Mathis Moulin, Raphael Lacôte, Samuel Bérion, Cléo Eiras, Jade Van Brabandt

---

TP6 - Privacy Database

Contexte : Ce TP a pour but d'illustrer les risques du partage d'une base de données ainsi que de présenter certaines techniques d'anonymisation de données permettant un partage plus sécurisé. Les exercices s'appuieront sur les différentes techniques vues en CM et en TD.
Pour ce TP vous avez accès à une base de données ainsi que d'un script python permettant de communiquer avec la base de données. Votre rôle sera de modifier le script python selon les besoins des exercices.
Vous n'aurez jamais besoin d'interagir avec la base de données directement.

Exercise 1 (K-Anonymity (10 points))

Rappel : https://en.wikipedia.org/wiki/K-anonymity

1. Vous savez qu'un certain monsieur **Geneva Preston** est présent dans la base de données. À l'aide du script python, quel est le diagnostic médical de monsieur **Geneva Preston** ? (1 point)

2. Le champ “Nom” dans la base de données est dit “identifiant” car il permet d'identifier une personne à l'aide de cette information seule. Implémentez une façon de rendre l'identification d'un individu plus compliquée, expliquez votre démarche. (3 point)

3. Monsieur **Geneva Preston** n'est désormais plus identifiable par son nom. Cependant, vous savez que monsieur **Geneva Preston** est un **homme (M)** de **24** ans et que son métier est **Ingenieur**. À l'aide de ces informations, retrouvez le diagnostic de monsieur **Geneva Preston**. (1 point)

4. En vous basant sur les principes de k-anonymity, rendez monsieur **Geneva Preston** non identifiable via ces trois attributs. (5 points)

Exercise 2 (L-diversity (5 points))

Rappel : https://en.wikipedia.org/wiki/L-diversity

1. Monsieur **Geneva Preston** est désormais introuvable dans la base de données. Cependant, vous savez que monsieur **Geneva Preston** fait partie de cette base de données, que pouvez vous deviner du diagnostic médical de monsieur **Geneva Preston** ? Pourquoi ? (2 points)

2. En vous basant sur les principes de l-diversity, rendez le diagnostic de monsieur **Geneva Preston** moins devinable. (3 points)

Exercise 3 (Differential privacy (5 points))

Rappel : https://en.wikipedia.org/wiki/Differential_privacy

1. A l'aide d'une nouvelle fonction, trouvez combien de personnes sont atteintes d'une grippe dans la base de données ? (1 point)

2. A partir d'un nombre de lignes retournés par une requête, pouvez vous déterminer si monsieur **Geneva Preston** est atteint d'une grippe ? (1 point)

3. Vous avez désormais trouvé une manière détournée de récolter des informations sur quelqu'un. Trouvez une façon d'empêcher ce genre d'attaque. Implémentez cette solution et vérifiez qu'il n'est plus possible de trouver ces informations. (3 points)