CLUSTER UNIVERSITY SRINAGAR
SYLLABUS (FYUP UNDER NEP - 2020)
Offered by Department of Information Technology
Semester 6th (Major Course – CT2)
Title: Data Science with Python

Course Code: UGICT22J602
Credits: 6 (Theory: 4, Practical: 2)
Contact Hrs: 120 (Theory: 60, Practical: 60)

Max. Marks: 150
Theory External: 80; Min Marks: 32
Theory Internal (Continuous Assessment): 20 Marks, Min Marks: 08
Practical Experimental Basis= 30, Min. Marks: 12
Practical Experimental (Continuous assessment) = 20, Min. Marks: 08

Course Objectives:
Understand the data science life cycle and its significance. Gain proficiency in data handling and manipulation using Python and libraries such as Pandas and NumPy.

Course Outcome:
Students will be able to effectively use Python for data science tasks, including data manipulation with Pandas and NumPy, data exploration and cleaning, and creating visualizations with Matplotlib and Seaborn to derive actionable insights from data.

Unit 1:   **(15 Hrs.)**
Introduction to Data Science: Definition, scope, and applications of data science, Data science workflow: Data collection, cleaning, exploration, modeling, and deployment, data science process (CRISP-DM), Ethical considerations in data science.

Unit 2:   **(15 Hrs.)**
Introduction to Python for Data Science: Basic syntax and data structures, Setting Up Jupyter Notebook, Data Structures:Lists, tuples, dictionaries,Sets and their operations,NumPy arrays,Pandas Series and DataFrames.

Unit 3:   **(15 Hrs.)**
Data Exploration: Reading data from various formats (CSV,Excel, JSON, etc.), Writing data to different formats, Data cleaning techniques: Handling missing values, duplicates, and data types, Summary statistics,

Unit 4:   **(15 Hrs.)**
Creating Informative Data Visualizations: data visualization with Matplotlib and Seaborn, Line plots, scatter plots, bar charts, Heatmaps and pair plots, Exploring Data Distributions and Feature Relationships: Histograms, density plots, Drawing Insights from Visualizations: Identifying trends, outliers, and patterns.

Lab Work (2 Credits: 60 Hours)

1. Write a Python script that demonstrates basic syntax, including variables, data types, and control structures (if statements, loops).
2. Create a program that uses lists, tuples, dictionaries, and sets to store, retrieve, and manipulate data (e.g., a simple contact management system).
3. Write a Python script that demonstrates basic NumPy operations, including creating arrays, performing mathematical operations, and manipulating array dimensions.
4. Create a DataFrame using sample data and perform operations such as filtering, grouping, and aggregating data.
5. Write a Python script that reads data from a CSV file and writes the cleaned data to a new CSV file, demonstrating file handling in Python.
6. Create a Jupyter Notebook that performs data cleaning on a sample dataset by handling missing values, duplicates, and converting data types.
7. Write a program that calculates summary statistics (mean, median, mode, variance, standard deviation) for a given dataset using Pandas.
8. Create a set of visualizations (e.g., histograms, bar charts) using Matplotlib for a sample dataset to illustrate various data distributions.
9. Write a program that generates line plots and scatter plots from a dataset, showcasing trends and relationships between variables.

10. Develop a Jupyter Notebook that creates histograms and density plots to explore the distribution of a continuous variable.

Recommended Books
- ❖ Data Science From Scratch: First Principles with Python, Second Edition (Greyscale Indian Edition) by Jake VanderPlas
- ❖ Data Science and Machine Learning using Python by Dr Reema Thareja
- ❖ Data Science and Analytics with Python by Sandhya Arora and Latesh Malik