40-Yard Dash Times and Draft Placement for NFL Skill Position Players

By Jason Tran (ht2354) and Patricia Guirao (ppg2023)

Abstract:

Our study aimed to explore the importance of the 40-yard dash, the NFL Combine's most important and anticipated event, and how it could influence a "skill-position" players (Quarterbacks [QB], Running Backs [RB], Wide Receivers [WR], and Tight Ends [TE]) placement in the NFL Draft. The research question we wanted to answer was: Does a faster 40-yard dash time correlate with higher draft position and which skill position does it influence the most? To do this we gathered 14 years of 40-yard dash times from the 2009-2023 NFL Combine for the skill positions we wanted and noted down what their 40-yard times were and where they ended up getting drafted. We then conducted a simple data profiling of the results which gave us a table of the average, min, max, std dev 40-yard dash time of each position of each round they were drafted in. We found that: Wide Receivers and Running Backs showed a clear trend where faster 40-yard dash times influenced a higher draft position evidenced by their mean times as well as a lower variability in times between each round — highlighting how important speed is those 2 skill positions. Contrastingly Quarterbacks and Tight Ends, while a faster 40-yard dash was beneficial in terms of earlier round draft placement, there was a lesser emphasis on speed as evidenced by higher average times and a larger variability between rounds compared to Wide Receivers and Running Backs — suggesting that while speed does play a role in being drafted higher, the nature of those positions do not have an emphasis on speed as much as Wide Receivers and Running Backs.

Introduction:

The NFL Combine is a highly anticipated 3-day event, held annually in late February at Lucas Oil Stadium in Indiana, that puts NFL draft prospects through a set of comprehensive physical, medical and psychological evaluations. The results from these tests significantly inform all 32 NFL teams' front offices about the potential draft prospects and oftentimes how well a player tests at the combine can have a significant influence on where they get drafted. The importance of the NFL Combine can not be understated, as millions of dollars can be earned or lost depending on how these prospects perform at the Combine. Being the 1st overall pick comes with a \$27 million signing bonus, while the last player drafted, "Mr Irrelevant" earns the league minimum of \$750,000. Given the importance of the NFL Combine and how much is at stake for these NFL hopefuls, we believed that studying the Combine's most important and watched event — the 40-yard dash — could further enhance our understanding of a player's speed in relation to their draft placement, beyond the simple notion that "faster is always better". Our aim was to

enhance our understanding of the relationship between a player's 40-yard dash time and their draft position, delying into the intricacies such as the most important positions for 40-yard dash times and the average time for each round for each position. We believe that the information and insights gathered from our studies can hugely benefit not only prospects preparing for the NFL Combine, providing them a high-level benchmark of what type of times they should be striving for, but also for youth players, who hope to play beyond the highschool level and coaches across all levels of the sport, who can use these insights to better optimize player development philosophies and strategies. Furthermore, this information can provide NFL General Managers, Owners and Coaches with more insight into the importance of the 40-yard dash, whether or not they should put more emphasis on it or whether it is just a simple test of speed that does not affect how well a player actually performs on the field. The NFL is a multi-billion dollar business for all parties involved: players, coaches, general managers, owners and the league. One player can change the entire direction of a team and a city, generating huge sums of revenues and profits on their own or by contributing to their team's success, and the path to finding that player all begins at the NFL Combine. With millions of dollars at stake and billions of dollars to gain, we believed that uncovering deeper insights about the NFL Combine and the 40-yard dash would be beneficial to all parties involved.

Figure 1.0 (below) illustrates how our group integrated our two data sources and datasets into a single table for analysis. We first utilized Hive to combine the five individual tables from Pro-Football-Reference (2019 - 2023 Combine Result) into one file, while Kaggle's 2009 - 2018 data was already consolidated. Then, Scala was employed for both data sets in order to standardize the format of each combine results table, allowing us to merge them again using Hive which created a cohesive, clean dataset of all NFL Combine results from 2009 - 2023, that was then used for our analytic code and to derive our findings.

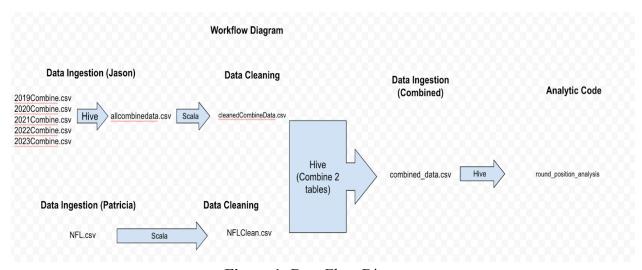


Figure 1: Data Flow Diagram

Motivation:

This study is important because it offers insights into how a player's speed can influence how high they get drafted in the NFL Draft. With a potential salary difference of over \$25 million between a 1st and 7th Round Pick riding on a player's draft position, understanding the physical test outcomes that heavily influence that allows us to help better guide and prepare draft prospects, better informing them to then help maximize their draft position and financial gains.

Jason:

The motivation for this research project comes from my interest in the application of data and analytics into sports, particularly football, my favorite sport. The aim was to provide valuable insights not just for the NFL, but for athletes, coaches, scouts, and fans at various levels, by establishing a quantifiable benchmark for what top-tier athletic performance looks like. I believe that simply having the awareness and understanding of these metrics can help refine training methods for the development of younger players as well as enhancing the likelihood for scouts to predict a player's success in the NFL. Overall, this project was just an exciting opportunity for me to combine my interest in data, analytics, and sports into one.

Patricia:

While I'm not particularly interested in the sports world, including football, I have more recently been interested in the NFL due to one of my all-time favorite artists, Taylor Swift, recently beginning a relationship with Professional Football player Travis Kelce. Although my interest in the NFL was not sparked in a conventional way, since gaining more awareness of the Professional Football world, I've actually become interested in how the process as a whole works, wanting to better understand the differences between each kind of player (and therefore what skills are important for each one to have) and overall, what skills can lead players to be successful in the draft.

Related Works:

The insights and findings from our study: that there does seem to be correlation between a skill position player's 40-yard dash time and where they get drafted, are consistent with and validated by the findings of more extensive published works into this topic matter.

Tucker, Raymond, and Willie Black. "Predictive Validity of the Physical Skills Test of the 40-yard Dash and Draft Placement in the NFL Draft." The Sport Journal, University of Houston at Victoria, 2023. — A study into how the 40-yard dash can influence all positions draft placement — supported our finding and also concluded that there "seems to be a correlation between a faster time in the physical test of the 40-yard dash and draft placement," (Tucker & Black, 2023) and that "significant correlations were found between faster times in the 40-yard dash and its predictive validity to forecast draft placement for the positions of Wide Receiver (WR), Tight End (TE), Offensive Lineman (OL), Linebacker (LB), and Safety (S)," (Tucker & Black, 2023).

Meil, Andrew James. "Predicting Success Using the NFL Combine." Master of Science in Clinical Psychology, California State University, Fullerton, 2018. — A much more extensive research paper into how combine results can be predict NFL Player success rather than draft position — was also able to support our finding that there is a much heavier emphasis and importance on the 40-yard dash time for a Running Back as in the study they found that "The three other position groups that had statistically significant prediction (defensive backs, running backs, and linebackers) highlight the necessity of a player's overall speed and lateral agility to be successful in the professional level" (Meil, 2018).

Description of Datasets:

The first dataset is a combination of datasets from Pro Football Reference. This dataset combines Combine results of participants from the 2019 to 2023 Combine. This dataset details each player's name (Player), position (Pos), school (School), college (College), height in meters (Ht), weight in kilograms (Ht), 40-yard dash time in seconds (40yd), vertical jump height in centimeters (Vertical), bench press repetitions of 225 pounds (Bench), broad jump length in centimeters (Board_Jump), 3 cone shuttle time in seconds (3ConeShuttle), and their team, the round they were drafted, their pick number, and the year of the combine (Drafted (tm/rnd/yr)).

The second dataset is from Kaggle. This dataset contains results of participants from the 2009 to 2019 Combine. This dataset details the year of the combine (Year), each player's name (Player), age (Age), school (School), height in meters (Height), weight in kilograms (Weight), 40-yard dash time in seconds (Sprint_40yd), vertical jump height in centimeters (Vertical_Jump), bench press repetitions of 225 pounds (Bench_Press_Reps), broad jump length in centimeters (Broad_Jump), 3 cone shuttle time in seconds (Agility_3cone), lateral shuttle time in seconds (Shuttle), their team, the round they were drafted, their pick number, and the year of the combine (Drafted..tm.rnd.yr.), body mass index in kg/m^2 (BMI), whether they're an offensive player, defensive player, or special teams player (Player_Type), what kind of broad classification of player they are (Position_Type), specific position (Position), and whether or not they were drafted (Drafted).

While we cleaned our datasets on our own, we made them so they could be combined together when getting our analytics.

Analytic Stages (Process)

Data Profiling (Numerical Data)

On our second dataset, we profiled numerical data using Scala. On the original dataset, we found the mean, median, and mode for each numerical column. Additionally, we found the standard deviation for the 40-yard dash time column.

```
Standard Deviation for Sprint_40yd: 0.30143134275087086
Mean for Column Year: 2013.8236985907392
Mean for Column Age: 21.983259309873592
Mean for Column Height: 1.8739677308024267
Mean for Column Weight: 109.7463934459305
Mean for Column Sprint 40yd: 4.769079624583718
Mean for Column Vertical Jump: 83.39240287769773
Mean for Column Bench Press Reps: 20.241057542768274
Mean for Column Broad_Jump: 291.6296980720278
Mean for Column Agility_3cone: 7.237415929203538
Mean for Column Shuttle: 4.40384253316216
Mean for Column BMI: 31.074416959209085
Median for Column Year: 2014.0
Median for Column Age: 22.0
Median for Column Height: 1.8796
Median for Column Weight: 104.7798375
Median for Column Sprint 40yd: 4.69
Median for Column Vertical Jump: 83.82
Median for Column Bench Press Reps: 20.0
Median for Column Broad_Jump: 294.64
Median for Column Agility_3cone: 7.14
Median for Column Shuttle: 4.36
Median for Column BMI: 30.12262555
Mode for Column Year: 2014.0
Mode for Column Age: 22.0
Mode for Column Height: 1.905
Mode for Column Weight: 95.70799007
Mode for Column Sprint_40yd: 4.5
Mode for Column Vertical Jump: 83.82
Mode for Column Bench_Press_Reps: 19.0
Mode for Column Broad_Jump: 304.8
Mode for Column Agility 3cone: 7.07
Mode for Column Shuttle: 4.28
Mode for Column BMI: 37.36882981
```

Data Profiling (Non-Numerical Data)

On our second dataset, we profiled non-numerical data using first Scala then Hive. Using our original dataset, we first used Scala to parse through the Drafted..tm.rnd.yr. Column and extracted just the team for each player, creating a new column with just each player's team. Furthermore, we dropped all numerical columns. Then, we used Hive to get unique values for each non-numerical column.

```
| Section | Contention | Fances | Fance
```

```
I playe_type |
| glarge_type |
| defense | |
| special_teams |
| special_teams |
| special_teams |
| playe_type |
| special_teams |
| special_teams |
| playe_type |
| special_teams |
| special_teams |
| position_type |
| offensive_lineman |
| ticking_special_tylenan |
| offensive_lineman |
```

Data Ingestion and Cleaning

For our first dataset from Pro Football Reference, the first step was combining the datasets from 2019-2023 using Hive. Then, to clean the dataset, Scala was used. First, we parsed through the Drafted (tm/rnd/yr) column and extracted just the round for each player, creating a new column with just each player's round number. From there, we then dropped all columns except the ones necessary for our analytics (we kept the player name, position, 40-yard dash time, and round).

Original Data Source 1 Example:

Player	Pos	School	College	Ht	Wt	40yd	Vertical	Bench	Broad Jump	3Cone	Shuttle	Drafted (tm/rnd/yr)
Johnathan Abram	s	Mississippi State	College Stats	11-May	205	4.45			116			Oakland Raiders / 1st / 27th pick / 2019

Cleaned Data Source 1 Example:

Player	Pos	40yd	Round	
Jonathan Adams	WR	4.59	Undrafted	

For our second dataset from Kaggle, we underwent a similar process. To clean the dataset, Scala was used. First, we parsed through the Drafted..tm.rnd.yr. column and extracted just the round for each player, creating a new column with just each player's round number. From there, we then dropped all columns except the ones necessary for our analytics (we kept the player name, position, 40-yard dash time, and round).

Original Data Source 2 Example:

Year	Player	Age School	Height	Weight	Sprint_40yd	Vertical_Jump	Bench_Press_Reps	Broad_Jump	Agility_3cone	Shuttle	Draftedtm.rnd.yr.	ВМІ	Player_Type	Position_Type	Position	Drafted
2009	Beanie Wells\WellCh00	20 Ohio St.	1.8542	106.594207	4.38	85.09	25	325.12	NA	NA	Arizona Cardinals / 1st / 31st pick / 2009	31.00419426	offense	backs_receivers	RB	Yes

Cleaned Data Source 2 Example:

Player	Pos	40yd	Round
Beanie Wells	RB	4.38	1

Data Analytics

We combined both data sources using Hive. We did this by creating individual tables for each dataset in Hive and loading each dataset into them. Then, we used Hive to get our analytics. First, we combined both tables in Hive. Then, we created a table for our data analysis – we aimed to get the number of players, mean 40-yard time, standard deviation 40-yard time, minimum 40-yard time, maximum 40-yard time, and number of missing 40-yard times for each position type within each round. Then using corresponding functions in Hive (ex. COUNT, AVG, STDDEV, MIN, MAX), we got our analytics from the combined data and loaded it into our round position analysis table.

Round	Pos	Players Drafted	Mean 40yd Dash	Std Dev 40yd Dash	Min 40yd Dash	Max 40yd Dash	Missing Values 40yd Dash	
1	QB	41	4.738157894736840	0.16381721622714100	4.33	5.04	3	
1	RB	23	4.460454545454550	0.0741299395243533	4.34	4.62	1	
1	TE	13	4.601538461538460	0.1118272124475390	4.42	4.8	0	
1	WR	56	4.434230769230770	0.08051149353797870	4.22 4.61		4	
2	QB	14	4.744285714285710	0.15683085910415400	4.53	5.1	0	
2	RB	36	4.507142857142860	0.07901433607974210	4.34	4.66	1	
2	TE	22	4.717619047619050	0.09596153802459280	4.51	4.88	1	
2	WR	69	4.452727272727270	0.09274509632623140	4.28	4.7	3	
3	QB	16	4.880000000000000	0.21213203435596400	4.52	5.37	1	
3	RB	42	4.527073170731710	0.09200328473560890	4.32	4.67	1	
3	TE	36	4.691764705882350	0.10495014469724300	4.46	4.89	2	
3	WR	70	4.456666666666670	0.09437309040842470	4.26	4.68	1	
4	QB	24	4.823333333333333	0.13034143197344800	4.61	5.11	0	
4	RB	54	4.50622641509434	0.09485294308389570	4.33	4.69	1	
4	TE	32	4.709375	0.11994627703684700	4.45	4.89	0	
4	WR	65	4.473492063492060	0.08696211052619750	4.28	4.67	2	
5	QB	17	4.8425	0.1724275210052040	4.56	5.12	1	
5	RB	39	4.52921052631579	0.09376290308250820	4.36	4.81	1	
5	TE	25	4.73904761904762	0.1375605411967080	4.52	5.04	4	
5	WR	53	4.490377358490570	0.09230716287239430	4.28	4.73	0	
6	QB	21	4.81777777777780	0.17472376787869600	4.47	5.02	3	
6	RB	44	4.547380952380950	0.09411390069407850	4.34	4.73	2	
6	TE	19	4.760000000000000	0.1369914839202300	4.43	4.97	1	
6	WR	60	4.491186440677970	0.080844792622103	4.33	4.66	1	
7	QB	17	4.821764705882350	0.15812263647590100	4.52	5.22	0	
7	RB	29	4.5253571428571400	0.09937239279085320	4.31	4.83	1	
7	TE	27	4.744814814814820	0.1429418522122010	4.4	4.97	0	
7	WR	46	4.513333333333333	0.09444575162494060	4.31	4.72	1	
Undrafted	QB	101	4.813800000000000	0.16110729344135900	4.5	5.26	1	
Undrafted	RB	185	4.602402234636870	0.10577902124612700	4.34	4.84	6	
Undrafted	TE	87	4.810361445783130	0.141126669438577	4.5	5.19	4	
Undrafted	WR	303	4.548892733564010	0.0917127619715195	4.34	4.85	14	

Graphs (Visual Representation of Analytics)

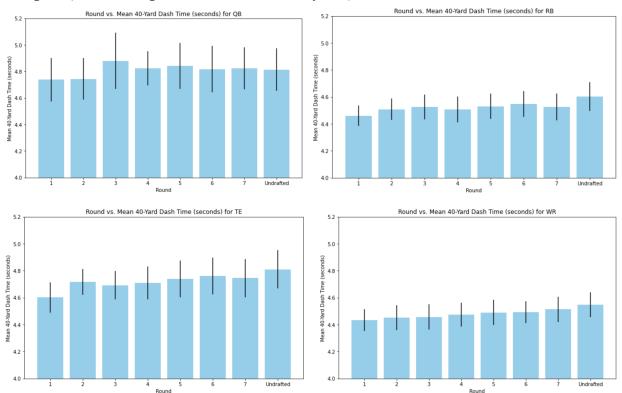


Figure 2: Mean 40-Yd Dash Times Across Rounds For Each Position (With Standard Deviation)

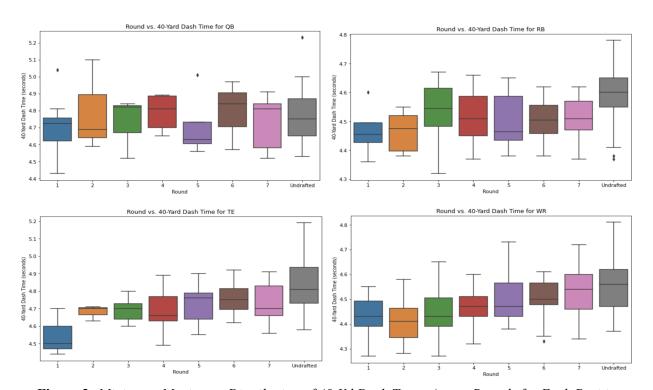


Figure 3: Minimum, Maximum, Distribution of 40-Yd Dash Times Across Rounds for Each Position

Conclusion

Insights:

Wide Receivers: Had the fastest mean 40-yard times across the 4 positions. With the range of average time being 4.43 seconds for a 1st Round Draft Pick to 4.55 seconds for Undrafted Players. Wide Receivers also had the least variability (0.037 seconds) across the board in 40-yard dash times, therefore suggesting a consistent emphasis on speed across all rounds for the position.

Running Backs: There also seems to be a correlation between speed and draft rounds as evidenced by Round 1 Running Backs having an average mean time of 4.46 seconds. and Undrafted Running backs having an average mean time of 4.60 seconds

Quarterbacks and Tight Ends: There is shown to be a lesser emphasis on speed. Quarterbacks show a wider range of acceptable times with means of 4.74 to 4.84 seconds. Tight Ends also exhibit the highest variability (0.061 seconds), indicating a more diverse range of acceptable speeds. While 1st Round QB's also had a large range of acceptable 40 yd dash times. Indicating that while speed can be an important factor, the nature of these positions does not necessarily emphasize speed as much as Wide Receivers and Running Backs.

Among these positions, Wide Receivers and Running Backs show a clear trend where 40-yard dash times are favored in higher draft rounds highlighted by their mean times across all rounds as well as very small variability between each passing round. Contrastingly, Quarterbacks and Tight Ends are evaluated with a much lesser emphasis on speed which given the nature of all these positions make sense.

Implications

These data insights that we got from our draft round-position analysis of 40-yard dash times are first of all potentially most important to those directly involved in the Professional Football world (ex. players, teams) because by knowing whether faster times generally correlate with higher draft picks (or not) and furthermore how that is influenced by each type of position, players themselves can know what skills to focus on training and to what extent (whether or not they should focus on improving speed as opposed to another skill) and teams can know for what positions they should prioritize what kind of skill. Furthermore, our data insights can be somewhat important to those just generally interested in the Professional Football world (but not directly involved in it) as they can be better informed how Combine results affect draft results.

Importance

Our findings are significant because Professional Football is both a difficult (but desired) field to be in and one that holds great interest here in the United States. By increasing awareness of what skills are potentially important to be successful in this world, players are able to better prepare themselves to do well, teams are able to better curate their roster, and viewers are able to better understand why certain things happen.

Other Angles

Given the extensive data collection at each combine, there are various other expanded ways of thinking about our research problem. For example, since there are several other kinds of tests conducted at the Combine, an expanded way of thinking about our research problem is considering how other tests (and therefore skills) affect draft rounds (for example, considering how the number of bench press repetitions affects draft round could tell us how strength/endurance affect draft round for each kind of position). Another way we can also think about our research problem is seeing how the results of other tests that also test speed in some way affect draft round (for example, lateral shuttle time).

References:

Tucker, Raymond, and Willie Black. "Predictive Validity of the Physical Skills Test of the 40-yard Dash and Draft Placement in the NFL Draft." The Sport Journal, University of Houston at Victoria, 2023.

https://thesportjournal.org/article/predictive-validity-of-the-physical-skills-test-of-the-40-yard-dash-and-draft-placement-in-the-nfl-draft/

Meil, Andrew James. "Predicting Success Using the NFL Combine." Master of Science in Clinical Psychology, California State University, Fullerton, 2018. https://scholarworks.calstate.edu/downloads/1z40kt64t