

COLLEGE FOOTBALL RECRUITING AND PORTAL ANALYSIS

JACOB YOUNT

DATA ANALYSIS

3/30/2023

DATA

Data Source

The data was sourced from [Kaggle](#). It is a public and open-sourced data set.

Data Contents

There are two different sets of data in this file. The first is the Transfer Portal Data, which represents college football athletes transferring to and from different colleges. The second set represents high school recruits and their rating and destination. Both are for colleges in the United States.

Limitations and Ethics

- Both data sets mention first and last name but are both public record. Don't think it is unethical to keep this information.
- Recruiting set contains height and weight, which will be eliminated as it's not needed.
- There are some formatting differences between the two sets, so combining them will take some formatting changes.
- Portal data is new and only dates to 2021 so there will be a limited sample size we can draw from.
- No data dictionary on file. Positions have initials. Will have to create a data dictionary and do additional research.

Data Profile

Portal Data Set:

Column	Description	Data Type	Time Variant
Season	Year or football seasons	Quantitative, discrete	yes
first_name	First name	Qualitative, nominal	No
Last_name	Last name	Qualitative, nominal	No
position	The player's position on field	Qualitative, nominal	No
origin	Starting school	Qualitative, nominal	No
destination	School transferring to	Qualitative, nominal	No
Transfer_date	Date of transfer with time	Quantitative, continuous	Yes
rating	Composite ranking of player performance	Quantitative, discrete	No
stars	The ranking of talent on a 1-5 scale	Quantitative, ordinal	No
eligibility	Determination if they can play	Qualitative, nominal	No
Transfer_date.1	Transfer date without time	Quantitative, continuous	Yes
Time	Time of Transfer	Quantitative, discrete	Yes

Recruiting Data Set:

Column	Description	Data Type	Time Variant
year	Year of football season	Quantitative, discrete	Yes
position	The player's position on field	Qualitative, nominal	No
school	Starting school	Qualitative, nominal	No

ranking	Ranking talent based on position group	Quantitative, discrete	No
rating	Composite ranking of player performance	Quantitative, discrete	No
stars	The ranking of talent on a 1-5 scale	Quantitative, ordinal	No
Committed_to	School they intend to go to	Qualitative, nominal	No
Recruit_type	Highschool, Junior college, or other	Qualitative, ordinal	No
height	Player height	Quantitative, continuous	No
weight	Player weight	Quantitative, continuous	No
city	City of school	Qualitative nominal	No
State_providence	State or providence of school	Qualitative, nominal	No
country	country	Qualitative, nominal	No
id	Unique id	Qualitative, discrete	No
Athlete_id	Unique id for each athlete	Qualitative, discrete	No
First_name	First name	Qualitative, nominal	No
Last_name	Last name	Qualitative, nominal	No
Last_name NULL	Last name with missing info	NA	No
Unnamed: 19	unknown	NA	No

Cleaning and Constancy Checks

- Checked for duplicates in both sets
- Checked for mixed data types in both steps
 - Changed the mixed types in each set
- Checked for missing values
 - Left blank as they could effect integrity of analysis

- Creating a yes/no column for committed and transferred columns to address NaN
- Merged two data sets on multiple keys (school, position, first name, last name, stars, rating)
- New Shape
 - (52181, 19)

Statistics of new data set

	season	rating	stars	ranking
count	52181.000000	48344.000000	51503.000000	46043.000000
mean	2017.656063	0.823153	2.763955	1638.368438
std	3.357293	0.056073	0.672115	1134.999116
min	2012.000000	0.606300	1.000000	1.000000
25%	2015.000000	0.781700	2.000000	599.000000
50%	2018.000000	0.820600	3.000000	1541.000000
75%	2021.000000	0.857100	3.000000	2536.000000
max	2023.000000	1.000000	5.000000	4259.000000

```
St. Thomas Aquinas    187
IMG Academy           185
Centennial            133
Iowa Western C.C.     127
St. John Bosco        125
...
Savannah State       1
Monticello Senior     1
Twin Falls Senior     1
Amherst               1
Redondo Union HS      1
Name: school, Length: 6787, dtype: int64
```

```
In [208]: df_merged['position'].value_counts()
```

```
Out[208]: WR      6873
          CB      4777
          OT      4261
          S      3950
          RB      3399
          ATH     3388
          OLB     2849
          DT      2708
          OG      2430
          TE      2351
          ILB     2019
          SDE     2013
          WDE     1943
          PRO     1690
          DUAL     1029
          LB      1028
          DL      1026
          QB       743
          IOL      650
          OC       632
          EDGE     595
          APB      569
          K        554
          P        335
          LS       189
          FB       160
          RET        1
          Name: position, dtype: int64
```

```
In [202]: df_merged['transfer_to'].value_counts().head(25)
```

```
Out[202]: Texas State    54
          UMass          48
          Arizona State  47
          SMU            47
          Western Kentucky 44
          USC            41
          Georgia Tech   40
          South Florida   40
          Memphis        40
          Oklahoma       39
          Indiana        39
          Jackson State   39
          Colorado       38
          Kansas         38
          Buffalo        36
          Michigan State  36
          Houston        35
          Florida State   35
          Louisiana Tech  35
          UCF            35
          Auburn         34
          Nevada         34
          UCLA           34
          Arkansas State  34
          Louisville     33
          Name: transfer_to, dtype: int64
```

```
In [200]: df_merged['committed_to'].value_counts().head(25)
```

```
Out[200]: Air Force      409
          Army          384
          Navy          336
          Alabama       319
          Texas A&M      317
          Georgia       315
          Washington State 313
          Minnesota     308
          Tennessee     307
          Virginia Tech  299
          Northern Illinois 297
          LSU           296
          Oklahoma      295
          Arizona       295
          Texas         295
          Arkansas      294
          Memphis       293
          Michigan      292
          Kentucky      290
          Florida       288
          South Carolina 288
          Indiana       287
          Mississippi State 285
          Iowa State     283
          Oregon        282
          Name: committed_to, dtype: int64
```

Questions to be answered.

What schools have higher transfer rates?

Is there a correlation between transfers and recruits?

Are certain positions more prone to transfer?

Are there schools that have higher rated recruits?

Is rank, rating, or star rating related to the transfer portal?

What schools are getting more recruits? Transfers?