



Data Glacier

Your Deep Learning Partner

Exploratory Data Analysis

G2M insight for Cab Investment firm

Submitted by: Negar

October 21, 2022

Agenda

Executive Summary

Problem Statement

Approach

EDA

EDA Summary

Recommendations

Executive Summary

- Provide useful information to assist the investor choose the best cab company to invest in.
- Cab companies:
 - Yellow Cab
 - Pink Cab

Problem Statement

- Due to remarkable growth in the Cab Industry in last few years and multiple key players in the market, XYZ 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 data available for two Cab companies to investigate. Thus, they need to know which Cab company can perform better to invest on.

Approach

- Look through each dataset and comprehend it
- Identify relationships across the datasets and create a master dataset
- Feature transformations
- Identify and remove duplicates
- Identify NA values and remove them
- Outlier detection
- Exploratory Data Analysis

Master Dataset:

- Time period of data is from 31/01/2016 to 31/12/2018

	Transaction ID	Date of Travel	Company	City	KM Travelled	Price Charged	Cost of Trip	Profit	Year	Month	Day	Customer ID	Payment_Mode	Gender	Age	Income (USD/Month)	Population	Users
0	10000011	2016-01-08	Pink Cab	ATLANTA GA	30.45	370.95	313.6350	57.3150	2016	1	8	29290	Card	Male	28	10813	814885	24701
1	10351127	2018-07-21	Yellow Cab	ATLANTA GA	26.19	598.70	317.4228	281.2772	2018	7	21	29290	Cash	Male	28	10813	814885	24701
2	10412921	2018-11-23	Yellow Cab	ATLANTA GA	42.55	792.05	597.4020	194.6480	2018	11	23	29290	Card	Male	28	10813	814885	24701
3	10000012	2016-01-06	Pink Cab	ATLANTA GA	28.62	358.52	334.8540	23.6660	2016	1	6	27703	Card	Male	27	9237	814885	24701
4	10320494	2018-04-21	Yellow Cab	ATLANTA GA	36.38	721.10	467.1192	253.9808	2018	4	21	27703	Card	Male	27	9237	814885	24701
...
359387	10307228	2018-03-03	Yellow Cab	WASHINGTON DC	38.40	668.93	525.3120	143.6180	2018	3	3	51406	Cash	Female	29	6829	418859	127001
359388	10319775	2018-04-13	Yellow Cab	WASHINGTON DC	3.57	67.60	44.5536	23.0464	2018	4	13	51406	Cash	Female	29	6829	418859	127001
359389	10347676	2018-07-06	Yellow Cab	WASHINGTON DC	23.46	331.97	337.8240	-5.8540	2018	7	6	51406	Card	Female	29	6829	418859	127001
359390	10358624	2018-08-02	Yellow Cab	WASHINGTON DC	27.60	358.23	364.3200	-6.0900	2018	8	2	51406	Cash	Female	29	6829	418859	127001

Exploratory Data Analysis

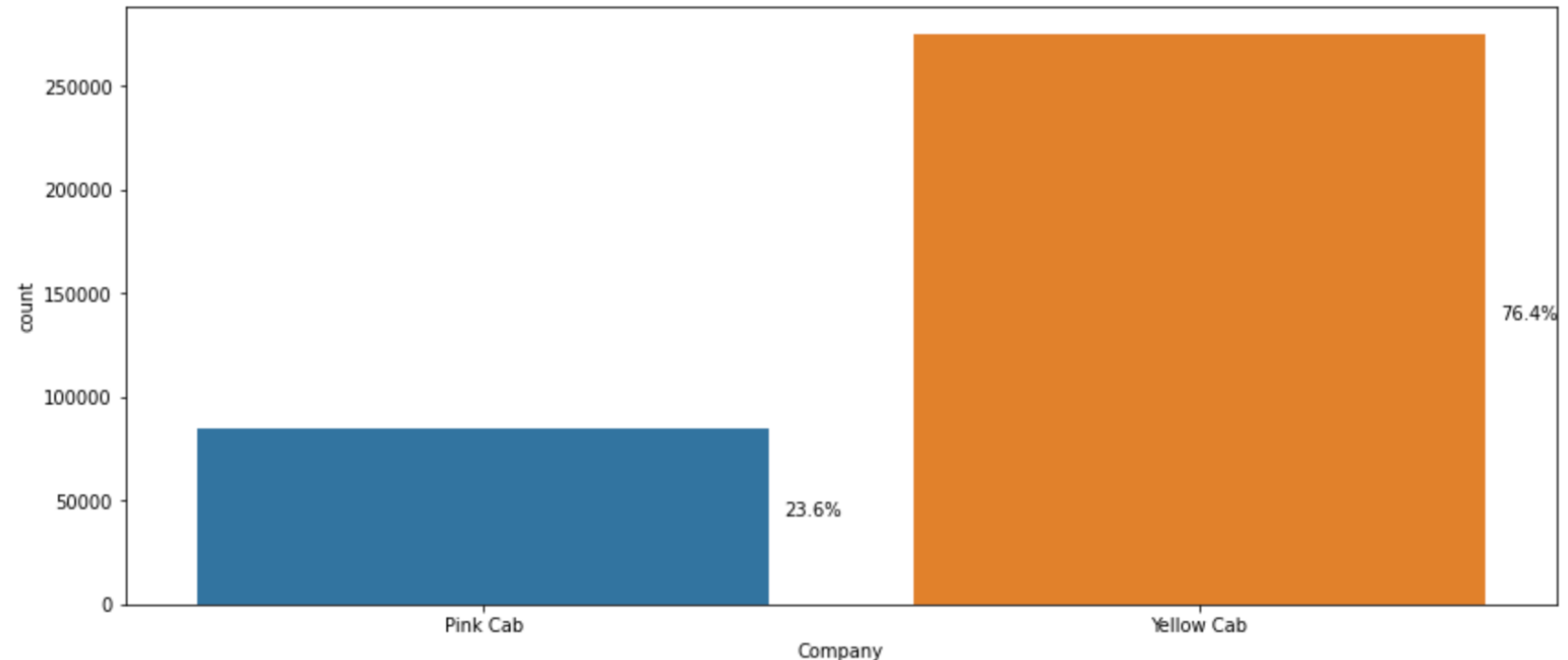
Correlation:

- The heatmaps show a strong relationship between 'KM Travelled,' 'Price Charged,' 'Cost of Trip' and 'Profit.' Additionally, and this is rather evident, there is a significant correlation between the 'Population' and 'Users.'

	Transaction ID	KM Travelled	Price Charged	Cost of Trip	Profit	Year	Month	Day	Customer ID	Age	Income (USD/Month)	Population	Users
Transaction ID	1.000000	-0.001429	-0.052902	-0.003462	-0.087130	0.941600	0.287009	0.012021	-0.016912	-0.001267	-0.001570	0.023868	0.013526
KM Travelled	-0.001429	1.000000	0.835753	0.981848	0.462768	-0.001072	-0.001903	0.001652	0.000389	-0.000369	-0.000544	-0.002311	-0.000428
Price Charged	-0.052902	0.835753	1.000000	0.859812	0.864154	-0.036409	-0.061116	-0.004021	-0.177324	-0.003084	0.003228	0.326589	0.281061
Cost of Trip	-0.003462	0.981848	0.859812	1.000000	0.486056	-0.001773	-0.008370	0.002091	0.003077	-0.000189	-0.000633	0.015108	0.023628
Profit	-0.087130	0.462768	0.864154	0.486056	1.000000	-0.060567	-0.096352	-0.008942	-0.306527	-0.005093	0.006148	0.544079	0.457758
Year	0.941600	-0.001072	-0.036409	-0.001773	-0.060567	1.000000	-0.030490	-0.020995	-0.002539	-0.000579	-0.001617	0.000003	-0.000771
Month	0.287009	-0.001903	-0.061116	-0.008370	-0.096352	-0.030490	1.000000	0.041412	-0.044656	-0.002226	0.000322	0.064571	0.036558
Day	0.012021	0.001652	-0.004021	0.002091	-0.008942	-0.020995	0.041412	1.000000	0.000650	-0.001275	0.001860	0.000970	0.000514
Customer ID	-0.016912	0.000389	-0.177324	0.003077	-0.306527	-0.002539	-0.044656	0.000650	1.000000	-0.004735	-0.013608	-0.647052	-0.610742
Age	-0.001267	-0.000369	-0.003084	-0.000189	-0.005093	-0.000579	-0.002226	-0.001275	-0.004735	1.000000	0.003907	-0.009002	-0.005906
Income (USD/Month)	-0.001570	-0.000544	0.003228	-0.000633	0.006148	-0.001617	0.000322	0.001860	-0.013608	0.003907	1.000000	0.011868	0.010464
Population	0.023868	-0.002311	0.326589	0.015108	0.544079	0.000003	0.064571	0.000970	-0.647052	-0.009002	0.011868	1.000000	0.915490
Users	0.013526	-0.000428	0.281061	0.023628	0.457758	-0.000771	0.036558	0.000514	-0.610742	-0.005906	0.010464	0.915490	1.000000

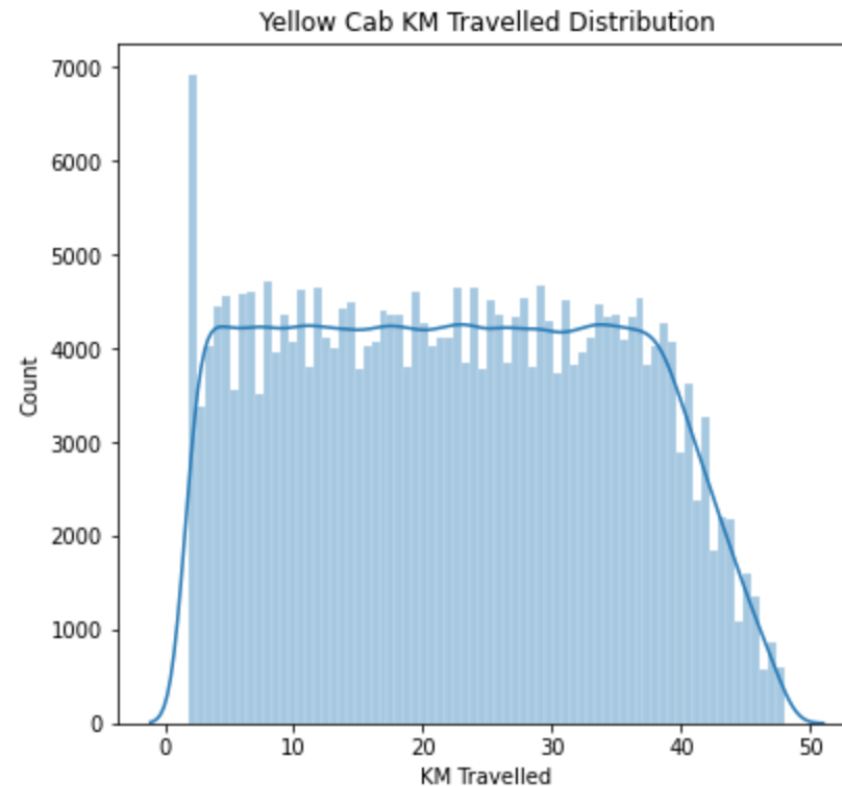
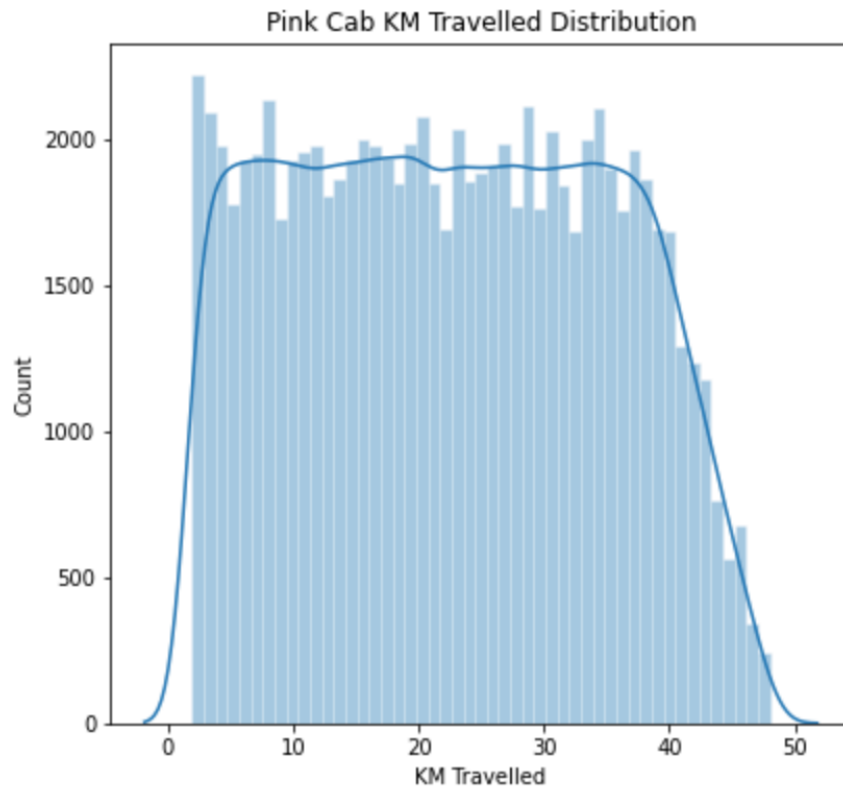
Distribution of Cabs

- About 75% of cabs come from the yellow cab company, and 25% from the pink cab company. Thus, It seems that consumers like the "Yellow Cab" over the "Pink Cab" more.



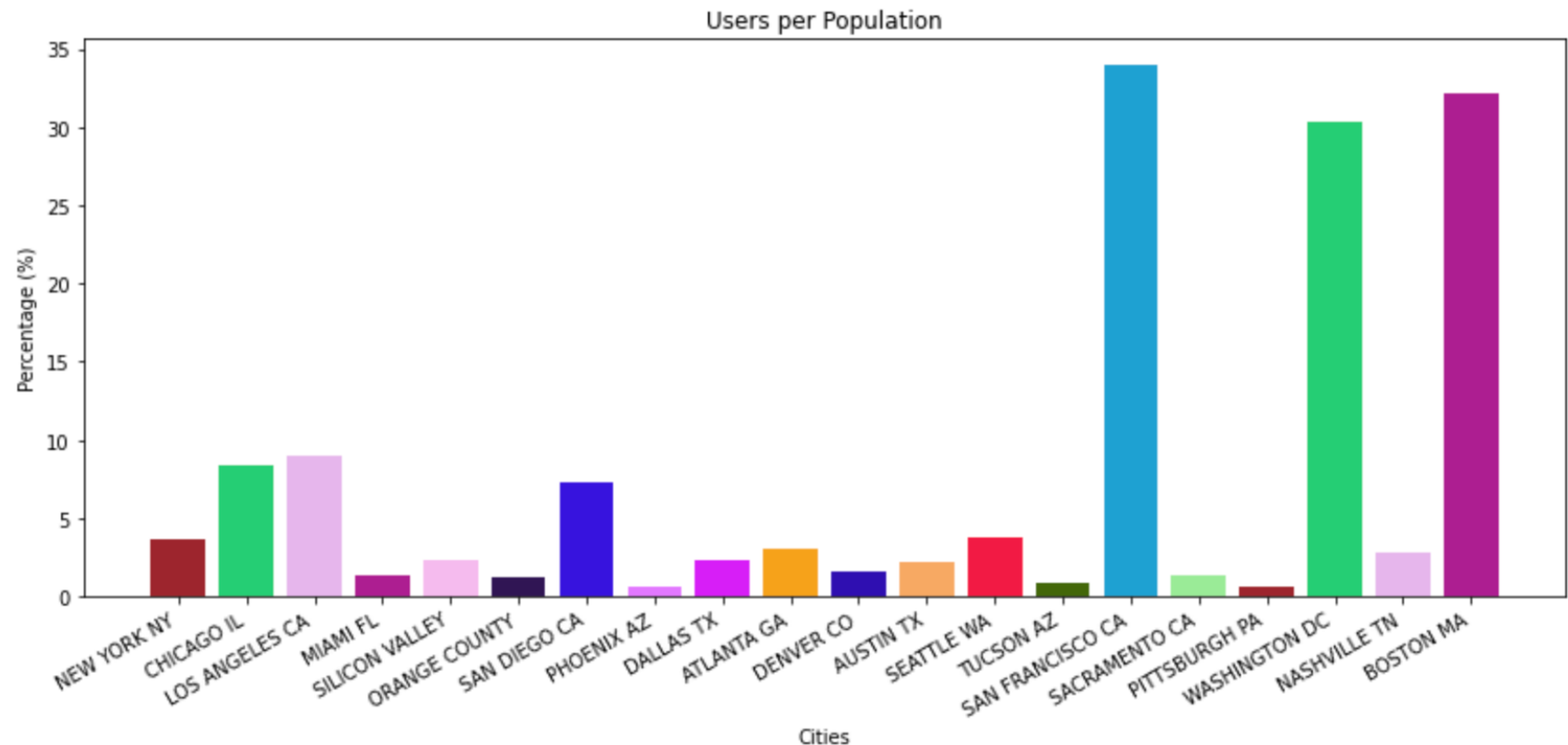
KM Travelled Distribution

- Most rides range in length from 2 to 48 kilometers for both Pink and Yellow cabs.



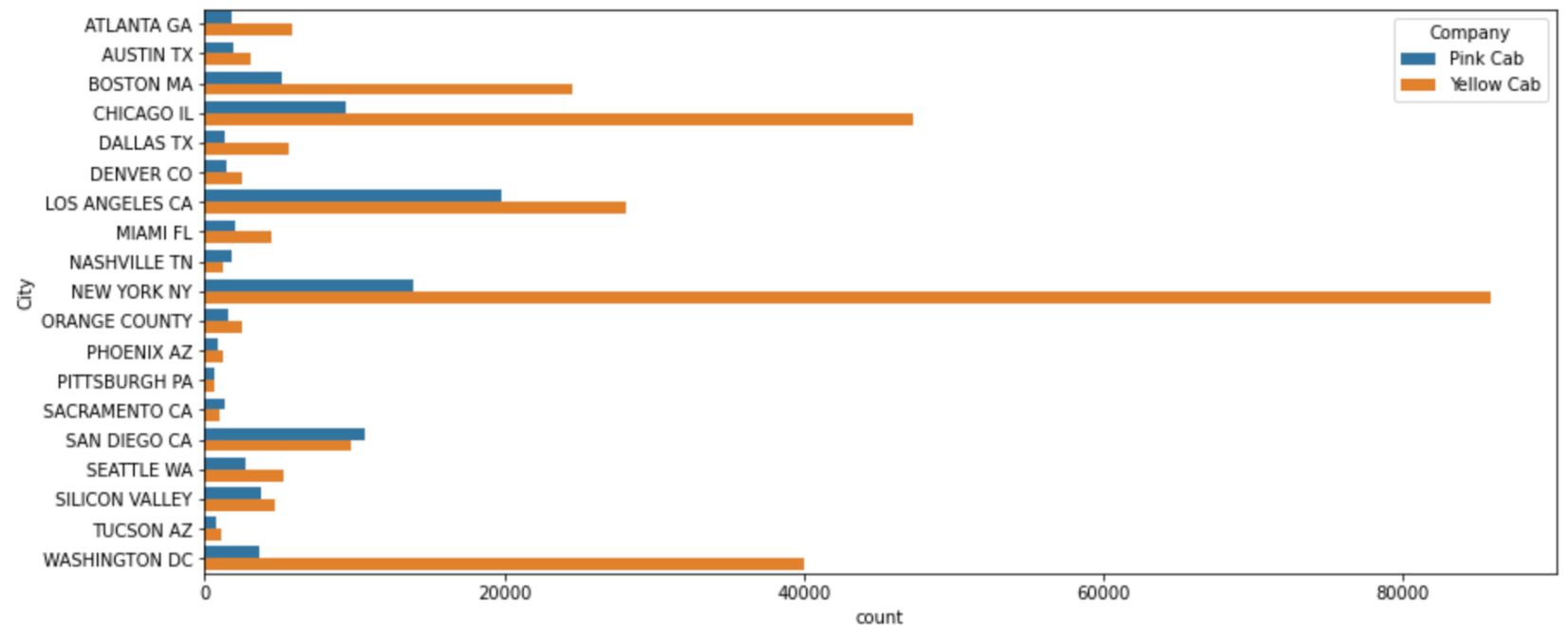
Users per population

- More than 30% of people in the cities of "San Francisco," "Washington," and "Boston" use taxis.



Pink and Yellow cab Users Distribution

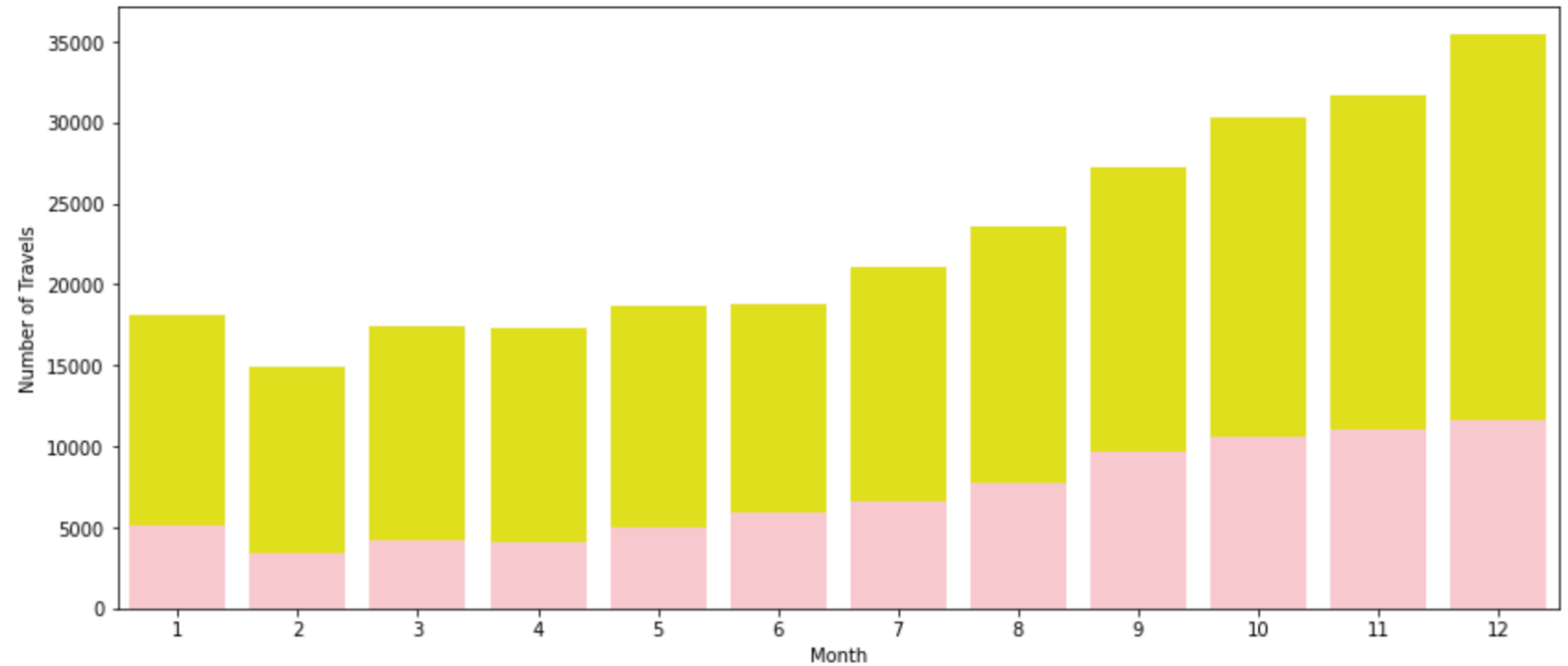
- People in the cities of "Boston," "Chicago," "Los Angeles," "New York" and "Washington," far prefer Yellow Cab rather than Pink.



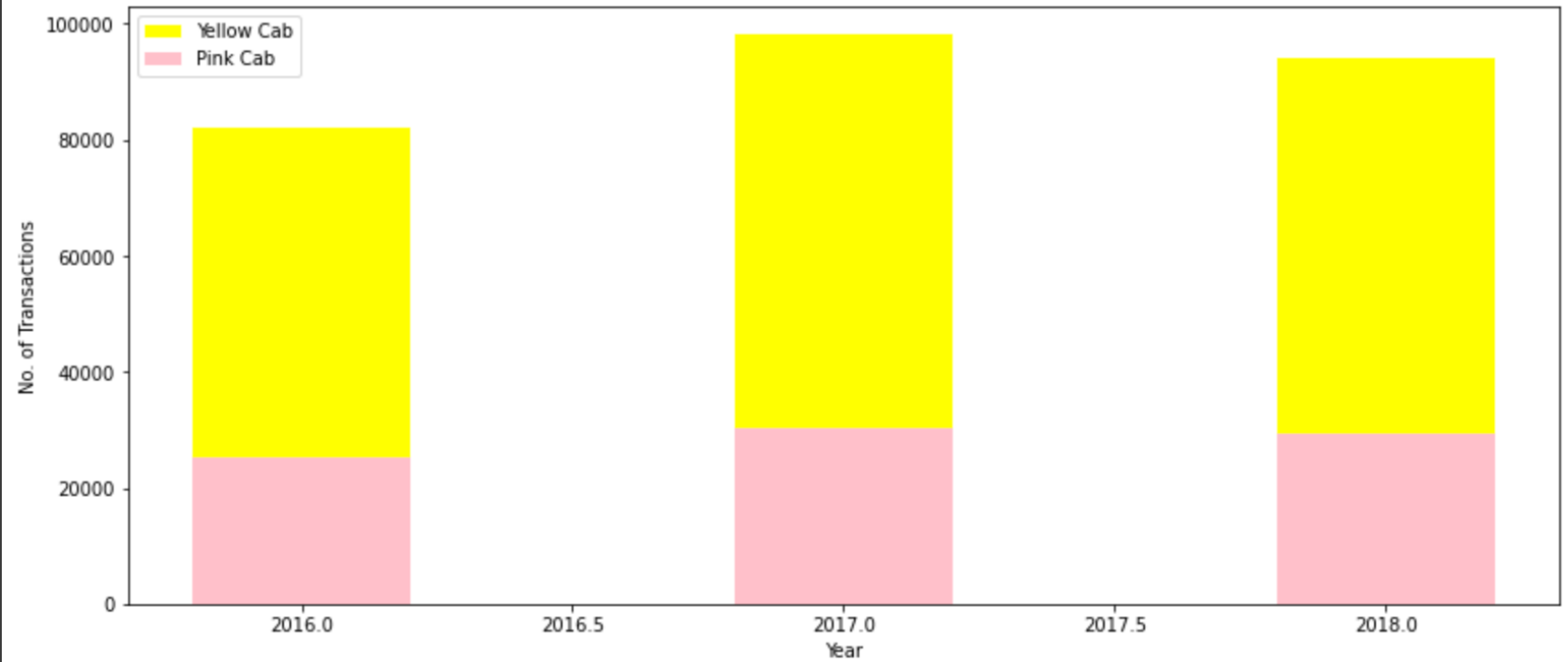
Which
company has
maximum cab
users at a
particular
time period?

- In all months, Yellow Cab sees a higher volume of passengers. In December, it had the most trips (35439), compared to Pink Cab's (11568).

Monthly Travel Distribution



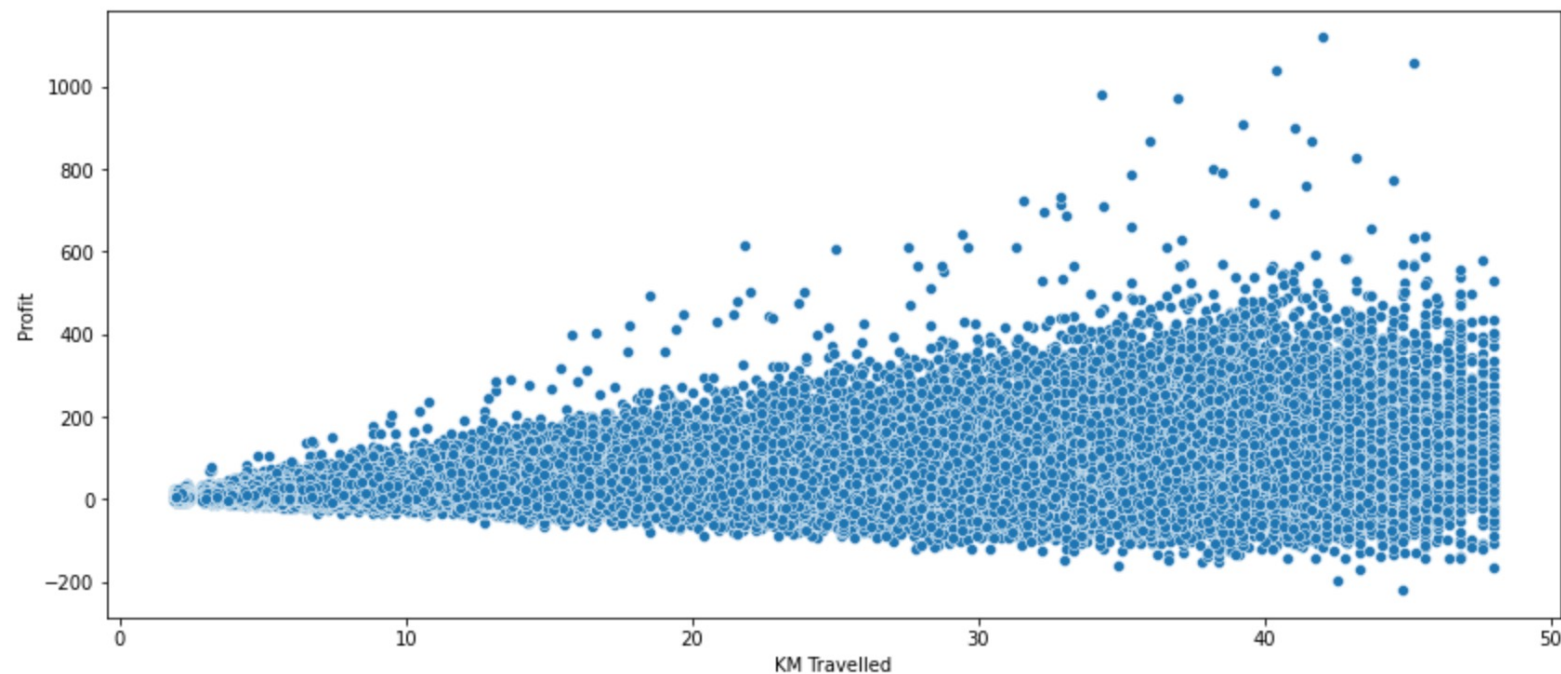
Yearly Travel Distribution



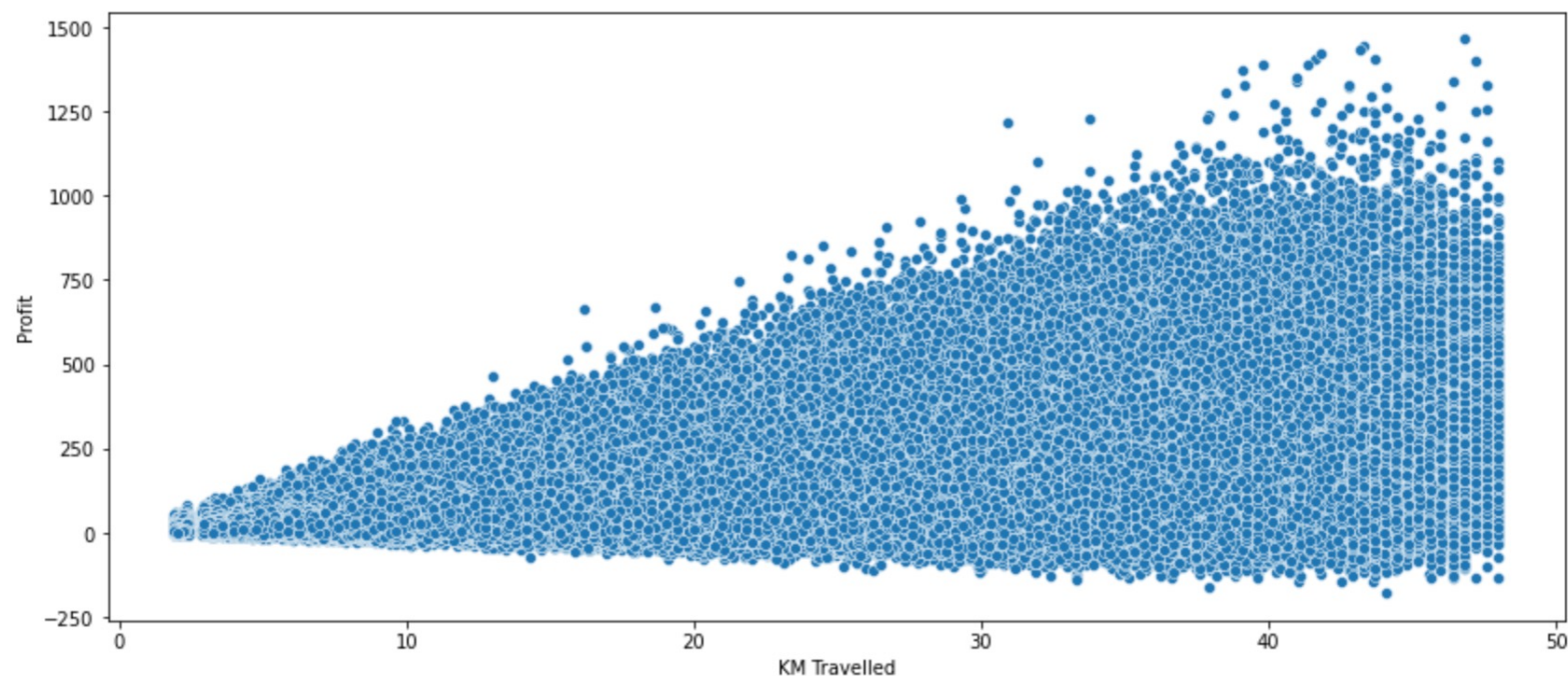
Does margin
proportionally
increase with
increase by
travelled
distance?

- According the following plots, not necessarily the margin will increase based on travelled distance. In some KMs the profit is negative that might be because of the discounts.

Margin based on KM Travelled for Pink Cab

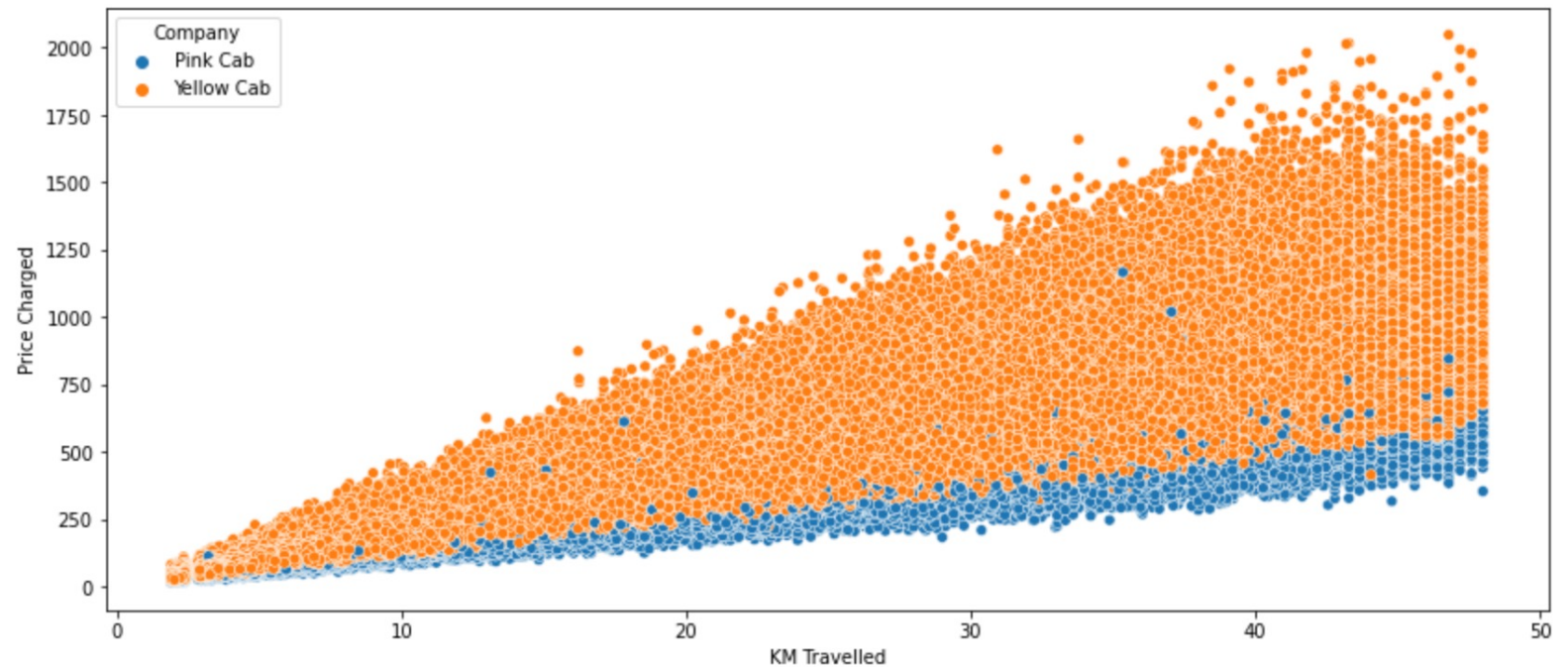


Margin based on KM Travelled for Yellow Cab



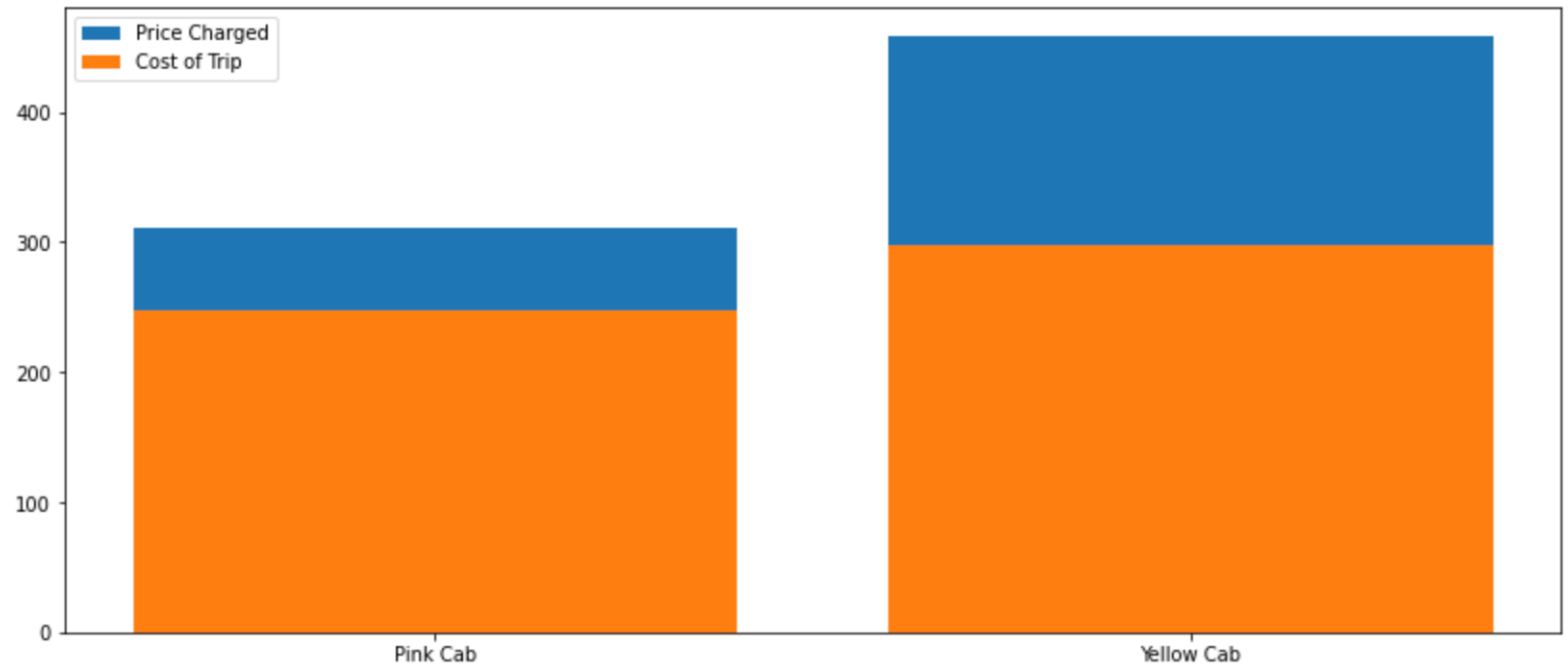
Price Charged based on KM Travelled

- the relationship between KM travelled and the price charged is linear.
- However, Yellow Cab's rates are higher than Pink Cab's.



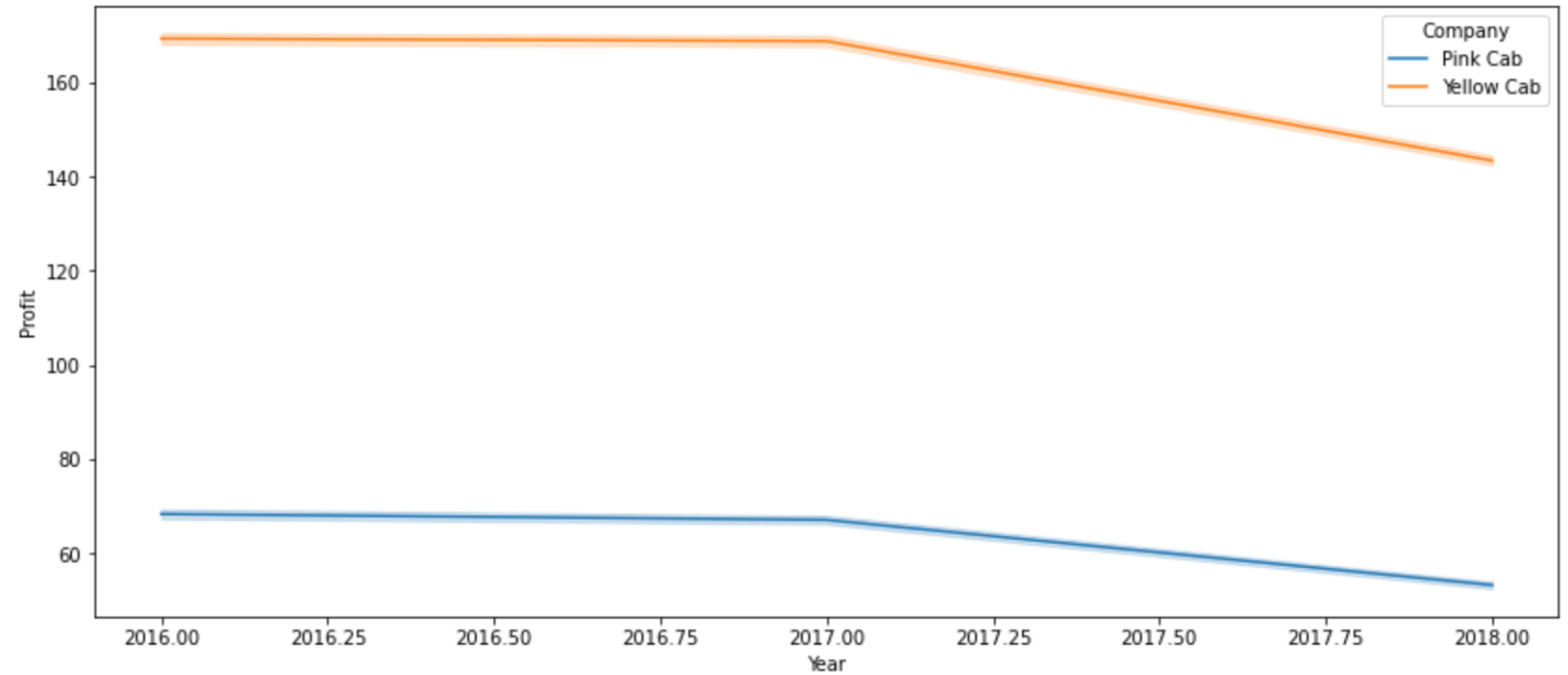
Price Charged based on KM Travelled

- In comparison to the Pink cab, the Yellow cab has a bigger profit margin.



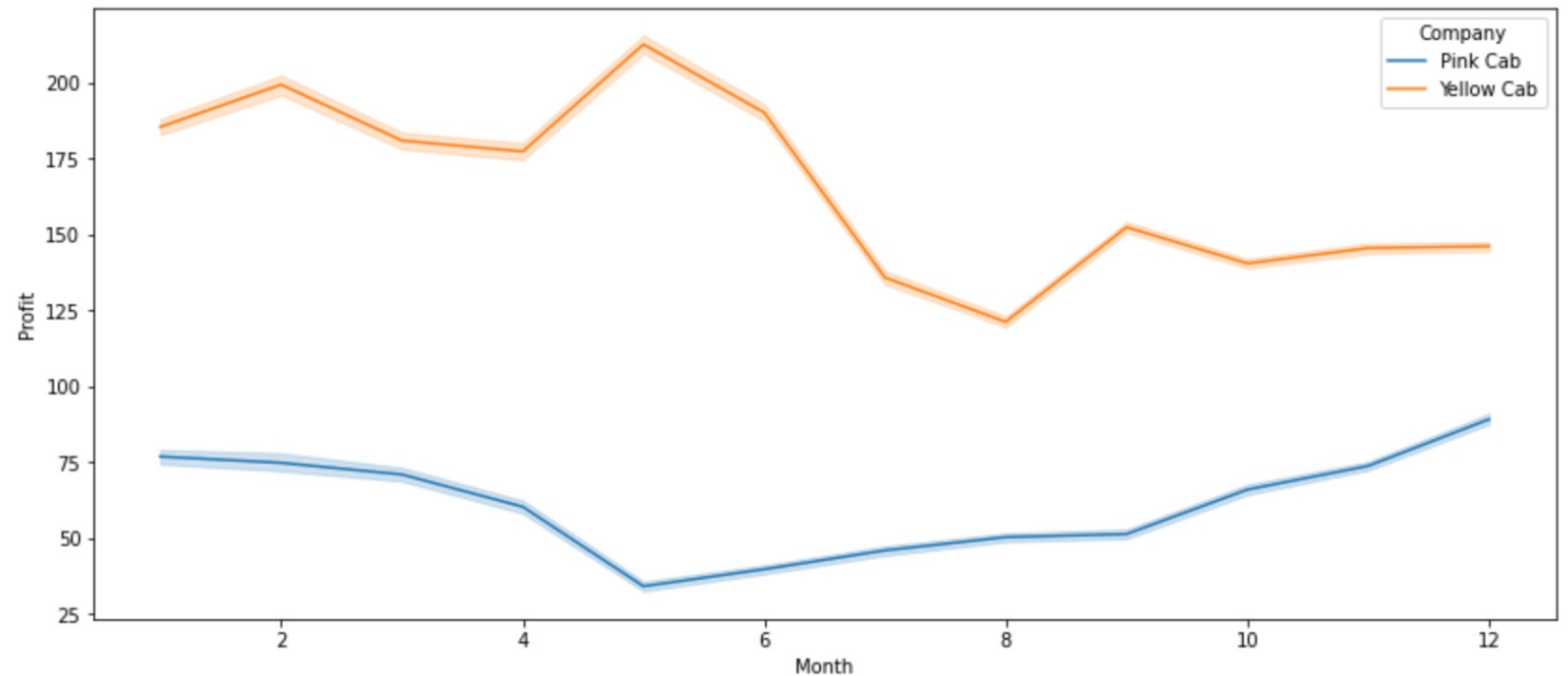
Profit Margin per Year

- The profit margin decreased over years.



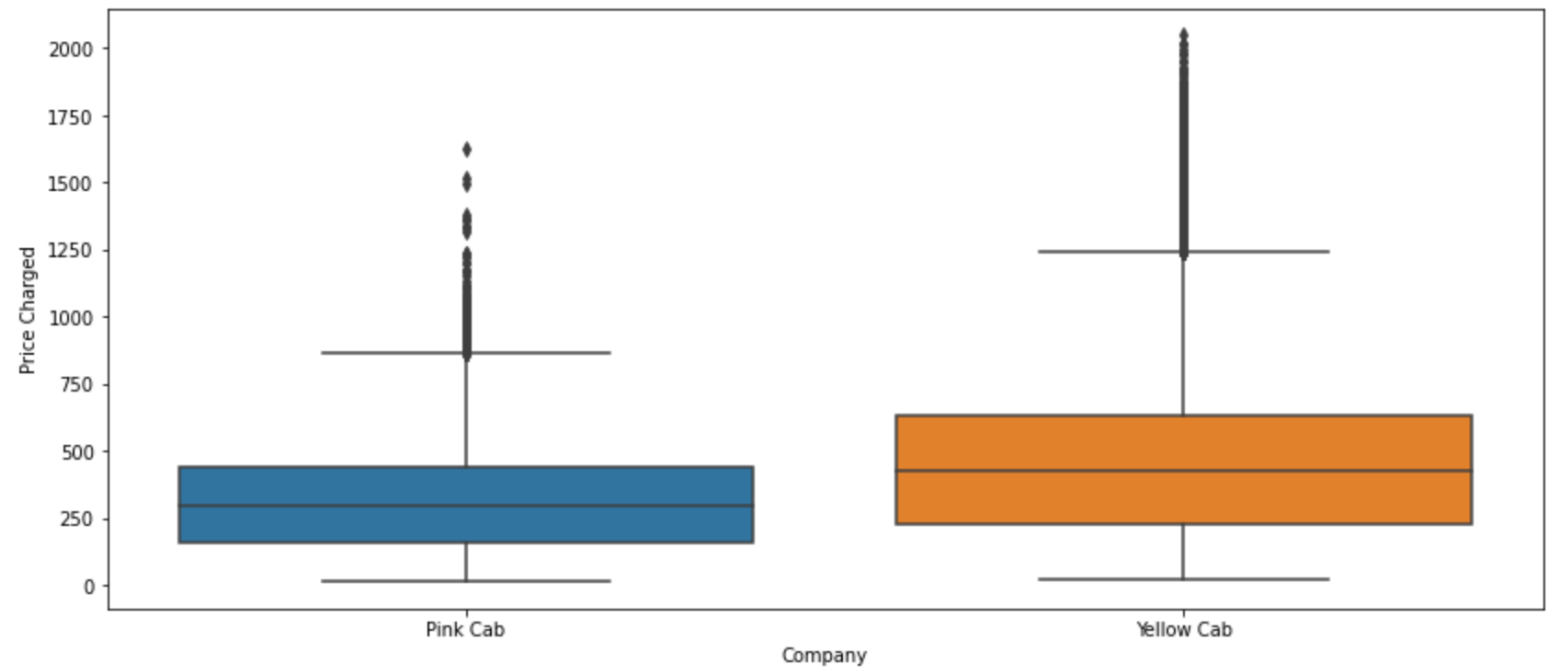
Profit Margin per Year

- Month to month, the profit margin changes.



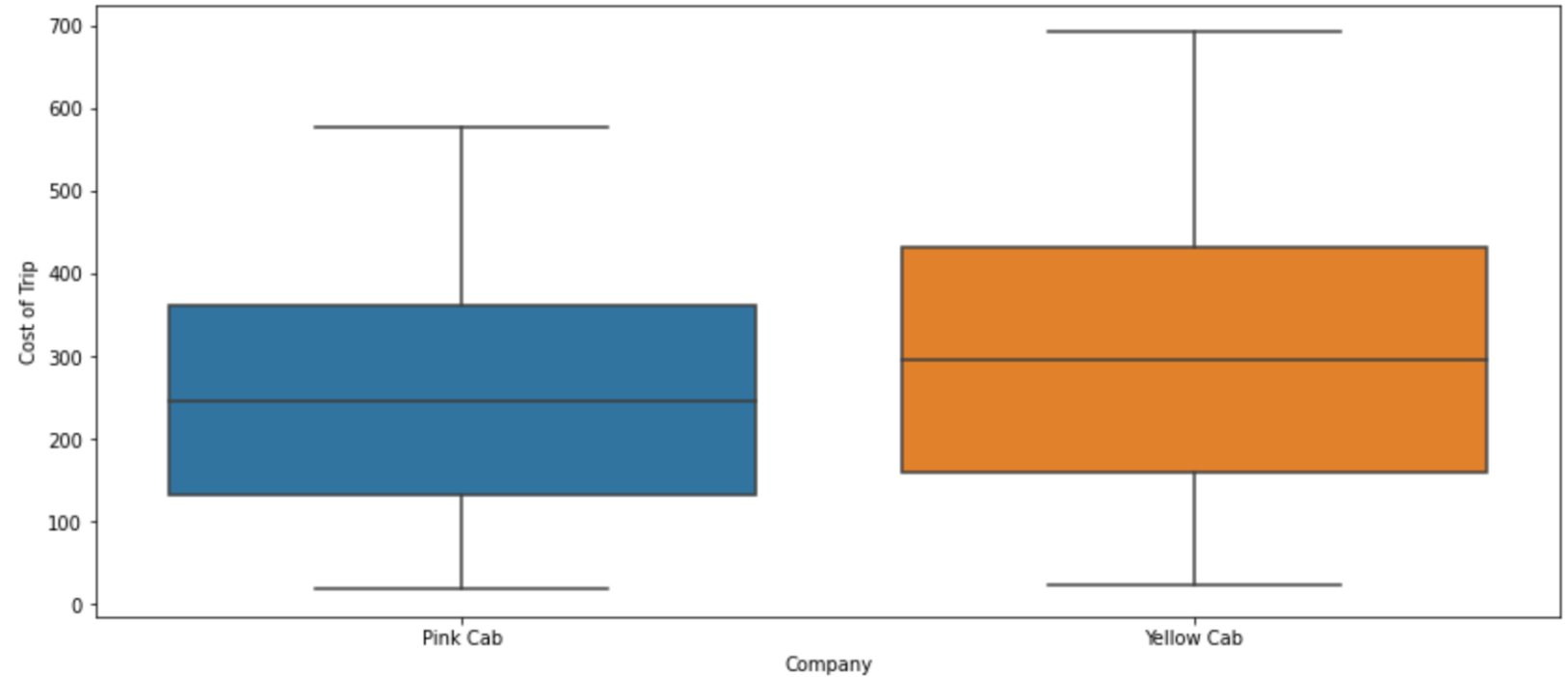
Price Charged

- "Yellow Cab" Price Charged is higher than "Pink Cab" and the average is almost double.



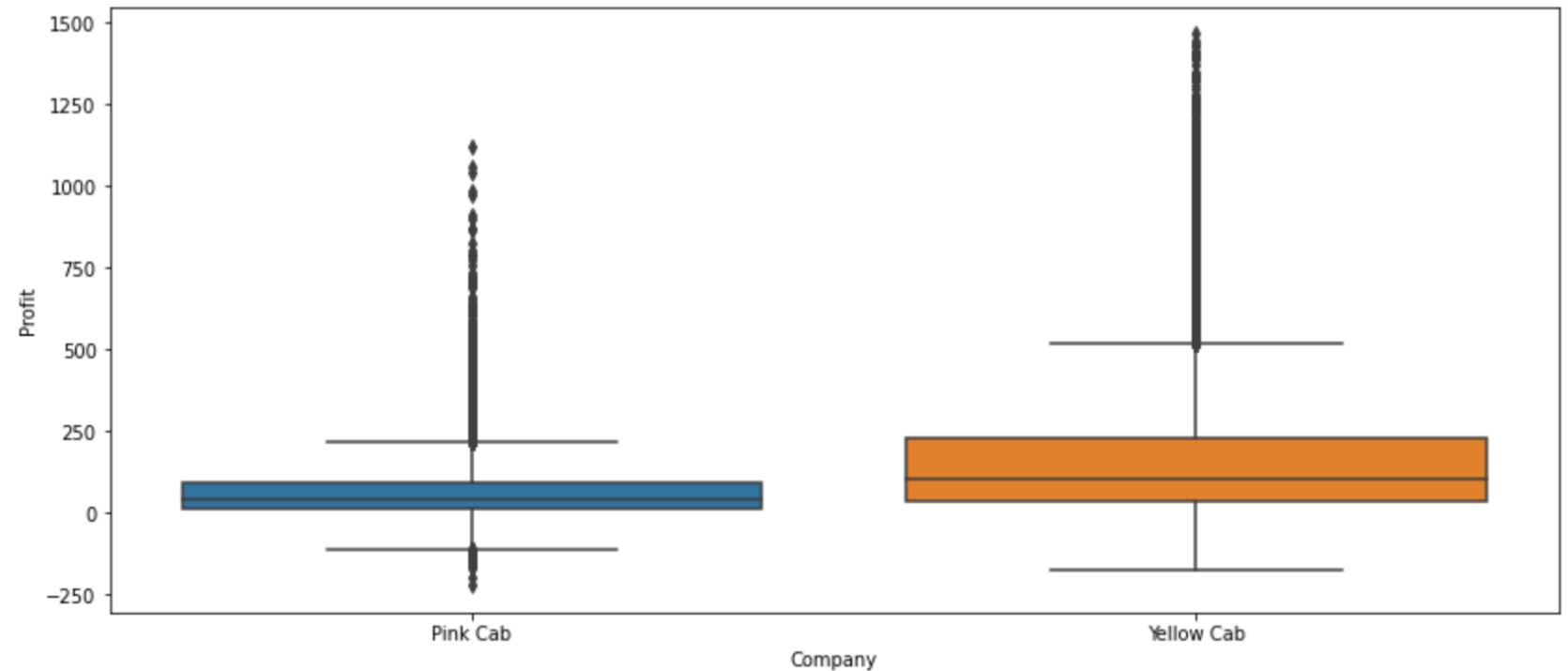
Trip Cost

- Overall "Yellow Cab" trip cost is higher than "Pink Cab".



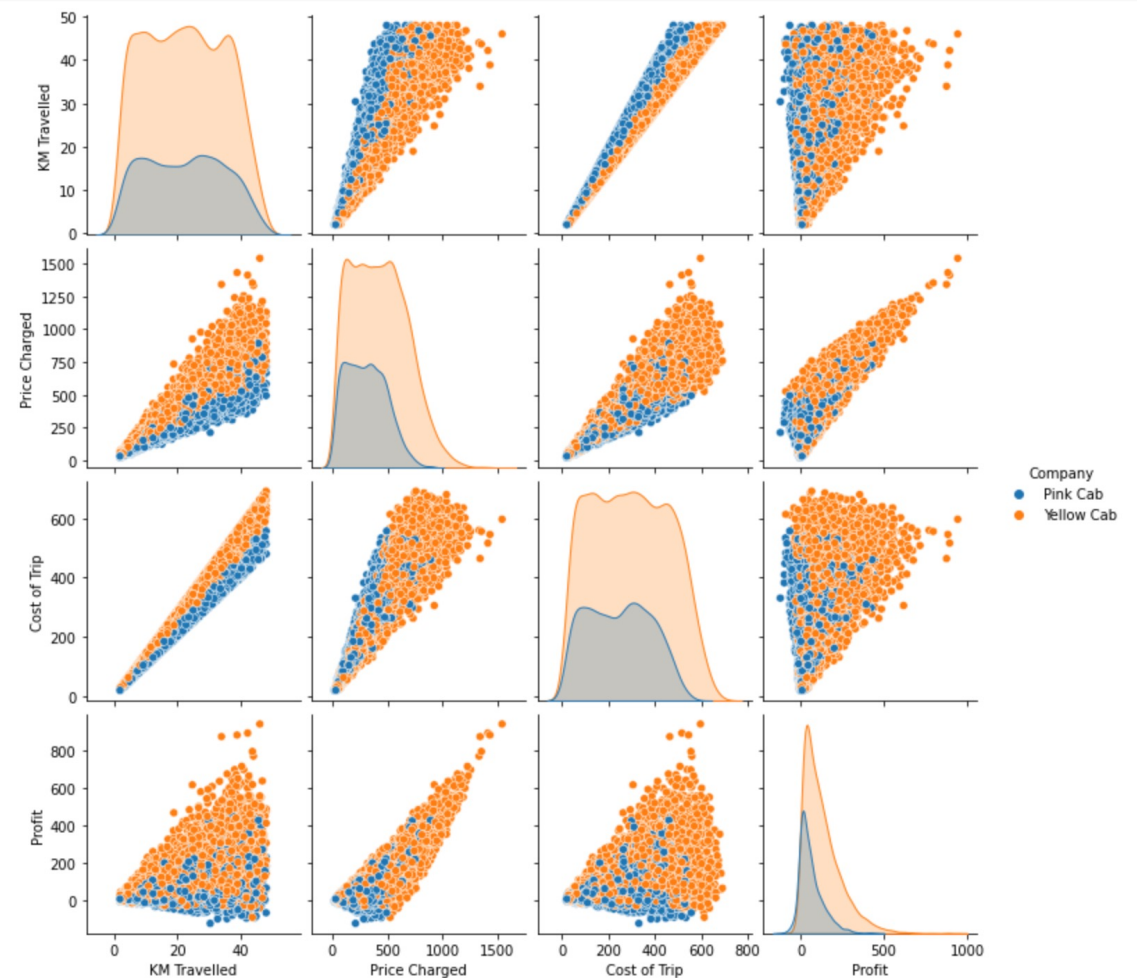
Average Profit

- As a result, the Yellow cab average profit is higher than the Pink cab.



EDA Summary

- Compared to Pink Cab, Yellow Cab is thought to be more profitable.
- The users like yellow cabs over pink cabs.
- The most travel occurs in December, which makes sense given that it's often a month of holidays.



Recommendations

- Yellow Cab Company would be a better investment than Pink Cab Company since it has a higher profit margin, more users, and more transactions annually.

Thank You