

Identifying the business problem

In this case, from the given company proposition, the challenge to be addressed is:

Helping the client make a conclusive decision on which cab company to invest in i.e. based analyzing and providing concrete actionable opinion built on the given data sets and field trends.

Hypothesis Formulation

In order to help bring out a thorough and conclusive findings to this investigative exercise the following questions will have to be satisfactorily answered:

- Which company made a larger profit based after calculating the difference between price per charge and cost of trip.
- Which company appears to have made the most transactions and consequently how much did it make.
- Which cities have a predominantly higher population-cab user ratio?
- Which cab is popular in which city.
- Which of the two cab companies seems to show a significantly higher ratio in the income-cab user frequency ratio (if any). (in other words do cab users seem to be of a higher earning regimen.)
- Is there a relationship (if any) between the average age group of cab users and their cab-use frequency. For instance are those frequently using the service distinctly associated within a certain age group e.g (18-64) Then establish, which of the two cab companies seems to show a significantly higher ratio in this regard.

Deduced findings:

In understanding each given dataset it was imperative to also establish the relations between the data features and how these were actually of any use in helping to answer or solve the problem at hand.

As such, this resulted in more columns being added or deduced from the already available data features in each set so as to help clearly bring out the hypothesis being proposed.

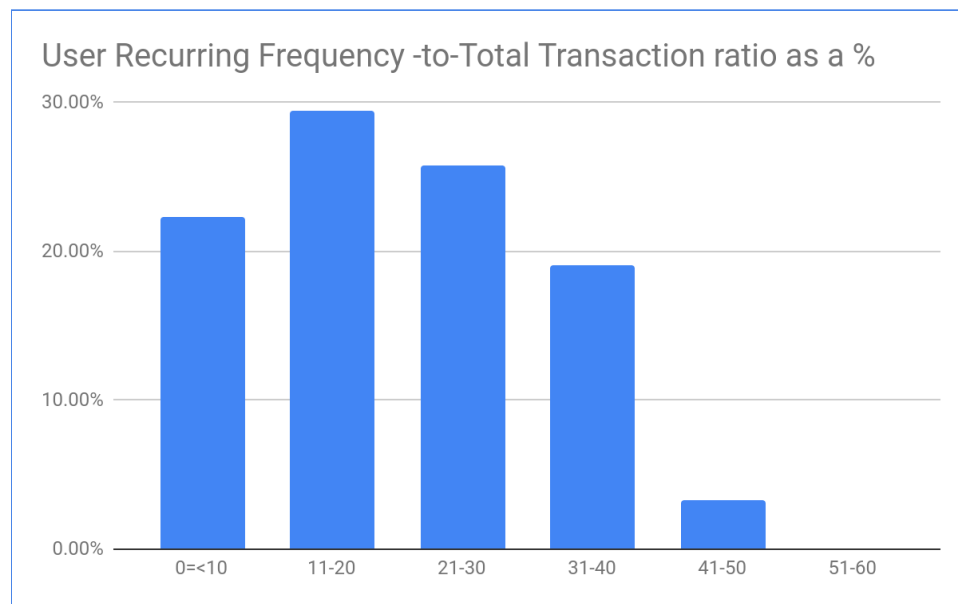
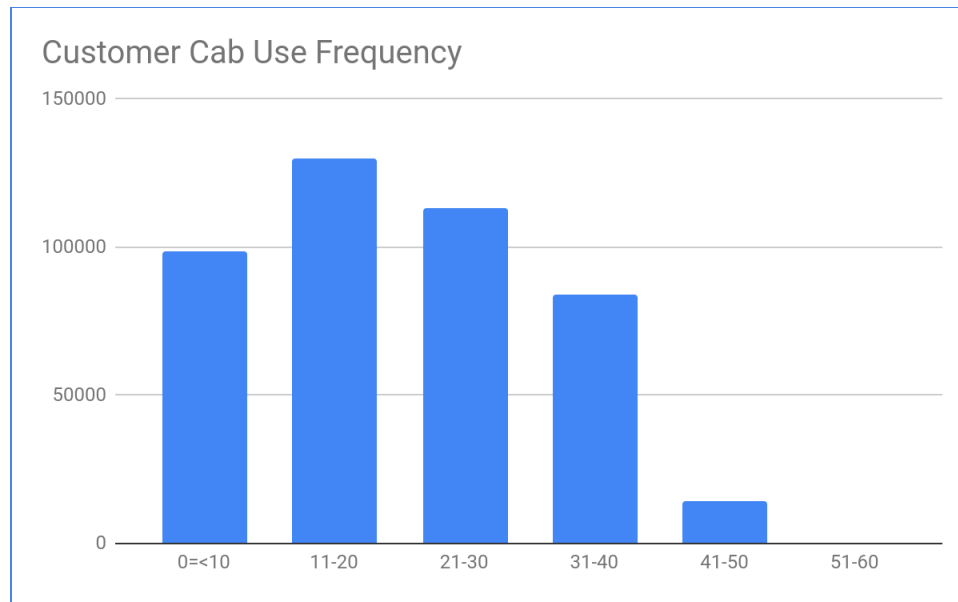
In each table, the added data feature columns were mentioned on the findings given below.

a) Findings from the Transaction table:

While considering the Transaction table, it was also interesting to actually note the frequency of transactions per customer.

This was vital as it helped to give an idea of the cab user frequency per user in the given time frame.

The idea is, the more frequent the transactions per user, the more chances the user is a dependant on the service hence a positive pointer regarding the profitability of the venture.

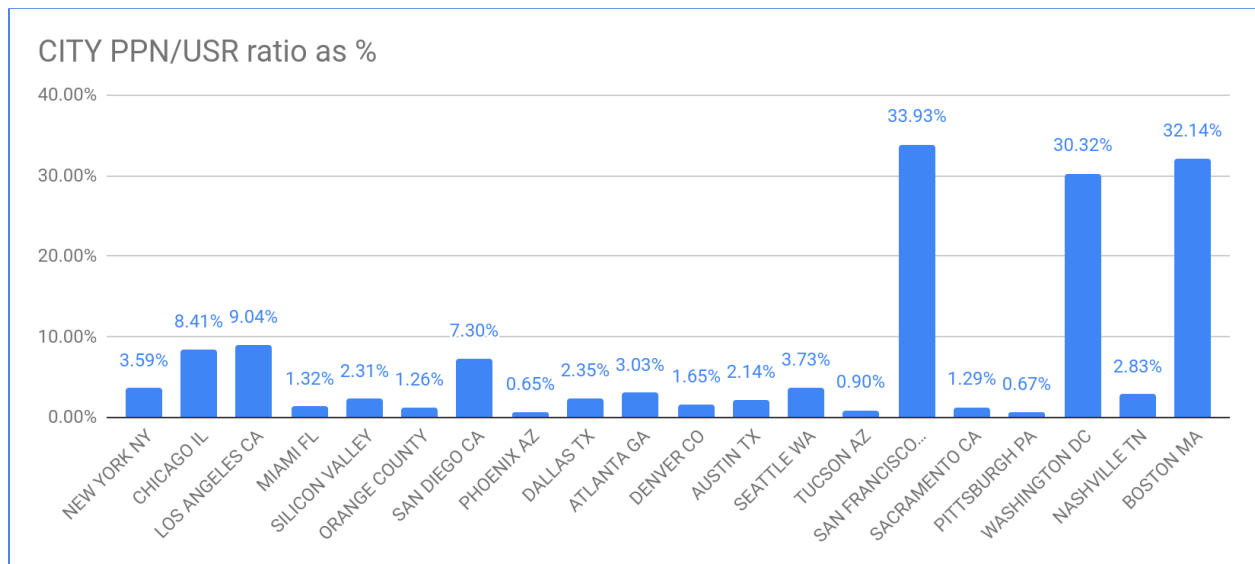


Nb: Column added- '**Customer_USER FREQUENCY**' this column was added to actually see the frequency of each user with regards to cab usage.

b) Findings from the City table:

From first glance the obvious takeaway is that a large city population does not necessarily mean a large cab user population as well.

Another interesting fact is all the seemingly investment lucrative cities have a population of less than a 1 000 000 million people. In fact, the smaller the population the higher the ratio.



1. SAN FRANCISCO CA approx. 34% to the city-cab user ratio of the given data set.
2. BOSTON MA approx. 32% to the city-cab user ratio of the given data set.
3. WASHINGTON DC approx. 30% to the city-cab user ratio of the given data set.

The three cities have a population/ user ratio above 25%, meaning over 25% of their population used a cab in the given time frame.

Consequently this is lucrative insight into understanding the potential market in these areas.

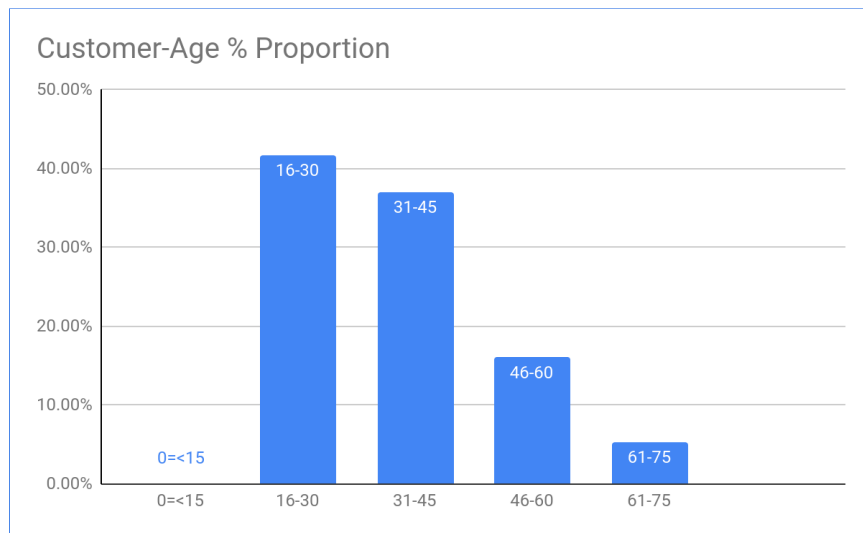
Some of the factors that could be contributing to this trend is:

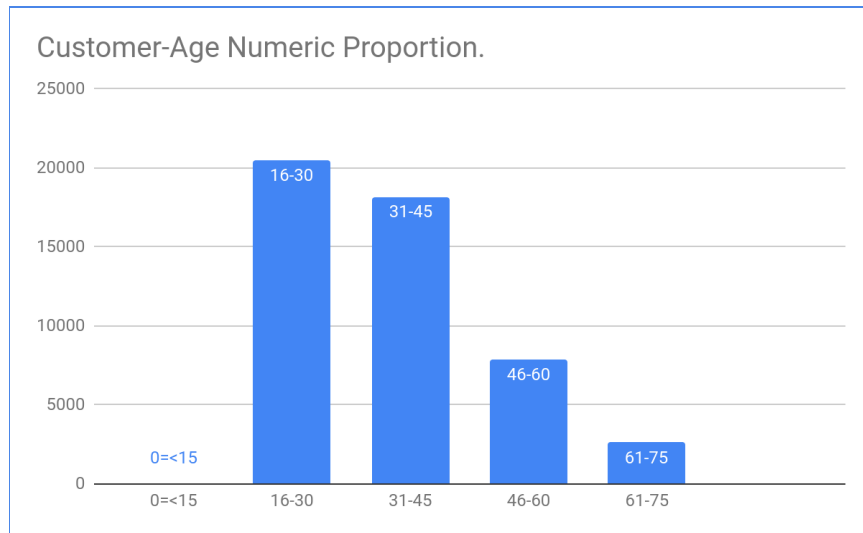
- all three are metro cities.
- all three have a robustly congested traffic network hence people are forced to rely more on cabs.
- all three have a largely expensive cost of living, hence income is relatively higher meaning the population can afford to use taxis not as a luxury but a necessity.

Nb: Column added- 'CITY POPULATION TO USER RATIO' this column was added to actually see the relation between the population and its allocated cab users per city.

c) Findings from the Customer_ID table:

From this table, it was also key to establish the specific age categories of all cab users so as to cluster them according to working age group/ independent population. This was insightful as it would help actually give an opinion on whether the majority of the population is independently capable of paying their cab fees, making them more likely to be employed and hence recurring users.

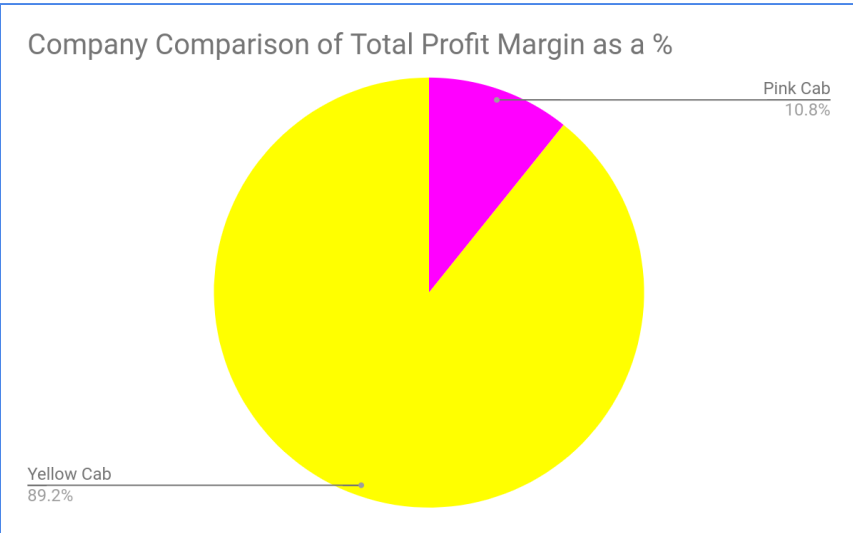
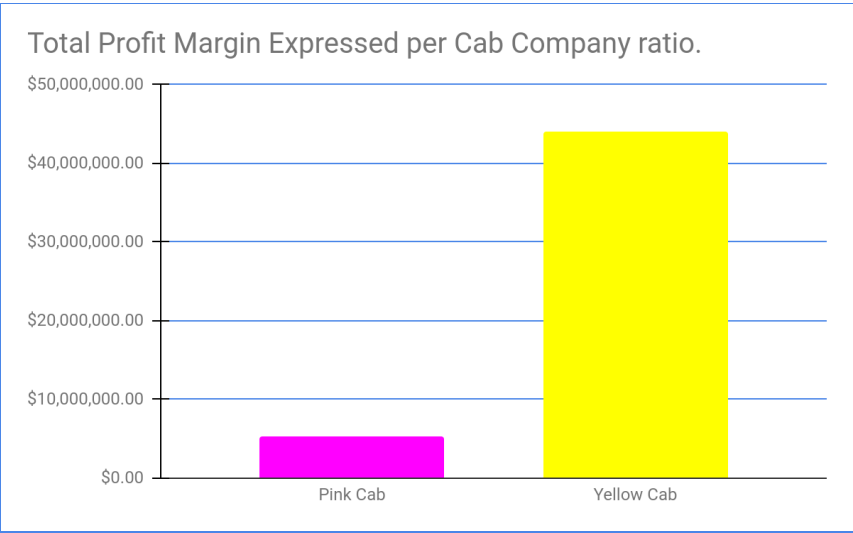


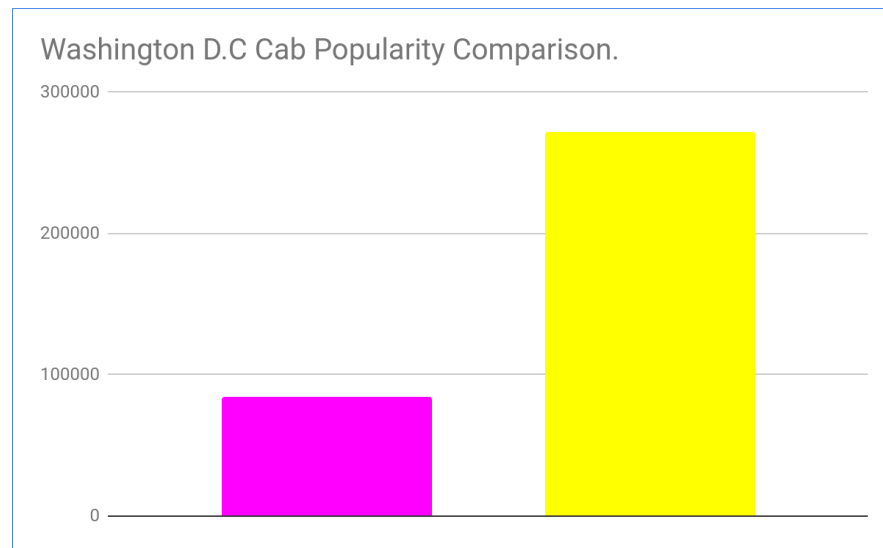
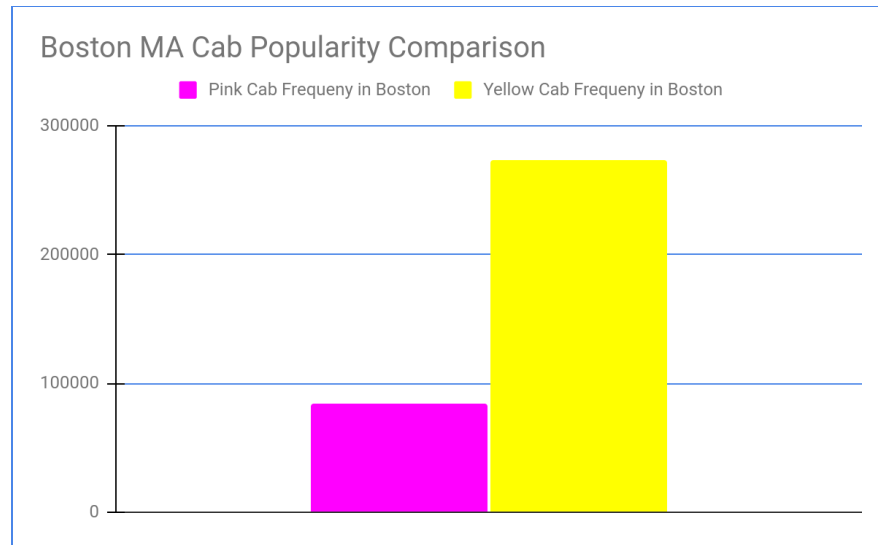


Observation-Interestingly over 75% of the population obtained for this survey wher between the ages of 16-45 which means they are still in the prime of their working years and hence would likely be counted in to use the service for linger.

d) Findings from the Cab_Data table:

The cab_data table was the most packe in terms of insights to be extracted. From this table the key questions to be established were 1. To establish the cab city popularity and make up the profit margin per cab company.





Observation- from the findings, the yellow cab company actually had more transactions and consequently made more profit. The yellow company was also the more popular of the two in both Boston and Washington DC.

Overall Recommendations

1. The cab service industry registered profits overall and hence seems to be a worthwhile business to invest in.
2. The majority of the cab users who participated in this survey were in the lower age range of the working age population and hence this means the business is generally a sustainable and viable investment opportunity as it is supported by a vibrant working population.

3. The data also showed that over 70% plus users of the service had used the service recurrently between 11-40 times. As such this proved that there definitely is a consistent customer base to support the business, making this a worthy investment offer.
4. The main three cab-user populous cities were Boston, San Francisco and Washington DC making these three the hotspot areas of developing and expanding the business to potential investors
5. Yellow cab company made close to 90 percent of the total profit in the given time frame, making it the most lucrative company to invest in.
6. Yellow cab company was also the most financially successful in the three major cities i.e Boston, Washington DC, San Francisco making it the worthy choice of investing in.