



Python3 - Version: 1
01 May 2021

Data Science with Python



Data Science with Python

Python3 - Version: 1

 **2 days Course**

Description:

Python is at the top of all other languages and is the most popular language used by data scientists. It is the preferred programming language for data scientists because it is very intuitive and easy to understand and has fast development time.

Data science developers in Python can use a lot of efficient, open-source and ready to use packages. There are also excellent interactive shells, notebooks, frameworks and IDEs for this topic

Data science is becoming more and more popular with the advent of Machine Learning that based on Data-Science development

In this course concentrated workshop we will start by introduction of Anaconda and jupyter notebook, continue with numpy arrays creation, access and manipulation, while emphasizing performance efficiency and memory management. We will continue with deep dive into the Pandas and its unique data structures, their Index preserving operations, broadcasting, combining datasets and aggregations. We will use matplotlib module for data visualizations.

Intended audience:

This course is intended for Python programmers and Project leads, Matlab programmers and data scientists with python background who want improve their data analysis process and data visualization

Prerequisites:

python base knowledge and base understanding of Matlab-like matrix data structure

Objectives:

Topics:

Module 01 – Introduction

- What is data science
- Getting Started

Module 02 – Ipython

- IPython - the Interactive Python
- The Jupyter notebook
- The Documentation and Auto-completion
- Magic Commands
- In and Out operators

Module 03 – Numpy

- Numpy array memory efficiency
- Numpy Array attributes
- Array Slicing and Fancy indexing
- Universal functions (ufunc)
- Array arithmetic
- Arrays Broadcasting
- Comparison operators on arrays and Boolean Masks
- View and Copies
- Array Aggregators and other functions

Module 04 – Pandas

- Pandas built-in types: Series and DataFrame
- Constructing Series and DataFrame
- Loading csv files
- Indexing, Selection and loc and iloc indexers
- Operators and ufuncs
- Index Alignment
- Working with NaN values

- Combining Datasets – concat
- Combining Datasets - merge
- groupby aggregations, filter and apply

Module 05 – Plotting

- Introduction to matplotlib
- plot function and parameters
- Scatter Plots
- Figure and subplots
- Histograms and 2d Histograms
- Pies
- Combine plots
- Text and Annotations
- Custom ticks names
- DateTime ranges, Date tickers and Date tick labels
- DateTime plots
- Intro to Three-Dimensional Plotting