



Continuous Assessment Test (CAT) – II OCTOBER 2025

Programme	: B.Tech. CSE (Data Science)	Semester	: Fall 2025-2026
Course Code & Course Title	: BCSE207L Programming for Data Science	Class Number	: CH2025260101356
Faculty	: Dr. N.G. Bhuvaneswari Amma	Slot	: A1
Duration	: 90 Minutes	Max. Marks	: 50

General Instructions:

- Write only your registration number on the question paper in the box provided and do not write other information
- Only non-programmable calculator without storage is permitted

Answer all questions

Q. No.	Description	Marks	CO	BT Level
1	A food delivery company keeps daily records of all completed orders in a CSV file named <code>delivery_data.csv</code>. The dataset contains these columns: OrderID, CustomerID, Restaurant, CuisineType, OrderDate, OrderTime, DeliveryTime, Amount, and Rating. The sample data is given as follows: O1000, C002, GreenBowl, Italian, 12-06-2024, 13:45:00, 14:09:00, 1184.48, 4.6 O1001, C015, PizzaWorld, Chinese, 29-01-2024, 11:24:00, 12:01:00, 538.69, 3 O1002, C003, BurgerTown, Chinese, 07-01-2024, 14:44:00, 15:47:00, 710.85, 3.1 O1003, C018, GreenBowl, Italian, 11-03-2024, 17:24:00, 18:01:00, 1034.99, 4.4 O1004, C011, SushiHouse, Italian, 03-03-2024, 13:52:00, 14:14:00, 219.5, 4.8 Perform the following tasks using R Programming: a) Identify restaurants whose daily revenue never drops below 70% of their overall daily average revenue across the entire dataset. Rank them by consistency score i.e., lowest variance in daily revenue. [3 Marks] b) Find customers who had a high average spend (top 20% of all customers) in the first half of the year, but no orders in the last two months. These customers are at risk of churn—calculate the potential revenue loss. [4 Marks] c) If the company wants to increase low-rated restaurant revenue (<code>average_rating < 3.5</code>), simulate offering a ₹50 discount per order for these restaurants and estimate the total discount cost and the potential increase in number of orders if each discount triggers a 5% order boost (assume constant customer base). [3 Marks]	10	3	K3

2	Assume that you are a data analyst at a retail company. The company has provided you with a dataset containing the monthly sales records of its stores. The dataset <code>sales_data</code> has the following columns: <code>StoreID</code>, <code>Region</code>, <code>Product</code>, <code>UnitsSold</code>, <code>UnitPrice</code>, <code>Month</code>, <code>Discount</code>. You are asked to perform the following tasks using R Programming: - Add a new column <code>TotalSale</code> which is the total revenue for each transaction ($\text{UnitsSold} \times \text{UnitPrice} \times (1 - \text{Discount}/100)$). [2 Marks] - Find all transactions where <code>TotalSale</code> exceeds 50,000. [2 Marks] - For each region, find the product with the highest total sales. [2 Marks] - Summarize the dataset to find total units sold and total revenue per month. Create a wide table where each <code>Month</code> becomes a column, showing total revenue. [2 Marks] - Rank stores based on total revenue across all months, and display the top 5 stores. [2 Marks]	10	3	K4
3	You are a data analyst in a company managing HR and payroll data. The company provides you with two datasets: - Employees: This dataset contains the following employee details: <code>EmpID</code>, <code>Name</code>, <code>Department</code>, <code>JoinDate</code>, and <code>ManagerID</code> - Payroll: This dataset contains the following salary payment information: <code>EmpID</code>, <code>Month</code>, <code>BaseSalary</code>, <code>Bonus</code>, and <code>Deductions</code> Perform the following using R Programming: - Merge employees and payroll so that each salary record includes employee details. [2 Marks] - Create a dataset containing only payroll records for existing employees. [2 Marks] - Create a dataset with all employees, including those who haven't received any salary. [2 Marks] - For each manager, find the total payroll ($\text{BaseSalary} + \text{Bonus} - \text{Deductions}$) for their team. [4 Marks]	10	2	K6
4	A hospital needs to automate the day-to-day process. The hospital maintains a MySQL database called <code>hospital_db</code> to manage patient records, doctors, appointments, and treatments. You are a data analyst at a large hospital and you have been asked to analyze patient visits, doctor performance, and treatment effectiveness. The database contains the following tables: - Patients, which contains the following columns: <code>PatientID</code>, <code>Name</code>, <code>DateofBirth</code>, <code>Gender</code>, <code>JoinDate</code> - Doctors, which contains the following columns: <code>DoctorID</code>, <code>Name</code>, <code>Department</code>, <code>JoinDate</code> - Appointments, which contains the following columns: <code>AppointmentID</code>, <code>PatientID</code>, <code>DoctorID</code>, <code>AppointmentDate</code>, <code>Diagnosis</code>, <code>Status</code> - Treatments, which contains the following columns: <code>TreatmentID</code>, <code>AppointmentID</code>, <code>TreatmentType</code>, <code>Cost</code>, <code>Outcome</code> Perform the following using R Programming: - Establish a connection to <code>hospital_db</code> using R. [1 Mark]	10	2	K5

	b) Fetch all patients who joined after 2023-01-01. [1 Mark] c) Retrieve all appointments and treatments for a specific patient, including doctor name, treatment type, and outcome. [2 Marks] d) Calculate the total revenue generated per department from treatments. [2 Marks] e) Find the number of successful treatments per doctor. [2 Marks] f) List all appointments scheduled in the next 7 days with patient and doctor details. [2 Marks]			
5	A social media analytics company collects daily engagement data from multiple social media platforms (e.g., Twitter, Facebook, Instagram) in JSON format. Each JSON file contains nested data about posts, likes, comments, shares, and user details. The following is the sample JSON structure for one day: <pre> { "date": "2025-09-01", "platform": "Twitter", "posts": [{ "PostID": "T1001", "User": { "UserID": "U101", "UserName": "Alice" }, "Content": "Just posted a new blog!", "Metrics": { "Likes": 120, "Comments": 15, "Shares": 10 } }, { "PostID": "T1002", "User": { "UserID": "U102", "UserName": "Bob" }, "Content": "Check out my new video.", "Metrics": { "Likes": 250, "Comments": 30, "Shares": 25 } }] } </pre> Perform the following using R Programming: a) Read the JSON file into R and explore its structure. [1 Mark] b) Convert the nested JSON into a tidy data.frame where each row represents a post, including columns: PostID, UserID, UserName, Content, Likes, Comments, Shares. [2 Marks]	10	3	K4

- | | | | | |
|--|--|--|--|--|
| | <p>c) Create a new column $\text{TotalEngagement} = \text{Likes} + \text{Comments} + \text{Shares}$ and filter the top 5 posts with the highest engagement. [3 Marks]</p> <p>d) Aggregate the data to calculate total likes, comments, shares, and total engagement per user. [3 Marks]</p> <p>e) Save the flattened and processed data into a CSV file for reporting. [1 Mark]</p> | | | |
|--|--|--|--|--|

*****All the best *****