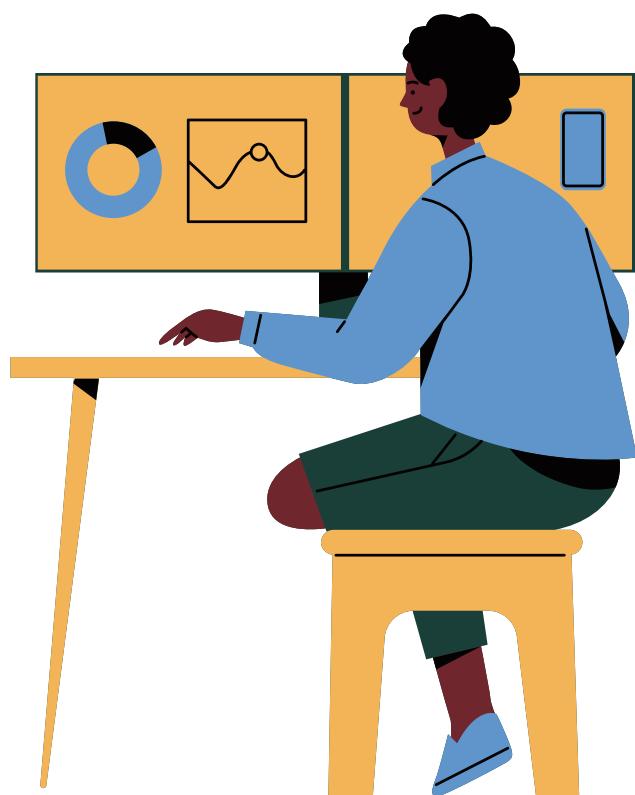


# DATA ANALYSIS PORTFOLIO

By Korada saikiran



# CONTENT DESCRIPTION

- 1 Professional Background
- 2 Data Analytics
- 3 Instagram user analytics
- 4 Operation analytics and investigating Metric Spike
- 5 Hiring Process Analytics
- 6 IMDB movie analysis
- 7 Bank Loan Case Study
- 8 Impact of Car Features



9

ABC Call Volume Trend

Conclusion

10

11

Appendix





# Professional Background

Currently, I'm pursuing third year of bachelors in technology, in Electronics & communication branch. Apart from my academic, my area of interest is a Data science. I really like to deal with data, analyzing data. And also like to create Machine Learning models using predictive & statistical analytics. Also I have hands on web-development which really helps to present projects or myself with ease.

In my current role as a Data Analytical trainee, I have developed a wide range of skills in data analysis. Before this training, I've explored the data using traditional methods like using pandas to perform Exploratory Data Analysis. As a data analyst, one should be well known for his skill to use analytical tools like SQL, Excel, PowerBI, tableau. In the training period I'm very much thankful that I have got to know about these analytical skills and now I'm well equipped with these tools.

I also got a chance to use and apply these taught skills in projects that have been assigned.

I've been assigned a total of 8 projects with which I got a hands on experience on the analytical skills. I demonstrated my ability to use data science to solve complex problems and drive business outcomes.

I am also passionate about staying up-to-date with the latest trends and technologies in data analytics and data science, and I am constantly seeking to learn and grow in my field. I believe that I am a good learner and also I am working to upskill myself to get ready for the Data Analytics role. I am learning and even ready to learn more, to enhance my skills and knowledge.

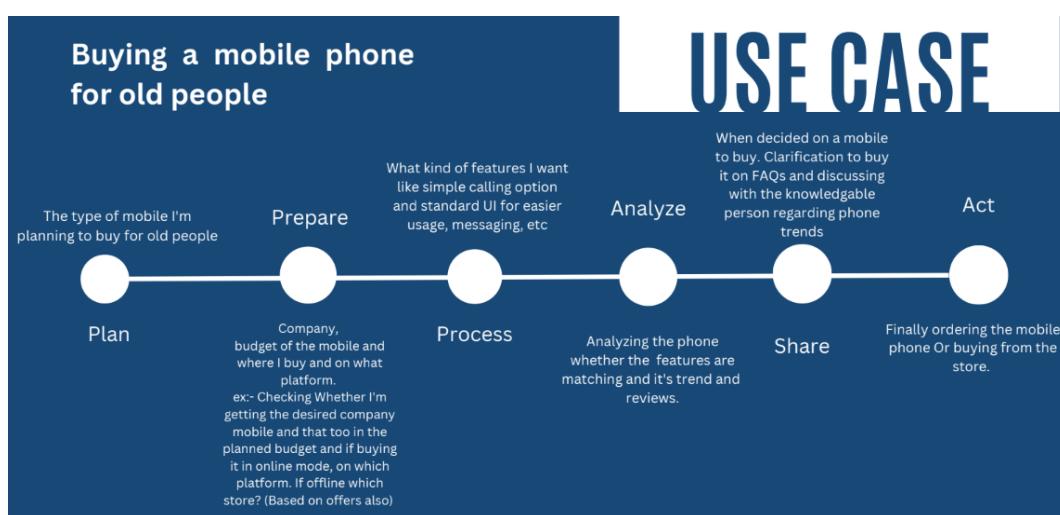
# Project - 01

## Data Analytics



### Project description

Data analytics is vast, and should be applied in a proper way in order to deal with a business problem. Any problem can be broken into chunks and can be solved. Here its a simple task, where we should observe the daily life problems where analysis is required and should provide the solution how we solve it using analytics step wise



## Insights & Findings

Most of the real-life problems require a proper analysis in order to solve it. Most of the Business problems require data analysis. This is where Data analysts come into picture. Got to know the requirement of data analysis in every domain and the exact way of solving a problem using analytics



## Project -02

# *Instagram user Analytics*

## Project Description

User analysis is the process by which we track how users engage and interact with our digital product (software or mobile application) in an attempt to derive business insights for marketing, product & development teams. Engagement of the Instagram users has to be analyzed and derive insights which helps to improve the business growth of the app. Determining User satisfaction and engagement of user in the app will help us further improvement of the application. Here in this project, SQL has to be used to extract the data and understand it in order to provide the insights.



## Findings & Insights

The Instagram's data is analyzed, many insights which improve the business of the app have been derived like no. of posts of a user, how often he is posting, how active he is, if he/she isn't making them active by sending them mails. By these we can explain the investors about the performance of the app. I have gained a lot of sample knowledge regarding how the analysis of the Instagram can be made using SQL especially.

The given questions are solved using the SQL queries and the required analysis has been made



## Some of the findings:

### Most loyal users

	id	username	created_at
▶	80	Darby_Herzog	2016-05-06 00:14:21
	67	Emilio_Bernier52	2016-05-06 13:04:30
	63	Elenor88	2016-05-08 01:30:41
	95	Nicole71	2016-05-09 17:30:22
*	38	Jordyn.Jacobson2	2016-05-14 07:56:26
	NULL	NULL	NULL

### Top 5 #Hashtags

tag_name
beach
beauty
concert
delicious
dreamy

**users who  
haven't  
posted yet a  
single time**

username	id
Aniya_Hackett	5
Kasandra_Homenick	7
Jadyn81	14
Rocio33	21
Maxwell.Halvorson	24
Tierra.Trantow	25

## Project -03

# *Operation Analytics and Investigating Metric Spike*



Operation Analytics is the analysis done for the complete end to end operations of a company. With the help of this, the company then finds the areas on which it must improve upon. You work closely with the ops team, support team, marketing team, etc and help them derive insights out of the data they collect.

This project is about using Operation analytics and Investigating Metric spike to derive certain insights about the product. Different datasets of XYZ company have been provided which can be further used to evaluate and understand the growth and other important factors of the company.

# Findings & Insights

Some of the insights and findings are shown below



Jobs per hour per day

day	time	no_jobs_reviewed
28	22	3
27	22	3
26	140	3
28	140	3
29	140	3
28	140	3
29	140	3
30	50	3
30	50	3
30	45	3
30	45	3
30	45	3
25	45	3

time_spent	day	rolling_thrput
45	25	1.0000
56	26	1.0000
11	27	1.0000
104	27	1.0000
22	28	1.5000
11	28	1.5000
20	29	2.0000
15	30	2.5000
25	30	2.5000

Seven day rolling average of throughput

# Percentage share of each language

english_percent	arabic_percent	persian_percent	italian_percent	hindi_percent
21.4286	14.2857	35.7143	7.1429	14.2857

user_engagement	user_id	day
5	10612	1
5	11037	1
6	11240	1
5	15352	1
4	11364	1
4	15136	1
3	11464	1
4	11497	1
3	15127	1
5	14857	1
5	11750	1
6	11768	1
3	11770	1

User engagement

product_growth	year	device
289	2014	acer aspire desktop
604	2014	acer aspire notebook
197	2014	amazon fire phone
778	2014	asus chromebook
658	2014	dell inspiron desktop
1154	2014	dell inspiron notebook
663	2014	hp pavilion desktop
288	2014	htc one
877	2014	ipad air
482	2014	ipad mini

Amount of users growing over time for a product.



## Project -04

# Hiring process Analytics

### Project description :

Hiring process is the fundamental and the most important function of a company. Here, the MNCs get to know about the major underlying trends about the hiring process. Trends such as- number of rejections, number of interviews, types of jobs, vacancies etc. are important for a company to analyse before hiring freshers or any other individual. Thus, making an opportunity for a Data Analyst job here too!

The detailed report was made based on the data given using excel where statistics knowledge, and different formulas are used to complete the project

### Insights & Findings:

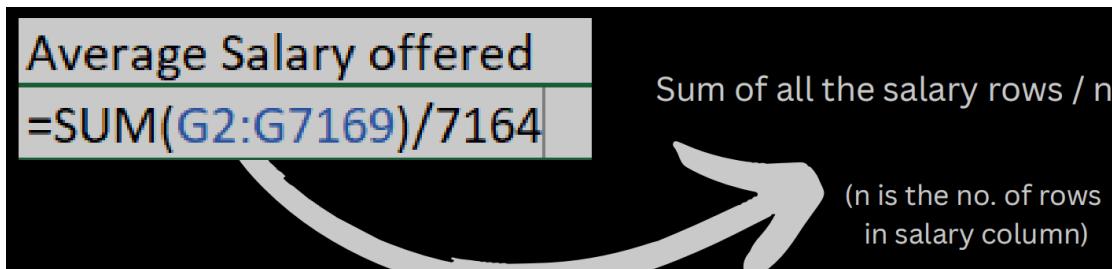
calculated number of people got hired (males & females). we are able to count the number of people got hired is **4695**

	Males & Females who got hired
Hired	=COUNTIF(C2:C7169, P7)



Average salary offered is **49878.3318**

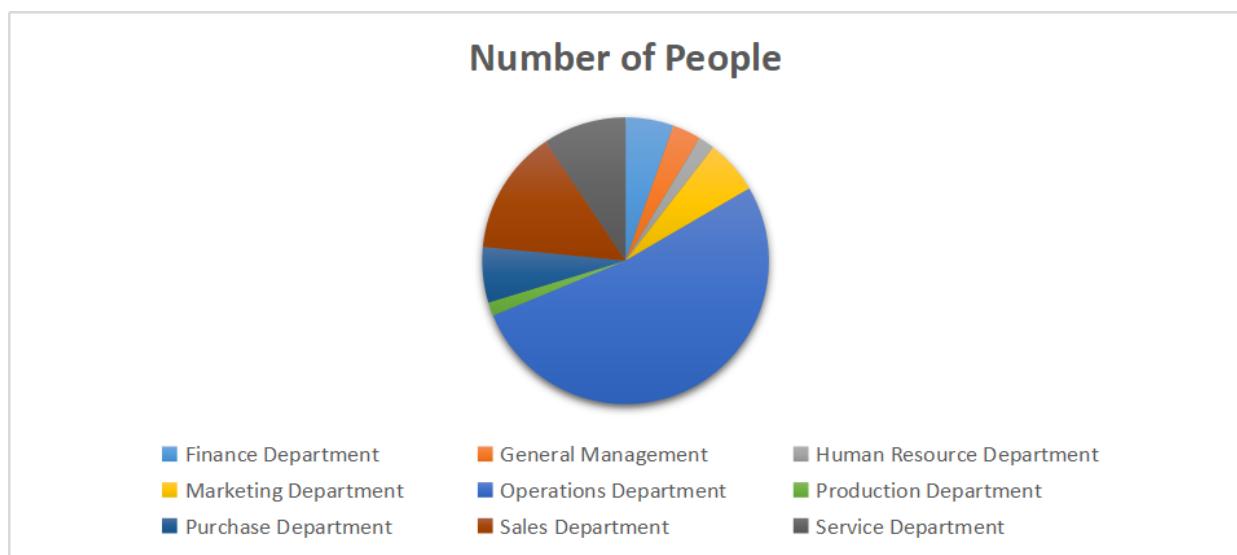
**12**



## CLASS INTERVALS

Upper class limit	99967
Lower class limit	100
Class_interval	99867

Proportion of people working in different department



# Project -05

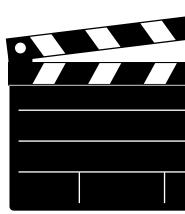
## ***IMDB movie Analysis***

### Project Description:

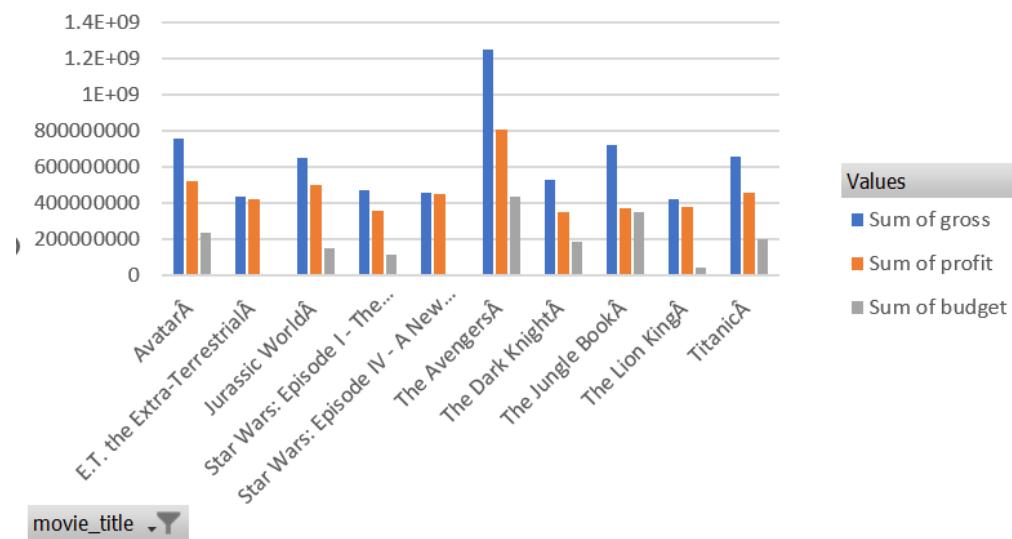
The data containing movies of various languages from IMDB is given. IMDB is the website which has various data of movies, like their cast, crew, year, ratings, etc. Detailed report based upon some of the questions has to be given using root cause analysis, developed by Sakichi Toyoda, founder of Toyota Industries.

So, the Root cause analysis (RCA) is the process of discovering the root causes of problems in order to identify appropriate solutions. RCA assumes that it is much more effective to systematically prevent and solve for underlying issues rather than just treating ad hoc symptoms and putting out fires.

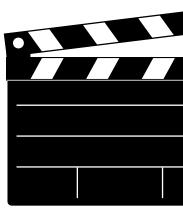
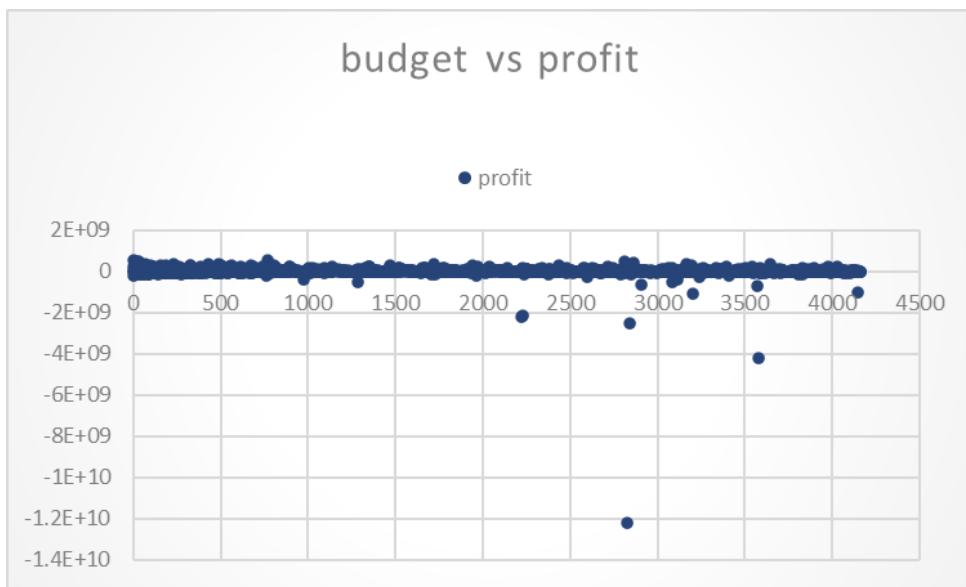
### Insights & Findings:



# Movies with highest profit:

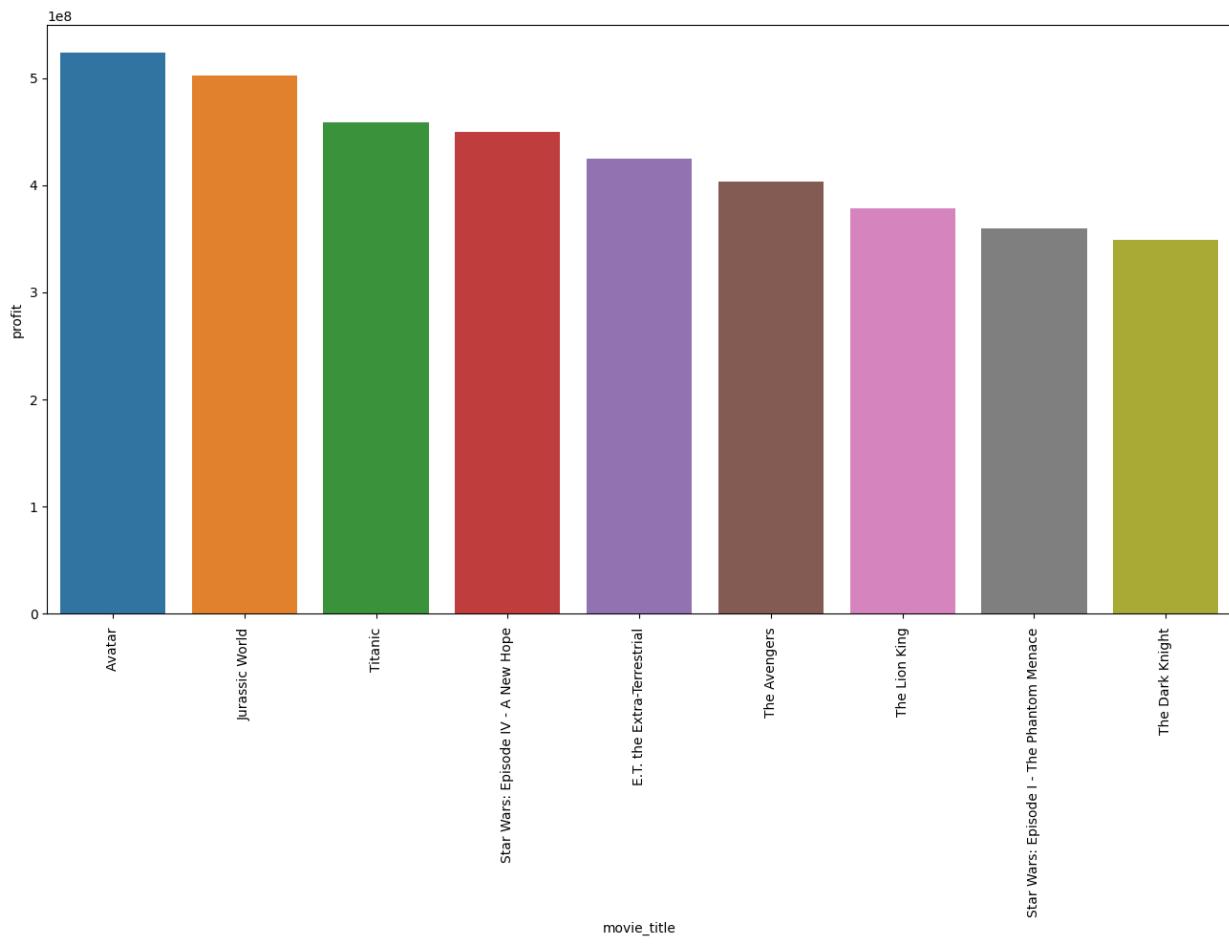


# Budget vs Profit:



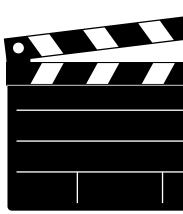
# Highest profitable movies:

15



# Top 250 based on IMDB rating:

Imdb_Top_250
The Shawshank RedemptionÂ
The GodfatherÂ
The Dark KnightÂ
The Godfather: Part IIÂ
The Lord of the Rings: The Return of the KingÂ
Pulp FictionÂ
The Good, the Bad and the UglyÂ
Schindler's ListÂ
InceptionÂ
Fight ClubÂ
Star Wars: Episode V - The Empire Strikes BackÂ
The Lord of the Rings: The Fellowship of the RingÂ
Forrest GumpÂ
Seven SamuraiÂ
City of GodÂ
Star Wars: Episode IV - A New HopeÂ

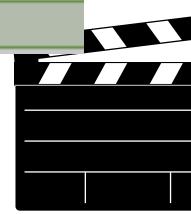


# Ranking of the films based on ratings & number of reviews(>25000)

Rank_of_films
1
2
3
3
5
5
5
5
9
9
9
9

Top Foreign language films other than English

Top_Foreign_Lang_films
Crouching Tiger, Hidden DragonÃ¢
AmÃ©lieÃ¢
Pan's LabyrinthÃ¢
Spirited AwayÃ¢
HeroÃ¢
HeroÃ¢
OldboyÃ¢
The Good, the Bad and the UglyÃ¢
City of GodÃ¢
Seven SamuraiÃ¢
Princess MononokeÃ¢
DownfallÃ¢
AkiraÃ¢
Das BootÃ¢
MetropolisÃ¢
Baahubali: The BeginningÃ¢
The Lives of OthersÃ¢

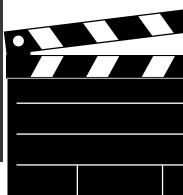


# Top 10 directors based on the average IMDB score

director_name	imdb_score
Tony Kaye	8.600000
Charles Chaplin	8.600000
Majid Majidi	8.500000
Damien Chazelle	8.500000
Alfred Hitchcock	8.500000
Ron Fricke	8.500000
Sergio Leone	8.433333
Christopher Nolan	8.425000
Marius A. Markevicius	8.400000
S.S. Rajamouli	8.400000

## Popular Genres

genres	imdb_score
Crime Drama Fantasy Mystery	8.50
Adventure Animation Drama Family Musical	8.50
Action Adventure Drama Fantasy War	8.40
Adventure Animation Fantasy	8.40
Adventure Drama Thriller War	8.40
Biography Drama History Music	8.30
Documentary War	8.30
Documentary Drama Sport	8.30
Adventure Animation Comedy Drama Family Fantasy	8.30
Adventure Drama War	8.25
Drama Mystery War	8.20
Biography Crime Documentary History	8.20
Drama Fantasy War	8.20
Adventure Drama Thriller Western	8.10
Action Animation Sci-Fi	8.10

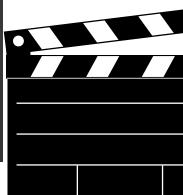


# Top 10 directors based on the average IMDB score

director_name	imdb_score
Tony Kaye	8.600000
Charles Chaplin	8.600000
Majid Majidi	8.500000
Damien Chazelle	8.500000
Alfred Hitchcock	8.500000
Ron Fricke	8.500000
Sergio Leone	8.433333
Christopher Nolan	8.425000
Marius A. Markevicius	8.400000
S.S. Rajamouli	8.400000

## Popular Genres

genres	imdb_score
Crime Drama Fantasy Mystery	8.50
Adventure Animation Drama Family Musical	8.50
Action Adventure Drama Fantasy War	8.40
Adventure Animation Fantasy	8.40
Adventure Drama Thriller War	8.40
Biography Drama History Music	8.30
Documentary War	8.30
Documentary Drama Sport	8.30
Adventure Animation Comedy Drama Family Fantasy	8.30
Adventure Drama War	8.25
Drama Mystery War	8.20
Biography Crime Documentary History	8.20
Drama Fantasy War	8.20
Adventure Drama Thriller Western	8.10
Action Animation Sci-Fi	8.10



## Project -06

### ***Bank Loan case study***

The loan providing companies find it hard to give loans to the people due to their insufficient or non-existent credit history. Because of that, some consumers use it as their advantage by becoming a defaulter.

EDA is performed on the given data records to find relationships, patterns in the data in order to make sure that the loan is availed to the customer who is capable of paying, and also making sure to avoid bias.



## APPROACH

Framed a 4 step analysis process



Understanding data

TRYED TO UNDERSTAND THE DATA GIVEN TO START ANALYZING THE PROBLEM



Performing EDA in excel

Analyzing the data using variate analysis which is a part of EDA, to recognize the patterns and understand relation between features, and also data cleaning.



Python usgae

Also used python for EDA for extensive understanding

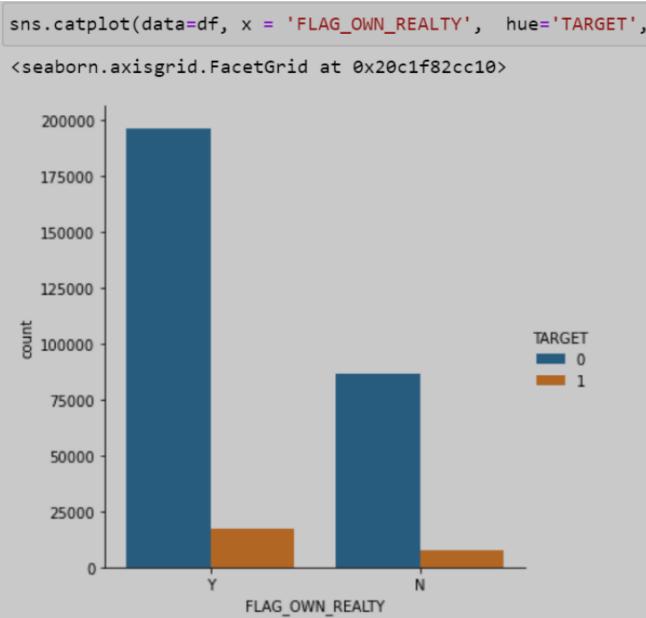


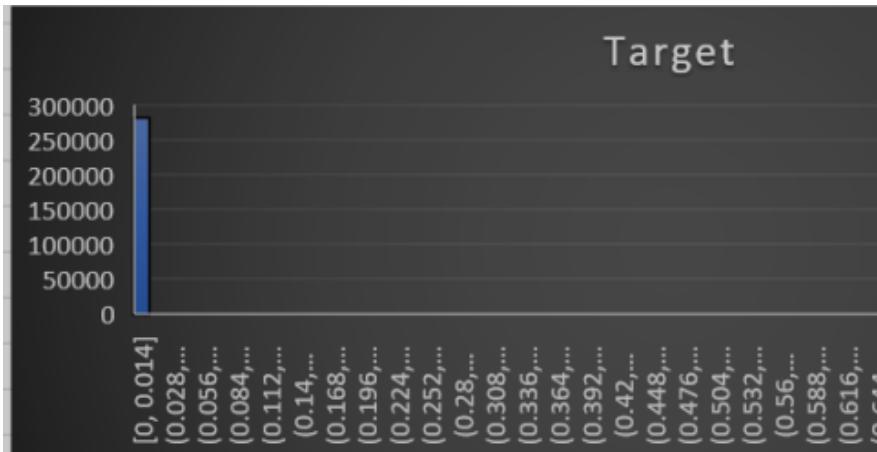
Summarize insights

Summarizing the insights from the analysis made

## Insights & Findings:

Analyzing all the FLAG variables, these features are heavily imbalance and not useful differentiator for the TARGET To reduce the data imbalance, this step is to be performed



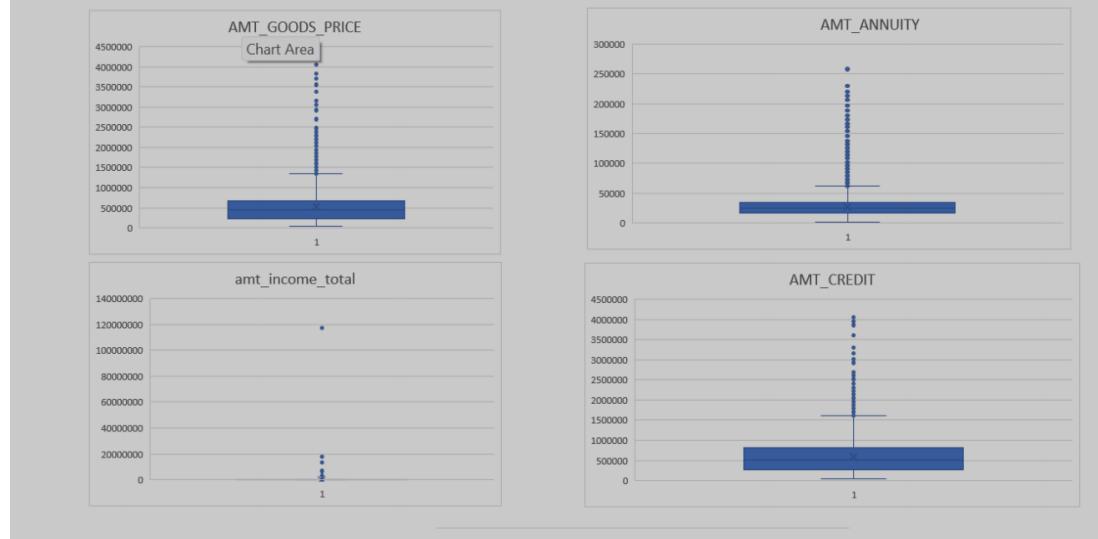


Target variable is having huge data imbalance

clients with payment difficulties are so less in number to say, whereas the other category are very high in number comparatively

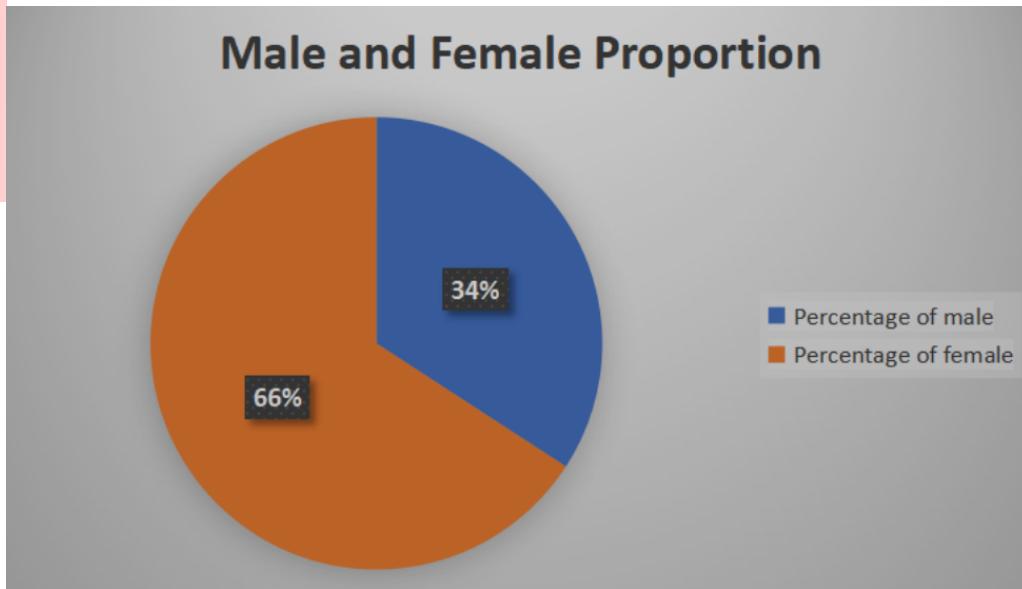
## Outlier analysis

### Outlier Analysis





As the Income is increasing, the credit amount also seems to be increasing



Female applicants are more than the male applicants comparatively





As we can see there's a huge imbalance between the categories of the target feature. The ratio of data imbalance is 8:92



# Project -07

## *Analyzing the Impact of Car Features on Price and Profitability*

### Project Description :

The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars.

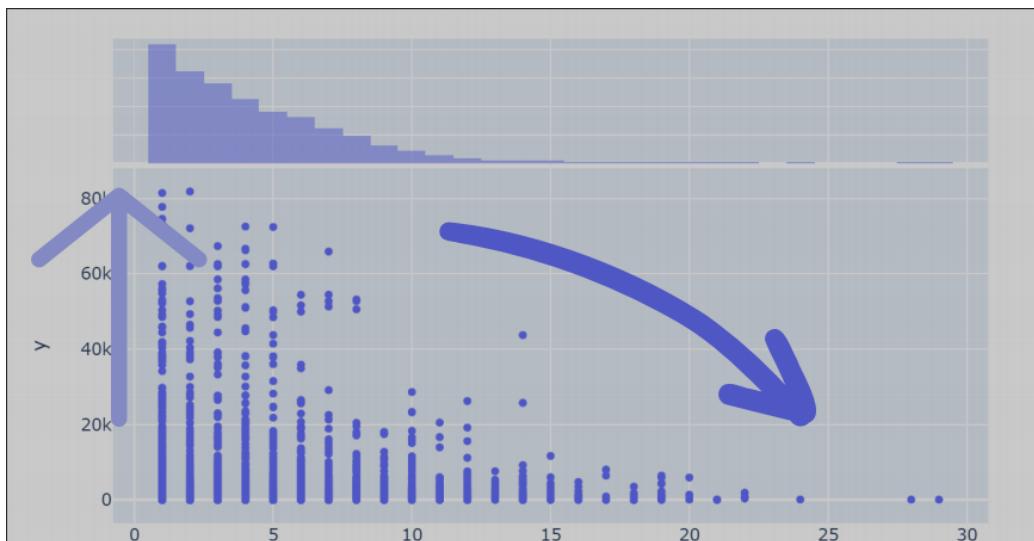
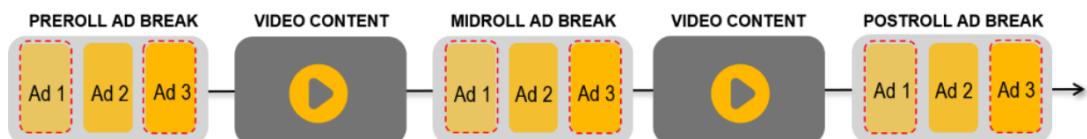
In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.

Using analysis and visualization, I have seen many observations. Like which car manufacturer is leading among the car companies, their strategy to stand in the leading position. Based on the type of advertisement given by each company, its duration, their spend, etc. Many influencing factors were found and were used to solve the problem. The problem was solved in a way such that, the manufacturers can optimize their pricing, spends to increase the business, product development decisions to maximize profitability while meeting consumer demand.

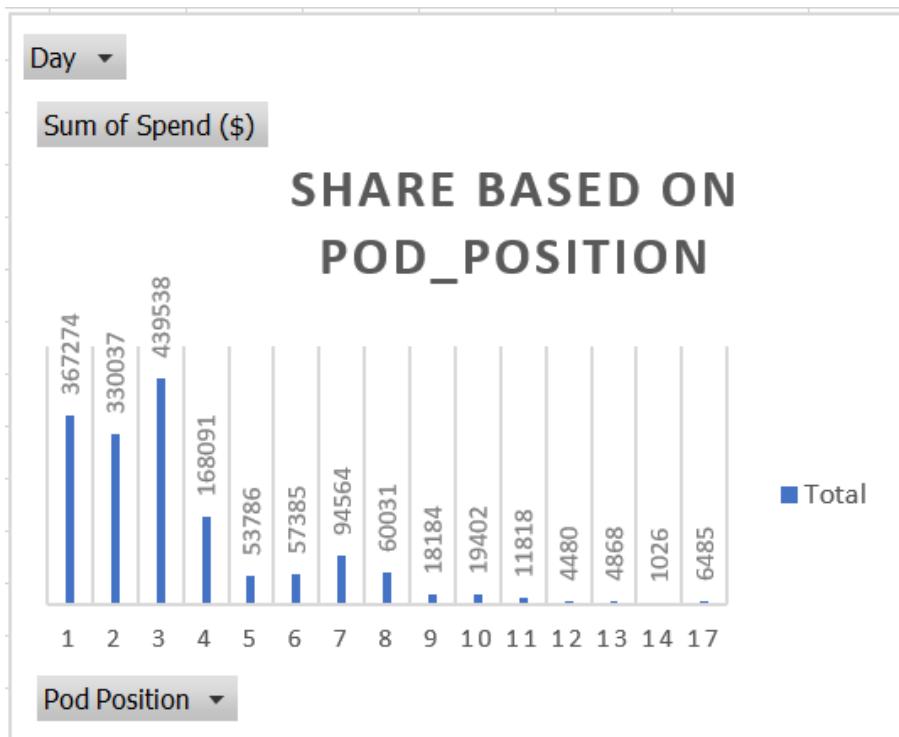


# Insights & Findings

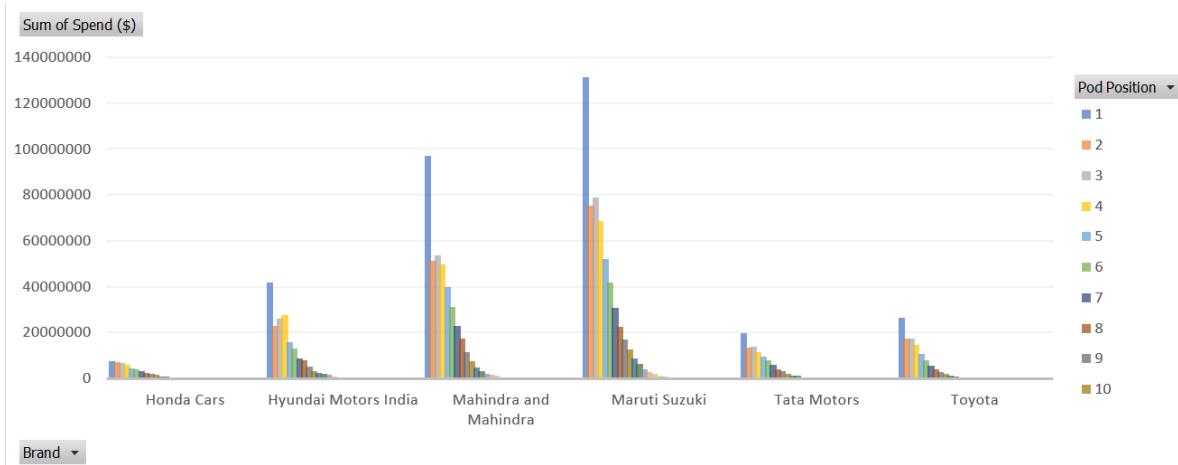
Pod Position is the position of the ad where the multiple ads are grouped together and are played one after another with a single break each time. Say you're watching DareDevil series in Hotstar, while watching ad appears while watching the episode, there will be three ads played back to back after then the episode resumes. You just sat through an ad pod



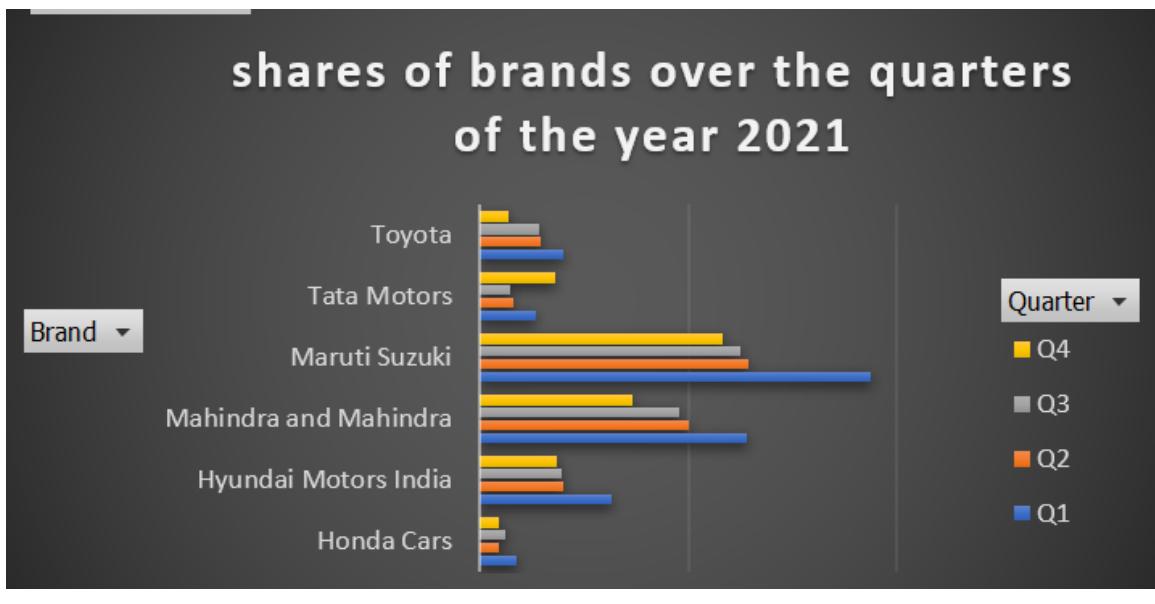
- we can see here the distribution as well as the graph plotted between Podposition(along horizontal axis), and Spend(\$) along the vertical axis. As pod position increases, the spend is increasing initially, but later on that relation became inverse. Like when pod position increasing, the spend is getting decreased.



- The distribution of spend seems to be normally distributed according the pod position.



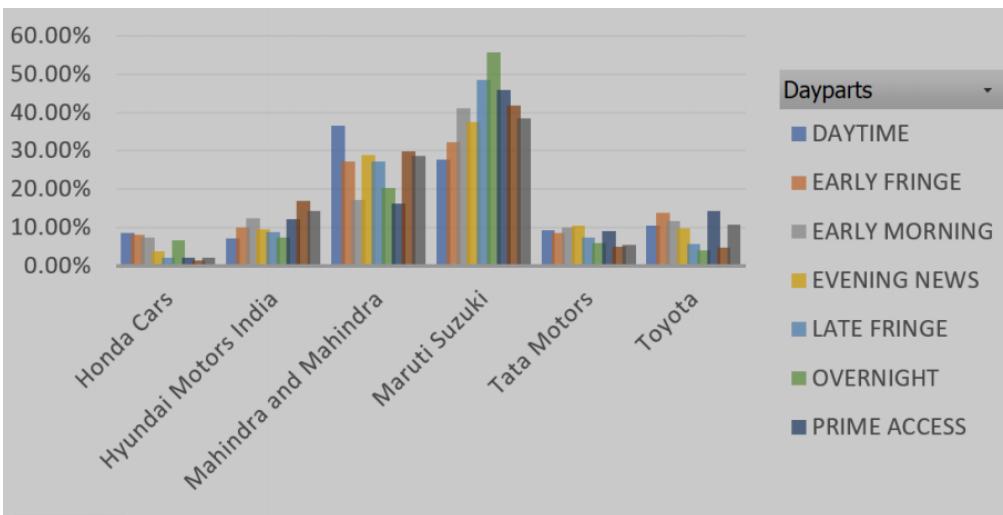
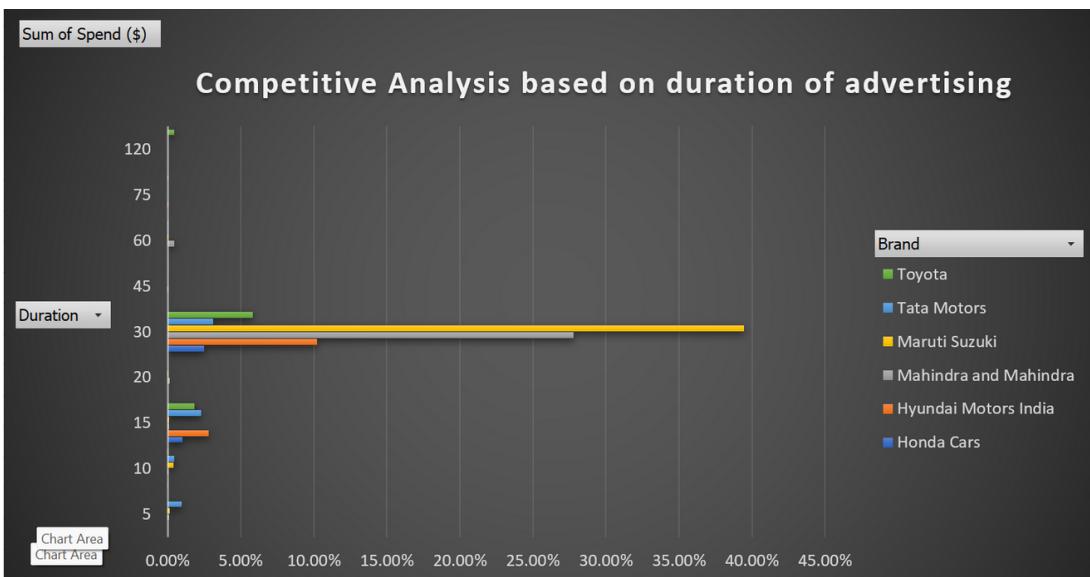
- For quarter wise share from Q1 to Q4, I created a new column named Quarter which has 4 quarters in which each quarter holds 3 months each from the start. (Q1 has Jan, Feb, Mar and so on)



When we go brand wise, and their share spent, observations:

- Spent Share of Honda is too less when compared to other brands
- Tatamotors & Toyota are other 2 brands which have less shares. The share is decreasing when going from Q1 to Q4 in 2021 but except for Tata motors where share is high in the Q4 than in Q1, Q2, Q3).
- The trend is almost similar for all the brands and Share of Suzuki is highest





In the most of the dayparts, Maruti Suzuki is the leading(except for daytime it seems, where Mahindra and Mahindra is leading in the daytime). Maruti spent most at overnight. Mahindra spent most at the daytime and also the Honda cars spent most at the day time whereas Hyundai at prime time



# Project -08

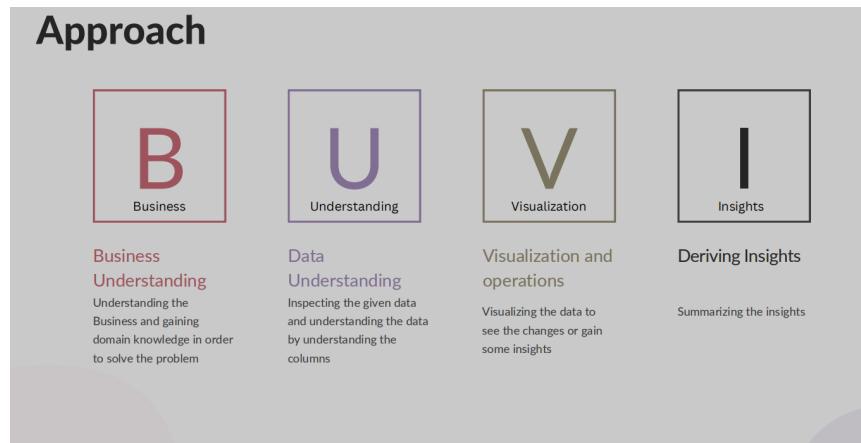
## ***Call Volume Trend Analysis***

### **Project Description :**

A customer experience (CX) team consists of professionals who analyze customer feedback and data, and share insights with the rest of the organization. Typically, these teams fulfil various roles and responsibilities such as: Customer experience programs (CX programs), Digital customer experience, Design and processes, Internal communications, Voice of the customer (VoC), User experiences, Customer experience management, Journey mapping, Nurturing customer interactions, Customer success, Customer support, Handling customer data, Learning about the customer journey.

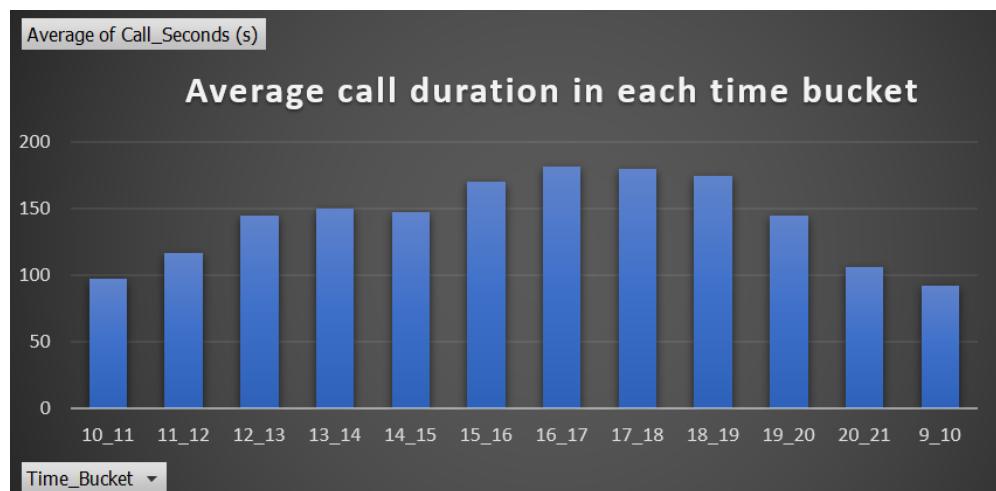
Objective is to analyze the calls data, and provide insights like number of customer calls are being answered at different time slots in a day, and providing solutions to increase this answering rate to customers by agents. Doing these improves the By solving your customers' problems and helping them achieve success using your product or service, you can delight your customers and turn them into a growth engine for your business. And this even helps to make aware of the product to the customer at low cost



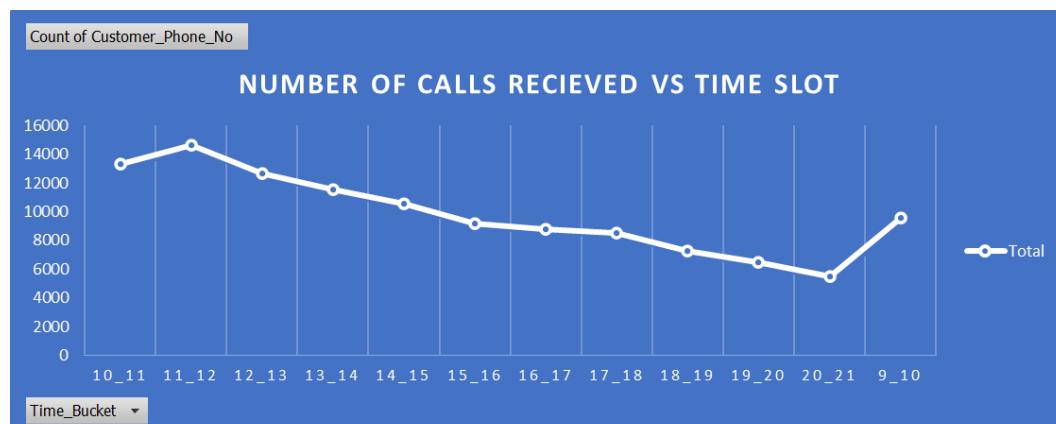


# Insights & Findings:

- The average call duration in each time slot/ time bucket



- The number of calls being received are decreasing over time when day makes a transition from day to night.



- The current abandon rate is 30% and objective is to decrease it to 10%. By having the minimum number of agents in each time bucket we can make it possible.

answering percentage	abandon percentage
69.88168288	30.11831712

- we have On average an agent occupied for 60% of his total actual working Hours i.e. for 6 hours out of actual working hours (9 hours)
- we can get the required number of agents for 90% answer rate if we can calculate the average calls received on a single day and total time spent.
- average calls on first day = (total number of agent on that day) \* (total time spent)
- where total agents required = total number of agents on that day
- total agents required = average calls on first day / total time spent
- As per the assumption, the time spent by agent will be 60% of 7.5 hours, i.e. 5 hours

For 60% occupancy agents required	37.59244444
For 90% occupancy agents required	?

56.38866667

- Hence the number of agents required are 56

Let's say customers also call this ABC insurance company in night but didn't get answer as there are no agents to answer, this creates a bad customer experience for this Insurance company. Suppose every 100 calls that customer made during 9 Am to 9 Pm, customer also made 30 calls in night between interval [9 Pm to 9 Am] and distribution of those 30 calls are as follows:



Distribution of 30 calls coming in night for every 100 calls coming in between 9am - 9pm (i.e. 12 hrs slot)												
9pm- 10pm	10pm - 11pm	11pm- 12am	12am- 1am	1am - 2am	2am - 3am	3am - 4am	4am - 5am	5am - 6am	6am - 7am	7am - 8am	8am - 9am	
3	3	2	2	1	1	1	1	3	4	4	5	

- By calculating the Average number of calls answered (per agent), we can get the additional agents required for the night time buckets i.e.
- **additional agents = percent of incoming calls/Average number of calls answered (per agent)**

type of count	count
answered count	82452
count of abandoned	34403
transfer count	1133
total count	117988

Total agents	65
Average number of calls ans	1268.492308
	1268

Therefore,

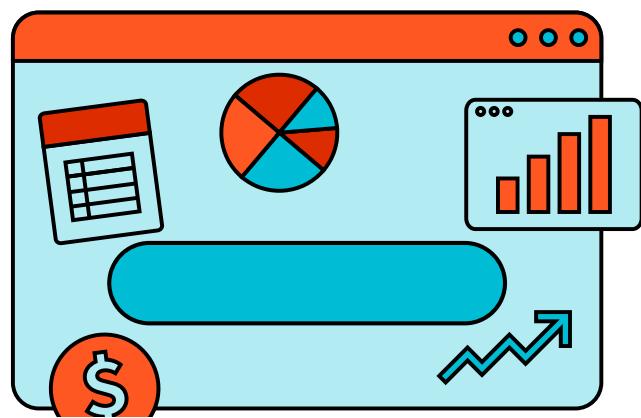
- Agents required for 90% answering rate are agents required = Incoming Calls/ Average number of calls answered (per agent)

Time Bucket	Incoming Calls	New Agents Required
9_10	9588	8
10_11	13313	11
11_12	14626	12
12_13	12652	10
13_14	11561	10
14_15	10561	9
15_16	9159	8
16_17	8788	7
17_18	8534	7
18_19	7238	6
19_20	6463	6
20_21	5505	5



# Conclusion

By doing the above case studies and projects, I got to realize how it will be when dealing with data and analyzing information. I got some hands on the real-time data and solve the problems using different methods, statistical analysis, analytical thinking. And also got some domain knowledge which is required to deal with business problems. I understood like controlling waves of data, making sense of the data that seems useless, drawing insights & conclusions from the data. Also got to understand the key aspects of analytical thinking like visualization, making strategy, decision-making, problem orientation. These projects involved identifying, defining a problem and then solving it by using data in a organized, step by step manner.



# Appendix

## Project-01

[https://drive.google.com/file/d/1Ylhd5eP9PRLZkaPZFtcHhF6RJktu4QeG/view?usp=share\\_link](https://drive.google.com/file/d/1Ylhd5eP9PRLZkaPZFtcHhF6RJktu4QeG/view?usp=share_link)

## Project-02

[https://drive.google.com/drive/folders/1ro7V\\_F9JJUFe7MSniALrwholvgo0AVvz?usp=share\\_link](https://drive.google.com/drive/folders/1ro7V_F9JJUFe7MSniALrwholvgo0AVvz?usp=share_link)

## Project-03

[https://drive.google.com/drive/folders/1OvW8guIV9kMZb5dlvm5HDUwJxBJOZQR?usp=share\\_link](https://drive.google.com/drive/folders/1OvW8guIV9kMZb5dlvm5HDUwJxBJOZQR?usp=share_link)

## Project-04

[https://drive.google.com/drive/folders/1wvlZZxwXWnS\\_Y9J2sIH5fjBx1KbJfvEw?usp=share\\_link](https://drive.google.com/drive/folders/1wvlZZxwXWnS_Y9J2sIH5fjBx1KbJfvEw?usp=share_link)

## Project-05

[https://drive.google.com/drive/folders/1wvlZZxwXWnS\\_Y9J2sIH5fjBx1KbJfvEw?usp=share\\_link](https://drive.google.com/drive/folders/1wvlZZxwXWnS_Y9J2sIH5fjBx1KbJfvEw?usp=share_link)

## Project-06

[https://drive.google.com/drive/folders/1wvlZZxwXWnS\\_Y9J2sIH5fjBx1KbJfvEw?usp=share\\_link](https://drive.google.com/drive/folders/1wvlZZxwXWnS_Y9J2sIH5fjBx1KbJfvEw?usp=share_link)

## Project-07

<https://drive.google.com/drive/folders/1jiJRMFvqsS3mrm0q2BbnXAsS9YmjE2kz?usp=sharing>

## Project-08

[https://drive.google.com/drive/folders/1VG1e8paVD8W7YJNjL\\_6TcPlaygmcYla?usp=share\\_link](https://drive.google.com/drive/folders/1VG1e8paVD8W7YJNjL_6TcPlaygmcYla?usp=share_link)