



**SRM Institute of Science and  
Technology College of Engineering and Technology  
School of Computing**

**Set-D**

SRM Nagar, Kattankulathur-603203, Chengalpattu District, Tamil Nadu

**Academic Year: 2024-25 (Even)**

Test: FT1

Date: 25-02-2025

Course Code & Title: 21CSS303T-Data Science

Duration: 50 Minutes

Year & Sem: III Year / VI Sem

Max. Marks: 25

Course Articulation Matrix:

Course Outcome	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO1	-	-	-	-	1	-	-	-	-	-	-	-
CO2	-	-	-	-	1	-	-	-	-	-	-	-

**Note:** CO1 - To understand the relationship between data

CO2 - Identify the different data structures to represent data

**Part- A**

(5x2= 10 Marks)

Answer ALL the questions

Q.No	Question	Marks	BL	CO	PO	PI.Code
1	<p>What are the uses of NumPy?</p> <p>NumPy is used for numerical computations in Python. It provides support for large, multi-dimensional arrays and matrices. It offers mathematical functions for linear algebra, statistical operations, and Fourier transforms. It enhances performance due to its efficient memory usage and vectorized operations</p>	2	1	1	5	5.6.1
2	<p>How do you search for a specific value in a NumPy array?</p> <pre>import numpy as np arr = np.array([10, 20, 30, 40, 50]) index = np.where(arr == 30) print(index) # Output: (array([2]),)</pre> <pre>result = arr[arr == 30] print(result) # Output: [30]</pre>	2	3	1	5	5.4.1
3	<p>Which function is used to join arrays along a specific axis?</p> <pre>import numpy as np a = np.array([[1, 2], [3, 4]]) b = np.array([[5, 6]]) result = np.concatenate((a, b), axis=0) print(result)</pre> <p>And hstack()</p>	2	2	2	5	5.4.1

	vstack()					
4	<p>List out the advantages of web scraping.</p> <p>Automates data collection from websites.</p> <p>Helps in price comparison and market analysis.</p> <p>Enables real-time data updates for applications.</p> <p>Assists in sentiment analysis and business intelligence.</p> <p>Extracts structured data for research purposes.</p>	2	2	2	5	5.6.1
5	<p>How do you sort a Pandas DataFrame based on multiple columns? Explain with an example.</p> <pre>import pandas as pd data = {'Name': ['Alice', 'Bob', 'Charlie', 'David'],         'Age': [25, 30, 35, 40],         'Salary': [50000, 60000, 70000, 80000]} df = pd.DataFrame(data) sorted_df = df.sort_values(by=['Age', 'Salary'],                            ascending=[True, False]) print(sorted_df)</pre> <p>This sorts first by Age in ascending order and then by Salary in descending order.</p>	2	3	2	5	5.4.1

**Part– B**  
(3x5= 15 Marks)

Q.No	Question	Marks	BL	CO	PO	PI.Code
6	<p>Explain the different phases in the Data Science Process. Discuss how each phase contributes to solving a real-world problem.</p> <p>Problem Definition: Identify the objective (e.g., predicting sales).</p> <p>Data Collection: Gather relevant data (e.g., customer transactions).</p> <p>Data Cleaning: Remove inconsistencies and handle missing values.</p> <p>Exploratory Data Analysis (EDA): Identify trends and patterns.</p> <p>Model Building: Train machine learning models.</p> <p>Model Evaluation: Validate accuracy using metrics like RMSE or accuracy score.</p> <p>Deployment &amp; Monitoring: Implement and refine based on real-world feedback.</p> <p>Description of phase to be included</p>	5	2	1	5	5.4.2
7	<p>You are developing a price comparison tool to track the price of a specific product (e.g., "iPhone 15" or "Samsung Galaxy S23") from multiple e-commerce websites such as Amazon, eBay, and Walmart. Explain the key steps involved in performing web scraping for this task, covering aspects such as identifying the target websites, extracting the relevant data, handling dynamic content, and storing the collected information for further analysis.</p> <p>Identify Target Websites: Select Amazon, eBay, Walmart, etc.</p>	5	3	2	5	5.5.1

	<p>Inspect Website Structure: Use browser developer tools to locate price-related HTML elements.</p> <p>Extract Data: Use Python libraries like BeautifulSoup and requests:</p> <pre>import BeautifulSoup import requests url = "https://www.example.com/product" response = requests.get(url) soup = BeautifulSoup(response.text, 'html.parser') price = soup.find('span', {'class': 'price'}).text print(price)</pre> <p>Handle Dynamic Content: Use Selenium if data is loaded via JavaScript.</p> <p>Store Data: Save in CSV, database, or cloud storage for analysis</p>					
9	<pre>df = pd.DataFrame({'ID': [101, 102, 103, 104],                   'Name': ['Alice', 'Bob', 'Charlie', 'David'],                   'Age': [25, 30, 35, 40],                   'Salary': [50000, 60000, 70000, 80000]})</pre> <ul style="list-style-type: none"> <li>• Select the rows where the 'Age' is greater than 30.</li> <li>• Select the 'Name' and 'Salary' columns for the first two rows.</li> <li>• Select all rows except for the last one.</li> </ul> <pre>import pandas as pd df = pd.DataFrame({'ID': [101, 102, 103, 104],                   'Name': ['Alice', 'Bob', 'Charlie', 'David'],                   'Age': [25, 30, 35, 40],                   'Salary': [50000, 60000, 70000, 80000]})</pre> <pre># Select rows where 'Age' &gt; 30 result1 = df[df['Age'] &gt; 30] print(result1)</pre> <pre># Select 'Name' and 'Salary' for the first two rows result2 = df.loc[:1, ['Name', 'Salary']] print(result2)</pre> <pre># Select all rows except the last one result3 = df.iloc[:-1] print(result3)</pre>	5	4	2	5	5.5.1

