ourdiu.com | Shared by S. M. Mahbub Rahman

Daffodil International University

Faculty of Science & Information Technology
Department of Computer Science & Engineering
Mid Examination, Fall 2025
Course Code: CSE328, Course Title: Introduction to Data Science
Level: 3 Term: 3 Batch: 63

Time: 01:30 Hrs **Marks: 25**

Answer ALL Questions

[The figures in the right margin indicate the full marks and corresponding course outcomes. All portions of each question must be answered sequentially.]

1.	**Scenario 1 :** An organization whose core operations depend on data specifically through the collection, storage, and processing of large-scale datasets has recently implemented a new system. This system now produces an extensive matrix containing millions of observations and thousands of variables that capture detailed aspects of customer behavior and operational performance. The CEO has outlined a broad business objective: "We need to understand our customers better in order to extract value from our data and make our marketing campaigns more effective." However, the data science team is uncertain where to begin and has not yet established a clear direction for the project. **Question: Apply** the concepts of **Data Science** to develop a **step-by-step plan** that the data science team could follow to achieve the CEO's goal.	[5]	**CO1**
2.	Scenario: A large research organization has been assigned to investigate the professional opinions of all IT professionals across an entire nation, defined as the population. However, due to limitations in time, manpower, and financial resources, conducting a full census survey is not practical. The organization must therefore select a representative sample of IT professionals. The lead statistician has proposed three possible sampling approaches: 1. Approach A: The statistician personally selects individuals they believe are most informative. No probability mechanism is used. 2. Approach B: The population is divided into homogenous groups based on years of experience (e.g., 0–5 years, 6–15 years, 16+ years), and samples are drawn separately from each group.	[5]	**CO2**

ourdiu.com | Shared by S. M. Mahbub Rahman

3. Approach C: Samples are selected by following a specific sequence, such as choosing every 5th IT professional listed in the national database.

The statistician need your help to find appropriate method among these.

Question: Analyze the three proposed sampling approaches (A, B, and C) and **Select** the most appropriate method for obtaining a representative sample of IT professionals with proper justification.

[NB. You must name the methods first]

3. [5]

Scenario: A data scientist is analyzing the number of daily support tickets logged for a new software feature over a two-week period (14 days). The daily ticket counts are recorded as follows:

12,15,13,11,14,16,10,14,13,15,12,14,30,11

Question: Examine the daily support ticket data, **Distinguish** any values that deviate from the typical pattern, and infer how these values could affect the interpretation of overall performance metrics.

4. [10] **CO3**

An investor is deciding which of four technology companies, **Innovatech, FutureCorp, TechSolutions,** and **NextGen,** represents the most stable investment. All four firms have achieved the **same average annual profit** over the past five years, but their year-to-year performance varies.

Year	Innovatech	FutureCorp	TechSolutions	NextGen
1	10	15	12	8
2	11	5	14	10
3	10	10	12	12
4	9	15	13	9
5	10	5	12	11

Question: Apply appropriate measures to determine the most stable investment, and **demonstrate** how your analysis can guide the investor's decision.

ourdiu.com