Python - data analysis (Pandas)

Course code: PYTHON_DATAN

The course is designed for all those who are looking for a flexible data analysis tool, those interested in data processing in the Python programming language, who plan to use it for data manipulation, analysis and visualization, respectively.

What we teach you:
- Course participants will learn to use the Pandas library and other supporting libraries that are needed to work with data, analyze and visualize it.

Who the course is for:
- Data Scientist, data analysts, especially in the Big Data environment, is the primary auditor for this intensive course.
- Software developers who are familiar with Python at least at intermediate and advanced levels and aim to create data-intensive allocations using the SPARK Big Data (Cloud) engine.
- Data architects

Required skills:
- Basic knowledge of Python at the PYTHON_INTRO level

Teaching methods:
- Professional explanation with practical samples and examples.

Teaching materials:
- Powerpoint handouts and module printouts.

Course syllabus:
Introduction to IPython
- Tools for online work with data (Jupyter Notebook and Jupyter Lab)
- Export output to PDF and other formats

Overview of data structures in Python
- Sheet, tuple, etc.

Pandas module
- Objects for working with tay (Series, DataFrame)

NumPy module
- Work with multidimensional data structures

Import data from different sources
- CSV
- HTML
- JSON
- SQL

Work with data
- Transformation
- Filling in missing values
- Mapping
- Replacement

Advanced work with data
- Concat merge join
- Pivoting, stacking, unstacking
- Melting

Aggregation and grouping of data

GOPAS®

- Split apply combine
- Transformation
- Filtering

Time lines
- Representation of date, time, and time intervals
- Timestamp, Timedelta, DateTimeIndex, and more

Visualization
- IPython graphing
- Loading data into charts from Pandas objects
- Work with charts of different types
- Data aggregation into graphs