



Alexandre Mélanger
Auren Moyo Kamdem
Paul Gravejal
Promo : 2025
SCIA-G

Rapport Data visualization

Foot and stats



Foot and stats

"Ce qui compte ne peut pas toujours être compté, et ce qui peut être compté ne compte pas forcément." Cet aphorisme d'Albert Einstein offre une perspective précieuse sur l'importance des données dans le contexte du football. La révolution des données a non seulement libéré les mesures quantitatives simples telles que les tirs, les passes et les dribbles, mais elle a également révélé l'étendue de ce qui échappe encore aux algorithmes. Dans le cadre de notre projet réalisé à trois, nous avons choisi de nous pencher sur les statistiques dans le football. Nous avons identifié deux ensembles de données pour notre analyse : le premier provient de la plateforme understat.com, offrant des analyses détaillées de chaque match des meilleures ligues européennes, tandis que le deuxième est tiré du site Kaggle, fournissant des statistiques sur les meilleurs joueurs sélectionnés sur une période de cinq ans [2016-2020]. Ce rapport vise à présenter les étapes et les méthodologies suivies pour répondre à chaque aspect du projet, ainsi que les résultats obtenus et les implications pratiques de nos analyses.

I. Analyse de la performance des équipes en fonction de leur capacité à convertir les occasions en buts :

Ici la question qu'on a laquelle on n'a essayé de répondre est Les équipes qui marquent plus de buts que leurs expected goals sont-elles généralement plus performantes dans différentes ligues UEFA au fil des saisons ?

II. Analyse de l'impact du pressing défensif sur la performance des équipes :

Ici la question qu'on a laquelle on n'a essayé de répondre est Les équipes qui montrent une activité et une intensité défensives plus élevées dans le camp adverse remportent-elles généralement plus de matches, et marquent-elles plus de buts ?

III. Analyse de l'efficacité des joueurs dans la conversion des occasions en buts :

Ici la question qu'on a laquelle on n'a essayé de répondre est Les joueurs ayant des expected goals (xG) plus élevés ont-ils tendance à marquer plus de buts dans différentes ligues.

IV. Analyse de la corrélation entre la performance des joueurs et celle de leur équipe:

Ici la question qu'on a laquelle on n'a essayé de répondre est Existe-t-il une corrélation significative entre l'efficacité des joueurs dans la conversion des expected goals (xG) en buts réels et les résultats globaux de leur équipe dans chaque ligue ?

V. Analyse de l'impact de la pression défensive sur la performance des équipes en fonction de la qualité des joueurs :

Les équipes capables d'exercer une pression défensive plus élevée, tout en ayant des joueurs de qualité supérieure, sont-elles généralement plus performantes en termes de conversion des expected goals (xG) en buts réels ?

La donnée

Pour effectuer des analyses sur les différents éléments d'étude, nous utiliserons deux jeux de donnée provenant du site kaggle: "players", contenant des informations et statistiques de 661 joueurs professionnels entre 2016 et 2020. Ainsi que "understat" qui contient des informations sur 114 équipes entre 2014 et 2019. Nous limiterons nos études entre 2016 et 2019 pour corrélérer ces données.

La donnée est directement très exploitable, après quelque recherches, il n'y a aucune valeur manquante ou problématique, les différents calculs potentiels entre les colonnes ont déjà été réalisés et nous permette de nous lancer rapidement dans l'analyse. Le seul changement nécessaire est le passage des noms de clubs sous forme d'abréviation dans le jeu de donnée en noms complets à l'aide d'un dictionnaire.

I. Analyse de la performance des équipes en fonction de leur capacité à convertir les occasions en buts :

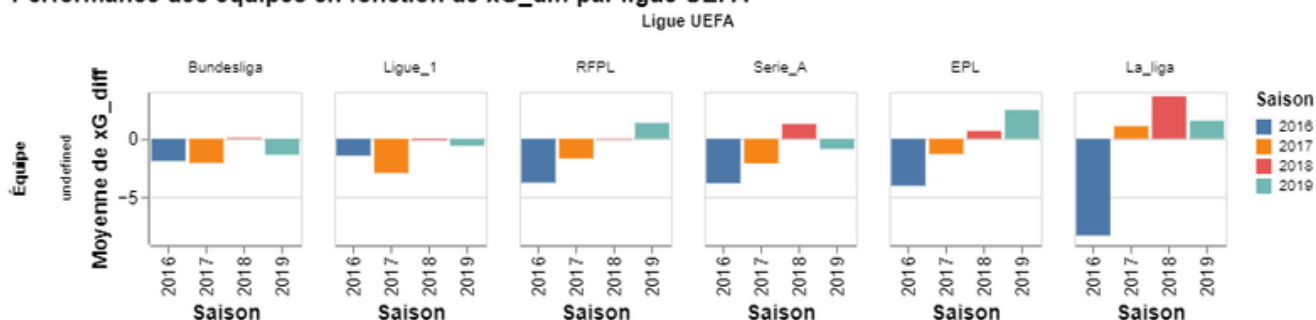
La capacité à convertir les occasion en buts se traduit par le xG_diff, cet variable représente une différence entre les positions théoriquement gagnantes, qui sont censées se concrétiser (expected goals xG, ou buts attendus en français) et les buts marqués on a donc:

$$\text{xG_diff} = \text{xG} - \text{scored}$$

Ainsi une équipe avec un xG_diff faible concrétise beaucoup et à l'inverse un xG_diff indique une difficulté à concrétiser

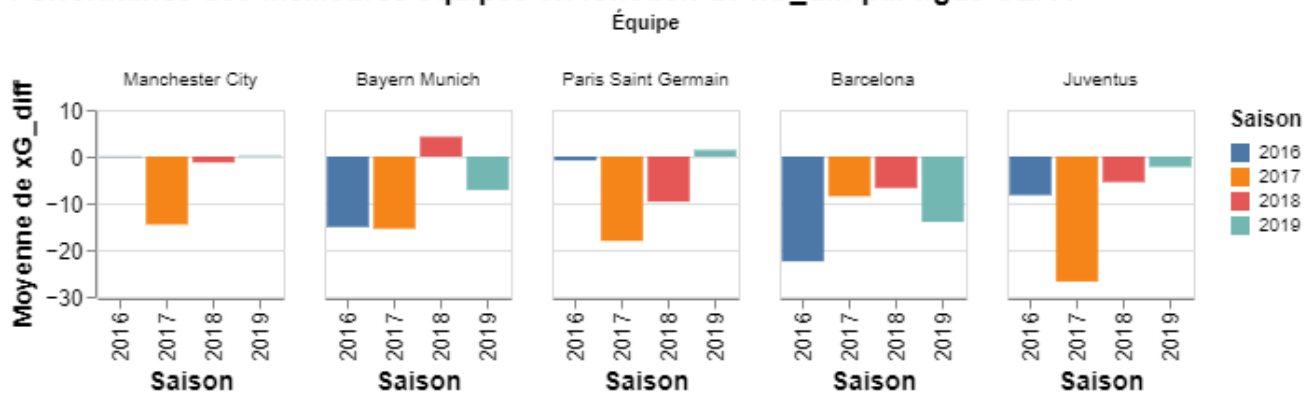
Le xG_diff va alors permettre de comparer le niveau des équipes, les leagues et les années.

Performance des équipes en fonction de xG_diff par ligue UEFA



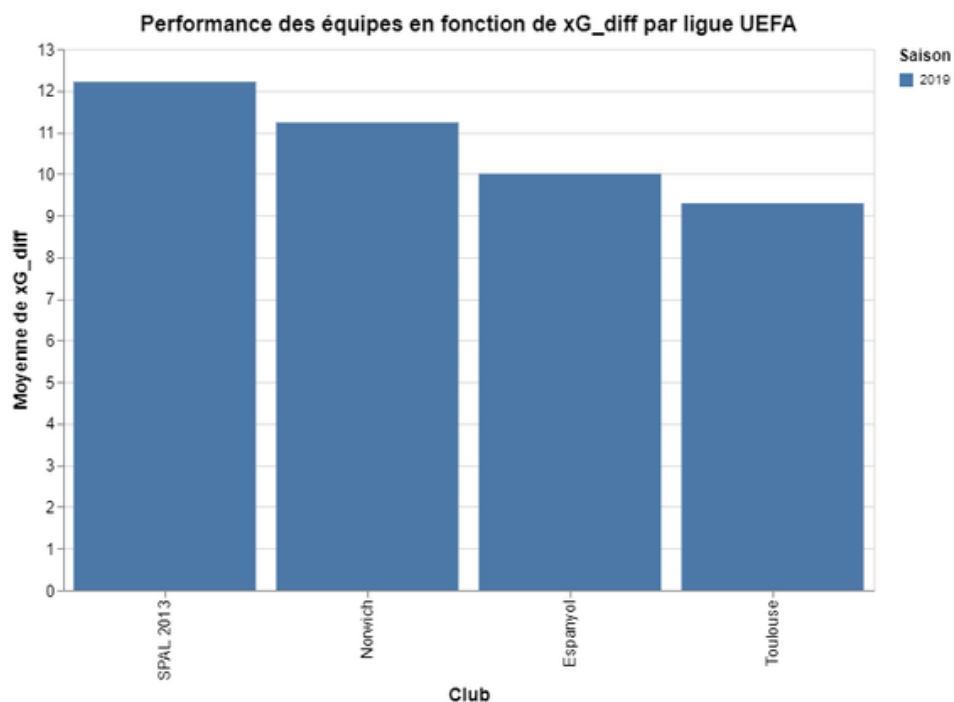
Ci dessus, on voit alors les différences de niveau en attaque des différentes leagues avec par exemple un niveau qui varie beaucoup d'une année à l'autre en liga.

Performance des meilleures équipes en fonction de xG_diff par ligue UEFA

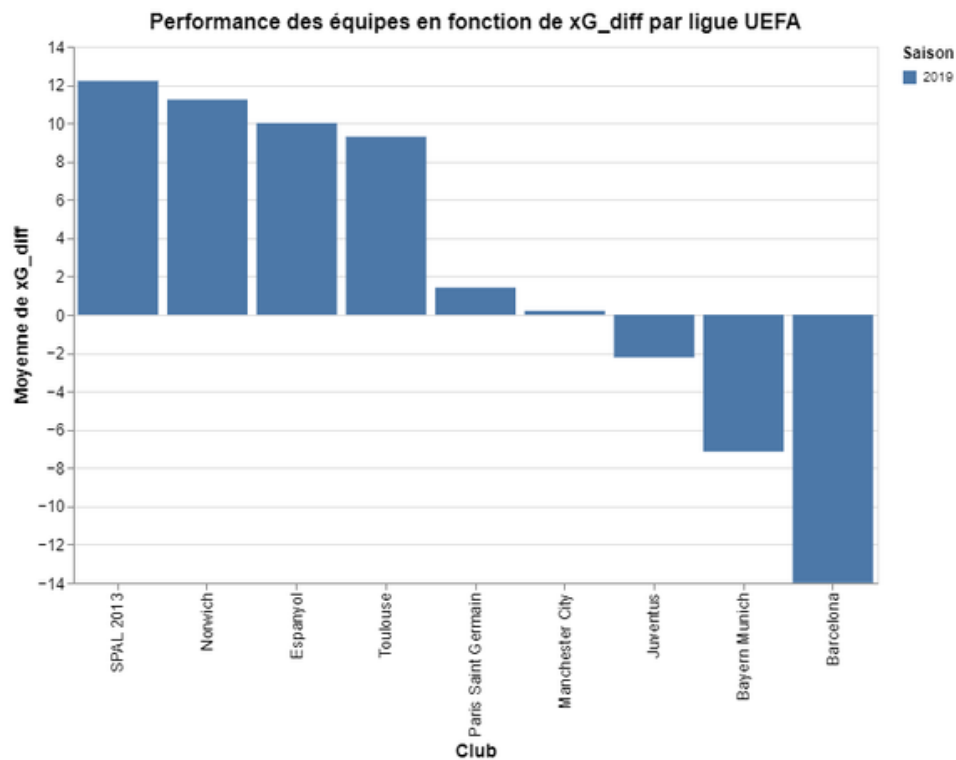


Avec cette meme analyse mais cette fois ci non plus sur les ligues mais les clubs, il est alors possible de faire des comparaisons sur les équipes pour avoir des informations sur les compositions qui concrétisait le mieux. Ceci a mettre en corrélation avec le départ des joueurs a mettre en corrélation avec ce graphique.

Maintenant, pour observer si le niveau de concrétisation d'une équipe a un réel impact sur ses performances générales nous avons pour les 4 pires équipes du classements en 2019 affiché leurs xG_diff que voici ci dessous.

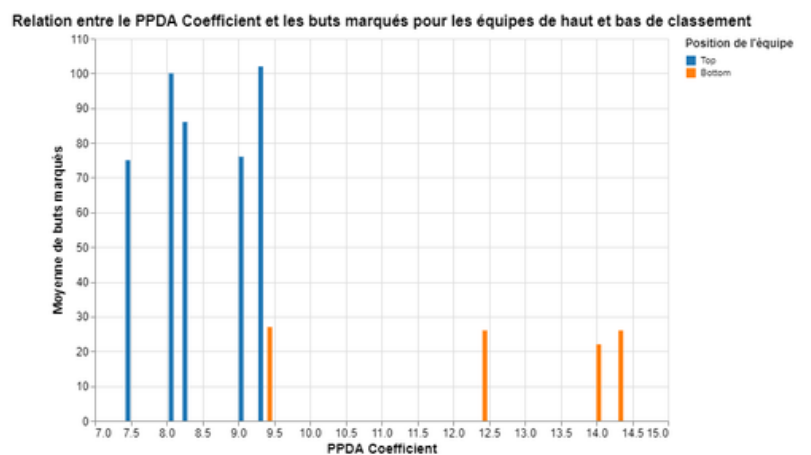
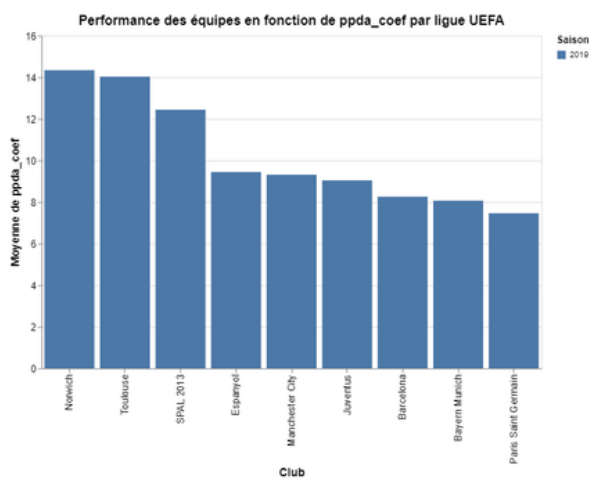


On voit bien une grosse différence avec les équipes étudiées précédemment. Pour afficher cette différence on met cote a cote les 4 pires équipes du classement avec les 5 meilleures et voici le résultat:



La différence frappante permet de dire que le xG_diff est une valeur representative des équipes.

Analyse de l'impact du pressing défensif sur la performance des équipes :



Dans notre analyse de l'impact du pressing défensif sur la performance des équipes, nous avons exploré si les équipes qui démontrent une activité et une intensité défensives plus élevées dans le camp adverse remportent généralement plus de matches et marquent plus de buts.

La visualisation que nous avons créée met en évidence les performances des meilleures équipes en fonction de leur moyenne de ppda_coef (Passes adverses par action défensive dans le camp adverse) pour chaque saison et chaque ligue UEFA. En examinant les barres sur le graphique, nous pouvons tirer plusieurs conclusions intéressantes :

Variation de ppda_coef au fil des saisons : Nous observons que la moyenne de ppda_coef des équipes peut varier d'une saison à l'autre. Par exemple, pour le FC Barcelone, nous remarquons une légère baisse de la moyenne de ppda_coef de 2017 à 2018, suivie d'une légère augmentation en 2019. Cette variation peut refléter les ajustements tactiques et les performances des équipes au fil du temps.

Comparaison entre les équipes : En comparant les différentes équipes, nous pouvons observer des différences dans leurs moyennes de ppda_coef. Par exemple, le FC Barcelone affiche généralement des moyennes de ppda_coef plus basses par rapport à d'autres équipes telles que la Juventus ou le Paris Saint-Germain. Cela suggère que le FC Barcelone adopte probablement une stratégie de pressing plus intense dans le camp adverse.

Corrélation avec les résultats de l'équipe : En analysant les performances des équipes en fonction de leur moyenne de ppda_coef, nous pouvons rechercher une corrélation avec les résultats de l'équipe, tels que le nombre de victoires ou de buts marqués. Par exemple, nous pourrions constater que les équipes ayant des moyennes de ppda_coef plus basses remportent généralement plus de matches ou marquent plus de buts, ce qui suggérerait un lien entre le pressing défensif et la performance globale de l'équipe.

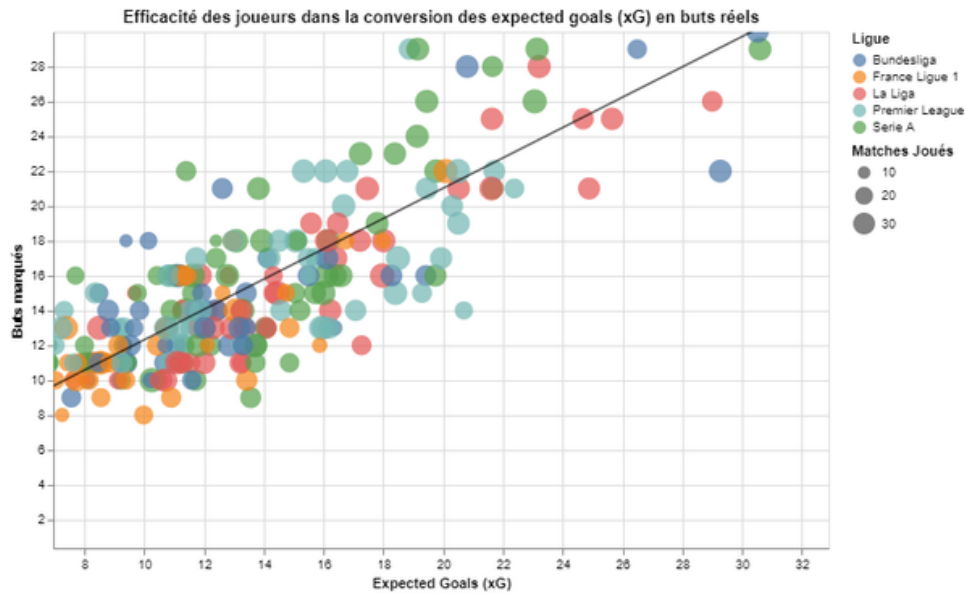
Impact sur la stratégie tactique : Enfin, nous pouvons utiliser ces données pour discuter de l'impact du pressing défensif sur la stratégie tactique des équipes. Par exemple, une équipe qui adopte un pressing intense dans le camp adverse peut chercher à récupérer rapidement le ballon et à créer des opportunités de buts en exploitant les erreurs de l'adversaire.

En conclusion, notre analyse de l'impact du pressing défensif sur la performance des équipes met en lumière l'importance de cette stratégie tactique dans le football moderne. Les données sur le ppda_coef nous permettent de comprendre comment les équipes abordent le jeu défensif et comment cela peut influencer leurs performances globales. Cependant, il est également important de prendre en compte d'autres facteurs, tels que la qualité de l'effectif, les tactiques de jeu et les conditions de match, pour une analyse complète et précise.

III. Analyse de l'efficacité des joueurs dans la conversion des occasions en buts :

Comment varie l'efficacité des joueurs dans la conversion des expected goals (xG) en buts réels dans différentes ligues au cours des dernières saisons ?

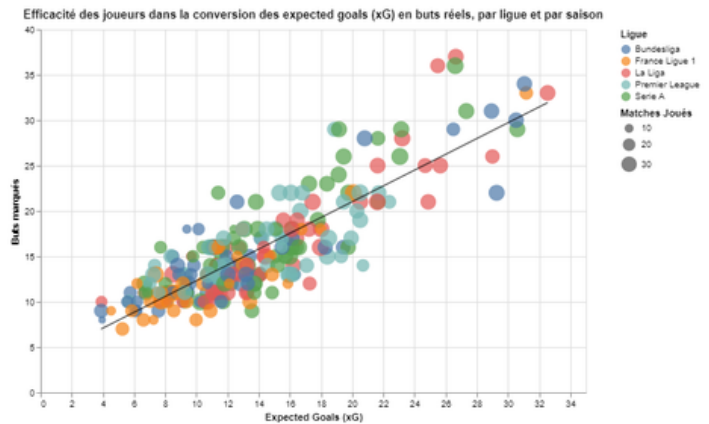
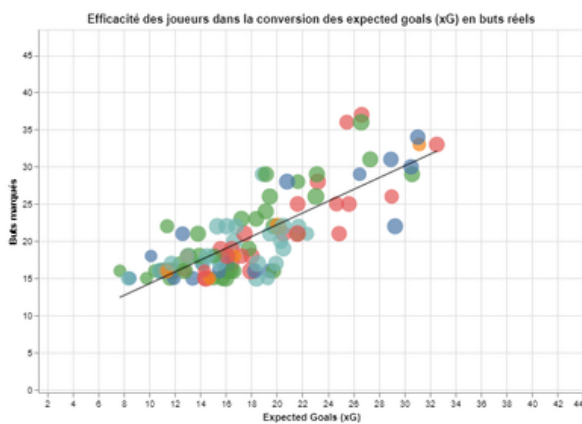
Cette question permet d'explorer la performance individuelle des joueurs par rapport à leurs attentes statistiques et de comparer ces performances entre les différentes compétitions.



Pour affiner la visualisation et explorer davantage les données

nous avons filtré les données pour ne montrer que les joueurs ayant joué un certain nombre de matches (20) ou ayant un certain nombre de buts marqués (15). Cela nous a permis de nous concentrer sur les joueurs les plus performants.

De plus on peut regrouper les données par ligue et par saison c'est à dire plutôt que d'afficher toutes les données en même temps, nous pourrions regrouper les données par ligue et par saison



Analyse de la corrélation entre la performance des joueurs et celle de leur équipe:

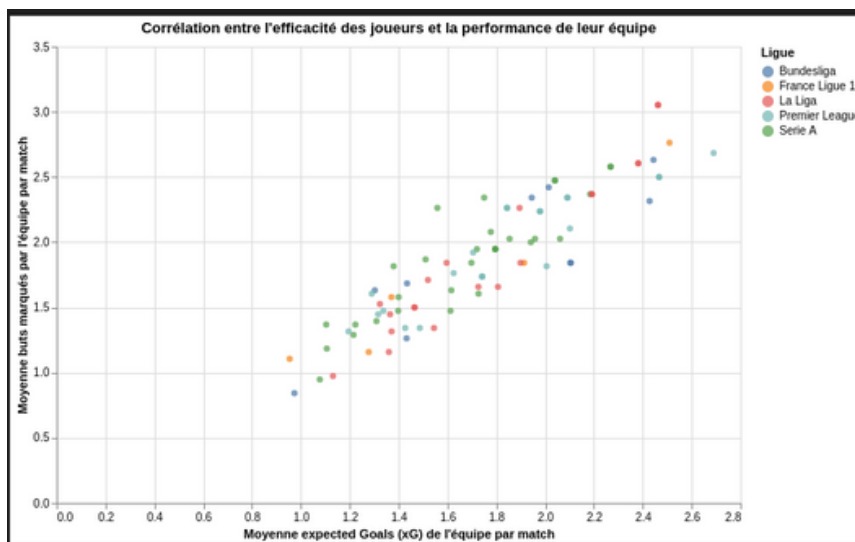
Dans notre analyse de la corrélation entre la performance des joueurs et celle de leur équipe, nous avons exploré si l'efficacité des joueurs dans la conversion des expected goals (xG) en buts réels est liée aux résultats globaux de leur équipe dans chaque ligue.

La visualisation que nous avons créée met en évidence plusieurs points intéressants. En examinant la dispersion des points sur le graphique, nous observons une tendance générale où les équipes ayant une moyenne plus élevée de xG par match ont également une moyenne plus élevée de buts marqués par match. Cela suggère qu'il existe une corrélation positive entre l'efficacité des joueurs dans la conversion des xG en buts réels et la performance globale de leur équipe.

En se concentrant sur des exemples concrets, nous pouvons observer des cas où cette corrélation est particulièrement forte. Par exemple, Antoine Griezmann à Barcelone affiche une moyenne élevée de xG par match (2.46) ainsi qu'une moyenne élevée de buts marqués par match (3.05). De même, Lionel Messi, également à Barcelone, présente des chiffres similaires avec une moyenne de xG par match de 2.46 et une moyenne de buts marqués par match de 3.05. Ces exemples illustrent comment l'efficacité des joueurs peut avoir un impact significatif sur la performance globale de leur équipe.

Cependant, il est important de noter qu'il existe des exceptions à cette tendance. Par exemple, Cristiano Ronaldo à la Juventus affiche une moyenne de xG par match élevée (2.46), mais sa moyenne de buts marqués par match n'est pas aussi élevée (3.05). Cela suggère que d'autres facteurs, tels que la qualité de l'opposition ou les tactiques de l'équipe, peuvent également influencer les résultats de l'équipe, en plus de l'efficacité individuelle des joueurs.

En conclusion, bien que notre analyse mette en lumière une corrélation positive entre l'efficacité des joueurs dans la conversion des xG en buts réels et la performance globale de leur équipe, il est important de prendre en compte d'autres facteurs pour une compréhension plus approfondie des résultats d'une équipe. Comme le souligne l'aphorisme d'Albert Einstein, "Ce qui compte ne peut pas toujours être compté, et ce qui peut être compté ne compte pas forcément." Cette nuance nous rappelle l'importance de considérer divers aspects du jeu pour une analyse complète et précise.



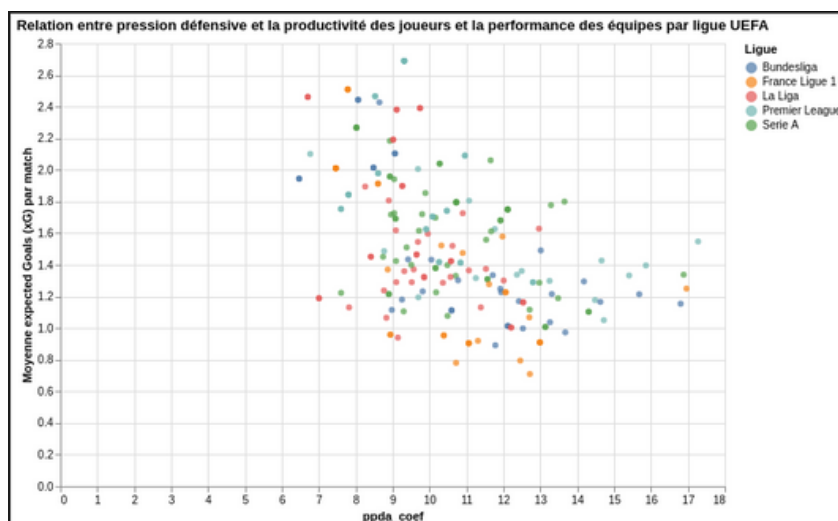
Analyse de l'impact de la pression défensive sur la performance des équipes en fonction de la qualité des joueurs :

Dans notre étude sur l'impact de la pression défensive sur la productivité des joueurs, nous avons exploré la corrélation entre la capacité des équipes à exercer une pression défensive, mesurée par le `ppda_coef`, et l'efficacité de leurs joueurs offensifs dans la conversion des expected goals (xG) en buts réels. Le `ppda_coef`, un indicateur disponible sur understat.com, confronte les passes adverses par action défensive dans le camp adverse, où un `ppda_coef` plus bas indique une pression défensive plus intense.

Notre visualisation a révélé des tendances intéressantes. En observant la dispersion des points sur le graphique, nous constatons une corrélation générale entre le `ppda_coef` et la moyenne des expected goals par match (`xG_avg_team`). Par exemple, les équipes avec un `ppda_coef` plus faible tendent à avoir une moyenne de xG plus élevée, suggérant que leur pression défensive plus intense peut favoriser les opportunités offensives et la conversion des chances en buts.

Cependant, il est important de noter quelques nuances dans notre analyse. En examinant des exemples concrets, tels que Cristiano Ronaldo à la Juventus et Lionel Messi à Barcelone, nous pouvons observer des différences significatives. Ronaldo, avec un `ppda_coef` de 10.16, présente une moyenne de xG par match de 1.70, tandis que Messi, avec un `ppda_coef` de 6.70, affiche une moyenne de xG par match plus élevée de 2.46. Cela suggère que malgré le fait que Ronaldo marque beaucoup de buts, la Juventus n'exerce peut-être pas une pression défensive aussi intense que Barcelone, où Messi brille également dans la conversion des occasions en buts.

Ainsi, bien qu'il existe une corrélation générale entre la pression défensive et la productivité des joueurs, ces exemples soulignent l'importance de prendre en compte les variations individuelles et contextuelles. Il est crucial de reconnaître que le football est un sport complexe où plusieurs facteurs interagissent pour influencer les performances des équipes et des joueurs.



Annexe disponible ci-dessous.