



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Case Study

03-March-2023

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

Based on Problem statement, I did Exploratory Data Analysis(EDA). Researched field are two cab companies, Pink Cab and Yellow Cab. Using provided datasets, Check the correlation, compare two companies in terms of gender ratio, price, age, and so on. I also check whether there is seasonality of demand.

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.

Based on above project, I researched these questions below,

- Is there any relationships between income and other terms?
- Is there difference between two company in terms of profit?
- Can we identify the key fact of demand?
- Is There seasonality ?
- Which city is most using cab?

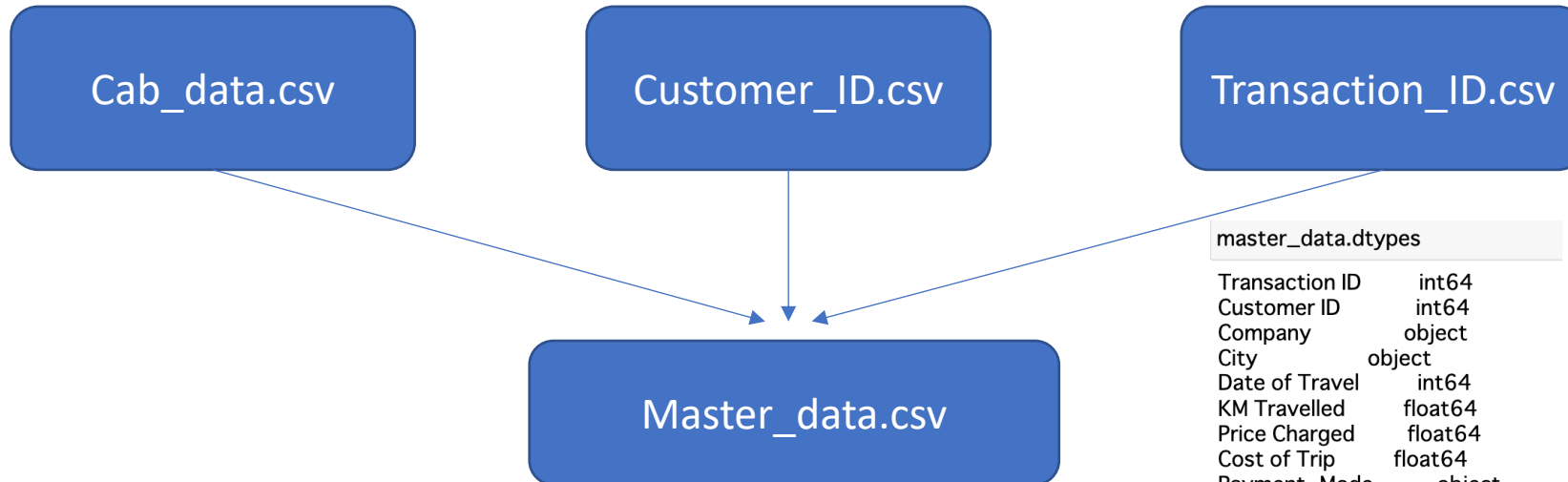
Dataset

Original datasets

- Cab_data.csv
- Customer_ID.csv
- Transaction_ID.csv
- City.csv

Processed Datasets

- Master_data.csv
- City.csv



To conduct analysis, merge three datasets with IDs and named it master data.

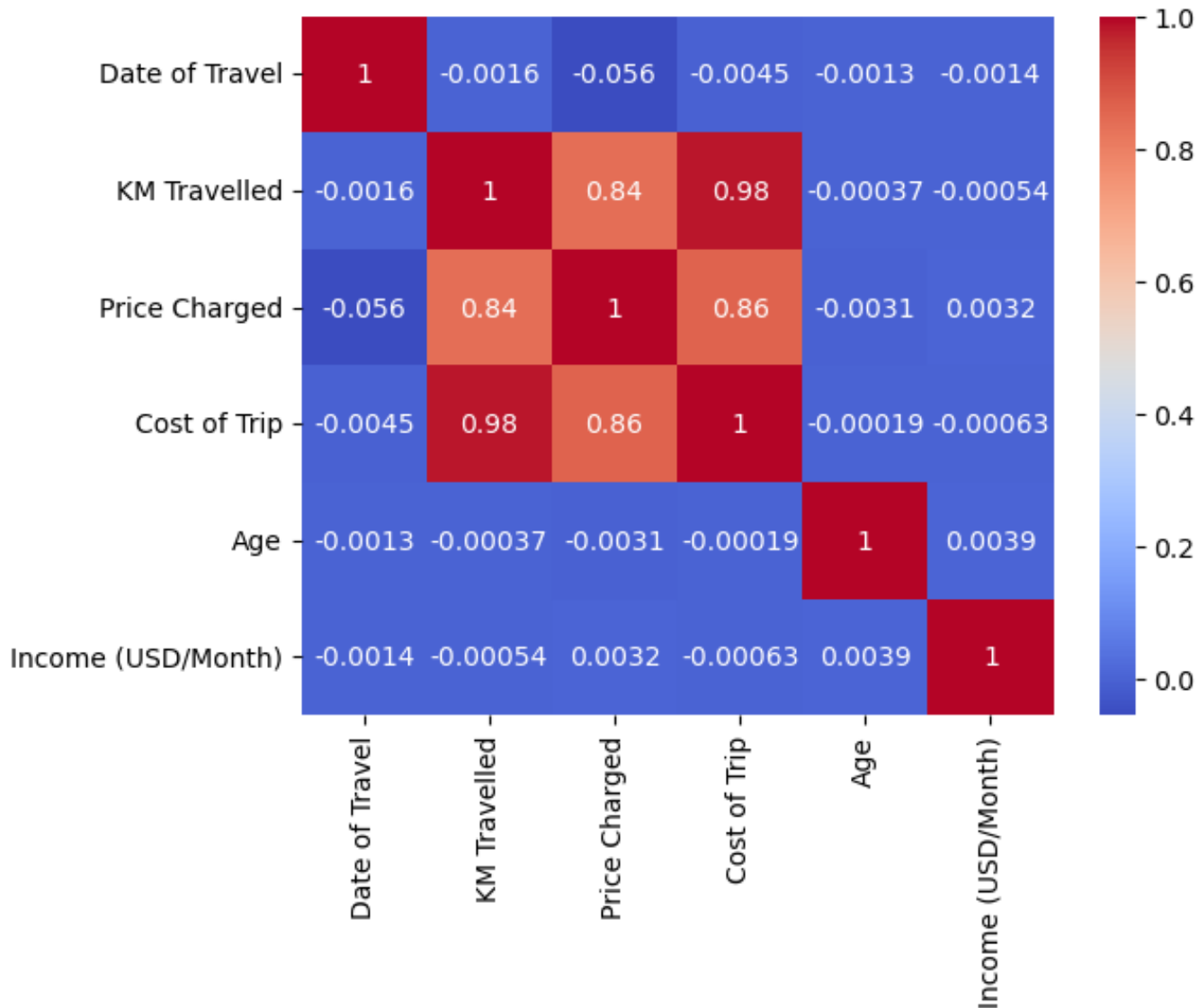
master_data.dtypes

Transaction ID	int64
Customer ID	int64
Company	object
City	object
Date of Travel	int64
KM Travelled	float64
Price Charged	float64
Cost of Trip	float64
Payment_Mode	object
Gender	object
Age	int64
Income (USD/Month)	int64
dtype:	object

Approach

- Check correlation with heatmap
- Comparison two company with users, gender, average cost and income
- Check users by city
- Time series analysis

EDA



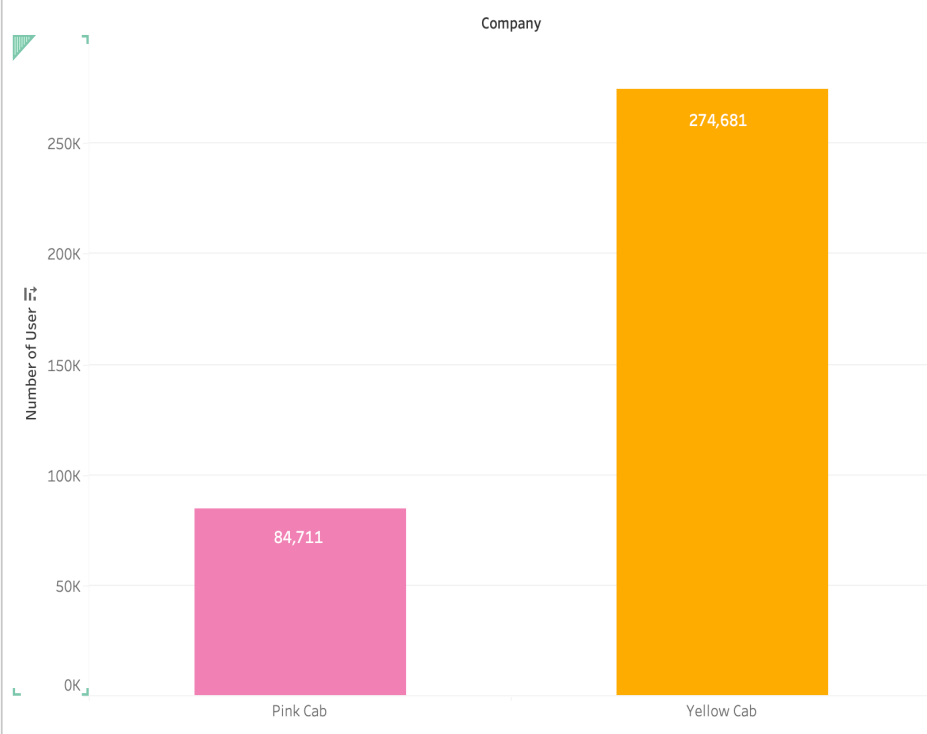
Check correlation with heatmap

- There is no strong correlation between income and other features.
- KM Travelled, Cost of Trip and Price charged has strong correlation.
- Price charged and Cost of Trip has strong correlation.

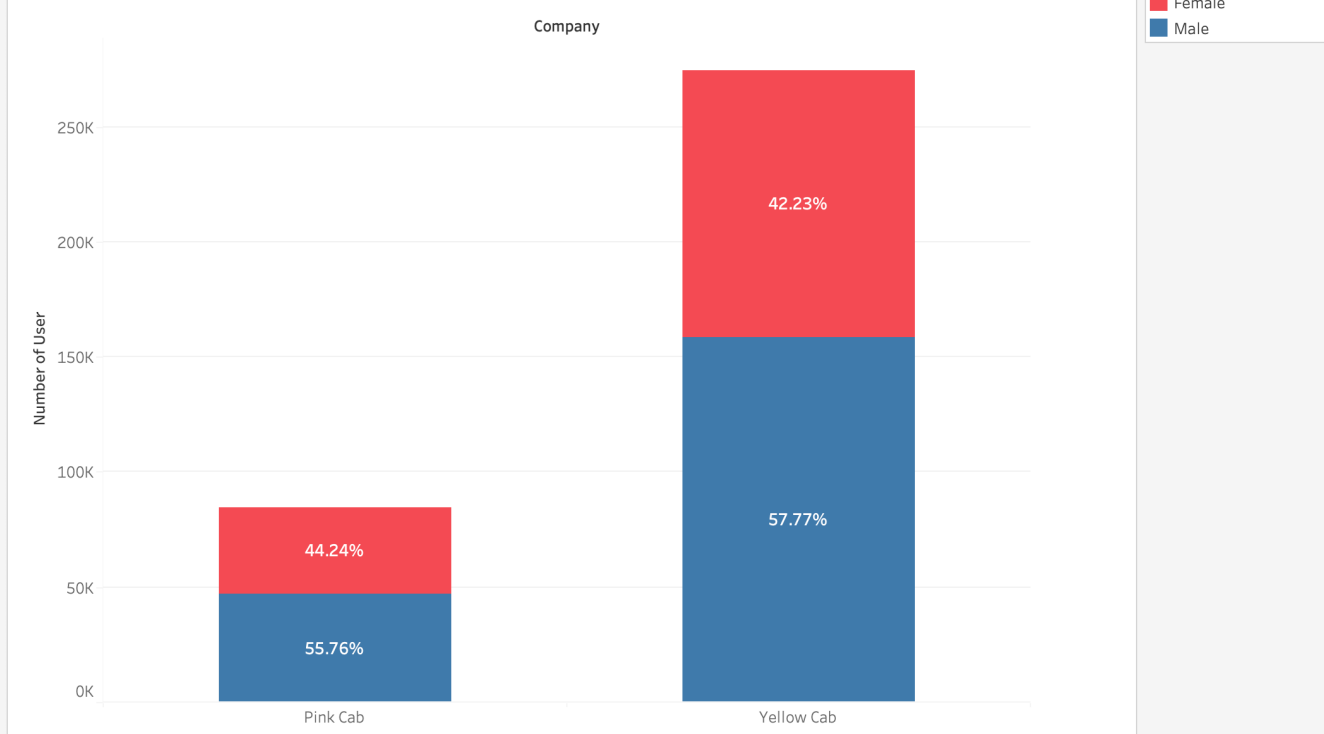
EDA

Comparison two company

The number of Users by company



The Number of Users by company and gender

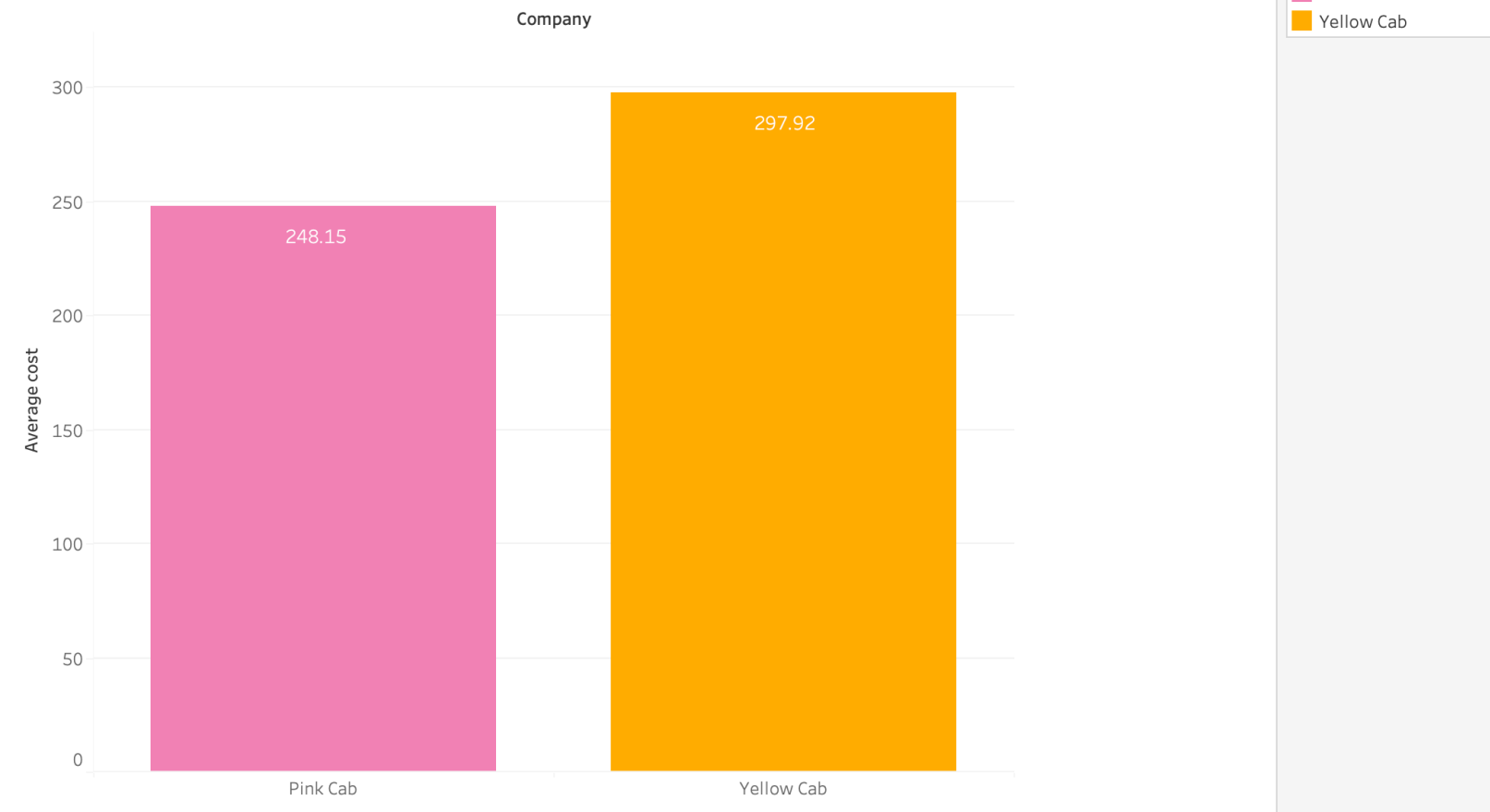


- Yellow Cab has more users than Pink Cab
- There is no significant difference in ratio of gender
- Yellow Cab is about three times as popular as Pink Cab

EDA

Comparison two company (cont)

Average Cost by company

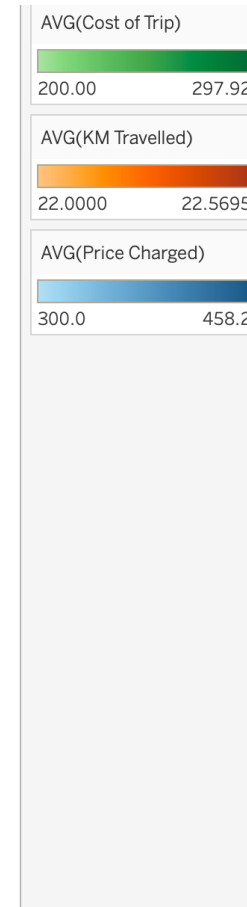
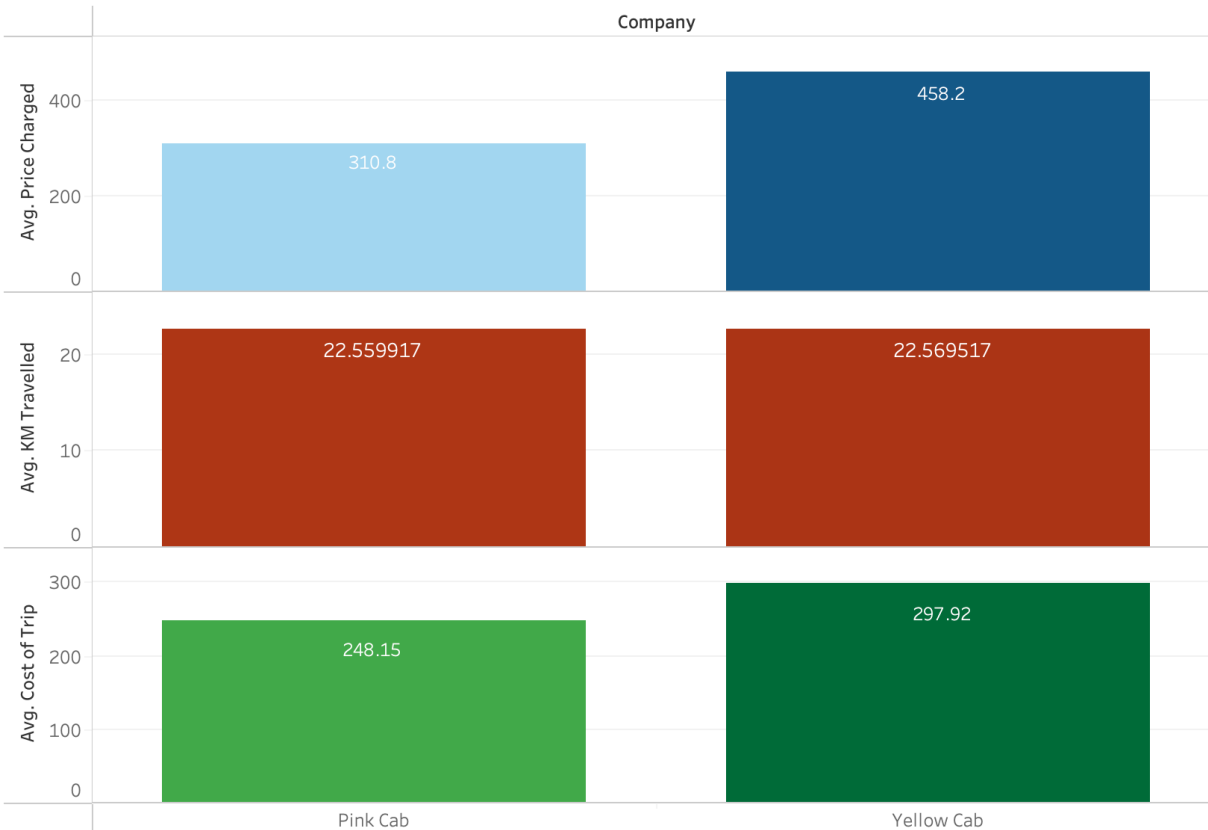


Average cost of Pink Cab is lower than the one of Yellow Cab.

EDA

Comparison two company (cont)

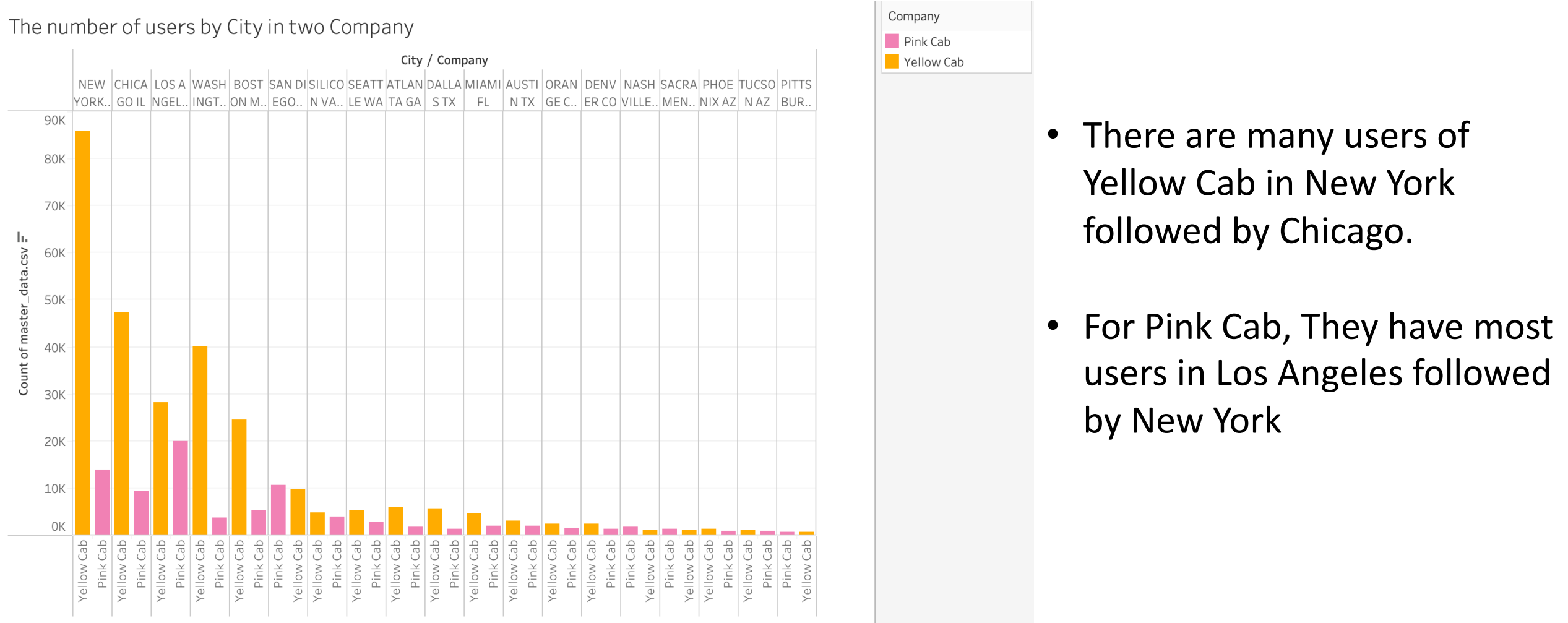
Average Price, KM travelle, and Cost by company



Although average KM travelled is almost same, average price charged and cost of trip in Pink Cab is lower than Yellow Cab.

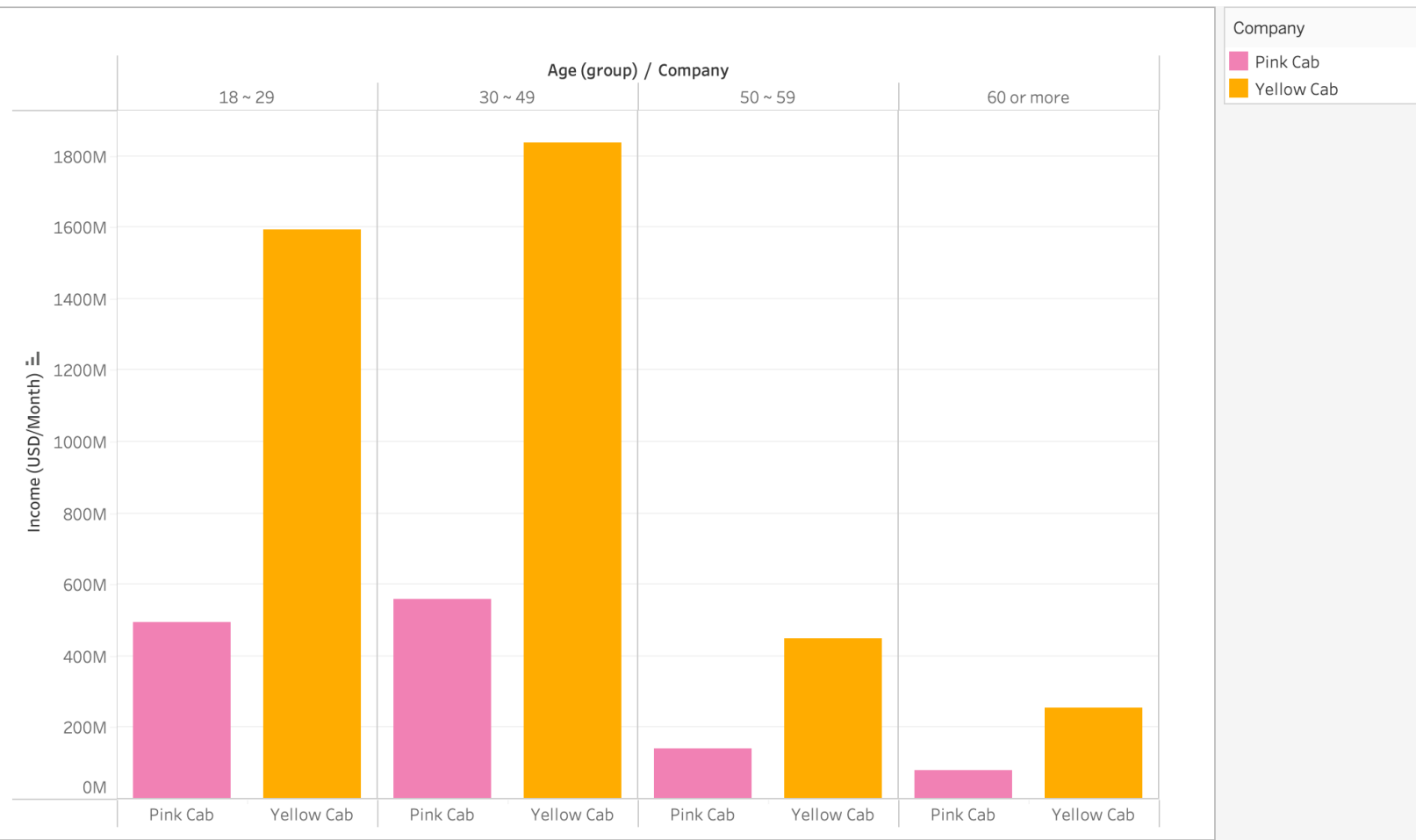
EDA

Comparison two company (cont)



EDA

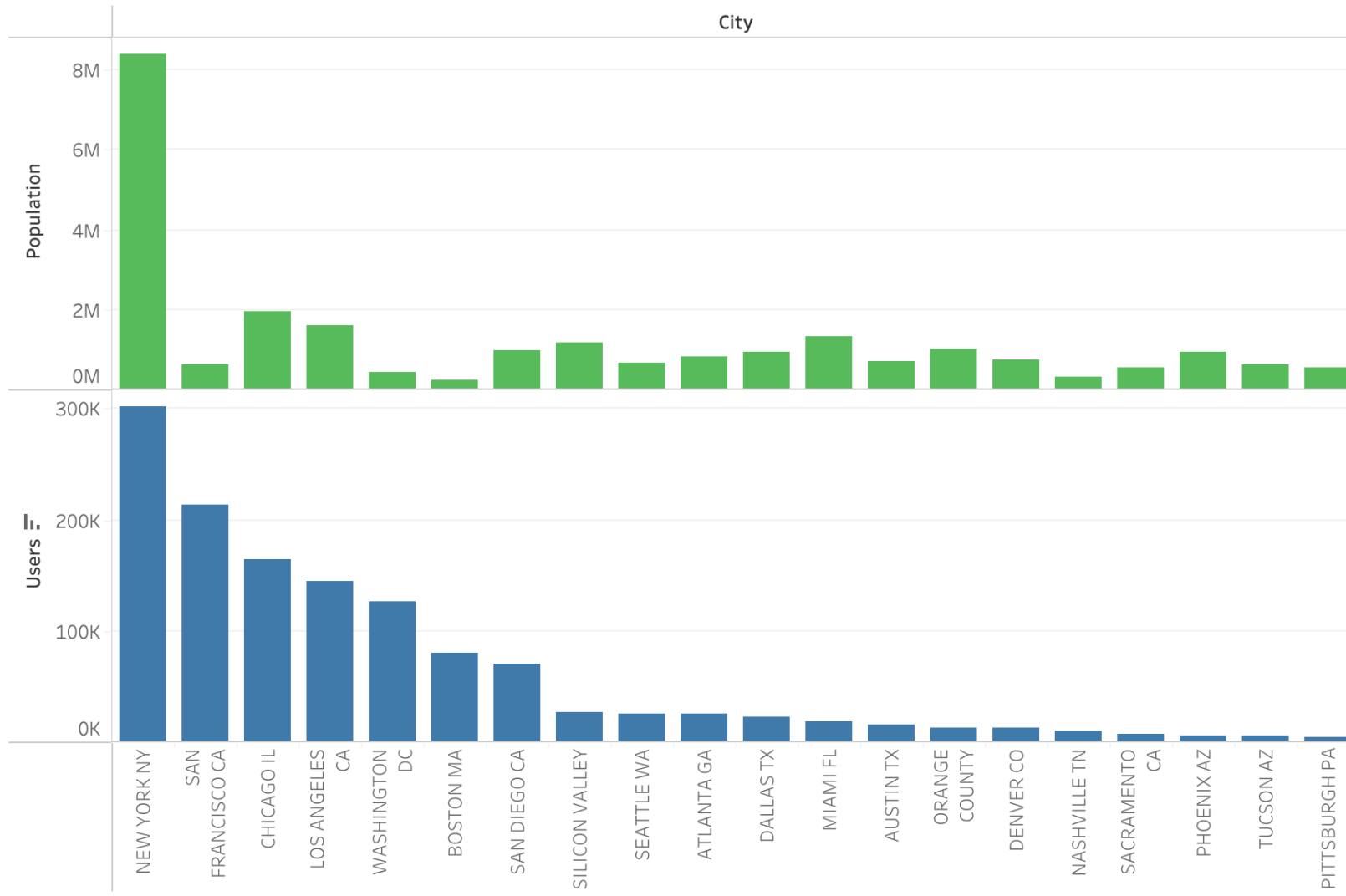
Comparison two company (cont)



Age 38 ~ 49 is primary users for both companies.

EDA

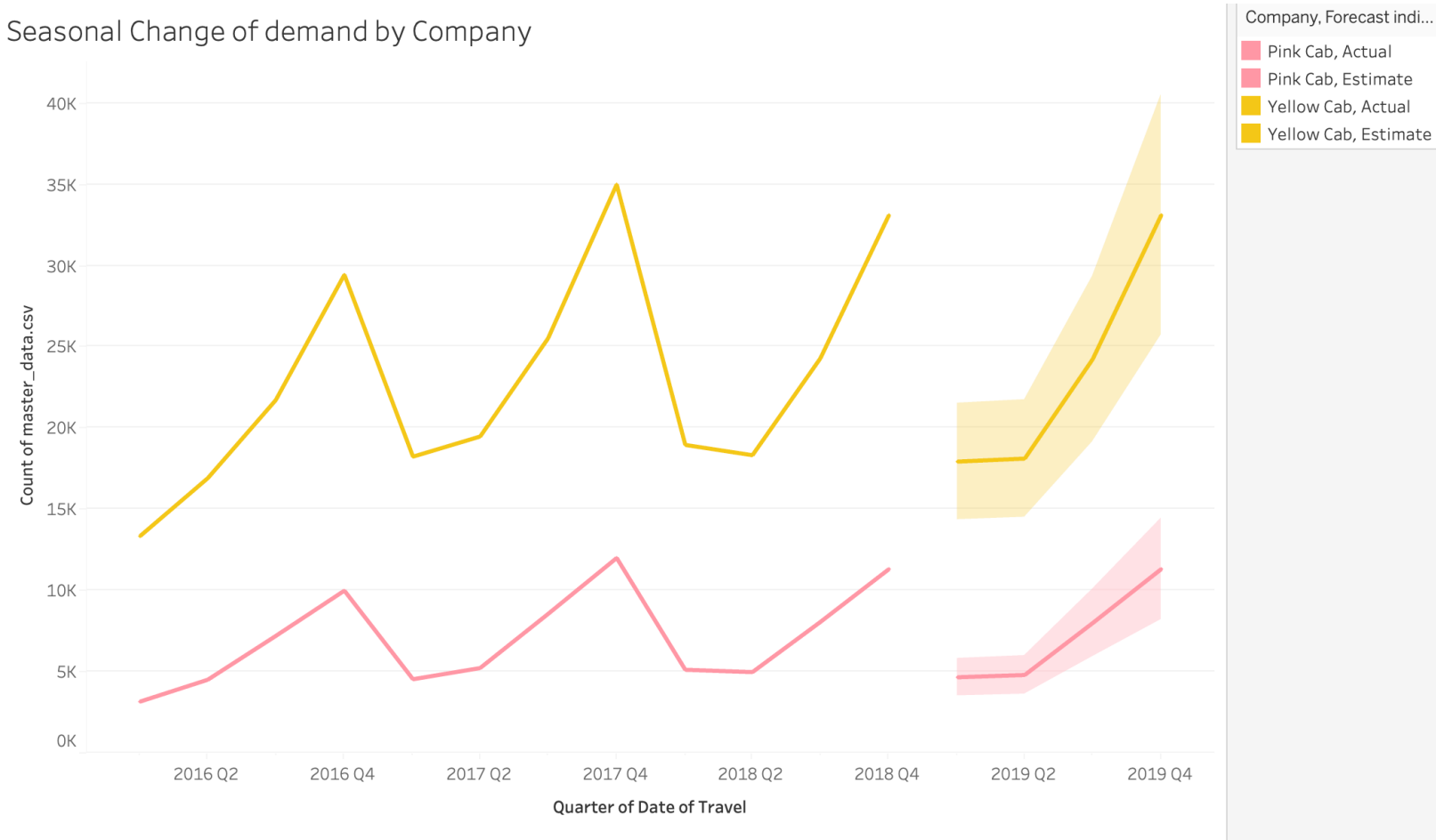
The number of Users by city



- There are many users in New York.
- Following New York, San Francisco has many users despite less population.

EDA

Time series analysis for demand (Quarter)



- There is a seasonality trend in demand of cab.
- Quarter 4 in every year has most demand.

EDA Summary

- KM Travelled, Cost of Trip and Price charged has strong correlation. Price charged and Cost of Trip also has strong correlation.
- Yellow Cab has about three times as many users as Pink Cab.
- Average cost of trip for Pink Cab is lower than the one of Yellow Cab.
- Although average distance of travelled in both companies is almost same, Pink Cab is cheaper than Yellow Cab and cost of trip slightly less. For this reason, Pink cab has nice cost performance.
- Yellow Cab has most customers in New York while Pink Cab has in Los Angeles.
- People 38 ~ 49 years old use cab much.
- In San Francisco, there are many cab users despite less population.
- In winter, demand of cab got high due to seasonality.

Reccomendations

- I found that Pink Cab has better cost performance than Yellow Cab so Pink Cab is promising company. However, Yellow Cab is three times as popular as Pink Cab, and I could not find the key fact to leverage the demand of Pink Cab
- There is high demand in San Francisco despite of less population, so cab company should more focus on San Francisco.
- In quarter 4 that is winter, the demand of cab is high so we can get better profit by devising the price setting in the season.
- People who are 50 ~ 59 years old are still working but the demand is low. I recommend to conduct campaign for the people to facilitate using
- In conclusion, I recommend to invest for Yellow Cab due to the popularity.

Appendix

Code for this analysis:

<https://github.com/KHUC1998/G2M-market/blob/main/Untitled.ipynb>

programming language:

Python

Used software:

Jupyter notebook Tableau

Thank You