

## **Module 01: Introduction to Data Science**

- Definition and importance
- Data science workflow
- Applications in various industries

## **Module 02: Data Analysis with Excel**

- Advanced Excel functions and formulas
- Data cleaning and preparation in MS Excel
- PivotTables and Pivot Charts

## **Module 03: Data Visualization with Power BI**

- Overview of Power BI interface
- Connecting to data sources
- Creating and customizing visualizations
- Building interactive dashboards
- Sharing and publishing reports

## **Module 04: Power BI with Tools**

- Power Query advanced applications
- Data modeling with Power Pivot
- DAX for data analysis

## **Module 05: Advanced Data Visualization with Tableau**

- Getting started with Tableau
- Data connections and preparation
- Creating advanced visualizations
- Dashboard design best practices
- Storytelling with data

## **Module 06: Fundamentals of Python Programming**

- Python programming basics (Clarity of What, why and how)
- Input, output concepts
- Variables, Loops Concepts
- Functions, Packages, Modules Concepts

## **Module 07: Programming for Data Science with Python**

- Key libraries (NumPy, pandas, matplotlib, seaborn, markdown)
- Data manipulation and analysis
- End to End Data analysis project

## **Module 08a: Foundation of SQL and Databases**

- What is SQL and why is it used?
- Flavors of SQL
- Tables rows and columns
- Concept of ERD (Entity Relationship Diagram)

## **Module 08b: Statements, Function and Analysis using SQL**

- Syntax Familiarity of Select Statements and Techniques of Ordering and filtering Results
- Using WHERE clause
- Sum, Count, min, max, Data types, Functions and Calculations
- SQL for Data Analytics (Project based Learning)

## **Module 09: Data Analysis and Visualization with R**

- Introduction to R programming
- Data manipulation with dplyr
- Data visualization with ggplot2
- Statistical analysis and modeling

## **Module 10: Introduction to Machine Learning**

- Definition, scope, and key concepts
- Real-world examples of ML applications
- Comparison with traditional programming
- Supervised, unsupervised, and reinforcement learning

## **Module 11: Executable Machine Learning / ML Model Development**

- Importance of reproducibility in ML projects
- Introduction to tools like Jupyter Notebook, Google Colab, or IDEs (e.g., PyCharm)
- Importing datasets using libraries like Pandas
- Cleaning and preprocessing data: Handling missing values, scaling, encoding categorical variables
- Feature selection and engineering techniques
- Train-test split and cross-validation setup

## **Module 12a: Machine Learning and CI/CD Pipeline for deployment**

- Continuous Integration: Automating the testing and integration of code
- Continuous Deployment: Automating the deployment process
- Overview of Git for version control:
  - Repositories, commits, branches, and pull requests
  - Setting up a GitHub repository for ML projects
- Collaborating on code and managing updates in ML workflow

## **Module 12b: Deployment of ML Models (Updated)**

- On-Premises Deployment: Running the model locally using Flask/Streamlit.
- Serverless platform for ML based application Deployment