

# Data Analysis Portfolio

Ashutosh Mishra  
Data Analyst

# Professional Background

- Seasoned sales professional with 10+ years of achievement in surpassing sales targets and fostering client relationships. Recently completed a Data Science and Machine Learning program, coupled with a Google Data Analytics certificate. 10-month Data Analysis intern Experience, applying analytical skills in real-world projects to bridge the gap between sales expertise and data-driven insights.
- **Highlights:**
- **Sales Excellence:** Proven track record in exceeding sales targets and managing diverse portfolios making a good story teller and communicator.
- **Data Science Intern:** Currently gaining hands-on experience in data cleaning, exploration, and visualization. Executed real-world projects to provide actionable insights.
- **Data Science Skills:** Proficient in Advance Excel, Python, SQL, Tableau, Power-bi, statistical analysis, and machine learning algorithms. Applied skills to extract valuable insights during internship projects.
- **Google Data Analytics Certification:** Demonstrates competence in data analysis, visualization, and interpretation.
- **Career Transition:** Aspiring to secure a Data Analyst role, combining professional experience with practical data analysis experience gained during the internship.
- I am eager to contribute my unique skill set to a data-centric role and translate complex data into actionable business insights.

# Table of Content

► Professional Background -----	02
► Table of Contents -----	03 - 04
► 1. Project 1 - Data Analytics Process	
► Description -----	05
► Design and Insights -----	06 - 08
► Result -----	09
► 2. Project 2 - Instagram User Analytics	
► Description and Insights -----	10
► Design and Insights -----	11-17
► Result and Insights -----	18
► 3. Project 3 - Operation and Metric Analytics	
► Description -----	19
► Design and Insights -----	20 - 31
► Result -----	32
► 4. Project 4 - Hiring Process Analytics	
► Description -----	33
► Design and Insights -----	34 - 36
► Result -----	37

# Table of Content (cont.)

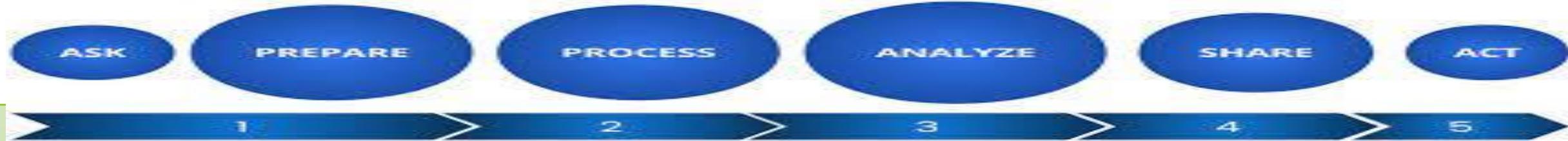
- ▶ 5. Project - IMDB Movie Analysis
- ▶ Description ----- 38
- ▶ Design and Insights ----- 39 - 44
- ▶ Result ----- 45
- ▶ 6. Project - Bank Loan Case Study
- ▶ Description ----- 46
- ▶ Design and Insights ----- 47 - 53
- ▶ Result ----- 54
- ▶ 7. Project - Impact of Car Features
- ▶ Description ----- 55
- ▶ Design and Insights ----- 55
- ▶ Result ----- 67
- ▶ 8. Project - ABC Call Volume Trend
- ▶ Description ----- 68
- ▶ Design and Insights ----- 69 - 77
- ▶ Result ----- 78
- ▶ 9. Final Project - 79 - 80
- ▶ 10. Learnings - 81 -82

# **Project 1**

# **Data Analytics Process**

## **Project Description:**

This project applies the key steps of the data analytics process to a hypothetical scenario of a retail store's Diwali promotional campaign. It demonstrates how data can be used to plan, execute, and optimize a marketing campaign.

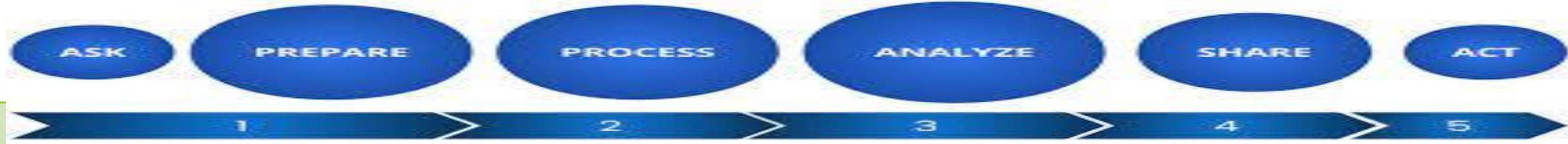


## Step 1: Plan

- **Data Analysis:** A2Z Stores begins by analysing historical sales data and market trends. They identify that during Diwali, there is a surge in demand for electronics, including smartphones, home appliances, and audiovisual equipment.
- **Budget Allocation:** Based on the data analysis, the store sets a budget for the Diwali promotion campaign. They allocate a larger budget for high-demand products and adjust their pricing strategy.
- **Optimal Inventory Levels:** Data analysis helps A2Z Stores determine the optimal inventory levels for Diwali. For example, they decide to stock a higher quantity of LED TVs and smartphones.

## Step 2: Prepare

- **Supplier Selection:** A2Z Stores uses data analytics to evaluate the performance of their electronics suppliers. They choose the most reliable ones based on past data, considering factors like delivery times and product quality.
- **Negotiating Terms:** Data-driven negotiations help the store secure favourable terms with suppliers. For instance, they leverage data to negotiate bulk discounts on popular electronics.
- **Data Collection Systems:** The store sets up data collection systems to track real-time sales data, inventory levels, and customer feedback. They use point-of-sale systems and online analytics tools.

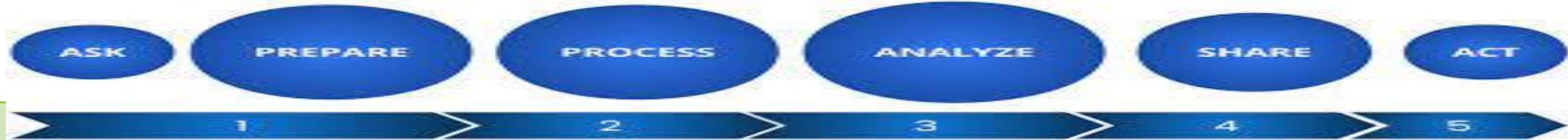


### Step 3: Process

- **Inventory Management:** A2Z Stores employs data analytics tools to monitor real-time inventory levels. This allows them to respond quickly to sales trends and reorder popular products.
- **Sales Tracking:** Real-time sales data is tracked continuously to optimize promotions. The store can see which products are selling well and make adjustments to their marketing strategy accordingly.
- **Crowd Management:** If the store is planning a Diwali sale event, data analytics helps estimate expected footfall, enabling better crowd management and resource

### Step 4: Analyse

- **Consumer Behaviour:** Data analytics is applied to understand customer behavior. A2Z Stores identifies which electronics are popular during Diwali and which customer segments are most interested.
- **Gift Preferences:** The store uses data analysis to provide personalized gift recommendations for shoppers based on their preferences and past purchases.
- **Festival Trends:** Trend analysis helps the store determine which electronic products are currently trending for Diwali gifts, helping them focus their promotions.



### Step 5: Share

- **Supply Chain:** A2Z Stores shares data with suppliers, allowing them to understand the high demand for specific electronics. This facilitates efficient restocking and supply chain coordination.
- **Event Planning:** If the store is planning a Diwali event or special sale days, they collaborate with sponsors, performers, and vendors based on data-driven insights on expected attendance and preferences.

### Step 6: Act

- **Promotions:** A2Z Stores launches data-driven promotions and discounts on popular electronics. They send personalized offers to customers, based on their preferences and past purchases.
- **Safety Measures:** If organizing a Diwali sale event, the store takes action to enhance safety measures based on data on crowd behaviour and historical incidents.
- **Waste Management:** The store uses data analytics to plan for waste management during the Diwali promotion by analysing previous years' data on the volume of waste generated.

## ❖ Result:

By applying data analytics at each step of the process, A2Z Stores significantly boosts its sales during the Diwali festival. They strategically stock high-demand products, offer personalized promotions, efficiently manage inventory, and ensure a safe and enjoyable shopping experience for their customers. This leads to increased revenue and customer satisfaction, making the Diwali promotion campaign a success.

# Project 2

## Instagam User Analytics

### Project Description

- This project is done for the purpose for to find ,how was users engage and interact with our digital platforms (Instagram) .
- This data below given is used for development our business and strategies to market our product.
- These insights given below are used by teams across the business to launch a new marketing campaign, decide on features to build for an app, track the success of the app by measuring user engagement and improve the experience altogether while helping the business grow.
- This objective of this project is defined as follows

#### For Marketing Team

1. To reward 5 oldest user of the instagram
2. To send email for those never posted a single photo in instagram .
3. To announce the contest winner.
4. To find the trending hashtag for our partner brand.
5. To launch a AD campaign in the best day where are in active.

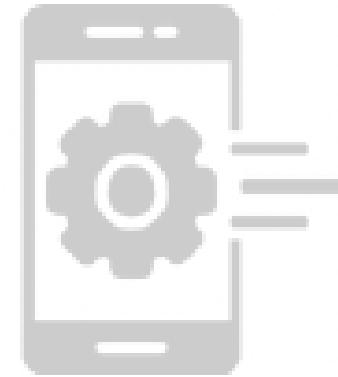
#### For Investor Metrics

1. To find how many of users using instagram and no.of photos posted in instagram and activities of users .
2. To find fake and dummy accounts in instagram.



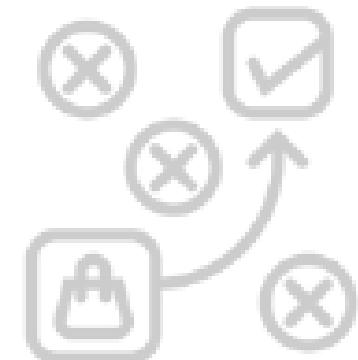
# Tech-Stack Used

- MySQL Workbench.  
Purpose : Used to filter and analyze the data.
- Canva.  
Purpose : Used to visualize the data.



# Approach

This Instagram user analytics project is so simple to do this. I had solved problems by using DML commands. I had faced some difficulty while solving "Best day when the users are active in Instagram". I have added graph, infographics and tables in this report. I had created this project report as simple to understand for development and marketing team.



# **SQL Queries for problems**

## **For Marketing Team**

### **1) Rewarding Most Loyal Users**

- `SELECT * FROM users ORDER BY created_at;`

### **2) Remind Inactive Users to Start Posting**

- `SELECT * FROM users WHERE id NOT IN(SELECT user_id FROM photos);`

### **3) Trending Hashtags**

- `SELECT tags.tag_name,tags.id , photo_tags.tag_id  
FROM tags INNER JOIN photo_tags ON  
tags.id = photo_tags.tag_id ,  
WHERE COUNT(*) AS total FROM photo_tags GROUP BY tag_id;`

### **4) Winner of the Contest**

- `SELECT user_id ,COUNT(*) AS total FROM likes GROUP BY user_id;`

### **5) Best day of launching AD Campaign**

- `SELECT DATE_FORMAT(created_at,'%a, %Y %M %e %H:%i:%s')AS  
format_registration_datetime FROM users;`

## For Investor Metrics

### 1) The Bots & Fake Accounts

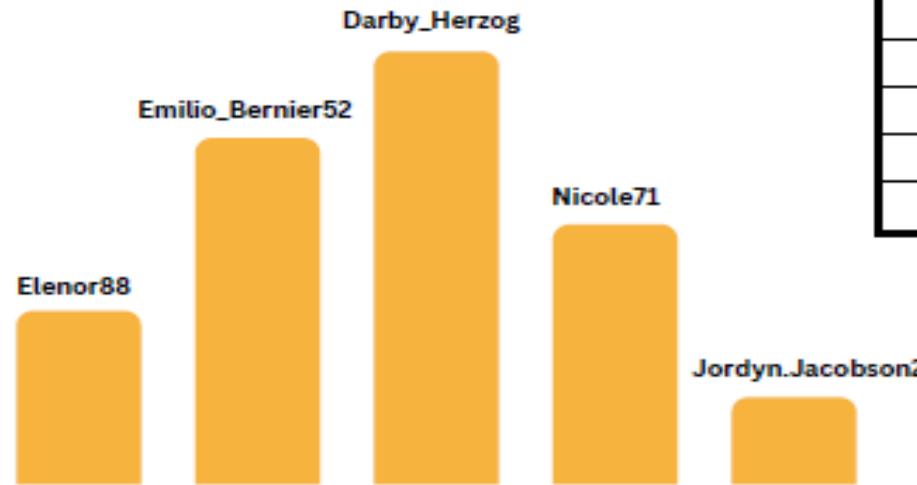
- `SELECT user_id ,COUNT(*) as total FROM likes GROUP BY user_id;`

### 2) User Engagement

- `SELECT AVG  
(SELECT user_id ,COUNT(*) as total FROM photos GROUP BY  
user_id) AS average FROM photos;`
- `SELECT COUNT(*) as Total_users FROM users;`
- `SELECT COUNT(*) as Total_photos FROM photos;`

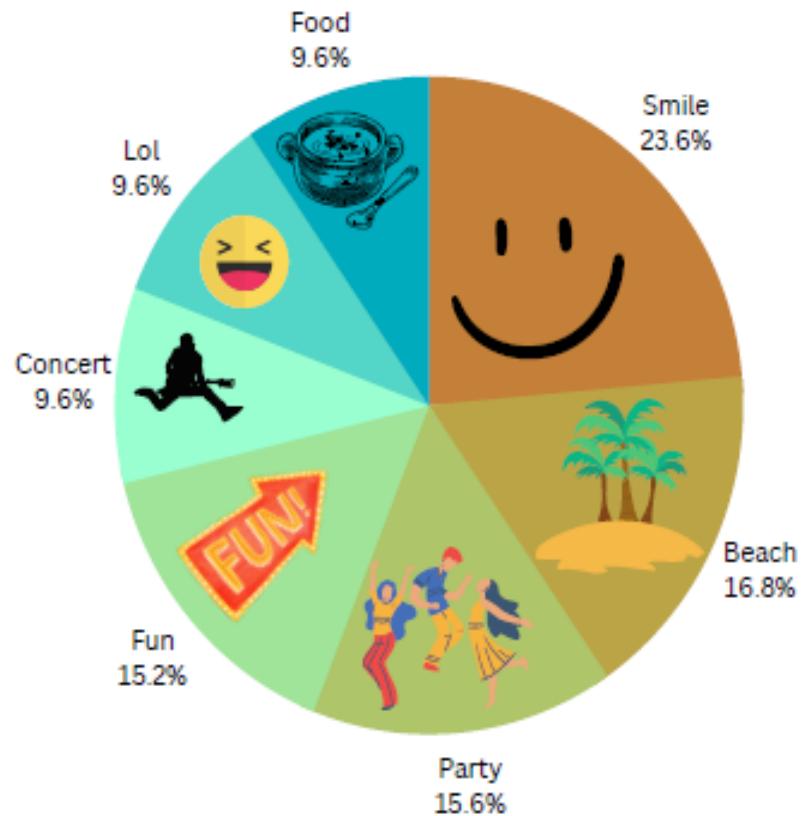
# Report for Marketing Team

## 1) Rewarding Most Loyal Users



Top Position	User id	Username	Created Date	Created Time
1	80	Darby_Herzog	5/6/2016	12:14:21 AM
2	67	Emilio_Bernier52	5/6/2016	1:04:30 PM
3	63	Elenor88	5/8/2016	1:30:41 AM
4	95	Nicole71	5/9/2016	5:30:22 PM
5	38	Jordyn.Jacobson	25/14/2016	7:56:26 AM

### 3) Trending Hashtags



Top position	Tag Name	Tag ID	Repeated Times
1	Smile	21	59
2	Beach	20	42
3	Party	17	39
4	Fun	13	38
5	Concert	5	24
	Lol	11	24
	Food	18	24

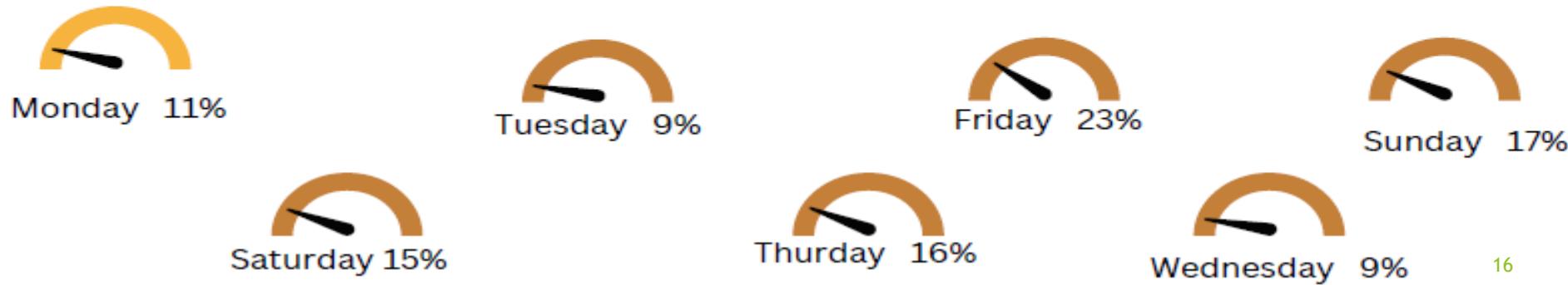
#### 4) Winner of the Contest



User ID : 52  
User Name : Zack\_Kemmer9  
User Created at : 31/1/2017 5:58  
Photo ID : 145  
Image URL : <https://jarret.name>  
Post Created At : 12/16/2022 21:32  
No.of Likes : 48

#### 5) Best day of launching AD Campaign

The Best day for launching AD Campaign is Friday



# Report for Investor Metrics

## 1) The Bots & Fake Accounts

S.no	Bot ID	Bot Name	Created At	
1	5	Aniya_Hackett	12/7/2016	1:04
2	14	Jaclyn81	2/6/2017	23:29
3	21	Rocio33	1/23/2017	11:51
4	24	Maxwell.Halvorson	4/18/2017	2:32
5	36	Ollie_Ledner37	8/4/2016	15:42
6	41	Mckenna17	7/17/2016	17:25
7	54	Duane60	12/21/2016	4:43
8	57	Julien_Schmidt	2/2/2017	23:12
9	66	Mike.Auer39	7/1/2016	17:36
10	71	Nia_Haag	5/14/2016	15:38
11	75	Leslie67	9/21/2016	5:14
12	76	Janelle.Nikolaus81	7/21/2016	9:26
13	91	Bethany20	6/3/2016	23:31

## 2) User Engagement

Total number of post in instagram : 257  
Total number of users in instagram : 100  
Total number of Original accounts : 87  
Total number of Fake accounts : 13

3-4 times does average user post on instagram

# **Result**

Through this project, the following achievements were realized:

**1. Meaningful Insights:**

Uncovered valuable insights into user behavior, engagement patterns, and content popularity.

**2. Informed Decision-Making:**

Provided actionable data for the marketing team's loyalty program, user engagement initiatives, and ad campaign scheduling.

**3. Enhanced Understanding:**

Improved understanding of user demographics and platform dynamics, contributing to future strategic planning.

The analysis conducted in this project has the potential to significantly impact the direction and success of Instagram by providing data-driven recommendations for feature development, marketing strategies, and user engagement initiatives. The insights gained will guide the product team in making informed decisions to enhance user satisfaction and drive platform growth.

# **Project 3**

## **Operation Analytics and Investigating Metric Spike**

### **Project Description**

- This analysis is about end to end working of company operation.
- This given data is used to find the department where the company wants to improve.
- By analyzing the data , I can predict the overall growth and decline of company's future.
- This objective of this project is defined as follows .

## **Case Study 1**

1. No.of jobs reviewed at November month.
2. No.of events happening per second.
3. The percentage of share of the each languages of users.
4. To find the duplicate rows.

## **Case Study 2**

1. To find the activeness of weekly user engagement.
2. User growth over a time for a product.
3. To find weekly retention of user-signup cohort.
4. To find activeness of user weekly engagement per device.
5. User engaging with the email service.

# Tech-Stack Used

- MySQL Workbench.  
Purpose : Used to filter and analyze the data.
- Canva.  
Purpose : Used to visualize the data.

# Approach

- I have solved operation and metrics analytics using DML commands.
- Using MySQL Workbench and I import files in new databases and started writing queries as the objective of this project.
- Executed all queries and finally , attached my code in this report.

# SQL Queries for creating database and schema

## Case Study 1

```
CREATE DATABASE Project;
USE project;
/*Users*/
CREATE TABLE users(
    DS TIMESTAMP DEFAULT NOW() NOT NULL,
    job_id INT NOT NULL,
    actor_id INT NOT NULL,
    event_name VARCHAR(255) NOT NULL,
    time_spent INT NOT NULL,
    Languages Varchar(255) Not null,
    org varchar(10) Not null
);
```

## Case Study 2

```
CREATE DATABASE TASK;
USE TASK;
/*Users*/
CREATE TABLE users(
    id INT PRIMARY KEY NOT NULL,
    created_at TIMESTAMP,
    company INT NOT NULL,
    languages VARCHAR(255) NOT NULL,
    activated_at TIMESTAMP NULL DEFAULT NULL,
    state VARCHAR(255) NOT NULL
);

/*events*/
CREATE TABLE events(
    users_id INT NOT NULL,
    occurred_at TIMESTAMP not null,
    event_type VARCHAR(255) NOT NULL,
    event_name VARCHAR(255) NOT NULL,
    location VARCHAR(255) NOT NULL,
    device_used VARCHAR(255) NOT NULL,
    user_type int,
    FOREIGN KEY(users_id) REFERENCES users(id)
);
```

```
/*email*/
CREATE TABLE email(
    users_id INT NOT NULL,
    occurred_at TIMESTAMP DEFAULT NOW(),
    email_action VARCHAR(255) NOT NULL,
    user_type int,
    FOREIGN KEY(users_id) REFERENCES users(id)
);
commit;
```

# SQL Queries & Result

## Case Study 1

### 1. No.of Jobs reviewed at november month

#### Query

```
with avgr(ds,times) as
(
  select ds,round((sum(time_spent)/3600)/(count(job_id)),4) as times from users where
  event_name in ('transfer','decision') group by ds
)
select round(sum(times)/count(times),4) as Timespent_perday_perhour from avgr;
```

#### Result

	Timespent_perday_perhour
▶	0.0117

## 2. Throughput No.of events happening per second.

### Query

```
with avgr (jey,job,ds) as
(
  select count(job_id) as jey, count(job_id) as job ,ds from users group by ds order by ds
)
select ds,job,avg(jey)over(order by jey) as rollingavg from avgr order by ds;
```

### Result

	ds	job	rollingavg
▶	2020-11-25 00:00:00	1	1.0000
	2020-11-26 00:00:00	1	1.0000
	2020-11-27 00:00:00	1	1.0000
	2020-11-28 00:00:00	2	1.3333
	2020-11-29 00:00:00	1	1.0000
	2020-11-30 00:00:00	2	1.3333

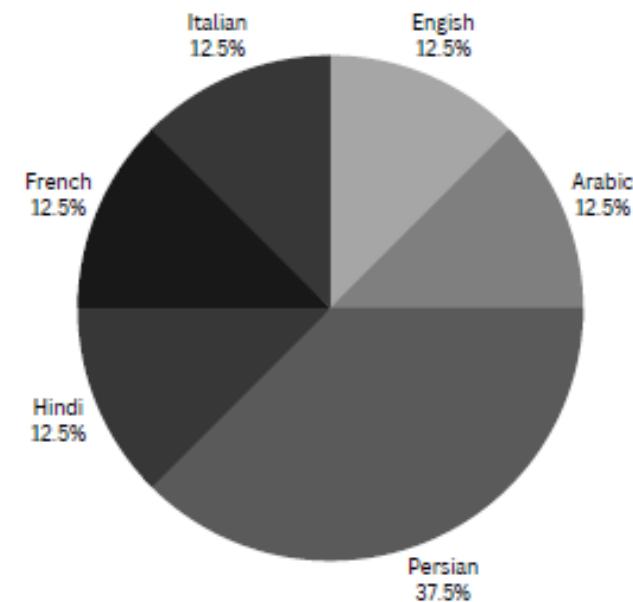
### 3. The percentage of share of the each languages of users.

#### Query

```
select Languages, count(job_id) *100 / (select count(*) from users) as percentshare from users group by languages;
```

#### Result

	Languages	percentshare
▶	English	12.5000
	Arabic	12.5000
	Persian	37.5000
	Hindi	12.5000
	French	12.5000
	Italian	12.5000



#### 4. To find the duplicate rows.

##### Query

```
select job_id,count(*) from users group by job_id having count(*)>1;
```

##### Result

	job_id	count(*)
▶	23	3

## Case Study 2

### 1. To find the activeness of weekly user engagement.

#### Query

```
select event_type as state, count(distinct users_id) as Num_users from events  
where users_id in (select id from users where state like('active')) group by event_type limit 1;
```

#### Result

	state	Num_users
▶	engagement	6142

## 2. User growth over a time for a product.

### Query

```
select job_id,count(*) from users group by job_id having count(*)>1;
```

### Result

	state	Num_users
▶	engagement	6142

### 3. To find weekly retention of user-signup cohort.

#### Query

```
select device_used ,count(distinct(users_id)) as total from events where  
users_id in (select id from users where state like('active')) group by device_used;
```

#### Result

	device_used	total
▶	acer aspire desktop	196
	acer aspire notebook	334
	amazon fire phone	88
	asus chromebook	351
	dell inspiron desktop	356
	dell inspiron notebook	671
	hp pavilion desktop	337
	htc one	196
	ipad air	470
	ipad mini	285
	iphone 4s	399
	iphone 5	1013
	iphone 5s	615

	device_used	total
	kindle fire	200
	lenovo thinkpad	1284
	mac mini	148
	macbook air	943
	macbook pro	1923
	nexus 10	270
	nexus 5	615
	nexus 7	350
	nokia lumia 635	210
	samsung galaxy tablet	106
	samsung galaxy note	115
	samsung galaxy s4	790
	windows surface	178

#### 4. To find activeness of user weekly engagement per device.

##### Query

```
select state, count(*) as user_per_device from users group by state limit 1;
```

##### Result

	state	user_per_device
▶	active	9381

## 5. User engaging with the email service.

### Query

```
select email_action,user_type,count(distinct users_id) as no_of_user from email group  
by email_action,user_type order by user_type;
```

### Result

	email_action	user_type	no_of_user
▶	email_clickthrough	1	1529
	email_open	1	1717
	sent_reengagement_email	1	892
	sent_weekly_digest	1	1217
	email_clickthrough	2	1529
	email_open	2	1701
	sent_reengagement_email	2	1071
	sent_weekly_digest	2	1098
	email_clickthrough	3	2219
	email_open	3	2509
	sent_reengagement_email	3	1690
	sent_weekly_digest	3	1796

# Result

- I have completed a visualization project on operation and investigation metric data.
- I have studied the relationship between the users and the company languages and email response.
- We use Canva to visualize and MySQL workbench analyze the raw data
- which has been processed to a easy-to use form and find interesting patterns.

# Project 4 HR Process Analytics

## Hiring Process Analytics Project Report

### Project Description

The project aimed to analyse the hiring process data at a multinational company, leveraging statistical and Excel-based analysis to derive actionable insights. The primary objectives included understanding trends in hiring, identifying outliers, handling missing data, and presenting findings to aid the improvement of the company's hiring process.

### Approach

The approach involved:

- **Data Collection:** Obtained a dataset comprising hiring records.
- **Data Cleaning:** Handled missing values, outliers, and combined columns for streamlined analysis.
- **Statistical Analysis:** Utilized Excel to calculate averages, gender distribution, salary analysis, and visualized data for better comprehension.
- **Insight Generation:** Derived meaningful insights to aid the hiring department's decision-making process.

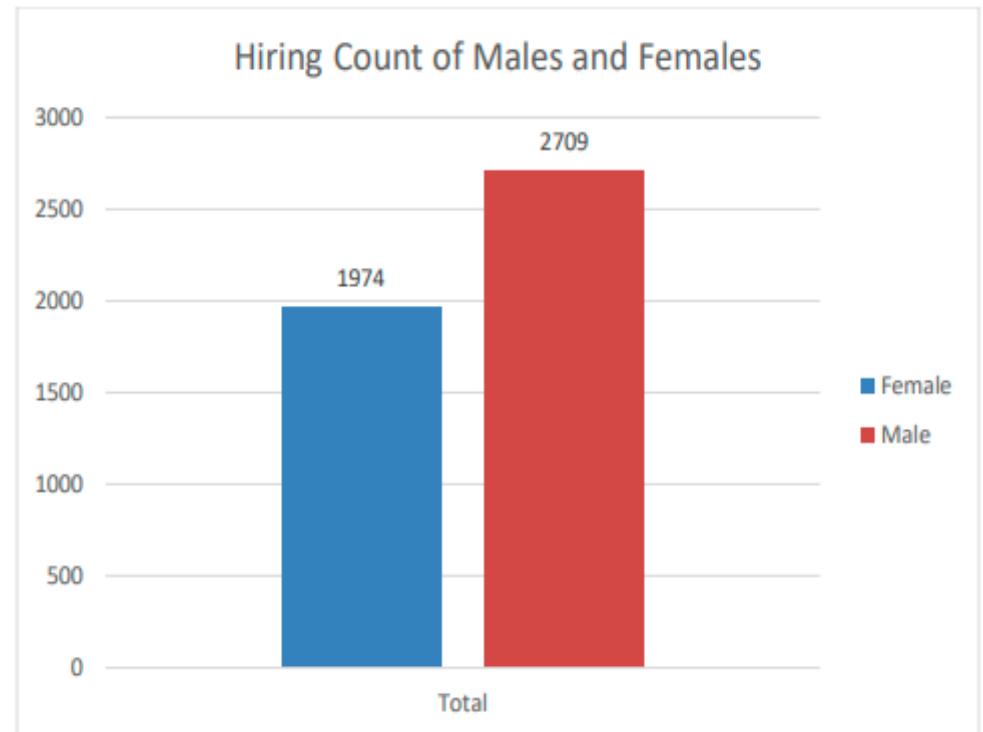
## Tech-Stack Used

- **Microsoft Excel 2022:** Employed for data cleaning, analysis, visualization, and insights derivation.

## Insights

### A. Hiring Analysis

- **Gender Distribution:** Out of the hires, 2709 were males and 1974 were females.



## B. Salary Analysis

- **Average Salary:** The average salary offered by the company is **49881**.

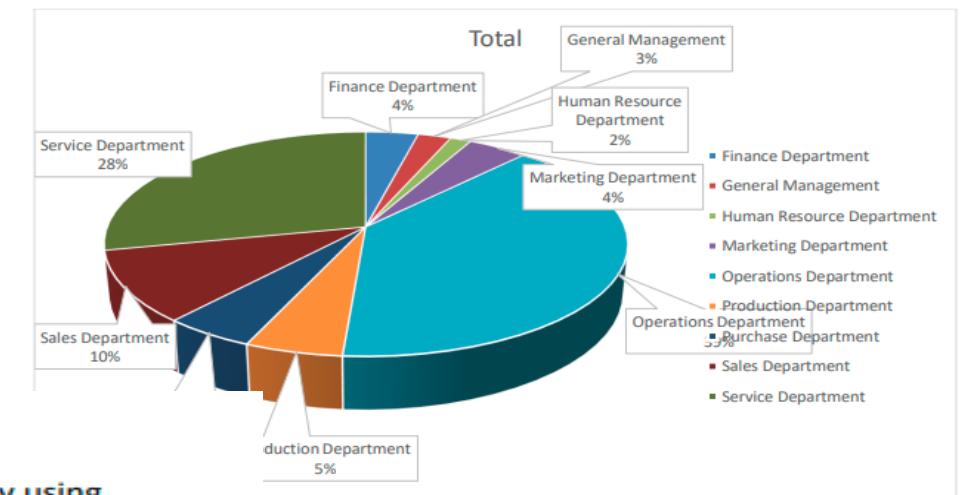
## C. Salary Distribution

- **Class Intervals:** Defined salary brackets to understand salary distribution across ranges.

Salary Brackets	Salary Class Intevals
0-20000	1405
20000-40000	1418
40000-60000	1532
60000-80000	1430
80000-100000	1364

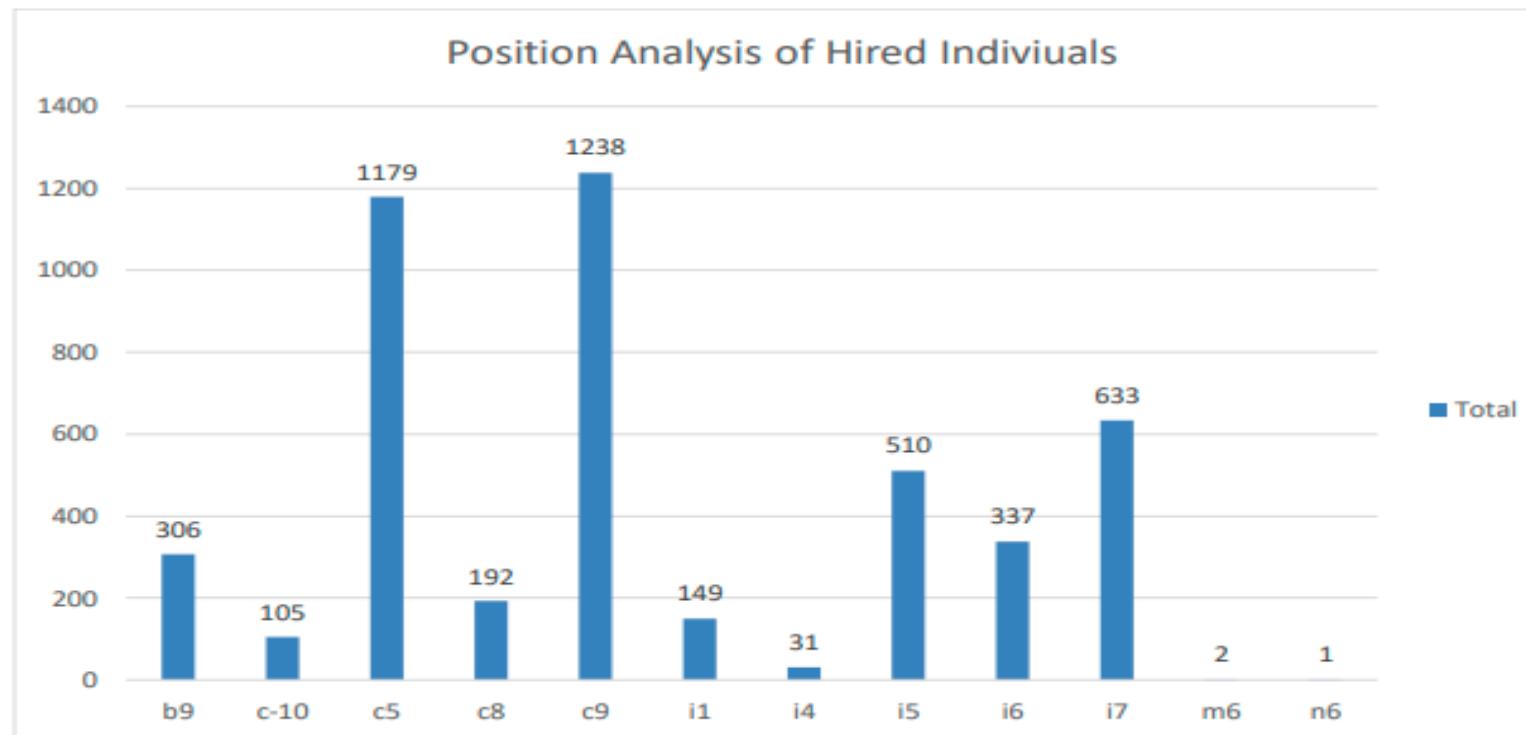
## D. Departmental Analysis

- Visualization:** Utilized Pie Charts to depict the proportion of employees across different departments.



## E. Position Tier Analysis

- Chart Representation:** Visualized the distribution of position tiers within the company using Column Chart Type.



## Results

### Achievements

The analysis offered comprehensive insights into the hiring process, including gender distribution, salary trends, and departmental structure. The findings contributed to a deeper understanding of the company's hiring landscape, potentially aiding future hiring strategies.

# **Project 5**

## **IMDb Movie Analysis**

### **Project Description**

IMDb provides a wealth of information, including:

- Basic information about a movie or TV show, such as its title, release date, director, and cast.
- User ratings and reviews, which can help you get a sense of what others thought of a particular movie or TV show.
- Box office data and other financial information, such as the budget and gross earnings of a movie.
- Technical information, such as the running time, aspect ratio, and sound information for a movie.
- The libraries for data analysis and visualization used in this project is Excel.

## **Approach:**

**Our approach involved comprehensive data analysis using Microsoft Excel 2022. We executed the project in phases, starting with data cleaning, where we sorted and organized the dataset to ensure accuracy. We then performed various statistical analyses, including correlation calculations using Excel functions like CORREL, AVERAGEIF, and MAX to assess relationships between budget, gross earnings, and director impact on ratings.**

## **Tech-Stack Used:**

- Microsoft Excel 2022: Utilized for data manipulation, statistical analysis, and visualization.**
- IMDB Movie Dataset: Main source for movie-related information and attributes.**

## Insights:

### A. Movie Genre Analysis:

- Unique Genres:

- Drama has the highest count with 1800 movies, followed by Comedy (1425) and Action (890).

#### Impact on IMDB Scores:

- Drama movies have a mean IMDB score of 6.5 and a median score of 6.9, indicating a consistent rating distribution with some higher-rated movies.
- Documentary films have a high variance (2.06), implying a wider range of IMDB scores, while Musical movies have the lowest variance (0.65), suggesting more consistent ratings.
- Mystery films exhibit an unusually high range of IMDB scores (43.5), indicating potential outliers affecting the overall range.

#### Variability and Consistency:

- Animation, Fantasy, Family, and Sci-Fi genres have relatively high standard deviations, indicating a wider spread of IMDB scores within these genres.
- Biography movies showcase a low standard deviation, suggesting more consistency in ratings across this genre.

## Median vs. Mean:

- Romance films show a higher median (7.6) compared to the mean (6.5), indicating a skewed distribution with several higher-rated movies.
- Thriller movies display a similar trend with a higher median (6.4) than the mean (7.4), suggesting some lower-rated movies affecting the mean.

Genre Type	Unique Counts	Mean	Median	STD	Range	Variance
Action	890	6.28	6.35	1.04	6.90	1.08
Adventure	725	6.56	6.60	1.13	6.60	1.27
Thriller	1031	4.80	6.40	0.97	6.30	0.94
Animation	193	6.74	6.80	1.00	5.80	0.99
Family	422	126.70	6.30	1.18	6.70	1.39
Fantasy	467	6.28	6.30	1.14	6.70	1.30
Romance	818	4.12	6.50	0.96	6.40	0.93
Crime	676	6.94	6.60	0.99	6.90	0.98
Comedy	1425	6.17	6.30	1.03	6.90	1.07
Drama	1800	6.85	6.90	0.88	7.20	0.78
Sci-Fi	462	99.79	6.40	1.18	6.90	1.39
Horror	357	5.78	5.90	0.98	6.30	0.97
Mystery	358	37.04	6.50	0.99	5.30	0.99
Western	55	8.10	6.80	0.95	4.20	0.90
History	143	138.00	7.20	0.66	3.40	0.44
Musical	94	6.60	6.70	1.11	6.40	1.23
Music	147	6.70	6.50	1.24	6.90	1.53
War	146	6.50	7.10	0.80	4.30	0.64
Biography	235	7.16	7.20	0.70	4.40	0.48
Sport	138	6.50	6.80	1.06	6.30	1.12
Documentary	45	6.80	7.40	1.38	6.90	1.92

## B. Movie Duration Analysis:

### Mean Duration:

The mean duration of movies in the dataset is 110.1 minutes. This value represents the average length of movies.

### Median Duration:

The median duration is 106 minutes. This value indicates the middle point of the dataset when arranged in ascending order of duration. It's noteworthy that the median is slightly lower than the mean.

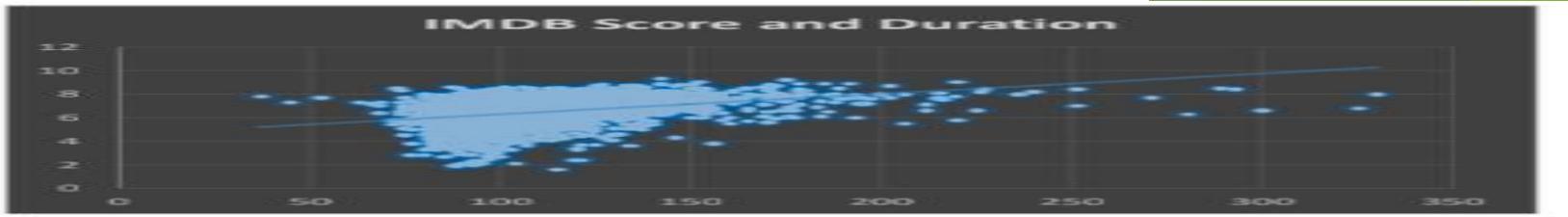
### Standard Deviation of Duration:

The standard deviation of 22.8 minutes shows the amount of variation or dispersion from the average duration. A higher standard deviation signifies that the data points are more spread out from the mean.

### Interpretation:

The mean duration being higher than the median suggests that there might be a few longer movies that are influencing the average to be higher.

The standard deviation being relatively high indicates that movie durations vary considerably from the average duration of 110.1 minutes. There's notable variability in movie lengths, with some significantly longer or shorter films compared to the average.



Descriptive Statistics	
Mean for Duration	110.09
Median for Duration	106.00
Standard Deviation for Duration	22.79

## C. Language Analysis:

### 1. Language and Ratings:

- ❑ English dominates the dataset with the highest count of movies (3405). It maintains a mean IMDb score of 6.4 and a median of 6.5, showcasing consistent ratings.
- ❑ Mandarin, Spanish, and French also have notable counts and relatively higher mean IMDb scores (7.0, 7.1, and 7.4, respectively).

### 2. Consistency Across Languages:

- ❑ Several languages, such as Filipino, Kazakh, Mongolian, and Russian, have limited movie counts (1 or 2) but display consistent mean and median IMDb scores.
- ❑ Aramaic, Bosnian, Zulu, Hebrew, Arabic, Vietnamese, and Romanian languages each have one movie represented, and their IMDb scores align closely with the dataset's overall standard deviation.

### 3. Higher-Rated Languages:

- ❑ Japanese, Danish, Portuguese, and Persian languages showcase movies with higher IMDb ratings, having mean scores ranging from 7.8 to 8.1.

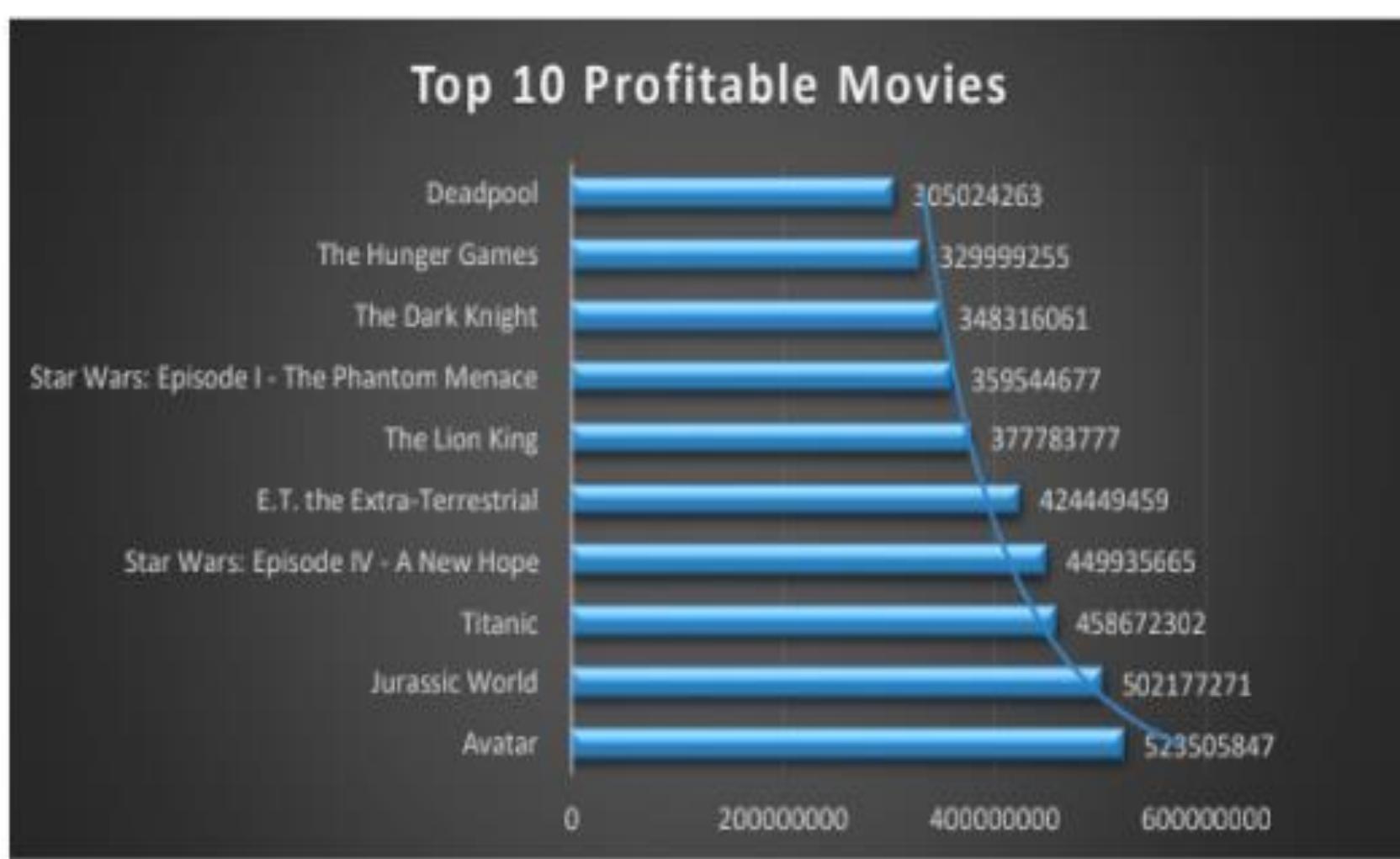
### Overall:

The dataset exhibits diverse languages with varied movie counts and IMDb ratings. While English dominates in volume, languages like Mandarin, Spanish, and French also

# D. Director Analysis

TOP 10 DIRECTOR NAMES As per AVG IMDB Ratings			
S.No	Director Names	Average IMDB Ratings	Percentile Comparison
1	Akira Kurosawa	8.7	ABOVE
2	Tony Kaye	8.6	ABOVE
3	Charles Chaplin	8.6	ABOVE
4	Alfred Hitchcock	8.5	ABOVE
5	Ron Fricke	8.5	ABOVE
6	Damien Chazelle	8.5	ABOVE
7	Florian Henckel von Donnersmarck	8.5	ABOVE
8	Majid Majidi	8.5	ABOVE
9	Sergio Leone	8.4	ABOVE
10	Christopher Nolan	8.4	ABOVE

# E. Budget Analysis



# Project 6

## Bank Loan Case Study



### Project Description:

This project involves conducting an Exploratory Data Analysis (EDA) on a dataset containing loan application information for a finance company. The primary objective is to identify patterns that indicate the likelihood of loan default among customers and understand the key factors influencing loan repayment.

Through this analysis, the aim is to support better decision-making in loan approval processes.

## **Approach:**

For this project, Microsoft Excel was utilized as the primary tool for data analysis. The following steps were undertaken:

### **1. Data Preprocessing:**

- Data cleaning involved handling missing values and outliers to ensure the accuracy of the analysis.

### **2. Exploratory Data Analysis:**

- Univariate, segmented univariate, and bivariate analyses were performed to understand variable distributions, compare scenarios, and explore relationships between variables and the target variable (likelihood of loan default).

### **3. Statistical Analysis:**

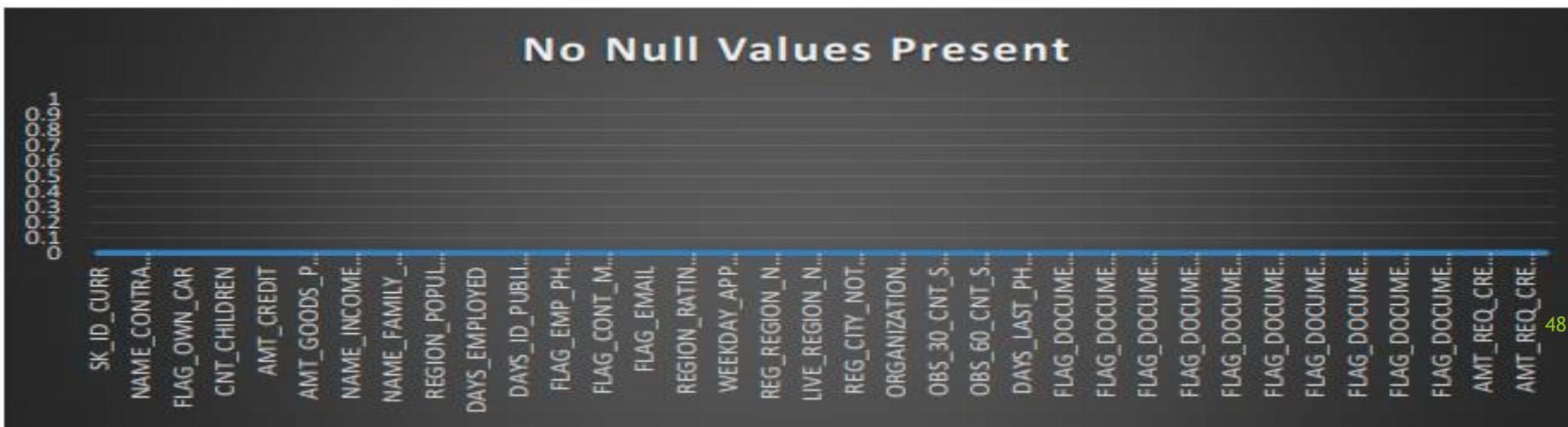
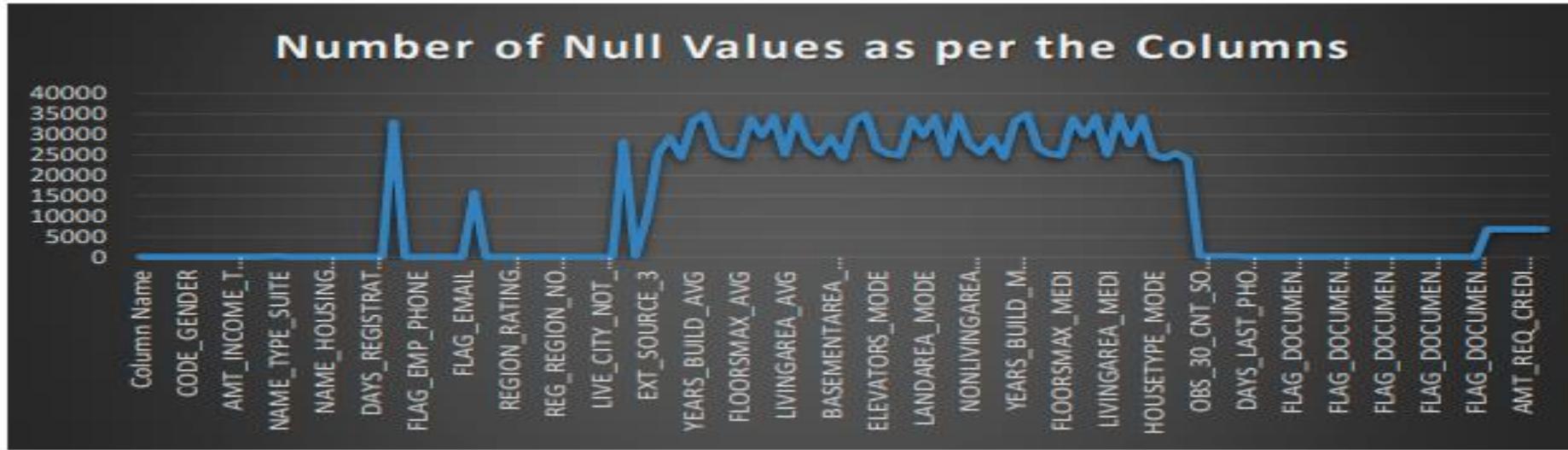
- Techniques such as correlation analysis were applied to identify relationships between variables and the target variable, aiding in determining indicators of loan default.

### **4. Visualization:**

- Graphs such as bar charts, box plots, scatter plots, and heatmaps were created in Excel to visualize data distributions, correlations, and relationships between variables.

# Insights:

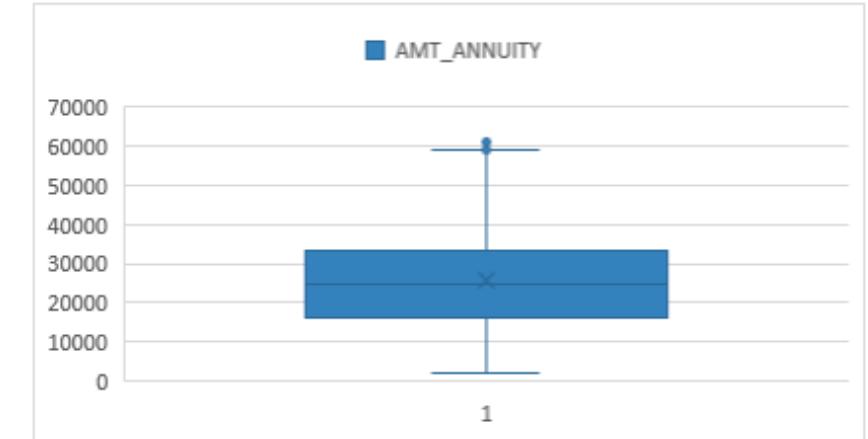
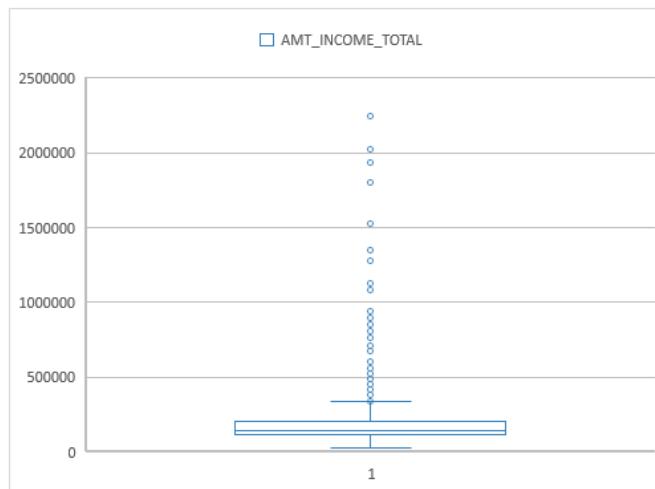
1. Identify Missing Data and Deal with it Appropriately: Identified missing values across variables and employed appropriate methods such as imputation using averages or medians.



**B. Identify Outliers in the Dataset:** Detected outliers in numerical variables using quartiles and interquartile range (IQR) methods, highlighting potential anomalies that may affect analysis.

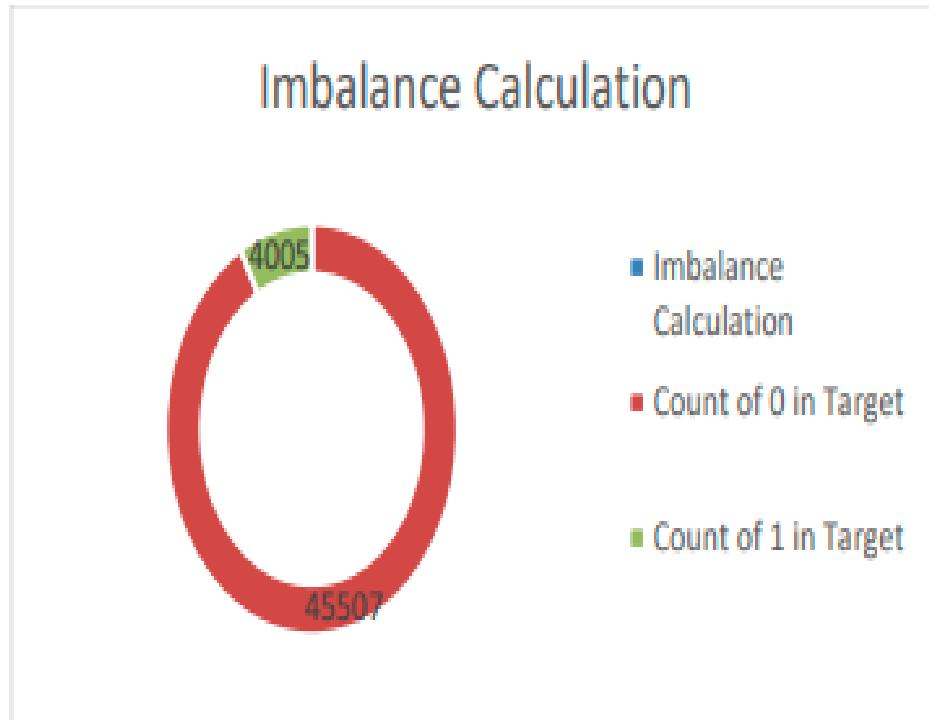
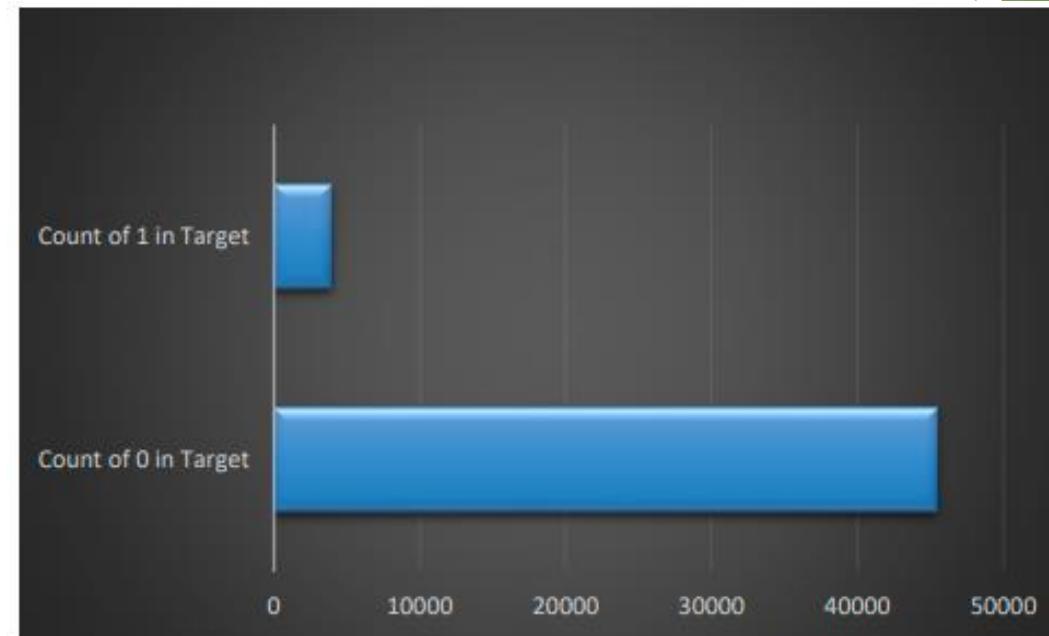
**Outlier Calculation of Relevant Columns:**

Outliers Calculation					
Column Name	Q1	Q3	IQR	Upper Bound	Lower Bound
AMT_INCOME_TOTAL	112500	202500	90000	337500	-22500
AMT_CREDIT	270000	808650	538650	1616625	-537975
AMT_GOODS_PRICE	238500	679500	441000	1341000	-423000
DAYS_BIRTH	-19643	-12382.75	-7260	-23272.75	-8753
DAYS_EMPLOYED	-2789.25	-294	2495	3448.875	-6532.125
DAYS_REGISTRATION	-7463.25	-1999	5464	6197.375	-15659.625
DAYS_ID_PUBLISH	-4297	-1722.75	2574	2138.625	-8158.375
AMT_ANNUITY	16522.875	34596	18073	61705.6875	-10586.8125



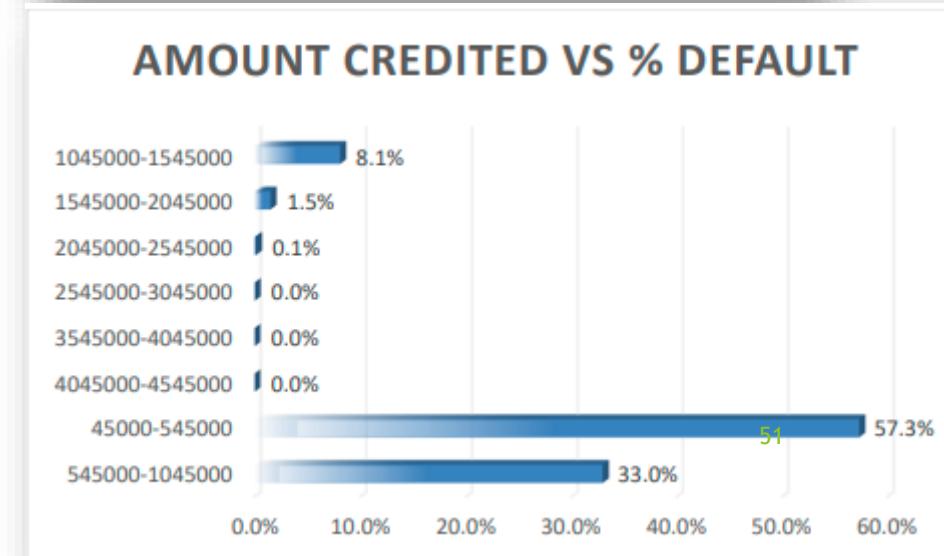
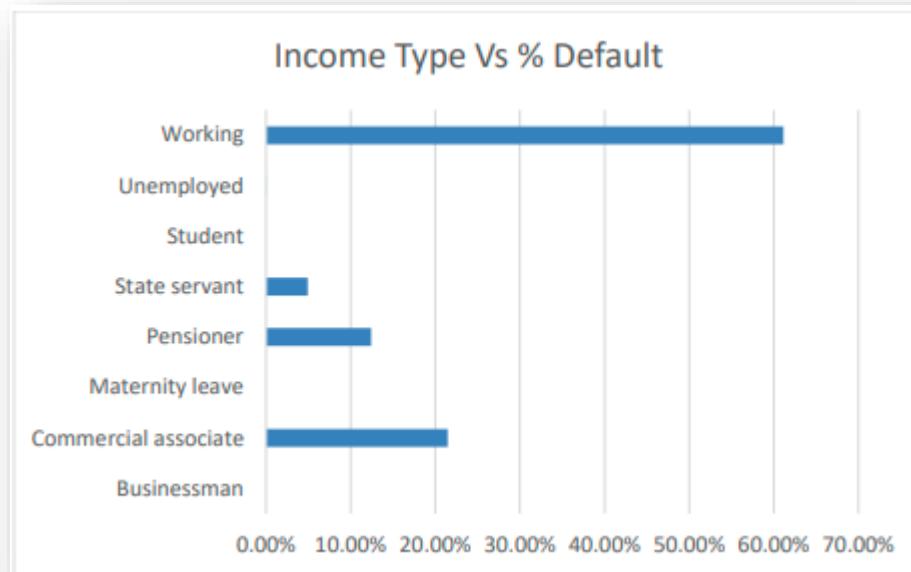
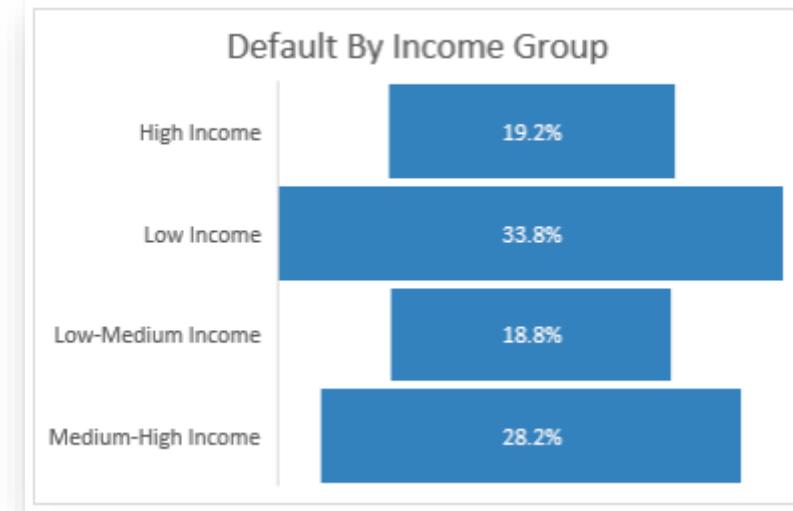
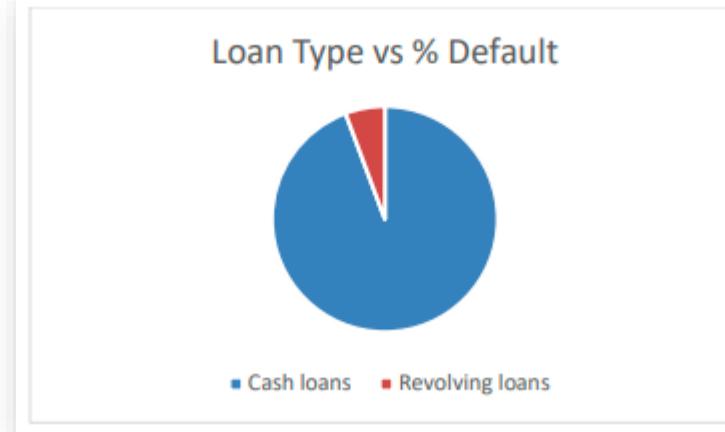
**C. Analyze Data Imbalance:** Observed data imbalance for the Target Variable, crucial for understanding the distribution and potential bias in the dataset.

Imbalance Calculation	
Count of 0 in Target	45507
Count of 1 in Target	4005
Imbalance Ratio	0.088

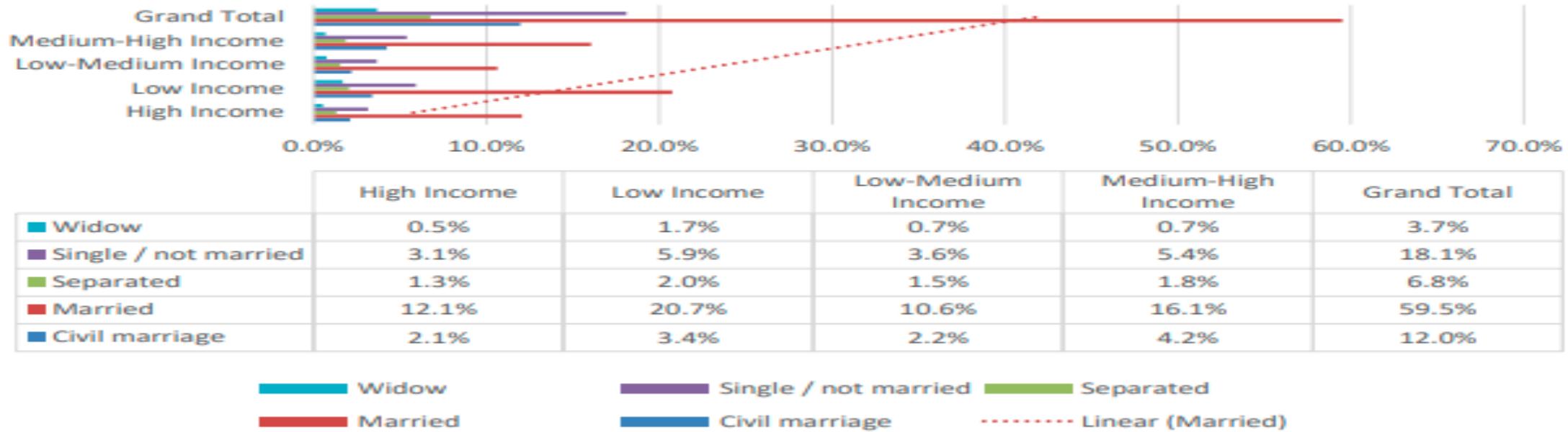


## D. Perform Univariate, Segmented Univariate, and Bivariate Analysis:

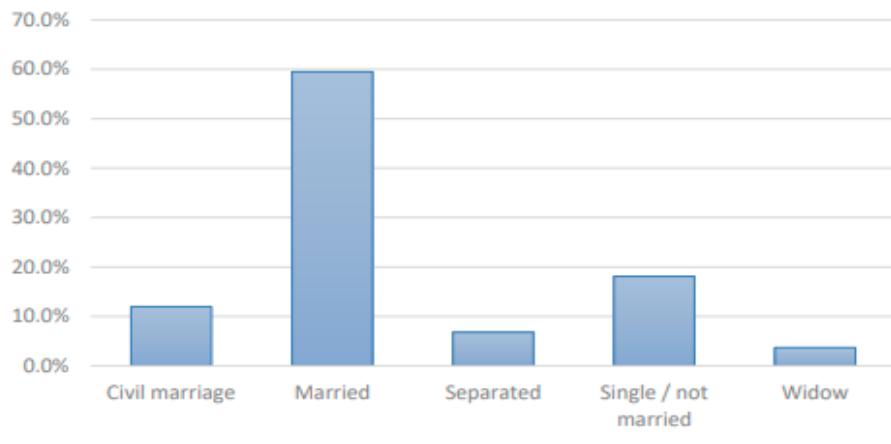
Identified key variables strongly correlated with the likelihood of loan default in different customer scenarios, providing insights into risk factors. Here are the graphical Representations of the same:



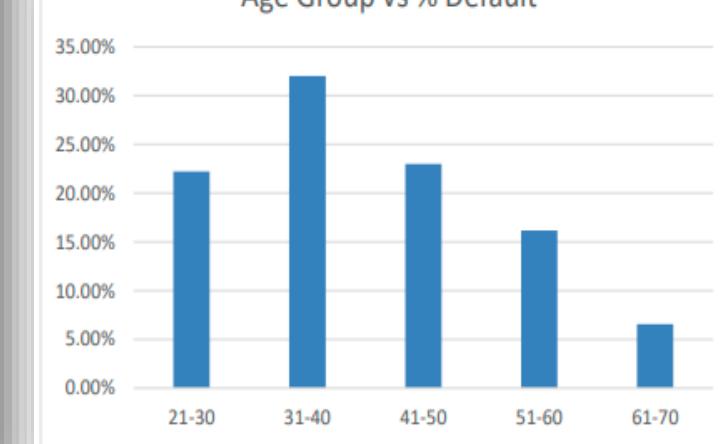
## Income and Family Status vs % Default



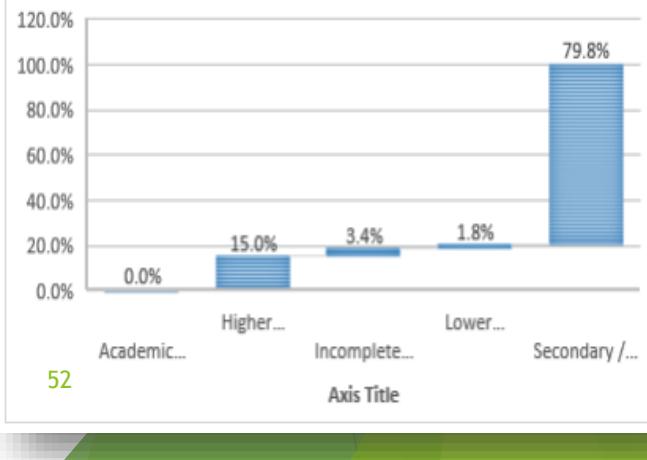
### Family Status vs % Default



### Age Group vs % Default



### EDUCATION VS % DEFAULT



## E. Identify Top Correlations for Different Scenarios:

### Positive Correlations:

- CODE\_GENDER, DAYS\_BIRTH, Employment Duration, DAYS\_REGISTRATION, DAYS\_ID\_PUBLISH, FLAG\_EMP\_PHONE, CNT\_FAM\_MEMBERS, REGION\_RATING\_CLIENT, REGION\_RATING\_CLIENT\_W\_CITY, REG\_CITY\_NOT\_LIVE\_CITY, REG\_CITY\_NOT\_WORK\_CITY, LIVE\_CITY\_NOT\_WORK\_CITY, DEF\_30\_CNT\_SOCIAL\_CIRCLE, DEF\_60\_CNT\_SOCIAL\_CIRCLE, and DAYS\_LAST\_PHONE\_CHANGE show positive correlations.
- Example: DAYS\_BIRTH and Age positively correlate, indicating an increased likelihood of loan default with age.

### Negative Correlations:

- FLAG\_OWN\_CAR, AMT\_INCOME\_TOTAL, AMT\_CREDIT, AMT\_ANNUITY, AMT\_GOODS\_PRICE, FLAG\_PHONE, HOUR\_APPR\_PROCESS\_START, FLAG\_EMAIL, FLAG\_WORK\_PHONE, FLAG\_CONT\_MOBILE, REG\_REGION\_NOT\_LIVE\_REGION, REG\_REGION\_NOT\_WORK\_REGION, LIVE\_REGION\_NOT\_WORK\_REGION, and DAYS\_EMPLOYED exhibit negative correlations.
- Example: AMT\_INCOME\_TOTAL, AMT\_CREDIT, and related financial variables negatively correlate, suggesting a lower likelihood of loan default with higher income or credit amounts.

### Weak Correlations:

- CNT\_CHILDREN, OBS\_30\_CNT\_SOCIAL\_CIRCLE, OBS\_60\_CNT\_SOCIAL\_CIRCLE have correlations close to zero, indicating a weak linear relationship with the likelihood of loan default.

**Results:** This project successfully identified crucial patterns and insights within the loan application dataset. The analysis highlighted significant variables influencing loan default scenarios. Insights gained from this project contribute to making informed decisions in loan approval processes, enabling the company to mitigate risks associated with loan default.

# Project 7

## Impact of Car Features

This project aimed to analyze the impact of car features on pricing and profitability for a car manufacturer. The analysis aimed to provide insights into optimizing pricing strategies and product development decisions while meeting consumer demand.

**Overview:** Explored trends in car features, pricing, and consumer preferences using a dataset containing over 11,000 car models and their specifications.

**Business Problem:** How to maximize profitability by understanding the relationship between car features, market categories, and pricing.

**Data Sources:** Utilized the "Car Features and MSRP" dataset from Cooper Union available in CSV format.

**Data Preprocessing:** Cleaned and prepared the dataset, handling missing values, and ensuring data consistency. Assumptions were made for missing data based on logical deductions.

# Approach:

The analysis involved descriptive statistics, regression analysis, and visualization techniques to understand various aspects of the automotive industry and address business problems effectively.

**Analytical Methods:** Employed regression analysis to determine the relationship between car features and pricing, while leveraging visualization techniques for insights.

**Tech Stack:** Utilized Microsoft Excel for data cleaning, PivotTables, regression analysis, and dashboard creation due to its versatility and familiarity.

**Challenges:** Encountered challenges with missing data

# Insights:

Insight Required 1: How does the popularity of a car model vary across different market categories?

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

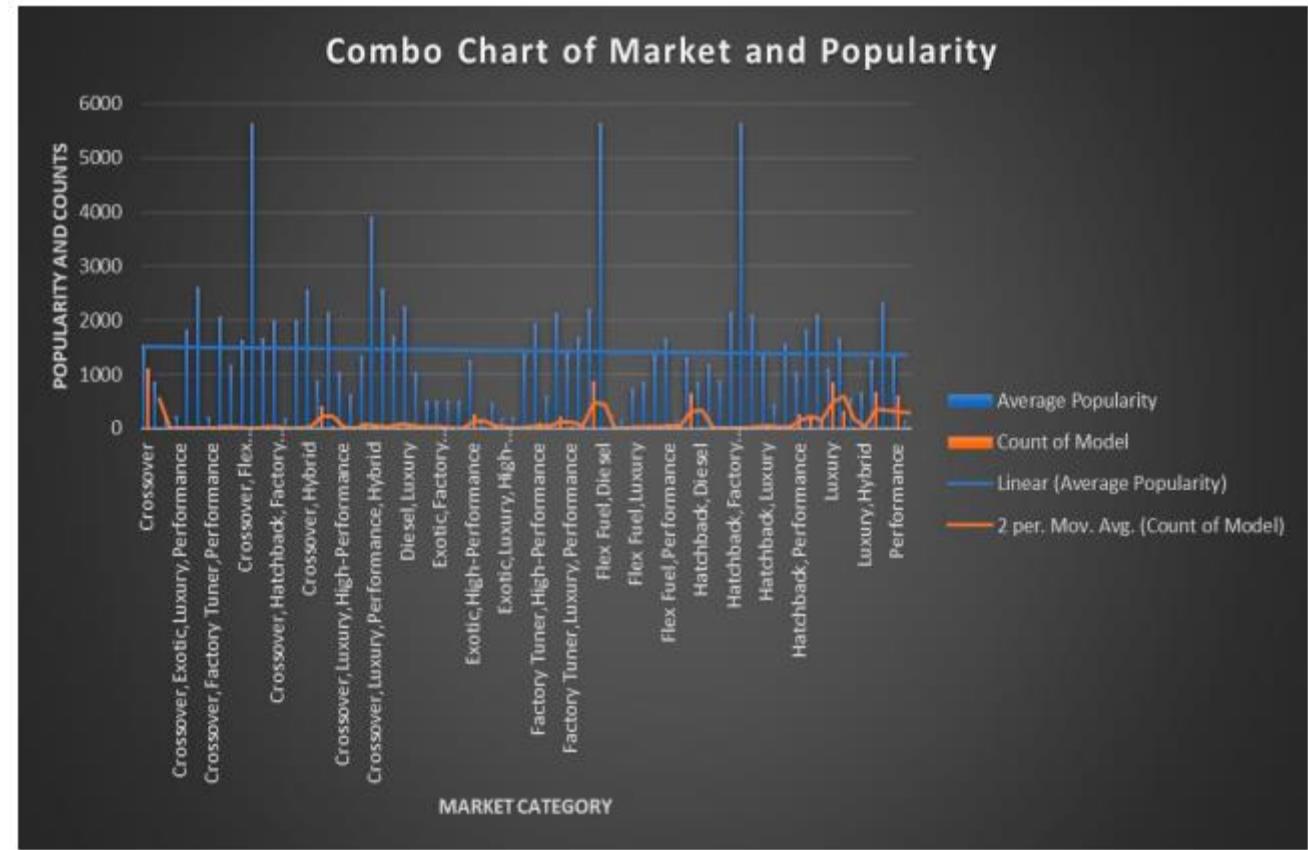
## Task 1.A and 1.B Solution:

Created a Pivot Table showcasing the count of car models within each market category alongside their corresponding popularity scores.

The table offers insights into how popularity varies across diverse market segments.

## The Key Insights Found are -

- Luxury and Performance Variants are Popular:** The categories that involve luxury or performance tend to have higher average popularity scores.
- Flex Fuel Variants Have Varied Popularity:** Flex Fuel variants show a diverse range of popularity scores across different configurations. While some configurations like "Flex Fuel" have a high count of models, the average popularity scores for these configurations vary widely.
- Exotic and Factory Tuner Categories Have Niche Appeal:** Categories like "Exotic" and "Factory Tuner" showcase niche segments with relatively lower counts of models and varying popularity scores. These segments likely cater to specific audiences seeking high-performance, specialized vehicles, hence their lower count but often higher average popularity scores.



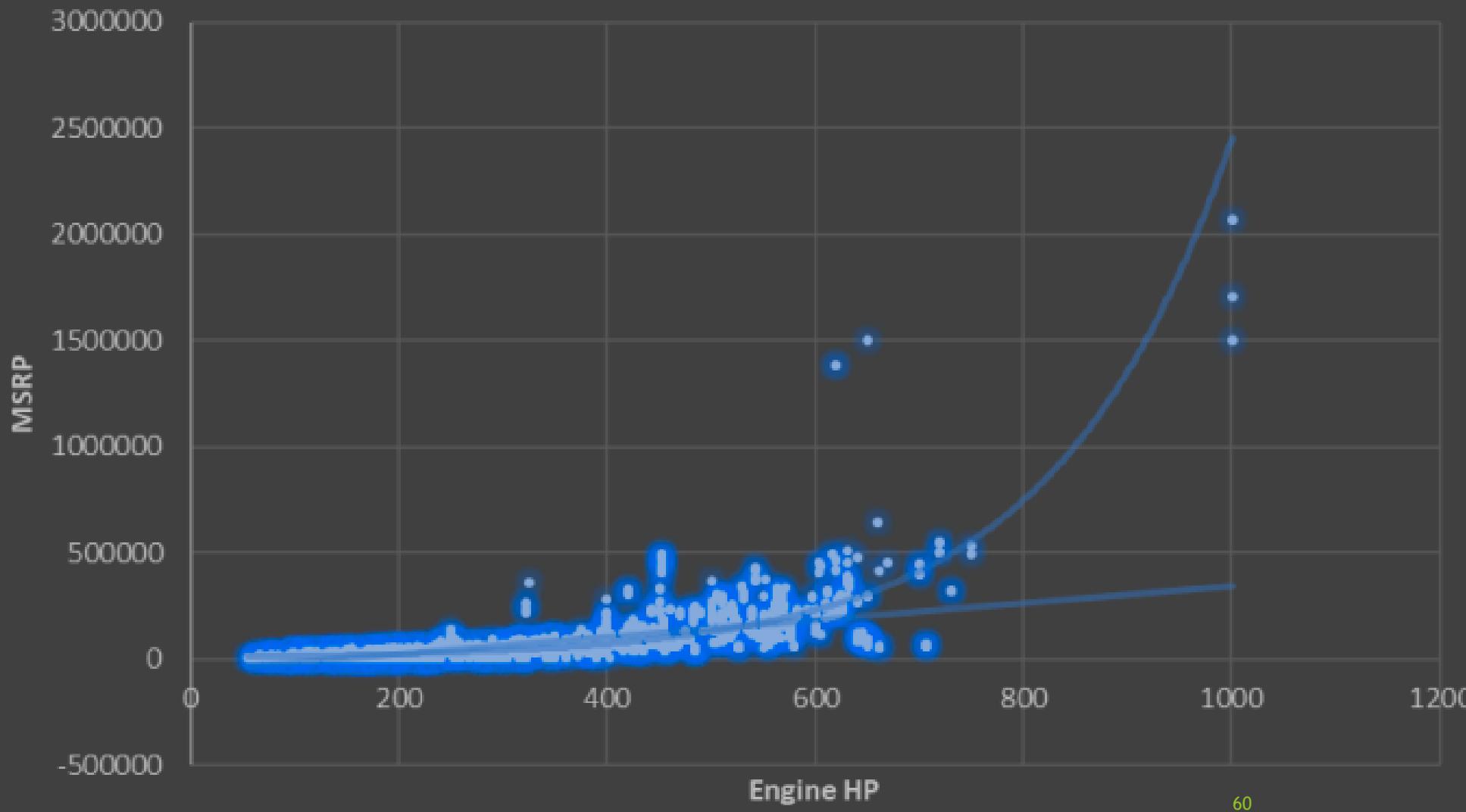
## Insight Required 2: What is the relationship between a car's engine power and its price?

- **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.

## Solution for Task 2:

1. Constructed a Scatter Chart plotting engine power on the x-axis against car prices on the y-axis.
2. Incorporated a trendline to visually depict the relationship between engine power and price, providing insights into their correlation.

## Scatter Chart of Engine HP vs MSRP



Insight Required 3: 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.

Solution for Task 3:

1. Utilized regression analysis to identify the most influential car features determining price.
2. Created a Bar Chart showcasing coefficient values for each variable, offering a visual representation of their relative importance in influencing car prices.

Regression Analysis	
Independent Variables	Coefficients
Engine HP	-2.530388699
highway MPG	-1520.641807

city mpg	-1337.784745
Popularity	400.8396044
Engine Cylinders	20973.22252
Number of Doors	-14247.98511

### 1. Negative Coefficients for MPG:

- Both highway MPG and city MPG have negative coefficients.
- Higher fuel efficiency (MPG) correlates with lower prices or sales.

### 2. Positive Coefficient for Popularity:

- "Popularity" coefficient is positive.
- Higher popularity is positively correlated with the dependent variable (e.g., price or sales).

### 3. Strong Influence of Engine Cylinders:

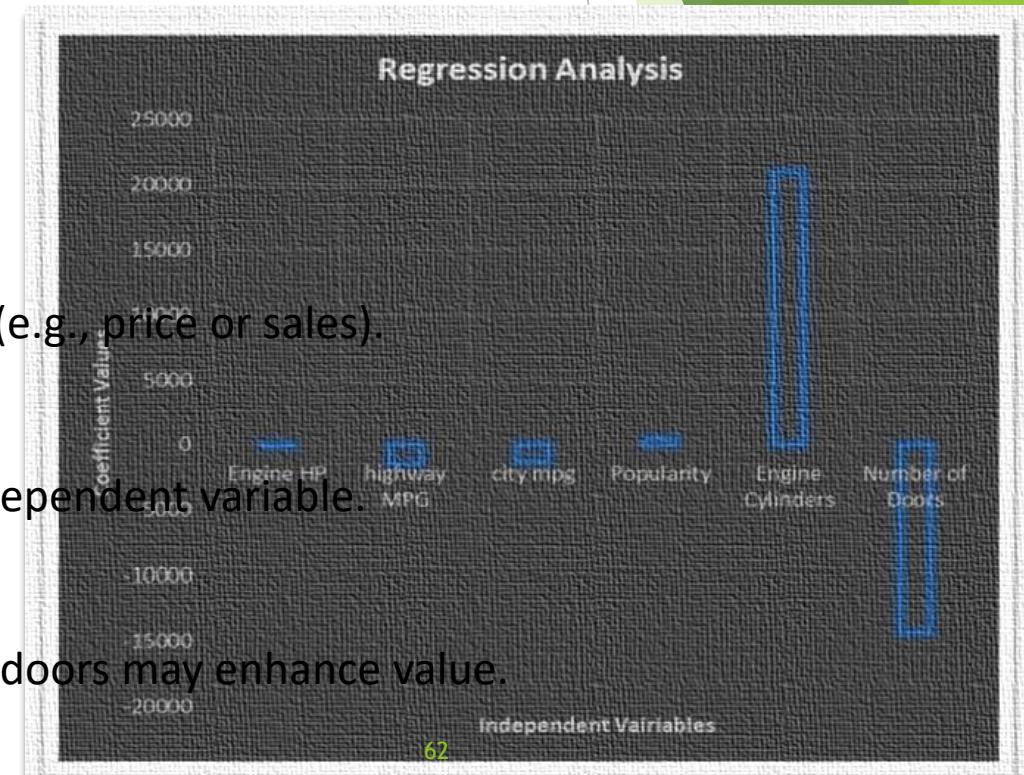
- "Engine Cylinders" coefficient is notably high.
- Strong positive correlation between the number of cylinders and the dependent variable.

### 4. Strong Impact of Number of Doors:

- "Number of Doors" coefficient is negative and substantial.
- More doors correlate with decreased prices or sales, suggesting fewer doors may enhance value.

### 5. Engine HP Impact:

- "Engine HP" coefficient is negative but less significant.
- Weaker negative correlation between engine horsepower and the dependent variable compared to other factors.



Insight Required 4: How does the average price of a car vary across different manufacturers?

- Task 4.A: Create a pivot table that shows the average price of cars for each manufacturer.
- Task 4.B: Create a bar chart or a horizontal stacked bar chart that visualizes the relationship between manufacturer and average price.

Solution for Task 4.A and 4.B:

Constructed a Bar Chart or Horizontal Stacked Bar Chart visualizing the relationship between manufacturers and their average prices.

The visualization aids in comparing and understanding the average prices offered by various manufacturers.

## Insights from Price Analysis

### 1. Luxury Manufacturers Dominate High Price Ranges:

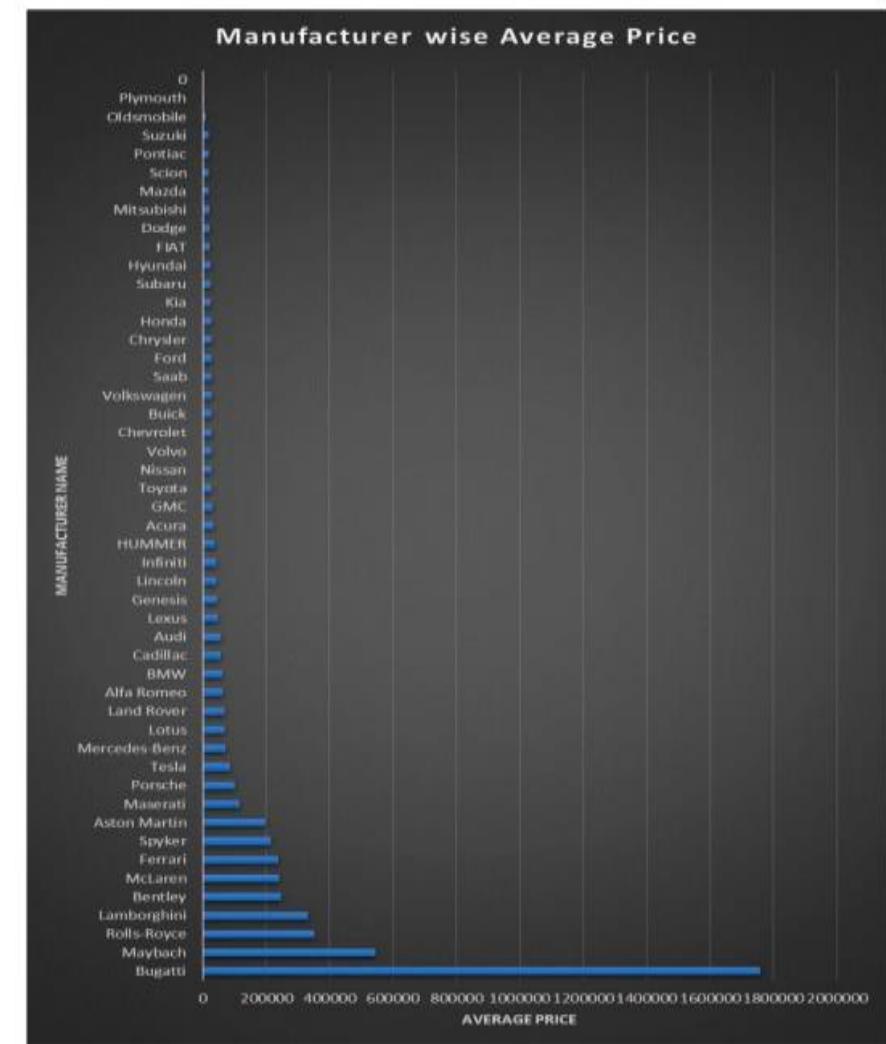
- Maybach, Rolls-Royce, Lamborghini, Bentley, and McLaren lead in average prices.
- Luxury and high-performance vehicles from these manufacturers command significantly higher prices, emphasizing the correlation between exclusivity and higher average prices.

### 2. Widespread Price Variation Across Brands:

- Substantial price disparity is evident across different manufacturers.
- The range spans from six-figure averages to more affordable options, reflecting diverse market segments and consumer preferences. This underscores the varied positioning of brands in terms of pricing.

### 3. Mainstream Brands and Average Prices:

- Mainstream manufacturers like Toyota, Nissan, Chevrolet, and Ford have lower average prices.
- Despite the lower pricing, these brands maintain larger market shares due to affordability and wider accessibility, appealing to a broader consumer base. The analysis highlights the balance between affordability and market share for mainstream brands.



Insight Required 5: 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

Solution of Task 5.A and 5.B:

Calculated the correlation coefficient between the number of cylinders and highway MPG.

This quantifies the strength and direction of the relationship between fuel efficiency and engine cylinders.

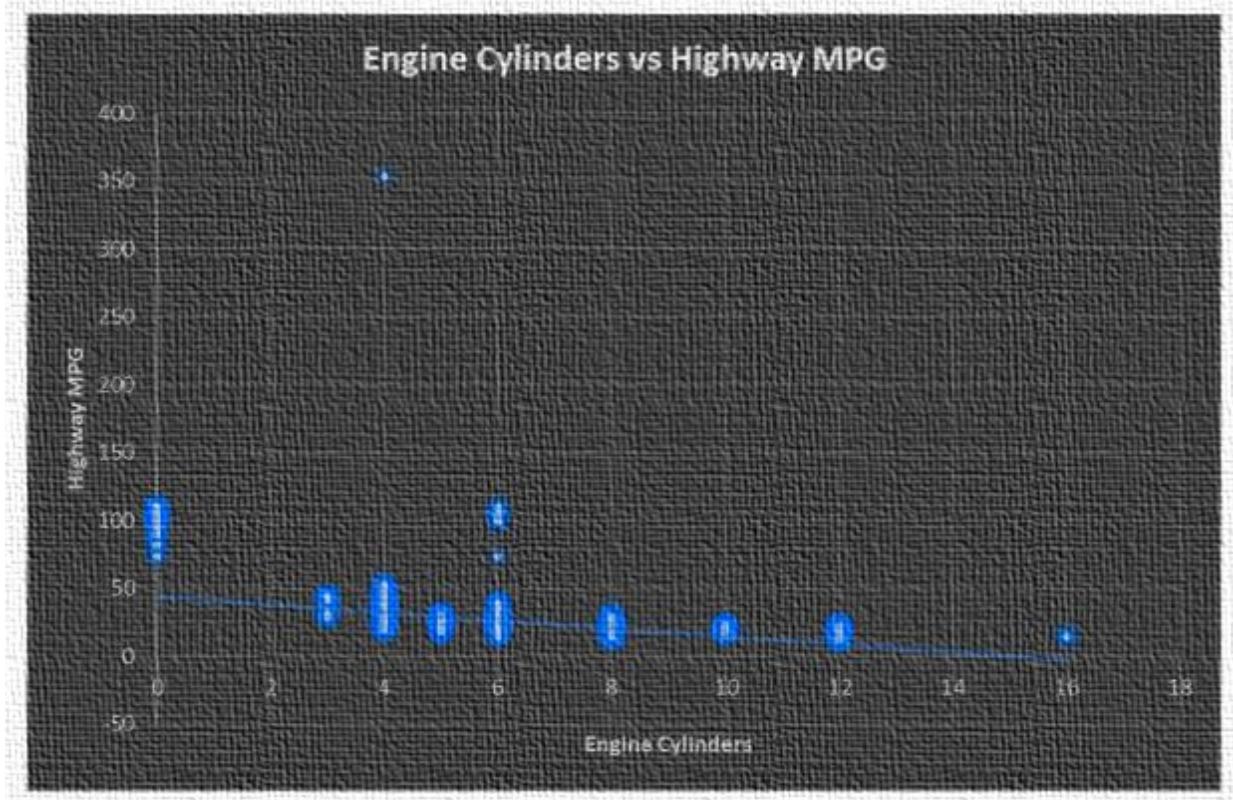
# Engine Cylinders vs. Highway MPG Analysis Insights

## 1. Negative Correlation:

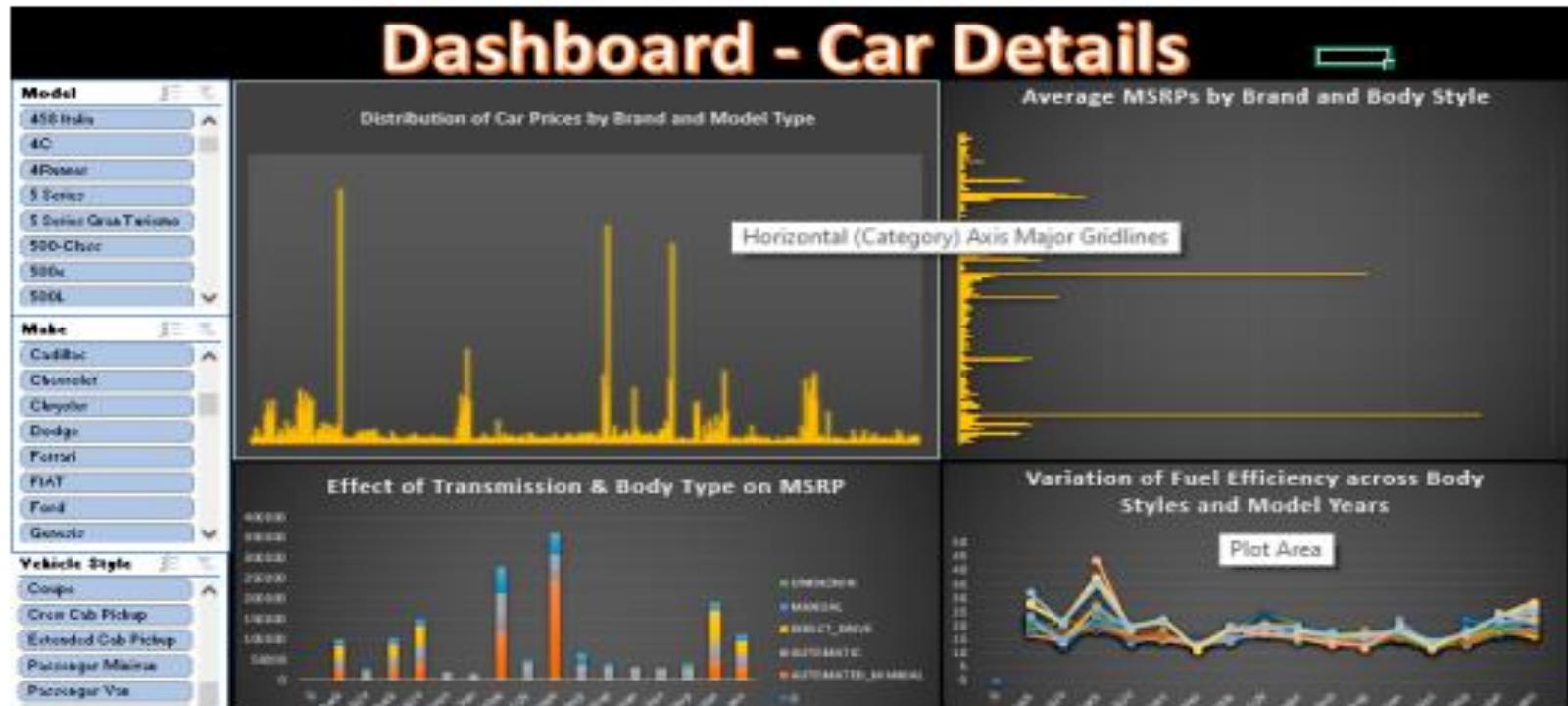
- There is a negative correlation between engine cylinders and highway MPG.
- As the number of cylinders increases, highway MPG tends to decrease.
- Larger engine displacement and increased power production in multi-cylinder engines contribute to higher fuel consumption.

## 2. Non-Linear Relationship:

- The relationship between engine cylinders and highway MPG is not linear.
- The decrease in highway MPG is not proportional to the increase in engine cylinders.
- For example, the difference in MPG between a 4-cylinder and a 6-cylinder engine may be more significant than the difference between a 6-cylinder and an 8-cylinder engine.



Visualization: Developed interactive dashboards using Excel, showcasing trends in car features, pricing, and consumer preferences.



## Conclusion:

The analysis provided valuable insights into consumer preferences, feature impact on pricing, and market segmentation. Recommendations include capitalizing on eco-friendly vehicle trends and optimizing pricing strategies to enhance profitability.

# Project 8

# ABC Call Volume Trend Analysis

## Project Description

- ABC Call Volume is a company which is used for customer experience by voice process.
- A customer experience team is made up of experts who evaluate customer feedback and data and communicate their findings with the rest of the business.
- This dataset consist of Inbound calls of a ABC company from the insurance category.
- Customer support is defined as the call centre responsible for processing customer incoming calls.
- Inbound calls are inbound voice calls from existing or prospective clients for your company.

# Data Understanding

- The dataset given below consist of
- **Customer Experience (CX)** Inbound calling team for 23 days.
- Data includes **Agent Name, Agent ID**
- **Queue Time** [duration for which customer have to wait before they get connected to an agent]
- **Time** [time at which call was made by customer in a day]
- **Time Bucket** [for easiness we have also provided you with the time bucket]
- **Duration** [duration for which a customer and executives are on call]
- **Call Seconds** [for simplicity we have also converted those time into seconds]
- **Call status** (Abandon, answered, transferred).

# **Tech-Stack Used**

- MS-Excel.  
Purpose : Used to filter and analyze the data.
- Canva.  
Purpose : Used to visualize the data.

# **Approach**

- Open data in MS-Excel.
- Firstly understood the problem and the data.
- Used the statistical formulas like count if , min , max , avg and like others.
- Analyzed data using pivot table and chart .

# Result & Insights

## 1. Calculate the average call time duration for all incoming calls received by agents (in each Time\_Bucket).

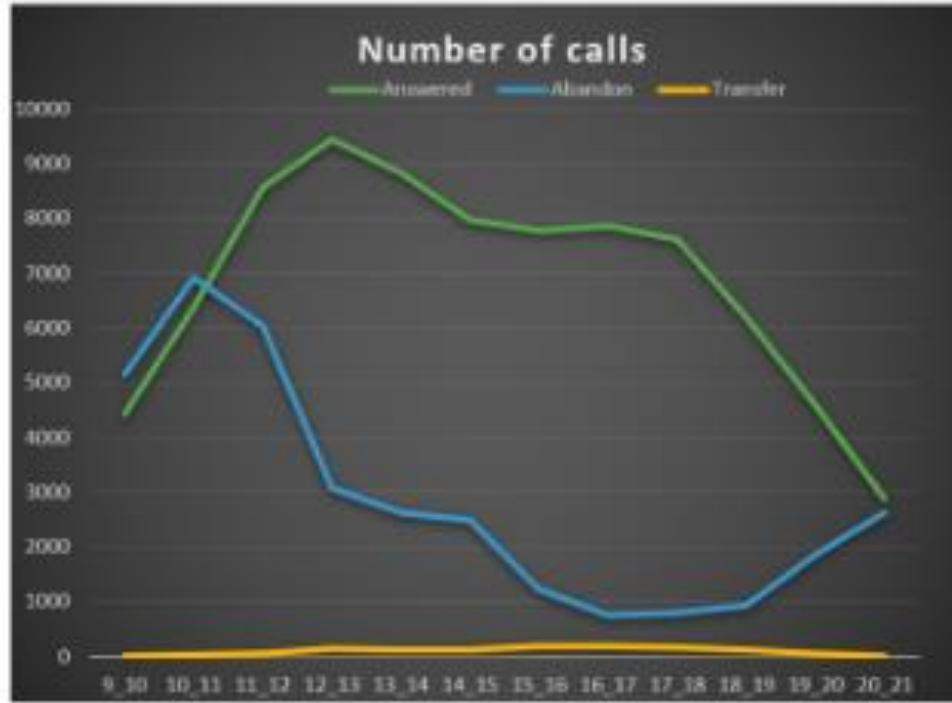
Time_Bucket	Average Seconds
9_10	92.01
10_11	97.42
11_12	116.78
12_13	144.73
13_14	149.54
14_15	146.97
15_16	169.9
16_17	181.44
17_18	179.72
18_19	174.32
19_20	144.58
20_21	105.95

- The average call time length for all phone calls received by agents is lengthy between 4 and 7 p.m.
- The interaction between customers and clients are high in evening.

## 2. Show the total volume/ number of calls coming in via charts/ graphs [Number of calls v/s Time].

Time_Bucket	Answered	Abandon	Transfer	Total
9_10	4428	5149	11	9588
10_11	6368	6911	34	13313
11_12	8560	6028	38	14626
12_13	9432	3073	147	12652
13_14	8829	2617	115	11561
14_15	7974	2475	112	10561
15_16	7760	1214	185	9159
16_17	7852	747	189	8788
17_18	7601	783	150	8534
18_19	6200	933	105	7238
19_20	4578	1848	37	6463
20_21	2870	2625	10	5505
Total	82452	34403	1133	117988

- The no of calls from customers to the call center is gradually increases from 9 am reaches to peak at 12 pm.
- After the peak time the call rate slowly decreases till night 10 pm.
- The abandon call rate by customer care is high in morning time 9 am - 12 pm.
- The no calls transferred is always maintained below at all day parts.



### 3. You have to calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.

Time_Bucket	No of Abandon calls	Abandon call rate
9_10	5149	53.7
10_11	6911	51.91
11_12	6028	41.21
12_13	3073	24.29
13_14	2617	22.64
14_15	2475	23.44
15_16	1214	13.25
16_17	747	8.5
17_18	783	9.18
18_19	933	12.89
19_20	1848	28.59
20_21	2625	47.68

- The Abandon call rate by agents is maximum above 30% in 9 am - 12 pm .
- The abandon called are also assumed to be agents was busy in another calls at time .
- So we want to increase no of agents working in the ABC Call company to decrease abandon call rate 10%.
- Agents Req Calculated =  $1.1 * \text{calls Each Day} * 2 * (\text{avg call duration} + \text{avg queue time})$ .
- So we want assume tolerance for no of calls and waiting time as 2 and 1.1 respectively.

Time_Bucket	No.of calls by customer	Average of Queue Times	Number of agents Answered	Average call duration	Calls on each day	New Agent required
9_10	9588	82.8627451	41	92.01	420	44.9
10_11	13313	83.25493878	50	97.42	580	64
11_12	14626	72.32421715	58	116.78	640	74
12_13	12652	41.66337338	59	144.73	550	62.6
13_14	11561	41.80313122	57	149.54	500	58.5
14_15	10561	43.59871224	59	146.97	460	53.6
15_16	9159	29.88361175	57	169.9	400	48.8
16_17	8788	23.53812016	57	181.44	380	47.6
17_18	8534	23.74548863	57	179.72	370	46
18_19	7238	34.08731694	58	174.32	310	39.5
19_20	6463	58.6941049	51	144.58	280	34.8
20_21	5505	75.27829246	26	105.95	240	26.6

- The average agent required in General shift is 50.
- If 50 agents newly added in call center the abandon call rate will become less than 10% percentage.

### 3. You have to calculate minimum number of agents required in each time bucket so that at least 90 calls should be answered out of 100.

Distribution of 30 calls coming in night for every 100 calls coming in between (9 pm - 9 am)	
Time_Bucket	Distribution of 30 calls
21_22	3
22_23	3
22-00	2
00_01	2
01_02	1
02_03	1
03_04	1
04_05	1
05_06	3
06_07	4
07_08	4
08_09	5

- The Shift is assumed as 8 hrs works for agents and 3 shifts is allotted in a day.
  - Call on each day for 9pm - 9am is calculated by formula given below.
  - $=ROUND(X * 30 / 70 * Y / 30, 0)$   
X = sum for calls on each day of time 9am - 9pm  
Y = Distribution of 30 calls.
  - Agent required calculation is same as previous slide above mentioned.
- 
- The average agent required in all 3 Shift is 70.

Shift Timing	Time_Bucket	Distribution of 30 calls	No.of calls by customer	Average of Queue Times	Number of agents Answered	Average call duration	Calls on each day	Agent required
Night Shift	00_01	2	3381	50.89		141.9	147	17.3
	01_02	1	1679	50.89		141.9	73	8.6
	02_03	1	1679	50.89		141.9	73	8.6
	03_04	1	1679	50.89		141.9	73	8.6
	04_05	1	1679	50.89		141.9	73	8.6
	05_06	3	5060	50.89		141.9	220	25.9
	06_07	4	6739	50.89		141.9	293	34.5
	07_08	4	6739	50.89		141.9	293	34.5
Morning Shift	08_09	5	8418	50.89		141.9	366	43.1
	09_10		9588	82.86	41	92.01	420	44.9
	10_11		13313	83.25	50	97.42	580	64
	11_12		14626	72.32	58	116.78	640	74
	12_13		12652	41.66	59	144.73	550	62.6
	13_14		11561	41.8	57	149.54	500	58.5
	14_15		10561	43.6	59	146.97	460	53.6
	15_16		9159	29.88	57	169.9	400	48.8
	16_17		8788	23.54	57	181.44	380	47.6
	17_18		8534	23.75	57	179.72	370	46
Evening Shift	18_19		7238	34.09	58	174.32	310	39.5
	19_20		6463	58.69	51	144.58	280	34.8
	20_21		5505	75.28	26	105.95	240	26.6
	21_22	3	5060	50.89		141.9	220	25.9
	22_23	3	5060	50.89		141.9	220	25.9
	22-00	2	3381	50.89		141.9	147	17.3

# Result

- I learned how to analyze and how to not analyze the Call center data .
- We should increase man power in all shift .
- We should avoid to Queue time for the customer.
- IVR Duration can be improvised little bit.

# Final Project and Appendix

- ▶ Final project is this report which compiles all the work done for different projects
- ▶ List of project in Data Analysis Project portfolio:
  1. Data Analytics Process
  2. Link:- <https://drive.google.com/file/d/1DT0ligEezkaJXrzXNySZSPHa3MPx9KMI/view?usp=sharing>
  3. Instagram User Analytics
  4. Link:- <https://drive.google.com/file/d/1Rxevl1xwd2LCoeOEAbjsGGFPkHbNi73B/view?usp=sharing>
  5. Operation and Metric Analytics
  6. Link:- <https://drive.google.com/file/d/1YeQU5pnNPzeFPwnzZXFAOHFXRccV9YI4/view?usp=sharing>
  7. Hiring Process Analytics
  8. Link:- <https://drive.google.com/file/d/14hW3Z31bfOFnjXfhenIJcFUr2rQEihvl/view?usp=sharing>
  9. IMDB Movie Analysis
  10. Link:- <https://drive.google.com/file/d/1xXCK3w-Um9alZV55mCA8BDFIXPFOvzkM/view?usp=sharing>

## 6. Bank Loan Case Study

Link:- [https://drive.google.com/file/d/11\\_eTtawKVKwMuLX50WmDajnC6Yq46JaR/view?usp=sharing](https://drive.google.com/file/d/11_eTtawKVKwMuLX50WmDajnC6Yq46JaR/view?usp=sharing)

## 7. Impact of Car Features

Link:- <https://drive.google.com/file/d/1w3tfzv8FAj5mTj-erv5bxhKZ6HU-zIUZ/view?usp=sharing>

## 8. ABC Call Volume Trend

Link:- <https://drive.google.com/file/d/1nwHbK1e6U0C8H6UihiboG3PZx5kprLECi/view?usp=sharing>

- ▶ Github Profile:- <https://github.com/Ashu2887/Trainity-Internship>
- ▶ LinkedIn Profile:- <http://www.linkedin.com/in/ashutoshmishra4>
- ▶ Google Drive: - <https://drive.google.com/drive/folders/10qZCJewzBd6P-c5burQRxzgVpiJ0bj-Y?usp=sharing>

# My Learnings

## ► Project 1 – Data Analytics Process

- *Learnings:* This project provided a comprehensive understanding of the end-to-end data analytics process, from data collection and cleaning to analysis and visualization. The insights gained in this project laid the foundation for subsequent projects, emphasizing the importance of a structured analytics workflow.

## ► Project 2 – Instagram User Analytics

- *Learnings:* Explored techniques for social media analytics using Python and Excel. Analyzed user behavior, engagement patterns, and content preferences. Insights gained contributed to a deeper understanding of audience targeting and content strategy.

## ► Project 3 – Operation and Metric Analytics

- *Learnings:* This project involved the analysis of operational data and metrics, showcasing the application of analytics in optimizing processes. Insights obtained highlighted areas for improvement in efficiency and resource allocation, emphasizing the strategic use of analytics in operations.

## ► Project 4 – Hiring Process Analytics

- *Learnings:* Explored the application of analytics in human resources. Analyzed hiring trends, time-to-fill metrics, and candidate success factors. This project provided insights into optimizing the hiring process, improving candidate experience, and making data-driven decisions in talent acquisition.

## **Project 5 – IMDB Movie Analysis**

•*Learnings:* Utilized Python for in-depth analysis of IMDB movie data. Explored trends in genres, ratings, and box office performance. Insights gained were crucial for understanding audience preferences and informed decision-making for future film projects or recommendations.

## **Project 6 – Bank Loan Case Study**

•*Learnings:* Analyzed data related to bank loans, focusing on risk assessment and customer profiling. Insights gained played a pivotal role in refining lending strategies, minimizing risks, and enhancing overall decision-making processes within the banking domain.

## **Project 7 – Impact of Car Features**

•*Learnings:* Explored the influence of various car features on customer satisfaction and sales. Utilized both Excel and Python for data analysis. Insights gained helped in tailoring marketing strategies, product development, and understanding customer preferences in the automotive industry.

## **Project 8 – ABC Call Volume Trend**

•*Learnings:* Analyzed call volume trends using Python and Excel, aiming to optimize customer support operations. Insights obtained facilitated improved resource allocation, reduced wait times, and enhanced overall customer satisfaction. The project demonstrated the significance of analytics in improving service efficiency.

These projects collectively showcase proficiency in data analysis using Excel and Python, emphasizing the application of insights to drive informed decision-making. The use of PowerPoint and Word in creating reports and presentations enhances communication and knowledge dissemination across various stakeholders.

# Thank You