



Module 8 - Sciences des données et éthiques

Exercices - Solutions

Exercice 1 :

La sécurité des données fait référence au résultat (le fait que les données soient protégées).
La protection des données a trait aux mesures destinées à assurer la sécurité des données.

Exercice 2 :

Dans le tableau de l'exercice 2, nous avons deux colonnes qui contiennent des informations d'identification à savoir le nom et l'âge de la personne. Nous pouvons tout simplement remplacer le nom par une étoile (*) et l'âge par une colonne contenant des groupes d'âges. Dans ce cas, nous avons trois groupes d'âges. Donc, $K=3$. Le tableau 1 montre donc le résultat de l'anonymisation :

TABLE 1 – Données démographiques anonymisées.

ID	Nom	Genre	Age	Province
1	*	F	$20 < \text{age} \leq 30$	QC
2	*	M	$\text{age} \leq 20$	ON
3	*	M	$40 < \text{age} \leq 50$	AB

Exercice 3 :

La méthode de K -anonymisation est vulnérable aux attaques de type "attribute linkage" :

- attaques basées sur l'homogénéité : ce type d'attaques pourrait se produire lorsque les valeurs sensibles (age, maladie, code postal, etc.) dans un ensemble de K enregistrements sont identiques. Ça devient donc très simples pour les attaquants de dés-identifier les données.

- attaques basées sur la connaissance à priori : la connaissance à priori pourrait être n'importe quel fait connu qui n'est pas en soi une divulgation de la vie privée, mais l'attaquant le combine avec d'autres informations pour tirer des conclusions plus précises sur les informations sensibles des individus.

Exercice 4 :

Le chiffrement symétrique possède beaucoup d'inconvénients, on peut citer par exemple :

- Le chiffrement symétrique n'assure que la confidentialité des données, contrairement au chiffrement asymétrique qui assure des principes de sécurité supplémentaire.
- Une clé symétrique correspond à un échange entre 2 personnes. Donc, pour communiquer avec d'autres personnes il faudra d'autres clés symétriques. Soit une clé pour chaque échange.