



Data Analysis **PORTFOLIO**

By **Basil L**

Prepared By Basil L

[Detailed Version]

PROFESSIONAL BACKGROUND

Basil L

MBA Candidate in Business Analytics and Finance

I am currently pursuing an MBA with a focus on Business Analytics and Finance, building upon my undergraduate degree in B.Com. My skill set includes expertise in Data Analysis, Java, and Excel, among others. I have successfully completed eight projects in MySQL, Excel, and Data Analytics, each demonstrating my strong analytical abilities and proficiency in data interpretation.



TABLE OF CONTENTS

05	INSTAGRAM USER ANALYTICS DESCRIPTION	31	HIRING PROCESS ANALYSIS
06	PROBLEM OF THE PROJECT	32	PROBLEM OF THE PROJECT
07	DESIGN	33	DESIGN
08-14	FINDINGS	34-38	FINDINGS
15	ANALYSIS	39	ANALYSIS
16	CONCLUSION	40	CONCLUSION
17	OPERATION & METRIC ANALYTICS	41	IMDB MOVIE ANALYSIS
18	PROBLEM OF THE PROJECT	42	PROBLEM OF THE PROJECT
19	DESIGN	43	DESIGN
20-28	FINDINGS	44-48	FINDINGS
29	ANALYSIS	49	ANALYSIS
30	CONCLUSION	50	CONCLUSION

TABLE OF CONTENTS 2

51 BANK LOAN CASE STUDY

52 PROBLEM OF THE PROJECT

53 DESIGN

54-58 FINDINGS

59 ANALYSIS

60 CONCLUSION

61 IMPACT OF CAR FEATURES

62 PROBLEM OF THE PROJECT

63 DESIGN

64-73 FINDINGS

74 ANALYSIS

75 CONCLUSION

76 ABC CALL VOLUME TREND

77 PROBLEM OF THE PROJECT

78 DESIGN

79-82 FINDINGS

83 ANALYSIS

84 CONCLUSION



INSTAGRAM USER ANALYTICS DESCRIPTION



This project focuses on analyzing user-generated content on Instagram to gain insights into popular tags used by users. By examining the frequency of tag usage across photos, the goal is to identify trending topics and interests among Instagram users.

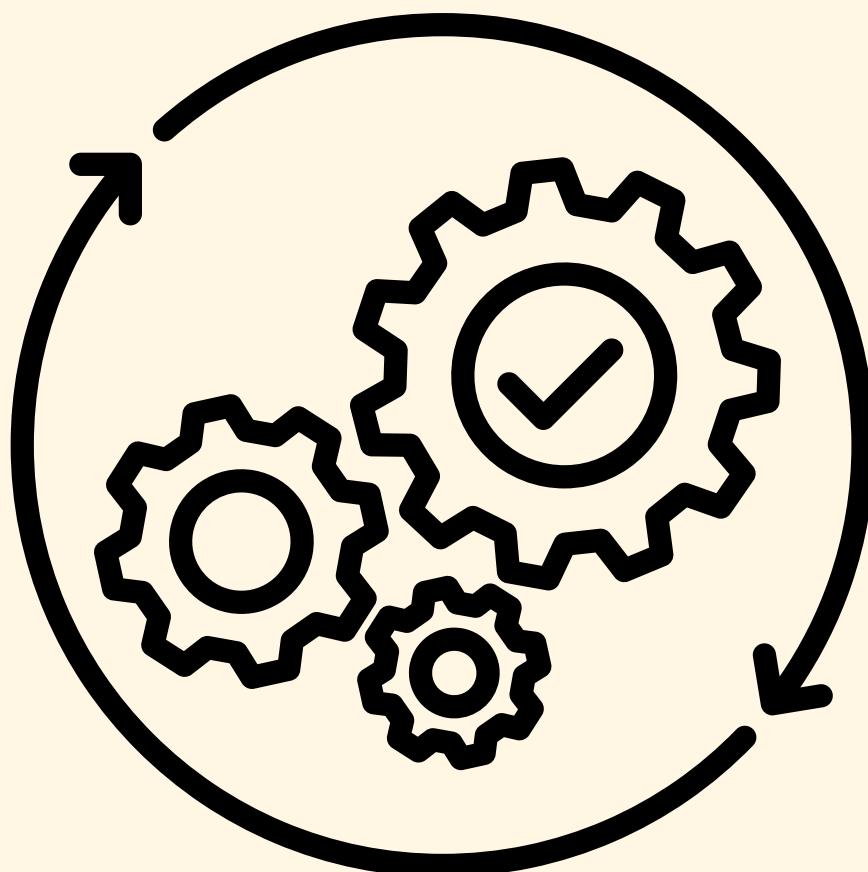


PROBLEM OF THE PROJECT

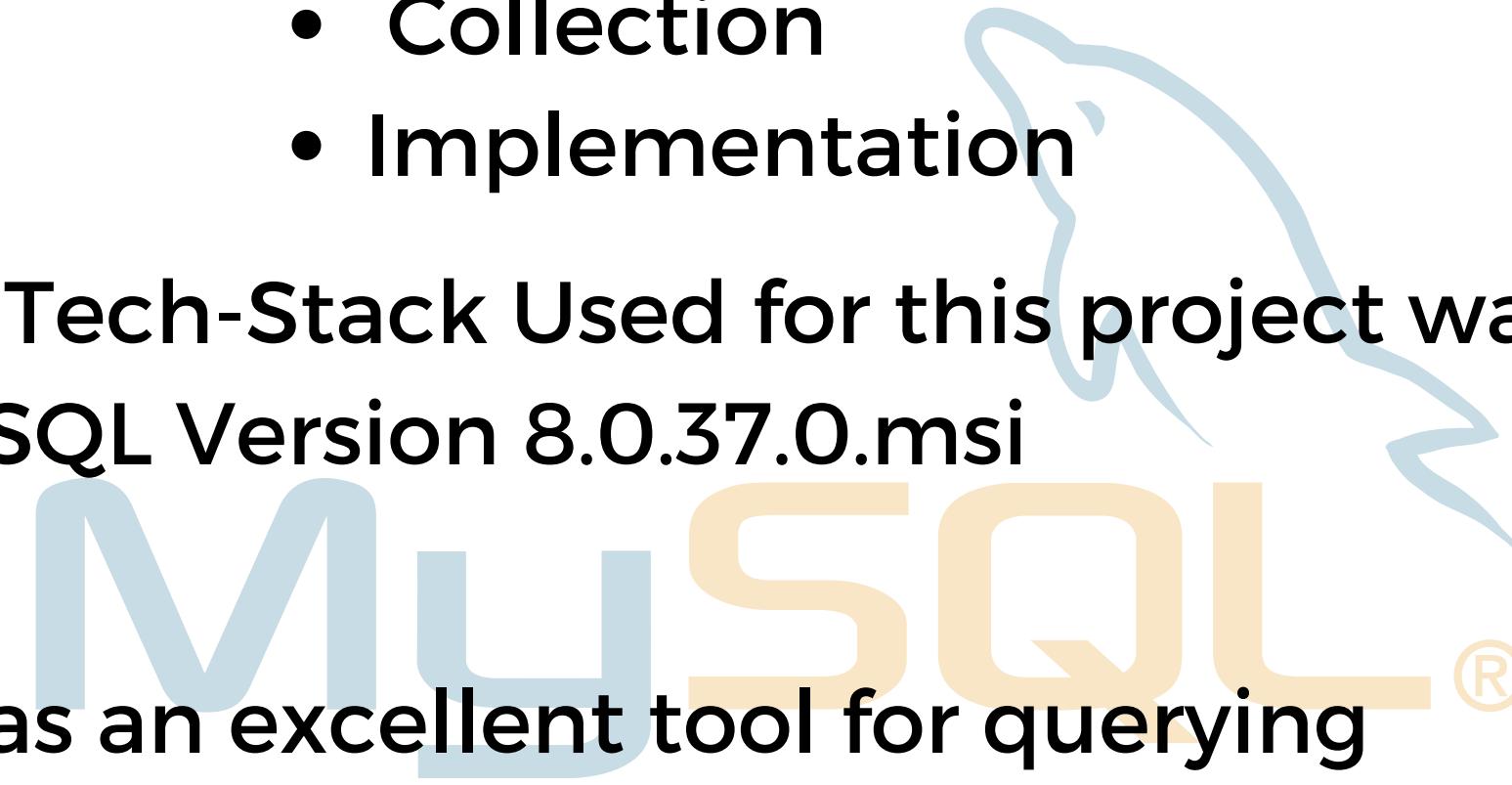


- The Problems are in Market Analysis and Investor Metrics
- The Problems in finding :-
 - **Loyal User Reward**
 - **Inactive User Engagement**
 - **Contest Winner Declaration**
 - **Hashtag Research**
 - **Ad Campaign Launch**
- And also....
 - **User Engagement**
 - **Bots & Fake Accounts**

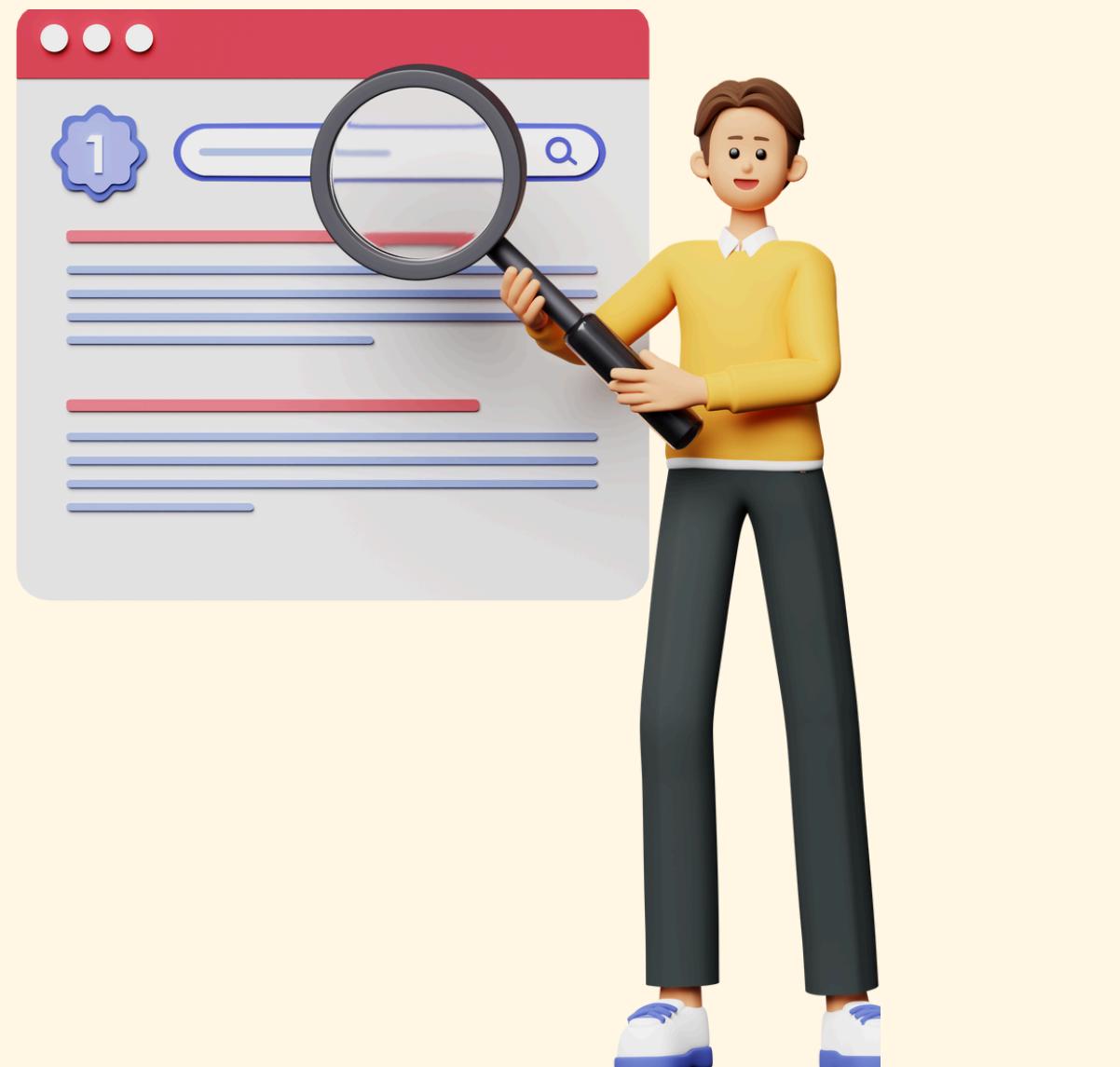
DESIGN



- This project was done by using MySQL Queries and it involves various steps like
 - Data Analysis
 - Collection
 - Implementation
- The Tech-Stack Used for this project was MySQL Version 8.0.37.0.msi
- it was an excellent tool for querying database and also used AI for some troubleshoots



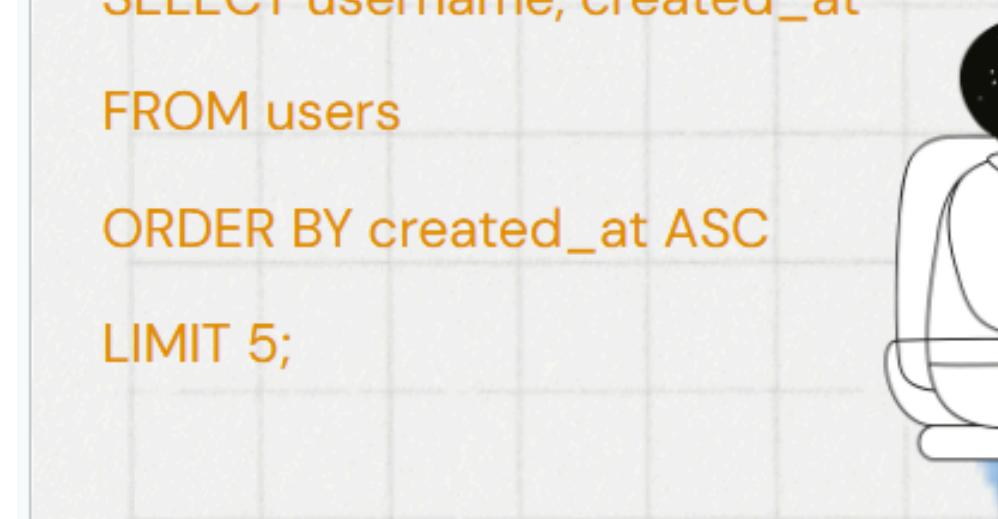
FINDINGS



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

CODE:

```
## 1-Identify the Five Oldest Users  
USE ig_clone;  
  
SELECT username, created_at  
FROM users  
  
ORDER BY created_at ASC  
  
LIMIT 5;
```



OUTPUT:

the five oldest users on Instagram from the provided database.

Darby_Herzog - 2016-05-06 00:14:21
Emilio_Bernier52 - 2016-05-06 13:04:30
Elenor88 - 2016-05-08 01:30:41
Nicole71 - 2016-05-09 17:30:22
Jordyn.Jacobson2 - 2016-05-14 07:56:26

FINDINGS



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

CODE:

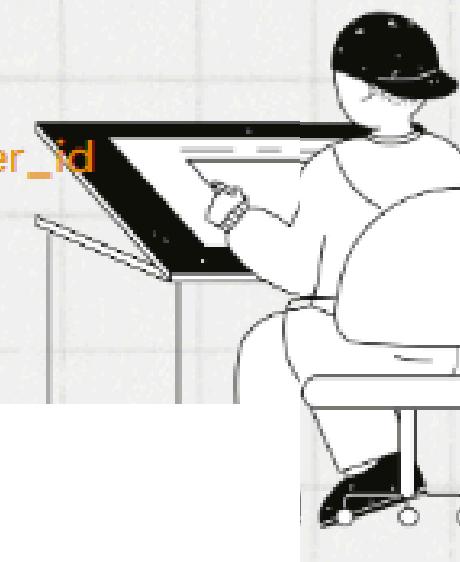
```
## 2. Identify Users Who Have Never Posted a
Single Photo
USE ig_clone;

SELECT u.username
FROM users u
LEFT JOIN photos p ON u.id = p.user_id
WHERE p.user_id IS NULL;
```

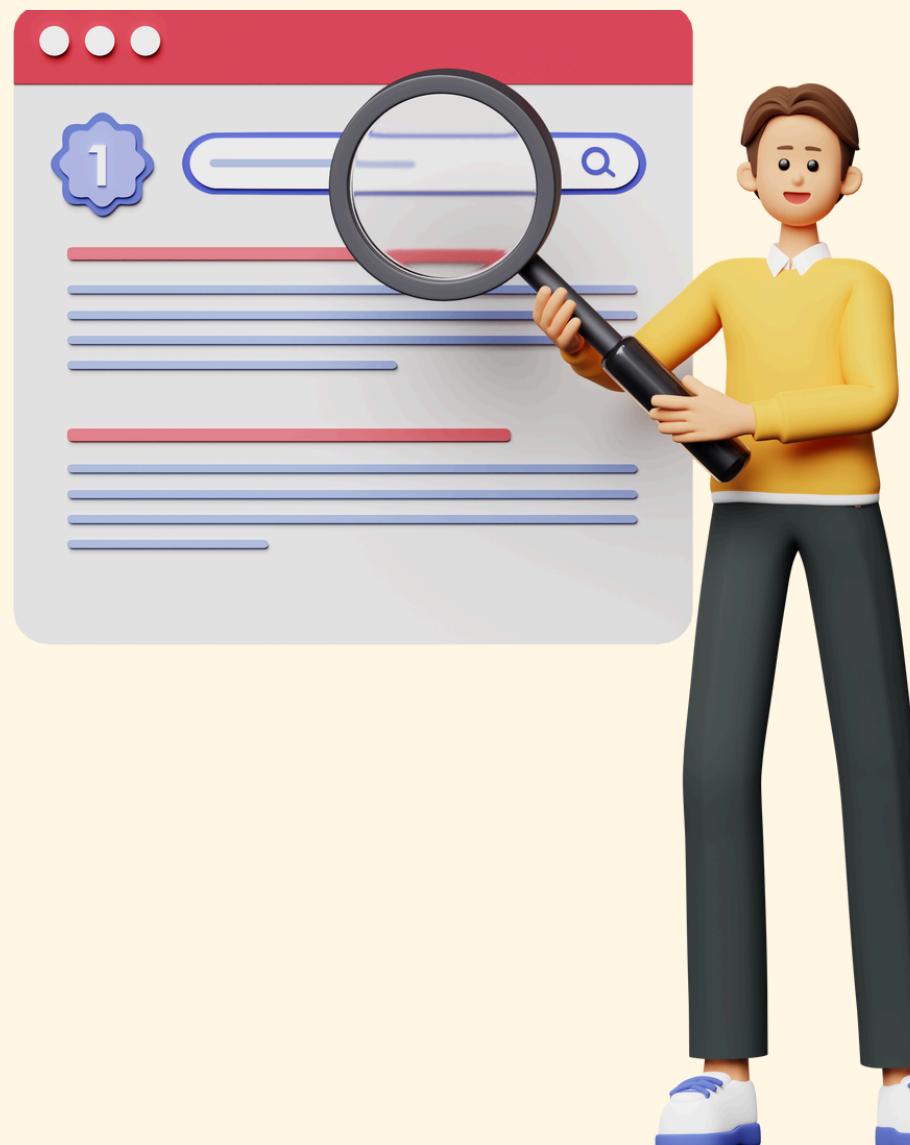
OUTPUT:

The users who have never posted a single photo on Instagram.

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



FINDINGS



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

CODE:

```
-- 3- Determine the Winner of the Contest  
(Most Likes on a Single Photo)  
USE ig_clone;  
  
SELECT p.id AS photo_id, p.user_id,  
u.username, COUNT(l.user_id) AS like_count  
FROM photos p  
JOIN likes l ON p.id = l.photo_id  
JOIN users u ON p.user_id = u.id  
GROUP BY p.id, p.user_id, u.username  
ORDER BY like_count DESC  
LIMIT 1;
```

OUTPUT:

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

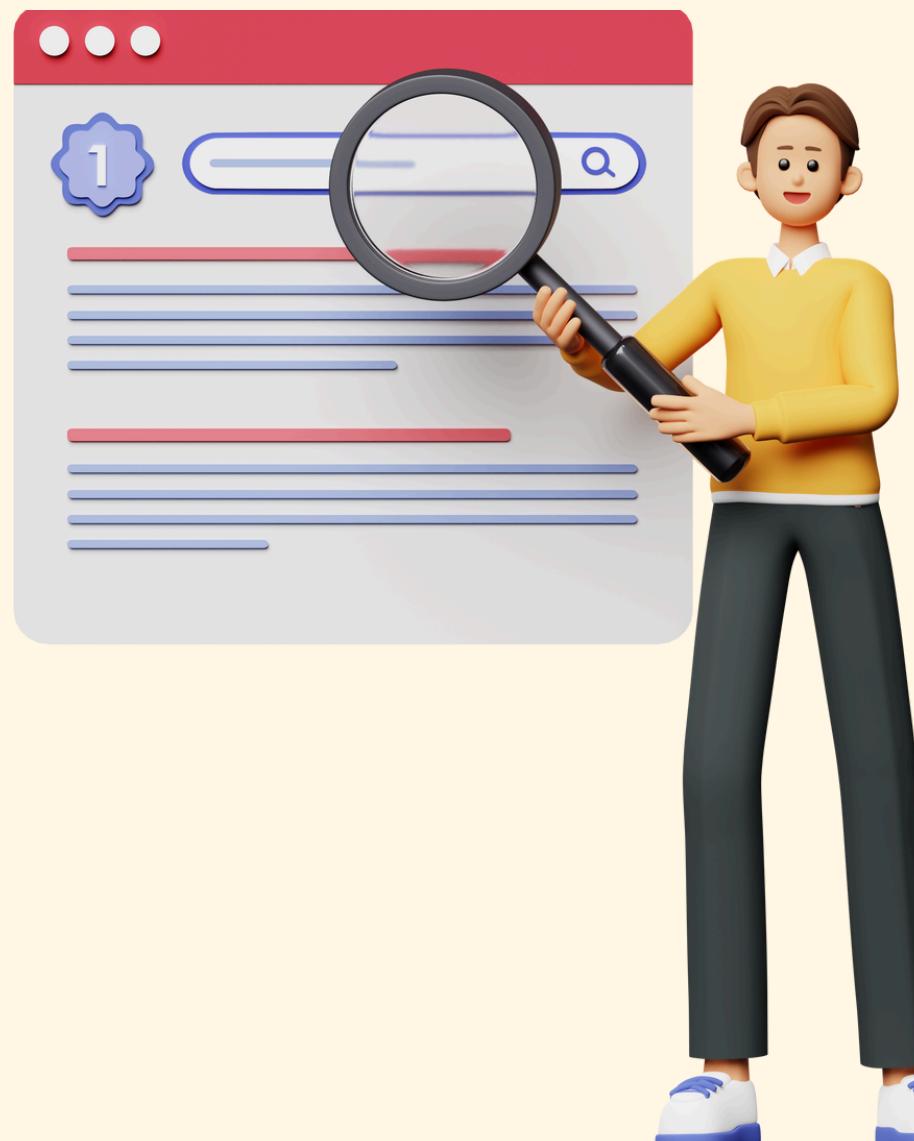
photo_id - 145

user_id - 52

username - Zack_Kemmer93

like_count - 48

FINDINGS



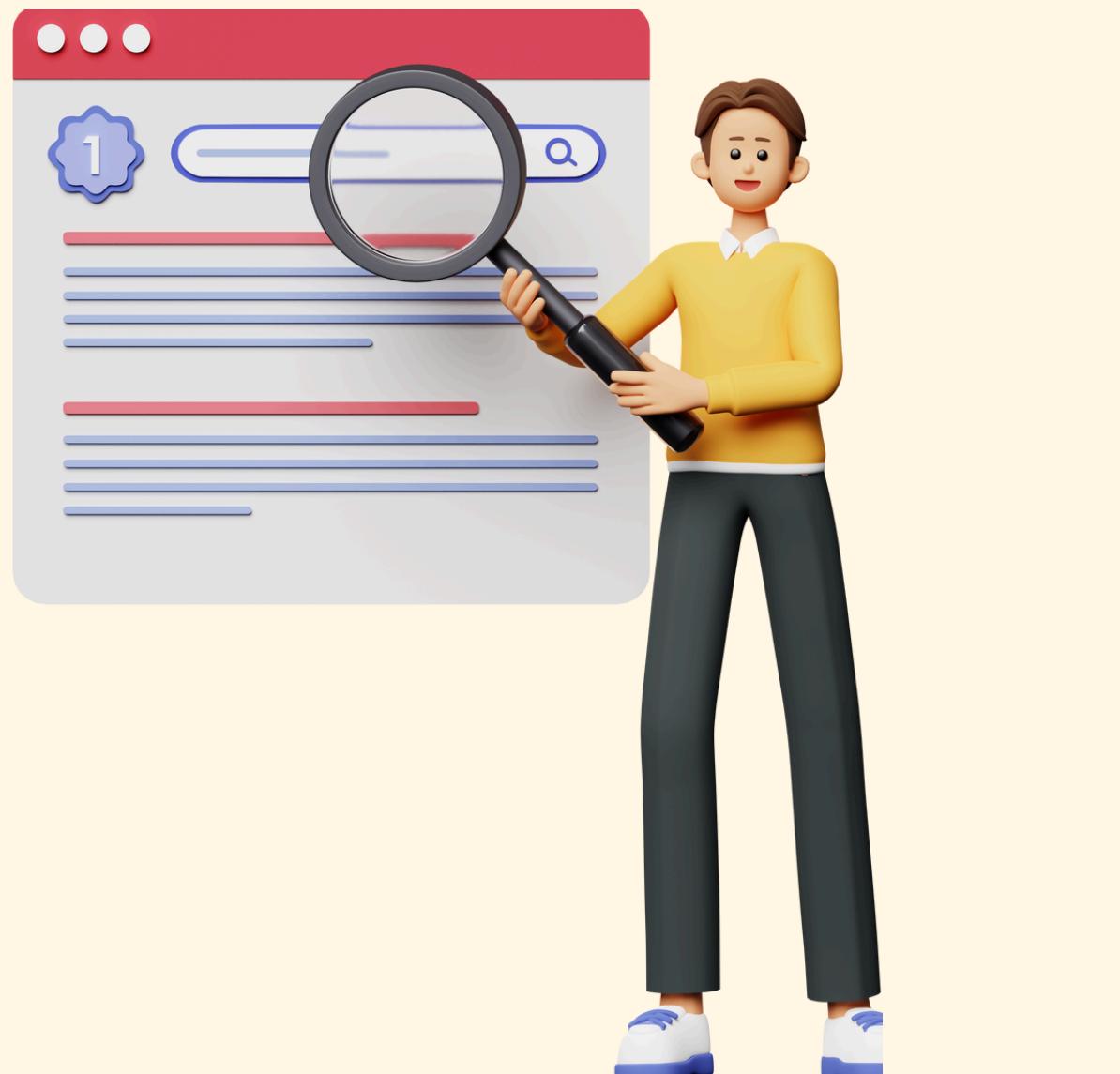
4. Identify and suggest the top five most commonly used hashtags on the platform.

CODE:

```
-- 4 Identify the top five most commonly used
hashtags
USE ig_clone;
SELECT
tags.tag_name,
COUNT(*) AS total
FROM photo_tags
JOIN tags
ON photo_tags.tag_id= tags.id
GROUP BY tags.id
ORDER BY total DESC
LIMIT 5
```



FINDINGS



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

CODE:

```
-- 5 - Determine the Best Day of the Week to
Launch Ads (Most Registrations)

USE ig_clone;

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

OUTPUT:

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

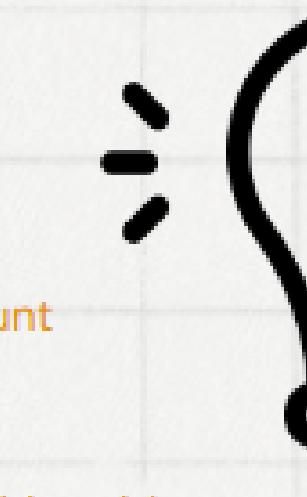
registration_day -Thursday

registrations - 16

FINDINGS



6. 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.

CODE: 

```
-- 1 Calculate the Average Number of Posts per User
SELECT AVG(post_count) AS avg_posts_per_user
FROM (
    SELECT COUNT(*) AS post_count
    FROM photos p
    INNER JOIN users u ON p.user_id = u.id
    GROUP BY u.username
) AS user_posts;
```

number

OUTPUT:

```
avg_posts_per_user
3.4730
```

FINDINGS



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

CODE:

```
-- 7 Total Number of Photos Divided by Total Number of Users  
  
SELECT  
  
    SUM(photoid) AS totalphotos,  
  
    COUNT(userid) AS total_users,  
  
    SUM(photoid) / COUNT(userid) AS photoperuser  
  
FROM (  
  
    SELECT  
  
        u.id AS userid,  
  
        COUNT(p.id) AS photoid  
  
    FROM users u  
  
    LEFT JOIN photos p ON p.user_id = u.id  
  
    GROUP BY u.id  
  
) AS base;
```

OUTPUT:

totalphotos - 257

total_users - 100

photoperuser - 2.5700

ANALYSIS



The analysis of the most commonly used tags provides valuable insights into user behavior and content trends on Instagram. By leveraging these findings, Instagram can enhance user engagement, improve content discoverability, and explore new monetization opportunities. The strategic implementation of these insights will help Instagram maintain its position as a leading social media platform and continue to grow its user base.

CONCLUSION



The analysis of commonly used tags on Instagram offers valuable insights that can drive strategic improvements in user engagement, content visibility, and monetization opportunities. By applying these findings effectively, Instagram can strengthen its position as a top social media platform and continue expanding its user base.

OPERATION & METRIC ANALYTICS

The project aims to provide actionable insights to different departments within a company, such as operations, support, and marketing. This project focuses on advanced SQL analysis to extract actionable insights from various datasets, aiding in operational efficiency and understanding sudden changes in key metrics.

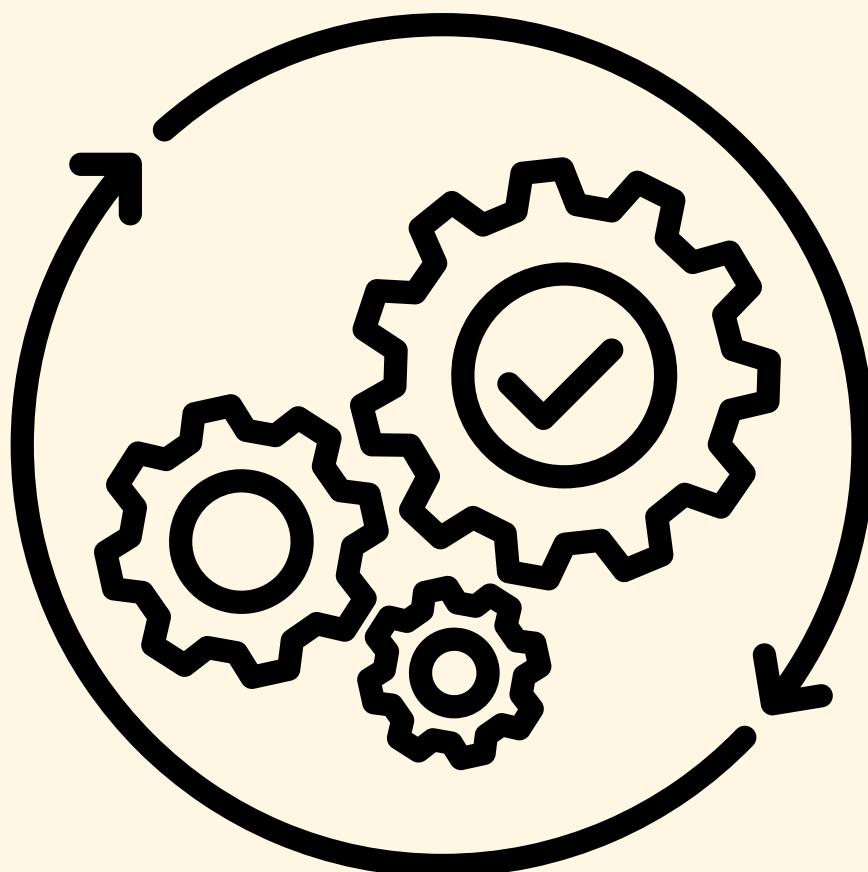


PROBLEM OF THE PROJECT

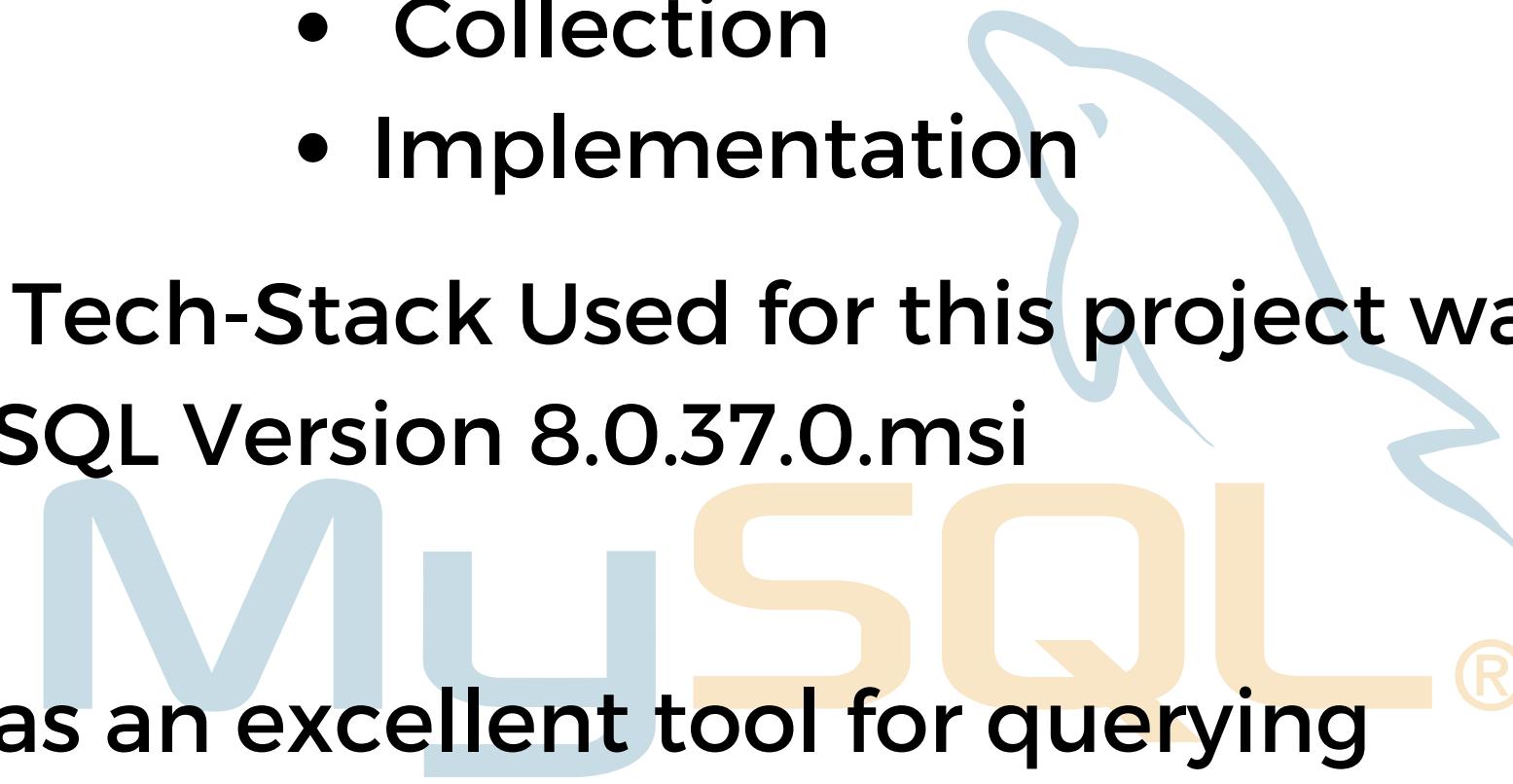


- The Problems in finding :-
 - Calculating the number of jobs reviewed per hour for each day
 - Calculate the 7-day rolling average of throughput (number of events per second).
 - Calculate the 7-day rolling average of throughput (number of events per second).
 - Calculate the percentage share of each language in the last 30 days.
 - Identify duplicate rows in the data.
 - ETC.....

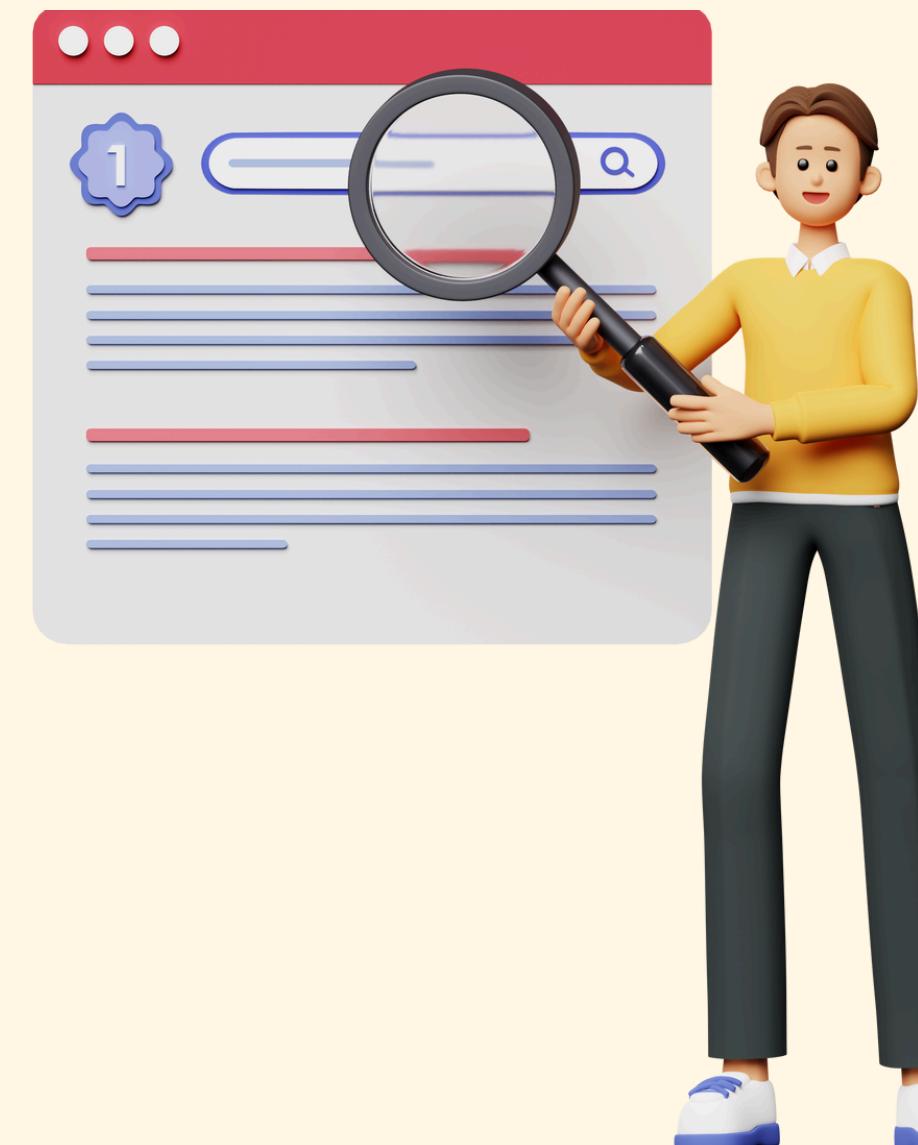
DESIGN



- This project was done by using MySQL Queries and it involves various steps like
 - Data Analysis
 - Collection
 - Implementation
- The Tech-Stack Used for this project was MySQL Version 8.0.37.0.msi
- it was an excellent tool for querying database and also used AI for some troubleshoots



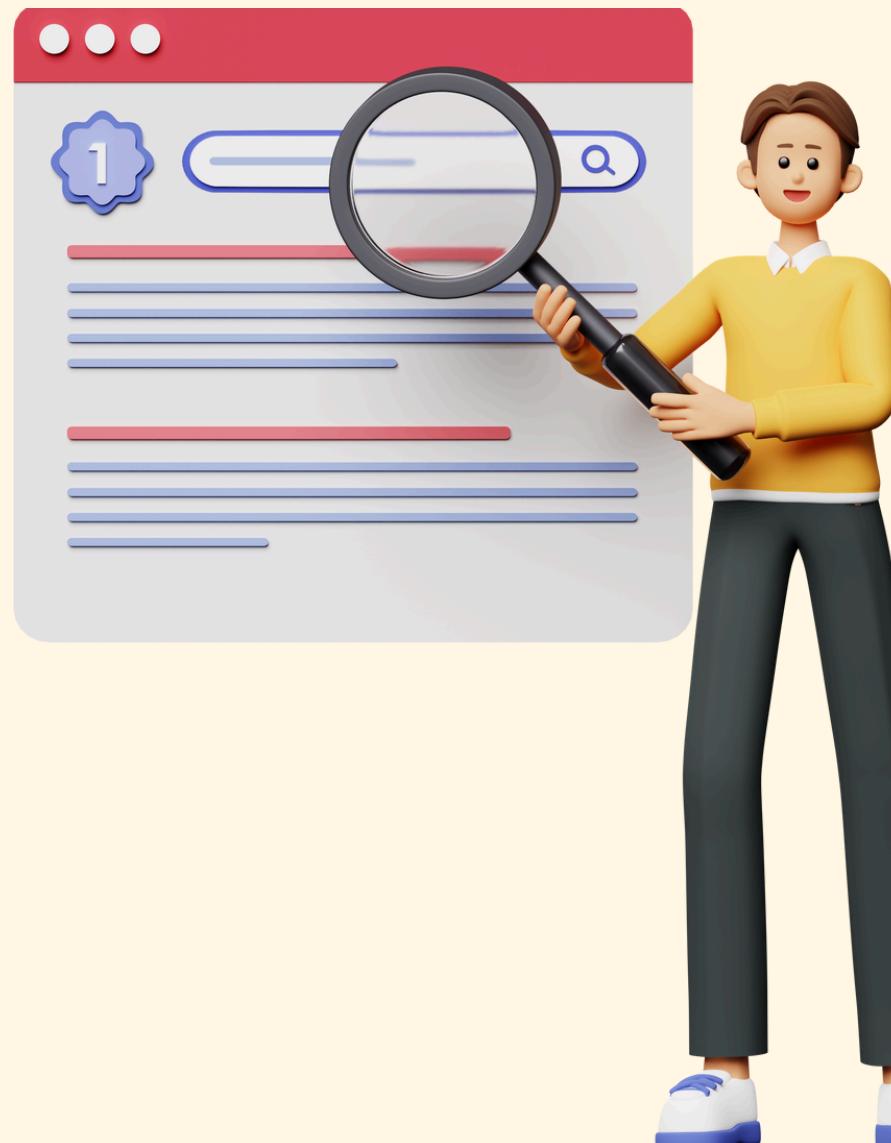
FINDINGS



1. Calculate the number of jobs reviewed per hour for each day in November 2020.

CODE:	
SELECT	
DATE_FORMAT(STR_TO_DATE(ds, '%m/%d/%Y'),	
'%m/%d/%Y %H:00:00') AS review_hour,	
COUNT(*) AS jobs_reviewed	
FROM	
job_data	
WHERE	
ds LIKE '11/%/2020'	
GROUP BY	
review_hour	
ORDER BY	
review_hour;	
OUTPUT:	
review_hour	jobs_reviewed
11/25/2020	00:00:00 1
11/26/2020	00:00:00 1
11/27/2020	00:00:00 1
11/28/2020	00:00:00 2
11/29/2020	00:00:00 1
11/30/2020	00:00:00 2

FINDINGS



2. Calculate the 7-day rolling average of throughput (number of events per second).

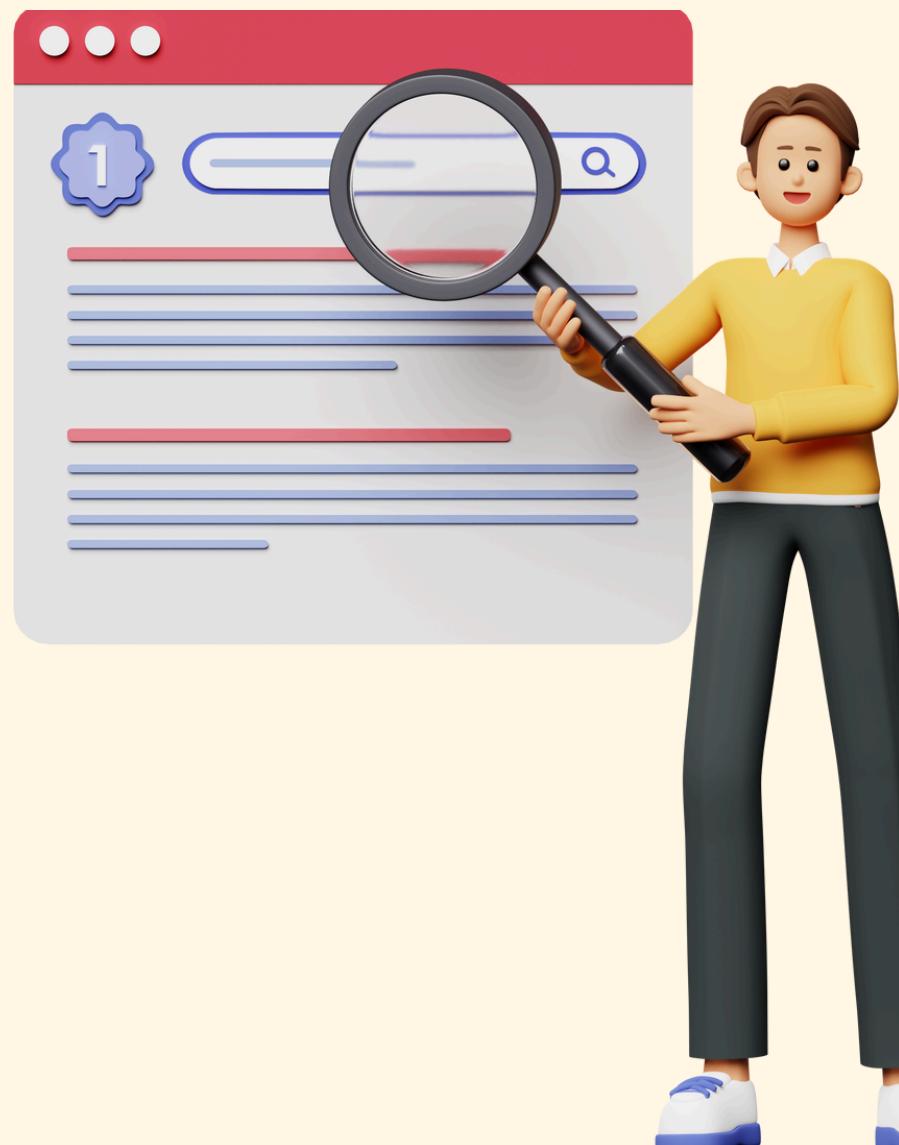
CODE:

```
SELECT ROUND  
(COUNT(event)/SUM(time_spent), 2) AS "Weekly  
Throughput" FROM job_data;  
  
SELECT ds AS Dates, ROUND  
(COUNT(event)/SUM(time_spent), 2) AS "Daily  
Throughput" FROM job_data GROUP BY ds  
ORDER BY ds;
```

OUTPUT:

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

FINDINGS



3. Calculate the percentage share of each language in the last 30 days.

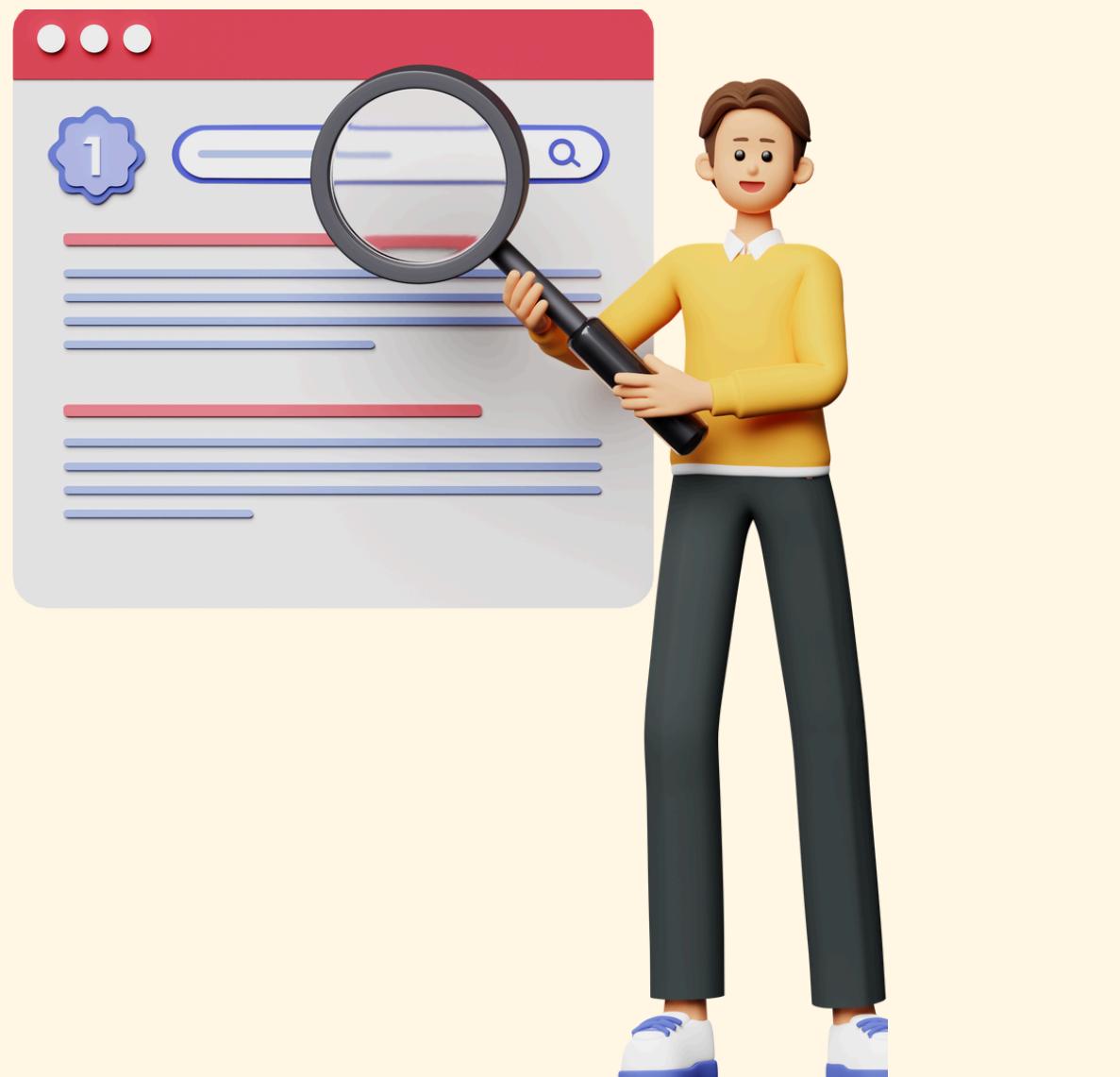
CODE:

```
SELECT language AS Languages,  
ROUND (100 * COUNT(*)/total, 2) AS  
Percentage, sub.total FROM job_data  
CROSS JOIN (SELECT COUNT(*) AS  
total FROM job_data) AS sub  
GROUP BY language, sub.total;
```

OUTPUT:

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

FINDINGS



4. Identify duplicate rows in the data.

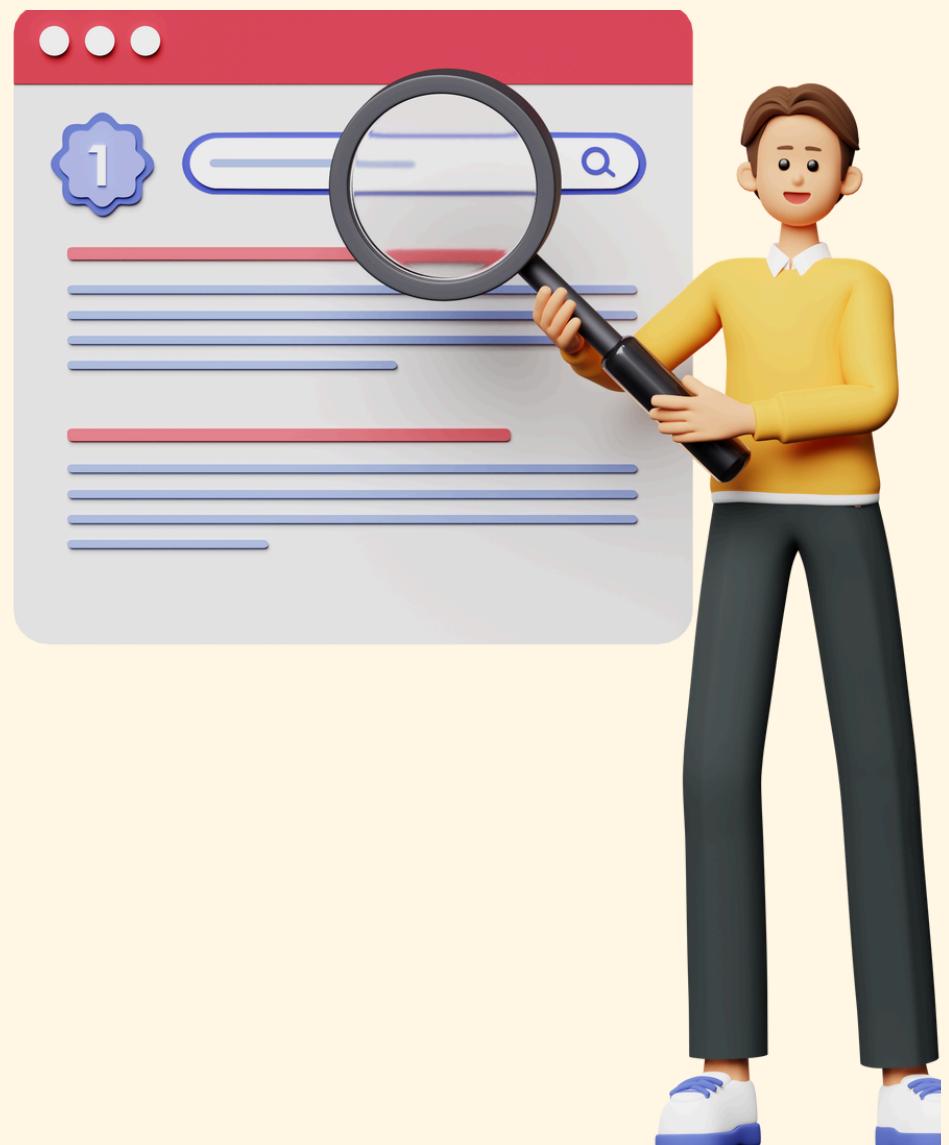
CODE:

```
SELECT actor_id, COUNT(*) AS  
Duplicates FROM job_data  
GROUP BY actor_id HAVING COUNT(*)  
>1;
```

OUTPUT:

actor_id	Duplicates
1003	2

FINDINGS



5. Measure the activeness of users on a weekly basis.

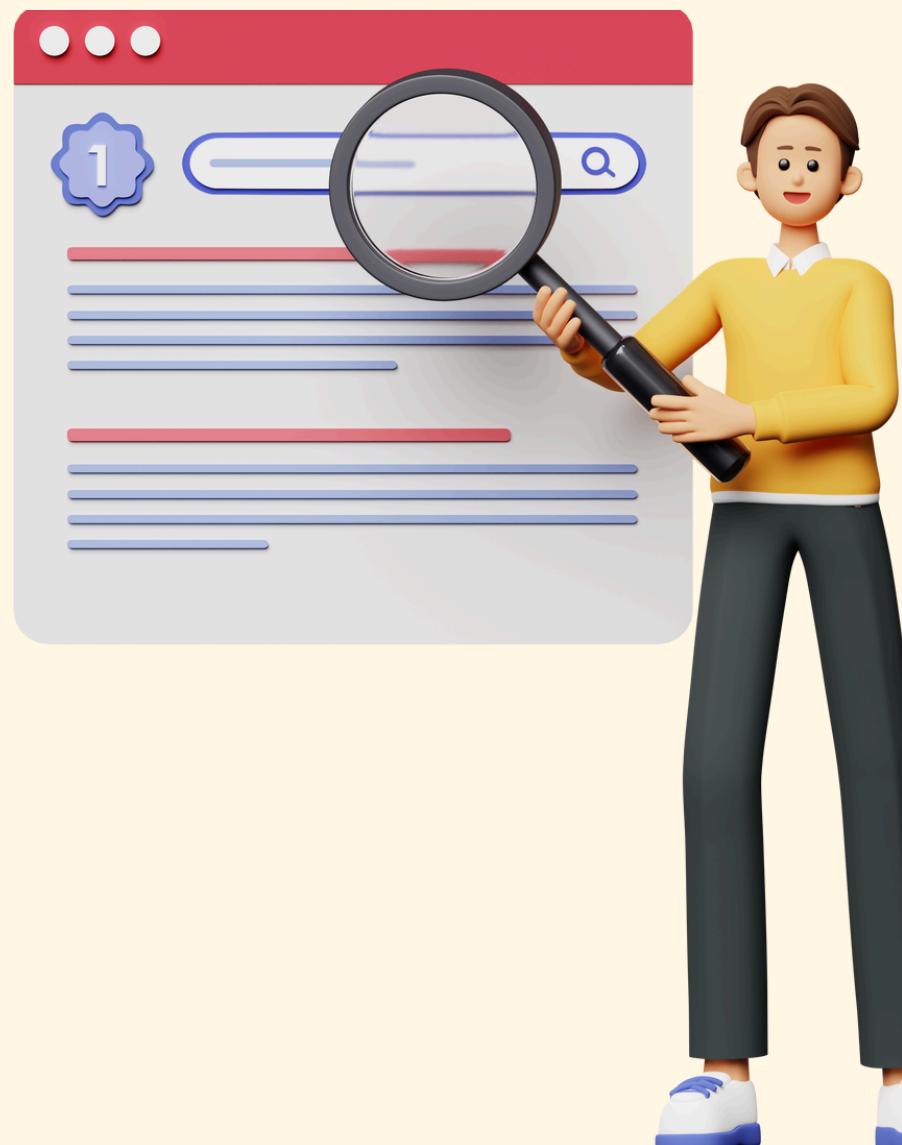
CODE:

```
SELECT
    YEARWEEK(occured_at, 1) AS week,
    user_id,
    COUNT(*) AS weekly_engagement
FROM
    events
GROUP BY
    YEARWEEK(occured_at, 1), user_id;
```

OUTPUT:

user_id	weekly_engagement
4	186
8	72
11	252
17	110
19	142
20	136
ETC.....	

FINDINGS



6. Analyze the growth of users over time for a product.

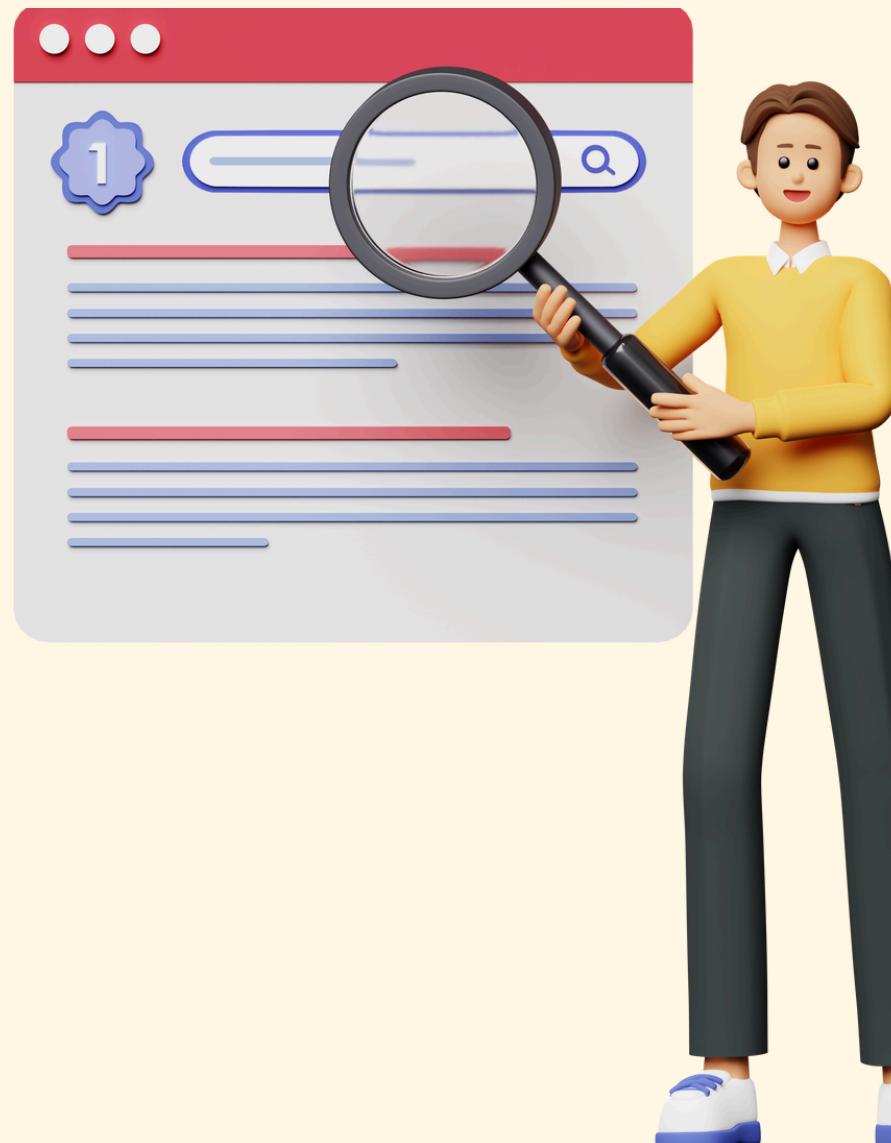
CODE:

```
SELECT  
    DATE_FORMAT(activated_at, '%d-%m-%Y')  
    AS signup_day ,  
    COUNT(*) AS user_count  
FROM users  
GROUP BY signup_day  
ORDER BY signup_day;
```

OUTPUT:

```
# signup_day  
'31-12-2013'  
user_count  
9381
```

FINDINGS



7. Analyze the retention of users on a weekly basis after signing up for a product

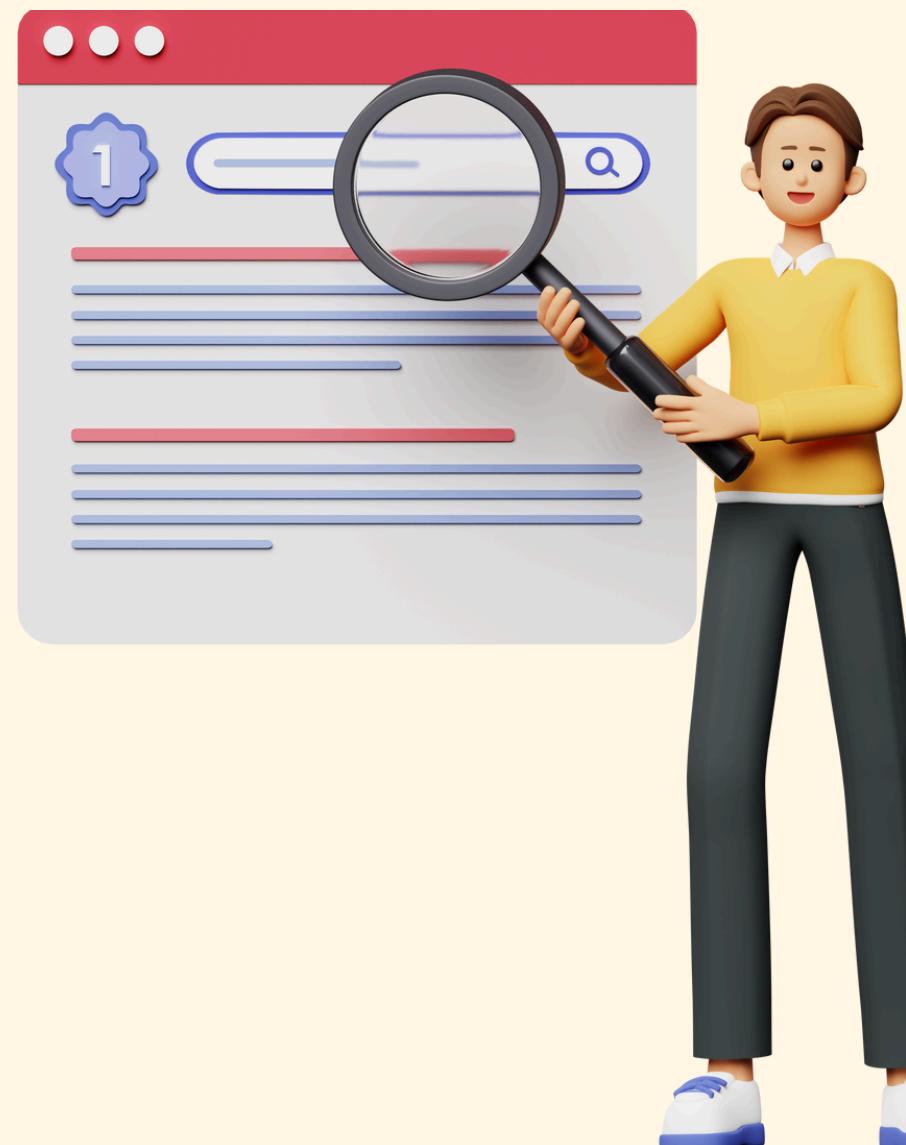
CODE:

```
SELECT
    YEARWEEK(u.created_at, 1) AS signup_week,
    YEARWEEK(e.occurred_at, 1) AS event_week,
    COUNT(DISTINCT u.user_id) AS retained_users
FROM
    users u
LEFT JOIN
    events e ON u.user_id = e.user_id
GROUP BY signup_week, event_week
ORDER BY signup_week, event_week
```

OUTPUT:

```
retained users
9381
```

FINDINGS



8. Measure the activeness of users on a weekly basis per device.

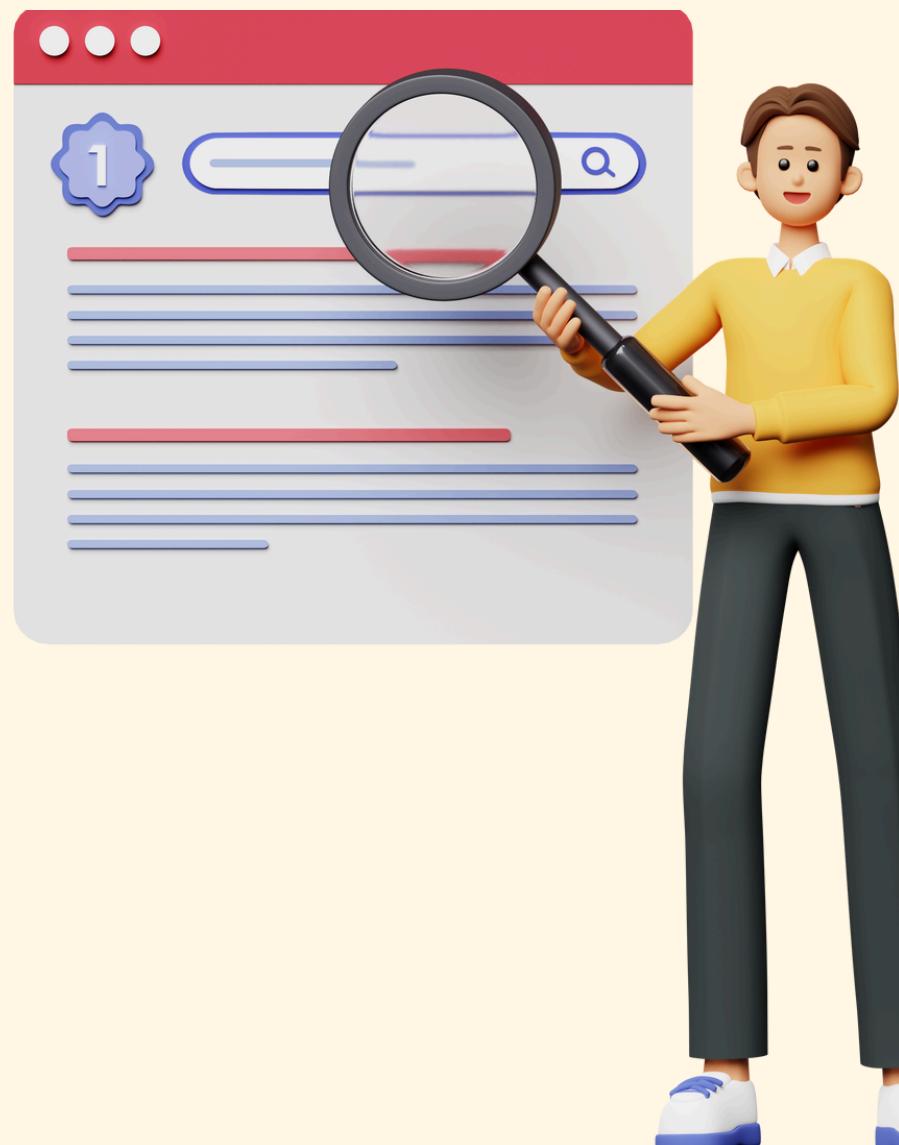
CODE:

```
SELECT
    YEARWEEK(occured_at, 1) AS week,
    device,
    COUNT(DISTINCT user_id) AS active_users
FROM events
GROUP BY YEARWEEK(occured_at, 1),
    device
ORDER BY week,
    device;
```

OUTPUT:

device	active_users
acer aspire desktop	198
acer aspire notebook	338
amazon fire phone	89
asus chromebook	355
dell inspiron desktop	360
ETC.....	

FINDINGS



9. Analyze how users are engaging with the email service

CODE:

```
SELECT
    SUM(CASE WHEN action = 'email_open' THEN 1 ELSE
0 END) AS total_emails_opened,
    SUM(CASE WHEN action = 'sent_weekly_digest' THEN
1 ELSE 0 END) AS total_emails_clicked,
    (SUM(CASE WHEN action = 'email_open' THEN 1 ELSE
0 END) /
    NULLIF(SUM(CASE WHEN action =
'sent_weekly_digest' THEN 1 ELSE 0 END), 0)) AS
open_to_click_ratio
FROM email_events;
```

OUTPUT:

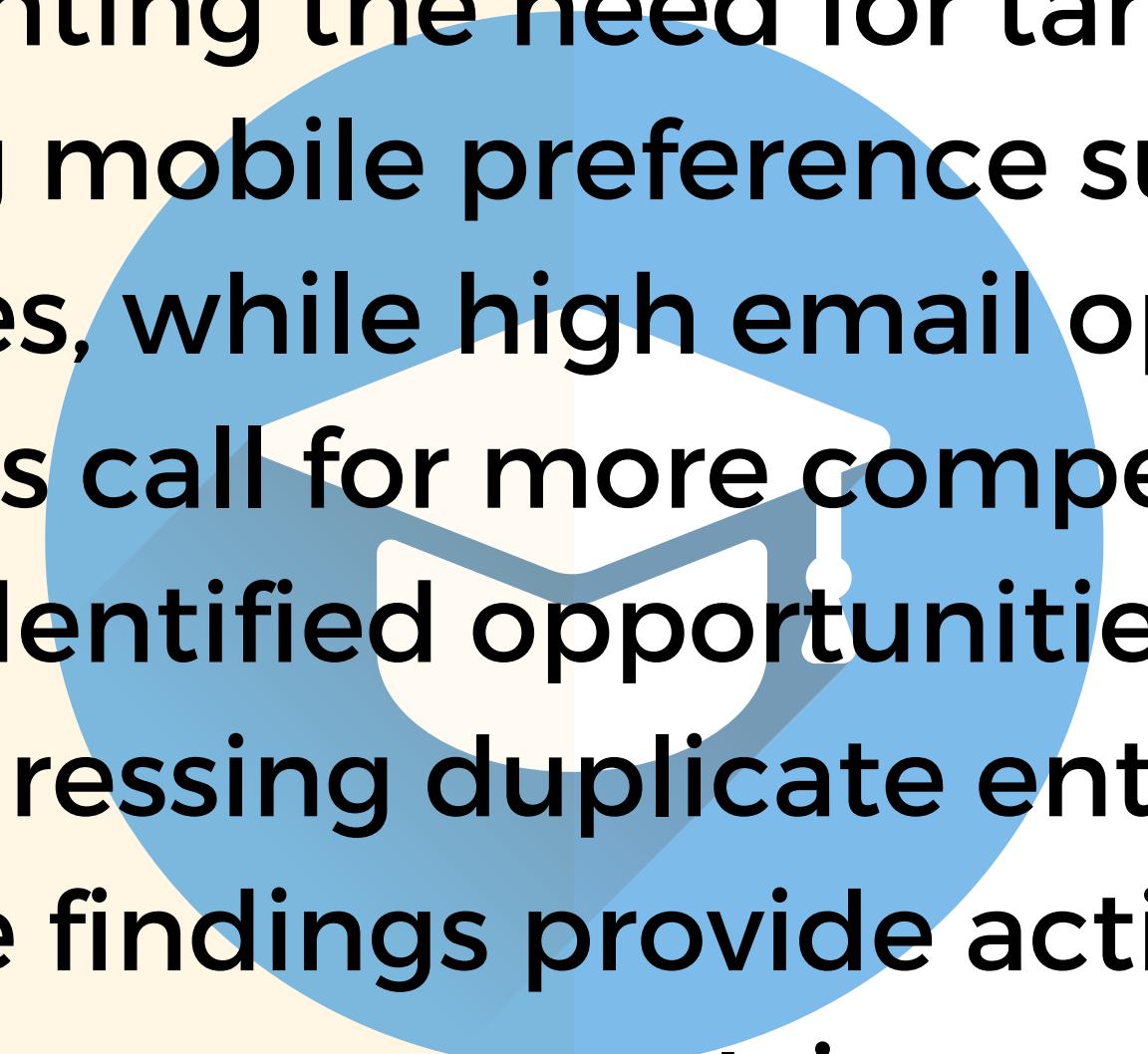
#	total_emails_opened
	20459
	total_emails_clicked
	57267
	open_to_click_ratio
	0.3573

ANALYSIS



In this project, we conducted a comprehensive analysis of user engagement, growth, and email effectiveness, as well as job data metrics. We identified key trends in weekly user activity, observed steady user growth with challenges in retention, and discovered a strong preference for mobile devices among users. Our email engagement analysis revealed high open rates but indicated opportunities for improving click-through rates. Additionally, we calculated job review metrics and analyzed language share and duplicate rows in job data, providing insights into operational efficiency. These findings offer actionable recommendations for enhancing user engagement, refining marketing strategies, and improving data handling processes.

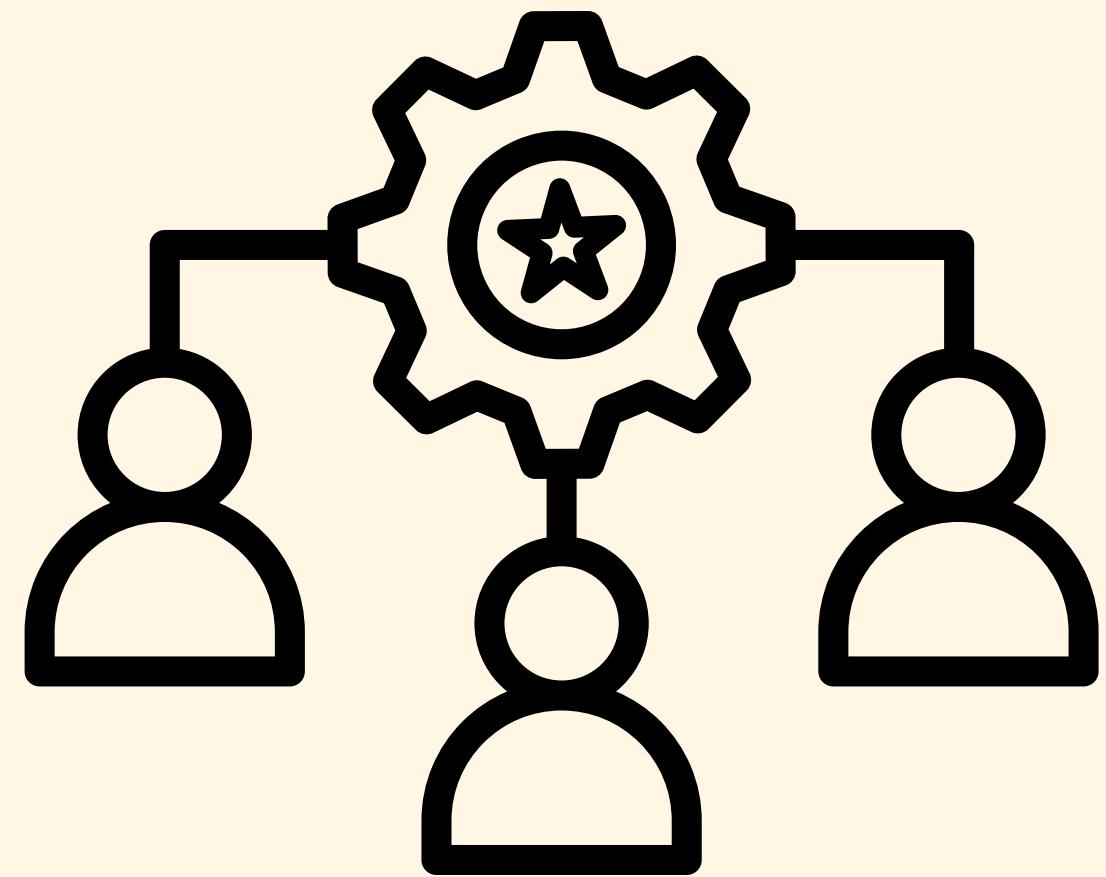
CONCLUSION



the analysis revealed steady user growth but retention challenges, highlighting the need for targeted engagement strategies. A strong mobile preference suggests optimizing mobile experiences, while high email open rates with low click-through rates call for more compelling content. Job data analysis identified opportunities to streamline operations by addressing duplicate entries and language distribution. These findings provide actionable insights to enhance user engagement, improve marketing effectiveness, and optimize data management.

HIRING PROCESS ANALYTICS

This project analyzes the hiring process data of a multinational company. The objective is to draw meaningful insights from the data to help the company improve its hiring processes.

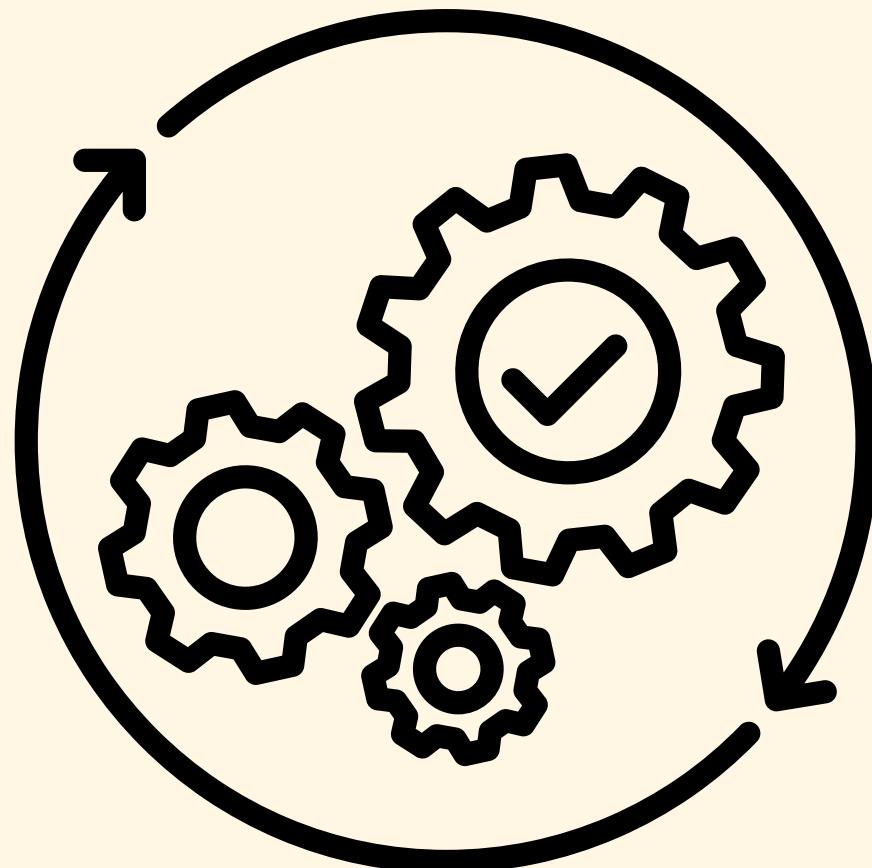


PROBLEM OF THE PROJECT



- The Problems in finding :-
 - **Gender distribution of hiring.**
 - **Average salary or each department.**
 - **Salary distribution of each department.**
 - **Size of each department**
 - **The position tier**

DESIGN



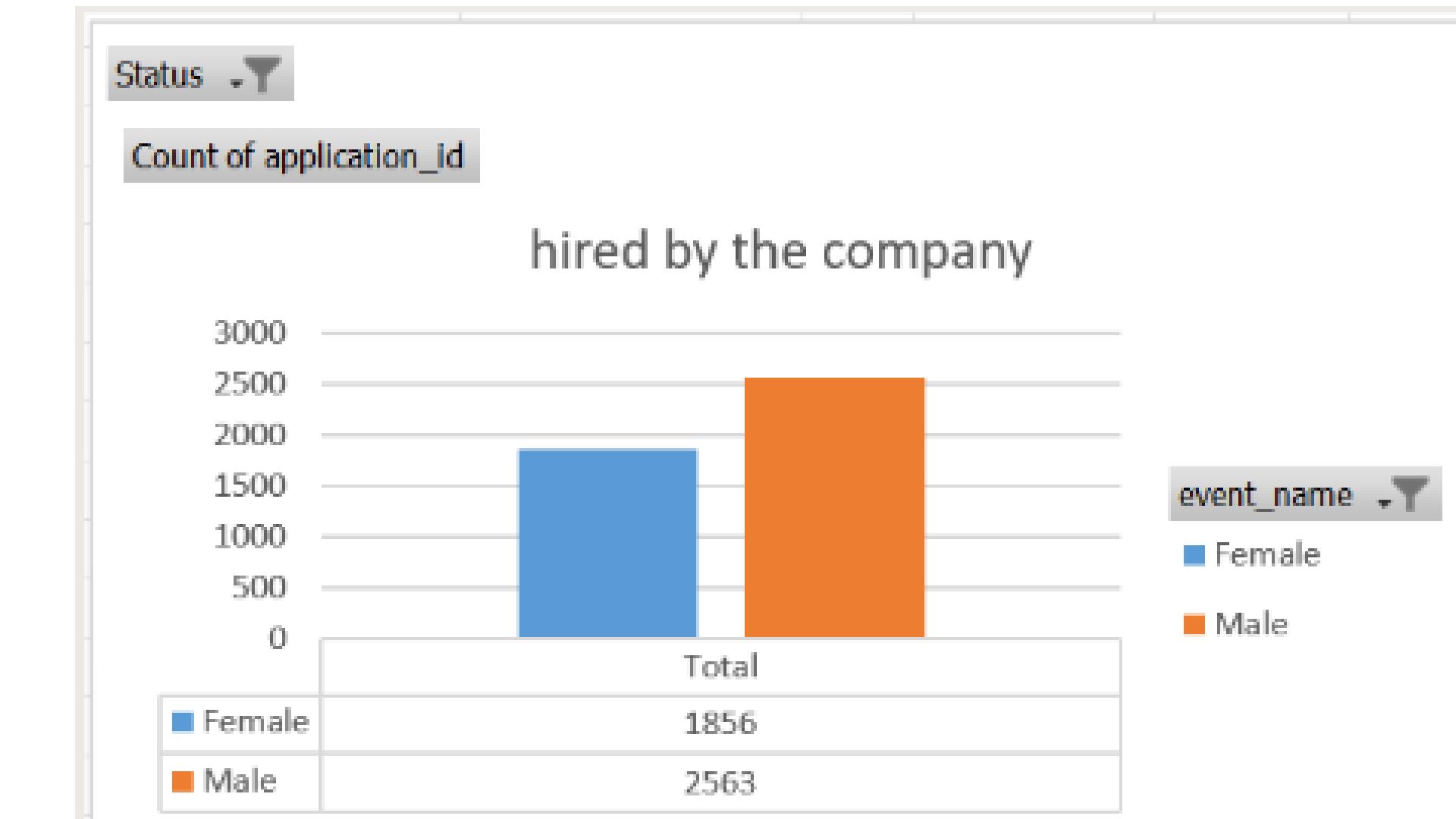
- This project was done by using MS EXCEL 2022
- XLS
 - Data Analysis
 - Collection
 - Implementation
- It is used for data analysis, creating pivot tables, and visualizations.

FINDINGS

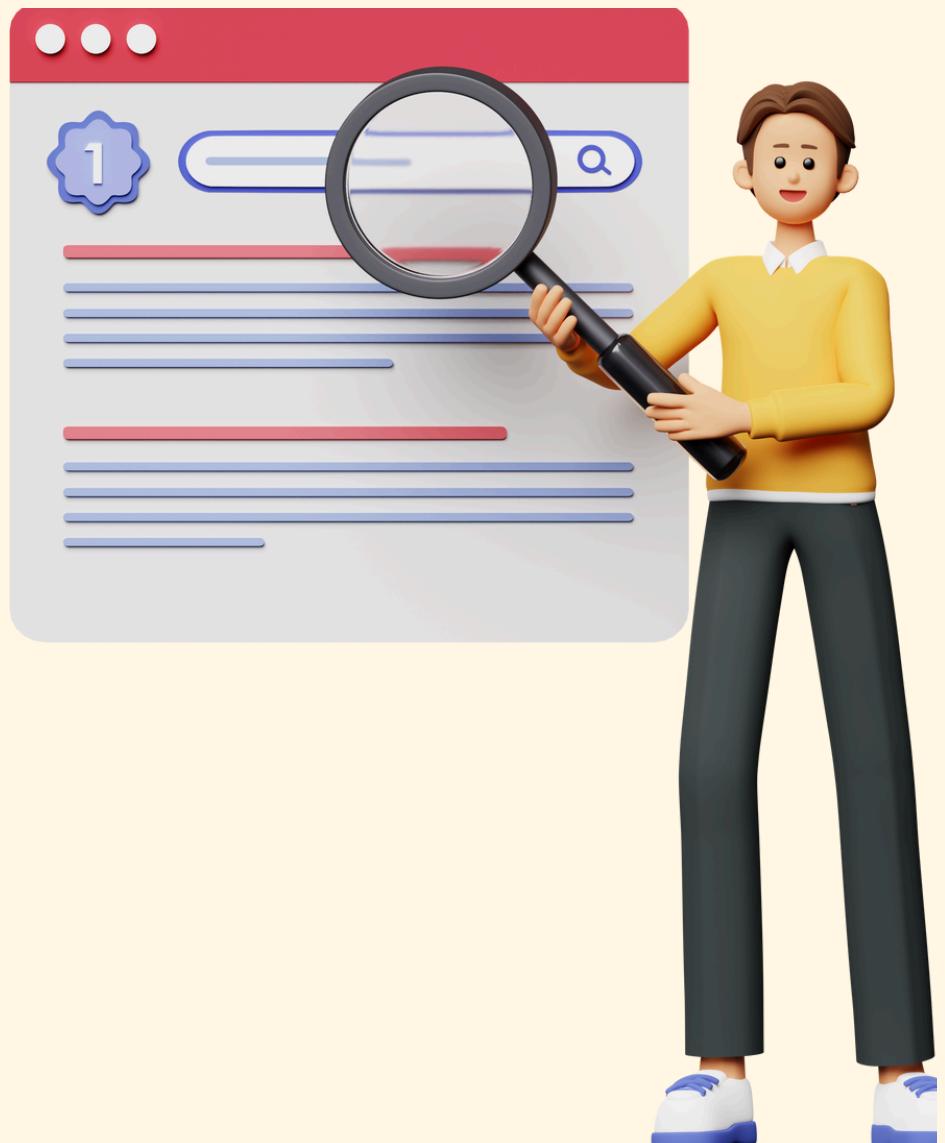


**1. Determine the gender distribution of hires.
How many males and females have been
hired by the company?**

Status	Hired	
Column Labels		
	Female	Male
Count of application_id	1856	2563
		Grand Total
		4419

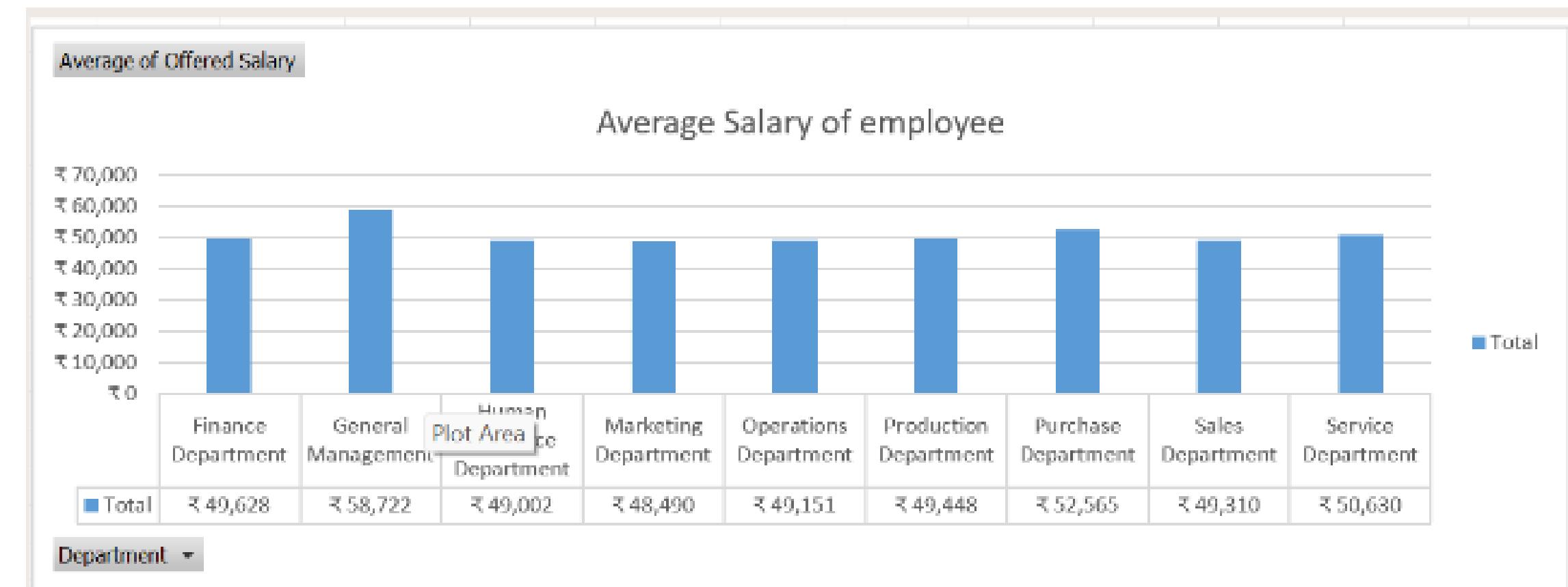


FINDINGS



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

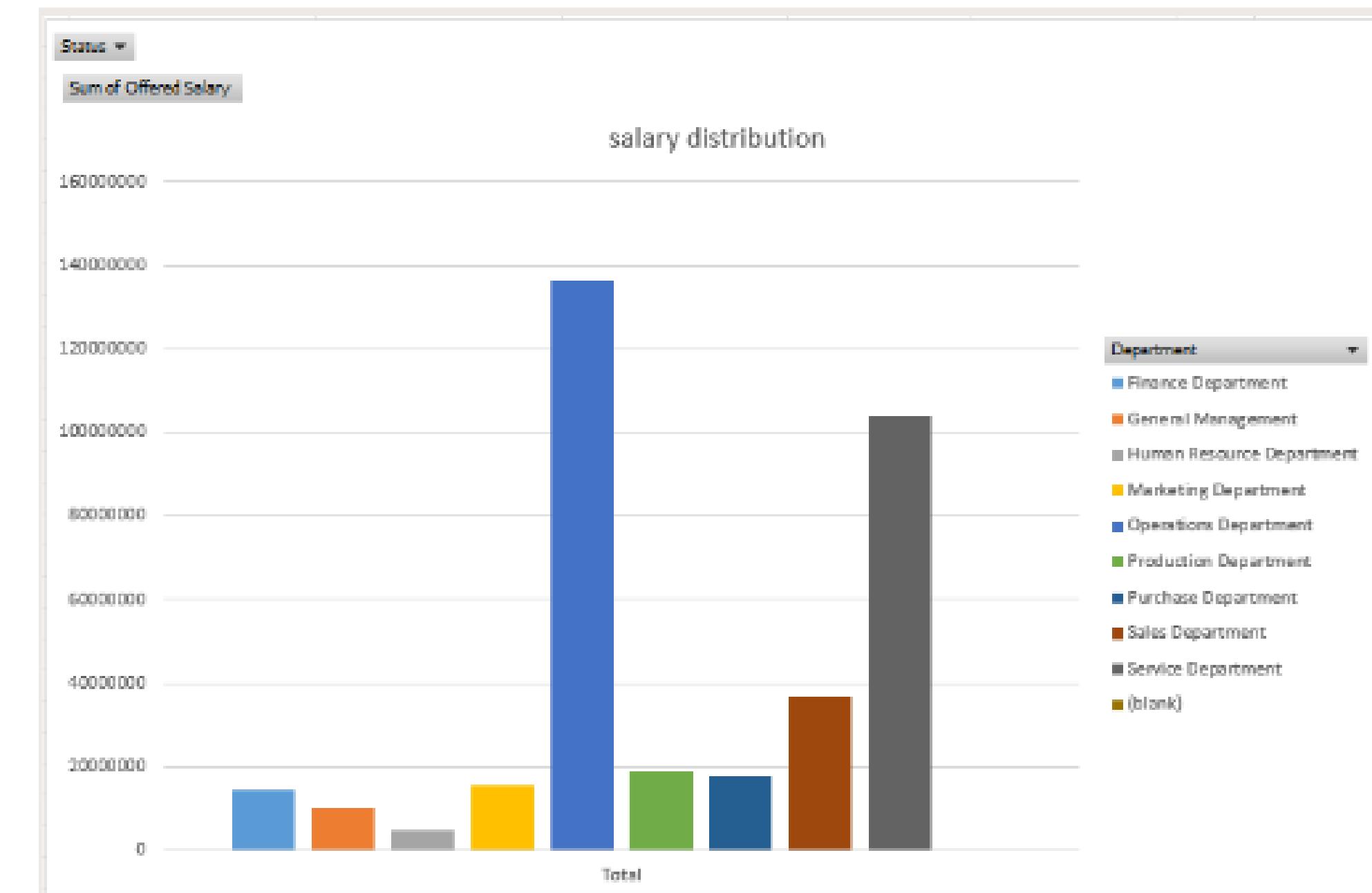
Row Labels	Average of Offered Salary
Finance Department	₹ 49,628
General Management	₹ 58,722
Human Resource Department	₹ 49,002
Marketing Department	₹ 48,490
Operations Department	₹ 49,151
Production Department	₹ 49,448
Purchase Department	₹ 52,565
Sales Department	₹ 49,310
Service Department	₹ 50,630
Grand Total	₹ 49,983



FINDINGS



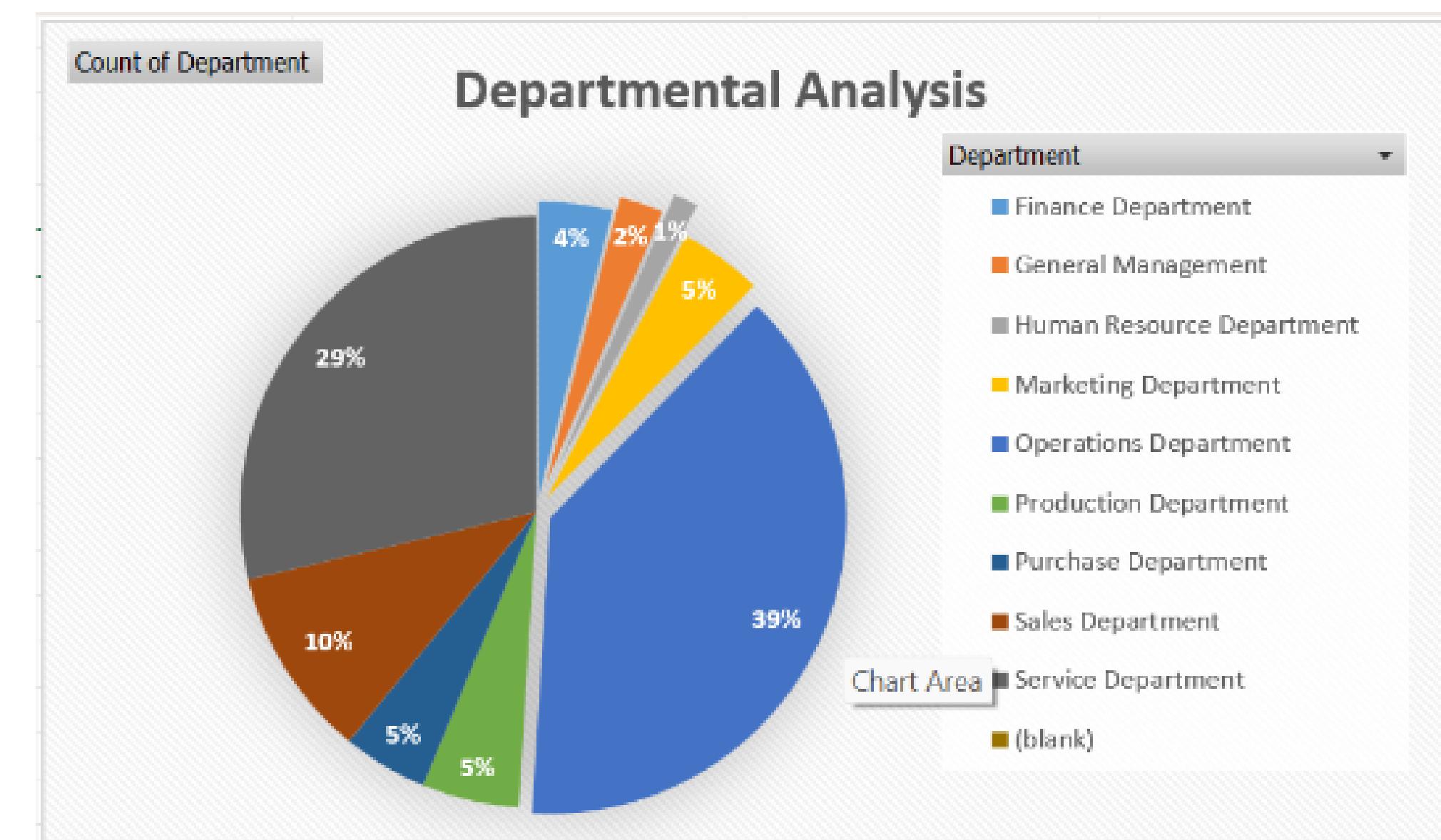
3. Create class intervals for the salaries in the company. This will help you understand the salary distribution.



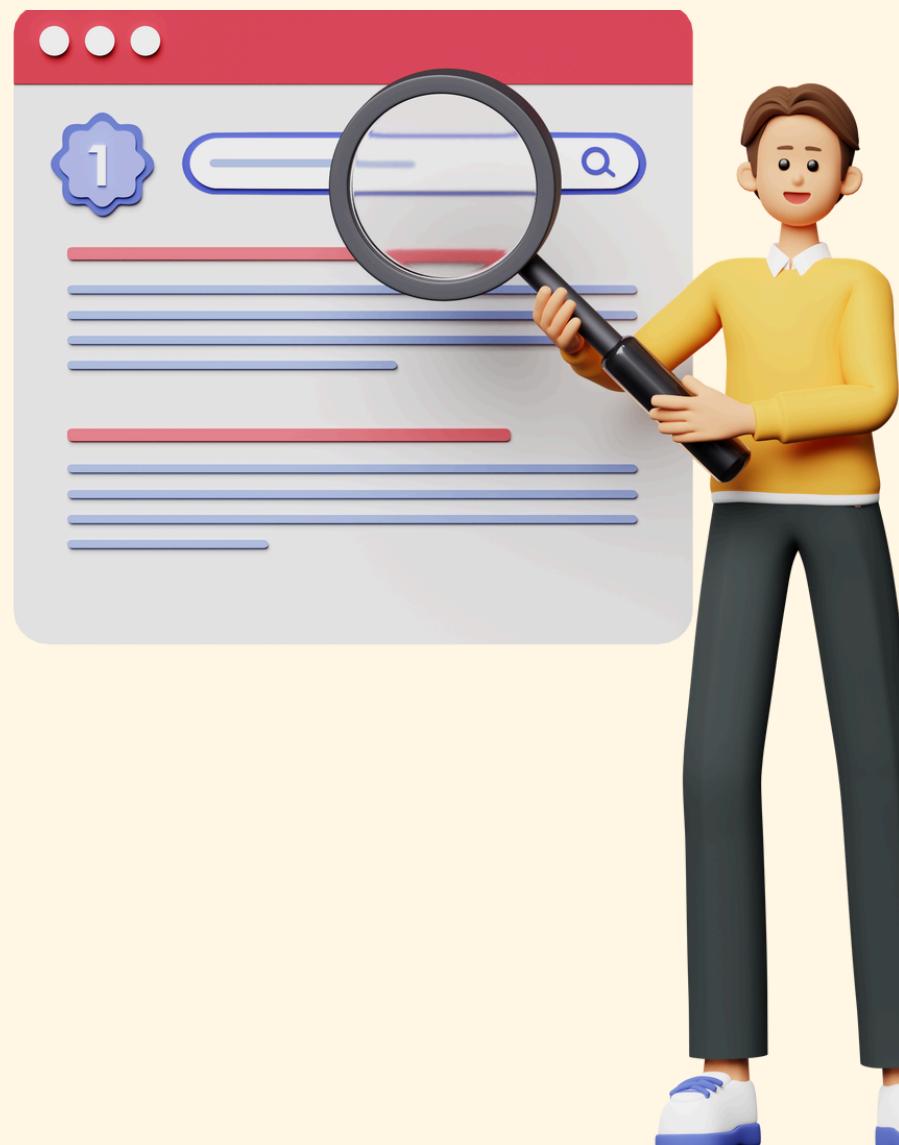
FINDINGS



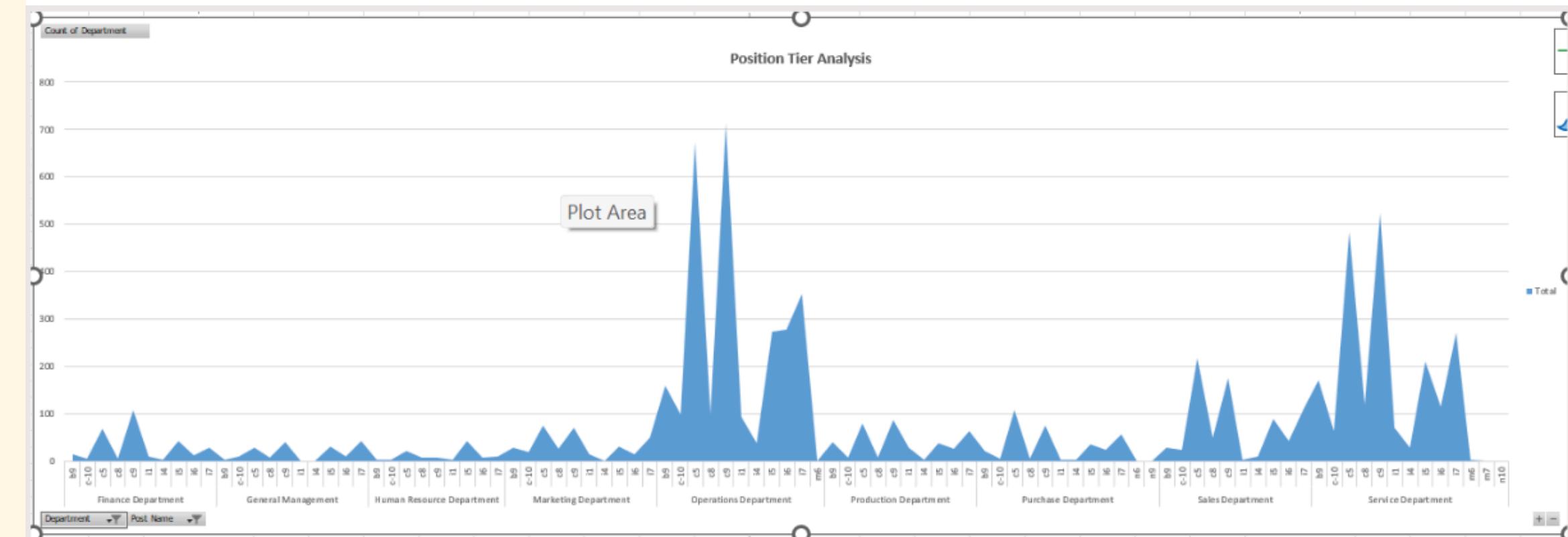
4. Use a pie chart, bar graph, or any other suitable visualization to show the proportion of people working in different departments.



FINDINGS



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



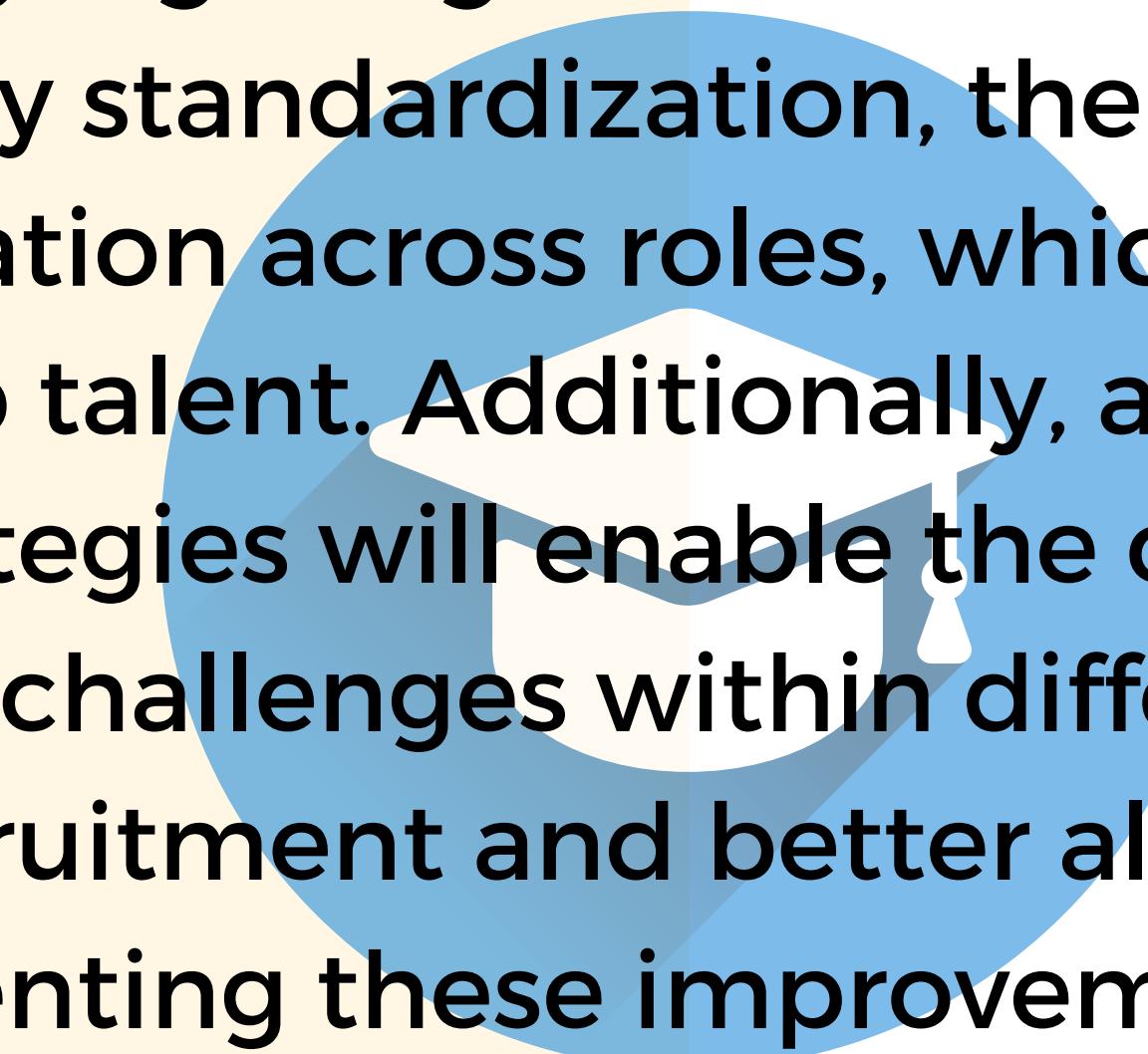
ANALYSIS



The analysis provided a comprehensive understanding of the hiring process. It identified key areas where the company can improve, such as salary standardization and department-specific hiring strategies.

This project has contributed to a deeper understanding of the hiring analytics and has provided actionable insights for the company's leadership to consider in future hiring decisions.

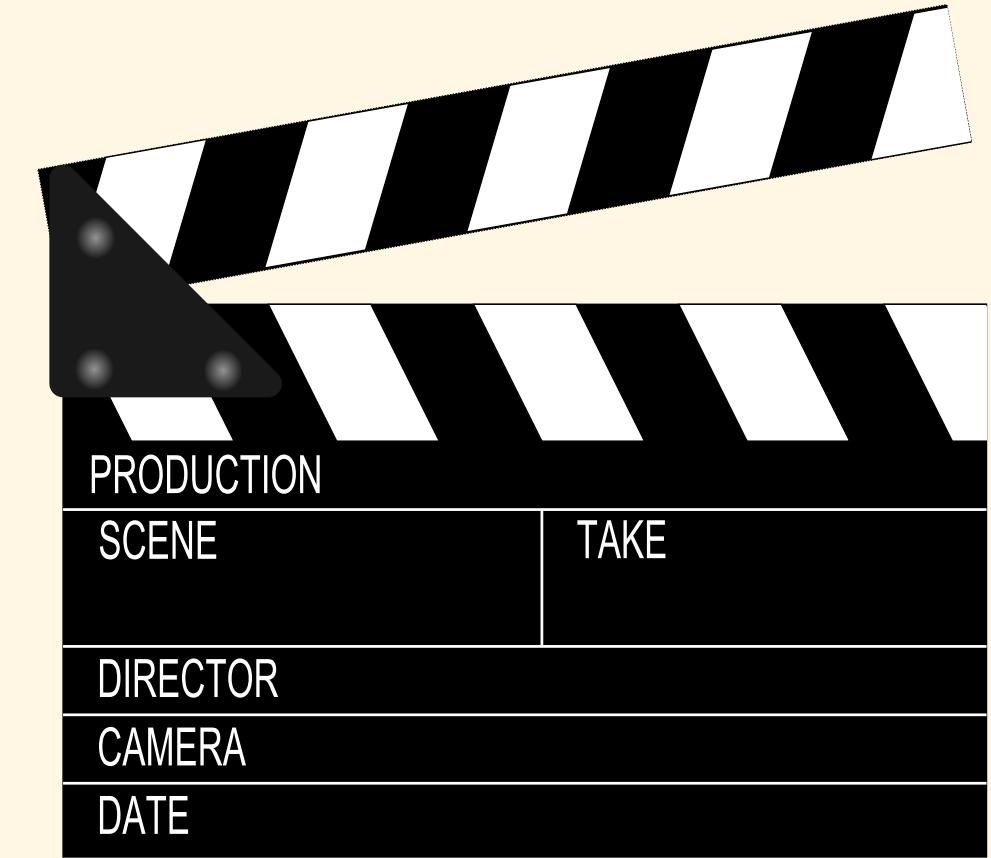
CONCLUSION



The analysis has delivered valuable insights into the company's hiring process, highlighting crucial areas for enhancement. By focusing on salary standardization, the company can ensure equitable compensation across roles, which may help in attracting and retaining top talent. Additionally, adopting department-specific hiring strategies will enable the organization to address unique needs and challenges within different teams, leading to more effective recruitment and better alignment with business objectives. Implementing these improvements will strengthen the overall hiring process, contributing to a more efficient and competitive workforce.

IMDB MOVIE ANALYSIS

The IMDB Movie Analysis project aims to investigate the factors that influence the success of movies on IMDB. Success is defined by high IMDB ratings, which are crucial for movie producers, directors, and investors to make informed decisions. This project analyzes various aspects such as genres, durations, languages, directors, and budgets to uncover meaningful insights about what contributes to a movie's success.

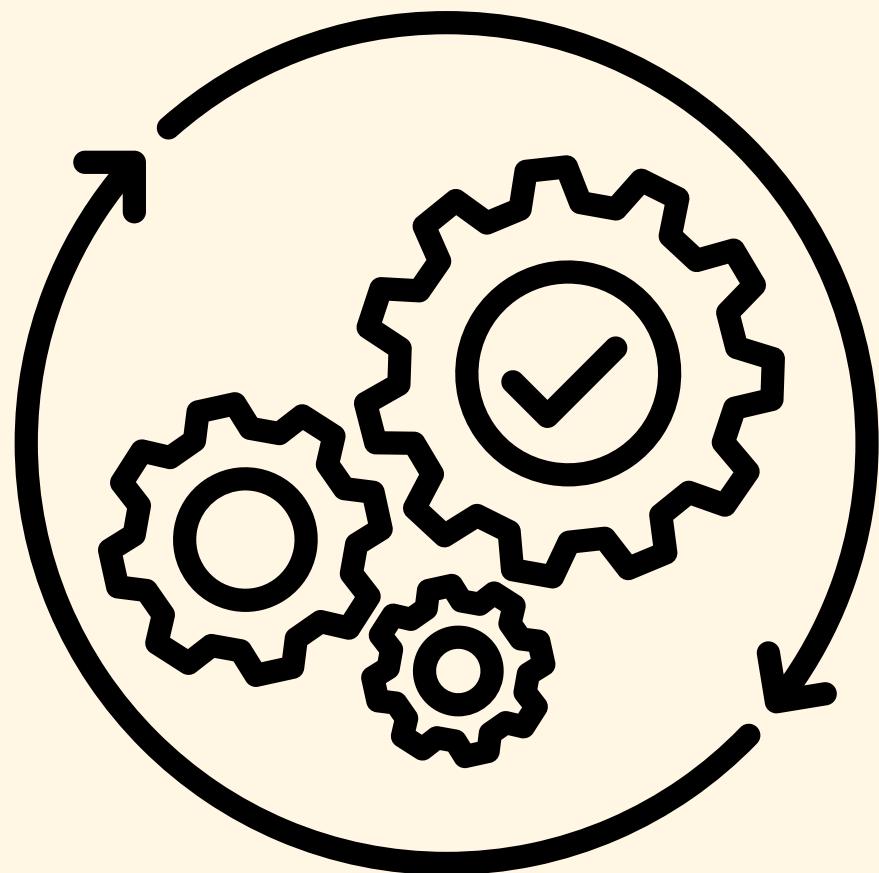


PROBLEM OF THE PROJECT



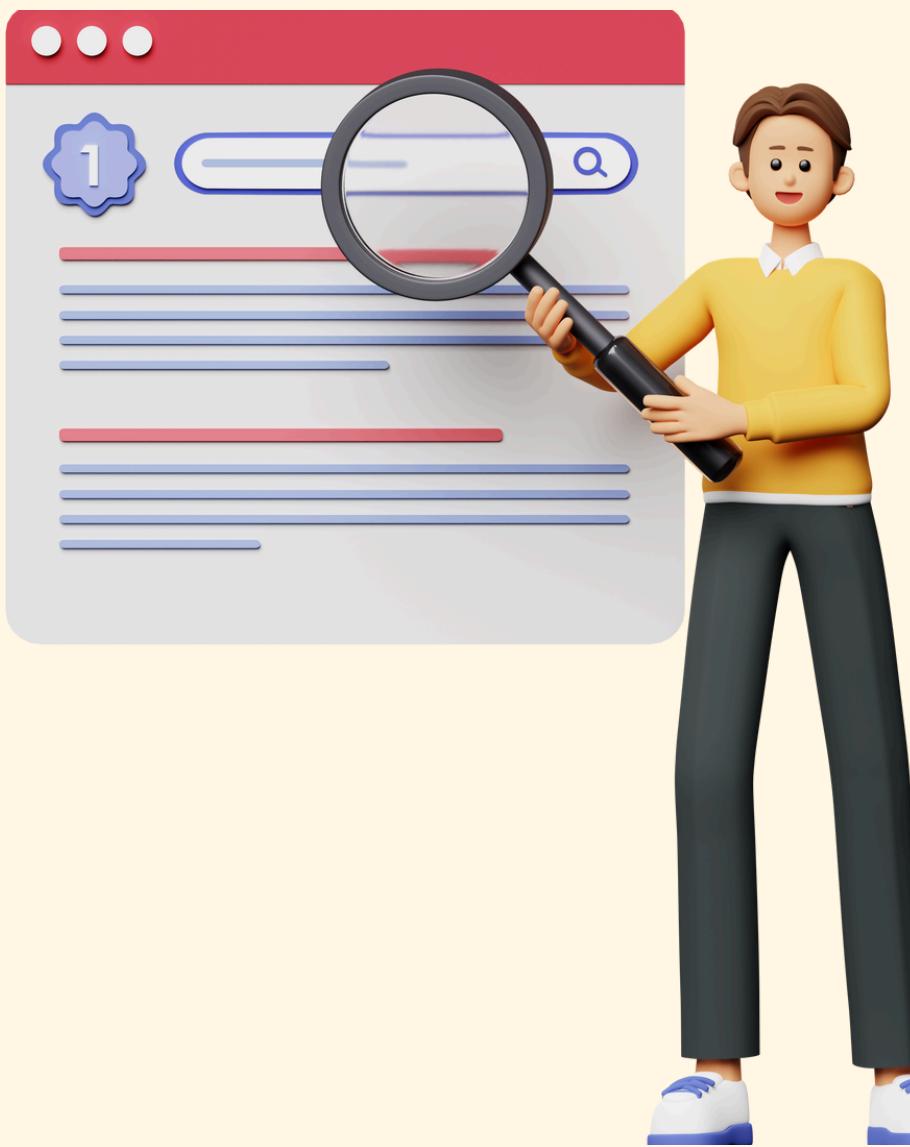
- The Problems in finding :-
 - The analysis revealed that genres play a significant role in a movie's success.
 - Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.
 - Determine the most common languages used in movies
 - Identify the top directors
 - Analyze the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

DESIGN



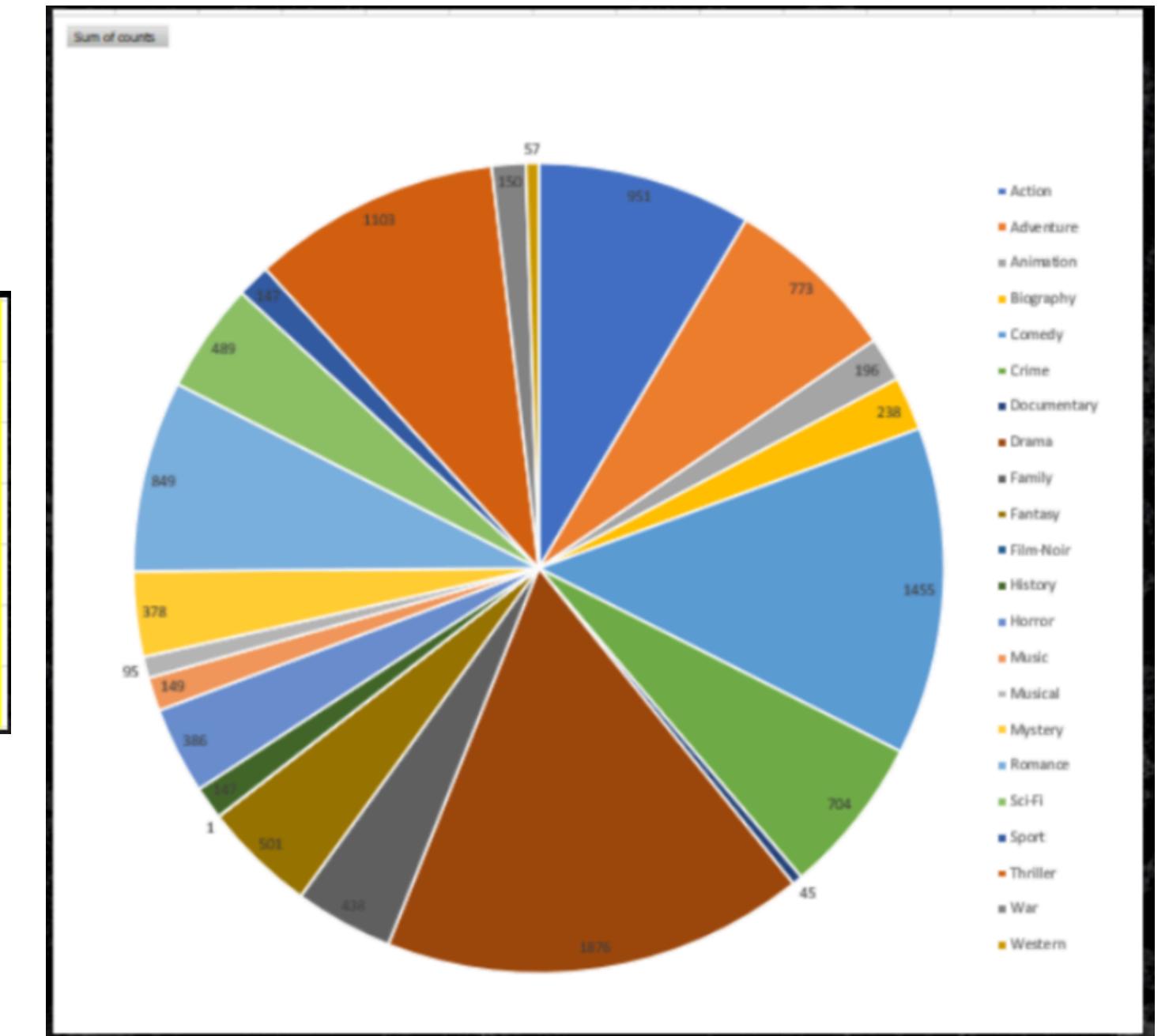
- This project was done by using Microsoft excel 2022
- Used for data cleaning, analysis, and visualization. Functions such as COUNTIF, AVERAGE, MEDIAN, STDEV.S, and CORREL were utilized to perform statistical analysis and derive insights.

FINDINGS



1.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.

avg	505.8181818
median	382
mode	147
max	1876
min	1
var	243511.2987
std	493.46864

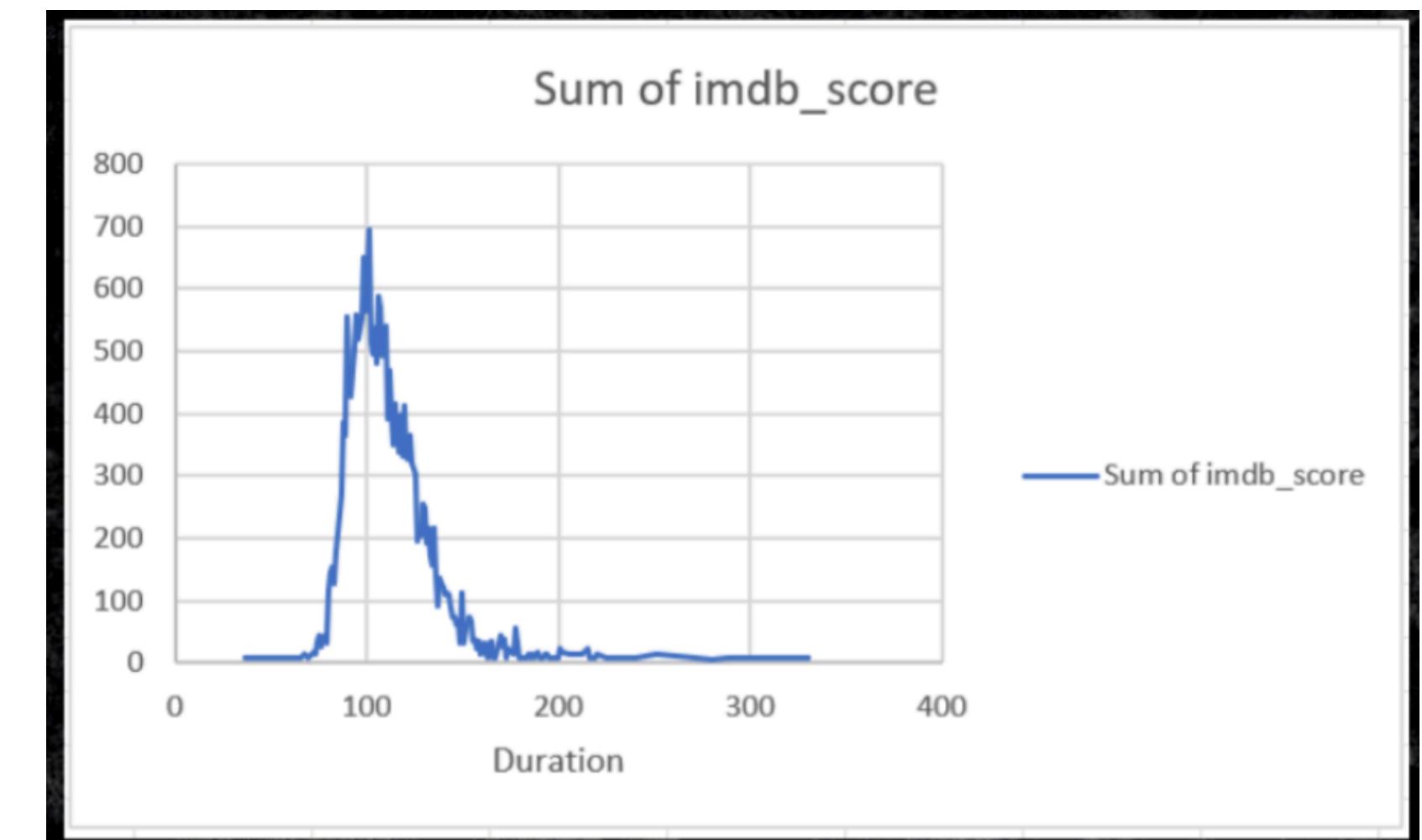


FINDINGS



2. Analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

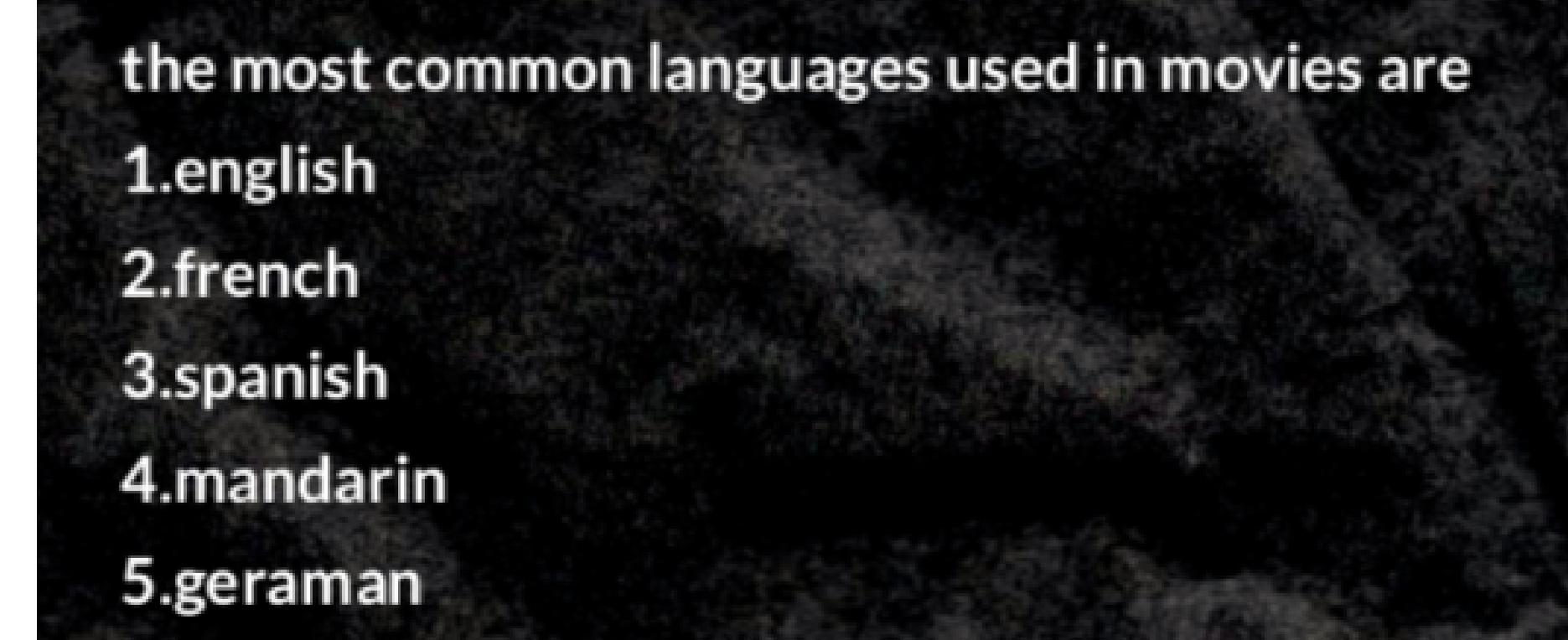
duration	Sum of imdb_score
330	8
325	6.8
300	6.6
293	8.4
289	8.5
280	6.3
271	7.7
251	15.4
240	8.2
236	8
227	8.4



FINDINGS



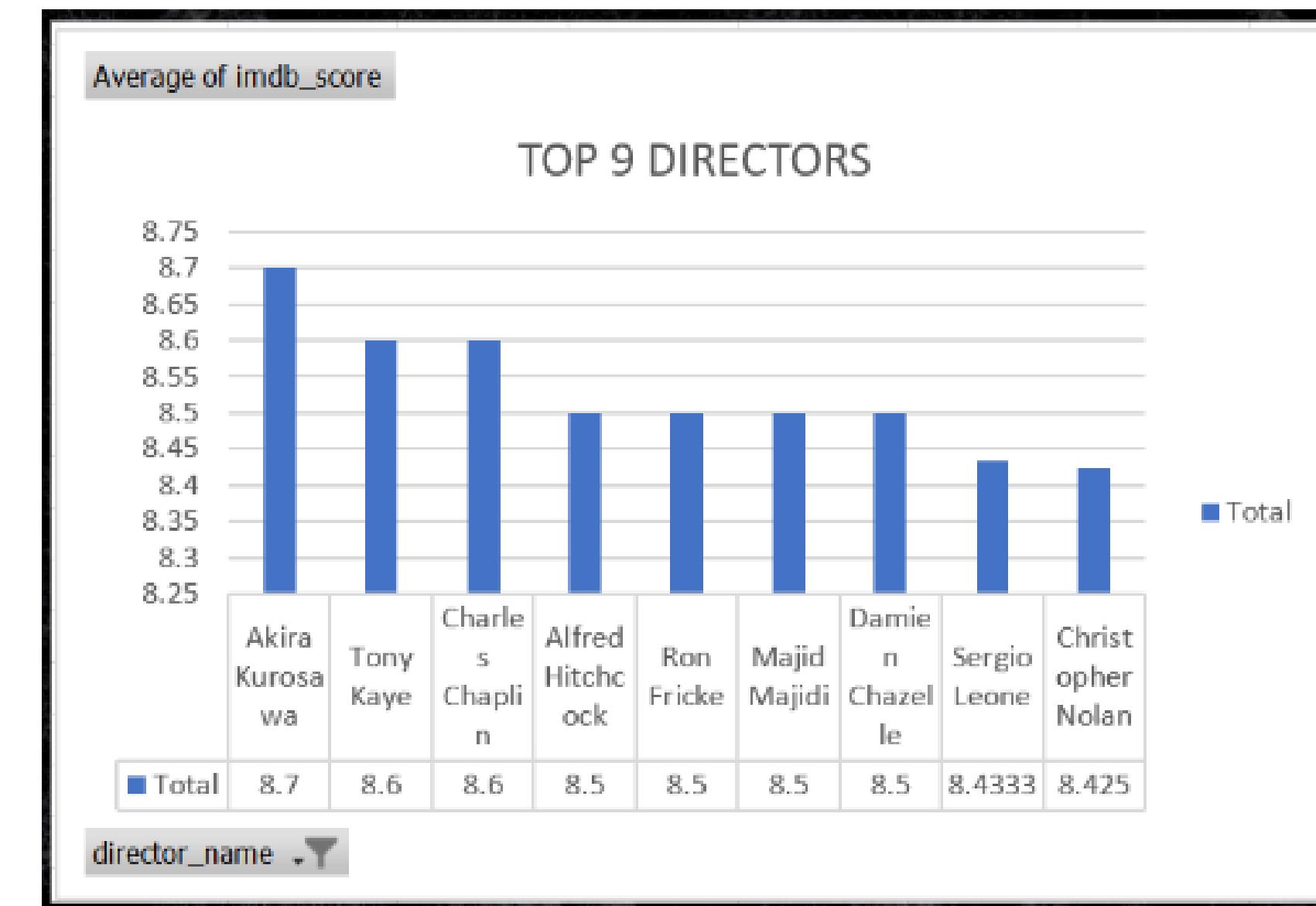
3.Determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.



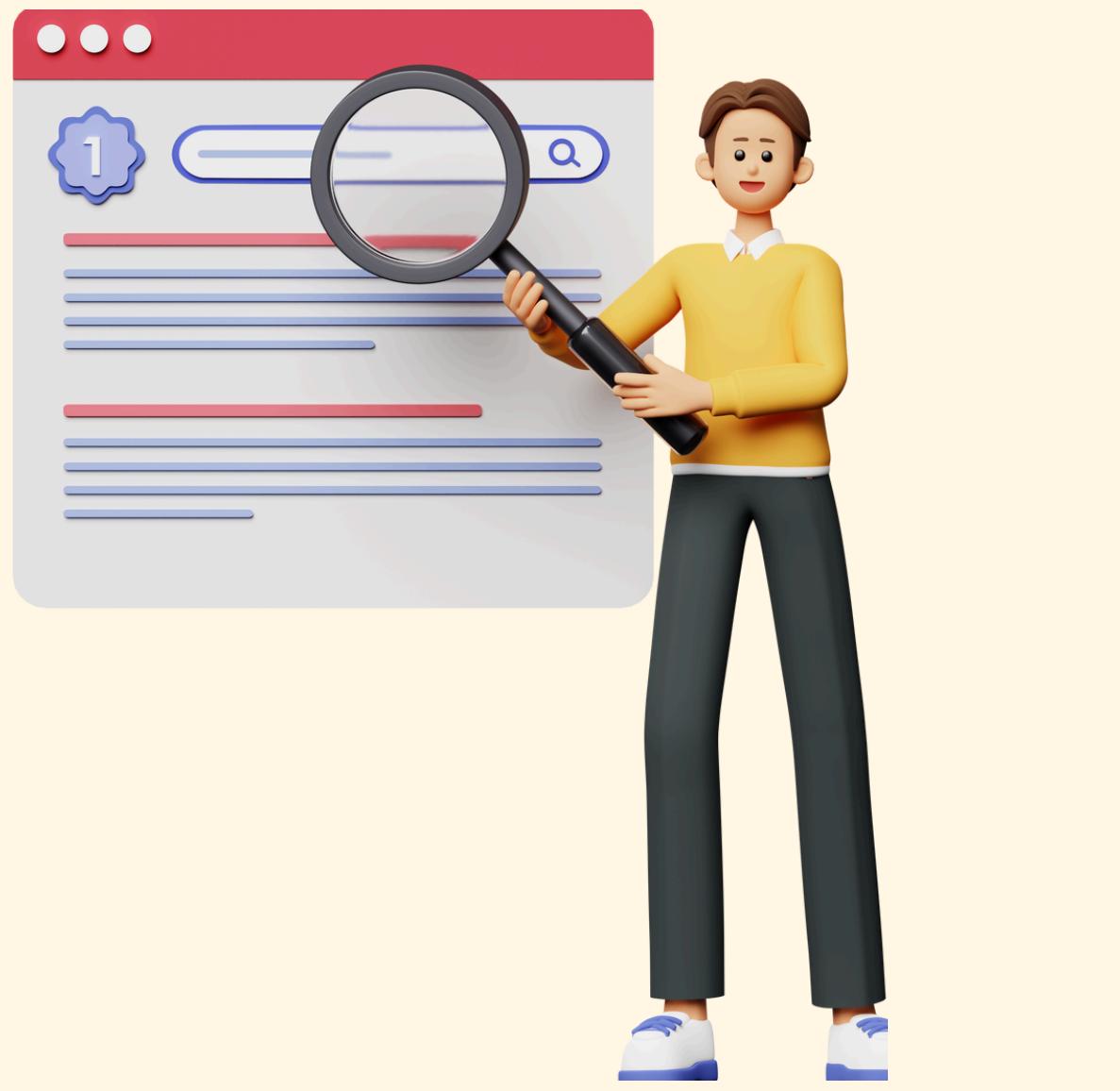
FINDINGS



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



FINDINGS



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

budget	Sum of gross	profit
\$ 218.00	\$ 5,92,014.00	\$ 5,91,796.00
\$ 1,100.00	\$ 85,222.00	\$ 5,90,914.00
\$ 4,500.00	\$ 1,36,007.00	\$ 5,87,514.00
\$ 7,000.00	\$ 24,65,680.00	\$ 5,85,014.00
\$ 10,000.00	\$ 8,54,263.00	\$ 5,82,014.00
\$ 14,000.00	\$ 12,667.00	\$ 5,78,014.00
\$ 15,000.00	\$ 10,79,17,283.00	\$ 5,77,014.00
\$ 23,000.00	\$ 12,27,508.00	\$ 5,69,014.00
\$ 25,000.00	\$ 1,31,03,222.00	\$ 5,67,014.00
\$ 40,000.00	\$ 3,87,769.00	\$ 5,52,014.00
\$ 42,000.00	\$ 2,41,816.00	\$ 5,50,014.00
\$ 46,000.00	\$ 11,27,331.00	\$ 5,46,014.00
\$ 60,000.00	\$ 14,43,24,408.00	\$ 5,32,014.00
\$ 65,000.00	\$ 1,19,19,172.00	\$ 5,27,014.00
\$ 70,000.00	\$ 30,084.00	\$ 5,22,014.00
\$ 83,532.00	\$ 6,17,18,000.00	\$ 5,08,482.00
\$ 1,00,000.00	\$ 3,94,77,173.00	\$ 4,92,014.00
\$ 1,20,000.00	\$ 4,69,947.00	\$ 4,72,014.00

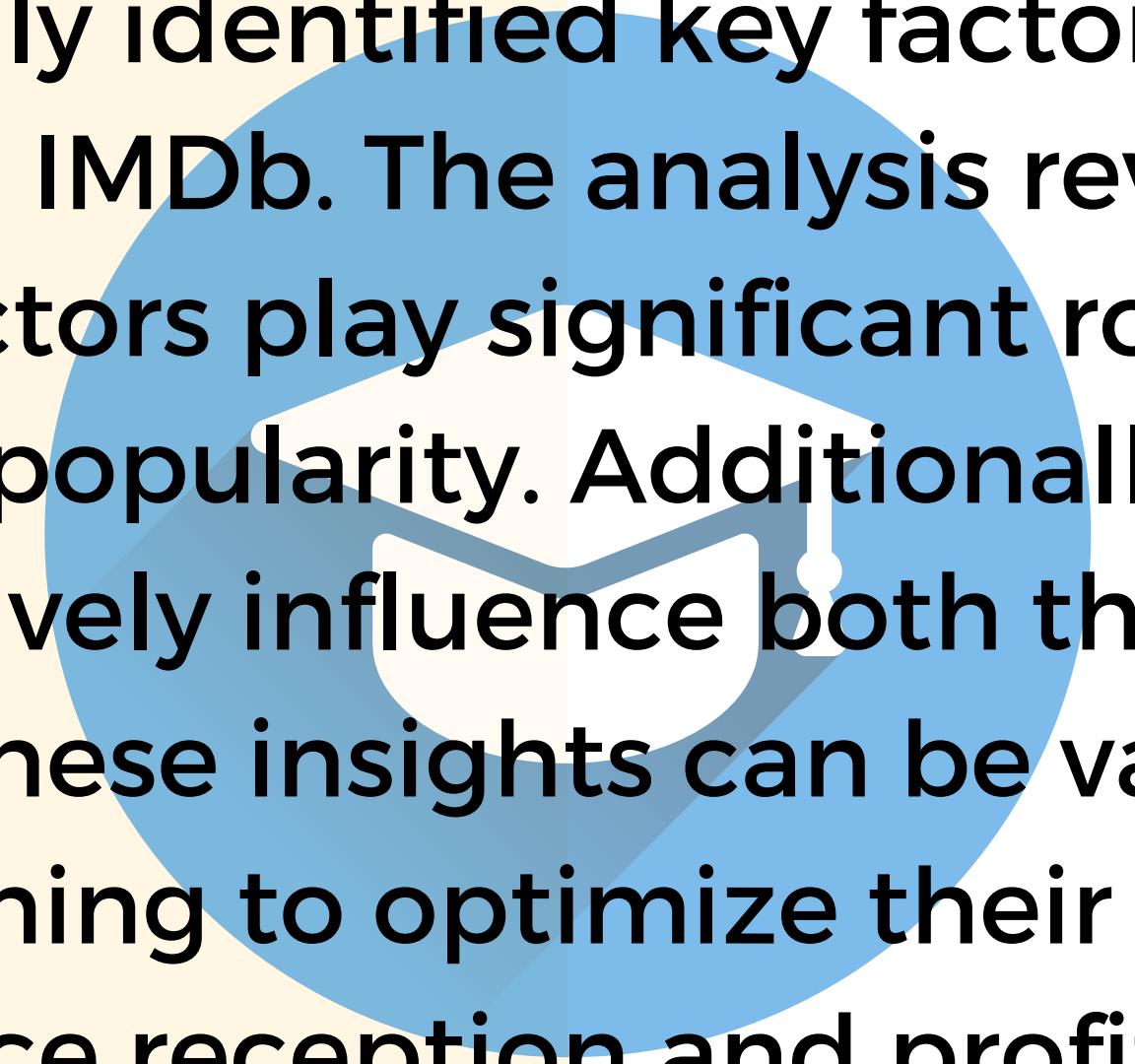
correlation	-0.030598094
high profit margin	\$ 7,44,21,71,803.00
min	\$ -12,21,49,07,986.00

ANALYSIS



The project provided a comprehensive understanding of the factors influencing movie success on IMDB. The analysis highlighted the importance of genres, languages, and directors, as well as the positive impact of higher budgets on movie ratings and financial success.

CONCLUSION



the project successfully identified key factors that contribute to a movie's success on IMDb. The analysis revealed that genres, languages, and directors play significant roles in determining a movie's rating and popularity. Additionally, it was found that higher budgets positively influence both the movie's ratings and its financial success. These insights can be valuable for filmmakers and producers aiming to optimize their projects for better audience reception and profitability.

BANK LOAN CASE STUDY

This project involves a detailed analysis of a loan application dataset to identify patterns and factors that influence loan defaults. As a data analyst at a finance company specializing in lending to urban customers, the primary objective is to use Exploratory Data Analysis (EDA) to ensure that capable applicants are not rejected, while mitigating financial losses due to defaults by high-risk applicants. This project will explore customer and loan attributes to identify indicators of potential payment difficulties, enabling more informed decision-making in loan approvals.

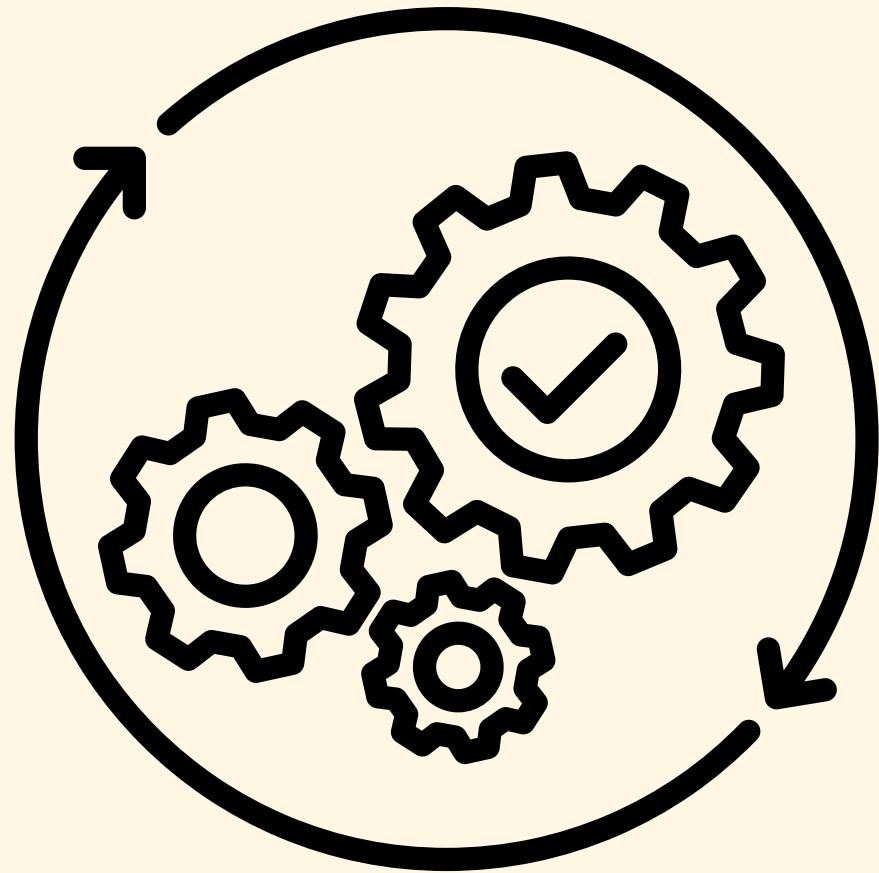


PROBLEM OF THE PROJECT



- The Problems in finding :-
 - Identify Missing Data and Deal with it Appropriately
 - Identify Outliers in the Dataset
 - Analyze Data Imbalance set
 - Perform Univariate, Segmented Univariate, and Bivariate Analysis
 - Identify Top Correlations for Different Scenarios

DESIGN



- This project was done by using MS EXCEL 2022

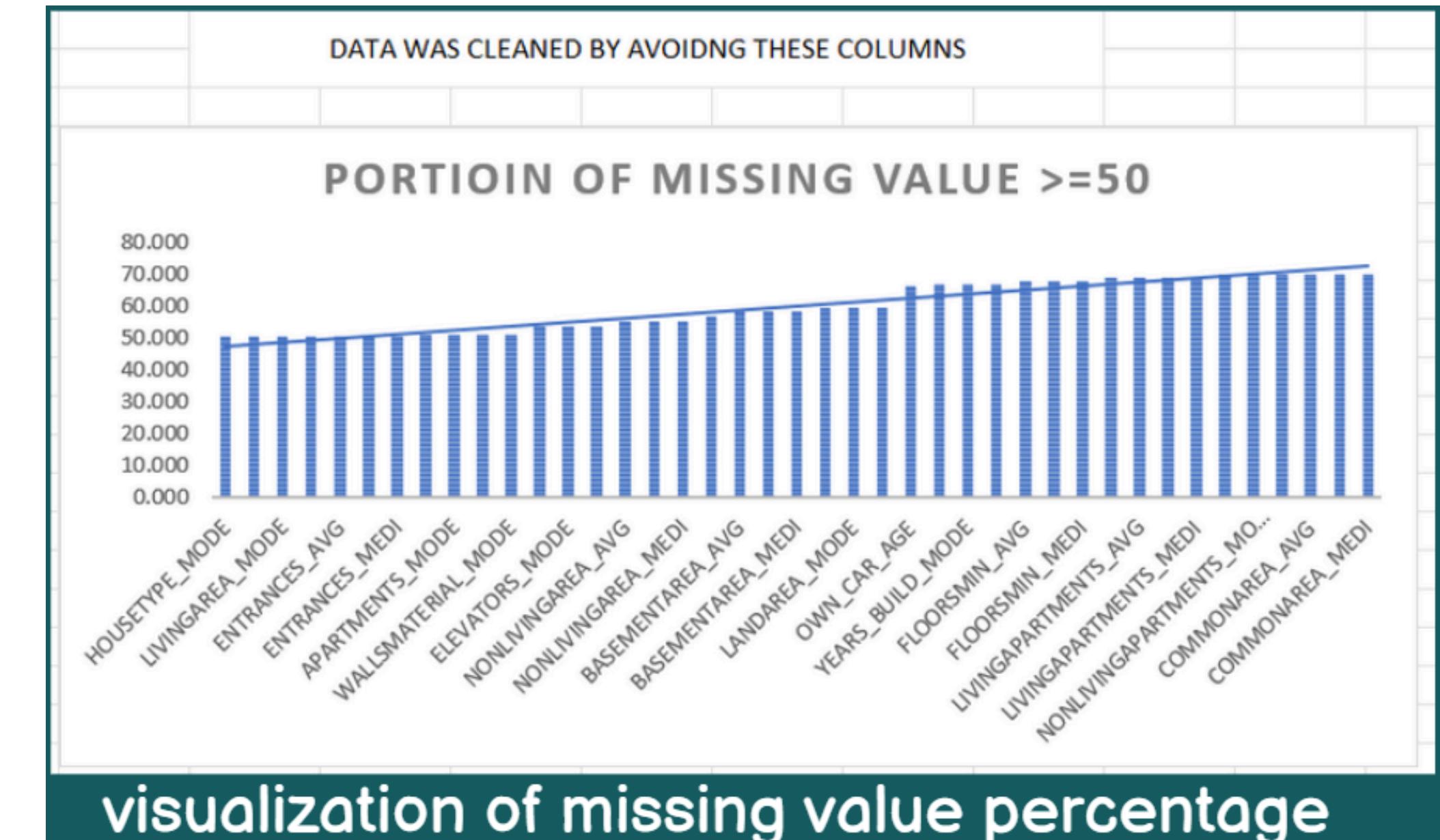


- Excel was used for data cleaning, analysis, and visualization. Its built-in functions and features such as pivot tables, charts, and conditional formatting

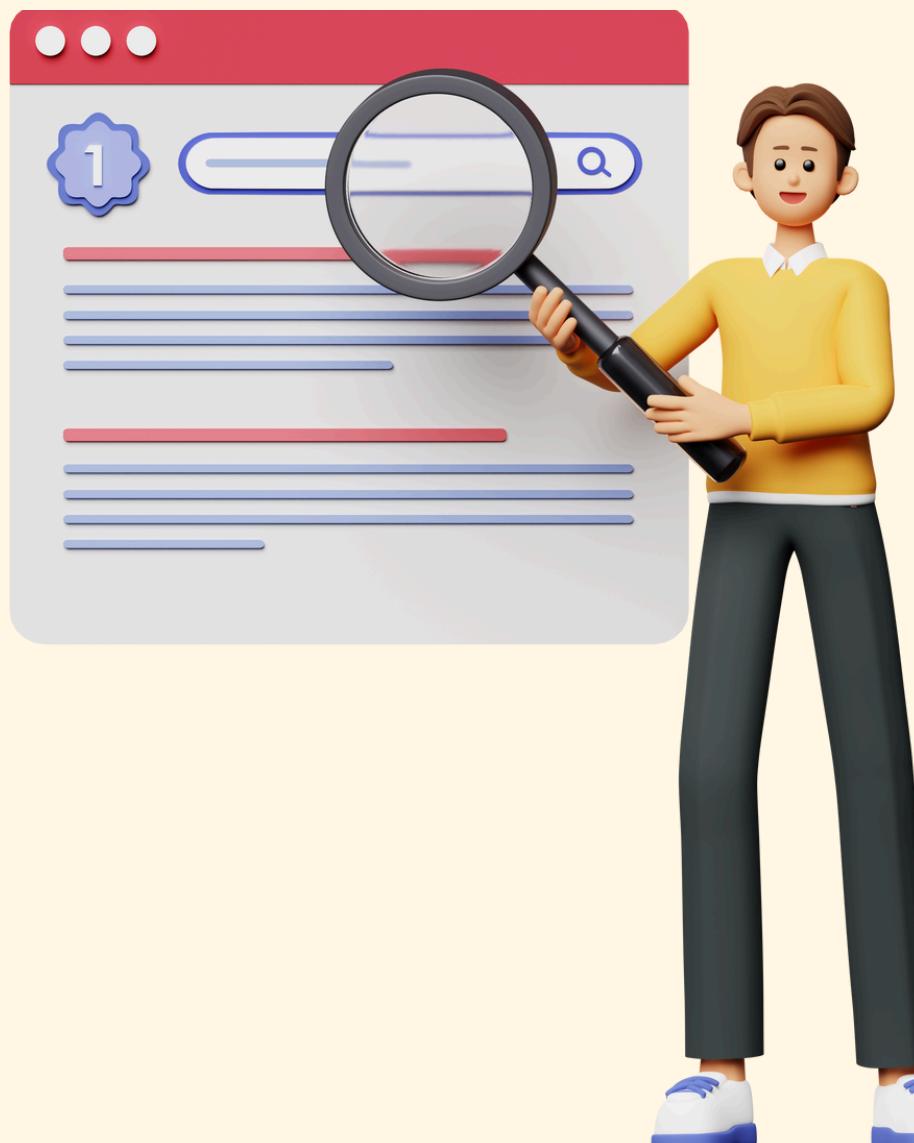
FINDINGS



1. Identify Missing Data and Deal with it Appropriately

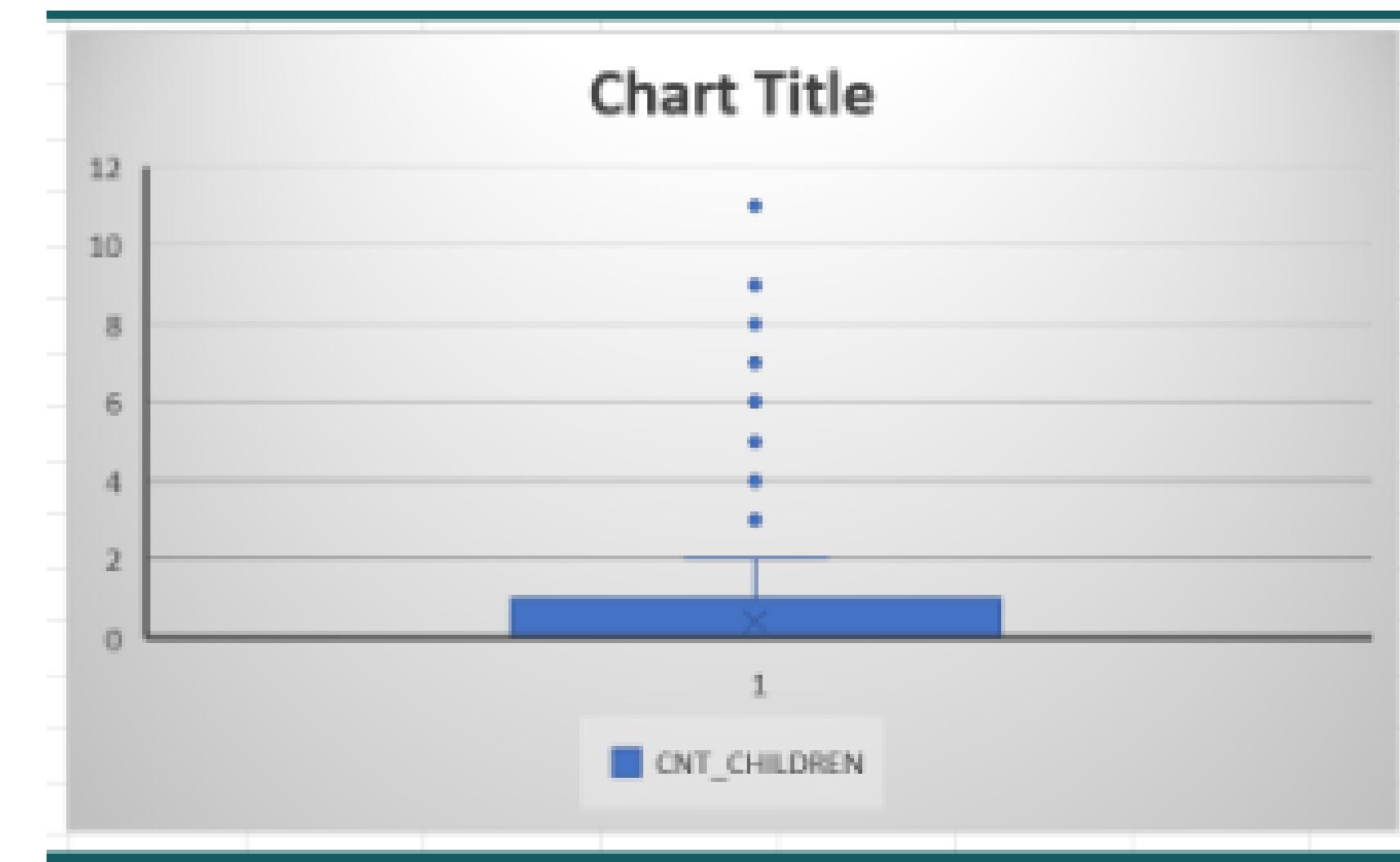


FINDINGS



2. Identify Outliers in the Dataset

	11	₹ 11,70,00,000.00	4050000	258026	4050000
MAXIMUM VALUE	11	₹ 11,70,00,000.00	4050000	258026	4050000
MINIMUM VALUE	0	₹ 25,650.00	45000	2052	45000
QURTILE 1 (FINDING 25 TH PORTION OF DATA)	0	₹ 1,12,500.00	270000	16456.5	238500
QURTILE 3 (FINDING 75 TH PORTION OF DATA)	1	₹ 2,02,500.00	808650	34596	679500
Interquartile Range (IQR)	1	₹ 90,000.00	538650	18139.5	441000
UPPER LIMIT	2.5	337500	1616625	61805.3	1341000
LOWER LIMIT	-1.5	-22500	-537975	-10752.8	-423000
OUTLIER BELOW LOWER LIMIT	0	0	0	0	0
OUTLIER ABOVE UPPER LIMIT	723	2295	1063	1188	2387
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODIE

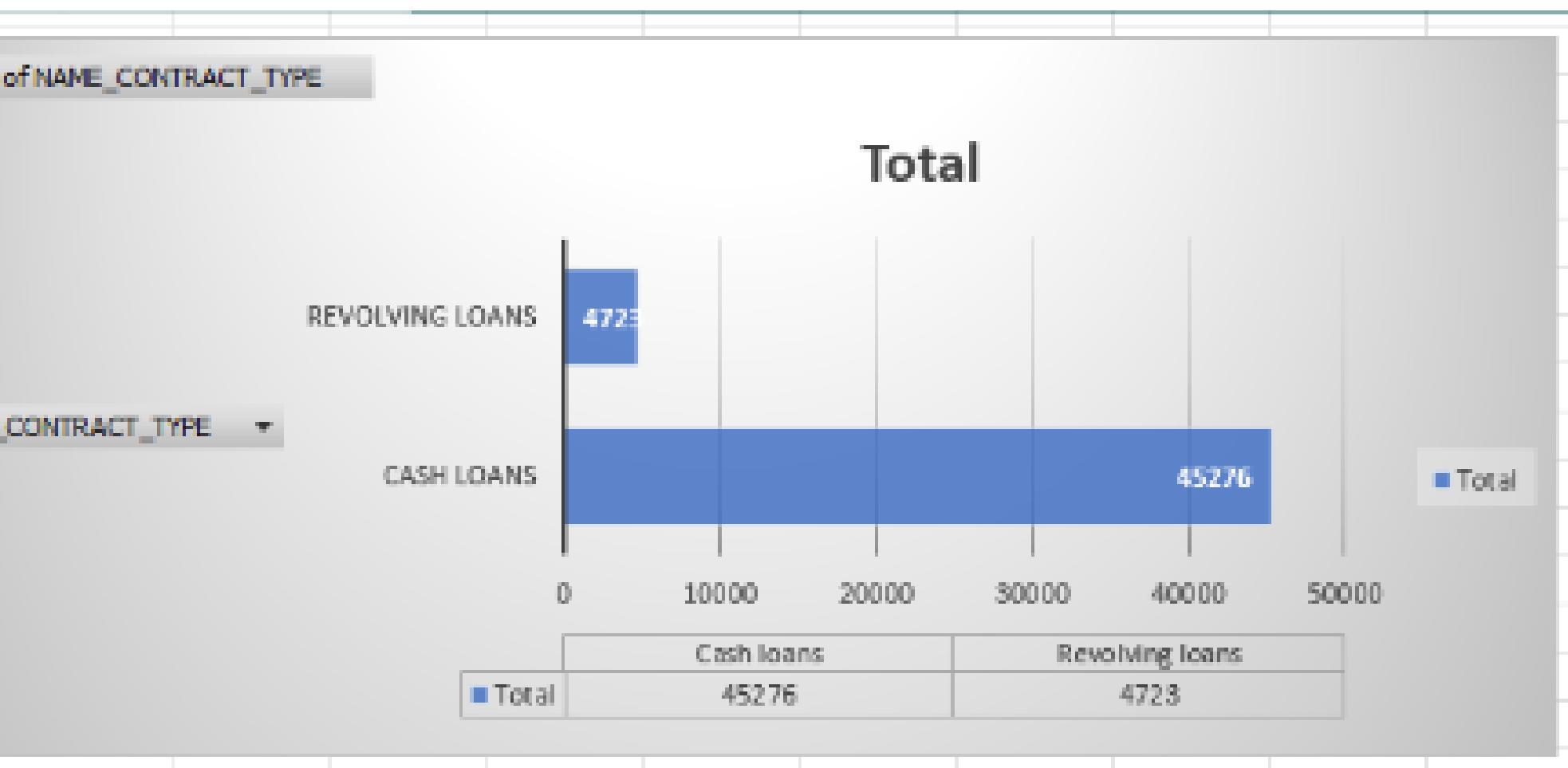


FINDINGS



3. Analyze Data Imbalance

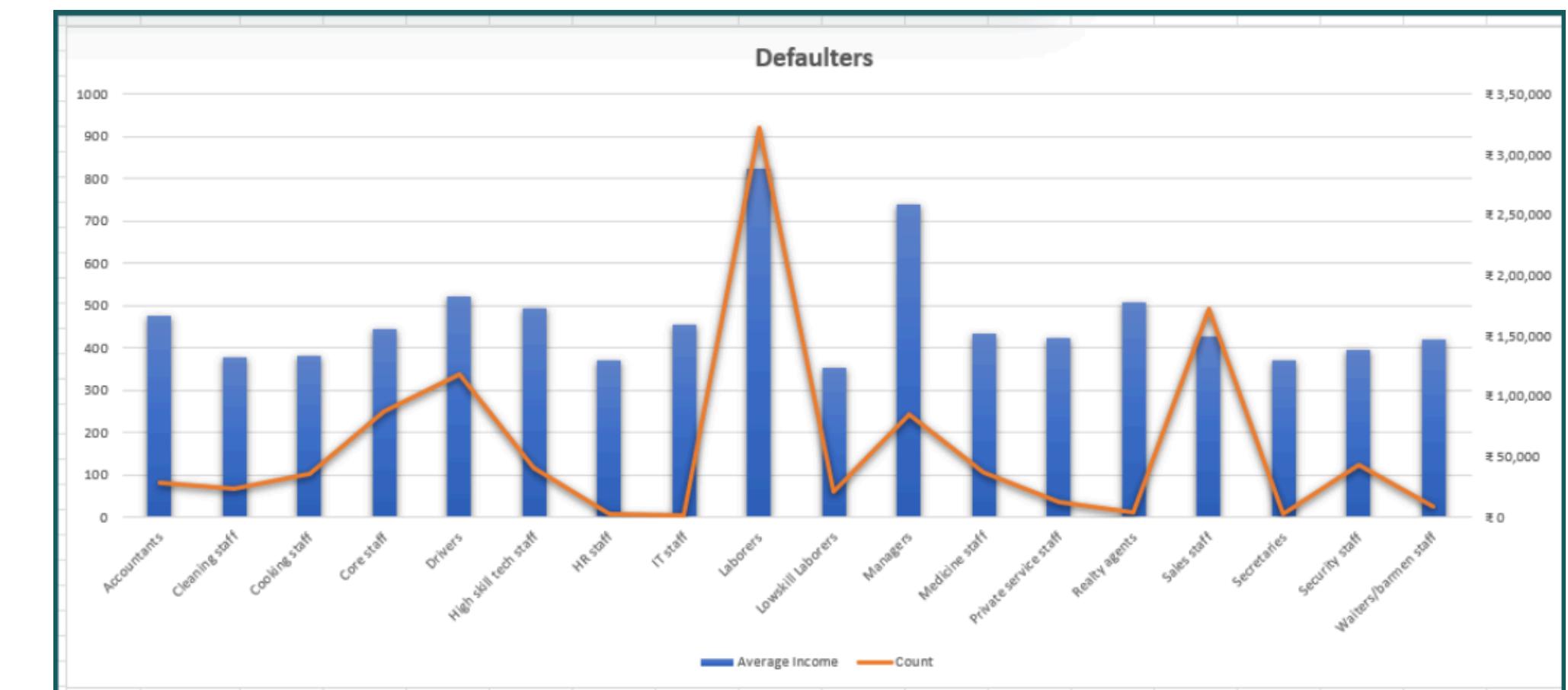
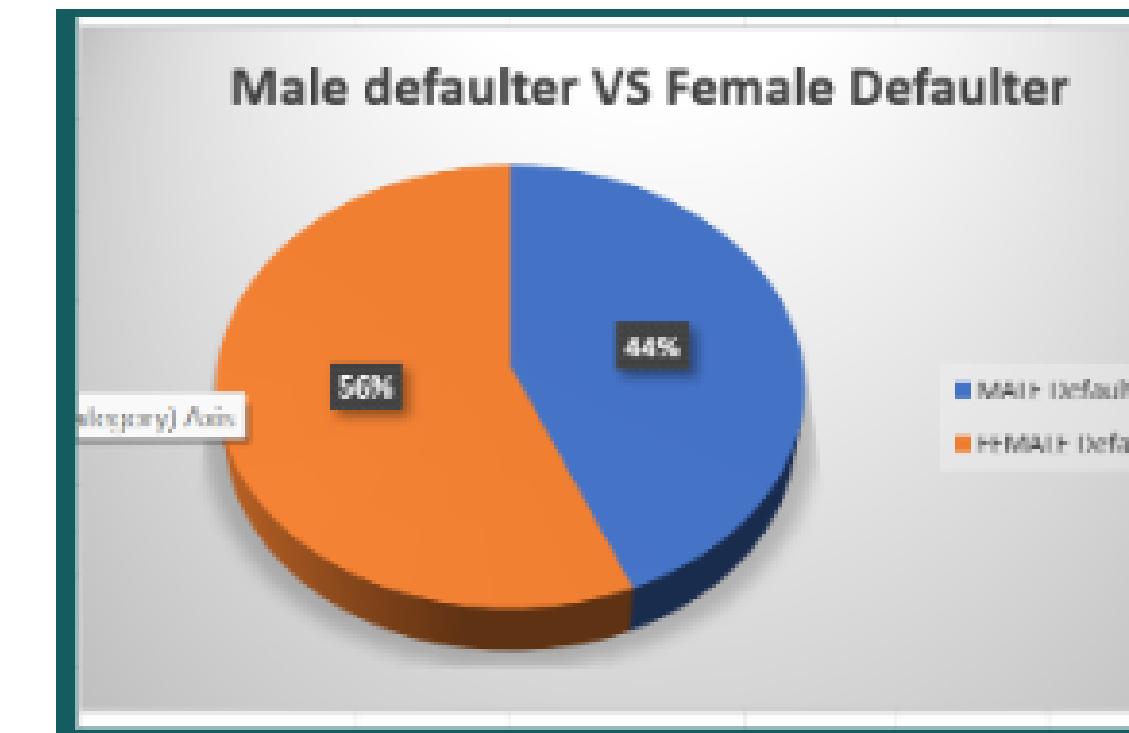
0 - difficulties in payment of loan		
1 - paid loan on time		
Sum of customers		
Type of customer	Count of TARGET	Data Imbalance Ratio
0	45973	11.41902633
1	4026	
Grand Total	49999	



FINDINGS



4. Perform Univariate, Segmented Univariate, and Bivariate Analysis



FINDINGS



5. Identify Top Correlations for Different Scenarios

DEFAULTERS

COLUMNS	CNT_CHIL	AMT_INCO	AMT_CRE	AMT_ANNU	AMT_GOC	REGION_P	Age	Days_Emp	Days_Regi	Days_ID_P	CNT_FAM_MEMBERS
CNT_CHILDREN	1	0.01011	0.007602	0.029173	-0.00112	-0.02036	-0.24967	-0.18977	-0.15211	0.042361	0.892521875
AMT_INCOME_TOTAL	0.01011	1	0.015271	0.018005	0.013266	-0.00618	-0.00903	-0.01176	0.009561	0.009122	0.013121678
AMT_CREDIT	0.007602	0.015271	1	0.749665	0.982432	0.067776	0.142506	0.018782	0.042844	0.043772	0.06124869
AMT_ANNUITY	0.029173	0.018005	0.749665	1	0.749705	0.073124	0.008752	-0.07811	-0.02158	0.021321	0.075838463
AMT_GOODS_PRICE	-0.00112	0.013266	0.982432	0.749705	1	0.076596	0.140996	0.023159	0.043371	0.049785	0.055103609
REGION_POPULATION_RELATIVE	-0.02036	-0.00618	0.067776	0.073124	0.076596	1	0.016469	0.00771	0.04613	0.005119	-0.017257146
Age	-0.24967	-0.00903	0.142506	0.008752	0.140996	0.016469	1	0.588243	0.288438	0.247897	-0.199141397
Days_Employed(YEARS)	-0.18977	-0.01176	0.018782	-0.07811	0.023159	0.00771	0.588243	1	0.192436	0.232662	-0.183362962
Days_Registration(years)	-0.15211	0.009561	0.042844	-0.02158	0.043371	0.04613	0.288438	0.192436	1	0.090291	-0.151786548
Days_ID_Publish(years)	0.042361	0.009122	0.043772	0.021321	0.049785	0.005119	0.247897	0.232662	0.090291	1	0.044037815
CNT_FAM_MEMBERS	0.892522	0.013122	0.061249	0.075838	0.055104	-0.01726	-0.19914	-0.18336	-0.15179	0.044038	1

NON-DEFAULTERS

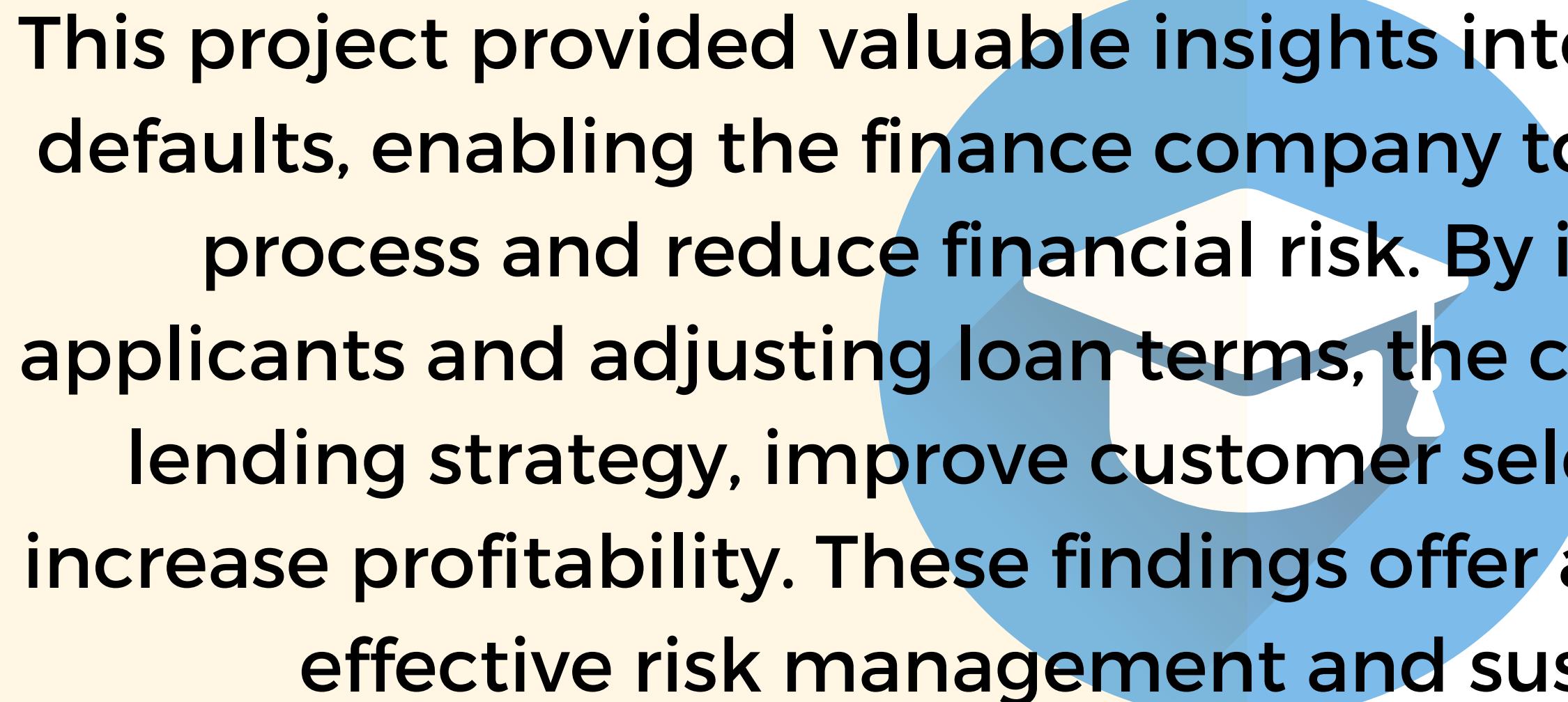
COLUMNS	CNT_CHIL	AMT_INCO	AMT_CRE	AMT_ANNU	AMT_GOC	REGION_P	Age	Days_Emp	Days_Regi	Days_ID_P	CNT_FAM_MEMBERS
CNT_CHILDREN	1	0.03632	0.00571	0.02638	0.00155	-0.0249	-0.3359	-0.2455	-0.1831	0.03254	0.87923936
AMT_INCOME_TOTAL	0.03632	1	0.37797	0.45114	0.38468	0.18194	-0.0738	-0.1617	-0.0689	-0.0323	0.041613404
AMT_CREDIT	0.00571	0.37797	1	0.77077	0.98724	0.09554	0.05108	-0.0747	-0.0081	0.00829	0.064877635
AMT_ANNUITY	0.02638	0.45114	0.77077	1	0.77614	0.11728	-0.0099	-0.1113	-0.0346	-0.0094	0.077891705
AMT_GOODS_PRICE	0.00155	0.38468	0.98724	0.77614	1	0.09897	0.0487	-0.0725	-0.0113	0.0093	0.062957956
REGION_POPULATION_RELATIVE	-0.0249	0.18194	0.09554	0.11728	0.09897	1	0.03044	-0.0068	0.0585	0.00224	-0.023006667
Age	-0.3359	-0.0738	0.05108	-0.0099	0.0487	0.03044	1	0.62347	0.33503	0.27007	-0.284384945
Days_Employed(YEARS)	-0.2455	-0.1617	-0.0747	-0.1113	-0.0725	-0.0068	0.62347	1	0.20885	0.27452	-0.234767657
Days_Registration(years)	-0.1831	-0.0689	-0.0081	-0.0346	-0.0113	0.0585	0.33503	0.20885	1	0.10355	-0.171485094
Days_ID_Publish(years)	0.03254	-0.0323	0.00829	-0.0094	0.0093	0.00224	0.27007	0.27452	0.10355	1	0.025058177
CNT_FAM_MEMBERS	0.87924	0.04161	0.06488	0.07789	0.06296	-0.023	-0.2844	-0.2348	-0.1715	0.02506	1

ANALYSIS



Through this project, a comprehensive understanding of the factors contributing to loan defaults was achieved. The analysis provided actionable insights that can be used to refine the loan approval process, reduce financial risk, and optimize customer selection criteria. By identifying high-risk applicants and adjusting loan terms accordingly, the finance company can improve its lending strategy and enhance overall profitability.

CONCLUSION



This project provided valuable insights into the factors driving loan defaults, enabling the finance company to refine its loan approval process and reduce financial risk. By identifying high-risk applicants and adjusting loan terms, the company can enhance its lending strategy, improve customer selection, and ultimately increase profitability. These findings offer a strategic path for more effective risk management and sustainable growth.

THE IMPACT OF CAR FEATURES ON PRICE AND PROFITABILITY

This Project involves a detailed analysis of Car Features and others detailed dataset to identify How the Car features that impact on Price and Profitability

Analyze car features, market categories, and pricing to balance consumer demand with profitability.

The Dataset given by the Trainity was used for this project analysis

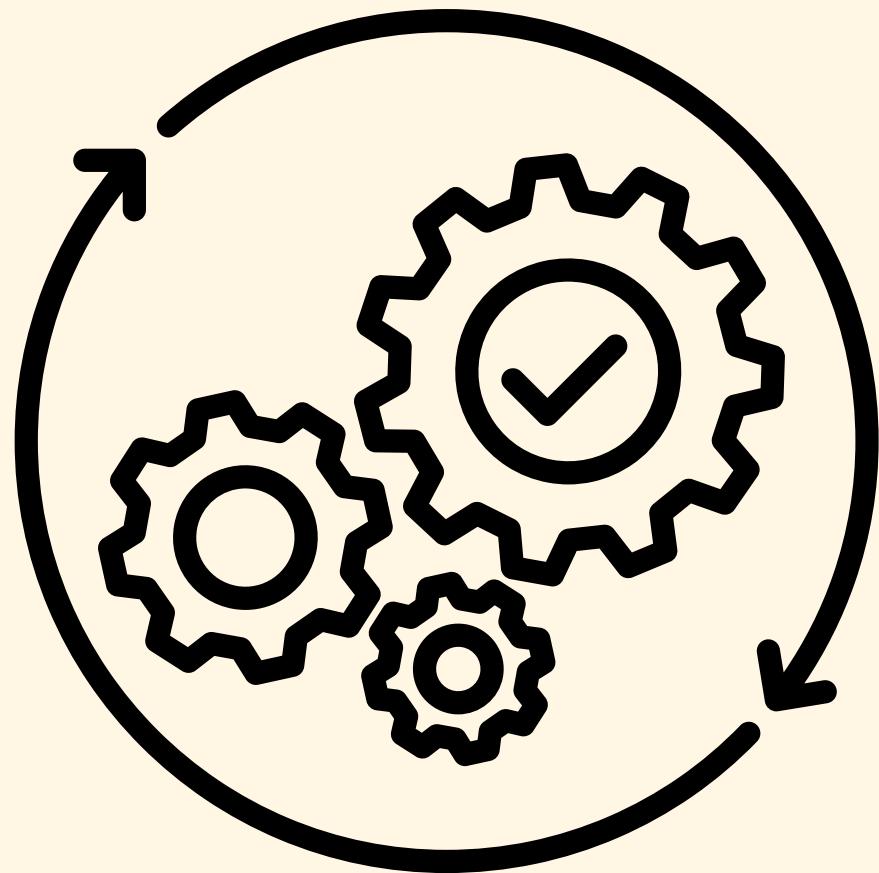


PROBLEM OF THE PROJECT



- The Problems in finding :-
 - Identify Missing Data and Deal with it Appropriately
 - Identify Outliers in the Dataset
 - Analyze Data Imbalance set
 - Perform Univariate, Segmented Univariate, and Bivariate Analysis
 - Identify Top Correlations for Different Scenarios

DESIGN

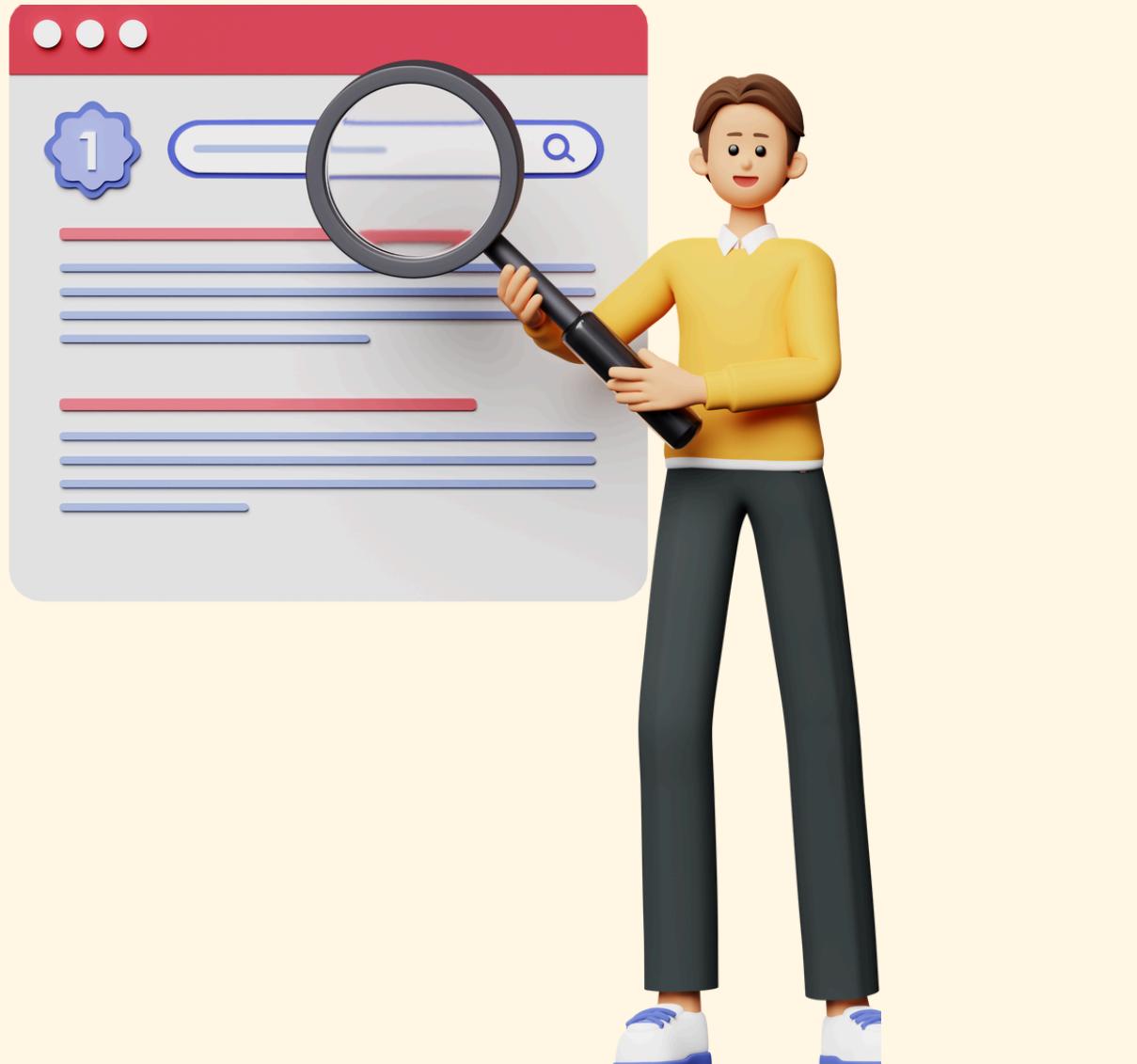


- This project was done by using MS EXCEL 2022



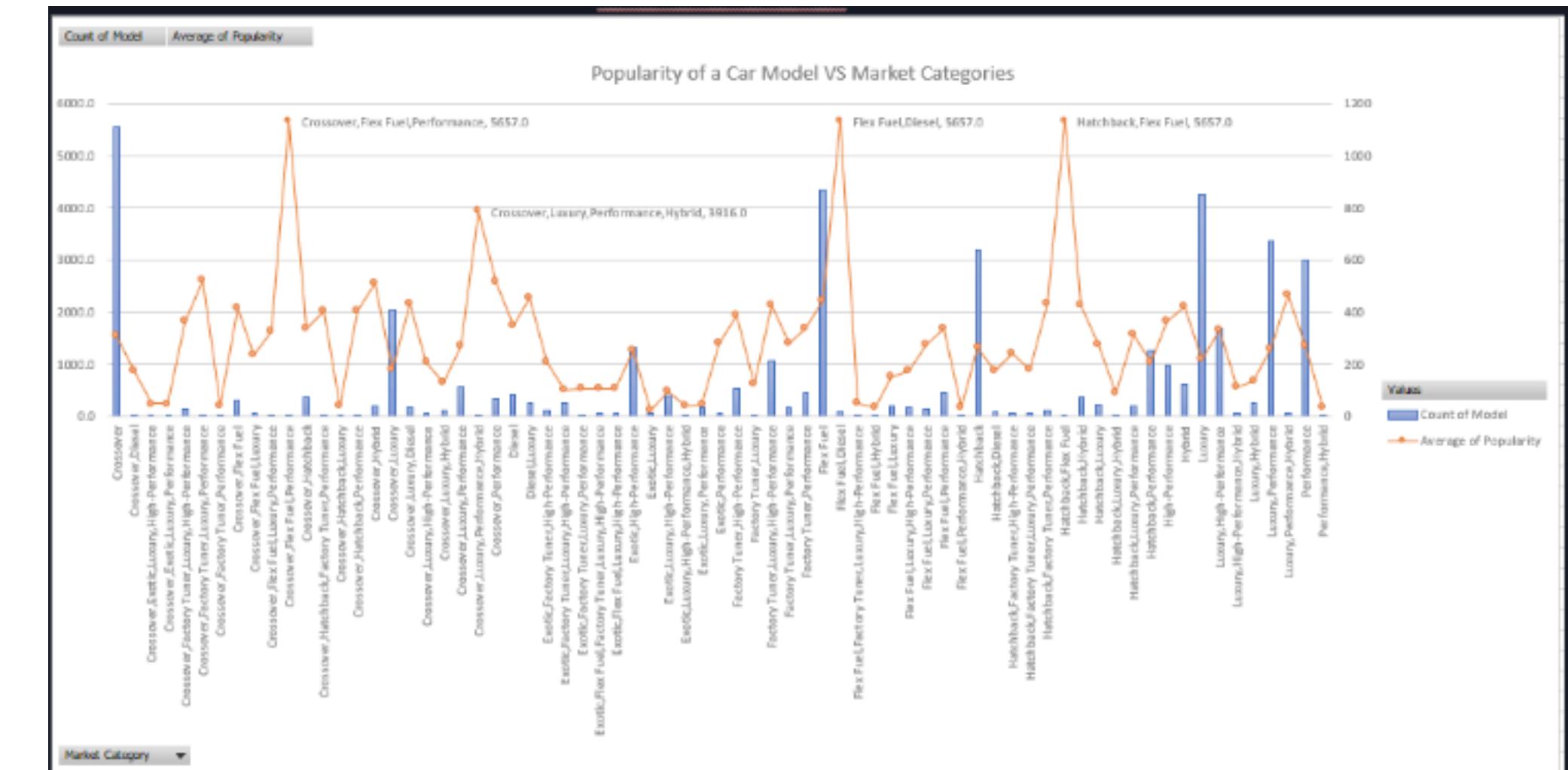
- Excel was used for data cleaning, analysis, and visualization. Its built-in functions and features such as pivot tables, charts, and conditional formatting

FINDINGS



1. HOW DOES THE POPULARITY OF A CAR MODEL VARY ACROSS DIFFERENT MARKET CATEGORIES?

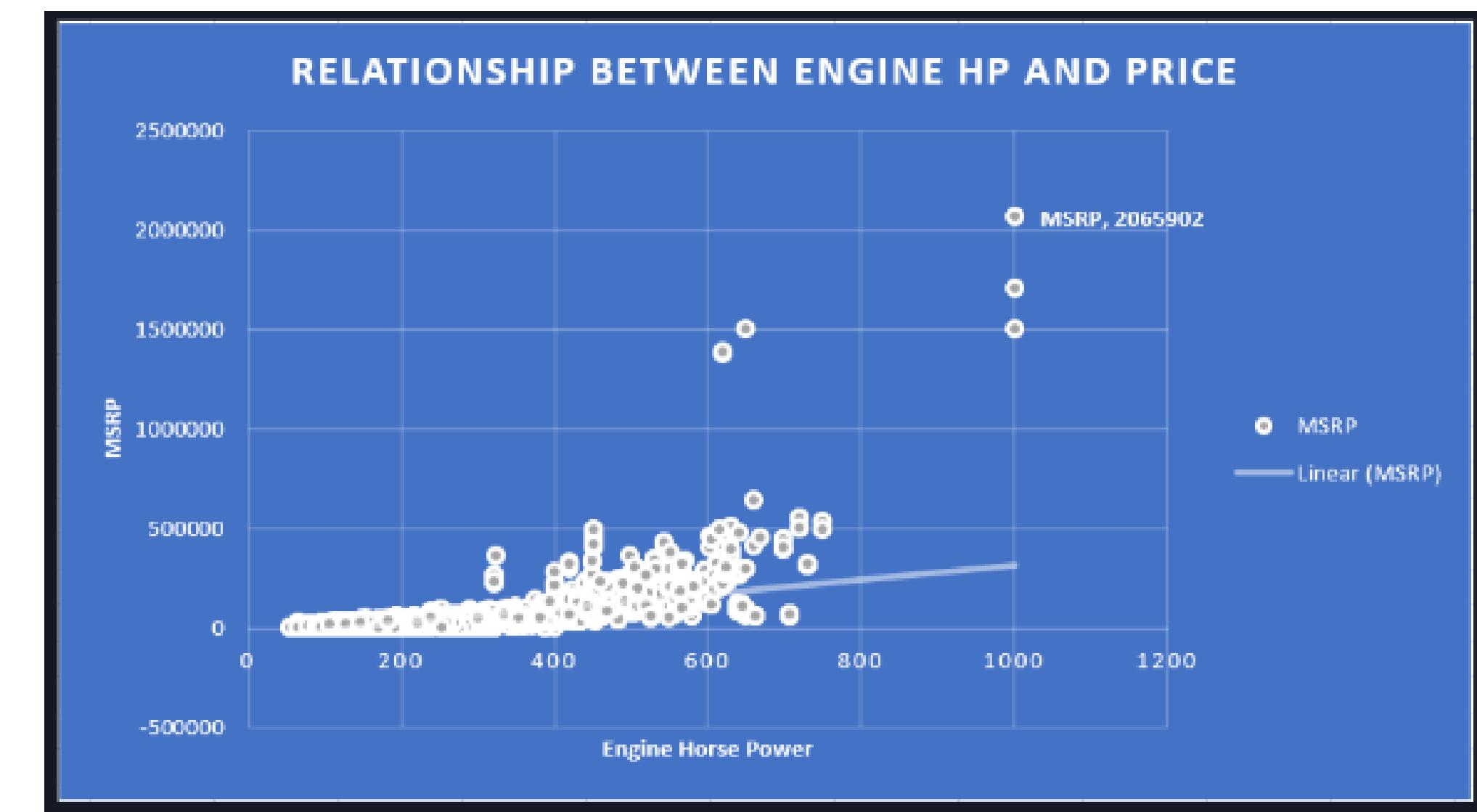
TOP 4 Popular Market Categories of Car		
Market Categories	Count of Model	Average of Popularity
Crossover,Flex Fuel,Performance	6	5657.0
Flex Fuel,Diesel	16	5657.0
Hatchback,Flex Fuel	7	5657.0
Crossover,Luxury,Performance,Hybrid	2	3916.0



FINDINGS



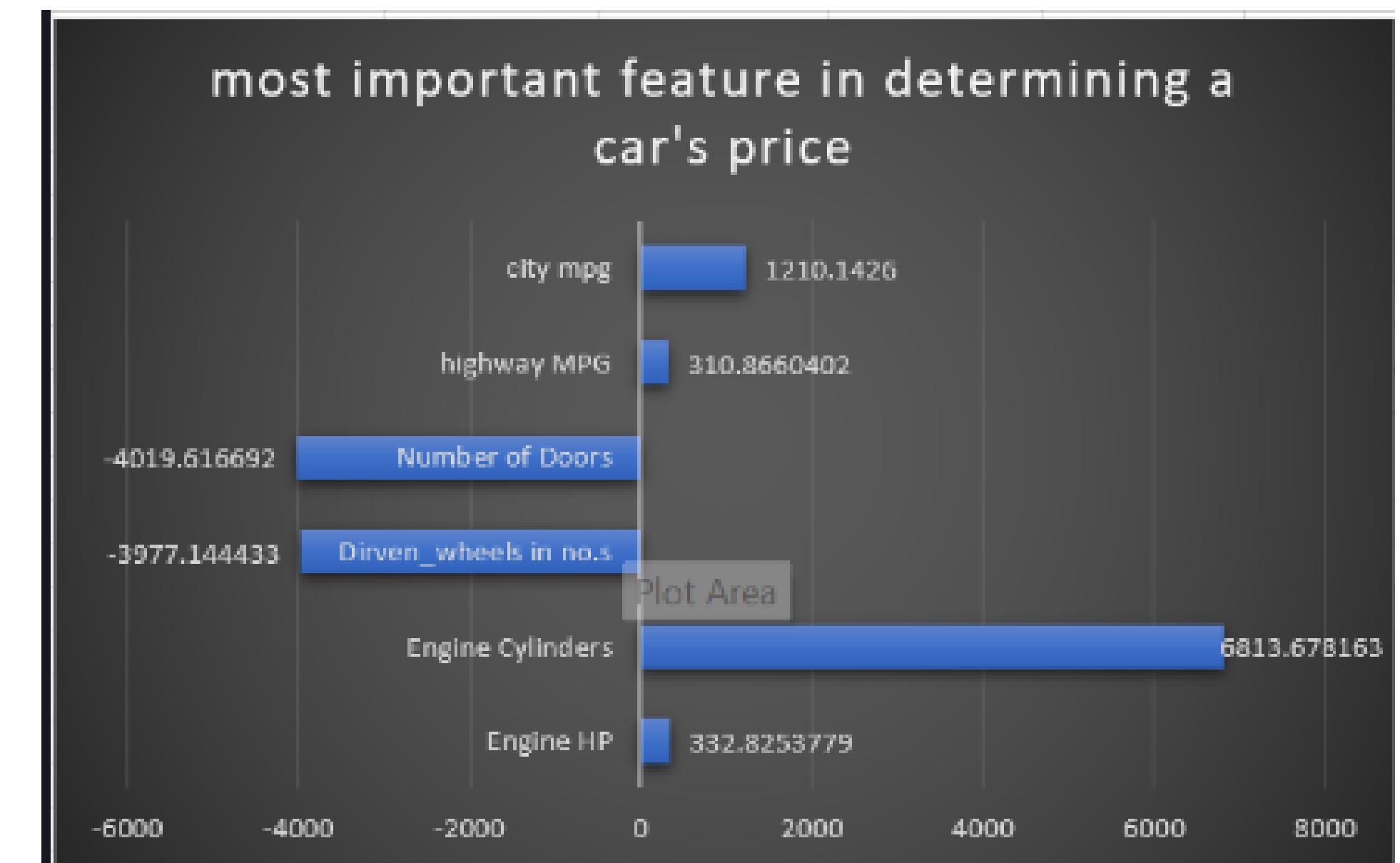
2. WHAT IS THE RELATIONSHIP BETWEEN A CAR'S ENGINE POWER AND ITS PRICE?



FINDINGS



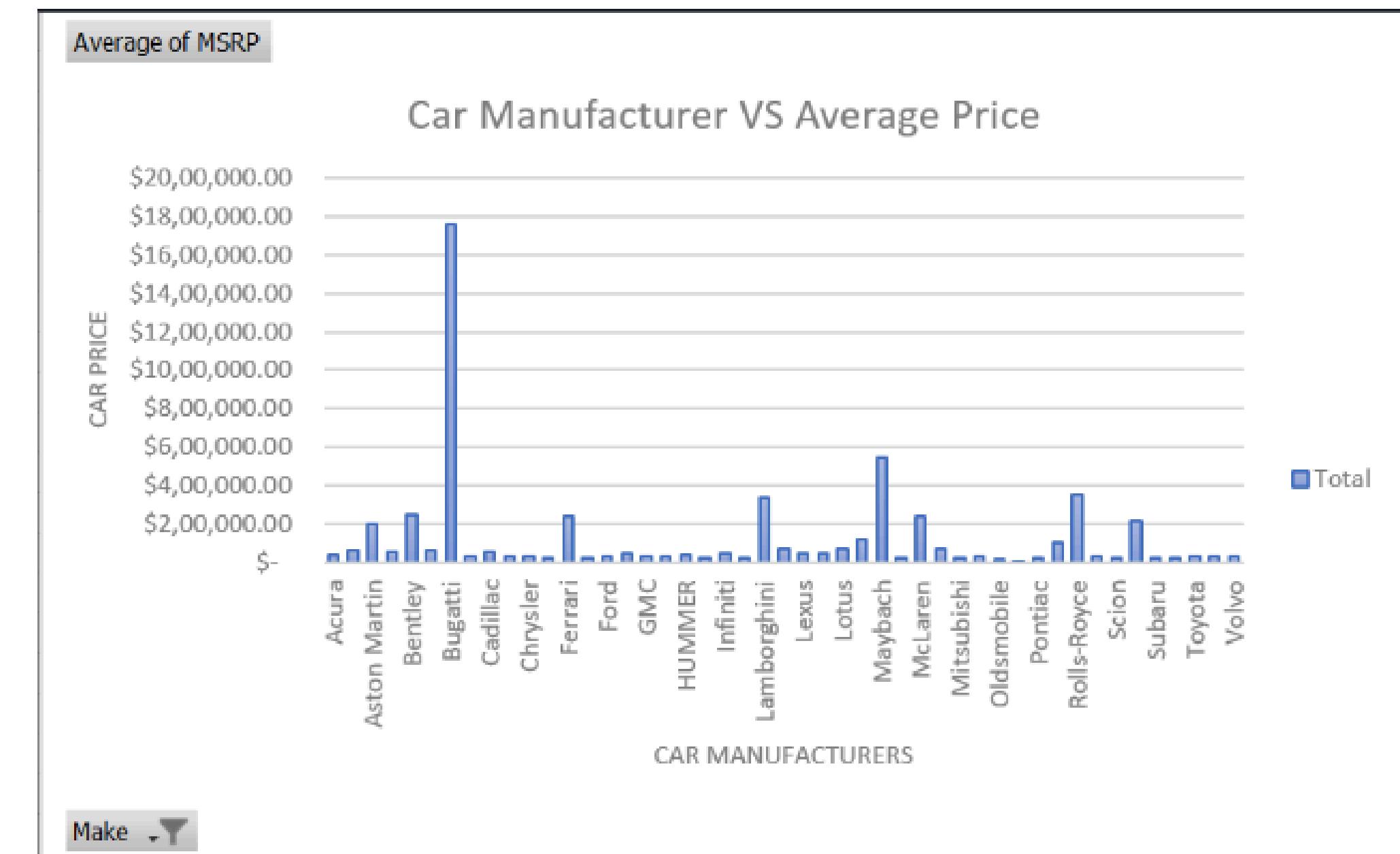
3.WHICH CAR FEATURES ARE MOST IMPORTANT IN DETERMINING A CAR'S PRICE?



FINDINGS



4. HOW DOES THE AVERAGE PRICE OF A CAR VARY ACROSS DIFFERENT MANUFACTURERS?

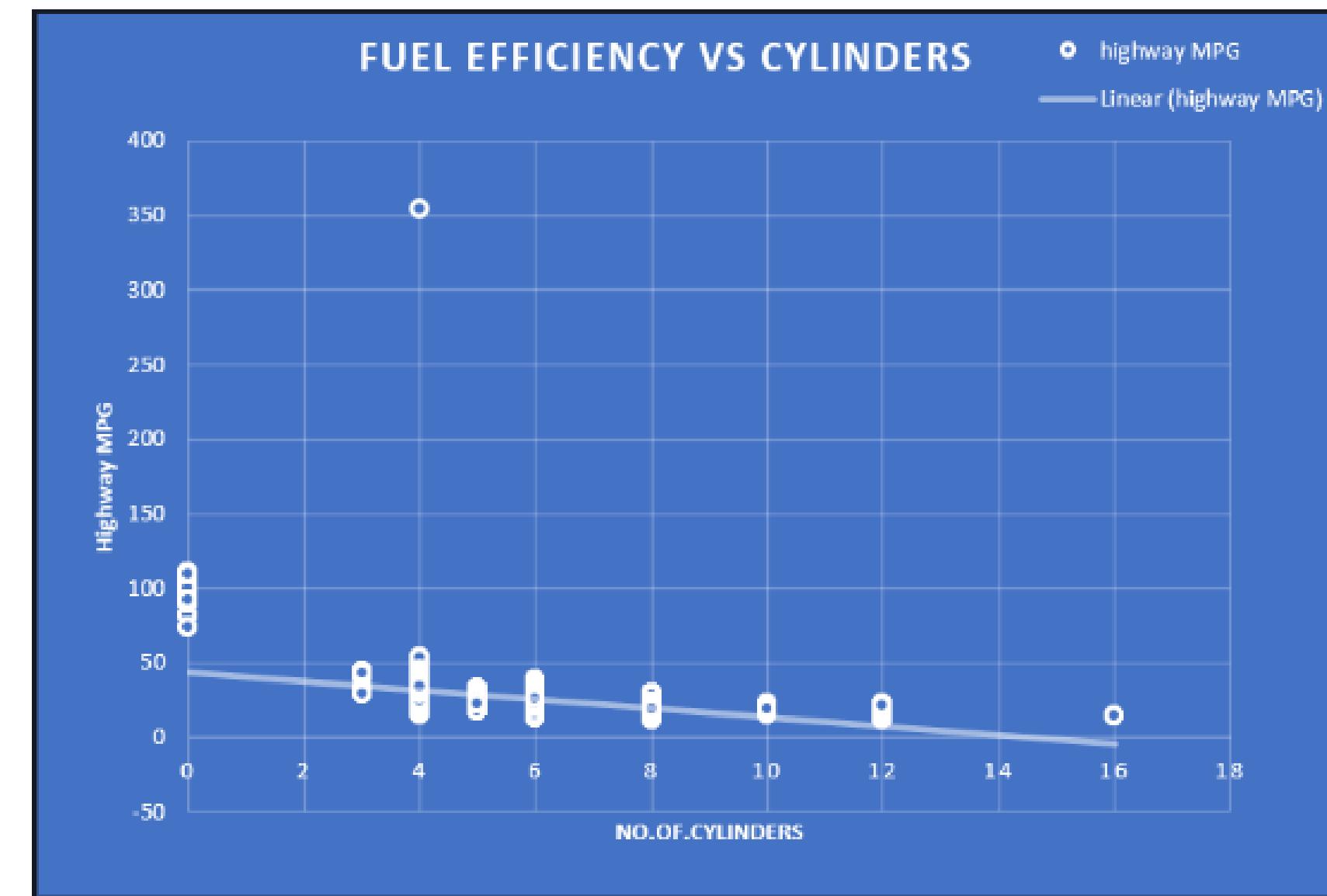


FINDINGS

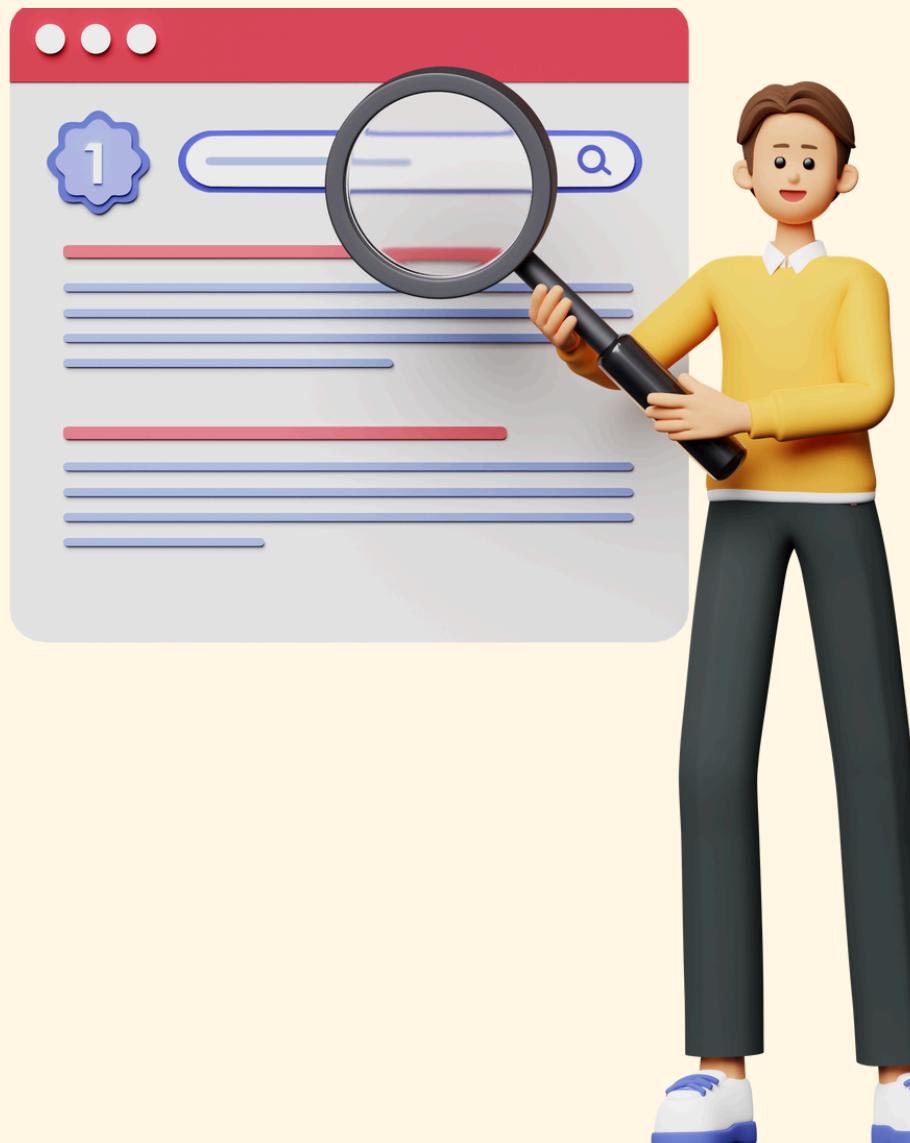


5.What is the relationship between fuel efficiency and the number of cylinders in a car's engine?

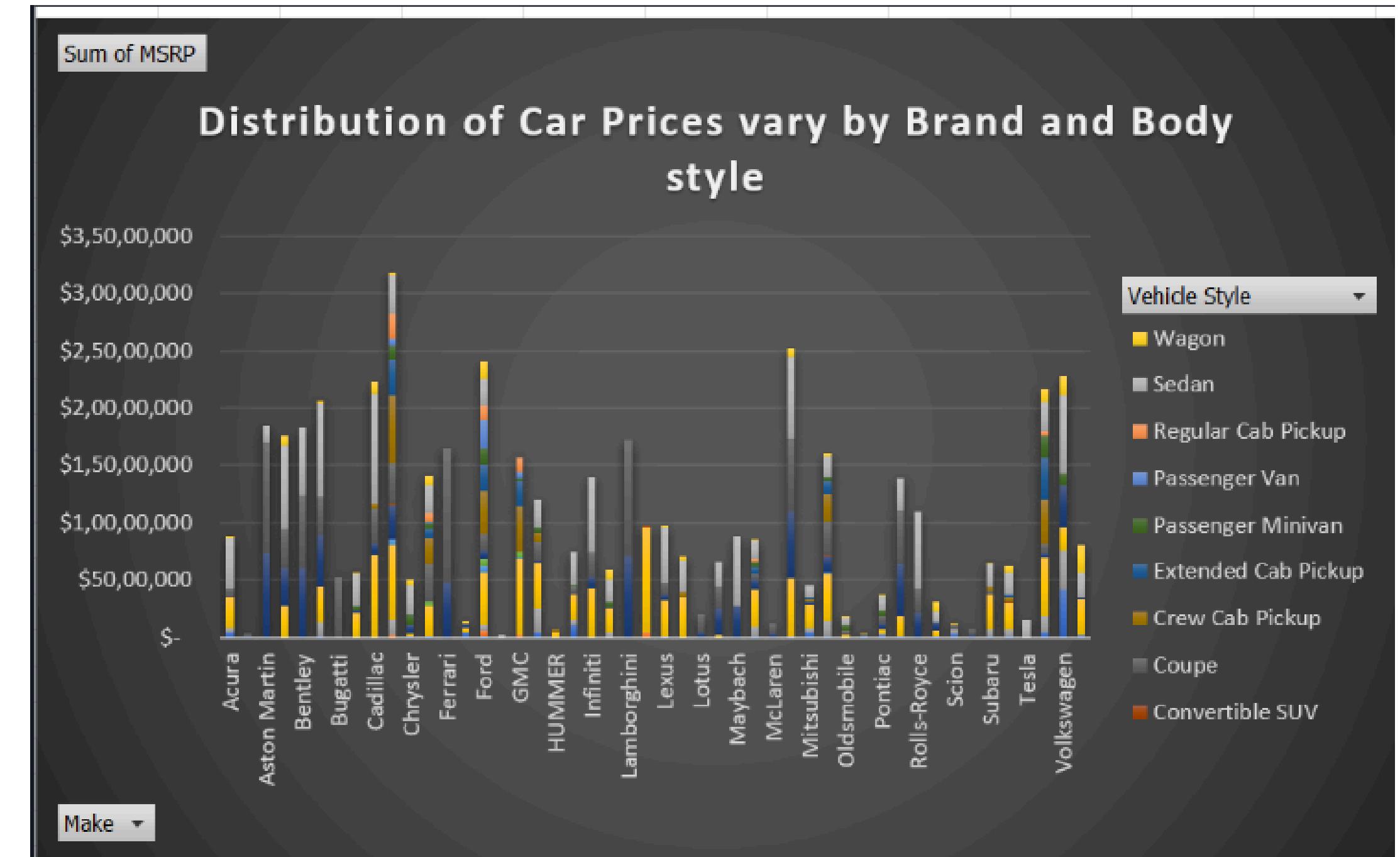
correlation coefficient
-0.621605733



FINDINGS



6. How does the distribution of car prices vary by brand and body style?

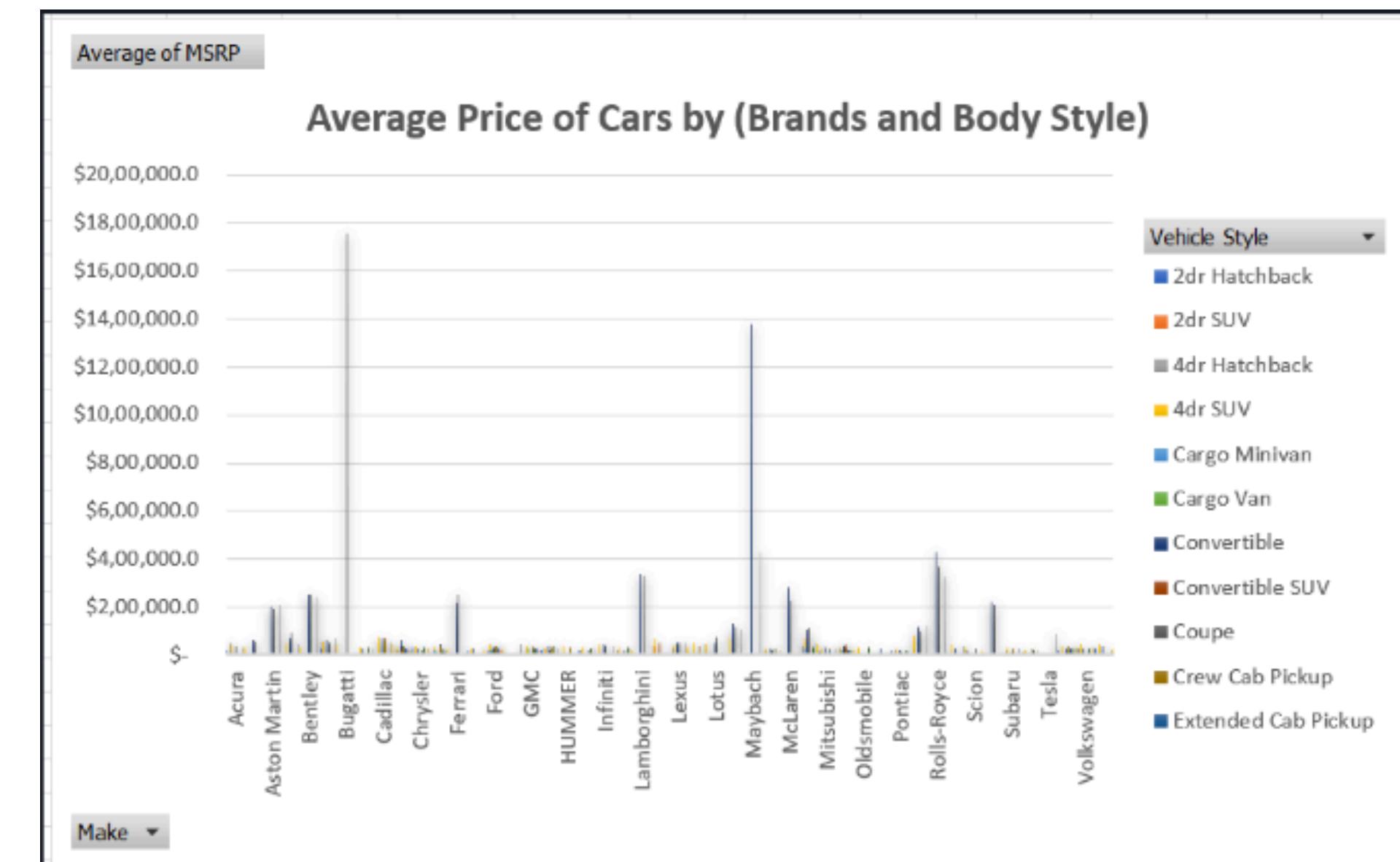


FINDINGS



7. Which car brands have the highest and lowest average MSRPs, and how does this vary by body style?

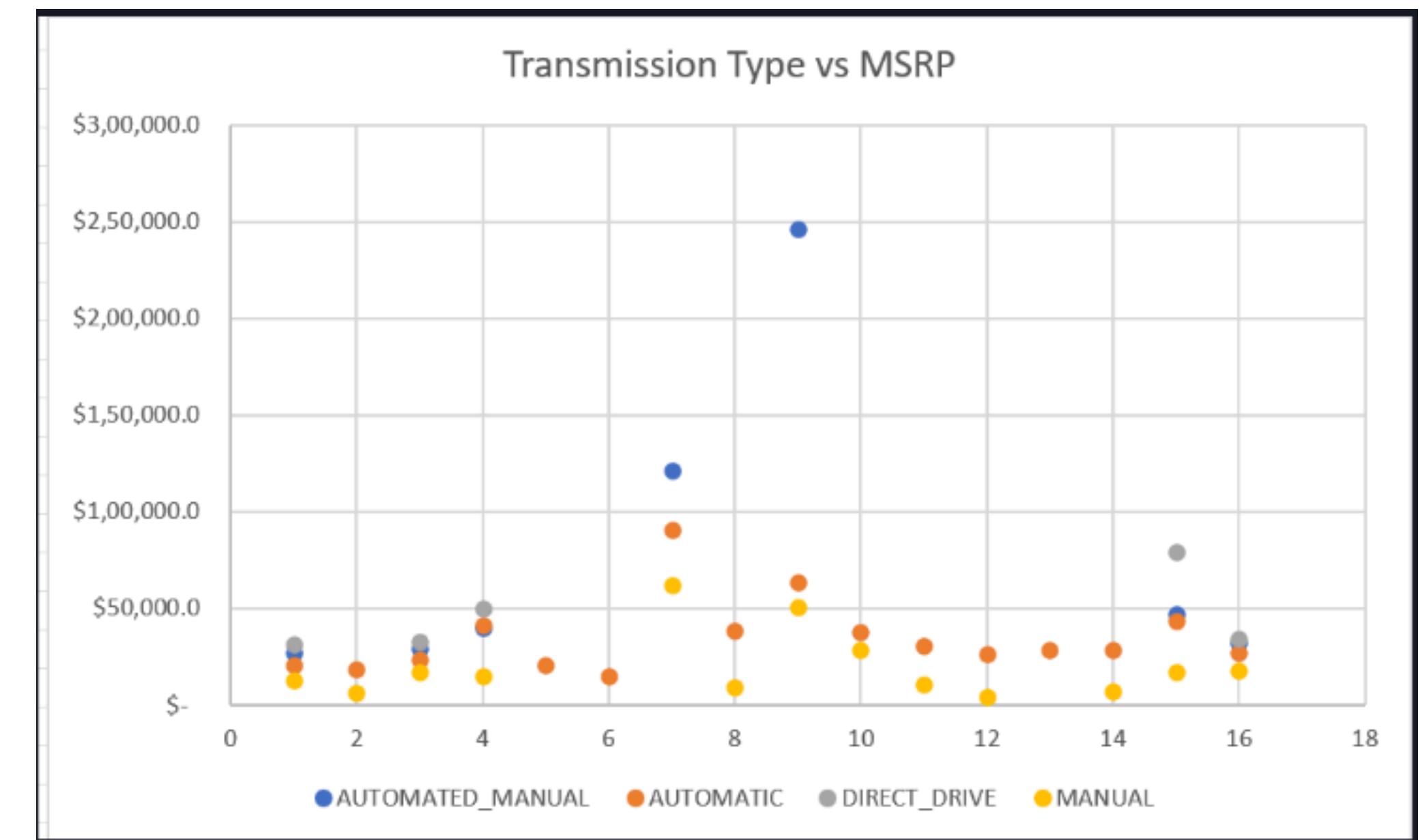
Highest	Bugatti	\$ 17,57,223.67
lowest	Plymouth	\$ 3,122.90



FINDINGS



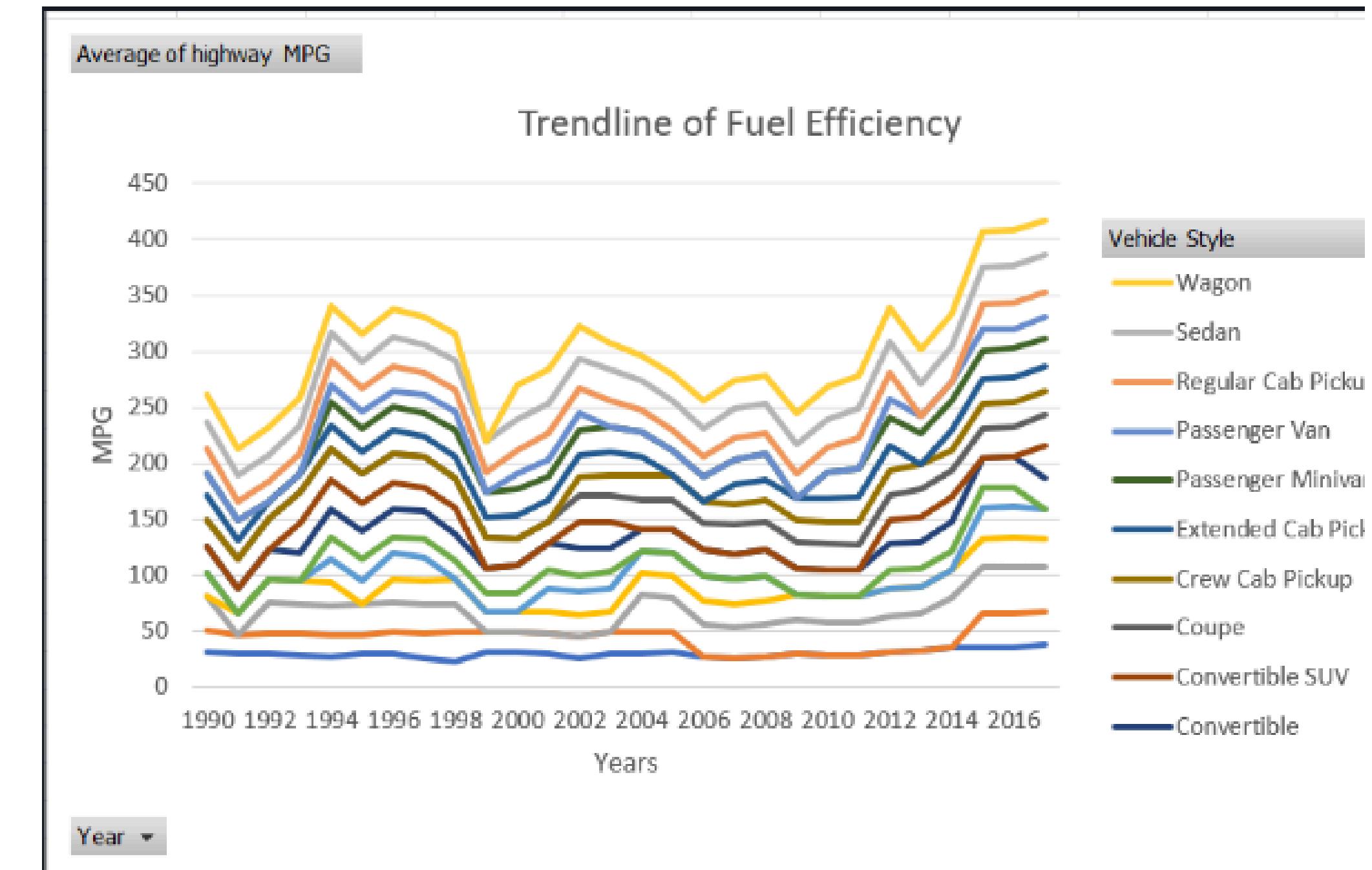
8. How do the different feature such as transmission type affect the MSRP, and how does this vary by body style?



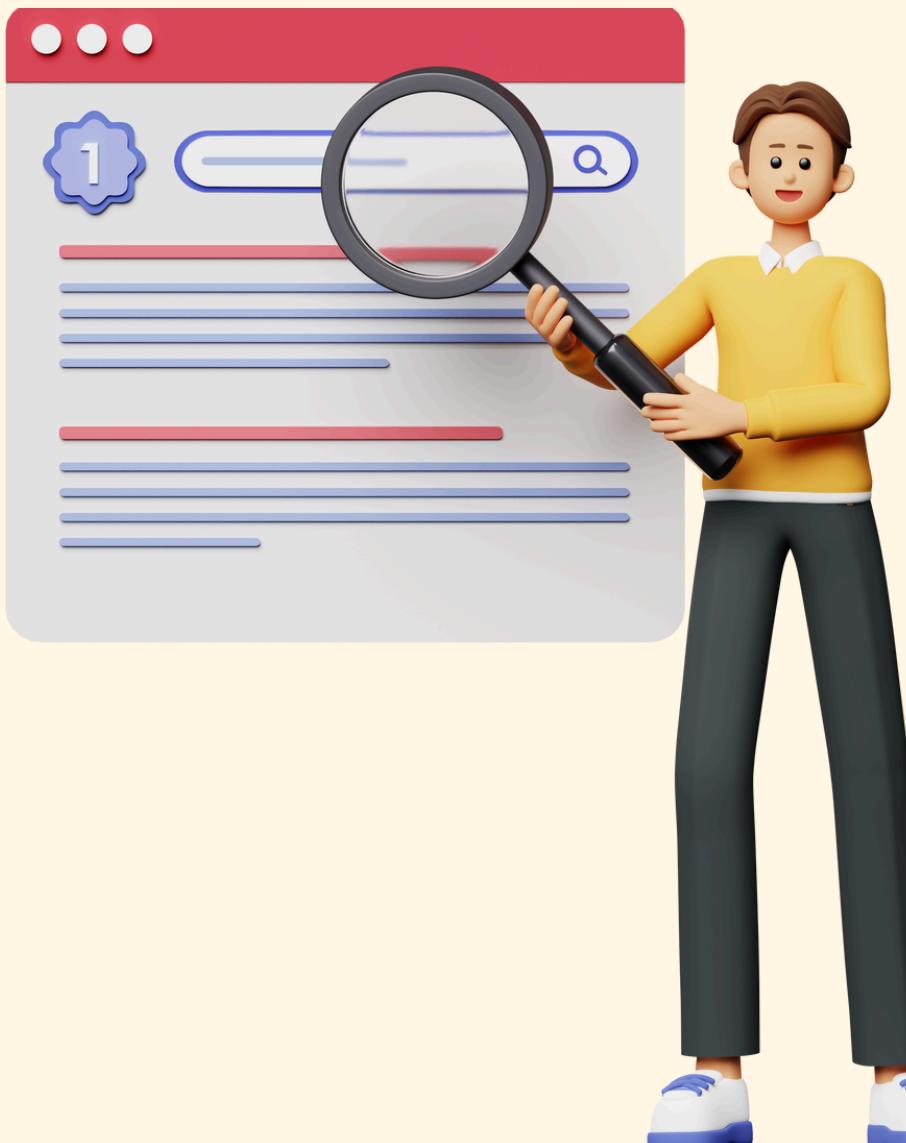
FINDINGS



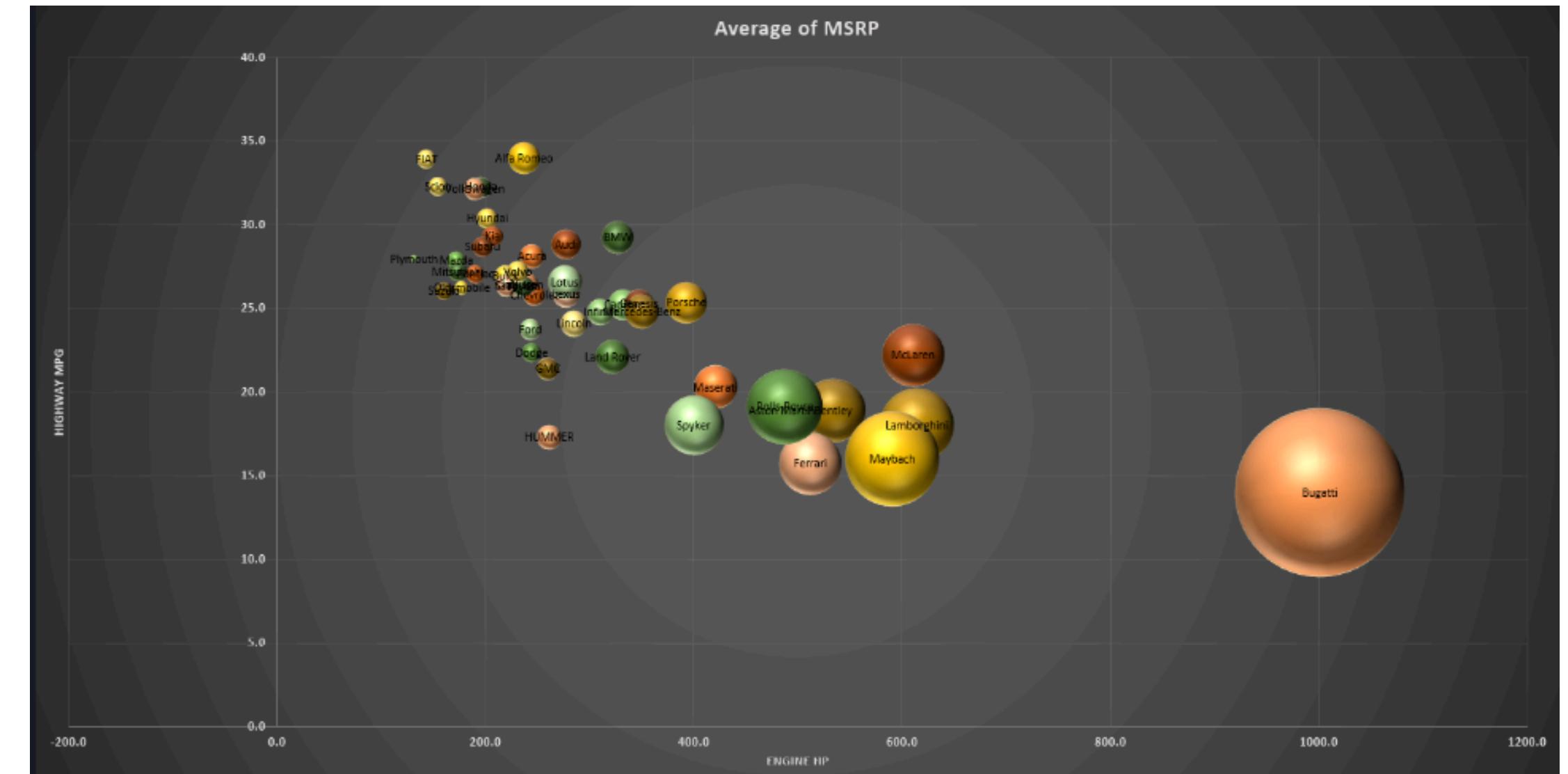
9. How does the fuel efficiency of cars vary across different body styles and model years?



FINDINGS



10. How does the car's horsepower, MPG, and price vary across different Brands?

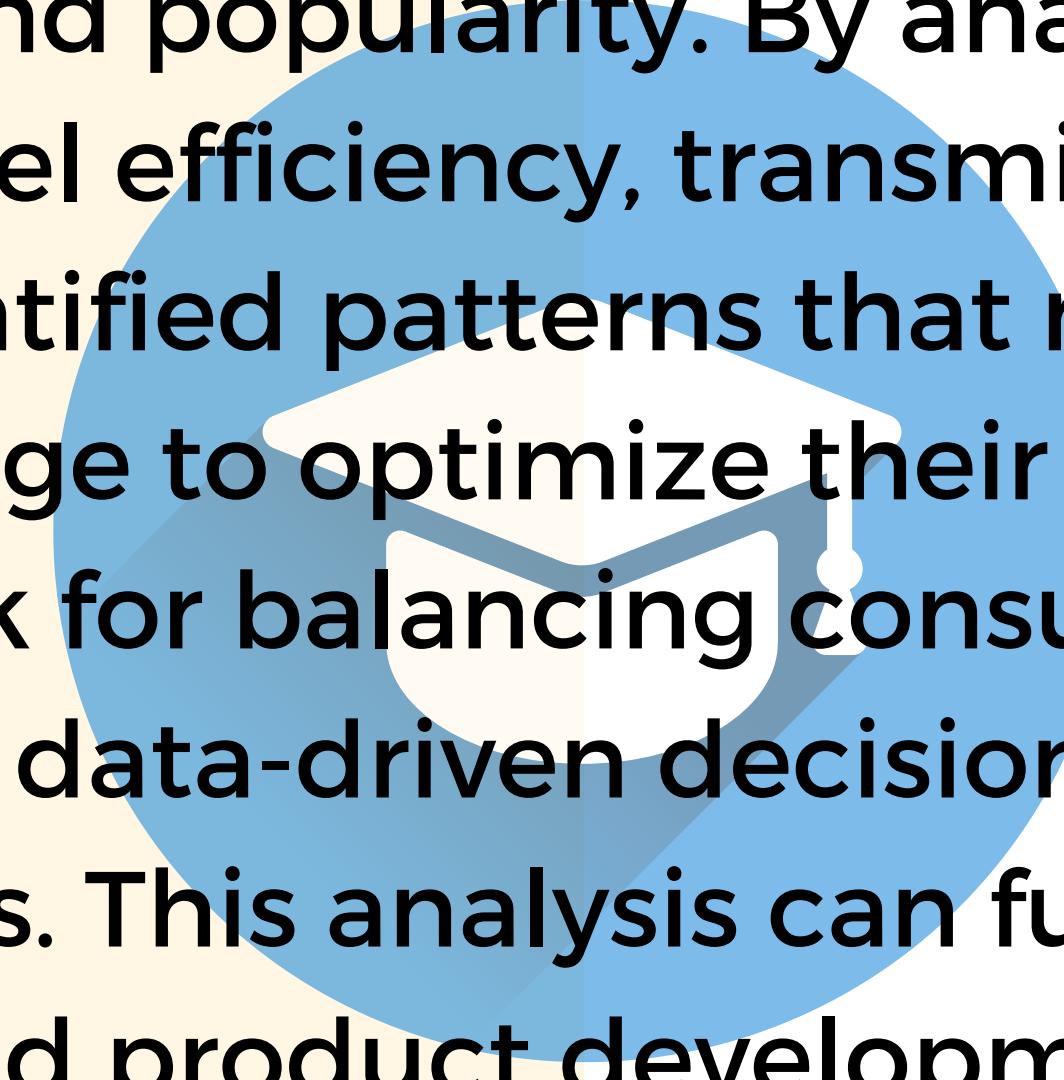


ANALYSIS



Through this project, a deep understanding of the features that impact the Car price and popularity was achieved. The analysis provided actionable insights that can be used to Analyze car features, market categories, and pricing to balance consumer demand with profitability.

CONCLUSION



This project has successfully uncovered the key factors that influence car price and popularity. By analyzing various features like engine type, fuel efficiency, transmission, and additional features, we identified patterns that manufacturers and dealerships can leverage to optimize their offerings. These insights provide a framework for balancing consumer preferences with profitability, enabling data-driven decisions in market positioning and pricing strategies. This analysis can further support targeted marketing efforts and product development in the automotive industry.

ABC CALL VOLUME TREND ANALYSIS

In this project, I analyzed the call volume trends for ABC, an insurance company, over 23 days. The goal was to understand call durations, volume patterns, and abandonment rates to improve manpower planning and reduce abandoned calls. I used Excel to analyze the data and suggest strategies for better staffing, particularly during peak hours and nighttime.

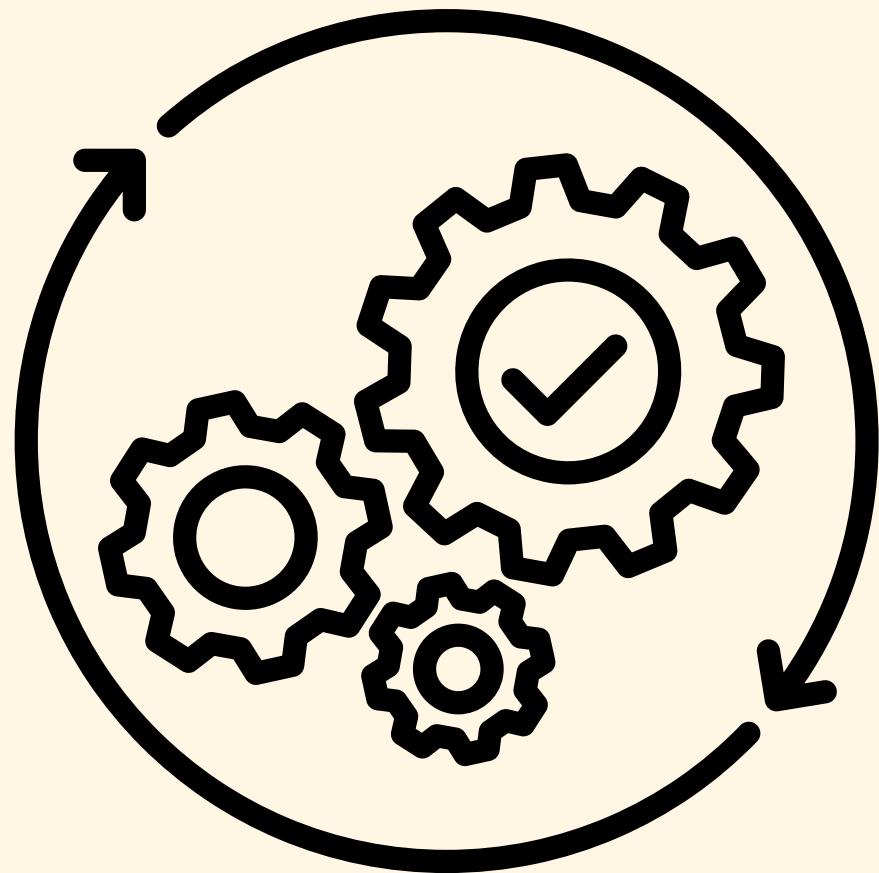


PROBLEM OF THE PROJECT



- The Problems in finding :-
 - Average Call Duration
 - Call Volume Trends
 - Manpower Allocation

DESIGN



- This project was done by using MS EXCEL 2022

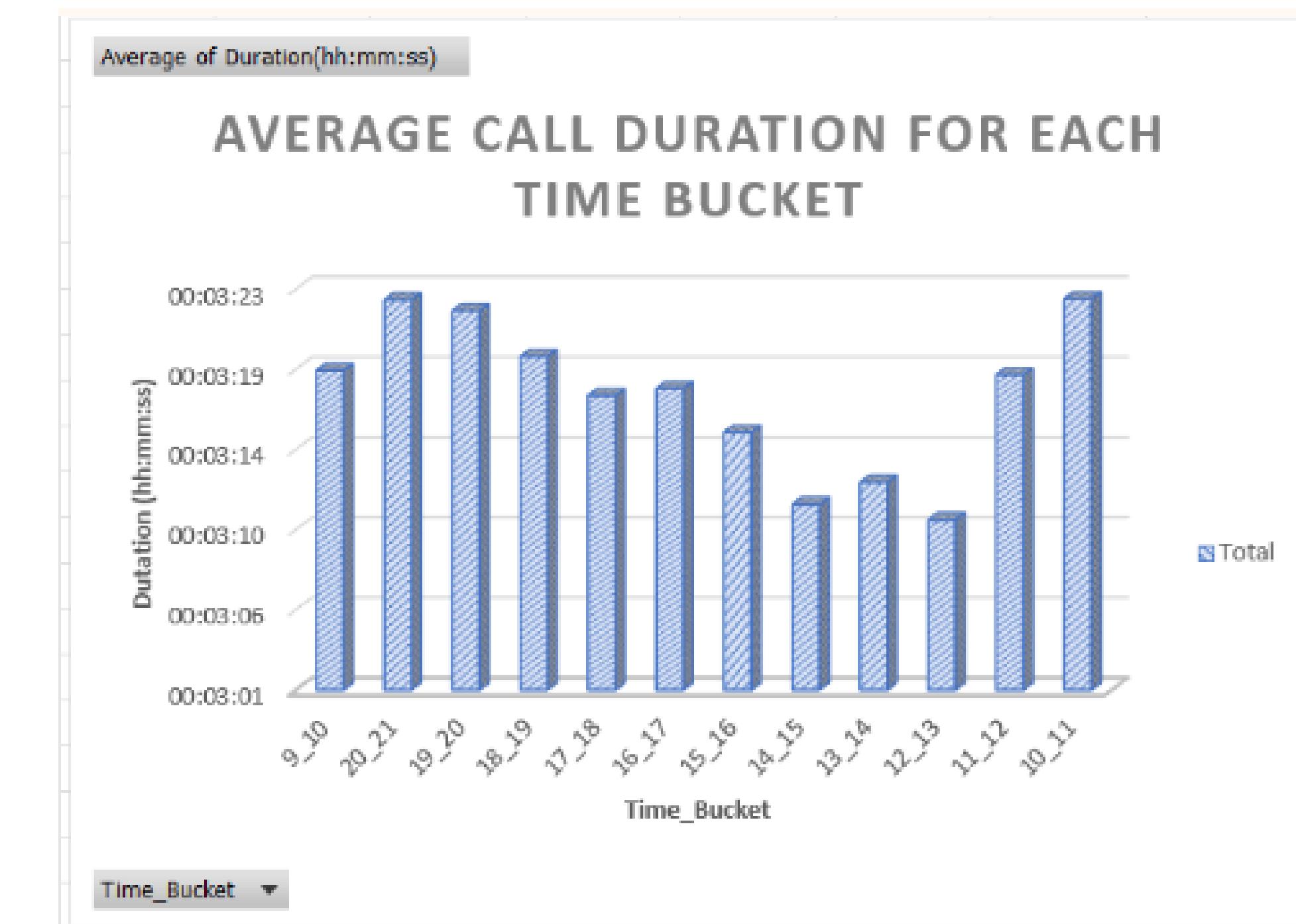


- Excel was used for data cleaning, analysis, and visualization. Its built-in functions and features such as pivot tables, charts, and conditional formatting

FINDINGS



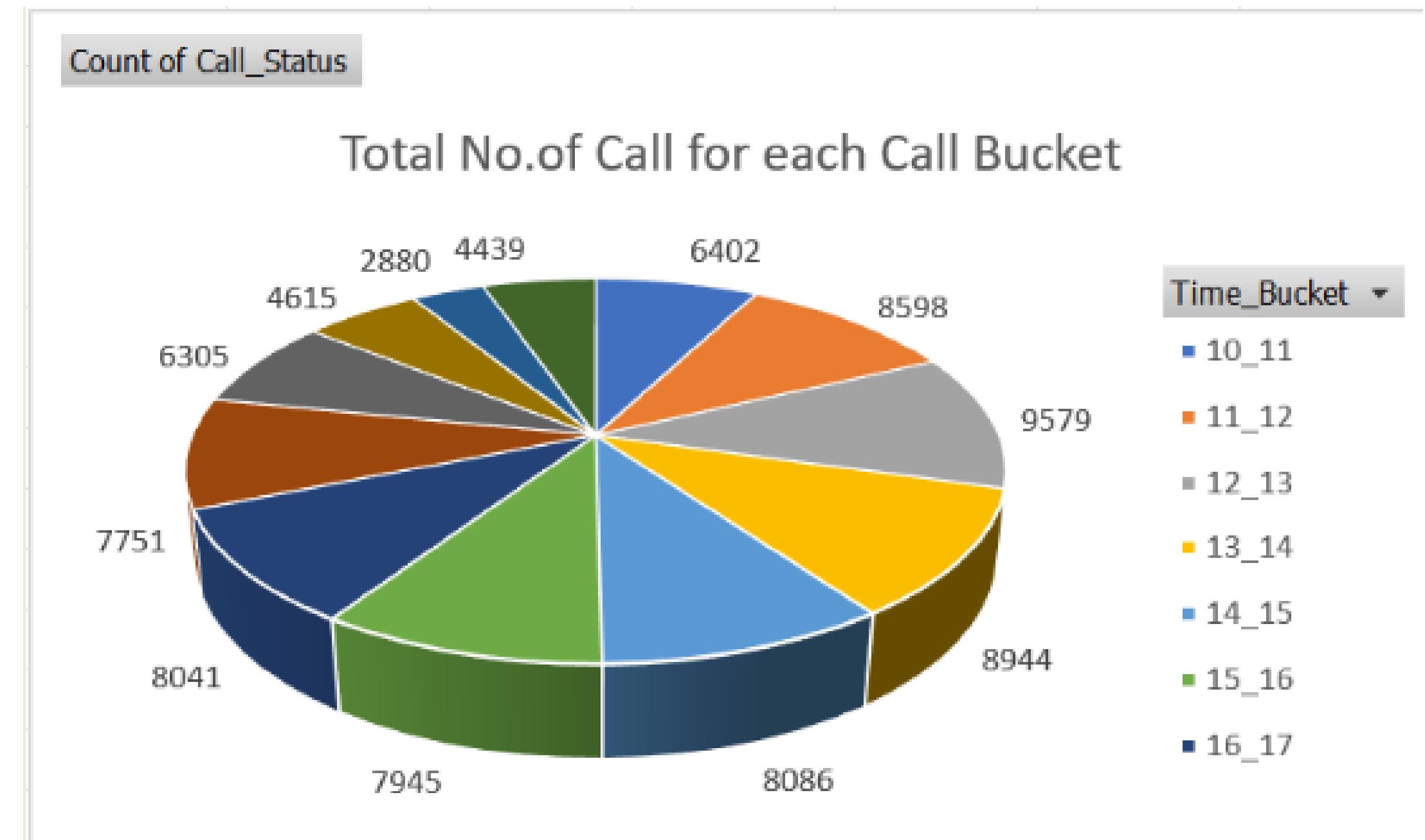
1.What is the average duration of calls for each time bucket?



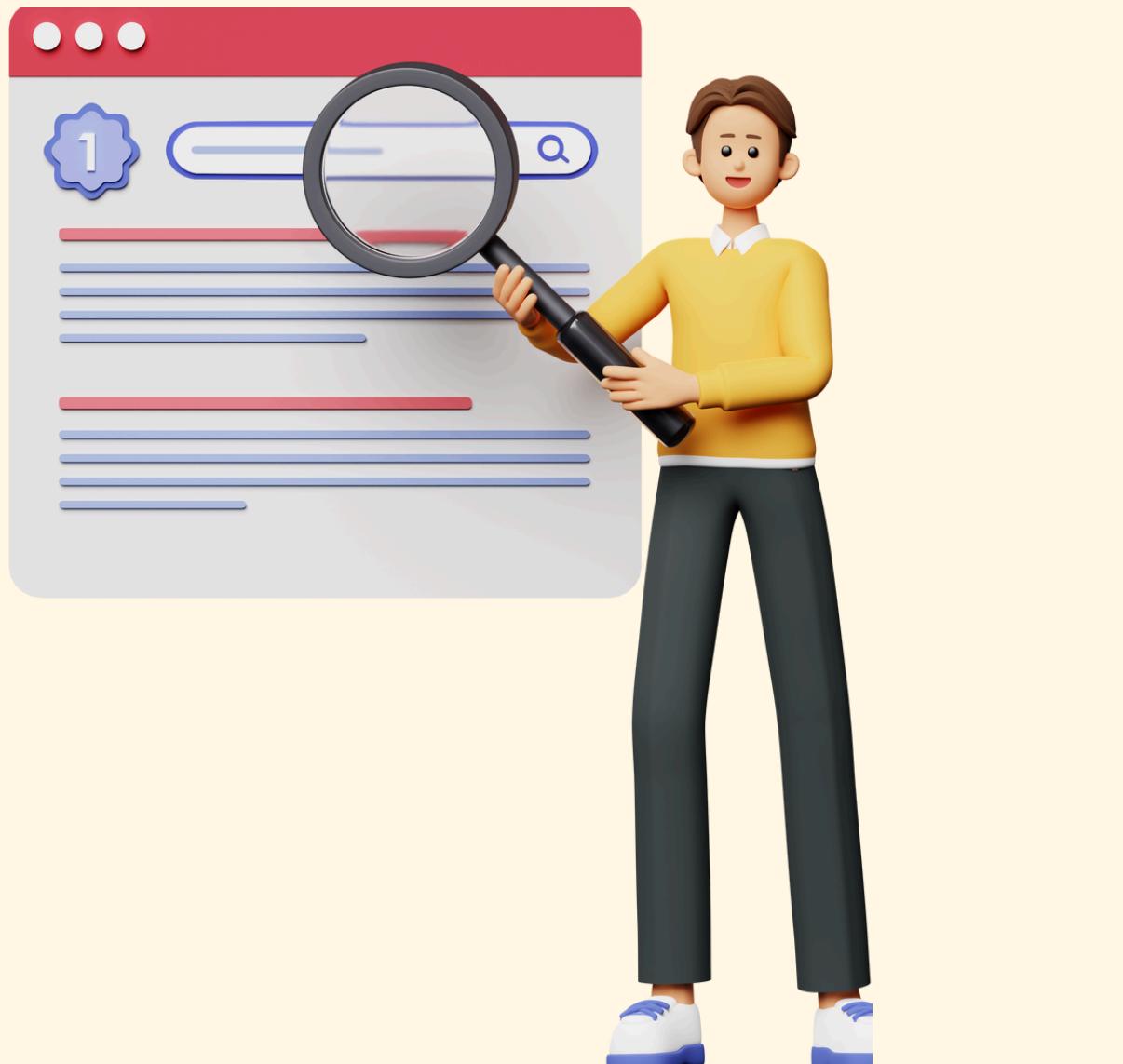
FINDINGS



2.the number of calls received in each time bucket?

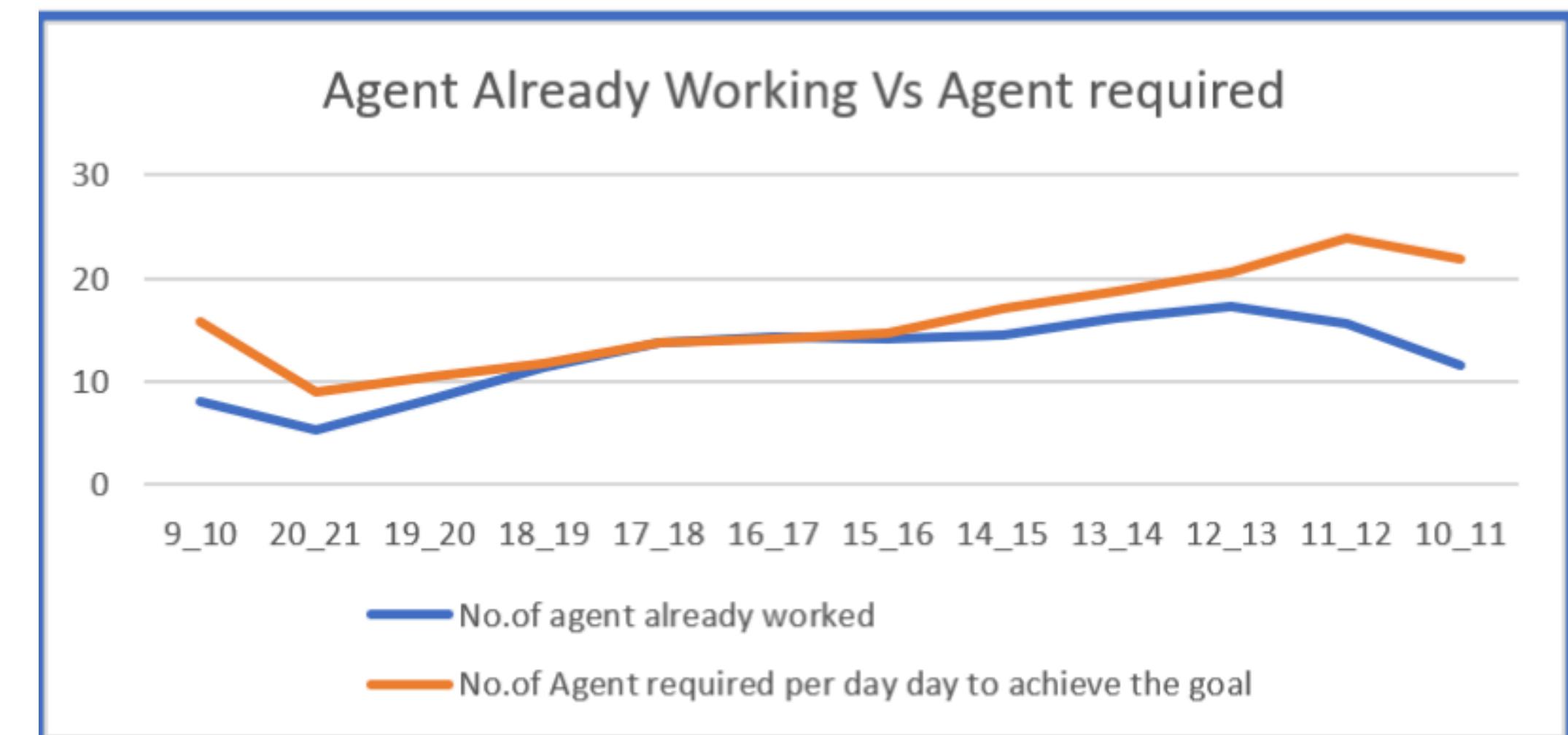


FINDINGS

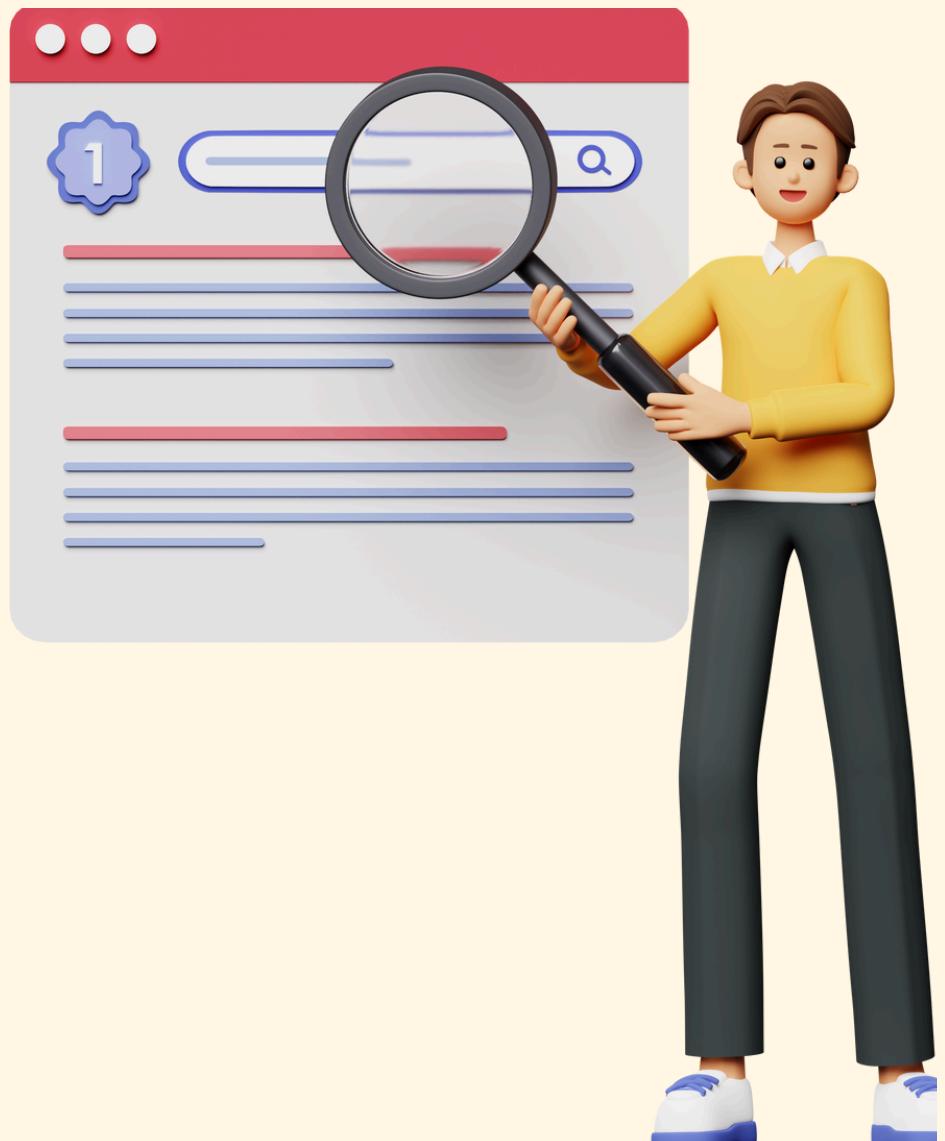


3.What is the minimum number of agents required in each time bucket to reduce the abandon rate to 10%?

the minimum number of agents required in each time bucket to reduce the abandon rate to 10%?		
Time_bucket	No.of Calls need to answered to reduce 30% to 10%	No.of Agent required per day day to achieve the goal
9_10	287	16
20_21	165	9
19_20	193	11
18_19	214	12
17_18	252	14
16_17	258	14
15_16	269	15
14_15	313	17
13_14	343	19
12_13	375	21
11_12	438	24
10_11	398	22
Grand total	3506	192

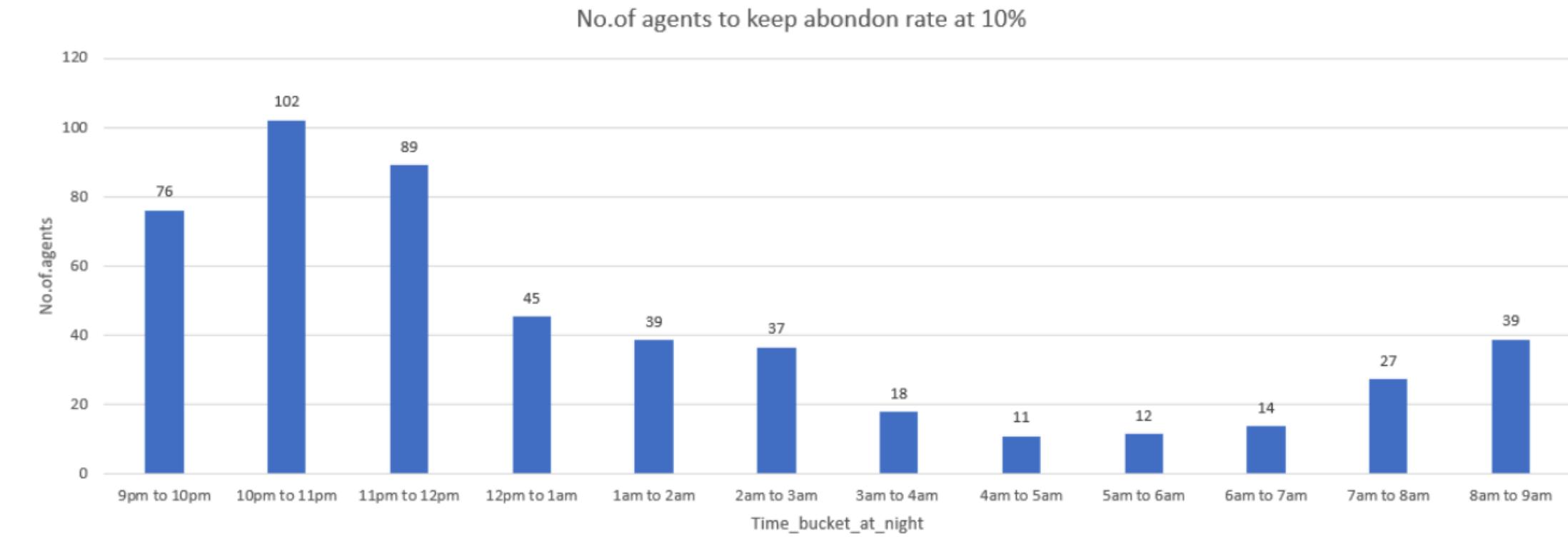


FINDINGS



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

Distribution of 30 Calls in night for every 100 calls coming in day(9am to 9pm)													Grand total
Time_bucket_at_Night	9pm to 10pm	10pm to 11pm	11pm to 12pm	12pm to 1am	1am to 2am	2am to 3am	3am to 4am	4am to 5am	5am to 6am	6am to 7am	7am to 8am	8am to 9am	
No.of agents required	76	102	89	45	39	37	18	11	12	14	27	39	508
No.of calls at night	1390	1866	1628	830	707	668	328	202	211	252	499	709	9289

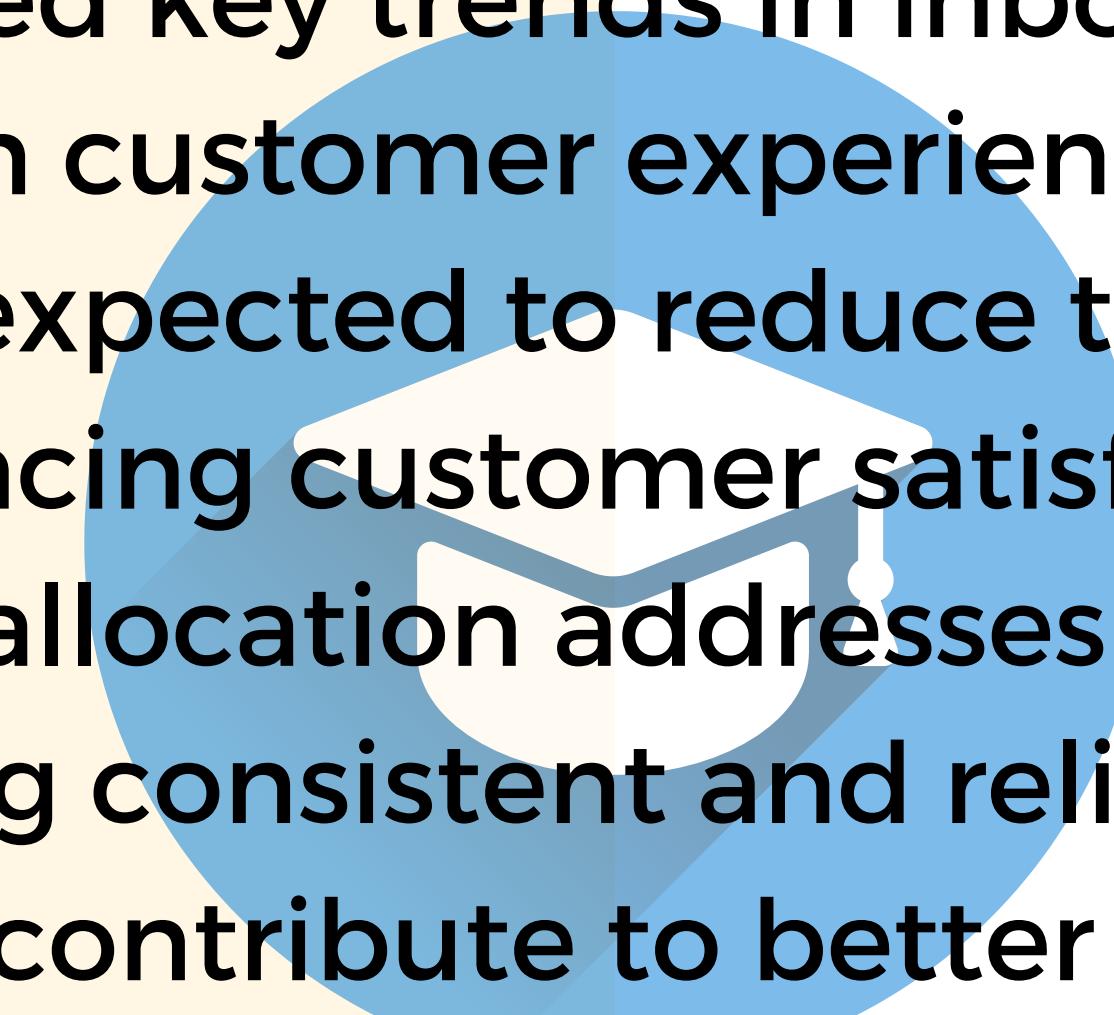


ANALYSIS



Through this project, I successfully identified the trends in inbound call volumes and their impact on customer experience. The insights gained led to a detailed manpower plan that, when implemented, should reduce the call abandonment rate to 10%, thereby improving overall customer satisfaction. Additionally, the proposed night shift allocation aims to address the current gap in after-hours customer support, ensuring a more consistent and reliable service experience for ABC's customers.

CONCLUSION



This project identified key trends in inbound call volumes and their impact on customer experience. The proposed manpower plan is expected to reduce the call abandonment rate to 10%, enhancing customer satisfaction. Additionally, the new night shift allocation addresses the gap in after-hours support, ensuring consistent and reliable service. These improvements will contribute to better overall service quality for ABC's customers.

APPENDIX:-

INSTAGRAM USER ANALYTICS

[https://docs.google.com/presentation/d/1K3ZOK_u2QMP4PZqqXH4mPCgCOO4jBB_t/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/1K3ZOK_u2QMP4PZqqXH4mPCgCOO4jBB_t/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

HIRING PROCESS ANALYTICS

[https://docs.google.com/presentation/d/1PcSqFmOCp0rnEs2Rfn06TtIt8Afxd6Ki/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/1PcSqFmOCp0rnEs2Rfn06TtIt8Afxd6Ki/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

OPERATION ANALYTICS AND INVESTIGATING METRIC SPIKE

[https://docs.google.com/presentation/d/10zNic9C7k6RGOSDwsLluElasPU1J1ZCY/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/10zNic9C7k6RGOSDwsLluElasPU1J1ZCY/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

IMDB ANALYSIS PROJECT

[https://docs.google.com/presentation/d/1hK89bwdoX1ek22fj9XldI_g3emGGoEQ-/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/1hK89bwdoX1ek22fj9XldI_g3emGGoEQ-/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

BANK LOAN CASE STUDY -LINK FOR PPT AND VIDEO EXPLANATION

[https://docs.google.com/presentation/d/1Vbpky5A7htrC5ACxwLm8IP7ACnaeSJks/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/1Vbpky5A7htrC5ACxwLm8IP7ACnaeSJks/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

[https://drive.google.com/file/d/15t_8oFfuv7eOD9Qb_05ntZnM0xQEoSfs/view?
usp=sharing](https://drive.google.com/file/d/15t_8oFfuv7eOD9Qb_05ntZnM0xQEoSfs/view?usp=sharing)

IMPACT OF CAR FEATURES

[https://docs.google.com/presentation/d/1OWCytVliaKFsb0RKPF8TOBBS9k0gO6Y/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/1OWCytVliaKFsb0RKPF8TOBBS9k0gO6Y/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

[https://drive.google.com/file/d/1OWwtDIJhVwunP6vagor4nwgBmasqTbL8/view?
usp=sharing](https://drive.google.com/file/d/1OWwtDIJhVwunP6vagor4nwgBmasqTbL8/view?usp=sharing)

ABC CALL VOLUME ANALYSIS

[https://docs.google.com/presentation/d/1NILRVHH_r4wzHGXOkLU89vbM6ZWnlazS/edit?
usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true](https://docs.google.com/presentation/d/1NILRVHH_r4wzHGXOkLU89vbM6ZWnlazS/edit?usp=sharing&ouid=104856657744948691536&rtpof=true&sd=true)

<https://drive.google.com/file/d/13GWA8gdETI2Y26FyEmwwyC4Y3eR7IHPb/view?usp=sharing>

THANK YOU

This Is My Detailed version of My Project
Portfolio

+91-8940944934

basiladour@gmail.com

By Basil L