ST.JOSEPH’S COLLEGE FOR WOMEN (AUTONOMOUS), VISAKHAPATNAM

V SEMESTER **COMPUTER SCIENCE**  Time: 3Hrs/Week

CS-E3-5604(3) **PYTHON FOR DATA SCIENCE** Max.Marks:100

w.e.f. 20AH Batch **SYLLABUS**

**COURSE OBJECTIVES: -**

* To learn various python concepts, functions and packages
* To enable data manipulation and analysis.
* To learn how to solve real-world problems using python.

**COURSE OUTCOMES:**

Students after successful completion of the course will be able to:

* Learning the need for data science using python.
* Solve basic problems using Python built-in data types and their methods.
* To demonstrate an application with user-defined modules and packages using OOP concept
* Implement data operations using NumPy arrays for easy storage.
* Applying data manipulation methods using Pandas.
* Implement data pre-processing and visualization of applications using Pandas

**UNIT-I: 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: 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 :NumPy Basics:** Arrays and Vectorized Computation- The NumPyndarray- 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.

**P.T.O.**

**CS-E3-5604(3) ::2::**

**UNIT–IV: 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: 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.

**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.

**\*\* \*\* \*\***