Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Insight for Cab Investment Firm

15/09/2023

Agenda

Data Glacier
Your Deep Learning Partner

The Client

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.

XYZ need actionable insights to help them identify the right company to make their investment.

Agenda

Data Glacier
Your Deep Learning Partner

The Approach

Data across 3 years relating to the cab companies was downloaded, and the data sets cleaned and relationship between data sets identified.

The customer base and profits for each company will be analysed, and other factors affecting profits determined so XYZ can be given information on which company to invest in, and the most effective way to make their investment.

The Data

Cab_data.csv

- This data set contains information on the transaction ID, date of travel, the company the cab was under, the city the journey took place in, the distance travelled, the cost of the journey and the price charged/.

City.csv

- This data set contains information on the population of each city and the number of cab users

Customer_ID.csv

- Contains information on the customer ID, gender, age, and income of each customer

Transaction_ID.csv

- Contains information on the transaction ID, customer ID and payment method of each transaction

Rain data

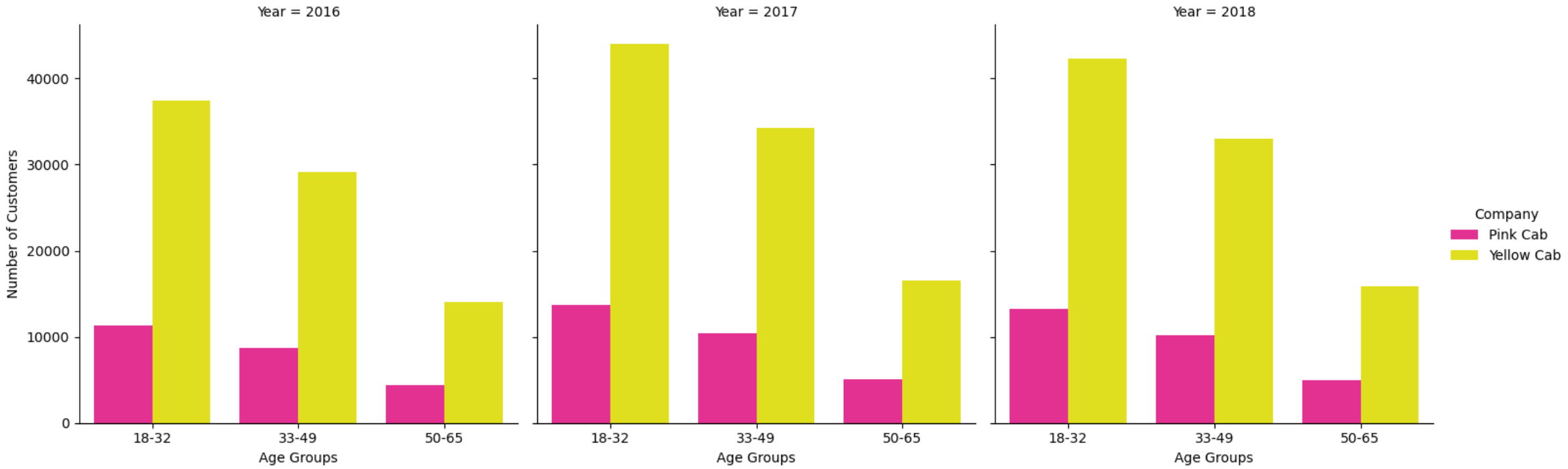
- From Climate New York - Temperature, Rainfall and Averages (usclimatedata.com), contains data on the average amount of rain per month in New York

EDA

Cleaning and Transforming the Data

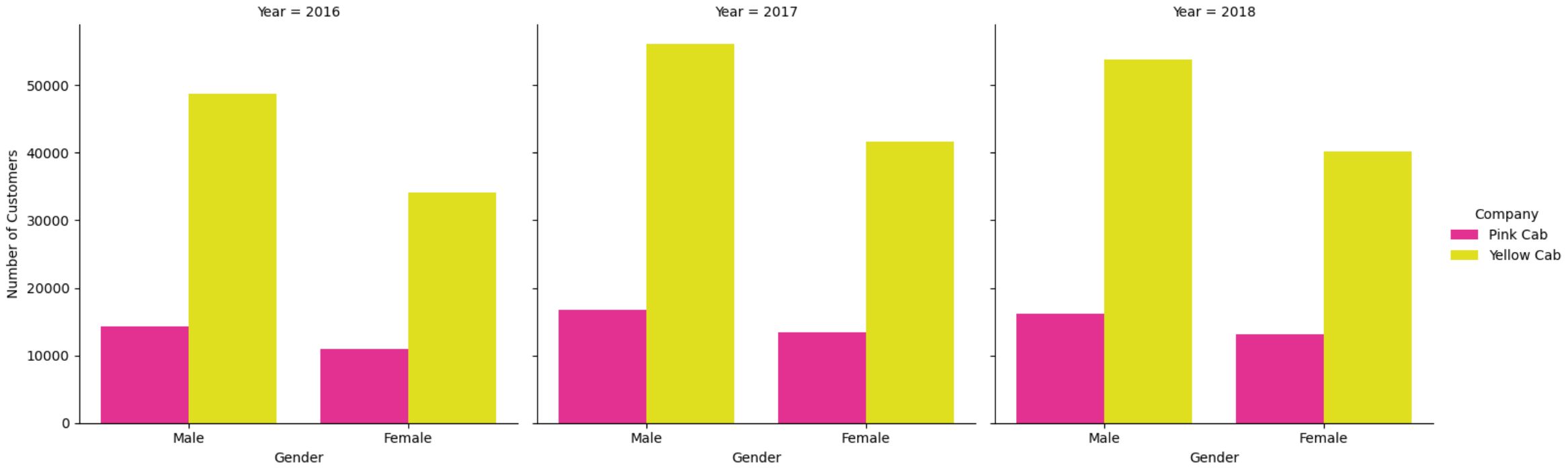
- First the data sets were downloaded and checked for any missing/duplicated values.
- The data was also checked for outliers. The only data which showed a very wide range of values was for the distance travelled, cost of the journey and price charged in Cab_data.csv. It is assumed these are all valid data points as the length of a taxi journey varying a lot if expected.
- The dates of transactions in the Cab_data dataset were converted in standard day-month-year format, the profit (price charged-cost of journey) for each journey computed and added to the data set.
- The Customer_ID and Transaction_ID datasets were merged so that all the customer demographic details were available alongside the transaction details, and this dataset merged with the Cab_data dataset so the demographic details and transaction details were available alongside all the journey details.
- This was the preliminary data analysis, after which more in depth analysis was performed

Customer Base Analysis – Age



From this we can see that across all 3 years and both genders the Yellow cab company dominated. The 18-32 age group had the highest number of users, followed by the 33-49 and 50-65 age groups.

Customer Base Analysis - Gender

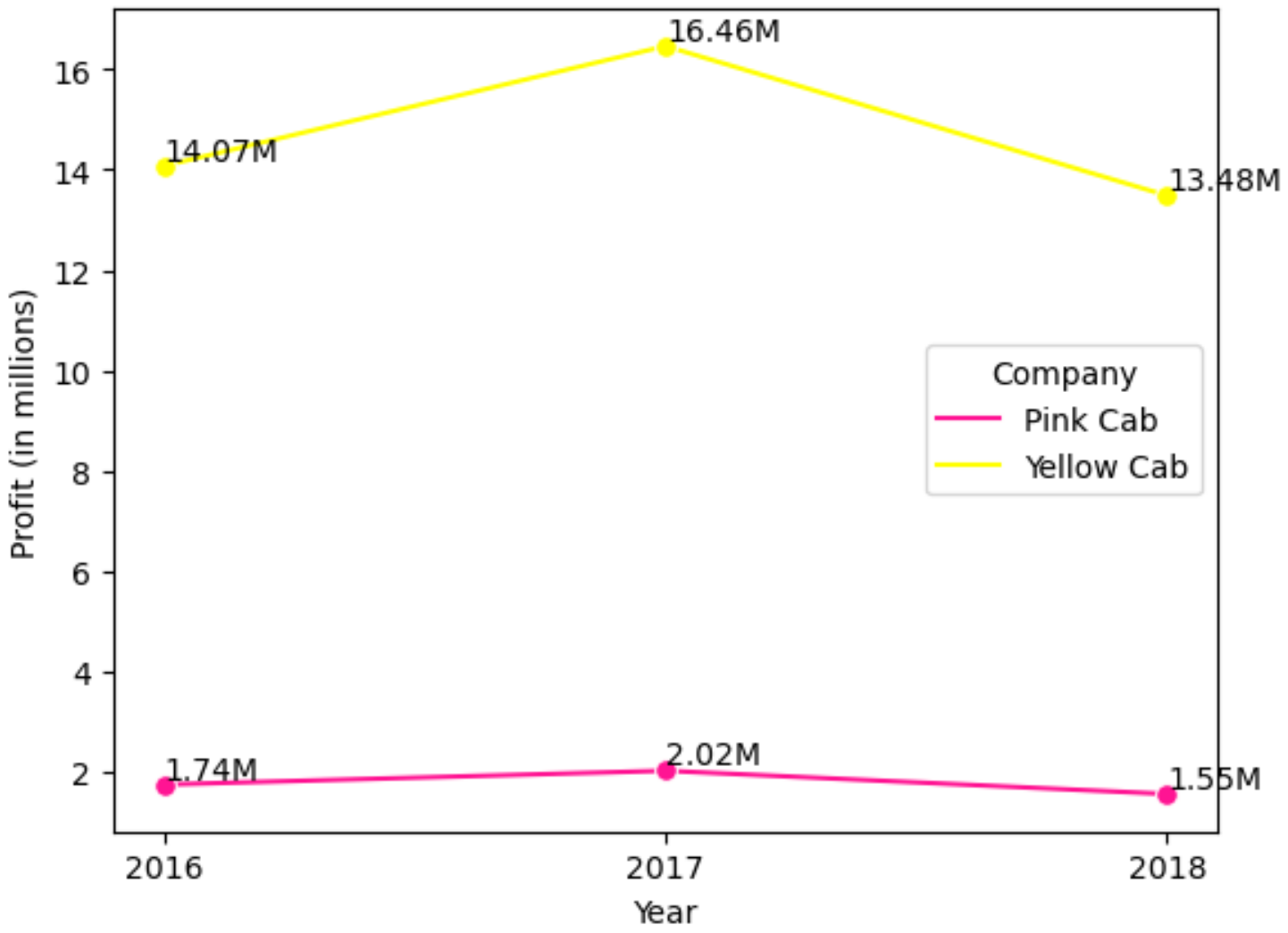


The Yellow Cab company also had more customers across both genders, with both companies having more male customers than females. This difference is, however, small.

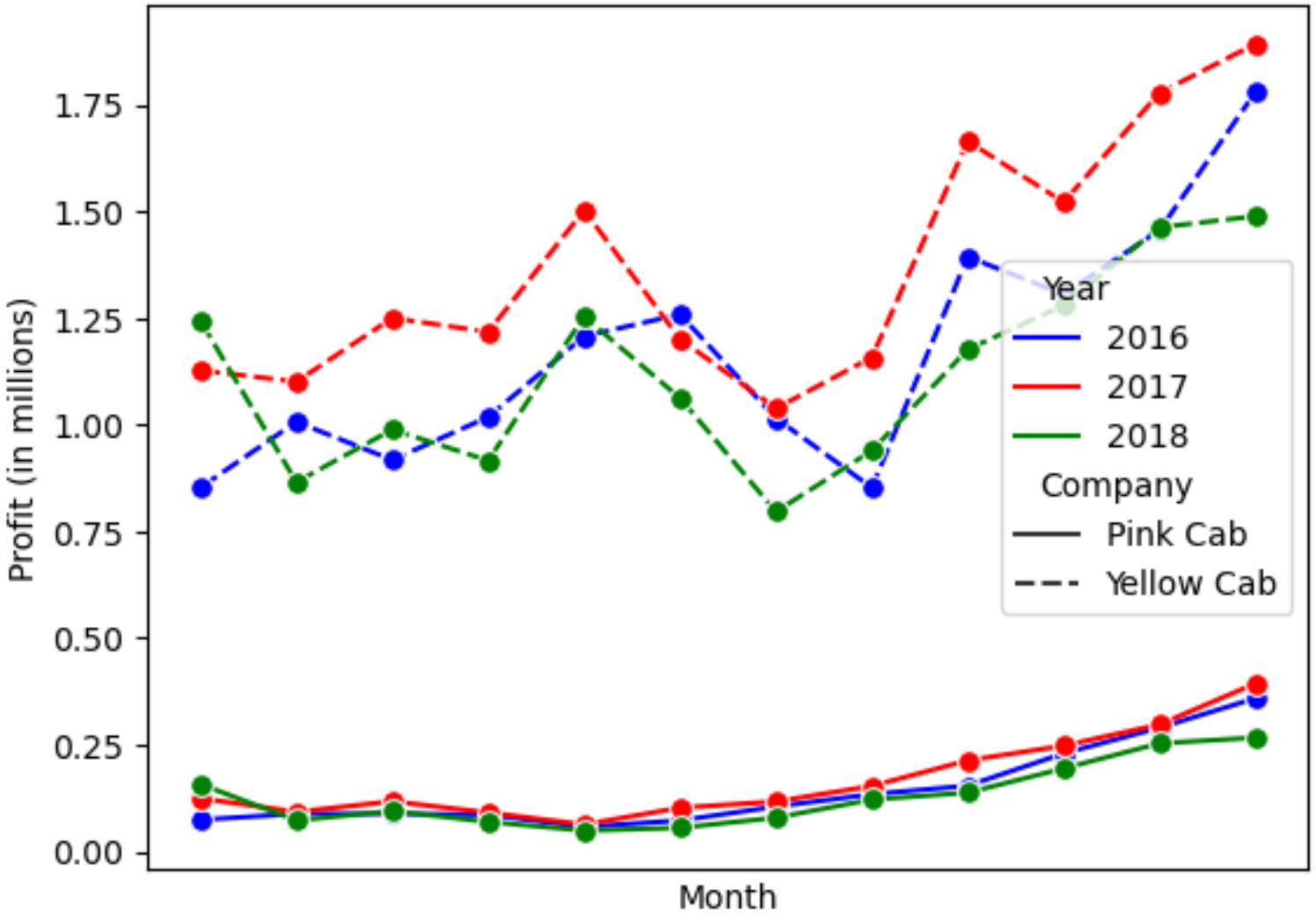
Profit per Company per Year

Company	Profit in Millions
Pink Cab	5.307328
Yellow Cab	44.020373

Yellow cab has a much higher total profit over the 3 years than Pink cab. Although Yellow cab's profit decreases from 2017 to 2018, we can see that Pink cab's does as well.

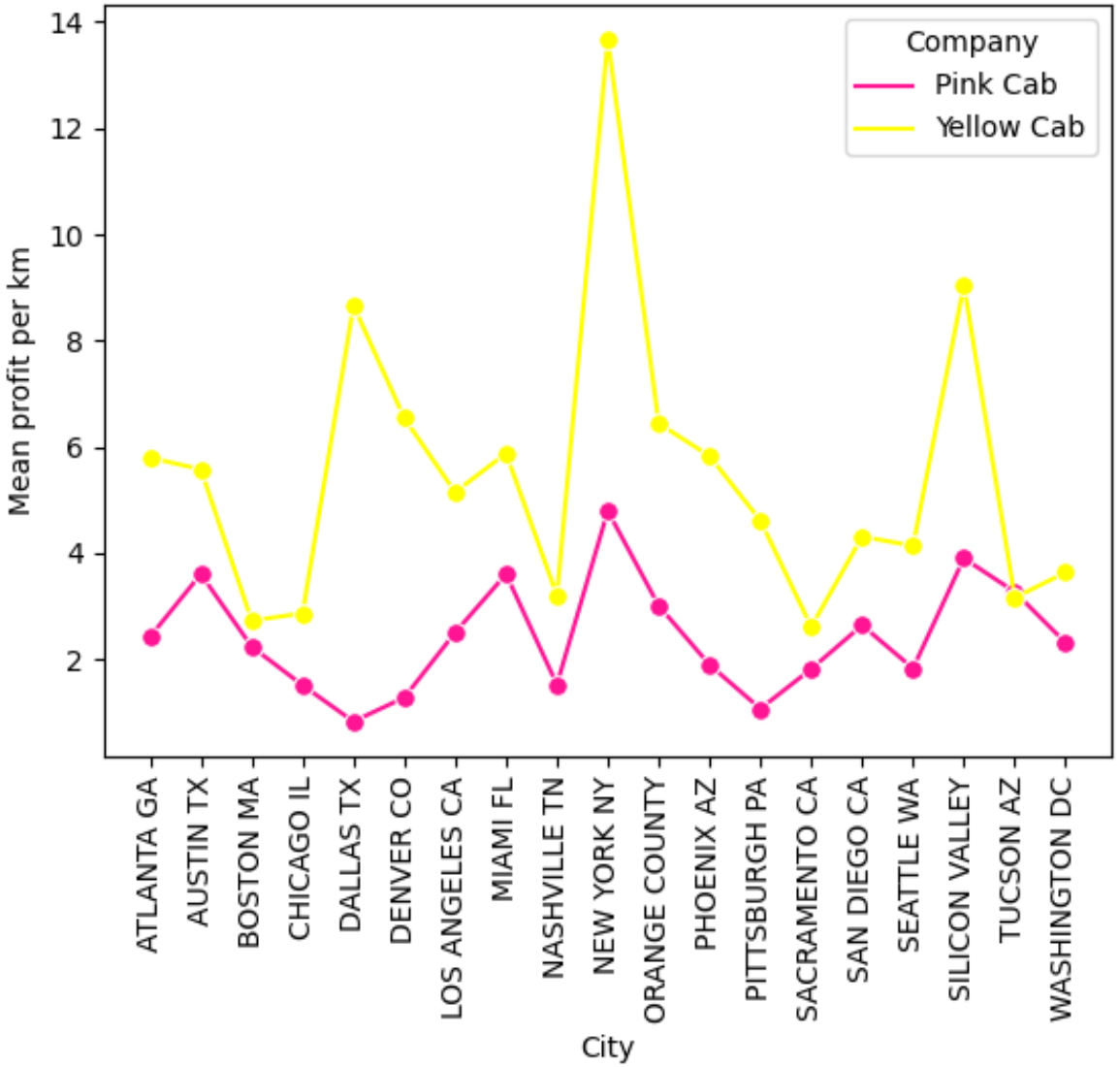


Profit per Company per Month



Yellow cab's profit is consistently higher than Pink cab's throughout the year. We can see that for both companies profit increases in the second half of the year.

Profit per City



Now looking at the profit per km for all the journeys, we can see that New York NY and Silicon Valley are the most profitable cities for both cab companies. Regardless of which cab company is chosen for investment, we recommend increasing business in these two cities as they are the most profitable.

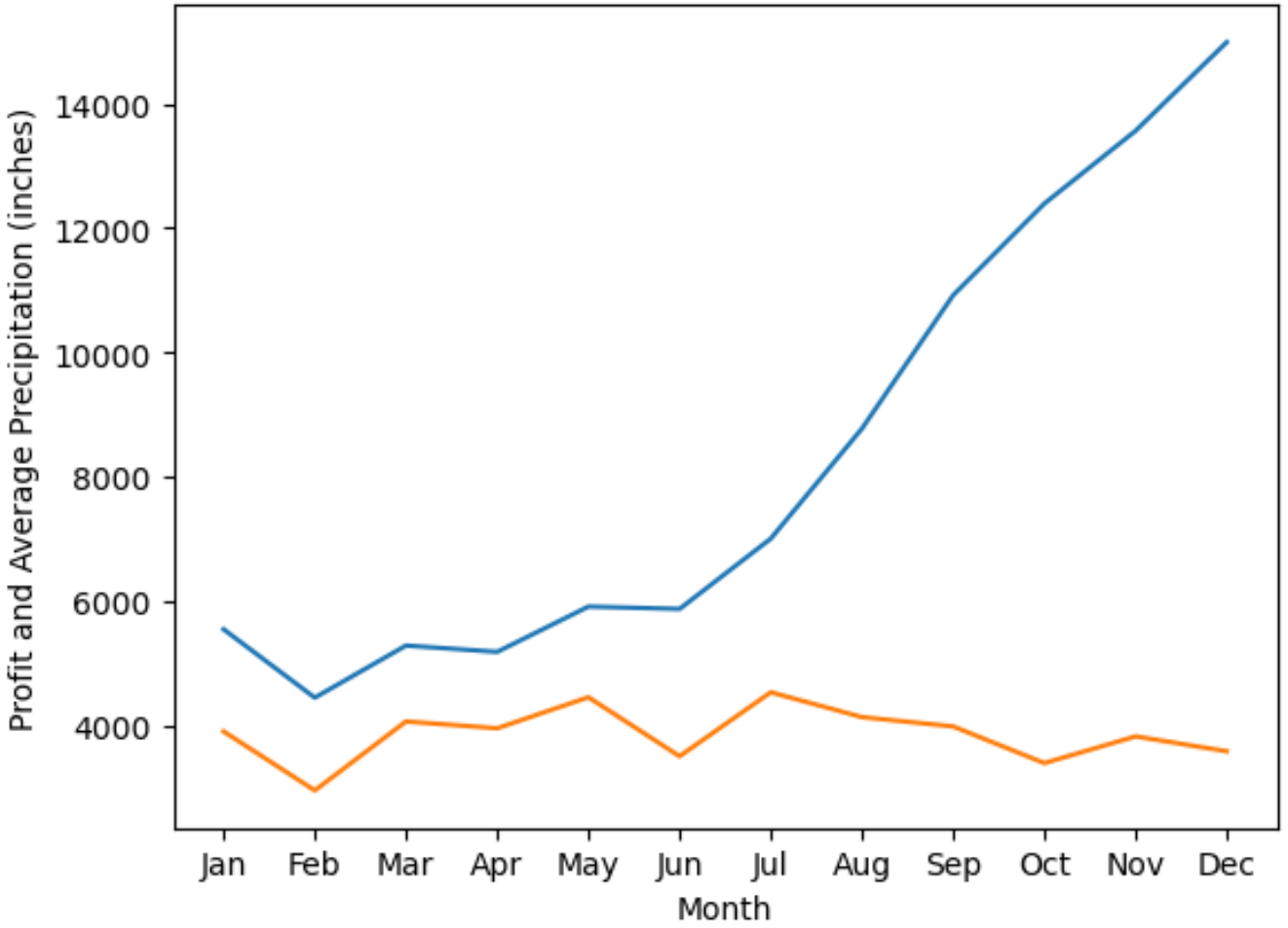
The mean profit per km is higher for Yellow cab for all the cities but 4, therefore we recommend investing in the Yellow cab company, especially in the cities New York NY, Silicon Valley and Dallas TX, as these are by far the three most profitable cities

Number of Returning Customers

Considering a returning customer to be one who has taken more than 3 rides with the company.

The number of returning customers is a strong indication of success for the company, as a company will attract not only new customers, but returning customers, and new customers who were recommended by returning customers.

How Precipitation affects Profits



Looking at the month profit over the past three years (blue line), and the average precipitation per month (orange line, scaled by 1000). We can see that for the first 5 months there is a strong correlation between average precipitation and profit, however this drops off after May.

Summary

- **Age:** Yellow cab has consistently more customers in all age ranges across the three years.
- **Gender:** Yellow cab has consistently more customers across both genders across the three years – the difference in male and female customers is small.
- **Yearly profit:** Yellow cab's yearly profit is higher than Pink cab's for all 3 years. Although Yellow cab's profit dips in 2018, so does Pink cab's.
- **Monthly profit:** Both cab companies' profit increases in the second half of the year, with the gradient of increase steeper for Yellow cab.
- **City:** Yellow cab has a greater profit per km for all cities bar 4, with New York, Silicon Valley and Dallas Texas having the highest profit per km for Yellow cab.
- **Returning customers:** Yellow cab has a much higher number of returning customers than Pink cab
- **Precipitation:** The monthly precipitation seems to affect the profits up until May, but after that there's no trend

Based off this information, we would recommend Yellow cab for investment as their number of customers and profits are much higher than Pink cab's for all demographics. We would also suggest focussing the investment in the cities of New York, Silicon Valley and Dallas Texas, and in the months of January-May looking at the weather forecast as the amount of precipitation will affect profits closely.

Thank You