**2020 NFL Draft Fan Sentiment Insights**

Group Project Report

MGT 6203

Fall 2023

https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-79

**TEAM 79:**

Hugh Hoagland

Raju Ahmed

Pablo Francisco Ramos Soszna

Alexander Martin Stetzer

Nicholas Weist

Table of Contents

**Background3**

**Objective3**

**Methodology3**

Comment Analysis3

Performance Analysis6

**Approach9**

**Conclusion9**

**Background**

In the ever-evolving landscape of professional sports in the United States, the National Football League (NFL) towers above the rest, both in terms of revenue generation and the unwavering dedication of its passionate fanbase. As of 2022, the NFL's annual revenue surpassed a staggering $11.9 billion (Forbes, 2023), solidifying its status as an economic powerhouse. The influence of the NFL extends far beyond the realm of sports, with entire industries, such as fantasy football, flourishing in its wake. At the heart of this lies the NFL draft, a pivotal event that serves to infuse new talent into the league and offers the opportunity for the keen manager to infuse their team with fresh talent that may be overlooked by others.

The NFL draft is a process characterized by notable outliers, showcasing remarkable success stories and stark failures. Examples like Tom Brady, who was selected as the 199th overall pick and went on to become the player with the most Super Bowl championship wins, and JaMarcus Russell, who was chosen as the first overall pick but quickly descended into the ranks of notable busts, highlight the complexity and uncertainty that surrounds this critical juncture in the NFL's operation. The NFL draft, as a pipeline for emerging football talent, represents a microcosm of the challenges and opportunities intrinsic to the league's success.

The Covid-19 pandemic introduced a unique twist to the 2020 NFL draft process by forcing the event to be fully remote. In lieu of in-person attendance, fans were encouraged to post to NFL-related subreddits and express their opinions on their respective teams' draft picks. These comments and reactions, spanning the entirety of the draft, were assigned sentiment scores, ranging from -1 (extremely negative) to 1 (extremely positive). This shift towards social media allowed the collection of data produced a natural experiment and provided valuable data to study the collective, crowd-sourced evaluation of draft picks similar to statistician Francis Galton’s observation of the "wisdom of crowds" (Vox Populi, 1907).

Through a comprehensive analysis of fan sentiment in Reddit comments during the NFL draft, this research aims to determine whether such sentiment can provide measurable benefits to NFL team managers in improving draft selections and subsequently impacting player success in the league.

**Objective**

Our research comprises three key components: accurate ingestion of Reddit comments to ensure data reliability, the development of performance metrics for objective player assessment, comparative analysis of fan sentiment, draft positions, and player performance over three NFL seasons with the ultimate goal of empowering decision-makers with actionable insights. By exploring the relationship between fan sentiment and player success, we aim to test the hypothesis that fan sentiment, as extracted from Reddit comments, can quantifiably benefit managers in improving draft selections.

**Overview of DataData Sources**

To test the hypothesis, our team relied on two categories of data sources. The first pertains to determining sentiment associated with draft decisions. To achieve this, we combined multiple datasets from Kaggle that provided the components of our draft analysis, timestamped draft picks and similarly timestamped Reddit comments and their sentiment scores. The second category focuses on player performance metrics. In our quest to assess player success, we examined various sources, but ultimately made the decision to construct a proprietary scoring system, with its basis being derived from Pro Football Focus' (PFF) popular "player grade" system.

**Exploratory Data Analysis**

1. **Comment analysis**

To achieve our project's goal of comprehending fan sentiment during the 2020 NFL draft and its potential impact on NFL teams, we cleaned and aligned Kaggle datasets. The first step taken was to combine the round data frames into a single data frame.

Based on our initial analysis of positive and negative sentiment for time stamps containing videos and images, we found that those rows show no clear trend, represent a small fraction of the data, and are not accurately captured by the sentiment scoring system (Vader). Therefore, they were omitted from the ultimate dataset.

To associate comments with draft pick, using a look up table based on the original picks data frame, a series of operations calculate the previous timestamp in the picks data frame and associate that pick’s timestamp, round, pick number, player name and team name with the comment’s timestamp. This dataset now contains a valid timestamp, round number, pick number, player name and selecting team value for each individual comment with which to predict performance.

Figure 3 shows the summary of associated picks and their respective number or rows indicating that 75% of the picks will have at least 440 rows of data. Since the number of rows associated with each pick is quite high, an analysis of player name frequency was necessary.

1. **Player Name Analysis**

The project aims to determine if fan sentiment towards drafted players correlates with their on-field performance. While Reddit post sentiment scores reflect comment sentiment, many comments are irrelevant to players or picks. To address this, we examined the number of comments mentioning a player's name. From the summary, we found that median player has about 9 comments with their name, while the top ten players have several hundred (Figure 4). This discrepancy suggests a potential variable for identifying draft superstars

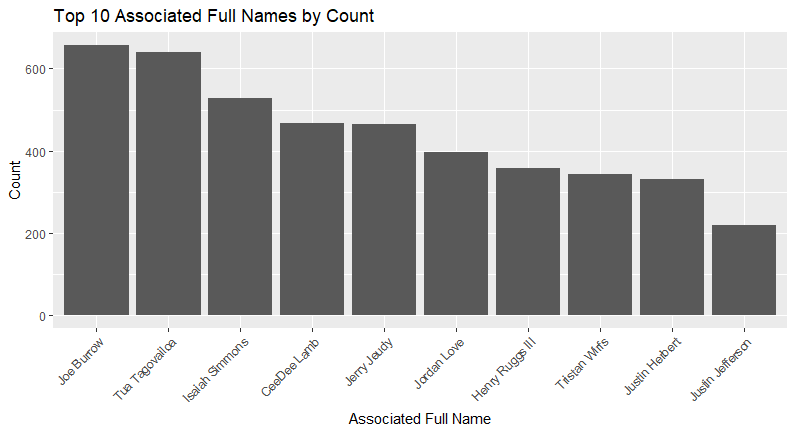


Figure 4: Top 10 payer full names by count

1. **Offensive Words and Fan Sentiment**

To explore correlation between the reddit comment and sentiment score, we have analyzed the contents of the text against a list of top 50 offensive/slang language words against the text. Finding such a library was challenging due to ethical concerns surrounding offensive words.. The goal of this analysis was to check whether these words can predict a bad sentiment score. These top slang/offensive words used for the library were collected from various dictionaries, social media platforms, news articles, and rap and hip-hop songs, the top 20 of which are shown with associated sentiments in table 1 below.

A table with numbers and letters

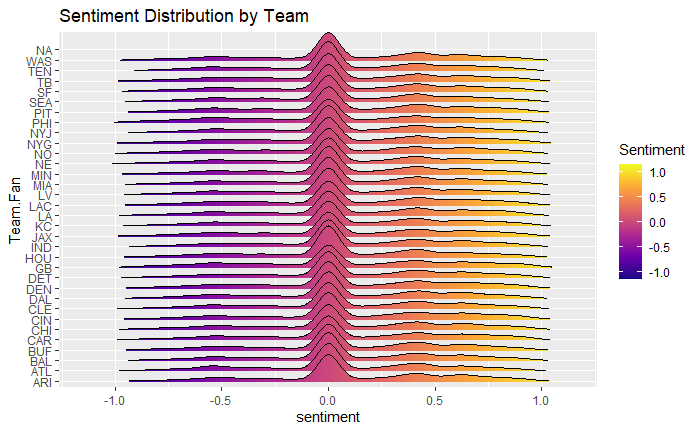
Description automatically generated

Analyzing the offensive words and sentiment score we discovered a few interesting facts, traditional offensive languages are generally associated with poor sentiment scores. The count of the offensive words with low median sentiment can therefore be an interesting variable to predict the draft pick.

**iv.Team Sentiment**

Beyond the player’s draft positions and social media vernacular, we had a final question of the draft data about how to compare the effect of each team’s fans on the comment collection process as most comments were generated by fans, and fans are generally loyal to a team as opposed to the entirety of the NFL.

By and large, the team’s sentiments had similar shapes clustered around zero but with a slight positive skew (Figure 7). This is in line with the general distribution (Figure 3) and confirms that no team is responsible for a disproportionate number of positive or negative comments.

A graph with a line and a line

Description automatically generated

Figure 7: Sentiment scores by team Figure 8: Proportion of comments to sentiment by team

One avenue we intend to explore is whether average sentiment and proportion of count both have predictive power independent of each other.

1. **Player Performance Data Methodology**

Table 2: Wins Above Replacement by Position

A table with numbers and letters

Description automatically generated

Player evaluation is a complex area of NFL analysis. We initially considered using fantasy football data but discarded it due to inconsistent scoring methodologies. We then turned to Pro Football Reference (PFR) and extracted data for various positions. However, data limitations for positions like punters and offensive linemen necessitated their exclusion from the initial analysis.

Seeking a more established metric, we explored Wins Above Replacement (WAR), analogous to baseball's WAR. However, a standardized WAR metric for football is unavailable. We instead utilized Eager and Chahrouri's "PFF WAR: Modeling Player Value in American Football" to adjust fantasy PPG based on positional value and variance as shown in Table 2.

This approach, while promising, presented challenges in balancing factors that contributed to the scaling factor. Our findings revealed model limitations and inconsistencies, exacerbated by data absence for certain positions. Consequently, we made the critical decision to reassess our methodology.

**Methodology: Utilizing PFF Player Grades as the Baseline Metric**

Our current methodology for assessing player performance utilizes PFF's "player grade" as the foundational metric. This selection is driven by PFF's comprehensive evaluation of individual contributions on the football field. Key benefits of PFF's grading system include:

• Contribution to production: PFF's grading system emphasizes a player's "contribution to production" on every play, disregarding inherent traits or measurable attributes.

• Position-tailored grading scale: PFF employs a position-specific grading scale ranging from -2 to +2 in 0.5 increments, ensuring an impartial assessment of each role's unique demands.

• Diverse grading panel: Over 600 analysts, including former players, coaches, and individuals with diverse backgrounds, contribute to the grading process. Final grades are determined by the organization's top analysts.

0-100 scale for easy player comparisons: PFF's grades are converted to a 0-100 scale at both the game and season levels, facilitating straightforward player comparisons.

**Data Collection and Cleaning Methodology**

.

To construct a comprehensive dataset, we gathered regular season player performance data from PFF (2015-2022) and draft summary information from Pro Football Reference (2015-2020). We standardized team names and player positions for data uniformity. This unified dataset lays the foundation for our analysis of player performance, draft outcomes, and their interconnected dynamics.

**Early Assessments of Draft Performance**

We have created a slightly modified metric "Raw Value Provided (RVP)" to assess the total value of a player within a specific season. The calculation is:



To comprehensively evaluate the performance of players from the 2020 NFL Draft, we initiated our analysis by establishing a baseline. This baseline was constructed by assessing players drafted in the years 2015 through 2019. Since we only have three full years of performance to look at for the 2020 class, only the first three subsequent seasons were considered for each draft pick. This is less than ideal as it does not consider each position having unique characteristics such as average career longevity and average "peak" year, but it is a necessary control to establish a baseline. The total RVP over three seasons was calculated for each member of the 2015-2019 draft. An aggregated average total 3-year RVP was then calculated for each possible position-round combination. (Figure 9)

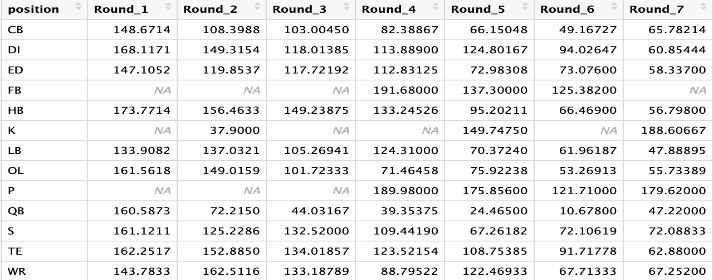


Figure 9: Average 3-year Raw Value Provided

Using the aggregate averages, it was then possible to determine the difference between expected and actual three year RVP for the 2020 draft class. The resulting top draft picks displayed a good mix of generational talents in the first few rounds and value picks in later rounds.

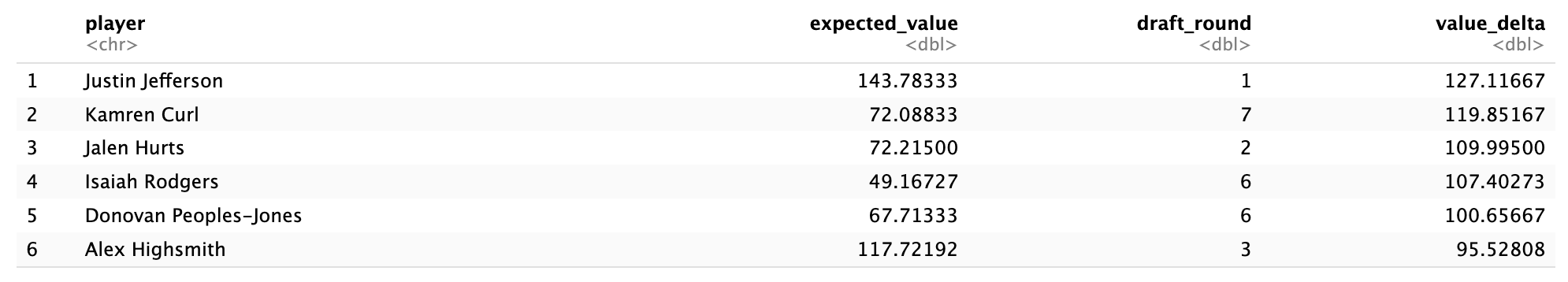


Figure 10: Expected Player Value for Top Draft Picks

Aggregating the total difference in expected value on the teams, we find that based on our current methodology the Titans had the worst 2020 draft with a total difference in expected value of -306.1, while the Bengals had the best draft with a difference of 285.4. This is consistent with many leading NFL draft analysts and platforms.

**Overview of Modeling**

**Data Sources and Code repository**

https://github.gatech.edu/MGT-6203-Fall-2023-Canvas/Team-79

**Works Cited**

Galton, Francis. “Vox Populi.” *Nature News*, Nature Publishing Group, www.nature.com/articles/075450a0. Accessed 3 Nov. 2023.

Ozanian, Mike. “NFL National Revenue Was Almost $12 Billion in 2022.” *Forbes*, Forbes Magazine, 14 July 2023, www.forbes.com/sites/mikeozanian/2023/07/11/nfl-national-revenue-was-almost-12-billion-in-2022/.

Eager, Eric, and George Chahrouri. "PFF WAR: Modeling Player Value in American Football." MIT Sloan Sports Analytics Conference, n.d., www.sloansportsconference.com/research-papers/pff-war-modeling-player-value-in-american-football

Dean, Jeff, et al. “Grading the 2020 NFL Draft Class.” Sports Info Solutions, 23 May 2023, www.sportsinfosolutions.com/2023/05/09/grading-the-2020-nfl-draft-class/.

“NFL Draft: Which Teams Have Gotten the Best Value since 2012?” ESPN, ESPN Internet Ventures, 22 Apr. 2022, www.espn.com/nfl/story/\_/id/33297949/nfl-draft-which-teams-gotten-best-value-2012-rank-all-32.

**Team Background**

* Hugh Hoagland (hhoagland6) has a BS in Statistics and Business from the University of Tennessee.
* Raju Ahmed (rahmed73) has BS & MS in Applied Physics, PhD in Materials Science and Engineering from Texas State University.
* Pablo Francisco Ramos Soszna (psoszna3) completed his Undergrad in Business Technology from the University of Miami.
* Alexander Martin Stetzer (astetzer3) completed his Undergrad in Physics from UW-La Crosse and Aerospace Engineering UM-Twin Cities.
* Nicholas Weist (nweist3) completed his Undergrad with a dual major in Math and Econ from the University of Connecticut.