

# DATA ANALYST

## Data Analyst Project Life Cycle

### Phase 1 - Data Collection

- After carefully evaluating the business case in a particular domain, data will be collected surrounding it.

### Phase 2 - Data Preparation

- Using SQL, a database will be created to store the data collected in the previous step.

### Phase 3 - Insights Generation And Dashboard Building

- Establish a connection between the database and Tableau/Python/R tools to extract the required data. Generate user-friendly reports according to the business needs and develop the dashboard using Tableau/Power BI.

## Projects

### Analysis Of Patient Data (Domain: Healthcare)

- This project requires learners to analyze the patient data of those suffering from different diseases across various summaries. The facility, chain organizations, and dialysis stations analysis is required to be carried out where the patients are undergoing dialysis. The project also focuses on the payment mode aspect wherein if any discounts or reduction in payments have happened then those are analyzed.

## **Loan Of Customers (Domain: Banking And Finance)**

- In this project, learners analyze the loan given by a financial institution to different customers of varied grades and sub-grade levels. The analysis needs to consider the loan disbursement reasons, funded amount, and revolving balance values for every customer in different states and geolocations. The project requires the customers payment modes and the last payment values.

## **Employee Retention (Domain: HR Analytics)**

- This HR-related project considers the attrition rate of employees working at an organization at different levels. The attrition rate analysis is done with respect to different factors such as monthly income, last promotion year, job role, and work-life balance of every employee of different departments

## **Industrial Combustion Energy Use (Domain: Energy)**

- The project requires learners to analyze the usage of different fuels in different facilities in different applications by finding the MMBTu and GWHt values. The fuels used for different geo-locations and for different primary titles are also taken into consideration while doing analysis.

## **Flights Delay Analysis (Domain: Aviation)**

- The primary aim of the project is to determine the different reasons behind the delay of flights of various airlines. The analysis needs to consider the number of flights in operation, the number of flights cancelled, and the statistical summary of week-wise, state-wise, and city-wise flight distributions.

## **Olist Store Analysis (Domain: ECommerce)**

- The market for a certain product is analyzed by considering a particular retail outlet which sells these products. The project involves statistical analysis on the payment distribution from different customers with the different modes of transactions across different product categories. The feedback from customers with respect to shipping days and other factors also needs to be considered while carrying out the analysis.

## Topics To Be Covered

- Excel: Basics to Advanced
- MySQL
- Python
- AWS Kinesis, AWS RDS, AWS GLUE , AWS DATA PIPELINE , AWS QUICKSIGHT , AWS REDSHIFT ,AWS ATHENA , AWS EMR
- Power BI

## Excel

### Introduction

- MS office Versions(similarities and differences)
- Interface(latest available version)
- Row and Columns
- Keyboard shortcuts for easy navigation
- Data Entry(Fill series)
- Find and Select
- Clear Options
- Ctrl+Enter
- Formatting options(Font,Alignment,Clipboard(copy, paste special))

# Referencing, Named ranges,Uses,Arithmetic Functions

- Mathematical calculations with Cell referencing(Absolute,Relative,Mixed)
- Functions with Name Range
- Arithmetic functions (SUM,SUMIF,SUMIFS,COUNT,COUNTA,COUNTIFS, AVERAGE,AVERAGEIFS,MAX,MAXIFS,MIN,MINIFS)

## Logical Functions

- Logical functions:IF,AND,OR,NESTED IFS,NOT,IFERROR
- Usage of Mathematical and Logical functions nested together

## Referring data from different tables: Various types of Lookup, Nested IF

- LOOKUP
- VLOOKUP
- NESTED VLOOKUP
- HLOOKUP
- INDEX
- INDEX WITH MATCH FUNCTION
- INDIRECT
- OFFSET

## MySQL

### Introduction to Mysql

- Introduction to Databases

- Introduction to RDBMS
- Explain RDBMS through normalization
- Different types of RDBMS
- Software Installation(MySQL Workbench)

## **SQL Commands and Data Types**

- Types of SQL Commands (DDL,DML,DQL,DCL,TCL) and their applications
- Data Types in SQL (Numeric, Char, Datetime)

## **POWER BI**

### **Power BI Introduction and Installation**

- Understanding Power BI Background
- Installation of Power BI and check list for perfect installation
- Formatting and Setting prerequisites
- Understanding the difference between Power BI desktop & Power Query

## **Fundamentals Of Python**

### **Anaconda Installation,Introduction to python,Data types,Opearators**

- Variables,data types(integer,Boolean,Float,List,tuple,string),Opearators in python

### **Data types Contd,Slicing the data,Inbuilt functions in python**

- Dictionaries,Sequence methods,Concatenate,Repetition,len,min,max functions,Index position,Addition and deletion of elements,Reverse,Sorting

## **Sets,Set Theory,Regular Expressions,Decision making statements**

- Sets,re module(findall,search,split,match),if,elifGetting input from user,Identity Operators

## **Loops,Functions,Lambda functions,Modules**

- For,While loops,Functions,Lambda functions,Math module,Calender module,Date & time module

## **Pandas,Numpy,Matplotlib,Seaborn**

- Data frame creation using different methods,Using Pandas anlysis on Universities,Salary data sets,Visualization using Matplotlib and Seaborn,Numpy introduction