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 analyzes attendance and performance trends in the Women's Super League (WSL) using data from the English Women's Football (EWF) Database. We examine the evolution of attendance over time and evaluate how key factors—such as goals scored, goals conceded, and league tier—influence both match-level outcomes and season-level team performance. Using regression modeling, we find that match attendance has grown substantially, correlating with higher probabilities of home-team success, suggesting a measurable 'home advantage.' These insights provide actionable information for enhancing fan engagement strategies and guiding team development, especially as the intrest in womens sports increases globally.

Table of contents

*Code and data are available at: github.com/RayanAlim/EnglishWomensFootballAnalysis/.

1 Introduction

Women's professional football in England (and globally) has experienced substantial growth in recent years, notably reflected in the Women's Super League (WSL). Increased media coverage, enhanced professionalism, and international success of national teams have all contributed to rising attendance and interest Eekeren (2022) Kitching (2022). Understanding the determinants of match attendance and performance outcomes is valuable for clubs, team analysts, league organizers, and stakeholders seeking to enhance fan engagement, team performances and competitive standards as demonstrated by Garcia (2002), attendance determinants can be significantly influenced by local factors.

This paper explores two central themes: the evolution of match attendance over time and the relationship between key performance indicators (e.g., goals scored, goals conceded) and both match-level outcomes (win, draw, loss) and season-level performance (final standings, points accrued). Using the English Women's Football (EWF) Database(RobC (2024))- a historical record of top-tier and second-tier women's football in England- we analyze how attendance trends correlate with on-field results and whether home teams benefit from larger crowds.

The estimand is to understand how factors like attendance, goals, and tier level influence match results and team points. Our findings reveal that attendance has steadily increased, especially around peak events and growth phases of the league. Higher attendance correlates with a greater likelihood of home-team success, suggesting a measurable "home advantage." Additionally, goals scored and conceded are strong predictors of a team's season-long performance. The analysis also finds that attendance has generally increased over time, with certain key events serving as catalysts for spikes in crowd size.

This research matters because it provides insights into what drives successful outcomes and robust fan engagement in women's football. It informs strategic decision-making for clubs aiming to improve both their on-field performances and the experiences they offer to supporters, for organizers to boost engagement, and for analysts interested in understanding sports dynamics. More broadly, these findings are particularly significant as women's sports experience unprecedented growth globally,this study contributes to the growing body of literature that underscores the economic and cultural value of women's football where robust data-driven evidence can shape policy, marketing, and development strategies.

The remainder of this paper is structured as follows: The Data Section 2 discusses the data sources and cleaning processes. The Model Section 3 outlines the model used to evaluate match outcomes. The Results Section 4 presents the key findings from the data analysis, and the Discussion Section 5 provides a summary of what we have learned and suggests potential areas for future research.

2 Data

2.1 Dataset Context and Variable Overview

Our data is derived from the English Women's Football (RobC (2024)) Database, which provides a detailed dataset of matches, team appearances, and standings in the Women's Super League(tier-1) and Women's Championship(tier-2) in England. The dataset includes records spanning multiple seasons, capturing key variables such as attendance, team results, player appearances, and seasonal standings. By analyzing these variables, we can better understand the factors driving attendance, match outcomes, and overall team performance in the league.Following the guidance provided by Alexander (2023), we considered how best to prepare and use this data for analysis in order to effectively tell a story of attendance and performance trends. The analysis was conducted using the statistical programming language R (R

Core Team 2023) and several libraries, including `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `knitr` (Xie 2024), `arrow` (Richardson et al. 2024), and `here` (Müller 2020) for data manipulation and visualization, as well as `janitor` (Firke 2023) and `kableExtra` (Zhu 2024) for data cleanaing and presentation.

This study utilizes three main datasets from the English Women's Football (EWF) Database (RobC 2024):

1. *ewf_matches*: Contains all matches played with details like date, attendance, team and scores.
2. *ewf_appearances*: Contains team-level information for each match, linking teams to their goals and results.
3. *ewf_standings*: Contains end-of-season standings, including points, goals for/against, and final positions.

While other datasets, such as FIFA or Opta's proprietary data, exist, they are not as accessible or as granular at the domestic match level. The EWF dataset is chosen for its open availability, rich detail, and focus on the English domestic league structure. This level of granularity (per-match attendance and scores, historical standings) is critical to the research questions here, making the EWF dataset the most suitable for our study.

2.2 Measurement and Variable Constructions

Each match played in the Women's Super League translates into one or more structured entries in the dataset. For instance, a match between Arsenal and Chelsea that took place on March 1, 2022, with a recorded attendance of 3,500 spectators and a final score of Arsenal 2 – Chelsea 1, becomes a record in the EWF matches file. The attendance figure is typically taken from official league sources, ensuring reliable measurement. The goals for each team are recorded from match officials' reports. Similarly, the final league standings at the end of the season, which reflect teams' cumulative performance and points, are aggregated from individual match results. This chain of measurement ensures that every number in the dataset corresponds to a specific, verifiable real-world event.

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.

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. These constructed variables allow easier interpretation of offensive and defensive strength.

Score Margins (home_team_score_margin, away_team_score_margin) are constructed by subtracting the opponent's goals from a team's goals in a specific match. This transformation provides a numeric sense of dominance or competitiveness in a game.

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.

2.2.1 Variables of Interest

1. *Season and Tier Variables:*

- Season ID (season_id): Identifies the season in a format like "2021–2022". It connects matches to a specific timeframe and competition format. This variable helps track temporal trends.
- Tier (tier): Indicates whether the match was played in the top tier (1) or second tier (2). This classification captures differences in competition level, resources, and potentially attendance patterns.

2. *Match Identifiers and Structure:*

- Match ID (match_id) and Match Name (match_name): : Unique identifiers specifying the teams involved. These ensure each match is distinguishable and can be cross-referenced with other variables.
- Date (date): The date the match was played. This variable is important for tracking seasonal and temporal trends.
- Home Team Score Margin (home_team_score_margin) and Away Team Score Margin (away_team_score_margin): Calculated as the difference between goals scored by the respective team and the opponent. These variables provide insight into match competitiveness.

3. *Attendance Metrics*

- Attendance (attendance): The number of spectators present at each match. No substantial transformations were needed, although matches without reported attendance were excluded from attendance-specific analyses.The attendance variable captures the number of spectators present at each match, serving as a proxy for fan engagement and the overall popularity of the Women's Super League. Analyzing attendance trends not only provides insights into the growth trajectory of women's football but also highlights key moments that have driven fan interest. Figure Figure 2 illustrates the evolution of attendance over time, which shows key shifts and spikes in audience behavior, we can see

from Figure 1 and Table 1 that attendance is most around the 1000-2000 atendees range. The notable spikes during certain seasons, often corresponding with major international tournaments or high-profile matches. For instance, attendance saw a significant increase following the FIFA Women's World Cup, indicating a spillover effect where global events bolster domestic league interest. This suggests that leveraging international exposure through strategic marketing and scheduling can have a profound impact on league engagement. Table 1 shows the summary of attendance.

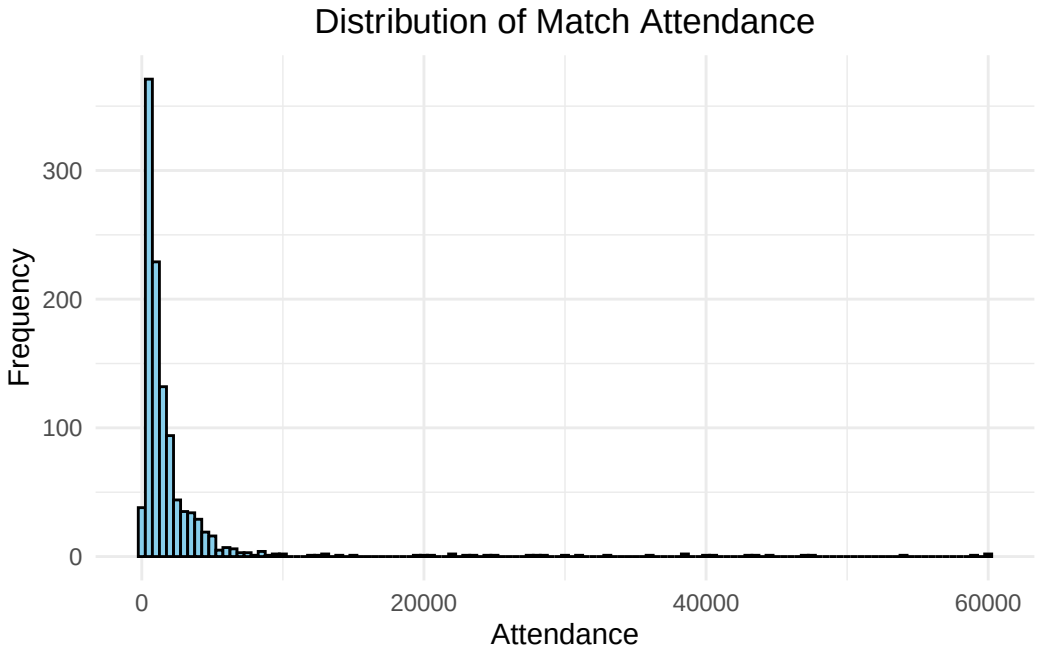

Figure 1: Histogram showing the distribution of match attendance in the Women's Super League.

Table 1: Summary statistics of match attendance in Women's Super League.

Total Matches Played	Average Attendance	Median Attendance	Highest Attendance	Lowest Attendance
1110	2492.245	1062.5	60160	103

4. *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. As shown in Figure Figure 3, the distribution of match attendance varies slightly across match

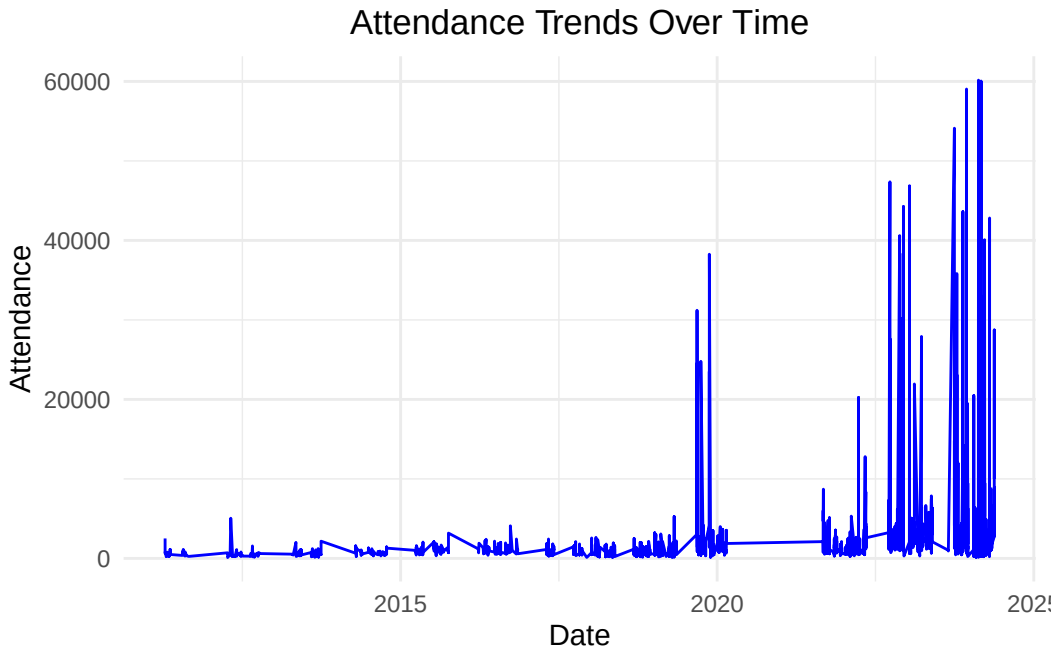

Figure 2: The trend in match attendance over time in the Women's Super League has increased overall with notable spikes often corresponding with major international tournaments or high-profile matches

outcomes. This plot excludes outliers and focuses on typical audience sizes, providing a clearer comparison between Home Wins, Away Wins, and Draws. Table 3 summarizes the average attendance and total matches for each match outcome in the Women's Super League, showing that there games where home team won, had more audience. Table 2 provides summary statistics of attendance, showing the minimum, median, mean, and maximum values across all recorded matches.

Table 2: Summary statistics for goals for and points across all seasons

Mean Goals For	Median Goals For	Mean Points	Median Points
27.39535	24	25.03101	23

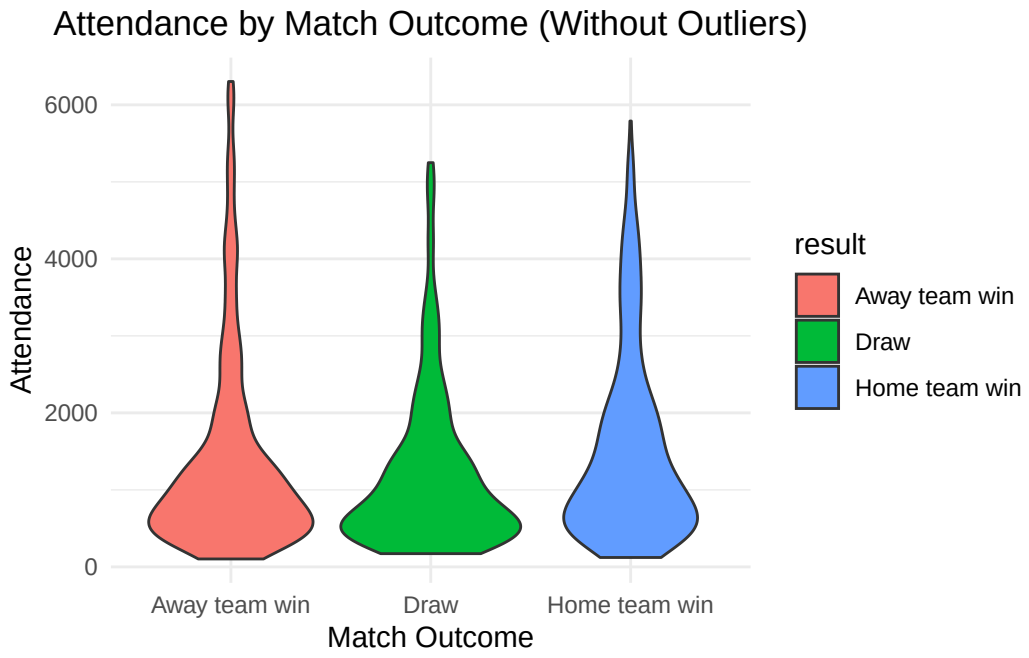

Figure 3: Violin plot illustrating the distribution of match attendance categorized by match outcomes (Home Win, Away Win, Draw) in the Women's Super League. Outliers are excluded to focus on typical audience sizes.

Table 3: Summary statistics for match outcomes by attendance and other predictors.

Match Outcome	Average Attendance	Total Matches
Away team win	2294.546	438
Draw	1910.373	193
Home team win	2907.472	479

3 Model

The primary goal of our modeling is to quantify the relationships between match-level factors—such as attendance, competitiveness (score margins), and tier—and key outcomes (match attendance levels, match results, and team points). Based on the data described previously, we emploed three models:

1. *Linear Regression Model for Match Attendance:*
   To understand which factors influence the number of spectators.
2. *Logistic Regression Model for Match Outcomes:*
   To estimate the probability of a home win, incorporating attendance and performance indicators.
3. *Linear Regression Model for Team Points:* To link season-long performance metrics (goals scored and conceded) to overall success (points earned).

These models allow us to quantify the impact of factors such as goals scored, goals conceded, and attendance on match success and fan engagement. All models were fitted using R (R Core Team 2023) and the `rstanarm` package for Bayesian regression with default priors (Cepeda et al. 2024).

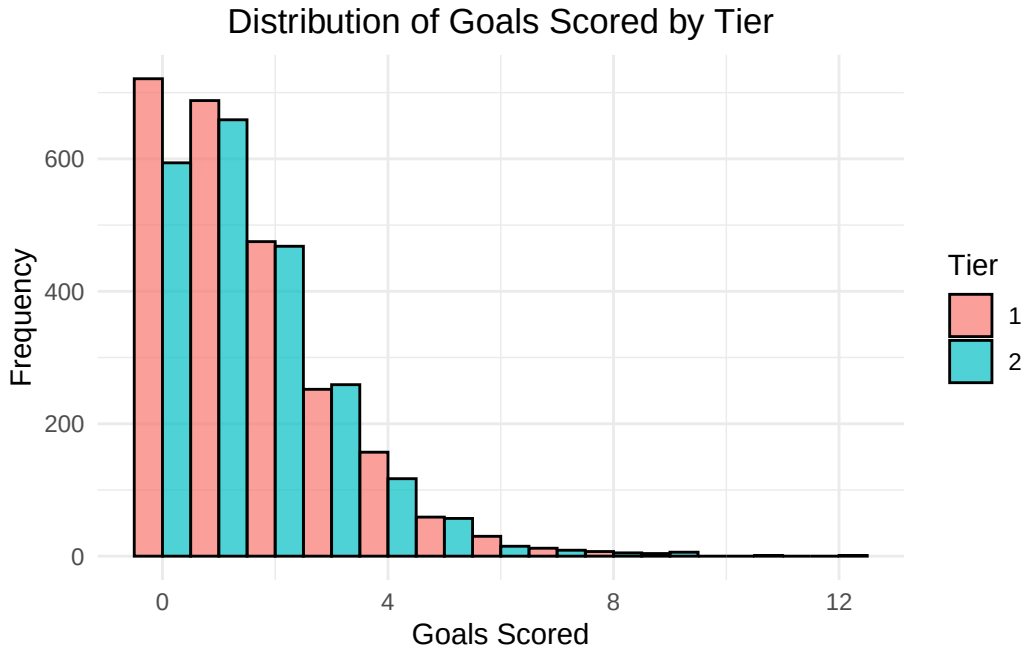

Figure 4: Distribution of Goals Scored and Goals Against by Team Tier

3.1 Predictor variables

This section discusses the predictor variables used in the models. These predictors are integral to understanding the drivers of attendance and match outcomes in the Women's Super League.

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.

3.2 Model set-up

3.2.1 Model 1: Predicting Match Attendance

The first model aims to predict match attendance based on several key variables. We use a linear regression model where attendance is the response variable and predictors include both home and away teams, as well as the score margins.

The first model aims to predict match attendance based on several key variables. We use a linear regression model where attendance is the response variable, and predictors include both home and away teams, as well as the score margins.

Let Y_i represent the match attendance for match i The linear regression model is represented as:

$$Y_i = \beta_0 + \beta_1 X_{home\_team\_i} + \beta_2 X_{away\_team\_i} + \beta_3 X_{home\_team\_score\_margin\_i} + \beta_4 X_{away\_team\_score\_margin\_i} + \epsilon_i$$

Where:

- Y_i: Match attendance for match i
- β_0: Intercept term
- $X_{home_team_i}$, $X_{away_team_i}$: Categorical variables representing the home and away teams
- $X_{home_team_score_margin_i}$, $X_{away_team_score_margin_i}$: Score margins for the home and away teams in match i

- ϵ_i: Error term, assumed to be normally distributed with mean 0

This model captures how the identity of the teams playing and the competitive nature of the match (score margins) influence attendance.

The analysis reveals that factors such as the teams playing and the score margin (both home and away) are significant predictors of match attendance. Popular teams and closely contested matches tend to attract more spectators.

3.2.2 Model 2: Predicting Match Outcome

The second model is a logistic regression model designed to predict the likelihood of a home win. This model uses attendance, score margins, and tier (league level) as predictor variables.

Let (p_i) be the probability that the home team wins match (i). The logistic regression model is given by:

$$\log\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta_1 X_{attendance_i} + \beta_2 X_{tier_i} + \beta_3 X_{home\_team\_score\_margin\_i} + \beta_4 X_{away\_team\_score\_margin\_i}$$

Where:

- p_i: Probability of a home win for match i
- α: Intercept term
- $X_{attendance_i}$: Attendance for match i
- X_{tier_i}: Tier level of the teams
- $X_{home_teams_core_margin_i}$, $X_{away_teams_core_margin_i}$: Score margins for home and away teams in match i

The logistic regression results show that higher attendance significantly increases the probability of a home team victory, suggesting the influence of crowd support. Furthermore, the score margins for both teams are critical factors in determining match outcomes.

3.2.3 Model 3: Predicting Team Performance

The third model is a linear regression model aimed at predicting team points based on goals scored, goals conceded, and the tier of the competition. This model helps us understand the impact of offensive and defensive performance on the overall points accumulated by a team.

Mathematical Representation

Let (P_i) represent the total points earned by team (i) over a season. The model is defined as follows:

$$P_i = \gamma_0 + \gamma_1 X_{goals\_for\_i} + \gamma_2 X_{goals\_against\_i} + \gamma_3 X_{tier\_i} + \epsilon_i$$

Where:

- P_i: Total points earned by team i
- γ_0: Intercept term
- $X_{goals_for_i}$: Goals scored by team i
- $X_{goals_against_i}$: Goals conceded by team i
- X_{tier_i}: Tier level for team i
- ϵ_i: Error term, assumed to be normally distributed with mean 0

The results indicate that scoring more goals positively impacts the number of points earned, whereas conceding goals has a negative effect. The tier variable also shows a significant effect, with teams in higher tiers performing differently compared to those in lower ones.

3.2.4 Model justification

The models chosen are well-suited for the goals of the analysis as linear regression for attendance and team points provides a straightforward method to determine the relationship between several continuous and categorical predictors and the response variables.

Logistic regression for match outcomes appropriately models a binary response variable, allowing us to estimate the probability of a specific result (home win). While the default priors are used here for simplicity, these could be refined if domain expertise suggested particular parameter distributions.

The inclusion of variables such as attendance, score margins, and tier helps capture the essential dynamics influencing match outcomes and team success. Including these features reflects the aspects discussed in the data section, where attendance serves as a proxy for crowd support, and score margins indicate competitive balance. While logistic regression was ultimately

chosen for predicting match outcomes, alternative methods such as Poisson regression explored for their suitability in modeling count-based phenomena like goals scored so it was not chosen due to the main interest being on outcomes and points rather than raw goal counts. Also hierarchical models could account for team-level variability, but would add complexity. The current approach focuses on fixed effects for clarity and interpretability. By considering these alternatives and justifying our chosen models, we show that the final approach is both reasonable and aligned with the research questions and the nature of the data.

3.2.4.1 Assumptions and Limitations

- **Linearity:** The linear models assume a linear relationship between predictors and the outcome. Non-linearities, if present, are not modeled here, but could be explored if diagnostics suggest it.
- **Independence of Observations:** Matches involving the same teams might be correlated. We do not model random effects here for simplicity, but this could be considered if team-level clustering emerges as important.
- **Distributional Assumptions:** For linear models, we assume normally distributed residuals. The logistic model assumes a Bernoulli outcome with a logit link. These are standard assumptions.

3.2.4.2 Model Validation and Diagnostics

Found in the Appendix :

- **Data Splits:** The dataset is divided into training and test sets. Models are fitted on training sets, and performance metrics (RMSE for linear models, accuracy/log-loss for logistic models) are evaluated on test sets.
- **Diagnostics:** Residual plots, posterior predictive checks, and chain convergence diagnostics (for Bayesian fitting) ensure model appropriateness.
- **Sensitivity Analyses:** Alternative specifications (e.g., dropping outliers or using different prior strengths) may be explored to assess robustness.

4 Results

The key results from the predictive models are summarized below:

- *Match Attendance*: Match attendance is positively influenced by popular teams and close score margins.

- *Match Outcomes*: Higher attendance is associated with an increased likelihood of a home win, indicating a potential "home advantage" driven by crowd support.
- *Team Performance*: Goals scored have a positive impact on points earned, while goals conceded have a negative impact. The tier of competition also affects overall team performance.

These findings are useful for team managers and league organizers in developing strategies to enhance both team performance and fan engagement.

5 Discussion

5.1 Big events Understanding the Factors that Drive Match Attendance

One of the key findings is that attendance has generally increased over time, with major events and international tournaments serving as pivotal moments that boost interest in women's football. Understanding what drives this interest allows league organizers to align promotional efforts with these catalysts.

The correlation between attendance and home team success suggests that fan presence can impact match outcomes, likely by providing a motivational boost to players. The findings here indicate that increasing audience engagement could have tangible benefits for home team performance.

5.2 Both offensive and defensive abilities are crucial for a team's performance

Another key observation is the direct impact of goals scored and goals conceded on team standings. Goals scored positively affect match outcomes, as expected, while goals conceded correlate negatively with team success. These results confirm the intuitive idea that both offensive and defensive abilities are crucial for a team's performance. It is not enough to simply score goals-preventing the opponent from scoring also plays an important role.

5.3 Home-Field Advantage: Larger crowds positively correlated with wins at Home turf

Moreover, the analysis of attendance reveals that larger crowds tend to coincide with better performances for the home team. This finding highlights the potential advantage that crowd support can provide, supporting the concept of "home-field advantage" in sports.

5.4 Implications for Teams and Organizers

The results also suggest that teams should focus on improving both offensive capabilities and defensive solidity to succeed. Beyond game tactics, teams and league organizers should continue to work on increasing fan attendance, as the presence of spectators has a clear impact on home team success. This can be achieved through better marketing, improved game day experiences, and leveraging international events to draw in larger audiences.

For organizers, these insights are valuable for shaping promotional strategies that increase attendance, leveraging key calendar events, and enhancing audience experience, thereby boosting overall league engagement and team performance.

5.5 Weaknesses and Next Steps

There are several limitations in the current study that should be addressed in future research. Firstly, the dataset used only spans from 2011-2018, which limitits the generalizability of these results. Extending the analysis to include more seasons would provide a better understanding of longer-term trends.

Another limitation is the absence of player-specific performance metrics. Incorporating individual-level data, such as player fatigue or injuries, would add depth to our understanding of what drives match outcomes. Future models could incorporate additional predictors to improve predictive power and capture the complex dynamics of match performance.

Further research should also consider qualitative factors like weather conditions or managerial changes, which could have significant effects on match outcomes but are not captured in the current quantitative analysis. It would also be useful to explore the impact of specific international events on audience engagement to better understand how these events can be utilized for marketing and promotion.

Appendix

A Survey, Sampling, and Observational Data

A.1 Data Collection Methodology

The data for this study was obtained from the English Women's Football (EWF) Database, which aggregates match-level information from official league sources. The database includes attendance figures, match outcomes, team standings, and player appearances. While the data appears robust, the processes used to collect these data merit closer scrutiny.

- **Attendance Data:**
  Attendance figures are typically recorded at match venues by event organizers. This data provides a direct measure of fan engagement but may suffer from inaccuracies due to:
  - Variability in counting methods (e.g., ticket sales vs. actual turnstile entries).
  - Missing data for lower-profile matches.
  - Potential underreporting or overreporting during high-demand matches.
- **Performance Metrics:**
  Match outcomes and performance data, such as goals scored and conceded, are derived from official match reports. These sources are generally reliable, but the possibility of reporting errors, particularly in historical records, cannot be ignored.
- **Survey Data:**
  While this study does not directly use survey data, similar analyses in the literature often rely on surveys to capture fan demographics, motivations, and preferences. Incorporating such surveys into future research could provide richer contextual insights.

A.2 Sampling Considerations

The dataset spans multiple seasons, capturing a range of tiers and divisions. However, it is important to evaluate the representativeness of the data:

- **Temporal Sampling:**
  Attendance and performance data are unevenly distributed across seasons, with higher-quality data available in more recent years. Earlier seasons may be underrepresented due to incomplete records.

- **Tier Bias:**
  Matches in the Women's Super League (tier 1) are more likely to have complete and reliable data compared to lower-tier matches. This could lead to an overemphasis on top-tier trends, potentially overlooking important dynamics in lower divisions.
- **Event Sampling Bias:**
  High-profile matches (e.g., derbies or post-World Cup fixtures) are disproportionately represented in the dataset due to their visibility and reporting completeness. This could skew attendance trends.

A.3 Observational Nature of the Data

The study relies on observational data, which introduces challenges in establishing causal relationships. Key issues include:

- **Confounding Variables:**
  Attendance and performance trends may be influenced by unobserved factors such as:
  - Weather conditions.
  - Managerial changes.
  - Sponsorship deals.
  - Local economic conditions.
- **Reverse Causality:**
  While high attendance may boost home team performance, it is also possible that strong teams attract more fans, creating a feedback loop.

A.4 Literature Linkages

The use of attendance as a proxy for fan engagement aligns with previous studies in sports economics. Key references include:

- **Borland and MacDonald (2003):** Discuss the determinants of demand for sport, emphasizing the role of team quality and competitive balance.
- **Allan and Roy (2008):** Explore the impact of broadcasting and high-profile events on match attendance in English football.
- **García and Rodríguez (2002):** Highlight the spillover effects of international tournaments on domestic leagues.

These studies underscore the validity of incorporating attendance and performance metrics into analyses of league dynamics. However, they also caution against overinterpreting observational data without controlling for potential confounders.

A.5 Recommendations for Future Surveys

To enhance the depth and applicability of findings, future research could incorporate survey data to complement observational records. Key survey themes might include:

- **Fan Demographics:**
  Age, gender, income, and regional distribution of attendees.
- **Motivations for Attendance:**
  Reasons for attending matches, including loyalty to teams, matchday experience, or interest in specific players.
- **Perceptions of Quality:**
  Fan opinions on the competitiveness and entertainment value of matches.

Survey results could be integrated with observational data through techniques like propensity score matching to reduce selection bias.

B Additional Data Details

The English Women's Football (EWF) Database provided detailed match, attendance, and standings data. This database includes metrics such as match outcomes, goals for and against, attendance, and standings, which were all used in the analysis. Data cleaning and filtering steps were performed to ensure consistency and reliability in the findings presented.

B.1 Model Details and Validation

B.1.1 Posterior Predictive Check

A posterior predictive check was conducted to evaluate the extent to which the model captures the patterns in the observed data. In **?@fig-ppcheckandposteriorvsprior**, the observed data is compared against replicated datasets generated from the posterior distribution of the model parameters. This diagnostic provides insights into areas where the model may underfit or overfit the data.

Additionally, a comparison of posterior distributions with prior distributions is included. This comparison highlights how the data has influenced the model's parameters, illustrating the extent to which prior beliefs have been updated by observed evidence.

B.1.2 Assumption Checks

Diagnostic plots were generated to check for the assumptions of linear regression including:

Residuals vs. Fitted Values: - This plot checks the linearity assumption by examining whether residuals have constant variance and are centered around zero Figure 5.

Q-Q Plot for Normality: - This plot assesses whether residuals are normally distributed, a critical assumption for inference in linear regression models Figure 6.

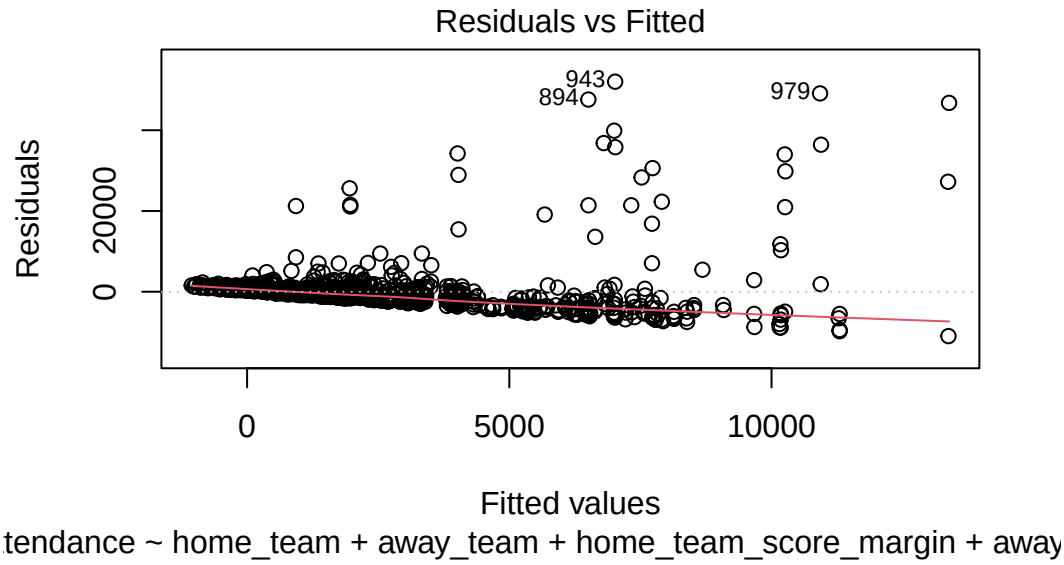

Figure 5: Diagnostic plot for linear regression assumptions: The Residuals vs. Fitted Values plot checks for linearity and homoscedasticity by assessing whether residuals are randomly dispersed around zero with constant variance.

B.1.3 Model Validation

The dataset was split into training (80%) and testing (20%) sets to validate the model's performance on unseen data. This approach ensures that the model generalizes well and is not overfitted to the training data, see Table 4.

1. **Train-Test Split:**

- Training data: Used to fit the model.
- Testing data: Used to evaluate the model's predictive accuracy.

2. **Performance Metrics:**

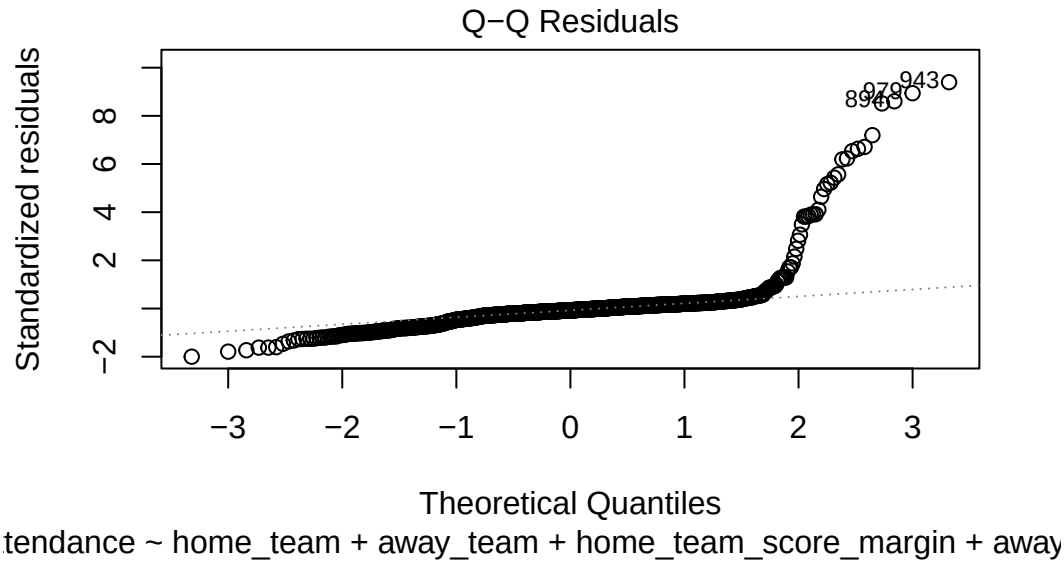

Figure 6: Diagnostic plots for linear regression assumptions: The Q-Q plot evaluates the normality of residuals, ensuring they follow a standard normal distribution critical for valid inference in linear regression.

- Root Mean Squared Error (RMSE): Measures the average prediction error in the same units as the response variable (attendance).
- Mean Absolute Error (MAE): Captures the average absolute prediction error.
- R-Squared (R^2) Indicates the proportion of variance in attendance explained by the predictors.

Table 4: Performance metrics for the linear model predicting attendance.

Metric	Value
Root Mean Squared Error (RMSE)	5265.9418526
Mean Absolute Error (MAE)	2523.6917492
R-Squared (R2)	0.1852367

B.1.4 Residual Analysis:

- Residuals plotted against each predictor to identify non-linear relationships or heteroscedasticity Figure 7.

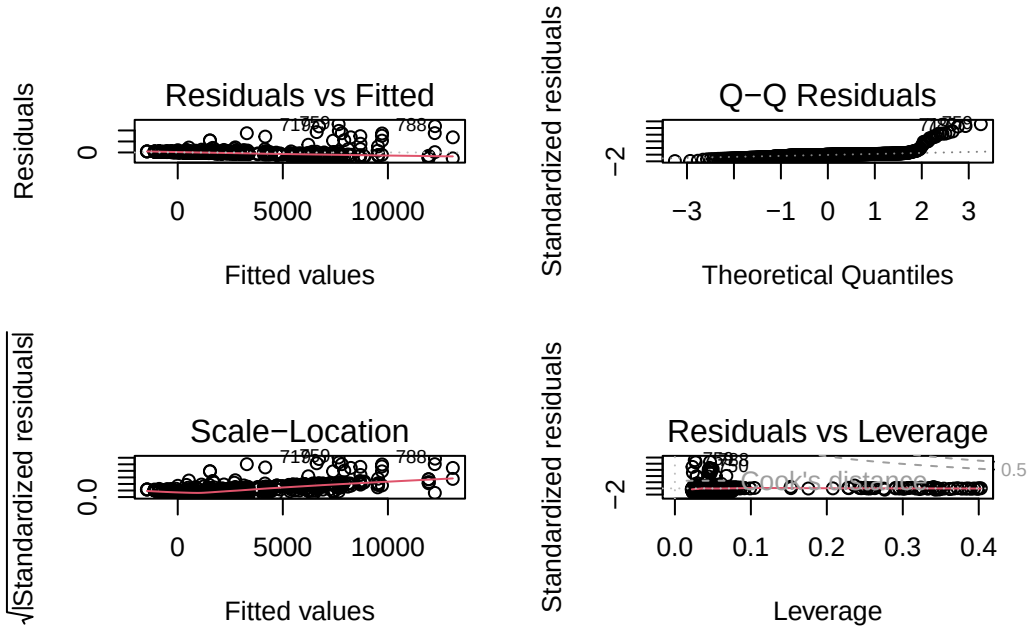

Figure 7: Residuals plotted against each predictor to identify non-linear relationships or heteroscedasticity.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Cepeda, Gabriel A., Ben Goodrich, Jonah Gabry, Rachael Meager, and Andrew Gelman. 2024. *Rstanarm: Bayesian Applied Regression Modeling via Stan*. https://mc-stan.org/rstanarm/.

Eekeren, ter Haar van, F. 2022. "The Rise and Development of Women's Football: A Scoping Review of Scholarly Literature." *International Journal of Sport Policy and Politics* 14 (2): 285–307. https://doi.org/10.1080/19406940.2022.2039702.

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

Garcia, Rodiguez, J. 2002. "The Determinants of Football Match Attendance Revisited: Empirical Evidence from the Spanish Football League." *Journal of Sports Economics* 3 (1): 18–38.

Kitching, Harvey, N. 2022. "Women's Football, Commercialization and Engagement: A Critical Review." *Sport in Society* 25 (6): 1093–1110. https://doi.org/10.1080/17430437.2021.1995707.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

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

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

RobC. 2024. "The English Women's Football (EWF) Database." https://github.com/probjects/ewf-database/tree/main.

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.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.