



The 40-Yard Dash

Exploratory Data Analysis

NFL Combine and Pro Data from 1987 -2021

Which (if any) metrics in the NFL Combine signal a strong 40-yard dash?



Lauren Chlebove • Shweta Ale • Tsering Lhamo • Philip Park





Introduction

The **NFL Combine** and **Pro Day Data** tracks key metrics and attributes that professional coaches and scouts deem to be crucial in making decisions on whom they should draft to their organization.

The NFL Combine is an invitation-only showcase event held annually at Lucas Oil Stadium in Indianapolis, Indiana for draft-eligible college athletes. These athletes are considered the top football prospects in the country.

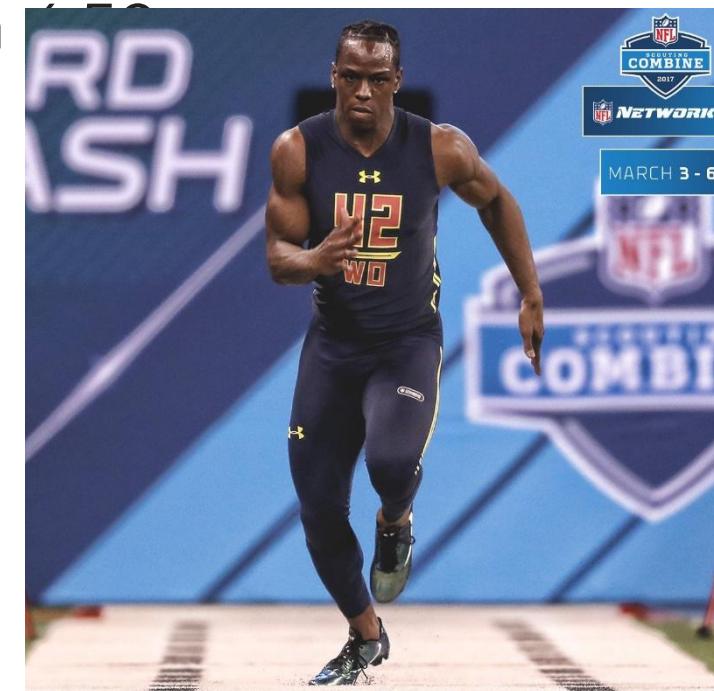
In addition, colleges and universities will hold their own "combine" in which their student athletes get to run a battery of tests in the same ilk as the NFL Combine.

Scouts are invited to assess these athletes. Within

Dash

The crowning jewel metric of the NFL Combine is the **40-yard dash**. The analysis for this dataset will revolve around identifying attributes that positively correlate to a fast 40-yard dash. Note that lower numbers for the speed categories (40-yard, Shuttle, and Cone) are viewed as high-performing.

For example, a 40-yard time of 4.30 is viewed more favorably than a



*Record-holder
John Ross ran
a 4.21 s / 40
yd dash*





Hypothesis

If an athlete posts an above average 40-yard dash time, we should expect to see an above average shuttle and 3 cone performance.

40_{Response Variable} Yard (sec)

Year

The year of the combine.

Wonderlic

College

The college the athlete attends.

The Attributes
POS

Bench Press

An assessment used to measure cognitive and

problem-solving skills.

Shuttle 3-point stance, 5 yards in one direction, 10

yards backwards in the other direction, lastly

5 yards through the starting point.

Height (in)

Vert Leap (in)

Weight (lbs.)

Broad Jump

3 Cone

A drill that consists of five-yard sprints and turns, that uses 3 cones to guide the athlete as they complete the L-shaped run.



Data Cleaning and Feature Engineering



Explain this here.

```
▶ import pandas as pd  
import seaborn as sns  
import numpy as np  
import matplotlib.pyplot as plt  
from sklearn.preprocessing import StandardScaler  
  
df = pd.read_csv("https://raw.githubusercontent.com/josedv82/public_sport_science_datasets/main/NFL%20Combine/NFL%20Combine%20and%20pro%20day%20data%20(1987%20-%202021).csv")  
print(df.shape)  
print(df.tail())
```

```
👤 (13230, 13)  
Year Name College POS Height (in) Weight (lbs) \  
13225 1987 Rod Woodson Purdue CB 72.0 202  
13226 1987 John Wooldridge Ohio State RB 68.4 193  
13227 1987 Dave Wyman Stanford ILB 74.0 235  
13228 1987 Theo Young Arkansas TE 74.0 231  
13229 1987 Jeff Zimmerman Florida OG 75.4 341
```

```
Wonderlic 40 Yard Bench Press Vert Leap (in) Broad Jump (in) \  
13225 NaN 4.33 10.0 36.0 125.0  
13226 NaN NaN NaN NaN  
13227 NaN 4.79 23.0 29.0 118.0  
13228 NaN 4.89 9.0 30.0 107.0  
13229 NaN 5.36 22.0 22.5 NaN
```

```
Shuttle 3Cone  
13225 3.98 NaN  
13226 NaN NaN  
13227 4.30 NaN  
13228 4.20 NaN  
13229 4.83 NaN
```

```
[3] #Add a metric description to a few column names for easier read.  
df = df.rename(columns={'40 Yard': '40 Yard (sec)', 'Shuttle': 'Shuttle (sec)', '3Cone': '3Cone (sec)'})
```

```
[4] #Here we are identifying the total number of nulls for each category.  
df.isnull().sum()
```





Info

- Put stuff in

Info





Info

Info

Explain

Put stuff in

Info

