**GANDONOU Sèdjro Amos**

**PROJET EN SYSTEME D’INFORMATIONS DECISIONNELLES**

# Données : 2021 Olympics in Tokyo

Les données utilisées pour ce projet (2021 Olympics in Tokyo) ont été récupérées sur **kaggle** au lien : <https://www.kaggle.com/datasets/arjunprasadsarkhel/2021-olympics-in-tokyo>.

# Présentation des données

Les données relatives aux jeux olympiques de Tokyo 2021 ont été élaborées par : **Arjun Prasad Sarkhel**. Elles donnent des informations sur les athlètes et équipes ayant compéties au cours du tournoi. Elles donnent les détails concernant leurs disciplines, leurs pays et les médailles gagnées.

**Le jeu de données**

Il est composé de 5 tables à savoir :

1. **Athletes**

Donne les informations relatives aux athlètes :

***NOC*** : Le comité national olympique de l’athlète

***Name*** : Prénom et Nom de l’athlète

***Discipline*** : Sa discipline (Basketball, Volleyball, …)

1. **Coaches**

Donne les informations relatives aux athlètes :

***NOC*** : Le comité national olympique de provenance

***Name*** : Prénom et Nom de l’entraineur

***Discipline*** : Sa discipline (Basketball, Volleyball, …)

***Event*** : La nature des équipes qu’il entraine (masculine, feminine, mixte)

1. **GenderEntries**

Présente une vue globale sur le genre des compétiteurs par discipline :

***Discipline*** : La discipline

***Female*** : Le nombre de femme dans la discipline

***Male*** : Le nombre d’hommes dans ladite discipline

***Total*** : Nombre de femmes et d’homme par discipline.

1. **Teams**

Cette présente les différentes équipes ayant participées aux jeux olympiques :

***Name*** : Pays de provenance de l’équipe

***Discipline*** : Discipline de l’équipe

***NOC*** : Comité National Olympique de l’équipe

***Event*** : Nature de l’équipe (masculine, féminine, mixte, ...)

1. **Medals**

Cette présente les résultats en termes de médailles remportées :

***Rank*** : Le classement de l’équipe à l’issue des jeux

***Team***/***NOC*** : Le comité national olympique de l’équipe

***Gold*** : Le nombre de médailles d’or remportées par l’équipe

***Silver*** : Le nombre de médailles d’argent remportées par l’équipe

***Bronze*** : Le nombre de médailles de bronze remportées par l’équipe

***Total*** : Le total cumulé des différentes médailles obtenues par l’équipe

***Rank*** ***by*** ***Total*** : Classement de l’équipe (par rapport au nombre total de

médailles)

# Quelques pistes d’analyse des données

# Les données étant relatives aux athlètes, aux équipes et aux prix remportées, nous restreignons notre champ aux données (tables) relatives aux équipes et aux distinctions remportées par ces dernières, d’où l’usage des tables *Teams* et *Medals.*

# Les pistes d’analyse envisageables au travers de la disposition de ces données sont entre autres :

# La recherche des 5 meilleures équipes en termes de :

# Nombre de médailles d’or

# Nombre de médailles d’argent

# Nombre de médailles de bronze

# Nombre de distinctions (tous médailles confondus)

# Les 5 meilleurs pays les plus médaillés

# Les disciplines les plus distinguées :

# Par pays : top 3 disciplines

# Etc.

# Mise en place d’un ETL avec Python

L’extraction, la transformation et le chargement dans un entrepôt de données ont été réalisé dans Python

* Le fichier ***etl\_extraction.py*** a servi a récupérer en ligne les 5 tables constitutives de nos jeux de données
* Le fichier ***etl\_exploration.ipynb*** nous a permis, de comprendre nos jeux de données en vue d’identifier les transformations nécessaires pour aboutir à une obtention des résultats en vue.
* Le fichier ***etl\_tranformation.ipynb*** comporte le code des transformations réalisées sur les tables notamment : l’affection de nouveaux noms aux colonnes
* Le fichier ***etl\_load.py*** nous a servi a insérer les différentes données de nos tables dans des tables équivalentes (préalablement créées) de notre entrepôt de données appelé : ***jeux\_olympique\_db***

# *Informations supplémentaires:*

# 

# Création des tables de la base de données jeux\_olympique\_db:

# CREATE TABLE teams (

# id INT AUTO\_INCREMENT PRIMARY KEY,

# country VARCHAR(30),

# discipline VARCHAR(20),

# committee VARCHAR(40),

# event VARCHAR(255)

# );

# 

# CREATE TABLE medals (

# id INT AUTO\_INCREMENT PRIMARY KEY,

# rank INT,

# committee VARCHAR(40),

# gold INT,

# silver INT,

# bronze INT,

# total INT,

# rank\_by\_total INT

# );

# 

# 

# *REPORTING:*

# Connection à la base de données:

# 

# 

# 

# 

# 