REG.NO.:

VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

NAME OF THE SCHOOL
CONTINUOUS ASSESSMENT TEST - I
WINTER SEMESTER 2024-2025

SLOT: D2

Programme Name & Branch	**: B.Tech – CSE**
Course Code and Course Name	**: BCSE331L and Exploratory Data Analysis**
Faculty Name(s)	**: Prof. Prakash M**
Class Number(s)	**: VL2024250502177**
Date of Examination	**: 30.01.2025**
Exam Duration	**: 90 minutes**

Maximum Marks: 50

General instruction(s):

- Answer All Questions
- M - Max mark; CO – Course Outcome; BL – Blooms Taxonomy Level (1 – Remember, 2 – Understand, 3 – Apply, 4 – Analyse, 5 – Evaluate, 6 – Create)
- Course Outcomes:
  CO1: Handle missing data in the real world data sets by choosing appropriate methods.
  CO2: Summarize the data using basic statistics. Visualize the data using basic graphs and plots.
  CO3: Identify the outliers if any in the data set.

Q. No	Question	M	CO	BL
1.	A fitness company that develops wearable devices to track physical activities. The company wants to improve customer engagement by understanding user activity patterns and recommending personalized fitness plans. The dataset that contains the columns: User_ID, Date, Steps, Calories_Burned, Active_Minutes, Sleep_Hours, Device_Type, Age_Group, Membership_Type. The company wants insights that can improve user activity levels and overall health outcomes. Explain the various steps would be taken in Exploratory Data Analysis.	10	1	2
2.	The following are the ages of 15 customers at a retail store: 30, 72, 20, 80, 35, 65, 50, 62, 60, 55, 40, 70, 45, 75, 25. Partition the data into three bins into a) Equi-depth b) Equal-distance c) Bin boundary and d) Bin median.	10	2	3
3.	Explain the difference between append, concatenate, merge, and join in data transformation. Provide examples of situations where each method would be appropriate using python code.	10	2	3
4.	Consider the following dataset of marks obtained by students in a class: [82, 90, 87, 74, 88, 91, 77, 85, 89, 92, 79, 85, 83, 90, 86] a) Construct a frequency distribution table with class intervals of 5 marks. b) Calculate the measures of central tendency. c) Create a box plot to visualize the spread of the data and identify any outliers.	10	3	4
5.	A company wants to analyze the relationship between advertising spending, social media engagement, product pricing, and total sales to optimize their marketing strategy.	10	3	3

CampaignID	AdSpending	SocialMediaEngagement	ProductPrice	TotalSales
1001	5000	15000	100	7000
1002	7000	20000	120	8000
1003	4000	12000	90	6000
1004	8000	25000	110	8500
1005	6000	18000	95	7500
...	...	...	...	...

Using the python code to

a) load the dataset and calculate statistical summaries.

b) use pairplots to visualize relationships between AdSpending, SocialMediaEngagement, ProductPrice, and TotalSales.and

c) create a heatmap to show correlations among the variables.

****