Exploratory Visualization of Football Match Dynamics

Using event-level data to explore contextual effects on performance, discipline and outcomes

Course:	Intelligent Data Analysis
Academic Year:	2025/2026
University:	Universidad Politécnica de Madrid
Delivery Date:	January 31, 2026

Author: Elías Cobas Pascual

Abstract: This report applies exploratory data visualization techniques to football match data in order to investigate how contextual factors shape match outcomes and on-field behavior. We analyze disciplinary patterns in matches with large score differences and the interaction between team experience and competitive context. The focus is on visual exploration, pattern identification and methodological transparency.

Universidad Politécnica de Madrid

1 Introduction

Football matches are complex dynamic systems in which outcomes and on-field behavior are shaped by many factors like team quality, match state, scoreline, experience, or critical in-game events. This report uses event-level data from Spanish LaLiga matches spanning multiple seasons (2000–2022) to explore how match context relates to performance and disciplinary behavior. The dataset, which is publicly available [2], contains detailed information on match results, referees, tactical formations, lineups, and a chronological log of in-game events such as goals, cards, substitutions, and VAR decisions. While the data lack advanced technical metrics (e.g. possession, shots or pressure), their temporal structure enables the study of situational patterns and short-term reactions within matches.

Two main research questions guide this analysis. First, we examine whether disciplinary behavior in situations of large score differences ("goleadas")[3] has changed over time, focusing on the frequency of yellow cards relative to match context. Secondly, we explore whether team experience, measured through average player age, is associated with different match outcomes in balanced and unbalanced matches when controlling for team quality using ELO ratings.

The goal of this work is not to establish causal relationships, but to identify patterns, asymmetries and potential effects through visualization and descriptive analysis. Throughout the report, we explicitly discuss data limitations, dead ends encountered during exploration, and outline how more rigorous statistical methods could be applied in future analyses to test the observed patterns.

2 Disciplinary behaviour in situations of heavy defeats

A recurrent narrative in football discourse is that the game in previous decades was more aggressive, especially in emotionally charged contexts such as heavy defeats. This idea was recently articulated by former Spanish international goalkeeper Santiago Cañizares[4], who argued that in the past, players were more likely to react violently or unsportingly when losing by large margins, leading to injuries, confrontations, and expulsions. In his words, after a 5–0 defeat, players would often "end up fighting" rather than showing sportsmanship [5].

Such statements raise an empirical question that can be explored through match data: *are disciplinary sanctions more frequent in situations of clear imbalance, and has this relationship changed over time?* While yellow cards cannot directly measure player behaviour or intent, they do provide an observable proxy for referee-sanctioned unsporting actions on the pitch.

2.1 Data preparation and manipulation

To analyse disciplinary behaviour in situations of heavy defeats, several design decisions were required to define comparable contexts and avoid misleading interpretations.

First, a situation of clear imbalance ("*goleada*") was defined operationally as any moment in which a team was losing by three goals or more (goal difference ≥ 3). Rather than classifying entire matches, the analysis was conducted at the minute level. Each yellow card was assigned to one of two mutually exclusive contexts depending on the scoreline at the exact minute in which it occurred: *goleada* (difference ≥ 3 goals) or *non-goleada* (difference < 3 goals). This strict separation allows a single match to contribute to both contexts if the scoreline changed over time, avoiding an overly broad match-level classification.

Second, all disciplinary counts were normalised by time of exposure. The total number of minutes played in goleada situations varies substantially across seasons, as does the number of minutes played outside such situations. Comparing absolute numbers of yellow cards would therefore be misleading. Instead, we computed the rate of yellow cards per minute for each context and season, to be able to calculate the ratio between these two rates.

Before analysing temporal trends, we performed exploratory checks to ensure that the results were not driven by a small number of extreme seasons or by insufficient data. Figure 1 shows the distribution of total minutes played in goleada situations per season, while Figure 2 shows the distribution of total yellow cards issued during goleadas. Both boxplots were inspected using the interquartile range (IQR) criterion to detect potential outliers. No single season was found to dominate the distributions, supporting the robustness of the subsequent visual analysis.

2.2 Plots and findings

Figure 3 shows the temporal evolution of the ratio between the rate of yellow cards per minute in goleada situations and the corresponding rate in non-goleada situations. A ratio above one indicates that yellow cards are issued more frequently, relative to playing time, during heavily unbalanced phases of the match, while a ratio below one indicates the opposite. The dashed horizontal line at one represents the neutral reference where both contexts exhibit the same disciplinary intensity. Contrary to the common narrative that modern football is more restrained in situations of heavy defeat, the figure does not reveal a clear downward trend in this ratio over time. In early seasons, the ratio is often below one, suggesting fewer yellow cards per minute during goleadas compared to more balanced situations. However, from approximately the mid-2010s onwards, the ratio frequently approaches or exceeds one, indicating that goleada situations are at least as likely, and in some seasons more likely, to result in disciplinary sanctions. This pattern suggests that heavily unbalanced scorelines have not become relatively "cleaner" over time in terms of referee-sanctioned behaviour.

To aid interpretation, Figure 4 presents the same comparison using an alternative but equivalent metric: minutes played per yellow card. Higher values correspond to fewer sanctions per unit of time. While the average time between yellow cards in non-goleada situations remains relatively stable across seasons, the corresponding values for goleada situations show a clearer downward tendency. This indicates that yellow cards have become more frequent during heavily unbalanced phases of play, even as disciplinary behaviour in more competitive contexts has remained largely unchanged.

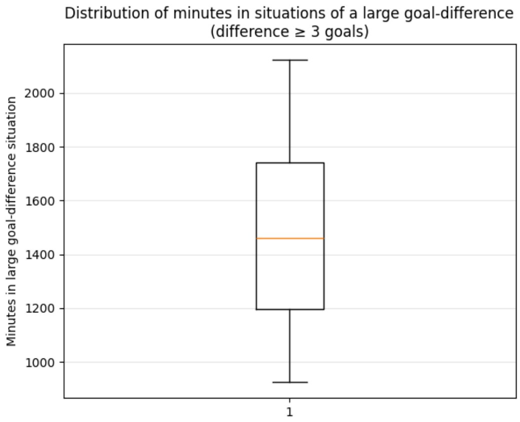

Figure 1: Distribution of total minutes played in goleada situations per season.

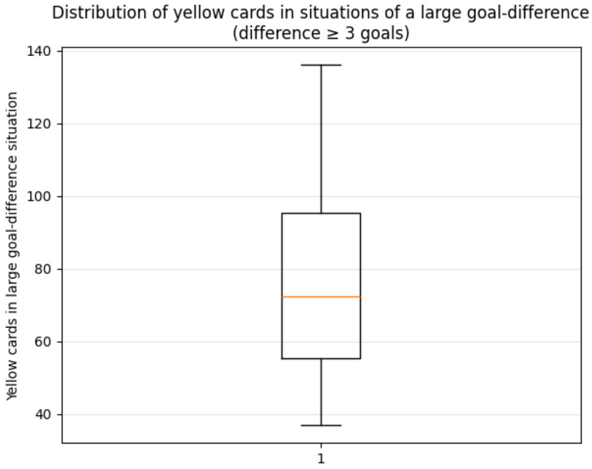

Figure 2: Distribution of yellow cards issued in goleada situations per season.

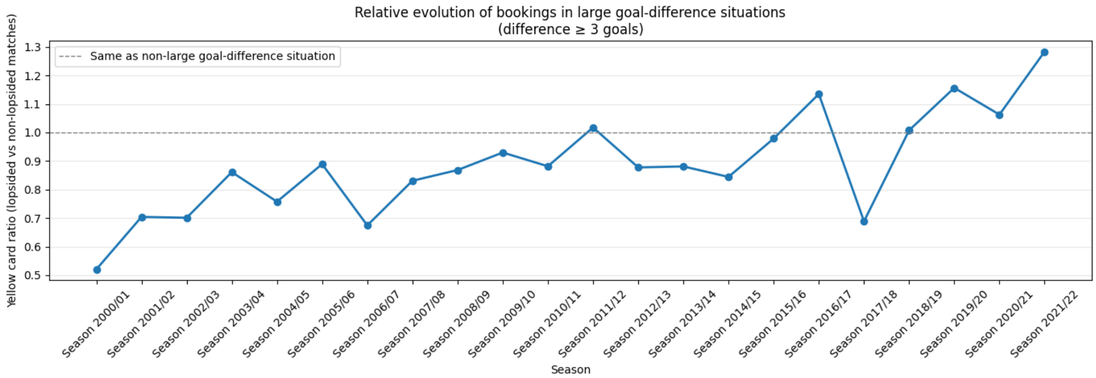

Figure 3: Relative evolution of yellow card rates in goleada versus non-goleada situations.

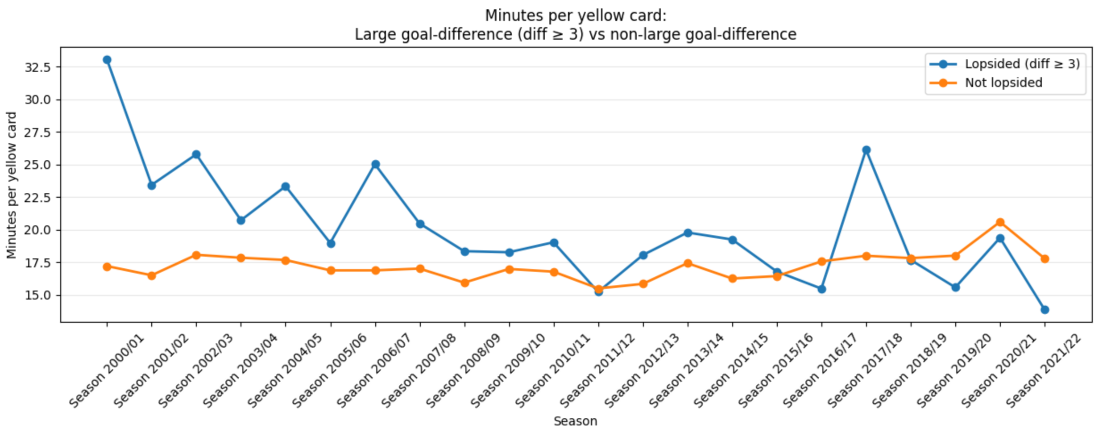

Figure 4: Minutes per yellow card in goleada and non-goleada situations.

Findings Overall, these plots suggest that, contrary to common belief, disciplinary behaviour in situations of heavy defeat has not become relatively more restrained over time. The absence of a downward trend in the relative frequency of yellow cards during goleadas indicates that modern players are not necessarily more passive or conciliatory when losing by large margins. At the same time, this pattern may also reflect a shift in refereeing standards, with officials potentially applying stricter disciplinary criteria in such situations. As a result, the observed stability or increase in relative sanction rates may be driven not only by player behaviour but also by changes in how unsporting actions are identified and penalised.

Data analysis plan To assess whether the observed patterns reflect a statistically meaningful change rather than random variation, a formal analysis could model the probability of a yellow card occurring at the minute level using regression-based approaches. For example, a logistic regression could be employed with the occurrence of a yellow card as the response variable and match context (goleada vs non-goleada), season, and their interaction as predictors. This would allow testing whether the relative effect of goleada situations on disciplinary sanctions has changed significantly over time. Additionally, bootstrapping or permutation tests could be used to evaluate the robustness of the observed trends without relying on strong parametric assumptions.

3 The impact of team age and experience on match outcomes

It is a common belief in football that teams with very young squads lack the maturity and experience required to consistently achieve good results, particularly in demanding competitive contexts. Young teams are frequently described as more error-prone, less tactically disciplined, or less capable of managing adverse situations during a match. In this section, we explore whether this common assumption is reflected in match outcomes by comparing the performance of particularly young teams against more experienced ones. Specifically, we analyze whether team age is associated with different results in balanced matches, as well as in unbalanced matches where teams either face stronger opponents or play with a clear quality advantage.

3.1 Data preparation and manipulation

To study the relationship between team age, experience and match outcomes, we relied on the information available in the dataset regarding starting lineups, players' dates of birth and individual ELO ratings (a skill measure on a 1–100 scale). From this information, we computed for each team and match the average age of the starting eleven and the average ELO, and derived relative measures by comparing each team to its opponent.

Age and ELO distributions. Before defining what we consider a *young* or *experienced* team, and what constitutes an *balanced* or *unbalanced* match, we first explored the empirical distributions of these variables. Figure 5 shows the distribution of ELO differences between teams. The median is zero, with the interquartile range concentrated roughly between −5 and +5 ELO points, while more extreme differences appear as outliers. Figure 6 displays the distribution of the average age of the starting elevens. Most teams fall between roughly 26.5 and 28 years.

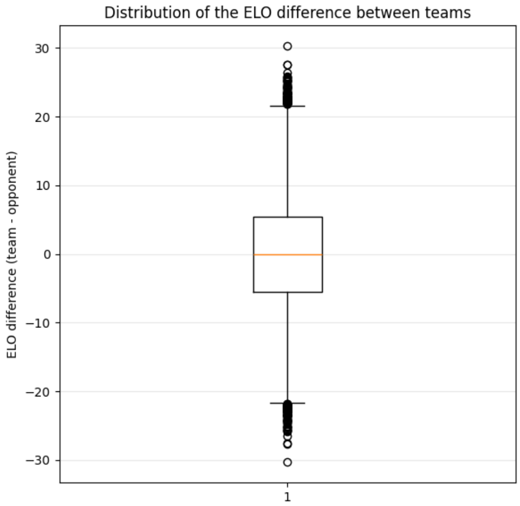

Figure 5: Distribution of ELO differences between teams across all matches.

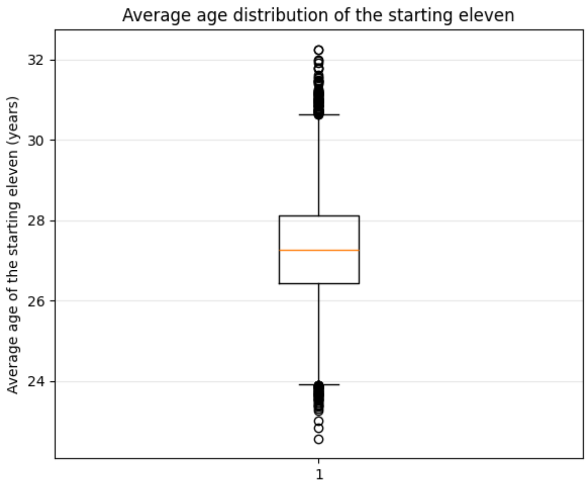

Figure 6: Distribution of the average age of the starting eleven.

Operational definitions. Based on the observed distributions, we adopted the following definitions:

- **Young vs experienced match**: a match in which one team has an average starting-eleven age ≤ 26 years and is at least 0.75 years younger than its opponent.
- **Experienced vs experienced match**: a match in which both teams have an average starting-eleven age ≥ 26.5 years.
- **Balanced match**: absolute ELO difference between teams $|\Delta\text{ELO}| \leq 5$.
- **Unbalanced match**: absolute ELO difference between teams $|\Delta\text{ELO}| \geq 7.5$.

Sample size checks. To ensure that the visual comparisons are meaningful, we verified that there were sufficient observations in each category. Tables 1 and 2 summarize the number of matches available for each combination of age group and competitive context.

The tables show that, while young teams are less frequent than experienced ones, especially in matches where they start with an ELO advantage, there are still several hundred observations in each relevant subgroup. This provides a sufficient basis for exploratory visualization and comparison.

ELO context	Match-up type	Matches
Disadvantage	Young vs experienced	522
Disadvantage	Experienced vs experienced	2389
Advantage	Young vs experienced	355
Advantage	Experienced vs experienced	2593

Table 1: Volume of unbalanced matches by age matchup and ELO context

Match-up type	Matches
Young vs experienced	1064
Experienced vs experienced	6333

Table 2: Volume of balanced matches by age matchup

3.2 Plots and findings

To evaluate whether team age and experience are associated with performance, we focus on the points obtained per match (loss: 0 points, draw: 1 point, win: 3 points) under different competitive contexts.

Figure 7 shows the distribution of match outcomes in **unbalanced games** ($|\Delta\text{ELO}| > 7.5$), separating matches where teams play *with advantage* (higher ELO than the opponent) and *with disadvantage* (lower ELO than the opponent). Results are further split by age match-up: *young vs experienced* and *experienced vs experienced*.

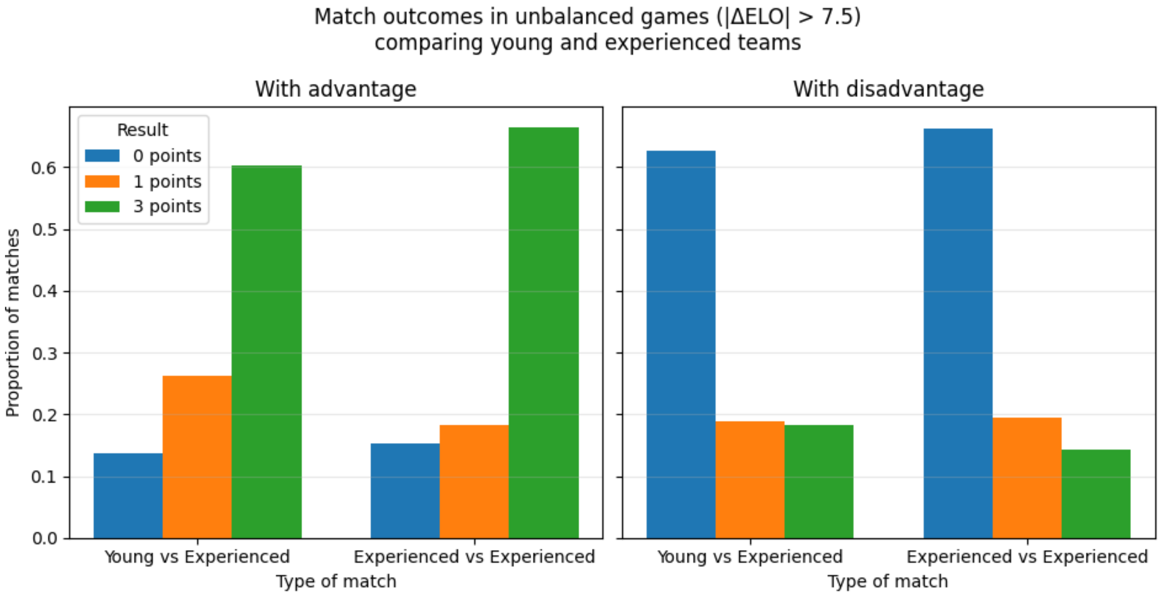

Figure 7: Match outcomes in unbalanced games ($|\Delta\text{ELO}| > 7.5$), comparing *young vs experienced* match-ups with matches between experienced teams. Results are shown separately for teams playing with ELO advantage and disadvantage.

When playing **with ELO advantage**, young teams display a slightly different outcome profile compared to experienced teams. While experienced teams convert a higher proportion of these matches into wins, young teams tend to draw more frequently and lose slightly less often. This suggests that, despite having a quality advantage, young teams may be less effective at fully capitalising on it, but they also appear less prone to outright collapse.

In matches played **with ELO disadvantage**, young teams again show a somewhat more resilient pattern: they lose slightly fewer matches and achieve a marginally higher proportion of wins compared to experienced teams facing similar disadvantage. Although these differences are small, they consistently point in the same direction, suggesting that younger squads may cope better with adverse competitive contexts.

Figure 8 presents the same analysis restricted to **balanced matches** ($|\Delta\text{ELO}| < 5$). In this setting, outcome distributions between young vs experienced match-ups and experienced vs experienced match-ups are remarkably similar. Young teams win marginally less often, but differences are minimal and well within what could be expected from random variation.

Findings Overall, the plots suggest that team age does not play a decisive role in balanced matches. In contrast, in unbalanced contexts, young teams appear to be slightly more resilient both when favoured and when unfavoured, although the magnitude of these differences is modest. These findings challenge the common narrative that younger teams systematically underperform due to lack of maturity, at least in terms of match outcomes measured by points.

Data analysis plan The visual patterns observed in these plots suggest that team age may be slightly associated with differences in performance, particularly in unbalanced matches. To formally assess whether these differences reflect a genuine effect rather than random variation or confounding factors, a follow-up analysis could employ regression-based models. For instance, logistic or multinomial regression could be used to model match outcomes as a function of team age group while controlling for ELO difference, home advantage and season effects. Additionally, hypothesis testing on estimated coefficients or bootstrapping techniques could be applied to evaluate the robustness and statistical significance of the observed trends.

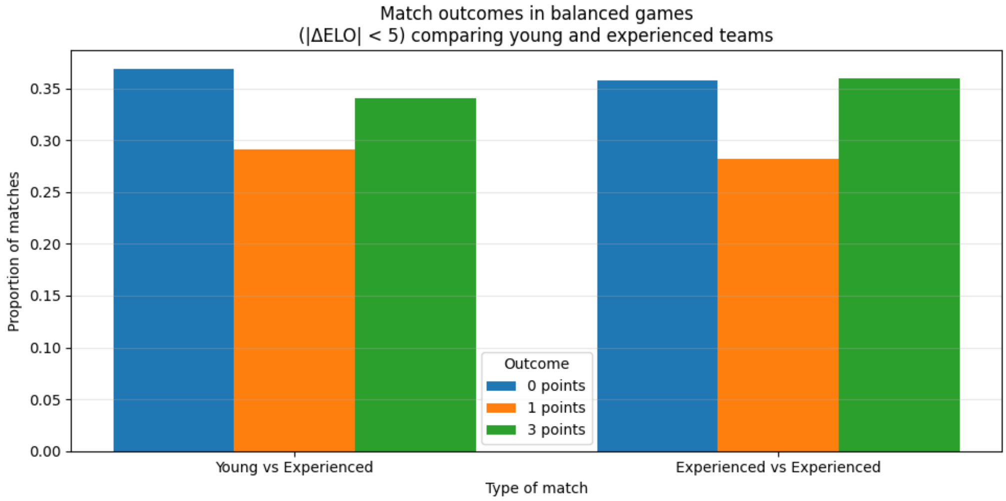

Figure 8: Match outcomes in balanced games ($|\Delta \text{ELO} < 5|$), comparing *young vs experienced* match-ups with matches between experienced teams. Outcome proportions (loss, draw, win) are shown for each match type.

4 Short-term impact of disallowed goals (VAR)

As an exploratory analysis, we investigated whether having a goal disallowed by the Video Assistant Referee (VAR) [1] affects a team's short-term defensive performance. The initial hypothesis was that the psychological impact of a disallowed goal could temporarily increase the likelihood of conceding shortly afterwards, compared to similar match situations without a disruptive event.

To explore this idea, we compared the cumulative probability of conceding a goal in the minutes following a VAR-disallowed goal with a control group of matches in equivalent match states (same scoreline, match minute and numerical situation) where no disruptive event occurred. Figure 9 shows the resulting cumulative curves.

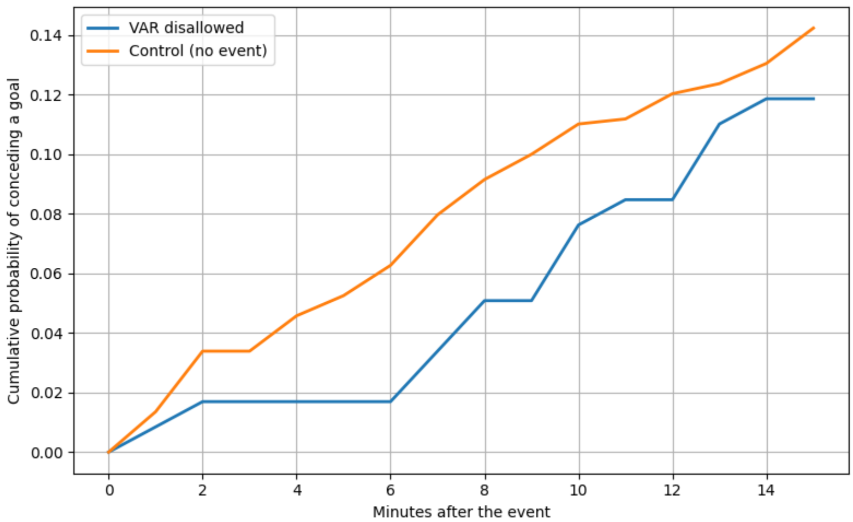

Figure 9: Cumulative probability of conceding a goal after a VAR-disallowed goal compared to a control group without disruptive events.

Contrary to expectations, teams whose goals were disallowed by VAR do not appear to concede more frequently in the short term; if anything, the control group exhibits a slightly higher cumulative probability. This result is likely driven by limitations in the available data rather than a genuine absence of effect. In particular, the dataset lacks detailed match statistics (e.g. possession, territorial dominance, shot quality or pressure metrics) that would be required to ensure that the compared situations are truly equivalent. As a result, this analysis is reported as a dead end, illustrating the importance of contextual richness when attempting to isolate short-term psychological effects in football matches.

References

[1] Video Assistant Referee (VAR). Wikipedia. https://en.wikipedia.org/wiki/Video_assistant_referee

[2] Football Match Event Dataset (LaLiga Seasons). Zenodo. https://zenodo.org/records/7341037

[3] Definition of *Goleada* in Association Football. Wikipedia (Spanish). https://es.wikipedia.org/wiki/Goleada#F%C3%BAtbol

[4] Cañizares, S. (Former professional footballer). Wikipedia. https://en.wikipedia.org/wiki/Santiago_CaÃšizares

[5] Santiago Cañizares on sportsmanship after heavy defeats, Sport.es (Spanish): https://www.sport.es/es/noticias/barca/santi-canizares-abrazo-joan-garcia-125482048