



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

<G2M Insight for Cab Investment Firm>

<September 21, 2022>

Sok Leng Chan

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

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

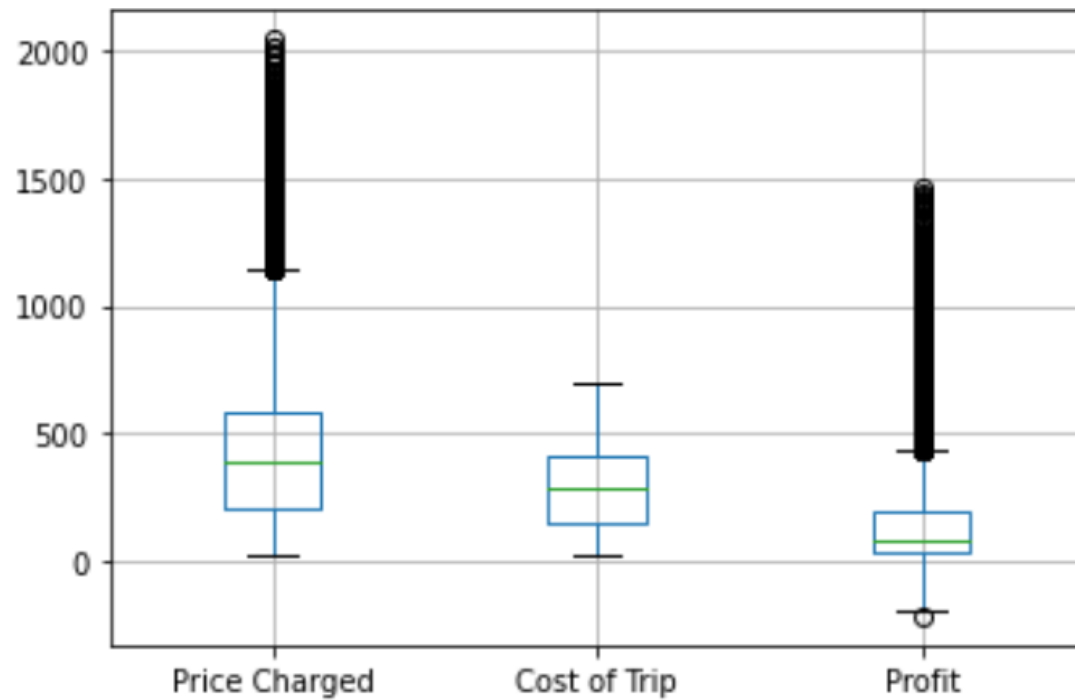
Problem Statement

- There are two cab company in the cab industry. After analyzed, we need to decide which cab is valued to invest.

Approach:

- Check the structure of the data and clean up the dirty data
- Check for missing value and duplicated value
- Check for the outliers

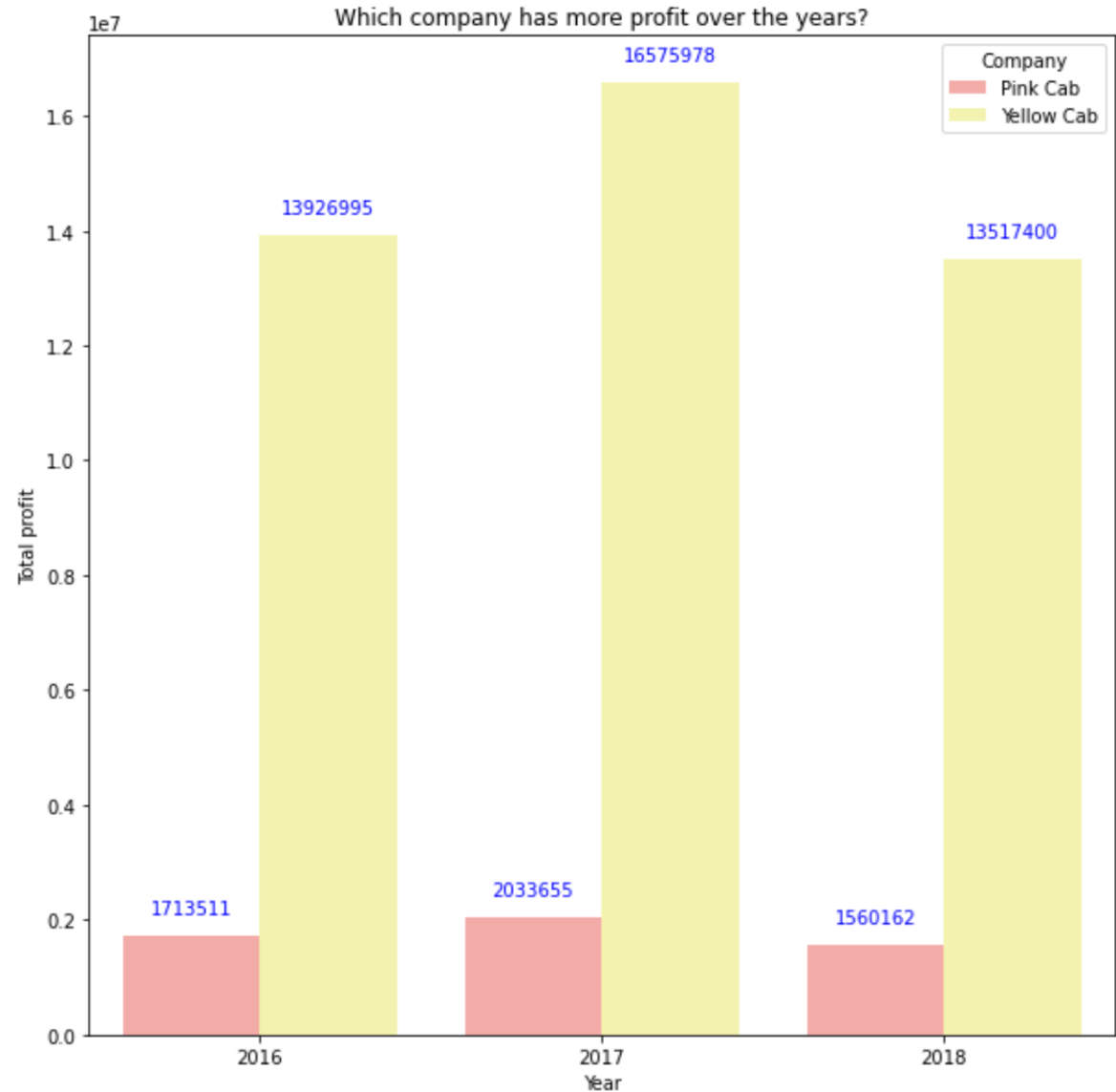
EDA (Data exploration Analysis)



I found outlier for the variables price charge, cost of trip and profit.

Which Company has more profit over the years?

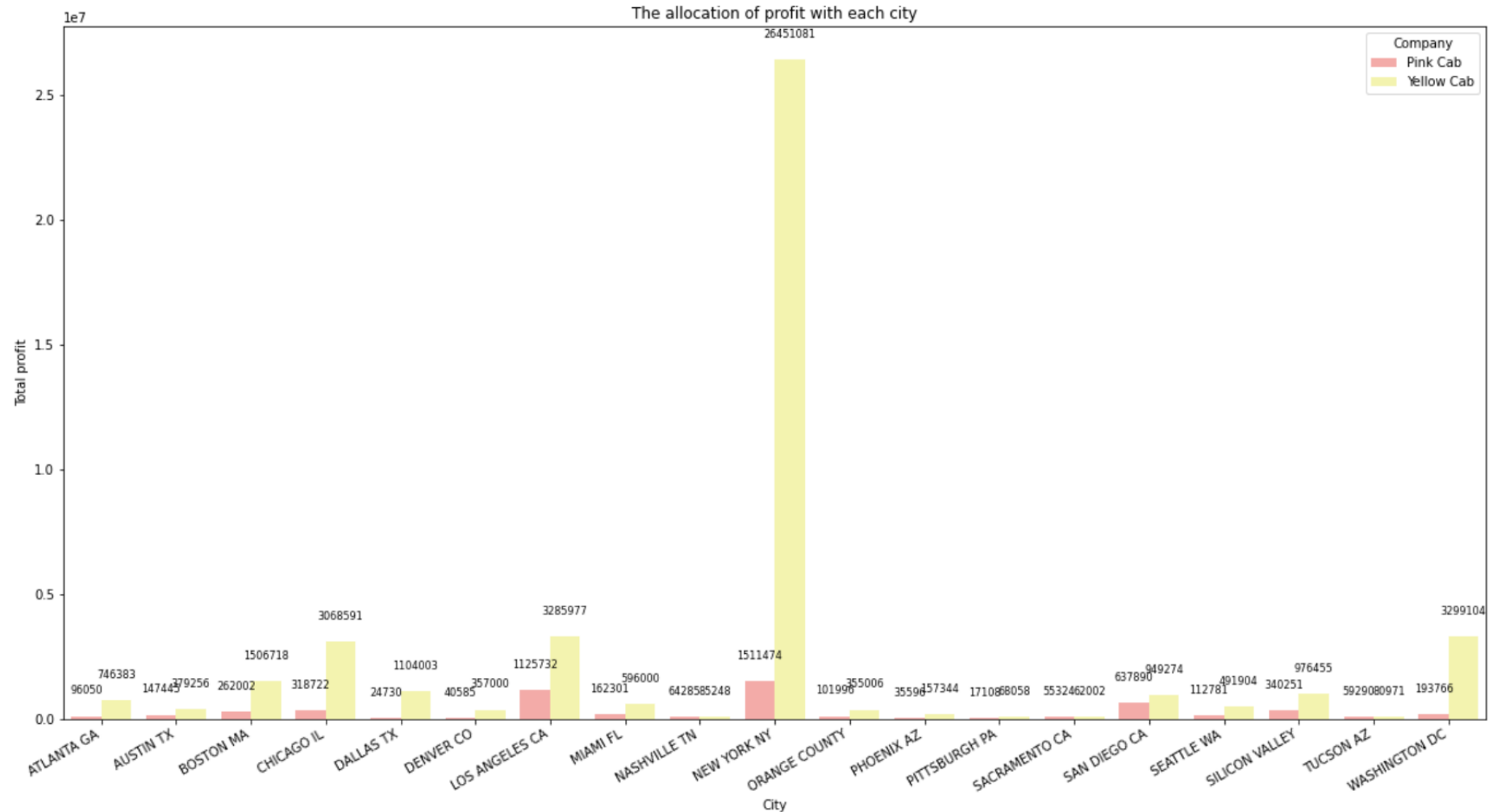
I found Yellow cab has more profit over the years. There is clear difference between yellow cab and pink cab. The profit of yellow cab is far more than that of pink cab.



Which City earn more profit?

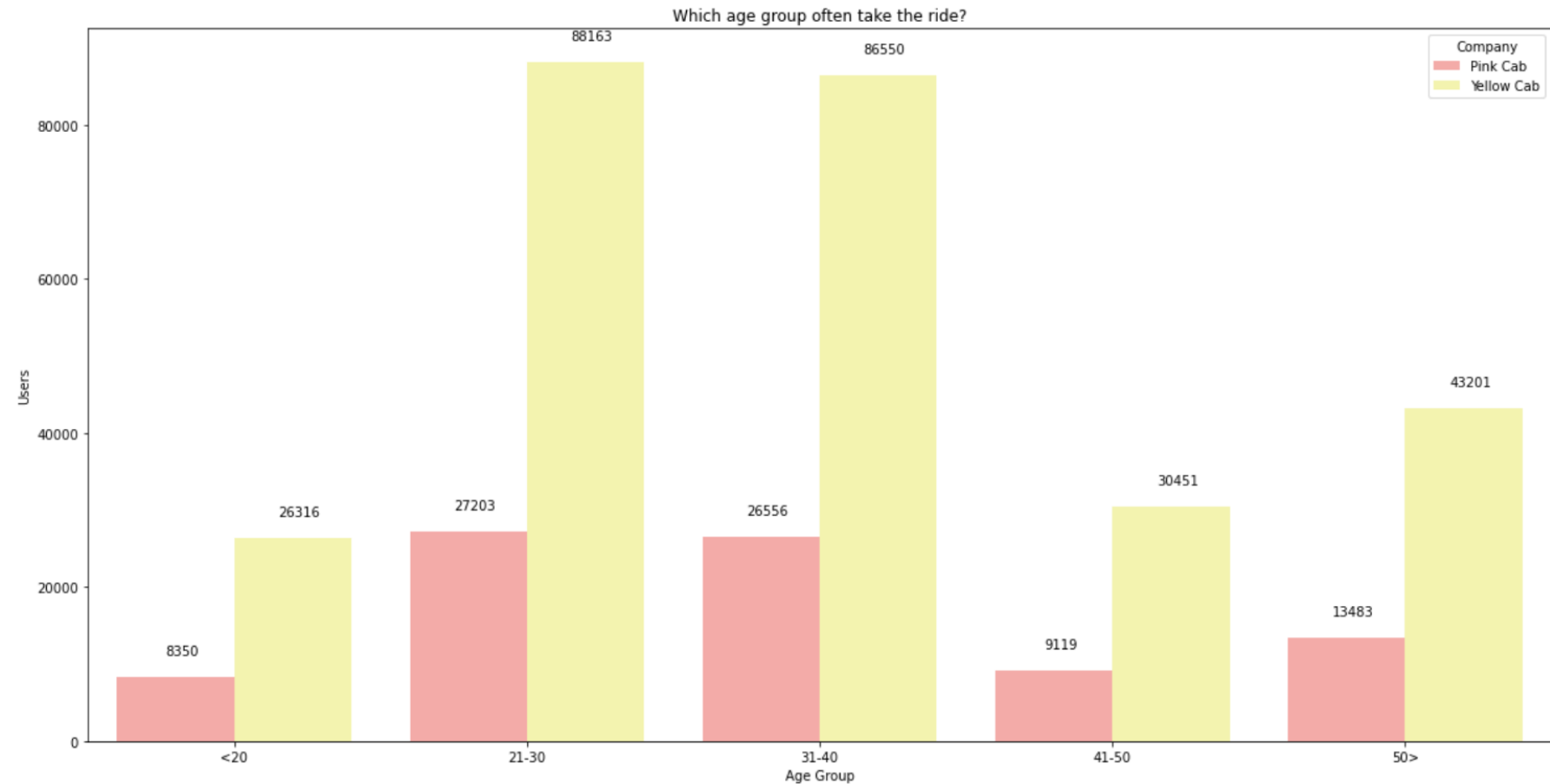
	Company	City	Total profit
28	Yellow Cab	NEW YORK NY	2.645108e+07
37	Yellow Cab	WASHINGTON DC	3.299104e+06
25	Yellow Cab	LOS ANGELES CA	3.285977e+06
22	Yellow Cab	CHICAGO IL	3.068591e+06
9	Pink Cab	NEW YORK NY	1.511474e+06

- New York City has higher profit for Yellow cab and Pink cab.
- Top 3 profit for the city as below:
- New York, Washington Dc, Los Angeles.
- Data reflected the most profit earn in New York, so it is best to increase the cab service in this region.



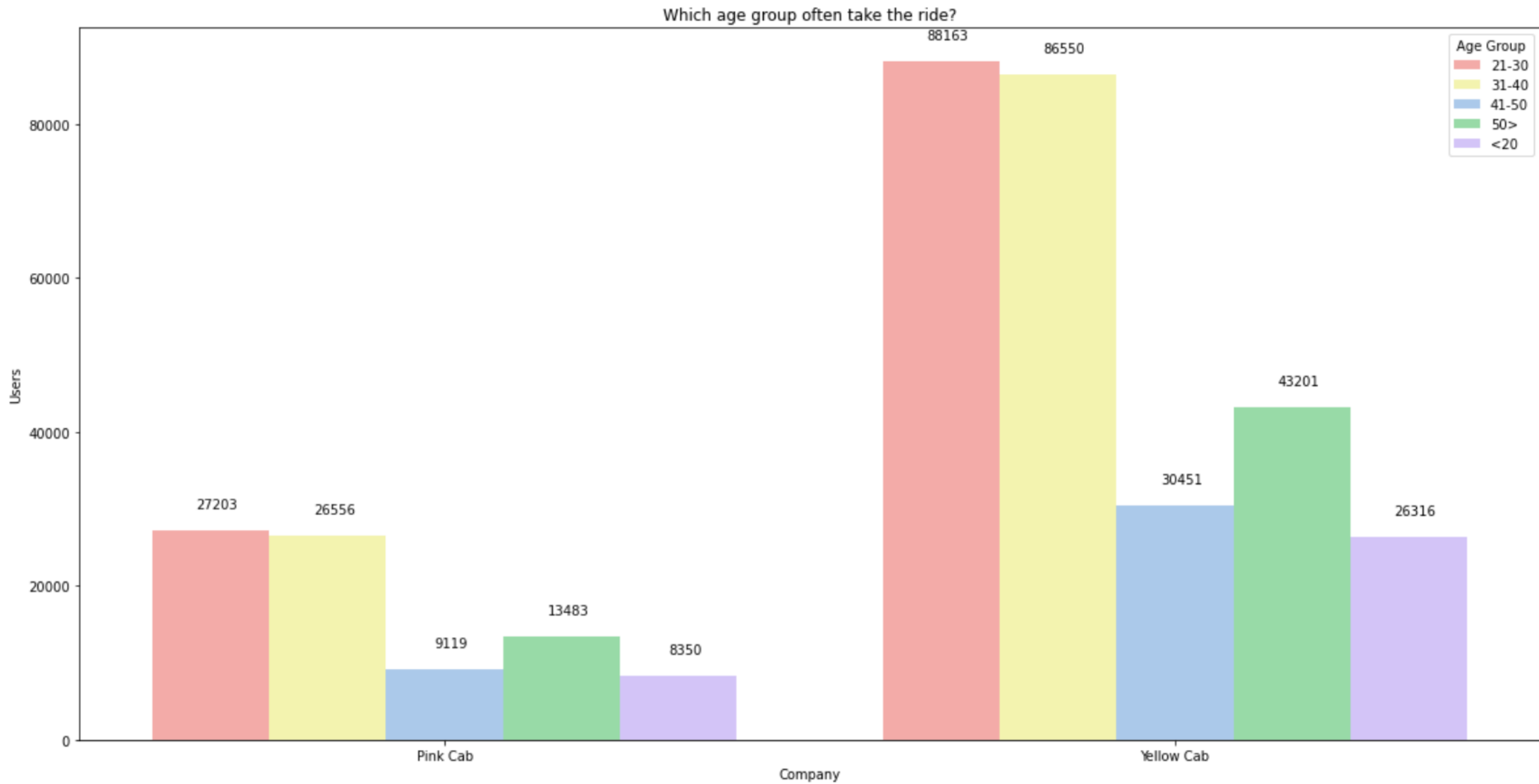
Which age group often take the ride?

	Age Group	Company	Users
0	21-30	Pink Cab	27203
1	21-30	Yellow Cab	88163
2	31-40	Pink Cab	26556
3	31-40	Yellow Cab	86550
4	41-50	Pink Cab	9119
5	41-50	Yellow Cab	30451
6	50>	Pink Cab	13483
7	50>	Yellow Cab	43201
8	<20	Pink Cab	8350
9	<20	Yellow Cab	26316

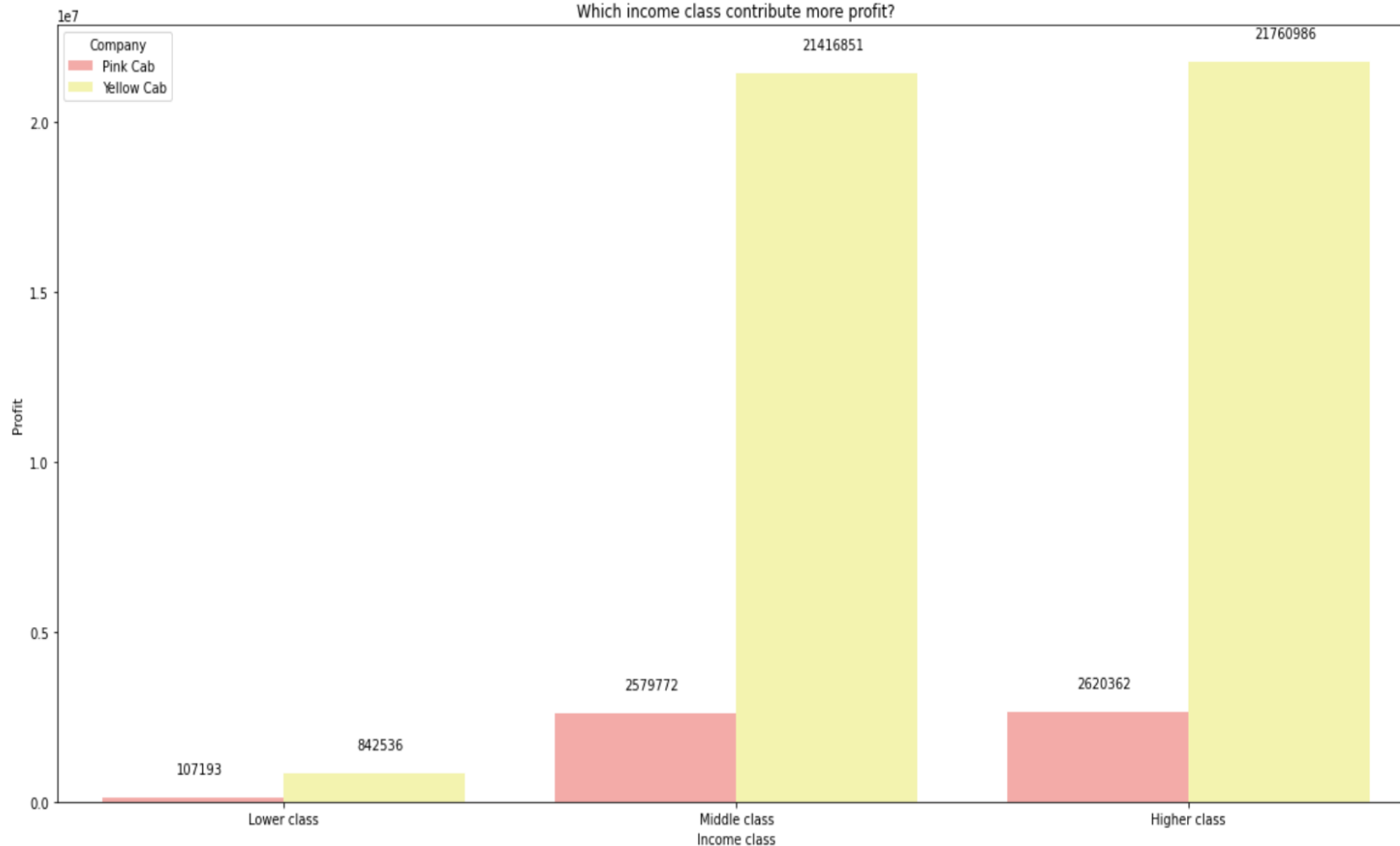


The users who most use both companies are in the age ranges between 21 and 40 years.

It reflect this age range often take the rides by cab and the majority age group for the service.



Which income class contribute more profit?



- As we see, the middle class and higher class are the majority group to take the yellow cab and pink cab. Thus, the yellow has much more profit than pink cab.

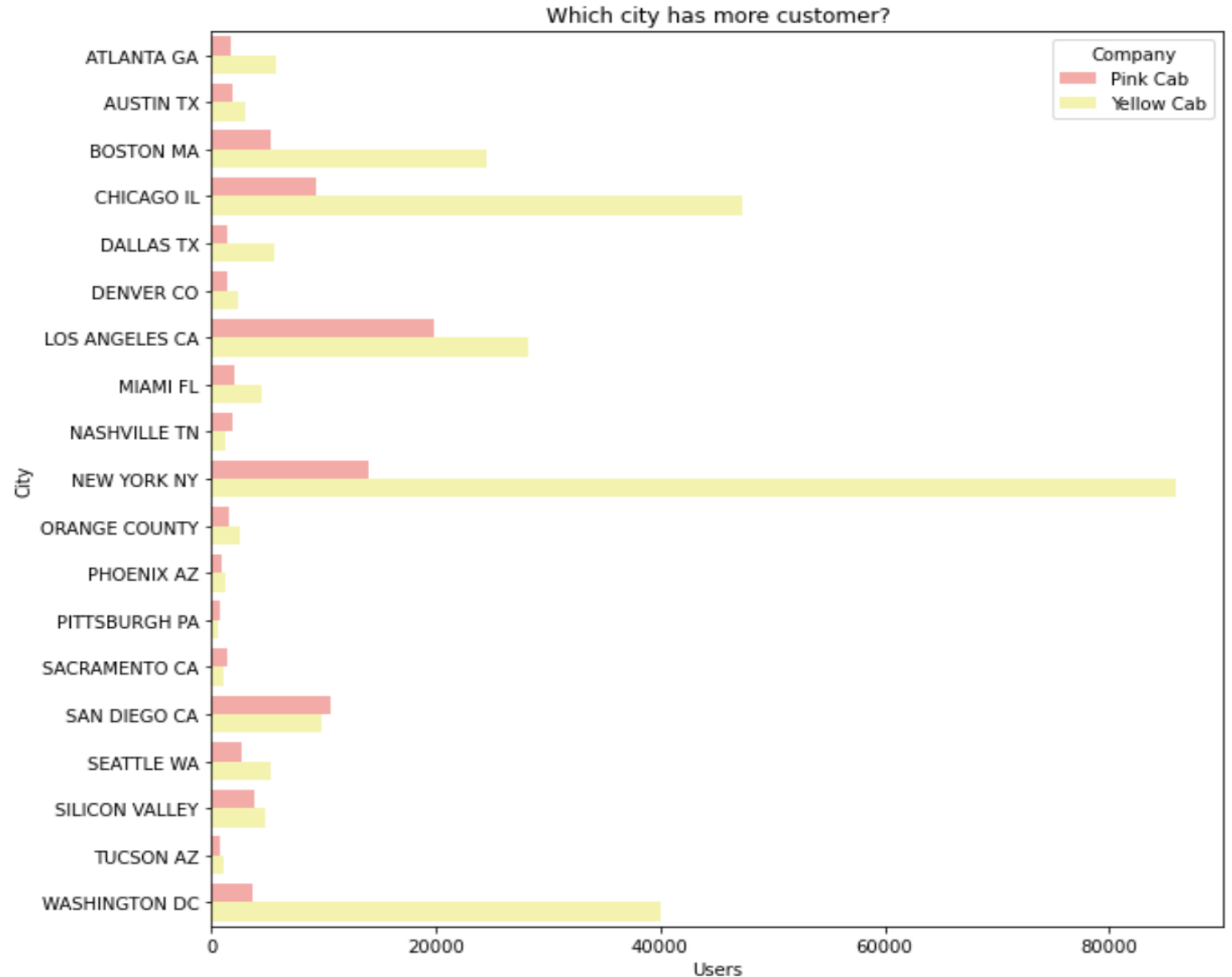
Which city has more customers?

- We can see that Yellow cab is the most users to ride and top 5 city which user take the ride as below: New York, Chicago, Washington DC, Los Angeles, Boston.
- For Pink Cab, Top 3 cities for ride are Los Angeles, New York, San Diego

	Company	City	Users
28	Yellow Cab	NEW YORK NY	85918
22	Yellow Cab	CHICAGO IL	47264
37	Yellow Cab	WASHINGTON DC	40045
25	Yellow Cab	LOS ANGELES CA	28168
21	Yellow Cab	BOSTON MA	24506
6	Pink Cab	LOS ANGELES CA	19865
9	Pink Cab	NEW YORK NY	13967
14	Pink Cab	SAN DIEGO CA	10672
33	Yellow Cab	SAN DIEGO CA	9816
3	Pink Cab	CHICAGO IL	9361

Which city has more customer?

- We can see the most customer to ride yellow cab in New York city and pink cab in Los Angeles. Although yellow cab is not the first choice for Los Angeles users, but it is still in the top 5 cities.



EDA summary:

- For exploration, I find out Yellow cab should be invested since it has more profit and more customers.
- After did analysis, we gained the below result:
 - -Yellow cab generate greater profit than pink cab over the years.
 - -Yellow cab generate more profit per year than pink cab
 - -Yellow cab has more customers than pink cab.
- In additional, I also found which city should be invested more in the future.
- After analyzed, New York City has more ride and more profit.
- -Top 5 city which has more customer are New York, Chicago, Washington , Los Angeles and Boston.
- -The age of customers are between 21 and 40 , they often take the ride no matter which company is.
- -Middle class and higher class are willing to spend money to take the cab.

Recommendations

- From EDA analysis, it seems that yellow cab has higher performance, and we can see all distributions of yellow cab is much higher than pink cab.
- For city coverages, it is obvious the yellow cab occupies very high percentage of coverage in cities. The most popular cities are New York, Chicago, Los Angeles, so I will suggest to increase more service in these regions and will have more profit in the future.

Thank You