# When Should NFL Teams Go For it On 4th Down?

By Tom Seifert

Intro:

Throughout a football game, each coach is faced with many decisions about what to do on 4th down plays. While converting a 1st down on a 4th down undoubtedly improves a team's likelihood of scoring on a given offensive drive, the possibility of failing to convert a 4th down introduces a significant risk in 4th down conversion attempts. When the offensive team is backed up in their own territory, for example, failing to convert on 4th down would put the opposing offense in an opportune position to score. When the offensive team is instead deep in opponent territory, failing to convert on 4th down results in a missed opportunity to score 3 points by settling for a field goal attempt.

These risks force football coaches to perform risk analysis in every 4th down scenario in order to evaluate whether the benefits of converting outweigh the negative impact of failing to convert and make a decision

accordingly. With the uptick in data-driven decision making in the NFL today, the way coaches make decisions on 4th down is changing. Here's a graph showing the change in the percentage of 4th downs in which a team attempted to convert over the past 11 full NFL seasons:

NFL 4th Down Attempt Rate by Year
(Meaningful Attempts)
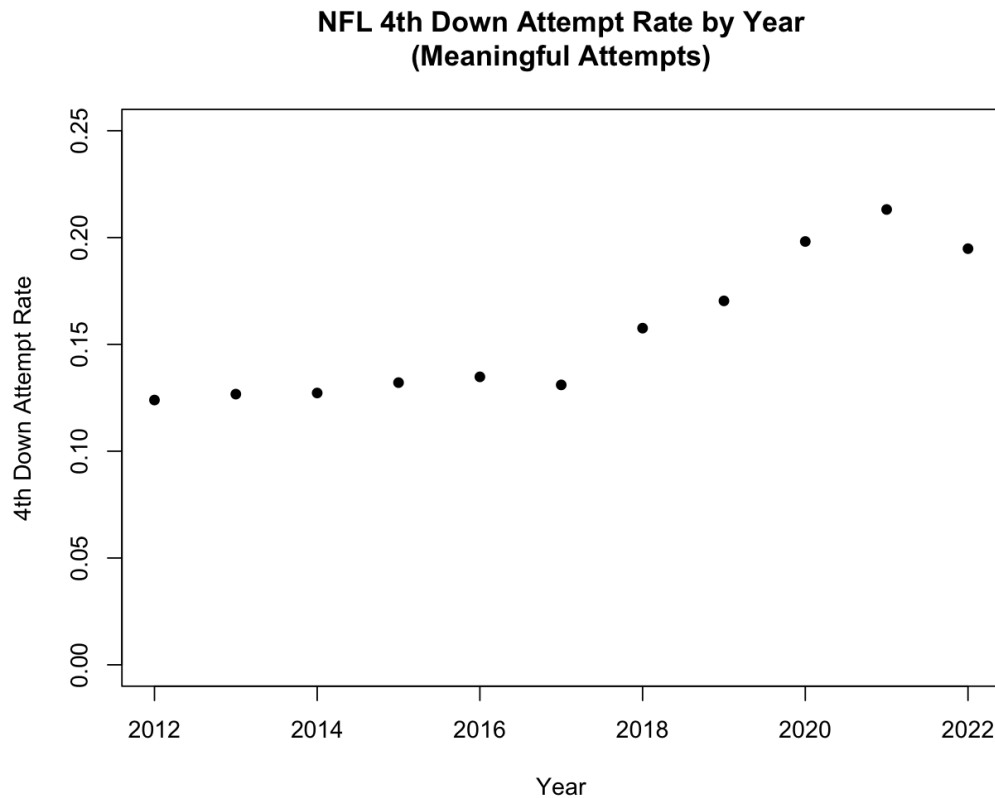
The last 10 years has brought upon a fairly steady increase in league-wide 4th down attempt rate, starting at 11.4% in 2012 and peaking at 20.6% in 2021. But why? What has changed about how NFL coaches decide whether to go for it on 4th down? Does the increase in attempts mean that teams are making better decisions when it comes to going for it on 4th down? This does not seem to necessarily be the case. Here is a graph that shows the league-wide 4th down conversion percentage over the last 11 full seasons:

NFL 4th Down Conversion Rate by Year
(Meaningful Attempts)
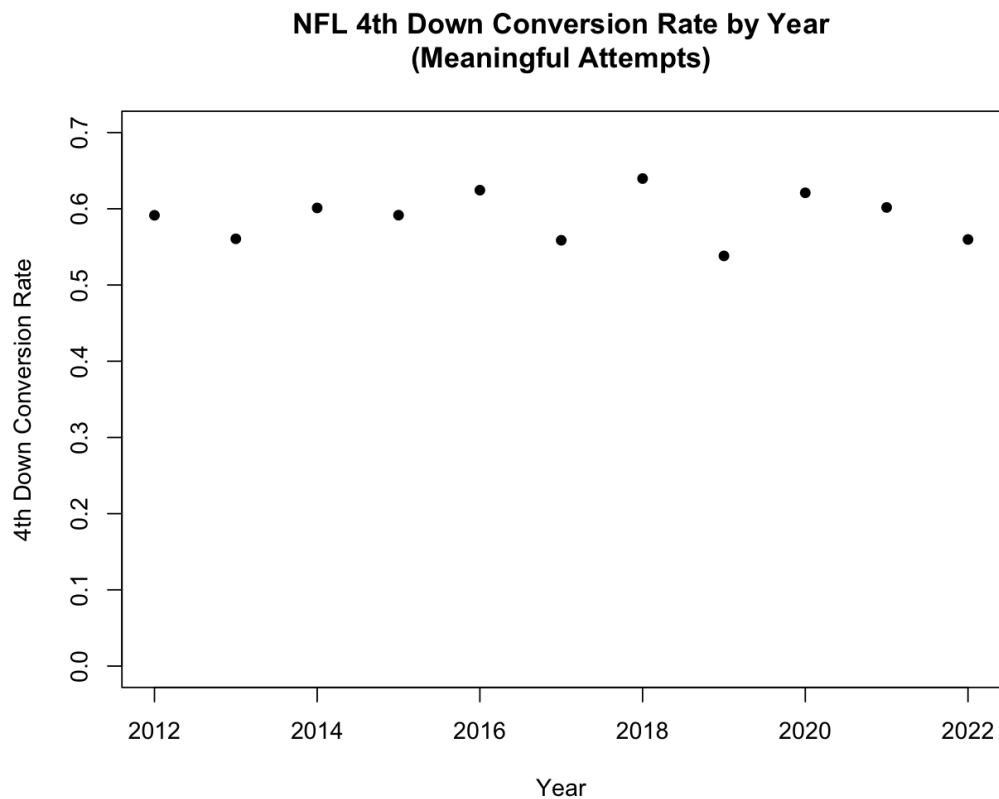
It appears that despite the increase in attempts over the past 10 years, the league-wide 4th down conversion rate has not significantly changed. This suggests that teams may not have improved their decision-making, and instead have only become more willing to accept the risk of going for it on 4th down given the benefits of converting. While there is no formula that coaches can follow to make the correct decision every time, the process by which coaches decide whether to go for it on 4th down can always be improved.

My goal is to build a model that can successfully decide whether an NFL team should or should not go for it on 4th down. To achieve this, I will use 4th down data from 2012 to 2022 to find statistics that correlate to the probability of converting a 4th down and making a field goal, as well as statistics that correlate to the added value of converting/failing a 4th down and making/missing a field goal. From there, I will create machine learning models with these statistics to estimate the effect of each 4th down decision on a team's chance of winning. I will then use the models to provide recommendations of what decision to make on all 4th downs from 2012-2022. The following definitions will be important in understanding the models:

- **Win Probability (WP)**: the estimated probability of the offensive team winning at a given point in the game. Calculated using score, time remaining, down & distance, and field position.

- **Win Probability Added (WPA)**: the change in win probability on a given play.

- **Yards to go**: the number of yards needed to convert a first down on a given play.

- **Score differential**: the difference between the scores of the offensive and defensive team on a given play.

- **Field position**: the position of the offensive team on the field on a given play, represented by the number of yards away they are from the endzone, between 0 and 100 (i.e. a field position of 88 indicates that the offensive team is 88 yards from the opponent's end zone).

"Meaningful" 4th Downs:

In order to ensure that my model can recommend reasonable 4th down decisions, I will only build it upon 4th down plays which I consider to be "meaningful." There are many situations where a team may take an extreme course of action on 4th down because of the low risk associated. For instance, when there are only a few seconds remaining in the half, the opposing team cannot recover possession if the 4th down attempt fails. Thus, the offensive team may elect to go for it on a 4th down that they have a very low chance of converting, because failing to convert has no associated risk. A similar mindset may guide decision-making towards the end of the game when the offensive team needs a score to have a chance at winning, or in a blowout where either team leads by an insurmountable amount of points.

I have established the following criteria for "meaningful" 4th downs in order to filter out these extreme scenarios, and I will only be using data that fit these criteria in my models:

- There are 30+ seconds remaining in the half at the start of the play
- There are 5+ minutes left in the game at the start of the play
- The offensive team's win probability is between 2% and 98% at the start of the play
- There are 25 or less yards to go to convert a first down

In addition, I will exclude all overtime plays because of how the rules differ from regulation game time, and I will only consider regular season games.

In every one of these models, I will split relevant data sets into two: a training set and a test set. The training set will contain 70% of the data chosen at random, and the test set will contain the remaining 30% of the data. I will use the training set to create the models (training the models), and I will use the test sets to test the effectiveness of the models in predicting certain outcomes.

Conversion/FG Probability Models:

The first two models I will construct will be logistic regression models; one to compute the probabilities of converting on 4th down, and one to compute the probability of making a field goal.

4th Down Conversion Probability Model

The conversion probability model will be trained and tested using data on all meaningful 4th down attempts from 2012-2022. The model is based only on yards to go because I anticipate that as yards to go increases, converting a 4th down becomes more difficult. Here is a visualization to show how the model predicts 4th down conversion probabilities based on existing data:
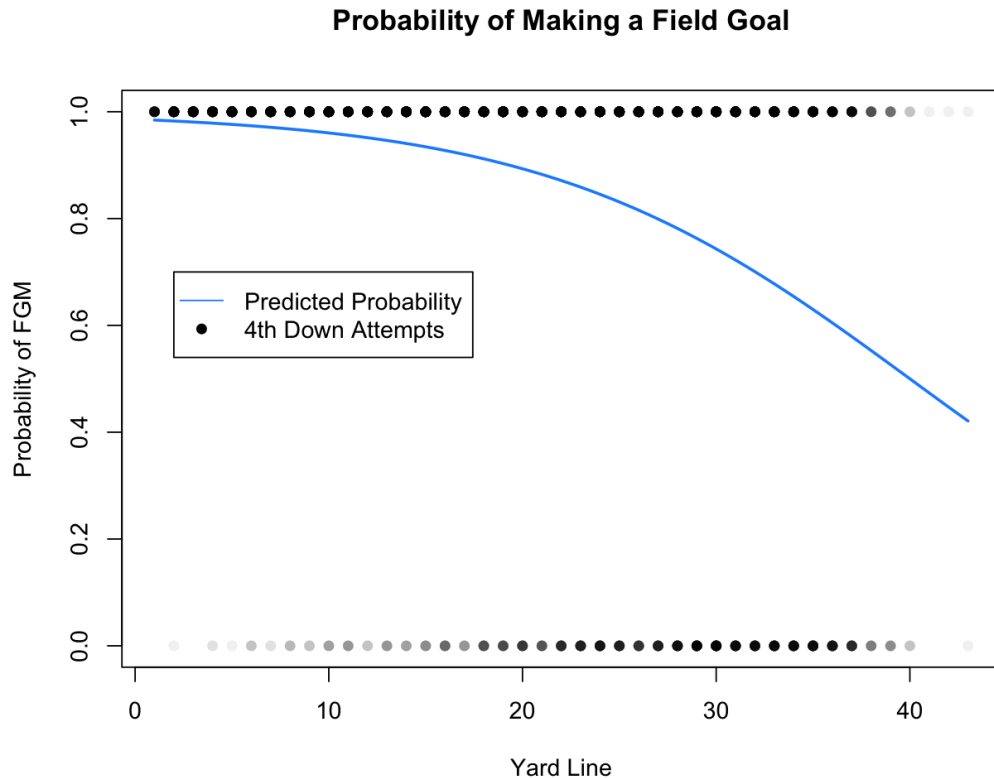
Probability of Converting a 4th Down
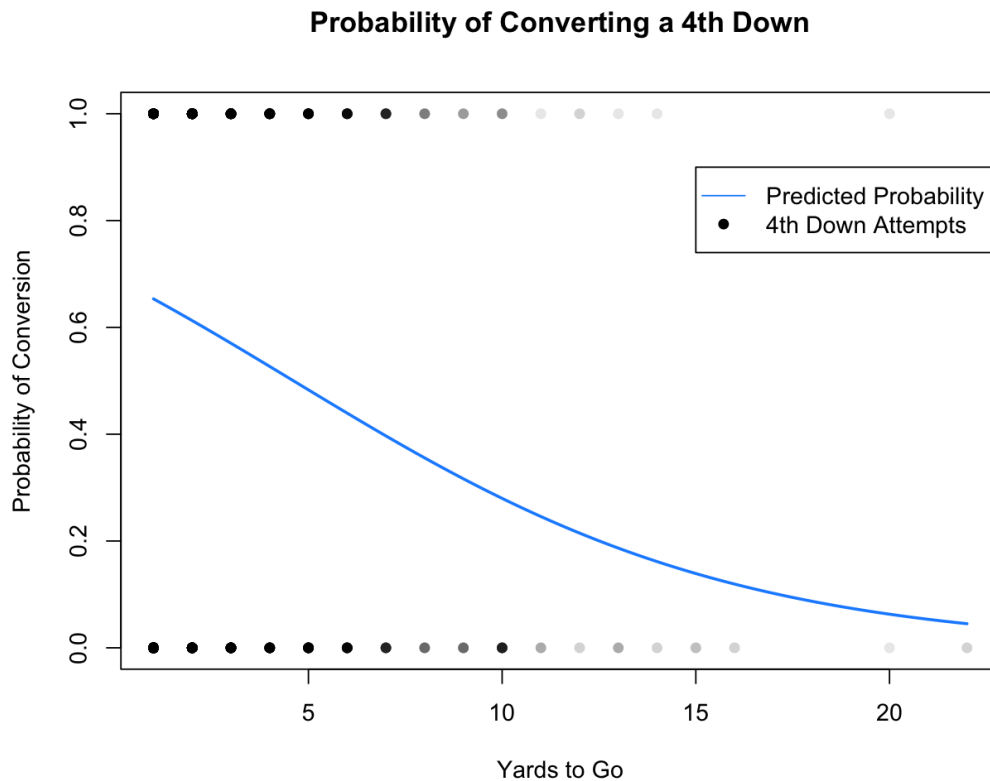
The points show the distribution of yards to go values on the existing 4th down attempts data plotted with their corresponding conversion probabilities. A conversion probability of 1 indicates a successful conversion while a conversion probability of 0 indicates a failed conversion attempt.

The blue curve represents the model's predicted conversion probabilities. The model follows the intuition that the probability of converting a 4th down decreases as the yards needed for a first down increases. Interestingly, a 1 yard difference in yards to go has less impact on conversion probability as yards to go increases.

Field Goal Made Probability Model

The field goal made probability model will be trained and tested using data on all attempted field goals on meaningful 4th downs from 2012-2022. The model is completely based on the yard line from which the field goal is attempted because as the offensive team's position on the field becomes further away from the end zone, making a field goal should become more difficult. Here is a visualization to show how the model predicts field goal conversion probabilities based on existing data:

Probability of Making a Field Goal

The points show the distribution of yard line values on the field goal attempts data plotted with their corresponding conversion probabilities. A conversion probability of 1 indicates a made field goal while a conversion probability of 0 indicates a missed field goal attempt.

The blue curve represents the model's predicted field goal conversion probabilities. The model follows the intuition that the probability of making a field goal decreases as the distance the offensive team is from the endzone increases. More specifically, the model's probability of making a field goal decreases exponentially as the distance of the kick increases.

There is still much more to explore beyond these models, as the probability of making a field goal or converting a 4th down does not fully describe what course of action a team should take on 4th down. For example, if a team is well within field goal range and facing 4th & 15 while down 12 points in the 4th quarter, they might go for the 4th down conversion instead of the field goal attempt in hopes of increasing the likelihood of scoring a touchdown and cutting the deficit to 6. In this case, the option that is more likely going to succeed, making a field goal, is not the play call chosen.

In general, there is much more than a play call's probability of success that contributes to choosing a play call on 4th down because the factors that make it worth the risk to kick or attempt to convert on a 4th down relate more to the context of the game. These factors include how close the game is, the offensive team's field position, how many yards the offensive team needs for a first down, and how much time there is left in the game.

Conveniently, there is an existing statistic, win probability, that attempts to summarize these factors into a single metric that represents the likelihood of the offensive team winning the game.

Win Probability and Win Probability Added:

Win probability (WP) is calculated using score, time remaining, the down, the yardage needed for a first down, and the offensive team's field position. This means that when a 4th down is converted, win probability will increase due to better field position and a new set of downs. However, this effect could be dampened or heightened by how much time is left in the game and what the score of the game is. Likewise, when a 4th down is failed, win probability will decrease due to the opposite team's offensive field position and fresh set of downs, and this effect could be altered by how close the game is as well as how much time is left.

For a better assessment of what makes it worth going for it or kicking it on 4th down, we can use win probability added (WPA) to see how certain plays affect a team's chance of winning a game. Due to the opposite effect that change in factors like field position and yards to go have on WPA on conversions versus failed attempts, I will have to construct several different models. This will include two separate models for 4th down attempts, one for 4th downs failed and one for converted 4th downs, and two separate models for field goal attempts, one for made field goal attempts and one for missed field goal attempts. Each model will predict WPA based on the variables used to calculate WP. However, I will omit yardage needed for a first down in every model, as I found that predicted WPA values heavily overestimated actual WPA values when taking yards to go into account.
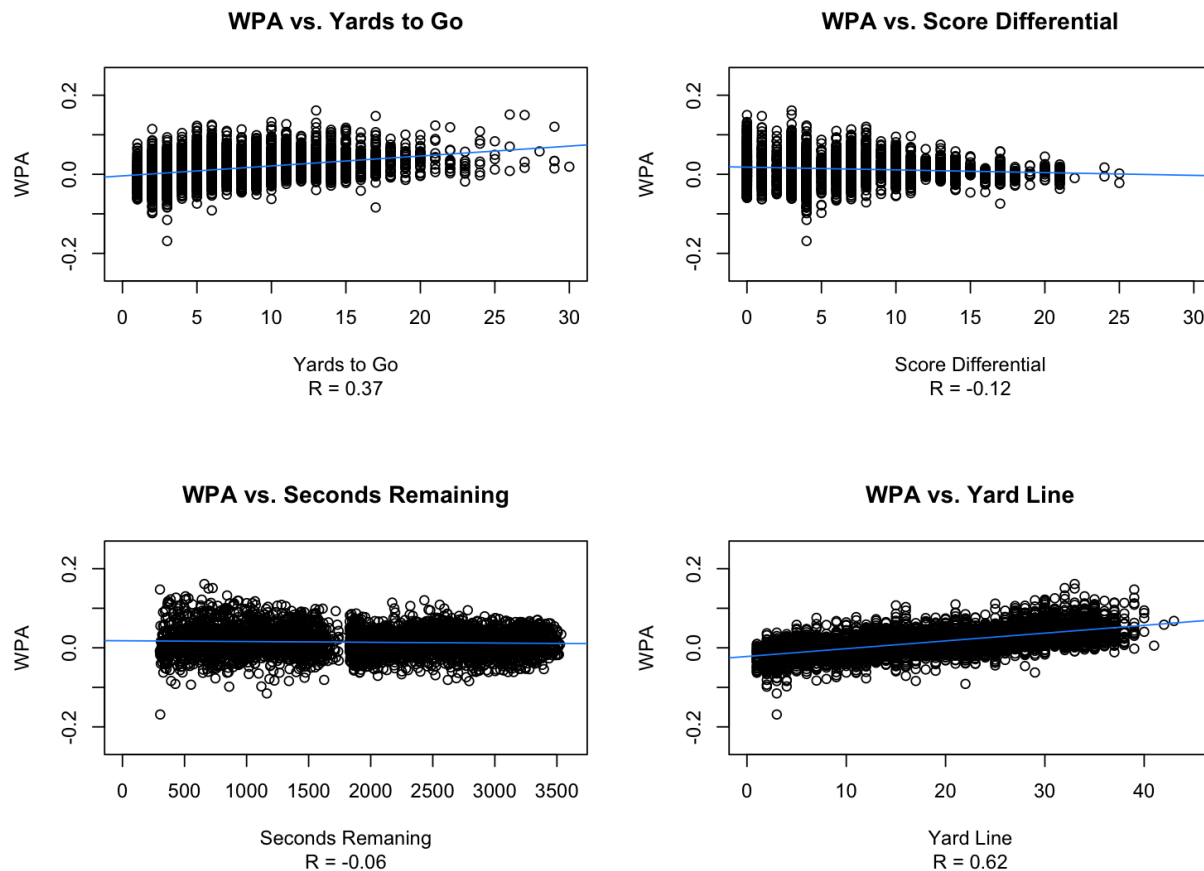
The motivation of these models is the creation of four statistics that then can be calculated for every meaningful 4th down:

- WPA if converted
- WPA if failed
- WPA if field goal made
- WPA if field goal missed

Since the data I am training these four models on (field position, score differential, and seconds remaining) is available for all 4th downs, and not just the specific cases in which the models are trained on (conversion attempts and field goal attempts), I will be able to predict these four statistics for any 4th down using these models.

Field Goal Make WPA Model

The field goal make WPA model will be constructed using data from all made field goals from 2012 to 2022. Here's how each statistic used to calculate WPA correlates with WPA on field goal makes in the training data set:

WPA vs. Yards to Go

WPA

Yards to Go
R = 0.37

WPA vs. Score Differential

WPA

Score Differential
R = -0.12

WPA vs. Seconds Remaining

WPA

Seconds Remaning
R = -0.06

WPA vs. Yard Line

WPA

Yard Line
R = 0.62

While these are by no means the strongest linear relationships by themselves, the intersection of these variables prove to correlate strongly to WPA, with a multiple R-squared of 0.42.

Once creating the model to find the correlation between these statistics and WPA for field goal makes based on the training data, the model can be used to predict WPA on field goal makes. Predicted WPA values will be generated using the testing data (not taken into account in the model's construction) and compared with the actual WPA values in the test set to assess the model's effectiveness:
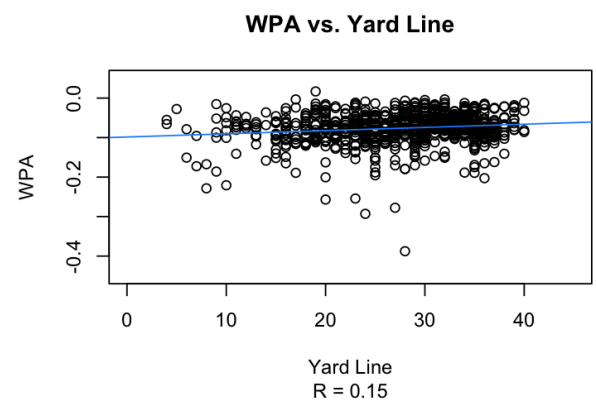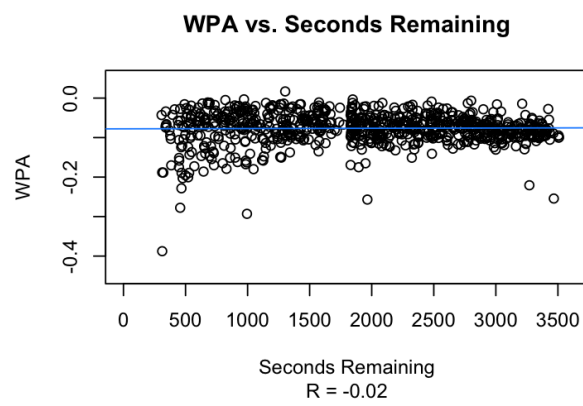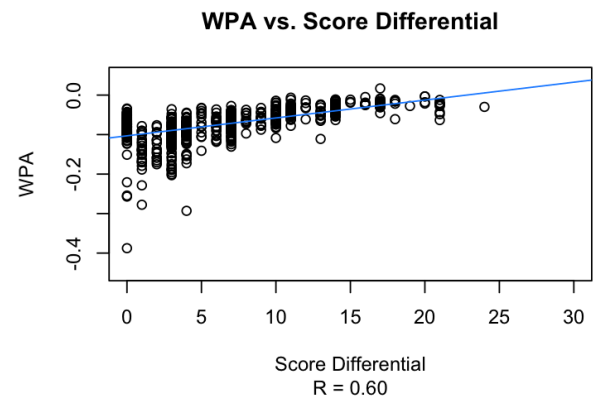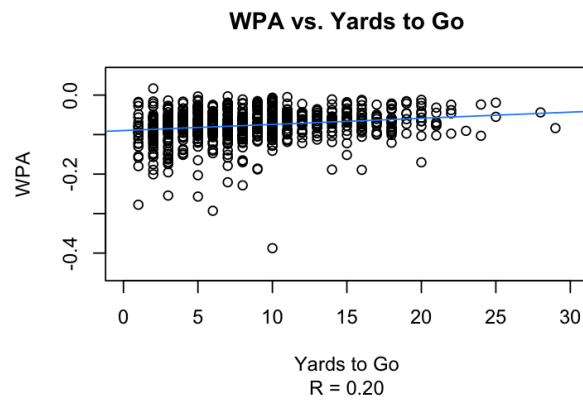
**Predicted WPA vs. WPA
(Field Goal Makes)**
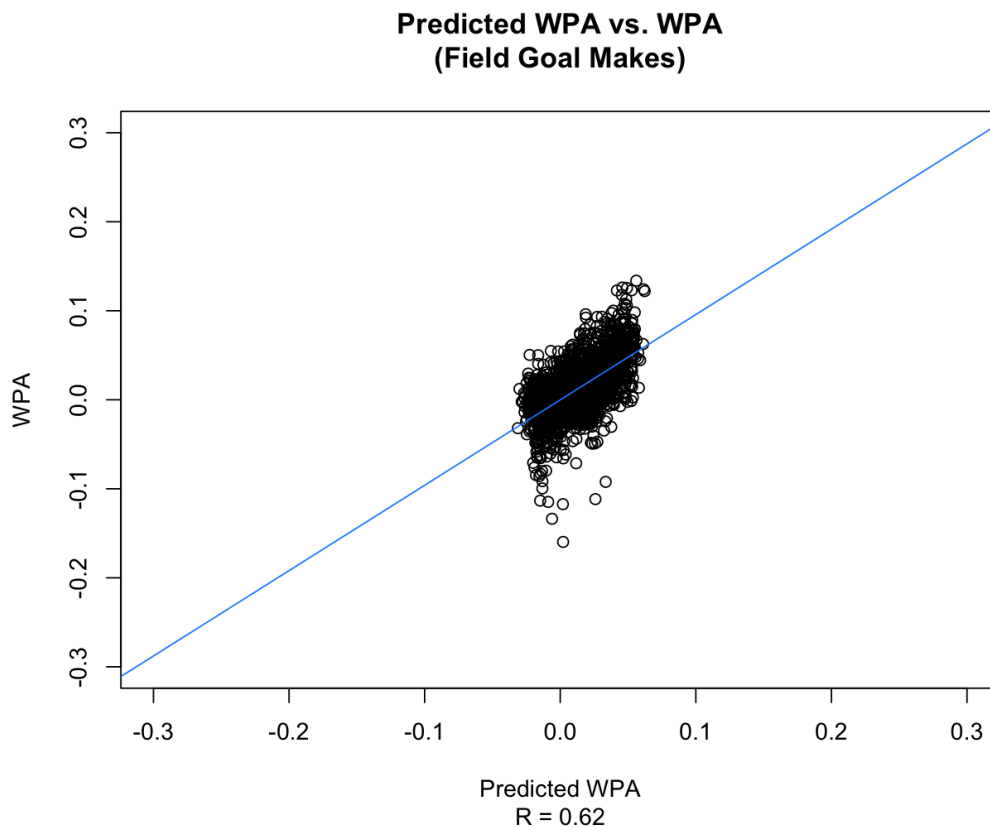
Predicted WPA
R = 0.62

Plotting predicted WPA with actual WPA for all field goal makes in the training set results in another fairly strong correlation with a correlation coefficient of 0.62, making these predictions sound.

I will repeat this process for each of the three remaining WPA models:
- Show correlation between statistics contributing to WPA and WPA for all relevant data (i.e. failed field goal attempts)
- Plot predicted WPA vs. real WPA to assess effectiveness of the model

Field Goal Miss WPA Model

Correlation of statistics contributing to WPA and WPA for all field goal misses:

WPA vs. Yards to Go

Yards to Go
R = 0.20

WPA vs. Score Differential

Score Differential
R = 0.60

WPA vs. Seconds Remaining

Seconds Remaining
R = -0.02

WPA vs. Yard Line

Yard Line
R = 0.15

Multiple R-Squared: 0.41 (fairly strong relationship)

Real WPA vs predicted WPA for field goal misses:
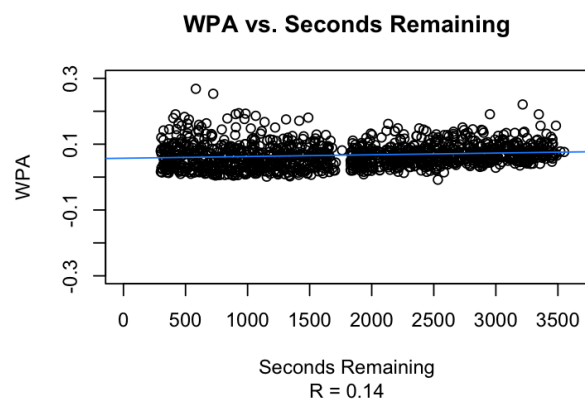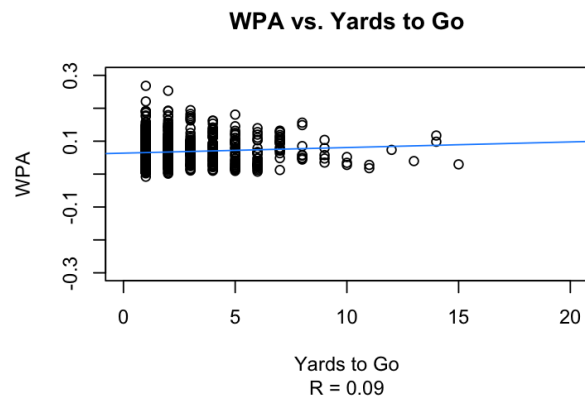
Predicted WPA vs. WPA
(Field Goal Misses)

WPA vs. Predicted WPA

R = 0.64
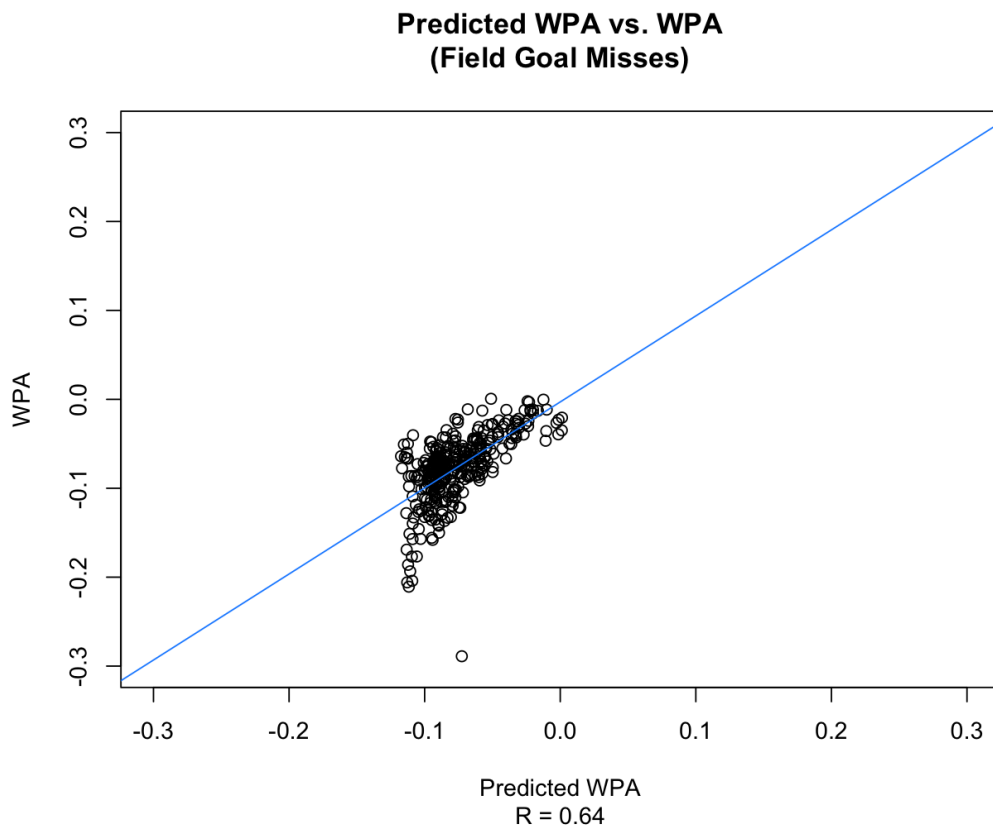
Correlation coefficient: 0.64 (solid relationship between predicted and actual WPA)

# Converted WPA Model

Correlation of statistics contributing to WPA and WPA for all converted 4th down attempts:

WPA vs. Yards to Go

WPA

Yards to Go
R = 0.09

WPA vs. Score Differential

WPA

Score Differential
R = -0.52

WPA vs. Seconds Remaining

WPA

Seconds Remaining
R = 0.14

WPA vs. Yard Line

WPA

Yard Line
R = -0.12
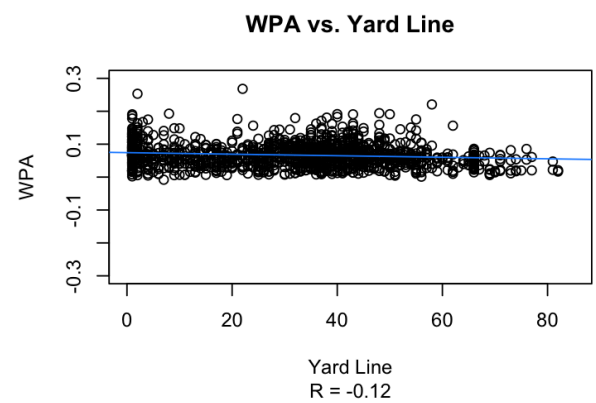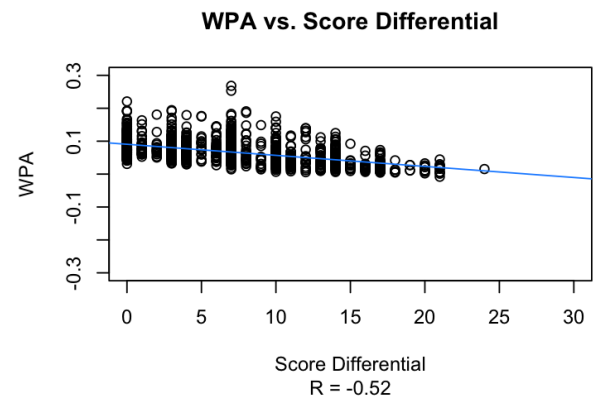
Multiple R-Squared: 0.29 (moderately strong relationship)

Real WPA vs predicted WPA for field goal misses:

Predicted WPA vs. WPA
(Fourth Down Conversions)
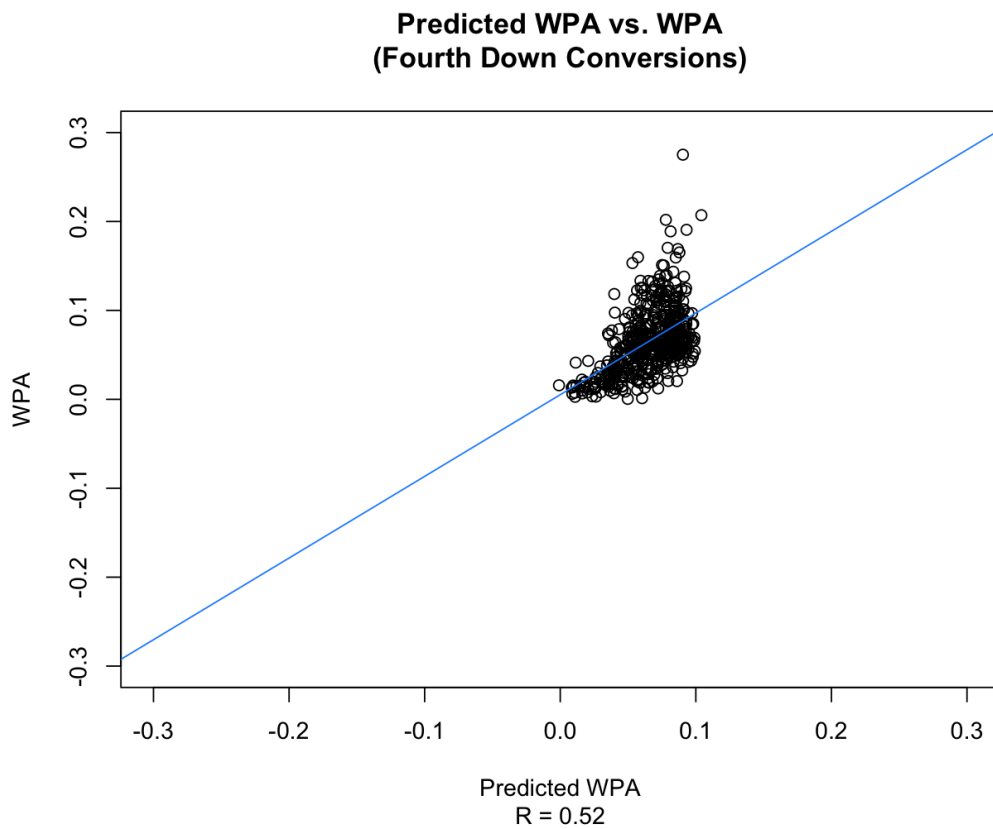
Correlation coefficient: 0.52 (moderate relationship between predicted and actual WPA)

Failed WPA Model

Correlation of statistics contributing to WPA and WPA for all failed 4th down attempts:

WPA vs. Yards to Go

Yards to Go
R = 0.36

WPA vs. Score Differential

Score Differential
R = 0.67

WPA vs. Seconds Remaining

Seconds Remaining
R = -0.21

WPA vs. Yard Line

Yard Line
R = 0.29
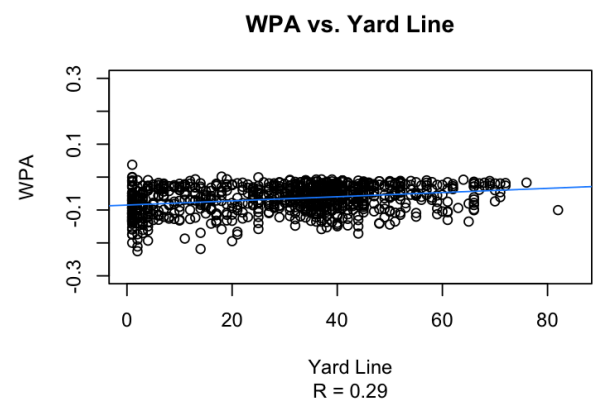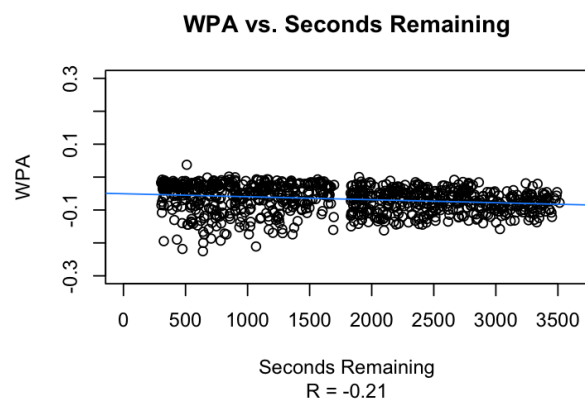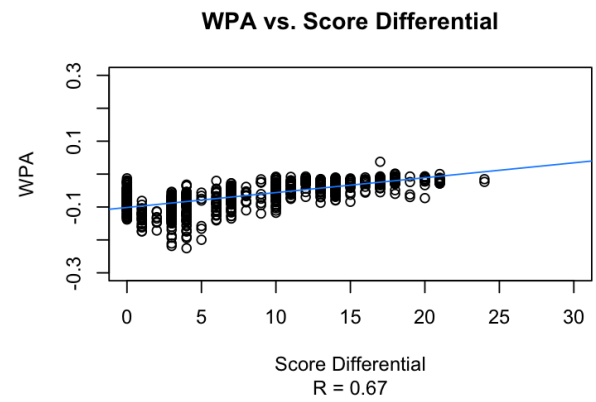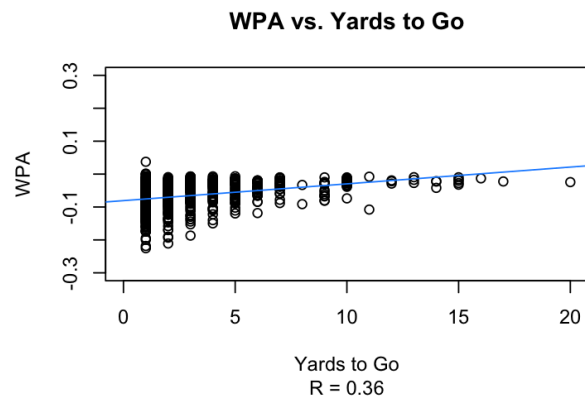
Multiple R-squared: 0.54 (strong relationship)

Real WPA vs predicted WPA for failed 4th down attempts:

Predicted WPA vs. WPA
(Failed Fourth Down Attempts)

Predicted WPA
R = 0.71
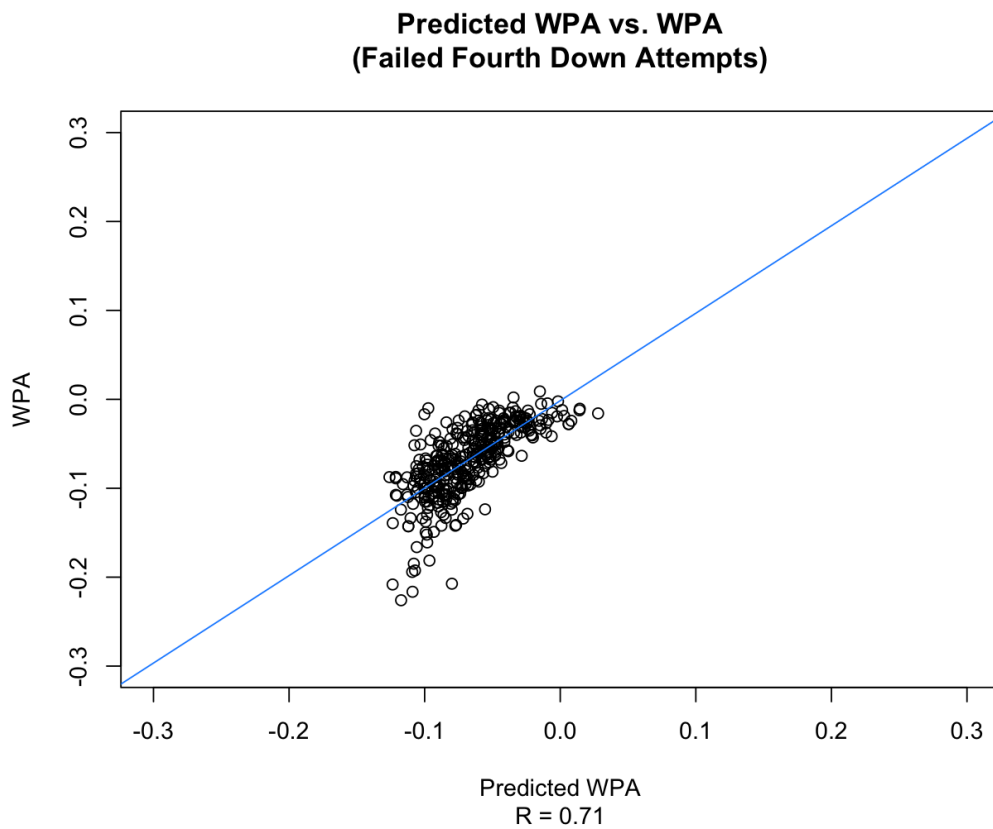
Putting it All Together:

Now that we have six models related to the likelihood of certain outcomes on the 4th down and their expected impact on the game, we can establish what it means when a team "should" go for it or attempt a field goal on 4th down.

To achieve this, I will create two statistics, Expected Field Goal WPA (eFGWPA) and Expected Attempt WPA (eAWPA), defined in the following ways:

Expected Field Goal WPA (eFGWPA):
- Definition: Expected WPA if a team attempts a field goal
- Equation: (WPA if FGM x FGM Prob) + (WPA if FG Miss x (1 - FGM Prob))

Expected Attempt WPA (eAWPA)
- Definition: Expected WPA if a team goes for it on 4th down
- Equation: (WPA if Conv x Conv Prob) + (WPA if Failed x (1 - Conv Prob))

In short, these statistics represent weighted averages for expected WPA on field goals and 4th down attempts separately. To present an example of how eAWPA and eFGWPA change with a statistic contributing to WPA, here is a graph of both statistics plotted against score differential:

Expected Win Probability Added vs. Score Differential
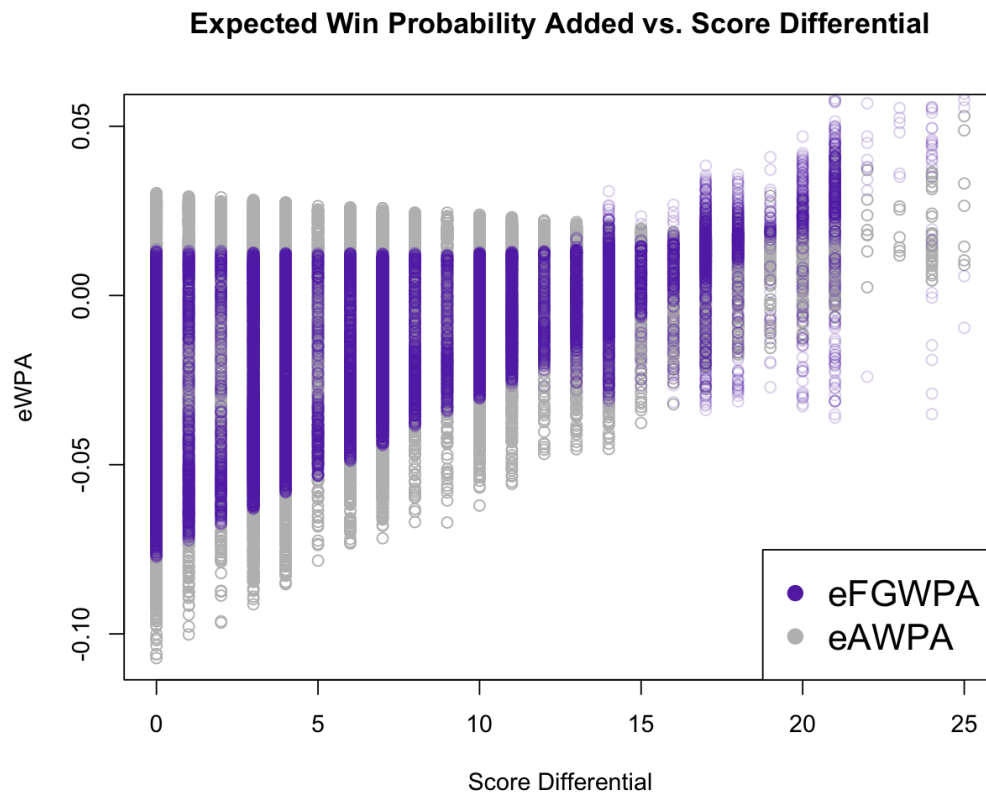
It appears that conversion attempts have a wider range of eWPA values when a game is close, while the opposite becomes true as the score differential increases. In other words, field goal attempts have less impact on a team's chance of winning the game on average than attempting to go for it on 4th down in a close game. Conversely, field goal attempts have on average a higher impact on a team's chance of winning in a more one-sided contest.

With all of these statistics available, we can analyze whether a team should go for it based on the expected contribution that each decision will make to the probability of the offensive team winning the game. For example, if eFGWPA exceeds eAWPA such that attempting a field goal yields a higher expected WPA than going for it on 4th down, then the team should kick a field goal. If instead eAWPA is higher, the team should go for it. However, if both eFGWPA and eAWPA are negative, such that both attempting a field goal and going for it on 4th down is expected to have a negative impact on the offensive team's win probability, then the team should punt. Using this rationale, the model can assign each 4th down a recommended decision:

Recommended Decision:
      Go for it:
          -    eAWPA is greater than eFGWPA
          -    eAWPA is positive
   Attempt a Field Goal:
          -    eFGWPA is greater than eAWPA
          -    eFGWPA is positive
    Punt:
          -    Both eAWPA and eFGWPA are negative

With these recommended outcomes, we can analyze how often NFL teams follow the model. Here's a graph showing the relative frequency of coaches following and not following the model's 4th down recommendation on all 4th downs from 2012-2022:

**Does the Model Agree With the Coach's 4th Down Decision?**
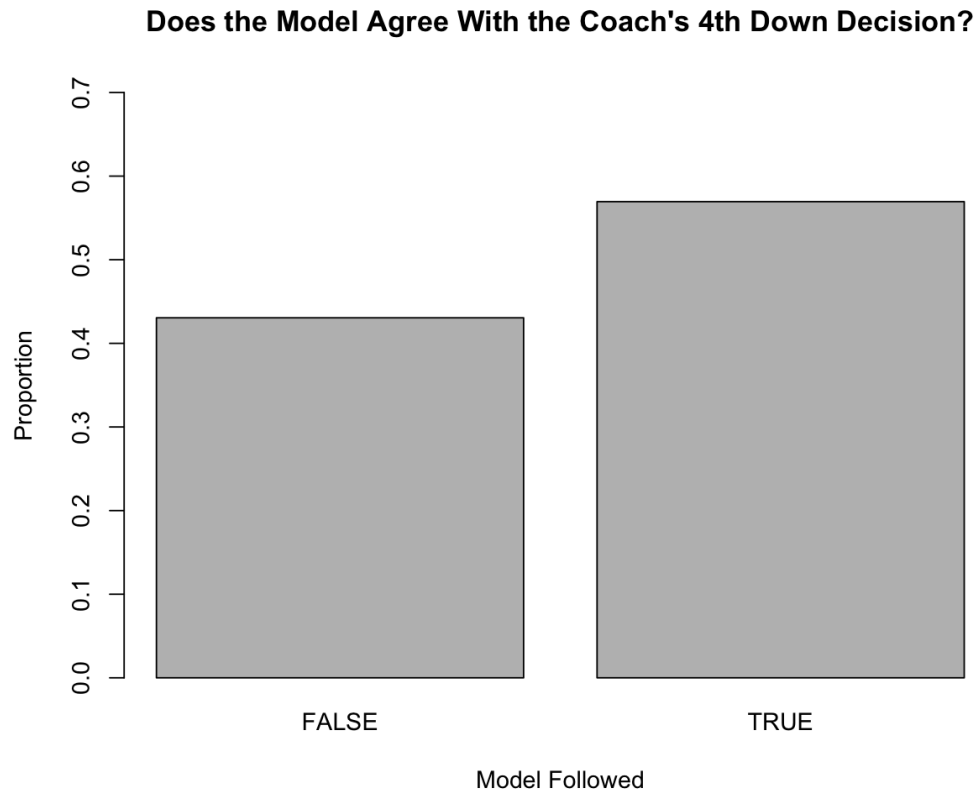
NFL coaches seem to consider more than just win probability when making 4th down decisions. However, this raises the question of whether they should take WPA into account to see better results on

4th down. In order to investigate this, we must compare the model's recommendations and coaches' decisions based on the rate at which they produce a positive WPA.

We find that the recommendations made by the model produce a positive WPA **62.8%** of the time, while the real 4th down data produces a positive WPA **54.3%** of the time. If coaches instead had followed the model strictly, they would see a higher proportion of positive WPA on 4th down decisions.

The model's recommendations not only differ from real NFL coaching decisions based on success rate, but also in the relative frequency of each play type. Here is a comparison between the frequency of play types called on all 4th downs and the frequency of the model's recommendation of each play type:

**4th Down Play Type Breakdown**
(2012-2022)
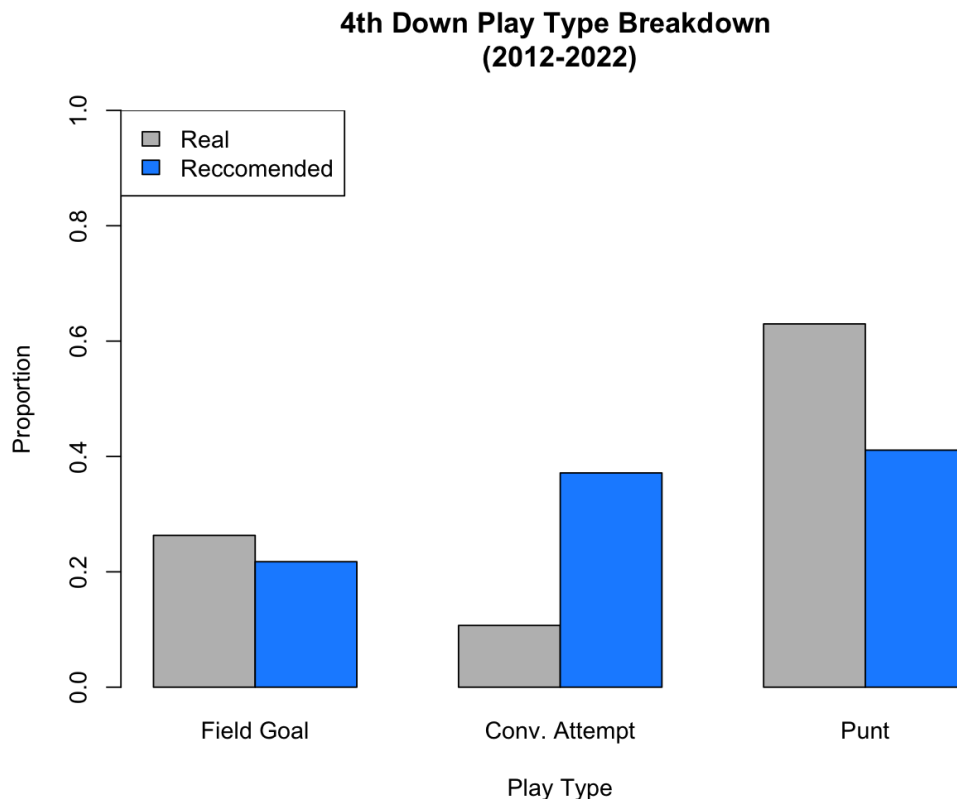
The model calls for a lot fewer punts, slightly fewer field goals and much more conversion attempts than the real data.

The 4th down conversion attempt rate of 37.2% suggested by the model would be a significant spike in the league average 4th down conversion attempt rate from 2012-2022, as shown below:

NFL 4th Down Attempt Rate by Year
**(Meaningful Attempts)**
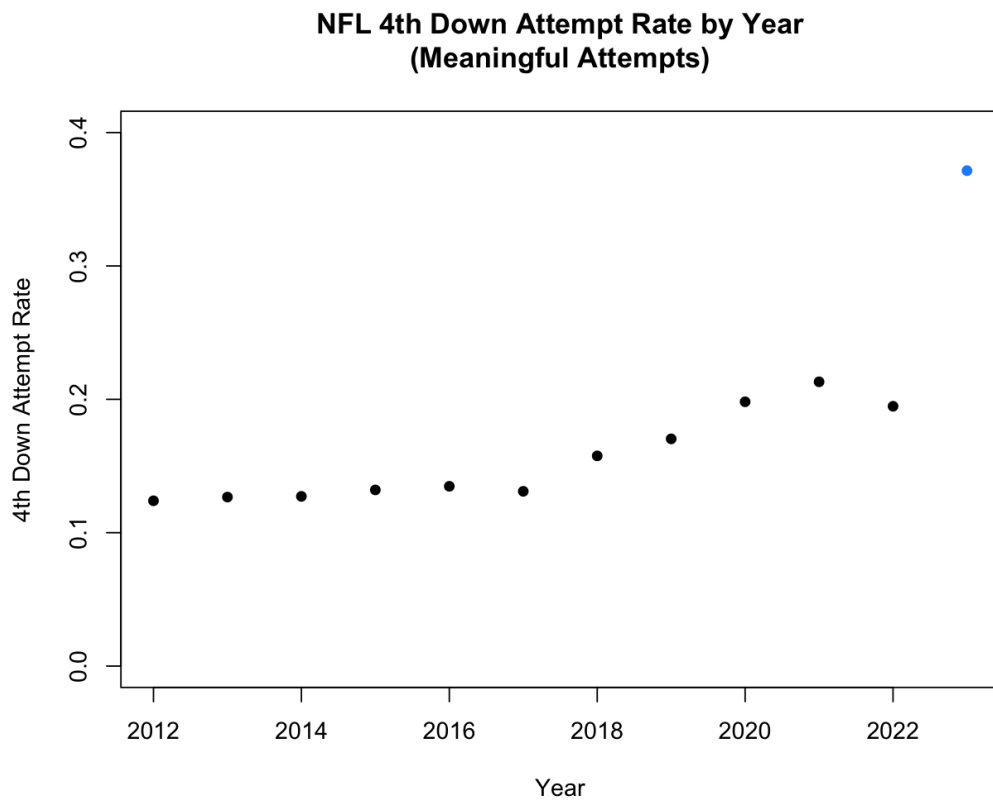
Thus, should coaches decide to follow the model more closely in the future, the league average 4th down attempt rate will likely increase. This could be the case in the near future, as the rate at which teams follow the model has increased steadily from 2012 to 2022:

League-Wide % of 4th Down Plays Following Model
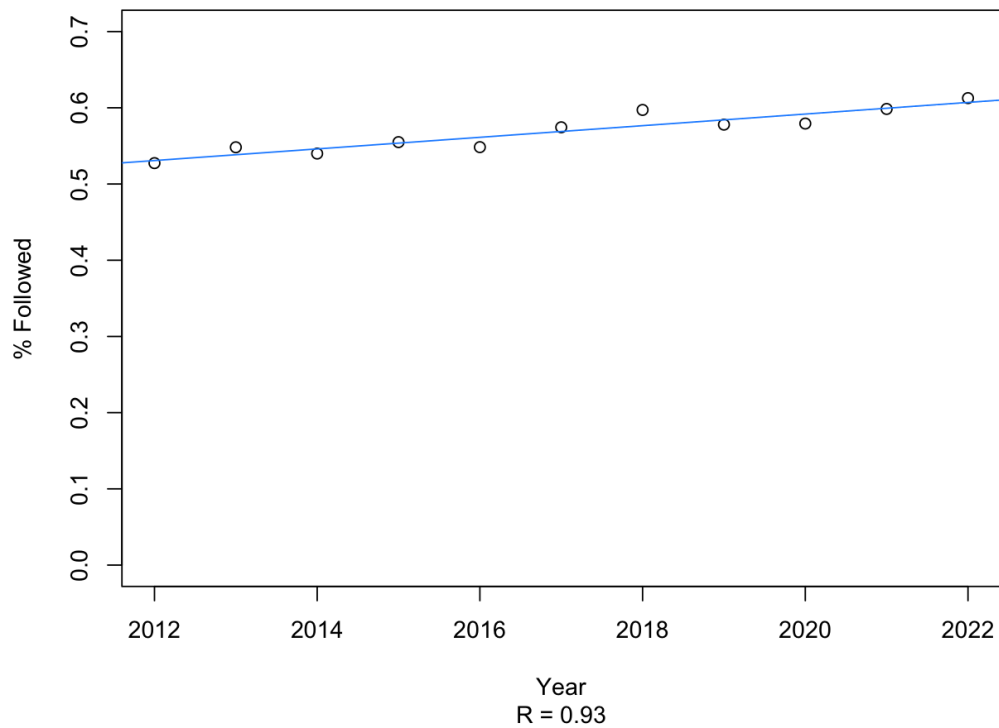
The positive linear relationship between year and % followed is very strong, with an R value of 0.93. This will likely lead to a higher conversion attempt rate, but it is unclear whether this change will translate to more offensive success on 4th down across the league.

In order to assess whether following the model contributes to success, here is a plot showing every team's win percentage plotted against their model followed % in 2022:

Team Win % vs. Model Followed %
(2022)
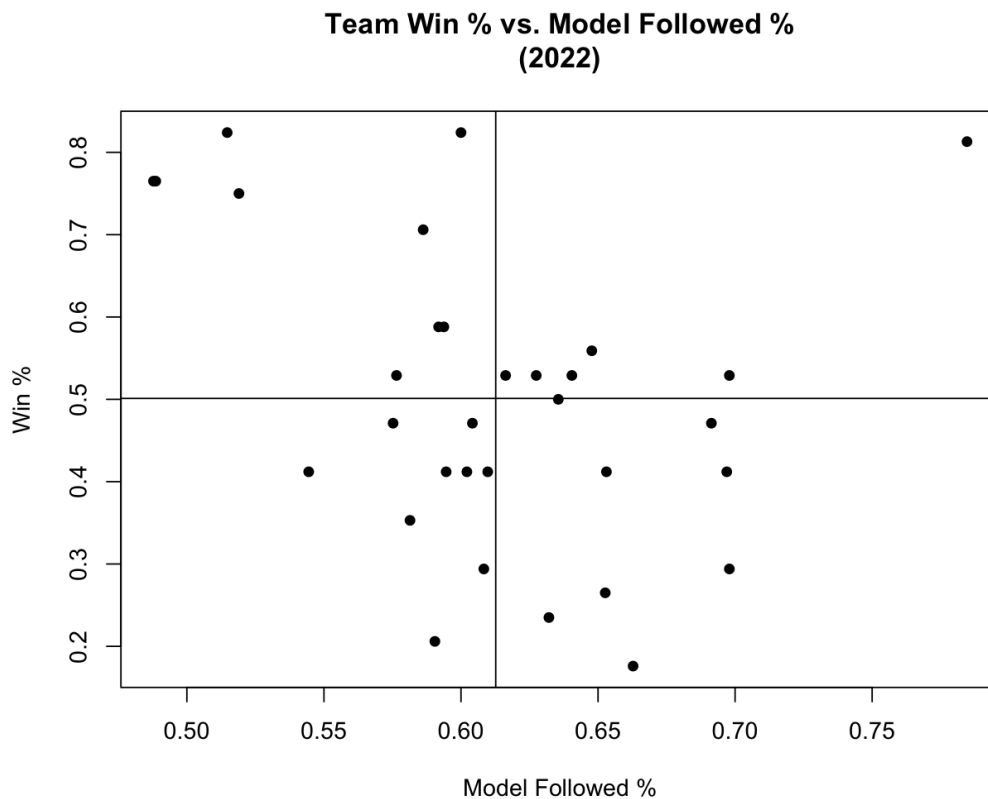
With a low correlation coefficient value of R = 0.31, following the model closely does not appear to contribute to success. It is interesting to note that despite this, the model continues to be followed more closely year by year.

Conclusion/Model Flaws and Further Direction:

While the culmination of these models offers an analytical framework from which to make educated 4th down decisions, by no means do these recommendations capture all that goes into making 4th down play calls. Possible weaknesses of these models include the following:

- Offensive styles, tendencies, and areas of strength of certain teams is not taken into account
- Coaching style is not considered—some game strategies lend itself to a more conservative approach, others to a more aggressive approach (i.e. strength of and reliance on defense to win the game)
- Player personnel is not taken into account— this might include the skill of the kicker, the effectiveness of linemen and backs on short yardage runs, reliability of receivers, and the strength of the defensive front or secondary
- There is subjectivity in the definition of "meaningful" 4th downs, which could be improved upon to make the model predict more accurately

- The models rely on the validity of WP and WPA stats, assuming that they are really representative of a team's win probability

Given more time and data at my disposal, I would analyze when the model under/over estimates WPA for certain play types in order to look for areas of weakness. In addition, I would look deeper into punting as a third option to assess how deciding to punt might also affect the probability of a team winning the game.

Sources:
- https://www.nflfastr.com/index.html