45 Days - Foundation in Data Science with Python

Module 1: Introduction to Data Science (1 Day)

- Overview of Data Science and its applications
- Introduction to key concepts and tools

Module 2: Python for Data Science (4 Days)

- Python basics: Variables, control flow, functions, and modules
- Libraries: NumPy, Pandas, Matplotlib, Seaborn

Module 3: Data Collection and Cleaning (5 Days)

- Data collection techniques: CSV, JSON, APIs
- Data preprocessing and cleaning, feature engineering

Module 4: Exploratory Data Analysis (5 Days)

- Summary statistics and visualizations
- Identifying trends and patterns

Module 5: Introduction to Machine Learning (5 Days)

- Overview of supervised and unsupervised learning
- Scikit-learn basics: Classification and regression

Module 6: Supervised Learning (10 Days)

- Linear and Logistic Regression
- K-Nearest Neighbors (KNN)
- Model evaluation metrics

Module 7: Unsupervised Learning (8 Days)

- K-Means Clustering, Hierarchical Clustering
- Principal Component Analysis (PCA)

Module 8: Real-Life Projects (7 Days)

- Hands-on projects like predictive modeling and clustering

90 Days - Intermediate Data Science with Python

Includes All 45-Day Modules

Additional Modules:

Module 8: Advanced Data Cleaning and Feature Engineering (10 Days)

- Handling large datasets and advanced feature extraction
- Working with time-series and categorical data

Module 9: Advanced Machine Learning Techniques (15 Days)

- Decision Trees, Random Forests, and Gradient Boosting
- Model interpretation and feature importance

Module 10: Model Optimization and Evaluation (10 Days)

- Cross-validation, hyperparameter tuning (GridSearchCV, RandomizedSearchCV)
- Performance metrics (confusion matrix, ROC-AUC curves)

Module 11: Introduction to Deep Learning (15 Days)

- Neural Networks and TensorFlow/Keras basics
- CNNs and RNNs introduction with small projects

Module 12: Real-Life Projects and Case Studies (15 Days)

- Projects on regression, classification, and clustering
- End-to-end implementation of data science pipelines