PREDICTIVE VALUE OF NFL COMBINE EVENTS FOR NFL QUARTERBACKS AND RUNNING BACKS

David Deniziuk and Jackson Filosa

Table of Contents

Introduction

Every year, the NFL Draft serves as an opportunity for NFL teams to select their next generation of players from the best that college football has to offer. Whether a team is bouncing off a winless season and attempting to sign a franchise-changing player or won the Super Bowl and needs to replace expensive free agents, the Draft provides opportunities for every team to build for the future. The Draft consists of seven rounds and 259 picks, with the worst teams in the previous season getting chances to select first in order to promote league parity (barring any trades). With limited picks and one chance to get it right, the NFL Draft can be a make-or-break event for the future decades of any NFL Team. It is an irreplaceable opportunity to sign inexpensive, young, and coachable players to a roster. A sleeper pick taken in the late rounds can turn into a superstar for a franchise, just as a bust taken in the early rounds can haunt a franchise for years to come. As Robert Kraft, owner of the New England Patriots, summarized: “In the end, if you want to have a good, consistent, winning football team, you can't do it in free agency. You have to do it through the draft” (Yang, 2021).

With the Draft being of the utmost importance, teams invest heavily in scouting and analytics to determine which players to select. In addition to weighing college performance, teams will factor in results from the NFL Combine to help them make their picks. The NFL Combine is an event that takes place before the NFL Draft every year, and tests prospective players in various forms of athletic feats. An outstanding performance at the NFL Combine can put a player from a poor college program on a scout’s radar, just as a lackluster performance can signify an alarming lack of athleticism and poise. Most importantly, it establishes another standard metric through which players can be evaluated and compared. Since its inception, there has been much controversy about the value of the NFL Combine. Common objections are that while it tests for athleticism, it does not emphasize football instincts and rewards players for their work in the weight room as opposed to the football field. A Bleacher Report article by Ty Schalter makes the case that many of the Combine's top performers underachieve in the NFL, and that the Combine is certainly "not a divining rod" for predicting a players NFL success (2013). For quarterbacks especially, less stock is put into the combine with critics saying that while baseline athleticism is important, even more so is the ability to read a defense, choose the right throw, and make accurate passes, which does not always come through from tests for speed, strength, and quickness.

In "On the Predictive Efficiency of Past Performance and Physical Ability: The Case of the National Football League," researchers from the University of Georgia found that college performance was a better indicator of NFL performance than the Combine. One notable exception to this found in the study was that performance in the 40-yard dash had some predictive value for NFL performance (Lyons et al., 2011). Building off this previous effort, our research question for analysis was: Which NFL Combine Events correlate with NFL Season Performance for quarterbacks and running backs? We sought to investigate whether anything had changed since this 2011 study, and whether quarterbacks specifically were rewarded for being better athletes in the changing landscape of an NFL that has featured more dynamic, running quarterbacks. For comparison, we also sought to investigate whether running backs would have differing events as significant predictors of NFL performance, as the position requires an entirely different skill set with more emphasis on power and speed.

Data Collection

For our analysis, we used three datasets found on Kaggle, a well-recognized repository for free and open data sharing. Our first dataset includes NFL Combine performances for all players from 2000 to 2018. This includes the results of the physical tests in the combine, and the round that these players were drafted in. Our second dataset contains metrics of quarterback in-season performance from 1996 to 2016, and our last contains metrics of running back season performance from all seasons of the NFL.

Data cleaning was consolidated into two workflows, which resulted in two separate final data frames: one for quarterbacks and one for running backs. In both cases, we started by filtering the combine data for observations which contained the necessary predictors. This means we selected for prospective players who completed the necessary combine drills for our analysis. We then removed unnecessary columns from the dataset.

After cleaning the data frame of predictor variables, we then started to filter the response data frames. This first involved filtering for the correct years so that this data lines up with the predictor data. We then transformed observations from player games to player years by taking averages and sums of required variables. For both quarterbacks and running backs, we only selected player seasons with at least 10 games, so that non-starting players or players who got injured would be removed.

At this point, we then merged the predictor data frame with the response data frame for both QBs and RBs. In both cases, the name variable being used to join was not formatted the same across data frames, so we utilized the agrep() function to do an approximate string match join. For the running back portion, we had to alter agrep()'s tuning parameter, string length, in order to match observations with relative success. Upon completing these joins after much experimentation and hard work, our final two data frames were created.

Descriptive Statistics

Using R, we were able to obtain visual representations of the data we were working with to aid in our analysis. The response variable we sought to investigate for quarterbacks was quarterback rating, or QBR, which is a metric developed by ESPN that uses a proprietary but respected algorithm to rate a quarterback's individual game performance on a scale of 0-100. For further comparison in the study, we decided to calculate a quarterback's average QBR across their entire career, as well as the QBR of their first season after the Combine. This could provide information about whether certain Combine events were able to predict which quarterbacks were ready to play immediately, while also allowing an average QBR variable to account for any above average or below average seasons a quarterback had to regress towards their mean performance. The distributions for these variables can be seen in Table 1 and Figure 1 below:

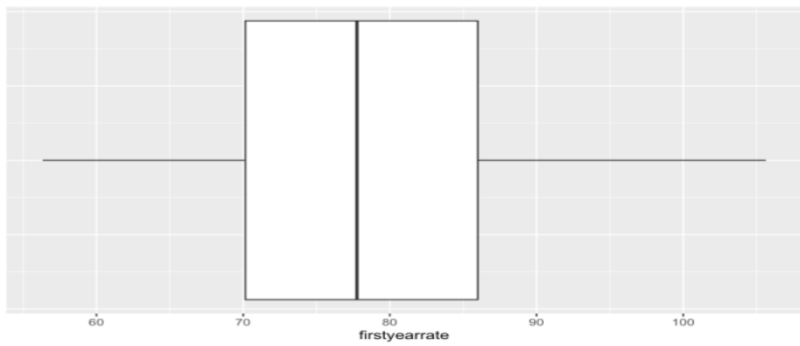
Table 1

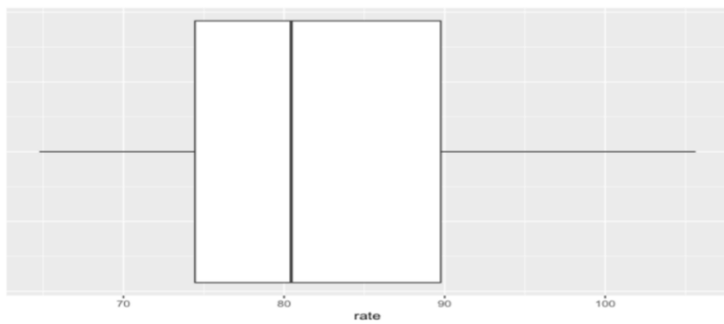
Summary Statistics of Quarterback Response Variables

Variable	Mean	Standard Deviation
First Season QBR	79.17	12.07
Career QBR	82.85	10.10

Figure 1

Boxplots Describing Average QBR (rate)





Note: Boxplots obtained using R.

As the visuals demonstrate, the mean for Career QBR (82.85) is higher than the mean for first season QBR (79.17), which makes sense as players will either improve throughout their careers or get replaced after a bad first season.

Next, we analyzed the predictor variables from the NFL Combine for quarterbacks. The first predictor was a player's 40-yard dash time, which measures the time in seconds that it takes a player to run forty yards. The next predictor variable was the broad jump, which measures the distance, in inches, a player can jump horizontally. The next predictor was the vertical jump, which measures the distance in inches a player can jump straight into the air. The final two predictors were the cone drill, which tests in seconds a player's ability to make rapid changes of direction between cones, and the shuttle run, which tests a player's time in seconds taken to move laterally across the field in different directions. The only predictor within our data that we did not use was the bench press, since many quarterbacks had missing values in the data (most quarterbacks can opt out of this event which is geared more towards linemen and defensive tackles). Table 2 shows the summary statistics for these predictor variables for quarterbacks, and Figure 2 shows the distribution of these variables in histograms.

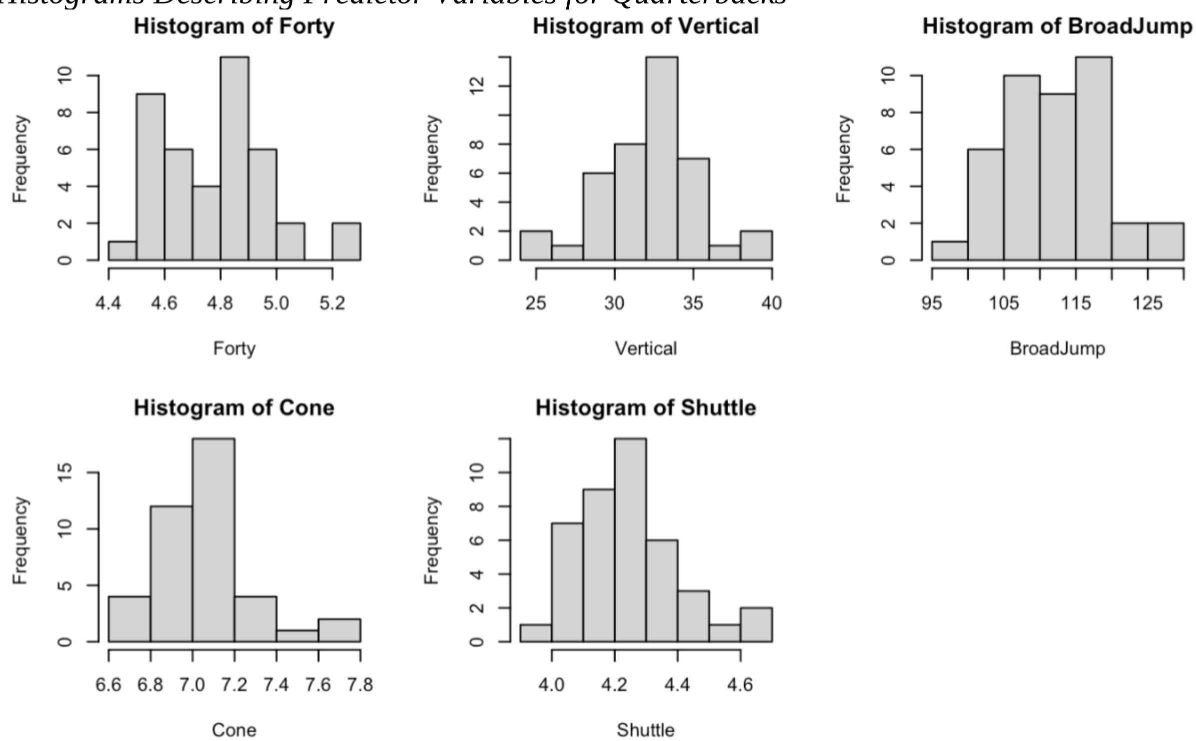
Table 2

Summary Statistics of Quarterback Predictor Variables

Variable	Mean	Standard Deviation
Forty (seconds)	4.77	0.19
Vertical (feet)	32.38	3.24
Broad Jump (feet)	112.41	6.85
Cone (seconds)	7.07	0.23
Shuttle (seconds)	4.25	0.16

Figure 2

Histograms Describing Predictor Variables for Quarterbacks



Note: Histograms obtained using R.

As is evidenced in Figure 2, most of the histograms are approximately normally distributed, with the exception of the forty-yard dash which is almost bimodal and peaks in the faster times as well.

We repeated this same process for running backs using the same predictor variables but different response variables. Instead of QBR as a response variable, we used Yards Per Carry, or the average amount of yards a running back will bring the ball each time they receive it.

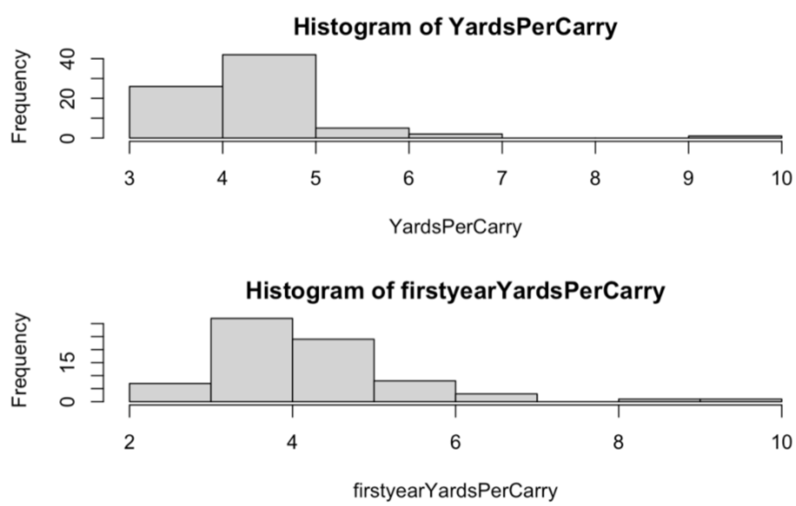
Table 3

Summary Statistics of Runningback Response Variables

Variable	Mean	Standard Deviation
First Season YPC	4.22	1.21
Career YPC	4.26	0.87

Figure 3

Histograms Describing Response Variables for Runningbacks



Note: Histograms obtained using R.

Table 3 and Figure 3 again demonstrate that Yards Per Carry increases after the first season for running backs as they continue to improve, and that standard deviation decreases. Next, Table 4 and Figure 4 display the summary statistics for the running back predictor variables from the NFL Combine, which were the same ones used for quarterbacks.

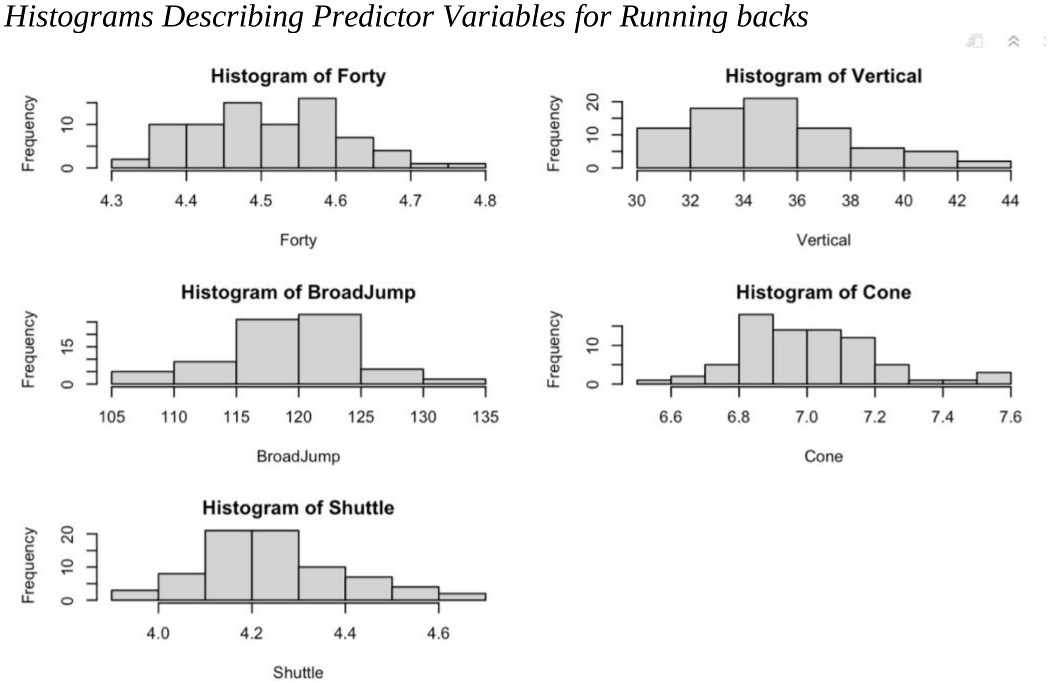
Table 4

Summary Statistics of Runningback Response Variables

Variable	Mean	Standard Deviation
Forty (seconds)	4.51	0.10
Vertical (feet)	35.36	3.07
Broad Jump (feet)	119.79	5.45
Cone (seconds)	7.01	0.20
Shuttle (seconds)	4.26	0.16

Figure 4

Histograms Describing Predictor Variables for Running backs



Note: Histograms obtained using R.

The means are all slightly improved from the quarterback means for these events, indicating that the running backs are more athletic on average. The histograms are approximately normally distributed. After visualizing and confirming no glaring errors with these descriptive statistics, we proceeded with our analysis.

Inferential Statistics

To answer our research question, we used the step model function within R, which utilizes AIC (Akaike Information Criterion) values to determine the optimal model given a set of possible predictors. Step AIC can be seen as an alternative to using a step function based on partial f-tests, which commonly uses thresholds of p=0.05 and p=0.15 for the addition and removal of predictors. AIC penalizes the addition of non-significant predictors into a model, while still selecting for significant predictors. The step model function repeatedly adds and drops predictors in a stepwise manner to select the set of predictors which minimizes the AIC value. This AIC-minimized model is then used as our final model. The result of the summary statistics for the stepwise regression for quarterbacks is shown below in Figure 5 (for average QBR) and Figure 6 (for first year QBR):

Figure 5
Quarterback Avg QBR Model Summary

```
Call:
lm(formula = rate ~ 1, data = final)

Residuals:
    Min      1Q  Median      3Q     Max 
-18.089  -8.399  -2.405   6.911  22.783 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   82.849      1.577   52.53   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.1 on 40 degrees of freedom
```

Figure 6

Quarterback First Year QBR Model Summary

```
Call:
lm(formula = firstyearrate ~ 1, data = final)

Residuals:
    Min      1Q  Median      3Q     Max 
-22.839  -9.022  -1.419   6.818  26.456 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   79.175      1.885      42   <2e-16 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.07 on 40 degrees of freedom
```

As is evidenced in both model summaries, none of the predictors from the Combine were significant in determining either average of first year QBR. In other words, the step model returned no significant predictors.

For running backs, we repeated the same process and obtained the following model summaries depicted in Figure 7 and Figure 8. Figure 7 displays the step model for running backs with first year yards per carry as a response variable, and Figure 8 displays the step model for running backs with average career yards per carry as a response variable.

Figure 7

Running back First Year YPC Model Summary

```
Call:
lm(formula = firstyearYardsPerCarry ~ 1, data = finalRB)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.7278 -0.7017 -0.2331  0.3273  5.1447 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.2153     0.1388   30.37   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.21 on 75 degrees of freedom
```

Figure 8

Running back Avg. Career YPC Model Summary

```
Call:
lm(formula = YardsPerCarry ~ Forty + BroadJump, data = finalRB)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.2034 -0.4139 -0.1120  0.2648  4.5312 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)   
(Intercept) 20.55166    6.09678   3.371   0.0012 **
Forty       -2.58074    1.07332  -2.404   0.0187 * 
BroadJump   -0.03865    0.01931  -2.002   0.0490 * 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8453 on 73 degrees of freedom
Multiple R-squared:  0.08986,   Adjusted R-squared:  0.06492 
F-statistic: 3.604 on 2 and 73 DF,  p-value: 0.03217
```

For the first model (which used first year yards per carry as a response variable), again we found that no predictors from the combine were significant, as the model returned zero significant predictors. However, when using average career yards per carry as a response variable, both the forty-yard dash and broad jump categories were returned as significant predictors. The forty-yard dash had a negative coefficient of -2.58 and a p-value of .0187, while the broad jump had a negative coefficient of -.03 and a p-value of .049.

Discussion/Conclusion

For quarterbacks, our analysis supports the pre-existing research that NFL Combine performance will not determine the success of a quarterback in the NFL. This can provide more incentive to teams to overlook sheer athleticism when searching for their next franchise quarterback and focus instead on their college performance. Ultimately, our analysis shows QBR is not dependent on NFL Combine performance, and will likely be higher for more precise and accurate quarterbacks who play with finesse, even if they are not the most physically imposing athletes. As anecdotal evidence of this, one needs look no further than Tom Brady, who is infamous for posting one of the slowest forty-yard dash times in combine history, but has gone on to win seven Super Bowls.

For running backs, however, our analysis shows that certain Combine events are significant predictors of success when examining their Yards Per Carry over their entire careers. The negative coefficient of -2.58 in our final model implies that for every additional second for a running back in the forty-yard dash, our model predicts 2.58 less yards per carry in the NFL over the course of a career. This makes sense as one second is a very long addition for a forty-yard dash (the standard deviation was 0.10 seconds), and so it makes sense that being this much slower would lead to a great decrease in average yards per carry (for which the standard deviation was (.87 yards). It also makes sense that faster players will average more yards per carry. Interestingly, the broad jump also had a negative coefficient (-.039), which means that for every increase of 1 inch in a player’s broad jump, we expect a running back’s yards per carry to decrease by .039 yards. This is surprising, as it penalizes players for better broad jump performances. It is worth noting the p-value of this is as high as it can be while remaining significant (.049). A potential explanation of this is could be that a long broad jump might be correlated with having longer legs, which could result in a bigger running back that is slower and easier to tackle.

These results could have repercussions for teams when evaluating running backs at the combine. Scouts and coaches might focus their attentions to the 40-yard dash times of their potential running backs while not putting as much stock into the other events and completely disregarding the broad jump.

Another interesting ramification of our analysis was that these predictors were only significant when predicting Yards Per Carry over an entire career, which means that the predictors are not significant in a running back's first season. This means that the Combine does not provide any significant predictors for a running back's first season, which is helpful for teams to know who are attempting to win immediately. While these predictors might indicate long term success, these teams are probably better off looking at college statistics as well.

Limitations faced by this study were the difficulties of obtaining data from more recent NFL seasons. Unfamiliar with more advanced data scraping techniques, we were limited only to what data we could already find available. It would have been interesting to see the results on a current analysis of the same procedure, as even more mobile quarterbacks such as Kyler Murray and Lamar Jackson have taken over the NFL, and determine whether these predictors in the Combine now hold more value. The data we used was only able to take into account quarterbacks and running backs through 2016. Another limitation of the study is that we had a very small sample of quarterbacks who met all the criteria. A final limitation of our study was the time we had to complete it, which prevented us from exploring analysis in more positions such as wide receiver, linemen, defensive tackle, and more.

Suggestions for further study would be to examine these additional positions to explore what positions are predicted most accurately by the combine. Furthermore, being able to incorporate new events in the Combine that test a quarterback's throwing ability (which were not included in our data set) could be interesting to examine. Lastly, it would be interesting to see if the results change when using a metric other than QBR (such as Yards Per Attempt or Completion Percentage) to evaluate quarterback performance.

Ultimately, we determined that the NFL Combine has some value to scouts for the running back position over their entire career, but based on our research, for quarterbacks within our data, their athletic metrics at the combine were insignificant to their NFL success.

Bibliography:

Lyons, B. D., Hoffman, B. J., Michel, J. W., & Williams, K. J. (2011). On the Predictive Efficiency of Past Performance and Physical Ability: The Case of the National Football League. *Human Performance*, *24*(2), 158–172. https://doi.org/10.1080/08959285.2011.555218

Schalter, T. (2013). *Does the NFL Combine Have Any Predictive Value to NFL Success?* Bleacher Report. Retrieved November 4, 2021, from https://bleacherreport.com/articles/1535606-does-the-nfl-combine-have-any-predictive-value-to-nfl-success

Yang, N., (2021). *Patriots owner Robert Kraft reiterates the importance of this NFL Draft—The Boston Globe*. BostonGlobe.Com. Retrieved November 6, 2021, from https://www.bostonglobe.com/2021/04/29/sports/patriots-owner-robert-kraft-reiterates-importance-this-nfl-draft/

Data Sets Used:

NFL Combine Data:

https://www.kaggle.com/savvastj/nfl-combine-data

NFL Quarterback Data:

https://www.kaggle.com/speckledpingu/nfl-qb-stats?select=QBStats_all.csv

NFL Running Back Data:
https://www.kaggle.com/kendallgillies/nflstatistics?select=Game_Logs_Runningback.csv