Course Name: Python for Data Science

Credits: 4

Objectives:

The objective of this course is to enable students to effectively use Python for performing key data science tasks. With an emphasis on Python's essential data structures and powerful libraries such as Pandas, NumPy, and Matplotlib, students will gain practical skills in data preparation, cleaning, exploratory data analysis, and visualization. The course aims to equip learners with the ability to derive insights from data and support decision-making through predictive and pattern discovery techniques.

Prerequisite: Programming concepts, Statistical and numerical methods

Contents:

Module 1: Getting started with Data Science and Python (5 Hrs)

Discovering the match between data science and python, Setting Up Python for Data Science, Understanding the tools: Using the Jupyter Console, Interacting with screen text, Getting Python help, Getting IPython help, Using magic functions, Discovering objects, Using Jupyter Notebook, Working with styles, Restarting the kernel, Restoring a checkpoint, Performing Multimedia and Graphic Integration: Embedding plots and other images, Loading examples from online sites, Obtaining online graphics and multimedia

Module 2: Numerical Computation using NumPy (11 Hrs)

Installing and Understanding Numpy, Understanding Data Types in Python: Fixed-Type Arrays in Python, Creating Arrays from Python Lists, Creating Arrays from Scratch, NumPy Standard Data Types, The Basics of NumPy Arrays: NumPy Array Attributes, Array Indexing: Accessing Single Elements, Array Slicing: Accessing Subarrays, Reshaping of Arrays, Array Concatenation and Splitting, Computation on NumPy Arrays: Universal Functions: The Slowness of Loops, Introducing UFuncs, Exploring NumPy's UFuncs, Aggregations: Min, Max, and Everything in Between, Summing the Values in an Array, Computation on Arrays: Broadcasting, Introducing Broadcasting: Rules of Broadcasting, Broadcasting in Practice, Fancy Indexing: Exploring Fancy Indexing, Combined Indexing, Modifying Values with Fancy Indexing, Sorting Arrays: Fast Sorting in NumPy: np.sort and np.argsort

Module 3: Data Manipulation using Pandas (11 Hrs)

Installing and Using Pandas, Introducing Pandas Objects: The Pandas Series Object, The Pandas DataFrame Object, The Pandas Index Object, Data Indexing and Selection: Data Selection in Series, Data Selection in DataFrame, Operating on Data in Pandas: Ufuncs: Index Preservation, UFuncs: Index Alignment, Ufuncs: Operations Between DataFrame and Series, Handling Missing Data, Trade-Offs in Missing Data Conventions: Missing Data in Pandas, Operating on Null Values, Hierarchical Indexing: A Multiply Indexed Series, Methods of MultiIndex Creation, Indexing and Slicing a MultiIndex, Data Aggregations on Multi-Indices, Combining Datasets: Concat and Append, Simple Concatenation with pd.concat, Aggregation and Grouping: Planets Data, Simple Aggregation in Pandas, GroupBy: Split, Apply, Combine, Pivot Tables: Motivating Pivot Tables, Pivot Tables by Hand, Pivot Table Syntax, Working with Time Series: Dates and Times in Python, Pandas Time Series: Indexing by Time

Module 4: Data Visualization using Matplotlib and Seaborn (13 Hrs)

Installing and using Matplotlib, Figures and Subplots, Colors, Markers, and Line Styles, Ticks, Labels, and Legends, Annotations and Drawing on a Subplot, Saving Plots to File, Exploring Matplotlib Plots: Line Plot, Bar Plot, Stacked Plot, Histogram (1-Dimensional and 2-Dimensional), Box Plot, Area and Scatter Plot, Hex and Pie Plot, Scatter Matrix, Bar Chart on Polar Axis, Line Plot on a Polar Axis, Scatter Plot on a Polar Axis, Animation Plot, Visualization with Seaborn: Installing and using Seaborn, Seaborn Versus Matplotlib, Setting the Plot Styles, Choosing an Appropriate Color Palette, Setting the Plot Themes, Exploring Seaborn Plots: Categorical Seaborn Plots, Scatter Plot, Line Plot, Displot (Histogram, KDE, ECDF and Rug Plots), Two Dimensional Displots, Catplot - Introduction, Strip Plot and Swarm Plot, Box Plot and Violin Plot, Bar Plot and Point Plot, Joint Plot (Scatter + Histogram), Pair Plot (Multiple Scatter + Histogram Plots), Regression or Linear Model Plot.

Reference Books:

1. Python for data science for dummies 2nd Edition, John Paul Mueller, Luca Massaron, Wiley

2. Programming through Python, M. T. Savaliya, R. K. Maurya, G. M. Magar, STAREDU

Solutions

3. Pandas for everyone :Python Data Analysis, Daniel Y. Chen, Pearson

4. Introducing Data Science: Big Data, Machine Learning, and More, Using Python Tools, Davy Cielen, Arno D.B. Meysman, et al., Minning

5. Applied Data Science with Python and Jupyter: Use powerful industry-standard tools to unlock new, actionable insights from your data, Packt

6. Data Analytics, Anil Maheshwari , McGrawHill

7. Data Science From Scratch: First Principles with Python, Joel Grus , SPD