

JAMIA HAMDARD
Department of Computer Science & Engineering
School of Engineering Sciences & Technology
New Delhi-110062

Examination: Even Semester Examination 2023 (Backlog)

Paper Name: Data Preparation and Analysis
Time: 3 Hours

Paper Code: MTCSE PE 231 (Backlog)
Maximum Marks: 75

Note: Mobile Phones or programmable calculator or any equipment with memory are not allowed.
– Follow proper numbering for answering the questions.

SECTION-A

7*5=35

Section A is compulsory and each question carries 7 marks.

Q.No	Question	CO	BL
1	What are the various file formats used for data storage? Explain the issues associated with them	CO2	2
2	Explain data visualization. What are the methods to visualize the data?	CO5	2
3	Imagine you were building a product to help people buy mobile phone online. How to generate and validate hypothesis for mobile purchase module?	CO4	6
4	What is clustering? Perform k means clustering for following data (1,2), (2,3),(5,4),(3,1) Just do upto two iterations and state the termination condition.	CO4	3
5	For a given data, explain how the following terms: a) Skewness, Left skewness and Right skewness. b) Tailed Distributon c) What does skewness tell us about data? You may use Diagram to explain the above. Write the conditions for both the skewness properly.	CO4	5

SECTION-B (4 x 10 = 40)

Attempt any four questions. Each question carries 10 marks

Q.No	Question	CO	BL
1	The real data in the world is not in the form to be directly processed. Why do we say so? Give valid explanations as per problems faced	CO3	5
2	Explain the various normalization techniques.	CO3	6
3	What is missing data? When and why is data missed? How to handle missing data in real time.	CO2+ CO3	4
4	Explain the concept of primary and secondary data. Explain any three data collection techniques.	CO2	5
5	What are the ethical issues associated with data?	CO6	2
6	Calculate the mean, median, and mode for the following data: 2, 7, 4, 8, 4, 1, 6, 9, 3. Also compare them on any two parameters.	CO4	2