Roll No.												

PRESIDENCY UNIVERSITY

BENGALURU

Mid - Term Examinations – October 2025

Date: 09-10-2025 **Time:** 09.30am to 11.00am

School: SOCSE	**Program:** B.Tech	
Course Code : CSE2500	**Course Name:** DATA ANALYTICS	
Semester: V	**Max Marks**: 50	**Weightage**: 25%

CO - Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**26**	**24**	**-**	**-**	**-**

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

Part A

Answer ALL the Questions. Each question carries 2marks. **5Q x 2M=10M**

1	Briefly explain about Nominal scale with an example.	**2 Marks**	**L1**	**CO1**
2	What are the machine generated unstructured data with examples?	**2 Marks**	**L1**	**CO1**
3	Write a short note on any two packages in R programming?	**2 Marks**	**L6**	**CO1**
4	How sparse matrix is useful in R programming? Justify with code.	**2 Marks**	**L1**	**CO2**
5	Explain briefly about the data manipulation in R.	**2 Marks**	**L2**	**CO2**

Part B

Answer the Questions. **Total Marks 40M**

<table>
<tr><td>6.</td><td>a.</td><td>i. Compressive strength (MPa) of 10 concrete samples: Strength = 30, 32, 35, 30, 34, 32, 36, 30, 35, 34. Find the Mean, Median and Mode. (5 marks)

ii. Distinguish between Data vs Information with an example. (5 marks)</td><td>10 Marks</td><td>L1, L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>Describe briefly about the Data Analytics and its types.</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>7.</td><td>a.</td><td>i. Rainfall (mm) over 6 days: Rainfall = 5, 12, 8, 20, 15, 10. Use Min-Max Normalization to normalize rainfall values between 0 and 1. (5 marks)

ii. Explain briefly about any 5 V’s of data. (5 marks)</td><td>10 Marks</td><td>L1, L2</td><td>CO1</td></tr>
<tr><td></td><td>b.</td><td>Describe the role of data preprocessing in a data analysis pipeline. Why is it considered a critical step?</td><td>10 Marks</td><td>L2</td><td>CO1</td></tr>
</table>

<table>
<tr><td>8.</td><td>a.</td><td>Demonstrate the techniques to handle the missing values in the dataset.</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
<tr><td></td><td>b.</td><td>A team collects strain data (microstrain) from 3 sensors on a bridge under two load conditions (Light Load, Heavy Load):

<table>
<tr><th>Sensor</th><th>Light Load</th><th>Heavy Load</th></tr>
<tr><td>S1</td><td>250</td><td>400</td></tr>
<tr><td>S2</td><td>220</td><td>350</td></tr>
<tr><td>S3</td><td>270</td><td>420</td></tr>
</table>

Tasks (2 marks each):

i. Store the numerical values only in a matrix.

ii. Create a data frame including Sensor names along with Light Load and Heavy Load.

iii. Add a new variable Condition = "Bridge Test" into the data</td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
</table>

<table>
<tr><td></td><td></td><td>frame.

iv. Create a list containing both (i) the matrix and (ii) the modified data frame.

v. Using R code, compute the average strain under Heavy Load from the data frame (extract from the list).</td><td></td><td></td><td></td></tr>
<tr><td colspan="6">Or</td></tr>
<tr><td>9.</td><td>a.</td><td>Illustrate the use of the ggplot2 package in R for data visualization. Explain its layered structure and create two suitable plots (a bar chart and a scatter plot) with example data.</td><td>10 Marks</td><td>L3</td><td>CO2</td></tr>
<tr><td></td><td>b.</td><td>Given, Tensile strength (MPa) of two types of steel (Mild Steel, High Strength Steel) under three heat treatment conditions (Normal, Quenched, Tempered):

<table>
<tr><th>Steel</th><th>Heat Treatment</th><th>Strength (MPa)</th></tr>
<tr><td>Mild</td><td>Normal</td><td>400</td></tr>
<tr><td>Mild</td><td>Quenched</td><td>550</td></tr>
<tr><td>Mild</td><td>Tempered</td><td>500</td></tr>
<tr><td>High</td><td>Normal</td><td>600</td></tr>
<tr><td>High</td><td>Quenched</td><td>750</td></tr>
<tr><td>High</td><td>Tempered</td><td>700</td></tr>
</table>

i. Create the dataset as a data frame in R.

ii. Plot a bar chart showing Steel type vs Strength, grouped by HeatTreatment.

iii. Create a boxplot comparing the strength distribution of Mild Steel and High Strength Steel.

iv. Explain the aesthetics used (x, y, fill, group) in ggplot2.

v. State one engineering insight gained from these plots.</td><td>10 Marks</td><td>L2</td><td>CO2</td></tr>
</table>