



1



Data Analysis Portfolio

Mr. Raj Rathod

Data Analyst



[Watch Video Presentation](#)





Professional Background

2

Introduction

- Greetings! I am Mr. Raj Rathod, an aspiring Data Analyst with a wealth of experience and a track record of excellent achievements.
- With a background in Electrical Engineering and notable experience in various roles, including Graduate Engineer Trainee and Supervisor, I bring a diverse skill set and valuable insights to the field of data analysis.
- Throughout my professional journey, I have received accolades such as the IMDb India's Star Contributor Award and certifications in Ecommerce Analytics from Google, reflecting my commitment to excellence and continuous learning.
- Join me as I share my background, experiences, and aspirations for a fulfilling career in data analysis.



Academic Credentials

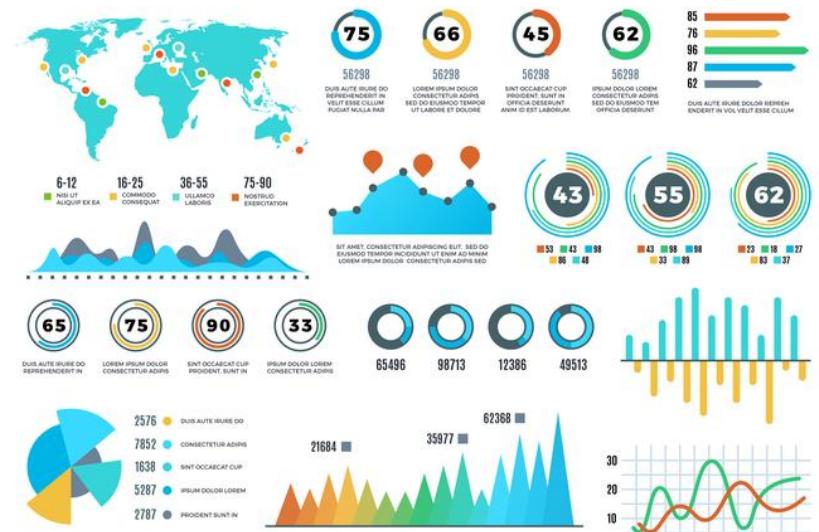
3

Educational Background :-

- Bachelor's Degree in Electrical Engineering
- Graduated with 9.31 CGPA

Skills :-

- Data Analysis
- Data Visualization
- SQL
- Microsoft Excel
- Microsoft Office
- Leadership
- Effective Communication
- Analytical and Problem Solving Skills



Professional Experience



Graduate Engineer Trainee at NTPC Ltd. (Leading Power Generation Company)

- Managed project timelines and budgets
- Designed maintenance programs for equipment
- Collaborated on engineering projects
- Increased operational efficiency
- Received Certificate of Merit



Data Capturer at IMDb.com

- Contributed to over 2,500 titles (movies/TV shows)
- Conducted 6,000+ comprehensive reviews
- Received India's Star Contributor Award by IMDb CEO
- Over 1 million views for contributions



Other Achievements



- Quora Content Writer: Views 2.4 Million+
- DesiDime Community Contributor: Areas - Finance, Technology, E-commerce
- Google Certification: Ecommerce Analytics
- IMDb India's Star Contributor Award by IMDb CEO
- Internship at MahaGENCO with Gold Medal
- Virtual Internship at Trainity
- Completed Data Analysis Course
- Certificate for 8 Live Projects of Data Analysis
- Certificate of Merit from NTPC



Data Analytics Certificates

trainity

Live Project Certificate

This is to certify

Raj Rathod

for successfully completing a [Data Analytics LIVE project](#) based on analysis of user data of a social media platform, with Trainity

13-01-2024

DATE

AUTHORIZED SIGNATORY

trainity



Certificate ID: 181919

This is to certify that
Raj Rathod

Has successfully completed the

8 Weeks Data Analytics

Specialization Training

5-02-2024

DATE

AUTHORIZED SIGNATORY



Raj Rathod

Add profile credential

203 followers • 9 following

Squealer

[Profile](#) 212 Answers 2 Questions 1 Post 203 Followers Following Edits Activity

Credentials & Highlights

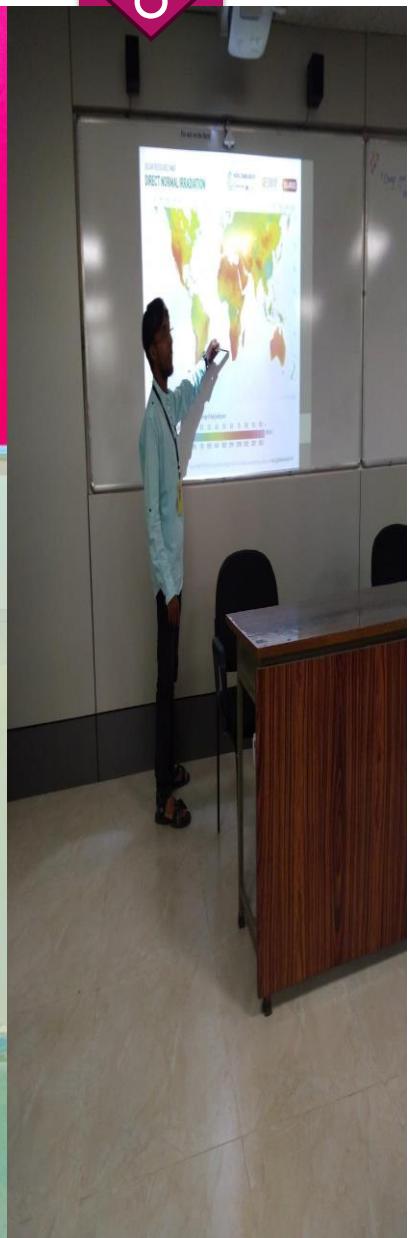
- Add employment credential
- Add education credential
- Add location credential
- 2.4M content views 8.2K this month
- Active in 1 Space
- Joined March 2016

↑ Quora

NTPC Excellence Award

MahaGenco Presentation (Gold MEDAL)

6





Career Transition

7

- Intention to transition into a Data Analyst role
- Motivation for the transition, such as passion for data analysis and desire for new challenges
- Relevant skills and experiences that support the transition, such as analytical skills, proficiency in Excel, and completion of a Data Analysis Course
- Commitment to continuous learning and professional development, demonstrated through certifications, internships, and completed projects
- Future career goals, including aspirations for growth and contribution to the field of data analysis



Future Plans



- Plan to actively pursue opportunities in the field of data analysis
- Intent to apply skills and knowledge gained from coursework, projects, and internships to real-world challenges
- Commitment to networking, further education, and skill development to enhance qualifications and career prospects
- Willingness to adapt and grow in response to industry trends and emerging technologies



Table Of Contents

9

Project Number	Project Name	Page No
1	Data Analytics Process - Doctor Selection	10
2	Instagram User Analytics	15
3	Operation Analytics and Investigating Metric Spike	27
4	Hiring Process Analytics	44
5	IMDb Movie Analysis	55
6	Bank Loan Case Study	72
7	Analyzing the Impact of Car Features on Price and Profitability	95
8	ABC Call Volume Analysis	115



Project 1 :Data Analytics Process

10

By : Raj Rathod
Data Analyst

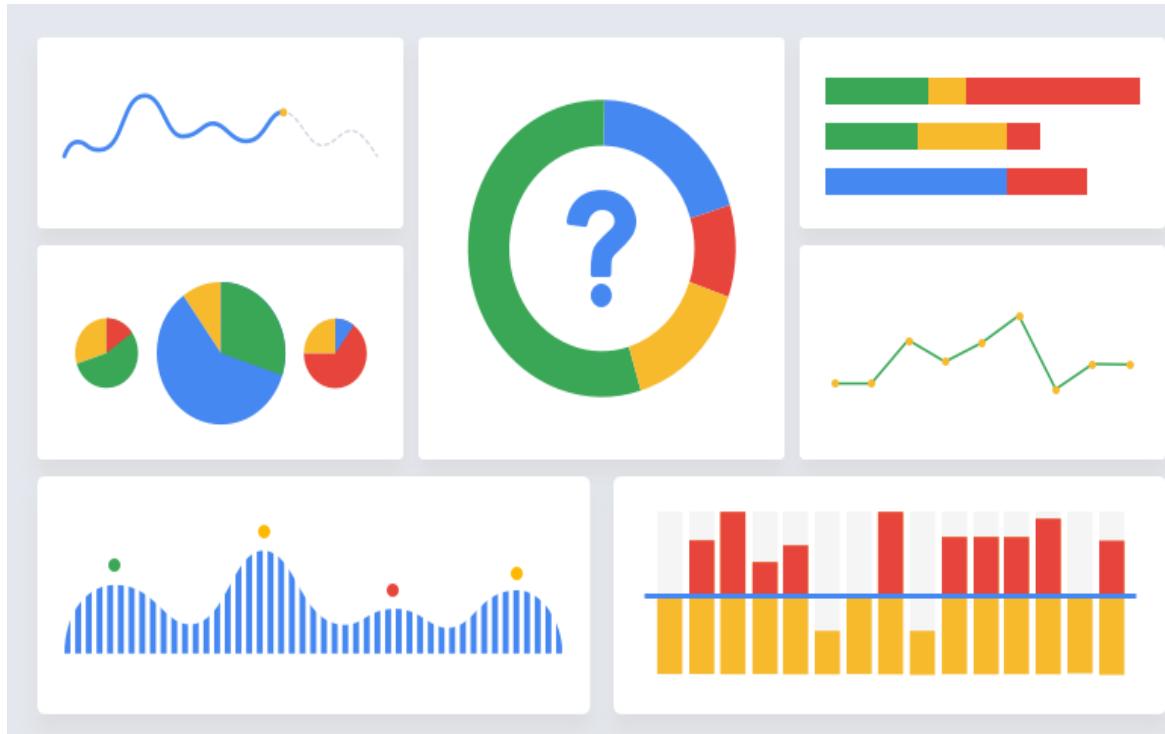
Doctor Selection Using Data Analytics

Description:

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

Task :

To Provide an example or examples of real-life situations where Data Analytics is utilized, and connect it with the Data Analytics process.



Project Approach

11

1. Ask/Plan:

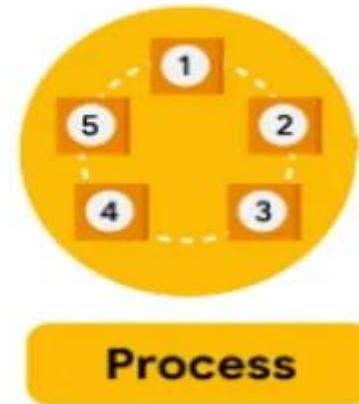
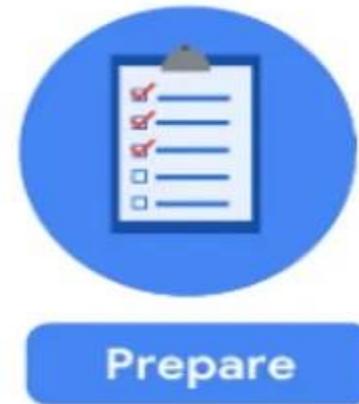
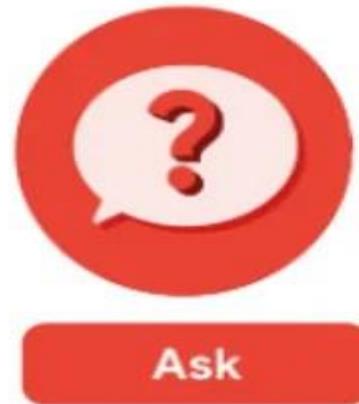
- Identify purpose: Treat hair fall issue
- Requirements: Qualified dermatologist, good clinic, trained staff

2. Prepare:

- Determine treatment budget
- Research advanced techniques
- Estimate time for results

3. Process:

- Gather data from online sources
- Ratings, reviews, clinic websites
- Check for latest techniques offered
- Look for evidence of successful treatments



Ask questions and define the problem.

Prepare data by collecting and storing the information.

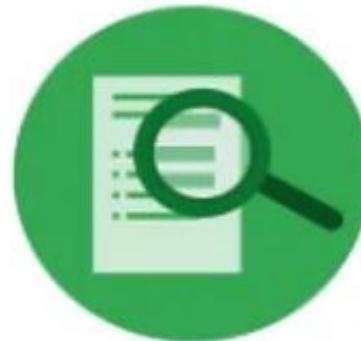
Process data by cleaning and checking the information.

Project Description and Conclusion

12

4. Analyze:

- Low ratings
- Unaffordable costs
- Lack of experience
- Poor infrastructure



Analyze



Share



Act

5. Share:

- Seek recommendations
- Friends, family
- Online forums
- Check past patient experiences

6. Act:

- Make an informed decision
- Select the best doctor/clinic
- Eliminate unsuitable options

Analyze data to find patterns, relationships, and trends.

Share data with your audience.

Act on the data and use the analysis results.

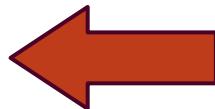
Project 1: Doctor Selection using Data Analytics

13

Projects Overview

- **Description:** Utilized data analytics to select the best dermatologist for treating hair fall issues.
- **Approach:** Gathered data from online sources, analyzed ratings and reviews, and made an informed decision based on evidence.
- **Outcome:**
 - ✓ Demonstrated the application of the data analytics process in a real-life scenario. (doctor selection for treating hair fall)
 - ✓ Highlighted the importance of data-driven decision-making for optimal outcomes





Access Project

Thank
you





Project 2: Instagram User Analytics

15

Introduction:

By : Raj Rathod
Data Analyst

Instagram



\$952
Photos



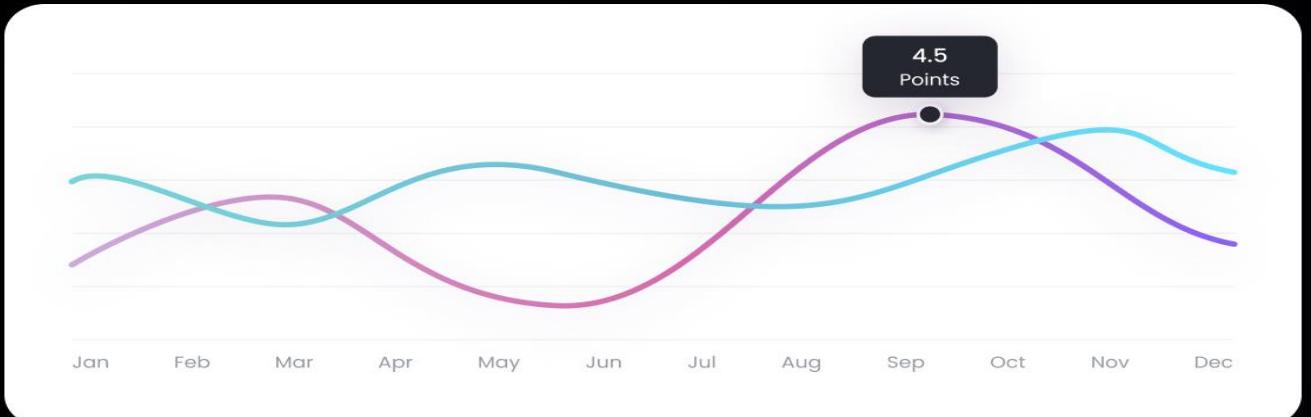
\$835
Internet



\$197
Movies



\$643
Music





Project Approach

16

Task:

Studying user sign-up behavior on different days of the week to determine the best day for running advertisements.

Output:

- **Data Import:** Utilized MySQL Workbench to load and verify the accuracy of provided data.
- **Exploratory Data Analysis:** Conducted analysis to understand user registration distribution over time and identify peak registration days.
- **SQL Queries:** Leveraged SQL to extract important insights, including the day with the highest user registrations.
- **Tech-Stack Used:**
 - ✓ MySQL Workbench: Chosen for its user-friendly interface and powerful SQL querying capabilities.
 - ✓ Familiarity with the tool facilitated efficient data exploration and manipulation.

A. Marketing Analysis:

17

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

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

Output : These are the five oldest users as per the provided Instagram Database.



MySQL Workbench

Instagram User Analytics Trial... Instagram User Analytics

File Edit View Query Database Server Tools Scripting Help

Schemas

ig_clone

zefla

rta

world

Trinity Instagram User Analytics

SQL File 1 SQL File 2 SQL File 3 SQL File 4 SQL File 5

1 * SELECT username, created_at
2 FROM users
3 ORDER BY created_at ASC
4 LIMIT 5;

Result Grid Filter Rows Export Wrap Cell Content

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

Administration Schemas Information

No object selected

user 1 x

Action Output

#	Time	Action	Message	Duration / Fetch
16	19:36:40	INSERT INTO photo_tag(photo_id,tag_id) VALUES (1,16),(1,17),(1,21),(1,19),(2,4),(2,3),(2,20,...)	501 rows affected Records: 501 Duplicates: 0 Warnings: 0	0.016 sec / 0.000 sec
17	22:16:28	SELECT username, created_at FROM users ORDER BY created_at ASC LIMIT 5	5 rows(s) returned	0.000 sec / 0.000 sec
18	22:34:41	SELECT user_id,username FROM users u LEFT JOIN photos p ON u.id = p.user_id WHERE p.user_id IS NULL	5 rows(s) returned	0.000 sec / 0.000 sec
19	22:47:24	SELECT tag_name,COUNT(*) as tag_count FROM tag JOIN photo_tag p ON tag.tag_id = photo_tag.tag_id GROUP BY tag_name	5 rows(s) returned	0.016 sec / 0.000 sec
20	23:20:14	SELECT tag_name,COUNT(*) AS tag_count FROM tag JOIN photo_tag p ON tag.tag_id = photo_tag.tag_id GROUP BY tag_name	5 rows(s) returned	0.000 sec / 0.000 sec
21	23:51:59	SELECT user_id,COUNT(*) AS like_count FROM likes GROUP BY user_id ORDER BY like_count DESC LIMIT 5	1 rows(s) returned	0.016 sec / 0.000 sec

Object Info Session

username	created at
Darby_Herzog	06-05-2016 00:14
Emilio_Bernier52	06-05-2016 13:04
Elenor88	08-05-2016 01:30
Nicole71	09-05-2016 17:30
Jordyn.Jacobson2	14-05-2016 07:56



INACTIVE

A. Marketing Analysis:

18

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

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

Output: These are the users who have never posted a single photo on Instagram. These are the users where the team has to encourage users by sending them promotional emails.

The screenshot shows the MySQL Workbench interface. On the left, the Navigator pane displays the schema 'ig_clone' with tables like 'users' and 'photos'. A SQL editor window contains the following query:

```
1 *   SELECT u.id, u.username
2   FROM users u
3   LEFT JOIN photos p ON u.id = p.user_id
4   WHERE p.user_id IS NULL;
```

The results grid shows the output of the query, listing user IDs and their corresponding usernames. The results are identical to the table on the right.

user_id	username
5	Aniya_Hackett
7	Kassandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

user_id	username
5	Aniya_Hackett
7	Kassandra_Homenick
14	Jaclyn81
21	Rocio33
24	Maxwell.Halvorson
25	Tierra.Trantow
34	Pearl7
36	Ollie_Ledner37
41	Mckenna17
45	David.Osinski47
49	Morgan.Kassulke
53	Linnea59
54	Duane60
57	Julien_Schmidt
66	Mike.Auer39
68	Franco_Keebler64
71	Nia_Haag
74	Hulda.Macejkovic
75	Leslie67
76	Janelle.Nikolaus81
80	Darby_Herzog
81	Esther.Zulauf61
83	Bartholome.Bernhard
89	Jessyca_West
90	Esmeralda.Mraz57
91	Bethany20

A. Marketing Analysis:

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

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

Output: The winner of the contest is “**user id – 5**” and he has the most liked on a single photo that is **257** across all users

The screenshot shows the MySQL Workbench interface. The top menu bar includes File, Edit, View, Query, Database, Server, Tools, Scripting, and Help. Below the menu is a toolbar with various icons. The left sidebar has a 'Navigator' section with 'SCHEMAS' and a tree view showing 'ig_clone' (Tables, Views, Stored Procedures, Functions) and other schemas like 'sakila', 'sys', and 'world'. The main area contains a SQL editor window with the following code:

```
1 * SELECT user_id, COUNT(*) AS like_count
2 FROM likes
3 GROUP BY user_id
4 ORDER BY like_count DESC
5
6
LIMIT 1;
```

Below the SQL editor is a 'Result Grid' window showing the output of the query:

user_id	like_count
5	257

The bottom status bar indicates 'Ready'.



A. Marketing Analysis:

20

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

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

Output: Top Five Most Commonly Used Hashtag on the Instagram are as below and it will help the partner brand for marketing purpose.

The screenshot shows the MySQL Workbench interface. In the top navigation bar, there are tabs for 'File', 'Edit', 'View', 'Query', 'Database', 'Server', 'Tools', 'Scripting', and 'Help'. Below the tabs, the 'Schemas' section is visible, showing a tree structure with 'Tables', 'Views', 'Stored Procedures', 'Functions', 'Tables', 'Views', and 'World'. The main area contains a SQL editor with the following query:

```
1 * SELECT tag_name, COUNT(*) as tag_count
2 FROM tags
3 JOIN photo_tags ON tags.id = photo_tags.tag_id
4 GROUP BY tag_name
5 ORDER BY tag_count DESC
6 LIMIT 5;
```

Below the SQL editor is a 'Result Grid' table with the following data:

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

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

A. Marketing Analysis:

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

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

Output: – “**Thursday**” is the best day of the week to launch ads. Team can schedule an ad campaign on Thursday because most users registered on Instagram on this date.



MySQL Workbench

Instagram User Analytics Trinity P.x Instagram User Analytics

File Edit View Query Database Server Tools Scripting Help

Navigator

SCHEMAS

ig_clone

Tables

Views

Stored Procedures

Functions

sakila

sys

world

project.sql.trinity

SQL File 2*

SQL File 3*

SQL File 4*

SQL File 5*

SQL File 6*

SQL File 7*

SQL File 8*

SQL File 9*

SQL File 10*

SQL File 11*

SQL File 12*

SQL File 13*

SQL File 14*

Dont List

1 • SELECT
2 DAYNAME(created_at) AS registration_day,
3 COUNT(*) AS user_count
4 FROM users
5 GROUP BY registration_day
6 ORDER BY user_count DESC
7 LIMIT 1,2

Result Grid | Filter Rows | Export | Wrap Cell Contents |

registration_day	user_count
Thursday	15

Administration Schemas Information

No object selected

Object Info Session Output

Result 1.x Ready Only



B. Investor Metrics :

22

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

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

Output: Average number of posts per Instagram user is 3.4730 and The Total Number of Photos on Instagram are 257 and the total number of users are 74

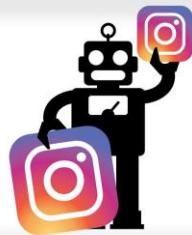
The screenshot shows the MySQL Workbench interface. In the top navigation bar, the title is "Instagram User Analytics Trinity P.x". The main area displays a SQL query:

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

The "Result Grid" pane shows the following data:

total_photos	total_users	ratio_photos_per_user
257	74	3.4730

total_photos	total_users	ratio_photos_per_user
257	74	3.473



B. Investor Metrics :

23

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

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

Output: This user id must be spam and there are 13 unique users (Potential Bots) who has liked every single posts on Instagram

The screenshot shows the MySQL Workbench interface. In the top-left, the 'Schemas' pane lists 'ig_clone' as the active database, containing tables like 'likes', 'photos', and 'users'. The main area displays a SQL query:

```
1 SELECT user_id
2 FROM likes
3 GROUP BY user_id
4 HAVING COUNT(DISTINCT photo_id) = (SELECT COUNT(DISTINCT photo_id) FROM likes);
```

The 'Results Grid' below shows the output of the query, listing 13 user IDs: 5, 14, 21, 24, 36, 41, 54, 57, 66, 71, 75, 76, and 91. The results grid has columns for 'user_id' and 'Count' (which is always 13). The bottom status bar indicates 'Read Only'.

user_id
5
14
21
24
36
41
54
57
66
71
75
76
91



Impact, Summary and Achievements

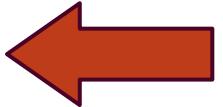
24

- **Impact:** The analysis provided deep insights into user engagement, bolstering my analytical skills and offering practical guidance for future decisions. It showcased the real-world impact of data analysis, aiming to optimize Instagram advertising strategies through techniques like data import, exploratory analysis, SQL querying, and MySQL Workbench utilization.
- **Summary:** In summary, the insights from the analysis not only enhanced my abilities as a data analyst but also positioned me to make meaningful contributions to the team's broader objectives and the success of the Instagram platform.
- **Achievements:** The project brought about significant achievements and turned out to be incredibly valuable. By successfully completing the analysis, I gained valuable insights that greatly improved my understanding of how users behave on Instagram.

Projects Result Summary

- **Description:** Analyzed user behavior on Instagram to optimize marketing strategies.
- **Approach:** Utilized MySQL Workbench for data import, conducted exploratory data analysis, and leveraged SQL queries for insights.
- **Findings:**
 - ✓ Determined the best day for ad campaigns, identified inactive users, and analyzed user engagement and growth.
 - ✓ Identified key insights for Instagram's marketing team, including:
 - ❖ Peak user registration day (Thursday) for scheduling ad campaigns
 - ❖ Top hashtags for partner brand collaborations
 - ❖ Users with no posts for targeted engagement strategies Quantitative Metrics
 - ❖ Determined the winner of the contest with the most liked photo (257 likes)
 - ❖ Identified 13 potential bot accounts that liked every single post





Access Project

26





Project 3: Operation Analytics and Investigating Metric Spike

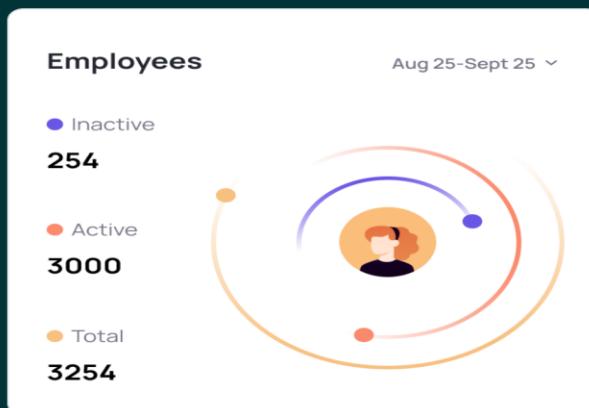
27

By : Raj Rathod
Data Analyst

Introduction:

trainity

Operation Analytics & Investigating metric spike case study





Project Description

28

- **Objective:** Thorough examination of company operations to identify improvement areas.
- **Collaboration:** Close collaboration with operations, support, and marketing teams to extract valuable insights from data.
- **Impact:** Predicts company's growth or decline, leads to enhanced automation and streamlined workflows.
- **Role:** As a Data Analyst Lead at Microsoft:
 - Analyze Job Data to provide insights and address inquiries from various departments.
 - Investigate Metric Spikes to understand fluctuations in engagement or sales performance.
- **Responsibilities:** Work with diverse datasets, collaborate with teams, derive actionable insights for business optimization



Project Approach

29

➤ Data Familiarization:

- Took time to understand the provided data and tables.
- Clarified concepts like job_id, actor_id, and event to grasp key aspects of the data.

➤ SQL Analysis:

- Utilized SQL to delve into the dataset provided by the management team.
- Developed the project using SQL Workbench for database creation and data loading.

➤ Analysis and Insights:

- Loaded data into SQL Workbench and performed analysis to extract actionable insights.
- Aimed to provide information to various teams such as operations, support, and marketing.
- Addressed daily questions like fluctuations in engagement or sales performance by investigating metric spikes.



Tech Stack Used

30

MySQL Workbench v8.0:

- Renowned for its versatility and user-friendly interface.
- Seamlessly integrates data modelling, SQL development, and database administration.
- Features a visual editor for building, editing, and running queries.
- Allows preview of changes before application.
- Includes a database administration suite for auditing databases, configuring servers, and viewing logs.
- Provides Performance Monitoring functionality for optimizing database performance.

Case Study 1 - Job Data Analysis

31

1. Jobs Reviewed Over Time

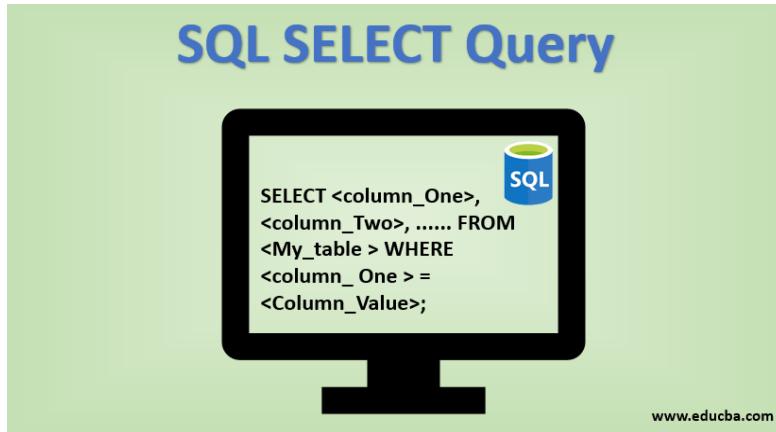
My Task : Calculate the number of jobs reviewed per hour for each day in November 2020?

Output : On date **28th November 2020** there is maximum number of jobs reviewed is **218**. Likewise, **on 29th and 30th of November 2020** Jobs reviewed were **180** and **on 25th Nov 2020** it was **80** and **on 26th Nov 2020** it was **64** and **on 27th Nov 2020** it was **35**.

Dates	Jobs Reviewed per Hour per Day
2020-11-30	180
2020-11-29	180
2020-11-28	218
2020-11-27	35
2020-11-26	64
2020-11-25	80

Job Data Analysis - SQL Queries

32



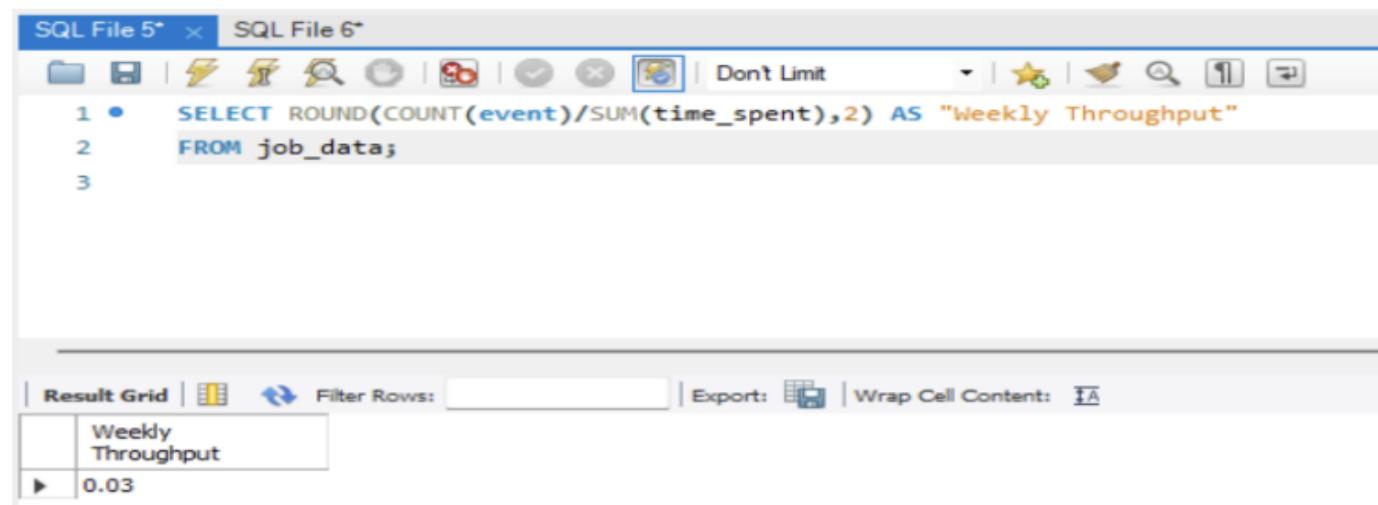
SQL Query: - Jobs Reviewed Over Time

```
SELECT ds AS Dates, ROUND((COUNT(job_id)/SUM(time_Spent))*3600) AS "Jobs
reviewed per Hour per Day"
FROM job_data
WHERE ds BETWEEN '2020-11-01' AND '2020-11-30'
GROUP BY ds;
```

2. Throughput Analysis:

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

Output: The Weekly throughput is 0.03



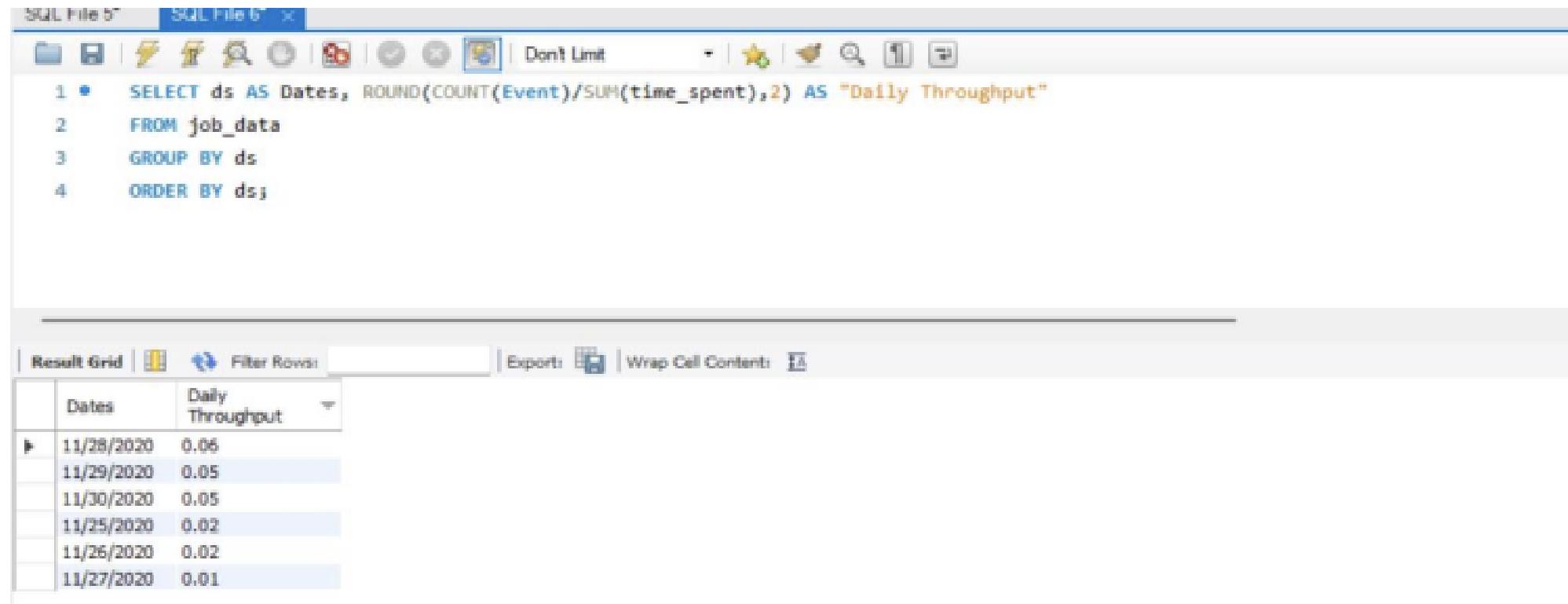
The screenshot shows the MySQL Workbench interface. The top tab bar has two tabs: "SQL File 5*" and "SQL File 6*". Below the tabs is a toolbar with various icons for database management. The main area contains the following SQL code:

```
1 •  SELECT ROUND(COUNT(event)/SUM(time_spent),2) AS "Weekly Throughput"
2   FROM job_data;
3
```

At the bottom, there is a "Result Grid" section showing the output of the query:

Weekly Throughput
0.03

Output: On date 2020-11-28 the throughput is highest 0.06. Metric values fluctuate regularly on both a weekly and daily basis. If you need quicker insights, you can obtain numbers daily or even by the minute. Consequently, rolling metrics excel at illustrating whether your metrics are exhibiting an upward or downward trend on a daily level



The screenshot shows a SQL query editor interface. At the top, there are two tabs: "SQL File 5" and "SQL File 6". Below the tabs is a toolbar with various icons. A dropdown menu labeled "Don't Limit" is open. To the right of the toolbar are several small icons. The main area contains a SQL query:

```
1 •   SELECT ds AS Dates, ROUND(COUNT(Event)/SUM(time_spent),2) AS "Daily Throughput"  
2     FROM job_data  
3    GROUP BY ds  
4   ORDER BY ds;
```

Below the query is a "Result Grid" table. The table has two columns: "Dates" and "Daily Throughput". The data is as follows:

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

3. Language Share Analysis: To Calculate the percentage share of each language in the last 30 days. My Task: - Write an SQL query to calculate the percentage share of each language over the last 30 days.

Output: The **Persian** language has the **highest percentage** share **37.5%**. Italian, Hindi, French, English, and Arabic are having the same percentage share at **12.5%**

The screenshot shows a SQL editor interface with multiple tabs at the top labeled "SQL File 5*", "SQL File 6*", "SQL File 7*", "SQL File 8*", "SQL File 9*", and "SQL File 10*". The "SQL File 10*" tab is active, displaying the following SQL code:

```
1 •  SELECT LANGUAGE,COUNT(LANGUAGE) AS LANGUAGE_COUNT,(COUNT(LANGUAGE)/ (SELECT COUNT(*)  
2 FROM JOB_DATA)) * 100 AS Percentage_Share FROM JOB_DATA GROUP BY LANGUAGE ORDER BY LANGUAGE DESC ;
```

Below the code, there is a "Result Grid" table with three columns: "LANGUAGE", "LANGUAGE_COUNT", and "Percentage_Share". The table contains the following data:

LANGUAGE	LANGUAGE_COUNT	Percentage_Share
Persian	3	37.5000
Italian	1	12.5000
Hindi	1	12.5000
French	1	12.5000
English	1	12.5000
Arabic	1	12.5000



4. Duplicate Rows Detection:

Identify duplicate rows in the data.

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

Output: These rows are having the Duplicate values which is 2 and 3.

The screenshot shows a SQL editor interface with multiple tabs at the top: SQL File 5*, SQL File 6*, SQL File 9*, SQL File 10*, SQL File 11*, SQL File 12* (highlighted in blue), SQL File 14*, and SQL File 15*. Below the tabs is a toolbar with various icons. The main area contains the following SQL code:

```
1 •  SELECT * FROM (SELECT *,ROW_NUMBER() OVER
2   (PARTITION BY JOB_ID) AS DUPLICATE_ROWS FROM JOB_DATA ) A_R
3   WHERE DUPLICATE_ROWS > 1;
```

Below the code is a "Result Grid" section with the following data:

ds	job_id	actor_id	event	language	time_spent	org	DUPLICATE_ROWS
11/28/2020	23	1005	transfer	Persian	22	D	2
11/26/2020	23	1004	skip	Persian	56	A	3



Case Study 2 - Investigating Metric Spike

37

1. Weekly User Engagement: To Measure the activeness of users on a weekly basis

My Task: To calculate the weekly user engagement.

Output:

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

SQL Query:

```
SELECT EXTRACT(WEEK FROM occurred_at) AS "Week  
Numbers", COUNT(DISTINCT user_id) AS "Weekly Active Users"  
FROM events  
WHERE event_type = 'engagement'  
GROUP BY 1;
```

2. User Growth Analysis:

Analyze the growth of users over time for a product

My Task: To calculate the user growth for the product.

Output: Given below is the Output

SQL Query :

```
SELECT Months, Users, ROUND(((Users/LAG(Users, 1) OVER (ORDER BY Mo
nths)-1)*100), 2) AS "Growth in %"
```

FROM

(

```
SELECT EXTRACT(MONTH FROM created_at) AS Months,
COUNT(activated_at) AS Users
```

FROM users

WHERE activated_at NOT IN("")

GROUP BY 1

ORDER BY 1)

) sub;

	Months	Users	Growth in %
▶	1	712	NULL
	2	685	-3.79
	3	765	11.68
	4	907	18.56
	5	993	9.48
	6	1086	9.37
	7	1281	17.96
	8	1347	5.15
	9	330	-75.50
	10	390	18.18
	11	399	2.31
	12	486	21.80

1. Number of Jobs Reviewed

- Analyzing the data on job reviews in November 2020, we found that, on average, there were 83% more distinct jobs reviewed per hour each day.

2. Throughput Analysis

- Utilizing the 7-day rolling average of throughput provides a comprehensive average spanning from day 1 to day 7, unlike the daily metric, which only reflects the average for that specific day.

3. Language Share Analysis

- In the past 30 days, the percentage share of languages is notable:
- Persian holds the highest at 37.5%
- Italian, Hindi, French, English, and Arabic each share an equal portion at 12.5%
- (You can use a simple bar chart or pie chart to visualize these percentages)



4. Duplicate Rows Detection

- Examining the data for duplicates, we found two identical rows when the data is partitioned by job_id. However, considering all columns collectively, each row is unique.

Insights - Case Study 2 - Investigating Metric Spike:

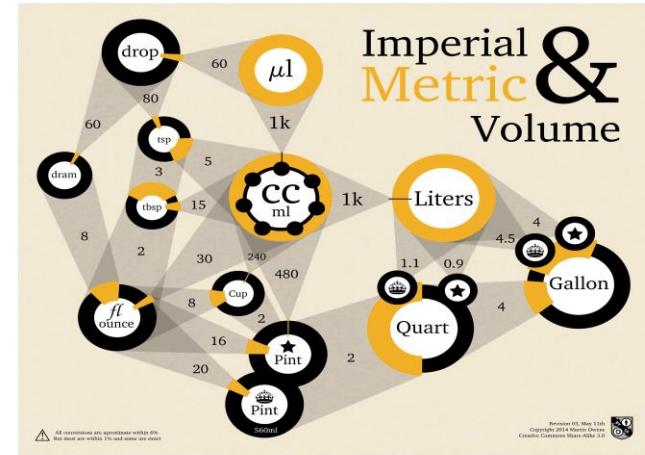
40

1. Weekly User Engagement (Use a bullet point)

- Users increased from week 18 to week 31
- After week 31, user numbers declined
- Potential quality issues with product/service

2. User Growth Analysis (Use a bullet point)

- Steady user base growth over time
- 9381 active users from week 17 (2013) to week 35 (2014) The user base for the product has grown steadily over time. From the 17th week of 2013 to the 35th week of 2014, there were a total of 9381 active users. (You can use a bar chart or a pictorial representation to show the user growth over time)



Project Results

41

Key Learnings:

- Gained proficiency in advanced SQL concepts
- Windows Functions
- Valuable real-world industry insights
- Enhancing SQL mastery
- Developed skills for:
- Posing relevant situational questions
- Determining critical columns for analysis
- Extracting insights for business growth
- Learned to identify and improve operational aspects
- Exploring metric spikes
- Understanding surges and dips



Overall Impact:

- Acquired knowledge in operation analytics
- Ability to investigate and analyze metric trends
- Enhanced data-driven decision-making capabilities

Project 3: Operation Analytics and Investigating Metric Spike

42

Project Result Summary

- **Description:** Conducted analysis to investigate metric spikes and fluctuations in engagement or sales performance.
- **Approach:** Utilized SQL Workbench for data analysis, aimed to provide insights to operations, support, and marketing teams.
- **Insights:**
 - ✓ Uncovered patterns in user engagement, user growth, language share, and duplicate rows detection.
 - ✓ Provided valuable insights for optimizing company operations, including:
 - ✓ Identified peak job review periods (28th November 2020 with 218 reviews)
 - ❖ Analyzed user engagement patterns and growth trends Quantitative Metrics:
 - ❖ Revealed a 83% increase in distinct jobs reviewed per hour each day in November 2020
 - ❖ Uncovered a steady user base growth from 9,381 active users (Week 17, 2013) to Week 35, 2014



Access Project

Thank You



Project 4: Hiring Process Analytics

44

Introduction:

By : Raj Rathod
Data Analyst

trainity

Company Statistics



Project Overview

45



- **Objective:** Analyze the company's hiring process data to draw meaningful insights
- **Focused Areas:** Number of rejections, interviews, job types, vacancies
- **Impact:** Improve the hiring department's processes and strategies

Dataset Overview

- **Description:** Dataset containing records of previous hires
- **Key Variables:** Rejections, interviews, job types, vacancies (and any other relevant variables)

Data Cleaning and Preparation

- Steps taken to clean and prepare the data for analysis
- Handling missing values, removing duplicates, etc.



Tech-Stack Used

Microsoft Excel 2021:

- Essential tool for data cleaning, analysis, and visualization.
- Sorting, filtering, and formula capabilities used for data cleansing and profit calculation.
- Functions employed for extracting top movies and grouping data.
- Charting features utilized for effective data visualization.
- User-friendly interface and diverse functionalities made it the ideal choice for the project.

Microsoft PowerPoint:

- For Presentation Purpose.





Analysis 1: Hiring Analysis

47

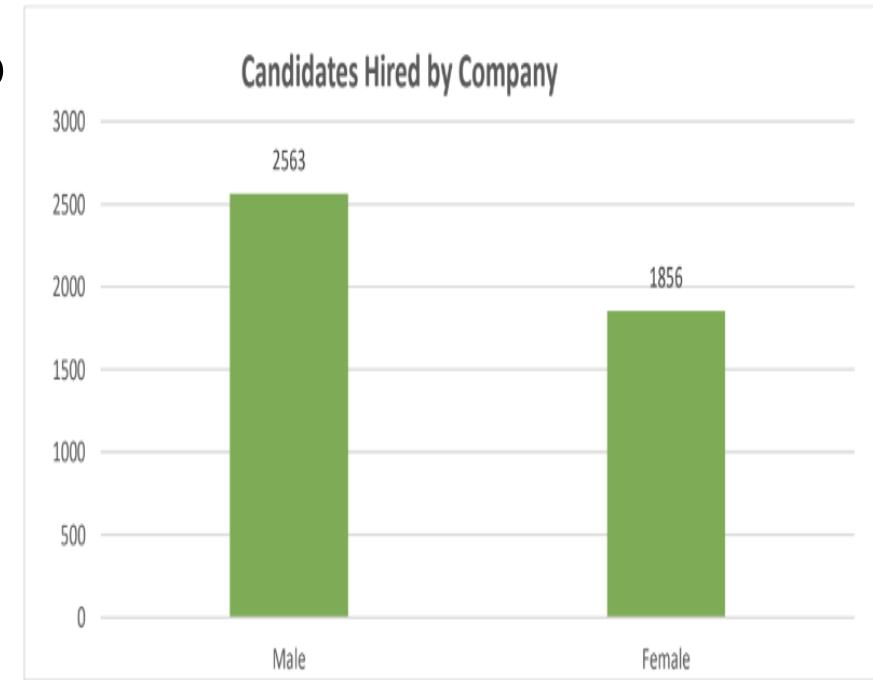
A. Task: Determine the gender distribution of hires. How many males and females have been hired by the company?

Insights: Female candidates should be hired equal to men so Neutral.

Output:

To Calculate the No of Candidates Hired by Company by Gender is,

- For Males: = COUNTIFS(D:D,"Male",C:C,"Hired")
- For Females: = COUNTIFS(D:D,"Female",C:C,"Hired")
- **'2563' Males and '1856' Females were hired by the Company.**
- The analysis revealed a breakdown of male and female hires, providing insights into potential gender imbalances within the organization.





Analysis 2: Salary Analysis

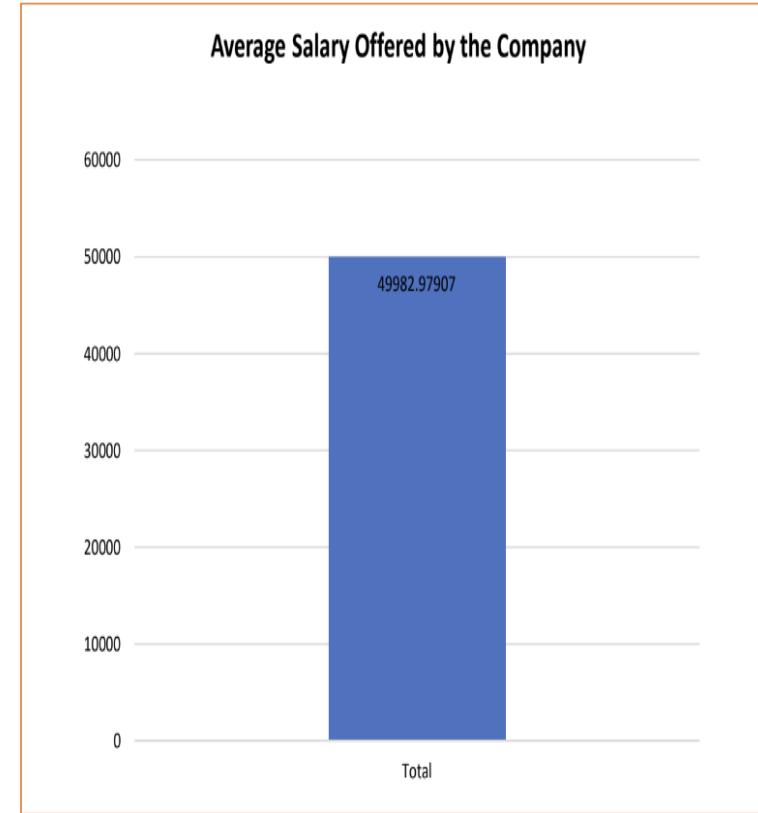
48

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.

My Solution:

- To Calculate the Average Salary Offered by the Company, I will be using AVERAGE function to calculate the averagesalary =AVERAGE(G:G)
- **The Average Offered Salary is “49878.86”.**
- The calculated average salary served as a benchmark for
- understanding the company's overall salary structure.





Analysis 3: Salary Distribution

49

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 bounds of the range.

Your Task: Create class intervals for the salaries in the company's dataset to understand the salary distribution.

My Solution:

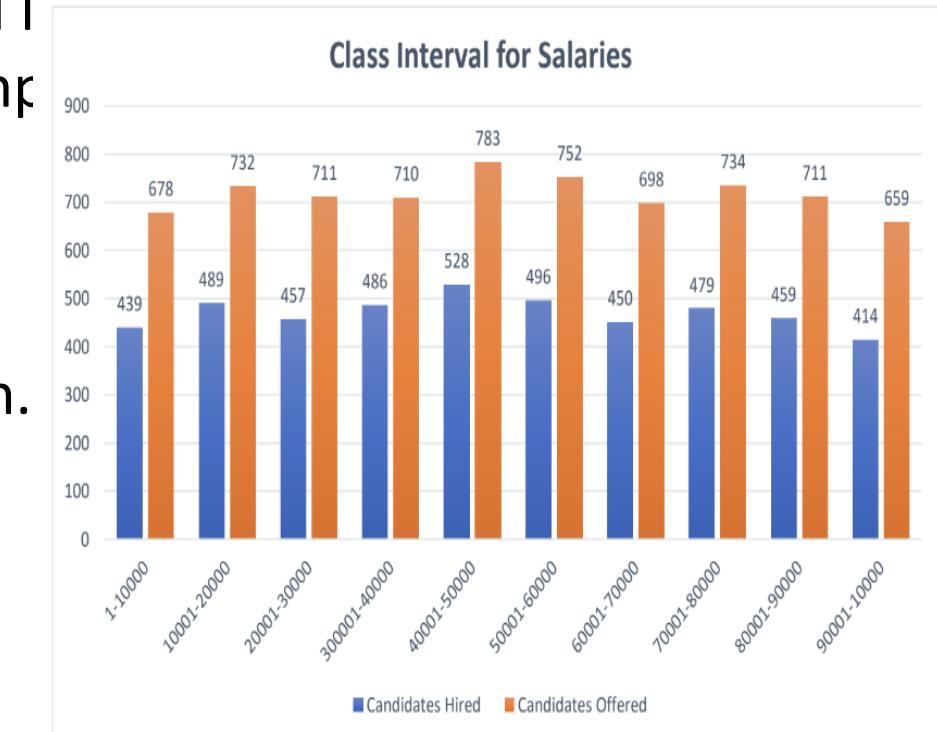
- Utilize Excel's HISTOGRAM function for Class Intervals.
- Histogram provides grouped salary data for visualization.

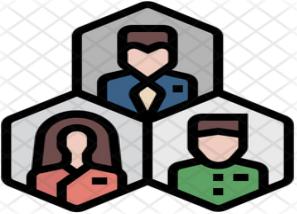
Salary Distribution Analysis:

- Max salary: 40,001–50,000 range.
- Min salary: 9,001–10,000 and 1–10,000 intervals.
- Implies higher demand for mid-level positions.

Candidate Recruitment Insights:

- Hired candidate salaries reflect similar trends.
- Focus on mid-level positions, fewer hires for seniors and entry-level roles.





Analysis 4: Departmental Analysis

50

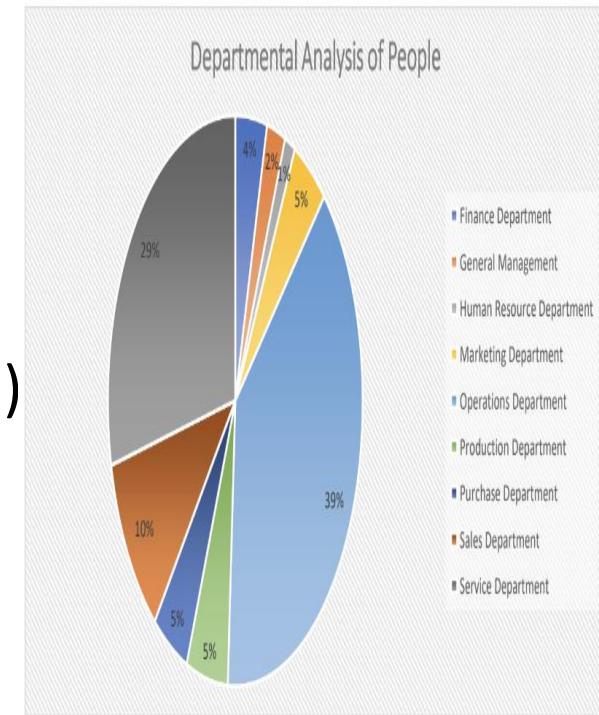
D. Departmental Analysis: Visualizing data through charts and plots is a crucial part of data analysis.

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

My Solution:

To Calculate, COUNTIF function to count the number of employees in each department.
=COUNTIFS(E:E,"Marketing Department",C:C,"Hired")
=COUNTIFS(E:E,"Service Department",C:C,"Hired"), likewise.

- Analyzing the pie chart above reveals that **the highest number of candidates is recruited in the Operations Department**, followed by the Services and Sales Departments.
- Conversely, **the Human Resource Department sees the fewest hires**. These figures potentially signify the team sizes and the significance of each department within the organization.



Analysis 5: Position Tier Analysis

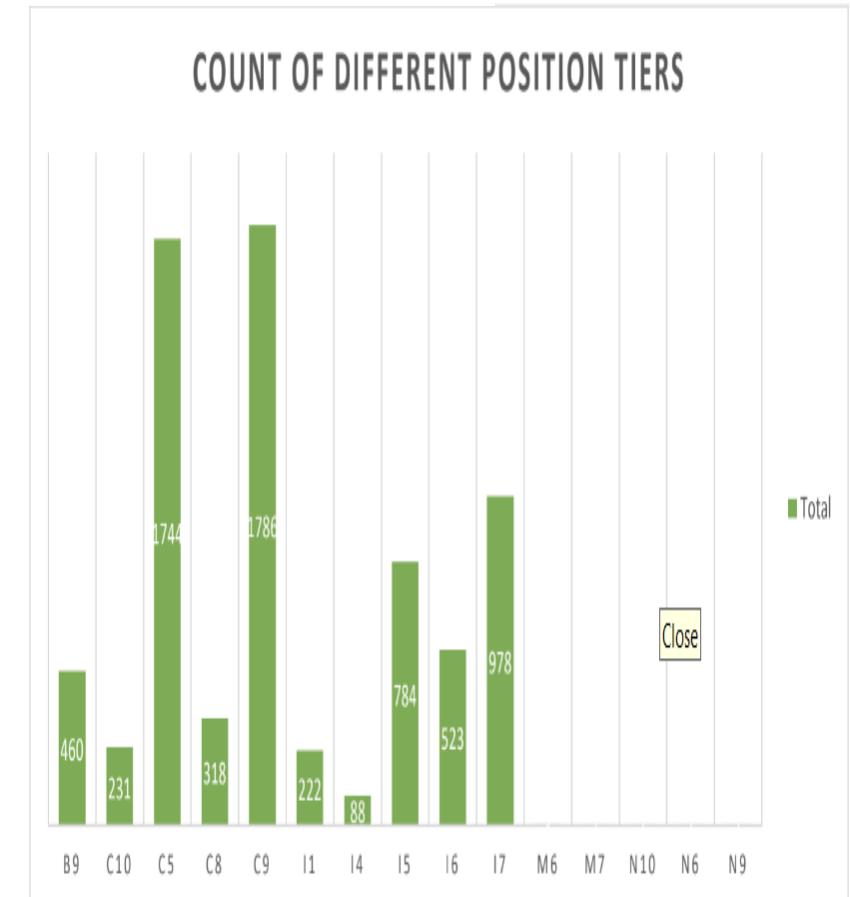
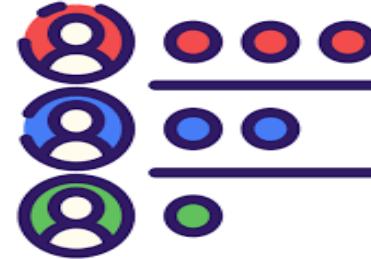
51

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.

My Solution:

- We Can use the Hired and Position Columns Relation and Filter it and Create a Chart by Visualization Option. Visual representations of position tiers offered a clear view of the organizational hierarchy and the distribution of roles
- In this observation, it is evident that the organization predominantly recruited candidates for the position tier c9, with c5 being the subsequent most common tier, followed by i7 at a considerable distance in third place.



➤ **Comprehensive Understanding:**

Project yields deep insights into hiring process analytics.

➤ **Actionable Insights:**

Provides valuable guidance for refining recruitment strategy.

➤ **Data-Driven Approach:**

Ensures informed decision-making.

➤ **Contribution to Efficiency:**

Enhances overall efficiency of the hiring process.

➤ **Appreciation:**

Gratitude for the valuable contribution of the project.





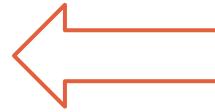
Project 4: Hiring Process Analytics

53

Project Result Summary

- **Description:** Analyzed different position tiers within the company to understand the distribution of roles.
- **Approach:** Used visualization tools to represent position tiers and gain insights into the organizational hierarchy.
- **Findings:** Revealed the distribution of positions across different tiers, providing valuable guidance for recruitment strategy.
 - ✓ Analyzed hiring data to identify potential biases and areas for improvement
 - ✓ Provided recommendations for strategic hiring and resource allocation
 - Quantitative Metrics:**
 - ✓ Revealed a gender imbalance with 2,563 male hires and 1,856 female hires
 - ✓ Calculated the average salary offered by the company as \$49,878.86





Access Project

Thank you!



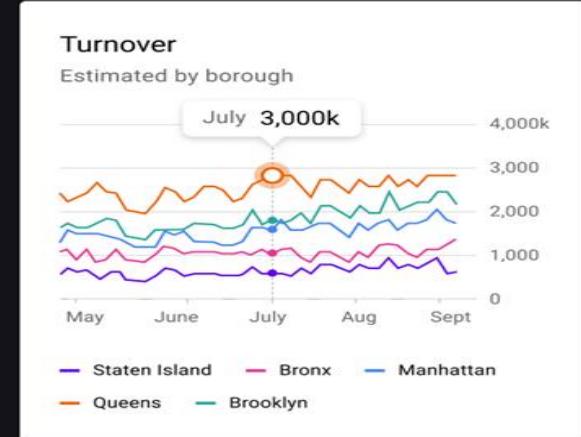
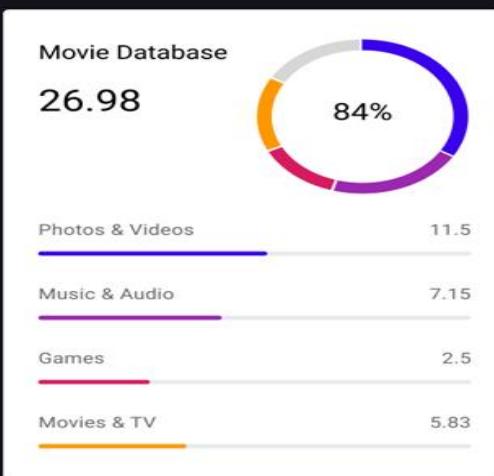
Project 5: IMDb Movie Analysis

55

Introduction:

By : Raj Rathod
Data Analyst

IMDb



Business Overview & Project Description

56

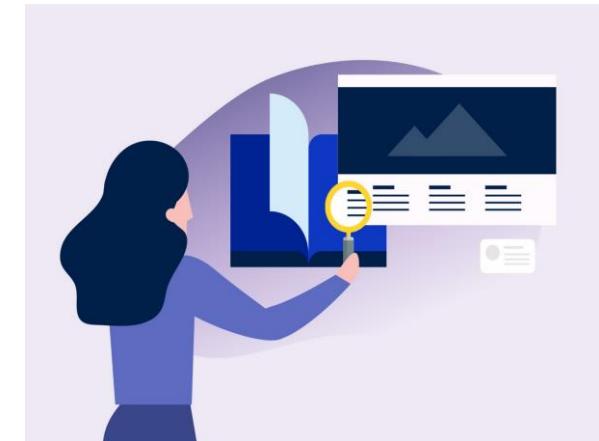


- IMDb, or Internet Movie Database, hosts a vast collection of movies, web-series, and TV shows.
- Users rate content on a scale of 1 to 10 stars, aiding others in assessing quality.
- Project Objective:
 - ✓ Uncover correlations between movie attributes and IMDb ratings.
 - ✓ Provide actionable insights for stakeholders to optimize future projects.
 - ✓ Importance:
 - ✓ Understanding factors influencing IMDb ratings guides strategic decisions in production, marketing, and distribution.
 - ✓ Optimization based on data-driven insights increases success likelihood.

Project Approach & Findings



- **Assessment:** Careful examination of dataset to outline tasks.
- **Data Cleansing:** Excel-based cleaning, removal of null values, and unnecessary columns.
- **Genre Identification:** Popular genres identified through analysis.
- **Movie Duration Analysis :** Calculated the Top Movies w.r.t Time and IMDb Ratings.
- **Language Analysis :** Sorted the Top 5 Language according to No of Movies.
- **Profit Calculation:** Computation of profit for each movie, sorting by profit.
- **Director Analysis:** Grouping data by director to determine top 10 with highest mean IMDb scores.





Microsoft Excel 2021:

- Essential tool for data cleaning, analysis, and visualization.
- Sorting, filtering, and formula capabilities used for data cleansing and profit calculation.
- Functions employed for extracting top movies and grouping data.
- Charting features utilized for effective data visualization.
- User-friendly interface and diverse functionalities made it the ideal choice for the project.



Microsoft PowerPoint:

- For Presentation Purpose.



Data Cleaning



Before Cleaning



Screenshot of Microsoft Excel showing the 'Before Cleaning' state of an IMDB_Movies dataset. The table contains 28 rows of movie information, including director names, critic reviews, duration, gross, genres, and actors. The data is presented in a raw, unstructured format.

	A	B	C	D	E	F	G	
1	director_name	num_critic_for_reviews	duration	actor_2_name	gross	genres	actor_1_name	movie_title
2	James Cameron	723	178	Joel David Moore	760505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar
3	Gore Verbinski	302	169	Orlando Bloom	309404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End
4	Sam Mendes	602	148	Rory Kinnear	200074175	Action Adventure Thriller	Christoph Waltz	Spectre
5	Christopher Nolan	813	164	Christian Bale	448130642	Action Thriller	Tom Hardy	The Dark Knight Rises
6	Andrew Stanton	462	132	Samantha Morton	73058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter
7	Sam Raimi	392	156	James Franco	336530303	Action Adventure Romance	J.K. Simmons	Spider-Man 3
8	Nathan Greno	324	100	Donna Murphy	200807262	Adventure Animation Comedy Family Fantasy Musical Romance	Brad Garrett	Tangled
9	Joss Whedon	635	141	Robert Downey Jr.	458991599	Action Adventure Sci-Fi	Chris Hemsworth	Avengers: Age of Ultron
10	David Yates	375	153	Daniel Radcliffe	301956980	Adventure Family Fantasy Mystery	Alan Rickman	Harry Potter and the Half-Blood Prince
11	Zack Snyder	673	183	Lauren Cohan	330249062	Action Adventure Sci-Fi	Henry Cavill	Batman v Superman: Dawn of Justice
12	Bryan Singer	434	169	Marlon Brando	20069408	Action Adventure Sci-Fi	Kevin Spacey	Superman Returns
13	Marc Forster	403	106	Mathieu Amalric	168368427	Action Adventure	Giancarlo Giannini	Quantum of Solace
14	Gore Verbinski	313	151	Orlando Bloom	423032628	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: Dead Man's Chest
15	Gore Verbinski	450	150	Ruth Wilson	89289910	Action Adventure Western	Johnny Depp	The Lone Ranger
16	Zack Snyder	733	143	Christopher Meloni	291021565	Action Adventure Fantasy Sci-Fi	Henry Cavill	Man of Steel
17	Andrew Adamson	258	150	Pierfrancesco Favino	141614023	Action Adventure Family Fantasy	Peter Dinklage	The Chronicles of Narnia: Prince Caspian
18	Joss Whedon	703	173	Robert Downey Jr.	623279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers
19	Rob Marshall	448	136	Sam Clafin	241063875	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: On Stranger Tides
20	Barry Sonnenfeld	451	106	Michael Stuhlbarg	179020854	Action Adventure Comedy Family Fantasy Sci-Fi	Will Smith	Men in Black 3
21	Peter Jackson	422	146	Adam Brown	25108370	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolation of Smaug
22	Marc Webb	599	153	Andrew Garfield	262030663	Action Adventure Fantasy	Emma Stone	The Amazing Spider-Man
23	Ridley Scott	343	156	William Hurt	105219735	Action Adventure Drama History	Mark Addy	Robin Hood
24	Peter Jackson	509	186	Adam Brown	25835534	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolation of Smaug
25	Chris Weitz	251	113	Eva Green	70083519	Adventure Family Fantasy	Christopher Lee	The Golden Compass
26	Peter Jackson	446	201	Thomas Kretschmann	218051260	Action Adventure Drama Romance	Naomi Watts	King Kong
27	James Cameron	315	194	Kate Winslet	658672302	Drama Romance	Leonardo DiCaprio	Titanic
28	Anthony Russo	516	147	Scarlett Johansson	407197282	Action Adventure Sci-Fi	Robert Downey Jr.	Captain America: Civil War

Pivot
Table/Formulas

Raw Data

Data Cleaning

After Cleaning



Screenshot of Microsoft Excel showing the 'After Cleaning' state of the same IMDB_Movies dataset. The data has been cleaned and organized. The first row is highlighted in yellow, and the second row is also highlighted in yellow, indicating specific items of interest or selection.

	A	B	C	D	E	F	G
1	director_name	num_critic_for_reviews	duration	gross	genres	actor_1_name	movie_title
2	James Cameron	723	178	760505847	Action Adventure Fantasy Sci-Fi	CCH Pounder	Avatar
3	Gore Verbinski	302	169	309404152	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: At World's End
4	Sam Mendes	602	148	200074175	Action Adventure Thriller	Christoph Waltz	Spectre
5	Christopher Nolan	813	164	448130642	Action Thriller	Tom Hardy	The Dark Knight Rises
6	Andrew Stanton	462	132	73058679	Action Adventure Sci-Fi	Daryl Sabara	John Carter
7	Sam Raimi	392	156	336530303	Action Adventure Romance	J.K. Simmons	Spider-Man 3
8	Nathan Greno	324	100	200807262	Adventure Animation Comedy Family Fantasy Musical Romance	Brad Garrett	Tangled
9	Joss Whedon	635	141	458991599	Action Adventure Sci-Fi	Chris Hemsworth	Avengers: Age of Ultron
10	David Yates	375	153	301956980	Adventure Family Fantasy Mystery	Alan Rickman	Harry Potter and the Half-Blood Prince
11	Zack Snyder	673	183	330249062	Action Adventure Sci-Fi	Henry Cavill	Batman v Superman: Dawn of Justice
12	Bryan Singer	434	169	20069408	Action Adventure Sci-Fi	Kevin Spacey	Superman Returns
13	Marc Forster	403	106	168368427	Action Adventure	Giancarlo Giannini	Quantum of Solace
14	Gore Verbinski	313	151	423032628	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: Dead Man's Chest
15	Gore Verbinski	450	150	89289910	Action Adventure Western	Johnny Depp	The Lone Ranger
16	Zack Snyder	733	143	291021565	Action Adventure Fantasy Sci-Fi	Henry Cavill	Man of Steel
17	Andrew Adamson	258	150	141614023	Action Adventure Family Fantasy	Peter Dinklage	The Chronicles of Narnia: Prince Caspian
18	Joss Whedon	703	173	623279547	Action Adventure Sci-Fi	Chris Hemsworth	The Avengers
19	Rob Marshall	448	136	241063875	Action Adventure Fantasy	Johnny Depp	Pirates of the Caribbean: On Stranger Tides
20	Barry Sonnenfeld	451	106	179020854	Action Adventure Comedy Family Fantasy Sci-Fi	Will Smith	Men in Black 3
21	Peter Jackson	422	146	164255108370	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolation of Smaug
22	Marc Webb	599	153	262030663	Action Adventure Fantasy	Emma Stone	The Amazing Spider-Man
23	Ridley Scott	343	156	105219735	Action Adventure Drama History	Mark Addy	Robin Hood
24	Peter Jackson	509	186	25835534	Adventure Fantasy	Aidan Turner	The Hobbit: The Desolation of Smaug
25	Chris Weitz	251	113	70083519	Adventure Family Fantasy	Christopher Lee	The Golden Compass
26	Peter Jackson	446	201	218051260	Action Adventure Drama Romance	Naomi Watts	King Kong
27	James Cameron	315	194	658672302	Drama Romance	Leonardo DiCaprio	Titanic
28	Anthony Russo	516	147	407197282	Action Adventure Sci-Fi	Robert Downey Jr.	Captain America: Civil War



Task 1. IMDb Movie Genre Analysis

60

- **Insight:** The Most Common Genres of Movies in the IMDb is ‘Drama’ followed by ‘Comedy’ and then ‘Thriller’. Also, The Descriptive Analysis Of Various Genres are as below

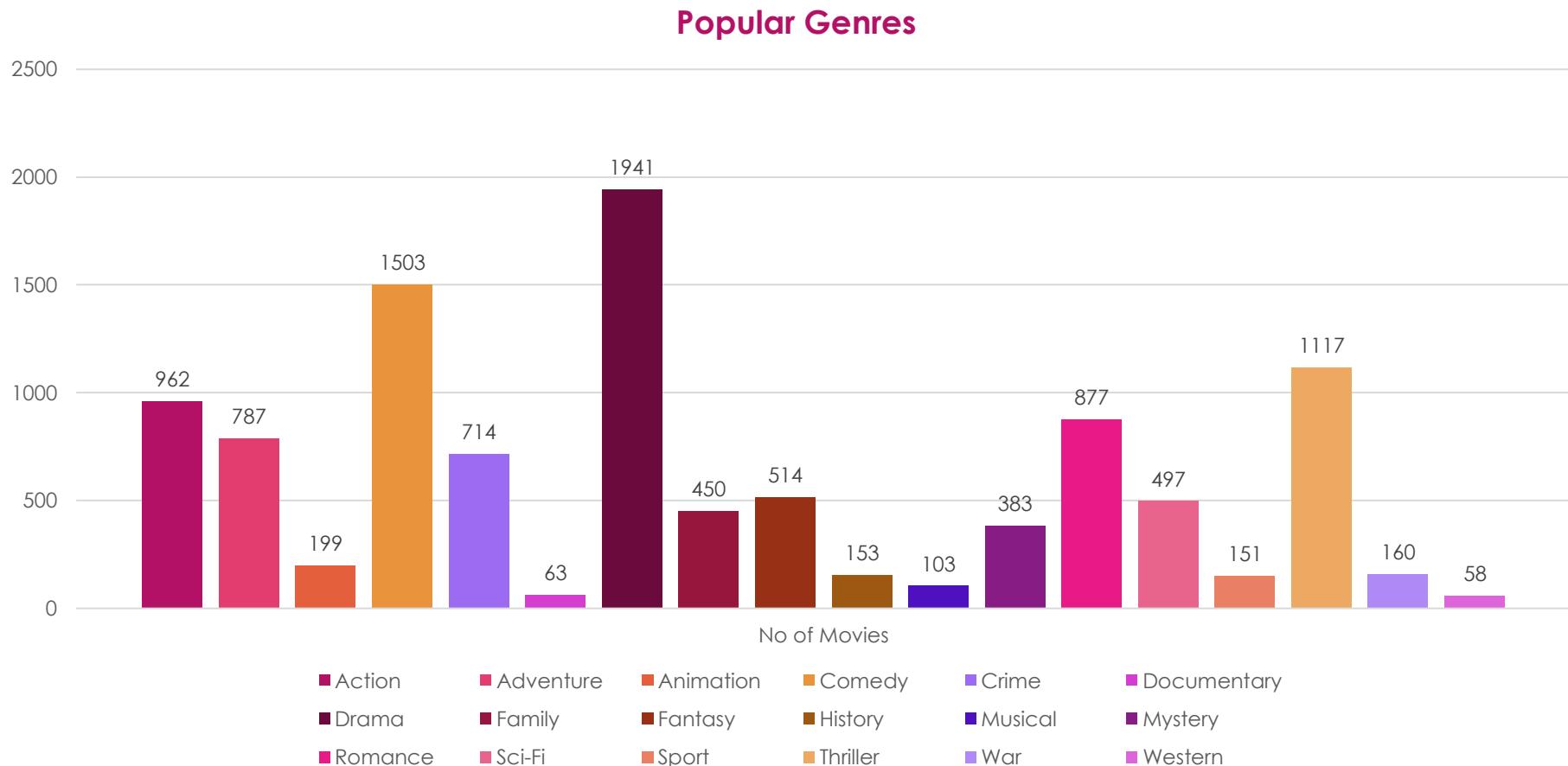


Task 1. Visualization and Descriptive Analysis

61

Descriptive Analysis for Various Genres						
Mean_IMDb	Mode_IMDb	Median_IMDb	Max_IMDb	Min_IMDb	StdDev_IMDb	Var_IMDb
590.67	129.00	473.50	1941.00	58.00	531.79	282805.65

Genre	No of Movies
Drama	1941
Comedy	1503
Thriller	1117
Action	962
Romance	877
Adventure	787
Crime	714
Fantasy	514
Sci-Fi	497
Family	450
Mystery	383
Animation	199
War	160
History	153
Sport	151
Musical	103
Documentary	63
Western	58



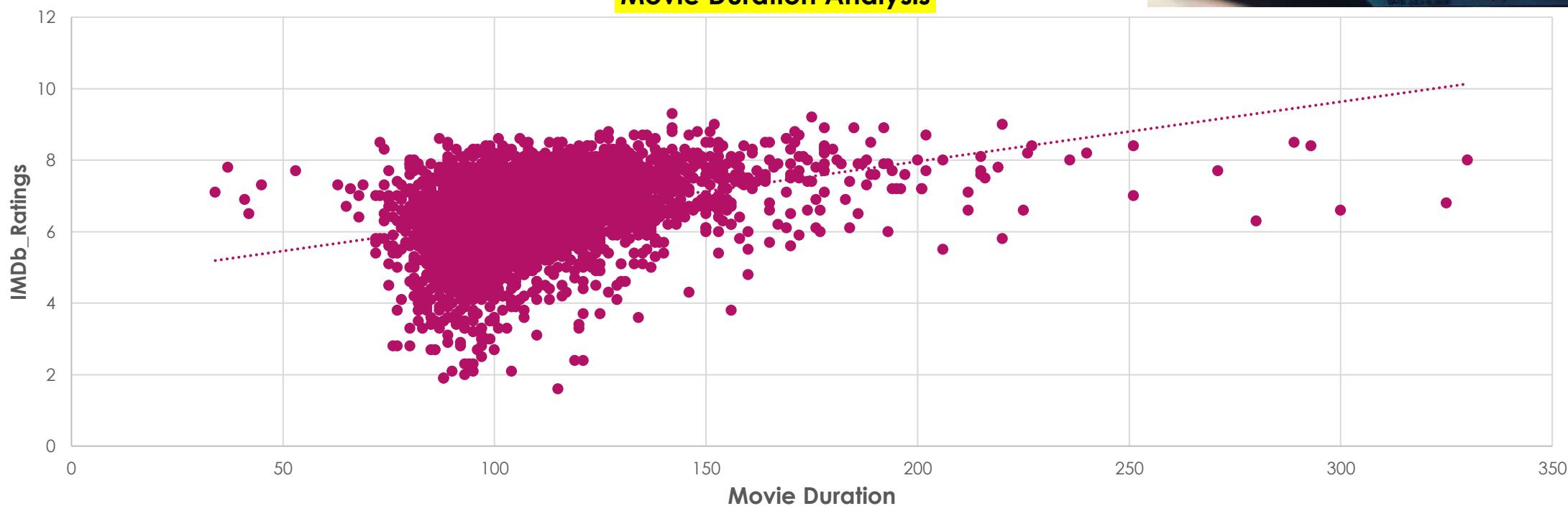
Task 2. Movie Duration Analysis

62

- Answer: The Average Movie in the Dataset is of 110 Minutes.

Average	Median	Standard Deviation
110	106	22.76

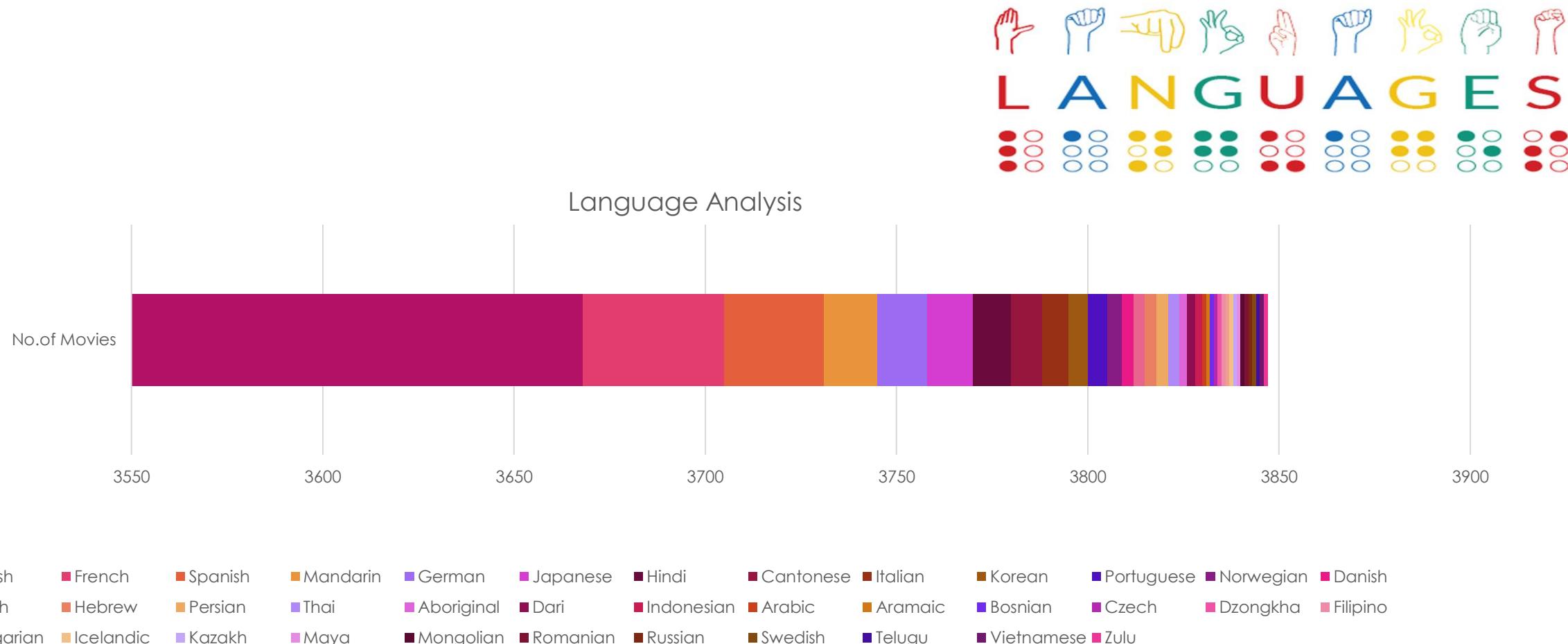
Movie Duration Analysis



Task 3. IMDb Language Analysis

63

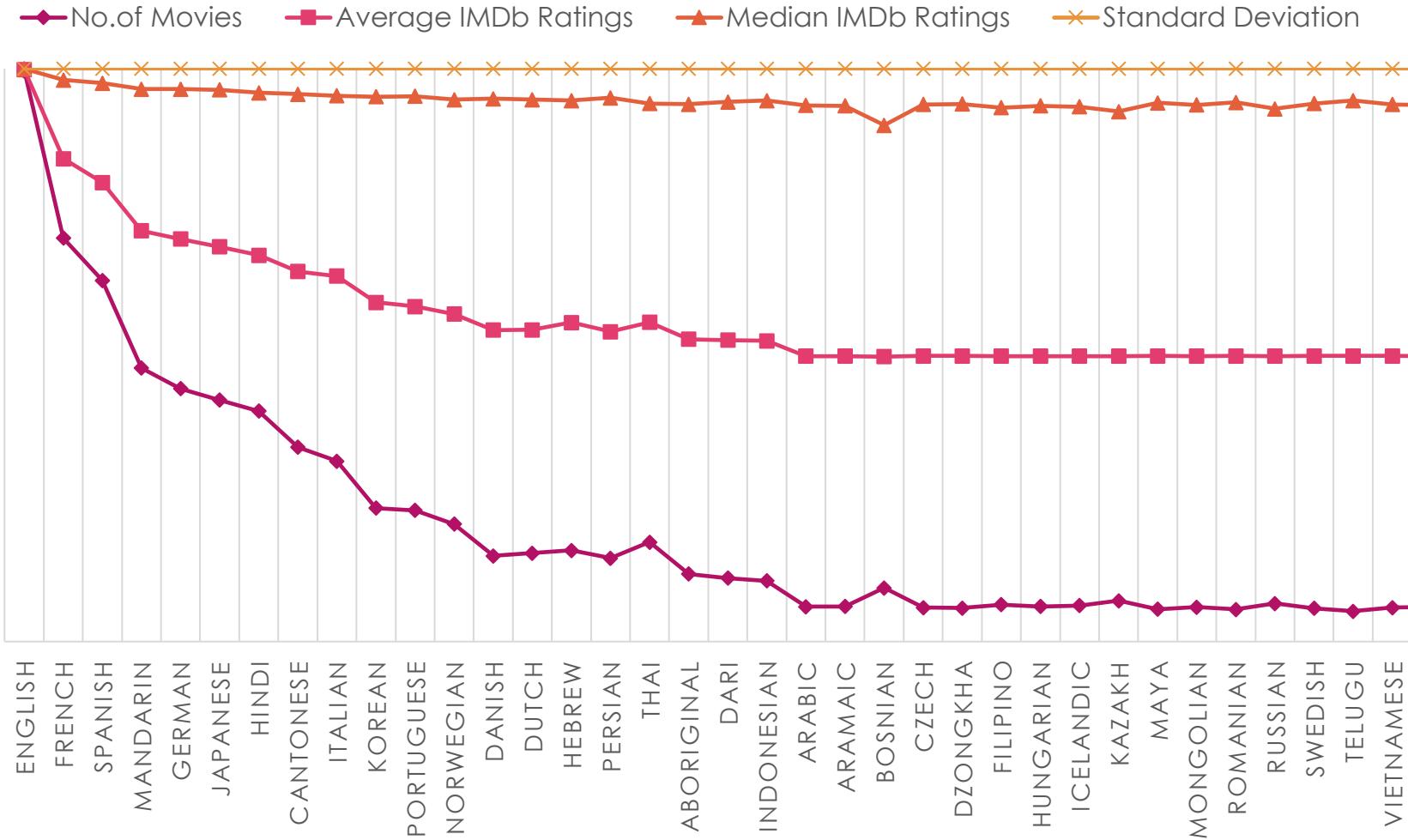
Insight – The Most Common Language used in Movies is English Language followed by French, Spanish, Mandarin.



Task 3. Visualization and Descriptive Analysis

64

DESCRIPTIVE LANGUAGE ANALYSIS

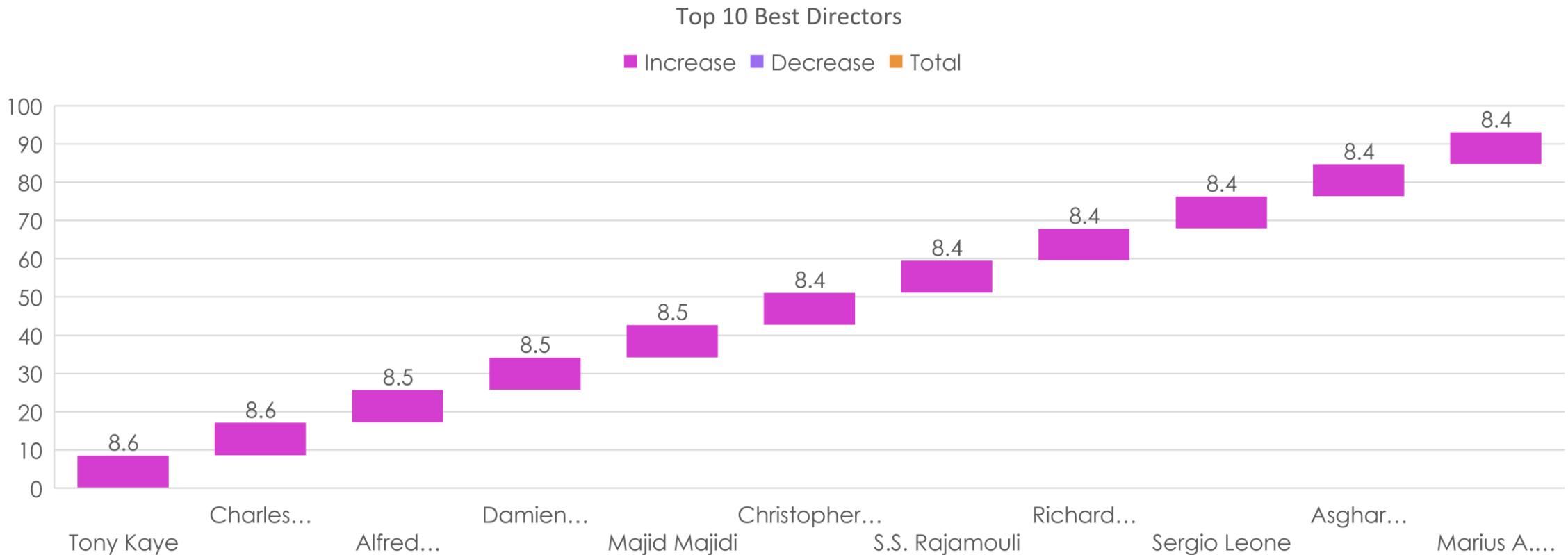




Task 4. IMDb Director Analysis:

65

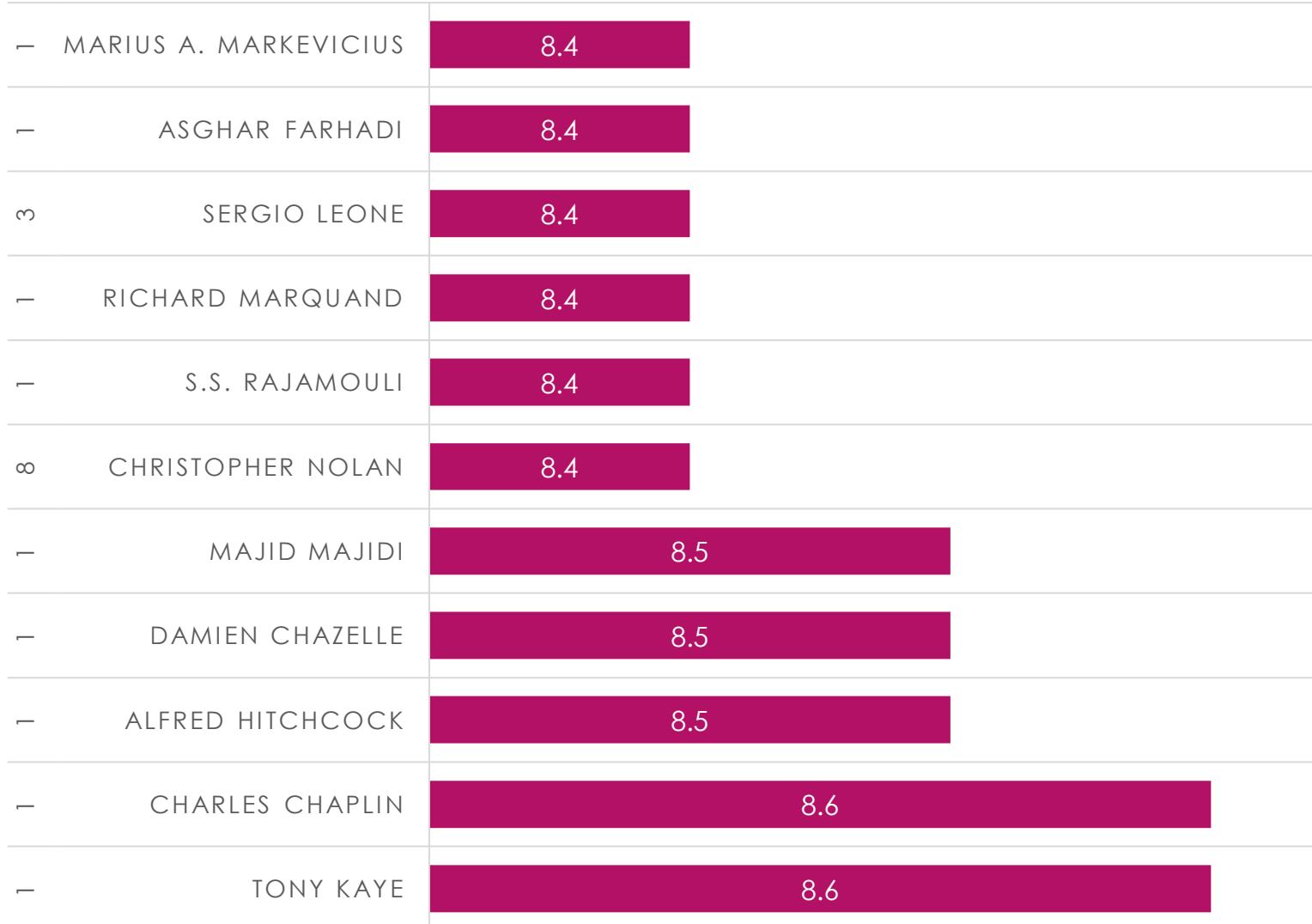
Insight – Top Directors based on IMDb Score is Tony Kayne with Mean Average Rating of 8.6 and Percentile of 8.6 amongst all Top Directors and Percentile Rank is 1.



Task 4. Visualization and Analysis

66

TOP 10 BEST DIRECTORS - IMDB RATINGS



VectorStock®

VectorStock.com/26751775

No of Movies	Top 10 Directors Name	Average IMDB Ratings	Percentile	Percentile Rank
1	Tony Kaye	8.6	8.6	1
1	Charles Chaplin	8.6	8.51	
1	Alfred Hitchcock	8.5	8.5	
1	Damien Chazelle	8.5	8.5	
1	Majid Majidi	8.5	8.44	
8	Christopher Nolan	8.4	8.4	
1	S.S. Rajamouli	8.4	8.4	
1	Richard Marquand	8.4	8.4	
3	Sergio Leone	8.4	8.4	
1	Asghar Farhadi	8.4	8.4	
1	Marius A. Markevicius	8.4	8.4	



Task 5. IMDb Movie Budget Analysis

67

The Correlation Coefficient is 0.100 which is Positive but If it could have near to 1,it could have been strong Possibility of High Budget Movies earns the Most.

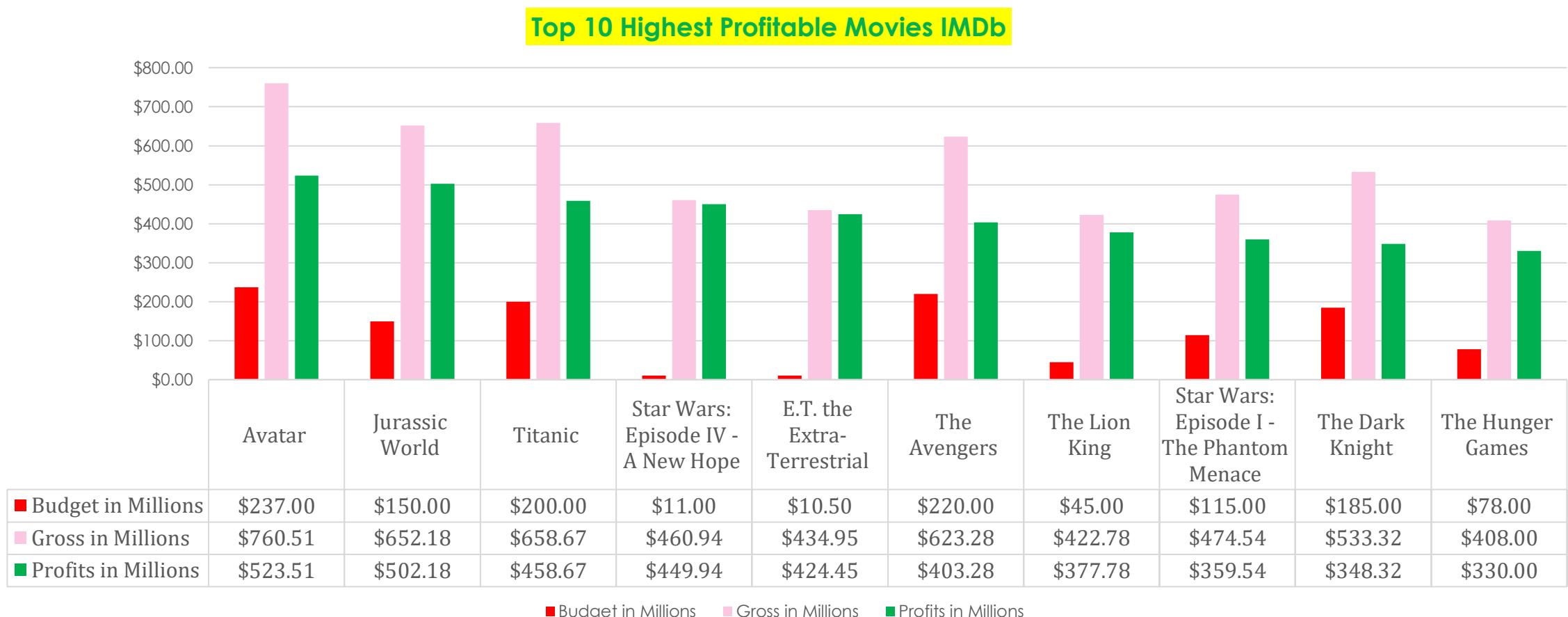
Correlation Coefficient
0.100823531
Maximum Gross Profitable Movie
\$523.51
Million USD



Task 5. Visualization and Analysis

68

Answer: The Top Highest Grossing Movies of IMDb is Avatar and Jurassic World both has Earned Profit over 500Million USD+





Result and Conclusion

69



- Enhanced Understanding of Data Analysis and Visualization Techniques
- ✓ Utilized Excel for data analysis, visualization, and organization
- ✓ Developed skills in cleaning data and identifying valuable insights

- Practical Experience in Real-World Data Analysis
- ✓ Analyzed real-world data to identify profitable movies
- ✓ Recognized top directors and explored popular genres

- Effective Presentation Skills
- ✓ Presented findings through charts and tables
- ✓ Improved ability to communicate insights succinctly





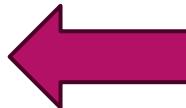
Project 5: IMDb Movie Analysis

70

Project Result Summary

- **Description:** Analyzed IMDb movie data to uncover correlations between movie attributes and ratings.
- **Approach:** Employed Microsoft Excel for data cleaning, genre identification, profit calculation, and director analysis.
- **Key Insights:**
 - ✓ Identified popular genres, analyzed movie duration, language distribution, and top directors based on IMDb scores.
 - ✓ Identified key factors influencing IMDb ratings for movie productions
 - ✓ Provided insights for optimizing future projects based on genre popularity and director impact Quantitative Metrics:
 - ✓ Determined the top 5 most popular genres (Drama, Comedy, Thriller, Action, Romance)
 - ✓ Revealed the top director (Tony Kaye) with an average IMDb rating of 8.6





Access Excel Dataset

71

By : Raj Rathod
Data Analyst

THANK YOU

A decorative banner featuring the words "THANK YOU" in large, bold, white letters on blue rectangular tags. The tags are suspended by thin grey vertical lines from a horizontal wire.

Watch Video Presentation





Project 6: Bank Loan Case Study

72

Introduction:

By : Raj Rathod
Data Analyst





Business Overview

73

- Discern patterns indicating installment payment challenges.
- Inform strategic decisions:
 - Loan denial
 - Adjusting loan amounts
 - Offering higher interest rates to high-risk applicants.
 - Unravel critical factors contributing to loan default.
 - Utilize Exploratory Data Analysis (EDA) to:
 - Understand customer characteristics and loan features' impact on default likelihood.
 - Enhance risk assessment through improved understanding.



Project Description

74

Data analyst at finance company focusing on urban lending.

- **Challenge:** Default risk due to insufficient credit history.
- **Task:** Use EDA to prevent capable applicants from rejection.

- **Risks:**
 - ✓ Rejecting capable applicants leads to lost business.
 - ✓ Approving incapable applicants results in financial losses.

- **Dataset Overview:**
 - ✓ Contains loan application info.
 - ✓ Scenarios:
 - ✓ Payment difficulties: Late payment > X days on first Y installments.
 - ✓ Other cases: Payments made on time.

- **Loan Application Outcomes:**
 - ✓ Approved, Cancelled, Refused, Unused Offer.

Tech Stack Used

75

- **Microsoft Excel 2021:**
Utilized for data cleaning, outlier detection, and conducting univariate and bivariate analysis using pivot tables and charts.

- **Microsoft PowerPoint:**
Employed for presentation purposes, summarizing key findings and insights derived from the bank loan case study analysis project.



Project Approach

76

- This case study involves two extensive datasets: the ‘application_data’ and the ‘previous_application’.
- These datasets contained numerous unnecessary columns and blank entries, which were deemed irrelevant for risk assessments, prompting an initial cleaning process.
- To assess this vast amount of data effectively, the first step involved cleaning the datasets, identifying and removing outliers.
- Subsequently, univariate and bivariate analysis was conducted using pivot tables and charts to delve deeper into the data and derive meaningful insights.



VectorStock®

VectorStock.com/39007697

Data Cleaning

77

- Deleted the Mentioned Columns for further Analysis as a part of Data Cleaning Process as it was not useful for my analysis.



Data Cleaning

Column1	Column2	Column3	Column4
FLAG_WORK_PHONE	FLAG_DOCUMENT_6	FLAG_DOCUMENT_13	FLAG_DOCUMENT_20
FLAG_COUNT_MOBILE	FLAG_DOCUMENT_7	FLAG_DOCUMENT_14	FLAG_DOCUMENT_21
FLAG_PHONE	FLAG_DOCUMENT_8	FLAG_DOCUMENT_15	EXT_SOURCE_2
FLAG_DOCUMENT_2-21	FLAG_DOCUMENT_9	FLAG_DOCUMENT_16	EXT_SOURCE_3
FLAG_DOCUMENT_3	FLAG_DOCUMENT_10	FLAG_DOCUMENT_17	
FLAG_DOCUMENT_4	FLAG_DOCUMENT_11	FLAG_DOCUMENT_18	
FLAG_DOCUMENT_5	FLAG_DOCUMENT_12	FLAG_DOCUMENT_19	

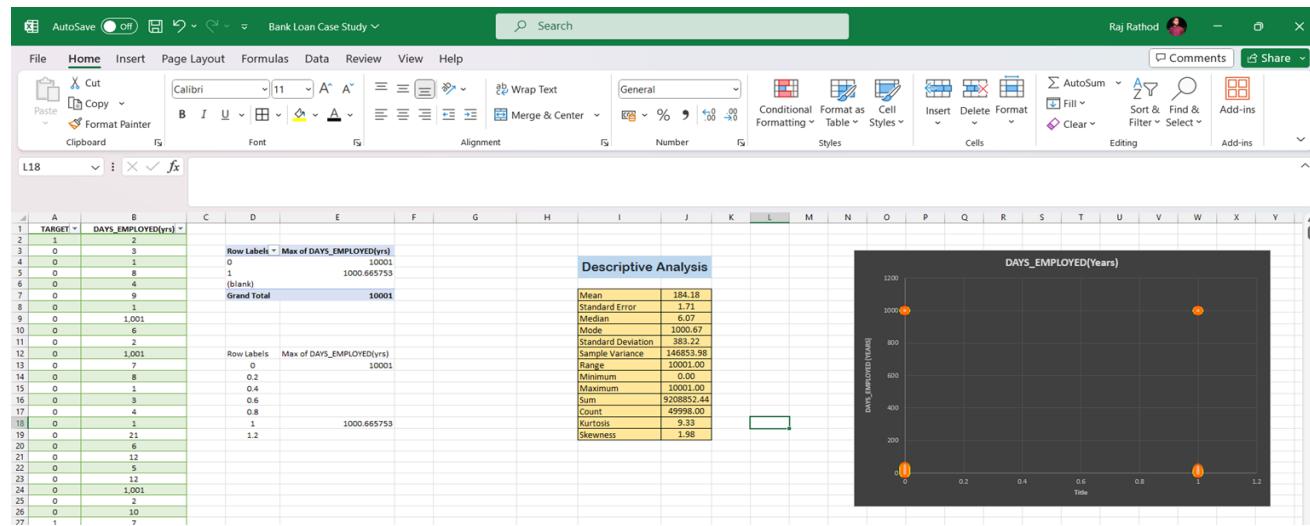
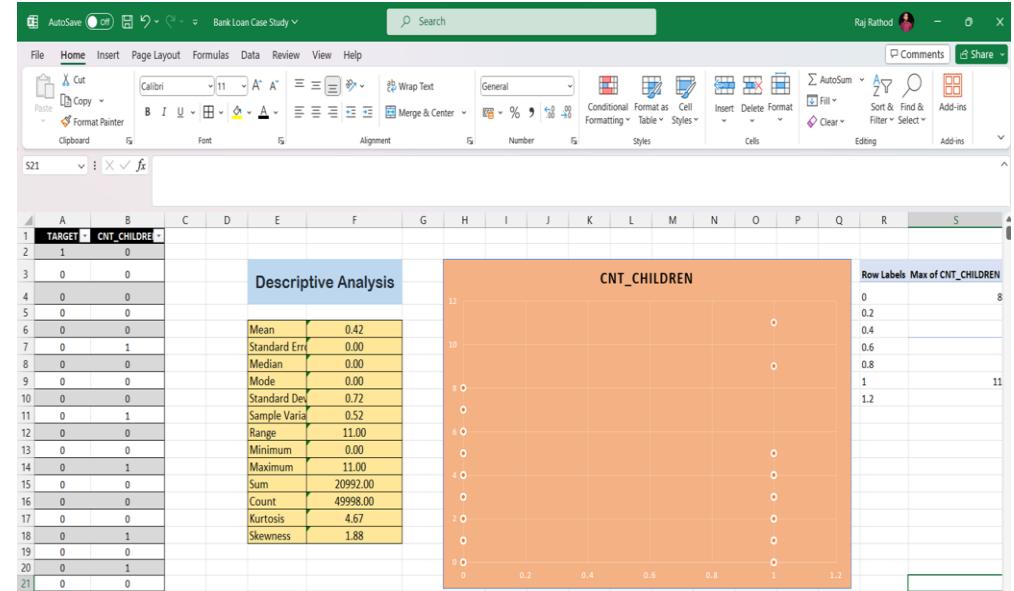
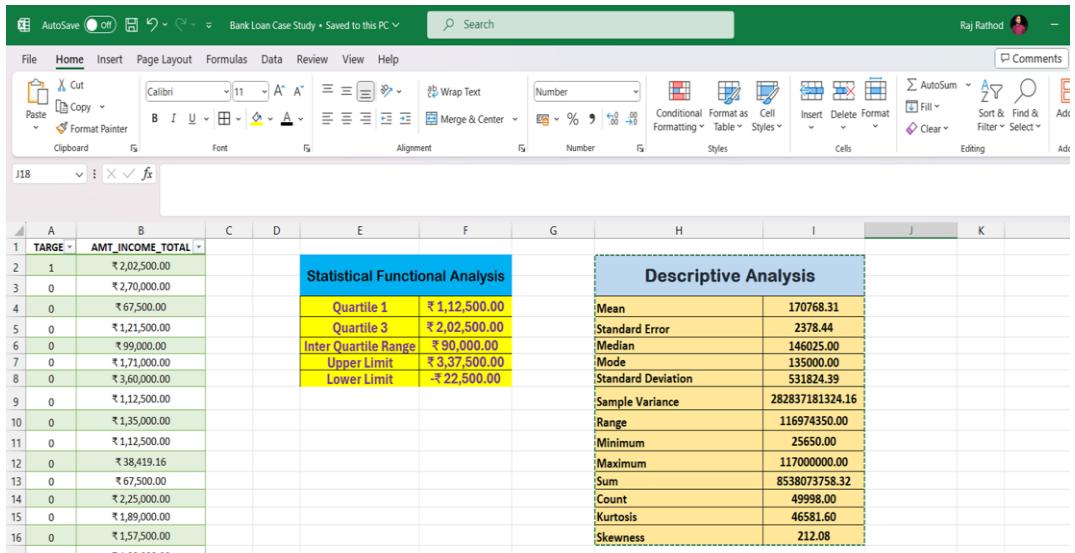
Task 1 : To Identify Missing Data & Deal with it Appropriately

78



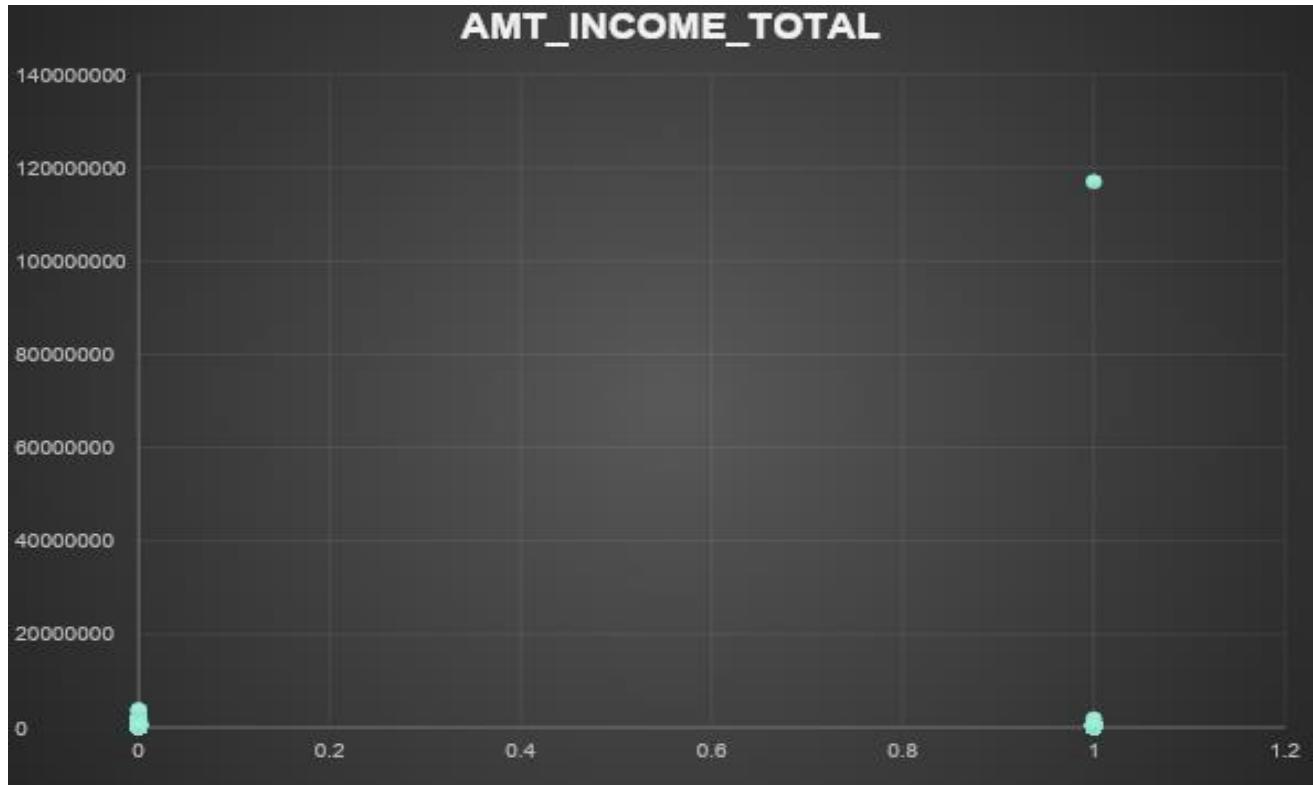
- Initially, I utilized the **COUNTA** function to determine the total number of rows in each column/variable. Then, I calculated the percentage of null values in each column using the formula: **(Total rows count for each column / Total rows count) * 100**.
- Subsequently, columns with a null value percentage exceeding 30% were eliminated, while those with less than 30% underwent distribution statistics analysis, including mean and mode, to address missing values.
- I retained only pertinent variables to extract meaningful insights and standardized the time units by converting days to years, dividing by 365.
- Furthermore, outliers were identified utilizing the interquartile range method, focusing on relevant variables.

Task 2. Identify Outliers in the Dataset



Task 2. (a) Outliers for AMT_INCOME_TOTAL

80



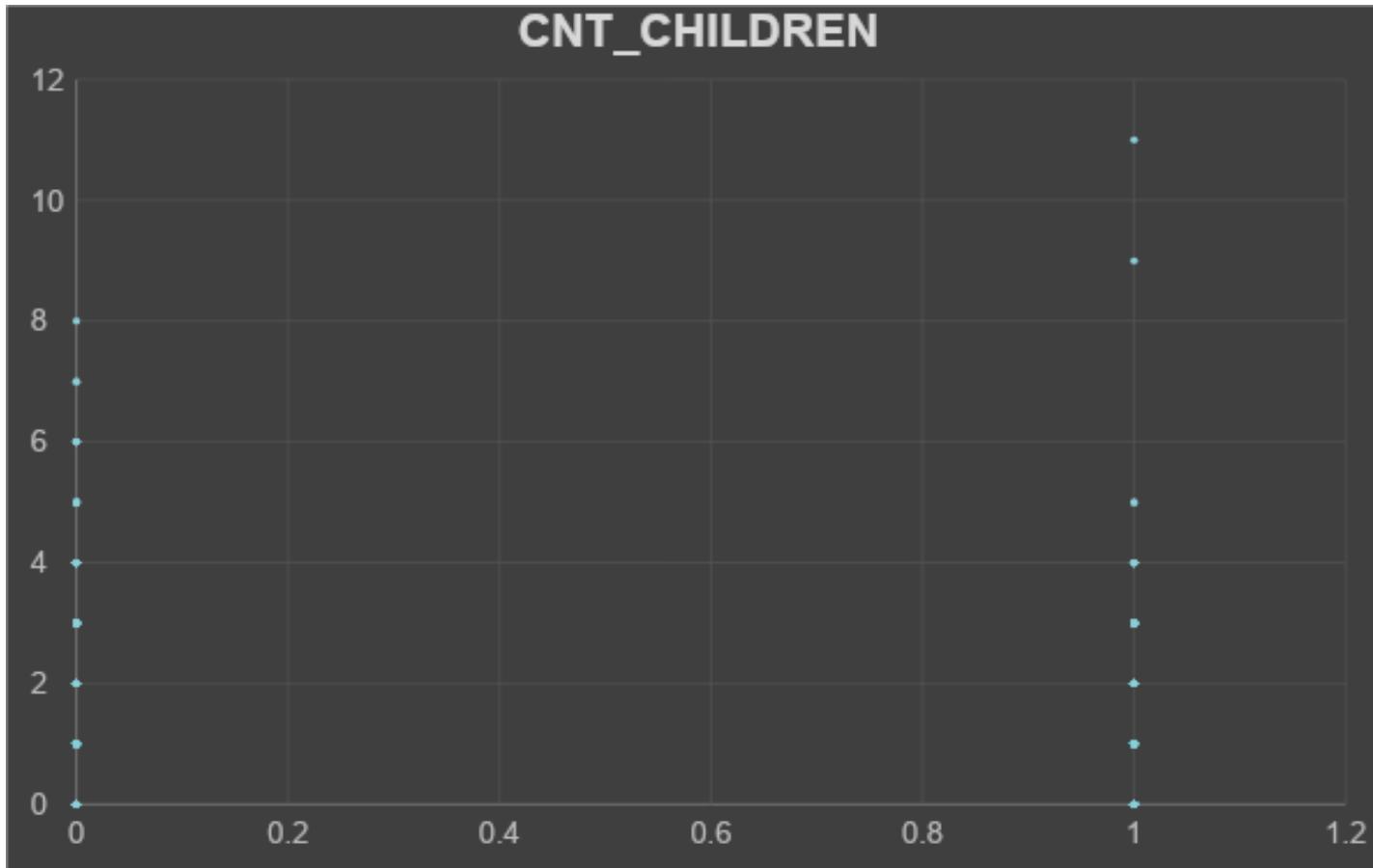
Descriptive Analysis	
Mean	170768.31
Standard Error	2378.44
Median	146025.00
Mode	135000.00
Standard Deviation	531824.39
Sample Variance	282837181324 .16
Range	116974350.00
Minimum	25650.00
Maximum	117000000.00
Sum	8538073758.3 2
Count	49998.00
Kurtosis	46581.60
Skewness	212.08

Statistical Functional Analysis	
Quartile 1	₹ 1,12,500.00
Quartile 3	₹ 2,02,500.00
Inter Quartile Range	₹ 90,000.00
Upper Limit	₹ 3,37,500.00
Lower Limit	-₹ 22,500.00

In this XY plotter, it's evident that for the target variable 1, there's a small number of applicants with an income of 11 crores, while the majority have incomes in the lakhs range.

Task 2. (b) Outliers for CNT_CHILDREN

81



Descriptive Analysis	
Mean	0.42
Standard Error	0.00
Median	0.00
Mode	0.00
Standard Deviation	0.72
Sample Variance	0.52
Range	11.00
Minimum	0.00
Maximum	11.00
Sum	20992.00
Count	49998.00
Kurtosis	4.67
Skewness	1.88

Row Labels	Max of CNT_CHILDREN
0	8
0.2	
0.4	
0.6	
0.8	
1	11
1.2	

In this XY plotter, it's evident that among the target variable 0 applicants, the maximum number of children observed is 8, a rarity in contemporary contexts. Conversely, for target 1 applicants, the maximum number of children observed is 11.

Task 2. (c) Outliers for DAYS_EMPLOYED(Years)

82



Descriptive Analysis	
Mean	184.18
Standard Error	1.71
Median	6.07
Mode	1000.67
Standard Deviation	383.22
Sample Variance	146853.98
Range	10001.00
Minimum	0.00
Maximum	10001.00
Sum	9208852.44
Count	49998.00
Kurtosis	9.33
Skewness	1.98

In this XY plotter, it's evident that a small number of applicants in both target groups 0 and 1 are shown as being employed for 1000 years, which is clearly impossible. Conversely, the majority of applicants appear to have employment durations ranging from around 80 to 90 years.

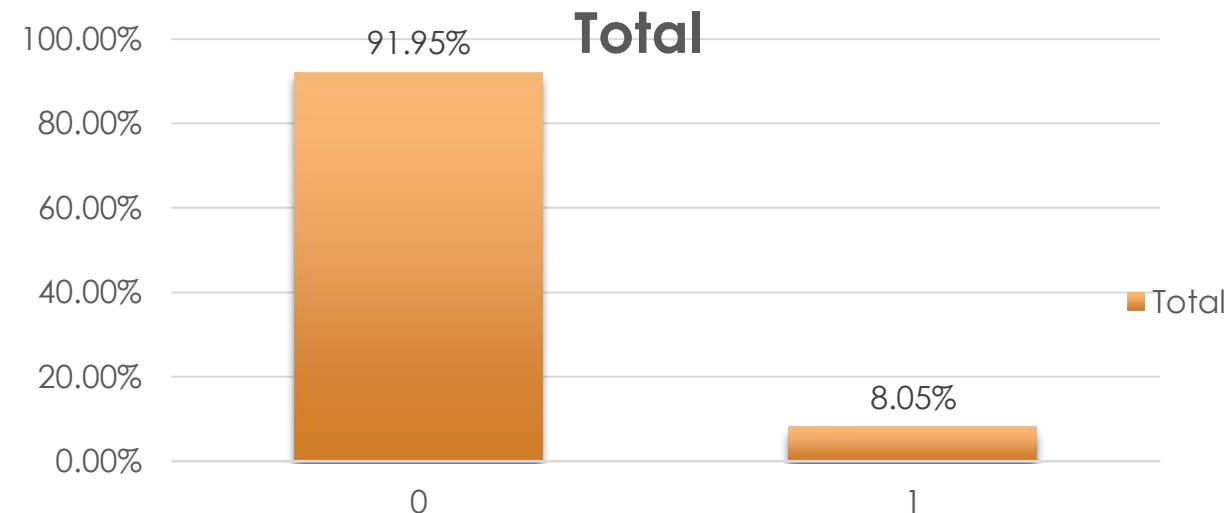
Row Labels	Max of DAYS_EMPLOYED(yrs)
0	10001
0.2	
0.4	
0.6	
0.8	
1	1000.665753
1.2	



Task 3. Analyse Data Imbalanced

83

- In the Excel file provided, the "Data Imbalanced" sheet illustrates the ratio between two categories: applicants facing payment difficulties (Target 1) and those making payments on time (Target 0), with a ratio of 11:41.
- Out of the total applicant pool of 49999, 91.95% successfully make payments on time, constituting the majority class. Conversely, the remaining 8.05% experience payment difficulties, forming the minority class.

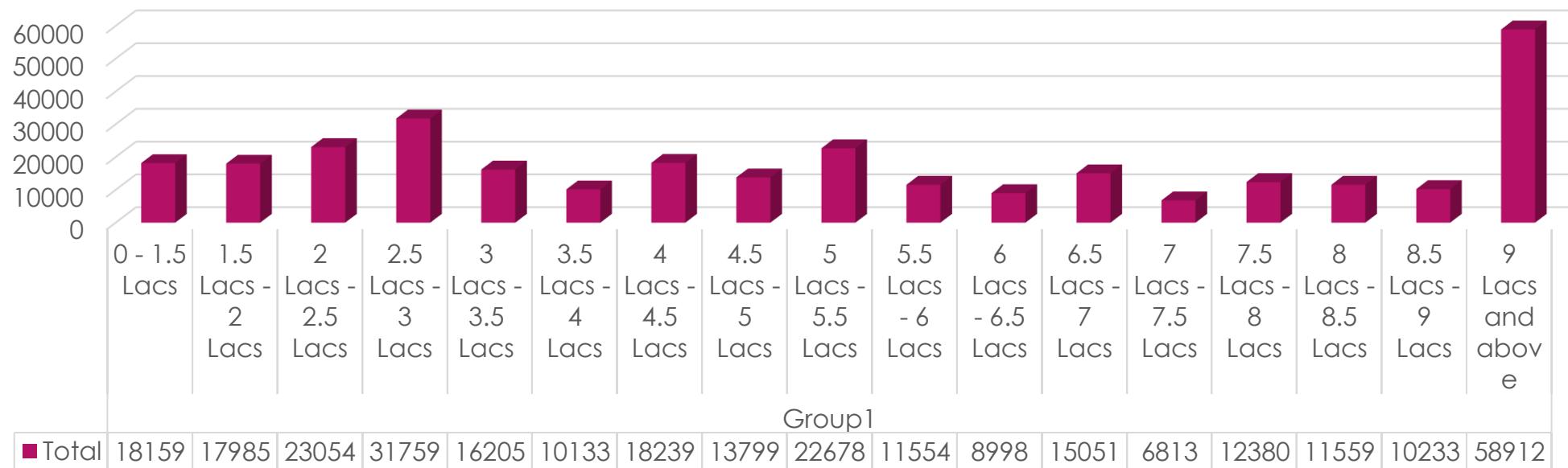


Task 4. To Perform : (a) Univariate Analysis

84

- Univariate Analysis involves examining data that consists of a single variable. It focuses on describing the data and identifying existing patterns, rather than exploring causes or relationships.
- The graph presented here illustrates univariate analysis by displaying the count of loan applicants (0 & 1) for various income brackets within the "AMT_CREDIT" column. The majority of applicants received loan approvals within the credit range of 9 lakhs and above.

APPLICANTS PER CREDIT RANGE

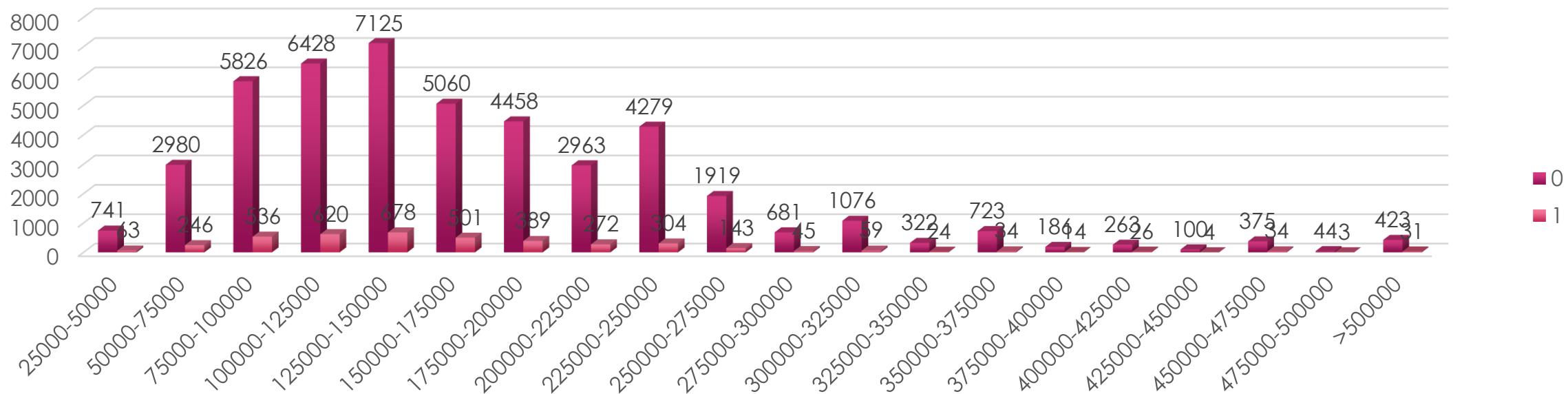


Task 4. To Perform: (b) Segmented Variate Analysis

85

- Segmented Univariate analysis involves examining data containing only one variable. In this context, segmented analysis refers to dissecting a variable into subsets for analysis.
- The graph above illustrates segmented analysis, revealing that the majority of applicants (0 & 1) earn between 1 lakh and 2.25 lakhs. Notably, very few target 1 applicants earn 5 lakhs or more, which could contribute to payment difficulties.

Target Applicants_Income Bins

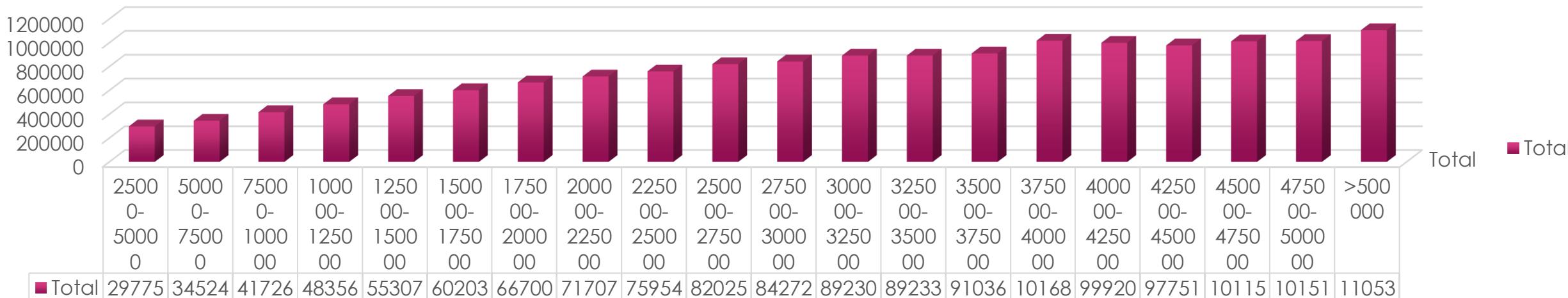


Task 4. To Perform: (c) Bi-Variate Analysis

86

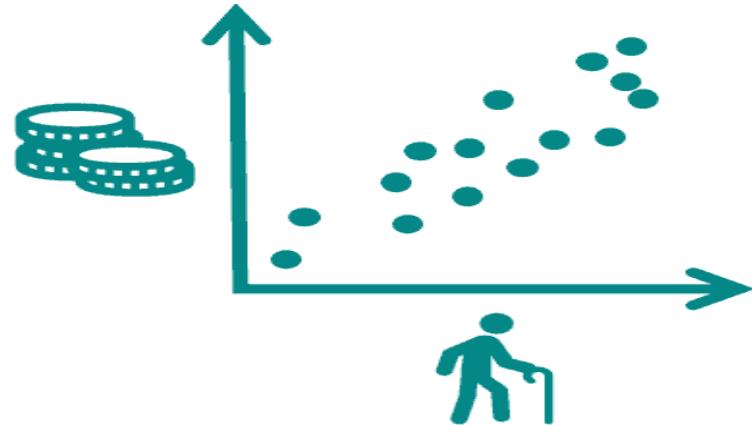
- Bivariate analysis involves examining data that includes two distinct variables. It explores the connections and causations between these variables, aiming to uncover their relationship.
- The graph above illustrates the correlation between applicants and various income brackets, demonstrating a direct proportionality between the two.
- Consequently, as income rises, the amount of credit also increases.

Bivariate Analysis - Amount Income to Amount Credit



Task 5. To Identify Top Correlations for Different Scenarios:

87



Understanding correlations between variables and the target variable can unveil significant indicators of loan default.

Task: Utilize Excel functions to segment the dataset based on various scenarios, such as clients experiencing payment difficulties and all other cases.

Identify the top correlations within each segmented dataset to discern key relationships contributing to loan default.

Task 5.(a)Correlation for Applicants with Payment Difficulties

88

	Correlation For Applicants With Payments Made On Time										
CNT_CHILDREN	1.000	0.036315621	0.006	0.026	0.002	-0.025	-0.336	-0.246	0.879	0.021	
AMT_INCOME_TOTAL	0.036	1.000	0.378	0.451	0.385	0.182	-0.074	-0.162	0.042	-0.205	
AMT_CREDIT	0.006	0.377963032	1.000	0.771	0.987	0.096	0.051	-0.075	0.065	-0.103	
AMT_ANNUITY	0.026	0.451137151	0.771	1.000	0.776	0.117	-0.010	-0.111	0.078	-0.130	
AMT_GOODS_PRICE	0.002	0.384573329	0.987	0.776	1.000	0.099	0.049	-0.072	0.063	-0.105	
REGION_POPULATION_RELATIVE	-0.025	0.181936304	0.096	0.117	0.099	1.000	Chart Area		-0.007	-0.023	-0.539
DAYS_BIRTH(yrs)	-0.336	-0.073764968	0.051	-0.010	0.049	0.030	1.000	0.623	-0.284	-0.009	
DAYS_EMPLOYED(yrs)	-0.246	-0.161685009	-0.075	-0.111	-0.072	-0.007	0.623	1.000	-0.235	0.041	
CNT_FAM_MEMBERS	0.879	0.041598095	0.065	0.078	0.063	-0.023	-0.284	-0.235	1.000	0.022	
REGION_RATING_CLIENT	0.021	-0.205032782	-0.103	-0.130	-0.105	-0.539	-0.009	0.041	0.022	1.000	
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	POPULATION_R	DAYS_BIRTH(yrs)	DAYS_EMPLOYED(yrs)	CNT_FAM_MEMBERS	REGION_RATING_CLIENT	

- The heatmap above illustrates the correlation among various variables for target 0, representing applicants who made payments on time.
- In the heatmap, red indicates the strongest correlation, while green indicates the weakest correlation between variables.
- Key correlations observed include:- AMT_TOTAL_INCOME to AMT_CREDIT- DAYS_BIRTH to DAYS_EMPLOYED- DAYS_EMPLOYED to DAYS_ID_PUBLISH

Task 5.(b) Correlation for Applicants with Payment Difficulties

89

Correlation For Applicants With Payment Difficulties

CNT_CHILDREN	1	0.010110177	0.007601905	0.029172977	-0.001079665	-0.020359154	-0.2496732	-0.189773227	0.892521875	0.055515557
AMT_INCOME_TOTAL	Chart Area 77	1	0.015271444	0.018004594	0.013269502	-0.006180303	-0.009033662	-0.011758681	0.013121678	-0.012846697
AMT_CREDIT	0.007601905	0.015271444	1	0.749665201	0.982267963	0.067775624	0.142506035	0.018782223	0.06124869	-0.045024534
AMT_ANNUITY	0.029172977	0.018004594	0.749665201	1	0.74950403	0.073123998	0.008751713	-0.078113894	0.075838463	-0.061578289
AMT_GOODS_PRICE	-0.001079665	0.013269502	0.982267963	0.74950403	1	0.076635488	0.141005898	0.023181572	0.055135807	-0.051296281
REGION_POPULATION_RELATIVE	-0.020359154	-0.006180303	0.067775624	0.073123998	0.076635488	1	0.016468731	0.007710059	-0.017257146	-0.430032303
DAYS_BIRTH(yrs)	-0.2496732	-0.009033662	0.142506035	0.008751713	0.141005898	0.016468731	1	0.588242824	-0.199141397	-0.045027112
DAYS_EMPLOYED(yrs)	-0.189773227	-0.011758681	0.018782223	-0.078113894	0.023181572	0.007710059	0.588242824	1	-0.183362962	-0.009237108
CNT_FAM_MEMBERS	0.892521875	0.013121678	0.06124869	0.075838463	0.055135807	-0.017257146	-0.199141397	-0.183362962	1	0.057279521
REGION_RATING_CLIENT	0.055515557	-0.012846697	-0.045024534	-0.061578289	-0.051296281	-0.430032303	-0.045027112	-0.009237108	0.057279521	1
	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	DAYS_BIRTH(yrs)	DAYS_EMPLOYED(yrs)	CNT_FAM_MEMBERS	REGION_RATING_CLIE

- The heatmap above illustrates the correlation among various variables for target 0, representing applicants facing payment difficulties.
- The heatmap colour scheme ranges from red, indicating the strongest correlation, to green, indicating the weakest correlation between variables.
- Consequently, the most significant correlations are observed between DAYS_BIRTH and DAYS_EMPLOYED, as well as between DAYS_ID_PUBLISH and DAYS_BIRTH.



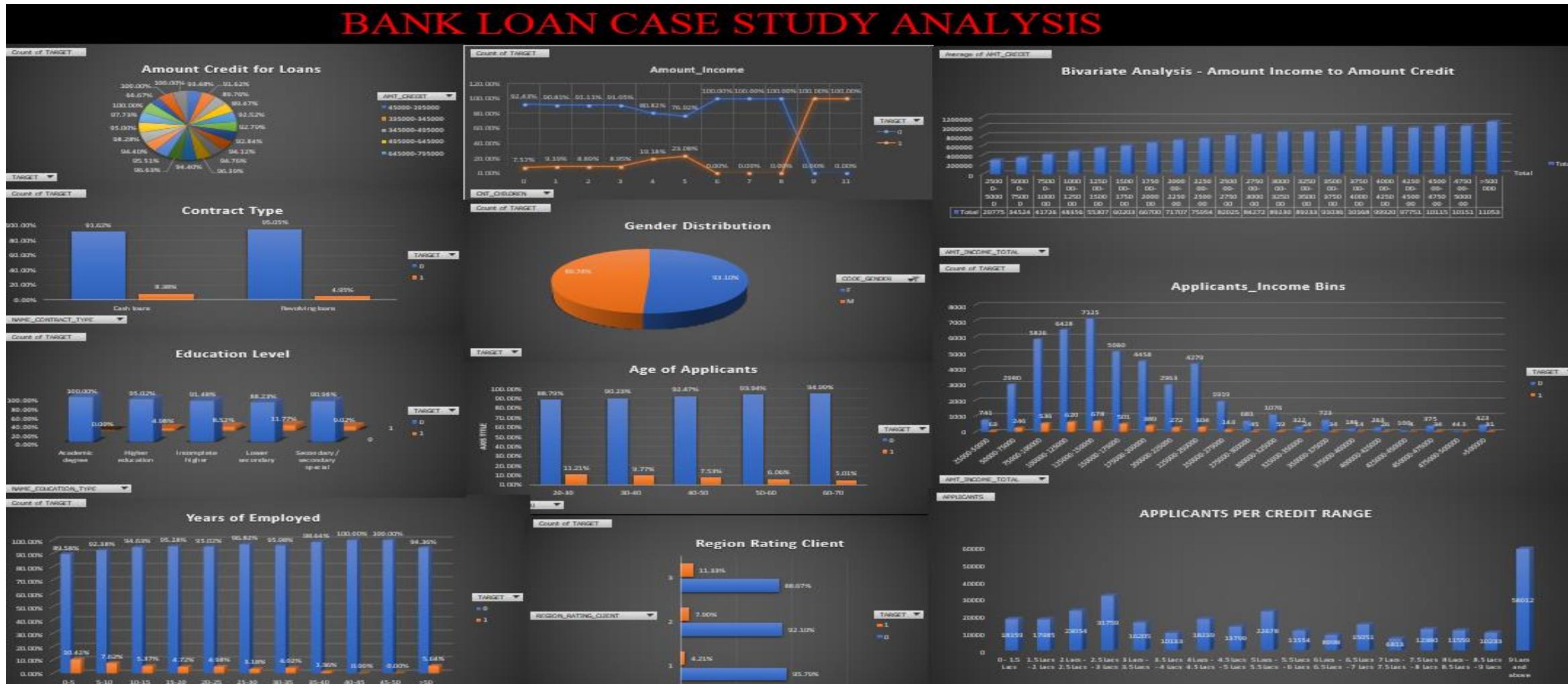
- Higher income correlates with higher credit offers.
- Majority approved for credit range ≥ 9 lakhs.
- Most applicants (0 & 1) earn 1-2.25 lakhs.
- Factors influencing default likelihood:
 - ✓ Age and experience inversely correlate with default.
 - ✓ Older, more experienced clients may be prioritized.
 - ✓ Higher education linked to lower default rates.
 - ✓ Males have higher default likelihood than females.
 - ✓ Region 3 has highest defaulters, suggesting stricter loan conditions.
 - ✓ Region 1 presents lowest default risk.
 - ✓ Loan amounts tend to increase with client age.
 - ✓ Older clients may pose lower default risks.
 - ✓ Clients with >2 children exhibit higher default frequencies.



Dashboard for Bank Loan Case Study

91

By : Raj Rathod
Data Analyst



Result

Result and Conclusion

92

- Practical Application of Exploratory Data Analysis (EDA) in Bank Loan Analysis.
- Acquired Fundamental Understanding of Risk Analytics in Banking and Financial Services.
- Addressed Challenges in Analysing Correlations Among Variables for Meaningful Insights.
- Gained Insights into Data Imbalances, Identifying Outliers, and Understanding Key Drivers.
- Visualization Played a Pivotal Role in Synthesizing Complex Datasets for Client Communication.





Project 6: Bank Loan Case Study

93

Project Result Summary

- **Description:** Conducted Exploratory Data Analysis (EDA) to prevent capable applicants from rejection due to insufficient credit history.
- **Approach:** Utilized Microsoft Excel for data cleaning, outlier detection, and univariate and bivariate analysis.
- **Findings:**
 - ✓ Identified correlations between income, credit amounts, age, education, and loan default likelihood.
 - ✓ Conducted exploratory data analysis to identify factors contributing to loan default
 - ✓ Provided recommendations for risk assessment and loan approval strategies
 - Quantitative Metrics
 - ✓ Identified Region 3 as having the highest defaulters, suggesting stricter loan conditions
 - ✓ Revealed clients with more than 2 children exhibited higher default frequencies



Access Excel Dataset



94

By : Raj Rathod
Data Analyst

THANK YOU!



Watch Video Presentation

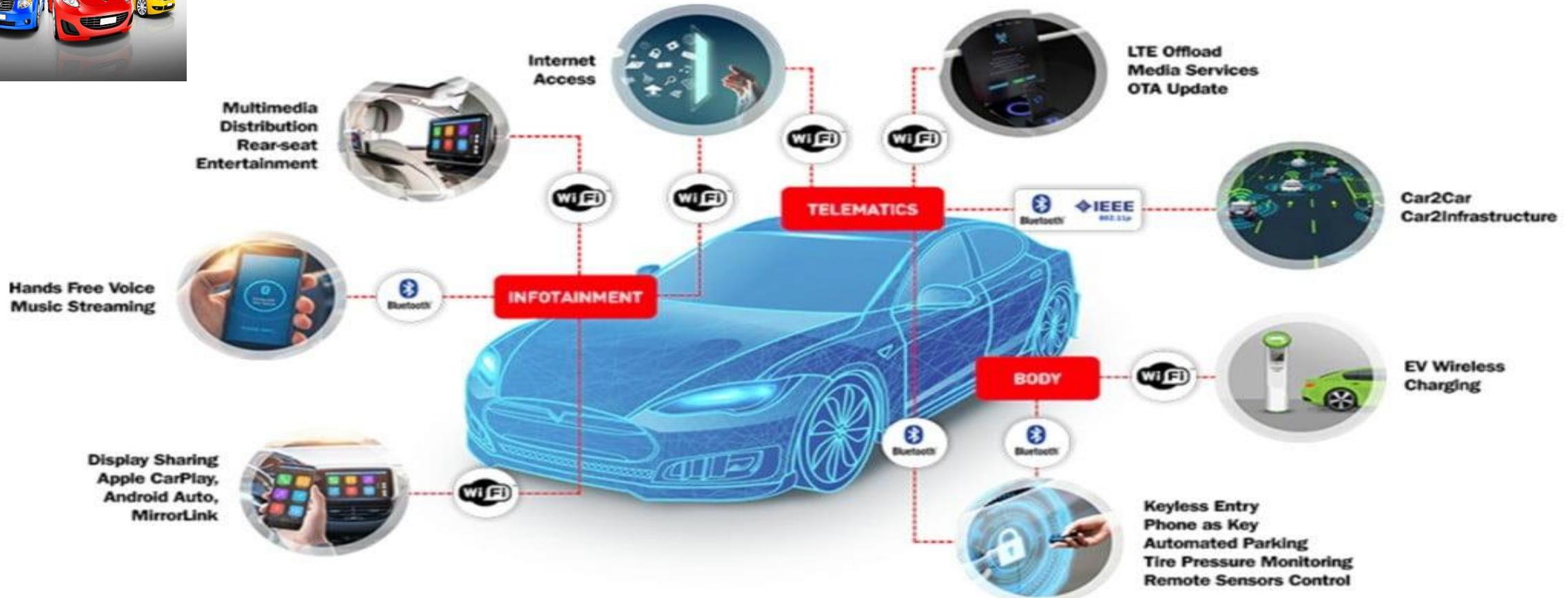


Project 7: Analyzing the Impact of Car Features on Price & Profitability

95

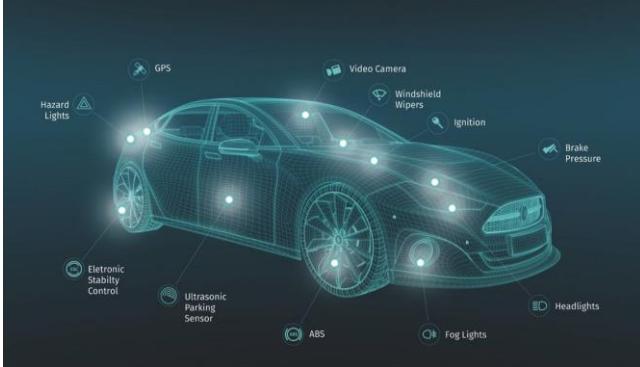
By : Raj Rathod
Data Analyst

Introduction:



Project Overview

96



➤ Automotive Industry Transformations:

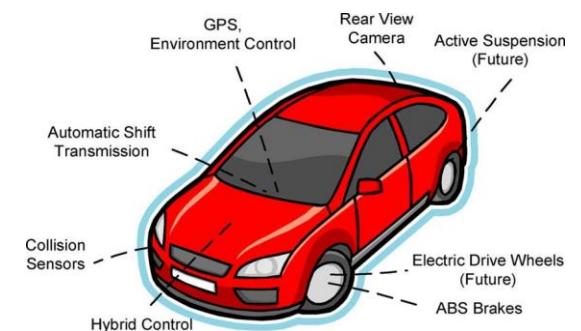
- ✓ Driven by fuel efficiency, environmental concerns, and technological advancements.
- ✓ Understanding consumer demand dynamics is crucial amidst competition and evolving preferences.

➤ Project Objective:

- ✓ Analyse relationship between car features, market category, pricing.
- ✓ Identify key factors affecting profitability and consumer demand.

➤ Client's Key Question:

- ✓ How to strategically set prices and develop products for enhanced profitability and meeting consumer demand?





Project Approach

97

- **Explore Trends:** Analyze car features and pricing over time.
- **Relationship Analysis:** Investigate engine power, fuel efficiency, and pricing correlations.
- **Feature Importance:** Identify key features influencing car prices.
- **Pricing Variations:** Examine price differences across manufacturers and market categories.
- **Methodology:**
- ✓ **Utilize Advanced Excel Skills:** Regression analysis and data visualization techniques.
- ✓ **Objective:** Provide valuable insights for pricing and product development decisions.

Microsoft Excel 2021:

- Essential tool for data cleaning, analysis, and visualization.
- Sorting, filtering, and formula capabilities used for data cleansing and profit calculation.
- Functions employed for extracting top movies and grouping data.
- Charting features utilized for effective data visualization.
- User-friendly interface and diverse functionalities made it the ideal choice for the project.

Microsoft PowerPoint:

- For Presentation Purpose.



Task 1. Popularity of Car Models

99

Insight :

Top 20 Market Category Popularity of Cars

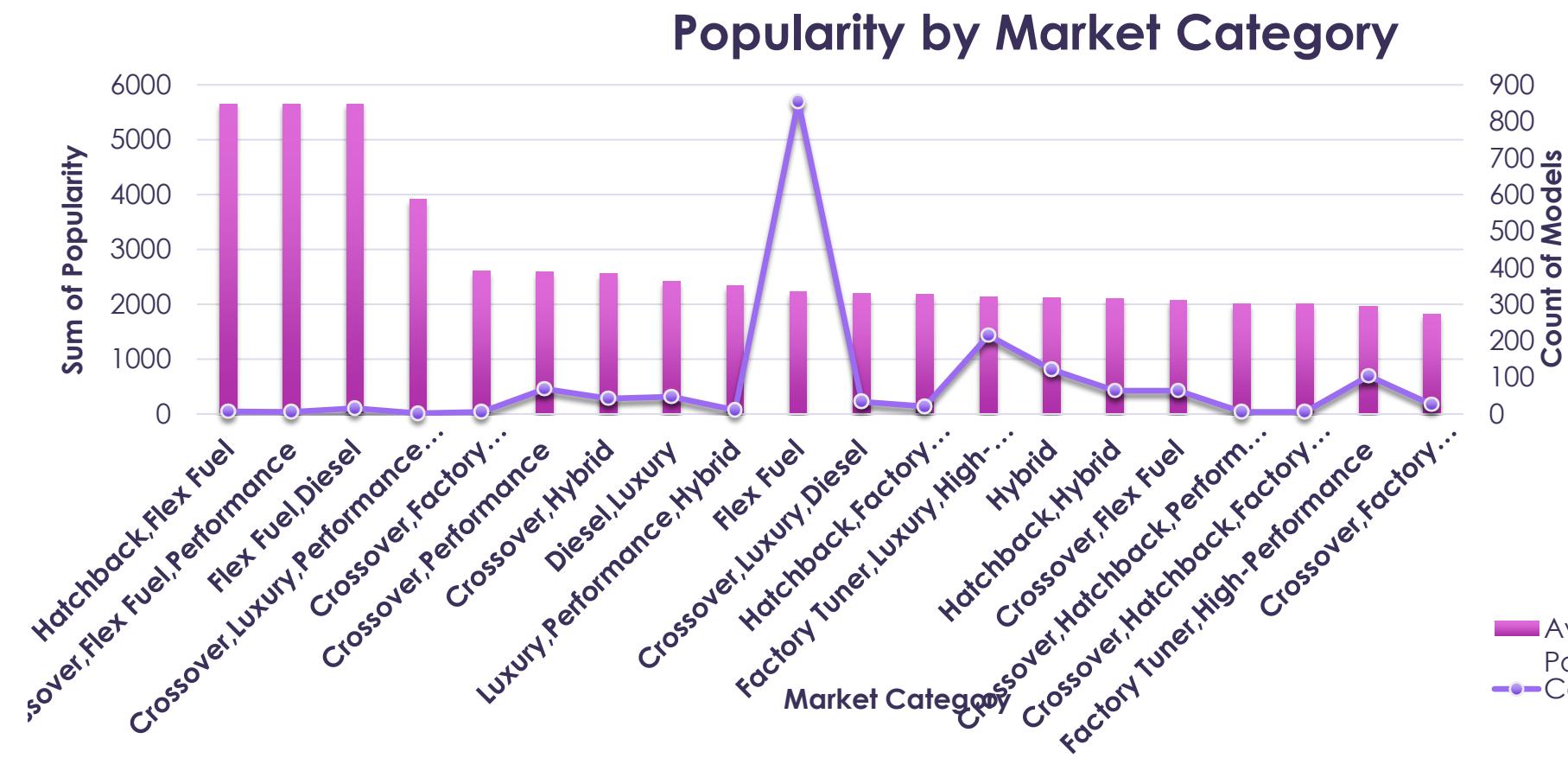
Market Category	Average of Popularity	Count of Model
Hatchback,Flex Fuel	5657	7
Crossover,Flex Fuel,Performance	5657	6
Flex Fuel,Diesel	5657	16
Crossover,Luxury,Performance,Hybrid	3916	2
Crossover,Factory Tuner,Luxury,Performance	2607	5
Crossover,Performance	2586	69
Crossover,Hybrid	2563	42
Diesel,Luxury	2416	47
Luxury,Performance,Hybrid	2333	11
Flex Fuel	2226	855
Crossover,Luxury,Diesel	2196	33
Hatchback,Factory Tuner,Performance	2174	21
Factory Tuner,Luxury,High-Performance	2133	215
Hybrid	2117	121
Hatchback,Hybrid	2111	64
Crossover,Flex Fuel	2074	64
Crossover,Hatchback,Performance	2009	6
Crossover,Hatchback,Factory Tuner,Performance	2009	6
Factory Tuner,High-Performance	1966	104
Crossover,Factory Tuner,Luxury,High-Performance	1823	26
Grand Total	2262	1720



Task 1. Popularity of Car Models

100

Insight : The Most Popular Category of Cars is Hatchback, Flex Fuel/Diesel, Crossover and Luxury Cars.

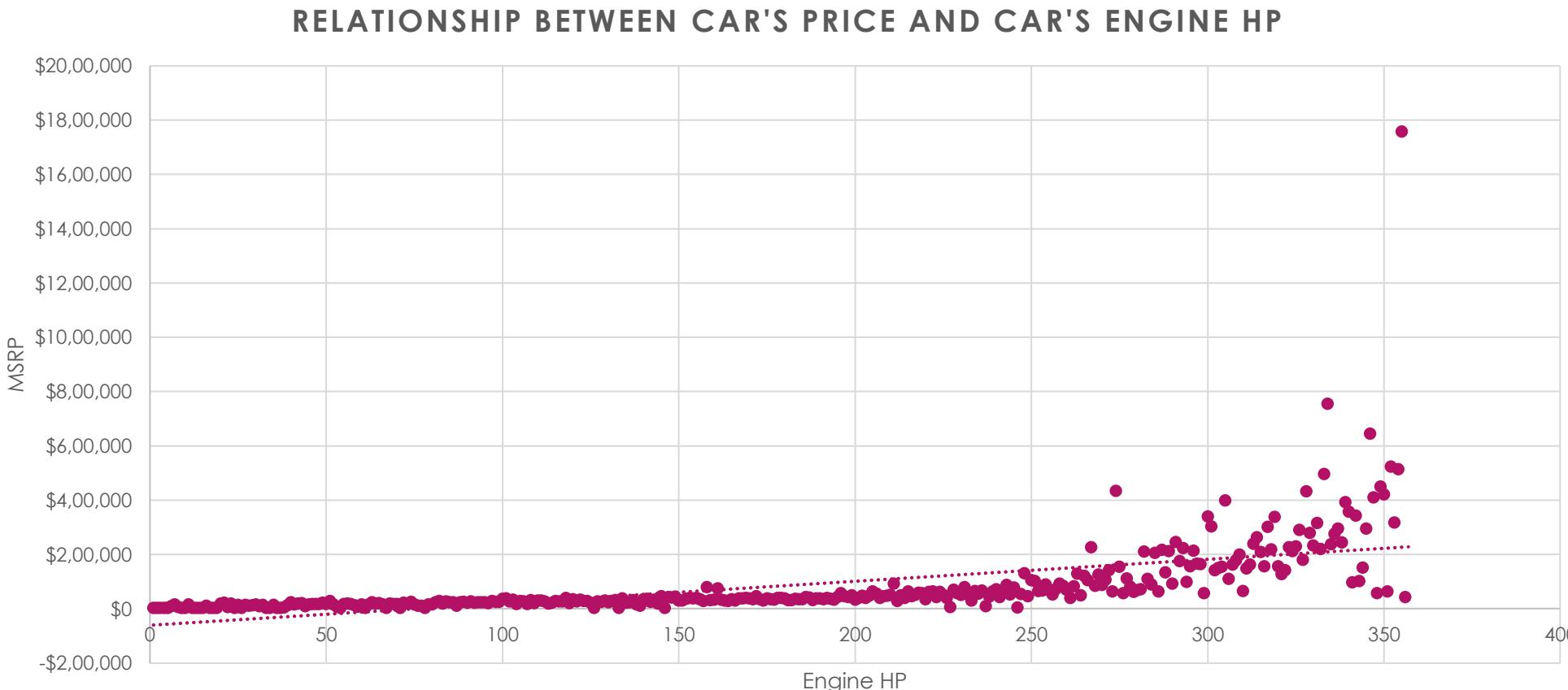


Average of
Popularity
Count of Model

Task 2. Relationship Between A Car's Engine Power And Its Price?

101

Insight : Below is the Scatter Chart Plot which shows that as the Engine Power i.e HP Increases Prices also increases and can say that it is directly proportional.

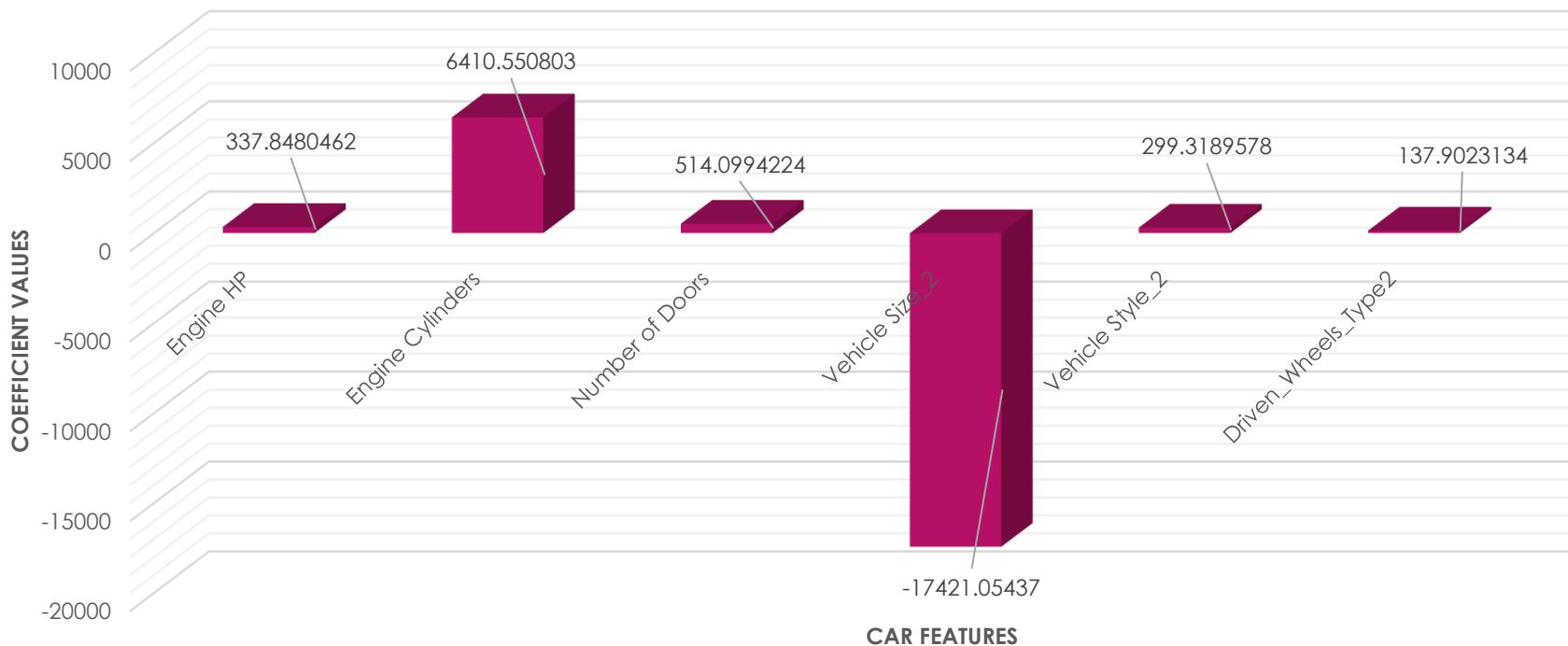


Task 3. Which Features Are Most Important For Determining Price?

102

Insight: The Features which is important with respect to price is Engine Cylinders, Engine HP (Power) and then Least is Vehicle Size. That is why Cars which are expensive like Sports Car are the Costliest.

Relative Importance of Car Features on Price



Task 4. Average Price By Manufacturer

103

Insight : Below is the List of Average Price of Cars of Each Manufacturer or Maker, where Bugatti is the Most Expensive Car Maker by Average Price.

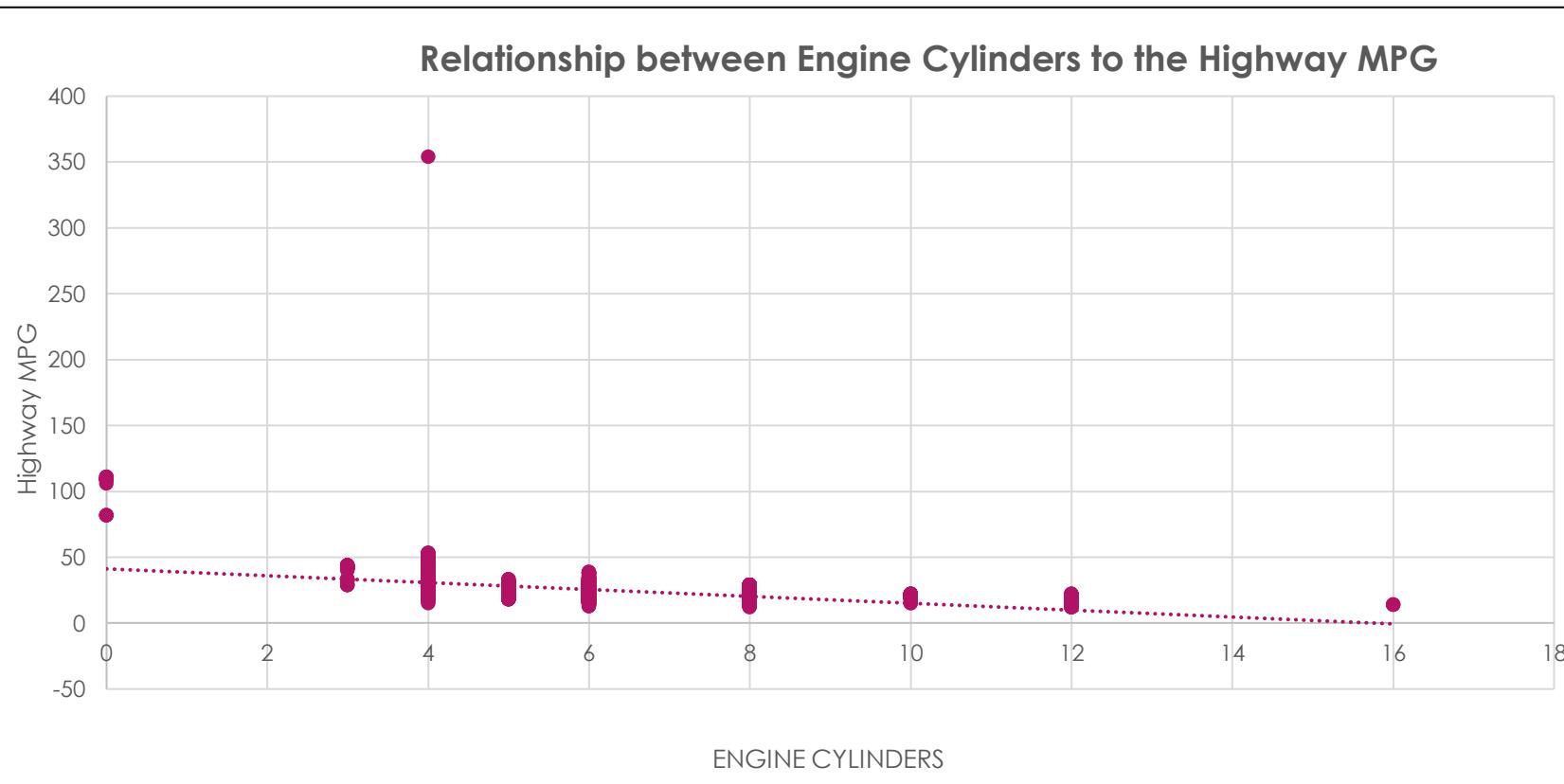
AVERAGE PRICE OF CAR ACROSS DIFFERENT MANUFACTURERS



Task 5. Relationship Between Fuel Efficiency And Engine Cylinders

104

Insight : As per the Analysis, When the Engine Cylinders are Higher then the Highway MPG is Lower because More Cylinders More Powerful Engine thus Lower MPG

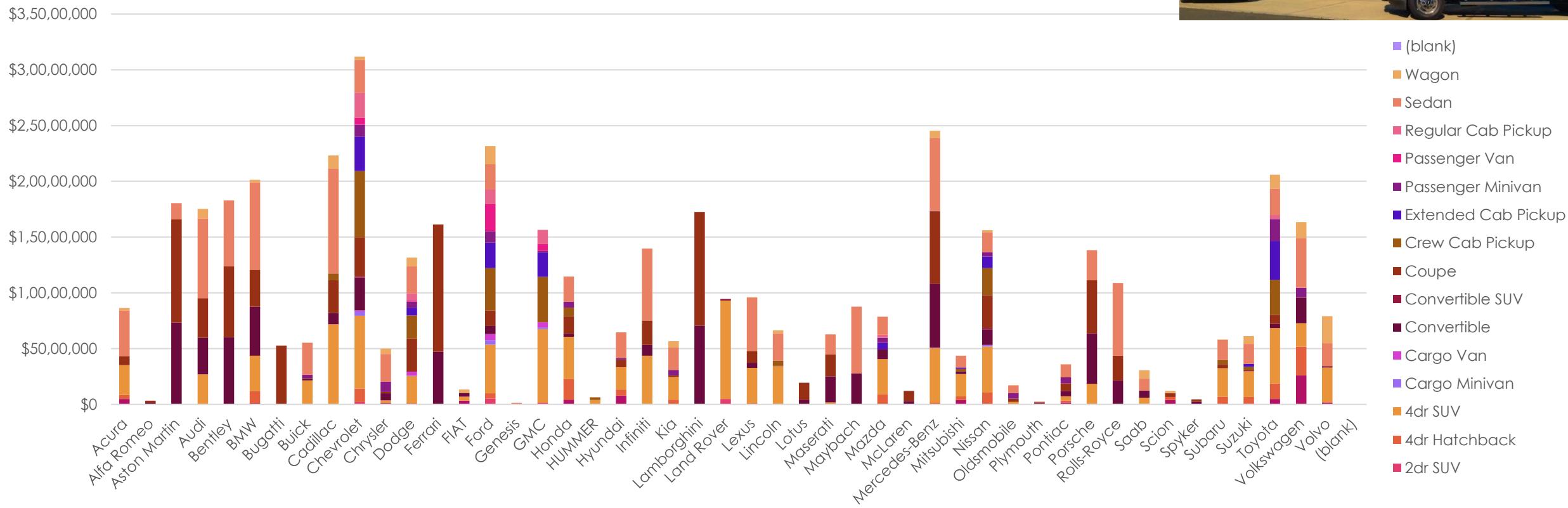


Task 6. Distribution of Car Prices

105

Insight: Chevrolet has the Highest Car Price Distribution Overall

Distribution of Car Price by Brand and Body Style



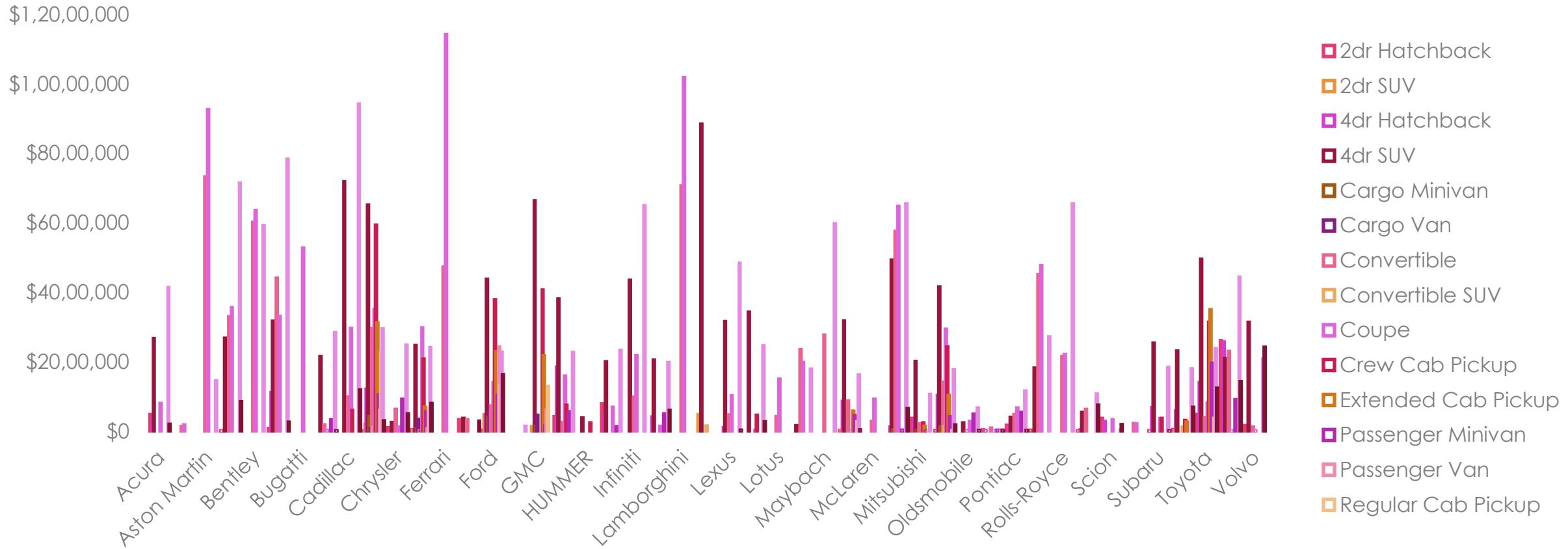


Task 7. Average Price of Car and Body Style

Insight: Ferrari has the highest average price of car 4dr Hatchback Body style.



Average Price of Car and Body Style

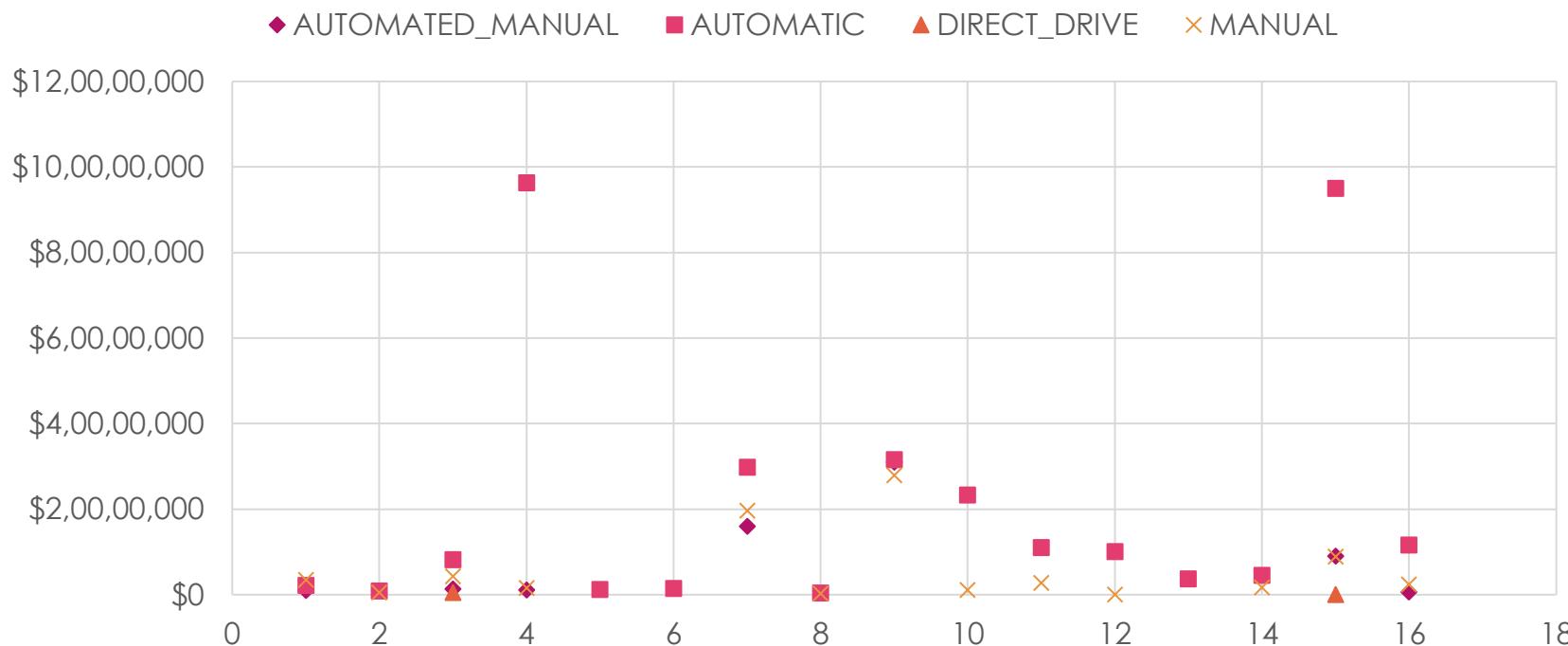


Task 8. Relationship of Price on Transmission Type

107

Insight: Greater Performance Corresponds to Higher Pricing, Especially in Transmission.

RELATIONSHIP BETWEEN MSRP AND TRANSMISSION TYPE



TYPES OF TRANSMISSION SYSTEM

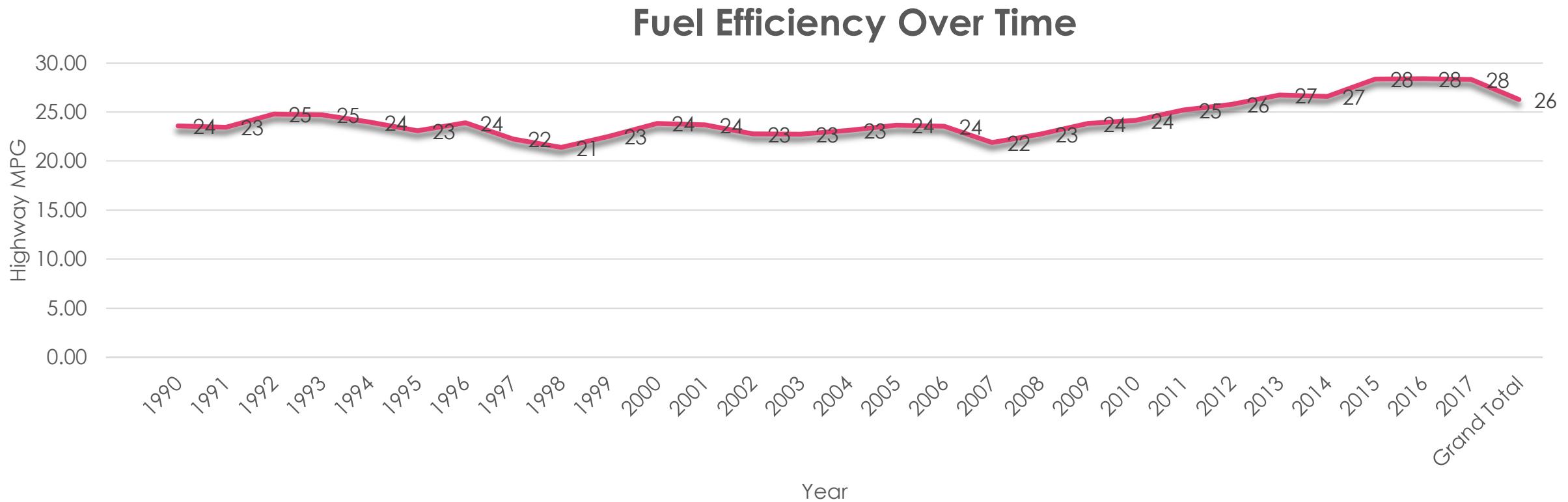




Task 9. Fuel Efficiency Over Time

108

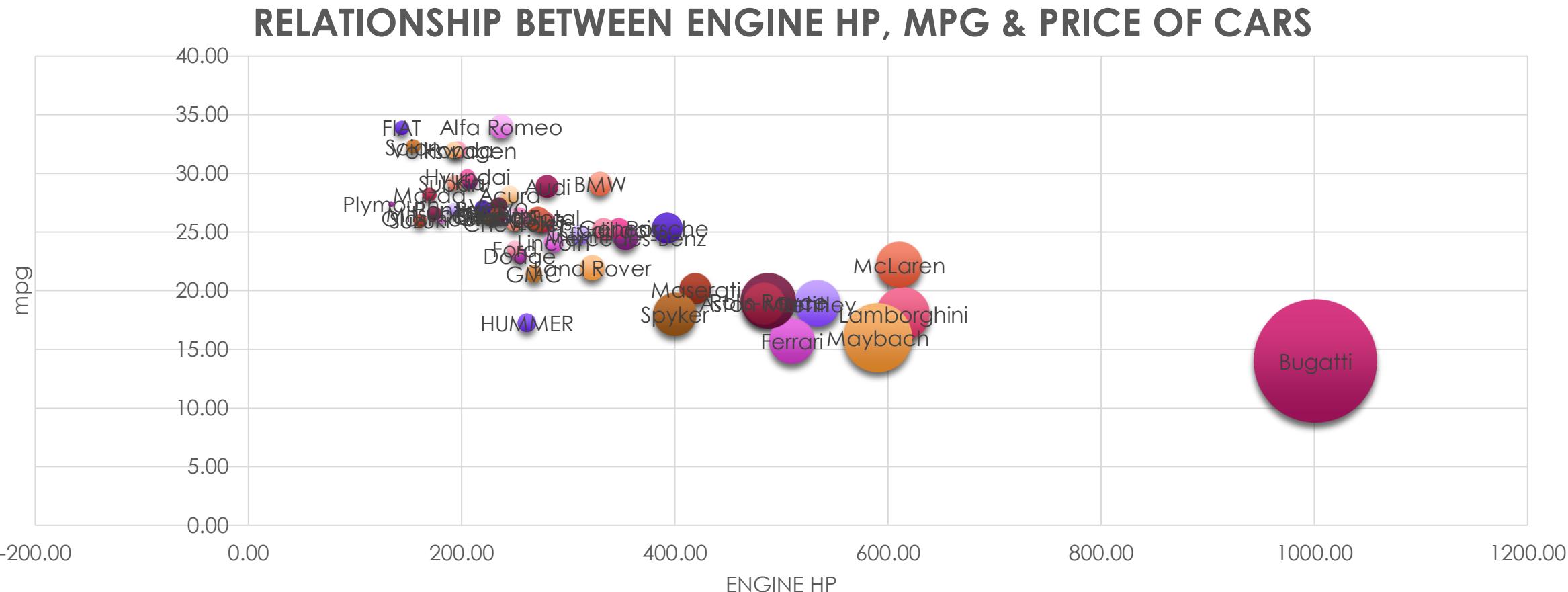
Insight: The ongoing improvement in engines is reflected in the increasing fuel efficiency observed annually, as depicted in the graph.



Task 10. Relationship between HP, MPG and Price

109

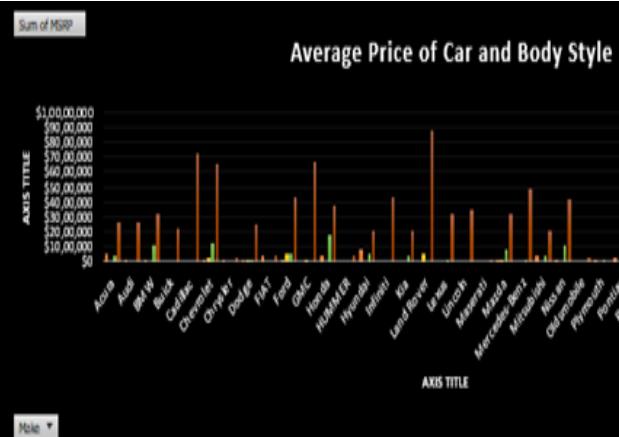
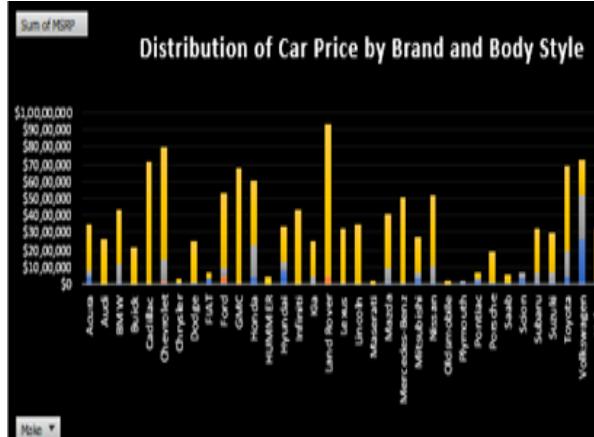
Insight: Bugatti has the highest average price of car which also has Lowest MPG (Mileage) and High Engine Horse Power



Dashboard for Impact of Car Features

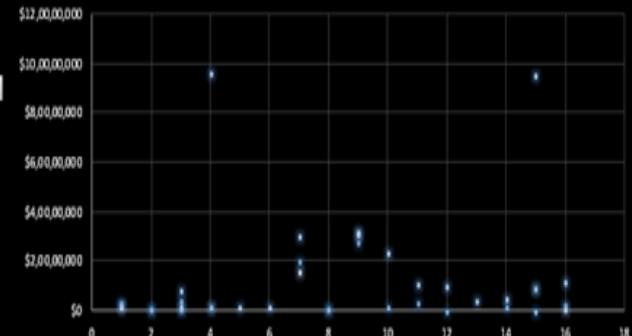
110

By : Raj Rathod
Data Analyst

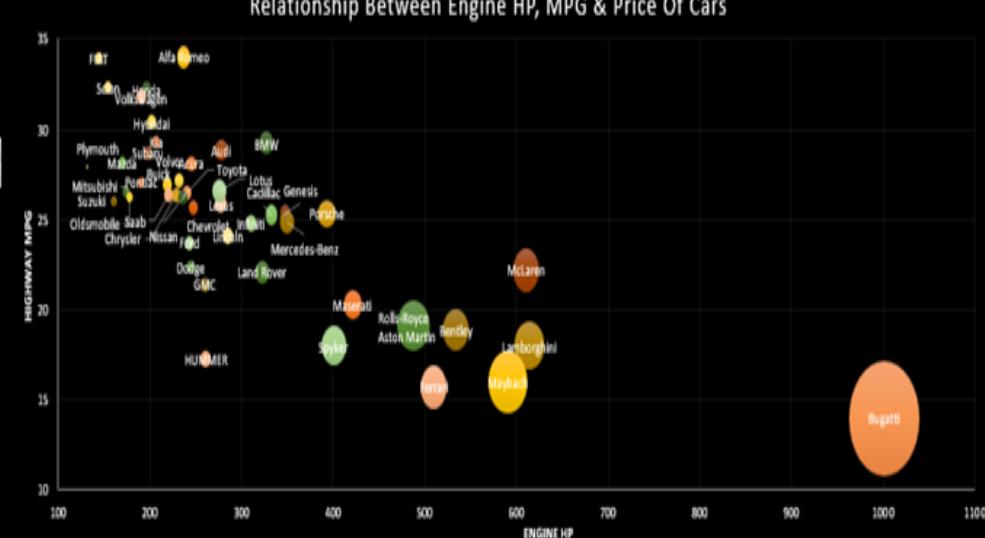
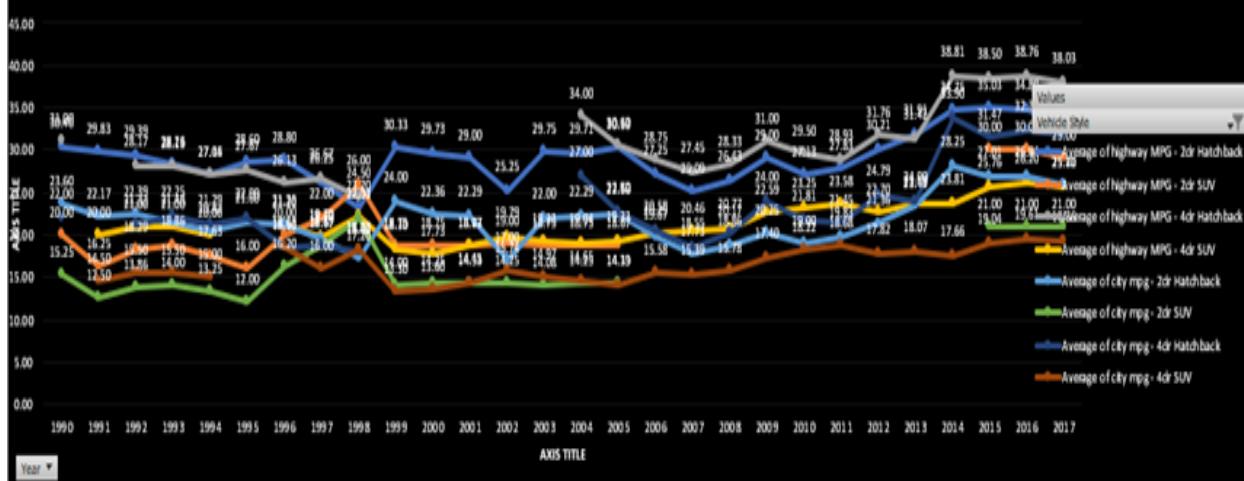


Relationship Between MSRP And Transmission Type

*AUTOMATED_MANUAL * AUTOMATIC * DIRECT_DRIVE * MANUAL



Fuel Efficiency Of Cars Vary Across Different Body Styles And Model Years





Result and Conclusion

111

- **Popular Categories:** Hatchback, Flex Fuel/Diesel, Crossover, Luxury Cars
- **Price vs. Engine Power:** Direct proportionality indicates consumers' preference for higher performance.
- **Key Price Influencers:** Engine Cylinders, HP, Vehicle Size. Sports Cars command higher prices.
- **Average Prices:** Bugatti: \$1,757,224 USD | Plymouth: \$3,297 USD
- **Engine Cylinders vs. MPG:** Higher cylinders correlate with lower highway MPG.

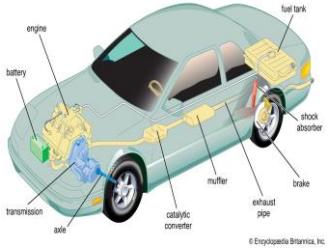


Result and Conclusion



- **Popular Category & Brand:** Sedan - Chevrolet tops.
- **Fuel Efficiency:** Passenger Vans (Avg. MPG: 12.53) vs. 4dr Hatchback Cars (Avg. MPG: 35.39).
- **Performance & Efficiency:** Bugatti (High HP, Low MPG), Fiat (Low HP), Alfa Romeo (High MPG), Plymouth (Most Affordable).
- **Conclusion:** Valuable insights for manufacturers to understand preferences, optimize pricing, and focus on key features.





Project 7: Analyzing the Impact of Car Features on Price & Profitability

113

Projects Result Summary

- **Description:** Analyzed the ratio between applicants facing payment difficulties and those making payments on time to enhance risk assessment.
- **Approach:** Utilized Excel for univariate analysis, examining the count of loan applicants for various income brackets.
- **Insights:**
 - ✓ Discovered correlations between income and credit amounts, providing insights for loan approval decisions.
 - ✓ Analyzed the relationship between car features, market categories, and pricing
 - ✓ Provided insights for strategic pricing and product development decisions Quantitative Metrics:
 - ✓ Determined the top market category for popularity (Hatchback, Flex Fuel) with 5,657 average popularity
 - ✓ Revealed Bugatti as the most expensive car manufacturer with an average price of \$1,757,224



Access Excel Dataset

114



Watch Video Presentation



Project 8: ABC Call Volume Analysis

115



Introduction:

By : Raj Rathod
Data Analyst





Introduction



116

- Customer Experience (CX) is a critical differentiator in today's competitive landscape
- ABC Insurance Company understands the importance of providing exceptional customer support
- This project focuses on analyzing inbound call volume trends to enhance CX strategy





About ABC Insurance Company

117

- Leading player in the insurance sector
- Committed to delivering exceptional customer experience (CX)
- Recognizes the importance of seamless customer support



- Leveraged a comprehensive dataset spanning 23 days
- Performed analytical tasks:
- Calculated average call duration per time bucket
- Visualized total call volumes using charts/graphs
- Proposed manpower allocation plans
- Utilized domain knowledge and statistical techniques



- **Microsoft Excel 2021:**
Utilized for data cleaning, outlier detection, and conducting univariate and bivariate analysis using pivot tables and charts.

- **Microsoft PowerPoint:**
Employed for presentation purposes, summarizing key findings and insights derived from the bank loan case study analysis project.





Project Objectives

120

- Analyze average call duration for each time bucket
- Visualize total call volumes across time buckets
- Propose manpower plan to reduce abandon rate (9 AM - 9 PM)
- Develop night-time manpower plan (9 PM - 9 AM)





Project Assumptions

121

- Agent works 6 days a week
- Average 4 unplanned leaves per agent per month
- 9-hour shift with 1.5 hours for lunch/snacks
- Agents spend 60% of working hours on calls
- 30 days in a month

Project Assumptions

Working Days	6
Working Hours	9
Break	1.5
Actual Working Hours	7.5
Occupancy	60%
Total Working Hours	4.5
Total Working Seconds	16200
Average Call Time Per Agent	199
Call Capacity of an Agent/Day	81
Call Capacity of an Agent/Hour	18



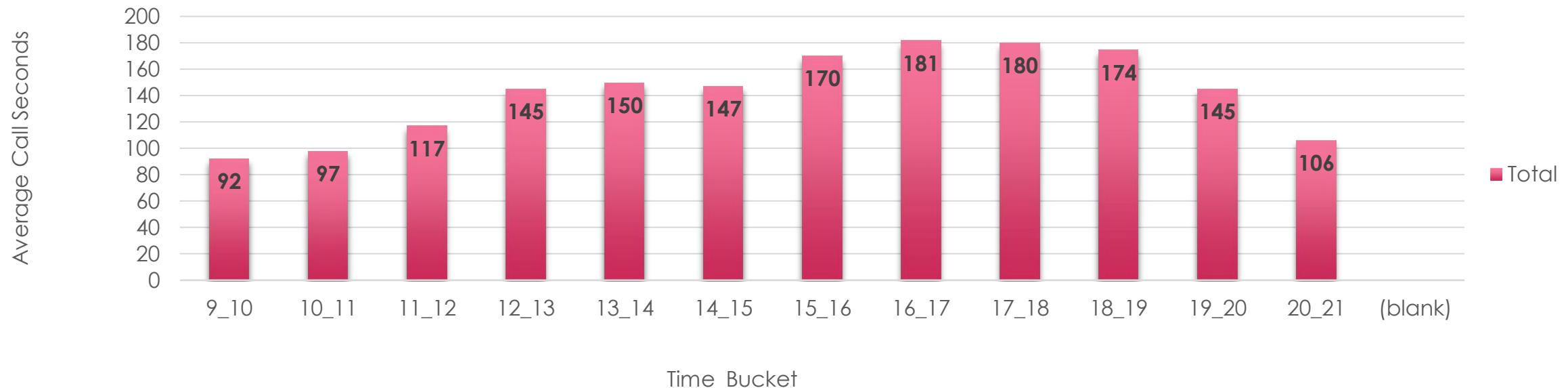


Task 1. Average Call Duration Analysis

122

Insights : Calculated average duration of incoming calls for each time bucket is 199 seconds.

Average Call Duration



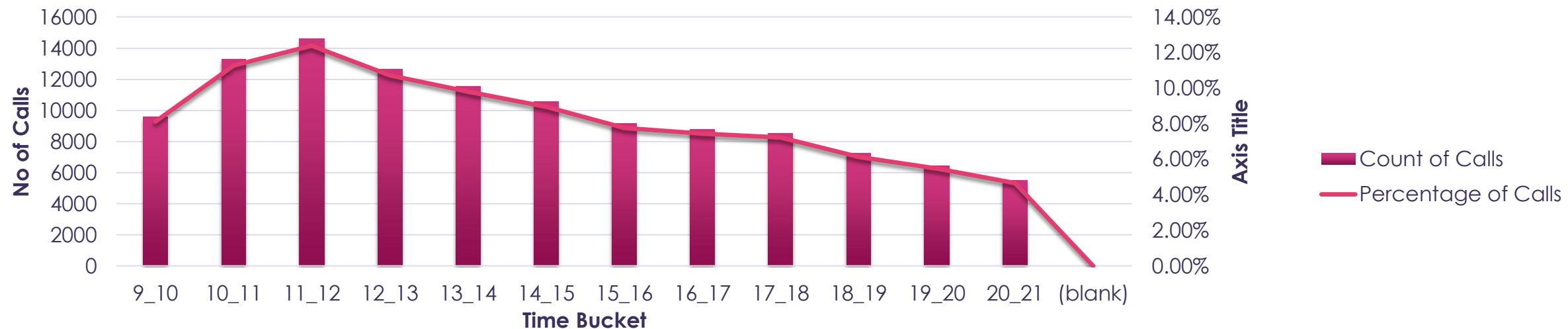
Task 2. Call Volume Analysis



Insight:

- Visualize total number of calls received across time buckets
- Identify call volume patterns and peak demand periods
- Peak Call Duration: 10 AM - 11 AM and 7 PM - 8 PM
- Lowest Call Duration: 12 noon - 1 PM

Call Volume Analysis



Task 3. Manpower Planning (9 AM - 9 PM)

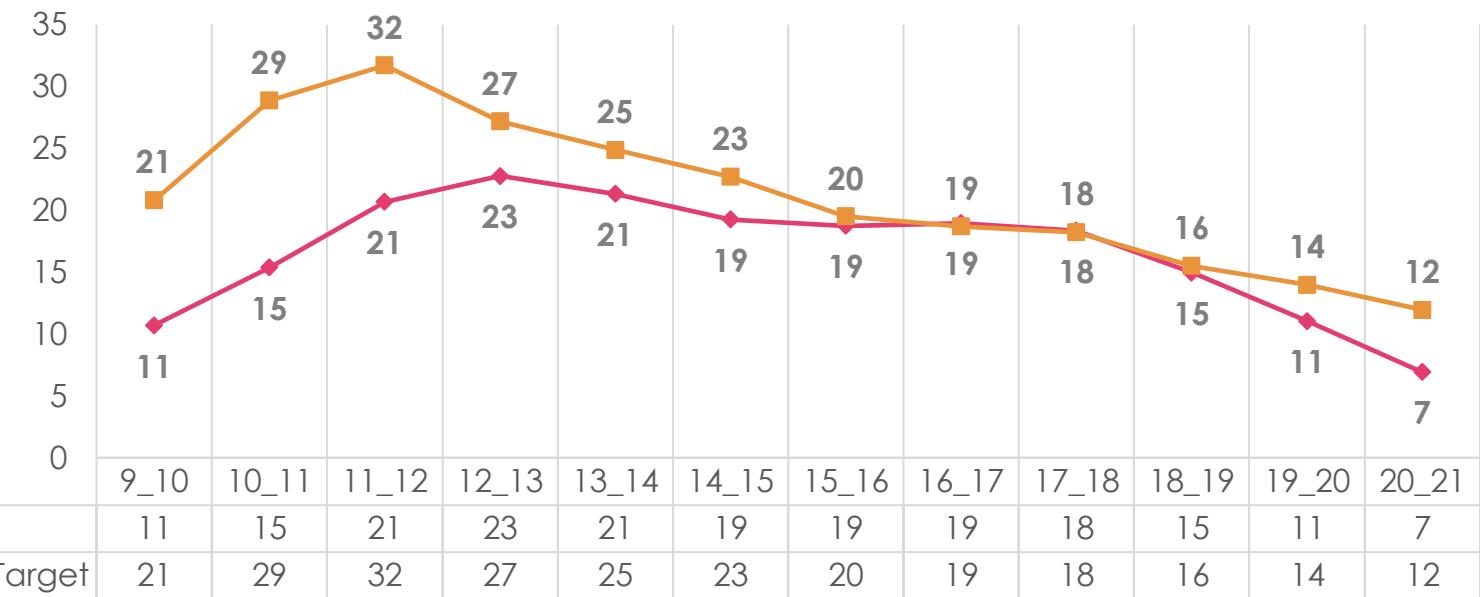
124

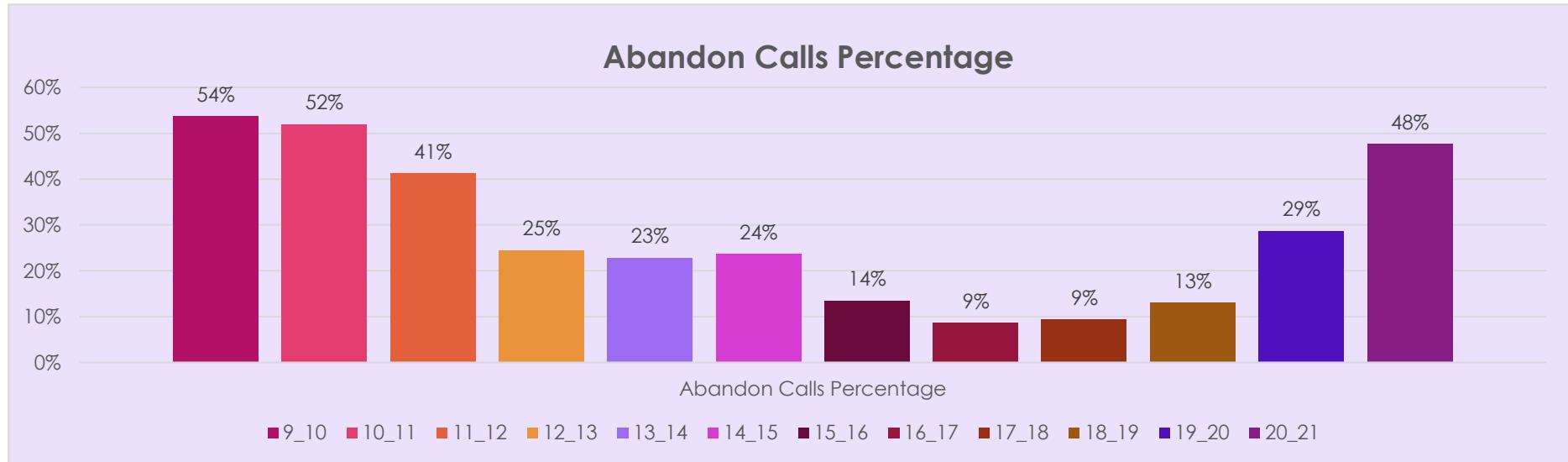
Insight:

- Current abandon rate: 30%
- Target: Reduce abandon rate to 10%
- Propose manpower allocation plan for each time bucket
- Ensure at least 90 out of 100 calls are answered promptly



AGENTS WORKING VS AGENT NEEDED



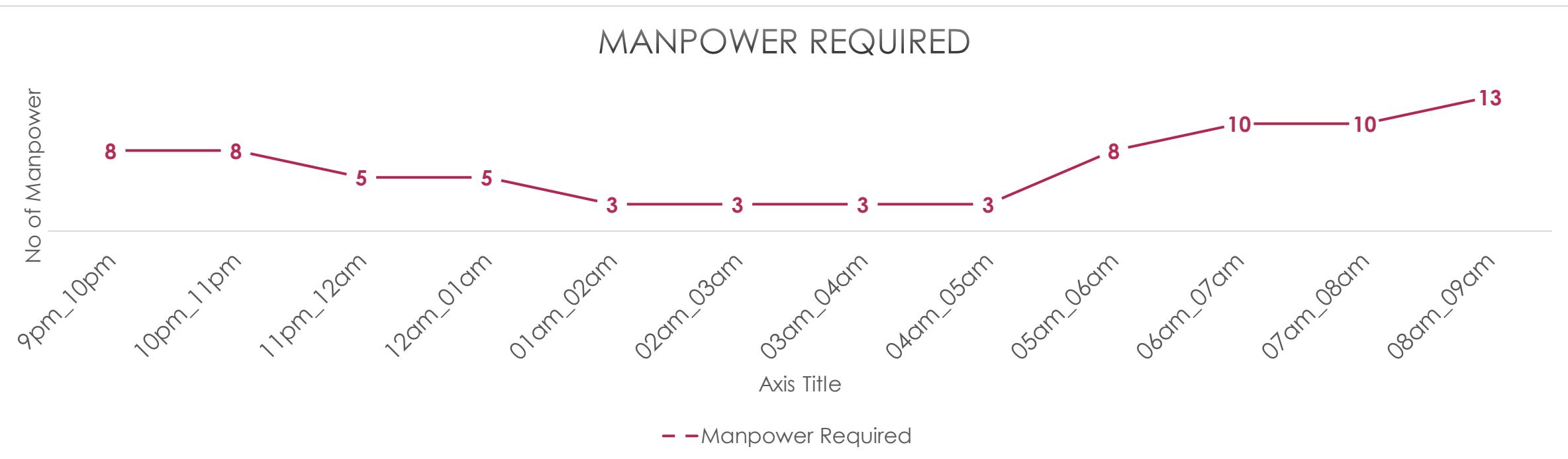


Time Bucket	Abandon Calls/23 Days	Answered /23 Days	Grand Total/23 Days	Abandon Calls/Day	Answered Calls/Day	Grand Total/Day	Abandon Calls Percentage	No of Calls Required to be answered to make Abandon % upto 10
9_10	5149	4428	9577	224	193	416	54%	375
10_11	6911	6368	13279	300	277	577	52%	520
11_12	6028	8560	14588	262	372	634	41%	571
12_13	3073	9432	12505	134	410	544	25%	489
13_14	2617	8829	11446	114	384	498	23%	448
14_15	2475	7974	10449	108	347	454	24%	409
15_16	1214	7760	8974	53	337	390	14%	351
16_17	747	7852	8599	32	341	374	9%	336
17_18	783	7601	8384	34	330	365	9%	328
18_19	933	6200	7133	41	270	310	13%	279
19_20	1848	4578	6426	80	199	279	29%	251
20_21	2625	2870	5495	114	125	239	48%	215
Grand Total	34403	82452	116855	1496	3585	5081	29%	4573

Task 4. Night Shift Manpower Planning (9 PM - 9 AM)

126

- Accommodate additional 30 calls for every 100 calls received during operational hours
- Develop strategic manpower plan for each time bucket
- Maintain maximum abandon rate at 10%





Task 5. Manpower Planning 9AM-9AM

127

- Propose a comprehensive manpower plan for the entire day
- Accommodate additional 30 calls for every 100 calls received during operational hours
- Develop strategic manpower allocation for each time bucket (24 hours)
- Maintain maximum abandon rate at 10%

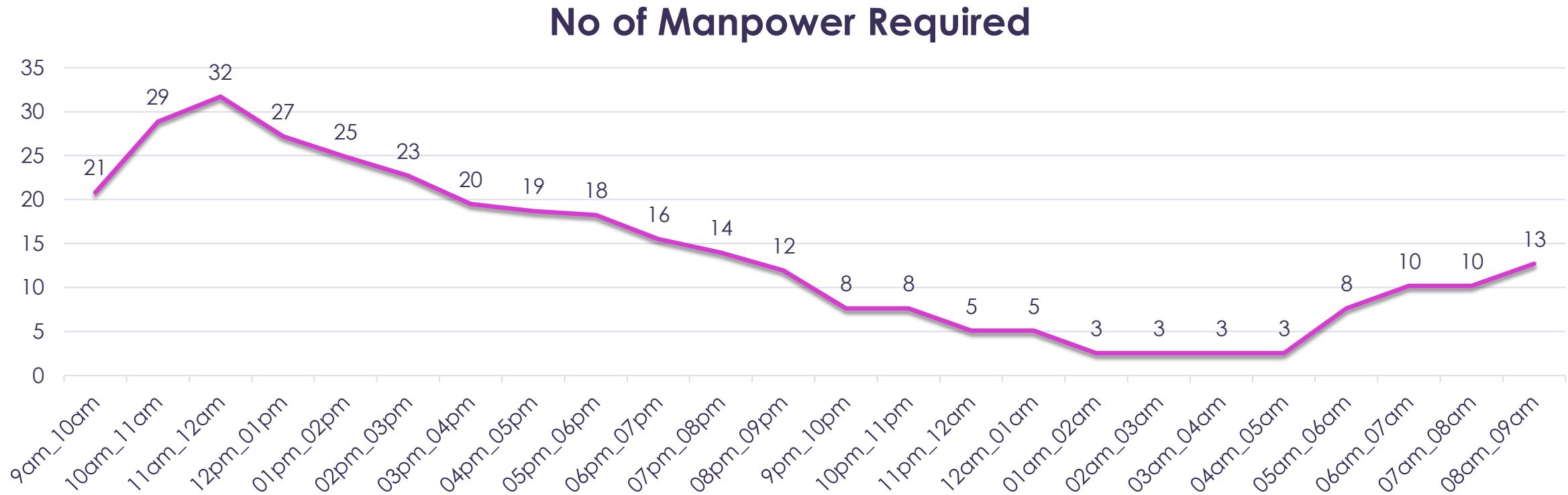
Time Bucket	No of Manpower Required
9am_10am	21
10am_11am	29
11am_12am	32
12pm_01pm	27
01pm_02pm	25
02pm_03pm	23
03pm_04pm	20
04pm_05pm	19
05pm_06pm	18
06pm_07pm	16
07pm_08pm	14
08pm_09pm	12
9pm_10pm	8
10pm_11pm	8
11pm_12am	5
12am_01am	5
01am_02am	3
02am_03am	3
03am_04am	3
04am_05am	3
05am_06am	8
06am_07am	10
07am_08am	10
08am_09am	13



Task 5. Manpower Required whole Day

128

Insight: 76 No of Manpower required to meet the Target Rate.



- Identified peak call volume periods and patterns
- Observed variations in average call duration across time buckets
- Recognized the need for strategic manpower allocation
- Gained understanding of customer engagement and operational efficiency

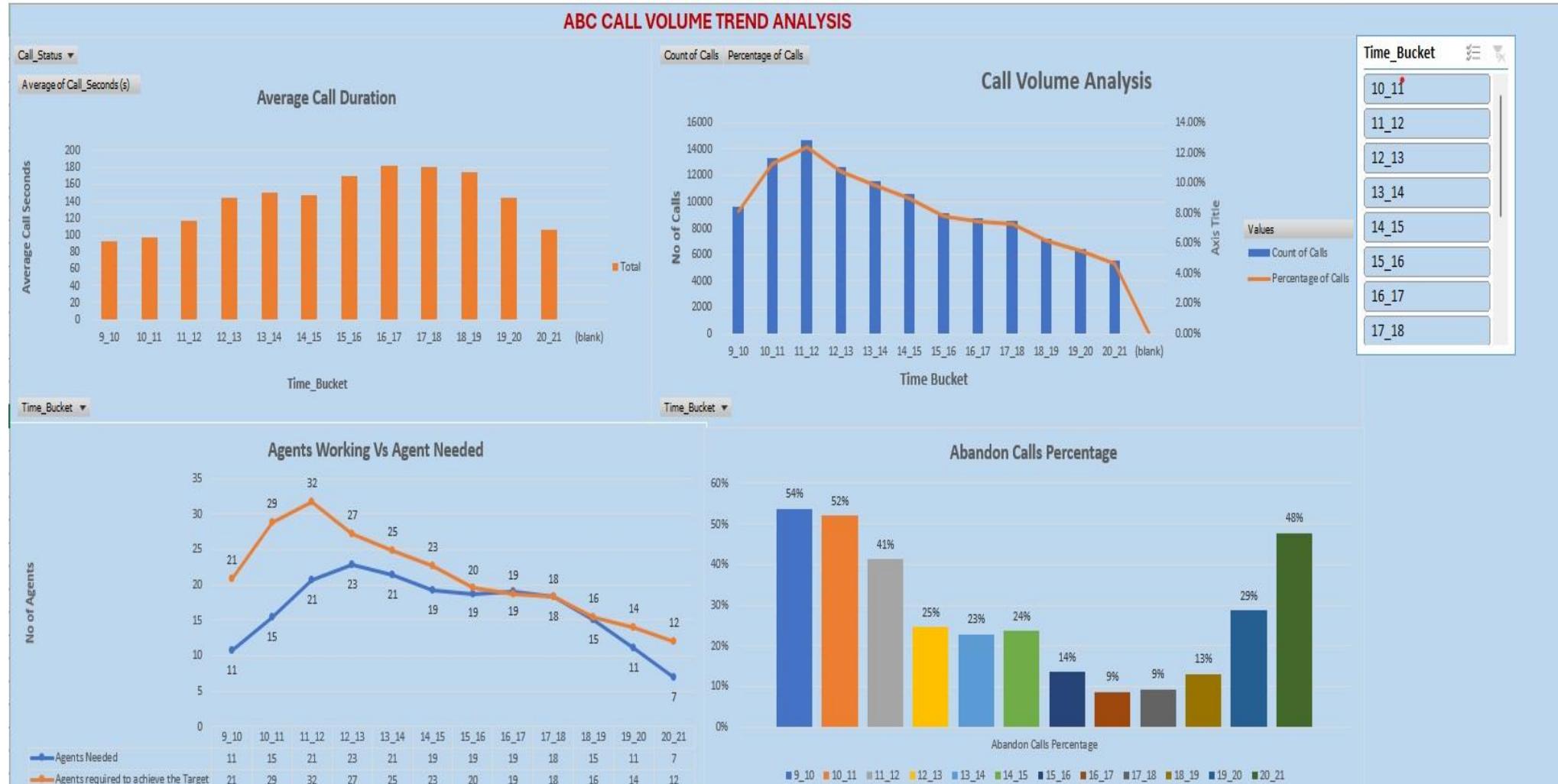
insight



Dashboard for ABC Call Volume Analysis

130

By : Raj Rathod
Data Analyst





Result and Conclusion



131

- Identified the pivotal role of data analysts in enhancing Customer Experience (CX) strategies for customer service departments.
- Recognized the significance of employing effective customer handling techniques, including AI-enabled Interactive Voice Response (IVR) systems, for efficient query resolution.
- Leveraged pre-processed data with time buckets and call duration metrics to streamline analysis efforts and improve operational efficiency.
- Explored behavioral analytics to uncover customer behavior patterns, identify trends, preferences, and opportunities for CX enhancement.
- Acquired comprehensive understanding of customer service operations and the integral role of analysts in driving customer satisfaction through data-driven insights.



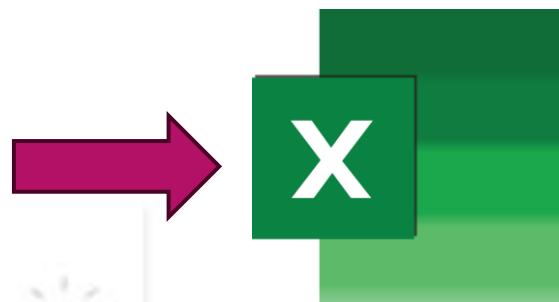
Project 8: ABC Call Volume Analysis

132

Project Result Summary

- **Description:** Explored correlations among variables to identify key indicators of loan default for different applicant scenarios.
- **Approach:** Utilized Excel functions to segment the dataset based on payment difficulties and all other cases.
- **Findings:**
 - ✓ Uncovered significant correlations like income to credit amounts, age to employment duration, and education level to default rates.
 - ✓ Analyzed inbound call volume trends to enhance customer experience strategy
 - ✓ Proposed strategic manpower allocation plans to reduce abandon rates Quantitative Metrics
 - ✓ Identified peak call volume periods (10 AM - 11 AM and 7 PM - 8 PM)
 - ✓ Recommended increasing manpower to 76 agents to maintain a maximum 10% abandon rate

Access Excel Dataset



133



Watch Video Presentation

