

# SAÉ 6 parcours C : Anonymisation d'une Base de données

## Expression du besoin

---

### Introduction

Ce document vous est fourni par une professeure de mathématique ne connaissant pas les méthodes et contraintes de l'anonymisation de données.

### Contexte

Vous travaillez pour le service informatique de protection des données de l'université de Rennes.

Une professeure de mathématique veut donner à ses étudiants de première année une base de données sur les Jeux Olympiques de 1896 à 2022 afin que ses élèves procèdent à son analyse.

Soucieuse du contexte scolaire de cette analyse elle doit procéder à l'anonymisation des données, pour ce faire elle demande à votre équipe de procéder à l'anonymisation irréversible des données.

Attention, ces données seront utilisées à des fins calculatoires, aussi le jeu de données anonymisé devra conserver intégralement son sens en vue de l'analyse statistique. Afin de vous aider dans votre tâche votre collègue vous indique les statistiques qu'elle souhaite réaliser.

### Les analyses qui seront réalisées

- Analyse des âges des joueurs d'après : leur pays, leur discipline, l'année des jeux
- Classement des joueurs d'après leur âge
- Résultats dans les disciplines d'après les pays
- Durée moyenne des jeux d'après leur année
- Corrélation entre le pays d'accueil des Jeux Olympiques et les médailles remportées
- Analyse entre la saison et le lieu d'accueil des Jeux (il serait aberrant de faire les jeux d'hiver au Sahara n'est-ce pas...)

## Les données

Le script de création et peuplement des tables vous est fourni ainsi que les fichiers CSV requis. Il y a 6 tables et une vue décrites par le schéma suivant.

