**Visualizing Home Team Advantages in Football: Exploring Score Trends, Goal Timing, and Penalty Benefits**

Student number: 240192149

Words count: 3130

**1. Knowledge building**

Home-field advantage is evident in many sports, analyzing the data from football matches brings out advantages that are in favor of the home team. Visualization helps to clarify whether the home team is better in terms of score, goals, and penalty kicks. Some advantages that could be attributed to familiarity with the venue, crowd support, and psychological edges (Wang & Qin, 2023), which has been well-documented in numerous sports (Bialkowski et al., 2014). This analysis, through visualizations, seeks to answer four fundamental questions using the 'results.csv' and 'goalscorers.csv' datasets:

**RQ1: What trends are observed in the scores of home and away teams?**

图表

低可信度描述已自动生成

This heatmap shows the number of goals scored by home and away teams. The x-axis shows the away team's goals, and the y-axis shows the home team's goals. Each grid in the plot represents a combination of home and away team goals, with shading indicating the frequency of each combination. We can find the home team tends to win with a small advantage, especially the score of 1:0, further emphasizing the home team's advantage (Wang & Qin, 2023).

**RQ2: How to further explain the home team's scoring advantage?**

**图表, 条形图

描述已自动生成**

This stacked bar chart shows match categories (e.g., "Win by 1 goal", "Win by 2 goals") on the x-axis, and the y-axis shows the win rates. Red and blue bars represent the win rates for the home and away teams. The chart reveals that the home team is more likely to win in matches with larger goal differences, meaning the proportion of home team victories increases as the goal margin widens. In contrast, away teams are less likely to win by large margins. This phenomenon is linked to home-field advantage, where the home team enjoys greater psychological and environmental benefits, making it harder for away teams to maintain a large lead (Hilton, Olfman, & Njunge, 2020).

**RQ3: What is the trend of goal timings and goal counts for home and away teams during regular match time (90 minutes)?**

**图表, 条形图

描述已自动生成**

This bar chart represents goals scored by home and away teams during the regular time of matches. The x-axis presents match time (from 0 to 90 minutes), while the y-axis shows the number of goals. Red bars represent home team goals, while blue bars represent away team goals. From the results, it indicates that at every point in the match, the home team scores more goals than the away team. Previous studies also showed a rise in goals in the second half, possibly due to factors such as physical fatigue and substitutions (Plakias et al., 2023). Moreover, the goal peaks at the 55th, 65th, 75th, 80th, and 85th minutes, possibly because substitutions affect attacking rhythm; after substitutions, goals may be scored within a short period of time (Plakias et al., 2023). Notably, at the 65th minute, the goal difference between home and away teams were the biggest, which might hint those substitutions around the 60th minute are most effective.

**RQ4:** **Does the home team have more of an advantage in penalty kicks?**

**图表, 图示, 饼图

描述已自动生成**

The above two pie charts compare the proportion of penalty goals to total goals for both teams. The orange sections represent penalty goals, while the green sections represent other goals. The results indicate that 6.5% of the home team's goals are penalty goals, which is higher than the away team's 5.8%. This suggests that the home team slightly benefits more from penalties. The subsequent pie charts show that in all matches with penalties, the home team benefited from penalties in 8.3% of games, compared to 5.8% for the away team. By the way, "benefiting from penalties" refers to matches where a team secured a draw or win due to penalty goals. For instance, if both teams score penalty goals, the penalty goals are adjusted by subtracting the away team's penalty goals, and the net effect is used to assess whether a team benefited from penalties.

The home team's higher proportion of benefiting from penalties may be attributed to referee bias or psychological factors associated with playing at home (Perin, Vuillemot, & Stolper, 2018).

**2. Theoretical Framework**

图表

低可信度描述已自动生成

**2.1 The Question to Answer:** **What trends are observed in the scores of home and away teams?**

This graph attempts to visualize the frequency of combinations of scores for home and away teams. The above heat map depicts the number of all possible different score combinations, to try to answer this question.

**2.2 ASSERT Framework:** The ASSERT framework, which consists of six stages (Ask, Search, Structure, Envision, Represent, and Tell), guides the creation of meaningful visualizations. Here's how each stage is applied to this project:

**A – Ask:** The central question is: "What trends are observed in the scores of home and away teams?"

**S - Search:** This visualization utilizes the data derived from the International Football Results dataset results.csv. This file includes scores for home and away teams in different matches.

**S – Structure:** The data is structured by filtering the scores of home and away teams in each match. This organization is done to understand whether home teams are more likely to score or if there are specific score combinations that occur more frequently.

**E – Envision:** The heatmap is the best choice to represent the frequency of combinations of scores by home and away teams. Heatmaps are great at visualizing large datasets and categorical data, such as scores, so the patterns and trends between the teams can be easily observed. The gradient from light to dark will represents differences, where the darker the shade, the greater the frequency of combinations.

**R – Represent:** Create heat map to shows the relationship between the scores of home and away teams. The cells in the matrix diagram represent different score combinations, and the shading of each cell represents how often that score combination occurred.

**T – Tell:** The heatmap clearly outlines the scoring performance of the home and away teams. The heatmap shows that home teams score more often than their away counterparts, and the most common scoring combination is 1:0. This observation is indicative of a home-field advantage, where the home team is more likely to win when the teams' abilities are relatively equal.

**2.3 Grammar of Graphics (GoG):**

This visualization follows the four elements of the Grammar of Graphics (GoG):

**Geometries:** The heatmap is the basic geometrical representation, where every single cell represents a specific score combination between the home and away teams, with the color gradient representing the frequency of such a combination.

**Aesthetics:** The size of the grid squares is equal. Add a dotted line on the grid of the tied score.

**Coordinate System:** The Cartesian coordinate system is used, where the x-axis represents away team goals, and the y-axis represents home team goals.

**Scales:** The color scale is continuous, with a gradient from light purple (low frequency) to dark purple (high frequency).

**2.4 Reasons for Choosing GoG Elements:**

**Geometries:** A heatmap is selected since it is the best way of showing the frequency of two variables, home and away team scores, as combinations. It is very good for the visualization of frequency data, where scoring patterns can be clearly reflected. For instance, Liu (2022) adopted a similar approach to visualize score combinations and reveal the frequency distribution in football matches.

**Aesthetics:** The reason why the grids are guaranteed to be equal squares is because the units of the x-axis and y-axis are the same, the squares are the most beautiful and intuitive. The dashed line is added to the diagonal of the heat map to highlight symmetric score positions (for instance, 1:0 vs. 0:1), making it easier for viewers to quickly understand the symmetric relationships in the data.

**Coordinate System:** The Cartesian coordinate system is used to make the comparison between home and away team scores clear and precise.

**Scales:** The color gradient is used to represent frequency because it provides an intuitive way to convey the density of data, allowing viewers to easily identify the most common score combinations.

**3. Accessibility**

**图表, 条形图

描述已自动生成**

In data visualization, accessibility means such content must be made understandable and usable by all users, visually impaired, color-blind, or having other cognitive impairments. That means, while designing the same, consideration of different types of user groups must be there so that accessing and understanding data becomes easier for them. Good accessibility design enables not only regular users but also assures that people with special needs can work effectively with the visualization (Basole & Saupe, 2016).

**3.1 Accessibility Analysis:**

The stacked bar chart displays the proportions of home and away team wins across different match categories, such as "Win by 1 goal," "Win by 2 goals," and so on. The x-axis represents the match categories, and the y-axis shows the proportions. The red and blue bars represent the proportions for home and away teams, respectively. From an accessibility perspective, the design helps convey clear information, but there are certain factors that may impact specific user groups.

**3.2 Design Choices that Support Accessibility:**

**Clear Structure of the Bar Chart**: The match categories are clearly listed on the x-axis, making it easy for users to understand the proportions of home and away team wins in different match categories.

**High Color Contrast**: The strong contrast between red and blue helps most viewers distinguish between home and away team proportions. This is effective for users with normal vision.

**Clear Horizontal Comparison**: The stacked bar chart allows for a clearer horizontal comparison across different score margins, which makes it easier to compare the relative proportions of home and away wins briefly. This feature enhances the visualization's clarity, enabling users to quickly identify patterns across different match categories.

**3.3 Design Choices that May Hinder Accessibility:**

**Color Usage**: Red and blue are among the most difficult color combinations for colorblind users to distinguish. This means that colorblind viewers may struggle to differentiate the proportions for home and away teams, which makes the chart less accessible for this group (Yuan & Du, 2021). To resolve this issue, other methods can be used to enhance accessibility. For example, adding patterns or different fill textures within the bars would have helped colorblind users differentiate between categories. Using shapes of different kinds in the legend-circles and squares-can help to further improve the accessibility of the chart.

**Lack of Specific Score Information**: While the chart effectively shows the proportions of home and away team victories across different goal margins (e.g., "Win by 1 goal" or "Win by 2 goals"), it does not display the specific score data. For example, matches with scores of 2:0 and 4:2 is grouped together under "Win by 2 goals," without distinguishing these specific score differences. This limitation restricts the chart's accessibility and prevents viewers from fully understanding why the home team is more likely to win by larger margins. To improve accessibility, combining this chart with another visualization, such as a heatmap, would help provide more specific score information. A heatmap can display the distribution of scores, offering more detailed insights and helping viewers better understand why home teams tend to win by larger goal margins (Basole & Saupe, 2016).

**Inability to Reflect Draws**: The chart cannot reflect draws because draws do not have a goal difference. In situations where both teams score the same number of goals, the absence of a goal difference prevents the chart from representing these cases clearly. As a result, viewers may overlook or misunderstand the occurrence of draws in the data.

To address this, a separate visualization focused specifically on draw games could be created. This visualization would highlight the proportion of games that ended in a draw, potentially using a distinct color or symbol to differentiate these cases from wins. This would allow users to see the full scope of outcomes (win and draw) in the dataset, while keeping the original chart focused on win proportions.

**4. Visualization choice**

**图表, 条形图

描述已自动生成**

**4.1 Justification for Choosing the Bar Chart**

For RQ3: What is the trend of goal timings and goal counts for home and away teams during regular match time (90 minutes)? I chose a bar chart to represent the goals scored by home and away teams during the match.

The bar chart is an appropriate choice for this data because it clearly shows the number of goals scored by both teams (home and away) at each minute of the match. The x-axis represents match time (from 0 to 90 minutes), and the y-axis shows the number of goals. The red bars represent home team goals, and the blue bars represent away team goals. This type of visualization helps achieve the goal of showing the trend in goal timings and counts because bar charts are very effective at displaying discrete data, such as the number of goals at specific time points (minute by minute). It allows for clear, direct comparison between home and away teams' goals throughout the match. This way, users can easily see at which points in the match the home team scores more goals. The chart also highlights goal peaks and troughs, making it easier to compare performance between home and away teams.

**4.2 Possible Alternatives**

While the bar chart is an effective choice, there are other types of visualizations that could also present this data. Below are two possible alternatives:

**Line Chart**

**图表, 直方图

描述已自动生成**

**Pros**: A line chart could effectively show the trend of goals scored by both the home and away teams throughout the match. It would, quite clearly, show how goals from each team change at each minute, especially for long-term trends or fluctuations in goal scoring.

**Cons**: However, the line chart is less suitable for comparing the goals of the home and away teams at the same point in time. Even though the line chart can reflect both teams' trends, it makes direct comparison of goal counts for both teams at the same minute difficult. This makes it difficult to easily see, from the chart, the comparison of home and away goals at any point in time.

**Scatter Plot**

**图表, 散点图

描述已自动生成**

**Pros**: A scatter plot could offer a cleaner, simpler way to show the goals scored at each minute by each team. It would represent the data as individual points, making it easier to see the distribution of goals over time.

**Cons**: While scatter plots are simple, they do not easily allow for the comparing of home and away team goals in the same minute. Since each data point represents a goal, it could be hard to compare quickly the total number of goals scored by each team at given times. Because the comparison structure is unclear, this may make it harder for users to grasp the trend between home and away teams.

**4.3 Discussion on the Visualization Goal**

This visualization is done to show how goals scored by home and away teams change as time progresses in a match, and if there is any pattern or trend in the data. The bar chart serves this purpose very well because it's easy to compare the number of goals at each minute, and it emphasizes for which minutes the home team tends to score more goals than the away team. It also highlights the sharp peaks in goals scored towards the end of the match, which could be due to factors such as physical fatigue, substitutions, and changes in game dynamics (Plakias et al., 2023).

**5. Implications and Improvements**

**图表, 图示, 饼图

描述已自动生成**

**Ethical Implications of Using the Visualization**

While visualizations can be a powerful way to communicate insights in data, if presented in a manner that is not properly framed, they may be misleading or even have ethical concerns. In this visualization, pie charts plot penalty goals on total goals for home and away teams, this kind of representation misleads the audience. A few possible misleading issues and recommendations for improvement are discussed below.

**Misleading Implication:** A higher probability of receiving penalties doesn't mean greater benefit from them: Although this may seem counterintuitive, a higher probability of receiving penalties does not necessarily mean a team benefits more from them. For example, the home team may receive more penalties in matches where they are winning by a large margin (because they dominate), but in closely contested matches, the probability of penalties awarded may be smaller, yet those penalties could have a more significant impact on the result. In other words, while the home team may have a higher probability of receiving penalties, it does not necessarily imply that penalties contributed more to the outcome of the match. To clarify this, we refer to the subsequent pie charts. These charts show the proportion of games in which the winning or drawing teams benefited from penalties—i.e., the teams that won or leveled the score with the help of penalties. These charts demonstrate the proportion of home and away teams benefiting from penalties in all matches with penalties, providing a more comprehensive view of the impact of penalties on match outcomes (Hopkins, 2021).

**Data Omission:** Missed penalties are not accounted, the current visualization does not include data on missed penalties, which limits our ability to fully understand the role of penalties. Simply considering penalty goals does not reflect the complete impact of penalties, as some penalties may have been missed. This omission is a significant gap in the analysis, preventing a full understanding of how penalties affect match outcomes.

To address this, data on missed penalties should be included in the visualization. This would allow for a more complete understanding of penalties' influence on the match, not just those that resulted in goals.

**Objective Factors Affecting Penalty Decisions**: This visualization is done to show the goals scored by home and away teams in a match as time progresses into the game and if there is any pattern or trend present within the data. This is very well served by this bar chart, as it is easier to compare how many goals at each minute and which minutes the home team usually scores more goals than the away team. It also underlines the peaks of goals scored sharply at the end of the game, possibly because of reasons such as physical fatigue, substitution, and changes in game dynamics (Plakias et al., 2023).

**Improvement Suggestions for the Visualization**

**Missed Penalties Data**: To ensure the completeness of the data, missed penalties should be incorporated into the visualization. By showing which penalties were missed, the audience can better understand whether the penalties played a decisive role in the match.

**Add More Contextual Information**: In addition to showing the proportion of penalty goals and penalty benefits, it would be helpful to provide more contextual information, such as the closeness of the match and the final score differences. This will help the audience understand whether penalties played a key role in critical moments, especially in tightly contested matches.

**Use Alternative Visualization Methods**: Consider using interactive charts that allow users to explore specific teams or penalty data over certain periods to draw more detailed conclusions. This will help avoid oversimplification and provide richer context for analysis.

**References**

1. Bialkowski, A., Lucey, P., Carr, P., & Yue, Y. (2014). Win at home and draw away: Automatic formation analysis highlighting the differences in home and away team behaviors. *Proceedings of the MIT Sloan Sports Analytics Conference*. Retrieved from <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=c932441f8ecc9348e2dd54ec749984c8f4abf224>

2. Hilton, B., Olfman, L., & Njunge, C. (2020). Home field advantage in major league soccer. *AIS Journal of Sports Analytics*. Retrieved from [https://scholar.archive.org/work/ps6hgpjyhvgg7bpjcv2f6o5l6i/access/wayback/https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1401&context=amcis2020](https://scholar.archive.org/work/ps6hgpjyhvgg7bpjcv2f6o5l6i/access/wayback/https:/aisel.aisnet.org/cgi/viewcontent.cgi?article=1401&context=amcis2020)

3. Plakias, S., Moustakidis, S., & Mitrotasios, M. (2023). Analysis of playing styles in European football: Insights from a visual mapping approach. *Journal of Sports Visualization*. Retrieved from <https://e-space.mmu.ac.uk/632444/1/Art169.pdf>

4. Perin, C., Vuillemot, R., & Stolper, C. D. (2018). State of the art of sports data visualization. *Computer Graphics Forum*. Retrieved from <https://hal.science/hal-01806107/document>

5. Wang, S., & Qin, Y. (2023). The impact of crowd effects on home advantage of football matches during the COVID-19 pandemic—A systematic review. *PLoS ONE*. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0289899>

6. Liu, N. (2022). *GeoVisualisation of Football Players Movement*. Retrieved from <https://theses.cz/id/mbeof5/Thesis_Liu.pdf>

7. O'Hare, A., McMenemy, P., & Mead, J. (2023). Expected goals in football: Improving model performance and demonstrating value. *PLoS ONE*. Retrieved from <https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0282295&type=printable>

8. Basole, R. C., & Saupe, D. (2016). Sports data visualization. *IEEE Computer Graphics and Applications, 36*(1), 72-83. Retrieved from <https://ieeexplore.ieee.org/abstract/document/7579431/>

9. Yuan, X., & Du, M. (2021). A survey of competitive sports data visualization and visual analysis. *Journal of Visualization, 24*(4), 679-692. Retrieved from <https://link.springer.com/article/10.1007/s12650-020-00687-2>

10. Hopkins, P. (2021). *Scotland's Islamophobia: Report of the inquiry into Islamophobia in Scotland by the Cross-Party Group on Tackling Islamophobia*. Retrieved from <https://drive.google.com/file/d/1oyAbWddyfwb4D9qJQKcikC40NSe5Xqo0/view>