SEMESTER-III

COURSE 5: PYTHON PROGRAMMING FOR DATA ANALYSIS

Theory Credits: 3 3 hrs/week

Aim and objectives of Course:

- To be able to Program in Python
- To know and understand the data Analysis phases
- To know the usage of all libraries

Learning outcomes of Course:

- Understands and learn all basic concepts of
- Python Program Data Analysis methods in Python
- Get used with Python Programming environments

UNIT I:

What is Data Analysis? Differences between Data Analysis and Analytics, What is Python, Why Python for Data Analysis? What is Library, Essential Python Libraries. Python Language basics, I Python and Jupyter Notebook. Python Language Basics.

UNIT II:

Built-in Data Structures, Functions, Files and Operating System. **NumPy Basics:** Arrays and Vectorized Computation, The Numpy ndarray, Universal Functions, Array-Oriented Programming with Arrays, File Input and Output with Arrays, Linear Algebra, Pseudorandom Number Generation.

UNIT III:

Getting Started with Pandas: Introduction to Pandas Data Structures, Essential Functionality, Summarizing and Computing Descriptive Statistics

Data Loading, Storage and File Formats: Reading and Writing Data in TextFormat, Binary Data Formats, Interacting with Web APIs, Interacting with Databases.

UNIT IV:

Data Cleaning and Preparation: Handling Missing Data, Data Transformation,String Manipulation.

Data Wrangling: Join, Combine and Reshape: Hierarchical Indexing, Combiningand Merging Datasets, Reshaping and Pivoting.

UNIT V:

Introduction to Modeling Libraries in Python: Interfacing between pandas andModel code, Creating model descriptions with Patsy, Introduction to stats models.

Plotting and Visualization: A brief matplotlib API Primer, Plotting with Pandas and Seaborn, Other Python visualization tools.