



G2M Case Study Virtual Internship

Soulaiman KACHA



Background

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.

- Objective : Provide actionable insights to help XYZ firm in identifying the right company for making investment.

The analysis has been divided into three parts:

- Data Understanding
- Forecasting profit and number of Users
- Finding the most profitable Cab company



Data Exploration

- 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

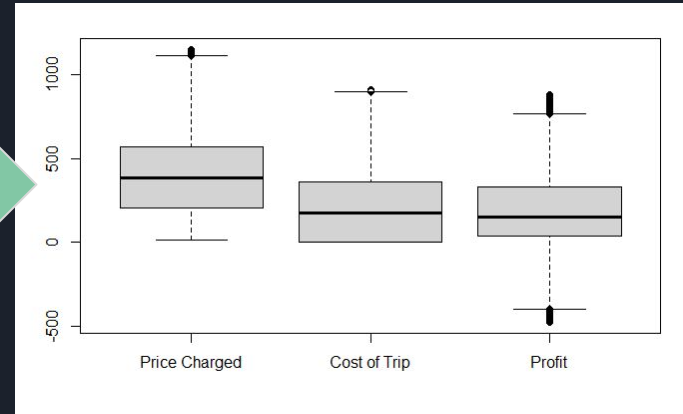
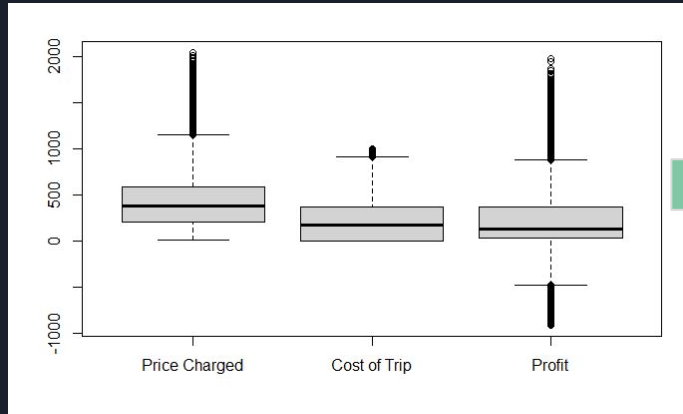
After exploring the four datasets understanding the names of the columns and identifying the relationships between them, we proceed to create a master dataset joining and ordering all the four previous dataset we have.



Data Exploration

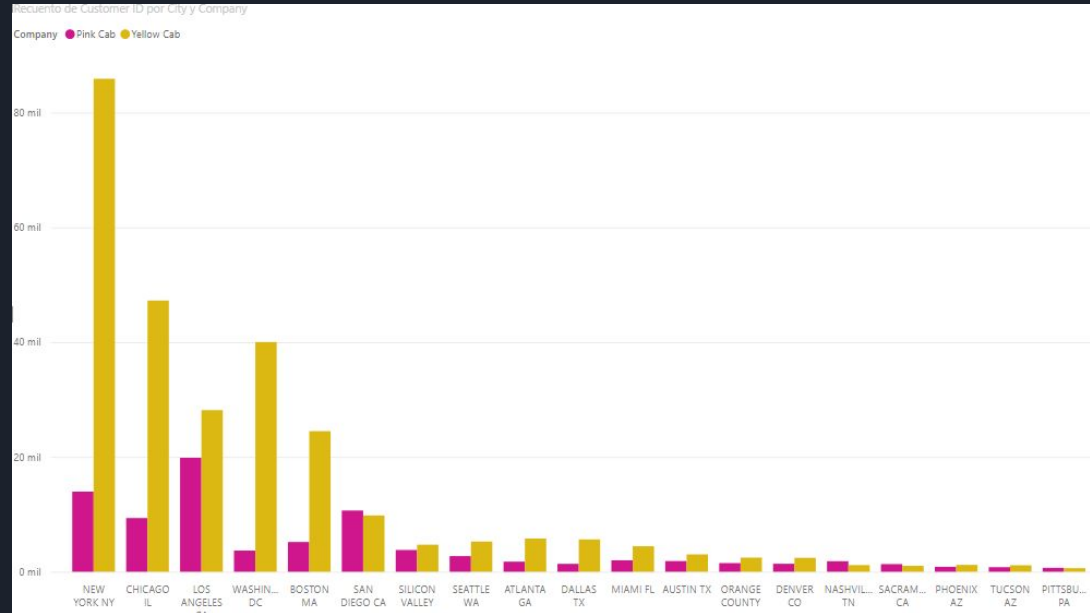
- Transaction ID and Customer ID: variables that are only used to identify each record, and that almost certainly will never be used in any statistical analysis (name or ID number of a person, for example).
- Gender: Qualitative Binary variable with two levels of the factor (Male/Female).
- Age: Quantitative Discrete variable
- Income: Quantitative Continuous variable. Income per month of each client in USD
- Company: Qualitative Variable (Binary). The Pink cab company and yellow cab company.
- City: Qualitative Nominal Variable. The cities where the two cab companies operate. 20 factors for this variable.
- Population: Quantitative Discrete variable. Number of people of each city.
- Users: Quantitative Discrete variable. Number of Users of the cab companies of each city.
- Date of travel: Quantitative continuous variable
- Km travelled: Quantitative continuous variable. KM travelled in each trip made.
- Price charged: Quantitative continuous variable. Price in USD charged in each Trip.
- Cost of Trip: Quantitative continuous variable. Cost in USD of each Trip.
- Payment Mode: Qualitative Binary variable. Method of payment (Cash/Credit card)

Outliers detection and treatment



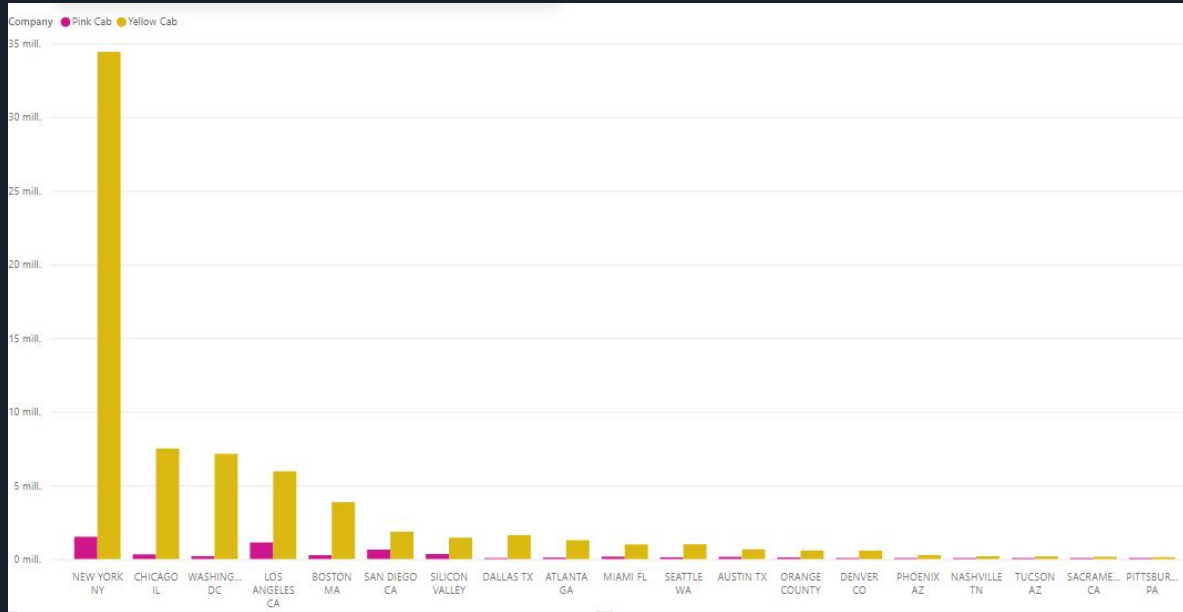
We found outliers for the variables Price charged, Cost of the trip and the variable we created called Profit.

Customer Presence of Yellow and Pink cab city wise



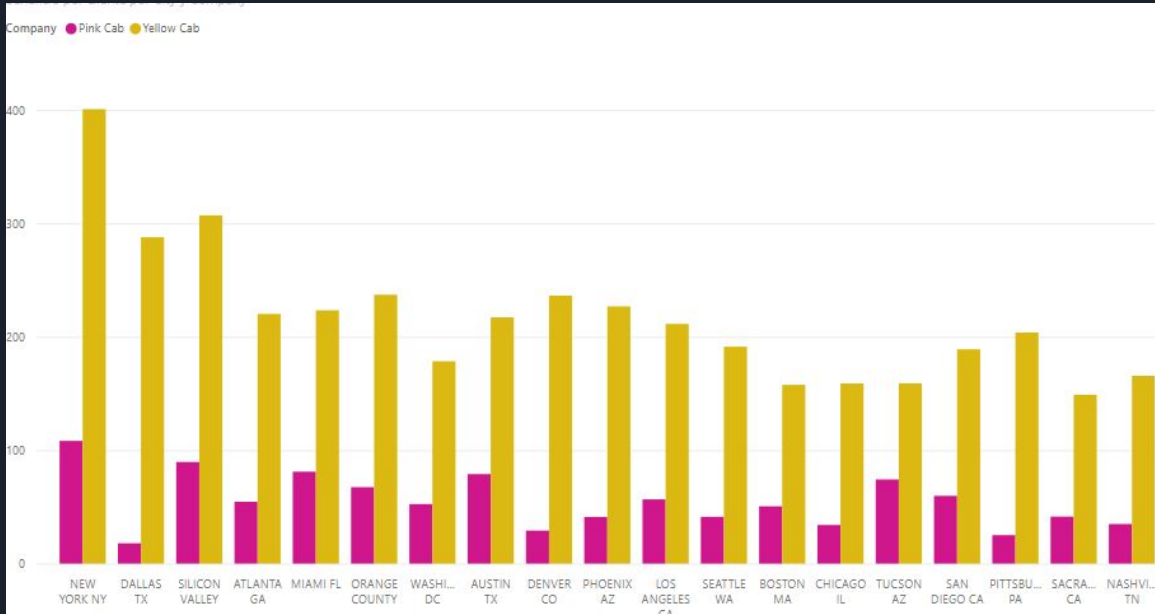
- We found that the Yellow company has way much customers per city than the pink company.
- New York, Los angeles and Washington are the cities with more Users for the yellow company
- Los angeles, New york and San Diego are the cities with more Users for the Pink company

Profit of Yellow and Pink cab city wise



- Yellow company seems to have a bigger profit per city than Pink company.
- The number of users are different for each company
- We have to study also the profit per user for each city

Profit per User of Yellow and Pink cab city wise

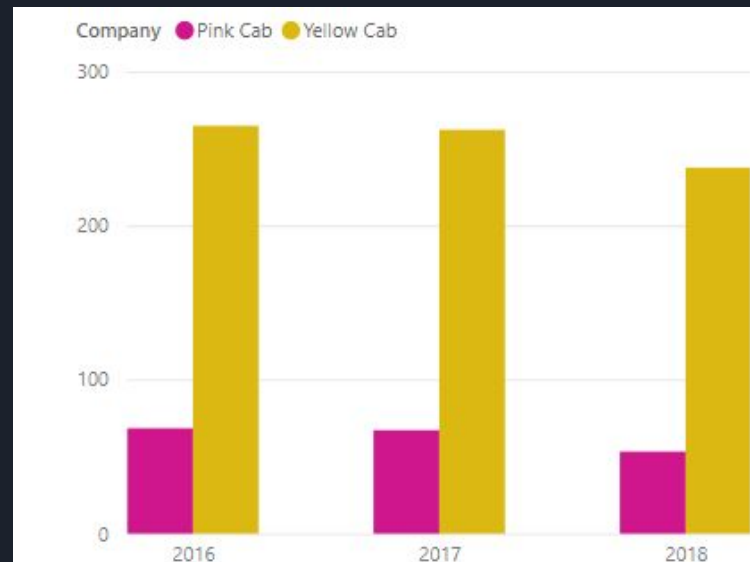


- Indeed the yellow company generates more Profit for each user than the Pink company

Profit and profit per client by Company Year wise

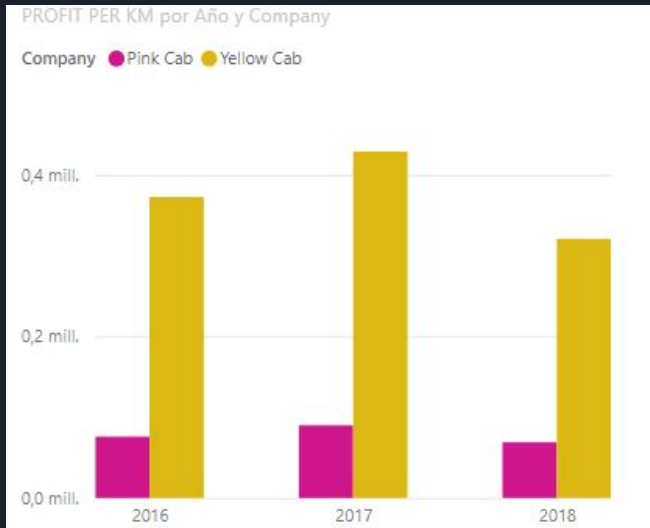


Profit by company Year wise



Profit per client by company year wise

Profit per KM year wise

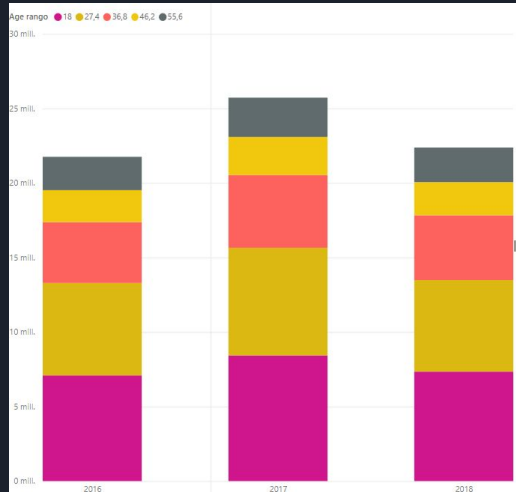


Profit per KM year wise

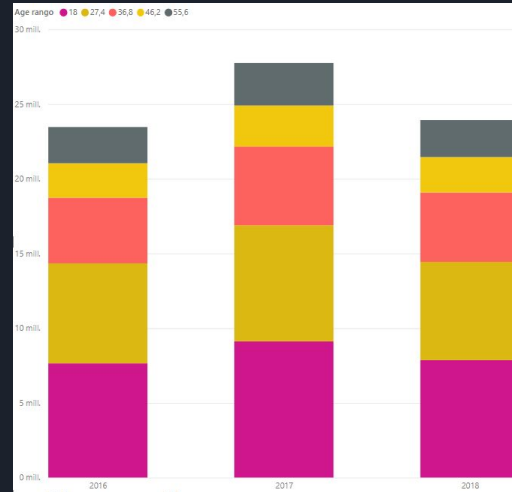
Año	Company	Profit	PROFIT PER KM	Beneficio por cliente
2016	Yellow Cab	21.753.828,25	372.609,16	264,52
2017	Yellow Cab	25.724.399,82	428.960,36	261,99
2018	Yellow Cab	22.374.973,54	320.935,63	237,39
2016	Pink Cab	1.713.511,22	75.912,47	68,32
2017	Pink Cab	2.033.654,91	89.837,57	67,07
2018	Pink Cab	1.560.162,19	68.891,61	53,23
Total		75.160.529,93	1.357.146,80	209,13

- Yellow company has bigger profit year wise then pink company
- Yellow company has bigger profit per Km year wise then pink company
- Yellow company has bigger profit per client year wise then pink company

Age GroupWise Profit And Customer Base



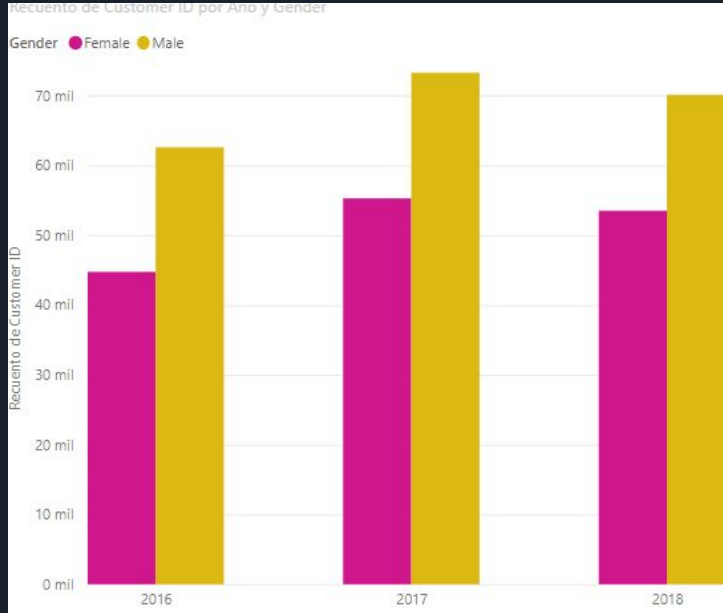
Yellow company



Pink company

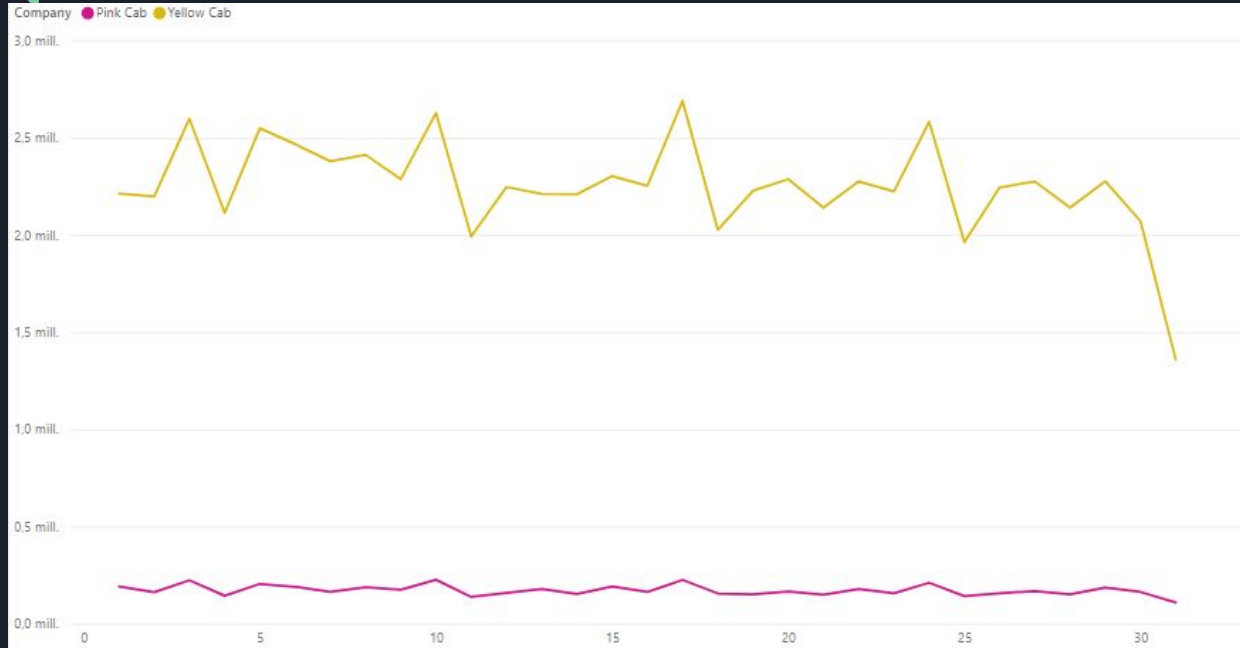
The users who most use both companies are in the age ranges between 18 and 35 years. As the age ranges increase, the use of both companies decreases, being the age range <65 the one that uses the least

Customer base Analysis Gender wise



- It seems that the gender that most uses transport services are men

Seasonality

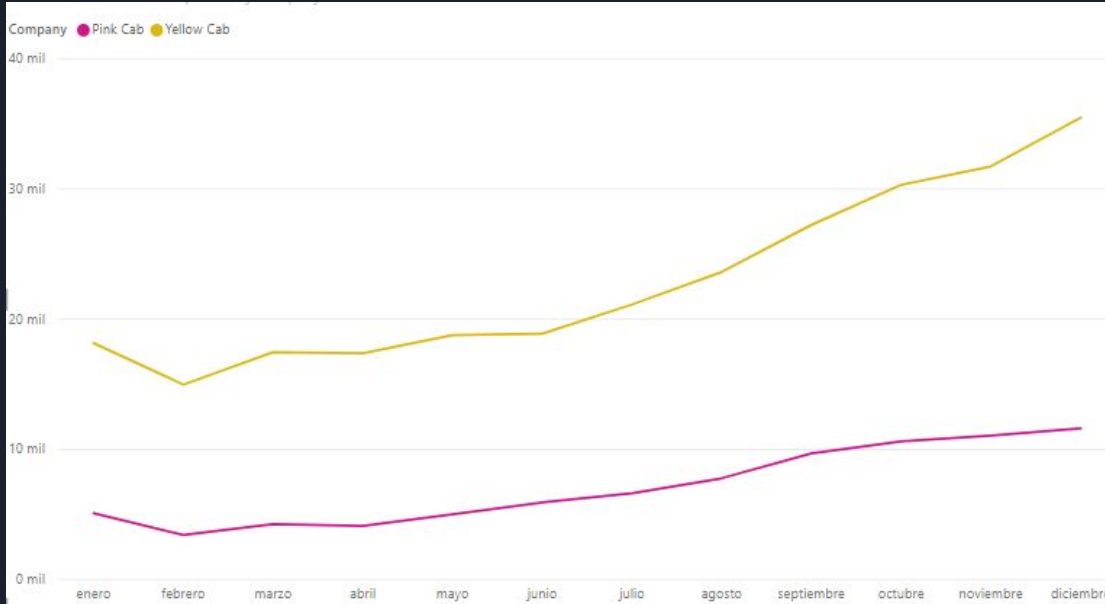


There seems to be a clear seasonality.

The pattern is repeated every 7 days where one of the days has the highest peak followed by the lowest peak.

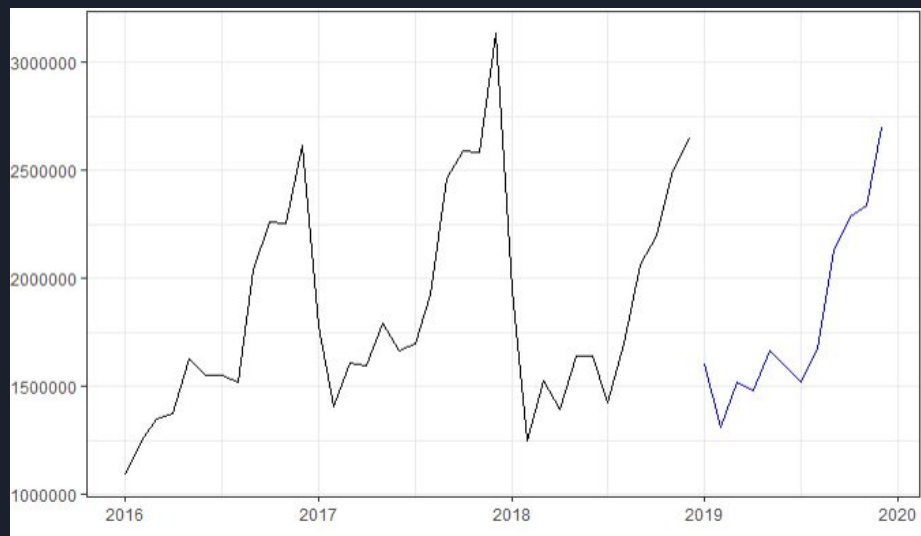
This makes us think of a very high use on Sundays, followed by a decrease on Monday,

Year production



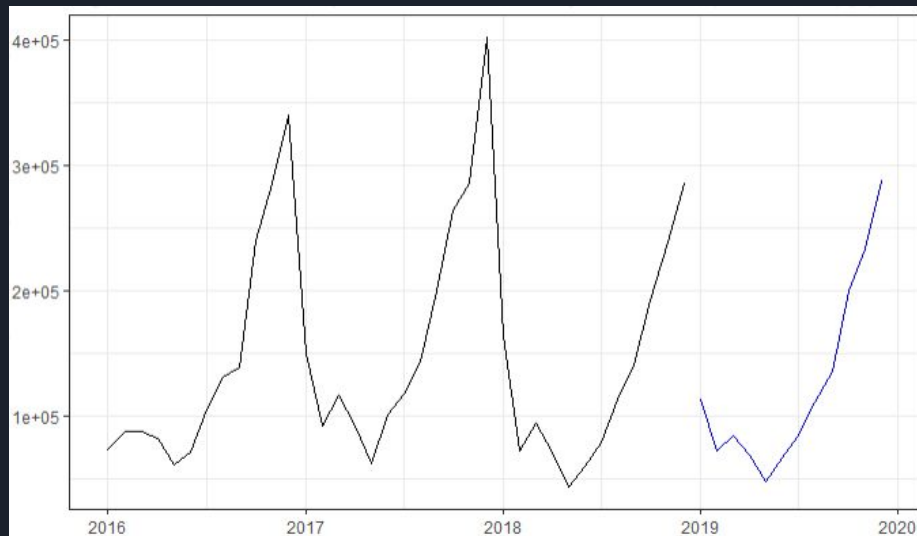
We can see that the months with the lowest production are those of the first four months of the year, especially February. The number of users increases in the second quarter and sees its greatest exponent in the last quarter of the year

Profit forecasts Yellow cab



	Point Forecast	Lo 95	Hi 95
Jan 2019	1603809	1267353	1940264
Feb 2019	1308462	1019557	1597367
Mar 2019	1520560	1168857	1872263
Apr 2019	1482863	1124967	1840759
May 2019	1664004	1246306	2081702
Jun 2019	1596642	1180975	2012308
Jul 2019	1521762	1111884	1931641
Aug 2019	1672033	1207091	2136975
Sep 2019	2131130	1520476	2741784
Oct 2019	2283516	1610389	2956644
Nov 2019	2339586	1631162	3048010
Dec 2019	2700728	1861818	3539638

Profit forecasts Pink cab



	Point Forecast	Lo 95	Hi 95
Jan 2019	112951.48	75989.44	149913.52
Feb 2019	72727.08	47517.07	97937.09
Mar 2019	84195.86	53458.59	114933.12
Apr 2019	68665.31	42389.25	94941.37
May 2019	47794.14	28698.06	66890.22
Jun 2019	66747.97	38994.42	94501.52
Jul 2019	85025.21	48338.25	121712.16
Aug 2019	111300.72	61586.14	161015.29
Sep 2019	135996.48	73247.31	198745.66
Oct 2019	199739.24	104717.83	294760.65
Nov 2019	233457.63	119138.53	347776.73
Dec 2019	287999.40	143053.13	432945.68



Recommendations

After conducting an exhaustive analysis, we have determined that the yellow company is more suitable to invest in it for the following reasons:

- Customer Reach : Yellow cab has higher customer reach in cities then Pink cab.
- Generated profit: The Yellow cab company generates more profit in most cities and over the years than the pink company
- Profit per KM: Yellow cab manage to generate more profit per Km than Pink cab
- Future Profit: Forecast of future profit shows the Yellow cab is going to generate much more profit than pink cab