PY010 - *Python Data Analytics with Pandas, NumPy, SciPy and Matplotlib*

DURATION: 2 Days; Instructor-led

WHAT YOU WILL LEARN
In a world increasingly centralized around information technology, huge amounts of data are produced and stored each day. Often these data come from automatic detection systems, sensors, and scientific instrumentation, or you produce them daily and unconsciously every time you make a withdrawal from the bank or make a purchase, when you record various blogs, or even when you post on social networks.

Python and its supporting packages are one of the famous tools in dealing with Data Analytics. This is the course to guide you how to use Python in Data Analytics.

This course will let you explore the latest Python tools and techniques to help you tackle the world of data acquisition and analysis. You will review scientific computing with NumPy and visualization with matplotlib.

AUDIENCE
This course is mostly for Python programmers who want to deal with the Data Analytics.

PREREQUISITES
A basic understanding of Python language.

A basic understanding of Data Analytics will be an added advantage.

METHODOLOGY
This program will be conducted with interactive lectures, PowerPoint presentation, discussions and practical exercise

COURSE OUTLINES

Module 1 - An Introduction to Data Analysis
- Data Analysis
- Knowledge Domains of the Data Analyst
- Understanding the Nature of the Data
- The Data Analysis Process
- Quantitative and Qualitative Data Analysis
- Open Data
- Python and Data Analysis

Module 2 – Introduction to the Python World
- The Language
- Python 2 and Python 3
- The IDEs for Python
- Packages for Data Analysis

Module 3 – The NumPy Library
- The History
- Ndarray: The Heart of the Library
- Basic Operations
- Indexing, Slicing, and Iterating
- Conditions and Boolean Arrays
- Shape Manipulation
- Array Manipulation
- General Concepts
- Structured Arrays
- Reading and Writing Array Data on Files

Module 4 – The pandas Library—An Introduction
- pandas: The Python Data Analysis Library
- Installation of pandas
- Testing Your pandas Installation
- Getting Started with pandas
- Introduction to pandas Data Structures
- Other Functionalities on Indexes
- Operations Between Data Structures
- Function Application and Mapping
- Sorting and Ranking
- Correlation and Covariance
- "Not a Number" Data
- Hierarchical Indexing and Leveling

Module 5 – pandas: Reading and Writing Data
- I/O API Tools
- CSV and Textual Files
- Reading Data in CSV or Text Files
- Reading and Writing HTML Files
- Reading Data from XML
- Reading and Writing Data on Microsoft Excel Files
- JSON Data
- The Format HDF5
- Pickle—Python Object Serialization
- Interacting with Databases
- Reading and Writing Data with a NoSQL Database: MongoDB

Module 6 – pandas in Depth: Data Manipulation
- Data Preparation
- Concatenating
- Data Transformation
- Discretization and Binning
- Permutation
- String Manipulation
- Data Aggregation
- Group Iteration
- Advanced Data Aggregation

Module 7 – Data Visualization with matplotlib
- The matplotlib Library
- The IPython and IPython QtConsole
- The matplotlib Architecture
- pyplot
- The Plotting Window
- Using the kwargs
- Adding Elements to the Chart
- Saving Your Charts

- Handling Date Values
- Chart Typology
- Line Charts
- Histograms
- Bar Charts
- Pie Charts
- Advanced Charts
- The mplot3d Toolkit
- Multi-Panel Plots