



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CAT-I Test

Fall Semester 2026-2027

SLOT: A1

Programme Name & Branch : B. Tech BCE/ BDS
Course Code and Course Name : BCSE331L & Exploratory Data Analysis
Faculty Name(s) : Dr. Vijayanand C, Dr. Prakash M, and Dr. Balaji G N
Class Number(s) : VL2026270102443, 2444, and 2446
Date of Examination : 09-Aug-2026 & 9:30 to 11:00 AM
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:
At the end of the course, the student will be able to
 1. Handle missing data in the real world data sets by choosing appropriate methods.
 2. Summarize the data using basic statistics. Visualize the data using basic graphs and plots.
 3. Identify the outliers if any in the data set.

Q. No	Question	M	CO	BL
1.	a) A college conducted a semester examination for first-year students. The examination data has been collected from different departments and stored in a CSV file named student.csv. While reviewing the dataset, the faculty observed that some records are duplicated, some students have missing scores or ranks. As a Data Analyst, the task is to analyze the dataset and prepare it for further machine learning and reporting. (7 marks) b) Compare EDA with Bayesian Analysis with a proper illustration. (3 marks).	10	1	2
2.	From the student.csv dataset 1, identify the datatypes for the attributes Student ID, Age, Subject, Score and Rank with sub-category and justify.	10	1	2
3.	From the student.csv dataset 1 given below, write a python code to perform data deduplication, replace values, handling missing data, filling missing values and display the result.	10	2	3
4.	Consider the following student scores: 52, 87, 71, 95, 63, 78, 84, 56, 91, 68, 74, 60, 89, 82, 97, 45 Using the above dataset, perform the following data discretization techniques by dividing into four bins. a) Equal depth b) Equal width c) Bin medians d) Bin boundary	10	2	3

5	Using the dataset 2 given below. a) Calculate the final Pearson correlation coefficient (r) rounded to two decimal places and write the Python code to perform this calculation. (7 Marks). b) What does this correlation coefficient tell you about the relationship between Book Value and P/E Ratio in this dataset? (3 Marks)	10	2	4
---	---	----	---	---

Dataset 1: student.csv

Student ID	Name	Age	Gender	Subject	Score	Rank
201	Anjali Nair	19	Female	Mathematics	88	5
202	Sreya Menon	20	Female	Science	91	2
202	Sreya Menon	20	Female	Science	91	2
203	Harsha Reddy	19	Male	Mathematics	NaN	?
204	Sai Teja	21	Male	English	69	8
205	Aparna Pillai	20	Female	Science	?	NaN
201	Anjali Nair	19	Female	Mathematics	88	5
206	Vishnu Nair	18	Male	Computer Science	95	1
207	Lakshmi Priya	19	Female	Mathematics	84	6
208	Karthik Reddy	20	Male	Science	NaN	9

Dataset 2:

Company	Book Value (X)	P/E Ratio (Y)
Tata Motors	120	15
ITC	150	12
Karnataka Bank	80	18
Firm A	200	10
Firm B	250	8