

590



VIT

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

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING (SCOPE)
CONTINUOUS ASSESSMENT TEST - I
WINTER SEMESTER 2024-2025

SLOT: C2

Programme Name & Branch :B.Tech(CSE)
Course Code and Course Name :BCSE318L
Faculty Name(s) :Prakash G, Devipriya A, Adaline Suji R, Geetha Mary A,
Padmavathy T, Jagalingam, Senthil Kumar K
Class Number(s) : 6630, 1609, 1588, 1573, 1605, 1720, 1569
Date of Examination :29-01-2025, 2.00pm to 3.30pm
Exam Duration :90 minutes Maximum Marks: 50

General instruction(s):

- Answer All Questions

Q. No	Question	M	CO	BL																														
1.	What is your understanding of data privacy, and why is it important in today's world? Discuss the data privacy ECO-system.	10	CO1	BL2																														
2.	A bank is preparing to share its financial transactions dataset with external analysts to study spending patterns and improve fraud detection algorithms. However, since the dataset contains sensitive information (e.g., transaction id, amount, customer details, and geolocations etc.), it is critical to ensure privacy protection through Statistical Disclosure Control (SDC) techniques. Identify the different stages in the SDC approaches, and infer the specific inputs and outputs associated with each stage on the above scenario.	10	CO1	BL2																														
3.	a. Calculate the global risk level for the micro data given below: (5 marks) <table border="1"> <thead> <tr> <th>ID</th><th>Age</th><th>Gender</th><th>Zip Code</th><th>Salary</th></tr> </thead> <tbody> <tr> <td>1</td><td>25</td><td>Male</td><td>12345</td><td>50,000</td></tr> <tr> <td>2</td><td>40</td><td>Female</td><td>12345</td><td>70,000</td></tr> <tr> <td>3</td><td>40</td><td>Female</td><td>12346</td><td>70,000</td></tr> <tr> <td>4</td><td>25</td><td>Male</td><td>12346</td><td>50,000</td></tr> <tr> <td>5</td><td>35</td><td>Female</td><td>12347</td><td>80,000</td></tr> </tbody> </table> b. Convert the above micro data into an anonymized dataset using any of the non-perturbative methods. (5 marks)	ID	Age	Gender	Zip Code	Salary	1	25	Male	12345	50,000	2	40	Female	12345	70,000	3	40	Female	12346	70,000	4	25	Male	12346	50,000	5	35	Female	12347	80,000	10	CO2	BL3
ID	Age	Gender	Zip Code	Salary																														
1	25	Male	12345	50,000																														
2	40	Female	12345	70,000																														
3	40	Female	12346	70,000																														
4	25	Male	12346	50,000																														
5	35	Female	12347	80,000																														



VIT

Vellore Institute of Technology
(Deemed to be University under section 3 of UGE Act, 1986)

REG.NO.:

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING (SCOPE)
CONTINUOUS ASSESSMENT TEST - I
WINTER SEMESTER 2024-2025

SLOT: C2

WINTER SEMESTER 2024-2025		SLOT: C2																																						
4	Apply the following perturbation techniques to the dataset provided below to ensure data privacy: a. Noise Addition (5 marks) b. Micro-aggregation (5 marks) <table><tr><th>ID</th><th>Age</th><th>Gender</th><th>Diagnosis</th><th>Treatment_Cost</th><th>Visit_Date</th></tr><tr><td>P1</td><td>45</td><td>Male</td><td>Diabetes</td><td>1500</td><td>10/01/24</td></tr><tr><td>P2</td><td>29</td><td>Female</td><td>Hypertension</td><td>2000</td><td>12/01/24</td></tr><tr><td>P3</td><td>37</td><td>Male</td><td>Asthma</td><td>1800</td><td>14/01/24</td></tr><tr><td>P4</td><td>50</td><td>Female</td><td>Heart Disease</td><td>2500</td><td>15/01/24</td></tr><tr><td>P5</td><td>33</td><td>Male</td><td>Migraine</td><td>1200</td><td>17/01/24</td></tr></table>	ID	Age	Gender	Diagnosis	Treatment_Cost	Visit_Date	P1	45	Male	Diabetes	1500	10/01/24	P2	29	Female	Hypertension	2000	12/01/24	P3	37	Male	Asthma	1800	14/01/24	P4	50	Female	Heart Disease	2500	15/01/24	P5	33	Male	Migraine	1200	17/01/24	10	CO2	BL3
ID	Age	Gender	Diagnosis	Treatment_Cost	Visit_Date																																			
P1	45	Male	Diabetes	1500	10/01/24																																			
P2	29	Female	Hypertension	2000	12/01/24																																			
P3	37	Male	Asthma	1800	14/01/24																																			
P4	50	Female	Heart Disease	2500	15/01/24																																			
P5	33	Male	Migraine	1200	17/01/24																																			
5	Imagine a healthcare organization wants to share their patient data with researchers to study disease patterns and improve treatments. The organization needs to remove or mask personal details like names and addresses to protect patient privacy. However, they also want the data to remain accurate enough for researchers to find meaningful insights. How can a healthcare organization share patient data with researchers in a way that protects privacy while keeping the information accurate enough for valuable analysis?	10	CO2	BL2																																				

<<<THE END>>>