



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M

21/12/2022

Agenda

Project description

Data description

EDA

Summary

Recommendations

Background

Situation:

XYZ is a private firm in US. They are deciding which company, Yellow Cab or Pink Cab, is better to invest in.

Goal:

Perform EDA and make recommendations for XYZ regarding their investment options.

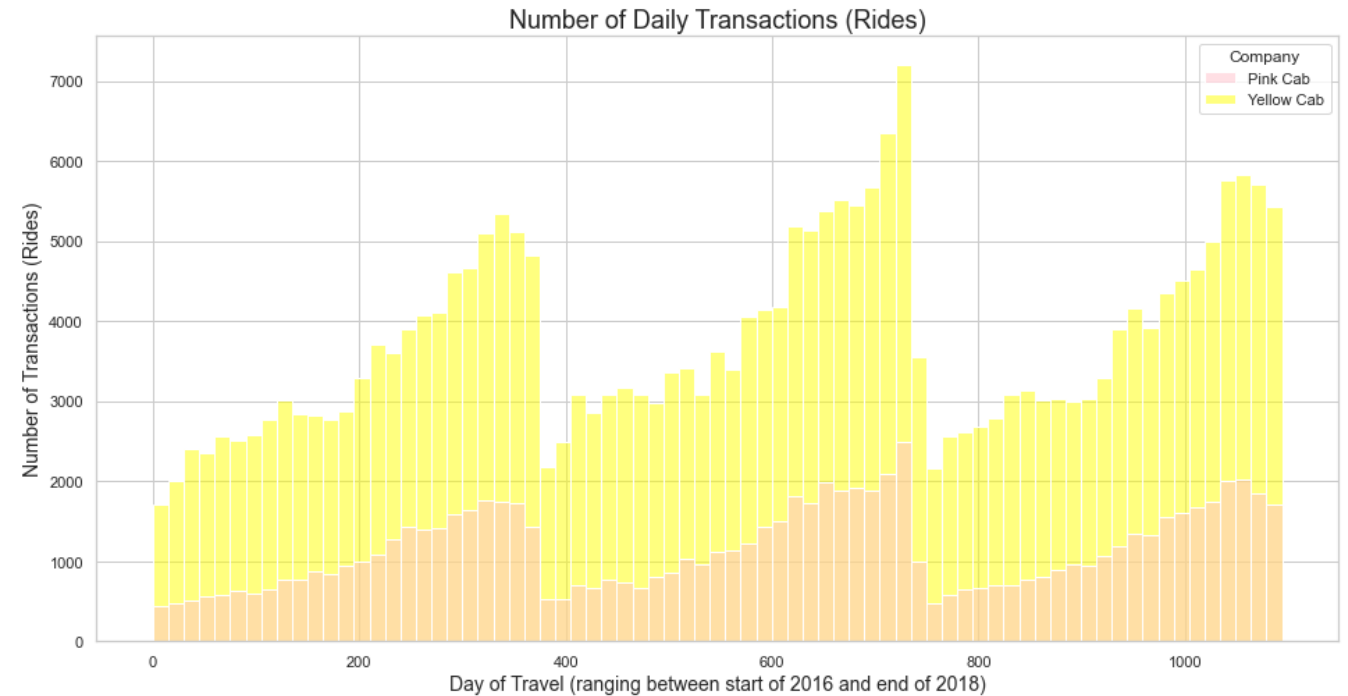
Data

Time period of data is from 01/01/2016 to 31/12/2018. There are 4 datasets:

- *Cab_Data.csv* – this file includes details of transaction for 2 cab companies.
- *Customer_ID.csv* – this is a mapping table that contains a unique identifier which links the customer's demographic details.
- *Transaction_ID.csv* – this is a mapping table that contains transaction to customer mapping and payment mode.
- *City.csv* – this file contains list of US cities, their population and number of cab users.

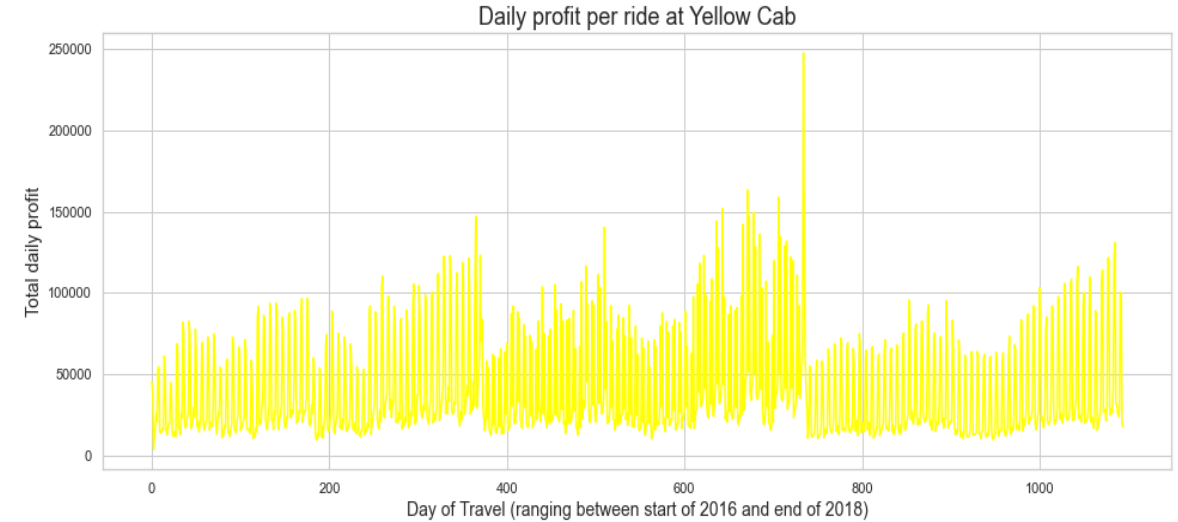
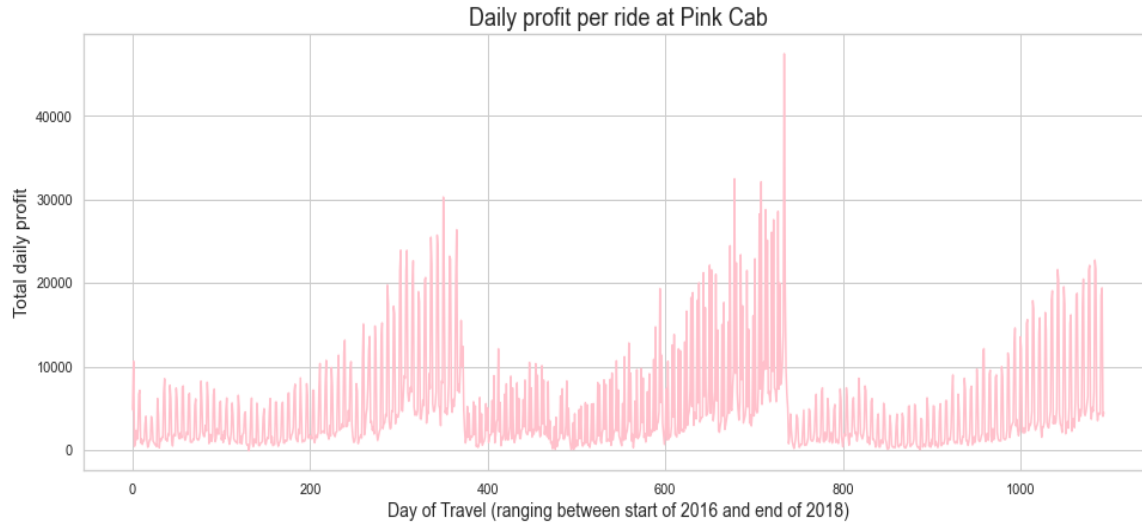
There is a total of 359,392 observations (cab rides recorded), out of which 274,681 rides were with the Yellow Cab and only 84,711 rides were with the Pink Cab.

Timeline analysis



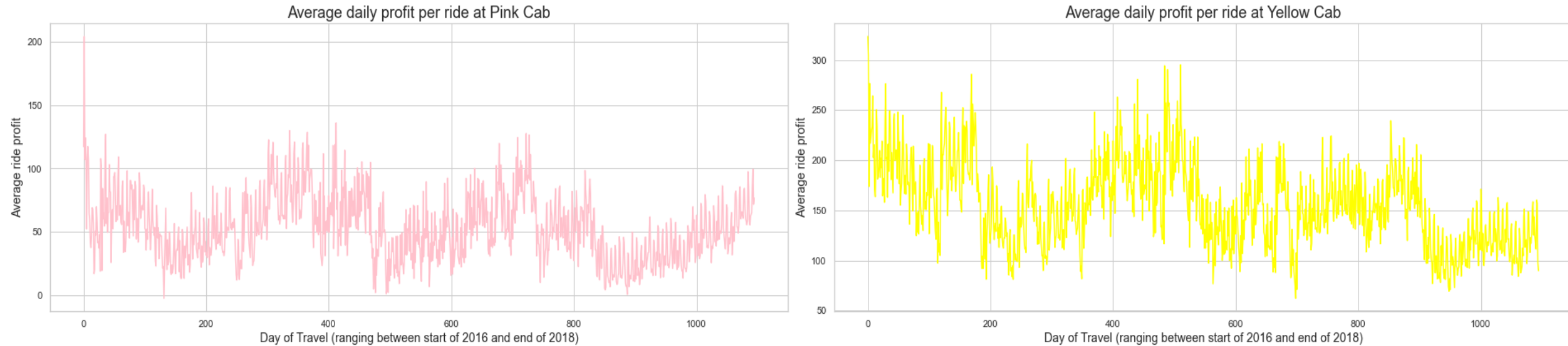
- There is some seasonality: the number of rides increase steadily at the beginning of each year until it peaks by the end of each year.
- Both companies follow the same pattern of number of daily transactions (rides) with Pink Cab having a significantly smaller part of the transactions every single date the data was collected.

Daily profit analysis



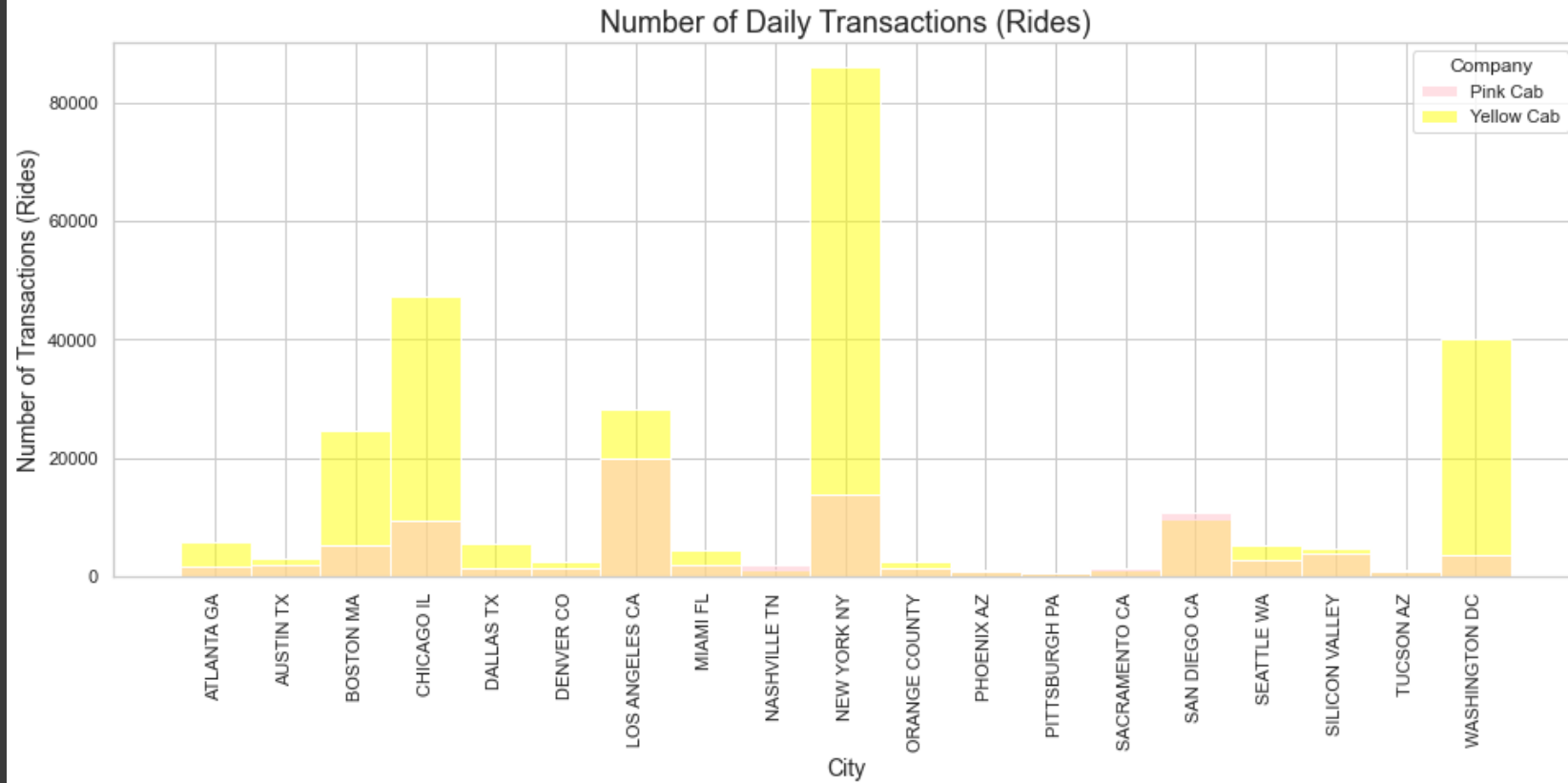
- Total profits of the Pink Cab seem to follow the pattern of the number of daily transactions, that is, the profit is very dependent on the number of rides that day and fluctuates a lot.
- The profits of the Yellow Cab are less dependent on the number of daily rides and more stable, although the pattern similar to one seen on the number of daily rides plot is still visible. Further, Yellow Cab is significantly more profitable.

Average daily profit analysis



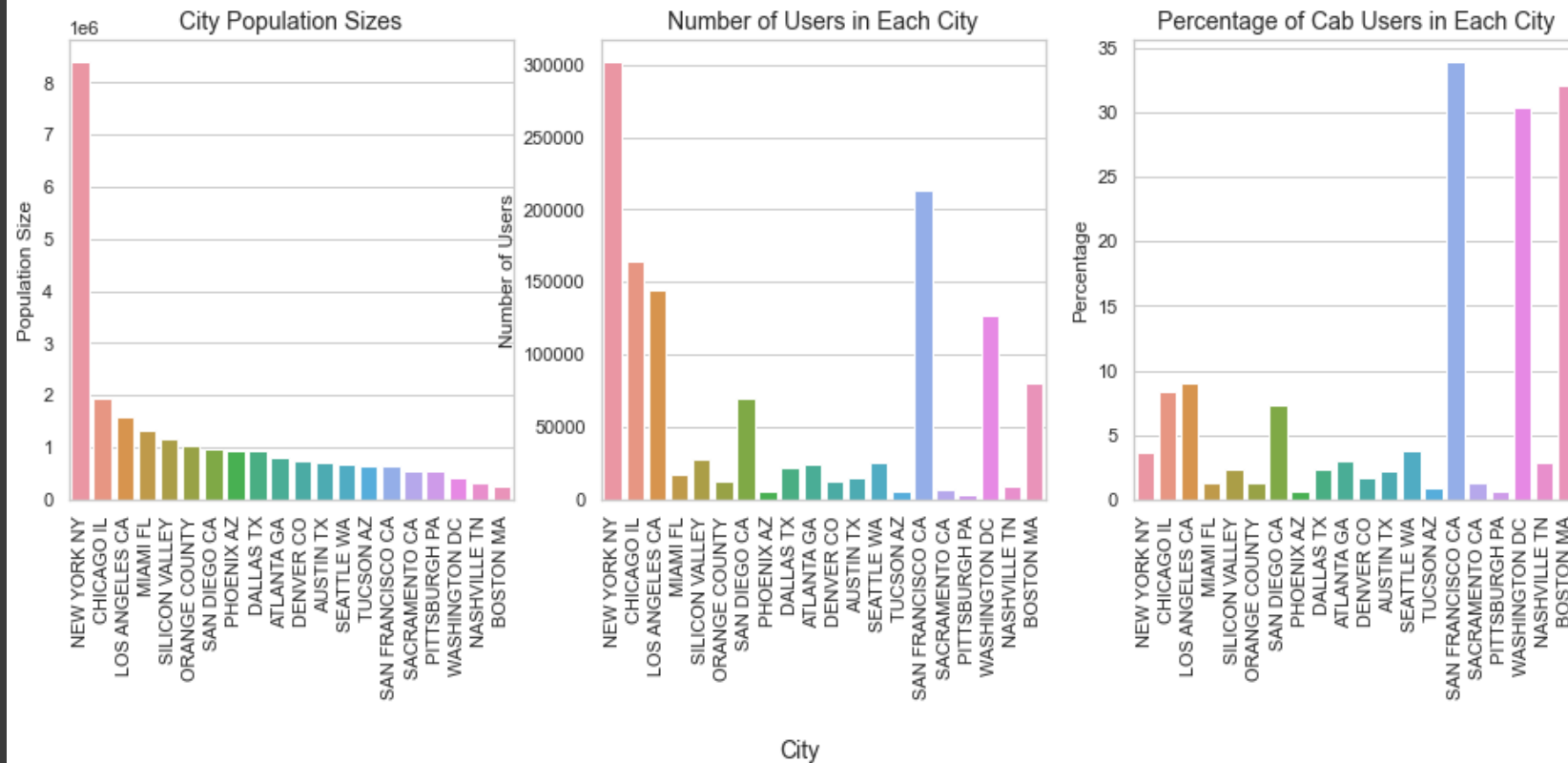
- We can see that the tendencies of average daily profit vary quite a bit daily, Yellow Cab makes greater daily average profits.
- Importantly, there is no increasing pattern in profits in either of the companies, in fact, Yellow Cab even shows potentially decreasing profitability. Thus we could not predict an increase in profits later after the investment is made, the profits should be similar as they are now for both companies.

Location analysis



- The Yellow Cab company dominates the market shared by these two companies, with a much larger market share in most cases.
- The only places where Pink Cab company has a larger market share than the Yellow Cab company are Sacramento (CA), San Diego (CA) and Nashville (TN), and even there, the difference in market shares is relatively small.

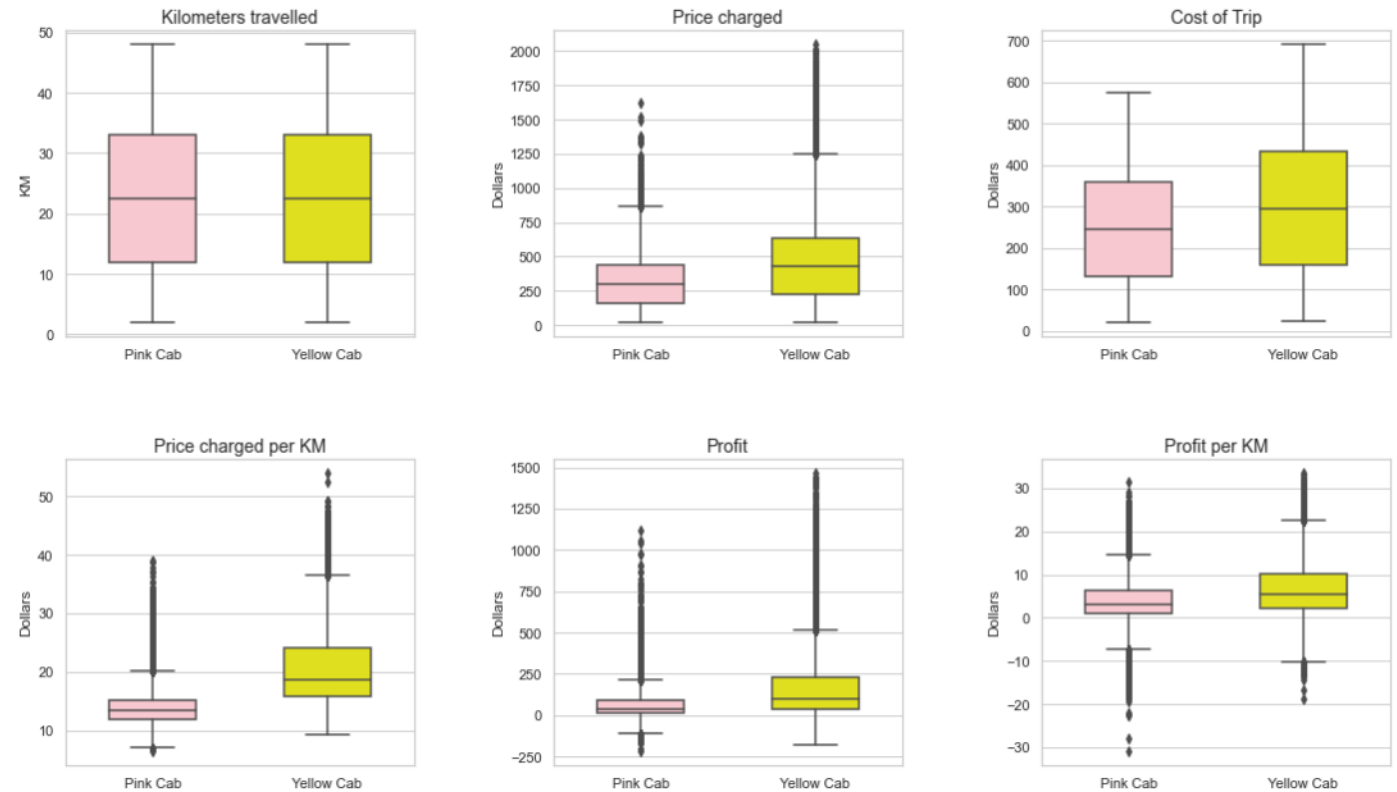
Location analysis



- Above, the leftmost plot shows the population of each city, the middle plot shows the number of cab users in each city, whereas the last plot puts it all together into perspective and shows the percentage of population in each city that are cab users.
- Even though only a small part (4%) of the New York population are cab users, they still are a significant part of all cab users in the US.
- Cab users are most frequent in San Francisco (CA), Washington (DC) and Boston (MA). Note that none of these cities were of important significance in the previous plot, in fact, in the previous plot San Francisco (CA) is not even displayed. This means that there neither Yellow Cab nor Pink Cab does business in San Francisco (CA), and, also, there are other cab companies that do well and have large market shares of cab users in these cities.

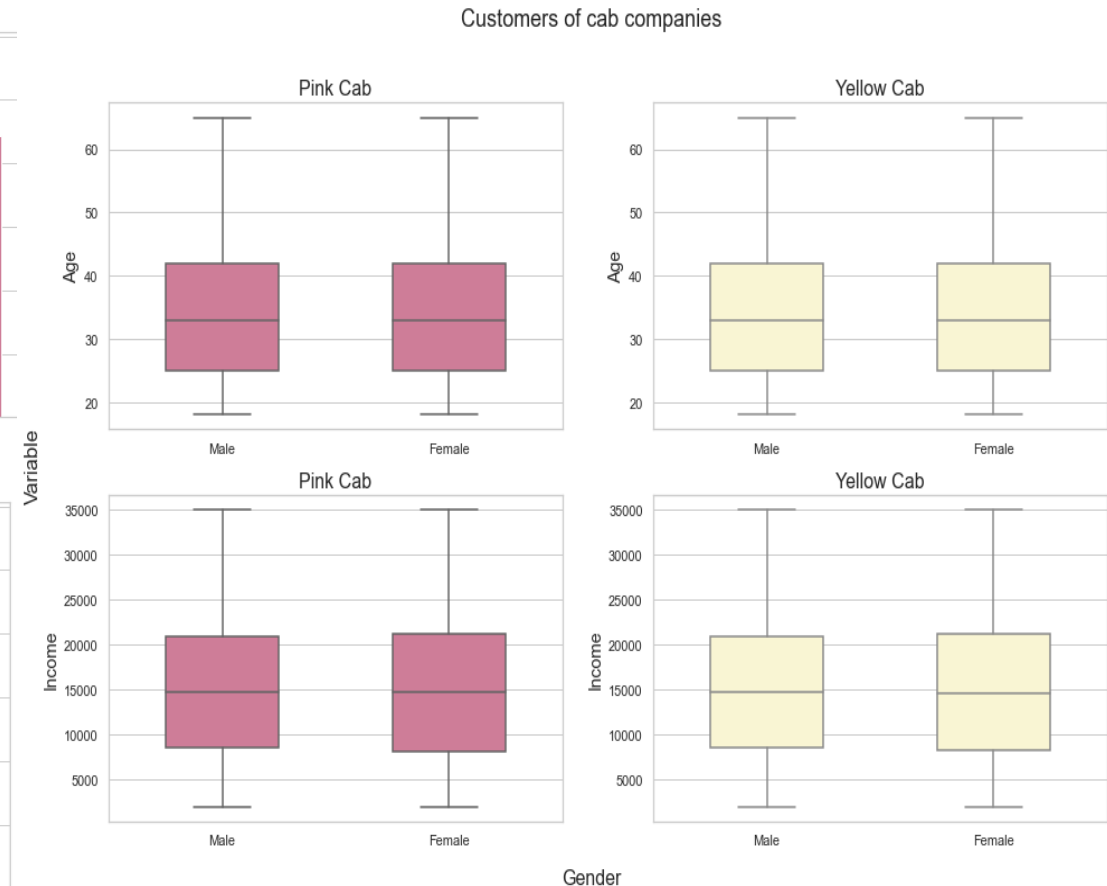
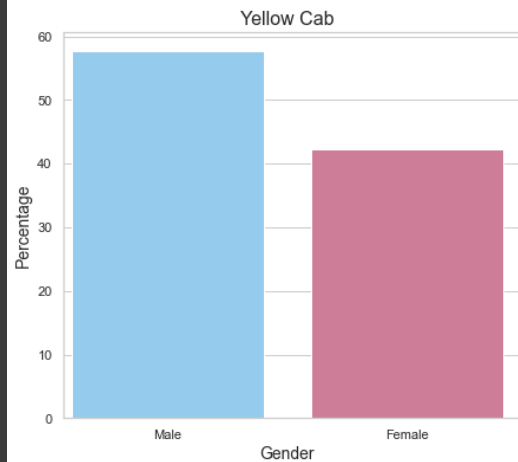
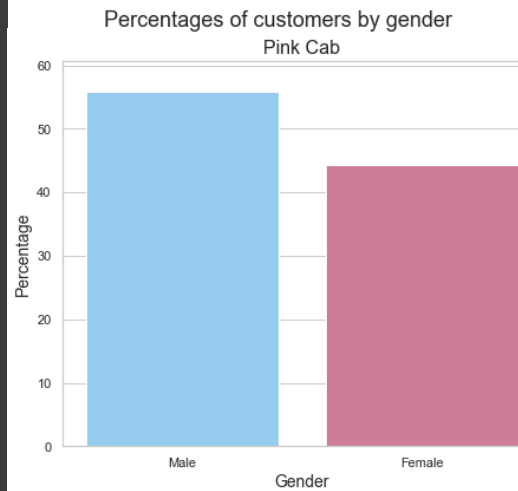
Ride cost analysis

Boxplots of ride information for each company



- We can see that the distribution of kilometres travelled per each ride (top left plot) is almost exactly the same for both companies. However, the rest of the plots differ significantly.
- Yellow Cab company tends to charge higher prices than Pink Cab company (top middle plot), however, their trips also cost more (top right plot), although the difference between costs amongst both companies is not as great as the difference between prices. Because of this, price charged per kilometre of the Yellow Cab is greater than that of the Pink Cab company, and, similarly the profit made (both overall and per kilometre) by the Yellow Cab is greater than the profit made by the Pink Cab.

Customer analysis



- We note that overall in our dataset there are 57% of male customers, and the rest are women. After looking into the gender distribution between Yellow Cab and Pink Cab companies, we can see that Pink Cab a greater proportion of female customers compared to the Yellow Cab. 42% of Yellow Cab rides had female customers, whereas 44% of Pink Cab rides had female customers.
- Clearly, the age or income of both women and men are of the same distributions and they do not vary between the two companies, suggesting that the target audience of both companies is exactly the same.

Summary

- The target customers of both companies are the same.
- The seasonality for both companies is the same.
- The profits of Pink Cab are very season dependent, whereas the profits of Yellow cab are not.
- The city with most cab rides is New York, it is also the most profitable city (higher prices than everywhere else, even though the costs are similar).
- Geographically, Yellow Cab dominates the market compared to the Pink Cab.
- The costs and profits of Yellow Cab are greater than those of Pink Cab.
- Price charged by Yellow Cab is on average higher than price charged of Pink Cab. Further, the Yellow Cab prices vary much more.

Recommendation

Yellow Cab

There is a lot of unknown information such as the quality of service, strategies of the companies, other competition, cars information, etc., which makes this decision rather uninformed. The only reliable and useful information we could extract is the profitability and market share of both companies, and, due to time limitations and lack of market research and data about cab markets, we make our decision only based on this data.

If we expect the market to stay the same (in terms of innovation and events such as the pandemic) and the companies to not change their operating strategies significantly, the evidence very strongly supports the investment into Yellow Cab company as opposed to the Pink Cab company.

Thank You