**NFL Combine Predictions**

Analysis of NFL Combine Results (2009 – 2019)

Marcin Maleski

December 17, 2022

**Background/Description:**

The NFL (National Football League) is one of the most prominent professional sports leagues in the United States. With so much to gain for being an NFL player it has become one most sought-after athletic jobs. Every year, hundreds of talented players are eagerly trying to land a spot on and NFL team roster. With so many players and so little teams, the NFL has created a performance-based test to evaluate every participant. This test is the NFL Combine.

The NFL Combine is a week-long showcase occurring every February. This is where the best potential NFL prospects come to showcase their abilities in a series of tests and performance-based evaluations. The NFL Combine gives NFL teams an in-depth evaluation of the best players coming out of college, so they can better prepare themselves for who they want to draft in the upcoming month.

**Problem Scenario/Business Issue:**

Drafting a player is a huge investment for NFL teams. With a potential NFL championship (Superbowl) on the line every year, it is essential that these teams select the best players .It is e hard to tell which players will succeed at the NFL level. Even with a extremely well performance at the NFL combine, success is never guaranteed.

With so much on the line at the NFL draft, what I want to know is what kind of thought goes into a drafting a player. Which Combine test results are the most impactful and what do these NFL scouts look at the closest when making their selections.

**Objective/ Goals of the Project:**

To peel back the layers of the NFL Combine and draft selection, I found a dataset from Kaggle that includes every NFL combine participant from 2009-2019 and their results. With this dataset I plan on breaking it down to give myself a better understanding of the makeup of these NFL players, and what NFL scouts are looking at when making their draft selections. Since Football is a very particular sport, I am going to try and break down the dataset by position type to show myself and the viewer a good insight on what kind of Combine performance best suits each individual player.

**Data Overview and Transformation:**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Column** | Height | Weight | Sprint 40yd Dash | Vertical Jump | Bench Press Reps | Broad Jump | Agility 3cone | Position Type | Drafted |
| **Description** | Height (ft) | Weight (lbs.) | How fast the participant can sprint 40 yards (sec) | Vertical Jump Height (cm) | Maximum reps of bench press reps (225 lbs.) | How far a player can jump horizontally (cm) | Base line agility test (sec) | What position the player is | Where they evidently drafted |

The original dataset included 18 columns with 3467 rows. The first thing I did was eliminate every column that I didn’t deem necessary for my project. I ended up with 9 columns that I found most important. After that I eliminated all rows with N/A values. I then changed some numeric values that were listed as metric system values to imperial system for better understanding. My final data set included 9 columns and 1755 rows.

-Below is a list of the columns and description of those columns

**Data Exploration/Visualization:**

To give myself a better understanding about what physical talents these NFL prospects possess, I first created a few important visualizations that represent a comparison of NFL combine test results per position.

To start, I wanted to see what Combine tests are highly correlated. To do this I created a heat map between all test results.

**Chart

Description automatically generated**

This heatmap breaks down the correlation between separate columns. When looking

at this heatmap, I was able to see which combine tests have the highest correlation. It gave me a good insight on what tests I should visualize.

-When I initially created this heatmap, I expected to see strong correlation between a few columns and weak correlation between others. What surprised me was some columns that I expected to have a strong correlation ended up having significantly weak ones, like the Vertical Jump test and Weight (-0.7).

Chart, scatter chart

Description automatically generated Since the correlation between Vertical Jump and Weight was so low, I decided to visualize it.

Although the heatmap indicated a low correlation between the two columns when creating a scatter plot you can see that there is a clear relation between the amount you weigh and how high you can jump.

When adding the position type to this scatter plot you can clearly see what players are preforming the best at the Vertical jump test. Defensive Backs are jumping the highest while Offensive lineman are jumping the lowest.

**Chart, box and whisker chart

Description automatically generated** To give us a better insight on what players are performing best at what positions I created a boxplot comparing the positional average of 40 yd dash times.

This boxplot contains very similar results to our first scatter plot. Again, Defensive Backs are the fastest players as they are averaging around 4.4 second for their 40 yard dash time, while Offensive Lineman are averaging the worst with around a 5.2 second 40 yard dash time. These last two plots show the disparity between positional performance. In the NFL . Defensive backs need to be some of the most agile and athletic players on the field, as their job requires them to cover a wide range of field. Offensive Lineman on the other hand don’t need to be as fast and agile. There position requires them to block any defensive player from reaching the quarter back. You can visualize the different talents required at different positions with these last two graphs.

The last two plots have shown us what defensive backs are excelling at in the NFL combine, so now let’s compare the Defensive Backs and Offensive Lineman in the strength test, a physical ability that Offensive Lineman rely on.

**Chart, histogram

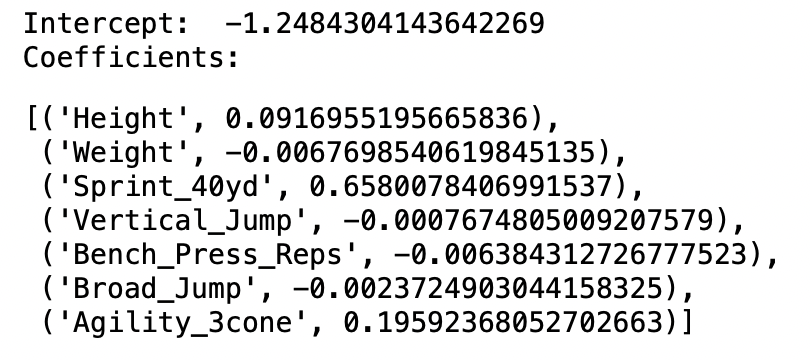
Description automatically generated**

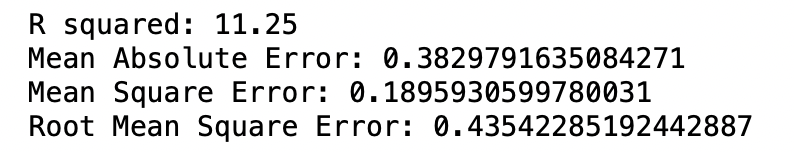
This plot shows us the comparison of defensive backs and offensive lineman in the strength column. You can see how offensive lineman are averaging much more on the bench press rep’s test. This is because it is required to be extremely strong as an offensive lineman.

**Multiple Regression Analysis:**

So we have seen the different performance of different positions of players in the NFL draft. The real question is overall what tests prove the best at predicting if someone is going to be drafted or not. In order to find this, I ran a multiple regression model to detect the weight of every variable.

Before completing and evaluating the model we need to build the model equation. This equation gives us a good insight on the weight of each model variable.



After reviewing the model equation there are some key takeaways. Height, 40 yd dash, and agility cone are the only variables that are positively correlated meaning they have the most weight on whether or not a player is drafted. This makes sense because football is a game of speed. Being more agile and faster than the rest of the field is hugely beneficial.

This model gave us a 11.25 value for r squared. This indicates that the model underperformed significantly as only 11.25% of the points fell within the regression line. Although our R-squared value was poor our MAE, MSE, and RMSE values were all extremely close to 0 which means that the model performed well.

**Conclusion**

NFL combine testing performance is extremely related to your position and NFL scouts will evaluate you differently based on your position. They are not expecting a 350 Offensive Lineman to have a really low 40 yard dash, but they are expecting him to be very strong. This goes for every position. When taking positions out of the picture, overall, your chances or getting drafted are much higher if you are faster and more agile than the rest of the field. We saw the weight of the 40 yard dash and agility test when constructing the model equation for our out multiple regression model. As a sport that is so dependent on physical abilities, this multiple regression model showed us that speed is the biggest factor when being drafted into the NFL.

**Sources**

[**https://www.kaggle.com/datasets/redlineracer/nfl-combine-performance-data-2009-2019**](https://www.kaggle.com/datasets/redlineracer/nfl-combine-performance-data-2009-2019)