

# Project Pacte

Presenter par: DOOH DOOH Jeff Williams







# PLAN

01

Définition du problème et introduction aux données

02

Objectifs

03

Prétraitement des données

04

Data Analysis

05

Transfert judicieux





01

## Définition du problème et introduction aux données



# Définition du problème et introduction aux données



## CONTEXTE DU PROJET :

*De nombreux clubs de football professionnel, qui veulent rivaliser avec les meilleurs clubs.*

- Les conseils d'administration des clubs savent comment l'analyse des données et l'apprentissage automatique peuvent les aider à en savoir plus sur les compétences qui doivent être chez leurs joueurs. les meilleurs clubs dans lesquels ils doivent concourir, et la meilleure position des joueurs en fonction de leurs compétences et de la similitude des joueurs dans leur équipe afin qu'ils puissent créer une équipe forte.*

## Description des données :

*Les données contiennent :*

- Tous les joueurs disponibles de FIFA 15 à FIFA 23*
- 110 attributs*
- Attributs des joueurs avec des statistiques telles que l'attaque, les compétences, la défense, la mentalité, les compétences GK, etc...*
- Les données personnelles des joueurs telles que la nationalité, le club, la date de naissance, le salaire, etc.*



## 02 Objectifs

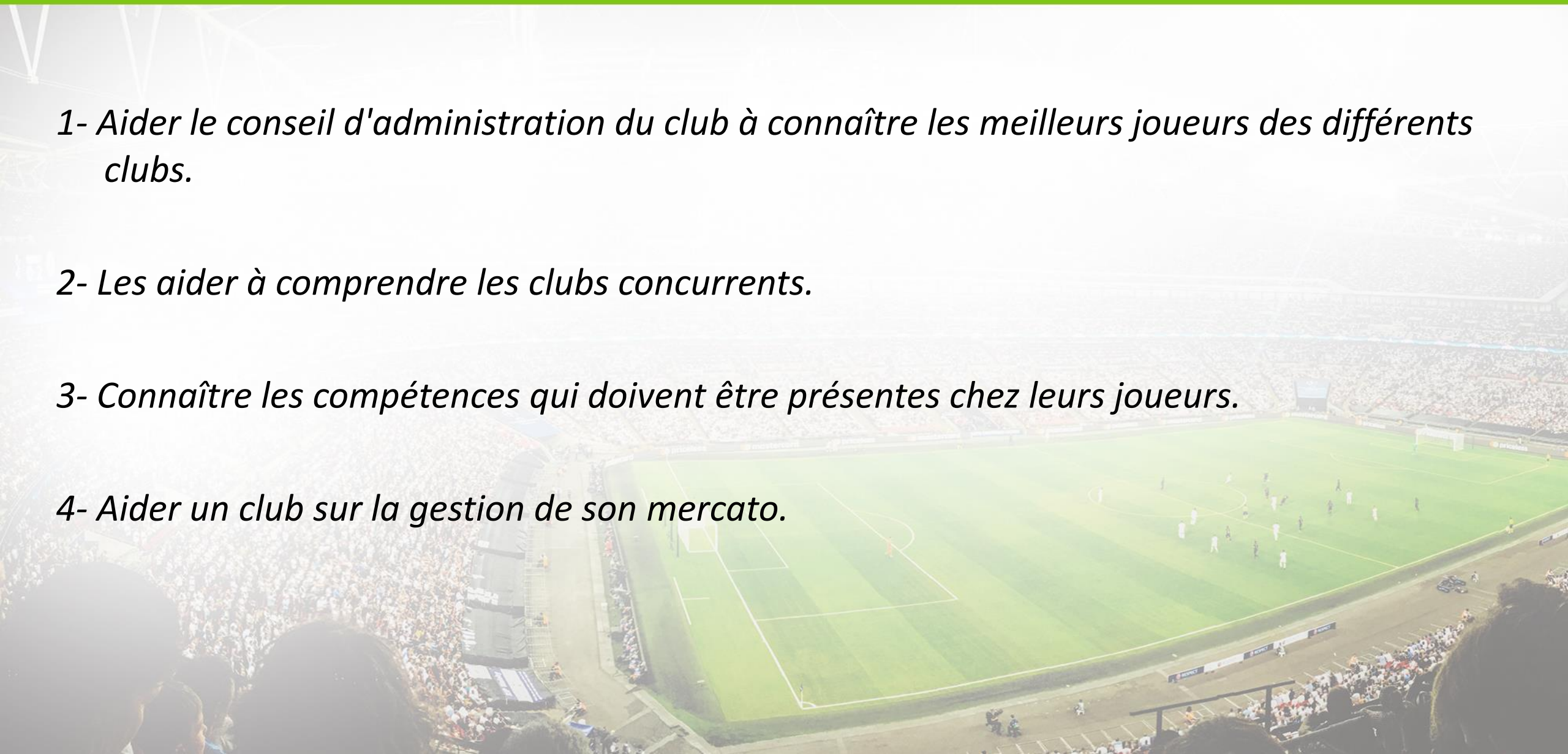




# Objectifs



- 1- Aider le conseil d'administration du club à connaître les meilleurs joueurs des différents clubs.*
- 2- Les aider à comprendre les clubs concurrents.*
- 3- Connaître les compétences qui doivent être présentes chez leurs joueurs.*
- 4- Aider un club sur la gestion de son mercato.*





# 03

## Prétraitement des données

Etapes :

1. Gestion du volume élevé de données et des valeurs dupliquées
2. Gestion des valeurs manquantes
3. Gérer les colonnes catégorielles



# Gestion de données et des valeurs dupliquées

- 1- Nous avons diviser le dataset en fonction de chaque version de FIFA (fifa\_version)*
- 2- Nous nous somme occuper des joueurs qui reviennent plusieurs fois c'est du a des mis a jour (fifa\_update)*
- 3- Nous avons choisie de ne retenir que les derniere version de chaque joueur*





# Gestion des valeurs manquantes

- 1- Nous avons fait une sélection de variable basse sur la connaissance des variables et nos objectifs dans ce projet des variable comme `player_traits` et `goalkeeping speed` ont été supprimer nous somme ainsi passe de 110 a 68 variable
  - 2- On a remarquer que dans chaque FIFA les variables `valeur_eur`, `wage_eur`, `league_name` et `club_name` on le même pourcentage de valeur manquante et il est tres faible ( $<3\%$ ), Il s'agit surement des joueur retraite on a décide de supprimer les ligne avec les valeurs manquante
- On as aussi des variables tel que `pace`, `dribbling`, `defending`, `physic`, `passing` et `shooting` qui elle aussi dans toute les version de FIFA avais des valeur manquante sensiblement égale (environs 10%) avec plus de recherche nous nous somme rendu compte qu'il s'agissait des gardiens nous avons décider de les stoker dans un autre dataset car nous allons que gérer les joueur de champs





# Gérer les colonnes catégorielles

- 1- Nous avons utiliser `LabelEncoder` de `sklearn` sur la variable `'preferred_foot'` pour remplacer gauche/droite par 1/0
- 2- La variable `player position` pose problème car il y as des joueur qui joue a plusieurs poste pour gérer ce problème nous avons utiliser le `One-Hot Encoding`
- 3- Il y as aussi des variable tel que `ls,st,rs,lw` etc... qui représente le niveau d'un joueur a diffèrent poste elle sont sous le format objet car elle sont sous la forme `'note (+/-) ajustement'` nous avons décider d'appliquer l'ajustement a la note





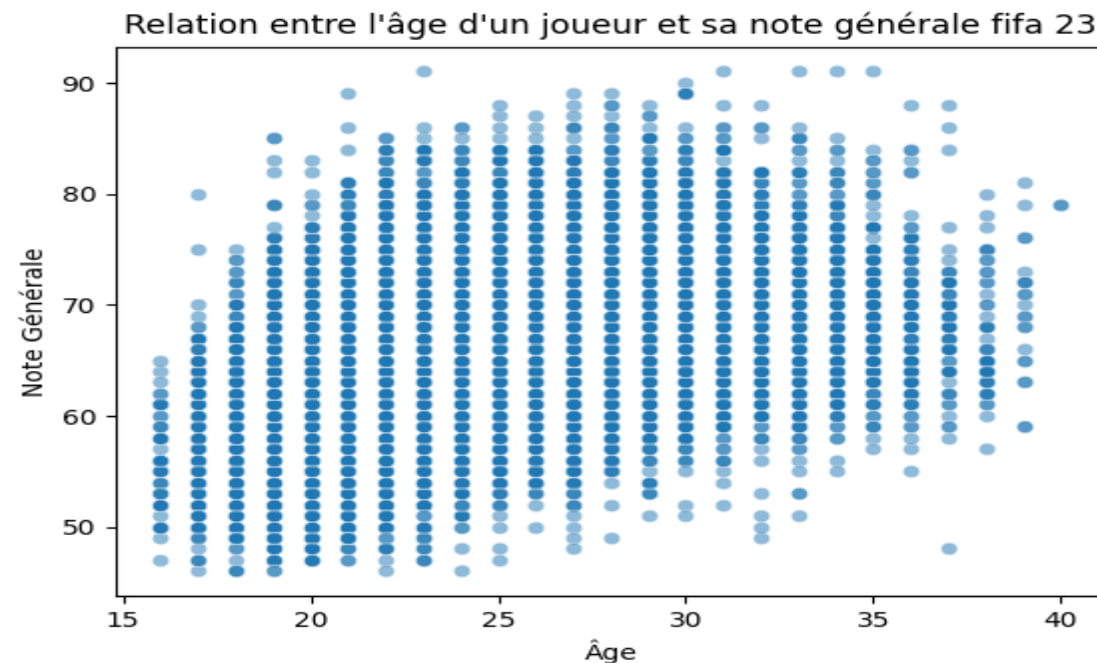
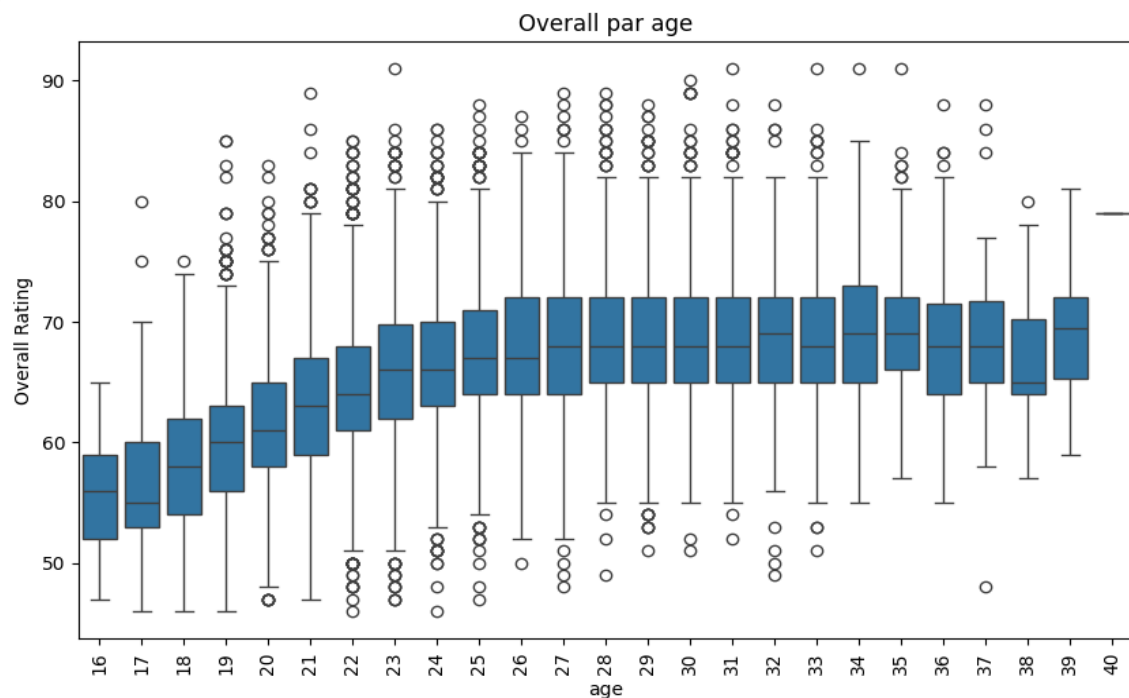
## 04 Data Analysis

1. L'âge du joueur affecte-t-il sa note général ?
2. Évolution des Caractéristiques des Joueurs à travers les Versions de FIFA
3. Comment la hauteur affecte différents facteurs du joueur ?
4. Relation entre le salaire et la note des joueurs ?
5. Déterminons s'il existe une relation entre la position du joueur et son salaire
6. Observation des nationalité des 100 joueur les mieux payer et des 100 joueur les mieux note ? Que peux observer ?
7. Quel sont les nationalité les plus présente dans le dataset
8. Voir les 50 meilleurs joueurs et leurs clubs

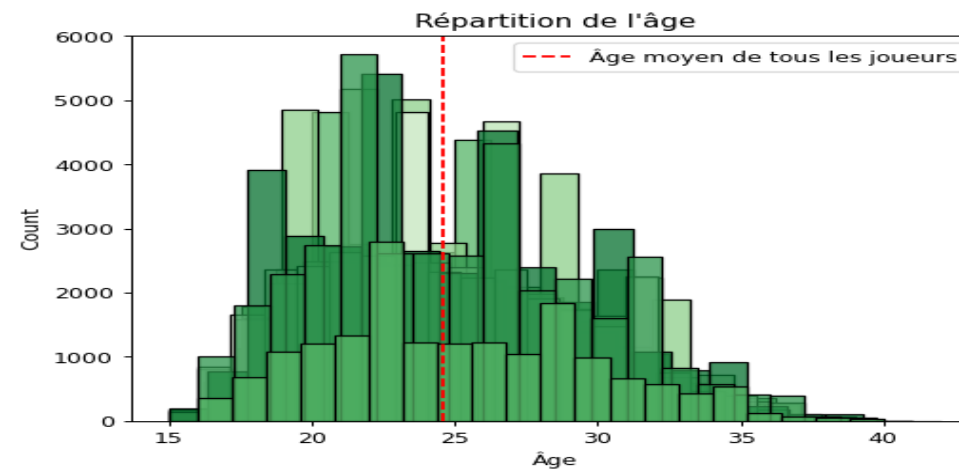




# L'âge du joueur affecte-t-il sa note général ?

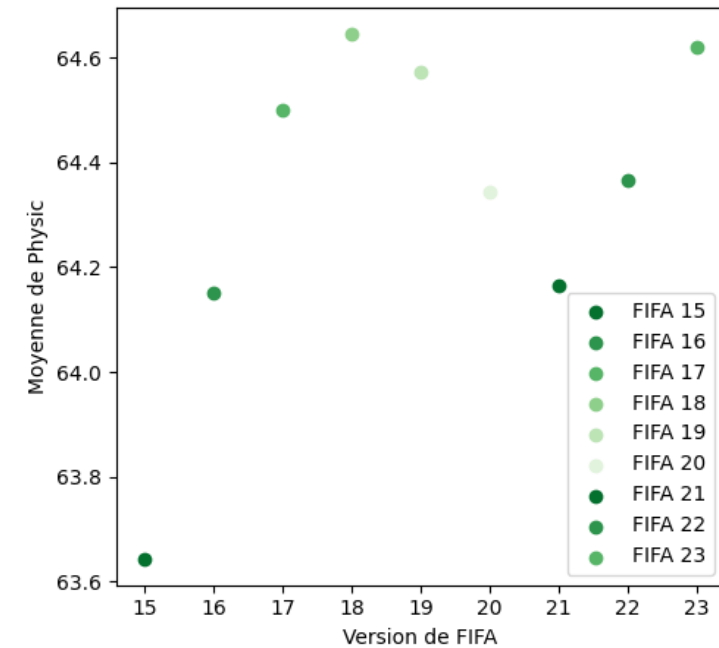
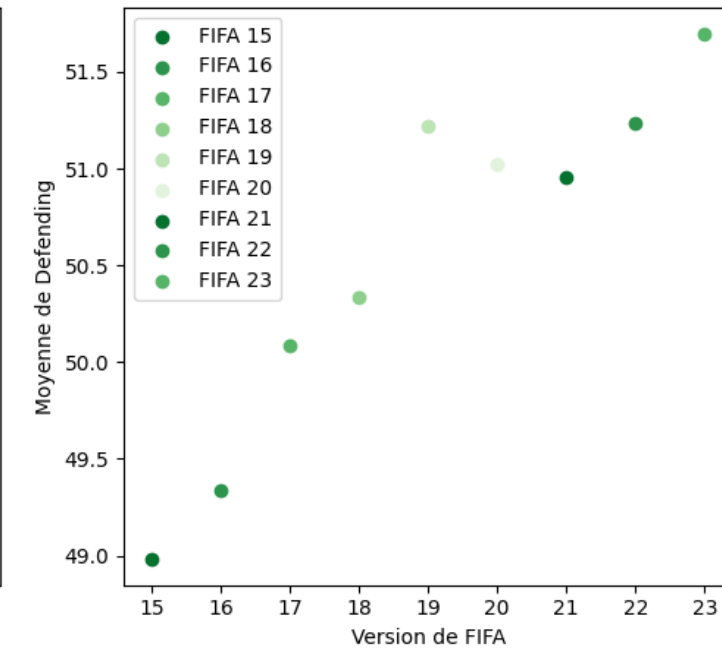
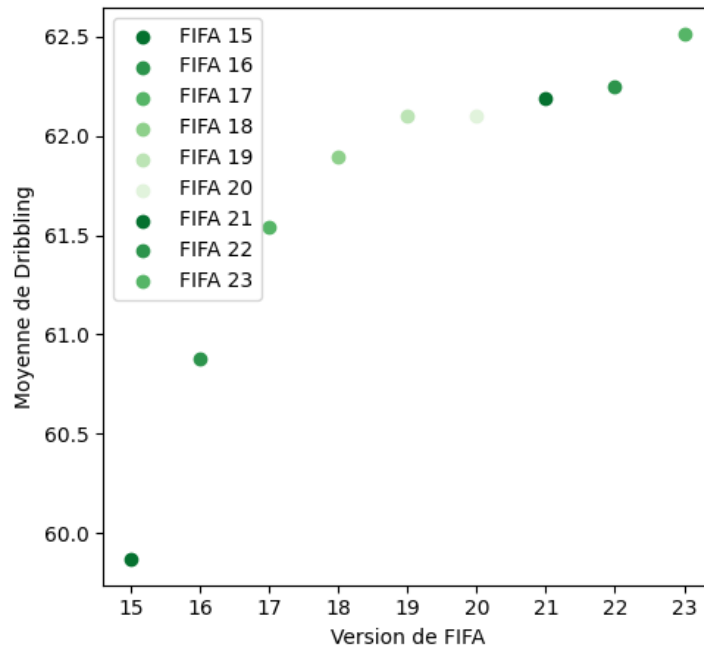
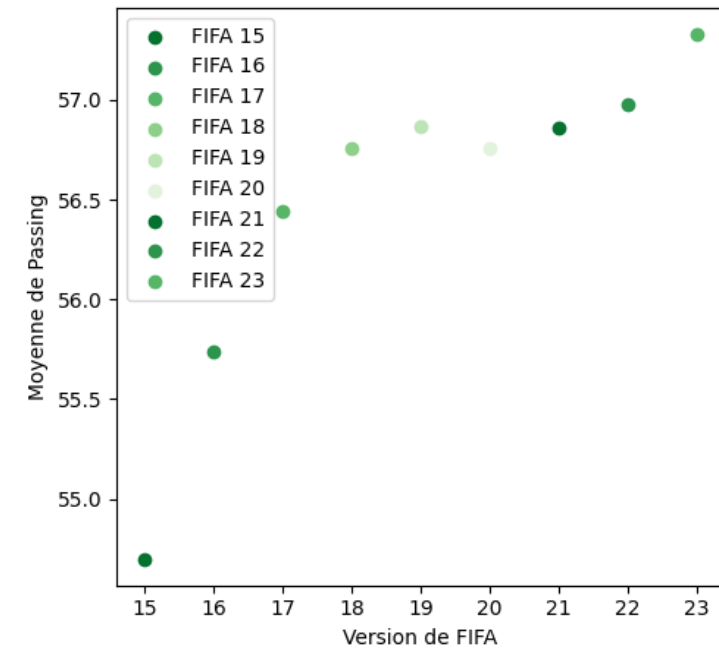
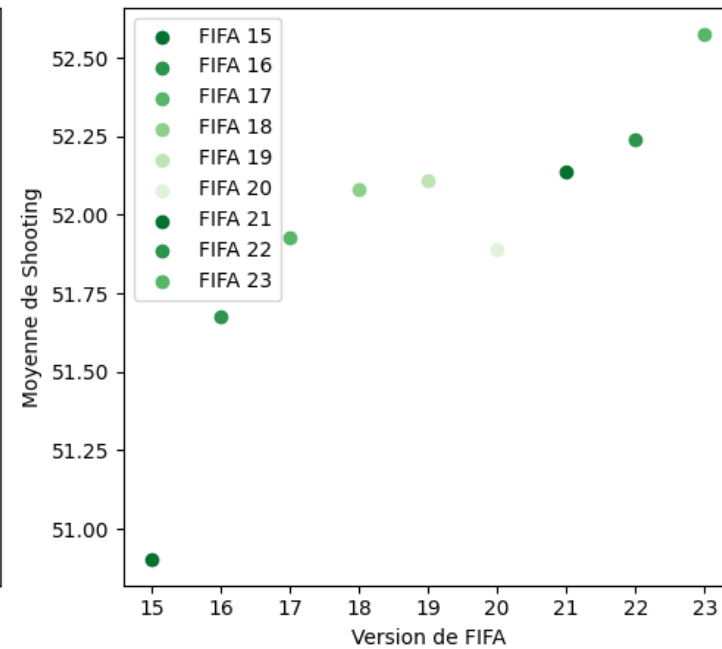
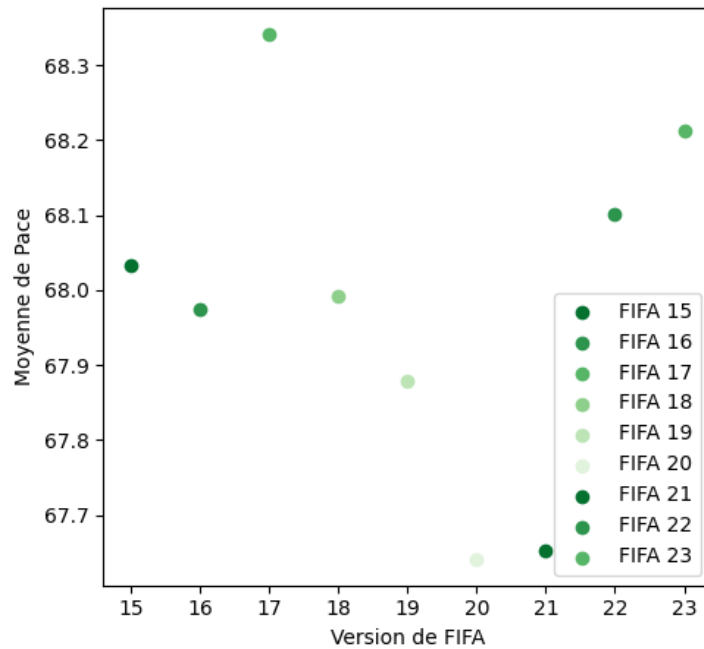


La corrélation est moyenne mais on peut quand même remarquer que après 26 ans le niveau se stabilise il pourrait donc être plus judicieux de recruter un joueur de moins de 25 ans



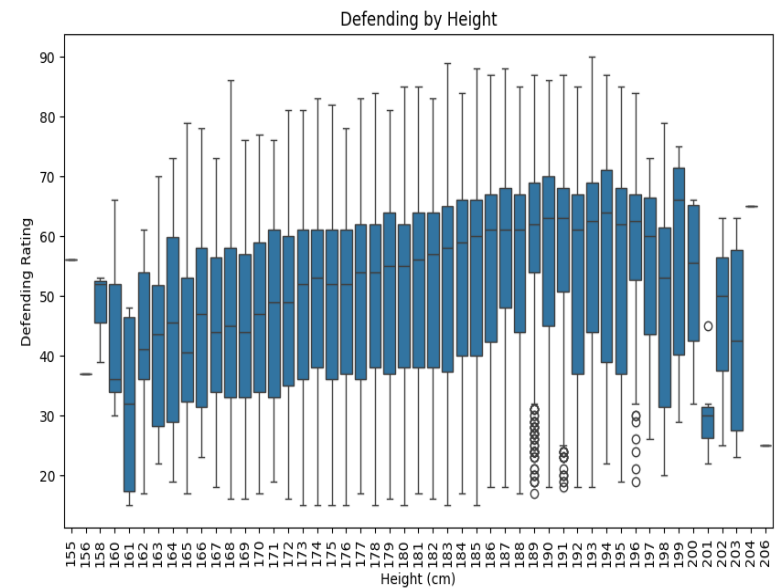
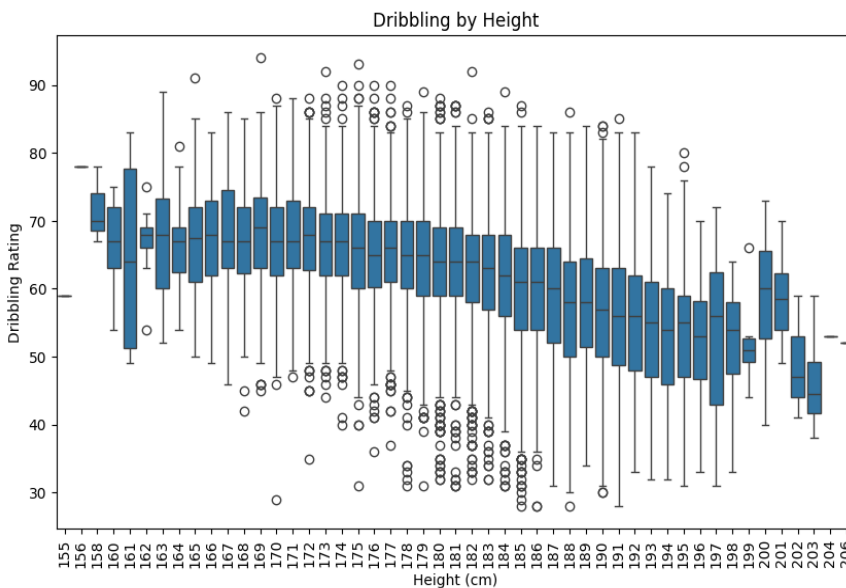
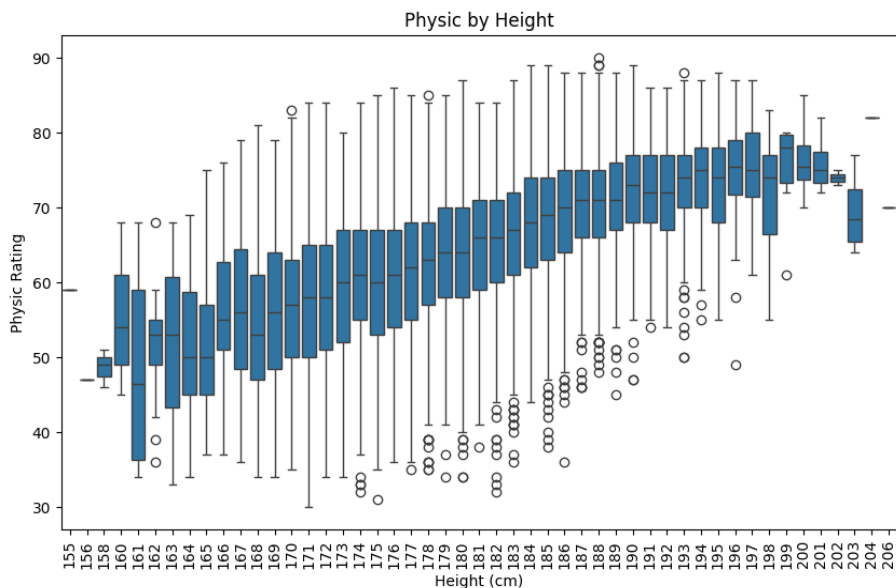
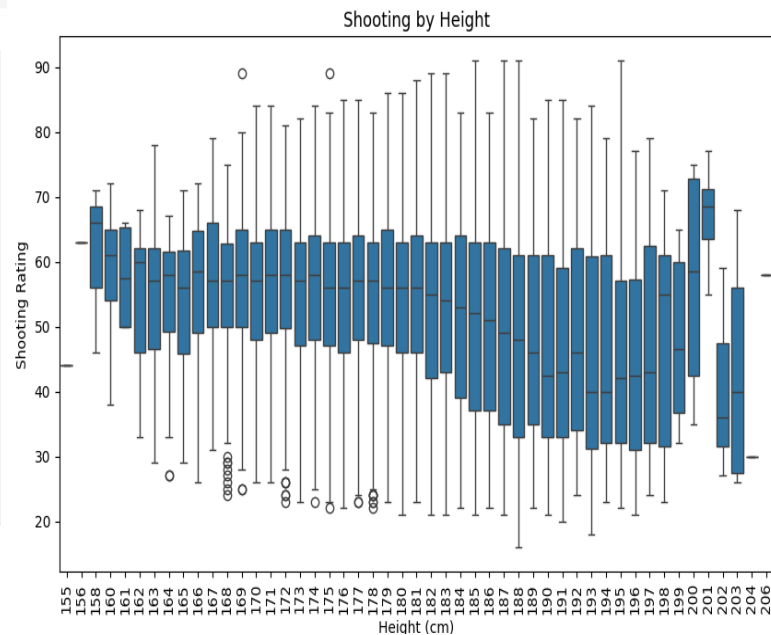
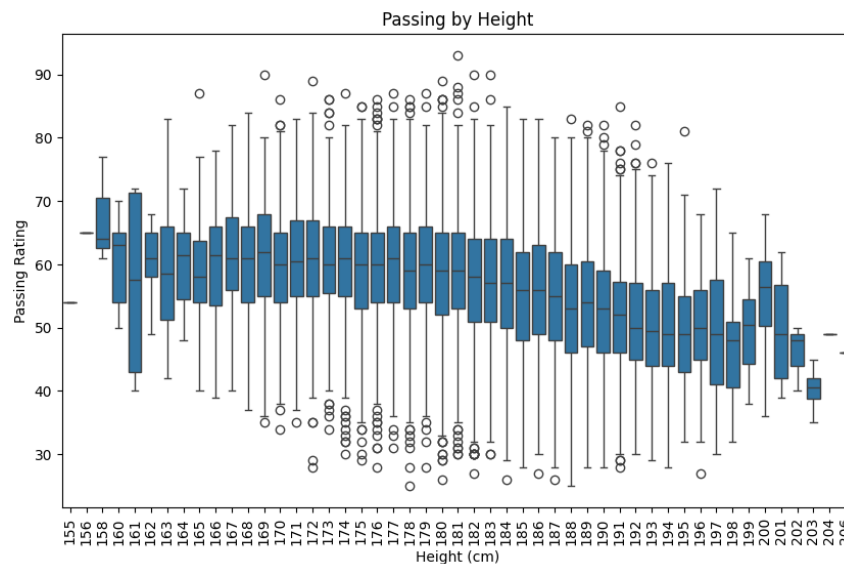
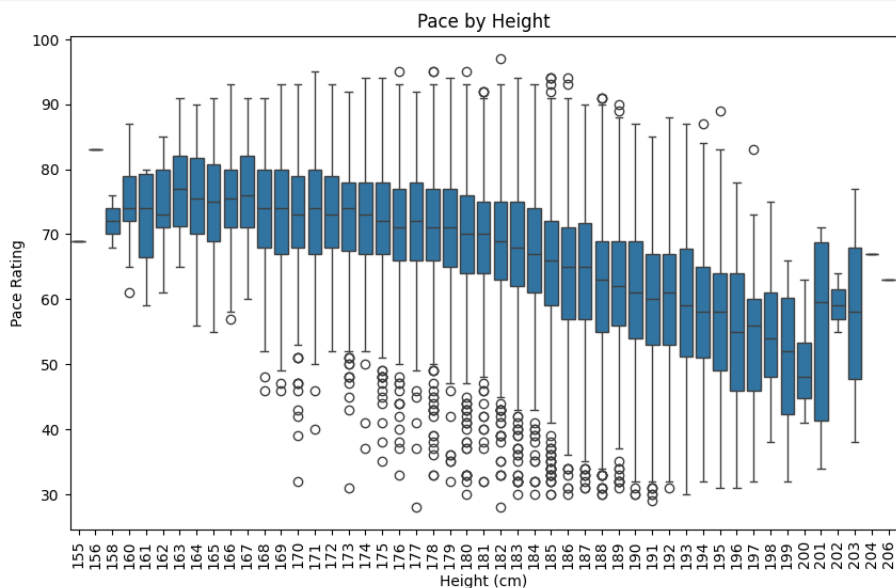


# Évolution des Caractéristiques des Joueurs à travers les Versions de FIFA



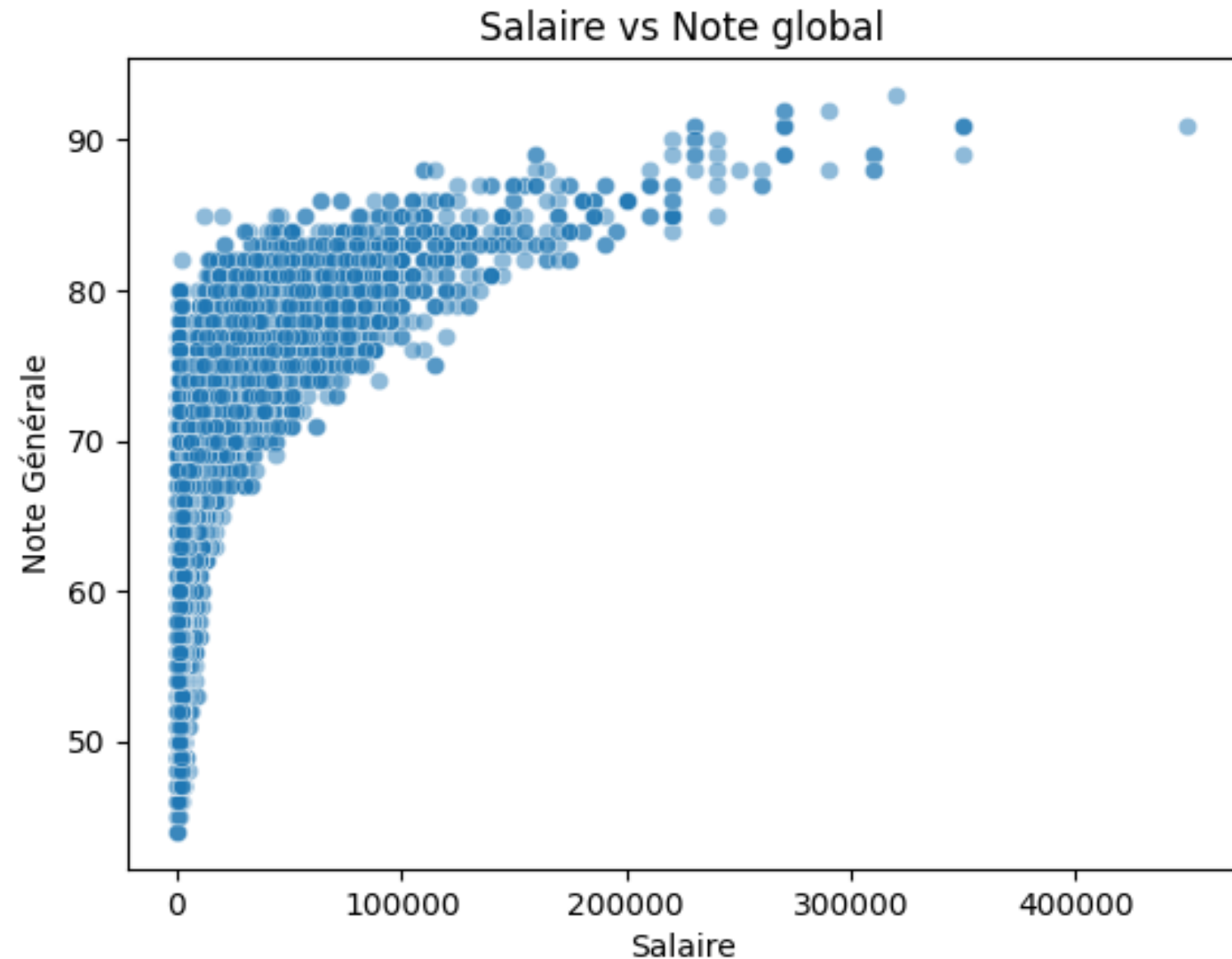


# Comment la hauteur affecte différents facteurs du joueur ?





# Relation entre le salaire et la note des joueurs ?

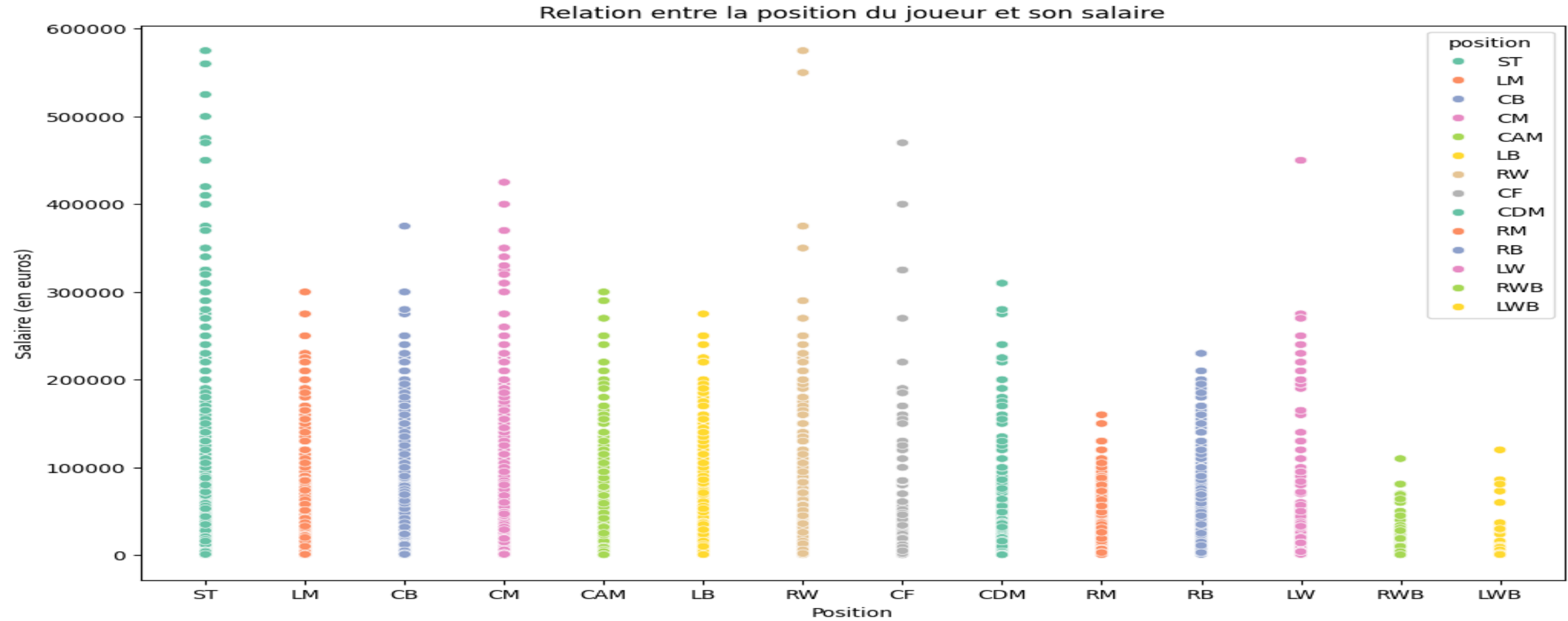


Plus le joueur a une bonne note plus il est bien payer



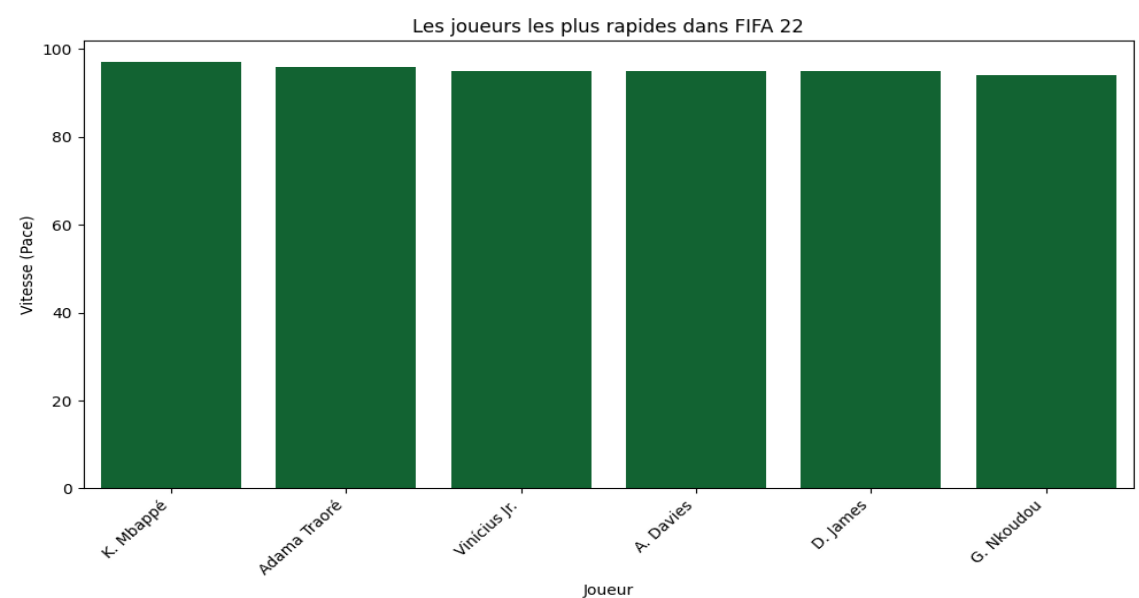
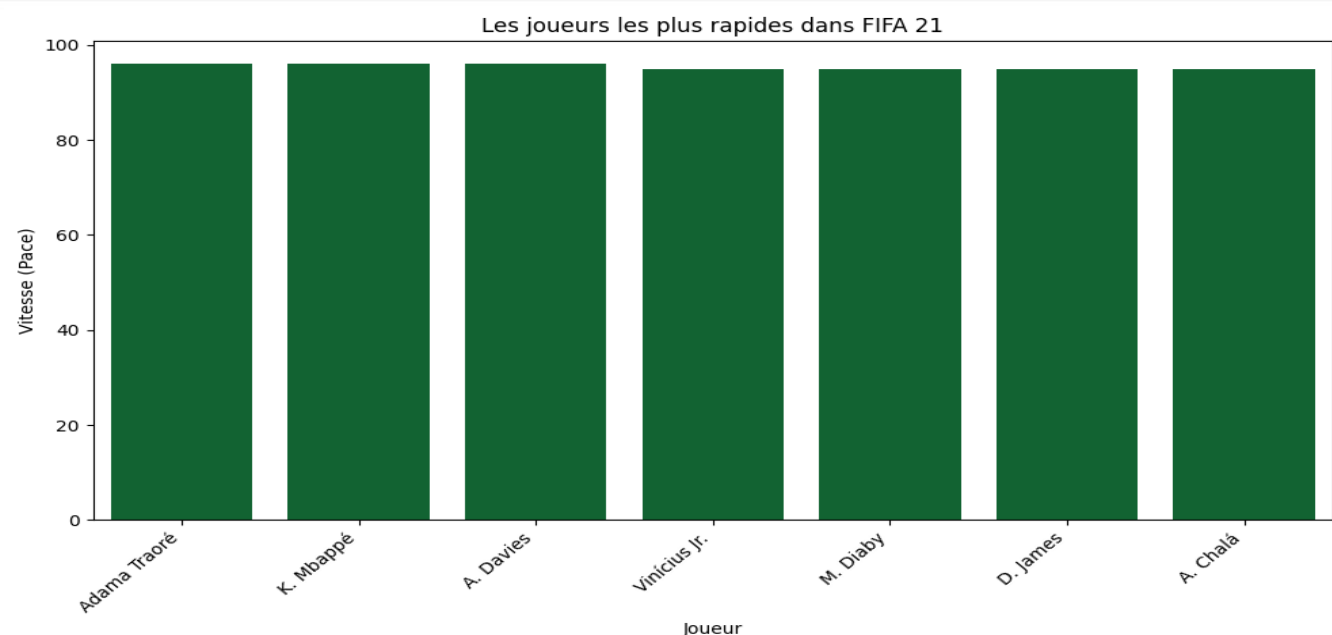
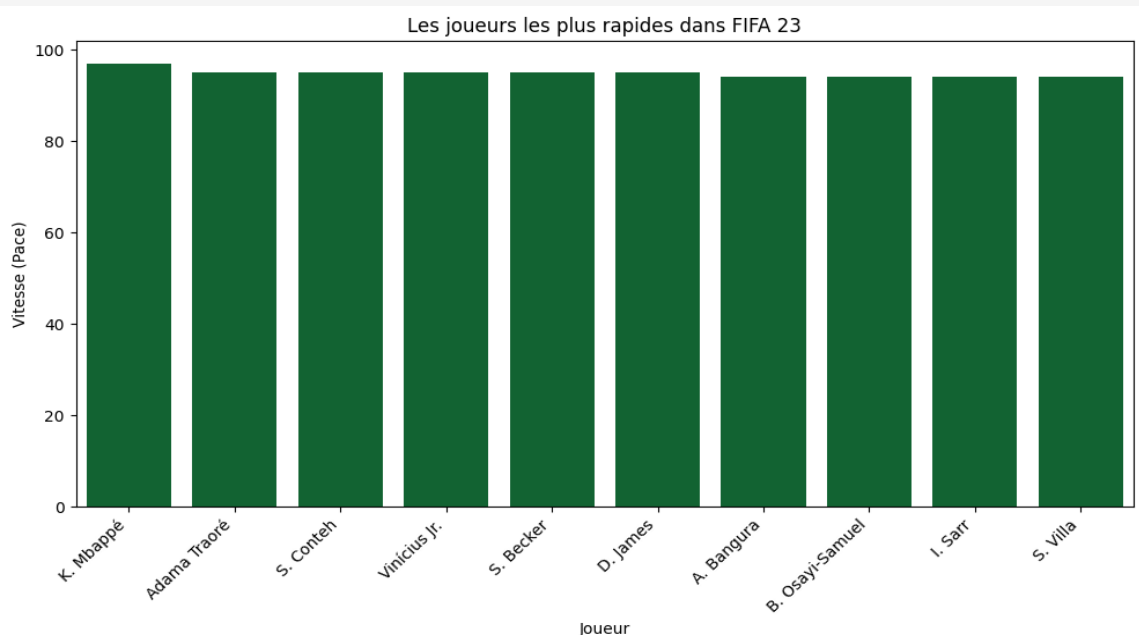


# Relation entre la position du joueur et son salaire





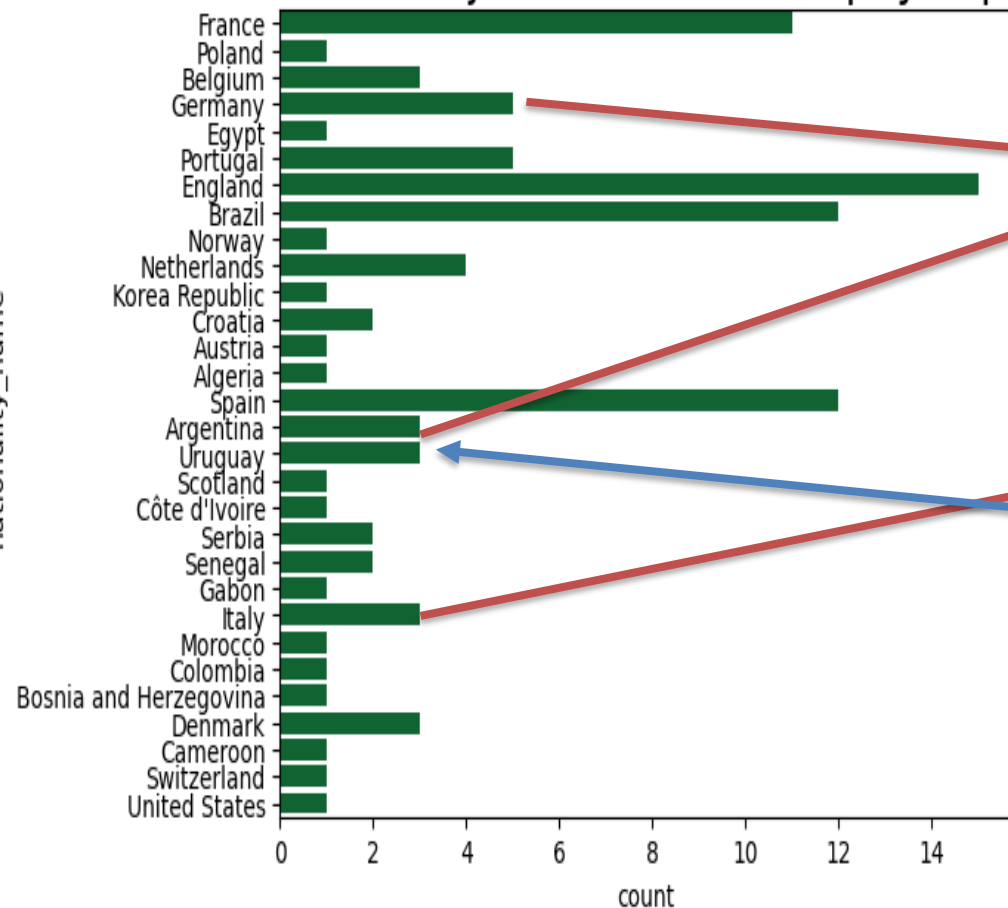
# Observons les joueur les plus rapides



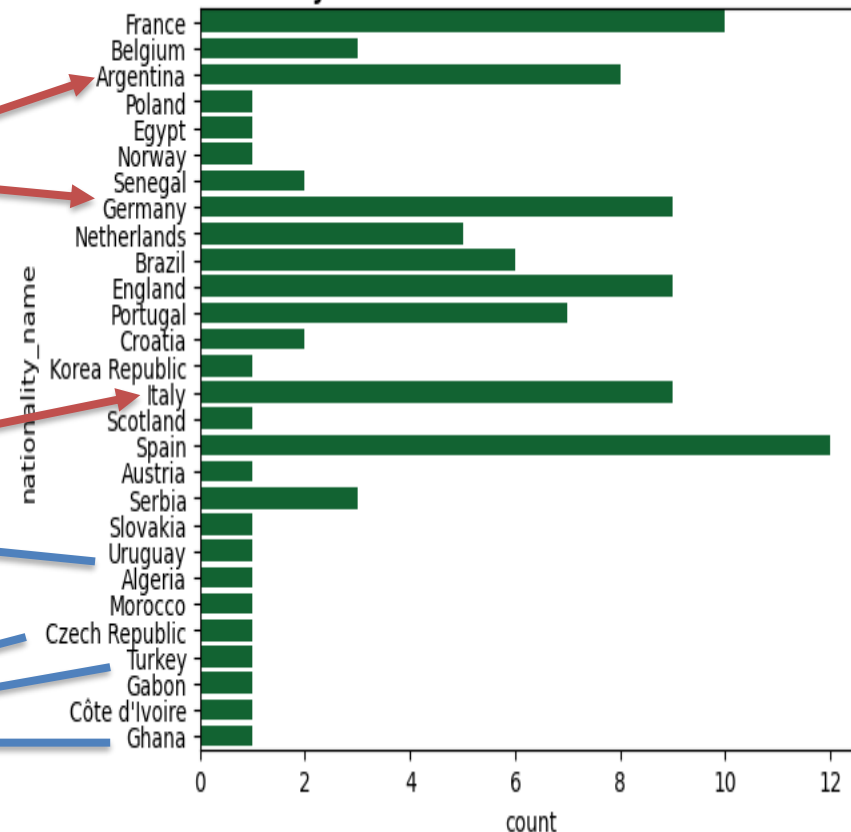


# Que peux on déduire d'un échantillons de joueur ?

Nationalité des 100 joueurs les mieux payés pour FIFA 23



Nationalité des 100 joueurs avec la meilleure note pour FIFA 23

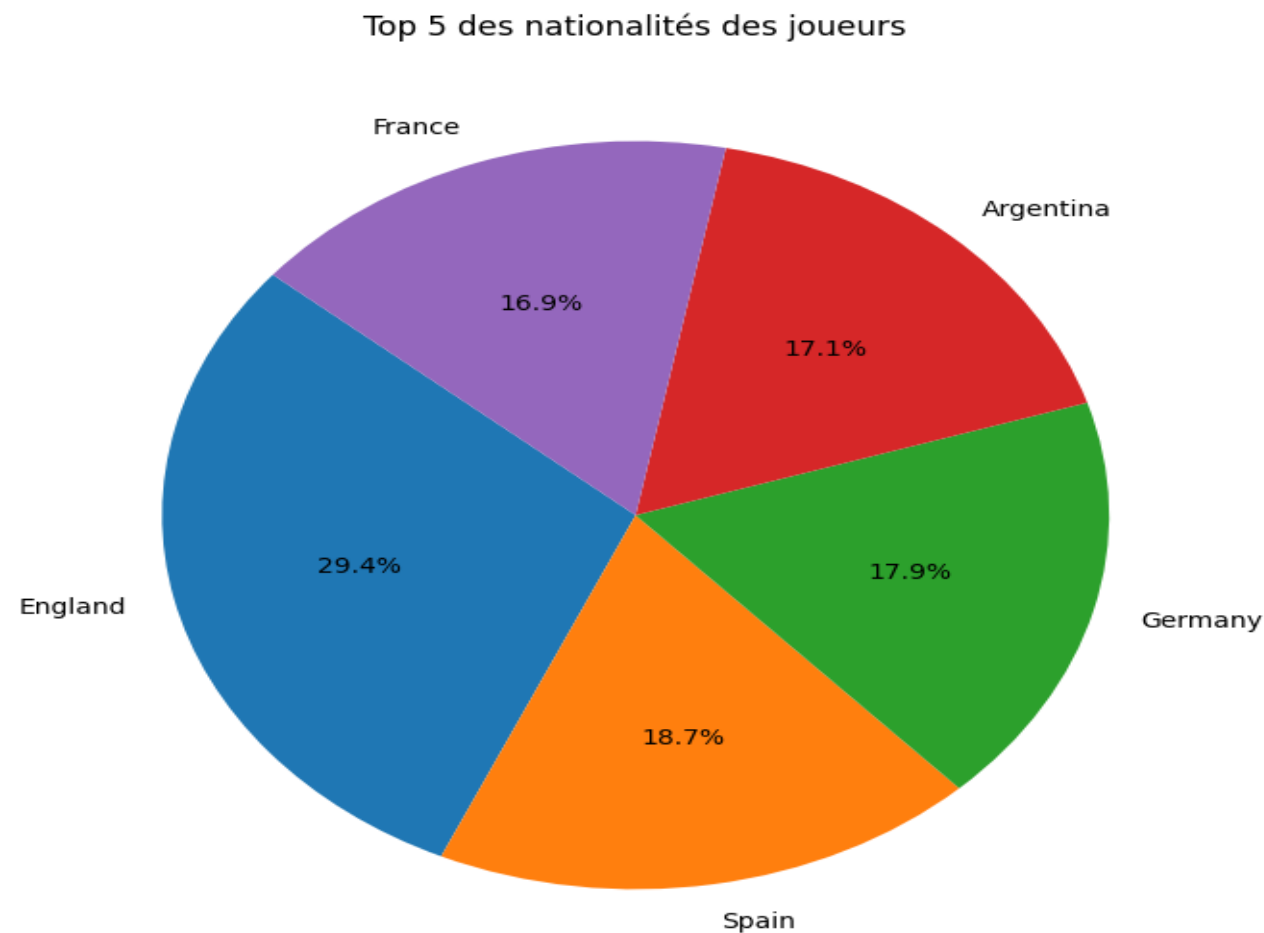
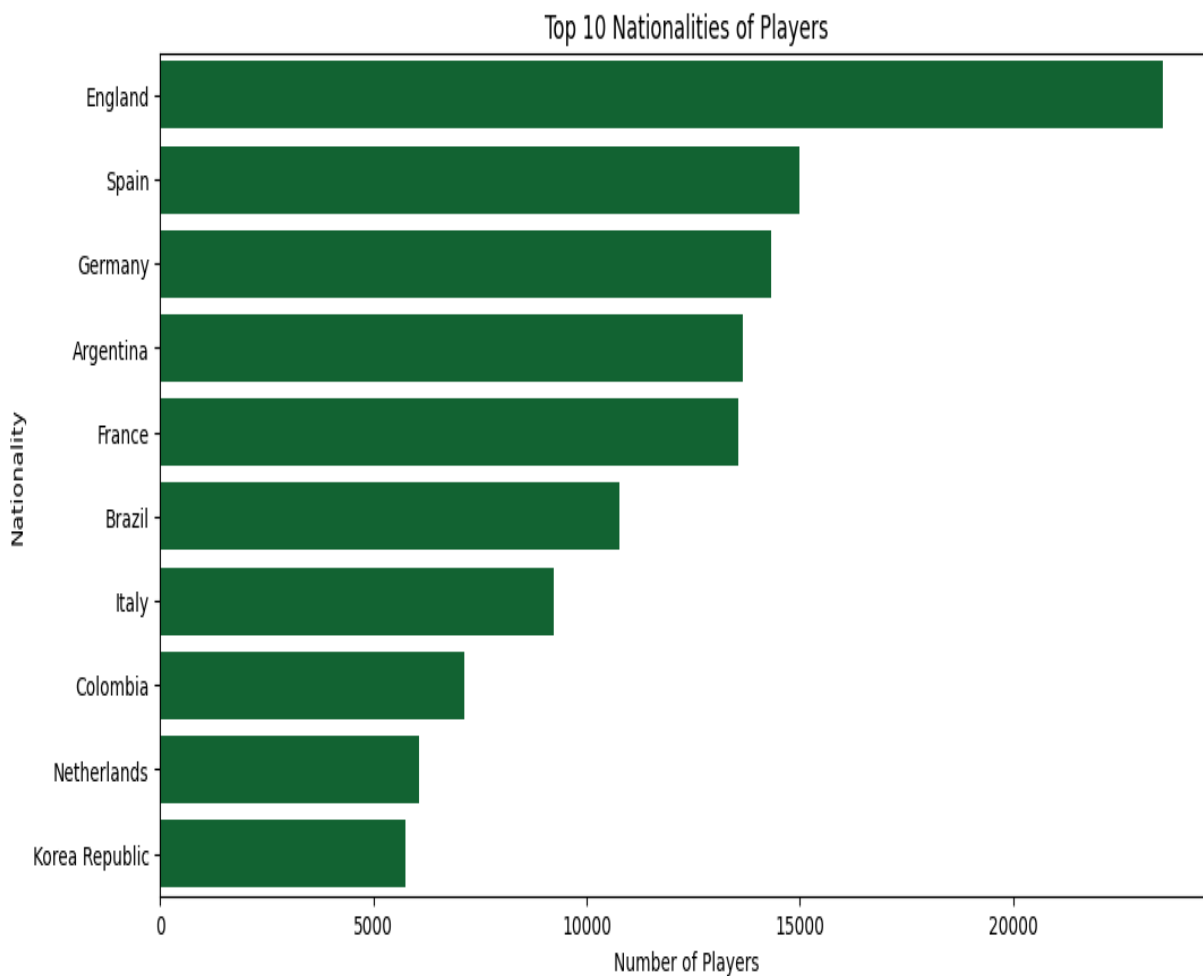


?

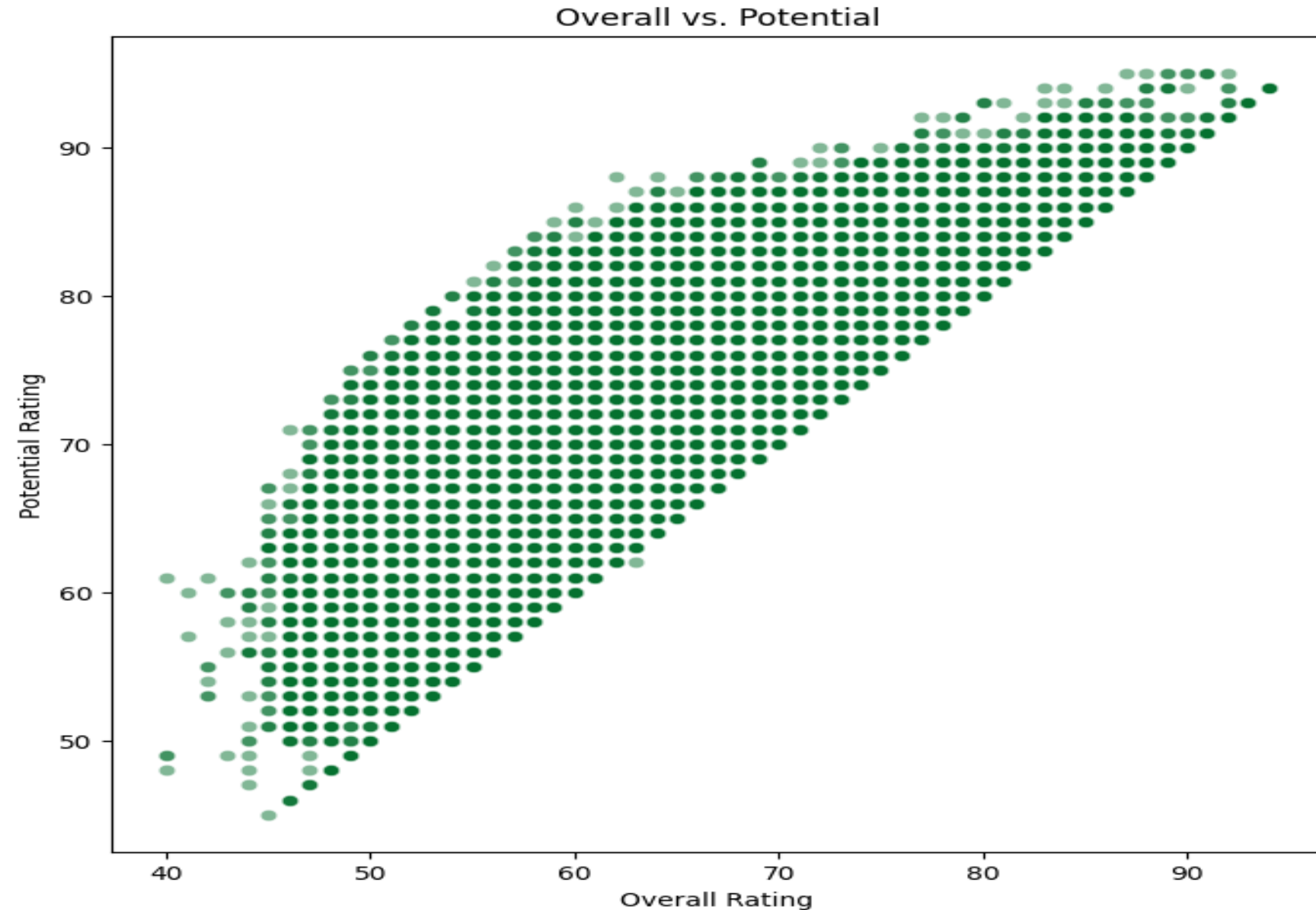




# Quel sont les nationalité les plus présente dans le dataset



# Relation entre Note et Potentiel du joueur ?

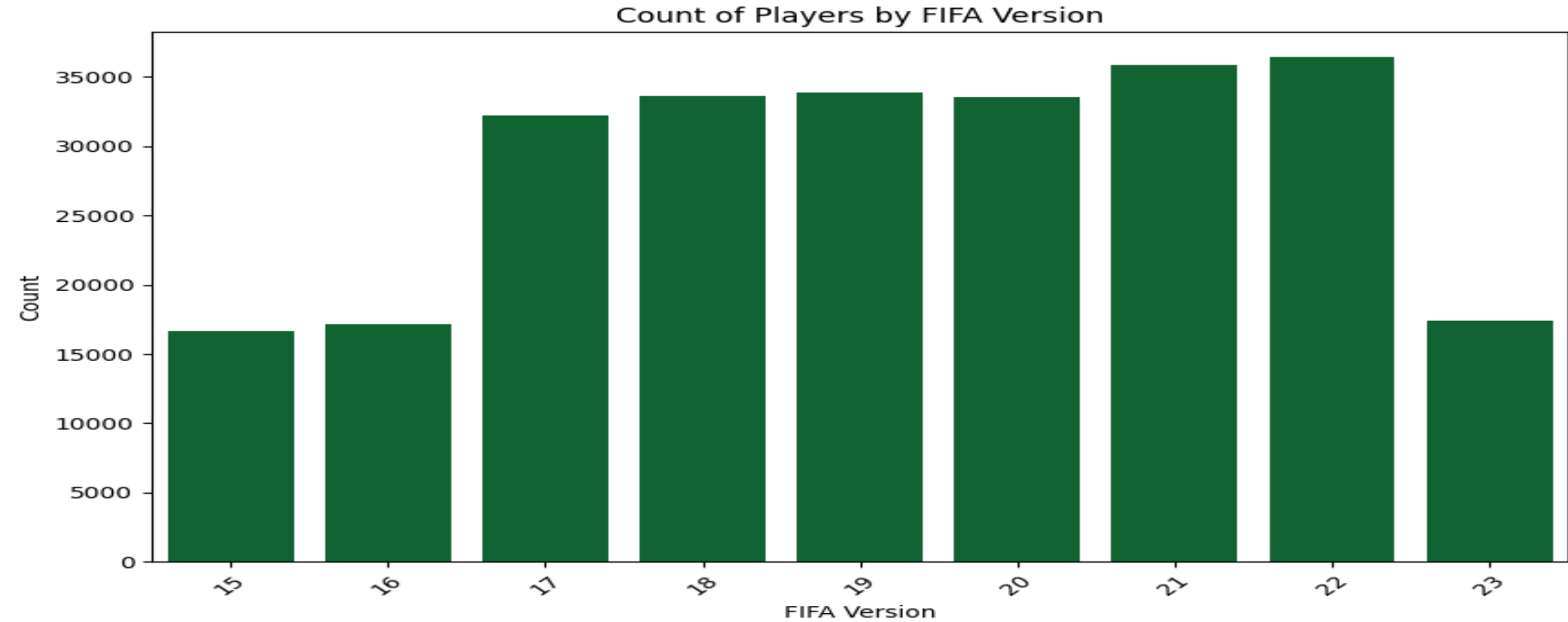


On observe une très forte corrélation entre note et potentiel



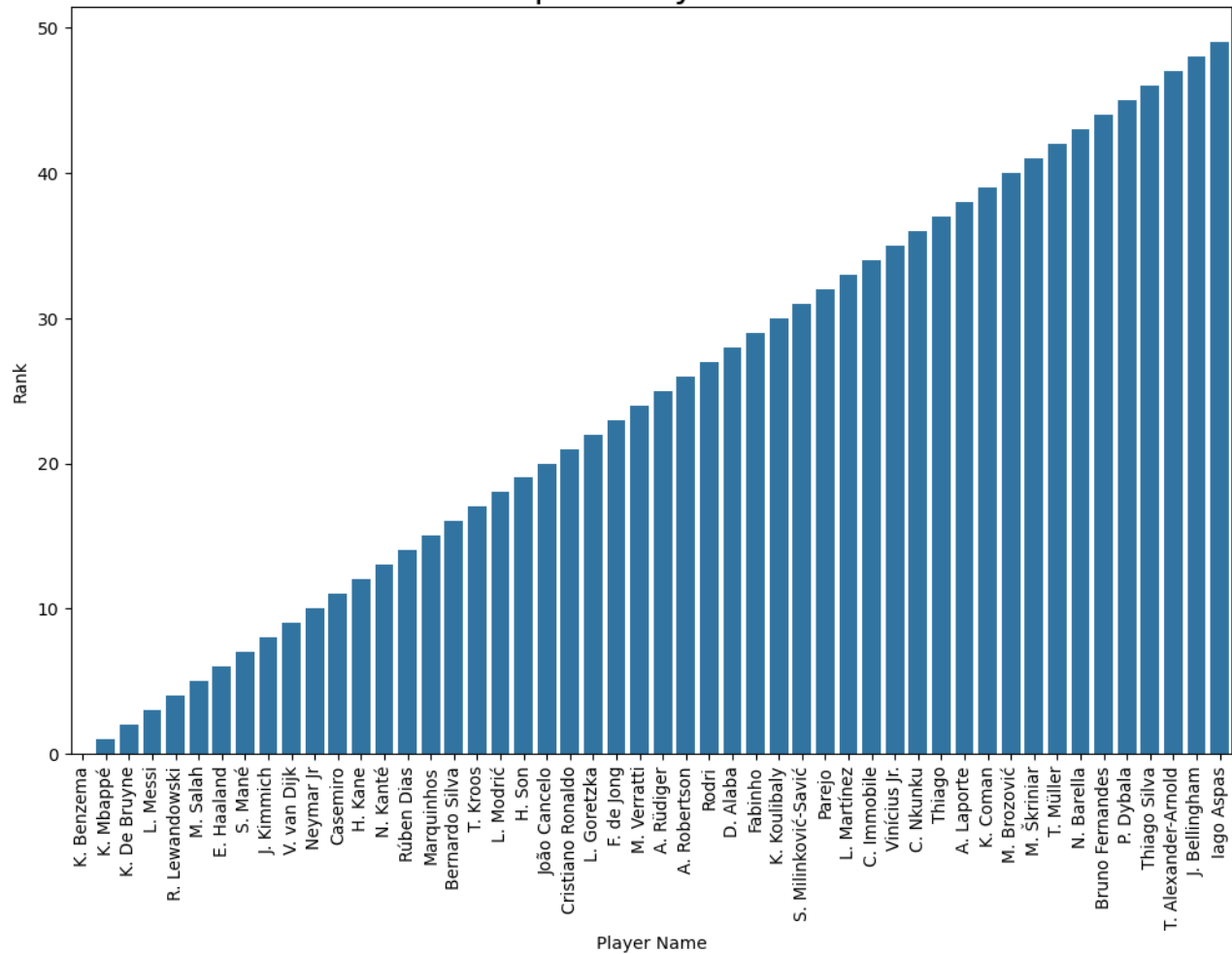


# Nombre de joueur par version de FIFA

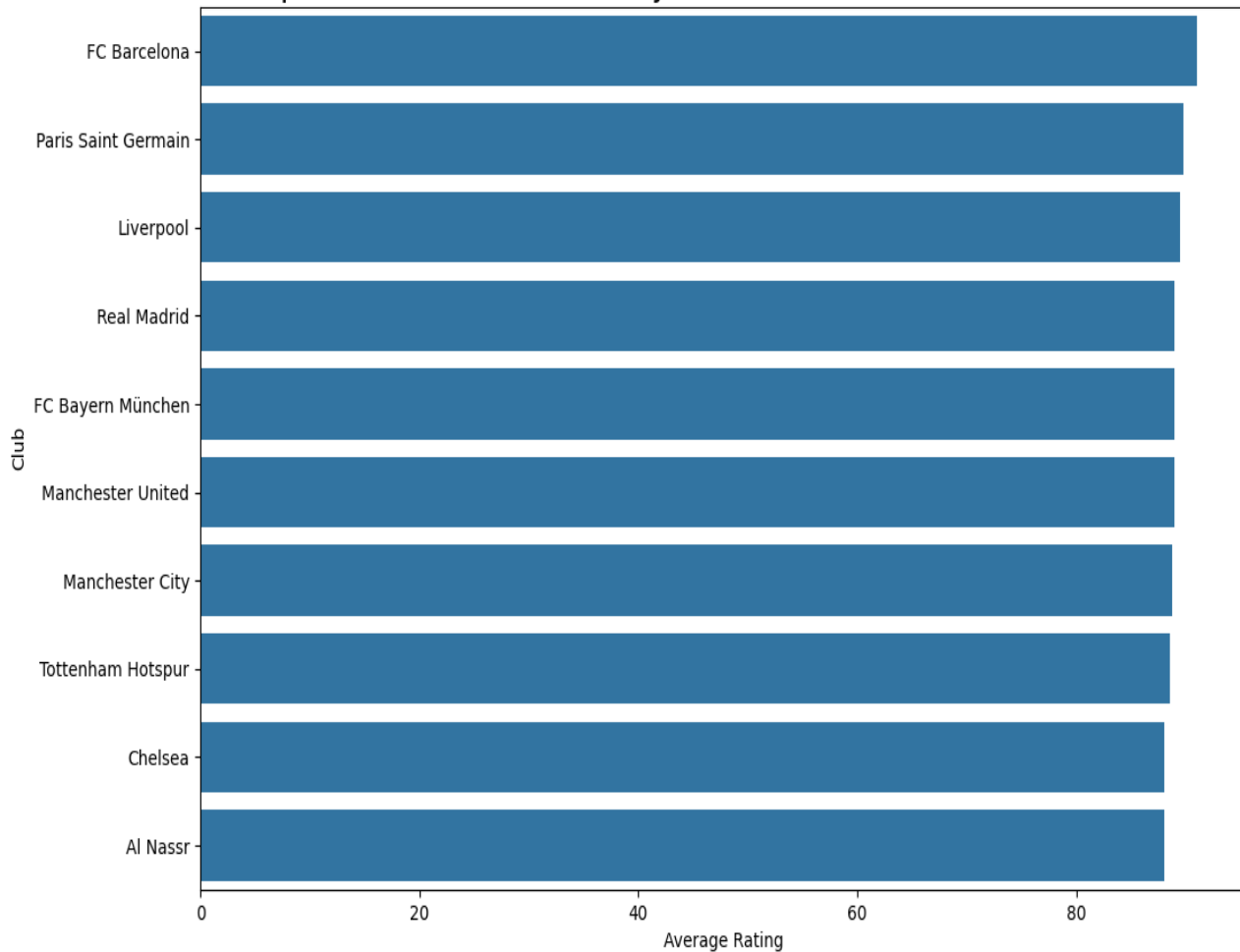


# Voir les 50 meilleurs joueurs et leurs clubs

Top 50 Players 23



Top 10 des clubs avec les joueurs les mieux notés de FIFA 23





# 05 Transfert judicieux





# Contexte



*Avec l'augmentation de la qualité et de la quantité des statistiques disponibles, le football est de plus en plus axé sur les données. Il est intéressant de noter qu'au moins en Premier League, la révolution des données n'a pas le plus grand impact en haut du classement. Alors que le top 6 est resté relativement stable, des clubs comme Brentford et Brighton ont pu exploiter les données pour frapper bien au-dessus de leur poids, Brentford obtenant des finitions confortables au milieu du tableau et Brighton se qualifie pour la compétition européenne pour la première fois de son histoire.*

*Dans cette optique, le public s'intéresse de plus en plus à la manière dont les équipes exploitent les données, et plus de clubs qui n'emploient jamais des directeurs de football pour superviser le recrutement. Bien que l'exactitude de la base de données de la FIFA puisse être discutable, elle constitue une abstraction utile des attributs des joueurs, e j'aimerais l'utiliser dans ce cas pour déterminer ce qui pourrait intéresser un directeur de football*



# Bibliothèque: Lazypredict



*Lazypredict est une bibliothèque Python qui simplifie le processus de formation et d'évaluation de modèles d'apprentissage automatique pour les tâches de classification. Elle fournit une interface simple et intuitive pour entraîner et évaluer un large éventail de modèles, notamment des classificateurs, des régressions et des ensembles.*

# INTRODUCTION DE L'ÉQUIPE



Pour cette analyse, je vais me pencher sur le FC Séville en Liga espagnole. Membre fondateur de la ligue, Séville est l'une des équipes les plus décorées de l'histoire de la ligue, avec 6 coupes d'Europe, 6 coupes de l'UEFA, 4 coupes d'Espagne, 6 supercoupes d'Espagne, 1 coupe du monde des clubs de la FIFA et 3 supercoupes de l'UEFA à son actif. Cependant, le dernier l'équipe passe par une phase tres compliquer actuellement.

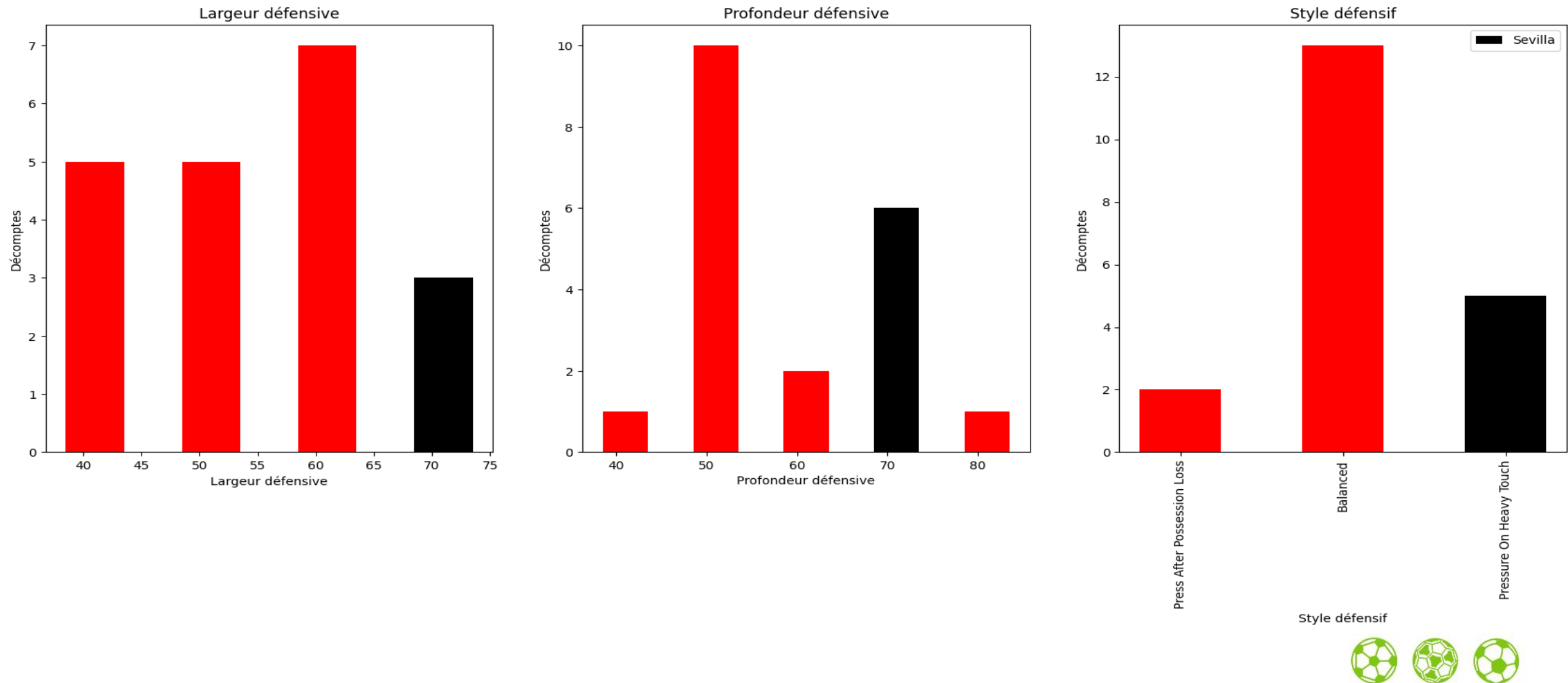


Formation de l'équipe:



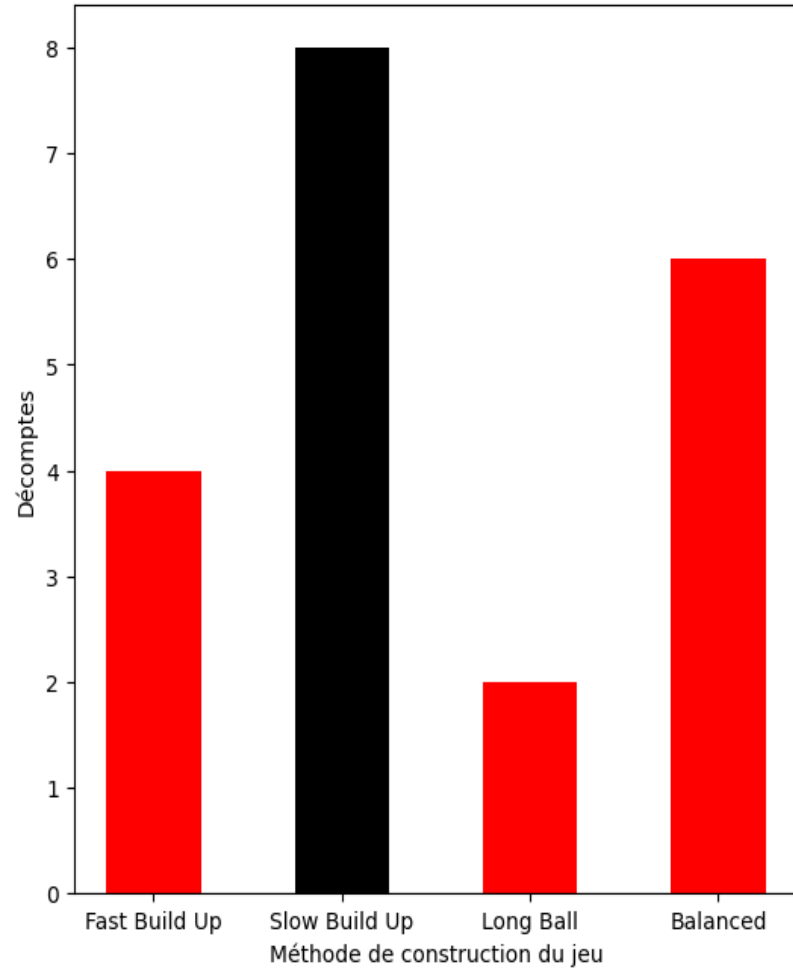


# Comment l'équipe défend ?

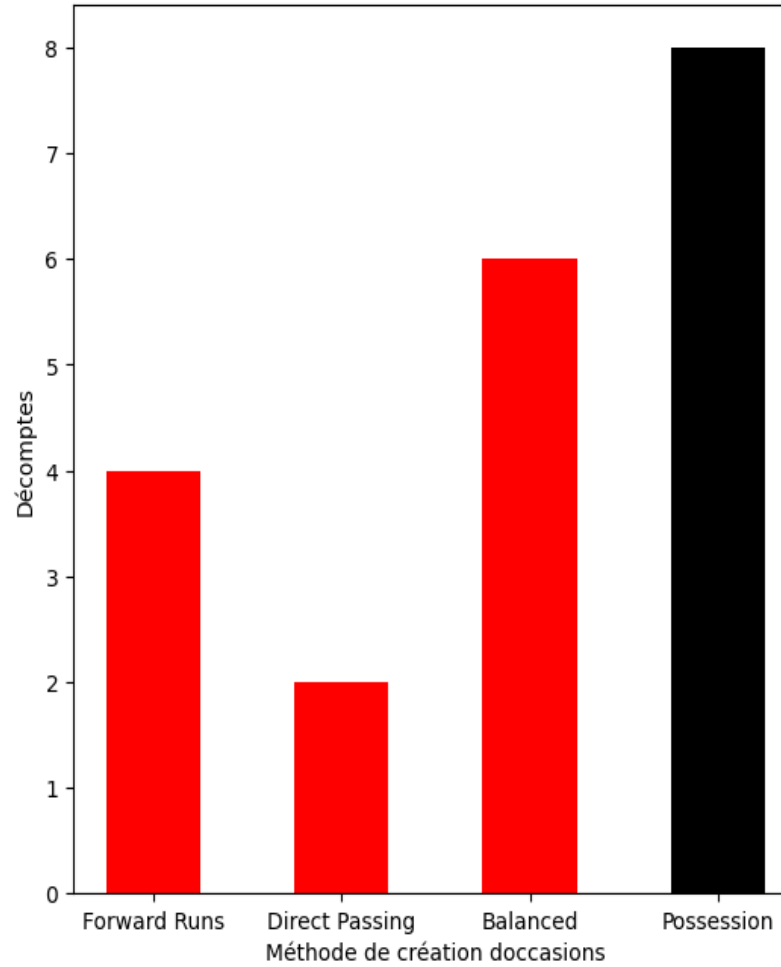


# Comment l'équipe attaque ?

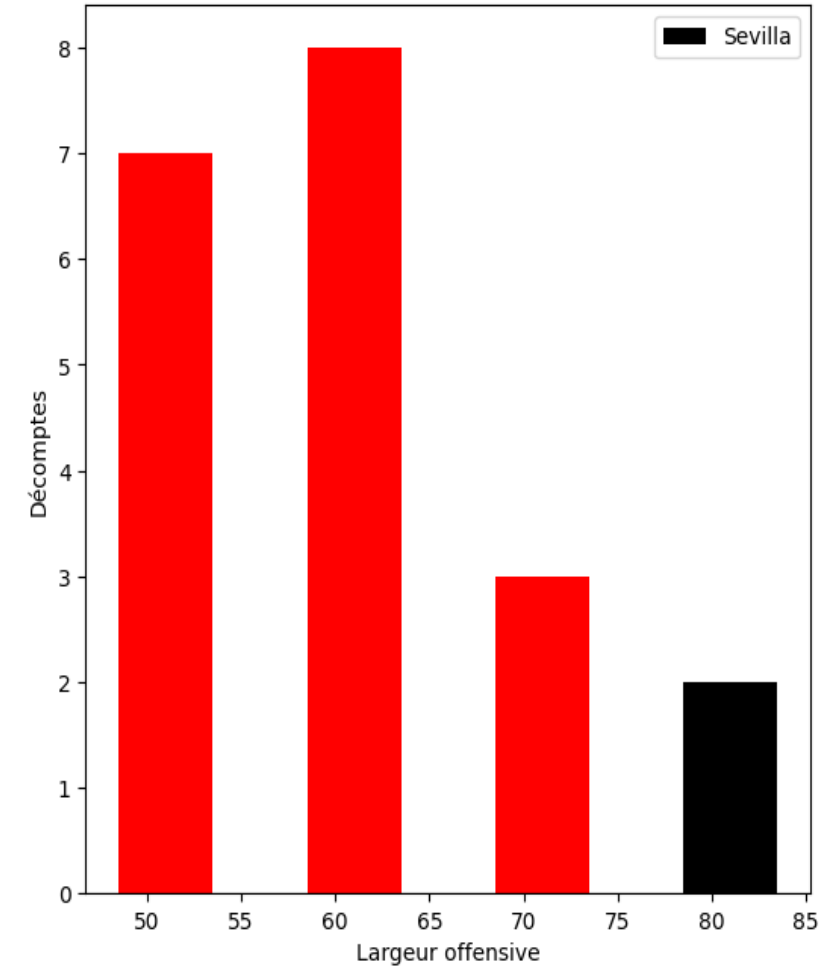
Méthode de construction du jeu



Méthode de création d'occasions



Largeur offensive





*Maintenant que nous en connaissons un peu plus sur l'équipe et sa philosophie de jeu, nous pouvons commencer à explorer des modèles d'entraînement basés sur les attributs des joueurs afin de prédire ceux qui seraient les meilleurs à recruter. Compte tenu des attentes du conseil d'administration, il n'y a pas de pression immédiate pour réussir instantanément. Nous pouvons donc nous permettre de signer des joueurs ayant des évaluations globales plus faibles maintenant, en espérant qu'ils s'intégreront au système et progresseront. Je pense que la note globale d'un joueur dans les futures versions de FIFA est largement déterminée par 3 facteurs principaux :*

- Note globale*
- Âge*
- Potentiel*



100%

<



*Étant donné que la fonctionnalité note globale est volatile et difficile à prédire en raison de la myriade de facteurs qui régissent le développement des joueurs, il peut être plus facile de prédire le développement des joueurs sur une base attribut par attribut. Puisque nous avons établi un profil des joueurs que nous aimerions recruter, nous pouvons prendre ces attributs et faire des pronostics sur eux. Puisque nous avons expliqué que nous aimerions que nos défenseurs jouent des passes longues précises avec une bonne qualité de contrôle de balle et une bonne qualité de passe commençons par prédire comment ces attribut au fil du temps :*

- skill\_long\_passing*
- skill\_ball\_control*
- passing*



100%| 30/30 [01:23<00:00, 2.79s/it]

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
GradientBoostingRegressor	0.81	0.81	4.31	0.56
HistGradientBoostingRegressor	0.81	0.81	4.32	0.20
MLPRegressor	0.81	0.81	4.35	5.42
NuSVR	0.81	0.81	4.36	4.76
ElasticNetCV	0.80	0.80	4.46	0.16
→ LassoCV	0.80	0.80	4.46	0.25
BayesianRidge	0.80	0.80	4.47	0.01
LassoLarsIC	0.80	0.80	4.47	0.02
LarsCV	0.80	0.80	4.47	0.06
Lars	0.80	0.80	4.47	0.02
LassoLarsCV	0.80	0.80	4.47	0.07
OrthogonalMatchingPursuitCV	0.80	0.80	4.47	0.02
LinearRegression	0.80	0.80	4.47	0.01
HuberRegressor	0.79	0.79	4.54	0.03

Model	Adjusted R-Squared	R-Squared	RMSE	Time Taken
HuberRegressor	0.79	0.79	4.54	0.03
LinearSVR	0.79	0.79	4.58	0.06
KNeighborsRegressor	0.78	0.78	4.64	0.03
PassiveAggressiveRegressor	0.78	0.78	4.71	0.02
LassoLars	0.78	0.78	4.73	0.03
Lasso	0.78	0.78	4.73	0.05
BaggingRegressor	0.77	0.77	4.76	0.16
ExtraTreesRegressor	0.77	0.77	4.83	1.21
OrthogonalMatchingPursuit	0.76	0.76	4.90	0.01
AdaBoostRegressor	0.76	0.76	4.94	0.16
ExtraTreeRegressor	0.70	0.70	5.50	0.05
DecisionTreeRegressor	0.69	0.70	5.52	0.04
ElasticNet	0.68	0.68	5.62	0.02
GammaRegressor	0.61	0.62	6.20	0.03
DummyRegressor	-0.00	-0.00	9.99	0.01
KernelRidge	-37.73	-37.70	62.15	18.04
GaussianProcessRegressor	-89.33	-89.24	94.90	52.15

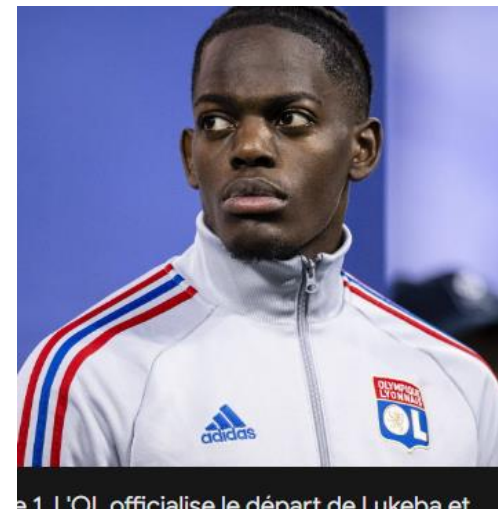
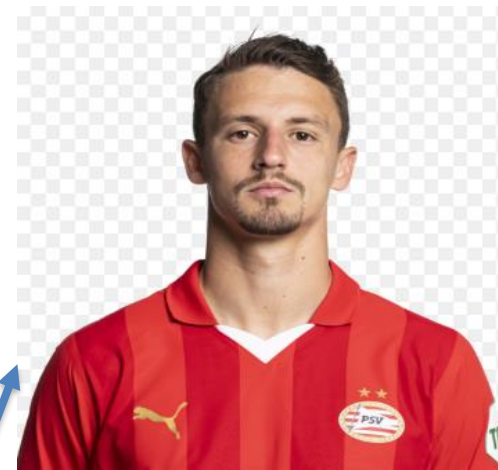




# Prédiction du modèle

```
def_df = def_df[['CB' in def_df['player_positions'].iloc[i] for i in range(def_df['player_positions'].shape[0])]]  
def_df = def_df[(def_df['value_eur'] < 30000000) & (def_df['wage_eur'] < 150000)]  
def_df = def_df[def_df['age'] < 25]  
def_df.head()
```

	short_name	overall	value_eur	wage_eur	age	nationality_name	club_name	player_positions	defender_attributes
5087	F. Dimarco	79	24500000.00	51000.00	24	Italy	Inter	['LWB', 'LM', 'CB']	78.67
12839	O. Boscagli	77	14500000.00	15000.00	24	France	PSV	['CB', 'LB', 'CDM']	76.67
4665	E. Ndicka	80	29500000.00	28000.00	22	France	Eintracht Frankfurt	['CB']	75.00
4904	Eric García	76	14000000.00	67000.00	21	Spain	FC Barcelona	['CB']	74.00
2811	C. Lukeba	76	15000000.00	25000.00	19	France	Olympique Lyonnais	['CB']	71.00





MERCI 😊

