NOVA IMS

Information Management School

MAA

Mestrado em Métodos Analíticos Avançados

Master Program in Advanced Analytics

Does sacking a coach really help? Evidence from a Difference-in-Differences approach

Gustavo de Souza Machado Fabrício

Dissertation presented as partial requirement for obtaining the master's degree in Advanced Analytics

NOVA Information Management School

Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

Does sacking a coach really help? Evidence from a Difference-in-Differences approach

By

Gustavo de Souza Machado Fabrício

This project is presented as partial requirement for obtaining the master's degree in Data Science and Advanced Analytics

Supervisor: Bruno Damásio

November 2021

Acknowledgements

I would like to express my gratitude to my supervisor, professor Bruno Damásio who gave me all the support and attention during the research and analysis, providing materials, his time to discuss about the methods, his patience, his vision that helped me to make this thesis a reality.

I am extremely grateful to my parents Mr. Jose Carlos Fabrício and Mrs. Solange Fabrício for their trust in me, for the time and effort invested in my education, without them none of this would be possible. Also, a special thank you to my life partner Danielle Newton for her patience and support during my master's.

Finally, my thanks to my master's friends that made this journey more interesting than it already was, with special mention to Victor Gonzalez my partner in most of the projects during the master's. Thanks for your support, and partnership.

To all of you, without whom I would not be who I am and what I have achieved, Thank You.

Abstract

This project looks to evaluate if football clubs should or should not change their coach in order to improve their performance in the national league. For this analysis I selected, three of the most important European football leagues, La Liga (Spain), Serie A (Italy) and Premier League (England).

The data used in this project was taken from the transfermarkt website, a large football platform. The data period is from season 2005-06 to season 2019-20 and has information about individual games results and squad value by player.

The steps before the analysis were a data cleaning and consolidation of the information, creation of new features as a performance measure and selection of cases of interest for this analysis based on club and coach profile. Numeric variables were standardized to be on the same scale and make different seasons comparable. A K-means was applied to identify clubs according to their investments which has a proportional correlation with performance.

Finally, a difference in differences analysis was applied to evaluate if a club would obtain a performance gain if they decided to sack their coach between game twelve and twenty-six of the season after a poor performance in consideration to squad price.

As a general conclusion, it is possible to consider that on average the clubs in the treatment group and comparison group recover their performance after a period of underperforming, but the recovery of the clubs that sack their coach is lower compared with the clubs that keep them.

Keywords

Difference in Differences; Statistics; Football; Data Scraping; Clustering

Index

List of Figures

List of Tables

List of abbreviations and acronyms

DiD	Difference in Differences
Performance	Total points won divided by total possible points
Comparison Group	Group of observations without intervention / treatment
Treatment Group	Group of observations that suffered intervention
Warning Performance	A performance percentage that clubs usually starts to sack their coach.

1. Introduction

In recent decades, in collective sports, the importance of the coach has been highlighted. The coach is the person who oversees training and team formation, preparing it for a good performance to obtain positive results. The coach is an expert in technical and tactical direction and in the player's psychological and physical development. In football this is no different, part of a successful season for a football club involves hiring the ideal coach to manage the team, forming the best squad, and frequently indicating players to hire.

In football history we have a few examples of coaches that stayed for decades at the same club, such as Guy Roux, who stayed for forty-four years (1961 - 2005) in charge of Auxerre-FRA, leading the club from the third division to the title of French League in 1995. In England, the most successful manager in the football history, Alex Ferguson, led Manchester United for twenty-seven years (1986 - 2013) winning a total of forty-nine titles.

But the history has more cases of clubs sacking their coaches during the season than examples of longevity. For example, in the Italian first division during the last five seasons (2015-16 – 2019-20) an average of twelve clubs per season kept their coaches for the duration of the season and in the same period in eleven occasions clubs there were three or more coaches during the same season.

When a club decides to sack its coach before the end of their contract it causes some inconveniences. One of them is a termination fine, that depends on the contract between both parties but when we talk about the richest leagues in the world this usually involves a large amount of money. As an example, when José Mourinho was sacked from Tottenham in April 2021, he had a contract until the end of the season 2022-23, and as a consequence of this breach of contract Tottenham had to pay a severance package of around fifteen million pounds to Mourinho according to the website football.london [13].

This project will analyse data from the top three leagues in Europe, La Liga – Spain, Serie A – Italy and Premier League - England between seasons 2005-06 and 2019-20. All leagues have almost the same structure. The entire season runs from August through to May and the league consists of twenty clubs playing against each other both at home and away. One season has three hundred and eighty matches, which results in thirty-eight matches for each team. At the end of the league the club with the highest number of points is the champion, and the bottom three are relegated to a second division.

The first part of this project consists of scrap data from a transfer market website about game results and player's information's of three of the biggest leagues in Europe mentioned before and uses this data to build analysis to extract insights about how clubs deal with low performance according to club expectation.

The second part consists of defining clusters of clubs based on their players values, calculating performance metrics, and filtering the cases of interest to this project and finally applying *Difference in Differences* to analyse the performance comparing the treated group and the comparison group.

1.1 Thesis Objective

The main objective of this thesis is to evaluate performance comparing clubs that sacked their coach (treatment group) and the clubs that DiD not (comparison group) and identify if changing the coach during the season is an efficient way to leverage performance by applying statistical technics as K-Means to perform a club segmentation based on squad price value and applying Difference in Differences to evaluate efficiency of treatment over time.

Main questions that this thesis will try to answer are, "Does sacking a coach in the middle of the season increase the club performance during the season?", "Is sacking a coach in the middle of the season more efficient than keeping them until the end of the season?"

2. Literature Review

Management change in sports and the impact on performance has been a theme of many articles in the recent years. In organisations, the performance of a manager/head coach in elite sport is critical to the success of the team and business (D. Bloyce et al. 2008), and this fact has been increasing the pressure on the coach over the years. The effective and efficient implementation of change is often required for both successful performance and management survival across a host of contemporary domains

(Diefenbach T, Klarner P, 2008). However, it is important to emphasize that these changes must be made with planning and through analysing several factors that involve the performance of the team. Considering that the coach contributes to a large percentage in relation to the success of a team on the field, this project seeks information in articles related to management change not only in sports but also in large companies and public sectors trying to identify properties to make a connection with football.

The analysis of efficiency in change managers during the season already have been theme of many articles, one study in particular analysed the effectiveness of in-season manager changes in English Premier League during seasons 2000-2001 to 2014-2015 and their conclusion was that some management changes are successful, and others are counterproductive but on average performance does not improve following coach replacement. The analysis was made comparing a treated group and a comparison group and comparing the accumulative surprise. This is the sum of the differences between the actual number of points and the expected number of points based on bookmaker odds (Lucas M. Besters, Jan C. van Ours and Martin A. van Tuijl, 2016).

Another study to analyse the impact of a mid-season change of manager on the sporting performance was made using data from the Spanish national league, using data envelopment analysis (DEA) to compare the performance of teams that changed coach during mid-season with teams that stayed with the same coach. The main result of this study was that changing coaches during mid-season improves sports performance but does not allow for as good a performance as teams that DiD not change coach mid-season (González-Gómez, Picazo-Tadeo & García-Rubio, 2011).

In the article related to change a coach on Basketball, an analysis was made to understand if a team changes his coach, will the next game be a win. Using logit models, the conclusion of this study was that the new coach has double the chance of winning his first game compared with the previous coach. Other factors that should also be considered are if the team is playing at home and if it is playing with a low-quality team, (Martínez, J.A. 2012).

Difference in differences is a very popular technique to evaluate performance in different periods, and this technique has many different variations. In 1994, David Card and Alan Kruegeur used DiD to evidence the effect of minimum wages on establishment level employment outcomes by comparing a minimum wage increase in New Jersey with the minimum wage in Pennsylvania analysing both cities before and after the minimum wage rise. In this paper they DiD not find evidence to confirm that the rise of minimum wage in New Jersey reduced the employment in fast food restaurants (Card & Krueger, 1994).

A more recent papers addressed the topic of DiD with multiple time periods, showing that a family of causal effect parameters are identified in staggered DiD setups, even if differences in observed characteristics create non-parallel outcome dynamics between groups, also suggesting different aggregation schemes to highlight treatment effects (Callaway, B. & Sant'Anna, 2020), (Goodman-Bacon, A. ,2021)

3. Theoretical background

3.1K-Means

K-Means is a prototype-based simple partitional clustering algorithm that attempts to find K non-overlapping clusters. These clusters are represented by their centroids (Junjie Wu, 2012) where the centroids are the mean of the data points in each cluster.
Before starting the clustering process, the number of clusters (K) needs to be defined by the user. The K-means process starts with the selection of K initial centroids, after that every data point is assigned to the closest centroid, Euclidian distances are the most common metric to calculate those distances. Every collection of data assigned to a centroid forms a cluster. The next step is, update the centroid based in the data points assigned to the cluster, after that once again the data points is assigned to the closest centroid. The process is repeated until no point changes clusters.
K-means is very simple and robust, highly efficient, and can be used for a wide variety of data types. Moreover, K-means as an optimization problem still has some theoretical problems. The emerging data with complicated properties, such as large-scale, high-dimensionality, and class imbalance, also require adapting the classic K-means to different challenging scenarios, which in turn rejuvenates K-means. (Junjie Wu, 2012).

3.2 Backward Stepwise

Backward stepwise selection is an efficient way to select the best subset of variables to run a linear model. This method consists in run the first model with all variables, and then removes the least useful predictor, one by one.

The backward selection approach searches though only $\frac{1+p(p+1)}{2}$ models, where p is the number of variables. Backward is not guaranteed to yield the best model containing a subset of the p predictors (Desboulets Loann, 2018).

Algorithm – Backward stepwise selection

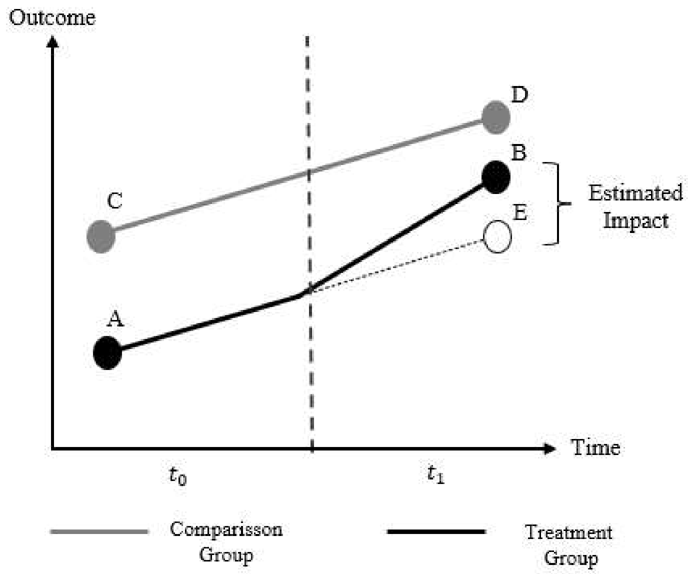
1. Let M_p denote the full model, which contains all p predictors.
2. For k = p, p-1, …,1:
   (a) Consider all k models that contains all but one of the predictors in M_k for a total of k-1 predictors.
   (b) Choosing the best among these k models and call it M_{k-a}. Here best is defined as having smallest RSS or highest R^2.
3. Select a single best model form among M_0,…, M_p using cross-validated prediction error, Cp (AIC), BIC, or adjusted R^2.

3.3 Differences in Differences

The difference-in-differences method compares the changes in outcomes over time between a population that had an intervention, this population can be named as *treatment group* and a population that had not named *comparison group*. DiD combines the two counterfeit estimates of the counterfactual, first is time, before and after, second are comparisons between treatment group and comparison group to produce a better estimate of the counterfactual. Treatment and comparison groups do not necessarily need to have the same conditions before the intervention but for DiD to be valid, the comparison group must accurately represent the change in outcomes that would have been experienced by the treatment group in the absence of treatment. (Gertler, Martinez, Premand, Rawlings & Vermeersch, 2016).

The figure bellow helps to illustrate de DiD method.

Figure 1 - Difference in diferences estimation



A represents the outcome for *treatment group* in t_0, before the intervention started, B represents the outcome for *treatment group* in t_1, after the intervention, C represents the outcome for *comparison group* before the intervention started and D after the intervention.

The two counterfeit estimates of the counterfactual are represented by:

- The difference of outcomes before and after the intervention for the treatment group (B – A)
- The difference of outcomes between treatment group and comparison group after the intervention (B – D)

The estimated impact is calculated by (B – A) – (D – C) that means assume if the treatment group did not have intervention, it will have the same tread as the comparison group this value is represented by E in figure 1. If trend is different between groups, the estimated impact obtained would be biased.

3.3.1 Notation and modelling

Let *Y* be the outcome of a population with two groups indexed by treatment status *T* = 0,1 where 0 is related to the group without treatment, i.e. comparison group and 1 the group that receive treatment, i.e. treatment group. Assuming that those groups are observed in two time periods, t = 0,1 where 0 is the period before the treatment group received treatment and 1 the time period after. Every observation is indexed by the letter i = 1, …, N; where individuals will have two observations each, one in t_0 and one in t_1.

The outcome Y_i is modelled by the equation:

$$Y_i = \alpha + \beta T_i + \gamma t_i + \delta (T_i \cdot t_i) + \varepsilon_i$$

Where α is the constant, β measure the average differences between treatment and control group, γ is the time trend common to control and treatment group and δ is the true effect of treatment DiD (Fredriksson & de Oliveira 2019).

3.3.2 Unbiased estimator assumptions

For δ be an unbiased estimator the parallel trends assumption is needed, that means if the treated group did not suffer an intervention, it would have the same trend as the control group. If treatment group also suffer impact of a different factor at the same time of the intervention considered in the analysis, DiD will not be able to separate out the different effects.

Thus, when DiD is applyed we need to assume that if there was not treatment the outcome on the treatment group would have moved in tandem with the outcome in the comparison group once there is no way to prove that assumption.

DiD has as key concept compare trends between treatment and comparison group instead of their outcomes after the intervention. DiD also cancel the effect of all the characteristics that are unique to that observation and do not change over time (Gertler, Martinez, Premand, Rawlings & Vermeersch, 2016).

In the last feel years some different approaches for DiD has been developed, (Callaway, B. & Sant'Anna, 2020) propose a new approach separating DiD in tree steps, identification of policy-relevant disaggregated causal parameters, aggregation of these parameters to form summary measures of the causal effects and estimation and inference about these different target parameters. This approach allows for estimation and inference on interpretable causal parameters allowing for arbitrary treatment effect heterogeneity and dynamic effects, thereby completely avoiding the issues of interpreting results of standard two-way fixed effects regressions as causal effects in DiD setups.

4. Methodology

This chapter focuses on explaining in depth the whole analysis process to answer the proposed questions in this thesis, starting from data scraping, data preparation, club's segmentation, filter cases of interest, descriptive analysis and DiD model application.

4.1 Data Scraping

The data used in this analysis was scraped from the web site transfermarkt.co.uk/ Transfermarkt offers the world's largest football database with all information on players, clubs, and competitions as well as one of the largest football communities and playing areas for anyone who wants to share and exchange ideas about football. An overview of player market values in addition to news, statistics, consultant information and fan insights. The platform was founded in May 2000 by Matthias Seidel.

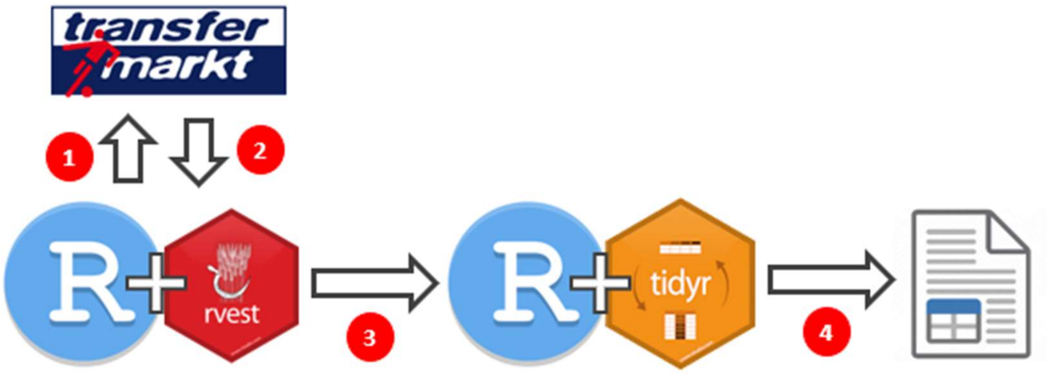
The scraping process was made with **R** using the library *rvest*, who efficiently scraped the data from seasons 2005-06 to 2019-2020 of the English, Spanish and Italian first division leagues. For each league there was a two-scrap process, one for the game results and one for player's information's.

The game results data consists of information such as date of the match, time, home team and away team, their position on the league table, coach and formation and attendance and result. Each line represents one game, an example of this table is available in the appendix at table 10. This information is fundamental to identify when a club sacked a coach during the season and calculate the performance.

The player's information consists of the players name, club, age, position, market value, best foot, and height. Each line represents one player during a specific season. This information will be helpful to identify club performance expectations based on the squad value and to create independent variables to explain performance. An example of this table is available in the appendix at table 11.

The diagram bellow represents the process of scrapping the data from transfermarkt using R, data cleaning and transformation to build the data tables for each league in this thesis.

Figure 2 - Scrapping and data preparation



Step 1: Using the library *rvest*, access the transfermarkt web site in a specific URL to access the information of interest from two main URL structures, one for matches information that was composed by four variables that changes according with club, season, and league. The second was to extract the players information and was composed by tree variables as season and two related to the club.

Step 2: Store that information into lists, where each list represents one specific information e.g. the score, name of the club playing home, name of the club playing away.

Step 3: Using the library *tidyr* clean variable out of the format and transforming lists into data tables, as mentioned before, one for matches and other for players.

Step 4: Export those data tables to txt file that going to be used on Python for the analysis

4.2 Data preparation

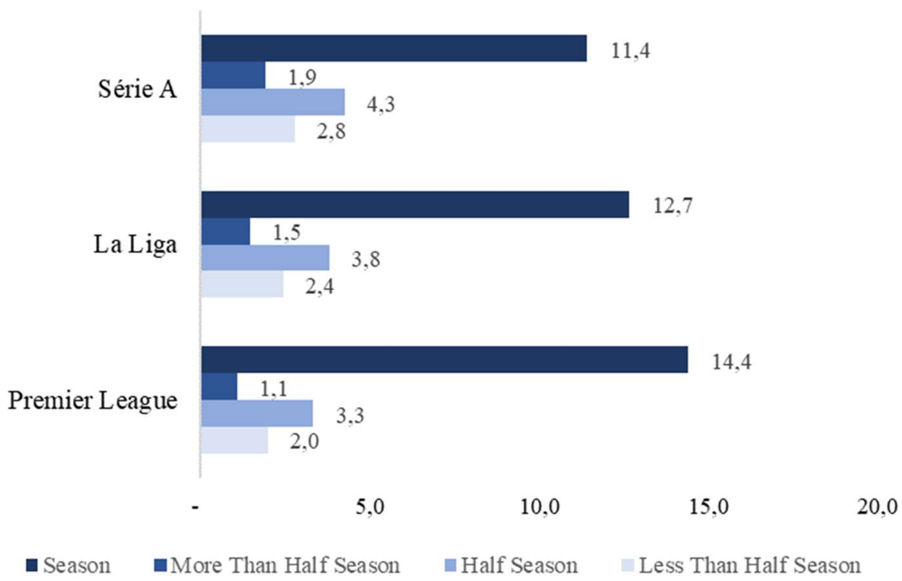
This step is essential to the project, has all consolidations, validations and criteria definitions were made to be used in the analysis. The idea is to aggregate the information by Season - Club – Manager to calculate performance, evaluate the cases of interest and define the three variables to apply DiD regression, Treated, Time and Treated x Time.

As described before, in the table each line represents a game of a specific club, that aggregation was made to create a table with club, season, coach, number of games, date of first game and last game of the coach. As an example, the table 12 in the appendix, shows that Alaves had three coaches during the season 2005-2006. That means each line represents a work period of a coach in a club during a specific season.

From the number of games in the aggregate table, it is identified which coach was sacked or not during the season and for how long they managed the club. The criteria are, if the coach had a work period of thirty-four games or more, it is considered that he stayed for the whole season, between twenty-seven and thirty-three games, the coach stayed for more than half of the season, between twelve and twenty-six games he stayed for half of the season and less than twelve games he stayed for less than half of the season.

According to this information, each coach's work period will be divided into two different groups, the ones whose work period lasted at least thirty-four games, these observations will represent the comparison group, which means observations without treatment during the season, and the work period that was interrupted during the season or started in some point after the start of the season, will be representing the treated group, observations that suffered an intervention during the season.

Figure 3- Average number of clubs by First coach quantity of games.



Note: Figure 3 represents the average number of clubs over the fifteen seasons that kept their coach by four different periods of time and league.

As we can see in figure 2, during a season in Serie A, only eleven clubs on average will maintain their coach until the end of the season, and seven clubs on average will sack their couch halfway through the season or before. Between the three leagues, Italy has the highest number of changes.

Evaluating the group *with treatment*, the ideal homogeneous behaviour for all clubs with treatment would be, club A during season one had coach X for the first nineteen games and then coach Y for the last nineteen games, where coach X represents t_0 and coach Y represents t_1. But in general, this does not happen, the data base has several cases of clubs with more than two coaches during the same season by consequence coaches with less than ten games in t_0 which will be considered a low number of games to evaluate performance. To make a fair evaluation, this thesis will only consider cases where the coach that started the season (t_0) had at least between twelve and twenty-six games, for the cases that the club had two or more coaches after the first coach, it will be considered one single work period, in other words, it will evaluate the performance of the coach that started the season and the performance of the club after the coach was sacked.

The comparison group also need to be split in t_0 and t_1, in order to compare performance in two different moments, in this thesis for the coaches who stayed the entire season this break will be the game with lowest cumulative performance between fifth and twenty-third game. With these criteria the season will be divided in the most critical moment that is a similar condition to what happens with the treated group.

The coach performance is calculated dividing the total points won by total possible points, that means if the coach won twenty-one points in fourteen games, his total possible points will be fourteen times three equals to forty-two, so his performance is twenty-one divided by forty-two equals to fifty percent. This metric will be the outcome of the model.

The other data set that will be used in this project is related to squad, each line has information about one player in a specific season and club like age, estimated value and position. This data set will be consolidated by season and club, with information of total, average and highest value by player's position, those new features will be used as independent variables to explain performance and create cluster of clubs/ season.

Analysing the evolution of the estimated market value of the player season over season we notice that this value has been rising over time, comparing the average estimated value of a player in season 2005-2006 against season 2019-2020 we see Serie A with a jump from 2.36M to 6.44M, with more than 170% growth, La Liga from 3.02M to

7.72M with 153% growth and Premiere League from 3.7M to 10.73M with 189% growth. In this scenario the price variables by position would not be effective to predict performance once performance is a fixed range between zero and one and the price have been growing season over season, the correlation between price and performance would depend on the year.

A solution for this issue was, once the data was consolidated, standardise the values for each season individually, in this way it is possible to have the same scale over the seasons and express the difference between clubs inside the season. The standardise technique was the standard score or Z - Score scale, this technique is better explained in the Theoretical Background. With this transformation the correlation between performance and total value of the squad standardized is 0.69 for Serie A, 0.74 for Premier League and 0.78 for La Liga, as expected a more valuable squad tends to have better results.

4.3 Clubs' segmentation

As was commented before, a more valuable squad tends to have better results, and consequently it implies greater pressure for coaches to have better results. Nobody expects that SPAL with an estimated squad value of ninety-seven million and four hundred thousand euros would have the same performance of Juventus with a squad value nine times higher in season 2018-19. In other words, if Massimiliano Allegri (Juventus coach) had the same performance of Leonardo Semplici (SPAL coach) of 37% probably he would be sacked during the season, whereas SPAL kept Semplici until the end of the season.

This segmentation is for the purpose of dividing clubs into clusters according to their squad value and after, identifying the critical percentage of performance of the clubs-season by cluster. The four most correlated variables with performance will be used to perform a K-means segmentation. The K-means was applied for three hundred observations (twenty clubs per season, times fifteen seasons.) The Elbow graph was used to determine the number of clusters, minimizing the variance inside the groups and maximizing the variance between group in order to have the most efficient number of cluster.

Table 1- Variables for K-means segmentation.

League	K-Means Variables - Standardized			
Serie A	Totat Value	Mean Value	Mean Value Defence	Total Value Attack
LA Liga	Mean Value	Totat Value	Mean Value Attack	Max Value Defence
Premier League	Totat Value	Mean Value	Total Value Defence	Mean Value Attack

The segmentation generated the following results:

Table 2 - K-means segmentation for Série A

Cluster	Totat Value	Mean Value	Total Value Defence	Mean Value Attack	N
0	3,05	2,85	2,63	2,83	43
1	0,38	0,39	0,36	0,42	156
2	4,40	4,42	4,40	4,26	35
3	1,42	1,30	1,22	1,37	66

As showed in the previous table, in Série A Cluster 1 has clubs with low investments e.g. Brescia, Pescara and Treviso. Cluster 3 has clubs with medium investments e.g. Udinese, Fiorentina, and Torino. Clusters 0 and 2 Clubs with high investments e.g Juventus, AC Milan, and Inter. Those two clusters have almost the same clubs in different seasons, for that reason Cluster 2 will be incorporated to Cluster 0. That happens because in some seasons those clubs had a large difference of investments compare to the others.

Table 3 - K-means segmentation for La Liga.

Cluster	Totat Value	Mean Value	Max Value Defence	Mean Value Attack	N
0	0,29	0,32	0,43	0,26	202
1	4,71	4,65	4,63	4,45	32
2	1,56	1,58	1,80	1,43	66

Cluster 0 has clubs with low investment e.g. Las Palmas, Levante, Maiorca and Numância. Cluster 2 has clubs with medium investment e.g. Sevilha, Valencia and Villarreal. Cluster 1 has clubs with high investment and is composed just with Barcelona, Real Madrid, and Atletico Madrid in some seasons.

Table 4 - K-means segmentation for Premier League

Cluster	Totat Value	Mean Value	Total Value Defence	Mean Value Attack	N
0	1,24	1,33	1,17	1,17	64
1	4,38	4,24	4,43	4,25	44
2	0,41	0,42	0,46	0,41	152
3	3,03	2,69	2,71	2,73	40

The Premier League is the league with more intersection of clubs between clusters, as an example Manchester City appears in all clusters, the reason is the growth of financial potential with investments over the years. Cluster 2, clubs with low investments e.g. Birmingham, Sheffield United and West Brom. Cluster 0, clubs with medium investments e.g. West Ham, Southampton and Wolverhampton; and Cluster 1 and 3 with clubs with high investments e.g. Manchester United, Liverpool and Chelsea. Those two clusters were grouped together due to the similarity of the clubs.

The next step is to merge the cluster information with the treated dataset, in doing that, it is possible to evaluate the warning performance percentage for each cluster by analysing the performance of the observations in t_0. Once the warning performance for each cluster is known these values will be used to select the observations to be used in the DiD regression. For those coaches whose performance is equal to or below the warning performance in t_0 means that they were with their position in risk, so it is reasonable to compare them with those who, in the same level of investments were sacked.

Table 5 - Performance of treated group in t_0 by League and Cluster.

League	Cluster	Performance				Qtt
		Max	P 90	Mean	Min	
Serie A	High	60%	57%	50%	39%	9
	Medium	46%	44%	38%	21%	16
	Low	37%	33%	25%	13%	44
La Liga	High	70%	70%	65%	60%	4
	Medium	57%	51%	42%	26%	11
	Low	38%	33%	28%	14%	38
Premier League	High	67%	66%	50%	31%	8
	Medium	56%	43%	33%	21%	10
	Low	41%	35%	28%	11%	32

As was expected, the clubs with high investments, had a higher standard for performance. On table 5 it is possible to notice that La Liga has a higher average performance for the treated observations when compared to the other leagues, which explains the difference of investments between the clubs. Just out of curiosity, between the seasons 2005-06 and 2019-20 only five different clubs finished the league in the top three, and the league had only three different champions.

4.4 Filtering cases of interest and data analysis

Once the treated club/season cases are already prepared and the performance by segment is known, the next step is filtering the non-treated observations in t_0 with a similar performance by cluster in the treated group in t_0 in order to have a fair comparison. For this filter the percentile 90% will be used. That means if a non-treated club A in season X of the Premier League that belongs to cluster medium investment with a performance equal or lower to 43% in t_0, this observation will be selected to the analysis.

In the same example above, it would not be fair to compare a non-treated observation that have 70% performance in t_0, because in normal conditions this performance level is outside of the performance range that clubs use to sack their coach.

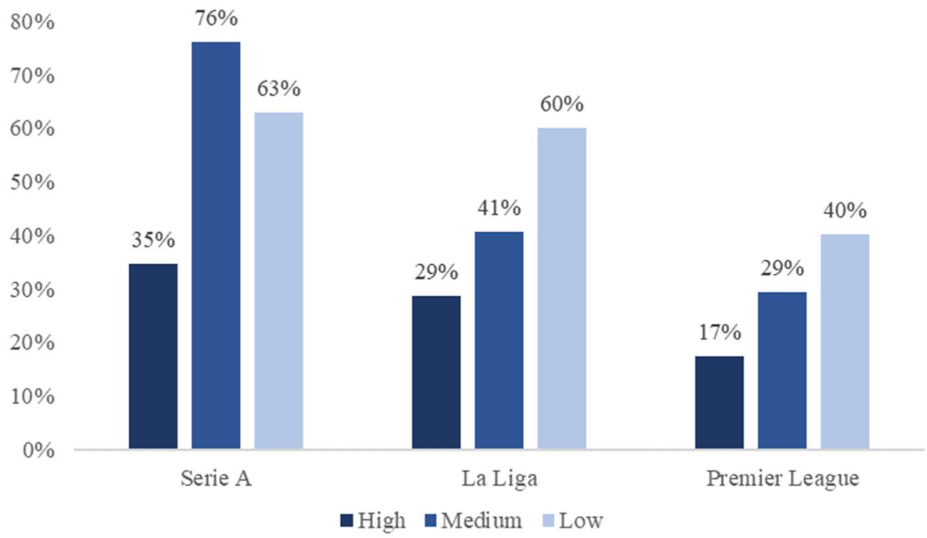
The final database for the analysis is composed by non-treated cases, comparison group, that performance in t_0 by cluster is lower than percentile 90% for the same cluster in the treated cases plus the treated cases, treatment group, that the manager at t_0 had between twelve and twenty-six games.

Table 6 - Quantity of observations by league, cluster and treated

Cluster	Treated	Qtt Observations in t_0		
		Serie A	La Liga	Premier League
High	Yes	9	4	8
High	No	17	10	38
Medium	Yes	16	11	10
Medium	No	5	16	24
Low	Yes	44	42	33
Low	No	26	28	49
Total		**117**	**111**	**162**

The table above shows that in the *cluster high* the number of observations in non-treated is higher than treated, indicating that clubs with high investments and consequently high performance, tend to keep their coach when the performance hits the warning percentage. On the other side, clubs with low investments are more likely to change their coach. A hypothesis for that behaviour can be related to the relegation, clubs with low investments and performance bellow the expectations tends to fight against relegation and sack a coach expecting a better performance is one of the most common effects. The plot below represents the percentage of sacked coaches with performance below the warning percentage of the cluster.

Figure 4 - Percentage of sacked coaches in the final data set by league and cluster.



Note: Figure 4 represents the percentage of sacked coaches in the final dataset by League and cluster.

Comparing leagues Serie A are more likely to sack their coach in relation to the other leagues in every cluster. Premier League has the lowest percentage with 17% for the clubs in the cluster High value. Also, can be notice that the clubs with lower investments are more likely to sack their coaches. A possible explanation is clubs with higher investments tends to hit their performance goals easier in comparison with clubs with lower investments.

The Boxplot below shows the distribution of performance by treated/non-treated case and time divided by cluster.

Figure 5 - Boxplot – Performance distribution for Série A by time and Cluster

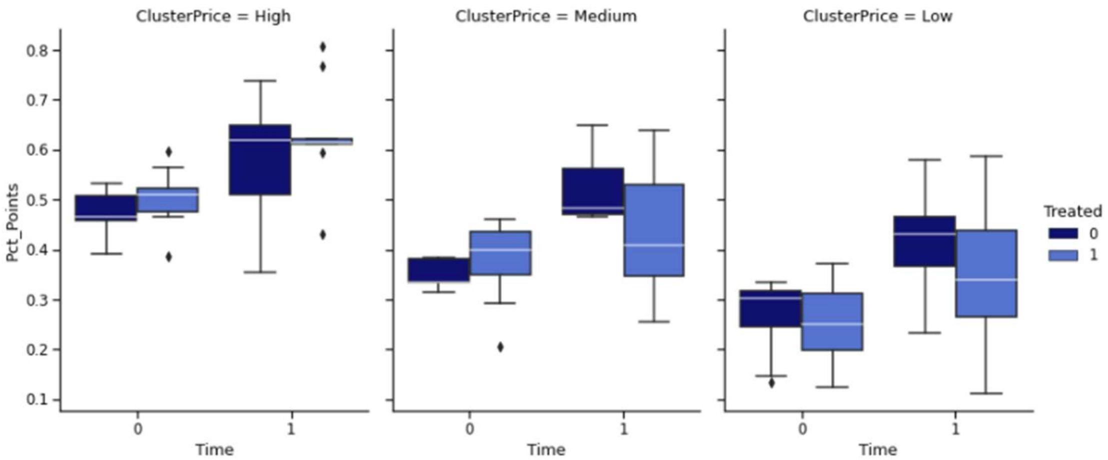


Figure 6 - Boxplot – Performance distribution for La Liga by time and Cluster

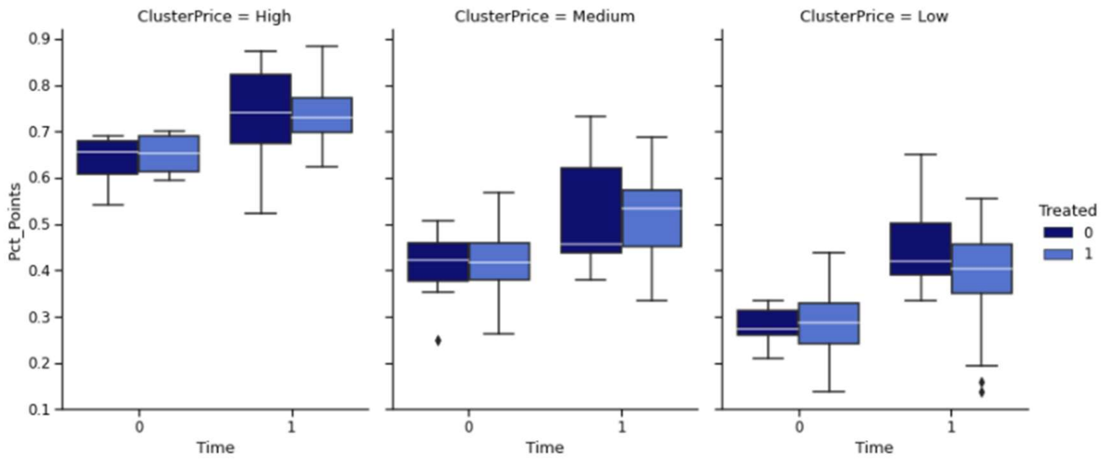
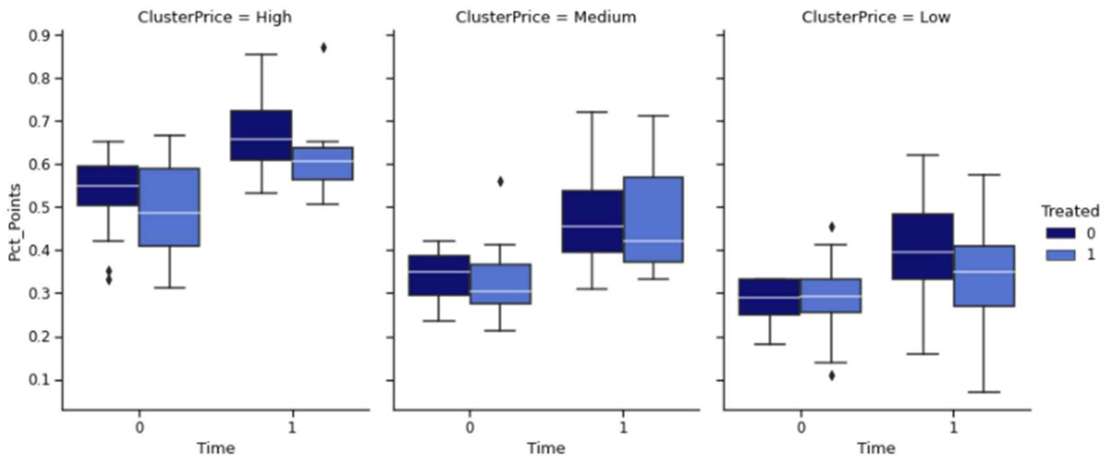


Figure 7 - Boxplot – Performance distribution for Premier League by time and Cluster.



Note: Figures 5, 6 and 7 represents the distribution of performance of club/season by treated and comparison group in t_0 and t_1 by cluster.

By the distribution above, it is easy to notice that time has a big influence for both cases treated and non-treated. When a club has a bad start to the season according to their expectations and capabilities, they tend to recover after a certain point. For every league and cluster, the median is higher in t_1, comparing with t_0. DiD will help to understand if there is performance gain by changing the coach.

4.5 Difference in differences model

After preparing the data in the previous steps a difference in differences model was applied for each league and the feature selection process was Backward Stepwise selection with alpha equal to 0.1. The variables Treated, Time and DiD will always be in the model in order to analyse the influence on the performance.

The final dataset contains dummy variables for cluster, Cluster_Low, Cluster_Medium and Cluster_High, mean age by position (goalkeeper, defence, side defence, midfield and attack), quantity players by position, maximum value standardized by position, mean value standardized by position, total value standardized by position, quantity of players in the squad, team average age, team total value standardized, team average value standardized, higher player value standardized and also Time, Treated, DiD.

5. Result Presentation

On the next pages the DiD models results for Serie A, La Liga and Premiere League will be shown with a description of the coefficients.

5.1 Serie A

The best model according to backward stepwise technique for Serie A is composed by five variables plus Time, Treated and DiD, (Cluster_Medium, Cluster_High, Tot_Games, max_value_SDEF_std, tot_value_std, Time, Treated, DiD).

Table 7 shows in detail the coefficient for every variable and the significance for the DiD model applied for the Serie A data set with two hundred thirty-four observations. The comparison group has ninety-six observations represented by forty-eight clubs/season and the treated group has one hundred thirty-eight observations represented by sixty-nine clubs/season.

Table 7- DiD Regression Results for Serie A.

Model:	OLS	**R-squared:**	0,65
Method:	Least Squares	**Adj, R-squared:**	0,64
No, Observations:	234	**F-statistic:**	54,38
Df Residuals:	224	**Prob (F-statistic):**	0,00
Df Model:	8	**Log-Likelihood:**	254,73
		AIC:	-491,50
		BIC:	-460,40

Var	coef	std err	z	P>|z|	[0,025	0,975]
const	0,184	0,03	6,076	0	0,124	0,243
Cluster_Medium	0,058	0,019	3,04	0,002	0,02	0,095
Cluster_High	0,075	0,041	1,828	0,068	-0,005	0,156
Tot_Games	0,004	0,002	2,313	0,021	0,001	0,007
max_value_SDEF_std	0,016	0,008	1,955	0,051	-4E-05	0,032
tot_value_std	0,038	0,013	2,881	0,004	0,012	0,065
Time	0,124	0,015	8,316	0	0,095	0,153
Treated	- 0,002	0,012	-0,139	**0,889**	-0,024	0,021
DiD	- 0,044	0,021	-2,076	0,038	-0,086	-0,002

Model Tests			
Omnibus:	0,76	Durbin-Watson:	1,651
Prob(Omnibus):	0,684	Jarque-Bera (JB):	0,477
Skew:	-0,057	Prob(JB):	0,788
Kurtosis:	3,189	Cond, No,	148

5.2 La Liga

The best model according to backward stepwise technique for La Liga is composed by four variables plus Time, Treated and DiD, (Cluster_Low, qtt_players_SDEF, max_value_DEF_std, max_value_MID_std, Time, Treated, DiD).

Table 8 shows in detail the coefficient for every variable and the significance for the DiD model applied for La Liga data set with two hundred twenty-two observations. The comparison group has one hundred eight observations represented by fifty-four clubs/season and the treated group has one hundred fourteen observations represented by fifty-seven clubs/season.

Table 8 - DiD Regression Results for La Liga

Model:	OLS	**R-squared:**	0,74
Method:	Least Squares	**Adj, R-squared:**	0,73
No, Observations:	222	**F-statistic:**	59,34
Df Residuals:	214	**Prob (F-statistic):**	0,00
Df Model:	7	**Log-Likelihood:**	251,93
		AIC:	-487,90
		BIC:	-460,60

Var	coef	std err	z	P>|z|	[0,025	0,975]
const	0,383	0,033	11,498	0	0,318	0,449
Cluster_Low	- 0,049	0,021	-2,345	0,019	-0,089	-0,008
qtt_players_SDEF	- 0,014	0,005	-2,948	0,003	-0,023	-0,005
max_value_DEF_std	0,035	0,01	3,522	0	0,015	0,054
max_value_MID_std	0,036	0,01	3,802	0	0,018	0,055
Time	0,137	0,016	8,509	0	0,106	0,169
Treated	0,011	0,012	0,945	0,345	-0,012	0,034
DID	- 0,039	0,022	-1,807	**0,071**	-0,082	0,003

Model Tests			
Omnibus:	6,959	Durbin-Watson:	1,786
Prob(Omnibus):	0,031	Jarque-Bera (JB):	7,423
Skew:	-0,309	Prob(JB):	0,0244
Kurtosis:	3,649	Cond, No,	36,6

5.3 Premier League

The best model according to backward stepwise technique for Premier League is composed by eight variables plus Time, Treated and DiD, (Cluster_Medium, Cluster_High, mean_age_MID, mean_age_SDEF, max_value_SDEF_std, mean_value_DEF_std, tot_value_GK_std, Time, Treated, DiD).

Table 9 shows in detail the coefficient for every variable and the significance for the DiD model applied for Premier League data set with three hundred twenty-four observations. The comparison group has two hundred twenty-two observations represented by one hundred eleven clubs/season and the treated group has one hundred two observations represented by fifty-one clubs/season.

Table 9 - DiD Regression Results for Premier League

Model:	OLS	**R-squared:**	0,72
Method:	Least Squares	**Adj, R-squared:**	0,71
No, Observations:	324	**F-statistic:**	97,39
Df Residuals:	313	**Prob (F-statistic):**	0,00
Df Model:	10	**Log-Likelihood:**	343,95
		AIC:	-665,90
		BIC:	-624,30

Var	coef	std err	z	P>|z|	[0,025	0,975]
const	0,233	0,087	2,682	0,007	0,063	0,404
Cluster_Medium	0,035	0,014	2,6	0,009	0,009	0,062
Cluster_High	0,147	0,027	5,371	0	0,093	0,201
mean_age_MID	- 0,007	0,003	-2,628	0,009	-0,012	-0,002
mean_age_SDEF	0,008	0,003	3,067	0,002	0,003	0,013
max_value_SDEF_std	0,016	0,006	2,577	0,01	0,004	0,028
mean_value_DEF_std	0,021	0,007	2,864	0,004	0,007	0,035
tot_value_GK_std	0,012	0,007	1,649	0,099	-0,002	0,027
Time	0,128	0,011	11,976	0	0,107	0,149
Treated	- 0,006	0,013	-0,455	0,649	-0,032	0,02
DID	- 0,048	0,023	-2,123	0,034	-0,092	-0,004

Model Tests

Omnibus:	4,454	Durbin-Watson:	1,907
Prob(Omnibus):	0,108	Jarque-Bera (JB):	5,684
Skew:	-0,047	Prob(JB):	0,0583
Kurtosis:	3,642	Cond, No,	630

5.4 Results Overview

Analysing the coefficients Time, Treated and DiD for the three leagues, it is notable that they have similar values and significance and even with some particularity as shown before such as Premier League Clubs having a higher percentage of clubs that maintain their coach compared to Serie A for example, the impact of sacking a coach is still similar.

- Time: All three leagues had time as a significant and positive variable which means that when clubs had a performance below the expectations, they tend to recover in the next games, independently if they belong to the comparison group or treated group. La Liga has the highest coefficient 0.137 which means that if a club hits the warning performance in t_0 they tend improve in almost 14% in t_1 on average.
- Treated: Once again the results are similar, for all three leagues, the coefficient is not significative, which means there is no average permanent differences between treatment and control.
- Time x Treated (DiD): Again, the result was similar for all three leagues, the parameters are significant and negative which means when a club sacks their coach during the season, the club tends to have a poorer performance in t_1 compared with the clubs that kept their coach. The DiD coefficients for all three

leagues is around -4% and Premier League had the coefficient with the biggest impact -4.8%, that means in the Premier League, when a club sacked his manager the performance in t1 decrease on average to 4.8% compared to the clubs that kept their coach.

6. Conclusion and future work

As a conclusion, based on this project analysing three of the most important leagues from season 2005-2006 to season 2019-2020 and comparing with results in the literature review, leagues share the same characteristics that is when a club has a performance under the expectation between games twelve and twenty-six, they tend to recover in the rest of the season, independently if the coach was sacked or not, and the DiD models show that with a positive and significative coefficient for time. Even both groups present better performance after an underperformance period the treatment effect is negative and significative what means that on average the clubs that sack their coach during the season has a worst performance after an underperformance period compared with the clubs that keep their manager.

As said before, these results are based on observations and the results is on average, that means is possible to find observations that are also against these conclusions. As a counterfactual for Serie A, it can be mentioned the case of Cesare Prandelli during season 2008-2009. In his first twenty-three games he had a performance of 45% and in the second part of 35%. Another example is Filippo Inzaghi on season 2018-2019, he was sacked after twenty-one games with a performance of 22% and after that Bolonha had a performance of 55%, much higher difference between t_0 and t_1 if we compare with the time coefficient for Serie A.

As future work, it would be interesting to add variables that would help understand why some management changes work well and others not that much and be able to identify when is the correct moment or correct situation to sack a coach in other to have better results keeping him.

Responding to the questions in the beginning "Does sacking a coach in the middle of the season increase the club performance during the season?" the answer is on average no. “Is sacking a coach in the middle of the season more efficient than keeping them until the end of the season?” the answer is also no, as explained before.

7. Limitations

Even though the models are giving good results, some variables that are important to measure results could be added to the model to improve these results, such as identifying the most important player(s) in the squad and having the information as to whether he/they played or not, quantity of games played at home and away and create a variable to measure the opponent difficulty.

A PCA could be implemented to reduce the dimensionality, variables correlated to value could be represented for lower quantity of components.

8. References

[1] D. Bloyce et al.(December 2008), Playing the Game (Plan): A Figurational Analysis of Organizational Change in Sports Development in EnglandEuropean Sport Management Quarterly, Vol. 8, No. 4, 359 - 378

[2] By RT, Diefenbach T, Klarner P (2008). Getting Organizational Change Right in Public Services: The Case of European Higher Education. J Change Manag. 2008;8(1):2 1-35.

[3] Besters, Lucas M., Jan C. van Ours, and Martin A. van Tuijl. (2016), Effectiveness of in-season manager changes in English Premier League football. De Economist 164: 335–56.

[4] González-Gómez, F., Picazo-Tadeo, A.J. and García-Rubio, M.Á. (2011), The impact of a mid-season change of manager on sporting performance, Sport, Business and Management, Vol. 1 No. 1, pp. 28-42.

[5] Martínez, J. A. (2012), Entrenador nuevo, ¿victoria segura? Evidencia en baloncesto /Chaning a coach, guarantee the win? Revista Internacional de Medicina y Ciencias de la Actividad Física y el Deporte 12 (48), 663-679.

[6] David Card & Alan B. Krueger (1994), Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania; The American Economic Review, Volume 84, Issue 4 (Sep. 1994), 772-793

[7] Callaway, B., & Sant'Anna, P. H. C. (2020), Difference-in-differences with multiple time periods. Journal of Econometrics.

[8] Book. Advances in K-means Clustering - A Data Mining Thinking - Hardback – 2012.

[9] Desboulets Loann (November 2018), A Review on Variable Selection in Regression Analysis Econometrics 6(4):45

[10] Book. Impact Evaluation in Practice (Second Edition) Paul J. Gertler, Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch - 2012

[11] Anders Fredriksson, Gustavo de Oliveira (October 2019), Impact evaluation using Difference-in-Differences RAUSP Management Journal 54(4):519-532

[12] Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics

[13] https://www.football.london/tottenham-hotspur-fc/news/cost-of-sacking-jose-mourinho-20416106

[14] https://www.transfermarkt.co.uk/

9. Appendix

Table 10 - Match data set example

Date	HomeTeam	HomeTeamPos	AwayTeam	AwayTeamPos	Formation	Manager	Attendence	Result	TeamMatches	Season
01/10/2000	AC Milan	6.0	Vicenza	16.0	3-4-3	Alberto Zaccheroni	46.836	02:00	AC Milan	2000
15/10/2000	Bolonha	17.0	AC Milan	2.0	3-4-3	Alberto Zaccheroni	34.631	02:01	AC Milan	2000
21/10/2000	AC Milan	8.0	Juventus	2.0	3-4-1-2	Alberto Zaccheroni	81.954	02:02	AC Milan	2000
01/11/2000	AC Parma	15.0	AC Milan	8.0	3-4-3	Alberto Zaccheroni	21.572	02:00	AC Milan	2000
05/11/2000	AC Milan	12.0	Atalanta	2.0	3-4-1-2	Alberto Zaccheroni	54.641	03:03	AC Milan	2000

Table 11 - Player's data set example

PlayerNames	Position	Number	Birth	Height	Foot	Joined	Value	Season	Team
Diego CavalieriDiego Cavalieri	Guarda-Redes	28	01/12/1982 (27)	1.89	esquerdo	01/08/2010	600 mil	2010	Cesana
Alex TeodoraniA. Teodorani	Guarda-Redes	91	21/09/1991 (18)	1.91	direito	01/07/2009	200 mil	2010	Cesana
Francesco AntonioliF. Antonioli	Guarda-Redes	1	14/09/1969 (40)	1.87	direito	07/07/2009	100 mil	2010	Cesana
Alex CalderoniA. Calderoni	Guarda-Redes	33	31/05/1976 (34)	1.82	esquerdo	07/01/2011	100 mil	2010	Cesana
Aldo SimonciniA. Simoncini	Guarda-Redes	86	30/08/1986 (23)	1.84	esquerdo	01/01/2011	100 mil	2010	Cesana

Table 12 - Coach consolidation data set

TeamMatches	Season	Manager	Match_qtt	FirstMatch	LastMatch
Alavés	2005	Chuchi Cos	18	27/08/2005	08/01/2006
Alavés	2005	Juan Carlos Oliva	5	15/01/2006	12/02/2006
Alavés	2005	Mario Luna	15	18/02/2006	13/05/2006
Alavés	2016	Mauricio Pellegrino	38	21/08/2016	20/05/2017
Alavés	2017	Luis Zubeldía	4	18/08/2017	17/09/2017