

# RAPPORT DE PROJET PYTHON



Léo-Paul Houitte  
07/01/24



**IUT de Paris - Rives de Seine**  
Université Paris Cité

## INTRODUCTION :

Ce projet prend racine dans ma passion pour le football, un sport universel qui suscite un intérêt mondial. En tant qu'amateur et supporter, l'idée de fusionner cette affection avec une analyse approfondie des données se présente comme une opportunité stimulante. Je suis convaincu qu'en travaillant sur un sujet qui m'inspire et dont j'ai beaucoup de connaissances donne un rendu intéressant. L'objectif est d'explorer les données de la première division de football français, la Ligue 1. Cette base de données regroupe tous les matchs disputés en Ligue 1 de 1999 à 2019, ce qui permet d'extraire des statistiques significatifs sur les performances des équipes.

Comment l'analyse des performances des clubs de ligue 1 de 1999 à 2019, peut-elle se révéler être un outil stratégique pour les dirigeants d'un club récemment promu en Ligue 1 ? En d'autres termes, comment cette analyse peut-elle guider les décisions tactiques et opérationnelles, permettant au club promu de garantir une adaptation réussie en optimisant ses chances de maintien et en rivalisant avec les clubs établis de la compétition ?

## LA SOURCE DU JEU DE DONNEES

J'ai récupéré la base de données « Ligue 1 Championship » sur Kaggle. Celui qui l'a partagée explique que les informations ont été extraites du site :

<https://www.football-data.co.uk/data.php>

Ce jeu de données est simple car je trouve que les variables sont faciles à comprendre et il n'y a aucune valeur manquante. Il contient chaque match de Ligue 1, de la saison 1999/00 à la saison 2018/19. Chaque ligne représente un match, et les colonnes sont les suivantes :

- **Season (Saison)** : Indique la saison durant laquelle le match a eu lieu.
- **Home team (Équipe à domicile)** : Désigne l'équipe qui joue sur son terrain.
- **Away team (Équipe à l'extérieur)** : Désigne l'équipe qui joue à l'extérieur.
- **Home team goals (Buts marqués par l'équipe à domicile)** : Indique le nombre de buts marqués par l'équipe à domicile dans le match.
- **Away team goals (Buts marqués par l'équipe à l'extérieur)** : Indique le nombre de buts marqués par l'équipe à l'extérieur dans le match.
- **Score (Résultat)** : Affiche le score final du match en combinant les buts marqués par les équipes à domicile et à l'extérieur.
- **Winner (Vainqueur)** : Identifie l'équipe gagnante du match.
- **Loser (Perdant)** : Identifie l'équipe perdante du match.

## ANALYSE DES DONNEES :

On commence par un résumé statistique des 3 variables numériques :

Variable	Moyenne	Mediane	Ecart Type
Home Team Goals	1.42	1	1.2
Away Team Goals	0.99	1	1.04
Score	2.41	2	1.57

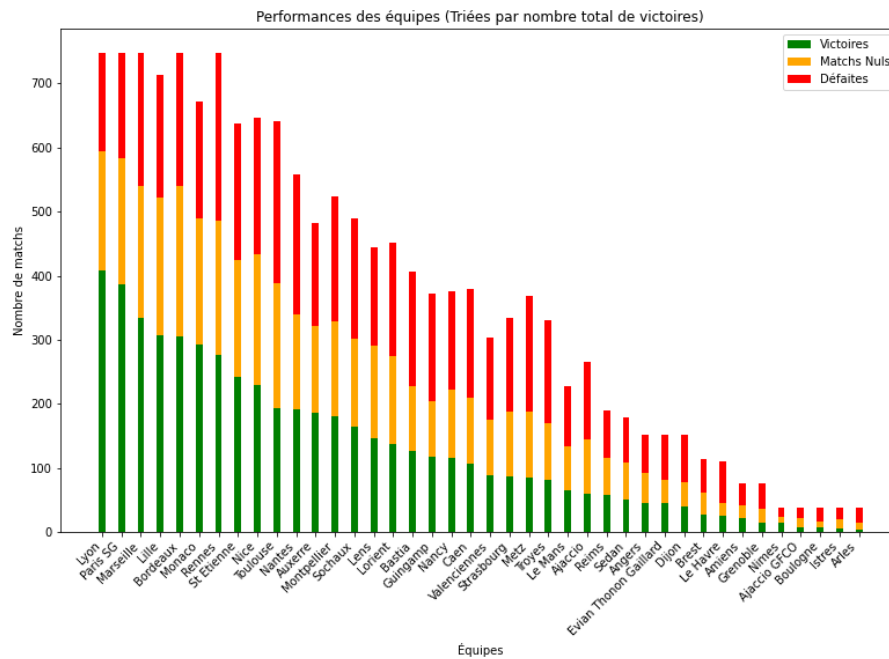
On observe que le nombre moyen de buts marqués par les équipes à domicile est 43% supérieur aux nombres moyens de buts marqués par les équipes à l'extérieur.

Tableau croisé entre 'Home Team Goals' et 'Away Team Goals' exprimé en % :

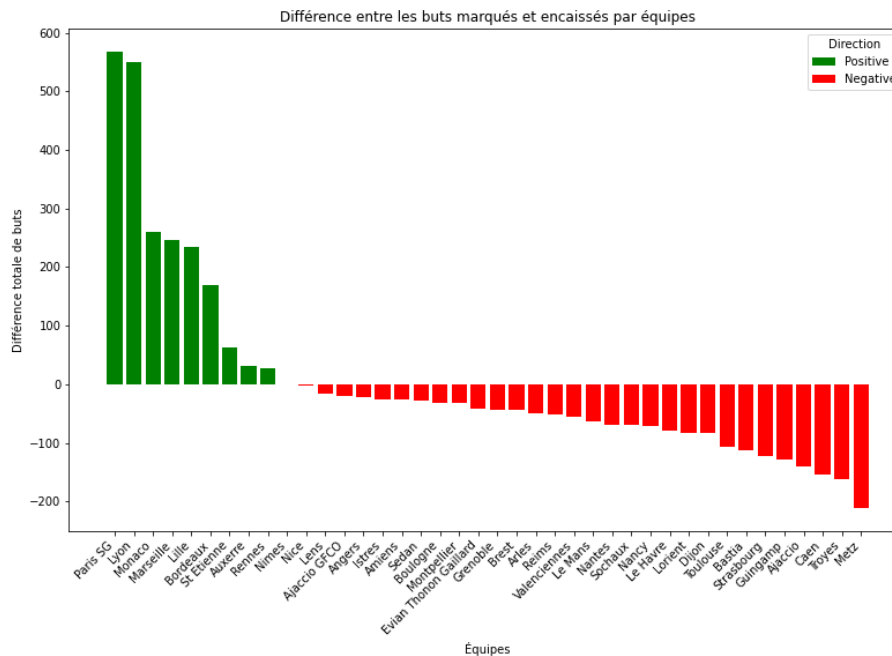
AWAY TEAM GOALS						
HOME TEAM GOALS		0	1	2	3	4
		Total				
0		10	8	4	2	1
1		13	13	6	2	1
2		9	8	5	1	0
3		4	4	2	1	0
4		2	1	1	0	0
5		1	0	0	0	0
Total		39	35	18	6	2

Ce tableau nous donne un aperçu du pourcentage des scores, par exemple les 3 scores les plus probables sont : 1-1 et 1-0 (13%) et 0-0 (10%)

## Graphique en bar empilés sur les performances des équipes :



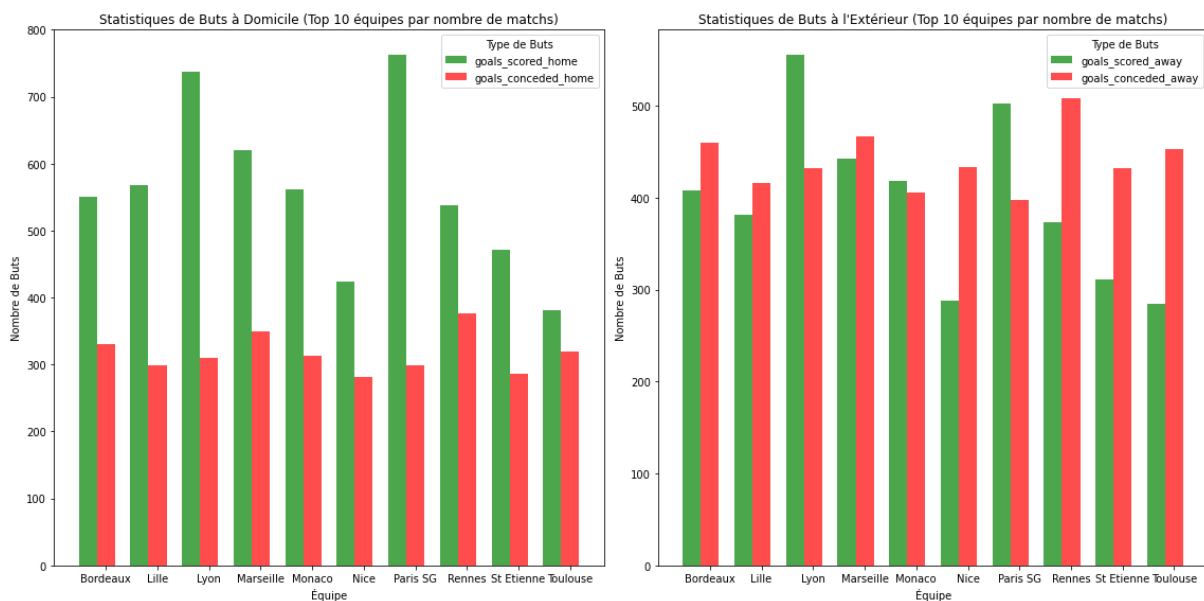
## Graphique sur le Goal average des équipes :



On remarque que seulement 9 équipes ont un goal average positif, soit 22,5 %.

Ces 2 graphiques nous permettent de voir que globalement Paris et Lyon surdominent le championnat, car ce sont les 2 meilleurs à chaque fois, et ont des stats impressionnantes.

## Double graphique sur la différence entre les buts marqués et concédés à domicile et à l'extérieur :

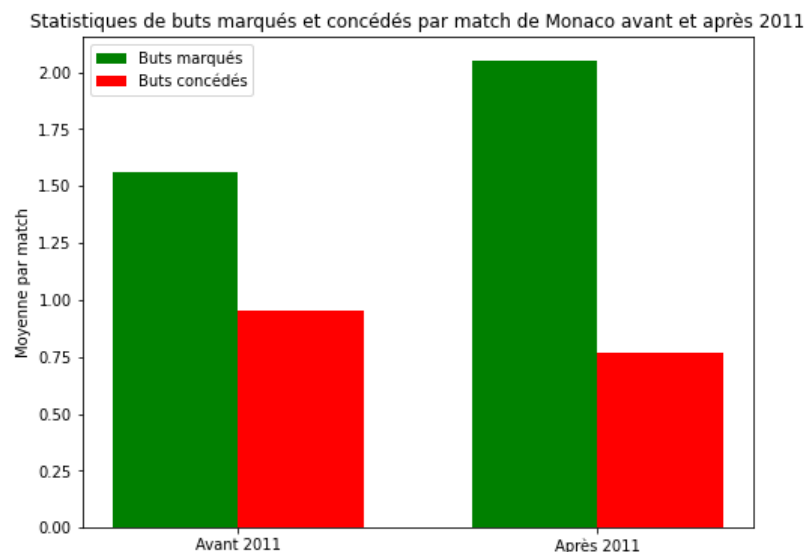
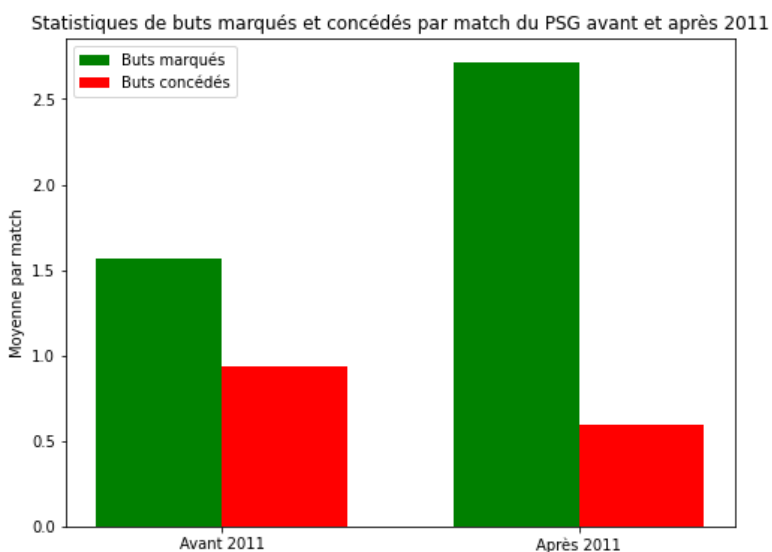


Pour une meilleure lisibilité, j'ai sélectionné uniquement les 10 équipes ayant joué le plus de matchs. A gauche, on observe les stats de buts quand l'équipe joue à domicile et à droite les stats de buts quand l'équipe joue à l'extérieur. Ce qui est très intéressant c'est qu'à domicile toutes les équipes marquent plus de buts qu'ils en encaissent alors qu'à l'extérieur, seulement 3 équipes réussissent cela. Plus globalement, on observe une vraie différence des statistiques en fonction du lieu où le match se joue et cela a un vrai impact sur les performances.

J'ai voulu vérifier si cette tendance se manifestait également pour les autres équipes, et voici ma conclusion : Sur les 394 saisons par équipes de Ligue 1, 22, soit environ 5,58%, ont affiché une performance à l'extérieur supérieure que celle à domicile au cours d'une saison. Cela confirme que la grande majorité des équipes performant mieux à domicile. Les facteurs peuvent être nombreux mais les principaux sont : l'influence positive du soutien des supporters, la familiarité avec le terrain et la fatigue du voyage pour l'équipe extérieur.

Il est à noter que Rennes a eu de meilleures performances à l'extérieur qu'à domicile lors de 4 saisons sur 5, de 2012 à 2016. Ce qui est une statistique extrêmement rare et importante à souligner, cependant je n'ai pas les données nécessaires pour pouvoir analyser grâce à quoi cela est dû.

On s'intéresse maintenant à l'impact positif du rachat d'un club, j'ai pris ici l'exemple du PSG et de Monaco racheté en 2011 par Qatar Sports Investments et par Dimitri Rybolovlev respectivement, ces 2 actionnaires sont milliardaires et ont dépensés des sommes astronomiques, ce sont les 2 plus gros rachats de l'histoire du football français, j'ai donc trouvé cela intéressant de **comparer les statistiques des équipes avant et après dans les 2 graphiques suivants** :



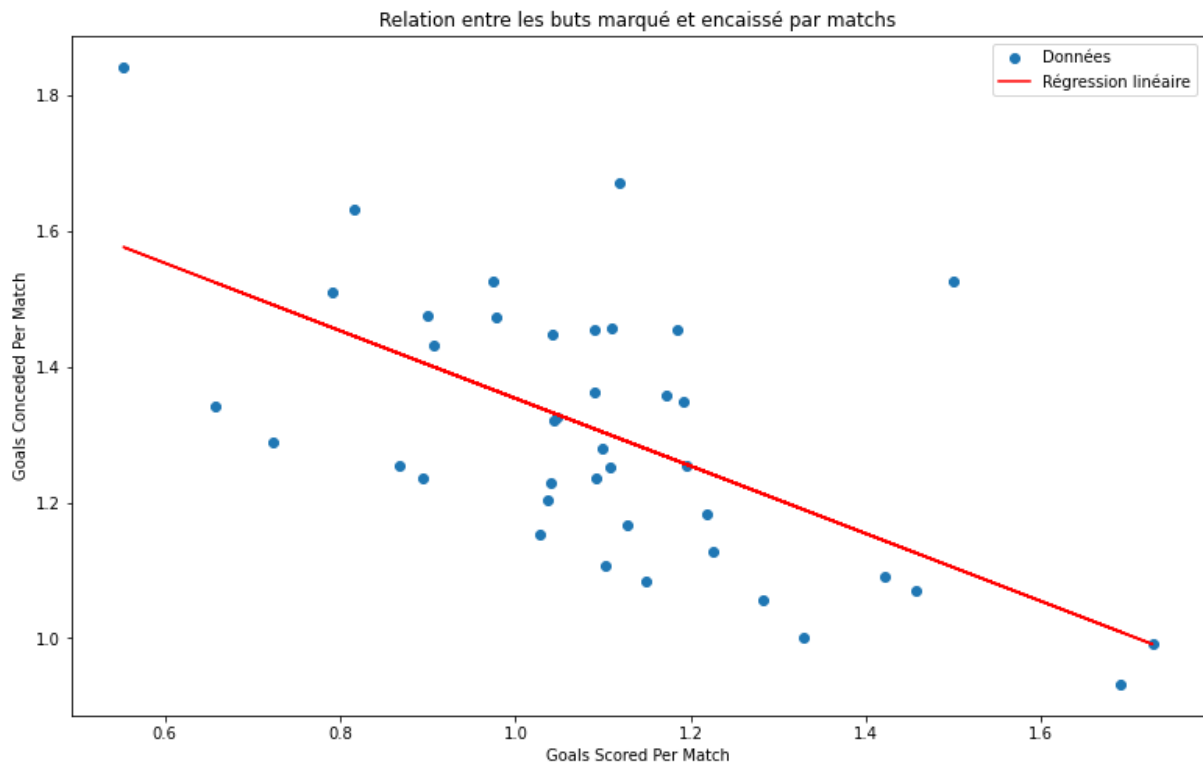
Premièrement pour le PSG, on observe une augmentation de 73% pour les buts marqués et une diminution de 36% des buts concédés.

Deuxièmement, Monaco affiche une augmentation de 31% des buts marqués, et une diminution de 19% des buts concédés.

Donc on observe grâce à ces 2 graphiques que le rachat d'un club, lorsqu'un énorme budget est injecté, que cela augmente forcément les performances d'une équipe car elle possède de meilleurs joueurs et peut investir dans de meilleures infrastructures.

## Corrélation entre les variables :

Nous commençons par étudier la corrélation entre les variables buts marqués par matchs et buts concédés par matchs dans le nuage de points ci-dessous :

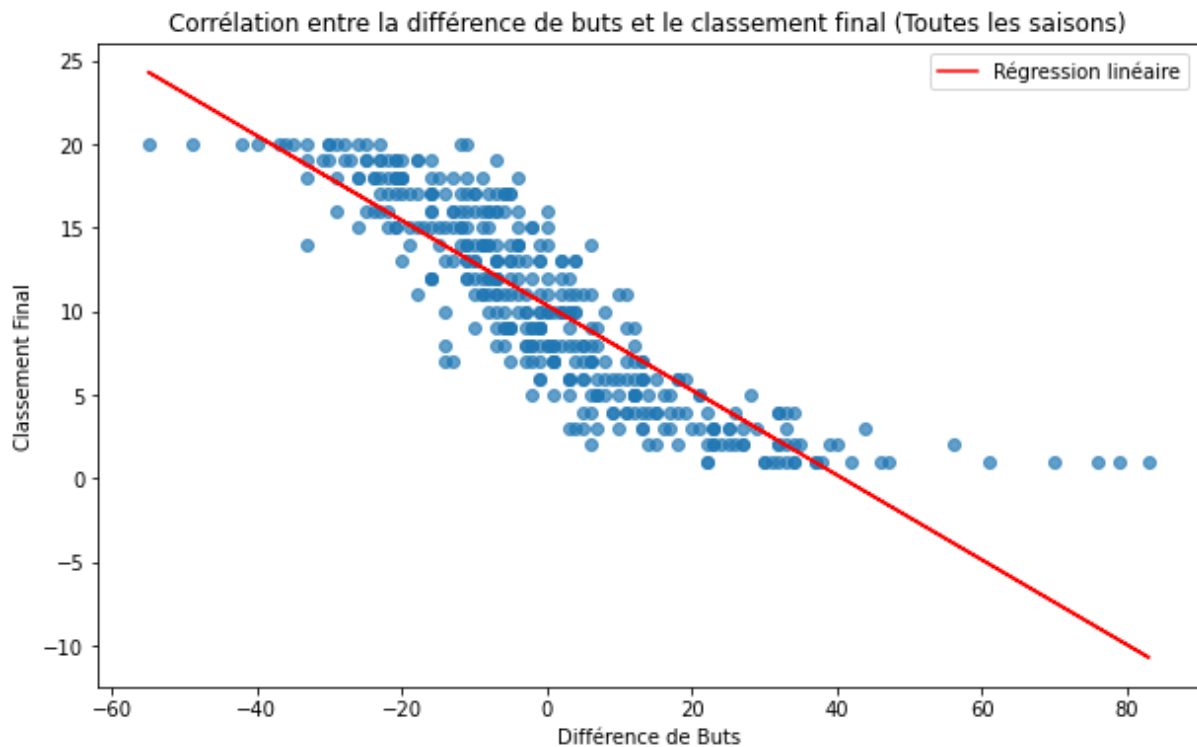


Corrélation entre Goals Scored Per Match et Goals Conceded Per Match :  
-0.6107569634530029

La corrélation de -0,61 suggère une relation significative entre les deux variables. Cette observation montre que les équipes performantes en attaque peuvent également démontrer une solidité défensive, et vice versa. Sur le graphique ci-dessus, il y a 40 données, ce sont toutes les équipes présente dans le jeu de données.

Pour réaliser le graphique ci-dessus, j'ai dû faire une boucle qui créer 20 dataframe, 1 pour chaque saison, qui répertorie les performances des équipes et en fonction du nombre de victoires et matchs nul calcule leurs classements finaux. Cela m'a permis de faire un graphique très intéressant et complet car il contient 394 données, ce sont toutes les saisons de toutes les équipes. On voit que le graphique sur l'axe du classement final va de -10 à 25, évidemment un classement de 0 à -10 ne correspond à rien mais c'est pour pouvoir visualiser la courbe de régression linéaire plus largement.

**Corrélation entre les variables différence de buts et classement final dans le nuage de points ci-dessous :**



**Corrélation entre Goals Difference et Ranking : -0.86**

La corrélation de -0,86 montrant qu'une meilleure différence de buts est fortement associée à un meilleur classement final. Cela suggère que les équipes qui marquent plus de buts que leurs adversaires ont tendance à obtenir un meilleur classement.

## **ANALYSE DES EQUIPES PROMUES :**

Dans notre étude, nous voulons aider une équipe promue à se maintenir en ligue 1, j'ai donc calculé le nombre d'équipes promues qui redescendent en ligue 2, la même saison ou la saison d'après :

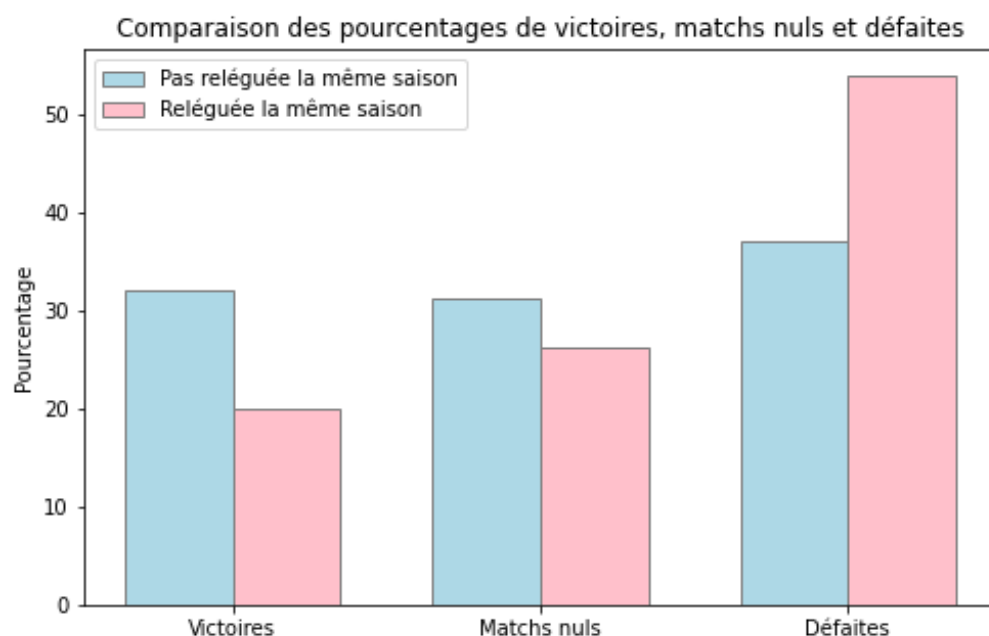
Pourcentage d'équipes promues qui ont été reléguées la même saison : 36.8%

Pourcentage d'équipes promues qui ont été reléguées la même saison ou la saison suivante : 59.6%

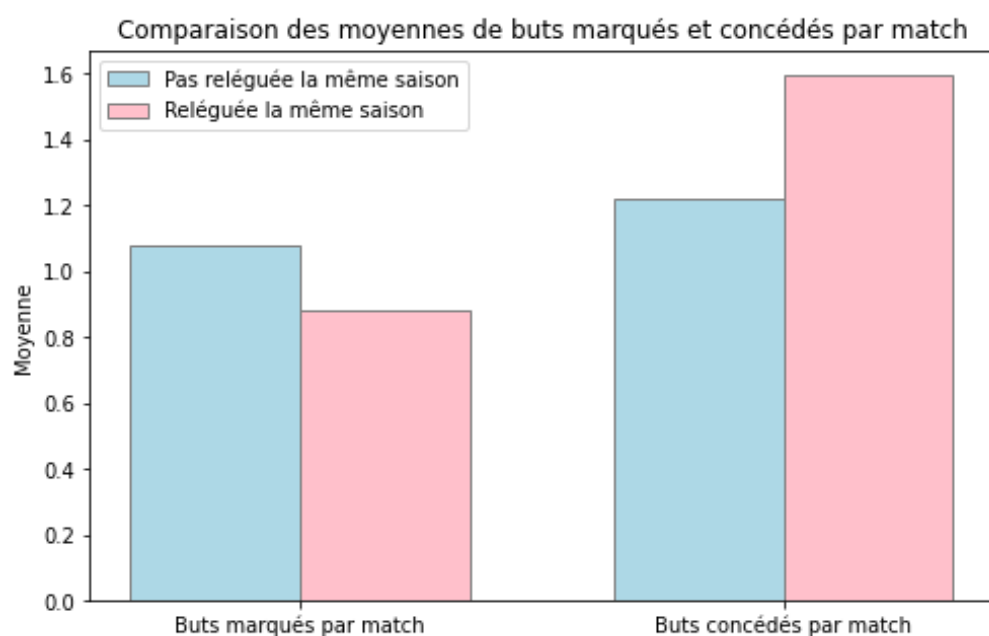
Ces pourcentages sont donc assez élevés car on se rend compte que plus d'un tiers des équipes promues, redescendent directement en Ligue 2. Et plus de la moitié sont reléguées avec une saison en plus, d'où l'importance d'étudier les statistiques pour ces équipes qui ne veulent surtout pas redescendre en Ligue 2.



Dans les 2 graphiques en bar ci-dessous, j'ai sélectionné uniquement les équipes promues, et j'ai ensuite trier en deux entre les équipes reléguées directement et les équipes qui ont réussis à se maintenir. Et cela permet de comparer les performances entre ces deux camps :



On remarque qu'il y a une vraie différence, notamment pour les défaites ou les équipes relégués ont une évolution de 45,9% de défaites en plus que celles pas relégués.



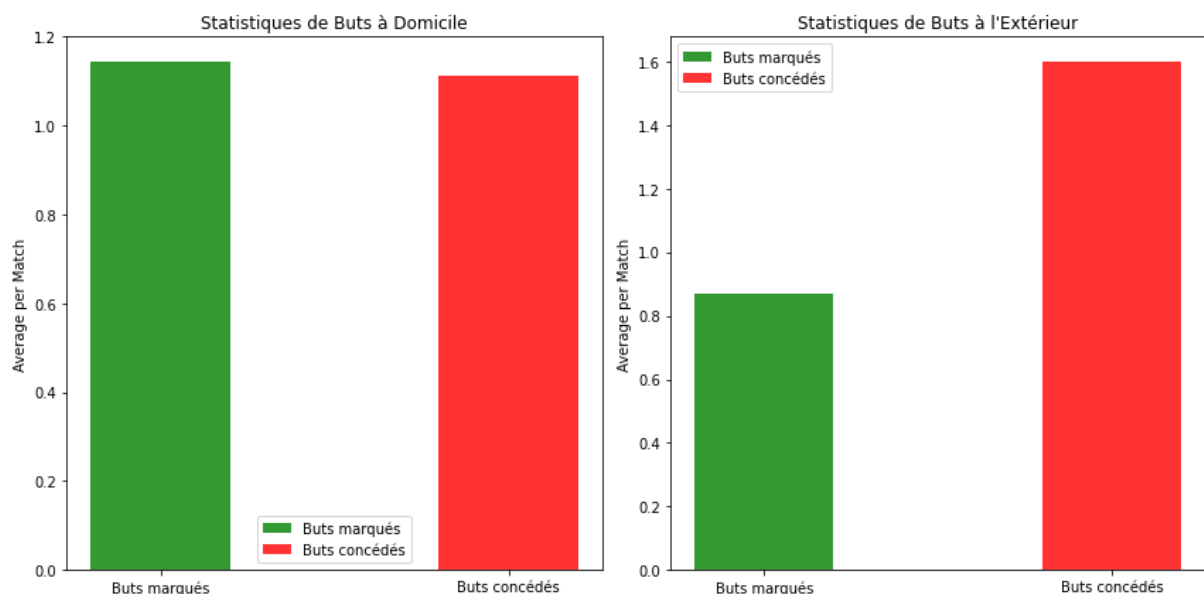
Pour les buts marqués et concédés, l'écart est encore flagrant avec des performances moins bonnes à chaque fois pour les équipes reléguées.

Ces 2 graphiques permettent à une équipe de visualiser facilement les performances à suivre tout au long de la saison pour ne pas être relégué.

En complément, **voici également les statistiques minimums auquel une équipe doit prétendre pour rester en Ligue 1**, j'ai donc récupéré les stats des équipes placés juste avant la zone de relégation et voici les moyennes des variables les plus intéressantes :

victories	9.3
draws	12
Goals Scored Per Match	1
Goals Conceded Per Match	1.4
Points	40
Goals Difference	-12.4

Enfin, le dernier graphique nous permet de visualiser la différence de performances entre les matchs à domicile et à l'extérieur pour les équipes promues face à toutes les équipes du championnat. C'est pour ça qu'il y a des statistiques différentes, si on comparait les équipes promues contre les autres équipes promues uniquement, bah il n'y aurait aucunes différences entre les statistiques à domicile et à l'extérieur :



Ce graphique en bar nous montre donc qu'à domicile les équipes promues ont une différence de buts légèrement supérieur, alors qu'à l'extérieur, ils encaissent deux fois plus de buts qu'ils n'en marquent. Ce qui semble logique, car c'est que nous observons depuis le début de l'étude : les équipes sont beaucoup plus performantes à domicile qu'à l'extérieur

Je me suis alors posé la question, qui s'en sort le mieux lors des matchs à domicile et à l'extérieur entre toutes les équipes et les équipes promues, cela pourrait être intéressant de voir si un certain type d'équipes ont plus de facilité ou difficulté en fonction du lieu.

Pour calculer les performances, nous utilisons les moyennes des buts marqués et des buts encaissés à domicile et à l'extérieur. La formule pour les performances à domicile et à l'extérieur est la suivante :

Buts marqués par match - Buts encaissés par match : goal average par match

- Performances à domicile pour toutes les équipes :  $0,65 - 0,54 = 0,10$
- Performances à domicile pour les équipes promues :  $1,14 - 1,11 = 0,03$
- Différence de performances à domicile :

$0,10 - 0,03 = 0,07$  de goal average en plus par match pour toutes les équipes à domicile

Nous remarquons que les équipes promues marquent en moyenne presque deux fois plus que toutes les équipes, mais comme elles encaissent également deux fois plus, elles ont finalement un goal average inférieur.

- Performances à l'extérieur pour toutes les équipes :  $0,45 - 0,76 = -0,31$
- Performances à l'extérieur pour les équipes promues :  $0,87 - 1,60 = -0,73$
- Différence de performances à l'extérieur :

$-0,31 - (-0,73) = 0,42$  de goals average en plus par match pour toutes les équipes à l'extérieur

Nous remarquons encore une fois que les équipes promues marquent en moyenne plus que toutes les équipes avec une augmentation de 93 %, mais en ce qui concerne les buts encaissés, elles encaissent 111 % de plus, ce qui leur donne un goal average encore inférieur.

En analysant ces résultats, nous constatons qu'en moyenne, toutes les équipes affichent de meilleures performances tant à domicile qu'à l'extérieur par rapport aux équipes promues. La différence est particulièrement marquée à l'extérieur, avec une performance moyenne supérieure de 0,42 but par match par rapport aux équipes promues, tandis qu'à domicile, l'écart est plus modéré, avec une performance moyenne supérieure de 0,07 but par match. Les dirigeants des clubs promus peuvent tirer des enseignements essentiels de cette analyse pour mieux appréhender les spécificités des performances à domicile et à l'extérieur. En se focalisant sur les aspects identifiés, tels que le renforcement des performances à l'extérieur, les clubs promus peuvent accroître leurs chances de maintien.

## **CONCLUSION :**

Ce projet, en analysant les performances des clubs de Ligue 1 de 1999 à 2019, a mis en lumière des aspects cruciaux du football français. L'étude des buts marqués à domicile et à l'extérieur a révélé l'influence significative du terrain local. Les variations dans les performances des équipes, accentuées par les rachats de clubs, ont été clairement démontrées à travers des analyses statistiques détaillées. Cela offre une vision précieuse de l'évolution des clubs au fil des ans et de l'impact de ces changements sur leurs résultats.

Une limite notable de ce projet réside dans la nature des données utilisées. La base de données se concentre exclusivement sur les scores des matchs, omettant des facteurs potentiellement influents comme le budget des clubs, les transferts de joueurs, ou d'autres éléments financiers et organisationnels. Ces facteurs, bien qu'importants, n'ont pas pu être inclus dans l'analyse en raison de leur absence dans les données disponibles. Leur intégration aurait pu fournir une compréhension encore plus riche et nuancée des dynamiques au sein de la Ligue 1.

Néanmoins, les résultats obtenus fournissent des insights précieux sur les stratégies et performances des clubs. Les clubs nouvellement promus peuvent en tirer des enseignements importants pour leur développement et leur adaptation dans l'élite du football français. Les résultats mettent également en évidence l'importance des décisions stratégiques et des investissements pour le succès à long terme.