

Python for Data Engineering & Machine Learning Foundations

Duration: 5 Days

Format: Theory + Hands-on Labs

Level: Beginner to Intermediate

DC: Koenig DC

Course Objectives

The objectives of this course are to:

1. Build a **strong foundation in Python programming** for data-driven applications.
 2. Enable learners to **connect, read, and manage data using MongoDB** from Python.
 3. Teach **robust data preprocessing techniques** required for machine learning workflows.
 4. Develop skills in **Exploratory Data Analysis (EDA)** to extract meaningful insights.
 5. Provide hands-on experience in **data visualization** for analysis and storytelling.
 6. Introduce **core Machine Learning concepts** to prepare learners for advanced ML and AI courses.
-

Course Learning Outcomes

By the end of this course, participants will be able to:

1. Write clean, efficient Python code using industry-standard practices.
 2. Connect Python applications to **MongoDB**, perform CRUD operations, and handle real datasets.
 3. Clean, preprocess, and transform raw data for machine learning readiness.
 4. Perform **Exploratory Data Analysis (EDA)** to identify trends, patterns, and anomalies.
 5. Create meaningful visualizations using Python libraries for effective data communication.
 6. Understand **machine learning fundamentals**, workflows, and evaluation concepts.
 7. Build a complete **data-to-model pipeline** using Python.
-

Detailed 5-Day Course TOC

DAY 1 – Python Programming Foundations (In Depth)

Topics Covered

- Introduction to Python for Data & ML
- Python syntax, indentation, and execution model
- Data types: int, float, string, boolean
- Control flow: if-else, loops
- Data structures:
 - Lists, tuples, sets, dictionaries
- Functions and modular programming
- File handling (CSV, TXT, JSON)
- Exception handling and debugging techniques
- Introduction to Numpy and Pandas

Hands-On Labs

- Python data manipulation exercises
 - File read/write operations
 - Error handling scenarios
 - Demo on Numpy and Pandas
-

DAY 2 – Data Connectivity with MongoDB

Topics Covered

- Introduction to databases and NoSQL concepts
- MongoDB architecture and document model
- Installing and setting up MongoDB
- Connecting Python with MongoDB using PyMongo
- CRUD operations:
 - Insert, read, update, delete documents
- Querying, filtering, sorting, and aggregation
- Handling nested and unstructured data
- Best practices for data ingestion pipelines

Hands-On Labs

- Build a Python–MongoDB data connector

- Load real-world data into MongoDB
 - Query and analyze stored data
-

DAY 3 – Data Preprocessing in Python

Topics Covered

- Importance of data preprocessing in ML
- Data cleaning techniques
- Handling missing values (mean, median, mode, imputation)
- Handling duplicates and inconsistent data
- Data type conversion and normalization
- Feature scaling (standardization, normalization)
- Encoding categorical variables (Label Encoding, One-Hot Encoding)
- Train-test split concepts

Hands-On Labs

- Data cleaning pipeline using Pandas
 - Feature transformation and preparation for ML
-

DAY 4 – Exploratory Data Analysis (EDA) & Visualization

Topics Covered

Exploratory Data Analysis (EDA)

- Descriptive statistics
- Understanding data distributions
- Correlation and covariance analysis
- Detecting outliers and anomalies
- Hypothesis generation from data

Data Visualization

- Principles of effective visualization
- Matplotlib fundamentals
- Seaborn for statistical visualization
- Plot types:

- Line, bar, scatter, histogram, boxplot, heatmap
- Choosing the right visualization

Hands-On Labs

- Full EDA on a real dataset
 - Visual storytelling using plots
-

DAY 5 – Machine Learning Concepts & Workflow

Topics Covered

- What is Machine Learning?
- Types of ML:
 - Supervised
 - Unsupervised
- ML workflow:
 - Data → Features → Model → Evaluation
- Overview of common algorithms:
 - Linear Regression
 - Logistic Regression
 - Decision Trees
 - K-Means Clustering
- Model evaluation concepts:
 - Accuracy, precision, recall
 - Overfitting vs underfitting
- Introduction to Scikit-learn

Hands-On Labs

- Build a simple ML model in Python
 - Evaluate model performance
 - End-to-end mini project
-

Who Should Attend

- Beginners in Data Science or ML

- Software developers moving into data roles
- Engineering students
- Professionals preparing for AI/ML programs