



DATA ANALYSIS PORTFOLIO

PREPARED BY: -
GIRADHAR KUMAR MAURYA

Professional Background

I am having 7.5 years of experience as an IT Support Analyst with a background in IT operations, successfully handling IT operations for the entire region. I have worked as customer support engineer for 5 years in NCR Corporation PVT. LTD., currently I am working as IT support analyst in Reliance retail and handling the escalations and query and proving remote support to end users. I have also worked as Data Analyst intern with Unified Mentors. Currently I am data analyst trainee at Trainity and doing internship remotely.

I am enthusiastic Data Analyst Trainee with a proven track record in data analysis and visualization. I am eager to apply my data analysis skills in a dynamic environment. I am Proficient in MS Excel, SQL, Power BI and data visualization. I have also studied Tableau and Python. Experienced in optimizing data processes, and improving decision-making through actionable insights.

At Trainity I have completed the 8 projects from easy label to hard label in SQL and Excel, learnt the data analysis, regression analysis, descriptive analysis, Exploratory Data Analysis using advance excel, created the charts and graphs, pivot tables, used the windows function, subquery, joins and many more commands in SQL query, to find out the key insights from the project. These all the project showcase my expertise in Data analysis and visualization. These projects improved my knowledge and proficiency in MS Excel and SQL. As a data analyst I got the experience from the internship by working on the live projects, learnt to preset the analysis by storytelling in front of audience.

Table Of Contents

PROFESSIONAL BACKGROUND -----	1
TABLE OF CONTENTS -----	2
INSTAGRAM USER ANALYTICS DESCRIPTION -----	3
THE PROBLEM, DESIGN-----	3
FINDINGS -----	3-5
ANALYSIS, CONCLUSION-----	6
OPERATION & METRIC ANALYTICS DESCRIPTION-----	7
THE PROBLEM, DESIGN-----	7
FINDINGS -----	8-12
ANALYSIS, CONCLUSION-----	12
HIRING PROCESS ANALYTICS DESCRIPTION-----	13
THE PROBLEM, DESIGN-----	13
FINDINGS -----	14-17
ANALYSIS, CONCLUSION -----	18
IMDB MOVIE ANALYSIS DESCRIPTION-----	19
THE PROBLEM, DESIGN-----	19
FINDINGS -----	20-23
ANALYSIS, CONCLUSION-----	24
BANK LOAN CASE STUDY DESCRIPTION-----	25
THE PROBLEM, DESIGN-----	25
FINDINGS -----	26-32
ANALYSIS, CONCLUSION-----	33
IMPACT OF CAR FEATURES DESCRIPTION-----	34
THE PROBLEM, DESIGN-----	34
FINDINGS -----	35-44
ANALYSIS, CONCLUSION-----	45
ABC CALL VOLUME TREND DESCRIPTION-----	46
THE PROBLEM, DESIGN-----	46
FINDINGS -----	47-50
ANALYSIS, CONCLUSION-----	51
APPENDIX-----	52

INSTAGRAM USER ANALYTICS

Description: -

This project aims to leverage the amounts of data generated on the platform to gain insights into user behavior, content performance, and engagement patterns. By analyzing data from various Instagram activities such as likes, comments, posts, and user interactions by using the database management tools. In this project, I am using SQL and MySQL Workbench as my tool to analyze Instagram user data and answer questions posed by the management team. My insights will help the product manager and the rest of the team make informed decisions about the future direction of the Instagram app.

Problem: -

As a data analyst working with the product team at Instagram. My role involves analyzing user interactions and engagement with the Instagram app to provide valuable insights that can help the business grow.

Design: - The following steps taken to clean the data.

- Downloaded the dataset provided then imported and removed the duplicates.
- Arranged the dataset in the proper way and improved the headers of each column.
- Handled the missing values and done the corrections
- Tools used for the analysis: MYSQL Workbench 8.0

FINDINGS

1. Identify the five oldest users on Instagram from the provided database.

Answer: -Following are the top 5 most loyal users who is using the Instagram from long time.

	username	created_at
▶	Darby_Herzog	2016-05-06 00:14:21
	Emilio_Bernier52	2016-05-06 13:04:30
	Elenor88	2016-05-08 01:30:41
	Nicole71	2016-05-09 17:30:22
	Jordyn_Jacobson2	2016-05-14 07:56:26

2. Identify users who have never posted a single photo on Instagram.

Answer: - The following users are inactive and not posted any photo on Instagram.

username
Aniya_Hackett
Bartholome.Bernhard
Bethany20
Darby_Herzog
David.Osinski47
Duane60
Esmeralda.Mraz57
Esther.Zulauf61
Franco_Keebler64
Hulda.Macejkovic
Jaclyn81
Janelle.Nikolaus81
Jessyca_West
Julien_Schmidt
Kassandra_Homenick
Leslie67
Linnea59
Maxwell.Halvorson
Mckenna17
Mike.Auer39
Morgan.Kassulke
Nia_Haag
Ollie_Ledner37
Pearl7
Rocio33
Tierra.Trantow

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

Answer: - The winner of the contest is “Zack_Kemmer93” having maximum total 48 likes

	username	id	image_url	Total_likes
▶	Zack_Kemmer93	145	https://jarret.name	48

4. Identify and suggest the top five most commonly used hashtags on the platform.

Answer: - Top 5 commonly used hashtags are smile, beach, party, fun and concert.

tag_name	Total_tags
smile	59
beach	42
party	39
fun	38
concert	24

5. Determine the day of the week when most users register on Instagram. Provide insights on when to schedule an ad campaign.

Answer: - On Sunday and Thursday most user's register on Instagram so these are the best day to launch ads.

	day	total_register
▶	Thursday	16
	Sunday	16
	Friday	15
	Tuesday	14
	Monday	14
	Wednesday	13
	Saturday	12

6. User Engagement: Investors want to know if users are still active and posting on Instagram or if they are making fewer posts.

Answer: - The average number of posts by user on Instagram is more than two.

	Avg_post
▶	2.5700

7. Investors want to know if the platform is crowded with fake and dummy accounts.

Answer: - These 13 users have liked total 257 photos, i.e., total photos, as this is not possible for normal user to like every photo so these are the fake users.

	username	Total_Photo_likes
▶	Aniya_Hackett	257
	Jaclyn81	257
	Rocio33	257
	Maxwell.Halvorson	257
	Ollie_Ledner37	257
	Mckenna17	257
	Duane60	257
	Julien_Schmidt	257
	Mike_Auer39	257
	Nia_Haag	257
	Leslie67	257
	Janelle.Nikolaus81	257
	Bethany20	257

ANALYSIS

After all the SQL query I have find these valuable insights that can help the business grow this analysis involves tracking how users engage with a digital product, such as a software application or a mobile app. These insights derived from this analysis can be used by various teams within the business like the marketing team might use these insights to launch a new campaign, the product team might use them to decide on new features to build, and the development team might use them to improve the overall user experience. Some of these analyses are follows

- **Loyal User Reward:** The oldest five users are identified to receive loyalty rewards.
- **Inactive User Engagement:** Users who have never posted are targeted for engagement campaigns.
- **Contest Winner Declaration:** The user with the most likes on a single photo is declared the winner.
- **Hashtag Research:** The top five hashtags are identified for brand recommendations.
- **Ad Campaign Launch:** The best day to launch an ad campaign is determined based on user registration trends.

CONCLUSIONS

This project has significantly contributed to a deeper understanding of user behavior, content performance, and influencer dynamics on the platform. The insights derived from this analysis have empowered stakeholders to make data-driven decisions, optimize their content strategies, and enhance user engagement. Personally, the project has provided valuable learning experiences, advanced my analytical skills, Query performing skill through SQL and enriched my knowledge in the field of social media analytics. Overall, the project's impact has been profound, driving more effective and strategic use of Instagram as a marketing and engagement tool.

OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

Description: -

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, my work is to closely work with various teams, such as operations, support, and marketing, helping them derive valuable insights from the data they collect. This project is for focusing and analyzing the data of the users about their interest and time spending over the events, email events and job data, this data is provided by the company. My task is to derive the key insight and the answer the question sent by the different department, through which department can predict the overall growth of the company.

Problem: -

In this project, I shall take on the role of a Lead Data Analyst at a company like Microsoft. I shall be provided with various datasets and tables, and my task will be to derive insights from this data to answer questions posed by different departments within the company. My goal is to use my 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.

Design: - The following steps taken to clean the data.

- Downloaded the dataset provided then imported and removed the duplicates.
- Arranged the dataset in the proper way and improved the headers of each column.
- Handled the missing values and done the corrections
- Tools used for the analysis: MYSQL Workbench 8.0

FINDINGS

1.Jobs Reviewed Over Time: Your Task: Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.

Insights: - Total number of jobs reviewed per hour for each day in November 2020 is 126.

	jobs reviewed per hour for each day in November 2020
▶	126. 18048333

2.Throughput Analysis: Write an SQL query to calculate the 7-day rolling average of throughput. Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and why.

	dates	Daily Throughput
▶	11/25/2020	0.02
	11/26/2020	0.02
	11/27/2020	0.01
	11/28/2020	0.06
	11/29/2020	0.05
	11/30/2020	0.05

	weekly throughput
▶	0.03

Insights: - Metrics going up and down on a weekly and daily basis. We will get numbers quickly every day or minute if we want. As a result, rolling metrics are good at showing if metrics are trending up or down on a daily level.

3.Language Share Analysis: Your Task: Write an SQL query to calculate the percentage share of each language over the last 30 days.

	language	percentage	total
▶	English	12.50	8
	Arabic	12.50	8
	Persian	37.50	8
	Hindi	12.50	8
	French	12.50	8
	Italian	12.50	8

Insights: - Persian language is highest share with 37.5% total language.

4.Duplicate Rows Detection: Your Task: Write an SQL query to display duplicate rows from the job_data table.

	actor_id	duplicates
▶	1003	2

Insights: - Actor ID 1003 having the duplicate row.

5. Weekly User Engagement: Your Task: Write an SQL query to calculate the weekly user engagement.

	week_number	active_user
►	17	663
	18	1068
	19	1113
	20	1154
	21	1121
	22	1186
	23	1232
	24	1275
	25	1264
	26	1302
	27	1372
	28	1365
	29	1376
	30	1467
	31	1299
	32	1225
	33	1225
	34	1204
	35	104

Insight: -These are the activeness of users on a weekly basis. On the week 30th maximum number of users are active whereas on week 35 user activeness is very low

6. User Growth Analysis: Your Task: Write an SQL query to calculate the user growth for the product.

	Months	Users	Growth in %
►	1	1424	NULL
	2	1370	-3.79
	3	1530	11.68
	4	1814	18.56
	5	1986	9.48
	6	2172	9.37
	7	2562	17.96
	8	2694	5.15
	9	660	-75.50
	10	780	18.18
	11	798	2.31
	12	972	21.80

Insights: - Over the time for a Product the growth in user are increases.

7. Weekly Retention Analysis: Objective: Analyse the retention of users on a weekly basis after signing up for a product.

Insights: - weekly retention of users based on their sign-up cohort. Are below.

	Week Numbers	Week 0	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8	Week 9	Week 10	Week 11	Week 12	Week 13	Week 14	Week 15	Week 16	Week 17	Week 18
►	17	663	472	324	251	205	187	167	146	145	145	136	131	132	143	116	91	82	77	5
	18	596	362	261	203	168	147	144	127	113	122	106	118	127	110	97	85	67	4	0
	19	427	284	173	153	114	95	91	81	95	82	68	65	63	42	51	49	2	0	0
	20	358	223	165	121	91	72	63	67	63	65	67	41	40	33	40	0	0	0	0
	21	317	187	131	91	74	63	75	72	58	48	45	39	35	28	2	0	0	0	0
	22	326	224	150	107	87	73	63	60	55	48	41	39	31	1	0	0	0	0	0
	23	328	219	138	101	90	79	69	61	54	47	35	30	0	0	0	0	0	0	0
	24	339	205	143	102	81	63	65	61	38	39	29	0	0	0	0	0	0	0	0
	25	305	218	139	101	75	63	50	46	38	35	2	0	0	0	0	0	0	0	0
	26	288	181	114	83	73	55	47	43	29	0	0	0	0	0	0	0	0	0	0
	27	292	199	121	106	68	53	40	36	1	0	0	0	0	0	0	0	0	0	0
	28	274	194	114	69	46	30	28	3	0	0	0	0	0	0	0	0	0	0	0
	29	270	186	102	65	47	40	1	0	0	0	0	0	0	0	0	0	0	0	0
	30	294	202	121	78	53	3	0	0	0	0	0	0	0	0	0	0	0	0	0
	31	215	145	76	57	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	32	267	188	94	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	33	286	202	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	34	279	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	35	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

8. Weekly Engagement Per Device: Objective: Measure the activeness of users on a weekly basis per device.

Insights: - weekly engagement per device are below.

	Week Numbers	Dell Inspiron Notebook	iPhone 5	iPhone 4S	Windows Surface	Macbook Air	iPhone 5S	Macbook Pro	Kindle Fire	iPad Mini	Nexus 7	Nexus 5
►	17	46	65	21	10	54	42	143	6	19	18	40
	18	77	113	46	10	121	73	252	27	30	30	73
	19	83	115	44	16	112	79	266	21	36	41	87
	20	84	125	55	21	119	79	256	23	32	32	103
	21	80	137	45	17	110	74	247	30	23	29	91
	22	92	125	45	15	145	71	251	21	34	45	96
	23	103	152	53	14	124	79	266	25	33	36	88
	24	99	142	53	22	152	79	255	25	39	49	87
	25	105	137	40	22	121	78	275	24	30	51	89
	26	89	152	50	21	134	94	269	26	43	46	87
	27	89	163	67	33	142	83	302	25	35	40	84
	28	103	151	61	33	148	93	295	31	35	39	85
	29	113	144	60	28	148	90	295	37	34	45	77
	30	127	152	65	19	159	103	322	25	35	62	84
	31	113	135	56	19	147	71	321	14	27	38	69
	32	104	119	34	10	125	67	307	12	30	25	67
	33	110	110	35	15	133	65	312	14	28	30	70
	34	105	101	50	18	136	70	292	13	25	33	70
	35	9	2	6	3	10	3	17	3	2	2	4

9. Email Engagement Analysis: Objective: Analyse how users are engaging with the email service.

Week	Weekly Digest Rate	Email Open Rate	Email Clickthrough Rate	Reengagement Email Rate
17	62.32	21.28	11.39	5.01
18	63.45	22.24	10.49	3.83
19	62.16	22.67	11.13	4.04
20	61.62	22.64	11.43	4.31
21	63.52	22.82	9.97	4.09
22	63.59	21.56	10.66	4.19
23	62.39	22.34	11.18	4.09
24	61.61	22.92	10.99	4.48
25	63.77	21.79	10.54	3.90
26	62.99	22.22	10.61	4.18
27	62.24	22.49	11.37	3.90
28	62.92	22.48	10.77	3.83
29	63.98	21.71	10.51	3.79
30	62.29	23.24	10.59	3.88
31	65.27	23.25	7.66	3.82
32	66.59	22.85	7.14	3.42
33	64.73	23.10	7.91	4.26
34	64.33	23.91	7.67	4.08
35	0.00	32.28	29.92	37.80

Insight: -Email engagement analysis is done on the weekly digest rate, email open rate, email clickthrough rate and re-engagement email rate.

ANALYSIS

- Operational Efficiency Improvements:** Through investigating metric spikes, we may identify bottlenecks or inefficiencies that can be addressed to improve overall performance.
- Early Warning System:** Establishing robust monitoring and analysis frameworks can serve as an early warning system for potential issues, reducing downtime and operational disruptions.
- Data-Driven Decision Making:** By understanding the root causes of metric spikes, we can make more informed decisions, leading to better resource allocation and strategic planning.
- Enhanced Customer Experience:** Addressing issues that lead to negative spikes can help improve customer satisfaction and retention.
- Strategic Opportunities:** Positive spikes might indicate new opportunities for growth or optimization that can be explored further.

CONCLUSIONS

The Operation Analytics and Investigating Metric Spike project is essential for maintaining the smooth operation of any organization. It requires a strong understanding of data analytics, operational processes, and strategic thinking. By successfully identifying and addressing metric spikes, the organization can improve efficiency, prevent potential issues, and capitalize on emerging opportunities.

HIRING PROCESS ANALYTICS

Description: -

This Project is about the hiring process analysis within the company, The purpose of this project is to analyze the current hiring process, identify inefficiencies, and propose improvements to streamline the recruitment and onboarding of new employees. In this project I have analyzed the average salary of the employee as per department, checked the proportion of employee in every department as well as created the class intervals on the basis of salary and done the analysis.

Problem: -

In this project, my 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. My job is to analyze this data and answer certain questions that can help the company improve its hiring process.

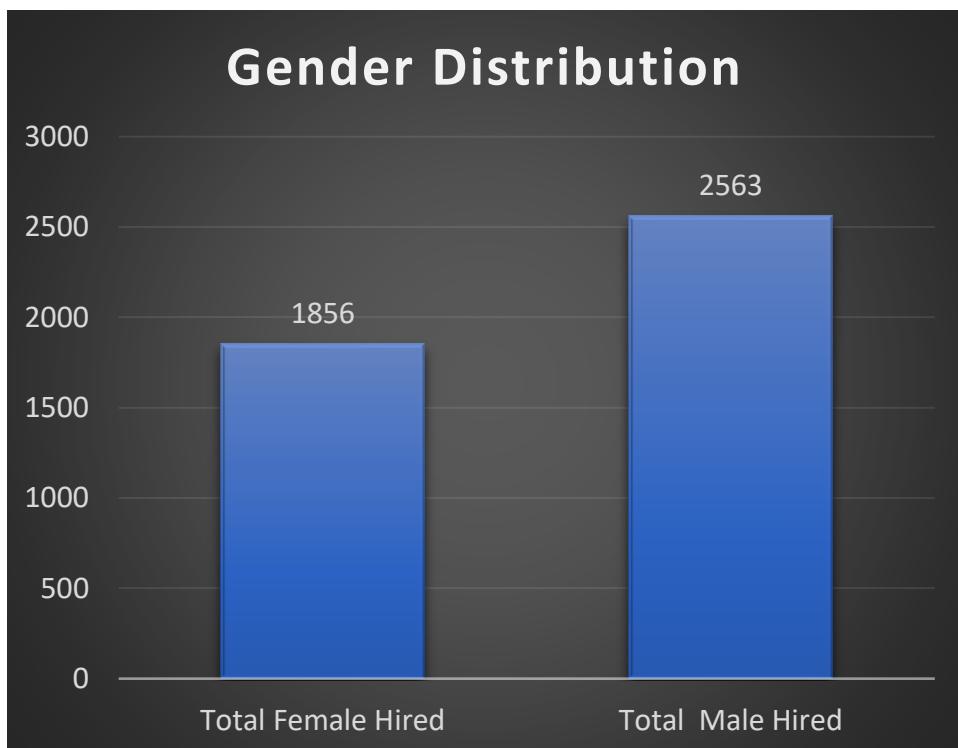
Design: - The following steps taken to clean the data.

- First Organize the data into a well-structured table, ensured each column has a clear and consistent header.
- Examine each column to understand the type of data it holds (e.g., dates, numerical values, categorical data).
- Utilize Excel's table functions to create a filterable table, allowing to easily sift through and analyze different segments of the data.
- Done the Exploratory Data Analysis (EDA) and perform the numerical calculations and visualize the data using charts and graphs.
- Prepare a presentation of key findings using Microsoft Excel's 2016 charting and graphing tools, and transfer to Microsoft PowerPoint 2016

FINDINGS

1.Determine the gender distribution of hires. How many males and females have been hired by the company?

Total Female Hired	1856
Total Male Hired	2563



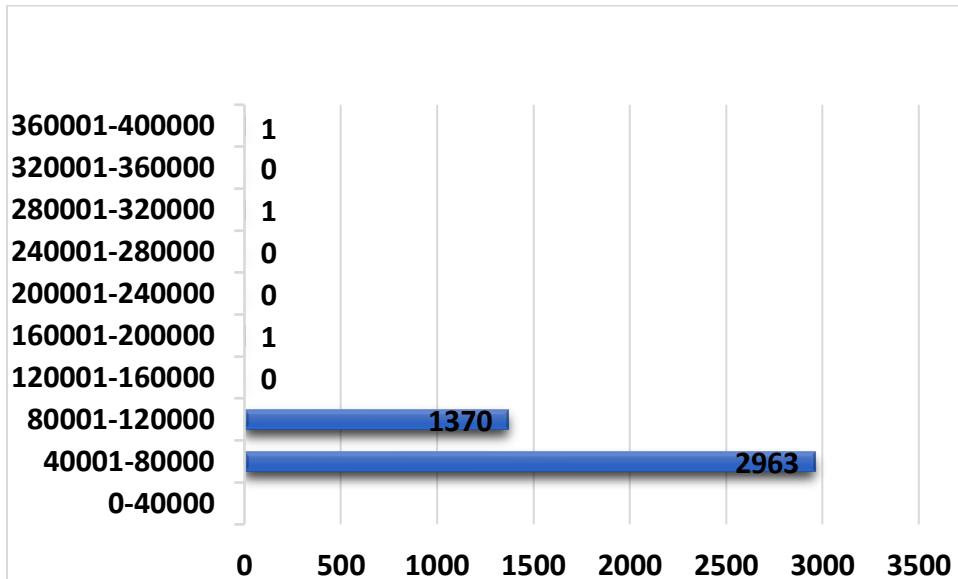
Insights: -Total number of females hired is 1856 and the total number of males hired is 2563

2.What is the average salary offered by this company? Use Excel functions to calculate this.

Average salary of Employee in company	49983.03
---------------------------------------	----------

Insights: - Total average salary provided by this company to employee is approx. Rs 49983

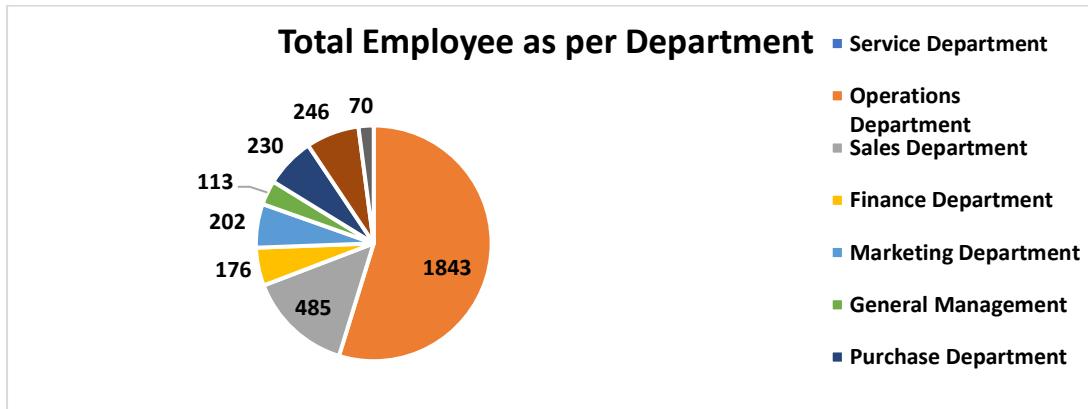
3.Create class intervals for the salaries in the company. This will help you understand the salary distribution.



Creating the class intervals(Range 0-40000)	Count
0-40000	
40001-80000	2963
80001-120000	1370
120001-160000	0
160001-200000	1
200001-240000	0
240001-280000	0
280001-320000	1
320001-360000	0
360001-400000	1
Total	4336

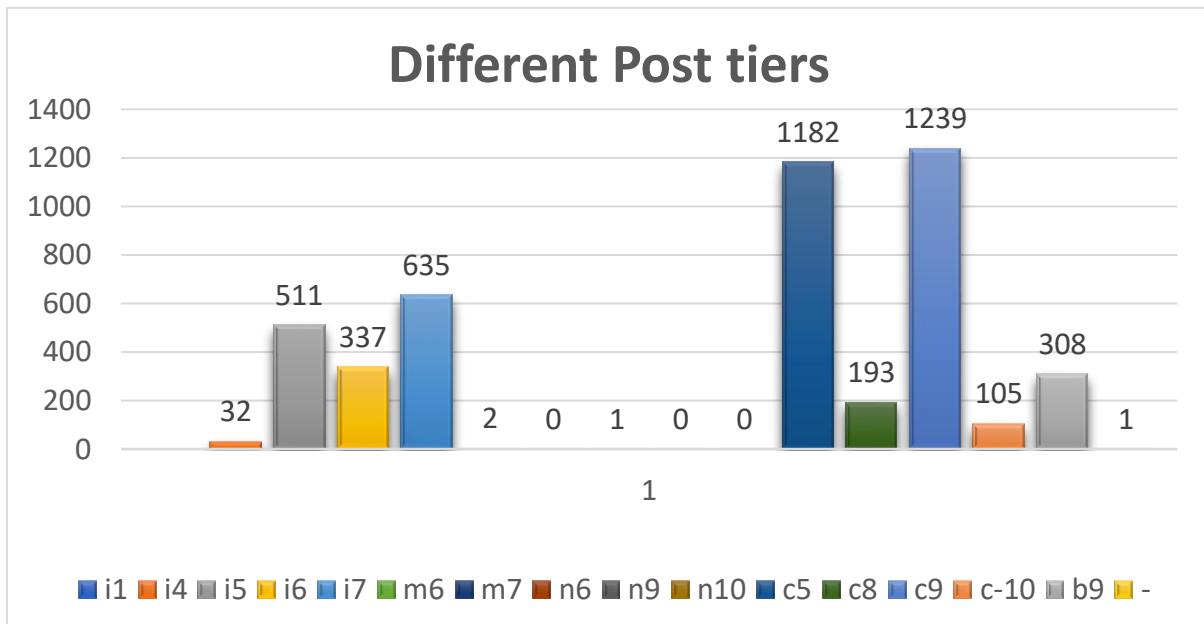
INSIGHTS: - Created the class intervals from range 0 to 40000 and then counted the total applicant lays in the salary range, maximum comes under 40k to 80k salary range, Total count is 7167.

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



Insights: - Maximum employee is working in Operations department and in human resource department there is less employee compare to all department.

5. Use a chart or graph to represent the different position tiers within the company. This will help you understand the distribution of positions across different tiers.



Post Name	Count
i1	
i4	32
i5	511
i6	337
i7	635
m6	2
m7	0
n6	1
n9	0
n10	0
c5	1182
c8	193
c9	1239
c-10	105
b9	308
-	1
Total	4546

Insights: - Used the countifs formula and calculated the total count of position across different tiers and visualized in bar chart.

ANALYSIS

- **Reduced Time-to-Hire:** By identifying and addressing bottlenecks, the organization can reduce the time it takes to hire, leading to faster onboarding and reduced costs.
- **Improved Candidate Quality:** Evaluating recruitment channels helps focus efforts on the most effective sources, improving the overall quality of hires.
- **Enhanced Candidate Experience:** Understanding candidate drop-off points and improving communication can lead to a better experience and higher offer acceptance rates.
- **Increased Diversity:** Regularly analyzing diversity metrics ensures that the organization is meeting its diversity and inclusion goals, fostering a more inclusive workplace.
- **Data-Driven Hiring Decisions:** Predictive analytics enable the HR team to make more informed decisions, reducing bias and increasing the likelihood of successful hires.

CONCLUSIONS

- Improved my ability to structure datasets in Excel, ensuring data integrity and ease of analysis.
- Conducted extensive EDA using Excel's in-built functions and visualization tools.
- Improving my skills in data visualization and using excel functions.
- This Hiring Process Analysis Project significantly deepened my understanding of the various facets of hiring process analytics.
- It provided practical experience in handling and analyzing real-world data, improving my technical skills in Excel and enhancing my ability to derive actionable insights from data.
- This project has equipped me with the knowledge and skills to contribute effectively to optimizing hiring processes in any organization.

IMDB MOVIE ANALYSIS

Description: -

This project aims to analyze and identify the key factors that influence the success of movies on IMDB. Success is primarily defined by high IMDB ratings. Understanding these factors can provide valuable insights for movie producers, directors, and investors, enabling them to make informed decisions about future projects. The analysis will involve exploring various attributes of movies, such as genre, director, budget, release date, runtime, and others, to determine their impact on IMDB ratings. In this project I have done the analysis on IMDB movies dataset. This dataset includes a variety of information about movies such as their ratings, genres, directors, budgets, runtimes, and more.

Problem: -

My goal is not just to answer questions but to provide insights that can drive decision-making and the analysis should aim to provide actionable insights that can help stakeholders make informed decisions.

Design: -

- ❖ Arranged the data and inserted into table then converted to normal range.
- ❖ Removed the column which is not required in the task given for analysis like (color, director_facebook_likes, actor_3_facebook_likes, actor_2_name, actor_1_facebook_likes, total_facebook_likes , actor_3_name, facenumber_in_poster ,plot_keywords, movie_imdb_link num_user_for_reviews, content_rating, title_year actor_2_facebook_likes ,aspect_ratio, movie_facebook_likes)
- ❖ Removed the blanks by selecting all and then go to special and selected the blanks then delete the blanks from data set.
- ❖ Removed the duplicate data from the dataset (Total 45 duplicate found and deleted)
- ❖ After all the process I have got the clean dataset of 3849 Rows and 14 columns.
- ❖ Ensure the data validation, handled the missing data, changed the non-printable character in proper format, (removing of letter Ä, ¥, © etc).

FINDINGS

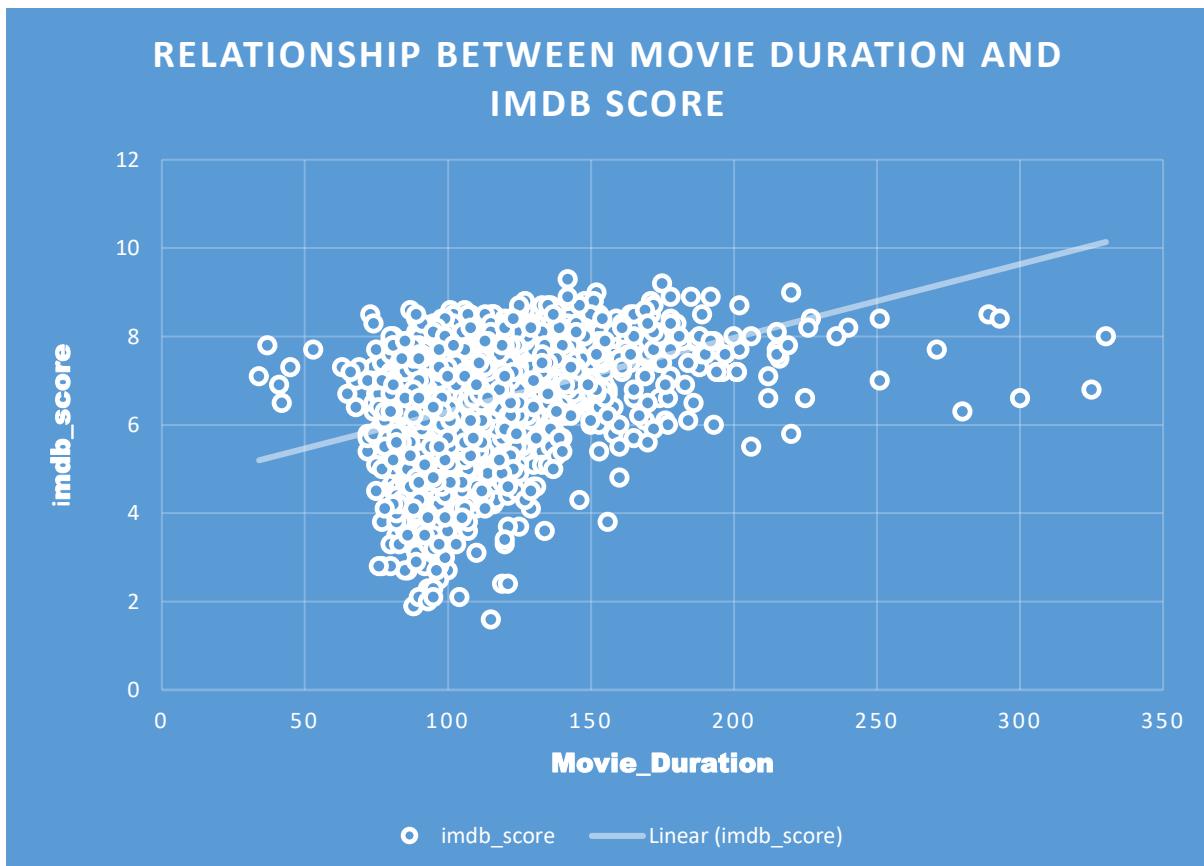
A. Movie Genre Analysis: Analyze the distribution of movie genres and their impact on the IMDB score.

Row Labels	Count of movie_title	Average of imdb_score	Max of imdb_score	Min of imdb_score	Var of imdb_score	StdDev of imdb_score	Median of imdb_score	Mode of imdb_score
Comedy	1029	6.165500486	8.8	1.9	1.074012906	1.036345939	6.3	6.4
Action	962	6.290748441	9	2.1	1.06375407	1.03138454	6.3	6.1
Drama	691	6.81447178	8.8	2.1	0.823732293	0.907596988	6.9	6.7
Adventure	375	6.5632	8.6	2.3	1.243508449	1.1151271	6.7	7.3
Crime	257	6.940856031	9.3	3.3	0.749691452	0.865847245	7	7.4
Biography	207	7.153623188	8.9	4.5	0.487547488	0.698246009	7.2	7
Horror	160	5.844375	8.5	2.3	1.053930425	1.026611136	5.9	5.9
Animation	46	6.763043478	8	4.5	0.945937198	0.972593028	7	7.1
Documentary	40	6.9175	8.5	1.6	2.140967949	1.463204684	7.3	7.5
Fantasy	37	6.281081081	7.9	4.3	0.799354354	0.894066191	6.5	6.8
Mystery	23	6.652173913	8.5	3.3	1.193517787	1.092482396	6.7	7.1
Sci-Fi	8	6.5875	8.2	5	1.064107143	1.031555691	6.35	0
Thriller	3	5.3	6.3	4.8	0.75	0.866025404	4.8	4.8
Western	3	6.766666667	8.9	4.1	5.973333333	2.444040371	7.3	0
Family	3	6.5	7.9	5.7	1.48	1.216552506	5.9	0
Musical	2	6.75	7.2	6.3	0.405	0.636396103	6.75	0
Romance	2	6.65	7.1	6.2	0.405	0.636396103	6.65	0

INSIGHTS:-

- The most popular genre is comedy, and total comedy movie count is 1029.
- Maximum average imdb score is 7.15 for genre biography and lowest liked genre is Thriller having avg imdb rating 5.3.
- Western genre are the more heterogeneous due to high variance 5.9 ,also it have the high stdDev of 2.4 where as Music and Romance genre are more homogeneous with low variance 0.4 and low stdDev 0.6.
- As per data highest average imdb score is 7.15 for the Biography genre followed by crime having avg imdb is 6.94 and avg imdb score is 6.9,with var 0.5, StdDev 0.7,median is 7.2 and Mode is 7.
- The lowest average imdb score belong to 5.3 for Thriller genre,with variance 0.8, stdDev 0.9, Median 4.8 and Mode is 4.8.

B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.



INSIGHTS: -

- The Mean for the Movie duration is 109.9
- Median for the movie duration is 106 and,
- Standard deviation for the movie duration is 22.75.
- The maximum imdb score is 9.3 having movie duration of 142 minutes, and the lowest imdb score is 1.6 having movie duration 115 minutes.
- The trend line in the scatter plot has slightly upward and it indicate that it has moderate positive relationship between duration and imdb score.
- Mean and median are close to each other and standard deviation is not too large, it means duration of movie data is normally distributed.

C. Language Analysis: Situation: Examine the distribution of movies based on their language

language	Total Movie Count >=10
English	3668
French	37
Spanish	26
Mandarin	14
German	13
Japanese	12
Hindi	10

INSIGHTS: -

- There is high difference between English movie count and other language movie count, so considered the movie having count ≥ 10 .
- As from charts English is most common language for the Movies.
- German and Japanese language movie having the highest mean 7.69, 7.63 and highest median 7.7, 7.8.
- Hindi movie language having the highest standard deviation 1.11 and mean is 6.76 and median 7.05.

D. Director Analysis: Influence of directors on movie ratings.

TOP DIRECTORS	PERCENTILE >=7.1
Tony Kaye	8.60
Charles Chaplin	8.60
Alfred Hitchcock	8.50
Ron Fricke	8.50
Damien Chazelle	8.50
Majid Majidi	8.50
Sergio Leone	8.43
Christopher Nolan	8.43
S.S. Rajamouli	8.40
Richard Marquand	8.40
Asghar Farhadi	8.40
Marius A. Markevicius	8.40
Lee Unkrich	8.30
Fritz Lang	8.30
Lenny Abrahamson	8.30
Billy Wilder	8.30
Pete Docter	8.23
Hayao Miyazaki	8.23

INSIGHTS: -

- First, I calculated the 90th percentile for each director by percentile formula and got the result 7.5, then counted the total directors having 90th percentile is greater or equal to 7.5
- Total count is 187 unique directors listed whose 90th percentile is greater or equal to 7.5 and these directors can be considered as top directors.
- Charles Chaplin having the highest average imdb score of 8.6 and 90th percentile is 8.5, followed by Majid Majidi having 8.5.
- S.S Rajamouli is the Indian director having 90th percentile is 8.4 and he is in the list of Top 10 director.

E. Budget Analysis: Explore the relationship between movie budgets and their financial success.

movie_title1	gross	budget	Profit Margin
Avatar	760505847	237000000	523505847
Jurassic World	652177271	150000000	502177271
Titanic	658672302	200000000	523505847
Star Wars: Episode IV - A New Hope	460935665	11000000	449935665
E.T. the Extra-Terrestrial	434949459	10500000	424449459

INSIGHTS: -

- Avtar Movie having the highest profit margins followed by Jurassic world.
- Correlation Coefficient calculated is 0.1008502
- Maximum profit margin is 523505847 of “AVATAR” Movie.
- Correlation is far from value 1, it means this is the weak relation of gross earning with budget.

ANALYSIS

- **Genres:** Certain genres such as Drama, Thriller, and Biography tends to receive higher ratings compared to genres like Comedy and Horror.
- **Director and Cast Popularity:** Movies directed by well-known directors and featuring popular actors generally receive higher ratings.
- **Budget:** There is a moderate positive correlation between a movie's budget and its IMDB rating, indicating that higher-budget movies tend to have better ratings, likely due to better production quality and marketing efforts.
- **Runtime:** Movies with a runtime between 90 to 120 minutes tend to perform better in terms of ratings.
- **Release Date:** Movies released in the holiday season or during major film festivals often receive higher ratings due to increased visibility and marketing.
- **Awards and Nominations:** Movies that have received awards or nominations tend to have higher IMDB ratings, reflecting the impact of critical acclaim on audience perception.

CONCLUSIONS

This project successfully identified and quantified the key factors that influence the success of movies on IMDB. By leveraging data analysis techniques, this project provides valuable insights and practical recommendations for movie producers, directors, and investors, helping them make informed decisions to enhance the chances of their movies achieving high ratings on IMDB. The findings from this analysis have the potential to guide future movie production and marketing strategies, leading to more successful movie projects in the industry.

BANK LOAN CASE STUDY

Description: -

This project aims to provide the key insights to bank about the loan applicant candidate, by finding the missing data, outliers, using the statics like average, median, finding the correlations, data imbalance etc. The primary objective of this project is to analyze and predict the likelihood of loan approval for applicants based on various factors. This helps the bank streamline the loan approval process, reduce defaults, and improve customer satisfaction.

Problem: - When a customer applies for a loan, there are four possible outcomes:

1. Approved: The company has approved the loan application.
2. Cancelled: The customer cancelled the application during the approval process.
3. Refused: The company rejected the loan.
4. Unused Offer: The loan was approved but the customer did not use it.

My goal in this project is to use EDA to understand how customer attributes and loan attributes influence the likelihood of default.

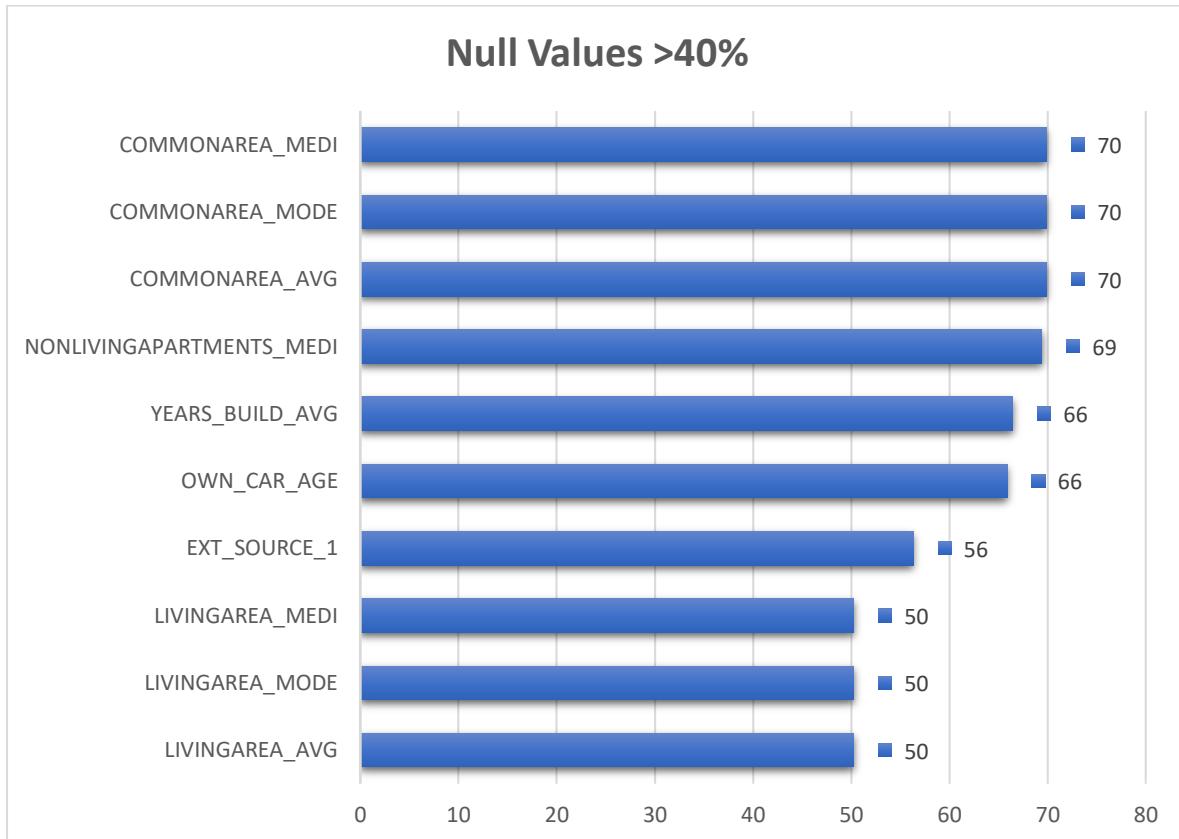
Design: - The following steps taken to clean the data.

- Downloaded the data from the sources.
- Identify and understand the factors affecting loan approval, predicting default risk.
- Address missing values, remove duplicates, and correct inconsistencies.
- Summarize the data to understand its basic characteristics.
- Done the Exploratory Data Analysis and created the charts and graph for proper understanding.
- Compile a comprehensive report detailing methodology, findings, and recommendations.

FINDINGS

A. Identify Missing Data and Deal with it Appropriately:

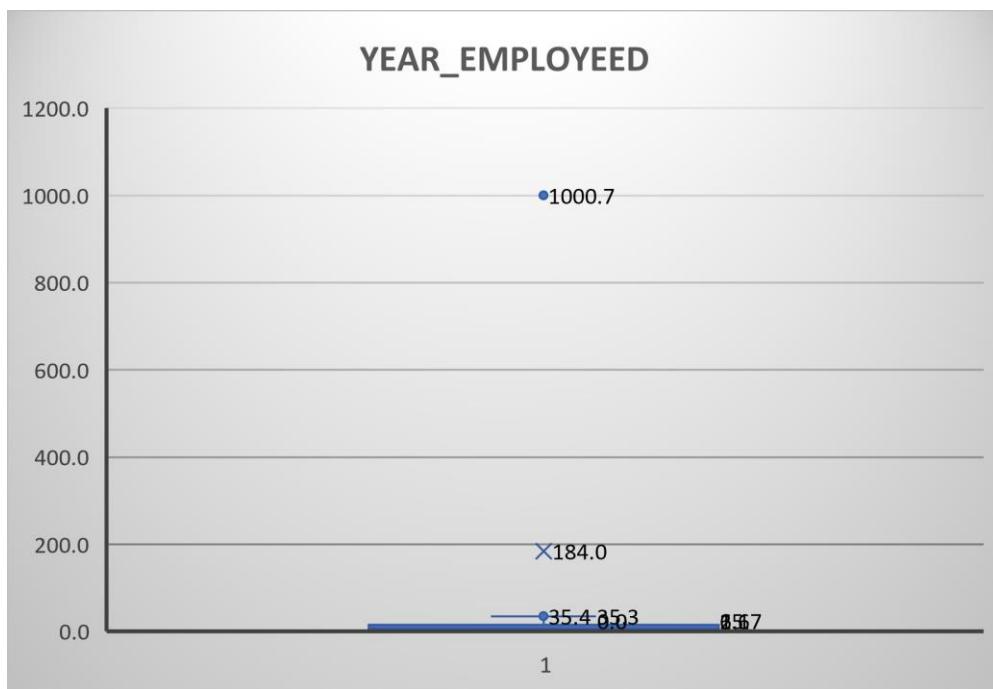
- **Task:** Identify the missing data in the dataset and decide on an appropriate method to deal with it using Excel built-in functions and features.



INSIGHTS: -

- “AMT_INCOME_TOTAL, AMT_CREDIT”, “AGE”, “YEAR_EMPLOYED”, “YEAR_REGISTRATION”, “YEAR_ID_PUBLISHED” Column have no missing data.
- “AMT_REQ_CREDIT_BUREAU_HOUR”, “AMT_REQ_CREDIT_BUREAU_DAY”, “AMT_REQ_CREDIT_BUREAU_YEAR” have 13% of missing data.
- “OCCUPATION_TYPE” column having the 31% of missing value.
- “COMMONAREA_AVG” field having the 70% of missing value so deleted the column.
- Column which is having the missing value greater than 40 % have been deleted for the proper analysis.
- For the imputation I have calculated the mean, median of each numerical column.

B. Identify Outliers in the Dataset: Outliers can significantly impact the analysis and distort the results. You need to identify outliers in the loan application dataset.



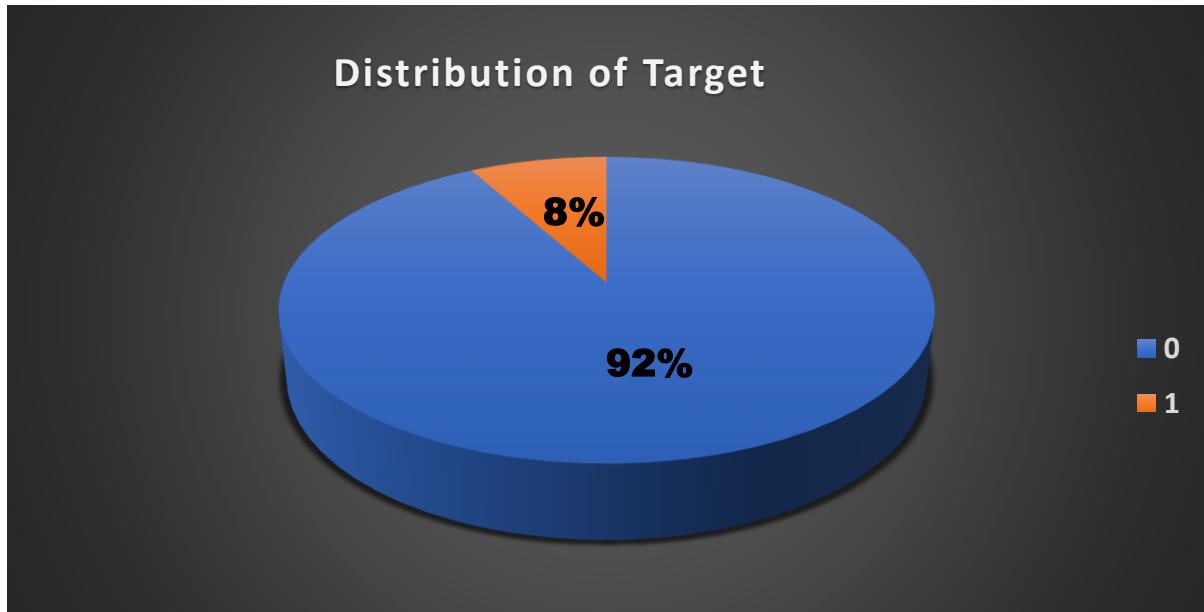
INSIGHTS: -

- The Box plot chart represent the variable name “YEAR_EMPLOYED, AND “CNT_CHILDREN”.
- The blue circle represents the employee working from in first chart, and in second chart it represents the count of children.
- Y axis indicated the value from 1 to 1007 year in fist box plot chart, and it is showing that employee who is working from 1000 of years, which is not possible.
- Same in the second box plot chart an applicant having the 11 child which is more in current time scenario.
- The X axis has one category at point 1, which is indicating the entire plot.
- Points in the plot indicating the outliers in the dataset, like employee working from 1000 year and count of children is 11.
- Outliers impact the statical analysis, so it's very important to consider them during exploratory data analysis.

C. Analyze Data Imbalance: Data imbalance can affect the accuracy of the analysis, especially for binary classification problems. Understanding the data distribution is crucial for building reliable models.

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

Row Labels	Count of TARGET	Row Labels	Count of TARGET
Cash loans	45276	0	45973
Revolving loans	4723	1	4026
Grand Total	49999	Grand Total	49999



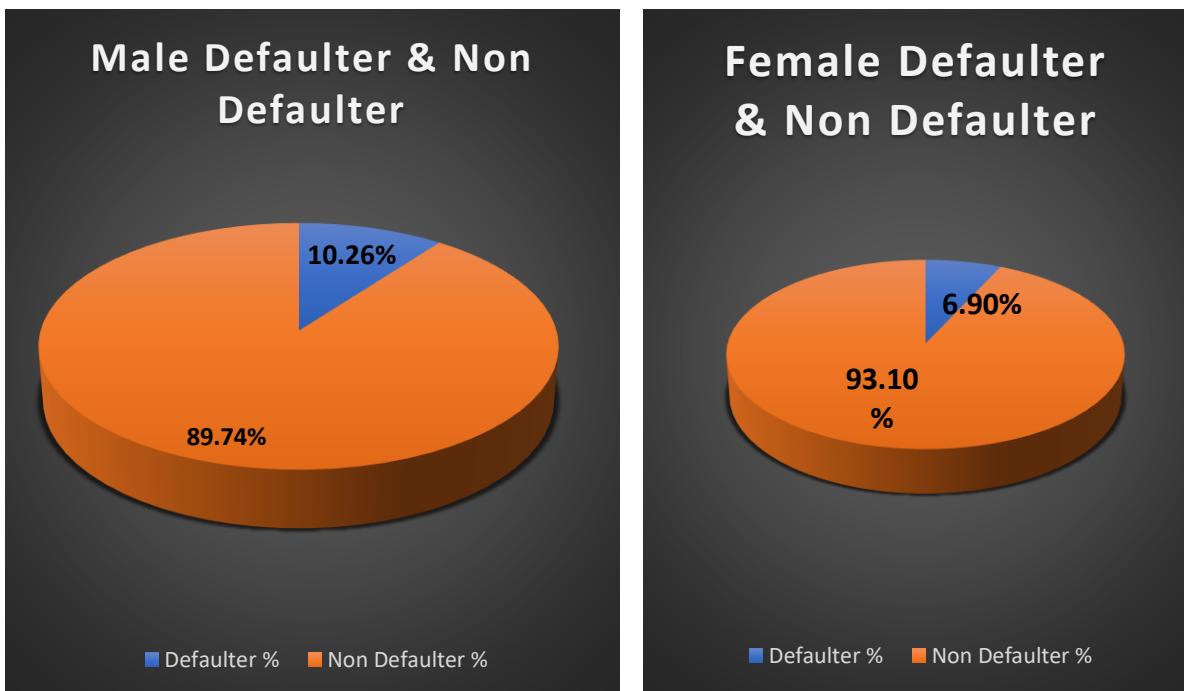
INSIGHTS: -

- The target column in the dataset indicates the defaulter and non-defaulter applicants.
- Total 45276 applicants applied for the cash loan which is 91 % of total loan types.
- Only 4723 applicants applied for the revolving loans which is 9 % of total applicants.
- As per the chart 92% of applicants are non-Defaulter and paying the loans on time.
- Whereas 8 % are the Defaulter and having difficulties in paying loan amount which is total 4026 of total applicants.
- Ratio of data imbalance is calculated 11.42 between the target 0 and target 1.

D. Perform Univariate, Segmented Univariate, and Bivariate Analysis: To gain insights into the driving factors of loan default, it is important to conduct various analyses on consumer and loan attributes.

Task: 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.

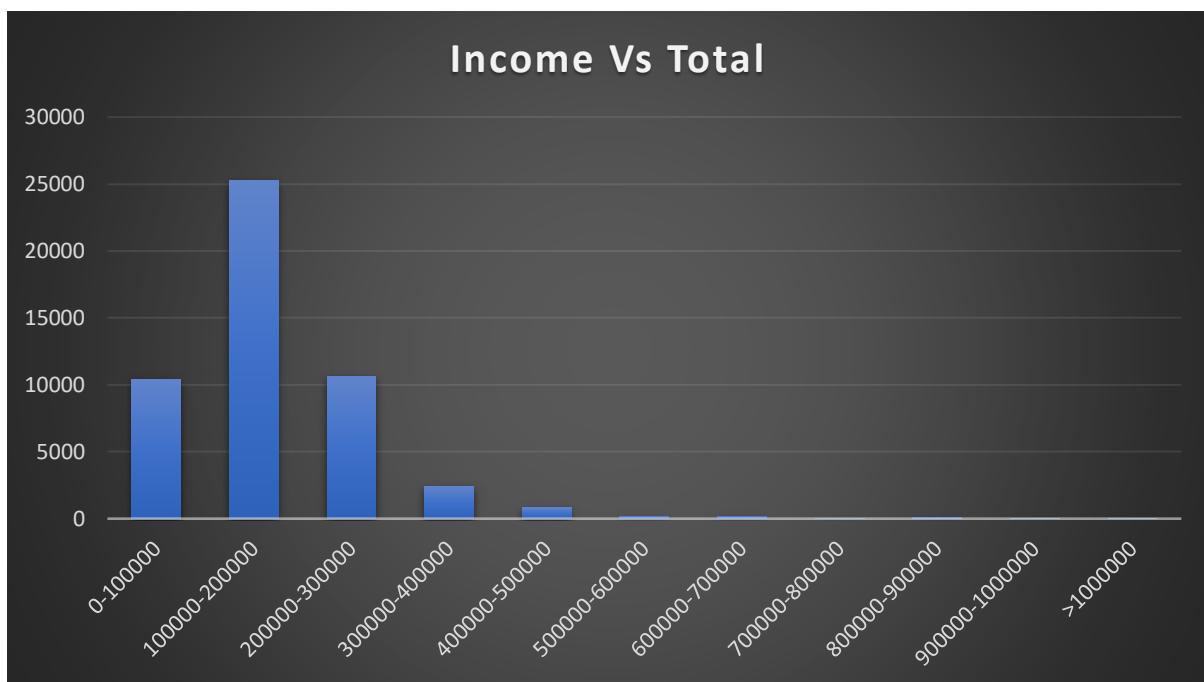
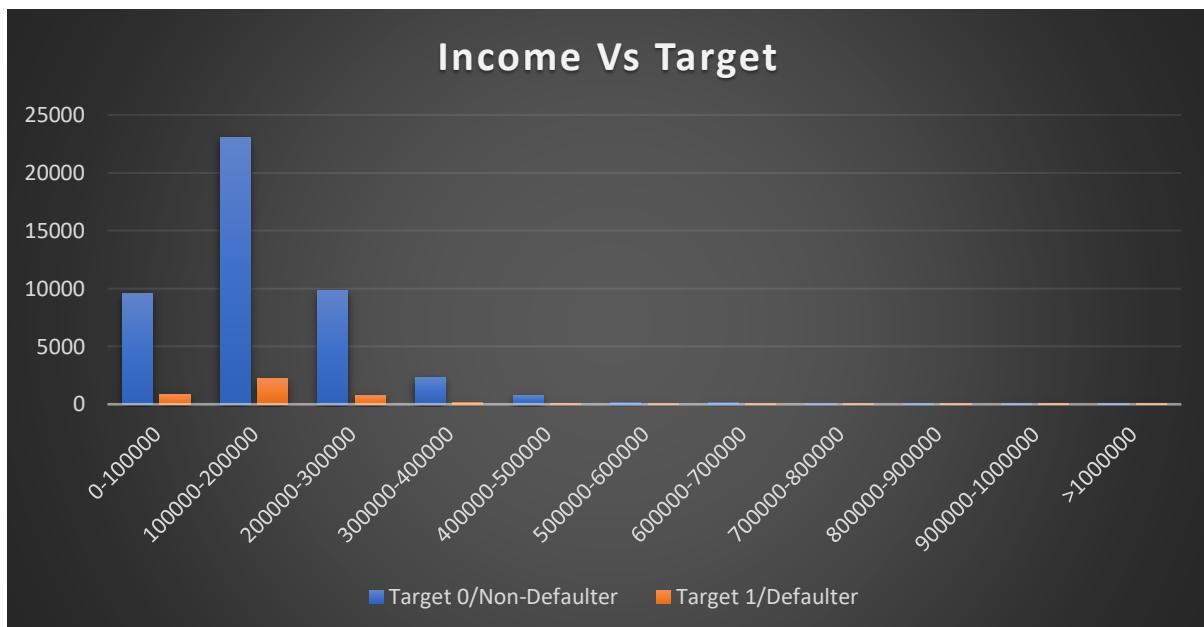
Univariate Analysis: - Analysis on the basis of Gender



INSIGHTS: -

- ❖ This univariate analysis done on the basis of gender, on the basis of family, on the basis of children count in family and family count.
- ❖ Total 10.26 % of the male are defaulter whereas rest 89.74 are non-defaulter.
- ❖ In Female 93.10 % belongs to non-defaulter and 6.90% are in defaulter category.
- ❖ When compare the defaulting percent Female are more i.e., 56 % and male are at 44%, it shows that Male are more active to pay the loan amount and having less difficulties than female.
- ❖ When we see on the basis of family status, Married people have taken the more loans comparing to the total family status.
- ❖ Married people are the most i.e., 59% is defaulting the loans and having issue is paying loans on time.

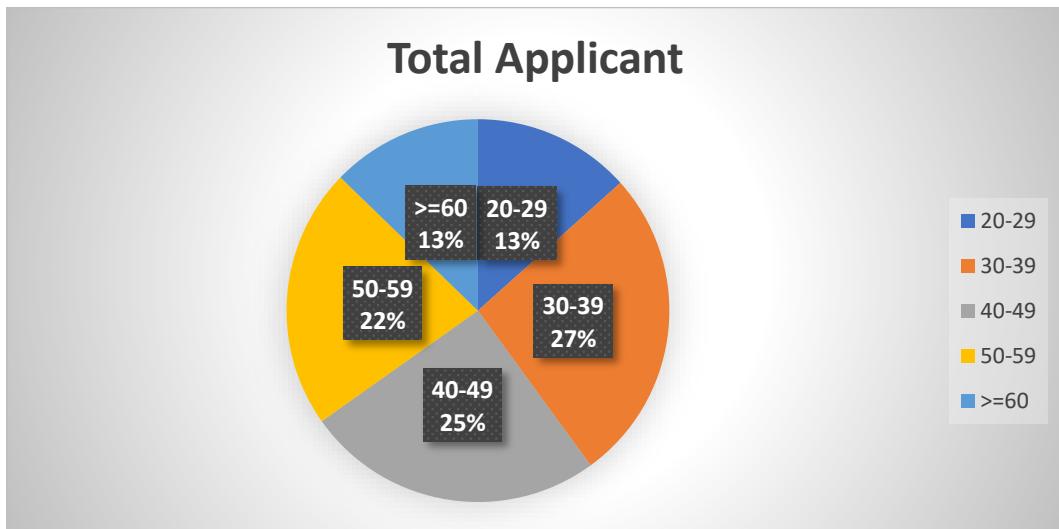
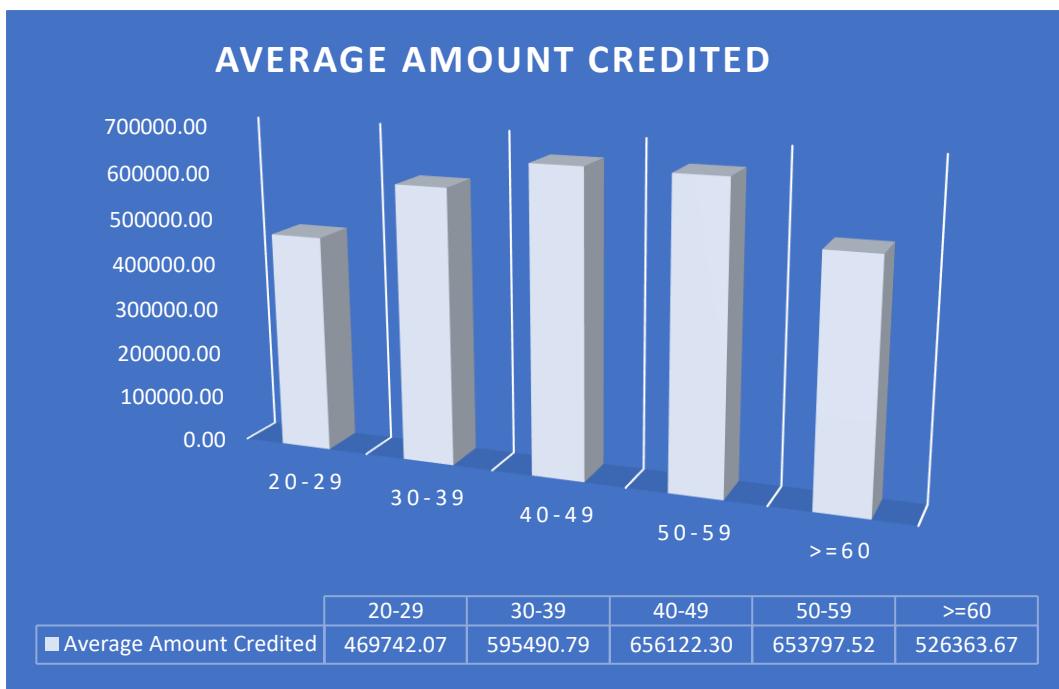
Segmented Univariate Analysis: - On the basis of income range



INSIGHTS: -

- ❖ The segmented univariate analysis is done on the basis of income range, amount of loan credited range and amount of annuity range.
- ❖ In the analysis, income range of 1 lac to 2 lac people are mostly applied for the loans and maximum, total 2188 of them having difficulties in payment and are defaulters.
- ❖ 2 lacs to 3 lacs of average amounts are credited to maximum 8849 applicant followed by the 8146-applicant taken the average amount more than 10 lac.

Bivariate Analysis: -On the basis of age group and loan amount taken.



INSIGHTS: -

- ❖ Bivariate Analysis DONE On the basis of age group and loan amount taken.
- ❖ Maximum 12042 applicants belong to age group 30 -39 year
- ❖ Maximum average amount 656122.30 credited to the age group of 40 to 49 years.
- ❖ Age group of more than 60 year are total 4026 defaulter, which is maximum.
- ❖ Maximum loan distributed to maternity leave that is average 640080.
- ❖ Maximum defaulter belongs to the working category that is 2132.

E. Identify Top Correlations for Different Scenarios: Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.

Task: 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.

NON-DEFALUTER CORRELATION/TARGET 0															DEFALUTER CORRELATION/TARGET 1														
COLUMN	CNT_CHIL	AMT_I	AMT_T_C	AMT_AN	AMT_GOOD	DAY_S_BI	year_employee	year_registration	year_d_pub	CNT_F_AM_M	CNT_F_EMBER	COLUMN	CNT_CHIL	AMT_I	AMT_T_C	AMT_AN	AMT_GOOD	DAY_S_BI	year_employee	year_registration	year_d_pub	CNT_F_AM_M	CNT_F_EMBER						
NAME	N	DRE	NCOME	REDI	NUIT	S_PRIC	RTH	d	published	S		NAME	N	DRE	NCOME	REDI	NUIT	S_PRIC	RTH	d	published	S							
CNT_C_HILDREN	1	0.03631	5705	3821	0.0155	5876	52151	0.1830	0.0325	0.87923	936	CNT_C_HILDREN	1	0.01011	7601	1729	0.00111	9673	77322	0.1521	0.0423	0.89252	1875						
AMT_I_NCOME	22	0.036	1	0.37	0.451	0.07	0.161	-	-	-	-	AMT_I_NCOME	77	0.010	1101	5271	0.0045	0.01326	9033	75868	0.0095	0.0091	0.01312	1678					
AMT_GOODS_PRICE	25	0.005	5752	1	65	4066	182	3	53758	90189	7635	AMT_GOODS_PRICE	1444	0.007	6019	1444	0.01527	6652	0.98243	2506	78222	0.0428	0.0437	0.06124	869				
AMT_ANNUITY	7	0.026	5696	0.77	0.770	0.05	0.074	-	-	-	-	AMT_ANNUITY	444	0.029	3821	4594	0.01800	9665	0.74	0.0428	0.0437	0.06124	8463						
AMT_DAYS_BIRTH	69	0.001	7244	1418	0.98	0.776	0.04	0.072	-	-	-	AMT_DAYS_BIRTH	6279	0.001	5500	5092	0.38467	2432	0.749	8751	11389	0.0215	0.0213	0.07583	1397				
year_emplotted	12	0.245	9098	443	43	5216	675	1	46476	16224	7657	year_emplotted	8681	0.189	5215	7375	0.16168	4733	0.00903	2506	7517	0.14099	24282	0.2884	0.2478	0.19914	2962		
year_registered	78	0.183	8053	6090	0.0034	0.294	0.07250	0.3474	0.2088	0.2745	0.23476	year_registered	1152	0.152	7244	1152	0.06893	8053	0.00956	2844	5816	0.04337	8437	0.192	0.1924	0.2326	0.18336		
year_id_published	21	0.032	6356	189	96	4005	313	4	48902	1	8177	year_id_published	206	0.042	5372	313	0.03228	8290	0.00912	3771	3210	0.04978	7896	66191	0.0902	0.04403	7815		
CNT_FA_M_MEMBERS	6	0.879	3404	635	05	7956	945	7	85094	58177	1	CNT_FA_M_MEMBERS	1678	0.06	5218	75	0.04161	4877	0.01312	1248	8384	0.05510	9141	36296	0.1517	0.0440	1		

INSIGHTS: -

Here I have calculated the top 10 correlation among the different variables, it shows the different correlation, strong correlation close to 1 or -1 indicate the linear relationship, whereas correlation close to 0 indicate the weak correlation and no linear relationship.

ANALYSIS: -

This project on bank loan case study aimed to analyze the factors influencing loan approval decisions, identify the characteristics of approved versus rejected applications, and build a predictive model to improve the accuracy of future loan approval decisions. Using data analysis techniques, I have explored the demographic and financial variables to understand their impact on loan outcomes. By understanding these factors, banks can enhance their risk assessment processes, offer more tailored financial products, and improve overall customer satisfaction. This predictive model developed in this study serves as a robust tool for future loan approval decisions, ensuring a balance between risk management and customer inclusivity. Implementing the recommendations can lead to more efficient and fair loan approval processes, and ultimately contributing to the bank's growth and customer trust.

CONCLUSIONS: -

Through this project I have learnt the lots of statics, analysis, and EDA, it enhanced my knowledge in advance excel as well as bank loan process and risk and challenge created after loan distribution. This project really vast my knowledge in excel by creating different types of charts, graphs, by calculating mean, median, correlation etc.

ANALYZING THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

Description: -

The automotive industry is highly competitive and continuously evolving, driven by consumer preferences, technological advancements, and regulatory changes. Understanding the impact of various car features on price and profitability is crucial for manufacturers, marketers, and stakeholders. Key features such as horsepower, fuel efficiency (MPG), car brand, and additional attributes like safety ratings, technological integrations, and design elements can significantly influence a car's market value and profitability.

Problem: - Analyzing the impact of car features on price and profitability is essential for automotive manufacturers and dealers to make data-driven decisions. Understanding which features drive higher prices and profitability can help in product development, marketing strategies, and sales tactics. This analysis will help in identifying which features contribute most to higher prices and profitability and how manufacturers can optimize their offerings.

Design: - The following steps taken to clean the data.

Analytical Methods: -

This project is for the analysis of automobile market and for this I have used the Descriptive Analysis, Exploratory Data Analysis, Predictive Analysis, regression analysis, coefficient correlation, data visualization, created the different types of charts like, scatter plot chart, stacked column charts etc.

Explanation: -

Analytical methods are the essential for various applications in business, healthcare, finance, engineering, and many other fields, providing a robust toolkit for data analysis and decision-making. Analytical methods encompass a wide range of techniques used to analyze data, understand patterns, and derive insights.

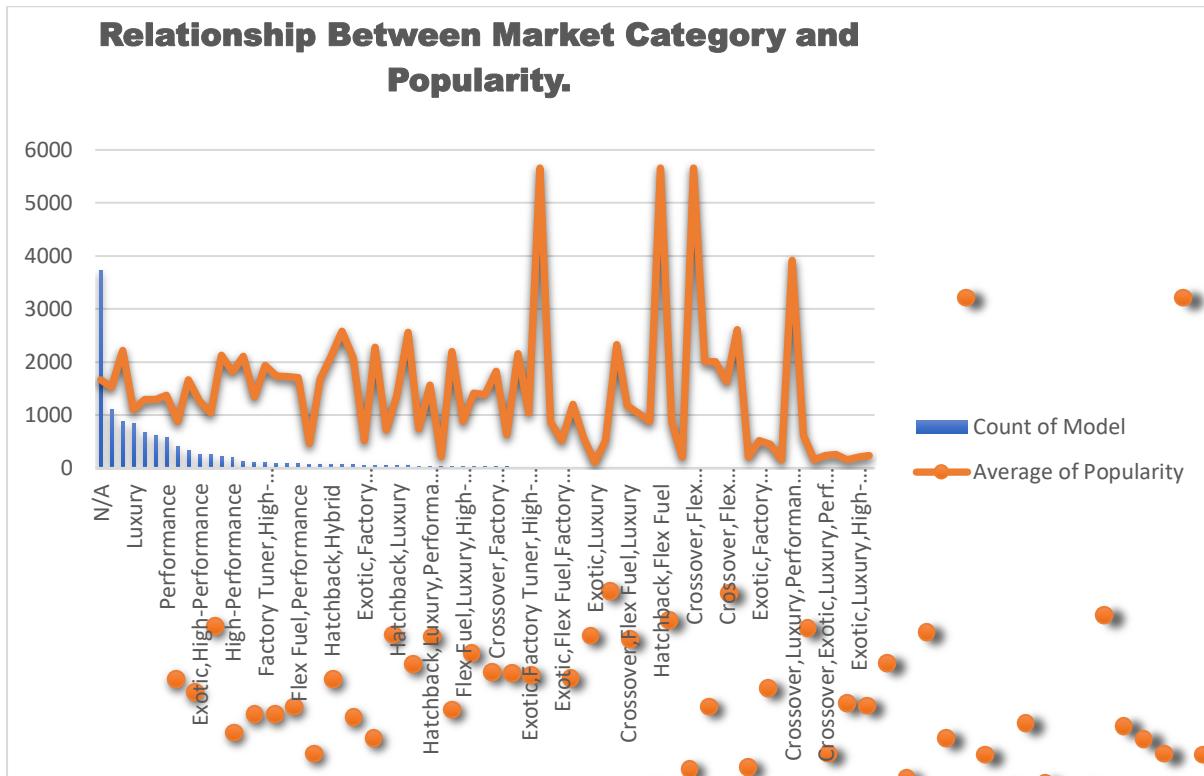
Description: -

I have used the Descriptive Analysis to find out the mean for the numerical data given in data set, used the Predictive Analysis and find out the linear regression analysis, Diagnostic Analysis for the Correlation Analysis, done the Exploratory Data Analysis (EDA) and visualize through scatter chart, bubble chart, distribution charts.

FINDINGS

Task 1.A: Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.

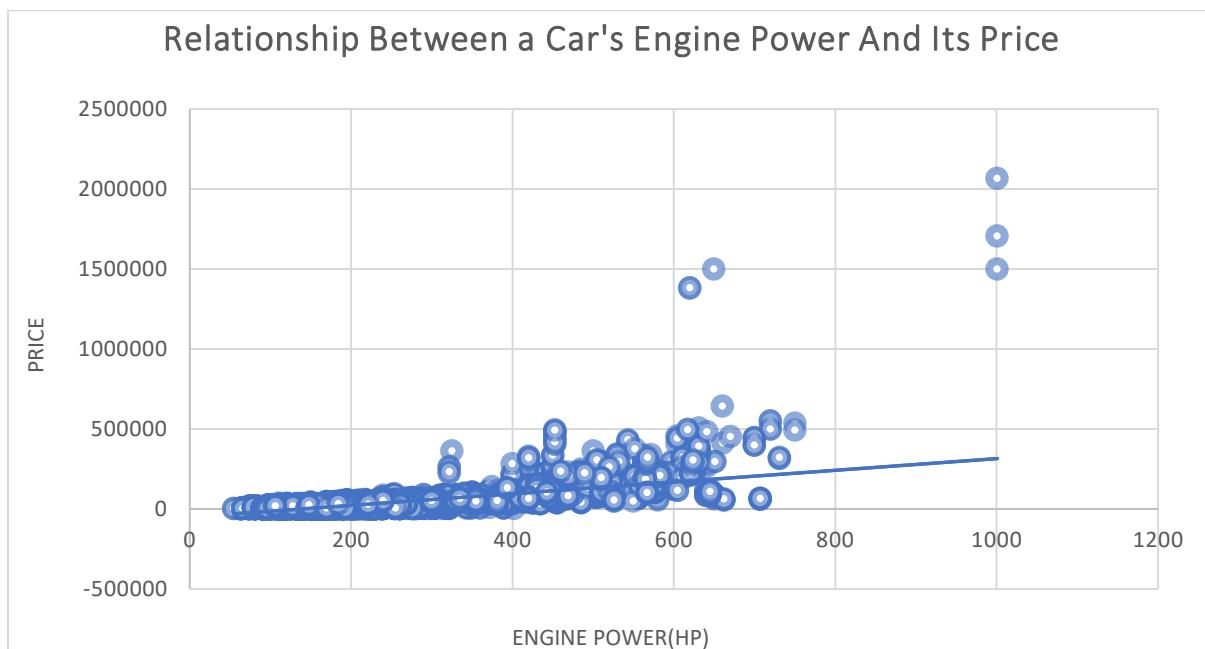
Task 1.B: Create a combo chart that visualizes the relationship between market category and popularity.



INSIGHTS:-

- ❖ ‘Hatchback, Flex, fuel’, ‘Flex Fuel, Diesel’, and Crossover, Flex Fuel, Performance model are very popular but available in few models.
 - ❖ However maximum 1103 models belong to crossover category in the market.
 - ❖ Exotic, Luxury Models are the very low in popularity, it also may be due to high price and only affordable for rich people.
 - ❖ Crossover, Luxury, Performance, Hybrid category have only 2 model available but it is very popular having popularity count is 3916.
 - ❖ ‘Crossover, Exotic, Luxury, High-Performance’, ‘Exotic, Luxury, High-Performance, Hybrid, Performance, Hybrid’, ‘Flex Fuel, Factory Tuner, Luxury, High-Performance’, ‘Crossover, Exotic, Luxury, Performance’ category belong to luxury models and have single model available in Market.

Task 2: 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.



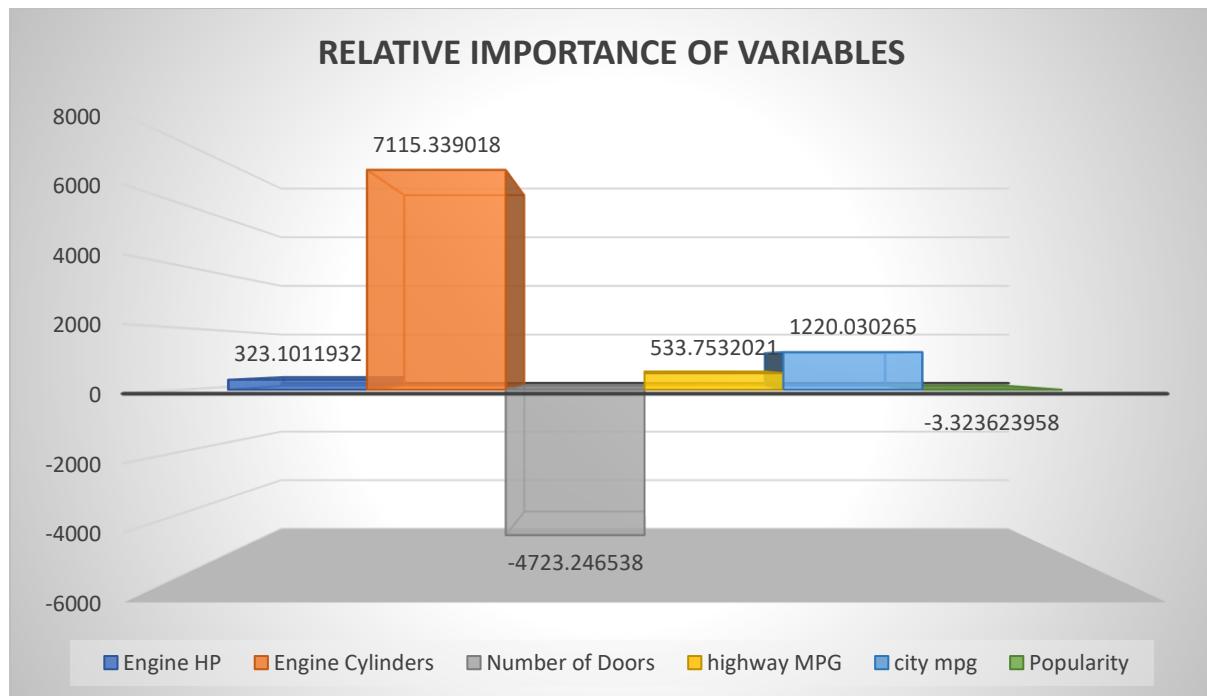
INSIGHTS: -

In the scatter plot graph trendline moving upward and it is showing the positive relationship between price and engine power, it means as the engine power increasing price is also increasing.

Task 3: Use regression analysis to identify the variables that have the strongest relationship with a car's price. Then create a bar chart that shows the coefficient values for each variable to visualize their relative importance.

REGRESSION ANALYSIS

<i>Column1</i>	<i>Coefficients</i>
Intercept	-96506.12134
Engine HP	323.1011932
Engine Cylinders	7115.339018
Number of Doors	-4723.246538
highway MPG	533.7532021
city mpg	1220.030265
Popularity	-3.323623958

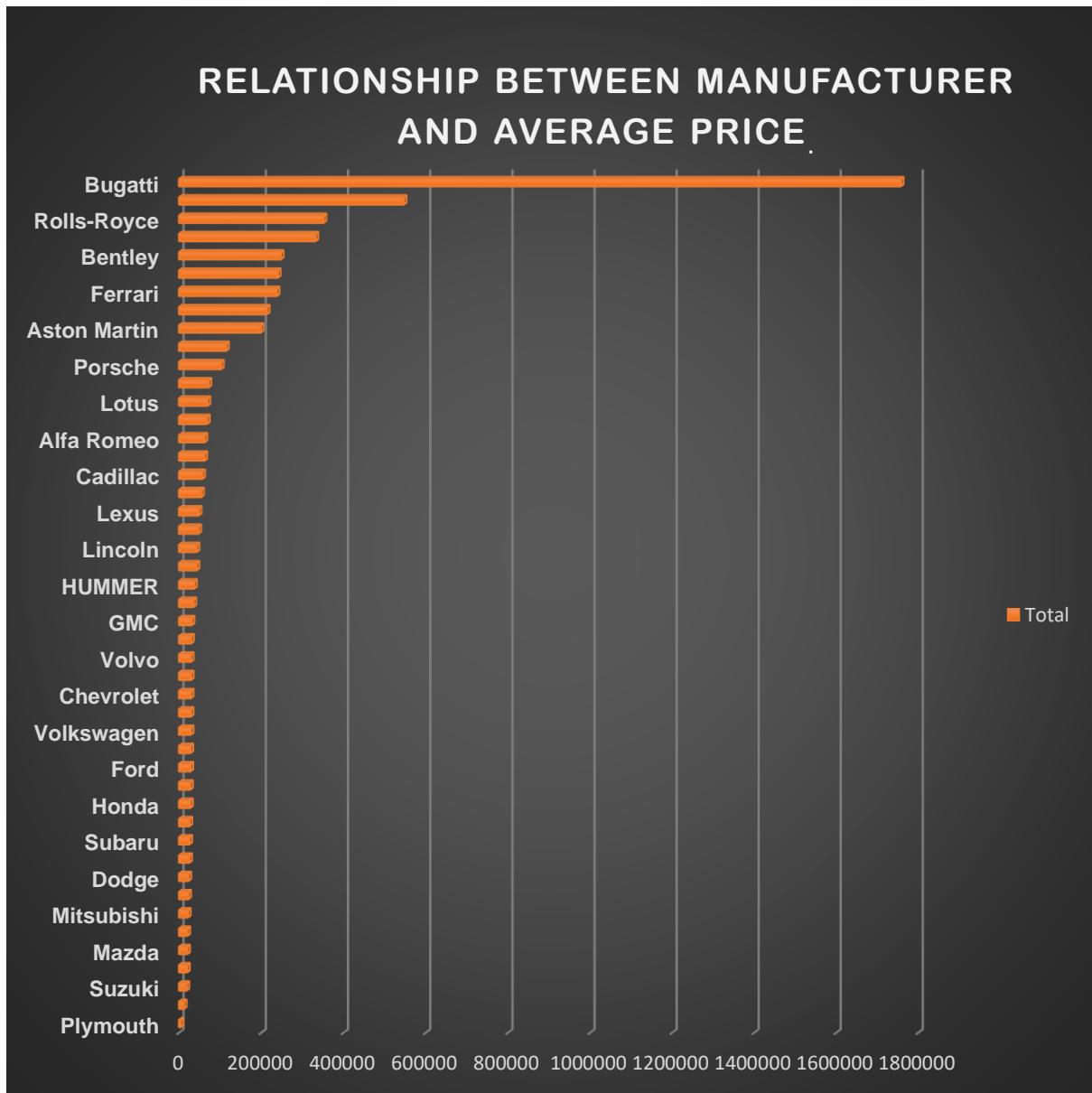


INSIGHTS: -

- ❖ On the basis of regression analysis, we can see that Engine HP, Engine Cylinders, Highway MPG, City MPS having the positive and significant relationship with car price.
- ❖ In the graph Engine cylinder having the most positive relation whereas Number of Doors having the most negative relation with Car Price, it means number of doors increasing then car price decreasing.

Task 4.A: Create a pivot table that shows the average price of cars for each manufacturer.

Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.



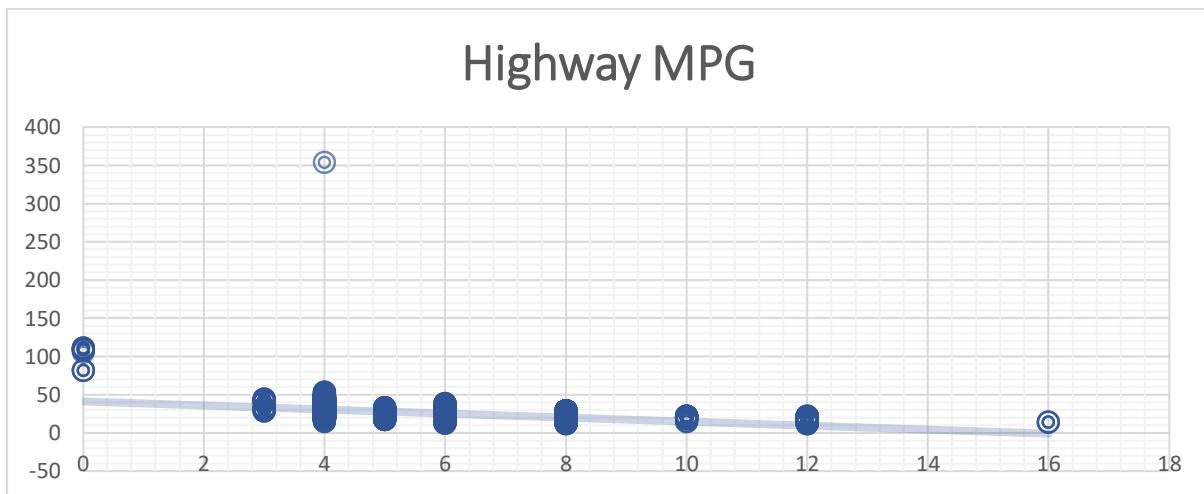
INSIGHTS: -

This analysis shows the average price difference of each car, some company like Bugatti having the very high price followed by Maybach, Rolls-Royce and Lamborghini, some company manufacture the less price car like Plymouth, Oldsmobile, Suzuki.

Task 5.A: Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.

Task 5.B: Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.

Correlation Coefficient		
	<i>Engine Cylinders</i>	<i>highway MPG</i>
Engine Cylinders highway MPG	1 -0.620312551	1

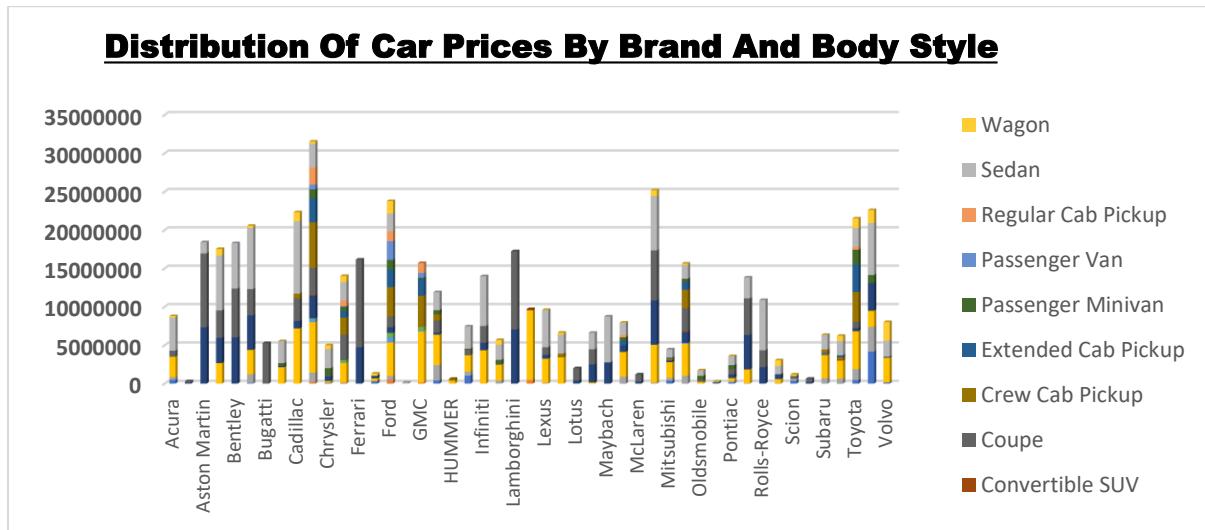


INSIGHTS: -

The correlation between Engine cylinder and Highway mpg is negative, its clearly showing that as the number of cylinders increase in car the Highway milage will also decrease and in the graph the trendline going downwards which is showing negative slope result is declining of Highway milage as per engine cylinder increase.

Task 1: How does the distribution of car prices vary by brand and body style?

- **Hints:** Stacked column chart to show the distribution of car prices by brand and body style. Use filters and slicers to make the chart interactive. Calculate the total MSRP for each brand and body style using SUMIF or Pivot Tables.

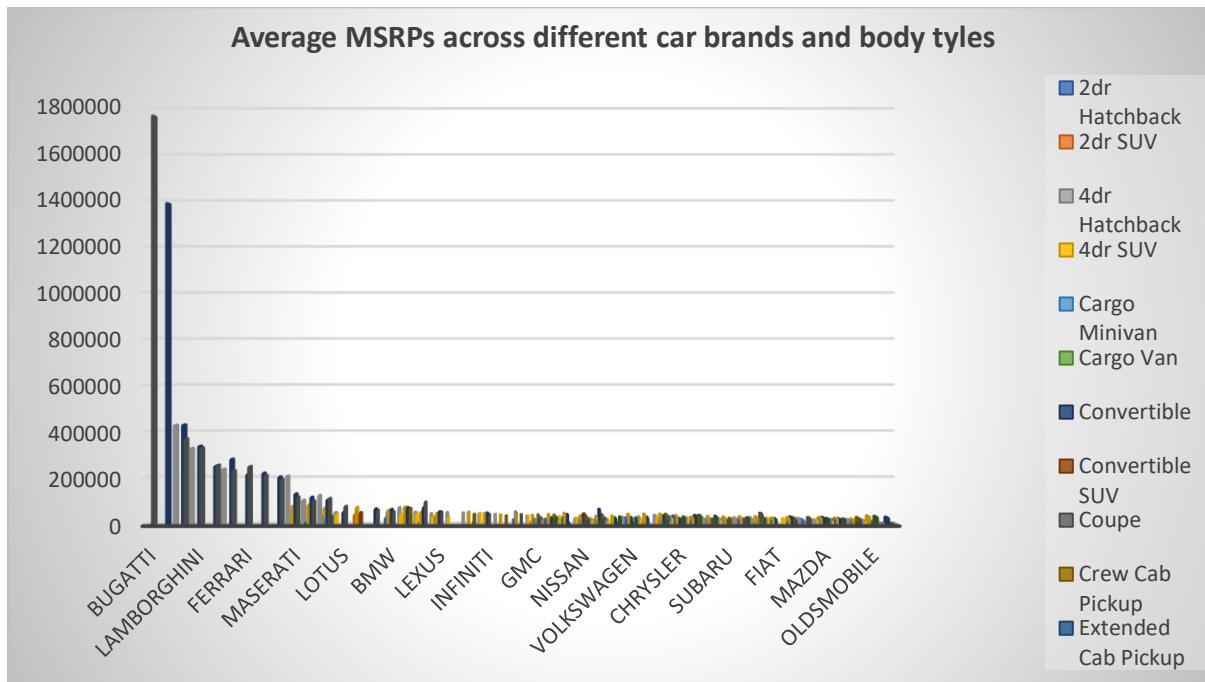


INSIGHTS: -

- ❖ In this task I have calculated the total MSRP as per body type and brand name.
- ❖ Draw the above stacked column charts which showing the distribution of car price by brand and body style.
- ❖ The brand Chevrolet have the maximum number of cars available in different body style then Mercedes-Benz cars having the maximum variant in body style.
- ❖ Bugatti having the single variant of body style

Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

- **Hints:** Clustered column chart to compare the average MSRPs across different car brands and body styles. Calculate the average MSRP for each brand and body style using AVERAGEIF or Pivot Tables.

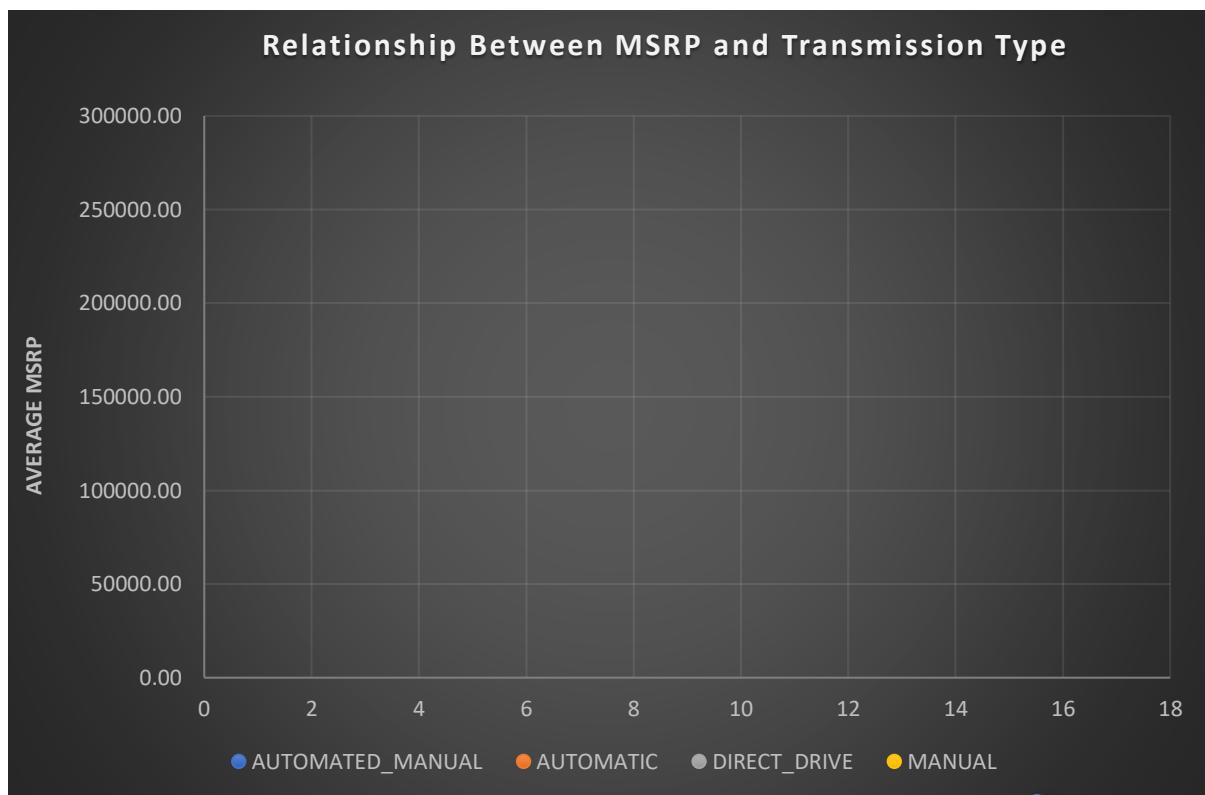


INSIGHTS: -

- ❖ Bugatti brand have the highest average price in Coupe style among all the brands followed by Maybach in Convertible body style.
- ❖ Plymouth brand having the lowest average price compare to all other brands.

Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

- **Hints:** Scatter plot chart to visualize the relationship between MSRP and transmission type, with different symbols for each body style. Calculate the average MSRP for each combination of transmission type and body style using AVERAGEIFS or Pivot Tables.

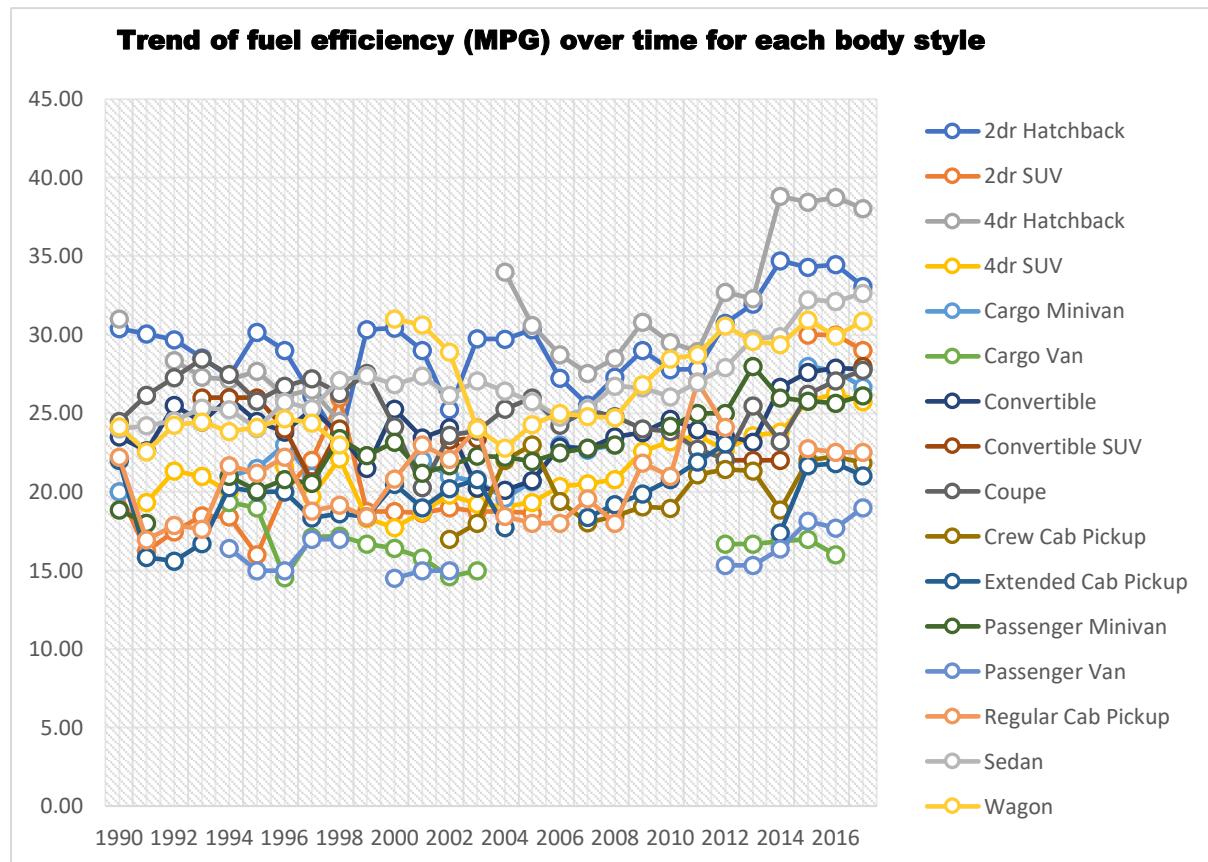


INSIGHTS: -

- ❖ The automated manual transmission having the highest average MSRP and have coupe, and convertible body style.
- ❖ Manual transmission has the lowest average MSRP and it is of body type Passenger minivan.
- ❖ From the graph we can see that the automatic transmission types having the higher average MSRP compare to the manual transmission.
- ❖ Automated manual transmission only available in the luxury brands like Bugatti, Maybach, Lamborghini, Ferrari, Bentley and also average MSRP is very high.

Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

- **Hints:** Line chart to show the trend of fuel efficiency (MPG) over time for each body style. Calculate the average MPG for each combination of body style and model year using AVERAGEIFS or Pivot Tables.

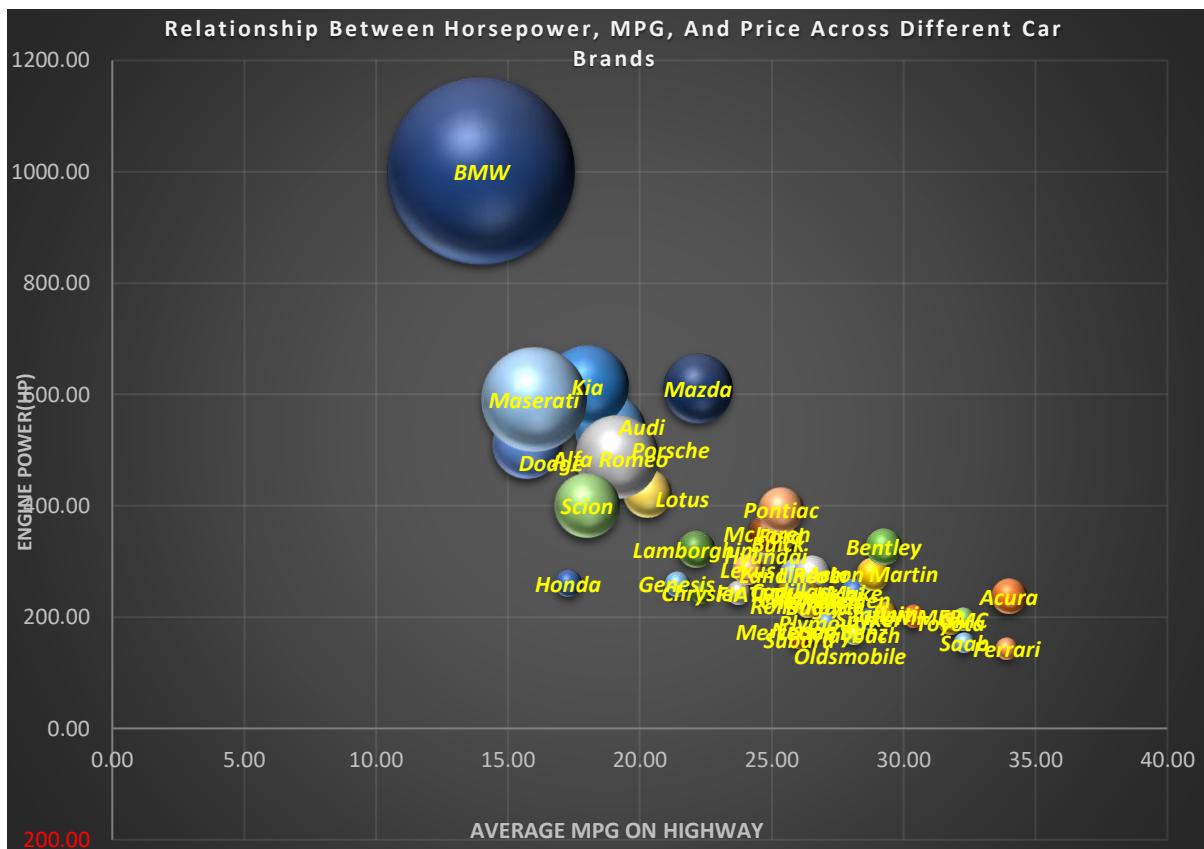


INSIGHTS: -

- ❖ By using the line chart analysis, it is observed that the '4dr Hatchback' body style have the highest milage approx. 39kmpl.
- ❖ At the second position the '2dr Hatchback' followed by sedan body style having the maximum milage.
- ❖ So, from the analysis we can say that these three, style are the most efficient in terms of milage.
- ❖ Passenger Van style are the highest fuel consumption style among all the body style and have no improvement during time.
- ❖ Fuel efficiency have significantly improved in 4dr hatchback during time and have very good milage.

Task 5: How does the car's horsepower, MPG, and price vary across different Brands?

- **Hints:** Bubble chart to visualize the relationship between horsepower, MPG, and price across different car brands. Assign different colors to each brand and label the bubbles with the car model name. Calculate the average horsepower, MPG, and MSRP for each car brand using AVERAGEIFS or Pivot Tables.



INSIGHTS: -

- ❖ Higher-end brands (e.g., Bugatti, Maybach) may have higher average horsepower compared to mid-range brands (e.g., Toyota, Suzuki).
 - ❖ Brands focused on fuel efficiency (e.g., Suzuki) might show higher average MPG.
 - ❖ Luxury brands (e.g., Bugatti, Lamborghini) typically have higher average prices.
 - ❖ Horsepower and price indicate that higher horsepower cars tend to be more expensive.
 - ❖ Correlation between MPG and price suggests that more fuel-efficient cars might be less expensive within some brands.
 - ❖ We can see that the brand BMW have the highest engine power but milage is very low.

ANALYSIS

The analysis provided valuable insights into how car features impact price and profitability. By leveraging these insights, automotive companies can make informed decisions to optimize their product offerings, pricing strategies, and marketing efforts to enhance profitability. This report summarizes the project results, highlighting the significant impact of car features on price and profitability, and provides actionable recommendations based on the findings.

Some of the findings are like

- ❖ Significant positive impact on price and profitability. Cars with higher horsepower are priced higher and yield better profit margins.
- ❖ Negative correlation with price; however, fuel-efficient cars appeal to a specific market segment.
- ❖ Strongest predictor of price and profitability. Higher brand prestige scores lead to significantly higher prices and better profit margins.
- ❖ Positive impact on price, as consumers are willing to pay a premium for safer vehicles.
- ❖ Newer models tend to have higher prices due to technological advancements and consumer preference for the latest models.

CONCLUSIONS

Overall, this project makes me learnt the various data analysis in excel, like regression analysis, creating the bubble charts, scatter charts, calculating the average, sum by using the pivot table etc. This project enabled my EDA skill to next level, also enhanced my knowledge to extract the meaningful insights from the data to take the progressive decision and contribute in meaningful outcomes.

ABC CALL VOLUME TREND ANALYSIS

Description: -

The main objective of this project ABC Call Volume Analysis is to examine and understand the trends in call volumes over a specified period for ABC Company. This analysis aims to identify patterns, peak periods, and potential areas for improvement in call handling and resource allocation. This analysis will help ABC Company better understand their call volume trends, leading to improved customer service and operational efficiency.

Problem: - Analyzing the impact of car features on price and profitability is essential for automotive manufacturers and dealers to make data-driven decisions. Understanding which features drive higher prices and profitability can help in product development, marketing strategies, and sales tactics. This analysis will help in identifying which features contribute most to higher prices and profitability and how manufacturers can optimize their offerings.

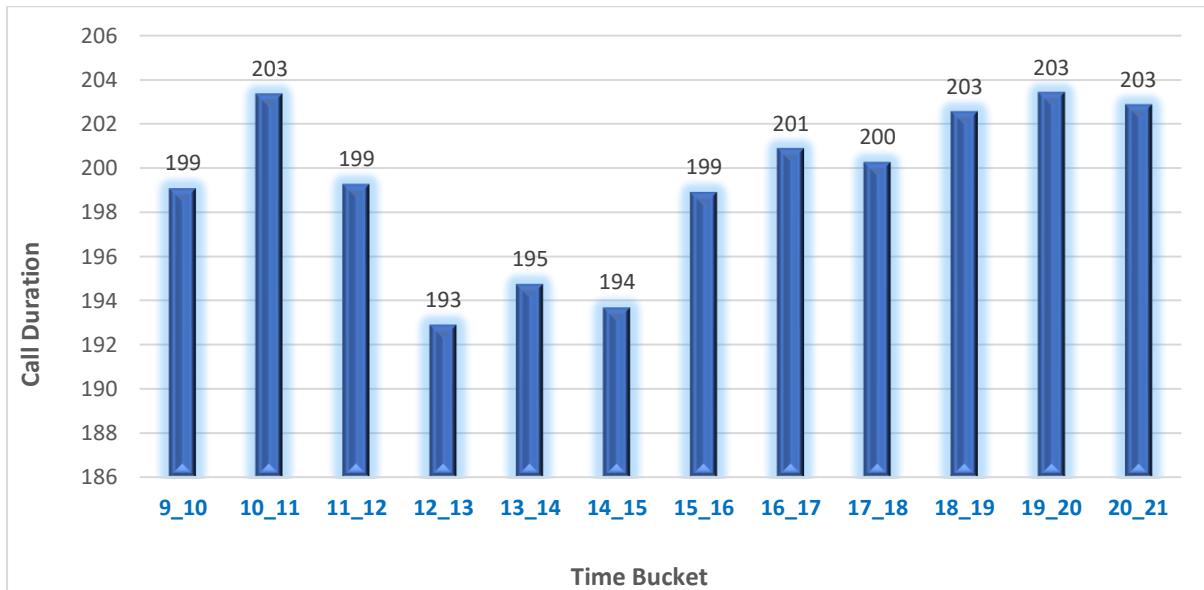
Design: - The following steps taken to clean the data.

I have used the approach that includes data collection, preprocessing, analysis, and visualization.

- **Cleaning Data:** Handle missing values, remove duplicates.
- **Time Series Formatting:** Convert data into a time series format where call volume is aggregated over the chosen time intervals.
- **Exploratory Data Analysis (EDA):** Done the EDA and created the graph and chart.
- **Data Visualization:** Visualize the data through graph
- **Dashboard Creation:** Develop an interactive dashboard stakeholder to explore call volume trends in detail.
- **Insights and Reporting:** Find out the insights and created the report.

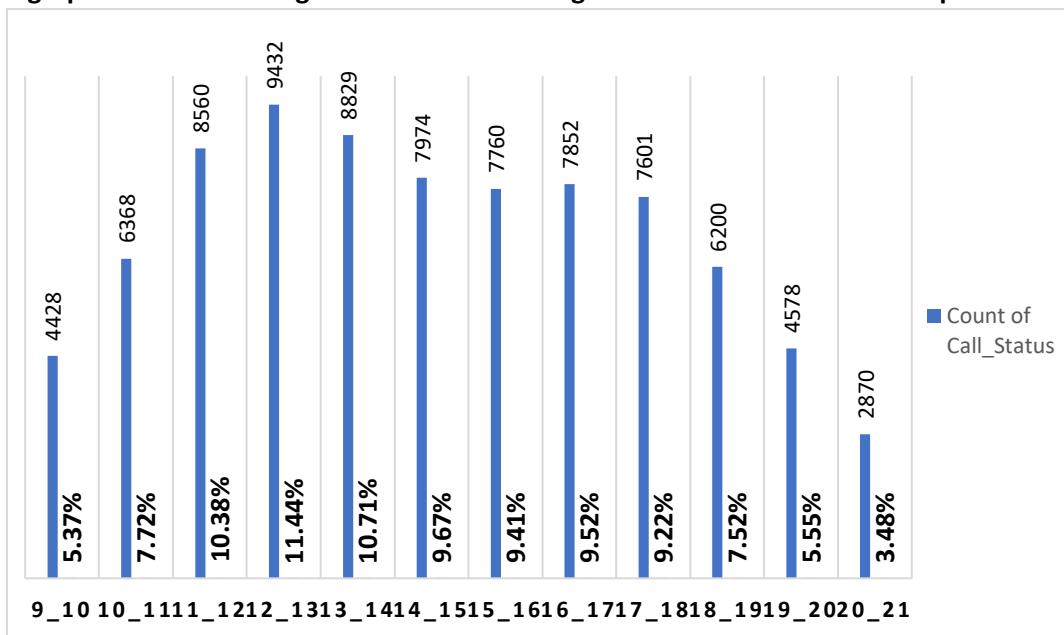
FINDINGS: -

- 1. Average Call Duration:** Determine the average duration of all incoming calls received by agents. This should be calculated for each time bucket.



- Maximum Call duration found in four time slot that is 10-11, 18-19, 19-20 and 20-21 that is total 203 Second.
- Call duration is minimum in time 12PM to 15PM.
- Total Average call duration in day shift from 9am to 9pm is 199 Sec
- By analysis the call volume as per time slots we can check the call duration and decide for manpower required.

- 2. Call Volume Analysis:** Visualize the total number of calls received. This should be represented as a graph or chart showing the number of calls against time. Time should be represented in buckets.

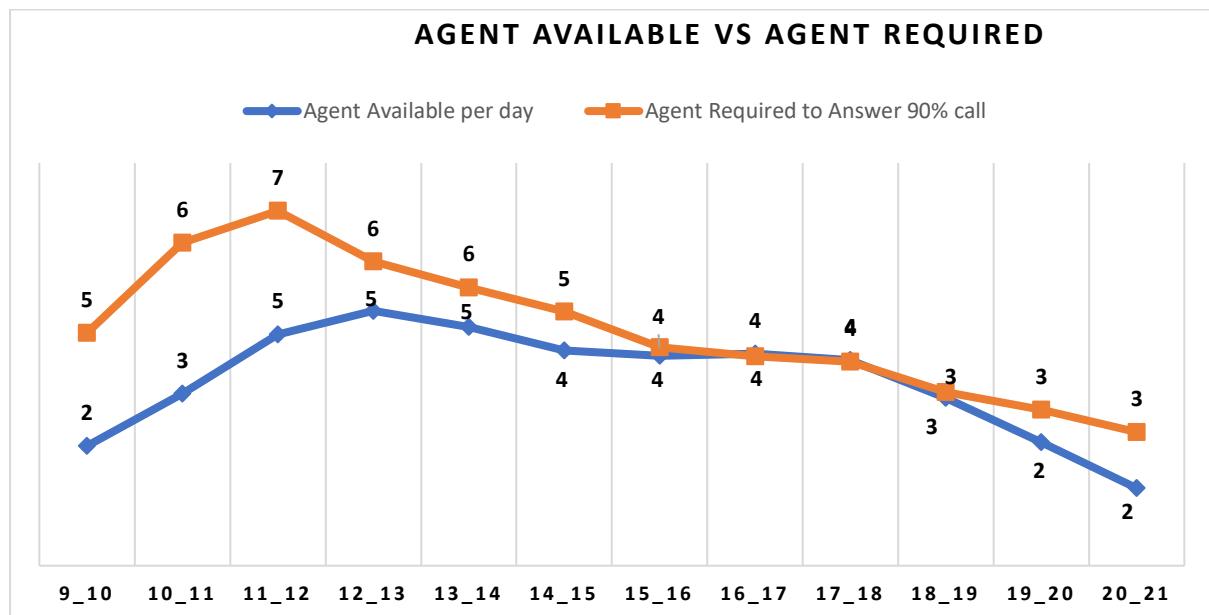


INSIGHTS: -

- Maximum Call received in time slot of 12-13 and 13-14 O'clock.
- Call count is minimum between 8 - 9 pm.
- Total Average call count in 23 day is 82452.
- Total call count is maximum between 11 to 14 which is approx. 11% of whole day call count.

3. Manpower Planning: The current rate of abandoned calls is approximately 30%. Propose a plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%. In other words, you need to calculate the minimum number of agents required in each time bucket to ensure that at least 90 out of 100 calls are answered.

Abandon call/23 Day	Answered Call/23 day	Total Calls/23 Day	Answered %/23day	Abandon % /23Day	Required to Answer call per day to achieve 90 %	Agent Available per day	Agent Required to Answer 90% call
224	193	416	46	54	375	2	5
300	277	577	48	52	520	3	6
262	372	634	59	41	571	5	7
134	410	544	75	25	489	5	6
114	384	498	77	23	448	5	6
108	347	454	76	24	409	4	5
53	337	390	86	14	351	4	4
32	341	374	91	9	336	4	4
34	330	365	91	9	328	4	4
41	270	310	87	13	279	3	3
80	199	279	71	29	251	2	3
114	125	239	52	48	215	2	3
1496	3585	5081	71	29	4573	44	56

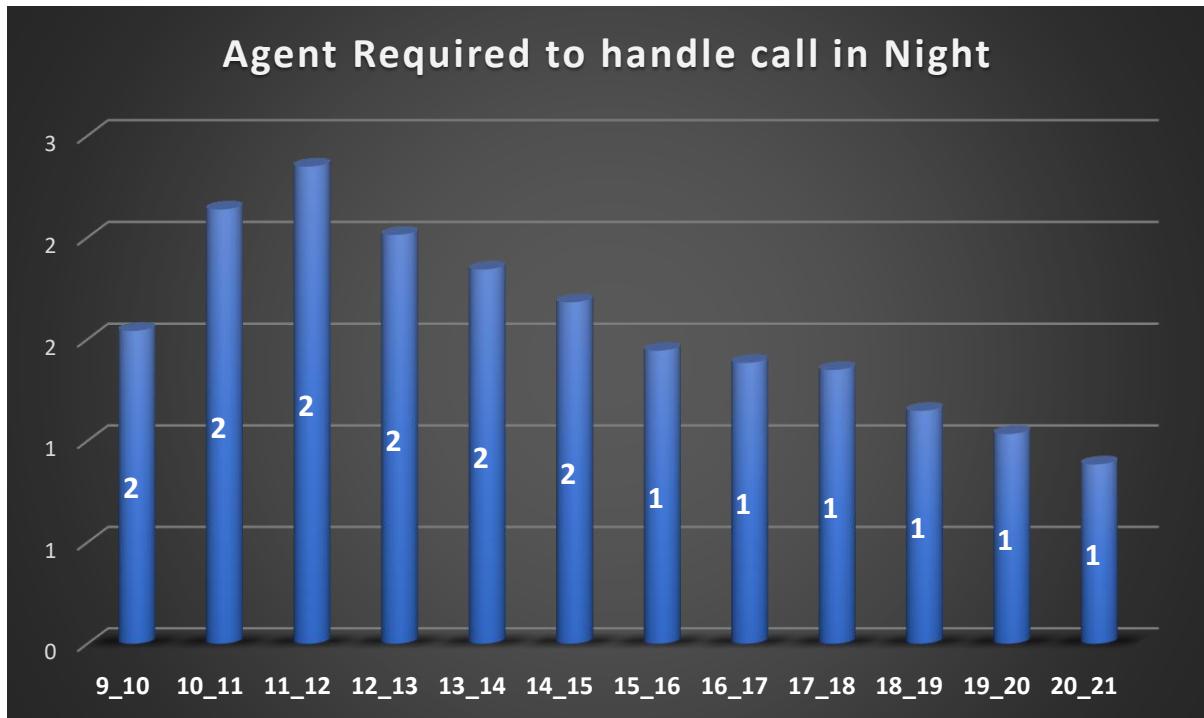


INSIGHTS: -

- Maximum agent required in morning time 9 to 12 PM due to call volume is high.
- In the time slot 3PM to 7PM no extra man power required.
- More 12 agents required to answer 90% of daily incoming calls.
- There is high abandon percent from 9 am to 12pm and in evening 8pm to 9pm

4.Night Shift Manpower Planning: Customers also call ABC Insurance Company at night but don't get an answer because there are no agents available. This creates a poor customer experience. Assume that for every 100 calls that customers make between 9 am and 9 pm, they also make 30 calls at night between 9 pm and 9 am. The distribution of these 30 calls is as follows:

9 PM TO 9 AM	No of Calls (Assumption)	Call Percent (Assumption)	Total Calls/23 Day	Call Received in Day time(70%)	Call Received in Night time(30%)	Agent Required to handle call in Night
9_10	3	10.00	416	291	125	2
10_11	3	10.00	577	404	173	2
11_12	2	6.67	634	444	190	2
12_13	2	6.67	544	381	163	2
13_14	1	3.33	498	348	149	2
14_15	1	3.33	454	318	136	2
15_16	1	3.33	390	273	117	1
16_17	1	3.33	374	262	112	1
17_18	3	10.00	365	255	109	1
18_19	4	13.33	310	217	93	1
19_20	4	13.33	279	196	84	1
20_21	5	16.67	239	167	72	1
TOTAL	30	100.00	5081	3556	1524	19

**INSIGHTS: -**

- Call Volume is maximum between 10 PM to 1 AM
- Total 19 agent required to handle the incoming call in the night.
- Total 5081 call received in a day in which 30% i.e., 1524 call received at night time.
- As per assumption call flow increased from 6AM and 17% call flow found at 8AM to 9AM.

ANALYSIS

- ❖ The analysis reveals that the call volumes exhibit trends over the analyzed period. Notable peaks were observed in specific time, indicating periods of higher customer engagement or service demand.
- ❖ There is a clear timing pattern, with call volumes peaking during 11AM to 2PM. This suggests a need for increased staffing or resource allocation during these periods.
- ❖ By analysis we can see that abandon rate is high during shift start and shift getting ended, so we have to utilize and motivate the agents to grab and answer the more calls.
- ❖ Based on the analysis, it is recommended to increase the staff during peak hours, implement more efficient call-handling processes, enhance customer self-service options, etc.
- ❖ Overall, this project enhanced my resource utilization planning, analysis through the Excel skill, Data analysis technique, also improved my ability to finding the key insights from the Data.

CONCLUSIONS

Daily Peaks: The data indicates that there are specific times of day when call volumes are consistently high. These peaks are likely aligned with customer work schedules, particularly during early morning hours and post-lunch periods.

High Traffic Seasons: The analysis may show distinct periods of increased call activity, possibly linked to seasonal promotions, product launches, or holiday periods.

Promotional Campaigns: Spikes in call volume can often be correlated with marketing efforts or new product launches, indicating effective customer engagement.

Call Duration and Resolution Times: The trends might show variations in call durations, possibly indicating complexity or customer satisfaction with the resolution process.

Staffing Needs: Understanding peak call times allows for better scheduling and resource allocation, ensuring adequate staff during high-demand periods.

APPENDIX

Google Sheet link for Instagram User Analytics Project: -

<https://docs.google.com/presentation/d/1tlisY3VwzDC961TKEpsiZPsg3BSNpqzG/edit?usp=sharing&ouid=103947946418359724622&rtpof=true&sd=true>

Google Sheet link for Operation & Metric Analytics Project: -

https://drive.google.com/file/d/1RWvrHH29afhWO5soJVqypxnREJFu5_j8/view?usp=sharing

Google Sheet link for Hiring Process Analytics Project: -

<https://docs.google.com/presentation/d/1Ph7D0C8Amq1oU3pDdp1RTrbh1eFM8vT/edit?usp=sharing&ouid=103947946418359724622&rtpof=true&sd=true>

Google Sheet link for IMDB Movie Analysis Project: -

<https://docs.google.com/presentation/d/1635P-5vGm89-037qtCPf7a2ICRKCFAGt/edit?usp=sharing&ouid=103947946418359724622&rtpof=true&sd=true>

Google Sheet link for Bank Loan Case Study project: -

<https://drive.google.com/file/d/1H3g6IS9DIB1glU7Byh8egWNs9V5hiVb2/view?usp=sharing>

Google Sheet link for Impact of Car Features project: -

<https://drive.google.com/file/d/1NLyL45SCroSvVly4GMnaH6aUpcrpEkAp/view?usp=sharing>

Google Sheet link for ABC Call Volume Trend Project: -

https://docs.google.com/presentation/d/1Mg-Bb6wxnvU57URkH6a_F1vpCzzdVZeP/edit?usp=sharing&ouid=103947946418359724622&rtpof=true&sd=true