



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

Virtual Internship

21 July 2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

Conclusion

Background/Overview of Case Study

Background:

- XYZ is a private firm based in the US. With an increase of the Cab Industry and key players in the market, XYZ is planning to invest in the Cab Industry.
- To make the best financial decision, XYZ uses a Go-to-Market strategy, a way to analyze the market.

Project Deliverable:

- Create an analysis between two cab companies to help XYZ choose the right company to invest in.

Analysis Divided by Two Parts:

- Income per company based on a specific variable.
- Count of rides per company based on a specific variable.

Approach

Given: 4 datasets (cab_data.csv, customer_id.csv, transaction_id.csv, and city.csv)

Method of Approach:

1. Check for outliers and missing data from all data sets.
2. Draw connections between data sets based on shared variables. Ex. Both cab_data.csv and transaction_id.csv had a transaction_id column, allowing for the combination of two data sets.
3. Create a master data frame using SQL.
4. Perform analysis on the data given using Python modules of Seaborn and Matplotlib.

Assumptions:

1. Data collected in the city.csv file is not specific to the two cab companies but can include others as well.
2. Income per cab ride is reliant on strictly the price charged and total cost of trip and does not incorporate other variables.

Final Dataset: A combination of all 4 data sets to create a master dataset.

Total Data Points: 359393

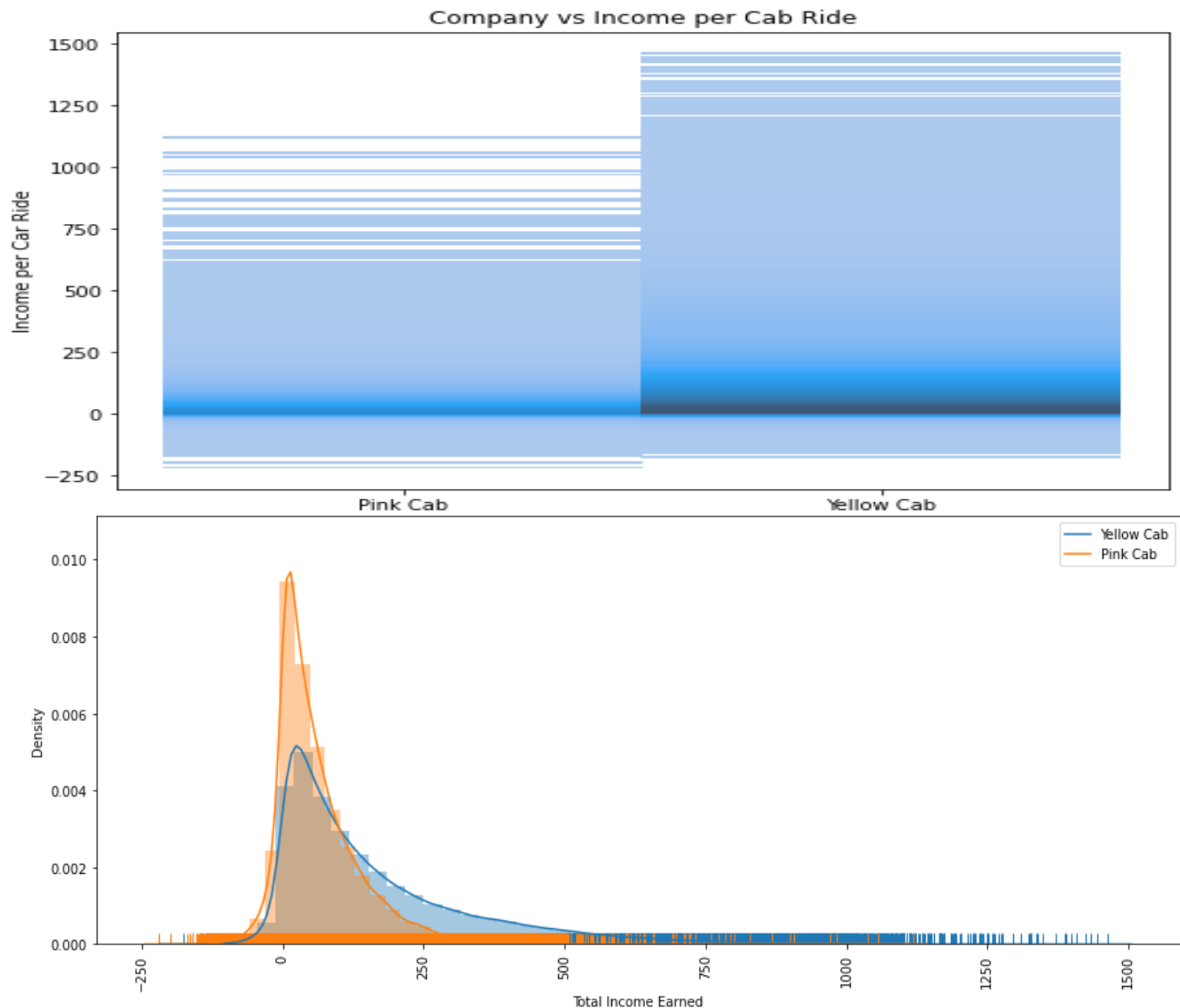
Company vs Income

The areas with a darker shade on the top graph represent where more rides are placed relative to the income per car ride.

Even though most of the rides for each company are centered around zero to a hundred dollars, the Yellow Cab has a wider span of income per car rides, the highest reaching around \$1,450 dollars.

This shows that the Yellow Cab has the potential of making higher amounts of money per cab ride than the Pink Cab, who only reaches around \$1,150 for the maximum income.

The graph on the bottom is just another look at the density of the income earned per car ride. It shows that the Yellow Cab has a higher density at all points but has a widespread reach.



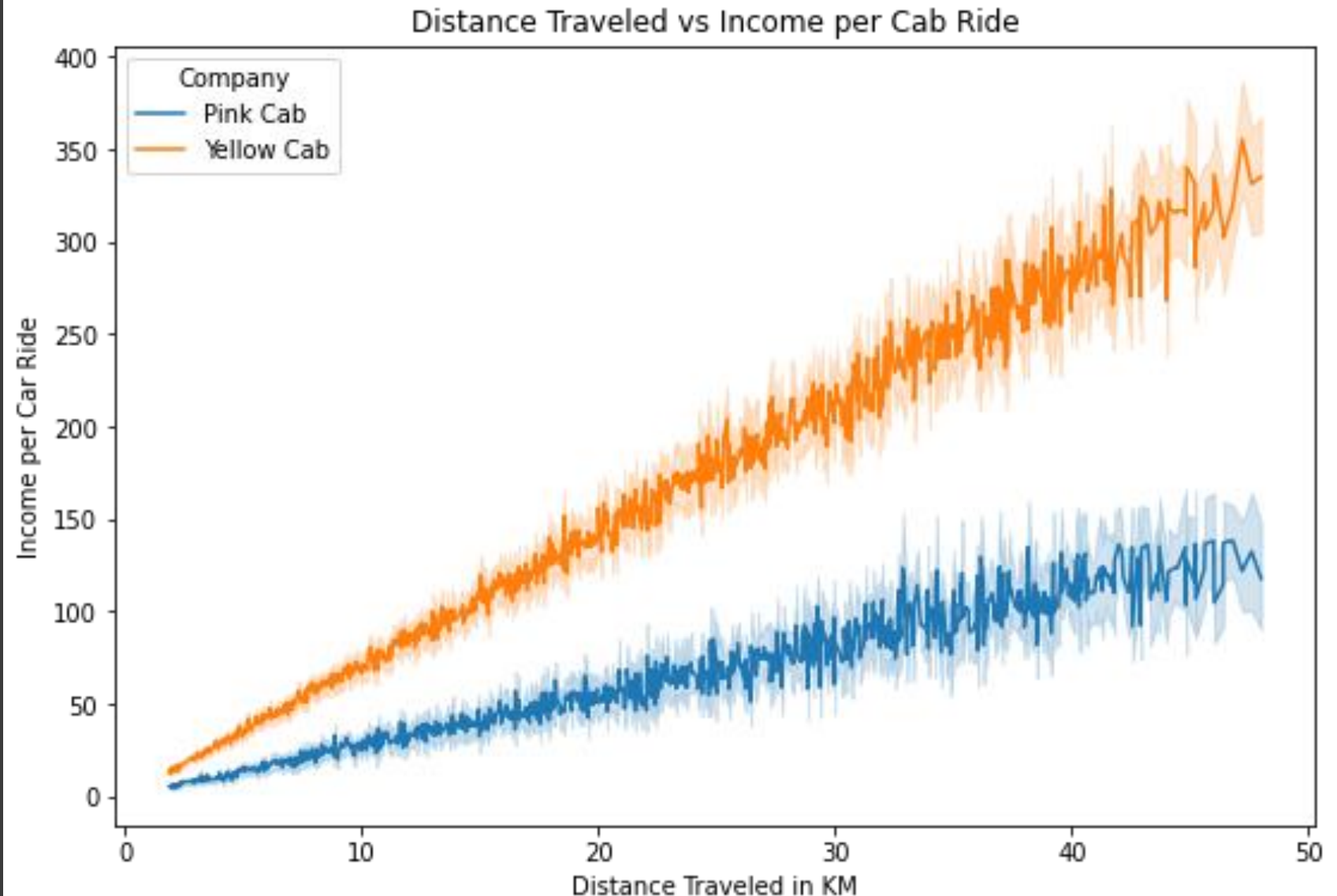
Distance vs Income

The graph to the right depicts how as the distance traveled per cab increases the Yellow Cab has the greater rate of change for income per car ride.

The Yellow Cab earns more as the distance increases than the Pink Cab.

At greater distances, the difference in income is wider than in smaller distances.

This shows that the Yellow Cab company gets more income per amount spent in costs such as gas and maintenance of vehicles.



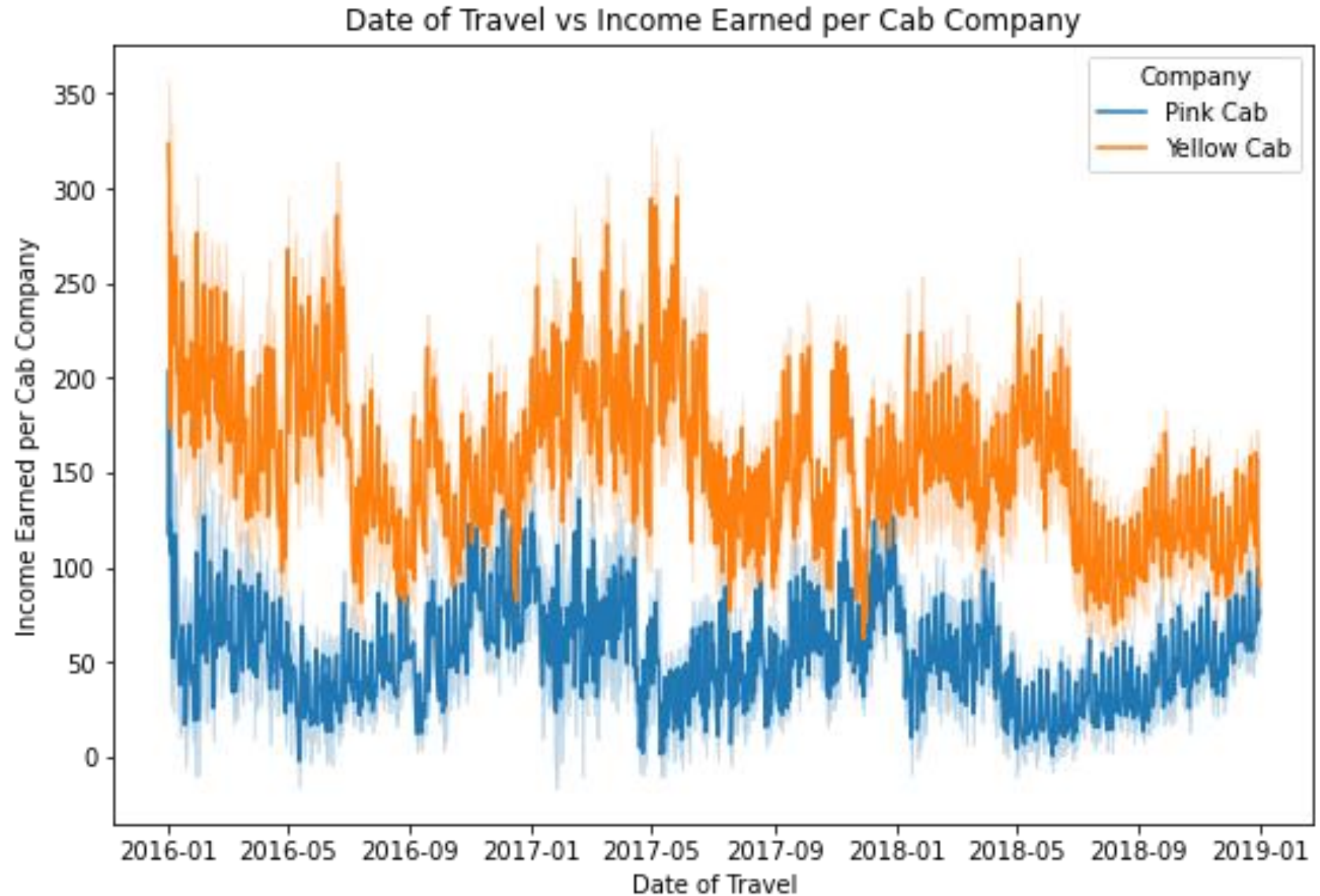
Date vs Income

In this graph, the Yellow Cab dominates the income earned per ride per timeframe.

It also looks as if when the Yellow Cab is successful, the Pink Cab has a lower income earned.

An interesting trend to note, however, is that both companies have been growing closer together from 2018 onwards, which may not lead to significant income increases.

Therefore, the Yellow Cab would have a greater income compared to the Pink Cab.



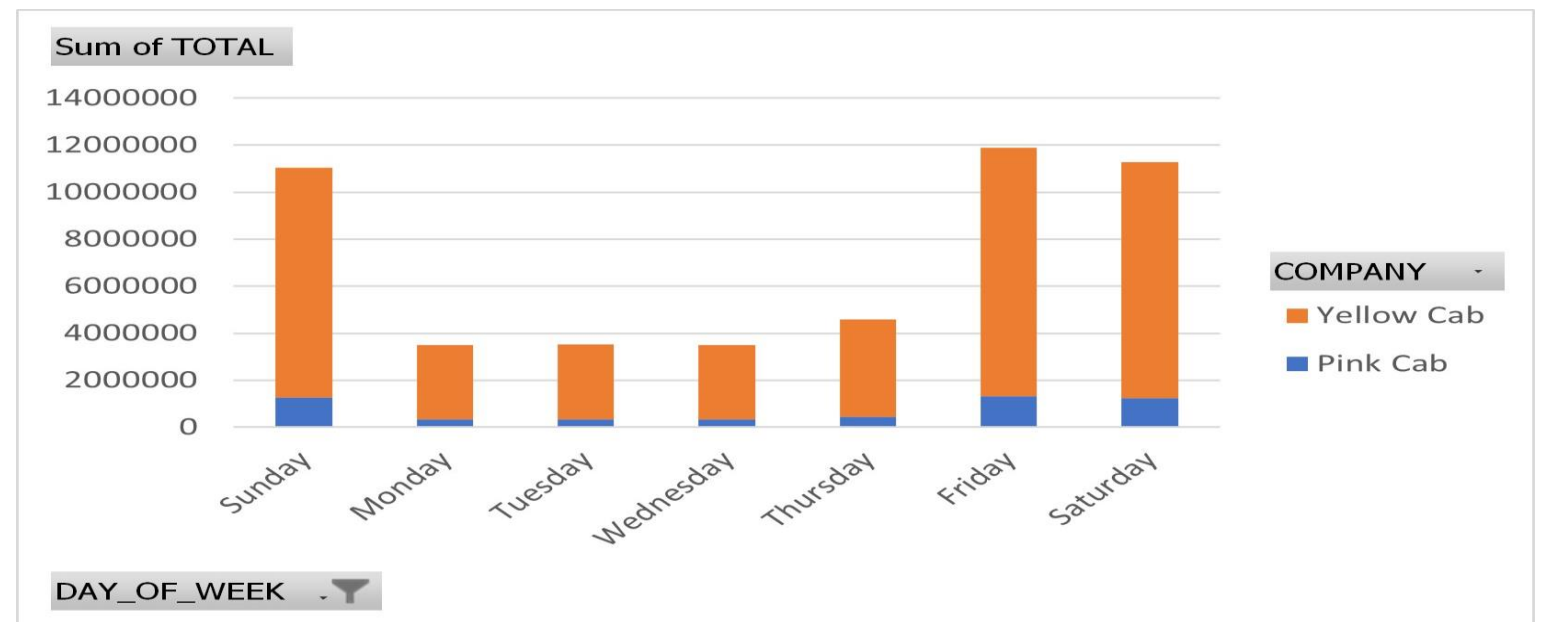
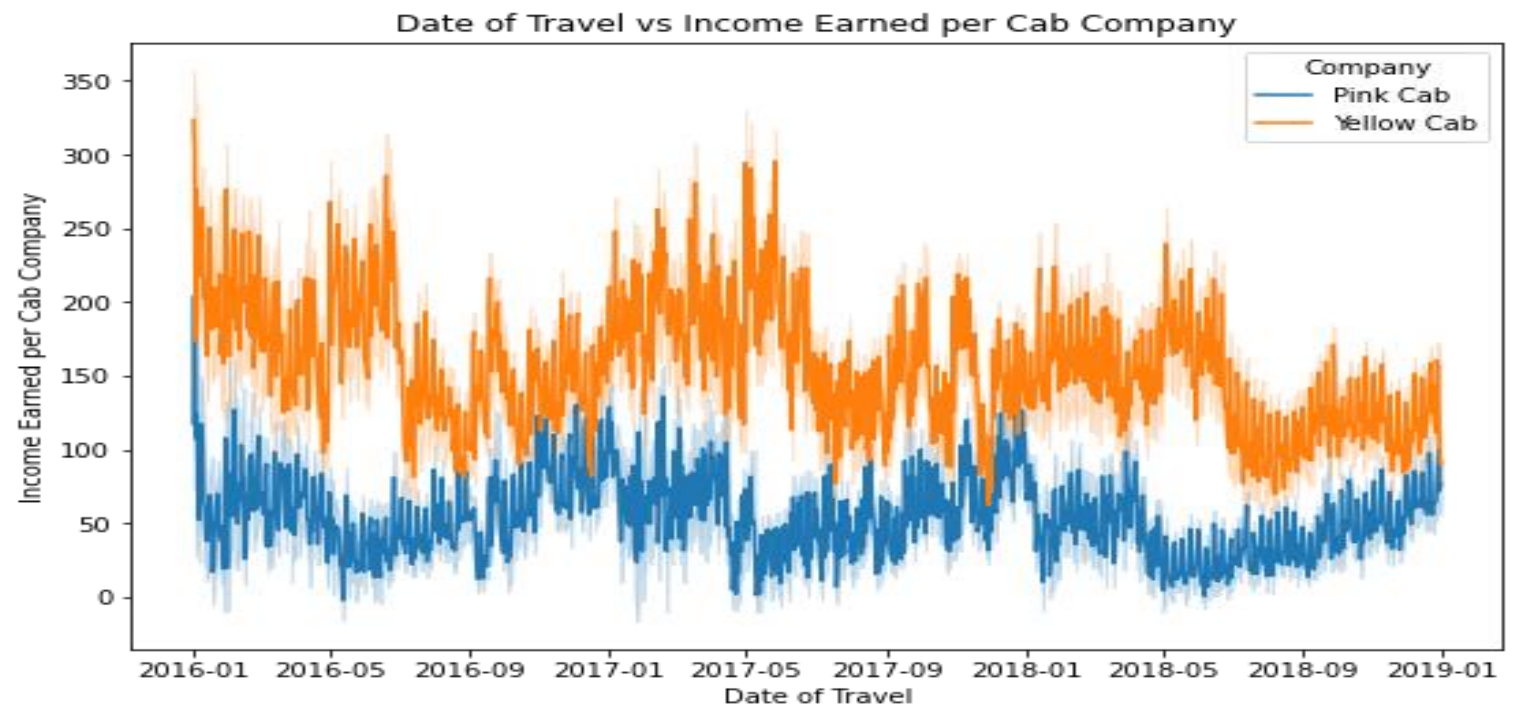
Date vs Income

In the graph on the top, the Yellow Cab dominates the income earned per ride per timeframe.

It also looks as if when the Yellow Cab is successful, the Pink Cab has a lower income earned.

An interesting trend to note, however, is that both companies have been growing closer together from 2018 onwards, which may not lead to significant income increases.

More specifically, the graph below shows that Friday, Saturday, and Sunday have the most demand for taxis.

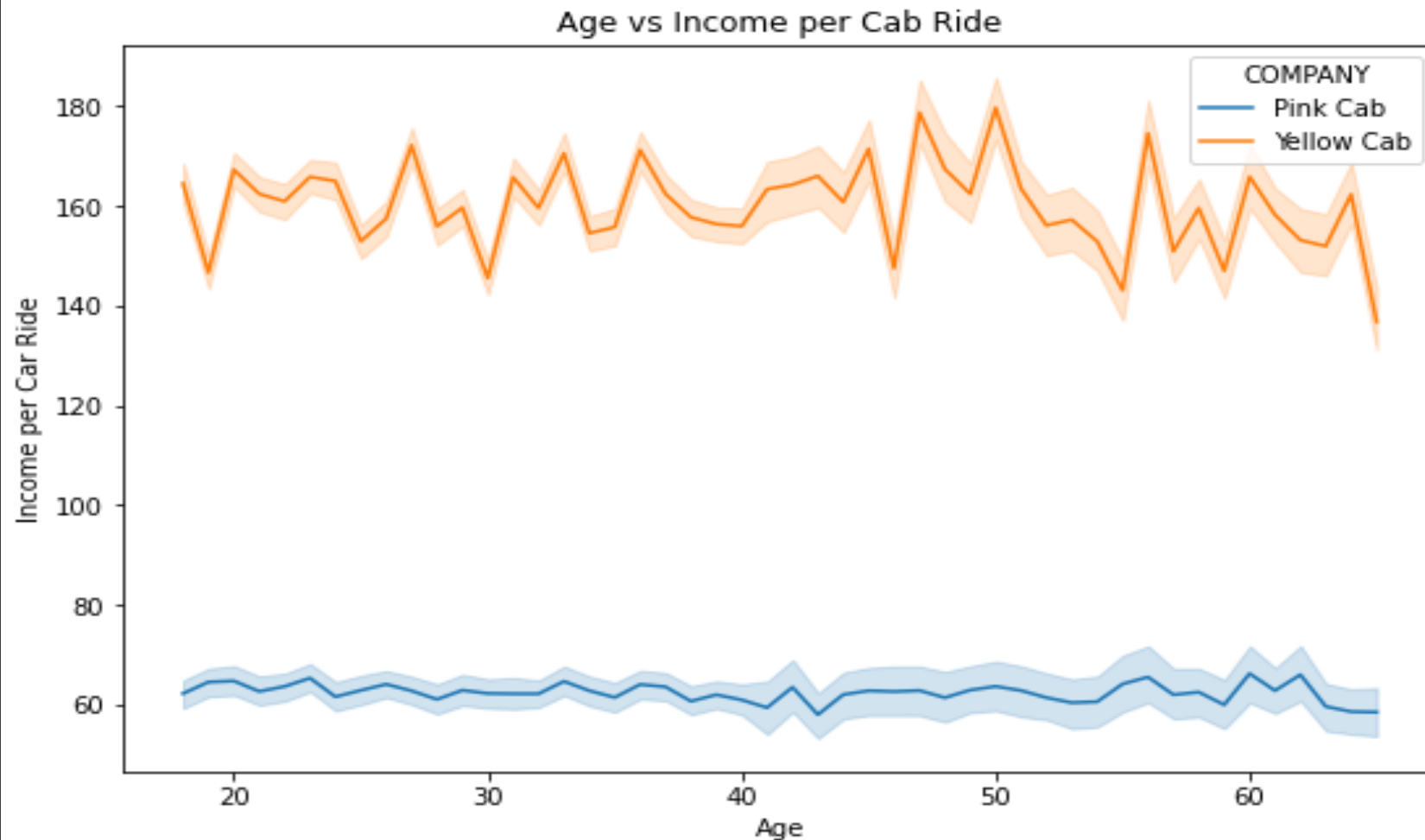


Age vs. Income

This graph represents that overall, no matter what age someone is they usually choose to take the Yellow Cab.

Even so, the Pink Cab company is more consistent, as the line is relatively smoother compared to the Yellow Cab company.

However, the Yellow Cab is consistently earning more income per car ride.



Income vs Customer Income

The graphs show the consumer income vs the profit of the cab ride. The Yellow Cab consistently earns more per cab ride, no matter the consumer's income.

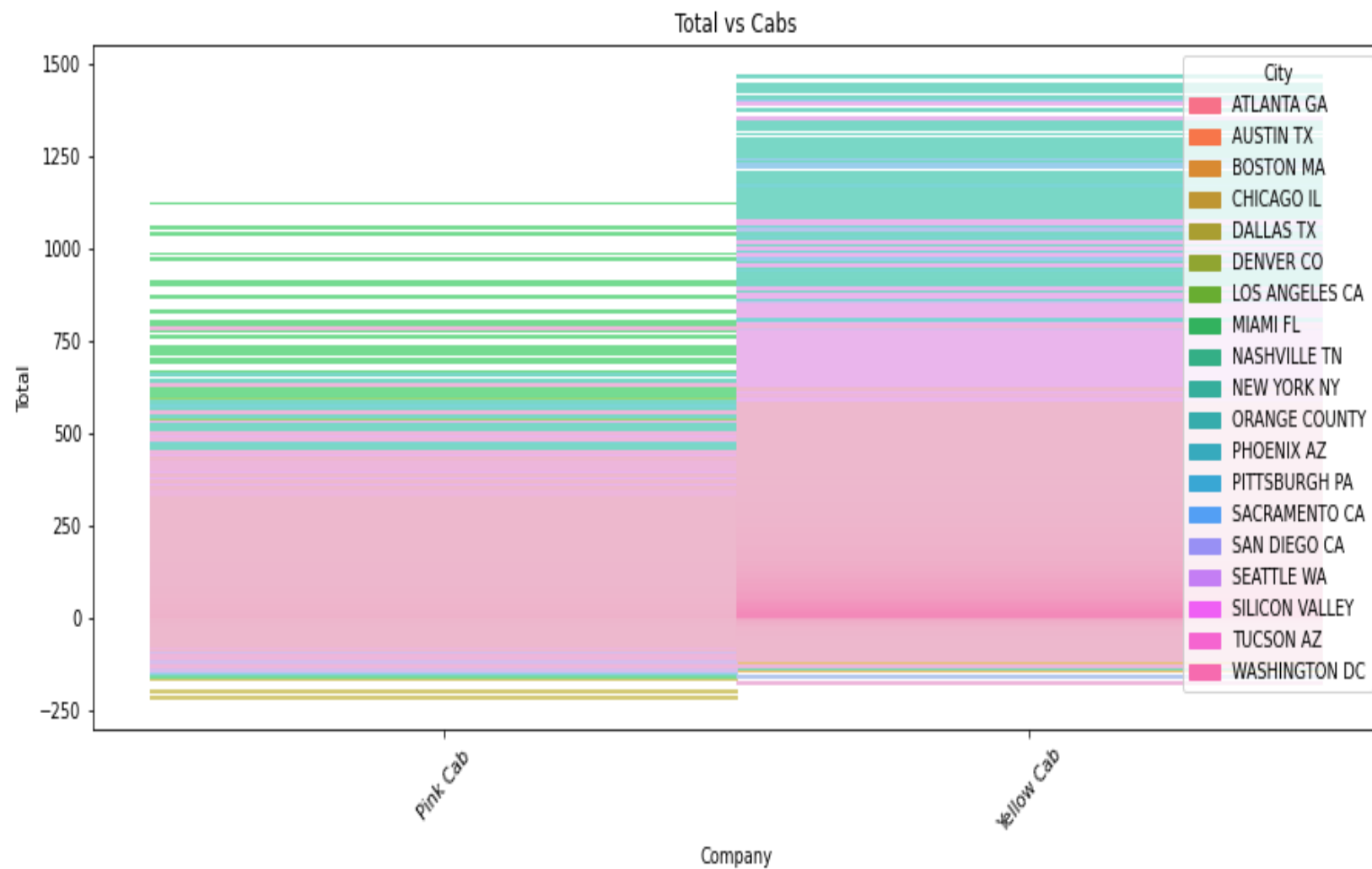
The Pink Cab also has more negative earnings than the Yellow Cab. The income earned per cab ride decreases at the very end of the chart, due to people who are usually richer not having to use taxis.



City vs. Income

This graph shows that the yellow company earns higher in the city of New York than the other cities it caters to.

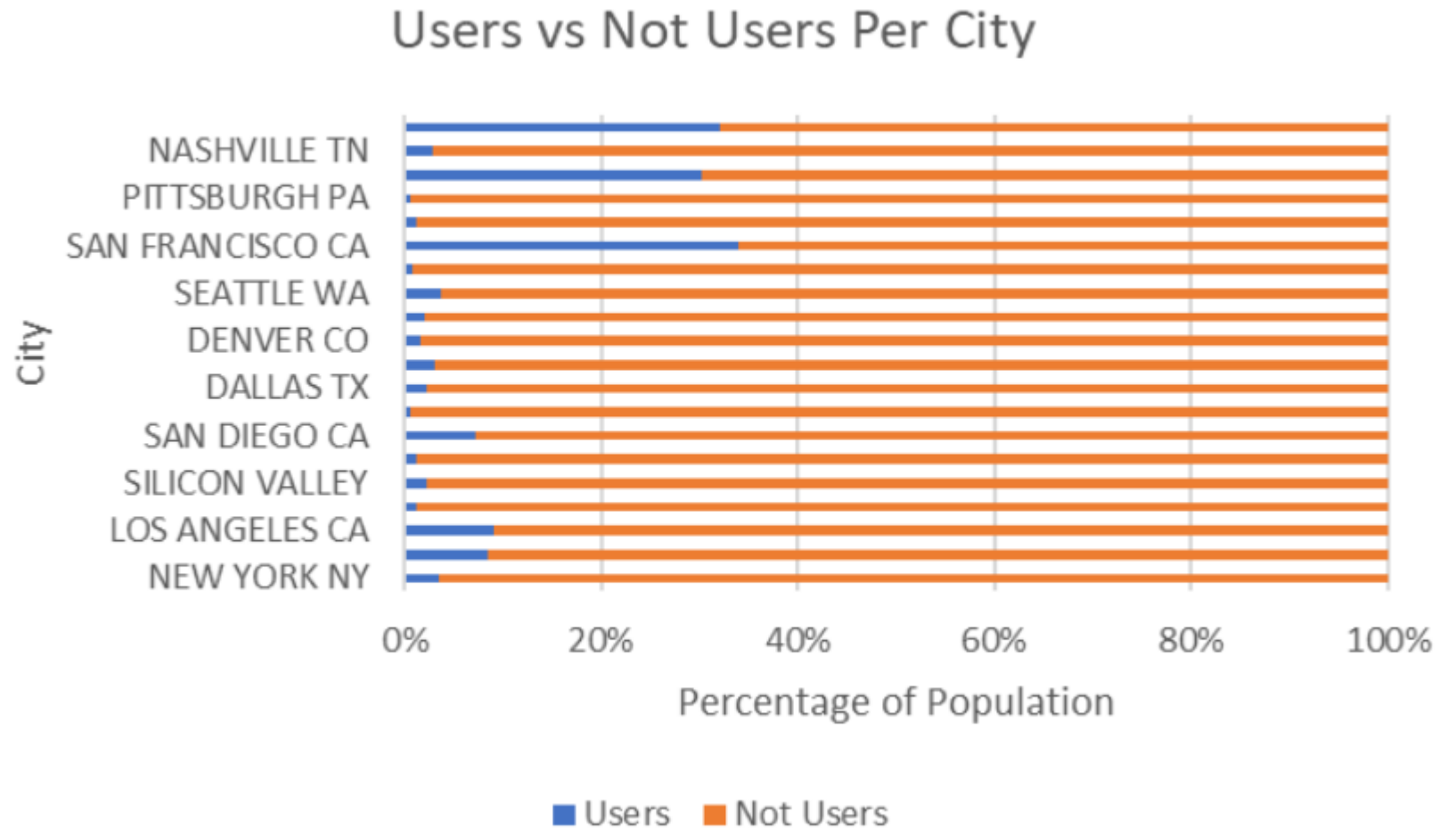
The difference in income suggests that the Yellow Cab company copy the structure of their New York franchise to other cities they would cater to, especially those with large populations.



Users vs Not Users

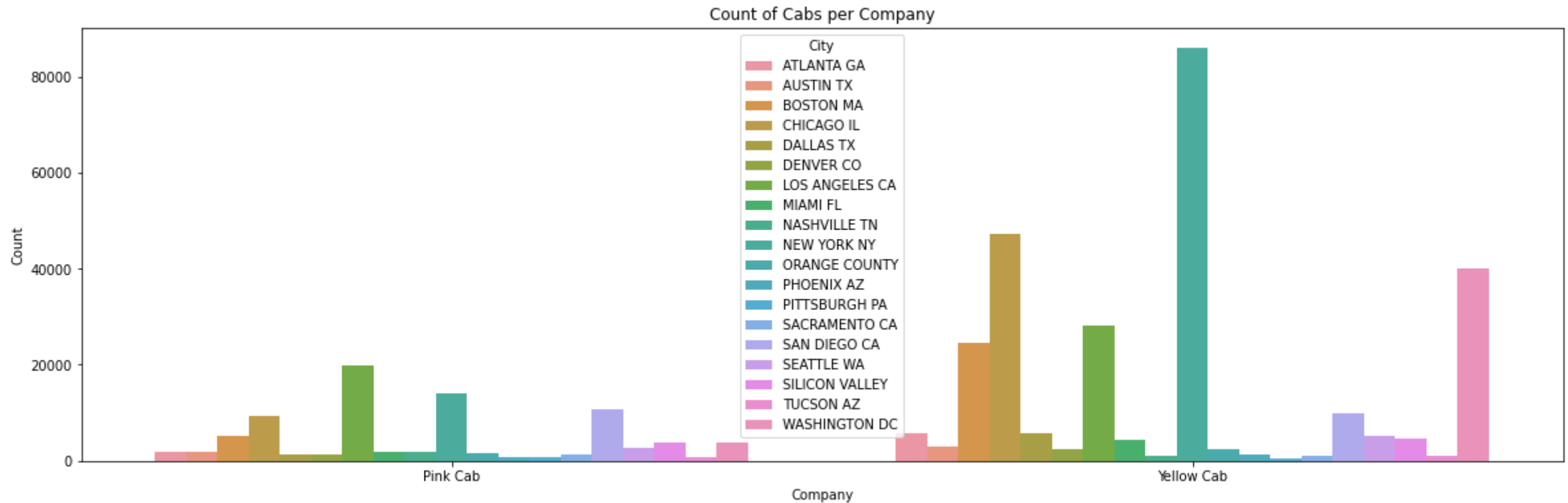
This graph depicts the percentage of users in major cities vs the percentage of people who do not use any taxi service in these cities.

Most of these cities do not have very high levels of taxi demands, so it would be important to emphasize the benefits of the companies via advertisements to customers.



Count vs City

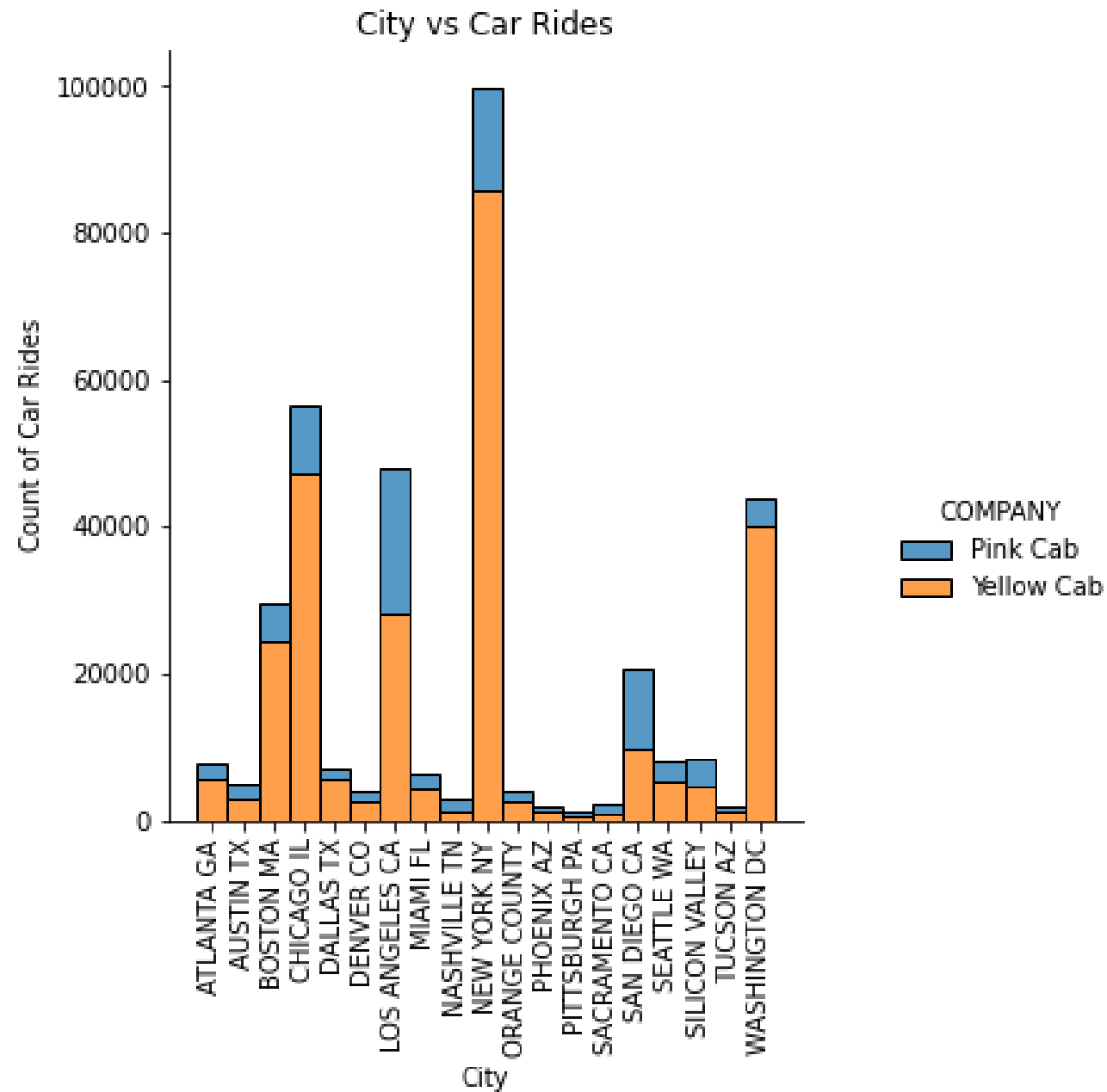
According to the graph below, the city with the most amount of rides per company is New York, who has the greatest amount of the Yellow Cab company's taxis. The Pink Cab company is mostly in LA, but the Yellow Cab company has more taxis in LA even still. The only city that the Pink Cab taxi has more coverage on is Atlanta, GA.



Count vs City

Once again, this graph shows how New York City has the most demand of taxis, where the Yellow Cab is most dominant.

Overall, every city has a greater amount of Yellow Cab rather than the Pink Cab.

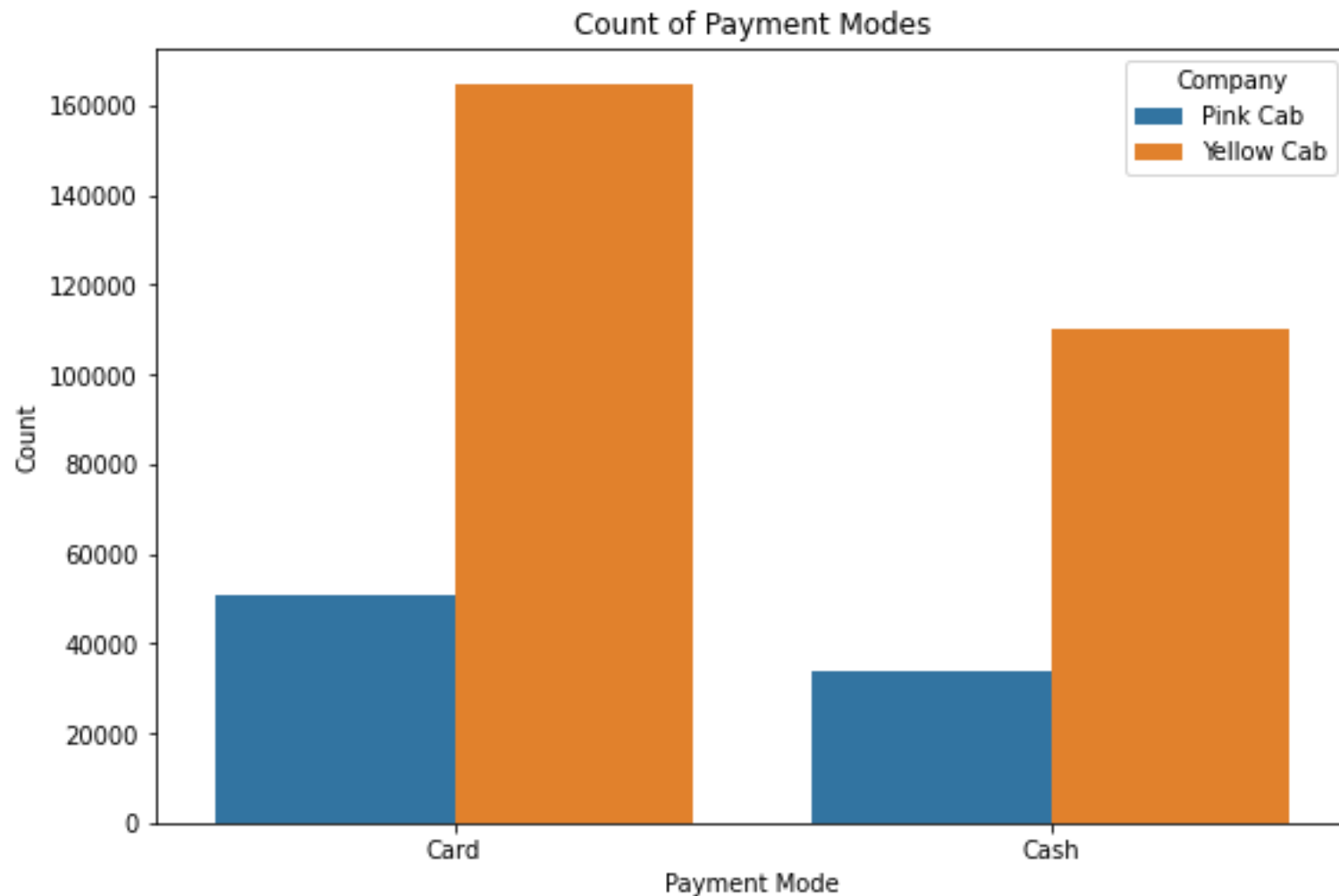


Count vs Payment Methods

This graph depicts the various payment methods the companies use and the total instances they were used.

It is clear that the Yellow Cab company had the most amount of transactions. Both companies have a greater amount of transactions with a card payment method.

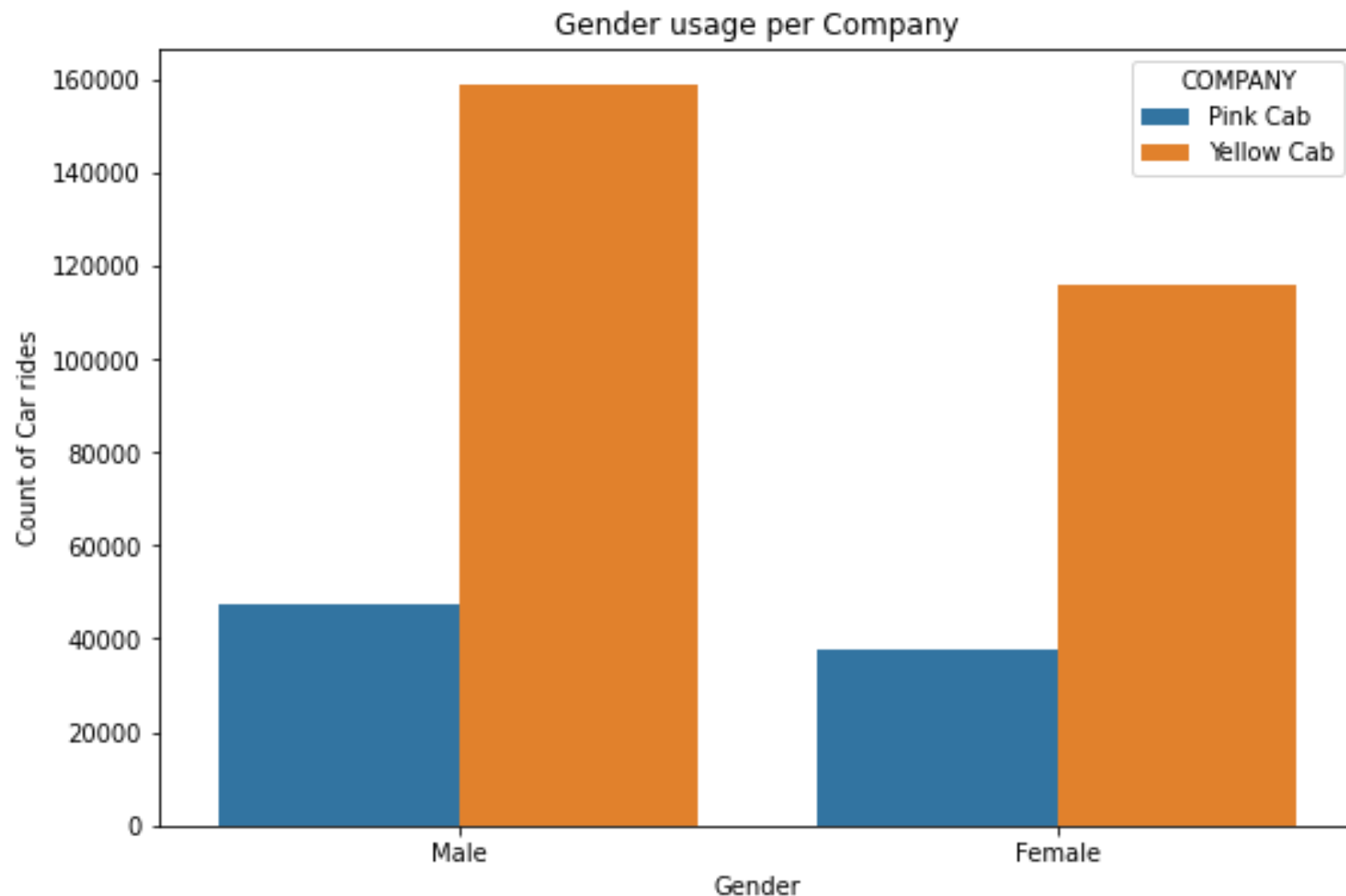
This suggests that the customer base of the companies are growing towards a more modern method of payment.



Count vs Gender Usage

This graph shows that male customers who use the Yellow Cab make up a majority of the population that uses taxis.

Female customers prefer using the Yellow Cab as well. Based on this, we can assume that both genders would use the Yellow Cab more than the Pink Cab, showing a strong customer base for the Yellow Cab company.



Conclusion

XYZ should invest in the Yellow Cab company instead of the Pink Cab company.

Per Company: Yellow Cab company dominates in income earned per car ride.

Per Date: Yellow Cab company successfully accounts for more people per day and during busy days of the week, it is the most chosen company.

Per Age: People consistently use the yellow company with no difference based on age.

Per Customer Income: No matter the customer income, people who use the Yellow Cab tend to spend more per car ride.

Per City: In major cities like New York, the Yellow Cab dominates in the number of rides wanted, showing a greater demand for it.

Per Payment Method: Card is the most preferred payment method, especially for the Yellow Cab Company.

Per Gender: Both genders consistently prefer using the Yellow Cab company.

Therefore, the Yellow Cab company will continue to remain more successful than the Pink Cab company.

Thank You