

KOENIG SOLUTIONS • COURSE OUTLINE

# Data Science with Python and Big Data Integration

**13**

MODULES

**92**

TOPICS

**2**

HANDS-ON ACTIVITIES

## LEARNING OBJECTIVES

- Explain core concepts of data science and data analytics
- Apply exploratory data analysis using graphical techniques
- Interpret statistics such as distribution, dispersion and correlation
- Implement mathematical computing with NumPy and SciPy
- Manipulate and clean data using Pandas DataFrames
- Build supervised and unsupervised machine learning models
- Apply natural language processing techniques using Scikit-Learn
- Develop visualizations with Matplotlib and scrape data with BeautifulSoup

# 13 Modules, 92 Topics

00

## Course Overview

### TOPICS

Course Overview

01

## Data Science Overview

### TOPICS

Introduction to Data Science · Different Sectors Using Data Science · Purpose and Components of Python

02

## Data Analytics Overview

### TOPICS

- Data Analytics Process
- Knowledge Check
- Exploratory Data Analysis (EDA)
- EDA-Quantitative Technique
- EDA - Graphical Technique
- Data Analytics Conclusion or Predictions
- Data Analytics Communication
- Data Types for Plotting
- Data Types and Plotting

03

## Statistical Analysis and Business Applications

### TOPICS

- Introduction to Statistics
- Statistical and Non-statistical Analysis
- Major Categories of Statistics
- Statistical Analysis Considerations
- Population and Sample
- Statistical Analysis Process
- Data Distribution
- Dispersion
- Histogram
- Testing
- Correlation and Inferential Statistics

04

## Python Environment Setup and Essentials

### TOPICS

- Anaconda
- Installation of Anaconda Python Distribution (contd.)
- Data Types with Python
- Basic Operators and Functions

**05**

## Mathematical Computing with Python (NumPy)

### TOPICS

Introduction to Numpy · Demo 01-Creating and Printing an ndarray · Knowledge Check · Class and Attributes of ndarray · Basic Operations · Copy and Views · Mathematical Functions of Numpy

### ACTIVITY

01 Sequence it Right

02 Slice It

**06**

## Scientific Computing with Python (SciPy)

### TOPICS

- Introduction to SciPy
- SciPy Sub Package - Integration and Optimization
- Knowledge Check
- SciPy sub package
- Demo - Calculate Eigenvalues and Eigenvector
- Knowledge Check
- SciPy Sub Package - Statistics, Weave and IO

**07**

## Data Manipulation with Pandas

### TOPICS

- Introduction to Pandas
- Knowledge Check
- Understanding DataFrame
- View and Select Data Demo
- Missing Values
- Data Operations
- Knowledge Check
- File Read and Write Support
- Knowledge Check-Sequence it Right
- Pandas Sql Operation

**08**

## Machine Learning with Scikit-Learn

### TOPICS

- Machine Learning Approach
- Understand data sets and extract its features
- Identifying problem type and learning model
- How it Works
- Train, test and optimizing the model
- Supervised Learning Model Considerations
- Knowledge Check
- Scikit-Learn
- Knowledge Check
- Supervised Learning Models - Linear Regression
- Supervised Learning Models - Logistic Regression
- Unsupervised Learning Models
- Pipeline
- Model Persistence and Evaluation

**09**

## Natural Language Processing with Scikit Learn

### TOPICS

NLP Overview · NLP Applications · Knowledge Check · NLP Libraries-Scikit · Extraction Considerations · Scikit Learn-Model Training and Grid Search

**10**

## Data Visualization in Python Using Matplotlib

### TOPICS

Introduction to Data Visualization · Knowledge Check · Line Properties · (x,y) Plot and Subplots · Knowledge Check · Types of Plots

**11**

## Web Scraping with BeautifulSoup

### TOPICS

Web Scraping and Parsing · Knowledge Check · Understanding and Searching the Tree · Navigating options · Demo 3 Navigating a Tree · Knowledge Check · Modifying the Tree · Parsing and Printing the Document

**12**

## Python Integration with Hadoop MapReduce and Spark

### TOPICS

- Why Big Data Solutions are Provided for Python
- Hadoop Core Components
- Python Integration with HDFS using Hadoop Streaming
- Demos
- Knowledge Check
- Python Integration with Spark using PySpark

#### 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](https://koenig-solutions.com)

✉ [info@koenig-solutions.com](mailto:info@koenig-solutions.com)