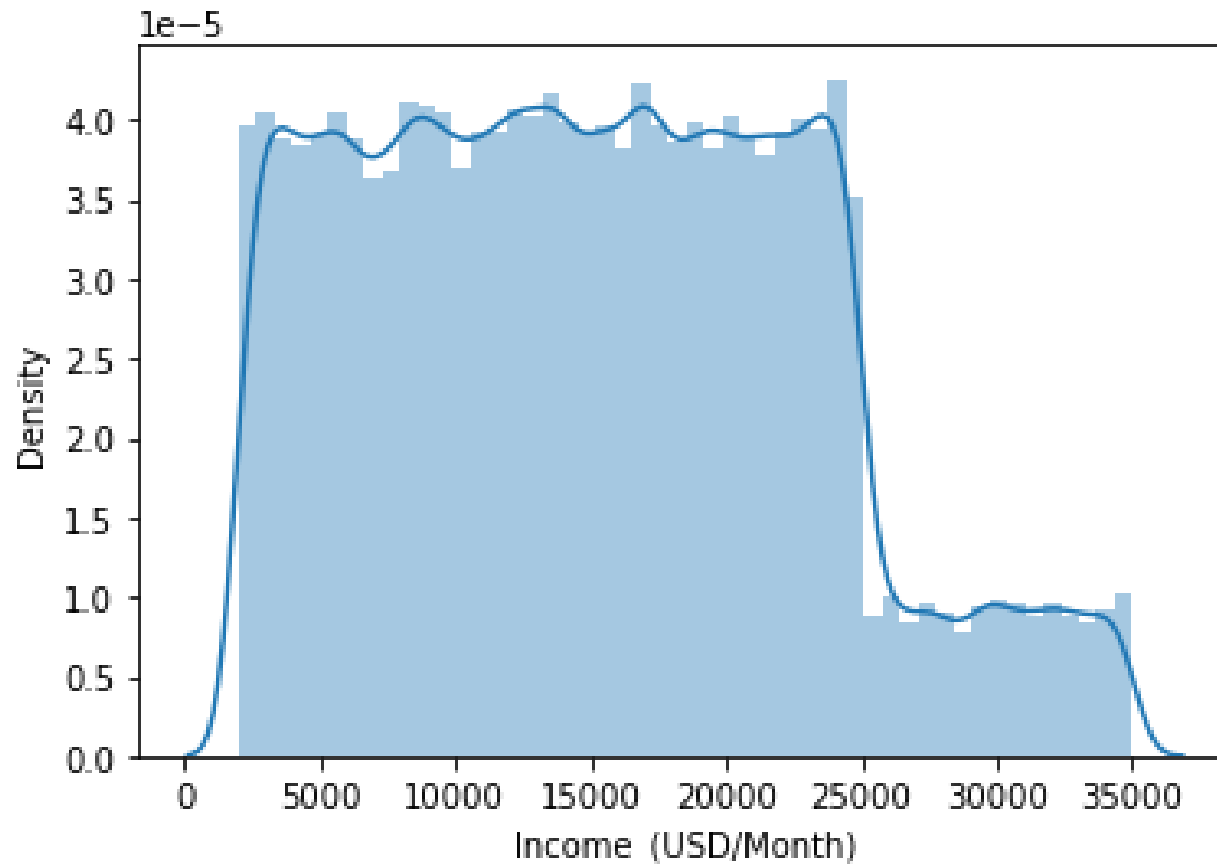KM Travelled by total customers

Price Charged ratio by both companies

➢ KM Travelled density is widely spread from around 2km to 49km.

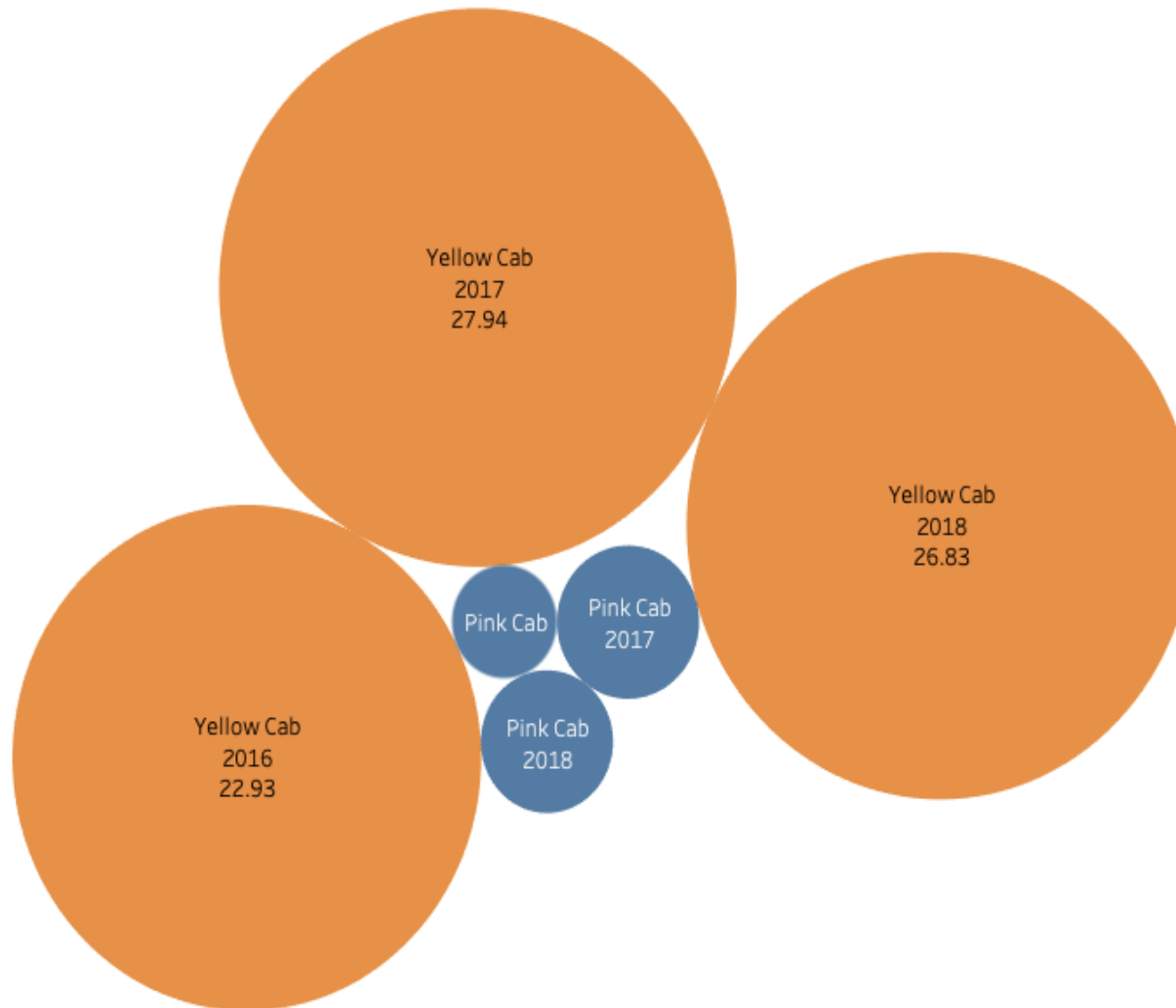➢ Price charged majority is in the range of 2 dollars to roughly around 800 dollars.

Univariate Plots Analysis

Income Distribution

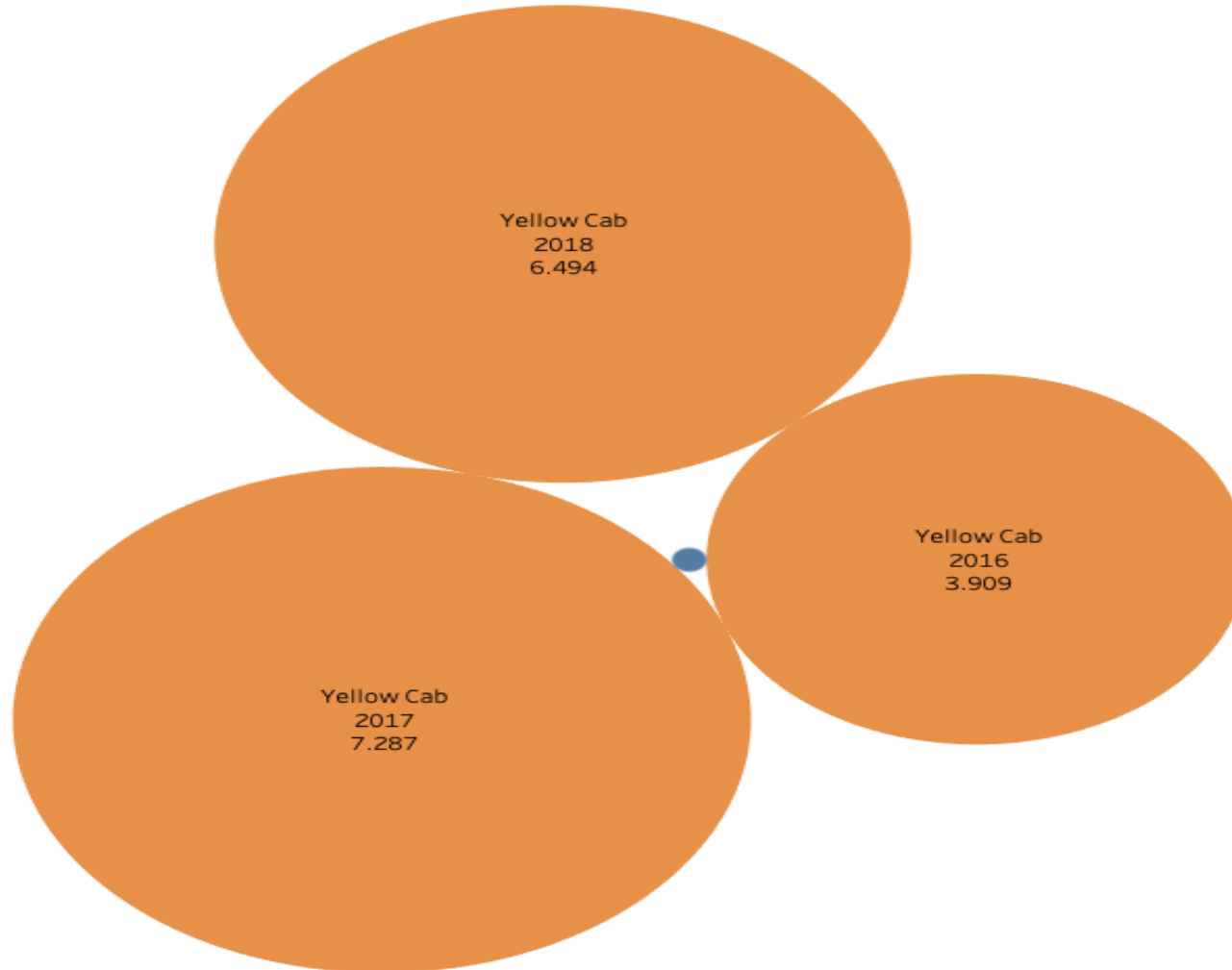➢ The income of the customers are mainly between the income bracket of around 200 to 24000.

Customer Retention

- This figure shows that both companies has a retention rate for customer using the same cab for 5 to 10 rides.

- Yellow Cab has the highest retention rate with an average of 25.9 over the three year period.

Customer Retention

- This figure shows that both companies has a retention rate for customer using the same cab for 10 rides or more.

- Yellow Cab has the highest retention rate with an average of 5.8967 over the three year period .

Yellow Cab
2018
6.494

Yellow Cab
2016
3.909

Yellow Cab
2017
7.287

Company

Pink Cab

Yellow Cab

Hypothesis Results

1. Yellow Cab has the most cab users on January 05, 2018.

2. Yes, margin proportionally increase with the increase in number of customers.

3. The attributes of the customer segments are Customer ID, Gender, Age and Income which are both user-defined numeric or discrete attributes.

4. According to the statistic the mean age of users are 35 years old.

5. There is a strongly positive correlation between price charged and cost of trip

Recommendation in support of investing in Yellow Cab

We have evaluated both the cab companies on following points and found Yellow cab better than Pink cab:

Based on the data analysis carried out here are the recommendation derived:

1.Earnings - While the Pink cab is making profits on rides the Yellow cab is almost making three times as much.

2.Customer - Based on the master data it shows that more customer ride with the Yellow cab on a constant basis.

3.City - Yellow cab is doing great in the top city (New York) which has the highest population.

Yellow Cab has the highest profit for the three year period, the highest retention rate and highest number of users. Therefore, it is safe to recommend Yellow Cab to XYZ as the right cab company to invest in for their G2M strategy.

Thank You ☺