Data Science & Data Analytics – Full Course Curriculum

Module 1: Introduction to Data Science & Analytics

- What is Data Science?

- Difference between Data Analyst, Data Scientist, Data Engineer

- Data Life Cycle & CRISP-DM

- Types of Data & Data Sources

- Real-World Use Cases

Module 2: Excel for Data Analysis

- Excel Basics to Advanced

- Lookup Functions (VLOOKUP, HLOOKUP, XLOOKUP)

- Pivot Tables, Pivot Charts

- Basic Statistical Functions

- Data Cleaning in Excel

- Excel Dashboards

Module 3: Python for Data Analysis

- Python Basics (Variables, Loops, Functions)

- Data Structures (List, Tuple, Dict)

- Working with Files

- NumPy for numerical operations

Module 4: Statistics & Probability

- Descriptive Statistics

- Probability Basics

- Distributions (Normal, Binomial, Poisson)

- Hypothesis Testing

- Correlation & Covariance

- Confidence Interval, p-value, t-tests, chi-square

Module 5: SQL for Data Analytics

- SQL Basics (DDL, DML)
- SELECT, WHERE, GROUP BY, HAVING
- Joins (INNER, LEFT, RIGHT, FULL)
- Subqueries
- Window Functions (RANK, OVER, PARTITION)
- SQL for Business Use Cases
- Creating Dashboards from SQL Queries

Module 6: Data Wrangling & Cleaning (ETL Basics)

- Understanding ETL Process
- Pandas for Data Manipulation
- Handling Missing Values
- Handling Outliers
- Data Formatting & Type Conversion
- Joining & Merging Datasets
- Working with JSON, CSV, Excel, Databases

Module 7: Data Visualization

- Matplotlib, Seaborn, Plotly
- Power BI Basics
- Data Import & Transform
- Creating Dashboards & Reports
- DAX for calculations
- Power BI Automation & Publishing

Module 8: Exploratory Data Analysis (EDA)

- Univariate, Bivariate, Multivariate Analysis
- Feature Distributions
- Correlation Heatmaps

- Identifying data patterns and insights

Module 9: Machine Learning Basics

- ML workflow

- Train/Test Split

- Bias-Variance concept

- Overfitting vs Underfitting

Module 10: Supervised Learning

- Linear Regression

- Logistic Regression

- Decision Trees

- Random Forest

- SVM

- Gradient Boosting (XGBoost, LightGBM)

Module 11: Unsupervised Learning

- K-Means Clustering

- Hierarchical Clustering

- PCA (Dimensionality Reduction)

- Association Rules (Apriori)

Module 12: Feature Engineering

- Scaling & Normalization

- Encoding Techniques

- Feature Selection Methods

- Handling Class Imbalance (SMOTE)

Module 13: Model Evaluation & Tuning

- Evaluation Metrics

- Cross-Validation

- Grid Search / Random Search

- Model Interpretability (SHAP/LIME)

Module 14: Time Series Analysis

- Time Series Components

- ARIMA Models

- Seasonal Decomposition

- Forecasting Methods

Module 15: Big Data & Cloud

- Hadoop & Spark Basics

- PySpark

- AWS, Azure, GCP Overview

Module 16: MLOps & Deployment

- Model Deployment using Flask/FastAPI

- Docker Basics

- CI/CD for ML Models

- Deploying on Cloud

Module 17: Capstone Project

- End-to-end project execution

- Documentation & Presentation

- Deployment (Optional)