A-1059

Total Pages : 4 **Roll No. ----------**

CDSA-102

Programming for Data Science

Certificate in Data Science & Applications

1st Semester Examination 2024 (June)

Time: 2:00 hrs Max. Marks: 100

Note : This paper is of Hundred (100) marks divided into Two (02) Section A and B. Attempt the questions contained in these sections according to the detailed instructions given therein. Candidates should limit their answers to the questions on the given answer sheet. No additional (B) answer sheet will be issued.

P.T.O.

Section-A (Long-Answer-Type Questions)

Note : Section 'A' contains Five (05) long-answer-type questions of Twenty Six (26) marks each. Learners are required to answer any Two (02) questions only.

[2x26=52]

Q.1. Describe the process of cleaning and preprocessing data in Python for a machine learning project. [26]

Q.2. a. Discuss the role of list comprehensions in Python and provide an example relevant to data science. [13]

b. Explain the concept of generators in Python and how they can be useful in handling large datasets. [13]

Q.3. a. Describe the concept of a lambda function in Python. Provide an example where a lambda function would be useful. [13]

b. Explain the difference between parameters and arguments in a Python function. [13]

Q.4. a. Explain the purpose of the filter (), select ( ) and mutate ( ) functions in the dplyr package. [13]

b. Describe the differences between a data frame and a matrix in R. [13]

Q.5. a. Explain different strategies for handling missing data in R. [13]

b. How do you install and load an R package? [13]

Section-B (Short-Answer-Type Questions)

Note : Section 'B' contains Eight (08) short-answer-type questions of Twelve (12) marks each. Learners are required to answer any Four (04) questions only. [4x12=48]

Q.1. Explain the various features of python. What are the application areas of python. [12]

Q.2. Explain the importance of data visualization in data science. Provide examples of situations where visualization is crucial. [12]

Q.3. Explain the concepts of mean, median, and mode. When would you use each measure of central tendency. [12]

Q.4. What is a Python dictionary, and how is it different from a list? Explain. [12]

P.T.O.

Q.5. Explain the difference between the if, else if and else statements in R through suitable example. [12]

Q.6. Illustrate the concept of hypothesis testing in R. Name and describe at least five built-in data types in Python. [12]

Q.7. Why is it recommended to use virtual environments in Python projects? Discuss the differences between a module and a package in Python. [12]

Q.8. Describe the basic steps involved in implementing a linear regression model in Python. [12]

****