

Course code: **PYTHON/BP**

Course title: **Data analysis in Python - best practices**

Days: 2

Description:

Course intended for

The course is designated for data analysts who wish to begin using a wide range of analytical libraries available in Python. The course is the most beneficial to people interested in programming and are already familiar with the analyst profession, as the course focuses on tools and not only techniques. The course aims to decrease the distance between the analyst and the programmer. In addition, the participants may better understand the use of analytical tools in the production language environment.

Course objective

The participants get familiarised with Python and the use of various analytical tools available in that language. The course features a combination of various tools, which in turn enables execution of more complex analyses using an extensive range of techniques. Furthermore, algorithms in Python may very often be easily transferred to the production environment in a relatively short time.

Course strengths

The course is conducted by people dealing with the problem of data analysis using Python in their everyday work and hold practical experience in that area. Therefore, the course often goes beyond the available materials. Moreover, the programme is regularly updated because of the fast development of solutions discussed during the course.

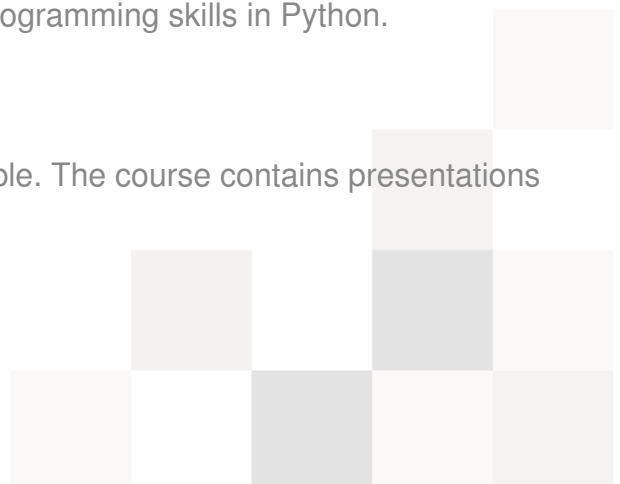
Requirements

The course requires basic programming skills in any language (a version including an introduction to Python - 1 additional day) or basic programming skills in Python.

Course parameters

2 working days, 2*7 working hours, group 8-10 people. The course contains presentations and coding workshops.

Course curriculum:



- Introduction
 - Purpose of the course
 - Python
 - History
 - Characteristics
 - Comparison with other languages
 - Why Python?
- Basic environment installation
 - Available versions
 - Python distributions
 - Python installation
 - Libraries installation
 - PyPI
 - Setuptools, Pip
 - VirtualEnv
 - interpreter
 - basic
 - IPython
 - Version control with Git
 - Installation
 - Basic operations
 - Tools



- Introduction to Python (optional day)
 - Basic characteristics of the language
 - Syntax
 - Data structures
 - Instructions
 - Functions
 - Objects
 - Modules and packages
 - Basic library
- IPython
 - Installation
 - Operating modes
 - Basic interactive command execution
 - History of commands
 - IPython Notebook
 - Advanced use
- NumPy
 - Installation
 - Boards, vectors, matrixes
 - Functions
 - Data analysis
 - Recording and reading files
 - Linear algebra



- SciPy
 - Installation
 - Basics
 - Selected packages
 - Statistics
 - Optimisation
 - Interpolation
 - Integration
- Matplotlib
 - Installation
 - Concept
 - Basic charts
 - Advanced functions
 - Image saving
- Pandas
 - Installation
 - Data structures
 - Basic operations
 - Working with data
 - Basic statistics
 - Charts
- NLTK
 - Installation



- Basics
- Selected functions
 - Division into tasks
 - Division into expressions (tokenisation)
 - Data
- Scikit-learn
 - Installation
 - Basics
 - Selected functions
 - Linear regression
 - Logistic regression
 - Clustering
 - Decision trees

