KOENIG SOLUTIONS · COURSE OUTLINE

Python for Data Engineering and Machine Learning (Python Institute)

Learn to build data engineering pipelines and machine learning models using Python, from data acquisition to clustering.

5 Days	12	36	12
DURATION	MODULES	TOPICS	LAB EXERCISES

COURSE OVERVIEW

This course covers data engineering fundamentals and the role of Python in building data systems, including data acquisition through web scraping, APIs and databases, followed by data wrangling, cleaning and the implementation of custom business logic within data pipelines. It then addresses data storage and retrieval across SQL, NoSQL and cloud services, before moving into Python's role in machine learning, covering data preprocessing, model selection and evaluation. Later modules focus on Python code development for data cleaning, dataframe manipulation and exploratory data analysis using Pandas, Numpy and Scipy, concluding with regression models in scikit-learn, supervised machine learning techniques, and clustering and unsupervised learning methods such as K-means, hierarchical and DBSCAN.

WHO THIS COURSE IS FOR

- **Data Engineers**
- **Machine Learning Engineers**
- **Data Analysts**
- **Python Developers**
- **Data Scientists**

LEARNING OBJECTIVES

- Explain the role of Python in data engineering and data systems
- Apply web scraping, API and database techniques for data acquisition
- Clean, transform and enrich data into high-quality data sets
- Implement custom business logic within data pipelines
- Configure data storage and retrieval across SQL, NoSQL and cloud services
- Apply data preprocessing techniques for machine learning
- Analyse datasets through exploratory data analysis with Pandas and Numpy
- Build regression, supervised and clustering models with scikit-learn

COURSE SYLLABUS

12 Modules, 36 Topics

01 Lab Exercise

Introduction to Data Engineering

TOPICS

Overview of Data Engineering · Role of Python in Data Engineering · Importance of Data Systems

02 Lab Exercise

Data Acquisition with Python

TOPICS

Web Scraping Techniques · Interacting with APIs · Connecting with Databases

03 Lab Exercise

Data Wrangling and Cleaning

TOPICS

Using Python Libraries for Cleaning · Transforming and Enriching Data · Creating High-Quality Data Sets

04 Lab Exercise

Implementing Custom Business Logic

TOPICS

Incorporating Complex Logic with Python · Executing Logic in Data Pipelines · Triggering Actions in Business Processes

05 Lab Exercise

Data Storage and Retrieval

TOPICS

Accessing Data in SQL and NoSQL Databases · Utilizing Cloud Storage Services · Serialization of Data for Efficiency

06 Lab Exercise

Python in Machine Learning

TOPICS

Overview of Python's Role in Machine Learning · Data Preprocessing in Machine Learning · Model Selection, Training, and Evaluation

07 Lab Exercise

Python Code Development

TOPICS

Cleaning and Preparing Data · Handling Missing Values · Formatting, Normalizing, and Binning Data

08 Lab Exercise

Data Manipulation and Pipelines

TOPICS

- Using Dataframes for Manipulation
- Summarizing Data and Understanding Distributions
- Building Efficient Data Pipelines

09

Lab Exercise

Exploratory Data Analysis with Python

TOPICS

Applying Analytical Techniques · Leveraging Pandas, Numpy, and Scipy · Real-world Dataset Exploration

10

Lab Exercise

Regression Models with scikit-learn

TOPICS

Building and Evaluating Regression Models · Prediction and Decision Making with Models · Practical Applications in Data Engineering

11

Lab Exercise

Supervised Machine Learning with Python

TOPICS

- Fundamentals of Supervised Learning
- Classification and Regression Techniques
- Implementing Supervised Models in Real-world Scenarios

12

Lab Exercise

Clustering and Unsupervised Learning

TOPICS

- Introduction to Clustering Algorithms
- Implementing K-means, Hierarchical, and DBSCAN Clustering
- Real-World Applications of Clustering

LAB EXERCISES

Practise What You Learn

#	MODULE	ACTIVITY
01	**Introduction to Data Engineering**	Setting Up Python Environment and Exploring Basic Data Structures
02	**Data Acquisition with Python**	Web Scraping Data from a Sample Website
03	**Data Wrangling and Cleaning**	Cleaning and Preparing a Messy Dataset
04	**Implementing Custom Business Logic**	Developing and Testing Custom Business Logic
05	**Data Storage and Retrieval**	Storing and Retrieving Data from a Database
06	**Python in Machine Learning**	Building a Simple Machine Learning Model with Python

#	MODULE	ACTIVITY
07	**Python Code Development**	Python Code for Data Cleaning and Preparation
08	**Data Manipulation and Pipelines**	Creating a Data Pipeline with Python
09	**Exploratory Data Analysis with Python**	Exploratory Data Analysis on a Given Dataset
10	**Regression Models with scikit-learn**	Building and Evaluating Regression Models with scikit-learn
11	**Supervised Machine Learning with Python**	Applying Supervised Learning Techniques on Datasets
12	**Clustering and Unsupervised Learning**	Clustering Analysis on Complex Datasets

WHAT IS INCLUDED

Everything You Need to Succeed

- Koenig course material for every module
- Delivery by certified instructors and subject matter experts
- Hands-on exercises throughout the course
- Course completion certificate from Koenig

ABOUT KOENIG

Who We Are

Established in 1993, Koenig Solutions is a globally recognised training provider with 35+ technology partnerships, 5,000+ training programmes and a network of certified instructors and subject matter experts, supporting individuals and enterprise teams worldwide.

FLEXIBLE DELIVERY

Learn the Way That Suits You

Classroom Training | Live Online Classes (ILO) | Fly-Me-A-Trainer (FMAT) | Flexi (Self-Paced) | 1-on-1 Training

Ready to Get Started?

Talk to a learning advisor about schedules, group training and customised delivery.

Visit koenig-solutions.com | info@koenig-solutions.com