Erasmus School of Economics

Master Thesis Data Science and Marketing Analytics

Does crowd support matter? Analyzing home advantage and referee bias in the Dutch Eredivisie between 2017 and 2023

Author: Maarten Jol

Student ID number: 457259

Supervisor: Prof.dr.ir R. Dekker

Second assessor: Dr. C.S. Bellet

May 31, 2023

Abstract

This master thesis analyzes to which extent the attendance restricting measures resulting from the COVID-19 pandemic have had an impact on the home advantage and referee bias in the Dutch Eredivisie. Data on the performance of teams and referees from the seasons 2017-2018 up to and including 2022-2023 is used. Using multilevel modeling we compared team performance and referees' decisions Before COVID, During COVID with Restricted Attendance, During COVID Without Attendance and After COVID across 1670 games. We ran six different regression models. One linear regression and two Poisson regressions are performed on various team performance measures as dependent variable. For referees' performance, also one linear regression and two Poisson regressions were run on three dependent variables measuring referees' performance. It was found that home advantage measured in goals and points decreased in the period during COVID where no attendance was allowed, which reflected the inferior performance of the home team. For points, evidence was found that this decrease in home team performance persisted after COVID. Also, in games without fans, home teams dominated less in game statistics such as corners, shots and shots on target. However, it is unclear whether this is a result of difference in schedule across COVID periods. Furthermore, regarding referees' performance, it was found that referee bias decreased in the period when attendance was absent, as opposed to the seasons Before COVID. Away teams received fewer yellow cards and committed less fouls without attendance present. It was unclear whether this decrease of referee bias persisted after COVID. Implications and future research directions are discussed.

Acknowledgements

First and foremost, I would like to thank my parents, my brother, my sister and my girlfriend for all their support and for helping me find the strength and motivation to go through the process of finishing this Masters's Thesis. In addition, I would like to thank my supervisor Rommert Dekker for his patience and for helping me by dedicating a lot of time and effort to me over the course of producing this thesis.

Key Words

Football, Impact, COVID-19, Eredivisie, Home Advantage, Referee Bias, Multilevel Modelling, Poisson Regression, Linear Regression.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Research Context

In January 2021, English football club Liverpool lost in the Premier League[1] at their famous Anfield Road stadium to then 15th placed Burnley. A surprising result in itself, but what stood out the most was that the defeat ended Liverpool's 68 game unbeaten run at home in the Premier League. This run lasted for more than three years and span from April 2017 up and until January 2021, with a win percentage of 81%. A record only beaten in the English Premier League by Chelsea's unbeaten run of 86 games, from February 2004 to October 2008. Like Chelsea and Liverpool, there are also clubs in the Netherlands that are famously hard to beat at their home ground. Frequently, this is mainly attributed to the support of the home fans. For example, in the Dutch *Eredivisie*[2], both Feyenoord - in their famous stadium *De Kuip* - and FC Twente are currently unbeaten at home for more than a year. Also, Feyenoord is on a run of 14 consecutive European home matches without a loss, including 12 wins and an impressive goal difference (44-12).

Feyenoord and FC Twente often credit their supporters, claiming that they play with a man more when playing in their own stadium. The term "twelfth man" (Buraimo et al., 2010) is sometimes used in team sports played with 11 players per side, like football, to reference the home crowd, who are presumed to have a positive influence on the performance of the home team. For Feyenoord, this is even embedded in the club culture. Since 1999, no player has played with the number 12 on the back of their shirt, because it belongs to the supporters. When the stadium announcer reads the line-up of the team before each match, the so-called "*Legioen*" is named as the player with number 12, as a sign of gratitude and appreciation for their everlasting support (Peters, 2017).

We take a closer look at Feyenoord's impressive record in European home matches over the past five seasons to illustrate the power of playing at home. 17 wins, five draws and only four losses, with wins at home against big European clubs such as Manchester United, Napoli, FC Porto, Olympique Marseille, Lazio, Shakhtar Donetsk and most recently, AS Roma. Feyenoord's ability to rise above it's own qualities is often contributed to the magic of De Kuip on European nights, which are remembered dearly by supporters. It is therefore very curious to note that two of their last four European home defeats have been in the 2020-2021 season. Precisely the season where supporters were not allowed into the

[1] The highest professional football competition in England.

[2] The highest professional football competition in the Netherlands.

stadium due to the COVID-19 pandemic. The same goes for Liverpool's unbeaten run, which was ended in an empty stadium a a result of COVID measures.

This rises some questions. Is this a coincidence? Or would it be the supporters that make the difference in tough matches? Furthermore, are the teams mentioned above so hard to beat at home because of the superior quality of their players, or do their stadium and their home fans play a significant role in the results? And what would happen if supporters were absent? This thesis aims to shed light on the impact of the absence of a crowd at a football match on both home team and referees' decisions, by analyzing various seasons of football results and in-game match statistics from the Dutch Eredivisie. The main focus of this thesis is therefore home team advantage and referee bias in professional football and the influence of the public on this phenomenon[3].

1.2 Problem description

In football, it is expected that crowd attendance has an influence on the performance of football teams. The home team (usually) plays for their home fans, whereby the vast majority of the stadium is occupied by fans of the home team. These particular home fans, most of the time, do everything in their power to positively impact the result of the home team. Hence, one of the main functions of a crowd at a football match, next to the pleasure of the crowd to be able to enjoy a moment of leisure time, is to influence the performance of the home team that they support, or at least are presumed to support. Clubs and players are benefiting from crowds that offer them some extra motivation, which can lead to a higher performance of the players on the team and therefore better results for the club. However, during the COVID-19 pandemic, public was not allowed at football matches for a period of time, resulting in matches being played without an attendance. We can use data on the attendance at matches and in game performance of teams to see which impact the absence of crowds have had on the performance of clubs and teams. The results will be relevant for football players, coaches and managers and football clubs.

[3]In this thesis, we use various different words for the crowd at a football match, which are considered to have the same meaning. We use for example audience, public, attendance, fans, supporters and crowd support to all indicate the same concept, namely that of crowd at a football match.

1.3 Research questions

To study the effect of COVID-19, and in particular the absence of crowds, on home advantage (HA) and referee bias, this thesis aims to find the answer to a central research question. Next to the main research question there are several sub-questions that are used to formulate a well rounded answer to the main research question. The main research question that will be answered is as follows:

How has the absence of fans in football stadiums as a result of the COVID-19 pandemic influenced the home advantage and referee bias in the Dutch Eredivisie in the seasons 2017-2018 up to and including 2022-2023?

The most important aspect of this research is thus to examine to which extent the attendance restricting measures resulting from the COVID-19 pandemic have had an impact on the home advantage and referee bias in the Dutch Eredivisie. To answer this question, we must assess if a change in home advantage has occurred during separate COVID-periods where attendance restrictions differed. The seasons in the data set will be divided in four COVID-19 related periods, which will be explained further in section 3.3 of the Data. The central research question is split up in five sub questions to arrive at a well rounded answer. To examine the impact of the absence of attendance in the various COVID-19 periods on the home advantage and home team performance, we have constructed the following three sub-questions:

How has absence and return of fans changed the home advantage measured in points in the various COVID-19 periods?

How has the absence and return of fans changed the home advantage measured in goals in the various COVID-19 periods?

How has the absence and return of fans changed the home advantage measured in home team match dominance statistics in the various COVID-19 periods?

Then, we will examine what the effect of the absence of supporters is on various referee-related performance measures. As measures of referees' performance, we use the indicators fouls, yellow cards and red cards. We will try to examine the influence of absence of public on referees by answering the following two sub-questions:

How has the absence and return of fans changed the referee bias measured in fouls in the various COVID-19 periods?

How has the absence and return of fans changed the referee bias measured in yellow and red cards in the various COVID-19 periods?

1.4 Relevance of the subject

1.4.1 Managerial relevance

This thesis is relevant for various stakeholders. Firstly, the managers or coaches of clubs in the Eredivisie who can act on the results of this thesis. We will examine the effect of absence of crowds on various predictors of home advantage and referees' decisions. When it is clear which effect playing at home and in front of a crowd has on the performance of the home team with respect to in game factors, and to referees with respect to their decisions, coaches of the away team can instruct their players differently. They can for example adopt a different tactical approach when playing away from home.

Secondly, this thesis is relevant for the Dutch government, especially the Dutch House of Representatives. During the COVID-19 pandemic the government decided on multiple occasions at different points in time to not allow fans at sport events, such as professional football matches (RIVM, 2020). This was done to prevent the spreading of the COVID-19 virus. However, the impact of not allowing fans into the stadiums are immense for football clubs, costing some clubs more than a million euro's of revenue per match (Witlox, 2022). With the stadiums closed for longer periods, clubs are losing millions in yearly revenue. The income from matches decreased from €106m and 20% of total revenue, to €25m and 5% of total revenue when comparing season 2019-2020 to 2020-2021 (Rozendaal, 2021). To add to this, with this thesis the consequences on the performance of clubs will be outlined, to show whether the absence of crowds is also detrimental to the home team performance. If a similar situation would occur again in the future, where the government would maybe see the need to exclude fans from matches, they have the full knowledge of the impact their decisions will have on the national professional football competition, both financially and on sporting grounds.

The same goes for the Dutch National Football Association (KNVB[4]). The KNVB, as the highest football governing organization in the Netherlands, has the power to punish clubs for misconduct, be it financially or otherwise. An example being the misconduct by football fans of a specific club. When this happens, to look for an appropriate punishment, they can use the results of this thesis. If the result would be that home advantage decreases

[4]Koninklijke Nederlandse Voetbalbond.

significantly without fans, an appropriate punishment for the misbehaviour of a club's fans could be a match without supporters present.

Lastly, the results of this thesis are relevant for the players, because they can use the results of this thesis to better prepare for matches. If they know that home fans are a main driver of home advantage, they can prepare themselves mentally for an away game. Also, when referee bias is present according to the thesis' results, away players can put more emphasis on the behaviour of the referee[5] to try to balance out the effect of the home crowd on the referee.

1.4.2 Academic relevance

The topics of home advantage and referee bias have been subject of plenty of research. However, home advantage in the Dutch Eredivisie specifically has only been researched intensively by one researcher (van Ours, 2018, 2019). The research by van Ours was also done before the COVID-19 pandemic. Also, the research of van Ours (2018; 2019) has specifically focused on the use of artificial pitches in the Dutch competition and its effect on the home advantage. Though its findings are considered interesting and informative, this thesis focuses more on the impact of the crowd on home advantage, which has not been researched intensively yet in the setting of the Dutch football competition.

Most of the published research on the impact of COVID-19 on HA has looked into the 2019-2020 season (Leitner et al., 2022), when in various European football leagues, so-called "ghost games" took place. There are a couple of remarks on these researches.

First, the majority of the 2019-2020 season was played *with* fans, since the season was only stopped in March. It would therefore be interesting to look into the season after, 2020-2021 that is, to see how a nearly full season without fans, played under a balanced schedule, has had an impact on home advantage. Then it can be researched what the effects are of the absence of fans for a longer period, and how this impacts the home advantage.

Second, the aim of this thesis is also to look at the return of supporters and how that has impacted the home advantage. In the seasons 2021-2022 and 2022-2023 in the Netherlands, fans were gradually allowed back into the stadiums again to attend football matches. It is interesting to look into these seasons and to research whether the home advantage immediately changed with the return of fans, and if there were differences compared to

[5] Even though this is is sometimes seen as unethical, it still happens in professional football.

the periods in COVID-19 without fans, and before COVID-19 with fans. Therefore, data of the 2021-2022 and 2022-2023 seasons were also used, as has not often been done before by previous research on the topic of COVID-19 and its influence on home advantage and referee bias.

Finally, as mentioned in the previous paragraphs, there is plenty of research on both home advantage as a phenomenon in itself, and the impact of the absence of a crowd on home advantage during COVID-19. However, both have not been researched only for the Dutch competition. Especially COVID-19 and the impact on home advantage in the Eredivisie has not been a sole topic of research. This is also because the Eredivisie has not continued its competition in the 2019-2020 season. All the more reason why it is academically relevant to fill the research gap in this paper, and to try to find the impact of the absence of a crowd on home advantage and referee bias in the Dutch Eredivisie.

1.5 Thesis structure

First, in the Literature Review, we will examine the literature on the topics of home advantage and referee bias. We will assess the most important factors of home advantage based on the literature. After, we look at the influence of the COVID-19 pandemic and the absence of fans on the home advantage and referee bias according to the literature. Next, in the Data section, we show the data we have used, the source of the data and how we altered the data from the source. Also, we define some key variables we use in further stages of the research.

Then, we discuss the methods and methodology we use. The following section shows and discusses the findings after applying the methods and methodology on our data set. In the last section, the Conclusions and Discussion, we evaluate our findings and compare them to the earlier found statements from the Literature Review. Also, we list implications of our research, give recommendations for future research and discuss limitations for our paper.

2 Literature Review

Home advantage in professional football is a well-known phenomenon, often cited by players, coaches and fans, who believe that playing at home can be beneficial to their team's performance. It has also been the subject of scientific research, for example to determine what home advantage is and how to quantify it, and what the main drivers of HA are. In this section, after carefully studying literature on home advantage in football, the most important findings will be discussed and compared. Then, our research will be positioned against the current literature, to see what the common grounds are, and what the gaps in the literature are. This thesis aims to contribute to fill those gaps.

It is important to explain how the articles that are used in this Literature Review are found and why they are used to construct the review. When starting with doing research among the available literature, we first used Google Scholar to search for key words on the topic. For example, using searches like *"Home Advantage"*, *"COVID-19"*, *"home supporters influence"*, *"influence COVID-19 on home advantage"*, *"referee bias"*, *"referee bias influence on home advantage"*. These searches provided some useful articles. Articles were then judged on various criteria, such as the type of journal the article was published in, the year in which the article was published, the author(s) and the number of citations. After some leading articles and authors were established, these leading articles and authors were used to examine which articles they referred to most in their text. Especially highly cited literature reviews/overviews were used in this process, to find as many relevant papers as possible. The total process of searching for literature resulted in an accumulation of approximately thirty to forty articles which were deemed relevant, and which were used to construct the Literature Review.

The Literature Review is structured as follows. First, we look at how home advantage is defined and measured. After that, the drivers of home advantage according to the literature are listed and discussed in subsections. Then, home advantage in (and after) COVID-19 times will be examined to see what the consequences of the pandemic were (and are) on the home advantage. In addition, we determine which findings from the literature are relevant for the context of this specific research. Lastly, we review which methods are used most in the literature, and we give a summary of the Literature Review as a whole.

2.1 Home advantage

Although some earlier articles and newspapers briefly mention the subject of home advantage, the existence of home advantage in team sports was first established by Schwartz and Barsky (1977). They researched various team sports in North America and found that the HA is most dominant in indoor sports, like for example basketball and ice hockey, and less so in outdoor sports like baseball and American football[6]. Also, they attributed the effect of HA mainly to the social support of home audience. This has been a subject of much attention in literature on home advantage in the following decades, and has both contradicted and agreed upon.

First, when quantifying home advantage we follow the definition of Pollard (1986), who defines home advantage as: "the number of games won by teams by playing at home expressed as a percentage of all games played"[7]. A figure of 50% thus indicates no home advantage and the higher the percentage above 50%, the stronger the home advantage is. This definition is agreed upon by Courneya and Carron (1992). Pollard (1986) further finds that the home advantage is the highest in football, contradicting earlier findings by Schwartz and Barsky (1977). Pollard (1986) also states some important facts regarding home advantage:

- It has been present ever since the emerging of organized football leagues at the end of the 19th century.
- It is a phenomenon which has been established worldwide, and it can vary significantly between continents and countries.
- It has declined in major professional football leagues in Europe over the past decades.

In agreement with this last finding from Pollard is the paper from Peeters and van Ours (2021), who find that home advantage in English professional football divisions has declined over time. Another research also shows the decrease of home field advantage for various European football leagues over time, including the Dutch Eredivisie (Maimone and Yasseri, 2021).

Furthermore, it has been concluded multiple times that home advantage can vary heavily

[6]To avoid confusion, in this thesis the term "football" will be used for the sport which people from North America would call "soccer". This is done because in Europe the term football is commonly used to refer to the sport North Americans would thus call soccer. The term "football" in Northern America would refer to the so called "American Football", as it is called in Europe.

[7]Provided that there is a balanced home and away schedule and all games are played to a finish.

between countries. In Europe, Pollard (2006a) finds that home advantage is the highest for Balkan countries (around 70-80%), and lowest (below 60%) in Northern Europe, comprising Scandinavia and the Baltic countries (Pollard, 2006a, 2006b). Pollard finds a figure of 61% home advantage for the Netherlands. This is approximately the same as close countries in Western Europe like Belgium, Germany, France, and England, who all have a value between 60-65%. Also, home advantage differs significantly between countries in Southern Europe, like France, Italy and Portugal, and is highest on the islands of Corsica, Siciliy and to lesser extent Sardinia (Pollard and Gómez, 2009). This higher advantage is attributed to the fact that those regions are more isolated and culturally distinct communities, which can lead to increased HA. Lastly, Marek and Vavra (2020) also looked at differences in HA between European countries, but with more recent data[8]. They found that the highest home advantage was in Greece, and the lowest in the fourth division of English football. For the Netherlands, they found that the home advantage is stable for the seasons in their data set, as opposed to some other countries in the study (Marek and Vávra, 2020).

2.2 What are the components of home advantage in the literature?

In a literature review by Courneya and Carron (1992), one of the most cited and referred to papers on the topic of home advantage in sport, four game location factors are identified, which are likely to affect the degree of home advantage. These factors are crowd factors, learning/familiarity factors, travel factors and rule factors. In the following sections, we will list various research on these factors to conclude the statements from the literature and to see if we can identify any gaps in knowledge. Also, we determine how relevant each factor is for our research setting. However, it is likely that home advantage is the result from a complex inter-relationship between all the to be named factors (Pollard, 2006a).

2.2.1 Crowd factors

The attendance at a football match is the most obvious factor to have an influence on the home advantage, and is perceived by football fans to be the most dominant factor (Wolfson et al., 2005). It is also the factor which is the main focus of the current research. Although, through the years, research has shown contradictory findings on the impact of crowds. It has not been clearly established if the impact of a crowd on home advantage comes from giving the home team an advantage, or the away team a disadvantage, and whether this

[8]Seasons 2007-2008 to 2016-2017.

is done directly by altering the behaviour of players, or by influencing decisions of the referee. Therefore, the term *referee bias* will also be considered under the crowd factor and will be discussed in the next subsection, as the two are closely connected. In this section, only research which has been carried out before the COVID-19 pandemic will be included, to see what the prior consensus in the literature was. In subsection 2.3, we will focus on the impact of COVID-19 in the light of the absence of crowds.

Before the pandemic created the opportunity for a natural experiment, there were very little football matches without attendance, so the effect of crowd support on home advantage was mainly researched through so called same-stadium derbies. These are games played between clubs who share a stadium[9]. By researching these matches, the effect of a crowd can be isolated, since learning/familiarity and travel factors do not play a role. In addition, some research uses matches in Italy where there were no fans allowed because stadiums of the home team did not comply with safety rules at the time (Van de Ven, 2011).

When researching same-stadium derbies, Ponzo and Scoppa (2018) found that the crowd support is of strong and significant impact on the performance of the home team. The home team scores more goals on average than the visiting team and has a higher probability of winning. They suggest that a large part of the advantage is due to the support of the crowd (Ponzo and Scoppa, 2018). Contrary to this finding, van de Ven (2011) claims that home crowd support is not a necessary precondition for the home advantage. He bases his claim on the fact that when home crowd was not present in his research, home teams still had an advantage. Also, in his research into same-stadium derbies, he found no home advantage, as opposed to Ponzo and Scoppa (2018). However, because in a regular football season it is highly unlikely that many matches are played without fans, the sample size of van de Ven is limited. Therefore, we must take this into account when interpreting the findings. Though the conclusion is, that before the pandemic, there was still no consensus on the exact effect of crowd support.

2.2.2 Referee bias

Something that goes hand-in-hand with the possible influence of a crowd on the home advantage, is *referee bias*. Referee bias is defined as the tendency for referees to be

[9]Mainly found in Italy between clubs from the same city, for example Lazio and Roma, and AC Milan and Internazionale. In these matches, the home team has more supporters due to the vast majority of tickets being owned by season ticket holders.

subconsciously influenced by (the noise of) a large crowd, which results in the favouring of the home team when making decisions (since it is usually the home team fans creating the noise). Usually the bias results in (subconsciously) awarding more penalties to the home team, and awarding less fouls and cards against the home team (Nevill et al., 2002). When researching the impact of crowd noise on referee decisions, the researchers found that observers in an experiment had a greater tendency to award fouls against away players when viewing the challenges with crowd noise compared to without. Observers refer to the crowd for guidance in their decision (Nevill et al., 1999). This is again concluded by Nevill, Balmer and Williams (Nevill et al., 2002), who find a dramatic effect of crowd noise on referees' decisions, expressing itself in significantly less fouls being called against the home team.

Looking back at the research of Ponzo and Scoppa (2018), they confirm the finding that the support of the crowd does in fact impact the referee. They find that referees are more prone to favour the home team, by awarding more penalties to the home team and handing out more cards to the visiting team. They argue that the impact of the crowd is thus twofold: it impacts both the players positively, as it influences the referee to be in favour of the home team. To conclude, we must take into consideration the referee bias when researching the effect of crowd on home advantage.

2.2.3 Learning/familiarity factors

According to Courneya and Carron (1992), learning or familiarity factors stem from the different physical characteristics of a playing facility. Examples of conditions whereby the home team could benefit due to more familiarity with the condition are the size of a football field, the texture of the field (artificial grass vs. natural grass) and the architecture of a stadium (open vs. closed). Visiting teams are likely to be less familiar with the (pitch) conditions as opposed to the home team, who regularly uses the facility.

The first research on the effect of artificial pitches was conducted in 1993 in England, at the request of the Football League (Barnett and Hilditch, 1993). Barnett and Hilditch (1993) found a statistically significant advantage in home match performance for English teams using an artificial pitch in the seasons 1981-82 to 1988-89. Clarke and Norman (1995) confirmed the findings of Barnett and Hilditch. When investigating the home ground advantage of individual clubs in English football, they found that the clubs playing on an artificial pitch during the eighties had a mean HA value that was significantly

higher than of clubs that had played their seasons on a natural pitch. The type of pitch was of significant impact on home advantage for individual clubs (Clarke and Norman, 1995). Because of these findings, the English Football Association (FA) banned the use of artificial pitches in 1988. In addition to Barnett and Hilditch (1993) and Clarke and Norman (1995), also Peeters and van Ours (2021) find that home advantage correlates with the use of artificial pitches, although the home advantage fluctuates from year to year.

When looking at the research setting of this thesis, being the Netherlands, the use of artificial pitches is still allowed and various clubs have been - and still are - playing on this surface for years. Based on research by van Ours (2019), who analyzes match data from the seasons 2014-2015 to 2017-2018, it can be concluded that teams in the Eredivisie who play on an artificial pitch have an additional home advantage over teams who use normal grass. It is already decided that in the future, artificial pitches will also be prohibited in the Netherlands, but this ban is from the summer of 2025 onward. Therefore, based on previous research, it is highly relevant for this research to include a variable which indicates whether the home teams has an artificial pitch or not. Table 3 in the Appendix shows which clubs played on an artificial pitch for each season in our data set.

2.2.4 Travel factors

It has been studied by various researchers whether distance travelled is of influence on the performance of athletes, although evidence is mostly inconclusive (Pollard, 2008; Nevill and Holder, 1999). The impact of traveling for the away teams would be through fatigue as a cause of traveling, and the breaking of familiar routines and/or habits.

Goumas (2014b) shows that in competitions where times zones are crossed by traveling for away matches, travel effects are present and can be of even more impact than other factors, like the crowd. This is supported by significant results from a study in Brazil, which also suggests climatic conditions might be of influence (Pollard et al., 2008). On the contrary, a study on home advantage in the European continental competitions[10] shows that the geographical distance and climatic differences are not of significant impact on the home advantage, suggesting that these dimensions of distance matter less in Europe (Van Damme and Baert, 2019). Van Damme and Baert (2019) do however find that the altitude of the playing facility is of significant impact: when the home team plays

[10]The UEFA Champions League and the UEFA Europa League.

at a higher altitude, they benefit substantially more from their home advantage. Also, a consistent finding is, that in (same-city) derbies, where (almost) no travel of the away team is required, home advantage is reduced (Pollard, 1986).

However, since the Netherlands is smaller than the countries from the above studies, has almost no climate differences, and is one of the most flat countries in the world, the found influences of travel distance, climate difference and altitude seem to be of no relevance to the Dutch football competition and are therefore not taken into account in this study.

2.2.5 Rule factors

Rule factors are found to have a minor contribution to the home advantage, and only in some sports, like baseball and ice hockey, they do play a role (Nevill and Holder, 1999). In football there are no particular rules that favour the home team, hence this factor is expected to be of little influence.

There are however some rules introduced in football in the last few years, that have the potential to affect the home advantage. Examples include the introduction of the Video Assistant Referee (VAR), and the possibility for coaches to substitute five players per match instead of the previous three. This rule change regarding substitutions was, at first, a temporary measure introduced in May 2020, to help the players and protect their health during the pandemic[11]. However, the rule was made permanent in 2022 by the International Football Association Board (IFAB) (Trouw, 2022). The rule is expected to favour richer teams, who can afford to make sure they have a surplus of good players to make more substitutions and thereby possibly impact matches. We however don't control for any rule factors in this research, since their impact is expected to be neglegible.

2.2.6 Other possible factors

In addition to the factors mentioned above, there are also various other factors which are mentioned in the literature:

- *Special playing tactics*: Teams often apply a more defensive tactic when playing away from home (Pollard, 2006a).
- *Psychological factors*: This factor is kind of a vicious cycle. If players keep believing in the existence of a home advantage, the will likely play home matches with more

[11] The 2019-2020 season would finish later than planned, because of a break due to COVID-19. The Eredivisie would never finish the season.

confidence, therefore contributing to the continuing existence of HA (Pollard, 2006a).

- *Territoriality*: Players' testosterone levels were higher before home matches than for away matches, indicating a sense of territoriality, which could result in better performance (Neave and Wolfson, 2003). This is also given as a potential explanation why HA in the Balkans is the highest in Europe (Pollard, 2006b).

2.3 Home advantage and the COVID-19 pandemic

Due to the COVID-19 pandemic that reached Europe in 2020, various professional football leagues had to deal with imposed restrictions by their country's respective government. The restriction that had maybe the most impact on football was the fact that fans could not be present at matches anymore. Because of that, numerous matches in the 2019-2020 and 2020-2021 seasons in several European professional football leagues had no (or little) attendance[12]. This situation however provided a very interesting natural experiment for researchers to investigate the impact of crowd support on home advantage and referee bias in football. Normally, the setting in which matches are played without the support of a crowd is very unusual, mainly because of the fact that crowds are presumed to be highly valuable to football clubs, both financially and competitively. Now, suddenly, a lot more data was available comprising football matches without attendance, and because of that the influence of crowds on the home advantage could be studied further. We already discussed some research on the impacts of crowds in subsection 2.2.1. In the current section, we will focus on research conducted in or after the COVID-19 pandemic, that use data from European football leagues during the pandemic to research the impact of COVID-19 and the resulting absence of attendance on the home advantage and referee bias.

Overall, when looking at the literature researching the influence of COVID-19 on home advantage, we find that most literature suggests that the absence of fans has a considerable impact on home advantage (Leitner et al., 2022). The majority of studies, 18 of the included 26 in the literature review by Leitner et. al. (2022), find a reduced or strongly reduced home advantage as a result of ghost games, explained mostly by reduced referee bias and the absence of motivation from the crowd. However, there are also six studies (of the 26) that do not find a change in home advantage, providing somewhat inconclusive

[12] In the Netherlands, there were even games in the 2021-2022 season without attendance due to COVID-19 related restrictions.

results. Though, there are no studies which decisively find an increase in home advantage as a result of the absence of a crowd. Zooming in further, by only considering peer-reviewed articles and the top four ranked leagues in Europe, 8 of 13 studies find a reduced or strongly reduced home advantage. These findings all point towards a decrease of home advantage. As a side note, the results can differ heavily between European countries.

There may be differences between European leagues, but if we take the researches that used data on more than ten different European leagues, they are pointing in the same direction as Leitner et. al. (2022): home team performance was reduced in COVID-19 games compared to previous seasons with spectators (Cueva, 2020; McCarrick et al., 2021). Only Benz and Lopez (2021) conclude mixed findings, indicating that the change of HA is league dependent. They found that in some leagues, evidence of a drop in HA was overwhelming, while in other leagues, sometimes HA even increased. In addition, three researches comprising more than ten leagues find a significant drop in cards awarded to away teams, and fouls awarded against away teams, indicating a decrease in referee bias, confirming the earlier suggestion by Leitner et. al. (2022) (Cueva, 2020; McCarrick et al., 2021; Bryson et al., 2021). This decrease in referee bias is likely a causal effect of the complete lack of crowd pressure on the referees, compared to the influence that the pressure of a crowd normally has (which has been outlined in section 2.2.2).

The majority of research on the impact of the absence of crowd attendance focuses on the ending of the 2019-2020 season, since in most European leagues, that is the season in which the most ghost games took place. Sors et. al. (2022) however also focus on the season after, the 2020-2021 season, using the top five ranked European leagues[13]. Due to the continuation of the pandemic, the majority of the matches were still played behind closed doors. This expanded earlier research with more matches without attendance, and the addition of a fully balanced home/away schedule, which wasn't present in the 2019-2020 season. They find, in agreement with earlier research, that the absence of spectators significantly reduced the HA compared to earlier season with spectators. Also, referee bias was eliminated, suggesting this mainly derives from the pressure normally exerted by spectators, again confirming earlier the suggestion of Leitner et al. (2022), among others (Sors et al., 2022).

When the pandemic reached Europe, around March 2020, the 2019-2020 Eredivisie was put to a halt. Instead of rescheduling the remaining matches, the KNVB decided to aban-

[13]That is, England, Spain, Italy, Germany and France according to UEFA rankings.

don the season, and the standings at the time were used as the final ranking. Hence, no ghost games were played, and little to no research on COVID-19 and HA has therefore included the Dutch Eredivisie. However, in the 2020-2021 and 2021-2022 season, there were still a lot of matches played in the Netherlands without attendance due to a continuation of the pandemic. Where Sors et. al. (2022) focused on the top five ranked countries, this research will focus only on the Netherlands, the sixth ranked European country according to UEFA rankings, as of May 2023. It will take into account also the seasons after 2019-2020, where there was limited to no attendance for parts of the seasons. By doing this, it will fill both the gap regarding the research on the effect of crowds on home advantage and referee bias in Dutch football, as will it shed more light on the impact of the continuation of empty stadiums on the HA, adding to the research of Sors et. al. (2022). An expansion is made to the research of Sors. et. al. (2022), by using both data of matches without attendance in a full season, as data on seasons where public was allowed back into the stadium again. By doing this, the situation before the pandemic can be compared to matches without crowd during the pandemic, and the situation where crowd was allowed again after the pandemic.

To conclude this section, we look at the impact of the pandemic and the resulting absence of crowds on the factors of home advantage we mentioned in section 2.2. We can conclude the following:

- *Crowd factors and referee bias*: Great impact. With the absence of crowd, home teams tend to perform worse. Also, referee bias decreased or disappeared as a result of the absence of a crowd.
- *Learning/familiarity factors*: The playing facilities did not change, so little impact of COVID-19 and the resulting absence of attendance.
- *Travel factors*: No impact on travel factors.
- *Rule factors*: There could be impact be via extra substitutions, but this is still unclear.
- *Other possible factors*: It is believed that the pandemic and the absence of crowd is of little to no impact.

2.4 Methods

Benz and Lopez (2021) include in their paper an overview with a comparison of already conducted post-COVID research on home advantage and referee bias in football. It it shown that most researches use correlation or Linear Regression as main method. Only three researches, being their own research, McCarrick et al. (2021), and Bryson et al. (2021), use Poisson regression modelling to find the effect of COVID on home advantage. Where Benz and Lopez (2021) use a bivariate Poisson regression and add expansions to this regression, McCarrick et al. (2021) and Bryson et al. (2021) make use of a univariate Poisson regression.

The benefits of the Poisson distribution are demonstrated by Goumas (2013). He explains that often in research into home advantage, the difference between the home and away score is used as dependent variable in an Ordinary Least Squares (OLS) regression, a method appropriate for continuous outcomes and where the dependent variable is normally distributed. This method also involves treating each match as one single observation in the data set. However, discrete performance measures such as goals scored by both teams in a match are count data which are not normally distributed, and are positively skewed. Hence, we use the Poisson regression for modeling count variables in our research. When using this type of regression to measure the home advantage, each match contributes to two observations. One for the home team and one for the away team, as opposed to the OLS method of regression which uses a single observation per match.

In this thesis, we will make use of the regression approach called multilevel modelling, the same methodology as shown in the paper of McCarrick et al. (2021), who also refer to Goumas (2013) for a similar approach. The most important reason for using the multilevel modeling approach is that it allows for the adding of confounding factors in the regression analysis. Adding these confounding factors allows us to provide the relative contribution of specific variables on home team performance and referee-related outcomes. In this research, we are interested in the variables COVID-period (Before COVID, During COVID with Restricted Attendance, During COVID without Attendance, and After COVID) which indicates the attendance of crowds at a match and in the Venue of a match (Home or Away). Their interaction is of main interest in this thesis, as also their interaction with various other variables. We use variables for the strength of opponents and the importance of a match for the home and away team to control for differences in schedule. All the variables which are named here will be explained in detail in the Data section.

2.5 Summary

Looking back at the findings from the Literature Review, there are several takeaways to keep in mind for the remainder of this research, and especially for our research context, which is the Dutch Eredivisie. We first have found that the HA in the Netherlands is stable over the last decade or so. Then, we examined the factors which have an influence on home advantage. The most important factors being crowd factors and the referee bias. From the rest of the factors, we only account for the learning/familiarity factor by taking into account artificial pitches, which are still of use in the Eredivisie and are shown to have an effect on the home advantage. Then we have looked at the impact of the COVID-19 pandemic and the resulting absence of attendance on the home advantage and referee bias. It was concluded by various research that the home advantage decreased in the period without attendance, and that the referee bias was diluted in multiple cases. These two concepts were however not researched in the Netherlands, which we aim to do in this thesis. We expect, based on previous research, that also in the Netherlands, the home advantage and referee bias in the period without attendance have decreased. To examine this, we use multilevel modelling, based on the methodology of research from McCarrick et al. (2021), whereby most variables are following a Poisson distribution.

3 Data

In this section, the data sets that were used in this paper are provided. The data sources are stated and it is explained how the data is manipulated or altered from the source. Also, we explain some variables and show how we have defined them. The processing of data sets and the further data manipulation was done by using R programming language (R Core Team, 2022).

3.1 Eredivisie and its format

First, to be able to understand the data and their context, the format of the Dutch Eredivisie is explained. There are 18 teams in the competition every season, whereby each team plays 34 games in total: 17 home games and 17 away games. This results in 306 matches in total in a normal season. A regular season is played from approximately August to May[14]. Each win is awarded three points, a draw accounts for one point and for a loss no points are given. Most of the matches are played on Saturday evening or Sunday afternoon: 1398 games, which is 83,7% of the total games in the data set. A few times per season there are rounds during the week, then games are also played on Tuesday, Wednesday or Thursday.

The eighteen clubs in the league vary each season because there is a regulation regarding promotion and relegation of teams. This particular regulation has also varied for the seasons in this data set. In the season 2017-2018 and 2018-2019, the 18th ranked team with the lowest amount of points is relegated directly to the second division. The number 16 and 17 after 34 matches play in a play-off system against various teams from the second division to determine whether they get relegated. Season 2019-2020 was a unique season, because it was stopped after 26 game weeks due to the COVID-19 pandemic. Therefore, it was decided by the KNVB that no team would get relegated, and there also were no teams promoted. From the 2020-2021 season onward the regulation has changed. Now, the two lowest placed teams are relegated directly, whereas the 16th placed team plays in a play-off system against teams from the second division to determine whether or not they will be relegated. Play-off games for promotion/relegation and for a ticket for European football are not included in the data set, as the data comprises only regular season games. Furthermore, each season there is a national cup competition, the *KNVB Beker*. Results for this competition are also not included in the data, because there is no

[14]The exception here is the COVID-19 season, which lasted until March.

balanced schedule and the fixtures differ each season. All teams in the data set and their ranking for each season can be seen in Table 3 and 4 in the Appendix. x

3.2 Data description

Six years of data from the Eredivisie is taken from `https://www.football-data.co.uk`. The data include the date of the match, the time of the match in Western European Time (UTC), the name of the home team and the away team, the full time result[15], the half time result, the goals scored by both teams at half time and at full time, the amount of shots of both teams, the amount of shots on target of both teams, the amount of fouls commited by both teams and the amount of yellow and red cards of both teams.

The data comprises six seasons: season 2017-2018 up and until the 33th match day of the 2022-2023 season. The data starts at 11-8-2017, the first match day of the 2017-2018 season, and ends on 21-05-2023, the 33th game week of the 2022-2023 season. The cut off point of the second to last match day is chosen to still be able to gather the most amount of data possible from after the COVID-19 pandemic, since the 34th match day still needs to be played at the time of writing this thesis.

The 2017-2018 and the 2018-2019 seasons both consist of 306 matches. Because of COVID-19, the 2019-2020 was put to a stop in March, which resulted in numerous matches not being played anymore. Therefore, this season only has 232 matches. To account for the fact that the data keeps a balanced home/away schedule whereby each team in a season plays each other home and away, 74 matches[16] in the 2019-2020 season are omitted from the data set, leaving this season with 158 matches in total. The 2020-2021 and 2021-2022 season have 306 matches again, where contrary to the 2017-2018 and 2018-2019 seasons before COVID, not all matches have been played with an attendance, because of attendance restricting measures still being present. From the 2022-2023 season 33 playing rounds are taken. This season therefore consists of 297 games. As we did with the 2019-2020 season, we also omitted nine matches from the 2022-2023 season which are the opposite fixtures of the matches that still have to be played, to keep a balanced schedule. This leaves the 2022-2023 season with 288 matches. In total, the full data set therefore consists of 1670 games, spanning across six seasons. Of these 1670 games, 1320 (79%) were played with the presence of a crowd, and 350 (21%) are played without

[15] Home team win, draw or away team win.

[16] These are the opposite fixtures of the matches in the last 7 to 8 game weeks which were never played anymore due to the pandemic.

attendance.

The total number of different clubs in our data set is 26. Only eight teams have played in all the seasons in the data set[17], whereas three teams have played in only one season[18]. An overview of all the clubs in the data, in which season they played in the Eredivisie and their respective ranking for each season can be seen in Tables 3 and 4 in the Appendix.

The data set from `https://www.football-data.co.uk` is supplemented with data from FiveThirtyEight, an American website that uses statistical analysis to tell compelling stories about various topics such as elections, politics, economics, science and sports. The data from FiveThirtyEight are retrieved via `https://data.fivethirtyeight.com/#soccer-spi` and include the rating (which is called SPI and is based on the Football Soccer Power Index) of each team and the importance of the match for each team. In subsection 3.4 we expand on these metrics and how they are calculated. For an expanded explanation, more information on the FiveThirtyEight data base can be found at `https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work/`.

3.3 COVID-19 and its impact on attendance

As earlier mentioned, there were periods in the Eredivisie in the 2020-2021 and 2021-2022 seasons when there were no supporters allowed inside the stadiums due to regulations from the Dutch government as a consequence of the COVID-19 pandemic. Because the regulations from the government were often altered, the periods are quite arbitrary and therefore hard to define. We have made the choice to split up the data in four different COVID-19 related periods, which we will explain here. An overview of these periods and their exact dates can also be seen in Table 5 in the Appendix:

- *Before COVID*: The seasons 2017-2018, 2018-2019 and 2019-2020 when there were no restrictions on the attendance and every match was played in front of a crowd with the possibility of maximum capacity. This period comprises **770** matches (already excluding the 74 matches from the 2019-2020 season which were not played both home and away), accounting for 46% of the data.
- *During COVID with Restricted Attendance*: There were periods in the pandemic when attendance was allowed, however this attendance was most of the times restricted to one third or two third of the maximum capacity of a stadium. There were

[17] Being Ajax, AZ, FC Groningen, FC Utrecht, Feyenoord, PSV, sc Heerenveen and Vitesse.
[18] Being De Graafschap, FC Volendam and Roda JC.

also periods in the pandemic where attendance wasn't allowed. Therefore, we have decided to split up the COVID period into a period with and one without attendance. The period of COVID as a whole is defined as follows. As begin point of the COVID period we have chosen the first match week of the 2020-2021 season, on the 12th of September 2020, which was the first match week when fans were not allowed into the stadiums. The end point of the COVID period is the moment when fans were allowed back into the stadiums again with maximum capacity. The end point therefore is match week 22 of the 2021-2022 season, on 13 February 2022. Because the regulations in the Netherlands regarding attendance at matches varied through the years where COVID was present, the periods without attendance and with restricted attendance are not chronological, but are rather intertwined. The exact dates and match weeks that cover both periods are specified in Table 5 in the Appendix. In total, there were **156** matches played in the COVID period where attendance was allowed, accounting for 9% of the data.

- *During COVID Without Attendance*: Matches played in the previously defined COVID period when no attendance was allowed. This comprises **348** matches in total, accounting for 21% of the data.
- *After COVID*: The period in which maximum capacity was allowed again in all the stadiums in the Eredivisie. The starting point of this period is match week 23 of the 2021-2022 season, on the 18th of February 2022. The end point is the 33th matchday of the current 2022-2023 season. In total, this period accounts for **396** matches, which is 24% of the data (already excluding the 9 matches from the 2022-2023 season which were not played both home and away).

There were five exceptions of cases were there was no attendance at the match, which were not COVID-related. They are handled as follows:

- Case 1 and 2: In game week 9 of the 2021-2022 season, on the 17th of October 2021, a part of the stadium of N.E.C. collapsed after their game against Vitesse. Their following two home matches in game weeks 11 and 12 were played without an attendance due to the stadium being investigated regarding safety. In these two matches, there was still a restricted attendance allowed. However, because these two games are in the period During COVID, we have chosen to still include them in the 'Without Attendance' category, even though the reason of absence of attendance is

not COVID-related.

- 3: The third case where there was no attendance, where attendance was allowed, was in game week 21 of the 2021-2022 season. In this game week, the government partly lifted the COVID measures and allowed one third of the capacity of the stadium to be filled. However, FC Twente - after discussing with its fans and sponsors - decided that it was all or nothing, and that the stadium would stay empty. This match is also added to the period "During COVID Without Attendance", since it was still During COVID period.
- Cases 4 and 5: After an incident in the KNVB Beker semi-final match Feyenoord - Ajax on 5 April 2023, the KNVB decided to impose new rules. When items were thrown on the field multiple times, the referee would abandon the match and the match had to be continued at a different time and day without supporters present. This happened twice in the remainder of the season, in game weeks 30 and 32 of the 2022-2023 season, both at a match of FC Groningen. The two matches were abandoned in the 18th' and 9th' minute respectively[19], and are therefore assigned 0 attendance. These matches were however in the period After COVID and are therefore not included in the matches without attendance, which were all in the period During COVID.

3.4 Football match statistics

In this section we briefly expand on some football metrics in our data, some of which are also used as dependent variables. The definitions are taken from van Leeuwen (2020) and from statistical website FiveThirtyEight via `https://fivethirtyeight.com/methodology/how-our-club-soccer-predictions-work/`.

- **Shots**: A shot is defined as an attempt to score a goal by kicking the ball towards the opponent's goal. The value is the total number of shots a team makes, both on and off target.
- **Shots on target**: The total number of shots on target a team makes. A shot is considered 'on target' when the ball goes in regardless of intent, or when the ball is heading towards goal and would have gone in if it were not for the intervention of a goalkeeper or defender. A higher number of shots on target can indicate that a team

[19] Rules prescribe the match to be continued at the minute it was abandoned, however without public present.

is creating more goal scoring opportunities.

- **Corners**: The total number of corners a team earns. A corner is a set piece kick taken from the corner of the field, awarded to the attacking team when the ball goes out of bounds over the goal line after being touched last by a defending player. A corner kick can result in a scoring opportunity for the attacking team.
- **Fouls**: The total number of fouls committed by a team. A foul is an illegal action committed by a player against an opponent or opposing team, assessed by the referee. When a foul is committed, the referee will award a free kick or penalty kick to the opposing team, depending on the severity of the foul and where it occurred on the field.
- **Yellow cards**: The total number of yellow cards shown to a team. A yellow card is a disciplinary action taken by the referee against a player for committing a minor infraction or unsportsmanlike conduct during the game. It serves as a warning to the player. When a player receives two yellow cards, he must leave the game and can not be replaced by a substitute, forcing his team to play with a fewer player.
- **Red cards**: The total number of red cards shown to a team. A red card is shown to a player who has committed a serious foul or misconduct during the game, as judged by the referee. The player must leave the pitch and can not be replaced, forcing his team to player with a fewer player.
- **Rating**: An estimate of a team's strength. FiveThirtyEight's team strength rating (SPI) is their best estimate of a team's overall strength. It is based on a team's previous results and the market value as calculated by Transfermarkt. During a season, the ratings are adjusted every match based on the teams performance during the match and the strength of the opponent. Since football is a low-scoring sport, it is also updated on other metrics[20] to mitigate the randomness of football. Values are between 0 to 100 and are comparable both within the league as across football leagues.
- **Importance**: How much the outcome of the match matters to the teams involved. The importance quantifies how much the outcome of a match will change each team's statistical outlook on the season. The importance is dependent on the team, as different teams play for different objectives: for example the title, qualifications for

[20] Namely adjusted goals, shot-based expected goals and non-shot expected goals.

international UEFA competitions or not to be relegated. The situation for each team in the league is taken into account, as probabilities of each team to reach a goal are calculated depending on the outcome of the game. Values are between 0 to 100 and are again comparable within the league and across different leagues.

3.5 Data cleaning and manipulating

The initial data frame collected from `https://www.football-data.co.uk` includes also a variable *Div*, which indicates the division in which the match is played. This value is the same for all observations in the data set and is therefore removed. Also, there are columns which include the betting odds for the teams in the data set prior to the match. Since this paper does not need them, all these columns are removed. The initial data set is expanded with the number of attendants per match, a dummy variable indicating the surface of the pitch (0 = normal grass, 1 = artificial grass) and a dummy variable indicating whether the home or away team was promoted (0 = no change, 1 = promoted). The attendance value for each match was obtained via match reports by the NOS, which is a Dutch national broadcaster of (sporting) news[21]. The data set was then merged with the FiveThirtyEight data on the variables Date, Home Team, Away Team and Season. For the FiveThirtyEight data set, there were 161 observations with missing values for the Importance of the match for both teams in the 2017-2018 season. Because this variable is important for the eventual analysis, and we didn't want to exclude the observations (because then also the return fixtures had to be excluded to keep the home/away balance), we imputed this variable and its missing values with the mean of the Importance columns for season 2017-2018, calculated without the missing values. We checked the median value of the variable before and after imputing, and it did not change a lot.

The merged data set was expanded with the amount of points for both home and away team as well as the goal difference of the match for both home and away team and the weekday the match took place, which were calculated using already present variables. We also added the maximum capacity of each team's stadium to be able to calculate the Occupancy Rate of each match. Stadium capacity for the stadiums of each club were obtained via the site Transfermarkt[22].

[21] See `https://nos.nl/artikel/2475118-feyenoord-verslaat-go-ahead-en-is-na-zes-jaar-weer-kampioen-van-nederland` for an example. The attendance can be found at the bottom of the page.

[22] Links are `https://www.transfermarkt.nl/eredivisie/stadien/wettbewerb/NL1` and `https://www.transfermarkt.nl/keuken-kampioen-divisie/stadien/wettbewerb/NL2`.

Each match was then given a unique game-ID, and the data set was duplicated so that each match appeared twice in the data set, one time coded from the perspective of the home team, and one time from the perspective of the away team, as will be explained in the methodology section. For each match, we indicated the Venue and in which COVID period the match was played. A Home match was coded as 0 (being the reference point), and an Away match was coded as 1. For the COVID periods, Before COVID is the reference point (coded as 0). During COVID with Restricted Attendance is coded as 1, During COVID Without Attendance as 2 and After COVID as 3.

We constructed three other variables. The first two are the rating difference, being the difference between the rating of the home team and away team on the date of the match and the importance difference, being the difference in importance of the match for the home and away team. Both differences are standardized, which means that the mean is zero and the standard deviation is one, to make the differences more interpretable when adding them to the multilevel models. Also, based on McCarrick et al. (2021), we compute a variable called *Dominance*, as an indicator for the attacking dominance of a team. This variable is the standardized mean of the corners, shots and shots on target of a team. The reasons for computing these variables is explained in the Methodology section.

3.6 Reliability of the data

It is expected that the data from `football-data.co.uk` is reliable, since various research papers discussing subjects in the football world have used the website (Razali et al., 2017; Zebari et al., 2021). Among others, one of the leading papers about home advantage in the COVID-19 period by Wunderlich et al. (2021). The FiveThirtyEight data is also used by a research paper (McCarrick et al., 2021), indicating its reliability. The same goes for the data from Transfermarkt, which is a website more often used in scientific research (Correia-Oliveira and Andrade-Souza, 2022; Fischer and Haucap, 2021)

Regarding the data from the NOS, it is also assumed that this data is reliable. The institution has a statutory obligation to publish news and sports programs for the Dutch public television channels and Dutch public radio services. NOS is funded by the Dutch government.

4 Methods and methodology

The methodology section is split up in various subsections, since we focus on two aspects of home advantage. On the one hand, we focus on the performance of the home team measured in various game outcomes and in-game outcomes, and on the other hand we focus on the performance of referees. The performance of a team is measured in the game outcomes goals and points. The performance of the home team with respect to in-game statistics is measured by an indicator called dominance, which is the standardized mean of the amount of corners, shots and shots on target of a team. The referees' performance is measured by the amount of fouls, yellow cards and red cards. All regression results, statistical tests, figures and tables were compiled using R programming software (R Core Team, 2022). For the regression results, we used the *glmer* package, and for the displaying of the regression results, we used the *sjPlot* package.

4.1 Home advantage reflected in home team performance

Home advantage was earlier quantified as the number of points (or goals) won at home as a percentage of the total points (or goals) home and away (Pollard, 1986). This way of quantifying HA has also been used and validated earlier by Goumas (2013), and tends to work well for a full season of play, where each team plays each other home and away. In our data set we also have two seasons where we don't have all games played so that not each team has played each other both home and away. However, we have accounted for the home and away balance by removing matches where this wasn't the case, to make sure that every team in the eventual data set does have played each other home and away for each season in the data.

As explained in the methods section in the Literature Review, we use multilevel modeling techniques to answer our research questions. The opportunity of adding possible confounding factors of home advantage is an important reason why we use this technique. Multilevel modeling is a regression approach where the individual matches are nested within individual teams and which allows for the inclusion of additional factors, including possible confounds. We are particularly interested in the factors COVID (Before COVID, During COVID with Restricted Attendance, During COVID Without Attendance and After Covid) and Venue (Home and Away), and most importantly, their interaction. If home advantage is influenced by attendance, we would expect to see a difference in values for the factors COVID, Venue and their interaction in the regressions with points accrued,

goals scored and dominance as dependent variables. Our main interest is therefore the interaction among the factors.

Also, in our analysis, we are comparing various different COVID periods where crowds were or were not present, which are spanning across multiple different seasons. Therefore, it is important to account for the schedule difficulty and importance of games of teams across COVID periods. Some teams could have had vastly different schedules in the four different COVID periods, playing for example against teams with more or less quality, which would bias our home advantage comparison between the four periods. To control for possible difference in playing schedule and in the importance of matches for the different attendance periods, we use the difference between home and away teams in the variables rating and importance, which are explained in detail in the subsection 3.4 of the Data. Both variables are added to the multilevel model as covariates. We are especially interested in their interaction with COVID and Venue factors, since a significant interaction would imply different schedule importance and different opponents according to team rating across COVID periods. We have also added variables such as the type of grass, whether the home team was promoted or not, and the Occupancy Rate to multiple different regressions. However, according to their p-value (using a 5% significance level), neither of the three variables were significant in the regressions we ran. Also their interactions with the Venue or COVID variables were not significant for the same significance level. Therefore, we have chosen to omit them from the models and not use the variables. The results of these variables are hence not included in the regression output.

Our final models for both points and goals uses all individual games, where each game is coded twice. As variables in the model, we include Venue (Home or Away), COVID Period (Before COVID, During COVID with Restricted Attendance, During COVID Without Attendance and After COVID), the interaction between both variables, the standardized rating difference, the standardized importance difference, the interaction between rating difference and Venue, and the interaction between rating difference and importance difference. The same models were run for points and goals.

Before executing the models, we first checked how the variables were distributed. We expected a Poisson distribution, since both variables are discrete occurrences which are rarely normally distributed. We examined a density plot, a QQ-plot and executed a Shapiro-Wilk test for both goals and points. An example of the plots for goals can be seen in Figures 10 and 11 in the Appendix. All results confirmed our expectations

and indicated that the variables were not normally distributed, but were in fact count variables, hence we used a Poisson distribution for both points and goals.

Regarding home team performance measured by in-game match statistics such as corners, shots and shots on target, we first checked the correlation between the three variables, using a Spearman correlation test. All correlations were relatively high, namely between 0.4 and 0.75. Therefore, and also based on research by McCarrick et al (2021), we made one variable called dominance, which is the mean of corners, shots and shots on target of a team. These indicator variables were however normally distributed (which we checked via the same plots and test as for the variables goals and points), hence we used a Gaussian distribution for the variable dominance. We standardized the values for the variable dominance. The variable is used as a proxy for the attacking dominance of a team as it indicates how much a team dominates its opponents. For the variable dominance, we ran the same multilevel model as for goals and points.

4.2 Home advantage reflected in referees' performance

When measuring the home advantage in referees' performance, we use fouls, yellow cards and red cards as indicators of referees' performance. Therefore, these variables are used as the dependent variables in three separate regressions. For fouls, we use a linear multilevel model, since this variable is normally distributed. We again tested this in the same way as for the variables in the previous subsection. The variables yellow cards and red cards are also examined but they are believed to be count variables, just as the variables goals and points, and are therefore modelled using a Poisson distribution. We ran the models for the referees' performance with the same variables as previously used in the models for home team performance, but we also added dominance as an indicator for the attacking tendencies of the home team. We do this because it is shown by Goumas (2014a) that the more dominant a team is, the more it is going to get fouled and therefore earn yellow and red cards for the opposing team. Therefore, when researching the performance of referees, we must control for attacking tendencies of a team.

4.3 Interpretation of estimates

When performing a Poisson regression, the expected value of the dependent variable E(Y) is related to the independent variables by the log link function. Therefore, the coefficients of a Poisson regression have to be interpreted differently as opposed to a regular linear

regression. For a Poisson regression, a one-unit increase in a predictor variable x_n, yields a factor equal to $\exp(\beta_n)$, which multiplied by the dependent variable, indicates the expected value of the dependent value due to a one-unit increase in x_n, holding everything else constant. For all regression results using Poisson regression (points, goals, yellow cards and red cards), we have provided incident rate ratios (IRR) instead of raw estimates, to make their interpretation easier. These IRR are already the exponentiated β's. If we look at Table 1 in the next section, for the points model, the IRR for Venue is 0.62. This means that the away team obtains 0.62 fewer points per game than the home team, since the home team is the reference category. Or else, for each point won by the home team, the away team wins 0.62 points, holding all other variables constant.

Furthermore, in the same model the IRR value for the interaction coefficient for Venue and During COVID Without Attendance is 1.44. This means that away teams during the COVID period without attendance gain 0.44 points per game compared to the same away teams in the Before COVID period, which is the reference period here. Therefore, the differences in home and away teams are 1.44 smaller, or 0.44 per full point, in the During COVID Without Attendance period than in the Before COVID period.

For fouls, we perform linear regression and therefore interpret the coefficients in Table 2 in the next section how we would normally do in a linear regression. For example 0.48 for the Venue variable means that on average, the away team commits 0.48 more fouls than the home side, holding all variables constant. The value of 0.31 for the interaction between Venue and COVID with Restricted Attendance gives information on the fewer amount of fouls away teams commit during the COVID period with Restricted Attendance as opposed to Before COVID, the reference category. Our last variable which is not modelled via Poisson regression is dominance. However, we already standardized the variable, which is why the estimates in Table 1 already represent the standardized estimates.

5 Results and Analysis

In this section, we evaluate the results of our models applying the methods and methodology explained in the previous section. We have divided the Results and Analysis section in two subsections. First, we evaluate our results for the variables we use to model team performance, namely points, goals and dominance. Secondly, we evaluate the results for the indicators of referees' performance, namely fouls, yellow cards and red cards. We use a 5% significance level to determine whether predictors are deemed to be of significant importance.

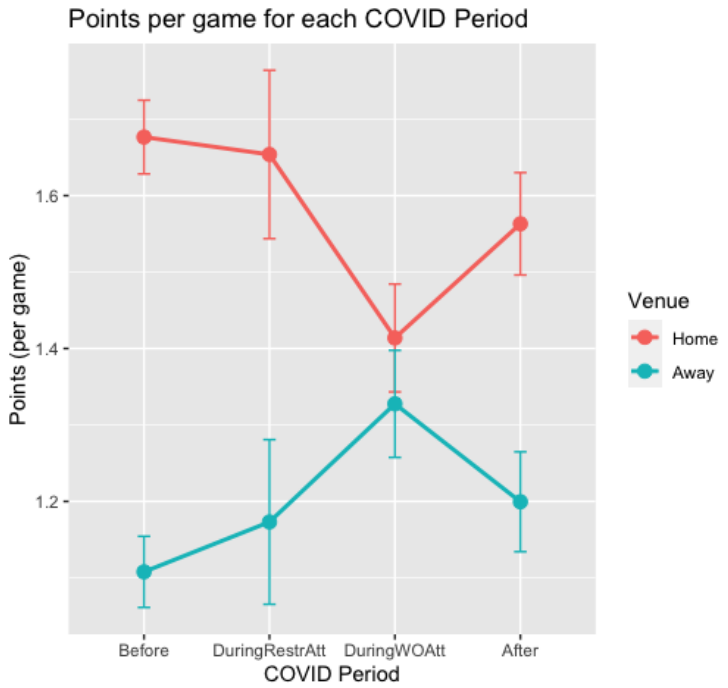
5.1 Team performance

We first looked at the change in points per game across all the data, separated per Venue (Home or Away) and split up for each COVID period. We checked the three separate seasons which belong to the Before COVID period to see whether there was already a trend Before COVID. This seasons indicated that the points per game for home teams increased from 2017-2018 to 2019-2020, and that the points per game for away teams only decreased in the same period. This indicates a growing home advantage before COVID. These separate seasons are not included in Figure 1, were we take the average value for points per game in the seasons belonging to the Before COVID period. Since we see points as the most important variable indicating home advantage, we have not looked at the separate seasons in the Before COVID period for other variables.

It can be seen in Figure 1 below that overall, on average more points are gathered by teams in home games, irrespective of the COVID period. However, a small drop in points per home game can be seen for the period Before COVID to During COVID with Restricted Attendance, whereas the mean points per game for away teams slightly increased. The difference between the mean points gathered at home and away then decreases more when looking at the period During COVID Without Attendance, where the difference between mean points per game for home teams and away teams is the smallest. The mean points per game decreased by 0.27 for home teams as opposed to Before COVID, and increased by 0.21 per game for away teams. The total difference in points per game between home and away teams has decreased by almost half a point per game, when comparing the period Before COVID to the period During COVID Without Attendance. This is a drop of nearly 85% in difference in mean points per game between home and away teams. After COVID, the figures are slowly returning back to Before COVID values for both home and

away teams. However, the smallest difference between home and away points per game was in the COVID period Without Attendance, indicating an influence of the absence of crowds on points per game.

Figure 1: Average points per game for each COVID period grouped by Venue.

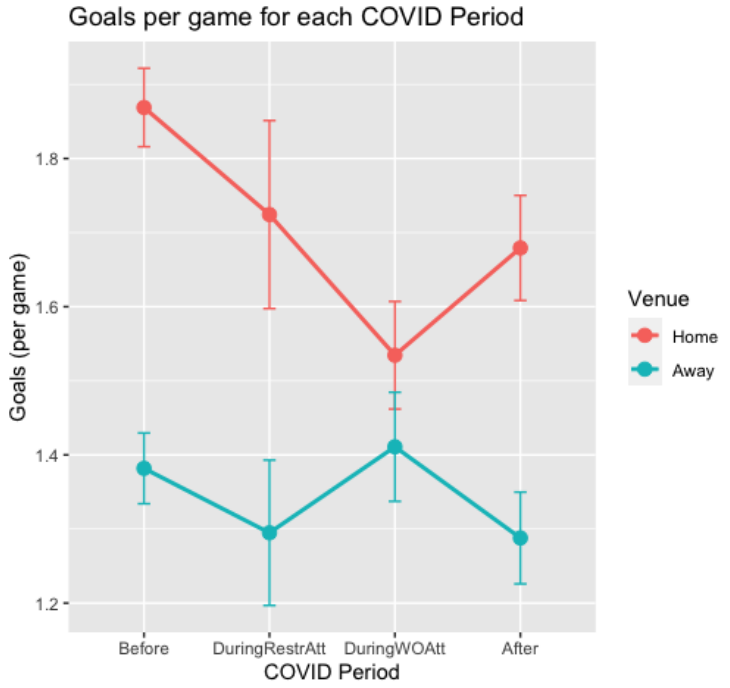


Note. The vertical lines around the mean values of points per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

When looking at the regression results in Table 1, we see that the coefficients for the interaction terms between COVID and Venue are significant for all periods except the period with Restricted Attendance. The IRR is 1.44 for the COVID period Without Attendance, indicating that away teams during the COVID period Without Attendance gain 0.44 point per game compared to the same away teams in the period Before COVID. This indicates a big decrease in the home advantage, thus confirming results from inspecting Figure 1. Also, effect of an audience seems not to be driven by the difference in schedule of teams, or difference in importance between home and away teams in the different COVID periods. The teams with a higher rating and importance won more points, but this was constant for home and away matches. Furthermore, the coefficient for the interaction between the Venue and the period After COVID was also positive (IRR = 1.17) and significant, indicating that away teams also accumulated more points per game in the period After COVID as opposed to the period Before COVID. Based on this we conclude that home

advantage measured in points did also not return immediately After COVID.

Figure 2: Average goals per game for each COVID period grouped by Venue.



Note. The vertical lines around the mean values of goals per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

When we look at the above Figure 2, showing the mean goals scored by home and away teams for the different COVID periods, we can see that goals follow a similar path as points. Goals per game for the home team first slightly dropped when attendance is still present, and then drop more when there is no attendance, going from 1.87 per match to 1.53, a drop of 0.34 per match for home teams. A pattern which follows that of points (see Figure 1). For away teams, the goals not entirely follow the points, seeing as the goals per game first decrease During COVID with Restricted Attendance, and then increase when the attendance is not allowed, indicating a drop in home advantage. This is because the mean goals per game increase for away teams, and the difference between average home and away goals is at its smallest in the period without attendance. However, the value for mean goals per game for away teams is quite similar to the Before COVID value, it is therefore mainly the mean goals scored per game for home teams that has dropped. This indicates a decrease in home advantage measured in goals. The mean goals for away teams decrease again with the return of fans to the stadiums After COVID and are even below the value Before COVID, whereas the mean goals per game for home teams

increase, indicating a positive influence of the returning fans on home team performance measured in goals.

Table 1: Regression models for points gained, goals scored and dominance across the Eredivisie seasons 2017-2018 to 2022-2023 for every COVID period and its interaction with Venue.

-	**Points**			**Goals**			**Dominance**		
Predictors	*IRR*	*SE*	*p*	*IRR*	*SE*	*p*	*std. β*	*SE*	*p*
(Intercept)	1.57	0.05	**<0.001**	1.75	0.05	**<0.001**	0.33	0.04	**<0.001**
Venue (Away)	0.62	0.03	**<0.001**	0.74	0.03	**<0.001**	-0.54	0.04	**<0.001**
COVID (DuringRestrAtt)	0.98	0.07	0.76	0.92	0.06	0.20	-0.17	0.07	**0.013**
COVID (DuringWOAtt)	0.84	0.04	**<0.001**	0.83	0.04	**<0.001**	-0.37	0.05	**<0.001**
COVID (After)	0.93	0.05	0.149	0.90	0.04	**0.033**	-0.10	0.05	**0.042**
Rating difference	1.48	0.03	**<0.001**	1.42	0.02	**<0.001**	0.53	0.02	**<0.001**
Importance difference	1.08	0.02	**<0.001**	1.03	0.01	**0.030**	0.01	0.02	0.648
Venue (Away) * COVID (DuringRestrAtt)	1.13	0.12	0.242	1.05	0.11	0.632	0.01	0.10	0.947
Venue (Away) * COVID (DuringWOAtt)	1.44	0.11	**<0.001**	1.24	0.09	**<0.003**	0.30	0.07	**<0.001**
Venue (Away) * COVID (After)	1.17	0.09	**0.038**	1.04	0.07	0.618	0.05	0.07	0.449
Venue (Away) * Rating difference	1.13	0.03	**<0.001**				-0.07	0.03	**0.015**
Rating difference * Importance difference	0.94	0.01	**<0.001**				0.02	0.01	0.150
Random effects									
σ^2	0.60			0.54			0.60		
τ_{00}	0.00_{Team}			0.00_{Team}			0.02_{Team}		
ICC	0.00						0.03		
N	26_{Team}			26_{Team}			26_{Team}		
Marginal R^2 / Conditional R^2	0.313 / 0.313			0.213 / NA			0.339 / 0.357		

Note. IRR = Incident Rate Ratios; Venue is coded 0 for Home and 1 for Away; COVID is coded as 0 for Before Covid, 1 for During Covid with Restricted Attendance, 2 for During COVID Without Attendance and 3 for After COVID. τ_{00} indicates the intercept variance, and ICC is the Intraclass Correlation Coefficient.

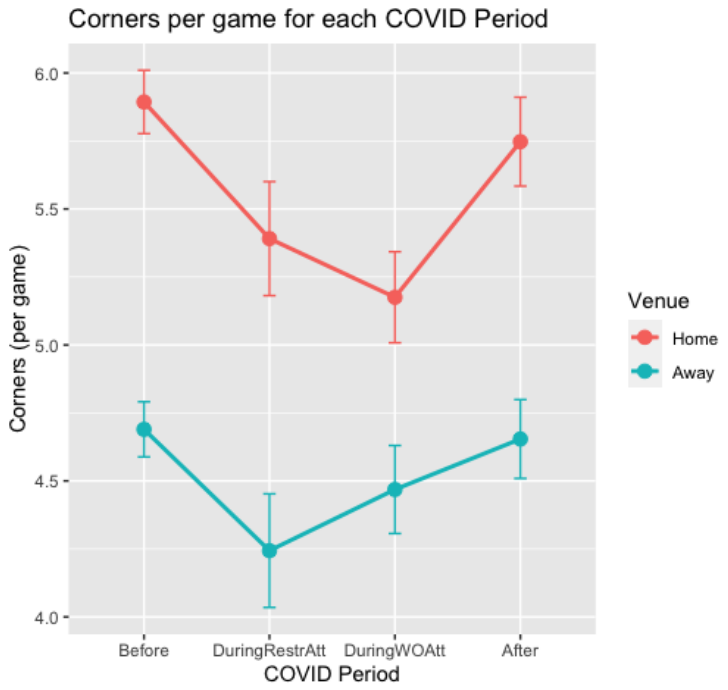
If we look at the regression for goals in Table 1, the IRR seem to follow the same pattern as with points. Again, the IRR for the interaction between Venue and the COVID period without fans is highly significant and has a value of 1.24, indicating that the way teams

gain 0.24 goals per game in the period without attendance compared to the period Before COVID with attendance present. This is however the only interaction between COVID period and Venue which is significant as indicated by the model. The interaction between Venue and the period After COVID is not significant, therefore we can not assume that the away teams performed better After COVID as opposed to before COVID, and can also not assume that the drop in home advantage measured in goals persisted After COVID. Both rating difference and importance difference were again significant and positive predictors, however any interactions were not present or significant. In other words, the effect of crowd on home advantage seems not to be influenced by differing schedules in COVID periods.

Next, we take a look at the indicators of team dominance, being corners, shots and shots on target, as well as their proxy dominance, and the change per game for each COVID period, grouped by either home or away venue. These can be seen in Figures 3, 4 and 5 depicted on the next page.

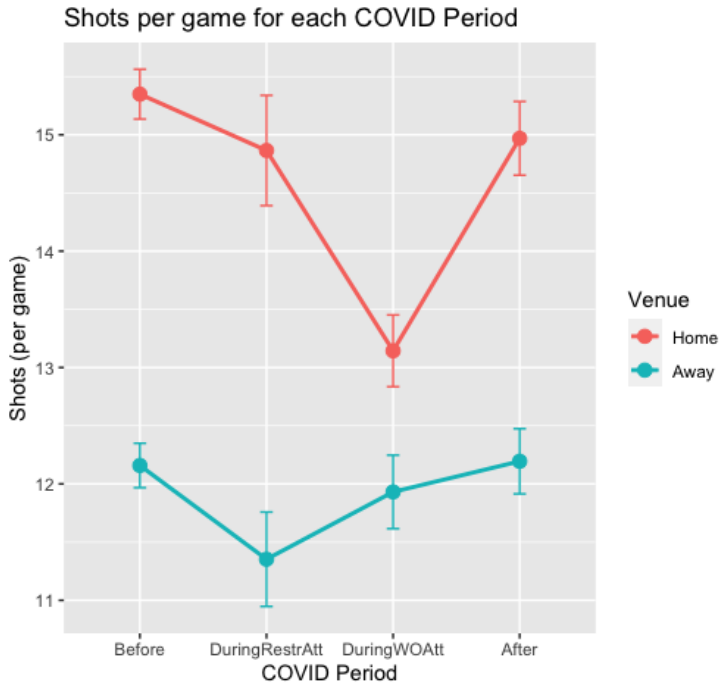
It can be seen that the home team always seems more dominant, having a higher mean value per game for each dominance indicator for every COVID period, as well as a higher value for the proxy of dominance overall. However, the mean difference is least pronounced in the COVID period Without Attendance, confirming the trend of the decrease of dominance for home teams without their fans. The indicators all follow the same pattern. For home teams, that means that the value first slightly decreases when going from Before COVID to COVID with Restricted Attendance, and further decreases going into the period During COVID Without Attendance. For all three indicators, their low point is in the COVID period Without Attendance, with a mean loss per game as opposed to the Before COVID period of 0.71 in corners, 2.21 in shots, and 0,76 in shots on target. With the return of the public After COVID all values increase again, but not to the same value as before the pandemic. For away teams, the values first decrease in the period with Restricted Attendance as opposed to Before COVID. Then, the mean value for all indicators rises for the period During COVID Without Attendance, indicating a drop in home advantage considering also the value of the home teams decreasing. With the return of the public, the mean values for away teams for corners and shots further increase slightly, as well as the dominance proxy, whereas the value for shots on target is the only indicator decreasing After COVID.

Figure 3: Average corners per game for each COVID Period grouped by Venue.



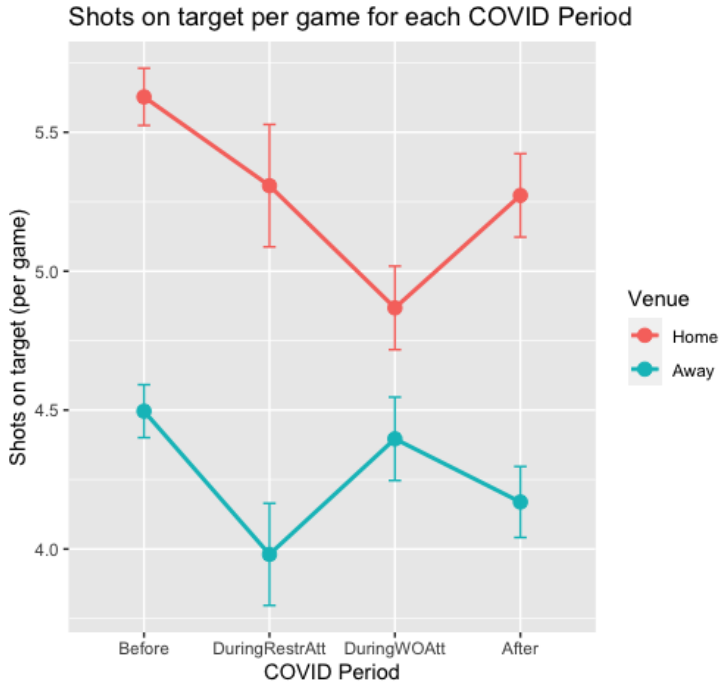
Note. The vertical lines around the mean values of corners per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

Figure 4: Average shots per game for each COVID Period group by Venue.



Note. The vertical lines around the mean values of shots per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

Figure 5: Average shots on target per game for each COVID Period group by Venue.



Note. The vertical lines around the mean values of shots on target per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

Figure 6: Average value for dominance for each COVID Period group by Venue

Note. The vertical lines around the mean values of dominance represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

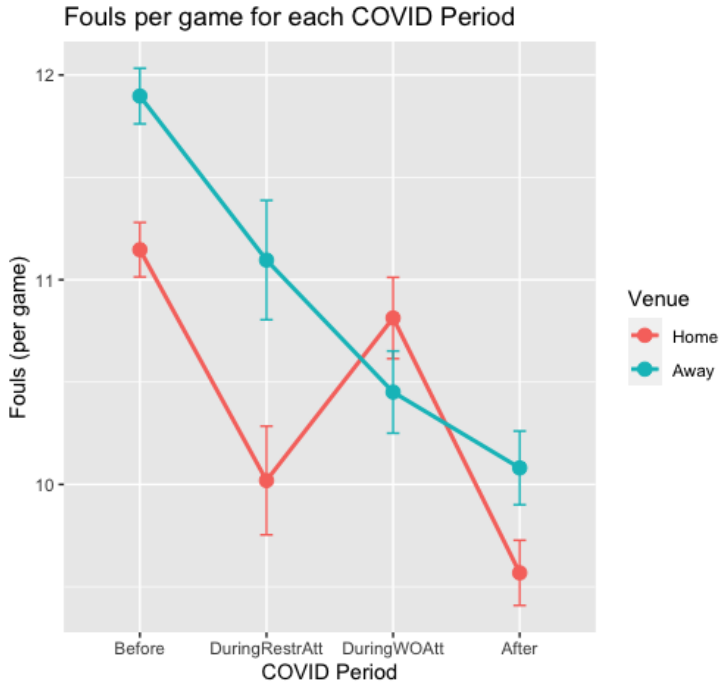
The multilevel modeling regression model on the dominance variable confirmed the negative effect of the absence of attendance for home teams, seeing that the Venue and COVID interaction was significant only for the COVID period Without Attendance, as depicted in Table 1. Away teams obtained a 0.30 higher score for dominance in games during the period Without Attendance than Before COVID, confirming the trends seen in the previous Figures 3, 4, 5 and 6. Also, the interaction between Venue and the After COVID period was not significant, therefore we can not assume that the better performance of away teams as opposed to Before COVID persisted in the period After COVID. However, the interaction between Venue and the rating difference was of negative impact and was significant, indicating a negative impact of the schedule difference on the dominance. Therefore, this needs to be taken into consideration when interpreting the results regarding dominance of teams.

5.2 Referees' performance

Looking at Figures 7, 8 and 9 below, we can see the mean values of fouls, yellow cards and red cards per game for each COVID period grouped by Venue. For fouls, the value of fouls per game for the home team is always lower than the away team, except for the value in the period During COVID Without Attendance. The sudden drop in fouls of the away team in the period Without Attendance, during which the mean amount of fouls per game are almost equal for home and away teams, indicates a possible referee bias. Also, the home team gets less cards in all periods where attendance is present. In the period where attendance is not allowed, the mean yellow cards per home and away team are equal, and therefore follow a similar pattern to the amount of fouls per game. Based on these visualizations, the referee bias seems to be diluted in the period Without Attendance. For red cards, the occurrence of events is smaller, hence why the differences are also smaller in mean value, and why the bars indicating the mean plus/minus the standard error are wider. The red cards however do follow quite a different pattern to the yellow cards. In the period Before COVID and COVID Without Attendance, mean value for red cards was higher for away teams. With the restriction of limited attendance, home teams received more red cards, which is explainable because the effect of public is expected to decrease. However, it is counter intuitive that away teams received on average more red cards in the period without attendance, because there was no public to influence the referee. After COVID, with the return of public, values are equal, which indicates no

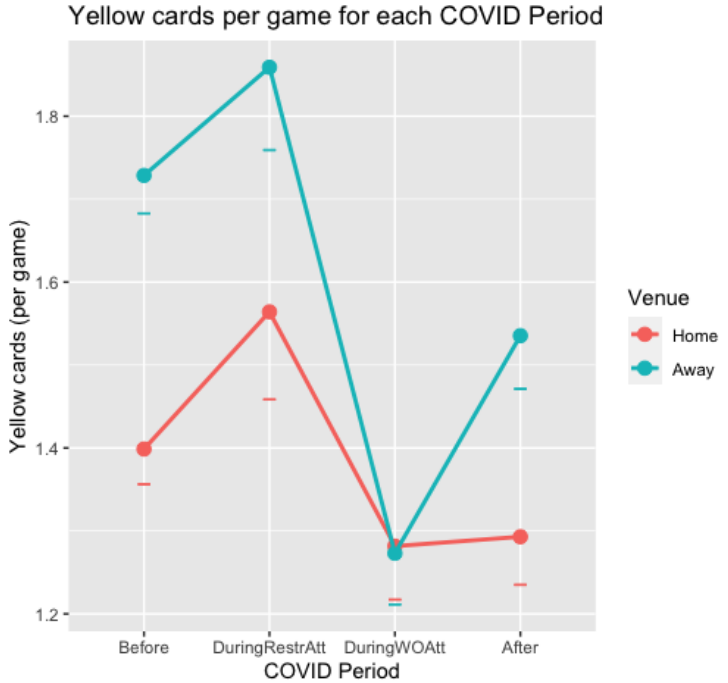
referee bias regarding red cards.

Figure 7: Average fouls per game for each COVID period group by Venue.



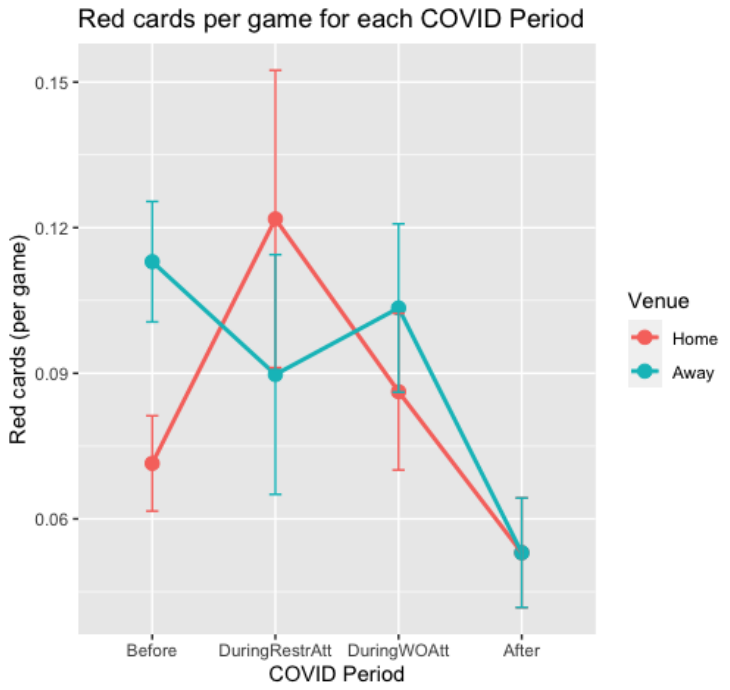
Note. The vertical lines around the mean values of fouls per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

Figure 8: Average yellow cards per game for each COVID period group by Venue.



Note. The vertical lines around the mean values of yellow cards per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

Figure 9: Average red cards per game for each COVID Period group by Venue.



Note. The vertical lines around the mean values of red cards per game represent the mean plus one standard error (maximum value) and minus one standard error (minimum value).

To examine the results from the figures above, multilevel regression analysis with fouls (linear), yellow and red cards as dependent variables are run, using Venue and COVID, and their interactions as predictors. The results confirm the descriptive results, where the interaction between Venue and the COVID period Without Attendance is significant for fouls, and has a value of -1.12. Hence, away teams get awarded 1.12 less fouls during the period Without Attendance than Before COVID. The value was also significant for the yellow cards, with an IRR of 0.80 indicating the away teams receive 0.80 times the yellow cards in the COVID period Without Attendance compared to the period Before COVID. For fouls and yellow cards however, the rating difference interaction with the Venue was also significant, indicating that differences in schedule play a role. For red cards we recorded only significant effects for the interaction between Venue and COVID with Restricted Attendance, instead of the period Without Attendance. This tells us that for red cards, during the period with Restricted Attendance, away teams received 0.46 less red cards than Before COVID.

Given the above results, one could think that the presence of the fans biases the decisions of referees against the visiting team, and that this was less in the COVID period With-

out Attendance for yellow cards and fouls. However, we have, in the above conducted regressions, not yet accounted for the attacking dominance of teams. Passive teams are not shy to adopt a defensive approach and are indicated by low values for dominance, since they record low amounts of shots (on target) and corners. They tend to foul more and receive more warnings as they play against dominant teams. The referees' decisions could therefore be a consequence of attacking tendencies of a team. If we take a look at Figure 6, it can be seen that the mean dominance value is always higher for home teams and decreases without the support of fans. Therefore, it is important to account for dominance by including the indicator in the model for referees' performance.

We include the dominance indicator variable in our multilevel regression models, as is shown in Table 2. We observe that the interactions between Venue and COVID period Without Attendance for fouls (Est = -0.96, SE = 0.33, p = 0.003) and yellow cards (IRR = 0.83, SE = 0.07, p = 0.015) remain significant. The effect sizes are only slightly weakened. These values therefore still indicate that in the COVID period Without Attendance, away teams committed less fouls and received less yellow cards as opposed to the period Before COVID. However, from the above results, we can not conclude that this decrease of referee bias persisted After COVID with the return of Attendance, seeing as the COVID and Venue interaction was only significant for the period Without Attendance. For fouls and yellow cards the rating difference interaction with the Venue was not significant anymore, which does not indicate differences in schedule for COVID periods.

For red cards, the interaction between Venue and COVID with Restricted Attendance also remains significant, and the effect size remains the same (IRR = 0.46, SE = 0.18, p = 0.049). This therefore still indicates that away teams received less red cards during the period with Restricted Attendance. Seeing as the effect sizes stayed almost the same when controlling for team dominance, this does therefore not seem to have an effect on the coefficients in the regressions on referees' performance. This means that, in our models, accounting for dominance is not presumed to be necessary when measuring the effect on referees' performance for various predictor variables.

Table 2: Regression models for referees' decisions. Fouls, yellow cards and red cards across the Eredivisie seasons 2017-2018 to 2022-2023 for every COVID Period and its interaction with Venue.

-	**Fouls**			**Yellow cards**			**Red cards**		
Predictors	*Estimates*	*SE*	*p*	*IRR*	*SE*	*p*	*IRR*	*SE*	*p*
(Intercept)	11.38	0.16	**<0.001**	1.43	0.05	**<0.001**	0.08	0.01	**<0.001**
Venue (Away)	0.48	0.19	**0.011**	1.18	0.05	**<0.001**	1.12	0.20	0.536
COVID (DuringRestrAtt)	-1.26	0.32	**<0.001**	1.11	0.08	0.148	1.57	0.42	0.091
COVID (DuringWOAtt	-0.59	0.23	**0.011**	0.89	0.05	**0.032**	0.95	0.22	0.806
COVID (After)	-1.70	0.23	**<0.001**	0.92	0.05	0.101	0.69	0.18	0.149
Dominance	-0.52	0.08	**<0.001**	0.91	0.02	**<0.001**	0.49	0.04	**<0.001**
Rating difference	-0.13	0.11	0.214	0.94	0.02	**0.010**	1.19	0.09	**0.021**
Importance difference	0.03	0.07	0.614						
Venue (Away) * COVID (DuringRestrAtt)	0.31	0.44	0.485	0.95	0.09	0.628	0.46	0.18	**0.049**
Venue (Away) * COVID (DuringWOAtt)	-0.96	0.33	**0.003**	0.83	0.07	**0.015**	0.91	0.28	0.762
Venue (Away) * COVID (After)	-0.21	0.31	0.504	0.97	0.07	0.643	0.66	0.23	0.241
Rating difference * Importance difference	-0.13	0.05	**0.010**						
Venue (Away) * Rating difference	0.24	0.12	0.055	1.05	0.03	0.068			
Random effects									
σ^2	12.67			0.52			2.55		
τ_{00}	0.18_{Team}			0.00_{Team}			0.00_{Team}		
ICC	0.01			0.00					
N	26_{Team}			26_{Team}			26_{Team}		
Marginal R^2 / Conditional R^2	0.068 / 0.082			0.055 / 0.058			0.153 / NA		

Note. IRR = Incident Rate Ratios; Venue is coded 0 for Home and 1 for Away; COVID is coded as 0 for Before Covid, 1 for During Covid with Restricted Attendance, 2 for During COVID Without Attendance and 3 for After COVID. τ_{00} indicates the intercept variance, and ICC is the Intraclass Correlation Coefficient.

6 Conclusions and Discussion

This section contains a summary of the research conducted in this thesis, which will be used to try to answer the initial research questions posed in the introduction. It is then followed by a discussion. Lastly, we provide managerial and academic implications, list limitations for our research and provide recommendations for future research.

6.1 Summary

In the summary, we look back on our main findings and try to use them to answer our earlier research question as well as possible, with the central research question being of main interest.

The first three sub questions addressed the home advantage by looking at points, goals and an indicator for dominance. For the first sub question, we looked at our results in Table 1, which indicate that away teams accumulated significantly more points in the COVID period Without Attendance than Before COVID, with 0.44 more points per game, and a significant interaction between the period Without Attendance and the Venue. This, along with Figure 1, leads us to believe that home advantage has decreased in the period without attendance, since away teams achieved significantly more points. Also, After COVID, away teams still performed better than compared to Before COVID, seeing as the interaction between the period Without Attendance and the Venue was also significant. This indicates that the home advantage did not immediately return with the return of attendance to the stadiums.

The second sub question had to do with the measure of goals as home advantage indicator. Looking at our regression results in Table 1, we see that, just as with points, away teams score on average more goals (0.24) in the COVID period Without Attendance as compared to the period Before COVID. This was the only COVID period of which the interaction with Venue was significant. However, from this result and looking at Figure 2 for mean goals per game, we conclude that the home advantage as measured in goals decreased in the COVID period Without Attendance as opposed to Before COVID. We have however no indication to believe that this drop in home advantage continued into the period After COVID, by looking at our regression results.

The third sub question had to do with the team dominance in a match, as indicated by the dominance variable, which was the standardized mean of the amount of corners, shots

and shots on target of a team. For this variable, the interaction between the Venue and the period During COVID Without Attendance was significant, indicating a 0.3 higher dominance of away teams in the period Without Attendance than Before COVID (see regression results in Table 1). As we look at Figure 6 and take into account the regression results, we see a decreasing dominance for home teams in the period Without Attendance, indicating a drop in home advantage. We can however not assume this decrease in home advantage measured in dominance persisted After COVID, as the interaction between Venue and the period After COVID was not significant. Also, for the dominance regression the interaction between the Venue and the rating difference is significant, which could indicate a different schedule for home and away teams. Therefore, we must take this into account when interpreting our results.

The last two sub questions were about the referees' performance, and were measured by the variables fouls, yellow cards and red cards, which were used as dependent variables. Results of the regression can be seen in Table 2. For both fouls and yellow cards, there was only a significant interaction between the Venue and the COVID period Without Attendance. Both indicated that the value for away teams in fouls (Est. = -0.96) and yellow cards (IRR = 0.83) was lower in this period Without Attendance than in the period Before COVID, indicating a decrease in referee bias measured in fouls and yellow cards. However, because there were no significant interactions between Venue and the period After COVID for both fouls and yellow cards, we can not conclude that the drop in both measures persisted for the period After COVID, and can therefore also not conclude a persisting decrease in referee bias with the return of the attendance. It could be that the return of public has an effect of the decisions of referees and that the referee bias returns. Figures 7 and 8 do indicate a lower average value for both fouls and yellow cards for home teams as opposed to away teams in the period After COVID. For red cards, the only significant interaction effect was between Venue and the COVID period with Restricted Attendance, where away teams received less red cards (IRR = 0.46). When accounting for the value of dominance, effect sizes of our regression coefficients were not weakened, meaning that accounting for dominance is not presumed to be necessary. Concluding, we have some evidence that the referee bias decreased in the period During COVID Without Attendance, measured by fouls and yellow cards. However, we have no evidence that the referee bias stays the same in the period After COVID.

Finally, we look back on our main research question, which was as follows:

How has the absence of fans in football stadiums as a result of the COVID-19 pandemic influenced the home advantage and referee bias in the Dutch Eredivisie in the seasons 2017-2018 up to and including 2022-2023?

Using the answers to our sub research questions, we can answer also the main research question of this thesis. To conclude, the results indicate that for points and goals, the values point to a decrease of home advantage in the Eredivisie when attendance is absent. Away teams gain more points and score more goals away from home in the period During COVID Without Attendance than Before COVID. For home advantage measured in points, we also conclude that the home advantage did not immediately return with the return of fans After COVID. For goals, we have however no indication to believe that the drop in home advantage continued. Also, the dominance of home teams seemed to decrease in the COVID period Without Attendance, which could indicate a drop in home advantage during this period. We could not conclude, based on our results, that the drop in home advantage measured in dominance persisted After COVID. Also, due to the significant interaction between the venue and rating difference, the results for the dominance indicator could also be due to the difference in schedule.

For the referee bias, we observed a decrease in fouls and yellow cards for away teams in the COVID period Without Attendance compared to the period Before COVID. This results indicate a decrease in referee bias. However, based on our results we could not conclude that this decrease persisted after COVID with the return of attendance. The referee bias measured in red cards was only lower in the COVID period with Restricted Attendance, indicating no persisted drop of referee bias in red cards After COVID. Also, accounting for the dominance of teams did not result in different results for our regressions on referees' performance.

6.2 Discussion

In the discussion, we compare the findings of our final research to the state of the literature we have outlined in the Literature Review, and see to which extent our findings confirm the literature or differ from it.

The results we found on the home advantage measured in points and goals are in line with our findings in the Literature Review, which indicate a drop in home advantage in the COVID period Without Attendance being present at matches. Our results are therefore in line with results from Cueva (2020) and McCarrick (2021), who also both

found a significant drop in home advantage when attendance is absent. Furthermore, also Leitner et al. (2022) find a (strongly) reduced home advantage for most studies which research the absence of crowds. Our results further support the idea that a decrease in home advantage persisted in the season after the emerging of COVID, as shown also by (Sors et al., 2022). Also, our findings on decreasing dominance of home teams was in line with results from McCarrick et al. (2021). However, they found that the diminishing performance of the home teams without their fans is not a consequence of the unbalanced schedules in the different COVID periods, as opposed to our results, which indicate that the drop in dominance of home teams could also have been due to the difference in schedule. The possible explanation for this difference in findings is that the research of McCarrick et al. (2021) measures the dominance in a different way. Or, the fact that they use different European leagues for their research, as opposed to this research which only uses data on the Dutch Eredivisie. It could be that the quality of teams in the Eredivisie is differently distributed accross the division, making schedule differences of importance.

As for the referee bias, we found in the literature that this was almost diluted in the COVID period Without Attendance due to the absence of the influence of the public on the referees' decisions (Sors et al., 2022). We found almost the same results. For our regression, fouls and yellow cards seemed to be decreased in the period During COVID Without Attendance, indicating a drop in referee bias. However, for red cards, we only found a significant decrease for away teams for the period with restricted attendance. This difference in results in referee bias measured by red cards could be due to the sparse occurrences of red cards in the data set.

However, this study has been unable to demonstrate that when accounting for the value of dominance when measuring referees' decisions, effect sizes or our regression coefficients shrink. When accounting for the value of dominance, effect sizes of our regression coefficients were not weakened, indicating no controlling effect for the variable dominance when measuring referee bias. This is not in line with earlier research, as this is opposed to results by McCarrick et al. (2021) and Goumas (2014a). It could be due to the fact that we measure the dominance indicator via a different way than McCarrick et al. (2021).

6.3 Managerial implications

The implications of our research can especially be useful for the KNVB. As absence of attendance does have an impact on home advantage, the KNVB can use this finding to

punish clubs for the misconduct of their fans, by for example banning supporters from matches. This would be an effective measure when home teams misbehave, seeing as a match without fans could impact the results of the home team, and therefore do what the punishment is intended to do.

It was already stated earlier that the KNVB has imposed new rules in the 2022-2023 season after an incident where objects were thrown onto the field. This resulted in the abandonment of several matches either temporarily or definitively. Therefore, the results of this thesis are very relevant and could be used in the evaluation of the imposed measures. Also, the results of this thesis could help in the decision making process regarding future punishments.

6.4 Academic implications and contribution

With this research, we add to the existing literature with respect to research on the impact of crowds on home advantage and referee bias. We fill various research gaps. For example, we add to the literature on home advantage in the Dutch Eredivisie, which has not yet focused on the influence of attendance on home advantage. Furthermore, we exploit new areas of research. We take into account the seasons after 2019-2020, namely 2020-2021 and 2021-2022, which also dealt with restricted or absence of attendance, to expand COVID research which often only focuses on the 2019-2020 season. Lastly, we also use the most recent season, 2022-2023, to see whether the home advantage and referee bias changed or returned with the return of fans to the stadiums.

6.5 Limitations and areas of future research

The limitations for our research mostly have to do with limitations for our data set. For 161 observations in the 2017-2018 season we had missing values for importance. This was solved by imputing the values with the mean importance over the rest of the season, but it would have been better if we would have had the full results. Also, the 2019-2020 season had less games than all the other seasons since it was stopped earlier. Therefore, we had to delete observations to keep a home and away balance. This could however not have been prevented, since the season is just shorter than others. Furthermore, we are missing one game week from the 2022-2023 season, since it is still underway at the moment of writing. Because of that, we also deleted observations for this season. However, we expect that one game week with 9 matches (and the 9 matches that we omitted) have little to no

impact on the final results. When conducting this research on a later moment, we advice to include the last match day.

For future research, it would be interesting to see what the pattern in the data would be when more seasons after COVID are included. Now, we have as much data as possible, but in a few years, we can hopefully look back at a longer period without COVID. Then, it can be investigated more in depth how the home advantage has changed over a longer period of time, and how detrimental the few years of COVID really were. Because with our current research, we only have one and a half seasons of play where attendance is back at the normal level. More full seasons with a home and away balance might lead to different or more clear results.

Also, the introduction of the VAR may have introduced a new dimension to referee bias. When a referee makes decisions based on VAR footage, the footage is evaluated during the match right next to the field of play. Therefore, a referee could also be influenced by the public when making decisions based on the VAR. It would be interesting for future research to add this dimension to the concept of referee bias.

Bibliography

Barnett, V. and Hilditch, S. (1993). The effect of an artificial pitch surface on home team performance in football (soccer). *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 156(1):39–50.

Benz, L. S. and Lopez, M. J. (2021). Estimating the change in soccer's home advantage during the covid-19 pandemic using bivariate poisson regression. *AStA Advances in Statistical Analysis*, pages 1–28.

Bryson, A., Dolton, P., Reade, J. J., Schreyer, D., and Singleton, C. (2021). Causal effects of an absent crowd on performances and refereeing decisions during covid-19. *Economics Letters*, 198:109664.

Buraimo, B., Forrest, D., and Simmons, R. (2010). The 12th man?: refereeing bias in english and german soccer. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 173(2):431–449.

Clarke, S. R. and Norman, J. M. (1995). Home ground advantage of individual clubs in english soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 44(4):509–521.

Correia-Oliveira, C. R. and Andrade-Souza, V. A. (2022). Home advantage in soccer after the break due to covid-19 pandemic: does crowd support matter? *International Journal of Sport and Exercise Psychology*, 20(4):1245–1256.

Courneya, K. S. and Carron, A. V. (1992). The home advantage in sport competitions: a literature review. *Journal of Sport & Exercise Psychology*, 14(1).

Cueva, C. (2020). Animal spirits in the beautiful game. testing social pressure in professional football during the covid-19 lockdown. osf preprints. *Center for Open Science*.

Fischer, K. and Haucap, J. (2021). Does crowd support drive the home advantage in professional football? evidence from german ghost games during the covid-19 pandemic. *Journal of Sports Economics*, 22(8):982–1008.

Goumas, C. (2013). Modelling home advantage in sport: A new approach. *International Journal of Performance Analysis in Sport*, 13(2):428–439.

Goumas, C. (2014a). Home advantage and referee bias in european football. *European journal of sport science*, 14(sup1):S243–S249.

Goumas, C. (2014b). Home advantage in australian soccer. *Journal of Science and Medicine in Sport*, 17(1):119–123.

Leitner, M. C., Daumann, F., Follert, F., and Richlan, F. (2022). The cauldron has cooled down: a systematic literature review on home advantage in football during the covid-19 pandemic from a socio-economic and psychological perspective. *Management Review Quarterly*, pages 1–29.

Maimone, V. M. and Yasseri, T. (2021). Football is becoming more predictable; network analysis of 88 thousand matches in 11 major leagues. *Royal Society Open Science*, 8(12):210617.

Marek, P. and Vávra, F. (2020). Comparison of home advantage in european football leagues. *Risks*, 8(3):87.

McCarrick, D., Bilalic, M., Neave, N., and Wolfson, S. (2021). Home advantage during the covid-19 pandemic: Analyses of european football leagues. *Psychology of sport and exercise*, 56:102013.

Neave, N. and Wolfson, S. (2003). Testosterone, territoriality, and the 'home advantage'. *Physiology & behavior*, 78(2):269–275.

Nevill, A., Balmer, N., and Williams, M. (1999). Crowd influence on decisions in association football. *The Lancet*, 353(9162):1416.

Nevill, A. M., Balmer, N. J., and Williams, A. M. (2002). The influence of crowd noise and experience upon refereeing decisions in football. *Psychology of sport and exercise*, 3(4):261–272.

Nevill, A. M. and Holder, R. L. (1999). Home advantage in sport: An overview of studies on the advantage of playing at home. *Sports Medicine*, 28:221–236.

Peeters, T. and van Ours, J. C. (2021). Seasonal home advantage in english professional football; 1974–2018. *De Economist*, 169(1):107–126.

Peters, G. (2017). Waarom sportclubs ervoor kiezen om rugnummers voor eeuwig vast te leggen. `https://www.volkskrant.nl/sport/waarom-sportclubs-ervoor-kiezen-om-rugnummers-voor-eeuwig-vast-te-leggen~bd3fe892/`. Accessed: 2023-04-09.

Pollard, R. (1986). Home advantage in soccer: A retrospective analysis. *Journal of sports sciences*, 4(3):237–248.

Pollard, R. (2006a). Home advantage in soccer: variations in its magnitude and a literature review of the inter-related factors associated with its existence. *Journal of Sport Behavior*, 29(2):169.

Pollard, R. (2006b). Worldwide regional variations in home advantage in association football. *Journal of sports sciences*, 24(3):231–240.

Pollard, R. (2008). Home advantage in football: A current review of an unsolved puzzle. *The open sports sciences journal*, 1(1).

Pollard, R. and Gómez, M. A. (2009). Home advantage in football in south-west europe: Long-term trends, regional variation, and team differences. *European Journal of Sport Science*, 9(6):341–352.

Pollard, R., Silva, C., and Medeiros, N. (2008). Home advantage in football in brazil: differences between teams and the effects of distance traveled. *Revista Brasileira de Futebol (The Brazilian Journal of Soccer Science)*, 1(1):3–10.

Ponzo, M. and Scoppa, V. (2018). Does the home advantage depend on crowd support? evidence from same-stadium derbies. *Journal of Sports Economics*, 19(4):562–582.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Razali, N., Mustapha, A., Yatim, F. A., and Ab Aziz, R. (2017). Predicting football matches results using bayesian networks for english premier league (epl). In *Iop conference series: Materials science and engineering*, volume 226, page 012099. IOP Publishing.

RIVM (2020). Tijdlijn van coronamaatregelen 2021. `https://www.rivm.nl/gedragsonderzoek/tijdlijn-van-coronamaatregelen-2021`.

Rozendaal, R. (2021). De financiële staat van eredivisie clubs. `https://www.voetbalfinancien.nl/resources/Voetbalfinancien_2021_Nederlands_Spreads.pdf`.

Schwartz, B. and Barsky, S. F. (1977). The home advantage. *Social forces*, 55(3):641–661.

Sors, F., Grassi, M., Agostini, T., and Murgia, M. (2022). A complete season with attendance restrictions confirms the relevant contribution of spectators to home advantage and referee bias in association football. *PeerJ*, 10:e13681.

Trouw (2022). Regel van vijf wissels per wedstrijd permanent gemaakt. `https://www.trouw.nl/sport/regel-van-vijf-wissels-per-wedstrijd-permanent-gemaakt~b70a1991/`. Accessed: 2023-04-13.

Van Damme, N. and Baert, S. (2019). Home advantage in european international soccer: which dimension of distance matters? *Economics*, 13(1).

Van de Ven, N. (2011). Supporters are not necessary for the home advantage: Evidence from same-stadium derbies and games without an audience. *Journal of Applied Social Psychology*, 41(12):2785–2792.

van Leeuwen, Q. (2020). Master thesis: Data science and marketing analytics - analyzing the impact of substitutions in football matches. Master's thesis, Erasmus University Rotterdam - Erasmus School of Economics.

van Ours, J. C. (2018). Artificial pitches and home advantage in professional football. *Tinbergen Institute*.

van Ours, J. C. (2019). A note on artificial pitches and home advantage in dutch professional football. *De Economist*, 167(1):89–103.

Witlox, M. (2022). Lege stadions zorgen voor lagere transfersommen: "de hele voetbalwereld geeft minder uit". `https://www.rtlnieuws.nl/economie/artikel/5281233/voetbal-eredivisie-transfersommen-corona-inkomsten-deel-1`. Accessed: 2023-04-15.

Wolfson, S., Wakelin, D., and Lewis, M. (2005). Football supporters' perceptions of their role in the home advantage. *Journal of sports sciences*, 23(4):365–374.

Wunderlich, F., Weigelt, M., Rein, R., and Memmert, D. (2021). How does spectator presence affect football? home advantage remains in european top-class football matches played without spectators during the covid-19 pandemic. *Plos one*, 16(3):e0248590.

Zebari, G. M., Zeebaree, S. R., Sadeeq, M. A., and Zebari, R. R. (2021). Predicting football outcomes by using poisson model: Applied to spanish primera división. *Journal of Applied Science and Technology Trends*, 2(04):105–112.

Appendix

Table 3: Clubs in the Eredivisie for each season in the data set.

Club	Season					
	2017-2018	2018-2019	2019-2020	2020-2021	2021-2022	2022-2023
ADO Den Haag	**X**	**X**	**X**	X		
Ajax	X	X	X	X	X	X
AZ	X	X	X	X	X	X
De Graafschap		X				
Excelsior	**X**	**X**				**X**
FC Emmen		**X**	**X**	**X**		**X**
FC Groningen	X	X	X	X	X	X
FC Twente	X		X	X	X	X
FC Utrecht	X	X	X	X	X	X
FC Volendam						**X**
Feyenoord	X	X	X	X	X	X
Fortuna Sittard		X	X	X	X	X
Go Ahead Eagles					X	X
Heracles Almelo	**X**	**X**	**X**	**X**	**X**	
N.E.C.					X	X
NAC Breda	X	X				
PEC Zwolle	**X**	**X**	**X**	X	X	
PSV	X	X	X	X	X	X
RKC Waalwijk			X	X	X	X
Roda JC	**X**					
SC Cambuur					**X**	**X**
sc Heerenveen	X	X	X	X	X	X
Sparta Rotterdam	**X**		**X**	**X**	**X**	X
Vitesse	X	X	X	X	X	X
VVV-Venlo	**X**	**X**	**X**	X		
Willem II	X	X	X	X	X	
Total clubs	18	18	18	18	18	18

Note: X denotes that a club played in the Eredivisie in that season. An **X** indicates that the specific club played on an artificial pitch that season.

Table 4: Clubs in the Eredivisie for each season in the data set and their final end rankings.

Club	Season					
	2017-2018	2018-2019	2019-2020*	2020-2021	2021-2022	2022-2023**
ADO Den Haag	**7**	**9**	**17**	18		
Ajax	2	1	1	1	1	3
AZ	3	4	2	3	5	4
De Graafschap		17				
Excelsior	**11**	**16**				**15**
FC Emmen		**14**	**12**	**16**		**16**
FC Groningen	12	8	9	7	12	17
FC Twente	18		14	10	4	5
FC Utrecht	5	6	6	6	7	7
FC Volendam						**14**
Feyenoord	4	3	3	5	3	1
Fortuna Sittard		15	16	11	15	13
Go Ahead Eagles					13	10
Heracles Almelo	**10**	**7**	**8**	**9**	**16**	
N.E.C.					11	11
NAC Breda	14	18				
PEC Zwolle	**9**	**13**	**15**	13	18	
PSV	1	2	4	2	2	2
RKC Waalwijk			18	15	10	9
Roda JC	**16**					
SC Cambuur					**9**	**18**
sc Heerenveen	8	11	10	12	8	8
Sparta Rotterdam	**17**		**11**	**8**	**14**	6
Vitesse	6	5	7	4	6	12
VVV-Venlo	**15**	**12**	**13**	17		
Willem II	13	10	5	14	17	
Total clubs	18	18	18	18	18	18

Note: A bold **ranking** indicates that the club played on an artificial pitch that season.
* Season 2019-2020 was stopped after the 26th gameweek due to the pandemic. The standings at the time were used as the end ranking.
** Season 2022-2023 is still underway as of writing, but the standings are with one game left to play. Feyenoord, FC Twente, Sparta Rotterdam, FC Utrecht and FC Emmen can not change their position in the table anymore, for the other 13 teams their final standings could change based on the last game week.

Table 5: Overview containing the different COVID we defined and explained in section 3.3 of the Data.

Period	**Season**	**Rounds**	**Start Date**	**End Date**	**#Matches**	**Total**
Before COVID:	2017-2018	1-34	11-08-2017	06-05-2018	306	
	2018-2019	1-34	10-08-2018	15-05-2019	306	
	2019-2020	1-25/26	02-08-2019	08-03-2020	158	
	Total matches Before COVID					*770*
COVID W/O Att.	2020-2021	4-29	02-10-2020	22-04-2021	235	
		31-34	01-05-2021	16-05-2021	36	
	2021-2022	11*	31-10-2021	31-10-2021	1	
		12*	06-11-2021	06-11-2021	1	
		13-20	20-11-2021	23-01-2022	74	
		21*	05-02-2021	05-02-2021	1	
	Total matches During COVID without Attendance					*348*
COVID restricted Att.	2020-2021	1-3**	12-09-2020	27-09-2020	26	
		30**	23-04-2021	25-04-2021	9	
	2021-2022	1-12**	13-08-2021	07-11-2021	104	
		21-22**	05-02-2022	13-02-2022	17	
	Total matches During COVID with restricted Attendance					*156*
After COVID	2021-2022	23-34	19-02-2022	15-05-2022	108	
	2022-2023	1-33*	05-08-2022	21-05-2023	288	
	Total matches After COVID					*396*
Total matches in the Data Set						**1670**

Note:
* As explained in subsection 3.3, three matches from game weeks 11, 12 and 21 in the 2021-2022 season, and two matches from game weeks 30 and 32 in the 2022-2023 season were played without attendance but for another reason than COVID restrictions.
** The matches in these game weeks were played with a restricted attendance, where full capacity was not allowed.

Figure 10: Density plot for the variable Goals.

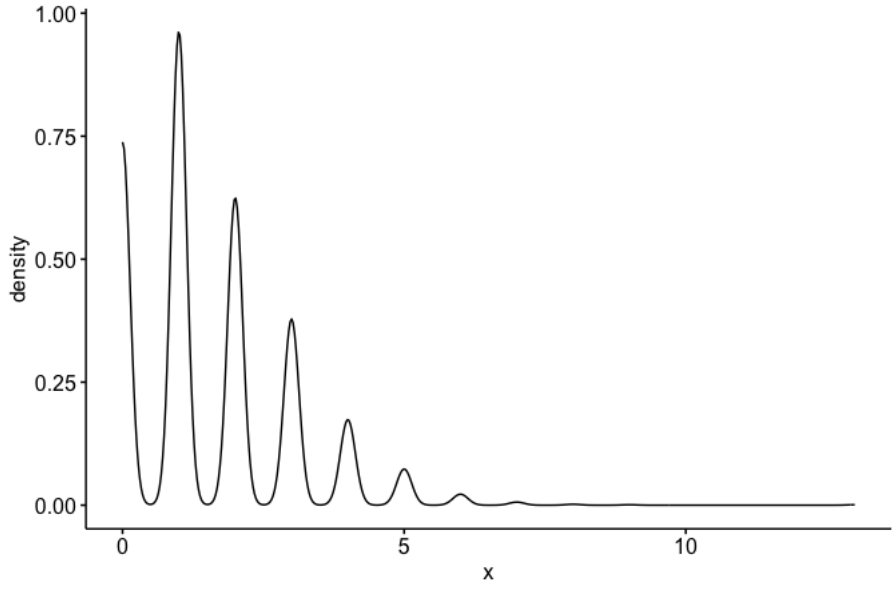


Figure 11: QQ plot for the variable Goals.