DATA ANALYSIS PORTFOLIO

Prepared by

Drustant Metar

PROFESSIONAL BACKGROUND

Dynamic and highly motivated, I hold a Master's degree in Information Technology with an exceptional 9.5 CGPA, complemented by a Bachelor's degree in Computer Science with an 8.71 CGPA. Adept at academic research, I have authored two papers published in "The International Research Journal of Modernization in Engineering Technology and Science," focusing on the "evolution of Intel processor generations “and "a systematic literature review of voice assistant systems".

In my academic pursuits, I have developed practical expertise through projects such as a Health Care Mobile Application, utilizing JAVA, XML, and PHP MySQL, as well as a Dora-Virtual Assistant using Python. These projects demonstrate my proficiency in software development and application across various platforms.

During my internship at Trainity, I honed my analytical skills through projects in SQL, Excel, and Machine Learning, showcasing my ability to derive valuable insights from data. As a fresher entering the corporate realm, I am eager to apply my theoretical knowledge to real-world challenges, leveraging my adaptability and enthusiasm for continuous learning to contribute meaningfully to organizational success.

With a solid foundation in technology, a passion for innovation, and a commitment to professional growth, I am poised to make significant contributions in the field of Information Technology and Data Analytics.

TABLE OF CONTENTS

Instagram User Analytics

Description:

The project's main objective is to get insights for investment metrics and marketing tactics by analyzing data from a platform that resembles Instagram. It entails doing things like finding loyal followers, interacting with inactive people, announcing contest winners, looking for hashtags, and spotting possible bots or fraudulent accounts. The purpose of these assignments is to offer insightful data about platform engagement, user behavior, and authenticity.

Problem:

Working with Instagram's product team, the product manager has requested that you provide your thoughts on the queries posed by the management group.

Design:

Steps taken to clean the data:

1. Importing the supplied datasets.
2. Combining them, eliminating blank cells and duplicates.
3. Enhancing the column headings with accurate values, locating and substituting accurate values for the data, and using SQL workbench as a visualization tool.

Findings:

A] Marketing Analysis

1. Loyal User Reward:

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

Solution:

Query:

```
select * from ig_clone.users order by created_at asc limit 5;
```

Query Result:

id	username	created_at
180	Darby_Herzog	2016-05-06 00:14:21
80	Darby_Herzog	2016-05-06 00:14:21
167	Emilio_Bernier52	2016-05-06 13:04:30
67	Emilio_Bernier52	2016-05-06 13:04:30
63	Elenor88	2016-05-08 01:30:41
NULL	NULL	NULL

The usernames and creation dates of the five users who have been on the site the longest will be chosen by this query and arranged in ascending order based on their creation dates.

2. Inactive User Engagement

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

Solution:

Query:

```
SELECT id, username FROM users WHERE id NOT IN (SELECT DISTINCT user_id FROM photos);
```

Query Result:

	id	username
▶	5	Aniya_Hackett
	7	Kasandra_Homenick
	14	Jaclyn81
	21	Rocio33

The users from the "users" table whose IDs and usernames do not have corresponding entries in the "photos" table a sign that they have never uploaded a photo to Instagram are chosen by this query.

To get the list of inactive users who have never posted any images, run this SQL query on your database. After that, you can utilize this data to send them emails with promotions to get them to start posting.

3. Contest Winner Declaration

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

Solution:

Query:

```
SELECT u.id AS user_id, u.username, p.id AS photo_id, COUNT(l.user_id) AS like_count FROM users u
JOIN photos p ON u.id = p.user_id LEFT JOIN likes l ON p.id = l.photo_id GROUP BY u.id, u.username, p.id
ORDER BY like_count DESC LIMIT 1;
```

Query Result:

	user_id	username	photo_id	like_count
▶	52	Zack_Kemmer93	145	48

The "users", "photos", and "likes" tables are joined in this query to determine the number of likes for each user's photo. The user with the most likes on a particular photo is represented by the top row, which is selected after the result determined by the like count is sorted in descending order. Run this SQL query over your database to find the contest winner and give the team their information.

4. Hashtag Research

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

Solution:

Query:

```
SELECT tag_name, COUNT(*) AS tag_count FROM tags JOIN photo_tags ON tags.id = photo_tags.tag_id
GROUP BY tag_name ORDER BY tag_count DESC LIMIT 5;
```

Query Result:

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

To count the likes for each photo, this query first joins the "photos" and "users" tables with a left join to the "likes" table. The results are grouped by user and photo, and then they are arranged in descending order of like count. The person with the most likes on a particular photo is represented by the top row, which is the final selection.

Run this SQL query over your database to find the contest winner and give the team their information.

5. Ad Campaign Launch

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

Solution:

Query:

```
SELECT DAYNAME(created_at) AS registration_day, COUNT(*) AS registration_count FROM users
GROUP BY registration_day ORDER BY registration_count DESC LIMIT 1;
```

Query Result:

registration_day	registration_count
Thursday	32

This query counts the number of registrations for each day and extracts the day of the week (using the DAYNAME function) from the created_at column of the user's table. The results are then grouped by day of the week and arranged in descending order of registration count. Lastly, it restricts the output to the day with the most registrations on top.

Run this SQL query on your database to find out which day of the week the majority of Instagram users sign up. With this information, the team can better plan when to run their advertising campaign to get the most exposure and interaction.

B] Investor Metrics

1. User Engagement

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.

Solution:

Average number of posts per user:

Query:

```
SELECT COUNT(*) / COUNT(DISTINCT user_id) AS avg_posts_per_user FROM photos;
```

Query Result:

	avg_posts_per_user
▶	3.4730

The average number of postings per user is calculated by taking the total number of photos and dividing it by the total number of unique user IDs.
Total number of photos divided by the total number of users:

Query:

```
SELECT COUNT(*) AS total_photos, COUNT(DISTINCT user_id) AS total_users, COUNT(*) / COUNT(DISTINCT user_id) AS photos_per_user FROM photos;
```

Query Result:

	total_photos	total_users	photos_per_user
▶	257	74	3.4730

To find the average number of photographs per user, this query counts the total number of photos and different user IDs. It then divides the total number of photos by the total number of users.

Run these SQL queries via your database to get information about Instagram user engagement. Investors can use these indicators to determine whether users are still using the site and making regular posts.

2. Bots & Fake Accounts

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

Solution:

Query:

```
SELECT user_id FROM ( SELECT user_id, COUNT(DISTINCT photo_id) AS total_photos_liked FROM likes GROUP BY user_id ) AS user_likes
WHERE total_photos_liked = (SELECT COUNT(*) FROM photos);
```

Query Result:

	user_id
▶	5
	14
	21

The number of unique photographs that each user has liked is first determined by this query, which then chooses user IDs that have liked the same number of unique photos as there are

overall on the website. Run your database through this SQL query to find any possible bots or fake accounts that have liked every single picture on the website. In order to ascertain whether these users are in fact bots or fraudulent accounts, more research into them may be necessary.

Analysis:

key Insights:

Marketing Analysis: To reward loyalty, the platform's five oldest users were identified. Identified inactive individuals for targeted engagement who have never uploaded a single photo. Found the winner of the contest whose photo had the most likes. Recommended the top five most popular hashtags to maximize reach. Based on patterns in user registration, identified the optimal day of the week to start advertising campaigns.

Investor Metrics: To determine user engagement, the average number of posts made by each user was calculated. Based on unusual like behavior, suspected bots or false accounts were identified.

Conclusion:

In conclusion, the project has successfully achieved its primary objective of gaining valuable insights into investment metrics and marketing tactics through a comprehensive analysis of data collected from a platform resembling Instagram. By focusing on key activities such as identifying loyal followers, engaging with inactive users, announcing contest winners, monitoring hashtag trends, and detecting potential bots or fraudulent accounts, the project has provided valuable data regarding platform engagement, user behavior, and authenticity.

Through the execution of these assignments, significant patterns and trends have emerged, shedding light on user preferences, interaction patterns, and the overall health of the platform's community. These insights can now inform strategic decision-making processes related to investment allocation, marketing strategies, and community management, ultimately driving growth and sustainability.

Moving forward, the project sets the stage for continued refinement and optimization of investment and marketing initiatives, leveraging the newfound understanding of user behavior and engagement dynamics. By continually monitoring and adapting to changes in the platform landscape, stakeholders can ensure the ongoing success and relevance of their efforts in an ever-evolving digital ecosystem.

Hiring Process Analytics Description

Description:

I took on the responsibility of examining our employment process data as a data analyst for our global organization in order to derive useful insights. The dataset offered a wealth of analytical data by supplying records of prior employment.

I started by carefully reviewing any missing data to make sure the information was reliable and comprehensive. I then combined pertinent categories inside columns to condense the dataset and make further analysis easier. In order to preserve the integrity of the study, outliers were then located and dealt with appropriately.

Ultimately, I provided an overview of the results using statistical analysis and visual aids, illuminating a number of facets of our hiring procedure, including job openings, frequency of interviews, and rejection rates. These revelations may help us make better recruiting decisions in the future and increase the efficacy and efficiency of our recruitment procedures by providing guidance for changes to our hiring strategy.

Design:

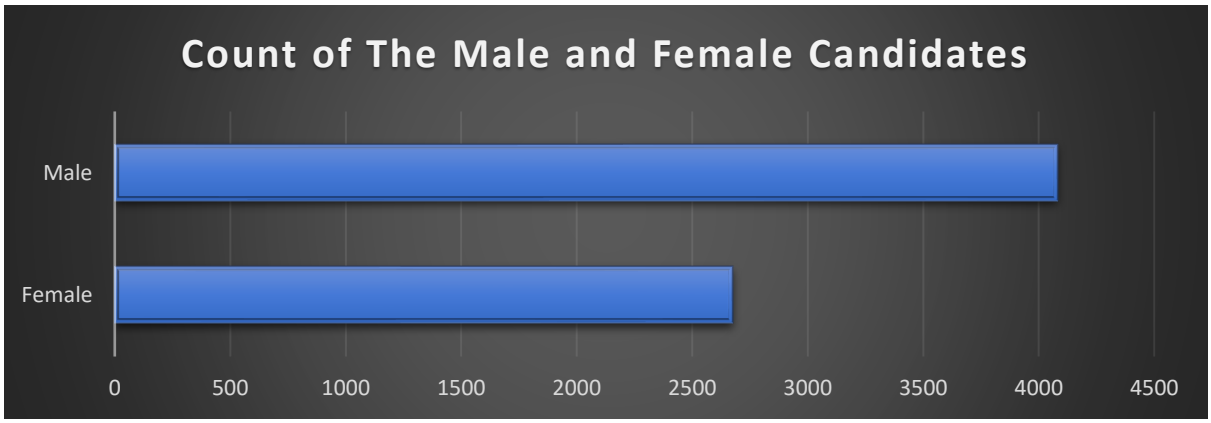
The project employs a methodical approach to exploratory data analysis (EDA), which entails comprehending the content of the data columns, looking for missing data, grouping columns with several categories for in-depth examination, recognizing and managing outliers, and producing a data summary. With the use of statistical expertise and Excel formulas, the initiative seeks to derive significant conclusions regarding employment patterns within the organization. The hiring department will receive actionable insights from the comprehensive report, which will help with decision-making and enhance the hiring procedure as a whole.

Findings:

Task 1:

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?



event_name	Count of event_name
Female	2675
Male	4085

Observation: The organization employed 4085 men as opposed to 2675 women. The company might carry out an assessment of gender diversity to guarantee fair recruiting procedures.

Task 2:

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.

Average Offered Salary	49983.03
Median of Offered Salary	49625
Minimum of Offered Salary	100
Maximum of Offered Salary	400000
The average salary offered by company is = 49983	

Obervation: The organization offers an average income of Rs. 49983. The business can evaluate if this average pay is in line with rivals' and industry norms. In order to draw and keep top talent, they might need to reconsider their compensation packages if the average is noticeably lower.

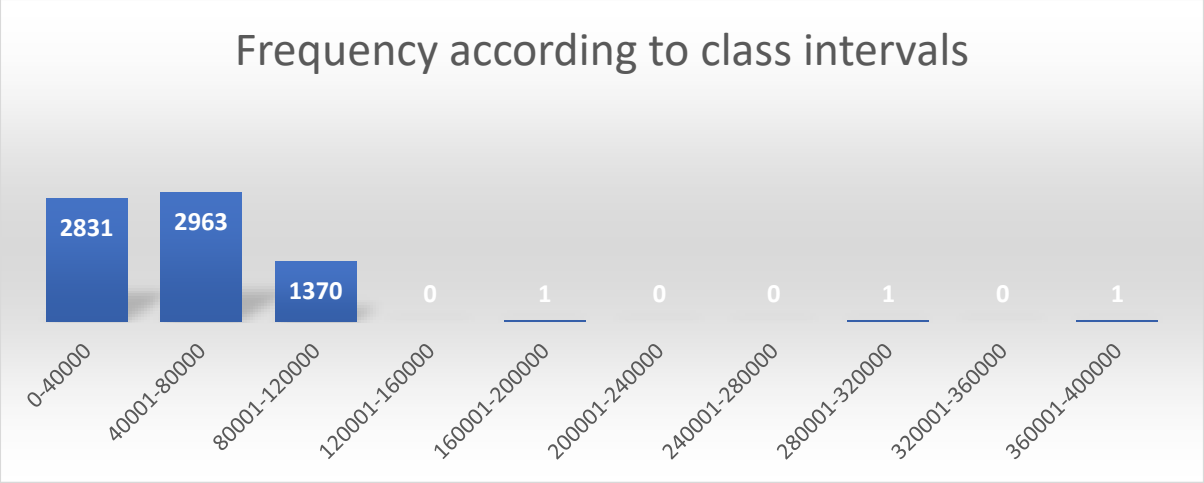
Task 3:

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.

class intervals	frequency
0-40000	2831
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



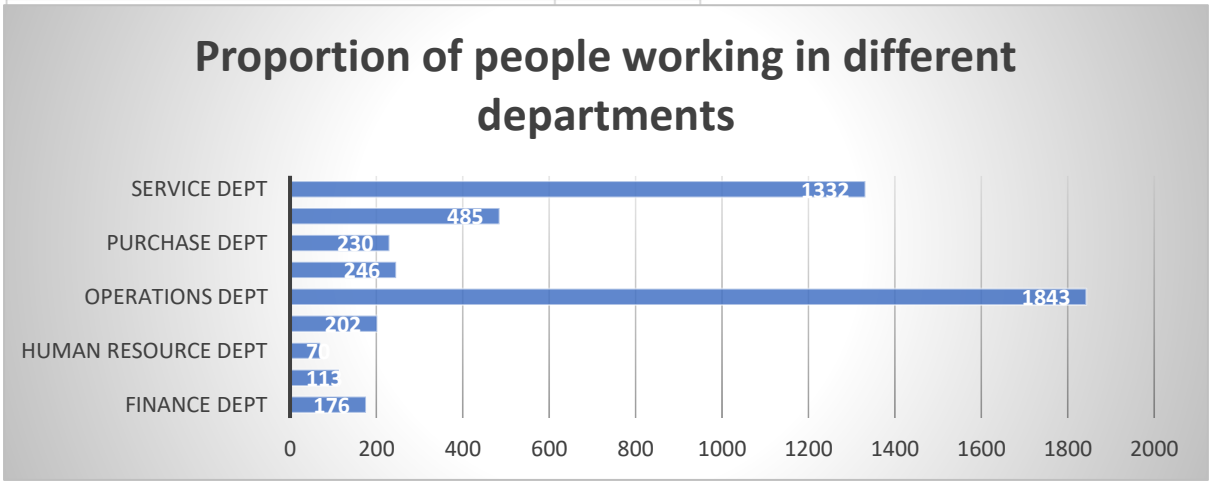
Observation: Most workers make between $0 and $80,000 annually. With the use of this data, the organization may examine how salaries are distributed and determine whether any changes are necessary to guarantee equitable and competitive pay, particularly for positions that are underrepresented in particular income categories.

Task 4:

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.

Department	Frequency
Finance Dept	176
General Management	113
Human Resource Dept	70
Marketing Dept	202
Operations dept	1843
Production Dept	246
Purchase Dept	230
Sales Dept	485
Service Dept	1332



Observation: The departments of "Operations" and "Service" employ the majority of workers. The organization can assess the distribution of departments and think about bolstering other divisions by modifying their hiring practices or providing incentives to draw in additional talent.

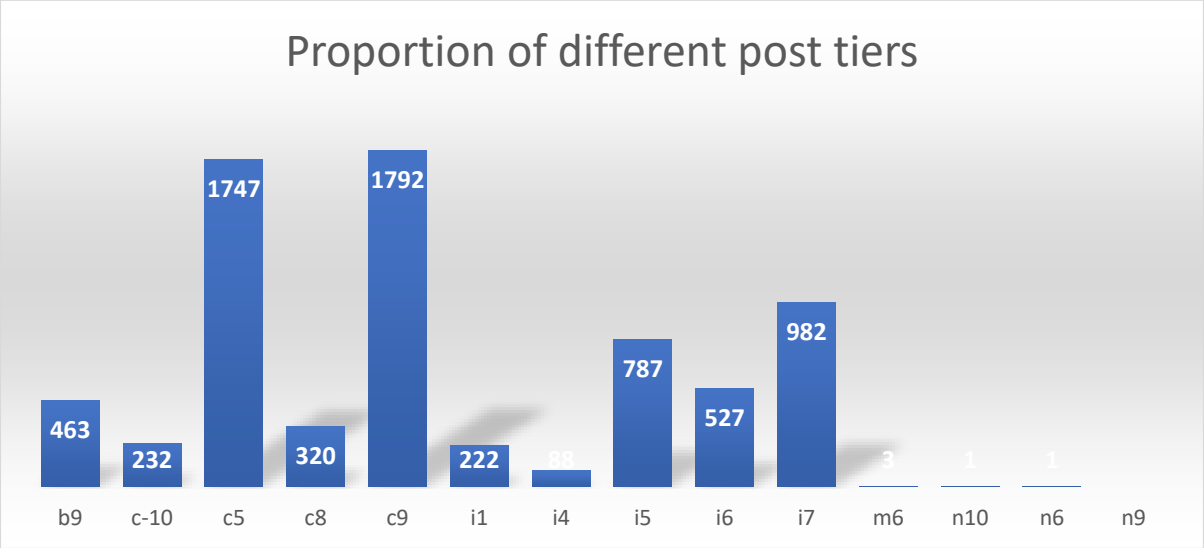
Task 5:

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.

proportion of different post tier	frequency
b9	463
c-10	232
c5	1747
c8	320
c9	1792
i1	222
i4	88
i5	787
i6	527
i7	982
m6	3
n10	1
n6	1
n9	



Observation: The majority of workers are in the "c9" post tier, with "i7" and "c5" following closely behind. The data can be utilized by the organization to assess the composition of their personnel and guarantee appropriate career advancement and expansion prospects. They might also examine if the post-tier distribution fits with the company's expansion objectives.

Analysis:

1. Missing Data: In order to ensure data integrity and dependability in later analyses, the dataset was inspected for missing values. Appropriate techniques were then put in place to address them.

2. Clubbing Columns: To expedite analysis and provide more lucid insights into hiring trends, categories within some columns were clubbed.

3. Outlier Detection and Removal: Appropriate techniques were employed to detect and eliminate outliers in order to keep them from distorting the analysis and improving the precision of inferences made from the data.

4. Data Summary: A thorough synopsis of the hiring process was produced using statistical metrics and visuals, offering a deeper knowledge of trends including rejection rates, interview frequency, job kinds, and openings.

Conclusion:

Several important insights have been discovered through the analysis of the hiring process data, and these can help the business improve its recruitment strategy. Important insights have surfaced to support decision-making by addressing missing data, clubbing pertinent columns, identifying and managing outliers properly, and summarizing the results.

IMDB Movie Analysis Description

Description:

The goal of this study is to look at what aspects affect a movie's IMDB success. In this case, having a high IMDB rating is considered successful. For filmmakers, investors, and producers of motion pictures who want to know what factors influence a film's success so they can plan ahead and make smarter decisions, this is a critical problem.

Design:

Cleaning The Data

- One of the most crucial things to do before continuing with the analysis is cleaning.
- To do this, apply the knowledge you have acquired thus far. (Removing null values, dropping columns, etc.)
- 1. Downloaded the dataset and made sense of the information offered.
- 2. During the download, a unique character emerged in the data set. Thus, I opened the Excel file in CSV format (comma separated files) and saved it as an Excel work book.
- 3. At this point, the analysis was conducted using Excel.
- 4. Removed duplicate values from the dataset after checking for them.
- 5. Next, examine the null values and remove any that don't belong in the necessary columns.

Findings:

Task 1: Movie genre analysis

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. (Top 20 most common genre).

1	Genre	Count of Movie
2	Comedy|Drama|Romance	147
3	Drama	143
4	Comedy	143
6	Comedy|Romance	132
8	Crime|Drama|Thriller	81
9	Action|Crime|Thriller	56
10	Action|Crime|Drama|Thriller	50
11	Horror	46
15	Drama|Thriller	43
16	Crime|Drama|Mystery|Thriller	41
18	Horror|Thriller	34
19	Action|Adventure|Sci-Fi|Thriller	34
20	Horror|Mystery|Thriller	31
21	Drama|Mystery|Thriller	30

Statistics Summary	Results
Mean	72.1
Median	46
Standard Deviation	44.91266379
Sample Variance	2017.147368
Range	117
Minimum	30
Maximum	147

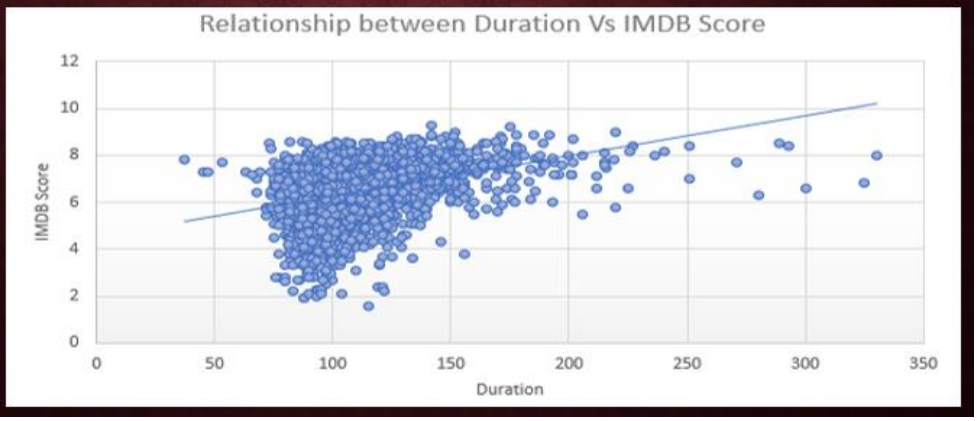
Task 2: movie duration analysis

Task: Analyze the distribution of movie durations and its impact on the IMDB score.

	duration	imdb_score
2	178	7.9
3	169	7.1
4	148	6.8
5	164	8.5
6	132	6.6
7	156	6.2
8	100	7.8
9	141	7.5
10	153	7.5
11	183	6.9
12	169	6.1
13	106	6.7
14	151	7.3
15	150	6.5
16	143	7.2
17	150	6.6
18	173	8.1
19	136	6.7
20	106	6.8
21	164	7.5

Total rows are 3782

Statistics Measure	Results
Average	110.05
Median	106.00
Std Dev	22.67



Task 3: language analysis

Task: Situation: Examine the distribution of movies based on their language.

	Movie Title	Count of movie_title
2	Aboriginal	2
3	Arabic	1
4	Aramaic	1
5	Bosnian	1
6	Cantonese	7
7	Czech	1
8	Danish	3
9	Dari	2
10	Dutch	3
11	English	3625
12	Filipino	1
13	French	34
14	German	10
15	Hebrew	1
16	Hindi	5
17	Hungarian	1
18	Indonesian	2
19	Italian	7
20	Japanese	10
21	Kazakh	1
22	Korean	5
23	Mandarin	14
24	Maya	1
25	Mongolian	1
26	None	1
27	Norwegian	4
28	Persian	3
29	Portuguese	5
30	Romanian	1
31	Russian	1
32	Spanish	23
33	Thai	3
34	Vietnamese	1
35	Zulu	1
36	Grand Total	3782

Language	mean	median	std dev
Aboriginal	6.95	6.6	1.07
Arabic	7.2	6.6	1.07
Aramaic	7.1	6.6	1.07
Bosnian	4.3	6.6	1.07
Cantonese	7.342857143	6.6	1.07
Czech	7.4	6.6	1.07
Danish	7.9	6.6	1.07
Dari	7.5	6.6	1.07
Dutch	7.566666667	6.6	1.07
English	6.415227586	6.6	1.07
Filipino	6.7	6.6	1.07
French	7.355882353	6.6	1.07
German	7.77	6.6	1.07
Hebrew	8	6.6	1.07
Hindi	7.22	6.6	1.07
Hungarian	7.1	6.6	1.07
Indonesian	7.9	6.6	1.07
Italian	7.185714286	6.6	1.07
Japanese	7.66	6.6	1.07
Kazakh	6	6.6	1.07
Korean	7.7	6.6	1.07
Mandarin	7.021428571	6.6	1.07
Maya	7.8	6.6	1.07
Mongolian	7.3	6.6	1.07
None	8.5	6.6	1.07
Norwegian	7.15	6.6	1.07
Persian	8.133333333	6.6	1.07
Portuguese	7.76	6.6	1.07
Romanian	7.9	6.6	1.07
Russian	6.5	6.6	1.07
Spanish	7.082608696	6.6	1.07
Thai	6.633333333	6.6	1.07
Vietnamese	7.4	6.6	1.07
Zulu	7.3	6.6	1.07

Task 4: director analysis

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

Top 10 highest director who has the average of imdb_score

Top 10 Director Name	Average of imdb_scor
Akira Kurosawa	8.70
Charles Chaplin	8.60
Michael Curtiz	8.60
Tony Kaye	8.60
Damien Chazelle	8.50
Majid Majidi	8.50
Ron Fricke	8.50
Sergio Leone	8.43
Christopher Nolan	8.43

```
=LARGE(Table1[Average of imdb_score],1)
=PERCENTRANK(Table1[Average of imdb_score],E16)
=PERCENTILE(Table1[Average of imdb_score],E17)
```

large	8.7
percen rank	1
percentile	8.7

Task 5: budget analysisTask: Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

corelation between gross and buget	0.0983181
highest profit margin	523505847

Total rows are 3722

	gross	budget	profit
2	760505847	237000000	523505847
3	309404152	300000000	9404152
4	200074175	245000000	-44925825
5	448130642	250000000	198130642
6	73058679	263700000	-190641321
7	336530303	258000000	78530303
8	200807262	260000000	-59192738
9	458991599	250000000	208991599
10	301956980	250000000	51956980

Analysis:

1. Determining the critical elements affecting a film's success on IMDB, such as the caliber of the production, the audience's taste in genres, and renowned directors or performers.

2. Applying the Five 'Whys' method to comprehend the causal links between these elements and high IMDB ratings.

3. Information on the tastes and behavior of viewers, particularly the role that favorable reviews have in promoting the success and acclaim of films.

4. Based on the discovered insights, possible methods for improving film success include making investments in high-quality production and catering to audience tastes.

Conclusion:

Important insights into the elements influencing the popularity of movies on IMDB have been obtained through thorough data analysis and interpretation. A thorough comprehension of the fundamental workings behind high IMDB ratings provides decision-makers in the film industry with useful information. Applying these findings going ahead can result in better production tactics, better financial choices, and ultimately, more success in the highly competitive film industry.

Bank Loan Case Study

Description:

Consider yourself a data analyst for a financial firm that focuses on providing urban clients with a range of loans. A problem exists for your business: certain clients who don't have enough credit history take advantage of this and stop making loan payments. It is your responsibility to do exploratory data analysis (EDA) to find trends in the data and make sure that qualified candidates are not turned away.

There are two dangers for your business when a consumer requests for a loan:

1. The business suffers if the applicant is able to repay the loan but is denied approval.
2. The business will suffer financial loss if the loan is authorized but the applicant is unable to repay it. You will be dealing with a dataset that includes loan application data.

There are two kinds of scenarios in it:

1. Clients have trouble making payments: These clients missed at least one of the first Y loan installments by more than X days in terms of payment.
2. Every other instance: In these instances, the payment was received on schedule. There are four things that can happen when a consumer asks for a loan:

In this project, you will utilize EDA to investigate how loan and customer attributes affect default probability.

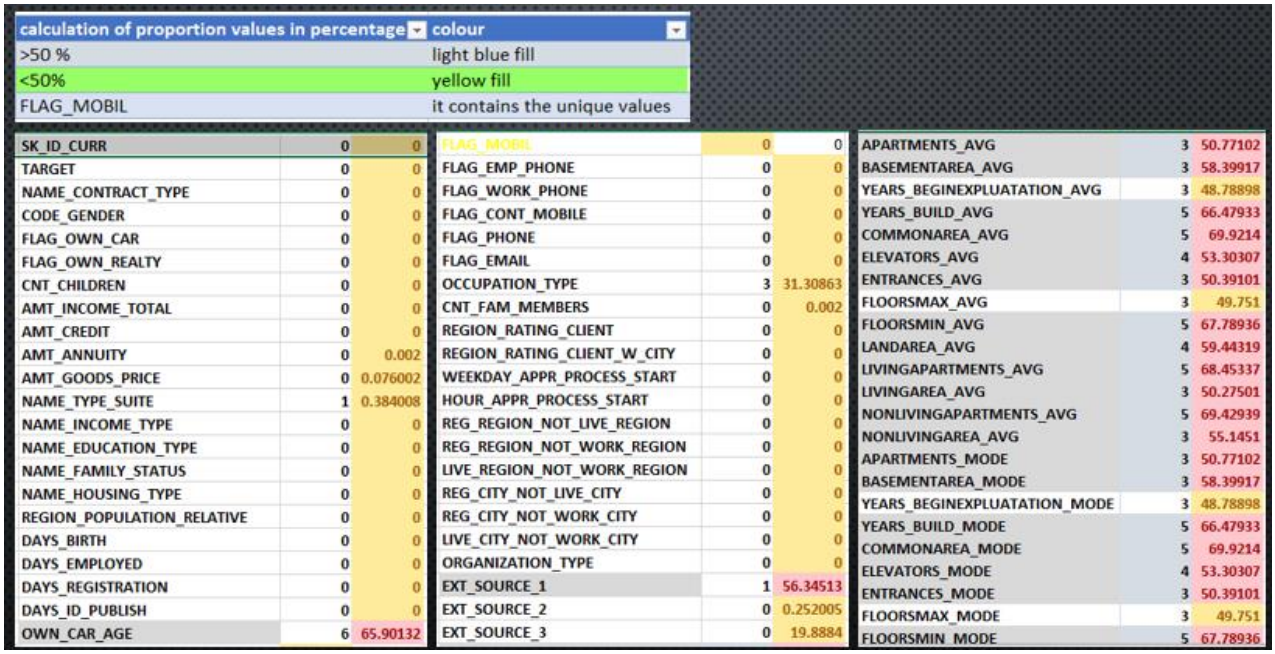
Design:

- Our collection contains tables with a wealth of information about bank loans. We thoroughly reviewed the dataset before executing a number of queries to extract relevant data from it.
- Find out the null values, blank values empty values and clean it.
- Use filter, pivot table, and other functions to give answer of asked questions.
- To make a proper solution and meaningful data representation.

Findings:

Task 1: 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.

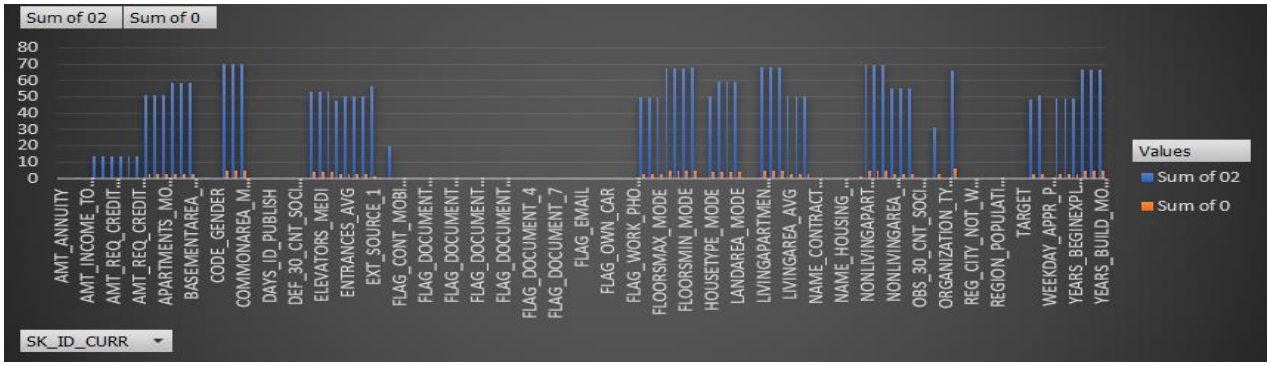


calculation of proportion values in percentage	colour
>50 %	light blue fill
<50%	yellow fill
FLAG_MOBIL	it contains the unique values

SK_ID_CURR	0	0
TARGET	0	0
NAME_CONTRACT_TYPE	0	0
CODE_GENDER	0	0
FLAG_OWN_CAR	0	0
FLAG_OWN_REALTY	0	0
CNT_CHILDREN	0	0
AMT_INCOME_TOTAL	0	0
AMT_CREDIT	0	0
AMT_ANNUITY	0	0.002
AMT_GOODS_PRICE	0	0.076002
NAME_TYPE_SUITE	1	0.384008
NAME_INCOME_TYPE	0	0
NAME_EDUCATION_TYPE	0	0
NAME_FAMILY_STATUS	0	0
NAME_HOUSING_TYPE	0	0
REGION_POPULATION_RELATIVE	0	0
DAYS_BIRTH	0	0
DAYS_EMPLOYED	0	0
DAYS_REGISTRATION	0	0
DAYS_ID_PUBLISH	0	0
OWN_CAR_AGE	6	65.90132

FLAG_MOBIL	0	0
FLAG_EMP_PHONE	0	0
FLAG_WORK_PHONE	0	0
FLAG_CONT_MOBILE	0	0
FLAG_PHONE	0	0
FLAG_EMAIL	0	0
OCCUPATION_TYPE	3	31.30863
CNT_FAM_MEMBERS	0	0.002
REGION_RATING_CLIENT	0	0
REGION_RATING_CLIENT_W_CITY	0	0
WEEKDAY_APPR_PROCESS_START	0	0
HOUR_APPR_PROCESS_START	0	0
REG_REGION_NOT_LIVE_REGION	0	0
REG_REGION_NOT_WORK_REGION	0	0
LIVE_REGION_NOT_WORK_REGION	0	0
REG_CITY_NOT_LIVE_CITY	0	0
REG_CITY_NOT_WORK_CITY	0	0
LIVE_CITY_NOT_WORK_CITY	0	0
ORGANIZATION_TYPE	0	0
EXT_SOURCE_1	1	56.34513
EXT_SOURCE_2	0	0.252005
EXT_SOURCE_3	0	19.8884

APARTMENTS_AVG	3	50.77102
BASEMENTAREA_AVG	3	58.39917
YEARS_BEGINEXPLUATATION_AVG	3	48.78898
YEARS_BUILD_AVG	5	66.47933
COMMONAREA_AVG	5	69.9214
ELEVATORS_AVG	4	53.30307
ENTRANCES_AVG	3	50.39101
FLOORSMAX_AVG	3	49.751
FLOORSMIN_AVG	5	67.78936
LANDAREA_AVG	4	59.44319
LIVINGAPARTMENTS_AVG	5	68.45337
LIVINGAREA_AVG	3	50.27501
NONLIVINGAPARTMENTS_AVG	5	69.42939
NONLIVINGAREA_AVG	3	55.1451
APARTMENTS_MODE	3	50.77102
BASEMENTAREA_MODE	3	58.39917
YEARS_BEGINEXPLUATATION_MODE	3	48.78898
YEARS_BUILD_MODE	5	66.47933
COMMONAREA_MODE	5	69.9214
ELEVATORS_MODE	4	53.30307
ENTRANCES_MODE	3	50.39101
FLOORSMAX_MODE	3	49.751
FLOORSMIN_MODE	5	67.78936



Output:

- The original dataset included 50000 rows and 125 columns, many of which had missing values.
- The count of the missing values in each column was obtained by using the countblank function, which gave rise to a preliminary comprehension of the data's integrity.
- The percentage of missing values in each column was determined since it was recognized that he absolute count of missing values would not be appropriate for analysis.

<u>Missing values proportion:</u>

- If A column has fewer than 50% missing values, imputation was done using average or median values for the numerical values in that column.

- If a column has more than 50% missing values. It was determined that it was not required to include it in the dataset.

Formulas:

- It count the blank values of the data set

  =COUNTBLANK(A49995:A50000)

- It counts the blank values proportion in the form of percentage

  =(COUNTBLANK(A2:A50000)/COUNT($A$2:$A$50000))*100

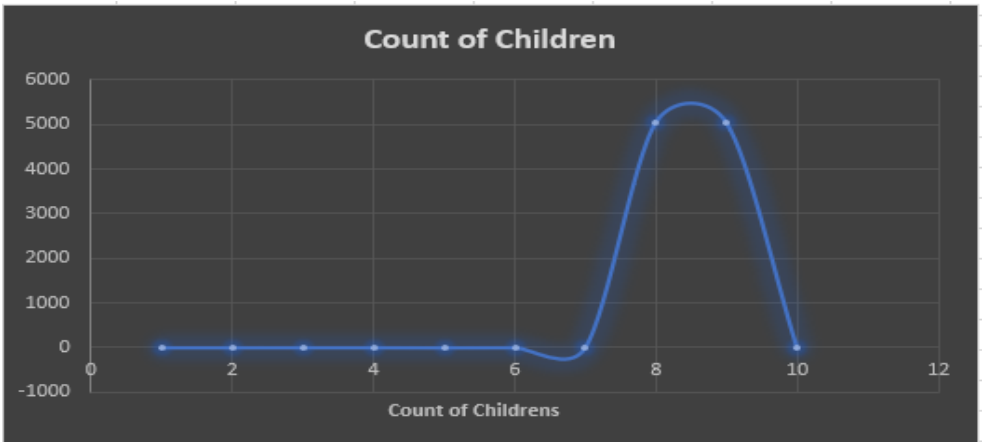
Task 2: identify outliers in the datasetTask:

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

The analysis done to find outliers is displayed in the table below:

Columns contained the numerical values	Q1	Q2	Q3	IQR=Q3-A1	lower bound=Q1-1.5*(Q3-Q1)	Upper bound=Q1+1.5*(Q3-1A1)	Outliers below the lower bound	Outliers above the upper bound	Total Outliers=Outliers below the lower bound + Outbounds above the upper bound	maximum value	checking wheather the outliers are valid data points
SK_ID_CURR	114570.5	129076	143438.5	28868	71268.5	157872.5	0	3	3	157875	no outlier found
TARGET	0	0	0	0	0	0	0	4026	4026	1	valid
CNT_CHILDREN	0	0	1	1	-1.5	1.5	0	5042	5042	11	not valid = 11 children it not a common in these days
AMT_INCOME_TOTAL	112500	145800	202500	90000	-22500	247500	0	6534	6534	117000000	not valid = it has the more income than expected to fit in
AMT_CREDIT	270000	514777.5	808650	538650	-537975	1077975	0	6662	6662	4050000	valid
AMT_ANNUITY	16456.5	24939	34596	18139.5	-10752.75	43665.75	0	6015	6015	258025.5	valid
AMT_GOODS_PRICE	238500	450000	679500	441000	-423000	900000	0	6813	6813	4050000	valid
REGION_POPULATION_R	0.010006	0.01885	0.028663	0.018657	-0.0179795	0.0379915	0	3517	3517	0.072508	valid
DAYS_BIRTH	-19644	-15731	-12378.5	7265.5	-30542.25	-8745.75	0	1354	1354	-7680	valid
DAYS_EMPLOYED	-2786	-1221	-292	2494	-6527	955	2788	8924	11712	365243	not valid= because estimated count the 365243 days is near about 1000 years and someone has the experience 1000 years so
DAYS_REGISTRATION	-7463.5	-4490	-1998	5465.5	-15661.75	734.75	96	0	96	0	valid
DAYS_ID_PUBLISH	-4297	-3261	-1722	2575	-8159.5	-434.5	0	2920	2920	0	valid
EXT_SOURCE_2	0.391722	0.565585	0.6634023	0.27167978	-0.01579719	0.799242161	0	86	86	0.854999666	no outlier found
EXT_SOURCE_3	0.37065	0.535276	0.6690567	0.29840706	-0.07696096	0.818260227	0	888	888	0.896009549	no outlier found
YEARS_BEGINEXPLUATAT	0.9767	0.9816	0.9866	0.0099	0.96185	0.99155	789	2411	3200	1	valid
FLOORSMAX_AVG	0.1667	0.1667	0.3333	0.1666	-0.0832	0.4166	0	1901	1901	1	valid
YEARS_BEGINEXPLUATAT	0.9762	0.9816	0.9866	0.0104	0.9606	0.9918	727	2165	2892	1	valid
YEARS_BEGINEXPLUATAT	0.9767	0.9816	0.9866	0.0099	0.96185	0.99155	786	2444	3230	1	valid
FLOORSMAX_MEDI	0.1667	0.1667	0.3333	0.1666	-0.0832	0.4166	0	1878	1878	1	valid
TOTALAREA_MODE	0.0415	0.0685	0.12705	0.08555	-0.086825	0.169825	0	4590	4590	1	valid
OBS_30_CNT_SOCIAL_CIR	0	0	2	2	-3	3	0	7060	7060	28	valid
DEF_30_CNT_SOCIAL_CIR	0	0	0	0	0	0	0	5642	5642	6	valid
OBS_60_CNT_SOCIAL_CIR	0	0	2	2	-3	3	0	6955	6955	28	valid
DEF_60_CNT_SOCIAL_CIR	0	0	0	0	0	0	0	4108	4108	5	valid
DAYS_LAST_PHONE_CHA	-1573	-755	-270	1303	-3527.5	381.5	63	0	63	0	valid
FLAG_DOCUMENT_2	0	0	0	0	0	0	0	2	2	1	valid
FLAG_DOCUMENT_3	0	1	1	1	-1.5	1.5	0	0	0	1	valid
FLAG_DOCUMENT_4	0	0	0	0	0	0	0	9	9	1	valid
FLAG_DOCUMENT_5	0	0	0	0	0	0	0	785	785	1	valid
FLAG_DOCUMENT_6	0	0	0	0	0	0	0	4335	4335	1	valid
FLAG_DOCUMENT_7	0	0	0	0	0	0	0	11	11	1	valid

FLAG_DOCUMENT_8	0	0	0	0	0	0	0	4038	4038	1	valid
FLAG_DOCUMENT_9	0	0	0	0	0	0	0	184	184	1	valid
FLAG_DOCUMENT_10	0	0	0	0	0	0	0	1	1	1	valid
FLAG_DOCUMENT_11	0	0	0	0	0	0	0	213	213	1	valid
FLAG_DOCUMENT_12	0	0	0	0	0	0	0	0	0	0	valid
FLAG_DOCUMENT_13	0	0	0	0	0	0	0	161	161	1	valid
FLAG_DOCUMENT_14	0	0	0	0	0	0	0	158	158	1	valid
FLAG_DOCUMENT_15	0	0	0	0	0	0	0	41	41	1	valid
FLAG_DOCUMENT_16	0	0	0	0	0	0	0	501	501	1	valid
FLAG_DOCUMENT_17	0	0	0	0	0	0	0	15	15	1	valid
FLAG_DOCUMENT_18	0	0	0	0	0	0	0	425	425	1	valid
FLAG_DOCUMENT_19	0	0	0	0	0	0	0	35	35	1	valid
FLAG_DOCUMENT_20	0	0	0	0	0	0	0	26	26	1	valid
FLAG_DOCUMENT_21	0	0	0	0	0	0	0	19	19	1	valid
AMT_REQ_CREDIT_BURE	0	0	0	0	0	0	0	295	295	3	valid
AMT_REQ_CREDIT_BURE	0	0	0	0	0	0	0	272	272	6	valid
AMT_REQ_CREDIT_BURE	0	0	0	0	0	0	0	1314	1314	6	valid
AMT_REQ_CREDIT_BURE	0	0	0	0	0	0	0	7140	7140	24	valid
AMT_REQ_CREDIT_BURE	0	0	0	0	0	0	0	8134	8134	8	valid
AMT_REQ_CREDIT_BURE	0	1	3	3	-4.5	4.5	0	4150	4150	25	valid



FORMULAS:

Quartile 1: =QUARTILE(B2:B50000,1)

Quartile 2: =QUARTILE(B2:B50000,2)

Quartile 3: =QUARTILE(B2:B50000,3)

IQR=Q3-Q1

lower bound=Q1-1.5*(Q3-Q1)

Upper bound=Q1+1.5*(Q3-1A1)

Outliers below the lower bound: =COUNTIF(B2:B50000,"<"&B50008)

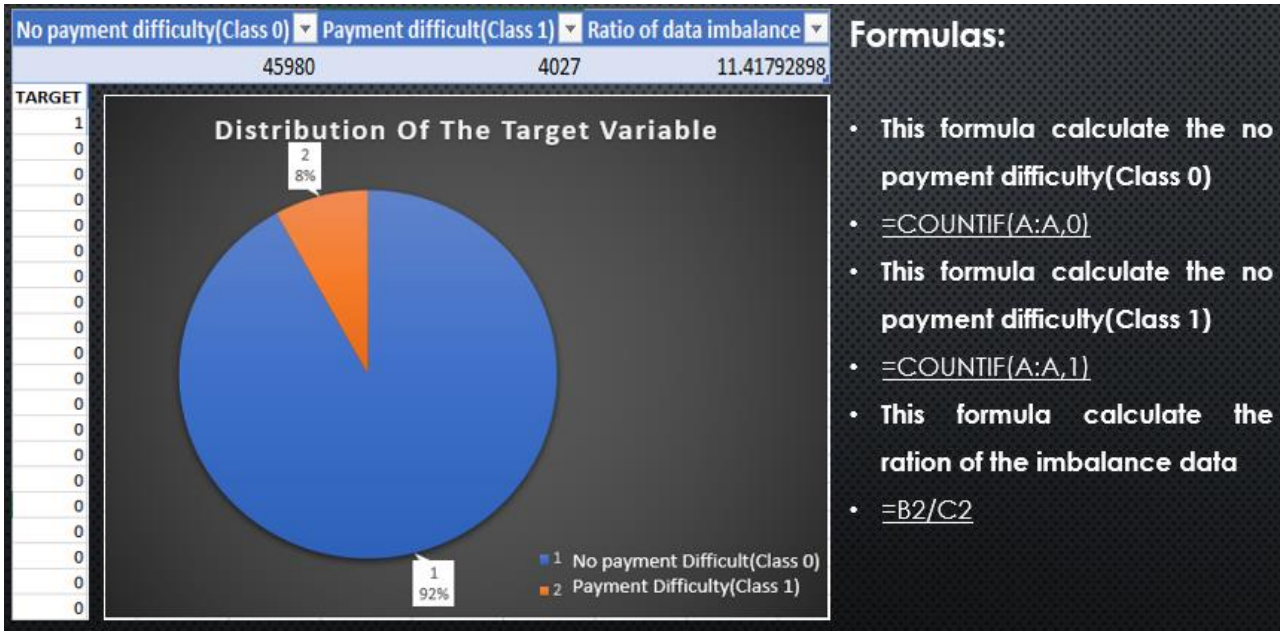
Outliers above the upper bound: =COUNTIF(B2:B50000,">"&B50009)

Total Outliers=Outliers below the lower bound + Outbounds above the upper bound:

maximum value: =MAX()

TASK 3: Analyze Data Imbalance

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



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

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: Univariate Analysis Summarize Only One Variable At A Time.

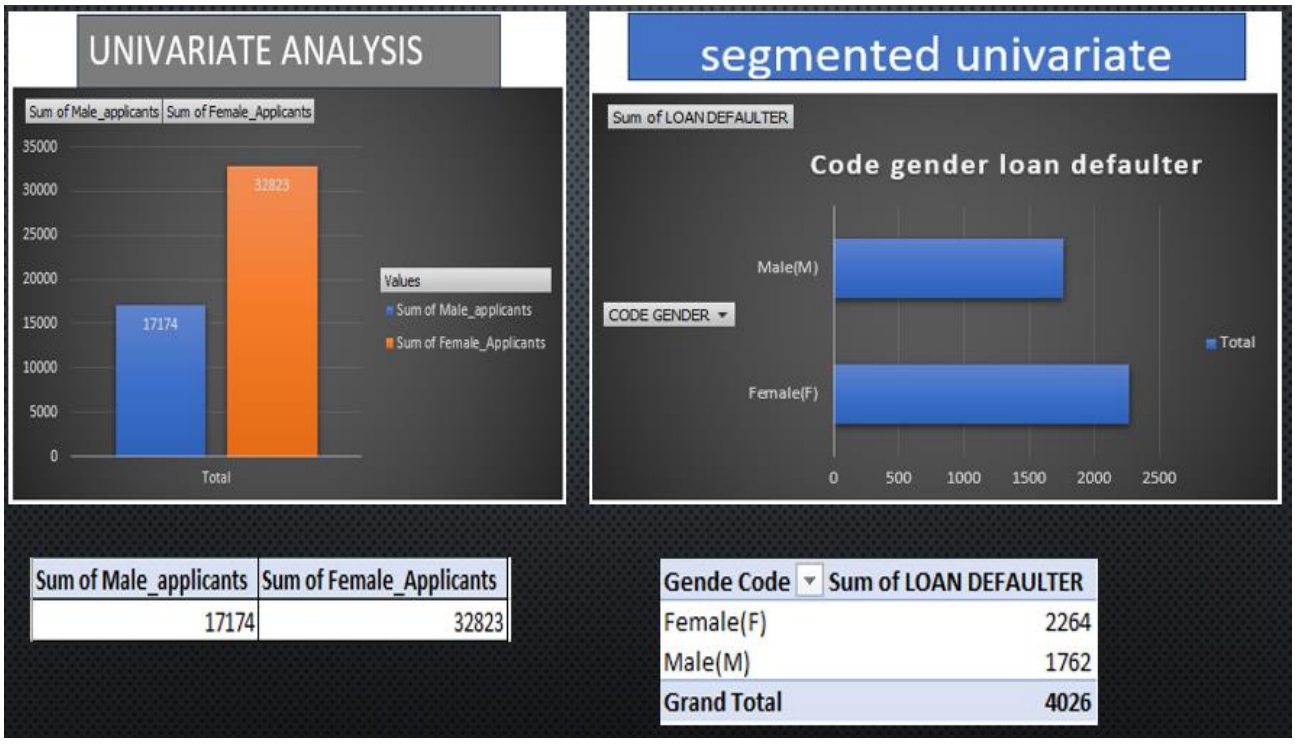
e.g. APPLICANTS PER CREDIT BINS and NAME-TYPE-SUITE

Segmented Univariate: Segment the categorical variable and then perform univariate analysis across its categorical.

e.g. TARGET APPLICANTS PER INCOME BINS

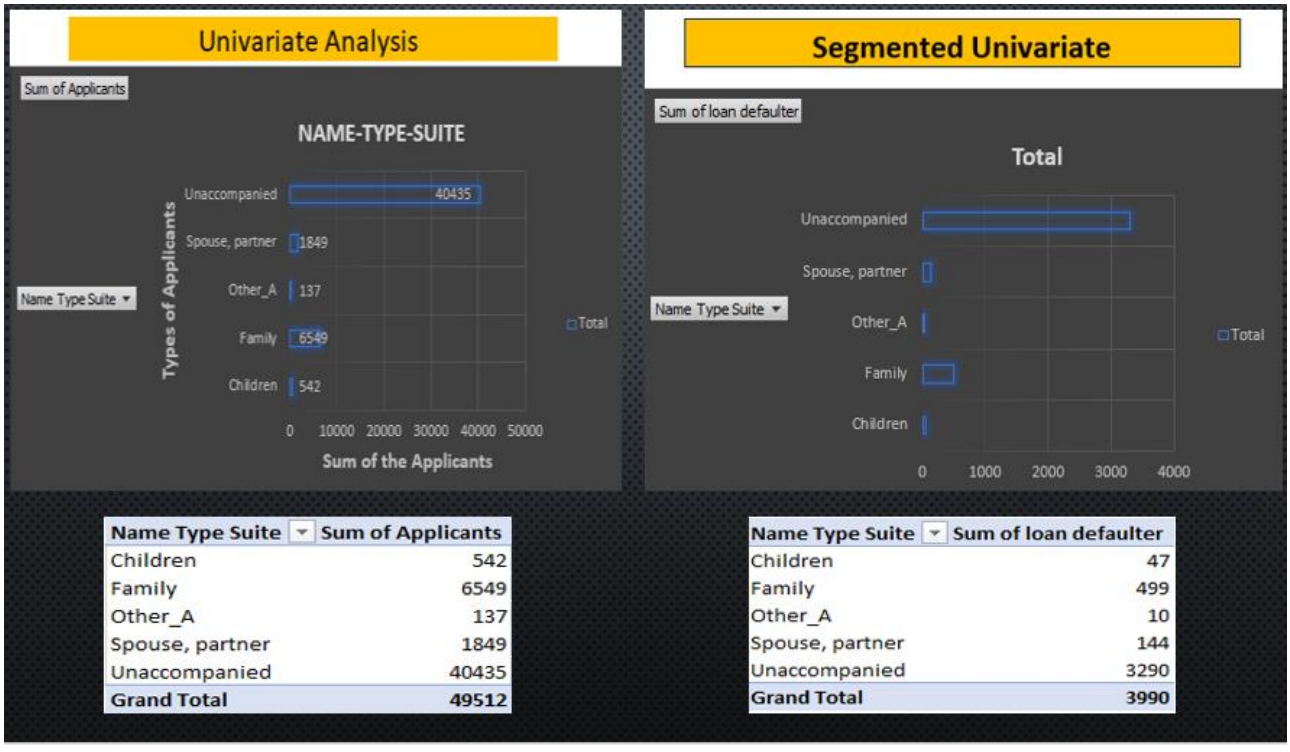
Bivariate Analysis: Bivariate statistics compare two variables.

e.g. AVERAGE CREDIT AMOUNT PER INCOME BINS



Sum of Male_applicants	Sum of Female_Applicants
17174	32823

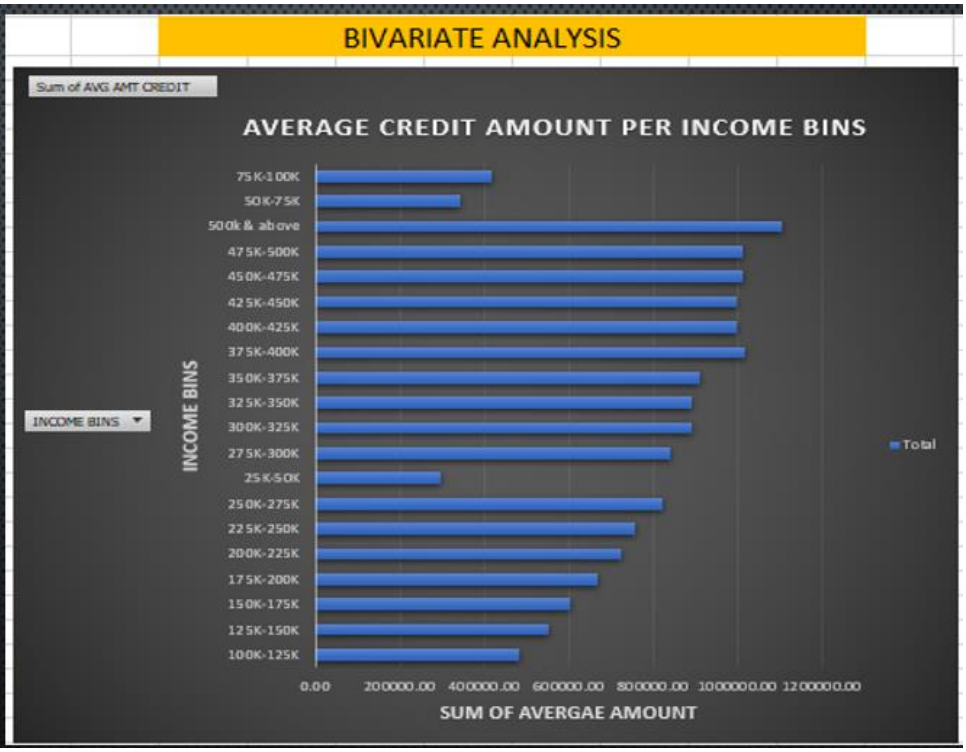
Gende Code	Sum of LOAN DEFAULTER
Female(F)	2264
Male(M)	1762
Grand Total	4026



Name Type Suite	Sum of Applicants
Children	542
Family	6549
Other_A	137
Spouse, partner	1849
Unaccompanied	40435
Grand Total	49512

Name Type Suite	Sum of loan defaulter
Children	47
Family	499
Other_A	10
Spouse, partner	144
Unaccompanied	3290
Grand Total	3990

INCOME BIN	Sum of AVG AMT CREDIT
100K-125K	483568.81
125K-150K	553042.16
150K-175K	602034.40
175K-200K	667004.42
200K-225K	727198.44
225K-250K	759541.38
250K-275K	820255.35
25K-50K	297752.08
275K-300K	842725.65
300K-325K	892300.07
325K-350K	892332.65
350K-375K	910363.05
375K-400K	1016814.38
400K-425K	999208.20
425K-450K	999153.64
450K-475K	1011521.84
475K-500K	1015150.40
500k & above	1105365.12
50K-75K	345240.36
75K-100K	417267.88
Grand Total	15357840.27



Output:

1. A through grasp of each variable's properties was made possible via univariate analysis.

2. which was crucial for locating possible predictors of loan default. compared variable distributions for several scenarios using segmented univariate analysis technique.

3. the study employed bivariate analysis to investigate correlation between the a target variable and other factors.

Formulas:

- **Calculate the maximum/minimum value from specified column**
- =MAX(AQ:AQ) =MIN(AQ:AQ)
- **Calculate the days in years** =ROUND(AP2/365.25,0)
- **It helps to count loan defaulter the values from the specified conditions**
- =COUNTIFS(AQ:AQ,">=20",AQ:AQ,"<=25",B:B,1)/COUNTIFS(AQ:AQ,">=20",AQ:AQ,"<=25")*100
- **Count the average of the observation with the help of specified condition**
- =AVERAGEIFS(Z:Z,M:M,">=25000",M:M,"<=50000")

Task 5: Identify Top Correlations for Different Scenarios

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.

Top correlation among loan applicants who meet the payment criteria

TASK E - 1						
CNT_CHILDREN	1	0.009588558	-0.025555665	0.012668609	0.007745347	-1.05885E-05
AMT_INCOME_TOTAL	0.009588558	1	0.029841469	-0.000520882	-0.0014739	-0.005587637
REGION_POPULATION_RELATIVE	-0.025555665	0.029841469	1	-0.004910106	0.000514692	0.006587724
DAYS_BIRTH(Years)	0.012668609	-0.000520882	-0.004910106	1	0.35155068	0.006520604
DAYS_EMPLOYED(Years)	0.007745347	-0.0014739	0.000514692	0.35155068	1	-0.007529606
DAYS_ID_PUBLISH(Years)	-1.05885E-05	-0.005587637	0.006587724	0.006520604	-0.007529606	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	REGION_POPULATION_RELATIVE	DAYS_BIRTH(Years)	DAYS_EMPLOYED(Years)	DAYS_ID_PUBLISH(Years)

Top correlation among loan applicants who fail to meet payment criteria

TASK E -2						
CNT_CHILDREN	1	0.009588558	-0.025555665	0.012668609	0.007745347	-1.05885E-05
AMT_INCOME_TOTAL	0.009588558	1	0.029841469	-0.000520882	-0.0014739	-0.005587637
REGION_POPULATION_RELATIVE	-0.025555665	0.029841469	1	-0.004910106	0.000514692	0.006587724
DAYS_BIRTH(Years)	0.012668609	-0.000520882	-0.004910106	1	0.35155068	0.006520604
DAYS_EMPLOYED(Years)	0.007745347	-0.0014739	0.000514692	0.35155068	1	-0.007529606
DAYS_ID_PUBLISH(Years)	-1.05885E-05	-0.005587637	0.006587724	0.006520604	-0.007529606	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	REGION_POPULATION_RELATIVE	DAYS_BIRTH(Years)	DAYS_EMPLOYED(Years)	DAYS_ID_PUBLISH(Years)

Analysis:

- Identification of key factors influencing movie success on IMDB, such as production quality, genre preferences, and notable directors or actors.
- Understanding the causal relationships between these factors and high IMDB ratings through the Five 'Whys' approach.
- Insights into viewer behavior and preferences, including the significance of positive reviews in driving movie popularity and success.
- Potential strategies for enhancing movie success based on the identified insights, such as investing in production quality and aligning with audience preferences.

Conclusion:

Important insights into the elements influencing the popularity of movies on IMDB have been obtained through thorough data analysis and interpretation. A thorough comprehension of the fundamental workings behind high IMDB ratings provides decision-makers in the film industry with useful information. Applying these findings going ahead can result in better production tactics, better financial choices, and ultimately, more success in the highly competitive film industry.

Impact Of Car Features

Description:

The objective of this project is to support an automobile manufacturer in making decisions about pricing and product development that will optimize profitability and satisfy consumer demand. Understanding consumer preferences and market dynamics is essential for success in the continuously changing automotive sector, which is marked by a shift towards electric and hybrid vehicles alongside traditional gasoline-powered cars.

Design:

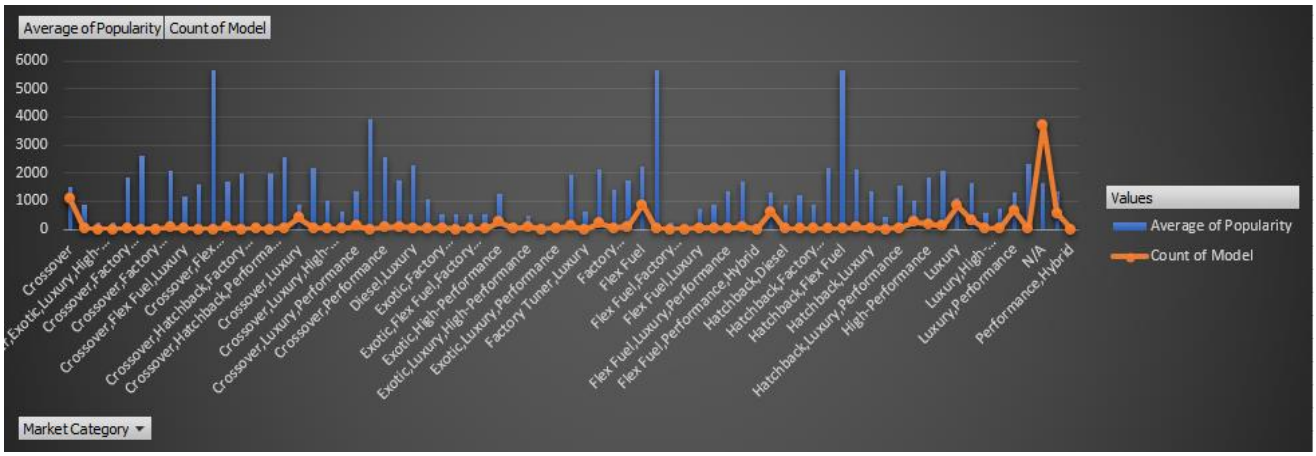
- The relationship between automotive characteristics, market classifications, and pricing will be examined in order to solve the issue.
- Regression analysis and market segmentation are two data analysis approaches that can be used to acquire insights into which features and categories are more popular with consumers and profitable for the manufacturer.
- This research will help determine which important product characteristics to focus on in subsequent product development initiatives and will also help design a pricing plan that balances profitability and consumer demand.

Findings:

Task 1:

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

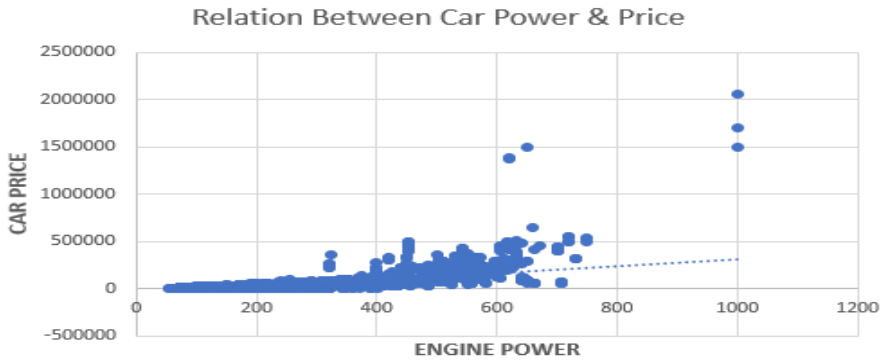
- **Task 1.A:** Create a pivot table that shows the number of car models in each market category and their corresponding popularity scores.
- **Task 1.B:** Create a combo chart that visualizes the relationship between market category and popularity.



Task 2:

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

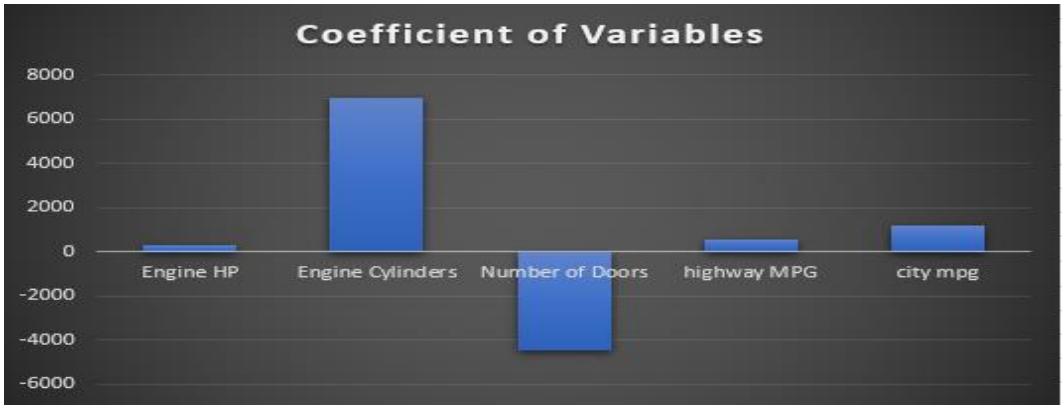
- **Task 2:** Create a scatter chart that plots engine power on the x-axis and price on the y-axis. Add a trendline to the chart to visualize the relationship between these variables.



Task 3:

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.



Task 4:

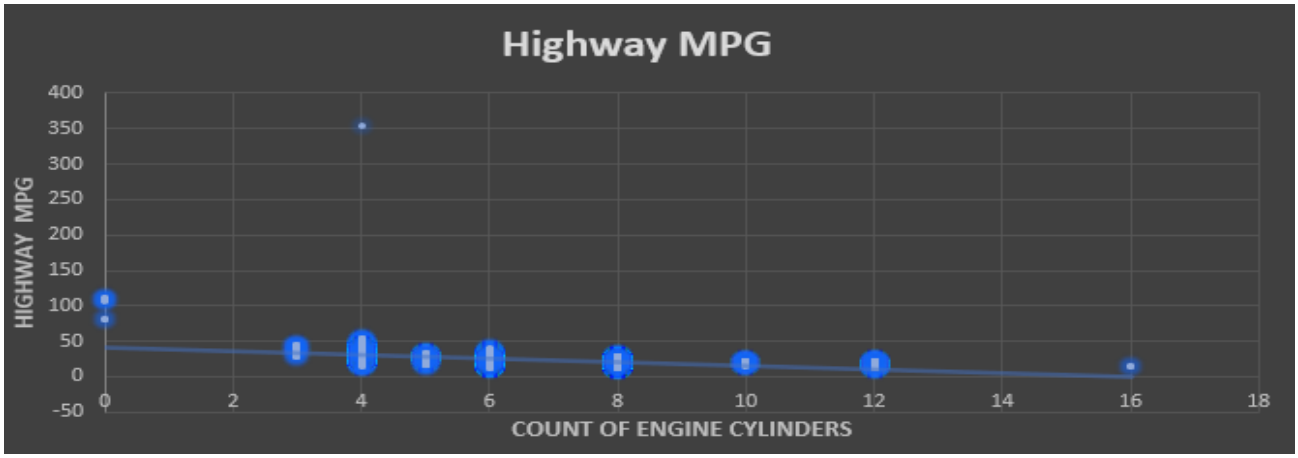
Insight Required: How does the average price of a car vary across different manufacturers?

- **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.

Task 5:

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

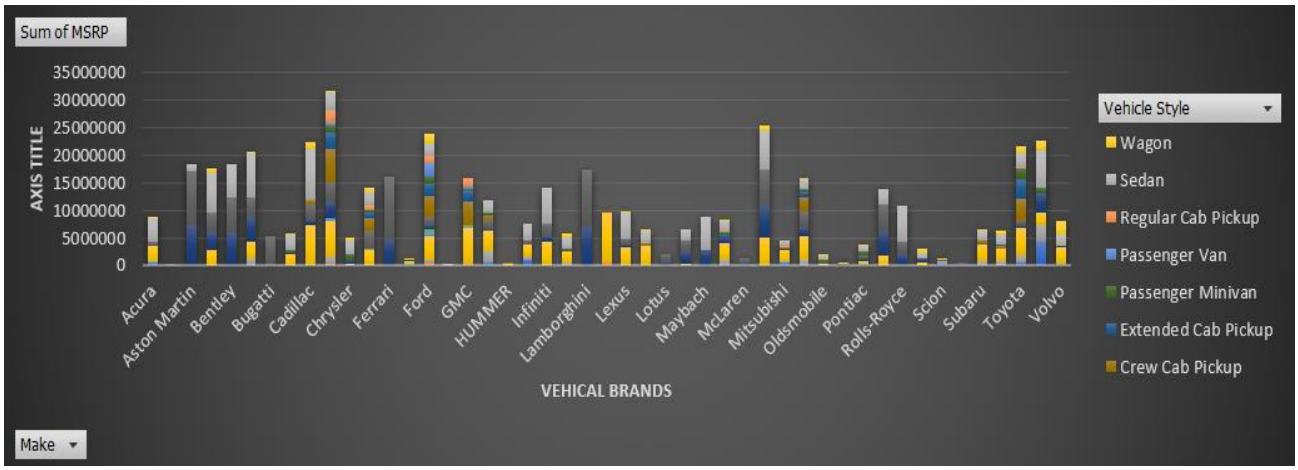
- **Task 5.A:** Create a scatter plot with the number of cylinders on the x-axis and highway MPG on the y-axis. Then create a trendline on the scatter plot to visually estimate the slope of the relationship and assess its significance.
- **Task 5.B:** Calculate the correlation coefficient between the number of cylinders and highway MPG to quantify the strength and direction of the relationship.



Building The Dashboard

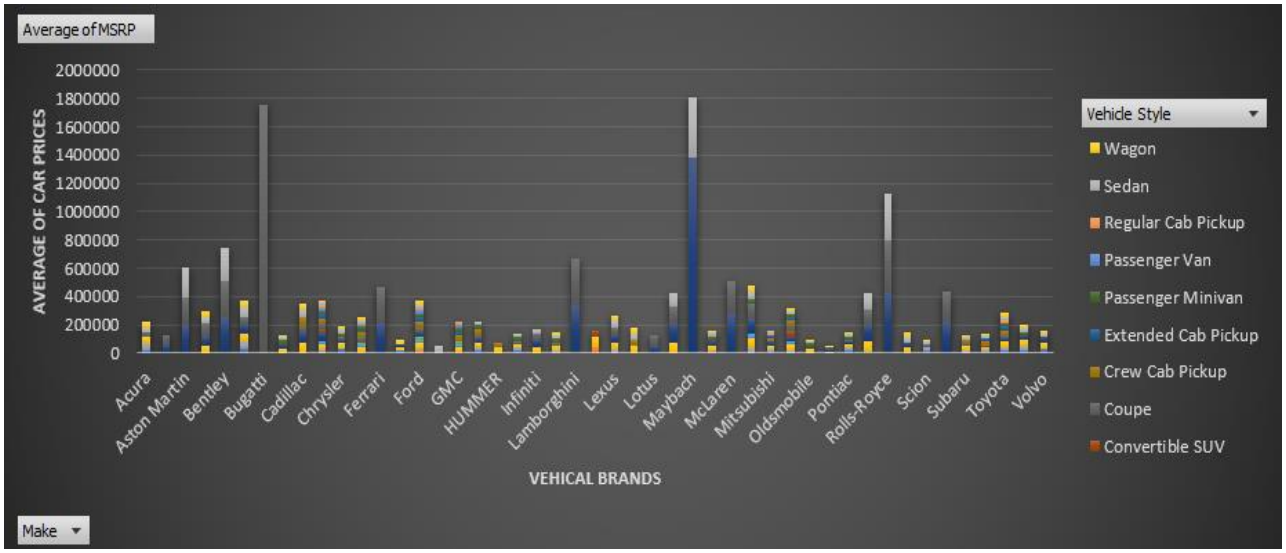
Task 1: How does the distribution of car prices vary by brand and body style?

- **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.



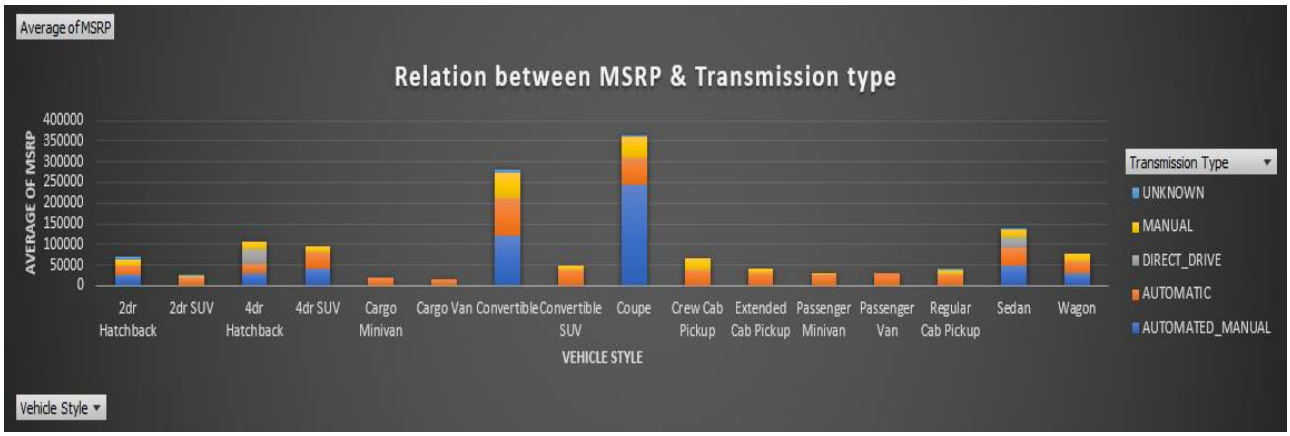
Task 2: Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

- **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.



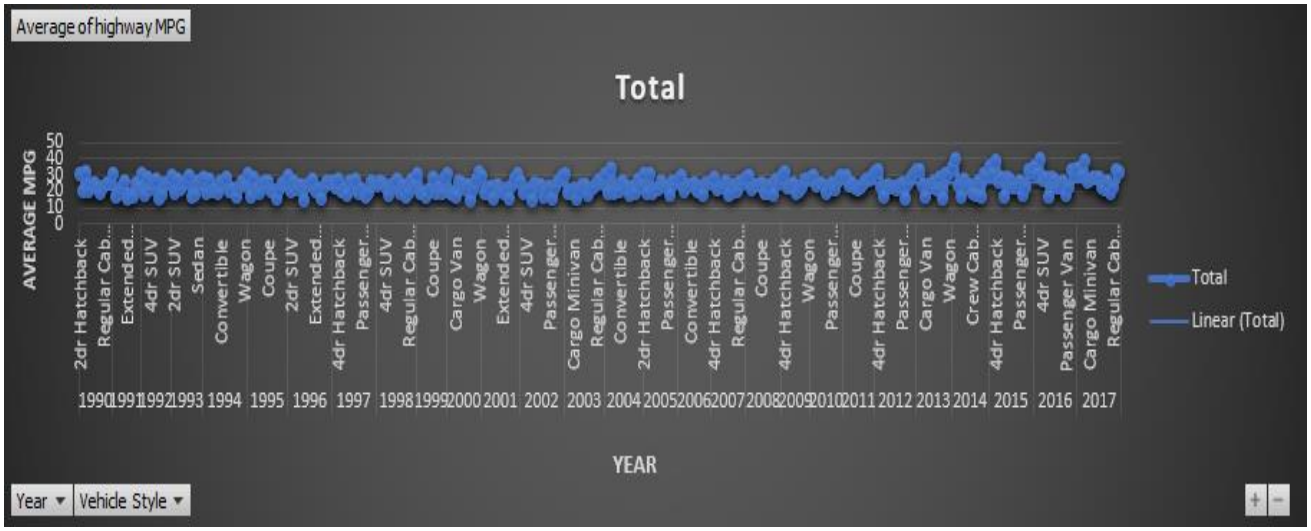
Task 3: How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?

- **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.



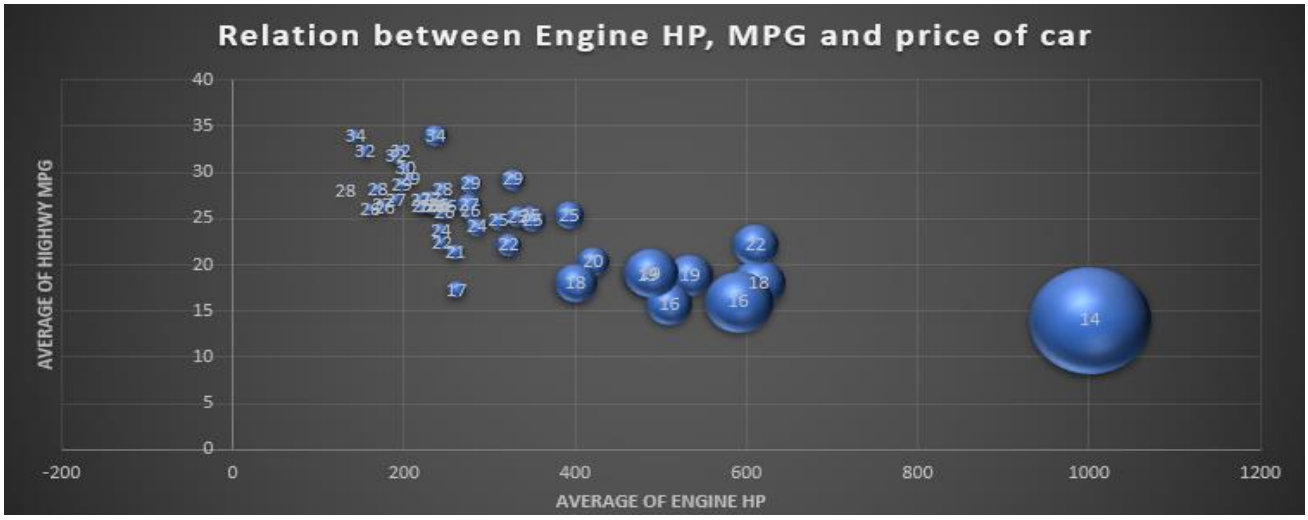
Task 4: How does the fuel efficiency of cars vary across different body styles and model years?

- **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.



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.



Analysis:

- **Pricing Sensitivity:** After looking at the data, we found that some car features and types of cars are more affected by changes in price compared to others. Knowing which features and types are sensitive to price helps us make a pricing plan that makes the most money while still staying competitive.
- **Product Development Focus:** We looked at what people want in cars by checking out the features, types of cars, and what people buy. This helps us decide what new features to add to future cars to make them more appealing to buyers and increase sales.
- **Balancing Profitability and Consumer Demand:** We need to make sure we make enough money while also giving customers what they want. By using data to

understand what customers are willing to pay and what they expect, we can set prices that make us money and keep customers happy.

- **Long-term Competitiveness:** To stay ahead in the market for a long time, we use data to make smart decisions about prices and product development. By always keeping an eye on what customers like and what's happening in the market, we can change our plans as needed to keep growing and making money.

Conclusion:

- In conclusion, data analysis provides valuable insights that drive informed decision-making in the automotive industry.
- By understanding pricing sensitivity, focusing on product development, balancing profitability and consumer demand, and prioritizing long-term competitiveness, manufacturers can optimize their strategies to thrive in an ever-changing market landscape.
- These insights serve as a roadmap for achieving sustainable growth and success in the automotive industry.

ABC Call Volume Trend Analysis

Description:

- The attached dataset is of Inbound calls of an ABC company from the insurance category consists of a Customer Experience (CX) Inbound calling team for 23 days. Data includes Agent_Name, Agent_ID, Queue_Time [duration for which customer have to wait before they get connected to an agent], Time [time at which call was made by customer in a day], Time_Bucket [for easiness we have also provided you with the time bucket], Duration [duration for which a customer and executives are on call, Call_Seconds [for simplicity we have also converted those time into seconds], call status (Abandon, answered, transferred).
- Professionals on a customer experience (CX) team examine data and comments from customers, then disseminate their findings to the rest of the company. Customer experience programs (CX programs), digital customer experience, design and processes, internal communications, voice of the customer (VoC), user experiences, customer experience management, journey mapping, nurturing customer interactions, customer success, customer support, handling customer data, and learning about the customer experience are just a few of the roles and responsibilities that these teams typically carry out.

Design:

- Our collection contains tables with a wealth of information about calls. We thoroughly reviewed the dataset before executing a number of queries to extract relevant data from it.
- Find out the null values, blank values empty values and clean it (If applicable).

- Use filter, pivot table, and other functions to give answer of asked questions.
- To make a proper decision-making solution.

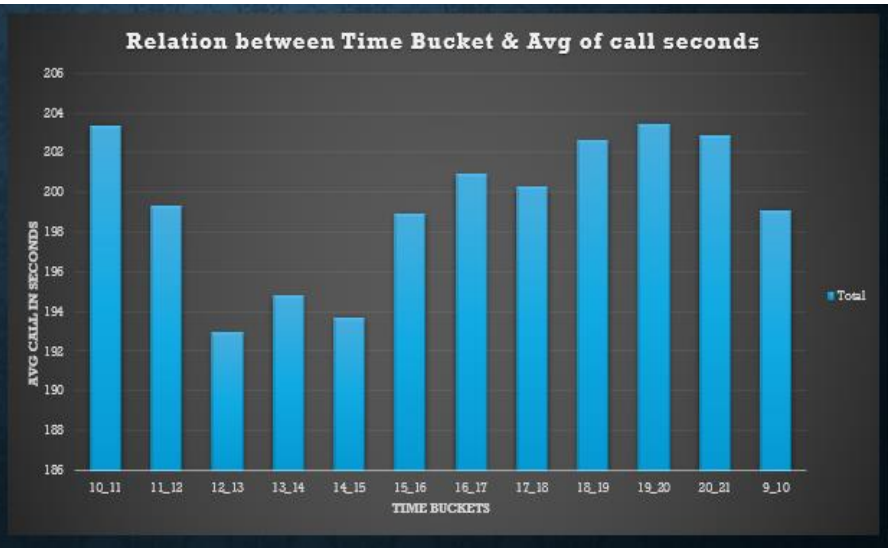
Findings:

Task 1: Average call duration

Task: What is the average duration of calls for each time bucket?

Determine the average duration of all incoming calls received by agents. This should be calculated for each time bucket.

Call_Status	answered
Time Bucke	Average of Call_Seconds (s)
10_11	203.3310302
11_12	199.2550234
12_13	192.8887829
13_14	194.7401744
14_15	193.6770755
15_16	198.8889175
16_17	200.8681864
17_18	200.2487831
18_19	202.5509677
19_20	203.4060725
20_21	202.845993
9_10	199.0691057
Grand Total	198.6227745

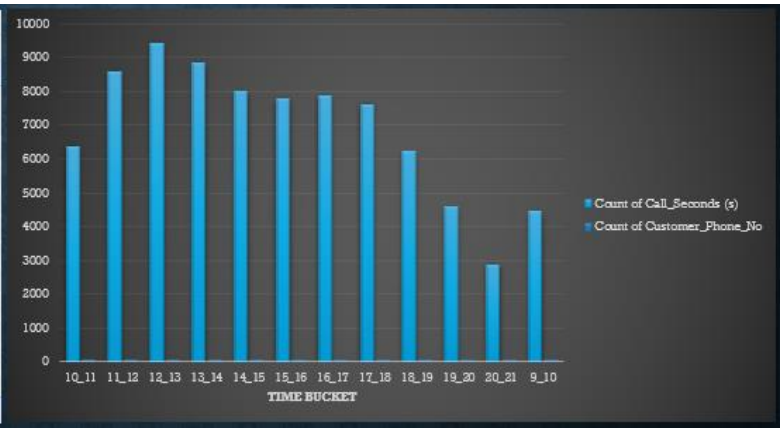


Task 2: Call Volume Analysis

Task: Can you create a chart or graph that shows the number of calls received in each time bucket?

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.).

Call_Status	answered	
Time Bucke	Count of Call_Seconds (s)	Count of Customer_Phone_No
10_11	6368	7.72%
11_12	8560	10.38%
12_13	9432	11.44%
13_14	8829	10.71%
14_15	7974	9.67%
15_16	7760	9.41%
16_17	7852	9.52%
17_18	7601	9.22%
18_19	6200	7.52%
19_20	4578	5.55%
20_21	2870	3.48%
9_10	4428	5.37%
Grand Total	82452	100.00%

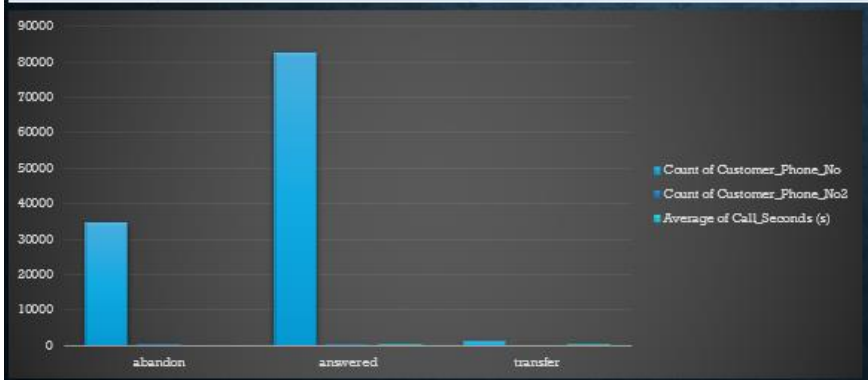


Task 3: Manpower Planning

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

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.

Call Status	Count of Customer_Phone_No	Count of Customer_Phone_No2	Average of Call_Seconds (s)
abandon	34403	29.16%	0
answered	82452	69.88%	198.6227745
transfer	1133	0.96%	76.14651368
Grand Total	**117988**	**100.00%**	**139.5321473**



Assumption

Total working hours = 9 hours

Break = 1.5 hours

IT downtime = 0.5 Hours

Meetings = 1 Hour

Actual working hours = 5 Hours

Task 4: Night Shift Manpower Planning

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

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:

Distribution of 30 calls coming in night for every 100 calls coming in between 9am - 9pm (i.e. 12 hrs slot)

9pm- 10pm	10pm - 11pm	11pm- 12am	12am- 1am	1am - 2am	2am - 3am	3am - 4am	4am - 5am	5am - 6am	6am - 7am	7am - 8am	8am - 9am
3	3	2	2	1	1	1	1	3	4	4	5

time bucke	call distribution	time distributic	agent requirement
9_10	3	10.00	1.10
10_11	3	10.00	1.10
11_12	2	15.00	0.73
12_1	2	15.00	0.73
1_2	1	30.00	0.37
2_3	1	30.00	0.37
3_4	1	30.00	0.37
4_5	1	30.00	0.37
5_6	3	10.00	1.10
6_7	4	7.50	1.47
7_8	4	7.50	1.47
8_9	5	6.00	1.83
Total	30	1	11.00

Formula = p14/q Formula = 15/Q

average call by asnwered=	3584.869565
answer call daily=	3564
for night 9pm-9am	1069.2
additional hour req=	53
additional agent req=	11
1 day=100 calls	
1 night=30 calls	

Count of Call_Status	Call Status			
Date	abandon	answered	transfer	Grand Total
01-Jan	684	3883	77	4644
02-Jan	356	2935	60	3351
03-Jan	599	4079	111	4789
04-Jan	595	4404	114	5113
05-Jan	536	4140	114	4790
06-Jan	991	3875	85	4951
07-Jan	1319	3587	42	4948
08-Jan	1103	3519	50	4672
09-Jan	962	2628	62	3652
10-Jan	1212	3699	72	4983
11-Jan	856	3695	86	4637
12-Jan	1299	3297	47	4643
13-Jan	738	3326	59	4123
14-Jan	291	2832	32	3155
15-Jan	304	2730	24	3058
16-Jan	1191	3910	41	5142
17-Jan	16636	5706	5	22347
18-Jan	1738	4024	12	5774
19-Jan	974	3717	12	4703
20-Jan	833	3485	4	4322
21-Jan	566	3104	5	3675
22-Jan	239	3045	7	3291
23-Jan	381	2832	12	3225
Grand Total	34403	82452	1133	117988

1. We first calculated the Time Distribution by dividing each calls distribution by total calls i.e., 30.
2. The number of agents required for each time bucket is calculated by 11 * Time Distribution
3. 11 is calculated above by dividing the additional hours required to answer the night calls by 5
   (actual working hours of agents).
4. Also, while calculating, the round figure is taken into consideration as there cannot be 1.5 men working.

Analysis:

- The data revealed that evening hours see the lowest frequency of client calls. This gives the business the chance to maximize its personnel by using fewer agents to handle calls during that period.
- The business may want to think about employing fifteen committed customer service representatives who are open at night in order to accommodate the night shift.
- Another option is to shift some of the day workers to the night shift, ensuring continuous coverage and efficient call handling throughout the day.
- By adjusting the shift timings of the employees, such as having some workers from 5 am to 2 pm and others from 2 pm to 11 pm, the company can maximize the number of calls answered during peak hours.
- To ensure round-the-clock availability, the company can divide its workforce into three shifts, enabling agents to be available 24/7 for addressing customer queries and concerns.
- It is significant to remember that outliers in the data were found during the analysis. Because these outliers may have influenced the results, eliminating them could potentially result in different answers and outcomes.
- These insights give the business practical plans for improving customer service effectiveness, allocating workers more efficiently, and guaranteeing constant availability to meet client demands.

Conclusion:

To sum up, this study explored Customer Experience (CX) analytics with a particular focus on a company's inbound phone staff. Through a 23-day dataset analysis that includes agent information, queue times, call durations, and call statuses, important insights were obtained to improve the customer experience as a whole.

It is impossible to overstate the importance of the CX team in any firm. The CX team, which is in charge of examining data and customer input, is essential to the advancement of organizational policies and projects meant to increase client loyalty and happiness. The team's capacity to provide outstanding customer support is further increased by utilizing AI-powered technologies like Intelligent Routing, Predictive Analytics, Interactive Voice Response (IVR), and Robotic Process Automation (RPA).

The customer service representatives, also known as call center agents, are essential to the operations of the CX team since they manage many types of support, including inbound customer support. The project's investigation of inbound customer assistance focuses on efficiently handling incoming calls from clients in an effort to draw them in, keep them interested, and win their advocacy over the long run.

It's critical to comprehend the highly competitive advertising market, where a wide range of businesses compete for target consumers' attention on several media channels. Businesses may find low-cost advertising channels with good conversion rates using analytical insights, which will eventually increase sales and brand awareness.

Organizations may effectively enhance their customer experience (CX) initiatives, simplify their processes, and obtain a competitive advantage in the current dynamic business environment by utilizing data-driven insights and analytical techniques. Going forward, companies looking to provide outstanding customer experiences and promote long-term growth will need to make consistent investments in CX analytics.

APPENDEX

Instagram User Analytics Description

PDF presentation link:
https://drive.google.com/file/d/1i2rZedYLgkVbQ3pDtrJ3rCAEmD14Y5Jo/view?usp=drive_link

Hiring Process Analytics Description

Power point presentation link: https://docs.google.com/presentation/d/1G_zLw-7Xf-VgpX8DQqZ-vDigI_gUaip3/edit?usp=drive_link&ouid=110876986073115360474&rtpof=true&sd=true

IMDB Movie Analysis Description

Power point presentation link:
https://docs.google.com/presentation/d/1ZF8vY9or2p519Irc5fPd64_uLdbbGXc_/edit?usp=drive_link&ouid=110876986073115360474&rtpof=true&sd=true

Video presentation link:
https://drive.google.com/file/d/1EWs3sqnpjfsOqGTsRDQmUHax4JVMyNe2/view?usp=drive_link

Bank Loan Case Study

Power point presentation link:
https://docs.google.com/presentation/d/1EoJLhwNbv7pVuczmKu2HkjFGPQJdV0db/edit?usp=drive_link&ouid=110876986073115360474&rtpof=true&sd=true

Video presentation link:
https://drive.google.com/file/d/1xP7LMK6m2d6cM4DvWQfx54qesUQLpLUa/view?usp=drive_link

Impact of Car Features

Power point presentation link: https://docs.google.com/presentation/d/1dExd-a0NBxJIZdW64pfZNNmacv0-N7c/edit?usp=drive_link&ouid=110876986073115360474&rtpof=true&sd=true

Video presentation link:
https://drive.google.com/file/d/1hjyFFtKZUYrzWB6SBL8415YopkmNjXPz/view?usp=drive_link

ABC Call Volume Trend Analysis

Power point presentation link:
https://docs.google.com/presentation/d/1GMF9O2rMA0VFUfOr443JR5tBTbGBpNm6/edit?usp=drive_link&ouid=110876986073115360474&rtpof=true&sd=true

Video presentation link: https://drive.google.com/file/d/1-IKANG8559uk_thxA8Oy1IOg1P873KiP/view?usp=drive_link