



Data Analysis Portfolio

Author: Sheevam Chakraborty

Professional Background

With a Bachelor of Technology in Computer Science and Engineering from the University of Engineering and Management, I have over three years of progressive experience at Tata Consultancy Services. I started as an Application Support Analyst and quickly advanced to the role of System Engineer, where I helped with system design and optimisation. In my most recent position as a BI Developer, I was responsible for leading business intelligence initiatives. I am skilled in a wide range of technologies, including Excel, SQL, Python, Power BI, and Azure services. I specialise in developing and implementing BI solutions, working with cross-functional teams to design optimised data models for actionable insights.

My certification as an Azure Data Fundamentals (DP-900) professional demonstrates my dedication to staying current in the rapidly changing field of data analytics. Aside from my professional endeavours, I have successfully completed personal data analytics projects, demonstrating my skills in data analysis and problem solving.

Outside of work, my love of dogs and habit of immersing myself in audiobooks and paperbacks reflect my diverse interests and dedication to personal development.

Table of Contents

- Data Analysis Process
- Instagram User Analytics
- Metric Spike Analysis and Operational Analytics
- Hiring Process Analytics
- IMDB Movie Analysis
- Bank Loan Case Study
- Analyzing the Impact of Car Features on Price and Profitability
- Call Volume Trend Analysis
- What did I learn from this?
- Reference



1

Data Analytics Process



In this project, we are putting on our Data Analyst caps for the first time here. We are needed to pick a chore or activity from our day to day lives and break it down

As a data analytics process.

This could give us a clear understanding how useful Data Analysis can be and how integral it is to every part of our life.

Problem: Visiting my parents place during the festive season

Plan:

1. Pick the trip dates.
2. Ensure that you have the time off by confirming your work schedule.
3. Take note of any travel arrangements, including scheduling flights or making travel plans.
4. Create a list of the things you must bring.

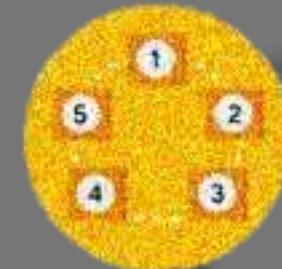


Prepare:

1. Make sure you have all the necessary clothing and personal stuff before you start packing.
2. Wrap and pack any presents you're bringing.
3. If you're driving, get the car ready for the trip.
4. Let your parents know you are coming and confirm the date and time of your visit.

Process:

1. Arrive at your parents' home on time.
2. Have fun and spend time with your family during the festivities.
3. Take part in any customs or activities related to the holiday season.
4. Assist with any necessary cooking or preparations.



Process

Analyze:

1. Consider your visit and the time you spent with your family in reflection.
2. Take into account any difficulties or gains you might have had in your journey.
3. Make a note of any changes or improvements you want to make for the next visit.



Analyze

Share:

1. Tell your friends and various family members about your experiences and images.
2. Thank your parents for having you stay with them during the holiday season.
3. Discuss on any upcoming vacation or family gathering plans.



Act:

1. Take action on any suggestions or enhancements detected during the analysis stage.
2. Set up plans for your upcoming trip or any other actions for retaining a good relation with your family





2

Instagram User Analytics



In this project, an Instagram data analyst is tasked with leveraging SQL and MySQL Workbench to analyze user interactions and engagement. The objective is to offer useful data that can direct choices for business growth, such as promotions, the development of new products, and improvements to the user experience. The project's importance lies in its capacity to shape Instagram's future, one of the largest networks of social media worldwide.

A. Marketing Analysis:

➤ Loyal User Reward:

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

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

```
1
2  #Identify the five oldest users on Instagram from the provided database.
3
4 • SELECT id, username
5    FROM ig_clone.users
6    ORDER BY created_at
7    LIMIT 5;
```

Result Grid | Filter Rows: [] | Edit: [] | Export/Import: [] | Wrap Cell Content: [] | Fetch rows: []

id	username
80	Darby_Herzog
67	Emilio_Bernier52
63	Elenor88
95	Nicole71
38	Jordyn.Jacobson2
NULL	NULL

➤ Inactive User Engagement:

The team wants to encourage inactive users to start posting by sending them promotional emails.

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

```
1
2  #Identify users who have never posted a single photo on Instagram.
3
4 • SELECT DISTINCT(id), username
5    FROM users
6    WHERE id NOT IN (
7      SELECT DISTINCT(user_id) FROM photos
8    );
```

Result Grid | Filter Rows: [] | Edit: [] | Export: [] | Wrap Cell Content: []

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

➤ Contest Winner Declaration:

The team has organized a contest where the user with the most likes on a single photo win. Your Task: Determine the winner of the contest and provide their details to the team

```
1
2  /*The team has organized a contest where the user with the most likes on a single photo wins.
3   Determine the winner of the contest and provide their details to the team*/
4
5 •  SELECT p.user_id, l.photo_id, count(l.user_id) AS total_likes
6     FROM likes l
7     JOIN photos p ON l.photo_id = p.id
8     GROUP BY l.photo_id
9     ORDER BY total_likes DESC
10    LIMIT 1;
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: [] | Fetch rows: []

user_id	photo_id	total_likes
52	145	48

➤ Hashtag Research:

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

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

```
1
2  #Identify and suggest the top five most commonly used hashtags on the platform
3
4 •  SELECT t.tag_name, count(p.photo_id) as total_tags
5     FROM photo_tags p
6     JOIN tags t ON
7         p.tag_id = t.id
8     GROUP BY p.tag_id
9     ORDER BY total_tags DESC
10    LIMIT 5;
```

Result Grid | Filter Rows: [] | Export: [] | Wrap Cell Content: [] | Fetch rows: []

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

➤ 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.

```
1      #Determine the day of the week when most users register on Instagram
2
3
4 •  SELECT DAYNAME(CREATED_AT) AS registered_on , COUNT(id) AS registered_users
5     FROM USERS
6     GROUP BY registered_on
7     ORDER BY registered_users DESC
8     LIMIT 1;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows: |

registered_on	registered_users
Thursday	16

B) Investor Metrics:

➤ 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.

```
1      #The total number of photos on Instagram divided by the total number of users.
2
3
4 •  SELECT (
5       SELECT count(*) FROM photos) /
6       (SELECT count(*) FROM users)
7       AS avg_of_photos_by_users;
8
9      #The average number of posts per user on Instagram
10
11 •  SELECT AVG(posts) AS avg_posts_per_user
12     FROM (
13       SELECT user_id, count(id) AS posts
14         FROM photos
15        GROUP BY user_id
16        ORDER BY posts DESC) a;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows: |

avg_posts_per_user
3.4730

➤ Bots & Fake Accounts:

The 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.

```
1
2      #Identify users (potential bots) who have liked every single photo on the site
3
4 •  SELECT u.id, u.username, count(*) AS posts_liked
5      FROM users u
6      JOIN likes l
7          ON u.id = l.user_id
8      GROUP BY l.user_id
9      HAVING posts_liked = (SELECT count(*) FROM photos);
```

Result Grid | Filter Rows: Export: Wrap Cell Content:

	id	username	posts_liked
▶	5	Aniya_Hackett	257
	14	Jadyn81	257
	21	Rodo33	257
	24	Maxwell.Halvorson	257
	36	Ollie_Ledner37	257
	41	Mckenna17	257
	54	Duane60	257
	57	Julien_Schmidt	257
	66	Mike.Auer39	257
	71	Nia_Haag	257
	75	Leslie67	257
	76	Janelle.Nikolaus81	257
	91	Bethany20	257



3

Metric Spike Analysis and Operational Analytics

At an organisation that resembles Microsoft, the lead data analyst is in charge of operational analytics. Working together across teams, the goal is to use advanced SQL skills to dig into sudden metric spikes. Finding new information is only one goal; the other is to improve operational effectiveness and understanding of unanticipated changes in metrics. The project highlights how important data analytics is for directing improvements in organisation and well-informed decision-making.

Case Study 1: Job Data Analysis

You will be working with a table named `job_data` with the following columns:

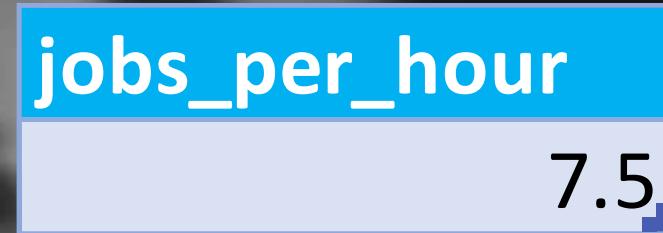
- **job_id**: Unique identifier of jobs
- **actor_id**: Unique identifier of actor
- **event**: The type of event (decision/skip/transfer).
- **language**: The Language of the content
- **time_spent**: Time spent to review the job in seconds.
- **org**: The Organization of the actor
- **ds**: The date in the format yyyy/mm/dd (stored as text).

```
• CREATE TABLE job_data (
    ds VARCHAR(100),
    job_id INT,
    actor_id INT,
    event VARCHAR(100),
    language VARCHAR(100),
    time_spent INT,
    org VARCHAR(50)
);

• LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/Uploads/job_data.csv"
  INTO TABLE job_data
  FIELDS TERMINATED BY ','
  ENCLOSED BY ""
  LINES TERMINATED BY '\n'
  IGNORE 1 ROWS;
```

A. Jobs Reviewed Over Time:

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



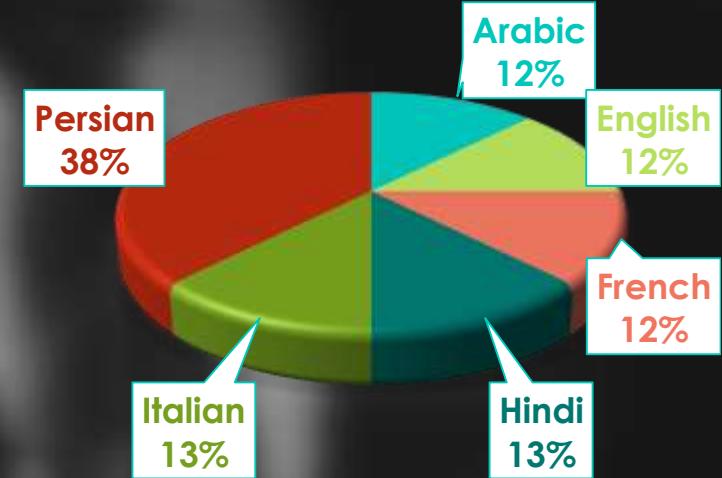
B. Throughput Analysis:

- Objective: Calculate the 7-day rolling average of throughput (number of events per second).
 - Your Task: Write an SQL query to calculate the 7-day rolling average of throughput.
- Additionally, explain whether you prefer using the daily metric or the 7-day rolling average for throughput, and why.



C. Language Share Analysis:

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



D. Duplicate Rows Detection:

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

job_id	actor_id	event	language	time_spent	org	ds	row_num
1	1	1	1	1	1	1	1

“Since there are no duplicates
In our dataset we get none”

Case Study 2: Investigating Metric Spike

You will be working with three tables:

- **users**: Contains one row per user, with descriptive information about that user's account.
- **events**: Contains one row per event, where an event is an action that a user has taken (e.g., login, messaging, search).
- **email_events**: Contains events specific to the sending of emails.

events

user_id	event_type	event_name	location	device	user_type	occured_at
10522	engagement	login	Japan	dell inspiron notebook	3	2014-05-02 11:02:00
10522	engagement	home_page	Japan	dell inspiron notebook	3	2014-05-02 11:02:00
10522	engagement	like_message	Japan	dell inspiron notebook	3	2014-05-02 11:03:00
10522	engagement	view_inbox	Japan	dell inspiron notebook	3	2014-05-02 11:04:00
10522	engagement	search_run	Japan	dell inspiron notebook	3	2014-05-02 11:03:00
10522	engagement	search_run	Japan	dell inspiron notebook	3	2014-05-02 11:03:00
10612	engagement	login	Netherlands	iphone 5	1	2014-05-01 09:59:00
10612	engagement	like_message	Netherlands	iphone 5	1	2014-05-01 10:00:00

email_events

user_id	action	user_type	occured_at
0	sent_weekly_digest	1	2014-05-06 09:30:00
0	sent_weekly_digest	1	2014-05-13 09:30:00
0	sent_weekly_digest	1	2014-05-20 09:30:00
0	sent_weekly_digest	1	2014-05-27 09:30:00
0	sent_weekly_digest	1	2014-06-03 09:30:00
0	email_open	1	2014-06-03 09:30:00
0	sent_weekly_digest	1	2014-06-10 09:30:00
0	email_open	1	2014-06-10 09:30:00

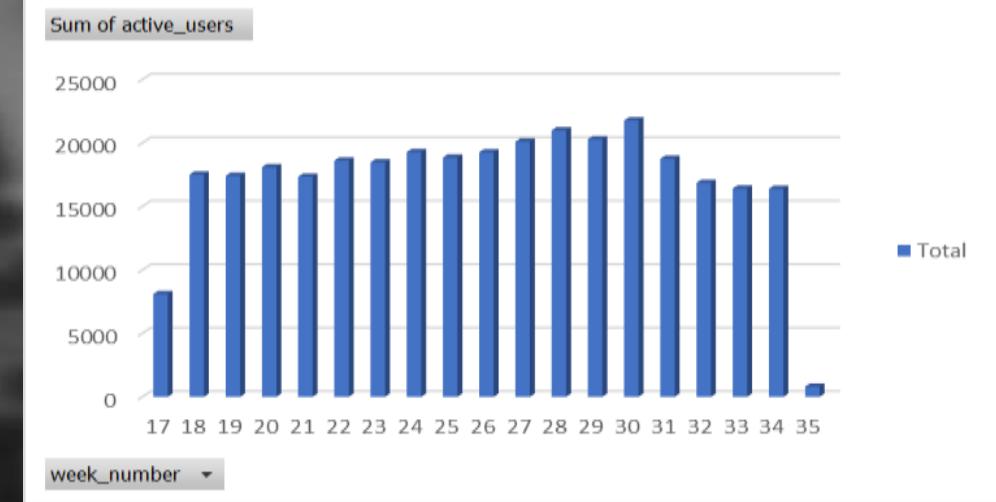
users

user_id	company_id	language	state	created_at	activated_at
0	5737	english	active	2013-01-01 20:59:00	2013-01-01 21:01:00
3	2800	german	active	2013-01-01 18:40:00	2013-01-01 18:42:00
4	5110	indian	active	2013-01-01 14:37:00	2013-01-01 14:39:00
6	11699	english	active	2013-01-01 18:37:00	2013-01-01 18:38:00
7	4765	french	active	2013-01-01 16:19:00	2013-01-01 16:20:00
8	2698	french	active	2013-01-01 04:38:00	2013-01-01 04:40:00
11	3745	english	active	2013-01-01 08:07:00	2013-01-01 08:09:00
13	4025	english	active	2013-01-02 12:27:00	2013-01-02 12:29:00

A. Weekly User Engagement:

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

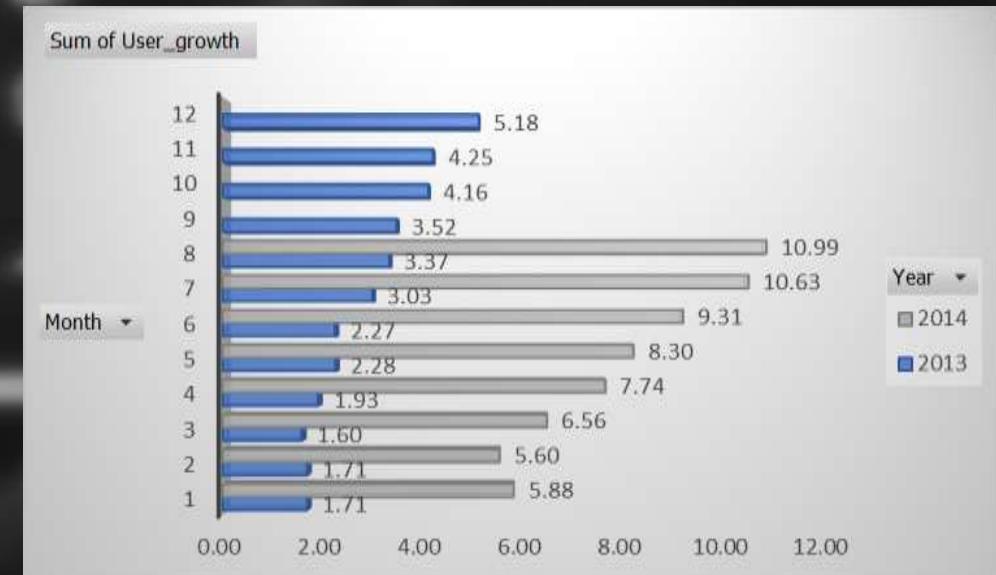
“Most user activity was recorded during the weeks 28 - 30 in 2014”



B. User Growth Analysis:

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

“We can see exponential user growth in 2014 than 2013. The best months being July and August”



C. Weekly Retention Analysis:

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

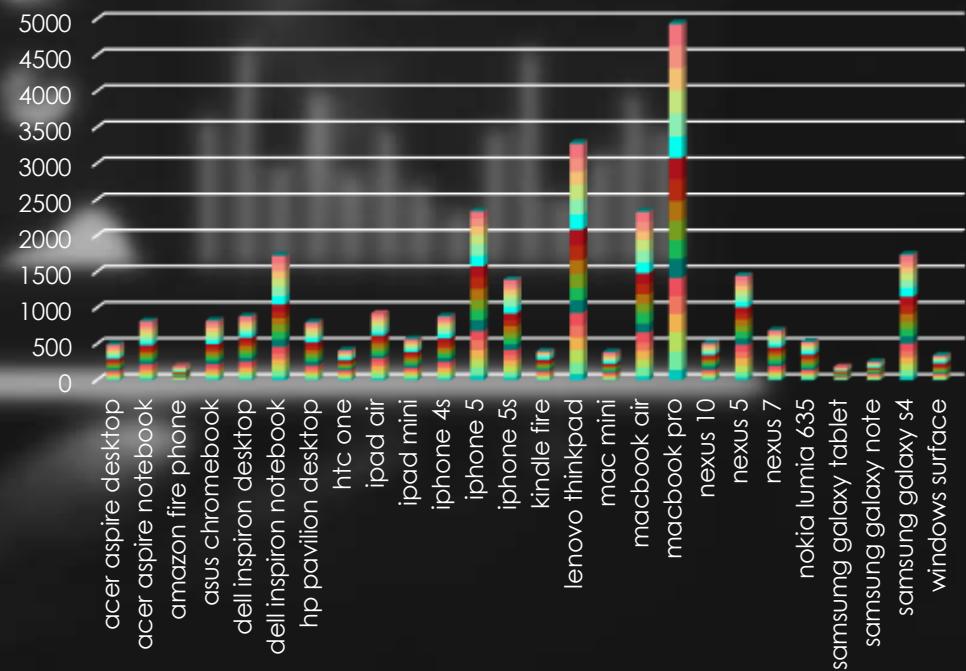
“There’s a massive dip in user retention between week 33 and 36.”

D. Weekly Engagement Per Device:

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

“MacBook pro has the highest user engagement per week (4900+)”

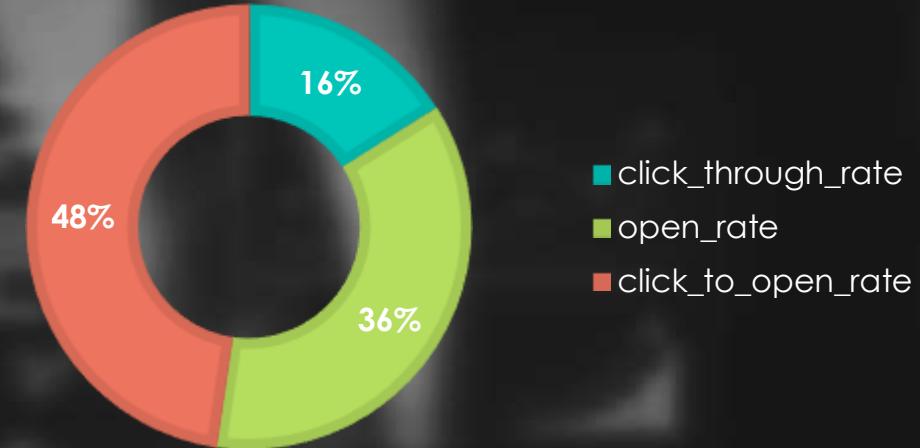
Weekly user retention



E. Email Engagement Analysis:

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

“The click to open ratio among users
Is at a record high value”



4

Hiring Process Analytics



My task as a data analyst at a global corporation like Google is to analyze hiring process data to extract insightful information. Understanding trends such as job types, vacancies, interviews, and rejections is the main goal.

Equipped with a dataset of previous recruits, my job is to address relevant inquiries meant to improve the hiring procedure and general productivity of the business.

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

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

Males Hired:

- 4083

Females Hired:

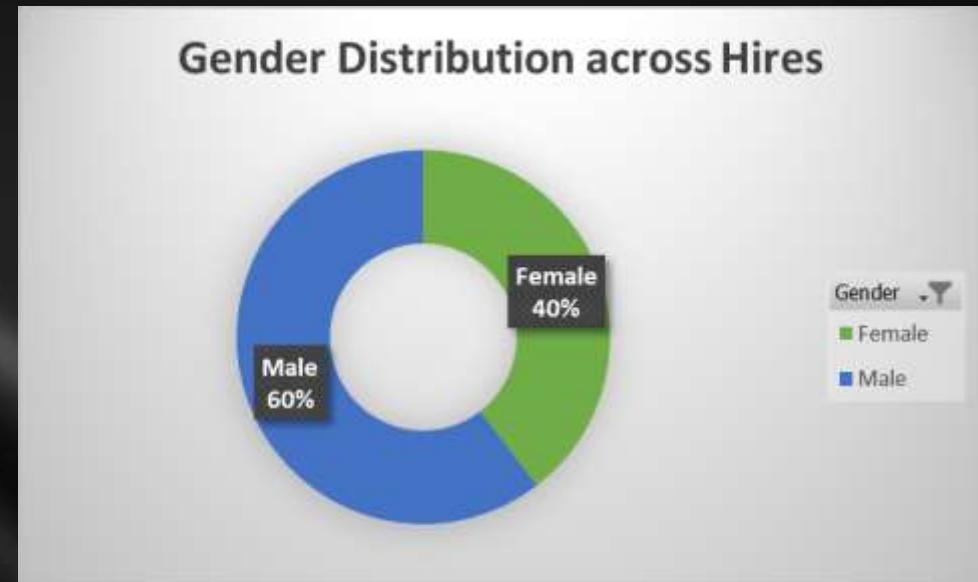
- 2675

B. Salary Analysis: The average salary is calculated by adding up the salaries of a group of employees and then dividing the total by the number of employees.

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

Average Salary Offered:

- 49986.07



Average Salary	49986.07
Median Salary	49625.00
Mode of Salary	72843.00

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

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

Most of the employees are clustered around the salary range of:

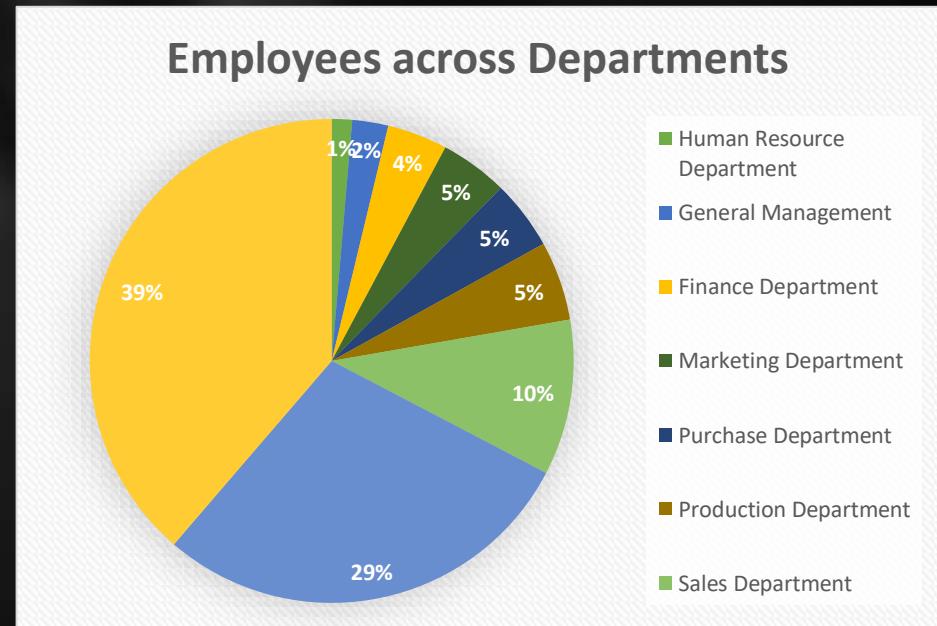
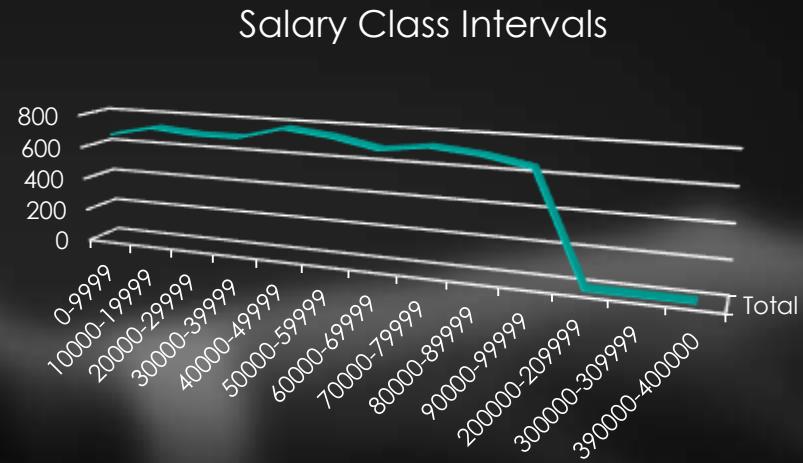
- 40,000 – 60,000

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.

Majority of Employees work in Operations Department:

- 39%

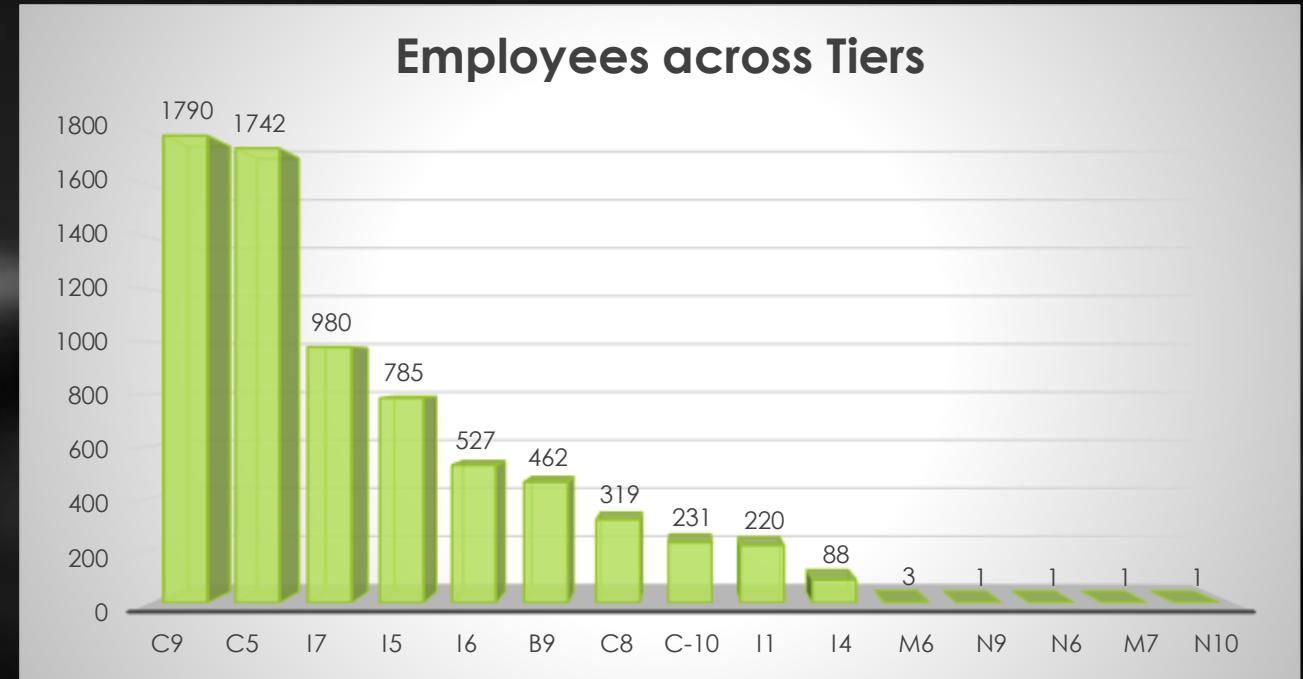


E. Position Tier Analysis: Different positions within a company often have different tiers or levels.

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.

Major tiers housing the maximum no. of employees:

- C9
- C5



5

IMDB Movie Analysis

The aim is to examine the IMDB Movies dataset in order to pinpoint important elements that contribute to a film's success, which is determined by its high IMDB rating. The goal of this study is to give investors, producers, and directors of motion pictures insightful information that will enable them to make better decisions and increase the probability that their next projects will succeed.

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

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

Comedy has the highest score among genres

- 9.5

The most common genre with highest number of movies is Drama

- 2506

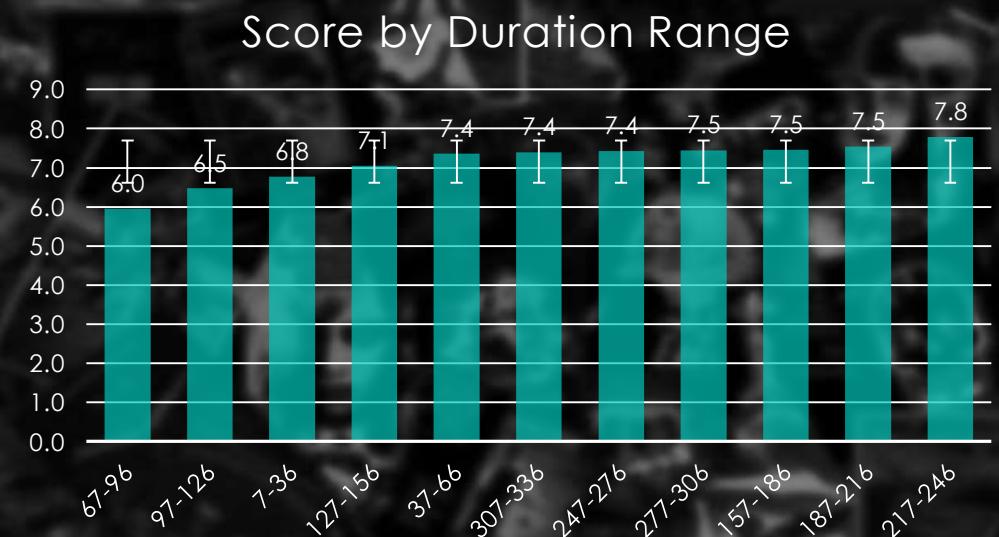
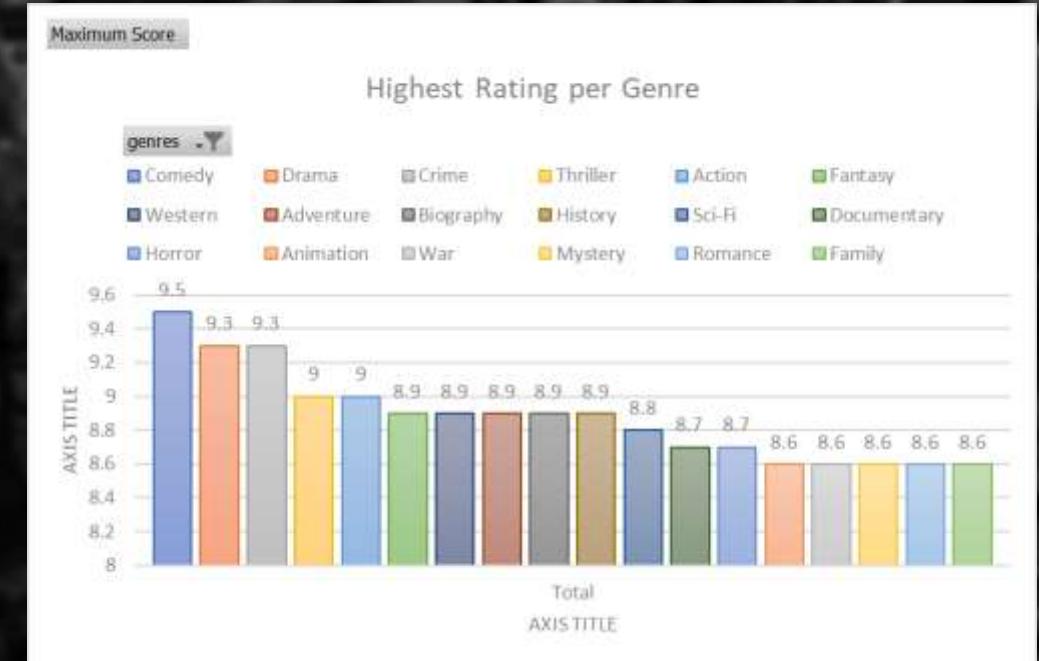
B. Movie Duration Analysis: Analyze the distribution of movie durations and its impact on the IMDB score.

Long movies have the highest ratings among viewers

- 7.8

Majority of the viewers enjoy movies with duration 110 – 140 minutes overall.

- Linear Trendline supports this



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

English is the most common language in movies

- 4593

More and more viewers are enjoying foreign language films in huge quantity

- Top ratings belong to foreign films

RATINGS BY LANGUAGE



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

Martin S., Steven S. and Robert Z. have the highest percentile ranks in directors

- 93%, 89%, 86%

Viewers are more keen to watch a movie directed by the top directors every time.

- Top ratings belong to the top directors

TOP DIRECTORS



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

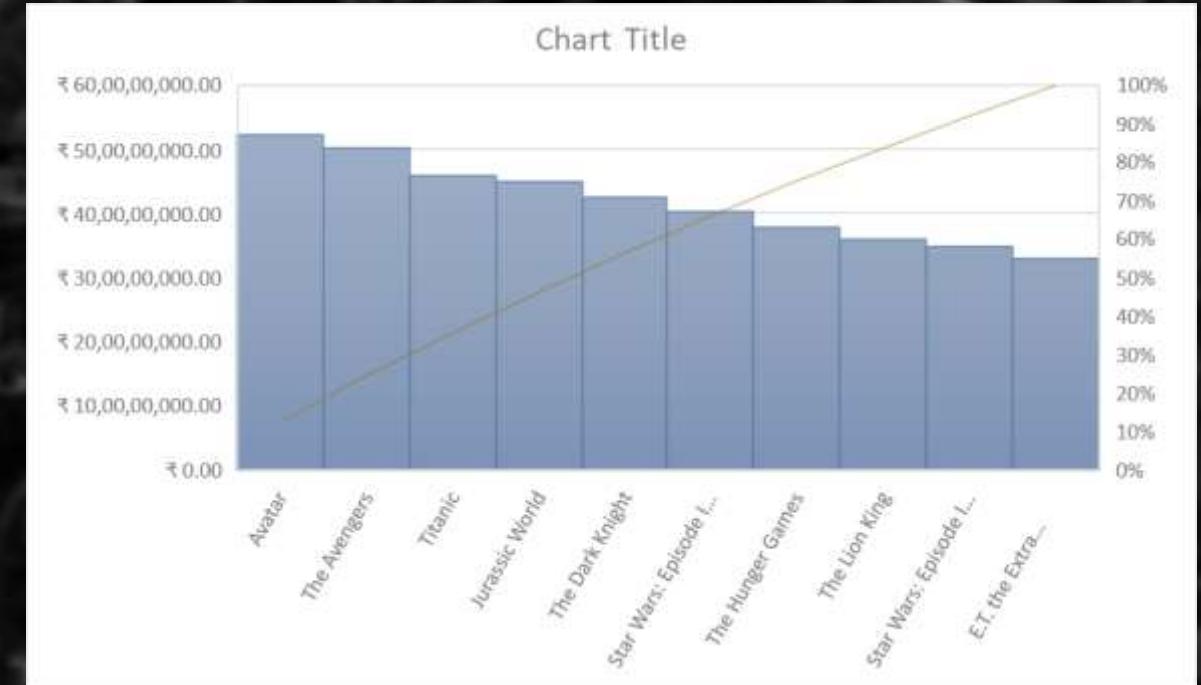
Task: Analyse the correlation between movie budgets and gross earnings, and identify the movies with the highest profit margin.

There is positive correlation between budget and gross earnings as correlation coefficient value is

- 0.0995482

Movie with maximum profit margin is Avatar, with profit ranging over

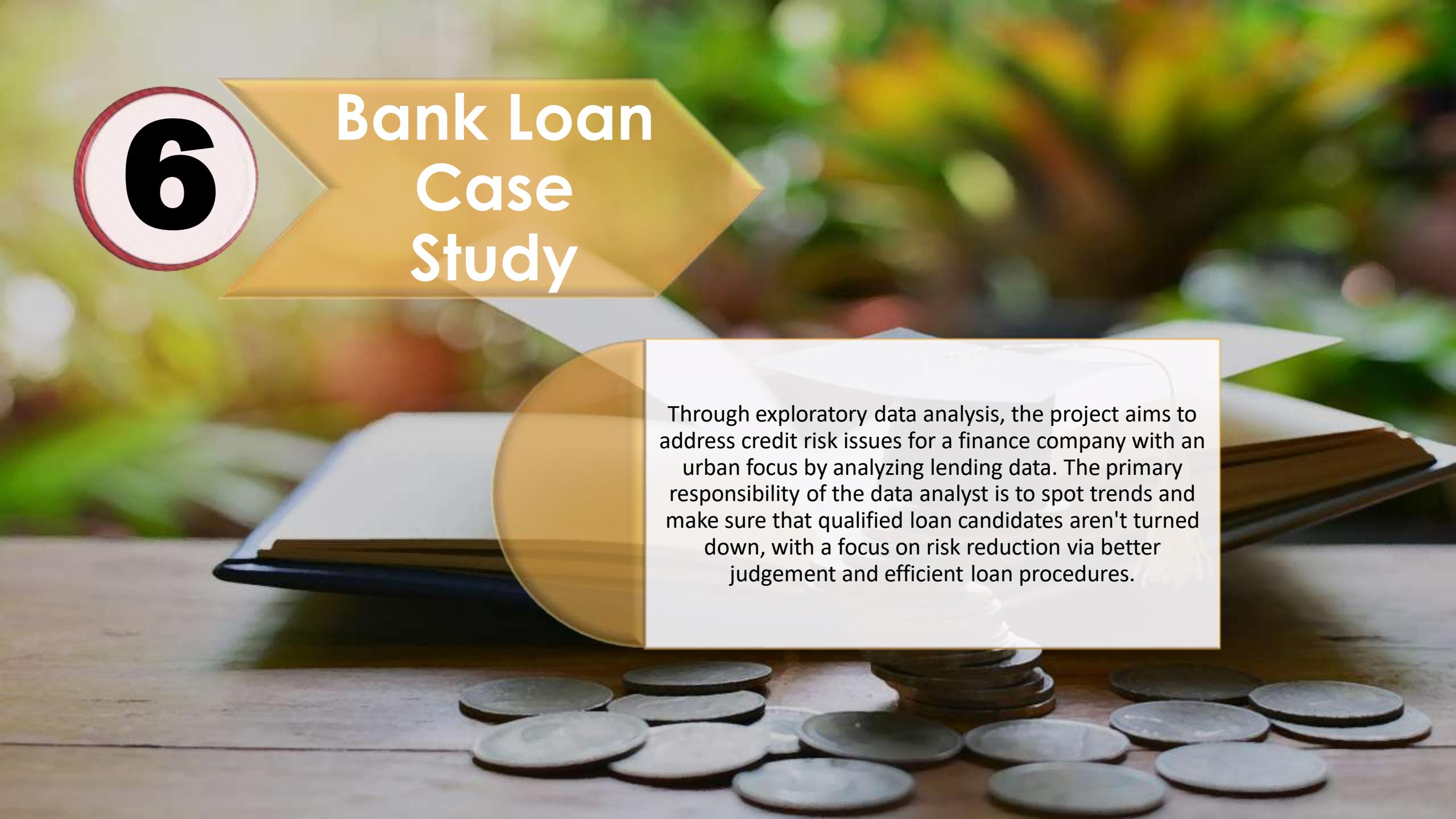
- 52 crore rupees





6

Bank Loan Case Study



Through exploratory data analysis, the project aims to address credit risk issues for a finance company with an urban focus by analyzing lending data. The primary responsibility of the data analyst is to spot trends and make sure that qualified loan candidates aren't turned down, with a focus on risk reduction via better judgement and efficient loan procedures.

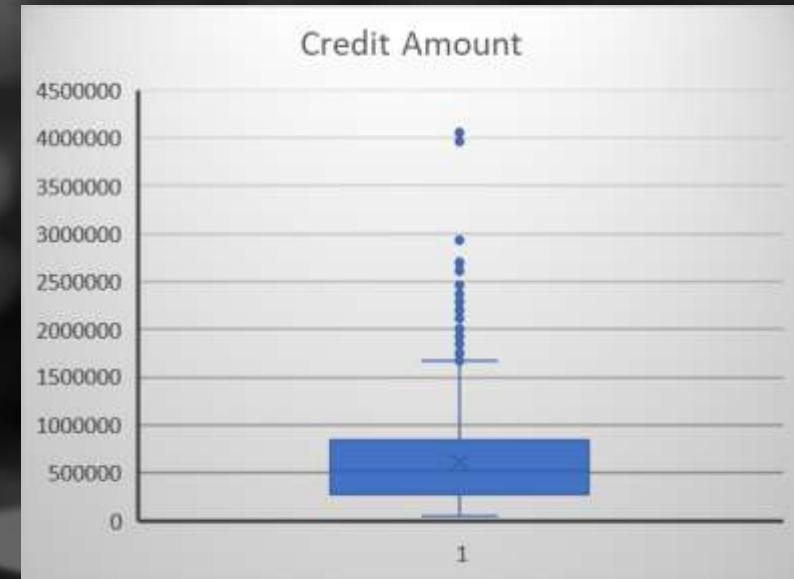
A. Identify Missing Data and Deal with it Appropriately:

As a data analyst, you come across missing data in the loan application dataset. It is essential to handle missing data effectively to ensure the accuracy of the analysis.

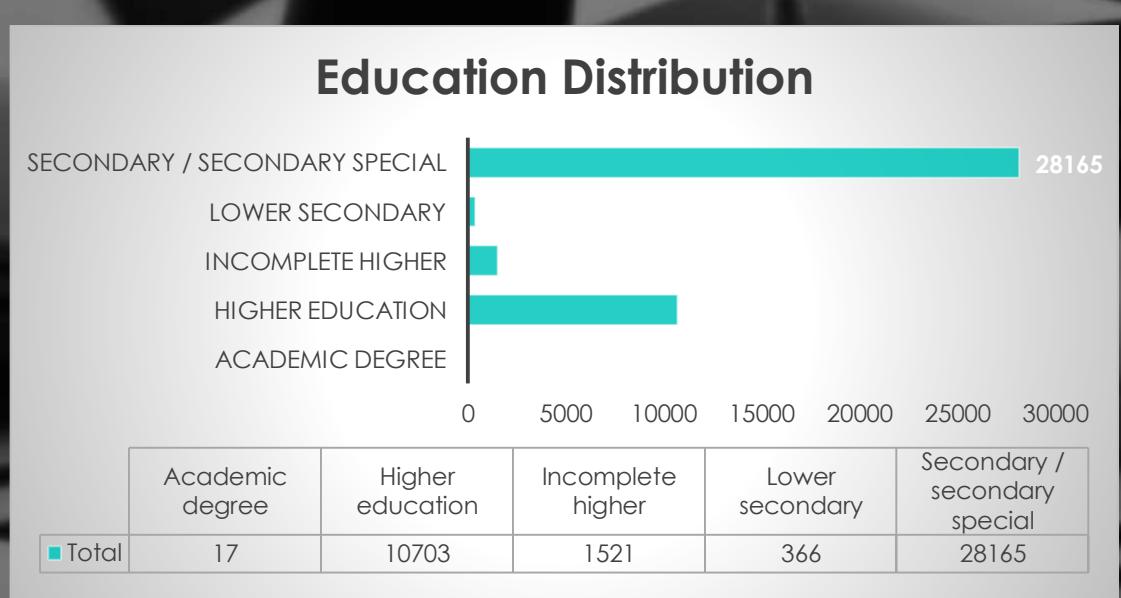
List of deleted columns from Application workbook

OWN_CAR_AGE FLAG_MOBIL FLAG_EMP_PHONE FLAG_WORK_PHONE FLAG_CONT_MOBILE OCCUPATION_TYPE REGION_RATING_CLIENT_W_CITY
WEEKDAY_APPR_PROCESS_START REG_REGION_NOT_LIVE_REGION REG_REGION_NOT_WORK_REGION LIVE_REGION_NOT_WORK_REGION
REG_CITY_NOT_LIVE_CITY REG_CITY_NOT_WORK_CITY LIVE_CITY_NOT_WORK_CITY EXT_SOURCE_1 EXT_SOURCE_2 EXT_SOURCE_3 APARTMENTS_AVG
BASEMENTAREA_AVG YEARS_BEGINEXPLUATATION_AVG YEARS_BUILD_AVG COMMONAREA_AVG ELEVATORS_AVG ENTRANCES_AVG FLOORSMAX_AVG
FLOORSMIN_AVG LANDAREA_AVG LIVINGAPARTMENTS_AVG LIVINGAREA_AVG NONLIVINGAPARTMENTS_AVG NONLIVINGAREA_AVG APARTMENTS_MODE
BASEMENTAREA_MODE YEARS_BEGINEXPLUATATION_MODE YEARS_BUILD_MODE COMMONAREA_MODE ELEVATORS_MODE ENTRANCES_MODE
FLOORSMAX_MODE FLOORSMIN_MODE LANDAREA_MODE LIVINGAPARTMENTS_MODE LIVINGAREA_MODE NONLIVINGAPARTMENTS_MODE
NONLIVINGAREA_MODE APARTMENTS_MEDI BASEMENTAREA_MEDI YEARS_BEGINEXPLUATATION_MEDI YEARS_BUILD_MEDI COMMONAREA_MEDI
ELEVATORS_MEDI ENTRANCES_MEDI FLOORSMAX_MEDI FLOORSMIN_MEDI LANDAREA_MEDI LIVINGAPARTMENTS_MEDI LIVINGAREA_MEDI
NONLIVINGAPARTMENTS_MEDI NONLIVINGAREA_MEDI FONDKAPREMONT_MODE HOUSETYPE_MODE TOTALAREA_MODE WALLSMATERIAL_MODE
EMERGENCYSTATE_MODE FLAG_DOCUMENT_2 FLAG_DOCUMENT_3 FLAG_DOCUMENT_4 FLAG_DOCUMENT_5 FLAG_DOCUMENT_6 FLAG_DOCUMENT_7
FLAG_DOCUMENT_8 FLAG_DOCUMENT_9 FLAG_DOCUMENT_10 FLAG_DOCUMENT_11 FLAG_DOCUMENT_12 FLAG_DOCUMENT_13 FLAG_DOCUMENT_14
FLAG_DOCUMENT_15 FLAG_DOCUMENT_16 FLAG_DOCUMENT_17 FLAG_DOCUMENT_18 FLAG_DOCUMENT_19 FLAG_DOCUMENT_20 FLAG_DOCUMENT_21
AMT_REQ_CREDIT_BUREAU_HOUR AMT_REQ_CREDIT_BUREAU_DAY AMT_REQ_CREDIT_BUREAU_WEEK AMT_REQ_CREDIT_BUREAU_MON
AMT_REQ_CREDIT_BUREAU_QRT AMT_REQ_CREDIT_BUREAU_YEAR

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



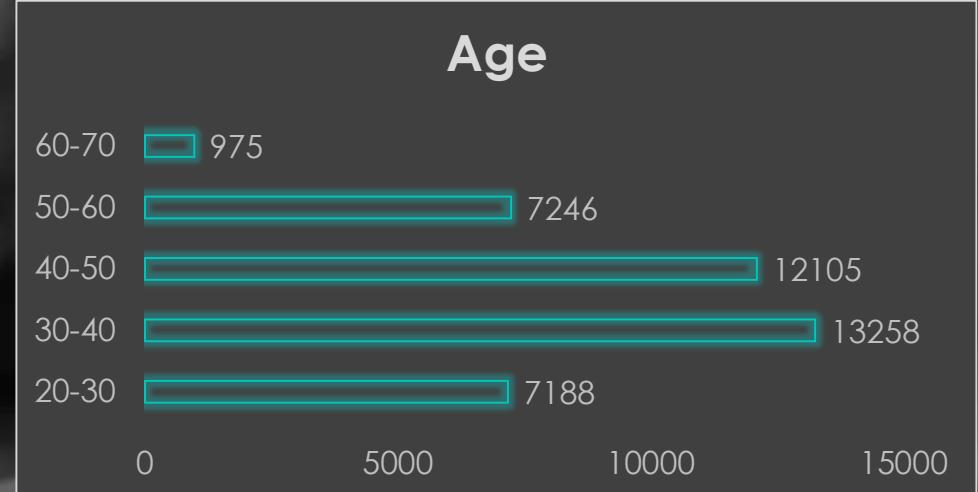
C. Analyze Data Imbalance: Data imbalance can affect the accuracy of the analysis, especially for binary classification problems. Understanding the data distribution is crucial for building reliable models.



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

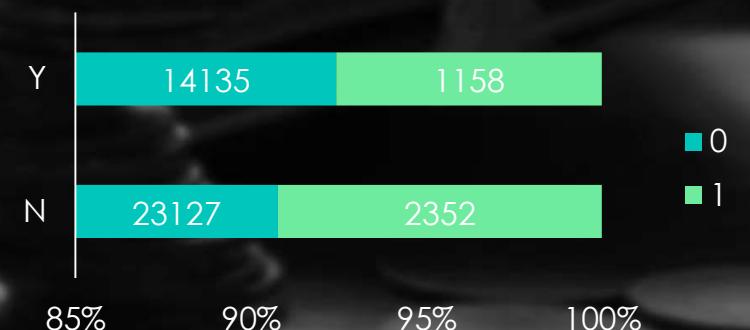
Univariate Analysis

	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE
<i>mean</i>	0.50	178473.04	612778.02	27877.52	550201.93
<i>median</i>	0.00	157500.00	521280.00	25960.50	450000.00
<i>mode</i>	0.00	135000.00	450000.00	9000.00	450000.00
<i>range</i>	11.00	116974350.00	4005000.00	255973.50	4005000.00
<i>std dev</i>	0.77	587451.67	405812.74	14742.28	373054.26
<i>variance</i>	0.59	345099464241.57	164683983341.20	217334961.49	139169479662.83



Bivariate Analysis

Target by Car owner



E. Identify Top Correlations for Different Scenarios:

Understanding the correlation between variables and the target variable can provide insights into strong indicators of loan default.

Task: Segment the dataset based on different scenarios (e.g., clients with payment difficulties and all other cases) and identify the top correlations for each segmented data using Excel functions.

Correlation Matrix with Heat Map

<i>Target 1 (DEFaulter)</i>	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELATIVE	Age	CNT_FAM_MEMBERS	REGION_RATING_CLIENT	DEF_30_CNT_SOCIAL_CIRCLE	DEF_60_CNT_SOCIAL_CIRCLE
CNT_CHILDREN	1										
AMT_INCOME_TOTAL	-0.030411677	1									
AMT_CREDIT	0.009645704	0.311445675	1								
AMT_ANNUITY	0.01123647	0.3713188	0.745055477	1							
AMT_GOODS_PRICE	0.002144239	0.313428484	0.982120427	0.746788213	1						
REGION_POPULATION_RELATIVE	-0.018820739	0.095908597	0.051861792	0.061329275	0.058418661	1					
Age	-0.162559089	0.089531201	0.197146316	0.088162963	0.19060081	0.015390652	1				
CNT_FAM_MEMBERS	0.900664182	-0.02214905	0.052652036	0.049491281	0.048115413	-0.020944955	-0.104287027	1			
REGION_RATING_CLIENT	0.061990007	-0.160207729	-0.046143285	-0.063209658	-0.052516777	-0.434272694	-0.052665197	0.064942398	1		
DEF_30_CNT_SOCIAL_CIRCLE	-0.006996512	-0.046072367	-0.037927318	-0.037351709	-0.032484936	0.016604761	0.001350976	-0.001395323	0.01541248	1	
DEF_60_CNT_SOCIAL_CIRCLE	-0.012498616	-0.033596939	-0.041024916	-0.041522425	-0.032871852	0.015001485	0.009293982	-0.004047456	9.49418E-05	0.891433576	1

7

Analyzing the Impact of Car Features on Price and Profitability

In the automotive industry, a data analyst's mission is to optimize profits through the investigation of relationships among vehicle characteristics, market niches, and expenses. The goal is to identify valuable features and categories among customers by utilizing data analysis techniques such as market segmentation and regression analysis. Strategic pricing is guided by the insights, which strike a balance between customer demand and profitability. Furthermore, by identifying critical product characteristics, the analysis aids in the development of future products and raises competitiveness and long-term profitability in the ever-changing automotive industry.

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

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

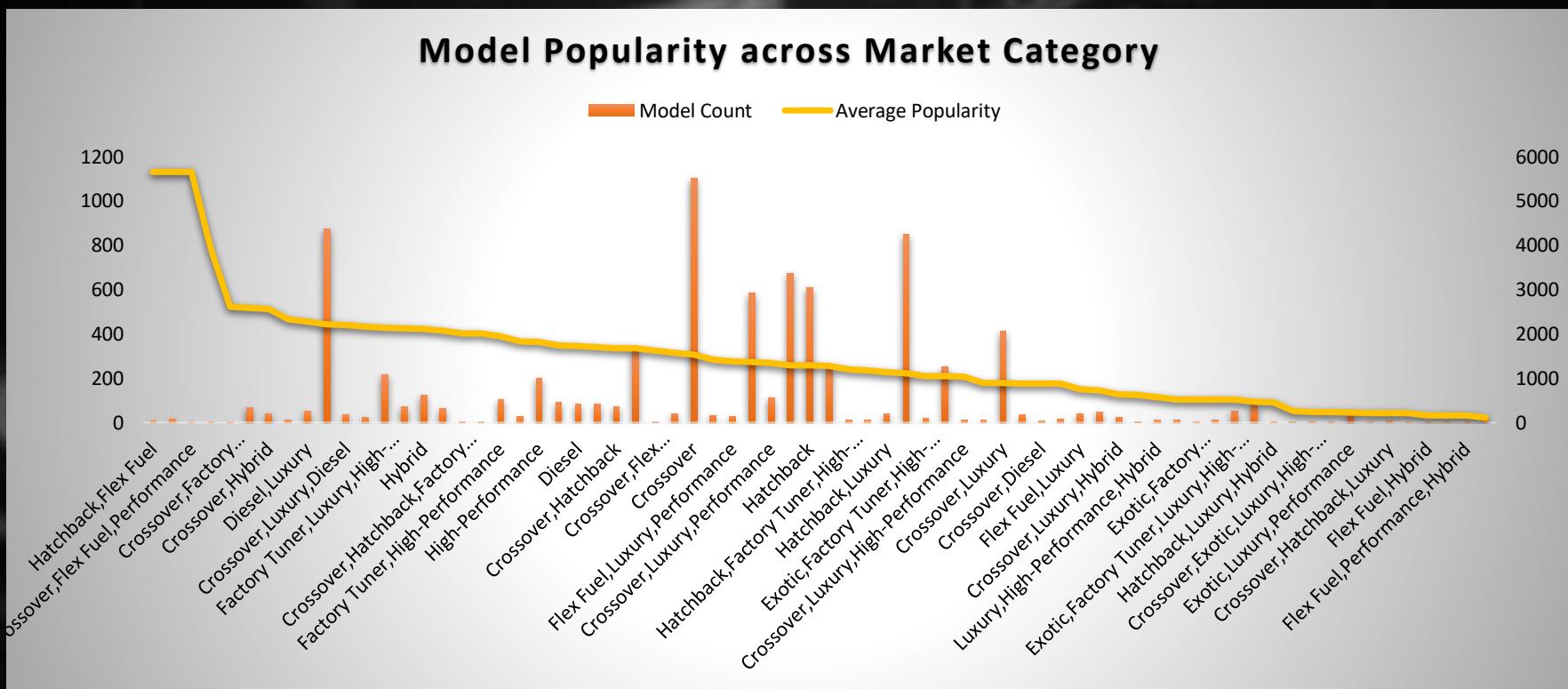
Task 1.B: Create a combo chart that visualizes the relationship between market category and popularity.

Hatchback, Flex Fuel is the most popular car category

- 5657 votes

Crossovers has the most units sold

- 1103 units



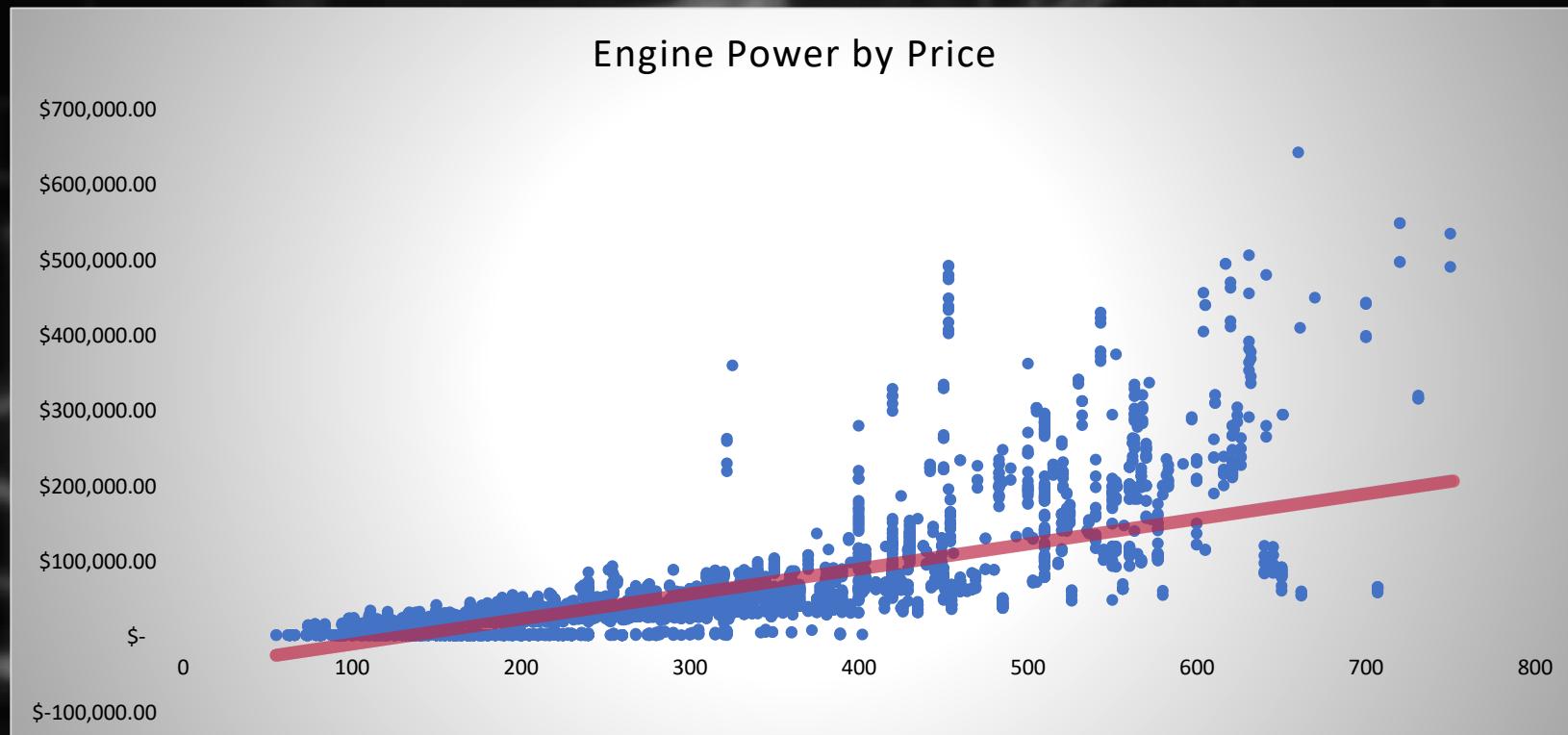
Insight Required: What is the relationship between a car's engine power and its price?

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

Strong positive correlation

- Positive incremental trendline

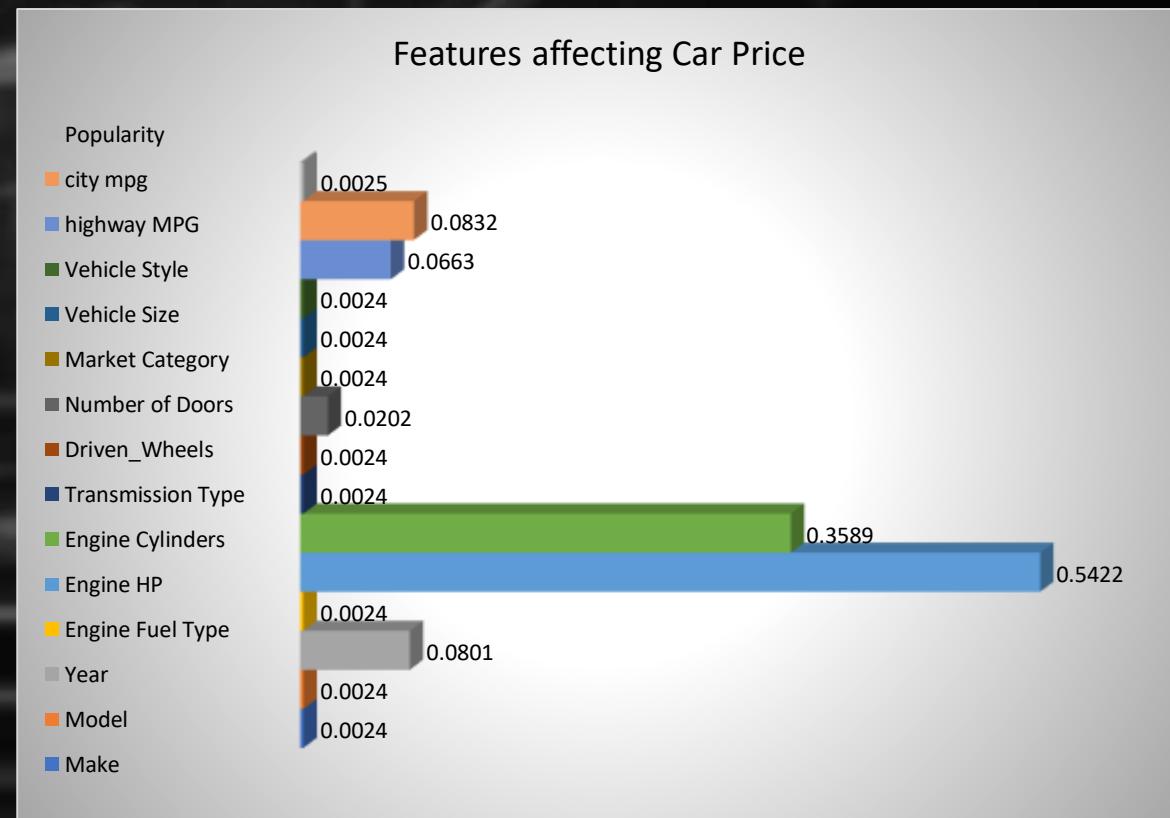
Engine power increases, Car Price increases



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

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

Values	Regression Coefficient
Make	0.0024
Model	0.0024
Year	0.0801
Engine Fuel Type	0.0024
Engine HP	0.5422
Engine Cylinders	0.3589
Transmission Type	0.0024
Driven_Wheels	0.0024
Number of Doors	0.0202
Market Category	0.0024
Vehicle Size	0.0024
Vehicle Style	0.0024
highway MPG	0.0663
city mpg	0.0832
Popularity	0.0025



Engine HP and Cylinders has strongest impact on Car Price

- 0.54, 0.36

Other factors affecting price would be MPG and year of release

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

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

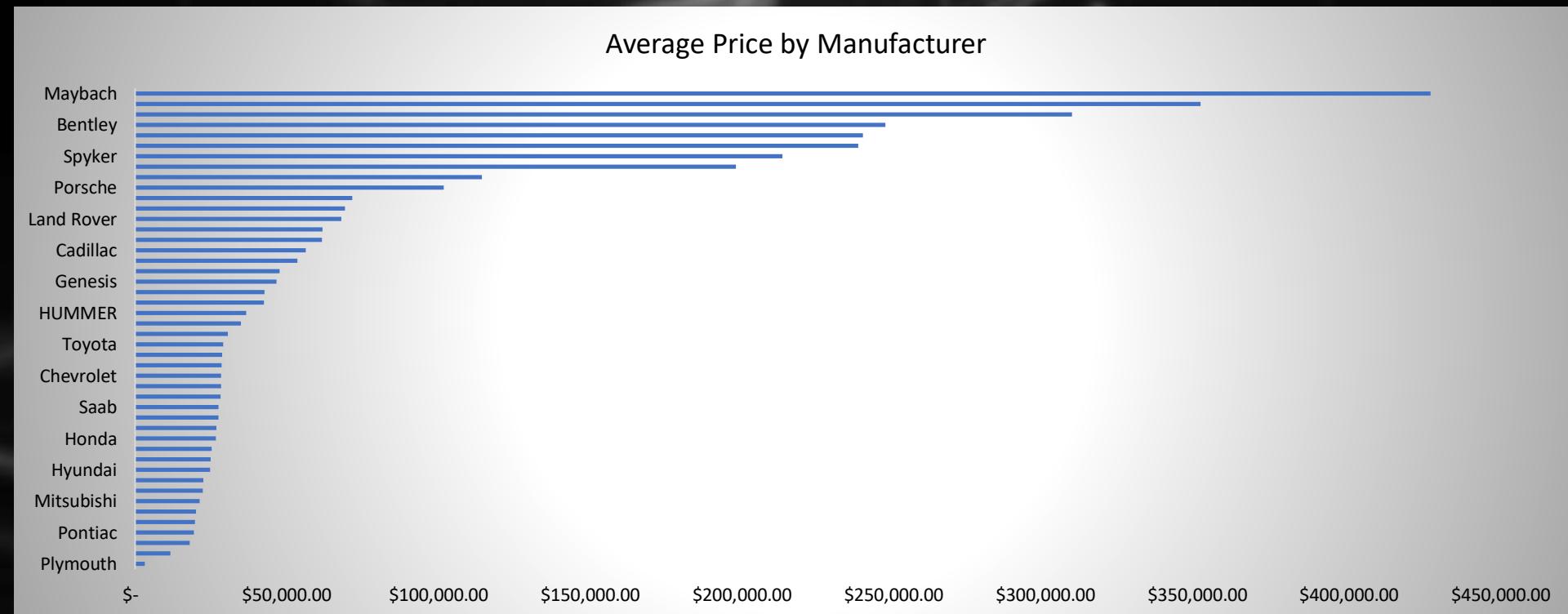
Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between the manufacturer and the average price

Premium manufacturers have higher Car Price

- Maybach, \$ 427k

Everyday use car manufacturers have way cheaper Car Prices

- Plymouth, \$ 3k



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

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

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

Strong negative relation

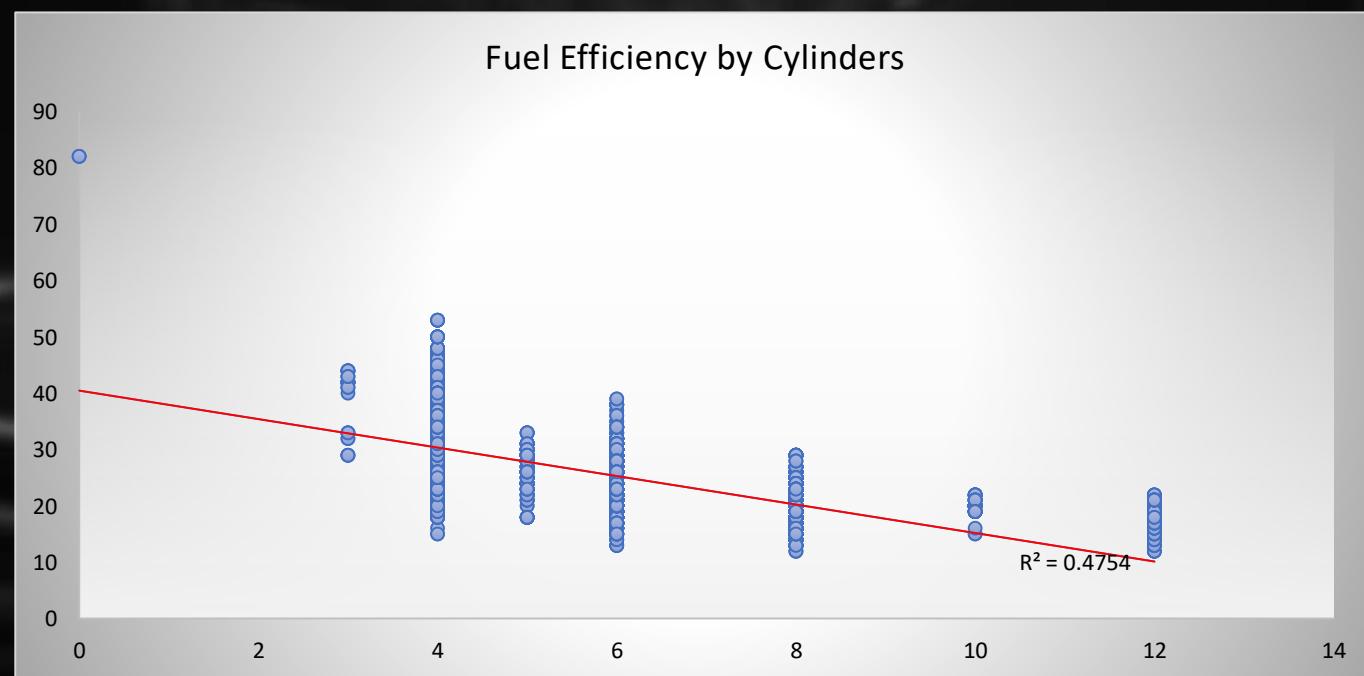
- Negative slope of trendline

Number of Cylinders increases,
Fuel efficiency decreases

- Correlation coefficient strongly negative

Correlation coefficient between Cylinders and Highway MPG

-0.68947



Inferences from Dashboard

○ Power BI Dashboard

- Chevrolet has the major market cap in Total Car Sales, Sedan body style is the most sold vehicle style
- Maybach has the Highest Average MSRP in cars, Plymouth has the Lowest Average MSRP
- Even in case of Average MSRP the Sedan style vehicles have a record high and low price
- Automatic transmission type cars have higher MSRP, Sedan and Cargo Styles record the most units sold
- Average MPG has increased over the years to a record high.
- This trend stands the same for different models from various brands
- There's a strong negative relation between MPG and Engine Power
- There's a strong positive correlation between MSRP and Engine Power
- Therefore we can deduce that as Engine Power Increases, Car Price Increases but the fuel efficiency decreases
- Ferrari Brand tops the charts when it comes to Engine Power whereas Mercedes Benz electric model offers the most MPG

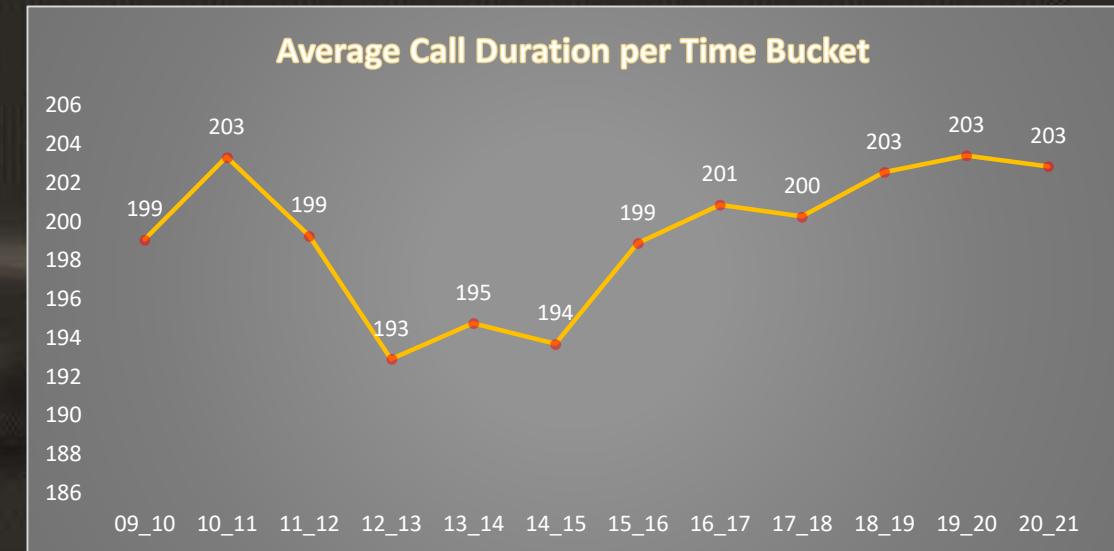
8

Call Volume Trend Analysis

The project focuses on the inbound calling team and examines Customer Experience (CX) analytics over a 23-day period. The CX team, which is critical to organizational success, oversees internal communications, experience programmed, and customer journey mapping. The project emphasizes the critical role of call center agents in AI-powered inbound customer support, with the primary objectives of improving the customer experience, engaging customers, and cultivating brand loyalty.

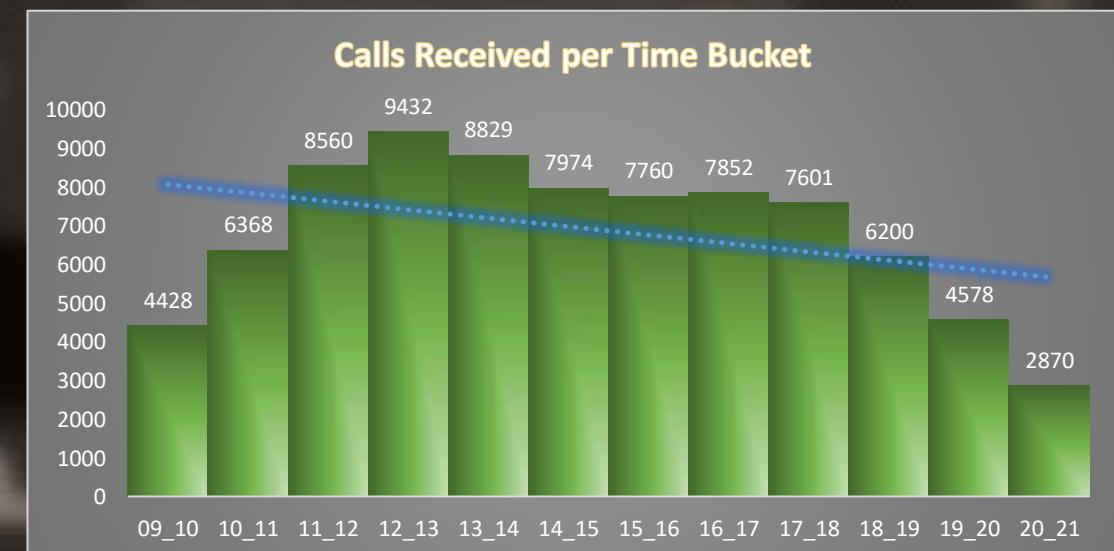
- Average Call Duration: Determine the average duration of all incoming calls received by agents. This should be calculated for each time bucket.

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



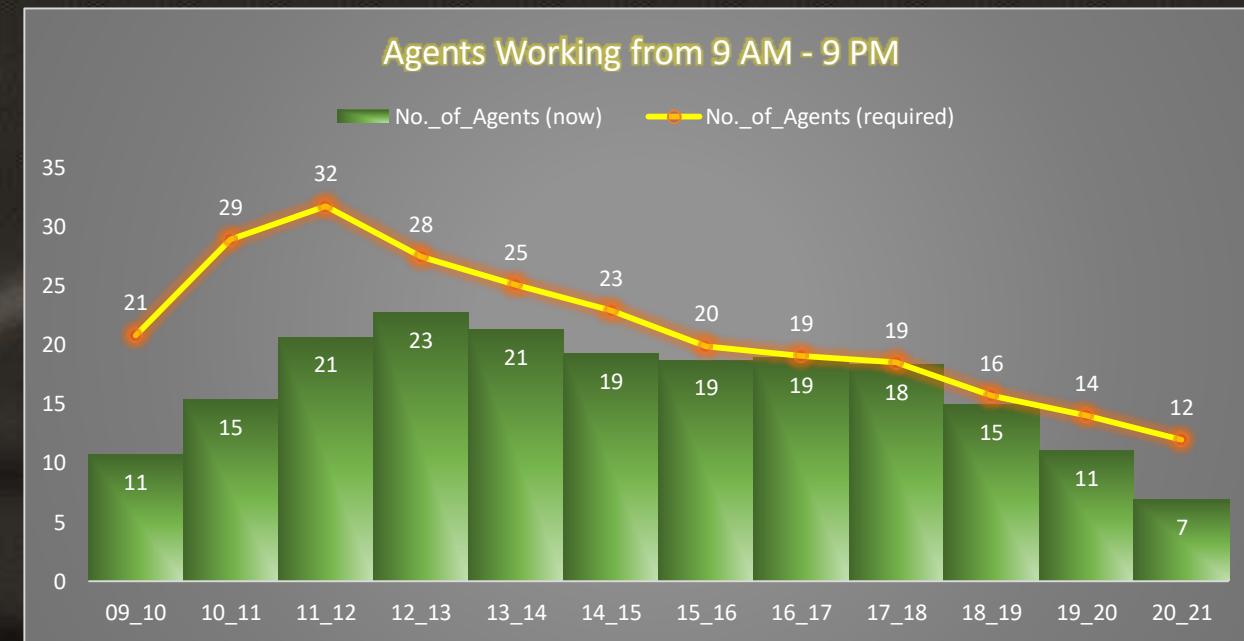
- Call Volume Analysis: Visualize the total number of calls received. This should be represented as a graph or chart showing the number of calls against time. Time should be represented in buckets (e.g., 1-2, 2-3, etc.).

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



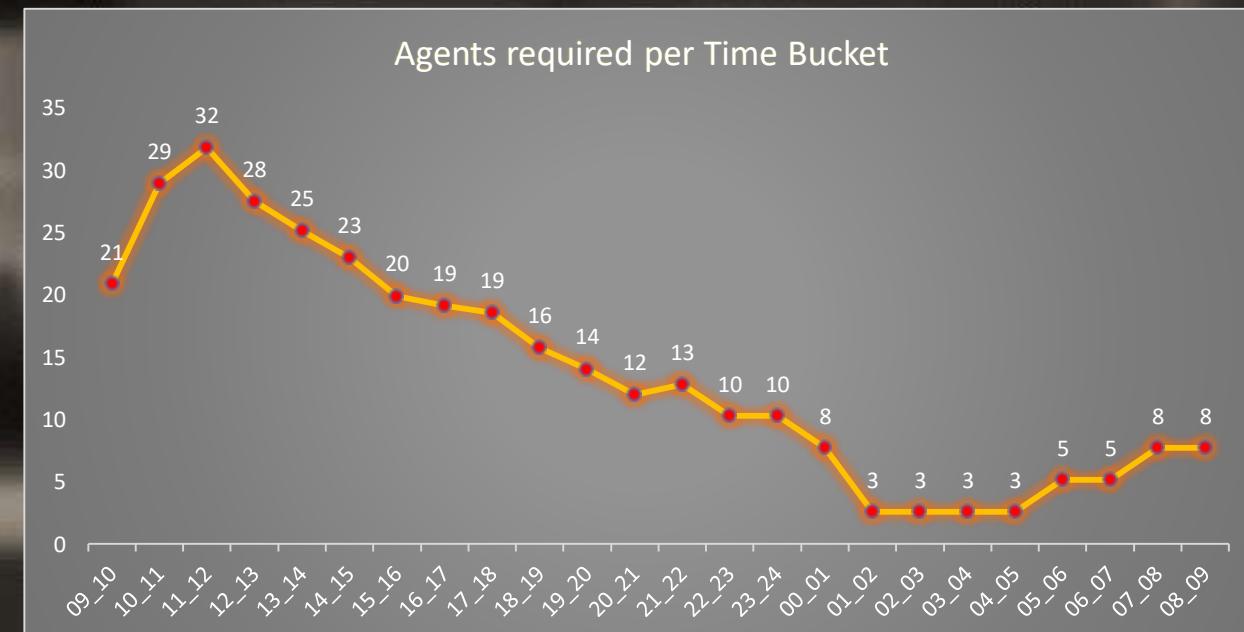
- Manpower Planning: The current rate of abandoned calls is approximately 30%. Propose a plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%. In other words, you need to calculate the minimum number of agents required in each time bucket to ensure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

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



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

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



What did I learn from this?

- I have learnt to think like a Data Analyst in my everyday life
- I have learnt how important a role SQL plays in deriving insights and meanings out of metrics data
- I have learnt about Operational Analytics, how vastly product based companies rely on Metrics/Operational analytics in decision making
- I now have a lot more experience regarding what goes around in the hiring process. Why a certain person gets hired and other doesn't, what are the factors responsible in affecting this, etc.
- For as long as I can remember I have been in love with movies. After performing the movies analysis, I got to know a lot. Things like RCA, what factors steer the audience, why a certain movie in spite of odds does great at the theatres

- Bank Loans project was one of the most challenging projects in this series, mostly due to the immense data that was available. I spent a lot of time cleaning the data, performing imputations where necessary, discarding unwanted data, identifying outliers. I learnt about correlation, univariate and bivariate analysis techniques
- Car Features and Price Analysis taught me lots about charts, graphs, statistics and dashboards. I learned regression analysis techniques as a part of this project
- Finally the call volume analysis gave me the opportunity to explore calculations, probabilities. I have learned how to work with assumptions and excel arithmetic functions
- All in All, this has been a journey of growth and challenges and further growth after overcoming those challenges
- Thanks a lot trinity for making this journey of practical learning so fun and encouraging

Reference

- [Github](#)
- [Google Drive](#)

Thank You

Thiaguion