U.S.N.										

B.M.S. College of Engineering, Bengaluru-560019

Autonomous Institute Affiliated to VTU

June 2025 Semester End Main Examinations

Programme: B.E. **Semester: V**

Branch: Artificial Intelligence and Machine Learning **Duration: 3 hrs.**

Course Code: 24AM5PCDAV **Max Marks: 100**

Course: Data Analysis and Visualization

Instructions: 1. Answer any FIVE full questions, choosing one full question from each unit.
2. Missing data, if any, may be suitably assumed.

Important Note: Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.

			UNIT - I	*CO*	*PO*	Marks
1	a)		Elaborate on different types of data analysis with an example for each.	*CO1*	*PO1*	**10**
	b)		Differentiate structured data from unstructured data with examples for each.	*CO1*	*PO1*	**5**
	c)		Predict the type of data analysis that can be performed for the given scenarios: i. A retail store owner wants to understand customer shopping patterns. ii. A finance team is trying to forecast next quarter's revenue. iii. An e-commerce platform is looking to personalize recommendations for individual users based on their browsing and purchase history. iv. A school wants to understand its student population. v. A restaurant wants to investigate why sales dropped last month.	*CO1*	*PO2*	**5**
			OR			
2	a)		Describe the various data collection methods with examples.	*CO1*	*PO2*	**8**
	b)		Outline the steps in data analysis with a neat diagram.	*CO1*	*PO2*	**6**
	c)		Compare i. Univariate Vs multivariate data. ii. Qualitative Vs Quantitative data.	*CO1*	*PO2*	**6**
			UNIT - II			
3	a)		Consider the given below paragraph and answer the following using string manipulation functions: "The customer Service was Excellent! However, the product did not meet my expectations. I received a wrong item. I had to	*CO1*	*PO4*	**5**

			contact support at support@example.com and feedback@example.org, which was quite disappointing. The response was Slow, and I had to wait for a long time. Overall, I am not satisfied with my experience, and I would not recommend this to my friends! " i. Remove the leading and trailing spaces in the paragraph. ii. Convert all the upper-case letters to lower case. iii. Check for the presence of keywords such as "disappointing" or "satisfied". iv. Extract the email addresses present in the paragraph using regular expressions. v. Determine how many times the word "support" appears in the paragraph?			
	b)	Summarize any two methods for handling missing data with examples.	*CO1*	*PO4*	**5**
	c)	The district headquarters aims to analyze student performance data of various schools from multiple spreadsheets. How data wrangling strategies can be applied to consolidate this information effectively? Justify.	*CO1*	*PO5*	**10**
		OR			
4	a)	Illustrate any 4 data transformation techniques with relevant python snippet.	*CO2*	*PO2*	**10**
	b)	Consider the student data base to answer the following questions: 	Employee ID	Name	Age	Department	Performance Score	Bonus Amount
---	---	---	---	---	---
101	Alice	28	Sales	88	1500
102	Bob	34	Marketing		1200
103	Charlie		HR	75	
104	David	29	Development		
105	Eve	42	Sales	90	2000
106	Frank	38		85	1800
107	Grace	30	HR	70	
108	Henry		Sales		1600 | i. Identify missing entries in the table,how do these gaps affect the overall analysis of student performance. ii. Recommend the possible imputation and interpolation methods could be applied to fill in the missing scores with syntax for each. iii. Determine the ways to drop the instances having missing data. iv. Delete only the instances having two or more missing data.	*CO2*	*PO5*	**10**
		UNIT - III			
5	a)	A healthcare organization wants to visualize patient data to identify trends in disease outbreaks. They have the following data sets: • Daily patient counts across various age groups • The relationship between temperature and reported cases of a particular disease. • Distribution of symptoms reported per week	*CO1*	*PO1*	**6**

			Tasks: I. Suggest appropriate charts for visualizing each dataset and justify your choices. II. Explain how **Gestalt principles** can be used to improve the visual clarity of these charts. III. Propose techniques to reduce **information overload** when displaying these data sets on a single dashboard.			
		b)	Two lines are represented by the following sets of data points: Line 1: (1,2),(2,4),(3,6),(4,8) Line 2: (2,1),(4,2),(6,3),(8,4) A new data point, (5,10), is introduced. Based on the Gestalt principle of continuity, determine which line the new point most likely belongs.	*CO1*	*PO1*	**6**
		c)	Elaborate the process of creating visual representations for data using the Visualization Reference Model. Highlight the significance of visual mapping and visual analytics in this process.	*CO2*	*PO1*	**8**
			OR			
	6	a)	Categorize various types of statistical charts based on the data type. Provide visual examples for each.	*CO2*	*PO1*	**10**
		b)	Infer the importance of Gestalt principles in visual perception and explain how they enhance data visualization. Provide an example for each principle to illustrate its application.	*CO2*	*PO1*	**10**
			UNIT - IV			
	7	a)	Explain the use of trees and groups in data visualization with examples.	*CO1*	*PO1*	**6**
		b)	Outline the advantages of metaphorical visualization over conventional charts or graphs.	*CO1*	*PO1*	**4**
		c)	Evaluate how visualization design choices (e.g., color schemes, chart selection) can create unintentional bias or misinterpretation. Suggest strategies to minimize misleading visualizations.	*CO1*	*PO2*	**10**
			OR			
	8	a)	A social media company wants to analyze user interactions. Explain how visualizations like graphs and clusters can be used to represent relationships (e.g., friends, followers). Provide tools or techniques to achieve this.	*CO2*	*PO1*	**8**
		b)	How can text documents be visualized? Provide examples.	*CO2*	*PO1*	**6**
		c)	A retail company is tracking sales trends. Suggest suitable visualization techniques for: • Sales over time (one-dimensional) • Sales by product category and region (two-dimensional) • Sales influenced by time, category, region, and customer demographics (multi-dimensional)	*CO2*	*PO1*	**6**

			UNIT - V			
	9	a)	What is volumetric data visualization? Explain with examples.	*CO1*	*PO1*	**6**
		b)	Describe the role of GIS systems in geographic visualization.	*CO1*	*PO1*	**7**
		c)	How do you create a collaborative visualization for a vector field dataset and explain its relevance?	*CO1*	*PO1*	**7**
			OR			
	10	a)	Infer the advantages of representing data in vectors with suitable examples.	*CO2*	*PO2*	**6**
		b)	Describe different types of maps. Illustrate how maps can be effectively visualized for data analysis.	*CO2*	*PO2*	**8**
		c)	Outline different types of vector fields with an example for each.	*Co2*	*PO2*	**6**

****