



HAUTES ÉTUDES EN INFORMATIQUE

4PDS – Projet Python Data Science

M1 FEDE – AIBD

Sujet de projet – Matches de football

2023-2024

Version 1.0
Dernière mise à jour 09/11/2023
Utilisateur : Étudiants
Auteur : Samuel CUELLA

Sommaire

I.	Introduction	1
II.	Import des données	2
III.	Historique des transferts	2
IV.	Requêtes	2
A.	Clubs	2
B.	Joueurs.....	3
C.	Matches	3
V.	Analyse et visualisation graphique.....	3
A.	Résultats en fonction du club	3
B.	Analyse des performances du joueur Messi	3
C.	Analyse des statistiques des clubs Liverpool et Arsenal	3
D.	Corrélation entre match à domicile et victoire.....	4
E.	Paris sportif	4
F.	Corrélation entre formation et victoire	4
VI.	Conditions	4
VII.	Deadline	5
VIII.	Barème	5

I. INTRODUCTION

Ce sujet est accompagné d'un jeu de données sur un ensemble de match de football, au format json. Chaque fichier représente un match et se présente sous la forme suivante :

```
{
  "id": "match_1059706",
  "type": "match",
  "dateMatch": "2019-08-11T13:00:00Z",
  "Home": {
    "id": "13",
    "club": "Leicester",
    "players": {
      "15033": {
        "stat": {},
        "info": {
          "idteam": "13",
          "idplayer": "15033",
          "lastname": "Morgan",
          "position": "Defender",
          "formation_place": 0,
          "formation_used": "4141",
          "sub": 1,
          "goals": 0,
          "own_goals": 0,
          "yellow_card": 0,
          "red_card": 0,
          "mins_played": 0
        }
      },
      "17745": {}
    }
  },
  "Away": {
    "id": "...",
    "players": {}
  },
  "quotationPreGame": {
    "home": "...",
    "Draw": "...",
    "away": "..."
  },
  "period": "FullTime",
  "matchTime": 95,
  "matchData": {
    "home": {
      "goals": [],
      "bookings": [],
      ...
    },
    "away": {
      "goals": [
        {
          "playerId": "90585",
          "type": "var",
          "time": "51",
          "side": "away",
          "lastname": "Boly"
        }
      ],
      "bookings": [
        {
          "playerId": "171317",
          "type": "yellow",
          "time": "60",
          "side": "away",
          "lastname": "R\u00faben Neves"
        },
        {
          "playerId": "114128",
          "type": "yellow",
          "time": "31",
          "side": "away",
          "lastname": "Jonny"
        }
      ],
      ...
    }
  }
}
```

II. IMPORT DES DONNEES

Afin de réaliser des requêtes, vous devez les importer dans une base de données SQL, ou des dataframes Pandas. Si vous choisissez d'utiliser Pandas, vous aurez un bonus.

Créez les tables suivantes (les clefs primaires sont en gras) :

- teams : **idteam**, name
- players : **playerid**, lastname
- matches : **matchid**, date, home_idteam, away_idteam, duration, period, championship, home_formation, away_formation, quotation_home, quotation_away, quotation_draw, home_score, away_score.
- highlights: **matchid**, **time**, playerid, type (ex: goal, carton,...)
- substitutions: **matchid**, **time**, off_playerid, on_playerid, reason
- match_players: **playerid**, **matchid**, team_id, position, formation_place, play_duration, final_mark_2015, quotation_player, player_stats (une colonne par champ: penalty_conceded, red_card, ...).

Écrivez un script python qui lit les fichiers du dataset et réalise l'import des données en base ou dans Pandas. Attention aux informations manquantes : Certains fichiers json n'ont pas de champ championship et/ou quotationPreGame. Certains joueurs n'ont pas de stats de match et les champs varient en fonction de la position du joueur (goal vs attaquant).

III. HISTORIQUE DES TRANSFERS

En utilisant les informations en base, créez et remplissez une table/dataframe permettant de tracer le passage des joueurs dans les différentes équipes. Cette table aura les colonnes : player_id, name, team, starting_date, end_date

IV. REQUETES

En utilisant SQL ou Pandas, effectuez les tâches suivantes :

A. Clubs

- Afficher la liste de tous les clubs (idteam et nom du club)
- Quel est le nom du club dont l'id est 43 ?
- Quel est le nombre total de clubs ?

B. Joueurs

- Afficher la liste des joueurs
- Quel est le nombre total de joueurs ?

C. Matches

- Afficher la liste des matchs joués
- Afficher tous les id des matchs où le club à domicile est Arsenal
- Afficher les information du/des match(s) qui se sont déroulés le 11-08-2019.
- Combien de matchs ont eu lieu en 2019 ?
- Combien de matchs a joué le club Liverpool ?

V. ANALYSE ET VISUALISATION GRAPHIQUE

A. Résultats en fonction du club

- Afficher le nombre de matchs par club pour le championnat 5 pour la saison 2019-2020 (dont victoire, défaite, math nul) (visualisation) Les dates du championnat 5 (Série A) pour la saison 2019-2020 sont 24 août 2019 - 2 août 2020
- Afficher les 10 meilleures équipes pour le championnat 2 de la saison 2020-2021 (celles avec le meilleur ratio victoire/matchs joués) Les dates du championnat 2 (Premier League) pour la saison 2020-2021 sont du 12 septembre 2020 au 23 mai 2021.

B. Analyse des performances du joueur Messi

- Quel est l'id de ce joueur ? (requête)
- A quel(s) club(s) appartient ce joueur ? (requête)
- Calculer les performances (note_final_2015) de ce joueur sur trois matchs consécutifs et afficher l'évolution en fonction du temps. (visualisation)
- Combien de cartons (rouge et jaune) a eu ce joueur ?
- Est-ce que ce joueur fait partie des meilleurs butteurs dans son championnat ?
- Pour cela, afficher les 10 meilleurs buteurs de son championnat (idplayer,nom, nombre de buts) (requête). (=> championnat = 3 et oui, il s'agit du meilleur butteur de son championnat)

C. Analyse des statistiques des clubs Liverpool et Arsenal

- Quels sont les team_id des clubs Liverpool et Arsenal
- Afficher la liste des joueurs du club Arsenal (id, nom)

- Est-ce qu'il existe un match opposant Liverpool et Arsenal ? Si oui, quel est le score de ce(s) match(s) ? Et afficher les joueurs (id, position) du club Liverpool d'un de ses matchs.
- Afficher l'évolution des notes (note_final_2015) moyennes des joueurs par position (defender, Midfielder...) en fonction du temps (visualisation) du club Liverpool. On pourra lier les postes forward et striker.
- Réaliser un graphique de l'évolution des écarts de résultats des clubs Liverpool et Arsenal (sur tous les matchs) en fonctions du temps (par ex : une victoire avec un score 3-2 correspondra à un score de 1 ; une défaite avec un score 1-4 correspondra à un score -3 ; un match nul correspond à un score 0).
- Quels est le nombre moyen de changement par match pour le club Arsenal ?

D. Corrélation entre match à domicile et victoire

Existe-t-il une corrélation entre match à domicile et victoire ? Afficher le nombre de victoire/défaite en fonction des équipes à domicile (home) et invité (away). Est-ce qu'on observe une corrélation ?

E. Paris sportif

Quel club nous aurait-il rapporté le plus si nous avions parié 1€ sur sa victoire à chaque match ? A partir des « quotations » de pré-match et des résultats finaux, nous souhaitons répondre à cette question. Pour y répondre, il commencer par faut calculer le gain pour chaque résultat de match.

F. Corrélation entre formation et victoire

Nous allons maintenant analyser la formation et la victoire. L'objectif est de savoir s'il existe une corrélation entre un certain type de formation et la victoire. Réalisez un histogramme avec le taux de victoire en fonction de la formation mise en place en match. Analysez le résultat obtenu.

VI. CONDITIONS

Ce projet est un projet individuel.

Les éléments du rendu sont les suivants:

- Code python
- Code SQL (si vous choisissez SQL)
- Graphiques réalisés et réponses tirées de la base le cas échéant (lorsqu'un id, un nom, une liste, etc. sont demandés).
- Code/requêtes ayant permis l'extraction des données demandées.

VII. DEADLINE

Votre rendu doit intervenir avant la date limite fixée sur SCE.

VIII. BAREME

Item	Points
Import des données	2
Historique des transferts	1
Requêtes	7
Analyse et visualisation graphique	10
Bonus	10
Total	30/20