

# Rapport d'analyse de données - Joueurs de football

Stanislas ROLLAND - Pierre LECHAT

Septembre 2025

## Table des matières

<b>Introduction</b>	<b>2</b>
<b>Présentation du jeu de données</b>	<b>2</b>
Description du jeu de données . . . . .	2
Exemple d'une entrée du jeu de données . . . . .	3
<b>Les techniques d'analyse de données</b>	<b>3</b>
Analyse quantitative des données : l'ACP . . . . .	3
Analyse catégorielle / qualitative des données : la CAH . . . . .	3
<b>Conclusion</b>	<b>3</b>

## Introduction

Dans le cadre de ce travail, nous nous intéressons à l'analyse statistique d'un jeu de données regroupant les performances de joueurs de football évoluant dans les cinq grands championnats européens (Premier League, La Liga, Serie A, Bundesliga et Ligue 1) pour la saison 2024–2025. L'objectif principal de cette analyse est de dégager des tendances et des profils types parmi les joueurs, en tenant compte à la fois de leurs statistiques individuelles (buts, passes décisives, temps de jeu, etc.) et de leurs caractéristiques qualitatives (poste, équipe, nationalité, etc.). Pour ce faire, nous avons choisi d'appliquer deux techniques complémentaires d'analyse de données :

- **L'Analyse en Composantes Principales (ACP)** : une méthode d'analyse multivariée appliquée aux données quantitatives, permettant de réduire la dimensionnalité tout en conservant un maximum d'information. Elle nous aidera à visualiser les relations entre les différentes variables numériques et à identifier des profils de joueurs.
- **Une méthode d'analyse de données qualitatives** : qui permettra de mettre en valeur les variables catégorielles (poste, compétition, équipe, etc.) et de mieux comprendre la répartition des joueurs selon leurs caractéristiques.

En combinant ces deux approches, nous espérons obtenir une vision globale et nuancée des performances des joueurs de football, en tenant compte à la fois de leurs statistiques individuelles et de leur contexte compétitif. Ce rapport détaillera les étapes de notre analyse, les résultats obtenus, ainsi que les interprétations que nous en tirerons.

## Présentation du jeu de données

### Description du jeu de données

Le jeu de données que nous avons choisi d'analyser est un ensemble de données concernant les statistiques des joueurs de football sur la saison 2024 - 2025 les cinq meilleures ligues européennes (Premier League, La Liga, Serie A, Ligue 1, Bundesliga).

Notre jeu de données comporte 197 entrées (une par joueur) et les 15 variables suivantes :

#### Variables qualitatives :

- **Player** : Nom et prénom du joueur.
- **Nation** : Nationalité du joueur.
- **Pos** : Position principale du joueur sur le terrain (ex : Attaquant, Défenseur, Gardien).
- **Squad** : Équipe actuelle du joueur.
- **Comp** : Compétition dans laquelle le joueur évolue (ex : Premier League, La Liga etc).

#### Variables quantitatives :

- **Id** : Identifiant unique du joueur.
- **Age** : Âge du joueur.
- **Born** : Année de naissance du joueur.
- **MP** : Nombre de matchs joués par le joueur.
- **Starts** : Nombre de matchs commencés comme titulaire.
- **Min** : Nombre total de minutes jouées.
- **90s** : Nombre de périodes de 90 minutes jouées (équivalent à des matchs complets).
- **Gls** : Nombre total de buts marqués.
- **Ast** : Nombre total de passes décisives.
- **G+A** : Somme des buts et des passes décisives.

### Exemple d'une entrée du jeu de données

```
1 Id,Player,Nation,Pos,Squad,Comp,Age,Born,MP,Starts,Min,90s,Gls,Ast,G+A
2 1,Bukayo Saka,eng ENG,"FW,MF",Arsenal,Premier League,22.0,2001,25,20,1729,19.2,6,10,16
```

FIGURE 1 –

## Les techniques d'analyse de données

Analyse quantitative des données : l'ACP

Analyse catégorielle / qualitative des données : la CAH

## Conclusion