



PRESIDENCY UNIVERSITY

BENGALURU

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

End - Term Examinations – MAY 2025

Date: 31-05-2025

Time: 01:00 pm – 04:00 pm

School: SOIS	Program: BCA	
Course Code: CSA1003	Course Name: Fundamentals of Data Science	
Semester: II	Max Marks: 100	Weightage: 50%

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

10Q x 2M=20M

1.	What is data cleaning, and why is it necessary?	2 Marks	L2	C01
2.	List any two benefits of Data Science.	2 Marks	L2	C01
3.	Differentiate between Data Mining and Data Warehousing	2 Marks	L2	C01
4.	What are the different types of data in statistics?	2 Marks	L2	C02
5.	Define categorical and numerical variables with one example each.	2 Marks	L2	C02
6.	Define correlation, what are types and correlation?	2 Marks	L2	C03
7.	What is regression, what are types of regression?	2 Marks	L2	C03
8.	What is the regression toward mean, give two examples	2 Marks	L2	C03
9.	What is supervised learning, What are its types	2 Marks	L2	C04
10.	Explain Reinforcement learning with an example	2 Marks	L2	C04

Part B

Answer the Questions.

Total Marks 80M

11.	a.	Discuss the process and importance of data cleaning and transformation.	6 Marks	L3	C01
	b.	Discuss the benefits and real-world applications of Data Science	6 Marks	L3	C01
	c.	Explain the Data Science Process with its key steps.	8 Marks	L2	C01
Or					
12.	a.	What are facets of data? Explain their importance in Data Science.	6 Marks	L2	C01
	b.	Differentiate between Data Mining and Data Warehousing with examples.	6 Marks	L3	C01
	c.	Explain the differences between structured, semi-structured, and unstructured data with examples.	8 Marks	L2	C01

13.	a.	Differentiate between qualitative and quantitative data with examples.	6 Marks	L3	C02
	b.	What is the range in statistics? How does it compare to variance and standard deviation in measuring data dispersion?	6 Marks	L2	C02
	c.	Calculate the variance and standard deviation for the dataset: 4, 8, 12, 16, 20.	8 Marks	L3	C02
Or					
14.	a.	Explain the concept of Mean, Median, and Mode, and their significance in data analysis.	6 Marks	L2	C03
	b.	Explain Relative frequency distribution and Cumulative frequency distribution.	6 Marks	L2	C03
	c.	Create Histogram graph for the following Data Distribution in Statistics: dataset: 5, 8, 12, 15, 18, 22, 25, 28, 30, 35, 40, 42, 45, 48, 50, construct a histogram using 5 class intervals. Calculate the class width and frequency for each interval.	8 Marks	L3	C03

15.	a.	Describe applications of correlation coefficient in data science	6 Marks	L2	C03
	b.	Discuss the Interpretation of R-square - Coefficient of Discovery with an example	6 Marks	L3	C04

	c.	<p>Describe the least square regression line, How are slope (m) and intercept(b) are computed.</p> <p>Find the regression equation for the following dataset</p> <table><tr><th>Study hours(X)</th><th>Marks Scored(Y)</th></tr><tr><td>2</td><td>50</td></tr><tr><td>3</td><td>55</td></tr><tr><td>5</td><td>60</td></tr><tr><td>7</td><td>70</td></tr><tr><td>9</td><td>85</td></tr></table>	Study hours(X)	Marks Scored(Y)	2	50	3	55	5	60	7	70	9	85	8 Marks	L3	CO4
Study hours(X)	Marks Scored(Y)																
2	50																
3	55																
5	60																
7	70																
9	85																

Or

16.	a.	Describe Computational Formula, why computational formula is necessary	6 Marks	L2	CO4																		
	b.	Discuss the Standard of Error Estimation(SEE) with an example	6 Marks	L3	CO4																		
	c.	Compute the Coefficient of correlation for the following data set	8 Marks	L3	CO4																		
		<table><tr><td>Student</td><td>Study Hours</td><td>Exam Score</td></tr><tr><td>1</td><td>2</td><td>50</td></tr><tr><td>2</td><td>3</td><td>55</td></tr><tr><td>3</td><td>5</td><td>65</td></tr><tr><td>4</td><td>7</td><td>70</td></tr><tr><td>5</td><td>9</td><td>85</td></tr></table>	Student	Study Hours	Exam Score	1	2	50	2	3	55	3	5	65	4	7	70	5	9	85			
Student	Study Hours	Exam Score																					
1	2	50																					
2	3	55																					
3	5	65																					
4	7	70																					
5	9	85																					

17.	a.	Discuss Reinforcement Learning. Describe the various terms used in Reinforcement learning and their significance	6 Marks	L3	CO4
	b.	Discuss Supervised Machine Learning with an example	6 Marks	L3	CO4
	c.	What is K Nearest Neighbour Algorithm ? Discuss the step by step working of K Nearest Neighbour algorithm .	8 Marks	L3	CO4

Or

18.	a.	Describe Hierarchical Clustering and its types.	6 Marks	L2	CO4
	b.	Explain any Five applications of Machine Learning	6Marks	L2	CO4
	c.	Explain what you mean by K- means clustering..	8 Marks	L3	CO4

		Divide the following data into two clusters using K - means clustering							
		X	1	2	2	3	4	5	
		Y	1	1	3	2	3	5	