Data Science Course Syllabus

1. Introduction to Data Science

- What is Data Science?
- Data Science lifecycle
- Types of data
- Roles of a Data Scientist
- Tools and technologies
- Introduction to Python

2. Python for Data Science

- Python fundamentals
- Variables, data types, operators
- Conditional statements and loops
- Functions
- Lists, tuples, dictionaries, sets
- File handling
- Exception handling
- Object-oriented programming basics

3. NumPy

- Arrays and dimensions
- Array indexing and slicing
- Mathematical operations
- Broadcasting
- Statistical functions
- Working with multidimensional data

4. Pandas

- Series and DataFrames
- Importing CSV/Excel/JSON data
- Data selection and filtering
- Sorting and grouping
- Handling missing values
- Merging and joining datasets
- Data transformation

5. Data Cleaning & Preprocessing

- Identifying data-quality problems
- Missing-value treatment
- Duplicate records
- Outlier detection
- Encoding categorical variables
- Feature scaling
- Feature engineering

6. Data Visualization(MATPLOTLIB)

- Matplotlib
- Charts and plots
- Histograms, box plots, scatter plots
- Correlation heatmaps
- Interactive visualization basics
- Data storytelling

7. Statistics for Data Science

- Descriptive statistics
- Mean, median, mode
- Variance and standard deviation
- Probability fundamentals
- Probability distributions
- Sampling
- Central Limit Theorem
- Confidence intervals
- Hypothesis testing
- Correlation and covariance

8. SQL for Data Science(MYSQL):-

- Relational databases
- SELECT, WHERE, ORDER BY
- GROUP BY and HAVING
- Aggregate functions
- JOINs
- Subqueries
- CTEs
- Window functions
- Data extraction for analysis

9. Machine Learning Fundamentals (SCIKIT-LEARN)

- What is Machine Learning?
- Supervised vs. unsupervised learning
- Training, validation and test sets
- Features and target variables
- Overfitting and underfitting
- Bias-variance tradeoff
- Cross-validation

10. Supervised Learning

- Linear Regression
- Logistic Regression
- Decision Trees
- Random Forest
- K-Nearest Neighbors
- Support Vector Machines
- Gradient Boosting
- Model evaluation

11. Unsupervised Learning

- Clustering
- K-Means
- Hierarchical clustering
- DBSCAN
- Principal Component Analysis (PCA)
- Dimensionality reduction

12. ADVANCE EXCEL

- Excel basics and interface
- Data entry, formatting & sorting
- Formulas and functions
- IF, SUMIF, COUNTIF, VLOOKUP/XLOOKUP
- Conditional formatting
- Data validation
- Pivot Tables & Pivot Charts
- Charts and dashboards
- Data cleaning
- Power Query basics
- Basic automation/macros

13. Power BI

- Power BI interface
- Importing data from Excel, CSV & databases
- Data cleaning with Power Query
- Data modeling & relationships
- DAX formulas
- Calculated columns & measures
- Charts, tables, filters & slicers
- Interactive dashboards
- Drill-down & drill-through
- Publishing and sharing reports

14. Projects

A good course should include **3–4 practical projects**, such as:

- Sales/data analytics dashboard
- House-price prediction
- Customer segmentation
- Sentiment analysis