Data Science Syllabus

Module 1: Beginners Python

- Variables, Keywords, comments
- Numbers, Strings
- Lists, Dictionaries, Sets, Tuples
- Python Operators
- Conditional Statements
- loop statements
- control statements
- Functions, Lambda Functions
- Modules (pip install)

Module 2: Advance Python

- Read, Write files
- Exception handling
- Classes, Objects
- OOPs concepts
- Regular Expressions
- Python Coding Questions

Module 3: Relational Database (MySQL) and SQL

- What is a Database? Relational vs. Non-Relational Databases
- Introduction to MySQL: History, Features, and Use Cases
- Installing MySQL and MySQL Workbench
- Overview of MySQL Architecture: Client-Server Model, Storage Engines

- Basic Terminology: Tables, Rows, Columns, Keys
- SQL Overview: DDL, DML, DCL, TCL
- Creating Databases and Tables (CREATE DATABASE, CREATE TABLE)
- Data Types: Numeric (INT, FLOAT), String (VARCHAR, TEXT), Date/Time
- Modifying Structures (ALTER TABLE, DROP TABLE)
- Constraints: Primary Key, Foreign Key, NOT NULL, UNIQUE
- Inserting Data (INSERT INTO)
- Querying Data (SELECT, WHERE, ORDER BY, LIMIT)
- Updating Data (UPDATE, SET)
- Deleting Data (DELETE FROM)
- Basic Joins: INNER JOIN, LEFT JOIN, RIGHT JOIN
- Advanced Queries: GROUP BY, HAVING,
- Aggregate Functions (COUNT, SUM, AVG)
- Subqueries and Nested Queries
- Indexes

Module 4: NumPy, Pandas, Data Visualization

- Overview of Python in Data Science
- Setting Up the Environment (Anaconda, Jupyter Notebook)
- Python Recap: Lists, Dictionaries, and Basic Operations
- Why NumPy, Pandas, and Visualization Matter
- Introduction to NumPy: Arrays vs. Lists
- Creating Arrays: np.array(), np.zeros(), np.arange(), np.linspace()
- Array Operations: Arithmetic, Broadcasting, Indexing, Slicing

- Multi-Dimensional Arrays: Shape, Reshape, Stacking
- Mathematical Functions: np.sum(), np.mean(), np.std(), Linear Algebra Basics
- Random Number Generation: np.random.rand(), np.random.randn()
- Introduction to Pandas: Series and DataFrames
- Creating Data Structures: pd.Series(), pd.DataFrame(), Reading Files (read_csv, read_excel)
- Data Inspection: head(), info(), describe()
- Data Cleaning: Handling Missing Values (fillna(), dropna()), Duplicates
- Indexing and Selection: .loc[], .iloc[], Filtering
- Data Operations: Grouping (groupby()), Merging (merge(), concat()), Pivot Tables
- Time Series Basics: DateTime Indexing, Resampling
- Introduction to Data Visualization: Purpose and Principles
- Matplotlib Basics: Line Plots, Bar Charts, Scatter Plots (plt.plot(), plt.bar())
- Customizing Plots: Titles, Labels, Legends, Colors
- Seaborn for Enhanced Visuals: Statistical Plots (sns.histplot(), sns.boxplot(), sns.heatmap())
- Advanced Visualization: Subplots, Pair Plots, Facet Grids
- Choosing the Right Plot: When to Use Bar, Line, Scatter, etc.
- Optional: Interactive Plots with Plotly

Module 5: Math for Data Science

- Descriptive vs. Inferential Statistics: Definitions and examples
- Data Types: Continuous vs. Discrete, Nominal vs. Ordinal
- Measures of Central Tendency: Mean, Median, Mode

- Measures of Dispersion: Variance, Standard Deviation
- Basic Plots: Histograms, Bar Charts, Pie Charts
- Vectors: Definition, Operations (Addition, Scalar Multiplication)
- Matrices: Definition, Operations (Addition, Multiplication)
- Dot Product: Definition, Geometric Interpretation
- Eigenvalues and Eigenvectors: Basics and significance
- Basics of Probability: Events, Rules, Independence
- Conditional Probability: Definition and Examples
- Bayes' Theorem: Applications in inference
- Probability Density: Continuous Distributions
- Entropy: Measuring uncertainty (used in LLMs)
- Maximum Likelihood Estimation (MLE): Fitting models to data
- Normal Distribution: Properties, Z-scores
- Correlation and Covariance: Measuring relationships
- Central Limit Theorem: Why averages are normal
- Advanced Plots: Scatter Plots for Correlation, Box Plots for Outliers
- Hypothesis Testing: Null vs. Alternative Hypotheses
- P-Value and Confidence Intervals: Interpreting significance
- Type I and Type II Errors: False Positives vs. False Negatives
- Z-Test: Testing means with known variance

Module 6: Excel

- What is Excel? Why is it important?
- Excel interface overview
- Workbooks vs Worksheets

- Basic navigation and shortcuts
- Saving, opening, and formatting files
- Data entry and editing
- Basic formulas: SUM, AVERAGE, MIN, MAX, COUNT
- Cell referencing: Relative, Absolute, Mixed
- Autofill, Flash Fill
- Formatting cells (Number, Text, Date)
- Conditional Formatting
- Find & Replace
- Data Validation (Dropdowns, Restrictions)
- Sorting & Filtering
- Freezing Panes, Grouping & Ungrouping
- Logical functions: IF, AND, OR, IFERROR, IFS
- Text functions: LEFT, RIGHT, MID, LEN, TRIM, CONCAT, TEXTJOIN
- Lookup & Reference: VLOOKUP, HLOOKUP, INDEX, MATCH,
- Date & Time functions: TODAY, NOW, DATEDIF, NETWORKDAYS, EDATE
- Sorting & Advanced Filtering
- Subtotals & Grouping
- Tables (Excel Tables)
- Named Ranges
- Creating charts: Bar, Line, Pie, Column, Area, Scatter
- Dynamic charts
- Chart formatting & customization
- Combo charts

- Creating Pivot Tables
- Drag and drop fields
- Grouping data in Pivot Tables
- Calculated fields and items
- Slicers and Timeline
- Creating Pivot Charts
- Removing duplicates
- Dealing with blanks
- Text to Columns
- Creating Dashboards in Excel

Module 7: Power BI

- What is Power BI?
- Importance and applications of Power BI
- Power BI Components (Power BI Desktop, Service, Mobile)
- Power BI vs Excel vs Tableau
- Power BI architecture
- Installing and setting up Power BI Desktop
- Interface walkthrough
- Connecting to data sources (Excel, CSV, SQL, Web, etc.)
- Data Loading and Transformation (Power Query Editor)
- Introduction to Power Query
- Cleaning and shaping data
- Merging & Appending Queries
- Creating custom columns

- Handling missing and duplicate data
- Data types and formatting
- Types of Visuals: Bar, Line, Pie, Card, Matrix, etc.
- Conditional Formatting
- Tooltips, Slicers, and Filters
- Bookmarking and Drillthroughs
- Custom Visuals from Marketplace
- KPI Visuals and Gauges
- Using Themes and Templates
- Advanced Chart Interactions
- Basics of DAX: Syntax and functions
- Calculated Columns vs Measures
- Common DAX Functions:
- SUM, AVERAGE, COUNT
- CALCULATE, FILTER, ALL
- IF, SWITCH, RELATED
- End-to-End Dashboard from Real Dataset

Module 8: Machine Learning

- What is Machine Learning? Definitions and Types (Supervised, Unsupervised)
- ML vs. Traditional Programming
- Key Concepts: Features, Labels, Models, Training, Testing
- Tools Setup: scikit-learn, Jupyter Notebook, Datasets (e.g., Iris)
- ML Pipeline Overview: Data to Prediction
- Importance of Data Preprocessing in ML

- Handling Missing Data: Imputation (fillna()), Deletion
- Feature Scaling: Normalization, Standardization
- Encoding Categorical Data: One-Hot Encoding, Label Encoding
- Outlier Detection and Treatment: Z-Score, IQR Method
- Feature Engineering: Creating New Features, Polynomial Features
- Data Splitting: Train-Test Split
- Linear Regression
- Logistic Regression
- Decision Tree
- Random Forest
- AdaBoost
- XGBoost
- Support Vector Machine (SVM)
- K-Nearest Neighbors (KNN)
- Naive Bayes
- K-Means
- Hierarchical Clustering
- DBSCAN
- Principal Component Analysis (PCA)

Module 9: Deep Learning

- What is Deep Learning?
- Difference between ML and DL
- Biological vs Artificial Neurons
- Architecture of a Neural Network

- Perceptron and Multilayer Perceptron (MLP)
- Activation Functions (ReLU, Sigmoid, Tanh, Softmax)
- Loss Functions (MSE, Cross-Entropy)
- Backpropagation and Gradient Descent
- Optimizers (SGD, Adam, RMSProp)
- Regularization (Dropout, L1/L2)
- Convolutional Neural Networks (CNNs)
- Recurrent Neural Networks (RNNs)

Module 10: Natural Language Processing (NLP)

- What is NLP?
- Applications of NLP
- NLP vs NLU vs NLG
- Text vs Speech Processing
- Challenges in NLP
- Text Cleaning: Lowercasing, punctuation, stopwords
- Tokenization (word, sentence)
- Stemming vs Lemmatization
- POS (Part-of-Speech) Tagging
- Named Entity Recognition (NER)
- Bag of Words (BoW)
- TF-IDF (Term Frequency–Inverse Document Frequency)
- N-grams
- Word Embeddings:
- Word2Vec

Complete Project Details:

- Exploratory Data Analysis of E-Commerce Sales Trends
- Customer Segmentation Using Retail Transaction Data

- Visualizing COVID-19 Impact on Global
- Earthquake Prediction Model
- Car Price Prediction Model
- Rainfall Prediction Model
- Weather Prediction Model
- Customer Churn Prediction Model
- Heart Disease Prediction Model
- Student Performance Prediction
- Loan Default Prediction
- House Price Prediction
- Clustering Supermarket Customers with K-Means
- Handwritten Digit Recognition
- Image Classification of Cats vs. Dogs
- Medical Image Segmentation for Tumor Detection