

# Projet Ligue 1 – météo

## Point d'avancement





# La météo influence-t-elle le nombre de buts marqués ?

- 1 Le projet et les ressources utilisées
- 2 La gestion de projet
- 3 L'avancement du projet
- 4 Points notables et axes d'amélioration

# Le projet

Déterminer l'impact de la météo sur la qualité d'un match, à des fins marketing (vente de sports publicitaires)

2 principales  
ressources

## Mission :

Etablir une représentation visuelle précise de l'évolution du nombre de buts marqués par équipe et par joueur, en fonction de la météo (pluviométrie, température)

## Champ d'observation :

Les matches de la Ligue 1 qui se sont déroulés du 22 août 2020 au 23 mai 2021



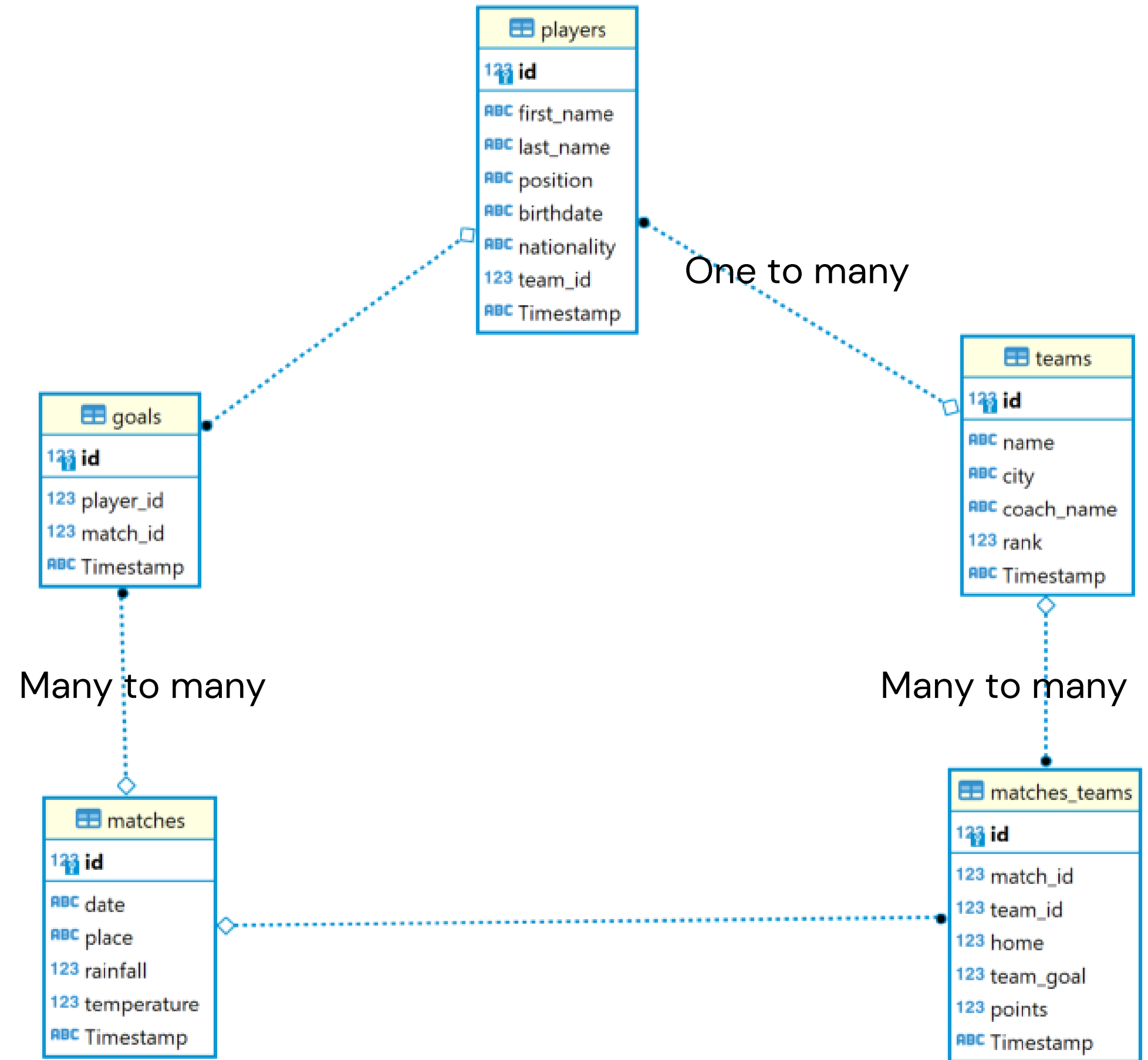
**Historique-Météo.net**  
quel temps faisait-il ce jour-là ?

# Le projet

De quelles données avons-nous besoin ?

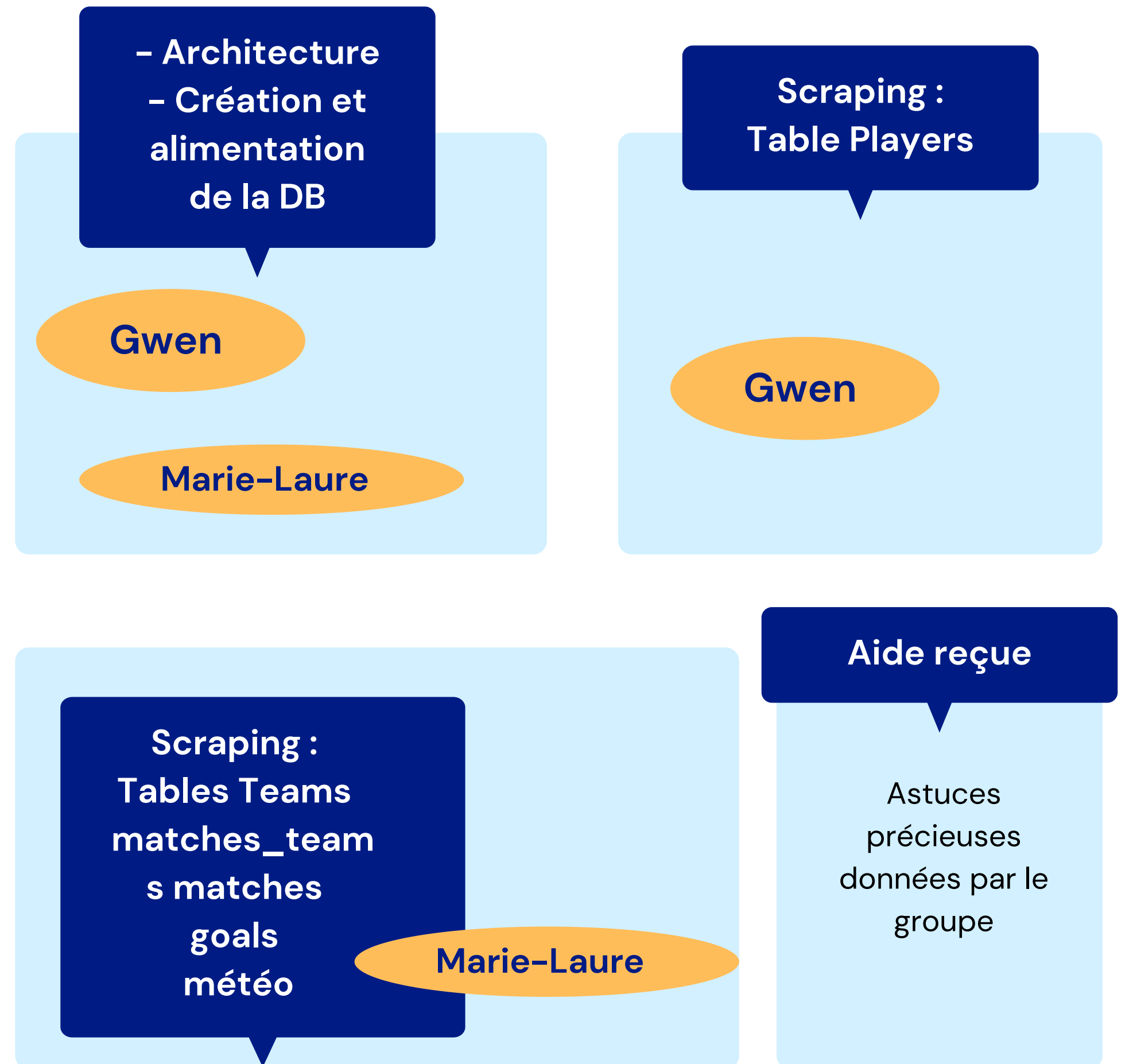
## Données nécessaires :

- Une équipe a plusieurs joueurs
- Une équipe a plusieurs rencontres
- Une rencontre est caractérisée, entre autres par sa pluviométrie et sa température
- Une rencontre a plusieurs buts
- Un joueur a plusieurs buts



# La gestion de projet

- 1 Définir l'architecture de la BD en fonction du besoin client
- 2 Concevoir et alimenter la base de données par des méthodes de scraping
- 3 A venir : statistiques et visualisation des données



# L'avancement du projet

**Les clubs de la Ligue 1 2021-2022**  
<https://www.lequipe.fr/Football/ligue-1/page-participants>



**Fiches club**  
<https://www.lequipe.fr/Football/FootballFicheClub9.html>

- Scraping à l'aide de la **méthode read\_html de pandas** (récupération sous forme de tableau)
- Dijon a dû être scrapé séparément car pas même structure de tableau
- Team\_ID : création de la clé étrangère - on relie le joueur à une équipe grâce à la city (récupéré grâce à la **bibliothèque BeautifulSoup et SQL**).
- Rajout des buteurs transférés en cours d'année (fichier csv)

players	
123	id
ABC	first_name
ABC	last_name
ABC	position
ABC	birthdate
ABC	nationality
123	team_id
ABC	Timestamp

2



**Fiches club**  
<https://www.lequipe.fr/Football/FootballFicheClub9.html>

1

teams	
123	id
ABC	name
ABC	city
ABC	coach_name
123	rank
ABC	Timestamp

- scraping nom du coach / city / club / classement : **méthode read\_html de pandas**
- city : bibliothèque **BeautifulSoup**



**Classement**  
<https://www.lequipe.fr/Football/ligue-1/saison-2020-2021/page-classement-equipes/general>

# L'avancement du projet

## Calendrier & résultats

<https://www.lequipe.fr/Football/ligue-1/saison-2020-2021/page-calendrier-resultats>

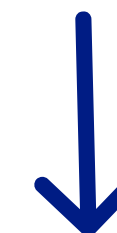
Extraction des liens (1 item par journée de championnat) grâce à **json**



Récupération des données avec **json** et sauvegarde dans un dataframe puis fichier .csv :  
dates / places / scores / domiciles / extérieurs / vainqueurs / scores\_exterieur / scores\_domicile

## Météo

<https://www.historique-meteo.net/france>



5

goals
123 id
123 player_id
123 match_id
ABC Timesta...

4

matches_teams
123 id
123 match_id
123 team_id
123 home
123 team_goal
123 points
ABC Timestamp

- Création de matches\_teams domicile en faisant : un boolean sur domicile / score en fonction de la colonne vainqueur / identification des buts domicile

- Création de matches\_teams extérieur en faisant : boolean sur domicile / score en fonction de la colonne vainqueur / identification des buts extérieur

- Scraping de la météo par jour et ville avec pd.read\_html de Pandas.

3

matches
123 id
ABC date
ABC place
123 rainfall
123 temperature
ABC Timestamp

# Récapitulatif des outils et techno utilisés



## Pour le scraping :

- méthode Pandas read\_html
- BeautifulSoup
- format json

## Pour la base de données :

- Création de la base de données dans le notebook à l'aide de Sqlite3 (connect / execute / commit)
- Alimentation de la base de données au fur et à mesure (on a utilisé la méthode zip pour alimenter la DB)
- Consultation et modification de la base de données avec la méthode read\_sql\_query de Pandas (pour vider les tables erronées lors des tests)



# Points notables

## Axes d'amélioration



- Mise à jour du site des pages classements/calendriers & résultats/clubs qui affichent désormais les informations de la saison 2021-2022

Solutions : enregistrement des données au préalable sous format .csv des tables teams et matches / changement url

- Météo : difficulté à scraper les url car les régions non présentes et problème avec les températures et la pluviométrie (% , décalage dans les lignes, km)

- 12 joueurs plus présents suite à des transferts



- En fonction des résultats obtenus (probants ou non) avec un seul championnat, évaluer s'il est nécessaire d'enrichir la base de données avec d'autres championnats (optionnel). Mesurer le coût et le temps passé.



Difficultés :

- création des clés étrangères
- limitation dans le code python
- tâche énorme pour des débutants