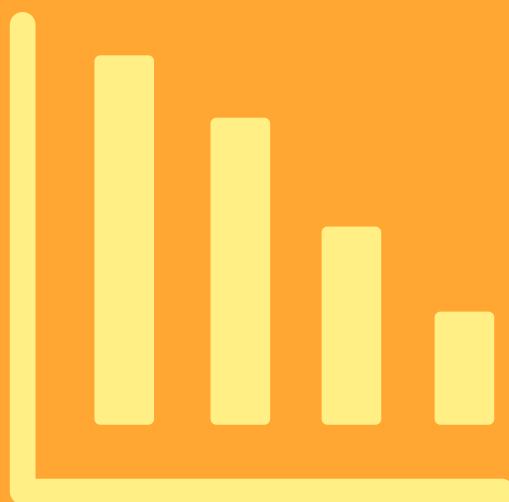Data Analyst

PORT FOLIO

Sumit Chaure

PORTFOLIO

2023

SUMIT CHAURE

Data Analyst

SUMIT CHAURE

Sumit Chaure, Data Analayst

EDUCATION

I have completed my *Bachelors in Engineering* from *Mumbai University* in 2018 with a 7,2 CGPA.

I have recently completed my Internship based certification course from **Trainity** in *Data analytics* where i learnt about the various tools, techniques used to manipulate and handle large set of data & find meaningful insights from them.

I have got hands-on working experience using MS-Excel, SQL, Python, BI Tools, Visualization Library, Story telling & report making.

Currently I am learning ML to get to know more about data domain and how to use data to predict insights.

WORK EXPEREINCE

I have 2.5 years of experience working as a software engineer.

I have worked on projects involving backend & database handling using Python, Go-lang, Mongodb, SQL and various cloud DevOps tools like Docker, Kubernetes, Git version control & services like AWS,GCP etc.

I have been constantly working on side projects to learn more about data domain alongside my work which involved python frameworks & library , this has lead to my decision to switch my career into the data domain for which i am currently exploring roles in data analytics or data science jobs.

Table of Contents

PROJECT INDEX

Project - 1

Data Analytics Process: Real World Application

trainity

Shopping & Use of 6 Step Data Analytics Process

DATA ANALYTICS PROCESS

Data Analytics Basics

Project 1 – **Data Analytics Process**
Application in Real Life Scenario Case Study

Objective : Give the example(s) of such a real-life situation where we use Data Analytics and link it with the data analytics process.

Example - *Purchasing a Laptop*
I will be taking the example of a laptop purchase to explain the steps involved in data analytics as a real world example of the same.

Steps Involved in Data Analytics Process -

1. *Plan* - Relevant questions to understand the requirement of purchase
- What is the use case of the laptop?
- What is the budget?
- Are you looking for a specific brand?
- How many years are looking to use the machine?
- Do you plan to sell the laptop in a couple of years (to see resale value)?
- What are the specifications necessary for your Usecase?
- Checking for warranty and after sales services.
- Any specific hardware requirements needed to consider?

2. *Prepare* - Collecting the necessary data to answer our previous question
- Checking online retailers for checking the latest models.
- Summarizing the specifications satisfying our needs(use case) and formulating best options for purchase.
- Planning the purchase by keeping in mind future demand and opting to buy a higher version if possible.
- Checking review sites for comparing between different models and brands not just by budget but comparing on hardware specifications and service quality for future needs.
- Look for multiple options in our price range and compare them on different sites and offline stores if possible (by visiting or calling them).
- Watching for discounts and offers if available to add products in our budget.

3. *Process*
- Making an excel sheet with relevant columns for details of the products gathered in the previous step.
- Sorting things in a budget category or depending on hardware specs.
- We can ask for reviews about brands from our friends to get more insights on the durability and actual usage or if someone has already bought the model we are looking at in our purchase.
- If possible, visiting a shop and using the model yourself and getting more details from a seller about the warranty and upgrades or other options would be a better choice to make our decision.

4. *Analyze*
- We can use excel to sort things and make a better visual view of the collected data.
- We can remove the options which we don't think are relevant or some good alternatives are present from the list.
- We can take help from certain online websites at these points to help ourselves to narrow down on the option by comparing the finalized list of options on full specifications and quality rating given by users and tech experts.
- Excel or a simple notepad can help us to narrow down our options after doing the necessary comparison either by using websites or from our own findings on excel sheets or from reviews from our friends.

5. *Share*
- After all the data gathering and comparison of offers, specifications and brand value and its after services we can share the details with the one buying the product or with the dealer to book the required product.
- If we are planning to buy from online stores we can search on multiple sites for the same model and look for offers and discounts and finalize from which to buy and see if we can get additional warranty with the purchase.

6. *Act*
- The final step is to purchase the selected model after considering all the best options available and after looking for festive offers or bank discounts and taking all the necessary additional things if needed with the final purchase.

1. **Plan** - To buy a laptop for studying - coding and data analytics so a mid-level would be a good option under the rs.50k price bracket.
2. **Prepare** - I will check for various models on online websites and compare them and look for the finance option for purchase.
3. **Process** - After collecting data and comparing them from various sites and offline store visits I will try to sort the data to make it more reliable and readable.
4. **Analyze** - The processed data can now be useful to make an insight for my purchase and available options.
5. **Share** - After my thorough research and comparison i will narrow down my choices to best laptop in my budget and which will match my future use and then share the options with my parents who will be helping me finance the purchase for same along with a lookout for nearby festive or online sales to get maximum return out of my purchase.
6. **Act** - After final comparison and decision i will go either to buy the products from online store with bank offers or from nearby offline dealers for festive offers and future service needs and extra goodies benefits like additional warranty or ram upgrades etc.

KEY FINDINGS & REPORT LINKS
- *The case study helped me to gain the knowledge about the 6 processed in the analysis process.*
- *The given task helped me to explore the core basics of data analysis - steps involved in data collection, data manipulation, data cleaning, visualization, insights from data and decision making approaches.*
- *The simple assessment helps us to understand how a data analyst needs to make decision while he works with data and presents the findings to stakeholders.*

📌 <u>Full Report</u> 🔗

Project - 2
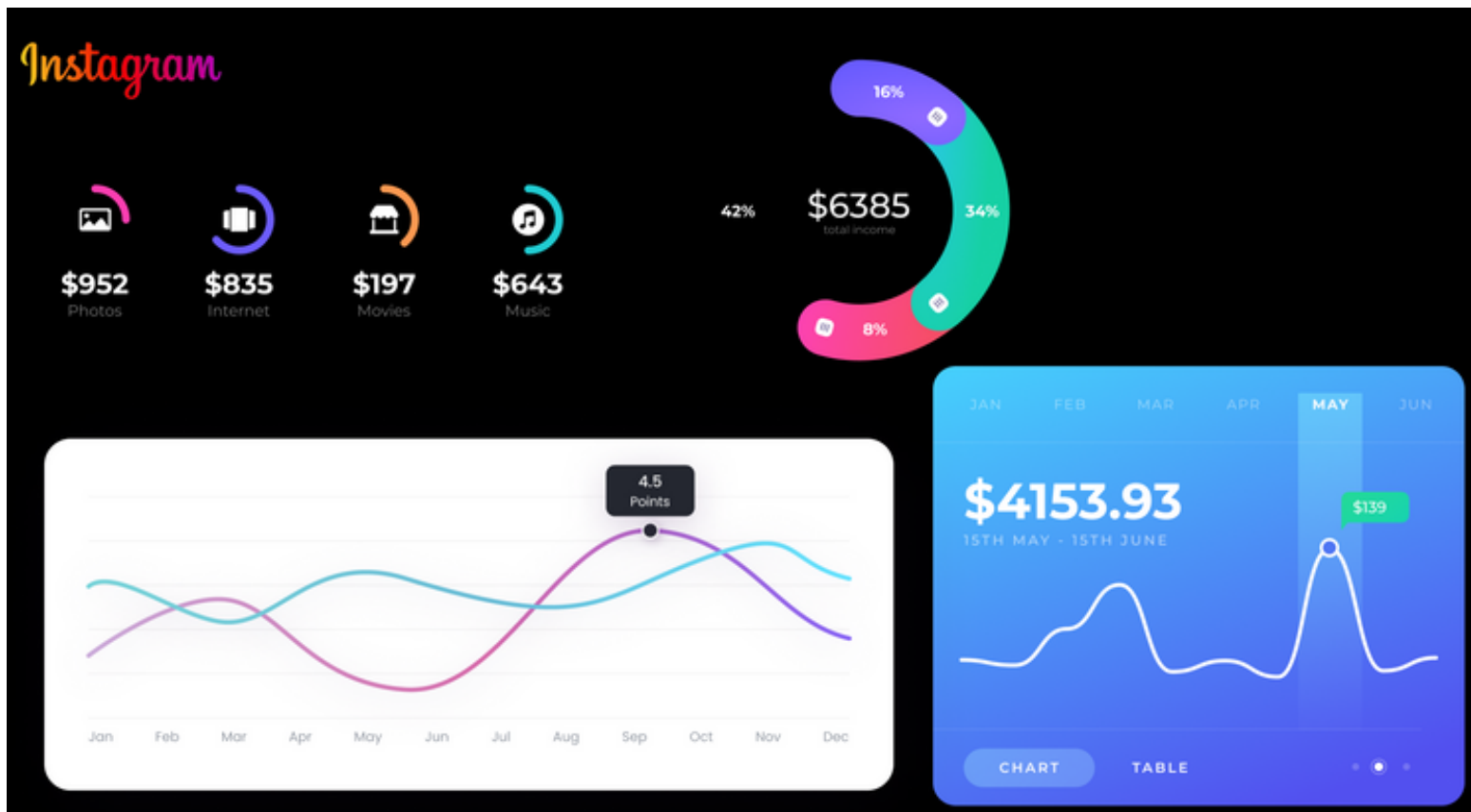
INSTAGRAM USER ANALYTICS

SQL Fundamentals

Project 2 – INSTAGRAM USER ANALYTICS

<u>Objective</u> : Imagine you're a data analyst working with the product team at Instagram. Your role involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

Task 1 : Loyal User Reward

<u>Question</u> : Identify the five oldest users on Instagram from the provided DB.

INSIGHTS :

- The oldest member on the site are shown using the given select query ordered By the account creation date, using the data the management can reward the loyal users for their love & support.
- The Query helps to understand the depth of accuracy that SQL can help to traverse across a large set of data just by setting few filters.

Task 2 : Inactive User Engagement

<u>Question</u> : Identify users who have never posted a single photo on Instagram.

INSIGHTS :

- The query helps us to find the user **(26 Users)** who has not posted a single post on the site indicating they are inactive.
- For a service provider it is important to know about the user activity to manage & scale the service to handle the traffic so such insights are impactful.

Task 3 : Contest Winner Declaration

Question : Determine the winner of the contest and provide their details to team.

INSIGHTS :

- The Query used here helps us to get knowledge of selecting data from multiple tables to get desired results.
- The User name "*Zack_Kemmer93*" has the most liked post in our finding.

Task 4 : Hashtag Research

Question : Identify & suggest the top 5 most commonly used hashtags on platform.

INSIGHTS :

1. The query helps us to get the most used hashtags *(smile, beach, party, fun, concert)* that people used to post on the site, this is a common thing today to have more reach using them so helps the analytics team to know whats trending on the platform.

Task 5 : Ad Campaign Launch

Question : Determine the day of the week when most users register on Instagram.

INSIGHTS :

- We found from our analysis that most new users registered on "Thursdays & Sunday" and on 2nd number is Friday when the Campaign teams can launch ad to grab more traffic and clicks.
- Knowing the most active days helps business to target customers for revenue generation.

Task 6 – User Engagement

Question : Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total no. of users.

```
75  •      SELECT
76             (SELECT
77                      COUNT(*) AS total_users    -- Total users
78             FROM
79                  users) AS Total_Users,
80             (SELECT
81                      COUNT(*) AS total_photos   -- Total posts
82             FROM
83                  photos) AS Total_Photos,
84             (SELECT
85                      COUNT(*)
86             FROM
87                  photos) / (SELECT
88                      COUNT(*)
89             FROM
90                  users) AS avg_posts_per_user;
91
```

Total_Users	Total_Photos	avg_posts_per_user
100	257	2.5700

	A	B	C
1		User Engagement Check	
2	Total_User	Total_Photo	avg_posts_per_use
3	100	257	2.57

KEY FINDINGS

1. The multi-select query helps us get the idea about the user activity & engagement ratio which is important factor for business. *(100 users , 257b posts , ratio = 2.57)*
2. Multi - Select query is useful to get data by sorting things in loop and then querying on it.

Task 7 – Bots & Fake Accounts

Question : Identify users (potential bots) who have liked every single photo on the site, as this is not typically possible for a normal user.

```
95      -- Total post/photos on platform
96  •  SELECT
97          COUNT(*) AS total_photos     -- total post = 257
98     FROM
99          PHOTOS;
100    -- Bot/Suspicious activity - liking every single photo
101 •  SELECT
102         id AS user_id,username, COUNT(*) AS Number_of_likes
103    FROM
104         users
105             INNER JOIN
106         likes ON likes.user_id = users.id
107    GROUP BY likes.user_id
108    HAVING Number_of_likes = (SELECT
109             COUNT(*)
110        FROM
111             photos);
```

user_id	username	Number_of_likes
5	Aniya_Hackett	257

	B	C	D
	Bot Activity Check - Using Post Likes		
	user_id	username	Number_of_likes
	5	Aniya_Hackett	257
	14	Jaclyn81	257
	21	Rocio33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	Mckenna17	257
	54	Duane60	257
	57	Julien_Schmidt	257
	66	Mike.Auer39	257
	71	Nia_Haag	257
	75	Leslie67	257
	76	Janelle.Nikolaus81	257
	91	Bethany20	257
	Suspicious Accounts Count	13	
	Total_Post	In Our check if the user who likes all the posts on platform can be considered as a bot	
	257		

KEY FINDINGS

From the given findings it was evident that around *13 suscipious accounts* have been identified as as they have liked every single post ie. **257 likes** which can be flagged as bot activities as in real life no one follows every single person or page on social media and then like there feeds.

PROJECT 2 - INSTAGRAM USER ANALYTICS

Project Summary: Imagine you're a data analyst working with the product team at Instagram. Your role involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.
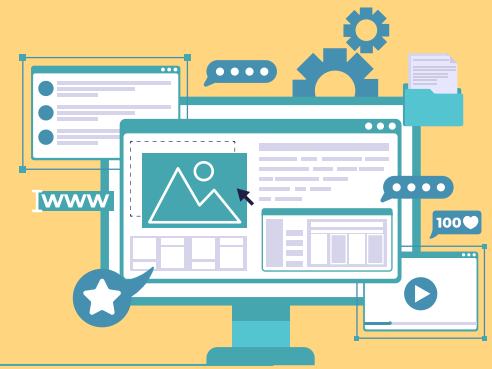
1 PROBLEM/PROJECT DESCRIPTION

- As a data analyst analyze the given dataset to get relevant insights for the questions asked by the stakeholders and management team to help the service develop & generate more revenues from ad leads & products for users.
- Showcase SQL knowledge for analysis.

2 APPROACH (RESEARCH)

1. Created database & tables from the provided data & performed the querying.
2. For Visualization purpose saved the results as csv & used MS excel for tables.
3. From the results crafted the reports with appropriate insights.

3 TECH-STACK USED

- *MySQL* - Querying & data manipulation
- *Excel* - Results generation (tables).
- *Word* - Report building for insights.
- *Google Drive* - Sharing the files used.

4 INSIGHTS

- The given dataset comprises of the users data on the social media site which tracks user activities.
- The data insights helps the developer team to make the site efficient & more engaging.
- For the marketing team it helps to device strategies to generate revenue from users via ad and services.
- For the stakeholders the investment plans can be formulated after looking at the progress reports.

5 RESULTS

1. *The project helped me to gain knowledge about SQL basics like table creation, data imports, querying, joins , etc. which helps the data analyst to track & monitor the live dataset as well as use the data for analysis purpose.*
2. The tasks assigned covered various aspects of basic SQL querying and data handling.

📌 <u>Full Report</u> 🔗
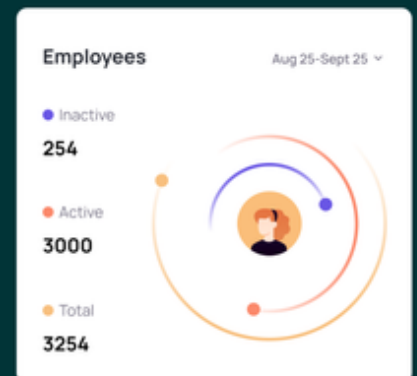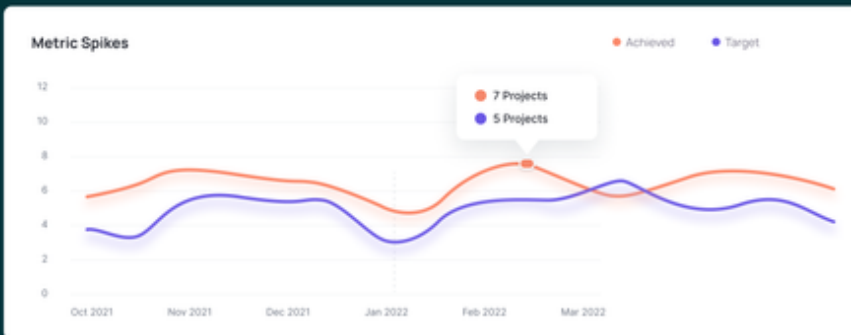
Project - 3

OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

Advanced SQL & Operational Analytics

Project 3 – **Operation Analytics and Investigating Metric Spike**

Objective : Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company. As a Data Analyst, you'll work closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect.

Task 1 : Jobs Reviewed Over Time

Question : Calculate the number of jobs reviewed / hr for each day in November 2020.

Task 2 : Throughput Analysis

Question : Write an SQL query to calculate the 7-day rolling average of throughput.

Img - Weekly Throughput shows a smooth distribution of events done for the week and distribution is more readable. Weekly metric helps us to understand which events holds account for the most work/reviews.
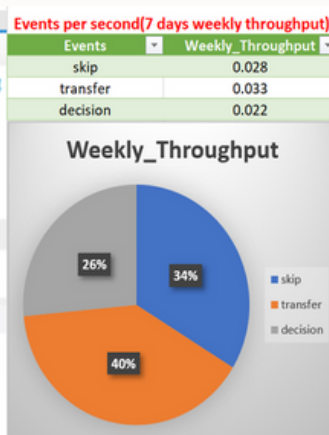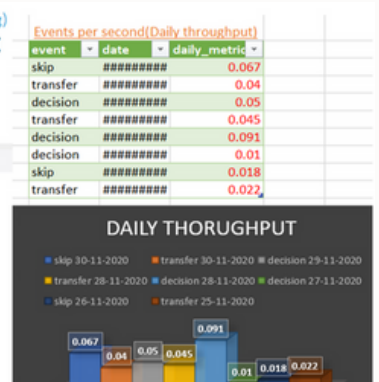
Img - The daily throughput based on the events has distributed the work over daily basis which is not that readable for the insights if we look at the event wise distributions too.

Task 3 : Language Share Analysis

Question : Write an SQL query to calculate the % share of each language over the last 30 days
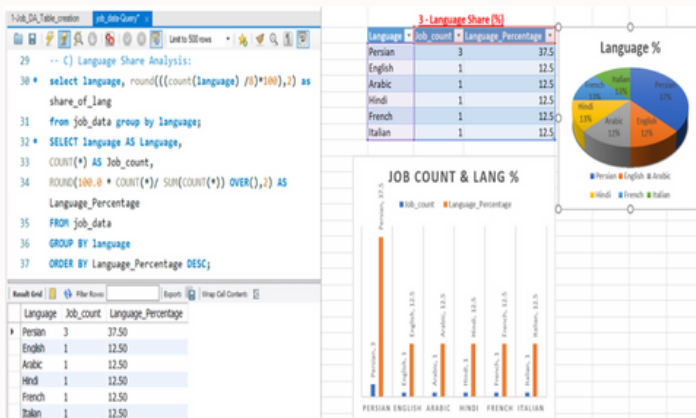
INSIGHTS :

- The persian language has the highest share over the last 30 days.
- The SQL query here uses aggregate function and grouping clause to get the desired results.

Task 4 : Duplicate Rows Detection

Question : Write an SQL query to display duplicate rows from the job_data table.
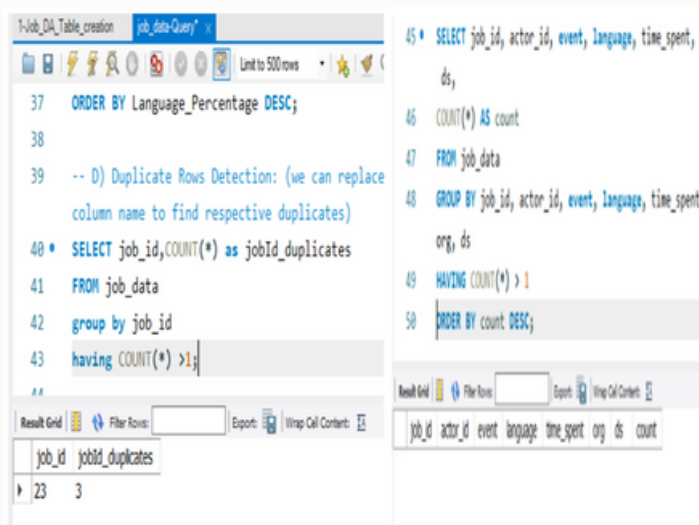
Img - 1st image considers job_id column while 2nd image is considering every column

INSIGHTS :

1. If we want to find the duplicates we can do the count function over the columns and see if duplicates is present in them.
2. If we want to find complete row duplication we need to group it over all the columns and check.
3. Duplicate detection in database helps the admins to check if someone has multiple entries, orders or accounts with the same id which will create error in future.
4. Treating such duplicates is important as data can be duplicate but the whole record should be unique & not complete same.

Task 5 : Weekly User Engagement

Question : Write an SQL query to calculate the weekly user engagement.
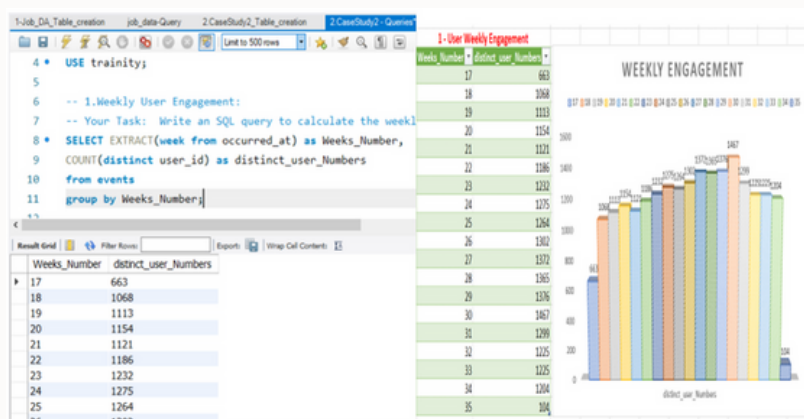
INSIGHTS :

- From the above query we get the user engagement that is unique user visit according to the week numbers.
- This demograph helps us to gather information about the events density and user uniqueness at the events.

Task 6 : User Growth Analysis

Question : Write an SQL query to calculate the user growth for the product.

```sql
15    -- 2.User Growth Analysis:
16    -- Your Task:  Write an SQL query to calculate the user growth for the p
17 •  select week_num, year_num,
18    sum(active_users) over (order by week_num, year_num
19    rows between unbounded preceding and current row) as cumulative_sum
20    from (
21    select extract(week from activated_at) as week_num,
22    extract(year from activated_at) as year_num,
23    count(distinct user_id) as active_users from users
24    where state= "active"
25    group by year_num, week_num
26    order by year_num, week_num) as alias;
27
28
```

INSIGHTS :

- From the query we get to find out about the number of users on the platform and the growing numbers according to week and year wise when we see at the active state of users column.
- *This data would help us determine strategy for future growth planning and ad or other revenue generation activities.*

Task 7 : Weekly Retention Analysis

Question : Write an SQL query to calculate the weekly retention of users based on their sign-up

INSIGHTS :

1. From the query we have found the number of users who have signed up into our cohort program by completing the signup flow and complete the signup process.
2. *The company can make plans to bring back the retained users on the platform to grow the business.*

Task 8 : Weekly Engagement Per Device

Question : Write an SQL query to calculate the weekly engagement per device

```sql
37    -- 4.Weekly Engagement Per Device:
38    -- Your Task:  Write an SQL query to calculate the
39 •  select * from events;
40 •  select device AS Device,
41    extract(week from occurred_at) as Weeks,
42    count(distinct user_id) as No_Of_Users
43    from events
44    where event_type="engagement"
45    group by Device, Weeks order by Weeks;
```
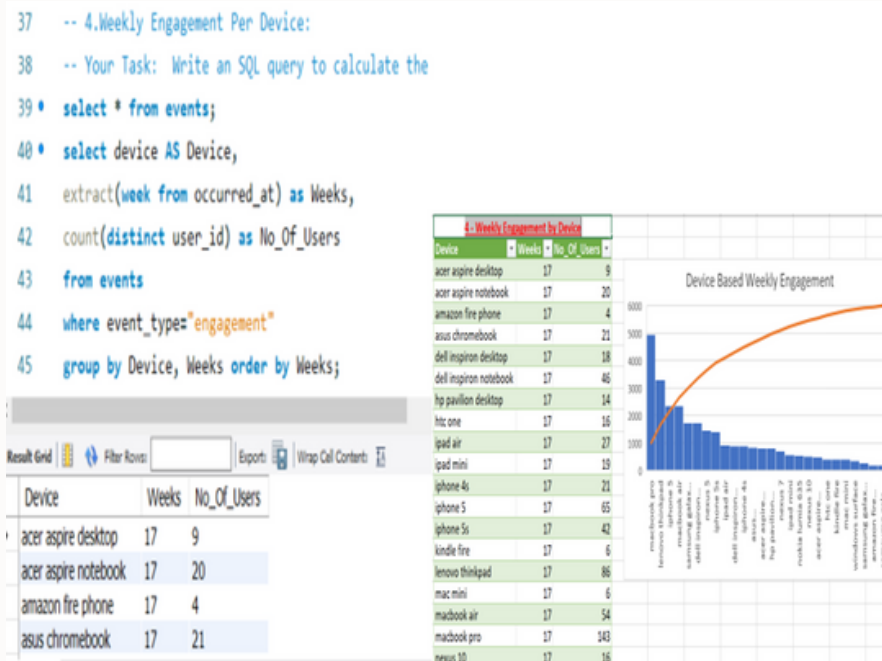
INSIGHTS :

- The insight shows us the weekly demographic of the various devices used to access the events page and the number of user using the device in a certain weeks period.
- These data helps us to optimize our site/page/app to suite to majority of the devices/OS and thereby increasing our reach.
- The data shows that majority of the users prefer to use *macbook pro.*

PROJECT 3 - OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

Project Summary: As a Data Analyst, you'll work closely with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect. Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations. This analysis helps identify areas for improvement within the company.
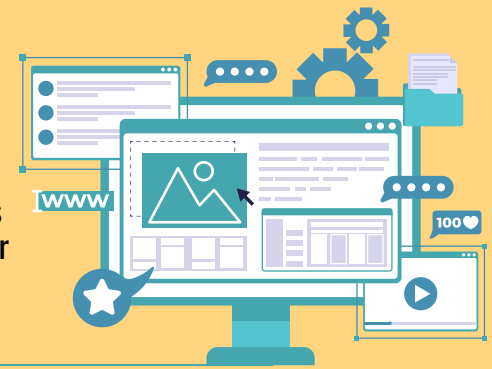
1. PROBLEM/PROJECT DESCRIPTION

- Your goal is to use your advanced SQL skills to analyze the data and provide valuable insights that can help improve the company's operations and understand sudden changes in key metrics.
- There are two Case Study in these assignment :
  > Job Data Analysis
  > Investigating Metric Spike

2. APPROACH (RESEARCH)

1. Created database & tables from the provided data & performed the querying.
2. For Visualization purpose saved the results as csv & used MS excel for tables.
3. From the results crafted the reports with appropriate insights.

3. TECH-STACK USED

- *MySQL* - Querying & data manipulation
- *Workbench* - To import & change the data to suit our needs.
- *Excel* - Results generation (tables).
- *Word* - Report building for insights.
- *Google Drive* - Sharing the files used.

4. INSIGHTS

- The given dataset comprises of the job data & user data to track the job application & users on our platform.
- The data insights helps the developer team to make the site efficient & more engaging and for the HR to keep a tab on the applicants data.
- The user data can be handled by team to generate revenue while the job data is for recruitment.

5. RESULTS

1. *The project helped me to gain knowledge about advanced SQL concepts like multi-select statements, date transformation, casting, joins, CTE & Window Functions etc. which helps the data analyst to track & monitor the live dataset as well as use the data for analysis purpose and dig deeper into the fundamentals of operational analysis.*
2. The tasks assigned covered various aspects of advanced SQL & operation metrics.

📌 Full Report 🔗
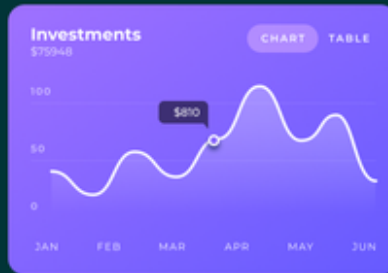
Project - 4

HIRING PROCESS ANALYTICS

Statistics

Project 4 – HIRING PROCESS ANALYTICS

Objective : Imagine you're a data analyst at a multinational company like Google. Your task is to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.

Task 1 : Hiring Analysis

Question : Determine the gender distribution of hires.
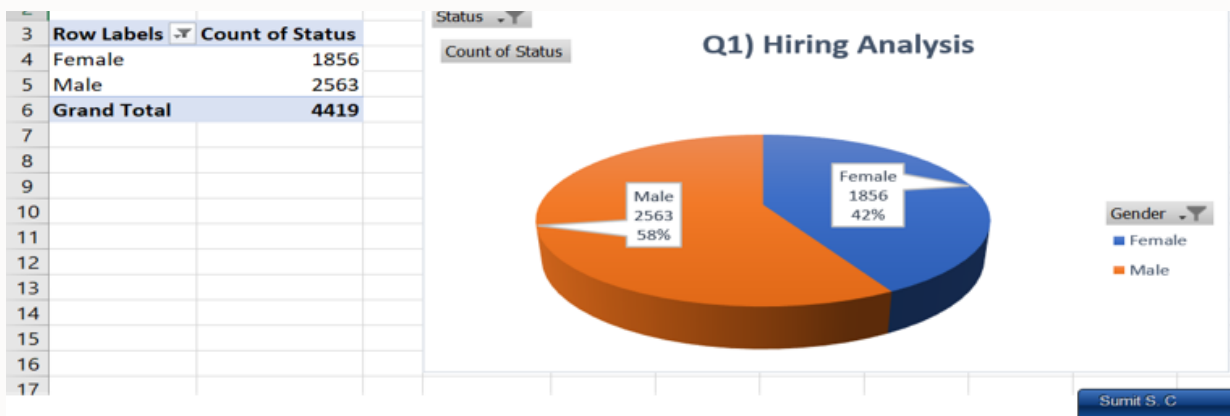
Figure 1) Q1 – *Hiring Analytics Link*

Insights : From The above data we can get a gist of the hiring data filtered by Hired Employee (so that only we get the Hired employees). We can see that around 58% male & 42% females have been hired in the company.

Task 2 : Salary Analysis

Question : What is the average salary offered by this company?

Figure 3) Q2 - *Salary Analysis (Genderwise & By Department))*

Insights :

From the above two pivot tables and bar chart we can see the average salary of the organization

1. Considering the overall average salary of the organisation of all the hired personal which comes around **49983**.

2. In 2nd observation I made slight change and calculated the average by genderwise and sorted them by department so we get the deep insights about the –

   a. Overall average by genderwise (when we zoom out the pivot)

   b. The distrbution of salary by gender in various department and get the average salary of the same.

Task 3 – Salary Distribution

Question : Create class intervals for the salaries in the company.

Salary	Employee Counts
1-20001	928
20001-40001	943
40001-60001	1021
60001-80001	929
80001-100001	873
180001-200001	1
280001-300001	1
380001-400001	1
Grand Total	**4697**
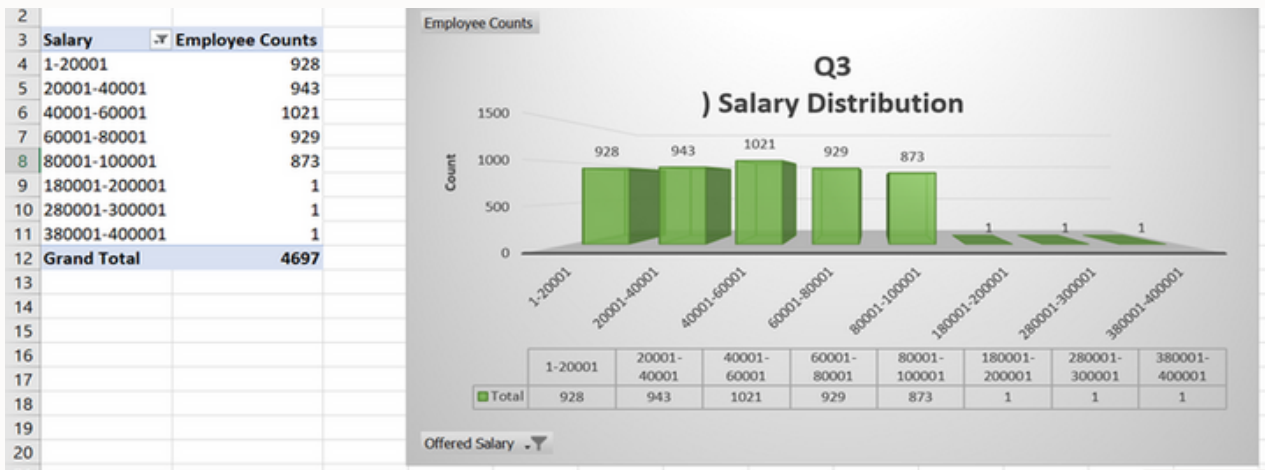
KEY FINDINGS

1. From the above insight we can have made few class interval on the salary data by the range of 20000 and then summarized the value of counts and got a pivot chart which shows the number of salary in that range.
2. The rough idea shows that there is a data outlier in the salary range above 1.8L,3L & 4L only single individual have the high salary while the graph indicates that majority of the salary lies between 40k-60k and same with other trend which are visible.

Task 4 – Departmental Analysis

Question : Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments

Status	Hired	
Row Labels	Count of Employee	
Finance Department	176	
General Management	113	
Human Resource Department	70	
Marketing Department	202	
Operations Department	1843	
Production Department	246	
Purchase Department	230	
Sales Department	485	
Service Department	1332	
Grand Total	**4697**	
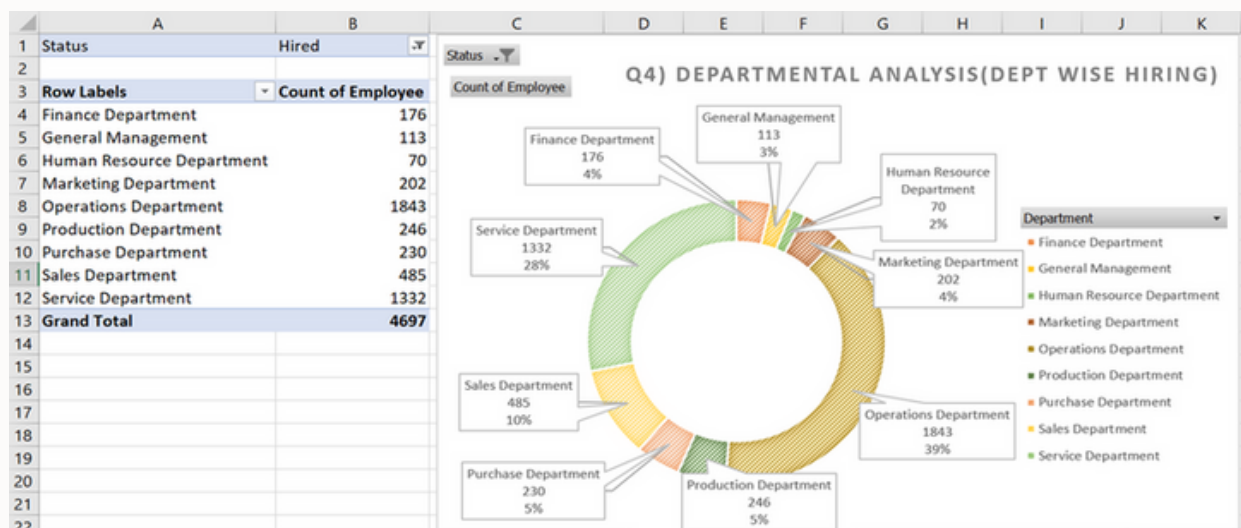
KEY FINDINGS

- From the above pie chart we can get the insights of the hired employees working in various departments of the organization.
- Pie charts help us to get the overall idea of data distribution in a particular group.

Task 5 – Position Tier Analysis

Question : Use a chart or graph to represent the different position tiers within the company

	A	B
3	Row Labels ▾	Employee Count
4	b9	291
5	c10	99
6	c5	1121
7	c8	182
8	c9	1158
9	i1	144
10	i4	29
11	i5	472
12	i6	325
13	i7	594
14	m6	2
15	n6	1
16	NA	1
17	Grand Total	4419
18		
19	Note : I changed - value with	
20	NA for reference in Postion	

Q5) POSITION TIER ANALYSIS

Employee Count

Post Name ▾

Post	Value
NA	
n6	
m6	
i7	594
i6	325
i5	472
i4	29
i1	144
c9	1158
c8	182
c5	1121
c10	99
b9	291

(x-axis: 0, 200, 400, 600, 800, 1000, 1200)
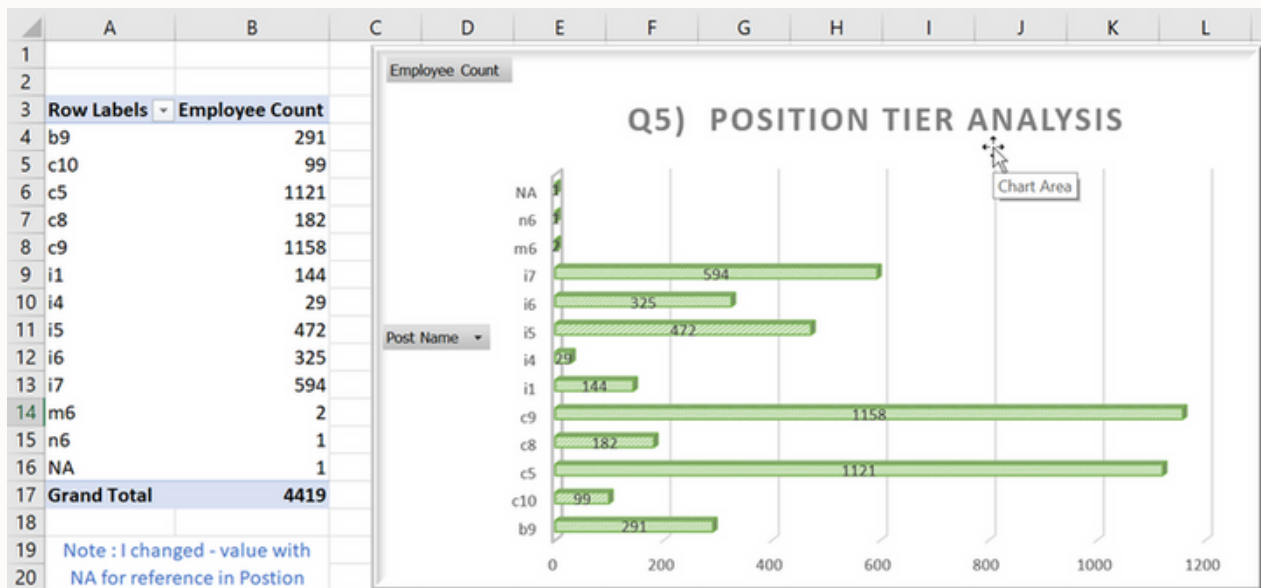
KEY FINDINGS

1. From the above observation we get to see the distribution of position at the organization.
2. The roles of c9,c5,i7 are among the top in the organization.
3. Pivot tables help us to drill down the data on various different points and also in forming various charts to represent the data.
4. Pivot makes the data sorting & filtering easy job.

PROJECT 4 - HIRING PROCESS ANALYTICS

Project Summary: As a data analyst, your task is to analyze the company's hiring process data and draw meaningful insights from it. The hiring process is a crucial function of any company, and understanding trends such as the number of rejections, interviews, job types, and vacancies can provide valuable insights for the hiring department.

1
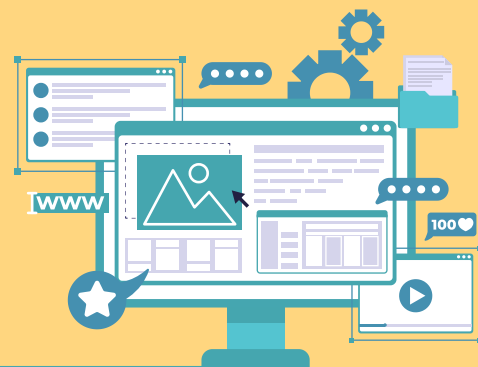
PROBLEM/PROJECT DESCRIPTION

- As a data analyst analyze the given dataset to get relevant insights for the questions asked by the hiring team & try to formulate meaningful & useful insights of the given dataset.
- Showcase use of Excel analysis using Pivot and charts.

2

APPROACH (RESEARCH)

1. Imported the given data in excel and then eye-balled to check for missing or random data.
2. Queried on the task and used appropriate excel function and pivot charts to analyze the data.
3. Formed the insights using visuals.

3

TECH-STACK USED

- *Google Sheets* - To get column stats & basic details of data like empty cells.
- *Excel* - Analysis and data visualization.
- *Word* - Report building for insights.
- *Google Drive* - Sharing the files used.

4

INSIGHTS

- The given dataset contains the details of Job hiring like salary range, gender ratio, departments etc.
- Using excel we analyzed the dataset on various useful parameters to get gist of the distribution of the organization.
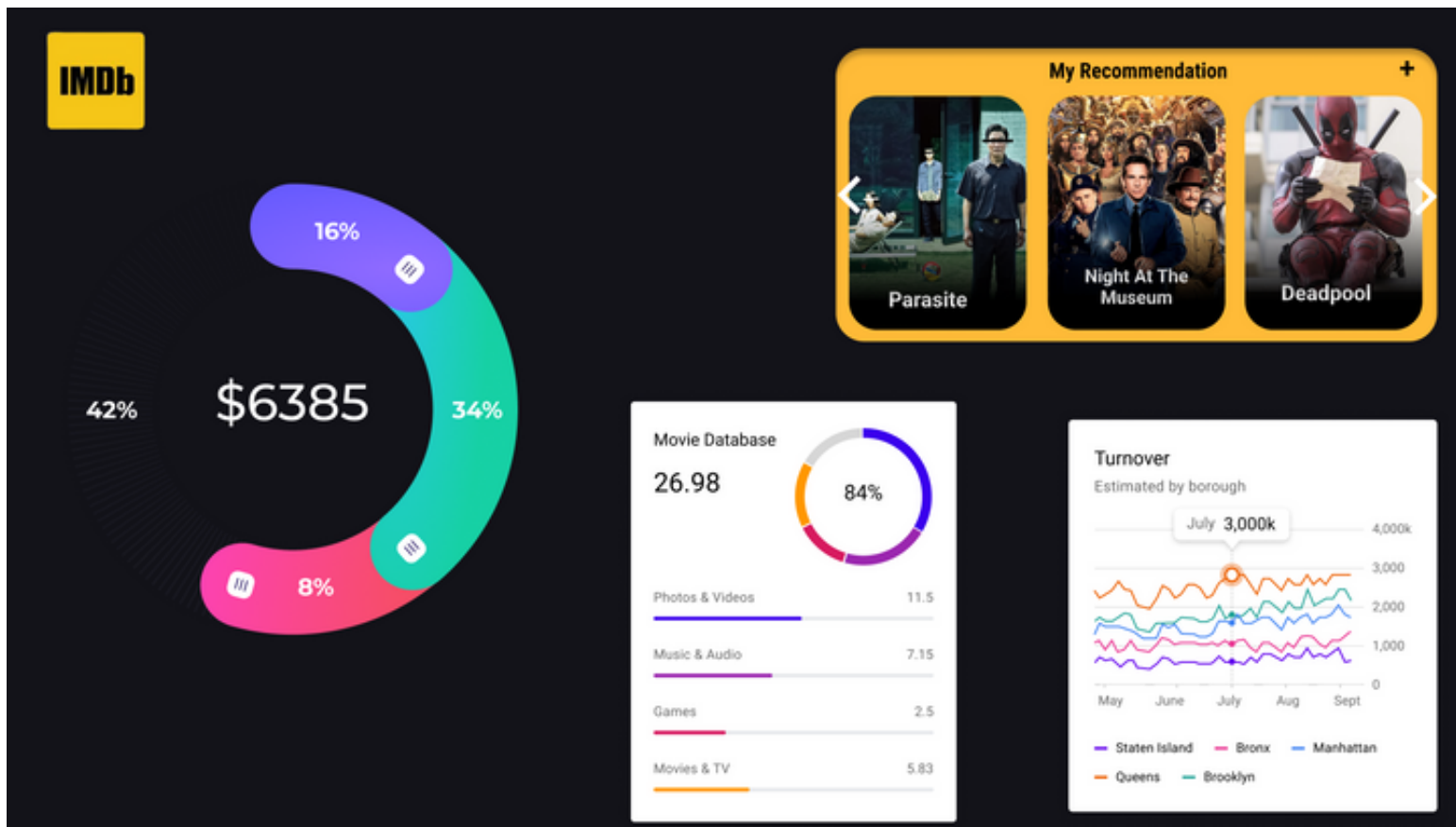
5

RESULTS

1. *The project helped me gain hands-on working on Excel sheet and the usefulness of charts and various functions.*
2. *Use of excel changes the plain simple data into meaningful piece of information.*

📌 <u>Full Report</u> 🔗

Project - 5

# IMDB MOVIE ANALYSIS

Descriptive Statistics

Project 5 – IMDB MOVIE ANALYTICS

Objective : The dataset provided is related to IMDB Movies. A potential problem to investigate could be: "What factors influence the success of a movie on IMDB?" Here, success can be defined by high IMDB ratings. The impact of this problem is significant for movie producers, directors, and investors who want to understand what makes a movie successful to make informed decisions in their future projects.

Task 1 : Movie Genre Analysis

Question : Determine the most common genres of movies in the dataset & impact on IMDB score. Then, for each genre, calculate descriptive statistics.

3	genres3	(All)					
4	genres42	(All)					
5	genres5	(All)					
6	genres6	(All)					
7	genres7	(All)					
8	genres8	(All)					
9	Primary Genre			Range of score			
10	Row Labels	Count of imdb_score	Min of imdb_score	Max of imdb_score	Mean of imdb_score	Var of imdb_score	StdDev of imdb_score
11	⊞ Action	935	2.1	9	6.285989305	1.078186788	1.038357736
12	⊞ Adventure	367	2.3	8.6	6.561307902	1.264345826	1.124431334
13	⊞ Animation	46	4.5	8	6.763043478	0.945937198	0.972593028
14	⊞ Biography	206	4.5	8.9	7.151941748	0.489337675	0.69952675
15	⊞ Comedy	1026	1.9	8.8	6.164424951	1.074567328	1.036613393
16	⊞ Crime	252	3.3	9.3	6.945238095	0.75475811	0.868768157
17	⊞ Documentary	42	1.6	8.4	6.914285714	1.994425087	1.412241158
18	⊞ Drama	675	2.1	8.8	6.822518519	0.821539642	0.90638824
19	⊞ Family	3	5.7	7.9	6.5	1.48	1.216552506
20	⊞ Fantasy	35	4.3	7.9	6.234285714	0.799378151	0.894079499
21	⊞ Horror	156	2.3	8.5	5.813461538	1.015624069	1.007781757
22	⊞ Musical	2	6.3	7.2	6.75	0.405	0.636396103
23	⊞ Mystery	23	3.3	8.5	6.586956522	1.23027668	1.109178381
24	⊞ Romance	2	6.2	7.1	6.65	0.405	0.636396103
25	⊞ Sci-Fi	8	5	8.2	6.5875	1.064107143	1.031555691
26	⊞ Thriller	3	4.8	6.3	5.3	0.75	0.866025404
27	⊞ Western	3	4.1	8.9	6.766666667	5.973333333	2.444040371
28	Grand Total	3784	1.6	9.3	6.461971459	1.117810651	1.057265648
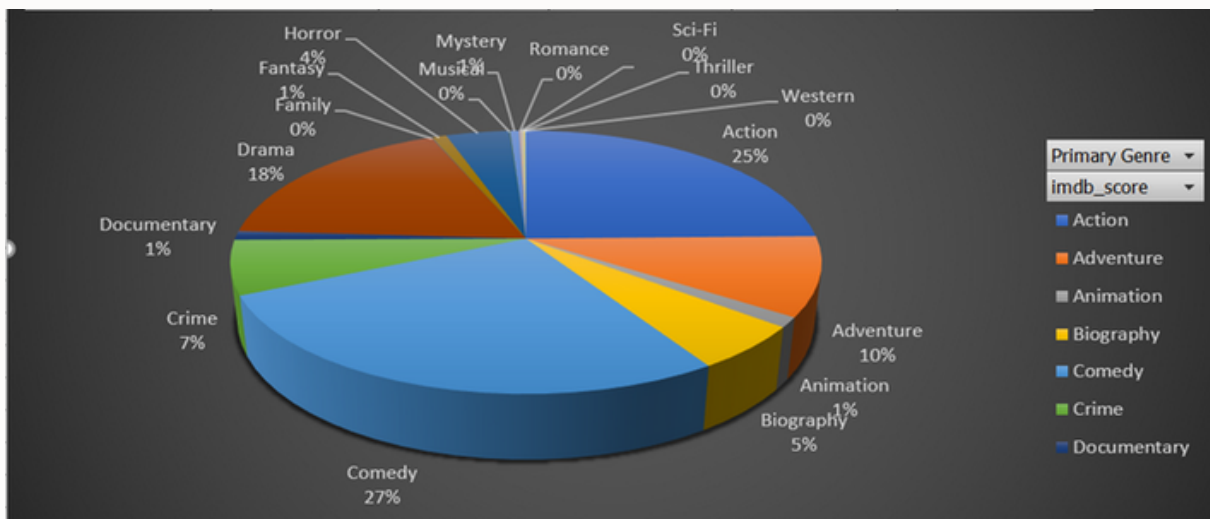
Descriptive Statistics (For primary Genre) [Using Data Analysis Fucntion in Excel]																
Statistics	Action	Adventure	Animation	Biography	Comedy	Crime	Documentary	Drama	Family	Fantasy	Horror	Musical	Mystery	Romance	Sci-Fi	Thriller	Western
Mean	5.876190476	6.035294118	6.54	6.965714286	5.370769231	7.085365854	6.786363636	6.105769231	6.9	6.183333333	5.97380952 4	7.2	6.686666667	6.65	6.814285714	6.3	8.1
Standard Error	0.235829465	0.223112098	0.200582485	0.186202332	0.241397405	0.199498181	0.300597922	0.223714923	1	0.20900616	0.20337054	0	0.267771521	0.45	0.329811325	0	0.8
Median	5.9	6.1	6.7	7	5.4	7.1	7.15	6.15	6.9	6.3	5.95	7.2	6.7	6.65	6.4	6.3	8.1
Mode	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Standard Deviation	1.871838348	1.593339076	1.002912426	1.101587853	1.946208098	1.277411636	1.40992923	1.613231255	1.414213562	1.02391689	1.317991732	#DIV/0!	1.03707464	0.636396103	0.872598746	#DIV/0!	1.13137085
Sample Variance	3.503778802	2.538729412	1.005833333	1.213495768	3.787725962	1.631780488	1.987900433	2.602515083	2	1.048405797	1.737102207	#DIV/0!	1.07552381	0.405	0.761428571	#DIV/0!	1.28
Kurtosis	-1.125806711	-0.939945976	-0.939097448	-1.032947109	-1.148817444	-0.786602835	2.34645897	-0.706332748	#DIV/0!	-0.892978657	-0.97450046	#DIV/0!	0.024181082	#DIV/0!	-1.075431446	#DIV/0!	#DIV/0!
Skewness	-0.063033373	-0.204231668	-0.399097448	-0.08743408	-0.038496732	-0.121222876	-1.512543695	-0.21550956	#DIV/0!	-0.20882416	0.076708978	#DIV/0!	-0.277952541	#DIV/0!	0.71753062 4	#DIV/0!	#DIV/0!
Range	6.6	5.8	3.3	3.9	6.8	5.1	5.7	6.7	2	3.6	4.9	0	3.9	0.9	2.3	0	1.6
Minimum	2.4	2.8	4.7	5	2	4.2	2.7	2.1	5.9	4.3	3.6	7.2	4.6	6.2	5.9	6.3	7.3
Maximum	9	8.6	8	8.9	8.8	9.3	8.4	8.8	7.9	7.9	8.5	7.2	8.5	7.1	8.2	6.3	8.9
Sum	370.2	307.8	163.5	243.8	349.1	290.5	149.3	317.5	13.8	148.4	250.9	7.2	100.3	13.3	47.7	6.3	16.2
Count	63	51	25	35	65	41	22	52	2	24	42	1	15	2	7	1	2
Largest(2)	8.9	8.5	7.9	8.7	8.6	9.2	8.3	8.7	5.9	7.6	8.2	1	8.1	6.2	7.7	6.3	8.9
Smallest(2)	2.7	3	4.8	5.1	2.1	5.1	4.1	3.4	7.9	4.4	3.9	1	4.6	6.2	5.9	6.3	7.3

Note: Mode is shown as NA in calculation as No imdb score seems to be repeated for the genre wise distribution

Task 2 – Movie Duration Analysis

- **Question :** Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

KEY FINDINGS

1. From the Pivot table it is evident that most movies are in the range of 84-108minutes long and also in range of 109-133 minutes on average count where people like to rate them on imdb.
2. The analysis shows that the movies in range of 84-133 minutes are among the most rated movies or we could assume are favoured by people more to watch & rate.

Task 3– Language Analysis

Question : Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.

KEY FINDINGS

1. From the analysis of the data it is evident that most of the rated movies on IMDB are from the **English origin (3605 Movie titles)**.
2. Other Prominent languages include **French (37), Spanish (26), Mandarin (14) , German (13) , Japanese (12)** among other languages.
3. The analysis shows that majority of the user watch Hollywood movies AKA English (US) and they are among the most watched languages.

Task 4 – Director Analysis

- **Question :** Identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.

Row Labels	Count of imdb_score	Average of imdb_score	Sum of imdb_score	Percentile (WholeData)	Ranking Acc to IMDB_Score
Steven Spielberg	25	7.544	188.6	0.77%	1
Woody Allen	19	7	133	0.54%	3
Clint Eastwood	19	7.205263158	136.9	0.56%	2
Ridley Scott	16	7.13125	114.1	0.47%	5
Martin Scorsese	16	7.675	122.8	0.50%	4
Spike Lee	15	6.733333333	101		6
Steven Soderbergh	15	6.68	100.2		7
Renny Harlin	15	5.746666667	86.2		12
Tim Burton	14	7.05	98.7		8
Robert Zemeckis	13	7.307692308	95	0.39%	9
Oliver Stone	13	6.907692308	89.8	0.37%	11
Ron Howard	13	6.930769231	90.1	0.37%	10
Robert Rodriguez	13	5.692307692	74	0.30%	20
Barry Levinson	13	6.576923077	85.5	0.35%	13
Tony Scott	12	6.791666667	81.5	0.33%	14
Michael Bay	12	6.616666667	79.4	0.32%	16
Joel Schumacher	12	6.341666667	76.1	0.31%	19
Shawn Levy	11	6.090909091	67	0.27%	28
Rob Reiner	11	7.018181818	77.2	0.32%	18
Richard Linklater	11	7.327272727	80.6	0.33%	15
Chris Columbus	11	6.654545455	73.2	0.30%	21
Wes Craven	10	5.97	59.7	0.24%	34
Sam Raimi	10	6.96	69.6	0.28%	23
Paul W.S. Anderson	10	5.99	59.9	0.24%	33

(Tooltip: Percentile (WholeData) — Value: 0.50% — Row: Martin Scorsese — Column: Percentile (WholeData))

Movie Director & its Impact On IMDB Score

KEY FINDINGS

1. Certain Directors like **Steven Spielberg, Client Eastwood, Woody Allen, Martin Scorsese** are among the highest rated director Movies and also have the most number of movies.
2. The extra rank is made according to the average rating of directors and table is sorted by the number of movies of each director – *Total directors are = 1751*

Task 5 – Budget Analysis

Question : Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

movie_title	title_year	director_name	imdb_score	language	country	budget	gross	Profit	Profit(In Million)
Avatar	2009	James Cameron	7.9	English	USA	237000000	760505847	523505847	523.51M
Jurassic World	2015	Colin Trevorrow	7	English	USA	150000000	652177271	502177271	502.18M
Titanic	1997	James Cameron	7.7	English	USA	200000000	658672302	458672302	458.67M
Star Wars: Episode IV - A New Hope	1977	George Lucas	8.7	English	USA	11000000	460935665	449935665	449.94M
E.T. the Extra-Terrestrial	1982	Steven Spielberg	7.9	English	USA	10500000	434949459	424449459	424.45M
The Avengers	2012	Joss Whedon	8.1	English	USA	220000000	623279547	403279547	403.28M
The Lion King	1994	Roger Allers	8.5	English	USA	45000000	422783777	377783777	377.78M
Star Wars: Episode I - The Phantom Menace	1999	George Lucas	6.5	English	USA	115000000	474544677	359544677	359.54M
The Dark Knight	2008	Christopher Nolan	9	English	USA	185000000	533316061	348316061	348.32M
The Hunger Games	2012	Gary Ross	7.3	English	USA	78000000	407999255	329999255	330M
Deadpool	2016	Tim Miller	8.1	English	USA	58000000	363024263	305024263	305.02M
The Hunger Games: Catching Fire	2013	Francis Lawrence	7.6	English	USA	130000000	424645577	294645577	294.65M
Jurassic Park	1993	Steven Spielberg	8.1	English	USA	63000000	356784000	293784000	293.78M
Despicable Me 2	2013	Pierre Coffin	7.5	English	USA	76000000	368049635	292049635	292.05M
American Sniper	2014	Clint Eastwood	7.3	English	USA	58800000	350123553	291323553	291.32M
Finding Nemo	2003	Andrew Stanton	8.2	English	USA	94000000	380838870	286838870	286.84M
Shrek 2	2004	Andrew Adamson	7.2	English	USA	150000000	436471036	286471036	286.47M
The Lord of the Rings: The Return of the King	2003	Peter Jackson	8.9	English	USA	94000000	377019252	283019252	283.02M
Star Wars: Episode VI - Return of the Jedi	1983	Richard Marquand	8.4	English	USA	32500000	309125409	276625409	276.63M
Forrest Gump	1994	Robert Zemeckis	8.8	English	USA	55000000	329691196	274691196	274.69M
Star Wars: Episode V - The Empire Strikes Back	1980	Irvin Kershner	8.8	English	USA	18000000	290158751	272158751	272.16M
Home Alone	1990	Chris Columbus	7.5	English	USA	18000000	285761243	267761243	267.76M
Star Wars: Episode III - Revenge of the Sith	2005	George Lucas	7.6	English	USA	113000000	380262555	267262555	267.26M
Spider-Man	2002	Sam Raimi	7.3	English	USA	139000000	403706375	264706375	264.71M
Minions	2015	Kyle Balda	6.4	English	USA	74000000	336029560	262029560	262.03M

Tabs: Cleaned-Data (Removed rows) | Sheet2 | Genre-Separation(Text-Column) | Q1) Genre Analysis

KEY FINDINGS

1. The top movie from our data is **Avatar** at a staggering **$535Million in profits** it had a budget of $237M so its *almost 2X of investment* (Note – the data is not upto mark as Avatar is supposed to be around $2.2B according to sources but our data might be just of US & of old origins).
2. The other top 25 movies are shown in the screenshots.

PROJECT 5 - IMDB MOVIE ANALYTICS

Project Summary: Imagine you're a data analyst working with the product team at Instagram. Your role involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

PROBLEM/PROJECT DESCRIPTION

1
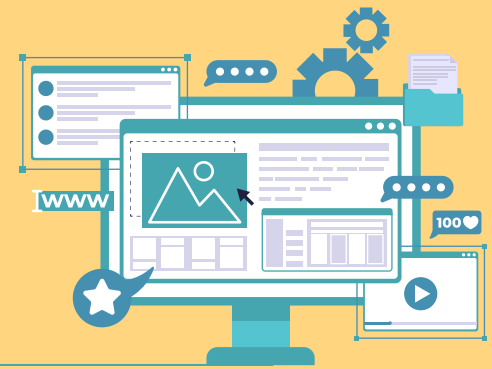
- The Given dataset contains a list of various movie from IMDB database we need to find out the factors impacting the scores and user choice of a movie.
- The dataset can be analyzed on factors like director, budget, critic scores, cast, genre, language etc.

APPROACH (RESEARCH)

2

1. Importing the data & glancing.
2. Removing Unknowns or blanks by treating using appropriate methods.
3. After final processing using excel or other tools to solve the tasks.
4. Using data visualization to show the results.

TECH-STACK USED

3

- *Google Sheets - To get column stats & basic details of data like empty cells.*
- *Excel - Analysis and data visualization.*
- *Word - Report building for insights.*
- *Google Drive - Sharing the files used.*
- *ShareX - Screenshots & Gif Recording.*

INSIGHTS

4

- The dataset contains large info about the factors impacting the movie popularity like genre, cast, director.
- The project focuses on advanced excel or other visualization library like (matplotlib , seaborn in python) to showcase the output.
- Using descriptive statistics we get to delve more useful insights from the dataset.

5

RESULTS

1. *The project helped me gain strong understanding of advanced excel functions , pivot charts, descriptive statistics and much more.*
2. *I also explored the dataset using python and visualization library to compare the results with excel cleaning & reporting.*
3. The tasks helped to understand about excel functions & overall about data analysis.

📌 <u>Full Report</u> 🔗

Project - 6

Loan Case Study

BANK LOAN CASE STUDY

Exploratory Data Analysis (EDA)

Project 6 – **Bank Loan Case Study**

Objective : Imagine you're a data analyst at a finance company that specializes in lending various types of loans to urban customers. Your company faces a challenge: some customers who don't have a sufficient credit history take advantage of this and default on their loans. Your task is to use Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.

Task 1 : Identify Missing Data and Deal with it Appropriately

Question : Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.
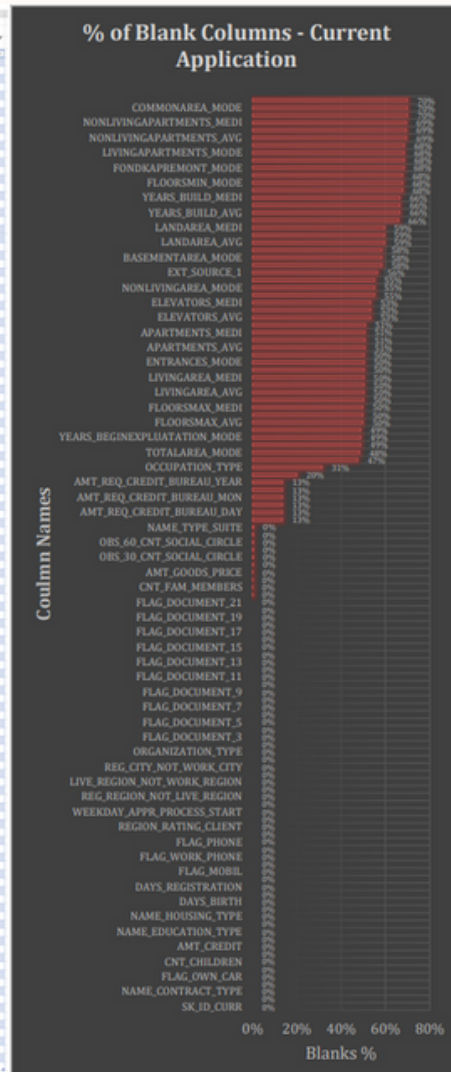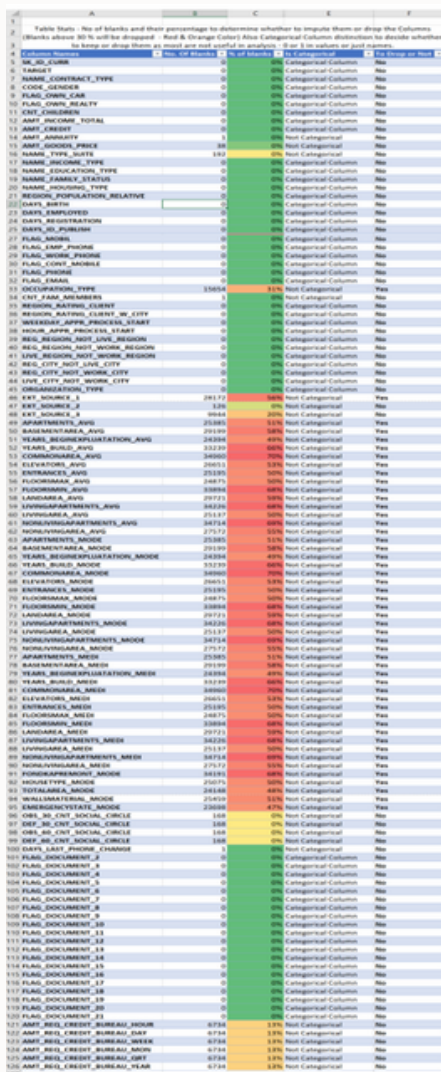
INSIGHTS :

- To clean the given dataset we first check the no. of blanks & its %.
- After decision we drop the columns having more than 35% blanks in them completely as filling them would change the analysis a lot.
- For the rest the blanks can be treated using appropriate descriptive stats.
- For filling the columns we can use the special > go to function for selecting all blanks in a column & then filling up with appropriate stats.
- Also duplicates or random things can be dropped instead of treating them.

Task 2 – Identify Outliers in the Dataset

Question : Detect and identify outliers in the dataset using Excel statistical functions and features, focusing on numerical variables.
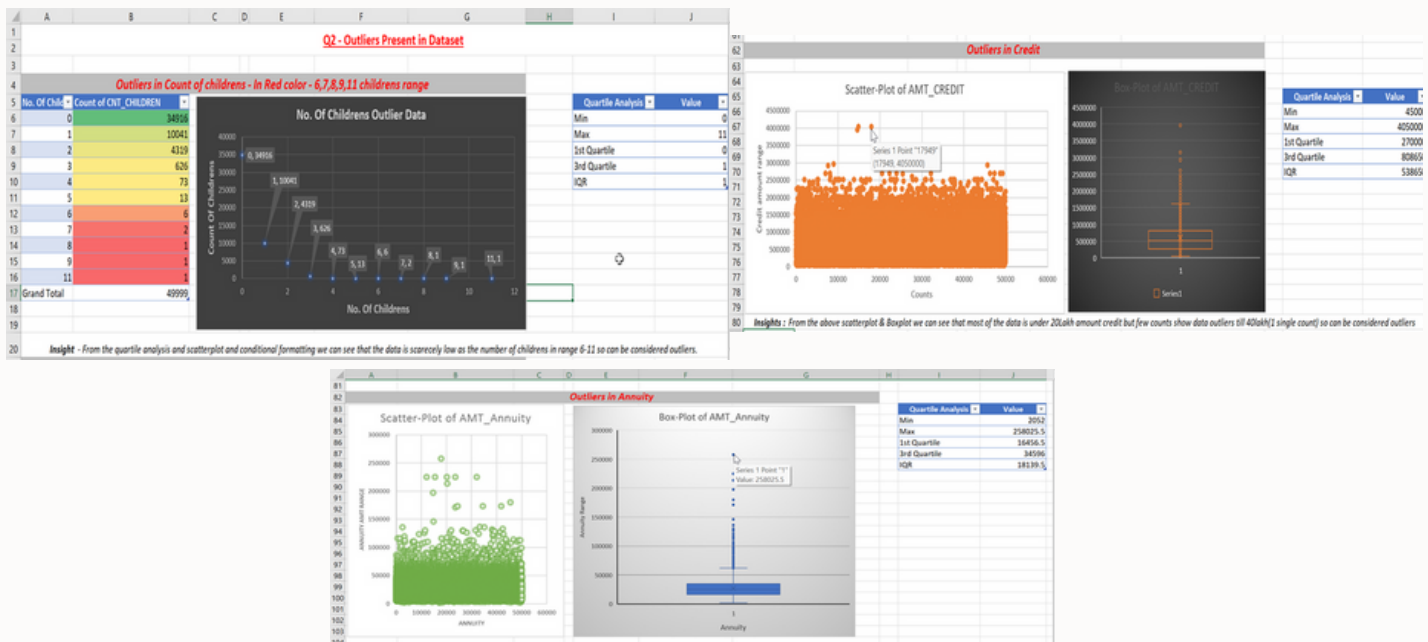
KEY FINDINGS

1. There are many outlier present in the dataset. (See attached Report for more)
2. The outlier needs to be capped for ideal analysis but removing them will drastically impact the dataset if we have columns like price or salary in it.

Task 3 – Analyze Data Imbalance

Question : Determine if there is data imbalance in the loan application dataset and calculate the ratio of data imbalance using Excel functions.
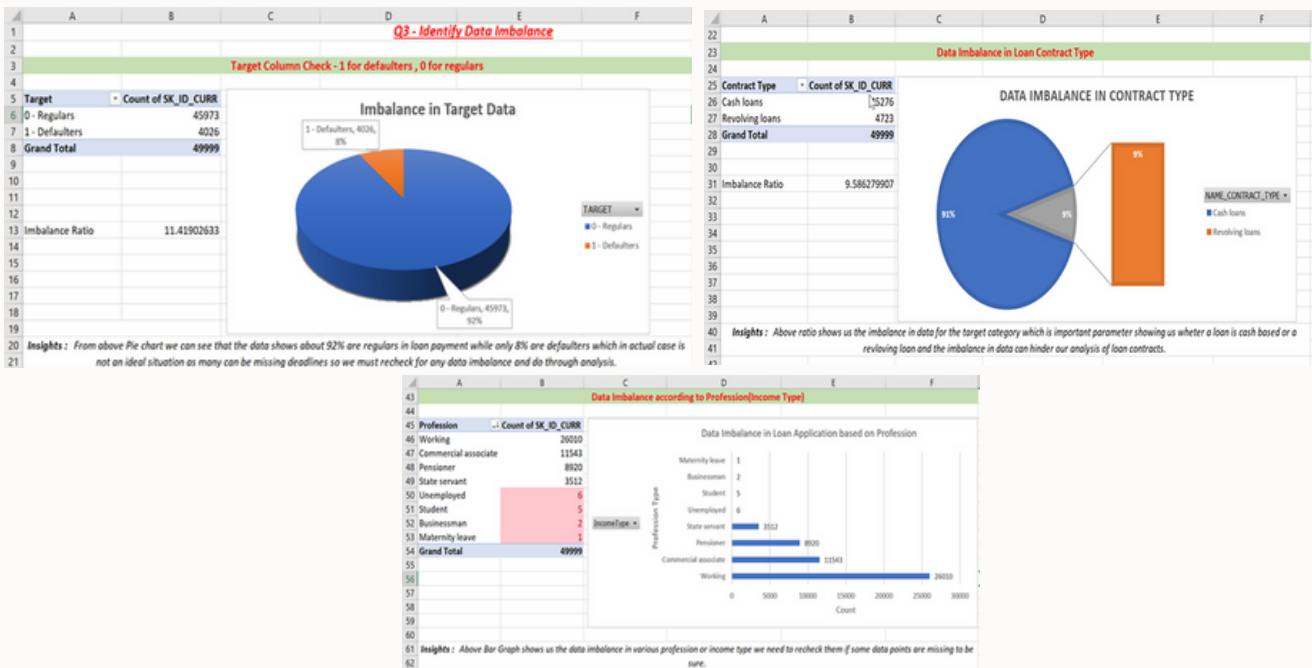
INSIGHTS :

1. From the above analysis which we did using pivot table and basic functions like COUNT, COUNTIF & SUM we got the data imbalance.
2. For better results and analysis we must recheck the data if the given data inputs are valid and are not having error will help us to get insights.

Task 4 – Perform Univariate, Segmented Univariate, and Bivariate Analysis

Question : Perform univariate analysis to understand the distribution of individual variables, segmented univariate analysis to 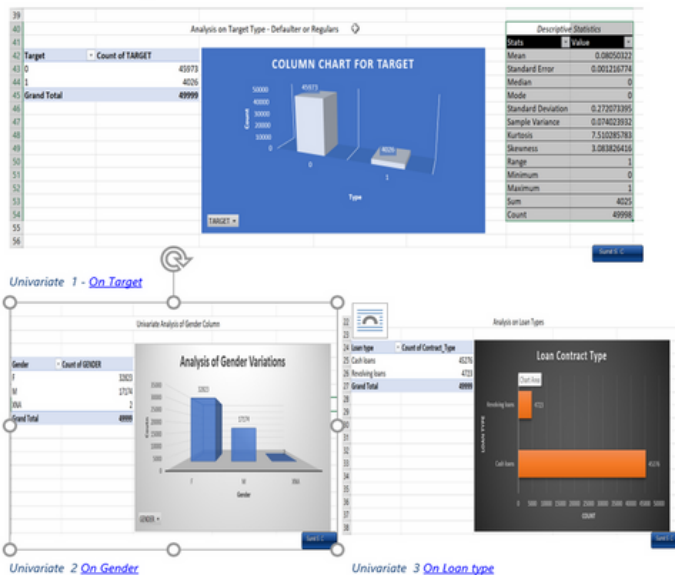compare variable distributions for different scenarios, and bivariate analysis to explore relationships between variables and the target variable using Excel functions and features.
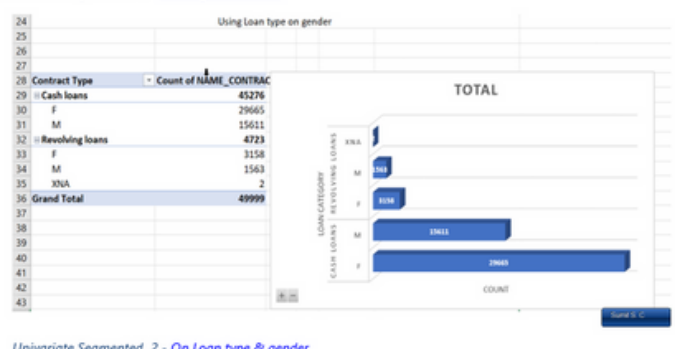

Univariate Analysis – When we examine and analyse a single variable in isolation and understand the distribution summary stats using descriptive stats and try to gain information its called as univariate analysis.

Univariate 1 - On Target

Univariate 2 On Gender

Univariate 3 On Loan type

Univariate 3 On Age

Univariate 4 On Loan Amount

Univariate 5 on estate ownership

Univariate 6 On Income Type

Segmented Univariate Analysis – When we examine and analyse a single variable with different segments or subgroups defined by another variable and understand the distribution summary stats using descriptive stats and try to gain information its called as segmented univariate analysis.
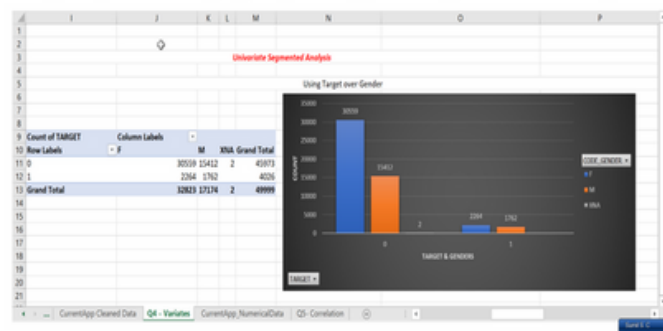
Univariate Segmented 1 - On Target & Gender

Univariate Segmented 2 - On Loan type & gender

Bivariate Analysis – When we examine and analyse two variables simultaneously and understand the distribution summary stats using descriptive stats and try to gain information its called as bivariate analysis. The purpose is to identify the relationship between 2 variables to understand the patterns that change the data corresponding to other data in comparisons.
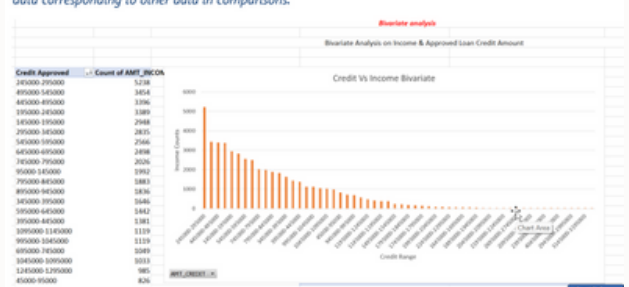
Bivariate 1 Column Bar GraphOn Credit amount & Credit approved
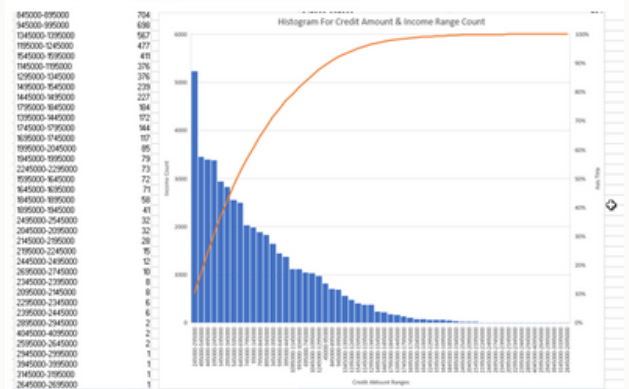
KEY FINDINGS

1. In the univariate analysis we have selected a single column and checked its descriptive stats and plotted a graph to see its distribution.
2. In Segmented univariate analysis we basically divide the single variable and further deeply analysis it says on target type on gender roles or say by loan type and drill the column for more better insights on the column.
3. In Bivariate analysis we compare two variables and compare them on a single feature to see the impacts of the comparison on particular analysis.

Task 5 – Identify Top Correlations for Different Scenarios

<u>Question</u> : Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.

KEY FINDINGS

1. In correlation comparison we do the comparison of each column with oneself and with other column in question and thus a matrix is formed of the same length as the column numbers and the diagonal is equal to 1 (since relation of same column comparison.)
2. From the above analysis we have found out about the top 15 factors that might impact the loan defaults as these factors play a role in individuals spending and savings habit etc.

PROJECT 6 -BANK LOAN CASE STUDY

Project Summary: Imagine you're a data analyst at a finance company that specializes in lending various types of loans to urban customers. Your company faces a challenge: some customers who don't have a sufficient credit history take advantage of this and default on their loans. Your task is to use Exploratory Data Analysis (EDA) to analyze patterns in the data and ensure that capable applicants are not rejected.
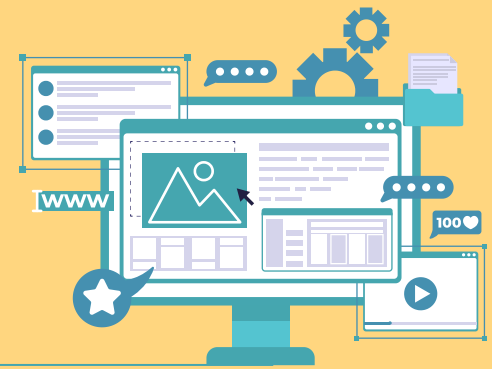
PROBLEM/PROJECT DESCRIPTION

- The given dataset contains the information about the loans filled by user.
- The dataset can be analyzed on factors like family background, no. of children, education, income, profession, loan types etc.
- We need to check for the eligibility & loan defaults of the applicants.

APPROACH (RESEARCH)

1. Importing the data & glancing.
2. Removing Unknowns or blanks by treating using appropriate methods.
3. After final processing using excel or other tools to solve the tasks.
4. Using data visualization to show the results & using Correlation to get gist of factors.

TECH-STACK USED

- *Google Sheets* - To get column stats & basic details of data like empty cells.
- *Excel* - Analysis and data visualization.
- *Word* - Report building for insights.
- *Google Drive* - Sharing the files used.
- ShareX - Screenshots & Gif Recording.

INSIGHTS

- The dataset contains large info about the factors impacting the loan approvals & defaulting.
- The project focuses on advanced excel or other visualization library like (matplotlib , seaborn in python) to showcase the output.
- Using EDA we get to know about the factors revolving around consumers that impact the loan and payments.

RESULTS

1. *The project helped me gain strong understanding of advanced excel functions , pivot charts, descriptive statistics and much more like the concepts of defaulting, correlation.*
2. *I also explored the dataset using python and visualization library to compare the results with excel cleaning & reporting.*
3. The tasks helped to understand about advanced excel functions & overall about data analysis.

📌 Full Report 🔗

*

Project - 7

CLICKS LAST 7 DAYS
6,113
↑ 111%

CTR TODAY
1.80%
↓ 27%

SOCIAL CTR TODAY
0.50%
↓ 44%

PEOPLE TAKING ACTION LAST 7 DAYS
49,212

SPENT TODAY
$ 204.45

METRICS OVERVIEW LAST 7 DAYS

KPIS	Last 7 Days
Reach	1.9M
Frequency	0.9
Social impressions	29,823
Actions	281,083
Clicks	39,190
CTR	1.80%
Total spent	$ 4,251.10
CPA	$ 0.01
CPC	$ 0.11

facebook Ads   My Website

Last sync an hour ago   Powered by ⬡ databox

# ANALYZING THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

Data Analysis & Dashboarding

Project 7 – **Analyzing the Impact of Car Features on Price and Profitability**

Objective : For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?

Task 1 : Popularity of Car

Question : Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores. Create a combo chart that visualizes relationship b/n market category & popularity
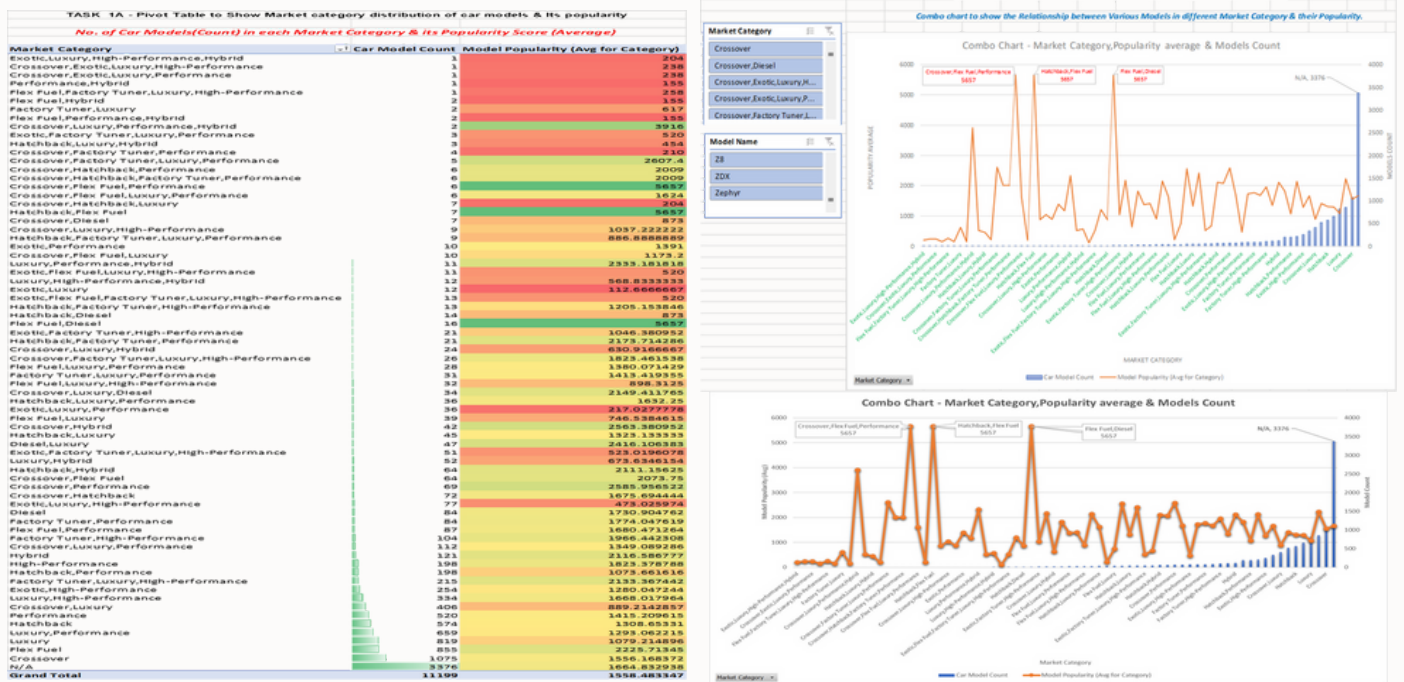
Task 2 : Cars Engine Power & its Price

Question : Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables
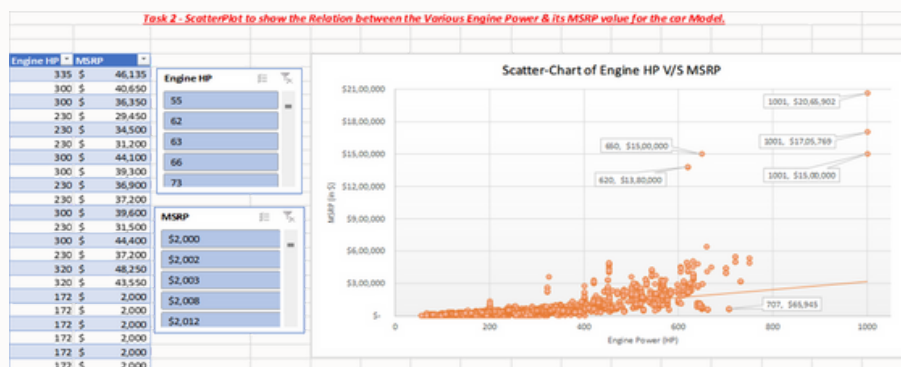
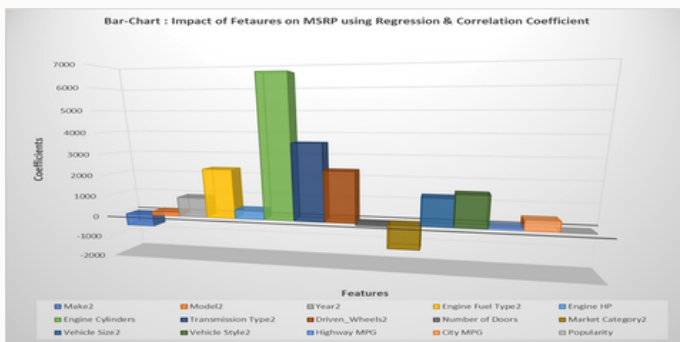### INSIGHTS :

- In the task , we analyzed the MSRP value by the engine HP using a scatterplot to check the distribution.
- It can be summarized that as the engine HP increases the price of car increases in linear relation.
- In our data Bugatti had outlier with 1001HP value and also the highest price at around 1.75M avg while the rest of data was below the 600HP line and below 0.6M MSRP line.

Task 3 : Features determining cars price

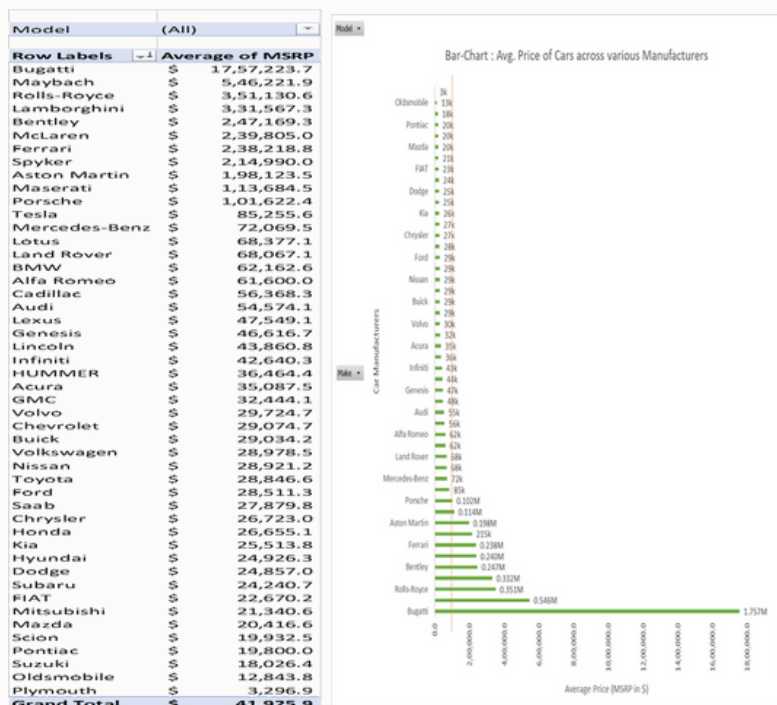**Question :** Determine the winner of the contest and provide their details to team.

INSIGHTS :

- From the new data and regression analysis we get the relationship between various features that indirectly impacts the price of cars.
- The basic bar-chart shows us that few values lie below the zero indicating that they might not be impacting much on the MSRP.

Task 4 : Price Across different Brands

Question : Create a pivot table that shows the average price of cars for each manufacturer. Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

Model	(All)	
Row Labels	**Average of MSRP**	
Bugatti	$	17,57,223.7
Maybach	$	5,46,221.9
Rolls-Royce	$	3,51,130.6
Lamborghini	$	3,31,567.3
Bentley	$	2,47,169.3
McLaren	$	2,39,805.0
Ferrari	$	2,38,218.8
Spyker	$	2,14,990.0
Aston Martin	$	1,98,123.5
Maserati	$	1,13,684.5
Porsche	$	1,01,622.4
Tesla	$	85,255.6
Mercedes-Benz	$	72,069.5
Lotus	$	68,377.1
Land Rover	$	68,067.1
BMW	$	62,162.6
Alfa Romeo	$	61,600.0
Cadillac	$	56,368.3
Audi	$	54,574.1
Lexus	$	47,549.1
Genesis	$	46,616.7
Lincoln	$	43,860.8
Infiniti	$	42,640.3
HUMMER	$	36,464.4
Acura	$	35,087.5
GMC	$	32,444.1
Volvo	$	29,724.7
Chevrolet	$	29,074.7
Buick	$	29,034.2
Volkswagen	$	28,978.5
Nissan	$	28,921.2
Toyota	$	28,846.6
Ford	$	28,511.3
Saab	$	27,879.8
Chrysler	$	26,723.0
Honda	$	26,655.1
Kia	$	25,513.8
Hyundai	$	24,926.3
Dodge	$	24,857.0
Subaru	$	24,240.7
FIAT	$	22,670.2
Mitsubishi	$	21,340.6
Mazda	$	20,416.6
Scion	$	19,932.5
Pontiac	$	19,800.0
Suzuki	$	18,026.4
Oldsmobile	$	12,843.8
Plymouth	$	3,296.9
Grand Total	**$**	**41,925.9**

INSIGHTS :

1. The pivot chart shows us the Task 4.A part to showcase the average MSRP for each manufacturer.
2. The bar chart in Task 4.B gives us a clear picture about the MSRP for various brands and we can even get the top brands and cheap brands directly from the categorization.
3. . The chart shows us that Bugatti brand is an ultra-luxury brand whose MSRP is almost double of the 2nd top brand, while the Plymouth is your daily average household secondhand cars range under $3k dollars.

# Task 5 : Relation Between Fuel Efficiency & no. of Cylinder Engines

Question : Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Also calculate the correlation coefficient .
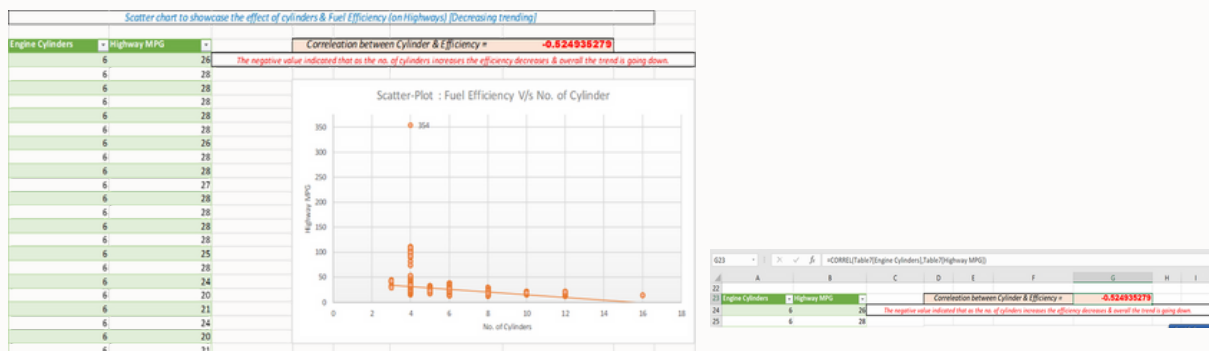
INSIGHTS :

- The task 5.A helped us analyze the relation between the no. of cylinder & fuel efficiency using the scatterplot.
- From the graph it is evident that as the number of cylinder increase it impacts the mileage of the cars.
- The task 5.B we did the CORREL of the two column values to get the coefficient of correlation which indicates us that there is negative trend .

# Dashboard 1 – Car Price (total) by brand & Body Style
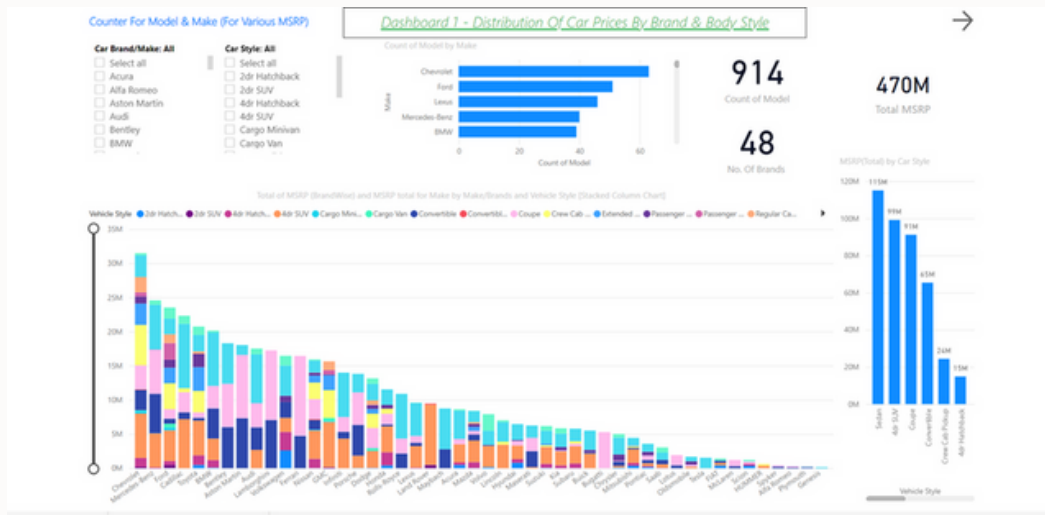
PowerBi Dashboard 🔗

KEY FINDINGS

1. The dashboard can be interacted to get the distribution of prices across brand & vehicle style.
2. We can see the impact of brands on MSRP showing loyalty as well as the preference for style when buying as a trend.
3. The stacked column takes into the various brands under style so we can see the dist. Neatly.

# Dashboard 2 : Car MSRP (Avg, Min, Max) & variation on body style.
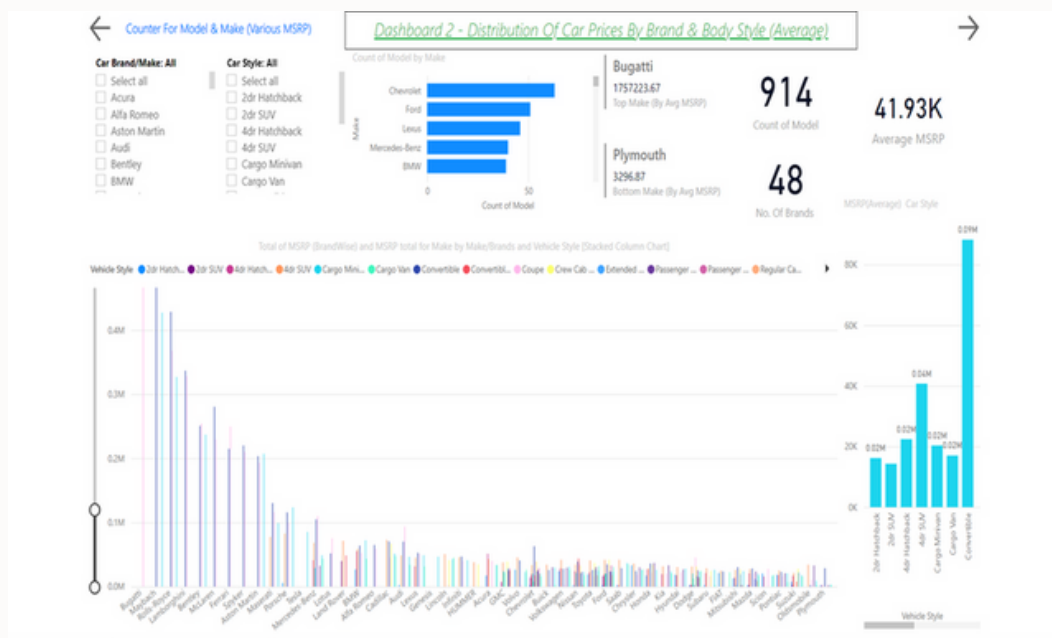
PowerBi Dashboard 🔗

KEY FINDINGS

1. The question helped us analyze the same analysis like previous but these time the difference is that we did the average MSRP calculation and used a bar chart which drills the data more and we get in depth knowledge.
2. Like previous insights we can toggle the filters & slicers to see the trends.

# Dashboard 3 - Effects of features on MSRP
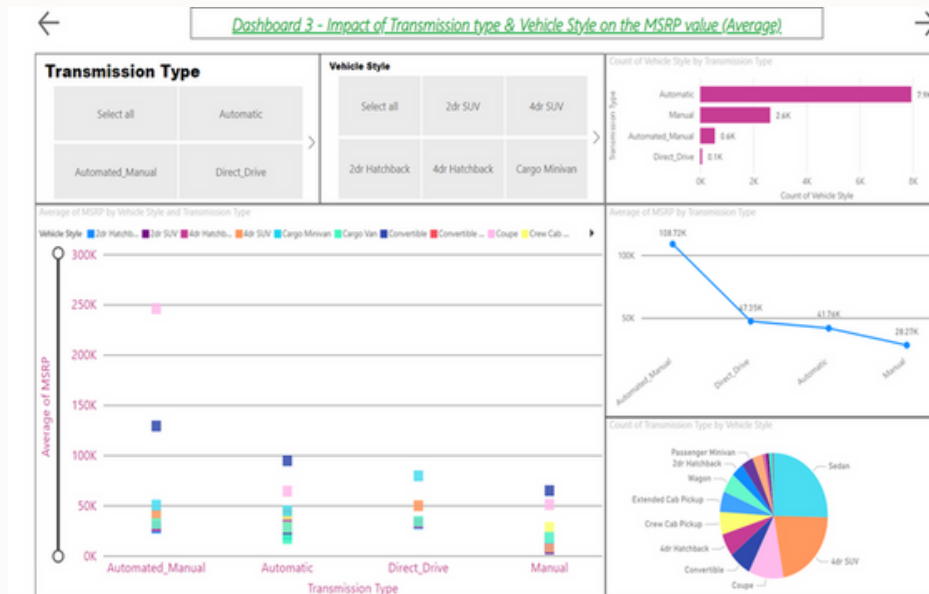
PowerBi Dashboard 🔗

KEY FINDINGS

1. The dashboard focus on seeing the distribution of MSRP based on vehicle type & transmission types.
2. It is interesting to see the distribution of cars based once using the transmission over scatterplot and once using the vehicle type the scattering changes shows us the variation in detail.
3.3. The other charts and counters bars helps us to get more refined insights.

Dashboard 4 : Fuel Efficiency across different body style & model years
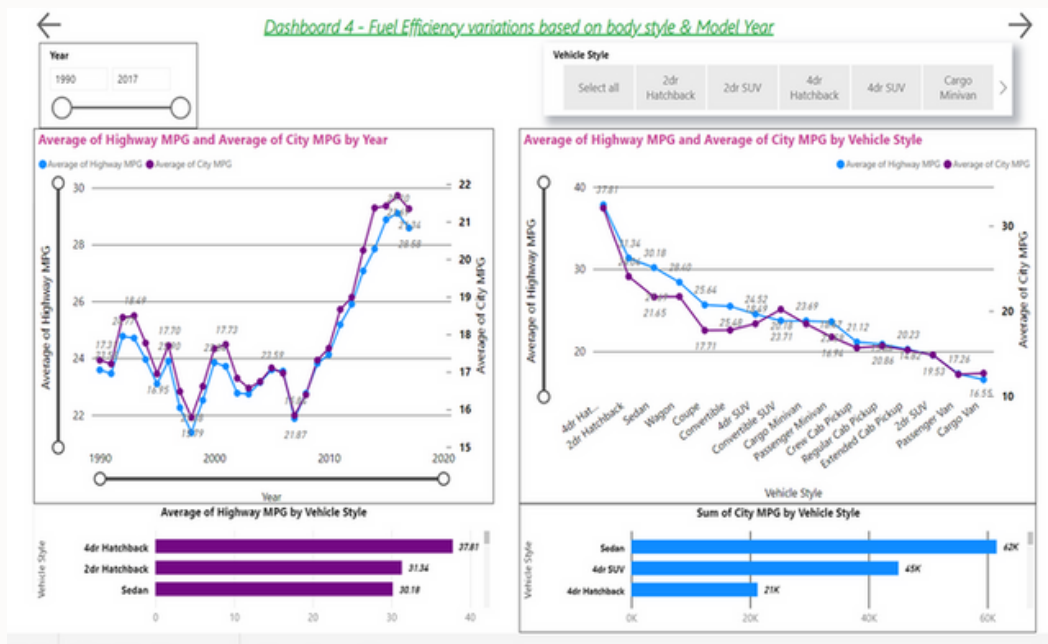

PowerBi Dashboard 🔗

KEY FINDINGS

1. The task helped us to make a dashboard to see the variation of fuel efficiency over the body style & year of the model of car.
2. The year slicer helps us to get to know how has the technology improved the mileage over the years and it does show significant changes.
3. The body style shows us the variation of mileage to showing us the impacts of different type of vehicles.

PROJECT 7 - ANALYZING THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

Project Summary: For the given dataset, as a Data Analyst, the client has asked How can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand?
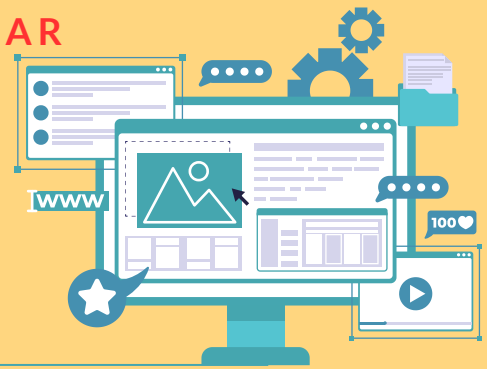
PROBLEM/PROJECT DESCRIPTION

- The given dataset contains the information about the various cars from Kaggle submission.
- The dataset can be analyzed on factors like brand, body style, no. of cylinders etc.
- We need to check for the factors impacting the overall MSRP..

APPROACH (RESEARCH)

1. Importing the data & glancing.
2. Removing Unknowns or blanks by treating using appropriate methods.
3. After final processing using excel or other tools to solve the tasks.
4. Using data visualization to show the results & making an interactive dashboard using BI Tool.

TECH-STACK USED

- *Google Sheets* - To get column stats & basic details of data like empty cells.
- *Excel* - Analysis and data visualization.
- *Word* - Report building for insights.
- *Power BI* - To make interactive dashboards.
- *Google Drive* - Sharing the files used.
- ShareX - Screenshots & Gif Recording.

INSIGHTS

- The dataset contains large info about the factors impacting the MSRP of cars & various factors that influences the choice of customers.
- The project focuses on advanced excel or other visualization library like (matplotlib , seaborn in python) to showcase the output.
- Using BI Tool make an interactive dashboard for the clients report.

RESULTS

1. *The project helped me gain strong understanding of advanced excel functions , pivot charts, descriptive statistics and much more like the concepts of regressions.*
2. *I also explored the dataset using python and visualization library to compare the results with excel cleaning & reporting.*
3. The tasks helped to understand about BI tools to create interactive dashboard to gain insights which the user or client can go through to get actual insights.

📌 <u>Full Report</u> 🔗

Project - 8

Tools to Optimize Your Customer Experience

Social Media Listening Tools: Listen to what customers are posting about your brand.

Suggestion Boxes: They don't have to be physical boxes, they can be an email address or a section of your support site.

Behavioral Analytics: Learn how customers react after visiting your website.

Customer Relationship Management (CRM): Easily track and manage customer relationships throughout their journey.

Surveys: Design questions that pertain to customers' unique journeys with your brand.

ABC CALL VOLUME TREND ANALYSIS

Predictive Analytics

Project 8 – ABC Call Volume Trend Analysis

Objective : Inbound customer support, which is the focus of this project, involves handling incoming calls from existing or prospective customers. The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.

Task 1 : Average Call Duration

- **Question :** What is the average duration of calls for each time bucket?
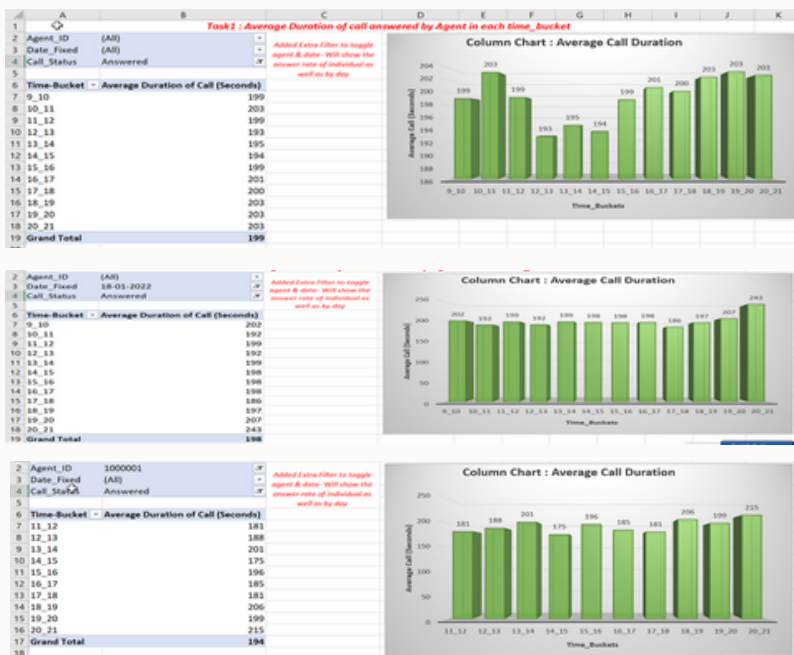
INSIGHTS :

- The task helped us drill down the average duration of calls received by the agents throughout the day. The average call duration of calls comes to be around 199 seconds.
- The slots 10_11,18-21 has higher average of 203 sec as compared with the overall average duration data.
- In our question it was mentioned to find the average of calls answered so we used a filter on status – "Answered" to get the call details of received calls by an agent, as an extra analysis I have added a filter on "Date" to so one can toggle the relevant details of each date if required.

Task 2 : Call Volume Analysis

Question : Create a chart or graph that shows the number of calls received in each time bucket?
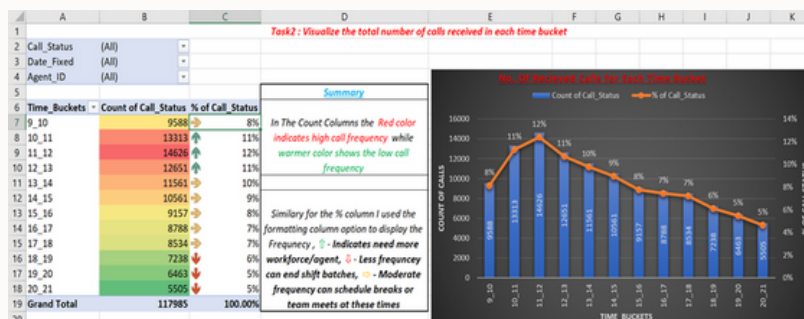
INSIGHTS :

- From the analysis we can see that the time period between 10am to 3pm are the busiest of all so we need to adjust the agents such that most are on line to answer the calls 11am-12pm being the busiest hr. of the shift.
- Also from the same analysis we can see that after 3 pm or so the call counts start decreasing so we can plan the lunch breaks and other activities like meetings, trainings etc. for the agents such that the incoming calls can be handled by the available agents with ease.
- Also, after 7pm the frequency of call drops so we could plan for shift changes or EOD times so as to conclude the day.
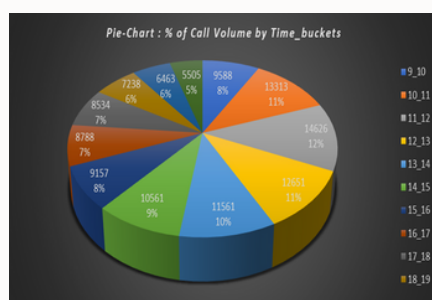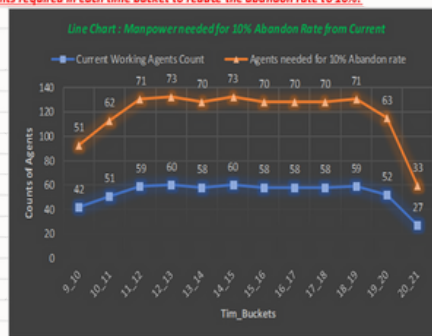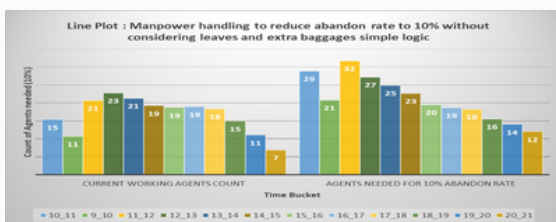
Task 3 – Manpower Planning

Question : What is the minimum number of agents required in each time bucket to reduce the abandon rate to 10%?
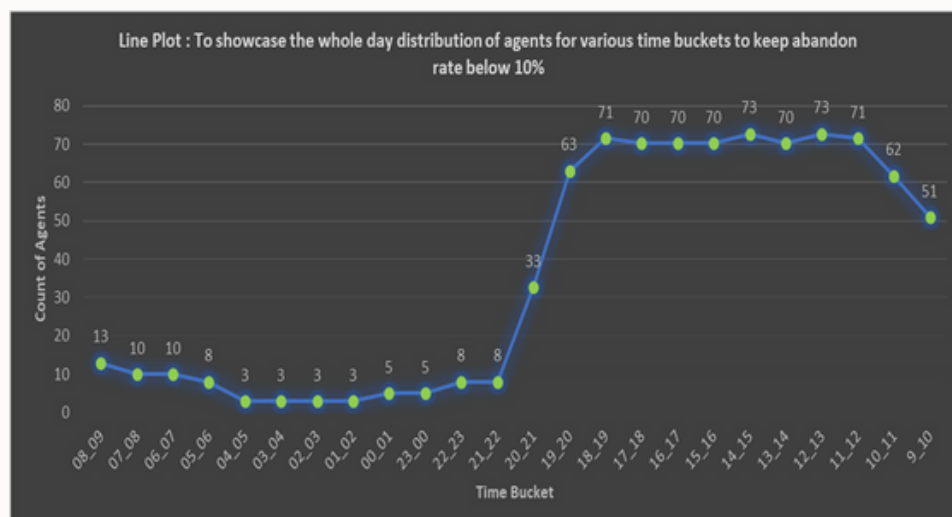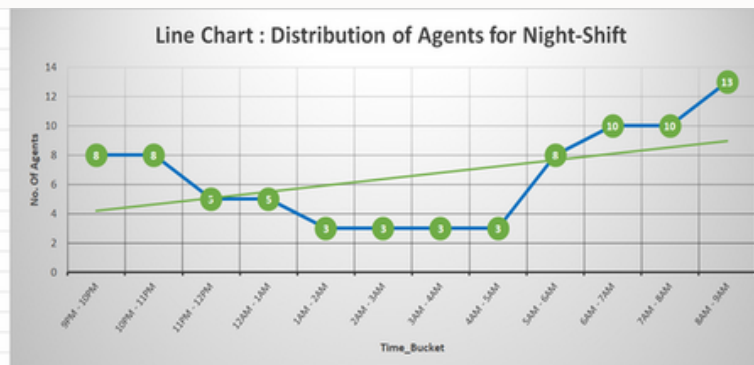

KEY FINDINGS

1. I have done 2 analysis 1 considering leaves taken by agent and compensating them over the whole data which brings the overall no. to 66 (distinct count of agent in data) & 80 avg needed agent from calculation (see distribution for deeper insights).
2. If we ignore the leaves and distinct count and instead use math's to get the avg call handling capacity of individual the current value comes to 17 agents for current data & 21 avg agents needed for 10% abandon rates (see charts for distributions.

Task 4 – Night Shift Manpower Planning

Question : Propose a manpower plan for each time bucket throughout the day, keeping the maximum abandon rate at 10%.

KEY FINDINGS

- In the given question 1st we calculated the night shift manpower or agents requirement by the given assumption that if 100 calls are made in a morning shift than for same shift at night 30 calls occur assuming this data we calculated the required numbers.
- For the next part of the question we were asked to plot the manpower of the whole day i.e. morning + night shift. We simply copied the data got from task3 & task4 in time_bucket and made few basic charts on the same.

PROJECT 8 - ABC CALL VOLUME TREND ANALYSIS

Project Summary: Inbound customer support, which is the focus of this project, involves handling incoming calls from existing or prospective customers. The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.

PROBLEM/PROJECT DESCRIPTION

1
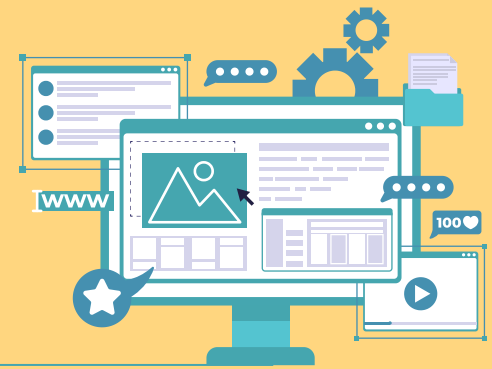
- The given dataset contains the information about the customer call data of an insurance company "ABC"
- The task is to analyze the customer data and formulate strategies to reduce the abandon rate of calls.
- The task is to do predictive analysis and device roosters and agent number to handle the customer requests.

APPROACH (RESEARCH)

2

1. Importing the data & glancing.
2. Removing Unknowns or blanks by treating using appropriate methods.
3. After final processing using excel or other tools to solve the tasks.
4. Using data visualization to show the results & making predictive analysis manipulation to suggest agents quantity to reduce abandon rate.

TECH-STACK USED

3

- *Google Sheets* - To get column stats & basic details of data like empty cells.
- *Excel* - Analysis and data visualization.
- *Word* - Report building for insights.
- Canva - Templates and formatting of report
- *Google Drive* - Sharing the files used.
- ShareX - Screenshots & Gif Recording.

INSIGHTS

4

- The dataset contains large info about the factors impacting the overall customer satisifaction regarding the service.
- The analysis of data & use of predictive techniques helped use understand a lot about the actual use of data manipulation to suggest forecast to the stakeholders.
- The project focuses on advanced excel or other visualization library like (matplotlib , seaborn in python) to showcase the output.

RESULTS

5

1. *The project helped me gain strong understanding of advanced excel functions , pivot charts, descriptive statistics and much more like the predictive analysis techniques.*
2. *I also explored the dataset using python and visualization library to compare the results with excel cleaning & reporting.*
3. The tasks helped to understand about raw customer churn dataset to predict the results of future forecast & handling scheudles and reducing call abandon rates using previous data.

📌 Full Report 🔗

SUMIT CHAURE

SOFTWARE ENGINEER

CONTACT

📞 +91-9167884727

@ sumitsc.work@gmail.com

🏠 Pune,🇮🇳 in 🐙

EDUCATION

2018
Mumbai University
Bachelors of Engineering
(Mechanical)   *7.21 CGPA*

2021
SM VITA, Mumbai
PG- DAC   *63 %*

SKILLS

SQL
Python
MS Excel
Power BI
Tableau
Story-telling/Reports
Communication

PROFILE

Software engineer with 2+ years of experience transitioning into the data domain, combining coding expertise in Python, SQL & cloud with a passion for data-driven insights.

EXPERIENCE

Data Analyst Intern
Aug 2023 - Now | Trainity

- Collected, cleaned, and analyzed data using Excel, Python and SQL.
- Created data visualizations to communicate insights to assigned tasks.
- Acquired proficiency in data tools & developed a strong understanding of data analysis techniques & terminologies.

Software Engineer
Apr 2021 - Aug 2023 |Cybage Software Pvt. Ltd.

- Software development experience with coding and problem-solving skills relevant to data roles.
- Applied analytical thinking and problem-solving for efficient data-driven backend job performance.
- Improved programming proficiency for transitioning to data-related roles.

RESOURCE LIST (DRIVE/GITHUB LINKS & POWERBI DASHBOARDS)

Google Drive Folder

The drive folder contains all the files used for the projects done and also the detailed reports(linked after each project)

GitHub Repo

The repo where the data for projects with the steps explained and alternatives if any to be explored, also case-study to study more about Data analytics.

PowerBI Dashboard

PowerBi Dashboard project for showcasing the interactive dashboards to the users and create a portfolio page.

Kaggle Profile

Portfolio page for following and learning from fellow Data enthusiast & learn from huge dataset & get into the domain to use large dataset to query & analyze for analysis & model making.

Thank You

Sumit Chaure

The End