

PROJET EN SYSTEME D'INFORMATIONS DECISIONNELLES

1. 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>.

2. 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 :

a) 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, ...)

b) Coaches

Donne les informations relatives aux athlètes :

NOC : Le comité national olympique de provenance

Name : Prénom et Nom de l'entraîneur

Discipline : Sa discipline (Basketball, Volleyball, ...)

Event : La nature des équipes qu'il entraîne (masculine, feminine, mixte)

c) 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.

d) **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, ...)

e) **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)

3. 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.

4. Mise en place d'un ETL avec Python

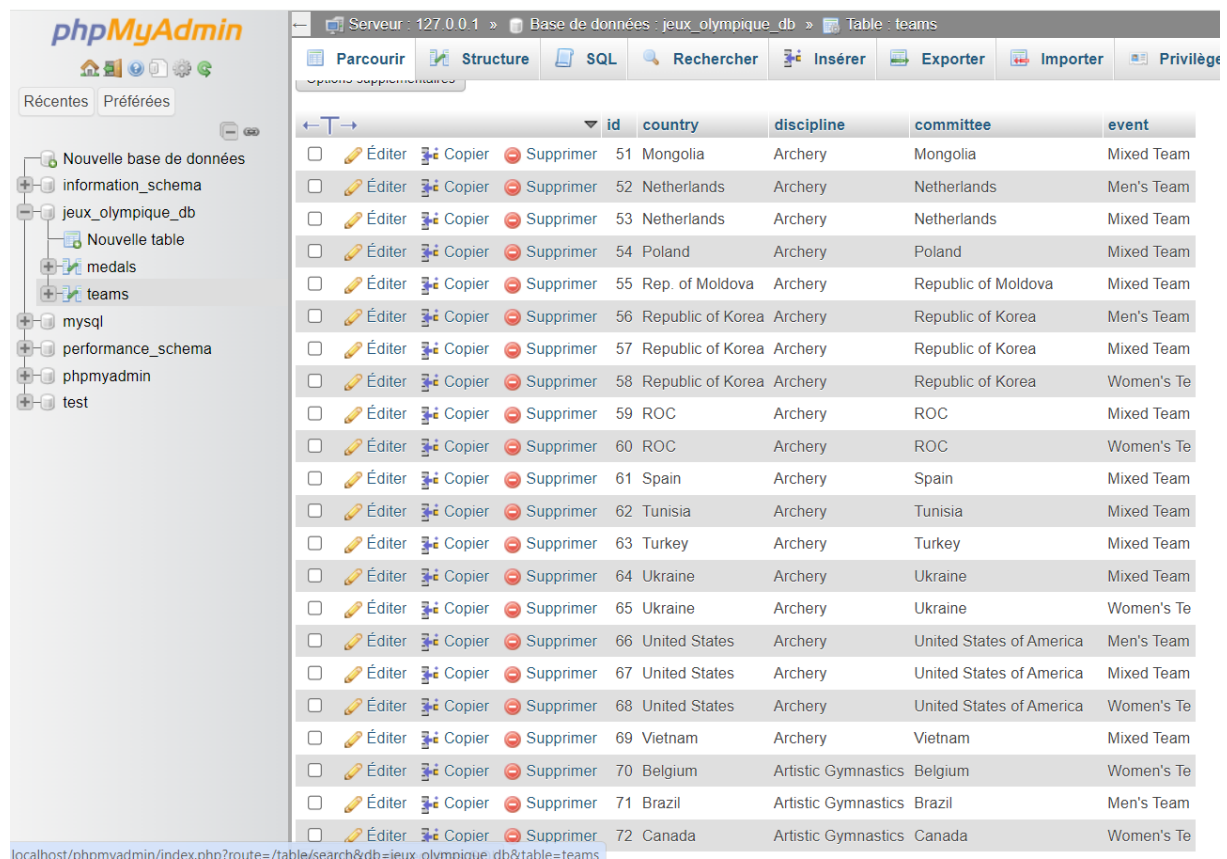
L'extraction, la transformation et le chargement dans un entrepôt de données ont été réalisé dans Python

- Le fichier **etl_extraction.py** a servi à 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_transformation.ipynb** comporte le code des transformations réalisées sur les tables notamment : l'affectation de nouveaux noms aux colonnes
- Le fichier **etl_load.py** nous a servi à 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)
);
```



	id	country	discipline	committee	event
<input type="checkbox"/>	51	Mongolia	Archery	Mongolia	Mixed Team
<input type="checkbox"/>	52	Netherlands	Archery	Netherlands	Men's Team
<input type="checkbox"/>	53	Netherlands	Archery	Netherlands	Mixed Team
<input type="checkbox"/>	54	Poland	Archery	Poland	Mixed Team
<input type="checkbox"/>	55	Rep. of Moldova	Archery	Republic of Moldova	Mixed Team
<input type="checkbox"/>	56	Republic of Korea	Archery	Republic of Korea	Men's Team
<input type="checkbox"/>	57	Republic of Korea	Archery	Republic of Korea	Mixed Team
<input type="checkbox"/>	58	Republic of Korea	Archery	Republic of Korea	Women's Te
<input type="checkbox"/>	59	ROC	Archery	ROC	Mixed Team
<input type="checkbox"/>	60	ROC	Archery	ROC	Women's Te
<input type="checkbox"/>	61	Spain	Archery	Spain	Mixed Team
<input type="checkbox"/>	62	Tunisia	Archery	Tunisia	Mixed Team
<input type="checkbox"/>	63	Turkey	Archery	Turkey	Mixed Team
<input type="checkbox"/>	64	Ukraine	Archery	Ukraine	Mixed Team
<input type="checkbox"/>	65	Ukraine	Archery	Ukraine	Women's Te
<input type="checkbox"/>	66	United States	Archery	United States of America	Men's Team
<input type="checkbox"/>	67	United States	Archery	United States of America	Mixed Team
<input type="checkbox"/>	68	United States	Archery	United States of America	Women's Te
<input type="checkbox"/>	69	Vietnam	Archery	Vietnam	Mixed Team
<input type="checkbox"/>	70	Belgium	Artistic Gymnastics	Belgium	Women's Te
<input type="checkbox"/>	71	Brazil	Artistic Gymnastics	Brazil	Men's Team
<input type="checkbox"/>	72	Canada	Artistic Gymnastics	Canada	Women's Te

```

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
);

```

phpMyAdmin

Récentes Préférences

- Nouvelle base de données
- information_schema
- jeux_olympique_db
 - Nouvelle table
 - medals
 - teams
- mysql
- performance_schema
- phpmyadmin
- test

Serveur : 127.0.0.1 » Base de données : jeux_olympique_db » Table : medals

Parcourir Structure SQL Rechercher Insérer Exporter Importer Privilèges

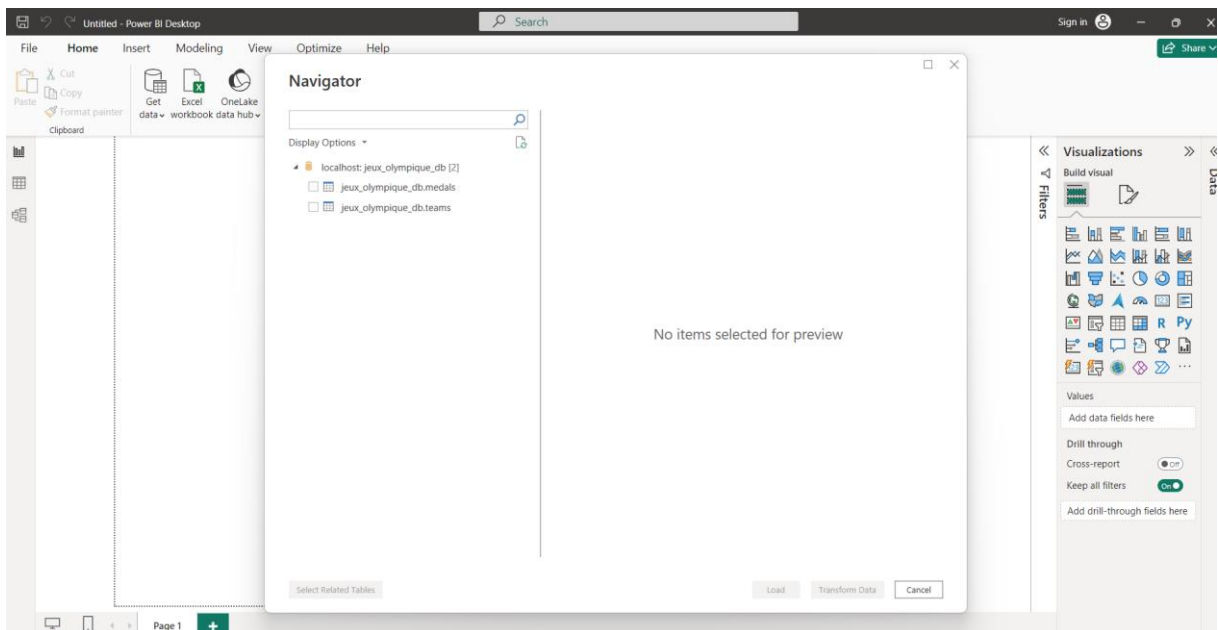
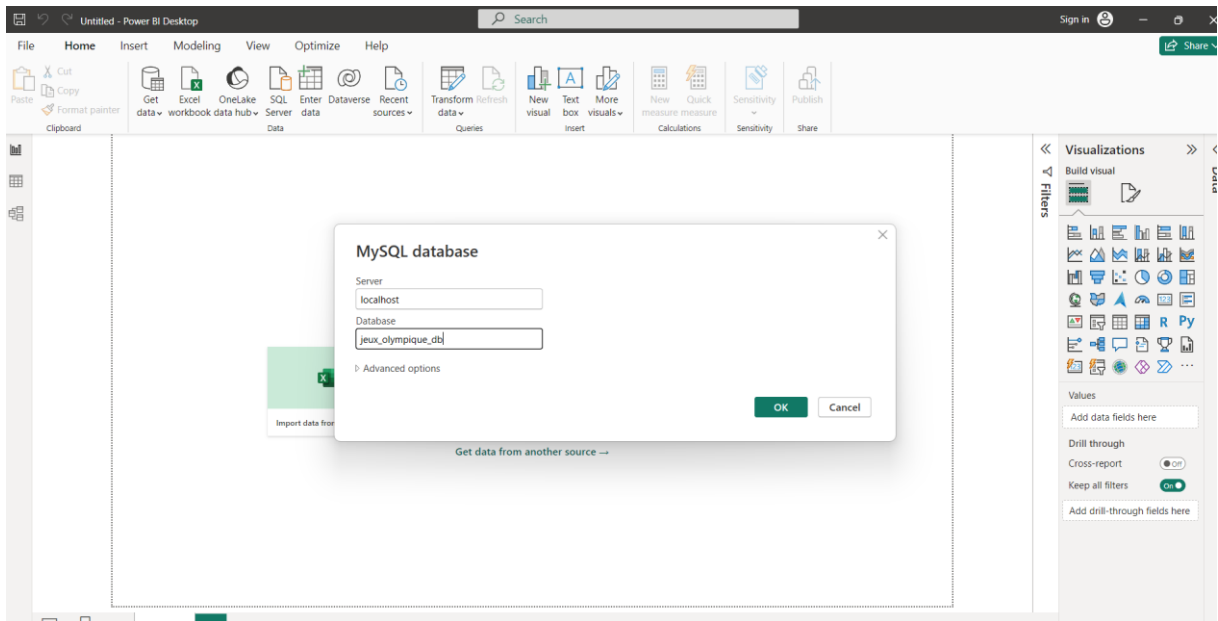
Options supplémentaires

				id	rank	committee	gold	silver	bronze	total	rank_by_total
<input type="checkbox"/>	Éditer	Copier	Supprimer	1	1	United States of America	39	41	33	113	1
<input type="checkbox"/>	Éditer	Copier	Supprimer	2	2	People's Republic of China	38	32	18	88	2
<input type="checkbox"/>	Éditer	Copier	Supprimer	3	3	Japan	27	14	17	58	5
<input type="checkbox"/>	Éditer	Copier	Supprimer	4	4	Great Britain	22	21	22	65	4
<input type="checkbox"/>	Éditer	Copier	Supprimer	5	5	ROC	20	28	23	71	3
<input type="checkbox"/>	Éditer	Copier	Supprimer	6	6	Australia	17	7	22	46	6
<input type="checkbox"/>	Éditer	Copier	Supprimer	7	7	Netherlands	10	12	14	36	9
<input type="checkbox"/>	Éditer	Copier	Supprimer	8	8	France	10	12	11	33	10
<input type="checkbox"/>	Éditer	Copier	Supprimer	9	9	Germany	10	11	16	37	8
<input type="checkbox"/>	Éditer	Copier	Supprimer	10	10	Italy	10	10	20	40	7
<input type="checkbox"/>	Éditer	Copier	Supprimer	11	11	Canada	7	6	11	24	11
<input type="checkbox"/>	Éditer	Copier	Supprimer	12	12	Brazil	7	6	8	21	12
<input type="checkbox"/>	Éditer	Copier	Supprimer	13	13	New Zealand	7	6	7	20	13
<input type="checkbox"/>	Éditer	Copier	Supprimer	14	14	Cuba	7	3	5	15	18
<input type="checkbox"/>	Éditer	Copier	Supprimer	15	15	Hungary	6	7	7	20	13
<input type="checkbox"/>	Éditer	Copier	Supprimer	16	16	Republic of Korea	6	4	10	20	13
<input type="checkbox"/>	Éditer	Copier	Supprimer	17	17	Poland	4	5	5	14	19
<input type="checkbox"/>	Éditer	Copier	Supprimer	18	18	Czech Republic	4	4	3	11	23
<input type="checkbox"/>	Éditer	Copier	Supprimer	19	19	Kenya	4	4	2	10	25
<input type="checkbox"/>	Éditer	Copier	Supprimer	20	20	Norway	4	2	2	8	29
<input type="checkbox"/>	Éditer	Copier	Supprimer	21	21	Jamaica	4	1	4	9	26
<input type="checkbox"/>	Éditer	Copier	Supprimer	22	22	Spain	3	8	6	17	17

Console de requêtes SQL

REPORTING:

Connexion à la base de données:



Power BI Desktop interface showing the Navigator pane. The selected table is **jeux_olympique_db.teams**. The table data is as follows:

id	country	discipline	committee
1	Belgium	3x3 Basketball	Belgium
2	China	3x3 Basketball	People's Republic of China
3	China	3x3 Basketball	People's Republic of China
4	France	3x3 Basketball	France
5	Italy	3x3 Basketball	Italy
6	Japan	3x3 Basketball	Japan
7	Japan	3x3 Basketball	Japan
8	Latvia	3x3 Basketball	Latvia
9	Mongolia	3x3 Basketball	Mongolia
10	Netherlands	3x3 Basketball	Netherlands
11	Poland	3x3 Basketball	Poland
12	ROC	3x3 Basketball	ROC
13	ROC	3x3 Basketball	ROC
14	Romania	3x3 Basketball	Romania
15	Serbia	3x3 Basketball	Serbia
16	United States	3x3 Basketball	United States of America
17	Australia	Archery	Australia
18	Australia	Archery	Australia
19	Bangladesh	Archery	Bangladesh
20	Belarus	Archery	Belarus
21	Brazil	Archery	Brazil
22	Canada	Archery	Canada
23	China	Archery	People's Republic of China

Power BI Desktop interface showing the Relationships pane. A relationship is established between the **teams** table (column: **committee**) and the **medals** table (column: **committee**). The relationship is set to **Many to one (N:1)**, **Yes** (Make this relationship active), and **Single** (Cross-filter direction).

teams table fields: committee, country, discipline, event, id, Collapse ^

medals table fields: bronze, committee, gold, id, rank, rank_by_total, silver, total, Collapse ^

