



PRESIDENCY UNIVERSITY

BENGALURU

Roll No.													
----------	--	--	--	--	--	--	--	--	--	--	--	--	--

End - Term Examinations – May / June 2026

Date: 27 -05-2026

Time: 09:30am – 12:30pm

School: SOIS	Program: BCA Data Science	
Course Code: CSA1203	Course Name: Essentials of Data Science	
Semester: II	Max Marks: 100	Weightage: 50%

CO - Levels	C01	C02	C03	C04	C05
Marks	26	24	24	26	-

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.	What is the primary goal of data science?	2 Marks	L1	C01
2.	Compare structured data with semi-structured data.	2 Marks	L1	C01
3.	What is the use of categorical variables?	2 Marks	L1	C01
4.	Find the median of the following data: 12, 7, 3, 9, 15	2 Marks	L1	C02
5.	Find the 25th percentile of the data: 10, 15, 20, 25, 30, 35, 40	2 Marks	L1	C02
6.	Define Linear Regression.	2 Marks	L1	C03
7.	Write the formulae to find Correlation Coefficient.	2 Marks	L1	C03
8.	What is the use of map function?	2 Marks	L1	C03
9.	Define reinforcement learning.	2 Marks	L1	C04
10.	Expand KNN.	2 Marks	L1	C04

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Illustrate different volumes of data in detail.	10 Marks	L3	CO1
	b.	Classify different types of data and variables with examples and neat diagram.	10 Marks	L3	CO1
Or					
12.	a.	Demonstrate key aspects of data mining and data warehousing.	10 Marks	L3	CO1
	b.	Predict Data Science Processes in detail.	10 Marks	L3	CO1

13.	a.	A college recorded the following information: Average monthly attendance percentage of students during an academic year Number of students enrolled in 6 different departments. Percentage distribution of students based on grade categories (Distinction, First Class, Second Class, Pass) Questions: a) Identify suitable charts (Line, Bar, Pie) for representing the above data. b) Explain why each chart is appropriate. c) Discuss how these visualizations assist the principal in academic planning and decision-making.	10 Marks	L2	CO2																				
	b.	From a statistics standpoint, the standard deviation of a dataset is a measure of the magnitude of deviations between the values of the observations contained in the dataset. From a financial standpoint, the standard deviation can help investors quantify how risky an investment is and determine their minimum required return on the investment. Solve the standard deviation for the following 12 weeks data set. <table><tr><td>Weeks</td><td>Expenditure</td></tr><tr><td>1</td><td>\$48.50</td></tr><tr><td>2</td><td>\$87.40</td></tr><tr><td>3</td><td>\$19.98</td></tr><tr><td>4</td><td>\$59.74</td></tr><tr><td>5</td><td>\$40.87</td></tr><tr><td>6</td><td>\$105.51</td></tr><tr><td>7</td><td>\$40.80</td></tr><tr><td>8</td><td>\$23.10</td></tr><tr><td>9</td><td>\$98.10</td></tr></table>	Weeks	Expenditure	1	\$48.50	2	\$87.40	3	\$19.98	4	\$59.74	5	\$40.87	6	\$105.51	7	\$40.80	8	\$23.10	9	\$98.10	10 Marks	L2	CO2
Weeks	Expenditure																								
1	\$48.50																								
2	\$87.40																								
3	\$19.98																								
4	\$59.74																								
5	\$40.87																								
6	\$105.51																								
7	\$40.80																								
8	\$23.10																								
9	\$98.10																								

			10	\$60.54				
			11	\$64.81				
			12	\$48.01				

Or

14.	a.	A consumer goods company has collected the following data for the year 2025: Monthly sales revenue of the company for 12 months. Sales quantity of 4 different products (A, B, C, D) in the month of December. Percentage contribution of each product to the total annual revenue. Advertising expenditure and corresponding monthly sales revenue. The management wants to analyze performance, profitability, and the effectiveness of advertising strategies. a) Select the most appropriate chart for each dataset from the following: b) Justify your choice of chart for each case with proper reasons.	10 Marks	L2	CO2													
	b.	Solve the standard deviation for the following data set. <table><tr><td>X</td></tr><tr><td>53.50</td></tr><tr><td>92.40</td></tr><tr><td>24.90</td></tr><tr><td>64.70</td></tr><tr><td>45.80</td></tr><tr><td>105.00</td></tr><tr><td>40.80</td></tr><tr><td>23.10</td></tr><tr><td>98.10</td></tr><tr><td>60.50</td></tr><tr><td>48.00</td></tr><tr><td>64.80</td></tr></table>	X	53.50	92.40	24.90	64.70	45.80	105.00	40.80	23.10	98.10	60.50	48.00	64.80	10 Marks	L2	CO2
X																		
53.50																		
92.40																		
24.90																		
64.70																		
45.80																		
105.00																		
40.80																		
23.10																		
98.10																		
60.50																		
48.00																		
64.80																		

15.	a.	Solve Using linear regression X = 10,12,14,16,18,20 and Y = 8, 9,10,11,12,13	20 Marks	L3	CO3
Or					
16.	a.	Calculate the correlation coefficient for the following data? X = 5 ,9 ,14, 16,18,22 and Y = 6, 10, 16, 20, 24, 28	20 Marks	L3	CO3

17.	a.	Perform KNN Algorithm on following data. Predict the class for New Entry. [Where K = 3]	20 Marks	L3	CO4																								
		<table><tr><td>BRIGHTNESS</td><td>SATURATION</td><td>CLASS</td></tr><tr><td>20</td><td>35</td><td>?</td></tr></table>				BRIGHTNESS	SATURATION	CLASS	20	35	?																		
		BRIGHTNESS				SATURATION	CLASS																						
		20				35	?																						
		<table><tr><td>BRIGHTNESS</td><td>SATURATION</td><td>CLASS</td></tr><tr><td>40</td><td>20</td><td>Red</td></tr><tr><td>50</td><td>50</td><td>Blue</td></tr><tr><td>60</td><td>90</td><td>Blue</td></tr><tr><td>10</td><td>25</td><td>Red</td></tr><tr><td>70</td><td>70</td><td>Blue</td></tr><tr><td>60</td><td>10</td><td>Red</td></tr><tr><td>25</td><td>80</td><td>Blue</td></tr></table>				BRIGHTNESS	SATURATION	CLASS	40	20	Red	50	50	Blue	60	90	Blue	10	25	Red	70	70	Blue	60	10	Red	25	80	Blue
		BRIGHTNESS				SATURATION	CLASS																						
		40				20	Red																						
		50				50	Blue																						
		60				90	Blue																						
		10				25	Red																						
70	70	Blue																											
60	10	Red																											
25	80	Blue																											
Or																													
18.	a.	Explain KNN Algorithm in detail.	10 Marks	L3	CO4																								
	b.	Compare Supervised learning and Unsupervised learning.	10 Marks	L3	CO4																								