Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

21.05.2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

XYZ, a private firm in US, planning to make an investment in Cab industry since the industry has developed highly in recent years. Considering multiple key players in the market, it is following Go-to-Market(G2M) strategy since they want to understand the market before taking final decision.

Through actionable insights from exploratory data analysis, the executive team can identify the right company to make their investment.

Problem Statement

A private firm, XYZ, is planning for an investment in Cab industry. Before taking final decision, they want to understand the market. In other words, they want to follow the Go-to-Market(G2M) strategy to make this investment.

For this, the firm needs to identify the right company to make their investment.

Approach

1. First, I created the master data by merging the given datasets.

2. Then I explored the master data to understand the features of the dataset.

3. Then I dropped some of the features that might not help me to find the right company for the investment.

4. After that, I used the columns of Price Charged and Cost of Trip to calculate the profit for the master dataset.

Approach

5. Then the “groupby” operation was applied to find the description of the companies. It is obtained that the Price Charged, Cost of Trip and overall Profit gained are comparatively higher for the yellow cabs than the pink ones.

6. I also determined the sum and mean of the features for each cab company. It is obtained that the sum and mean of the features are higher for the yellow cabs than the pink ones due to both popularity and a large number of transactions.

So, it can be decided from the findings that the yellow cabs will be the right company to invest for the XYZ firm.

EDA

After creating the master dataset, I performed a few basic Exploratory Data Analysis. First, I applied the describe() method which returns the description of the data in the DataFrame. Then I displayed the first and last five rows of the dataset. I applied the shape method to print out the shape of the DataFrame, which is (440098 X 12) indicating the rows and columns respectively. To know the layout of each array in the dataset, I applied dtype method.

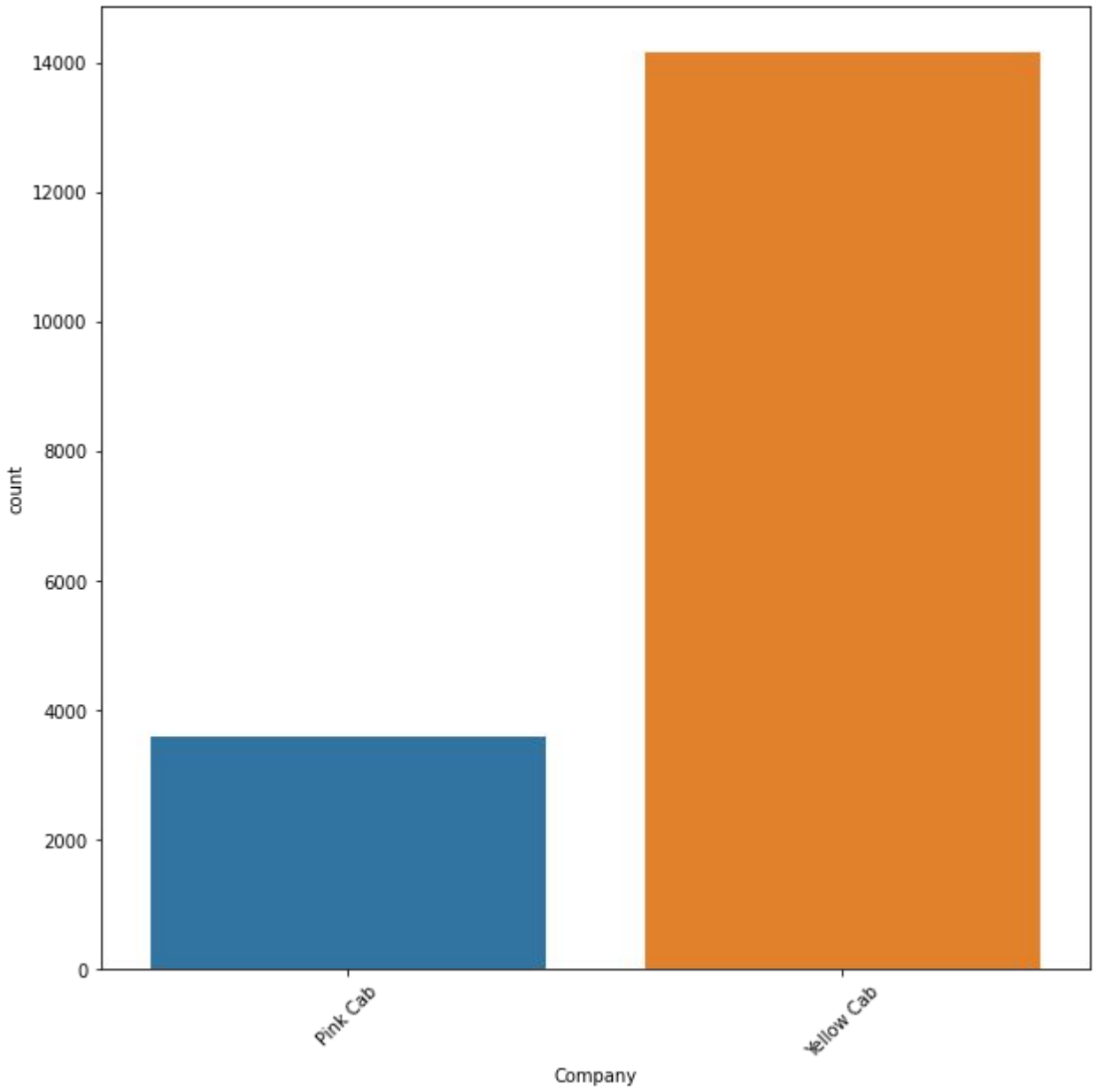
EDA

After applying some basic exploratory analysis, I found out the number of duplicates in DataFrame. It was obtained that, there is no duplicate row found in the dataset. As a result, the shape of the dataset remained the same after dropping the duplicate rows.

After identifying the duplicate rows, I found out the missing and null values for each features. It was found that there were some missing or null values for the Date of Travel, Company, City, KM Travelled etc, in the dataset. Since, the rows with missing or null values are of no or less importance to find insights, those rows were dropped. As a result, the shape of the dataset became smaller, which is (17763 X 12).

EDA

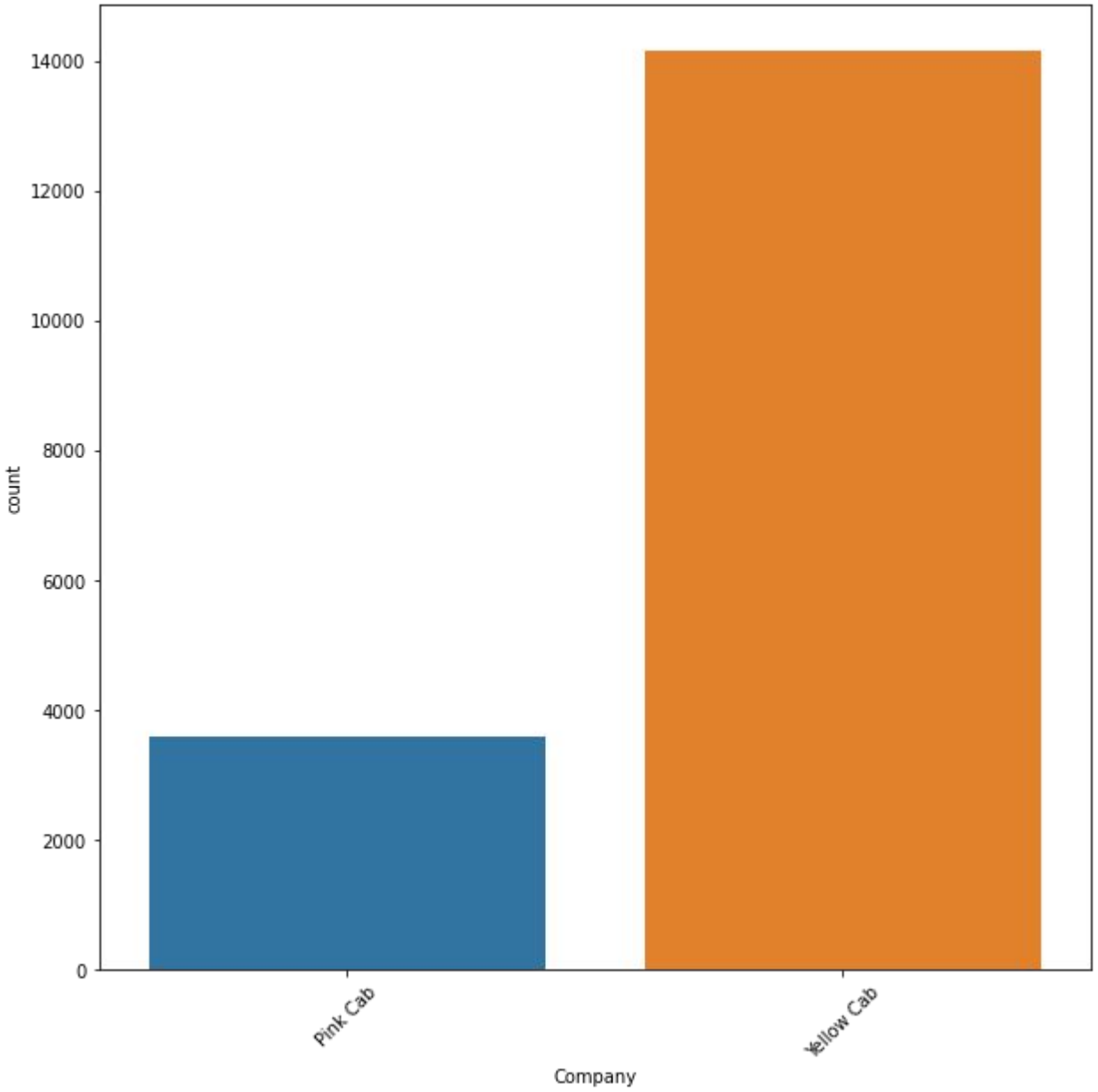
After the exploring the basic overview of the dataset, the number of customers of each cab company was counted using a bar chart.



Data Glacier
Your Deep Learning Partner

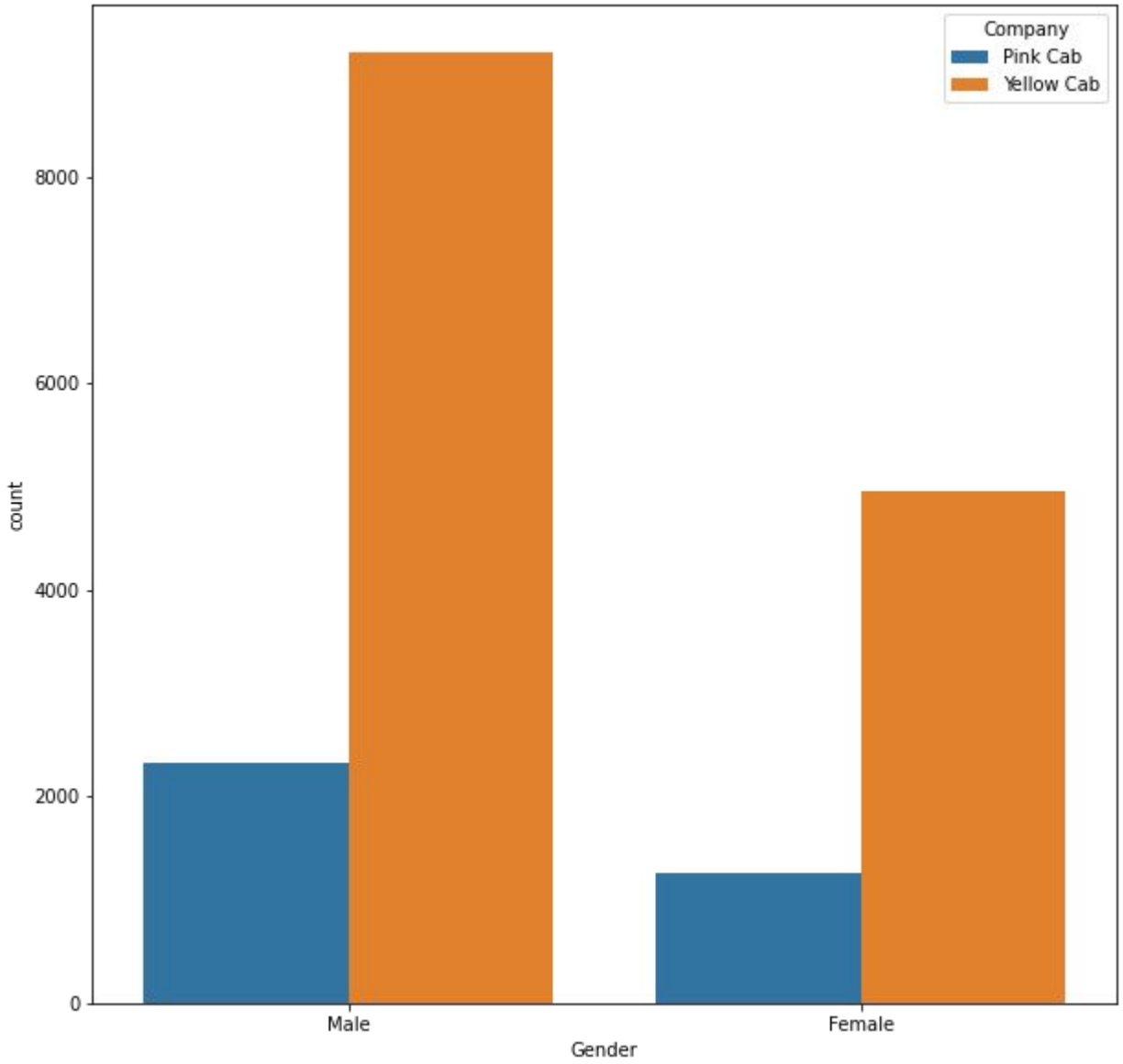
EDA

Then the number of transactions of each cab company was counted using a bar chart, which was found almost similar as the previous bar chart. The chart shows that the number of transactions occurred for the yellow cab is much higher than the pink ones.

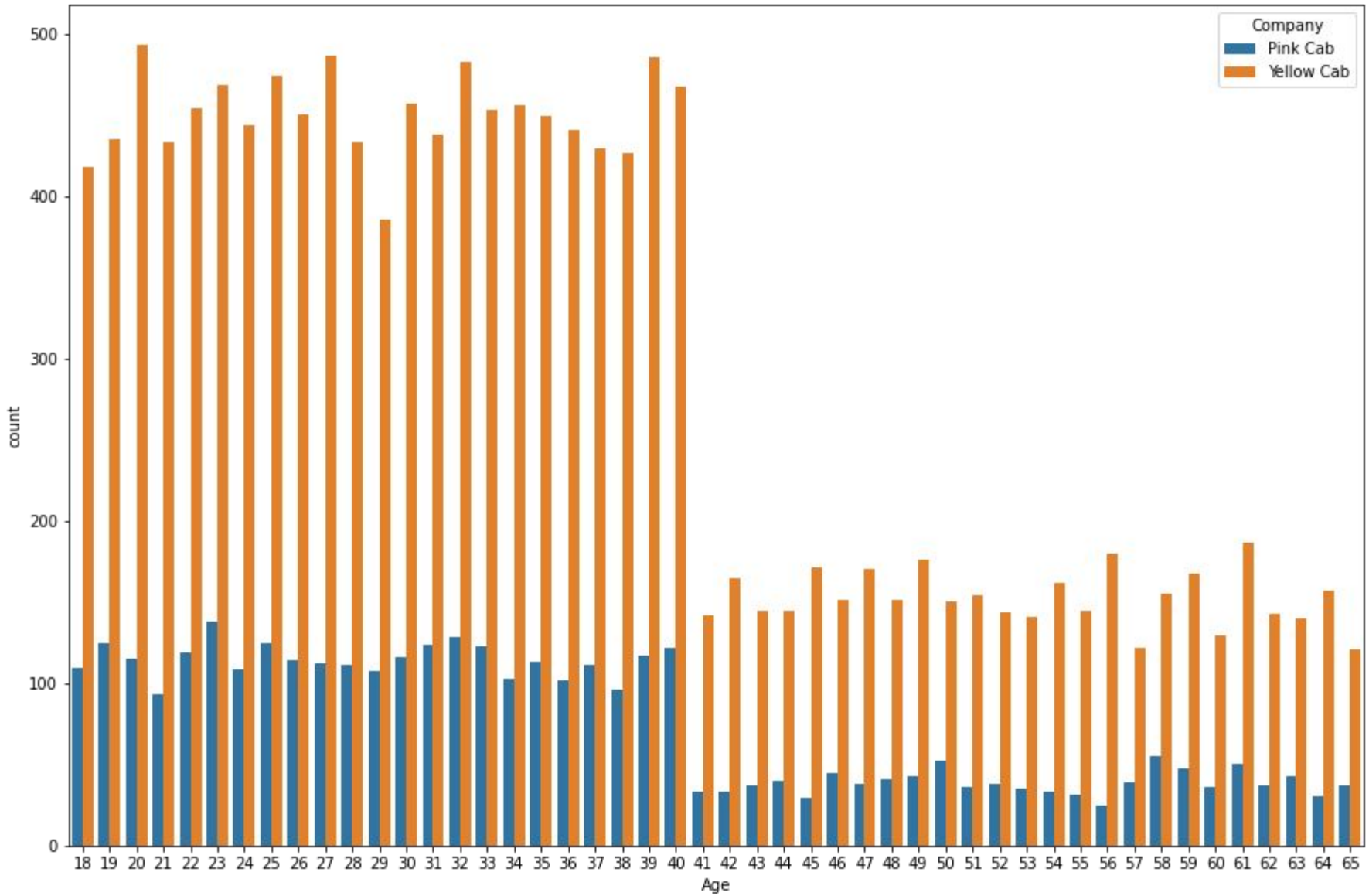


EDA

Moreover, the yellow cabs were popular for both male and female customers.

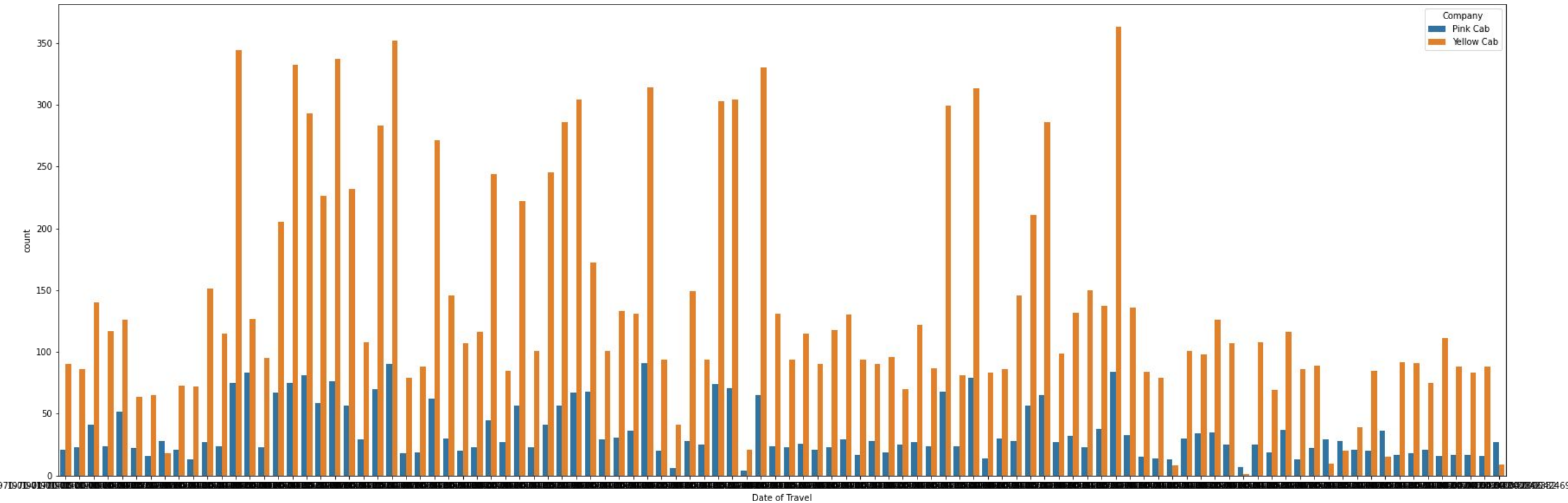


Data Glacier
Your Deep Learning Partner


Company
Pink Cab
Yellow Cab
count
Age
0
100
200
300
400
500
18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65

EDA

It was found that yellow cabs were more popular and commonly used by all age groups of customers.

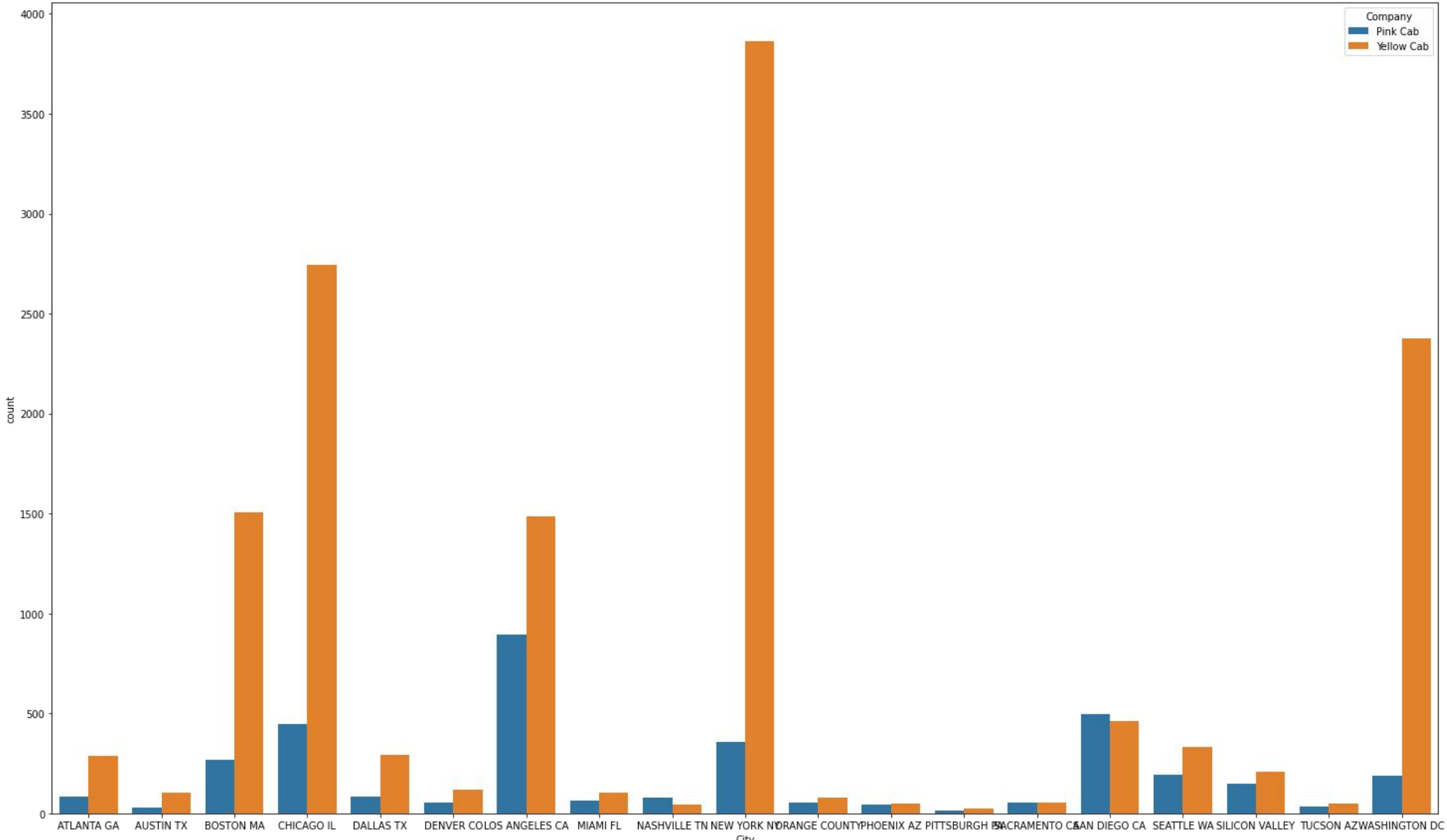

Company
Pink Cab
Yellow Cab
count
0
50
100
150
200
250
300
350
Date of Travel

EDA

It was also found that yellow cabs were more popular than the pink ones during most of the travel dates.

EDA

The popularity between the cab companies were also obtained for every city in US.


Company
Pink Cab
Yellow Cab
count
4000
3500
3000
2500
2000
1500
1000
500
0
ATLANTA GA
AUSTIN TX
BOSTON MA
CHICAGO IL
DALLAS TX
DENVER CO
LOS ANGELES CA
MIAMI FL
NASHVILLE TN
NEW YORK NY
ORANGE COUNTY
PHOENIX AZ
PITTSBURGH PA
SACRAMENTO CA
SAN DIEGO CA
SEATTLE WA
SILICON VALLEY
TUCSON AZ
WASHINGTON DC
City

EDA

It was found out yellow cab is more popular than pink cab in most of the cities in US, though pink cab is insignificantly popular in San Diego and Nashville.

EDA Summary

In conclusion, it was found that the firm will gain comparatively higher profit if it makes the investment in the yellow cab company.

Recommendation

It is recommended to the XYZ firm to invest in the yellow cab company since the EDA proved that yellow cabs are more popular and more used by the customers than the pink ones. It is also obtained that the yellow cab company makes more profit than the other one.

Thank You