
Jeux Olympiques

1896 à 2016



Introduction et choix des données et le but final

- Elaborer l'aspect historique des Jeux Olympiques à travers les chiffres essentiels de la compétition.
- Analyser l'évolution de cet événement mondial depuis les JO de 1896

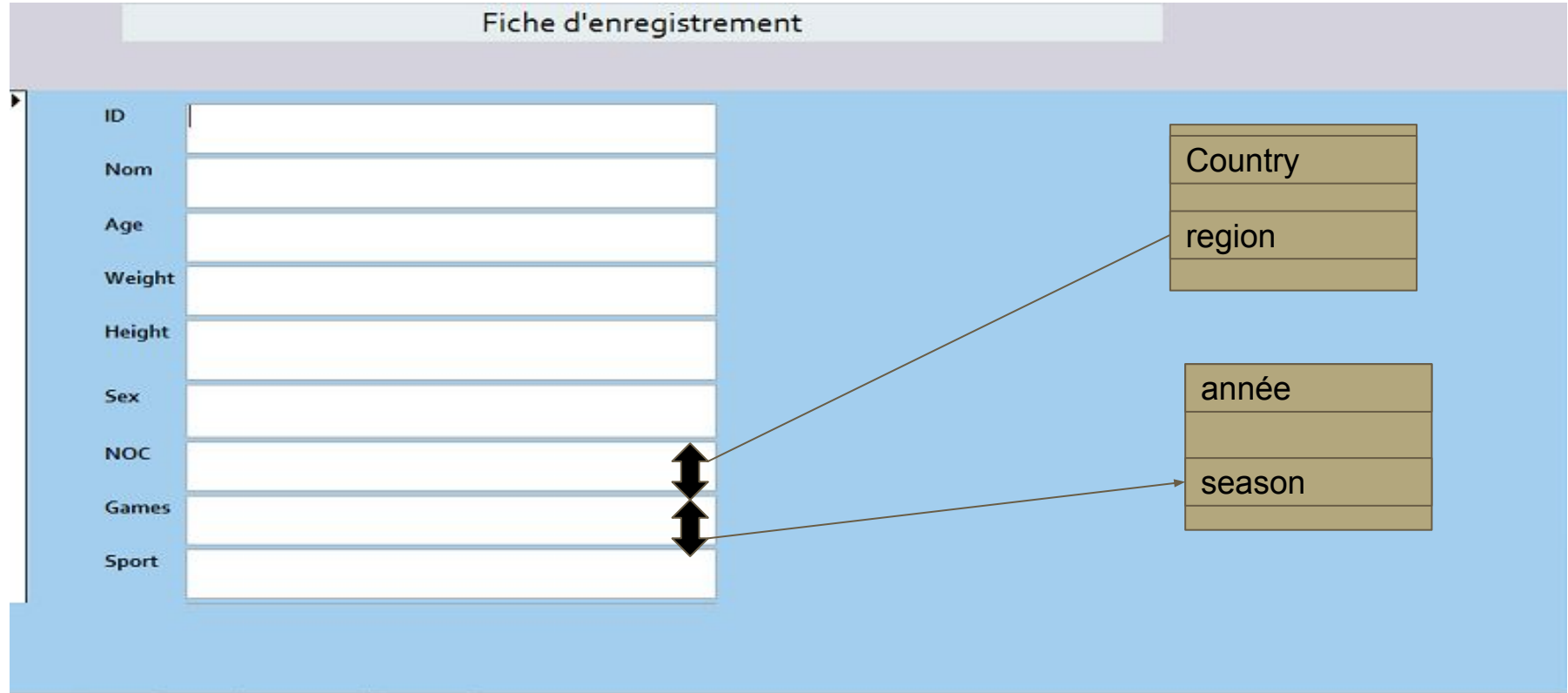


Organisation du projet et outils utilisés

1. Elaboration du Modèle conceptuel des données
2. Création des tables games,athletes et régions
3. Importation des données vers les tables citées ci-dessus
4. Les Relations entre les tables
5. Importation des librairies nécessaires au fonctionnement du notebook
6. Etablir une connection entre le serveur SQL et Python
7. Transformation des données en DataFrames Pandas
8. Analyse



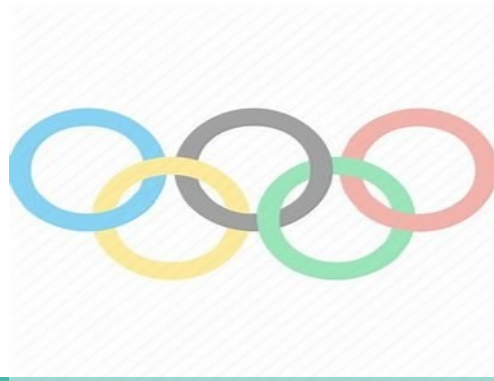
Modèle conceptuel des données



Création des tables dans MySQL Workbench

Création de la table games :

```
CREATE TABLE jo.games  
(  
  name VARCHAR(15) PRIMARY KEY NOT NULL,  
  year INT,  
  season VARCHAR(8),  
  city VARCHAR(30)  
);
```



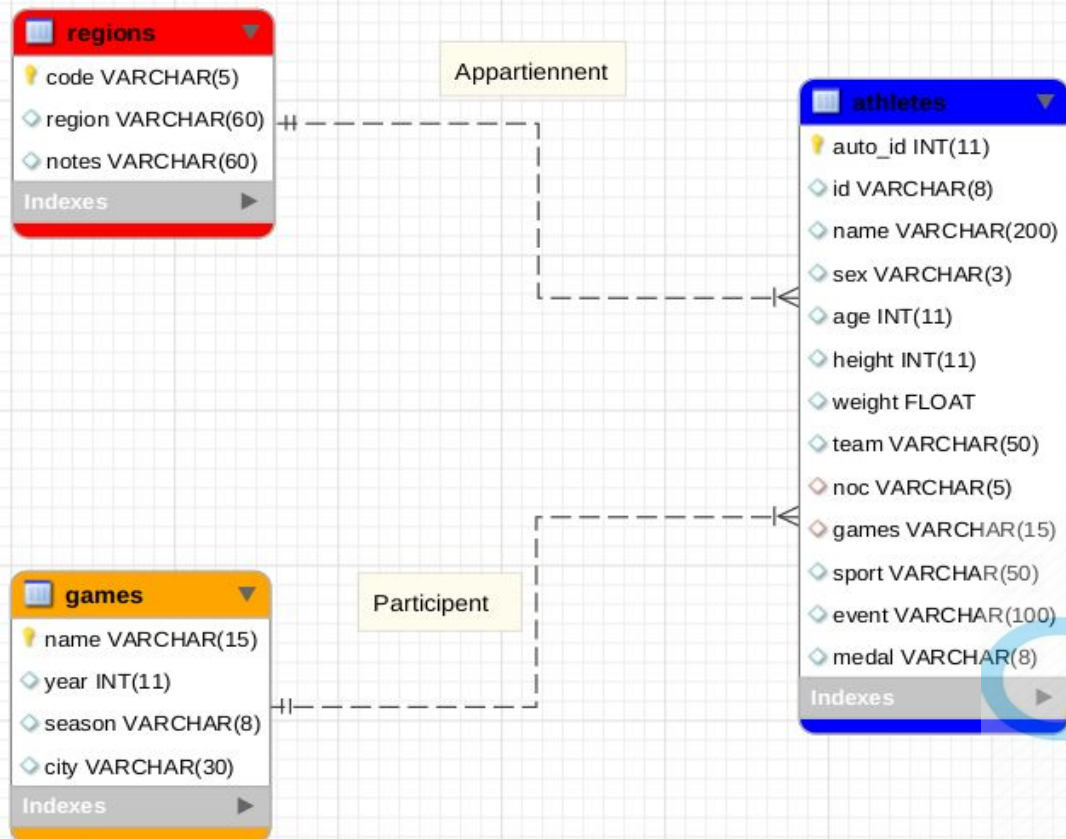
Importation des données avec MySQL Workbench

Importation des données vers la table athletes. Modification du fichier CSV au préalable pour enlever les colonnes non voulues.

```
LOAD DATA INFILE "C:/ProgramData/MySQL/MySQL Server 8.0/ Uploads  
/athlete_events .csv"  
INTO TABLE jo.athletes  
FIELDS TERMINATED BY ","  
ENCLOSED BY ""  
LINES TERMINATED BY "\r\n"  
IGNORE 1 LINES  
(ID,Name,Sex,Age,Height,Weight,Team,NOC, Games, Sport, Event, Medal);
```

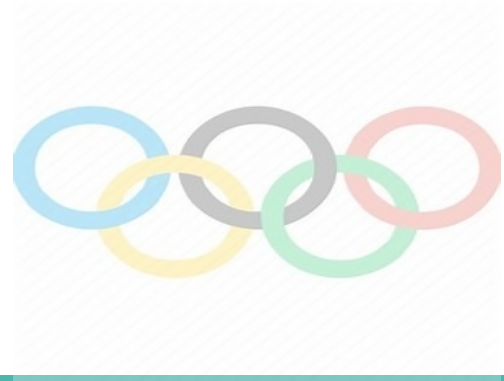


Les relations



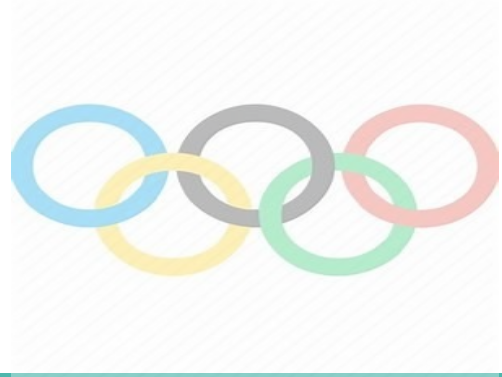
Librairies utilisées

- Dyplyr de R
- Pandas
- Pyplot de Matplotlib
- PyMySQL
- SQLAlchemy
- Seaborn



Importation des librairies

- > import sqlalchemy as db
- > import pandas as pd
- > import pymysql
- > import seaborn as sns
- > from matplotlib import pyplot as plt



Etablir une connection entre le serveur SQL et Python

sqlalchemy:

```
> connection =  
db.create_engine('mysql+pymysql://[user_name]:[password]@localhost/db').connect()
```

Transformation des tables en DF Pandas :

```
> athlete = pd.read_sql('SELECT * FROM jo.athletes', con = connection, index_col  
= ["auto_id"])
```



Pistes d'améliorations du projet

La partie concernant les sports et les disciplines n'a pas du tout été abordée dans ce projet et permettrait d'apporter de nouvelles analyses ou d'affiner celles déjà effectuées.

La popularité des sports ou de certaines disciplines pourrait ainsi être étudiée mais cela pourrait également permettre de mieux comprendre l'évolution de la taille et du poids au fil des années (nouveaux sports admis qui font augmenter la moyenne ?).

