Football Radar Task

Table of contents

Introduction

Task

The task is to answer this question: “*Based on the data in this dataset, do red cards seem to lead to more goals?*”

We mean more goals in general here - we might expect the team with the red card to get fewer goals than we would have otherwise expected, and their opponent to get more, but the question is whether these effects combined lead to more goals overall.

About the data and some initial EDA:

The dataset referenced in the task description comes from games.csv and events.csv files provided. We opened with some exploratory data analysis in order to get to grips with what data we had available to us and how much of it.

Our games data set contains 20370 games from across 10 different European leagues of varying standards. Each game has a unique *game_id*. Other columns included *home_team, away_team, competition, date, home_goals,* and *away_goals.* I updated our data frame immediately with a *total_goals* column, adding home and away goals.

Our events dataset contains every occurrence of a goal or red card from the games in the previous dataset. The events dataset has columns *game_id, minute, side,* and *type.* Side is the team it occurred for and type is one of “*red_card”* or “*goal*”. Using the events, I added a column named “*any_red”* to the games dataset specifying whether a red card had occurred in that particular game.

We found 3731 red cards across 3273 games, meaning multiple reds occurred in some games. 16.08% of games had at least one red card. 0.1832 red cards per game on average, nearly 1 every five games. 54019 goals were scored in total, giving an average of 2.65 goals per game.

Explanation of approach

My logic for approaching this question was to ask the most obvious questions first and to incrementally introduce granularity and complexity to analysis, thereby hopefully increasing the precision and robustness of our answers in return.

For instance, our first analysis below asks the very basic question of whether games in which at least one red card occurs have more goals in general than games without, while our final mode of analysis applies a more complex survival analysis method to take account for the time red cards occur in games.

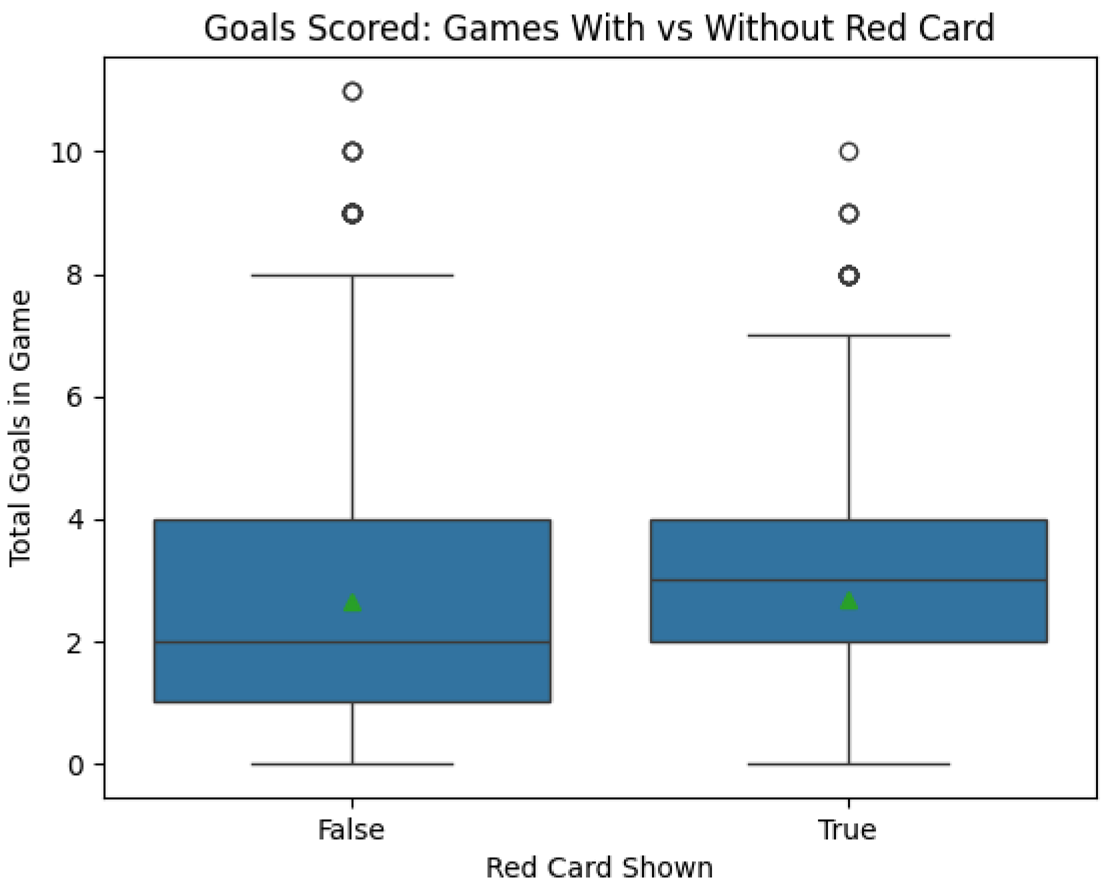
Analysis 1: Games With Versus Without Red Cards

Set up

The question we aim to answer in this initial analysis is whether games in which there is at least one red card tend to have higher goal counts than games without any red cards. We will do this by totaling the goals scored in each game, taking each game as a separate event. As specified in the task outline we're not concerned with who scores more or less goals, we are concerned with whether more goals are scored in general. We can then visualise the distribution of goals scored in each game for each group, games with red cards and games without red cards.

Results

Our findings are visualised in the following box-plot.



Median for games without reds are one lower (2) than with reds (3). However the means are quite close, with games with reds scoring 0.05 goals per game more (2.69) than games without (2.64). There seems to be a much higher upper quartile for the games without reds as well as some extreme outliers. These might explain why the mean is so much closer and might provide us with enough reason to doubt the small difference in means.

Hypothesis test

We employ a hypothesis test to answer our question of whether games in which there is at least one red card tend to have higher goal counts than games without any red cards. We chose a T-test for the means of two independent samples of scores because we make the reasonable assumption that games are independent of each other.

Hypothesis Test Summary

- Null Hypothesis (H_0): Games with at least one red card do not have more goals on average than games without red cards.
- Alternative Hypothesis (H_1): Games with at least one red card have more goals on average than games without.
- Test Used: One-tailed independent t-test

The test gave us a ***test-statistic* of 1.63** and a ***p-value*** of **0.0512**.

Conclusions

Using the usual confidence of level of 95%, we need the p-value to be less than 0.05 to accept the alternative hypothesis and reject the null hypothesis. So in our case, there is not enough evidence to say more goals are scored on average in games with at least one red card.

Limitations

1. Red cards may be more common in games that are *already* wild and higher scoring. This is called reverse causality.
2. We could have accounted for the uneven sample sizes better.
3. The timing of red cards are not accounted for here. For example, if the red happens late in the game, it may not have had time to influence the number of goals scored.
4. Similarly, because most reds occur later in the game (average of 63 minutes in), games that have a red are still played predominantly without any reds.

We will mitigate these limitations in the second analysis.

Analysis 2: Minutes With Versus Without Red Cards

Set up

In this second analysis, we focus on the goal-scoring rate during the periods of matches when red cards are in effect. This allows us to isolate the impact of red cards on scoring more precisely than in the previous analysis, where entire matches were grouped based on red card presence, regardless of when the red card occurred.

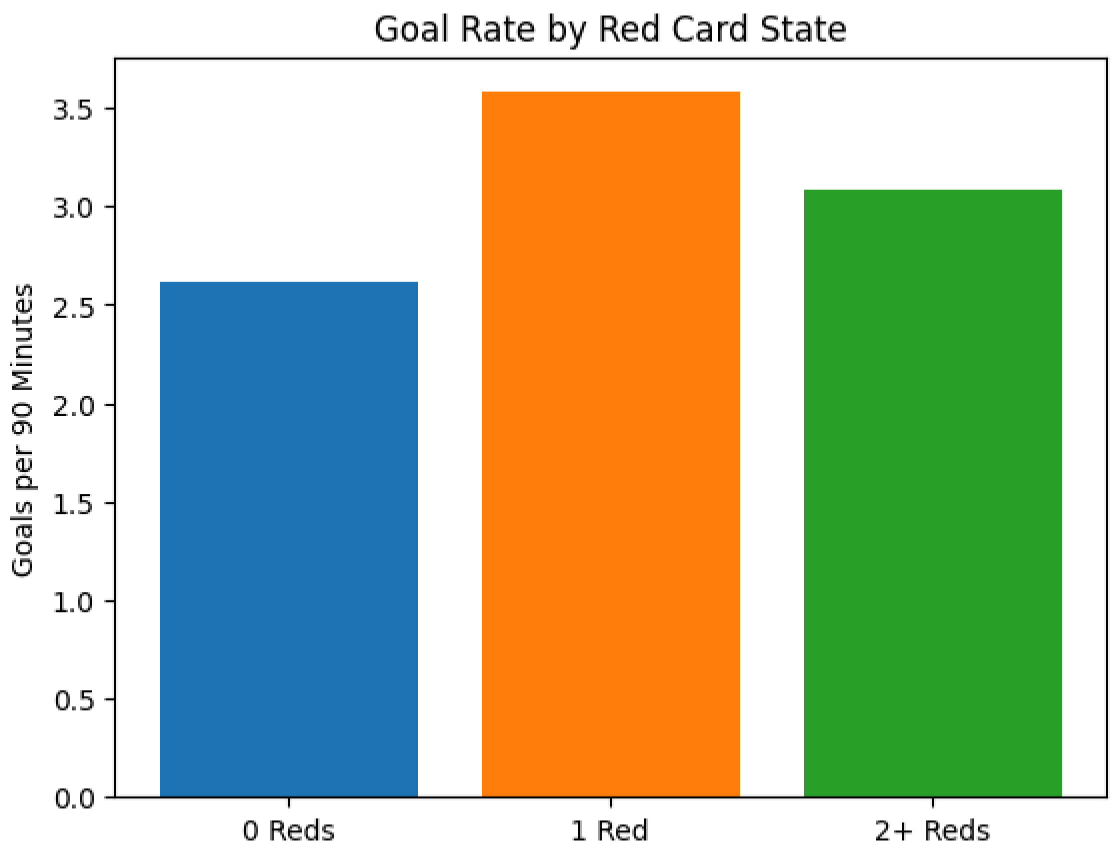
Steps:

- We calculated the total number of minutes played and goals scored while 0, 1, and 2+ red cards were active on the pitch.
- We calculated goals per minute and goals per 90 minutes for each red card state. As football is played over 90 minutes, this second representation is a bit more intuitive.
- We chose not to break down further the 2+ red cards group, as only 0.24% of games had 3 or more red cards, and these games are both rare and atypical. This would have made results for these individual bands very noisy.

We can make a more direct statement about the impact of red cards on scoring by comparing goals per minute (and per 90) across these red card states.

Results

Our findings are visualised in the following bar chart:



Now we begin to see a clear difference. **Average goals per 90** minutes of **2.61, 3.58**, and **3.09** are recorded in the bar chart above. A clear increase in average when one red is in place. Somewhat surprisingly the average drops to 3.09 when we go to 2+ red cards. This could be due to the rebalancing of numbers each team has on the field. It could also be that minutes with 2 red cards occur even later than the first and are almost certainly in quite a chaotic game. These averages seem to suggest a higher average number of goals scored across minutes played with a red card than without. They also suggest more red cards do not suggest a bigger increase in goals scored. For this reason, for the remainder of our paper, we'll only consider two groups: games/minutes in which there was *at least one* red card present and games/minutes in which there was *no* red card present. The average goals per 90 for minutes with 1+ red cards is 3.54.

Hypothesis test

We utilise a two-proportion Z-test by assuming independence across minutes of football played. As such each minute is treated as an independent trial. This is a simplification of course and limitation of this analysis. As the old saying goes, anything can happen in football, but one minute is certainly independent from the previous one. For example, a goal in one minute can change tactics and energy levels in the following minutes. However, this assumption is reasonable at scale and allows us to compare goal-scoring rates in periods with and without red cards. Each goal is a success for we find proportions and compare for each group.

Hypothesis Test Summary

- Null Hypothesis (H_0): Minutes played with at least one red card have the same number of goals on average as minutes played without red cards.
- Alternative Hypothesis (H_1): Minutes played with at least one red card have more goals on average than games without.
- Test Used: Two-proportion Z-test

The test gave us a ***test-statistic* of 17.314** and a ***p-value*** of **0.00001**.

Conclusions

We found extremely strong statistical evidence that goals are more likely to be scored on average in minutes where at least one red card is in effect. The p-value was effectively 0, suggesting the observed difference is almost impossibly due to chance.

Combining the results of analysis 1 with analysis 2, we could argue that while games with red cards aren't inherently more open or higher scoring, the empirical evidence suggests more goals are definitively scored in the minutes following the first red than in minutes without a red card.

Limitations

1. **Minutes as independent events:** Our analysis treats minutes as independent events when in reality a goal in one minute certainly affects the probability of a goal in the following minutes.

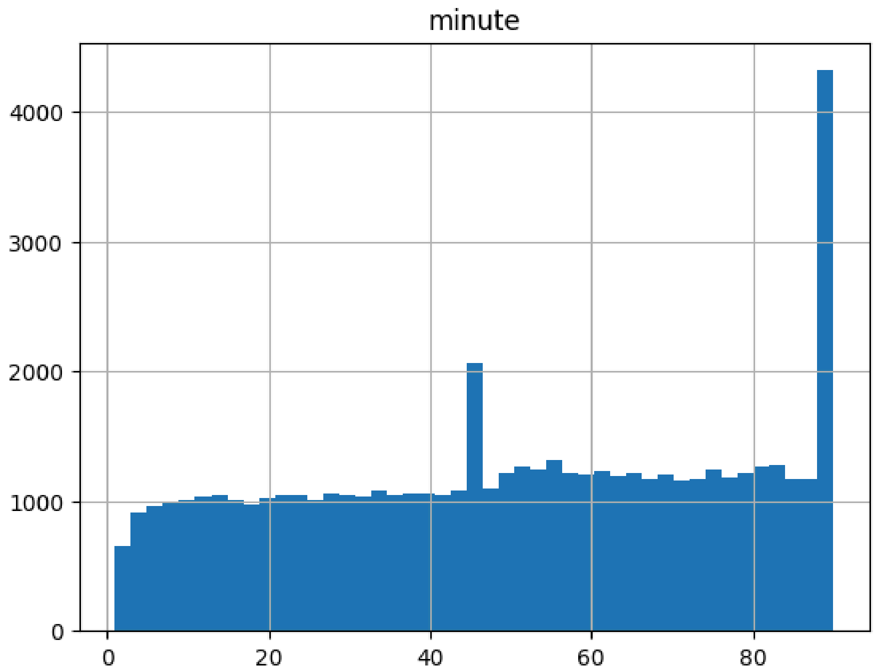
2. **Timing of red cards:** Most red cards occur in the second half meaning most minutes played with a red card are also in the second half. This means non-red carded minutes from the first half are of little use for comparison.

We mitigate this second limitation in Analysis 3.

Analysis 3: Relevant Minutes Comparison

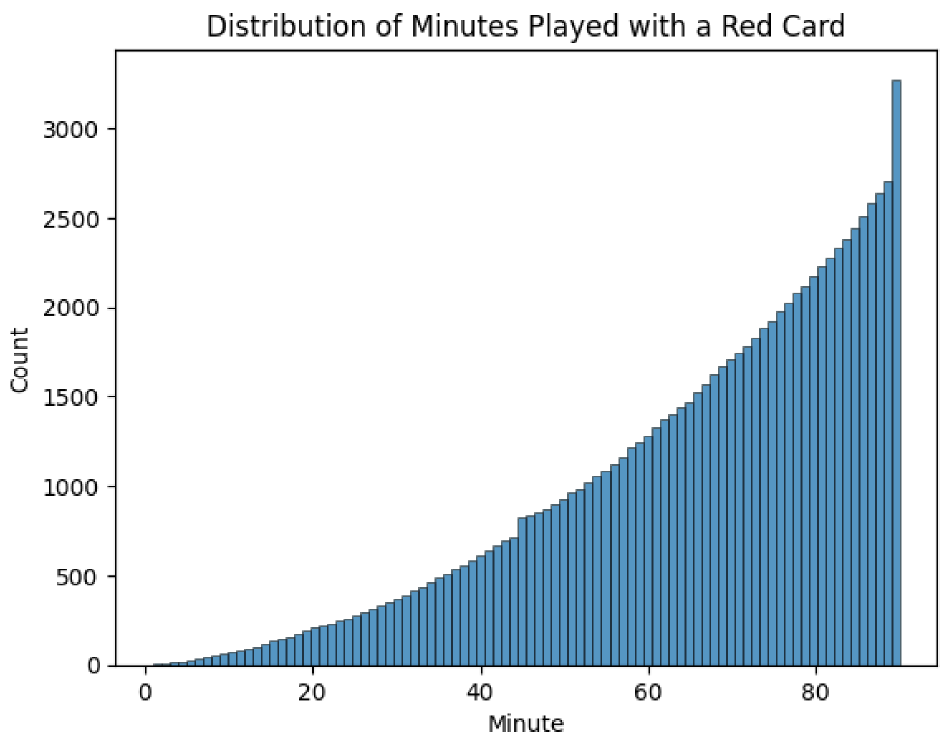
Set up

The previous analysis suggests conclusively that in general more goals are scored in the minutes played with a red card than in the minutes where no red card has been given. However, this could be due to many factors. For instance, take this point: **the average first red card time is 63 minutes** with standard deviation 23.6 (taken from our data). This means that the majority of minutes played with a red card are the minutes towards the end of a game. Maybe more goals are scored in the later stages of games anyways and this factor is conflating our results? We found this to be true, as shown by the histogram which shows the number of goals scored in each minute bin throughout the game. We found spikes at 90 and 45 minutes because of the limitations specified in the task description: minutes are capped at 45 and 90 minutes for each half so any even occurring in injury time will be included in these minute bins.

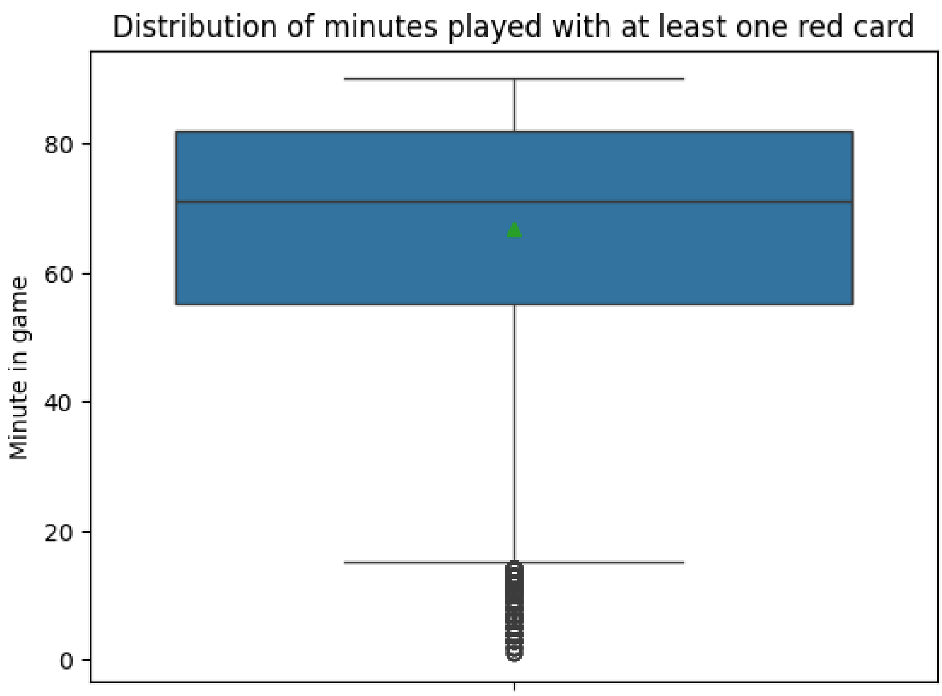


To mitigate this factor, instead of taking all minutes played without a red, let's now only take minutes from a similar time frame as those played with at least one red.

Let’s first look at the distribution of the minutes played with red cards.



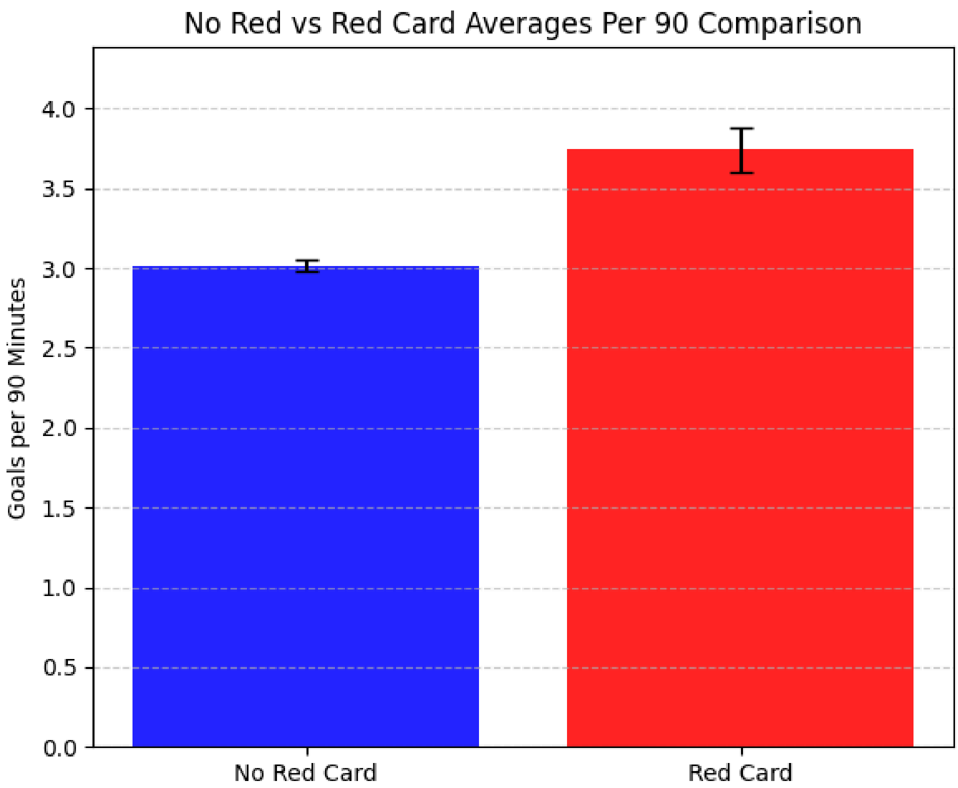
The counts in the histogram rise exponentially. We can see the distribution of these minutes further in the box plot. It shows: Lower quartile= 55, Higher quartile= 82.0, Median= 71.0



This means 75% of the minutes played with a red are after the 55th minute. Therefore to make a fairer comparison, we decided to only count goals and minutes from after this minute for both red and non-red groups. We then compared and tested the hypothesis similarly to in the previous analysis. We also ran the analysis for a range of cut-off points from 55 to 90 minutes which we've included full statistics on in the appendix.

Result 1: Comparing Goal Scoring Rates

The following is a bar chart visualising our results:



The average goals per 90 for minutes without red cards rises from 2.61 to 3.01 when we only consider after the 55th minute. The same average for red carded minutes increase from 3.54 to 3.74 when we apply the same cut-off. The percentage increase from non-red to red minutes is 24.1%. Certainly a sizable increase still.

Hypothesis test

Hypothesis Test Summary

- Null Hypothesis (H_0): Minutes played with at least one red card have the same number of goals on average as minutes played without red cards.
- Alternative Hypothesis (H_1): Minutes played with at least one red card have more goals on average than games without.
- Test Used: Two-proportion Z-test

Applying the same two-proportion Z-test as in analysis two, we calculate a **test-statistic** of **10.705** and a **p-value** of **0.00001**.

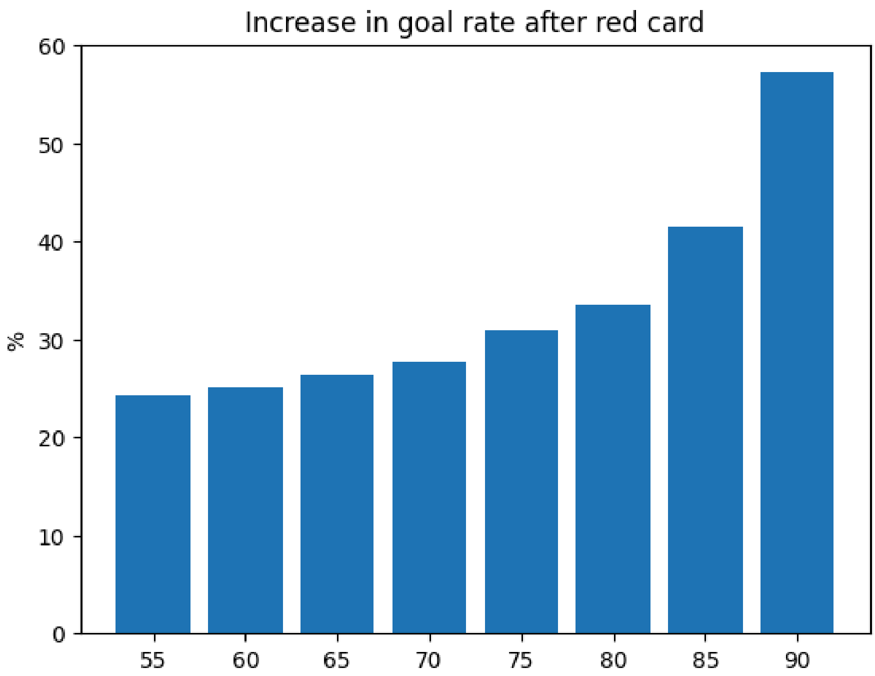
Conclusions

While the margin narrows somewhat when we apply a cut-off at the 55th minute, this analysis still conclusively suggests that the likelihood of goals increases in general when there is a red card present.

Result 2: Comparing Increase Rates by Varying Cut-Off Points

We calculated the percentage increase from non-red minutes to red minutes for a range of different cut-off points, as in the minute from which we consider minutes and goals for both groups. When we increase the cut-off point we see a definite rise in increase-rate from non-red to red minutes. With the cut-off at 55 minutes, the increase rate was 24%. At 85 minutes, the increase rate was 41%.

These results are visualised as follows.



Conclusions

This seems to suggest that red cards have an even greater effect the closer to the end of the game we get. More goals are scored in general the closer to the end of games we get. However, surprisingly, the rate of scoring increases much more in red card games than in non-red card games.

In general, our sample size for games without a red was much bigger than games with a red. Because we count the goals in the 90th minutes but not the extra minutes, we would actually expect the opposite of our findings here. We'd expect the goal rate for non-red games to be more inflated the closer we got to 90 minutes for the cut-off point. More games means more minutes missed, meaning more times including this 90th minute, meaning a bigger scoring rate. This only adds to the weight of our findings, and suggests the effect of the red card is more pronounced the later in the game we get.

That said, games with a red card in them also tend to have more injury time added on. A red card is a huge disruption in the game and is often followed by a tactical rebalance or substitutions by the team for whom it was received.

Further investigations and possibly more precise data would be needed to ascertain the reasons for the results we see above.

Limitations

- **Injury Time Binning Bias:** As noted, goals scored during injury time are binned into minute 45 and 90. This artificially inflates goal counts in those minutes and prevents precise time-based comparisons, especially in the final phase of matches.
- **Exclusion of Injury Time Minutes:** Since actual injury time minutes aren't tracked, our goal rate calculations exclude those extra minutes. This introduces a **bias particularly for games with red cards**, where longer added time is common (due to stoppages, fouls, substitutions). This may undercount the actual minutes played with a red, underestimating goal rate during red card periods.
- **Imbalance in Sample Sizes:** There are far more minutes played without a red card than with one. Even after applying a 55-minute cutoff, the groups may not be fully balanced in terms of total minutes or number of unique matches, potentially affecting variance and statistical power.

Analysis 4: Difference between leagues

Set up

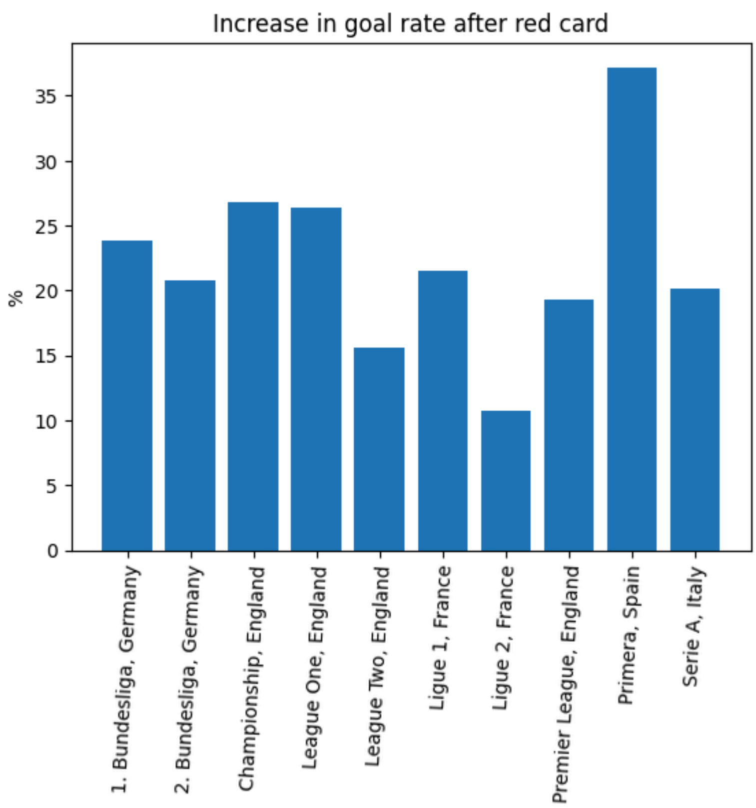
Next we want to assess whether the effect of red cards on scoring changes based on the league we're in. Does the standard of football affect the effect a red card has on the scoring? We'll repeat our analysis from part 3 taking only goals and minutes from the 55th minute.

Results

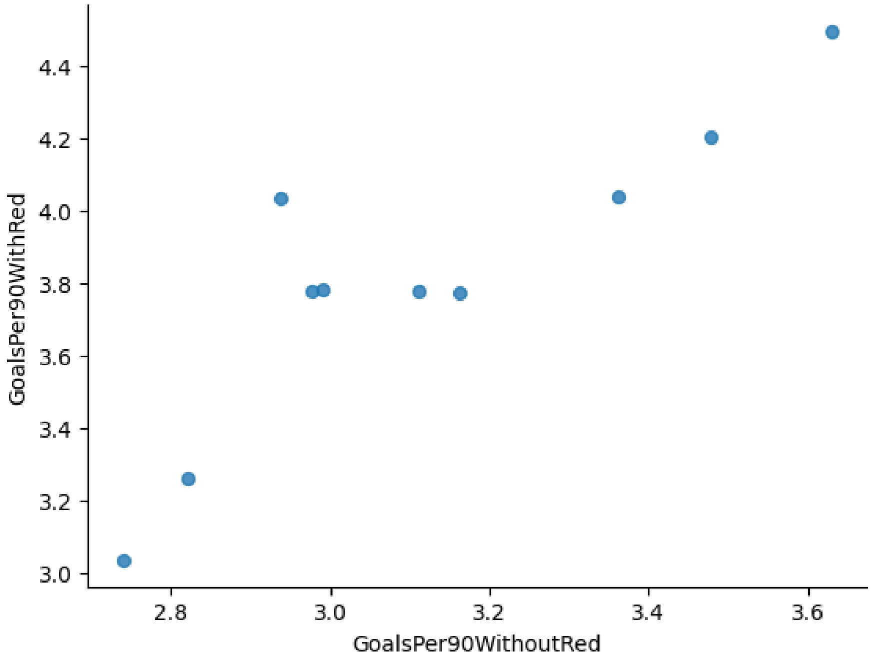
Results are summarised in the following table.

competition	GamesWithRedCard	TotalGames	% of overall games	GoalsPer90WithoutRed	GoalsPer90WithRed	GoalsPer90	percIncrease
1. Bundesliga, Germany	184	1530	7.511046	3.629659	4.494055	3.690756	23.814799
2. Bundesliga, Germany	264	1530	7.511046	3.478970	4.203297	3.547899	20.820157
Championship, England	317	2759	13.544428	2.977791	3.776630	3.029048	26.826564
League One, England	341	2608	12.803142	2.990766	3.779736	3.049628	26.380190
League Two, England	378	2648	12.999509	2.821011	3.261963	2.857898	15.630996
Ligue 1, France	400	1799	8.831615	3.110656	3.779257	3.203208	21.493860
Ligue 2, France	409	1797	8.821797	2.739837	3.035114	2.780348	10.777175
Premier League, England	203	1900	9.327442	3.162021	3.773292	3.203459	19.331656
Primera, Spain	412	1900	9.327442	2.938445	4.031528	3.043759	37.199350
Serie A, Italy	365	1899	9.322533	3.361360	4.039143	3.421801	20.163956

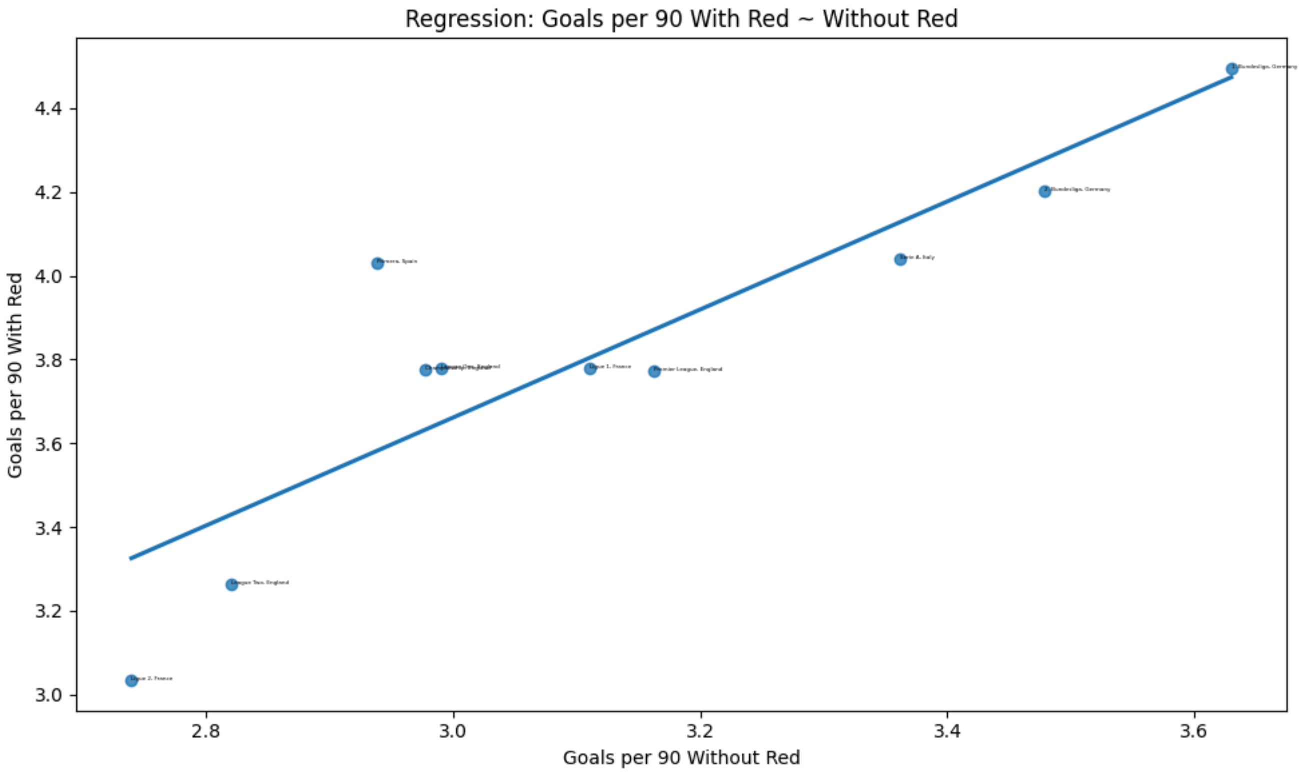
We visualise the increase in goal rate for each league from non-red minutes to red minutes:



We see Ligue 2, France has the lowest increase, while the Primera, Spain has the highest as the potentially only outlier. Most of the others lie in a similar range, suggesting the effects of a red card follow a general pattern that is roughly consistent throughout the leagues. This is further emphasised in the following scatter plot graphing the goal rate with a red versus goal rate with a red card for each league.



The points almost form a straight line if we were to exclude the Primera, Spain. We ran a regression model on this data for each league and were able to fit it to the graph.



The slope of the regression was approximately **1.29** which is the expected increase in average goals per 90 for minutes **with a red card** for every 1-unit increase in goals per 90 **without red.** This was calculated with a p-value of 0.001 which makes it significant as it's highly unlikely to have occurred by chance. We calculated the y-intercept as -0.2084 but this isn't significant given the associated p-value was 0.797, meaning it is a highly unreliable statistic.

We found a R-squared value of 0.769 meaning 76.9% of the variation in goals per 90 with reds is explained by our goals per 90 without reds metric. This is quite a high value that strongly suggests red cards increase scoring across all the European leagues.

Conclusions

Given the visualisations and the regression above, we can say that the effect of red cards across the different leagues is systemic and isn't just random. This is especially typified by our regression with its slope p-value of 0.001. It implies that red cards amplify the already existing goal rates in effect. The effect is multiplicative rather than a constant increase.

The strength of the regression analysis suggests it could be used as a predictive model in the future for similar leagues without all the same data. This could be useful for a management team looking to maximise their chances of success.

This analysis also shows we were correct to include all leagues in our previous analysis.

Limitations

- This analysis has largely the same limitations as the previous one as it is set up the same way, just with some segmentation in the data.

Analysis 5: Survival Analysis

Set up

A major limitation of the previous analysis in this report is that it assumes all minutes are uniform and that the effect of a red card on a game of football is the same in say the 30th minute as it is in the 75th. This is obviously not true, the changes in the dynamics of the game can be very different depending on the time of the card. Furthermore, previous analysis also did not account for uneven sample sizes; there were far more minutes without reds than with red in these. Our survival analysis will deal with both of these limitations.

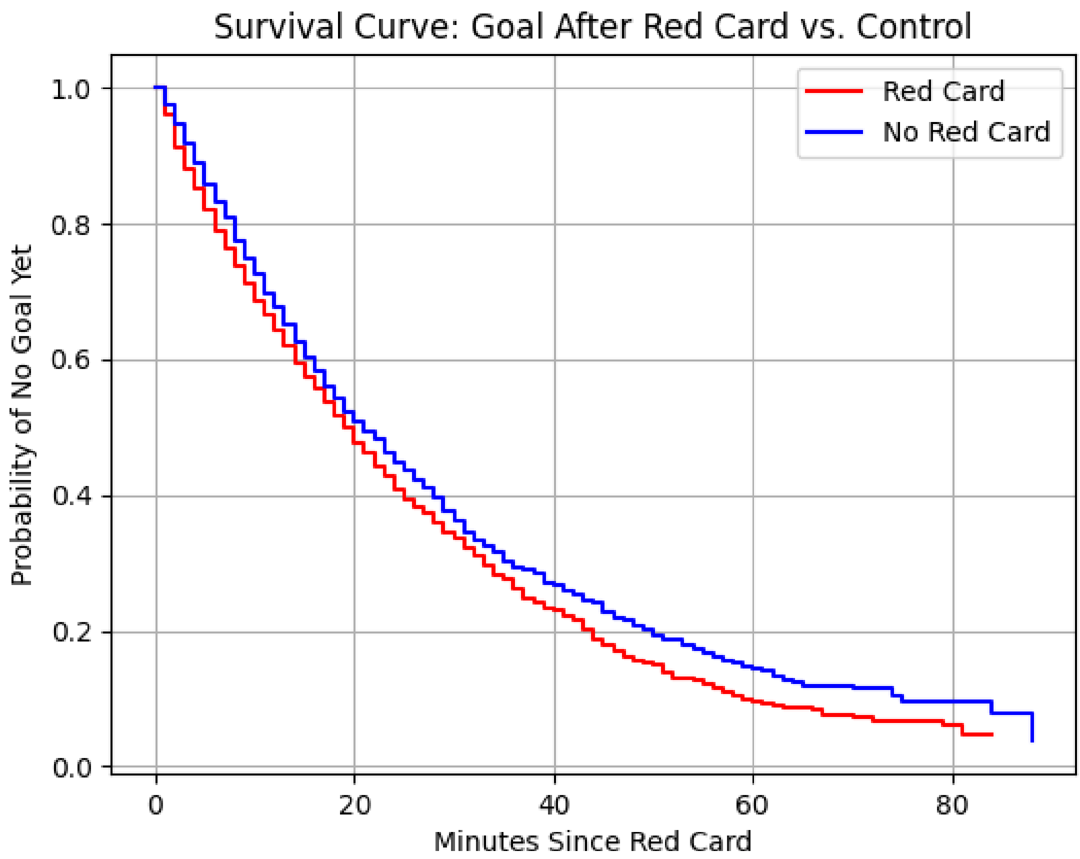
Survival analysis, or time-to-event analysis, is a branch of statistics where the time from a particular moment to an event occurring is our measured variable. It is commonly used in medicine for time to disease relapse or time to death analysis, but has a wide variety of uses in engineering and social sciences too. For our use case, the minutes elapsed from the red card to the first goal after it will be our measured variable. This analysis aims to answer the question of **whether goals happen more quickly after red cards than usual**.

We collate our data for this analysis as follows. We want two data from two groups: games with a red card and games without. The games without a red card will be our control group. Our measured variable is supposed to be the minute of a red card subtracted from the minute of the first goal after, but what do we do when there are no red cards? We decided to use the same times as in the red card games for a fair comparison. There were 3273 games with red cards, so we sampled 3273 games without replacement from the total of 17097 games without red cards and joined the actual red card times as our pseudo-red card times. This gave us the data we needed to calculate the duration time for all. The survival analysis method deals with situations where no goal is scored after the red card time so we leave these data points in. We call these censored observations. So after joining and manipulating the data in games.csv and events.csv to get what we described above, we end up with a dataframe with 3 columns: duration (time difference from red to goal), event (boolean for whether any goal was scored after the red), and group (whether this game had a red card or not).

We will use the Kaplan-Meier method for this analysis. It is suitable for comparing groups of interest to control groups where we don't have other predictor values or covariates, which is exactly our situation. It is also nonparametric, meaning we don't need an assumption of how data is distributed to use it. At each minute after a red, the Kaplan-Meier method estimates the probability of no goal having been scored based on the data it has been provided. In example, it calculates this probability as the fraction of games for each group where no goals have been scored. We censored our games without goals after reds at duration = 91 - red card time. Each of these are included in the calculations until this duration in the graph below.

Results

The following is the Kaplan-Meier survival curve generated on our data:



It is clear that the probability of no goal being scored after a red card decreases more quickly for games where a red occurred than games where they did not.

We can quantify this assertion using a log rank test, a common test used in survival analysis to compare two survival curves. It tests the null hypothesis that there is no difference in the time-to-goal distributions between the red-card and no-red-card groups. Utilising python a **chi-squared test-statistic** of **4.95** and a **p-value** of **0.03** were calculated. The test statistic from the log-rank test measures how different the two survival curves are. The test statistic of 4.95 indicates that the observed difference between the groups is larger than we'd expect by random chance. While it's not extremely large, it's enough to be considered statistically significant, with a p-value of 0.03 that means there is a 3% chance we'd calculate a difference this big by chance. This suggests a meaningful difference in how quickly goals are scored after red cards compared to similar non-red periods.

Conclusion

We can say within a 95% confidence interval that goals are scored more quickly than usual after red cards. This result was visualised above and is significant as it takes into account the timing of the red cards. This survival analysis anchors each case to the time of the red card, allowing us to model the likelihood of a goal occurring afterwards, which shows a definite increase in the chance of a goal happening afterwards.

In Summary

Let’s remind ourselves of the initial question:

“Based on the data in this dataset, do red cards seem to lead to more (total) goals?”

Our “Analysis 1” showed that goal rates were higher in games with at least one red card but could not conclusively show that more goals are scored on average in games with at least one red card. Our second analysis built further granularity by considering time on a minute to minute basis and tallying goals scored in red carded minutes as well as goals scored in non-red carded minutes. This changed the complexion of our picture quite a bit, as we found overwhelming evidence to accept the alternative hypothesis goals are more likely to be scored on average in minutes where at least one red card is in effect.

Further inquisition into the data aroused suspicions in this result however: it turned out far more goals are scored in general in the later stages of games, precisely the stages of games most of our red-carded minutes occurred. To level the playing field, in “Analysis 3” we only considered minutes after a certain cut off point of 55 minutes, deciding on this minute based on the distribution of minutes played with red cards. Although the margins did narrow from “Analysis 2”, we still found strong evidence to accept the alternative hypothesis that goals are more likely to be scored on average in minutes where at least one red card is in effect. We also found the difference in goal rates to increase with the cut off point. Why is excellent grounds for further study but is perhaps influenced by the fact that red card games have more injury time, thereby maybe more goals on average.

In “Analysis 4”, we compared the same data for different leagues to see if effects were systemic and we found that they were through our regression analysis. We found that red cards seem to amplify the already existing goal rates in effect. The effect is multiplicative rather than a constant increase. We also found no correlation between the standard of league and the effect of red cards on goal scoring rates, as shown by the 4 English Leagues.

Finally, we ran a survival analysis, applying a method more commonly used in medicine and engineering to football data. We found strong evidence to suggest goals are scored more quickly than usual after red cards.

Overall Conclusion

My overall conclusion is that we have sufficiently shown that red cards do increase total goals scored. Across multiple modes of analysis we showed that goal scoring rates increased in general and across the leagues. We strengthened this argument with our survival analysis by showing goals occurred more quickly after red cards, also accounting for the exact timing of the red cards.

More General Limitations and Revisions

1. **45/90 minute caps:** Led to huge spikes of goals at these minutes. Could have tried to mitigate this by adding a reasonable average injury time constant for each game and spreading events evenly throughout these added pseudo-injury time minutes
2. **Uneven sample sizes in earlier analysis:** In the first 4 modes of analysis there was an imbalance in sample sizes between the two groups which could have perhaps led to biases. Perhaps these sections could have been done differently to even sample sizes. We dealt with this issue in the survival analysis.
3. **Team/league strength or game state:** We didn't control for these aspects, which might yield more interesting results that could be leveraged by a football management team.

Further Research Possibilities

1. The regression model proved a strong estimator in predicting average goal scoring rates after red cards by league based on goal scoring rates without reds. Perhaps this could be used more locally intra games.
2. Further analysis into who usually does more scoring (the side with the red or the other team) would be interesting and I'm sure very useful to a management team looking to maximise their chances of victory.

Appendix

Analysis 3 results for different cut-off points:

Results:
Taking 55 minutes as the cut off:
- 649123 minutes were played without a red card in this time frame.
- 63827 minutes were played with at least one red card.
- 21741 goals were scored without a red card.
- 2653 goals were scored with at least one red card.
- 0.033492881934548614 goals per minute without a red card.
- 0.04156548169270058 goals per minute with a red card.
- 3.0143593741093753 goals per 90 without a red card.
- 3.7408933523430523 goals per 90 with at least one red card.
- 24.252491694352177 percentage increase

Taking 60 minutes as the cut off:
- 553103 minutes were played without a red card in this time frame.
- 57997 minutes were played with at least one red card.
- 18810 goals were scored without a red card.
- 2468 goals were scored with at least one red card.
- 0.03400813230085536 goals per minute without a red card.
- 0.04255392520302774 goals per minute with a red card.
- 3.060731907076982 goals per 90 without a red card.
- 3.8298532682724966 goals per 90 with at least one red card.
- 25.12867459633308 percentage increase

Taking 65 minutes as the cut off:
- Taking the cut off minute as: 65
- 458070 minutes were played without a red card in this time frame.
- 51180 minutes were played with at least one red card.
- 15990 goals were scored without a red card.
- 2259 goals were scored with at least one red card.
- 0.034907328574235376 goals per minute without a red card.
- 0.04413833528722157 goals per minute with a red card.
- 3.141659571681184 goals per 90 without a red card.
- 3.9724501758499415 goals per 90 with at least one red card.
- 26.444322983224435 percentage increase

Taking 70 minutes as the cut off:
- Taking the cut off minute as: 70
- 364074 minutes were played without a red card in this time frame.
- 43326 minutes were played with at least one red card.
- 13244 goals were scored without a red card.
- 2012 goals were scored with at least one red card.
- 0.036377220015711094 goals per minute without a red card.

- 0.046438628075520474 goals per minute with a red card.
- 3.2739498014139983 goals per 90 without a red card.
- 4.179476526796843 goals per 90 with at least one red card.
- 27.658540304794936 percentage increase

Taking 75 minutes as the cut off:

- Taking the cut off minute as: 75
- 271171 minutes were played without a red card in this time frame.
- 34379 minutes were played with at least one red card.
- 10562 goals were scored without a red card.
- 1753 goals were scored with at least one red card.
- 0.03894959269243393 goals per minute without a red card.
- 0.05099043020448529 goals per minute with a red card.
- 3.5054633423190533 goals per 90 without a red card.
- 4.589138718403676 goals per 90 with at least one red card.
- 30.913898399739455 percentage increase

Taking 80 minutes as the cut off:

- Taking the cut off minute as: 80
- 179445 minutes were played without a red card in this time frame.
- 24255 minutes were played with at least one red card.
- 7867 goals were scored without a red card.
- 1419 goals were scored with at least one red card.
- 0.043840731143247234 goals per minute without a red card.
- 0.058503401360544216 goals per minute with a red card.
- 3.945665802892251 goals per 90 without a red card.
- 5.26530612244898 goals per 90 with at least one red card.
- 33.44531406054224 percentage increase

Taking 85 minutes as the cut off:

- Taking the cut off minute as: 85
- 88981 minutes were played without a red card in this time frame.
- 12869 minutes were played with at least one red card.
- 5095 goals were scored without a red card.
- 1042 goals were scored with at least one red card.
- 0.057259414931277466 goals per minute without a red card.
- 0.08096977232108167 goals per minute with a red card.
- 5.153347343814972 goals per 90 without a red card.
- 7.287279508897351 goals per 90 with at least one red card.
- 41.408661646755014 percentage increase

Taking 89 minutes as the cut off:

- Taking the cut off minute as: 89
- 17668 minutes were played without a red card in this time frame.
- 2702 minutes were played with at least one red card.
- 3099 goals were scored without a red card.
- 745 goals were scored with at least one red card.
- 0.17540185646366313 goals per minute without a red card.

- 0.2757216876387861 goals per minute with a red card.
- 15.786167081729682 goals per 90 without a red card.
- 24.814951887490746 goals per 90 with at least one red card.
- 57.19428129080581 percentage increase

References

1. Lifelines: Survival Analysis in Python – https://lifelines.readthedocs.io/en/latest/
2. Scipy Stats: t-tests and Proportions – https://docs.scipy.org/doc/scipy/reference/stats.html
3. Statsmodels: Generalized Linear Models (Poisson Regression) – https://www.statsmodels.org/stable/glm.html
4. Kaplan-Meier Estimator – https://en.wikipedia.org/wiki/Kaplan–Meier_estimator
5. Log-Rank Test – https://en.wikipedia.org/wiki/Log-rank_test
6. Proportions Z-Test – https://www.statsmodels.org/dev/generated/statsmodels.stats.proportion.proportions_ztest.html