

**Course 7C: Python for Data Science**  
(Skill Enhancement Course (Elective), Credits: 05)

**I. Learning Outcomes:** Students after successful completion of the course will be able to:

1. Identify the need for data science and solve basic problems using Python built-in data types and their methods.
2. Design an application with user-defined modules and packages using OOP concept
3. Employ efficient storage and data operations using NumPy arrays.
4. Apply powerful data manipulations using Pandas.
5. Do data pre-processing and visualization using Pandas

**II. Syllabus:** (Total Hours: 90 including Teaching, Lab, Field training, Unit tests etc.)

**Unit - I (10 hours)**

Introduction to Data Science - Why Python? - Essential Python libraries - Python Introduction- Features, Identifiers, Reserved words, Indentation, Comments, Built-in Data types and their Methods: Strings, List, Tuples, Dictionary, Set - Type Conversion- Operators. Decision Making- Looping- Loop Control statement- Math and Random number functions. User defined functions - function arguments & its types.

**UNIT –II (10 hours)**

User defined Modules and Packages in Python- Files: File manipulations, File and Directory related methods - Python Exception Handling.  
OOps Concepts -Class and Objects, Constructors – Data hiding- Data Abstraction- Inheritance.

**UNIT –III (10 hours)**

NumPy Basics: Arrays and Vectorized Computation- The NumPy ndarray- Creating ndarrays- Data Types for ndarrays- Arithmetic with NumPy Arrays- Basic Indexing and Slicing - Boolean Indexing-Transposing Arrays and Swapping Axes.  
Universal Functions: Fast Element-Wise Array Functions- Mathematical and Statistical Methods-Sorting- Unique and Other Set Logic.

**UNIT –IV (10 hours)**

Introduction to pandas Data Structures: Series, Data Frame and Essential Functionality: Dropping Entries- Indexing, Selection, and Filtering- Function Application and Mapping- Sorting and Ranking.  
Summarizing and Computing Descriptive Statistics- Unique Values, Value Counts, and Membership. Reading and Writing Data in Text Format

**UNIT –V (10 hours)**

Data Cleaning and Preparation: Handling Missing Data - Data Transformation: Removing Duplicates, Transforming Data Using a Function or Mapping, Replacing Values, Detecting and Filtering Outliers- String Manipulation: Vectorized String Functions in pandas.  
Plotting with pandas: Line Plots, Bar Plots, Histograms and Density Plots, Scatter or Point Plots.

### III. References

1. Y. Daniel Liang, "Introduction to Programming using Python", Pearson, 2012.
2. Wes McKinney, "Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython", O'Reilly, 2nd Edition, 2018.
3. Jake VanderPlas, "Python Data Science Handbook: Essential Tools for Working with Data", O'Reilly, 2017.
4. Wesley J. Chun, "Core Python Programming", Prentice Hall, 2006.
5. Mark Lutz, "Learning Python", O'Reilly, 4th Edition, 2009.
6. Web resources:
  - a. <https://www.edx.org/course/python-basics-for-data-science>
  - b. <https://www.edx.org/course/analyzing-data-with-python>
  - c. <https://www.coursera.org/learn/python-plotting?specialization=data-science-python>
  - d. <https://www.programmer-books.com/introducing-data-science-pdf/>
  - e. <https://www.cs.uky.edu/~keen/115/Haltermanpythonbook.pdf>
7. Other web sources suggested by the teacher concerned and the college librarian including reading material.

### IV. Co-Curricular Activities:

**a) Mandatory:** (*Training of students by teacher in field related skills: (lab:10 + field: 05):*)

1. **For Teacher:** Field related training of students by the teacher in laboratory/field for not less than 15 hours on collecting the data, analyzing the data and presenting the data using Python language with some real time data.
2. **For Student:** Students shall (individually) visit any of the agencies like Agriculture dept, statistical cell, irrigation department, Ground water department, CPO office, Rural Water Supply and Sanitation department etc or search online to get real time data like Aids database, weather forecasting database, social networking data, etc and identify any one database, implement and present the necessary charts in Python language and submit a handwritten Fieldwork/Project work/Project work/Project work/Project work Report not exceeding 10 pages. Example: Identifying a database, get the data, present the data in required charts and to predict the future instances if possible.
3. Max marks for Fieldwork/Project work/Project work/Project work/Project work Report: 05.
4. Suggested Format for Fieldwork/Project work/Project work/Project work/Project work: *Title page, student details, index page, and details of place visited, observations, method of data collection, database identified, and implementation in Python language, other findings and acknowledgements.*
5. Unit tests (IE).

#### **b) Suggested Co-Curricular Activities**

2. Training of students by related industrial experts.
3. Assignments
4. Seminars, Group discussions, Quiz, Debates etc. (on related topics).
5. Presentation by students on the topics within and outside the syllabus.