

# 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: Computer Science and Engineering**

**Duration: 3 hrs.**

**Course Code: 22CS5PEDEV**

**Max Marks: 100**

**Course: Data Exploration 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.

		UNIT - I	CO	PO	Marks
<b>Important Note:</b> Completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages. Revealing of identification, appeal to evaluator will be treated as malpractice.	1	a) What are the steps involved in Exploratory Data Analysis (EDA), and why are these stages considered pivotal in data analysis?	CO1	PO1	<b>6</b>
		b) Compare EDA with classical and Bayesian analysis.	CO1	PO1	<b>6</b>
		c) Identify the measurement scale? <ul style="list-style-type: none"> <li>i. What is your favorite color?</li> <li>ii. Which country are you from?</li> <li>iii. Rate your satisfaction with our service on a scale from 1 to 5.</li> <li>iv. Rank the following in order of preference: Action, Comedy, Drama.</li> <li>v. What is the temperature today in Celsius?</li> <li>vi. On a scale of 1 to 10, how happy are you today?</li> <li>vii. How many hours did you study for the exam?</li> <li>viii. What is your annual income?</li> </ul>	CO1	PO1	<b>8</b>
		OR			
	2	a) i. What is categorical data? Explain the key characteristics of categorical data and how it is handled during EDA <ul style="list-style-type: none"> <li>ii. Classify the following data into discrete, continuous, or categorical:               <ul style="list-style-type: none"> <li>• Number of books in a library.</li> <li>• Weight of students in a class.</li> <li>• Types of beverages sold in a café.</li> <li>• Temperature readings in a city</li> </ul> </li> </ul>	CO1,2	PO1,2	<b>10</b>
		b) Discuss the four types of measurement scales (Nominal, Ordinal, Interval, and Ratio) with examples. Highlight the key differences between them			<b>10</b>
		UNIT - II			
	3	a) Define Data transformation. Explain any four transformation activities.	CO1	PO1	<b>10</b>
		b) Describe how to Fill Missing values using Fillna() with example program.	CO1	PO1	<b>10</b>
		OR			
	4	a) You are given the following dataset: ID      Name      Age      City 101     Alice     25      New York	CO2	PO2	<b>6</b>

		<p>102 Bob 30 Chicago      101 Alice 25 New York      103 Charlie 28 Boston      102 Bob 30 Chicago</p> <p>Write a Python script to create a dataframe given and demonstrate how to remove duplicate records while keeping the first occurrence</p>																		
	b)	<p>The following dataset contains missing values:</p> <table> <tr><td>Name</td><td>Age</td><td>City</td></tr> <tr><td>Alice</td><td>25</td><td>New York</td></tr> <tr><td>Bob</td><td>None</td><td>Chicago</td></tr> <tr><td>Charlie</td><td>28</td><td>None</td></tr> </table> <p>Write a Python script to:</p> <ol style="list-style-type: none"> <li>Replace missing numerical values with the mean of the Age column.</li> <li>Replace missing categorical values in City with Unknown</li> </ol>	Name	Age	City	Alice	25	New York	Bob	None	Chicago	Charlie	28	None	CO2	PO2	6			
Name	Age	City																		
Alice	25	New York																		
Bob	None	Chicago																		
Charlie	28	None																		
	c)	Explain Forward and Backward filling with an example	CO1	PO1	8															
		<b>UNIT - III</b>																		
5	a)	Explain three measures of Central Tendency in details.	CO1	PO1	6															
	b)	Demonstrate the quartiles and its points with a suitable example (using panda's libraries).	CO3	PO3	6															
	c)	<p><b>Scenario:</b></p> <ul style="list-style-type: none"> <li>You are a data analyst working for a multinational corporation that manufactures electronic devices. You have a dataset containing information about the sales of different electronic products across various regions over the past year.</li> </ul> <p><b>Dataset:</b></p> <ul style="list-style-type: none"> <li>'Region': ['North', 'North', 'South', 'South', 'North', 'South'],      'Product_Category': ['Laptops', 'Smartphones', 'Laptops',      'Smartphones', 'Laptops', 'Smartphones'], 'Sales': [5000, 7000, 8000,      6000, 9000, 5500], 'Units_Sold': [50, 70, 80, 60, 90, 55]</li> </ul> <p><b>Question:</b></p> <ol style="list-style-type: none"> <li>Create a pivot table to calculate the total sales for each product category in different regions.</li> <li>Create another pivot table to find the average selling price for each product category in different regions (average price per unit sold).</li> </ol>	CO3	PO3	8															
		<b>(OR)</b>																		
6	a)	<p>i. Explain the purpose of the groupby() function in pandas. How does the groupby() operation work?</p> <p>ii. Given the dataset:</p> <table> <tr><td>Department</td><td>Employee</td><td>Sales</td></tr> <tr><td>HR</td><td>Alice</td><td>100</td></tr> <tr><td>HR</td><td>Bob</td><td>150</td></tr> <tr><td>IT</td><td>Charlie</td><td>200</td></tr> <tr><td>IT</td><td>David</td><td>250</td></tr> </table> <p>Write a Python script to calculate the total sales by department</p>	Department	Employee	Sales	HR	Alice	100	HR	Bob	150	IT	Charlie	200	IT	David	250	CO1,2	PO1,2	10
Department	Employee	Sales																		
HR	Alice	100																		
HR	Bob	150																		
IT	Charlie	200																		
IT	David	250																		
	b)	Explain correlation and its type with an example and justify this statement "Correlation does not imply causation"	CO1	PO1	10															
		<b>UNIT - IV</b>																		
7	a)	Outline different types of Multi Variate Analysis Technique.	CO1	PO1	10															

	b)	What is the fundamental purpose of Univariate Analysis and mention the methods used to characterize the patterns observed in univariate data?	CO1	PO1	5
	c)	Demonstrate the importance of Bivariate analysis with an example.	CO3	PO3	5
		<b>OR</b>			
8	a)	Explain the Characteristics of Time series data.	CO1	PO1	10
	b)	Consider a dataset containing monthly sales data for a retail store over a period of two years. The dataset has two columns: "Date" and "Sales." Your task is to perform time series analysis on this data. 1. Load the dataset and convert the "Date" column to the datetime format. 2. Set the "Date" column as the index of the DataFrame. 3. Plot the time series to visualize the sales pattern over the two-year period. 4. Calculate and plot the 7-day rolling mean of sales to identify trends. 5. Determine whether there are any seasonal patterns in the data.	CO3	PO3	10
		<b>UNIT - V</b>			
9	a)	How do you map data on aesthetics? Explain its types of variables with descriptions.	CO1	PO1	10
	b)	Illustrate how pie chart visualizing portions into stacked bars.	CO2	PO2	5
	c)	Describe coordinate system and Aspects Ratio in visualization with an example.	CO1	PO1	5
		<b>OR</b>			
10	a)	How Aesthetic data can be visualized? Explain it with types of data.	CO1	PO1	5
	b)	Illustrate with chart Relationship between Cartesian and polar coordinates.	CO2	PO2	5
	c)	Demonstrate Empirical Cumulative Distribution Functions and Q-Q Plots with chart.	CO2	PO2	10

\*\*\*\*\*