

DATA ANALYSIS PORTFOLIO

BY MANSHOO GULATI



Table of content

S.NO.	TOPIC	Page NO.
1	Professional background	4
2	Data analytics process	5 - 7
	Description	5
	Design	6 - 7
	Conclusion	7
3	Instagram user analytics	8 - 16
	Description	9
	Problems	10 - 11
	Design	11
	Findings	12 - 14
	Analysis	15
	Conclusion	16
4	Operation and metric analytics	17 - 26
	Description	18
	Problems	19 - 20
	Design	21
	Findings	21 - 24
	Analysis	24 - 25
	Conclusion	26

5	Hiring process analytics	27 - 34
	Description	28
	Problems	28 - 29
	Design	30
	Findings	30 - 32
	Analysis	32 - 33
	Conclusion	34
6	IMDB movie analysis	35 - 41
	Description	36
	Problems	36 - 36
	Design	37
	Findings	38 - 40
	Analysis	40
	Conclusion	41
7	Bank loan analysis	42 - 53
	Description	43
	Problems	43 - 45
	Design	45
	Findings	45 - 51
	Analysis	52 - 53
	Conclusion	53
8	Impact of car features	54 - 63
	Description	55
	Problems	55 - 56
	Design	57
	Findings	57 - 61
	Analysis	62 - 63
	Conclusion	63
9	ABC Call volume trend analysis	64 - 73
	Description	65
	Problems	66
	Design	67
	Findings	67 - 70
	Analysis	71 - 72
	Conclusion	73
10	Appendix	74 - 75

PROFESSIONAL BACKGROUND

Name : Manshoo Gulati

Academic pursuit : Btech CSE with specialization in
Data Science

Year – 2nd

Skills:

Excel Mastery: Proficient in advanced Excel
functionalities for data manipulation and analysis.

SQL Proficiency: Skilled in querying databases for data
extraction and analysis.

Programming languages : C , Java , Python

DATA ANALYTICS PROCESS

DESCRIPTION

We use Data Analytics in everyday life without even knowing it.

For eg : Going to a market to buy something .

DESIGN

PLAN –

The first basic thing to be decided is what I wanna have depending upon my cravings, is it something sweet I am craving for, something spicy and crispy, or something soft and bland.

PREPARE -

Next, asking for permission and the amount of money I need to get for it from my parents and then asking my friends or my brother to join me.

PROCESS -

If I wanna go for a light meal or a healthier option either I can go for subway or a café preparing boiled chicken/chicken breast or MC'D for fillet-o-fish burger.

ANALYZE -

I need to check if the beverage I wanna drink will make a good combination with the meal chosen. (it is advised to avoid milk with fish and me being a coffee lover , I'll most probably go for a coffee). OK so I decided to go at 'SUBWAY' and grab a coffee from 'THE WAFFLE COMPANY' nearby. Then after paying I can ask for the time needed for the 'SUB' to be prepared and meanwhile I can go at the waffle company and grab my favourite frappe.

SHARE -

Then at the subway I've to communicate with the submaker about the customisation of the sub i.e. length of the bread , veg/non veg sub , the sauces and the vegetables I prefer in my sub.

ACT -

I can finally enjoy my meal.

CONCLUSION

Hence, we have seen how we can use the 6 steps of data analytics (plan , prepare , process , analyze , share , act) while making any decisions in real life in the example of choosing a meal to satisfy the cravings

INSTAGRAM USER ANALYTICS

DESCRIPTION

As a data analyst collaborating with Instagram's product team, my role centered around delving into user interactions within the app. Utilizing SQL and MySQL Workbench, I conducted a thorough analysis of user data to extract invaluable insights.

The objective was to empower decision-making across diverse teams. From aiding marketing endeavors in crafting targeted campaigns to guiding product development by suggesting new features, the findings served as a compass for the entire business.

My work involved tracking user engagement patterns, deciphering trends, and presenting actionable recommendations. This project not only showcased my proficiency in SQL but also underscored my ability to transform raw data into strategic insights. The outcomes contributed to informed decisions steering the future evolution of one of the world's most popular social media platforms.

PROBLEM

A) Marketing Analysis:

1.Loyal User Reward: The marketing team wants to reward the most loyal users, i.e., those who have been using the platform for the longest time.

Your Task: Identify the five oldest users on Instagram from the provided database.

2.Inactive User Engagement: The team wants to encourage inactive users to start posting by sending them promotional emails.

Your Task: Identify users who have never posted a single photo on Instagram.

3.Contest Winner Declaration: The team has organized a contest where the user with the most likes on a single photo wins.

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

4.Hashtag Research: A partner brand wants to know the most popular hashtags to use in their posts to reach the most people.

Your Task: Identify and suggest the top five most commonly used hashtags on the platform.

5.Ad Campaign Launch: The team wants to know the best day of the week to launch ads.

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

B) Investor Metrics:

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

Your Task: Calculate the average number of posts per user on Instagram. Also, provide the total number of photos on Instagram divided by the total number of users.

2. Bots & Fake Accounts: Investors want to know if the platform is crowded with fake and dummy accounts.

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

DESIGN

Steps taken to load the database

1. Using “create db” function of MYSQL to create a database.
2. Adding columns and tables
3. Adding the values into the columns using “insert into” function of MYSQL
4. Retrieving the data via “select” command.

Software used → MYSQL Workbench 8.0 CE

FINDINGS

1. These are the 5 oldest users

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

2. These are the people who have never posted on Instagram.

id	username	count(image_url)	-	U
5	Aniya_Hackett	0	66	Mike.Auer39 0
7	Kassandra_Homenick	0	68	Franco_Keebler64 0
14	Jadyn81	0	71	Nia_Haag 0
21	Rocio33	0	74	Hulda.Macejkovic 0
24	Maxwell.Halvorson	0	75	Leslie67 0
25	Tierra.Trantow	0	76	Janelle.Nikolaus81 0
34	Pearl7	0	80	Darby_Herzog 0
36	Ollie_Ledner37	0	81	Esther.Zulauf61 0
41	Mckenna17	0	83	Bartholome.Bernhard 0
45	David.Osinski47	0	89	Jessyca_West 0
49	Morgan.Kassulke	0	90	Esmeralda.Mraz57 0
53	Linnea59	0	91	Bethany20 0
54	Duane60	0		
57	Julien_Schmidt	0		

3. Usernames with most liked posts

	username	id	image_url	total
▶	Zack_Kemmer93	145	https://jarret.name	48
	Adelle96	182	https://dorcus.biz	43
	Malinda_Streich	127	https://celestine.name	43
	Seth46	123	http://shannon.org	42

4. 5 Most commonly used hashtags

	tag_name	totalusage
▶	smile	59
	beach	42
	party	39
	fun	38
	concert	24

5. Day of the week when most users register

	day	total_users
▶	Thursday	16
	Sunday	16

6. Avg posts per user

	avg_photos_posted
▶	2.5700

7. Potential bots (who have liked every single picture)

	username	num_of_likes
▶	Aniya_Hackett	257
	Jadyn81	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 performing the analysis , these are the following outcomes-

- Most loyal users (oldest users) are –

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

- 26 out of 100 users haven't posted anything on Instagram.
- Zack_kemmer93 is the winner of the contest with most number of likes on a post.
- #smile , #fun , #concert , #beach , #party are the most commonly used hashtags.
- Most users register on Instagram on Thursdays and Sundays.
- Average posts per user = 2.57
- Out of 100 users , 13 are potentially bots as it is not practically possible for a human to like each and every post on Instagram.

CONCLUSION

I would like to conclude that In the world of social media and businesses, not just Instagram, but many companies dig into customer data like a treasure hunt.

Every week, month, and year, they're doing some serious homework on your likes, clicks, and swipes. They want to find the folks who'll stick around, love their stuff, and make the company shine.

It's not just about selling things. They're playing the long game, trying to be smart about spending less money while making more of it. So, they're like scientists, looking at all the data under a microscope, figuring out who's going to be the MVP (Most Valuable Person) for the company.

So, next time you see a cool new feature or a catchy ad, remember, it might just be the result of this behind-the-scenes detective work, all thanks to your data and the magic of analysis!

OPERATION AND METRIC ANALYTICS

DESCRIPTION

Operation analytics involves analyzing and optimizing business operations to improve efficiency, productivity, and customer satisfaction. Metric analytics focuses on tracking, measuring, and interpreting key performance indicators (KPIs) to gain valuable insights.

As a Lead Data Analyst, imagine being the detective of a company like Microsoft. Your mission? To analyze the nitty-gritty details of how the entire operation runs. You're not just crunching numbers; you're helping teams like operations, support, and marketing spot areas to level up.

In this project, you're handed datasets from different corners of the company, and your superpower is advanced SQL skills. Your goal? Uncover insights that can boost how the company works and explain those unexpected jumps and dips in the numbers. You're not just a data analyst; you're the data superhero of Microsoft!

PROBLEMS

CASE STUDY 1

A.Jobs Reviewed Over Time:

- A. Objective: Calculate the number of jobs reviewed per hour for each day in November 2020.
- B. Your Task: Write an SQL query to calculate the number of jobs reviewed per hour for each day in November 2020.

B.Throughput Analysis:

- A. Objective: Calculate the 7-day rolling average of throughput (number of events per second).
- B. Your Task: 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.

C.Language Share Analysis:

- A. Objective: Calculate the percentage share of each language in the last 30 days.
- B. Your Task: Write an SQL query to calculate the percentage share of each language over the last 30 days.

D.Duplicate Rows Detection:

- A. Objective: Identify duplicate rows in the data.
- B. Your Task: Write an SQL query to display duplicate rows from the job_data table.

CASE STUDY 2

A. Weekly User Engagement:

- A. Objective: Measure the activeness of users on a weekly basis.
- B. Your Task: Write an SQL query to calculate the weekly user engagement.

B. User Growth Analysis:

- A. Objective: Analyze the growth of users over time for a product.
- B. Your Task: Write an SQL query to calculate the user growth for the product.

C. Weekly Retention Analysis:

- A. Objective: Analyze the retention of users on a weekly basis after signing up for a product.
- B. Your Task: Write an SQL query to calculate the weekly retention of users based on their sign-up cohort.

D. Weekly Engagement Per Device:

- A. Objective: Measure the activeness of users on a weekly basis per device.
- B. Your Task: Write an SQL query to calculate the weekly engagement per device.

E. Email Engagement Analysis:

- A. Objective: Analyze how users are engaging with the email service.
- B. Your Task: Write an SQL query to calculate the email engagement metrics.

DESIGN

Steps taken to load the database.

- Downloading the dataset
- Creating a database
- Importing the dataset in the database.
- Using “select” command to retrieve the data required as per the task.
- Visualizing the final report using powerpoint
- Software used for data retrieval → MYSQL workbench 8.0

FINDINGS

CASE STUDY 1

A. Jobs reviewed over time

	avg jobs reviewed per hour	avg jobs reviewed per second
▶	126.18048333	0.03505000

B. Throughput analysis

review_date	jobs_reviewed	7_rolling_days_avg
2020-11-25	1	1.0000
2020-11-26	1	1.0000
2020-11-27	1	1.0000
2020-11-28	2	1.2500
2020-11-29	1	1.2000
2020-11-30	2	1.3333

C. Language share analysis

	Languages	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

D. Duplicate rows detection

	actor_id	Duplicates
▶	1003	2

CASE STUDY 2

A. Weekly user engagement

Year	weekNo	User_Engagement	2014	26	1302
2014	17	663	2014	27	1372
2014	18	1068	2014	28	1365
2014	19	1113	2014	29	1376
2014	20	1154	2014	30	1467
2014	21	1121	2014	31	1299
2014	22	1186	2014	32	1225
2014	23	1232	2014	33	1225
2014	24	1275	2014	34	1204
2014	25	1264	2014	35	104

B. User growth analysis

year	weeknum	new_active_user	cum_active_user
2013	0	23	23
2013	1	30	53
2013	2	48	101
2013	3	36	137
2013	4	30	167

C. Weekly retention

user_id	total_users	retention
11919	2	0
11920	1	0
11924	1	0
11926	8	1
11928	8	0
11929	1	0
11931	6	1
11933	6	1
11936	3	0
11939	3	1
11940	4	1
11942	7	1
11944	3	1
11947	2	1

163
Rows
Returned

D. Weekly engagement per device

year	week_num	device	users
2014	17	acer aspire desktop	9
2014	17	acer aspire notebook	20
2014	17	amazon fire phone	4
2014	17	asus chromebook	21
2014	17	dell inspiron desktop	18

491 rows
returned

E. Email engagement analysis

emails_rate_of_opening	emails_rate_of_opening
25.14039	12.88431

ANALYSIS

- No. of jobs reviewed per hour = 126.180
- For the dates 25 , 26 , 27 , 28 , 29 , 30 – 7 days rolling average is 1 , 1 , 1 , 1.25 , 1.20, 1.3333 respectively.
- Persian language has the highest contribution.
- There are 2 duplicate rows.
- Weekly user engagement is the highest for week number 30 that is 1467.
- Emails rate of opening = 25.140
- Emails clicking rate = 12.88

I have used Why's approach to gain more insights.

1. Why is the operation and metrics analysis important?

→ It helps us spot patterns, ensuring that we can manage workload efficiently, especially during peak hours, and potentially optimize processes for better user experience.

2. Why analyze user engagement regularly?

→ Regular analysis ensures we stay attuned to user behavior, enabling us to adapt and enhance the product based on evolving user preferences and needs.

3. Why measure the activeness of users on a weekly basis?

→ Weekly measurement provides a consistent and meaningful timeframe for assessing user engagement trends, helping us understand user behavior over time.

4. Why do we need to identify duplicate rows in the data?

→ Identifying duplicates ensures data accuracy, prevents errors in analysis, and maintains the integrity of our database

5. Why calculate the percentage share of each language in the last 30 days?

→ We need to adapt strategies, allocate resources efficiently, and cater to the preferences of our diverse user base based on recent language distribution trends.

CONCLUSION

In conclusion, the regular operation analysis and investigation of metric spikes are imperative for any business. The frequency of these analyses, whether daily, weekly, monthly, quarterly, or yearly, should align with the dynamic needs of the firm. This adaptive approach ensures that the business remains agile, responding swiftly to emerging trends and challenges.

Moreover, emphasizing email engagement is paramount. Crafting attention-grabbing headlines paired with enticing discounts and coupons serves as a potent strategy to not only retain existing customers but also to expand the customer base. Effective email engagement establishes a direct line of communication, fostering a meaningful connection with the audience.

Furthermore, recognizing the importance of potential customers who left the sign-up process midway is a strategic move. Establishing a dedicated department to address their concerns and guide them through the process can be transformative. This proactive approach not only seeks to understand the reasons behind their initial hesitation but also aims to convert them from mere visitors into loyal customers. It's a customer-centric initiative that values every interaction and aims for sustained business growth..

HIRING PROCESS ANALYTICS

DESCRIPTION

Imagine yourself as a data analyst in a company . Your primary task involves delving into the company's hiring process data to extract valuable insights. The hiring process is a critical function, and by understanding trends such as rejection rates, interviews, job types, and vacancies, you play a key role in helping the hiring department make informed decisions.

You'll be working with a dataset containing records of previous hires, and your goal is to analyze this data to answer specific questions that can contribute to improvements in the company's hiring process. Your role is essential in deciphering patterns and providing actionable insights, ensuring that Google continues to refine and optimize its approach to talent acquisition.

PROBLEMS

A. Hiring Analysis: The hiring process involves bringing new individuals into the organization for various roles.

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

B. Salary Analysis: The average salary is calculated by

adding up the salaries of a group of employees and then dividing the total by the number of employees.

Your Task: What is the average salary offered by this company? Use Excel functions to calculate this

.

C. Salary Distribution: Class intervals represent ranges of values, in this case, salary ranges. The class interval is the difference between the upper and lower limits of a class.

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

D. Departmental Analysis: Visualizing data through charts and plots is a crucial part of data analysis.

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

E. Position Tier Analysis: Different positions within a company often have different tiers or levels.

Your Task: 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.

DESIGN

Before starting the tasks –

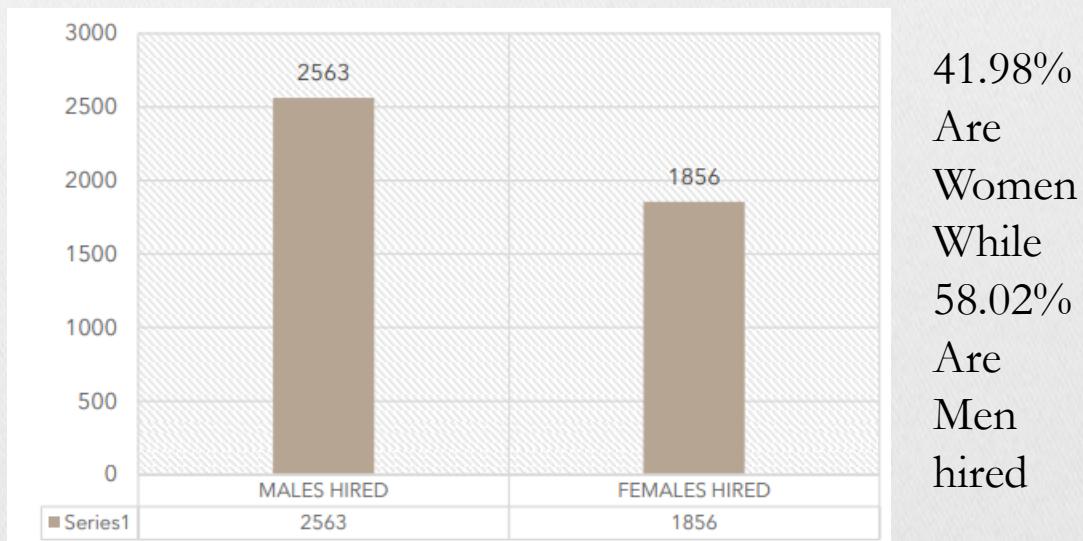
- Made a copy of the original dataset
- Looked for blanks/duplicates and removed them.
- Removed irrelevant columns which aren't needed for the tasks.

Used the excel functions to perform the tasks. Then visualizing them in powerpoint

FINDINGS

A. Gender distribution of hires

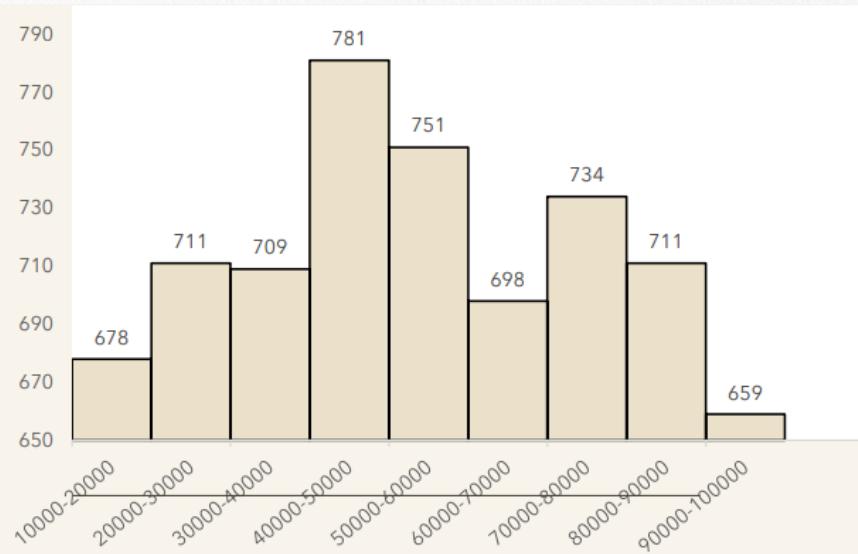
GENDER	STATUS	TOTAL NO. OF PEOPLE
MALES	HIRED	1855
FEMALES	HIRED	2563



B. Average salary offered

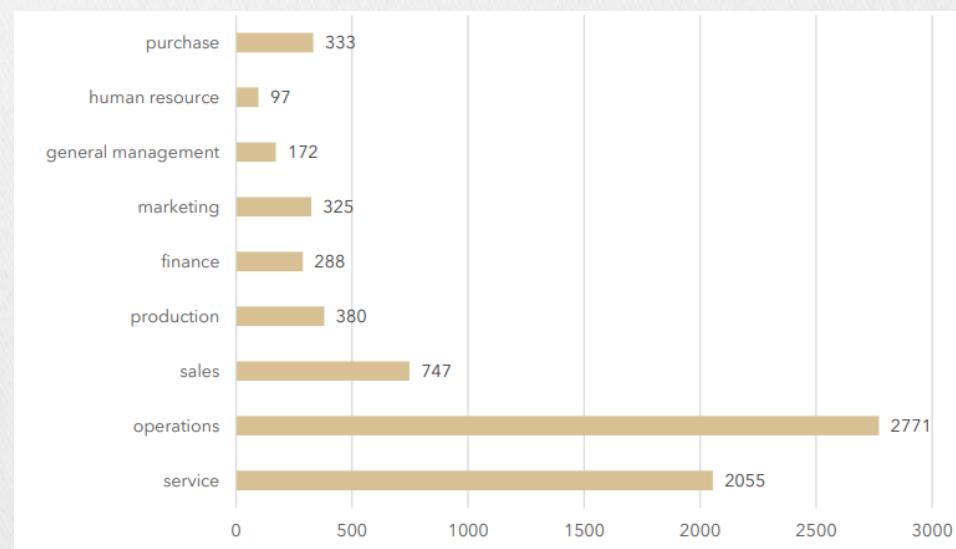
OUTPUT -
49983.02902

C. Salary distribution



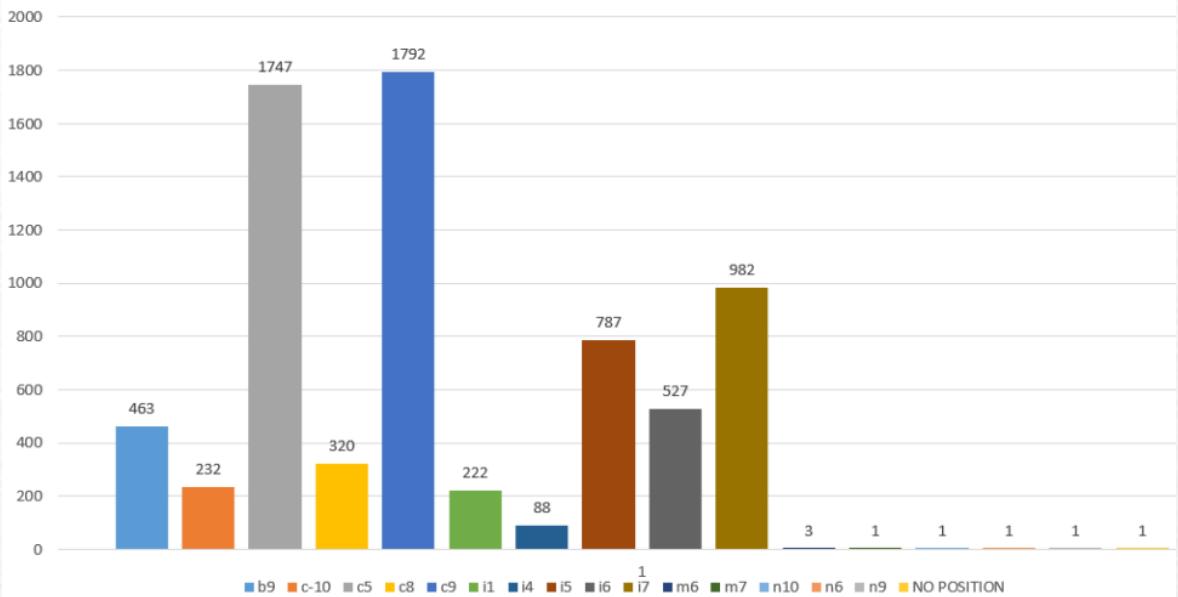
Class interval
40k-50k
Has the highest no. of people i.e. 781

D. Departmental analysis



Operations dept has the highest no. of people – 2771 which is 39% of the total people hired

E. Position tier analysis



“C9” position has the highest number of people -1792.

ANALYSIS

5 why's approach –

1. Why determine the gender distribution of hires?

→ To assess and ensure gender diversity within the organization, promoting an inclusive and balanced work environment.

2. Why is evaluating competitiveness important?

→ It helps attract and retain top talent, ensuring the company remains attractive in a competitive employment landscape.

3. Why create class intervals for the salaries in the company?

→ To organize and visualize salary data effectively, providing insights into the distribution and helping identify salary trends.

4. Why use charts and plots for visualizing departmental data?

→ Visualizations provide a clear, concise representation of complex data, making it easier to identify patterns and trends.

5. Why is understanding the distribution of positions across tiers important?

→ It provides insights into the organization's hierarchy, aids in succession planning, and identifies potential gaps in talent development.

CONCLUSION

I would like to conclude that hiring process analytics plays an important part for all the companies and firms to decide the job openings for near future.

It is done monthly , quarterly or yearly based on the needs and policies of the company.

For any company the operations department has the highest number of workforce due to the workload on the department as the department acts as a central hub for all the executive tasks carried out.

For any company there will be employees who have high salary packages than other employees due to their skills and years of experience in a particular field.

It helps the company to decide the salary for freshers joining the company also it tells requirement of workforce for each department.

IMDB MOVIE ANALYSIS

DESCRIPTION

Uncover the determinants of high IMDB ratings for movies. This investigation holds paramount importance for producers, directors, and investors, offering invaluable insights to shape informed decisions and elevate the success of upcoming projects.

PROBLEMS

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

•**Task:** Determine the most common genres of movies in the dataset. Then, for each genre, calculate descriptive statistics (mean, median, mode, range, variance, standard deviation) of the IMDB scores.

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

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

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

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

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

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

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

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

DESIGN

After downloading the dataset , creating a copy of the original dataset so that it doesn't get affected if make any changes. Before performing the tasks -

- Removing duplicate rows
- Removing rows with major column values blank
- Removing errors

Using excel functions to perform the tasks and then visualizing them in powerpoint.

FINDINGS

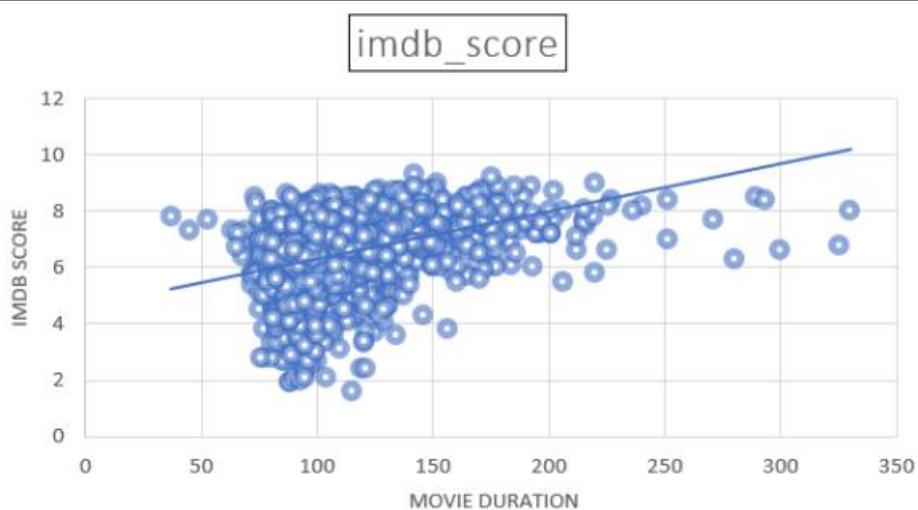
A. Movie genre analysis

GENRES	COUNT
Action	954
Documentary	53
Adventure	781
Drama	1905
Animation	198
Comedy	1481
Mystery	381
Fantasy	509
Crime	707
Biography	239
Sci-Fi	495
Horror	388
Romance	864
Thriller	1107
Game-Show	0

Family	445
Music	156
Western	57
Musical	98
Film-Noir	1
History	149
Sport	150
War	153
News	0
Reality-TV	0
Short	0
TOTAL	11271

Most Loved Genres Are Drama And Comedy

B. Relation b/w movie duration and imdb score



Avg Movie Duration Is 110.05 minutes.

Most
Common
Language
In
Movies
Is
English

Language	Count of language	MEAN
Aboriginal	2	6.95
Arabic	1	7.2
Aramaic	1	7.1
Bosnian	1	4.3
Cantonese	7	7.34286
Czech	1	7.4
Danish	3	7.9
Dari	2	7.5
Dutch	3	7.56667
English	3625	6.42568
Filipino	1	6.7
French	34	7.35588
German	10	7.77
Hebrew	2	7.65
Hindi	5	7.22
Hungarian	1	7.1
Indonesian	2	7.9
Italian	7	7.18571
Grand Total		3783

D. Impact of directors on movie's success

Director name	Average of imdb_score
Akira Kurosawa	8.7
Tony Kaye	8.6
Charles Chaplin	8.6
Alfred Hitchcock	8.5
Ron Fricke	8.5
Majid Majidi	8.5
Damien Chazelle	8.5
Sergio Leone	8.433333333
Christopher Nolan	8.425
Marius A. Markevicius	8.4
Richard Marquand	8.4
Asghar Farhadi	8.4
Grand Total	8.466666667

Akira Kurosawa
Is the
Most succesfull
Director with
The highest
Average imdb
Score of 8.7.

E. Correlation b/w movie budgets and gross earnings.

correlation	high profit margin
0.100005839	523505847

The movie with highest profit margin is “AVATAR”
Directed by “James Cameron” in the year 2009.

ANALYSIS

1. Why do movies with higher budgets tend to have higher ratings?
→ They can afford better production quality.
2. Why does better production quality lead to higher ratings?
→ It enhances the viewer's experience.
3. Why does an enhanced viewer experience lead to higher ratings?
→ Viewers are more likely to rate a movie highly if they enjoyed watching it.

4. Why are viewers more likely to rate a movie highly if they enjoyed watching it?"

→ Positive experiences lead to positive reviews.

5. Why do positive reviews matter?

→ They influence other viewers' decisions to watch the movie, increasing its popularity and success.

CONCLUSION

Ultimately, this analysis serves as a compass for navigating the intricate landscape of movie success, empowering industry professionals to craft compelling narratives, assemble effective casts, and allocate resources judiciously. As the cinematic landscape continues to evolve, leveraging these insights will be instrumental in steering the course toward creating movies that resonate and thrive on the IMDb platform.

In conclusion, IMDb movie analysis isn't just a tool for filmmakers; it's equally crucial for investors and stakeholders in the pre-production phase. Notably, the highest IMDb rating doesn't guarantee the utmost success. Understanding audience fatigue, we observe a preference for movies in the comedy/drama genre. This analysis stands as a key compass for decision-makers, guiding them in creating engaging and resonant films that transcend the boundaries of mere ratings.

BANK LOAN CASE STUDY

DESCRIPTION

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.

When a customer applies for a loan, your company faces two risks:

1. If the applicant can repay the loan but is not approved, the company loses business.
2. If the applicant cannot repay the loan and is approved, the company faces a financial loss.

PROBLEMS

A. Identify Missing Data and Deal with it Appropriately:
As a data analyst, you come across missing data in the loan application dataset. It is essential to handle missing data effectively to ensure the accuracy of the analysis.

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.

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.

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

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

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.

E. Identify Top Correlations for Different Scenarios:

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.

DESIGN

- Downloading the dataset.
- Creating a copy of the dataset so that the original dataset doesn't get affected.
- Using excel functions for performing the tasks and visualizing the data via graphs/charts.
- Using powerpoint for the final visualization.

FINDINGS

A. Handling data

→ Deleting the columns with more than 30% missing values

V	W	X	Y	Z	AA	AB	AC	AD
17050	49999	49999	49999	49999	49999	49999	34346	31.31%
65.90%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%		
OWN_	FLAG_N	FLAG_E	FLAG_V	FLAG_C	FLAG_F	FLAG_E	OCCUPATION_TYPE	CN
1	1	0	1	1	1	0	Laborers	0.00%

→ Converting days to years

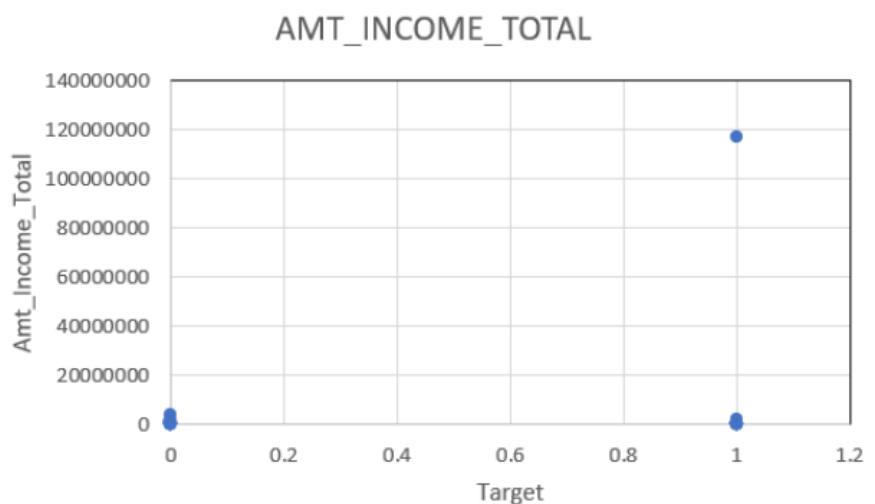
DAYS_BIRTH	DAYS_BIRTH(yrs)
-9461	26
-16765	46
-19046	52
-19005	52
-19932	55

→ Mean, median , mode imputation

J	K	L	AN	AO	M	BO	BP	BQ	BR	BS	BT
AMT_ANNUITY	AMT_G	N	EXT_SO	EXT_SO	OBS	AMT_RI	AMT_RI	AMT_RI	AMT_RI	AMT_RI	AMT_RI
24700.5	351000	Ur	0.26295	0.13938		0	0	0	0	0	1
35698.5	1129500	Fa	0.62225	0.51188		0	0	0	0	0	0
6750	135000	Ur	0.55591	0.72957		0	0	0	0	0	0
29686.5	297000	Ur	0.65044	0.51188		0	0	0	0	0	0
21865.5	513000	Ur	0.32274	0.51188		0	0	0	0	1	1
27517.5	454500	Sp	0.35422	0.62123		0	0	0	1	1	2
41301	1395000	Ur	0.724	0.49206		0	0	0	0	0	0
42075	1530000	Ur									

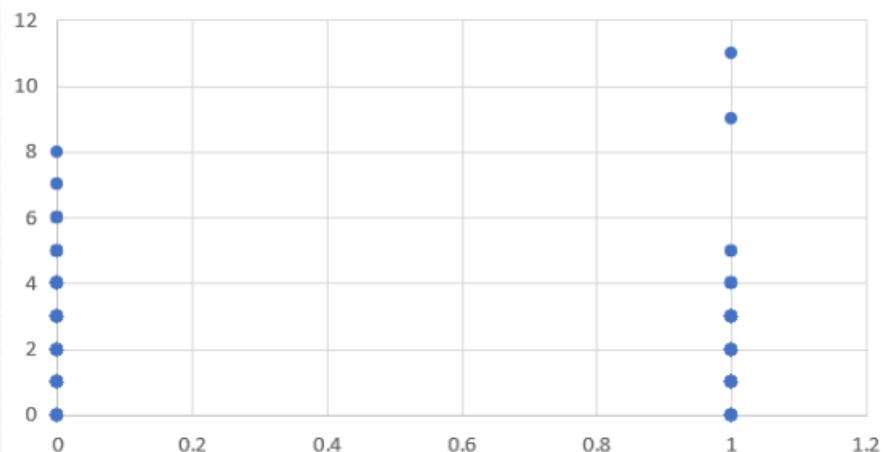
B. Finding outliers

quartile 1
112500
quartile 3
202500
interquartile range
90000
upper limit
337500
lower limit
-22500



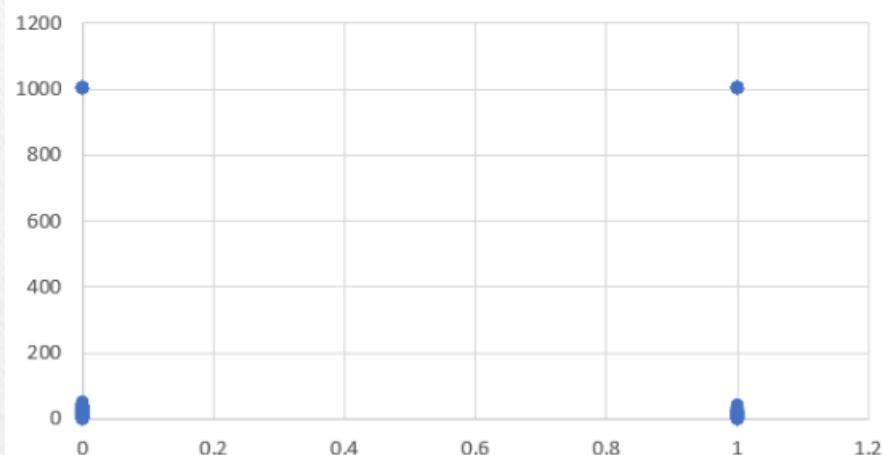
for amt_income_total and target , we can clearly identify the outlier at (1,120000000)

CNT_CHILDREN



For
cnt_children ,
we can
clearly see
outliers at
(1,9) ,(1,11).

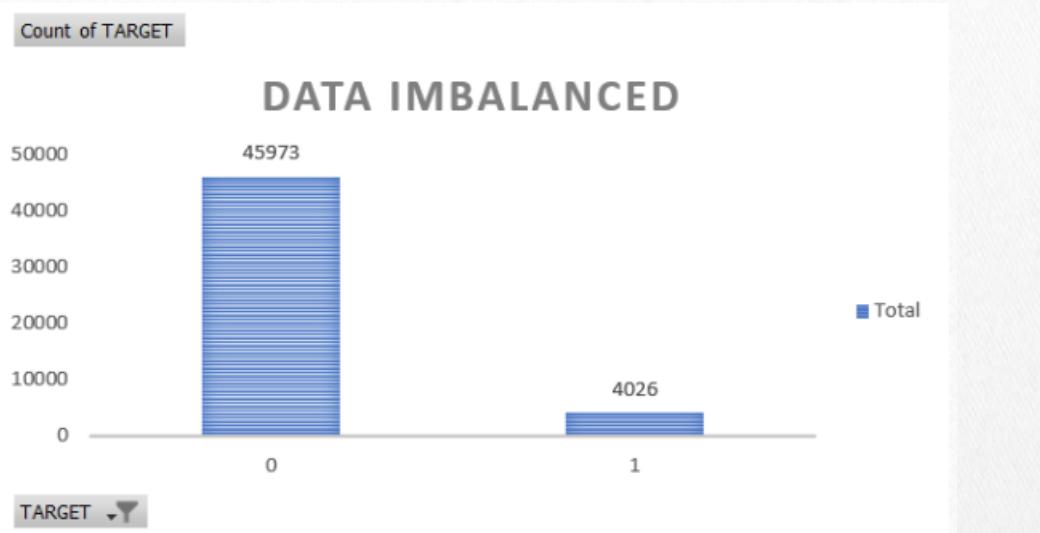
DAYS_EMPLOYED(yrs)



For days
employed ,
outliers can
be seen at
(0,1001) ,
(1,1000)

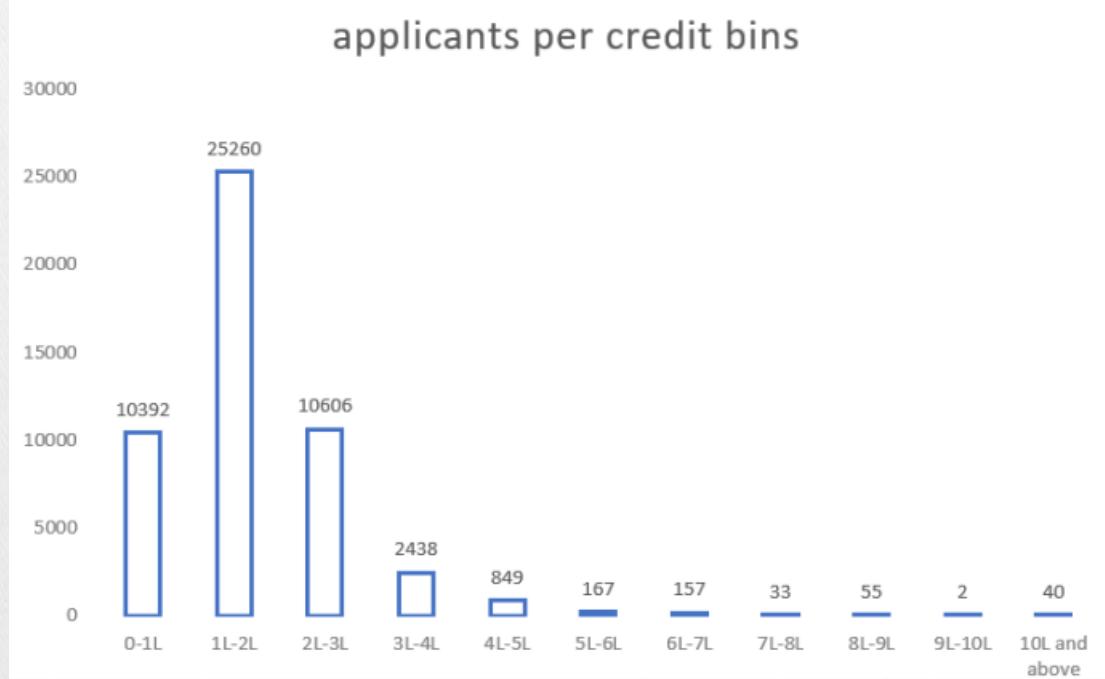
C. Data imbalance

TARGET	Count of TARGET	TARGET	CONTRIBUTION
0	45973	0	92%
1	4026	1	8%
Grand Total	49999		



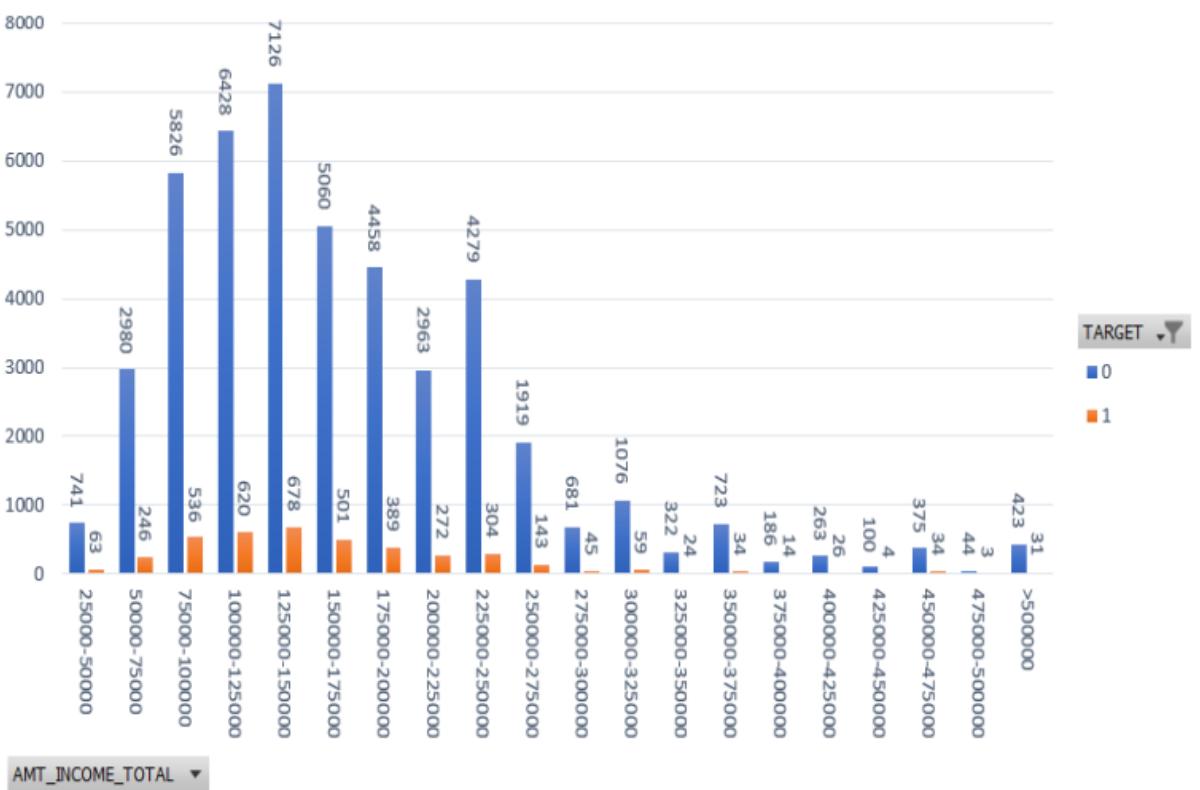
D. (i) Univariate analysis

mean	170767.5905
median	145800
mode	135000
stdev	531819.0951
min	25650
max	117000000
variance	2.828315E+11



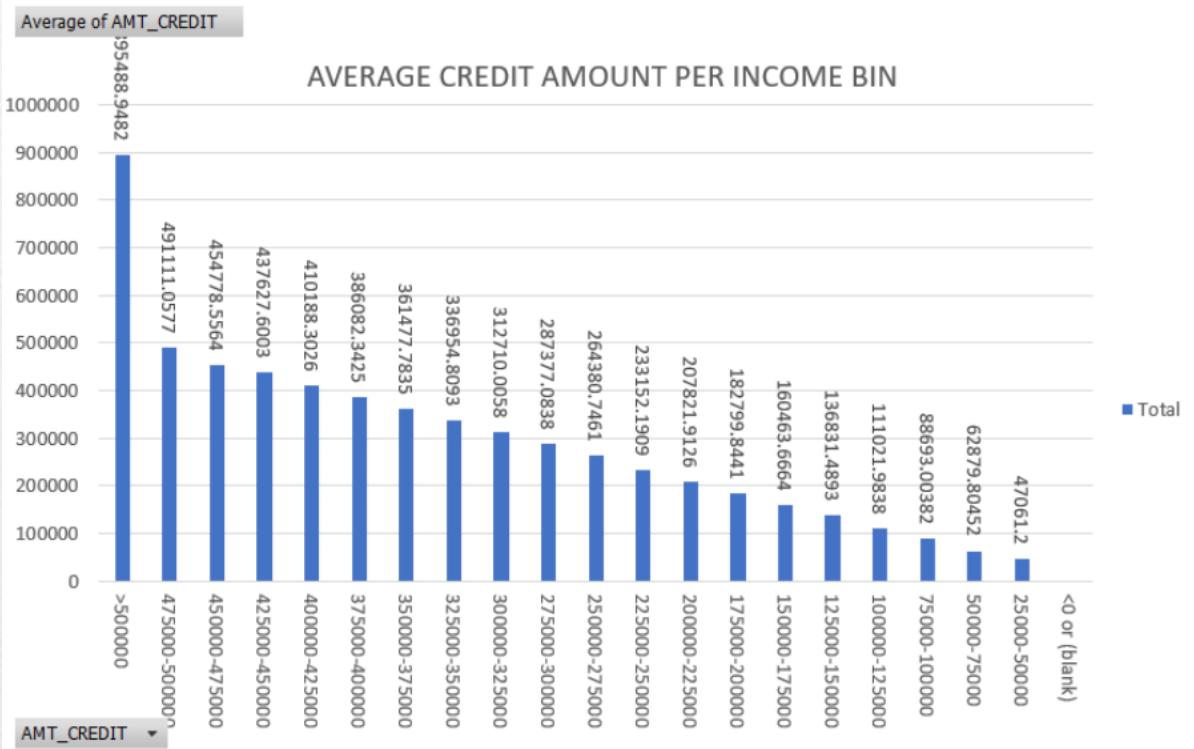
D. (ii) Segmented univariate analysis

CREDIT BINS	TARGET		Grand Total
	0	1	
25000-50000	741	63	804
50000-75000	2980	246	3226
75000-100000	5826	536	6362
100000-125000	6428	620	7048
125000-150000	7126	678	7804
150000-175000	5060	501	5561
175000-200000	4458	389	4847
200000-225000	2963	272	3235
225000-250000	4279	304	4583
250000-275000	1919	143	2062
275000-300000	681	45	726
300000-325000	1076	59	1135
325000-350000	322	24	346
350000-375000	723	34	757
375000-400000	186	14	200
400000-425000	263	26	289
425000-450000	100	4	104
450000-475000	375	34	409
475000-500000	44	3	47
>500000	423	31	454
Grand Total	45973	4026	49999



D. (iii) Bivariate analysis

CREDIT BINS	Average of AMT_CREDIT
>500000	8,95,488.95
475000-500000	4,91,111.06
450000-475000	4,54,778.56
425000-450000	4,37,627.60
400000-425000	4,10,188.30
375000-400000	3,86,082.34
350000-375000	3,61,477.78
325000-350000	3,36,954.81
300000-325000	3,12,710.01
275000-300000	2,87,377.08
250000-275000	2,64,380.75
225000-250000	2,33,152.19
200000-225000	2,07,821.91
175000-200000	1,82,799.84
150000-175000	1,60,463.67
125000-150000	1,36,831.49
100000-125000	1,11,021.98
75000-100000	88,693.00
50000-75000	62,879.80
25000-50000	47,061.20
<0 or (blank)	
Grand Total	599700.5815



Correlation with variables (TARGET = 0 i.e. all other cases)

CORRELATION FOR TARGET 0 WITH VARIABLES	
CNT_CHILDREN	1
AMT_INCOME_TOTAL	0.036
AMT_CREDIT	0.006
REGION_POPULATION_RELATIVE	-0.025
DAY_S_BIRTH(yrs)	0.378
DAY_S_EMPLOYED(yrs)	0.182
DAY_S_REGISTRATION(yrs)	0.096
DAY_S_ID_PUBLISH(yrs)	0.182
REGION_RATING_CLIENT	0.1
CNT_CHILDREN AMT_INCOME_TOTAL AMT_CREDIT REGION_POPULATION_RELATIVE DAY_S_BIRTH(yrs) DAY_S_EMPLOYED(yrs) DAY_S_REGISTRATION(yrs) DAY_S_ID_PUBLISH(yrs) REGION_RATING_CLIENT	1

Correlation with variables
(TARGET = 1 i.e. clients with payment difficulties)

CORRELATION FOR TARGET 1 WITH VARIABLES	
CNT_CHILDREN	1
AMT_INCOME_TOTAL	0.010
AMT_CREDIT	0.008
REGION_POPULATION_RELATIVE	-0.020
DAY_S_BIRTH(yrs)	0.015
DAY_S_EMPLOYED(yrs)	0.006
DAY_S_REGISTRATION(yrs)	0.015
DAY_S_ID_PUBLISH(yrs)	0.009
REGION_RATING_CLIENT	0.013
CNT_CHILDREN AMT_INCOME_TOTAL AMT_CREDIT REGION_POPULATION_RELATIVE DAY_S_BIRTH(yrs) DAY_S_EMPLOYED(yrs) DAY_S_REGISTRATION(yrs) DAY_S_ID_PUBLISH(yrs) REGION_RATING_CLIENT	1

E. Correlation with other variables

ANALYSIS

Used five Why's approach to gain more insights-

1. Why is it crucial to handle data effectively in bank loan analysis?

→ Effective data handling ensures accurate and reliable insights, preventing biases, errors, and misinterpretations in the loan assessment process.

2. Why are we finding outliers in the bank loan data?

→ To identify and understand any extreme values that may impact the overall analysis, risk assessment, and decision-making process.

3. Why address data imbalance in the context of bank loan analysis?

→ Balancing the data is vital for fair and unbiased model training, preventing skewed results that might disproportionately influence loan approval or rejection.

4. Why delve into bivariate analysis in the context of bank loan assessment?

→ Bivariate analysis explores relationships between pairs of variables, providing a comprehensive understanding of how different factors interact and influence loan approval outcomes.

5. Why is it essential to correlate loan data with other variables in the analysis?

→ Correlating loan data with other variables uncovers potential dependencies and relationships, enhancing the predictive power of the model and refining risk assessment strategies.

CONCLUSION

I would like to conclude that Bank loan analysis plays a vital role in evaluating the financial health of individuals and businesses. By understanding the types of loans, the approach to analysis, and the importance of this process, we can make informed decisions and ensure compliance. As technology continues to advance, the future of bank loan analysis holds even greater efficiency and accuracy for all.

Machine learning algorithms and artificial intelligence are poised to streamline the evaluation process, providing quicker insights and enhancing risk assessment capabilities. Moreover, the integration of big data analytics offers a more comprehensive understanding of financial behaviors, paving the way for personalized and predictive lending models.

Analyzing the Impact of Car Features on Price and Profitability

DESCRIPTION

In the ever-changing auto industry, understanding what consumers want is crucial. As a Data Analyst, the task is to help a car manufacturer optimize pricing and product development. By analyzing the dataset, we aim to uncover the ideal balance between a car's features, market category, and pricing. This involves exploring the popularity and profitability of different features and categories using techniques like regression analysis and market segmentation. The goal is to craft a pricing strategy that aligns with consumer preferences while maximizing profitability. Ultimately, these insights will guide the manufacturer to stay competitive and boost profitability in a rapidly evolving market.

PROBLEMS

Insight Required: How does the popularity of a car model vary across different market categories?

- **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.

Insight Required: What is the relationship between a car's engine power and its price?

•**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.

Insight Required: Which car features are most important in determining a car's price?

•**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.

Insight Required: How does the average price of a car vary across different manufacturers?

•**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.

Insight Required: What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

•**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.

DESIGN

Downloading the dataset.

Copying the original dataset so that the original dataset get affected while making changes.

Handling the columns with missing values.

Visualizing the tasks via graphs/charts

FINDINGS

TASKS

A. Popularity of car model vs Market categories

Row Labels	Average of Popularity	Count of Model
Crossover	1529.030825	1103
Crossover,Diesel	873	7
Crossover,Exotic,Luxury,High-Performance	238	1
Crossover,Exotic,Luxury,Performance	238	1
Crossover,Factory Tuner,Luxury,High-Performance	1823.461538	26
Crossover,Factory Tuner,Luxury,Performance	2607.4	5
Crossover,Factory Tuner,Performance	210	4
Crossover,Flex Fuel	2073.75	64
Crossover,Flex Fuel,Luxury	1173.2	10
Crossover,Flex Fuel,Luxury,Performance	1624	6
Crossover,Flex Fuel,Performance	5657	6
Crossover,Hatchback	1675.694444	72
Crossover,Hatchback,Factory Tuner,Performance	2009	6
Crossover,Hatchback,Luxury	204	7
Crossover,Hatchback,Performance	2009	6
Crossover,Hybrid	2563.380952	42
Crossover,Luxury	884.5487805	410
Crossover,Luxury,Diesel	2195.848485	33
Crossover,Luxury,High-Performance	1037.222222	9
Crossover,Luxury,Hybrid	630.9166667	24
Crossover,Luxury,Performance	1344.849558	113
Crossover,Luxury,Performance,Hybrid	3916	2
Crossover,Performance	2585.956522	69
Diesel	1730.904762	84

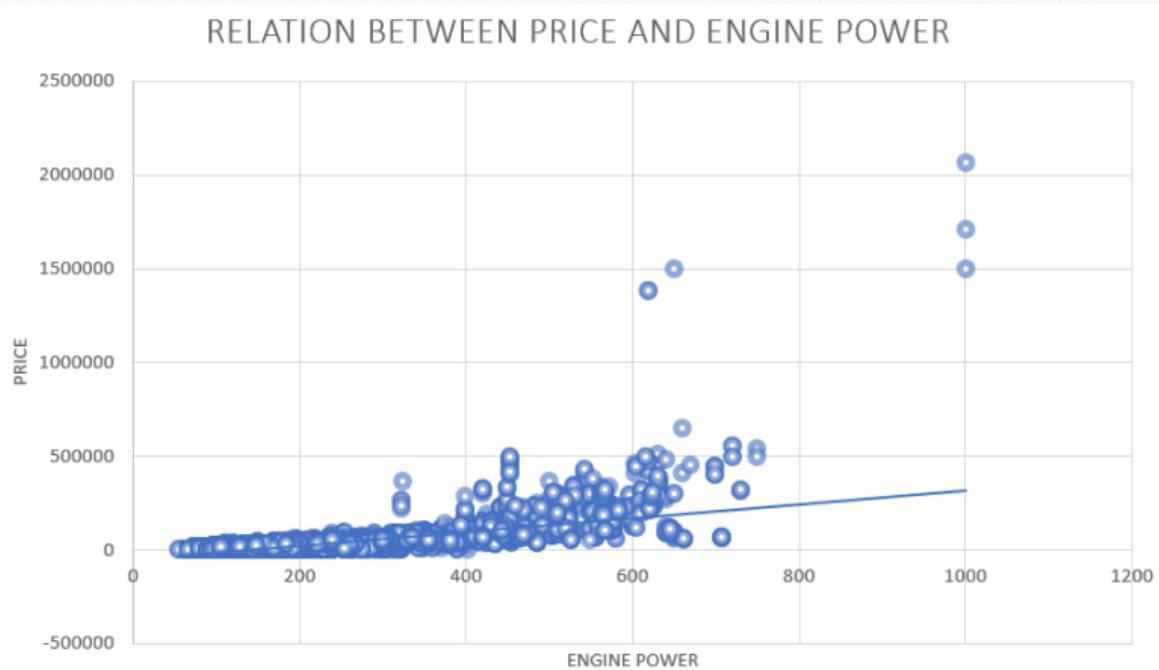
Diesel,Luxury	2275	51
Exotic,Factory Tuner,High-Performance	1046.380952	21
Exotic,Factory Tuner,Luxury,High-Performance	517.5384615	52
Exotic,Factory Tuner,Luxury,Performance	520	3
Exotic,Flex Fuel,Factory Tuner,Luxury,High-Performance	520	13
Exotic,Flex Fuel,Luxury,High-Performance	520	11
Exotic,High-Performance	1261.571429	252
Exotic,Luxury	112.6666667	12
Exotic,Luxury,High-Performance	467.0759494	79
Exotic,Luxury,High-Performance,Hybrid	204	1
Exotic,Luxury,Performance	217.0277778	36
Factory Tuner,High-Performance	1941.415094	106
Factory Tuner,Luxury	617	2
Factory Tuner,Luxury,High-Performance	2133.367442	215
Factory Tuner,Luxury,Performance	1413.419355	31
Factory Tuner,Performance	1733.101124	89
Flex Fuel	2217.302752	872
Flex Fuel,Diesel	5657	16
Flex Fuel,Factory Tuner,Luxury,High-Performance	258	1
Flex Fuel,Hybrid	155	2
Flex Fuel,Luxury	746.5384615	39
Flex Fuel,Luxury,High-Performance	878.9090909	33
Flex Fuel,Luxury,Performance	1380.071429	28
Flex Fuel,Performance	1702.358025	81
Flex Fuel,Performance,Hybrid	155	2

Hatchback	1292.998371
Hatchback,Diesel	873
Hatchback,Factory Tuner,High-Performance	1205.153846
Hatchback,Factory Tuner,Luxury,Performance	886.8888889
Hatchback,Factory Tuner,Performance	2159.045455
Hatchback,Flex Fuel	5657
Hatchback,Hybrid	2121.25
Hatchback,Luxury	1379.5
Hatchback,Luxury,Hybrid	454
Hatchback,Luxury,Performance	1566.131579
Hatchback,Performance	1039.646825
High-Performance	1821.447236
Hybrid	2105.569106
Luxury	1107.553467
Luxury,High-Performance	1668.017964
Luxury,High-Performance,Hybrid	568.8333333
Luxury,Hybrid	724.6875
Luxury,Performance	1292.615156
Luxury,Performance,Hybrid	2333.181818
N/A	1671.388144
Performance	1371.080479
Performance,Hybrid	155
(blank)	
Grand Total	1553.679902

Row Labels
 Crossover,Flex Fuel,Performance
 Flex Fuel,Diesel
 Hatchback,Flex Fuel

These market categories have highest avg popularity of 5657.

B. Relation between price and engine power

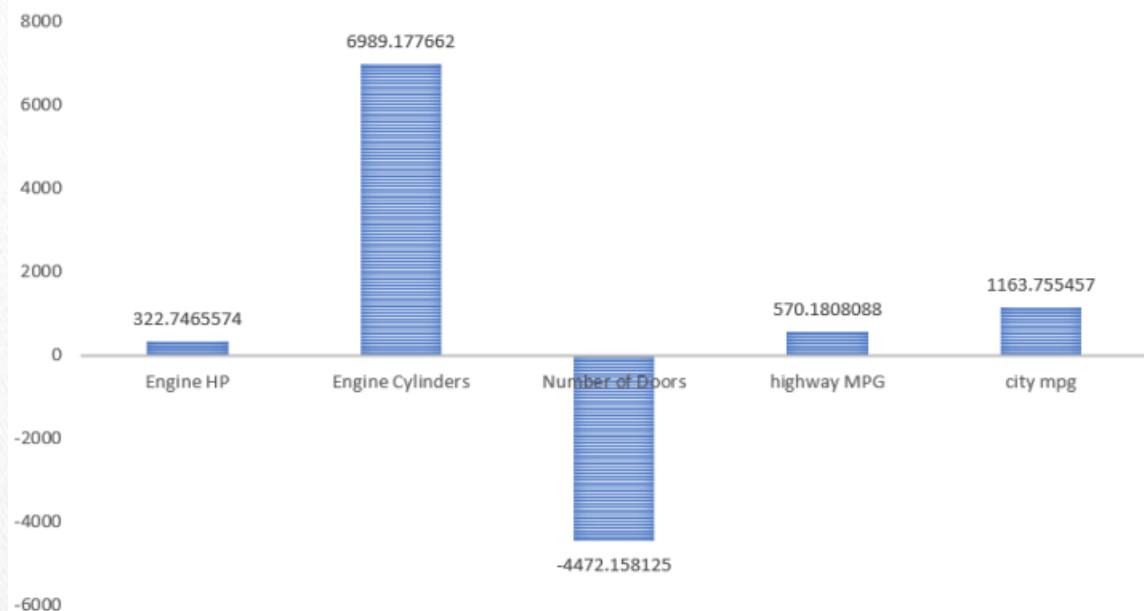


As engine power increases , price increases.

C. Regression analysis

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.680708							
R Square	0.463364							
Adjusted R	0.463136							
Standard E	44170.78							
Observatio	11812							
ANOVA								
	df	SS	MS	F	Significance F			
Regression	5	1.99E+13	3.98E+12	2038.799	0			
Residual	11806	2.3E+13	1.95E+09					
Total	11811	4.29E+13						
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95% Upper 95%	Lower 95% Upper 95%
Intercept	-101602	3684.352	-27.5766	2.8E-162	-108824	-94379.8	-108824	-94379.8
Engine HP	322.7466	6.017674	53.63311	0	310.9509	334.5422	310.9509	334.5422
Engine Cyls	6989.178	439.645	15.89732	2.54E-56	6127.401	7850.954	6127.401	7850.954
Number of	-4472.16	465.7181	-9.60272	9.35E-22	-5385.04	-3559.27	-5385.04	-3559.27
highway M	570.1808	105.784	5.390049	7.18E-08	362.8268	777.5349	362.8268	777.5349
city mpg	1163.755	121.9978	9.53915	1.72E-21	924.6196	1402.891	924.6196	1402.891

COEFFICIENT OF VARIABLES

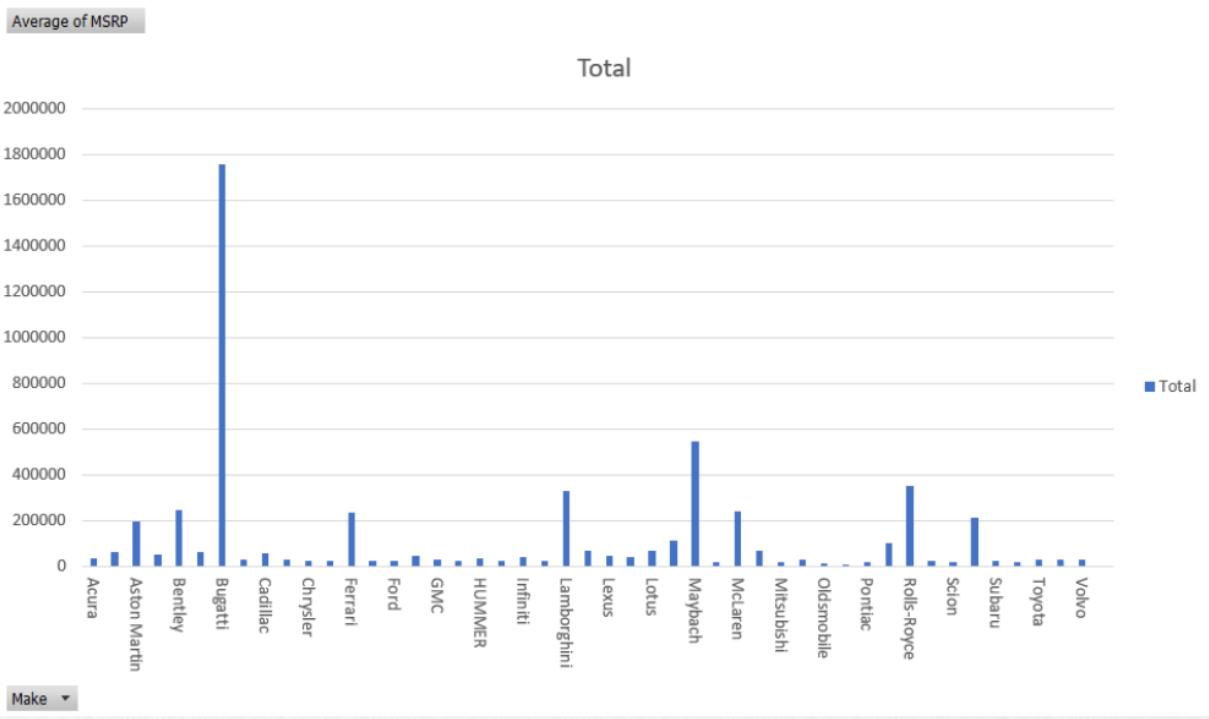


D. Avg of price vs Manufacturer

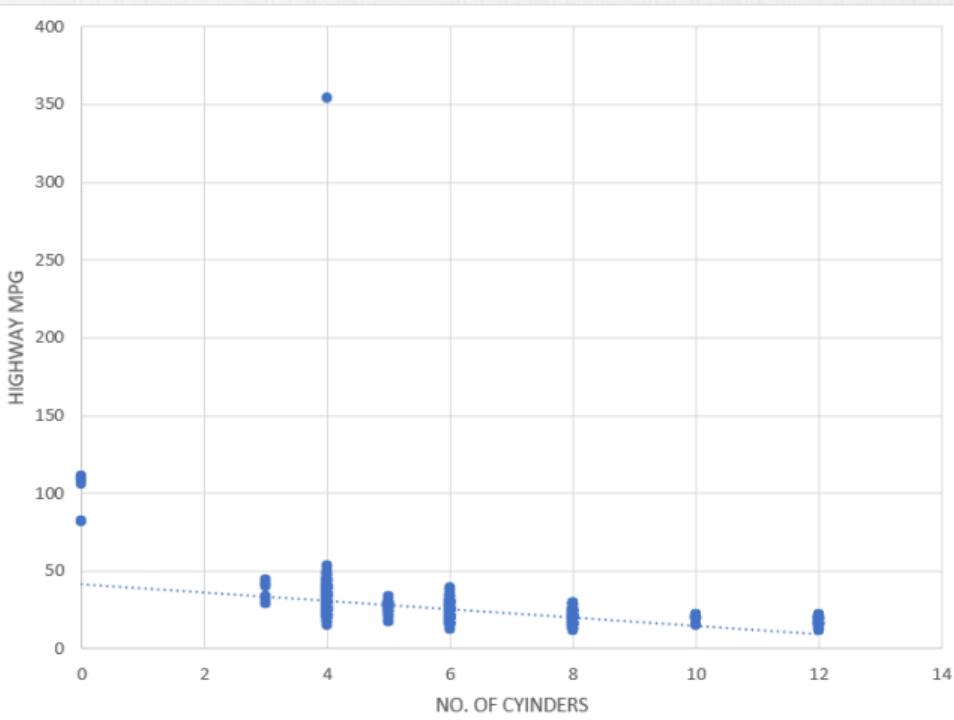
Manufacturer	Average of MSRP		
Acura	34887.5873	Lexus	47549.06931
Alfa Romeo	61600	Lincoln	42494.37179
Aston Martin	197910.3763	Lotus	69188.27586
Audi	53452.1128	Maserati	114207.7069
Bentley	247169.3243	Maybach	546221.875
<u>BMW</u>	<u>61546.76347</u>	Mazda	19719.05707
Bugatti	1757223.667	McLaren	239805
Buick	28206.61224	Mercedes-Benz	71537.80966
Cadillac	56231.31738	Mitsubishi	21215.47143
Chevrolet	28273.35695	Nissan	28513.36679
Chrysler	26722.96257	Oldsmobile	11542.54
Dodge	22390.05911	Plymouth	3122.902439
Ferrari	237383.8235	Pontiac	19321.54839
FIAT	22206.01695	Porsche	101622.3971
Ford	27393.42051	Rolls-Royce	351130.6452
Genesis	46616.66667	Saab	27413.5045
GMC	30493.29903	Scion	19932.5
Honda	26629.81879	Spyker	213323.3333
HUMMER	36464.41176	Subaru	24827.50391
Hyundai	24597.0363	Suzuki	17900.9569
Infiniti	42394.21212	Toyota	28946.15343
Kia	25112.38938	Volkswagen	28076.2
Lamborghini	331567.3077	Volvo	28541.16014
Land Rover	67823.21678	(blank)	
Grand Total		40559.93532	

Highest average of MSRP = 1757223.66
OF BUGGATI

Lowest average of MSRP = 3122.9
OF PLYMOUTH



5. Fuel efficiency vs no. of cylinders



ANALYSIS

Used Five Why's approach –

1. Why analyze the popularity of car models within market categories?

→ To understand consumer preferences and buying patterns, helping manufacturers tailor their offerings to align with market demands.

2. Why investigate the relationship between price and engine power?

→ To identify if there's a correlation between these factors, helping consumers make informed decisions and assisting manufacturers in pricing strategies.

3. Why calculate the coefficients of variables in regression analysis?

→ Coefficients quantify the strength and direction of relationships, providing a precise understanding of how each variable influences the model.

4. Why analyze the average price of cars based on manufacturers?

→ To comprehend how pricing varies across manufacturers, aiding consumers in brand choices and guiding manufacturers in pricing strategies.

5. Why is knowing the impact of the number of cylinders on fuel efficiency important for consumers?

→ It enables consumers to make environmentally conscious choices and optimize fuel costs based on their driving needs.

CONCLUSION

After analyzing the impact of various car features on price and profitability, we can conclude that certain features play a significant role in determining these factors. Our study provides valuable insights for car manufacturers and industry professionals looking to make informed decisions regarding vehicle pricing and profitability.

Our research found that features such as fuel efficiency, safety ratings, and luxury amenities have a significant impact on the price of a car. Additionally, these features can also influence the profitability of a vehicle for a manufacturer.

For example, our analysis of fuel efficiency features revealed that higher gas mileage ratings can increase both the price and profitability of a car. This is because consumers are willing to pay more for cars that offer better fuel economy, while manufacturers can benefit from lower production costs and increased demand for fuel-efficient vehicles.

ABC CAR VOLUME TREND ANALYSIS

DESCRIPTION

In this project, you'll be diving into the world of Customer Experience (CX) analytics, specifically focusing on the inbound calling team of a company. You'll be provided with a dataset that spans 23 days and includes various details such as the agent's name and ID, the queue time (how long a customer had to wait before connecting with an agent), the time of the call, the duration of the call, and the call status (whether it was abandoned, answered, or transferred).

A Customer Experience (CX) team plays a crucial role in a company. They analyze customer feedback and data, derive insights from it, and share these insights with the rest of the organization. This team is responsible for a wide range of tasks, including managing customer experience programs, handling internal communications, mapping customer journeys, and managing customer data, among others.

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.

PROBLEMS

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

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 (e.g., 1-2, 2-3, etc.). **Your Task:** Can you create a chart or graph that shows the number of calls received in each time bucket?

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

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

DESIGN

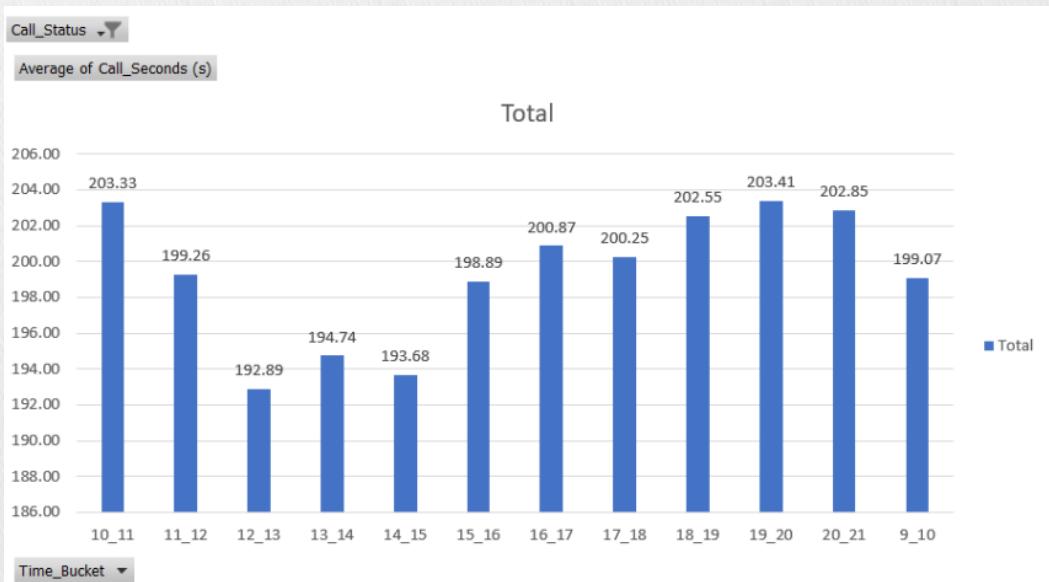
- Downloading the dataset.
- Performing the tasks using excel functions and visualizing them using graphs/charts.
- Final report is visualized using powerpoint.

FINDINGS

A. Avg call duration

7-8 pm has highest average Call_seconds while 12-1 pm has the lowest.

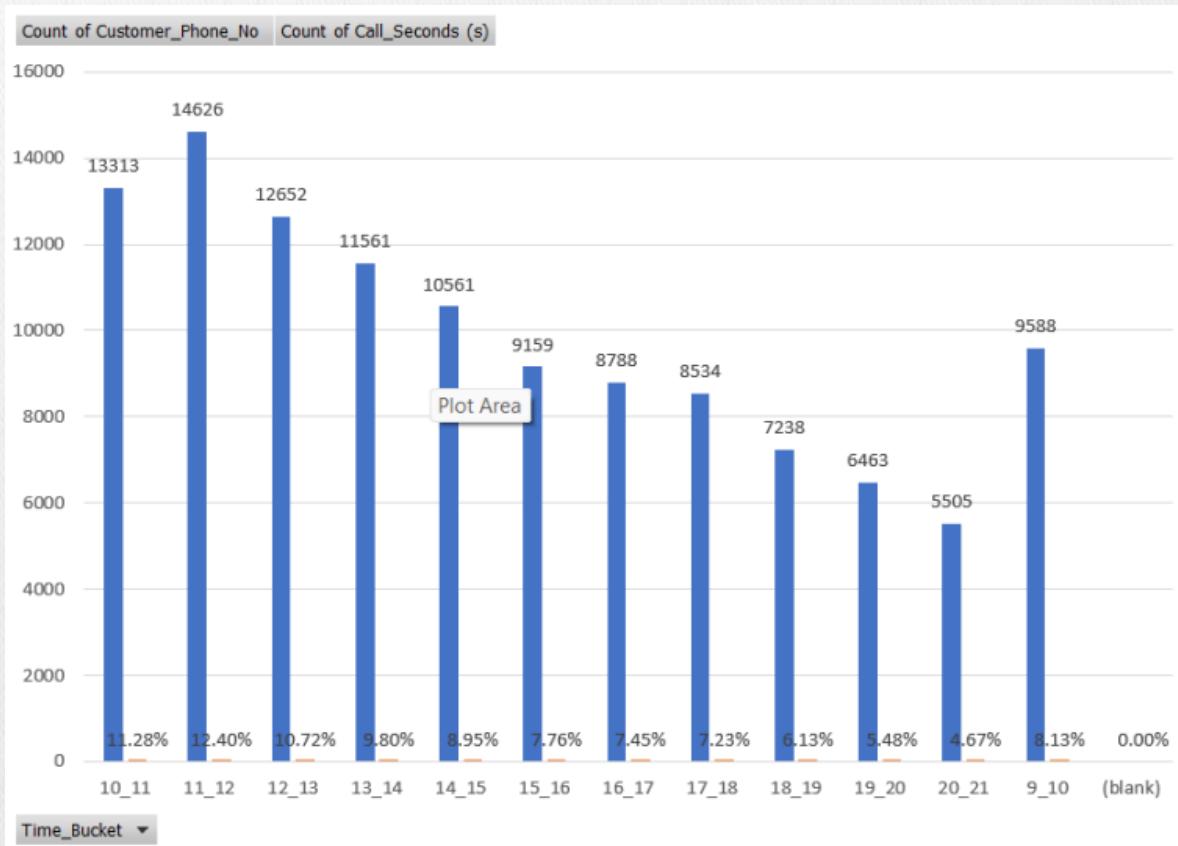
Time_Bucket	Average of Call_Seconds (s)
10_11	203.33
11_12	199.26
12_13	192.89
13_14	194.74
14_15	193.68
15_16	198.89
16_17	200.87
17_18	200.25
18_19	202.55
19_20	203.41
20_21	202.85
9_10	199.07
Grand Total	198.62



B. Call volume analysis

Time_Bucket	Count of Customer_Phone_No	Count of Call_Seconds (s)
10_11	13313	11.28%
11_12	14626	12.40%
12_13	12652	10.72%
13_14	11561	9.80%
14_15	10561	8.95%
15_16	9159	7.76%
16_17	8788	7.45%
17_18	8534	7.23%
18_19	7238	6.13%
19_20	6463	5.48%
20_21	5505	4.67%
9_10	9588	8.13%
(blank)		0.00%
Grand Total	117988	100.00%

Highest number of calls are made during 11am to 12 pm while lowest number of calls are made during 8-9pm.

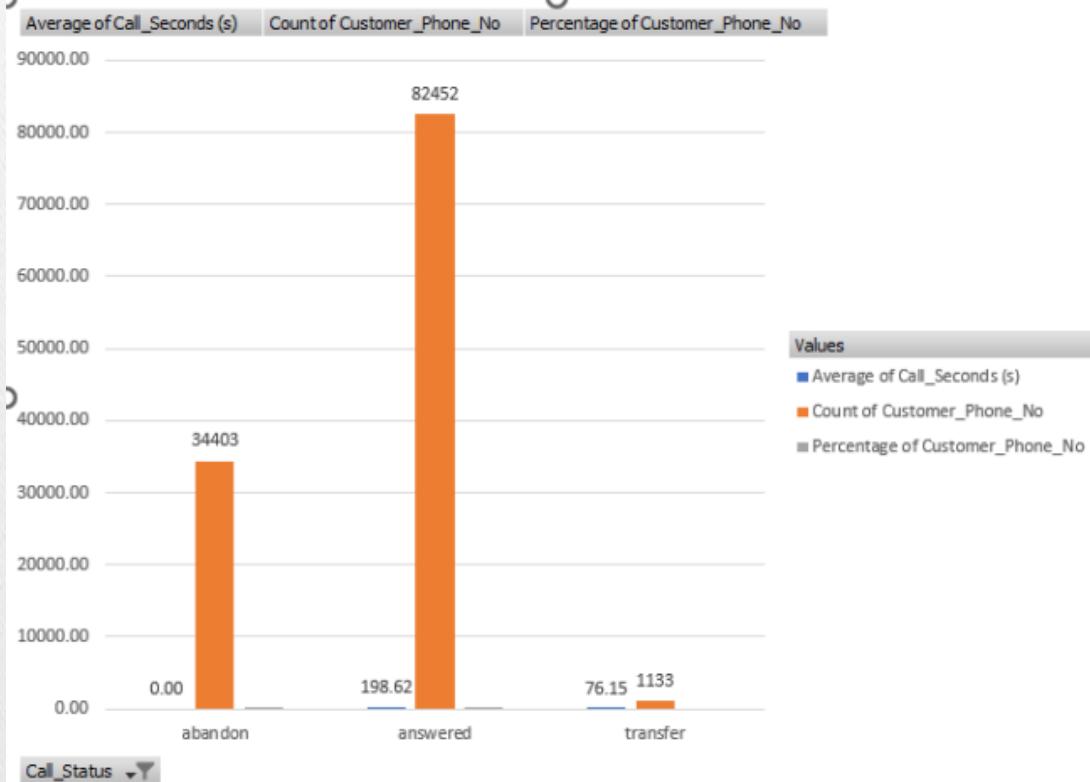


C. Manpower planning

work days	6
unplanned leaves	4
work_days_in_a_month	20
duty hours	9
lunch break	1.5
actual working hrs	4.5
total days in a month	30

31% calls are abandoned on an average

Call_Status	Average of Call_Seconds (s)	Count of Customer_Phone_No	Percentage of Customer_Phone_No
abandon	0.00	34403	29.16%
answered	198.62	82452	69.88%
transfer	76.15	1133	0.96%
Grand Total	139.5321473	117988	100.00%



Row Labels	Count of Call_Seconds (s)	Count of Call_Seconds (s)2	Agents Required
10_11	11.89%	0.12	7
11_12	10.49%	0.10	6
12_13	10.10%	0.10	6
13_14	8.03%	0.08	4
14_15	9.32%	0.09	5
15_16	7.90%	0.08	4
16_17	7.13%	0.07	4
17_18	8.01%	0.08	4
18_19	7.26%	0.07	4
19_20	5.06%	0.05	3
20_21	7.08%	0.07	4
9_10	7.73%	0.08	5
Grand Total	100.00%	1.00	56

D. Night shift manpower planning

average calls in a day	additional hrs required
5130	76
calls between 9pm-9am	agents required
1539	15

Time bucket	call distribution	time distribution	agents required
9_10	3	10.00	2
10_11	3	10.00	2
11_12	2	15.00	1
12_1	2	15.00	1
1_2	1	30.00	1
2_3	1	30.00	1
3_4	1	30.00	1
4_5	1	30.00	1
5_6	3	10.00	2
6_7	4	7.50	2
7_8	4	7.50	2
8_9	5	6.00	3
TOTAL	30	201.00	15

ANALYSIS

Used five Why's approach to gain more insights –

1. Why analyze the average call duration in the "ABC Call Volume Trend Analysis" project?

→ To understand the typical duration of calls, which helps in assessing customer engagement and optimizing resources based on expected call lengths.

2. Why conduct a call volume analysis in the "ABC Call Volume Trend Analysis" project?

→ To unravel patterns and trends in call volume, enabling proactive management strategies to handle peak times and enhance overall service efficiency.

3. Why is manpower planning a key component of the "ABC Call Volume Trend Analysis" project?

→ Manpower planning ensures the right number of agents are available to handle incoming calls, aligning with service level objectives and maintaining operational efficiency.

4. Why specifically plan for manpower during night shifts in the "ABC Call Volume Trend Analysis" project?

→ Night shifts often have different call volume patterns; planning ensures adequate staffing to meet nighttime service demands.

5. Why is dedicated manpower planning for night shifts important for the project's objectives?

→ It ensures 24/7 service availability, meets customer expectations, and addresses potential challenges unique to nighttime operations, such as fatigue and reduced staff availability.

CONCLUSION

In conclusion, the ABC Call Volume Trend Analysis is a crucial tool for understanding and analyzing call volume patterns. By examining the trends and fluctuations in call volume over a specified period, organizations can gain valuable insights that can inform their strategic decisions.

One of the key benefits of conducting a call volume trend analysis is the ability to identify patterns and forecast future volumes. By analyzing historical data, organizations can identify recurring trends and use them to predict future call volumes. This forecasting capability can help organizations allocate resources effectively, ensuring that they have the right number of agents available to handle incoming calls.

Furthermore, the ABC Call Volume Trend Analysis can provide insights into the effectiveness of marketing campaigns and customer service initiatives. By examining the relationship between call volume and specific events or activities, organizations can determine the impact of their

strategies on call volume. This information can guide future marketing and customer service efforts, helping organizations improve their overall performance.

Another important aspect of the ABC Call Volume Trend Analysis is its role in identifying anomalies or outliers. By examining the data for unusual spikes or drops in call volume, organizations can investigate the underlying causes and take appropriate action. For example, if a sudden increase in call volume is observed, it may indicate a problem with a product or service that needs to be addressed promptly.

In summary, the ABC Call Volume Trend Analysis is a powerful tool for understanding call volume patterns, forecasting future volumes, and making strategic decisions. By conducting this analysis, organizations can gain valuable insights that can inform their resource allocation, marketing strategies, and customer service initiatives. It is an essential component of any comprehensive call center analytics strategy.

APPENDIX

DATA ANALYTICS PROCESS

- Link for the shared pdf -
<https://drive.google.com/file/d/1EhacMMPIidWqTlTZOzc8tXLPrem2VDMB/view?usp=drivesdk>

INSTAGRAM USER ANALYTICS

- Link for the shared pdf for tasks performed -
<https://drive.google.com/file/d/1cv9gT2VZ6g7rv0-D5iW1IgZ0tJnOK3Al/view?usp=drivesdk>
- Link for the shared sql query file -
https://drive.google.com/file/d/1q3eMy3IBEKDpIJf5YE_kxru38yYjlDC1X/view?usp=drivesdk

OPERATION AND METRIC ANALYTICS

- Link for the shared pdf for tasks performed –
https://drive.google.com/file/d/1MxTUPWpj1DcP_dGAjVhmxFggjDXlsN18/view?usp=drivesdk
- Link for sql query file -
https://drive.google.com/file/d/1b9muGWELwz_bX852s_oJ0g66XnQje0bX/view?usp=drivesdk

HIRING PROCESS ANALYTICS

- Link for the shared pdf for tasks performed –
<https://drive.google.com/drive/folders/1-jNkcFGCUGXSejHGTNajLMIqw3SfaeQ1>
- Link for the excel –
https://docs.google.com/file/d/1OUFQP8H3FK_yZ9R8wiXYBUn4D7On0x3G/edit?usp=docslist_api&filetype=msexcel

IMDB MOVIE ANALYSIS

- Link for the shared pdf for tasks performed –
<https://drive.google.com/file/d/1myANorlQmBPcSmZpb2kc4BrB1FluZKE/view?usp=drivesdk>

BANK LOAN ANALYSIS

- Link for the shared pdf for tasks performed –
<https://drive.google.com/file/d/1g5GmB2CUGwj8OFxA1urKO8KN-TjcSqPl/view?usp=drivesdk>
- Link for the excel –
https://docs.google.com/file/d/1mEfyoFLgmvBe5BaURuuGroWk3zKLzg8L/edit?usp=docslist_api&filetype=msecel

IMPACT OF CAR FEATURES -

- Link for the shared pdf for tasks performed –
<https://drive.google.com/file/d/14yKbs4PHuXFRTpcLLhKYnOGpBM7ciuWh/view?usp=drivesdk>
- Link for the excel –
https://drive.google.com/drive/folders/10-CwlzFfnAUpsXXjz_UH5tz_yB9UjJo6

ABC CALL VOLUME TREND -

- Link for the shared pdf for tasks performed –
<https://drive.google.com/file/d/1vpRJ36Vhmxad0LOeZKJdfuGr5jIFzXDb/view?usp=drivesdk>
- Link for the excel –
<https://drive.google.com/drive/folders/1062z8AuIxs778up6oS6PBWGeTvUqx0CK>