

# Head coach dismissal effect on football team performance

Mathis Derenne<sup>1</sup>, Ewann Boismartel<sup>1</sup>, Scott Thomas<sup>1</sup>, and Romain Biguet<sup>1</sup>

<sup>1</sup>University of Rennes

---

## Abstract

This paper aims to explore the relationship between the team's performance and sports coach. To do that, we use traditional statistical methods applied to Men's Football First Divisions of major leagues between 2015 and 2023. We investigate this relation on three different aspects: the effect of head coach tenure, the effect of regular renewal of sports coach by club, and the effect of changing clubs on the coach's performance. Our results show that there is a positive correlation between head coach tenure and team's performance, while the regular renewal of a sports coach by a club has a negative relation on the team's performance. We also find that head coach who regularly change club is negatively correlated with team's performance. These results suggest that stability and continuity are important factors for the success of a football team.

---

**Keywords** coach dismissal, team performance

## 1. INTRODUCTION

Le football est un sport populaire qui attire des millions de fans à travers le monde. Les équipes de football sont composées de joueurs talentueux, mais elles sont également dirigées par des entraîneurs sportifs qui jouent un rôle crucial dans le succès de l'équipe. Les entraîneurs sportifs sont responsables de la stratégie de jeu, de la tactique, de la motivation des joueurs et de la gestion de l'équipe. Ils sont souvent considérés comme les architectes du succès d'une équipe.

Au cours des dernières années, de nombreuses équipes de football ont connu des changements fréquents d'entraîneurs sportifs. Ces changements peuvent être dus à divers facteurs, tels que les mauvais résultats de l'équipe, les conflits internes, les différends avec les joueurs ou la direction du club, ou simplement le désir de renouveler l'équipe. Cependant, il est important de comprendre l'impact de ces changements d'entraîneurs sur la performance de l'équipe.

Dans cette étude, nous allons analyser les données des ligues de football masculin de première division de cinq pays européens (Angleterre, Espagne, Italie, Allemagne et France) sur la période 2015 - 2023.

Cette analyse vise à mieux comprendre la relation entre la performance d'une équipe de football et l'entraîneur sportif. Nous allons nous intéresser à trois aspects différents de cette relation : l'effet de l'ancienneté de l'entraîneur sur la performance de l'équipe, l'effet du renouvellement régulier de l'entraîneur par le club sur la performance de l'équipe, et l'effet du changement de clubs sur la performance de l'entraîneur.

### 1.a. Source des données

Les données utilisées au cours de cette analyse sont extraites de deux sites spécialisés dans les statistiques de football : [Fbref](#) et [Transfmarkt](#).

Published Apr 14, 2024

## Data Availability

L'ensemble des fichiers et données relatif à ce travail sont disponible en accès libre sur le [dépot GitHub](#) sous licence MIT.

FBref met à disposition une grande quantité de données sur les matches, les joueurs, les équipes et les compétitions de football. Il fournit des informations détaillées sur les performances des joueurs, les statistiques de match, les classements des équipes, les résultats des matches et d'autres données liés au football.

Fbref est détenu par l'entreprise Sport Reference qui gère également d'autres sites spécialisés dans les statistiques sportives, comme Baseball-Reference et Basketball Reference.

Transfermarkt est un site spécialisé dans les transferts de joueurs de football. Il fournit des informations sur les transferts de joueurs, les valeurs marchandes des joueurs, les contrats des joueurs, les rumeurs de transferts et d'autres détails liés aux transferts de joueurs.

Ces sites sont utilisés par les fans de football, les médias, les clubs de football et les professionnels du football pour obtenir des informations footballistique.

### *1.b. Fiabilité des données*

Ces sites sont considérés comme fiables. Ils sont régulièrement mis à jour et vérifiés pour assurer la qualité et la précision des données. Les données sont collectées à partir de sources officielles, comme les ligues de football, les clubs de football, les fédérations de football et d'autres sources fiables. Les données sont ensuite vérifiées et validées par des experts en statistiques sportives avant d'être publiées sur les sites.

### *1.c. Références et outils utilisés*

L'intégralité du travail de récupération, de pré-traitement, d'analyse et visualisation des données a été réalisé au sein de Notebook Jupyter.

La récupération des données footballistique a été effectuée à l'aide du package R [WorldFootBallR](#). Ce package est régulièrement mis à jour et implémente des outils de web scraping afin d'extraire les données des principaux sites footballistiques.

Le pré-traitement, l'analyse et la visualisation des données a été effectué sous Python à l'aide de bibliothèques standards : Pandas, Numpy, Matplotlib, Seaborn, Scipy.

L'écriture de ce papier a été réalisé dans un fichier au format Markdown.

[MyST](#) fait partie d'un écosystème d'outils qui cherchent à améliorer le travail de communication scientifique en favorisant le développement d'une science reproductible et indexable. Cet outil a été utilisé pour permettre la diffusion de ce papier de recherche au format d'un [site statique](#) et d'un [PDF](#) répondant aux exigences de qualité scientifique.

MyST permet de réutiliser les entrées et les sorties des Notebooks Jupyter. Ainsi l'ensemble des figures, tableaux et une partie des variables présentes dans ce papier sont directement issus des Notebooks Jupyter. À titre d'exemple, il est possible de renouveler l'intégralité de l'étude à d'autres ligues ou d'autres périodes en modifiant simplement les valeurs des variables utilisées dans les Notebooks Jupyter :

```
country <- c("ENG", "ESP", "ITA", "GER", "FRA")
year <- c(2015:2022)
```

Enfin, la lecture de Fundamentals of Data Visualization (Wilke, 2019) a permis d'améliorer la qualité des graphiques et de la présentation des données en les rendant plus clairs et informatifs.

## 2. PRÉPARATION DES DONNÉES

Il est essentiel de préparer les données avant de les analyser. Cette étape consiste à nettoyer les données, à les transformer et à les organiser de manière à ce qu'elles soient exploitables pour l'analyse statistique. De plus, il est important de vérifier l'intégrité et la complétude des données afin de s'assurer de leur cohérence et de leur fiabilité dans le cadre de l'analyse statistique.

Ces différentes étapes de préparation des données sont réalisées de manière itérative et suivent rarement un chemin linéaire. L'évolution des besoins de l'étude, les contraintes imposées par les données disponibles et les données d'intérêts et la détection d'incohérences dans les données sont autant de facteurs qui amènent à faire évoluer la pipeline de préparation des données. Davantage de détails sur le traitement des données sont disponibles dans les Notebooks Jupyter associés à ce papier.

### 2.a. Données extraites

Un premier jeu de donnée concernant les matchs est récupéré à partir de [Fbref](#). Il contient des informations sur les matchs de football, notamment les équipes qui ont joué, le score final, le lieu du match et la date du match.

```
tablex(columns: 9, header-rows: 1, repeat-header: true, ..tableStyle, [ ], [ League ], [ Country ],
[ Season ], [ Date ], [ Home ], [ HomeGoals ], [ Away ], [ AwayGoals ], [ 0 ], [ Premier
League ], [ England ], [ 2015 ], [ 2014-08-16 ], [ Manchester Utd ], [ 1.0 ], [ Swansea City ],
[ 2.0 ], [ 1 ], [ Premier League ], [ England ], [ 2015 ], [ 2014-08-16 ], [ Stoke City ], [ 0.0 ],
[ Aston Villa ], [ 1.0 ], [ 2 ], [ Premier League ], [ England ], [ 2015 ], [ 2014-08-16 ], [ Leicester
City ], [ 2.0 ], [ Everton ], [ 2.0 ], [ 3 ], [ Premier League ], [ England ], [ 2015 ], [ 2014-08-16 ],
[ QPR ], [ 0.0 ], [ Hull City ], [ 1.0 ], [ 4 ], [ Premier League ], [ England ], [ 2015 ],
[ 2014-08-16 ], [ West Ham ], [ 0.0 ], [ Tottenham ], [ 1.0 ], )
```

Table 1: Extrait du jeu de donnée sur les matchs

Un second jeu de donnée concernant les mandats des entraîneurs chef est récupéré à partir de [Transfermarkt](#). Il contient la date de début et de fin du mandat d'un entraîneur sportif, le club concerné, ainsi que nom et la performance agrégé de l'entraîneur lors de ce mandat.

```
tablex(columns: 12, header-rows: 1, repeat-header: true, ..tableStyle, [ ], [ Team ], [ League ],
[ Country ], [ HeadCoach ], [ Appointed ], [ EndDate ], [ Tenure ], [ Matches ], [ Wins ],
[ Draws ], [ Losses ], [ 0 ], [ Chelsea FC ], [ Premier League ], [ England ], [ Mauricio
Pochettino ], [ 2023-07-01 ], [ NaT ], [ 296 ], [ 44 ], [ 22 ], [ 9 ], [ 13 ], [ 1 ], [ Chelsea FC ],
[ Premier League ], [ England ], [ Graham Potter ], [ 2022-09-08 ], [ 2023-04-02 ], [ 206 ], [ 31 ],
[ 12 ], [ 8 ], [ 11 ], [ 2 ], [ Chelsea FC ], [ Premier League ], [ England ], [ Thomas Tuchel ],
[ 2021-01-26 ], [ 2022-09-07 ], [ 589 ], [ 100 ], [ 63 ], [ 19 ], [ 18 ], [ 3 ], [ Chelsea FC ], [ Premier
League ], [ England ], [ Frank Lampard ], [ 2019-07-04 ], [ 2021-01-25 ], [ 571 ], [ 84 ], [ 44 ],
[ 15 ], [ 25 ], [ 4 ], [ Chelsea FC ], [ Premier League ], [ England ], [ Maurizio Sarri ],
[ 2018-07-14 ], [ 2019-06-30 ], [ 351 ], [ 63 ], [ 40 ], [ 11 ], [ 12 ], )
```

Table 2: Extrait du jeu de donnée sur les mandats des entraîneurs sportif

### 2.b. Données d'intérêt

On exclut des données de résultat de matchs, les matchs ne concernant pas les équipes de première division.

On exclut des mandats des entraîneurs, les entraîneurs sportifs n'ayant pas été actifs entre le début de la saison 2015 et la fin de la saison 2023.

On exclut des mandats des entraîneurs, les entraîneurs sportifs ayant eu moins de 5 matchs à la tête d'une équipe. En effet, ces entraîneurs sont soit des entraîneurs de substitution, qui sont intervenus pour une courte période avant d'être remplacé par un entraîneur permanent, soit des entraîneurs qui ont été licenciés rapidement en raison d'une faute professionnelle ou d'une incompatibilité entre l'équipe et l'entraîneur sportif.

### 2.c. Données manquantes ou incohérentes

Certains matchs ne possèdent pas d'informations sur le résultat final. Ces matchs correspondent en réalité à des matchs qui ont été annulés ou reportés notamment en lien avec le COVID-19. Ces matchs sont exclus de l'analyse.

Certains enregistrements de mandats d'entraîneurs ne possèdent pas d'information sur le pays et la ligue du club concerné. Ces informations ont été ajoutées manuellement en se basant sur le nom du club avant le filtrage sur les données d'intérêts.

Un nombre important d'équipe a des informations incomplètes sur les mandats des entraîneurs sportifs. Certaines de ces données manquantes correspondent à des périodes de changement d'entraîneurs, durant laquelle le club n'a plus d'entraîneur sportif en poste. Toutefois, un nombre significatif d'équipes ne possèdent aucune information sur l'entraîneur sportif à la tête du club entre 2015 et 2023. L'absence de mandat entre 2015 et 2023 concerne 95 équipes sur les 161 équipes d'intérêts. Le problème provient de la liste de liens TransferMarkt retournée par la fonction `tm_league_team_urls(country_name, start_year)` de `worldfootballR` qui ne contient pas l'ensemble des équipes des pays d'intérêts.

Ce problème a été détecté à la fin de l'analyse et n'a pas pu être corrigé. Une des solutions consisterait à ajouter manuellement les liens TransferMarkt des 95 équipes pour lesquelles il n'y a aucun enregistrement d'entraîneurs sportifs.

Nous avons trouvé 11 enregistrements de mandats d'entraîneurs sportifs qui avaient lieu alors qu'un autre chef-entraîneur assurait l'entraînement de l'équipe. Ces enregistrements concernent les équipes suivante : 'FC Empoli, Hellas Verona, LOSC Lille, Newcastle United, Rayo Vallecano, SSC Napoli, Stade Reims, Stade Rennais FC, Torino FC, West Ham United'. Ces enregistrements sont exclus du jeu de données.

```
tablex(columns: 5, header-rows: 1, repeat-header: true, ..tableStyle, [ ], [ Team ], [ Appointed ],
[ EndDate ], [ Overlap ], [ 3423 ], [ Stade Reims ], [ 2016-07-01 ], [ 2017-05-22 ], [ False ],
[ 3422 ], [ Stade Reims ], [ 2017-05-22 ], [ 2021-05-25 ], [ True ], [ 3421 ], [ Stade Reims ],
[ 2018-07-01 ], [ 2019-03-30 ], [ False ], [ 3420 ], [ Stade Reims ], [ 2021-06-23 ], [ 2022-10-13 ],
[ False ], )
```

Table 3: Exemple d'inconsistences dans les données des entraîneurs sportifs

Les résultats de matchs contiennent 161 équipes et les mandats de coaches contiennent 67 équipes. Cependant, le nom de certaines équipes est différent entre les deux jeux de données. Par exemple, « Liverpool » dans les résultats de match devient « Liverpool FC » dans les mandats des entraîneurs.

Il existe 59 équipes présentes dans les mandats de coaches qui n'ont pas de correspondance dans les résultats de match.

Team in head coach records	Team in match results
Newcastle United	Newcastle Utd

Crystal Palace	Crystal Palace
Chelsea FC	Chelsea
Arsenal FC	Arsenal
Liverpool FC	Liverpool

Table 4: Exemple de correspondance des noms d'équipes

Nous avons utilisé l'algorithme de la distance Levenshtein (Levenshtein, 1965) afin de faire correspondre les équipes des mandats d'entraîneurs aux équipes des résultats de matchs. Nous restreignons par ailleurs la recherche d'une équipe correspondante à la liste des équipes du pays correspondant afin de limiter le nombre de correspondances possibles. L'ajustement du score de Levenshtein a permis de rapidement vérifier la validité des correspondances de plus faible certitude.

```

from thefuzz import process

team_name_mapping = {}

# For each country
for country in coach_teams_by_country.index:
    # Get teams for this country
    coach_teams = coach_teams_by_country[country]
    match_teams = match_teams_by_country.get(country, [])

    # For each team in coach_teams
    for coach_team in coach_teams:
        # Find the best match in match_teams
        matching_scores = process.extract(coach_team, match_teams, limit=1)

        if len(matching_scores) != 0 and matching_scores[0][1] >= 60:
            team_name_mapping[coach_team] = matching_scores[0][0]
        else:
            team_name_mapping[coach_team] = None
            print(f"No match found for {coach_team} among {match_teams} in
{country}")

```

Code 1: Utilisation de l'algorithme de la distance Levenshtein

#### 2.d. Transformation des données

Les [Table 1](#) sont modifiés de manière à ce que chaque match soit divisé en deux lignes : une pour chaque équipe. Ainsi, chaque ligne représente le résultat d'une équipe lors d'un match : la date du match, le résultat final, la présence ou non de l'équipe à domicile, le score final de l'équipe ainsi que le nom de l'équipe.

De plus, nous ajoutons aux données de match l'ancienneté de l'entraîneur au sein de l'équipe lorsque ce match a été joué.

```

tablex(columns: 10, header-rows: 1, repeat-header: true, ..tableStyle, [ ], [ League ],
[ Country ], [ Date ], [ Team ], [ Goals ], [ Result ], [ isHome ], [ HeadCoach ], [ DaysInPost ],
[ 0 ], [ Premier League ], [ England ], [ 2014-08-16 ], [ Manchester Utd ], [ 1.0 ], [ loss ],
[ True ], [ Louis van Gaal ], [ 33.0 ], [ 1 ], [ Premier League ], [ England ], [ 2014-08-16 ],
[ Arsenal ], [ 2.0 ], [ win ], [ True ], [ Arsène Wenger ], [ 6528.0 ], [ 2 ], [ Premier League ],
[ England ], [ 2014-08-17 ], [ Liverpool ], [ 2.0 ], [ win ], [ True ], [ Brendan Rodgers ], [ 777.0 ],
[ 3 ], [ Premier League ], [ England ], [ 2014-08-17 ], [ Newcastle Utd ], [ 0.0 ], [ loss ], [ True ],

```

[ Alan Pardew ], [ 1347.0 ], [ 4 ], [ Premier League ], [ England ], [ 2014-11-01 ], [ Newcastle Utd ], [ 1.0 ], [ win ], [ True ], [ Alan Pardew ], [ 1423.0 ], )

Table 5: Extrait du jeu de donnée sur les matchs après pré-traitement

Nous ajoutons également aux [Table 2](#) le nombre de clubs pour lequel l'entraîneur a travaillé parmi les clubs des ligues et divisions d'intérêts.

### 3. PRÉSENTATION DES DONNÉES

Nous avons collecté les résultats des matchs et les mandats des entraîneurs sportif chef de la première division masculine de football pour les saisons 2015 - 2023 et pour les ligues suivantes : 'Premier League, La Liga, Ligue 1, Bundesliga, Serie A'.

Cela correspond à un total de 14361 matchs joués par 161 équipes. Parmi ces 161 équipes, 95 équipes n'ont pas d'information sur les mandats des entraîneurs chefs pour tout ou partie des saisons 2015 à 2023. En proportion de match c'est '46' % des matchs qui ne comportent pas d'information sur l'entraîneur chef.

Les données des mandats des entraîneurs sportifs contiennent 381 mandat d'entraîneurs chefs parmi 249 entraîneurs uniques.

```
tablex(columns: 5, header-rows: 2, repeat-header: true, ..tableStyle, [ ], [ ], [ Number of Teams ], [ Number of Matches ], [ Percentage of Matches without Head Coach ], [ League ], [ Country ], [ ], [ ], [ ], [ La Liga ], [ Spain ], [ 31 ], [ 6145 ], [ 48.23% ], [ Ligue 1 ], [ France ], [ 32 ], [ 6043 ], [ 51.98% ], [ Serie A ], [ Italy ], [ 34 ], [ 5927 ], [ 34.64% ], [ Premier League ], [ England ], [ 33 ], [ 5582 ], [ 53.78% ], [ Bundesliga ], [ Germany ], [ 31 ], [ 5026 ], [ 38.32% ], )
```

Table 6: Summary of the match data

#### 3.a. L'avantage de jouer à domicile

En moyenne, l'équipe qui joue à domicile marque '1.50' but contre '1.18' but pour l'équipe jouant à l'extérieur ('22' % moins de buts). Ainsi, les équipes jouant remportent '43' % de leurs matchs contre '29' % pour les équipes jouant à l'extérieur. Les matchs nuls représentent '25' % des matchs.

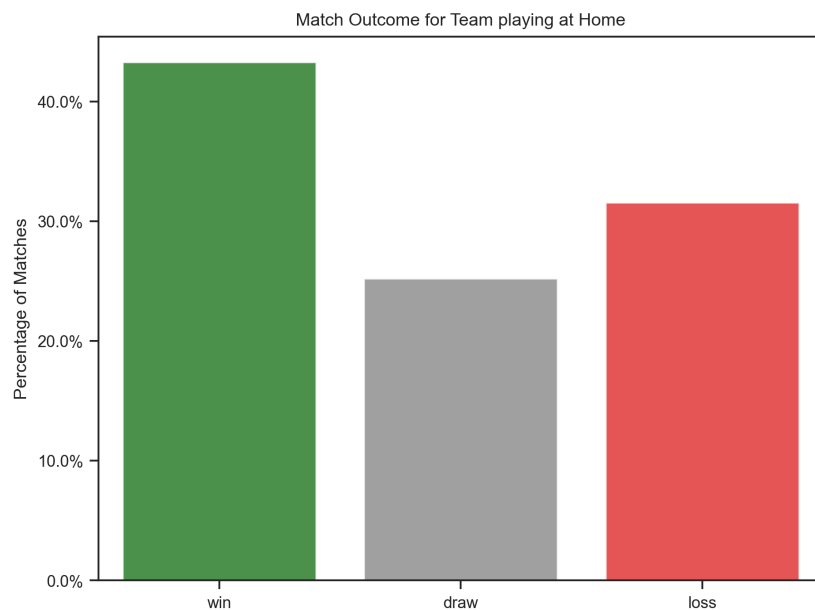


Figure 1: Venue effect on team's performance (2015 - 2023)

### 3.b. Les dynamiques temporelles

Les saisons de football sont divisées en deux périodes : la saison régulière et la saison hors-saison. La saison régulière est la période pendant laquelle les équipes jouent des matchs de championnat et de coupe, tandis que la saison hors-saison marque une période de pause permettant à l'équipe de s'entraîner et de se préparer pour la saison suivante.

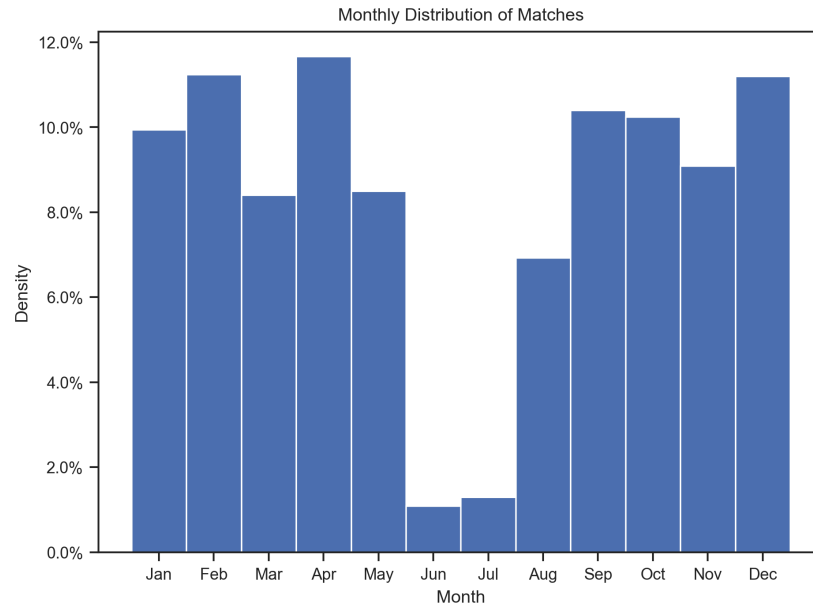


Figure 2: Monthly Distribution of Matches (2015 - 2023)

Une proportion importante des nominations et licenciement des entraîneurs sportifs a lieu en période de hors-saison ((Figure 3) et Figure 4). En effet, les clubs profitent de cette période pour revoir leur stratégie et renouveler certains membres de l'équipe. Le changement d'entraîneur au cours de la saison régulière est plus rare et risqué, car l'équipe est déjà engagée dans des compétitions et le changement d'entraîneur peut perturber la dynamique de l'équipe. Le licenciement d'un entraîneur en cours de saison pourrait indiquer une situation critique pour le club, comme une série de mauvais résultats ou des conflits internes.

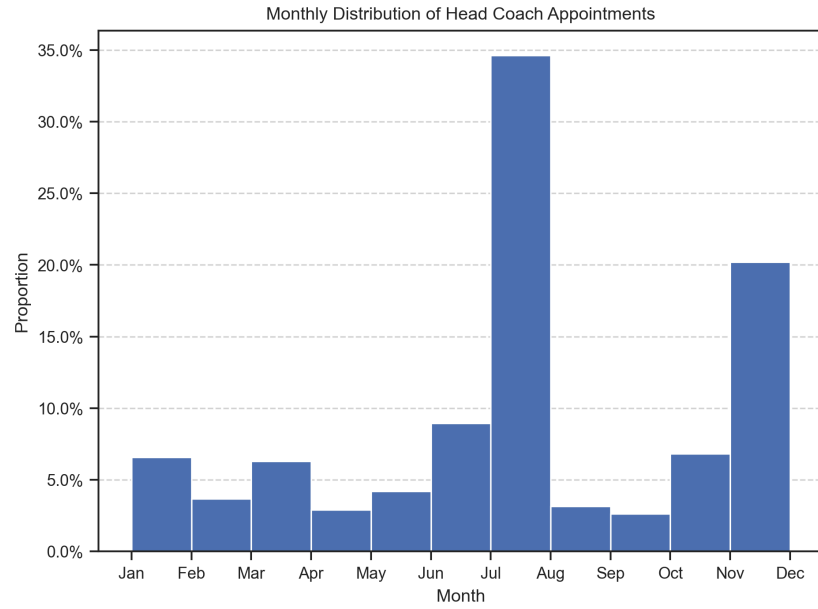


Figure 3: Monthly Distribution of Head Coaches Appointments

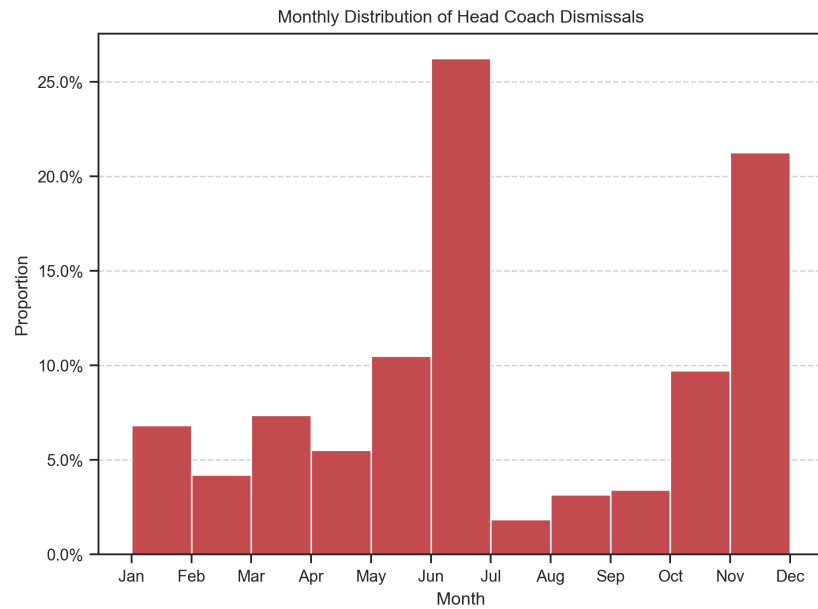


Figure 4: Monthly Distribution of Head Coaches Dismissals

Les périodes de nominations et de licenciement des entraîneurs sportifs sont relativement stables au cours des saisons, mais aussi entre les ligues (Figure 5).



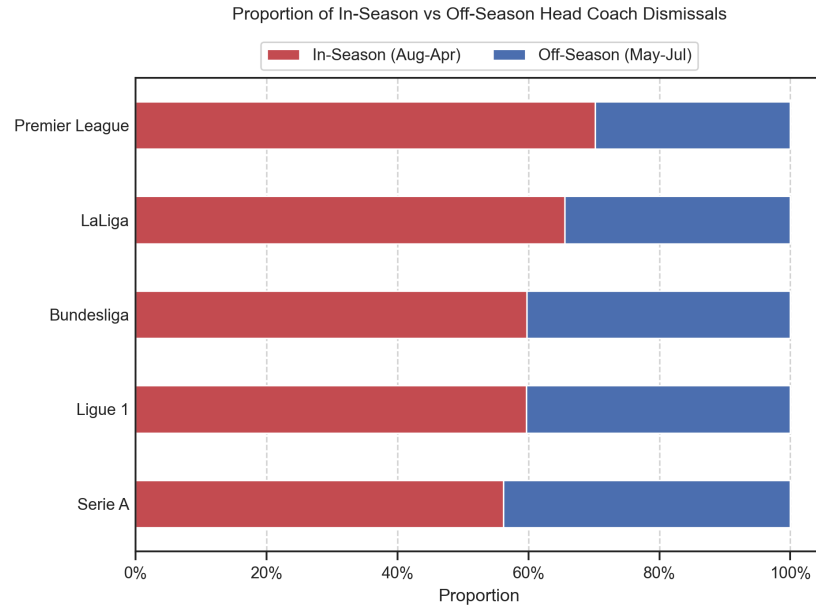


Figure 5: Proportion of in-season vs off-season head coach dismissal per league

## 4. ANALYSE STATISTIQUE

### 4.a. Définitions

Le coefficient de corrélation de Pearson (valeur de  $r$ ) est une mesure statistique qui évalue la force et la direction de la relation linéaire entre deux variables continues. Il est souvent utilisé pour quantifier la relation entre deux variables.

- Lorsque  $r = 1$ , cela indique une corrélation linéaire parfaite positive.
- Lorsque  $r = -1$ , cela indique une corrélation linéaire parfaite négative.
- De plus, lorsque  $r = 0$ , cela signifie qu'il n'y a aucune corrélation linéaire entre les deux variables.

La p-value (valeur de  $p$ ) est une mesure statistique utilisée pour déterminer la signification statistique d'un résultat. Dans le contexte de la corrélation de Pearson, la p-value est utilisée pour évaluer si la corrélation observée entre deux variables est statistiquement significative ou non. Si la p-value est inférieure à un seuil (ici  $p < 0.05$ ), on rejette l'hypothèse selon laquelle il n'y a pas de corrélation dans la population, et on conclut qu'il y a une corrélation significative entre les deux variables.

### Caution

Une p-valeur faible dans le cadre de la corrélation de Pearson ne signifie pas nécessairement qu'il existe une relation de cause à effet entre les deux variables observées. Elle indique simplement que la corrélation observée est statistiquement significative. Ainsi, les hypothèses explicatives fournies sont simplement des hypothèses plausibles et non des conclusions définitives.

### 4.b. Renouvellement de l'entraîneur et performance du club

Dans cette partie, nous allons étudier les effets de la fréquence de renouvellement des coaches sur la performance d'une équipe. L'objectif est de voir les effets sur les résultats obtenus lorsque qu'une équipe change régulièrement d'entraîneur.

La [Figure 6](#) nous informe quant à la distribution du nombre d'entraîneurs employés par les clubs durant la période 2015 - 2023. On observe que plus de 85 % des clubs ont employés au moins 3 entraîneurs différents suggérant qu'il y a un renouvellement régulier des entraîneurs dans les clubs de football.

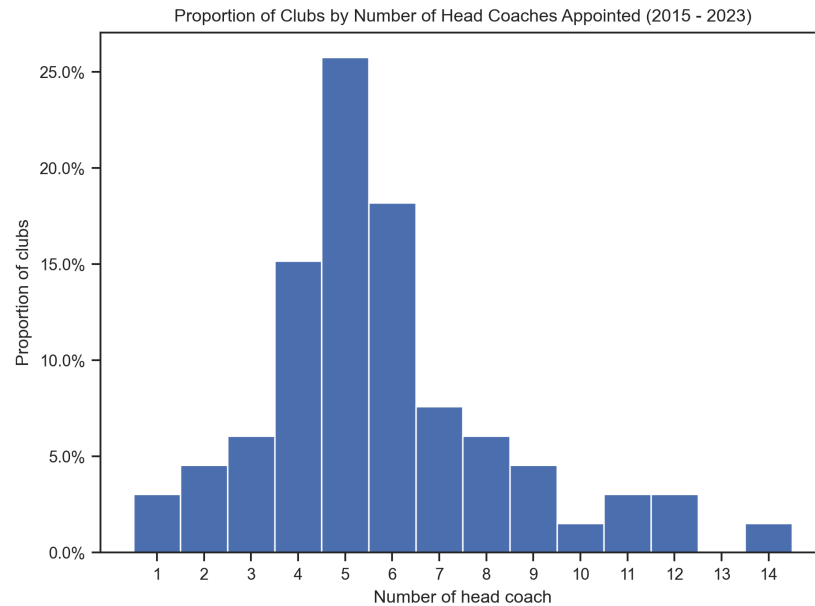


Figure 6: Proportion of Clubs by Number of Head Coaches Appointed (2015 - 2023)

La [Figure 7](#) montre que les entraîneurs de la Premier League restent en poste plus longtemps que ceux des autres ligues. De plus, les équipes de la Premier League changent moins souvent d'entraîneur que celles des autres ligues. À l'inverse, la LaLiga renouvelle fréquemment ses entraîneurs.

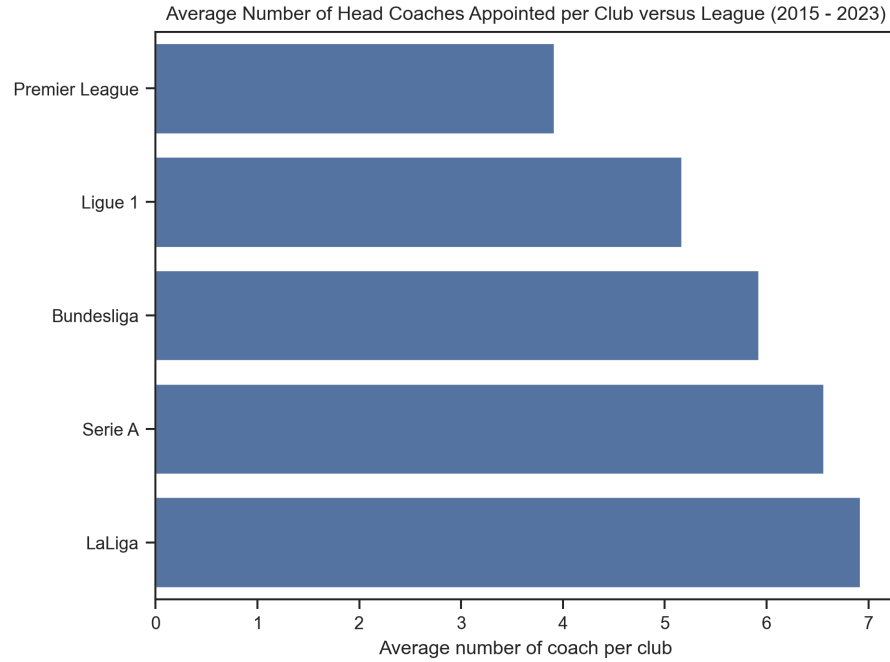


Figure 7: Average Number of Head Coaches Appointed per Club versus League (2015 - 2023)

La [Figure 8](#) s'intéresse aux ratios de victoires, de matchs nuls et de défaites des clubs en fonction du nombre d'entraîneurs nommés à la tête de l'équipe au cours de la période 2015 - 2023. Ces graphiques permettent de visualiser la relation entre la fréquence de renouvellement des entraîneurs et la performance des équipes.

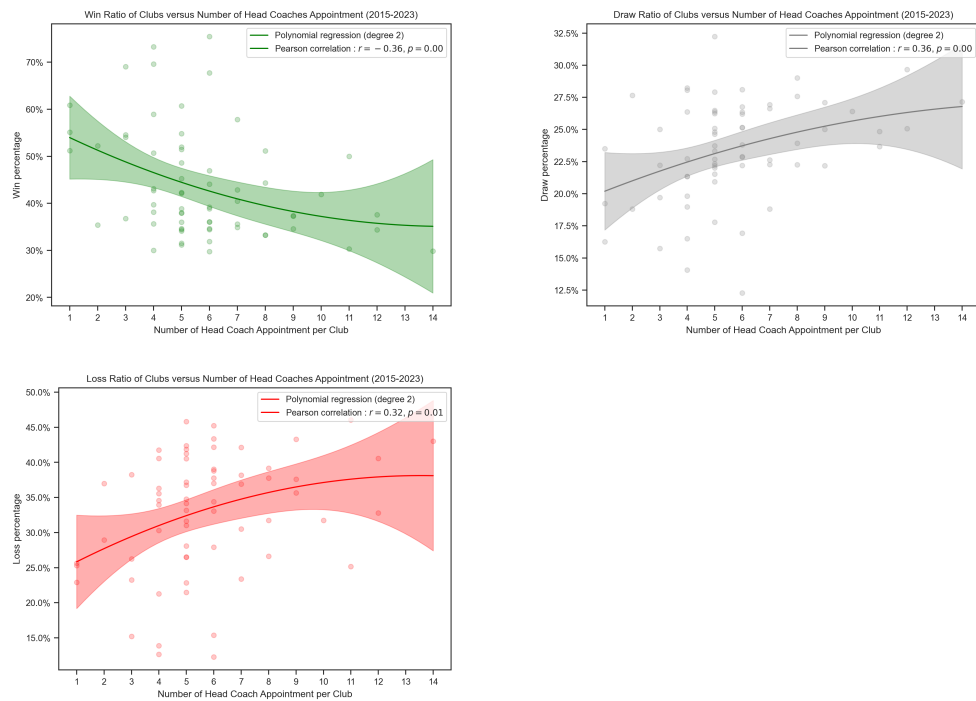


Figure 8: Clubs Results versus Number of Head Coaches Appointed by Club

Ainsi, nous observons :

- [Figure 8a](#) une corrélation négative modérée ( $r = -0.36$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre d'entraîneurs nommés par le club et son ratio de victoires.
- [Figure 8b](#) une corrélation positive modérée ( $r = 0.36$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre d'entraîneurs nommés par le club et son ratio de matches nuls.
- [Figure 8c](#) une corrélation positive faible ( $r = 0.32$ ) statistiquement significative ( $p = 0.01$ ) entre le nombre d'entraîneurs nommés par le club et son ratio de défaites.

Plusieurs hypothèses peuvent être émises pour expliquer ces résultats :

1. La performance de l'équipe comme facteur explicatif de la fréquence de renouvellement du coach :
  - Une équipe qui affiche de bons résultats a tendance à garder son entraîneur plus longtemps car il est une partie intégrante de la stabilité et de la performance actuelle de l'équipe. Aussi les entraîneurs choisissent de rester plus longtemps au sein d'un club qui affiche de bons résultats (avantages financiers, reconnaissance, satisfaction personnelle).
  - Une équipe qui affiche de mauvais résultats a tendance à renouveler ses entraîneurs plus fréquemment pour tenter d'améliorer ses performances. Aussi, les entraîneurs qui ne parviennent pas à obtenir de bons résultats sont plus susceptibles de quitter le club ou d'être licenciés.
2. La fréquence de renouvellement du coach comme facteur explicatif de la performance de l'équipe :
  - Les clubs qui gardent le même entraîneur pendant une longue période bénéficient de la stabilité et de la continuité dans la stratégie de jeu et la gestion de l'équipe, ce qui peut contribuer à améliorer les performances de l'équipe.
  - Les clubs qui choisissent de renouveler fréquemment leurs entraîneurs peuvent souffrir d'un manque de continuité et de stabilité dans la stratégie de jeu et la gestion de l'équipe, ce qui peut affecter négativement les performances de l'équipe.
3. Existence de facteurs tiers qui affecte à la fois le renouvellement du coach et la performance de l'équipe
  - Les clubs renouvellent fréquemment leurs entraîneurs en raison de problèmes internes ou de conflits avec les joueurs ou la direction du club, ce qui peut affecter négativement les performances de l'équipe.
  - Les clubs qui affichent de moins bon résultats sont moins attractifs pour les entraîneurs et les joueurs

#### 4.c. *Mobilité des entraîneurs et performance des équipes*

Ici, nous allons observer la relation entre les performances des entraîneurs et le nombre de clubs par lesquels ils sont passés. Cette relation nous permettra de voir si le nombre de clubs qu'un entraîneur a faits a un impact sur les résultats de l'équipe.

Au cours de la période 2017 - 2022, plus de 60 % des entraîneurs sportifs n'ont entraîné qu'un seul club. Environ 20 % des entraîneurs ont entraîné 2 clubs et seulement 10 % des entraîneurs ont entraîné plus de 3 clubs au cours de cette période (voir [Figure 9](#)).

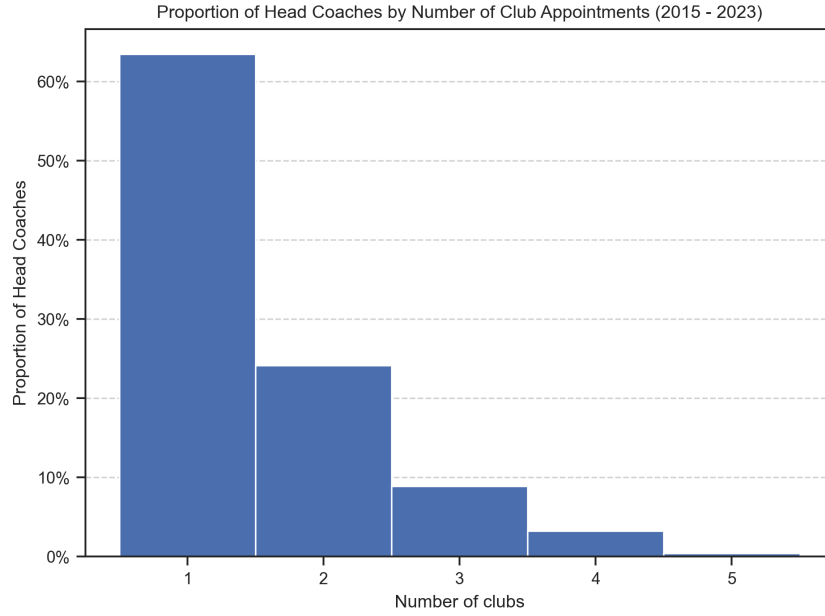


Figure 9: Proportion of Head Coaches by Number of Club Appointments (2015 - 2023)

La [Figure 10](#) s'intéresse aux pourcentages de victoires, de matchs nuls et de défaites de chaque mandat des entraîneurs par rapport au nombre d'équipe qu'ils ont entraîné jusqu'à présent au sein des équipes d'intérêts pour la saison 2015 - 2023. Ces graphiques permettent de visualiser la relation entre l'expérience acquise par l'entraîneur au cours de son passage successif au sein de plusieurs clubs et l'impact sur la performance de leurs équipes.

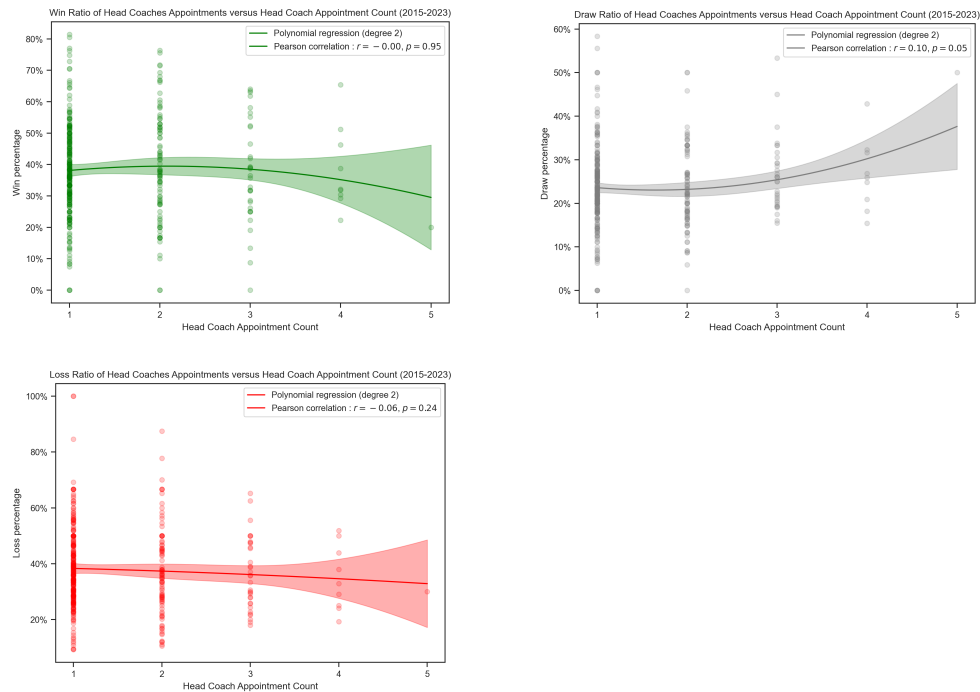


Figure 10: Head Coach Appointment Performance versus Appointment Count

Il n'existe pas de corrélation statistique significative entre le nombre de clubs pour lesquels un entraîneur a travaillé et son pourcentage de victoires ( $p = 0.95$ ), de matchs nuls ( $p = 0.05$ ) ou de défaites ( $p = 0.24$ ).

Il est possible que l'absence de corrélation statistiquement significative soit dû à l'absence d'une information complète sur l'expérience passée des entraîneurs.

La Figure 11 suivante diffère quant à elle car elle s'intéresse à la performance agrégée des entraîneurs au cours de leurs carrières au sein des équipes et saisons d'intérêt par rapport au nombre total de club pour lesquels ils ont travaillé durant la saison 2015 - 2023. Ces graphiques permettent de visualiser la relation entre la mobilité des entraîneurs et la performance de leurs équipes.

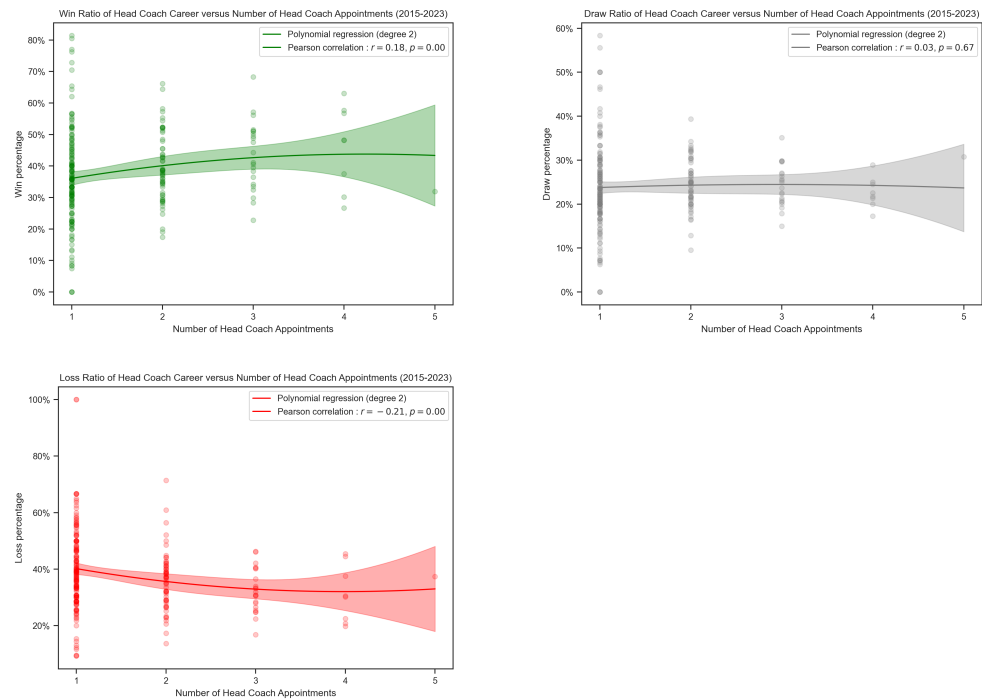


Figure 11: Performance of Head Coaches versus Number of Clubs Appointments

Ainsi, nous observons :

- Figure 11a une corrélation positive faible ( $r = 0.18$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre de clubs entraînés par le coach et son ratio de victoires.
- Figure 11b une corrélation positive très faible ( $r = 0.03$ ) mais statistiquement non significative ( $p = 0.67$ ) entre le nombre de clubs entraînés par le coach et son ratio de matchs nuls.
- Figure 11c une corrélation négative faible ( $r = -0.21$ ) statistiquement significative ( $p = 0.00$ ) entre le nombre de clubs entraînés par le coach et son ratio de défaites.

Plusieurs hypothèses peuvent être émises pour expliquer ces résultats :

1. La performance de l'équipe comme facteur explicatif de la mobilité du coach :

- Les entraîneurs qui obtiennent de bons résultats sont plus susceptibles d'être sollicités par d'autres clubs, ce qui peut expliquer pourquoi ils ont travaillé pour un plus grand nombre de clubs.
  - Les entraîneurs qui obtiennent de mauvais résultats sont potentiellement relégués à des ligues de division 2 ne faisant pas partie de nos ligues d'intérêts.
2. La mobilité du coach comme facteur explicatif de la performance de l'équipe :
    - Les entraîneurs qui ont travaillé pour un plus grand nombre de clubs ont acquis une expérience plus diversifiée et ont pu développer des compétences et des connaissances qui leur permettent d'obtenir de meilleurs résultats.
    - Les entraîneurs qui ont travaillé pour un plus petit nombre de club n'ont pas eu l'opportunité de développer des compétences et des connaissances aussi diversifiées, ce qui peut affecter négativement leurs performances et donc la performance de l'équipe.
  3. Le biais des données : en effet, nous n'avons collecté que les données des entraîneurs qui ont été actifs entre 2015 et 2023 et pour les ligues d'intérêts. Il est possible que les entraîneurs qui ont réussi leurs mandats soient plus susceptibles d'être sollicités par d'autres clubs faisant partie des ligues d'intérêts.

#### 4.d. Ancienneté de l'entraîneur et performance de l'équipe

Dans cette partie, nous cherchons une éventuelle relation entre l'expérience qu'un entraîneur a pu accumuler avec son ancienneté et les performances de l'équipe lorsqu'il est à sa tête. L'ancienneté pourrait bien être un facteur important au sein d'une équipe, car l'entraîneur gagne en expérience au fil des clubs par lesquels il est passé. Mais cette expérience a-t-elle une réelle incidence sur les résultats obtenus ?

La [Figure 12](#) montre la distribution de l'ancienneté des entraîneurs sportifs au sein de leur club. On observe que plus de 50 % des entraîneurs sportifs sont renouvelés après 1 an de mandat. Ce pourcentage augmente à 80 % après 2 ans de mandat et à 90 % après 3 ans de mandat.

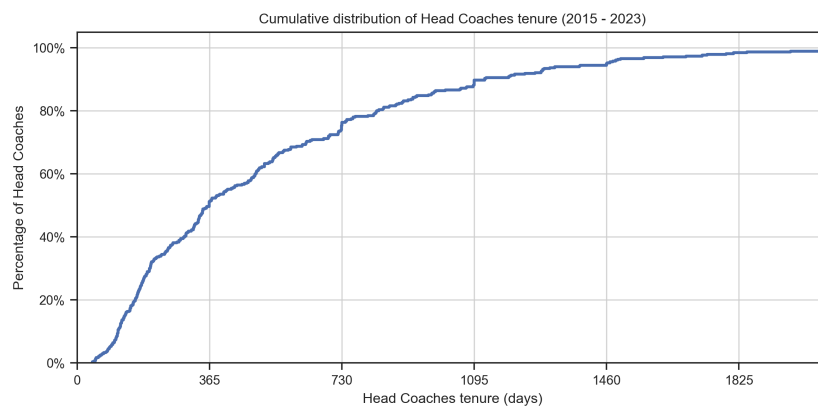


Figure 12: Empirical Cumulative Distribution Function of Head Coaches Tenure For Completed Appointments

La [Figure 13](#) montre que les entraîneurs de la Premier League restent en poste plus longtemps que ceux des autres ligues. De plus, les équipes de la Premier League changent moins souvent d'entraîneur que celles des autres ligues. À l'inverse, la LaLiga renouvelle fréquemment ses entraîneurs.

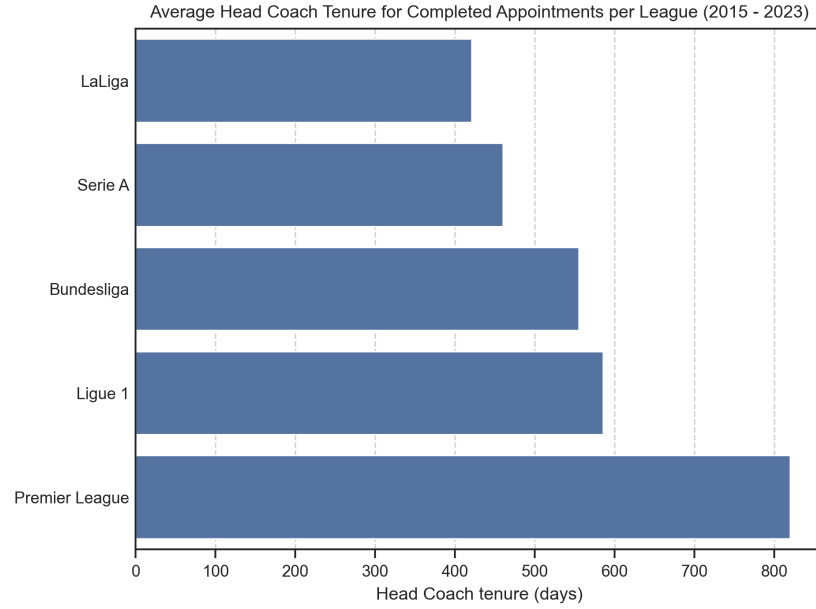


Figure 13: Average Head Coach Tenure for Completed Appointments per League

La [Figure 14](#) s'intéresse aux résultats des équipes durant le mandat d'un entraîneur en fonction de la durée du mandat de l'entraîneur. Cette analyse permet de mettre en évidence la relation entre l'ancienneté de l'entraîneur et la performance de l'équipe.

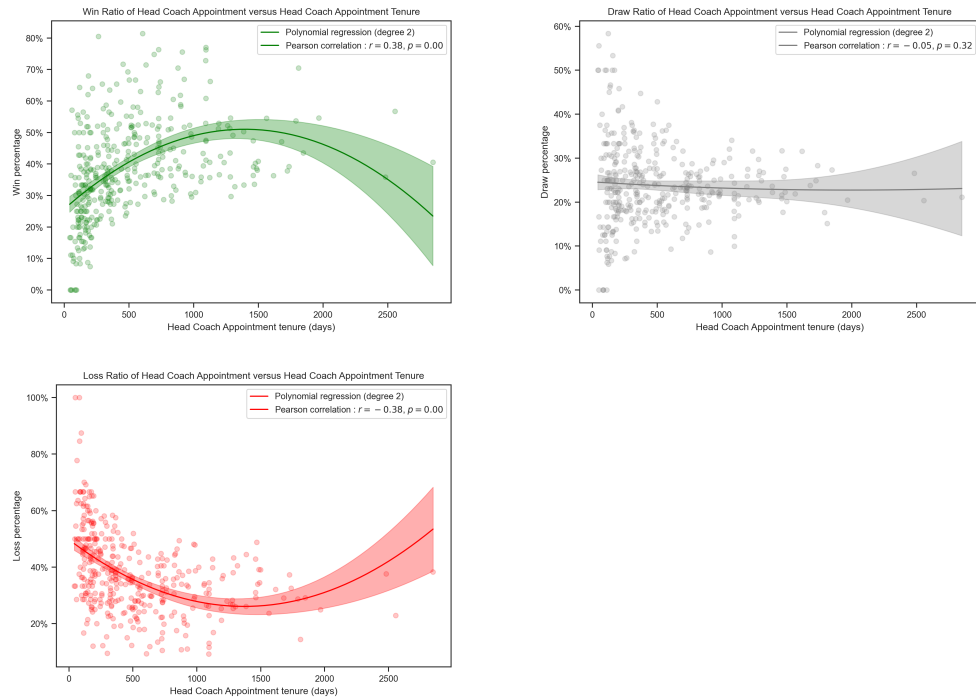


Figure 14: Team Results of Head Coach Appointment versus Appointment Tenure

Nous observons :

- [Figure 14a](#) une corrélation positive modérée ( $r = 0.38$ ) statistiquement significative ( $p = 0.00$ ) entre la durée du mandat de l'entraîneur et le pourcentage de victoires.



- [Figure 14b](#) une corrélation négative faible ( $r = -0.05$ ) mais statistiquement-non significative ( $p = 0.32$ ) entre la durée du mandat de l'entraîneur et le pourcentage de matchs nuls.
- [Figure 14c](#) une corrélation négative modérée ( $r = -0.38$ ) statistiquement significative ( $p = 0.00$ ) entre la durée du mandat de l'entraîneur et le pourcentage de défaites.

Les hypothèses explicatives du lien entre la durée du mandat de l'entraîneur et la performance de l'équipe sont similaires aux hypothèses fournies pour expliquer le lien entre la fréquence du renouvellement des entraîneurs par les clubs et la performance du club :

1. La durée du mandat comme facteur explicatif de la performance de l'équipe :
  - Les entraîneurs qui restent plus longtemps à la tête de l'équipe ont tendance à mieux connaître les joueurs et à mieux comprendre les forces et les faiblesses de l'équipe, ce qui peut contribuer à améliorer les performances de l'équipe.
  - Les entraîneurs qui restent moins longtemps à la tête de l'équipe n'ont pas eu le temps de mettre en place leur stratégie de jeu et de s'adapter à leur nouvel environnement, ce qui peut affecter négativement les performances de l'équipe.
2. La performance de l'équipe comme facteur explicatif de la durée du mandat de l'entraîneur :
  - Les équipes qui obtiennent de bons résultats ont tendance à garder leur entraîneur plus longtemps, car il est une partie intégrante de la stabilité et de la performance actuelle de l'équipe.
  - Les équipes qui obtiennent de mauvais résultats ont tendance à renouveler leurs entraîneurs plus fréquemment pour tenter d'améliorer leurs performances.
3. Existence de facteurs tiers qui affecte à la fois la durée du mandat de l'entraîneur et la performance de l'équipe.

#### 4.e. Ancienneté de l'entraîneur lors du match

La [Figure 2](#) montre la distribution des matchs en fonction de l'ancienneté de l'entraîneur lors du match. On observe qu'environ 63 % des matchs sont supervisés par un entraîneur ayant moins de 500 jours d'ancienneté au sein du club.

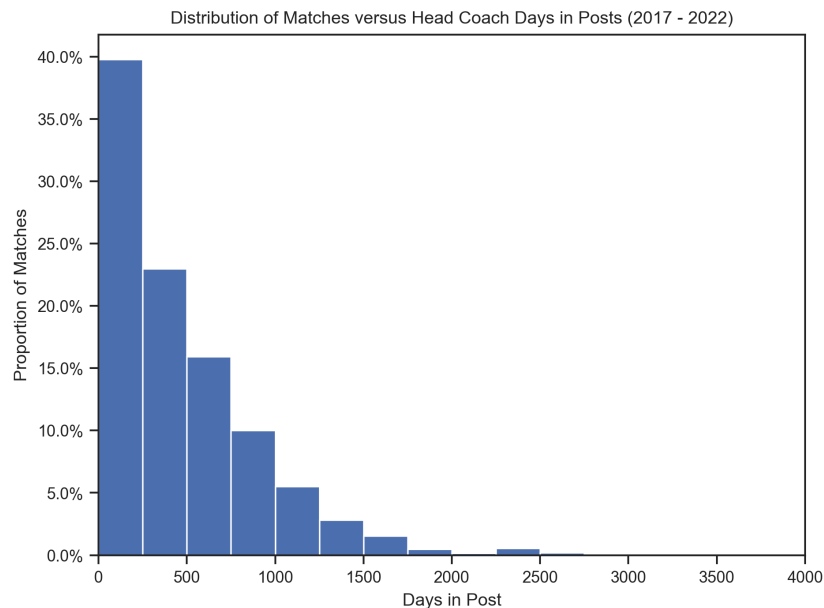


Figure 15: Distribution of Matches versus Head Coach Tenure on Match Day

La [Figure 16](#) montre le résultat individuel des matchs en fonction de l'ancienneté du coach lors du match. Ce graphique permet de visualiser la relation entre l'ancienneté du coach lors du match et la performance de l'équipe lors de ce match. Néanmoins, aucune des corrélations observées n'est statistiquement significative.

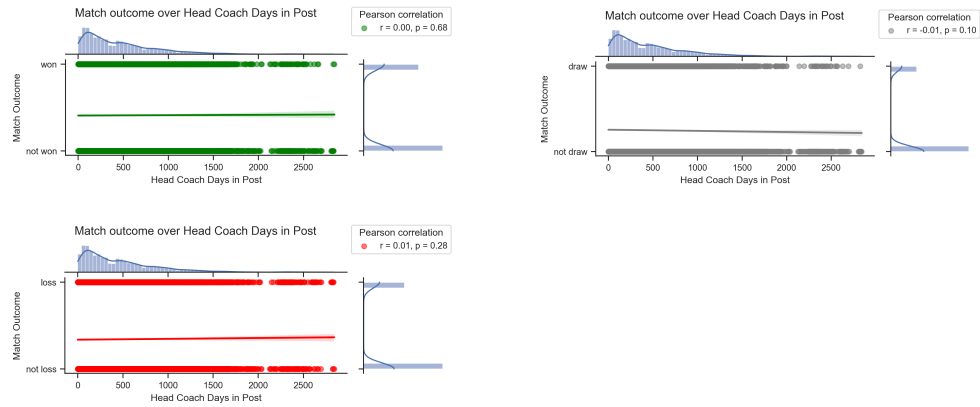


Figure 16: Match Outcome versus Head Coach Tenure on Match Day

Parmi l'ensemble des matchs où l'on possède des informations sur l'entraîneur sportif et où l'entraîneur sportif avait moins de 1500 jours d'ancienneté lors du match :

- le pourcentage de match gagné est de '41.14' %.
- le pourcentage de match nul est de '25.01' %.
- le pourcentage de match perdu est de '33.85' %.

La [Figure 17](#) illustre la proportion des résultats des matchs en fonction de l'ancienneté de l'entraîneur principal lors du match. Cette proportion est calculée à l'aide d'une moyenne pondérée sur une période de 100 jours. La zone verte représente les victoires, la zone grise les matchs nuls et la zone rouge les défaites. On n'observe pas de tendance claire dans les résultats des matchs en fonction de l'ancienneté de l'entraîneur lors du match.

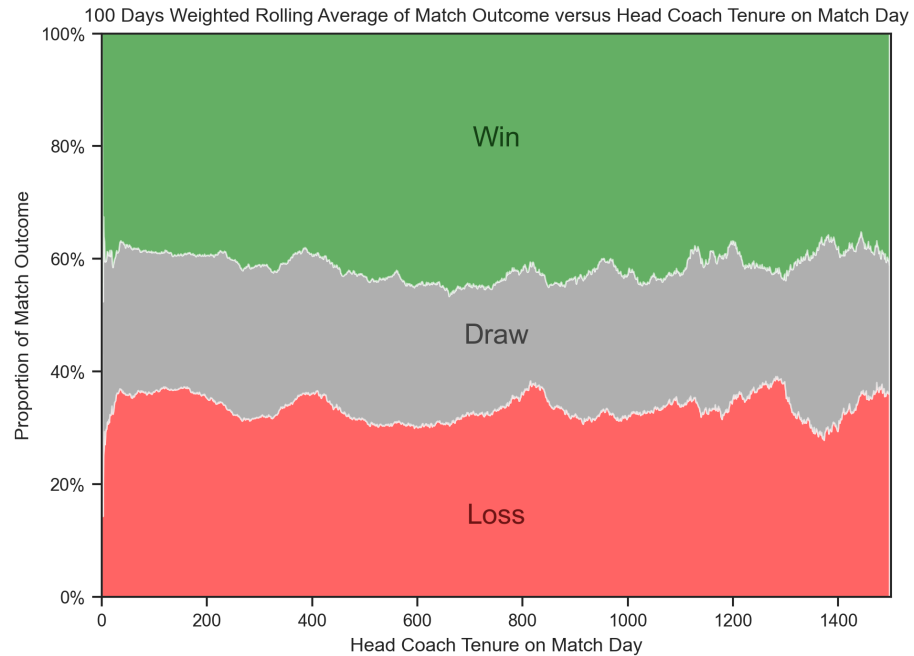


Figure 17: Weighted Rolling Average of Match Outcome versus Head Coach Tenure on Match Day

```
import numpy as np

def weighted_rolling_mean(data, weights, window_size=30):
    def weighted_mean(x):
        return np.average(data.loc[x.index], weights=weights.loc[x.index])

    return data.rolling(window_size, min_periods=1).apply(weighted_mean,
raw=False)
```

Code 2: Calcul des moyennes mobiles pondérées

## 5. PISTES D'AMÉLIORATIONS

Voici plusieurs pistes afin d'améliorer et compléter cette étude :

- Ajouter les données manquantes des mandants des entraîneurs sportifs
- Agrandir le cadre d'étude en incluant davantage de saison et de pays
- Améliorer les données sur l'expérience des entraîneurs en incluant des informations sur leur carrière avant 2015
- Observer les causes et conséquences d'un licenciement en cours de saison sur la performance de l'équipe le reste de la saison

## 6. CONCLUSION

L'analyse des données a permis de mettre en évidence plusieurs corrélations :

- La fréquence de renouvellement des entraîneurs par les clubs est corrélée négativement avec la performance de l'équipe.
- La mobilité des entraîneurs est corrélée positivement avec la performance des équipes entraînées.

- La durée du mandat de l'entraîneur est corrélée positivement avec la performance de l'équipe.

Ces résultats suggèrent que la stabilité et la continuité sont des facteurs importants de la performance sportive d'une équipe. Les entraîneurs qui restent longtemps à la tête de l'équipe ont tendance à mieux connaître les joueurs et à mieux comprendre les forces et les faiblesses de l'équipe, ce qui peut contribuer à améliorer les performances de l'équipe. En revanche, les entraîneurs qui changent régulièrement de club ont tendance à voir une amélioration de la performance de l'équipe. Aussi, les entraîneurs qui changent régulièrement de club semblent acquérir une expérience plus diversifiée qui leur permet d'obtenir de meilleurs résultats.

Cependant, il est difficile de tirer des conclusions définitives sur la causalité de ces relations, car il existe de nombreux autres facteurs qui peuvent influencer la performance d'une équipe de football. Par exemple, la qualité des joueurs, la stratégie de jeu, la gestion du club et d'autres facteurs peuvent également jouer un rôle important dans la performance de l'équipe. Il est donc important de prendre en compte ces facteurs lors de l'analyse des données et de ne pas tirer de conclusions hâtives.

## REFERENCES

- Levenshtein, V. I. (1965). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics. Doklady*, 10, 707–710. <https://api.semanticscholar.org/CorpusID:60827152>
- Wilke, C. O. (2019). *Fundamentals of Data Visualization*. O'Reilly Media. <https://clauswilke.com/dataviz/>