



# Project Portfolio : Data Analytics Course

---

**Submitted By**

**Swathy Mugundan**

**Contact : +974 - 7107 2857**

**Email : [swat.mugund97@gmail.com](mailto:swat.mugund97@gmail.com)**

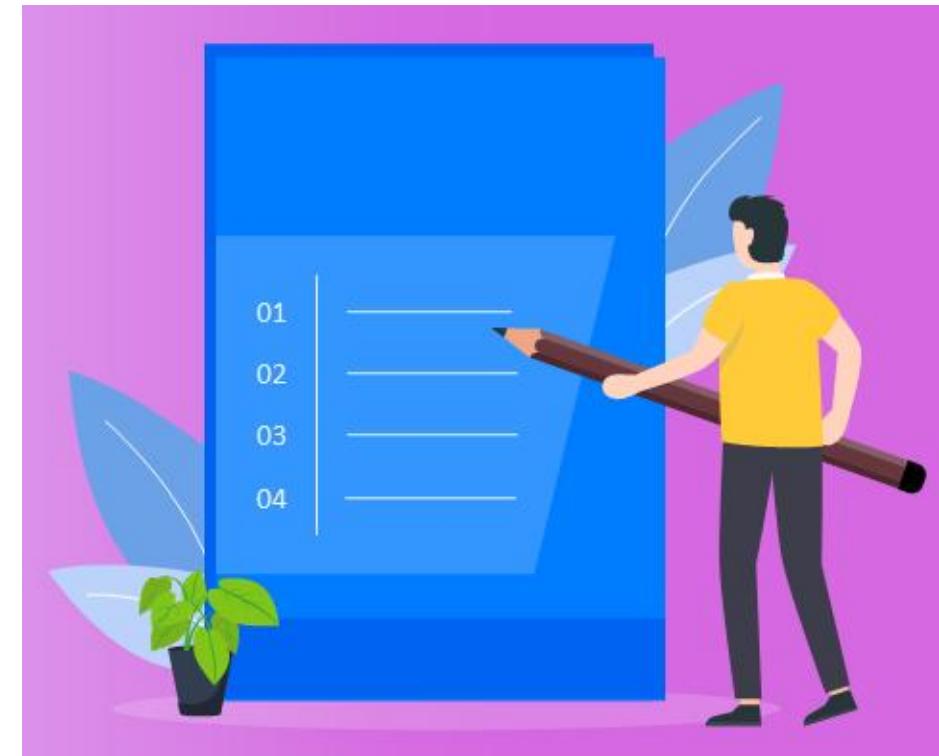




# Project Portfolio : Contents

---

- Professional Background
- Module Descriptions
  - Objective
  - Implementation Screenshots
  - Key Takeaways
- Module References
- Conclusion



# PROFESSIONAL BACKGROUND

# Professional Background



- CS Aspirant - looking for a position to enhance skills in various domains. Adept to user experience for Innovative ideas. Punctual and pursue good communication skills. Keen interest in updating knowledge and have exhibited leadership skills in organising various events.
- Highly enthusiastic and eager to meet challenges and quick to learn, assimilate new concepts and ideas. Have completed my B.Tech (CSE) and M.Tech (CSE) with majoring in Distributed Computing Systems. I have also cleared the UGC-NET Examination in December 2021.
- Have contributed to research through several publications and mini-projects during my course of study.
- Having 2 years of Industry experience as a software developer and 1 Year of experience as Assistant Professor.
- Hoping to be a part of passionate work environment, where I can showcase technical skills to accomplish organisational goals.
- **Currently relocated to Doha, Qatar and hence looking for opportunities at Qatar or for Work from Home Jobs in India.**

# Professional Background - Work Experience



<b>Assistant Professor (CSE)</b> PES University, Bangalore	<b>June 2022 - June 2023</b>
Worked as Assistant Professor in CSE Department with class coordinator and mentor responsibilities. Have been a part of curriculum meetings. Subjects: Python Programming, C Programming, Statistics for DS, Web Development and Lab Components	
<b>Software Developer</b> ZF-Wabco, Bangalore	<b>Jan 2021 - June 2022</b>
Worked as a Front-End Software Developer for developing UI applications for autonomous vehicles. Tech Stack : ReactJS, React Native, HTML, CSS, GIT, Agile, SQL	
<b>Software Developer - Intern</b> HP Inc., R and D, Bangalore	<b>Jan 2019 - June 2019</b>
Worked as Software Developer Intern for Printer Drivers to manage multiple print jobs simultaneously in a generic driver schema with synchronisation mechanisms with C++	
<b>Algorithm Development</b> CENTAC, Govt. of Puducherry	<b>April 2018 - June 2018</b>
Worked as a member of Developer team to implement online admission system for the UG admissions through common merit scheme for State Govt. Counselling System with PHP	
<b>Website Developer - Internship</b> DAT, Govt. of Puducherry	<b>Dec 2017 - Jan 2018</b>
Worked as a developer to implement automatic pay calculations of IAS and IPS officials with Visual C and MySQL	
<b>Core JAVA</b> Internshala - Online Internship	<b>December 2016</b>
Online Internship Course which comprises of Java basics, OOPS, Swing and AWT	

## LIST OF MINI-PROJECTS

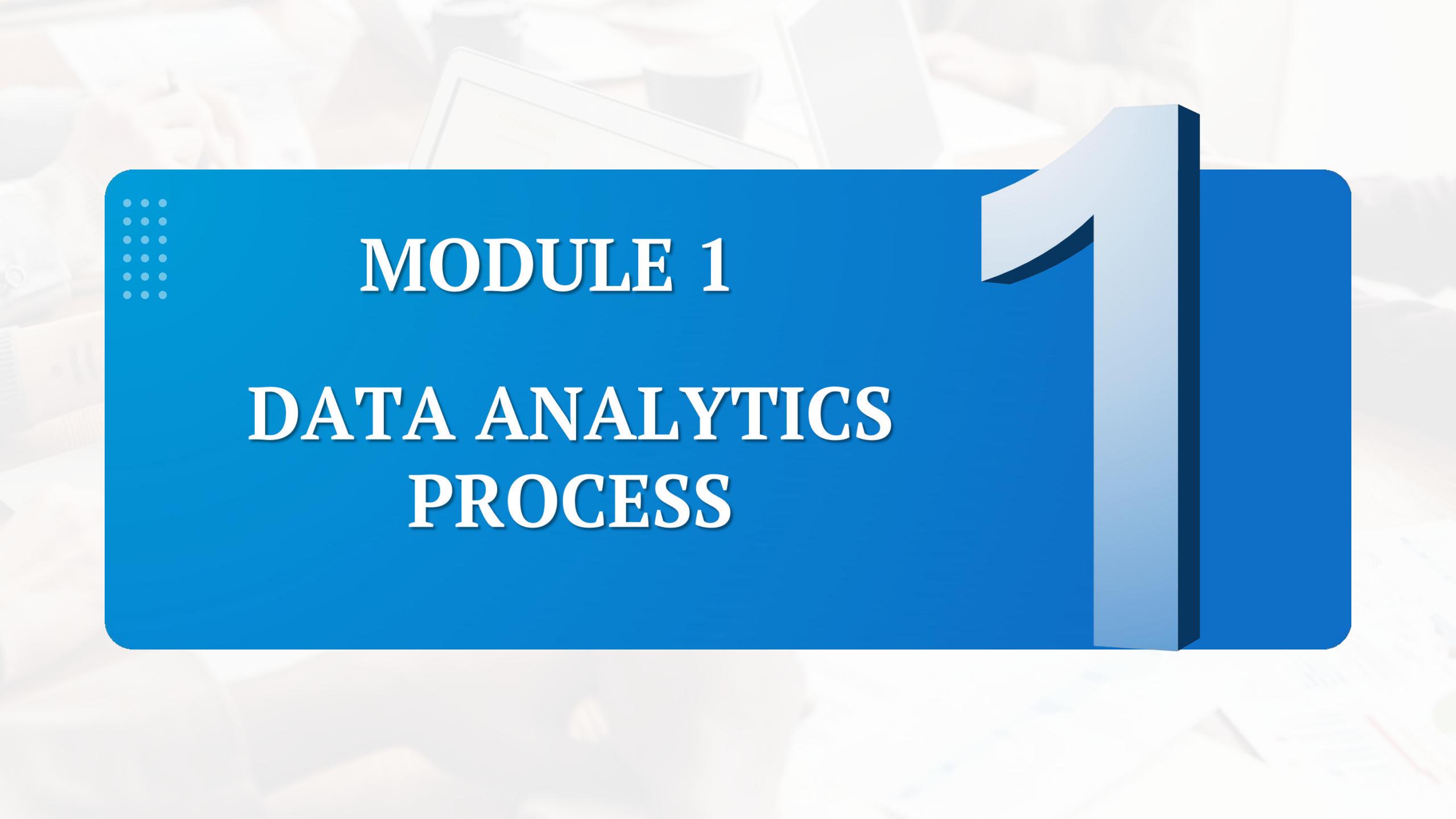
<b>Prognosis Prediction Model</b> PG - Final Year Project	<b>2019 - 2021</b>
Developed for CVD Prediction using Feature Fusion Techniques, Attention Mechanism and Deep Neural Networks	
<b>Predictive Analytics</b> UG - Final Year Project	<b>2019</b>
Developed for Heterogeneous Web-Application Vulnerability Detection and Prevention	
<b>Burglar Alarm</b> Embedded Systems Project (H/W)	<b>2018</b>
Developed for security purpose using Arduino Kit and Sensors	
<b>MASTECH</b> Android App Development	<b>2018</b>
Developed for on-campus placement training - Aptitude and Logical Reasoning, Core CS Concepts MCQ's with Android Studio	
<b>Feedback System</b> Web Development	<b>2017</b>
Developed for automating feedback system procedures of CS Department using PHP, MySQL, JS	

# Professional Background - Publications



## Publications

- A Finite State Machine Based Model for Attack Detection and Vulnerability Assessment, *International Journal of Pure and Applied Mathematics*, Vol.116, November 2017: 649-651.
- A Prediction Model for Heterogeneous Service Composition Using Decision Making and Capability Assessment Algorithm *International Journal of Pure and Applied Mathematics*, Vol.116, Nov'17:325-339.
- Validation Model for Heterogeneous Service Composition using Constraint Satisfaction Algorithm- Advances in Computational System *InderScience Journal*, Communicated.
- Effective AppState Maintenance of RESTful Web Services using Semantic Rules and Syntax Analysis *IEEE Journal*, Communicated.
- A Comparative Study of Classification and Prediction of Cardio-Vascular Diseases (CVD) using Machine Learning and Deep Learning Techniques *SCIE Indexed, Elsevier*, <https://doi.org/10.1016/j.icte.2021.08.021> (Impact Factor:4.17).



# **MODULE 1**

# **DATA ANALYTICS**

# **PROCESS**

# Module 1 : Data Analytics Process

## Objective / Learning

- Importance of Data Analytics (DA)
- Various approaches to DA
- Different stages of DA Process
- Understanding a problem statement and applying the process of DA.



# Module 1 : Data Analytics Process



## Step : 1 : Plan (Ask)

Define the problem by effectively asking questions and collaboration



## Case Study Implementation

Decide the subject and the standard for whom the syllabus and classes has to be scheduled.  
(Eg. Computer Science - Class 9)



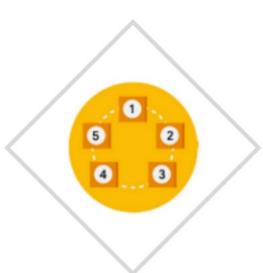
## Step : 2 : Prepare

Prepare data by collecting and storing information based on previous step



## Case Study Implementation

Check for how many hours per week the subject is scheduled for the particular class.  
Calculate the theory and practical hours for the subject.



## Step : 3 : Process

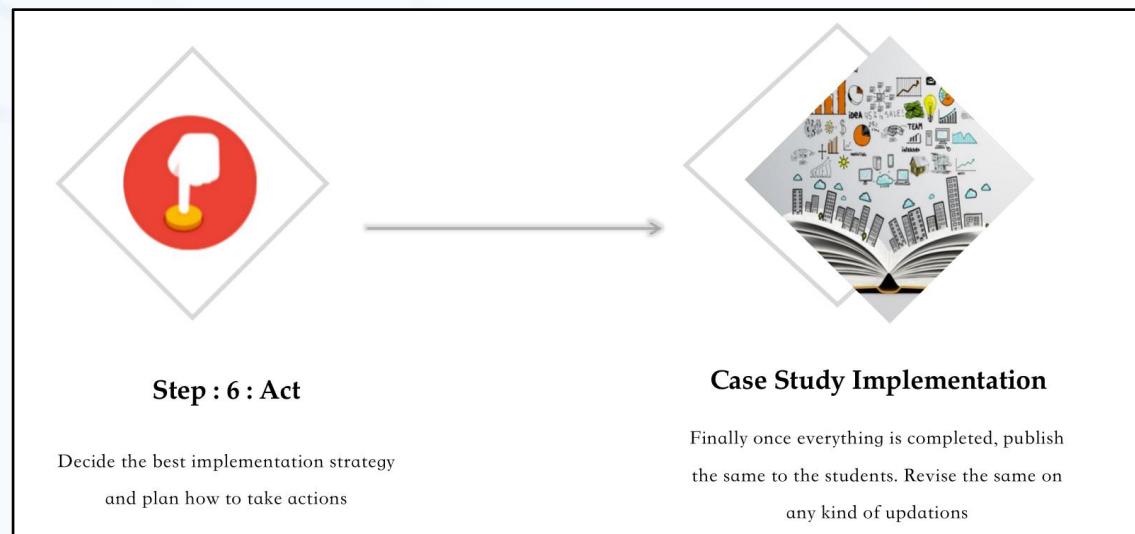
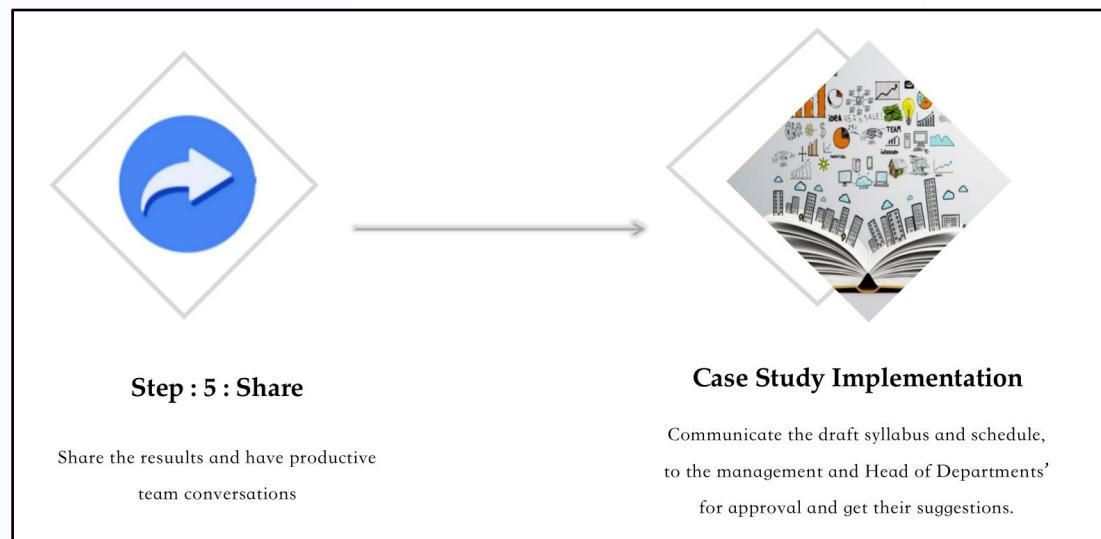
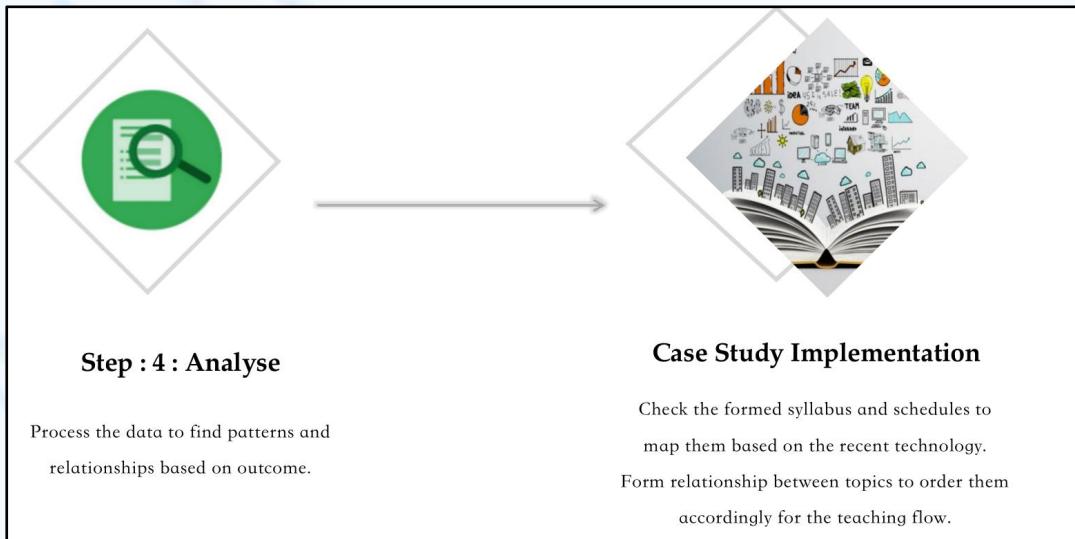
Process the data based on the outcome



## Case Study Implementation

Check the outcome of the subject for the particular class. (Eg. Computer Science - What sub-domain to be chosen ? MS-Office / Python /Web Development / Data Science/ etc.,?)

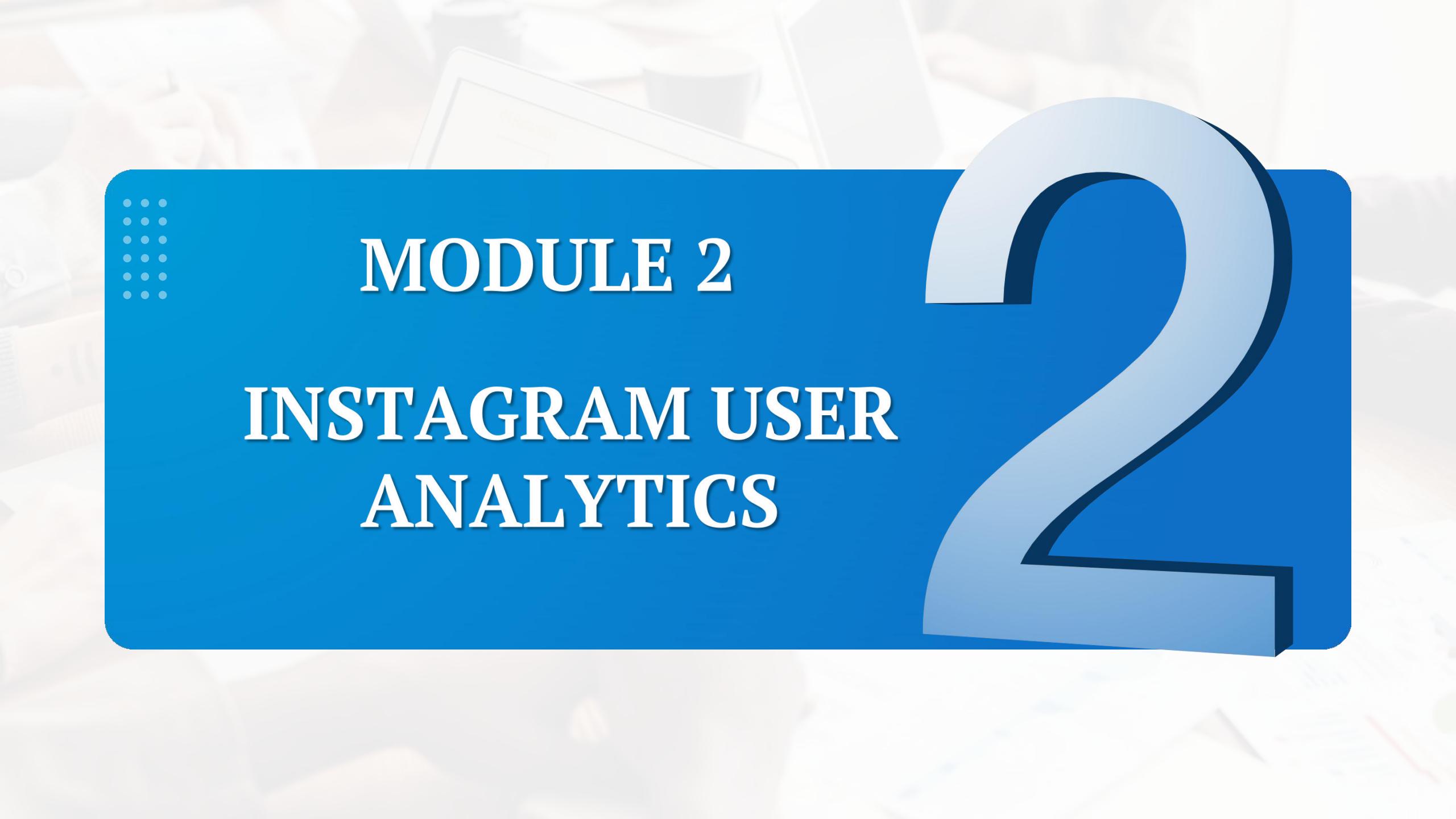
# Module 1 : Data Analytics Process



# Module 1 : Key Takeaways



- Data analytics is the science of analyzing raw data to make conclusions about that information.
- It helps a business optimize its performance, perform more efficiently, maximize profit, or make more strategically-guided decisions.
- The techniques and processes of data analytics have been automated into mechanical processes and algorithms that work over raw data for human consumption.
- Various approaches to data analytics include looking at what happened (descriptive analytics), why something happened (diagnostic analytics), what is going to happen (predictive analytics), or what should be done next (prescriptive analytics).
- Data analytics relies on a variety of software tools including spreadsheets, data visualization, reporting tools, data mining programs, and open-source languages for the greatest data manipulation.



MODULE 2

# INSTAGRAM USER ANALYTICS

2

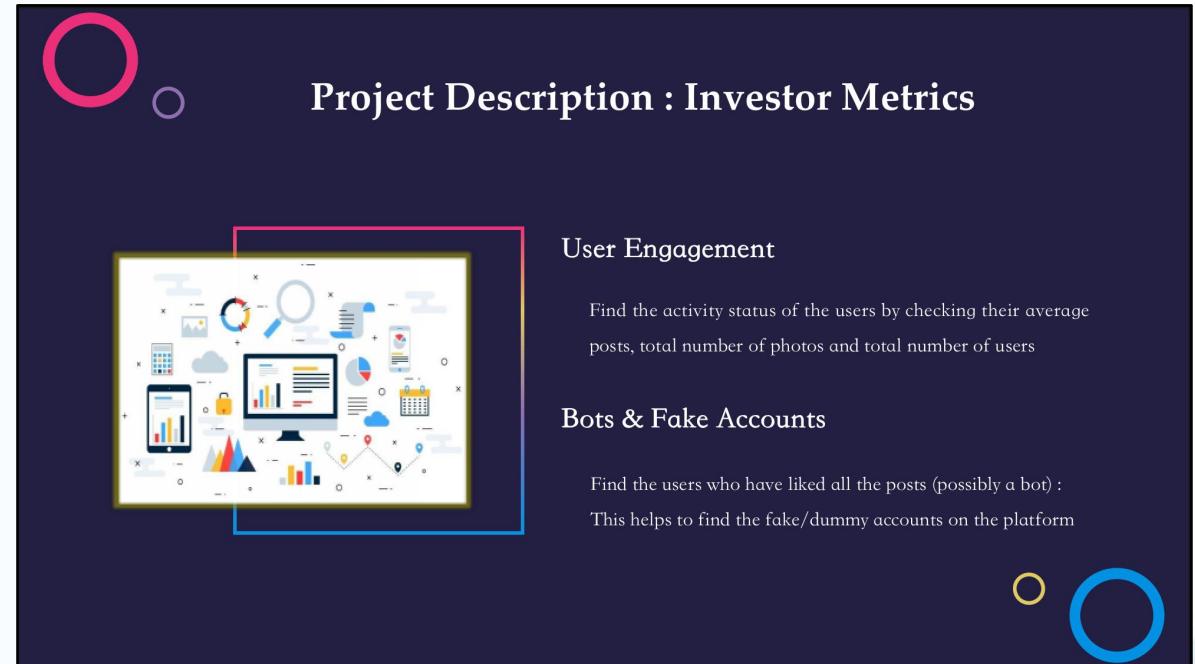
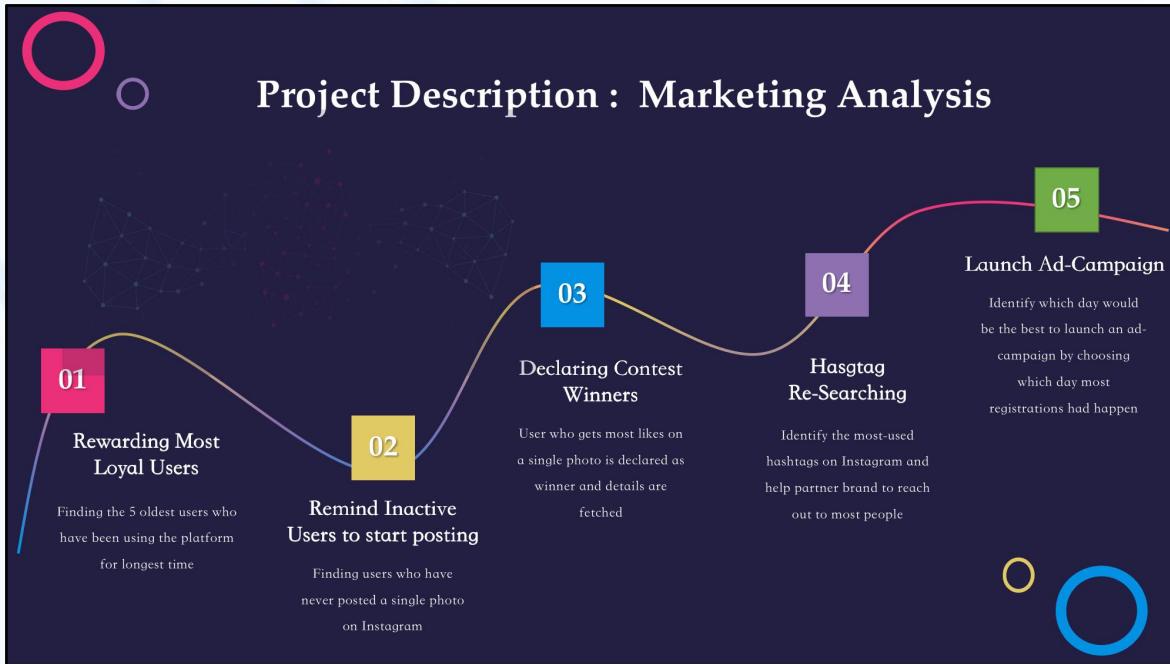
# Module 2 : Instagram User Analytics

## Objective / Learning

- Understanding MySQL and Queries
- Managing data through MySQL Workbench
- CURD Operations
- Performing two tasks : Marketing Analysis and Investor Metrics
- Analysing result sets for inferences



# Module 2 : Instagram User Analytics



## PROJECT DESCRIPTION AND TASKS

### MARKETING ANALYSIS AND INVESTOR METRICS

# Module 2 : Instagram User Analytics



Project 2 : DB-Tables Creation Screenshots

Comments Table

Likes Table

Follows Table

Tags Table

Photo Tags Table

11 12 13 14 15

## PROJECT TASKS SAMPLE SCREENSHOTS IN SLIDE SORTER VIEW

Job Data Analysis : Jobs Reviewed over Time

```
select ds as Date, count(ds) as Jobs, sum(time_spent)/3600 as Hours_Spent from job_data group by ds;
```

NOTE: The given time was in seconds. Hence, converted to hours by dividing it by 3600. Also, all the data are in month of November. So where clause between the date is omitted.

16

Job Data Analysis : Throughput Analysis

```
SELECT ds as Date, job_id as Job_ID, AVG(time_spent) OVER (ORDER BY ds ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS Rolling_Average FROM job_data;
```

NOTE: In order to track the trends in analysis, Rolling Averages are preferred as the daily metric can have unwanted noise in the data. Also, here to track the 7-day pattern (specific time series) as mentioned rolling average is best to be used to spot the trend.

17

Job Data Analysis : Language Share Analysis

```
SELECT language as Lang, count(language) as COUNT, 6/count(language) as SHARE from job_data group by language;
```

NOTE: The total number of languages is 6.

18

Job Data Analysis : Duplicate Rows Detection

```
SELECT job_id, actor_id FROM job_data GROUP BY job_id, actor_id HAVING (COUNT(job_id)>1);
```

NOTE: Returns an empty result-set depicting NULL values. From this, we can conclude that there are no duplicate rows in the job\_data table.

19

Case Study 2 : Investigating Metric Spike Database Creation Screenshots

20

# Module 2 : Key Takeaways



- MySQL is an open-source relational database management system that can be used when working with structured or unstructured data across platforms and operating systems.
- Through a combination of products and services, such as **MySQL Workbench**, this database can be used for data modeling, design, visualization, and development.
- Inclusive of multiple data types, MySQL is also widely recognized as the top tool for data scientists and developers.
- Data Analysts use SQL to manipulate the data, understand the data, and access, read and analyze the data before storing it in the database.
- It consists of a set of actions, which when executed, assemble a set of data, and rely on targeted techniques performed over the data to get valuable insights.
- The set of general actions include **Select, Update, Delete, Add, Modify, Alter**, and some clauses.
- Analysts don't have to copy data into other programs because they can access enormous amounts of data immediately where it's kept.
- Data Analysis done in SQL is easy to audit and reproduce when compared to spreadsheet methods which eliminates the need for analysts to search for the cell containing the formula error.



# MODULE 3

# OPERATION AND METRIC ANALYTICS



# Module 3 : Operation & Metric Analytics

## Objective / Learning

- Understanding MySQL and Queries
- Managing data through MySQL Workbench
- CURD Operations
- Performing two tasks : Job Data Analysis and Investigating Metric Spike
- Analysing result sets for inferences



# Module 3 : Operation and Metric Analytics




**Jobs Reviewed Over Time**

To calculate the number of jobs reviewed per hour for each day in November 2020.

**Throughput Analysis**

To calculate the 7-day rolling average of throughput (number of events per second).

**Language Share Analysis**

To calculate the percentage share of each language in the last 30 days.

**Duplicate Rows Detection**

Identify duplicate rows in the data.

**Weekly User Engagement**

Measure the activeness of users on a weekly basis.

**Weekly Retention Analysis**

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

**Email Engagement Analysis**

Analyze how users are engaging with the email service.

**User Growth Analysis**

Analyze the growth of users over time for a product on weekly basis

**Weekly Engagement (per Device)**

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

## PROJECT DESCRIPTION AND TASKS

### JOB DATA ANALYSIS AND INVESTIGATING METRIC SPIKE

# Module 3 : Operation and Metric Analytics



7 Database Description

- Table name `job_data` will be used for the analysis.
- The attributes are explained as follows :
  - `job_id`: Unique identifier of job.
  - `actor_id`: Unique identifier of actor.
  - `event_id`: Type of event (decision/skip/transfers).
  - `language`: The language of the content.
  - `time_spent`: Time spent to review the job in seconds.
  - `org`: The organization of the actor.
  - `dt`: The date in the format yyyy/mm/dd (stored as text).

Job Data CSV Screenshot

8 Project Approach

Job Reviewed Over Time

Thesaurus Analysis

Language Share Analysis

Duplicate Rows Detection

9 Job Data Analysis : Database & Table Creation Screenshot

A Database named "Job\_Data" is created for the Case Study 1. Job\_Data and the data is loaded using the Table One Import Wizard from the file.

10 Job Data Analysis : Job Reviewed Over Time

select as Date, count(job\_id) as Jobs, sum(time\_spent)/3600 as Hours\_Spent from job\_data group by date;

NOTE: The given time was in seconds. Hence, converted that to hours by dividing by 3600. Also, all the data set is from November. So when clause between the date is omitted.

11 Job Data Analysis : Throughput Analysis (Rolling Average)

SELECT as Date, job\_id, Rolling\_Average, COUNT(language) as COUNT, COUNT(language)\*100/6 as SHARE from job\_data group by language;

NOTE: In order to track the trends in analysis, Rolling Average is preferred as the daily metric has more fluctuations. Hence, it is better to use the rolling average of the 7-day pattern (specific time series) as mentioned rolling average is best to be used to spot the trend.

12 Job Data Analysis : Language Share Analysis

Result Grid

Result Inference: Persian Language has a maximum share in the data.

## PROJECT TASKS SAMPLE SCREENSHOTS IN SLIDE SORTER VIEW

13 Case Study 2 : alter email\_events table datatype text to datetime

```
SURSQL$MSH_UPDATES = 0;
alter table email_events add column temp_datetime;
update email_events set temp_datetime = TO_DATE('1970-01-01 00:00:00', 'YYYY-MM-DD HH:MI:SS');
alter table email_events drop column occurred_at;
alter table email_events change column temp_occurred_at DATETIME;
select * from email_events;
```

14 Case Study 2 : Weekly User Engagement

```
select extractweek from occurred_at as Week_Num,
count(event_id) as Users_Count
from events
where
  event_type = 'Engagement'
  AND event_name = 'Topic'
group by Week_Num;
```

15 Case Study 2 : User Growth Analysis

```
select week_num as Week, year_num as Year, active_users as Users,
sum(active_user) over (order by year_num, week_num) rows between unbounded preceding and current row) as Cumulative
from (select count(DISTINCT user_id) as active_users,
extractweek from occurred_at as week_num,
extractyear from occurred_at as year_num
from users where state='Active'
group by year_num, week_num, order by year_num, week_num);
NOTE: The result set describes the number of users who have logged in weekly basis. The maxteam is week_40 and minimum is week_35.
```

16 Case Study 2 : User Growth Analysis - Result Set(1)

THE RESULT SET OF THE QUERY IS TOO LARGE ATTACHED SCREENSHOT IS IN THE FOLLOWING SLIDES PLEASE REFER NEXT 2 SLIDES

17 Case Study 2 : User Growth Analysis - Result Set(2)

RESULTS FOR THE YEAR 2014

18 Case Study 2 : Weekly Retention Analysis

```
select LOGIN_WEEK,
  SEMICAR_WHEN_RETENTION = 0 THEN 1 ELSE 1-END-SWEEK_5,
  SEMICAR_WHEN_RETENTION = 1 THEN 1 ELSE 1-END-SWEEK_4,
  SEMICAR_WHEN_RETENTION = 2 THEN 1 ELSE 1-END-SWEEK_3,
  SEMICAR_WHEN_RETENTION = 3 THEN 1 ELSE 1-END-SWEEK_2,
  SEMICAR_WHEN_RETENTION = 4 THEN 1 ELSE 1-END-SWEEK_1,
  SEMICAR_WHEN_RETENTION = 5 THEN 1 ELSE 1-END-SWEEK_0,
  SEMICAR_WHEN_RETENTION = 6 THEN 1 ELSE 1-END-SWEEK_1,
  SEMICAR_WHEN_RETENTION = 7 THEN 1 ELSE 1-END-SWEEK_2,
  SEMICAR_WHEN_RETENTION = 8 THEN 1 ELSE 1-END-SWEEK_3,
  SEMICAR_WHEN_RETENTION = 9 THEN 1 ELSE 1-END-SWEEK_4,
  SEMICAR_WHEN_RETENTION = 10 THEN 1 ELSE 1-END-SWEEK_5;
```

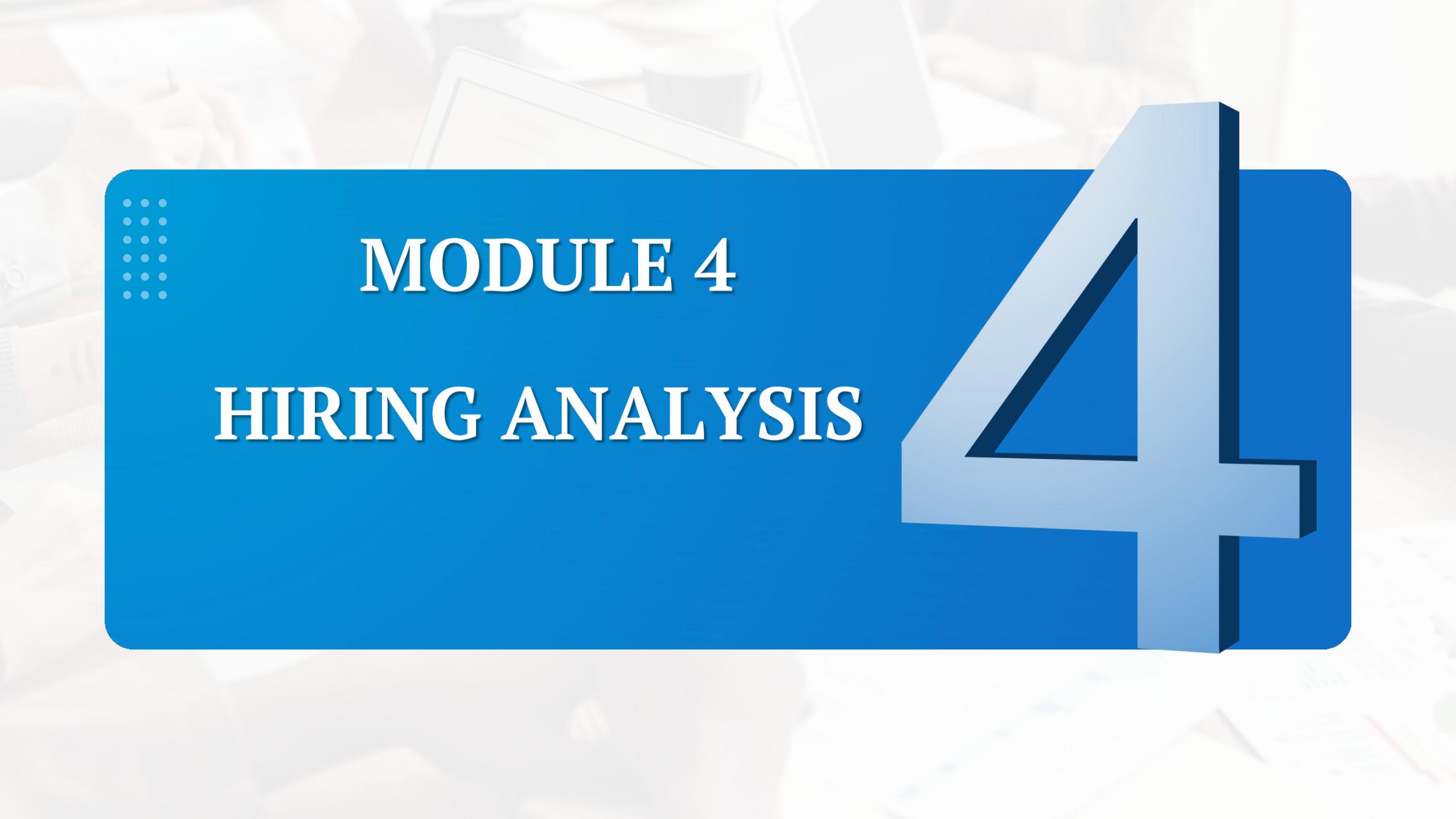
QUERY BREAKDOWN: CROSSTAB FORMATION WITH THE RETENTION USERS  
(Refer Next Slide for the Sub-Query)

```
select user_id AS User_ID, Signup_Week AS SIGNUP_WEEK, Login_Week AS LOGIN_WEEK,
  Signup_Week - Login_Week AS RETENTION from
```

# Module 3 : Key Takeaways



- Operational Analytics is a crucial process that involves analyzing a company's end-to-end operations.
- This analysis helps identify areas for improvement within the company.
- This can be used to automate these requests and sync that data directly to your operational systems.
- One of the key aspects of Operational Analytics is investigating metric spikes.
- This involves understanding and explaining sudden changes in key metrics, such as a dip in daily user engagement or a drop in sales.
- With Operational Analytics, your business teams don't have to worry about getting access to your data because all the information they need is directly at their fingertips, allowing them to build personalized experiences for your customers at scale.



# MODULE 4

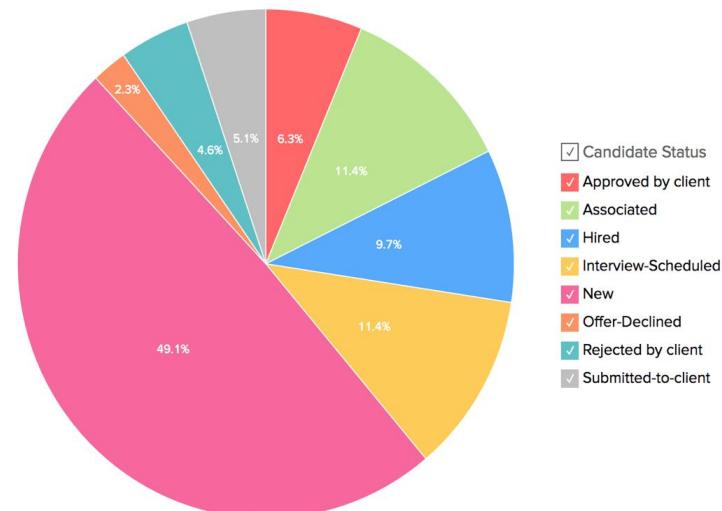
# HIRING ANALYSIS



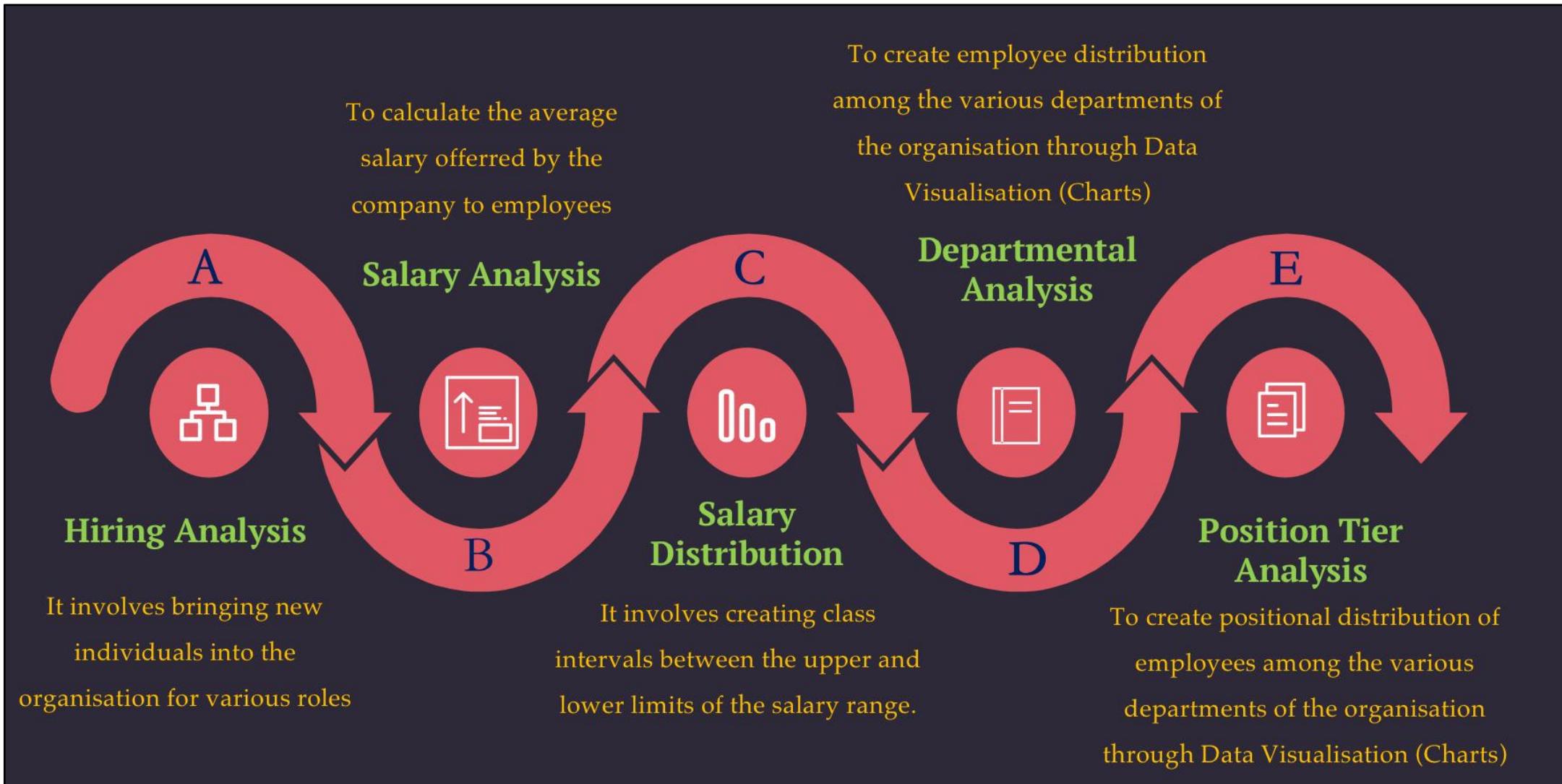
# Module 4 : Hiring Analysis

## Objective / Learning

- Understanding Statistics and MS-Excel
- Managing data through MS-Excel
- Performing data analysis through mathematical formulas
- Analysing result for improving hiring process and make informed decisions in future



# Module 4 : Hiring Analysis - Tasks



# Module 4 : Hiring Analysis



**Case Study : Hiring Analysis (Q.A)**

[CLICK HERE FOR THE SHEET](#)

**Objective :** To find the total number of male and female employees hired by the company

**Approach 1 :** Formula is used to count the number of male and female employees in the range of data.

Formula : `=COUNTIF($B:$B,"Male") + COUNTIF($B:$B,"Female")`

**Result**

Total Male Hired : 2563
Total Female Hired : 1856

**Case Study : Hiring Analysis (Q.A)**

[CLICK HERE FOR THE SHEET](#)

**Objective :** To find the total number of male and female employees hired by the company

**Approach :** A pivot-table has been created based on the fields shown in the image. This pivot table is used to find the count of male and female employees. Based on this data, a suitable pie-chart is made to depict the distribution among various departments.

Note : The formula is applied to cells J5 and J6. Sheet Name : Input\_Data

**Result**

Total Male Hired : 2563
Total Female Hired : 1856

**Case Study : Hiring Analysis (Q.A)**

[CLICK HERE FOR THE SHEET](#)

**PivotTable Fields**

**MALE EMPLOYEE**

Status	Hired
Count	2563
Department	Count
Finance Department	354
General Management	113
Human Resource Department	28
Marketing Department	28
Operational Department	673
Production Department	104
Purchase Department	230
Sales Department	371
Service Department	403
Grand Total	1856

**PIVOT - TABLE SCREENSHOTS**

[Click Here : Sheet Name : Q.A](#)

[Pivot Table and Chart](#)

**Case Study : Hiring Analysis (Q.B)**

[CLICK HERE FOR THE SHEET](#)

**Objective :** To find the average salary offered by this company

**Approach 1 :** Formula is used to get the average of the offered salaries column from the data.

Formula : `=AVERAGE(J4:J6)`

**Result**

Average Salary = 49983.02902

Note : The formula is applied to cell J11. Sheet Name : Input\_Data

## PROJECT TASKS SAMPLE SCREENSHOTS IN SLIDE SORTER VIEW

**Case Study : Hiring Analysis (Q.B)**

[CLICK HERE FOR THE SHEET](#)

**PIVOT - TABLE SCREENSHOT**

Department	Sum	Average
Finance Department	14292866	69028.0004
General Management	10103020	58722.09303
Human Resource Department	4753221	49002.27835
Marketing Department	15795222	48491.30338
Operational Department	13104030	54348.34385
Production Department	18790424	69448.68421
Purchase Department	17904070	52564.77277
Sales Department	36785544	49310.3807
Service Department	10404442	50629.88418
<b>Grand Total</b>	<b>388783109</b>	<b>49983.02902</b>

[Click Here : Sheet Name : Q.B](#)

[Pivot Table and Chart](#)

**Case Study : Hiring Analysis (Q.C)**

[CLICK HERE FOR THE SHEET](#)

**Objective :** To create class intervals for the salaries in the company.

**Approach :** Frequency Formula is used to get count in the each interval. The class interval is formed manually by calculating the minimum and maximum salaries.

Formula : `=FREQUENCY(G:G,M15:M27)`

Data Array represents the salary data and the bins array represents the class intervals for which count is being calculated.

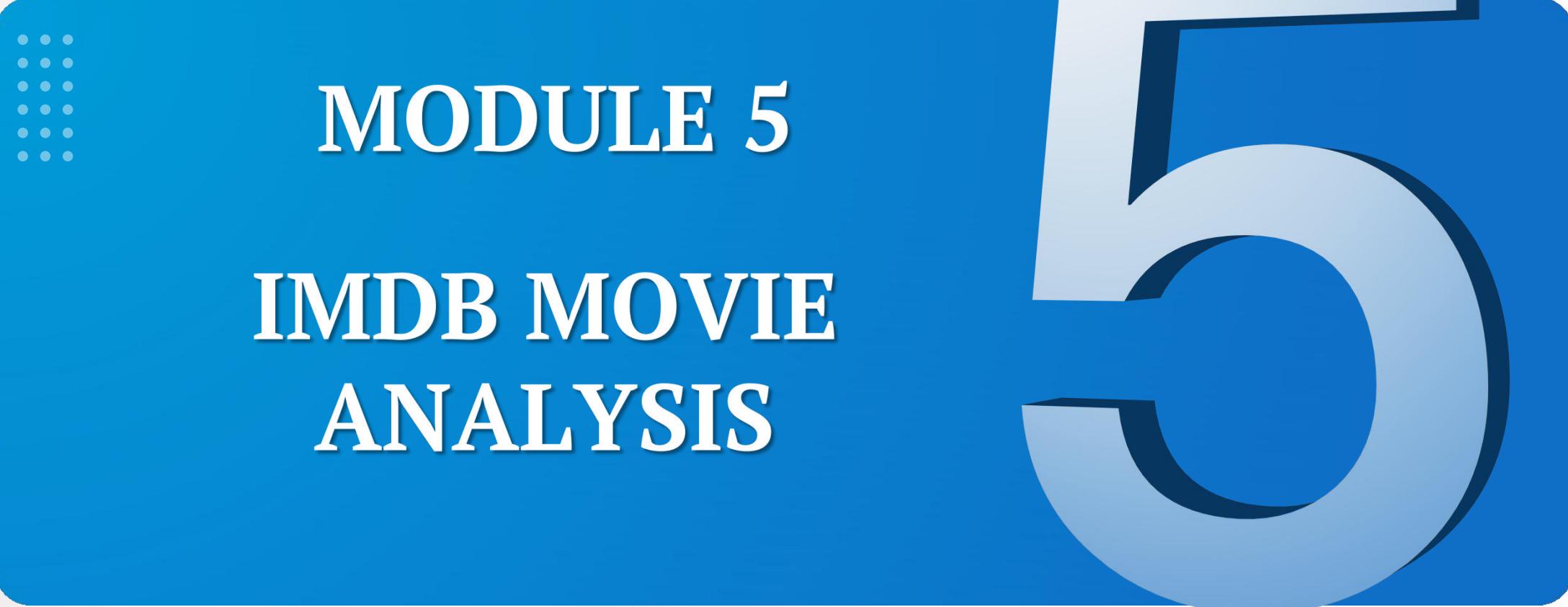
**Result**

Class Intervals	Frequency
0-100000	732
100000-200000	731
200000-300000	731
300000-400000	731
400000-500000	750
500000-600000	699
600000-700000	699
700000-800000	711
800000-900000	659
900000-1000000	1
1000000-1100000	1
1100000-1200000	1
1200000-1300000	1
1300000-1400000	1
1400000-1500000	1
1500000-1600000	1
1600000-1700000	1
1700000-1800000	1
1800000-1900000	1
1900000-2000000	1
2000000-2100000	1
2100000-2200000	1
2200000-2300000	1
2300000-2400000	1
2400000-2500000	1
2500000-2600000	1
2600000-2700000	1
2700000-2800000	1
2800000-2900000	1
2900000-3000000	1
3000000-3100000	1
3100000-3200000	1
3200000-3300000	1
3300000-3400000	1
3400000-3500000	1
3500000-3600000	1
3600000-3700000	1
3700000-3800000	1
3800000-3900000	1
3900000-4000000	1
4000000-4100000	1
4100000-4200000	1
4200000-4300000	1
4300000-4400000	1
4400000-4500000	1
4500000-4600000	1
4600000-4700000	1
4700000-4800000	1
4800000-4900000	1
4900000-5000000	1
5000000-5100000	1
5100000-5200000	1
5200000-5300000	1
5300000-5400000	1
5400000-5500000	1
5500000-5600000	1
5600000-5700000	1
5700000-5800000	1
5800000-5900000	1
5900000-6000000	1
6000000-6100000	1
6100000-6200000	1
6200000-6300000	1
6300000-6400000	1
6400000-6500000	1
6500000-6600000	1
6600000-6700000	1
6700000-6800000	1
6800000-6900000	1
6900000-7000000	1
7000000-7100000	1
7100000-7200000	1
7200000-7300000	1
7300000-7400000	1
7400000-7500000	1
7500000-7600000	1
7600000-7700000	1
7700000-7800000	1
7800000-7900000	1
7900000-8000000	1
8000000-8100000	1
8100000-8200000	1
8200000-8300000	1
8300000-8400000	1
8400000-8500000	1
8500000-8600000	1
8600000-8700000	1
8700000-8800000	1
8800000-8900000	1
8900000-9000000	1
9000000-9100000	1
9100000-9200000	1
9200000-9300000	1
9300000-9400000	1
9400000-9500000	1
9500000-9600000	1
9600000-9700000	1
9700000-9800000	1
9800000-9900000	1
9900000-10000000	1
10000000-10100000	1
10100000-10200000	1
10200000-10300000	1
10300000-10400000	1
10400000-10500000	1
10500000-10600000	1
10600000-10700000	1
10700000-10800000	1
10800000-10900000	1
10900000-11000000	1
11000000-11100000	1
11100000-11200000	1
11200000-11300000	1
11300000-11400000	1
11400000-11500000	1
11500000-11600000	1
11600000-11700000	1
11700000-11800000	1
11800000-11900000	1
11900000-12000000	1
12000000-12100000	1
12100000-12200000	1
12200000-12300000	1
12300000-12400000	1
12400000-12500000	1
12500000-12600000	1
12600000-12700000	1
12700000-12800000	1
12800000-12900000	1
12900000-13000000	1
13000000-13100000	1
13100000-13200000	1
13200000-13300000	1
13300000-13400000	1
13400000-13500000	1
13500000-13600000	1
13600000-13700000	1
13700000-13800000	1
13800000-13900000	1
13900000-14000000	1
14000000-14100000	1
14100000-14200000	1
14200000-14300000	1
14300000-14400000	1
14400000-14500000	1
14500000-14600000	1
14600000-14700000	1
14700000-14800000	1
14800000-14900000	1
14900000-15000000	1
15000000-15100000	1
15100000-15200000	1
15200000-15300000	1
15300000-15400000	1
15400000-15500000	1
15500000-15600000	1
15600000-15700000	1
15700000-15800000	1
15800000-15900000	1
15900000-16000000	1
16000000-16100000	1
16100000-16200000	1
16200000-16300000	1
16300000-16400000	1
16400000-16500000	1
16500000-16600000	1
16600000-16700000	1
16700000-16800000	1
16800000-16900000	1
16900000-17000000	1
17000000-17100000	1
17100000-17200000	1
17200000-17300000	1
17300000-17400000	1
17400000-17500000	1
17500000-17600000	1
17600000-17700000	1
17700000-17800000	1
17800000-17900000	1
17900000-18000000	1
18000000-18100000	1
18100000-18200000	1
18200000-18300000	1
18300000-18400000	1
18400000-18500000	1
18500000-18600000	1
18600000-18700000	1
18700000-18800000	1
18800000-18900000	1
18900000-19000000	1
19000000-19100000	1
19100000-19200000	1
19200000-19300000	1
19300000-19400000	1
19400000-19500000	1
19500000-19600000	1
19600000-19700000	1
19700000-19800000	1
19800000-19900000	1
19900000-20000000	1
20000000-20100000	1
20100000-20200000	1
20200000-20300000	1
20300000-20400000	1
20400000-20500000	1
20500000-20600000	1
20600000-20700000	1
20700000-20800000	1
20800000-20900000	1
20900000-21000000	1
21000000-21100000	1
21100000-21200000	1
21200000-21300000	1
21300000-21400000	1
21400000-21500000	1
21500000-21600000	1
21600000-21700000	1
21700000-21800000	1
21800000-21900000	1
21900000-22000000	1
22000000-22100000	1
22100000-22200000	1
22200000-22300000	1
22300000-22400000	1
22400000-22500000	1
22500000-22600000	1
22600000-22700000	1
22700000-22800000	1
22800000-22900000	1
22900000-23000000	1
23000000-23100000	1
23100000-23200000	1
23200000-23300000	1
23300000-23400000	1
23400000-23500000	1
23500000-23600000	1
23600000-23700000	1
23700000-23800000	1
23800000-23900000	1
23900000-24000000	1
24000000-24100000	1
24100000-24200000	1
24200000-24300000	1
24300000-24400000	1
24400000-24500000	1
24500000-24600000	1
24600000-24700000	1
24700000-24800000	1
24800000-24900000	1
24900000-25000000	1
25000000-25100000	1
25100000-25200000	1
25200000-25300000	1
25300000-25400000	1
25400000-25500000	1
25500000-25600000	1
25600000-25700000	1
25700000-25800000	1
25800000-25900000	1
25900000-26000000	1
26000000-26100000	1
26100000-26200000	1
26200000-26300000	1
26300000-26400000	1
26400000-26500000	1
26500000-26600000	1
26600000-26700000	1
26700000-26800000	1
26800000-26900000	1
26900000-27000000	1
27000000-27100000	1
27100000-27200000	1
27200000-27300000	1
27300000-27400000	1
27400000-27500000	1
27500000-27600000	1
27600000-27700000	1
27700000-27800000	1
27800000-27900000	1
27900000-28000000	1
28000000-28100000	1
28100000-28200000	1
28200000-28300000	1
28300000-28400000	1
28400000-28500000	1
28500000-28600000	1
28600000-28700000	1
28700000-28800000	1
28800000-28900000	1
28900000-29000000	1
29000000-29100000	1
29100000-29200000	1
29200000-29300000	1
29300000-29400000	1
29400000-29500000</	

# Module 4 : Key Takeaways



- The data analytical tasks were made using the concept of “Pivot” Tables in Excel. These are highly interactive as we can move, edit and delete the fields and achieve on the desired results. These are also dynamic and easy to update.
- The Data Visualisation given to the respective tasks, gives us an enhanced understanding of the data. These help us to find remarkable results, patterns in the data, and also can be used for detecting errors in some cases. They also help us to identify relationships among the data.
- The results achieved through the tasks will be highly helpful for the company for its hiring process and also gives an heads up to emerge in the lagging areas.
- On a whole, the entire solutions helps the company for faster decision making and enhances the further activities



# MODULE 5

## IMDB MOVIE ANALYSIS

5

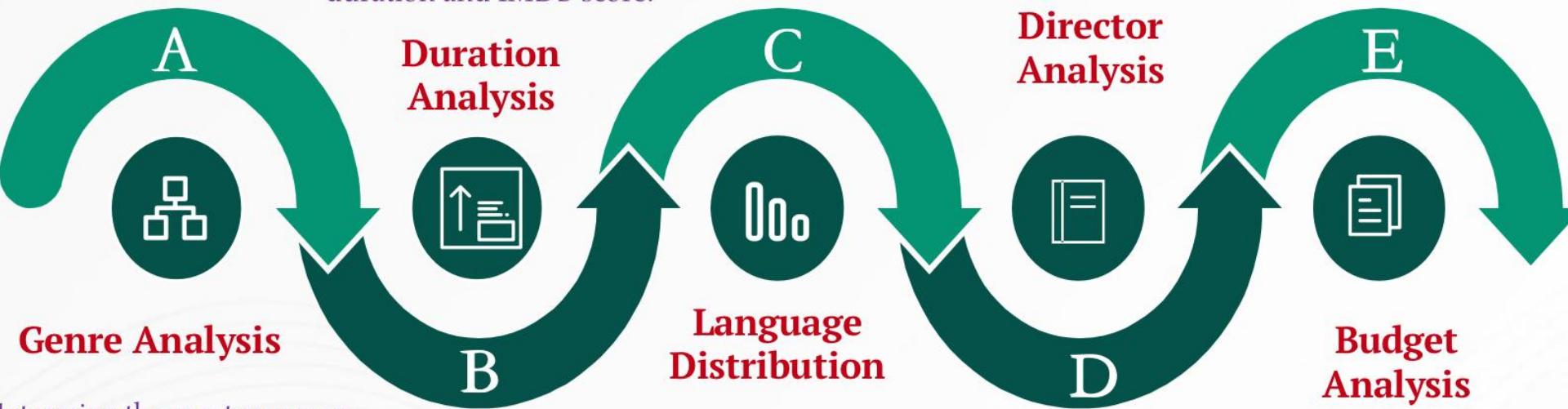
# Module 5 : IMDB Movie Analysis

## Objective / Learning

- Implementing Descriptive Statistics using MS-Excel to investigate the success of a movie.
- Performing data analysis to provide insights that can drive decision-making.
- Analysing results for actionable insights that can help the stakeholders make informed decisions.



# Module 5 : IMDB Movie Analysis - Tasks



To determine the most common genres of movies in the dataset and calculate descriptive statistics based on their IMDB Scores

To analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score.

To identify the top directors based on their average IMDB score and analyze their contribution to the success of movies

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

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

# Module 5 : IMDB Movie Analysis



<b>Task 1 : Movie Genre Analysis</b> <p><b>Objective :</b> To 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.</p> <p><b>Approach :</b></p> <ul style="list-style-type: none"><li>✓ Use text-split function to split the individual genre of a movie</li><li>✓ Using countif, count the number of movies under each genre</li><li>✓ Calculate the descriptive statistics (using filter and formulas) for each genre</li><li>✓ Compare the statistics with ratings and observe the impact</li></ul> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	<b>Task 1 : Movie Genre Analysis</b> <p><b>Descriptive Statistics</b></p> <table border="1"><thead><tr><th></th><th>Average/Mean</th><th>Median</th><th>Variance</th><th>Std. Deviation</th><th>Maximum</th><th>Minimum</th><th>Range</th></tr></thead><tbody><tr><td>Genre</td><td>547.2307692</td><td>306.5</td><td>390777.6246</td><td>625.1220878</td><td>2537</td><td>1</td><td>2536</td></tr></tbody></table> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>		Average/Mean	Median	Variance	Std. Deviation	Maximum	Minimum	Range	Genre	547.2307692	306.5	390777.6246	625.1220878	2537	1	2536	<b>Task 1 : Movie Genre Analysis</b> <p><b>No. of Movies Genre Wise</b></p> <p>The pie chart illustrates the proportion of movies across various genres. The largest categories are Action (24.2%), Drama (18.2%), and Thriller (13.2%). Other significant genres include Romance (10.2%), Mystery (8.2%), and Crime (7.2%). Smaller proportions are seen in Science Fiction, Fantasy, Horror, and War genres.</p> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	<b>Task 1 : Movie Genre Analysis</b> <p><b>IMDB Ratings</b></p> <p>This bar chart displays the average IMDB rating for each genre. Action movies have the highest rating at approximately 7.2, followed by Thriller (7.0), Drama (6.8), and Mystery (6.7). Crime, Romance, and Fantasy genres show the lowest ratings around 6.0.</p> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	<b>Task 2 : Movie Duration Analysis</b> <p><b>Objective :</b> To analyze the distribution of movie durations and identify the relationship between movie duration and IMDB score</p> <p><b>Approach :</b></p> <ul style="list-style-type: none"><li>✓ Calculation of descriptive statistics for the movie duration column</li><li>✓ Creating a scatter plot to visualise the relationship between the movie durations and the IMDB Score of the movies</li><li>✓ Observing the trend of the data spread using trendline</li></ul> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>					
	Average/Mean	Median	Variance	Std. Deviation	Maximum	Minimum	Range																		
Genre	547.2307692	306.5	390777.6246	625.1220878	2537	1	2536																		
11	12	13	14	15																					
<b>Task 2 : Movie Duration Analysis</b> <p><b>MOVIE DURATION ANALYSIS</b></p> <p><b>Metric</b>      <b>Values</b></p> <table border="1"><thead><tr><th>Metric</th><th>Values</th></tr></thead><tbody><tr><td>Average</td><td>106.776482</td></tr><tr><td>Median</td><td>103</td></tr><tr><td>Mode</td><td>90</td></tr><tr><td>Std Dev</td><td>25.92551132</td></tr><tr><td>Variance</td><td>672.1321372</td></tr><tr><td>Maximum</td><td>251</td></tr><tr><td>Minimum</td><td>0</td></tr></tbody></table> <p>R<sup>2</sup> is an statistical measure in a regression model that determines the proportion of variance in the dependent variable that can be explained by the independent variable.</p> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	Metric	Values	Average	106.776482	Median	103	Mode	90	Std Dev	25.92551132	Variance	672.1321372	Maximum	251	Minimum	0	<b>Task 3 : Language Analysis</b> <p><b>Objective :</b> To determine the most common languages used in movies and analyze their impact on the IMDB score using descriptive statistics.</p> <p><b>Approach :</b></p> <ul style="list-style-type: none"><li>✓ Using the Pivot Table, the number of movies which belong to a particular language is calculated</li><li>✓ The descriptive statistics is calculated for IMDB Scores for each language</li><li>✓ Observing the results to understand the relationship between language and movie ratings</li></ul> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	<b>Task 3 : Language Analysis</b> <p><b>PivotTable Fields</b></p> <p>The dialog box shows the configuration of a PivotTable. The "Rows" field is set to "Language", and the "Values" field is set to "IMDB Score". Other fields like "Genre", "Title", "Year", "Rating", "Score", and "Score(Standard Deviation)" are listed under "Available Fields".</p> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	<b>Task 3 : Language Analysis</b> <p><b>Language Analysis</b></p> <p>This bar chart compares the average IMDB scores for movies in different languages. English is the clear leader with a score of approximately 7.2. Other prominent languages include Spanish, French, German, and Italian, all with scores around 6.8 to 7.0. Languages like Chinese, Japanese, and Korean show significantly lower average scores, around 5.5 to 6.0.</p> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	<b>Task 4 : Director Analysis</b> <p><b>Objective :</b> To identify the top directors based on their average IMDB score and analyze their contribution to the success of movies using percentile calculations.</p> <p><b>Approach :</b></p> <ul style="list-style-type: none"><li>✓ Using the Pivot Table, the average of IMDB Scores are calculated</li><li>✓ Using the % Grand Total, the contribution to the total distribution is figured out.</li><li>✓ Compare the scores of the directors to the overall distribution and analyse the influence of directors on movie ratings.</li></ul> <p><a href="#">CLICK HERE TO VIEW THE EXCEL SHEET</a></p>	16	17	18	19	20
Metric	Values																								
Average	106.776482																								
Median	103																								
Mode	90																								
Std Dev	25.92551132																								
Variance	672.1321372																								
Maximum	251																								
Minimum	0																								

**PROJECT TASKS SAMPLE SCREENSHOTS IN  
SLIDE SORTER VIEW**

# Module 5 : IMDB Movie Analysis - Results



## Genre Analysis

It is observed that, genre has an impact in the IMDB Ratings. Although most of the genre has a maximum rating above 8, based on the number of movies released on a particular genre, the maximum and minimum scores are impacted.



## Duration Analysis

Based on the R2 value generated on the trendline, it is observed that, duration has 6% impact upon the ratings. Based on the scatter plot, its visualised that the data points does not fall on the regression line completely and it depicts that the variance is small.



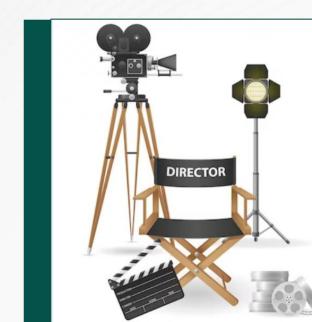
## Language Analysis

It is observed that, language has a very big impact on the IMDB Ratings. The graph plotted for Variance and SD shows that, more the number of movies, there is a good fit for data based on the ratings. Also, variance greater than 1 depicts that the values are very far from mean.



## Budget Analysis

Based on the CORREL Function, it is observed that there is a weak relationship between the gross and budget attributes. The list of movies are categorised as hit or flop based on the profit values. Using Max function, the highest profit margin was found to be Avatar.



## Director Analysis

- The top 10 directors based on their average IMDB Scores are figured out using the pivot table.
- But, this does not give a good influence on the ratings.
- The directors with single movie and having high rating are being the top-most which does not behave as a good way of predictions.
- Hence, one more pivot table was added to find out the top 10 directors based on the average IMDB Scores and the number of movies directed.
- The correlation between the number of movies and the average scores are calculated as **0.125 (~0.13)**
- This states that, there is a weak correlation between the director and the average ratings.
- It can be concluded that, there is **no much influence** on the ratings based on the director.

# Module 5 : Key Takeaways



- Descriptive statistics refers to a branch of statistics that involves summarizing, organizing, and presenting data meaningfully and concisely. It enables researchers or analysts to gain insights and understand patterns, trends, and distributions within the dataset.
- This summary typically includes measures such as central tendency (e.g., mean, median, mode), dispersion (e.g., range, variance, standard deviation), and shape of the distribution (e.g., skewness, kurtosis).
- Descriptive statistics also involves a graphical representation of data through charts, graphs, and tables, which can further aid in visualizing and interpreting the information.
- Data Cleaning involves preprocessing the data to make it suitable for analysis.
- It includes handling missing values, removing duplicates, converting data types if necessary, and possibly feature engineering.
- It involves identifying and removing any missing, duplicate, or irrelevant data. This ultimately increase overall productivity and allow for the highest quality information in decision-making. It also helps in removal of errors when multiple sources of data are at play.



**MODULE 6**

**BANK LOAN - CASE  
STUDY**

# Module 6 : Bank Loan - Case Study

## Objective / Learning

- Implementing Exploratory Data Analysis (EDA) to analyse patterns and trends in the loan data.
- Performing data analysis to provide insights about customers and their previous loan details.
- Analysing results to make informed loan approval decisions.
- Python is used as the dataset is too large to be handled through MS-EXCEL.



# Module 6 : Bank Loan Case Study



## Exploratory Data Analysis (EDA)

- Exploratory Data Analysis (EDA) refers to the method of studying and exploring record sets to apprehend their predominant traits, discover patterns, locate outliers, and identify relationships between variables.
- It is normally carried out as a preliminary step before undertaking extra formal statistical analyses or modeling.
- It refers back to the method of analyzing and analyzing information units to uncover styles, pick out relationships, and gain insights.
- There are various sorts of EDA strategies that can be hired relying on the nature of the records and the desires of the evaluation.



## Project : Tech Stack Used

- **Why Python over Excel ?** As a data scientist, we often work with lots of different types of data from different sources and in large amounts. While Excel can manage data from multiple sources, Python has libraries that allows to easily access and process data from lots of other sources.
- **Python :** Python libraries and packages can level-up how to analyze, visualize, and understand complex data. It is easy to learn and simple to read, you can start mastering more complicated concepts quicker. To incorporate Machine Learning Predictions, it is much more straightforward with Python
- **Jupyter Notebook :** The Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text.



## Project : Tech Stack Used - Python Libraries



- **Pandas** is an open-source library in Python that is made mainly for working with relational or labeled data both easily and intuitively. It provides various data structures and operations for manipulating numerical data and time series. Pandas is fast and it has high performance & productivity for users.
- **NumPy** is a general-purpose array-processing package. It provides a high-performance multidimensional array object and tools for working with these arrays. It is the fundamental package for scientific computing with Python. It is open-source software.
- **Math** module provides various the value of various constants . Having such constants saves the time of writing the value of each constant every time we want to use it and that too with great precision. Constants provided by the math module are – Euler's Number, Pi, Tau, Infinity, Not a Number (NaN).
- **Matplotlib** is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.
- **Seaborn** is an amazing visualization library for statistical graphics plotting in Python. It provides beautiful default styles and color palettes to make statistical plots more attractive. It provides dataset-oriented APIs so that we can switch between different visual representations for the same variables for a better understanding of the dataset.

# Module 6 : Bank Loan - Case Study



PROJECT TASKS SAMPLE SCREENSHOTS DEPICTING THE  
UNIVARIATE AND BI-VARIATE ANALYSIS  
IN SLIDE SORTER VIEW

# Module 6 : Key Takeaways



- Exploratory Data Analysis (EDA) refers to the method of studying and exploring record sets to apprehend their predominant traits, discover patterns, locate outliers, and identify relationships between variables. It refers back to the method of analyzing and analyzing information units to uncover styles, pick out relationships, and gain insights.
- The dataset is observed completely before analysis for the pre-processing steps. The various attributes are learnt and identified.
- Missing Values and outliers are handled appropriately for the required attributes analysis.
- Through the univariate and bivariate analysis, different attribute relationships are identified which is used for analysing the risk associated with a customer.
- Correlation Analysis is made to identify the closely related attributes and the top 10 attributes are listed.
- Combined Dataset EDA is performed and the relationship are identified between the attributes for more detailed conclusions.
- Based on these evaluations, a list of Categorical and Numerical Attributes are identified which can be used for prediction of risk customers.



# MODULE 7

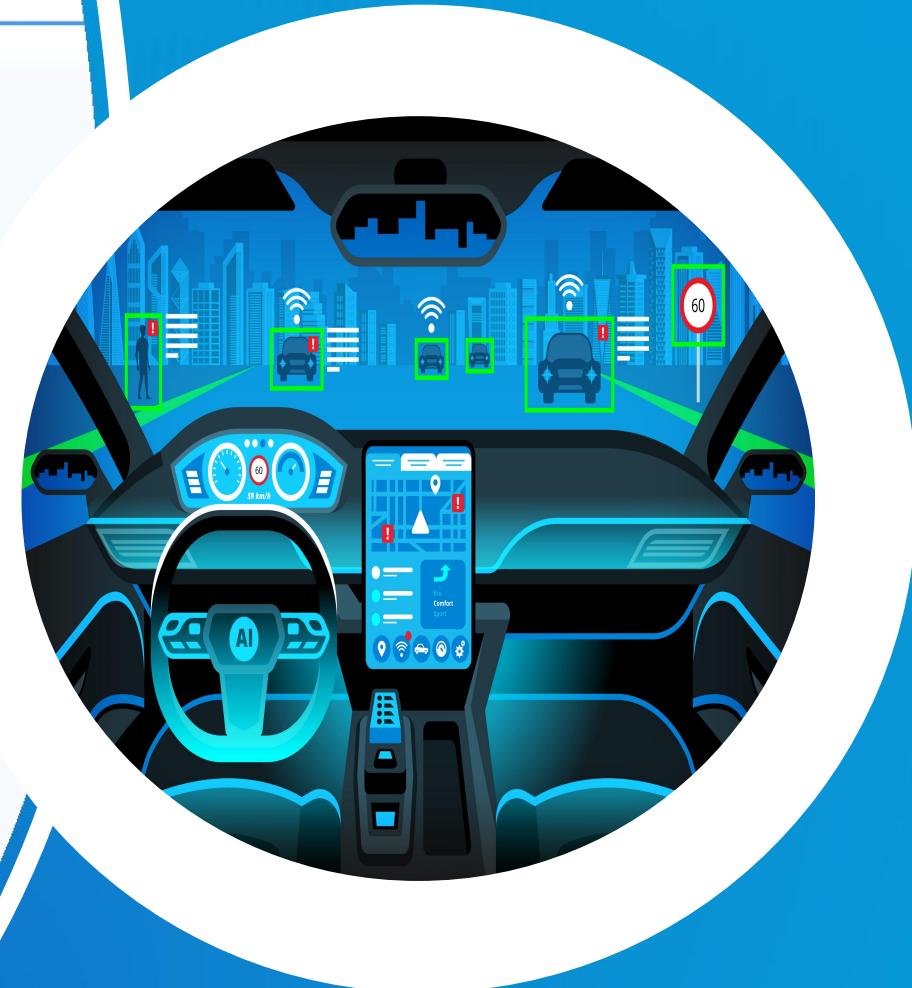
ANALYSING THE IMPACT OF  
CAR FEATURES ON PRICE  
AND PROFITABILITY



# Module 7 : Car Feature Analysis

## Objective / Learning

- To observe how can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand.
- Perform Business Analytics and Data Analytics tasks to visualise charts and dashboards.
- Tableau is used for analytical tasks.



# Module 7 : Anaysing Car Features



## Problem Statement



- The automotive industry has been rapidly evolving over the past few decades, with a growing focus on fuel efficiency, environmental sustainability, and technological innovation. With increasing competition among manufacturers and a changing consumer landscape, it has become more important than ever to understand the factors that drive consumer demand for cars. In recent years, there has been a growing trend towards electric and hybrid vehicles and increased interest in alternative fuel sources such as hydrogen and natural gas. At the same time, traditional gasoline-powered cars remain dominant in the market, with varying fuel types and grades available to consumers.
- For the given dataset, as a Data Analyst, observe how can a car manufacturer optimize pricing and product development decisions to maximize profitability while meeting consumer demand.

## Project : Tech Stack Used

- Tableau is a leading data visualization tool used for data analysis and business intelligence.
- Gartner's Magic Quadrant classified Tableau as a leader for analytics and business intelligence.
- Tableau supports powerful data discovery and exploration that enables users to answer important questions in seconds
- No prior programming knowledge is needed.
- It can connect to several data sources that other BI tools do not support.
- Tableau enables users to create reports by joining and blending different datasets
- Tableau Server supports a centralized location to manage all published data sources within an organization



**PROJECT TASKS SAMPLE SCREENSHOTS DEPICTING THE  
PROBLEM STATEMENT AND THE TECH-STACK (TABLEAU)**

# Module 7 : Anaysing Car Features



## Task 1 : Popularity of Car Models over Categories

- Insight Required :** To understand how popularity of a car model vary across different market categories.
- Approach :**
  - To create chart that shows the number of car models in each market category and their corresponding popularity scores.
  - To create a combo chart that visualizes the relationship between market category and popularity.
- Tableau Link :** [https://public.tableau.com/views/Project\\_7\\_10978475300/Dashboard1?embed=true&publish=yes&showSheetLinks=true](https://public.tableau.com/views/Project_7_10978475300/Dashboard1?embed=true&publish=yes&showSheetLinks=true)
- Sheet Name :** Task 1A, Task 1B

## Task 1 : Popularity of Car Models over Categories



Bar charts and Combo Charts are created for Market Category, Car Model and Popularity Score. Filters are used for Market Category. Dual Axis concept was used for combo chart.

7

## Task 2 : Relationship between Price and Engine Power

- Insight Required :** To understand the relationship between a car's engine power and its price.
- Approach:**
  - To 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.
- Tableau Link :** [https://public.tableau.com/views/Project\\_7\\_10978475300/Dashboard2?embed=true&publish=yes&showSheetLinks=true](https://public.tableau.com/views/Project_7_10978475300/Dashboard2?embed=true&publish=yes&showSheetLinks=true)
- Sheet Name :** Task 2

## Task 2 : Relationship between Price and Engine Power



As observed in the trend line, there is a positive correlation existing between the car price and engine power. As the engine power of vehicle increases, the car price also increases. Hence, they both are directly correlated.

9

## Task 3 : Regression Analysis on Various Attributes

- Insight Required :** To understand which our features are most important in determining a car's price.
- Approach:**
  - To 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.
- Tableau Link :** [https://public.tableau.com/views/Project\\_7\\_10978475300/Dashboard3?embed=true&publish=yes&showSheetLinks=true](https://public.tableau.com/views/Project_7_10978475300/Dashboard3?embed=true&publish=yes&showSheetLinks=true)
- Sheet Name :** Task 3

6

8

10

# BUSINESS ANALYTICS - PROJECT TASKS

## SAMPLE SCREENSHOTS IN SLIDE SORTER VIEW

## Task 3 : Regression Analysis on Various Attributes



As observed in the trend line, there is a positive correlation existing between the car price with Engine HP and Cylinders. Hence, they are directly correlated. As the trendline is downline with Highway and City MPG, they are Negatively correlated with price. There doesn't exist any correlation between price and popularity. Hence, they are not correlated.

Sheet Name : Task 3

## Task 4 : Average Price of Car for Various Manufacturers

- Insight Required :** To understand how the average price of a car vary across different manufacturers.
- Approach:**
  - To create a chart that shows the average price of cars for each manufacturer.
  - To visualize the relationship between car manufacturer and average price.
- Tableau Link :** [https://public.tableau.com/views/Project\\_7\\_10978475300/Dashboard4?embed=true&publish=yes&showSheetLinks=true](https://public.tableau.com/views/Project_7_10978475300/Dashboard4?embed=true&publish=yes&showSheetLinks=true)
- Sheet Name :** Task 4

12

## Task 4 : Average Price of Car for Various Manufacturers



As observed in the bar chart, we can see that Bugatti has higher average price than other manufacturers.

13

## Task 5 : Relationship between Fuel Efficiency and No. of Cylinders

- Insight Required :** To understand relationship between fuel efficiency and the number of cylinders in a car's engine.
- Approach:**
  - To create a chart that shows the average price of cars for each manufacturer.
  - To visualize the relationship between car manufacturer and average price.
- Tableau Link :** [https://public.tableau.com/views/Project\\_7\\_10978475300/Dashboard5?embed=true&publish=yes&showSheetLinks=true](https://public.tableau.com/views/Project_7_10978475300/Dashboard5?embed=true&publish=yes&showSheetLinks=true)
- Sheet Name :** Task 5

14

## Task 5 : Relationship between Fuel Efficiency and No. of Cylinders



Correlation Coefficient formula is added as a "Calculated Field" and is obtained between Cylinders and Highway MPG of cars. It is observed that, the trendline is downward, and hence, there exists a "Negative Correlation" between them.

15

11

# Module 7 : Anaysing Car Features



Dashboard 1 : Price vs Car Make vs Car Style



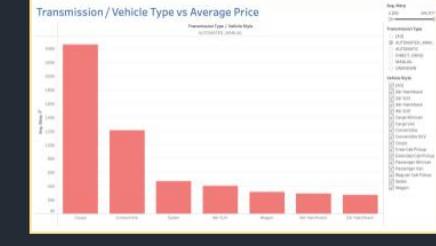
17

Dashboard 2 : Average Price vs Car Make vs Car Style



18

Dashboard 3 : Average Price vs Transmission / Vehicle Type



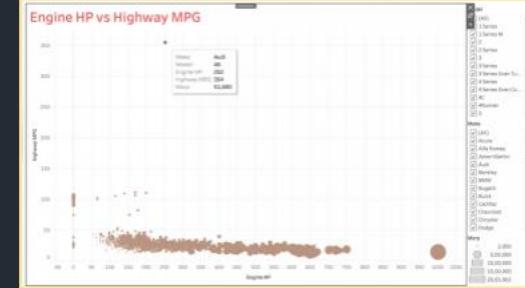
19

Dashboard 4 : Yearwise Highway MPG



20

Dashboard 5 : Engine HP vs Highway MPG on Price



21

**DATA ANALYTICS (DASHBOARDS) - PROJECT TASKS : SAMPLE SCREENSHOTS**

# Module 7 : Key Takeaways



- Analyzing trends in car features and pricing over time: By examining the variables in the dataset, car features and prices have changed over time. This could help manufacturers make informed decisions about product development and pricing.
- Comparing the fuel efficiency of different types of cars: By looking at the MPG variables in the dataset, the fuel efficiency of different types of cars are compared and the most efficient is identified. This could help consumers make informed decisions about which car to purchase.
- Investigating the relationship between a car's features and its popularity: By examining the popularity variable in the dataset, the features which are most popular among consumers are identified. This could help manufacturers make informed decisions about product development and marketing.
- Predicting the price of a car based on its features and market category: By using the various features and market category variables in the dataset, a model is developed to predict the price of a car. This could help manufacturers and consumers understand how different features affect the price of a car and make informed decisions about pricing and purchasing.
- Overall, this analysis could be a valuable resource for exploring various aspects of the automotive industry and could provide insights that could inform decisions related to product development, marketing, and pricing.



# MODULE 8

## ABC CALL VOLUME TREND ANALYSIS



# Module 8 : ABC Call Volume Trend Analysis

## Objective / Learning

- Perform Customer Experience (CX) analytics, specifically focusing on the inbound calling team of a company
- To analyze customer feedback and data, derive insights from it, and share these insights with the rest of the organization.
- The goal is to attract, engage, and delight customers, turning them into loyal advocates for the business.
- Tableau and MS-Excel is used for analytical tasks.



# Module 8 : ABC Call Volume Trend Analysis



## Customer Experience Analytics (CX)



- Customer Experience (CX) Analytics is a process that allows businesses to understand their customers' interactions and experiences with their products or services.
- It involves the collection, analysis, and interpretation of customer data to gain insights into customer behavior, preferences, and expectations.
- This data-driven approach helps businesses improve their customer service, enhance their product offerings, and make informed business decisions.

## Importance of CX Analytics



- **Understanding Customer Behavior:** CX analytics provides insights into how customers interact with a business. This includes their browsing patterns, purchase history, product usage, and feedback. Understanding these behaviors helps businesses tailor their offerings to meet customer needs.
- **Personalization :** With CX analytics, businesses can create personalized experiences for their customers. This can lead to increased customer satisfaction and loyalty.
- **Predictive Analytics:** CX analytics can also be used to predict future customer behavior. This can help businesses anticipate customer needs and take proactive measures to meet them.
- **Improving Customer Service:** By analyzing customer feedback and interactions, businesses can identify areas of improvement in their customer service. This can lead to improved customer satisfaction and retention.
- **Driving Business Decisions:** The insights gained from CX analytics can inform business decisions. This can lead to improved product development, marketing strategies, and overall business performance.

**PROJECT TASKS SAMPLE SCREENSHOTS DEPICTING THE CUSTOMER EXPERIENCE ANALYTICS (CX-ANALYTICS) AND ITS IMPORTANCE**

# Module 8 : ABC Call Volume Trend Analysis



## Data Analytics - Tasks



### Average Call Duration

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



### Call Volume Analysis

To visualize the total number of calls received. This should be represented as a graph or chart showing the number of calls against time.



### Manpower Planning

To propose a plan for manpower allocation during each time bucket (from 9 am to 9 pm) to reduce the abandon rate to 10%.



### Night Shift Manpower Planning

To propose a plan for manpower allocation during night time, like they also make 30 calls at night between 9 pm and 9 am.

## PROJECT TASKS - DATA ANALYTICAL TASKS

# Module 8 : ABC Call Volume Trend Analysis



**Average Call Duration**

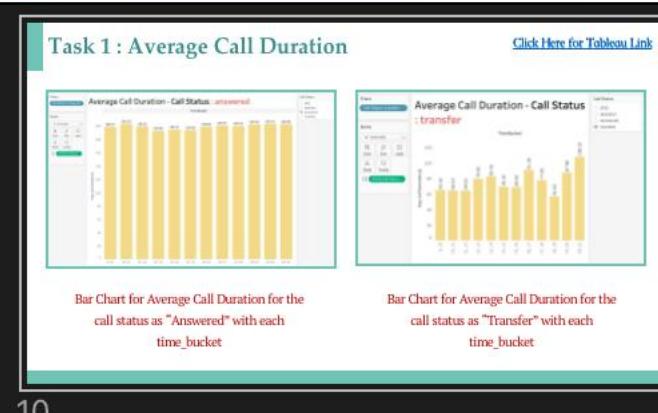
**Objective :** To observe the average call duration for each time bucket.

**Approach :** In tableau, the time bucket is taken as dimension and the call seconds is taken as measure. For the measure, the average value is calculated. Additionally, Call Status values are added as filters for more understanding.

**Data Visualization :** Bar Graph is plotted for the above scenario to visualise the average call duration for each time bucket. Average Values are used as Labels.

**Tableau Sheet Name :** Task 1

01



**Call Volume Analysis**

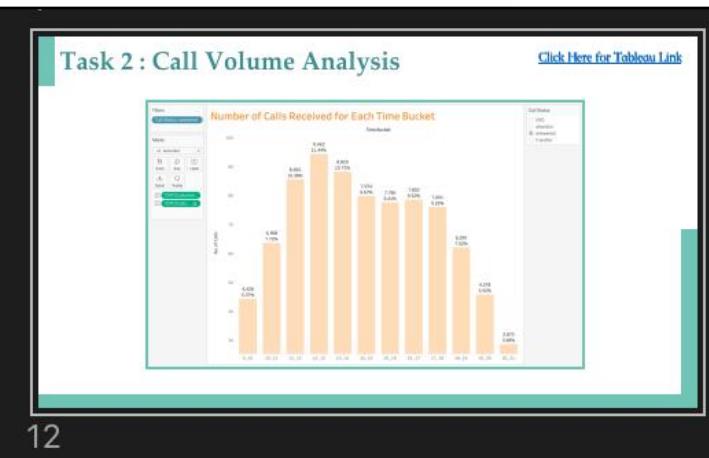
**Objective :** To observe the total number of calls for each time bucket.

**Approach :** In tableau, the time bucket is taken as dimension and the count of customer phone number is taken as measure. For the measure, the percentage of proportion is calculated. Additionally, Call status values are added as filters for more understanding.

**Data Visualization :** Bar Graph is plotted for the above scenario to visualise the average call duration for each time bucket. The count and percentage values are used as labels.

**Tableau Sheet Name :** Task 2

02



## DATA ANALYTICS TASKS (TABLEAU) : SAMPLE SCREENSHOTS

# Module 8 : ABC Call Volume Trend Analysis



**Manpower Planning**

**Objective :** To propose a manpower allocation for the time bucket to reduce the abandon rate to 10%.

**Approach :** In tableau, the distribution of various call status are observed. Based on this, the number of agents required to answer 90 out of 100 calls are calculated to reduce the abandon rate.

**Data Visualisation :** Pie Chart, Bar Graph, Stacked Bar Chart are plotted for the above scenario to visualise the average call duration for each time bucket. The count and percentage values are used as labels.

**Tableau Sheet Name :** Task 5(A), Task 5(B), Task 5(C), Task 5

13

**Task 3 : Manpower Planning**

**Pie Chart for the different call status is plotted along with the percentage of proportion**

**Stacked Bar Chart is plotted for observing the proportion of various call status over each time\_bucket**

14

**Task 3 : Manpower Planning**

**Text Table is constructed for visualising the count of calls, its respective percentage and the number of agents required to reduce abandon rate to 10% over time\_bucket**

**Calculated Field**

**Number of agents required to achieve 10% abandon rate is 57.**

**Using this value, the number of agents required for each time\_bucket is calculated.**

15

**Night Shift : Manpower Planning**

**Objective :** To propose a manpower allocation for the time bucket to reduce the abandon rate to 10% and manage the night calls.

**Approach :** In tableau, the distribution of various call status are observed. Based on this, the number of agents required to right calls with 10% abandon rate is identified. Using Excel, the data is tabulated.

**Data Visualisation :** Using MS-Excel formulae, the time distribution is made and the agents are figured out. Bar chart is plotted for number of agents.

**Tableau Sheet Name :** Task 4, Excel Table

17

**Task 4 : Night Shift : Manpower Planning**

**Text Table is constructed for visualising distribution of calls over each time\_bucket. The number of calls for each call status is observed.**

18

**Task 4 : Night Shift : Manpower Planning**

**Assumption: An agent works for 8 days a week. On average, each agent takes 4 unanswerable calls per month. An agent's total working hours is 8 hours per day. An agent spends 60% of their total active working hours (i.e. 480) in handling calls with customers. The total number of calls in a month is 300.**

**Total Average Night Calls: 1130 For every 100 calls, 30 calls are in night shift**

**Additional Hrs: 76.4113% Average hours per day: 198.6**

**Agents Reqd: 15.28227**

**Based on the assumptions given in the problem, the average of night calls, the additional hours required by the agents and the number of agents for each time bucket are calculated respectively.**

19

**Task 4 : Night Shift : Manpower Planning**

**Table representing the time\_bucket with additional agents required for night-shift planning to reduce the abandon rate to 10%**

**Bar Chart representing the night shift manpower planning with number of agents**

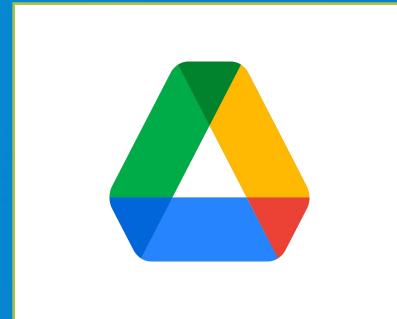
20

## DATA ANALYTICS TASKS (TABLEAU) : SAMPLE SCREENSHOTS

# Module 8 : Key Takeaways



- Customer Experience (CX) Analytics is a process that allows businesses to understand their customers' interactions and experiences with their products or services. It involves the collection, analysis, and interpretation of customer data to gain insights into customer behavior, preferences, and expectations.
- This data-driven approach helps businesses improve their customer service, enhance their product offerings, and make informed business decisions. Understanding these behaviors helps businesses tailor their offerings to meet customer needs.
- The dataset is observed completely before the analysis. Data Pre-processing is done for removal of irrelevant data.
- 9am - 12pm is the slot which receives the highest number of calls and also has a great impact in the company's abandon rate.
- Manpower planning will help to manage the agents efficiently. It will help to reduce the abandon rate and improve the customer experience. This can also lead to improved customer satisfaction and retention.
- On a whole, this analysis would help the company to attract, engage and delight customers, turning them as loyal advocates for the business. The insights gained from CX analytics can inform business decisions. This can lead to improved product development, marketing strategies, and overall business performance.



# MODULE REFERENCES

# MODULE REFERENCES



- **Module 1 : Data Analytics Process**  
[https://drive.google.com/file/d/1s0SlcqOhYhp9qeTQHgtvmzmPF7YRwRP/view?usp=drive\\_link](https://drive.google.com/file/d/1s0SlcqOhYhp9qeTQHgtvmzmPF7YRwRP/view?usp=drive_link)
- **Module 2 : Instagram User Analytics**  
[https://drive.google.com/file/d/1Hydzrdvs\\_ry\\_ovAsxumZCMSgtGC-E3xs/view?usp=drive\\_link](https://drive.google.com/file/d/1Hydzrdvs_ry_ovAsxumZCMSgtGC-E3xs/view?usp=drive_link)
- **Module 3 : Operation and Metric Analytics**  
[https://drive.google.com/file/d/1fIW-thIagXonoOn8DigOI6bSdG8BjLFK/view?usp=drive\\_link](https://drive.google.com/file/d/1fIW-thIagXonoOn8DigOI6bSdG8BjLFK/view?usp=drive_link)
- **Module 4 : Hiring Analysis :** [https://drive.google.com/file/d/1iQzwfznOEnDyULwlwTFBwHjGz4xxMk\\_9/view?usp=drive\\_link](https://drive.google.com/file/d/1iQzwfznOEnDyULwlwTFBwHjGz4xxMk_9/view?usp=drive_link)
- **Module 5 : IMDB Movie Analysis :**  
[https://drive.google.com/drive/folders/157A0I0wwh5gI0DGLtrmCOOivgLuRG2ME?usp=drive\\_link](https://drive.google.com/drive/folders/157A0I0wwh5gI0DGLtrmCOOivgLuRG2ME?usp=drive_link)
- **Module 6 : Bank Loan - Case Study**  
[https://drive.google.com/drive/folders/1kGQraFwReBuFPJ0hwGWVNHKBqg8wt397?usp=drive\\_link](https://drive.google.com/drive/folders/1kGQraFwReBuFPJ0hwGWVNHKBqg8wt397?usp=drive_link)
- **Module 7 : Analysing Impact of Car Features on Price and Profitability**  
[https://drive.google.com/file/d/1xpueoJqJbZmaQ8fjiWlEn1fK4P4uqSV6/view?usp=drive\\_link](https://drive.google.com/file/d/1xpueoJqJbZmaQ8fjiWlEn1fK4P4uqSV6/view?usp=drive_link)
- **Module 8 : ABC Call Volume Trend Analysis**  
[https://drive.google.com/drive/folders/1CjG7hxEzGGvKZTJCo6PBllccmUiM5H7YP?usp=drive\\_link](https://drive.google.com/drive/folders/1CjG7hxEzGGvKZTJCo6PBllccmUiM5H7YP?usp=drive_link)



# Project Portfolio : Data Analytics Course

---

Submitted By

Swathy Mugundan

Contact : +974 - 7107 2857

Email : [swat.mugund97@gmail.com](mailto:swat.mugund97@gmail.com)

