

# Football Statistique

CHIRINA Rania, Gonzalez Emmanuel, AKHLOUFI Khalil, Martin Samuel,  
VITOFFODJI Adjimon

Département MIASHS, UFR 6 Informatique, Mathématique et Statistique Université Paul  
Valéry, Montpellier 3

2024



- ① Problématique et objectifs de l'étude
- ② Méthodologie
- ③ Présentation des résultats

- ❶ Problématique et objectifs de l'étude

## Problème posé

- L'analyse statistique des performances des joueurs de football est devenue un outil incontournable pour les entraîneurs, les analystes et les passionnés du ballon rond.
- Des joueurs emblématiques tels que Lionel Messi et Cristiano Ronaldo ont continué à dominer les classements statistiques, tandis que de nouveaux talents ont émergé pour laisser leur empreinte dans l'histoire du football.
- Cependant, malgré l'abondance de données disponibles, il reste des questions cruciales à explorer pour mieux comprendre les facteurs qui influent sur la performance des joueurs et pour anticiper les tendances à venir.

## Problème posé

- Comme l'a souligné Johan Cruyff, légendaire joueur et entraîneur de football, "Les statistiques sont comme un bikini : elles montrent beaucoup de choses, mais elles cachent l'essentiel."
- Cette citation emblématique résume parfaitement la complexité des chiffres dans le contexte du football. Alors que les statistiques peuvent offrir des indications précieuses sur les performances individuelles et collectives des joueurs, elles ne capturent jamais entièrement l'essence du jeu, son flair artistique et son imprévisibilité.

# Problématique (3/3)

## Problème posé

- Quelles sont les variables qui ont une influence sur la performance des joueurs ?
- Qui est-ce qui seront les potentiels meilleurs buteurs de chaque championnat durant l'année 2024?

# objectifs (1/1)

## Objectifs général

L'objectif général de cette étude est de mener une analyse approfondie des performances des joueurs de football de 2016 à 2023, en se concentrant sur les principales ligues européennes

## Objectifs spécifique

- $O_1$  : d'examiner les données statistiques des joueurs pour identifier les variables clés qui influent sur leur performance individuelle et collective.
- $O_2$  : anticiper les tendances futures en matière de performances des joueurs, en utilisant des techniques de modélisation prédictive basées sur les données
- $O_3$  : prédire les potentiels meilleurs buteurs de chaque championnat pour l'année 2024, en tenant compte des caractéristiques des joueurs, des stratégies des équipes et des dynamiques du jeu.

## ② Méthodologie



Ce travail a été réalisé en utilisant les logiciels :

- Excel
- PhpMyAdmin
- Rstudio version 4.2.2
- Rmarkdown pour la rédaction
- Rmarkdown à travers Beamer pour la présentation

## Natures et Sources des données

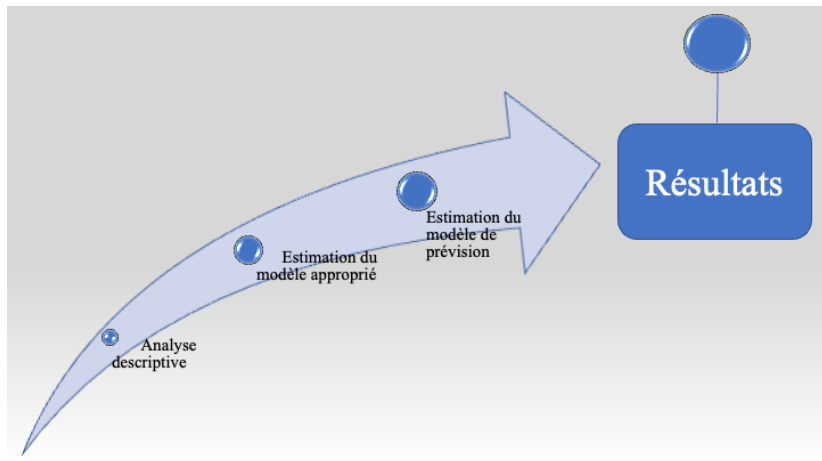
- Les données étutilisé dans la présente études couvrent la période de 2016 à 2023 et concernent les 5 grandes ligues européennes.
- Ces données proviennent de la base de sofascore sur le site de [sofascore.com](https://www.sofascore.com)

## Identification des variables de l'étude

- **La variable endogène** : Le nombre de but marqué par chaque joueur au cours d'une année;
- **Les variables exogènes** : Nombre de match jouer, Le nombre de minutes joué pendant la saison, Le nombre de tire tanté pendant la saison, Le nombre de tire cadré pendant une saison, Le ratio de but par match, Grosses occasions créées durant la saison, Identifiant associé à chaque joueur

# Methodologie (4/6)

## Processus d'analyse



# Methodologie (5/6)

## Modèle d'analyse

Notre modèle d'analyse repose sur une fonction de type régression linéaire multiple qui se présente comme suit:

### Modèle 1

$$\text{But} = \beta_0 + \beta_1 \text{Tirs\_Total} + \beta_2 \text{Gross\_occas} + \beta_3 \text{tirs\_cadrés} + \beta_4 \text{matches} + \beta_5 \text{minutes\_jouées} + \beta_6 \text{Ratio\_but\_par\_match} + \beta_7 \text{id\_nom} + \varepsilon$$

### Modèle 2

$$\text{But} = \beta_0 + \beta_1 \text{tirs\_cadrés} + \beta_2 \text{matches} + \beta_3 \text{minutes\_jouées} + \beta_4 \text{Ratio\_but\_par\_match} + \beta_5 \text{id\_nom} + \varepsilon$$

### Modèle 3

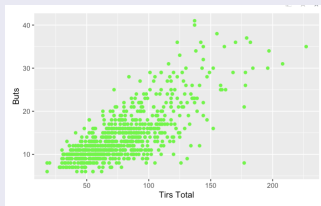
$$\begin{aligned}\hat{But}_t = & \hat{\beta}_0 + \hat{\beta}_1 tire_c adrés_t + \hat{\beta}_2 matchs_t + \hat{\beta}_3 minutesjouées_t \\ & + \hat{\beta}_4 Ratio_{but_{match}}_t + \hat{\beta}_5 id_{nom}_t + \epsilon_t\end{aligned}$$

## ③ Présentation des résultats

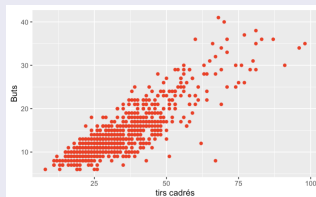
# Présentation des résultats (1/9)

## Analyse descriptive

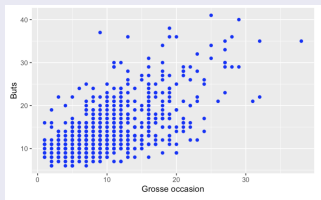
### But et Tirs total



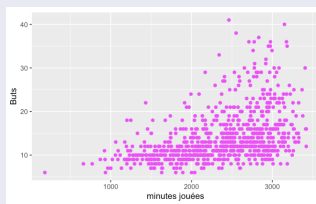
### But et tirs cadrés



### But et Grosse occasion



### But et minutes jouées





# Présentation des résultats (2/9)

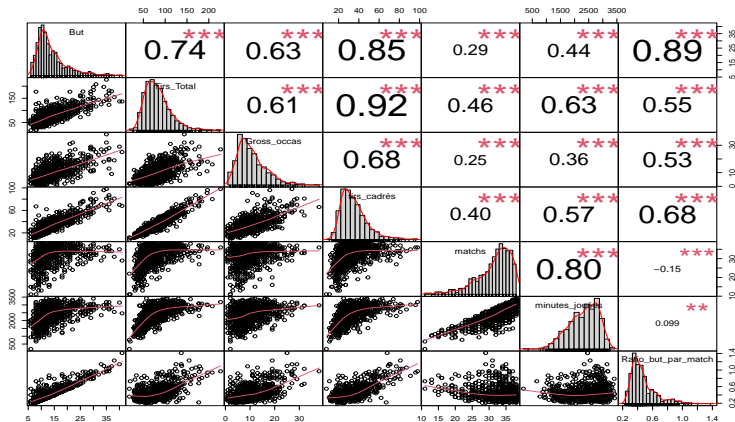
## Analyse descriptive

Variables	Statistique W	p-value	Conclusion
<b>But</b>	0,85561	2,2e-16	Absence de normalité
<b>Tirs_Total</b>	0.94029	2.2e-16	Absence de normalité
<b>Gross_occas</b>	0.93052	2.2e-16	Absence de normalité
<b>tirs_cadrés</b>	0.91627	2.2e-16	Absence de normalité
<b>matches</b>	0.89984	2.2e-16	Absence de normalité
<b>minutes_jouées</b>	0.96931	6.374e-12	Absence de normalité
<b>ratio_but_par_match</b>	0.86488	2.2e-16	Absence de normalité

# Présentation des résultats (3/9)

## Analyse descriptive

### Représentation graphique de la liaison entre les variables (Matrice de corrélation de Spearman)



# Présentation des résultats (4/9)

## Estimation du modèle

### Estimation du modèle 1

	But	
	T-Statistics	Coefficients
<b>Tirs_Total</b>	-1,050	0,2939
<b>Gross_occas</b>	-1,050	0,2695
<b>tirs_cadrés</b>	4,974	<b>8,05e-07 ***</b>
<b>matches</b>	43,464	<b>2e-16 ***</b>
<b>minutes_jouées</b>	-2,346	<b>0,0192 *</b>
<b>ratio_but_par_match</b>	99,589	<b>2e-16 ***</b>
<b>ld_Joueur</b>	1,542	0,1235
<b>C</b>	-54,704	<b>2e-16 ***</b>
<b>R<sup>2</sup>:</b>	0,9803	
<b>F-statistic:</b>	5680	
<b>p-value:</b>	2,2e-16	

\*\*\* et \* indiquent que les variables sont significatives aux seuils de 1/1000 et 5% respectivement.

# Présentation des résultats (5/9)

## Estimation du modèle

### Estimation du modèle 2

	But	
	T-Statistics	Coefficients
tirs_cadrés	3,116	<b>0,001899 **</b>
matches	-17,277	<b>2e-16 ***</b>
minutes_jouées	-3,497	<b>0,000496 ***</b>
ratio_but_par_match	-34,402	<b>2e-16 ***</b>
Id_Joueur	1,223	0,221614
C	61,631	<b>2e-16 ***</b>
$R^2$ :	0,8353	
F-statistic:	811,6	
p-value:	2,2e-16	

\*\*\* et \*\* indiquent que les variables sont significatives aux seuils de 1/1000 et 10% respectivement.

# Présentation des résultats (6/9)

## Estimation du modèle de prévision

### Modèle 3

$$\hat{But}_t = 61,631 + 3,116tire_cadrés_t - 17,277matchs_t + -3,497minutesjouées_t \\ + -34,402Ratio_{butmatch}_t + 1,223id_{nom}_t + \epsilon_t$$

# Présentation des résultats (7/9)

## Interprétation de prévision

### Pévision

Prévisions 2024		
id_nom	nom	But
68	Erling Haaland	35
1	Harry Kane	29
61	Ivan Toney	20
30	Mohamed Salah	19
41	Callum Wilson	18

# Présentation des résultats (8/9)

## Interprétation de prévision



### FOOTBALL STAT

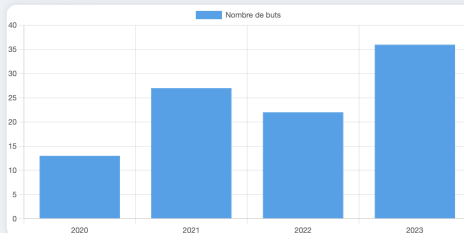


**Erling Haaland**

[Ajouter](#)

[Prediction](#)

Nationalité	Norvégienne
Clubs	Borussia Dortmund,Manchester City
Ligue	Bundesliga,Premiere league



# Présentation des résultats (8/9)

## Interprétation de prévision



### FOOTBALL STAT

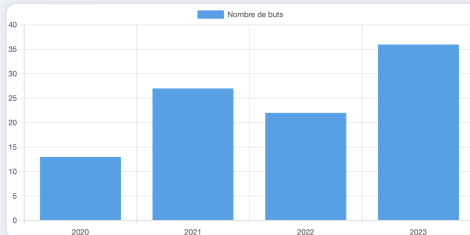


**Erling Haaland**

Ajouter

Prediction

Nationalité	Norvégienne
Clubs	Borussia Dortmund,Manchester City
Ligue	Bundesliga,Premiere league



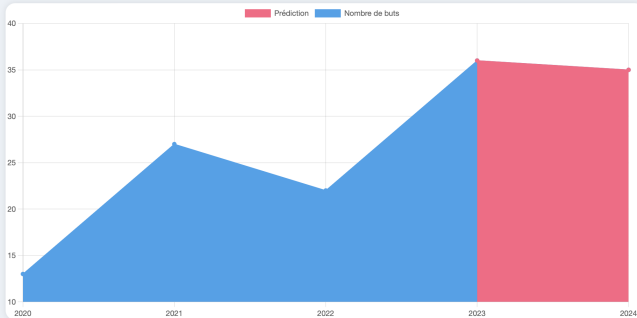


# Présentation des résultats (9/9)

## Interprétation de prévision



### FOOTBALL STAT



[Infos légales](#)

[Contact](#)

[Conditions d'utilisation](#)

# Conclusion

Merci pour votre attention