Bases de Données Avancées
Extensions au k-anoymat

Jean-François COUCHOT
couchot[arobase]femto-st[point]fr

1er mars 2022

QID			SA	
Gender	**ZIP Code**	**Age**	**Disease**	**Salary**
Male	400071	35	bronchitis	10k
Male	400182	37	pneumonia	11k
Male	400095	39	stomach cancer	12k
Female	440672	54	gastritis	12k
Female	440123	58	Flu	15k
Male	440893	54	bronchitis	16k
Male	400022	41	gastric ulcer	16k
Male	400135	46	gastritis	17k
Female	400182	44	stomach cancer	18k

TABLE 1 – Données originales

Gender	Zip Code	Age	Disease	Salary
*	400***	[30, 40[	bronchitis	10k
*	400***	[30,40[	pneumonia	11k
*	400***	[30, 40[	stomach cancer	12k
*	400***	[40, 50[	stomach cancer	18k
*	400***	[40, 50[	gastric ulcer	16k
*	400***	[40, 50[	gastritis	17k
*	440***	[50, 60[	gastritis	12k
*	440***	[50, 60[	Flu	15k
*	440***	[50, 60[	bronchitis	16k

TABLE 2 – Une version 3-diverse

Exercice 0.1 (l-diversité). *La table 1 reprend un jeu de donné minime issu de* [1]

1. *Quelles stratégies de généralisation proposez-vous ?*
2. *On considère la proposition donnée dans le tableau 2.*
   (a) *Est-elle 3-anonyme ?*
   (b) *Pourquoi est-elle 3-diverse ?*
   (c) *Quelle est sa valeur de Loss ?*

Exercice 0.2 (t-proximité). *On reprend le même jeu de donné que précédemment.*

1. *Calculer l'indice de proximité de la proposition donnée.*
2. *Existerait-il une autre proposition (différente de la suppression de tous les QID) qui aurait un indice t de proximité plus faible ? Dans ce cas, quel serait sa valeur de Loss ?*

1. Elabd, E., Abdulkader, H., & Mubark, A. (2015). L–diversity-based semantic anonymaztion for data publishing. IJ Information Technology and Computer Science (IJITCS), 10, 1-7.