Python for Data Engineering and Machine Learning

Course Duration: 05 Days

Chapter 01: Introduction to Data Engineering
- Overview of Data Engineering
- Role of Python in Data Engineering
- Importance of Data Systems
- Lab Exercise: Setting Up Python Environment and Exploring Basic Data Structures

Chapter 02: Data Acquisition with Python
- Web Scraping Techniques
- Interacting with APIs
- Connecting with Databases
- Lab Exercise: Web Scraping Data from a Sample Website

Chapter 03: Data Wrangling and Cleaning
- Using Python Libraries for Cleaning
- Transforming and Enriching Data
- Creating High-Quality Data Sets
- Lab Exercise: Cleaning and Preparing a Messy Dataset

Chapter 04: Implementing Custom Business Logic
- Incorporating Complex Logic with Python
- Executing Logic in Data Pipelines
- Triggering Actions in Business Processes
- Lab Exercise: Developing and Testing Custom Business Logic

Chapter 05: Data Storage and Retrieval
- Accessing Data in SQL and NoSQL Databases
- Utilizing Cloud Storage Services
- Serialization of Data for Efficiency
- Lab Exercise: Storing and Retrieving Data from a Database

Chapter 06: Python in Machine Learning
- Overview of Python's Role in Machine Learning
- Data Preprocessing in Machine Learning
- Model Selection, Training, and Evaluation
- Lab Exercise: Building a Simple Machine Learning Model with Python

Chapter 07: Python Code Development

- Cleaning and Preparing Data
- Handling Missing Values
- Formatting, Normalizing, and Binning Data
- Lab Exercise: Python Code for Data Cleaning and Preparation

Chapter 08: Data Manipulation and Pipelines

- Using Dataframes for Manipulation
- Summarizing Data and Understanding Distributions
- Building Efficient Data Pipelines
- Lab Exercise: Creating a Data Pipeline with Python

Chapter 09: Exploratory Data Analysis with Python

- Applying Analytical Techniques
- Leveraging Pandas, Numpy, and Scipy
- Real-world Dataset Exploration
- Lab Exercise: Exploratory Data Analysis on a Given Dataset

Chapter 10: Regression Models with scikit-learn

- Building and Evaluating Regression Models
- Prediction and Decision Making with Models
- Practical Applications in Data Engineering
- Lab Exercise: Building and Evaluating Regression Models with scikit-learn

Chapter 11: Supervised Machine Learning with Python

- Fundamentals of Supervised Learning
- Classification and Regression Techniques
- Implementing Supervised Models in Real-world Scenarios
- Lab Exercise: Applying Supervised Learning Techniques on Datasets

Chapter 12: Clustering and Unsupervised Learning

- Introduction to Clustering Algorithms
- Implementing K-means, Hierarchical, and DBSCAN Clustering
- Real-World Applications of Clustering
- Lab Exercise: Clustering Analysis on Complex Datasets