

P-1159

Total Pages : 4

Roll No.

CDSA-101

Foundations of Data Science and Data Analysis Tools

Certificate in Data Science & Applications (CDSA)

1st Semester Examination, 2023 (June)

Time : 2 Hours]

Max. Marks : 100

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

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.

(2×26=52)

1. What is Data Analysis? List out the differences between data analysis and data analytics. Differentiate between univariate, bivariate, and multivariate analysis. [26]

2. (a) What is the need of data visualization? What are the different techniques? Explain with suitable example. [13]
- (b) Explain general techniques for handling large volumes of data? [13]
3. (a) What's the computational complexity of finding a document's most frequently used words? If you're given 10 TBs of unstructured customer data, how would you go about finding extracting valuable information from it? [13]
- (b) What qualification is required to become a data scientist? Explain various types of jobs and their responsibilities in the field of data scientist. [13]
4. What is python? What are the benefits and scope of python? Explain in detail about arrays and its related concepts in python? [26]
5. (a) Explain at least seven statistical functions of MS-Excel. [13]
- (b) Describe about plotting and visualization concepts in python? [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.

(4×12=48)

1. Define and explain the differences between clustered and non-clustered indexes. [12]
2. What are the data frames? Write its significance in R Language. [12]
3. Explain the common built-in data types of python. [12]
4. How many data formats are available in MS-Excel. Explain them using suitable example. [12]
5. Explain the merits and demerits of Database approach and file based approach. [12]
6. What is the need of an ER-Diagram? What are common symbols used in ER-Diagram? [12]
7. What are the advantages and disadvantages of SPSS to run descriptive statistics? [12]

8. Explain the following SPSS functions: crosstab, compute, descriptive and recode. [12]
-