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

In this section, discuss the articles you have read on the subject by giving references. This is a narrative citation Ghahfarokhi (2022). This one is a parenthetical citation ((**ghahfarokhi?**);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 articles does not directly refer to the home advantage or the difference that playing a match at home or away could make. However, the Poisson models used in the article are structured to predict the outcomes of matches by taking into account the probabilities of the home and away teams scoring goals. Therefore, the results of the models take into consideration the performances of the home and away teams separately. The home field advantage has been studied for over a century, and there is general consensus that it exists and gives the home team an advantage of around 55-60% over their opponents. Four factors contribute to this advantage: crowd involvement, travel fatigue, familiarity of facilities, and referee bias. Soccer has one of the highest home field advantages among major sports, and the advantage tends to deteriorate as athletes play longer on an away field. The article also discusses how home field advantage can vary based on the number of games per season, the strength of the teams, and the rules of the sport.

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 dataset used for the analysis includes the Premier League season from 2021-2022, focusing on information related to home teams and away teams. The dataset is cross-sectional, as it captures the outcomes of matches played within a specific period.

To investigate whether playing at home provides an advantage in terms of scoring, a statistical technique called the Two Sample t-Test is employed. Specifically, the analysis focuses on comparing the performance of two teams, Arsenal and Liverpool, in their home and away matches.
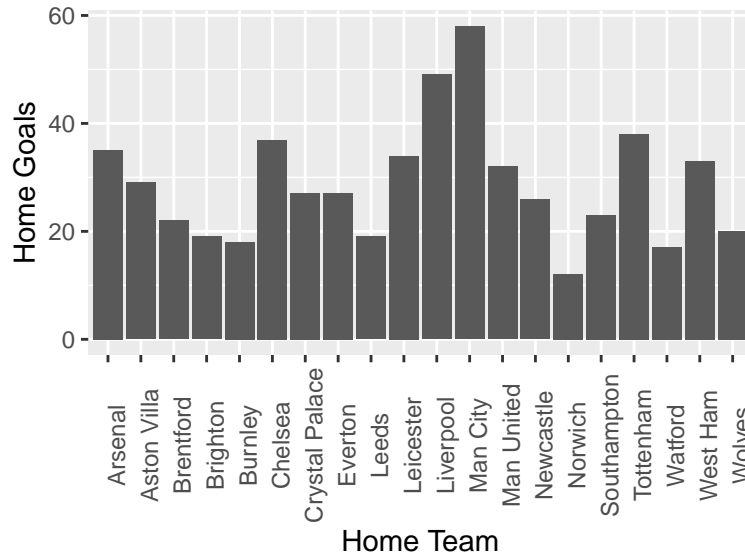
The main hypothesis to be tested is whether playing at home leads to a higher-scoring performance. The hypothesis states that there is a significant difference in the average goals scored by Arsenal and Liverpool when they play at home compared to when they play away.

The analysis controls for other variables or factors that may influence the scoring performance, such as the opponent team, match conditions, or other contextual factors. These control variables ensure that any observed differences in scoring between home and away matches can be attributed to the home advantage rather than other confounding factors.
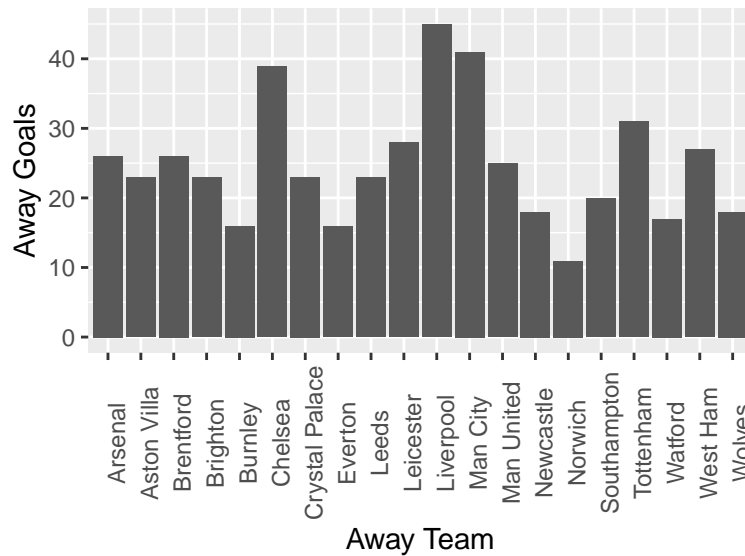
By conducting the Two Sample t-Test and comparing the mean goals scored by Arsenal and Liverpool in their home and away matches, the analysis aims to determine if there is statistically significant evidence to support the hypothesis that playing at home is more advantageous in terms of scoring.

Other statistical techniques, such as correlation analysis, are not explicitly mentioned as relevant to the research question and objective of the study. However, if there are additional relationships or associations of interest, they can be explored using appropriate techniques as needed.
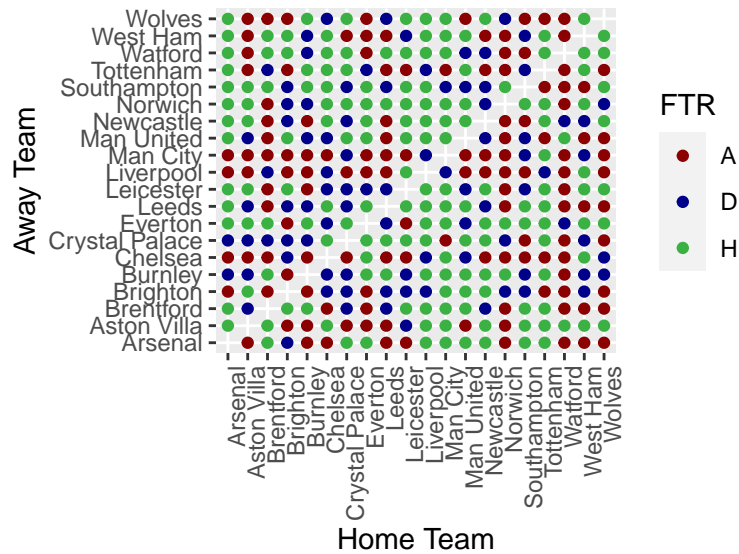
Overall, the analysis focuses on comparing the scoring performance of two specific teams, Arsenal and Liverpool, in their home and away matches using a Two Sample t-Test and controlling for relevant factors. The goal is to determine whether there is a statistically significant advantage to playing at 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?".
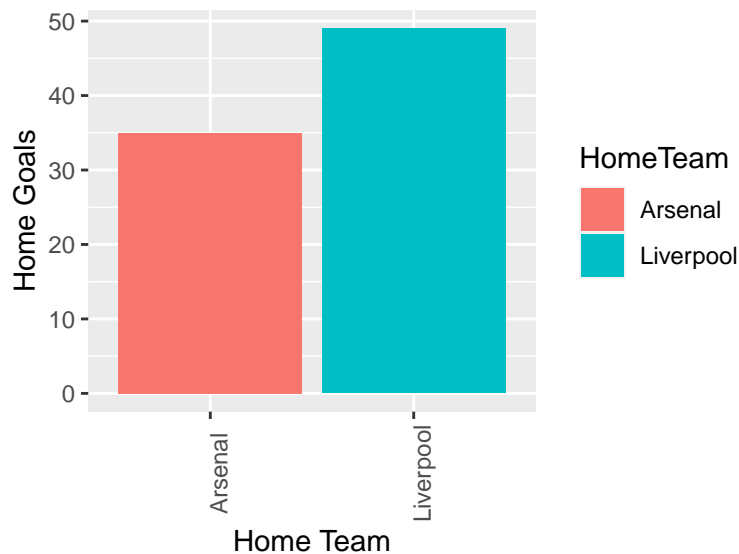

Unlike the first one, this chart shows the Premier League Teams' goal numbers when they play away games. The main reason why I make this chart is, show to the difference between the goal gap between the home game and the away game.

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."
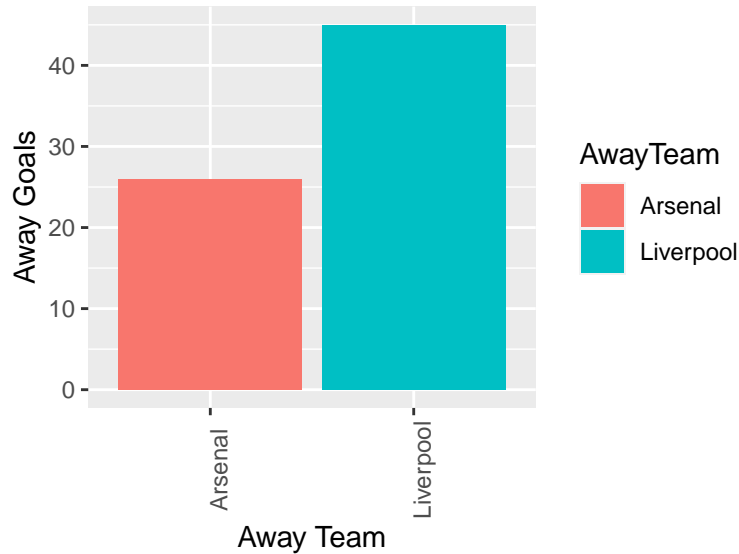
***Before looking at the samples, my hypothesis was;

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