

```
In [1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
```

```
In [2]: data = pd.read_csv("ZCruit_Cleaned_Data.csv")

/Users/ryanaihara/anaconda3/lib/python3.7/site-packages/IPython/core/interactiveshell.py:3063: DtypeWarning: Columns (15,17,19,21,23,25,27,29,31,33) have mixed types.Specify dtype option on import or set low_memory=False.
  interactivity=interactivity, compiler=compiler, result=result)
```

```
In [3]: data
```

Out[3]:

	Full Name	Last Name	Grad Year	Position	Height	Height Source	Weight	Weight Source	Zcruit Rating	Committed To	...	Hand Size	Powerball Toss
0	Korey Foreman	Foreman	2021	DE	6' 4"	nike_opening	248.0	nike_opening	96.6970	USC	...	NaN	
1	J.T. Tuimolou	Tuimolou	2021	DE	6' 4"	national_preps	275.0	national_preps	96.6970	NaN	...	NaN	
2	JC Latham	Latham	2021	OT	6' 6"	national_preps	279.0	nike_opening	96.6970	Alabama	...	NaN	
3	Emeka Egbuka	Egbuka	2021	WR	6' 0.5"	nike_opening	190.0	nike_opening	96.3939	Ohio State	...	national_combine	
4	Dallas Turner	Turner	2021	DE	6' 2.5"	rivals_camp	223.0	rivals_camp	96.3939	Alabama	...	rivals_camp	
...	
71365	Carl Collins III	Collins III	2022	WR	4' 8"	rivals	86.0	rivals	NaN	NaN	...	NaN	
71366	Antrone\t Coston	Coston	2024	WR	5' 1"	rivals	85.0	rivals	NaN	NaN	...	NaN	
71367	Gerald\t Lampkin	Lampkin	2024	WR	4' 11"	rivals	84.0	rivals	NaN	NaN	...	NaN	
71368	Jaeden\t Hamm	Hamm	2024	WR	4' 10"	rivals	80.0	rivals	NaN	NaN	...	NaN	
71369	Eugene Thomas	Thomas	2022	RB	4' 9"	rivals	71.0	rivals	NaN	NaN	...	NaN	

71370 rows x 39 columns

```
In [4]: data.columns
```

```
Out[4]: Index(['Full Name', 'Last Name', 'Grad Year', 'Position', 'Height', 'Height Source', 'Weight', 'Weight Source', 'Zcruit Rating', 'Committed To', 'Commitment Date', 'Signed To', 'Offers', 'Visits', 'Forty Yard Dash', 'Forty Yard Dash Source', 'Shuttle', 'Shuttle Source', 'Vertical Jump', 'Vertical Jump Source', 'Broad Jump', 'Broad Jump Source', '3 Cone', '3 Cone Source', 'Wingspan', 'Wingspan Source', 'Arm Length', 'Arm Length Source', 'Hand Size', 'Hand Size Source', 'Powerball Toss', 'Powerball Toss Source', 'SPARQ Rating', 'SPARQ Rating Source', 'GPA', 'ACT', 'SAT', 'Unnamed: 37', 'Unnamed: 38'], dtype='object')
```

```
In [5]: clean = data.drop(columns = ['Full Name', 'Height Source', 'Weight Source', 'Committed To', 'Commitment Date', 'Signed To', 'Offers', 'Visits', 'Forty Yard Dash Source', 'Shuttle Source', 'Vertical Jump Source', 'Broad Jump Source', 'Broad Jump Source', '3 Cone Source', 'Wingspan Source', 'Arm Length Source', 'Hand Size Source', 'Powerball Toss Source', 'SPARQ Rating Source', 'GPA', 'ACT', 'SAT', 'Unnamed: 37', 'Unnamed: 38'])
```

```
In [6]: clean
```

Out[6]:

	Last Name	Grad Year	Position	Height	Weight	Zcruit Rating	Forty Yard Dash	Shuttle	Vertical Jump	Broad Jump	3 Cone	Wingspan	Arm Length	Hand Size	Powerball Toss	SPARQ Rating
0	Foreman	2021	DE	6' 4"	248.0	96.6970	4.880	4.700	30.4	104.0	7.758	NaN	NaN	NaN	44.5	
1	Tuimolou	2021	DE	6' 4"	275.0	96.6970	4.750	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	Latham	2021	OT	6' 6"	279.0	96.6970	5.460	4.870	26.8	NaN	NaN	NaN	NaN	NaN	NaN	
3	Egbuka	2021	WR	6' 0.5"	190.0	96.3939	4.859	4.200	35.1	112.0	7.083	75.0	NaN	9.00	41.0	
4	Turner	2021	DE	6' 2.5"	223.0	96.3939	4.991	4.487	26.0	115.0	7.752	83.5	34.25	9.84	NaN	
...	
71365	Collins III	2022	WR	4' 8"	86.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71366	Coston	2024	WR	5' 1"	85.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71367	Lampkin	2024	WR	4' 11"	84.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71368	Hamm	2024	WR	4' 10"	80.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71369	Thomas	2022	RB	4' 9"	71.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

71370 rows x 16 columns

```
In [7]: clean['Position'].value_counts()
```

```
Out[7]: WR      12359
RB       9080
LB       8832
QB       6345
DE       4229
CB       3976
OL       3781
DT       2994
S        2871
OT       2599
OG       2421
TE       2301
ATH      2247
DB       2172
DL       2060
OC       1304
K         783
FB        326
LS        277
P         165
Name: Position, dtype: int64
```

```
In [8]: """
for i in ['WR', 'RB', 'LB', 'QB', 'DE', 'CB', 'OL', 'DT', 'S', 'OT', 'OG', 'TE', 'ATH', 'DB', 'DL', 'OC', 'K', 'FB', 'LS', 'P']:
    clean_ = clean[clean['Position'] == i]
"""
#FAILED ATTEMPT, CORRECT ATTEMPT (WITH DICTIONARIES) BELOW
```

```
Out[8]: "\nfor i in ['WR', 'RB', 'LB', 'QB', 'DE', 'CB', 'OL', 'DT', 'S', 'OT', 'OG', 'TE', 'ATH', 'DB', 'DL', 'OC', 'K', 'FB', 'LS', 'P']:\n    clean_ = clean[clean['Position'] == i]\n"
```

```
In [9]: list_of_positions = ['WR', 'RB', 'LB', 'QB', 'DE', 'CB', 'OL', 'DT', 'S', 'OT', 'OG', 'TE', 'ATH', 'DB', 'DL', 'OC', 'K', 'FB', 'LS', 'P']
```

```
In [10]: d = {}
for position in list_of_positions:
    d[position] = pd.DataFrame()
#or simply
#d = {position: pd.DataFrame() for position in list_of_positions}
```

```
In [11]: #for me to look at what is iterated through
"""
for position, df in d.items():
    print(position)
    print('done')
    print(df)
    print('end')
"""
```

```
Out[11]: "\nfor position, df in d.items():\n    print(position)\n    print('done')\n    print(df)\n    print('end')\n"
```

```
In [12]: for position, df in d.items():
    d[position] = clean[clean['Position'] == position]
```

```
In [13]: #IT WORKS!
d['DE']
```

Out[13]:

	Last Name	Grad Year	Position	Height	Weight	Zcruit Rating	Forty Yard Dash	Shuttle	Vertical Jump	Broad Jump	3 Cone	Wingspan	Arm Length	Hand Size	Powerball Toss	SPARQ Rating
0	Foreman	2021	DE	6' 4"	248.0	96.6970	4.880	4.700	30.4	104.0	7.758	NaN	NaN	NaN	44.5	
1	Tuimolou	2021	DE	6' 4"	275.0	96.6970	4.750	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
4	Turner	2021	DE	6' 2.5"	223.0	96.3939	4.991	4.487	26.0	115.0	7.752	83.5	34.25	9.84	NaN	
7	Sawyer	2021	DE	6' 4"	197.0	96.3939	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
11	Smith	2021	DE	6' 5.25"	334.0	96.0909	5.630	4.690	22.2	94.0	8.621	84.0	36.00	7.87	NaN	
...	
70639	Knight	2021	DE	5' 6"	136.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
70879	Hare Jr	2021	DE	5' 3"	131.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
70987	Moidel	2023	DE	5' 8"	128.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71001	Paz	2024	DE	5' 6"	128.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71232	Cea	2024	DE	5' 5"	118.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

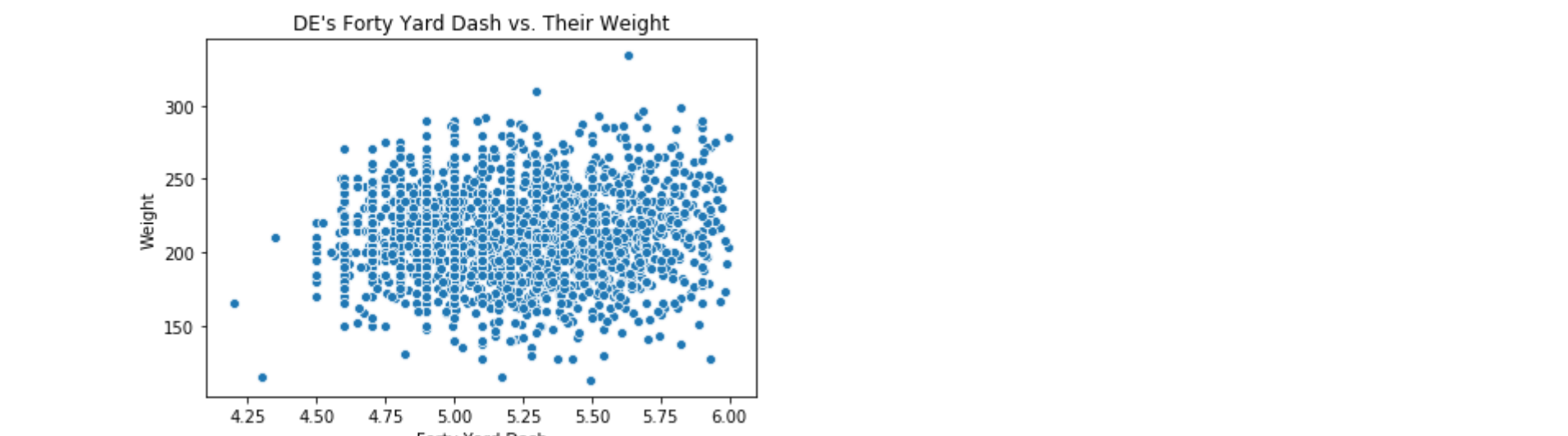
4229 rows x 16 columns

```
In [14]: #d['WR']['Height']
```

```
In [17]: #WR = d['WR'][d['WR']['Forty Yard Dash'] < 6.0]
DE = d['DE'][d['DE']['Forty Yard Dash'] < 6.0]
```

```
In [20]: sns.scatterplot(DE['Forty Yard Dash'], DE['Weight'])
plt.title("DE's Forty Yard Dash vs. Their Weight")
```

```
Out[20]: Text(0.5, 1.0, "DE's Forty Yard Dash vs. Their Weight")
```



```
In [22]: #ATH = d['ATH'][d['DE']['Forty Yard Dash'] < 6.0]
d['ATH']
```

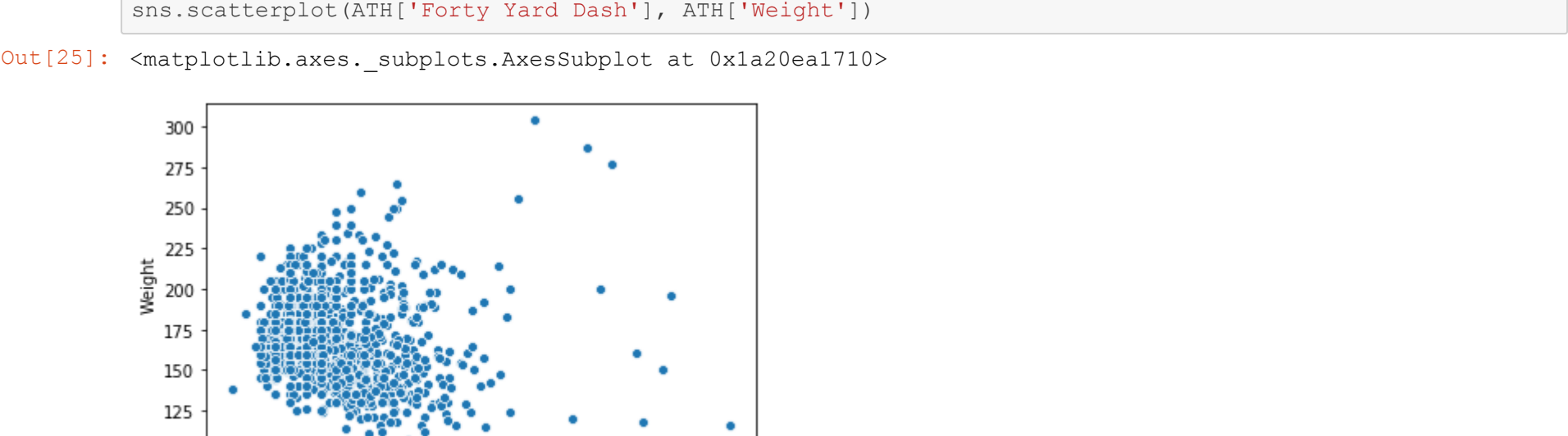
Out[22]:

	Last Name	Grad Year	Position	Height	Weight	Zcruit Rating	Forty Yard Dash	Shuttle	Vertical Jump	Broad Jump	3 Cone	Wingspan	Arm Length	Hand Size	Powerball Toss	SPARQ Rating
8	McKinstry	2021	ATH	6' 1"	175.0	96.3939	4.67	4.12	33.7	NaN	NaN	NaN	NaN	NaN	NaN	
42	Bowman	2021	ATH	5' 10"	185.0	95.1212	4.50	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
75	McCall	2022	ATH	6' 2"	180.0	94.2121	4.59	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
103	Starks	2022	ATH	6' 3"	205.0	93.5152	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
107	Bouie	2022	ATH	5' 11"	165.0	93.5152	4.62	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	
71276	Katzenmeyer	2021	ATH	5' 4"	115.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71277	Thomas	2021	ATH	5' 4"	115.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71301	Vidito	2021	ATH	5' 0"	112.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71336	Jones	2021	ATH	5' 1"	106.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
71352	Matarelli	2021	ATH	5' 1"	98.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

2247 rows x 16 columns

```
In [25]: #For data cleaning & visual aesthetics purpose
ATH = d['ATH'][d['ATH']['Forty Yard Dash'] < 10]
#We then visualize
sns.scatterplot(ATH['Forty Yard Dash'], ATH['Weight'])
```

```
Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x1a20ea1710>
```



```
In [ ]:
```