

Northeastern University
Mechanical and Industrial Engineering Department
IE 6200: Engineering Probability and Statistics
Prof. Rehab Ali
Section 7- Fall 2024

**EVALUATING THE IMPACT OF NEW TIMEKEEPING RULES ON MATCH DYNAMICS
IN THE ENGLISH PREMIER LEAGUE**

Final Report



**Northeastern
University**

Group 2

Team Members

Akshat Hiten Shah	002345822
Mayur Mahavir Bijarniya	002059856
Jay Gherwada	002475075
Pratyush Wasnik	002336460
Rushabh Vora	002320956
Savan Awanti	002448893

PROBLEM STATEMENT

The 2023-24 English Premier League season introduced new timekeeping rules aligned with IFAB guidance, aiming to increase the accuracy of additional time calculation and enhance ball-in-playtime. This means timers in stadiums will now be left running until the completion of each half, including additional time in both the first and second half. Match officials are committed to ensuring a more accurate calculation of additional time, as well as an improvement in the amount of time the ball is in play. Therefore, the exact time lost when certain match events occur will now be added, as opposed to the previous policy of a nominal period being added for particular game incidents.

These match events include:

- Goals and subsequent celebrations
- Substitutions
- Injuries and treatment time (if required)
- Penalties (from moment of offence to the whistle for the penalty kick)
- Red cards (from the moment of offence to when the player leaves the field)

This study will focus on several critical aspects of the Premier League's new timekeeping rules and their effects on match dynamics. The likelihood and frequency of goals occurring during extended playtime compared to previous seasons will be the key statistic of this study. This will provide insights into how these changes have altered late-game scoring patterns. The research will also examine injury rates to assess the physical impact of longer playing times.

In-scope:

1. Analyze the likelihood and frequency of goals occurring during extended playtime compared to regular play periods.
2. Assess the relationship between increased match duration and player injuries.
3. Assess the relationship between increased match duration and disciplinary actions (yellow and red cards issued during matches).
4. Assess the relationship between increased match duration and substitutions made.

Out of scope:

1. Analysis of viewership trends and fan engagement metrics.
2. Evaluation of financial implications, including revenue changes for clubs and the league.

The project will deliver a comprehensive report detailing the impact of new timekeeping rules on goal-scoring patterns and player injuries in the Premier League. This report will include statistical analysis and data visualizations, providing insights into how extended playtime affects goal distributions and injury rates.

PROJECT GOALS

The goals of this project align closely with the in-scope elements, focusing primarily on the impact of extended playtime on goal-scoring patterns and player injuries in the Premier League. Our main objectives are to:

1. Determine how new timekeeping rules impact goal-scoring frequencies, analyze the effect of extended playtime, and compare regular and extended playtime distributions across seasons. This divides our scope, and we aim to examine the likelihood and frequency of goals occurring during extended playtime. An analysis of how the new timekeeping rules have influenced late-game scoring opportunities and overall goal distributions.
2. Assess the relationship between increased match duration and player injuries, examining injury rates with extended playtime. This goal corresponds to our in-scope focus on injury rates correlated with match duration and player fatigue indicators for consistent teams. An analysis of how the longer playing times affect player health and performance, potentially identifying any increased risks associated with extended playtime.
3. Examine the impact on team tactics like substitutions, particularly on goal-scoring opportunities and injury prevention strategies. This aligns with our in-scope analysis of how teams adapt their play during extended playtime, which may influence goal-scoring patterns and injury management.
4. Evaluate the number of cards issued by match officials, specifically during extended playtime. Time wasting is one of the ways to get a yellow card. This goal fits our in-scope analysis of how players and teams adapt to the new timekeeping rules.

By focusing on these goals, a comprehensive analysis of how the new timekeeping rules have affected critical aspects of Premier League matches, particularly regarding goal-scoring dynamics and player welfare, while staying within the project's defined scope. The accuracy and statistical significance of our goal-scoring probability models, the comprehensiveness of our injury correlation analysis, and the clarity and actionability of our insights regarding team tactics and competitive balance will measure the success of this project. If successful, this project is expected to yield significant benefits, including a data-driven understanding of the rule's impact on fairness and competitiveness, which can inform the refinement of league policies. Our insights into injury patterns could improve player welfare strategies while analyzing goal-scoring probabilities, and tactical adaptations could help clubs develop more effective game strategies. Ultimately, these outcomes could lead to a more exciting, fair, and sustainable Premier League, benefiting players and the broader football community.

DATA IDENTIFICTION, COLLECTION AND PREPARATION

1. Model Assumptions:

- The data provided on the Official Premier League website [1], Fotmob website [2], and The New York Times website [3] is trusted.
- The data gathered is up-to-date, accurate, and complete.
- Statistical software such as RStudio, MiniTab and MS Excel will perform as expected for statistical analysis.
- Statistical data of extended playtime added after 45 minutes is not considered.

2. Data Preparation:

- The data preparation phase will involve collecting datasets from the 2021-22, 2022-23, and 2023-24 seasons [1] [2] [3]. This will include detailed match statistics such as goal timings, injury occurrences, substitutions, disciplinary actions, and extended playtime details.
- The preparation process will ensure that all data is consistent across seasons, allowing for meaningful comparisons before and after implementing the new timekeeping rules. Any anomalies or missing values will be appropriately addressed through data imputation techniques or by consulting additional sources to fill gaps.
- This rigorous preparation will create a robust dataset for analyzing the impact of extended playtime on goal-scoring likelihood and player injuries in the Premier League.

3. Data Source:

This dataset was created using data from the below sources:

- *Official Premier League website* [1]: A reliable resource for official statistics, including goals scored, yellow and red cards, injuries, and substitutions.
- *Fotmob Website* [2]: Offered how much additional time was added after 90 mins.
- *The New York Times website* [3]: Provided data on the number of injuries.

To ensure accuracy and completeness, cross-referencing this data with official match reports and statistics provided by the Premier League may be done. These sources ensure data accuracy and consistency, providing a robust foundation for detailed performance analysis.

4. Data Handling and Cleaning:

- The data for this analysis was collected using web scraping techniques from two primary sources: the official Premier League website [1] and the Fotmob website [2].
- The Python library “Beautiful Soup” was used for scraping the Premier League site [1], while “Selenium” was utilized for the Fotmob website [2], allowing for efficient and accurate data extraction.
- The data will be cleaned to ensure completeness and accuracy, particularly for match event timings. Special attention will be given to extended playtime.
- Through RStudio, data inconsistencies were checked and verified with data types. This method filled data gaps and reduced bias in our analysis, ensuring our results' reliability and precision.

5. Target Population:

- 3 seasons of English Premier League across the 2021-22, 2022-23, and 2023-24 seasons, each consisting of 380 matches and 20 teams.
- Players are a primary focus, as the study will investigate the effects of extended play on their performance, as well as injury rates.
- Interdisciplinary actions and substitutions are dependent on match officials and coaches, respectively. However, the study emphasizes how extended playtime impacted these parameters.

6. Performance Measures:

- *Extra time:* Minutes added by match officials after 90 minutes of play.
- *Goals Scored:* Total number of goals scored by teams, both at home and away.
- *Goals Scored in added time:* Total number of goals scored by teams, both at home and away, after 90 minutes of regulation time.
- *Disciplinary Actions:* Yellow and red cards were issued during matches, reflecting team discipline.
- *Substitutions made in added time:* Number of substitutions made by teams after 90 minutes in each match.
- *Injuries:* Number of players injured for each team.

DATA VISUALIZATION:

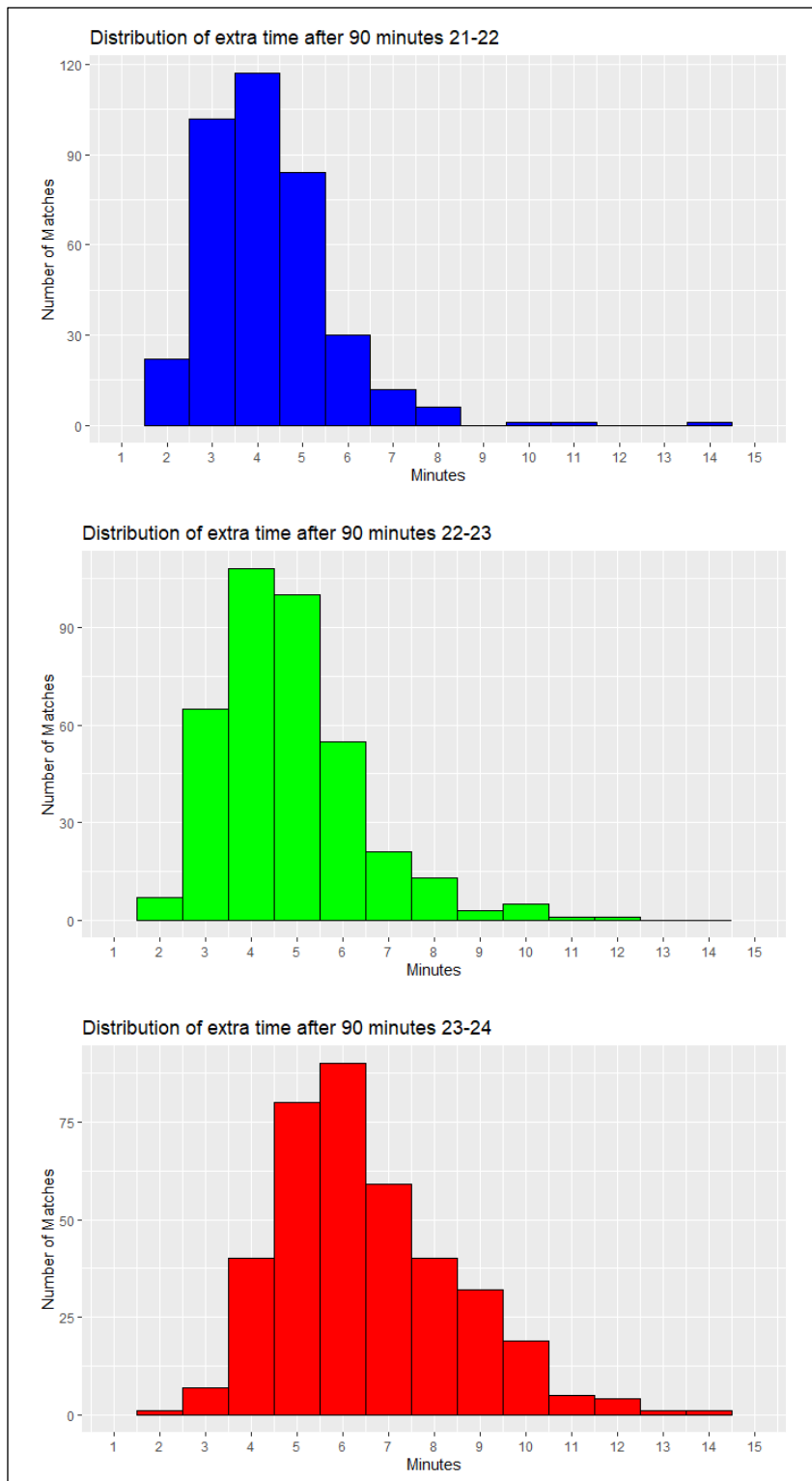


Figure 1: Comparison of minutes added across 2021-22, 2022-23 and 2023-24 seasons

Following the rule change to timekeeping since the 2023-24 season, it was expected the minutes of extra time added by match officials to increase. Data was collected from the FotMob website [2] to verify this.

Figure 1 compares the number of minutes added after 90 minutes of play across the three seasons. The X-axis represents the minutes added, while the Y-axis shows the number of matches.

For the 2023-24 season, the data shows a higher number of added time compared to the 2021-22 and 2022-23 seasons. The mode for both 2021-22 and 2022-23 seasons is 4 minutes, whereas it is 6 minutes for 2023-24 season. These indicate that there were higher instances of long added time in 2023-24 season.

Visually, 2021-22 and 2022-23 seasons have a clear positive skew. The 2023-24 season on the other hand has a slight positive skew.

There are outliers present for the 2021-22 and 2022-23 seasons on the extreme right, implying adding a higher amount of time was unusual. The 2023-24 season, however does not show any outliers.

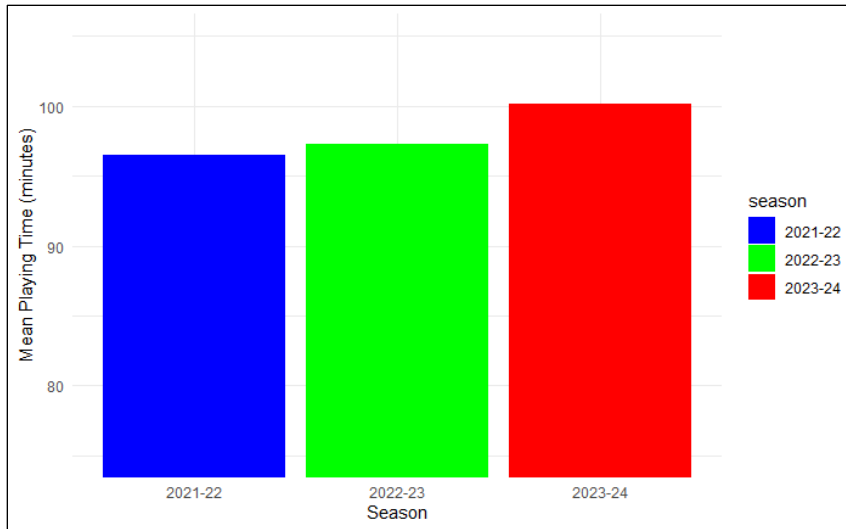


Figure 2: Average minutes of playing time per season

Furthermore, as shown in figure 2, the mean playing time in the 2023-24 season is higher than that for both the 2021-22 season and the 2022-23 season. This reflects a clear upward trend, suggesting that, on average, matches in 2023-24 had longer periods of added time than in the previous two seasons. This highlights a notable shift directly because of the rule change introduced in 2023-24 season.

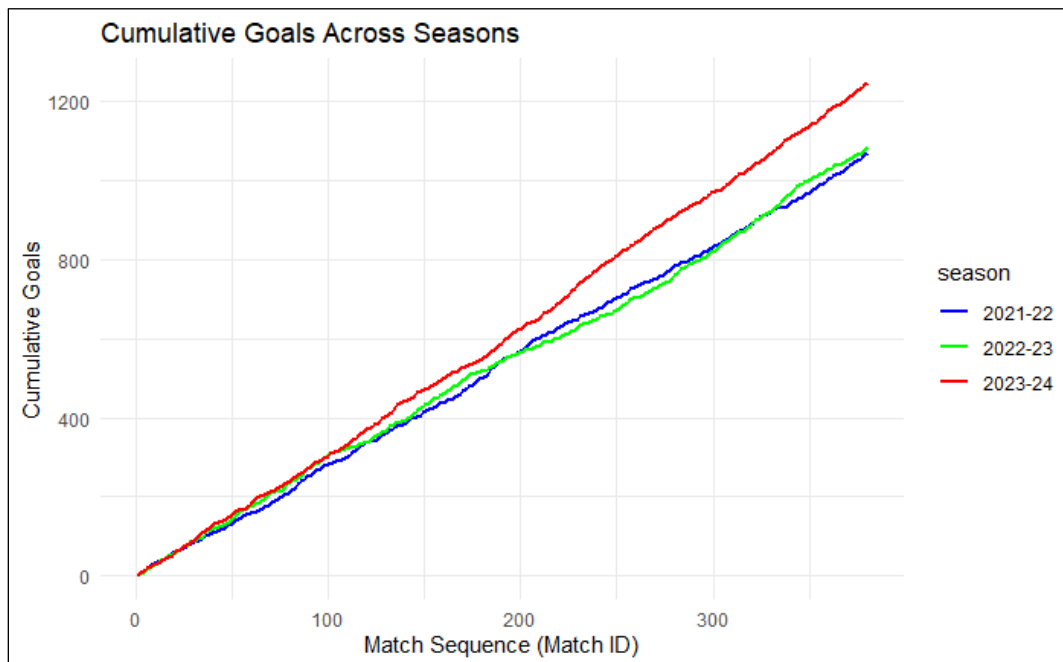


Figure 3: Line graph comparing total goals scored in 2021-22, 2022-23 and 2023-24 seasons

The increase in playing time has impacted on the number of goals scored throughout the season, as shown in figure 3 above. The data comparing 2021-22, 2022-23, and 2023-24 seasons was collected from the official premier league website [1]. Figure 3 shows the cumulative goals scored on the X-axis and match sequence (total of 380 matches in a season) on the Y-axis. It is evident that the total number of goals has seen a jump for the 2023-24 season compared to the previous two seasons, which trace a similar line.

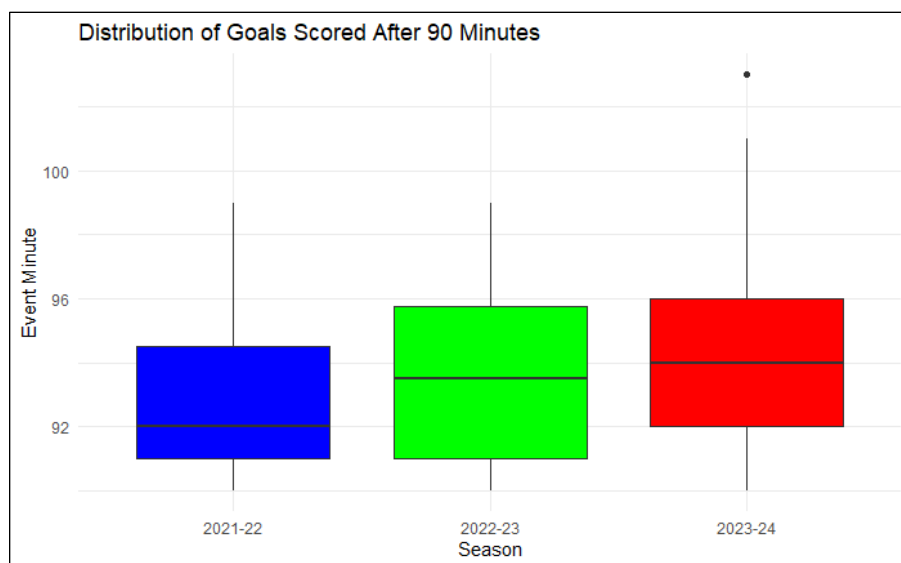


Figure 4: Boxplot comparison of the distribution of goals scored after 90 minutes in 2021-22, 2022-23 and 2023-24 seasons

As shown in figure 4, the box plot compares the distribution of goals scored after 90 minutes across 3 premier league seasons (2021-22, 2022-23, 2023-24) [1]. The 2023-24 season has the highest median minute event for goals scored. The range is higher and there is an outlier present for the 2023-24 season. This implies that goals were scored later in the 2023-24 season compared to the previous two seasons.

The graph illustrates the total number of injuries among 14 Premier League teams over three seasons: 2021-22, 2022-23, and 2023-24. These 14 teams were selected for analysis due to changes in team rankings related to the promotion and relegation rule[1]. The number of injuries for these 14 teams is compared based on data gathered from the New York Times article [3]. Since it is not in the scope to make inferences for each of these teams, the data has been simplified and represented as total injuries in a season in figure 5. The number of injuries decreased in the 2022-23 season from the 2021-22 season. However, there was a rise in the 2023-24 season, resulting in the highest number of recorded injuries during this period. The timekeeping rule change, which resulted in longer playing time could be one of the factors contributing to the increase in injuries.

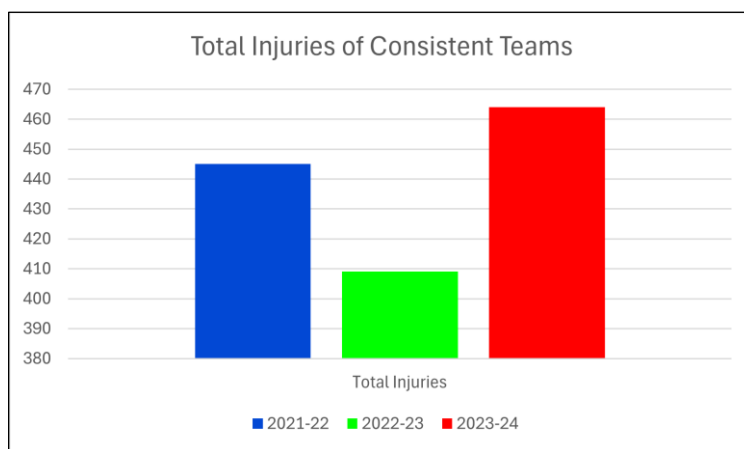


Figure 5: Total injuries for 14 teams combined in 2021-22, 2022-23 and 2023-24 season

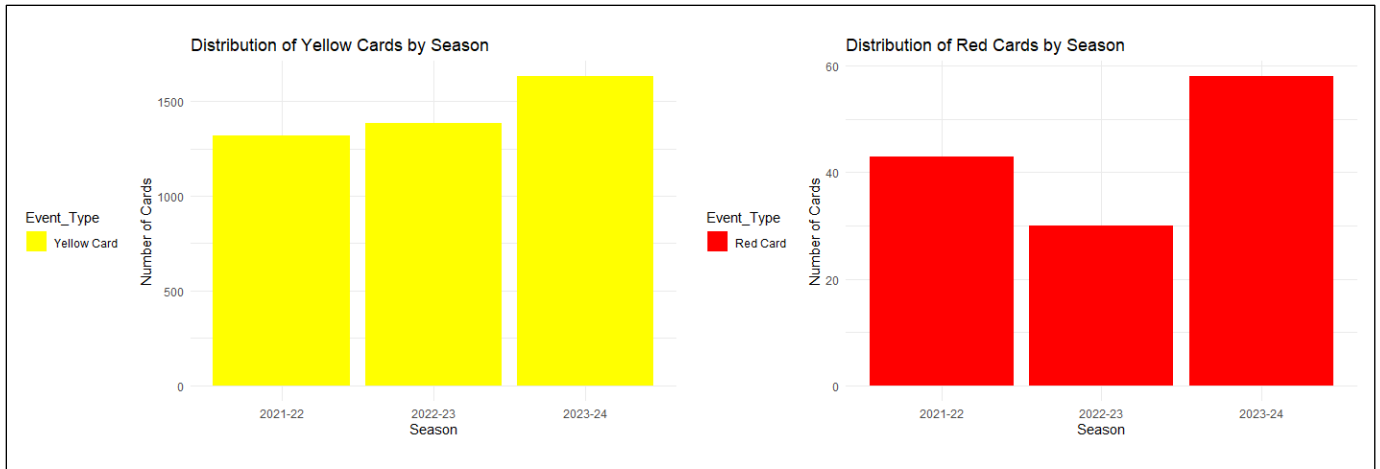


Figure 6: Number of red cards and yellow cards issued in 2021-22, 2022-23, and 2023-24 season

While tactical and reckless fouls are the most common offenses resulting in disciplinary actions, time-wasting and dissent are the offenses that players tend to commit in the final minutes of the match.

Figure 6 above compares the total number of yellow cards and red cards issued in each season [1]. There is an upward trend for both yellow and red cards. There is no way to conclude the exact reason for this. But players could waste time to counter the increase in playing time. Another theory is that players lose focus due to longer playing time and commit reckless fouls.

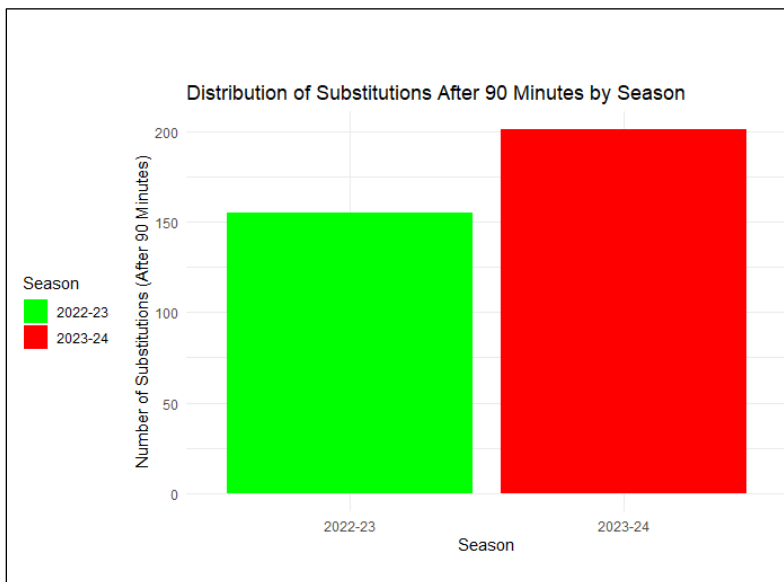


Figure 7: Comparison of number of substitutions made after 90 minutes

Since the rules changed in 2022 to allow five substitutions instead of three[1], data is considered from two seasons. The bar graph in figure 7, with the X-axis representing the seasons and the Y-axis indicating the number of substitutions made, clearly illustrates a significant increase in the number of substitutions during the 2023-2024 season compared to the 2022-23 season, particularly after the 90-minute mark. This trend shows that substitutions have increased during the later stages of the game. With the new rule allowing for more added time, teams have had to adjust their gameplay to maintain the tempo of the match.

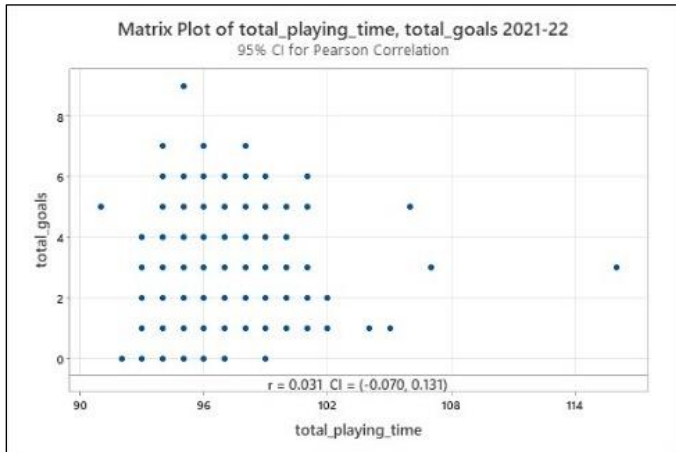


Figure 8: Correlation plot for 2021-22 using MiniTab

The correlation coefficient between total playing time and total goals is 0.031 for the 2021-22 season, indicating a very weak positive correlation. This suggests that total playing time has almost no linear relationship with the total goals scored in each match.

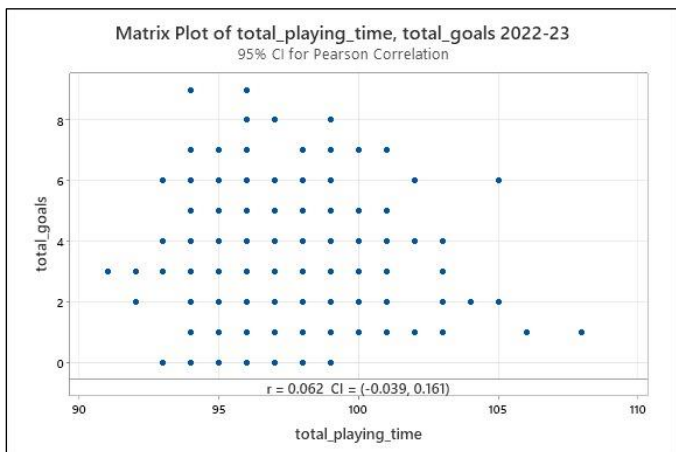


Figure 9: Correlation plot for 2022-23 using MiniTab

The correlation coefficient between total playing time and total goals is 0.062 for the 2022-23 season, indicating a weak positive correlation. This suggests that total playing time has almost no linear relationship with the total goals scored in each match.

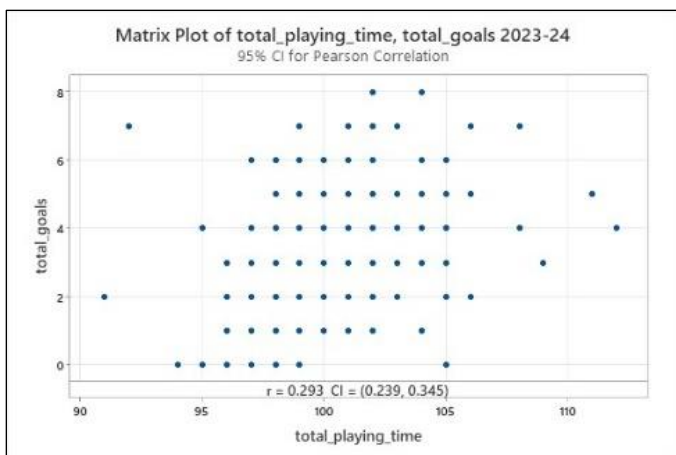


Figure 10: Correlation plot for 2023-24 using MiniTab

The correlation coefficient of time and total goals is 0.293 for the 2023-24 season, indicating a positive but weak correlation between total playing time and total goals. This means there is a slight tendency for goals to increase with extended play, though the effect is not strong.

STATISTICAL ANALYSIS:

The purpose is to study the impact of extended playtime on goals scored, player injuries, disciplinary actions, and substitutions. We used statistical analysis methods like confidence interval and hypothesis testing to get evidence-based inferences about the data collected and visualized. All results are from either R or MS Excel and calculated by formula. Hence verified.

A. Impact on goals scored

We calculated the confidence interval for overall goals scored per match at 95% confidence level.

$$CI = \bar{x} \pm 1.96 \cdot s / \sqrt{n}$$

$$CI = 2.835526 \pm 0.121741$$

For 2021-22 & 2022-23, **CI=(2.713785,2.957267)**

For 2023-24, **$\mu = 3.278947$**

```
## 95 percent confidence interval:
##  2.713785 2.957267
## sample estimates:
## mean of x
##  2.835526
```

Figure 11: Confidence Interval of overall goals scored per match using R

The Mean of season 2023-24 is outside the range of 2021-22 & 2022-23 season, indicating an increase in number of goals scored in season 2023-24.

We calculated the Hypothesis for goals scored after 90 minutes at 95% confidence level.

Parameter of interest: $\mu_1 - \mu_2$

Null Hypothesis: There is no significant increase in goals $H_0 : \mu_1 - \mu_2 = \Delta_0$

Alternative Hypothesis: $H_a : \mu_1 - \mu_2 > 0$

$$Z \text{ statistic: } z = \frac{(\bar{x}_1 - \bar{y}_1) - \Delta_0}{\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{m}}}$$

where:

\bar{x}_1 : Average of 2023-24 season,

\bar{y}_1 : Average of season 2021-22 & 2022-23

$$n = 760, m = 380$$

$$s_1 = 0.5215199, s_2 = 0.3904923$$

$$\Delta_0 = \mu_1 - \mu_2 = 0$$

$$\bar{x}_1 = 0.2710526, \bar{y}_1 = 0.1592105$$

$$\alpha = 0.005$$

After calculating, $z = 3.6946$

p-value = 0.0001101

```
## z = 3.6946, p-value = 0.0001101
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.06204942      NA
## sample estimates:
## mean of x mean of y
## 0.2710526 0.1592105
```

Figure 12: Two sample hypothesis using z-test for goals scored after 90 minutes using R

$$\alpha > \text{p-value}$$

We reject the null hypothesis and accept the alternative hypothesis.

This indicates an increase in the number of goals scored after 90 minutes in 2023-24 season.

B. Impact on player injuries

We calculated confidence interval for injuries per 14 consistent teams at 95% confidence level.

$$\bar{x} = 30.5, s = 8.5613256, n = 28$$

$$CI = \bar{x} \pm 1.96 \cdot s / \sqrt{n}$$

$$CI = 30.5 \pm 3.171101$$

For 2021-22 & 2022-23, **CI = (33.671101, 27.328899)**

For 2023-24, **$\mu = 33.142857$**

Combined average of 2021-22 and 2022-23	30.5
Standard Deviation	8.561326
z value	1.959964
Half-width	3.171101
Upper	33.6711
Lower	27.3289
Average of 2023-24	33.14286

Figure 13: Confidence interval of overall injuries per team using MS Excel

The Mean of season 2023-24 is within the range of 2021-22 & 2022-23 season indicating no significant change in number of injuries in season 2023-24.

C. Impact on disciplinary actions

We calculated confidence interval for overall cards issued per match at 95% confidence level
 $\bar{x} = 3.613158$, $s = 1.998771$, $n = 760$

$$CI = \bar{x} \pm 1.96 \cdot s / \sqrt{n}$$

$$CI = 3.61.3138 \pm 0.1421059$$

For 2021-22 & 2022-23, **CI = (3.471032,3.75526)**

For 2023-24, **$\mu = 4.368421$**

```
95 percent confidence interval:
 3.470828 3.755488
sample estimates:
mean of x
 3.613158
```

Figure 14: Confidence Interval of overall goals scored per match using R

The Mean of season 2023-24 is outside the range of 2021-22 & 2022-23 season indicating an increase in number of cards issued in season 2023-24.

D. Impact on substitutions

We calculated confidence interval for substitutions made per match after 90 minutes at 95% confidence level

$$\bar{x} = 0.4078947$$
, $s = 0.7404205$, $n = 380$

$$CI = \bar{x} \pm 1.96 \cdot s / \sqrt{n}$$

$$CI = 0.4078947 \pm 0.0744462$$

For 2022-23, **CI = (0.3332114,0.4825781)**

For 2023-24, **$\mu = 0.5289474$**

```
## 95 percent confidence interval:
##  0.3332114 0.4825781
## sample estimates:
## mean of x
## 0.4078947
```

Figure 15: Confidence interval of substitutions made per match after 90 minutes using R

The Mean of the 2023-24 season is outside the range of the 2022-23 season indicating an increase in a number of substitutions made per match after 90 minutes in 2023-24.

RESULTS AND CONCLUSIONS:

The goal of the project is to study the impact of new timekeeping rules introduced in the 2023-24 season in the Premier League. Comparing the data of the two seasons prior to the rule change, the study evaluated the relationship of this rule change with various parameters like goals scored, player injuries, disciplinary actions and substitutions using data visualization and statistical analysis.

The data showed that the frequency of more minutes added after 90 minutes had increased from previous seasons. This resulted in a higher average duration of the match, thus confirming the assumption that new timekeeping rules had increased the playing time.

The increased playing time led to an increase in overall goals scored per match as seen in the line graph and correlation plot. This is further verified using a confidence interval where the average of the 2023-24 season falls outside the confidence interval of previous seasons. Additionally, two sample hypothesis using a z-test for goals scored after 90 minutes were conducted revealing that there was a significant increase in goals scored after 90 minutes in 2023-24 season.

The data visualization showed an increase in player injuries for the 2023-24 season. However, the average of 2023-24 falls within the confidence interval of previous seasons, meaning the change was not significant. Therefore, extended playtime did not significantly increase player injuries based on statistical analysis.

Data visualization shows that the number of cards issued per match saw a jump as well. The average of the 2023-24 season falls outside the confidence interval of previous seasons. This concludes that a significantly higher number of disciplinary actions took place per match post the rule change.

The extended playtime impacted the number of substitutions made after 90 minutes as seen in the data. The statistical analysis confirms the same using confidence interval concluding significant increase in 2023-24 season.

STRENGTHS:

- Integrating datasets of three seasons containing 380 matches each was a significant challenge. Before data visualization and statistical analysis, these datasets had to be combined and cleaned to ensure data consistency.
- Statistical analysis was executed successfully, giving valuable insights into how the new rules impacted various parameters. While there was a significant increase in goals scored, disciplinary actions, and substitutions, player injuries did not see a significant change. This shows the importance of statistical analysis in addition to data visualization.







WEAKNESSES:

- It has only been one complete season since the introduction of the new rule. Thus, data was limited to one season.
- This study does not consider the statistics of events occurring in extended time after 45 minutes.

PROPOSED NEXT STEPS AND FUTURE SCOPE:

The implementation of the extended playtime rule in the Premier League in 2023-24 season changed team and game management. Aimed at promoting fairness by accounting for stoppages, this rule has influenced goal-scoring patterns, disciplinary actions, injury rates, and substitution patterns. While initial analysis has shed light on some of these impacts, the relatively short duration of this rule's application means much more to be explored and understood. Analyzing more seasons of data will help verify the consistency of observed patterns and determine whether initial trends remain stable or evolve as teams adapt to the rule over time. One important aspect to consider in future research is the potential impact of team budgets on squad depth and player management strategies. Teams with larger budgets have more financial resources to build deeper squads, which can reduce mental and physical stress on players, especially during the added time at the end of match. Teams can also hire additional personnel to analyze and strategize how to get most of the rule change. The connection between increased playtime in the Premier League and higher revenue is complex. Longer matches can enhance viewer engagement by creating more thrilling moments and attracting more viewers and ad revenue. This could also boost matchday revenue from attendance, concessions, and merchandise and raise the value of broadcasting rights. Additionally, extended playtime may create more sponsorship opportunities. However, the outcome isn't certain, as factors like match quality, team performance, and fan sentiment will impact any potential revenue gains.

COLLABORATION:

Team Member	Contributions	Percentage	Signature
Akshat Hiten Shah	Data Identification, Problem Statement, Data Visualization, Report Formatting, PowerPoint Presentation	16.67	
Mayur Mahavir Bijarniya	Problem Identification, Data Extraction, R Programming, Data Visualization, Statistical Analysis	16.67	
Jay Gherwada	Problem Identification, Problem Statement, Data Visualization, Report Formatting	16.67	
Pratyush Wasnik	Project Goals, Data Collection, Data Visualization, PowerPoint Presentation	16.67	
Rushabh Vora	Problem Identification, Data Identification, Data Visualization, Report Formatting, Statistical Analysis	16.67	
Savan Sangamesh Awanti	Problem Identification, Data Cleaning, RStudio programming, Data Visualization, Statistical Analysis	16.67	

BIBLIOGRAPHY:

[1] Premier League, "Results". Accessed Nov. 6, 2024. [Online].

Available: <https://www.premierleague.com/results>

[2] The New York Times, "Injuries 2022-2023". Accessed Nov. 6, 2024. [Online].

Available: <https://www.nytimes.com/athletic/4612629/2023/07/02/injuries-premier-league-2022-23-data/>

[3] FotMob, "Premier League 2023/2024". Accessed Nov. 6, 2024. [Online].

Available: <https://www.fotmob.com/leagues/47/overview/premier-league?season=2023-2024>