Exploratory Data Analysis
G2M FOR CAB INVESTMENT FIRM CASE STUDY

15-March-2021

Name : Buse Güngör
Location: Turkey
Team: Data and Analytics

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Data Glacier
Your Deep Learning Partner

Executive Summary –G2M case study

- 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.

- **Provide actionable insights to help XYZ firm in identifying the right company for making investment.**

The analysis has been divided into four parts:

- Data Understanding

- Examining the background of the data and its effect on results

- Finding the most profitable Cab company

- Recommendations for investment

Data Exploration

- 17 Features
- Timeframe of the data: 2016-01-31 to 2018-12-31
- 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.

- 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)

Cab_Data.csv   Customer_ID.csv   Transaction_ID.csv   City.csv
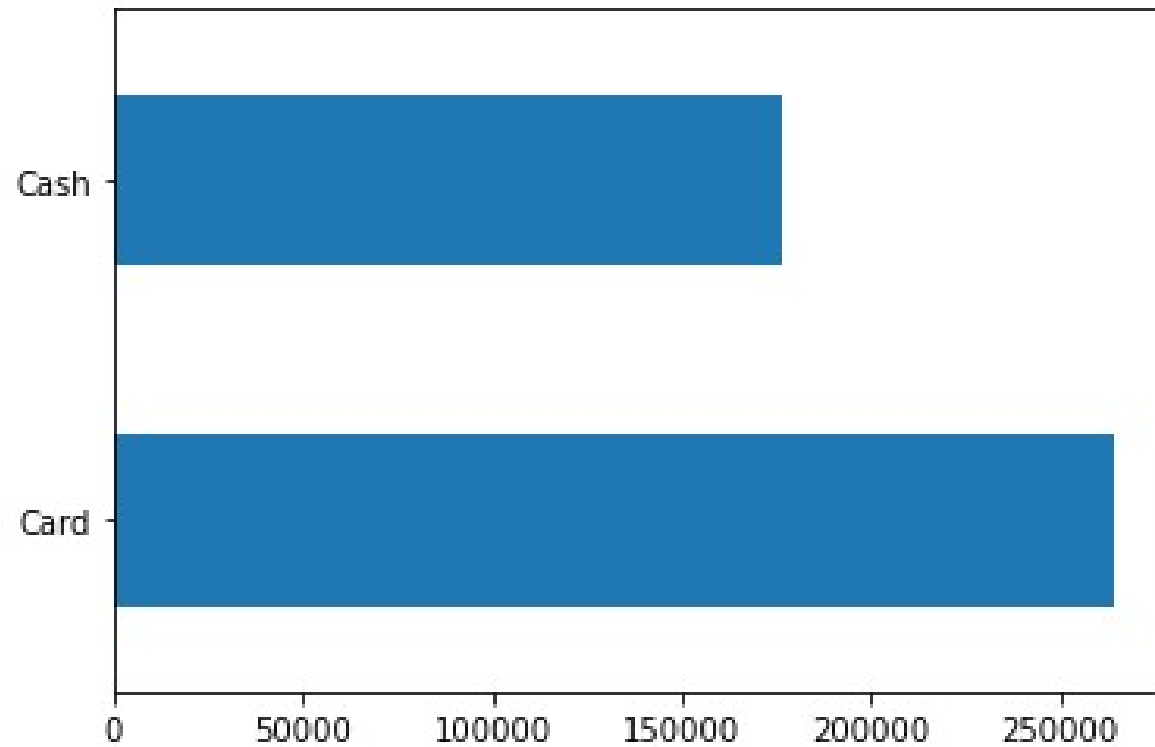
DF = Merged Data
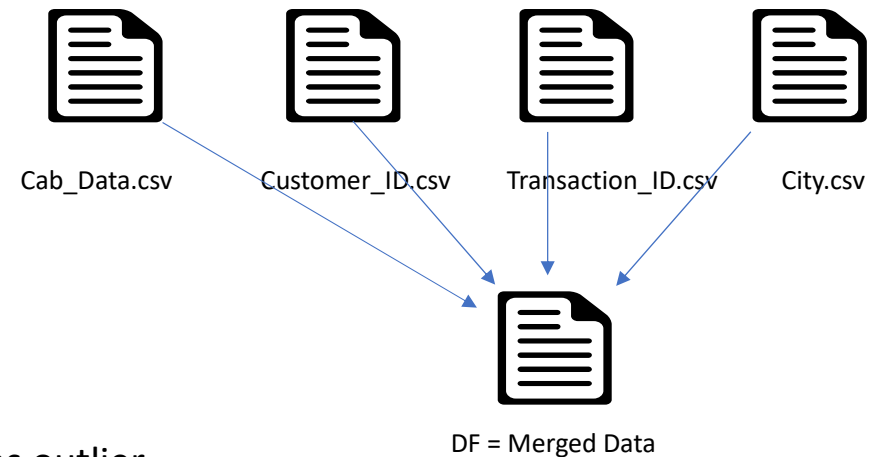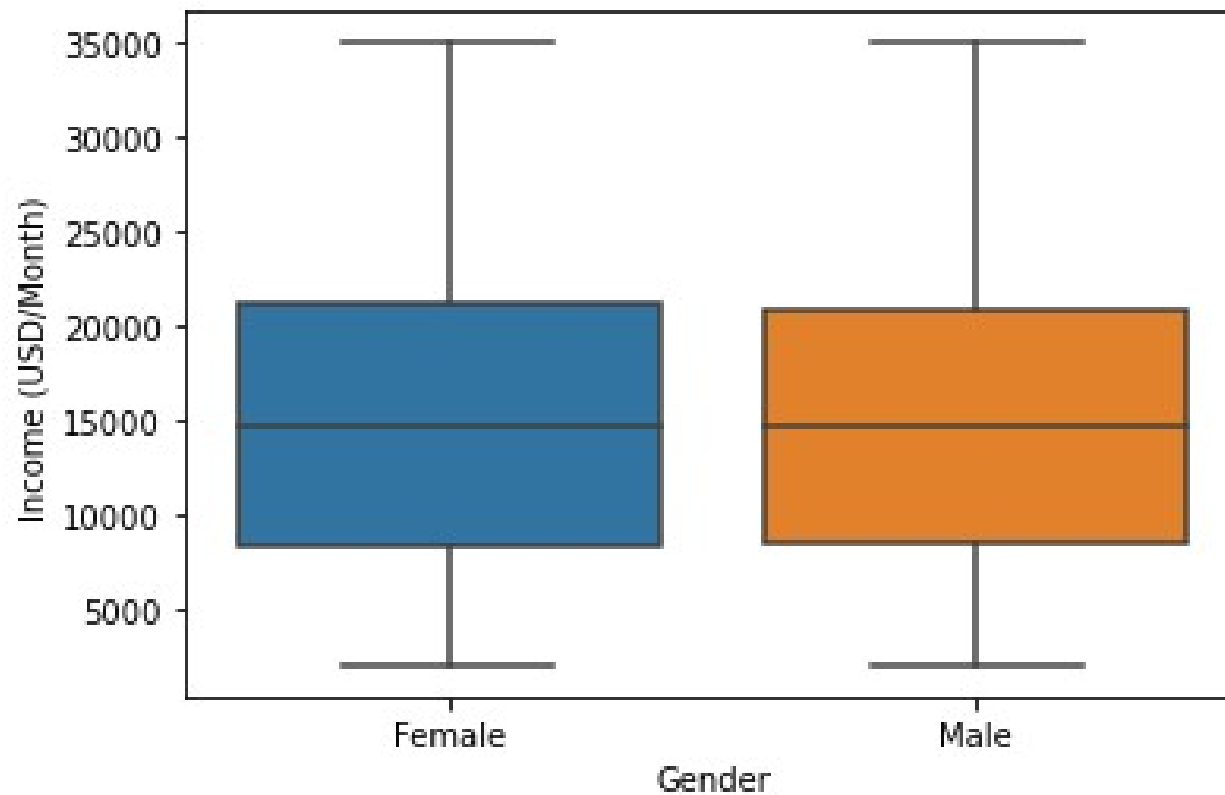
Frequency of Using Taxi Companies
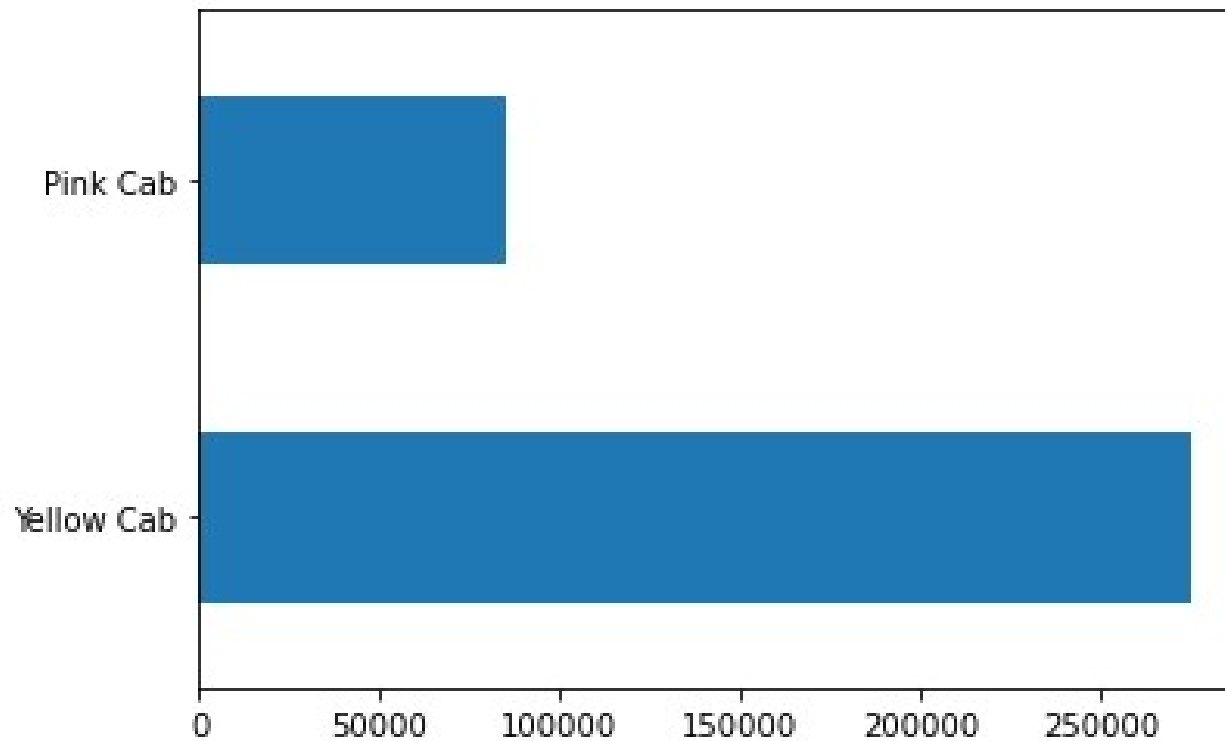
Frequency of Using Taxis in Cities
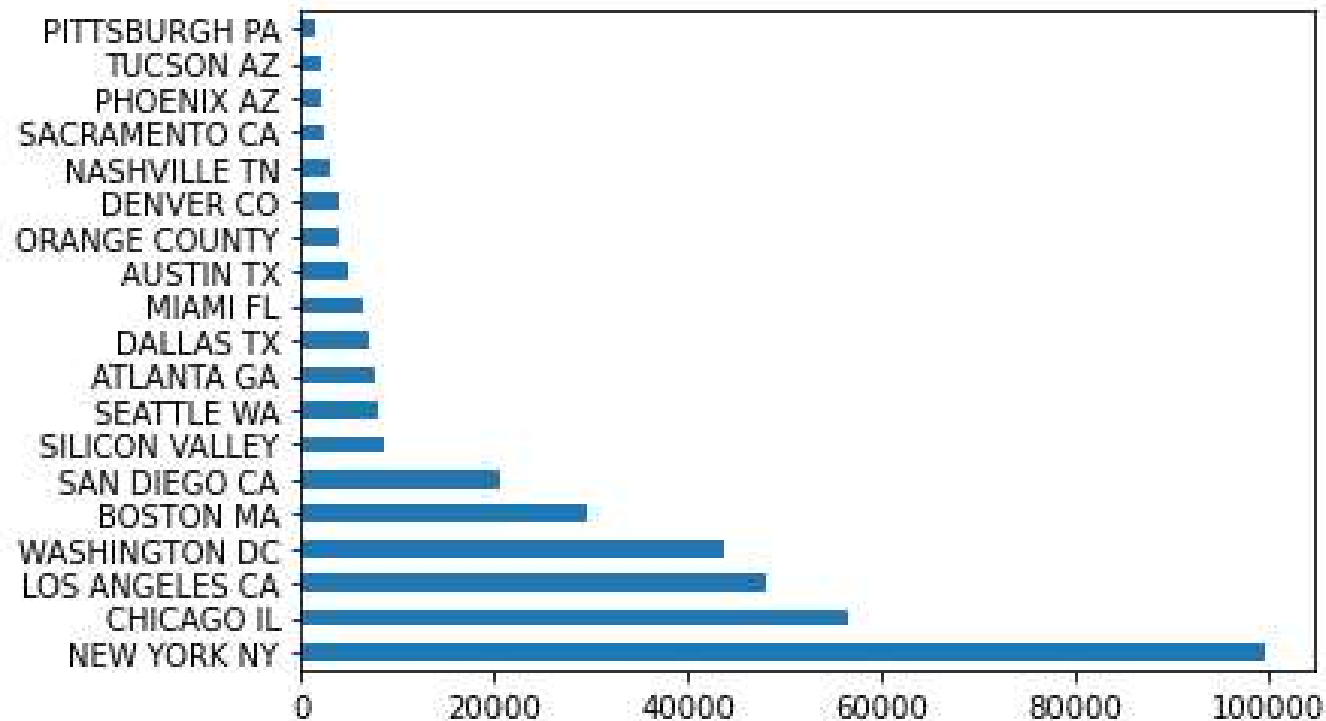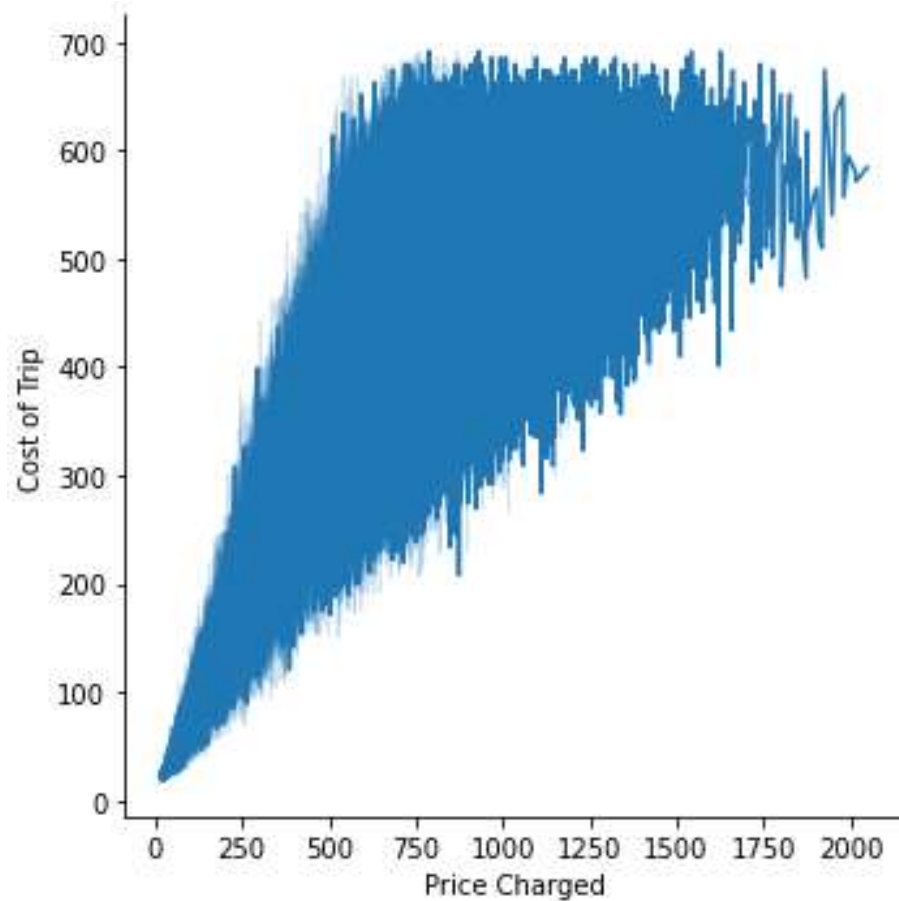
Payment Method in Taxis

Income Distribution by Gender

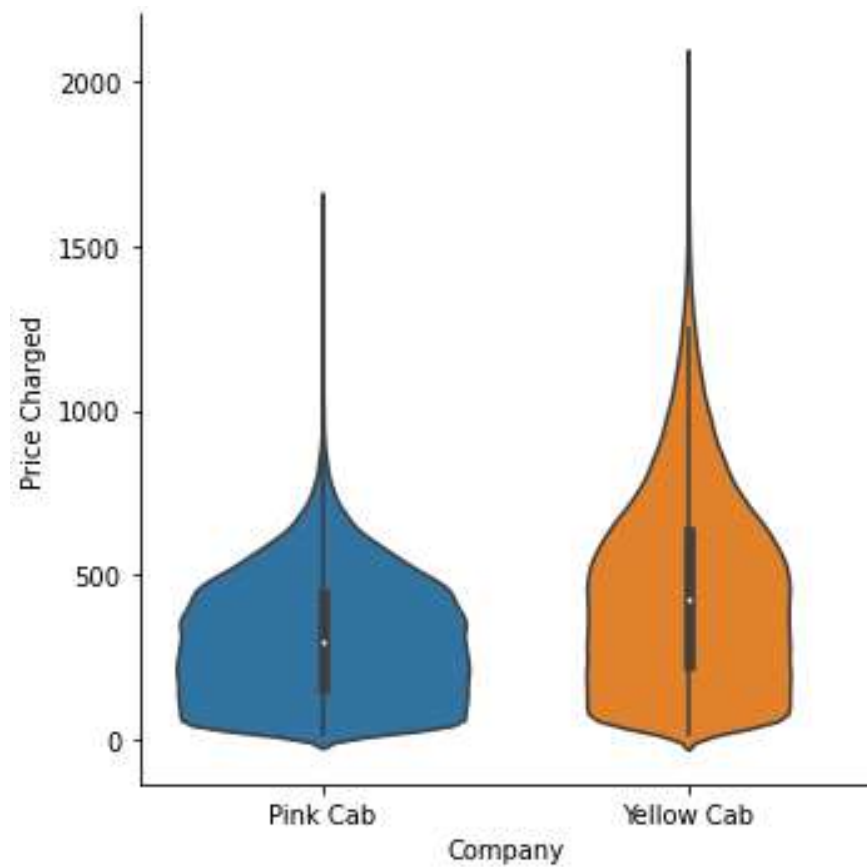Both genders have almost equal income distribution.

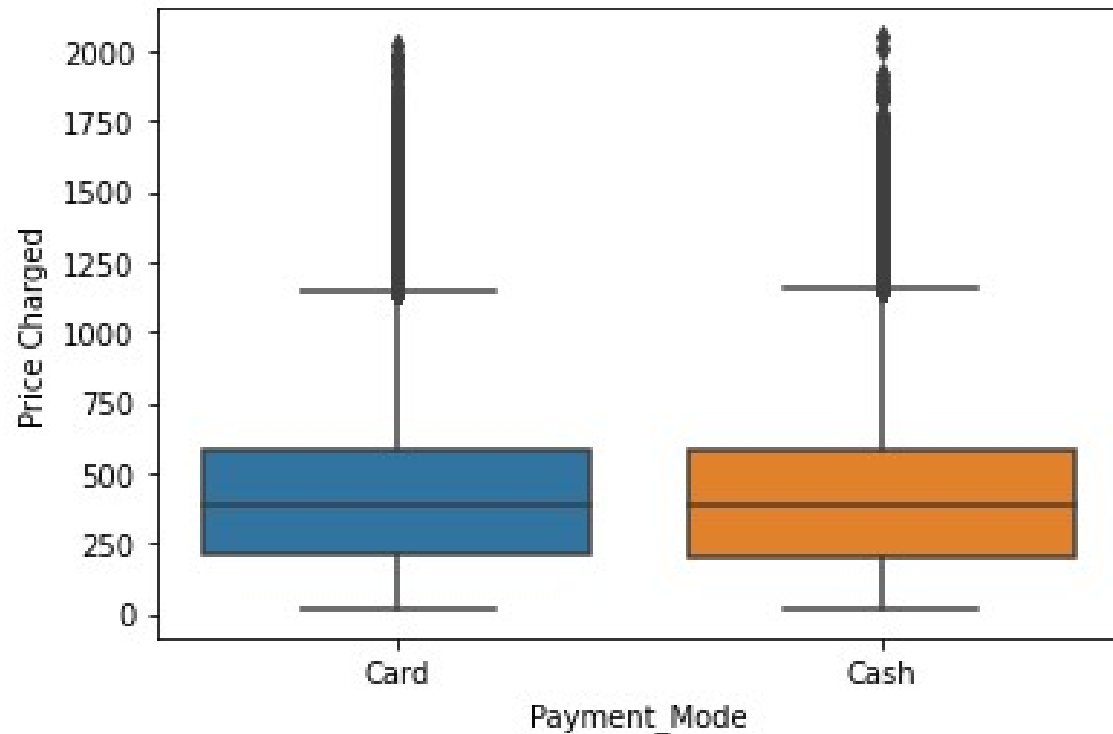# Relationship Between Cost of Trip and Price Charged

Cost of Trip increased, Price Changed naturally increased, but while Cost of Trip remained the same at $ 600 and beyond, Price Changed increased.

Distribution of The Price Charged by Taxi Companies

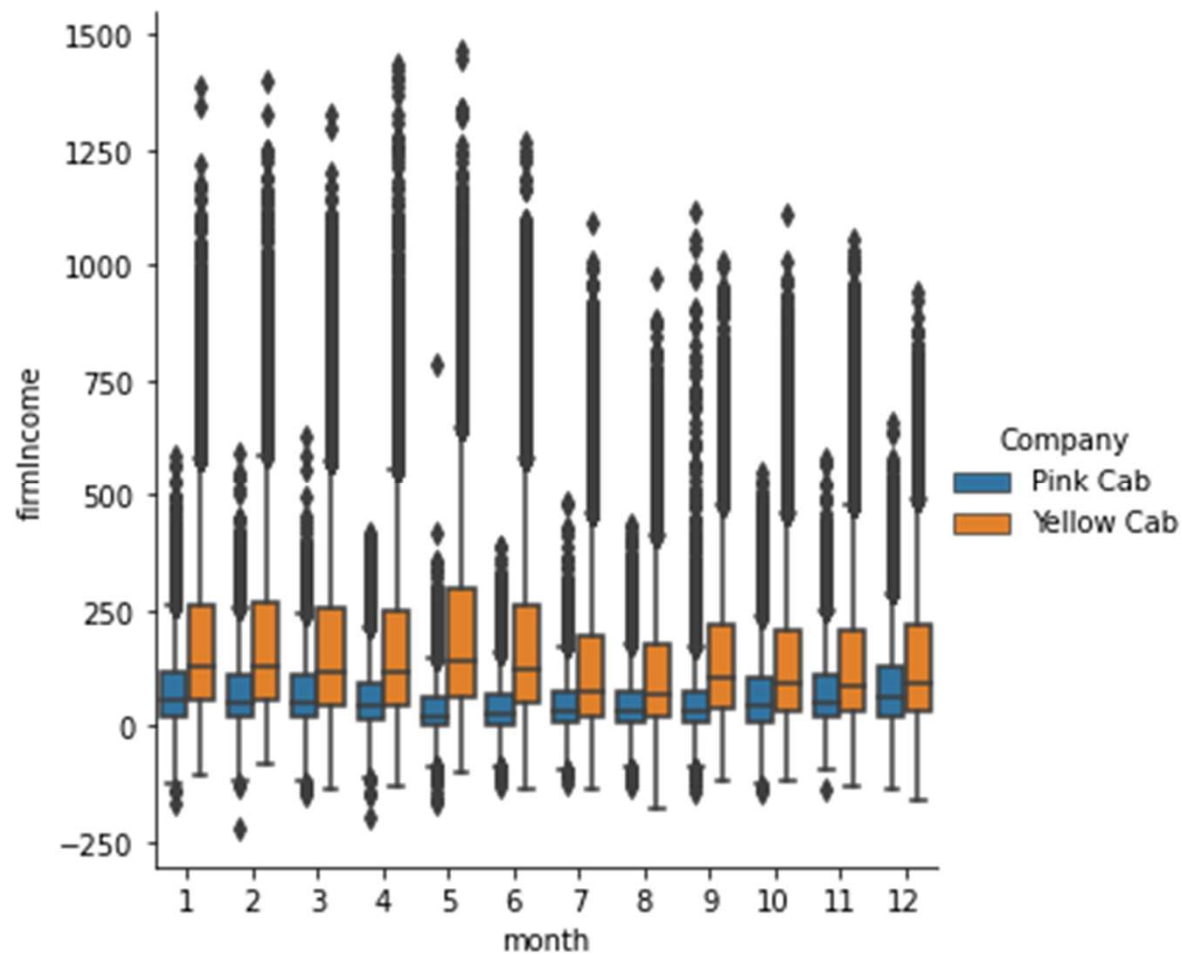# Distribution of The Price Charged According to Payment Mode

We can say that both payment methods are the same and both are close to normal distribution. And we see that there are outliers in both.

# Companies are Preferred According to The Income of The Users in The Cities

What is intended to be explained in this graph is which companies are preferred according to the income level of the users in the cities. Accordingly, the cities with high income (between 25000 and 30000 per month) salary are: Boston, Chicago, Los Angeles, New York City, Seattle, Washington DC, San Diego and Silicon Valley. The company preferences here are clearly seen as "Yellow Cab".

# Distribution of Firm' Income by Months

Although the reason for the outliers comes from the price charged, roughly it is the yellow taxi company that earns the most revenue per month.

Distribution of Firm' Income by Cities
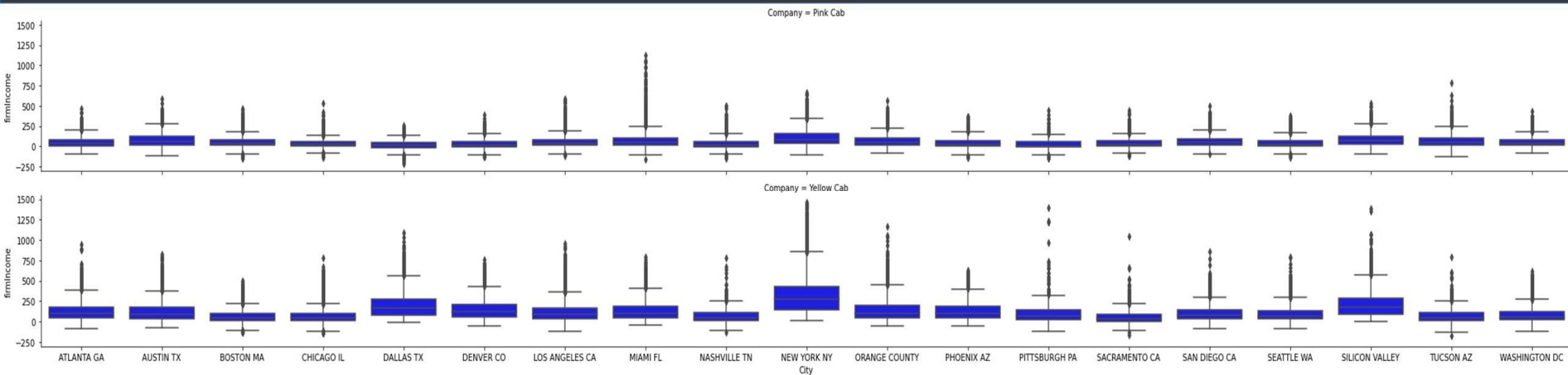
Showed how much the income of the firms was in which cities. The yellow taxi firm generated the most revenue in New York, Dallas and Silicon Valley, respectively. In the pink taxi company, the highest revenue was obtained in New York and Austin. But the yellow taxi company is more profitable in almost every city than the pink taxi company.

Companies Average Income by Month

Companies Average Income by Month
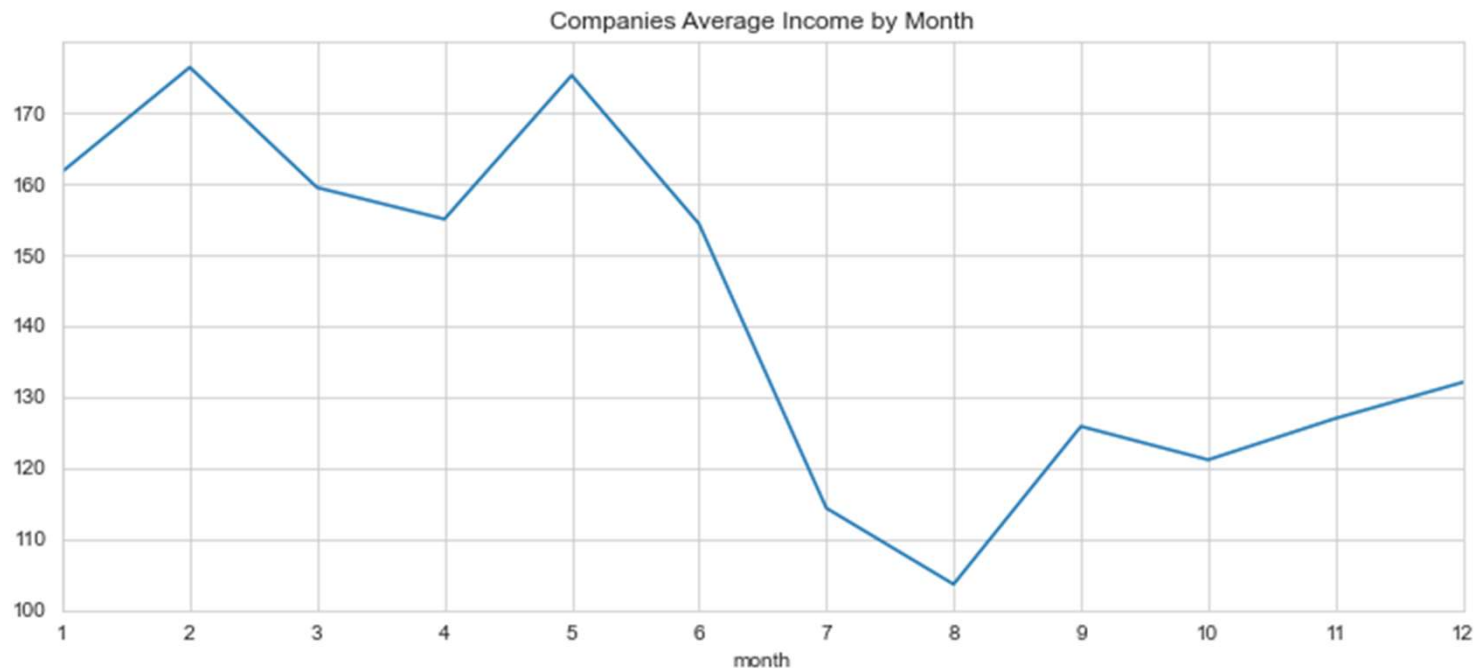
We see that the taxis earn the most income in the winter and spring seasons. The reason may be the cool and rainy weather in these months. We can attribute the decrease in income in July and August to the holiday season.

Yearly Average Companies Income in Each City

Yearly Average Companies Income in Each City
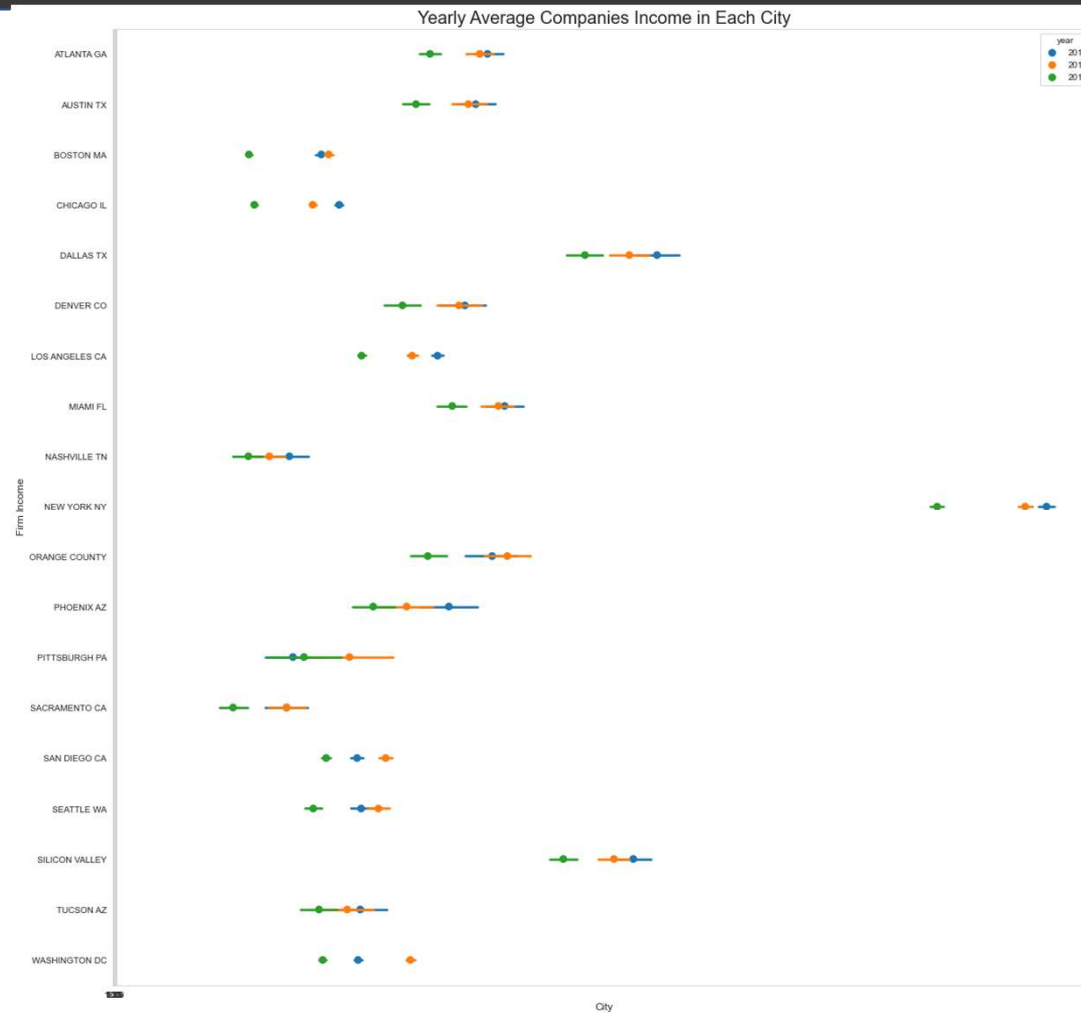
Mean of Company's Income (monthly-weekly-daily)
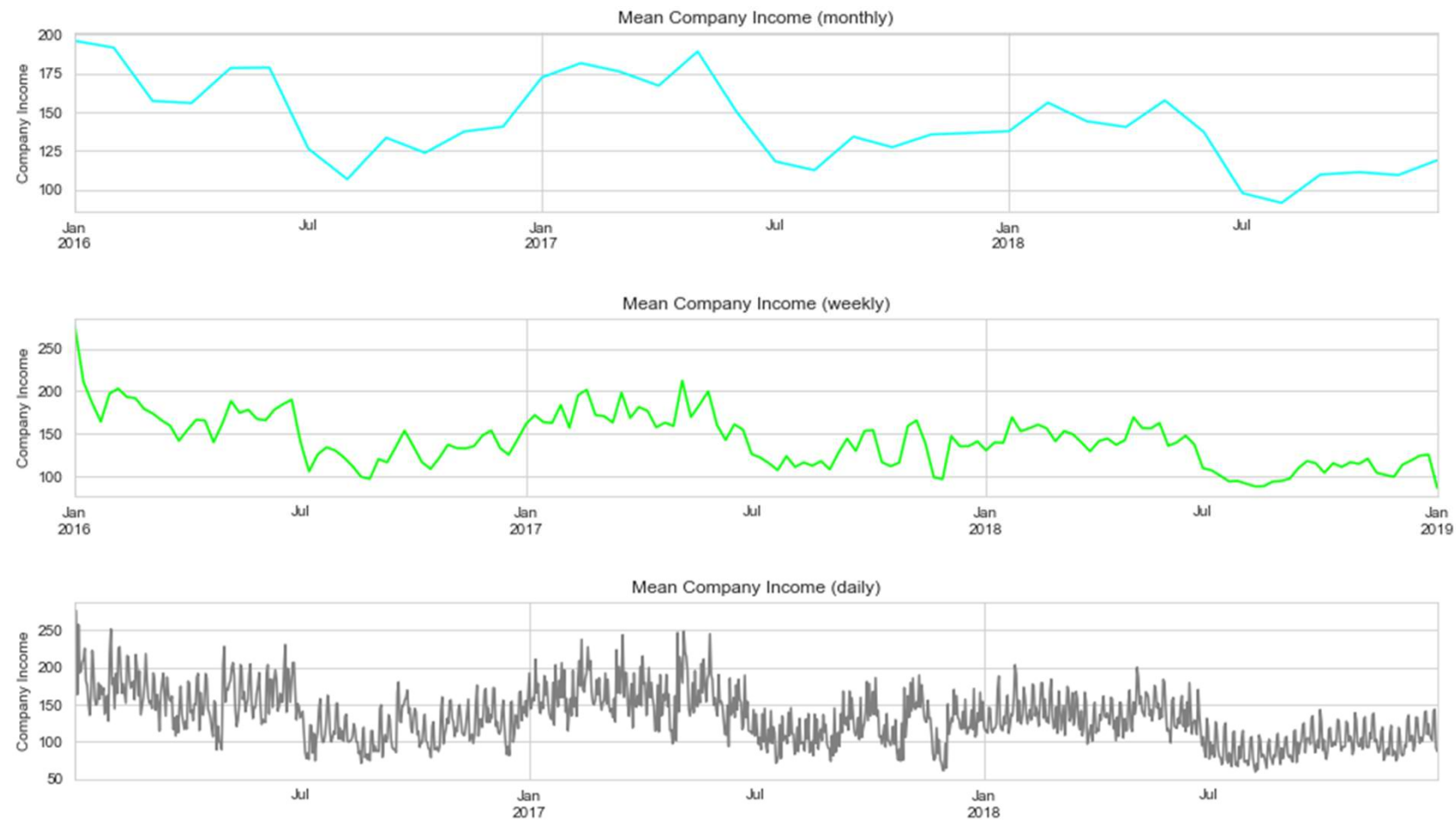
Recommendations

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

- **Customer Reach :** Yellow cab has higher customer reach in 25 cities while Pink cab has higher customer reach in 4 cities. We have also observed that Yellow cab is doing good in covering other cab users as compared to Pink cab.

- **Customer Retention:** We have analyzed this in two segments : at least 5 drive and at least 10 drive with the same cab company. And we found that Yellow cab is doing far better than Pink cab in both these segments.

- **Age wise Reach :** As can be seen in the correlation, age groups have no effect on firms' incomes.

- **Average Profit per KM:** Yellow cab's average profit per KM is almost three times the average profit per KM of the Pink cab.

- **Income wise Reach :** Both the cabs are very popular in high and medium income class but here also Yellow cab is performing better than Pink cab in offering their services to all the three income class group (low, medium and high)

- **Ride count and Profit Forecasting :** Although the revenues of the two companies remained the same in 2016 and 2017 during the years examined, the revenues of the companies decreased in 2018. but the revenue of the yellow taxi firm decreased more than that of the pink taxi firm.

On the basis of above point , we will recommend Yellow cab for investment.

Thank You

Data Glacier
Your Deep Learning Partner