Pragati Software Pvt. Ltd.

QUALITY, AGILITY, ADAPTABILITY

Corporate IT Training House

Course Title: Python Essentials for Data Science	Duration : 2 days

Python is one of the most popular language for Data Science. Its vast collection of libraries, tools and utilities make it the ideal language for aspiring Data Scientists. Libraries such as Pandas make working with data in python easy and engaging. Dealing with complex statistical analysis requirements is possible using relatively simple python scripts. The python community and support of major corporations in development of python platform ensures its long-term viability for data enthusiasts

This is a 2-day program aimed at developers, database administrators and data modelers who want to step into the role of a data scientist or data analyst by using python as the platform of choice. This program is an ideal blend of hands-on training experience along with required theoretical fundamentals

Course Objectives

- Profound understanding of fundamentals of Python
- Detailed understanding of libraries in python used for Data analysis
- Hands on experience of Data Analysis, pre-processing, and statistical operations using python
- Understanding Data Visualization with various types of graphs, modifying aesthetics, parameters of plots, layouts, etc

Audience

- Familiarity with programming/ scripting terminologies
- Basic understanding of command line/ terminal to execute scripts
- Basic idea of file formats for storage of data in files

Pragati Software Pvt. Ltd.

QUALITY, AGILITY, ADAPTABILITY

Corporate IT Training House

Course Outline

Module 1: Getting Started with Python Scripting

- Introduction to Python
- Setting up a Python environment (IDE, editors, environment variables, package manager , etc)
- writing a *"Hello World"* script in Python
- Input/Output functions (print,input, sep,end arguments for print , etc)
- Introduction to basic data types in python
- if, if-else, if-elif, else blocks
- Loops – *for loop* and *while loop,* variations of these loops and how to implement them
- Understanding expressions and statements

Module 2: Strings, List and Dictionaries

- Introduction to strings and string functions
- Use of split,find,replace and other functions for strings
- Utilizing *chr()* and *ord()* functions
- *raw* strings, *byte* strings, *f-strings* in python with multiple examples
- Understanding *unicode* representation and dealing with such strings in python
- Lists in Python and their properties
- List functions(*copy,append,remove,extend,*etc) and operations using lists
- List comprehension and how to write comprehension expressions
- Dictionaries in Python and their characteristics
- Keys, values and nested dictionaries
- Dictionary functions (*get,fromkeys,popitem,*etc)

Module 3: Functions and Lambda expressions

- Functions in python
- Nested functions and syntax for definition and calling such nested functions
- Lambda functions in python
- Lambda as key for sorting

Module 4: Data analysis, manipulation and visualization in python

- Getting started with Pandas library

- Reading Data from various sources(CSV, Excel, Databases, etc)

- Creating data frames from python objects

- Series and Data Frame basics

- Extracting rows, columns from data frames

- Functions to obtain results of common statistical operations(mean, standard deviation , etc)

- Applying conditions to extract matching rows and columns from data frames

Module 5: Advanced Operations in Pandas

- Using Yahoo finance API for real time data analysis

- Setting of Index to desired attributes and dealing with date time objects

- Re-sampling of Data for appropriate analysis

- Generating Date range and time series analysis

- Handling missing data using *drop* and *fillna* methods

- Pandas merge function to emulate database join queries

- Concatenation of data frames along various axis

Module 6: Numpy library in Python

- Introduction to features of *numpy*

- Creating multi-dimensional arrays using *numpy*

- Indexing and slicing of *numpy* arrays

- Important functions in *numpy* (Sorting, arithmetic operations, Broadcasting ,etc)

- Comparisons, masks, boolean logics in *numpy*

Module 7: Data Visualization in Python

- Introduction to Data visualization using *Matplotlib* and *Seaborn* modules in python

- Plotting various types of graphs including :

  - Line Plots

  - Scatterplots

  - Violin Plots

  - Box plots

  - Swarm Plots

- Using *Seaborn* to change template styles and contexts for plots

- Addition features of *Seaborn* – x and y axis ticks, orientation , color palette , etc

- Adjusting figure size, width, height, legend , etc in graphs

- Using facet grids to plot multiple graphs in grids