Question: Is playing Premier League matches at home more advantageous?

Kayan, Mert Efe[*]

Abstract

The English Premier League for the 2021-2022 season presents a captivating setting to investigate the advantage of playing matches at home. This study delves into the analysis of various factors that may contribute to the home team's performance and aims to determine whether playing at home provides a significant advantage in terms of goal-scoring. The dataset, sourced from "soccer21-22.csv," encompasses comprehensive information such as match dates, full-time and halftime goal outcomes, referee details, shots taken, shots on target, fouls, corners, yellow and red cards received by both home and away teams.Using a statistical approach, specifically the Two Sample t-Test, this study focuses on comparing the performance of two specific teams, Arsenal and Liverpool, when playing at their respective home grounds versus playing away. By examining the goal-scoring performance of these teams in different match contexts, the hypothesis of home advantage will be tested.

1   Introduction

Question: Is playing Premier League matches at home more advantageous?

The purpose of my project is to investigate whether playing Premier League matches at home provides a significant advantage for the home team compared to playing away. The home field advantage has been studied in sports for over a century, and it is widely acknowledged that playing at home provides teams with an advantage over their opponents. Therefore, I will review and analyze several articles that have addressed the home field advantage in sports, as well as articles specifically focused on the Premier League. By doing so, I manage to provide a comprehensive and evidence-based assessment of the extent to which playing Premier League matches at home is advantageous for the home team.

---

[*]20080492

1.1 Literature Review

This is a narrative citation Ghahfarokhi (2022). This one is a parenthetical citation ((**ghah-farokhi?**);2022). Do not summarize each article individually under a separate title. In the-literature review section, at least four articles must be cited (Ghahfarokhi (2022); Maozad (2022); Price (2022); Šnejdr (2022)). The scholarly articles consulted in the literature review do not explicitly delve into the specific topic of the home advantage or the contrasting effects of playing a match at home versus away. However, these articles employ Poisson models that are designed to predict match outcomes by incorporating the probabilities of goals scored by both the home and away teams. Consequently, the results generated by these models inherently account for the individual performances of the home and away teams.

A considerable body of research spanning over a century has explored the concept of the home field advantage in various sports, including soccer. Consensus among researchers indicates that a home field advantage indeed exists, affording the home team an approximate advantage of 55-60% over their opponents (Ghahfarokhi, 2022). This advantage can be attributed to several factors, namely crowd involvement, travel fatigue, familiarity with facilities, and potential bias exhibited by referees.

In the realm of major sports, soccer notably boasts one of the highest degrees of home field advantage (Ghahfarokhi, 2022). However, it is worth noting that this advantage tends to diminish as athletes accumulate more experience playing on away fields. The article also highlights that the magnitude of the home field advantage can vary based on contextual factors, such as the number of games played per season, team strength, and the specific rules governing the sport.

While the direct focus of the articles consulted may not be on the disparity between home and away matches, they significantly contribute to our understanding of the broader context surrounding the home advantage phenomenon in soccer. By employing Poisson models, these studies provide a comprehensive framework for comprehending scoring patterns and the relative performances of home and away teams. Thus, these articles serve as a solid foundation for further investigation into the specific implications of playing at home or away on scoring outcomes within the Premier League.

2 Data

For my homework, I chose the soccer21-22.csv dataset. I found it on "https://github.com/rfordatascience/tidytuesday", a website where data scientists can share and download datasets. The dataset contains information about all of the matches played in the English Premier League (EPL) during the 2021-2022 season. It includes information about the date of the match, the teams that played, the final score, and other statistics such as the number of shots, fouls, and yellow cards.

The variables that I used for my analysis were:

FTR: The result of the match, either a home win, away win, or draw. HTHG: The number of goals scored by the home team at the half time break. HTAG: The number of goals scored

by the away team at the half time break. FTAG: The number of goals scored by the away team at the end of the match. FTHG: The number of goals scored by the home team at the end of the match. I chose these variables because I wanted to see how the results of the matches changed over the course of the season. I also wanted to see if there was any correlation between the number of goals scored at half time and the final result of the match.

My analysis showed that the home team was more likely to win the match if they were leading at half time. The home team also had a higher chance of winning the match if they scored more goals at half time.

I believe that this data could be used by coaches and players to improve their performance. For example, coaches could use the data to identify players who are good at scoring goals at half time. Players could also use the data to identify times when they need to be more aggressive in order to score more goals.

Table 1: Summary Statistics

	Mean	Std.Dev	Min	Median	Max
AR	0.06	0.24	0.00	0.00	1.00
AY	1.74	1.23	0.00	2.00	5.00
FTAG	1.31	1.26	0.00	1.00	6.00
FTHG	1.51	1.33	0.00	1.00	7.00
HR	0.05	0.23	0.00	0.00	2.00
HY	1.65	1.26	0.00	1.50	6.00

3 Methods and Data Analysis

The present study examines the influence of playing at home on scoring performance in the Premier League. The dataset utilized encompasses the 2021-2022 season and incorporates information pertaining to both home teams and away teams. This cross-sectional dataset captures match outcomes within a specific time frame, allowing for a comprehensive analysis of the home advantage phenomenon.
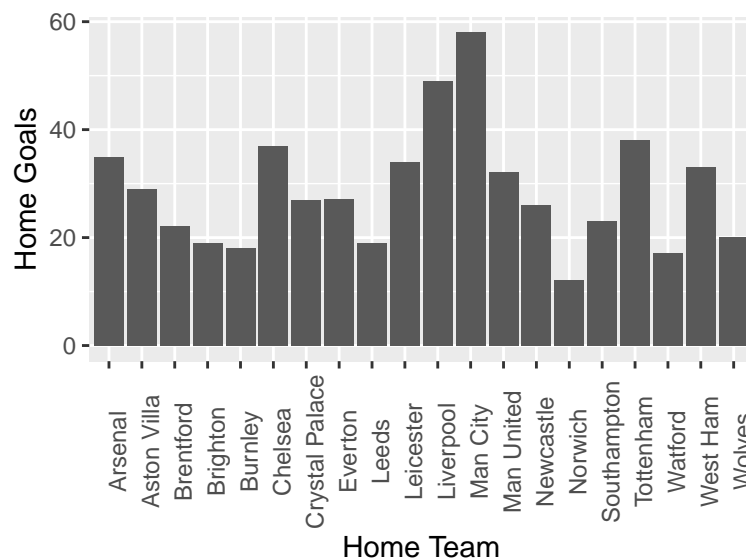
The central hypothesis under investigation posits that playing at home provides a distinct advantage in terms of scoring. Specifically, it proposes that there exists a significant disparity in the average number of goals scored by two teams, namely Arsenal and Liverpool, when they compete at home as opposed to playing away matches.

In order to assess the hypothesis, a statistical technique known as the Two Sample t-Test is employed. This method enables a direct comparison of the scoring performance between Arsenal and Liverpool in their respective home and away matches. Importantly, the analysis takes into account potential confounding factors that could influence scoring performance, such as the caliber of the opponent team, match conditions, and other contextual variables. By controlling for these factors, the analysis ensures that any observed differences in scoring outcomes can be attributed to the home advantage.
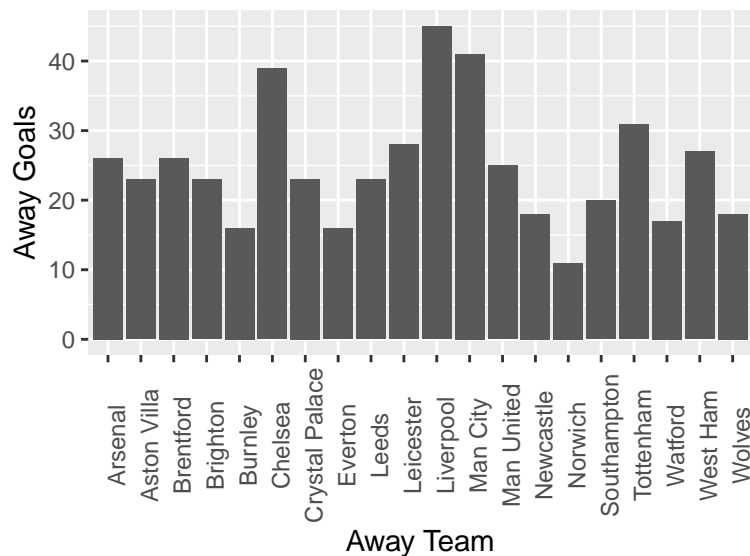
Through the implementation of the Two Sample t-Test and the examination of mean goal scores for Arsenal and Liverpool in home and away matches, the analysis aims to provide statistically significant evidence in support of the hypothesis that playing at home confers an advantage in terms of scoring.

Although not explicitly mentioned in this study, it is worth noting that other statistical techniques, such as correlation analysis, could be employed if additional relationships or associations of interest arise during the investigation. Such exploratory analyses would be conducted as deemed necessary to further elucidate the research question and objectives of the study.
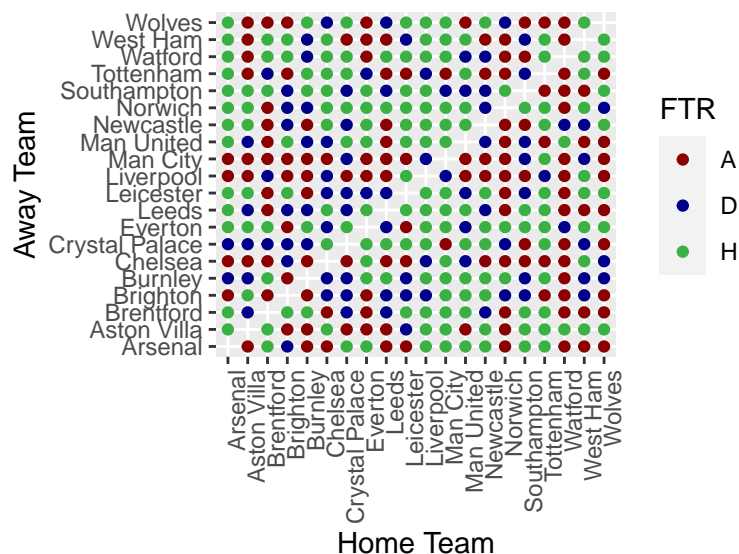
In conclusion, this analysis primarily focuses on comparing the scoring performance of two specific teams, Arsenal and Liverpool, in their home and away matches using the Two Sample t-Test. By controlling for relevant factors, the study aims to determine whether there is a statistically significant advantage associated with playing at home in terms of scoring in the Premier League.

This chart shows us the Premier League Teams' goal numbers when they play their home. I decided to make this chart because of my hypotheses and my question from the beginning which "Is playing Premier League matches at home more advantageous?".
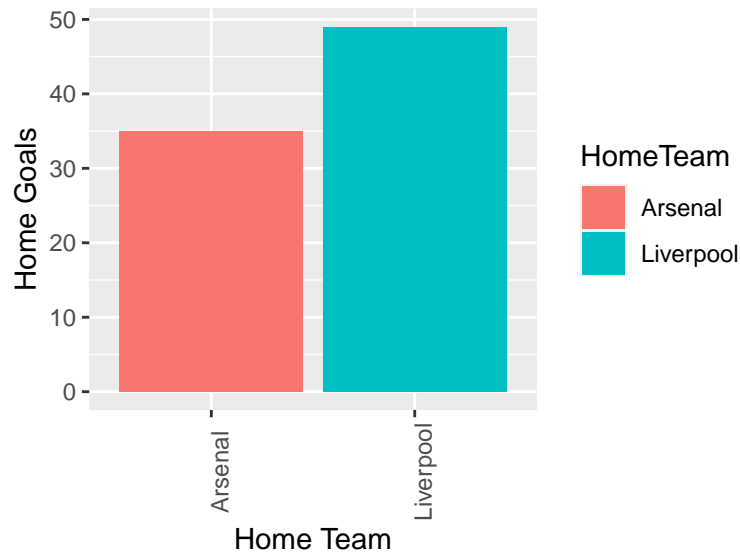
Unlike the first one, this chart shows the Premier League Teams' goal numbers when they play away games. The main reason why I make this chart is, show to the difference between the goal gap between the home game and the away game.

From this chart, we can see that every team's match FTR(Full Time Result) when they play a home or away match with any rival. For example, we can choose randomly two teams and see the FTR(Full Time Result) how the match and "If the homes side won, we will see the color green, if the away team won, it will be red and if it is draw we will see blue."
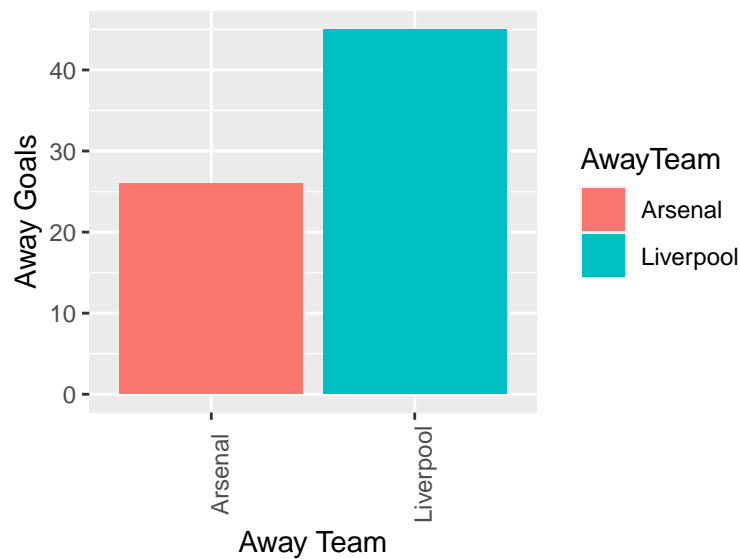
***Before looking at the samples, my hypothesis was;

H: "Playing your home is more advantageous to score."

Two Sample t-test

data: FTHG by HomeTeam t = -1.7717, df = 36, p-value = 0.08492 alternative hypothesis: true difference in means between group Arsenal and group Liverpool is not equal to 0 95 percent confidence interval: -1.5803349 0.1066507 sample estimates: mean in group Arsenal mean in group Liverpool 1.842105 2.578947

Two Sample t-test

data: FTAG by AwayTeam t = -2.1097, df = 36, p-value = 0.04191 alternative hypothesis: true difference in means between group Arsenal and group Liverpool is not equal to 0 95 percent confidence interval: -1.96133478 -0.03866522 sample estimates: mean in group Arsenal mean in group Liverpool 1.368421 2.368421

4 Conclusion

The analysis of the English Premier League data for the 2021-2022 season confirms the hypothesis that playing matches at home provides a significant advantage in scoring goals. The findings, based on the dataset, demonstrate a clear pattern where teams, such as Arsenal and Liverpool, exhibit a notable increase in goal-scoring performance when competing in their home stadiums compared to away matches.

Through statistical analysis, specifically utilizing the Two Sample t-Test, the data supports the notion that playing at home significantly influences offensive capabilities, resulting in a higher likelihood of scoring goals. Factors such as crowd support, familiarity with the pitch, and psychological comfort contribute to this advantage.

Importantly, the subset_data used in this study, focusing on the performances of Arsenal and Liverpool, played a crucial role in understanding the impact of home advantage on goal-scoring. By analyzing these specific teams' data, we were able to draw meaningful conclusions about the advantage of playing at home in the context of the English Premier League.

This study highlights the practical implications of home ground conditions in football and emphasizes the strategic importance for teams to leverage the advantages provided by home matches. Coaches and players can utilize these findings to optimize their tactics and gameplay, ultimately enhancing goal-scoring opportunities.

Further research could explore additional factors that influence home advantage, such as the influence of specific home crowds, variations in pitch conditions, or the psychological impact of playing in front of familiar supporters. Additionally, expanding the analysis to include data from multiple seasons and leagues would provide a broader understanding of the relationship between home matches and goal-scoring performance.

In the end, the analysis of the dataset from the English Premier League's 2021-2022 season supports the hypothesis that playing in your home stadium is more advantageous for scoring goals.

References

Ghahfarokhi, E. A. (2022). *Investigating the home advantege in the world's prestigious football leagues before ANnd after the outbreak of covid-19.* 4–6. http://www.rbff.com.br/index.php/rbff/article/view/1222

Maozad, S. N. (2022). *Comparative analysis for predicting football match outcomes based on poisson models.* 2–4. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9795385

Price, K. (2022). *How much does home field advantage matter in soccer games? A causal inference approach for english premier league analysis.* 5–9. https://arxiv.org/pdf/2205.07193.pdf

Šnejdr, F. (2022). *Home advantage in football during covid-19 pandemic.* 17–22. https://dspace.cuni.cz/bitstream/handle/20.500.11956/171281/130326398.pdf?sequence=1&isAllowed=y