

Head coach dismissal effect on football team performance

Mathis Derenne¹, Ewann x¹, Scott daz¹, and Romain dcv^{1†} University of Rennes

Abstract

The goals of this paper is to investigate the effect of coach dismissal on team performance. To do that, we will use traditional statistical method that we apply to football teams.

Keywords coach dismissal, team performance

1. INTRODUCTION

1.a. Cadre de la problématique

Rappeler le rôle du coach dans le football et l'importance de son rôle dans la performance de l'équipe.

Sujet du TER : Comprendre l'effet du changement de club sur les performances du coach ET NON, comme le sujet initial (Rocaboy & Pavlik (2020)) Comprendre l'effet du changement de coach sur les performances du club Idée du prof : toutes choses étant égales par ailleurs (ceteris paribus) (idée d'un club représentatif), quelles sont les variations de performances d'un coach au cours du temps et lorsqu'il change de club

Impossible de créer "ce club égal par ailleurs" :

La création d'un club égal par ailleurs nécessite l'intervention d'un modèle qui permettrait, à partir des données du club (masse salariale, budget, performance passé du club, etc.) de normaliser la performance du club afin d'étudier précisément l'impact du coach sur cette performance

Ceci pose plusieurs problèmes :

1. Les variations de performances du coach sont difficilement observable au travers la performance de l'équipe (détailler)
2. Impossible de respecter l'hypothèse d'uncounfoundness requise par de nombreux modèles statistiques corrigeant les externalités (ex: propensity score / PSM) (citer papier propensity score + expliquer l'idée du propensity score pour artificiellement recréer un groupe contrôle et test artificiel, expliquer l'hypothèse d'uncounfoundness et pourquoi elle est nécessaire)
3. Biais de causalité (point le plus important !) : on suppose que c'est la performance du coach qui fait varier la performance de l'équipe or, dès lors que cette causalité n'est plus vérifiée on se mord la queue dans la création du modèle explicatif :

Supposons que ce soit la performance de l'équipe qui causent les variations de performance du coach. Le modèle explicatif, censé créer ce club égale par ailleurs, va être amené à normaliser plus fortement un club qui performe bien par le passé. Or si c'est la performance de l'équipe qui cause la performance du coach on est en train de normaliser les variations de performance du coach. (mentionner l'existence de test d'inversion de la causalité + référence au papier) (expliquer ce que sont les fuites de données (data leakage) et que l'absence de cette hypothèse de causalité provoque des fuites de données entre les externalités et la variable d'intérêt (la performance du coach)).

1. Le peu de donnée (retrouver le chiffre sur le nombre de club avec au moins 2 ou 3 changements de coach) (expliquer que dans la problématique initiale il y a bien plus de donnée car il y a davantage de club qui ont vu passer de coachs que de parcours individuel de coach au sein de clubs)

2. Problème de temporalité : les données sur les budgets des équipes, masse salariale ou valeur marchande des équipes ne sont pas disponible sous forme temporelle : impossible de savoir si la hausse de performance de l'équipe est due à la hausse du budget de l'équipe ou inversement.
3. Faible qualité des variables exogènes permettant l'analyse du système :
 - Manque " d'objectivabilité " des variables externes : masse salariale (pas représentative, ex : sous-traitance), valeur marchande des joueurs (hautement subjectif)
 - Manque de diversité des variables

Conclusion : Lors de l'analyse des effets dans un système, on raisonne généralement à petite entité égales par ailleurs
Exemple : On parle d'agent économique représentatif et rarement d'une économie représentative :

- On observe l'effet de l'économie sur un agent économique
- et NON l'effet d'un agent économique sur l'économie

(à nuancer pour ne pas déplaire aux micro-économistes et rappeler le cadre statistiques de l'étude d'effets quantifiables !).

La lecture de Wilke (2019) a permis l'amélioration de la qualité des graphiques et de la présentation des données.

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/mercatos.

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. Les 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, Statsmodels et Scikit-learn.

La création d'un tableau de bord interactif a été réalisé à l'aide de la librairie ipywidgets.

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 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 paramètres des fonctions utilisées dans les Notebooks Jupyter :

```
country <- c("ENG", "ESP", "ITA", "GER", "FRA")
year <- c(2018, 2019, 2020, 2021, 2022)
```

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 joués, le score final, le lieu du match et la date du match. Un second jeu de données concernant les coachs sportifs est récupéré à partir de [Transfermarkt](#). Il contient des informations sur les coachs 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	2018	2017-08-11	Arsenal	4.0	Leicester City	3.0
1	Premier League	England	2018	2017-08-12	Watford	3.0	Liverpool	3.0
2	Premier League	England	2018	2017-08-12	Crystal Palace	0.0	Huddersfield	3.0
3	Premier League	England	2018	2017-08-12	West Brom	1.0	Bournemouth	0.0
4	Premier League	England	2018	2017-08-12	Chelsea	2.0	Burnley	3.0

Table 1: Extrait du jeu de donnée des résultats de matchs

Since we are not interested in match opponents but rather individual team's result, we will modify this dataframe by splitting the match results into two separate rows, one for each team. This will allow us to calculate the statistics for each team separately.

	Team	League	Country	HeadCoach	Appointed	EndDate	Tenure	Matches	Wins	Draws	Losses
0	Manchester City	Premier League	England	Pep Guardiola	2016-07-01	NaT	2838	461	341	56	64
1	Manchester City	Premier League	England	Manuel Pellegrini	2013-07-01	2016-06-30	1095	166	101	27	38

2	Manchester City	Premier League	England	Roberto Mancini	2009-12-12	2013-05-13	241	191	113	38	40
3	Manchester City	Premier League	England	Mark Hughes	2008-06-02	2009-12-19	63	77	37	15	25
4	Manchester City	Premier League	England	Sven-Göran Eriksson	2007-07-02	2008-06-03	32	45	19	11	15

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

On filtre dans un premier temps les coachs qui n'ont pas été actif entre 2018 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 inconsistent record of head coach in teams.

	Team	Appointed	EndDate	Overlap
3497	Stade Reims	2017-05-22	2021-05-25	True
3496	Stade Reims	2018-07-01	2019-03-30	False
3495	Stade Reims	2021-06-23	2022-10-13	False
3494	Stade Reims	2022-10-13	2022-12-31	False

Table 3: Example of inconsistency in the head coach data

On exclu ces enregistrements de coachs.

3.a. Joining head coach with match results

We would like to add information about how long head coach has been in charge of the team when the match was played. This will allow us to see if the head coach's tenure have any impact on the match result.

However, when trying to join the two dataframes, we found that team names are not consistent between the two dataframes. We will need to fix this before we can join the two dataframes. In total, match_results dataset contains teams and head_coach dataset contains teams. However some teams name are different between the two datasets. For example 'Liverpool' in match_results is 'Liverpool FC' in head_coach. This is problematic as we will need to join data on team's columns.

In total, there is teams present in head coach records that are not in match results and teams present in match results but not in head coach records.

We will use Levenshtein distance to find the closest team of *match_results* that match each team in head coach records. We will then manually check the results to ensure that the matches are correct. L'algorithme de la distance Levenshtein (Levenshtein, 1965) a été utiliser 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 coach_team in coach_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}")

```

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

	Team in head coach records	Team in match results
0	Manchester City	Manchester City
1	Juventus FC	Juventus
2	Crystal Palace	Crystal Palace
3	Villarreal CF	Villarreal
4	Udinese Calcio	Udinese

Table 4: Exemple de correspondance des noms d'équipes

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

	League	Country	Date	Team	Goals	Result	isHome	HeadCoach	DaysInPost
0	Premier League	England	2017-08-11	Arsenal	4.0	win	True	Arsène Wenger	7619.0
1	Premier League	England	2017-08-12	Chelsea	2.0	loss	True	Antonio Conte	407.0
2	Premier League	England	2017-08-12	Brighton	0.0	loss	True	Chris Hughton	955.0
3	Premier League	England	2017-08-13	Newcastle Utd	0.0	loss	True	Rafael Benítez	520.0
4	Premier League	England	2017-08-13	Manchester Utd	4.0	win	True	José Mourinho	408.0

Table 5: Extrait du jeu de donnée sur les matchs transformés

4. PRÉSENTATION DES DONNÉES

We collected matches results and head coach records from Men's Football First Divisions during - seasons for the following leagues .:

This amount to a total of matches across teams. Out of these we don't have any records of head coach for teams.

For certain team information about head coach is present but no throughout the study period. From match results this represent% of the matches.

There is a total of unique head coaches in the dataset and records of head coach appointments.

		Number of Teams	Number of Matches	Average Goals per Match
League	Country			
Bundesliga	Germany	27	3026	1.53
La Liga	Spain	28	3943	1.31

Ligue 1	France	28	3620	1.36
Premier League	England	28	3706	1.37
Serie A	Italy	28	3953	1.44

Table 6: Summary of the match data

4.a. L'avantage de jouer à domicile

In average, team playing at home scored goals while away team scored {away_goals:.2f} goals (% less goals).

This resulted in matches won by team playing at home vs for the team playing away (% less wins). Draw matches accounted for % of the matches. This effect is called home advantage.

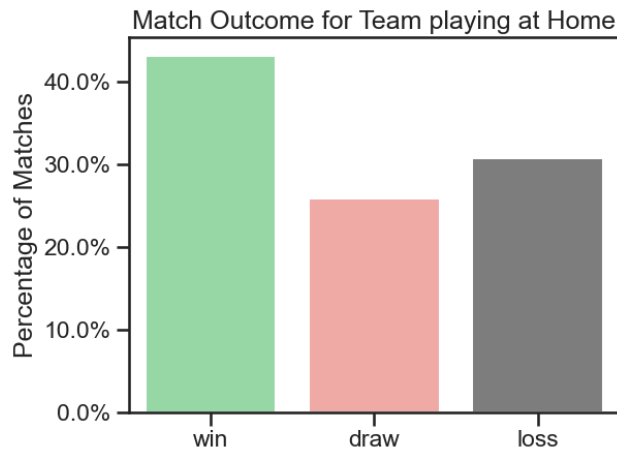


Figure 1: Venue effect on team's performance (2017 - 2022)

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.

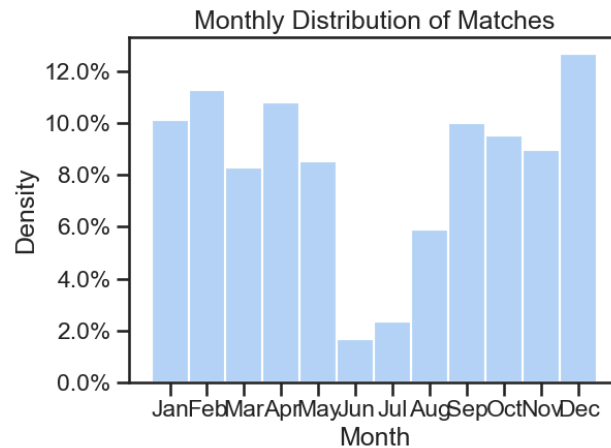


Figure 2: Monthly Distribution of Matches (2017 - 2022)

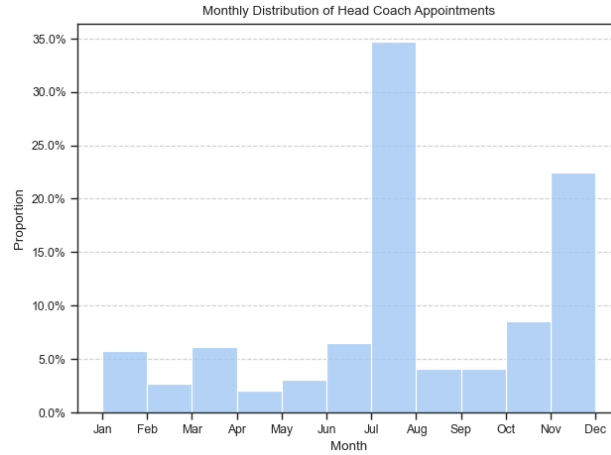


Figure 3: Monthly Distribution of Head Coaches Appointments

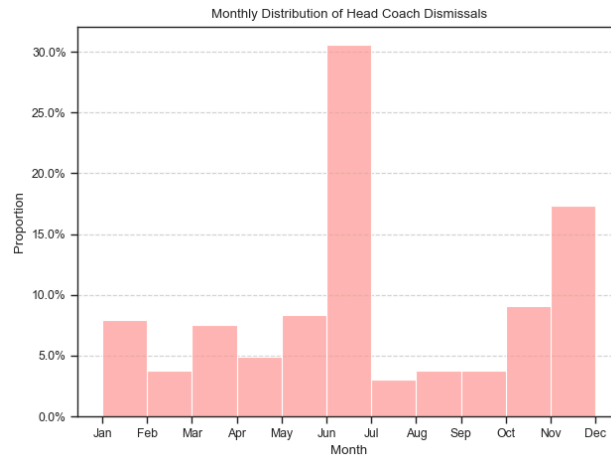


Figure 4: Monthly Distribution of Head Coaches Dismissals

Les licenciements de coachs sont plus fréquents en fin de saison (voir [Figure 3](#)), tandis que les nominations de coachs 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 STATISTIQUES

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. C'est un coefficient variant entre -1 et 1 .

- Lorsqu'il est de 1 , il indique une corrélation linéaire parfaite positive, ce qui signifie que lorsque la valeur d'une variable augmente, la valeur de l'autre variable augmente proportionnellement.
- Lorsqu'il est de -1 , il indique une corrélation linéaire parfaite négative, ce qui signifie que lorsque la valeur d'une variable augmente, la valeur de l'autre variable diminue proportionnellement. -
- De plus, lorsqu'il est de 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 d'une analyse statistique. 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 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. L'effet du changement de club sur la performance du coach

Au cours de la période 2017 - 2022, plus de 55% des coachs sportifs n'ont entraîné qu'un seul club. Environ 30% des coachs ont entraîné 2 clubs et seulement 10% des coachs ont entraîné plus de 3 clubs au cours de cette période (voir Figure 5).

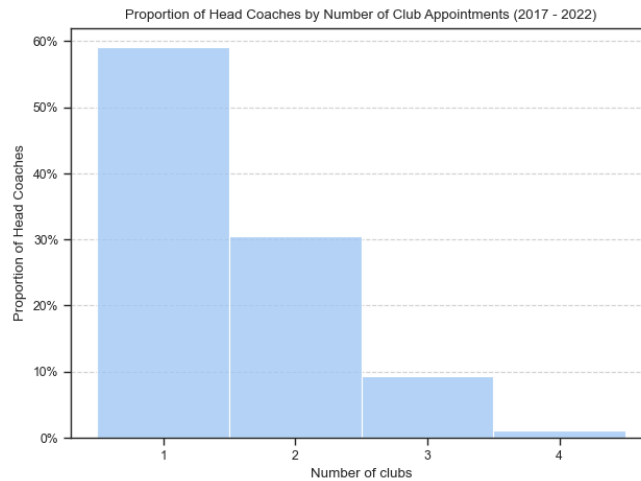


Figure 5: Proportion of Head Coaches by Number of Club Appointments (2017 - 2022)

Les Figure 6 et Figure 7 et Figure 8 aux ratios de victoires, de matchs nuls et de défaites en fonction du nombre de clubs différents dans lesquels l'entraîneur est passé. Dans ces régressions linéaires, nous examinerons l'effet de l'expérience de coaching d'un entraîneur sur ses résultats. Voici les coefficients obtenus :

- Pour les victoires, $r=0.19$ et $p=0.01$. Nous observons une corrélation positive de faible à modérée entre les deux variables, qui est statistiquement significative.
- Pour les matchs nuls, $r=-0.09$ et $p=0.25$. Cela indique une corrélation très faible et non significative entre les deux variables, ce qui ne nous permet pas de conclure.
- Pour les défaites, $r=-0.18$ et $p=0.02$. Nous constatons donc une corrélation négative faible mais statistiquement significative entre les deux variables examinées.

Nous observons un coefficient de Pearson positif pour les victoires, ce qui suggère que plus l'entraîneur a d'expérience (nombre de clubs entraînés), plus il y a de victoires. En revanche, pour les matchs nuls et les défaites, moins l'entraîneur a d'expérience dans différents clubs, plus il y a de matchs nuls et de défaites. On peut donc conclure que l'expérience de l'entraîneur est un facteur important pour la performance lors des matchs.

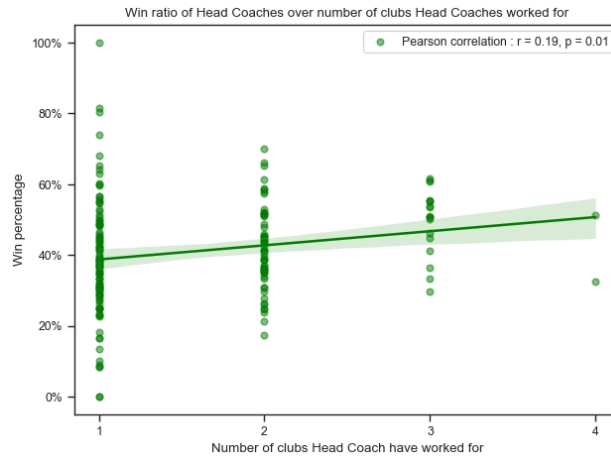


Figure 6: Win Ratio of Head Coaches versus Number of Clubs Appointments

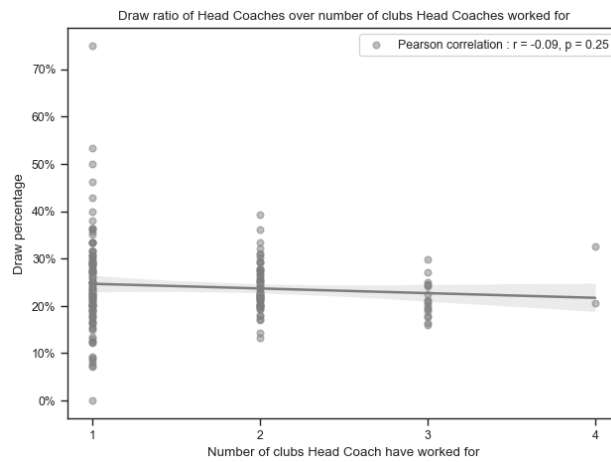


Figure 7: Draw Ratio of Head Coaches versus Number of Clubs Appointments

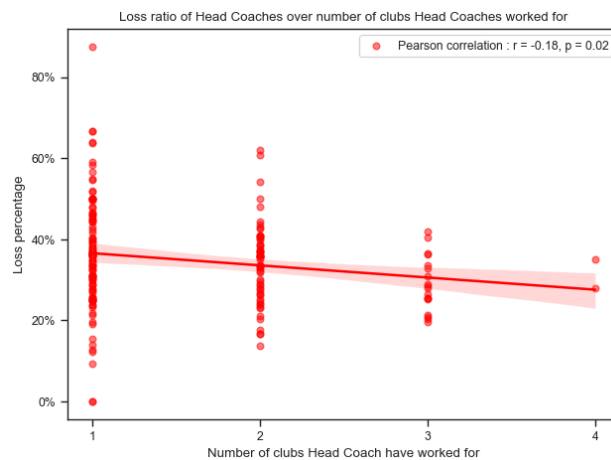


Figure 8: Loss Ratio of Head Coaches versus Number of Clubs Appointments

5.c. *L'effet du renouvellement régulier du coach sur la performance de l'équipe*

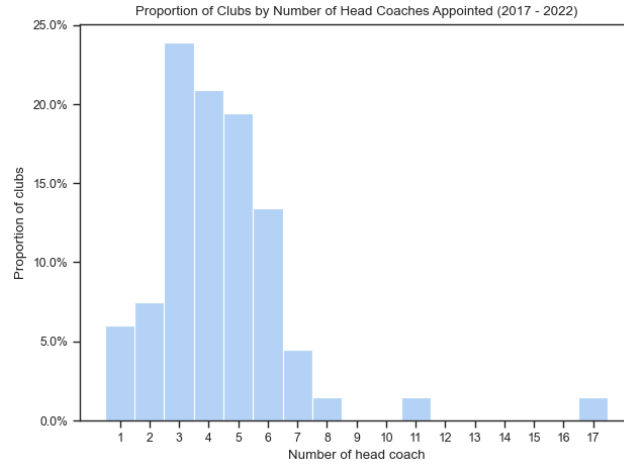


Figure 9: Proportion of Clubs by Number of Head Coaches Appointed (2017 - 2022)

Lorsque l'on s'intéresse au nombre de coach employés par les clubs durant la période 2017 - 2022, on observe que plus de 85% des clubs ont employés au moins 3 coachs différents (voir [Figure 9](#)).

La [Figure 10](#) nous montre que les coachs de la Premier League ont une ancienneté plus longue que les coachs des autres ligues. De plus, les clubs de la Premier League ont tendance à nommer moins de coachs que les clubs des autres ligues. Inversement, c'est LaLiga qui a la plus faible ancienneté moyenne des coachs et qui nomme le plus de coachs.

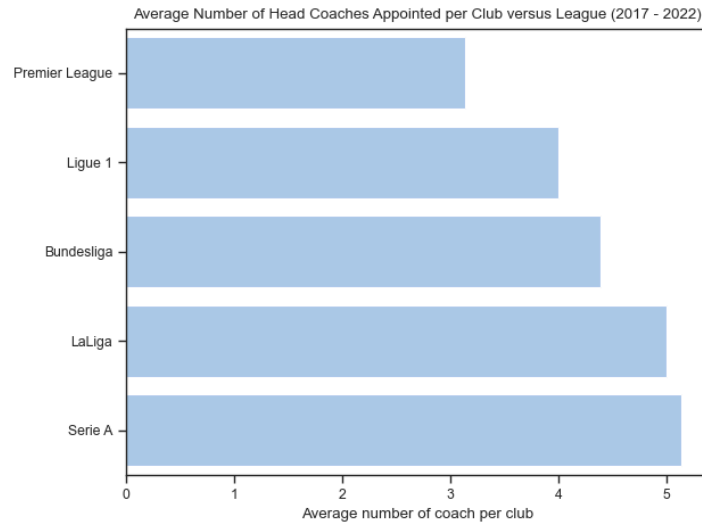


Figure 10: Average Number of Head Coaches Appointed per Club versus League (2017 - 2022)

Les [Figure 11](#) et [Figure 12](#) et [Figure 13](#) s'intéressent aux ratios de victoires, de matchs nuls et de défaites en fonction du nombre d'entraîneurs nommés à la tête de l'équipe. Ces relations pourraient montrer l'effet d'un changement régulier d'entraîneur sur les résultats de l'équipe. Les coefficients sont :

- Pour les victoires, $r=-0.27$ et $p=0.03$. Ceci indique une corrélation négative de faible à modérée entre les deux variables, qui est statistiquement significative.
- Pour les matchs nuls, $r=0.25$ et $p=0.04$. Cela indique une corrélation positive de faible à modérée entre les deux variables, qui est statistiquement significative, mais pas très forte.

- Pour les défaites, $r=0.24$ et $p=0.05$, ce qui indique une corrélation positive de faible à modérée entre les deux variables, qui est statistiquement significative.

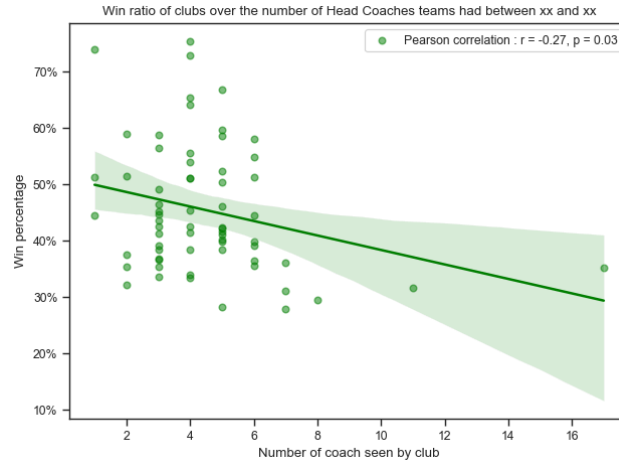


Figure 11: Win Ratio of Clubs versus Number of Head Coaches Appointed by Club

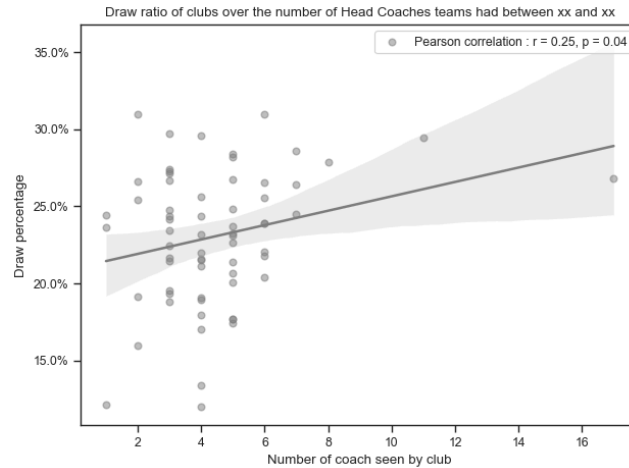


Figure 12: Draw Ratio of Clubs versus Number of Head Coaches Appointed by Club

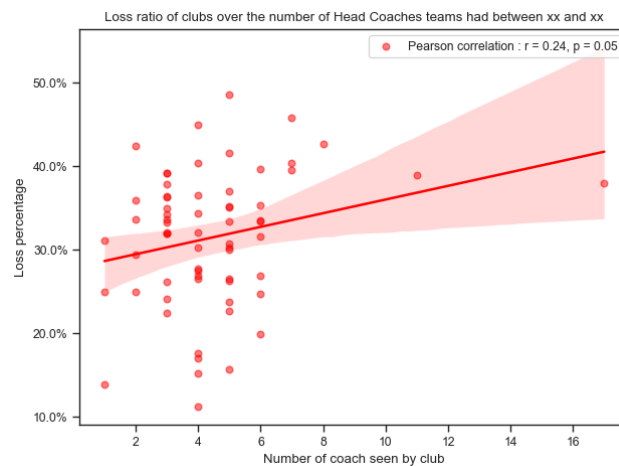


Figure 13: Loss Ratio of Clubs versus Number of Head Coaches Appointed by Club

On remarque qu'un coefficient de Pearson négatif pour les victoires nous montre que plus il y a d'entraîneurs différents dans un club pour une période donnée (potentiellement courte), moins il y aura de victoires, nous avons ici une relation négative entre les deux variables. En revanche, nous avons une relation positive pour les nuls et les défaites, ce qui signifie que plus il y aura d'entraîneurs, plus il y aura de matchs nuls et de défaites. Ainsi, changer d'entraîneurs n'est pas forcément le bon moyen pour gagner davantage, mais c'est potentiellement contre-productif pour l'équipe car pour chaque nouvel entraîneur, il y a un cadre de vie différent, avec des règles et des situations propres à chacun.

5.d. L'effet de l'ancienneté du coach sur la performance de l'équipe

Plus de 50% des coachs 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 (voir [Figure 14](#))

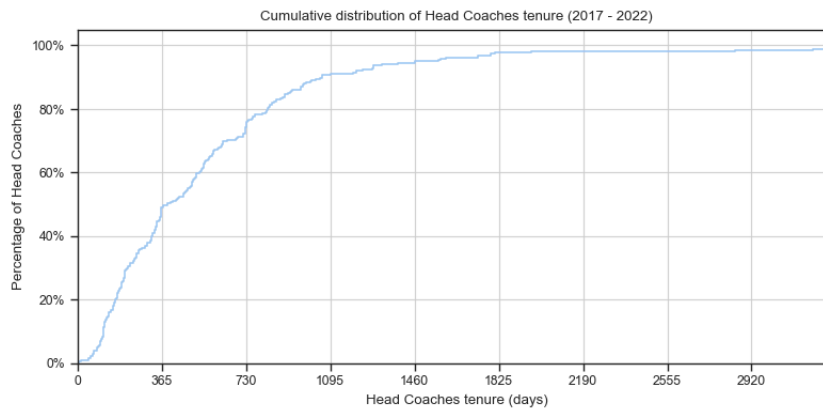


Figure 14: Empirical Cumulative Distribution Function of Head Coaches Tenure For Completed Appointments

Les [Figure 15](#) à l'ancienneté des coachs sportif et au renouvellement des coachs sportifs par rapport aux ligues d'intérêt.

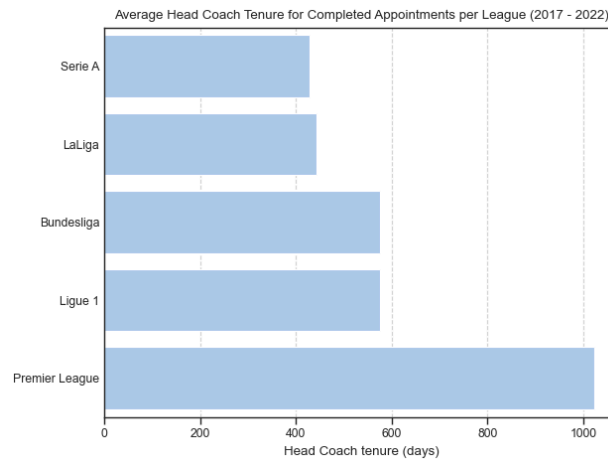


Figure 15: Average Head Coach Tenure for Completed Appointments per League

Les [Figure 16](#) et [Figure 17](#) et [Figure 18](#) 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.

- Les valeurs des coefficients pour les victoires sont : $r=0.36$ et $p=0.00$. Cela indique une corrélation positive modérée et statistiquement significative (avec $p < 0.05$) entre les deux variables étudiées.
- Pour les matchs nuls, $r=-0.12$ et $p=0.05$. Ainsi, il existe une corrélation linéaire faible et négative entre les deux variables, ce qui est statistiquement significatif selon le seuil de significativité choisi (ici, $p=0.05$).
- Quant aux défaites, $r=-0.35$ et $p=0.00$, cela indique une corrélation modérée négative et statistiquement significative entre les deux variables étudiées.

Il est à noter que les trois corrélations sont statistiquement significatives, étant donné que les valeurs de p sont inférieures (ou égales) 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, ne jouent pas en faveur du coach, étant considérés comme un résultat moyen voire mauvais (r négatif). Par conséquent, plus il y aura de matchs nuls, plus la durée du coach en tant que leader de l'équipe diminuera.

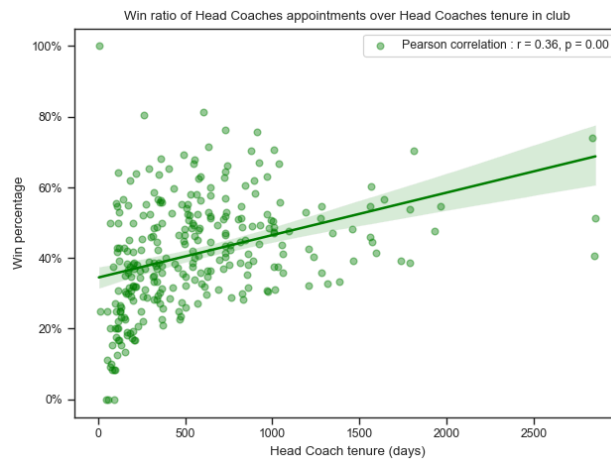


Figure 16: Win Ratio of Head Coaches Appointments versus Head Coach Tenure

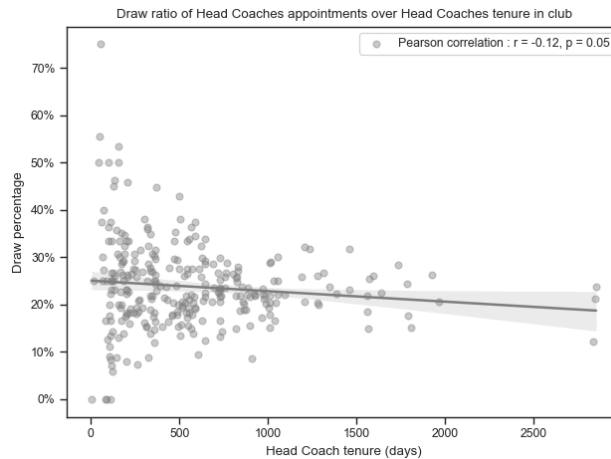


Figure 17: Draw Ratio of Head Coaches Appointments versus Head Coach Tenure

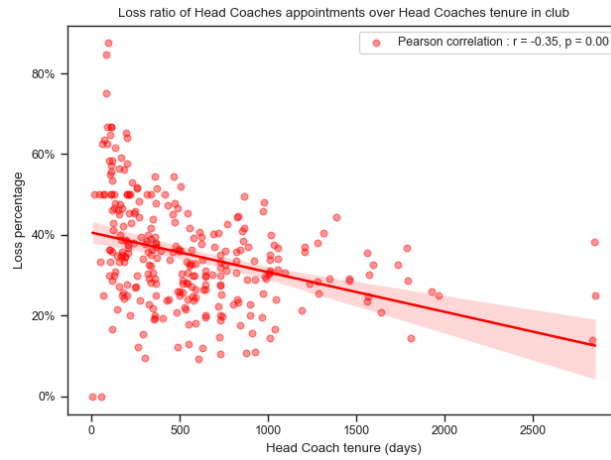


Figure 18: Loss Ratio of Head Coaches Appointments versus Head Coach Tenure

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

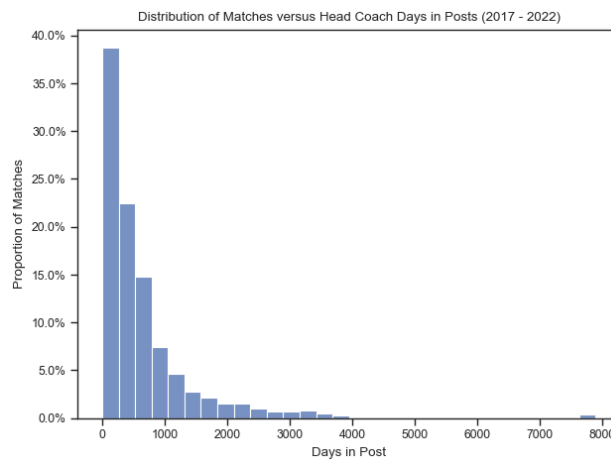


Figure 19: Distribution of Matches versus Head Coach Tenure on Match Day

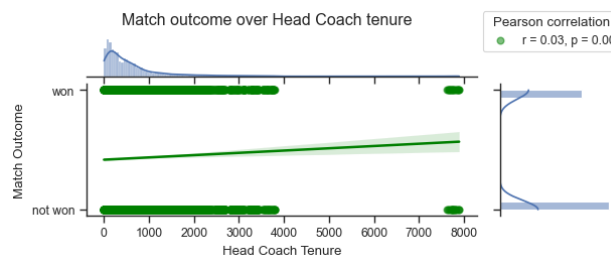


Figure 20: Match Win Outcome versus Head Coach Tenure on Match Day

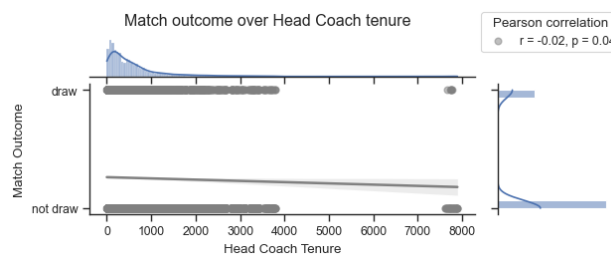


Figure 21: Match Draw Outcome versus Head Coach Tenure on Match Day

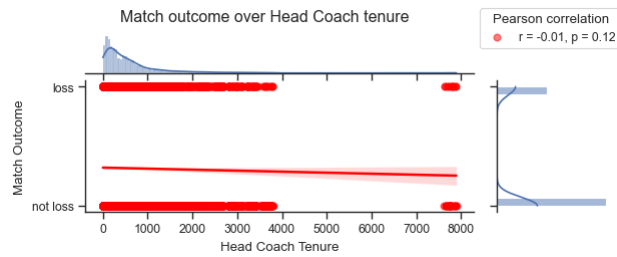


Figure 22: Match Loss Outcome versus Head Coach Tenure on Match Day

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

L'ancienneté, tout coachs confondus a un effet positif sur la performance de l'équipe. Cela peut s'expliquer par le fait que les coachs ont besoin de temps pour s'adapter à leur nouvel environnement et pour mettre en place leur stratégie de jeu. De plus, les coachs 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 coachs plus longtemps, ce qui peut expliquer en partie la corrélation positive entre l'ancienneté du coach et la performance de l'équipe.

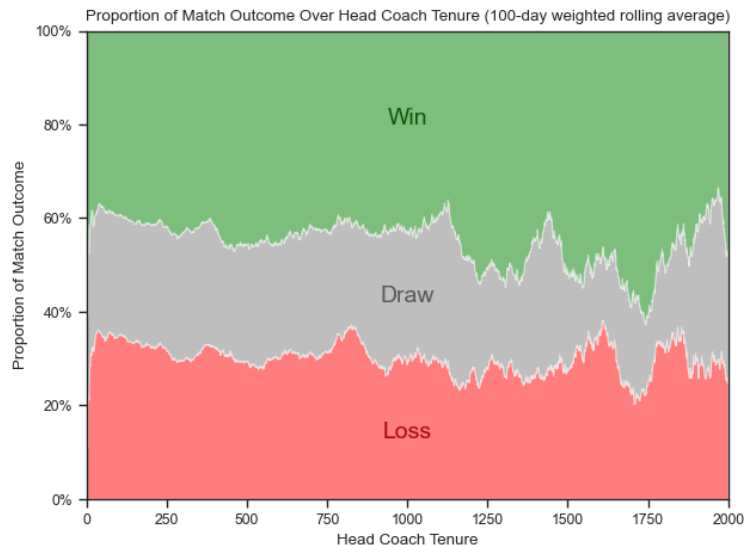


Figure 23: Weighted Rolling Average of Match Outcome versus Head Coach Tenure on Match Day

Explique que graph utilise les moyenne mobile pondérés sur une fenêtre de 30 jours :

```
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 l'effet du licenciement d'un coach sur la performance de l'équipe. 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. Ces résultats suggèrent que la stabilité et la continuité sont des facteurs importants pour la réussite d'une équipe de football. Dans la seconde série de graphiques nous avons montrée que les clubs qui renouvelle régulièrement leur coach ont tendance à voir une dégradation de leurs performances. Dans la troisième série de graphiques nous montrons que les coaches 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 coaches 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 [dépôt 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>
- Rocaboy, Y., & Pavlik, M. (2020). Performance Expectations of Professional Sport Teams and In-Season Head Coach Dismissals—Evidence from the English and French Men's Football First Divisions. *Economies*, 8(4), 82–83. <https://doi.org/10.3390/economies8040082>
- Wilke, C. O. (2019). *Fundamentals of Data Visualization*. O'Reilly Media. <https://clauswilke.com/dataviz/>