

Exploring Attendance and Performance Trends in Women's Super League*

An analysis of match attendance, key performance factors, and trends in the English Women's Football

Rayan Awad Alim

December 3, 2024

This paper uses data from the English Women's Football (EWF) Database to explore trends in match attendance and the factors influencing match outcomes in the Women's Super League. We found that attendance has generally increased over time, with major events acting as catalysts for spikes in attendance. Additionally, match attendance seems to influence match outcomes, suggesting that crowd support can impact team performance. These findings provide useful insights for teams, analysts, and league organizers regarding fan engagement and team strategies..

1 Introduction

The Women's Super League (WSL) has seen significant growth since its inception, both in terms of attendance and popularity. This paper aims to explore trends in WSL attendance and identify key performance factors that determine match outcomes, using the English Women's Football (EWF) Database. Understanding these trends is essential for teams, analysts, and league organizers seeking to enhance fan engagement and improve team performance.

The estimand of this study is to determine how factors like attendance, team strength, and historical performance affect match outcomes in the Women's Super League. By quantifying these relationships, we aim to provide insights into what drives successful match results and how attendance figures have evolved over time.

The analysis finds that attendance has generally increased over time, with certain key events serving as catalysts for spikes in crowd size. Additionally, higher attendance is correlated with

*Code and data are available at: [<https://github.com/RayanAlim/EnglishWomensFootballAnalysis>].

an increased likelihood of home team success, suggesting a potential impact of crowd support on team performance.

This research matters because it helps identify the factors that contribute to successful outcomes in women’s football, providing valuable information for teams to improve strategies, for organizers to boost engagement, and for analysts interested in understanding sports dynamics.

The remainder of this paper is structured as follows. Section ?? discusses the data sources and cleaning processes. Section ?? outlines the model used to evaluate match outcomes. Section ?? presents the key findings from the data analysis, and ?@sec-discussion provides a summary of what we have learned and suggests potential areas for future research.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) to perform our analysis. Our data is derived from the English Women’s Football (EWF) Database, which provides a comprehensive dataset of matches, team appearances, and standings in the Women’s Super League and Women’s Championship. Following the guidance provided by Alexander (2023), we consider how best to prepare and use these data for analysis in order to effectively tell a story of attendance and performance trends.

This study utilizes three main datasets from the EWF Database:

ewf_matches: Contains all matches played with details like attendance, score, and outcomes.

ewf_appearances: Contains information about team appearances in each match.

ewf_standings: Contains end-of-season standings for each team.

2.2 Measurement

The process of measurement involves translating real-world events into structured data entries. In the context of our analysis, this means taking observable phenomena such as the attendance of a football match, the outcome of a game, and individual team performances and converting them into numerical or categorical data points.

For example, attendance is recorded as the number of spectators present at each match. This variable captures the level of audience engagement and is an indicator of the popularity of the match. Attendance figures are sourced from official league records, ensuring reliability. However, missing values in the attendance data required filtering to maintain consistency in the analysis.

Similarly, match outcomes are recorded as categorical variables, with values such as “Home Win,” “Away Win,” and “Draw.” These categories are derived directly from match results and are used to evaluate performance trends. This structured approach allows us to quantify and model the likelihood of different outcomes based on a variety of predictors.

The team standings data, recorded at the end of each season, includes metrics such as points earned, goals scored, and final league position. These measurements help to contextualize team performances over multiple seasons and allow for comparative analysis between teams and over time.

2.3 Outcome variables

The outcome variables of interest in this study include match attendance and match outcomes. To understand these outcomes comprehensively, we provide graphical and tabular representations.

2.3.1 Attendance

The attendance variable captures the number of spectators present at each match. This outcome variable is important as it reflects the engagement level of fans and has potential implications for team performance. Plotting the trend of attendance over the seasons reveals notable spikes in attendance that often align with major international events, indicating the impact of broader football activities on domestic league engagement (Figure ??).

```
# Fixing the column reference to correct data
ewf_matches_cleaned <- cleaned_data %>%
  mutate(date = as.Date(date, format="%Y-%m-%d"))

attendance_plot <- ggplot(ewf_matches_cleaned, aes(x = date, y = attendance)) +
  geom_line(color = "blue") +
  labs(title = "Attendance Over Time in Women's Super League", x = "Date", y = "Attendance") +
  theme_minimal()

attendance_plot
```

Table 1: ?(caption)

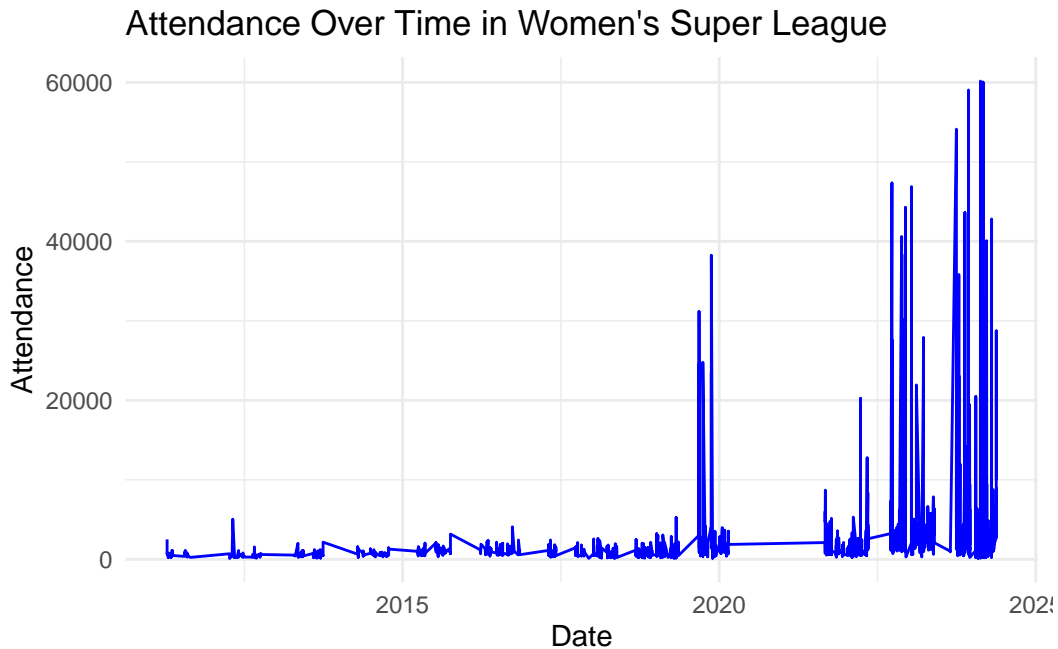


Figure 1: Attendance Over Time in Women's Super League, showing trends and notable spikes during key events.

2.3.2 Match Outcomes

The match outcomes variable is a categorical outcome representing whether the home team won, the away team won, or if the match ended in a draw. This outcome helps to assess the impact of various predictors, such as attendance, on the likelihood of different results.

```
# Summarizing match outcomes
match_outcome_summary <- cleaned_data %>%
  group_by(result) %>%
  summarise(
    avg_attendance = mean(attendance, na.rm = TRUE),
    total_matches = n()
  )

# Display table using tinytable for better formatting
# tinytable::tt(match_outcome_summary, colnames = TRUE, rownames = FALSE)
```

2.4 Predictor variables

This section discusses the predictor variables used in the model. These predictors include:

1. Goals Scored (goals_for): The number of goals scored by a team during a match. Goals are key indicators of team performance and are directly related to the likelihood of winning a match.
2. Goals Against (goals_against): The number of goals conceded by a team. Fewer goals conceded generally indicates a stronger defense and contributes to better match outcomes.
3. Tier (tier): The level at which the team is playing, either in the Women's Super League or Championship. Teams at different tiers may show varied performance due to differences in competitiveness.

```
# Load correct dataset for predictor variables
predictor_data <- read_parquet(here::here("data", "02-analysis_data", "ewf_appearances_clean.parquet"))

# Plotting distribution of goals scored by tier
predictor_plot <- predictor_data %>%
  filter(!is.na(goals_for), !is.na(tier)) %>%
  ggplot(aes(x = goals_for, fill = as.factor(tier))) +
  geom_histogram(binwidth = 1, alpha = 0.7, position = "dodge", color = "black") +
  labs(title = "Distribution of Goals Scored by Tier", x = "Goals Scored", y = "Frequency")
  theme_minimal()

predictor_plot
```