Exploratory Data Analysis

Project : G2M insight for Cab Investment firm

Submitted by : Amima Shifa

Date : 22/10/2022

Agenda

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Problem Statement

XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Objective:

To help them identify the right company to make their investment by providing them actionable insights

Approach

The approach is divided into the following sections:

- Data Understanding
- Data Evaluation
- Data Analysis
- Hypothesis creation and investigation

Data Summary

- 24 Features(including 9 derived features)
- Timeframe of the data: 2016-01-31 to 2018-12-31
- Total data points :355,032

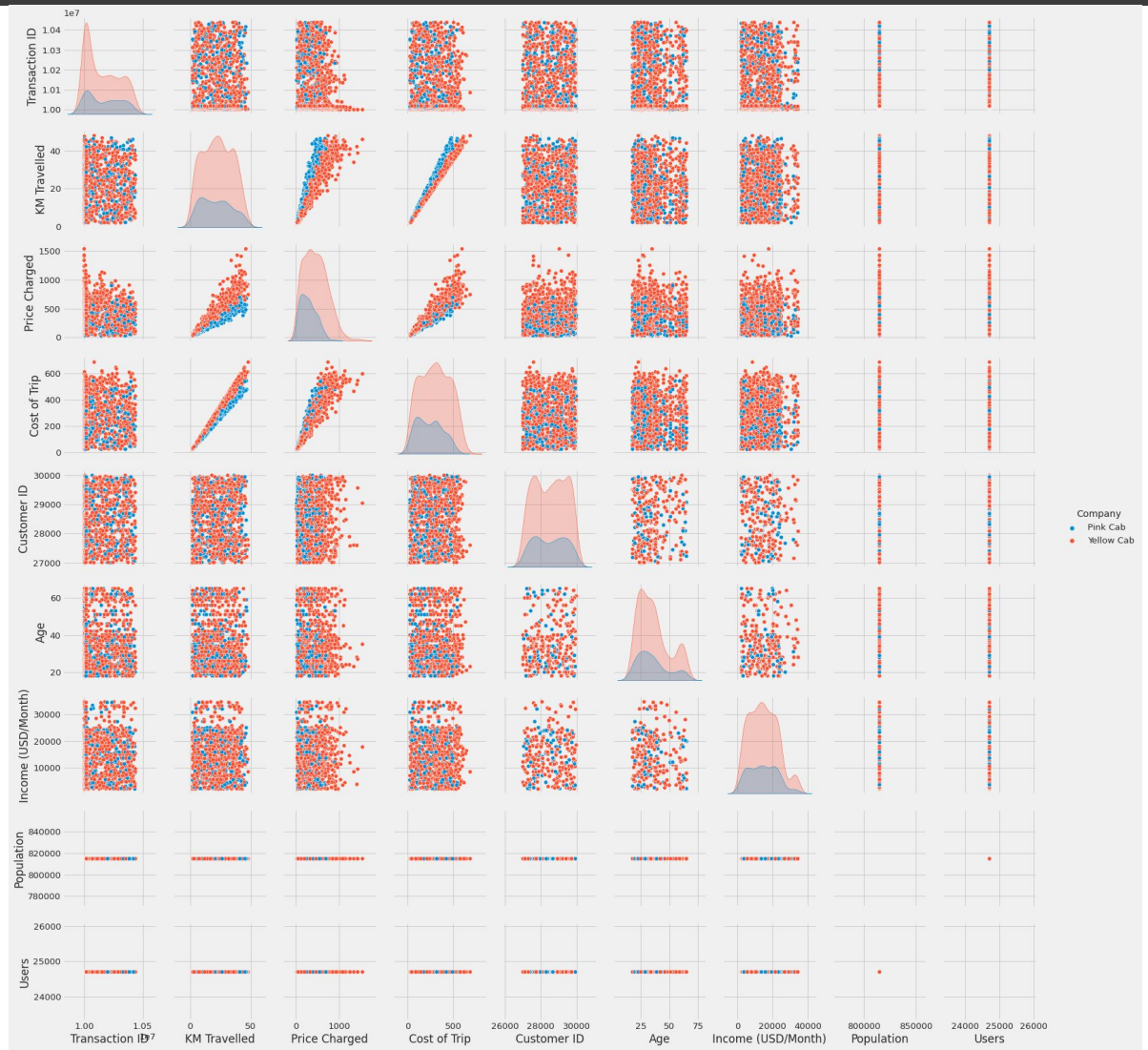
Assumptions:

- 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.
- Users feature of city dataset is treated as number of cab users in the city. We have assumed that this can be other cab users as well (including Yellow and Pink cab)

EDA

Data Glacier
Your Deep Learning Partner

Relationship between variables



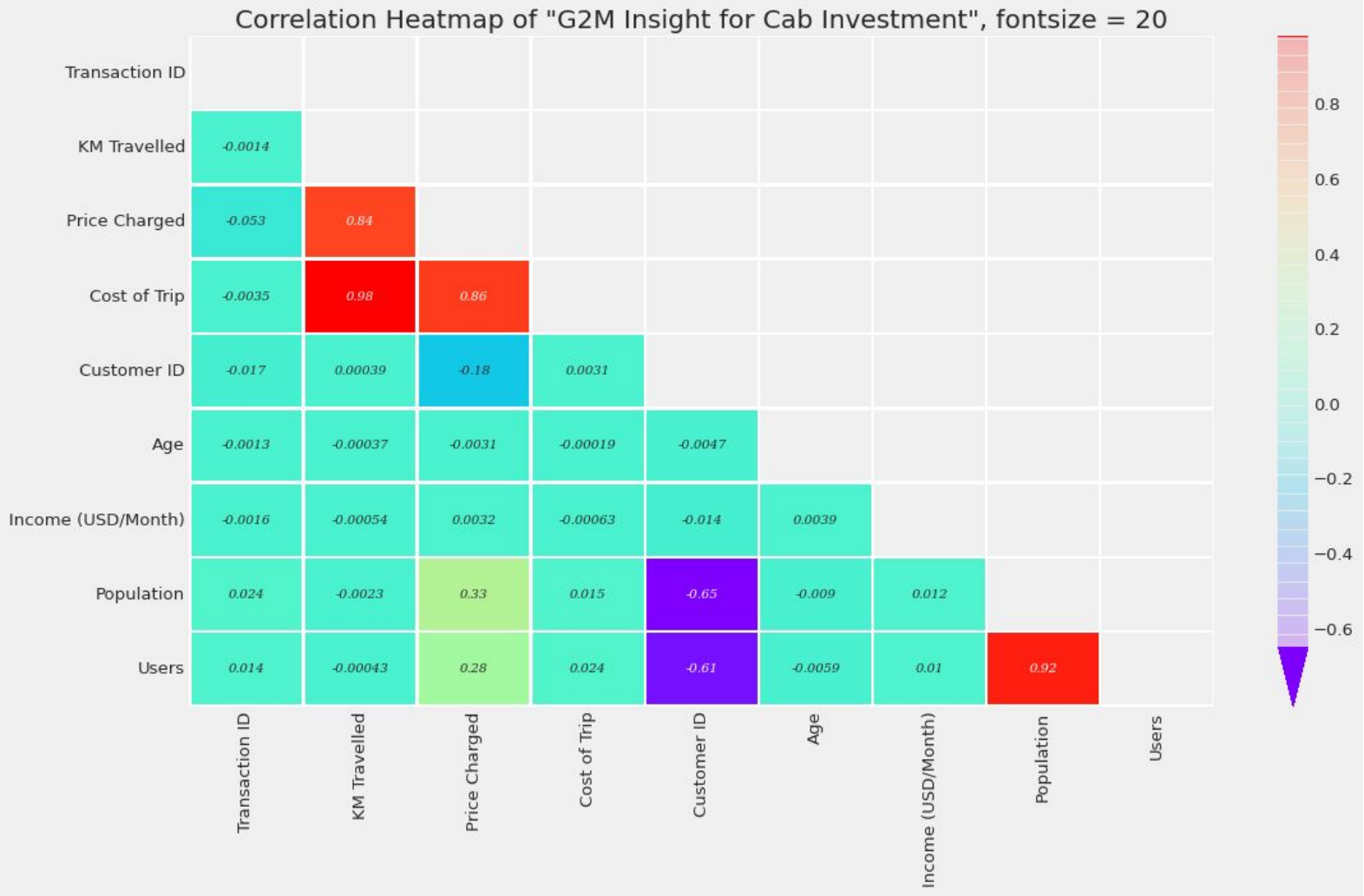
Data Glacier
Your Deep Learning Partner

Correlation Analysis

	Transaction ID	KM Travelled	Price Charged	Cost of Trip	Customer ID	Age	Income (USD/Month)	Population	Users
Transaction ID	1.000000	-0.001429	-0.052902	-0.003462	-0.016912	-0.001267	-0.001570	0.023868	0.013526
KM Travelled	-0.001429	1.000000	0.835753	0.981848	0.000389	-0.000369	-0.000544	-0.002311	-0.000428
Price Charged	-0.052902	0.835753	1.000000	0.859812	-0.177324	-0.003084	0.003228	0.326589	0.281061
Cost of Trip	-0.003462	0.981848	0.859812	1.000000	0.003077	-0.000189	-0.000633	0.015108	0.023628
Customer ID	-0.016912	0.000389	-0.177324	0.003077	1.000000	-0.004735	-0.013608	-0.647052	-0.610742
Age	-0.001267	-0.000369	-0.003084	-0.000189	-0.004735	1.000000	0.003907	-0.009002	-0.005906
Income (USD/Month)	-0.001570	-0.000544	0.003228	-0.000633	-0.013608	0.003907	1.000000	0.011868	0.010464
Population	0.023868	-0.002311	0.326589	0.015108	-0.647052	-0.009002	0.011868	1.000000	0.915490
Users	0.013526	-0.000428	0.281061	0.023628	-0.610742	-0.005906	0.010464	0.915490	1.000000

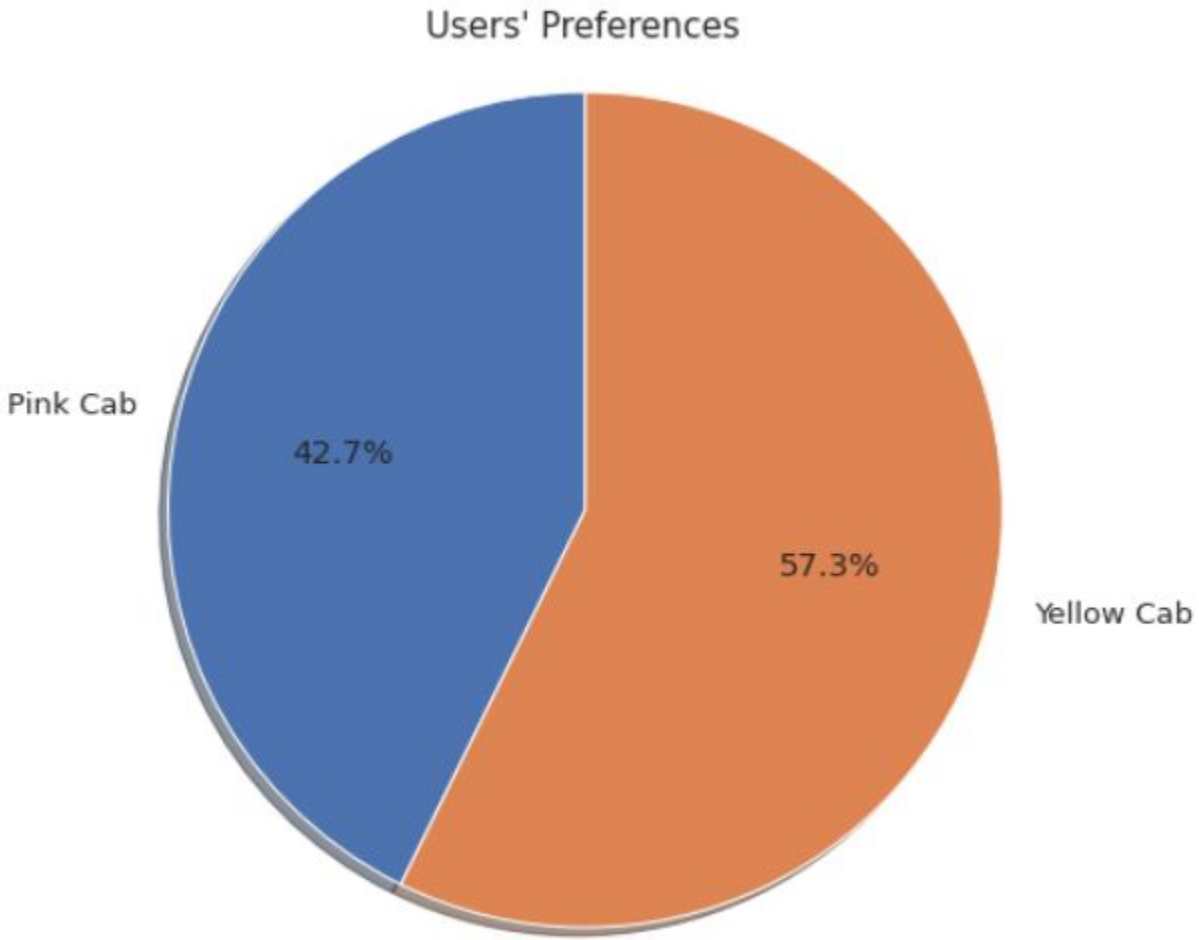
Data Glacier
Your Deep Learning Partner

Correlation Analysis



Data Glacier
Your Deep Learning Partner

Company Popularity



More users prefer to travel with 'Yellow Cab' more as compared to 'Pink Cab'

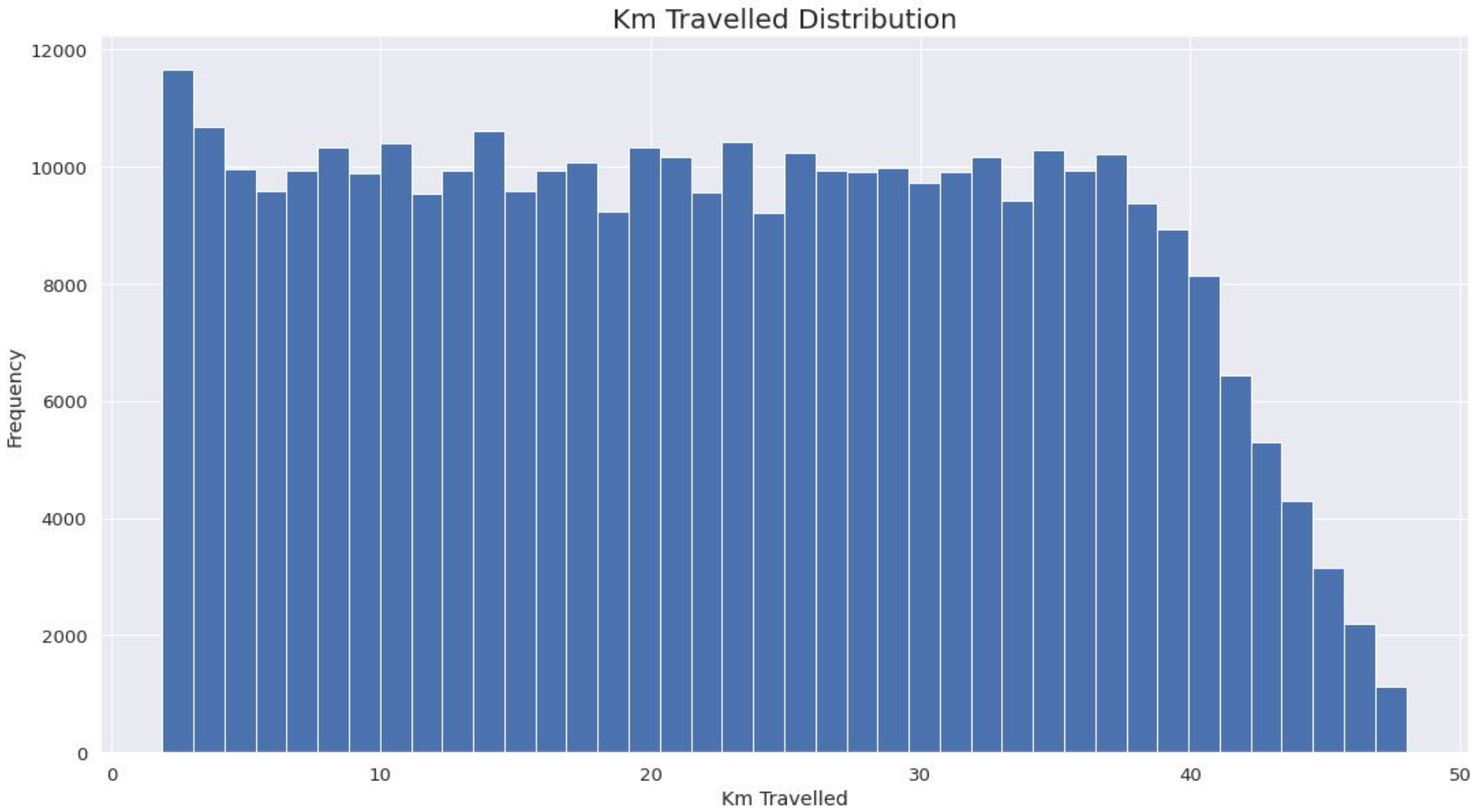
Data Glacier
Your Deep Learning Partner

Price Analysis

It can be observed that, price charged by 'Yellow Cab' is high compared to 'Pink Cab'.

Data Glacier
Your Deep Learning Partner

KM Travelled Distribution



Distance of travel for which cabs are frequently used varies from 2 km to 48 km approximately.

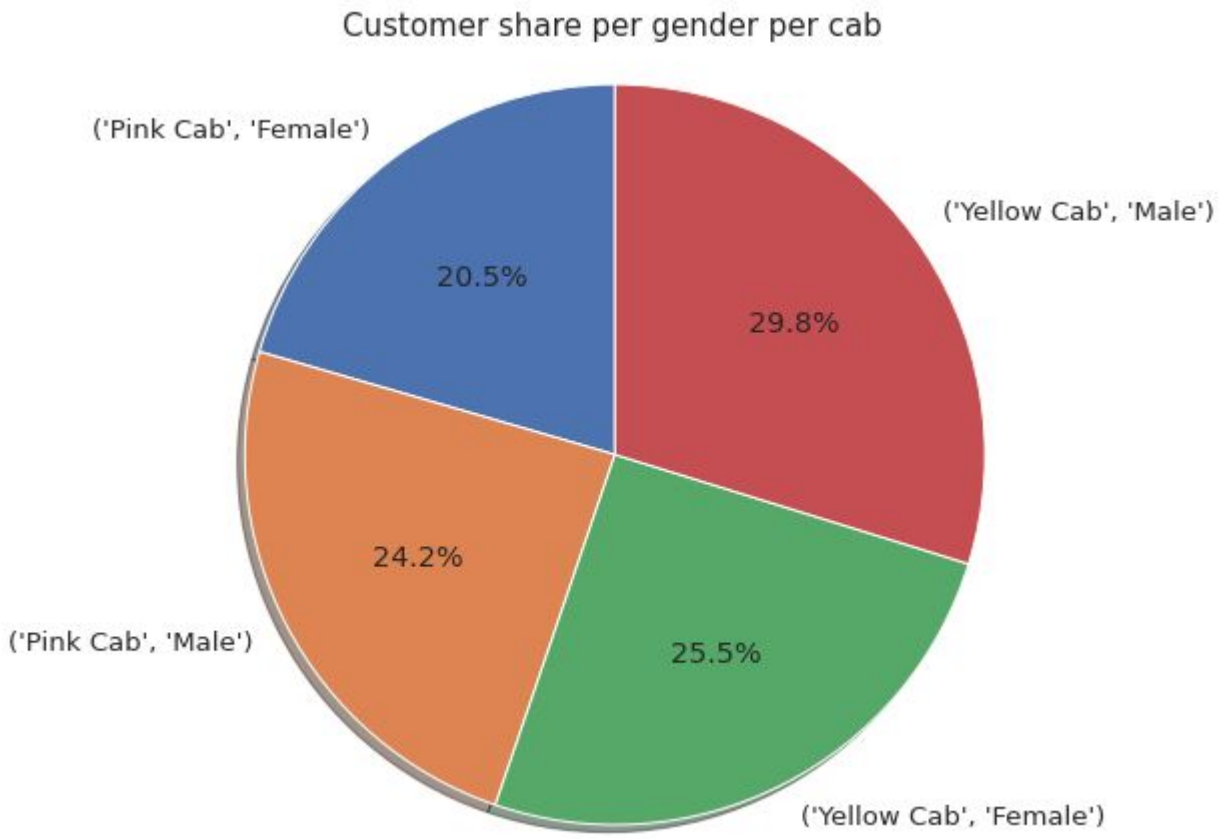
Data Glacier
Your Deep Learning Partner

Price Charged per km

It can be observed that 'Yellow Cab' charges more per km compared to 'Pink Cab'

Data Glacier
Your Deep Learning Partner

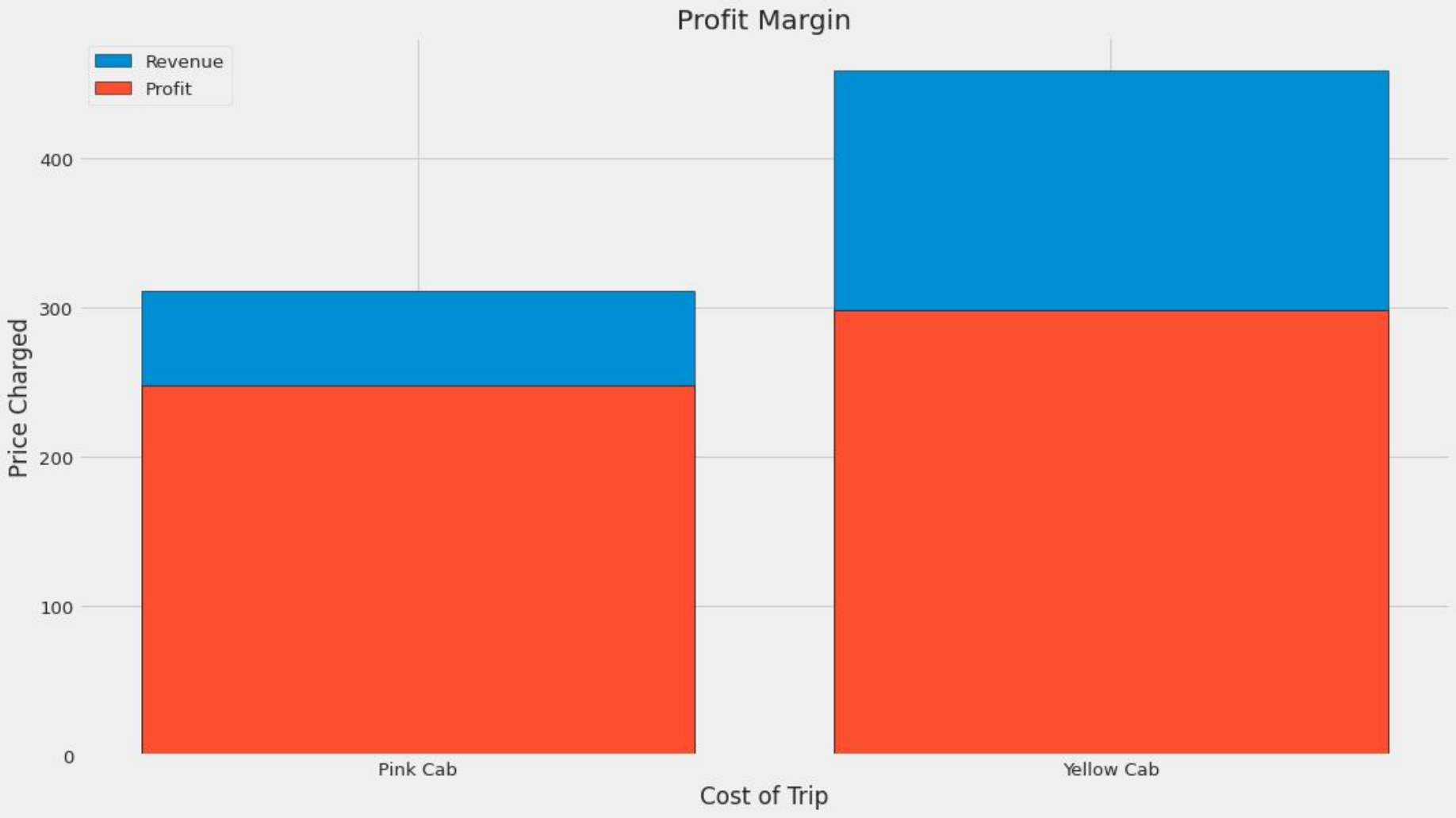
Customer share per gender per cab



It is observed that males prefer to travel to cab more than females for both the cab companies.

Data Glacier
Your Deep Learning Partner

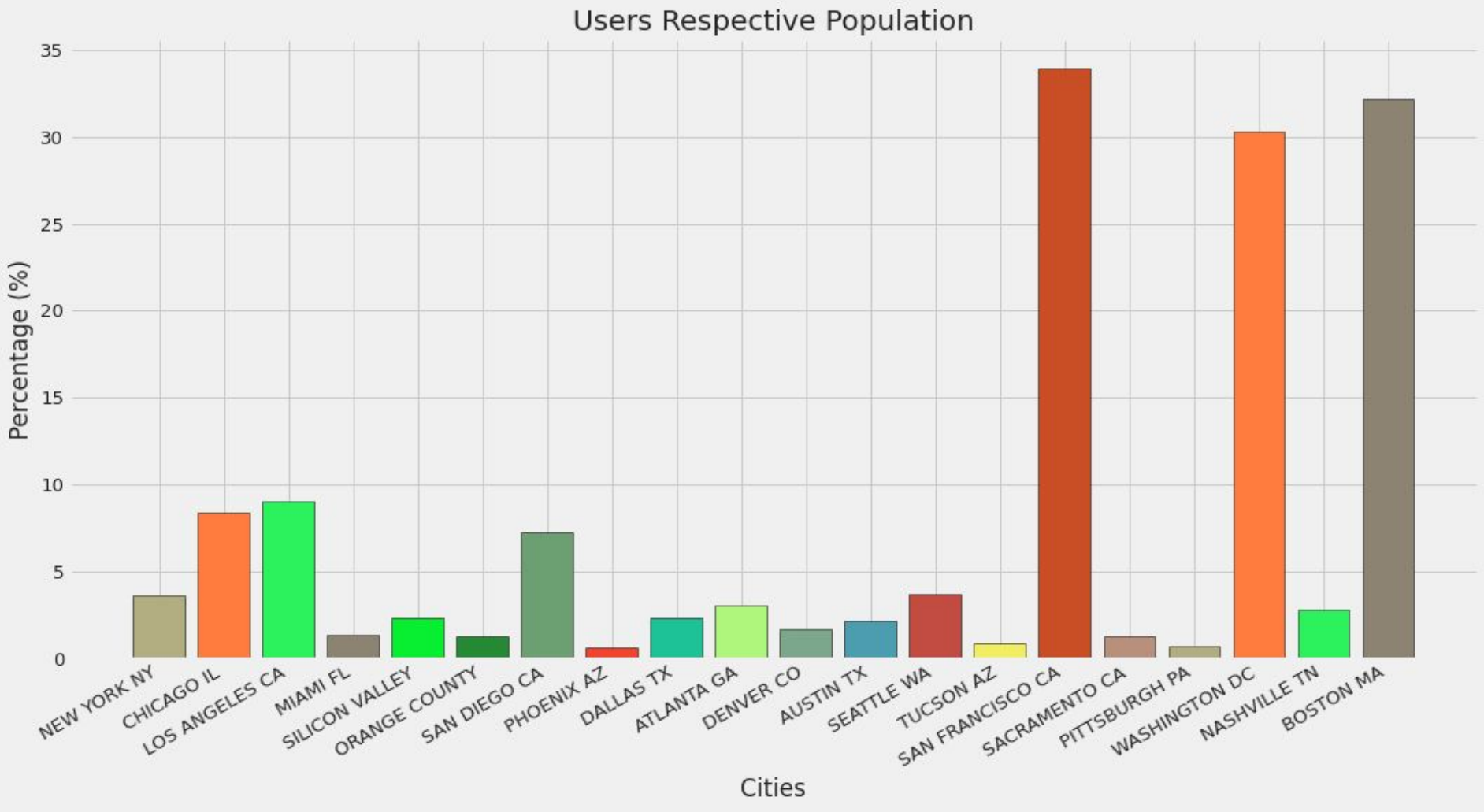
Profit Margin



It is observed that profit margin of 'Yellow Cab' is more than that of 'Pink Cab'.

Data Glacier
Your Deep Learning Partner

Users' Population Citywise



High percentage of population i.e. more than 30% in San Francisco, Washington DC and Boston MA use cabs.

Hypothesis Testing

Data Glacier
Your Deep Learning Partner

Hypothesis : Is there any difference in profit regarding Gender?

H0 : There is no difference regarding Gender in both cab companies.

H1 : There is difference regarding Gender in both cab companies.

We accept null hypothesis (H0) that there is no difference regarding gender for Pink Cab.

We accept alternative hypothesis (H1) that there is a difference regarding gender for Yellow Cab.

Data Glacier
Your Deep Learning Partner

Hypothesis : Is there any difference in Profit regarding Age?

H0 : There is no difference regarding Age in both cab companies.

H1 : There is difference regarding Age in both cab companies.

We accept null hypothesis (H0) that there is no difference regarding age for Pink Cab.

We accept alternative hypothesis (H1) that there is a difference regarding age for Yellow Cab.

Data Glacier
Your Deep Learning Partner

Hypothesis : Is there any difference in Profit regarding Payment mode ?

H0 : There is no difference regarding Payment Mode in both cab companies.

H1 : There is difference regarding Payment Mode in both cab companies.

We accept null hypothesis (H0) that there is no difference regarding payment mode for Pink Cab.

We accept null hypothesis (H0) that there is no difference regarding payment mode for Yellow Cab.

Conclusion

Data Glacier
Your Deep Learning Partner

Recommendation

Based on exploratory data analysis it is concluded that it is highly recommended to invest in Yellow Cab Company rather than Pink Cab Company as Yellow Cab Company has higher profit margin, more users, better age wise and income wise reach.

Thank You