College Football REcruiting and portal Analysis

Jacob Yount  
Data Analysis

3/30/2023

# Data

**Data Source**

The data was sourced from [Kaggle](https://www.kaggle.com/datasets/dubradave/college-football-portal-and-recruiting-statistics). 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 seaosn | Quantitative, discrete | yes |
| first\_name | First name | Qualitative, nominal | No |
| Last\_name | Last name | Qualitative, nominal | No |
| position | The players 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 players 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 |
| Commited\_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 athelte | 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)

**Table

Description automatically generated with medium confidenceTable

Description automatically generatedStatistics of new data set**

Table

Description automatically generated

Table

Description automatically generated

Table

Description automatically generated

**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?