



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M Insight For Cab Investment

03/21/2023

Agenda

Executive Summary

Approach

EDA

EDA Conclusions

Recommendations



Data Glacier

Your Deep Learning Partner

Executive Summary

Client:

XYZ is a private firm in US. Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, it is planning for an investment in Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before taking final decision.

Objective:

XYZ is interested in using your actionable insights to help them identify the right company to make their investment.

Five Main Analysis Pieces:

Profits, Locations, Age Groups, Payment method, and Customer Retention.

Approach

Data Exploration:

There are 359,352 total rows in the dataset. There are 20 total features that was given and an additional 4 that were derived from the dataset. The time period is between 2016 and 2018.

Assumptions:

Data analysis were performed for both companies under the assumption of external noise beyond the data that was provided.

Data analysis were performed for both companies under the assumption that the dataset were time constrained between 2016 and 2018.

The datasets were drawn from random selection.

Payment by cash or card are the only payment methods considered.

Price Charged - Cost of Trip were the only factors when calculating profit.

Profit Analysis

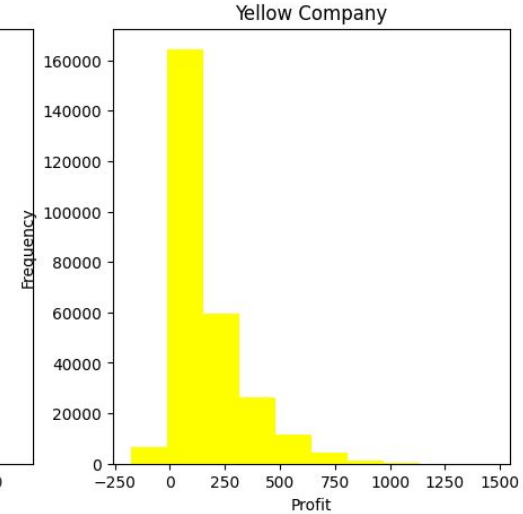
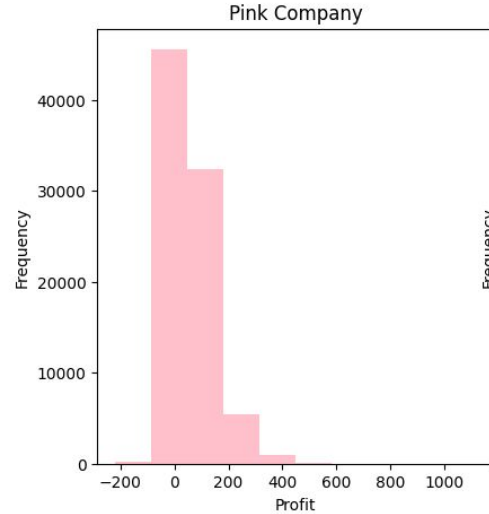
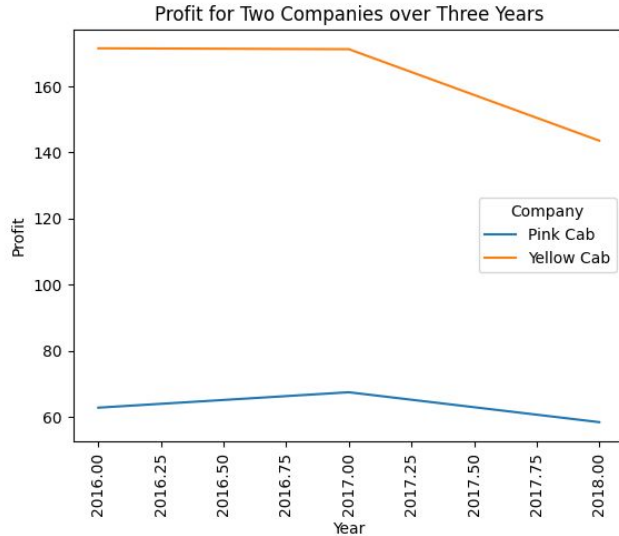
Motivation:

The first strategy of differentiating the two company's value is to figure out their profit margins. The company that shows a higher profit margin will have a greater value.

Tests:

The first test was how much total profit was made between the two companies in all four quarters between 2016 and 2018. The second test was to show the frequency of profit margins per a singular transaction.

Profit Analysis Results



Conclusion:

The first graph shows that in all four quarters between 2016 and 2018, Yellow Cab always made more total profit. However, this isn't really that telling, considering that Yellow Cab has more data provided (assuming the number of customers are about the same for both). Which is why I performed the second test. The results show that Pink Cab's singular transaction profit margins mainly hover around 0, while Yellow Cab's singular transaction profit margins mainly hover around 0-250. Thus, Yellow Cab is valued more, since its profit margins are higher based on frequency basis.

Location Analysis

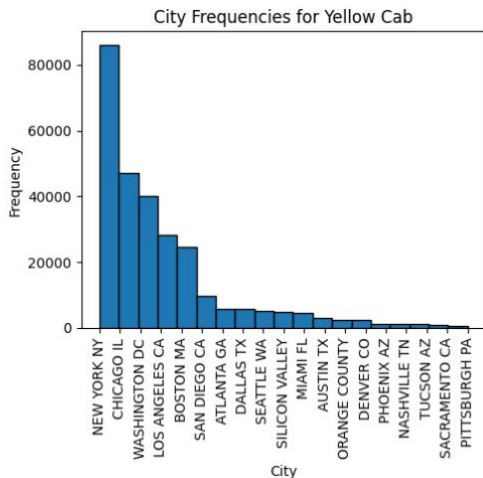
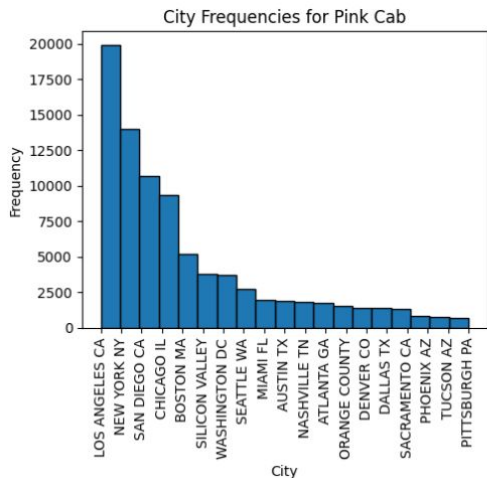
Motivation:

For this part, I wanted to see how the profits were distributed between the top three cities that produced the most profits for the two companies. Having a more equal distribution of profits coming from different locations can be safer during local economics downturns, meaning they will less likely take a hit to their profits.

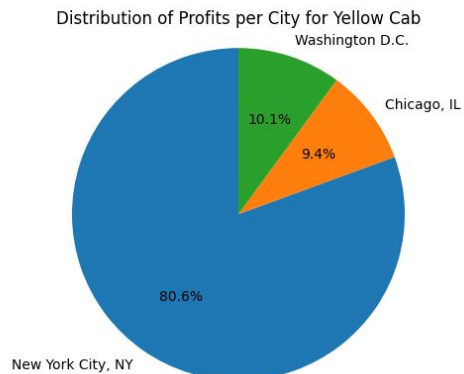
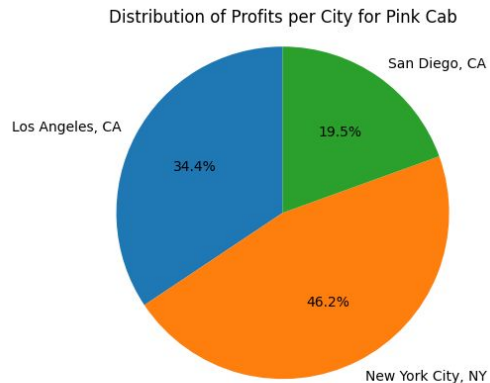
Tests:

The first test was making an histogram of the number of transactions recorded in each city. The second test was making a pie chart of the distribution of profits of the top three cities for each company.

Location Analysis



Conclusion: Yellow Cab is very top-heavy, as New York City contributed to 80% of their profits. Pink Cab is more equal among its top three cities compared to Yellow Cab. Thus, this analysis favors Pink Cab slightly more, as they are less likely to take losses if there is a local economic downturn in their top cities.



Age Groups Analysis

Motivation:

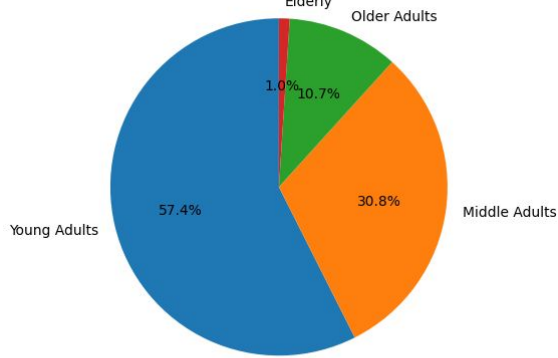
For this part, I wanted to see how the age groups were distributed among the two companies based on their customers. I wanted to see if one company had more younger customers among their distribution, because younger customers are more likely to keep using the service for a longer period of time.

Tests:

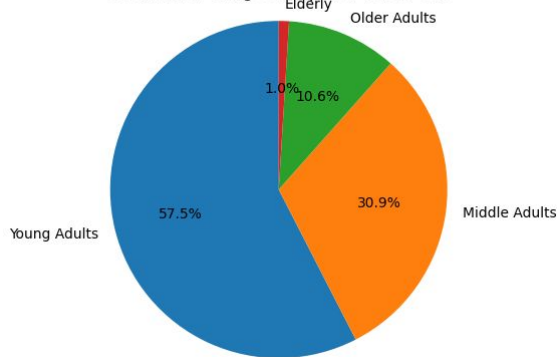
The first test was making a pie chart of four different age groups: young adults, middle adults, older adults, and elderly. The second test was making a pie chart of the total profit distribution in terms of these four age groups.

Age Groups Analysis Conclusion

Distribution of Age Groups for Pink Cab



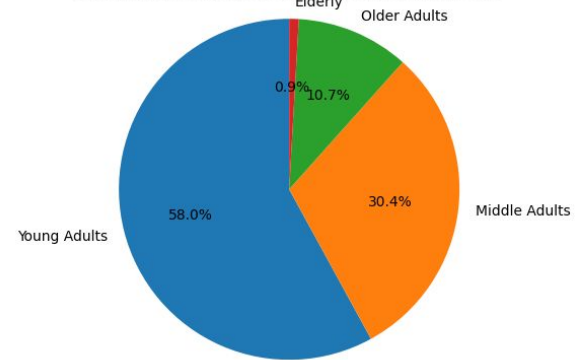
Distribution of Age Groups for Yellow Cab



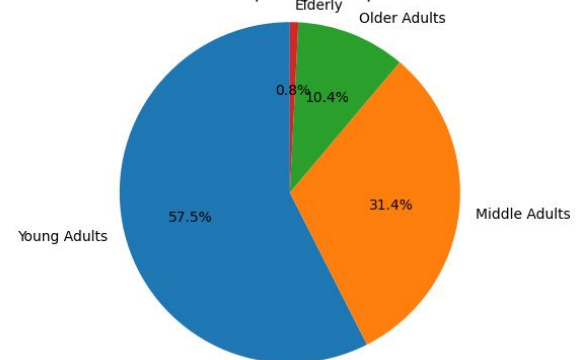
Conclusion:

The results are astonishingly similar. The distribution of age group distributions are basically the exact same. From this results, it wasn't surprising that the total profit distribution based on these age groups were also incredibly similar. Thus, this analysis proved to be important, in the sense that it gave us a better idea of how the age groups are distributed among the customer base of the two companies. However, it does not give us a conclusive answer as to which company performed better for this analysis.

Distribution of Profits per Age Groups for Pink Cab



Distribution of Profits per Age Groups for Yellow Cab



Payment Method Analysis

Motivation:

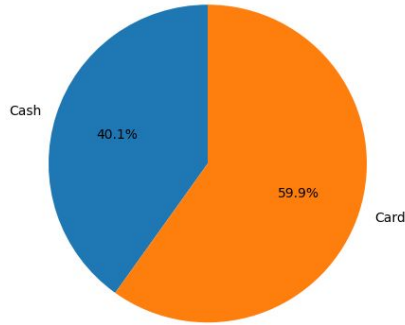
For this part, I wanted to see if there was a difference of cash and card users between the two companies. More card transactions would mean that the company has taken steps to adapt to changing technology and this will eventually become the more dominant payment method.

Tests:

The first test was making a pie chart of cash and card users for the two companies. The second test was making a line graph how percentage of card transactions changed between 2016 and 2018 for the two companies.

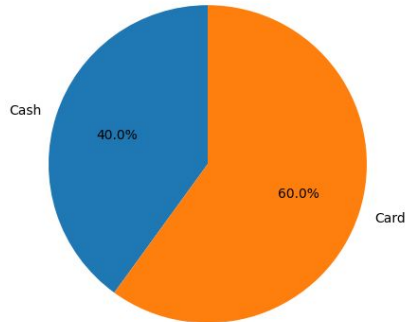
Payment Method Analysis

Distribution of Payment Method for Pink Cab



Yellow Cab number of cash transactions: 109896
Yellow Cab number of card transactions: 164785

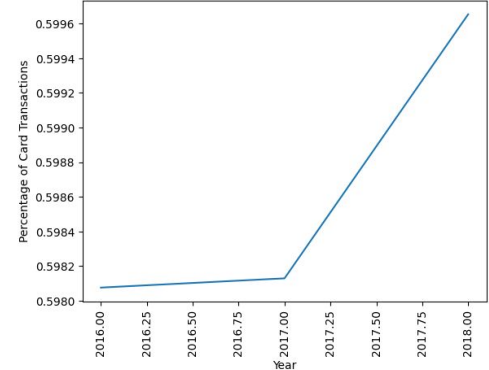
Distribution of Payment Method for Yellow Cab



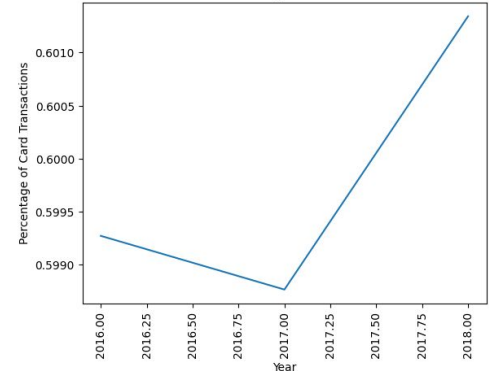
Conclusion:

The results are incredibly similar between the two companies. Both companies show about 40% as cash users and about 60% as card users. This means that there is no considerable difference between the two companies. The line graph also suggests the same results, where the increase in card transactions is incredibly minimal between 2016 and 2018 for both companies. Thus, this analysis did not differentiate the two companies enough, where a conclusion can be drawn, in terms of who performed better.

Pink Cab Percentage of Card Transactions



Yellow Cab Percentage of Card Transactions



Customer Retention Analysis

Motivation:

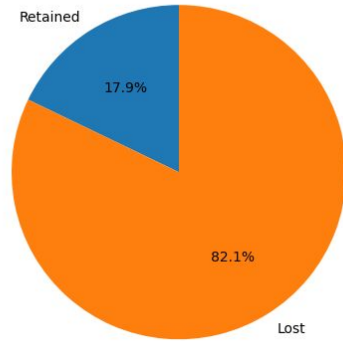
For this analysis, I wanted to find how often the same customers wanted to continue to use each company's cab services. The two main pieces that I looked at was the % of customers that used the cab service in 2016, also used the cab service in 2018. The other piece is how the average number of cab service calls from 2016 to 2018 changed for the same customer. This will show how each company is doing, in terms of their customer retention abilities, which is a high indicator of how valuable the company is.

Tests:

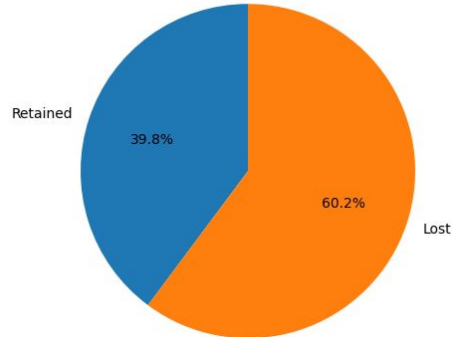
The first test was making a pie chart of how many customers were retained between 2016 and 2018. The second test was making a line graph that showed how the average number of cab service calls from 2016 to 2018 changed for the same customer.

Customer Retention Analysis Conclusion

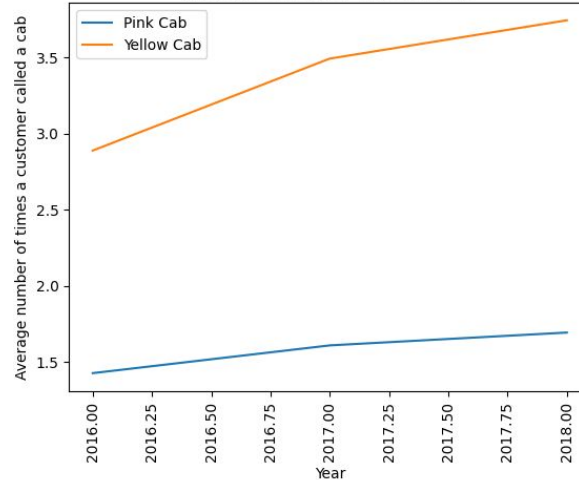
Pink Cab % of Customers Retained/Lost (2016-2018)



Yellow Cab % of Customers Retained/Lost



Average Customer Calls per Year



Conclusion: The data shows that the average transactions by the same customer per year between 2016 and 2018 grew about 18.7% for Pink Cab. The data shows that the average transactions by the same customer per year between 2016 and 2018 grew about 29.6% for Yellow Cab. This means that Yellow Cab provided a cab service that customers wanted to increase their usage from 2016 to 2018. This shows that Yellow Cab has better quality of services. The data shows that Pink Cab was only able to retain 17.9% of its customers between 2016 and 2018. While Yellow Cab was able to retain 39.8% of its customers between 2016 and 2018. This means that Yellow Cab provided a cab service that a higher percentage of their customers wanted to continue using their service. Thus, this means that Yellow Cab has better quality of services that was more incentivizing to the customers.

Recommendations

The five main categories that I analyzed were: Profits, Locations, Age Groups, Payment method, Customer Retention.

General Overview: Out of those, age groups and payment method were shown to be extremely similar between the two companies. However, locations, profits, and customer retention were heavily one-sided.

Location Analysis: For this part, I wanted to see how the profits were distributed between the top three cities that produced the most profits for the two companies. Yellow Cab was shown to be extremely top-heavy as nearly 80% of their profits came from New York City. Whereas, Pink Cab has a more even distribution. In my opinion, I think it is generally safer to have a more even distribution of profits per locations. Thus, I think that this analysis sides with Pink Cab more. Furthermore, I do think that this is the weakest determinant out of the three categories that were one-sided.

Profits: This analysis favors Yellow Cab in every way. Higher total profits in the three years and higher profit margins per transaction. Thus, this was a pretty straightforward interpretation. This sides with Yellow Cab.

Customer Retention: I wanted to find how often the same customers wanted to continue to use each company's cab services. The two main pieces that I looked at was the % of customers that used the cab service in 2016, also used the cab service in 2018. The other piece is how the average number of cab service calls from 2016 to 2018 changed for the same customer. From the conclusions from the previous slides, both pieces support Yellow Cab.

Overall, Yellow Cab clearly performed better on profit analysis and customer retention, while Pink Cab performed better on how location profits are spread out. Considering that the first two categories are better determinants of how a company will be doing in the future, I strongly conclude that Yellow Cab is the better investment.

Final Recommendation: Invest in Yellow Cab.

Thank You



Data Glacier

Your Deep Learning Partner