Data Science

Course Modules & Topics

📊 Module 1: Introduction to Data Science

- What is Data Science?
- Roles: Data Scientist vs Data Analyst vs ML Engineer
- Lifecycle of a Data Science Project
- Tools: Jupyter, Anaconda, Google Colab

📈 Module 2: Python for Data Science

- NumPy: Arrays, Broadcasting, Mathematical Functions
- Pandas: DataFrames, Series, Indexing, Merging, Grouping
- Matplotlib & Seaborn: Data Visualization Basics
- Handling Missing Data, Filtering, Sorting

📉 Module 3: Statistics & Probability

- Descriptive Statistics (Mean, Median, Mode, Variance)
- Probability Distributions (Normal, Binomial, Poisson)
- Hypothesis Testing & Confidence Intervals
- Correlation & Covariance

Module 4: Data Preprocessing & Cleaning

- Data Types & Conversion
- Outlier Detection and Handling
- Normalization & Standardization
- Encoding Categorical Data (Label, One-Hot)

Module 5: Machine Learning Basics

- Supervised vs Unsupervised Learning
- Regression (Linear, Logistic)
- Classification (KNN, Decision Tree, SVM)
- Clustering (K-Means, Hierarchical)

📈 Module 6: Model Evaluation & Tuning

- Train-Test Split, Cross Validation
- Metrics: Accuracy, Precision, Recall, F1, ROC
- Confusion Matrix
- Hyperparameter Tuning (Grid Search, Random Search)

📂 Module 7: Real-world Project

- End-to-end Data Science Project
- Data Collection → Cleaning → Modeling → Deployment
- Use case: e.g., Titanic Survival Prediction, Sales Forecasting, etc.

🌐 Module 8: Capstone and Deployment

- Creating a Portfolio Project
- Deploying with Streamlit or Flask
- Git & Version Control
- Sharing Projects on GitHub / Kaggle