G2M Case Study

Date: May 21st, 2024

Presented by: Jackson Taylor

Cohort: LISUM33

Background

- XYZ is a private firm in the US
- They want to invest in the cab industry
- Needs to choose between two cab companies to invest in: (Pink Cab and Yellow Cab)
- Objective: Provide supporting insights to make a cab company recommendation for XYZ to invest in
- The default datasets provided for this task (Cab_Data.csv, Customer_ID.csv, Transaction_ID.csv, City.csv) include data for cab transactions, customers, and residing cities for both Pink Cab and Yellow Cab from the start of 2016 to the end of 2018

Assumptions

- In Cab_Data.csv, for each transaction, the profit for that transaction is (Price Charged - Cost of Trip)
- Profit outliers are removed because they skew the overall dataset in a way that matters more than a variable like cost or charge
- Removing profit outliers has clearly more of a dampening effect on the Yellow Cab data than the Pink Cab data since Yellow Cab was found to have zero low profit outliers and a high number of high profit outliers which dominate Pink Cab's high profit outliers
- This fact further supports evidence in the following slides in favor of investing in Yellow Cab
- In Cab_Data.csv, Date of Travel is number of days starting from January 1, 1900
- In the City.csv, the number of users is the total number of cab users in the city including from Pink Cab, Yellow Cab, and other cab companies
- A ride can be servicing multiple people/transactions at once. The cost is adjusted to only count the cost once for multiple people being driven the same route

Content Summary

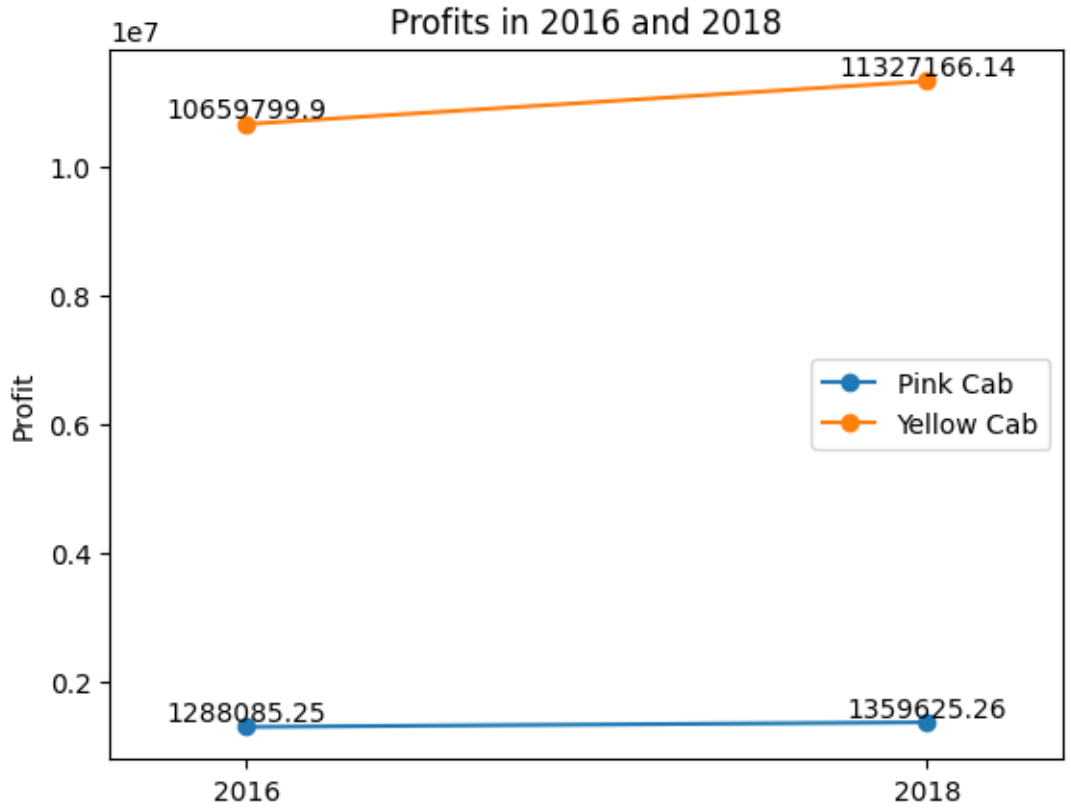
Analysis angles:

- Raw profit analysis (slides 5-7)
- Average profit per transaction analysis (slides 8-9)
- User retention analysis (slide 10)
- Cost Charge analysis (slide 11)
- Userbase analysis (slides 12-14)

Conclusion:

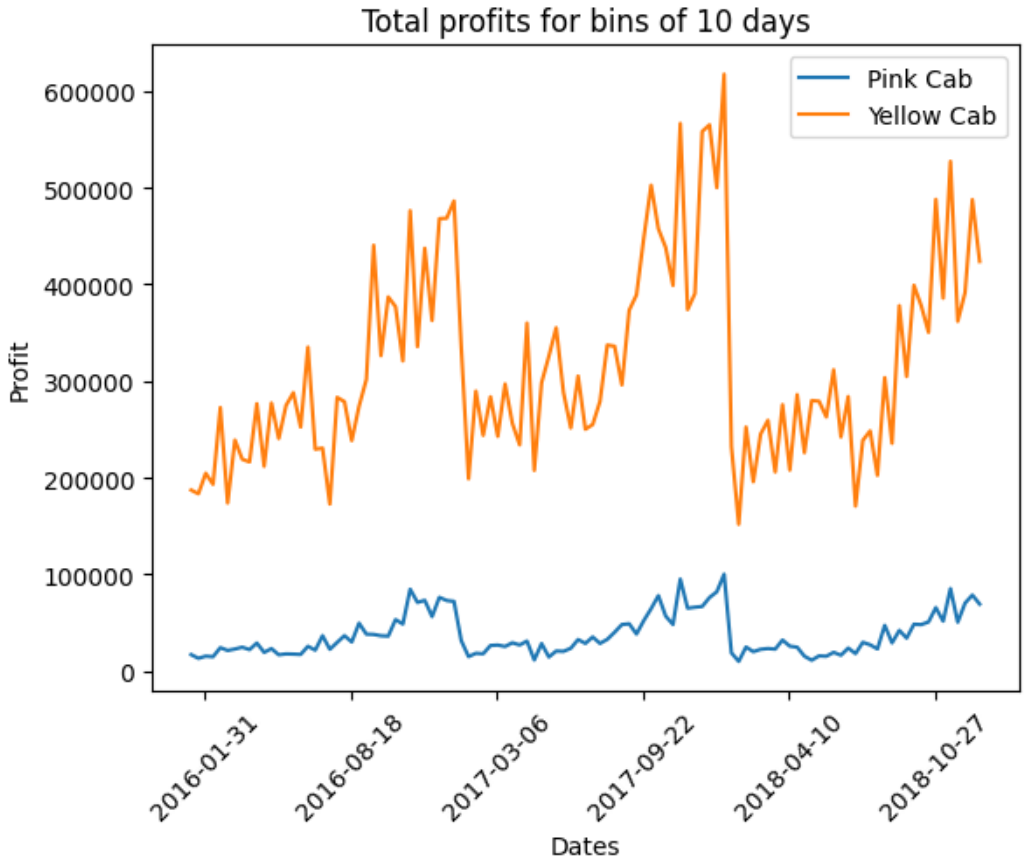
- Summary insights
- Recommendation

Company Projection



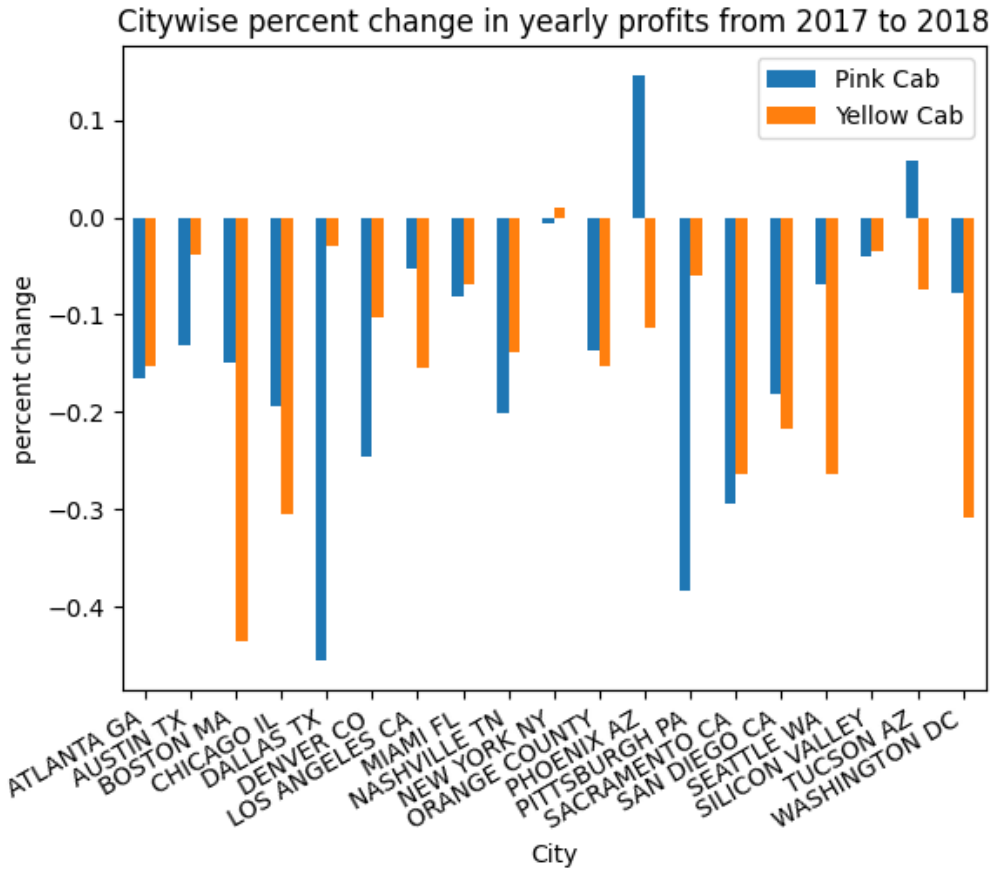
- Comparing the start and end years of the company data
- The 2016 to 2018 yearly profits percentage increase is higher for Yellow Cab than it is for Pink Cab (6.26% vs 5.55%)

Profit details over the entire timespan



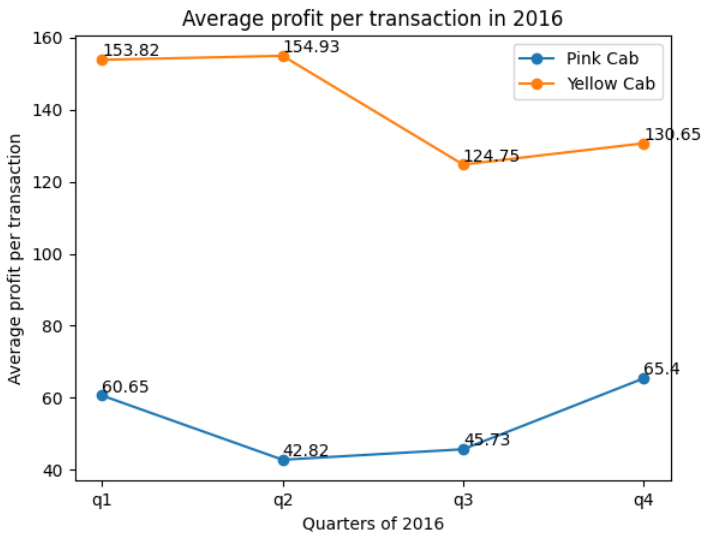
- Total profits for every distinct group of ten days for Pink Cab and Yellow Cab
- This shows yearly seasonally in profits being made
- Similar trends are being made for both companies to some degree
- Yellow Cab's profits are always higher than Pink Cab's profits

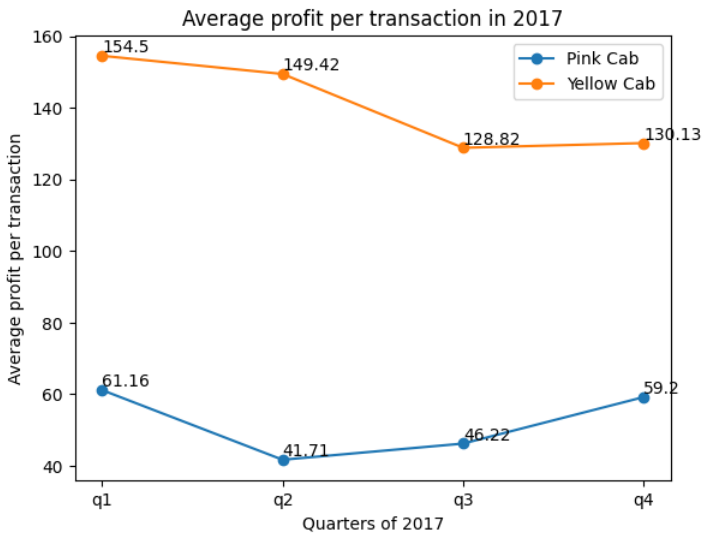
City wise percent changes in yearly profit from 2017 to 2018

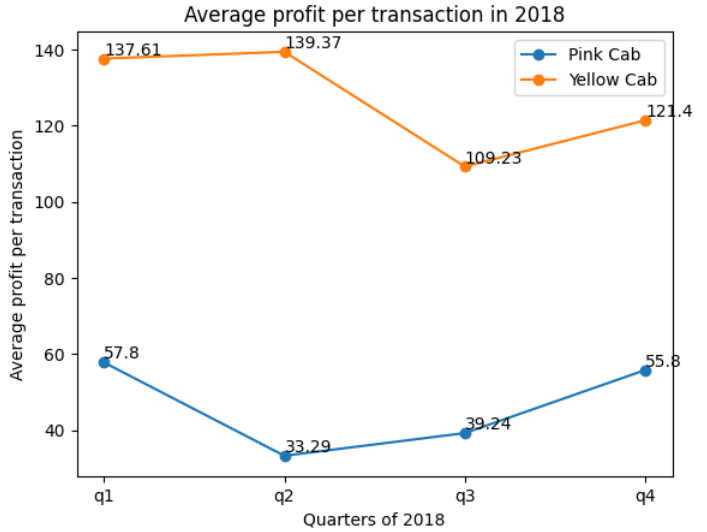


- Yellow Cab has better percent changes in yearly profits from 2017 to 2018 in most cities (10 out of 19)
- Although there was an overall decrease in yearly profits from 2017 to 2018 for both cab companies
- This is excluding San Francisco CA where there are no services for Pink Cab or Yellow Cab

Seasonality in average profit per transaction

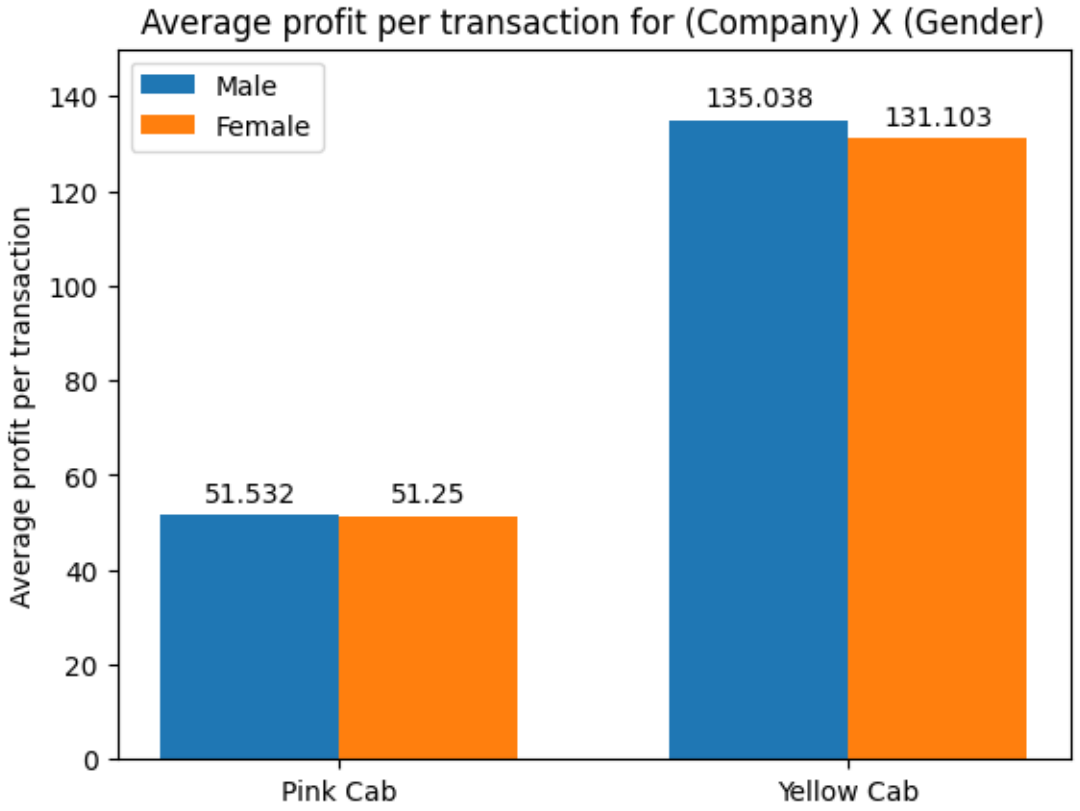






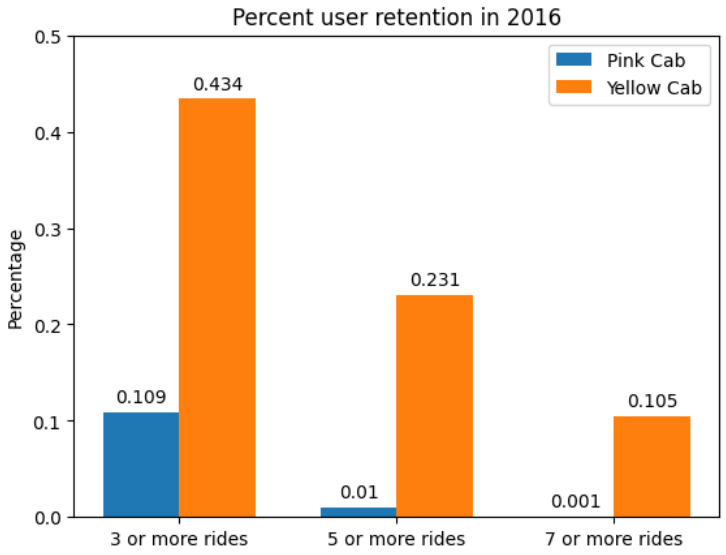
- This series of graphs shows yearly seasonality in the average profit per transaction
- There is significant difference in the average profit per transaction between Yellow Cab and Pink Cab in every quarter
- This is highly valuable information when combined with the fact that Yellow Cab also has significantly more transactions year-round

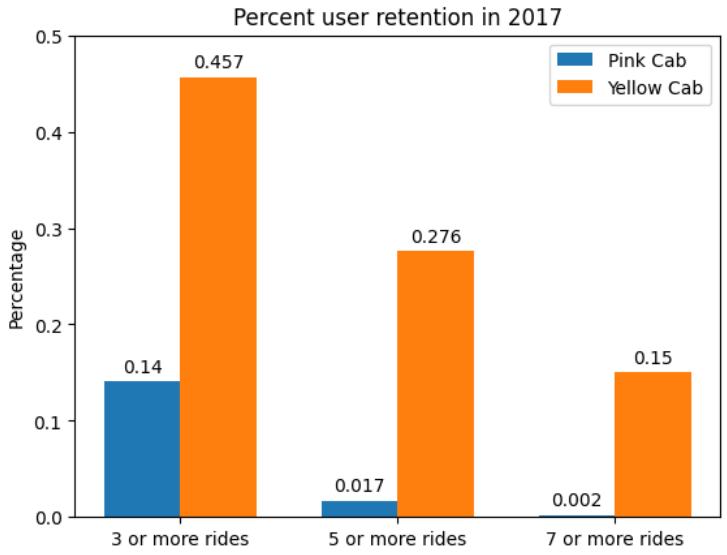
Gender's factor in average profit per transaction

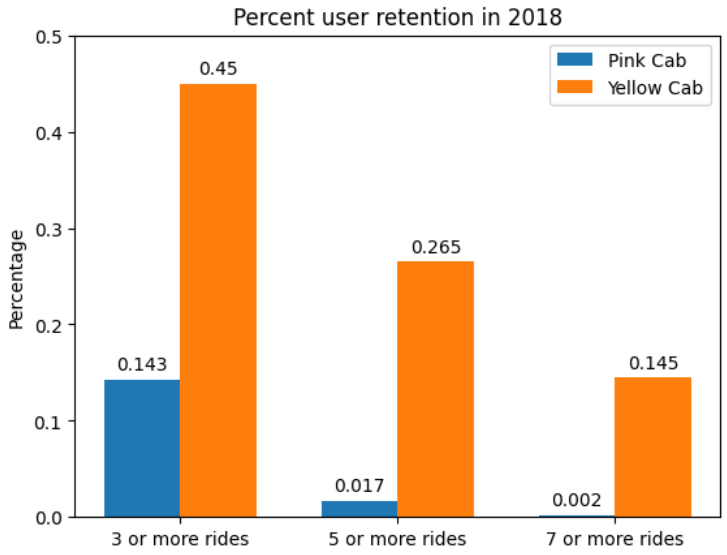


- Tests have been done to show that there is no significant difference in profit per transaction between genders for Pink Cab
- However, there is significant difference in profit per transaction between genders for Yellow Cab
- Still, segmenting by gender does not undermine Yellow Cab's dominance in any combination

User Retention

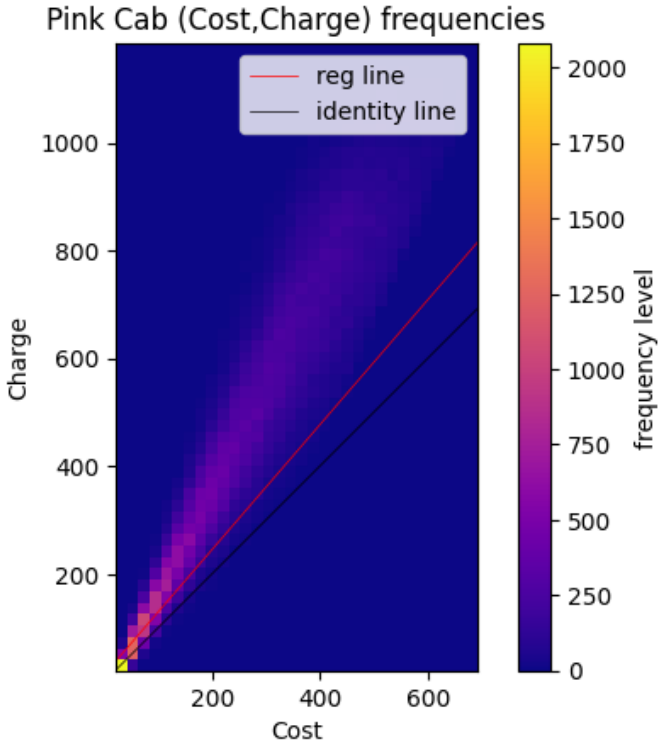


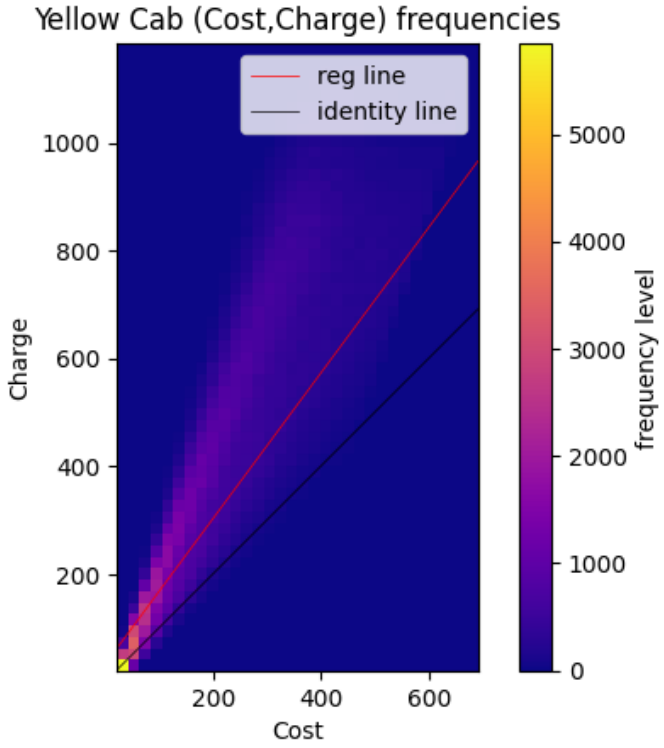




- It can be easily inspected that Yellow Cab has significantly higher percent user retention across all retention levels for all years
- When Pink Cab's percent user retention is close to zero at the (7 or more rides) retention level, Yellows Cab's percent user retention is greater than .1 and is at least ~84 times larger

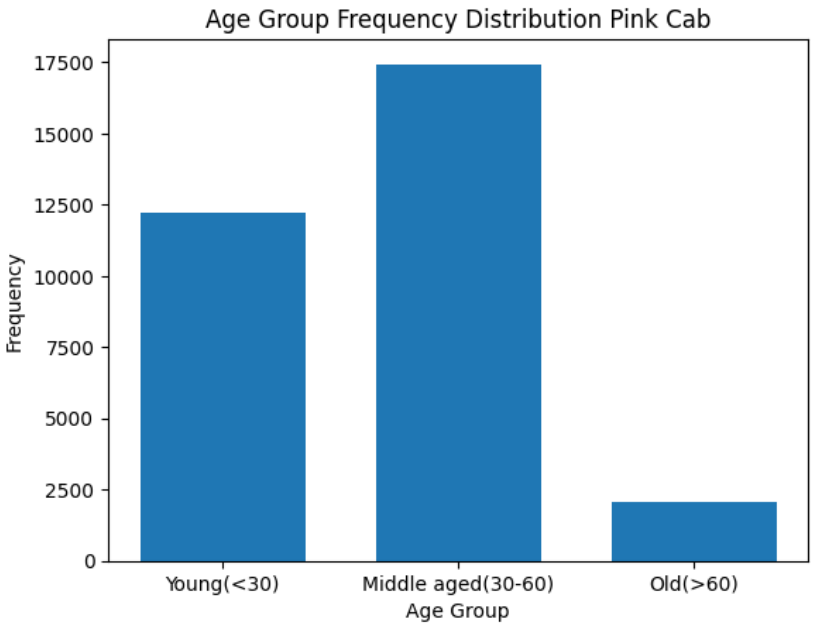
Cost and Charge relationship

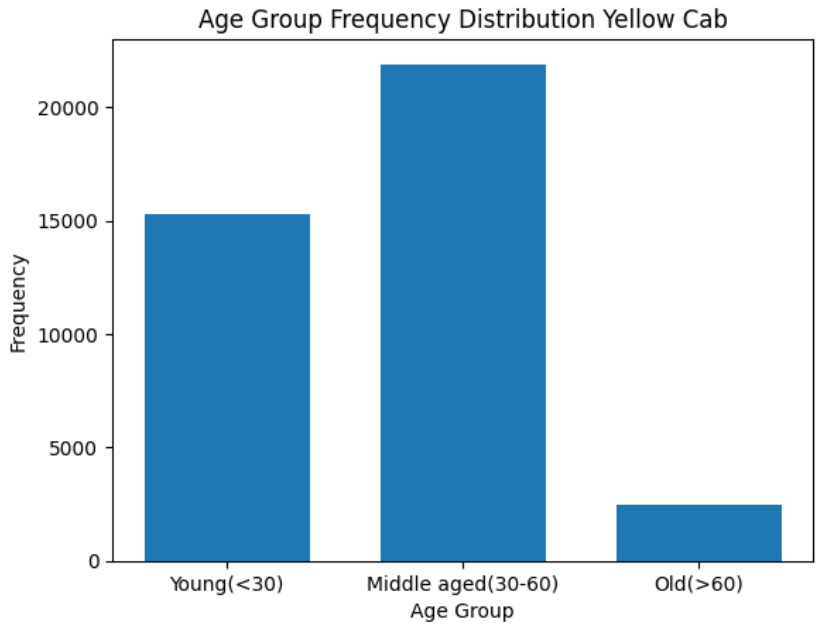




- As costs increase, charges increase linearly for both Yellow Cab and Pink Cab, each modeled by its respective regression line
- The identity line gives the intuition that the regression lines represent net profit for each cab company
- Yellow Cab's regression line has a higher y intercept and a higher slope than Pink Cab's regression line
- This implies that it is making more profit per transaction and the higher the cost the more profit it makes compared to Pink Cab

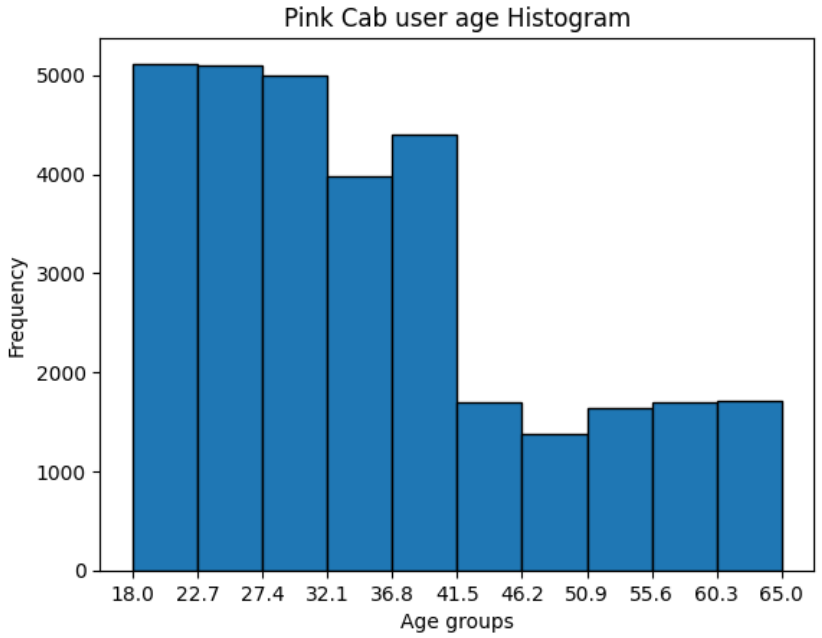
Trends in age groups between companies

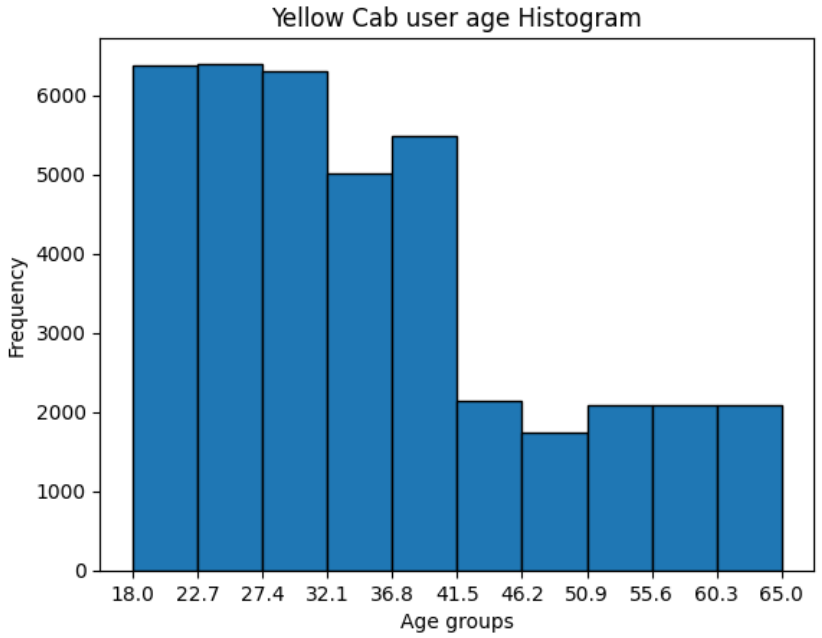




- Yellow Cab has higher number of transactions for each age group
- Both Pink Cab and Yellow Cab follow a trend between age categories
- Specifically, multiplicative scaling from young to middle aged is relatively similar between cab companies
- As is the scaling between all other age group pairs

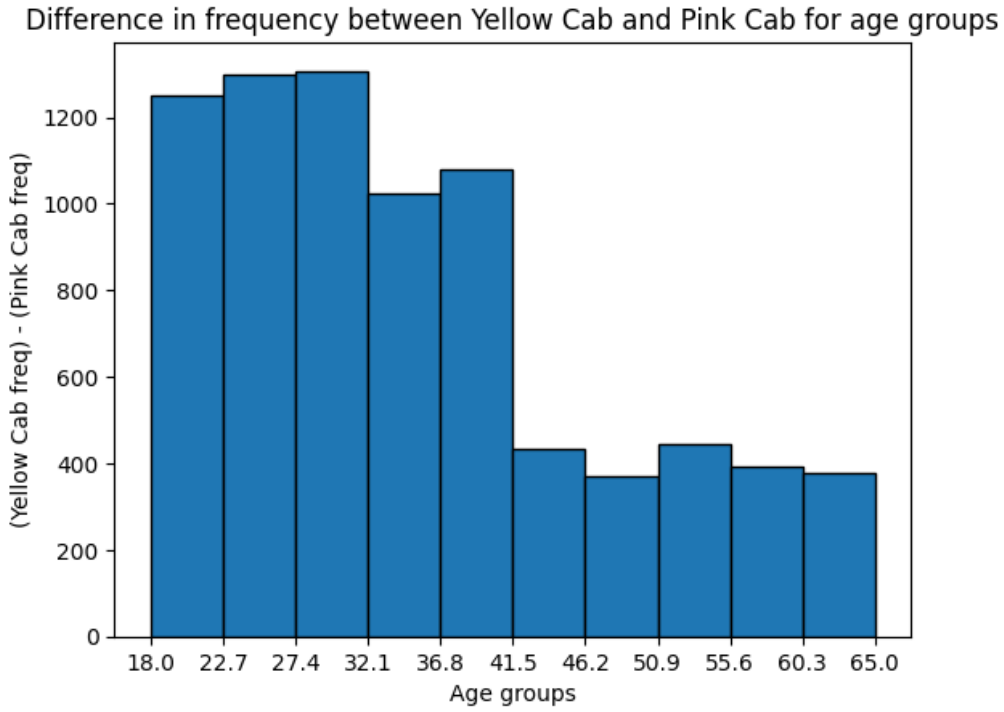
More age groups





- As seen in the previous case with wider age groups, the same kind of scaling persists between a large amount of equally sized age bins

Difference in age group frequencies between cab companies



- This is the difference between the user frequencies of Yellow Cab and Pink Cab
- The age group frequency difference scales with the frequency of the corresponding age groups for Both Yellow Cab and Pink Cab
- This means that Yellow Cab is not being overperformed in any age group and there is no standout weakness
- Yellow Cab successfully caters to customers of all age groups

Summary and Recommendation

- The 2016 to 2018 yearly profits percentage increase is higher for Yellow Cab than it is for Pink Cab
- For every ten days over the course of the 3 years Yellow Cab made more profits than Pink Cab
- Yellow Cab has better percent change in yearly profits in most cities between 2017 and 2018
- Yellow Cab has higher profit per transaction across all data points when segmented by both gender and quarters for each year. This is powerful knowledge combined with the fact that Yellow Cab has more transactions year-round than Pink Cab
- Yellow Cab has higher percentage user retention across all years and retention levels
- Yellow Cab has a superior linear relationship between cost and charge yielding more profits than Pink Cab as the cost per transaction increases
- Across all age groups there are more users for Yellow Cab than Pink Cab. The age group frequency difference scales with the frequency of the corresponding age groups for Both Yellow and Pink Cab. There is no significant exception to this pattern that favors Pink Cab

Based on the above insights, I recommend Yellow Cab over Pink Cab for investment.

Thank You