Python for Data Science (PDS)

Course Description:

Python is an easy to learn, powerful programming language. You can use Python when your data analysis tasks need to be integrated with web apps or if statistics code needs to be incorporated into a production database. Being a full-fledged programming language, it's a great tool to implement algorithms for production use.

Learning Outcomes:

After completing this course, you will be able to:

- Write your first python program
- Navigate your way through a python environment
- Perform basic operations related to data analysis
- Work with loops and functions
- Work with Numpy arrays and Pandas DataFrames
- Visualize data using various plots from seaborn and matplotlib
- Understand the distribution and spread of data
- Check for anomalies in data

Pedagogy:

The course is a mixture of classroom lectures, quizzes, assignments, and mini projects. Jupyter notebook will be the medium of coding in python.

Session 1: Introduction to Python

- Python Basics
  - Introduction to Jupyter Notebook
  - Comments
  - Data Types
  - Basic Type casting
  - Print()
  - input()
  - Arithmetic Operators
  - Relational Operators
  - Logical Operators
  - Membership operators
  - Assignment operators
  - Slicing
  - Python Strings - methods and functions
  - Pseudocode and flow chart
  - Python Flow control

- Python Containers
  - Container Data Type
  - List - Methods and Functions
  - Tuple - Methods and Functions
  - Dictionary - Methods and Functions
  - Sets - Methods and Functions

Session 2: Conditional statements and loops

- Conditional Statements (If, If Else, Elif)
- Loops (For loop, While loop, List Comprehension)
- Break, Continue

Session 3: Python Functions

- User Defined Functions
- Variable Scope
- Lambda Function
- Recursive functions

Session 4: Numpy

- Introduction to Numpy
- Array
  - Array creation with random numbers
  - Multidimensional arrays with Numpy
  - Array attributes
  - Array methods
  - Array indexing and slicing
  - Comparison and arithmetic operators for arrays
- Numeric functions
- Stacking and concatenation of arrays
- Array splits
- Iterating through array

Session 5: Pandas Series and DataFrame

- Introduction to Pandas
- Components of Pandas
- Pandas series - creation, accessing
- Filtering, Sorting and ranking
- Null value analysis
- Arithmetic operations
- Pandas DataFrame
  - Creation
  - Reading data from different sources
  - Understanding data - shape, info(), head()
  - Accessing the DataFrame elements
- Sorting and ranking

Session 6 : Advanced Functions with pandas

- Concatenating, series and DataFrames
- Joining and merging DataFrames

- Stack and unstack
- Cross tables and Pivot tables
- Managing Duplicate records
- Drop()
- Map() and replace()
- Groupby()
- Statistical operations

Session 7: Visualization

- Matplotlib visualization
  - line plot,
  - scatter plot,
  - bar plot,
  - pie chart,
  - histogram,
  - Box plot
- Seaborn visualization
  - strip plot,
  - violin plot,
  - swarm plot, pair plot, distribution plot,
  - count plot, regression plot, point plot, joint plot
  - Heat map
- Plotly Visualization (optional/self-read, not part of assessment)