Graphic Era Hill University

Mid Term (Even) Semester Examination May-June 2025

Roll no...............................

Name of the Program and semester: MCA IV
Name of the Course: ***Data Science Using R***
Course Code: TMC-401

Time: 1.5 hour Maximum Marks: 50

Note:

- (i) All the questions are compulsory.
- (ii) Answer any one sub questions from a and b in each main question.
- (iii) Total marks for each question is 10 (ten).

Q1. (1X10=10 Marks)

a. Compare the standard Data-Science Process with the CRISP-DM framework. Discuss two advantages of following a formal framework. CO1

b. Identify three ethical or societal risks that can emerge at different stages of a data-science project (e.g., during data collection or model deployment). CO1

Q2. (1X10=10 Marks)

a. The weekly online-sales figures (₹) for a start-up over eight weeks are: 8 200, 9 100, 9 300, 9 400, 9 500, 9 550, 15 000, 15 200. Calculate the mean, median, mode, range, variance, and coefficient of variation. CO1

b. List and explain methods to handle the missing values and state the importance of this process. CO1

Q3. (1X10=10 Marks)

a. List five distinct data-preparation tasks (e.g., feature scaling) each with an example. CO2

b Explain exploratory data analysis (EDA). Also state the importance of the EDA with respect to Data Science. CO2

Q4. (1X10=10 Marks)

a. Discuss the concept of bias–variance trade-off in model building. How do model complexity and sample size influence this trade-off. CO2

b. Discuss the difference between classification and Clustering and state and explain any one model for each one of them. CO2

Q5. (1X10=10 Marks)

a. Identify elements that help to influence effective data-visualisation design and explain them. CO1

b. Select three built-in R datasets that are especially suitable for :• Descriptive statistics & • Classification modelling. CO2