CSCI1312	INTRODUCTION TO DATA SCIENCE	L	T	P	S	J	C
		0	0	4	0	0	2
Pre-requisite	Python Programming						
Co-requisite	None						
Preferable Exposure	Python						

Course Description:

"Introduction to Data Science" is a comprehensive course designed for undergraduate students from non-computer science backgrounds. This course provides an introduction to the field of Data Science, covering its significance, applications, and essential concepts. Students will learn Python programming, focusing on functions, variable scope, and data structures using libraries such as NumPy and pandas. The course also covers data acquisition, cleaning, preparation, and visualization techniques using tools like Matplotlib and Seaborn. Through hands-on exercises and practical programming tasks, students will develop the skills necessary to analyze, visualize, and interpret data effectively.

Course Educational Objectives:

- To understanding the Fundamentals of Data Science
- To developing Python Programming Skills for Data Science
- To mastering Data Handling and Preparation Techniques
- To gaining Proficiency in Data Visualization

List of Experiments

S.no	Topic
1	What is Data Science? Introduction to the field, significance, and application areas. Facets of Data Science: Overview of the data science process and key concepts. Python Programming
2	Functions: Defining and calling functions, Parameters and return values, Scope and lifetime of variables
3	Program 1: Basic Function Definition and Call Write a program to create a function named greet that takes a name as a parameter and prints a greeting message.
4	Program 2: Function with Return Value Write a functions called add/subtract/multiply, divide that takes two numbers as parameters and returns their sum/subtract / multiply, divide
5	Program 3: Multiple Return Values Write a program to create a function named stats that takes a list of numbers and returns the minimum, maximum, and average of the list.
6	Program 4: Variable Scope Write a program to demonstrate the difference between local and global variables in a function.
7	Program 5: Lifetime of Variables Write a program to show how the lifetime of a local variable is limited to the function’s execution. Importing a data set using Kaggle /Github /UCI
8	Introduction to NumPy, installing and importing NumPy, Creating and manipulating NumPy arrays, basic array operations (join, split, search, sort),Array attributes, indexing, slicing, and iterating through arrays.
9	Introduction to pandas: Overview of pandas, installing and using the library. Series and Data Frames: Creating and manipulating. Basic Data Frame operations: indexing, re indexing, dropping entries, selecting entries, and summary statistics.
10	NumPy Basics Program 1: Introduction to NumPy Arrays
11	Program 2: Joining and Splitting Arrays Write a program shows how to join and split NumPy arrays.

12	Program 3: Searching and Sorting Arrays Write a program demonstrates how to search for elements and sort arrays in NumPy. pandas Basics
13	Program 4: Introduction to pandas DataFrames Write a program shows how to create pandas DataFrames and perform basic DataFrame operations.
14	Program 5: Handling Missing Data and Summary Statistics Write a program covers data manipulation techniques such as handling missing data and computing summary statistics.
15	Data Handling and Preparation Data Acquisition Techniques
16	Gathering data from various sources: APIs, open data sources, and web scraping (introduction). Data Acquisition Techniques Gathering data from various sources: APIs, open data sources, and web scraping (introduction).
17	Data Acquisition Techniques Program 1: Fetching Data Using an API Write a program demonstrates how to use the requests library to fetch data from a public API.
18	Program 2: Introduction to Web Scraping Write a program provides a basic introduction to web scraping using the Beautiful Soup library.
19	Data wrangling with pandas Cleaning, transforming, merging, and reshaping data. Handling large datasets with efficient techniques and general programming tips. Data Cleaning and Preparation: Dealing with missing data, data transformation techniques, and handling outliers and noise. Data normalization, discretization, and binning for data standardization.
20	Program 1: Merging and Concatenating Data Frames Write a program demonstrates how to merge and concatenate multiple pandas DataFrames.
21	Program 2: Reshaping Data with Pivot Tables Write a program uses pivot tables in pandas to reshape data, which is helpful in data analysis.
22	Data Cleaning and Preparation It may be discussed with domain specific task to be given Program 3: Handling Missing Data Write a program shows various techniques to handle missing data in pandas.
23	Program 4: Data Transformation and Outlier Detection Write a program demonstrates how to normalize data and detect outliers.
24	Program 5: Data Discretization and Binning Write a program demonstrates how to discretize a continuous variable into categorical bins.
25	Program 5: Data Discretization and Binning Write a program demonstrates how to discretize a continuous variable into categorical bins.
26	Basic Plotting with Matplotlib Program 1: Line Plot Write a program demonstrates how to create a simple line plot using Matplotlib.
27	Program 2: Bar Chart Write a program shows how to create a bar chart to compare quantities.
28	Program 3: Whisker-Box Plot Write a program shows how to create a whisker box plot to identify five-number summary and outliers.
29	Program 3: Histogram Write a program demonstrates how to create a histogram to visualize the distribution of a dataset.
30	Program 4: Scatter Plot Write a program illustrates how to create a scatter plot to visualize the relationship between two variables.
31	Advanced Plotting and Introduction to Seaborn Program 5: Subplots and Customization Write a program demonstrates creating multiple subplots and customizing plots.
32	Visualizing Data Creating advanced visualizations to understand data better. Using plots to explore and present data insights. Python for Data Science: Looping and conditional constructs to process and analyse data. Writing functions for repetitive data analysis tasks.

33	Advanced Data Visualizations Program 1: Pair Plot Write a program uses Seaborn to create a pair plot, which is useful for understanding the relationship between multiple variables in a dataset.
34	Program 2: Heatmap Write a program demonstrates how to create a heatmap for visualizing data correlation or any other metrics that can be displayed in a matrix form.
35	g and Functions Program 3: Analyzing Data with Loops Write a program demonstrates how to use loops to process and analyze data stored in a list.
36	Program 4: Writing Functions for Data Analysis Write a program shows how to write a function to compute the median of a dataset, a common data analysis task.
37	Program 5: Conditional Statements to Filter Data Write a program uses conditional statements to filter out specific data points from a dataset.

Textbook(s):

1. Jake VanderPlas,, *Python Data Science Handbook*, 2nd Edition, ed., O'Reilly Media,, 2022.
2. Field Cady, *The Data Science Handbook*, 1st Edition ed., 2017.

Reference(s):

1. Peter Bruce and Andrew Bruce, *Practical Statistics for Data Scientists: 50 Essential Concepts Using R and Python*, 2nd Edition ed., O'Reilly Media, 2020.
2. Lillian Pierson and Jake Porway, *Data Science for Dummies*, 2nd ed., 2017.

Course Outcomes:

1. Comprehension of Data Science Fundamentals
2. Proficiency in Python Programming for Data Science
3. Data Handling and Preparation Skills
4. Data Visualization and Basic Data Analysis

Course Articulation Matrix:

B.A. Economics / B.A. English / B.A. History / B.A. LL.B. (Hons.) / B.A. Mass Communication / B.A. Political Science / B.A. Psychology / B.A. Sociology / B.A. Visual Communication / B.A. with major in Cinematic Arts and Design / B.A. with major in Economics / B.A. with major in English / B.A. with major in History / B.A. with major in Journalism and Media Studies / B.A. with major in Political Science / B.A. with major in Psychology / B.A. with major in Sociology / B.B.A. LL.B. (Hons.) / B.Com. (ACCA) / B.Sc. Biochemistry / B.Sc. Electronics / B.Sc. Mathematics

	POs				PSOs	
CO	**1**	**2**	**3**	**4**	**1**	**2**
1	3	2	1	1		
2	3	2	2	1		
3	3	2	1	1		
4	3	2	1	1		

3 – High, 2 – Medium & 1 – Low Correlation

APPROVED IN MEETINGS HELD ON:

BOS : 08-06-2024 Academic Council Number: 29 Academic Council : 04-07-2024

SDG No(s). & Statement(s) :

4 & Quality Education : Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.

SDG Justification(s):

Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all.