Data Glacier
Your Deep Learning Partner

Exploratory Data Analysis

G2M Cab Case Study

20-Nov-2022

Executive Summary

- XYZ is a private firm in the US that is looking to invest in the cab industry given its growth and the key players involved
- They are not familiar with the market

- Objective: Given four data sets provide a recommendation for the most profitable investment for XYZ to make

Approach

Using the data provided, my approach was to compare the average values of many different variables to determine which city and cab company would be the smartest investment for XYZ. I am specifically looking for highest volume of cab trips as well as the highest price charged for each trip. Along with this, I will be using outside data to determine if there are any external factors that makes one place more profitable than another.

Assumptions:
- There were no significant outliers in the data used, so it can be assumed that there is no need to account for the data being skewed in any way

Total Trip Volume Analysis
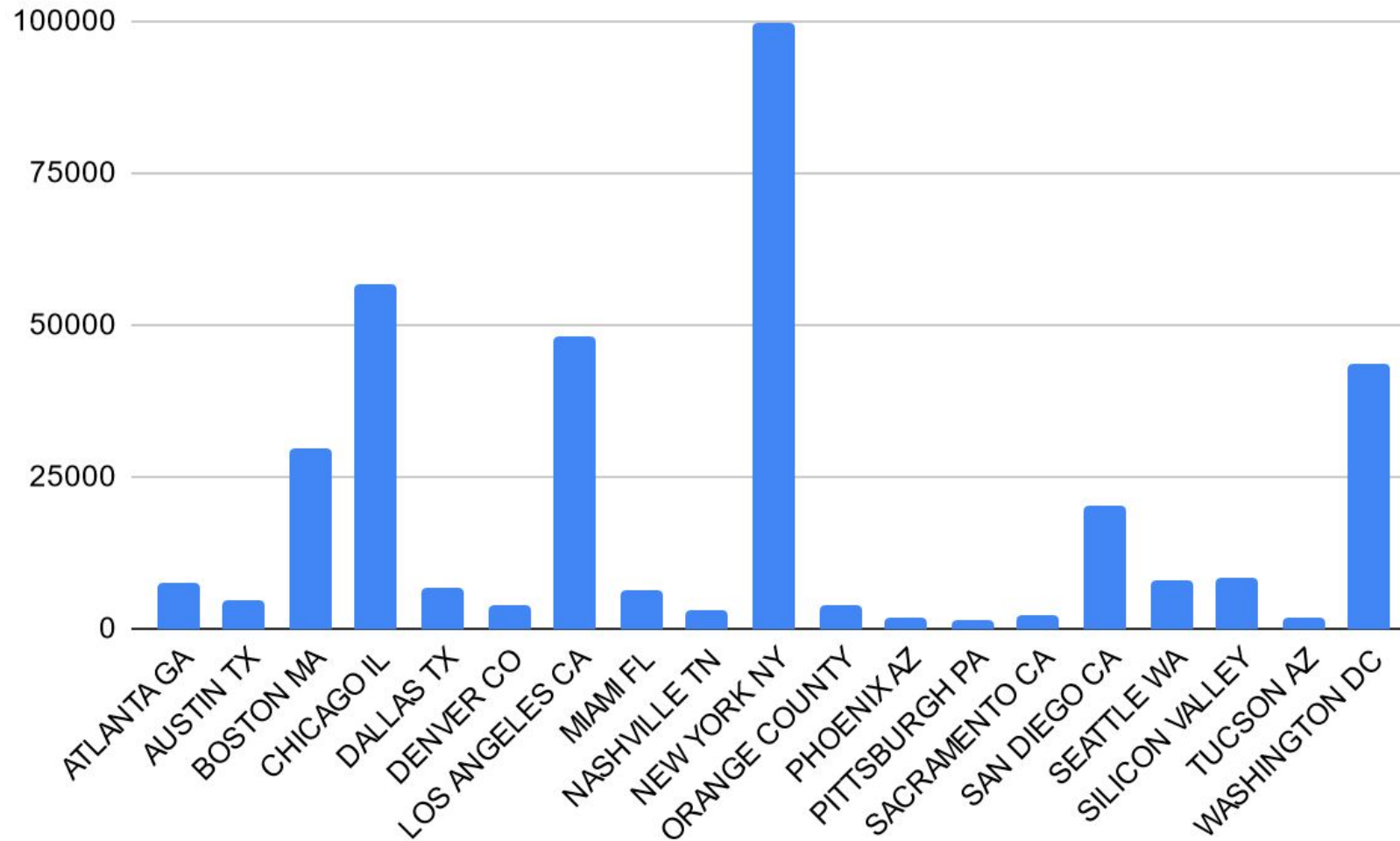
- This graph shows the total number of cab trips recorded in each city

- New York clearly has the highest volume of cab trips out of all the cities recorded

- Pittsburgh, PA has the lowest

Profit Analysis (Per City)
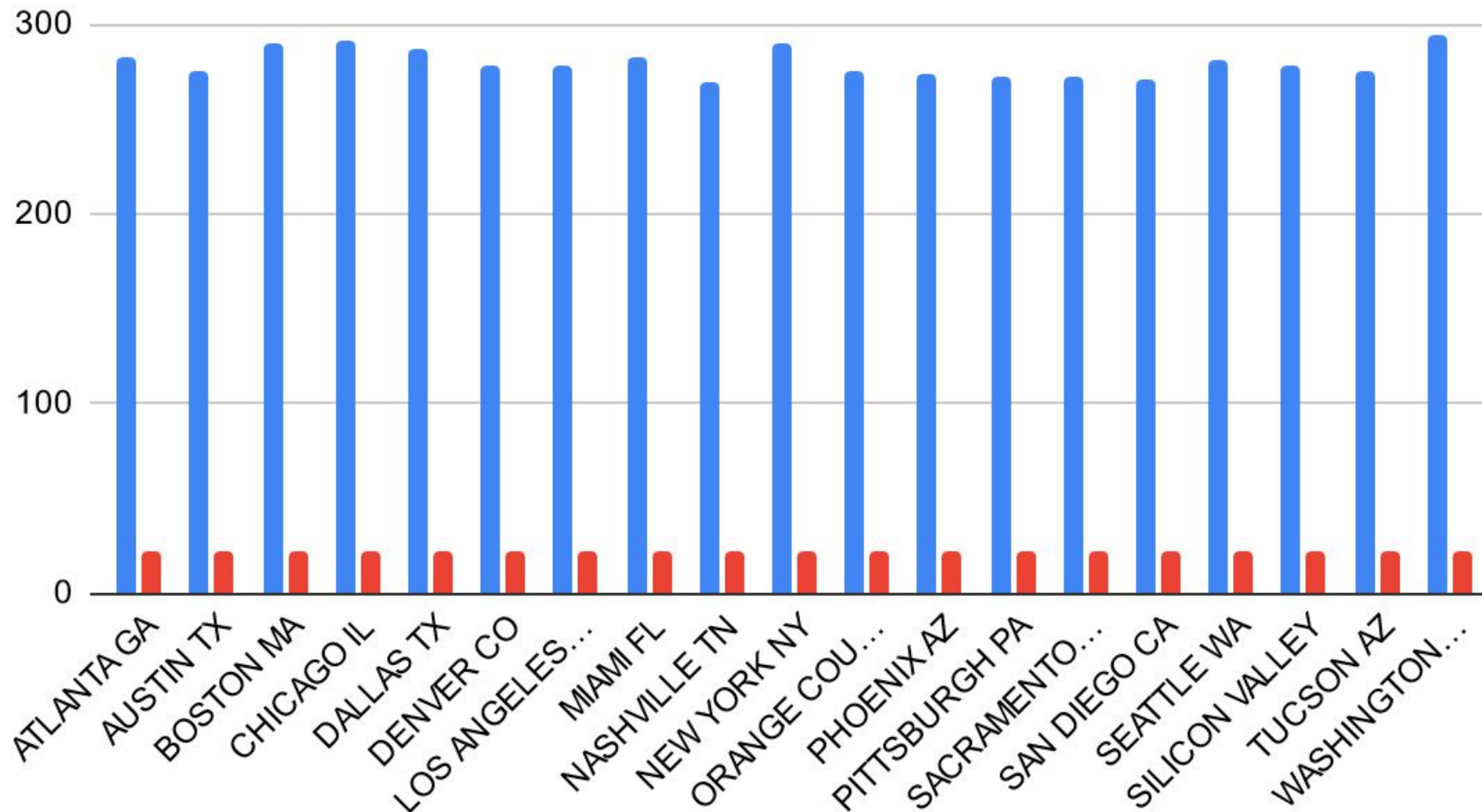
- This graph shows the average total cost of each trip in each city (blue) as well as the average distance of each trip in each city (red)

- There is no notable difference between the average distance traveled in each city, so this can be disregarded when analyzing the cost per trip

- Washington DC and New York have the two highest average costs. Chicago and Boston are also very high

- Nashville, Tucson, and San Diego have relatively low costs compared to the rest of the cities

Profit Analysis (Per Company)
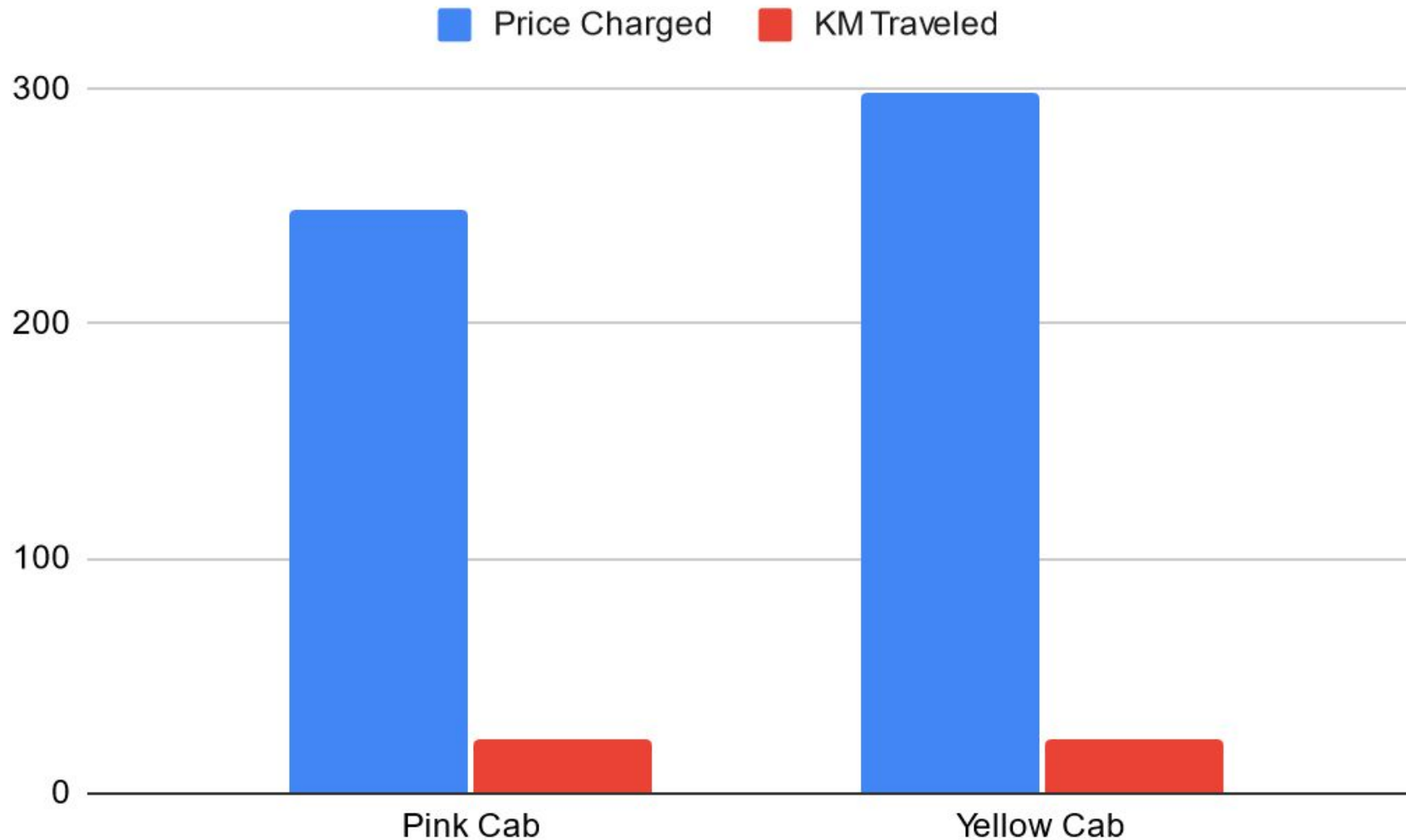
- This graph shows the average cost charged during each trip by each company (blue) as well as the average distance traveled per trip for each company (red)

- Again, there is no obvious difference in the average distance traveled between the companies so this factor can be disregarding when comparing costs

- Yellow Cab company clearly generates more money on average for each trip

Trip Volume Separated by Company

Number of Trips per Company per City
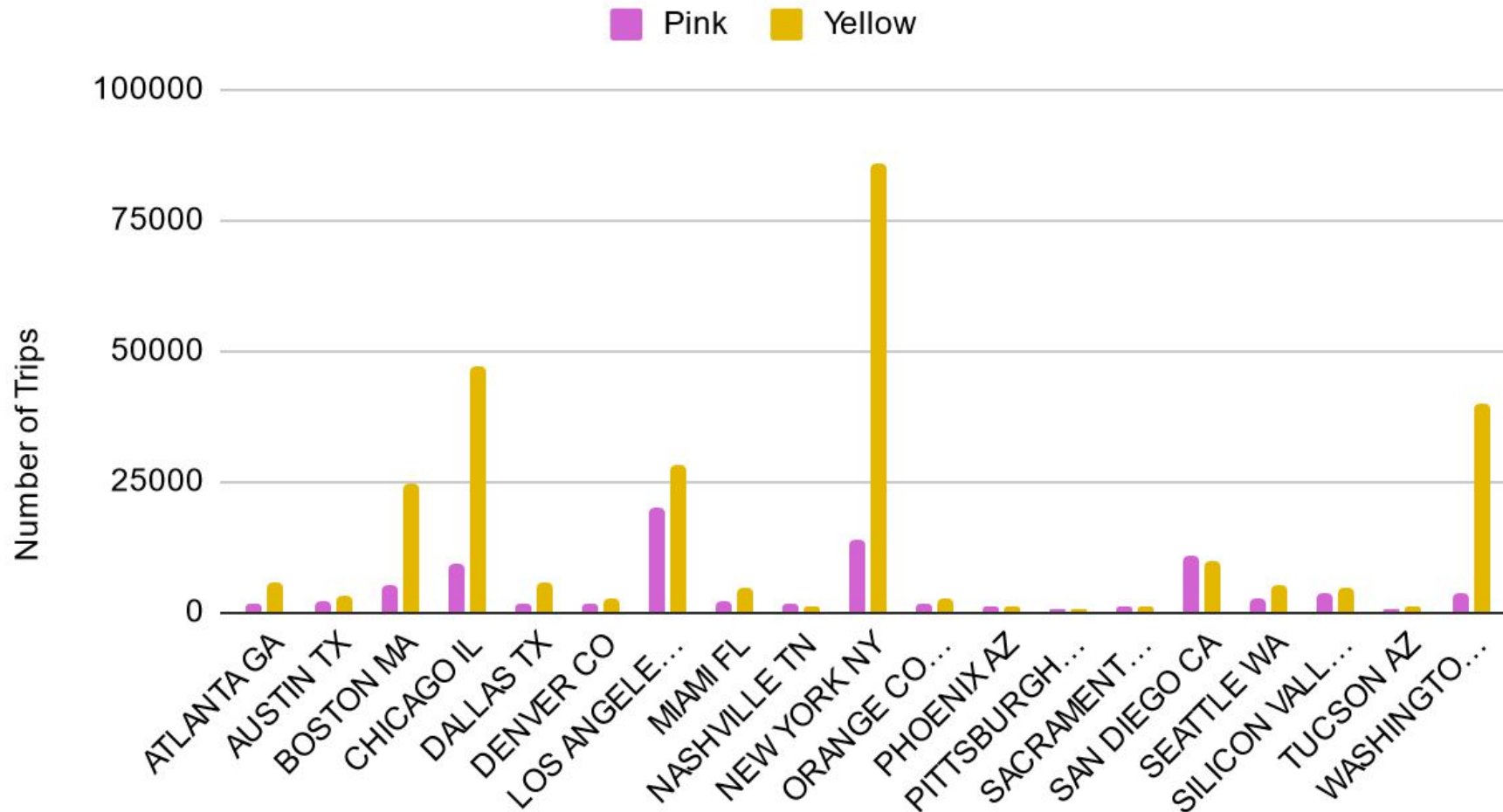
- This graph shows the total number of trips recorded in each city for each company

- New York has a much higher volume of trips from the Yellow Cab company than any other city and the second highest number of trips through the Pink Cab company

- Yellow Cab company has more trips in every city except Nashville, Pittsburgh, and Sacramento

Seasonal Demand for Cabs
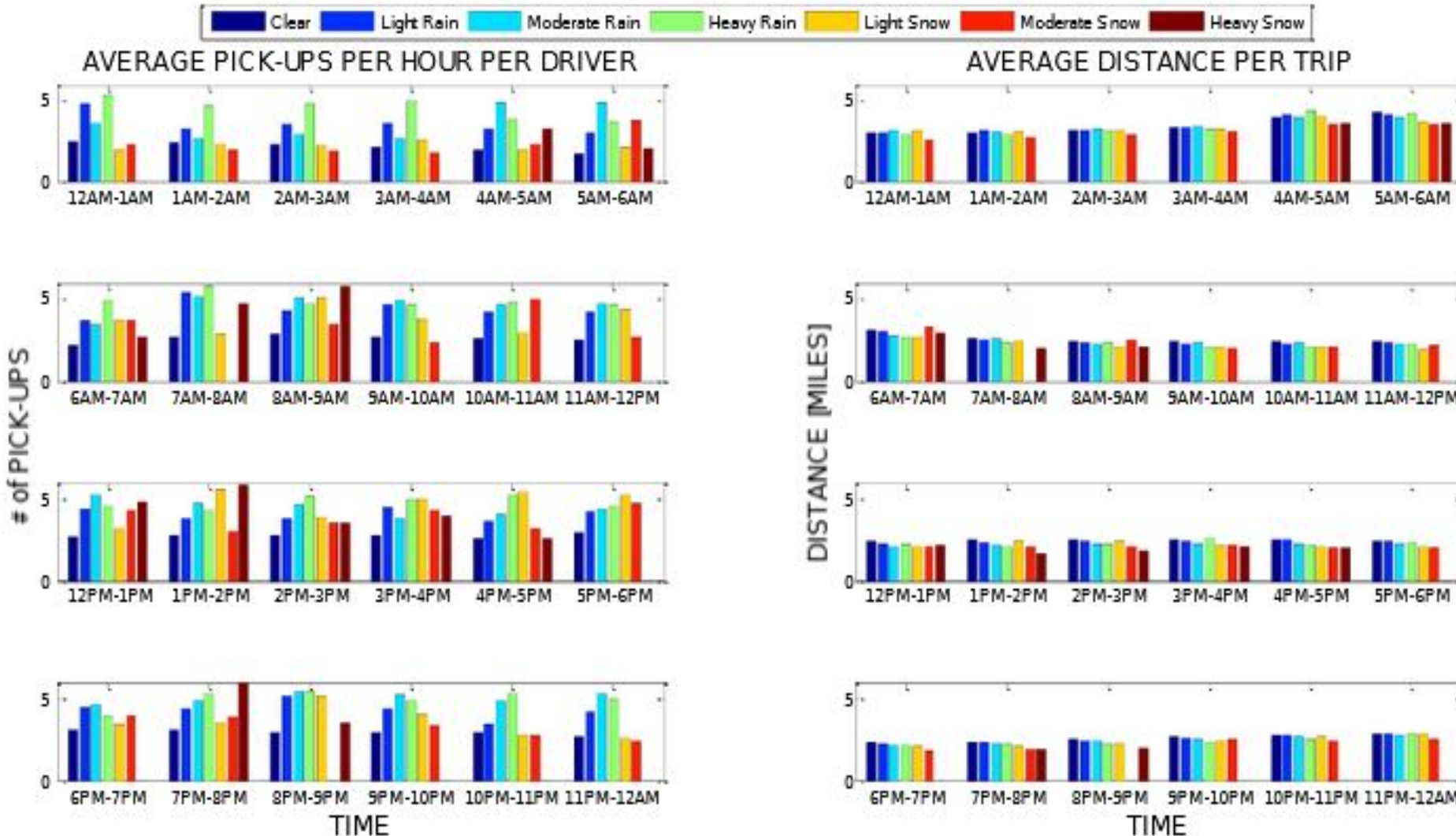
- There is a clear correlation between the average number of pick-ups per hour and the weather conditions

- People are more likely to use cabs in rain and light snow

- This makes sense given my previous data analysis, as the cities with the most recorded cab trips have colder climates than those with the lower number of trips

Recommendation

After evaluating the data, I would recommend investing in the Yellow Cab company, specifically in New York

- **Profit:** The Yellow Cab company has higher costs charged per trip overall
  - Yellow Cab also reports higher prices charged relative to distance traveled

- **Customer Reach:** 17 out of 20 cities recorded more trips through the Yellow Cab company than the Pink Cab Company
  - Higher costs for each trip in addition to greater trip volume results in an even greater ROI
  - New York and Chicago are the two cities with the highest overall demand for cabs

- **Cab Demand in New York:** In New York, Yellow Cab has 6.15 times the amount of trips than Pink Cab
  - New York has a cold climate which will also lead to higher cab usage

- **Cab Demand in Chicago:** In Chicago, Yellow Cab has 5 times the amount of trips than Pink Cab
  - Like New York, Chicago is also in a cold climate so there will be more overall cab pick-ups

Based on the data presented and the analysis above, I recommend XYZ to invest in the Yellow Cab company in New York.

Thank You