Roll No.											

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – December 2025

Date: 19 / 12/ 2025 **Time: 01:00pm – 04:00pm**

School: SoCSE	**Program: CSD, CSG**	
Course Code: CSD2002	**Course Name: Introduction to Data science**	
Semester: V	**Max Marks**: 100	**Weightage**: 50%

CO- Levels	CO1	CO2	CO3	CO4
Marks	26	26	24	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. **10Q x 2M=20M**

1.	Define Data Science. List two of its real-world applications.	2 Marks	L1	CO1
2.	Differentiate between structured and unstructured data with examples.	2 Marks	L2	CO1
3.	List any two Python libraries used for Data Science and mention their purpose.	2 Marks	L1	CO1
4.	Define Data Wrangling? State why is it important?	2 Marks	L1	CO2
5.	Differentiate between Data Cleaning and Data Transformation.	2 Marks	L2	CO2
6.	Describe outliers? Name two methods to detect outliers.	2 Marks	L2	CO2
7.	Describe the purpose of Feature Engineering in Machine Learning?	2 Marks	L2	CO3
8.	Define R-squared in Linear Regression and explain its significance.	2 Marks	L1	CO3
9.	Explain data privacy? Mention one example of how it can be violated in data science applications.	2 Marks	L2	CO4
10.	List two ethical concerns in Data Science applications.	2 Marks	L1	CO4

Part B

Answer the Questions. **Total Marks 80M**

<table>
<tr><td>11.</td><td>You are given the following dataset relating Study Hours (X) and Exam Scores (Y):
<table>
<tr><td>X</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td>Y</td><td>50</td><td>55</td><td>65</td><td>70</td><td>75</td></tr>
</table>
Calculate the slope (m) and intercept (b) using the least squares regression method. Write the regression equation and Predict the exam score if a student studies for 6 hours. Comment on the correlation between study hours and exam scores.</td><td>20 Marks</td><td>L3</td><td>CO3</td></tr>
<tr><td colspan="5">Or</td></tr>
<tr><td>12.</td><td>Using the following confusion matrix from the Titanic dataset:
<table>
<tr><td></td><td>Predicted Survived (1)</td><td>Predicted Not Survived (0)</td></tr>
<tr><td>Actual Survived (1)</td><td>65</td><td>15</td></tr>
<tr><td>Actual Not Survived (0)</td><td>10</td><td>90</td></tr>
</table>
Calculate Accuracy, Precision, Recall, and F1-Score. Interpret what these metrics reveal about the model's performance. Discuss why accuracy alone may not be the best measure for evaluating models.</td><td>20 Marks</td><td>L3</td><td>CO3</td></tr>
</table>

<table>
<tr><td>13.</td><td>Explain the Data Science Lifecycle with suitable diagrams and examples. Explain the applications of Data Science in Healthcare, Finance, and Marketing.</td><td>20 Marks</td><td>L2</td><td>CO1</td></tr>
<tr><td colspan="5">Or</td></tr>
<tr><td>14.</td><td>Define Data Science? Describe its major components. Explain how Data Science impacts business decision-making with case examples.</td><td>20 Marks</td><td>L2</td><td>CO1</td></tr>
</table>

<table>
<tr><td>15.</td><td>Explain how feature engineering and data transformation improve model accuracy. Perform an Exploratory Data Analysis (EDA) on a dataset of your choice. Mention the steps.</td><td>20 Marks</td><td>L2</td><td>CO2</td></tr>
<tr><td colspan="5">Or</td></tr>
<tr><td>16.</td><td>Discuss the difference between Normalization and Standardization with code examples. How outliers are detected using IQR method standard deviation method? Explain with an example.</td><td>20 Marks</td><td>L2</td><td>CO2</td></tr>
</table>

<table>
<tr><td>17.</td><td>Discuss ethical challenges faced by data scientists in real-world projects. Illustrate the applications of Data Science in Government, Education, and Manufacturing sectors.</td><td>20 Marks</td><td>L3</td><td>CO4</td></tr>
<tr><td colspan="5">Or</td></tr>
<tr><td>18.</td><td>Discuss the role of data ethics and privacy in responsible AI and analytics. Determine how Data Science is transforming the healthcare and e-commerce industries.</td><td>20 Marks</td><td>L3</td><td>CO4</td></tr>
</table>