An Analysis of NFL Head Coaches

Sayom Ghosh-Dastidar

Table of Contents

1) Background and Motivation

In the National Football League, head coaches are their team's most public facing representatives. For the most part, fans do not have insight into how teams are run, so coaches are lavished with praise when their team wins and receive the brunt of the scorn when they lose, regardless of how much responsibility they actually bear. Team owners understand how public perception can affect their team's reputation, ticket sales, and merchandise sales, so they are quick to fire coaches to pacify the fanbase when the team is struggling, even if the coach is not truly to blame.

A prime example of this phenomenon is Gary Kubiak, former head coach of the Houston Texans. The Texans are the most recent organization to join the NFL, having been formed in 2002, and had not yet made even a single playoff appearance when Kubiak became their head coach in 2006. After a few more middling years, Kubiak and the Texans finally broke through in 2011, making the playoffs for the first time in franchise history, and repeated the accomplishment the following season in 2012. It seemed as though Kubiak had cemented himself as one of the best coaches in the league. However, the 2013 season saw a sharp fall from grace for the team. Despite opening the season with 2 come-from-behind victories, the Texans proceeded to drop their next 14 contests and finish the year with the worst record in the entire NFL. The most obvious explanation for this steep drop off was a sharp decline in quarterback play. Quarterbacks touch the ball on every offensive snap and are commonly regarded as the most important players on the field. Matt Schaub, the team's star quarterback, experienced severe regression in his play in 2013 following an excellent 2012 campaign, leading to his benching and a carousel of uninspiring options as the Texans shuffled between Schaub and his backups each week. Gary Kubiak was not even afforded the dignity of finishing out the season. He was fired with 3 games remaining. Coaches are fired every year in the NFL, but the vast majority of those firings happen at the end of the year. Kubiak was fired *midseason*, which is a particularly stinging insult. After all, if the season is lost anyway, how bad do you have to be for ownership to decide that they need to usher you out the door *right now*, as opposed to when the season ends?

This, however, was not the end of Gary Kubiak's story. After spending the 2014 season as an assistant coach for the Baltimore Ravens, in 2015 Kubiak received a second chance when the Denver Broncos tapped him to be their next head coach. In his first season with the Broncos, Kubiak led the team on a miracle run that culminated in a Super Bowl championship. In spite of a myriad of injuries and generally poor play at the quarterback position, Denver rode one of the most statistically dominant defenses in league history to a title. Needless to say, this was an embarrassing development for the Houston Texans, who have not experienced that same success since Kubiak's departure. In fact, to this day the Texans have never even been one of the final four teams still alive in the playoffs, let alone won a championship.

These events lead us to an important question: if Gary Kubiak was a good enough coach to lead a team to the pinnacle of success, why didn't the Texans see that in him? Why did they retain Rick Smith, the general manager who built the failing roster that tanked the 2013 season, but fire Kubiak, the head coach? Rick Smith eventually did leave the team after the Texans bottomed out again in 2017, but by then it was too late to undo the mistake they made with Kubiak.

This example brings me to my key questions for this paper:

-Why do head coaches get fired in the NFL? What criteria are they failing?
-Can we predict coach firings using publicly available information?
-Which head coaches fired in the past were undeservedly let go?
-Which current day coaches are on the brink of firing and which appear to be safe?

My primary goal with this project is to employ detailed statistical analysis to derive well-informed answers to each of these questions. Each of the four key questions will have their own subsection in the Methodology chapter, where I will go over in detail how I plan to use the data I gathered to answer them.

Sandia National Labs has requested that the project be analyzed through the context of “healthy” versus “degraded” signals. As an energy research company, Sandia spends significant resources analyzing aging hardware components to determine whether they are still working or whether they have become defective. Physically testing each component is expensive, both monetarily and temporally, so Sandia has made efforts to use machine learning to develop a method to determine automatically which components have become defective based on the signals outputs they produce.

Though an analysis of NFL head coaches is a far cry from an analysis of hardware components, the “healthy versus degraded signals” logic can be repurposed for this research as well. In the context of this paper, healthy signals will be given by head coaches who are either still actively coaching, or are no longer coaching but left their post *voluntarily*, whether for retirement, to accept another role, or for any other reason. The degraded signals will be given by head coaches who have been fired. I will examine all NFL head coaching tenures from 2005 to present day – a span of 20 years – and find which predictor variables are most closely related to head coach firing, and use that data to then answer the key questions. Any head coaching tenure that began before 2005 but continued until 2005 or beyond will be included in the data set as well. There is one caveat to this: the data set does not include any *interim* head coaches. These are coaches who entered the season in a different role and were thrust into the head coaching position midseason due to the original head coach being fired or leaving for another reason. Interim head coaches are appointed with the understanding that they will only be in charge until the end of the year, and an exhaustive search for the next coach will follow.

2) Data Explanation

Securing the data used for this project proved to be much more difficult than I anticipated. My goal is to work in the sports statistics industry, and thus in the past have used class projects as an opportunity to use what I have learned to explore trends in sports like football, baseball, and basketball. However, this is my first time analyzing *coaches* rather than *players*, and I quickly learned that statistics on coaches are far less robust, nor are they conveniently located in one place. The usual data sources like Kaggle did not turn up anything useful, nor was I able to find much from the scholarly articles I perused – I had hoped that one might use a data set for their analysis that might also be usable for mine. In the end, I decided that I was unlikely to stumble across a suitable data set, so I manually compiled my own. The bulk of the data came from Pro Football Reference at https://www.pro-football-reference.com/, but I also had to consult other websites, as well as some news articles, to round out the data set. Pro Football Reference has strict paywalls preventing users from downloading csv files without paying, so I transcribed most of the data myself. Below are the variables in the data set, with an explanation of their significance and where I sourced them from:

Name (*name*)- This is not a predictor variable. This variable captures the names of the NFL head coaches who were active from 2005 to 2024 for the purposes of identification. Fortunately, no two coaches from that time period have the same first and last name – this was a complication that arose a few times in previous projects where I conducted player analysis – so the name variable is sufficient to differentiate each coach. This variable was found on Pro Football Reference.

Team (*team*)- Once again, this is not a predictor variable. This variable captures the team each head coach worked for. It is important to capture this because there are a few coaches from 2005 to 2024 who coached for multiple teams during that time. Their names appear in the data set multiple times, so including the team in another column is necessary to keep their multiple tenures separate. Additionally, it is also possible that associations exist between the team variable and the probability of a coach being fired, as different teams have different owners with varying levels of patience, but I am uncertain whether there will be enough data points for each team to draw any definitive conclusions. The data for this variable was also found on Pro Football Reference.

Fired (*fired*)- This is the response variable. It separates our healthy signals from our degraded ones. Healthy signals (coaches who are still coaching *or* are not coaching but left voluntarily) have fired = 0, and degraded signals (coaches who have been fired) have fired = 1. Unfortunately, to find this data, I was forced to do something that I try to avoid when working on an analytics project – I had to consult Wikipedia. The reason for this is because Pro Football Reference only notes that a coach has been fired if they were fired *midseason*, but if they are fired at the end of a season, this is not recorded anywhere. For the purposes of this project, both midseason and end-of-year firings are degraded signals, so I had to find that data elsewhere or the project's findings would be unusable. Wikipedia had what I needed, and linked the original sources from where they found the information. For the most part, the information came from various news articles. Even though I effectively used those news articles as sources, indirectly through Wikipedia, I decided not to list them individually in the References section. As you will see, there are more than 150 coaching tenures in the data set, and linking every single one of Wikipedia's sources would have ballooned the References section to comprise more than 10 pages. When looking through the Wikipedia articles to determine if a coach had been fired or not I looked for language such as "fired," "dismissed," and "let go." In the vast majority of cases, it was very unambiguous whether or not a coach had been fired. I could find all the active head coaches on Pro Football Reference, and for the inactive coaches who left for reasons other than firing, Wikipedia clearly stated those reasons.

Fired midseason (*fired_midseason*)- This stat is exactly as it sounds – it is a binary variable that takes the value 1 if the coach was fired midseason and 0 otherwise. It cannot be used as a predictor because it is perfectly related to the response – if a coach is fired midseason then he must also have fired = 1. Instead, I am including this variable in case comparing the predictors to this variable yields different relationships than we find when comparing the predictors to the response variable. I am able to find this data on Pro Football Reference as the website notes when a coach has been fired before the end of the year.

Active (*active*)- This is a binary variable that takes the value 1 if the coach is still actively coaching for a team, and 0 otherwise. It is necessary to include this variable because not all of the coaches with fired = 0 are created equal. There are likely several active coaches who are on the proverbial "hot seat," being on the brink of firing, whereas the predictor variables likely paint very different pictures of the retired coaches who never got fired. The data for this variable is easily found on Pro Football Reference.

Tenure (*tenure*)- This is a discrete variable that captures the tenure length of each coach in years. I anticipate that we will observe a lower probability of firing as a coach gains more experience. Coaches who do not perform at a high enough standard are typically let go very early into their tenures. A coach with a longer tenure is less likely to get fired at all, as their tenure length is an indicator of their prowess. This data was also found on Pro Football Reference.

Winning percentage (*win_pct*)- This variable captures each coach's winning percentage during their tenure with a team. Coaches who have worked for multiple teams in our 2005-2024 window have their multiple tenures split into separate rows, so their winning percentage in each row only captures their record with that specific team. Though I hope to find more complex relationships than simply firing probability decreases as winning percentage increases, it would be ignorant not to acknowledge that on-field performance is possibly the single biggest factor in coach evaluation. Pro Football Reference again made finding this data very easy.

Final year winning percentage (*final_yr_win_pct*)- This variable captures each coach's winning percentage in their *final year with a team only*. There have been occasional instances where long-tenured coaches are let go after the team experiences a significant downturn in performance after a period of prolonged success. Their overall winning percentage over their tenure would likely still be high, and thus not accurately capture this downtick in the team's fortunes, so the final year winning percentage hopefully will help us parse these cases. Pro Football Reference provided the data here.

Super Bowl champion (*sb_champ*)- This is a binary variable that takes the value 1 if a coach has won a Super Bowl championship in their tenure with a team and 0 otherwise. Winning the Super Bowl is each team's ultimate goal, so we would expect that winning one would buy a coach a lot of job security. Pro Football Reference notes clearly for us when a coach has won a championship with a team.

Runner up (*runner_up*)- This stat is a little bit convoluted, so I will do my best to explain clearly. Every year, the 2 teams that win their respective conferences compete against each other for the Super Bowl championship. This variable is to capture which coaches won at least one conference championship but did not win any Super Bowls. In other words, this variable is looking for the second place finishers. Here is how the values will be assigned:

-Coach has won the conference at least once, but never the Super Bowl → *runner_up* = 1
-Coach has never won the conference → *runner_up* = 0
-Coach has won the conference at least once, but has also won at least one Super Bowl → *runner_up* = 0

This means that *runner_up* and *sb_champ* are mutually exclusive. A coach can be one, the other, or neither, but never both. Pro Football Reference notes these instances clearly for us.

Scandal (*scandal*)- I was unsure whether or not to include this variable, as it is somewhat subjective. Ultimately, I decided to include it. Scandal is a binary variable with the value 1 if a coach has been embroiled in a scandal during their tenure, and 0 otherwise. What constitutes a scandal? Unfortunately, this is where the subjectivity comes into play. It is inarguable that in certain cases a coach's off-field problems contributed to their firing, and may have even been the primary cause. For example, head coach Mike Tice was fired after the 2005 season by the Minnesota Vikings, despite the team playing to a very respectable 9-7 record, after a raucous boat party in October led to 4 players on the team being charged [10]. There are also other cases of that nature, and I felt that leaving that context out of the data set would be to leave out crucial information. However, what issues were worthy of being considered a scandal and what were not were ultimately up to my discretion. I researched news articles on each coach and gave coaches a 1 for this variable if anything came up that could realistically have caused or contributed to a firing down the line, *even if they did not ultimately end up being fired*. Any news article I found about a coach that caused me to give them a 1 is included in the references section. Several different sources contributed to the data gathering for this variable.

Minority (*minority*)- This variable takes the value 1 if the coach belongs to a racial minority, and 0 if they are Caucasian. For the purposes of this project, a coach who is biracial or multiracial will still be considered a minority, even if they are partially Caucasian. I chose to include this variable because of research I came across that concluded that NFL teams exhibit different hiring and firing patterns for Caucasian head coaches versus minority head coaches. If curious, the aforementioned studies are linked in the references section: *Differences in the Success of NFL Coaches by Race, 1990-2002: Evidence of Last Hire, First Fire* by J.F. Madden [6], and *Differences in the Success of NFL Coaches by Race: A Different Perspective* by Malone, Couch, and Barrett [7]. I was able to find demographic information for the majority of coaches by examining Lapchick's *The 2021 Racial and Gender Report Card National Football League* [8]. For the few coaches who were not listed in the paper – the few who began coaching in 2022 or later – I resorted again to Wikipedia, which filled in the gaps.

Pro Bowl QB (*pro_bowl_qb*)- As explained in the background section, quarterbacks are the players who have the most influence over the game. They touch the ball on every offensive play and are responsible for making many executive decisions for the other players. It is rare for teams to have prolonged success without having good quarterback play. The stat Pro Bowl QB is a proxy for that. The Pro Bowl is the NFL's equivalent of an all star game, so, in theory, the quarterbacks who are selected to play in the pro bowl in a given year are the best quarterbacks that season. Since a coach's success is rarely independent of the quality of his quarterback, I thought it necessary to take QB ability into account. This is a binary variable with the value 1 if the team's quarterback made the pro bowl during the *final year of the coach's tenure*, and 0 if not. Pro Bowl appearances are noted clearly on Pro Football Reference.

First tenure (*first_tenure*)- This is a binary variable that takes the value 1 if a head coach has never served in that role before their current tenure, and 0 if they have previously been a head coach on another team. Some team owners may prefer to take chances on young, up-and-coming coaches who have never held the top job before, whereas other owners may prefer to stick with the tried-and-true experienced hires. I wanted to account for this in case teams behave differently in firing decisions for first-time head coaches versus experienced head coaches. The data for this variable was also found on Pro Football Reference.

Coach of the Year (*coty*)- This is a binary variable that takes the value 1 if a head coach has won the Associated Press' NFL Coach of the Year award at any point in their tenure, and 0 if not. Similar to how winning a Super Bowl championship is the pinnacle of success for a team, winning a Coach of the Year award is the height of individual recognition for a coach. Pro Football Reference has a page listing all the winners of this award.

3) Exploratory Data Analysis

Now that I have walked through each of the variables, the logic behind including them, and where I found them, I would like to perform some data exploration to investigate if any early trends emerge. Let's begin with some basic summary statistics.

The data set has dimensions of 167x14. We went over the 14 variables in detail above. From 2005 to 2024 there have been 167 separate coaching tenures in the NFL but only 130 *unique* coaches. This discrepancy arises because many coaches have coached for multiple teams over this span of time. There are also some coaches who only have 1 tenure recorded in the data set but still have first_tenure = 0 because they previously coached for another team before 2005 – before our window of interest.

Next, I find that 119 of the 167 (71.26%) tenures analyzed in the data set ended in firing. At first glance, it seems as though NFL head coaches are roughly 2.5 times more likely to get fired than they are to leave their post in any other way.

Then, I began testing to see if any trends emerged. I first created contingency tables between *fired* and all of the other binary predictors, and then fed those contingency tables into a chi-squared test for independence to see if they were related in a statistically significant way. Below are the results:

Table 1: Chi-squared Independence Test Results; *fired* vs Other Binary Variables

Variable	Chi-squared value	p-value
sb_champ	12.17	0.0005
runner_up	0.83	0.3625
scandal	1.61	0.2042
minority	0.06	0.8140
pro_bowl_qb	8.08	0.0045
first_tenure	0.26	0.6111
coty	7.36	0.0067

There are some interesting takeaways right away. To begin, we see that at the alpha = 0.05 level, *fired* has statistically significant relationships with *sb_champ*, *pro_bowl_qb*, and *coty*. Frankly the first one is expected, as it represents one of the most significant achievements a coach can attain. Intuitively, you would expect coaches experiencing that level of success to earn a lot of job security with their teams. I was interested however, to see that *pro_bowl_qb* has a statistically significant relationship as well. It is generally understood that teams with good quarterbacks have a better chance to win, and with winning comes job security for coaches, but even so I was not expecting to see such a significant association. Lastly, *fired* and *coty* having a significant relationship is not a surprise either. *coty* is the variable denoting whether or not a coach has won a Coach of the Year award – winning one typically means a coach is very capable and would be expected to earn them some job security.

The variables for which we do *not* observe a statistically significant relationship with *fired* are interesting as well. These variables are *runner_up*, *scandal*, *minority*, and *first_tenure*. Looking at the contingency tables, we can see that 6/18 (33.3%) coaching tenures where the coach won a Super Bowl championship ended in firing, but that figure balloons to 8/14 (57.1%) for coaches who have finished as runner-ups. Evidently, finishing in second place does not buy you nearly the same job security as winning a championship does. Next, to be completely transparent, I was disappointed not to find a noteworthy relationship between *fired* and *scandal*, as I had spent a great deal of time researching for this data. In the end, however, only *seven* of the 167 rows in the data set had *scandal* = 1. Out of those 7 tenures, 3 ended in firing, while 4 did not. It makes sense given this information that a statistically significant association could not be found. With only 4.2% of the data points having *scandal* = 1, and those few data points being split 3 to 4 in either direction, there simply was not enough data to derive anything useful from *scandal*. The results for *first_tenure* were interesting as well, as we found no significant difference in firing patterns between first-time head coaches and veteran head coaches (defined as a coach who previously served as the head coach of another team). Finally, *minority* was the variable I was most interested in. Despite the research I found that concluded that hiring and firing patterns are markedly different depending on the race of the coach, I was not able to confirm this with this quick test, as I found no statistically significant relationship between *fired* and *minority*. I went back to look at the contingency table, and found that out of 35 tenures by minority head coaches, 26 (74.3%) ended in firing, whereas for Caucasian head coaches that figure was 93 out of 132 (70.5%). Evidently this was not a large enough gap to suggest a strong association.

Out of curiosity, I ran another test, but this time I compared *minority* to *fired_midseason*, rather than *fired*. Even if coaches get fired at a similar rate regardless of race, could there be a difference in the *timings*? The results were inconclusive on this as well. The p-value was a lot lower at 0.1990, but still did not suggest a statistically significant relationship at the alpha = 0.05 level. From the contingency table, I found that 5 out of 35 (14.3%) minority head coaching tenures ended in a midseason firing, whereas for Caucasian head coaches that figure was 35 out of 132 (26.5%). If anything, this loosely suggests that minority head coaches are actually *less likely* to suffer the indignity of a midseason firing than their Caucasian counterparts. Given the discourse around the NFL's diversity practices, I expected to see the opposite.

Now, let's move on to the numeric variables. To compare *fired*, a binary variable, to the numeric predictors, I used point-biserial correlation. The sign of the correlation figure will tell us whether *fired* and the predictor are positively or negatively associated, the magnitude will tell us the strength of the relationship, and the p-value will tell us whether the relationship is statistically significant. Below are the results:

Table 2: Point-biserial Correlation; *fired* vs Numeric Variables

Variable	Correlation	p-value
tenure	-0.24	0.0015
win_pct	-0.46	0.0000
final_yr_win_pct	-0.55	0.0000

These results are a lot more straightforward than the ones we observed when comparing *fired* to the other binary variables. All 3 numeric variables in our data set are found to have weak to moderate negative associations with *fired*, and all 3 relationships are statistically significant. Intuitively this makes sense. Better coaches get longer tenures, and better coaches are less likely to be fired. Better coaches win more games, and better coaches are less likely to be fired.

There is one concern that I have about these variables: possible multicollinearity between *win_pct* and *final_yr_win_pct*. I am worried about this because *final_yr_win_pct* is already indirectly captured in a coach's overall winning percentage. I ran a quick Pearson correlation test between these two variables and received a correlation value of 0.664. Typically I would use 0.7 as a cutoff for multicollinearity, but in a case like this I think it would be prudent to select one variable and leave the other out. Even though *final_yr_win_pct* seems to have a stronger relationship with *fired* – having a larger absolute value for the correlation (0.55) than *win_pct* (0.46) – *win_pct* covers a coach's entire tenure and provides more information, so I will use *win_pct* going forward and discard *final_yr_win_pct*.

Thus, as we move towards the methodology section, where we will discuss modeling methods, the variables still in play are *sb_champ*, *pro_bowl_qb*, *coty*, *tenure*, and *win_pct*.

I also double checked any potential multicollinearity by testing the VIF for each predictor variable. VIF should tell us how much the variance of a regression coefficient is affected by multicollinearity with other variables. I generally want to see VIF < 5 for my predictors, and if the VIF > 10 then I would seriously consider leaving it out of the models. Fortunately, none of the predictors produced a VIF above 3, so we can comfortably continue forwards with the 5 predictor variables mentioned above.

4) Methodology

Since there are four key questions we have to answer, I am going to discuss them one by one and walk through how I plan to tackle them.

Why do head coaches get fired in the NFL? What criteria are they failing?

We know from data exploration that the variables with the strongest associations with *fired* are *sb_champ*, *pro_bowl_qb*, *coty*, *tenure*, and *win_pct*. However, we can use modeling approaches to take this even further. There are two methods I would like to try.

First, I will run a logistic regression to predict firing probability. Important note: *the model will only be trained on coaches who are not currently active – active coaches have incomplete information, as a large percentage of them will eventually be fired, but all have fired = 0 for the time being*. The coefficients of the predictors for this model will provide us a lot of information. The signs of the coefficients will tell us whether they are directly or inversely related to firing probability, and the magnitudes will tell us how much influence they have.

Second, I will run a random forest, using Gini Impurity when training the model, and then use permutation importance to determine how relevant each predictor is when attempting to predict firing probability. This model will also be trained using only the coaches who are no longer active. I anticipate that results will be similar to that of the logistic regression, but if not that will provide interesting discussion points for the Results and Analysis section.

Can we predict coach firings using publicly available information?

Answering this question will build off of both of the models created in the previous part. For each of the logistic regression and random forest models, I will test both the cross-validated accuracy as well as the cross-validated AUC. These are two slightly different tests, so I feel that it would be useful to do both. Testing the raw accuracy will tell me how accurate the models are given a specified probability threshold, and testing the AUC will tell me how accurate the models are regardless of threshold.

However, it is important to keep in mind that these tests cannot tell me *everything*. If one or both models score well, I can reasonably say that the high performing model(s), when given these variables as inputs, are good at predicting coach firings. It would also follow logically that yes, we can predict coach firings reasonably accurately with publicly available information. But, if the models score poorly, I cannot necessarily say the opposite. All I would be able to say definitively is that the chosen models do not work well given the data I provided – there is plenty more publicly available information that I chose not to include. It is possible that any number of those features that were left out would be able to improve model accuracy. There is also a chance that further tuning model parameters would improve predictive power, as well as potentially using different models altogether.

Thus, in summary, if one or more models perform well, I can conclude that coach firings are predictable with publicly available information, but if the models perform badly, I still cannot conclude with certainty that coach firings are completely *un*predictable. In addition, there are very likely many factors that go into these decisions that we cannot observe from the outside looking in – a coach's command of the locker room, their play design prowess, how well they delegate labor to their assistants, etc. – but I could not say with certainty that this sort of proprietary information is necessary to build a successful model.

Which head coaches fired in the past were undeservedly let go?

This question can also be answered by working with the previously created models. I will use both the logistic regression and random forest models to find which past coaches were fired despite having a very low predicted probability of firing. Any coaches who are ranked highly by both models will be particularly noteworthy, as those are coaches whose firings were very unexpected.

Model output can also be combined with other qualitative factors to try and explain why these firings occurred. For instance, in the data explanation section, when going over the *scandal* feature, I provided an example of Vikings coach Mike Tice, who was fired despite generally good performance after a boat party resulted in 4 players being charged. If he comes up as a "surprise firing" on one or both of these models, we will know why. Other surprise firings will need to be investigated as well to determine if they are truly shocking firings, or whether there is a clear explanation that the models are not picking up.

Which current coaches are on the brink of firing and which appear to be safe?

I plan to tackle this question by using the created models to predict on the active coaches, whose data will not be used to train the models. Since random forest has a risk of overfitting, I am hoping to avoid using the model to predict on data that it has already been trained on – doing so may give us an overly optimistic estimation of the model's predictive ability. Active coaches assigned a high firing probability can be assumed to be on the "hot seat," while coaches with a low firing probability are likely safe.

Another strategy I considered for this question, but decided not to move forward with, was survival analysis. Time-to-event modeling would have added an interesting wrinkle. Not only would I have been able to predict if a coach would eventually be fired, but I could also estimate when it would happen. The roadblock I ran into however, that eventually caused me to abandon this idea, was that I would need season-by-season data for every coach as opposed to just 1 row of data for each. As I collected most of this data manually, breaking every coach out into 3, 5, 10, or even 15 rows, depending on how long they coached, would have been a prohibitively time expensive exercise.

Having season-by-season data also would have allowed us to look for some other trends. For example, even if a coach had a healthy winning percentage overall during his tenure, if the season-by-season data showed that the team was getting worse, rather than improving, before his firing, that is context that only season-by-season data could capture. That type of information is lost in the data that I am using. It pained me to leave this avenue unexplored, but I had to be realistic about how much work I would be able to get done as just a 1-person group.

5) Results and Analysis

Similarly to the Methodology section, I will split the results into four sections, one for each of the key questions, and discuss them separately. To some degree the analysis for a previous question should flow smoothly into the next one, so I feel that this will be a logical way to present my findings.

Why do head coaches get fired in the NFL? What criteria are they failing?

The results indicate that both the logistic regression and random forest are in agreement here. I will present the output first, and then analyze it step-by-step:

Table 3: Logistic Regression Output

Variable	Coefficient	Odds ratio	Note
sb_champ	-0.179	0.836	Binary
pro_bowl_qb	-0.019	0.982	Binary
coty	-0.090	0.914	Binary
tenure	-0.106	0.899	Numeric
win_pct	-1.171	0.310	Numeric

I will discuss how to interpret this output carefully, as there is a different interpretation for the binary variables versus the numeric ones.

Binary Variables

For binary variables, we are finding the change that occurs when toggling the switch from 0 to 1 for each variable. We know that each of them are inversely related to firing probability, because each of the coefficients are negative. However, this is still not in a very interpretable form. First, we can find the odds ratio by exponentiating the coefficient. From there, we can find the *percent change in the predicted response* when the switch is toggled from 0 to 1 for each variable by using the formula (*odds_ratio* - 1) * 100. Doing so gives us the following output:

-Changing *sb_champ* = 0 to *sb_champ* = 1 leads to a -16.4% change in predicted firing probability

-Changing *pro_bowl_qb* = 0 to *pro_bowl_qb* = 1 leads to a -1.8% change in predicted firing probability

-Changing *coty* = 0 to *coty* = 1 leads to a -8.6% change in predicted firing probability

This seems to indicate that among the binary variables, *sb_champ* is the most important when it comes to affecting a coach's firing probability, followed by *coty* in the middle, and *pro_bowl_qb* bringing up the rear.

Numeric Variables

For the numeric variables, *ordinarily* the results would tell us how much a 1 unit shift in the feature variable affected the prediction. However, because *tenure* and *win_pct* are on different scales, where *tenure* can grow infinitely large whereas *win_pct* must be between 0 and 1, I standardized the two variables to both have mean 0 and standard deviation 1. Standardization changes the interpretations. With the variables now standardized, the output actually tells us how the prediction changes for a 1 *standard deviation* change in the feature variable. I took a moment to go back to the original data set and calculate the standard deviation for each of the 2 numeric predictor variables, and then using the same formulas laid out above to derive the percentage change, I received the following output:

-Increasing *tenure* by 1 standard deviation – 3.72 years – decreases predicted firing probability by 10.1%

-Increasing *win_pct* by 1 standard deviation – 0.145 – decreases predicted firing probability by 69.0%

These results show us that *win_pct* dwarfs *tenure* in terms of impact. Both intuitively make sense. The output for *win_pct* tells us that the more a coach wins, the likelier they are to survive. That almost goes without saying. The output for *tenure* tells us that the longer a coach has already survived, the more likely they are to continue surviving. This also follows logically. However, the percentage changes in the output tell us that *win_pct* has a much greater impact than *tenure*. Both have a relationship with the response, but not of equal magnitude.

Now, we know that the ranking for the binary variables is *sb_champ* first, *coty* second, and *pro_bowl_qb* last. We also know that for the numeric variables, *win_pct* is first and *tenure* is second. Where we run into a little bit of a roadblock is how to compare the binary variables to the numeric ones. Recall that the results for the binary variables tell us the expected change in the response when each predictor is toggled from 0 to 1, whereas for the numeric variables the results tell us the expected change in the response when each predictor is increased by 1 standard deviation. These 2 things cannot be compared 1 to 1, and that is where the random forest model comes into play. Using permutation importance, I am able to compare each of the predictor variables directly, regardless of whether they are binary or numeric. Below are the results:

Table 4: Random Forest Permutation Importance

Variable	Mean	Standard Deviation
win_pct	0.1063	0.0163
tenure	0.0169	0.0082
coty	0.0065	0.0031
pro_bowl_qb	-0.0046	0.0041
sb_champ	-0.0055	0.0032

Permutation importance is a fascinating measurement of feature importance, as it calculates how the model's accuracy changes when the value of each feature is randomly shuffled across all the rows, while keeping everything else the same. The table is ordered with the most important feature listed first, and descending to the least important feature. From this, we can see that *win_pct* was by far the most important variable. Over the 50 repetitions that were performed to gather this data, randomly shuffling *win_pct* reduced the random forest's accuracy by 10.63% on average, with a standard deviation of 1.63%.

With this context we see that the other variables pale in comparison to *win_pct* in terms of impact on the prediction. *tenure* is a distant second at 1.69% performance drop on average, while none of the other three variables finish above 1%. In fact, two variables, *sb_champ* and *pro_bowl_qb*, actually *increase* the accuracy of the model when they are shuffled. This means that for the random forest, *sb_champ* and *pro_bowl_qb* just served to add random noise. They did not help performance at all, and only served as a distraction from the more important predictors.

sb_champ being in this group of unimportant predictors is actually quite shocking, as it represents a major disagreement between the logistic regression and the random forest. I am inclined to side with the logistic regression in this dispute. As a reminder, *sb_champ* is a binary variable that tracks whether or not a coach won a championship! The idea that winning a championship would have no bearing on whether or not a coach is eventually fired seems absurd.

So why did *sb_champ* come up as such an unimportant variable in the random forest? I cannot say with 100% certainty, but I can offer a plausible theory. When I look back over the data, I realize that Super Bowl champion head coaches are *very* rare. This makes sense, as there simply are not very many championships to go around. Only one can be won each year.

This issue is further exacerbated by the fact that I did not train the model on any active coaches. There are several active head coaches who have Super Bowl trophies on their mantles, and none of them were used for model training. Out of the 142 rows of data that *were* used for training, only 13 had *sb_champ* = 1. That is a paltry rate of just 9.2%. It is entirely possible that the effect of the *sb_champ* variable dissipated in the random forest just because of how sparsely used it is.

There is also a second problem, which is *reverse survivorship bias*. Intuitively, it makes sense that coaches who have won championships are likely the best of the best, so they are more likely to have longer coaching tenures and more likely to eventually voluntarily leave rather than get fired. Thus, Super Bowl winning coaches are more likely to still be active than ones who have not won a championship. When I filter the data, this is exactly what I observe. Only 13 of the 142 inactive coaching tenures won a championship, a paltry 9.2% rate, but 5 of the 25 active coaching tenures have won a title, ballooning that rate to 20%. In other words, the way I decided to train the models inadvertently filtered out a disproportionate amount of the high achievers from the training data. My logic at the time was the active coaches had incomplete data, as many will eventually be fired but currently have *fired* = 0, but this approach was clearly flawed as well.

Zooming out, I see the exact same phenomenon with *coty* and *pro_bowl_qb*. 14/142 (9.9%) inactive head coaching tenures won a Coach of the Year award, but that rate is 5/25 (20%) for active head coaching tenures. For *pro_bowl_qb*, the rate is 18/142 (12.7%) for inactive head coaching tenures, and 8/25 (32%) for active head coaching tenures.

In summary, it feels reasonable to say that both models agree that *win_pct* is the most important variable for these predictions, followed by *tenure*, and after that they begin to clash with each other. I am of the opinion that, because of a flaw in my approach, *sb_champ*, *coty*, and *pro_bowl_qb* are more important than they are given credit for in the models – with *sb_champ* and *coty* possibly being even more important than *tenure*. We observed significant relationships between those 3 variables and the response, *fired*, when doing data exploration using the full data set, and those relationships fizzled out when we trained the models using only the inactive coaching tenures. Now, let us move on to key questions 2 through 4.

Can we predict coach firings using publicly available information?

The short answer to this question is yes, but with a caveat. I evaluated both the logistic regression and random forest models using 10-fold cross-validation and recorded both their accuracy and AUC. The cross-validation was done to ensure that they were not being tested on the same data points that they had trained on. For the random forest in particular, I also used a mesh grid to tune hyperparameters, specifically number of trees, maximum depth, and minimum numbers of nodes per leaf. The best configuration came back as 100 trees, no maximum depth, and minimum 4 nodes per leaf. Below are the accuracy and AUC results for each model.

Logistic regression accuracy: 0.845
Logistic regression AUC: 0.798

Random forest accuracy: 0.880
Random forest AUC: 0.845

The random forest edges out the logistic regression model in both fields by a small margin, but overall it seems as though both perform relatively well. The accuracy tells us what percentage of the predictions they got correct, and the AUC tells us how good the models are at separating the two classes, regardless of threshold used. The accuracy numbers seem good at first glance, and the AUC numbers seem positive as well given the context that AUC = 1 means perfect discrimination, and AUC = 0.5 is akin to random guessing. Given these results, it appears reasonable to say that we can fairly accurately predict coach firings using just publicly available data.

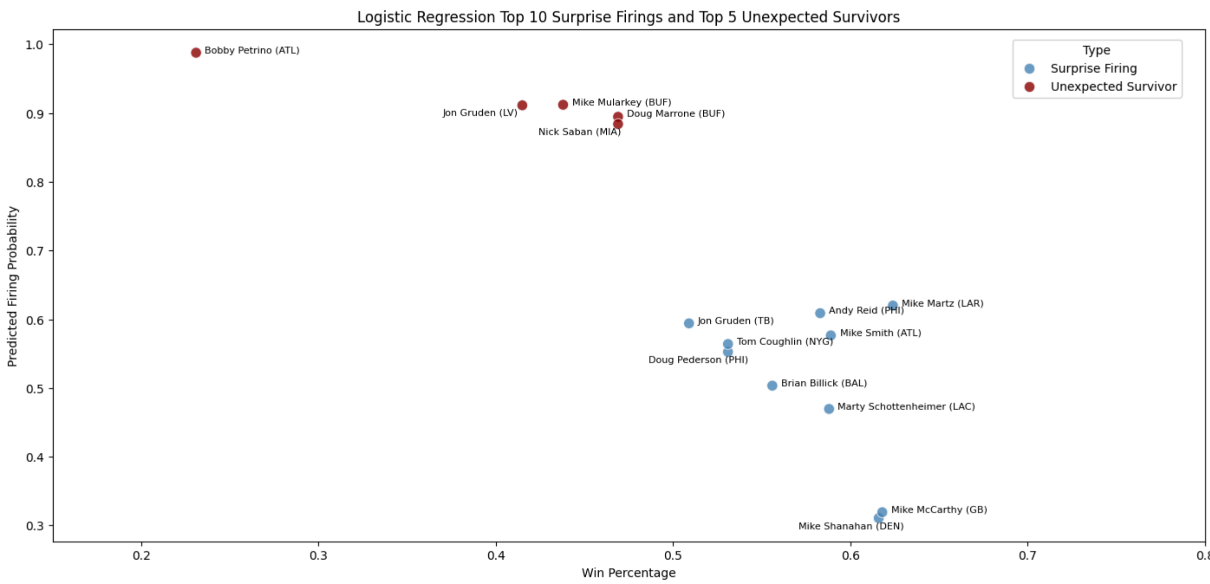
However, this is where the caveat comes in. The data is *imbalanced*, and I did not take this into account when building my strategy. 119 of the 142 inactive head coaches (83.8%) ended up being fired. This means that a dummy classifier that always guesses "fired" would be correct 83.8% of the time. With this context, the 84.5% accuracy from the logistic regression and 88% accuracy from the random forest do not seem as impressive. The high AUC indicates that the models truly are good at learning patterns to differentiate between the two classes, but this does not result in substantially better performance than a dummy classifier.

One way I could have potentially improved the models is by tuning the prediction threshold. By default, this threshold is set at 0.5 – if above 50% probability, predict "fired," if below, predict "not fired" – but given how many more fired coaches there were than safe coaches, it may have made sense to experiment with lowering the threshold, perhaps to 0.4, 0.3, or even lower. The models should probably tend towards predicting fired, even when the probability appears to be below 50%, simply because fired coaches make up an overwhelming majority. Therefore, I feel a reasonable conclusion is that we can predict coach firings with good accuracy using publicly available data, but the models still have room for improvement. Let us move on to key question 3.

Which head coaches fired in the past were undeservedly let go?

For the sake of variety, I decided to plot my answers to this question rather than put them in another table. I used the logistic regression and random forest models to predict on each of the inactive coaches to figure out not only which were undeservedly let go (fired = 1 with low firing probability), but also which coaches survived when they likely should have been fired (fired = 0 with high firing probability). In order to make the predictions, I used the *cross_val_predict* function from *scikit-learn* to make out-of-bag estimates on each of the points. It would not have made sense to train the model on the same points on which I was later predicting. I then plotted the points with predicted firing probability on the y-axis and *win_pct* on the x-axis, and I color-coded the points to distinguish between surprise firings and unexpected survivors. Below are the results for logistic regression:

Figure 1: Logistic Regression Top 10 Surprise Firings and Top 5 Unexpected Survivors

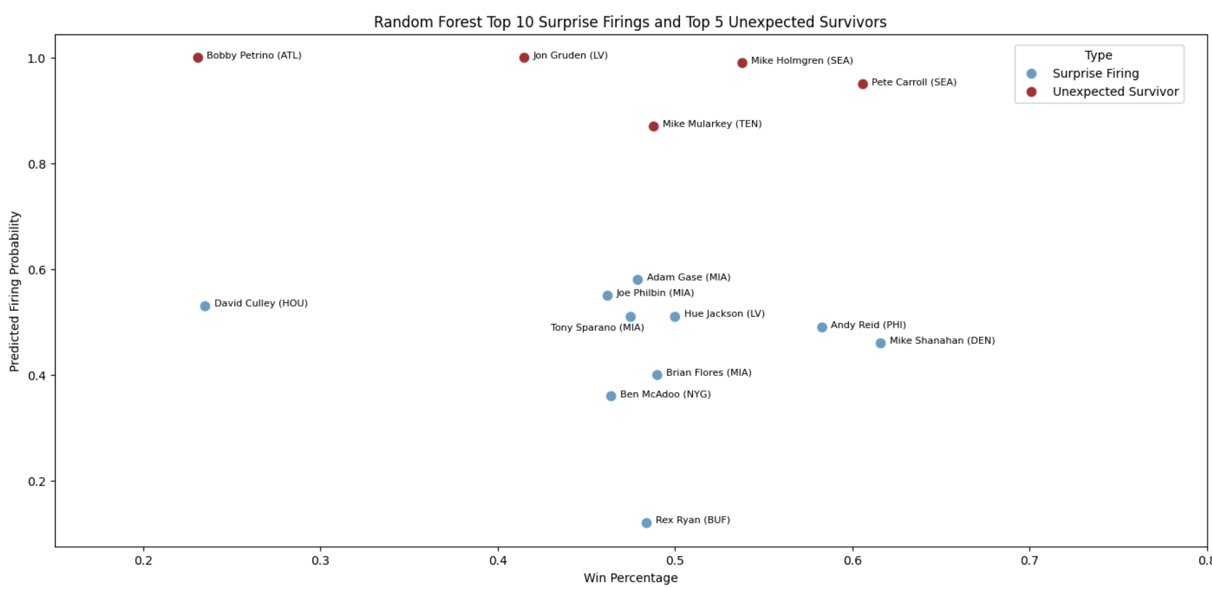


The logistic regression graph shows us roughly what we expect to see. We observe a clear trend of predicted firing probability decreasing as win percentage increases, leaving the unexpected survivors at the top left and the surprise firings at the bottom right. There are certainly some interesting names here. Bobby Petrino, the lone dot by himself in the corner, is a special case. He almost certainly would have been fired if he did not voluntarily quit before the season ended. His Falcons were among the worst teams in the league, and reporters had heard grumblings from players about his lackluster leadership skills. Knowing he was likely to be fired when the season ended, Petrino went job hunting early, and left for the head coaching position at the University of Arkansas with 3 games still to go for the Falcons [11]. Petrino did not even remain long enough to let his players know in person that he was leaving – all they received from him was a note. Another interesting unexpected survivor is Nick Saban, former coach of the Miami Dolphins. Saban is living proof of how different of a skill set it requires to coach successfully at the NFL ranks versus the college ranks. Before joining the Dolphins, Nick Saban was the head coach at Louisiana State University, where he won a national championship. He parlayed that success into a lucrative NFL job with the Dolphins, but after 2 mediocre seasons he resigned to return to the college ranks and took the head job at the University of Alabama. 20 years later, Nick Saban is known by many as the greatest coach in college football history, having won six more championships with Alabama – in addition to the one he won earlier at LSU – and his mediocre NFL tenure is largely forgotten.

The surprise firings are no less interesting. In the motivation section, I gave the example of Gary Kubiak, a coach fired by the Houston Texans who latched on with another team, the

Denver Broncos, and took them all the way to a Super Bowl title. Here we observe an even more *extreme* case in this vein. Andy Reid was fired by the Philadelphia Eagles following the 2012 season, after a string of mediocre years. He was immediately hired by the Kansas City Chiefs, where he has since won *three* Super Bowl championships and cemented his legacy as one of the greatest coaches in NFL history. Tom Coughlin and Mike Shanahan are also tragic cases, as both were let go by teams whom they led to not one, but *two* Super Bowl championships. Compelling arguments can be made that the New York Giants and Denver Broncos fired the greatest coaches in the history of the teams by letting go of Coughlin and Shanahan respectively. Finally, Marty Schottenheimer is a very curious case. He was fired after losing a power struggle with the team's general manager, following a season where the Chargers won 14 games and were the best team in the NFL. In the 18 years since Schottenheimer's firing, the Chargers have never again had a season that successful, and several NFL teams have never won 14 games in a season in their entire histories. Now, let's move on to the random forest and see where this model is similar and where it is different.

Figure 2: Random Forest Top 10 Surprise Firings and Top 5 Unexpected Survivors



Immediately, we can see that this plot looks very different from the one we analyzed above. Despite the random forest's feature importance rankings placing *win_pct* at the top, we do not see a clear trend of *win_pct* lowering predicted firing probability, as we did with the logistic regression plot. With this oddity comes some strange projections. In the unexpected survivors section, we see former Seahawks coach Pete Carroll. Looking at the win percentages, you can see that he has the 2nd best winning percentage of any coach on this chart, and yet, the model has assigned him greater than a 90% chance of firing. This seems to be in part because the random forest model does not put any weight into the *sb_champ* stat, as we discussed earlier. Pete Carroll led the Seahawks to a championship win in 2013, but in the eyes of the random forest model, winning a championship does not buy you any job security. This is clearly a spot where the model deviates from how teams behave in real life. In reality, Pete Carroll coached the Seahawks for another 10 seasons after winning the championship. Some of those seasons were quite mediocre, and it is hard to believe that Carroll could have retained his job for that long without winning the Super Bowl earlier.

Funnily enough, we also see Bobby Petrino on this graph. Despite the differences between the two models, they agree wholeheartedly on how bad Bobby Petrino was for the Atlanta Falcons. Both models assigned him the highest firing probability of any coach in the data set. Another unexpected survivor that appears on both charts is Mike Mularkey, but this is where things get a little bit strange. Mike Mularkey coached in short stints for 3 different teams in the 2005-2024 window that we are studying, and the logistic regression and random forest models each graphed a *different team that Mularkey coached for*. The logistic regression ranked Mularkey's tenure with the Buffalo Bills as an unexpected survivor, where the random forest charted Mularkey's tenure with the Tennessee Titans. Somehow, Mularkey evaded firing in *two separate instances* when the models believed he should have been axed.

Moving onto the surprise firings, we see Andy Reid and Mike Shanahan as agreements between the two models. Another interesting note is that former Miami Dolphins coaches make up *four* of the top 10 most surprising firings. Current owner Stephen Ross purchased the Dolphins in 2008. Tony Sparano, Joe Philbin, Adam Gase, and Brian Flores, all of whom registered on the random forest chart as surprise firings, were, in order, the *first four Dolphins head coaches who served under Ross*. Mike McDaniel, the current Dolphins coach, is the fifth one Ross has hired. That means every head coach firing he has ever made has registered as a surprise firing, according to the random forest. This points to Ross as an extremely impatient and aggressive owner. As far as we can tell, Ross and the Dolphins have not been rewarded for this behavior either. The Dolphins have not been particularly good, and have not even reached the 2nd round of the playoffs since Ross took over. Teams certainly should not hold on to coaches who are doing a poor job, but on multiple occasions the Dolphins have let go of coaches during periods of respectability for the team – there is an argument to be made that maintaining staff continuity could have helped them.

Which current coaches are on the brink of firing and which appear to be safe?

As explained in the methodology section, to answer this question I used the models trained on *all* the inactive coach data to predict on all the head coaches who are currently active. There is a concern about reverse survivorship bias, as discussed earlier, as the models were trained on data that disproportionately filtered out higher achieving coaches. The makeup of the data between inactive and active head coaches is fundamentally different, so the models' predictions on the active coaches should not be trusted completely at face value. With that said, the results for both models are below:

Table 5: Logistic Regression Top 5 Most At-Risk Coaches and Top 5 Safest Coaches

Coach	Team	Predicted Firing Probability	Winning percentage	Tenure length (years)
Brian Callahan	Tennessee Titans	0.99	0.18	1
Dave Canales	Carolina Panthers	0.97	0.29	1
Jonathan Gannon	Arizona Cardinals	0.95	0.35	2
Brian Daboll	New York Giants	0.92	0.36	3
Raheem Morris	Atlanta Falcons	0.88	0.47	1
Andy Reid	Kansas City Chiefs	0.26	0.73	12
John Harbaugh	Baltimore Ravens	0.34	0.62	17
Nick Sirianni	Philadelphia Eagles	0.35	0.71	4
Mike Tomlin	Pittsburgh Steelers	0.39	0.63	18
Sean McVay	Los Angeles Rams	0.44	0.61	8

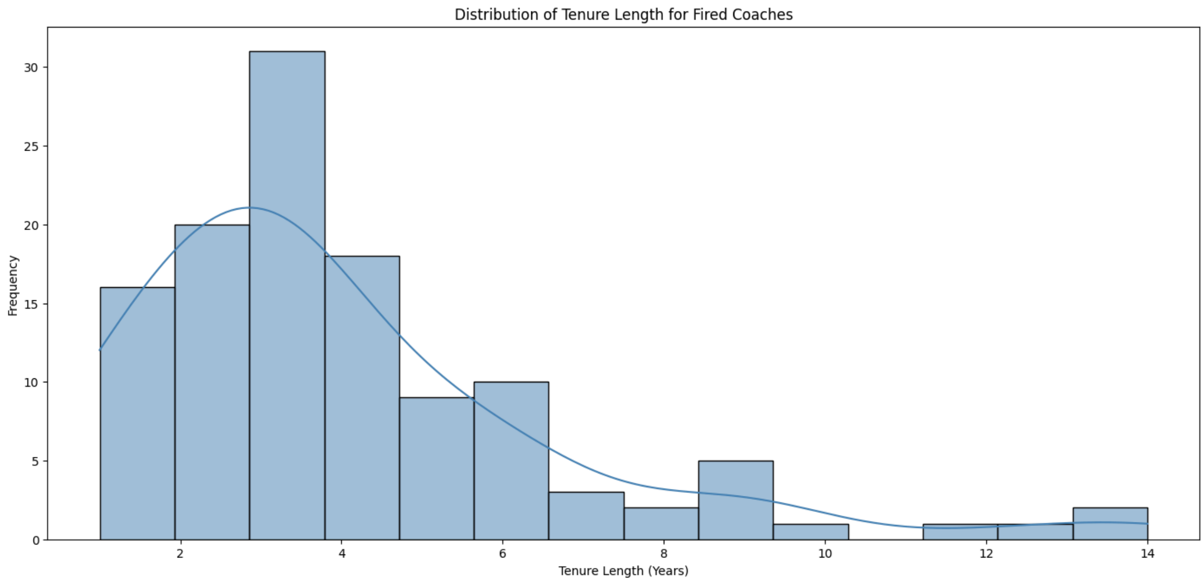
Table 6: Random Forest Top 5 Most At-Risk Coaches and Top 5 Safest Coaches

Coach	Team	Predicted Firing Probability	Winning percentage	Tenure length (years)
Jonathan Gannon	Arizona Cardinals	0.99	0.35	2
Dave Canales	Carolina Panthers	0.94	0.29	1
Brian Callahan	Tennessee Titans	0.94	0.18	1
Brian Daboll	New York Giants	0.94	0.36	3
Kyle Shanahan	San Francisco 49ers	0.91	0.53	8
Mike Tomlin	Pittsburgh Steelers	0.17	0.63	18
John Harbaugh	Baltimore Ravens	0.19	0.62	17
Kevin O'Connell	Minnesota Vikings	0.26	0.67	3
Nick Sirianni	Philadelphia Eagles	0.27	0.71	4
Andy Reid	Kansas City Chiefs	0.29	0.73	12

My first takeaway here is that with way fewer names to choose from – 25 active coaching tenures as opposed to 142 inactive ones – the logistic regression and random forest are now way more aligned. For both the at-risk coaches and the safe coaches they overlapped on 4 of the 5 names.

I also notice that among the at-risk coaches, 50% of the names between the two models are coaches who just finished their first seasons. This is not reflective of how teams actually behave. First year head coaches are typically walking into bad situations – most of the time the previous coach was fired, and the previous coach would not have been fired if the team was performing well. This means that these new head coaches are taking over teams lacking in talent, and fixing that in just one offseason can be very difficult. In acknowledgement of this reality, with a few exceptions, teams will give coaches grace if they stumble out of the gate in year 1. We know why this is happening with the logistic regression model. *tenure* has a negative coefficient, which means that as *tenure* increases, predicted firing probability decreases. The random forest is a little more opaque and less interpretable, but we can infer from the results that something similar is happening there too. In reality, firing probability does *not* monotonically decrease as tenure increases. For a visualization, look at the chart below:

Figure 3: Distribution of Tenure Length for Fired Coaches



This graph bears out an important truth that the models may not be grasping: all else equal, firing peaks in years 2 through 4, with a major spike in year 3. In reality, coaches are not very likely to get fired after year 1, but the models are proving unable to handle this sort of distribution, where the maximum/minimum are not at the ends of the distribution. This makes sense for the logistic regression, which cannot handle non-monotonic relationships very well, but I hoped for a little bit better from the random forest, which is supposed to provide a better fit as the tradeoff for the reduction in interpretability.

A possible counterargument to this interpretation is that the models are technically predicting whether the coaches will *ever* be fired, not that they will be fired right now. For instance, both models predicted Dave Canales of the Panthers as an at-risk coach. They would be proven correct even if he maintained his job for a few more years and was fired in 2029. He does not necessarily have to be fired this year for the models to be correct. However, recall how the logistic regression model uses the *tenure* variable. All else equal, they would project a lower firing probability for Dave Canales for every extra year he coaches. In reality, the longer he coaches while the team flounders in mediocrity would *increase* the probability he is fired, not decrease it as the logistic regression expects.

6) Conclusions

There are certainly some things we can take away from this analysis, but their usefulness to teams in real-world situations is debatable. I came into this hoping to find a more interesting relationship than what we actually observed: *win_pct* dominates *fired*, and while a few other variables have some influence, *win_pct* appears to be the strongest explanatory variable by far. Essentially, we found that wins and losses rule the day when it comes to a coach's fate. There is definitely some value in backing up conventional wisdom with numbers, but if all you can do is confirm what is already being done, no one's behavior will change as a result of your analysis. This is not necessarily a bad thing, as my job is to find the truth, not bend the numbers to fit a narrative that I am happier with.

I feel that there is a second step here that I would have liked to take, but was forced to limit the scope because of time. I answered (with flawed models, as explained in depth earlier) *why* coaches get fired, but I did not find out whether this was the best way to evaluate them. I feel that this would have been a logical next step. Yes, coaches get judged primarily on wins and losses, but *should* they be judged that way? How do we isolate their impact on a team's wins and losses from all the other factors (e.g. roster quality, owner's influence, etc.)? Are teams that behave in this manner better off in the long run? Those are good questions, and remain areas for further research.

Throughout working on this project, I learned a tremendous amount regarding the "do's and don'ts" of statistical modeling, arguably more than I learned about the subject matter that I was actually studying. As explained in detail in the Results and Analysis section, the trustworthiness of the models may have been compromised because of the flawed approach I used when training them. By training them only on inactive coaches, I inadvertently filtered out a disproportionately high percentage of good coaches, which possibly made the models focus on the wrong variables. For example, the random forest ranked *sb_champ* as the least important variable, as random noise with no predictive power. It is hard to believe that winning a championship would have no effect on a coach's job security. Furthermore, I did not tune the threshold for the logistic regression, but in hindsight I should have, realizing now that the training data was heavily skewed towards *fired* = 1, with 83.8% of the coaches falling there. The default 50% threshold may have made sense if the data was split closer to 50/50, but given that it was actually 84/16 I should have experimented with different thresholds. Perhaps with more robust data exploration I would have noticed these differences between the inactive and active head coaches.

7) Reflection on Practicum

This semester will be my last at Georgia Tech, and I had a lot of fun using the skills I've gained over the course of this degree to work on a project I am passionate about, with the freedom to go in whatever direction I pleased. When the semester began, each of us had the option to do a solo project or create groups as large as 3 people. I had already done several group projects to this point – 1 in each of the last 4 semesters – so this time I decided to see what I could do if I worked by myself. This came with some good and some bad. Working alone gave me full creative control of the project. I was able to choose the topic as well as the methods to deploy, whereas had I worked with other people I would have had to compromise on some aspects, whether it be the topic, methods, or anything else. However, the tradeoff was that I did not have a second pair of eyes throughout the whole process to review my work at every stage. I did receive assistance from both my project advisor at Sandia and Georgia Tech's TAs, but this is not a substitute for having a partner who works with you all semester and is as intimately familiar with the material as you are. Perhaps some of the mistakes I made, if not all, could have been avoided if I had worked in a larger group. Another groupmate may have caught my mistakes and stopped me.

There is also the matter of scope. As mentioned in the conclusions section, there were some other avenues of exploration that I would have liked to get to but was unable to. Being able to divide the work with 1 or 2 more people likely would have allowed me to explore these other areas more deeply. Data collection was also a major time sink, and though I feel that my data set search was exhaustive, it is possible that a groupmate could have suggested a data source that had not occurred to me, thus saving a lot of time.

Ultimately, I am pleased with how the project turned out, even if there were some speed bumps along the way. I have learned a lot, and I hope that my passion for this project shines through as you read all the material. Thank you for reading!

8) Supplemental Code Files Walkthrough

In case you have further questions after reading this, all the code files are heavily commented as well, so they may help clear things up.

Modeling.py – The output of this file provides the answers to the 4 key questions I laid out at the beginning of this paper. The packages used are *numpy*, *pandas*, *scikit-learn*, *seaborn*, and *matplotlib*. With them, this file:

-Loads the data
-Trains the models (tuning hyperparameters for random forest)
-Prints out interpretations of the coefficients for the logistic regression
-Prints out feature importance rankings for the random forest
-Calculates surprise firings and unexpected survivors from inactive coaches subset
-Plots surprise firings and unexpected survivors
-Calculates most at-risk coaches and safest coaches from active coaches subset

Data Exploration.py – This file produces most of the data used in the "Exploratory Data Analysis" section. The packages used are *pandas* and *scipy*. With them, this file:

-Loads the data
-Prints contingency tables for *fired* vs all binary predictors
-Runs Chi-squared independence test for *fired* vs all binary predictors
-Calculates point-biserial correlation for *fired* vs all numeric predictors
-Prints contingency table and runs independence test for *fired_midseason* vs *minority*
-Calculates correlation for *win_pct* vs *final_yr_win_pct*

VIF.py - This is a very simple file that only exists to calculate the variance inflation factor for each of the predictor variables. The packages used are *pandas* and *statsmodels*.

-Loads the data
-Calculates and prints the VIF for every predictor

Tenure Histogram.py - This is another simple program that only serves one function. It filters the data set by inactive coaches only and then plots their tenure lengths in a histogram. The packages used are *pandas*, *matplotlib*, and *seaborn*.

-Loads the data
-Filters out active coaches
-Plots remaining coaches (inactive only) in a histogram by tenure length

9) References

[1] Foster, G., O'Reilly, N., Best Devereux, J. and Shundi, M. (2024). *General manager and head coach exits in the NBA and the NFL*. Sport, Business and Management, Vol. 14 No. 1, pp. 1-18. https://doi.org/10.1108/SBM-09-2022-0081

[2] Farnell, A., Berri, D., O'Sullivan, V., & Simmons, R. (2021). *An analysis of hiring and firing in the NFL*. Lancaster University. https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/lums/economics/working-papers/LancasterWP2021_003.pdf

[3] Holmes, P. (2010). *Win or Go Home: Why College Football Coaches Get Fired*. Journal of Sports Economics, 12(2), 157-178. https://doi.org/10.1177/1527002510378820

[4] Dixon, Patrick. (2017). *The Differentiating Personal and Tactical Dimensions of Head Coaches in the NFL: Why is Bill Belichick Successful When Others Are Not?*. CMC Senior Theses. 1465. https://scholarship.claremont.edu/cmc_theses/1465

[5] Braig, Kevin P. (2008). *Quantifying NFL Coaching: A Proof of New Growth Theory*. Journal of Quantitative Analysis in Sports: Vol. 4: Iss. 3, Article 1

[6] Madden, J. F. (2004). *Differences in the Success of NFL Coaches by Race, 1990-2002: Evidence of Last Hire, First Fire*. Journal of Sports Economics, 5(1), 6-19. https://doi.org/10.1177/1527002503257245 (Original work published 2004)

[7] Malone, K. D., Couch, J. F., & Barrett, J. D. (2008). *Differences in the Success of NFL Coaches by Race: A Different Perspective*. Journal of Sports Economics, 9(6), 663-670. https://doi.org/10.1177/1527002508317474 (Original work published 2008)

[8] Lapchick, R. E. (2021). *The 2021 Racial and Gender Report Card National Football League*. Squarespace. https://static1.squarespace.com/static/673fe9f4925aa40afa36e72f/t/6787408cea338268b0ac17bf/1736917134533/2021+APSE+RGRC.pdf

[9] Keane, S. (2021, August 26). *8 Scandals That Rocked the NFL*. History.com. https://www.history.com/articles/nfl-scandals-controversy

[10] (2005, December 15). *Four Vikings Charged with Party Boat Misdemeanors*. ESPN. https://www.espn.com/nfl/news/story?id=2260019

[11] Newberry, P. (2007, December 11). *Petrino Resigns as Falcon Coach After 13 Games, Heads to Arkansas*. Statesboro Herald. https://www.statesboroherald.com/sports/petrino-resigns-as-falcon-coach-after-13-games-heads-to-arkansas/

[12] Smith, C. (2012, May 2). *NFL's Punishment for Saints Bounty Program is the Costliest in Sports History*. Forbes. https://www.forbes.com/sites/chrissmith/2012/05/02/nfls-punishment-for-saints-bounty-program-is-the-costliest-in-sports-history/

[13] Schuster, B. (2020, June 28). *Timeline of Spygate, Deflategate, Major Patriots Scandals After Video Punishment*. Bleacher Report. https://bleacherreport.com/articles/2898081-timeline-of-spygate-deflategate-major-patriots-scandals-after-video-punishment

[14] Odum, C. (2006, December 15). *Mora Apologizes for Comments About University of Washington Job*. Statesboro Herald. https://www.statesboroherald.com/sports/mora-apologizes-for-comments-about-university-of-washington-job/

[15] Jones, L. H. (2012, December 18). *Belcher to Chiefs: I Have Hurt My Girl; I Can't Go Back*. USA Today. https://www.usatoday.com/story/sports/nfl/chiefs/2012/12/18/jovan-belcher-kansas-city-chiefs-kasandra-perkins-murder-suicide/1777359/

[16] Gutierrez, P. (2021, October 8). *Report: Jon Gruden Used Racist Comment in Referring to NFLPA Head DeMaurice Smith in 2011 Email*. ESPN. https://www.espn.com/nfl/story/_/id/32362483/report-jon-gruden-used-racist-comment-referring-nflpa-head-demaurice-smith-2011-email

[17] Hayes, A. (2022, October 13). *Reverse Survivorship Bias: What it is, How it Works, Example*. Investopedia. https://www.investopedia.com/terms/r/reverse-survivorship-bias.asp