

G2M Cab Investment Case Study

DATA GLACIER VIRTUAL INTERNSHIP

LISUM25-CHRISTEENA JOHN

A solid orange horizontal bar at the bottom of the slide.

BACKGROUND STUDY

INTRODUCTION-XYZ is a private company in the United States. Due to the spectacular rise in the Cab Industry in recent years and the presence of numerous significant players in the market, it is planning to invest in the Cab Industry, and as part of their Go-to-Market (G2M) strategy, they want to understand the market before making a final choice.

AIM & OBJECTIVE

AIM-To Provide actionable insights to assist XYZ firm in picking the best company for investment

To attain the aim different objectives have selected.

- 1.Loading the libraries
- 2.Dataset Understanding
- 3.Data Preprocessing
- 4.Exploratory Data Analysis

DATA EXPLORATION

- 24 Features (9 Derived Features Included)
- Data was collected from 2016-01-31 to 2018-12-31.
- 355,032 total data points
- 4 Datasets including:
 1. **Cab_Data.csv** - file includes details of transaction for 2 cab companies
 2. **Customer_ID.csv** - mapping table that contains a unique identifier which links the customer's demographic details
 3. **Transaction_ID.csv** - a mapping table that contains transaction to customer mapping and payment mode
 4. **City.csv** -file contains list of US cities, their population and number of cab users

METHODOLOGY

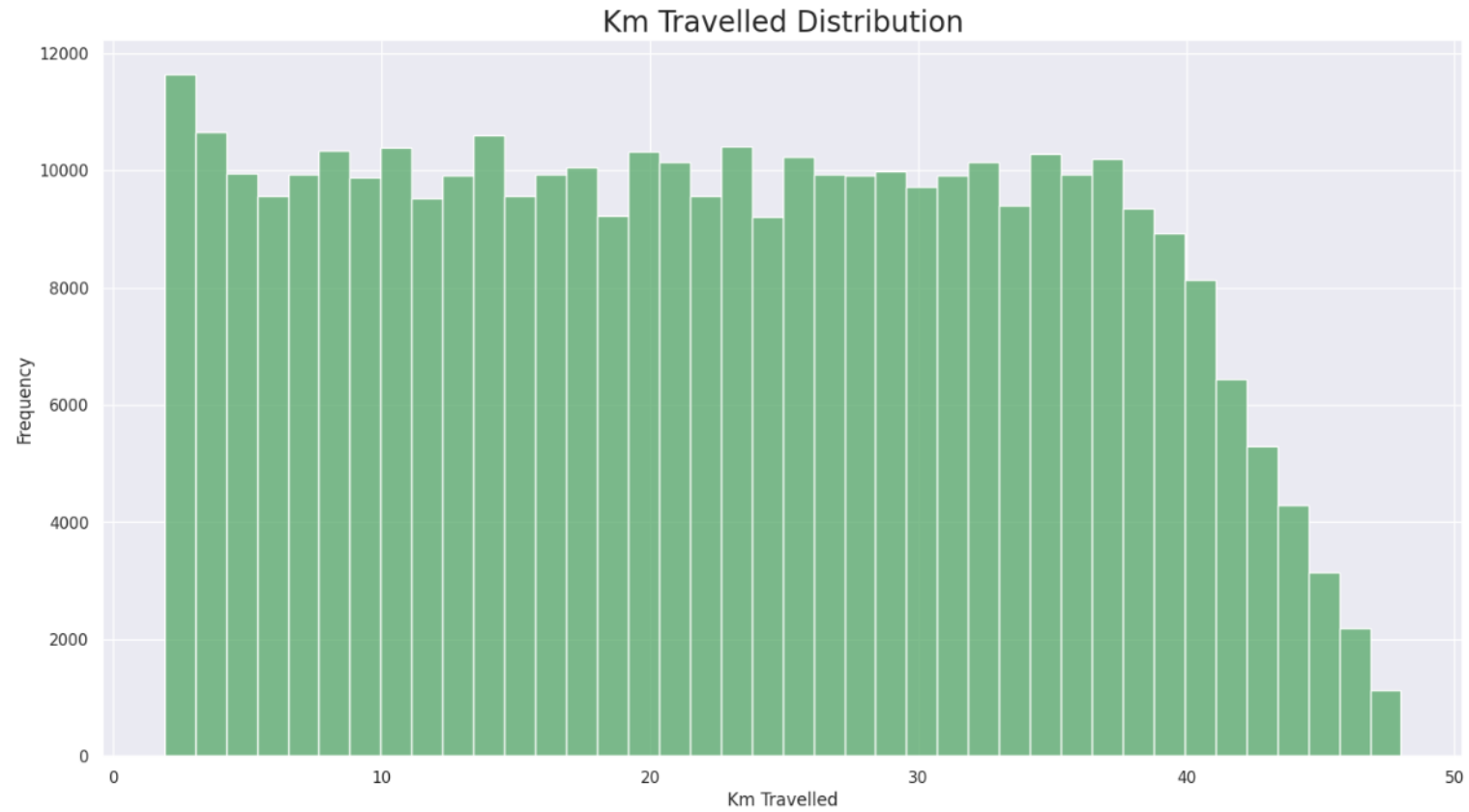
- First the libraries are loaded to the google Collab
- Then the 4 datasets are loaded and understanding of each dataset
- Data Preprocessing is the third step. The steps include:
 1. MASTERDATA CREATION
 2. Data quality assessment
 3. Data Cleaning
 4. Checking for Missing Values
 5. Identifying single-valued columns
 6. Checking duplicates
 7. Finding Outliers
 8. Data transformation

METHODOLOGY

After the data preprocessing Explanatory Data Analysis is done .The Exploratory Data Analysis Include:

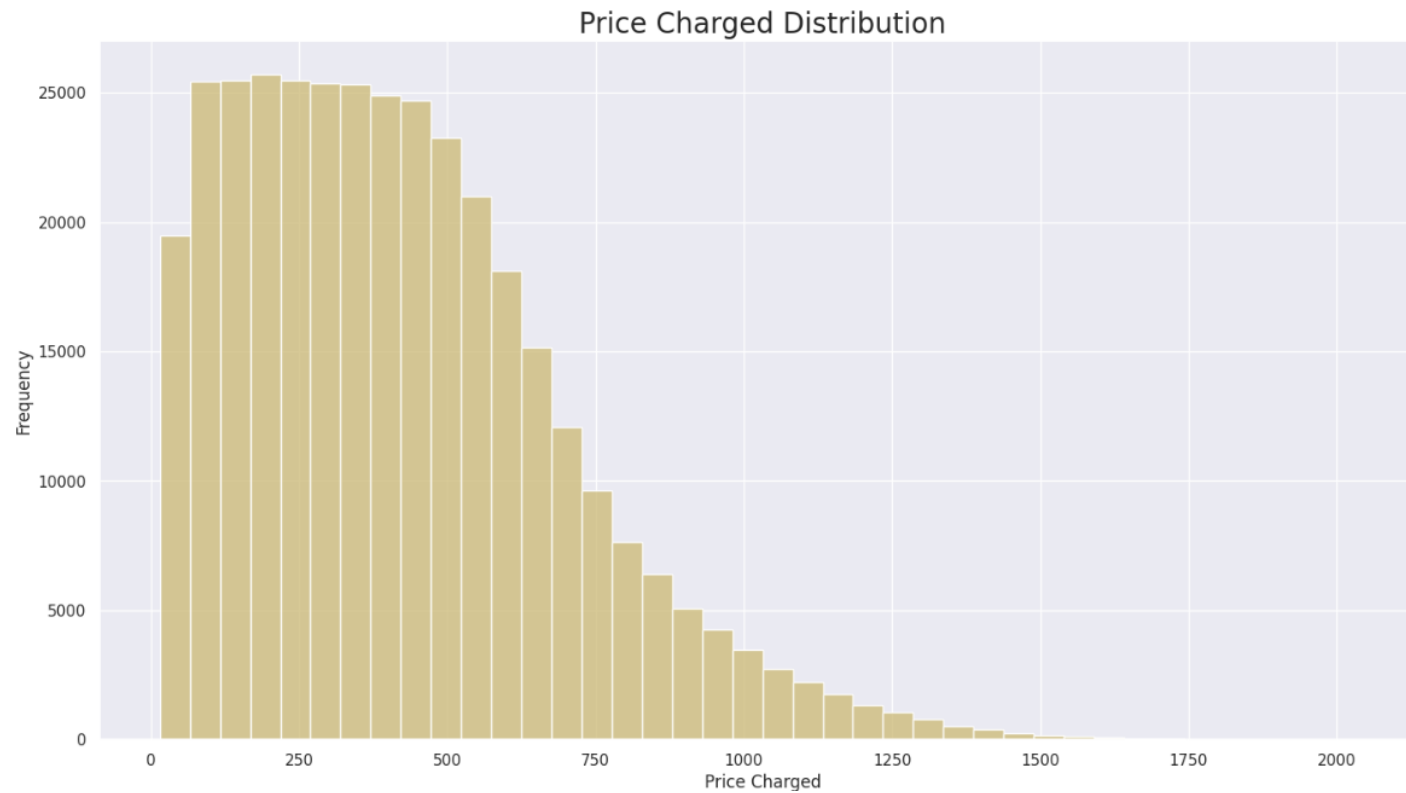
1. Features Analysis
2. Kilometers Travelled Distribution
3. Price Distribution
4. Cost of Trip Distribution
5. Age Distribution
6. COST per KM
7. Customer Payment Mode preference
8. Gender Preference for Cab Booking
9. Time series
10. Numerical features
11. Client Analysis
12. Profit Analysis

RESULTS



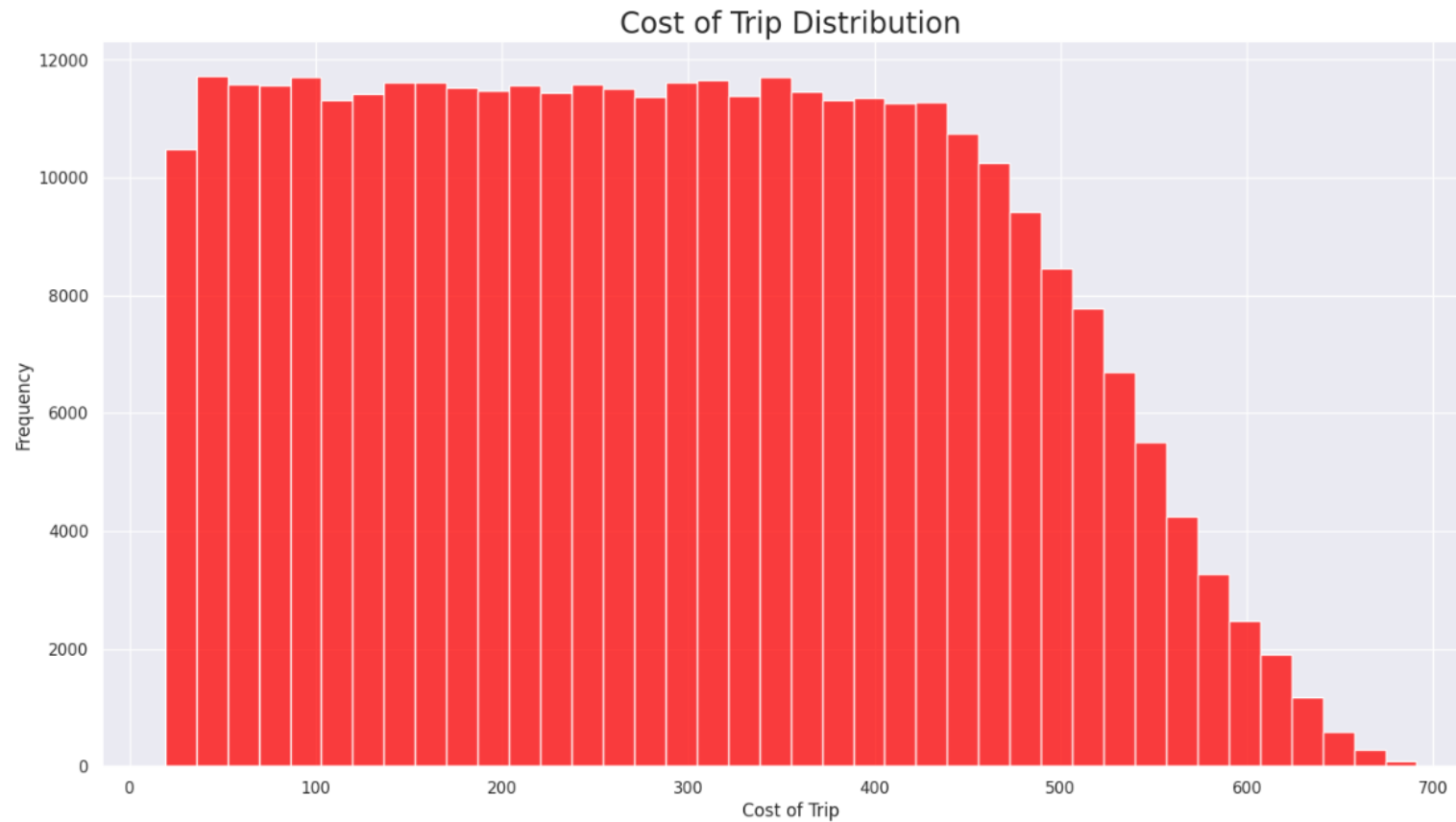
After 40KM there is gradual linear decrease in number of rides

RESULTS



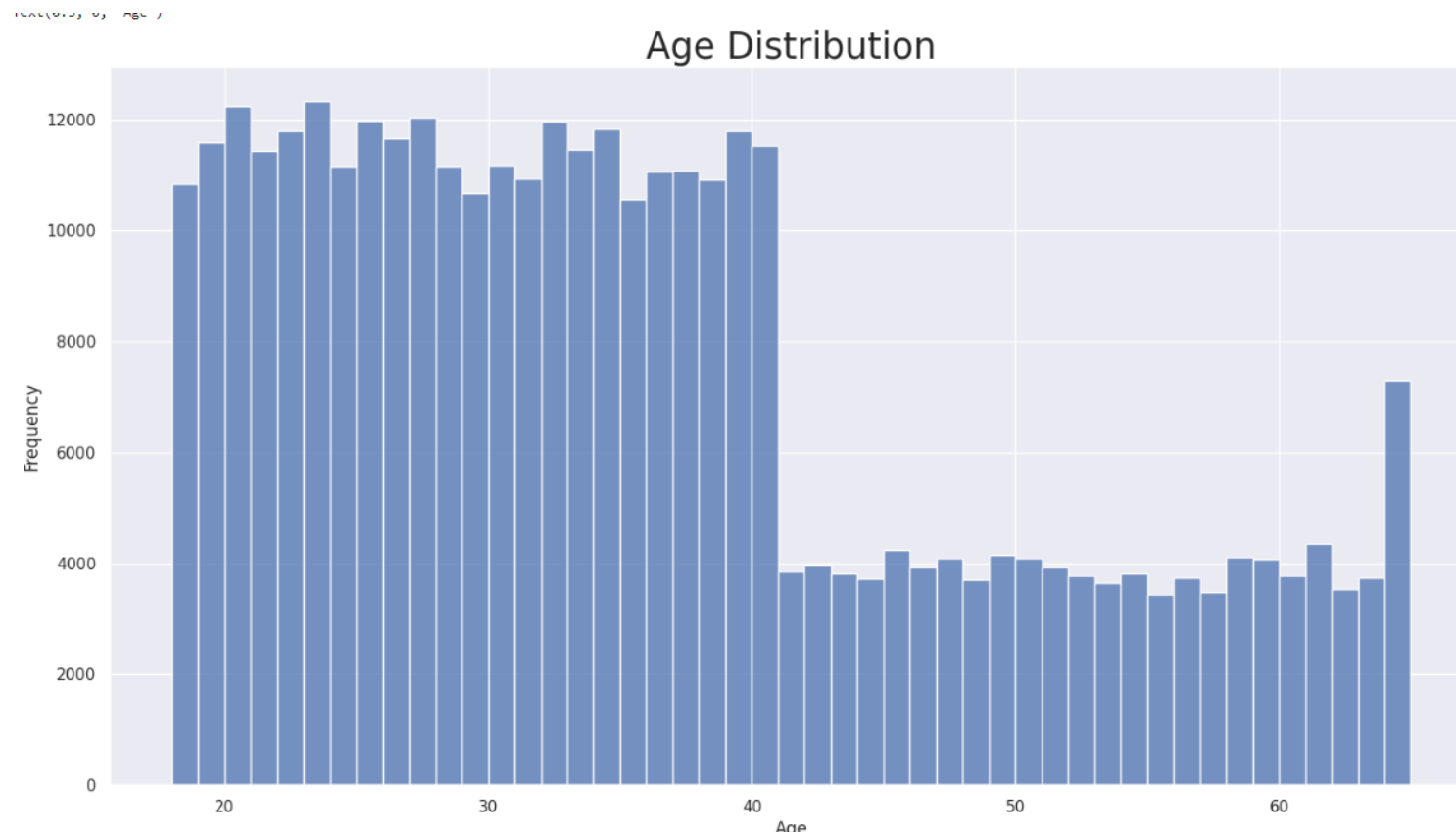
After 400, there is an exponential decrease in number of rides

RESULTS



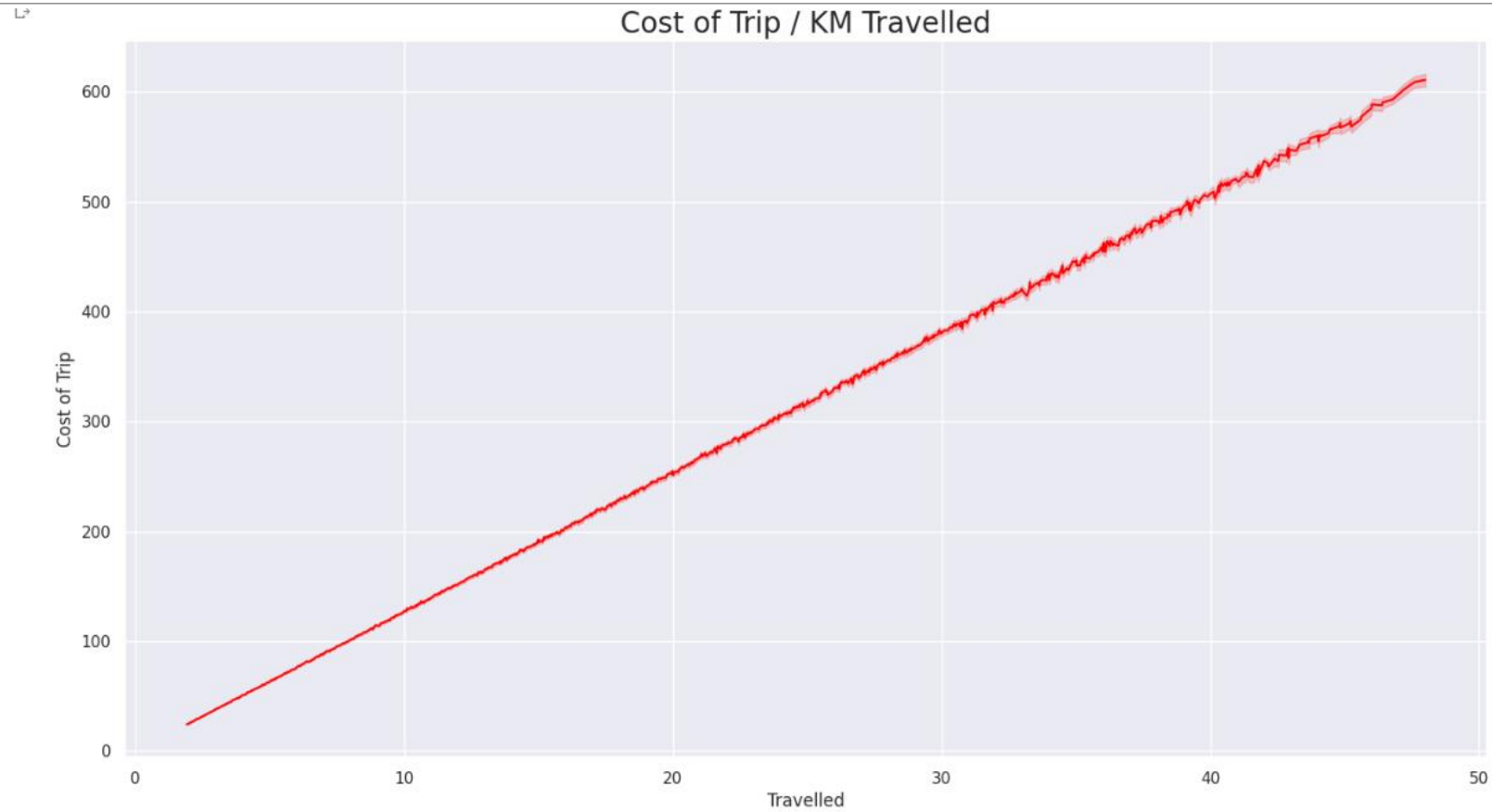
After around 420 there is linear decrease in number of rides till it reaches 0 around 700

RESULTS



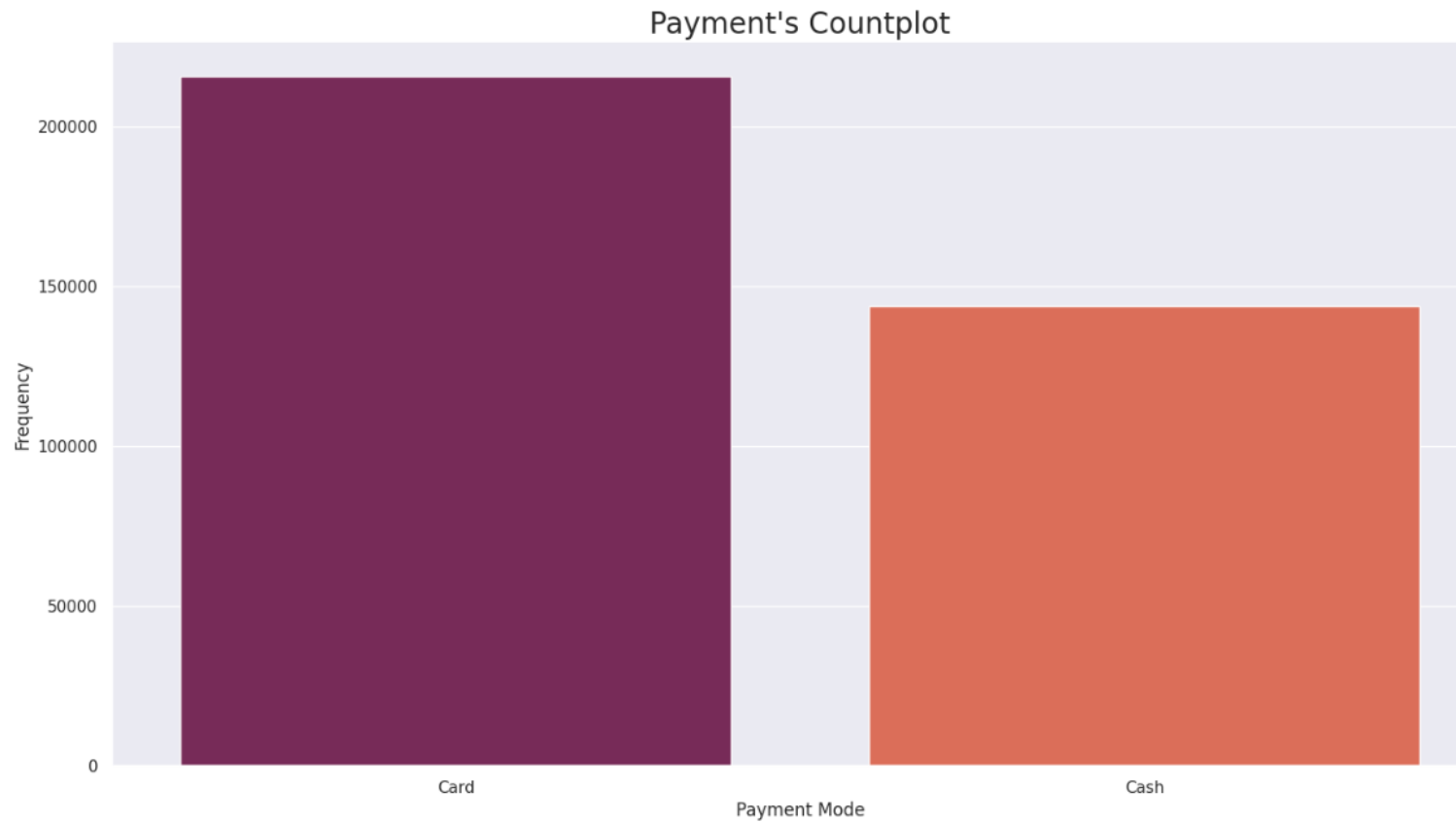
After the age of 40 there is a sudden drop in number of passengers. People below 40 years tend to travel a lot

RESULTS



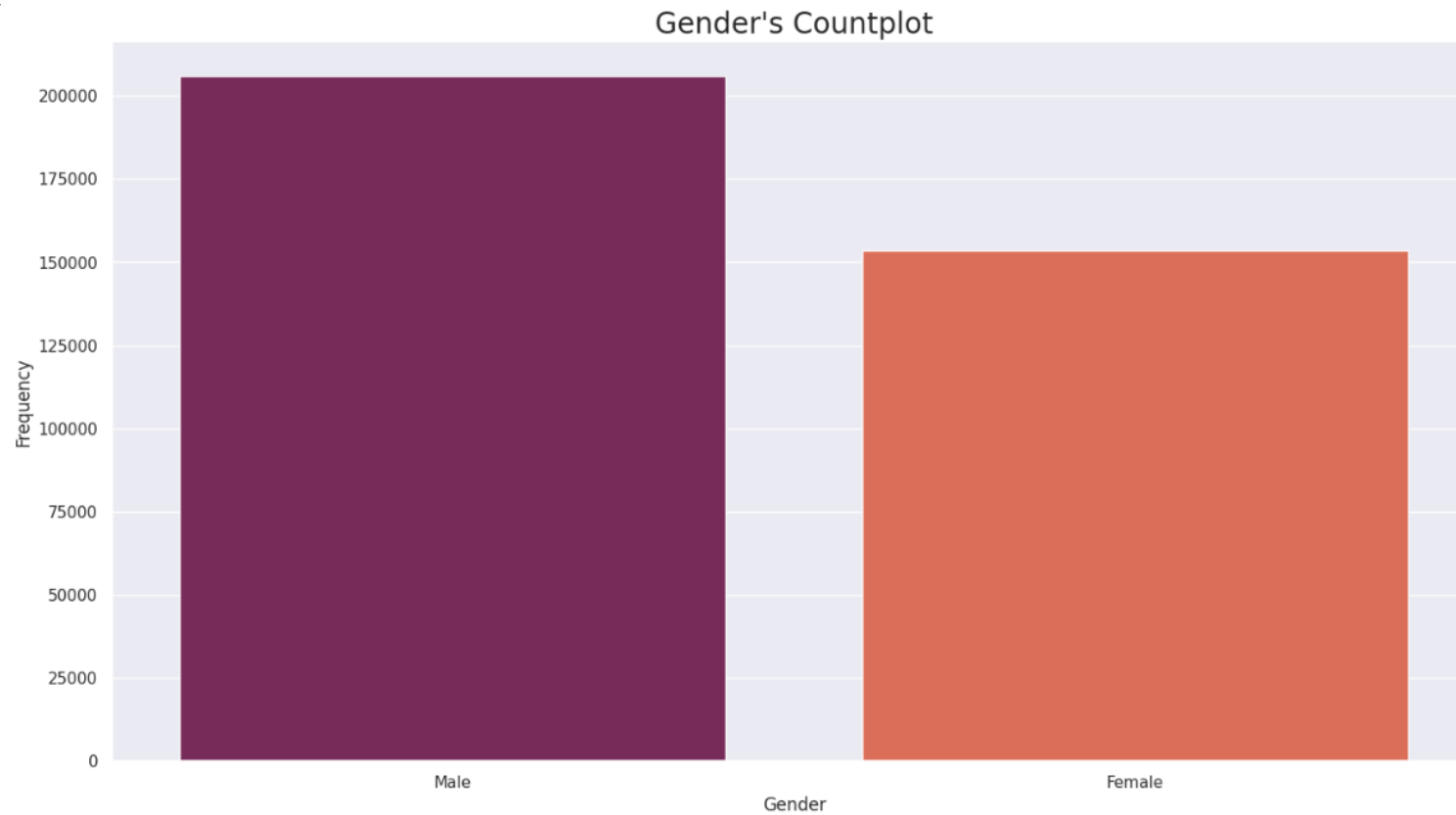
Cost of the trip tends to increase with the kilometers travelling

RESULTS



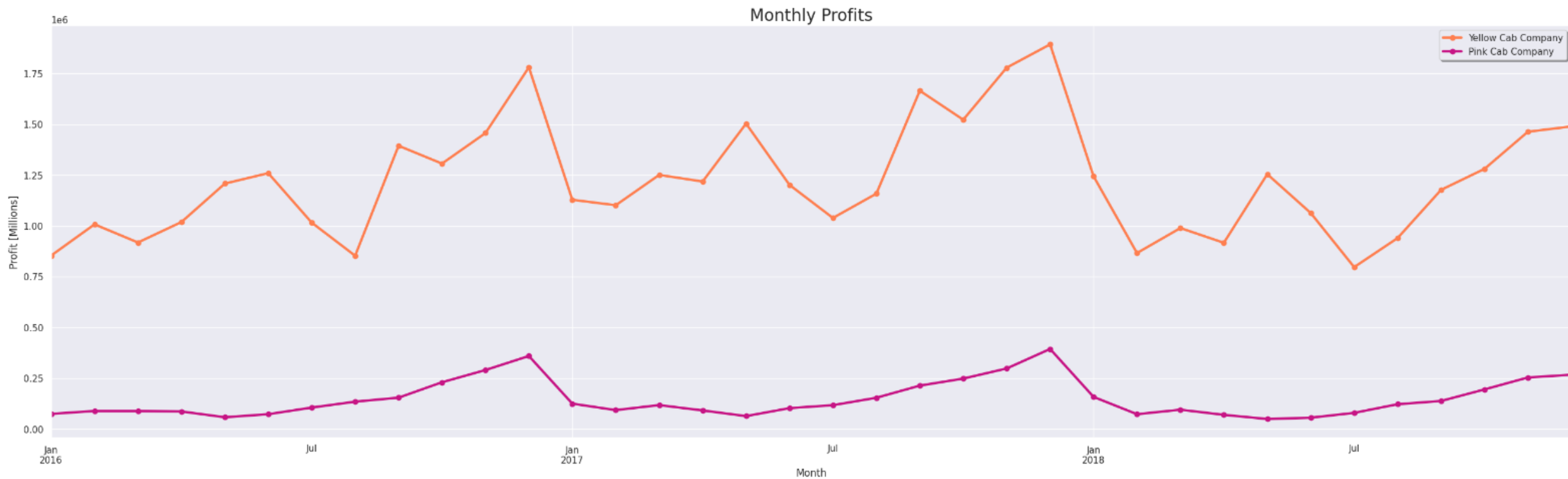
Most of the customers preferred for card payment

RESULTS



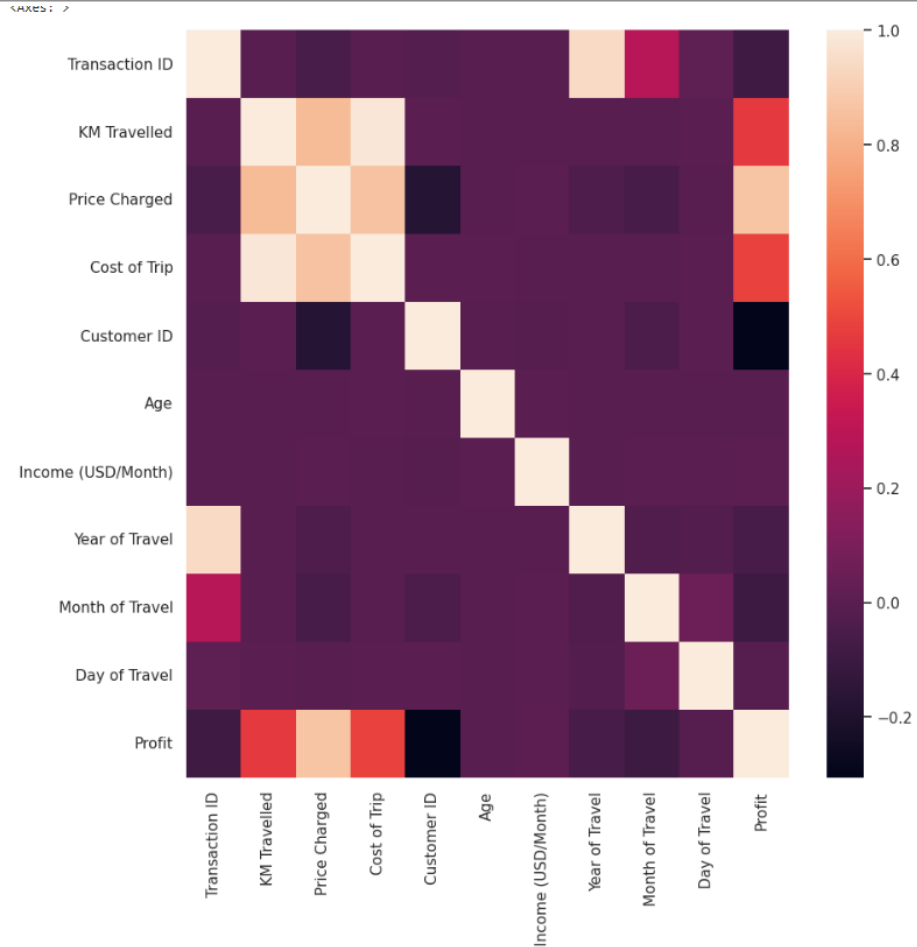
Male customers are more preferred for booking cab than female customers

RESULTS



YellowCab Company's earnings are more stable, with fluctuations of 23,08%, while those of the PinkCab Company's vary in the order of 61,22%

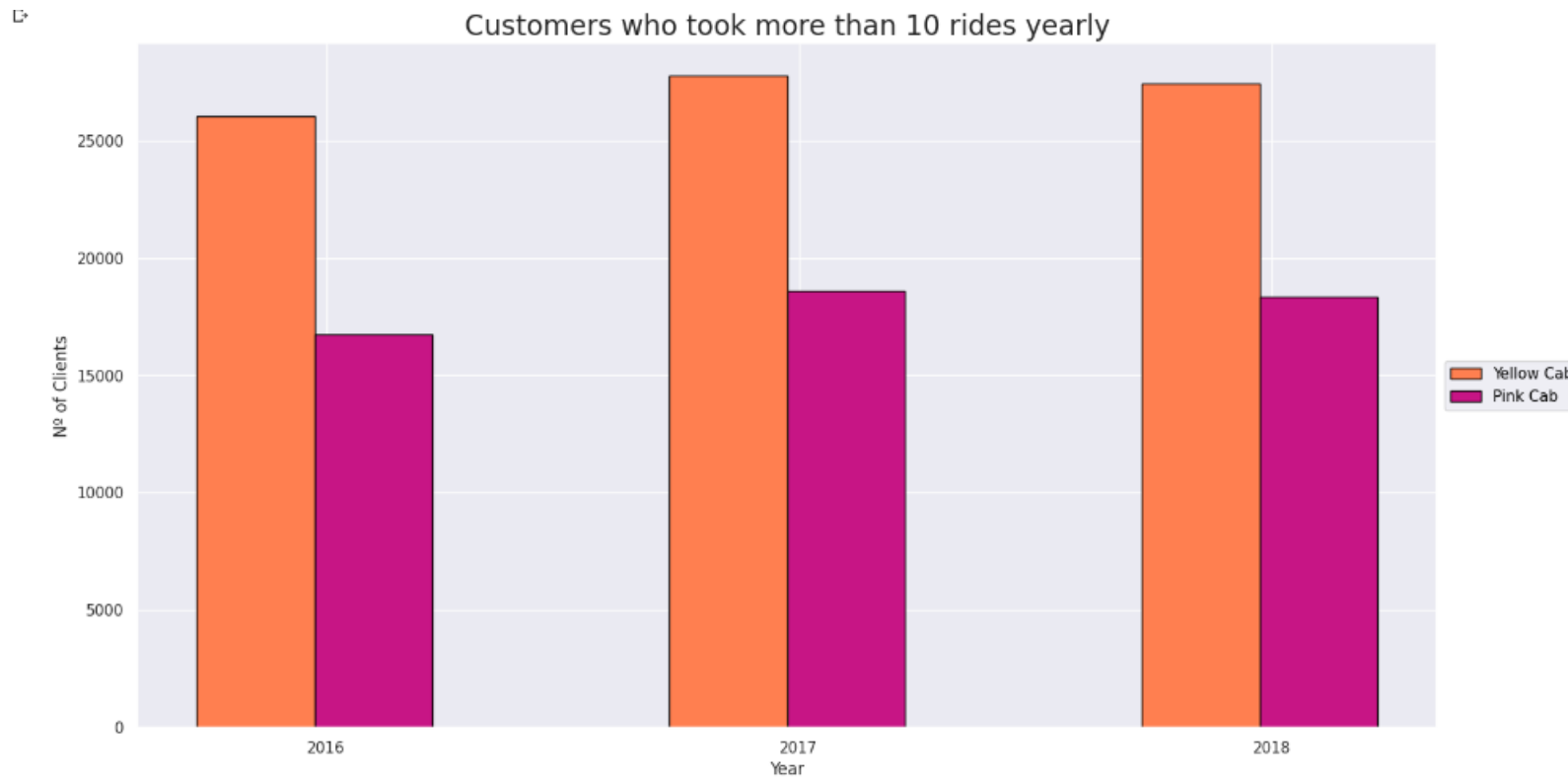
RESULTS



List of highly correlated attributes

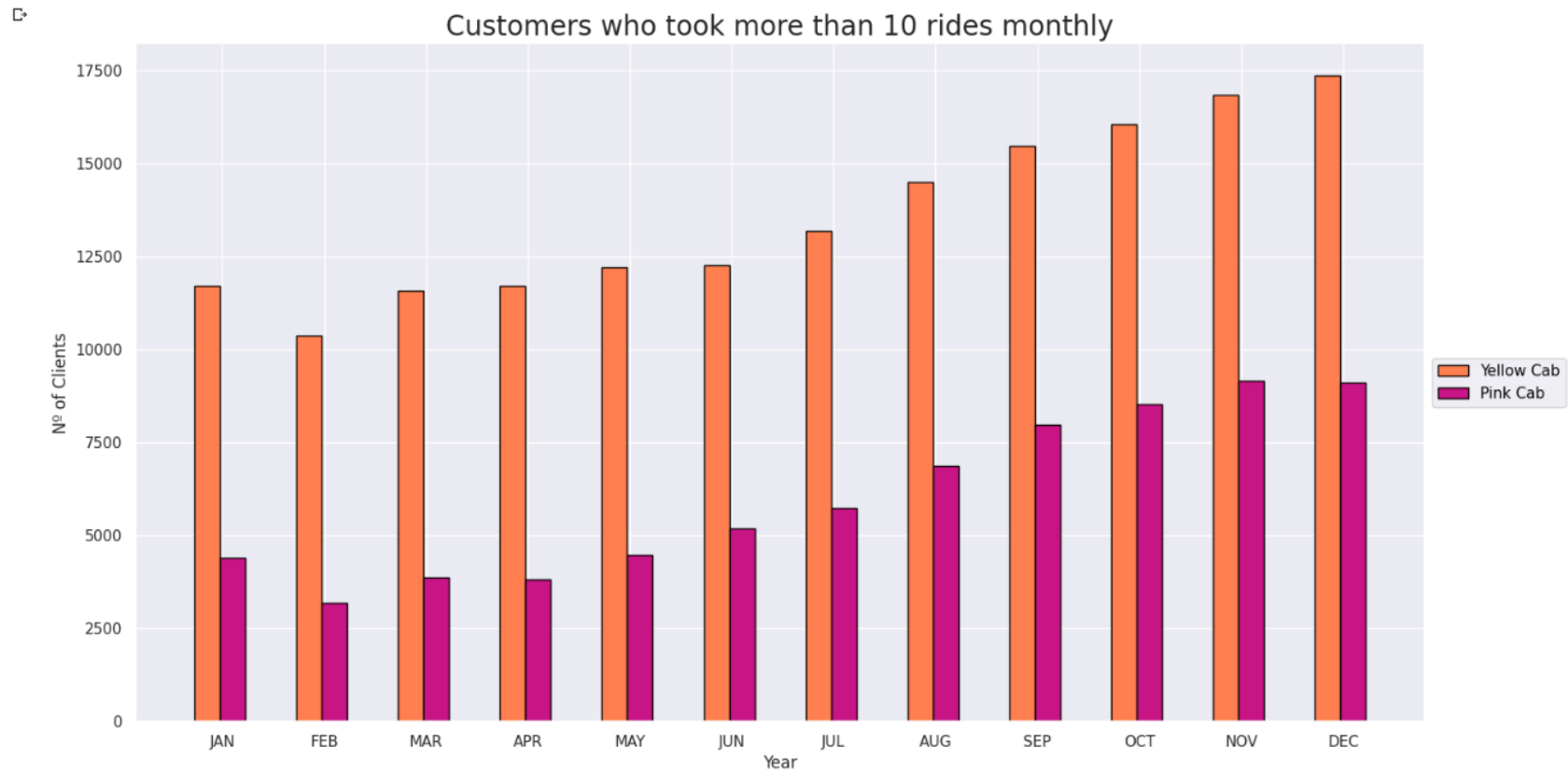
1. KM_Travelled and Price_Charged
2. KM_Travelled and Cost_of_Trip
3. Price_Charged and Cost_of_Trip

RESULTS



Customers who took more than 10 rides yearly booked for Yellow Cab

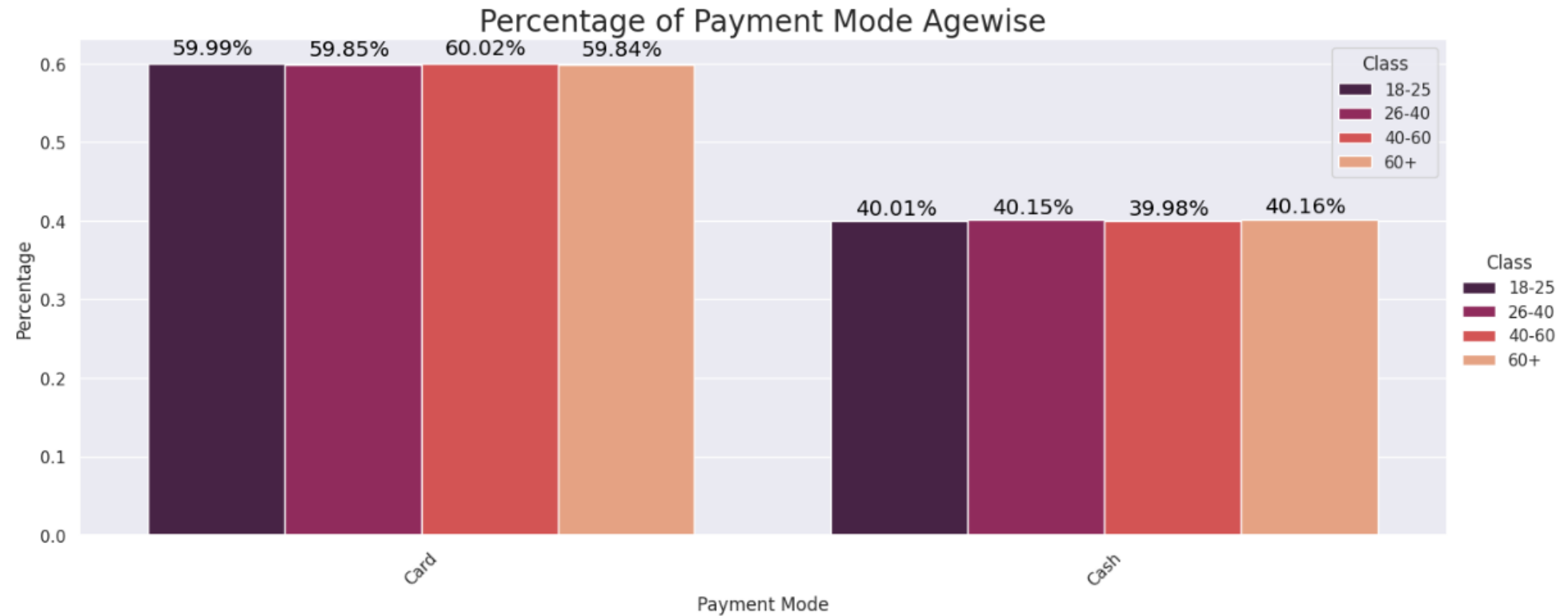
RESULTS



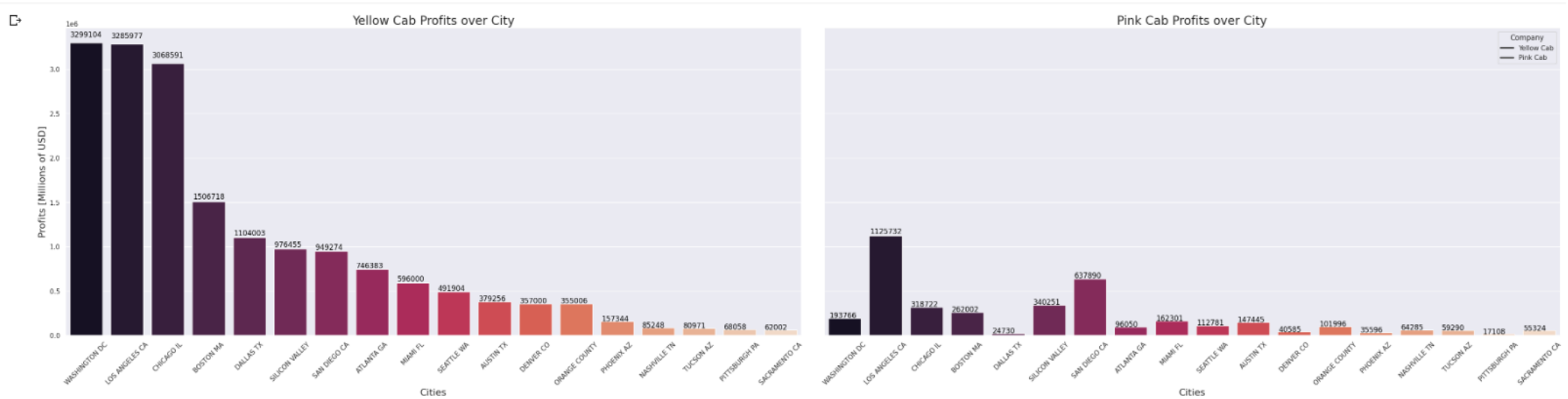
Customers who took more than 10 rides monthly are also booked for Yellow cab

RESULTS

<Figure Size 1500X/100 WITH 0 AXES>

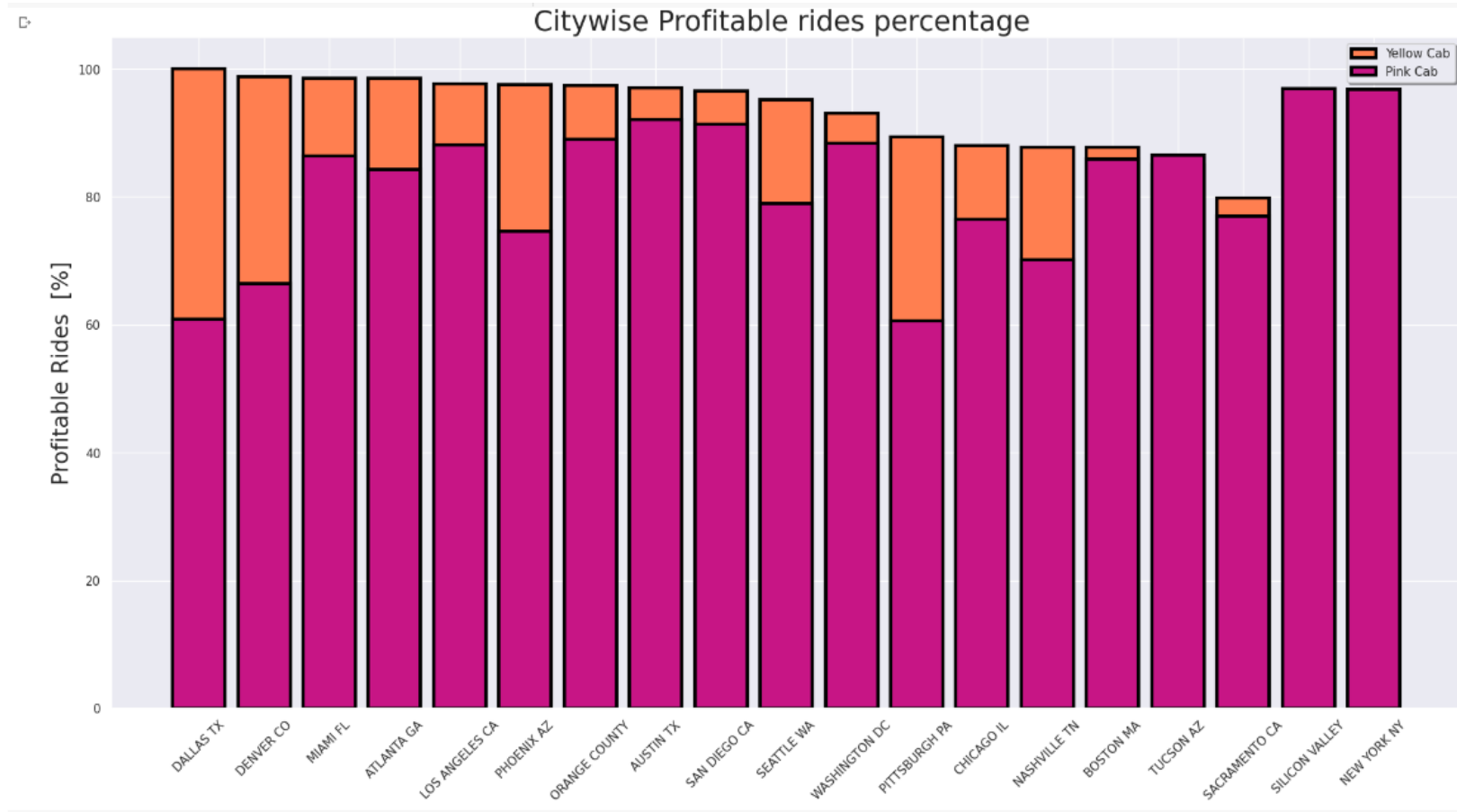


Around 60 % of customers card payment mode and only 40% people depend on cash payment mode on all age group



WASHINGTON DC provide the highest profit for yellow cab and LOS ANGELS CA provide the highest profit for pink cab

RESULTS



RESULTS



Total Profit	
Company	
Pink Cab	5307328.0
Yellow Cab	44020373.0

YellowCab's profits over the last 3 years are eight times higher than PinkCab's Company

CONCLUSIONS

Yellow Cab is used more than Pink Cab. 76.4% of the rides are by Yellow Cab

November and December (10 and 12) have the most rides. Jan and Feb being the least

2018 has the highest cab rides but the 'Year of Travel' column approximately has fairly distributed data

NEW YORK NY has the highest number of rides vs PITTSBURGH PA with lowest

After 40KM there is gradual linear decrease in number of rides

Both the companies have same average KM travelled

On an average Pink cab contributes to 23% in total cab count for all three years

Mean Profit for Pink Cab on an Average for the three years is around 17-18

Mean Profit for Yellow Cab on an Average for the three years is around 28-31

Highest Profit Percentage of Pink Cab was for the months- March and December, lowest for the months- June and July

Highest Profit Percentage of Yellow Cab was for the months- May, Feb Jan and June, lowest for the months- August and July

THANK YOU