Data Glacier

Your Deep Learning Partner

G2M Case Study

LISUM 15

Nov 16th 2022

Guangyan Huang
guangyan_huang@ucsb.edu

Agenda

Business Problem Background

Data Properties and Assumptions

Hypothesis Summary

EDA and EDA Summary

Investment Recommendations

Background: G2M insight for Cab Investment firm

- 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.

- There are four datasets provided for performing market analysis:
- **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

- Each dataset includes information of two companies in the Cab Industry: Pink Cab and Yellow Cab.

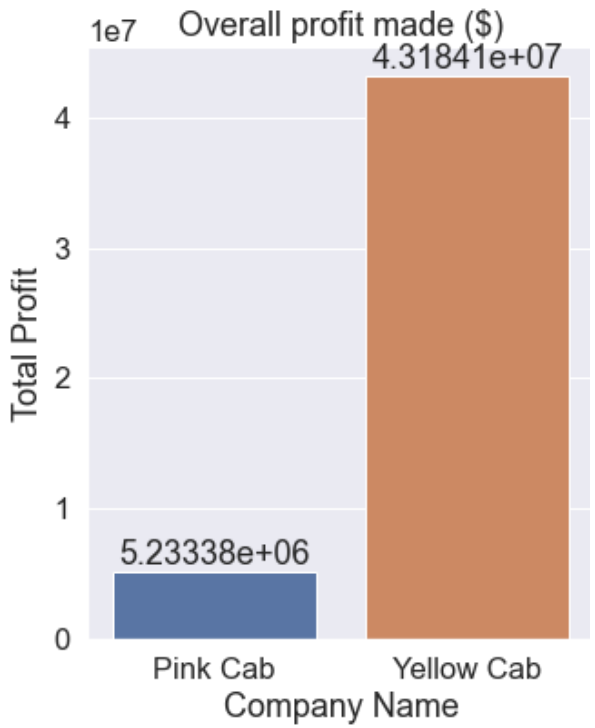
Properties of data provided

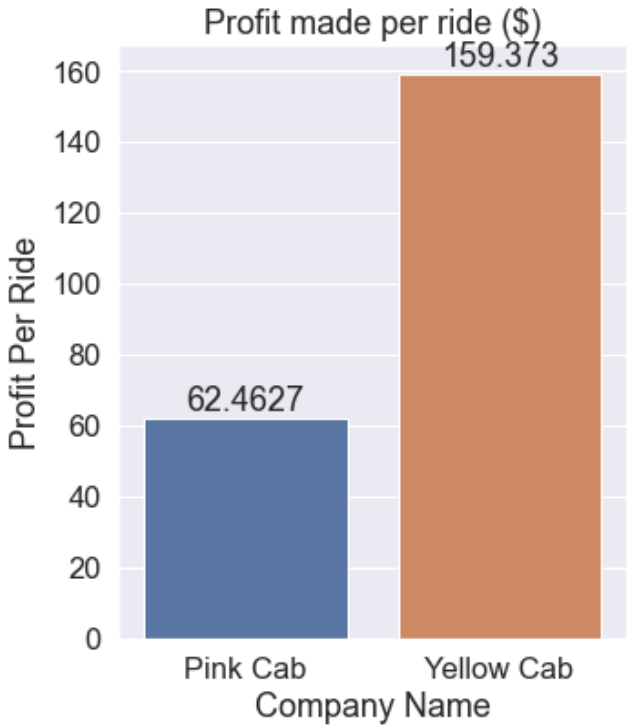
- Initially the dataset Cab_Data includes all transactions ranging from 2016/1/1 to 2018/12/31, after data cleaning, there are 354746 data points representing transactions took place between 2016/1/31 and 2018/12/31.
- Assumptions:
- No data point is removed given the transaction falls within the target time interval and no feature of the data point is null.
- When converting ‘date of transaction’, given 1095 distinct values, the minimum value corresponds to the date 2016/1/1 and the maximum value corresponds to the date 2018/12/31.

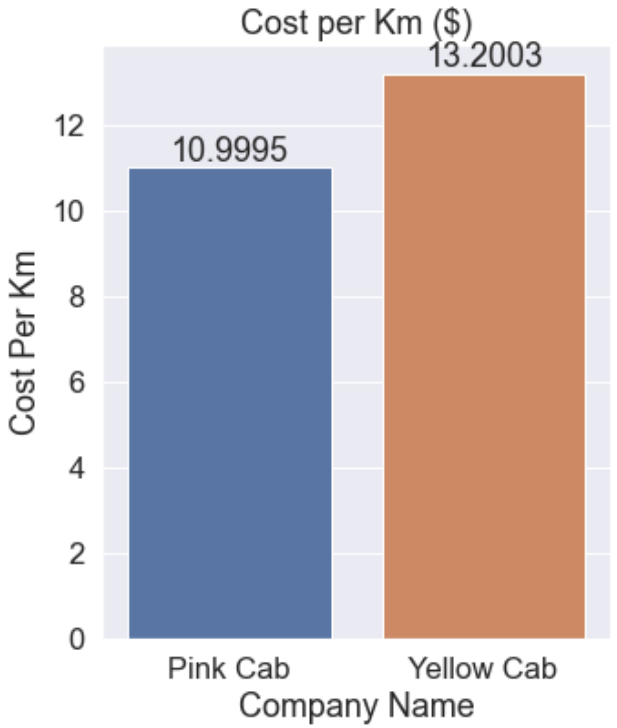
Hypothesis summary

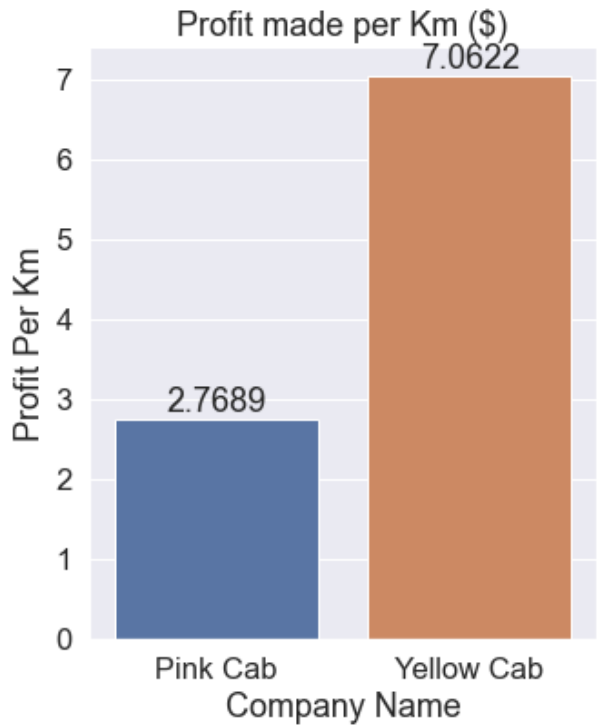
- Hypothesis 1: Is there any difference between the two cab companies, including the profit made annually, the average profit made from each ride, each Km covered, and the cost-profit efficiency (Given fixed cost, how much profit can be made)?
- Hypothesis 2: Does the profit made by the two cab companies differentiate in its distribution with respect different years, how about the profit distribution with respect to different time intervals in a year?
- Hypothesis 3: Is there any pattern for the spatial profit distribution of the two companies (profit V.S city)?
- Hypothesis 4: Is there other aspect affecting the profit making of the two companies? For example, payment mode, gender, age and income level of customers, and is there any difference between characteristics of customers for both companies?
- Hypothesis 5: Is there any difference between the capacity of the two companies in developing long term customers?

Profit and Cost Analysis (All unit as $)





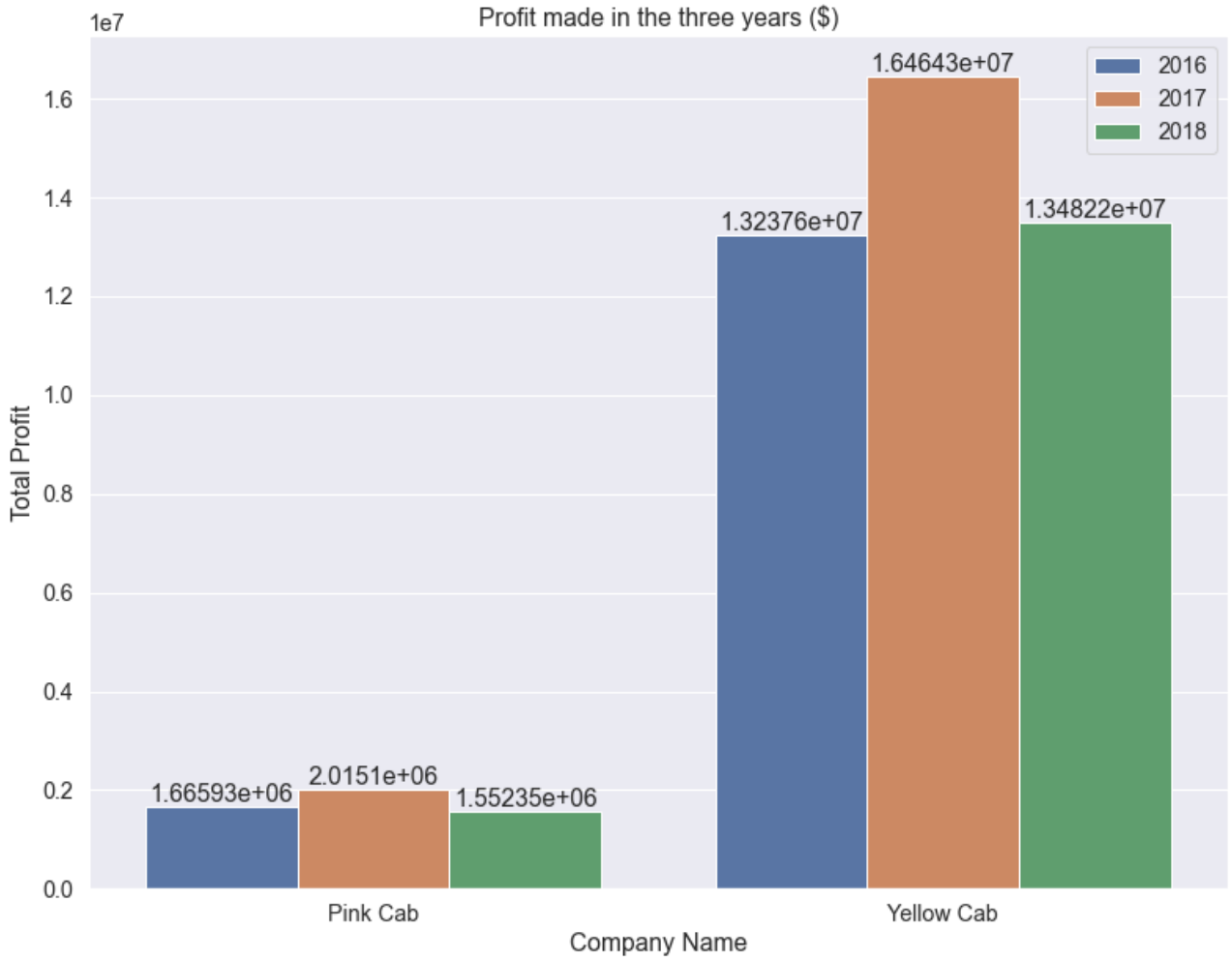




Company Name	Total Profit	Profit Per Ride	Profit Per Km	Cost Per Km	Cost per ride	Profit made per 100$ of cost
Pink Cab	5.233375e+06	62.462707	2.768902	10.999493	248.133762	25.172998
Yellow Cab	4.318407e+07	159.373141	7.062204	13.200348	297.892961	53.500137

The magnitude of business for Yellow Cab outweighs Pink Cab, given the profit ratio is approximately 8:1. Also, given the profit/ride ratio is approximately 2.5/1, the implication is Yellow Cab has transaction time more than three times of Pink Cab's transaction time. One thing to point out is, Pink Cab seems to be more efficient given less cost per km of each ride.

Analysis of Profit with spatial distribution (All unit as $)

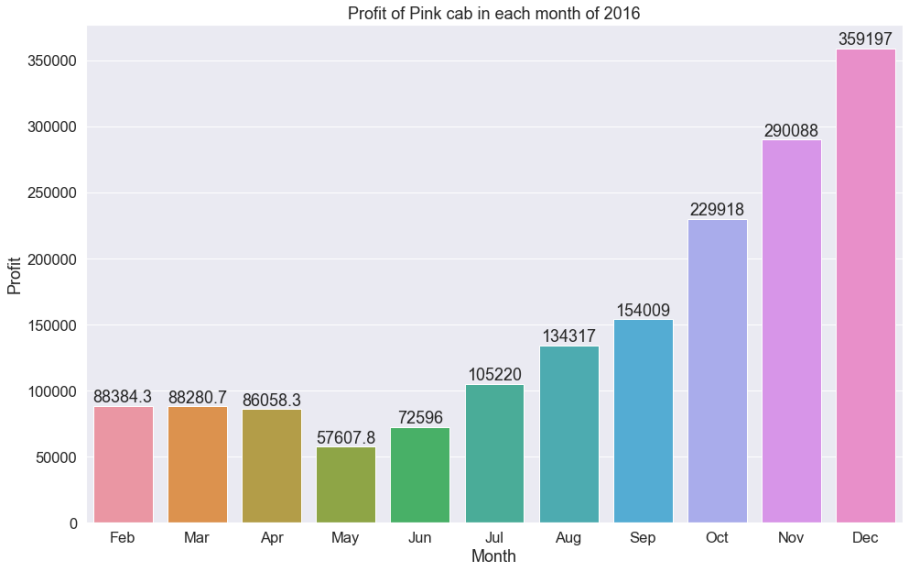


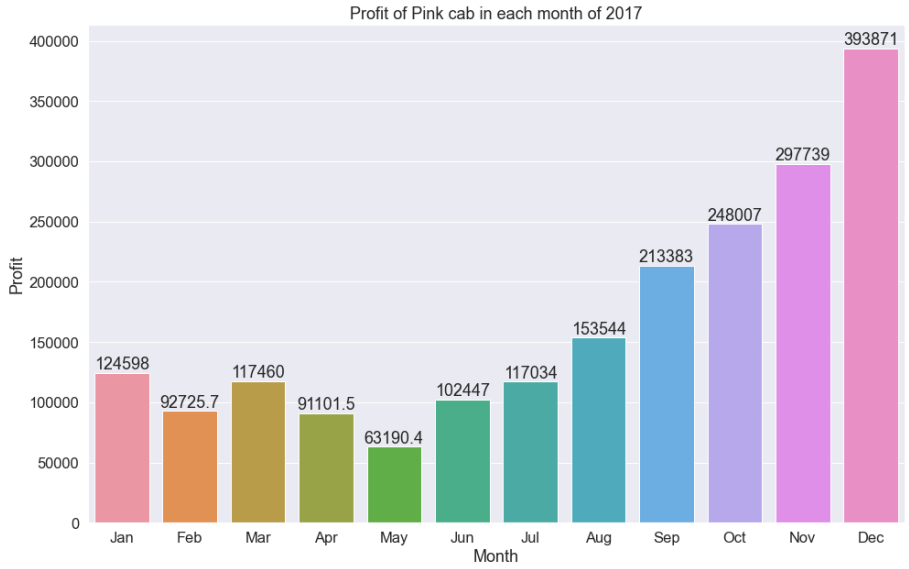
The distribution of profit V.S year for Pink Cab and Yellow Cab follow the similar pattern: increment in 2017 from 2016 and recession from 2018 to 2017.

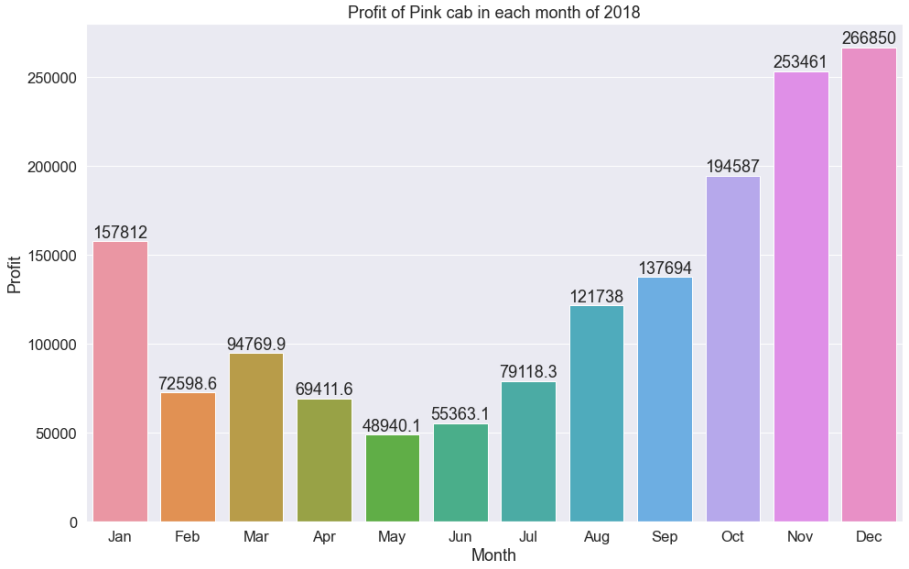
In 2017, Yellow Cab experienced an increment of 25% in Profit made while Pink Cab experienced an increment of 20%.

In 2018, Yellow Cab's annual profit decreased by 18%, while Pink Cab's annual profit decreased by 23%. This can be the outcome derived from the fact that smaller firms are more prone to economic recession.

Profit distribution for different months (Pink Cab)

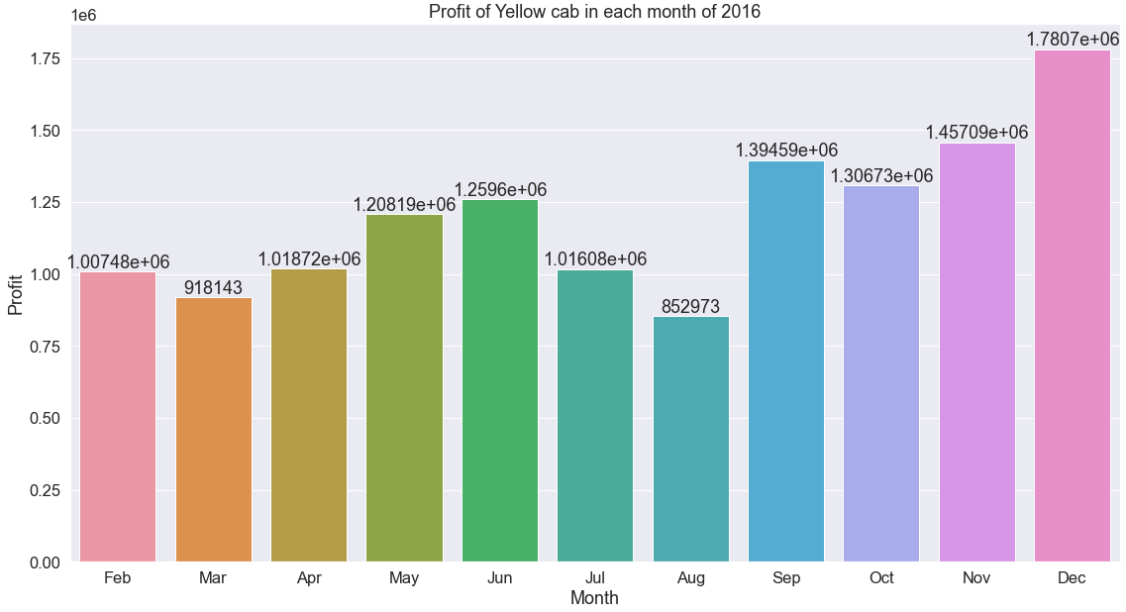


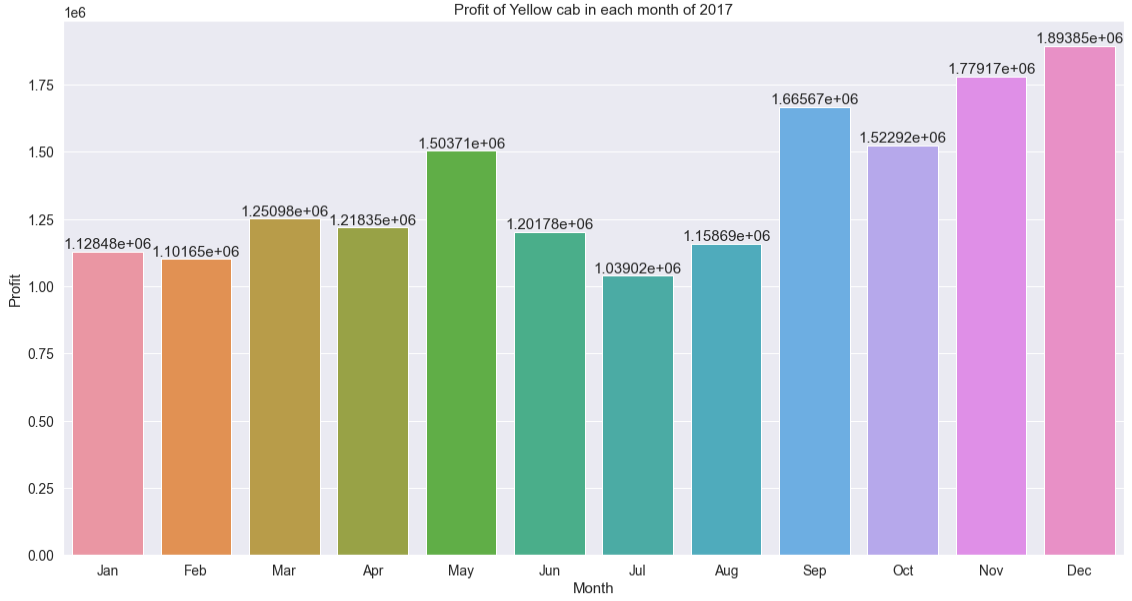


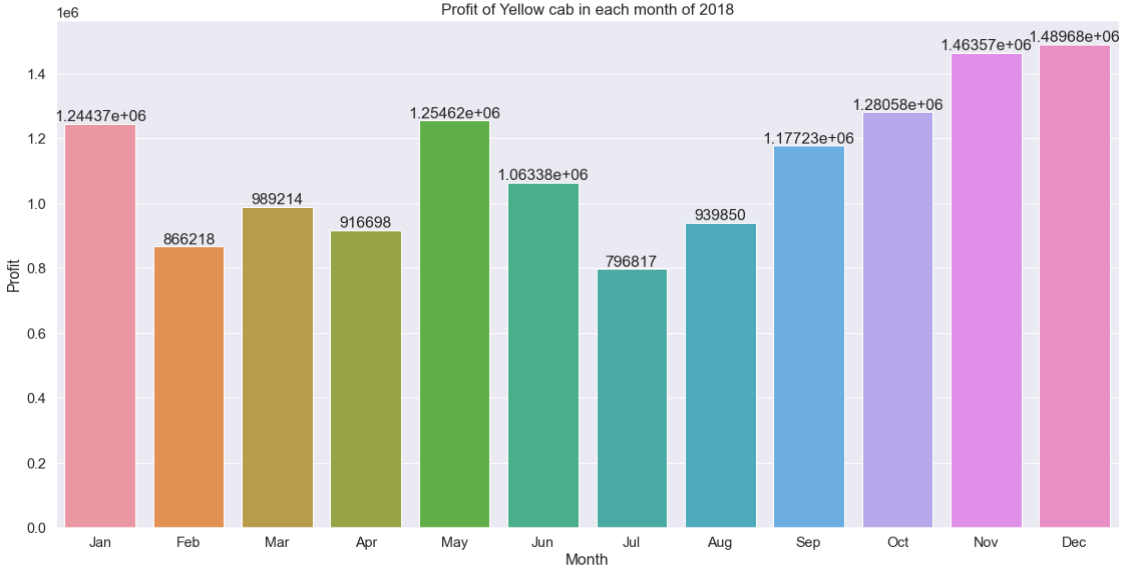


From the profit V.S month plot for each year, Pink Cab's profit is mainly generated in the last three months of each year.

Profit distribution for different months (Yellow Cab)

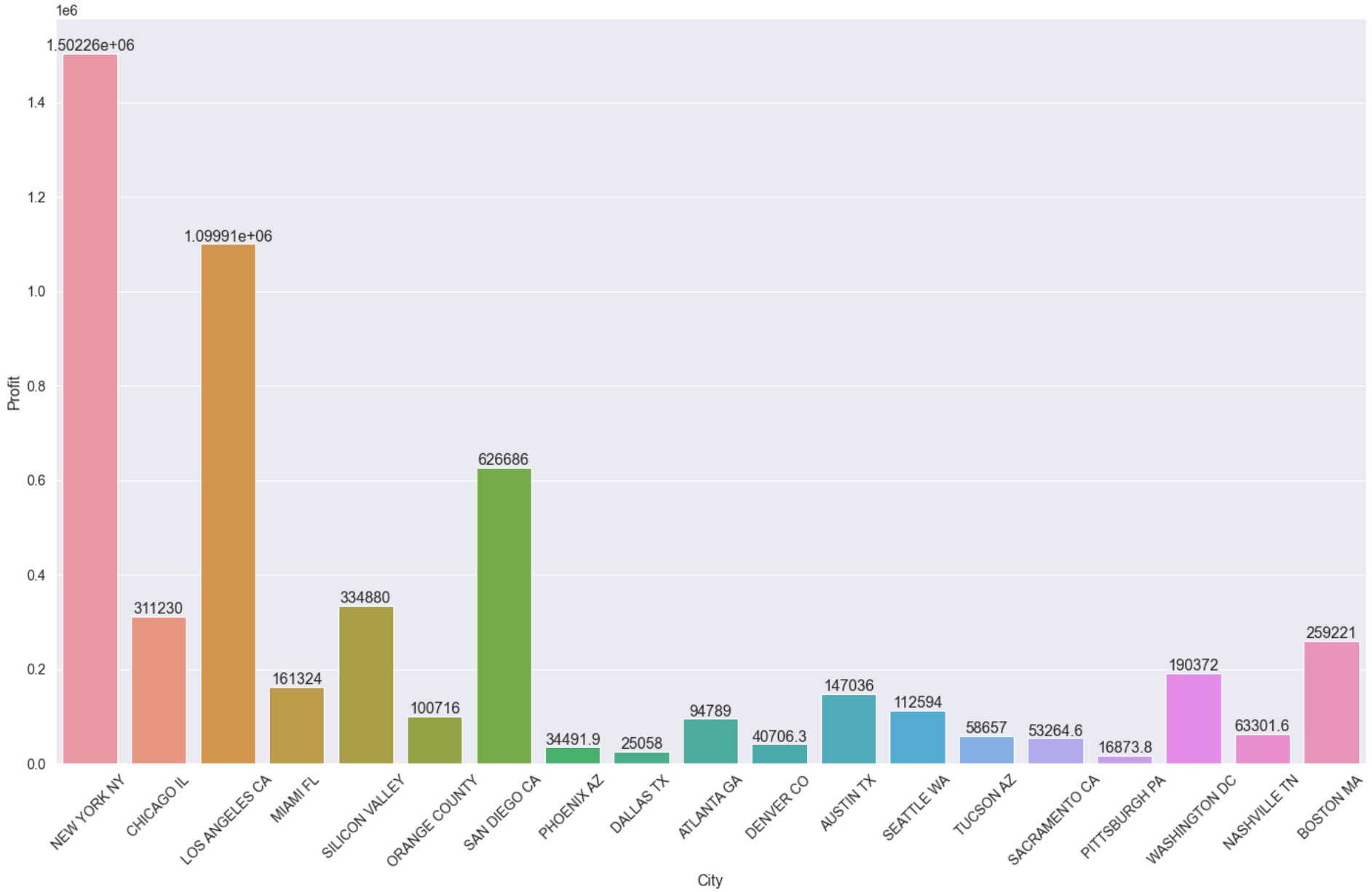




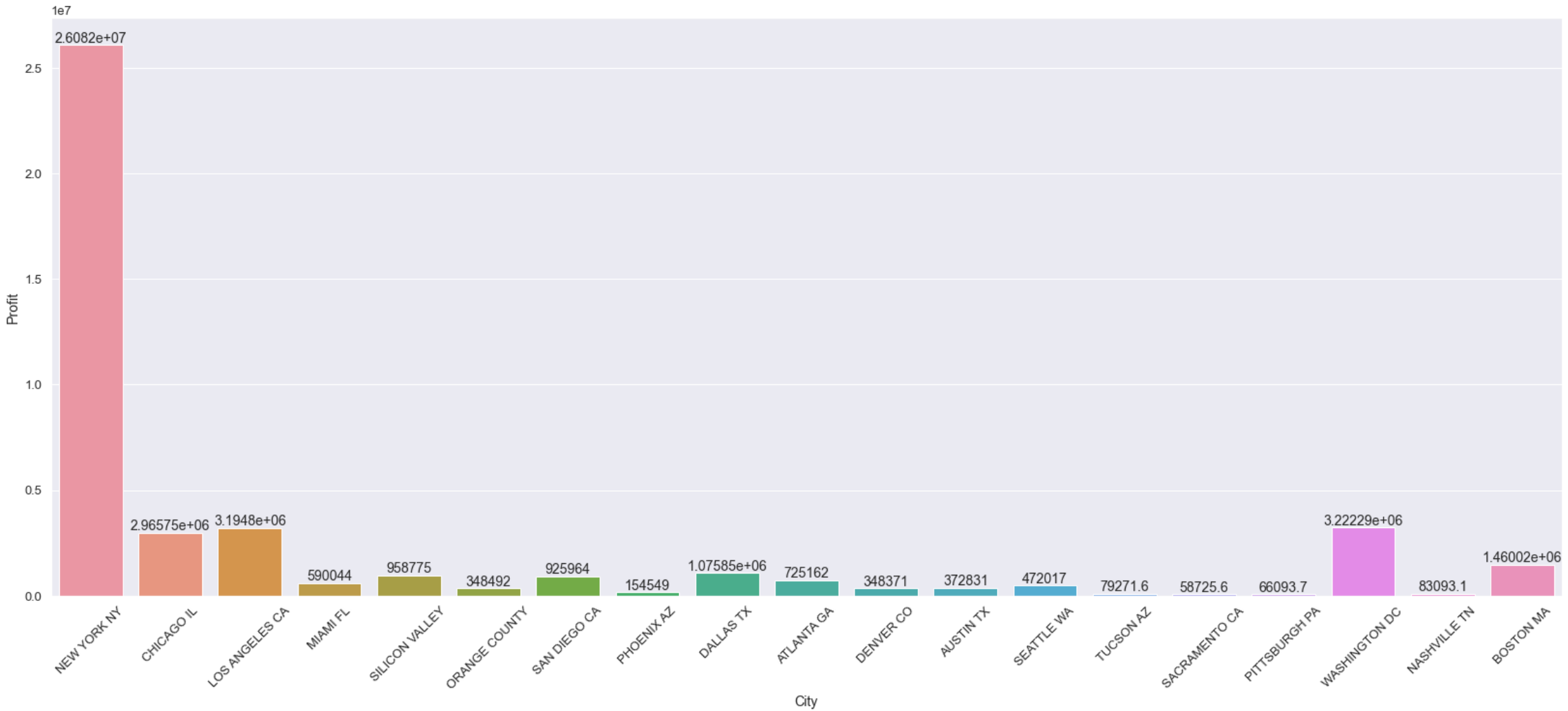


From the profit V.S month plot for each year, Yellow Cab's profit for each year is distributed in a more equal way. Nevertheless, the last three months of each year is still an outstanding profit source for Yellow Cab. One thing to point out is, comparing to Pink Cab with least profit in May, Yellow Cab's capacity of generating profit in May is much better.

Profit spatial distribution (Pink Cab)



Profit spatial distribution (Yellow Cab)

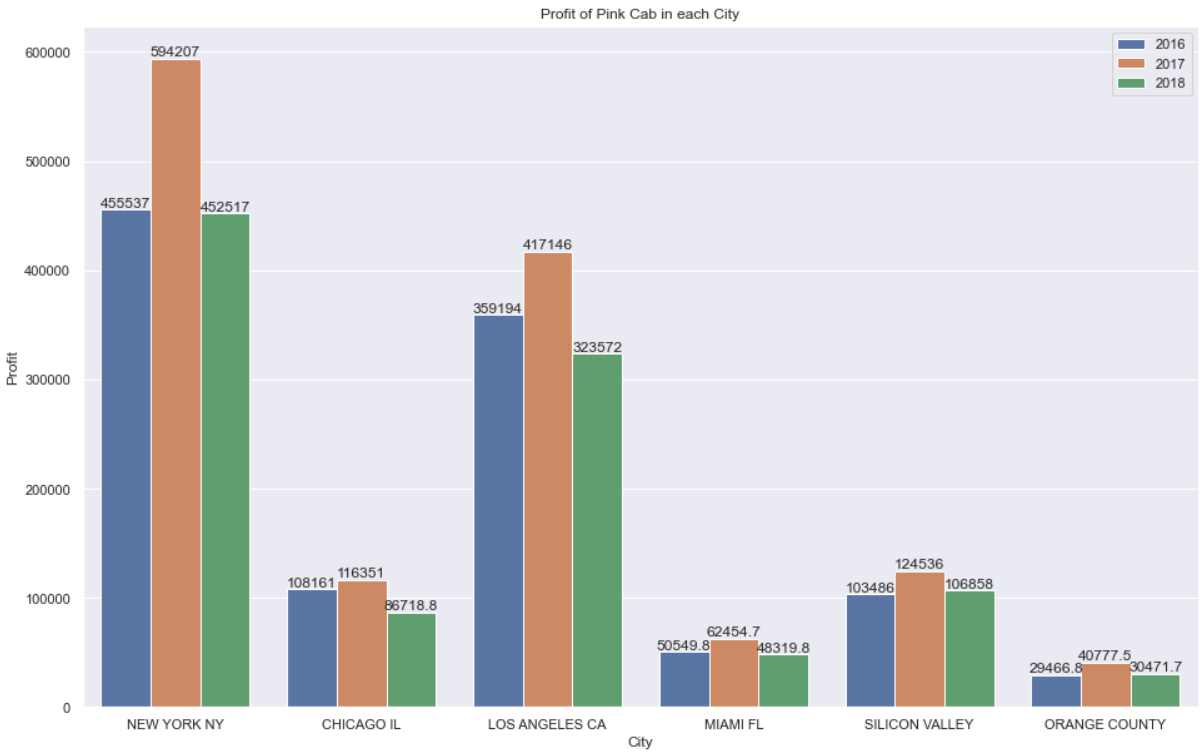


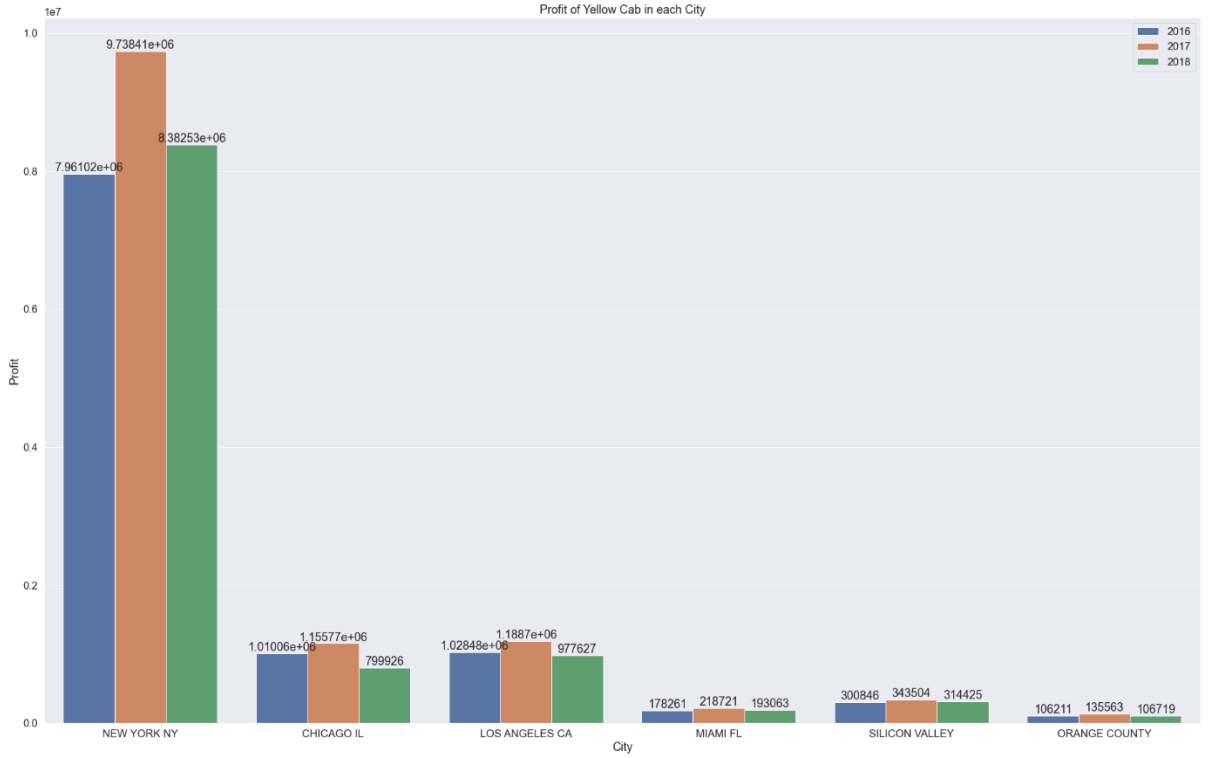
Profit spatial distribution

- From the observation above, the main profit source of both companies are New York. However, in addition to New York, Los Angeles as well as San Diego are important profit sources for Pink Cab as well, while Yellow Cab focuses exclusively on New York.
- One phenomenon to point out is, while overall profit ratio between Yellow Cab and Pink Cab is approximately 8:1, in some cities, the profit ratio is close to 3:1, or even 3:2. Surprisingly, the list of this city is Los Angeles, Silicon Valley, Orange County, San Diego and Sacramento, which all belong to California.

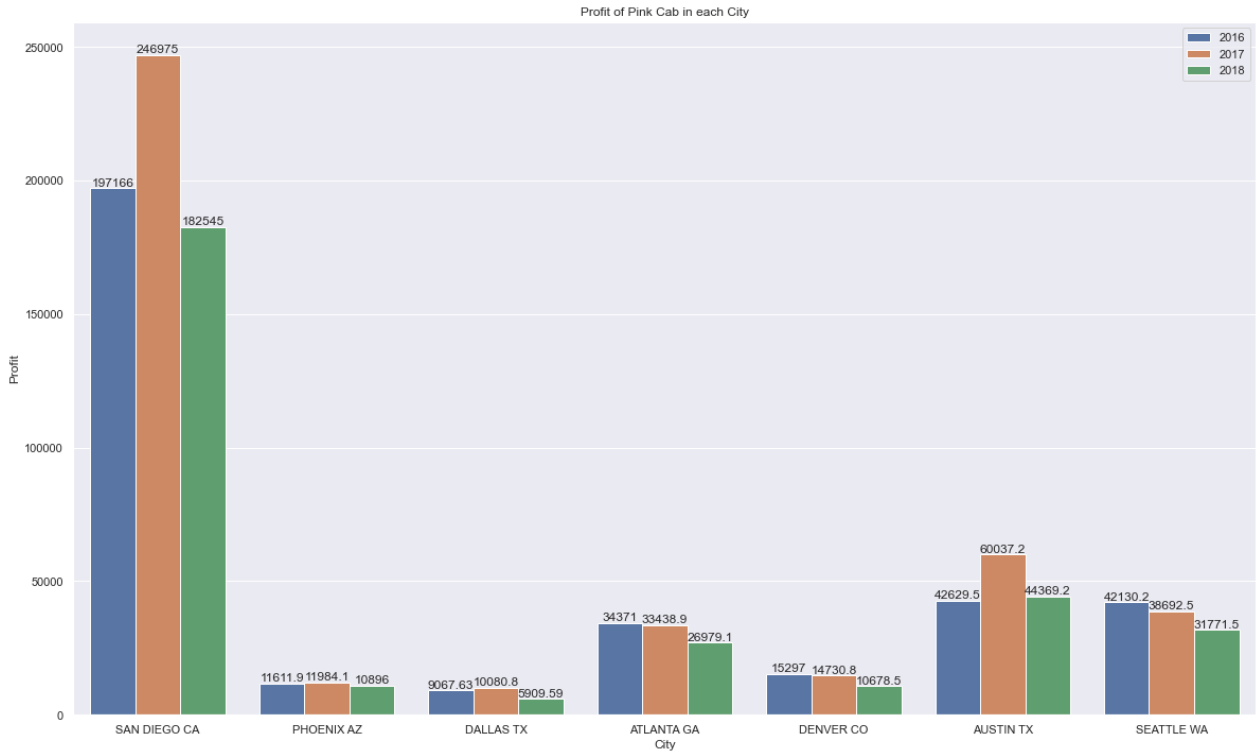
	City	Profit Pink Cab	Profit Yellow Cab	Yellow/Pink Profit Ratio
0	NEW YORK NY	1502261.393	2.608196e+07	17.361798
1	CHICAGO IL	311230.261	2.965753e+06	9.529128
2	LOS ANGELES CA	1099911.927	3.194800e+06	2.904596
3	MIAMI FL	161324.284	5.900443e+05	3.657504
4	SILICON VALLEY	334879.568	9.587753e+05	2.863045
5	ORANGE COUNTY	100716.008	3.484920e+05	3.460145
6	SAN DIEGO CA	626686.400	9.259642e+05	1.477556
7	PHOENIX AZ	34491.928	1.545495e+05	4.480744
8	DALLAS TX	25058.039	1.075847e+06	42.934222
9	ATLANTA GA	94788.955	7.251619e+05	7.650279
10	DENVER CO	40706.332	3.483707e+05	8.558144
11	AUSTIN TX	147035.904	3.728308e+05	2.535645
12	SEATTLE WA	112594.219	4.720171e+05	4.192197
13	TUCSON AZ	58656.980	7.927163e+04	1.351444
14	SACRAMENTO CA	53264.643	5.872564e+04	1.102526
15	PITTSBURGH PA	16873.783	6.609369e+04	3.916946
16	WASHINGTON DC	190372.377	3.222294e+06	16.926269
17	NASHVILLE TN	63301.566	8.309312e+04	1.312655
18	BOSTON MA	259220.891	1.460023e+06	5.632350

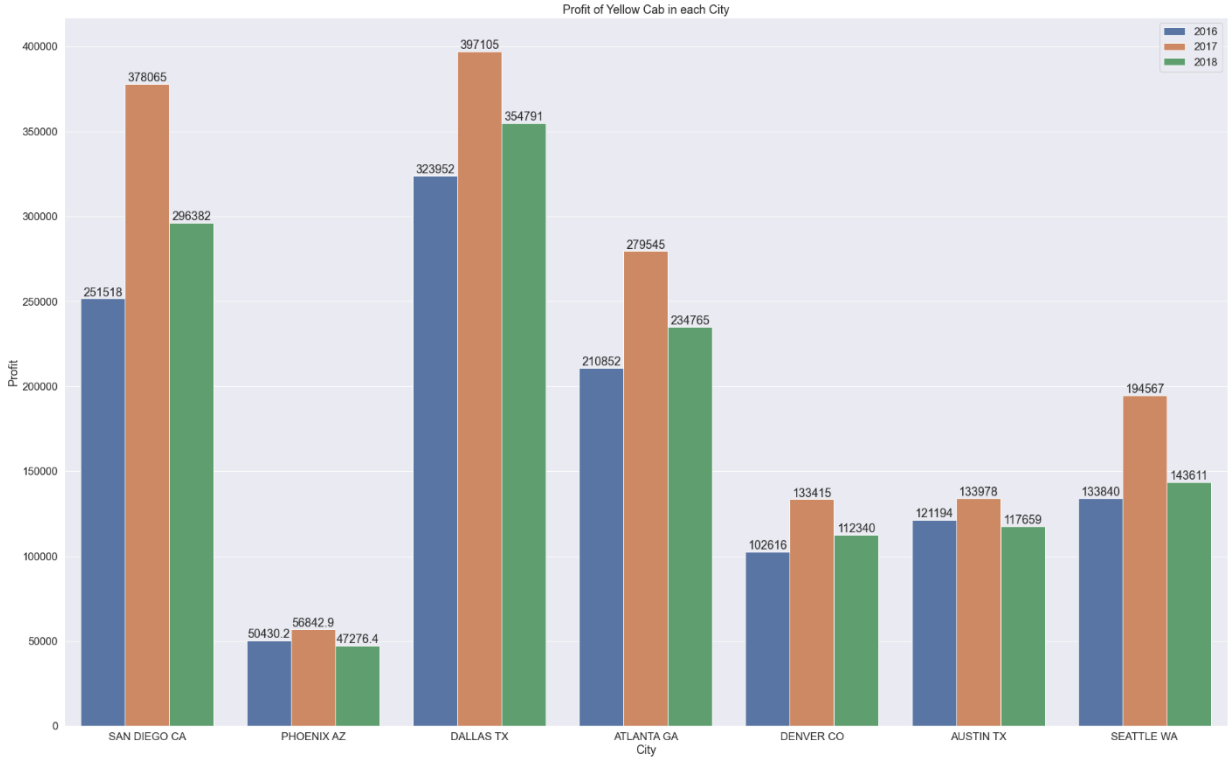
Profit spatial - time distribution for both companies



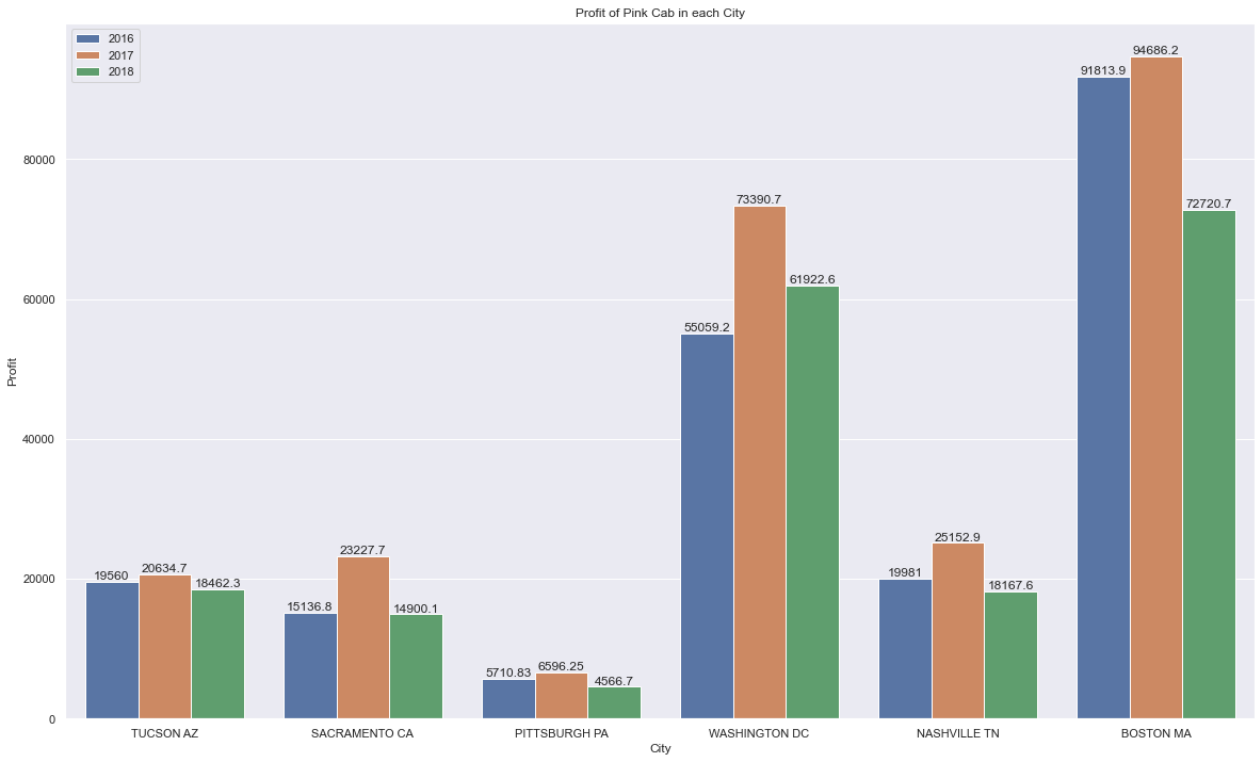


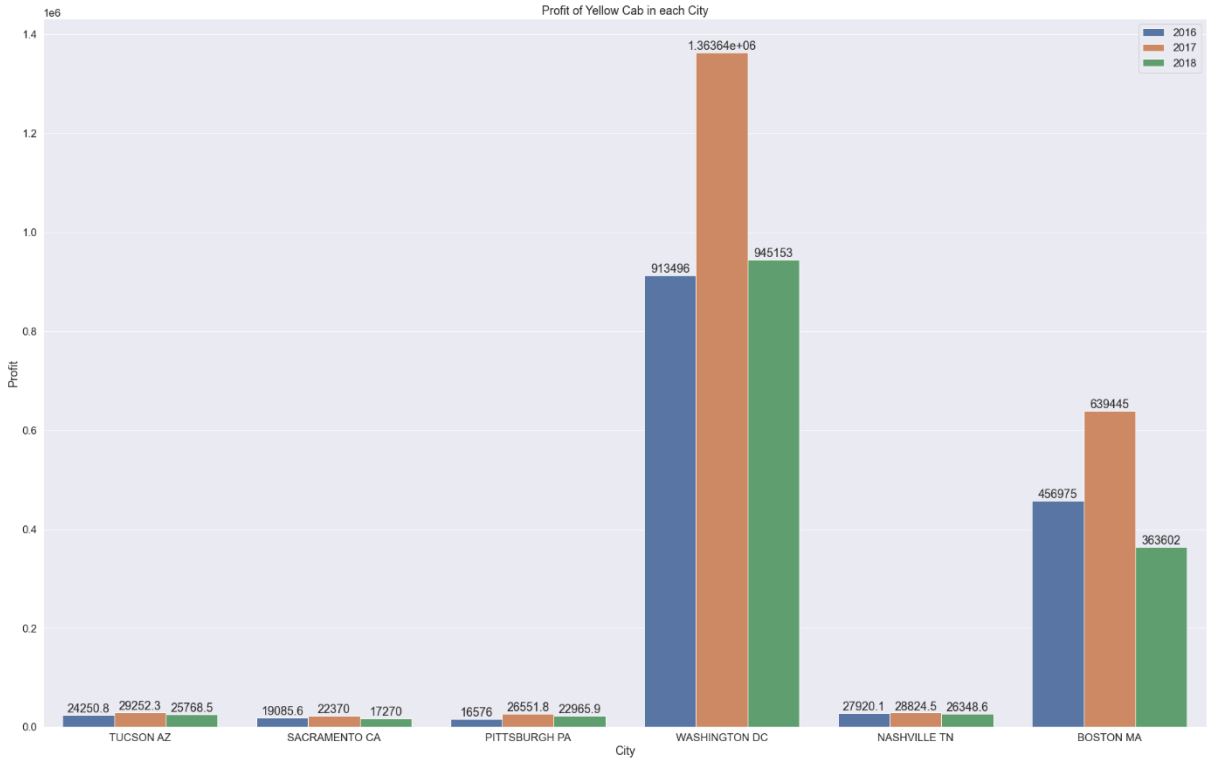
Profit spatial - time distribution for both companies





Profit spatial - time distribution for both companies

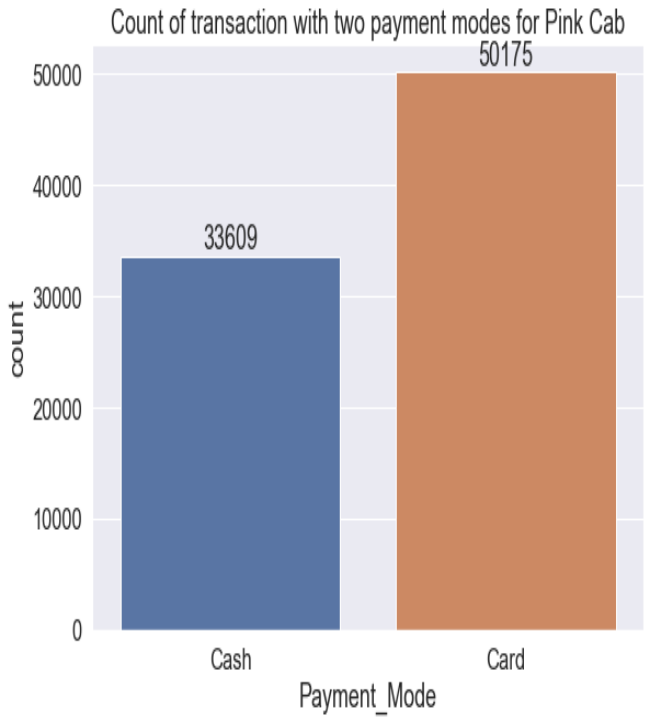


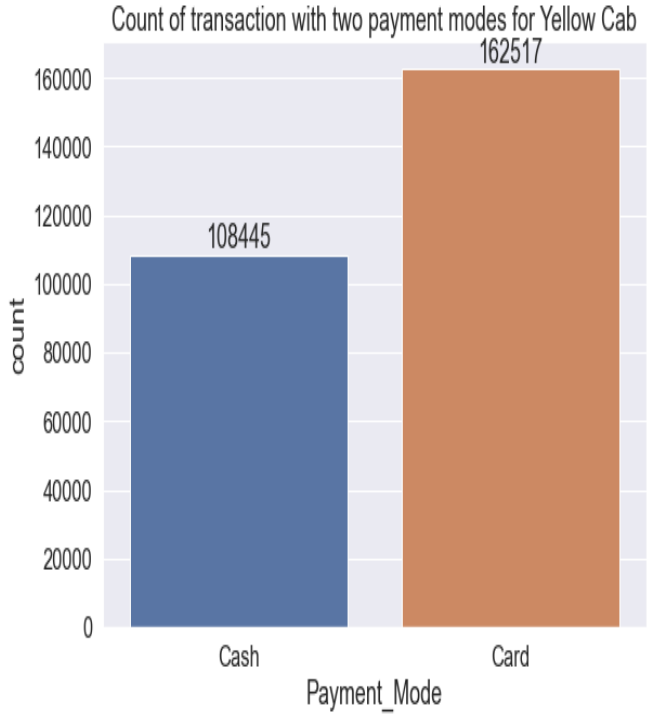


The yearly fluctuation of profit made in all 19 cities for both companies follow the same trend: increased in 2017 and decreased in 2018.

Analysis of other aspects affecting the profit making of the two companies

- Analysis of payment mode’s roll on profit making

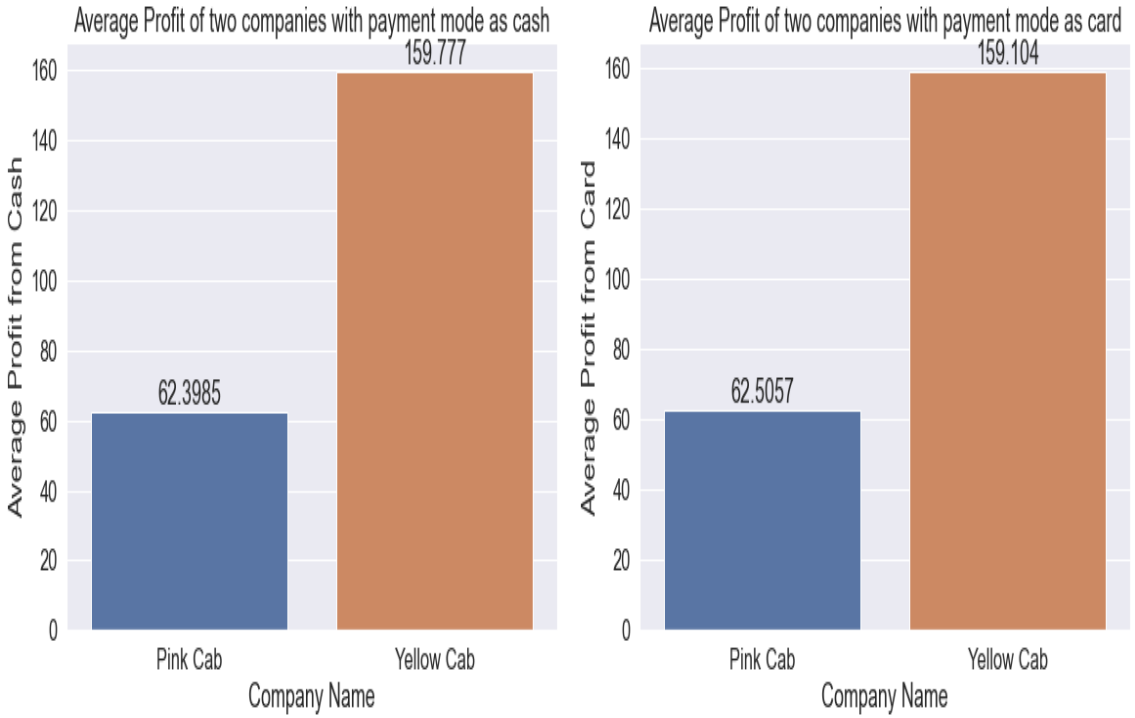




The Cash : Card ratio of customers for both companies are all close to 2 : 3.

Analysis of other aspects affecting the profit making of the two companies

- Analysis of payment mode's roll on profit making

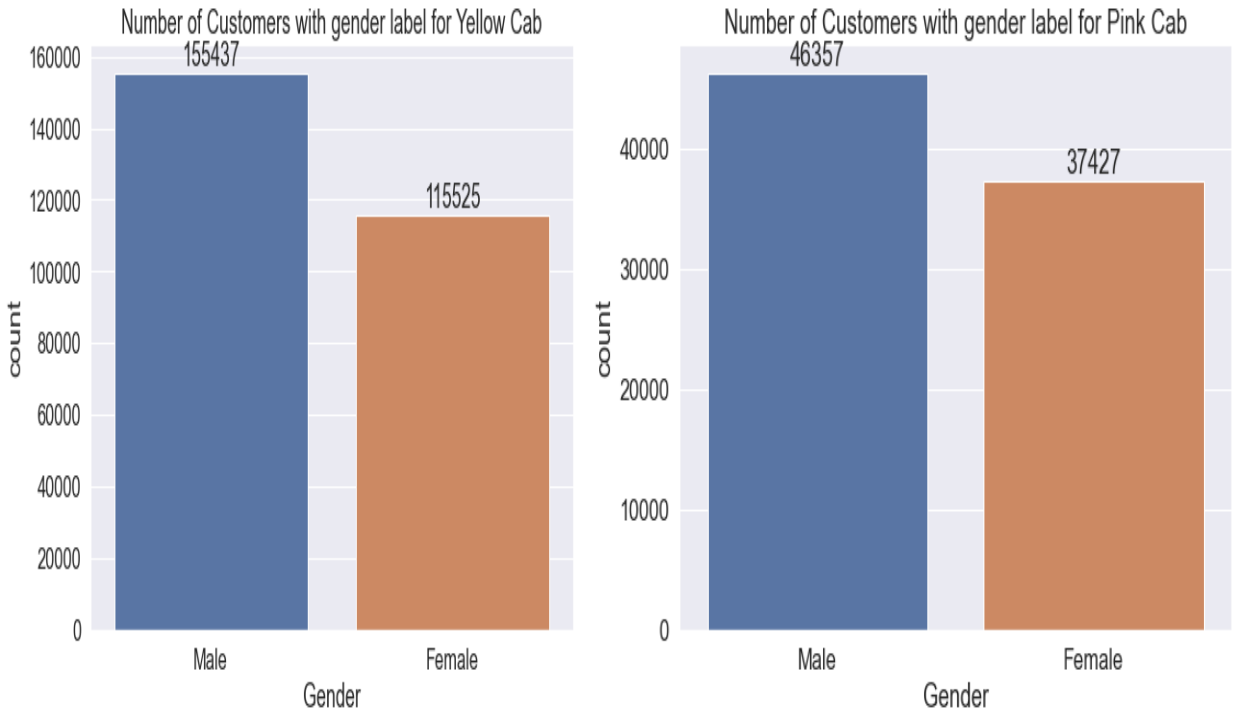


There is no significant difference between average profit per ride V.S payment mode for both companies.

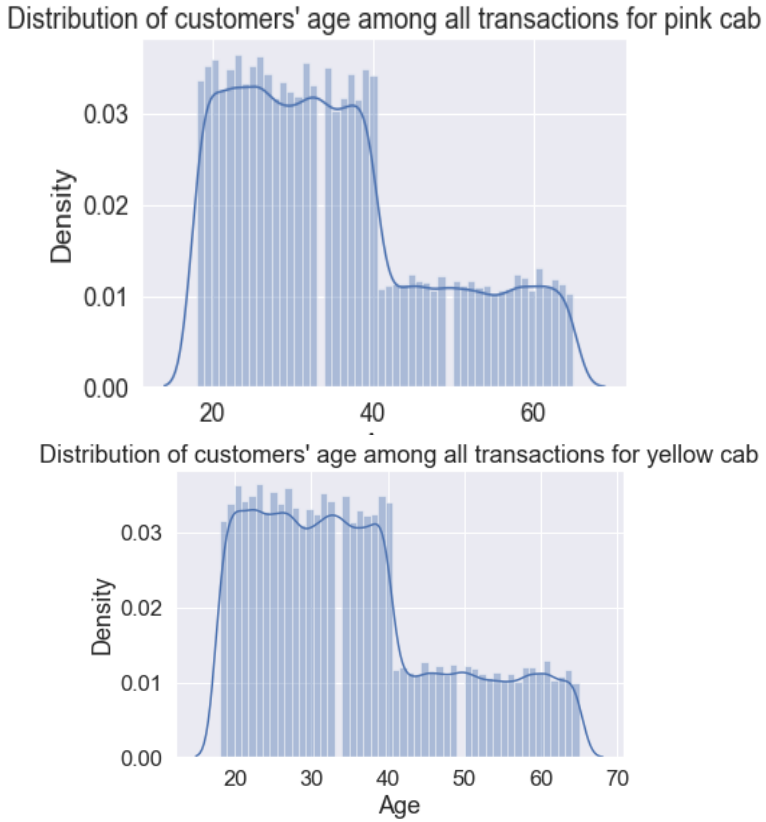
The ratio of profit made by cash V.S profit made by card for both companies are nearly 1 : 8.

Analysis of other aspects affecting the profit making of the two companies

- Analysis of Gender, Age's roll on profit making



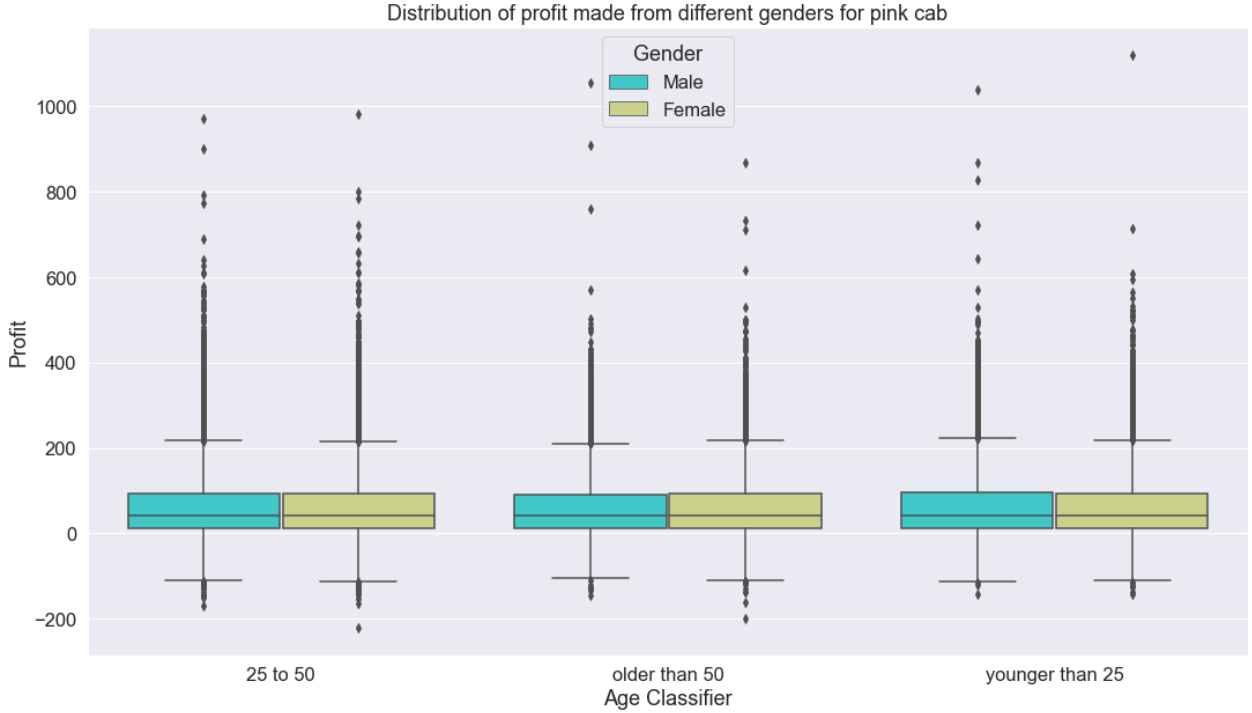
The Male : Female ratio of customers for both companies are all close to 57 : 43.

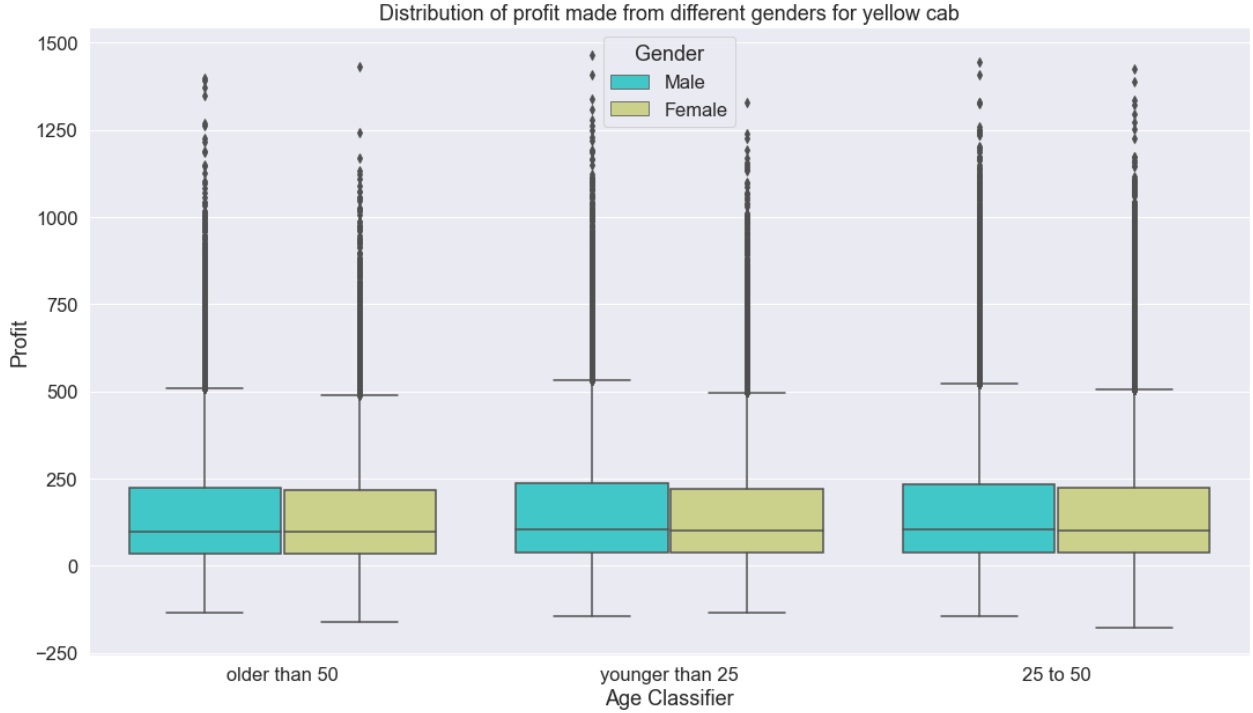


Among all transactions, the distribution of customers' age for two companies have no significant difference.

Analysis of other aspects affecting the profit making of the two companies

- Analysis of Gender, Age's roll on profit making

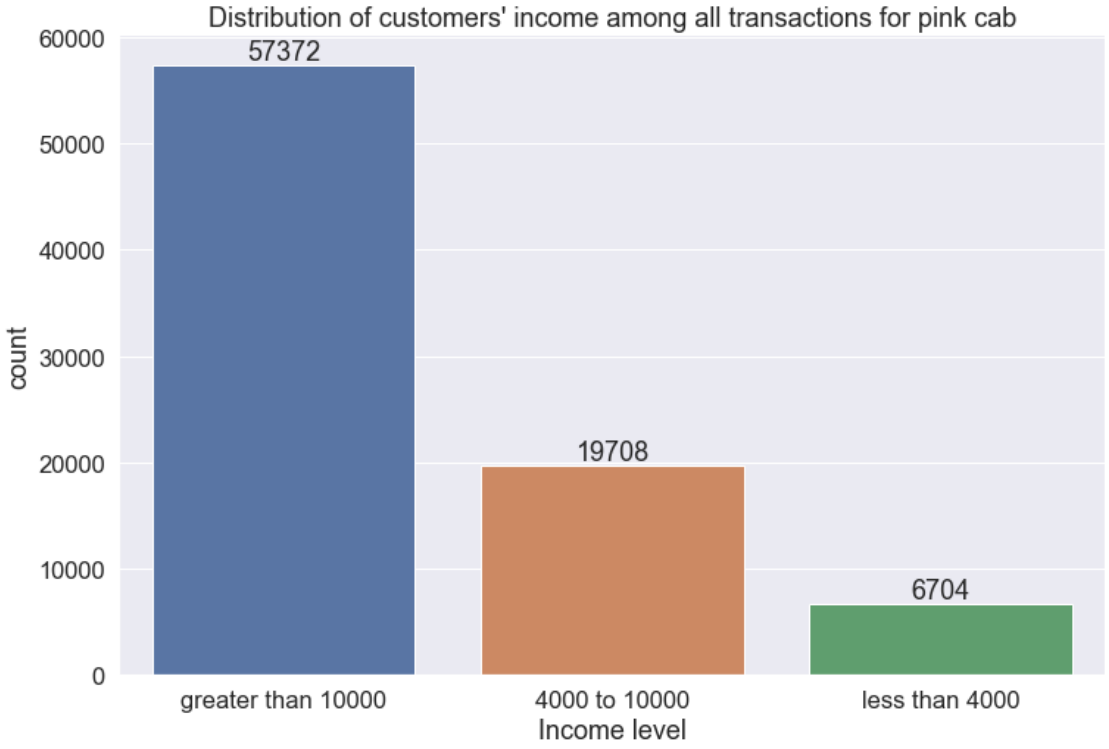


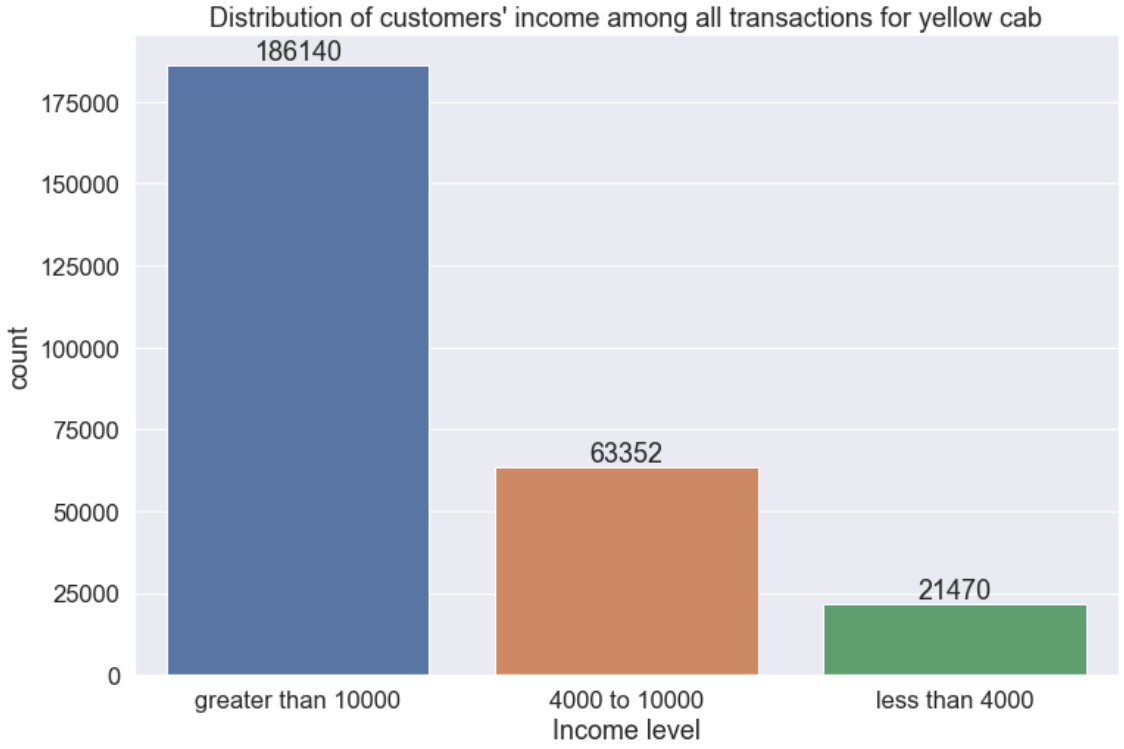


For Pink Cab and Yellow Cab, gender and age are not significant factors affecting profit per ride.

Analysis of other aspects affecting the profit making of the two companies

- Analysis of Income level's roll on profit making

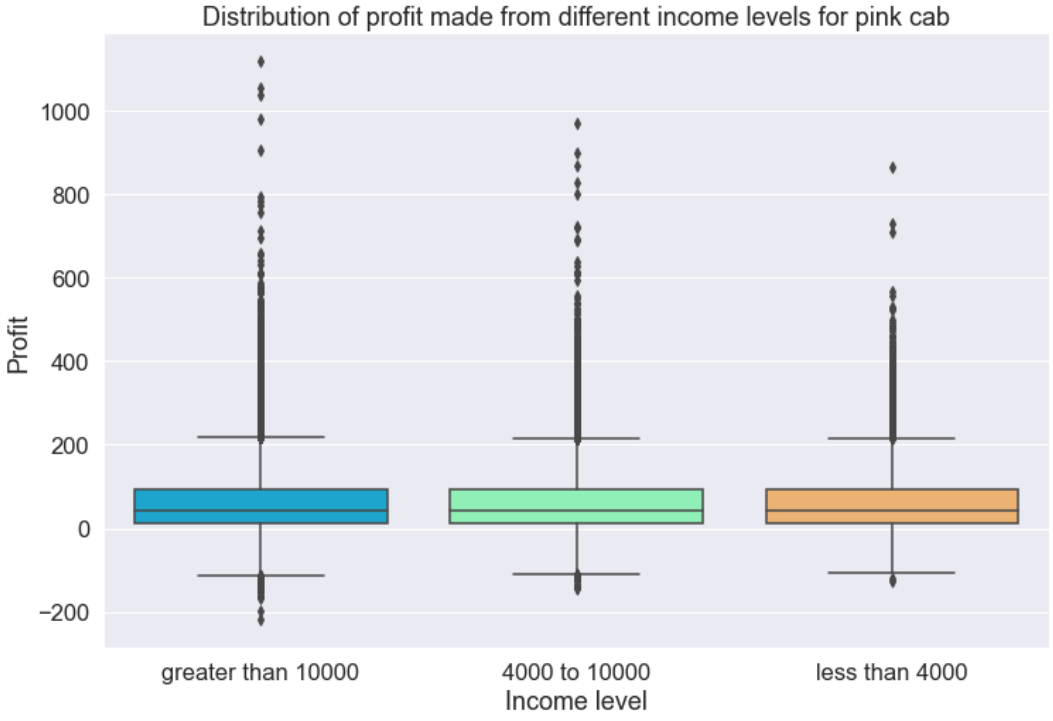


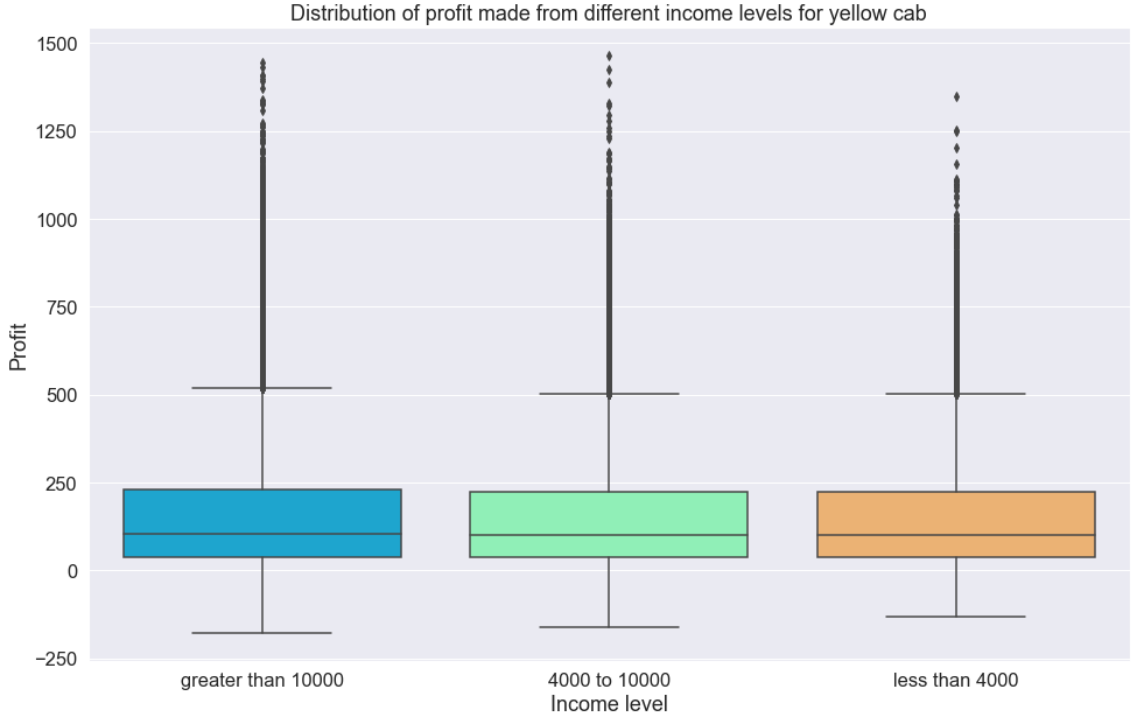


For Pink Cab and Yellow Cab, the distribution of customer's income level for transactions have the same pattern too: the ratio of above 10000, 4000-10000 and below 4000 is 9:3:1 for both companies.

Analysis of other aspects affecting the profit making of the two companies

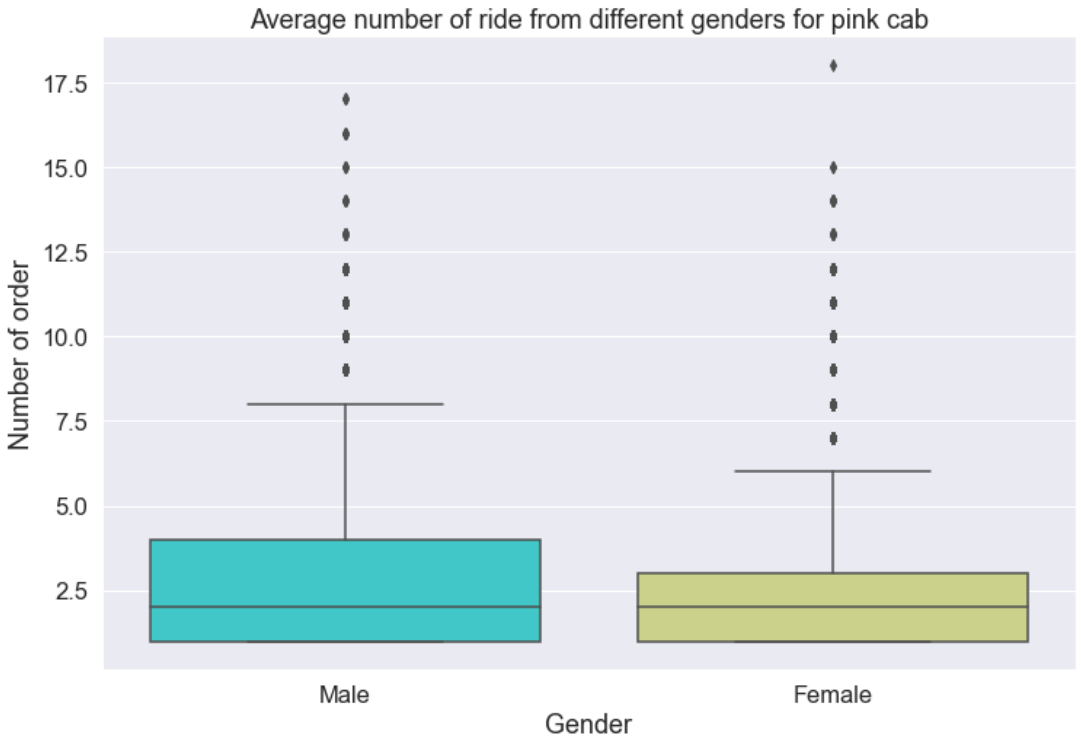
- Analysis of Income level's roll on profit making

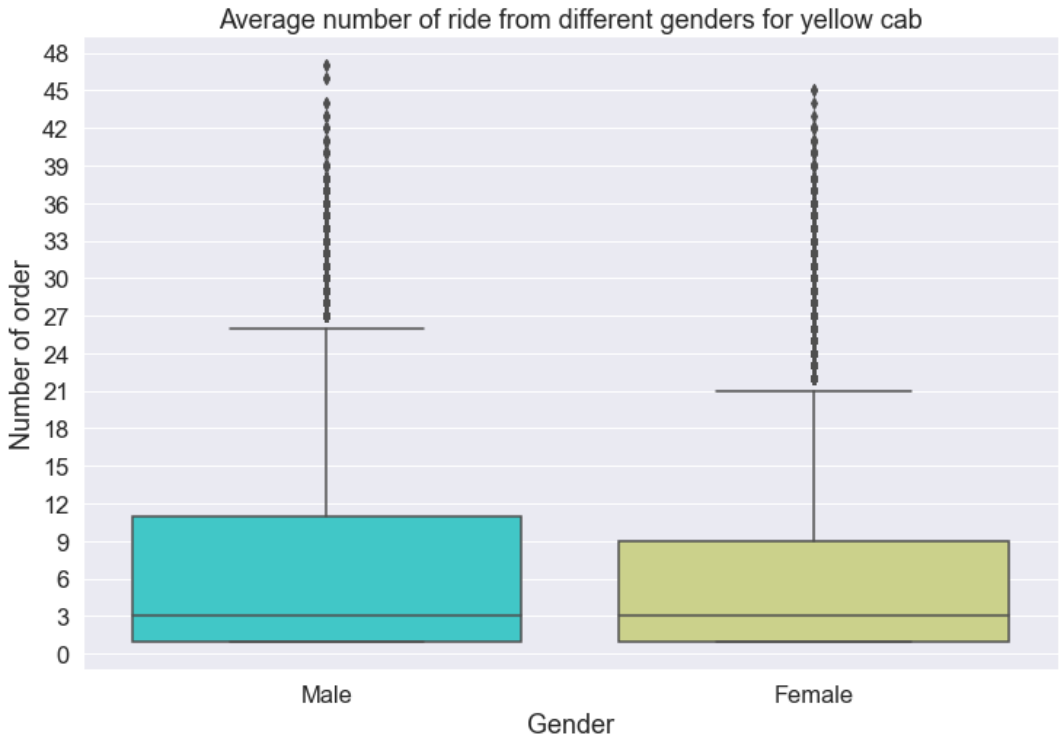




Surprisingly, the difference between average profit per ride for different income groups is ignorable for both Pink Cab and Yellow Cab.

Analysis of customer retention





Average ride of each cab user demanded from Pink Cab is approximately 2.5, while from Yellow Cab this number becomes 3. Thus Yellow Cab has better capacity in developing long term customers.

Investment Recommendation

- Here comes the final evaluation outcome:
- For hypothesis one, we can conclude that the scale of business for Yellow Cab outweighs Pink Cab drastically. For both amount of transaction as well as profit per ride, Yellow Cab outmatches Pink Cab, that is, given fixed input, Yellow Cab is more likely to generate greater profit.
- For hypothesis two, both companies experienced approximately same percentage of growth and recession in the given time interval, while Pink Cab's profit is more concentrated in Oct, Nov and Dec, Yellow Cab's profit is generated consistently throughout the year.
- For hypothesis three, though smaller in business scale, Pink Cab's profit in Californian cities are actually comparable to Yellow Cab's profit, which is surprising.
- For hypothesis four, there isn't any significant difference between customers of Yellow cab and Pink cab in factors including gender, age, income level, and these factors doesn't contribute to any difference in profit made.
- For hypothesis five, Yellow Cab has a better capacity in developing long term customers.
- The investment recommendation for our client XYZ is, if it already made detailed plans about stepping in the market of cab industry, and is determined to generate profit from it, choosing Yellow Cab, especially subdivision in east coast area, will be the most optimal option. However, if its business plan is still 'having a shot before making further decisions', then it can start from Pink Cab's subdivisions in California.

Thank You