Name:.....
Roll No:.....

MAR ATHANASIUS COLLEGE OF ENGINEERING (AUTONOMOUS),

KOTHAMANGALAM - 686666

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

B.TECH FIRST SERIES EXAMINATION, AUGUST 2026 (S5 DS,AIML)

B24CD3T01 DATA ANALYTICS

Max. Marks: 40 Duration: 1.5 Hours

PART A

Answer all questions. Each question carries 3 marks.

1. Given two objects represented by the tuples (22, 1, 42, 10) and (20, 0, 36, 8) Compute the Manhattan distance between the two objects. (CO1)
2. What is the Data Analytics Life Cycle? Why is it important for an analytics project?. (CO1)
3. Consider the dataset: Marks = 45, 56, NULL, 72, 68, NULL, 80. State and explain any three methods used to treat missing values in the dataset. (CO2)
4. The monthly expenditure of families is grouped as follows: Calculate the mean monthly expenditure. (CO2)

Expenditure (rs)	**5,000–10,000**	**10,000–15,000**	**15,000–20,000**	**20,000–25,000**	**25,000–30,000**
Number of Families	8	12	20	15	5

PART B

Answer all questions. Each question carries 14 marks.

5. (a) Data objects are typically described by attributes Illustrate the type of an attributes determined by the set of possible values. (CO1) (7 marks)

   (b) A total of 150 customers of a petrol station are asked about their satisfaction with their car and motorbike insurance. The results are summarized below: Determine Pearson's $\chi 2$ statistic, Cramer's V, (CO1) (7 marks)

	Satisfied	Unsatisfied	Total
Car	33	25	58
Car(Diesel Engine)	29	31	60
Motobike	12	20	32
Total	74	76	150

VISION

Mould competent professionals to serve as valuable resources for society in pursuit of excellence.

MISSION

- Provide quality education in Computer Science to equip with cutting- edge technologies.
- Mould disciplined and Socially Committed engineers capable of professional leadership.
- Promote theoretical and practical understanding, along with critical thinking and logical reasoning.

6. (a) Explain the data cleaning process in data analytics with suitable examples. (CO2) (7 marks)

(b) Apply the following normalization techniques to normalize the temperature value 30 from the given Dataset: 12, 18, 24, 30, 36, 42, 48 (CO2) (7 marks)

i. Min-Max Normalization (range 0–1)

ii. Z-Score Normalization

iii. Decimal Scaling