Roll No. ……..... **Total Pages: _ _**

Paper ID: BSD303 **SET-A** **Course Code:** BSDB32303T

Examination (January - 2024)
Bachelor of Science (B.Sc. – Data Science)
Semester - III

Data Preparation

Time Allowed: 3 Hours **Max. Marks: 70**

Instructions for the Students

1. Attempt any 2 questions out of 4 from Section – A (Each question carries 10 marks)
2. Attempt any 2 questions out of 4 from Section – B (Each question carries 10 marks)
3. Attempt any 10 questions out of 15 from Section – C (Each question carries 03 marks)

Section - A **2*10=20**

Q1. Write a note on sources of data, data collection, data storage and fixing data.
Q2. Differentiate between primary and secondary data. Also explain the process of dealing with outliers and missing values with example.
Q3. Explain the process of data cleaning, data parsing and data segmentation.
Q4. Define cluster analysis. Explain k-means clustering algorithm with example.

Section - B **2*10=20**

Q5. Explain the importance of chi-square, t and f distributions.
Q6. What is sampling? Explain and compare between simple, stratified and systematic sampling.
Q7. Differentiate between null and alternate hypothesis? Explain its use in difference of two means.
Q8. Explain in detail additive and multiplicative models for the analysis of time series.

Section - C **10*3=30**

Q9. Short Answer Questions (Attempt any 10 questions)

a) Data collection
b) Outliers
c) Missing values
d) Real time
e) Consistency checking
f) Data formats
g) Research design
h) Sampling
i) Stratified sampling
j) Variance
k) Heterogeneous data
l) proportions
m)Hypothesis
n) P-value
o) Time series