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 world. With so much to gain for being an NFL player it has become one most sought-after athletic jobs. With so many players and so little teams, the NFL has created a performance-based test to evaluate every incoming prospect. 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 for the upcoming season.

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 to help them achieve their goal. With so much on the line at the NFL draft, what I want to know is what kind of thought goes into 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. I used this dataset to break down NFL combine results and to give myself a better understanding of the physical abilities of these NFL prospects. 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:

The original dataset included 18 columns with 3467 rows. First, I eliminated every column that I didn't deem necessary. 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.

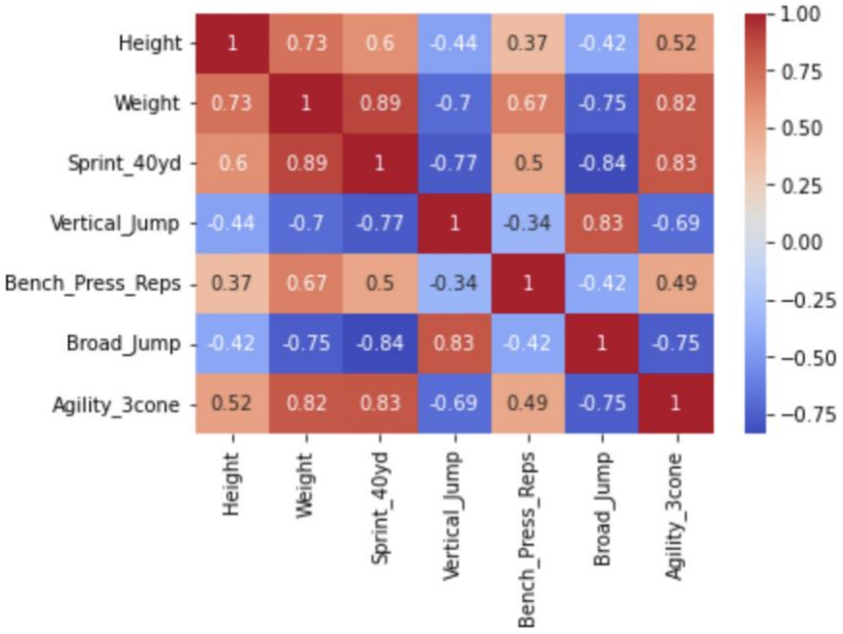
-Below is a list of the columns and description of those columns

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

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.

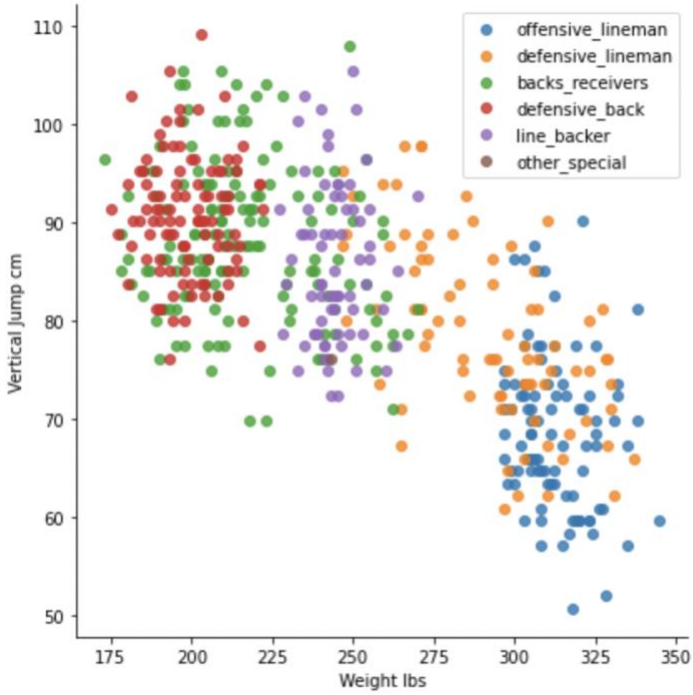
To start, I wanted to see what Combine tests are highly correlated. To do this I created a heat map between all test results.



This heatmap shows the correlation between separate columns. When looking at this heatmap, I was able to see which combine tests have the highest correlation between each other. 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).

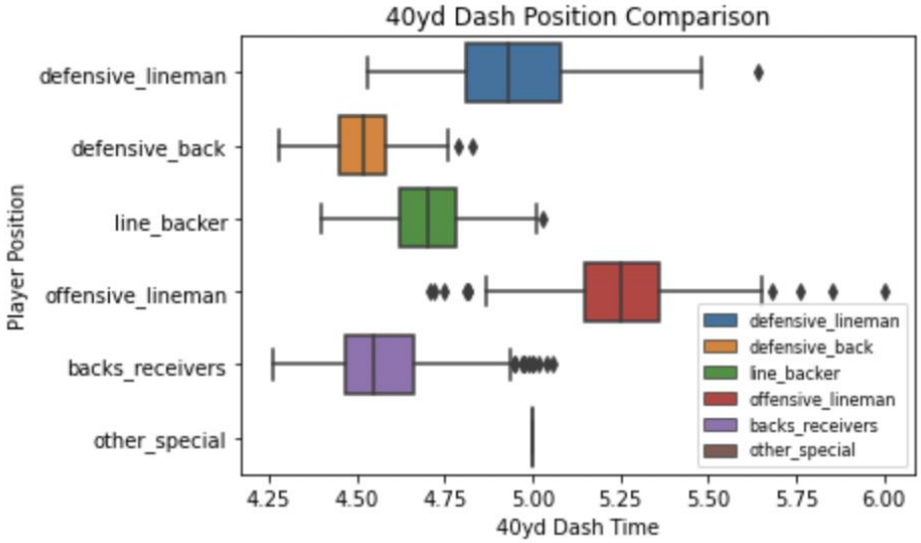
Since the correlation between Vertical Jump and Weight was so low, I decided to visualize it.



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 how much these players weight and how high they jump.

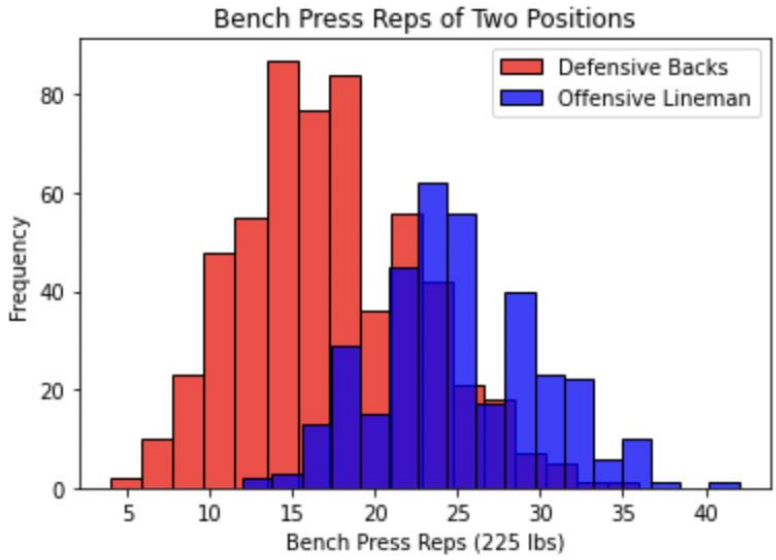
When labeling the individual data points by their position type you can see what position types excel at the vertical jump test. Defensive Backs are jumping the highest while Offensive lineman are jumping the lowest.

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 the slowest, with an average of 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 most athletic players on the field. Their job requires them to defend the largest part of the football field. In order to do that they need to be fast. Offensive Lineman don't need to be as fast and agile. Their position is limited to a small section of the field. You can visualize the different physical attributes 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.



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. 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.

```
Intercept:  -1.2484304143642269
Coefficients:

[('Height', 0.0916955195665836),
 ('Weight', -0.0067698540619845135),
 ('Sprint_40yd', 0.6580078406991537),
 ('Vertical_Jump', -0.0007674805009207579),
 ('Bench_Press_Reps', -0.006384312726777523),
 ('Broad_Jump', -0.0023724903044158325),
 ('Agility_3cone', 0.19592368052702663)]
```

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 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.

```
R squared: 11.25
Mean Absolute Error: 0.3829791635084271
Mean Square Error: 0.1895930599780031
Root Mean Square Error: 0.43542285192442887
```

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. Setting positional requirements aside, the multiple regression model showed your chances or getting drafted are much higher if you perform well on the speed and agility test. 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