Exploratory Data Analysis

G2M Case Study

Halit Ayberk DEMIR

20.03.23

Data Glacier

Your Deep Learning Partner

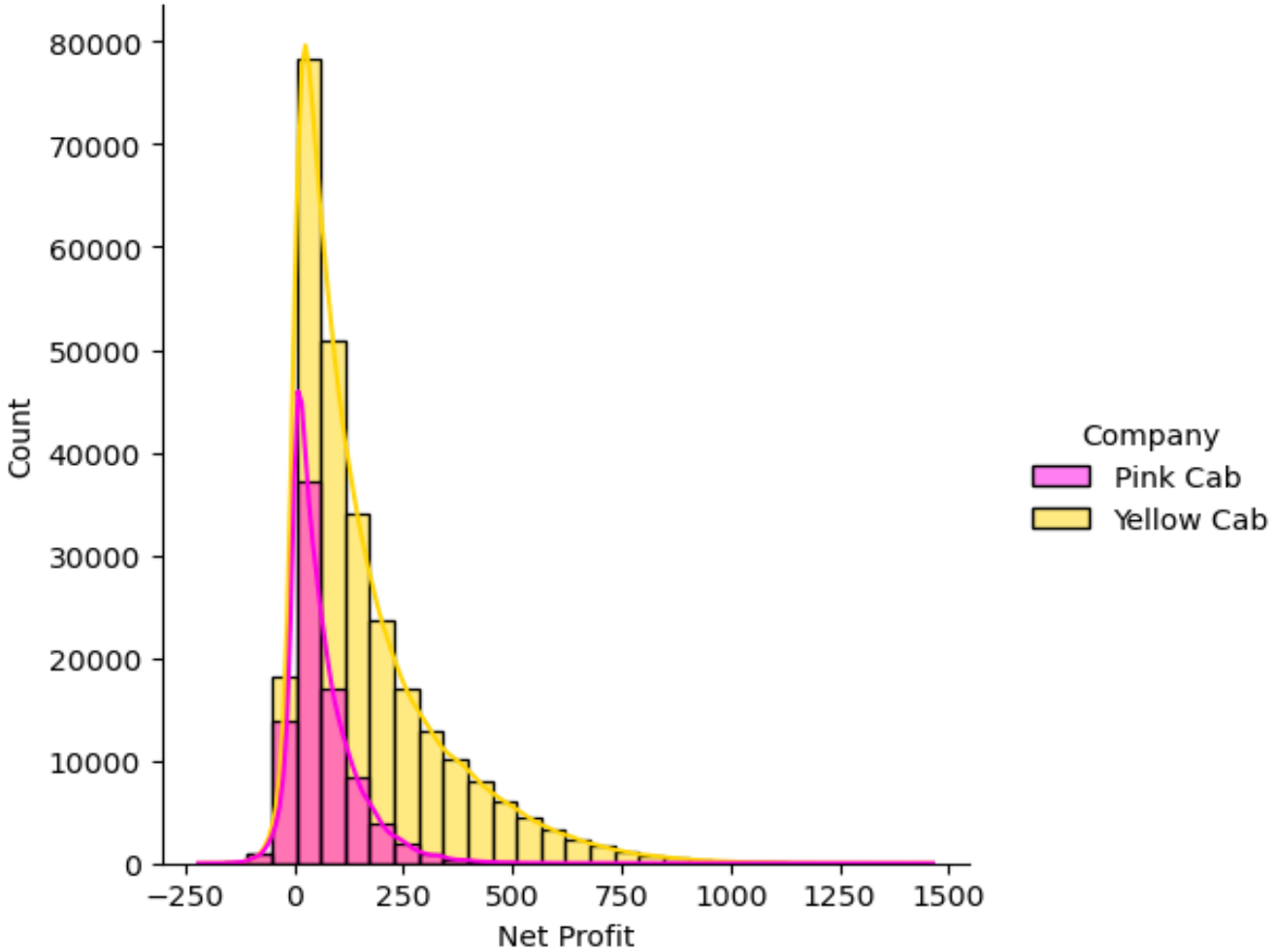
Introduction

- At this project Pink Cab and Yellow Cab company insights will be investigated;
  - Difference of the Net Profits in the companies
  - The firms' net profit in different cities
  - Seasonality of the drives
  - Both cab firms' reach for unique customers
  - Their profits in differently populated cities
  - Their profits related to customers economic status
  - Their profit trend in 3 years

Univariate Analysis

Data Glacier
Your Deep Learning Partner

Net Profit Distribution For the 2 Cab Firms.

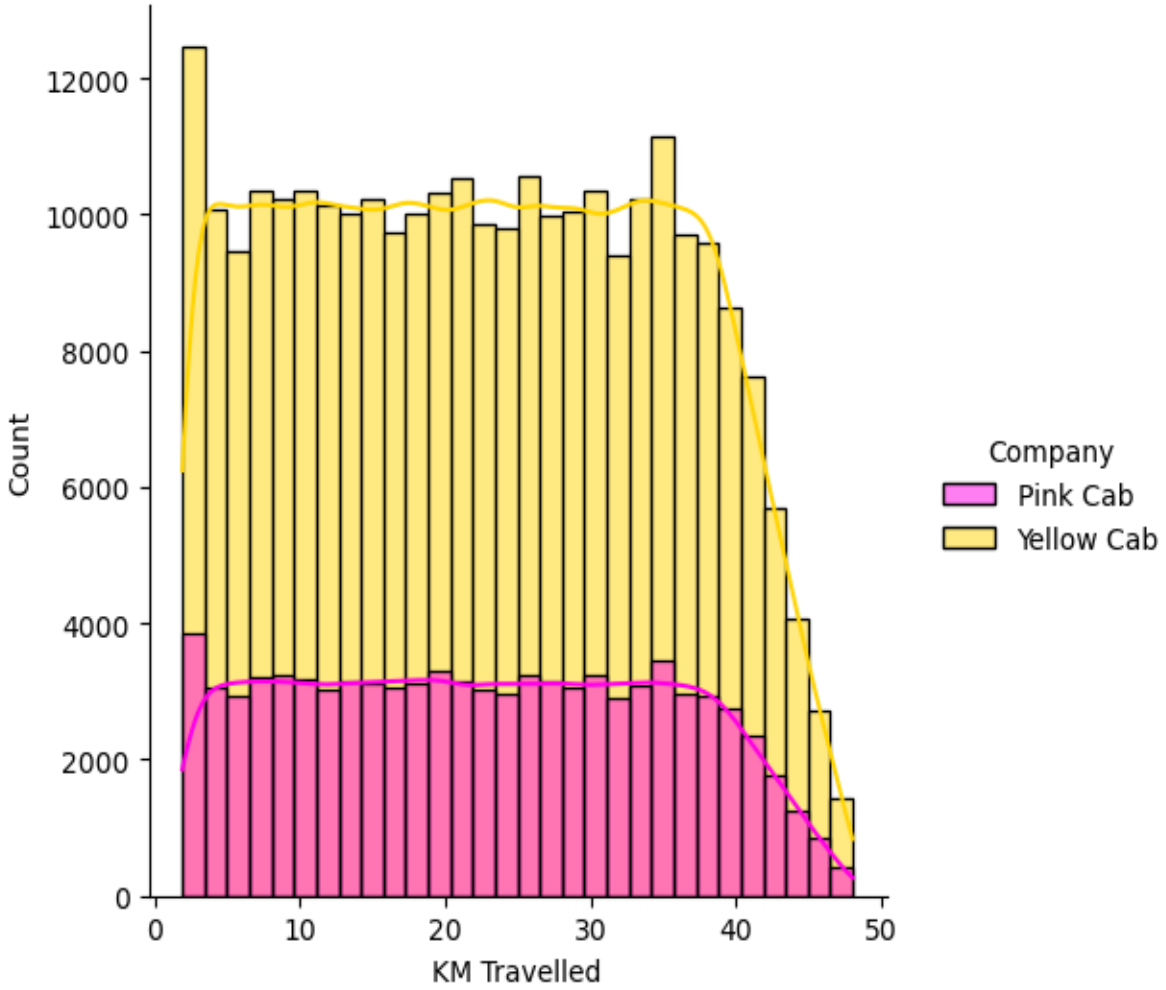


- Mean profit of the Yellow Cab per drive is 160.26$
- Mean profit of the Pink Cab per drive is 62,3$
- Median profit of the Yellow Cab per drive is 102.0$
- Median profit of the Pink Cab per drive is 102.0$
- Both of the Data is positively skewed, Pink Cab has lower mean profit then the Yellow Cab firm.

22.55

Data Glacier
Your Deep Learning Partner

Distance Travelled Distribution For the 2 Cab Firms.



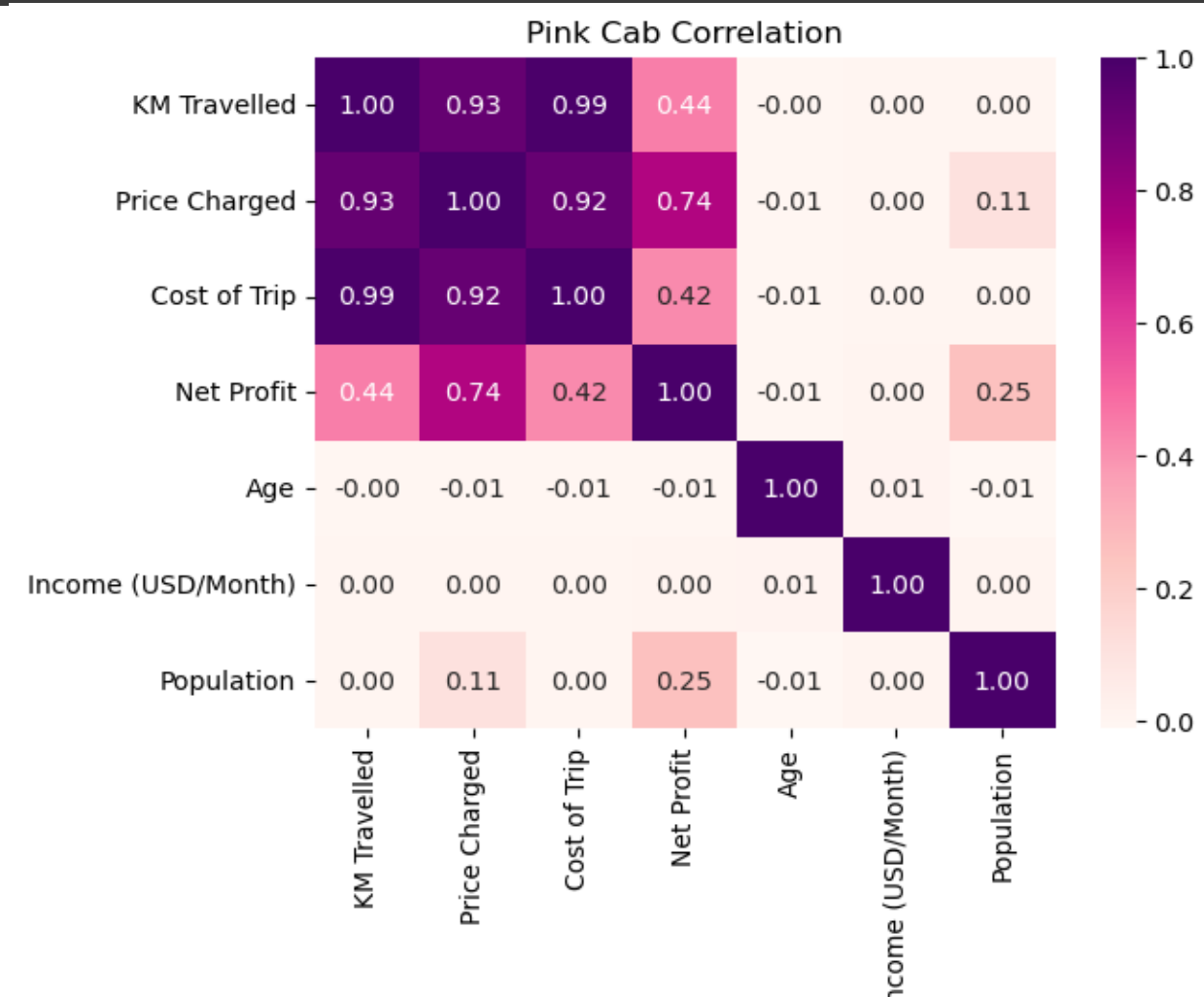
- Mean distance travelled of the Yellow Cab per drive is 22.57km
- Mean distance traveled of the Pink Cab per drive is 22.56 km.
- Median distance traveled of the Yellow Cab per drive is 22.44 km
- Median distance traveled of the Pink Cab per drive is 22.44 km
- Pink and Yellow cab has close distance travelled in average, but count of the measures is a lot more in Yellow Cav

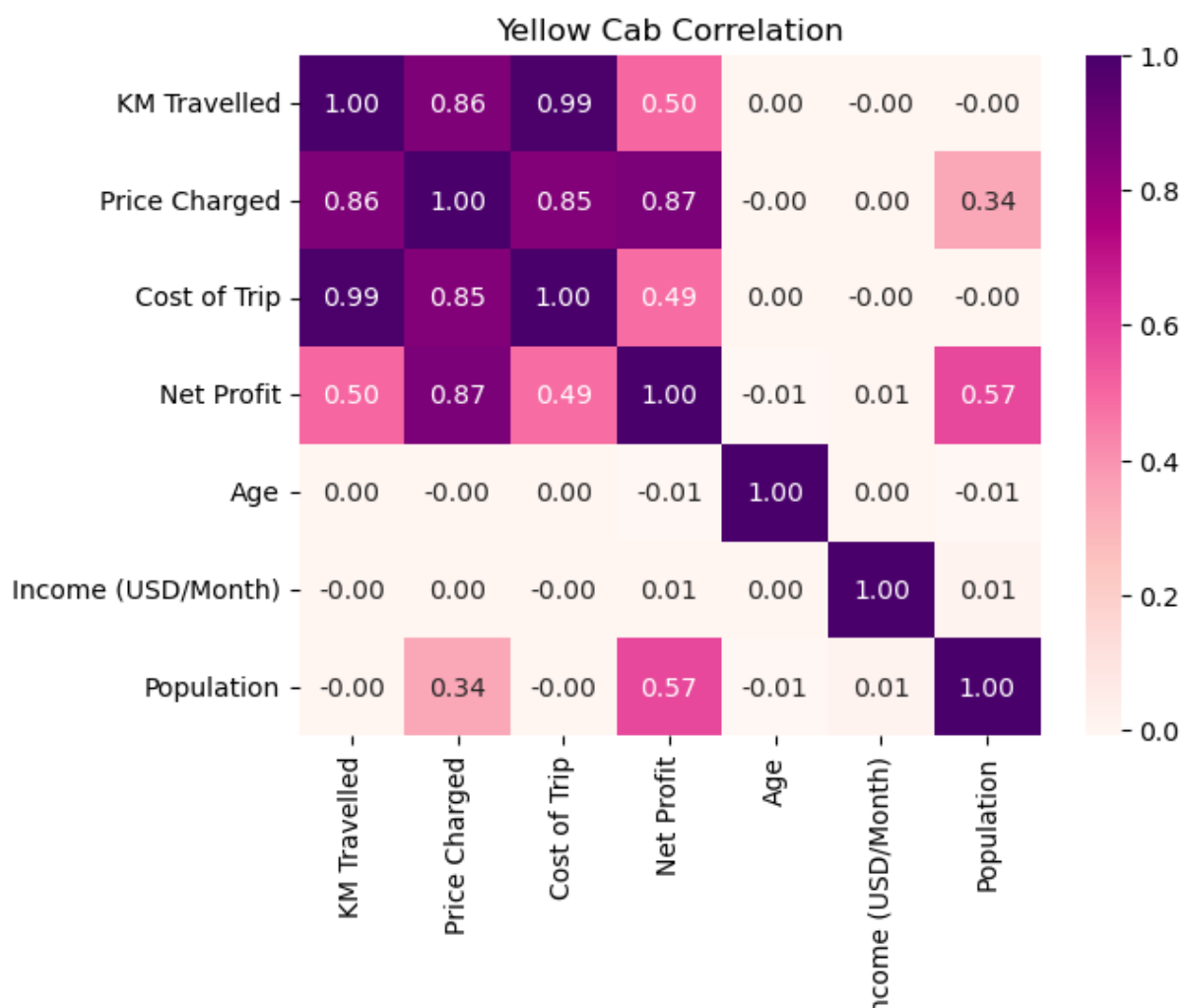
Bivariate Analysis

22.55

Data Glacier
Your Deep Learning Partner

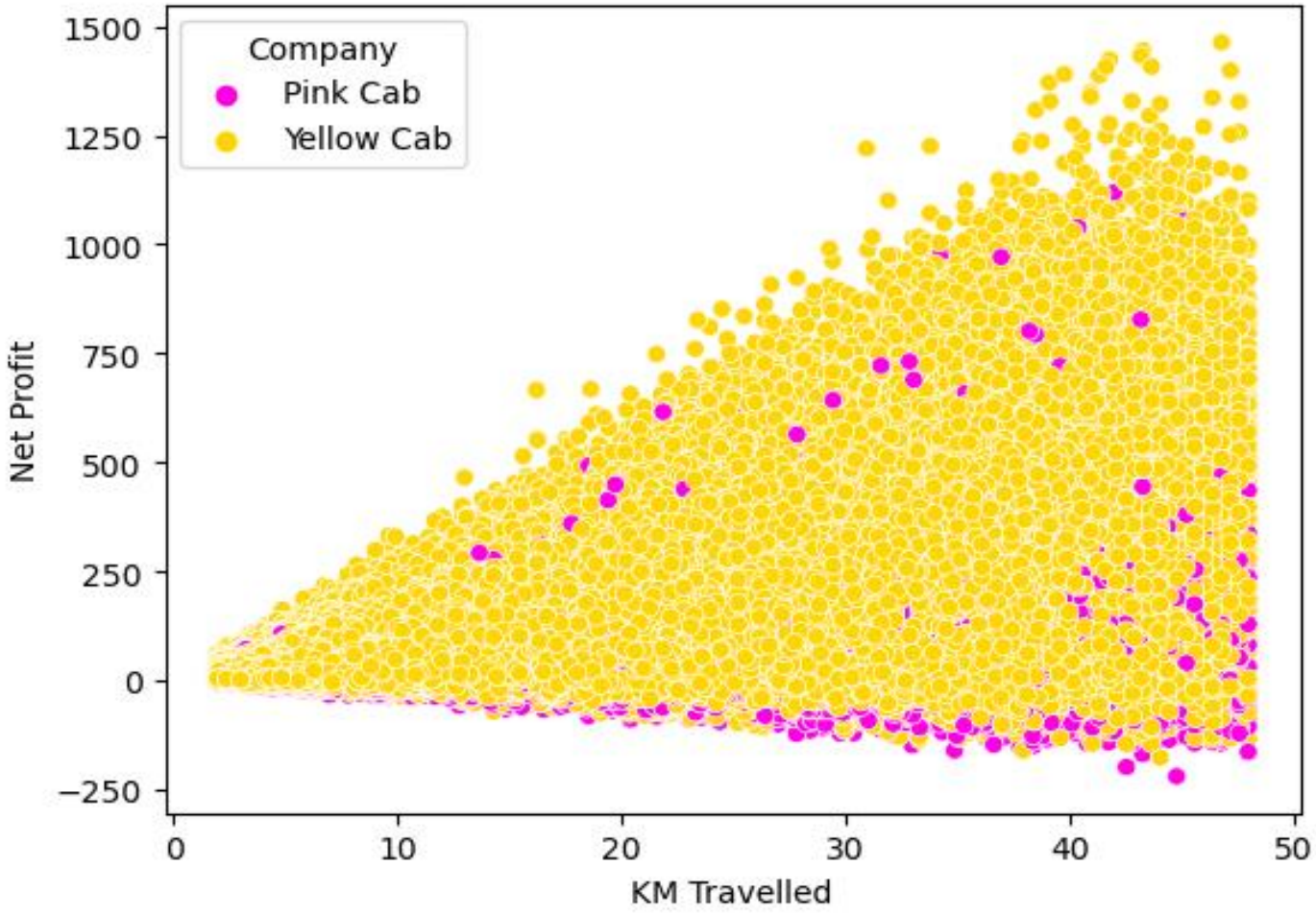
Correlation Matrices For Both of the Cab Firms.





Data Glacier
Your Deep Learning Partner

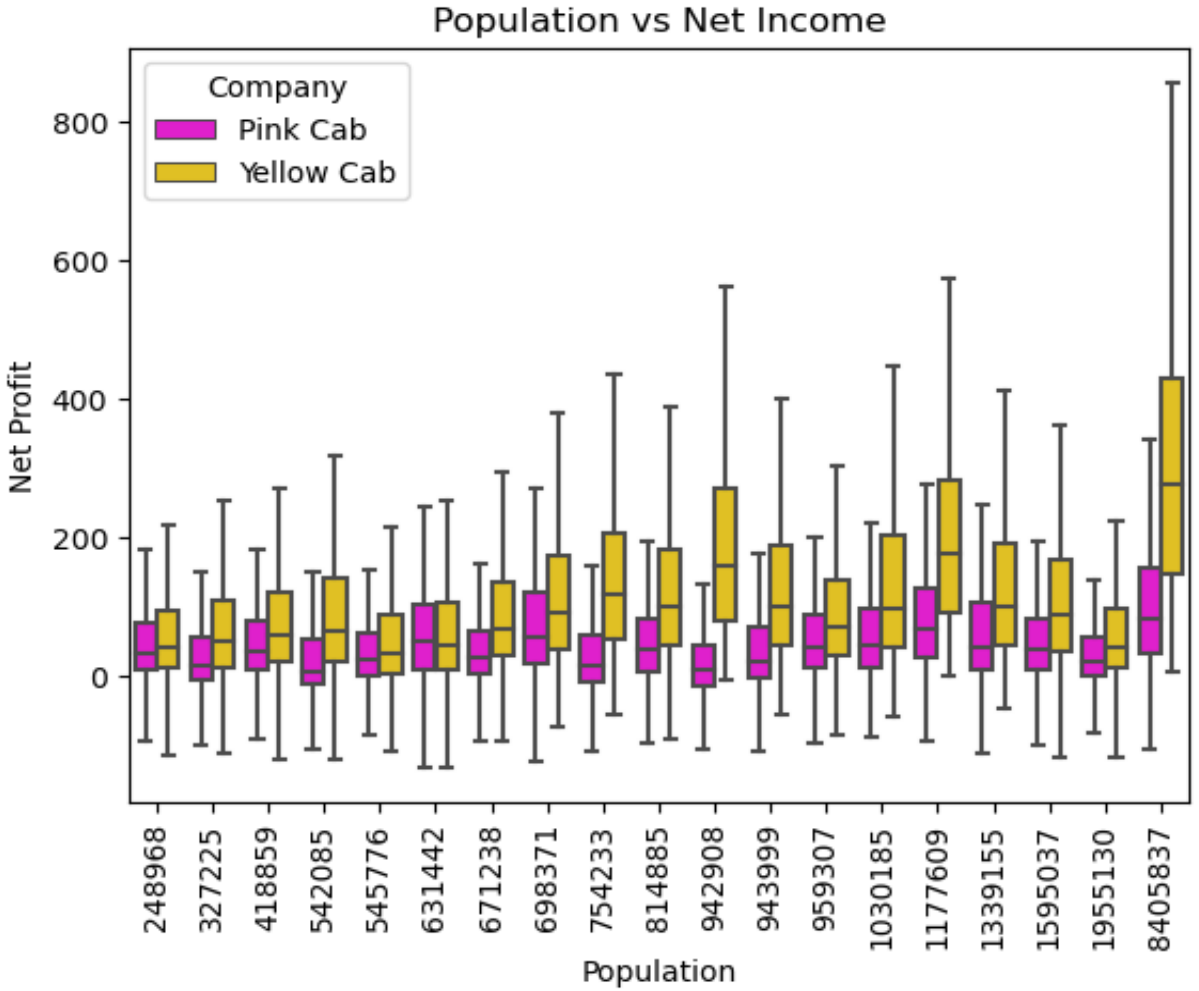
Distance Travelled and Net Profit Comparison



- Km traveled and net profit can be highly correlated.
- The correlation can be seen from this scatter graph
- Yellow cab's Net profit correlation with the KM travelled is more than the Pink Cab.

Data Glacier
Your Deep Learning Partner

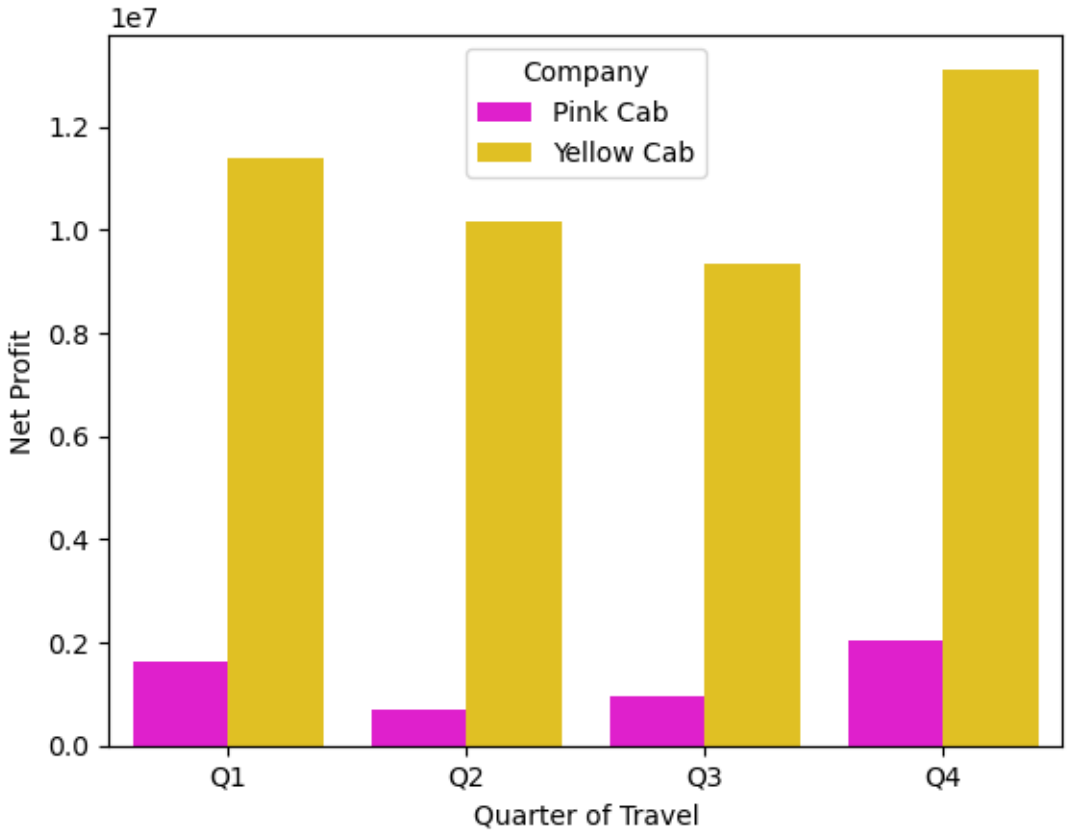
Net Profit's Dependency of Population



- Also Yellow Cab's Net profit correlation with the population is 0.57 , whereas Pink Cab's Net profit has a correlation with the population 0.25
- Boxplot shows Yellow Cab gains more net profit in the more populated cities
- Pink Cab's net profit does not increase as much as the population increases.

Data Glacier
Your Deep Learning Partner

Sum of Net Profit's in Different Quarters

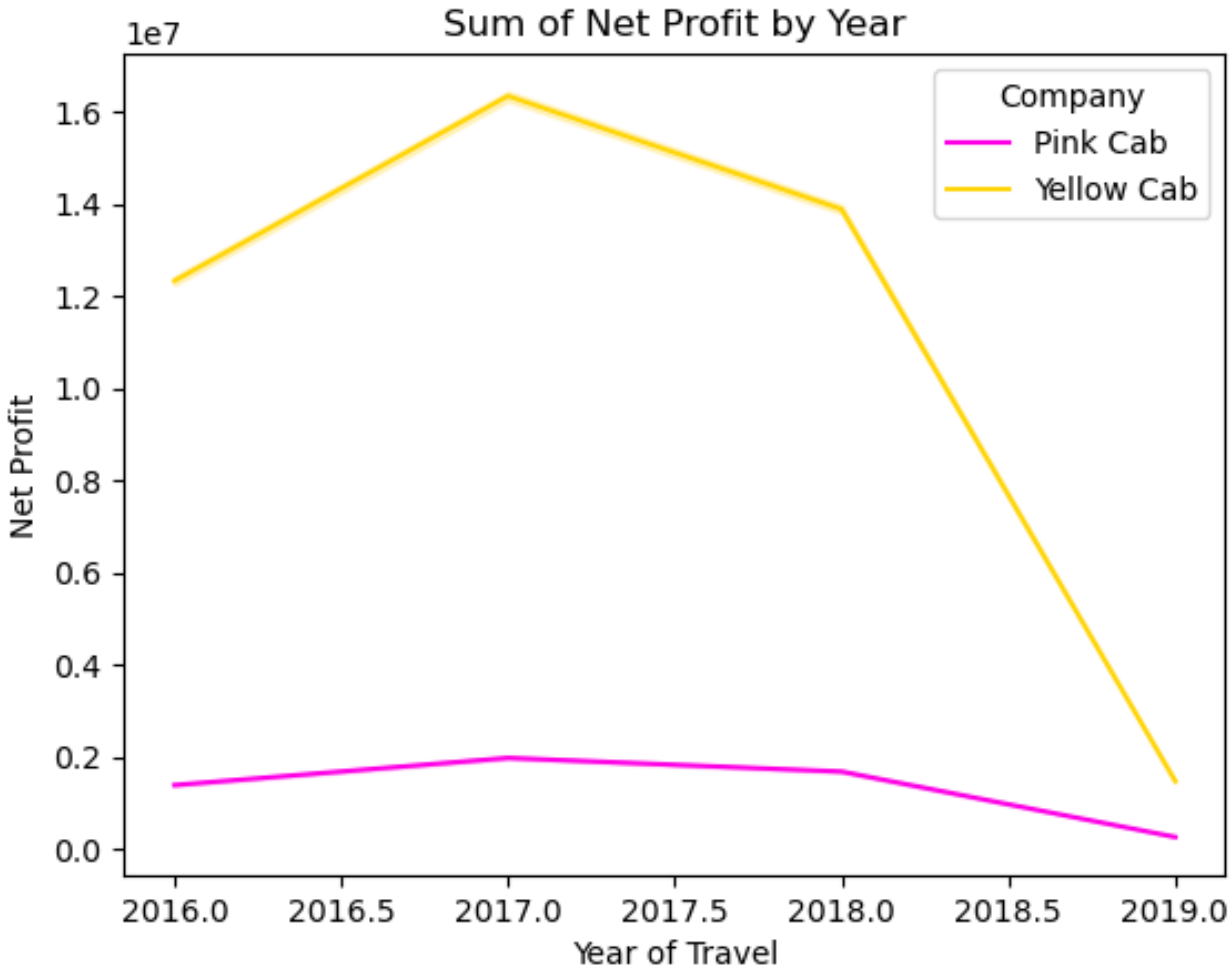


- Sum of the Net Profit vs quarters.
- Net Profit has seasonality.
- In winter Net Profits are rising for both Pink and Yellow Cab Company.

22.55

Data Glacier
Your Deep Learning Partner

Distance Travelled Distribution For the 2 Cab Firms.

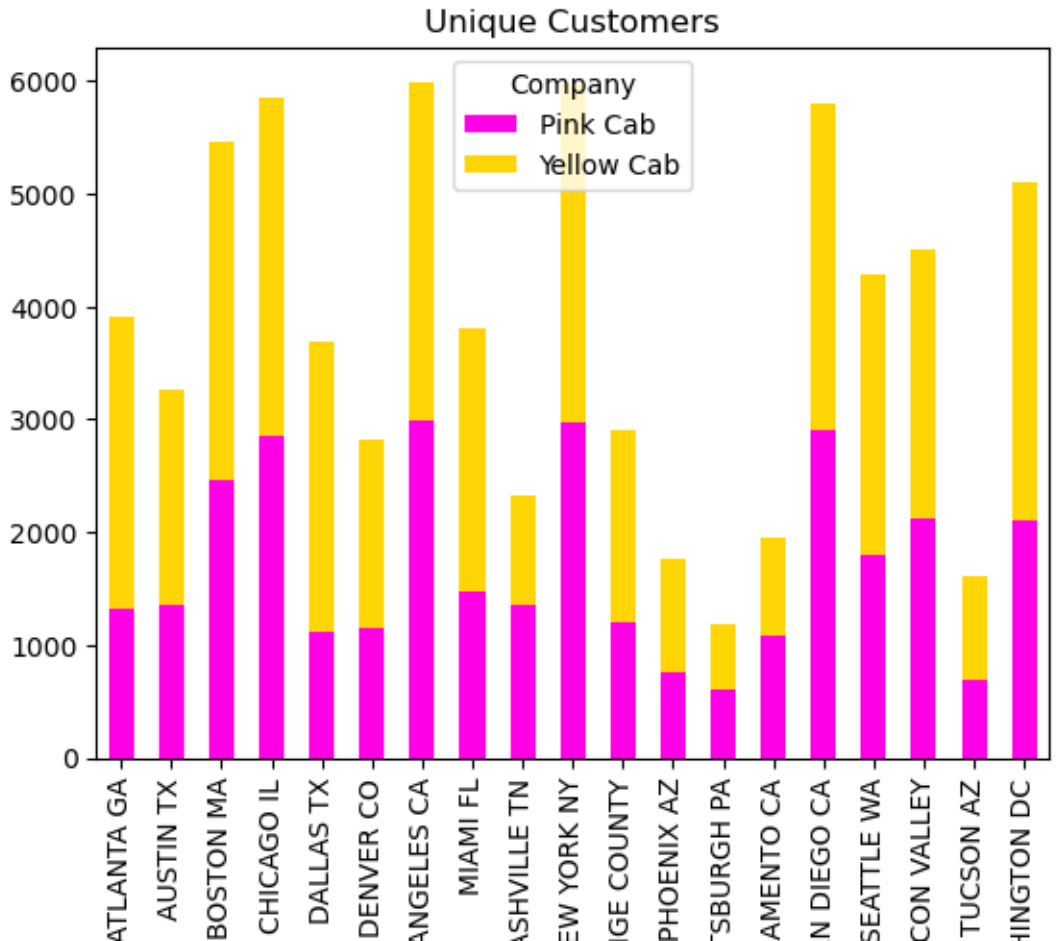


- Decrease in 2019 caused by lack of data.
- There is a decrease for the both cab firms in 2018
- Sum of Net profit is more than the pink cab firm in all years
- More Data is needed to see the trend in years.

22.55

Data Glacier
Your Deep Learning Partner

Unique Customer Reach For Both of the Cab Firms



- Both Cab firm's reach for the unique customers are close but still yellow leading in the reach for the customers
- This means that the unique users use yellow cab firm more frequently than the pink cab.

Hypothesis Testing

Data Glacier

Your Deep Learning Partner

Hypothesis Testing

- In hypothesis Testing it is proven that;
  - Average net profit of the company is higher than the pink cab
  - Net profit depends on the distance travelled in the cab
  - Net profit gained and the economic status of the customer is related
  - The average profits are different between quarters therefore data is seasonal

Data Glacier

Your Deep Learning Partner

Conclusion

- Yellow Cab is a smarter choice to invest because;
  - Average net profit of the company is higher than the pink cab
  - Has a wider customer reach, and frequency
  - Yellow cab has net more profit in more populated cities
  - The average profits are different between quarters therefore data is seasonal
  - Also for more km travelled yellow cab gains more profit

H. Ayberk
DEMIR

h.ayberk.demir.34@gmail.com

Thank You For Your Attention!