Home Team Advantage in International Soccer

Michael Nichols, Evan Weller

2024-12-11

Table of contents

Introduction:

Research Question

Does a 'home field advantage' exist in international soccer?

Specifically, is there a measurable advantage for home teams in terms of the percentage of matches they win?

Data Source

The data, *International Football Results from 1872 to 2024*, was sourced from Kaggle. This data set contains the results and statistics from over 47,000 international soccer matches from as far back as 1872.

Provenance

The dataset was curated by Mart Jürisoo to analyze trends and performance in international football. The data was sourced from multiple reputable platforms, including Wikipedia, rsssf.com, and official websites of individual football associations. It comprises over 47,000 international soccer matches from more than 200 nations, encompassing friendlies, qualifiers, and tournament matches. Notably, the dataset does not include Olympic matches.

FAIR Principles

The dataset aligns with the FAIR principles:

- **Findable:** Available on Kaggle with well-documented metadata for easy discovery.
- **Accessible:** Freely available for download in CSV format.
- **Interoperable:** Designed to work seamlessly with standard data analysis tools.
- **Reusable:** Provided with clear licensing, allowing for academic and non-commercial use.

CARE Principles

The dataset does not include sensitive or personal information, minimizing ethical concerns. It is based entirely on publicly available sports data, ensuring compliance with ethical data usage standards.

Results and Discussion

Home Team Win Percentage

Figure 1 shows the win percentage of home teams across all matches, revealing that home teams win approximately **50.72%** of the time.

At first glance, this may not seem like a significant "home team advantage." However, when considering the outcomes of matches—home team wins, away team wins, and draws—a clearer pattern emerges. Away teams win about **26.4%** of the time, while draws occur in **22.88%** of matches. This indicates that when a team plays on their home field, they are nearly twice as likely to win as they are to lose to the away team.

While the data demonstrates a clear home-field advantage in terms of **avoiding defeat**, it does not necessarily translate to a strong advantage for **outright winning** the game, as the home team still fails to win in roughly half of all matches.

Average Goals Scored

Figure 2 compares the average goals scored by home teams, which is **1.78**, with the average goals scored per team in matches, calculated as **1.45**.

These averages are derived using different methods. The average goals scored by home teams is calculated by taking the mean of the "home team goals" column. To determine the average goals per team in a match, the total goals scored in a match, which averages **2.92**, is divided by two to account for both teams.

The data highlights an increase of **0.33 goals** in favor of home teams compared to the per-team average. While this difference may appear minor, it is derived from a dataset of approximately **47,000 games**, making the margin statistically significant. This difference is particularly significant in the context of soccer, a sport where matches generally feature fewer goals than many others. The increased scoring by home teams supports the idea of a "home team advantage," especially in terms of goal production, even if the overall margin is relatively small.

Data Visualizations

Table 1: Football Match Results (Filtered)

home_team	away_team	home_score	away_score	outcome	date
Scotland	England	0	0	Draw	1872-11-30
England	Scotland	4	2	Home Win	1873-03-08
Scotland	England	2	1	Home Win	1874-03-07
England	Scotland	2	2	Draw	1875-03-06
Scotland	England	3	0	Home Win	1876-03-04
Scotland	Wales	4	0	Home Win	1876-03-25
England	Scotland	1	3	Away Win	1877-03-03
Wales	Scotland	0	2	Away Win	1877-03-05

home_team	away_team	home_score	away_score	outcome	date
Scotland	England	7	2	Home Win	1878-03-02
Scotland	Wales	9	0	Home Win	1878-03-23

Figure 1: Home Team Win Percentage

A bar plot of match outcomes (win/loss/draw) categorized by match type.

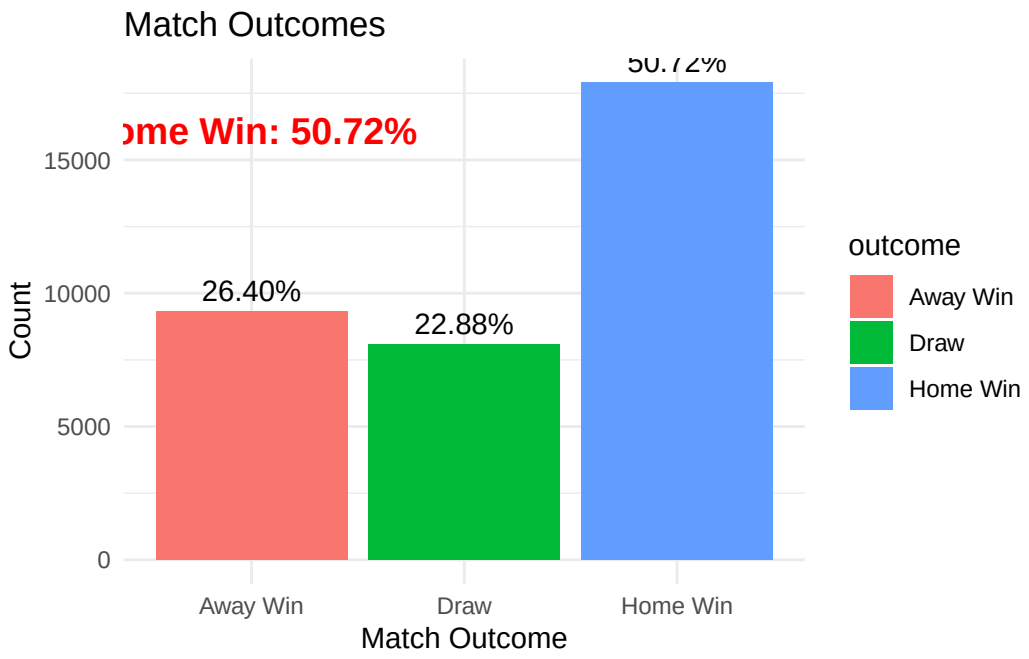
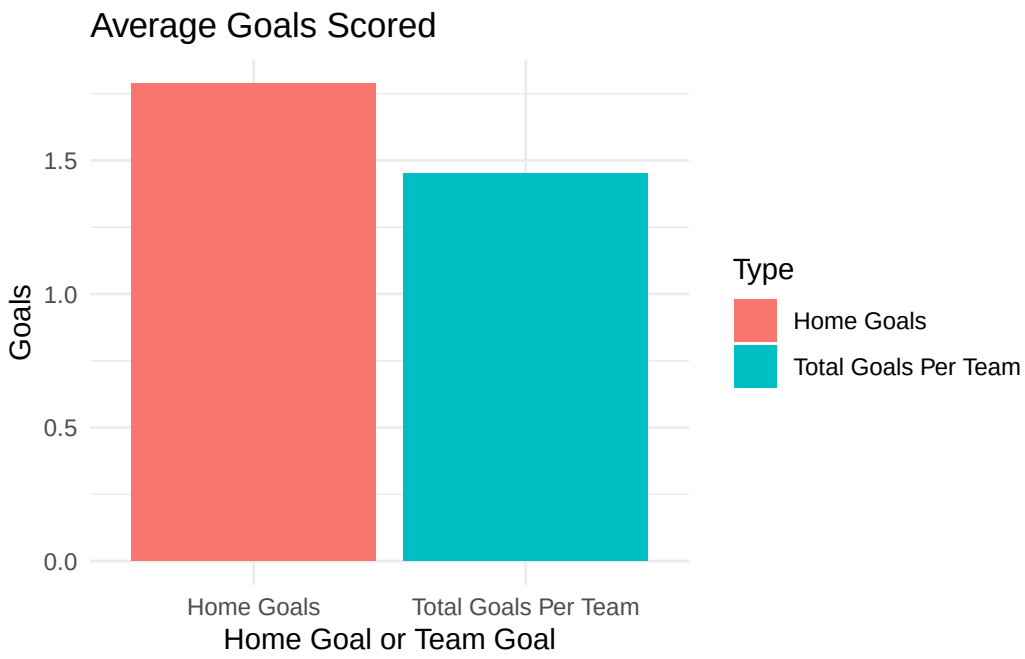


Figure 2: Average Goals Comparison

A side-by-side bar chart comparing home goals and total match goals.



Conclusion

The data reveals a clear "home team advantage" in soccer, demonstrated by the fact that home teams are twice as likely to avoid losing compared to away teams and tend to score more goals on average. While every match is influenced by the skill and effort of the players, these findings suggest that playing at home provides a meaningful edge.

Factors not directly measurable in this dataset, such as the impact of a supportive home crowd, likely contribute to this advantage. A home stadium filled with cheering fans can boost a team's morale and motivation, as many athletes attest to the energizing effect of crowd support. Additionally, the familiarity of practicing and competing on a home field may also play a role in enhancing performance. These intangible factors, combined with the statistical trends, reinforce the existence of a significant home-field advantage in soccer.

Appendix: Code

```r
#|label = Cleaning Data
# Load libraries

library(tidyverse)
library(knitr)

# Load data
football_data <-read_csv("results.csv")

# Filter out neutral matches
football_clean <- football_data %>%
  filter(neutral == FALSE)

# Add match outcome column
football_clean <- football_clean %>%
  mutate(
    outcome = case_when(
      home_score > away_score ~ "Home Win",
      home_score < away_score ~ "Away Win",
      TRUE ~ "Draw"
    )
  )

# Home win percentage
home_win_rate <- as.integer(mean(football_clean$outcome == "Home Win") * 100)

# Average goals
avg_home_goals <- mean(football_clean$home_score)
avg_total_goals_per_team <- mean((football_clean$home_score + football_clean$away_score)/2)
football_clean %>%
  select(home_team, away_team, home_score, away_score, outcome, date) %>%
  head(10) %>%  # Display the first 10 rows for clarity
  kable(caption = "Football Match Results (Filtered)")
#|Label = Figure 1
library(ggplot2)
library(dplyr)

# Calculate percentage of home wins
win_data <- football_clean %>%
  mutate(total_matches = n()) %>%
  count(outcome, total_matches) %>%
```

```r
  mutate(percentage = (n / total_matches) * 100)

# Extract the percentage of home wins for annotation
home_win_percentage <- win_data %>%
  filter(outcome == "Home Win") %>%
  pull(percentage)

# Create the bar plot with annotation
win_plot <- win_data %>%
  ggplot(aes(x = outcome, y = n, fill = outcome)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Match Outcomes",
    x = "Match Outcome",
    y = "Count"
  ) +
  theme_minimal() +
  geom_text(aes(label = sprintf("%.2f%%", percentage)), vjust = -0.5, size = 4) +  # Add perc
  annotate(
    "text",
    x = 1, y = max(win_data$n) * 0.9,  # Position near the top of the Home Win bar
    label = sprintf("Home Win: %.2f%%", home_win_percentage),
    color = "red",
    size = 5,
    fontface = "bold"
  )

# Save the plot
ggsave("win_plot_with_percentage.pdf", win_plot, width = 8, height = 7)

print(win_plot)
#|Label = Figure 2
# Plot 2: Average goals comparison
goals_plot <- tibble(
  Type = c("Home Goals", "Total Goals Per Team"),
  Average = c(avg_home_goals, avg_total_goals_per_team)
) %>%
  ggplot(aes(x = Type, y = Average, fill = Type)) +
  geom_bar(stat = "identity") +
  labs(
    title = "Average Goals Scored",
    x = "Home Goal or Team Goal",
```

```
    y = "Goals"
  ) +
  theme_minimal()
ggsave("goals_plot.pdf", goals_plot, width = 8, height = 6)
print(goals_plot)
```