Data Analytics Course Content

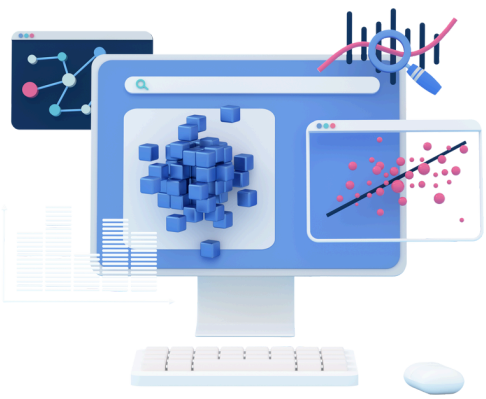
Module 1: Introduction to Data Analytics

- What is Data Analytics?
- Roles and Responsibilities of a Data Analyst
- Data Analytics vs. Data Science vs. Data Engineering
- Data Analytics Process Overview: Collection, Cleaning, Analysis, and Visualization
- Key Tools and Technologies in Data Analytics

Module 2: Data Collection and Preparation

- **Data Collection Methods:**
  - Primary vs. Secondary Data
  - Data Sources: Databases, APIs, Web Scraping

- **Data Cleaning and Preprocessing:**
  - Identifying Missing Values, Outliers, and Duplicates
  - Data Transformation Techniques (Normalization, Standardization)
  - Handling Categorical Data (Encoding, Binning)
  - Data Imputation and Data Validation



Module 3: Data Analysis with Excel

- Introduction to Excel for Data Analysis
- Excel Overview: Interface, data entry, and basic formatting
- Excel Functions: SUM, AVERAGE, COUNT, VLOOKUP, INDEX, MATCH
- Data Visualization in Excel (Charts and Graphs)
- Pivot Tables and Pivot Charts
- Advanced Excel Features: Power Query, Power Pivot, Conditional Formatting
- Data Analysis ToolPak Add-ins: Regression, Hypothesis Testing
- What-If Analysis: Goal Seek and Scenario Manager for forecasting
- Data Validation: Setting up data entry rules, drop-down lists, and preventing incorrect data input.

Module 4: Statistical Analysis for Data Analytics

- **Descriptive Statistics:**
  - Measures of Central Tendency: Mean, Median, Mode
  - Measures of Dispersion: Variance, Standard Deviation, Range

- **Inferential Statistics:**
  - Probability, Hypothesis Testing, Confidence Intervals
  - T-tests, Chi-square tests, ANOVA

- **Correlation and Regression Analysis:**
  - Pearson's Correlation Coefficient
  - Simple Linear Regression, Multiple Linear Regression

Module 5: Data Visualization

- **Data Visualization Principles:**
  - Best Practices for Visualizing Data
  - Types of Visualizations: Bar, Line, Scatter Plots, Histograms, Pie Charts

- **Advanced Data Visualization Tools:**
  - Tableau:
    - Data Connection and Preparation in Tableau
    - Building Interactive Dashboards
    - Filters, Parameters, and Calculated Fields
    - Data Blending and Joins in Tableau
  - Power BI:
    - Data Modeling and Transformation in Power BI
    - Creating Reports and Dashboards in Power BI
    - DAX Functions and Measures

Module 6: SQL for Data Analysis

- **SQL Basics:**
  - Introduction to Databases and SQL
  - SELECT, WHERE, JOIN, GROUP BY, ORDER BY
  - Aggregation Functions (COUNT, SUM, AVG, MIN, MAX)

- **Advanced SQL:**
  - Subqueries and Nested Queries
  - Window Functions (RANK, ROW_NUMBER, SUM OVER)
  - Complex Joins (LEFT JOIN, RIGHT JOIN, INNER JOIN)
  - SQL Optimization Techniques

- **Data Extraction and Transformation (ETL):**
  - Extracting Data from Databases
  - Data Transformation with SQL Queries
  - Writing Efficient SQL Queries for Large Datasets

- **Data Security in SQL:**
  - Encryption of Sensitive Data
  - User Authentication and Authorization

Module 7: Python for Data Analytics

- **Python Basics for Data Analysts:**
  - Introduction to Python and IDEs (Jupyter, VS Code)
  - Variables, Data Types, Functions, Loops, and Conditionals
  - Working with Data Structures (Lists, Tuples, Dictionaries)

- **Data Manipulation with Pandas:**
  - DataFrames, Series, and Data Indexing
  - Data Cleaning and Preprocessing with Pandas
  - Merging, Joining, and Concatenating DataFrames
  - GroupBy and Aggregations

- **Data Visualization with Matplotlib and Seaborn:**
  - Creating Static Plots (Line, Bar, Histogram)
  - Advanced Plotting (Boxplots, Heatmaps, Pairplots)

- **NumPy for Numerical Data Analysis:**
  - Array Manipulation and Broadcasting
  - Mathematical Operations and Vectorization

Module 8: Machine Learning Basics for Data Analysts

- **Introduction to Machine Learning:**
  - What is Machine Learning? Supervised vs. Unsupervised Learning
  - Overview of Algorithms: Linear Regression, Decision Trees, k-NN, k-Means

- **Model Evaluation:**
  - Training, Testing, and Validation Sets
  - Performance Metrics: Accuracy, Precision, Recall, F1-Score, ROC Curve

- **Basic Machine Learning in Python:**
  - Using Scikit-learn for Model Building
  - Feature Engineering and Selection
  - Training and Evaluating Models

Module 9: Advanced Topics in Data Analytics

- **Big Data Concepts:**
  - Introduction to Big Data and Hadoop Ecosystem
  - Working with Large Datasets (HDFS, MapReduce)

- **Data Analytics in Cloud:**
  - Cloud Platforms for Data Analytics (AWS, Google Cloud, Azure)
  - Using BigQuery, AWS Redshift, or Azure Data Lake

- **Data Automation:**
  - Automating Data Workflows with Python
  - Using Apache Airflow for Scheduling Jobs

Module 10: Data Analyst Tools & Collaboration

- **Collaboration with Stakeholders:**
  - Communicating Insights and Reporting
  - Presenting Findings to Non-Technical Audiences

- **Version Control with Git:**
  - Introduction to Git for Data Analysts
  - Using GitHub for Version Control and Collaboration

- **Data Analytics Tools and Platforms:**
  - Overview of Industry Tools (Alteryx, SAS, R)
  - Using Jupyter Notebooks for Data Analysis

Module 11: Project Work

- **Mini Projects:**
  - Sales Data Analysis and Reporting
  - Customer Segmentation Analysis using Clustering
  - Website Traffic Analysis

- **Capstone Project:**
  - Build a Real-World Data Analytics Project (e.g., Financial Forecasting, Marketing Campaign Analysis)
  - Data Collection, Cleaning, Analysis, and Visualization
  - Presenting Insights and Recommendations