# Analyzing the Dynamics of Home Advantage: A Statistical Investigation into Goals, Corners, and Shots on Target in Premier League Football*

Jerry Lu (Yen-Chia Lu)

April 18, 2024

This study investigates the phenomenon of home advantage in the Premier League for the 2022-2023 campaign by examining the correlation between goals, corners, and shots on target for the home club. Advanced statistical studies are performed using a season-long dataset to find trends and insights that clarify the significance of playing at home. Significant correlations that have practical consequences for clubs and coaching techniques, as well as contributing to the scholarly discourse on sports analytics, are highlighted by the findings. In light of the most recent season, this research provides fresh perspectives on the tactical and psychological factors affecting football game results.

## Table of contents



*Code and data are available at: https://github.com/Jerryluuuu/-Premier-League.git

# 1 Introduction

Since its founding in 1992, the Premier League has grown to become one of the world's most competitive and watched football competitions. Each of the twenty teams that compete each year plays 38 games, half of which are played at their home stadium and the other half away. The league is renowned for its competitive matches, fervent fandoms, and the exceptional skill level of its clubs and players. This competitive atmosphere creates an ideal setting for investigating several aspects of a team's performance, especially the home field advantage.

In sports, the idea of home team advantage is widely accepted, and the Premier League is no different. Playing at home is frequently linked to a higher chance of success, which can be related to things like comfortable surroundings, the backing of local supporters, and the lack of travel fatigue. Metrics that are essential for assessing a team's attacking prowess at home games include Full Time Home Team Goals, Home Team Shots on Target (HST), and Home Team Corners (HC). Understanding how home advantage might affect a team's attacking dynamics and overall performance in a match depends on these factors.

Not to be overlooked in this situation is the referee's involvement. Although it is the referees' responsibility to maintain fair play, their judgement have a big influence on how the game goes and how it ends. There is interest in the idea of "home bias," which holds that choices may unintentionally benefit the home team, for example, by giving them extra time off from injuries or more lenient rulings for fouls. Understanding the psychological effects of home advantage and the impact of outside variables on game outcomes can be gained by examining the relationship between referee decisions and home team performance indicators like goals, shots on target, and corners.

Furthermore, a key factor in maximizing home advantage is the home team's confidence. When combined with the encouragement of home fans, the psychological boost of playing in a famil-

iar environment can improve team performance. Higher percentages of shots on target and in corners are signs of an aggressive offensive strategy that frequently reflects this confidence. We can measure the effect of home confidence on a team's performance by looking at these variables, which provides a greater understanding of the psychological and strategic components that contribute to the Premier League's home advantage occurrences.

# 2 Data

The dataset was collected by the FootBall-Data.co and obtained from United kingdom Football-Data.co.uk (2023). This dataset excerpts from major football divisions and events. The main research object of this paper is to study the England Premier League from 2022-2023 season, it will be discussing the differences between home and away games and how the home field advantage affected the team competition. I analysis the data in R(R Core Team (2023)), with additional tools for support the analysis, including `tidyverse`(Wickham et al. 2019),`here`(Müller 2020), `dplyr`(Wickham et al. 2023), `readr`(Wickham, Hester, and Bryan 2023), `modelsummary`(Arel-Bundock 2022), `janitor`(Firke 2023),`tibble`(Müller and Wickham 2023), `ggplot2`(Wickham 2016), and research method is from Alexander (2023).

## 2.1 Variables

The dynamics of the English Premier League 2022–2023 season are the main focus of this study. I have carefully chosen a range of variables that are important to game outcomes in order to obtain insight into the subtleties of team performance both at home and away. They include "Full Time Team Goals," "Team Shots on Target," "Team Corners," as well as "Referees." These factors were picked because they have the capacity to have a big impact on how successful a team is. The fact that each team's performance can differ significantly between home and away games highlights the significant influence that venue dynamics have on football games. We hope to clarify the nuances of team dynamics and highlight the essential part that home and away games play in determining the results of matches by examining these elements.

## 2.2 Measurements

This project's metrics will address home-court advantage in sporting events, with a focus on Premier League games. Home field advantage is blatantly evident in any competitive sports match, including basketball, football, and basketball. Players typically perform better at home than away, on average. Supporter encouragement can help the home team perform better, whether it's by lowering mental strain. Consequently, the purpose of this study is to investigate, at the data level, the veracity of people's opinions regarding "home advantage." The team's success in scoring goals both at home and away, shooting on target both at home

and away, how well the team performs at corner kicks both at home and away, and the much-maligned refereeing all have an impact on the team's winning percentage both at home and away.

## 2.3 Data Visualizations and Summary Statistics

In England Premier League, I assume that each team will have different performances at home and away games, from the average number of full time goals, number of corner kicks, and number of shot on target, which can greatly affect the game. The Table 1 is the table shows the average of full time home team goals, average number of home team corners and the average number of home team shot on target for each of the 20 teams in league. The Table 2 is the table shows the teams at away games that have huge difference compared to the games at home. For example, the Man City, Arsenal, and Liverpool are famous teams in the world, the full time team goals, team corners, and shot on target are all drop at away games.

Table 1: Premier League Home Team data in 2023 season

| Team | Avg FTHG | Avg HC | Avg HST |
|---|---|---|---|
| Man City | 3.2 | 7.3 | 6.2 |
| Arsenal | 2.8 | 7.0 | 6.5 |
| Liverpool | 2.4 | 6.8 | 5.9 |
| Brighton | 1.9 | 7.5 | 7.1 |
| Tottenham | 1.9 | 6.2 | 6.1 |
| Man United | 1.9 | 5.9 | 6.2 |
| Newcastle | 1.9 | 8.2 | 5.8 |
| Brentford | 1.8 | 4.9 | 5.2 |
| Aston Villa | 1.7 | 4.8 | 4.6 |
| Fulham | 1.6 | 5.6 | 4.3 |
| Nott'm Forest | 1.4 | 2.6 | 3.7 |
| Leeds | 1.4 | 5.8 | 4.2 |
| West Ham | 1.4 | 5.6 | 4.1 |
| Leicester | 1.2 | 3.5 | 4.2 |
| Crystal Palace | 1.1 | 5.3 | 3.5 |
| Chelsea | 1.1 | 6.2 | 4.1 |
| Bournemouth | 1.1 | 4.3 | 3.8 |
| Southampton | 1.0 | 4.6 | 4.2 |
| Wolves | 1.0 | 5.3 | 3.9 |
| Everton | 0.8 | 5.3 | 4.7 |

Table 2: Premier League Away Team data in 2023 season

| Team | Avg FTAG | Avg AC | Avg AST |
|---|---|---|---|
| Arsenal | 1.8 | 4.7 | 4.3 |
| Brighton | 1.8 | 4.7 | 5.1 |
| Man City | 1.8 | 5.4 | 5.3 |
| Tottenham | 1.7 | 4.5 | 4.5 |
| Newcastle | 1.7 | 6.1 | 4.6 |
| Liverpool | 1.5 | 5.6 | 5.3 |
| Leicester | 1.5 | 3.6 | 3.8 |
| Fulham | 1.3 | 4.2 | 3.6 |
| Brentford | 1.2 | 3.7 | 3.3 |
| Leeds | 1.2 | 4.7 | 3.4 |
| Man United | 1.2 | 4.3 | 5.1 |
| Crystal Palace | 1.0 | 4.5 | 3.6 |
| Aston Villa | 0.9 | 3.9 | 3.3 |
| Chelsea | 0.9 | 4.8 | 4.3 |
| Everton | 0.9 | 3.9 | 3.3 |
| Bournemouth | 0.9 | 3.3 | 3.2 |
| Southampton | 0.9 | 3.6 | 3.3 |
| West Ham | 0.8 | 5.3 | 3.6 |
| Wolves | 0.6 | 4.5 | 2.6 |
| Nott'm Forest | 0.6 | 4.1 | 2.5 |

### 2.3.1 The relationship between win rates and team corners

The two scatter plots examine the correlation between a football team's percentage of wins both at home and away and its corner kicks. The first graph Figure 1 demonstrates a positive correlation, suggesting that teams with a greater average of corner kicks win more often at home. This implies that the number of corners could be a useful indicator of home game superiority. While there is still a positive tendency in the second graph Figure 2, which focuses on away games, it is not as noticeable, suggesting that if corners are linked to wins on the road, the relationship may not be as strong as it is in home games. This could imply that in away games, other elements become more important. Teams are represented by color-coded dots in both graphs, which makes it simpler to see how each team is doing in terms of corner kicks and victory rates.

### 2.3.2 The relationship between win rates and teams shots on target

The association between a football team's the win percentage for home and away Premier League games and the number of shots on target is displayed in these two scatter plots. The

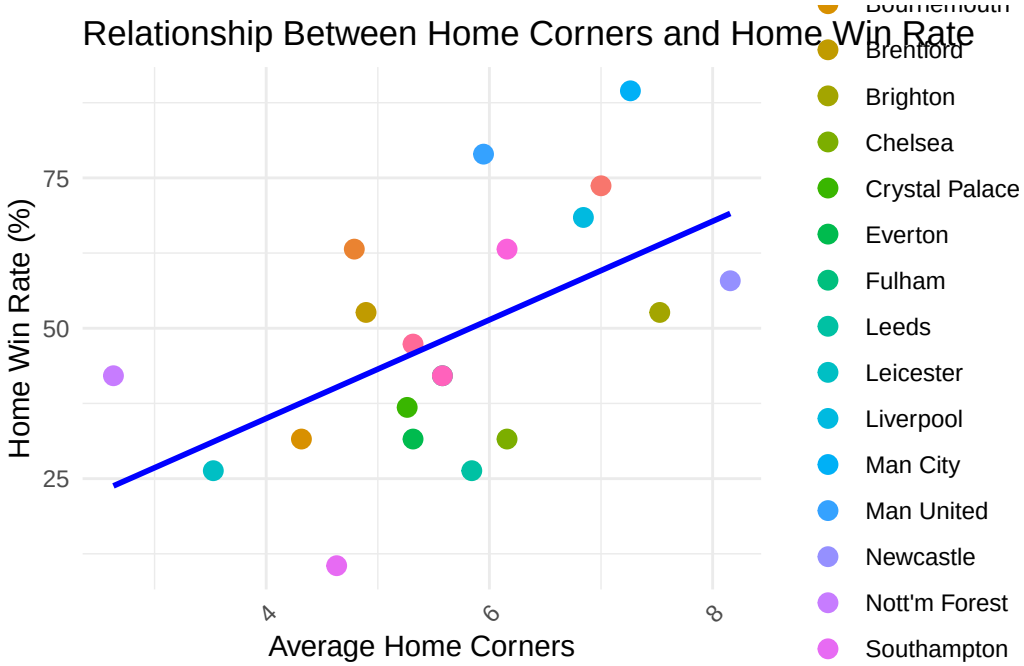


Figure 1: Home Corners and Win rate relationship

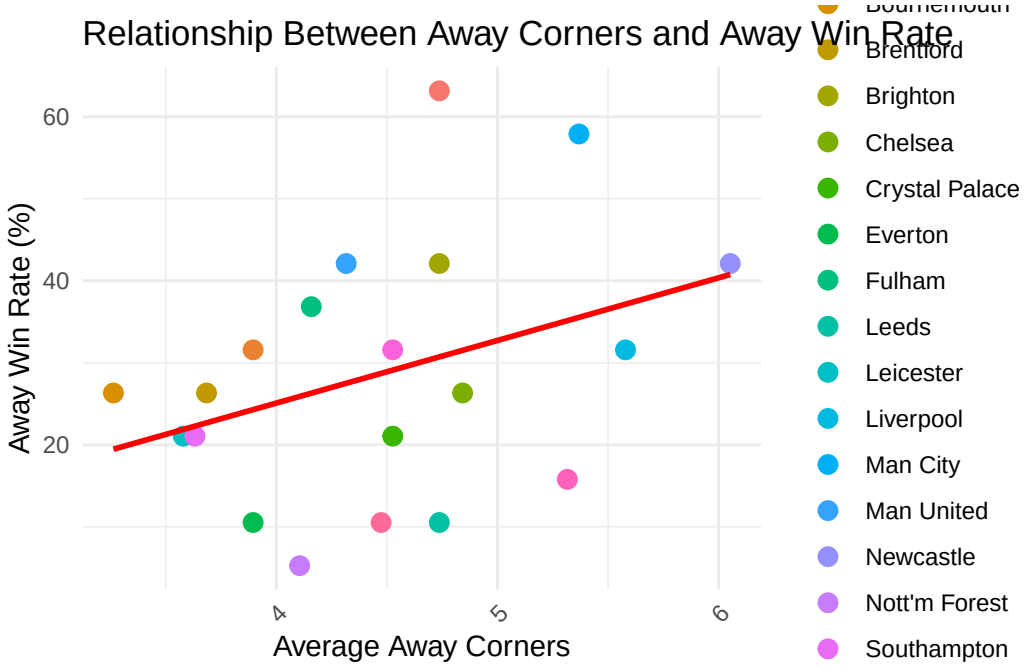


Figure 2: Away Coners and Win rate relationship

average home shots on target and the home victory rate have a positive link, as seen in the first plot Figure 3. It suggests that a better chance of winning in home games is correlated with increased precision. This could imply that a team's performance is greatly influenced by accurate attacking play at home.

The association between average away shots on target and away victory rates is displayed in the second plot Figure 4. Similar to the first plot, there is a clear positive association, indicating that teams that are able to strike the target more often when playing away from home also have a tendency to win more. This upward trend is shown by the red trend line, and the distribution of the points points to a somewhat constant link between various teams.

Teams are shown as color-coded dots in both plots, with team names and colors corresponded by the legend on the right. Although it's evident from both graphs that shooting accuracy and winning are correlated, it's huge to remember that correlation does not imply causation and that other factors might also play a role in a team's performance. When analyzing the variables influencing a team's performance, analysts may find these visualizations to be helpful resources.

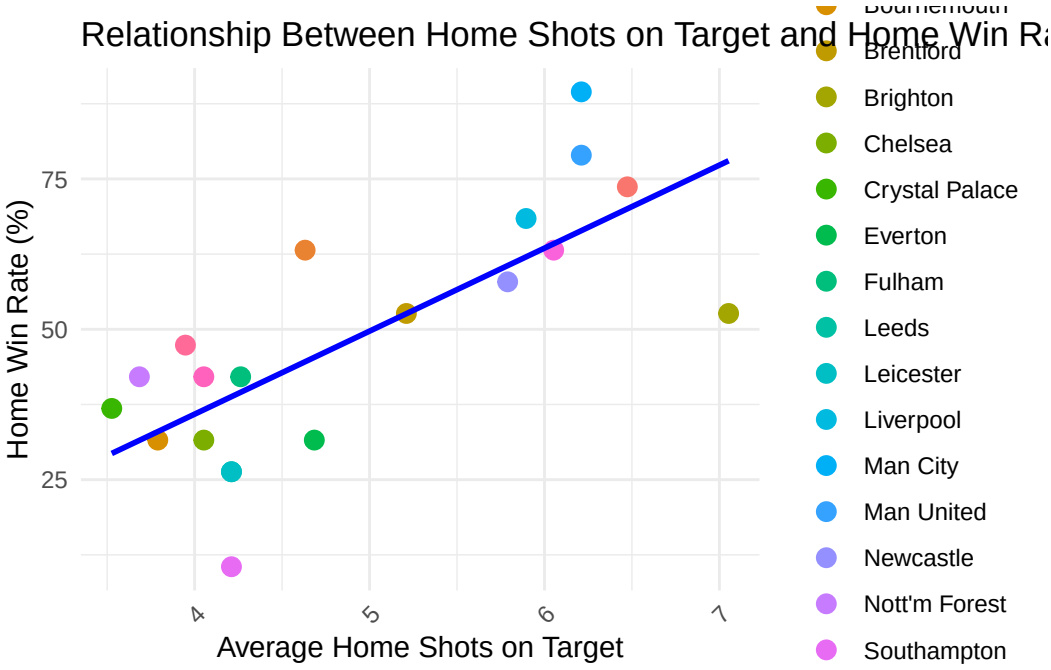


Figure 3: Home Team Shot on Target and Win rate relationship

### 2.3.3 Teams win rate on home games and away games

This bar graph gives a visual comparison of different football teams' home and away victory percentages. The teams are arranged along the x-axis, while the victory percentages are

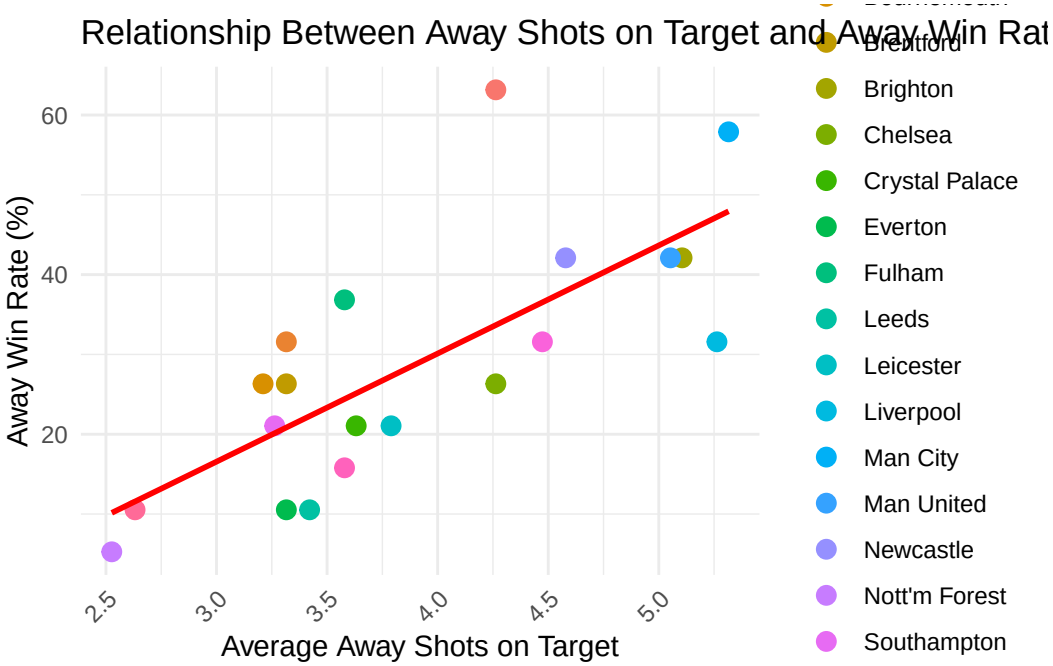


Figure 4: Away Team Shot on Target and Win rate relationship

displayed as percentages on the y-axis. Each team is represented by two bars, one in red for the away win rate and one in blue for the home win rate.

We can quickly compare each team's home and away results thanks to the graphic. As a general rule, teams with more blue bars than red bars tend to win more often at home than away, which is consistent with the notion of "home advantage" in sports. Nonetheless, several teams exhibit comparable home and away victory percentages, indicating that they consistently play well at all venues. On the other hand, a team's heavy dependence on home advantage may be indicated by notable disparities between their home and away bars.

In addition to being a simple tool for coaches, analysts, and fans to assess team performance, this bar chart may elicit more in-depth inquiries regarding the factors that influence each team's success both at home and away.

### 2.3.4 Can Referee influence the game results?

The bars from Figure 6 show how each referee affects the win rates at home (blue) and away (red). If a blue bar stretches upward, it indicates that the home team has a better win percentage than usual when this referee is refereeing. On the other hand, a blue bar that extends downward would indicate a home victory rate that is below average. For away teams, the red bars show the same information.

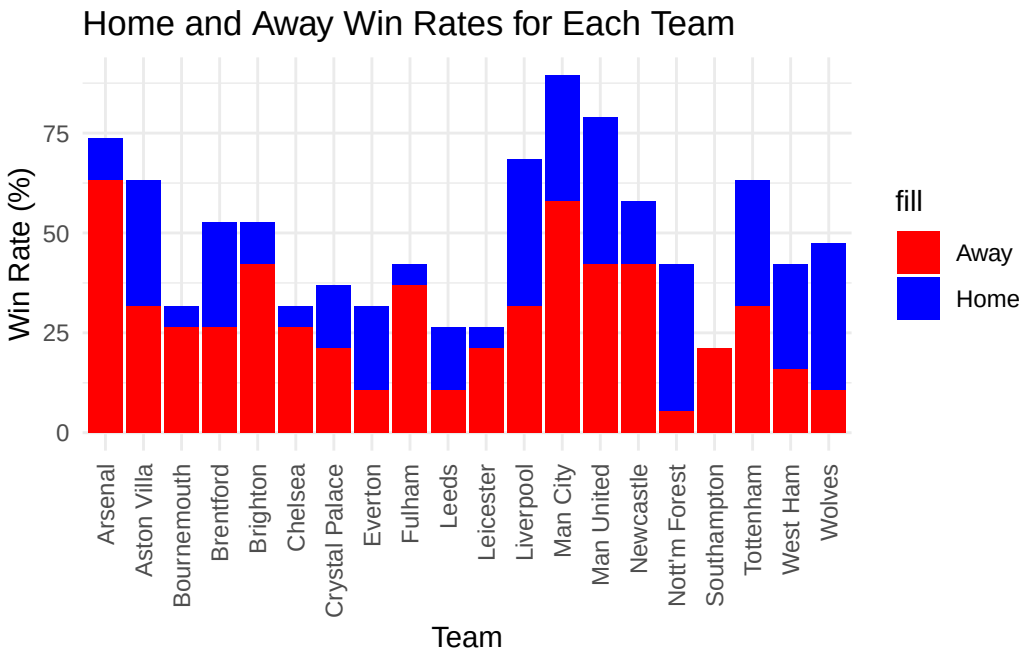


Figure 5: Home ad Away win rates for each team in 2023 season

We can see from the bar chart that there are differences in the relationships between the performance of the home and away teams and the presence of different referees. For example, given the limited sample size of games each referee officiates, some referees show a noticeable positive variance for home teams, which could indicate a "home bias" or simply be a coincidence. Other referees, on the other hand, seem to have a detrimental effect on home teams' winning percentages. Additionally, other referees' impacts are more evenly distributed, with no variation for either the home or away team, according to the chart, indicating that the location of the game may not have a major influence on the result of the match.

## 3 Model

The statistical model used in this part of the analysis evaluates how specific home game factors affect Premier League scoring outcomes using a Bayesian framework. The model specifically looks into the link between two important variables, home team shots on target and home team corners, and the total number of goals scored by the home team (full time home team goals). This link points to the need to concentrate on evaluating the ways in which offensive plays, like as corner kicks and shots on goal, affect the scoring dynamics of home games.

According to the model, the home team's goal totals have a normal distribution, with the predictors' linear relationship to the distribution mean $\mu_i$. A relationship of this kind can be expressed as follows: $\mu_i = \beta_0 + \beta_1$ * Home Team Shots on Target + $\beta_2$ * Home Team Corners,

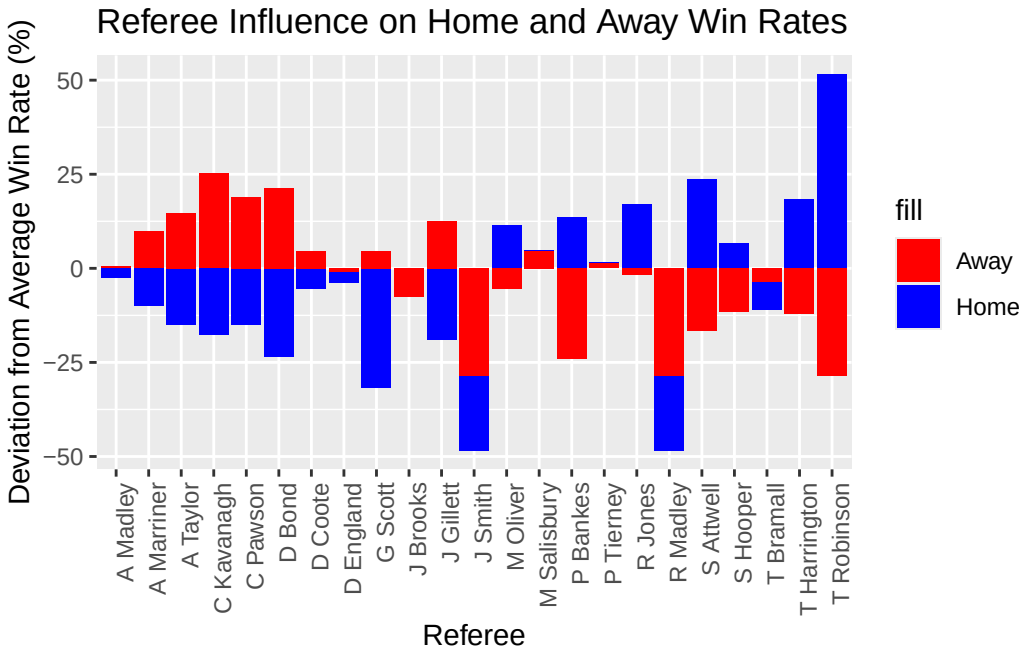


Figure 6: The relationship between refuree and win rate

where $\beta_0$ is the intercept and $\beta_1$ and $\beta_2$ are the coefficients that indicate how each predictor affects the scoring result. These coefficients are important because they show how, when other things stay the same, the projected change in the number of goals scored per unit increase in shots on target and corners will be. $\beta_0$ with a zero mean, and $\beta_1$ with 10 means to avoid the negative of goals, also a 2.5 standard deviation, the model's priors for these coefficients and the intercept are normal distributions, which represent a balanced belief in the available prior knowledge about these parameters. ## Model set-up

$$y_i|\mu_i \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \beta_0 + \beta_1 x_i \tag{2}$$
$$\beta_0 \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta_1 \sim \text{Normal}(10, 2.5) \tag{4}$$
$$\sigma \sim \text{Exponential}(1) \tag{5}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2023). We use the default priors from `rstanarm`.

### 3.0.1 Model justification

The particulars of football data and the research questions asked make the choice of a Bayesian linear regression model appropriate for this investigation. Precisely expressing the degree of uncertainty in forecasts is essential in sports analytics, and the Bayesian framework performs admirably by combining observational data with past knowledge. This method is especially helpful for assessing how shots on target and corners might affect the result of scoring in home games. The model can adjust based on the data from the current season since normal priors, which are centered at zero, guarantee that there is no initial bias towards positive or negative impacts.

Additionally, the exponential prior set for the residual standard deviation accounts for the possibility of occasionally higher variances that are typical in sporting situations, with an expectation of normally moderate departures from expected values. This configuration takes into account the unpredictable aspects of football games that are missed by traditional measurements. Using this model allows us to calculate the effects of home advantage elements and gain a grasp of the uncertainty associated with these calculations, both of which are important for making strategic decisions in sports-related situations.

## 4 Results

The visualizations from Figure 7 presented illustrate the positive correlation between the number of goals scored in Premier League matches and the home team's shots on target. This correlation is the outcome of the Bayesian linear regression model. A scatter plot with a regression line superimposed on top illustrates this relationship by showing that there is a tendency for the number of goals scored to rise in tandem with the home team's shots on target. The model's coefficient for home team shots on target, which is 0.362, further quantifies this tendency by suggesting that every extra shot on target is linked to a fractional rise in the expected number of goals scored by the home team.

On the other hand, from the model summary Table 3 it shows the coefficient for home team corners is -0.061, indicating a marginally negative correlation between the total number of corners and the home team goals at the end of the game. The model's R-squared value of 0.364 means that roughly 36.4% of the variability in the home team's goals can be explained, which is a significant percentage considering the wide range of variables influencing football match outcomes. The R-squared adjusted score and the log-likelihood number, which is -586.446 and provides a gauge of the model's general plausibility in light of the observed data, both further reinforce the model's predictive power and fit. The findings offer valuable insights into the attacking dynamics of Premier League home games, as shots on target show up as a stronger indicator of successful scoring than the quantity of corner kicks.

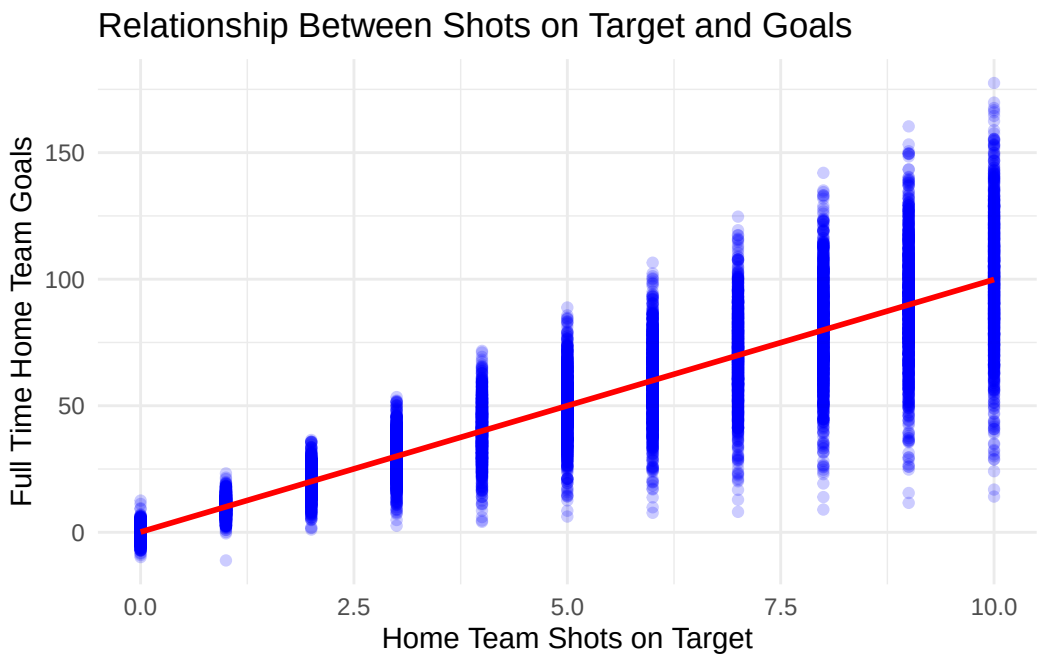


Figure 7: The relationship of Full Time Home Team Goals on the Home Team shot on Target and Home Corners

Table 3: Explanatory models of Premier League Full Time Home Team Goals on Home Team Shots on Target and Home Team Corners

| | (1) |
|---|---|
| (Intercept) | 0.199 |
| Home Team Shots on Target | 0.362 |
| Home Team Corners | −0.061 |
| Num.Obs. | 380 |
| R2 | 0.364 |
| R2 Adj. | 0.355 |
| Log.Lik. | −586.446 |
| ELPD | −590.1 |
| ELPD s.e. | 17.5 |
| LOOIC | 1180.3 |
| LOOIC s.e. | 35.0 |
| WAIC | 1180.3 |
| RMSE | 1.13 |

# 5 Discussion

## 5.1 Teams are more confident to shoot on targeet at home field

The primary goal of this paper is to investigate if home advantage exists for the League Premier League in football games. The study's findings demonstrate the critical role that shooting plays in a team's success, and they also show a positive relationship between goals scored and a team's capacity to make shots on goal. It demonstrates that making shots on target not only increases one's edge in the game but also plays a significant role in gaining home field advantage. This research can give the squad some excellent training direction so that players can hone their shooting techniques and gradually dismantle the opponent's defensive scheme. Coaches of the squad should therefore think about providing training to enhance players' ability to create and convert offensive possibilities. The team's shooting statistics can also be used to assess the team's offensive efficiency and trends, which can help opponents develop focused defensive strategies, as well as establish whether the team is attacking in a methodical manner. According to the data, the 20 teams' average number of shots on target while playing at home is higher than when playing away. This indicates that the team can attack more effectively when playing at home, and it also subtly demonstrates that having a home court undoubtedly has certain benefits.

## 5.2 Teams Corners may not be considered home advantage

Corner kicks are a important offensive tactic in modern football. Corner kicks are necessary to open up a situation when there is a heated situation in the game or when the score is tied or close. Even though some teams aren't as good as their opponents overall in many games, they may always prevail when they get corner kicks at massive times. This article will examine the effect of corner kicks on the game to see whether you can improve your performance at home, as corner kicks have the greatest overall winning rate among different set kicks. According to the data, the impact of corner kicks on scoring in home games is surprisingly minor. This result casts doubt on the conventional wisdom of the game, which holds that corners present important chances for scoring. This essay obviously does not cover the opponent's defensive strategy, corner kick execution quality, or psychological pressure. However, these same characteristics might also account for the fact that corner kicks have no statistically significant impact on the outcome of the game. To fully comprehend the connection between corner kicks and the game, more research is needed on two of these: tactical arrangement and the impact of corner kick execution.

## 5.3 Referee and teams advantage

Football forums have a strong tradition of discussing home advantage, and referees are a vital part of this discourse. Almost every team performs better at home than away, according to

the research. According to a different set of data, certain referees significantly affect both the home and away game winning percentage. The team's penalty given by the referee throughout the match will significantly affect how the game turns out. The happiness of the home crowd with the referee will also be a factor. The outcome of a set kick, like a corner kick, can alter the direction of play. Evidently, referee prejudice has a significant negative effect on the team. The referee may typically impact the team's strategy by allowing them to choose between more aggressive attacking tactics and tougher defensive techniques based on the referee's penalty scale. VAR and other cutting-edge football technology can reduce the arbitrators' subjectivity. Sometimes the system will make a mistake, but the referee's unaided vision won't. Thus, it is true that referees have the power to influence the result of a game, particularly when biased officials are assigned to a team. The game will have easy control over how the game turns out.

## 5.4 Weaknesses and next steps

### 5.4.1 Weaknesses

The potential limitation of the data set itself, which might not include all the subtle factors that affect home game performance—like player fatigue, team morale, or particular game circumstances like the weather or the significance of the game—is one of the study's weaknesses. Additionally, the model makes the assumption that predictors and outcomes follow a linear relationship, which could oversimplify the intricate dynamics of football games. Additionally, the effect of corner kicks (as indicated by the negative coefficient) begs the question of whether more complex measurements are required to accurately represent set-piece efficiency or if the quality of the underlying data is sufficient. The model's ability to explain things is another flaw. For a sport as unpredictable as football, an R-squared value of about 36% is rather large, but the model is still unable to explain a sizable fraction of the variation. This highlights the need for more data collecting and analysis techniques by implying that additional unmeasured factors might be important in deciding match outcomes.

### 5.4.2 Next Steps

The data set would need to be expanded to cover additional seasons and possibly more detailed information at the player or play level in order to address these shortcomings. Compiling information on player performance, tactical alignments, and in-game occurrences like possession shifts could enhance the examination and facilitate a more understanding of the elements that lead to home advantage. Future study may employ more advanced statistical models, such as machine learning algorithms that may capture non-linear connections and interactions between variables or hierarchical models that account for team-specific effects, in terms of analytically methodologies. These additions would not only increase the findings' robustness but also offer clubs wanting to maximize their home advantage or offset the home team's advantages when

playing away useful insights. Further investigation into this topic may also result in a better comprehension of how psychological elements such as referee bias affect the results of matches and the steps that may be taken to guarantee fair play.

## References


Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. https://tellingstorieswithdata.com/.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Football-Data.co.uk. 2023. "Football-Data.co.uk - Premier League Data." urlhttps://www.football-data.co.uk/englandm.php.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2023. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://tibble.tidyverse.org/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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 Grolemund, 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, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.