skillsoft global knowledge™

NEXT LEVEL PYTHON FOR DATA SCIENCE | WORKING WITH LIBRARIES, FRAMEWORKS, AND VISUALIZATION TOOLS

Course Code: 100671

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.

What You'll Learn

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.

Who Needs to Attend

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

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.

NEXT LEVEL PYTHON FOR DATA SCIENCE | WORKING WITH LIBRARIES, FRAMEWORKS, AND VISUALIZATION TOOLS

Course Code: 100671

VIRTUAL CLASSROOM LIVE	$3,504 CAD	5 Day

Virtual Classroom Live Outline

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 MLib

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

- Statistics in MLib

Spark Streaming

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

Sep 29 - Oct 3, 2025 | 10:00 AM - 6:00 PM EST

Dec 8 - 12, 2025 | 10:00 AM - 6:00 PM EST

NEXT LEVEL PYTHON FOR DATA SCIENCE | WORKING WITH LIBRARIES, FRAMEWORKS, AND VISUALIZATION TOOLS

Course Code: 100671

PRIVATE GROUP TRAINING	5 Day

Visit us at www.globalknowledge.com or call us at 1-866-716-6688.

Date created: 7/30/2025 7:49:15 AM