

PROJET : FIFA

TRAVAIL FAIT PAR :

-DOOH DOOH Jeff Williams

SOMMAIRE

I.	CONTEXTE DU PROJET :	3
1.	Description des données :	3
II.	Objectifs :	3
III.	Prétraitement des données	3
1.	Gestion du volume élevé de données et des valeurs dupliquées	3
2.	Gestion des valeurs manquantes	4
3.	Gestion des colonnes catégorielles	4
IV.	Analyse exploratoire des données :	5
1.	L'âge du joueur affecte-t-il sa note générale ?	5
2.	Quel est l'évolution des caractéristiques des Joueurs à travers les versions de FIFA ?	7
3.	Comment la taille affecte différents facteurs du joueur ?	8
4.	Relation entre le salaire et la note des joueurs ?	9
5.	Déterminons-s'il existe une relation entre la position du joueur et son salaire ?	10
6.	Observation des nationalités des 100 joueur payer et celle des mieux des 100 joueur les mieux note ? Que peut-on observer ?	10
7.	Quel sont les nationalités les plus présente dans le dataset ?	11
8.	Voir les 50 meilleurs joueurs et leurs clubs ?	12
V.	Transfert judicieux	13
1.	Contexte	13
2.	Bibliothèque: Lazypredict	13
3.	INTRODUCTION DE L'EQUIPE	14
a)	Comment l'équipe défend ?	15
b)	Comment l'équipe attaque ?	16
VI.	Prédiction du modèle	20
VII.	Conclusion	Error! Bookmark not defined.

I. CONTEXTE DU PROJET :

Des clubs de football professionnel, qui veulent rivaliser avec les meilleurs clubs. Les conseils d'administration de ces 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 peuvent concourir afin qu'ils puissent créer une équipe forte et s'assurer que chaque joueur jouera et évoluera efficacement dans le club.

Nous allons travailler sur un dataset Kagel qui contient des joueurs FIFA de 2015 à 2023

1. Description des données :

Les données contiennent :

- Tous les joueurs disponibles dans FIFA de 2015 à 2023
- 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, le salaire, etc.

II. Objectifs :

- 1- Aider le conseil d'administration du club à connaître les meilleurs joueurs dans les différents clubs.
- 2- Les aider à comprendre les clubs de leurs concurrents.
- 3- Connaître les compétences qui doivent être chez leurs joueurs.
- 4- Les aider sur la gestion d'un budget lors de la période de mercato. (Définition : la période de mercato est une période dans l'année ou les clubs peuvent acheter, vendre et prêter des joueurs dans le but d'atteindre un objectif)

III. Prétraitement des données

1. Gestion du volume élevé de données et des valeurs dupliquées

Nous avons divisé le dataset en fonction de chaque version de FIFA avec la variable fifa_version. Nous avons pris en charge les joueurs qui reviennent plusieurs fois dans chaque nouveau petit dataset, ces joueurs reviennent plusieurs fois parce que dans un jeux FIFA il y as plusieurs mises à jour ceci est montré par la variable fifa_update Nous avons choisi de ne retenir que les derniere version de chaque joueur.

2. Gestion des valeurs manquantes

Premièrement grâce à une analyse préalable et méthodique des 110 variable nous avons fait une sélection de variable basse sur la connaissance des variables et nos objectifs dans ce projet des variables comme player_traits et goalkeeping speed ont été supprimer nous somme ainsi passe de 110 à 68 variables ce qui nous a faciliter la gestion des valeurs manquante.

Ceci étant fait on a remarqué que dans chaque FIFA les variables valeur_eur, wage_eur, league_name et club_name qui représente respectivement la valeur, le salaire, le nom de la ligue et le nom du club du joueur on le même pourcentage de valeur manquante qui est tres faible (<3%). Il s'agit surement des joueur retraite on a donc décidé de supprimer les ligne avec les valeurs manquante.

On a aussi des variables telles que pace, dribbling, defending, physic, passing et shooting qui représentent respectivement la vitesse, le dribble la défense le physique et la passe qui elles aussi dans toutes les versions de FIFA avaient des valeurs manquantes sensiblement égales (environ 10 %) avec plus de recherche nous nous sommes rendu compte qu'il s'agissait des gardiens nous avons décidé de les stocker dans un autre dataset car nous allons que gérer les joueurs de champ dans ce projet. Plutôt que de les supprimer, nous les conservons car cela pourrait être nécessaire de travailler avec les gardiens à l'avenir.

3. Gestion des colonnes catégorielles

Nous avons utilisé LabelEncoder de sklearn sur la variable preferred_foot pour remplacer gauche/droite par 1/0 ce qui transforme cette variable catégorielle en variable numérique.

La variable player_position pose problème car il y a des joueurs qui joue à plusieurs postes ce qui nous posera un problème du au gros volume de valeur unique pour gérer ce problème nous avons utilisé le One-Hot Encoding. L'encodage one-hot ou encodage 1 parmi n consiste à encoder une variable à n états sur n bits dont un seul prend la valeur 1, le numéro du bit valant 1 étant le numéro de l'état pris par la variable. En gros cette technique crée plusieurs nouvelles colonnes ayant pour titre chaque poste ces colonnes sont constituer de 0 et de 1, 0 a pour signification le joueur ne joue pas à ce poste et 1 a pour signification le joueur joue à ce poste

Il y a aussi des variables tel que ls, st, rs, lw etc... qui représente le niveau d'un joueur a diffèrent poste elles sont sous le format objet car elles sont sous la forme 'note (+/-) ajustement' nous avons décidé d'appliquer l'ajustement à la note cette à dire appliquer l'addition ou la soustraction pour avoir la valeur finale.

IV. Analyse exploratoire des données :

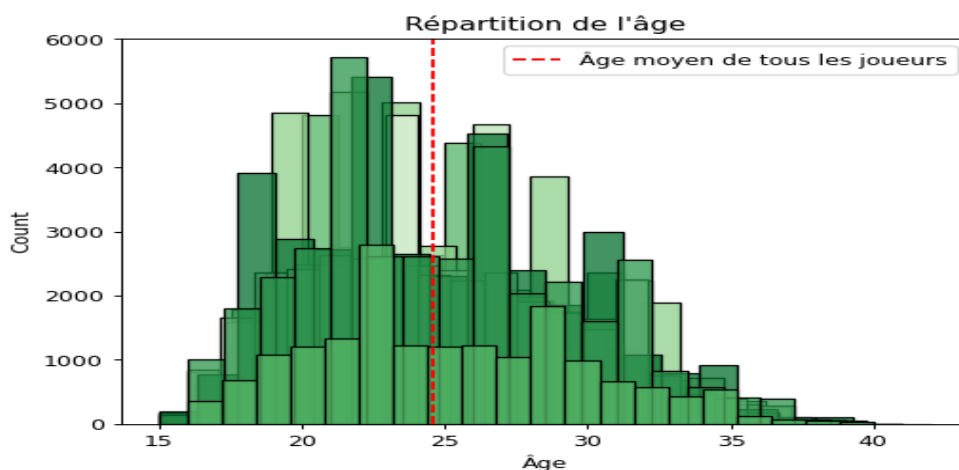
Cette étape est celle où nous allons faire de l'analyse de donnée qui nous permettra de répondre à certaine question

Questions auxquelles nous allons répondre :

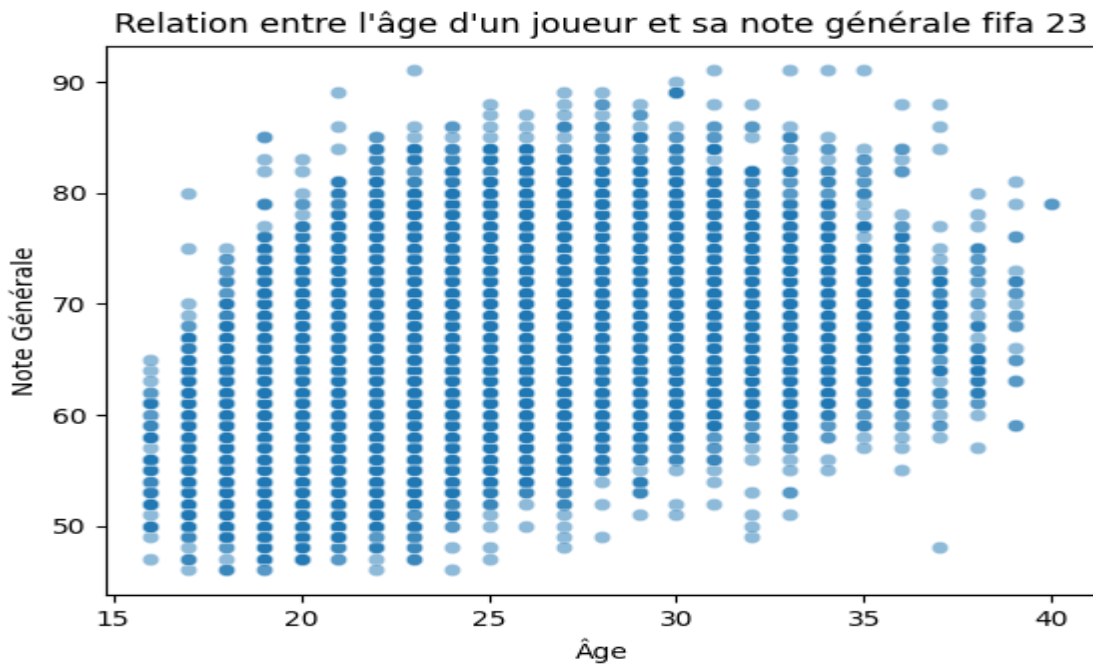
- L'âge du joueur affecte-t-il sa note générale ?
- Quel est l'évolution des caractéristiques des Joueurs à travers les Versions de FIFA ?
- Comment la taille affecte différents facteurs du joueur ?
- Relation entre le salaire et la note des joueurs ?
- Déterminons-s'il existe une relation entre la position du joueur et son salaire
- Observation des nationalités des 100 joueur payer et celle des mieux des 100 joueur les mieux note ?
- Que peut-on observer ?
- Quel sont les nationalités les plus présente dans le dataset ?
- Voir les 50 meilleurs joueurs et leurs clubs ?

1. L'âge du joueur affecte-t-il sa note générale ?

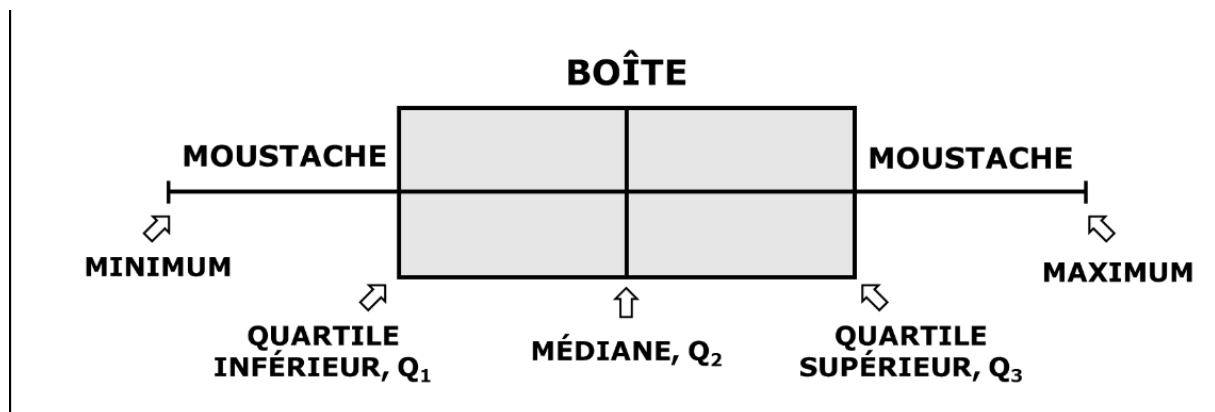
Premièrement nous avons d'abord visualiser la répartition de l'âge

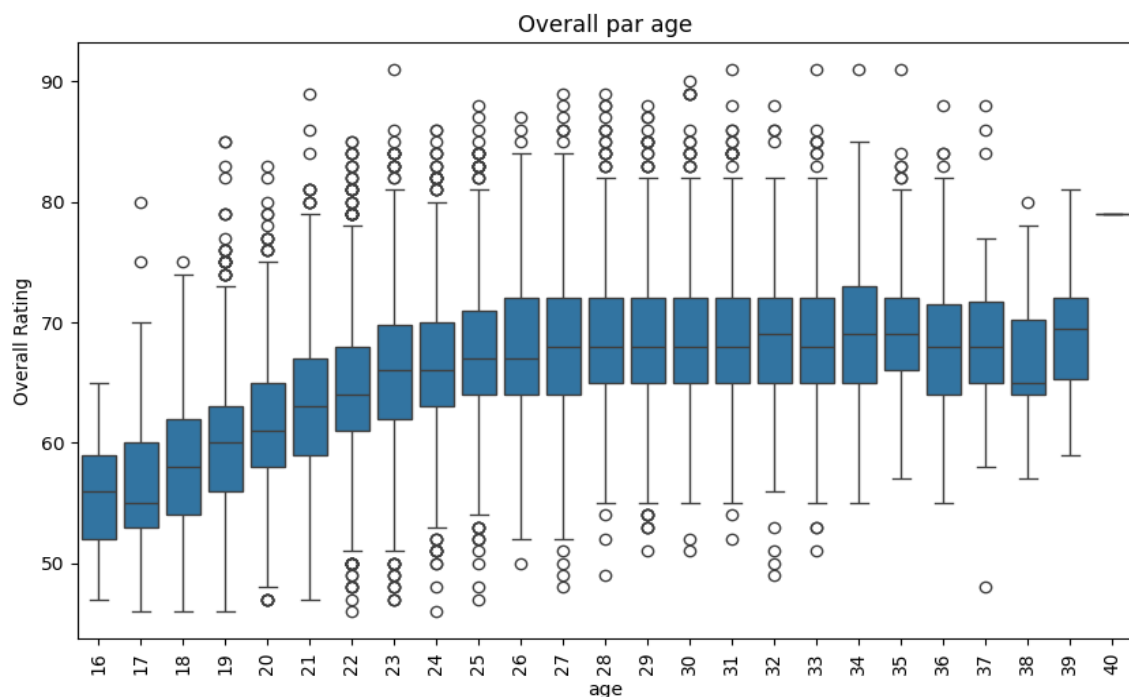


On se rend compte que l'âge moyen de tout le joueur est sensiblement égal à 25 ans pour répondre à notre question nous utiliserons un nuage de point. Un nuage de points est grandement utilisé pour présenter la mesure de deux ou plusieurs variables liées. Le nuage de points est particulièrement utile lorsque les valeurs des variables sur l'axe des y dépendent des valeurs de la variable de l'axe des x.



On se rend compte que le nuage de point n'est pas très explicite avec une corrélation moyenne nous allons donc essayer un autre graphe soit les boîtes à moustache. Une boîte à moustaches montre la distribution des données pour une variable continue. la boîte à moustache a plusieurs caractéristiques intéressantes comme nous le montre cette image mais celle qui nous intéresse est la médiane (La médiane est le point milieu d'un jeu de données, de sorte que 50 % des unités ont une valeur inférieure ou égale à la médiane et 50 % des unités ont une valeur supérieure ou égale.) qui compare à la moyenne est beaucoup moins sensible aux valeurs aberrantes

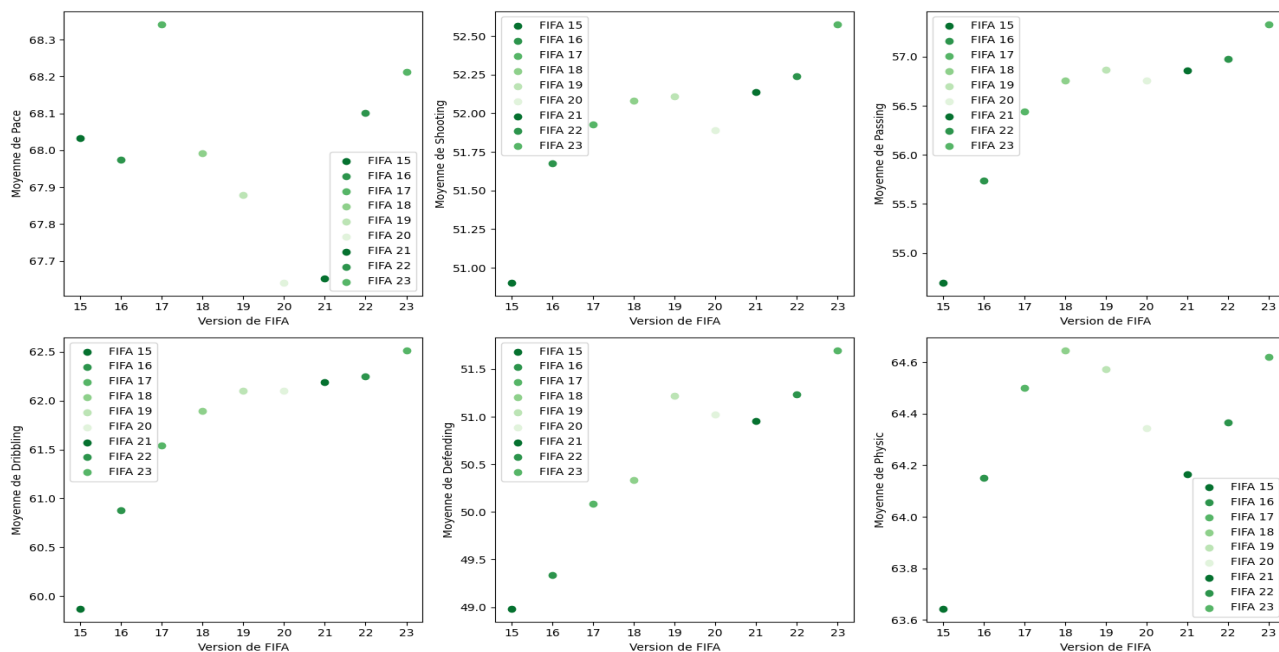




Avec ce graphique plus explicite on voit l'évolution de la valeur de la médiane en fonction de l'âge ce qui nous permet de conclure qu'après 26 ans le niveau se stabilise. Il pourrait donc être plus judicieux de recruter un joueur de moins de 25 ans comme cela on a un joueur qui est moins chère et qui évoluera sûrement et s'il n'atteint pas le niveau requis nous pourrions le revendre dans l'objectif de faire une plus-value. (Relation âge et salaire sera montrée à la question 4)

2. Quel est l'évolution des caractéristiques des Joueurs à travers les versions de FIFA ?

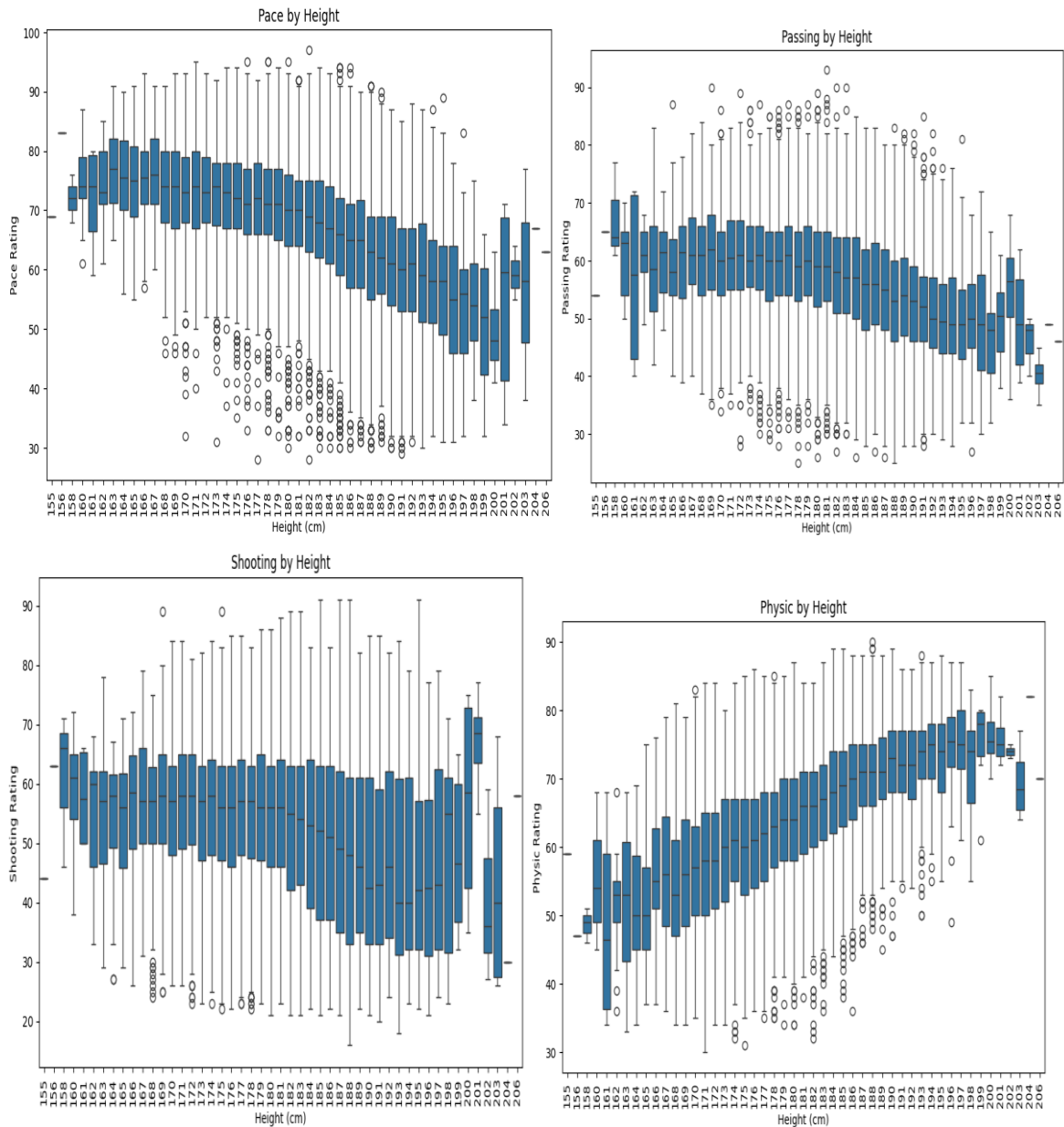
Évolution des Caractéristiques des Joueurs à travers les Versions de FIFA

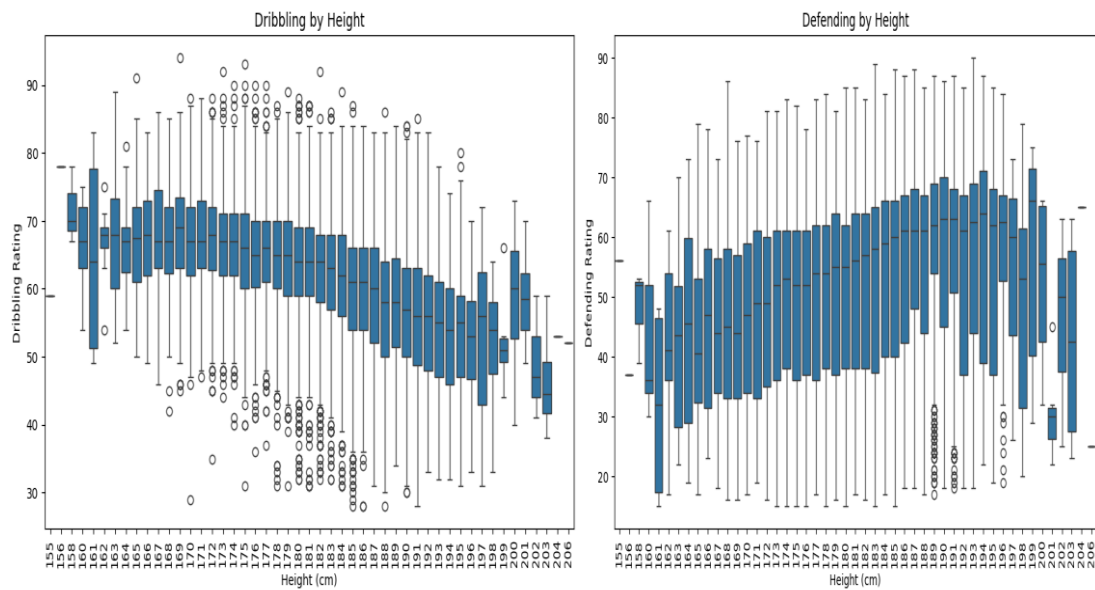


Avec ce graphe on observe l'évolutions des caractéristiques en fonction des FIFA on se rend vite compte que toutes les caractéristiques ont évoluer avec le temps ce qui peut sous-entendre que plus le foot évolue plus les joueurs sont performant ceci peut être du a de meilleur entrainement de meilleure connaissance tactique et technique dans le domaine du foot en général.

3. Comment la taille affecte différents facteurs du joueur ?

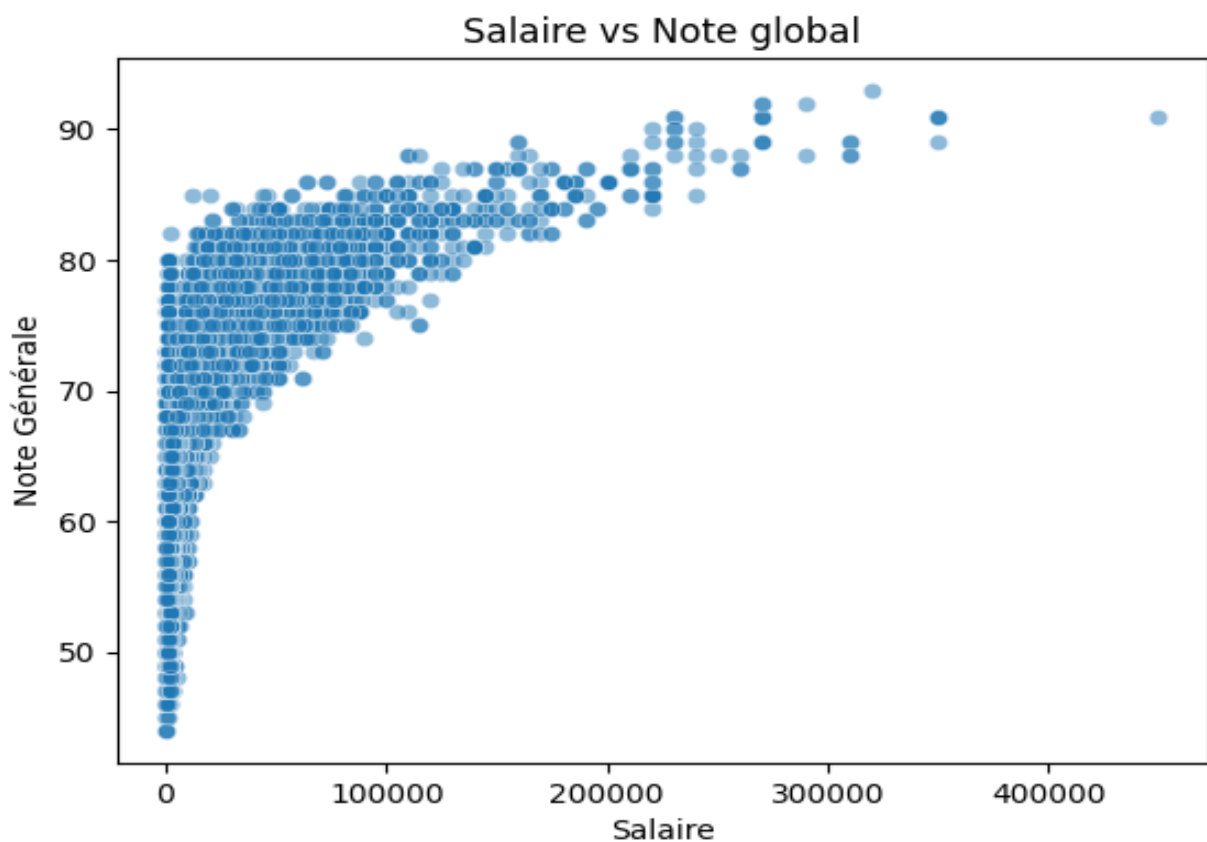
Pour répondre à cette question nous allons utiliser le même graphe qu'a la question 1





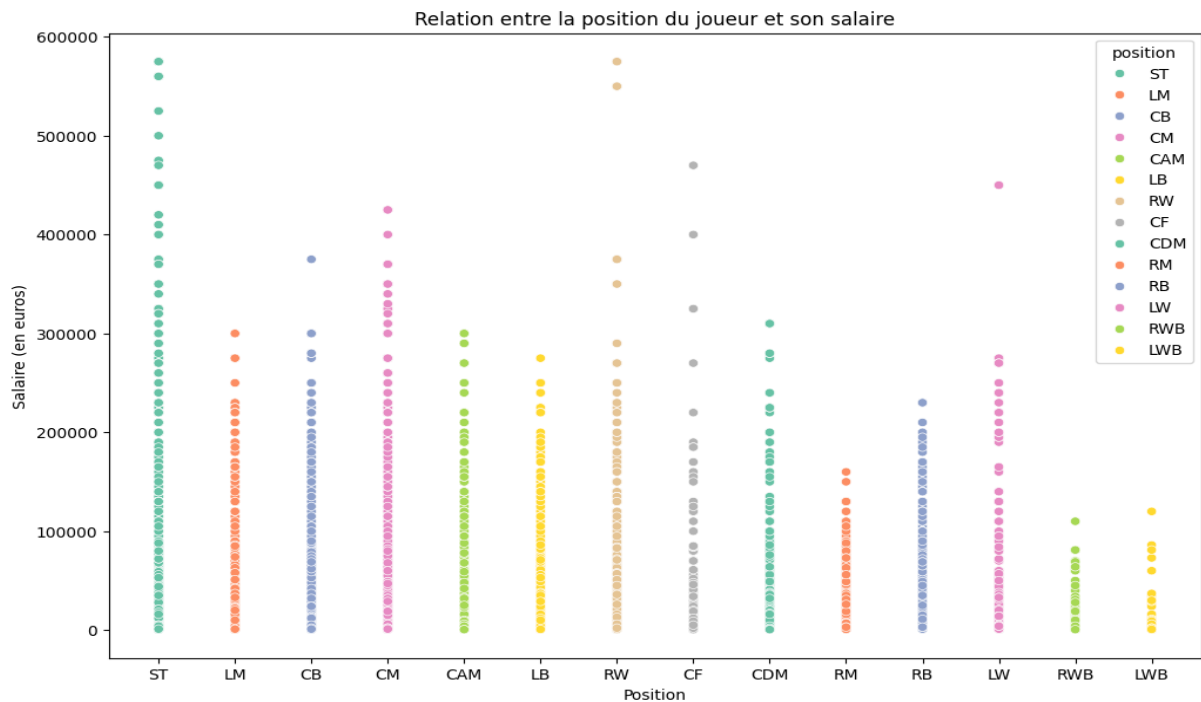
On se rend compte que des fait logique tel que plus un joueur est grand moins il est rapide, moins il a de bonne capacite en drible et inversement plus un joueur est grand plus il est physique et plus il a de bonnes compétences en défense sont démontré par ces graphes.

4. Relation entre le salaire et la note des joueurs ?



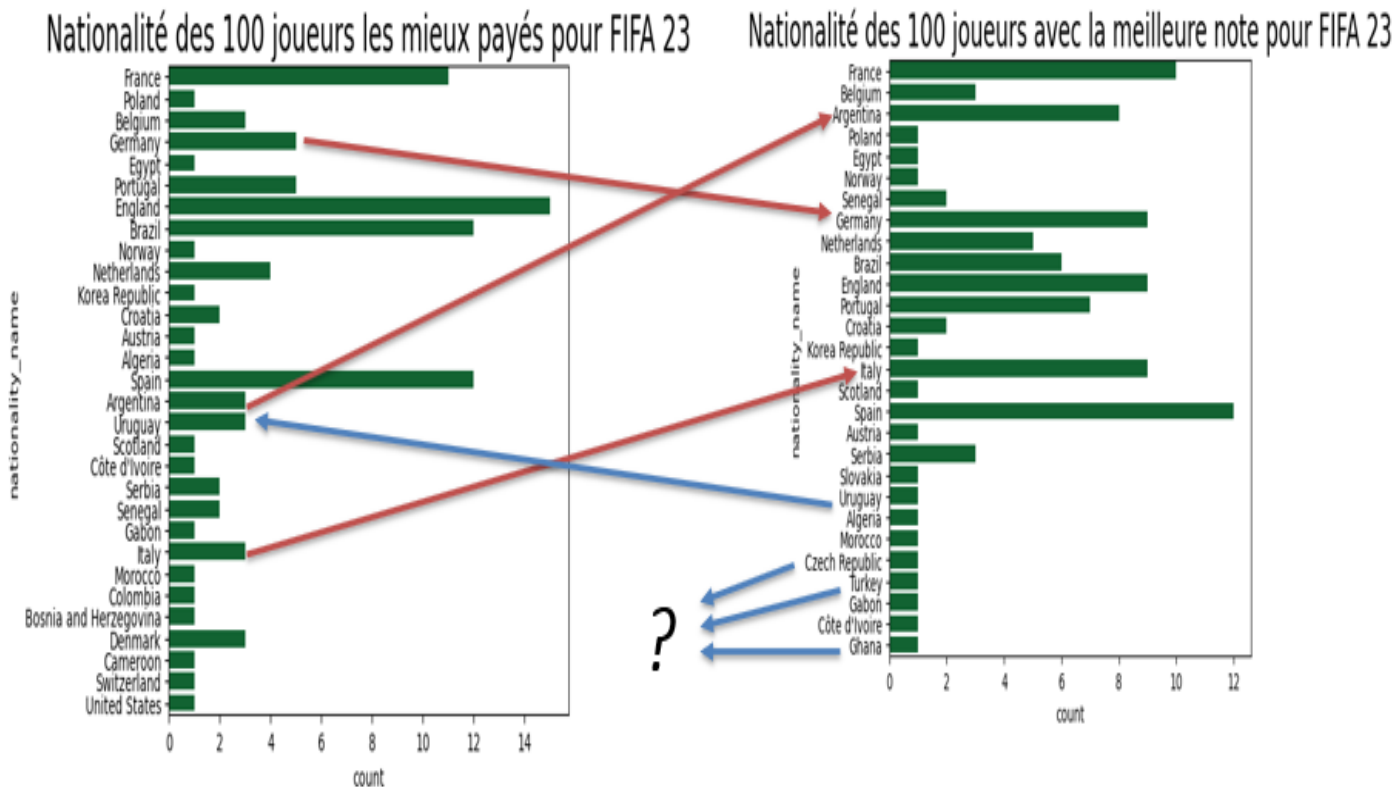
Dans ce cas le nuage de point est suffisamment explicite pour conclure que plus un joueur est bien note plus il est performant d'où la conclusion faites à la question 2

5. Déterminons-s'il existe une relation entre la position du joueur et son salaire ?



Nous pouvons ainsi observer que les attaquants sont les mieux payés, nous pouvons ainsi conclure que pour un club qui veut acheter un attaquant, il doit prévoir un budget plus conséquent.

6. Observation des nationalités des 100 joueurs les mieux payés et celle des 100 joueurs les mieux notés ? Que peut-on observer ?

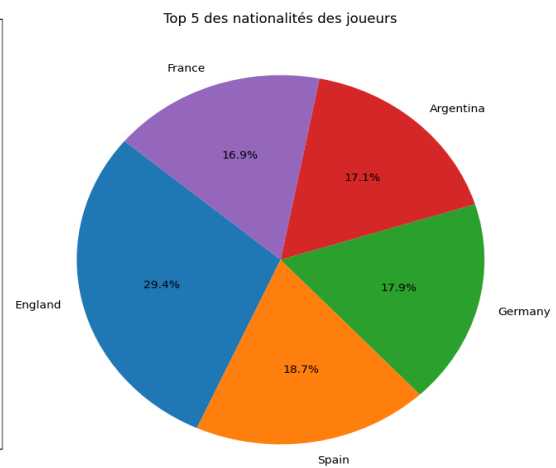
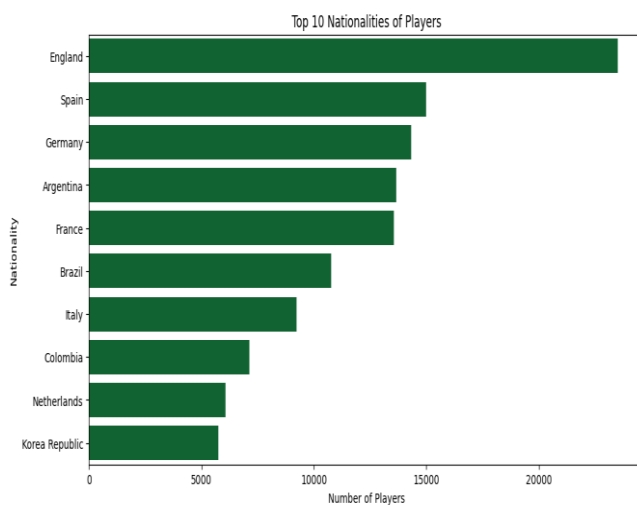


On peut remarquer que les pays tel que la Belgique l'argentine et l'Italie ont moins de joueur dans le top 100 joueur les mieux paye que dans le top 100 joueur les mieux note ceci peut sous-entendre que les argentin, les belge et les italiens sont sous évalué sur le marché cette à dire qu'à compétences égale vaut moins chère qu'un français par exemple ce qui pour un petit budget est une énorme opportunité

On peut aussi inversement remarquer que des pays tel que Uruguay et le Ghana ont moins de joueur dans le top 100 joueur les mieux note que dans le top 100 joueur les mieux paye peut sous-entendre que les joueurs de ces pays sont surévalués sur le marché cette à dire qu'à compétences égale ils valent plus chère

ATTENTION : ce ne sont pas des conclusions justes des hypothèses qui reste à confirmer avec plus d'études

7. Quel sont les nationalités les plus présente dans le dataset ?

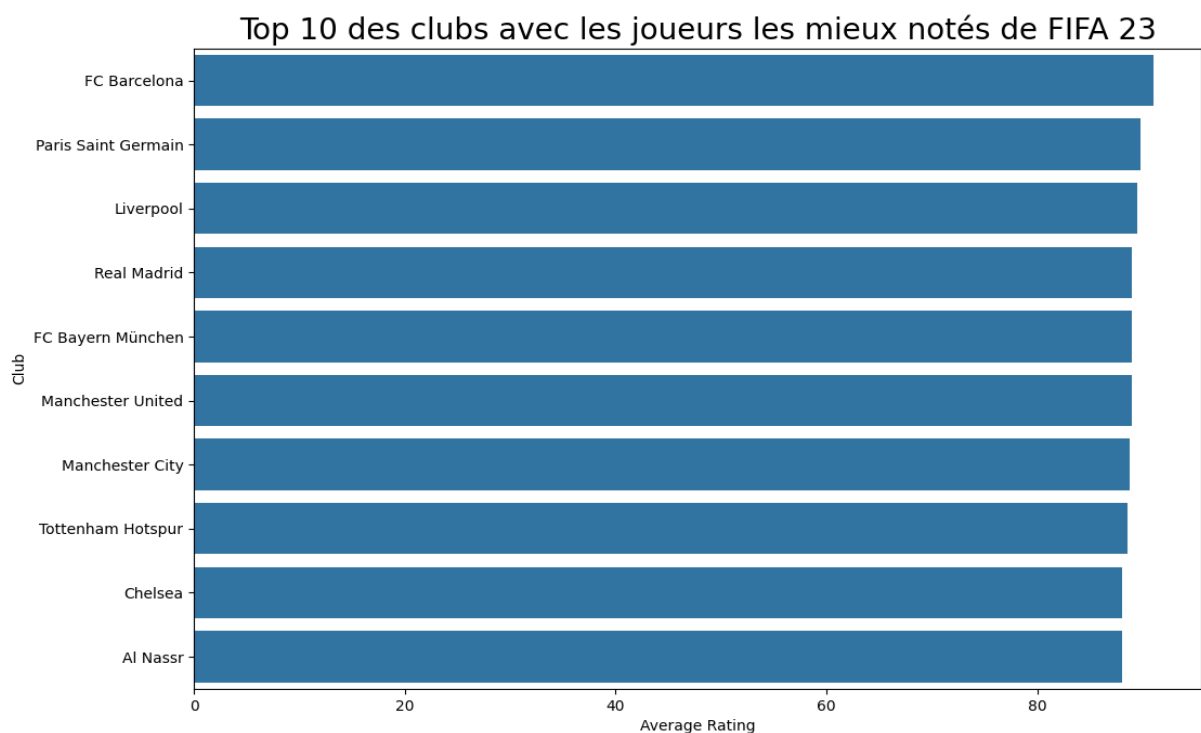
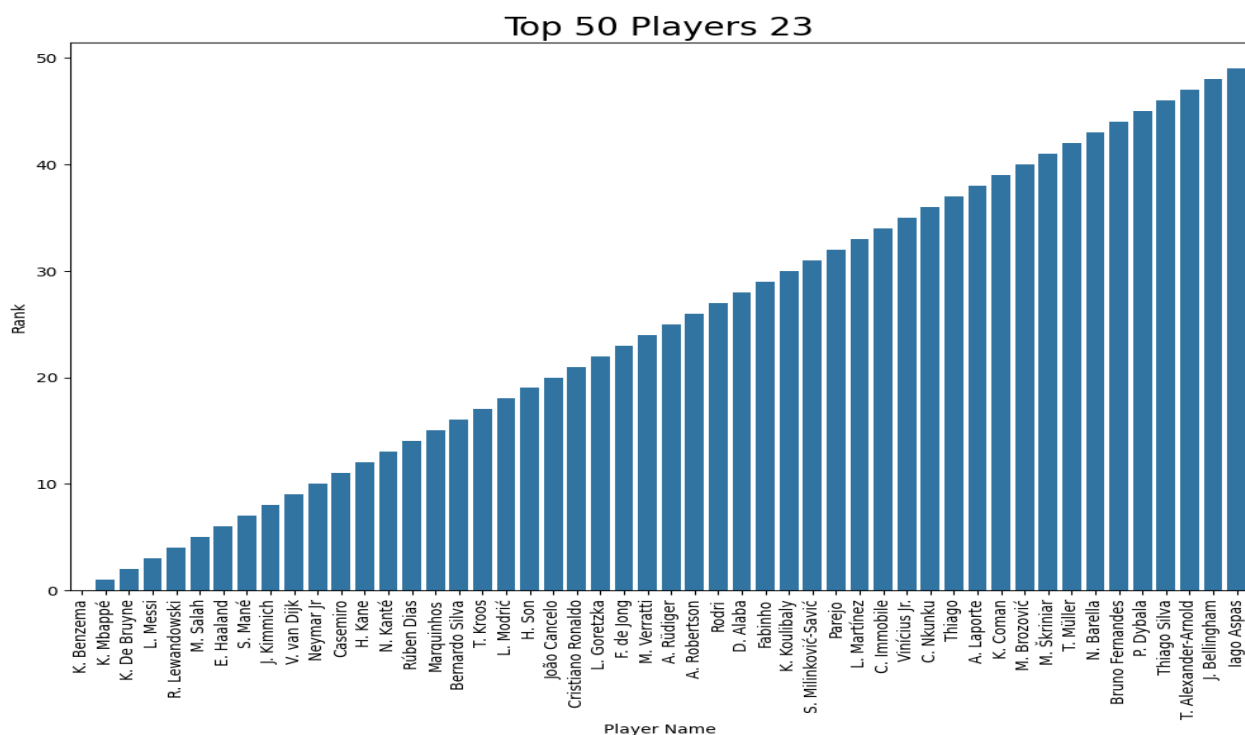


Les pays les plus présent dans le dataset sont :

- L'Angleterre
- L'Espagne
- L'Allemagne
- L'Argentine
- La France
- Le Brésil

8. Voir les 50 meilleurs joueurs et leurs clubs ?

Observons les 50 meilleurs joueurs et leurs clubs



V. Transfert judicieux

1. 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, et j'aimerais l'utiliser dans ce cas pour déterminer ce qui pourrait intéresser un directeur de football

2. 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.

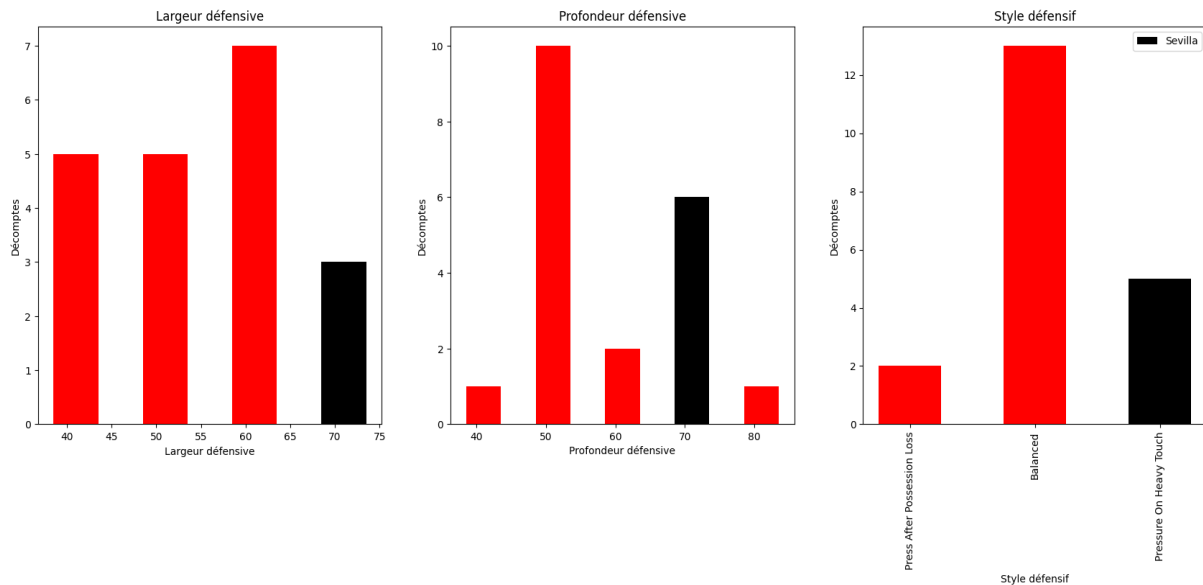
3. INTRODUCTION DE L'EQUIPE



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 Supercoupe d'Espagne, 1 coupe du monde des clubs de la FIFA et 3 Supercoupe de l'UEFA à son actif. Cependant, le dernier l'équipe passe par une phase très compliquer actuellement. Cette équipe joue en 4-3-3



a) Comment l'équipe défend ?



1. Graphique de la Largeur Défensive :

2.

Ce graphique présente la distribution de la largeur défensive pour toutes les équipes de La Liga (en rouge). Une barre noire est ajoutée pour représenter la valeur de la largeur défensive spécifique au FC Séville. Cette barre est mise en évidence par rapport aux autres équipes.

3. Graphique de la Profondeur Défensive :

De manière similaire, ce graphique illustre la distribution de la profondeur défensive pour l'ensemble des équipes de La Liga. Une barre noire représente la profondeur défensive spécifique au FC Séville, mise en évidence par rapport aux autres équipes.

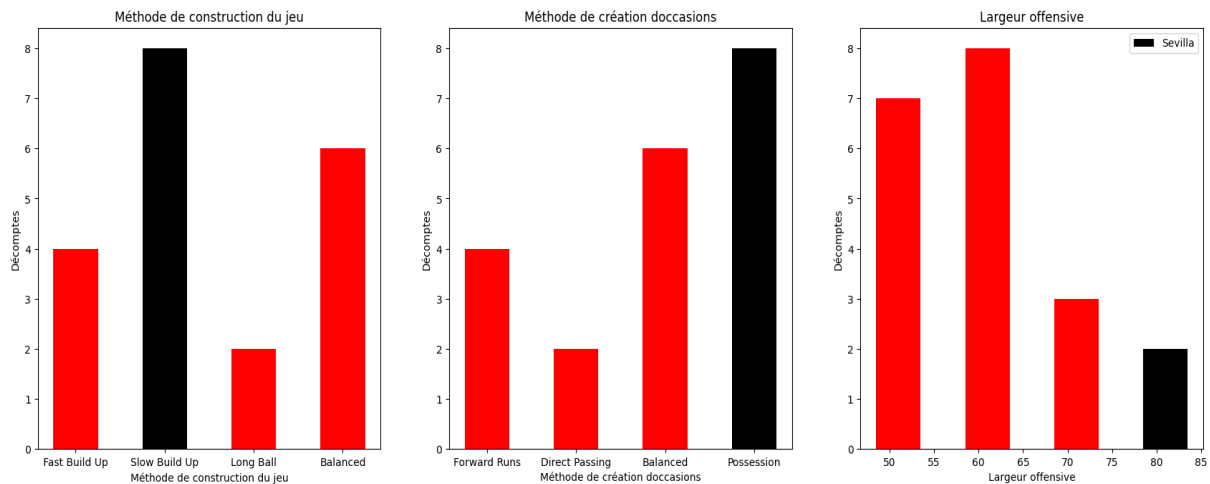
4. Graphique du Style Défensif :

Ce graphique affiche la distribution du style défensif pour l'ensemble des équipes de La Liga. Une barre noire est ajoutée pour indiquer la valeur du style défensif spécifique au FC Séville, mettant en évidence sa position par rapport aux autres équipes.

Légende :

Les barres rouges représentent la distribution des valeurs pour toutes les équipes de La Liga. La barre noire représente la valeur spécifique au FC Séville, mise en évidence dans chaque graphique. Ces graphiques permettent de comparer visuellement les caractéristiques défensives du FC Séville par rapport aux autres équipes de La Liga, pour les aspects de largeur, profondeur et style défensif.

b) Comment l'équipe attaque ?



En réunissant ces informations, nous pouvons commencer à former une image du style de jeu souhaité par le FC Séville. Cela nous permettra également d'établir un profil du joueur que nous souhaitons recruter :

Avec ou sans ballon, le FC Séville adopte un positionnement plutôt large. Nos joueurs devront donc être capables de contrôler et de circuler le ballon dans des zones où l'espace est important. L'équipe se crée des occasions en jouant la possession du ballon et en utilisant des passes courtes et précises pour créer des décalages. Les attaquants devront être techniques, dotés d'une bonne vision du jeu et capables de se créer des occasions en un contre un.

L'équipe a pour volonté de ne pas jouer dans sa propre moitié de terrain. Les défenseurs devront donc être capables de jouer le ballon long avec précision et d'offrir des options de relance. En tenant compte de ces éléments, voici un profil type du joueur recherché par le FC Séville :

- Joueur technique et à l'aise dans les grands espaces.
- Bon passeur et capable de créer des occasions pour ses coéquipiers.
- Rapide et doté d'une bonne endurance pour participer aux contres et à la possession du ballon.
- Bon jeu de tête pour les ballons longs et les situations de jeu aérien.
- Défenseur précis dans ses passes longues et bon relanceur.
- En recrutant des joueurs correspondant à ce profil, le FC Séville pourra mettre en place son style de jeu et atteindre ses objectifs.

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(overall)
- Âge(age)
- Potentiel(potential)

100%

</

Après avoir entraîné le modèle initial, nous obtenons un résultat quelque peu inattendu. Les caractéristiques que nous avons présentées comme étant importantes pour déterminer la progression d'un joueur ne semblent pas être une bonne mesure de tout. Cela peut s'expliquer par plusieurs raisons. La croissance des joueurs peut être non linéaire, ce qui signifie que le modèle linéaire que nous avons choisi ne peut pas facilement prendre en compte les progressions irrégulières. De plus, nous avons supposé que la caractéristique "potentiel" est une mesure précise de la capacité ultime d'un joueur, ce qui n'est peut-être pas le cas

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 attributs au fil du temps :

-skill_long_passing
-skill_ball_control
-passing

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

CO MP TE_

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



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

Dès le départ, ce modèle semble beaucoup plus utilisable. Passons les données de FIFA 23 au modèle et voyons comment il prédit la croissance future des joueurs.

VI. 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



Pour un défenseur âgé de moins de 25 ans avec une valeur inférieure à 30.000.000 et un salaire inférieur à 150.000 on obtient plusieurs joueurs intéressants dont deux qui sont Lukeba et Boscagli

Lukeba est un jeune français de 19 ans qui joue à l'olympique lyonnais qui malgré son jeune âge a déjà un très bon niveau et aura sûrement une très bonne marge de progression

Boscagli aussi est un jeune français de 24 ans qui joue au PSV ayant lui aussi un bon niveau il est en plus de sa polyvalence il peut jouer à plusieurs postes