



Python for Data Analysis



Objective:

This course aims to introduce participants to using Python (a general purpose programming language), which is commonly used in data science by companies to gain insights from harvested data for competitive advantage.

Participants will learn how to prepare data for analysis, perform simple statistical analysis, create meaningful data visualizations, present the data and predict trends, patterns and correlations from the data as well as using key Python libraries and data science tools used for data analysis.

Who should Attend?

This course is targeted at an individual and those who want to learn data analysis using Python, such as students, working professionals and PMETs.

Equipment:

Require 01 x Windows-based desktop/notebook computer with an IDLE Python shell installed.

Pre-requisite:

An understanding of Python basic. Participants who are unfamiliar or have no programming experience must take up basic foundation Course : **"Python Programming"** prior to attempting this course as basic fundamentals will not be cover again here.

Course Outline

Introduction

- What is Data Analysis
- Why is it important
- Different sources and type of Data
- Familiar with Python libraries used

Getting Data into and out of Python

- Introduction to Pandas and NumPy
- Type of file format (CSV, Excel, SQL)
- Loading and saving difference file format.

Data Cleaning

- Removal of unwanted data such as duplicate and erroneous ones.
- Standardize the dataset
- Organizing the data

Finding meanings from Data

- Computing Aggregate Statistics
- Sorting data
- Perform Correlation and Regression
- Basic Pivot Table

Data visualization

- Generate data quality report
- Graph a Dataset (Line, Bar, Pie etc)

Programming Exercises & Assignment

- Analysis Exercises
- Mini Project