Continuous Assessment Test (CAT) – I - JANUARY 2025

Programme	:	B.Tech. CSE	Semester	:	Winter
Course Code	:	BCSE206L	Class Number	:	CH2024250502049 CH2024250502551 CH2024250502047
Course Title	:	FOUNDATIONS OF DATA SCIENCE			
Faculty	:	Dr. Saranyaraj D Dr. Arthi M Dr. Dominic Savio M	Slot	:	C1+TC1
Time	:	90 minutes	Max. Marks	:	50

Answer all the Questions

Q. No.	Sub-division	Question Text	Marks
1		A retail company, ShopEase, operates a chain of supermarkets across multiple cities. The company is facing challenges in managing its inventory and understanding customer purchasing patterns, leading to frequent stockouts and excess inventory. To address these issues, they plan to utilize a data-driven approach. As a data scientist, you have been tasked with the following: Analyze customer purchasing patterns to identify high-demand products, predict future sales trends to optimize inventory levels, and provide actionable insights to improve the company's overall business intelligence strategy. a. Identify the steps in the data science process you would follow to address the company's challenges. Briefly explain the role of each step. (4) b. Analyse and identify the types of data would you need from ShopEase, and how would you preprocess this data to ensure its quality? (3) c. Data visualization techniques can be used to support business intelligence for inventory optimization. Justify your answer. (3)	10
2		A startup, HealthTrack, is building a platform to monitor patient health using wearable devices. These devices generate real-time data on heart rate, sleep patterns, and physical activity. The startup aims to analyze this data to provide personalized health insights to users. To kickstart this project, HealthTrack is hiring a data scientist to lead the initiative. The company needs someone skilled in handling and analyzing large datasets. Developing machine learning models for predictive analytics, Visualizing insights effectively for non-technical stakeholders. a. List the essential technical and non-technical skills a data scientist should have to excel in this role. (5) b. Suggest three programming tools or software platforms suitable for handling and analyzing the large volumes of real-time data generated by the wearable	10

		devices. (5)	
3.		You are an analyst for an online mobile company. The company wants to analyze sales data to understand customer behavior, product performance, and revenue trends. The database contains the following schema: **Customers** (CustomerID (Primary Key), CustomerName, Email, Country) **OrderInfo** (OrderID (Primary Key), CustomerID (ForeignKey), OrderDate, TotalAmount) **OrderDetail** (OrderDetailID (Primary Key), OrderID (Foreign Key), ProductID, Quantity, UnitPrice) **ProductInfo** (ProductID (Primary Key), ProductName, Category, Price) Solve the below queries using appropriate SQL functions. i. Identify suitable function to merge the common features from the databases Order_detail and Order_Info respectively. (2) ii. Identify the products which generated the highest revenue. (2) iii. Apply suitable window function and group the customers into 3 groups based on product price. (2) iv. Identify the top 5 customers by total spending. (2) v. Find the revenue generated by each product category. (2)	10
4		You are a data analyst for a hospital chain, tasked with analyzing patient data, revenue trends, and performance metrics across departments and doctors. Consider the hospital database schema given below: **Patients** (PatientID (Primary Key), PatientName, Age, Gender) **Appointments** (AppointmentID (Primary Key), PatientID (Foreign Key), DoctorID (Foreign Key), AppointmentDate, FeesCharged) **Doctors** (DoctorID (Primary Key), DoctorName, Specialization, Department) ***Departments*** *(DepartmentID (Primary Key), DepartmentName, Questions)* *Write SQL query for the statements for the following statements:* *i. Rank the doctors based on the total revenue they generate from appointments.* *ii. Calculate the cumulative revenue for each department over time, grouped by appointment date.* *iii. Calculate the monthly number of appointments and their cumulative totals.* *iv. Calculate the average revenue per appointment for each doctor and rank doctors within their specialization based on this metric.* *v. Calculate the total number of visits for each patient and the average number of days between their visits.*	10
		a) An e-commerce platform is analyzing customer purchase patterns to optimize its inventory. The company uses descriptive analytics to summarize past sales data but struggles with overstocking and understocking. Evaluate why descriptive analytics alone may not suffice. Propose an advanced data analytics approach, detailing how it can provide actionable insights for inventory optimization. (5) b) For each of the NoSQL databases for Data Science, give a most appropriate application. (5)	10

***** ALL THE BEST *****