

# You Miss 100% of The Shots You Don't Take

An Analysis of Soccer Expected Goals

Meredith Brown

December 8, 2021

## 1 Introduction

Sports can be extremely frustrating. When the game is decided by only a few runs, goals, or baskets, the better team often loses. Even when the better team does win, it can be difficult to discern the winning's team dominance with a simple scoreline.

In this vein, in 2003, Michael Lewis released his book, *Moneyball: The Art of Winning An Unfair Game*, which details how baseball statistics were outdated and didn't truly reflect the potential value of a player [Lewis, 2003]. Lewis' book highlighted a new strategy adopted by the Oakland Athletics (the As), by which the As sought out players undervalued by traditional metrics but valued highly by metrics like on-base percentage. In this way, the As leveraged one of the smallest player budgets into one of the top teams in Major League Baseball.

The "Moneyball philosophy" inspired the development of new metrics across sports. In the world of soccer, this became the concept of "expected goals." Invented by Matthew Benham, this is the idea that, "In a low-scoring sport that is skewed by randomness and luck, the quality and quantity of chances created during a match mattered more" [Mather, 2015]. Rather than focusing on wins and losses, Benham focused on players creating (or stopping) quality chances. In employing this philosophy for his club, Brentford FC, Benham raised the team from the fourth division to the English Premier League, the top flight for soccer in England [xGPhilosophy, 2021].

Previous uses and studies of expected goals (xG) have focused on teams' and players' expected goals for a game, across a season, or throughout a career [Maurath, 2020]. These studies aggregate across thousands of shots in hundreds of games, looking at expected goals across long periods of time. However, at its heart, expected goals is a measure of shot quality.

Thus, this study aims to examine the concept of expected goals at its core, assessing what factors make any given shot a likely goal. Due to the nature of how goals are scored, I hypothesize shots taken later in the game, in the attacking half near the goal, and with either of a player's feet will be more likely to result in a goal. Furthermore, this study will analyze whether there are differences in shot qualities over time and across the top leagues in Europe.

### 1.1 Data

The dataset used in this study comes from webscraping performed by Alin Secareanu [Secareanu, 2017]. It provides a granular view of 9,074 games, resulting in 941,009 events from the biggest 5 European football (soccer) leagues: England, France, Germany, Italy, and Spain from the 2011/2012 season to the 2016/2017 season (through January 25, 2017). The creator of the dataset stated this data has been previously used to "create predictive models for football games in order to bet on football outcomes, make visualizations about upcoming games, and build expected goals models and compare players" [Secareanu, 2017]. Thus,

this dataset is appropriate for use in this study, and the aim of this study has not yet been performed on this data.

Of the 941,009 events identified in these games, 229,135 were identified as “attempts” or shots. Of these, observations were removed on the basis of missing data such as the location from which the shot was taken, the final location of the ball, or the outcome of the shot. Further, as the dataset ceased collection halfway through the 2016/2017 season, shots from this season were removed to prevent irregularities when accounting for season as a predictor. The final dataset contains information on 202,577 unique shots.

In this analysis, the response variable is a binary representation of the outcome of the shot: goal or no goal. The data contains other potential response variables, such as the aforementioned final location of the ball and a more particular characterization of the outcome of the shot. However, these variables were determined to be too specific. As the primary aim of this study is to assess what factors make a shot successful, the nuances of where the shot ultimately failed were deemed unnecessary. Further, the binary response variable allows for ease of interpretability and easy application to real world scenarios.

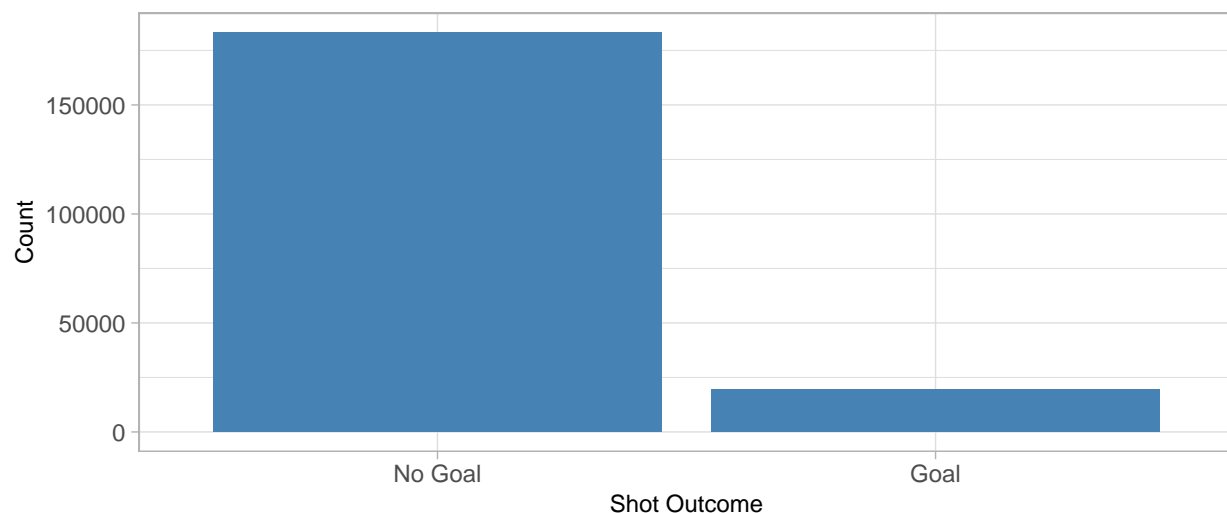
Predictors of interest are:

- Side: whether the player attempting the shot is on the home or away team
- Location: from where the shot was attempted
- Body part: whether the shot was attempted with a player’s left foot, right foot, or head
- Assist: a variable created to reflect whether the shot was assisted
- Time: the elapsed game time in minutes
- Situation: whether the shot occurred during open play, a corner kick, or a set piece
- Season: the soccer season during which the shot occurred
  - Note: Seasons are written in terms of the year in which the season, which typically lasts from August to May, ends
- Country: the country in which the league the shot occurred in is located

Time is a continuous variable, and season is a discrete ordered variable, while the remaining variables are categorical. The variation in type of predictor was thought to be important, as it allows for a diverse range of relationships to be explored.

## 1.2 Exploratory Data Analysis

**Figure 1 – Distribution of Shot Outcome**



In Figure 1, we observe that most of the shots taken did not result in goals - in fact, only 9.7% resulted in goals. This is within the range of my expectations, as in a typical season, the best teams can have a goal conversion ratio - the number of goals over the total number of shots - of just over 11% while the worst teams can have below 4% [Transfer Markt, 2021].

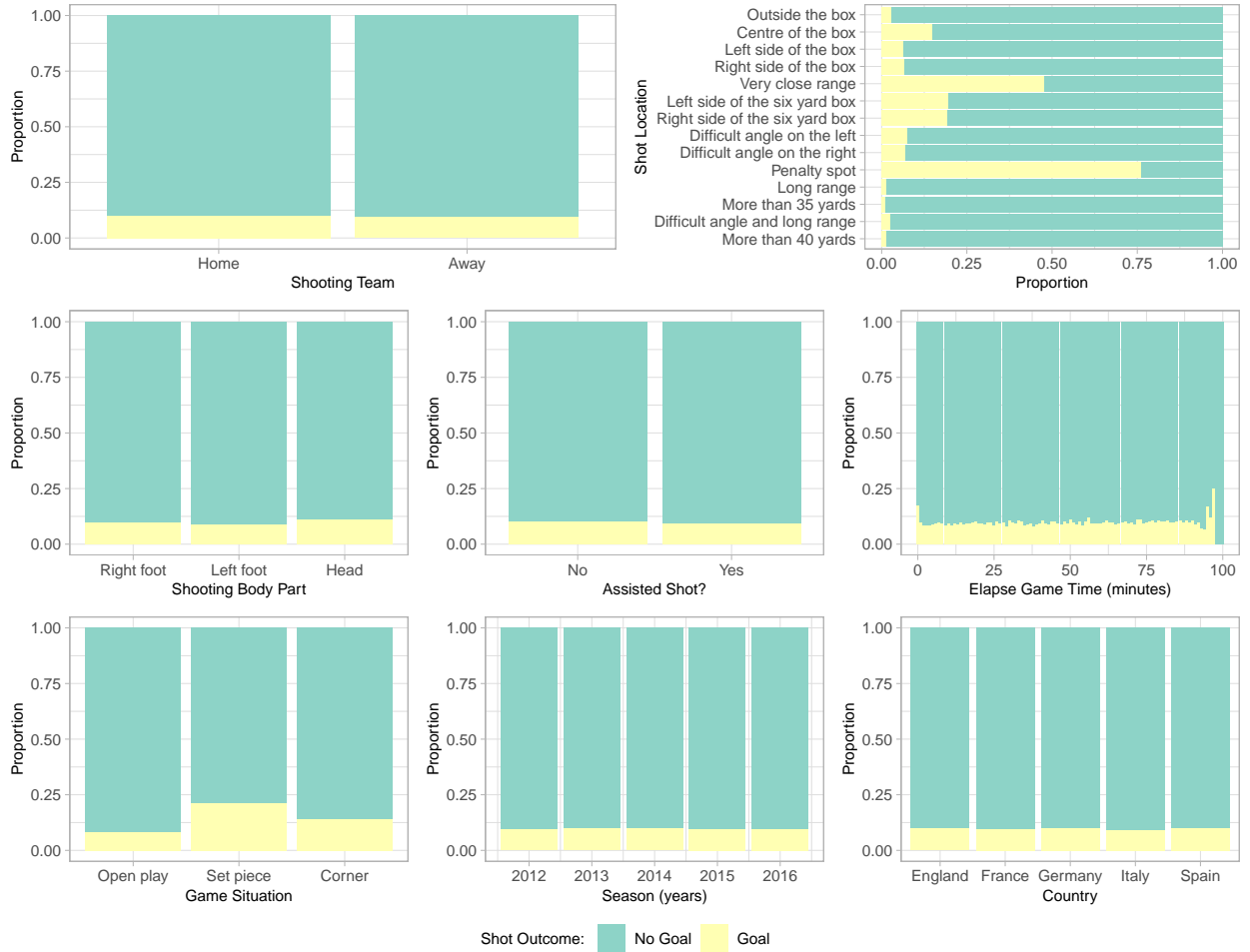
**Figure 2 – Count of Shot Outcome by Shot Characteristics**



Figure 2 visualizes the underlying trends in characteristics of shots, as well as the preliminary relationships between these characteristics and the outcome of interest: whether a goal was scored. For trends in the nature of a shot, there appear to be more shots attempted by home teams; most of the shots seem to be taken near the box, whether that be the left, right, or center of the box or just outside; most shots are attempted with players' right feet; most shots are assisted; more shots are attempted as the game goes on; most shots are attempted in the regular run of play; more shots are attempted in more recent season; and the most shots were attempted in the Italian Serie A, while the fewest were attempted in the English Premier League. Although Figure 2 aids in assessing these trends, it presents challenges in discerning the differences in the goal conversion ratio of the different levels of these categories.

Figure 3 aids in addressing that challenge. Figure 3 recreates the 8 visualizations of Figure 2, but the values are now presented in terms of proportions of each category, rather than the overall counts. Through this, Figure 3 serves to elucidate the relationship between shot characteristic and outcome. Here, the plots highlight there is little advantage to being the home or away team, to having an assist, to changing the season, or to switching countries. However, it does appear that shots attempted with players' heads, from set pieces, in the waning minutes of a game, or from within the 18-yard box increase the likelihood of the shot resulting in a goal. These visualizations provided compelling evidence for factors that may contribute to a shot's effectiveness. Let us now model these relationships to explore this further.

**Figure 3 – Proportion of Shot Outcome by Shot Characteristics**



## 2 Methodology

To assess whether characteristics of a shot affect its success, or its likelihood to become a goal, this study will utilize a logistic regression model to predict the log-odds a shot is a goal rather than not a goal. The dataset provides this binary outcome for each shot, as well as the team that attempted it, its location, the body part it was attempted with, and other characteristics of the shot. Thus, a logistic regression model will allow this study to explore the relationship between these different characteristics and the binary outcome of interest.

As mentioned previously, the binary outcome of goal or not goal for each shot was chosen over other outcome variables for its interpretability and ease of application. A multinomial logistic regression model would allow for modeling of the shot outcome stratifying by “On target”, “Off target”, “Blocked”, or “Hit the bar.” However, this information is a bit superfluous and introduces unnecessary complexity. It is more important to differentiate between shots that go in or don’t than shots that were blocked or hit the cross bar on the goal. Further, a multinomial logistic regression model could be used to model the final placement of the shot, placements such as “Bit too high”, “Bottom right corner”, “Misses to the right”, from which the four aforementioned categories aggregate. However, this provides even further unnecessary nuance. A potential weakness of this approach is its simplicity. These more complex categories could allow for more insight into shots that are “close” to being a goal. However, this is beyond the scope of this study, so a simple logistic regression model predicting the binary outcome will suffice.

Further, logistic regression is a suitable technique for exploring relationships between predictors that are both categorical and continuous, which is the case for this dataset. Another potential technique would be to perform Bayesian regression, where priors are set based on beliefs about the data and sampling is then performed from the posterior distribution. However, as this is a new field of study and there is little known about what makes a specific shot effective, there is not sufficient information to justify the use of or adequately establish priors for this analysis. A third technique could be Beta regression. Beta regression models probabilities on the open interval (0,1). Although it could be useful to explore what predictors increase the probability of a shot succeeding, the response variable of interest is a binary 0 and 1. Since Beta regression does not include these bounds, it again doesn't make sense to use this technique. Thus, logistic regression is concluded to be the best technique for modeling the data to address the question of interest.

Necessary assumptions for logistic regression include independence, sufficient sample size, and linearity. As each observation is a unique shot, it is reasonable to assume these are independent. Although shots by the same player or the same team may be related in some way, these factors are not considered in this model, so we can assume a level of independence. To further confirm a lack of collinearity, Appendix B features the variance inflation factors for this model. No value is concerning, so we can conclude independence to be met. The data used for this analysis contains over 200,000 observations, so there is certainly a sufficient sample size. Lastly, Appendix A includes visualizations of the binned residuals vs. predictors used in the model. The mean binned residuals for each level of each categorical predictor is close to 0, and there is no obvious deviation from linearity. Thus, we can conclude all assumptions for logistic regression are met.

With many categorical variables with myriad levels, the logistic regression model will take the following form:

$$\log\left(\frac{\mathbb{P}(isgoal_i=1)}{1-\mathbb{P}(isgoal_i=1)}\right) = \beta_0 + \beta_1\mathbb{I}(side_i = Away) + \beta_2\mathbb{I}(location_i = Difficult\ angle\ long\ range) + \beta_3\mathbb{I}(location_i = Difficult\ angle\ on\ the\ left) + \beta_4\mathbb{I}(location_i = Difficult\ angle\ on\ the\ right) + \beta_5\mathbb{I}(location_i = Left\ side\ of\ the\ box) + \beta_6\mathbb{I}(location_i = Left\ side\ of\ the\ six\ yard\ box) + \beta_7\mathbb{I}(location_i = Right\ side\ of\ the\ box) + \beta_8\mathbb{I}(location_i = Right\ side\ of\ the\ six\ yard\ box) + \beta_9\mathbb{I}(location_i = Very\ close\ range) + \beta_{10}\mathbb{I}(location_i = Penalty\ spot) + \beta_{11}\mathbb{I}(location_i = Outside\ the\ box) + \beta_{12}\mathbb{I}(location_i = Long\ range) + \beta_{13}\mathbb{I}(location_i = More\ than\ 35\ yards) + \beta_{14}\mathbb{I}(location_i = More\ than\ 40\ yards) + \beta_{15}\mathbb{I}(bodypart_i = Left\ foot) + \beta_{16}\mathbb{I}(bodypart_i = Head) + \beta_{17}\mathbb{I}(assist_i = Yes) + \beta_{18}time_i + \beta_{19}\mathbb{I}(gamesituation_i = Set\ piece) + \beta_{20}\mathbb{I}(gamesituation_i = Corner) + \beta_{21}season_i + \beta_{22}\mathbb{I}(country_i = France) + \beta_{23}\mathbb{I}(country_i = Germany) + \beta_{24}\mathbb{I}(country_i = Italy) + \beta_{25}\mathbb{I}(country_i = Spain)$$

Because of the length of this model, it was decided not to include interaction terms. Interaction terms provide further complexity. As the point of this analysis is to gain a baseline understanding of the characteristics associated with successful shots, it may be unnecessary to incorporate further complexity that may prevent ease of interpretability and application. Thus, interaction terms were left out of this model.

### 3 Results

For interpreting significance, this analysis will utilize a significance level of  $\alpha = 0.05$ . The model baseline is a shot from a player on the home team, taken from the center of the box, with the player's right foot, unassisted, at the beginning of the game (0th minute), during the normal run of play, during the 0th season, and in a game in the English Premier League.

Table 1: Logistic Regression Model Results

Term	Coefficient	Std. Error <sup>1</sup>	Lower CI <sup>2</sup>	Upper CI <sup>2</sup>	P-Value
Intercept	32.728	12.319	8.583	56.873	0.008
Away Team	-0.024	0.017	-0.057	0.008	0.148
Location = Difficult angle and long range	-2.201	0.170	-2.534	-1.869	<0.001
Location = Difficult angle on the left	-0.942	0.081	-1.099	-0.784	<0.001
Location = Difficult angle on the right	-1.087	0.085	-1.253	-0.921	<0.001
Location = Left side of the box	-1.185	0.034	-1.253	-1.118	<0.001
Location = Left side of the six yard box	0.326	0.048	0.232	0.419	<0.001
Location = Long range	-2.780	0.198	-3.168	-2.392	<0.001
Location = More than 35 yards	-3.073	0.231	-3.525	-2.621	<0.001
Location = More than 40 yards	-2.893	0.581	-4.031	-1.754	<0.001
Location = Outside the box	-2.127	0.025	-2.176	-2.077	<0.001
Location = Penalty spot	2.247	0.062	2.126	2.368	<0.001
Location = Right side of the box	-1.144	0.034	-1.211	-1.078	<0.001
Location = Right side of the six yard box	0.287	0.048	0.192	0.382	<0.001
Location = Very close range	1.669	0.030	1.611	1.728	<0.001
Body part = Left foot	-0.062	0.019	-0.100	-0.024	0.001
Body part = Head	-0.970	0.025	-1.019	-0.921	<0.001
Assist = Yes	0.122	0.021	0.080	0.164	<0.001
Time (minutes)	0.001	0.000	0.000	0.001	0.008
Game situation = Set piece	0.502	0.033	0.438	0.566	<0.001
Game situation = Corner	0.301	0.028	0.246	0.356	<0.001
Season (years)	-0.017	0.006	-0.029	-0.005	0.005
Country = France	0.012	0.029	-0.044	0.069	0.668
Country = Germany	0.054	0.030	-0.004	0.112	0.069
Country = Italy	-0.012	0.029	-0.068	0.044	0.683
Country = Spain	0.051	0.029	-0.005	0.107	0.076

<sup>1</sup> Std. Error = Standard Error<sup>2</sup> CI = 95% Confidence Interval Bounds

Holding all else constant, when a shot is taken by a player on the away team rather than a player on the home team, the odds of the shot being successful, i.e. of scoring a goal, are expected to multiply by a factor of 0.976 (decrease). This is consistent with our initial data analysis, as although it seems an away team player is less likely to score on any given shot, the odds only decrease by about 3.3%, which is not large.

The coefficient for each level of location is significant. Thus, each follows the following formula: Holding all else constant, when a shot is taken from location  $x$  rather than from the center of the box, the odds of scoring from that shot are expected to multiply by a factor of  $\beta_x$ , where  $x$  is the level and  $\beta_x$  is the associated coefficient. Based on the coefficients, the odds increase when the shot is taken from the left side of the six yard box, the right side of the six yard box, penalty spot, or from very close range. Intuitively, this makes sense and aligns with this study's hypothesis. Taking a shot from the center of the box provides the most room for error when shooting from middle distance. Thus, shooting from any closer should increase your odds, even if to the side, whereas any further should decrease your odds.

Holding all else constant, when a shot is taken with a player's left foot or head rather than a player's right foot, the odds of the shot being successful are expected to multiply by a factor of 0.940 and 0.379, respectively. The first makes sense, as players that are professionals should have relatively similar skill with both feet. However, the second factor diverges from our expectations. In the exploratory data analysis, it was shown that a higher proportion of shots taken with a player's head were goals than those taken with a player's right foot. Thus, this may be a result of the difference in the sample sizes, as highlighted in Figure 2.

Holding all else constant, when a shot is assisted rather than unassisted, the odds of the shot resulting in a goal are expected to multiply by a factor of 1.13 (increase). This again somewhat supports our hypothesis from the initial data analysis, as it appeared there was little difference between a shot being assisted or unassisted. However, again there was a much larger sample size of shots that were assisted than unassisted, so this could influence the model.

Holding all else constant, for each additional minute elapsed in the game, the odds of a shot resulting in a goal are expected to multiply by a factor of 1.001 (increase). Thus, although this is a statistically significant coefficient, this predictor does not have much impact on a change in the odds, which lines up with our hypothesis.

Holding all else constant, when a shot is attempted from a set piece or corner kick rather than during the run of play, the odds of a shot being successful are expected to multiply by a factor of 1.65 and 1.35, respectively. These factors make sense intuitively, as set pieces and corners are typically near the goal, can be practiced exactly, and are entirely focused on scoring, whereas run of play shots can be from anywhere, cannot be set-up for practice, and can sometimes be taken as last-ditch efforts.

Holding all else constant, for each additional season, the odds of a shot being a goal are expected to multiply by a factor of 0.983 (decrease). In reviewing Figure 3, there appears to be little to no difference in the goal conversion ratio of each season, so it makes sense that the odds would only decrease about 1.7% each season.

Finally, none of the coefficients for country of the league in which the shot was taken were deemed statistically significant at the  $\alpha = 0.05$  level. This makes sense, as these countries have the top 5 leagues in the world, each of which employs the best soccer players in the world. Thus, it follows that there would be no significant difference in the odds of a shot being a goal between the different leagues.

## 4 Discussion

At the outset, this study hypothesized little change in the odds of a shot being successful from being the home or away team, having an assist, changing the season, or switching countries. Conversely, it was hypothesized that shots attempted with players' heads, from set pieces, in the waning minutes of a game, or from within the 18-yard box increase the likelihood of the shot resulting in a goal.

The results of the model both confirmed and contradicted these hypotheses. The model found a statistically significant association between success and shots taken from set pieces, in the waning minutes of a game, and from within the 18-yard box, with shots taken closer to the goal having even higher odds than those taken from the center of the box. These results match what any latent soccer viewer can observe. Set pieces are designed for scoring goals. Teams often push offensively near the end of the game in hopes of securing victory. Shots taken closer to the goal have less room for error.

However, the model contradicted this study's hypotheses on the relationship between success and the presence of an assist or the body part used to shoot. Although assist appeared to have negligible impact on a shot success's from the exploratory data analysis, it was found to increase the odds by approximately 13%. This could be due to the fact that teams with more possession of the ball and better passing may be better overall, and, thus, their shots end up succeeding at a higher rate. Similarly, it might be that the final pass to the shooter moves the defensive unit just enough to increase the odds of a goal. On the other hand, it was found that attempting a shot with the head decreases the odds of success relative to attempting a shot with the right foot. One reason for this could be the discrepancy in sample sizing. There were over 3 times more shots attempted with the right foot than with the head. Thus, with right foot as the baseline, a header could seem a more unlikely way to score. In future analysis, it would be useful to explore a binary variable of header and non-headers to examine this relationship to a greater extent.

Interestingly, the model confirmed the negligible effects of season and country on the odds of success of a shot. This can be rationalized by the knowledge that the best players in the world are not concentrated in one country or league, and they move around from season to season as new players are signed. So, it follows

that no one nation's league would be supremely better, nor would one particular season have higher odds of goals.

Overall, this analysis is limited by the discrepancy in sample size between the two outcomes. As noted in the assessment of the binned residuals in Appendix A, over 90% of the data reflected shots not resulting in goals. Thus, the model could easily overfit or train itself to reflect a limited sample space. To assess this, in Appendix C, the model was used to predict the success values of each shot in the dataset. Here, the model was found to be 91.13% accurate, which, although impressive, is not much better than labeling every shot in the dataset as a failure. Furthermore, in plotting the ROC curve, the model was found to have an AUC of 0.7833. This confirms that although the model is better than guessing at random, it is nowhere near perfect. If this analysis were performed again, it would implement random undersampling and oversampling from the two outcomes to develop a more balanced dataset. This would address the class imbalance and would hopefully train a more robust model.

To improve upon the current model, future work could incorporate interaction effects, such as between side and time to examine if a home team gains an advantage near the end of a game or between assist and body part to examine if there's a relationship between teamwork and shot type. Likewise, future analysis could implement a multinomial logistic regression, utilizing one of the variables plotted in Appendix D to gain more insight into where shots are successful and where they are failing. Such a model could be particularly useful for soccer clubs looking to develop players, as it could allow coaches to quantify where their players score and miss the most and then teach ways to improve.

Future improvements to this analysis could also include additional webscraping to gather background information on the clubs and players attempting each shot. It most certainly would change a shot's likelihood of success were Lionel Messi shooting rather than a random Duke student. This knowledge would be useful in developing a better and more thorough model. In this vein, it could be helpful to use this model to develop prior beliefs for future Bayesian analysis. With these baseline results, future work could apply Bayesian analysis to more recent seasons' data, even potentially progressing to a model that can update each week after the newest games.

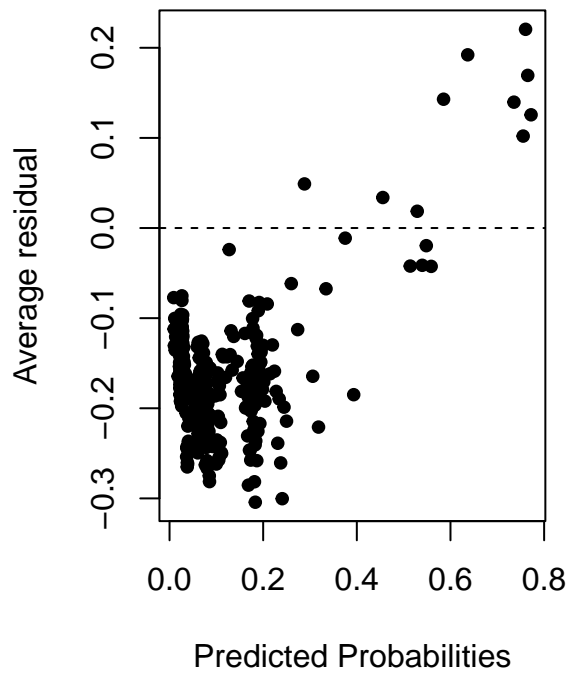
Overall, it was found that right-footed, assisted shots taken near the goal, in the crunch time of a game, and from a set piece or corner are the most likely to result in a goal, no matter the season or league.



## 5 Appendix

### 5.1 Appendix A: Diagnostic Plots

**Binned Residual vs. Predicted Value**



**Binned Residuals vs. Time**

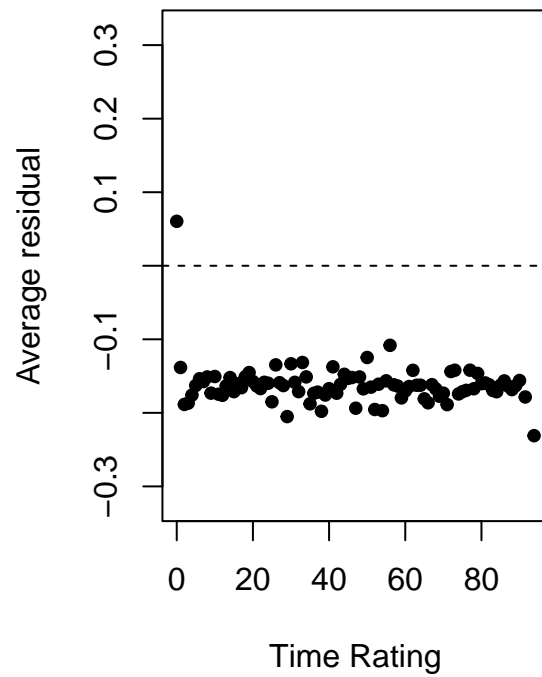


Table 2: Average Residuals per Team

Shooting Team	Average Residual
Home	-0.1632279
Away	-0.1632322

Table 3: Average Residuals per Location

Location	Average Residual
Centre of the box	-0.1872624
Difficult angle and long range	-0.1519554
Difficult angle on the left	-0.1929266
Difficult angle on the right	-0.1910551
Left side of the box	-0.1888369
Left side of the six yard box	-0.1705985
Long range	-0.1247867
More than 35 yards	-0.1117739
More than 40 yards	-0.1201506
Outside the box	-0.1537174
Penalty spot	0.1576347
Right side of the box	-0.1904972
Right side of the six yard box	-0.1727094
Very close range	-0.0143838

Table 4: Average Residuals for Body Part

Body Part	Average Residual
Right foot	-0.1575454
Left foot	-0.1609761
Head	-0.1870872

Table 5: Average Residuals for Assist

Assist	Average Residual
No	-0.1477784
Yes	-0.1687315

Table 6: Average Residuals per Situation

Situation	Average Residual
Open play	-0.1664142
Set piece	-0.1210487
Corner	-0.1675252

Table 7: Average Residuals per Season

Season	Average Residual
2012	-0.1693264
2013	-0.1582379
2014	-0.1575900
2015	-0.1666057
2016	-0.1647940

Table 8: Average Residuals per Country

Country	Average Residual
England	-0.1624496
France	-0.1636188
Germany	-0.1644571
Italy	-0.1617129
Spain	-0.1639660

These plots and tables serve to address the linearity assumption necessary for logistic regression. Each of the tables includes the mean binned residual value for the levels of the given categorical variable. All values are within 0.2 of 0, although all of the mean binned residual values are below 0. This presents some cause for concern, as these values should be scattered around 0. However, the binned residuals versus the predicted values is relatively scattered around 0, despite a slight linear pattern in the residuals. However, these trends could be due to the fact that a large portion of the dataset has a response value of 0, so the model may be overfitting to this value. Thus, we can tentatively conclude that the linearity assumption is met and proceed with caution.

## 5.2 Appendix B: Variance Inflation Factors

Table 9: Variance Inflation Factors

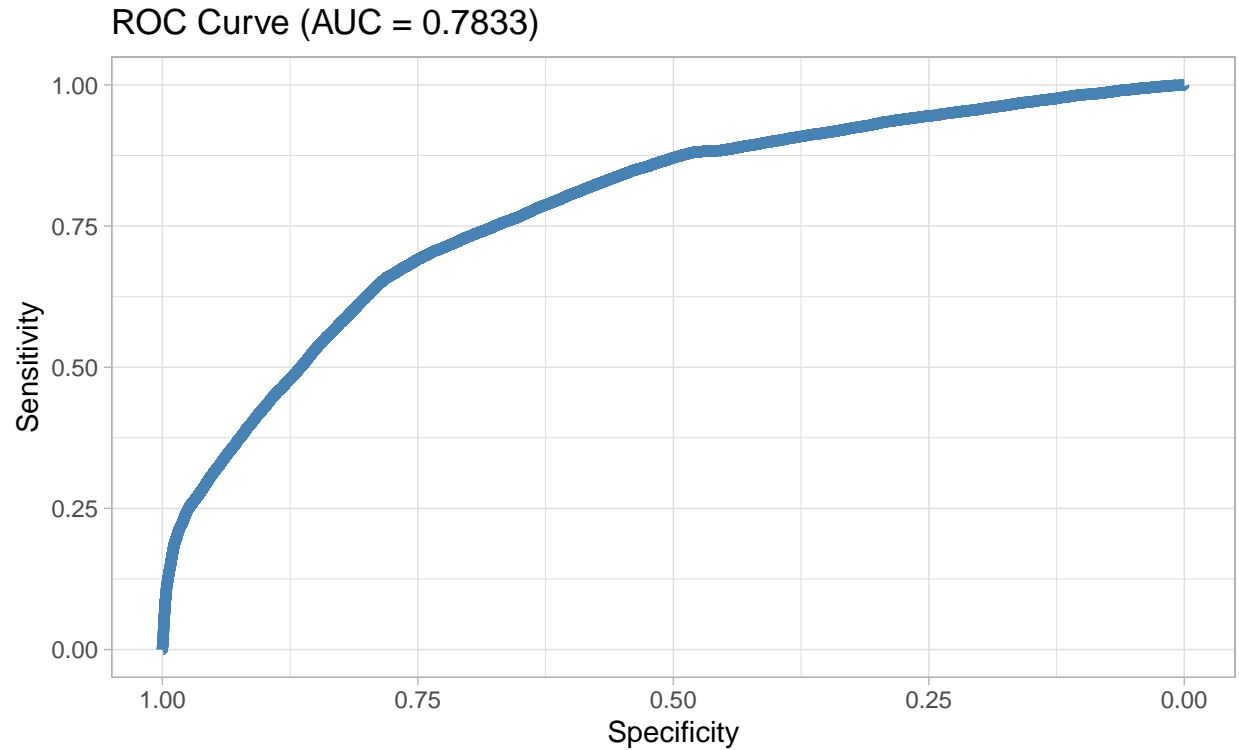
Variable	VIF
Side	1.001370
Location	1.734518
Body Part	1.392134
Assist	1.195823
Time	1.001392
Situation	1.602075
Season	1.076043
Country	1.080395

We confirm that no variance inflation factor is above 2, let alone above 10. So, we verify there is no concern for multicollinearity among the predictors used in the model.

### 5.3 Appendix C: Prediction

Table 10: Confusion Matrix for Logistic Model Predictions

	No Goal	Goal
No Goal	181129	1981
Goal	15987	3480

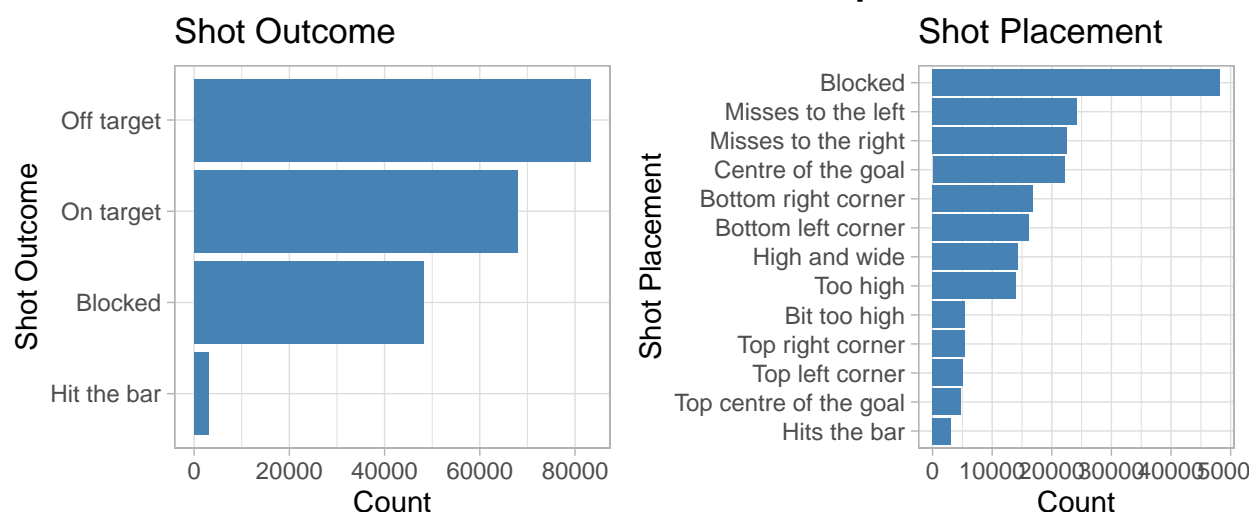


With a prediction threshold of 0.5, the model correctly predicts whether a shot is a goal or not 91.13% of the time. While this is great accuracy, it is worth reflecting on the true rate of goals in the dataset, which is approximately 9.6%.

The AUC of the model is 0.7833. This is better than random guessing, which would reflect an AUC of 0.5, but it not as close to 1 as would be desired. This can likewise reflect some of the failures of the sample size of the two outcome groups.

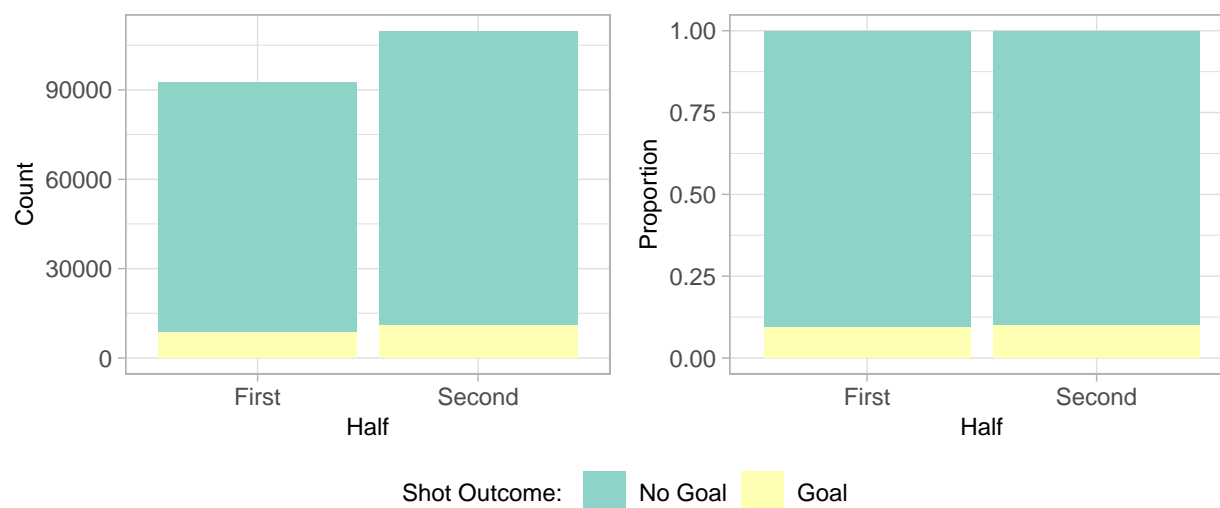
## 5.4 Appendix D: Other Exploratory Data Analysis

### Shot Outcome: More and More Specific



These plots represent two levels of specificity within the binary response variable representing goal or no goal. The left plot breaks down those shots that are unsuccessful into “Off target”, “Blocked”, and “Hit the bar”. The right plot includes these divisions, but also further breaks down “On target” and “Off target” into specific parts of the goal that are found or on which side the goal was missed, respectively.

### Shot Outcome by Half of Game



These plots represent the count and proportion of shot outcome in each half. Thus, this is an aggregation of the elapsed time variable into a categorical variable reflecting in which half of the game the shot was attempted. Here, we see that more shots are attempted in the second half, but about roughly the same proportion of shots is successful in each half.

## 6 References

- Lewis, Michael. 2003. *Moneyball: The Art of Winning an Unfair Game*. New York, W.W. Norton & Company.
- Mather, Victor. 2015. “Advanced Soccer Statistic Shows Better Team Doesn’t Always Win.” *The New York Times*. <https://www.nytimes.com/2015/07/04/sports/soccer/advanced-soccer-statistic-shows-better-team-doesnt-always-win.html>
- Maurath, Lars. 2020. “xG Model - Design and Implementation with R Tidymodels.” <https://www.thesignificantgame.com/portfolio/expected-goals-model-with-tidymodels/>
- Secareanu, Alin. 2017. “Football Events.” *Kaggle*. <https://www.kaggle.com/secareanualin/football-events/version/1?select=events.csv>
- Transfer Markt. 2021. “Premier League Conversion Rate.” <https://www.transfermarkt.us/premier-league/chancenverwertung/wettbewerb/GB1>
- @xGPhilosophy (The xG Philosophy) et al. 2021. “Brentford’s use of xG to master the transfer market.” *Twitter*. <https://twitter.com/xgphilosophy/status/1398232614201446401?lang=en>