

Amjad Zoubari  
Master 2 MIAHS  
parcours Web Analyst



## **Projet de fin d'études des Web-Analystes:**

*Comparer les statistiques de deux joueurs de la Ligue des champions saison 2020-2021 à partir du site sofascore.*





















Année universitaire 2020-2021

## Sommaire

I) Les objectifs du projet	3
II) Le déroulement de sa réalisation	4
III) La structure du code source	6
IV) L'adéquation du projet avec les objectifs d'évaluation	7
Remerciements	10

## 1) Les objectifs du projet

Avant de réfléchir au sujet sur lequel j'allais travailler, je me suis d'abord intéressé aux objectifs qui étaient attendus. Le choix du sujet a peut-être été pour moi l'étape la plus difficile car j'ai hésité entre plusieurs sujets à savoir l'un qui porterait sur la musique et l'autre sur le football. Mes échanges avec les enseignants du parcours m'ont permis de faire un choix et j'ai donc décidé que mon projet porterait sur la comparaison entre deux joueurs de la Ligue des champions grâce à leurs statistiques pour la saison 2020-2021 à partir du site Sofascore. Sofascore<sup>1</sup> est une application de sport qui nous permet d'obtenir en direct les scores, les résultats des matchs et les statistiques des joueurs et des équipes.

PLAYER STATISTICS								
SUMMARY ATTACK DEFENCE PASSING GOALKEEPER DETAILED								
Accumulation		All ▾	All players Players with minimum appearances ⓘ					
#	Team	Name	Goals	Succ. dribbles	Tackles	Assists	Accurate passes %	Rating ▾
1		Ruslan Neshcheret	0	0	0	0	90.91	9.70
2		Gianluigi Buffon	0	0	0	0	85.71	8.50
3		Joshua Kimmich	1	2	9	3	87.25	8.47
4		Lionel Messi	3	23	3	2	89.38	8.45
5		Kevin De Bruyne	0	1	5	4	85	8.40
6		Neymar	6	21	3	0	81.44	8.08
7		Erling Braut Haaland	6	6	0	0	74.63	8.08
8		Ciro Immobile	5	4	1	1	81.13	7.98
9		Álvaro Morata	6	9	4	1	72.52	7.92
10		Manuel Neuer	0	0	0	0	85.71	7.84
11		Alex Sandro	0	5	9	1	85.34	7.83
12		Bruno Fernandes	4	3	9	1	73.97	7.80
13		Adrián	0	0	0	0	62.07	7.80
14		Cristian Romero	1	1	12	1	84.51	7.78
15		Marcos Acuña	0	7	7	1	83.52	7.70
16		Ousmane Dembélé	3	16	2	2	83.06	7.68
17		Reece James	0	1	7	1	86.14	7.65
18		Caoimhin Kelleher	0	0	0	0	67.09	7.65
19		Alassane Pléa	5	4	2	3	77.37	7.63
20		Toni Kroos	0	1	7	1	94.04	7.63
< Previous			1 2 ... 36			Next >		

<sup>1</sup> <https://www.sofascore.com>

Les objectifs de mon projet sont les suivants:

1. Scraper le site Sofascore en utilisant des api qui contiennent les informations cherchées en format JSON.
2. Récupérer ces données et les mettre dans une base de données sur un serveur à distance.
3. Faire un client (page php) qui nous laisse choisir le nom de deux joueurs. Ensuite, il communique avec cette base de données pour monter les statistiques de ces joueurs en mettant en avant le joueur qui a les meilleurs résultats. On peut également voir les 5 joueurs les plus recherchés sur le site et voir quelques statistiques propres à chacun des joueurs.

## II) Le déroulement de sa réalisation

### Etape 1:

Avant de pouvoir commencer à travailler, je me suis fait bloqué à de nombreuses fois sur le site quand j'ai essayé de le scraper. J'ai dû imiter autant que possible le header du site (*referer, host, origin*) et surtout, il fallait changer les *User-Agent* à chaque nouvelle requête. Pour cela, j'ai cherché dix *User-Agent* différents en ligne.

### Etape 2:

J'ai scrapé les pages de tous les joueurs participant à la Ligue des champions. Cela fait au total **37 pages** et **724** joueurs.

### **Etape 3:**

J'ai mis toutes les données récupérées dans une base de données SQL sur un serveur à distance. Ces deux étapes ont été réalisées en utilisant Python sur un notebook. Même si ces données changent après, il suffit de redémarrer ce code pour obtenir les nouvelles données.

### **Etape 4:**

J'ai créé une page PHP sur le même serveur avec la base de données car je n'avais accès qu'à un seul serveur. Cette page permet à l'utilisateur de sélectionner le nom de deux joueurs et de montrer le résultat de cette requête. L'utilisateur va avoir les informations suivantes: le nom de chaque joueur, leur équipe, les buts, les dribbles passés avec succès, les passes décisives et les passes précises. On peut désormais cliquer sur un bouton pour afficher un graphe qui montre l'évaluation de chaque joueur.

J'ai mis en vert le meilleur chiffre et en rouge le moins bon pour chacune des informations fournies. On a également la possibilité de cliquer sur un autre bouton pour ouvrir une nouvelle page PHP sur laquelle on trouve deux graphes avec des informations générales récupérées à partir de la base de données (top 5 des meilleurs butteurs et top 5 des joueurs ayant les meilleures évaluations).

### III) La structure du code source

Pour scrapper et exporter les données sur la base de données, j'ai utilisé un notebook que je vous ai communiqué. Concernant la page PHP, vous le trouverez sur mon github en cliquant sur le lien suivant: <https://github.com/Amjad-Z/ProjectVer2> et vous pourrez visiter la page à l'adresse suivante: <http://172.28.100.206/comparing.php>

## IV) L'adéquation du projet avec les objectifs d'évaluation

Comme pour les objectifs du projet, j'ai pris le temps de lire et de bien interpréter les critères d'évaluation afin d'éviter le hors sujet. Pour rappel, les critères d'évaluation sont les suivants:

- **La qualité du rendu final:**

J'ai réussi à atteindre mon objectif principal qui était de comparer les statistiques sportifs de deux joueurs de la Ligue des champions. J'ai également pu faire des graphes en utilisant la librairie **chart.js** pour que les résultats soient plus lisibles et plus esthétiques. Cependant, j'aurais voulu rendre la page plus esthétique en utilisant plus de CSS.

- **La qualité du code source (commentaires, structure du code, ect...)**

Côté Scrapping et exportation de la base de données, je pense que le code est pratique et s'adapte facilement aux éventuels changements. Pour ce qui est du PHP, j'aurais voulu mieux structurer la page mais j'ai décidé d'opter pour une fonctionnalité simple mais qui fonctionne parfaitement.

- **L'originalité du projet**

Le site Sofascore ne propose pas ce service, c'est pourquoi je considère que mon projet est intéressant et pratique pour toute personne qui s'intéresse aux joueurs de football et à leurs statistiques.

Avant, il fallait aller sur chaque joueur pour voir ses performances et les résultats de son équipe. Ce que je propose, permet d'avoir sur une seule page les résultats recherchés. J'ai choisi de comparer uniquement deux joueurs mais on aurait très bien pu comparer 3 ou plusieurs joueurs. C'est un projet qui peut encore être développé et perfectible.

- **La qualité des analyses**

En mettant en couleur les résultats du joueur qui a les meilleurs scores et en ajoutant les graphes JS, cela donne une rapide idée quant à la comparaison entre les deux joueurs choisis. Cela est un gain de temps non négligeable et facilite la lecture des résultats.

- **L'adéquation avec les objectifs d'évaluations et les remarques**

Suite à mon oral portant sur la première version de mon projet de fin d'études, j'ai pris en considération les remarques qui m'ont été faites afin de répondre aux objectifs d'évaluation.



- Les améliorations effectuées:

- 1) Côté base de données, j'ai ajouté un champ (*searched*) qui permet de traquer le nombre de fois qu'un joueur est recherché.

Voici le schéma de cette base de données

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/> 1	<b>pid</b>	int			No	None			Change  Drop  More
<input type="checkbox"/> 2	<b>name</b>	varchar(255)	utf8_general_ci		No	None			Change  Drop  More
<input type="checkbox"/> 3	<b>team</b>	varchar(255)	utf8_general_ci		No	None			Change  Drop  More
<input type="checkbox"/> 4	<b>goals</b>	int			No	None			Change  Drop  More
<input type="checkbox"/> 5	<b>succ_dribb</b>	int			No	None			Change  Drop  More
<input type="checkbox"/> 6	<b>tackles</b>	int			No	None			Change  Drop  More
<input type="checkbox"/> 7	<b>assists</b>	int			No	None			Change  Drop  More
<input type="checkbox"/> 8	<b>acc_passes</b>	float			No	None			Change  Drop  More
<input type="checkbox"/> 9	<b>rating</b>	float			No	None			Change  Drop  More
<input type="checkbox"/> 10	<b>searched</b>	int			No	0			Change  Drop  More

- 2) J'ai ajouté une fonction AJAX qui envoie une requête toutes les seconds à la base de données pour récupérer les 5 joueurs les plus recherchés et les classer en fonction du nombre de recherches effectuées pour chacun des joueurs.
- 3) J'ai utilisé la librairie chart.js pour réaliser un graphe montrant l'évaluation globale des deux joueurs sélectionnés.
- 4) Toujours en utilisant la librairie chart.js, j'ai créé deux graphes sur une nouvelle page (top 5.php) qui listent le top 5 des joueurs par rapport à leurs évaluations et le top 5 des buteurs. Cette fois-ci, j'ai récupéré les données en utilisant le format JSON.

## Remerciements

Tout d'abord, je tiens à remercier M. Louis Bigo, Mme Petra Rahme, M. Dorian Baudry, M. Mohamed Elati et bien sûr M. Charles Paperman pour leur suivi, leur écoute et leurs conseils mais surtout de m'avoir offert la possibilité d'améliorer mon projet de fin d'études.

Ensuite, merci à mes camarades de m'avoir aidé à surmonter certaines difficultés et de m'avoir apporté tout leur soutien.