|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

****

**Presidency University**

**Bengaluru**

**School Of Computer Science and Engineering & Information Science**

**Summer Term End-Term Examinations, Aug 2024**

**Date**: 09-08-2024

**Time**: 9.30 am -12.30 pm

**Max Marks**: 100

**Weightage**: 50%

**Odd Semester**: 2023 - 24

**Course Code**: CSE2027

**Course Name**: Fundamentals of Data Analytics

**Department: CSE**

**Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *Do not write any matter on the question paper other than roll number.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Q.No** | **Questions** | **Marks** | **CO** | **RBT** |
| 1 | 1. **Define**  data and information by giving examples of everyday situations | 4 | CO1 | L1 |
| 1. **Differentiate between** categorical variable and a continuous variable. | 6 | CO1 | L2 |
| c.**Explain** any 5 Vs of Big Data | 10 | CO1 | L3 |
| **OR** | | | | |
| 2 | a.**Name** the different V's of Data. | 4 | CO1 | L1 |
| b. You are given a dataset with the following values: 12, 15, 18, 20, and 22. **Calculate** the mean (average) of this dataset | 6 | CO1 | L2 |
| **c.Describe** nominal, ordinal, and interval scales of measurement in data analysis | 10 | CO1 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | **a.**Describe the advantages and disadvantages of using bar charts and line graphs for representing different data sets. | 4 | CO2 | L1 |
| **b.Compare** the characteristics of questionnaires and schedules used in research. Highlight the key differences between these two data collection methods. | 6 | CO2 | L2 |
| **c.Describe** the different types of Probability sampling techniques | 10 | CO2 | L3 |

**OR**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4 | 1. **Identify** the mean, mode and median of the given sets of data: 7,9,11,13,15,17,6,8, 12, and 4 | 4 | CO2 | L1 |
| 1. **List** the main differences between structured and unstructured data types? | 6 | CO2 | L2 |
| 1. **Solve** the interquartile range formula, calculate the range of the following set of data: {4, 17, 7, 14, 18, 12, 3, 16, 10, 4, 4, 11} | 10 | CO2 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 | 1. **Classify** Quasi Structured Data with Semi Structured Data. | 4 | CO3 | L1 |
| 1. **Demonstrate** the standard error for the sample data: 30, 40, 50, 60, 65. | 6 | CO3 | L2 |
| 1. **Explain** different types of Data Analysis | 10 | CO3 | L3 |

**OR**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 6 | 1. Define Snowball Sampling | 4 | CO3 | L1 |
| 1. Differentiate between Primary Data and Secondary Data | 6 | CO3 | L2 |
| 1. Find the harmonic mean for data 2, 5, 7, and 9 | 10 | CO3 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 7 | 1. Define Schedule | 4 | CO4 | L1 |
| 1. What are the essentials of a good questionnaire | 6 | CO4 | L2 |
| 1. Explain different types of Observation Method | 10 | CO4 | L3 |

**OR**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 8 | 1. **Define** Enumerator also explain merits and demerits of enumerators | 4 | CO4 | L1 |
| 1. **Demonstrate** the standard error for the sample data: 10, 20, 30, 40, 45. | 6 | CO4 | L2 |
| 1. **Explain** Correlation and types of correlation | 10 | CO4 | L3 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 9 | 1. **Explain** different sources of data | 4 | CO1 | L1 |
| 1. **Describe** Non-Probability Sampling | 6 | CO1 | L2 |
| 1. **Interpret** the following scores for students are 40, 45, 49, 53, 61, 65, 71, 79, 85, 91. Calculate the percentile for score 71? | 10 | CO1 | L3 |

**OR**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 10 | 1. Define Data Collection and what are the purpose of data collection | 4 | CO2 | L1 |
| 1. Define questionnaire with its advantages and disadvantages | 6 | CO2 | L2 |
| 1. A machine runs on an average of 125 hours/year. A random sample of 49 machines has an annual average use of 126.9 hours with standard deviation 8.4 hours. Does this suggest to believe that the machines are used on the average more than 125 hours annually at 0.05 level of significance | 10 | CO2 | L3 |