NFL Combine Statistics and Draft Pick

Tyler Albert, Jake Angelucci

Introduction

Every year, NFL teams spend tons of time and money scouting out players, hoping it will translate to drafting strong prospects and building a competitive team. But what if the real investment isn't in scouting and interviews but in analytics? By focusing on the combine metrics that truly matter for specific positions, teams could identify hidden gems, draft smarter, and ultimately save money by targeting high-performing players in later rounds. Doing this would save teams millions as the minimum salary difference from the first to the second round is nearly $6 million, according to an article from The Big Lead[1] . Analytics might be able to not only shape better teams, but it could also completely change the economic strategies behind the draft.

In this report we will explore which NFL combine metrics have the most correlation with draft position for skill and non-skill position players, as well as which positions are prioritized in the draft, and how this data could improve draft efficiency. By determining which tests are the strongest indicators of success, teams could prioritize impactful metrics, avoid overvaluing less relevant ones, and find undervalued players in later rounds. Ultimately, analytics has the potential to change the draft process for teams, saving them millions of dollars along the way.

Data

For this project, we collected our data from two main sources: Kaggle[2], which contained our player's name, school, age, height, weight, position, and all their NFL draft combine metrics from all participants from 2009-2019. Our second source of data was Pro Football Reference[3] , which contained the year, the player, what pick in the draft they were, and their position.

2.1 Combine Statistics

To obtain our combine statistics we used a Kaggle dataset called *NFL Combine – Performance Data (2009 – 2019)*. Kaggle is a public platform which holds a bunch of different datasets users can use for various projects. The specific dataset we used contained almost 3,500 unique values for every player who participated in the combine during these years, drafted or not. The dataset was provided on Kaggle as a csv which made it easy to clean and work with.

To clean the data, we removed some unneeded columns, Position_Type which contained the position groups of the players and then Player_Type which said whether the player was on offense or defense. We then cleaned the Player column; it contained a player or ID or something after the name (Ndamukong Suh\SuhxNd99), so we dropped the second part and renamed the column to Player_Name. Next, we replaced missing values for age and performance metrics with

[1] https://www.thebiglead.com/posts/nfl-draft-salary-pick-round-list-01gywhx5ahr3
[2] https://www.kaggle.com/datasets/redlineracer/nfl-combine-performance-data-2009-2019
[3] https://www.pro-football-reference.com/draft/

the position averages. We decided this was reasonable for Age because most of the time every participant in the combine is around the same age, they are coming out of college. We still split age by position though because different positions have different age ranges, for example a decent number of Punters come out of Australia where they attend college at an older age than most Americans. We used the same method for performance metrics because again based on the position certain metrics can be very different, a WRs 40-yd dash will be much faster than a lineman. Also, typically when a player has a missing value for a drill it is not because they performed super well or bad, it is usually because of an injury or certain positions such as QBs just don't do all the drills as some are not relevant to the position.

Next, we replaced undrafted players with a draft position of 999. Doing this would make our final correlations appear a bit weaker than they are, but we only run our model on drafted players. Later, we merge our data frames to include just drafted players, so this won't affect our outcome. The last thing we did before uploading the cleaned data to a data frame was drop our Draft_Round, Draft_Period, Overall_Pick, and Drafted columns since we will get similar information from scraping our other data source.

2.2 Draft Pick

For our draft pick data, we used a website called Pro Football Reference, which houses a bunch of historical data, results, and statistics for football across history. We wrote a code to scrape our data from 11 different web pages, each containing draft data from a different year, 2009-2019.

We stored the scraped data into a csv which had year, overall pick, player name, and position for 2,809 unique players. It makes sense it is less than our previous data because not every player who participates in the combine gets drafted. Last, we reviewed the csv to make sure there were no errors or anything that needed to be cleaned and we found it was all scraped correctly.

2.3 Combining Combine Statistics and Draft Pick

Both data frames we had contained year, and Player Name, so we merged our data into a new data frame on both the Year and Player Name. At first, we were going to just do it on player name, but we realized there are currently two Josh Allens playing in the NFL and this may be the case with others, so we merged on both columns. When merging we wanted to perform our analysis just on players who performed in the combine and were also drafted, so we only included rows with a Player Name and Year matching value for both data frames.

To clean our merged data the first thing we did was remove repeated columns that were in both data frames and therefore were duplicated when merging. The next thing we did was rename some columns just so all the column names in the merged data frame had consistent formatting. Then because we wanted to analyze skill positions and non-skill positions separately, we split our data into two separate data frames for those two categories. The skill positions data frame had 1,170 rows and 16 columns while non-skill positions data frame had 1,038 rows and 16 columns. Together this is 2,208 rows which is less than our draft picks data, but this makes sense because not every player drafted participates in the combine the year they are drafted due to injury or other factors. Our two tables had all the same columns so the descriptions of the variables for both can be found in the Data Dictionary (Table 1) below.

Table 1 – Data Dictionary

Column	*Type*	*Source*	*Description*
Year	Date	Both	Year the player was Drafted to the NFL
Player Name	Text	Both	Name of the Player
Age	Numeric	Kaggle	Age of the Player
School	Text	Kaggle	School the player attended before getting drafted
Height	Numeric	Kaggle	Height of the player (in meters)
Weight	Numeric	Kaggle	Weight of the player (in kilograms)
Sprint_40yd	Numeric	Kaggle	Speed (in seconds) in which the player completed the 40-yard dash
Vertical_Jump	Numeric	Kaggle	Height (in cm) in which the player jumped from a stand-still
Bench_Press_Reps	Numeric	Kaggle	How many repetitions the player successfully completed bench pressing 102.1 Kg
Broad_Jump	Numeric	Kaggle	Distance (in cm) the player jumped forwards from a stand-still
Agility_3cone	Numeric	Kaggle	How fast a player can go in a triangle of cones places 5 yards apart, a test for agility and speed
Shuttle	Numeric	Kaggle	How fast a player can move laterally 5 yards one way, switch directions for 10 yards, and switch again for 5 more yards
BMI	Numeric	Kaggle	A calculation that estimates body fat based on a individual's height and weight

Position	Text	Both	The position the player is identifying as during the combine
Team	Text	Kaggle	The NFL team that the player was drafted to
Overall Pick	Numeric	Pro Football Reference	What number pick in the NFL draft the player was

Analysis

From our analysis we wanted to determine which combine metrics were the most important when looking at NFL draft prospects and their performance in the combine. To do this we investigated correlation between different metrics and draft positions for both skilled and non-skilled positions. Understanding these correlations can help teams optimize their draft strategies by locking in on these key metrics to predict a player's success.

We also explored whether certain positions are more or less prioritized in the early rounds of the draft. This lets us see how teams value different roles and how this affects drafting decisions. This approach allows us to understand not only the individual metrics that matter but also the strategic priorities that shape the draft results.

3.1 Combine Statistics and Draft Pick for Skilled Positions

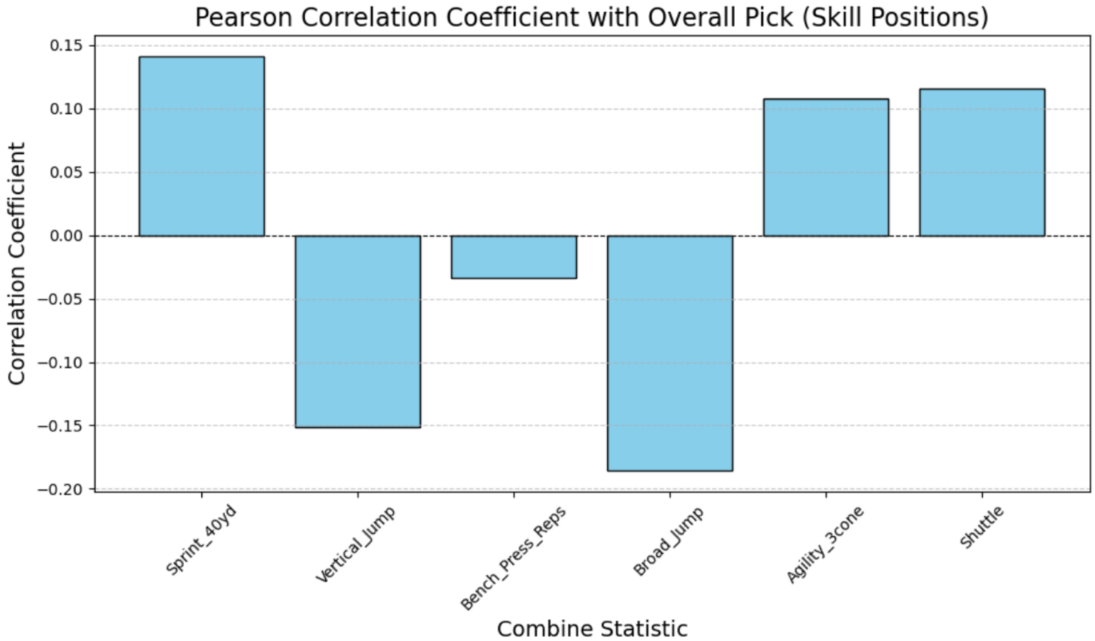
The first part of our analysis aimed to determine the correlation coefficients between each combine metric and the skilled position players' draft picks. When calculating the correlation coefficients, we dropped rows with an N/A value at combine metrics for that given calculation. To do this we used the Pearson correlation coefficient method, our results can be found below in Table 2.

Table 2 – Correlations and P-values for Skill Position Players

Combine Statistic	*Pearson Correlation Coefficient*	*P-value*
40-yard dash	*0.1411*	*1.2571e-06*
Vertical Jump	*-0.1517*	*1.8603e-07*
Bench Press Reps	*-0.0334*	*0.2537*
Broad Jump	*-0.1857*	*1.5560e-10*
3 Cone Agility	*0.1076*	*0.0002*
Shuttle	*0.1152*	*7.7821e-05*

After we had our results, we decided to put them into a chart to show a more clear and visual representation of the relationship between each combine metric and the draft outcome. This makes it easier to quickly compare correlation coefficients, the chart can be seen below in Figure 1.

Figure 1 – Correlation Coefficients for Skilled Position Players



The table and graph show that, due to their highest correlation coefficients, our top three measures for forecasting the overall selection of a skilled position player are broad jump, vertical jump, and 40-yard dash. Broad jump had the strongest correlation coefficient with –0.1857, this shows the moderate negative correlation with draft pick. This means that players with a better (longer) performance tend to be picked earlier in the draft, and the low p-value shows the relationship is statistically significant. Next is vertical jump with a correlation coefficient of –0.1517, this suggests that a higher vertical jump is associated with being selected earlier in the draft. This relationship is also statistically significant indicated by the low p-value. The 40-yard dash had the third-highest correlation coefficient, at 0.1411. The positive correlation shows that faster (lower) sprint times are linked with earlier draft picks. And we have a low p-value showing statistically significant correlation. These results show the importance of explosive power and speed for skill positions in the draft.

3.2 Combine Statistics and Draft Pick for Non-Skilled Positions

For the next part of our analysis, we did the same process but this time for the non-skilled position players. Our results can be found below in Table 3.

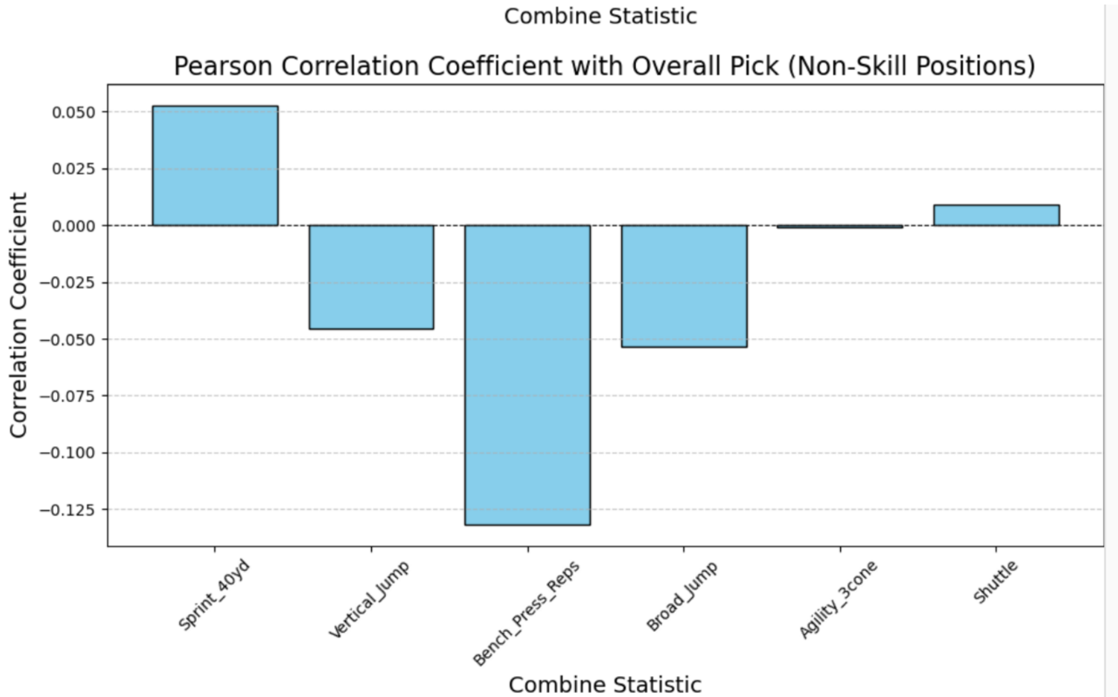
Table 3 – Correlations and P-values for Non-Skilled Position Players

Combine Statistic	***Pearson Correlation Coefficient***	***P-value***
40-yard dash	*0.0527*	*0.0895*
Vertical Jump	*-0.0455*	*0.1433*
Bench Press Reps	*-0.1319*	*2.0221e-05*

Broad Jump	*-0.0534*	*0.0856*
3 Cone Agility	*-0.0010*	*0.9748*
Shuttle	*0.0090*	*0.7747*

Again, to make our results clearer and to give a visual representation we displayed the correlation coefficients in a chart, this can be found below in Figure 2.

Figure 2 - Correlation Coefficients for Non-Skilled Position Players



From the table and graph we can see that because of their strongest correlation coefficients our top 3 metrics for predicting the overall selection for a non-skilled position player are Bench Press Reps, 40-yd Dash, and Broad Jump. Bench Press Reps had the strongest correlation coefficient with –0.1319, this negative correlation indicates that higher bench press reps are associated with an earlier draft pick. We also have a low P-value suggesting the relationship is statistically significant. Our next most significant correlation was for the 40-yard dash, it had a correlation coefficient of 0.0527, this shows that faster sprint times are slightly associated with earlier draft picks. Our p-value indicates it is not a statistically significant relationship, but it is not far above the significance threshold so we may still consider it as an alternative metric. Similarly, the broad jump has a correlation coefficient of –0.0534 which shows a weak negative correlation. This means better broad jump performance is slightly related to an earlier draft pick. Although the P-value shows the correlation is not statistically significant it is near the significance threshold suggesting it could be an alternative metric.

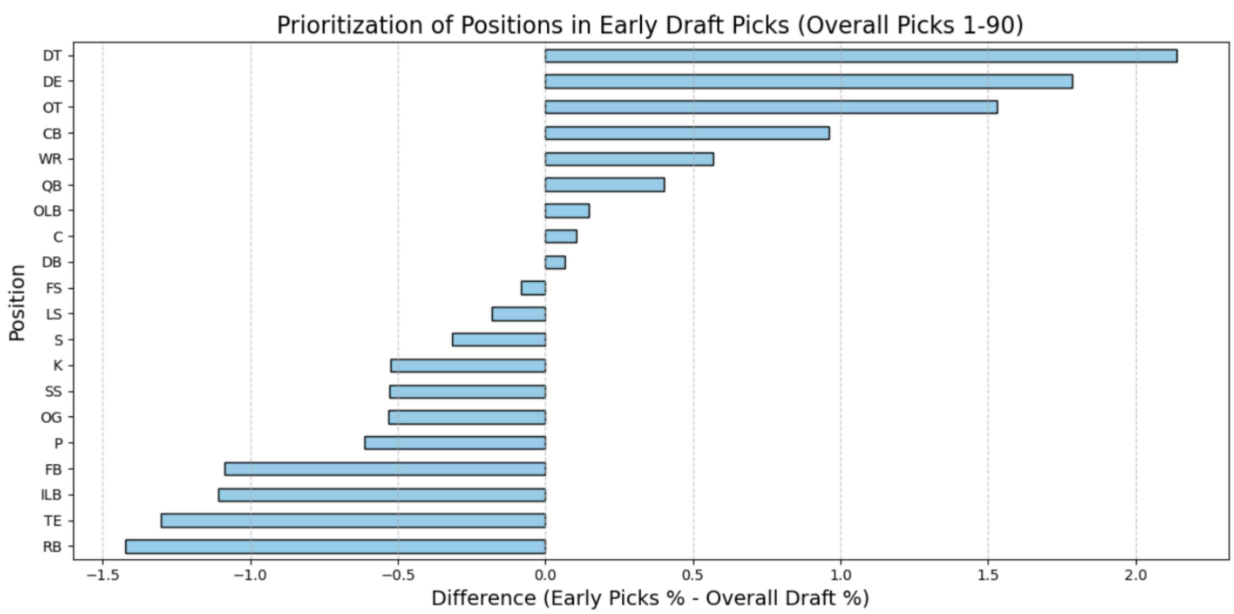
This analysis reveals strength, represented by bench press, is the most significant predictor of draft position of all the combine metrics. However, sprint and broad jump, despite not being statistically significant in this analysis, show relatively low p-values compared to other metrics. These characteristics could potentially serve as a secondary indicator for evaluating prospects, especially when primary metrics are close among potential prospects.

3.3 Positional Prioritization in Early Draft Rounds

In our third question we wanted to explore which positions are prioritized early in the NFL draft when compared to their selection frequency across the entire draft. We decided to use the first 90 picks as an "early pick" because this is approximately the first 3 rounds out of 7, we did not use the exact number of picks because there are times when teams get picks taken away for violating rules, so it is not the same amount every year.

Understanding which positions are prioritized early in the draft is important for a few main reasons. First, it provides insights into the strategies and the preferences of NFL teams when selecting top talent. This can help teams refine their own draft strategies by picking up on trends in position prioritization. This analysis also contributes to a broader understanding of our other questions. By understanding which combine metrics significantly impact draft picks, we can better comprehend how these metrics influence the prioritization of certain positions. This gives a more holistic view of the other factors and strategies that shape draft decisions. To answer this question, we made a graph to show the difference between the early pick percentage and the overall draft percentage for each position, this is shown below in Figure 3.

Figure 3 – Comparing Early Round Emphasis to Overall Draft Trends



From this graph we can see how various positions are prioritized early in the draft relative to the entire draft. The X-axis shows the difference between the percentage of early round selections and selections across the entire draft for each position, which is shown on the Y-axis. In the chart positive values indicate positions that are prioritized in early rounds while negative values show positions that are picked less frequently in early rounds compared to the overall draft.

We can see linemen are prioritized early in the draft, indicating teams think building strong lines is important. Strong linemen are crucial for both offense and defense which explains their high draft value, based on this, teams may prioritize looking at bench press for non-skilled positions early on. We can see that CBs, WRs, and QBs receive early draft attention as well. Our analysis shows speed and athleticism are important for these positions, so teams may prioritize those metrics for these positions over other skill positions early in the draft.

We can see RBs and TEs have the most negative values, showing they are less prioritized in early rounds. So, teams may want to wait to prioritize performance metrics for these skill positions until later on in the draft. There are also some positions like FB, P, C, and LS which have negative or close to zero difference because they have less impact due to other positions, or because traditionally they are not drafted early in the draft, so teams are able to wait to select those positions. For these positions teams can wait to use metrics until later in the draft or may use them after the draft to find an undrafted player.

Conclusion

For this project we analyzed three key aspects for NFL draft prospects to help teams improve their draft strategy and ultimately save millions of dollars. We analyzed the most important combine metrics for skill position players, the most important combine metrics for non-skilled position players, and the prioritization of positions in early NFL draft rounds. From these analysis questions we found the following results:

1. *What are the most important combine metrics for skill position players?*

   We found the broad jump, vertical jump, and 40-yard dash to be the most significant metrics for skill positions. Broad jump and vertical jump showed negative correlations with draft pick indicating higher score is associated with a higher draft pick. The 40-yard dash showed a positive correlation with draft pick meaning lower (faster) times are linked with earlier draft picks.

2. *What are the most important combine metrics for non-skill position players?*

   For non-skilled position players, we found bench press showed a significant negative correlation with draft pick, highlighting the importance of strength. Other metrics like 40-yard dash and broad jump exhibited weaker correlations, but their values were close to the confidence threshold, so they could be considered as alternative metrics.

3. *Which positions are prioritized early in the NFL draft relative to their selection frequency throughout the entire draft?*

   Our analysis identified that linemen are heavily prioritized in early draft rounds, reflecting their crucial roles in building strong lines. While positions such as CB, WR, and QB also receive early attention due to their specialized and impactful roles. We also saw positions like RB, TE, ILB, and special teams' positions are less prioritized early, because they can be effectively filled later in the draft

Understanding which combine metrics are important for certain positions, along with the positional prioritization in early draft rounds provides a holistic view of NFL draft strategies. For skill positions the importance of explosive power and speed is emphasized, while for non-skill positions strength is crucial. The prioritization analysis complements these findings by showing the strategic importance of certain positions in the early rounds, informed by their physical performance metrics.

This project has a few limitations, including potential variability in combine metric performance and draft strategies changing over time. Future projects could involve a larger data set containing metrics about a player's performance throughout their career, which could be used to determine combine metrics' long-term impact on player performance in the NFL. Another limitation is that the model does not understand team-specific draft strategies that change year by year or the context of each draft. For example, the Chiefs, who have probably the best QB in football would not rely on this model for that position in the upcoming draft. Additionally, if there is a generational prospect in a draft, teams would likely prioritize drafting them over following this model, given the opportunity.