|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Register Number** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



**SRM Institute of Science and Technology**

Set -

**College of Engineering and Technology**

**School of Computing**

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

# Academic Year: 2024-25 (EVEN)

Test: FT4 Date: 29-04-2025

Course Code & Title: 21CSS303T-Data Science Duration: Two periods

Year& Sem: III Year /VI Sem Max.Marks:50

Course Articulation Matrix:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Course  Outcome | PO1 | PO2 | PO3 | PO4 | PO5 | PO6 | PO7 | PO8 | PO9 | PO10 | PO11 | PO12 |
| CO3 | - | - | - | - | 1 | - | - | - | - | - | - | - |
| CO4 | - | - | - | - | 1 | - | - | - | - | - | - | - |
| CO5 | - | - | - | - | 1 | - | - | - | - | - | - | - |

**Note:** CO3 – To identify data manipulation and cleaning techniques using pandas

CO4 – To constructs the Graphs and plots to represent the data using python packages

CO5 – To apply the principles of the data science techniques to predict and forecast the outcome of real- world problem

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – A** (10 x 1 = 10 Marks)  *Instructions:*  1) Answer **ALL** questions.  2) The duration for answering Part A is **15 minutes** (this sheet will be collected after 15 minutes).  3**) Encircle the correct answer**. | | | | | | |
| S.No | Question | Marks | BL | CO | PO | PI Code |
| 1 | Which of the following methods is used to remove duplicate rows from a Data Frame in pandas?  a) drop()  b) drop\_duplicates() c) unique()  d) remove\_duplicates() | 1 | 1 | 3 | 5 | 2.1.3 |
| 2 | What function is used to fill missing values in a pandas Data Frame?  a) fillna() b) replace\_null() c) na\_fill() d) fill() | 1 | 1 | 3 | 5 | 2.1.3 |
| 3 | Which of the following is NOT a method for handling missing data?  a) Deletion b) Imputation c) Forward/Backward fill d) Duplicating | 1 | 1 | 3 | 5 | 2.1.3 |
| 4 | When preparing data for modeling, why is scaling important?  a) To hide patterns b) To reduce memory c) To ensure equal importance of features d) To convert text to numbers | 1 | 2 | 3 | 5 | 2.1.3 |
| 5 | Which of these is NOT a standard data cleaning step?  a) Handling missing values b) Removing duplicates c) Building machine learning models d) Correcting data types | 1 | 2 | 3 | 5 | 2.1.3 |
| 6 | Which function is used to set the x-axis label in matplotlib?  a) plt.labelx() b) plt.xlabel()  c) plt.xaxis() d) plt.set\_xlabel() | 1 | 1 | 4 | 5 | 2.1.3 |
| 7 | Which method is used to add a legend to the plot?  a) plt.add\_legend() b) plt.show\_legend() c) plt.legend() d) plt.make\_legend() | 1 | 1 | 4 | 5 | 2.1.3 |
| 8 | Which function is used to display multiple plots in one figure?  a) plt.split() b) plt.multi\_plot() c) plt.subplot()  d) plt.div() | 1 | 1 | 4 | 5 | 2.1.3 |
| 9 | What function is used to add custom text annotations to a plot?  a) plt.comment() b) plt.annotate() c) plt.tag() d) plt.label() | 1 | 2 | 5 | 5 | 2.1.3 |
| 10 | Which function sets the size of the overall figure in matplotlib?  a) plt.resize() b) plt.figure(figsize=(w, h))  c) plt.set\_size() d) plt.figsize() | 1 | 2 | 5 | 5 | 2.1.3 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Register Number** |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |



**SRM Institute of Science and Technology**

Set -

**College of Engineering and Technology**

**School of Computing**

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

# Academic Year: 2024-25 (EVEN SEM)

Test: FT4 Date:29-04-2025

Course Code & Title: 21CSS303T-Data Science Duration: Two periods

Year& Sem: III Year /VI Sem Max.Marks:50

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – B** (4 x 5 = 20 Marks)  Instructions: Answer **ANY FOUR** Questions | | | | | | |
| Q.  No | Question | Marks | BL | CO | PO | PI Code |
| 11 | Discuss the various methods of handling missing data in the dataset. | 5 | 2 | 3 | 5 | 2.1.2 |
| 12 | Explain various data transformation techniques used in data preprocessing. | 5 | 3 | 3 | 5 | 2.1.2 |
| 13 | Write a Python program that accepts a sentence from the user and performs the following string operations:   1. Display the total number of words in the sentence. 2. Convert the entire sentence to title case (first letter capitalized). 3. Find and display the number of times the word 'the' appears (case insensitive). 4. Replace all occurrences of the word 'and' with '&'. | 5 | 2 | 3 | 5 | 2.2.3 |
| 14 | Explain the concept of subplots in Matplotlib with suitable examples. | 5 | 3 | 4 | 5 | 2.2.3 |
| 15 | Define annotations in the context of data visualization using Matplotlib and briefly explain the types of annotations used. | 5 | 3 | 5 | 5 | 2.2.3 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Part – C (2 x 10 = 20 Marks)**  Instructions: Answer ALL questions. | | | | | | |
| Q.  No | Question | Marks | BL | CO | PO | PI  Code |
| 16 a | Discuss the major challenges encountered while working with large datasets and how these challenges impact data preprocessing, storage, and analysis. | 10 | 2 | 3 | 5 | 2.2.3 |
| **(OR)** | | | | | | |
| 16 b | Explain the concept of data wrangling and discuss the key steps involved in the data wrangling process and the importance of each step. | 10 | 3 | 3 | 5 | 3.3.1 |
|  | | | | | | |
| 1. a | 1. Write a Python program to Create a pie chart using Matplotlib showing the percentage distribution of students enrolled in different courses (e.g., Python, Java, C++, AI). 2. Write a Python program to draw a simple line graph using Matplotlib to represent the number of visitors to a website over 7 days. | 10 | 2 | 4 | 5 | 2.2.3 |
| **(OR)** | | | | | | |
| 17 b | 1. Define Seaborn? How does it differ from Matplotlib? Write a Python program to draw a scatter plot using Seaborn showing the relationship between height and weight of individuals. 2. Write a Python program using Seaborn to create a histogram that displays the distribution of students' exam scores. Customize the bin size and add color. | 10 | 3 | 5 | 5 | 3.3.1 |

**Course Outcome (CO) and Bloom’s level (BL) Coverage in Questions:**

|  |  |
| --- | --- |
|  |  |