

Next Level Python For Data Science

Duration: 5 Days **Course Code: TTPS4876** **Delivery Method: Company Event**

Overview:

Learn advanced Python skills for data analysis, visualizations, and big data processing.

This course explores using Python for data scientists to perform exploratory data analysis, complex visualizations, and large-scale distributed processing using Big Data. In this course you'll learn about essential mathematical and statistics libraries such as NumPy, Pandas, SciPy, SciKit-Learn, along with frameworks like TensorFlow and Spark. It also covers visualization tools like matplotlib, PIL, and Seaborn.

Target Audience:

Data Scientists, Data Engineers, and Software Engineers who are experienced with basic Python and data science.

Objectives:

- Join an engaging hands-on learning environment, where you'll learn:
- How to work with Python in a Data Science context
- How to use NumPy, Pandas, and Matplotlib
- How to create and process images with PIL
- How to visualize with Seaborn
- Key features of SciPy and SciKit Learn
- How to interact with Spark using DataFrames
- How to use SparkSQL, MLlib, and Big Data streaming
- This course has a 50% hands-on labs to 50% lecture ratio with engaging instruction, demos, group discussions, labs, and project work.

Prerequisites:

Before attending this course, you should have:

- A solid data analytics and data science background
 - Python experience
- Topics are covered in-depth and are geared for experienced students who have taken one of the prerequisite courses below or have practical hands-on experience.
- Python for Data Science Primer
 - JumpStart to Python for Data Science
 - Applied Python for Data Science and Engineering

Content:

Python Review

- Python Language
- Essential Syntax
- Lists, Sets, Dictionaries, and Comprehensions
- Functions
- Classes, Modules, and imports
- Exceptions

iPython

- iPython basics
- Terminal and GUI shells
- Creating and using notebooks
- Saving and loading notebooks
- Ad hoc data visualization
- Web Notebooks (Jupyter)

NumPy

- NumPy basics
- Creating arrays
- Indexing and slicing
- Large number sets
- Transforming data
- Advanced tricks

SciPy

- What can SciPy do?
- Most useful functions
- Curve fitting
- Modeling
- Data visualization
- Statistics

SciPy subpackages

- Clustering
- Physical and mathematical Constants
- FFTs
- Integral and differential solvers
- Interpolation and smoothing
- Input and Output
- Linear Algebra
- Image Processing
- Distance Regression
- Root-finding
- Signal Processing
- Sparse Matrices
- Spatial data and algorithms
- Statistical distributions and functions
- C/C++ Integration

pandas

- pandas overview
- Dataframes
- Reading and writing data
- Data alignment and reshaping
- Fancy indexing and slicing
- Merging and joining data sets

matplotlib

- Creating a basic plot
- Commonly used plots
- Ad hoc data visualization
- Advanced usage
- Exporting images

The Python Imaging Library (PIL)

- PIL overview
- Core image library
- Image processing
- Displaying images

seaborn

- Seaborn overview
- Bivariate and univariate plots
- Visualizing Linear Regressions
- Visualizing Data Matrices
- Working with Time Series data

SciKit-Learn Machine Learning Essentials

- SciKit overview
- SciKit-Learn overview
- Algorithms Overview
- Classification, Regression, Clustering, and Dimensionality Reduction
- SciKit Demo

TensorFlow Overview

- TensorFlow overview
- Keras
- Getting Started with TensorFlow

PySpark Overview

- Python and Spark
- SciKit-Learn vs. Spark MLlib
- Python at Scale
- PySpark Demo

RDDs and DataFrames

- DataFrames and Resilient Distributed Datasets (RDDs)
- Partitions
- Adding variables to a DataFrame
- DataFrame Types
- DataFrame Operations
- Dependent vs. Independent variables
- Map/Reduce with DataFrames

Spark SQL

- Spark SQL Overview
- Data stores: HDFS, Cassandra, HBase, Hive, and S3
- Table Definitions
- Queries

Spark MLlib

- MLlib overview
- MLlib Algorithms Overview
- Classification Algorithms
- Regression Algorithms
- Decision Trees and forests
- Recommendation with ALS
- Clustering Algorithms
- Machine Learning Pipelines
- Linear Algebra (SVD, PCA)
- Statistics in MLlib

Spark Streaming

- Streaming overview
- Integrating Spark SQL, MLlib, and Streaming

Further Information:

For More information, or to book your course, please call us on Head Office 01189 123456 / Northern Office 0113 242 5931

info@globalknowledge.co.uk

www.globalknowledge.com/en-gb/

Global Knowledge, Mulberry Business Park, Fishponds Road, Wokingham Berkshire RG41 2GY UK