Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 05/10/2026 **Time:** 02:00 PM - 03:30 PM

Course Code: CSE2500 **Course Name:** Data Analytics

Semester: Fifth Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	80	100

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	A survey records **Gender, Satisfaction Level (Low/Medium/High), Age, and Income**. Identify the appropriate data scale for each variable	2	CO1	BT3
2	A student wants to analyze a file named sales.csv in R. Write the R statement to import the file and store it in a variable called sales.	2	CO1	BT3
3	A student dataset contains missing values in the Age column. Name one R function that can be used to identify missing values.	2	CO2	BT3
4	A dataset contains monthly sales in separate columns. Explain why reshaping may be required before analysis.	2	CO2	BT2
5	Write a `ggplot2` command to create a scatter plot using Age on the x-axis and `Salary` on the y-axis from a data frame named `employee`.	2	CO2	BT3

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	**i.** The rainfall (in mm) recorded over 6 days is: 5, 12, 8, 20, 15, 10. Normalize these rainfall values between 0 and 1 using Min-Max Normalization. (5 marks) **ii. Explain** briefly about any 5 V's of data. (5 marks)	10	CO1	BT2
7	**Explain** Data Analytics and its types.	10	CO1	BT2

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	**Illustrate** the importance of data preprocessing in a data analysis pipeline. Why is this step considered essential?	10	CO1	BT2
9	**Explain** different scales of data and how each level can be applied in real-world research or business contexts.	10	CO1	BT2

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
10	**Explain** how missing data can be identified in R. Illustrate their usage with a suitable example.	10	CO2	BT2

11	Explain in detail about the data manipulation techniques. The following dataset represents the number of patients admitted to a hospital over 10 days: 42, 45, NA, NA, 51, NA, 56, NA, NA, 63 1. Replace all missing values using Forward Fill. 2. Calculate the mean of the completed dataset. 3. Replace all missing values using Backward Fill. 4. Calculate the mean of the backward-filled dataset. 5. Compare the two means and explain the difference.	10	CO2	BT2

PART E: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
12	Given, Tensile strength (MPa) of two types of steel (Mild Steel, High Strength Steel) under three heat treatment conditions (Normal, Quenched, Tempered): Steel Heat Treatment Strength (MPa) Mild Normal 400 Mild Quenched 550 Mild Tempered 500 High Normal 600 High Quenched 750 High Tempered 700 **i. Illustrate** how to create the dataset as a data frame in R. **ii. Demonstrate** how to plot a bar chart showing Steel type versus Strength, grouped by HeatTreatment. **iii. Compare** the strength distribution of Mild Steel and High Strength Steel using a boxplot. **iv. Explain** the role of aesthetics (x, y, fill, group) in ggplot2. **v. State** one engineering insight gained from these plots.	10	CO2	BT2
13	A team collects strain data (microstrain) from 3 sensors on a bridge under two load conditions (Light Load, Heavy Load): Sensor Light Load Heavy Load S1 250 400 S2 220 350 S3 270 420 Tasks (2 marks each): **i. Classify** the numerical strain values by storing them in a matrix (excluding sensor names). **ii. Illustrate** how to create a data frame that includes sensor names along with their strain readings under Light Load and Heavy Load. **iii. Show** how to extend the data frame by adding a new variable Condition = "Bridge Test". **iv. Demonstrate** how to construct a list that contains both (a) the matrix from part (i), and (b) the modified data frame from part (iii). **v.** Using R code, compute the average strain under Heavy Load from the data frame (extract from the list).	10	CO2	BT2