

# Le Machine Learning au coeur du foot

---

M1 - M2 Data Science

Théo BAUDON - Adrien GOGUE -  
Lucas GURY - Jean-Phillipe Cédric N'DRI -  
Luca MANGINI - Solène MARTINI -  
Héline MOREIRA





# Sommaire



## Partie 1

 Contexte

## Partie 2

 Etapes

## Partie 3

 Scraping

## Partie 4

 Préparation  
des données

## Partie 5

 Data Viz

## Partie 6

 Machine  
Learning



# Contexte

- Utilisation du Machine Learning
- Application au domaine du sport : Football
- Etapes :
  - Scraping
  - KPI data viz
  - KPI machine learning





# Etapes

*Scraping*



*Préparation  
des données*



*Data Viz*



*Machine  
Learning*



# Scraping

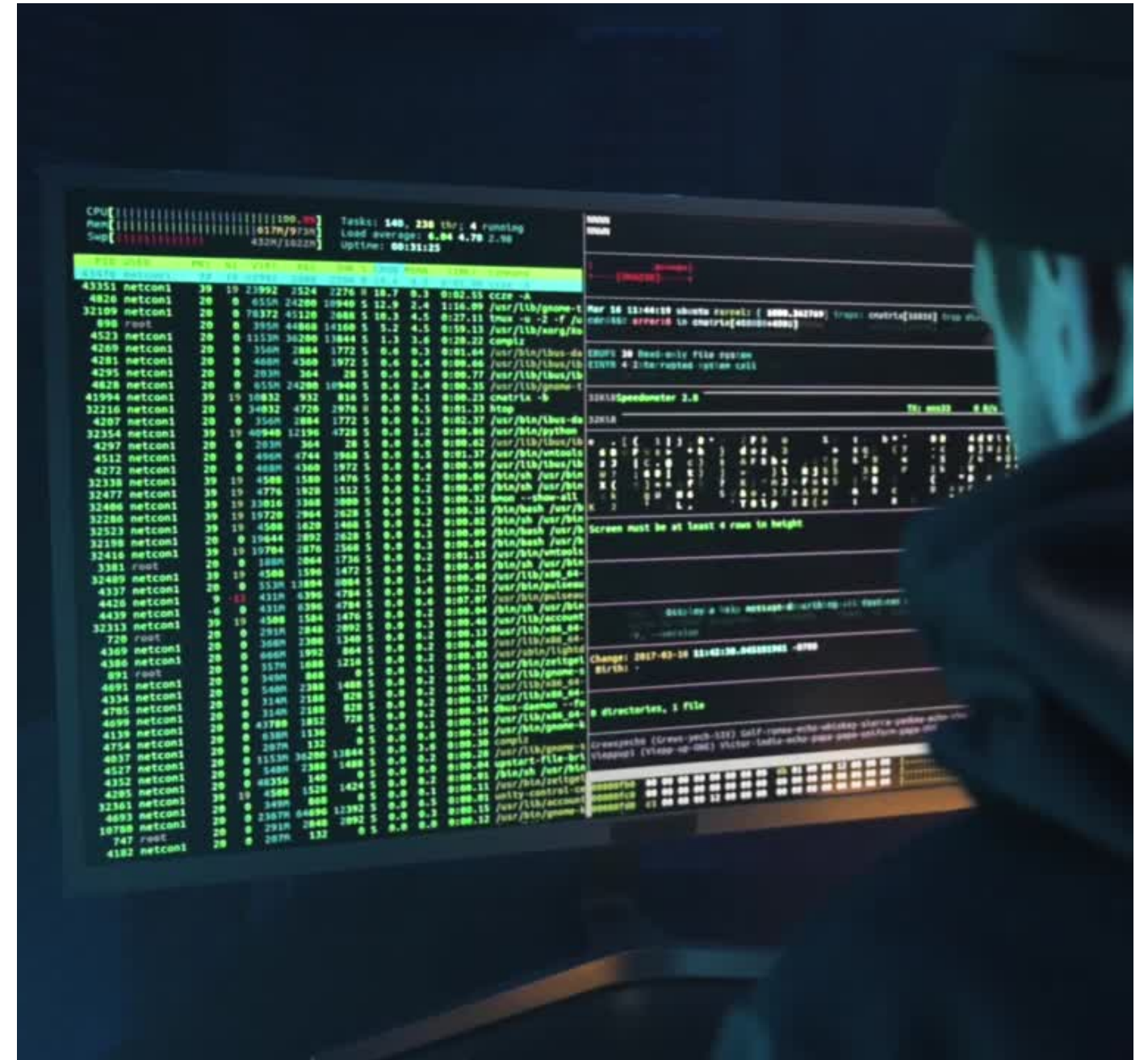
## Récupération des données :

- **FBref :**

- 5 plus grands championnats  
(France/Italie/Espagne/Allemagne/Angleterre)
- 2017 → Aujourd'hui
- Statistiques sur les joueurs et les équipes

- **Tranfer Markt :**

- 1992 → Aujourd'hui
- Poste des joueurs
- Valeurs des joueurs



# Préparation des données



## Fichier Fbref

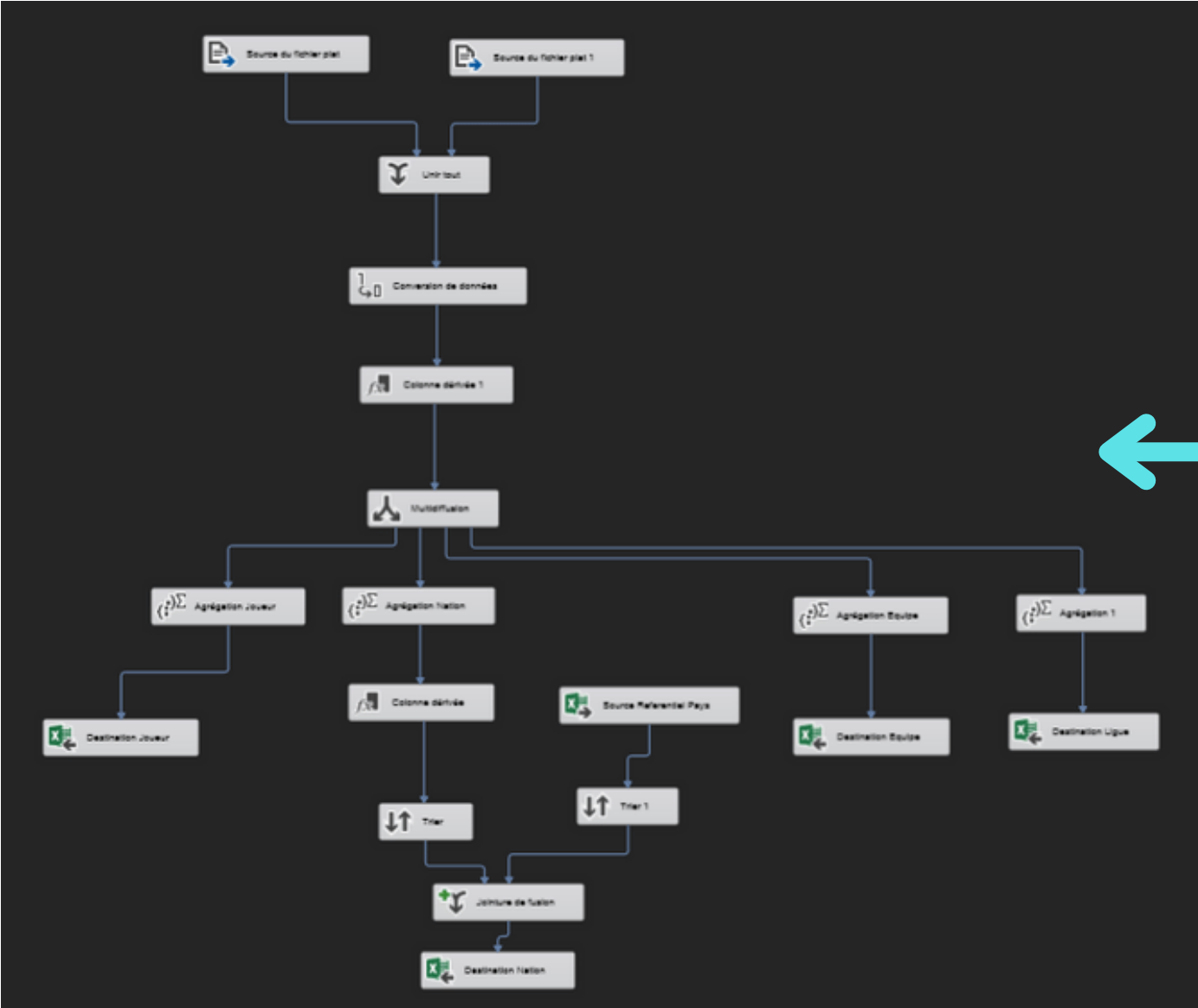
- Explication des différentes variables
- Choix des variables : cartons  
jaunes/rouges, tirs, buts, passes  
courtes/longues etc...
- Contrôle des NA
- Normalisation sur 90min

## Fichier Transfer Market

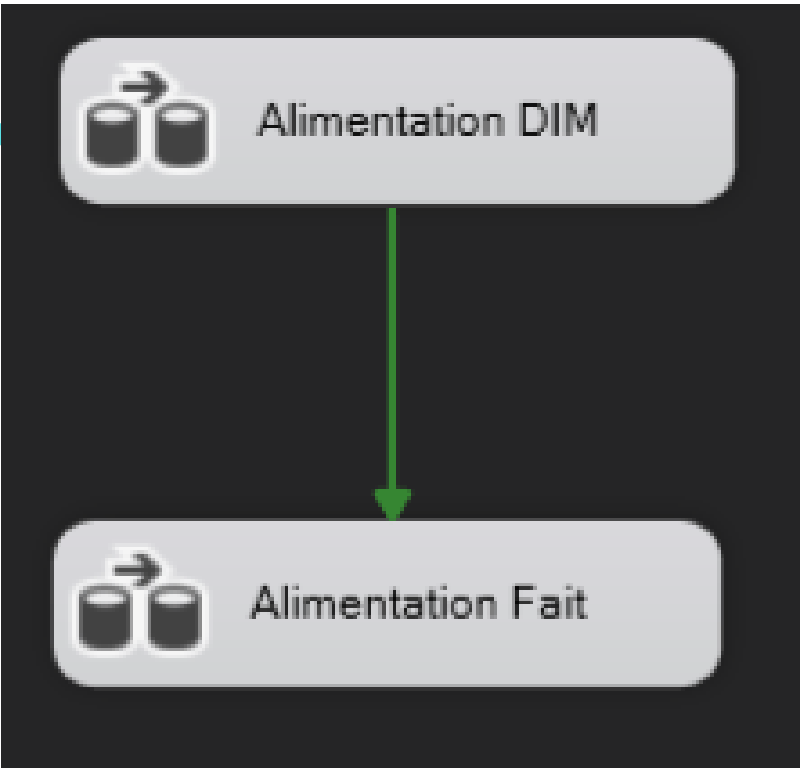
- Explication des différentes variables
- Concatenation des différents fichiers
- Choix des variables : suppression variables  
inutiles (date/ville de naissance, etc..)
- Moyenne de la valeur par saison de chaque  
joueurs



# Transformation des données Dataviz



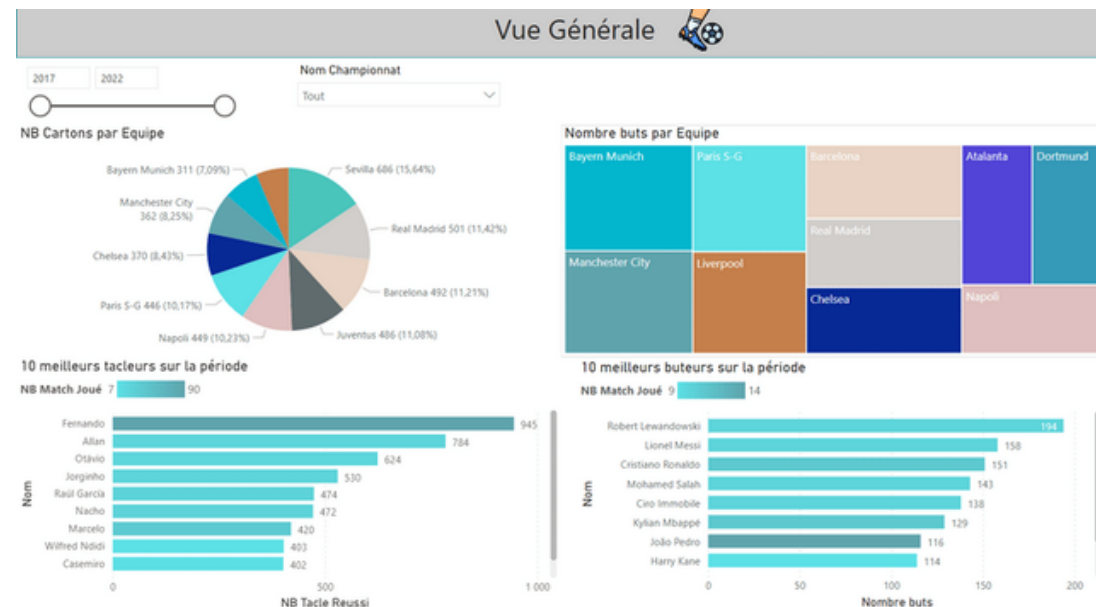
1



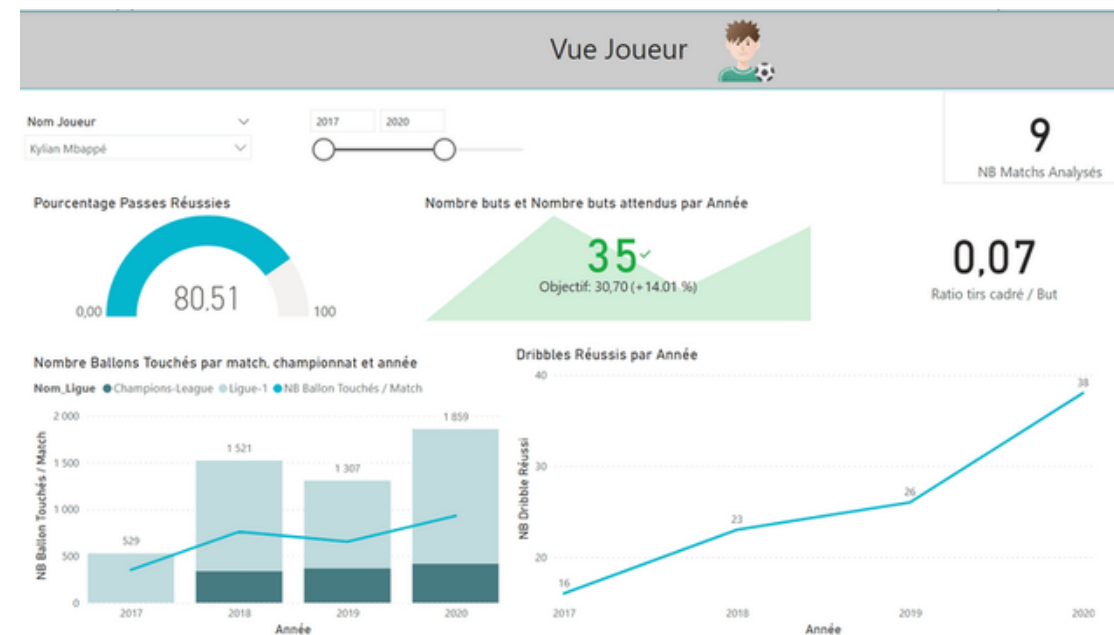
2



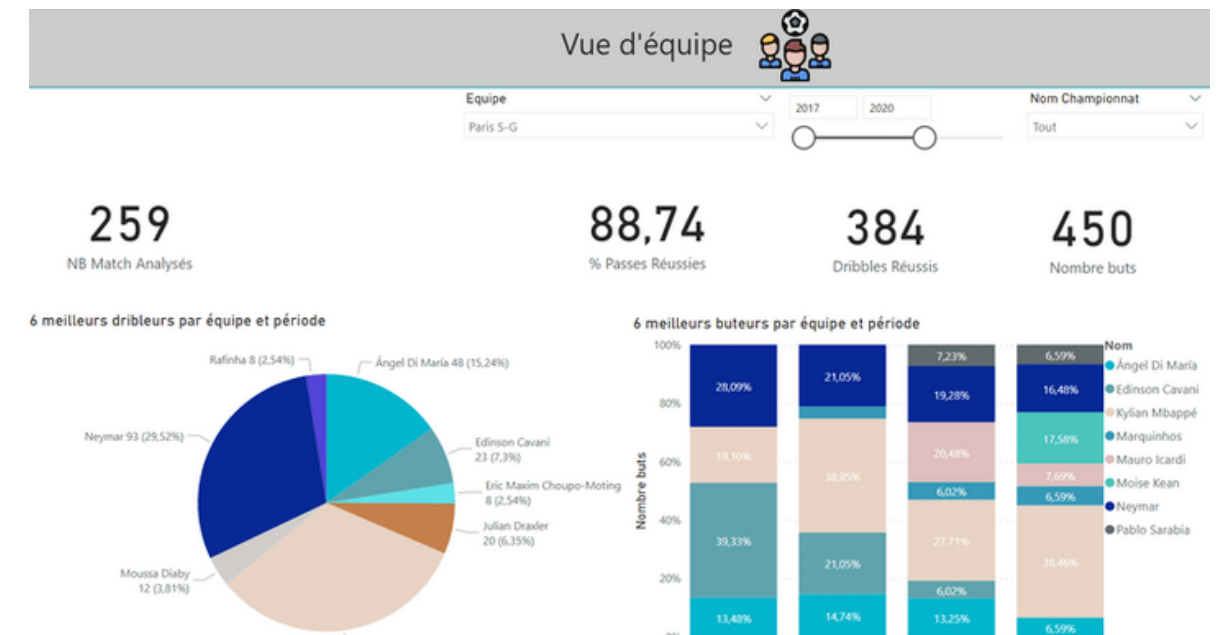
# Visualisation des données



# Vue Générale



# Vue Joueur



*Vue Equipe*



# KPI Machine Learning

## *Clustering*



Joueurs

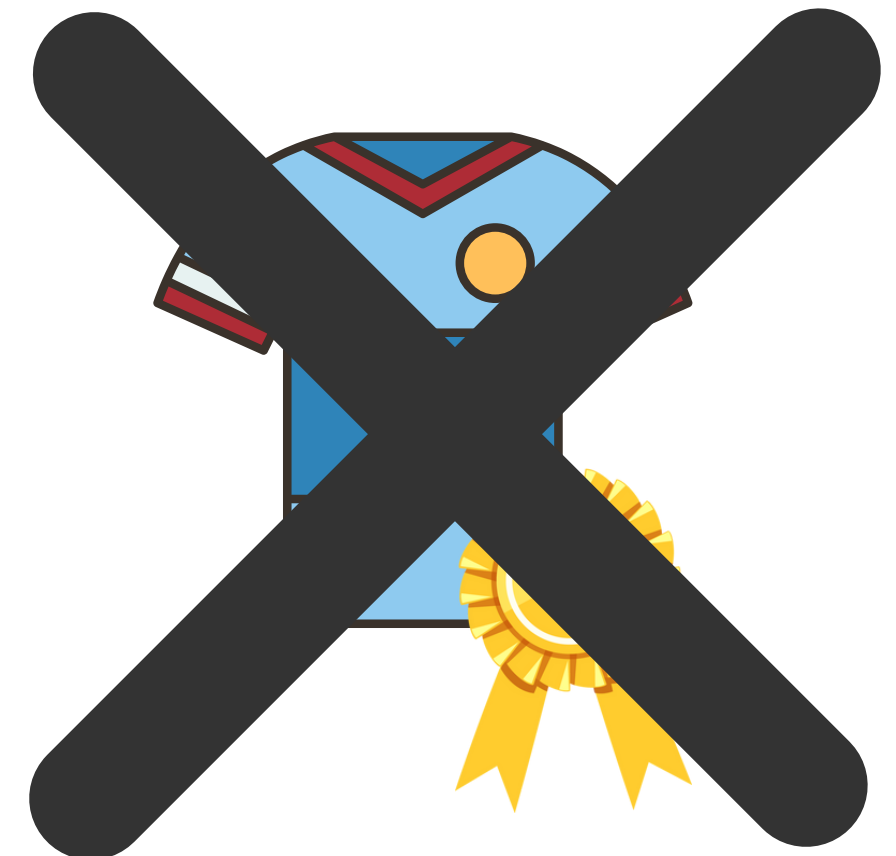


Précision  
sur les postes



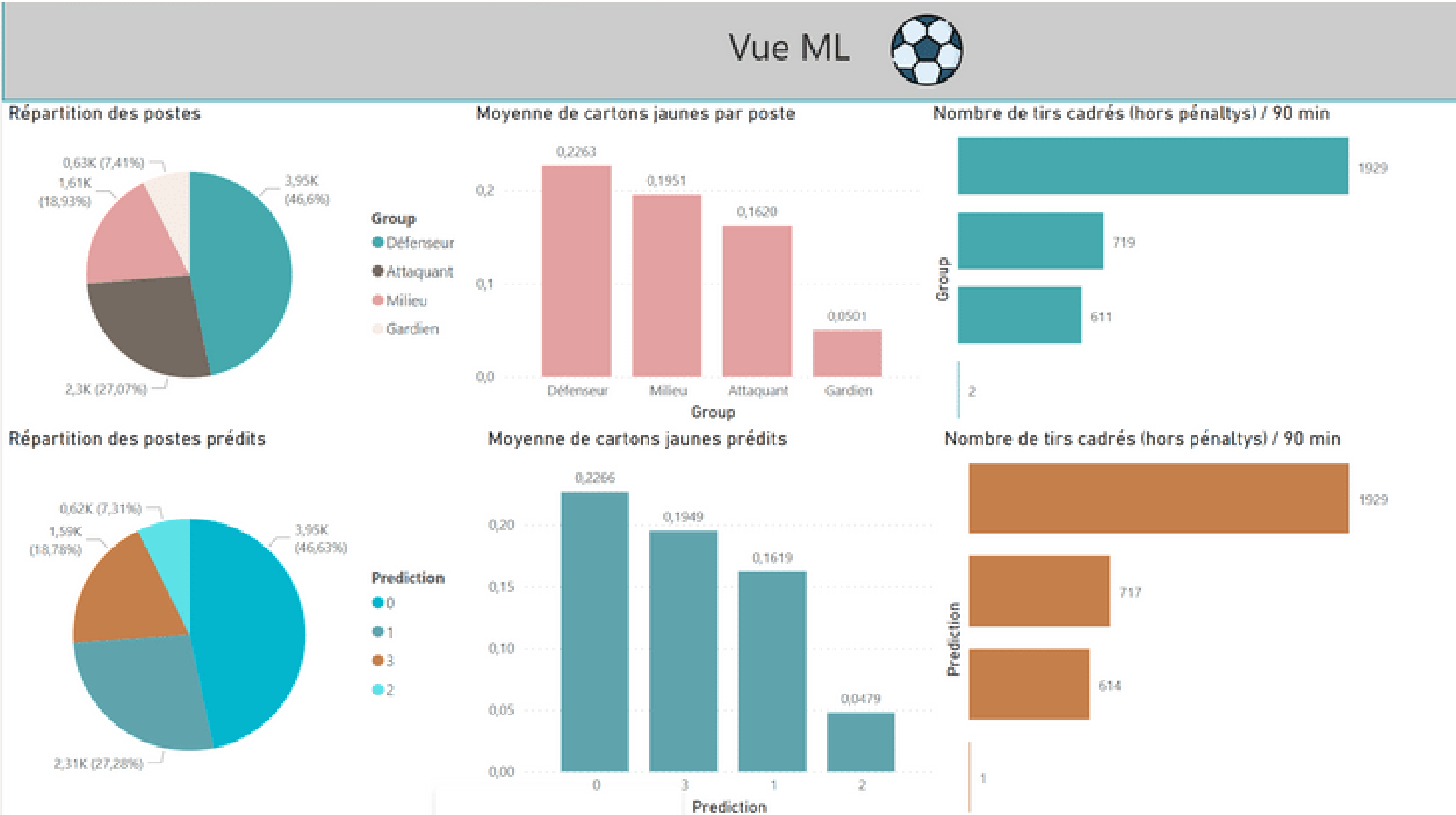
Equipes

## *Prédiction*



Valeur des joueurs

# KPI Machine Learning



Vue ML



# Partie Prédiction



Nous avons eu des problèmes dans nos jointures pour relier les données du fichier Fbref et du fichier Transfer Market.

Les noms des clubs étaient différents ainsi que les noms des joueurs, à cause de ça nous perdons environ 70% de nos données.

Nous n'avons donc pas pu faire notre partie prédiction sur la valeur des joueurs.

**POWER BI**