Spencer Slote

MS Practicum Report

Introduction

Every year, hundreds of NFL prospects gather for one of the most impactful talent evaluations in all of sports: the NFL Combine. Among the drills performed at this combine, none draws more attention than the 40-yard dash. This drill is typically seen as a great indicator of a player's raw speed and explosiveness, with a player's 40 time often dramatically influencing their draft stock and perceived upside. While the 40-yard dash receives this heavy emphasis from scouts and media alike, it's still sometimes treated as a standalone measurement, and at times is evaluated without considering a player's full profile. In reality, a player's performance in this drill is very likely influenced by a combination of physical traits and their performance in other drills, among other factors. This then raises an important question: can we accurately predict a player's time in the 40-yard dash based on the rest of their combine performance / profile? The main goal I set for this project takes on that question by developing a predictive model using data from over the two decades of NFL Combine results.

Being able to accurately predict a player's 40-yard dash time has real value from not only an evaluation perspective, but also a decision-making perspective as well. Not every player runs the 40 at the combine, and even when they do, their raw time may not always tell the full story. A predictive model can help fill in gaps for players who skip or miss out the drill due to injury or perhaps a strategic choice. Not only that, but it can also provide scouts with additional context to determine whether a player's 40 time is in line with any expectations based on a player's overall profile. In short, it has the potential to make player evaluation more consistent, more objective, and less reliant on a single timed run. To tackle this question, I utilized the website Football Reference (2025 Combine), which contained NFL Combine results from 2000 through 2025 separately. These included player measurements such as height and weight, results from all core athletic drills (40-yard dash, vertical jump, broad jump, bench press, shuttle, and 3-cone), as well as position and draft information. Once I had each year's data, I compiled them into one complete dataset to be used for my model and project overall. After cleaning and refining the data to remove incomplete or non-comparable entries, as well as fixing errors in certain columns, I shifted my focus to building the predictive model that would use these variables to estimate a player's 40-yard dash time.

The approach I decided to use for this involved training a machine learning model known as a random forest regression. This method allowed me to identify which factors most influence sprint speed and to evaluate how accurately a player's 40 time could be predicted. The reason I chose this as the modeling technique for this project was because it handles the complexity and variability of NFL Combine data extremely well. Unlike linear models, which assume a straight-line relationship between predictors and the outcome, random forest can capture nonlinear patterns and interactions between variables. This is especially important in athletic performance data, where traits like weight, vertical jump, or position may influence 40-yard dash time in ways that aren't perfectly linear or consistent across all players. Random forest is also robust to any missing or noisy data and can also handle both numeric and categorical variables, making it a great fit for a dataset that includes player positions and a variety of physical metrics. Not only this, but it can also provide insight into variable importance, which allows us to identify which factors contribute most to predicting speed.

In addition to the predictive modeling, this project also takes a look at the full combine dataset to explore additional patterns and insights. One area of interest involves analyzing the relationships between different drills to see which events tend to be most correlated with each other.

For example, it's worth examining whether a strong vertical jump usually goes hand-in-hand with an explosive broad jump, or if agility-based drills like the shuttle and 3-cone tend to overlap in what they measure. I also looked to identify any outliers in the data, such as players who posted elite performances across multiple drills. These kinds of outliers would help highlight what truly rare athleticism looks like in the context of over two decades of combine results. In addition, I explored how performance in each event varies across position groups to see which roles demand the most speed, power, or agility at a high level. Finally, the last additional area of analysis I chose was examining whether strong combine performance actually has any relationship with draft position, and whether there's measurable value in excelling in certain drills over others, or if teams have their mind set on players regardless of their performance at the combine.

Data Overview

The final dataset used for this project includes over 8,000 observations, with each row representing a player who participated in the NFL Combine between 2000 and 2025, except for 2021, where there was no NFL Combine due to COVID-19. For the 2021 observations, each row corresponds to a player's measurements and performance in their own pro-day. The dataset contains physical measurements for each player such as height and weight, as well as performance results from the six drills, position / college information, and draft outcomes. Several new variables were also created to enhance the analysis, including grouped position categories (e.g., offensive line, defensive backs, skill positions) to make comparisons across roles cleaner and more meaningful. This dataset served as the main foundation for both the predictive modeling and the exploratory analysis conducted throughout this project. Another important consideration with this dataset was how to manage observations with missing values. The most common way this happened was through players who voluntarily or involuntarily did not take part in certain drills. Depending on the analysis I was performing, I made a number of different versions of this dataset. For example, when constructing the predictive model for 40-yard dash times, every null value in the '40yd' column was removed, as players who didn't take part in the 40-yard dash would provide no additional insight into predicting another player's performance.

Exploratory Data Analysis

Before I got into modeling, I wanted to get a better feel for the overall data. I started by looking at how players performed across the different combine drills and how those results varied by position group. I also checked to see if any of the events were closely related, like whether players who do well in the vertical jump also tend to have strong broad jumps. Grouping players by position helped reveal which roles typically excel in certain drills, and which areas may matter more for different types of players. I also looked at how certain drills might be connected, like whether players who perform well in agility based tests also tend to do well in speed based ones. This exploration definitely helped set the foundation for the model and gave me a more clear picture of some of the patterns within the dataset.

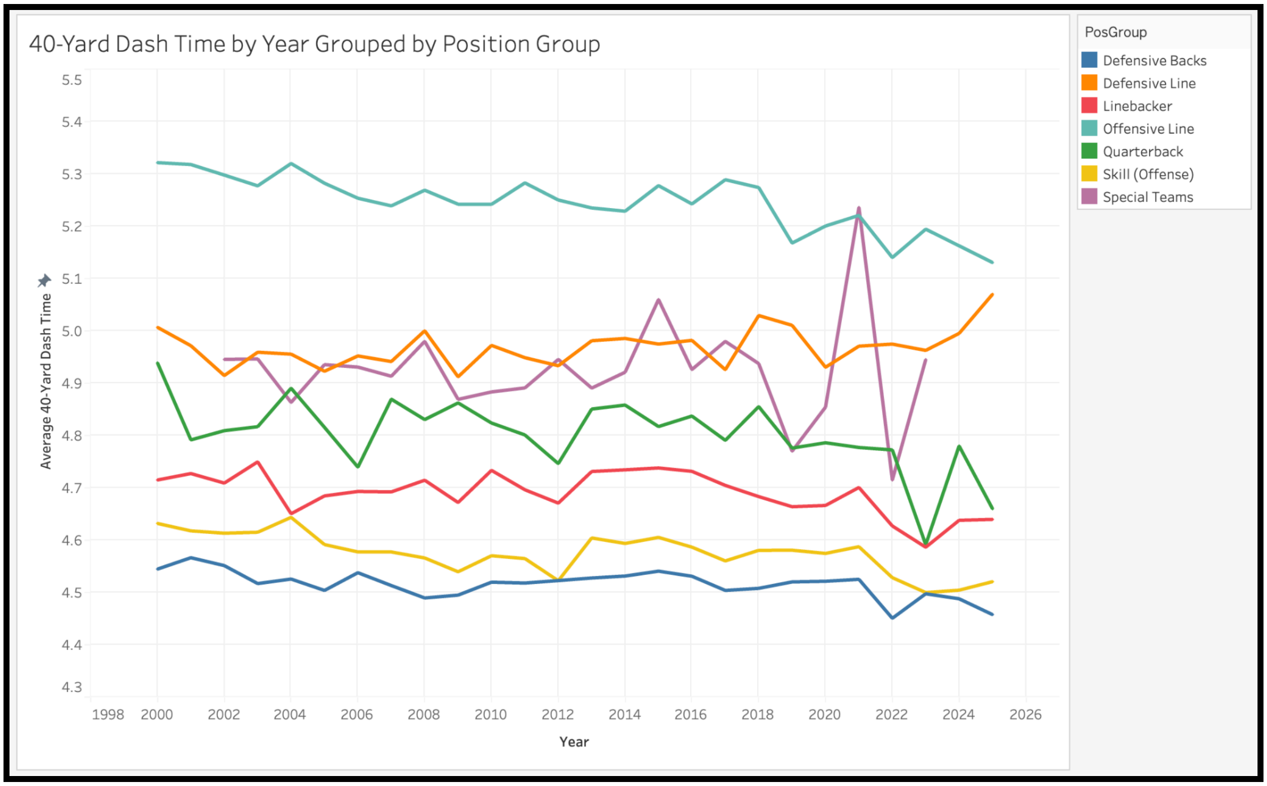

Figure 1: 40-Yard Dash Time by Year Grouped by Position Group

To get a sense of how 40-yard dash performance has changed over time, I looked at the average 40 time for each year, grouped by position type. As expected, wide receivers and defensive backs consistently posted the fastest times, typically hovering around 4.5 to 4.6 seconds. Offensive linemen remained at the high end of the range, averaging over 5.2-5.3 seconds in the early 2000s. However, we can actually see a gradual improvement for this group over time, with offensive linemen averaging closer to 5.1-5.2 seconds in recent years. While there are of course fluctuations year by year, the overall trends seemed to stay relatively stable for most of the other groups, with no major signs of players across positions getting significantly faster or slower over the last two decades. However, grouping by position still allowed us to clearly distinguish which position groups performed significantly better for not only the 40-yard dash, but for any of the main drills.

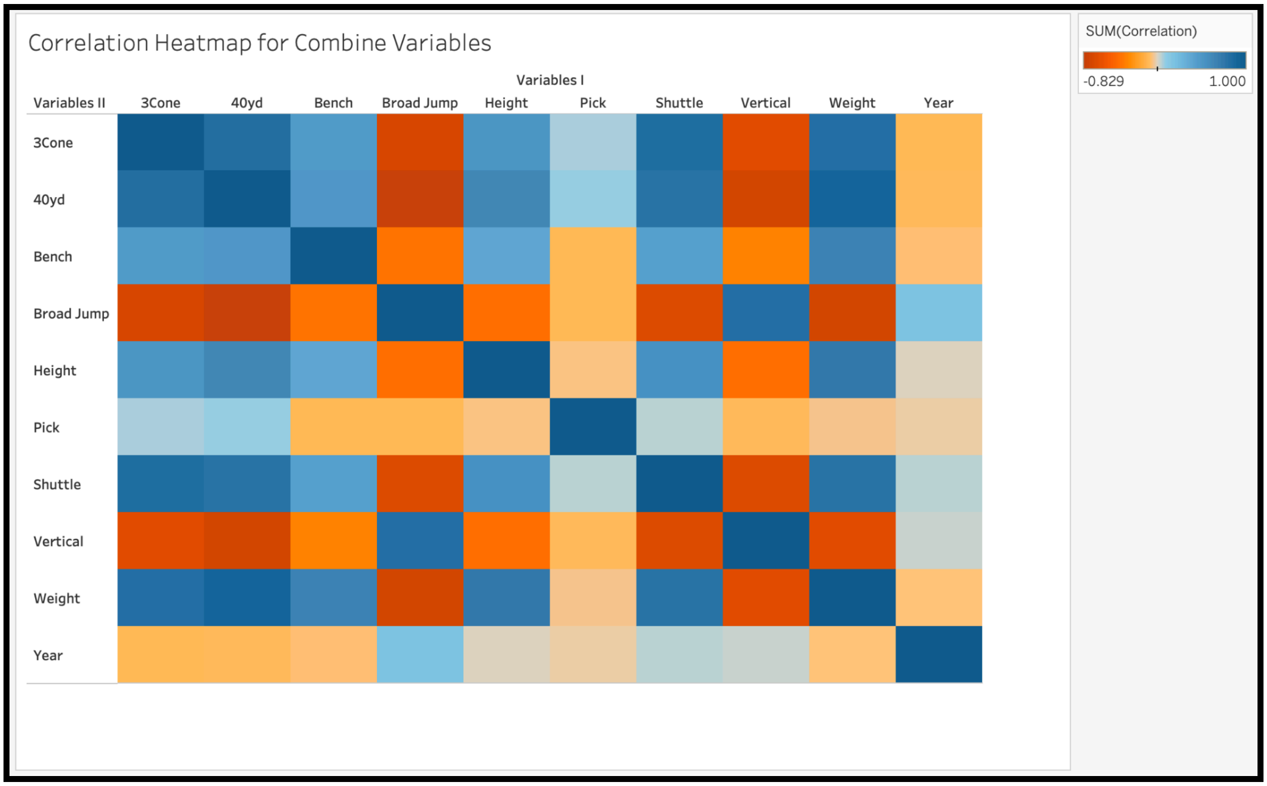

Figure 2: Correlation Heatmap for Combine Variables

Correlation Matrix for Combine Variables

SUM(Correlation): -0.829 to 1.000

Variables II	3Cone	40yd	Bench	Broad Jump	Height	Pick	Shuttle	Vertical	Weight	Year
3Cone	1.000	0.824	0.466	-0.741	0.523	0.060	0.845	-0.677	0.814	-0.112
40yd	0.824	1.000	0.497	-0.829	0.623	0.082	0.783	-0.756	0.897	-0.102
Bench	0.466	0.497	1.000	-0.425	0.381	-0.109	0.427	-0.358	0.657	-0.075
Broad Jump	-0.741	-0.829	-0.425	1.000	-0.438	-0.117	-0.717	0.816	-0.759	0.193
Height	0.523	0.623	0.381	-0.438	1.000	-0.059	0.548	-0.449	0.739	-0.006
Pick	0.060	0.082	-0.109	-0.117	-0.059	1.000	0.041	-0.098	-0.052	-0.030
Shuttle	0.845	0.783	0.427	-0.717	0.548	0.041	1.000	-0.695	0.784	0.038
Vertical	-0.677	-0.756	-0.358	0.816	-0.449	-0.098	-0.695	1.000	-0.694	0.019
Weight	0.814	0.897	0.657	-0.759	0.739	-0.052	0.784	-0.694	1.000	-0.069
Year	-0.112	-0.102	-0.075	0.193	-0.006	-0.030	0.038	0.019	-0.069	1.000

Figure 3: Correlation Matrix for Combine Variables (To Visualize Coefficients)

To get a better sense of how different combine events relate to each other, I created a correlation heatmap, as well as a matrix which shows the correlation coefficients using all the main drills, along with height, weight, draft pick, and year. Not surprisingly, there were a number of events that were strongly correlated with one another. Some of the strongest positive relationships were between the vertical jump and broad jump, the 3-cone and the 40-yard dash, as well as between the shuttle and 3-cone drills. Each of these three pairs had a correlation coefficient of over 0.8. Since these are all measuring similar types of explosiveness or agility, it would make sense that they're closely connected. The 40-yard dash also showed a decent negative correlation with vertical and broad jump, which also makes sense, as it shows that players who test well in those tend to run a faster 40. Bench press also stood out for not being very strongly correlated with anything else, with none of the correlation coefficients with bench press being over 0.5 except for weight. Draft pick had only weak correlations with all of the other variables, and the year variable didn't show much of a pattern either, which suggests there hasn't been a big shift in overall combine performance over time. This also connects with what we saw in the 40-yard dash performance over time in figure 1, and also shows that there is not a significant trend in the other events either.

Predictive Modeling Approach

The goal of this model was to predict a player's 40-yard dash time using the rest of their NFL Combine results. I used a mix of physical measurements like height and weight, along with performance in the other drills from the dataset. I also included player position as a variable, using the individual positions instead of grouping them together like I originally did in the exploratory analysis. I initially grouped positions rather than individual positions in order to potentially simplify the model. However, when I tested the model performance with each individual position instead, it actually ended up performing better, so I stuck with that. Rather than narrowing down the list of inputs for the model so early on, I decided to include all of the combine metrics from the dataset in the model. The goal was to give the model as much relevant information as possible and let it determine which traits were most useful in predicting 40-yard dash performance, as I wanted the model to do more than simply be the most accurate when predicting 40-yard dash times.

After testing a few different options, I decided to use random forest regression for the model. One of the main reasons I chose it is because it works well with a mix of variable types, including both numeric inputs as well as categorical ones like player position. This method also doesn't assume a linear relationship between the inputs and the target variable, which is important in this case since a trait like weight can impact 40-yard speed differently depending on position or other traits. Random forest is also good at handling the interactions between variables and it doesn't require a lot of manual feature selection. Since my goal wasn't just to make the most accurate predictions, but also to understand which factors were the most impactful, it helped that random forest was able to provide variable importance scores as part of the output. Overall, it was a good fit for a dataset like this, where the relationships between variables may not always be basic or straightforward.

Once I chose the model type, I split the data into a training set and a test set using an 80/20 split. The training set was used to build and tune the model, and the testing set was used afterwards to see how well the model performed on new data. Before training the model, I set up a preprocessing 'recipe' to clean and prepare the data. I started by filling in missing values for the numeric columns (vertical jump, shuttle, etc.) by replacing them with the median value for each column. After that, I converted the player position variable, which is categorical, into dummy variables so the model could actually work with it. This meant that each position (WR, QB, OT, etc.) was turned into its own binary column. Since random forest models can handle variables on different scales, there was no need to scale or normalize the data. Once these preprocessing steps were all set, I combined the recipe with the model into one single workflow. This made it easier to run everything together without having to do each step manually.

After setting up the model, I had to figure out which settings would help it perform best. I focused on tuning two main ones, with the first being *mtry*, which controls how many variables the model looks at when it decides where to split the data. The second setting, *min_n*, sets the minimum number of observations which are needed in a group before the model can split it again. Getting these settings right helps the model make better predictions without overfitting. To do this, I used five-fold cross-validation, meaning the training data was split into five parts. The model trained on four of them and tested on the fifth, and this rotated until each part had been used as the test set one time. This gives a much more reliable picture of how well the model would perform on new data. I also tested ten different combinations of *mtry* and *min_n*, and for each one, I tracked how well the model did using three metrics: RMSE (Root Mean Squared Error), MAE (Mean Absolute Error), and R^2. RMSE shows how far off the predictions were, on average. MAE does the same but in a simpler way, and R^2 tells how much of the variation in 40-yard dash times the model was able to explain. After testing all the combinations, I went with the one that had the lowest RMSE, since that meant it was the most accurate overall.

Once I identified the best combination of hyperparameters, I was able to finalize the model and retrain it using the entire training subset. This allowed the model to learn from as much as possible before being evaluated on new observations. With the finalized model, I generated predictions for the testing set, which were compared to the actual 40-yard dash times, making it possible to measure how accurate the model actually was. This step was essentially the final check to see how well the model could generalize past the data it was trained on initially. After this, I evaluated the model's performance, which showed a RMSE of 0.114 and an MAE of 0.088. These are both relatively low given that most 40-yard dash times fall between 4.2 and 5.5 seconds, meaning the model's predictions were generally very close to the actual times. The R^2 value for the model was 0.855, meaning it was able to explain just over 85% of the variability in 40-yard dash times based on the input variables. These results suggest that the model performs quite well, both in terms of accuracy and consistency, and it could be strong enough to be used for estimating 40 times for situations when a player didn't run the drill.

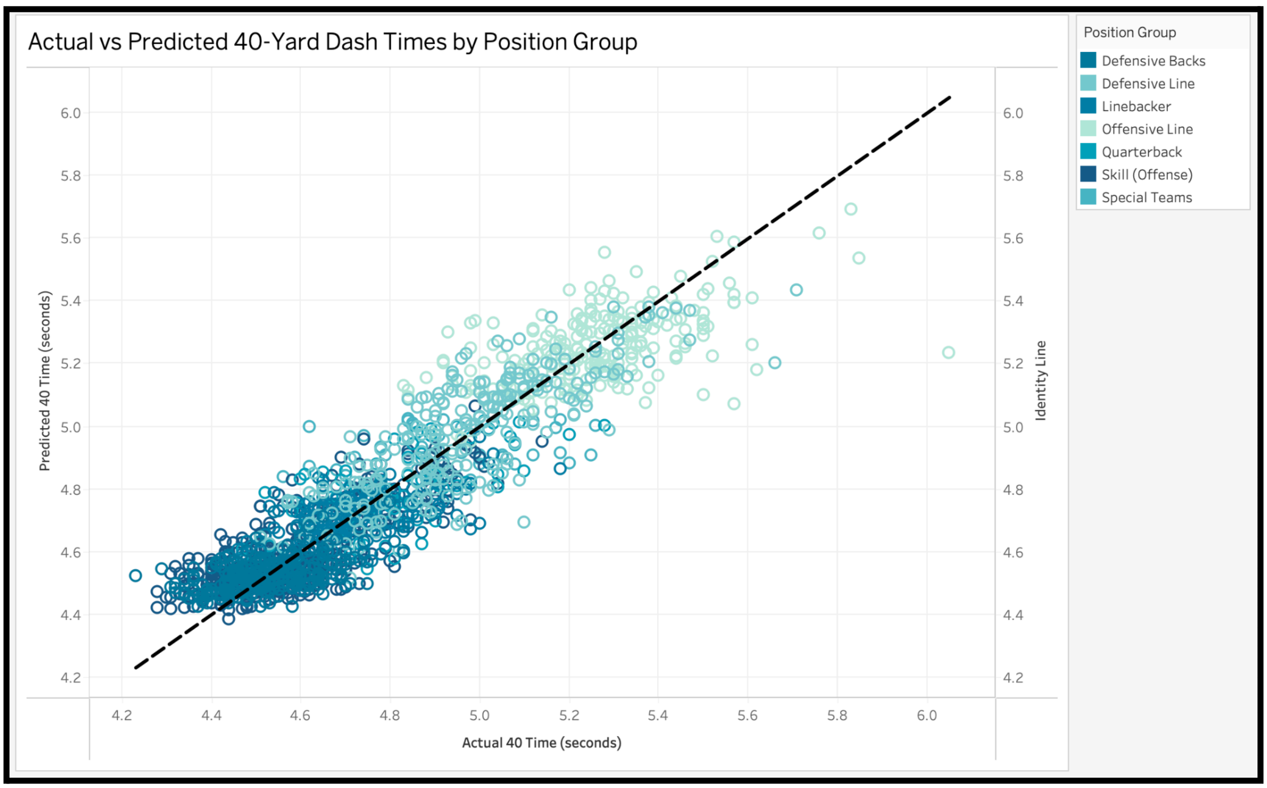

Figure 7: Actual vs Predicted 40-Yard Dash Times by Position Group

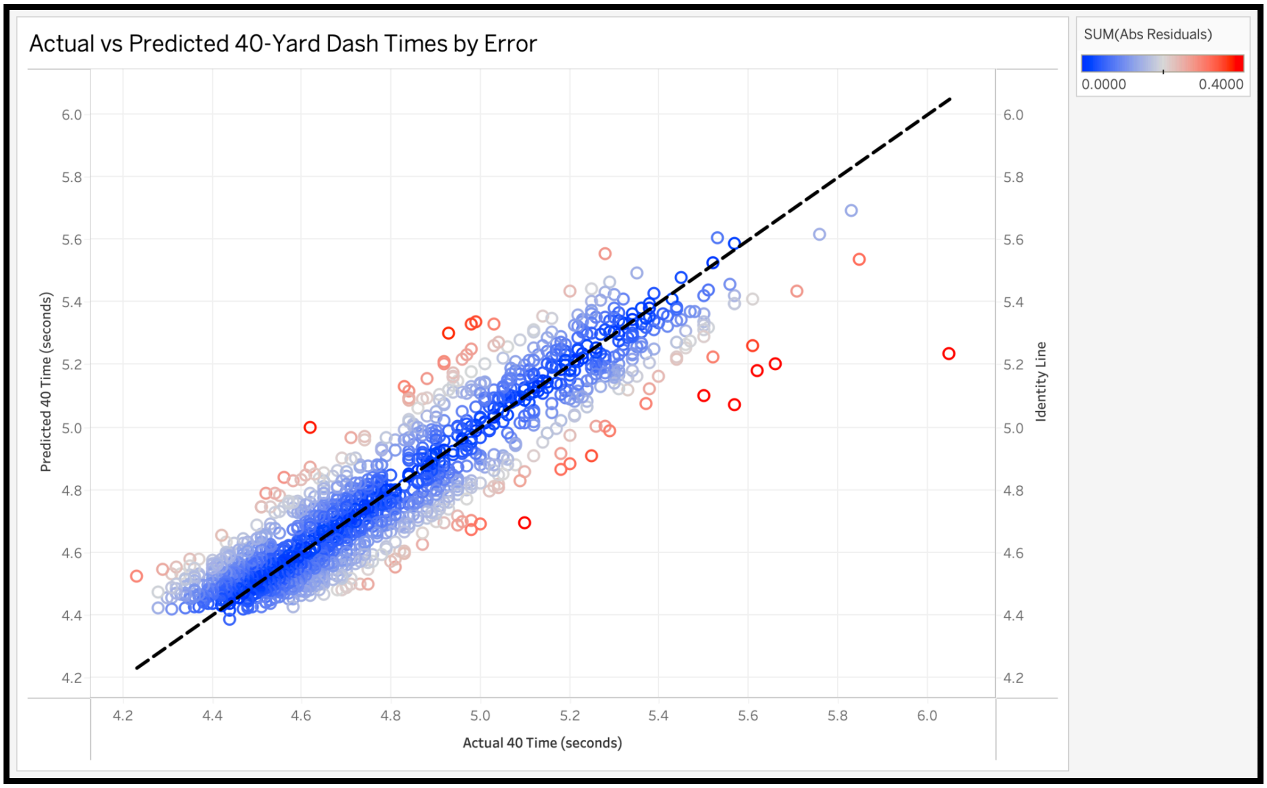

Figure 8: Actual vs Predicted 40-Yard Dash Times by Error

To further evaluate how well the model performed, I also visualized the predicted versus actual 40-yard dash times using two different scatterplots. In the first chart, each point represents a player, plotted by their actual and predicted 40 time, with the shade of blue / teal used to indicate their position group. The dashed line represents a perfect 1:1 prediction, so the closer the points are to this line, the more accurate the model's prediction. This chart helped visualize whether certain position groups were predicted more or less accurately overall, and it seemed to show that groups with slower average 40 times, such as offensive and defensive linemen, tended to have a few more 'larger' outliers, suggesting potentially a bit more variability in how those players were predicted. The second scatterplot added another layer to the visualization by using color to show the size of each prediction error (residual). Blue dots indicate players where the prediction was very close, while red dots represent the larger misses. What stood out in this graph is the tight clustering of points around the line across most observations, suggesting the model was able to consistently predict sprint times with strong accuracy. This helped confirm that the model was generally reliable, but also pointed out a few individual instances who may have run significantly faster or slower than expected, possibly due to unique traits the model couldn't fully account for. Together, these charts give a more complete view of how the model performed not just on average, but on a player-by-player basis.

Beyond measuring how accurate the model was, I also looked at which variables from the combine dataset had the most influence on the predictions. One of the other advantages of using random forest for the model is that it provides variable importance scores, which show how much each variable actually contributed to improving the model's accuracy. These scores don't show a direct cause and effect relationship, but they give a very useful view into which traits the model relied on most when estimating 40-yard dash times. Looking at the importance values helped confirm which factors matter most for speed, and also revealed a few interesting observations for some variables.

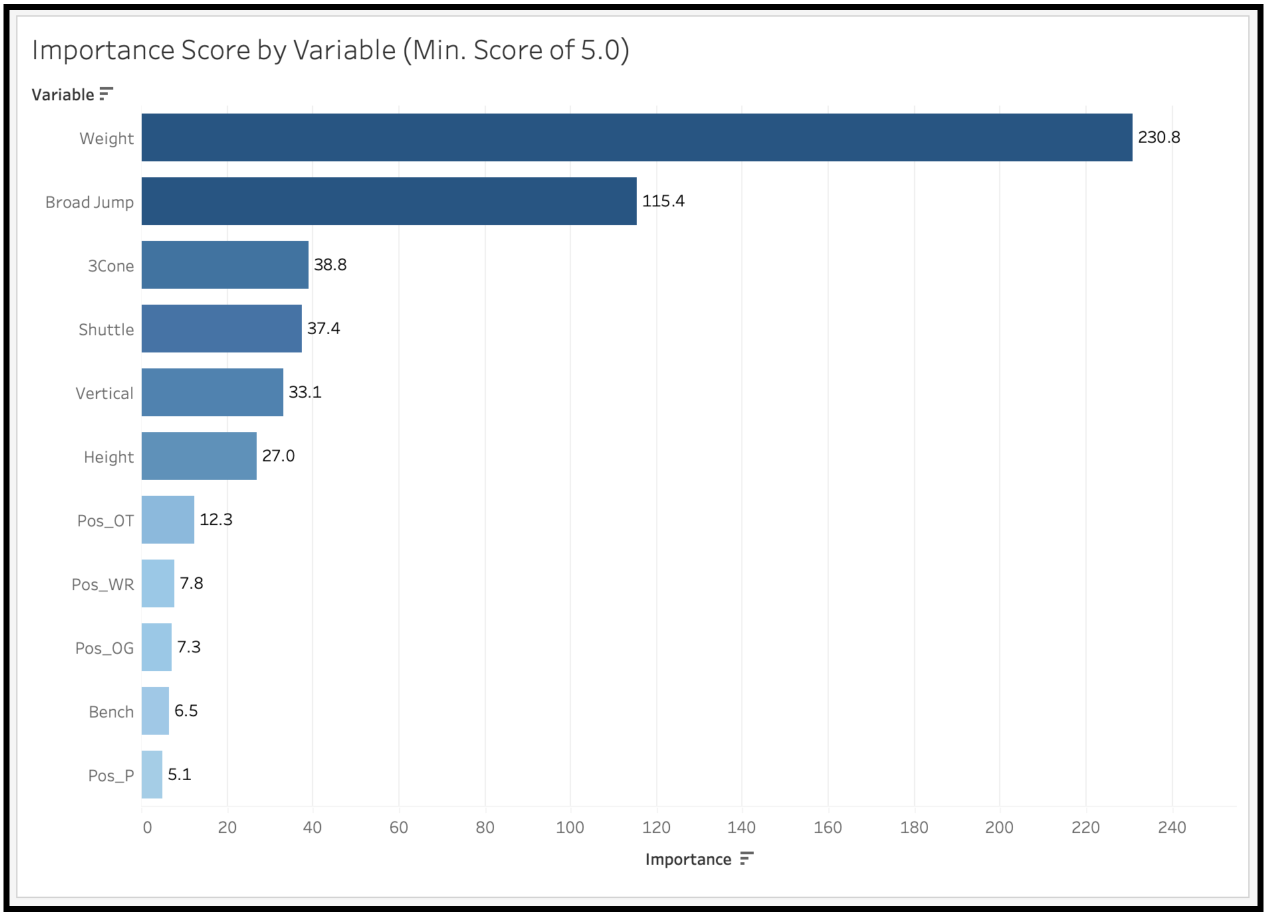

Figure 4: Bar Chart of Variable Importance Scores

The variable importance scores showed that weight was by far the most influential factor in predicting 40-yard dash times, which was expected for the most part. It had nearly double the importance of the next highest variable, broad jump, which also ranked highly. Other strong predictors included 3-cone, shuttle, and vertical jump, which are all tied to agility and lower body explosiveness, making sense given the nature of the 40-yard dash. Height was also moderately important, though noticeably less than weight. Interestingly, position variables like offensive tackle (OT), wide receiver (WR), and guard (OG) also showed up with moderate importance, suggesting that the model picked up on trends tied to certain positions. On the other hand, variables like bench press (compared to the other drills), as well as many of the remaining position indicators had relatively low impact, which also contributes to the idea that upper body strength and certain roles don’t contribute much when it comes to straight-line speed.

Overall, the model ended up performing very well and provided useful insight into which traits are most connected to 40-yard dash performances. Not only was it accurate in predicting these times, but it also showed which factors the model leaned on the most when making those predictions. Most of the results aligned with expectations, such as weight and explosiveness-based drills being top predictors, whereas others like bench press and certain positions had less impact than some people may have initially thought. The model isn’t meant to be perfect, and it doesn’t capture *everything* that goes into sprinting ability, but it adds a very valuable layer of analysis. It shows that we can accurately predict performance in a highly scrutinized drill like the 40-yard dash using the rest of a player's combine profile, which could be especially useful when times are missing or potentially taken out of context.

Additional Analysis

Although the main focus of this project was building a predictive model for 40-yard dash times, I also spent time exploring the full NFL Combine dataset to see if I could uncover any additional insights. Since the data covered over two decades of combine results, it gave me a chance to look at how players and position groups performed across different drills and how results have potentially varied over time. I also looked for players who put together ‘elite’ performances across multiple drills, as well as explored whether strong combine results showed any real connection to draft outcomes or even overall career success. The additional analysis helped put the 40-yard dash modeling into a bigger context, showing not just how individual traits predict sprint speed, but how combine athleticism can (or can't) affect how players are evaluated.

Figure 6: Chart of Elite Combine Performers

One of the more interesting side analyses I worked on was identifying players who delivered truly 'elite' performances across multiple combine events. To find these players, I kept track of anyone who scored in the top 5% in at least four of the six core combine drills. Out of more than 8,000 players, only 12 met this threshold, which showed just how difficult it is to perform at such a high level across the board at the NFL Combine. These players represent some of the best raw athletes the event has seen, with the majority excelling in primarily speed and explosiveness based drills. While a few recognizable names like Ja'Marr Chase appeared in this group, most of the list was actually made up of players who are far less well known, especially to casual fans. The main standout was Troy Apke, a safety from Penn State who posted elite scores in *five* of the six drills , which was more than any other player in the dataset. Apke's showing at the combine was incredible, but his NFL career so far shows that even the most impressive combine performances won't always guarantee long-term success. I also noticed from this analysis that every player was a wide receiver or a defensive back. This also makes sense, as the combine has a greater number of speed / agility drills, which offensive and defensive skill players excel at, rather than pure strength drills, which other position groups would potentially perform better at. Overall, this analysis helped to highlight both the strength, as well as the limitations of using combine results alone to predict future performance.

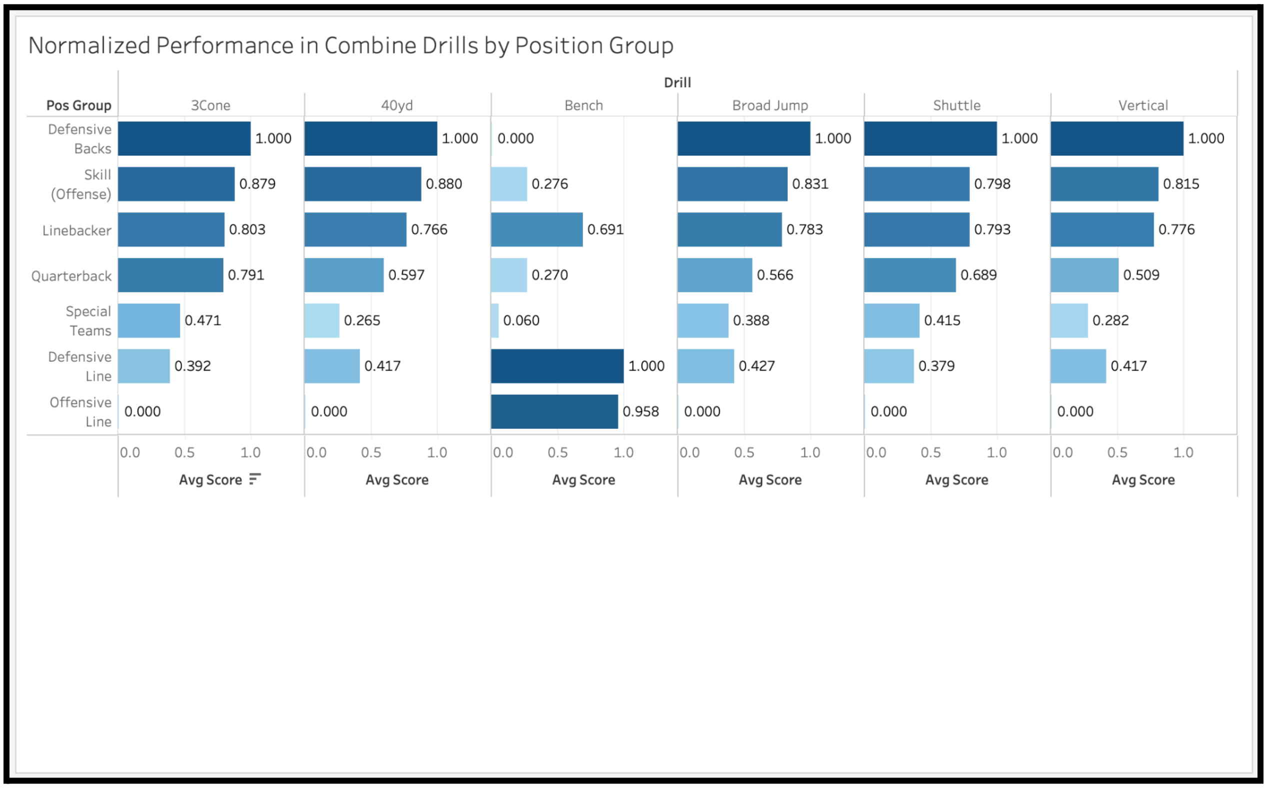

Figure 5: Bar Chart of Normalized Performance in Combine Drills by Position Groups

To wrap up the additional analysis, I compared normalized combine performances across the six main drills to see how each position group performed in comparison to one another. Unsurprisingly, defensive backs and offensive skill players (WRs, RBs, and TEs) stood out with the strongest overall scores in the speed and explosiveness events. Defensive backs actually posted a normalized score of 1.0 (out of 1.0) across every drill except for bench press, where they ranked the lowest among all groups. Skill players weren't far behind, with a normalized score of 0.88 in the 40-yard dash, a 0.88 in the 3-cone drill, and a 0.83 in the broad jump. On the opposite end, offensive linemen ranked near or at the bottom for almost every drill except for bench press, where they scored a very strong 0.96. Linebackers and quarterbacks showed more balanced results, with linebackers scoring at the higher end for every single drill. Defensive linemen and special teams players generally performed worse in most events, although defensive linemen performed extremely well in the bench press with a perfect normalized score of 1.0. These results lined up well with expectations based on general assumptions, as well as the physical demands of each position group. Overall, these additional analyses helped give a broader view of the patterns and outliers from the NFL Combine, providing more context around athletic performance beyond just the 40-yard dash.

Conclusion

The main question driving this project was whether a player's 40-yard dash time could be predicted using the rest of their NFL Combine results and overall profile. From pulling together over two decades of data and training a random forest model, the answer turned out to be yes, and with pretty solid accuracy. Not only did the model give strong predictions, but it also offered a clear picture of which traits truly matter most when it comes to sprint speed. Weight stood out as the most important factor, followed by performance in other explosive drills such as the broad jump and 3-cone, which makes sense given what the 40-yard dash is designed to measure. That being said, the model isn't perfect. Obviously, from the data it was trained on, it can't capture *everything* that goes into a player's sprinting ability, such as running form, past injuries, how well they were trained, etc.

Still, it gives a strong, consistent way to estimate 40-yard dash times when those numbers are missing or when you just want to put a raw time into better context. It's an extremely helpful tool for adding more perspective to a number that often simply gets taken at face value. Beyond the main model itself, the other analyses helped round out the story as well. Looking at combine performance across position groups, spotting trends between events, and identifying the outlier players with elite all-around results added more depth to the dataset overall. These extra pieces not only supported much of the modeling work, but they helped show what athleticism looks like across different roles, and how rare it is to see someone truly dominate a number of drills.

Altogether, this project truly showed me how much you can truly learn from NFL combine data when you step back and look at it as a whole. Building the main predictive model gave me a way to connect the dots between different traits and 40-yard dash performance, and the extra analysis helped put some of those numbers into context. On top of that, it gave me a chance to build a real machine learning model from the ground up, which was both a challenge and a genuinely fun problem to work through.

Citations

NFL Combine Data:

https://www.pro-football-reference.com/draft/2025-combine.htm

Random Forest Information:

https://www.ibm.com/think/topics/random-forest

Random Forest in R:

https://www.listendata.com/2014/11/random-forest-with-r.html

Chat GPT for Grammar & Structuring (Primarily in Intro / Outro):

https://chatgpt.com/