



SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

Programme Name & Branch : B. Tech (Specialization with Data Science)
Course Code and Course Name : BCSE207L-Programming for Data Science
Faculty Name(s) : Prof. Prabakaran N, Prof. Deepika J, Prof. Vijayan S
Class Number(s) : VL2025226052539, 2542, 2544
Date of Examination : 18/03/2026
Exam Duration : 90 minutes
Maximum Marks: 50

General instruction(s):

- Answer All Questions
- M-Maximum marks , CO – Course Outcome , BL- Blooms Taxonomy
- Co2 : Design a suitable form for analysis from real-time data.
- Co3 : Formulate insights from the data through statistical inferences.

Q. No	Question	M	CO	BL																																																							
1.	An e-commerce company maintains two datasets related to customer orders and product information. The following sample records are provided. (2.5 x 4 = 10) Orders Table <table><tr><th>Ord_ID</th><th>Cu_ID</th><th>Pr_ID</th><th>Quantity</th><th>OrderDate</th></tr><tr><td>1</td><td>C101</td><td>P1</td><td>2</td><td>2024-01-01</td></tr><tr><td>2</td><td>C102</td><td>P2</td><td>1</td><td>2024-01-02</td></tr><tr><td>3</td><td>C101</td><td>P3</td><td>4</td><td>2024-01-03</td></tr><tr><td>4</td><td>C103</td><td>P1</td><td>2</td><td>2024-01-05</td></tr><tr><td>5</td><td>C101</td><td>P2</td><td>3</td><td>2024-01-06</td></tr><tr><td>6</td><td>C102</td><td>P3</td><td>1</td><td>2024-01-07</td></tr><tr><td>7</td><td>C104</td><td>P4</td><td>2</td><td>2024-01-08</td></tr></table> Product Table <table><tr><th>Pr_ID</th><th>Category</th><th>Price</th></tr><tr><td>P1</td><td>Electronics</td><td>200</td></tr><tr><td>P2</td><td>Clothing</td><td>50</td></tr><tr><td>P3</td><td>Grocery</td><td>30</td></tr><tr><td>P4</td><td>Electronics</td><td>150</td></tr></table> a) Display the both table content b) Combine the two datasets and compute the total order value for each transaction. c) To determine customers who placed more than two orders and display their Cu_ID and Order Count. d) Calculate the total revenue generated for each product category and present the top two categories based on revenue.	Ord_ID	Cu_ID	Pr_ID	Quantity	OrderDate	1	C101	P1	2	2024-01-01	2	C102	P2	1	2024-01-02	3	C101	P3	4	2024-01-03	4	C103	P1	2	2024-01-05	5	C101	P2	3	2024-01-06	6	C102	P3	1	2024-01-07	7	C104	P4	2	2024-01-08	Pr_ID	Category	Price	P1	Electronics	200	P2	Clothing	50	P3	Grocery	30	P4	Electronics	150	10	2	5
Ord_ID	Cu_ID	Pr_ID	Quantity	OrderDate																																																							
1	C101	P1	2	2024-01-01																																																							
2	C102	P2	1	2024-01-02																																																							
3	C101	P3	4	2024-01-03																																																							
4	C103	P1	2	2024-01-05																																																							
5	C101	P2	3	2024-01-06																																																							
6	C102	P3	1	2024-01-07																																																							
7	C104	P4	2	2024-01-08																																																							
Pr_ID	Category	Price																																																									
P1	Electronics	200																																																									
P2	Clothing	50																																																									
P3	Grocery	30																																																									
P4	Electronics	150																																																									
2.	A logistics company records delivery information for shipments processed through different warehouses. (2.5 x 4 = 10) <table><tr><th>ShipmentID</th><th>Warehouse</th><th>DeliveryTime</th><th>DeliveryStatus</th><th>Distance</th></tr><tr><td>S1</td><td>W1</td><td>5</td><td>OnTime</td><td>100</td></tr><tr><td>S2</td><td>W2</td><td>8</td><td>Late</td><td>120</td></tr><tr><td>S3</td><td>W3</td><td>4</td><td>OnTime</td><td>80</td></tr><tr><td>S4</td><td>W4</td><td>9</td><td>Late</td><td>150</td></tr><tr><td>S5</td><td>W5</td><td>6</td><td>OnTime</td><td>110</td></tr><tr><td>S6</td><td>W6</td><td>10</td><td>Late</td><td>160</td></tr></table> a) Find the average delivery time for each warehouse. b) Computer the percentage of late deliveries for each warehouse. c) Identify warehouses whose average delivery time is higher than the overall average delivery time. d) Produce a summary table containing Warehouse, Average Delivery Time, and Late Delivery Percentage	ShipmentID	Warehouse	DeliveryTime	DeliveryStatus	Distance	S1	W1	5	OnTime	100	S2	W2	8	Late	120	S3	W3	4	OnTime	80	S4	W4	9	Late	150	S5	W5	6	OnTime	110	S6	W6	10	Late	160	10	2	5																				
ShipmentID	Warehouse	DeliveryTime	DeliveryStatus	Distance																																																							
S1	W1	5	OnTime	100																																																							
S2	W2	8	Late	120																																																							
S3	W3	4	OnTime	80																																																							
S4	W4	9	Late	150																																																							
S5	W5	6	OnTime	110																																																							
S6	W6	10	Late	160																																																							

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING
CONTINUOUS ASSESSMENT TEST - II
WINTER SEMESTER 2025-2026

3. An educational dataset is provided in table format. (2.5 x 4 = 10)
- | School | Math_2022 | Science_2022 | Math_2023 | Science_2023 |
|--------|-----------|--------------|-----------|--------------|
| S1 | 80 | 75 | 85 | 78 |
| S2 | 70 | 65 | 72 | 68 |
| S3 | 90 | 88 | 92 | 90 |
- A student using the below code attempted to reshape the dataset but produced incorrect results.
- ```
library(tidyr)
long_data <- gather(data, key="Subject", value="Score",
Math_2022, Science_2022)
```
- Identify the issue in the given code and correct the statement so that all subject-year columns are converted to long format.
  - Separate the Subject and Year components into two columns.
  - Convert the transformed dataset back to wide format using suitable tidyr functions.
  - Present the expected output structure after reshaping

10 2 4

4. A weather API returns the following JSON data. (2.5 x 4 = 10)
- ```
{ "city": "Chennai",
  "temperature": 32,
  "humidity": 70,
  "wind_speed": 12,
  "timestamp": "2024-03-01 10:00" }
```
- A student using the below code to retrieve and process this data using R.
- ```
library(httr)
library(jsonlite)
response <- GET("https://api.weather.com/data")
data <- fromJSON(response)
```
- Identify the problem in the code used to parse JSON data.
  - Retrieve the API response and convert it into a structured dataframe.
  - Extract only temperature and humidity values for the last seven observations.
  - Present the expected structure of the final dataframe obtained after processing the API response.

10 3 5

5. A small digital library maintains information about books, authors, and borrowing activities. The following sample records are available. (2.5 x 4 = 10)

**Books**

| Bk_ID | Book Title         | Book Price |
|-------|--------------------|------------|
| B1    | Data Science       | 500        |
| B2    | Machine Learning   | 600        |
| B3    | R Programming      | 450        |
| B4    | AI Basics          | 550        |
| B5    | Data Visualization | 520        |

**Borrowing**

| Bor_ID | Bk_ID | BorrowDate | ReturnDate |
|--------|-------|------------|------------|
| BR1    | B1    | 2024-01-01 | 2024-01-05 |
| BR2    | B2    | 2024-01-02 | 2024-01-06 |
| BR3    | B1    | 2024-01-04 | 2024-01-08 |
| BR4    | B3    | 2024-01-05 | 2024-01-09 |
| BR5    | B1    | 2024-01-07 | 2024-01-10 |
| BR6    | B4    | 2024-01-08 | 2024-01-12 |

- A data analyst plans to explore these tables using R and relational database operations.
- Demonstrate how the above datasets can be stored as tables in a temporary database environment from within R.
  - Retrieve the titles of books that appear most frequently in the borrowing records, along with the number of times they were borrowed.
  - Determine the total revenue generated from all borrowing transactions, assuming the price of each book represents the rental fee per borrowing.
  - Produce a table showing Book Title, Book Price and borrow count.

10 3 5