



**Data Glacier**

Your Deep Learning Partner

# G2M Case Study

Virtual Internship

June 2021

# Background – G2M (cab industry) case study

## Business problem:

- XYZ is an American private equity firm. Due to the recent growth of the cab industry, it is planning to invest in the cab industry.
- Goal: investigate the datasets pertaining to 2 cab companies, Yellow Cab and Pink Cab, and explain which company would make for a more lucrative investment.

## The analysis involves:

- Data overview
- Investigating no. of customers and transactions for each company (e.g., for each city)
- Understanding customer demographics
- Investigating number of quarterly transactions, profit and margin (and determining the most profitable company), and relationships between relevant variables
- Studying seasonality
- Summarising results and recommending a company for investment

# Overview of data and assumptions

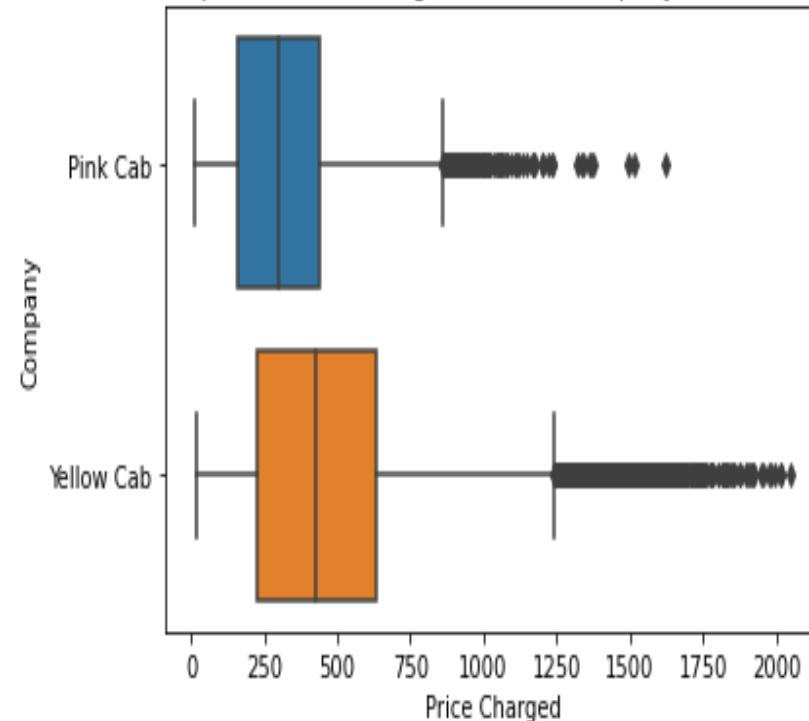
## Overview:

- There are 16 features (14 original, 2 created). This does not include features that are summed or averaged over a time period (e.g., monthly/quarterly profit, average profit)
- Time frame: 02/01/2016 to 31/12/2018
- Master data was created by merging City, Customer ID, Transaction ID, and Cab Data datasets.
- Rows in master data: 359,356 (after merging)
- About 80706 rows were lost after merging the original files

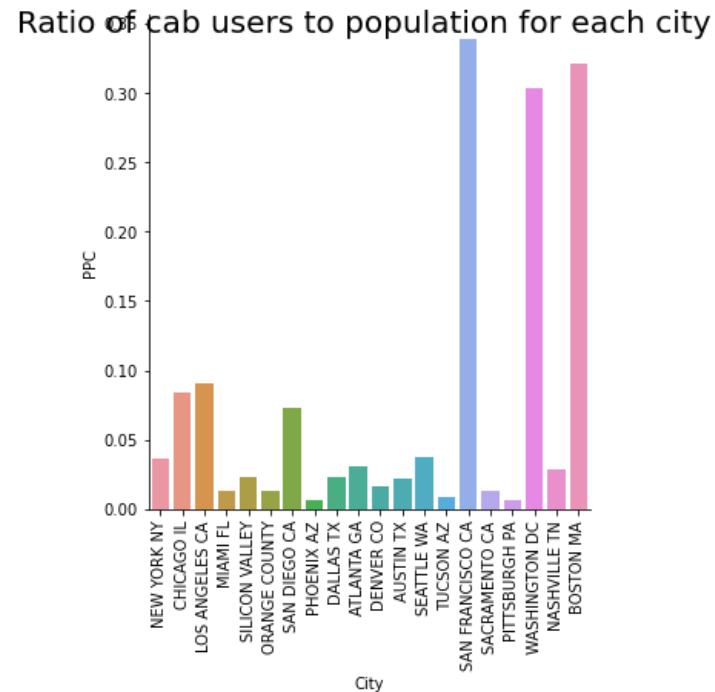
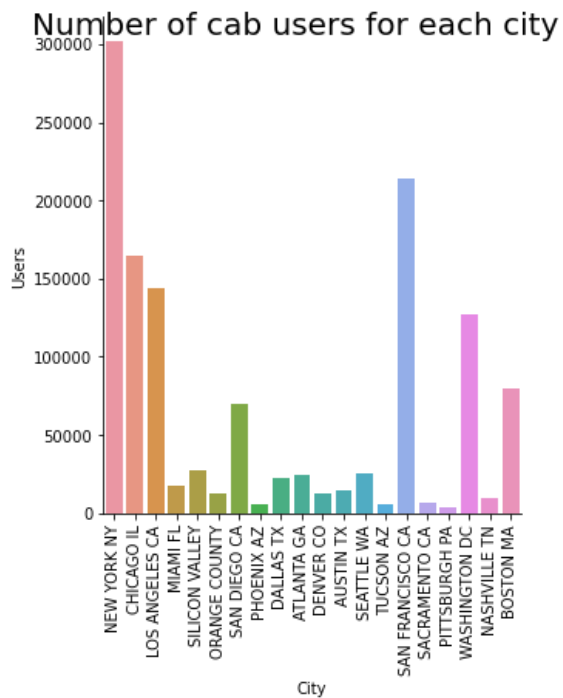
## Assumptions:

- Outliers are present in
- Profit is calculated with  $\text{Price\_Charged} - \text{Cost\_of\_Trip}$
- Margin can be calculated by sum of profit/sum of Price\_Charged
- The 'users' variable of the City dataset is treated as number of cab users in the city, including those for Yellow Cab and Pink Cab.

Box plot for Price Charged for each company, with outliers



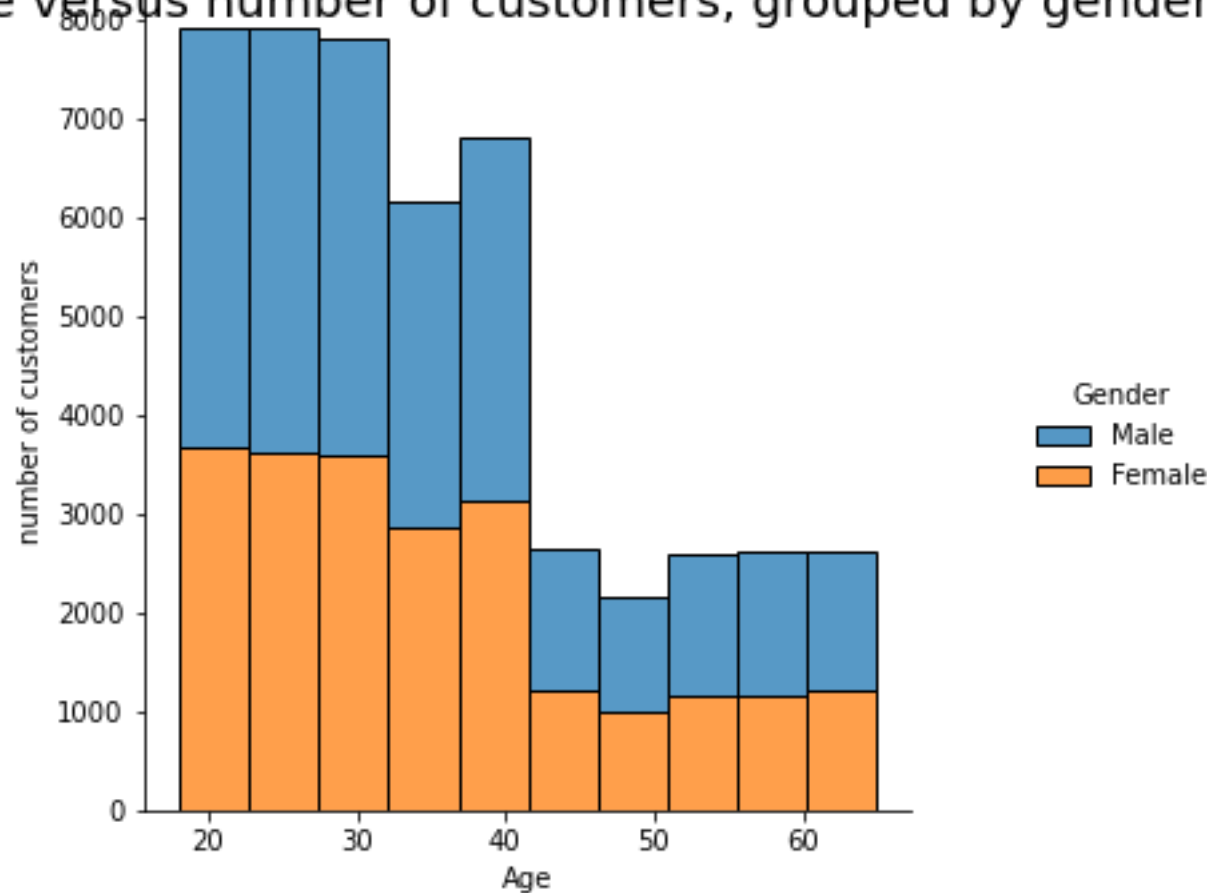
# Distributions of cab users for each city



New York has the most cab users (~300,000) and Pittsburgh has the lowest. San Francisco has the highest proportion of cab users to its population (~0.35).

# Distribution of cab users, grouped by age and gender

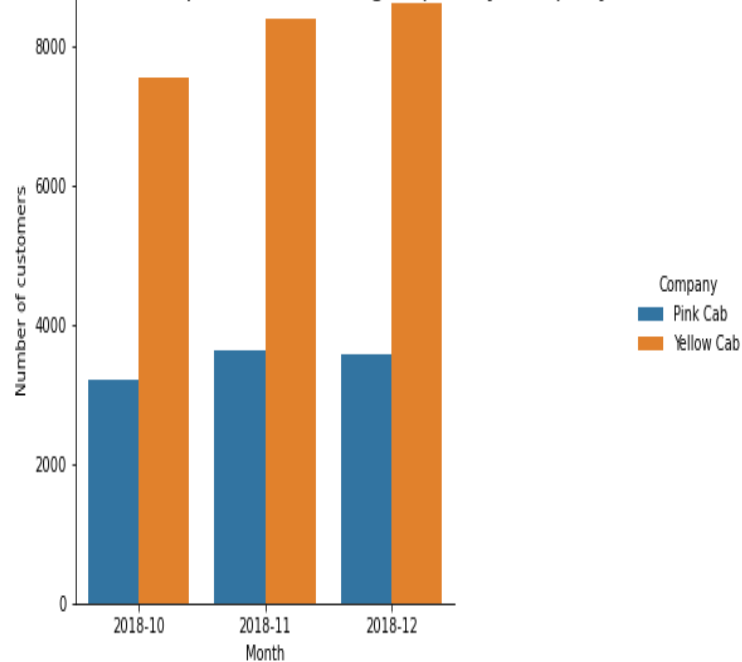
Age versus number of customers, grouped by gender



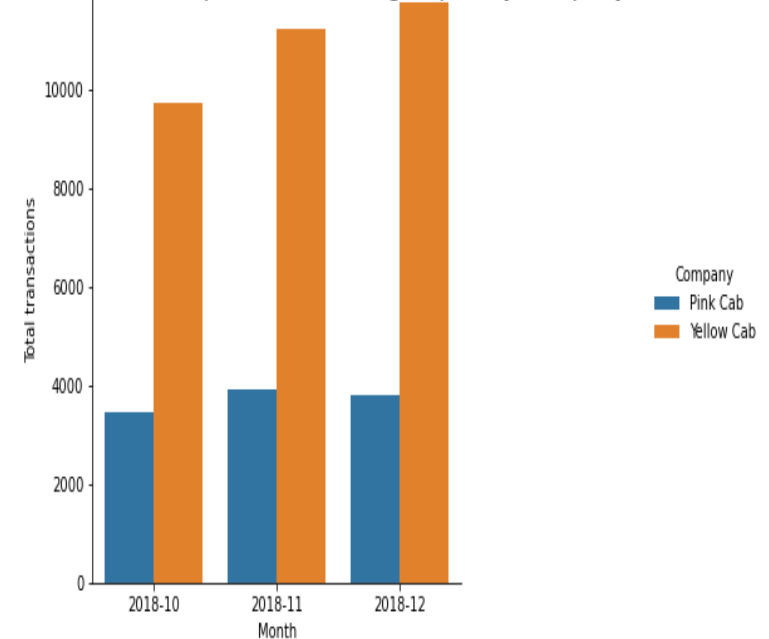
There are a lot more cab users in the lower age ranges (from 18-21 to 38-41) compared to the higher ranges (41+ ranges). There are approximately the same number of males and females in each age group.

# Number of customers / transactions for each company during last quarter of 2018

Number of customers for last quarter of 2018, grouped by company and month

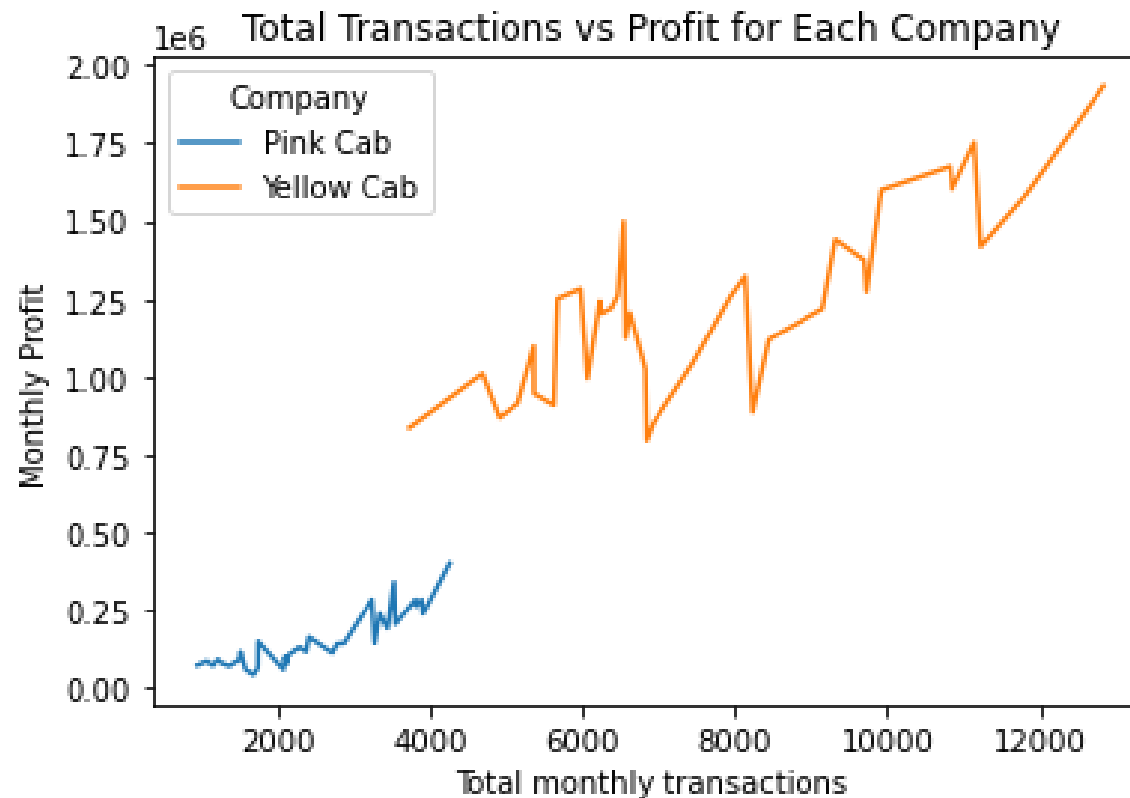


Number of transactions for last quarter of 2018, grouped by company and month



It's useful to see the recent performance of the cab companies. In the last quarter of 2018, Yellow Cab had considerably more transactions and cab users compared to Pink Cab (over twice as much for each).

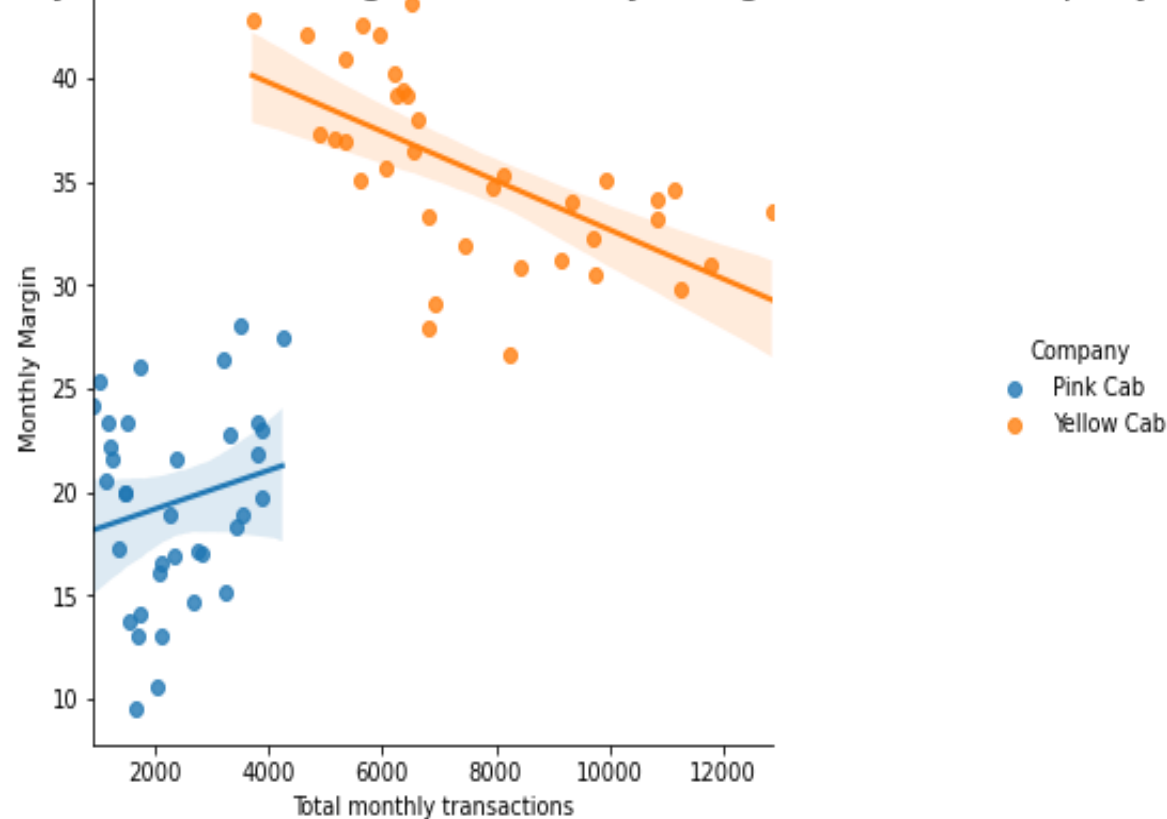
# Number of customers for each company during last quarter of 2018



As we would expect, the monthly profit generally increases with increased number of transactions for both companies.

# Total monthly transactions vs monthly margin for each company

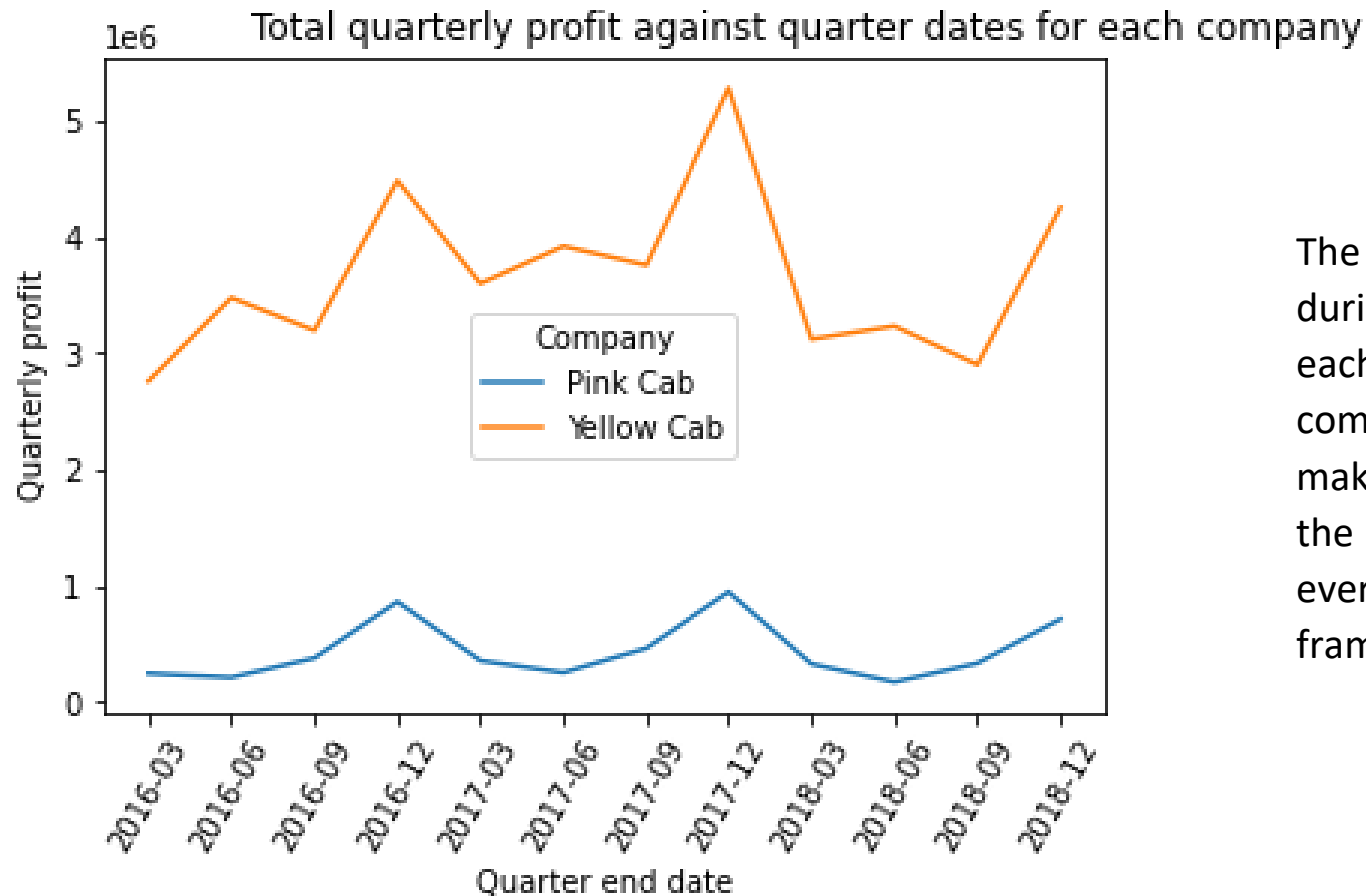
Total monthly transactions against monthly margin for each company



For Yellow Cab, as the number of monthly transactions increases, the monthly margin decreases. For Pink Cab, monthly margin tends to increase with the number of transactions, but the association is weaker.

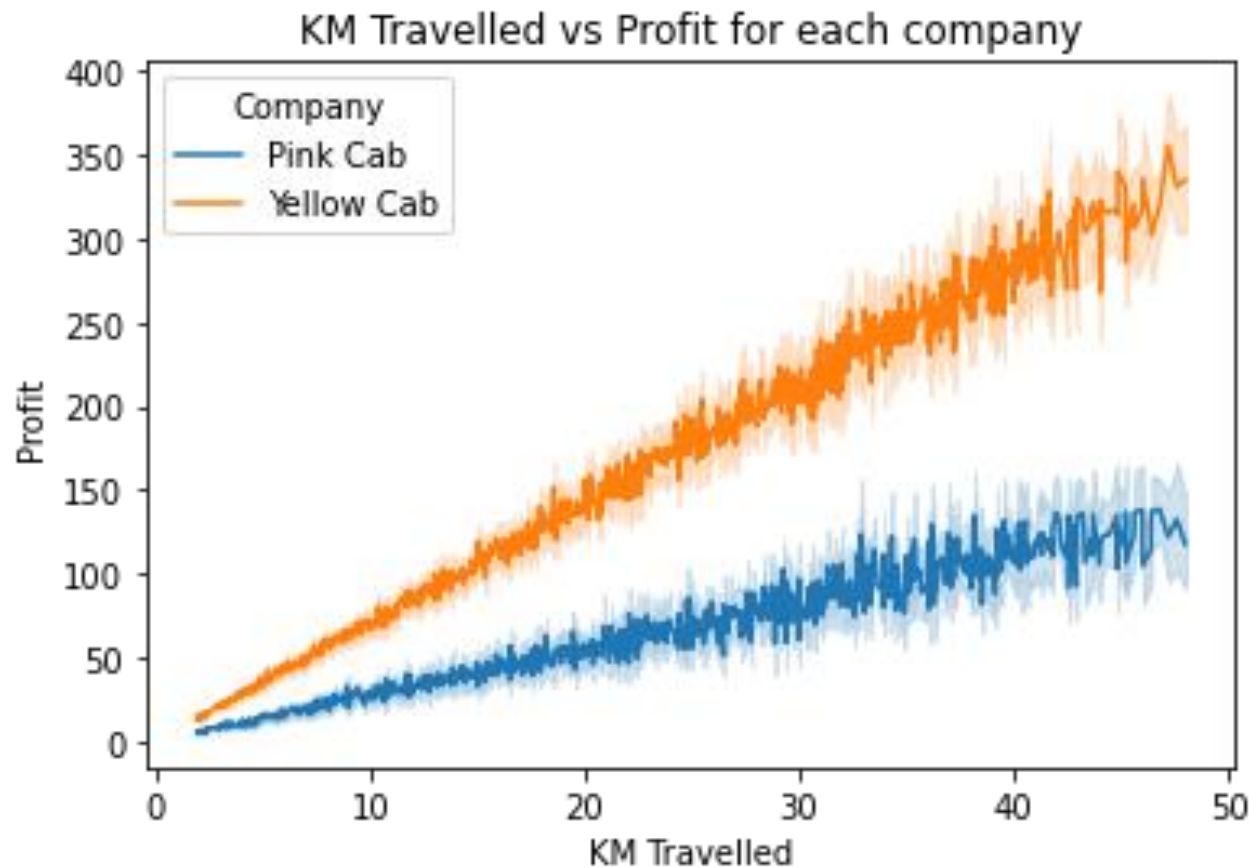


# Profit for each company per annual quarter



The profit tends to spike during the last quarter of each year for both companies. Yellow Cab is making at least 3 times the profit of Pink Cab for every quarter in the time frame.

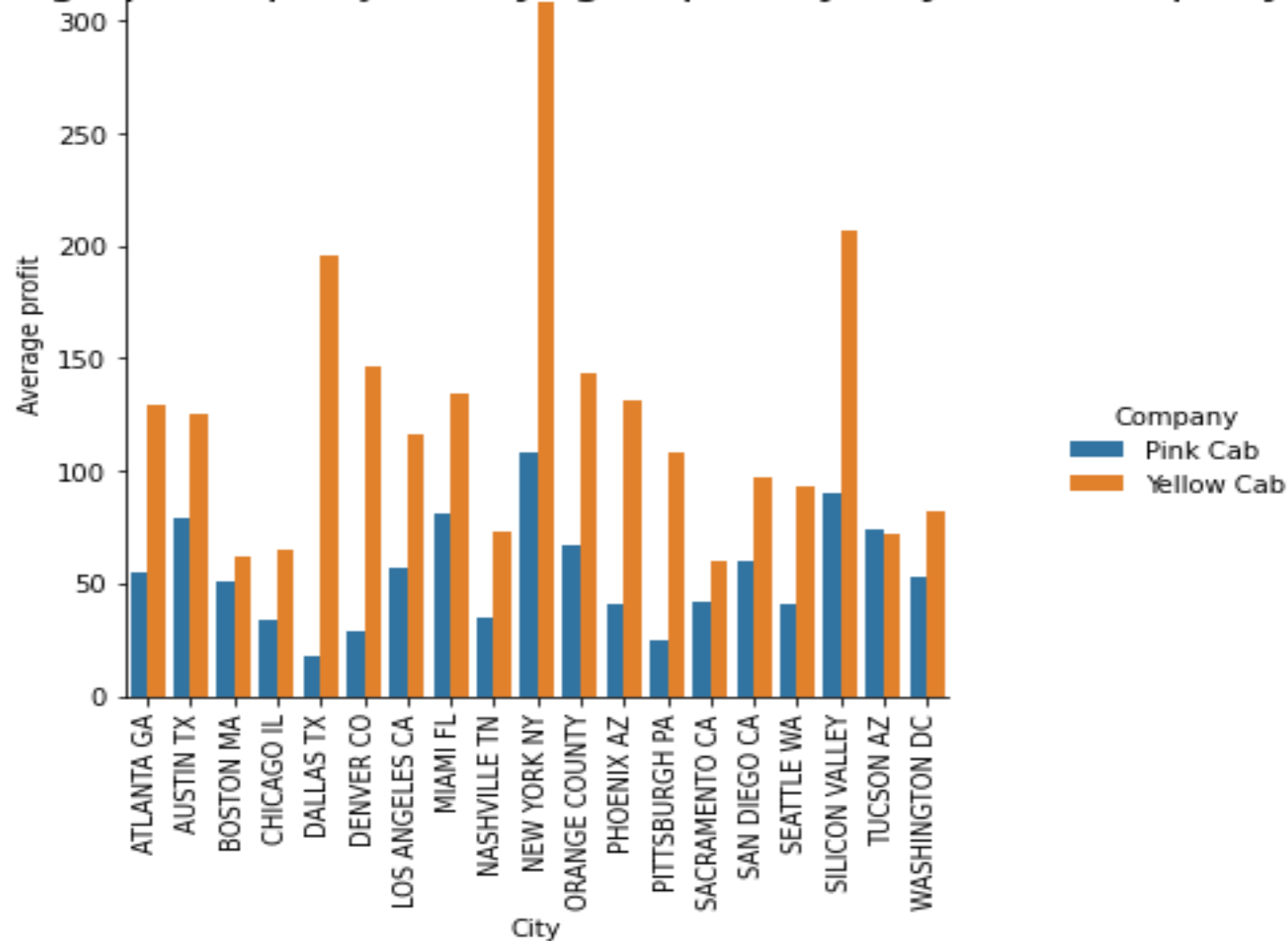
# Relation between distance travelled (KM) and profit per journey?



As we would expect, longer journeys are more profitable for both companies. Yellow Cab is making considerably more profit for the same KM travelled though.

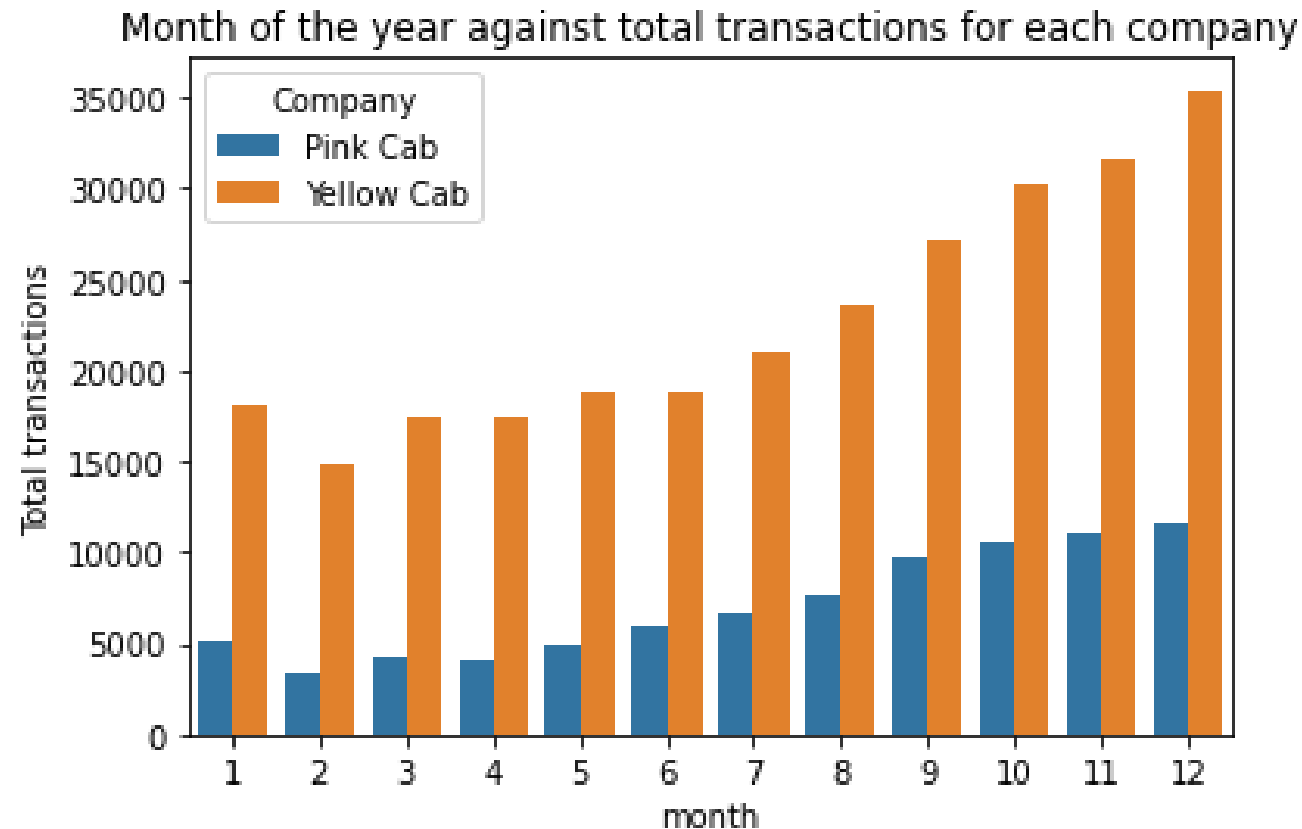
# Average profit per journey vary between cities for each company

Average profit per journey, grouped by city and company



Yellow Cab is making considerably more profit per journey on average in every city in the dataset, except for Tuscon AZ.

# Seasonality in cab usage



For both companies, the general trend is that in later months of year, there are considerably more transactions than earlier months. This is especially noticeable from July to December.

# Recommendations

From the analyses, it is clear that Yellow Cab is considerably more profitable than Pink Cab. Yellow Cab has a much larger customer base and higher number of transactions. Yellow Cab is more popular and more profitable on average in almost every city, and is generating much more profit per KM travelled. We can confidently say that Yellow Cab is a safer and more lucrative investment than Pink Cab.

Therefore, we recommend Yellow Cab for investment.

# Thank you



**Data Glacier**

Your Deep Learning Partner