G2M insight for Cab Investment firm

Darshan Patil

Background

XYZ, a private US-based firm, has recognized the significant growth experienced by the cab industry in recent years, as well as the presence of numerous key players in the market. As part of their Go-to-Market (G2M) strategy, XYZ is considering investing in the cab industry and seeks to gain a comprehensive understanding of the market landscape before making a final decision.

Datasets

Cab_Data.csv – This dataset contains transaction details for two competing cab companies.

Customer_ID.csv – A mapping table featuring a unique identifier that links customers' demographic information.

Transaction_ID.csv – A mapping table that connects transactions to customers and payment methods.

City.csv – A dataset that includes a list of US cities along with their respective populations and the number of cab users.

Using these following datasets, master data is created by merging the files all the columns to find the relationship between the columns.

Correlation Matrix

A correlation coefficient of -1 or 1 generally signifies a strong linear association between two variables, while a coefficient of 0 indicates an absence of linear connection. It is crucial to remember that correlation does not equate to causation, and other factors might influence the relationship between variables. Additionally, the Pearson correlation coefficient presumes a linear relationship between variables and might not be suitable for evaluating non-linear relationships.

Histogram

Utilizing the information encapsulated within the correlation matrix, I have devised a histogram to visually elucidate the data contained within the column.

Which Company has maximum cab user?

In order to come to this conclusion, the following steps had to be performed using python:

- Group the data by company and count the number of unique customers
- Sort the results in descending order
- Print the company with the maximum cab users

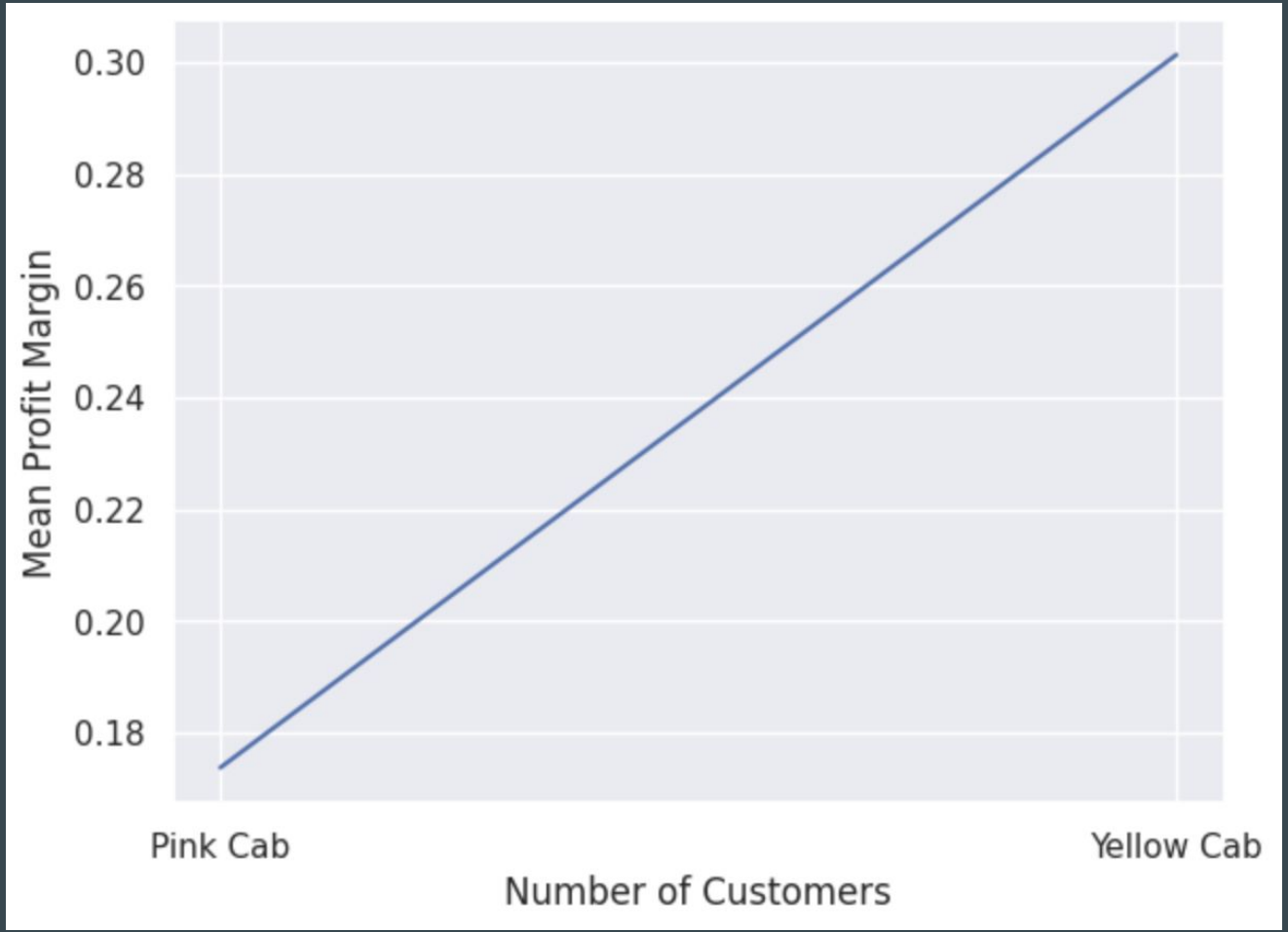
My deduction from this section of code is that Yellow Cab has the most users.

Does margin proportionally increases with increase of customer?

In order to come to this conclusion, the following steps had to be performed using python:

- Calculate the profit margin
- Group the data by the number of customers and calculate the mean profit margin for each group
- Plot the mean profit margin against the number of customers

Line Graph was used to visually represent the insight from the data found. (Visual on next slide)


0.30
0.28
0.26
0.24
0.22
0.20
0.18
Mean Profit Margin
Pink Cab
Yellow Cab
Number of Customers

What are the attributes of customer segments?

In order to come to this conclusion, the following steps had to be performed using pythons:

- Group the data by payment mode and perform aggregations
- Calculate the profit margin for each payment mode
- Print the grouped data

A data frame was used to visually represent the insight from the data found:

Payment_Mode	Price Charged	Cost of Trip	KM Travelled	Profit Margin
Card	91246079.25	6.168432e+07	4863981.15	0.323978
Cash	60936059.27	4.117012e+07	3246509.43	0.324372