



**Data Glacier**

Your Deep Learning Partner

# G2M Case Study: Cab Investment Firm

Data Glacier Virtual Internship

**9/16/2023**

# Agenda

Background of Case Study

Understanding the Data

EDA

EDA Summary

Recommendations

# Background of Case Study

- 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.
- Goal: XYZ is interested in using actionable insights to help them identify the right company to make their investment. Thus, this presentation's goal is to recommend the best company based on the data.
- There are 2 cab companies to look at:
  - **Pink Cab Company**
    - **Yellow Cab Company**
- Now, let's move on to understanding the data we've been given.

# Understanding the Data

XYZ provided 4 individual datasets:

- **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
- 
- Thus, this data was merged into one master data set based on the Cab data as it contained both the Transaction\_ID and the City for each transaction.
  - The time period of data we are interested in is from January 31<sup>st</sup>, 2016 to December 31<sup>st</sup>, 2018.

# Assumptions About the Data

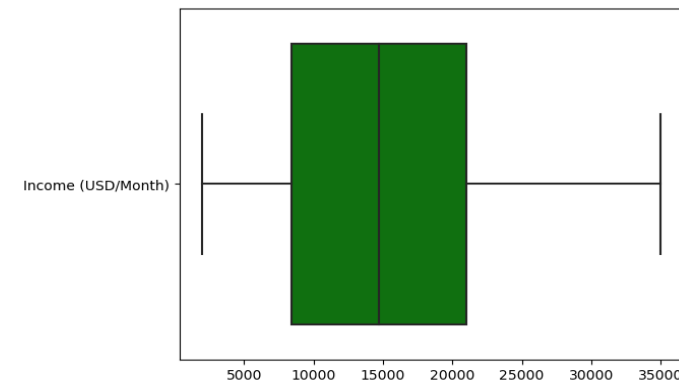
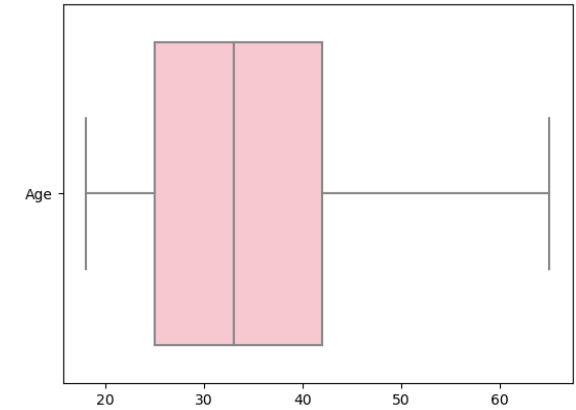
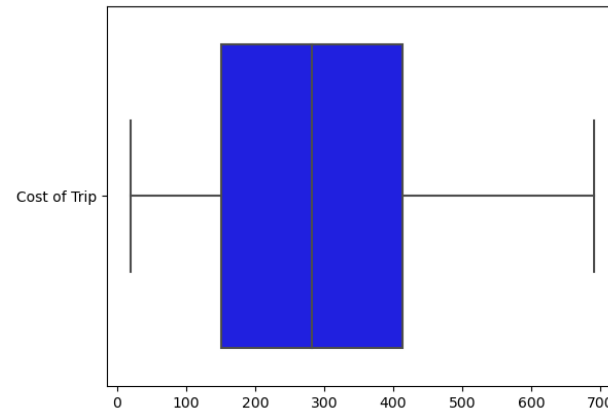
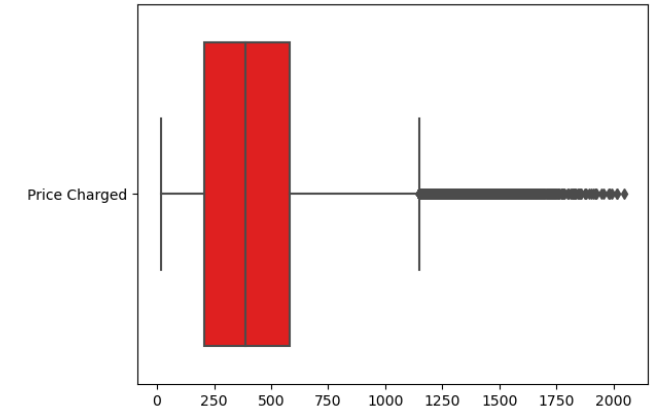
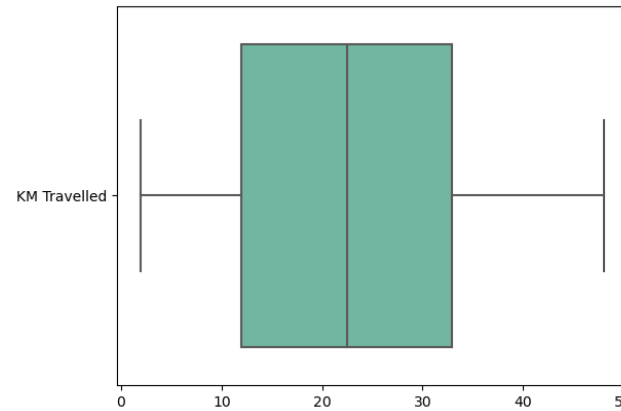
1. For the city data set, the column "Users" will be treated as the number of Cab users within that city.
2. The "Date of Travel" Column is in days from 1900 (1899, 12, 31 since 1900 was a leap year) which is the excel default original timestamp. Thus, the Date was converted and filtered the data for the time period specified and removed NA values for if the data had no date.
3. After filtering the master data set to the specified dates of the data we want to look at there are no other NA value columns. This means that the master data without the NAs give us a clearer picture of who and what we are dealing with.
4. The data recorded by each company was accurate and the missing values from one company were due to a lack of records.

# Checking for Outliers

We can see that most of the variables for the data don't have outliers.

It is reasonable to assume for the analysis that most of the data is normally distributed.

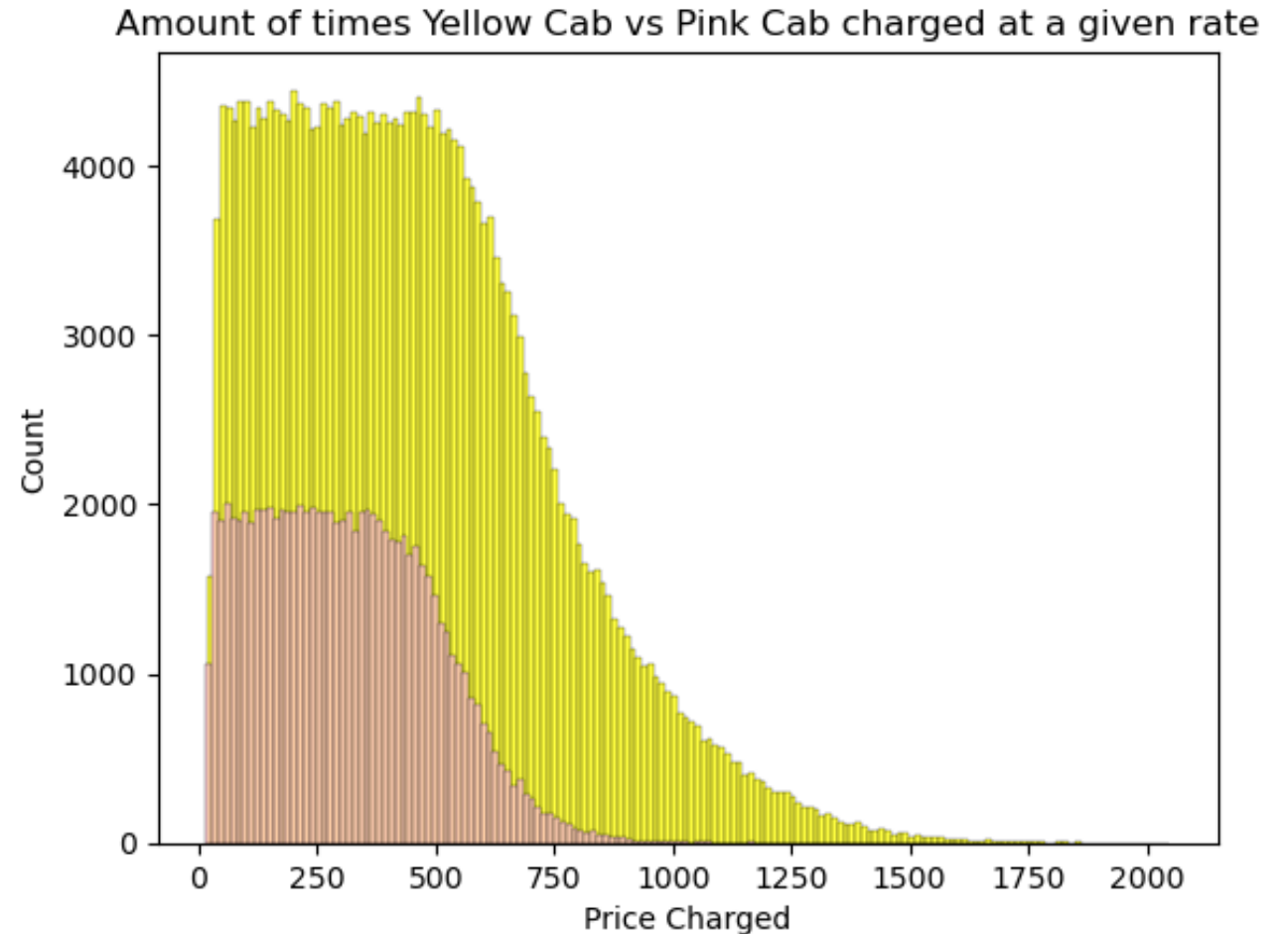
All of this is true except for the Price Charged variable which has a ton of outliers. Thus the price charged does go all the way into the \$2000 dollar range. However, since we don't know the full details of how long each trip was, I won't be treating them as outliers.



We can see from the graph to the right that not only does yellow have more volume of transactions, but also they charge higher rates to people.

This is important for our analysis since it means that yellow overall recorded better transactions than pink did. Thus, pink had a lot of NA values for their data which makes it unreliable for analysis.

Yellow already seems like the better company due to their volume of data alone, but let's break down the Profit between each company.



# Analysis (Per Company)

To calculate the profit of each company we need to first look at what each company is doing in terms of their Kilometers travelled, price charged, and cost of trip.

We can see that Yellow Cab on average is travelling around the same amount of Kilometers but charging more and spending more per trip.

To calculate profit for this analysis, I used the formula below:

$$\text{Profit} = \text{Sum of Price Charged} - \text{Sum of Cost of Trip}$$

Company	KM Travelled		Price Charged		Cost of Trip	
	sum	mean	sum	mean	sum	mean
Pink Cab	1890944.90	22.56	2.603852e+07	310.63	20799282.80	248.13
Yellow Cab	6122880.43	22.57	1.241268e+08	457.53	80824199.52	297.92



# Profit Analysis (Yearly)

Yearly analysis shows below that the profit per year is drastically different between each company with the Yellow Cab company profiting 10x the amount that Pink Cab makes. 2017 was the best year for both companies making an extra 25% of their annual income from the previous year.

We can also see that for each KM travelled Yellow Cab Company is making 3 times the Profit per KM for Pink Cab Company. These numbers are staggering as Pink Cab only 1-2 million dollars in profit per year while Yellow makes 13-16 million. Next, let's break this down by quarter.

		KM Travelled	Price Charged	Cost of Trip	Profit	Profit Per Month	Profit Per Day	Charge Per KM	Profit Per KM
Company	Date of Travel								
Pink Cab	2016	540694.55	7573094.36	5943731.79	1629362.57	135780.21	4464.01	14.01	3.01
	2017	685635.03	9577146.48	7542978.22	2034168.26	169514.02	5573.06	13.97	2.97
	2018	664615.32	8888278.86	7312572.80	1575706.06	131308.84	4317.00	13.37	2.37
Yellow Cab	2016	1773413.72	36529301.94	23411909.94	13117392.00	1093116.00	35938.06	20.60	7.40
	2017	2217474.69	45871862.15	29278065.45	16593796.70	1382816.39	45462.46	20.69	7.48
	2018	2131992.02	41725647.09	28134224.12	13591422.97	1132618.58	37236.78	19.57	6.37

# Profit Analysis (Quarterly)

For each quarter we can see that each companies profit is different.

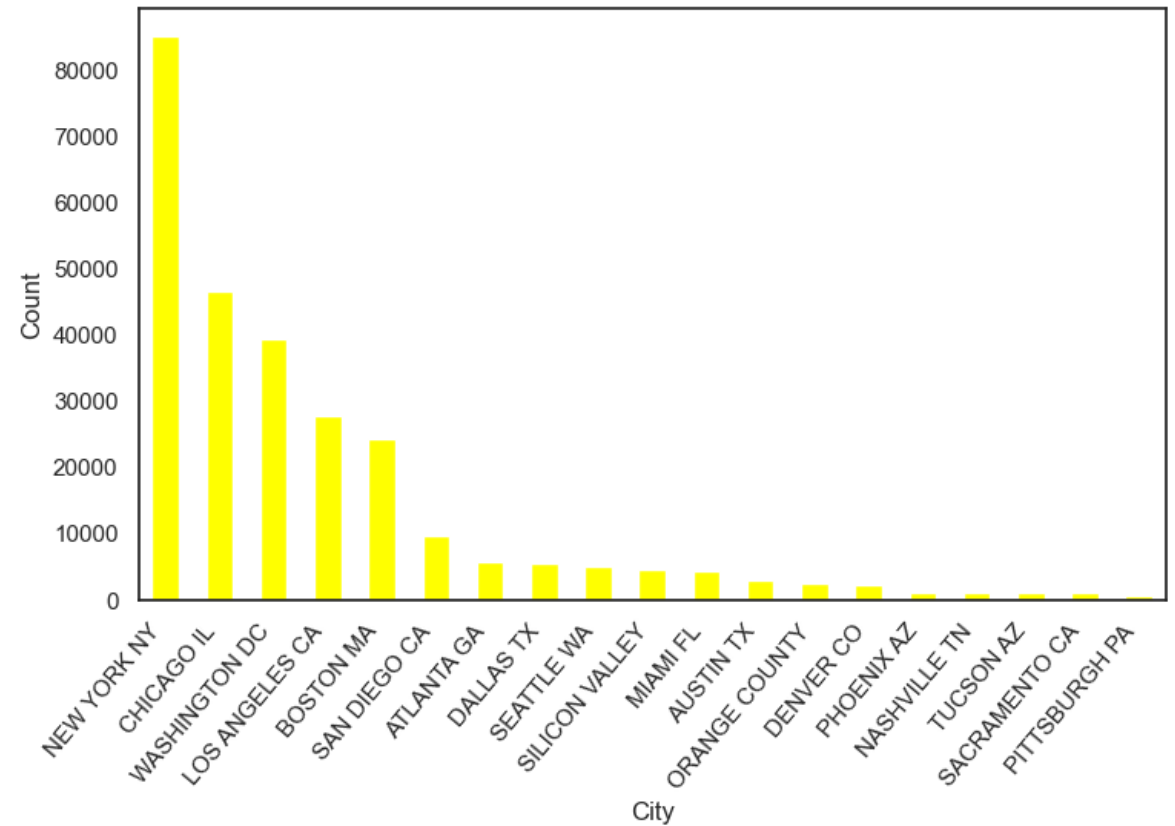
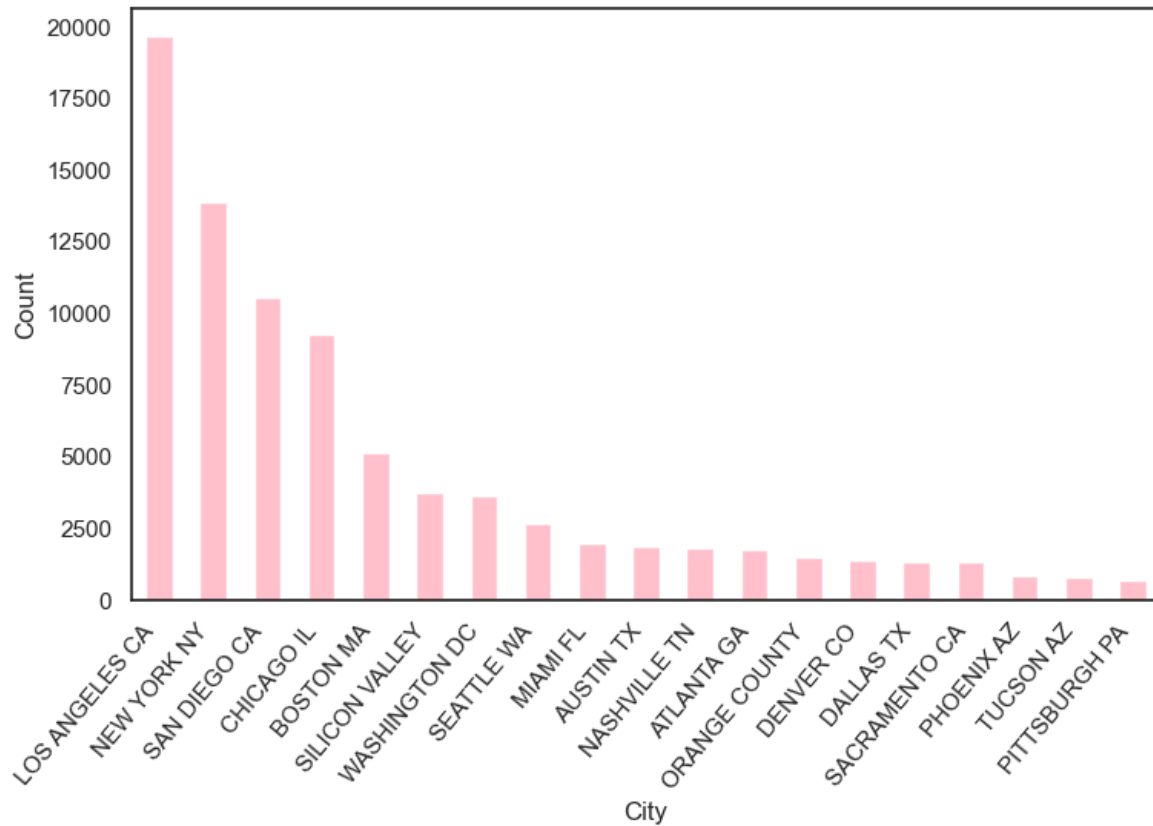
Yellow Cab company tends to do better in quarters 2 and 4 while Pink Cab company tends to make more profit in quarters 3 and 4.

This means that Yellow Cab is probably located in cities where intense weather occurs while Pink Cab has a tendency to do well in colder months.

Company	Date of Travel	KM Travelled				Price Charged		Cost of Trip		Profit	
Pink Cab	1	272	100.11	3890	898.46	2992	280.92	8986	17.54		
	2	333	964.91	4320	401.92	3674	023.74	6463	78.18		
	3	541	788.39	7150	825.35	5957	938.46	1192	886.89		
	4	743	091.49	1067	6393.97	8175	039.69	2501	354.28		
Yellow Cab	1	1074	193.22	2298	3729.63	1418	0374.95	8803	354.68		
	2	1236	594.38	2697	1472.67	1632	0199.99	1065	1272.68		
	3	1629	961.36	3145	2148.46	2151	9604.32	9932	544.14		
	4	2182	131.47	4271	9460.42	2880	4020.26	1391	5440.16		

# City Analysis

Here is the City breakdown of our data. Yellow cab does the best in New York City with over 80000 transactions while Pink Cab did the best in Los Angeles with approximately 20000 observations. Thus, Yellow Cab tends to gain the most data and profit in New York compared to most other cities.





# Profit Analysis (by City)

		KM Travelled	Price Charged	Cost of Trip	Profit	Profit Per KM
City	Company					
DALLAS TX	Pink Cab	29969.03	354559.42	329498.85	25060.57	0.836216
PITTSBURGH PA	Pink Cab	15359.66	186066.70	169192.92	16873.78	1.098578
DENVER CO	Pink Cab	31064.52	382356.02	341758.82	40597.20	1.306867
CHICAGO IL	Pink Cab	207930.59	2598708.02	2286456.39	312251.63	1.501711
NASHVILLE TN	Pink Cab	41235.00	515698.58	452394.01	63304.57	1.535215
SACRAMENTO CA	Pink Cab	29841.78	381586.44	328321.80	53264.64	1.784902
PHOENIX AZ	Pink Cab	18571.26	238753.83	204250.92	34502.91	1.857866
SEATTLE WA	Pink Cab	60418.26	778100.99	665491.86	112609.13	1.863826
BOSTON MA	Pink Cab	116530.44	1540983.77	1281139.87	259843.90	2.229837
WASHINGTON DC	Pink Cab	82150.77	1096267.45	904451.47	191815.98	2.334926
ATLANTA GA	Pink Cab	39209.93	525463.64	430727.46	94736.18	2.416127
LOS ANGELES CA	Pink Cab	445013.08	5995489.09	4894468.81	1101020.28	2.474130
SACRAMENTO CA	Yellow Cab	23287.74	365859.56	307012.45	58847.11	2.526957
SAN DIEGO CA	Pink Cab	237912.62	3245623.79	2617599.59	628024.20	2.639726
BOSTON MA	Yellow Cab	545632.90	8676316.95	7204327.13	1471989.82	2.697766
CHICAGO IL	Yellow Cab	1055027.42	16900943.45	13925153.94	2975789.51	2.820580
ORANGE COUNTY	Pink Cab	33341.87	467730.80	367128.11	100602.69	3.017308
NASHVILLE TN	Yellow Cab	26660.60	435578.76	352171.52	83407.24	3.128483
TUCSON AZ	Yellow Cab	24999.48	409831.91	330591.10	79240.81	3.169698
	Pink Cab	17893.89	256211.30	197606.42	58604.88	3.275134

AUSTIN TX	Pink Cab	40803.28	595306.30	448561.39	146744.91	3.596400
WASHINGTON DC	Yellow Cab	892943.12	15025223.01	11788212.61	3237010.40	3.625103
MIAMI FL	Pink Cab	44444.88	649931.87	488168.94	161762.93	3.639630
SILICON VALLEY	Pink Cab	85458.67	1275387.47	940149.82	335237.65	3.922804
SEATTLE WA	Yellow Cab	116626.99	2010554.27	1536626.48	473927.79	4.063620
SAN DIEGO CA	Yellow Cab	217150.19	3794926.36	2865556.80	929369.56	4.279847
PITTSBURGH PA	Yellow Cab	14172.55	252441.85	185904.02	66537.83	4.694838
NEW YORK NY	Pink Cab	313795.37	4954294.22	3451915.36	1502378.86	4.787766
LOS ANGELES CA	Yellow Cab	630749.05	11536591.48	8328527.62	3208063.86	5.086118
AUSTIN TX	Yellow Cab	67706.13	1267498.80	893258.52	374240.28	5.527421
ATLANTA GA	Yellow Cab	127274.68	2406772.46	1679652.72	727119.74	5.712996
PHOENIX AZ	Yellow Cab	26681.02	506928.89	352502.21	154426.68	5.787885
MIAMI FL	Yellow Cab	100177.05	1913616.95	1323770.08	589846.87	5.888044
ORANGE COUNTY	Yellow Cab	54746.37	1070456.06	721552.56	348903.50	6.373089
DENVER CO	Yellow Cab	53825.22	1059064.01	710824.41	348239.60	6.469822
DALLAS TX	Yellow Cab	125351.50	2730881.23	1652580.42	1078300.81	8.602217
SILICON VALLEY	Yellow Cab	106091.54	2361501.09	1401098.83	960402.26	9.052581
NEW YORK NY	Yellow Cab	1913776.88	51401824.09	25264876.07	26136948.02	13.657260

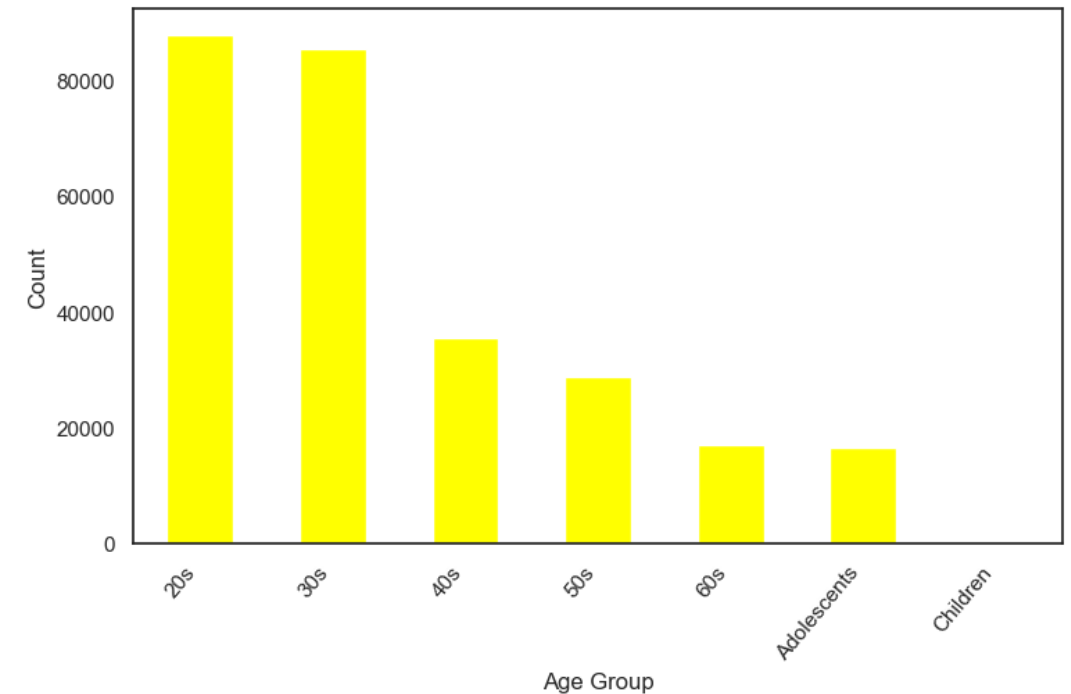
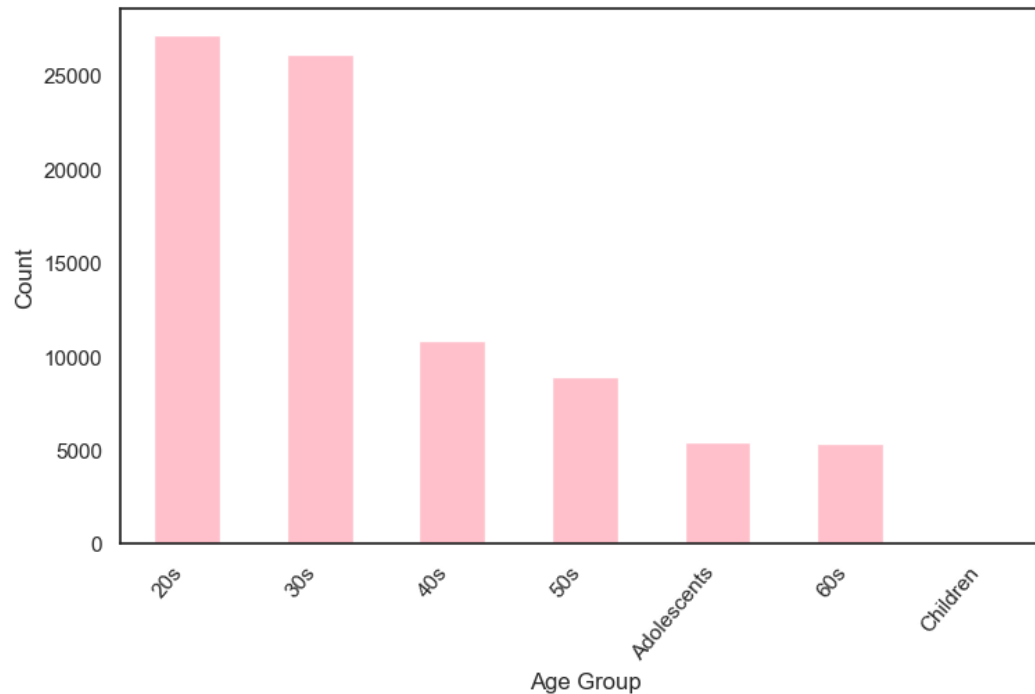


# Age

Finally, there seems to be a trend in the way younger people are taking Ubers to get around.

But is that true?

Taking a look at the graph below, we can see that people in their 20s and 30s are spending the most of cab rides for both companies. New York, for example, is not a place that tends to have older people due to it not being a great place for retirement.



# Recommendations

After looking at the numbers and graphs, we believe that Yellow Cab would be a better investment for a couple of reasons:

- Locations: Not only does Yellow Cab company have more locations that they operate in, but also they tend to dominate in those markets.
- Quarterly Profit: Due to their location base, Yellow cab tends to do better in quarters 2 and 4. Since they dominate the New York market and since it gets hot and cold in the summer and winter months, people tend to take cabs more to get out of the cold which results in more money.
- Yearly Profit: Yellow Cab within the two years made around 42 million dollars while Pink Cab only made 5 million.
- Volume: Yellow Cab not only has more volume, but charges more per transaction than Pink Cab does.
- Age Group: Yellow Cab just like Pink Cab appeals to younger audiences more than older audiences, but for people in their 20s Yellow cab made 80000 transactions while Pink Cab only made 20000 transactions for the two years.
- Kilometers travelled and Price Charged: Both companies on average are travelling the same distance, but Yellow Cab is making more profit by charging more.

Thus, we believe that Yellow Cab Company would be the best company to invest in for the future.

# Thank You