PRESIDENCY UNIVERSITY
BENGALURU

SCHOOL OF ENGINEERING
MID TERM EXAMINATION - APR 2023

Semester : Semester IV - 2021
Course Code : CSE2027
Course Name : Sem IV - CSE2027 - Fundamentals of Data Analytics
Program : CAI,CEI,CST

Date : 13-APR-2023
Time : 2:00PM - 3:30PM
Max Marks : 50
Weightage : 25%

Instructions:
(i) Read all questions carefully and answer accordingly.
(ii) Question paper consists of 3 parts.
(iii) Scientific and non-programmable calculator are permitted.
(iv) Do not write any information on the question paper other than Roll Number.

PART A

ANSWER ALL THE QUESTIONS (5 X 2 = 10M)

1. List any three data analysis types.

    (CO2) [Knowledge]

2. Mention the difference between descriptive and inferential statistics

    (CO2) [Knowledge]

3. Define skewness and types of skewness

    (CO2) [Knowledge]

4. Write any one data transformation technique.

    (CO1) [Knowledge]

5. Give any two examples of human generated unstructured data

    (CO1) [Knowledge]

PART B

ANSWER ALL THE QUESTIONS (4 X 5 = 20M)

6. A company wants to start analysing the online data generated per day. Discuss in detail, the 5 major charateristics the company has to take care about big data.

    (CO1) [Comprehension]

7. A Data Analyst wants to do sampling tofigure out the given data by combining more than one sampling techniques. Discuss in detail the appropriate Sampling Technique applicable.

    (CO1) [Comprehension]

8. Calculate the Harmonic mean for the given set of data.
X = 23, 9, 17, 4

(CO2) [Comprehension]

9. The weekly price in INR of a commodity in the store was observed and is given as : 48.4, 29.5, 17.2, 51.7, 83.1, 37.8. Find the fluctuation in the price of the item.

(CO2) [Comprehension]

PART C

ANSWER ALL THE QUESTIONS   **(2 X 10 = 20M)**

10. The population of all verbal KCET scores are known to have a standard deviation of 8.5. The XY Psychology department hopes to receive applicants with a KCET scores over 210. This year, the mean KCET scores for the 42 applicants was 212.79. Using a value of alpha= 0.05 is this new mean significantly greater than the desired mean of 210?

Z	Area between mean and Z	Area beyond Z
......	....	....
1.62	0.4474	0.0526
1.63	0.4484	0.0516
1.64	0.4495	0.0505
1.65	0.4505	0.0495
1.66	0.4515	0.0485

(CO1) [Application]

11. Suppose a study of sppeeding violations and the drivers who used cell phones resulted in following virtual data:

	Speeding Violations	No Speeding Violations	Total
Cell Phone User	25	280	305
Not a cell phone user	45	405	450
Total	70	685	755

Find:

    a. Probability that a person is a cell phone user
    b. Probability that no violation in the last year AND the person was a cell phone user
    c. Probability that a person is a cell phone user given he had violation last year
    d. Probability that a person had no violation last year given he was not a cell phone user
    e. Probability that a person is a cell phone user OR person had no violations last year

(CO2) [Application]