

Module Title : Python for Data Science

Duration : 5 days

Overview

This course aims to be a comprehensive course in using the power of Python to read data from various sources, prepare the data for analysis, create exploratory visualizations and an introduction to machine learning algorithms. At the end of the course participants will be sufficiently equipped to start their journey towards becoming a full fledged Data Scientist.

Audience

This short course intended audiences are IT professionals, Data Analyst, Data Scientists and professionals who want to learn about Python programming. The programme is suitable for personnel with some experience in programming, data management and business reporting. The effective number of participants for this course is between 10 – 12.

Pre-requisite

Basic knowledge of programming is preferable.

Objectives

After completing this module, participants will learn the following

1. Identify types of data that can be extracted
2. Cleaning and preparing data
3. Common approach in analyzing data
4. Visualization using Python
5. Python machine learning

Course Outline

The course content will be delivered according to this schedule (E.g. typically a short course has a maximum of 7 hours per day, 9:00 am to 5:00 pm with an hour for lunch)

DAY1	
0900 – 0930	Introduction to Python and Data Science
0930 – 1040	Installing and setting up the Python environment
1040 – 1100	Morning Break
1100 – 1200	Basic Python syntax and data types
1200 – 1300	NumPy, sets, list, and tuples data storage
1300 – 1400	Lunch
1400 – 1440	Control structures in Python
1440 – 1540	Creating and using functions, date and time functions
1540 – 1600	Afternoon Break
1600 – 1700	Laboratory exercise – My first Python program

DAY2	
0900-1000	Comprehension of the use of lists and indexing data with dictionaries
1000-1040	Sub-setting (Lists, Matching, Handling Missing Values)
1040-1100	Morning Break
1100-1300	Reading and writing real-world data (text file, CSV and SQL)
1300-1400	Lunch
1400-1500	Introduction to pandas and managing Data Frames
1500-1540	Using pandas for data conditioning
1540-1600	Afternoon Break
1600-1700	Laboratory exercise – Working with Python pandas data frame

DAY3	
0900-1000	Shaping data for analysis
1000-1040	Handling raw text, using bag of words, TF-IDF and DTM shaping
1040-1100	Morning Break
1100-1200	Sub-setting observations & variables, summarizing data
1200-1300	Laboratory exercise – Getting, format (sub-setting) and store data.
1300-1400	Lunch
1400-1500	Feature creation, combining variables
1500-1540	Understanding binning, discretization and data distribution
1540-1600	Afternoon Break
1600-1700	Laboratory exercise – Preparing extracted data analysis

DAY4	
0900-1000	Introduction to Matplotlib
1000-1040	Exploratory graphs in Python
1040-1100	Morning Break
1100-1200	Visualizing data
1200-1300	Laboratory exercise – Visualizing data
1300-1400	Lunch
1400-1500	Wrangling data and reducing dimensionality (feature selection)
1500-1540	Basic clustering of data and handling outliers
1540-1600	Afternoon Break
1600-1700	Laboratory exercise – Presenting analytics using Python

DAY5	
0900-1000	Regression
1000-1040	Association Rule and Frequent Itemset Mining
1040-1100	Morning Break
1100-1200	Knn, Decision Trees and Random Forest
1200-1300	Support Vector Machines
1300-1400	
1400-1500	Text processing using NLTK
1500-1540	Laboratory exercise – Text classifier
1540-1600	Afternoon Break
1600-1700	Wrap-up