

# An Analysis on NFL Draft Prospects and Career Performance in Relation to Collegiate School Size

Jensen Judkins

*Computer Science, Utah State University*

**Abstract**— This paper gives an analysis on the current state of the NFL draft and its subsequent athletes who are picked from said draft. With a large majority of resources, picks, and news coverage relaying information only from the largest and top athletic programs in the country, we wish to uncover if there is a tangible truth to picking athletes from larger collegiate programs than from those of their smaller peers. Using data analysis of the last decade of draft picks selected, their collegiate career performance, and their current NFL career stat lines we will see that there is little to no difference in picking from a ‘medium’ sized college and a ‘small’ college.

**Pick** - In reference to the selection made by an NFL team indicating that their professional career will start with that team.

**Receiver** - A position in football in which they catch the ball and progress towards their goal of obtaining a touchdown or gaining more yards.

**Draft** - Annual Player Selection Meeting, is an annual event which serves as the most common source of player recruitment in the National Football League.

**Round** - Each round contains 32 picks. Each one of those 32 picks goes to a single NFL team, so every team gets a chance to pick in every round.

**Bust** - A term used to describe a player who was expected to perform very well in the NFL, however when playing professionally has failed to meet those expectations.

## I. INTRODUCTION

Since 1936 the National Football League has held a Draft in which teams are rewarded picks based on their prior performance and can be exchanged for players from the current year's selection of collegiate athletes. The NFL draft is a way of balancing teams and keeping the sport competitive. If a team performs poorly, then they are awarded picks ahead of other teams in hopes that their team will perform better the coming season. Recent standouts of this team balancing in action include the Houston Texans who in 2022 won three games, lost thirteen games, and tied once selected C.J. Stroud with the second overall pick. With the addition of C.J. and other coaching changes, the Texans won ten games and only lost seven the following year. Drafting the right player at the right time is crucial in turning an

organization's team around and can have huge implications overall.

As expected with any tradable object, picks have rough assigned values depending on the position of that pick. The first overall pick in the draft has a value of three thousand as compared to the thirty second pick having a value of 590. These values continue to decrease as the draft continues. In addition, each team is allotted seven picks each year. Thus, teams want to get the most production out of every pick that they can and have as few busts as possible. One theory of a way to mitigate busts is to prioritize athletes who have played football at larger schools than smaller schools.

Thus, my motivation for the project, I argue that there are many players being overlooked in the NFL Draft purely because they compete at a smaller collegiate school, and that their potential and current production is being undervalued. Using collegiate receiver data, along with receiver data from the NFL website, and a record of every draft pick in the last decade I found that on average players from small and medium sized schools have negligible different stat lines while still being sizably beaten by their larger school competitors.

## II. METHOD

First, we must set a baseline for what constitutes a large, medium, and small school. These definitions for the purpose of this paper can be found below. This is a subjective ordering and thus may yield different results for those replicating this experiment with different groupings.

### A. Collegiate Groupings

**Small** = 'Air Force', 'Akron', 'Appalachian State', 'Arkansas State', 'Army', 'Ball State', 'Boise State', 'Bowling Green', 'Buffalo', 'Central Michigan', 'Charlotte', 'Cincinnati', 'Coastal

Carolina', 'Colorado State', 'Connecticut', 'Duke', 'Eastern Michigan', 'East Carolina', 'Florida Atlantic', 'Florida International', 'Fresno State', 'Georgia Southern', 'Georgia State', 'Hawaii', 'Houston', 'Illinois', 'Indiana', 'Kent State', 'Liberty', 'Louisiana', 'Louisiana Tech', 'Louisiana-Monroe', 'Marshall', 'Memphis', 'Miami (OH)', 'Middle Tennessee', 'Navy', 'Nevada', 'New Mexico', 'New Mexico State', 'North Texas', 'Northern Illinois', 'Ohio', 'Old Dominion', 'Rice', 'San Diego State', 'San Jose State', 'SMU', 'South Alabama', 'South Florida', 'Southern Miss', 'Temple', 'Texas State', 'Toledo', 'Troy', 'Tulane', 'Tulsa', 'UAB', 'UCF', 'UNLV', 'UTEP', 'UTSA', 'UTah State', 'UTah State', 'Western Kentucky', 'Western Michigan', 'Wyoming'

**Medium** = 'Arizona', 'Arizona State', 'Arkansas', 'Boston College', 'BYU', 'California', 'Clemson', 'Colorado', 'Florida State', 'Georgia Tech', 'Iowa', 'Iowa State', 'Kansas', 'Kansas State', 'Kentucky', 'Louisville', 'Maryland', 'Miami', 'Michigan State', 'Minnesota', 'Mississippi State', 'Missouri', 'Nebraska', 'North Carolina', 'North Carolina State', 'Northwestern', 'Oklahoma State', 'Oregon State', 'Pittsburgh', 'Purdue', 'Rutgers', 'South Carolina', 'Stanford', 'Syracuse', 'TCU', 'Tennessee', 'Texas A&M', 'Texas Tech', 'UCLA', 'Utah', 'Virginia', 'Virginia Tech', 'Washington', 'Washington State', 'West Virginia'

**Large** = 'Alabama', 'Auburn', 'Florida', 'Georgia', 'LSU', 'Michigan', 'Notre Dame', 'Ohio State', 'Oklahoma', 'Oregon', 'Penn State', 'Texas', 'USC', 'Wisconsin'

### B. Data Collection

Two data collection methods were used to collect all that was needed. For collecting college football statistics, I used collegefootballdata.com [1] and created a simple API python program which would save the Json results to a csv file. I requested all players from 2011 to 2022 with receiving yardage stats. Following that I used the same API to collect all draft picks from 2011 to 2022 from the respective colleges listed above and save them by year. During this method of collection, the player is also recorded which college they played for.

The second data collection method was using a scraper written by 'MarcLinderGit' [2] who wrote it in python. His script scrapes pages from the official NFL website and saves them to csv files. Using the web scraper I downloaded all NFL receiver statistics from 1970-present and then discarded the unused years.

### C. Data Preprocessing

Processing the data into a format that is easily analysed was by enlarge the biggest portion of the

project. The following steps were what was used to come to the conclusions as found later in this paper.

1. Combine all college receiving data into one dataframe and assign years to each row of data indicating the year the player recorded those specific stats.
2. Sum each player's career in college and create a new column indicating the number of years that player played as a college athlete.
3. Combine player draft information to their college career stats.
4. Create a new column indicating the size of the school dependant on which grouping that college correlates to as seen above.
5. By this point each player should have 27 columns indicating the following
  - 5.1. name - Player's name.
  - 5.2. playerId - Collegiate player ID.
  - 5.3. nflAtheleteId - NFL player ID.
  - 5.4. college - College they played for.
  - 5.5. nflTeam - Team that drafted them.
  - 5.6. school\_size - College size grouping.
  - 5.7. conference - Collegiate conference.
  - 5.8. LONG, REC, TD, YDS, YPR - College performance stat lines.
  - 5.9. draftYear - The year player was drafted.
  - 5.10. position - Position player played.
  - 5.11. height - Height of player.
  - 5.12. weight - Weight of player.
  - 5.13. round - Round that player was drafted.
  - 5.14. pick - Pick number they were drafted within round.
  - 5.15. overall - Overall pick number.
  - 5.16. preDraftRanking - Ranking before the NFL draft.
  - 5.17. preDraftPositionRanking - Ranking based on position.
  - 5.18. preDraftGrade - grading NFL scouts gave as a score of potential.
  - 5.19. hometownInfo - Information about the player's hometown.
6. Moving to making an NFL career dataframe, we combine all NFL stats based on playerId and then create a new column indicating the

number of years the player has played.  
Columns include:

- 6.1. nfl\_REC - Number of NFL receptions.
- 6.2. nfl\_YDS - Career NFL yards.
- 6.3. nfl\_TD - Touchdowns in NFL.
- 6.4. nfl\_LNG - YDS from 20+ yard receptions.
- 6.5. nfl\_Rec\_FUM - Career count of receptions leading to fumbles.
- 6.6. nfl\_Rec\_YAC/R - Average yards after catch per reception in NFL.
- 6.7. nfl\_TGTS - NFL targets.
- 6.8. years\_in\_nfl - Career length in years.
7. Following that we merge the NFL career dataframe to the college career dataframe and have 31 columns to every one player.

If a cell contained no information, that player would be subsequently dropped from the dataframe. We end up with 569 players with thirty-one columns each. Lastly, to grade a player's overall value within their NFL career I created a 'player\_value' column which is calculated by the following function. Descriptions of the weighting assignments are also as follows:

W1 = .3 [Yards overall are most important]  
W2 = 0.25 [Yards after receiving the ball are important as it signifies breaking tackles]  
W3 = 0.2 [Touchdowns will signify red zone competence and threat levels]  
W4 = 0.15 [Years in NFL shows resilience and toughness]  
W5 = 0.1 [Receptions confirming the player gets open and catches the ball]  
W6 = -0.1 [Negative for reception followed by a fumble]

Normalized by dividing by 6.

$Player\_value = [nfl\_YDS]*W1 + [nfl\_Rec\_YAC/R]*W2 + [nfl\_TD]*W3 + [years\_in\_nfl]*W4 + [nfl\_REC]*W5 + [nfl\_Rec\_FUM]*W6/6$

#### D. Initial Data Analysis

Initial analysis of the data indicates that large and medium schools obtain a large number of draft picks over their small school counterparts as seen in Figure 1.

Percentage of Drafted Wide Receiver Players by School Size

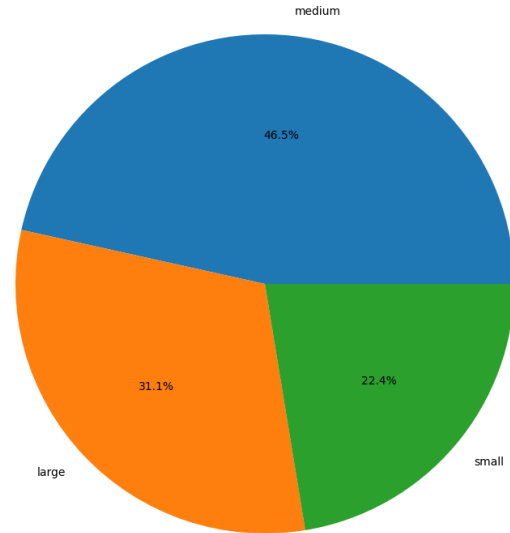


Fig. 1 A pie chart showing that large and medium sized schools make up for over 77% of the draft picks for wide receivers from years 2011-2022.

Diving deeper into this analysis we see that the most popular round to take an athlete from a small school is the fourth round, in which less than 30 have been picked in the last 11 years. Compared to the competing medium size where each round has over thirty players taken in the last 11 years with the fourth round again being the favorite at just under 60 players. One would think that as the players from large and medium sized schools get chosen in the earlier rounds, which would leave room for standout players from smaller schools to be picked as the rounds progressed. But in reality, we see the opposite, with the number of picks from small schools declining after round 4 (Fig. 2).

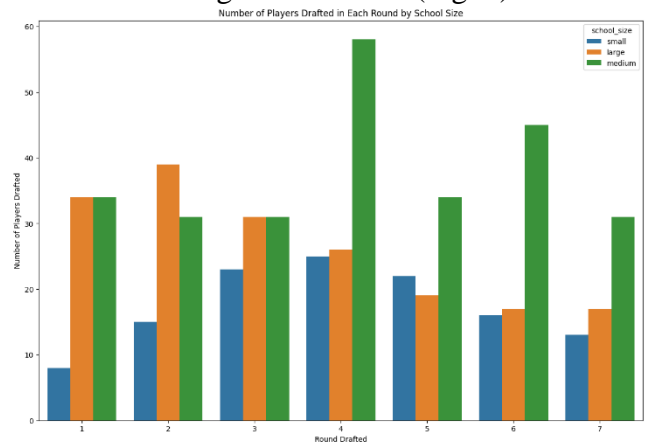


Fig.2 Number of draft picks from each corresponding round years 2011-2022

Finally, we see that the large majority of small schools have less than 5 draft picks in total as compared to the likes of Alabama who has had over 30 receivers drafted in the last 11 years. The leader in small schools is Utah State with 13 receivers drafted in the last 11 years.

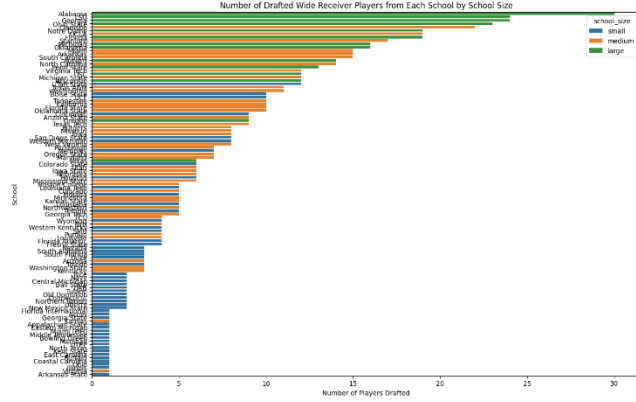


Fig. 3 Graph indicating number of drafted wide receivers in last 11 years by school.

As we see there is a large disparity in draft picks being favoured for large and medium sized schools over those found in our small grouping despite the small grouping being significantly larger in size than both the medium and large school grouping combined. We can confirm that there is a bias towards larger collegiate school size when it comes to drafting, but is there a valid bias in this case?

To answer the prior question, I used the value grading system discussed in the data preprocessing portion of this paper. However, in order to collect an accurate reading, we need to normalize the number of datapoints such that each school size grouping has the same number of players. I did this by using the smallest grouping of drafted players as my count and took samplings of the other two groupings according to the smallest size. In this dataset the smallest grouping was small schools with a count of 129 players, thus I took sample sizes of 129 players from both medium and large groups to properly compare them. In my tests I used an input of random state set to 15. The following is a graphic showing the average player value grouped by college size (Fig. 4).

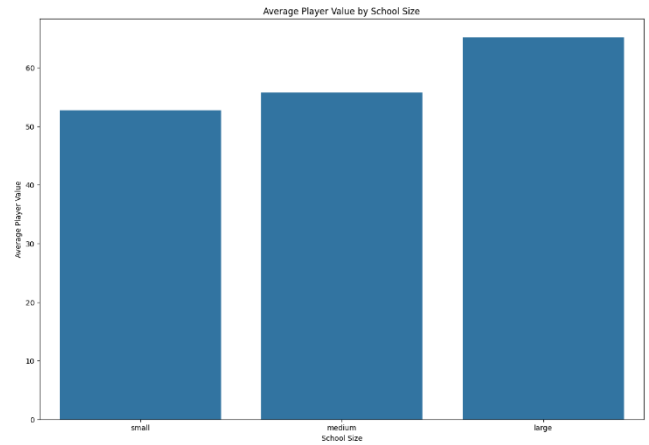


Fig. 4 Graphic showing average player value to college size.

We see that medium schools have a marginal improvement of only 5.735% in overall value over small schools. Athletes coming from large schools however produce 23.577% more value than that of small schools. Thus, we can conclude that medium and small school college draft ranking should be interchangeable based on actual NFL performance. We should thus see more schools trying to draft outstanding players from small schools then we currently see.

### III. EXPERIMENTAL RESULTS

As an additional test to conclude that no one grouping is better than another outstandingly, a training test model should be used. Using sklearn's `train_test_split` and a logistic regression model, we aimed to see if there was a specific column found within an NFL players stat line that would allow the model to successfully predict which school size the player was from.

#### A. Featured Columns of Logistic Regression Model

Feature selection was simple as we wanted to know if there were defining characteristics or traits found in one type of player to another based on school size. Thus, we used all columns available to us in each player's professional career. These columns are the same as those listed in C. Data Preprocessing 6.1-6.8.

#### B. Hypothesis

If the model can successfully predict which school size the player is from based on their stat line, then something must be missing that we are not acknowledging. However, if the model has a low

accuracy, then we can assume that differentiating player performance with respect to which college size they come from is much more difficult, and thus can conclude that players can often produce similar outputs regardless of collegiate school size. In laments terms, more standout players from smaller schools should be drafted, because they have the potential to blend in similarly in value to those drafted from larger schools.

### C. Results

Using a test size of 20%, and a training size of 80%, along with a random state of 1, we see out results come back with an accuracy score of 33.78%. Comparable results are present without the use of a random state such that the logistic regression model is unable to successfully predict with accuracy the college size in which a player originated from.

## IV. CONCLUSION

In conclusion we learned that there is a strong bias correlated between a school size and the number of draft picks based on that school size. We see that even though players from small schools on average produce similar value ranges with athletes from medium schools, medium schools find upwards of twice as many picks as small schools. We found that using a logistic regression model and statistics from players real NFL stat lines, a model has poor accuracy in predicting which school size the player had originated from. Thus, a conclusion in that players from smaller schools are continually being overlooked year over year, NFL scouts should commit more time, effort, and funding in scouting smaller schools for potential additions to NFL franchises.

## V. FUTURE WORK

Potential future work is plentiful with this project. Since my project only covered wide receivers there is more research to be done on the productivity, value, draft bias and more of every other position in American football. A position of particular interest should of course be quarterback as it often has the highest boom or bust valuation when regarding a teams performance. As we saw with C.J. Stroud and the Texans, does drafting a player from a big-name

school have a profound effect? Or will a player like Jordan Love from Utah State be regarded as better overall than C.J.? Another potential addition to this project is a larger dataset. Using the last 11 years of receiver data I thought would be adequate, however after all the data preprocessing and the dropping of rows with missing values, I only had 569 players with recorded stats to work with. Expanding the dataset could yield other interesting results such as a notable shift or lack thereof of drafting from large schools.

## VI. REFERENCES

- [1] College Football Data API - <https://collegefootballdata.com/>
- [2] NFL Data Web Scraper - [https://github.com/MarcLinderGit/NFL\\_Stats](https://github.com/MarcLinderGit/NFL_Stats)