

## Introduction

XYZ is a private firm in the US. Due to remarkable growth in the Cab Industry in the last few years and multiple key players in the market, it is planning for an investment in the Cab industry and as per their Go-to-Market(G2M) strategy they want to understand the market before making a final decision. So the aim of this case study is to make analysis and provide actionable insights to help XYZ firm in identifying the right company to make investment.

## Dataset

There are four datasets being provided, which each contains information about Cab companies, population and users in each city, customer information, and transaction information. The time period of the data is from 2016 to 2018. The four datasets are merged into one single dataframe for better analysis.

The raw data contains 14 variables and 359392 observations, and missing variables and duplicates are checked to ensure data representativeness. The variable “Date of Travel” has been rearranged and transformed into actual date time, starting from 2016-01-01 to 2018-12-31, and the variable “Profit” is added to the data as the target variable for the modeling analysis. Profit is calculated by subtracting the price charged for each trip from the cost of the trip.

There are some outliers in the features “Price Charged” and “Cost of Trip”, but because the details of the trips are not available, these values will not be removed as outliers to ensure the integrity of the data.

## Profit Analysis

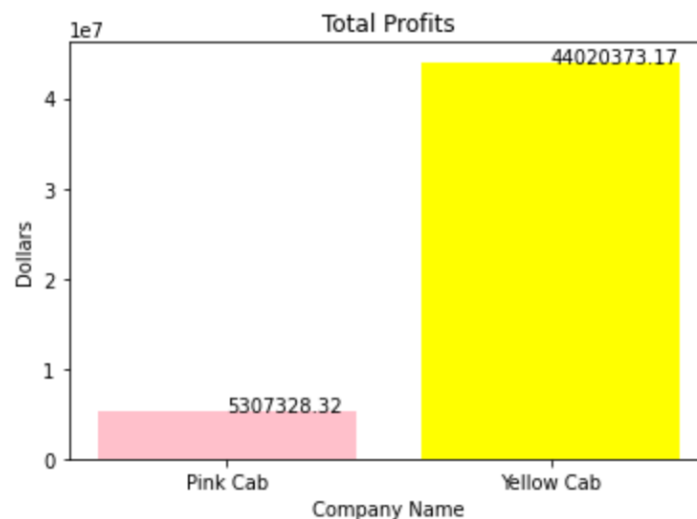


Figure 1.1

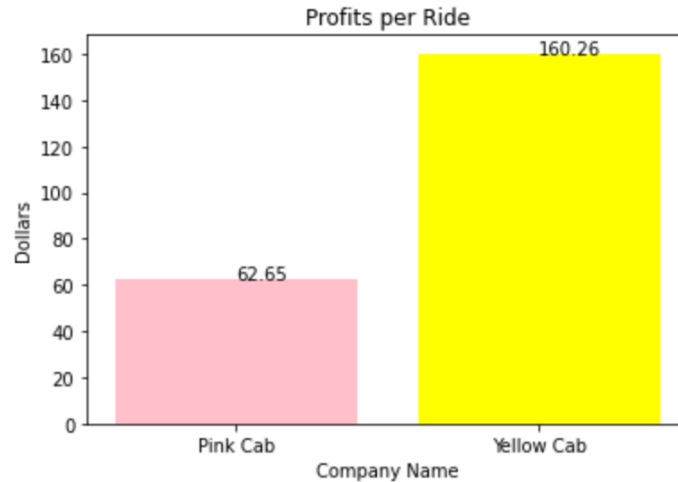


Figure 1.2

Figure 1.1 shows the profits earned for the two companies, with Pink Cab Company earning a total profit of 5,307,328.32 dollars, and Yellow Cab Company earning 44,020,373.17 dollars.

Figure 1.2 is the barplot that compares the profits of each ride between Pink Cab Company and Yellow Cab Company. The figure indicates that Pink Cab Company earns 62.65 dollars per ride on average, while Yellow Cab Company earns 160.26 dollars per ride on average. Ride is defined as the number of travels that occur or the transaction that has been made on record.

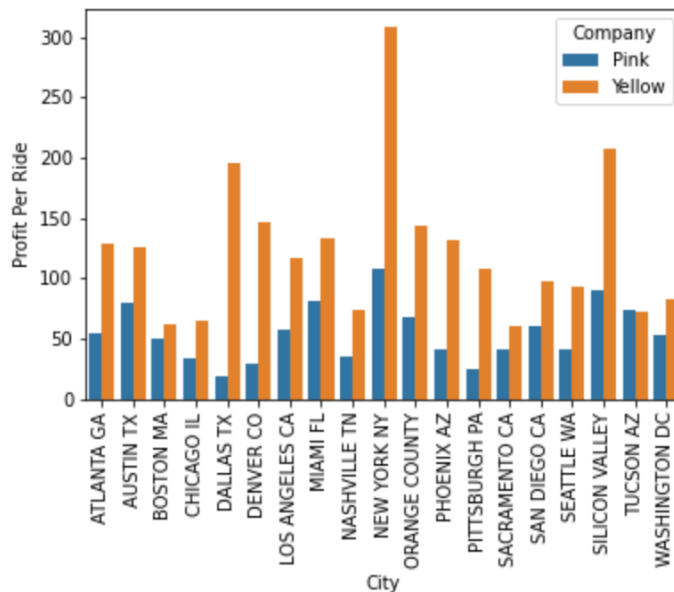


Figure 1.3

Figure 1.3 shows the profits gained per ride for different cities of the two companies, and the plot indicates that Yellow Cab Company is leading in all of the cities except for Tucson.

However, this plot may not be representative, because, from figures 1.1 and 1.2, Yellow Cab Company earns more total profits than Pink Cab Company does, so normally Yellow Cab Company should have higher profits earning in each city as well. A more detailed analysis should be done with the profits percentage, as profit percentage plot can give a better idea of how each city contributes to the total profits of each company.

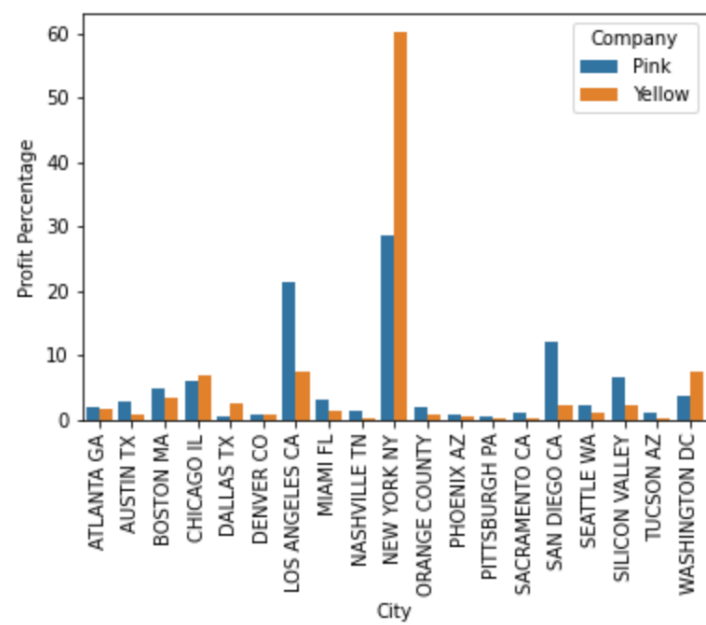


Figure 1.4

Figure 1.4 is the barplot that determines the profit percentage in each city for the two companies, and the profit percentage is the profit percentage of one specific city to the total profits earned by the company. The orange bars represent the Yellow Cab Company, and the blue bars represent the Pink Cab Company. The plot indicates that Yellow Cab Company earns most of its profits from trips in New York, and Pink Cab Company earns most of its profits from Los Angeles and San Diego. This is crucial for the investors if they want to make investments based on cities. More importantly, it evinces that Pink Cab Company may put more resources in Los Angeles and San Diego, while Yellow Cab Company puts more efforts in New York.

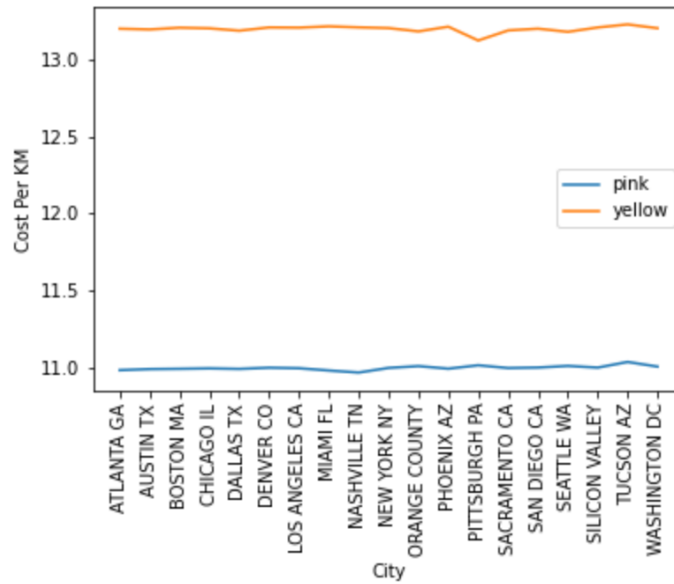


Figure 1.5

Figure 1.5 shows the costs per kilometer traveled in each city, with the blue line representing the Pink Cab Company and orange line representing the Yellow Cab Company. The costs are similar in each city, and Yellow Cab shows a higher cost, above 13 dollars on average, than the Pink Company, which is around 11 dollars. The Yellow Cab Company on average has higher profits, more rides, and higher costs for trips, and the costs may include gas, petroleum, services, costs of the driver, and etc. So a higher cost may indicate that the Yellow Cab Company spends more money on services to customers, treatments to the drivers, and maintenance of the vehicles. Higher costs may also indicate that Yellow Cab Company is larger in terms of business sizes, so they have more drivers and vehicles on duty.

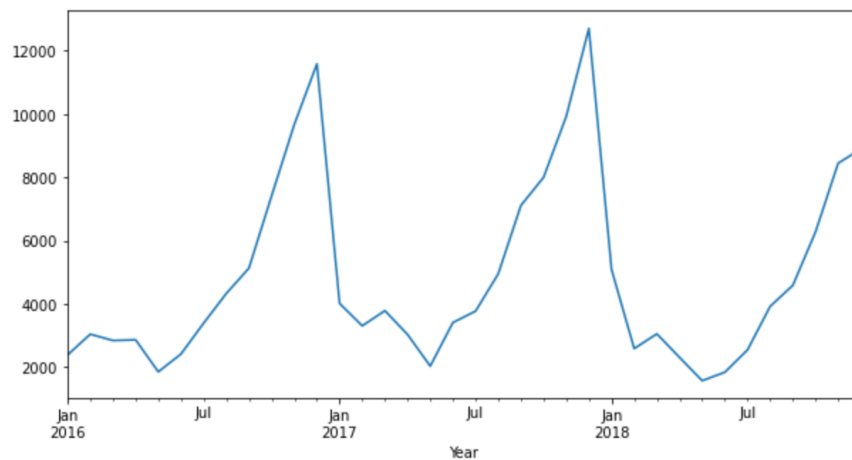


Figure 1.6

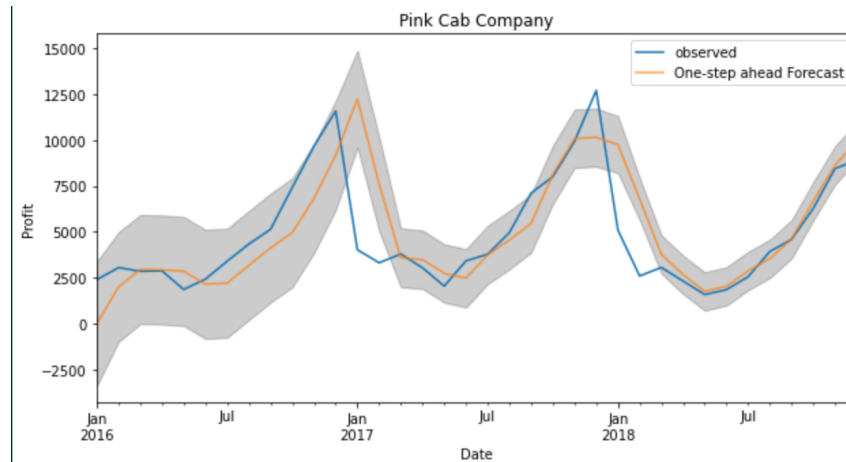


Figure 1.7

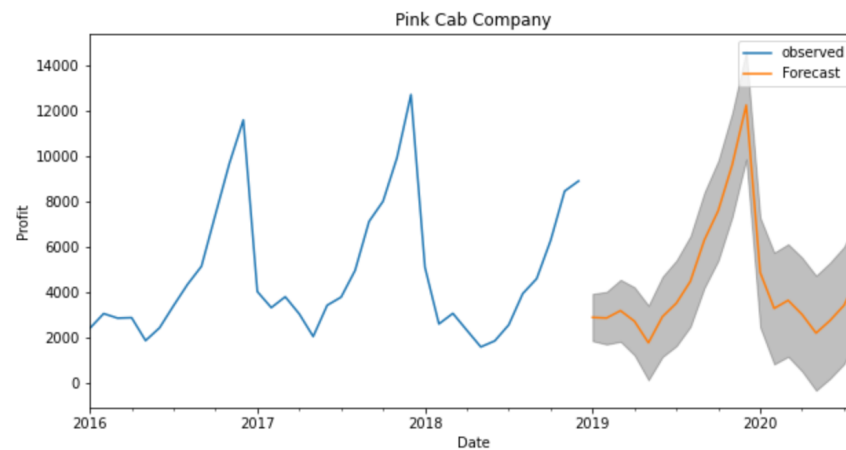


Figure 1.8

Figure 1.6 shows the profit fluctuation of the Pink Cab Company from 2016 to 2018, and it indicates that profit starts to decrease at the beginning of each year and reaches the bottom at the middle of each year, and then it starts to increase and reaches its peak at the end of each year. Then figure 1.7 is the forecast versus observed profit plots, with orange line representing the one-step ahead forecast, and the method used for forecasting is ARIMA, which is an autoregressive integrated moving average, and it is a statistical analysis model that uses time series data to understand the dataset and forecast the future trend. Figure 1.8 shows the 20 step forecasts with the ARIMA model, and it follows the trend of peaking at the end of each year, and the peak is getting higher over years. So it is an overall increasing trend.

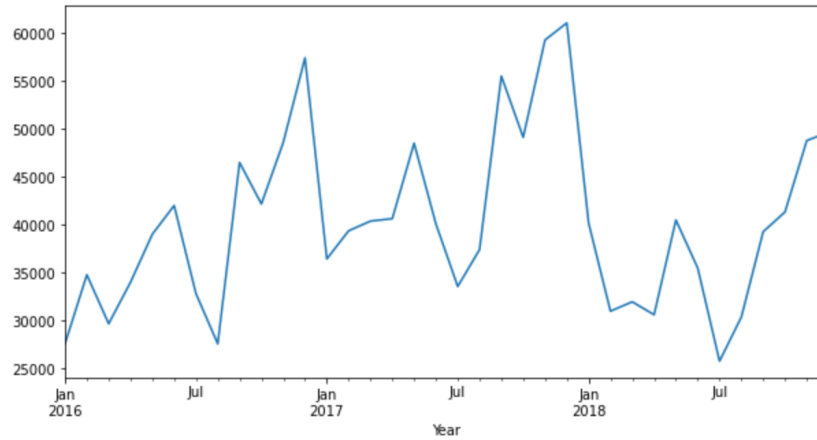


Figure 1.9

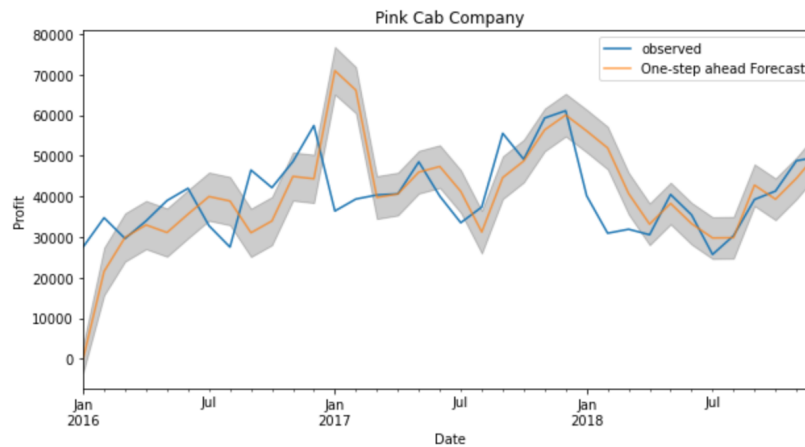


Figure 1.10

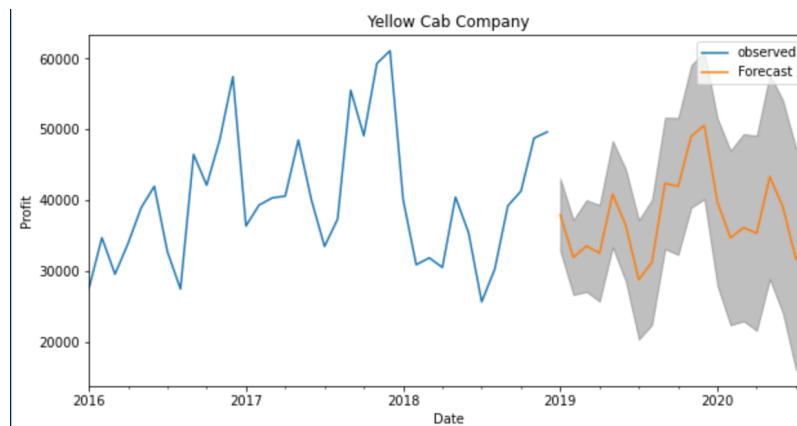


Figure 1.11

Figure 1.9 shows the profit trend for Yellow Cab Company from 2016 and 2018, and the pattern is not clear as the pattern shows for the Pink Cab Company; however, it still shows a relatively higher profit gains close to the end of each year, and it fluctuates around the middle of the year and then starts to increase until the end of the year. Figure 1.10 shows the one-step ahead forecast, and figure 1.11 shows the forecast if the profit for the future two years, and the

forecast shows a decreasing trend in terms of profit for the Yellow Cab Company, even though it has a higher total profit compared to the Pink Cab Company.

## Customer Analysis

	coef	std err	t	P> t	[0.025	0.975]
-----	-----	-----	-----	-----	-----	-----
const	-1.542e+04	6213.679	-2.482	0.013	-2.76e+04	-3235.142
Company	7041.4722	459.724	15.317	0.000	6139.931	7943.014
Customer	163.1934	1.484	109.987	0.000	160.284	166.103
Average KM Travelled	439.6323	136.808	3.213	0.001	171.344	707.920
Average Age	-96.0266	127.346	-0.754	0.451	-345.759	153.706
Average Income	0.0743	0.212	0.350	0.726	-0.342	0.491

Figure 2.1

Figure 2.1 shows the results of the coefficients performed by the linear regression analysis, and the target variable is the profit of each month. Among the predictors, the variable company is a binary categorical variable with 0 being the Pink Cab Company and 1 being the Yellow Cab Company. The variable customer is the average number of customers for each company by months. The average kilometers traveled variable is calculated by companies and every month of the year. The calculation is the same for the average age and average income of the customers. The results indicate that Yellow Cab Company will gain 7041.47 more dollars in terms of profits each month, and an increase in one customer will increase the profit by 163.19 dollars each month. Obviously, more kilometers traveled, more profits gained. Whereas, younger customers bring more profits, with a 96.03 dollars decrease on average for one-year growth of the customer. The variable income seems to have a little impact on the profits analysis.

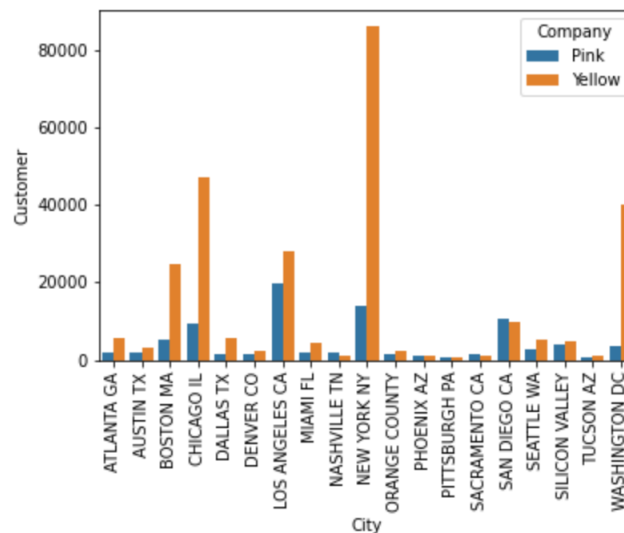


Figure 2.2

Figure 2.2 shows the customer presence of Pink Cab Company and Yellow Cab Company city wise, and based on the plots, Yellow Cab Company has more customers in total compared to the Pink Cab Company. Yellow Cab Company has customers concentrated mainly in New York

and Chicago, while Pink Cab Company has customers concentrated in Los Angeles and New York. Other than this, Pink Cab Company has more customers in San Diego compared to Yellow Cab Company.

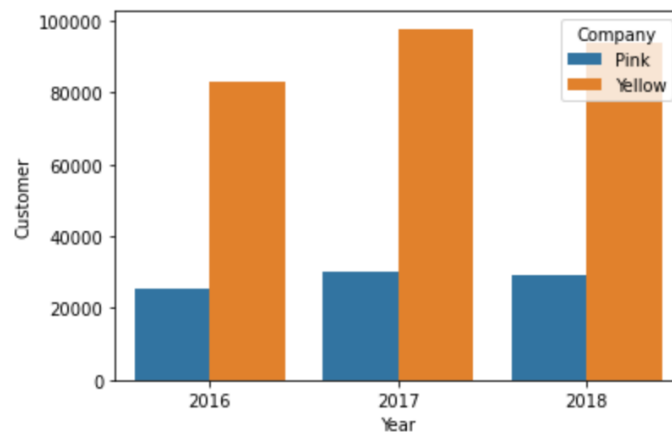


Figure 2.3

Figure 2.3 shows the customer base of Pink Cab Company and Yellow Cab Company year wise, with orange bar represents the Yellow Cab Company and blue bar represents the Pink Cab Company, both companies have more customers in 2017 compared to the other 2 years, and the customer base has decreased slightly in 2018.

### Recommendation

Yellow Cab Company obviously gains more profit compared to Pink Cab Company, even though Yellow Cab Company shows a decreasing trend in the future two years, the pattern is relatively stable compared to the fluctuation of Pink Cab Company. However, the costs per kilometer traveled for Yellow Cab Company is higher than Pink Cab Company, but the detailed number is not provided, so the purposes of the costs are not examined. Even though the costs are high, the profits will cover these costs. Yellow Cab Company mainly earns its profits in New York, and Pink Cab Company mainly earns its profits in Los Angeles and San Diego. However, Yellow Cab Company shows more customers in most of the areas except for San Diego. Moreover, the regression analysis indicates more customers lead to more profits, and Yellow Cab Company has more customers over year compared to Pink Cab Company. Therefore, I would recommend XYZ company to invest in the Yellow Cab Company in general for more profits. If XYZ company wants to make investments area wise, I would recommend XYZ company to invest in the Yellow Cab Company if they want to mainly focus on New York, and if they want to put more effort in Los Angeles or San Diego, I would probably recommend them to invest in the Pink Cab Company.