

Head coach dismissal effect on football team performance

Mathis Derenne¹, Ewann Boismartel¹, Scott Thomas¹, and Romain Biguet^{1†}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 2017 and 2022. 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

1.a. Cadre de la problématique

1.b. Source des données

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

- Fbref offre une gamme complète de données statistiques sur les joueurs, les équipes, les ligues et les compétitions de football du monde entier. Il propose des informations détaillées telles que les buts marqués, les passes décisives, les tirs au but, les interceptions et bien d'autres statistiques.
- Transfermarkt est une ressource en ligne majeure pour tout ce qui concerne les transferts de joueurs de football, les rumeurs de transferts, les valeurs marchandes des joueurs ainsi que les informations sur les contrats. Il offre une base de données exhaustive des joueurs, des clubs et des agents, ainsi que des détails sur les transferts passés et actuels.

Ces sites sont utilisés par les amateurs de football, les journalistes et les professionnels pour rester informés sur les évolutions au cours de la saison ou pendant les trêves/mercato.

1.c. Fiabilité des données

Ces sites sont très utilisés et considérés comme fiable. Fbref est entretenu par l'entreprise Sport Reference qui gère également d'autres sites spécialisés dans les statistiques sportives, comme Baseball-Reference et Basketball Reference. Les données sur Fbref sont souvent vérifiées et mises à jour régulièrement, ce qui contribue à leur fiabilité. Pour Transfermarkt, c'est aussi un site très utilisé pour les rumeurs de transferts et les transferts en général, il a une réputation de site fiable. Le site recueille des données sur les transferts, les valeurs marchandes des joueurs et d'autres détails liés aux contrats à partir de diverses sources, y compris les médias et les communiqués officiels des clubs. Cependant, c'est un site reliant des rumeurs de transferts, donc il peut y avoir des inexactitudes ou des spéculations qui ne se concrétisent pas toujours. Il est donc conseillé de vérifier les informations avec d'autres sources fiables, notamment lorsqu'il s'agit de transferts non confirmés

1.d. 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 effectué 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 librairies standards : Pandas, Numpy, Matplotlib, Seaborn, Scipy.

L'écriture de ce papier a été réalisé dans un fichier 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 majorité 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(2018, 2019, 2020, 2021, 2022)
```

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

2. EXTRACTION DES DONNÉES

L'extraction se fait aisément à l'aide de [WorldFootballR](#).

Un premier jeu de données 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. Un second jeu de données concernant les entraîneurs sportifs est récupéré à partir de [Transfermarkt](#). Il contient des informations sur les entraîneurs de football, notamment leur nom, leur date de naissance, leur nationalité, les dates de début et de fin de leur mandat, ainsi que des statistiques sur les matchs qu'ils ont dirigés.

3. PRÉ-TRAITEMENT DES DONNÉES

	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 des résultats de matchs

Étant donné que nous ne sommes pas intéressés par les résultats des équipes, mais plutôt par le résultat de chaque équipe, nous modifierons cette base de données en divisant les résultats du match en deux lignes distinctes, une pour chaque équipe. Cela nous permettra de calculer les statistiques pour chaque équipe séparément.

	Team	League	Country	HeadCoach	Appointed	EndDate	Tenure	Matches	Wins	Draws	Losses
0	Chelsea FC	Premier League	England	Mauricio Pochettino	2023-07-01	NaT	284	42	21	9	12
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	2022-09-03	389	100	63	19	18
3	Chelsea FC	Premier League	England	Frank Lampard	2019-07-01	2021-01-25	71	84	44	15	25
4	Chelsea FC	Premier League	England	Maurizio Sarri	2018-07-12	2019-06-30	51	63	40	11	12

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

On filtre dans un premier temps les entraîneurs qui n'ont pas été actif entre 2017 et 2022.

De plus, en vérifiant la qualité des données, nous avons remarqué qu'il existait parfois plus d'un coach pour une même période donnée :

There is 11 inconsistent record of head coach in 'FC Empoli ,Hellas Verona ,LOSC Lille ,Newcastle United ,Rayo Vallecano ,SSC Napoli ,Stade Reims ,Stade Rennais FC ,Torino FC ,West Ham United' teams.

	Team	Appointed	EndDate	Overlap
3395	Stade Reims	2014-07-01	2015-04-07	False
3394	Stade Reims	2015-04-07	2016-04-23	False
3393	Stade Reims	2016-07-01	2017-05-22	False
3392	Stade Reims	2017-05-22	2021-05-25	True
3391	Stade Reims	2018-07-01	2019-03-30	False
3390	Stade Reims	2021-06-23	2022-10-13	False
3389	Stade Reims	2022-10-13	2024-01-14	False

Table 3: Example of inconsistency in the head coach data

On exclu ces enregistrements du jeu de donnée.

Nous ajoutons aux données de match l'ancienneté du coach au sein de l'équipe lorsque ce match a été joué. Cela nous permettra de voir si l'ancienneté de l'entraîneur au sein d'une équipe a un impact sur sa performance.

Cependant, le nom des équipes dans les deux jeux de données sont différents. Nous devons corriger cela avant de pouvoir joindre les deux dataframes. Au total, les résultats de matchs contiennent 31 équipes et les mandats de coaches contiennent `{eval}`len(coach_team)` 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 58 équipes présente dans les mandats de coaches qui n'ont pas de correspondance dans les résultats de match.

Nous utiliserons la distance de Levenshtein afin de faire correspondre les équipes des mandats d'entraîneurs aux équipes des résultats de matchs. Nous restreignons 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 permet de rapidement vérifier la validité des correspondances de moins bonne certitudes. L'algorithme de la distance Levenshtein (Levenshtein, 1965) a été utilisé pour faire correspondre les noms des équipes. Cet algorithme permet de calculer la distance entre deux chaînes de caractères en mesurant le nombre minimum d'opérations nécessaires pour transformer une chaîne en une autre.

```
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

Team in head coach records	Team in match results
Newcastle United	Newcastle Utd
Crystal Palace	Crystal Palace
Chelsea FC	Chelsea
Tottenham Hotspur	Tottenham
Arsenal FC	Arsenal

Table 4: Exemple de correspondance des noms d'équipes

L'ancienneté du coach sportif au sein de l'équipe lors du match est ajouté à chaque ligne des données de résultat de match. Le tableau suivant est ainsi obtenu :

	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

4. PRÉSENTATION DES DONNÉES

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

Cela correspond à un total de 14873 matchs parmi 161 équipes. Parmi ces 161 équipes, nous n'avons pas d'information quant aux coachs sur une période des saisons 2015 - 2023 pour 96 de ces équipes. Cela représente '45%' des matchs sans information sur l'entraîneur sportif.

Les données des mandats des entraîneurs sportifs contiennent 398 enregistrements de 257 entraîneurs uniques.

		Number of Teams	Number of Matches	Average Goals per Match
League	Country			
Bundesliga	Germany	31	5263	1.49
La Liga	Spain	31	6276	1.30
Ligue 1	France	32	6445	1.33
Premier League	England	33	5667	1.26
Serie A	Italy	34	6096	1.36

Table 6: Summary of the match data

4.a. L'avantage de jouer à domicile

En moyenne, l'équipe qui joue à domicile marque '1.51' buts contre '1.18' buts pour l'équipe jouant à l'extérieur ('22%' moins de buts).

Cela se traduit par '43%' de matchs gagnés par les équipes jouant à domicile contre seulement '29' pour les équipes jouant à l'extérieur ('14' % moins de victoires). Les matchs nuls représentaient '0' % des matchs. Cet effet est appelé avantage à la maison.

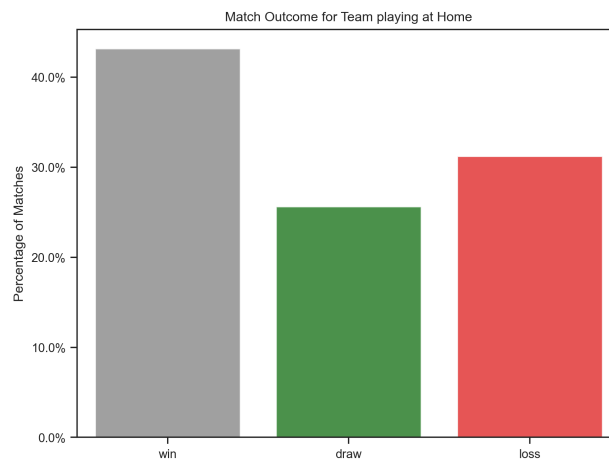


Figure 1: Venue effect on team's performance (2017 - 2022)

4.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 est la période pendant laquelle les équipes se préparent pour la saison suivante, notamment en recrutant de nouveaux joueurs et en changeant d'entraîneur.

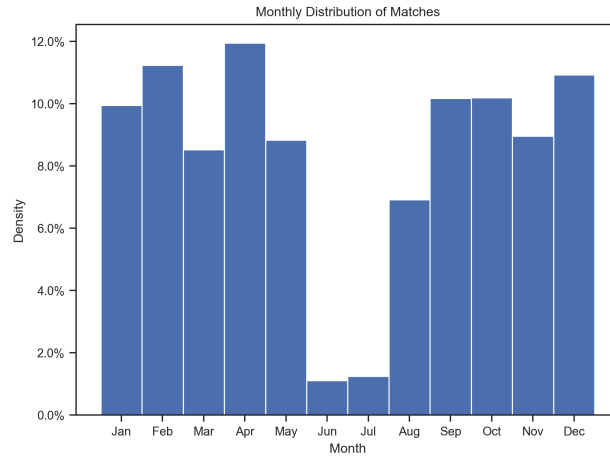


Figure 2: Monthly Distribution of Matches (2017 - 2022)

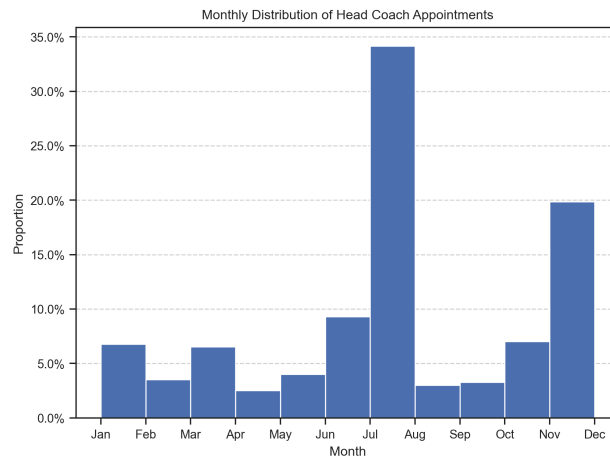


Figure 3: Monthly Distribution of Head Coaches Appointments

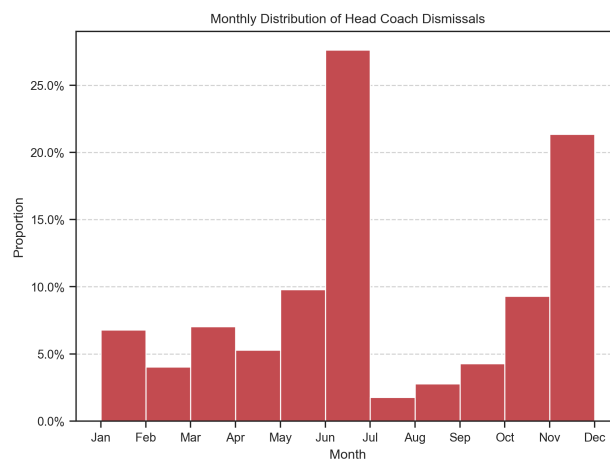


Figure 4: Monthly Distribution of Head Coaches Dismissals

Les licenciements d'entraîneurs sportif sont plus fréquents en fin de saison (voir Figure 3), tandis que les nominations d'entraîneurs sportif sont plus fréquentes en début de saison (voir Figure 4). Cela peut s'expliquer par le fait que les clubs cherchent à renouveler leur effectif et à se donner les meilleures chances de succès pour la saison suivante.

5. ANALYSE STATISTIQUE

5.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 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.

5.b. Fréquence de renouvellement du coach sur la performance du club

La Figure 5 nous informe quant à la distribution du nombre d'entraîneurs employés par les clubs durant la période 2017 - 2022. 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.

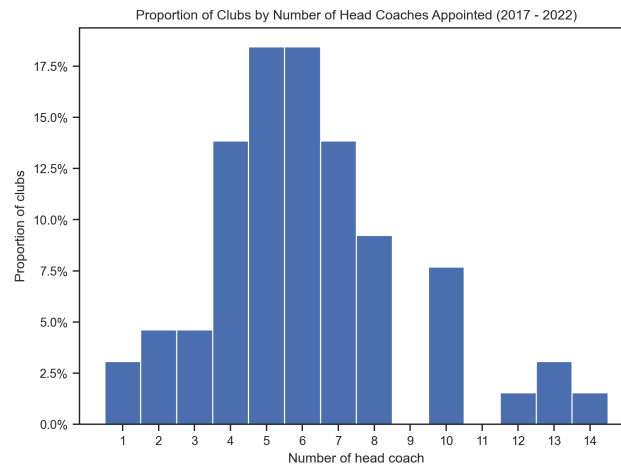


Figure 5: Proportion of Clubs by Number of Head Coaches Appointed (2017 - 2022)

La Figure 6 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.

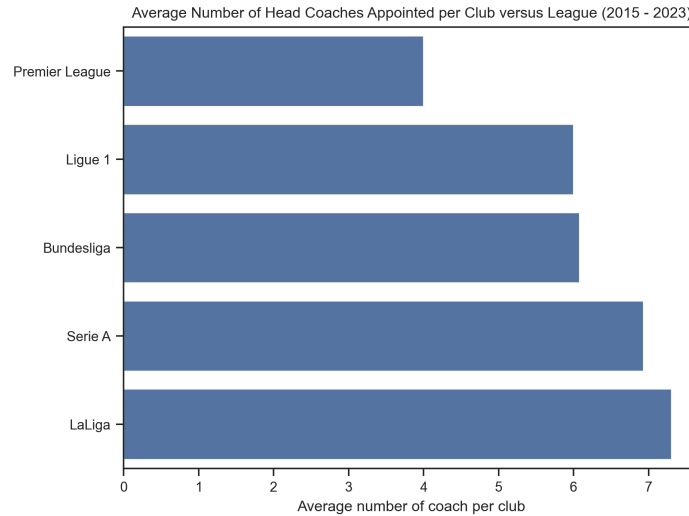


Figure 6: Average Number of Head Coaches Appointed per Club versus League (2017 - 2022)

Les [Figure 7](#), [Figure 8](#) et [Figure 9](#) s'intéressent 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 durant la période 2017 - 2022. Ces relations pourraient montrer l'effet de la fréquence de remplacement d'un coach sur les performances de l'équipe.

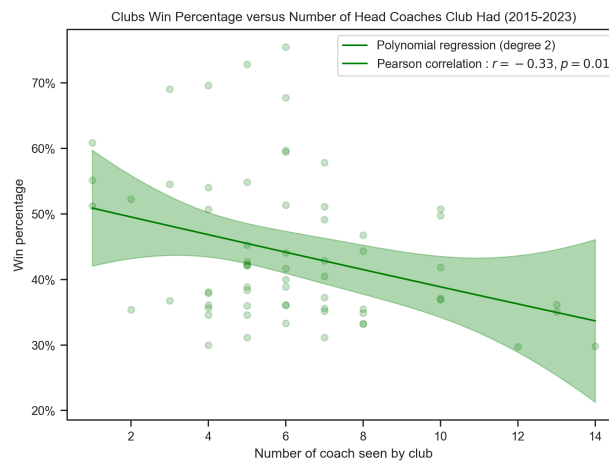


Figure 7: Win Ratio of Clubs versus Number of Head Coaches Appointed by Club

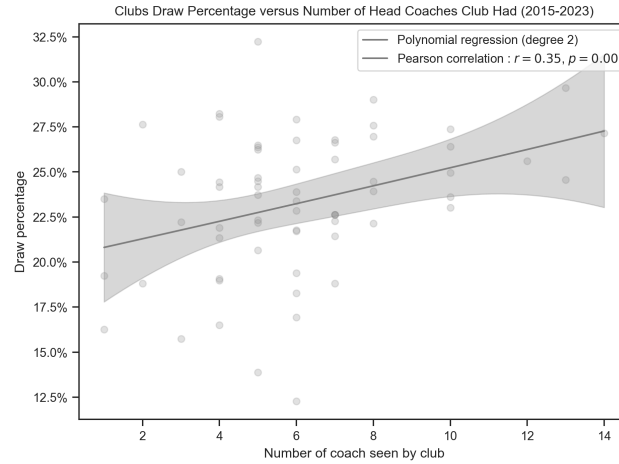


Figure 8: Draw Ratio of Clubs versus Number of Head Coaches Appointed by Club

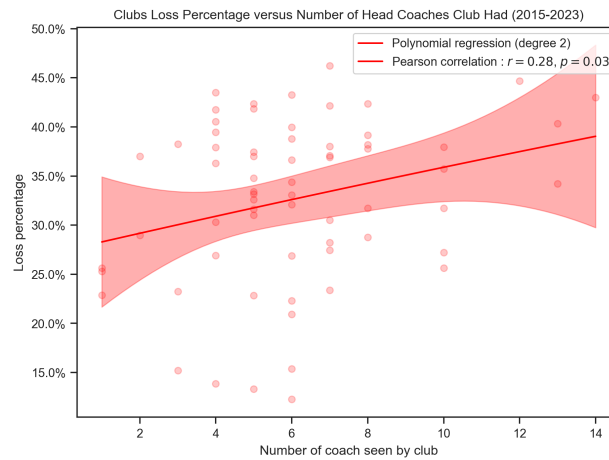


Figure 9: Loss Ratio of Clubs versus Number of Head Coaches Appointed by Club

Ainsi, nous observons :

- [Figure 7](#) une corrélation négative modérée ($r = -0.33$) statistiquement significative ($p = 0.01$) entre le nombre d'entraîneurs nommés par le club et son ratio de victoires.
- [Figure 8](#) une corrélation positive modérée ($r = 0.35$) statistiquement significative ($p = 0.00$) entre le nombre d'entraîneurs nommés par le club et son ratio de matchs nuls.
- [Figure 9](#) une corrélation positive faible ($r = 0.28$) statistiquement significative ($p = 0.03$) entre le nombre d'entraîneurs nommés par le club et son ratio de défaites.

Ces résultats suggèrent que le renouvellement régulier des entraîneurs a un impact négatif sur les performances des équipes. Cela pourrait s'expliquer par le fait que les entraîneurs ont besoin de temps pour s'adapter à leur nouvel environnement et pour mettre en place leur stratégie de jeu. De plus, 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.

5.c. Mobilité des coachs sur la performance des équipes

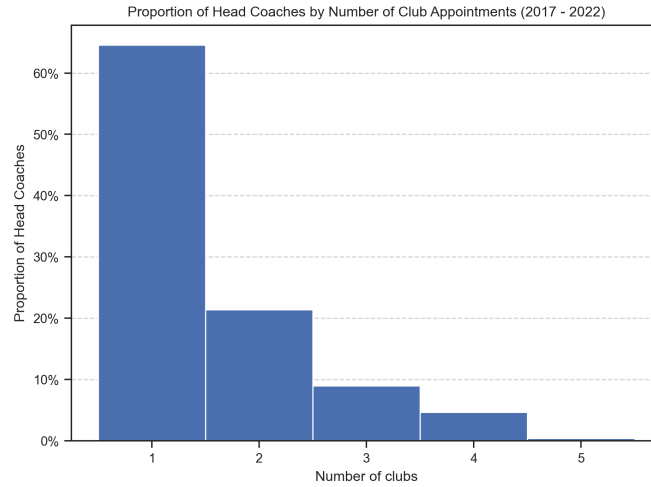


Figure 10: Proportion of Head Coaches by Number of Club Appointments (2017 - 2022)

Au cours de la période 2017 - 2022, plus de 55% des entraîneurs sportifs n'ont entraîné qu'un seul club. Environ 30% des entraîneurs ont entraîné 2 clubs et seulement 10% des entraîneurs ont entraîné plus de 3 clubs au cours de cette période (voir [Figure 10](#)).

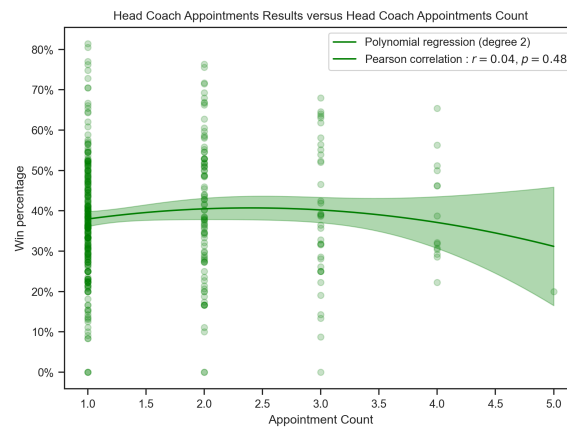


Figure 12: Head Coach Appointment Win Ratio versus Appointment Count

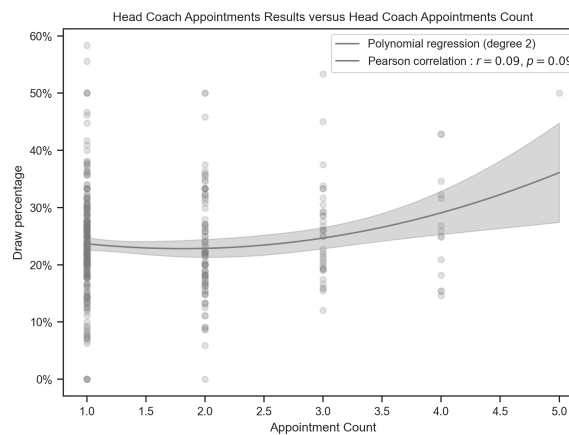


Figure 13: Head Coach Appointment Draw Ratio versus Appointment Count

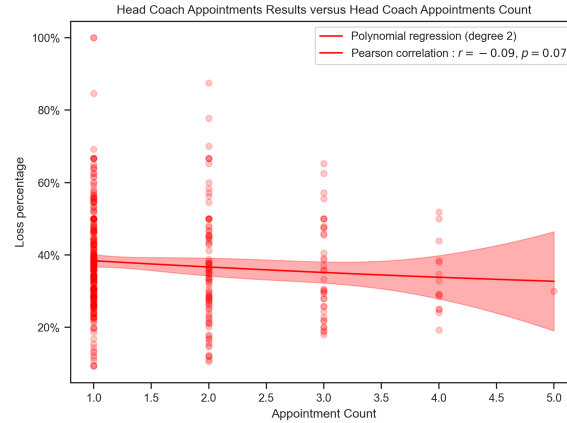


Figure 14: Head Coach Appointment Loss Ratio versus Appointment Count
Figure 11: Head Coach Appointment Performance versus Appointment Count

La Figure 11 s'intéresse aux pourcentage de victoires, de matchs nuls et de défaites des mandats individuel des entraîneurs par rapport au nombre de clubs pour lesquels ils ont travaillé durant la saison 2017 - 2022. Ces graphiques permettent de visualiser la relation entre l'expérience acquise par l'entraîneur et l'impact sur la performance de leurs équipes.

Nous observons qu'il n'existe pas de corrélation entre le nombre de clubs pour lesquels un entraîneur a travaillé et son pourcentage de victoires ($p = 0.48$), de matchs nuls ($p = 0.09$) ou de défaites ($p = 0.07$). Cela suggère que l'expérience acquise par l'entraîneur en travaillant pour plusieurs clubs n'a pas d'impact statistiquement observable sur la performance de l'équipe.

La Figure 15 suivante diffère quant à elle car elle s'intéresse à la performance agrégée des entraîneurs par rapport au nombre total de club pour lesquels ils ont travaillé durant la saison 2017 - 2022.

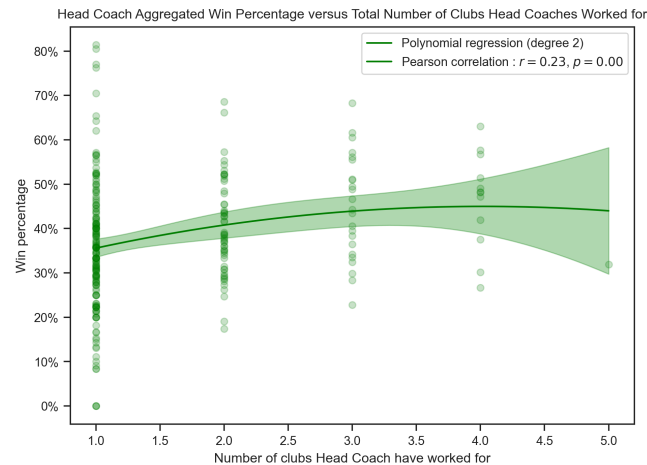


Figure 16: Head Coach Aggregated Win Ratio versus Total Number of Appointments

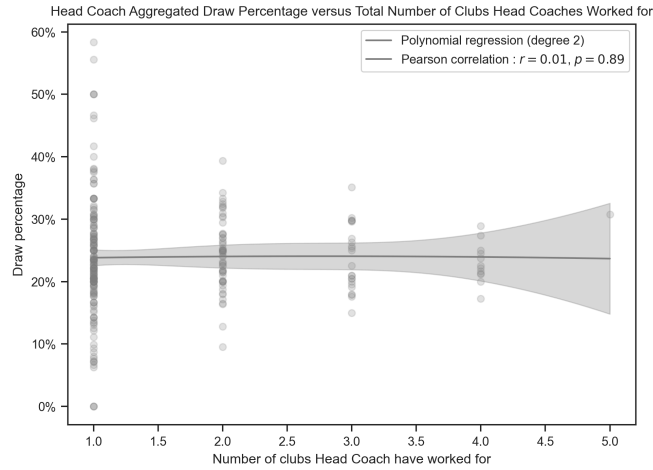


Figure 17: Head Coach Aggregated Draw Ratio versus Total Number of Appointments

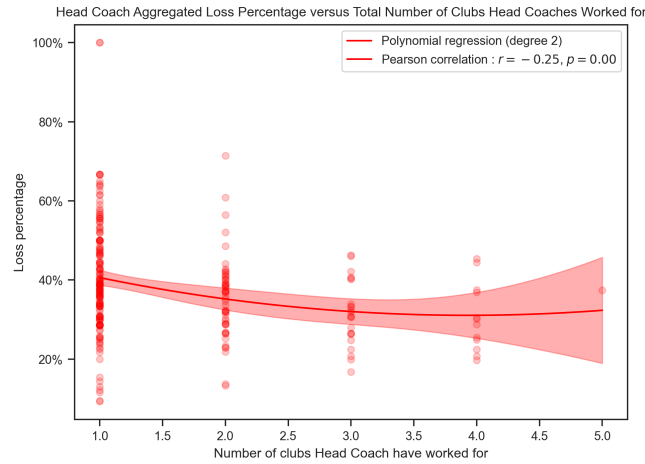


Figure 18: Head Coach Aggregated Loss Ratio versus Total Number of Appointments

Figure 15: Performance of Head Coaches versus Number of Clubs Appointments

Ces graphiques permettent de visualiser la relation entre la mobilité des entraîneurs par rapport à la performance de leurs équipes. Ainsi, nous observons :

- [Figure 16](#) une corrélation positive faible ($r = 0.23$) statistiquement significative ($p = 0.00$) entre le nombre de clubs entraînés par le coach et son ratio de victoires.
- [Figure 17](#) une corrélation positive très faible ($r = 0.01$) mais statistiquement non significative ($p = 0.89$) entre le nombre de clubs entraînés par le coach et son ratio de matches nuls.
- [Figure 18](#) une corrélation négative faible ($r = -0.25$) statistiquement significative ($p = 0.00$) entre le nombre de clubs entraînés par le coach et son ratio de défaites.

Ces graphiques ne suggèrent pas que le nombre de clubs pour lesquels un entraîneur a travaillé a un impact positif sur la performance des équipes qu'il entraîne car l'on a précédemment montré avec la [Figure 11](#) qu'il n'existait pas de corrélation statistiquement significative. Cependant, ce graphique suggère que les entraîneurs qui ont montré une performance supérieure à celle de leurs pairs ont tendance à travailler pour un plus grand nombre de clubs.

Cela pourrait s'expliquer par le fait que les entraîneurs performant soient davantage reconnus et donc sollicités par d'autres clubs, favorisant ainsi leur mobilité entre les clubs.

5.d. Ancienneté du coach sur la performance de l'équipe

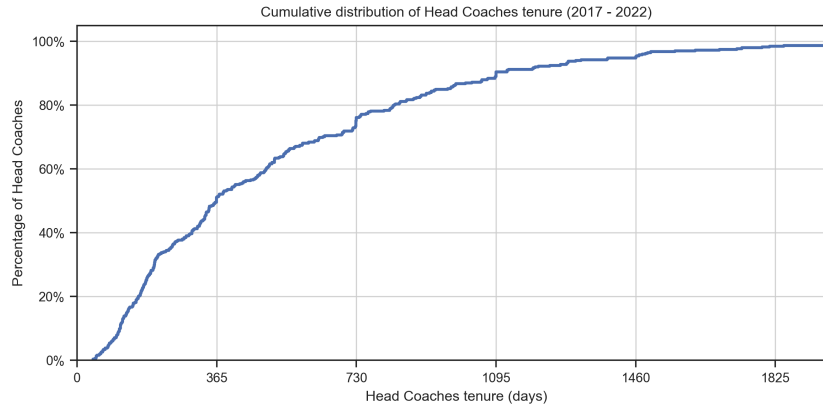


Figure 19: Empirical Cumulative Distribution Function of Head Coaches Tenure For Completed Appointments

La Figure 19 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.

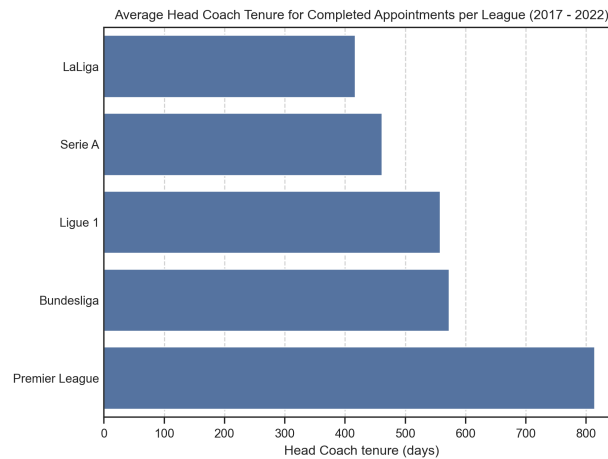


Figure 20: Average Head Coach Tenure for Completed Appointments per League

La Figure 20 s'intéresse à l'ancienneté des entraîneurs sportif et au renouvellement des entraîneurs sportifs par rapport aux ligues d'intérêt.

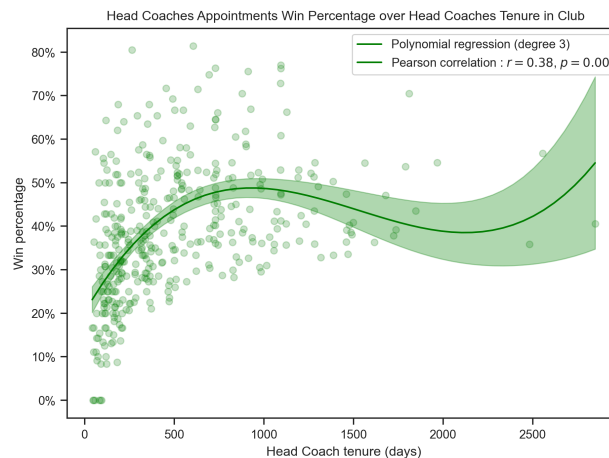


Figure 21: Win Ratio of Head Coaches Appointments versus Head Coach Tenure

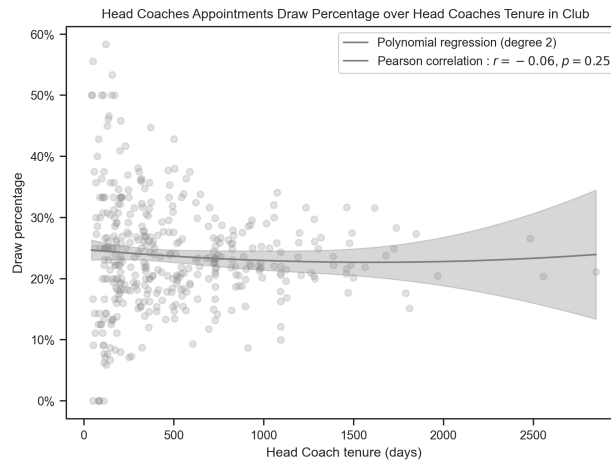


Figure 22: Draw Ratio of Head Coaches Appointments versus Head Coach Tenure

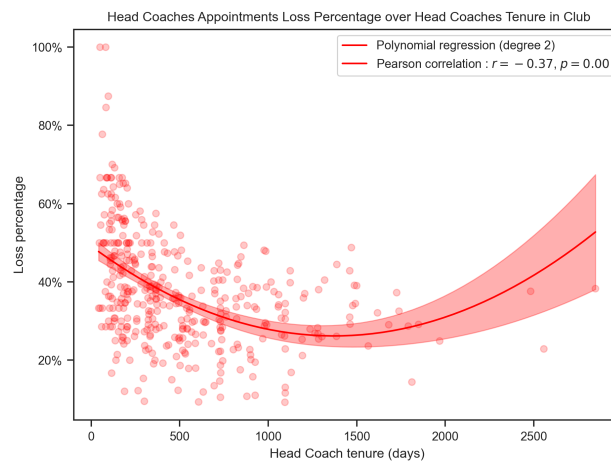


Figure 23: Loss Ratio of Head Coaches Appointments versus Head Coach Tenure

Les Figure 21 et Figure 22 et Figure 23 s'intéressent aux différents ratios de victoires, de matchs nuls et de défaites en fonction de la durée du coach au sein du club. Cette analyse permet de mettre en évidence le lien existant entre les résultats directs du coach et la durée de son mandat à la tête de l'équipe.

- Figure 21 une corrélation positive modérée ($r = 0.38$) statistiquement significative ($p = 0.00$) entre la durée du coach et le ratio de victoires.
- Figure 22 une corrélation négative faible ($r = -0.06$) mais statistiquement non significative ($p = 0.25$) entre la durée du coach et le ratio de matchs nuls.
- Figure 23 une corrélation négative modérée ($r = -0.37$) statistiquement significative ($p = 0.00$) entre la durée du coach et le ratio de défaites.

Il est à noter que deux des trois corrélations sont statistiquement significatives, étant donné que les valeurs de p sont inférieures à 0.05, qui est notre seuil d'acceptation. Ainsi, selon le coefficient de corrélation de Pearson r , nous observons que lorsque qu'un entraîneur gagne ses matchs, il augmente en même temps sa durée à la tête de l'équipe. En revanche, pour les défaites, plus il y en a, moins longtemps il reste coach de l'équipe. Les résultats neutres, représentés par les matchs nuls, non pas de signification statistique au vue du coefficient $p > 0.05$.

5.d.i. *Lien plus fin entre ancienneté du coach et performance de l'équipe:*

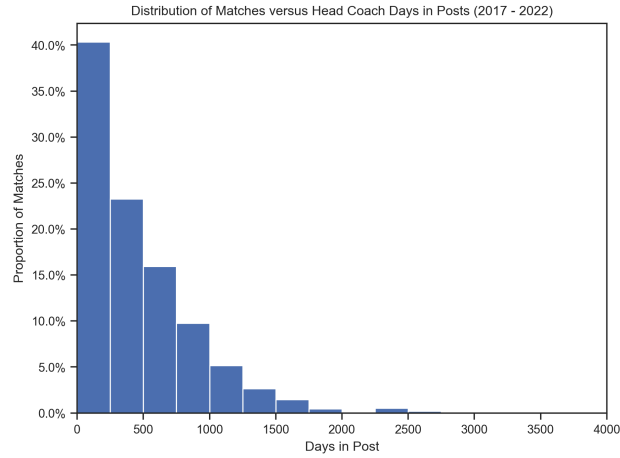


Figure 24: Distribution of Matches versus Head Coach Tenure on Match Day

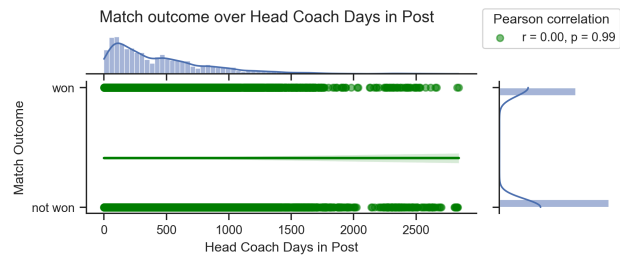


Figure 25: Match Win Outcome versus Head Coach Tenure on Match Day

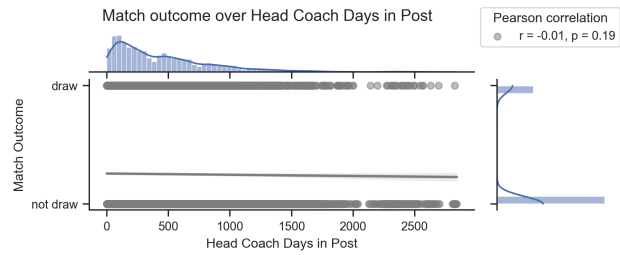


Figure 26: Match Draw Outcome versus Head Coach Tenure on Match Day

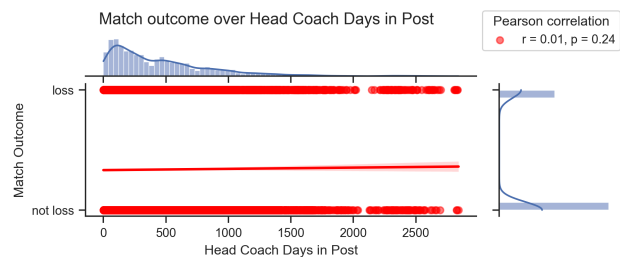


Figure 27: Match Loss Outcome versus Head Coach Tenure on Match Day

L'ancienneté, tout entraîneurs confondus a un effet positif sur la performance de l'équipe. Cela peut s'expliquer par le fait que les entraîneurs ont besoin de temps pour s'adapter à leur nouvel environnement et pour mettre en place leur stratégie de jeu. De plus, 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. Néanmoins, il est aussi probable que les équipes qui ont de bons résultats ont tendance

à garder leur entraîneurs plus longtemps, ce qui peut expliquer en partie la corrélation positive entre l'ancienneté du coach et la performance de l'équipe.

5.d.ii. Une visualisation graphique de l'effet de l'ancienneté du coach sur la performance de l'équipe:

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 '41.28' %.
- le pourcentage de match nul est de '25.28' %.
- le pourcentage de match perdu est de '33.44' %.

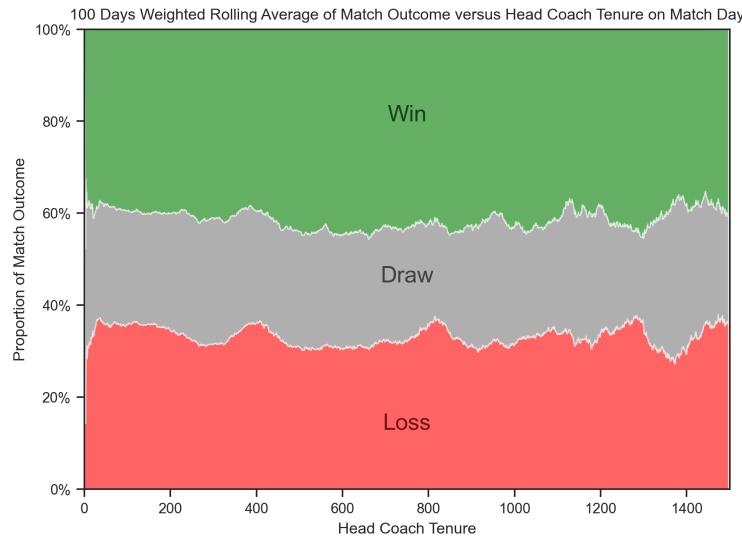


Figure 28: Weighted Rolling Average of Match Outcome versus Head Coach Tenure on Match Day

Le [Figure 28](#) illustre la proportion des résultats des matchs en fonction de l'ancienneté de l'entraîneur principal lors du match. Cette proportion est calculé à l'aide d'une moyenne pondérée sur une période de 100 jours. La zone verte représente les victoires, la zone grise représente les matchs nuls et la zone rouge représente les défaites. On peut observer que les victoires tendent à augmenter avec l'ancienneté de l'entraîneur, tandis que les défaites ont tendance à diminuer.

```
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

6. CONCLUSION

L'analyse des données a permis de mettre en évidence trois corrélation entre la performance d'une équipe et l'entraîneur sportif. Les résultats montrent que l'ancienneté du coach au sein de l'équipe est corrélée positivement avec la performance de l'équipe. En d'autres termes, plus un coach reste longtemps à la tête de l'équipe, meilleures sont les performances de l'équipe. De plus, le renouvellement régulier d'un coach sportif est corrélé négativement avec la performance de l'équipe. Enfin, les entraîneurs qui changent régulièrement de club ont tendance à voir une amélioration de la performance de l'équipe.

Ces résultats suggèrent que la stabilité et la continuité sont des facteurs importants pour la réussite d'une équipe de football. 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. Cela semble indiquer qu'un renouvellement régulier des entraîneurs peut être bénéfique pour l'équipe, mais que la stabilité et la continuité d'un coach peuvent également avoir un impact significatif sur les performances de l'équipe, en particulier à court terme.

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 sur la relation entre le licenciement d'un coach et la performance de l'équipe.

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

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/>