



Ref. No. _____

Dated _____

DATA SCIENCE WITH PYTHON

Python Essentials

An understanding of how to use the Python standard library to write programs, access various tools, and document and automate analytical processes.

- Types (strings, lists, dictionaries, and more)
- Control Flow (if-then statements, looping)
- Organizing code (functions, modules, packages)
- Reading and writing files
- Overview of Object-Oriented Programming (OOP)

NumPy and 2D Plotting

Introduction to NumPy and 2D plotting. The NumPy package is presented as a tool for rapidly manipulating and processing large data sets. 2D plotting is introduced with matplotlib.

- Understanding the N-dimensional data structure
- Creating arrays
- Indexing arrays by slicing or more generally with indices or masks
- Basic operations and manipulations on N-dimensional arrays
- Plotting with matplotlib

Pandas: Python's Workhorse Toolkit for All Things Data Analysis

Built on top of NumPy arrays, the Python Data Analysis Library (Pandas) is a powerful and convenient package for dealing with tabular datasets. Participants will learn about its powerful data aggregation and reorganization capabilities for data set explorations, including support for labeling data along each dimension, dealing with missing values, and time series manipulations.

An expert instructor will support students as they work through a typical real-world data analysis project step-by-step using Pandas. This course develops the deep knowledge and skills that will enable students to tackle their own projects with Pandas immediately when they get back to work on Monday morning.

Accessing Data From Multiple Sources

- Reading and writing data from local files (.txt, .csv, .xls, .json, etc)
- Reading data from remote files



Ref. No. _____

Dated _____

- Scraping tables from web pages (.html)
- Making the most of the powerful read_table method

Cleaning and Preparing Data

- Working with Pandas data structures: Series and DataFrames
- Accessing your data: indexing, slicing, fancy indexing, boolean indexing
- Data wrangling, including dealing with dates and times and missing datas
- Adding, dropping, selecting, creating, and combining rows and columns

Database Access and Data Wrangling

- Database access with DB-API2 and SQLAlchemy
- Executing SQL commands from Pandas
- Loading database data into a DataFrame
- Combining and manipulating DataFrames: merge, join, concatenate

Data Visualization

- Understanding the structure of a Figure
- Data visualization: scatter plots, line plots, box plots, bar charts, and histograms with matplotlib
- Customizing plots: important attributes and arguments

Data Analysis

- Split-apply-combine with DataFrames
- Data summarization and aggregation methods
- Pandas powerful groupby method
- Reshaping, pivoting, and transforming your data
- Simple and rolling statistics

Real-World Modeling and Problem Solving

- Deep learning of the data analysis tools through lectures, Q&A, and hands-on exercises
- Develop transferable skills through application to authentic data sets
- Predict the future with time series analysis
- And more!