

Similar to Final Exam Question

COMP3335 Fall 2024 Class Exercise 10

~~Final~~

Question 1. Make this Table 4-anonymous (Quasi-identifiers: ZipCode, Age, Nationality). *(Quasi-sensitive attribute)*

	<u>Zip code</u>	<u>Age</u>	<u>Nationality</u>	<u>Condition</u>
1	27609	18	Chinese	Heart Disease
2	27615	19	American	Heart Disease
3	26724	50	Indian	Cancer
4	26724	55	Chinese	Heart Disease
5	27615	21	Japanese	Viral Infection
6	26725	47	American	Viral Infection
7	27609	23	American	Viral Infection
8	27609	31	American	Cancer
9	27615	36	Japanese	Cancer
10	26725	49	American	Viral Infection
11	27609	37	Indian	Cancer
12	27615	35	American	Cancer

Question 2. Suppose the following medical record dataset has been published.

SSN	Name	Race	DateOfBirth	Sex	ZIP	Marital Status	HealthProblem
	asian	09/27/64	female	94139	divorced	hypertension	
	asian	09/30/64	female	94139	divorced	obesity	
	asian	04/18/64	male	94139	married	chest pain	
	asian	01/15/64	male	94139	married	obesity	
	black	03/13/63	male	94138	married	hypertension	
	black	03/18/63	male	94138	married	shortness of breath	
	black	09/13/64	female	94141	married	shortness of breath	
	black	09/07/64	female	94141	married	obesity	
	white	05/14/61	male	94138	single	chest pain	
	white	05/08/61	male	94138	single	obesity	
	white	09/15/61	female	94142	widow	shortness of breath	

Considering the publicly available voter registration list below, answer the following questions.

Name	Address	City	ZIP	DOB	Sex	Party
.....
.....
Sue J. Carlson	900 Market St.	San Francisco	94142	9/15/61	female	democrat
.....

- a) Answer the following questions:
 - i. What kind of anonymization has been performed on the released dataset before being published, if any?
 - ii. Explain if it is possible to infer any privacy-sensitive information from the released data using the voter list.
- b) Consider the medical record dataset above. Given attributes *Race*, *DOB*, and *Sex* together as the quasi-identifier, and attribute *Health Problem* as the sensitive attribute, create and report a 2-anonymous version of the medical dataset (k -anonymity where $k=2$). How many equivalency classes are there in the result dataset?
- c) Repeat the previous exercise for attributes *Sex*, *ZIP*, and *Marital Status* as quasi-identifier with $k = 3$.

(Q1) Making Table 4- chomous

	Zip Code	Age	Nationality	Condition
1	27609 -	18	Chinese -	Heart Disease -
2	27615 -	19	American -	Heart Disease -
3	26724 -	50	Indian -	Cancer -
4	26724 -	53	Chinese -	Heart Disease -
5	27615 -	21	Japanese -	Vircl Infection -
6	26725 -	47	American -	Vircl Infection -
7	27609 -	23	American -	Vircl Infection -
8	27609 -	31	American -	Cancer -
9	27615 -	36	Japanese -	Cancer -
10	26725 -	49	American -	Vircl Infection -
11	27609 -	37	Indian -	Cancer -
12	27615 -	35	American -	Cancer -



Making Table 4- chomous

	Zip Code	Age	Nationality	Condition
1	276**	<30	*	Heart Disease -
2	276**	<30	*	Heart Disease -
3	2672*	≥ 40	*	Cancer -
4	2672*	≥ 40	*	Heart Disease -
5	276**	<30	*	Vircl Infection -
6	2672*	≥ 40	*	Vircl Infection -
7	276**	<30	*	Vircl Infection -
8	276**	3*	*	Cancer -
9	276**	3*	*	Cancer -
10	2672*	≥ 40	*	Vircl Infection -
11	276**	3*	*	Cancer -
12	276**	3*	*	Cancer -

- (Q2) a) i) The name and SSN has been anonymized with cloning
 blank Naive anonymization, only explicit identifier removed
 ii) Yes, the name, DOB, address and city has been shown for
 full

a Certain user

k-Anonymity

K=2

As listed

Yes, because only one medical record matches the highlighted info in voter registration table based on quasi-identifiers DOB, SEX, ZIP is last one. So one can know Sue J. Carlson's health problem and marital status
 (mention ZIP/DOB as quasi-identifier, health problem as sensitive field)

	DOB	Sex
①	09/*/64	*
②	09/*/64	✗
③	04/*/64	✗
④	04/*/64	✗
⑤	03/*/63	✗
⑥	03/*/63	✗
⑦	09/*/64	✗
⑧	09/*/64	✗
⑨	19 61	✗
⑩	19 61	✗
⑪	19 61	✗

⑤ Equivalent classes

SEX	ZIP	Marital Status	Health Problem
female	94139	D	
F	94139	D	
M	94139	m	
M	94139	m	
M	94138	m	
M	94138	m	
M	94138	m	
F	94141	m	
F	94141	m	
M	94138	s	
M	94138	s	
F	94142	w	

↓

SEX	ZIP	Marital Status	H P
M	9413*	m	
F	941**	been married	
F	941**	been married	
F	941**	been married	
F	941**	been married	
F	941**	been married	
F	941**	been married	
M	94138	not-released	
M	94138	not-released	

①

②

③

Shortness of breath