Predicting NFL Rushing Yards

Applied Data Analysis Project

Logan Chalifour, Nathan Brouwers & Riley Demanche

MA 252 - 001

Bentley University

Professor Reagan Mozer

BIG DATA BOWL

Contents

Executive Summary

There are many elements that determine the success of a run play in a professional league that represents the world’s biggest, fastest, and strongest athletes. Many of these factors are outside of the control of the ball-carrier themselves, relying rather on coaching decisions, schemes, opposing defenses, and random variability through blown assignments.

This report will focus on the relationship between yards gained on a running play and the variables that are indicative in predicting those yards. Using data acquired from Next Gen Stats of 2017 and 2018 seasons in the NFL, the goal of this report is to propose a model that accurately predicts how many yards an NFL player will gain after receiving a handoff. Through various forms of data analysis, we determine which of these factors can help coaches and front office executives answer the question of “What makes for a ‘good’ run play?”

The model includes variables such as speed, acceleration, defenders in the box, down, and distance to predict yards gained. It will then be described how this model can be utilized and applied to help franchises improve their team and prepare for opponents. Additionally, the model will be tested for the assumptions of regression and a practical example will be provided. The details of these variables and findings are explored throughout this report.

Section 1 - Introduction

The National Football League is a multi-billionaire dollar industry that emphasizes winning more than anything. Due to the complexity and highly competitive outlook at the professional level, there are a plethora of ways that teams try to gain an edge. For teams to have success it is essential to be systematically sound across the organization. Due to the variation in player positioning, roles, and schematic game planning each week, more time and resources have been spent on developing statistical models to maximize player efficiency and value. The single most important objective is to outscore the opponent in any way possible. For the offense, this entails developing a game plan to sustain long-lasting drives that rack up yardage in the process, while the defense has the reverse goal to prevent scoring and long yardage plays.

In the NFL, nearly a third of the yards gained come from rushing plays. Over the past decade, the league has undoubtedly transitioned to a heavier aerial attack league with high octane offenses that rely on the pass more than the run. Nonetheless, having a successful running game with valuable ball carriers remains a staple for winning games in the NFL.

These ball carriers are paid millions of dollars each year, thus representing a significant investment to any franchise. Before paying out these lucrative salaries, NFL executives certainly want to make sure they are signing a player who will be a successful ball carrier that can be relied on to rack up a lot of rushing yards in order to help move the ball and score points. But what exactly makes for a successful ball carrier and more specifically a successful run play?

Just recently, the NFL extended this problem to the public through the NFL Big Data Bowl. Hosted on Kaggle, the competition utilizes NFL Next Gen Stats for all the rushing plays in the 2017 and 2018 seasons (see Appendix A to visit competition homepage). The goal of the competition is to construct a predictive model for yards gained on any particular rush play based on the data provided before the handoff. By analyzing nearly fifty different variables for over 500 thousand observations, anyone from reputable data scientists to college students are encouraged to generate a model that could be utilized by some of the most prestigious NFL teams.

In this report, we will attempt to solve the daunting task of predicting rush yards for any particular play by constructing a regression model with wide-scale practicality. Additionally, we will evaluate some relationships of interest between some potential explanatory variables and our outcome of yards:

- Down & Distance
- Number of defenders in the box
- Position of the ball carrier
- Speed & Acceleration
- Game Script (i.e. score spread)

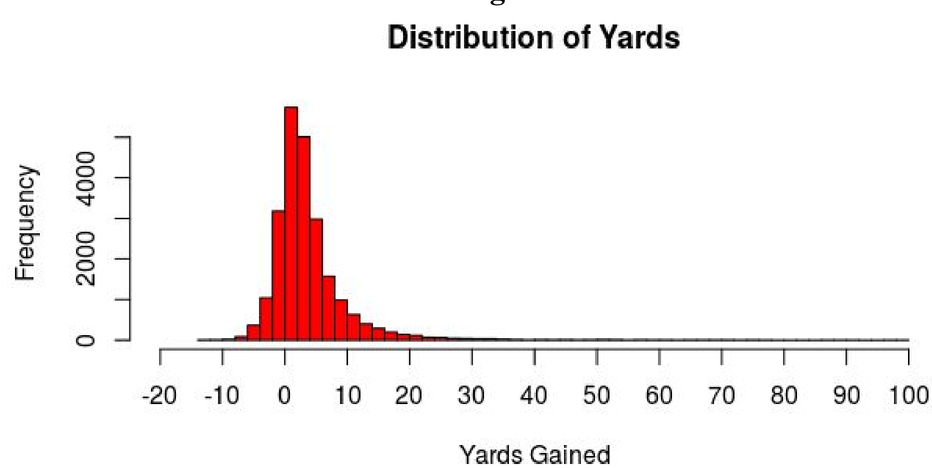
In Section 2, we present and explain the variables used for the analysis, and patterns observed in the data itself. Within Section 3, the most effective model that was constructed will be proposed and interpreted. Section 4 focuses on the evaluation of the model for its usability while also testing for regression assumptions and providing a practical example. To conclude these findings and results, Section 5 contains a summary that ties together any loose ends but also reiterates the practicality of the problem. For more information on the specific data mining process for selected topics, please reference the appendices in Section 6.

Section 2 - Data Characteristics

As previously mentioned, all the data used in this report was provided by the NFL Big Data Bowl on Kaggle which utilizes NFL Next Gen Stats from the 2017 and 2018 seasons. Powered by Amazon Web Services, NFL Next Gen Stats "captures real-time location data, speed, and acceleration for every player, every play on every inch of the field." The dataset from the NFL Big Data Bowl that we are utilizing in this report contains over 500 thousand of such observations for every run play, with information for every player on the field. For each player, we are provided the variables known at the time of handoff in order to forecast the rushing yards gained on that play. Since our dataset provided much extraneous information, we first decided to eliminate all the players from the dataset who were not the ball carrier. This brought the sample size down from over 500 thousand to around 23 thousand.

Before proceeding, it is important to get an idea of the rushing yards we are trying to predict with our model. In general, yards are measured as a discrete variable, although in reality, it is entirely possible to get half of a yard or any other fraction of a yard. As seen in Figure 1 below, the majority of rushes resulted in yardage between 0 and 5 yards with a mode of 2. It is evident that the distribution has a slight right skew with a median of 3 which is just lower than the mean of 4.21. Because of the slight right skew, we explored the practicality of a logarithmic transformation of the data, however, the results did not help normalize the distribution as we had hoped (see Appendix B for logarithmic distribution of yards). The transformation also added an extra level of complexity for interpreting results, so ultimately we decided to proceed with the untransformed data seen below since the distribution was mostly normal.

Figure 1

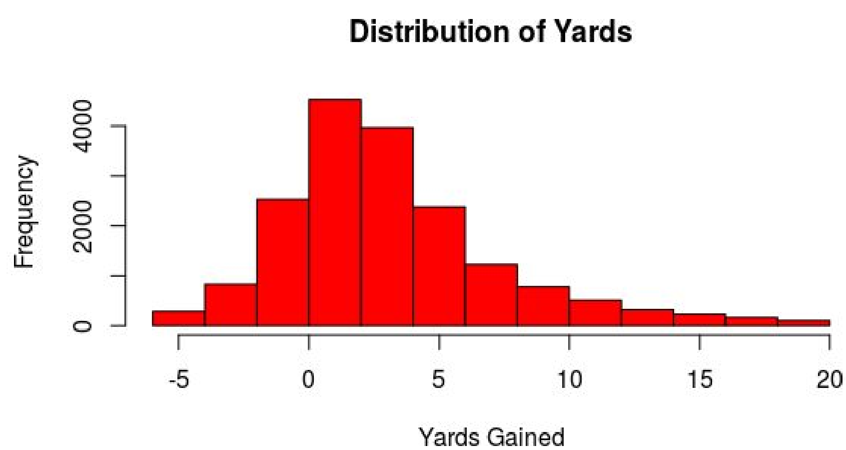


Another interesting characteristic of the data was the wide range of yardage. Within our dataset, we saw losses as extreme as 14 yards on an end-around by Josh Doctson. On the flip side, we saw Derrick Henry gallop the entire length of the field for 99 yards. Because of the unpredictable nature of such plays, you never know when a ball carrier will rip off a long run, and you never know when they will get tackled for a significant loss. This introduced a new problem to our dataset that we had to account for. Less than 3% of rushes go for more than 20 yards (which the NFL classifies as a "long-run") and similarly less than 1% of runs go for less than -5 yards (extreme loss). This means that approximately 96% of rushes result in yardage between -5 and 20 yards. Since the data is so heavily concentrated around the mean of 4.21, these outliers became hard to predict with an overarching model. To build the best predictive model moving forward, we decided to subset the data to eliminate these outliers, only including plays where the resulting yardage was between -5 and 20. This obviously poses some limitations to our model that will be discussed in later sections.

Up to this point, we have not introduced the variables that will be utilized to predict rushing yards. The dataset included forty-eight raw variables in addition to yardage (see Appendix C for a list of all the variables provided in the dataset along with their descriptions). Of these variables, there are many that intuitively we believed impacted yardage (i.e. defenders in the box); these relationships will be examined in further detail in Section 3. Others, however, naturally had no predictive value (i.e. Game ID). In addition to the variables provided, we constructed new variables such as score spread (the difference between ball carrier score and opponent score) as well as the age of the ball carrier.

Some columns also had null values which proved problematic in model construction. To account for this, we eliminated all rushes where a value was null. Ultimately, this helped cleanse the dataset even further to a total of 17,882 observations. The final subsetted distribution of yards can be seen below in Figure 2 with the same general pattern as Figure 1. Ranging from -5 to 20 yards, the data still has a slight right skew but is mostly normal.

Figure 2



Lastly, we implemented a basic test-train split with 80% of the data being used for model construction and 20% of the data being used to test that model. This 20% was excluded from the data used to construct the model and was revisited later to evaluate the accuracy of our model. Ultimately, this cross-validation method ensured that we were not overfitting for our data by comparing the accuracy of the model across the test and train data separately (see Appendix D for a comparison of test and train results for our best model derived later).

Section 3 - Model Selection and Interpretation

In this section, the process used to create and select the regression model will be introduced as well as its interpretation in context. We will evaluate and include the variables that impact yardage gained on any given rushing play. A proper model will be chosen on this basis:

- The model must predict rushing yardage for a play based on inputs prior to the handoff
- Only the variables deemed individually significant will be included in our best model
- The model must consider the interaction between variables when appropriate
- Our predictions will minimize error, however, random variability from human elements, randomness, and other immeasurable factors may hinder our ability to do so
- The model must have practical usability for coaches and front offices with a limited number of input variables

To get a general sense of the variables in our dataset and how useful they were in predicting rushing yards, we first constructed a full model that considered nearly all fifty variables provided in the dataset and the ones constructed. This full model provided us with a baseline of how much variation in yardage could be explained by including all of the variables together in one model. With a multiple R^2 of .1348 and an adjusted R^2 of .0928, this indicates that even a lengthy model that considers all the variables provided does not even account for 15% of the variation in yardage (see Appendix E for more information regarding the full model).

Next, our group was tasked with discovering which explanatory variables were individually significant, as well as jointly significant in predicting rushing yards. This was done by manually examining the relationship of each of the variables with respect to rushing yardage. All but one of the variables initially highlighted as potential predictors of yardage were deemed significant. As seen in the table below, each of the hypothesized predictors except for the score spread was significant in predicting yardage. It was also found that other variables such as yard line and player weight were significant. Other categorical variables such as down, quarter, and position were also significant for most classifications.

Explanatory Variable	Individual p-value	Significant at $\alpha = .05$?
Speed	$2*10^{-16}$	Yes
Acceleration	$2*10^{-16}$	Yes
Yard Line	$2*10^{-16}$	Yes
Defenders in the Box	$2*10^{-16}$	Yes
Distance	$2*10^{-16}$	Yes
Player Weight	$7.6*10^{-8}$	Yes
Score Spread	.7939	No

While many of these variables were individually significant, they were not always significant when combined in a regression model. Therefore, we knew it was necessary to consider the interaction between some variables. With our football knowledge, we were able to consider some interaction between variables and saw positive results as this enhanced the R^2 value while also reducing the residual standard error.

Furthermore, our group performed stepwise regression through backward elimination to reduce the Akaike's Information Criterion (AIC) and find a better predictive model. This stepwise regression technique starts with the full model and works backward by eliminating variables sequentially to minimize AIC. Using this technique we were able to eliminate insignificant variables, but also gathered a better understanding for which variables had meaningful interactions. After performing stepwise regression, many of the original variables that were considered proved significant in the overall model. The regression equation of the proposed model for predicting rushing yards on a given play is shown below (see Appendix F for the specific coefficients for this model).

Predicted Yards = (Speed) * (Acceleration) +
(Yard Line) * (Defenders In The Box) * (Down) * (Distance) +
(Position) * (Player Weight)

This model highlighted several of our hypotheses of interest and allowed us to begin answering the question of "What makes for a 'good' run play?" Some of the most significant predictors included the number of defenders in the box, speed, acceleration, and distance for gaining a first down.

Speed, as well as the acceleration of the ball carrier, are two of the most intriguing variables that are strongly correlated, thus we considered the interaction between them. Typically players with higher speed have higher accelerations and vice versa. This comes as no surprise though; teams have known since football's inception that faster and quicker ball carriers are more likely to be successful on the ground. Both of these showed to be individually significant and jointly significant and were thus incorporated into the model with interaction.

Our group considered a four-way interaction between yard line, defenders in the box, down, and distance. Ball carriers can be expected to rip off longer runs near midfield on earlier downs with minimal defenders in the box and long distances to go. Each was deemed individually significant and with our football knowledge, we found it appropriate to include interaction terms between all four of them. Each of these variables tends to impact each other based off of the game scenario and play design.

The last prong of our model is the ball carrier position and weight. Both of these variables were deemed individually and jointly significant. We believe these variables interacted since a ball carrier's position can be expected to impact their weight. It was found that lighter players at non-running back positions such as wide receiver usually fair better on the ground. This is likely because these players run less often so the defense is less prepared for it.

This process allowed us to eliminate certain variables that may have been individually significant but proved less relevant in joint significance for predicting yards within the model. By removing some of these less significant variables, it allowed for a more simplistic and practical model that can be utilized and interpreted by coaches, scouts and front offices across the league.

Section 4 - Model Evaluation

Throughout this section, our model will be evaluated for its practicality, accuracy and potential error. This will also cover the assumptions for a regression model and measure if this model can be utilized in the real world.

The model that we created has a multiple R^2 value of .08477 and an adjusted R^2 value of .08231. This indicates that roughly 8.2% of the variation in yards can be explained by the explanatory variables in the model. While this may appear very low, this does not necessarily mean that our model is "bad" or inaccurate. Rather, there is a lot of unexplained variation since this model is predicting human behavior which is very difficult to predict. Specifically, we see elements such as ball carrier vision, football instincts, offensive line assignments, and blown plays affect our ability to construct a highly accurate model. If there was a way to measure these variables and incorporate them into our model then we would have much more accurate predictions.

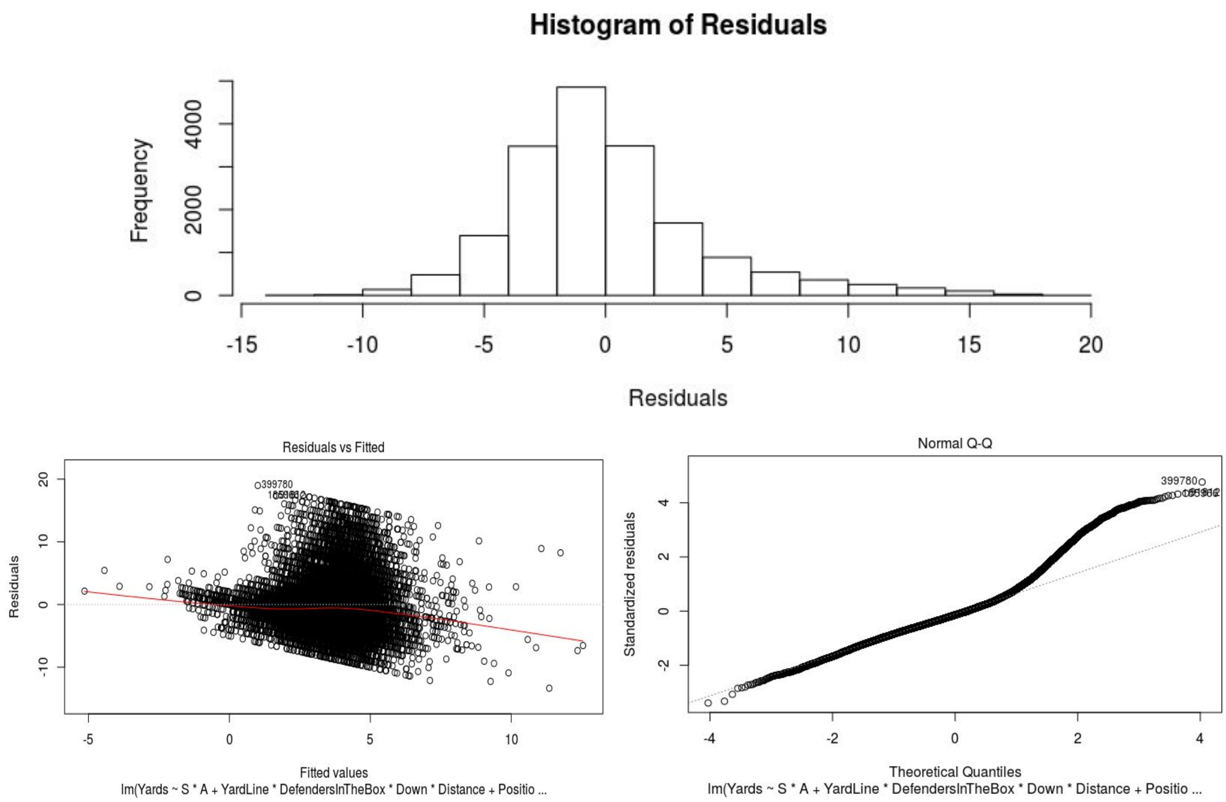
The full model that was analyzed in Section 3 indicates this was not an easy problem to solve. With all explanatory variables considered in the full model, the adjusted R^2 value was only slightly greater than our model. While the adjusted R^2 of the full model may be higher than our model, the usability of this full model is limited. With nearly 50 variables utilized in the full model, our model proves much more practical. Any interested party can utilize this model to predict the yardage that an NFL ball carrier would gain on a given run play.

The proposed model has an AIC of 62,246.14. Through backward elimination, we were able to reduce AIC from the full model and show that it was a stronger fit for predicting yardage gained. With a very low p-value of $2.2e^{-16}$, this shows that the model itself is statistically significant in predicting the yards as well. Incorporating interactions between variables increased the R^2 value greatly from the first models that were created using only several variables.

After finalizing the model selection process, it is essential to prove if this particular model is both useful and accurate. Although there are statistics that can prove these characteristics, visualizations will show that this model represents plausible predictions. Several assumptions need to be met before a model can be proven useful. The requirements of linearity, constant variability, independent observations, approximate normality of residuals, and no instances of multicollinearity all need to be satisfied.

By examining the residual analyses in Figure 3, we will evaluate how the histogram of residuals, residual plot, and QQ-Plot all represent proper residuals that satisfy all assumptions. Residuals represent the difference between actual results and predicted values. For a model to be useful, the assumptions of these residuals must be met. First, the histogram shows a normal distribution that is centered around zero. Similarly, the residual plot reflects constant variation but also does not show evidence of any sort of pattern or "fan" shape within the model. Lastly, by observing the QQ-Plot for the residuals, a linear relationship is shown for a majority of the distribution with a slight skew as we get further from the mean. In addition to the distribution of the residuals, two assumptions take on a more intuitive approach. The independent observations assumption is qualified since no play is impacted by another. While there are examples of plays that share similar characteristics such as down and distance, there is no impact from one play to the next for yards gained. For multicollinearity, there were some examples such as speed and acceleration for a ball carrier, but our model accounts for this through interaction terms. By examining several plots of the residuals for the model, there is clear and obvious evidence of normal and appropriate distributions which allows us to proceed with confidence that our model meets the assumptions of regression analysis.

Figure 3



Practical Example:
After meeting the assumptions for regression, we decided to move forward and look at an example. Using the proposed model above, our group chose reasonable values for each input variable and entered them into the regression equation. These values were entered as such:

- Weight = 220 Lbs
- Position = RB
- Down = 2nd down
- Distance = 5 Yards to Go
- Yardline = 30
- Speed = 4 Yards/sec
- Acceleration = 3 Yards/sec^2
- Defenders In The Box = 6

Predicted Yards = 4.22

The output of 4.22 yards is reasonable and shows that our model was not completely inaccurate due to the unexplained variation within it. Our group used this prediction to demonstrate the practicality of the model for coaches, front offices, or anyone else of interest.

Section 5 - Summary and Concluding Remarks

Certainly, there is no single model that can predict all the variability in rushing yards. As mentioned earlier, teams around the NFL have dedicated a tremendous amount of time and resources in an attempt to solve this problem which has proved quite difficult. We have seen the increased importance of "sabermetrics" in several professional sports – including American football – in an attempt to maximize ball carrier efficiency and value. By looking deeper into the statistical elements of any given ball carrier or play, teams may find hidden stars while increasing their ability to identify opportune run scenarios. Similarly, teams will have a better idea of how to game plan and manage rosters in order to be successful on the ground which will ultimately help win games.

In this report, we construct a predictive model for rushing yards that can be utilized by franchises across the league. Based on the results of our regression analysis, it appears there are some pre-handoff characteristics that improve rushing efficiency. The primary variables discussed in this study are the number of defenders in the box, speed, acceleration, down, and distance as well as other secondary factors. Other variables such as score spread were found insignificant in predicting rushing yards.

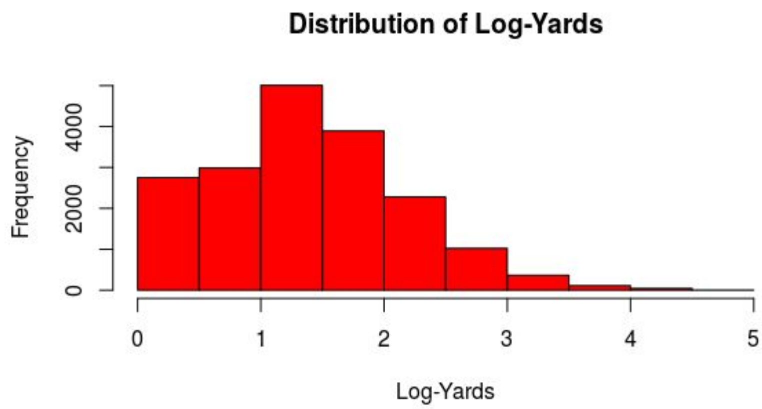
We are confident that our model provides a solid baseline for predicting rushing yards but we know there is much room for improvement and expansion in the future. Many elements of a successful run play are immeasurable or heavily attributed to chance. If there was a way to measure variables such as ball carrier vision, football intelligence, or frequency of blown plays, then we would have been able to build a much more accurate predictive model. Because of this, our model is limited to predicting rushing plays that result in yardage between -5 and 20 yards.

As the NFL continues to evolve both on the field and in front offices, this problem will only grow in size and scope. Maybe someday the league will have a conclusive answer for "What makes for a 'successful' rushing play?" Until then, it seems that the NFL will continue to entice fans with its unpredictable nature while coaches and executives attempt to harness data that will improve their team on the field.

Section 6 - Appendices

Appendix A: Dataset Homepage on Kaggle
https://www.kaggle.com/c/nfl-big-data-bowl-2020/overview

Appendix B: Logarithmic Distribution of Yards



Appendix C: List of Variables Provided in NFL Big Data Bowl with Descriptions

1. GameId - a unique game identifier
2. PlayId - a unique play identifier
3. Team - home or away
4. X - player position along the long axis of the field.
5. Y - player position along the short axis of the field.
6. S - speed in yards/second
7. A - acceleration in yards/second^2
8. Dis - distance traveled from prior time point, in yards
9. Orientation - orientation of player (deg)
10. Dir - angle of player motion (deg)
11. NflId - a unique identifier of the player
12. DisplayName - player's name
13. JerseyNumber - jersey number
14. Season - year of the season
15. YardLine - Yard line of the line of scrimmage
16. Quarter - game quarter (1-5, 5 == overtime)
17. GameClock - time on the game clock
18. PossessionTeam - team with possession
19. Down - the down (1-4)
20. Distance - yards needed for a first down
21. FieldPosition - which side of the field the play is happening on
22. HomeScoreBeforePlay - home team score before play started
23. VisitorScoreBeforePlay - visitor team score before play started

24. NflIdRusher - the NflId of the rushing player
25. OffenseFormation - offense formation
26. OffensePersonnel - offensive team positional grouping
27. DefendersInTheBox - number of defenders lined up near the line of scrimmage, spanning the width of the offensive line
28. DefensePersonnel - defensive team positional grouping
29. PlayDirection - direction the play is headed
30. TimeHandoff - UTC time of the handoff
31. TimeSnap - UTC time of the snap
32. Yards - the yardage gained on the play (you are predicting this)
33. PlayerHeight - player height (ft-in)
34. PlayerWeight - player weight (lbs)
35. PlayerBirthDate - birth date (mm/dd/yyyy)
36. PlayerCollegeName - where the player attended college
37. Position - the player's position (the specific role on the field that they typically play)
38. HomeTeamAbbr - home team abbreviation
39. VisitorTeamAbbr - visitor team abbreviation
40. Week - week into the season
41. Stadium - stadium where the game is being played
42. Location - city where the game is being played
43. StadiumType - description of the stadium environment
44. Turf - description of the field surface
45. GameWeather - description of the game weather
46. Temperature - temperature (deg F)
47. Humidity - humidity
48. WindSpeed - wind speed in miles/hour
49. WindDirection - wind direction

Appendix D: Comparison of Test-Train Results for Our Best Model

	Training Data	Testing Data
Adjusted R^2	0.085	0.079
Residual Standard Error	4.027	4.015

Appendix E: Full Model With No Interactions

While creating the full model in R we ran into some difficulty as it was unable to run the regression for a model with all interactions of the fifty variables due to the size and scope of the dataset. Thus, this full model did not incorporate all interactions but allowed us to get a better understanding of the data we were dealing with and how to approach the creation of our model.

```
> mod.full=lm(Yards~.-DisplayName,data=NFL.train)
Residual standard error: 4.01 on 10546 degrees of freedom
  (3247 observations deleted due to missingness)
Multiple R-squared:  0.1348,    Adjusted R-squared:  0.09276
F-statistic: 3.208 on 512 and 10546 DF,  p-value: < 2.2e-16
```

Appendix F: List of coefficients for best predictive model

```
Coefficients: (1 not defined because of singularities)
                                           Estimate Std. Error t value Pr(>|t|)    
(Intercept)                               1.723e+02  6.952e+01   2.478  0.01323 *  
S                                         6.544e-01  8.453e-02   7.741 1.07e-14 ***
A                                         1.447e+00  1.488e-01   9.727  < 2e-16 ***
YardLine                                  2.232e-02  1.314e-01   0.170  0.86512    
DefendersInTheBox                        -2.429e-02  1.982e-01  -0.123  0.90246    
Down2                                     5.647e+00  2.148e+00   2.629  0.00858 ** 
Down3                                     4.188e+00  2.438e+00   1.718  0.08579 .  
Down4                                     9.009e+00  9.112e+00   0.989  0.32281    
Distance                                  6.556e-01  2.012e-01   3.258  0.00113 ** 
PositionDT                                9.918e+01  4.206e+01   2.358  0.01840 *  
PositionFB                               -1.640e+02  7.043e+01  -2.328  0.01994 *  
PositionHB                               -1.867e+02  6.989e+01  -2.671  0.00758 ** 
PositionQB                               -1.857e+02  7.062e+01  -2.629  0.00857 ** 
PositionRB                               -1.775e+02  6.953e+01  -2.553  0.01069 *  
PositionTE                               -1.265e+02  8.083e+01  -1.565  0.11767    
PositionWR                               -1.732e+02  6.958e+01  -2.489  0.01282 *  
PlayerWeight                             -9.038e-01  3.675e-01  -2.460  0.01392 *  
S:A                                      -1.344e-01  3.287e-02  -4.089 4.36e-05 ***
YardLine:DefendersInTheBox                1.021e-02  1.971e-02   0.518  0.60439    
YardLine:Down2                           -2.437e-02  1.414e-01  -0.172  0.86313    
YardLine:Down3                            2.682e-02  1.466e-01   0.183  0.85484    
YardLine:Down4                           -2.125e-01  4.009e-01  -0.530  0.59602    
DefendersInTheBox:Down2                  -5.153e-01  2.636e-01  -1.954  0.05069 .  
DefendersInTheBox:Down3                  -4.261e-01  2.960e-01  -1.439  0.15005    
DefendersInTheBox:Down4                  -1.021e+00  1.126e+00  -0.907  0.36444    
YardLine:Distance                        -1.227e-03  1.349e-02  -0.091  0.92753    
DefendersInTheBox:Distance               -4.893e-02  2.581e-02  -1.895  0.05808 .  
Down2:Distance                           -5.829e-01  2.867e-01  -2.033  0.04204 *  
Down3:Distance                           -7.140e-01  2.618e-01  -2.727  0.00639 ** 
```


```
YardLine:Down4                           -2.125e-01  4.009e-01  -0.530  0.59602    
DefendersInTheBox:Down2                  -5.153e-01  2.636e-01  -1.954  0.05069 .  
DefendersInTheBox:Down3                  -4.261e-01  2.960e-01  -1.439  0.15005    
DefendersInTheBox:Down4                  -1.021e+00  1.126e+00  -0.907  0.36444    
YardLine:Distance                        -1.227e-03  1.349e-02  -0.091  0.92753    
DefendersInTheBox:Distance               -4.893e-02  2.581e-02  -1.895  0.05808 .  
Down2:Distance                           -5.829e-01  2.867e-01  -2.033  0.04204 *  
Down3:Distance                           -7.140e-01  2.618e-01  -2.727  0.00639 ** 
```


```
Down4:Distance                            1.348e+00  6.265e+00   0.215  0.82967    
PositionDT:PlayerWeight                          NA         NA      NA       NA    
PositionFB:PlayerWeight                   8.549e-01  3.703e-01   2.309  0.02098 *  
PositionHB:PlayerWeight                   9.498e-01  3.689e-01   2.575  0.01005 *  
PositionQB:PlayerWeight                   9.417e-01  3.716e-01   2.534  0.01129 *  
PositionRB:PlayerWeight                   9.076e-01  3.675e-01   2.470  0.01353 *  
PositionTE:PlayerWeight                   7.045e-01  4.030e-01   1.748  0.08051 .  
PositionWR:PlayerWeight                   8.902e-01  3.677e-01   2.421  0.01549 *  
YardLine:DefendersInTheBox:Down2         -6.626e-03  2.096e-02  -0.316  0.75187    
YardLine:DefendersInTheBox:Down3         -1.063e-02  2.146e-02  -0.495  0.62042    
YardLine:DefendersInTheBox:Down4          1.350e-02  5.795e-02   0.233  0.81583    
YardLine:DefendersInTheBox:Distance      -1.108e-03  2.025e-03  -0.547  0.58437    
YardLine:Down2:Distance                   3.238e-03  1.544e-02   0.210  0.83389    
YardLine:Down3:Distance                   1.077e-02  1.520e-02   0.709  0.47840    
YardLine:Down4:Distance                  -1.573e-03  3.160e-01  -0.005  0.99603    
DefendersInTheBox:Down2:Distance          5.000e-02  3.977e-02   1.257  0.20866    
DefendersInTheBox:Down3:Distance          8.787e-02  3.944e-02   2.228  0.02589 *  
DefendersInTheBox:Down4:Distance         -1.839e-01  8.391e-01  -0.219  0.82653    
YardLine:DefendersInTheBox:Down2:Distance  6.341e-04  2.311e-03   0.274  0.78380    
YardLine:DefendersInTheBox:Down3:Distance -1.112e-03  2.346e-03  -0.474  0.63546    
YardLine:DefendersInTheBox:Down4:Distance  3.714e-03  4.813e-02   0.077  0.93849    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```