



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Cab Company

7/17/2022

Agenda

Executive Summary
Problem Statement
Approach
EDA
EDA Summary
Recommendations

Executive Summary

- XYZ is a private firm in the US looking to capitalize on the rising growth of the Cab Industry and invest in a Cab company
- Given a data set with 2 different cab companies, and analyze the data

Problem Statement

- Most of the data will surround the Hypothesis Statements, which will be shown later
- However, the ultimate goal for this report is to analyze and determine which Cab Company (Pink or Yellow) is the best company to invest in, whilst also providing the necessary information on why
- It is important to notice that ultimately, each cab company will be better than the other in something. The goal of this data report is to analyze these nuances and come up with a reasonable conclusion.

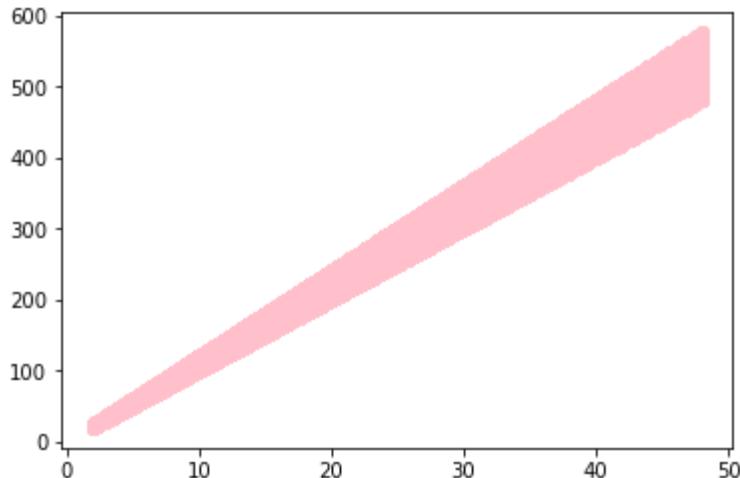
Approach

- We will be using a process known as Exploratory Data Analysis to analyze the two cabs
- For each hypothesis (in this case it is the factors we look into), we will look at each company, as well as the combined total as well, to come up with our conclusions
- Note: There has been a new column added, which represents the difference between the amount the customer pays for the trip, and the amount the cab owner/company pays for the trip. If the values are negative, then that means they are ultimately losing money for that trip.

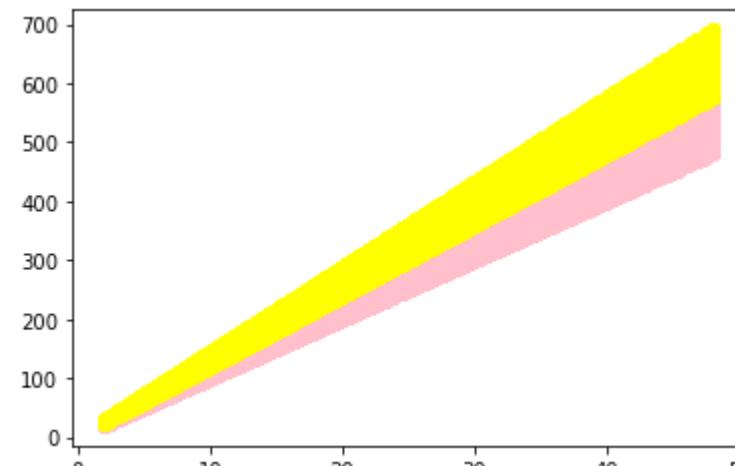
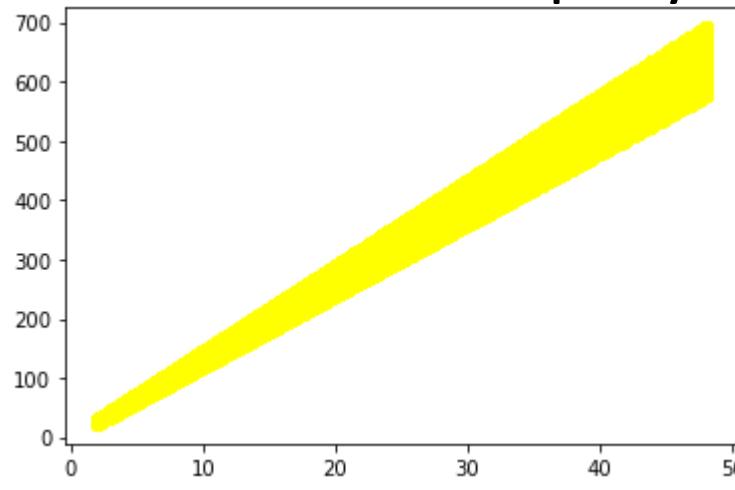
EDA

Hypothesis #1: Is there a linear correlation between the total kilometers traveled (x-axis) and the total cost of the trip (y-axis)? What about them separately based off of the two cab companies?

- Pink Cab Company



- Yellow Cab Company

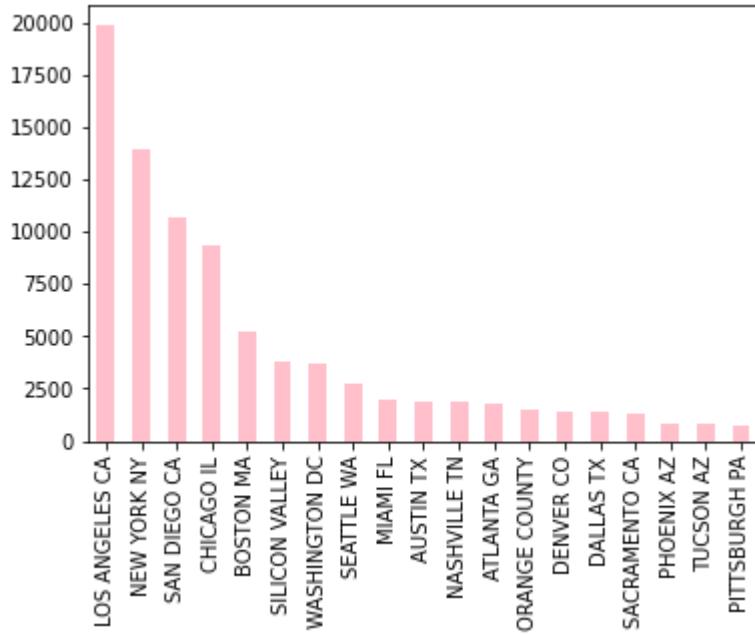


Conclusion

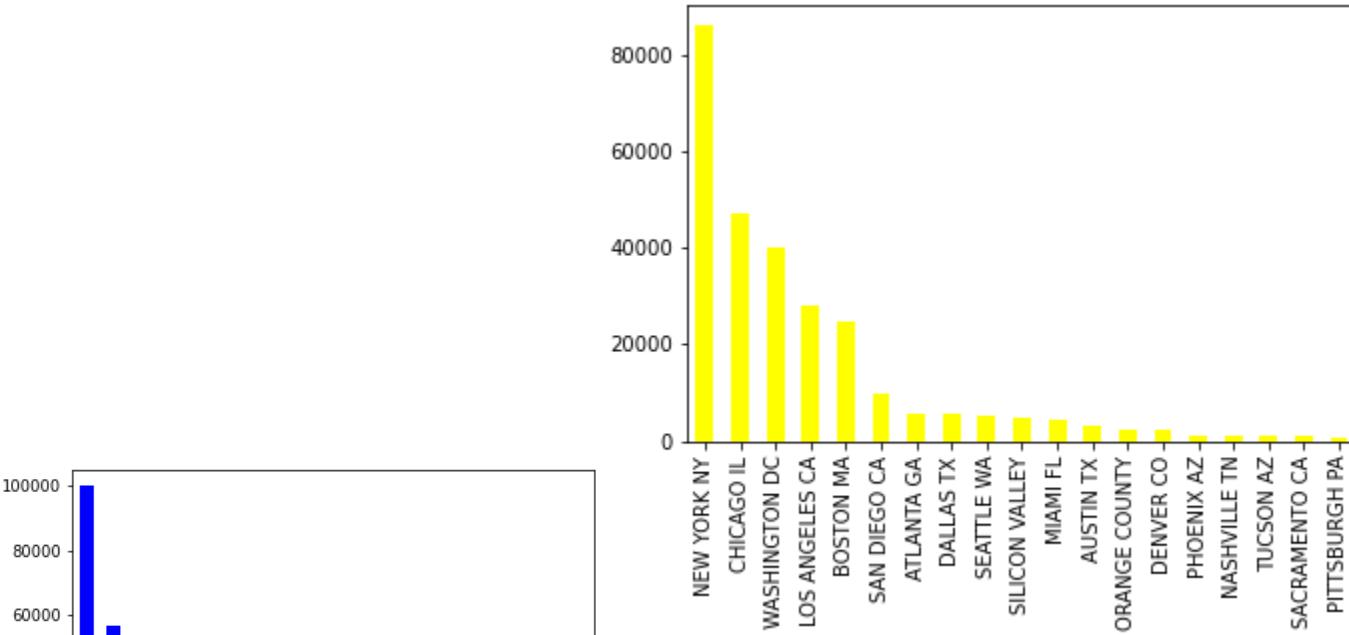
- There seems to be a general linear correlation with the number of kilometers for both cabs
- For the Pink Cab company, especially by the longer distances traveled, is usually less expensive than the Yellow Cab company
- However, one thing to note is that there appears to be more of a variance of the total cost the longer the distance is traveled.

Hypothesis #2: What is the frequency of each cab in each location?

- Pink Cab Company



- Yellow Cab Company

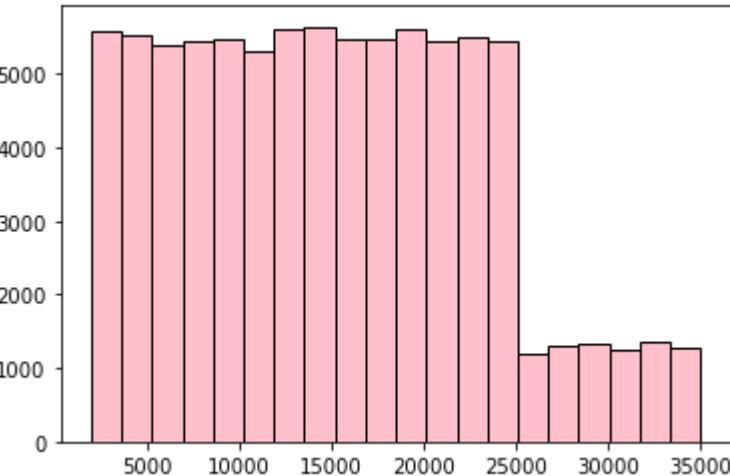


Conclusion

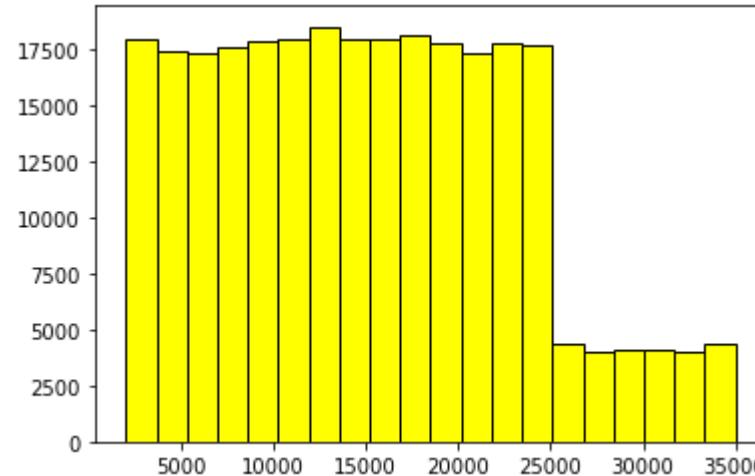
- A majority of the Pink Cab Company resides in Los Angeles, California (followed by New York City, New York and San Diego, California)
- A majority of the Yellow Cab Company resides in New York City, New York (followed by New York City, New York and Chicago, Illinois)
- A majority of the Total Cabs reside in New York City, New York (followed by Chicago, Illinois and Los Angeles, California)

Hypothesis #3: What is the income distribution between the users of both cabs?

- Pink Cab



- Yellow Cab



Conclusion

- Both Customers of both companies have relatively the same income
- As a result, there does not seem to be an inherent bias towards either company in that regard, regardless of expense

Hypothesis #4: Which company is making more of a profit?

- Pink Cab

Approximately \$5,307,328

- Yellow Cab

Approximately \$44,020,373

Conclusion

- The Yellow Cab makes more of a profit than the Pink Cab
- However, do note that according to the initial look at the data, there are more data points towards the Yellow Cab, meaning that there could be a bias

Recommendations

The Yellow Cab seems to be the best option for XYZ to invest in

Reasons

- Despite being a lot more expensive, especially as the trip distance increases, the customer income distribution indicates no real bias or tendency, meaning that almost the same base income consumers would be willing to take either cab
 - As an aside, Los Angeles, the most popular region for the Pink Cab, is 502 miles squared, compared to New York City's 302 miles squared. This means that the Pink Cab will actually be covering a lot more distance on average, making them more often to be more expensive than the Yellow Cab company
- In general, the Yellow Cab Company makes a large profit compared to the Pink Cab company.

Thank You