

The Impact of Home Advantage, Goals Scored, Shots on Target on Premier League Outcomes*

How do home advantage, goals scored and shots on target increase the winning rate for the premier league teams

Jeongwoo Kim

April 17, 2024

This study investigates the impacts of home advantage, goals scored, and shots on target on the winning rate of Premier League Football teams. Through multiple linear regression analysis using current season data, we found that home advantage, goals scored and shots on target significantly increase winning rates. These findings highlight the strategies for football teams to focus on winning by applying more offensive tactics at home game. This paper further contributes to a deeper understanding of the effects of home advantage and the essence of offensive quality in modern football.

1 Introduction

Football is arguably one of the most loved sports around the world and the English Premier League is one of the most popular professional football leagues. Professional footballers and managers in this league earn substantial popularity, fame and money. Given the low scoring nature of football, the moments when a player scores goals are the climax of the event. Fans enjoy watching football matches and often go and cheer for their team especially in their home game to give them support. In the competitive arena of the English Premier League, understanding the variables that affect team success is important. This study aims to investigate how home advantage, goals scored, and shots on target correlate with winning rates. These factors, though analyzed individually in past research are often not integrated into a unified model. This paper fills this gap by employing the most recent dataset from the ongoing 2023/24 Premier League season, reflecting the most recent trends in football.

*Code and data are available at: https://github.com/Kjeongwoo99/Premier_League_Home_Advantage

Home advantage refers to the tendency where sports teams generally achieve better results at their own stadiums compared to playing away. This concept is widely recognized across various team sports, including football (Goumas 2014). It occurs worldwide, regardless of the level of the game, and about 60% of points are gained at home at the elite football level (Goumas 2014). Factors such as home crowd’s support on player’s performance and referee’s decision-making, adverse effects of travelling, jet lag and fatigue, and psychologically defensive reaction to protect territory have been identified to contribute to home advantage (Goumas 2014). However, the extent to which home advantage influences football game remains ambiguous, especially in the context of goal scored and shot accuracy. This study uses multiple linear regression model to explore these relationships, providing quantifiable insights into how each factor contributes to the winning rates.

Our analysis reaffirms home advantage and highlights the importance of strategic offensive play evidenced by positive correlation between goals scored, shots on target and winning rates. Interestingly, our model has indicated a non-significant or even negative relationship between the proportion of home games and winning rates suggesting that the more number of home games do not guarantee higher winning rates. This suggests other underlying factors could play in home advantage.

These insights are crucial for board directors, managers, coaches, scouts and players who want to develop their team and improve team performance in one of the world’s most watched football leagues. By understanding the specific contributions of these factors to winning, teams can manage and adjust their training, match preparation, and tactical decisions more effectively. Moreover, this study contributes to the broader field of team sports by highlighting the importance of offensive plays.

The remainder of the paper is structured to first detail the data and cleaning process used in our analysis in Section 2. This is followed by an in-depth analysis of the specifics of the variables, and what we consider important in our data. Section 3 introduces the specifics of the multiple linear regression model. We then present the findings of the data in Section 4 and discussion of the implications of our findings in Section 5, confirming existing theories in sports performance analysis and conclude with recommendations for future research.

1.1 Estimand

The estimand of this study is the expected change in the winning rates by the three key variables: home advantage, the average goals scored, and the average shots on target.

2 Data

The data used in this study is the Premier League match results data from <https://www.football-data.co.uk/>. It contains Full Time/Half Time results and match statistics of different football

leagues and divisions around the world. I have chosen the current season (2023/24) of the Premier League Division to reflect the most recent data applicable for this analysis because sports, especially football is fast-changing in terms of individual performances and tactics (n=323). For instance, it has been almost 15 years since the tiki-taka style of Barcelona with Lionel Messi, Andrés Iniesta, Xavier Hernández and Sergio Busquets under Pep Guardiola shocked the world and changed the paradigm of modern football. Just five years later, the tiki-taka era came to an end when the Spanish were eliminated after a shocking defeat against the Netherlands in the group stage in the 2014 Brazil World Cup. As such, in modern football, trend is changing rapidly. What dominates the world football right now may become old and tired in a couple of years later. Using the current season’s data of the best league in the world can prevent this issue.

This dataset contains match results of all the Premier League matches in the ongoing season 2023-24. This includes variables such as Date (Match Date (dd/mm/yy)), Time (Time of kick off), HomeTeam (Home Team), AwayTeam (Away Team), FTHG (Full Time Home Team Goals), FTAG (Full Time Away Team Goals), FTR (Full Time Result (H=Home Win, D=Draw, A=Away Win)), HTHG (Half Time Home Team Goals), HTAG (Half Time Away Team Goals), HTR (Half Time Result (H=Home Win, D=Draw, A=Away Win)), Attendance (Crowd Attendance), Referee (Match Referee), HS (Home Team Shots), AS (Away Team Shots), HST (Home Team Shots on Target), AST (Away Team Shots on Target) and many others. This dataset was chosen because it provided the most complete data of home team, away team of every single match and the team’s performance metrics.

This study focuses on the relationship between the winning rate and home advantage, goals scored, and shots on target for the Premier League Teams. Of these variables, what are considered important are **Home Team**, **Away Team**, **Full Time Home Team Goals**, **Full Time Away Team Goals**, **Full Time Result**, **Home Team Shots on Target**, and **Away Team Shots on Target**. **Home Team** and **Away Team** are categorical variables that identifies each home/away team. **Full Time Home Team Goals** and **Full Time Away Team Goals** are the average number of goals scored by the home/away team per game, which indicates the offensive strength of the team at home/away. **Home Team Shots on Target** and **Away Team Shots on Target** are the average number of shots on target per game by home/away team. This metric is an indicator of both the team’s offensive quality and the accuracy of their attempts at goal. In professional football (soccer), around 10% of all shots taken result in a goal, and a mere 1% of all attacking plays lead to scoring (Gabriel Anzer 2021). Therefore, an increase in the number of shots on target taken by a team increases the scoring opportunity greatly due to the low scoring nature of football. I have also created a new column **Home Win** being 1 when Home Team won and being 0 if home team drew or lost. A similar column **Away Win** is also made that equals 1 if away team won and 0 if away team drew or lost.

Data was collected and analyzed using R statistical programming software (R Core Team 2023), with additional packages like tidyverse (Wickham et al. 2019), knitr (Xie 2020), dplyr (Wickham et al. 2023) here (Müller 2020), kableExtra (Zhu 2020), modelsummary (Arel-Bundock 2022), tibble (Wickham and Müller 2023), scales (Wickham and Seidel 2023), gt

(Iannone, Cheng, and Schloerke 2023), and ggplot2 (Wickham 2016).

Table 1: Premier League Home Team Match Data

Home Team Statistics
Summary of games played, wins, and shooting performance

| Home Team | Games Played | Wins | Winning rate (%) | Goals Scored | Shots on Target |
|------------------|--------------|------|------------------|--------------|-----------------|
| Liverpool | 17 | 13 | 76.47% | 2.53 | 8.71 |
| Newcastle | 17 | 11 | 64.71% | 2.53 | 6.41 |
| Man City | 17 | 12 | 70.59% | 2.53 | 6.18 |
| Aston Villa | 16 | 11 | 68.75% | 2.50 | 5.75 |
| Arsenal | 16 | 12 | 75.00% | 2.38 | 6.44 |
| Chelsea | 16 | 8 | 50.00% | 2.19 | 6.19 |
| Tottenham | 16 | 12 | 75.00% | 2.12 | 6.56 |
| Brighton | 15 | 7 | 46.67% | 1.87 | 6.47 |
| Fulham | 16 | 9 | 56.25% | 1.81 | 4.81 |
| Brentford | 17 | 5 | 29.41% | 1.59 | 4.53 |
| Man United | 15 | 8 | 53.33% | 1.53 | 5.53 |
| Wolves | 15 | 7 | 46.67% | 1.53 | 4.33 |
| West Ham | 17 | 6 | 35.29% | 1.53 | 4.12 |
| Luton | 16 | 4 | 25.00% | 1.50 | 4.38 |
| Nott'm Forest | 17 | 5 | 29.41% | 1.47 | 4.06 |
| Crystal Palace | 15 | 4 | 26.67% | 1.40 | 4.20 |
| Bournemouth | 17 | 6 | 35.29% | 1.35 | 5.18 |
| Everton | 15 | 4 | 26.67% | 1.07 | 5.20 |
| Sheffield United | 16 | 2 | 12.50% | 1.06 | 3.44 |
| Burnley | 17 | 2 | 11.76% | 1.00 | 3.82 |

Table 1 is a summary table for home teams in the English Premier League in 2023-24. It lists the teams along with their performance metrics for matches played at home, sorted by Goals Scored then Shots on Target in descending order. We can see that Liverpool has the highest Winning Rate with 76.47%, with Arsenal and Tottenham both 75%, Man City 70.59% and Aston Villa 68.75%, indicating the top 5 teams in terms of a strong home performance. As of 15 April, the top five teams in the league are Man City, Arsenal, Liverpool, Aston Villa, and Tottenham in that order. This table shows a strong correlation between home winning rate and the league position. In terms of Goals Scored, the top five teams are Liverpool, Newcastle, Man City, Aston Villa, and Arsenal. In the league, they are ranked 3rd, 6th, 1st, 4th, and 2nd respectively, which shows a strong correlation between winning rate and goals scored. Liverpool, being top of the list also has the most shots on target with 8.71 per match, followed by Newcastle at 6.41 and Man City at 6.18. We can find a similar pattern down the bottom of the list. The bottom five teams are Crystal Palace, Bournemouth, Everton,

Sheffield United, and Burnley that are ranked 14th, 13th, 16th, 20th, and 19th in the league table as of 32 match days. They have all scored less than 1.5 goals with less than 5.20 shots on target per home game. The bottom two teams (Sheffield United and Burnley) show very similar statistics with 12.5% and 11.76% winning rate, 1.06 and 1.00 goals scored and 3.44 and 3.82 shots on target per home game. These are by far the lowest among the teams and they are ranked 20th and 19th in the league respectively. From this table, we can find out that there seems to be a tendency that teams with a higher number of shots on target or goals scored to have a higher winning rate, suggesting a positive correlation between offensive play quality and winning outcomes.

Table 2: Premier League Away Team Match Data

Away Team Statistics
Summary of games played, wins, and shooting performance

| Away Team | Games Played | Wins | Winning Rate (%) | Goals Scored | Shots on Target |
|------------------|--------------|------|------------------|--------------|-----------------|
| Arsenal | 16 | 10 | 62.50% | 2.31 | 5.19 |
| Man City | 15 | 10 | 66.67% | 2.20 | 8.27 |
| Tottenham | 16 | 6 | 37.50% | 1.94 | 4.69 |
| Liverpool | 15 | 8 | 53.33% | 1.93 | 4.87 |
| Chelsea | 15 | 5 | 33.33% | 1.73 | 4.93 |
| Newcastle | 15 | 4 | 26.67% | 1.73 | 4.13 |
| Aston Villa | 17 | 8 | 47.06% | 1.65 | 4.82 |
| West Ham | 16 | 7 | 43.75% | 1.62 | 3.69 |
| Bournemouth | 15 | 5 | 33.33% | 1.60 | 4.53 |
| Brighton | 17 | 4 | 23.53% | 1.41 | 5.29 |
| Man United | 17 | 7 | 41.18% | 1.41 | 4.18 |
| Wolves | 17 | 5 | 29.41% | 1.35 | 4.18 |
| Luton | 17 | 2 | 11.76% | 1.29 | 2.94 |
| Brentford | 16 | 3 | 18.75% | 1.25 | 4.31 |
| Fulham | 17 | 3 | 17.65% | 1.18 | 4.88 |
| Nott'm Forest | 16 | 2 | 12.50% | 1.06 | 3.75 |
| Burnley | 16 | 2 | 12.50% | 1.00 | 3.12 |
| Crystal Palace | 17 | 4 | 23.53% | 0.94 | 3.88 |
| Everton | 17 | 5 | 29.41% | 0.94 | 3.65 |
| Sheffield United | 16 | 1 | 6.25% | 0.81 | 2.94 |

Table 2 is a summary table for away teams. What seems significant by first glance is the winning rate. The highest away winning rate is by Man City with 66.67%; this is about 10% lower than the highest home winning rate by Liverpool, suggesting that there is a general trend that away teams face greater challenge in winning. Similarly, Sheffield United at the bottom of the table with the lowest winning rate with 6.25%, compared the Table 1's lowest

winning rate of 11.76% by Burnley further strengthens this view that teams perform better at a home game in general. Top teams like Arsenal and Liverpool show a significant decrease in their winning rate at away games with 62.5% and 53.33% winning rate, which is around 13% and 23% lower than their home winning rate respectively. Similarly, the number of goals scored are also reduced for teams at an away game with Man City by 0.31, Tottenham by 0.18, Liverpool by 0.6, Newcastle by 0.8 and Aston Villa by 0.85. Most teams score less in away matches, which is intuitive because if you score fewer goals, you are less likely to win a match in football. Like Table 1, we can also identify a similar positive correlation between the winning rate and goals scored, and the winning rate and shots on target.

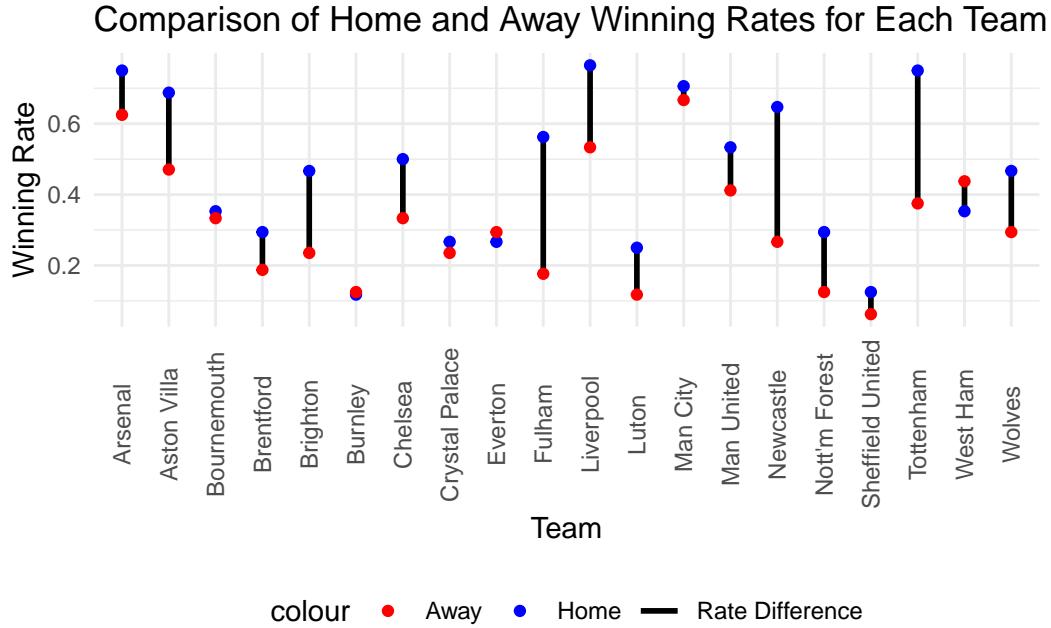


Figure 1: Comparison of Home and Away Win Rates for Each Team

Figure 1 below is a paired column chart that links the home (blue) and away (red) winning rates for each team with a line, making it easier to see the differences between the two for each team. Each team has two colored circles: blue represents the home winning rate, and red represents the away winning rate. The black line connecting the circles indicates the distance between the home and away winning rates for each team. The length of the black lines between the red and blue dots for each team shows the variability between home and away performance. A longer line suggests a greater disparity between home and away winning rates.

A common pattern is that the blue dots (home winning rate) are generally higher than the red dots (away winning rate), indicating a typical home advantage where teams tend to win more often when playing at their home venue. The only teams that play better at away games are Burnley, Everton, West Ham but with minimal differences. Except these 3 teams, the rest of

17 teams, which is 85% of the league play better at home. Teams with both high home and away winning rates could be considered strong overall. We can observe this trend in Arsenal and Man City where both have a winning probability higher than 60% at both home and away matches. On the other hand, teams with a significant gap in the winning rate when playing away might be more reliant on home advantage. Fulham is a typical example of this. Its winning rate is around 55% at home but it drops significantly to just less than 20% at an away game. Similarly, Tottenham plays much better at home with 65% winning rate but less than 40% at an away game. In general, the winning rates for both home and away games appear to range from approximately 20% to 60%, with home winning rates generally being higher.

3 Model

In this study, we aim to predict the winning rates based on variables: home advantage, goals scored and shots on target. This is because our data shows a strong correlation between the winning rate and home games, the winning rate and goals scored and the winning rate and shots on target. We saw earlier the linear relationships between the winning rate and goals scored and the winning rate and shots on target from Table 1 and Table 2. We also saw a consistent gap between the winning rate between home game and away game from Figure 1. Therefore, we are trying to understand how home game, an increase in goals scored and an increase in shots on target affect the winning rate for the football teams in the Premier League. To achieve this, we will run a multiple linear regression model with the dependent variable: the winning rate and the independent variables: proportion of home games, average goals scored and average shots on target.

The multiple linear regression takes the form of:

$$y_i = \beta_0 + \beta_1 x_{home} + \beta_2 x_{goal} + \beta_3 x_{shot} \quad (1)$$

In equation 1, each β represents the coefficients of the model. β_0 measures our dependent variable: the winning rate, representing the baseline winning rate when all other predictors are zero. The coefficients β_1 , β_2 and β_3 indicate the chance of winning by an additional increase in each variable; in this case, proportion of home games, scoring one more goal, and shot on target by one respectively.

The limitations of our model lie in the complex nature of football itself. It is a complex game with many influencing factors like player's skills, team cohesion, fitness levels, mentality, tactical strategies: formation, in-game adjustments, pitch conditions, injuries and even weather. This model with three variables may be an oversimplification of complex real world situation. However, this model can still provide some insights in some ways. By highlighting the impact of goals scored and shots on target, it directly gives attention to attacking abilities to win games. This is not only useful for teams to focus on developing their attacking strategies, but it can also develop the modern football tactically and attract more viewers and crowds since scoring

Table 3: Coefficients from the Model

| term | estimate | std.error | statistic | p.value | conf.low | conf.high |
|-----------------------|------------|-----------|-----------|-----------|------------|-----------|
| (Intercept) | 0.3447511 | 0.3287090 | 1.048803 | 0.3098537 | -0.3520809 | 1.0415832 |
| Proportion_Home_Games | -1.3168632 | 0.6802276 | -1.935915 | 0.0707499 | -2.7588813 | 0.1251549 |
| Avg_Goals_Scored | 0.2972522 | 0.0678685 | 4.379826 | 0.0004664 | 0.1533774 | 0.4411270 |
| Avg_Shots_on_Target | 0.0448461 | 0.0275649 | 1.626929 | 0.1232824 | -0.0135888 | 0.1032810 |

goals are the essence of watching football. Moreover, by quantifying the impact of playing at home, teams can adjust their training, match preparations and line-up accordingly.

4 Results

Table 3 shows the estimates of the coefficients of our multiple linear regression that aims to understand the influence of home games, average goals scored, and average shots on target on the winning rate in Premier League football.

The intercept (0.2360462) represents the predicted winning rate when all other variables are at zero. This is the baseline winning rate for a team without taking into account home games, goals scored, or shots on target. In reality, the variables cannot be zero but this value can be understood as the baseline winning rate before considering the effects of other variables in the model. With a high p-value (0.4956470), the intercept is not statistically significant. This means there is no significant prediction for the winning rate based solely on the intercept.

The variable Proportion_Home_Games (-1.0920396) indicates that on average, a 109.20% drop in the winning rate is expected when the team has one more home game than away game in the season. This outcome is counter-intuitive given the well-documented phenomenon of home advantage where teams are expected to perform better at home. A negative coefficient implies that, among all teams considered for this model, having more home games led to lower winning rate. This shows that this variable could have been constructed unfairly. However, the coefficient on proportion of home game of the regression line from Table 3 suggests otherwise. This contrast shows the mishandling of the dataset. This increase in proportion of more home games perhaps does not fully capture the home vs away variable. Instead of merging the home and away game into one, creating a binary variable with 1 = home game, 0 = away game and then incorporating this into the linear model to see the effect of home game could have captured the situation better. In addition, the high p-value (.1371363) suggests that this is not statistically significant.

The variable Avg_Goals_Scored (0.2918261) means that the winning rate increases by 29.18 percentage points for each additional goal scored. This is intuitive because scoring more goals directly increases the chances of winning matches given the lower scoring nature of football. In reality, this finding underscores the importance of a strong offense and the ability to score

goals as critical to a team's success in the league. The low p-value (0.0005106) indicates that it is highly statistically significant.

Lastly, Avg_Shots_on_Target (0.0461157) suggests that the winning rate increases by 4.61 percentage points with an additional increase in the average shot on target per game. While the effect is much smaller than for average goals scored, it suggests that having more shots on target, which are more likely to result in goals, can also contribute to increasing the winning rate. This, again, reinforces the importance of attacking forces in football.

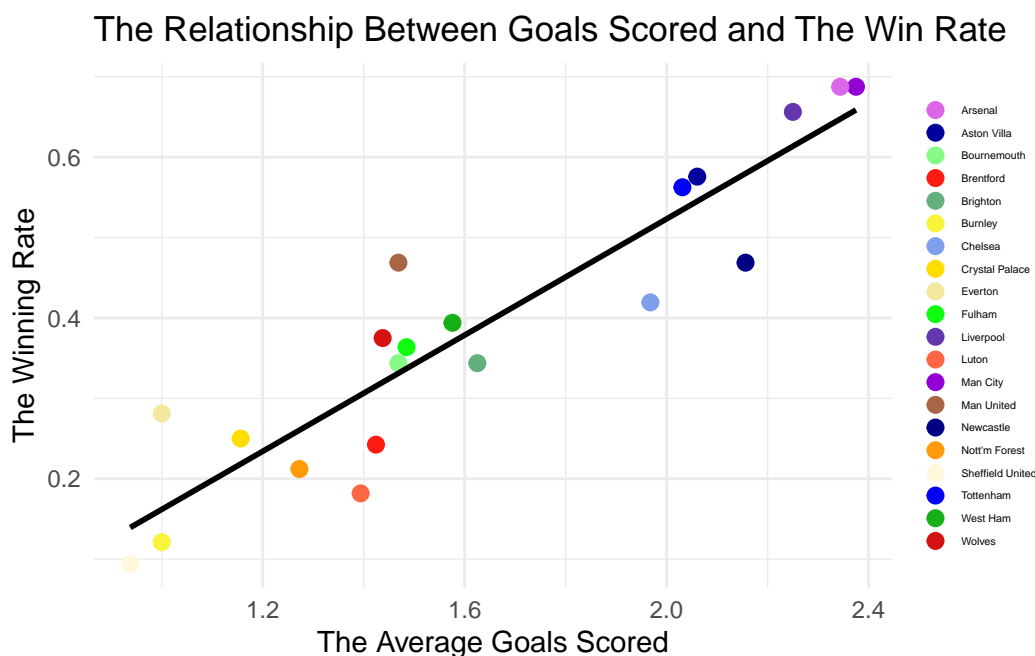


Figure 2: Goals Scored and The Winning Rate Relationship

Figure 2 is a scatter plot with a regression line, illustrating how the average goals scored by teams relate to their winning rates. Each team is represented by a colored dot on the graph on the rainbow color scale. The color gradient starts with yellow for the team with the lowest average goals scored and gradually darkens to violet for the team with the highest average goals scored. The color legend on the right allows readers to identify which color corresponds to which team.

There is a positive trend shown by the regression line that suggests that as the average number of goals scored by a team increases, their winning rate tends to increase as well. This is consistent with the general understanding that scoring more often leads to winning more matches. The x-axis (the average goals scored) shows that most teams score between 1.2 and 2.4 goals per game on average, with only 6 teams scoring more than 2 goals. They are the top teams in the league, competing for the title and the Champions League ticket. The y-axis, representing the winning rate as a percentage, varies from slightly above 20% to just below

60% for the majority. As teams score more goals, their winning rates increase accordingly. It shows that only the top teams like Liverpool, Arsenal and Man City record a winning rate of more than 60%, which shows how competitive the league is. The scatter of points around the regression line shows the variance in the data. Teams such as Everton and Man United that are above the line have a higher winning rate than predicted by their average goals scored, while those below the line such as Newcastle and Luton have a lower winning rate.

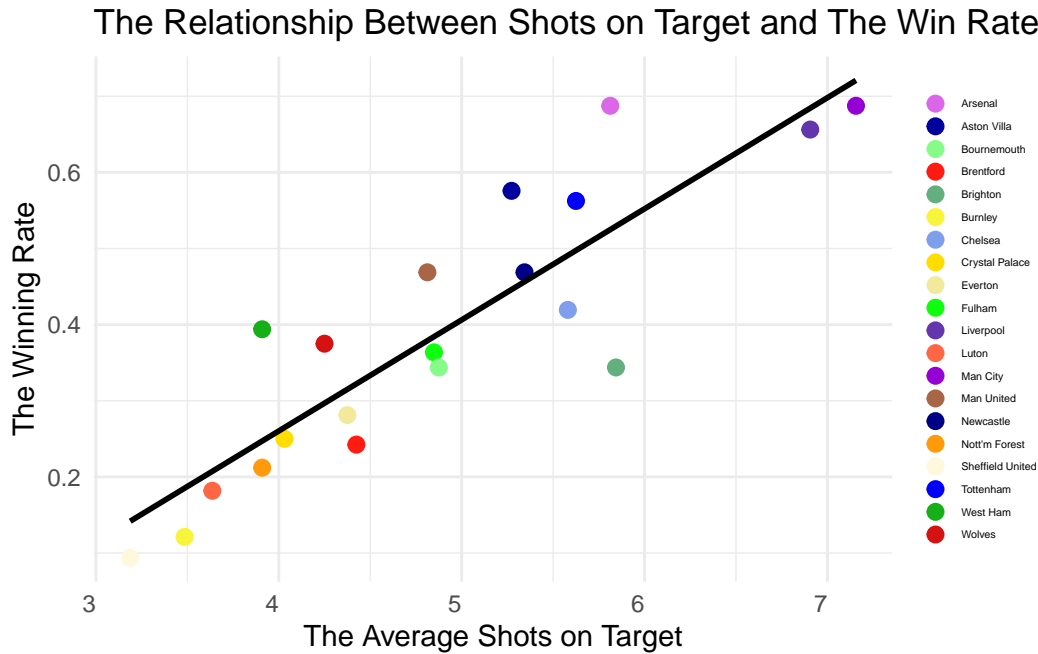


Figure 3: Shots on target and The Winning Rate Relationship

Figure 3 is another scatter plot that shows how the average number of shots on target by teams relates to their winning rates. Like the previous graph, this one also shows a positive correlation, as indicated by the upward slope of the regression line. This suggests that teams with more shots on target tend to have a higher winning rate, implying a beneficial impact of shooting accuracy and/or frequency on game outcomes. The data points are somewhat scattered around the regression line, indicating a reasonable fit of the model. From the plot, we can see that teams take about 3 to 7 shots on target on average, with top teams close to 7 shots on target and challengers close to 3 shots on target. This wide range suggests a more significant difference in attacking abilities, strategies, or execution between teams. Teams, such as 'Man City' and 'Liverpool', represented by dark and light purple dots, are at the higher end for both shots on target and winning rates, suggesting a strong attacking performance.

5 Discussion

5.1 Home Advantage

Home advantage has been a prominent feature of team sports and football is no exception to this. Regardless of division level, the home advantage is worth over 0.5 of a goal on average (Stephen R. Clarke 1995). It is evident that home advantage, which plays a crucial role in football, especially in tournaments, indeed exists. However, the extent of its impacts varies across teams. Our analysis reveals that while some teams play better at home, others do not capitalize on it as expected. Although the regression coefficients did not fully capture the home advantage, the disparities in the winning rates between home teams and away teams in Table 1 and Table 2 suggest that in the Premier League, home advantage exists and is quite important. Factors such as home crowd's support on player's performance and referee's decision-making, adverse effects of travelling, jet lag and fatigue, and psychologically defensive reaction to protect territory have been identified to contribute to home advantage (Goumas 2014). Past research suggests that referees can be influenced by social pressure from the home supporters (Marek Endrich 2020). This suggests that many factors beyond the physical venue can affect home advantage. One of the implications of the findings is that teams should do their best to win in their home match. Considering its benefits, a draw or a defeat at home is crucial especially for those competing for titles or avoiding relegation. This further suggests that the teams in the Champions League or Europa League should design their tactics to aggressively secure home wins to maximize this advantage.

5.2 Goals scored and The Winning Rate

The relationship between the winning rate and goals scored is one of the most straightforward metrics in football. Past studies suggest that winning teams show stronger results in goals scored than losing and drawing teams (Carlos Lago Peñas 2010). Our findings coincide with this view, which is shown effectively in Figure 2. It clearly shows that an increase in the average number of goals scored by a team is significantly correlated with an increase in their winning rate. This correlation goes back to the basic strategy of football: to win games by outscoring the opponent. The data shows that teams with higher average goals scored are not just winning more frequently but are often dominating the league. These teams are Man City, Liverpool and Arsenal from the 2023-24 season match data. This reinforces the need for teams to develop strong offensive capabilities as a critical component for their match strategy. This further leads to team tactics and training. To play better football offensively, you need to have higher shot quality, better positioning, better ball possession and build-up play into the opposition box to create goal scoring opportunities. Some very effective tactical managers like Pep Guardiola of Man City, Mikel Arteta of Arsenal, Jürgen Klopp of Liverpool, Xabi Alonso of Leverkusen and Roberto De Zerbi of Brighton & Hove Albion are dominating the world football in terms of new tactics. They show progressive tactics in their games, and other teams try to find ways to copy or defend their tactics. In the future, tactical planning and the

execution will be more important in modern football where players are now playing like a unit of a perfect mechanism compared to the past where an individual player's talent and skills were more important. Strikers and midfielders with creativity and work rate to consistently contribute to create more scoring opportunities will be more valuable to their teams. Defenders with great ball skills and passing abilities will be more valuable in the future.

5.3 Shots on Target and The Winning Rate

Among scholars and professionals, shooting in football has been a subject of keen interest and analysis (Rathke 2017). Previous research suggest that winning teams made more shots and shots on target than the opponents (Carlos Lago Peñas 2010). Figure 3 confirms a positive correlation between the number of shots on target and the winning rate, suggesting that increasing shots on target can increase a team's probability of winning. This is in line with the relationship between goals scored and the winning rate. The better ability to break down the opposition's defense and create scoring opportunities, the better chance of winning in football. This underscores the importance of the strategy of attacking. It is not just about quantity but the quality of shots because to shoot on target, it requires both precision and power in their attacks. Detailed training should focus on shooting accuracy, decision-making in the final third, and tactical movements that can lead to the creation of shooting opportunities. This metric can also be a evaluation tool for individual players especially for forwards, wingers and attacking midfielders. Players who have higher shots on target ratio are the ones more likely to convert chances into goals.

5.4 Weaknesses and Implications for Future Research

Despite these findings, the model, while insightful, relies on oversimplification to represent the complexity of football. Factors such as players' skills, team cohesion, fitness levels, mentality, tactical strategies: formation, in-game adjustments, pitch conditions, injuries, referee decisions are not captured but could significantly affect the outcomes of matches.

The negative coefficient associated with the proportion of home games suggests that this coefficient does not accurately capture home advantage given the 85% of teams performing better at home in our data. In the future, the study of home advantage could look into more specific aspects like crowd, referee's decision-making in home versus away games or psychological factors.

The study lacks in differentiation between different types of goals such as open play, set-pieces, penalties and solo, each of which might have different implications on the effect of goals scored and the winning rates. Looking into these details may provide better understanding. Integration of the impact of defensive tactics and pressure, assists, player movement, different types of passes can help understand the build-up play that leads to more goals. This can

provide a more comprehensive view of how teams can increase their scoring opportunities to win the match.

References

- Arel-Bundock, Vincent. 2022. “modelssummary: Data and Model Summaries in R.” *Journal of Statistical Software* 103 (1): 1–23. <https://doi.org/10.18637/jss.v103.i01>.
- Carlos Lago Peñas, Alexandre Dellal, Joaquin Lago Ballesteros. 2010. “Game-Related Statistics That Discriminated Winning, Drawing and Losing Teams from the Spanish Soccer League.” *J Sports Sci Med*. <https://pubmed.ncbi.nlm.nih.gov/24149698/>.
- Gabriel Anzer, Pascal Bauer. 2021. “A Goal Scoring Probability Model for Shots Based on Synchronized Positional and Event Data in Football (Soccer).” *Front. Sports Act. Living*. <https://www.frontiersin.org/articles/10.3389/fspor.2021.624475/full>.
- Goumas, Chris. 2014. “Tyranny of Distance: Home Advantage and Travel in International Club Football.” *International Journal of Performance Analysis in Sport*. <https://www.tandfonline.com/doi/epdf/10.1080/24748668.2014.11868698?needAccess=true>.
- Iannone, Richard, Joseph Cheng, and Barret Schloerke. 2023. *Gt: Easily Create Presentation-Ready Display Tables*. <https://gt.rstudio.com/>.
- Marek Endrich, Tobias Gesch. 2020. “Home-Bias in Referee Decisions: Evidence from ‘Ghost Matches’ During the Covid19-Pandemic.” *Economics Letters*. <https://www.sciencedirect.com/science/article/pii/S0165176520303815?via%3Dihub>.
- Müller, Kirill. 2020. *here: A Simpler Way to Find Your Files*. <https://CRAN.R-project.org/package=here>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rathke, Alex. 2017. “An Examination of Expected Goals and Shot Efficiency in Soccer.” *6th ISPAS International Workshop*. <https://www.semanticscholar.org/paper/An-examination-of-expected-goals-and-shot-in-soccer-Rathke/23dc10f6479f0e61b33c7b1ba1c9c408e873cd97>.
- Stephen R. Clarke, John M. Norman. 1995. “Home Ground Advantage of Individual Clubs in English Soccer.” *Journal of the Royal Statistical Society. Series D (The Statistician)*. <https://www.jstor-org.myaccess.library.utoronto.ca/stable/2348899?seq=6>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org>.
- Wickham, Hadley, and Kirill Müller. 2023. *Tibble: Simple Data Frames*. <https://tibble.tidyverse.org/>.
- Wickham, Hadley, and Dana Seidel. 2023. *Scales: Scale Functions for Visualization*. <https://scales.r-lib.org/>.
- Xie, Yihui. 2020. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.
- Zhu, Hao. 2020. *kableExtra: Construct Complex Table with ‘Kable’ and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.