**Analyzing the NFL Draft**

CISC 6325 Final Project

by Henry Gorelick

Table of Contents

[**Introduction** 2](#_Toc532751945)

[The Problem 2](#_Toc532751946)

[Part 1: Sustained Success 2](#_Toc532751947)

[The Outlier 2](#_Toc532751948)

[Part 2: The Draft 3](#_Toc532751949)

[Background 3](#_Toc532751950)

[The Rules 3](#_Toc532751951)

[The Issues 3](#_Toc532751952)

[**Claim** 4](#_Toc532751953)

[**Data Collection and Processing** 4](#_Toc532751954)

[Data Collection 4](#_Toc532751955)

[Python Packages 4](#_Toc532751956)

[Sources 4](#_Toc532751957)

[Preprocessing 4](#_Toc532751958)

[Processing 4](#_Toc532751959)

[Python Packages 4](#_Toc532751960)

[The Database – *PFR.db* 5](#_Toc532751961)

[Sample Results 5](#_Toc532751962)

[**Results** 6](#_Toc532751963)

[Single Player Results 6](#_Toc532751964)

[Cam Newton 6](#_Toc532751965)

[Derek Anderson 7](#_Toc532751966)

[All Players Results 8](#_Toc532751967)

[Real Pick vs. Average Difference in Predicted Pick 8](#_Toc532751968)

[Number of Correct Predictions by Position 9](#_Toc532751969)

[**Future Work** 10](#_Toc532751970)

[Include NFL Scouting Combine Data 10](#_Toc532751971)

[Use Data to Predict NFL Production 11](#_Toc532751972)

[Include Players Drafted After 2015 11](#_Toc532751973)

[**Appendix** 12](#_Toc532751974)

[Pick Value Formula 12](#_Toc532751975)

[Award Score Formula 12](#_Toc532751976)

# Introduction

## The Problem

### Part 1: Sustained Success

Consider the following facts:

* The Carolina Panthers have never had consecutive winning seasons in 22 tries
  + And will likely extend that streak to 23 seasons: after an 11-5 campaign in 2017, the team is 6-7 in 2018 with three games to play.
* The Atlanta Falcons went 42 seasons before putting together back-to-back seasons winning more than half their games.
* Since the NFL expanded its playoffs from ten to twelve teams in 1991, the average number of teams to return to the postseason after qualifying in the previous season is 6.4[[1]](#footnote-1).

It’s hard to win consistently in the National Football League.

##### The Outlier

Within this trend of mediocrity lies the New England Patriots. The Patriots are enjoying a 19th consecutive season winning 8 or more games. With one more victory the team will extend its streak of 15 consecutive seasons winning 10 or more games. With two and three more victories the Patriots can extend its streak of 8 consecutive 11- and 12-win seasons, respectively.

While experts debate the reasons for New England’s dynastic success, everyone agrees on one thing: Tom Brady. The legendary quarterback’s list of accolades is, well, legendary:

* 13x Pro Bowl (All-Star selections)
* 3x 1st Team All-Pro (more selective and prestigious than the Pro Bowl)
* 3x MVP
* 2x Offensive Player of the Year
* 5x Super Bowl Champion

Alone, Brady’s accomplishments make him arguably the greatest football player of all time, but he also holds the title as the greatest steal of all time. Brady was selected in the 6th round of the 2000 NFL Draft—he was the 199th overall pick.

### Part 2: The Draft

##### Background

At the end of April, nearly three months removed from the Super Bowl (the culmination of the NFL season), the NFL holds its annual NFL Draft. The Draft keeps the NFL going: it brings in new players and sparks hope in each of the 32 fan bases that maybe this year will be their year.

##### The Rules

* *The Picks*
  + 7 rounds of picks
  + By default[[2]](#footnote-2), each team receives 1 pick per round (32 picks per round[[3]](#footnote-3))
  + Total of ~254 picks
* *The Order*
  + Reverse standings of the previous season
    - The team with the worst record picks 1st in each round, the team that won the Super Bowl picks 32nd in each round.
* *The Players*
  + Teams may pick eligible players to join their team in the upcoming season
  + Eligibility
    - “To be eligible for the draft, players must have been out of high school for at least three years and must have used up their college eligibility before the start of the next college football season. Underclassmen and players who have graduated before using all their college eligibility may request the league’s approval to enter the draft early.”[[4]](#footnote-4)

##### The Issues

Each team has many, nuanced goals when beginning the Draft. Teams have scouting departments dedicated to evaluating each draft eligible player, and it’s up to the General Manager and Head Coach to leverage this scouting data when making their selections. Although Tom Brady is the most extreme outlier, he’s not the only “all-time great” to be selected late in the Draft. Furthermore, there are outliers on the other end of the spectrum: players selected early in the Draft who underachieve relative to their draft position.

###### *Bust*

A player who underperforms relative to his draft position

###### *Steal*

A player who overachieves relative to his draft position

While there are players who fall between *Busts* and *Steals*, common sense holds that the teams that finds steals and avoids busts are more likely to be successful.

# Claim

You might think that the first players selected in the Draft were also the best players during their time in college. That sort of thinking led the Denver Broncos and Cleveland Browns selecting Tim Tebow and Johnny Manziel, respectively, in the first round. Each quarterback won the Heisman Trophy during his college career, awarded to the best player at any position in college football. Neither were successful in the NFL with Tebow’s career lasting only two years, and Manziel’s one year.

**Therefore, I will demonstrate that there is no correlation between collegiate success and draft position.**

# Data Collection and Processing

## Data Collection

### Python Packages

* SQLite3
* BeautifulSoup
* Selenium Chromedriver

### Python Modules

* pfr\_scraper.py
  + Contains the PFRScraper web-scraper class with member functions that built PFR.db
* write\_pfr\_db.py
  + Contains all methods for building tables in PFR.db

### Sources

I used the following sources to build PFR.db

* [www.pro-football-reference.com](http://www.pro-football-reference.com)
* [www.sports-reference.com/cfb](http://www.sports-reference.com/cfb)

### Preprocessing

There was not much preprocessing required. For the most part, the websites provided clean, organized data in easily scrapable tables. The preprocessing required included:

* Removing players with little to no statistics
* Recalculating percentage statistics for more accurate significant digits

## Processing

### Python Packages

* Pandas
* Sklearn

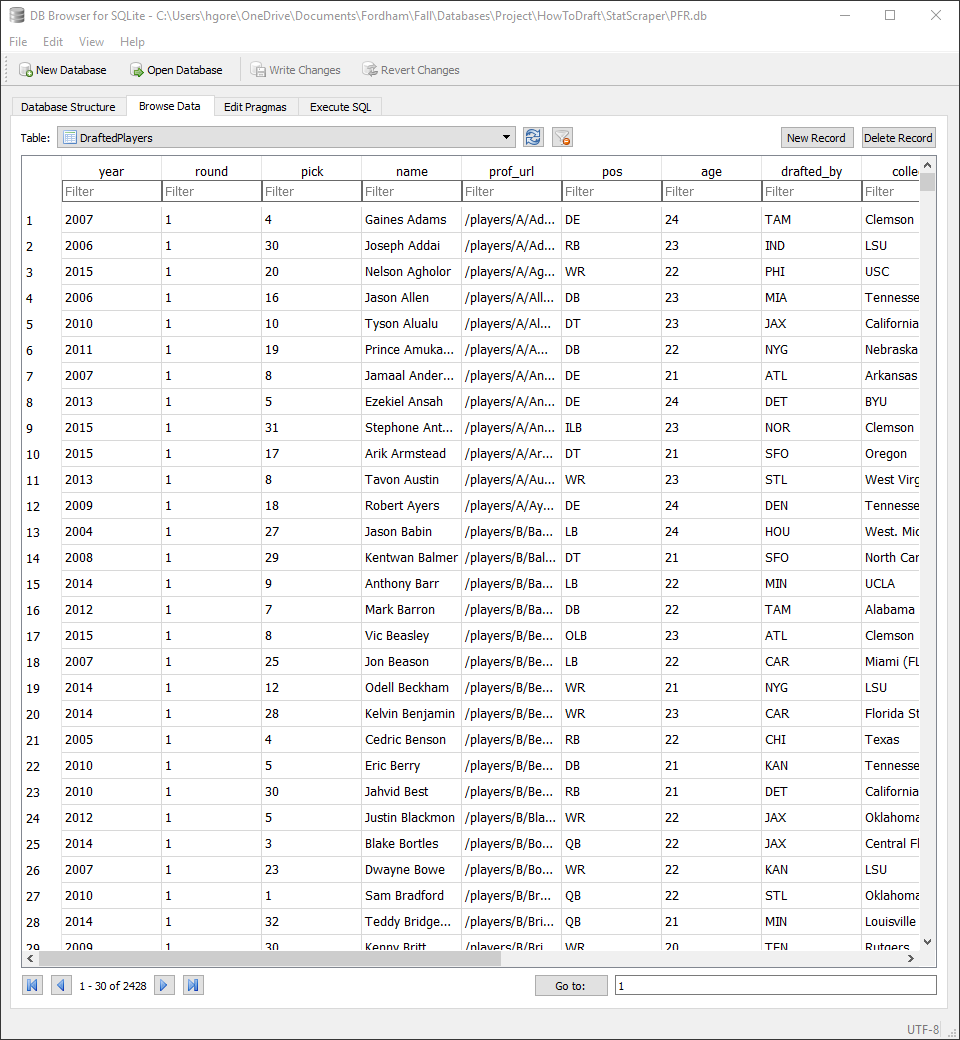
### Python Modules

* Processor.py
  + Contains the key functions for training the DecisionTreeClassifier, which predicts a player’s draft position given his collegiate stats.
  + Contains all graphical plotting functions

### The Database – *PFR.db*

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Name** | **Rows** | **Description** | **Used / Unused** |
| DraftedPlayers | 2,428 | All players drafted between 2004 – 2015 | Used |
| PickPredictionAccuracies | 581 | Contains every player drafted in the first two rounds between 2004 – 2015 and his predicted draft position | Used |
| CollegeStats | 2,021 | The collegiate career statistics of each player in DraftedPlayers | Used |
| CombinePlayers | 2,174 | NFL Scouting Combine data for every player who participated between 2004 – 2015 | Unused |
| NotOnDrafted | 807 | Players from DraftedPlayers who were not still on the team that drafted them after four seasons. | Unused |
| PlayersPickValues | 2,305 | Each player in DraftedPlayers with an [award score and pick value](#PickValue) | Unused |

### Sample Results



# Results

Unfortunately, I was not able to use all the data I collected. But I still arrived at interesting results.

## Single Player Results

Single player results include a bar graph depicting the weights of the Decision Tree’s five most important features in predicting a player’s draft position.

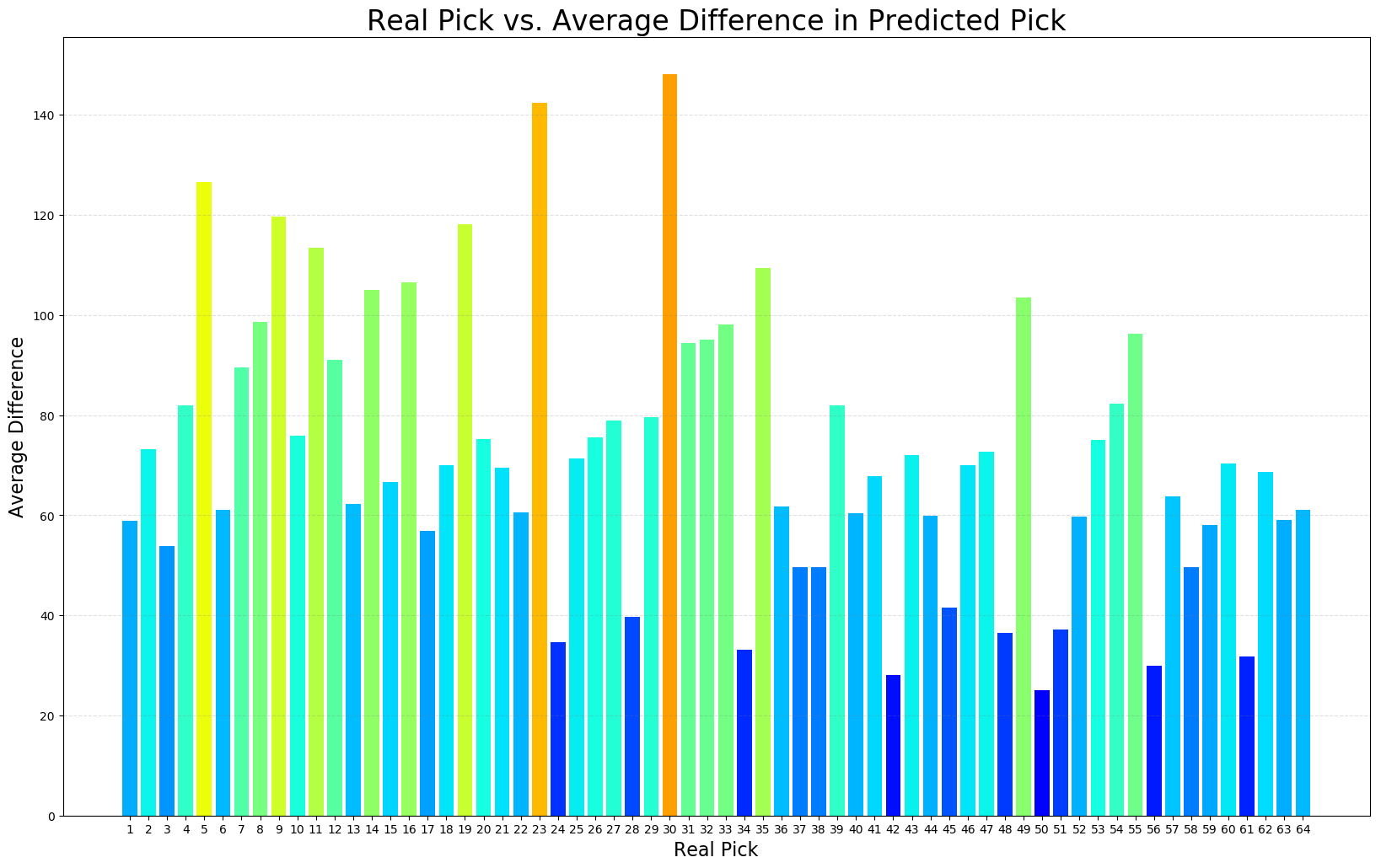
### Cam Newton

### Derek Anderson

## All Players Results

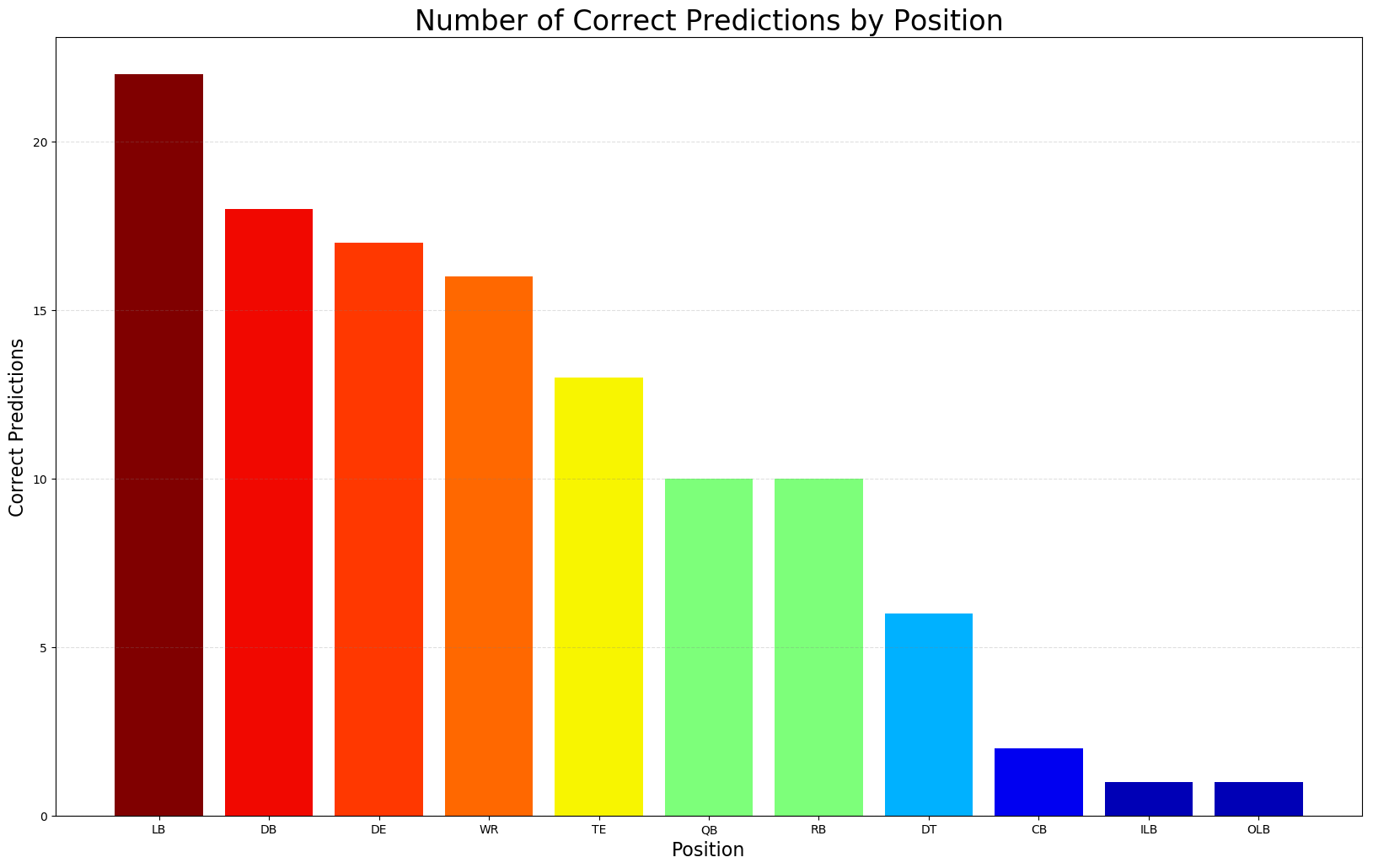
To further demonstrate the lack of correlation between collegiate success and draft position, I generated the following two bar charts:

### Real Pick vs. Average Difference in Predicted Pick



### Number of Correct Predictions by Position

Since correct prediction percentage is low, it’s more valuable to look at total correct prediction.



# Future Work

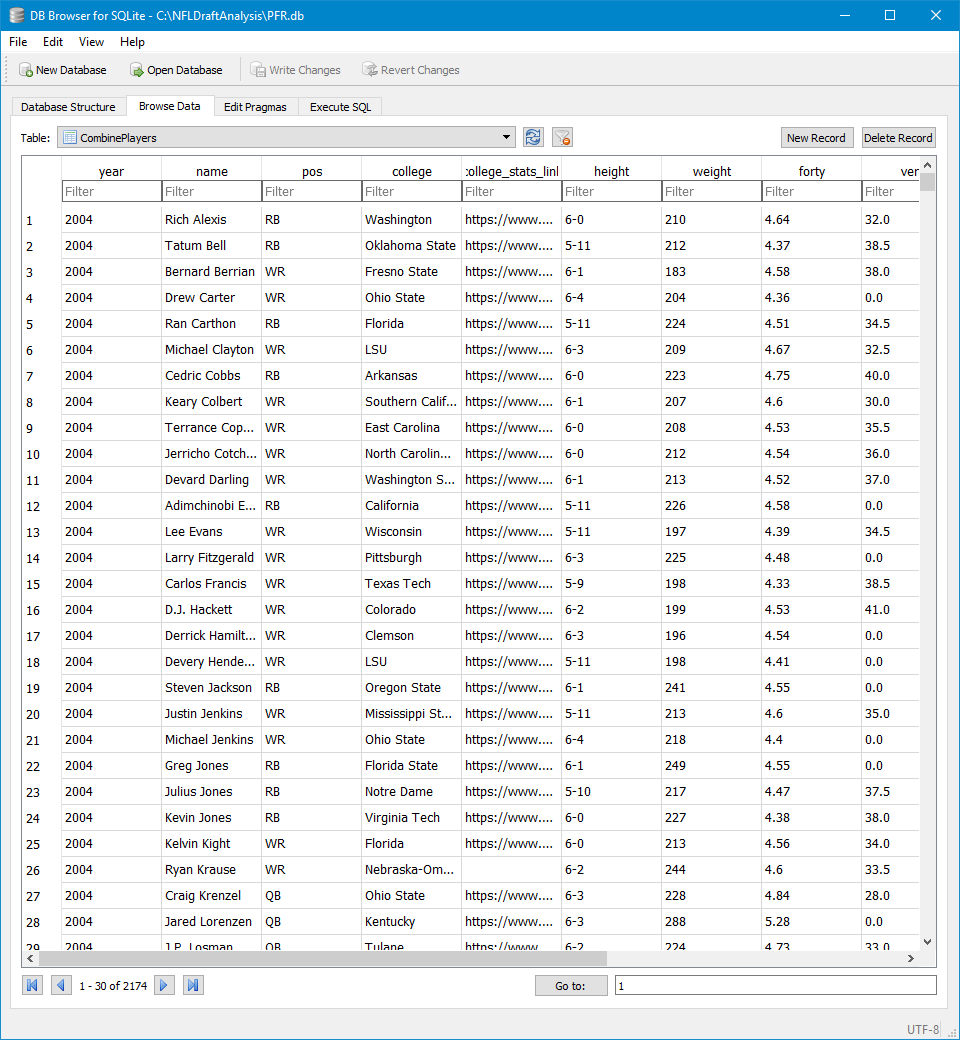
## Include NFL Scouting Combine Data

Scouts rely heavily on the results of the NFL Scouting Combine. At the Combine, players are broken into position groups, and then participate in six different tests to measure athletic ability:

* 40-Yard Dash
* Vertical Jump
* Bench Press
* Broad Jump
* 3-Cone Drill
* 20-Yard Shuttle

A player who demonstrates elite athletic ability at the Combine can drastically improve his draft stock. Players can go from a projected 3rd – 5th round pick to a 1st round pick. Including these results, which I collected and wanted to include, would likely make my model accurate.

### Combine Sample Results



## Use Data to Predict NFL Production

It would be extremely valuable to find a relationship between a player’s collegiate statistics and his professional statistics. This is ultimately what the draft is about: finding players who will produce at an elite level in the NFL.

## Include Players Drafted After 2015

I chose to not include players drafted after 2015 because they have not reached the end of their first contracts.

# Appendix

## Pick Value Formula

## Award Score Formula[[5]](#footnote-5)

# Sources

## Data

[www.pro-football-reference.com](http://www.pro-football-reference.com)

[www.sports-reference.com/cfb](http://www.sports-reference.com/cfb)

## Machine Learning Help

<https://medium.freecodecamp.org/using-machine-learning-to-predict-the-quality-of-wines-9e2e13d7480d>

1. https://thecomeback.com/nfl/nfl-2018-playoffs-retain-four-teams-last-year-tied-since-1991.html [↑](#footnote-ref-1)
2. Teams may trade their picks for essentially anything including, but not limited to: other picks, players, coaches, money, etc. [↑](#footnote-ref-2)
3. In March before the draft, teams are awarded compensatory picks for players lost in Free Agency. These picks range from the 3rd to 7th rounds. [↑](#footnote-ref-3)
4. <https://operations.nfl.com/the-players/the-nfl-draft/the-rules-of-the-draft/> [↑](#footnote-ref-4)
5. <http://www.nfl.com/news/story/0ap3000000810192/article/best-nfl-draft-picks-of-the-century-top-selections-by-school> [↑](#footnote-ref-5)