Ignacio Dowling

STAT 301

December 6, 2023

Professor Sass

STAT 301 Final Project: Executive Summary

GitHub Repo Link: <https://github.com/stat301-1-2023-fall/final-project-1-idowling11>

By cleaning, merging, and exploring different datasets covering NFL Draft Data – on prospects’ physical information, statistical information, and career accolades and outcomes – I looked to explore whether some of those metrics and factors influenced their career successes and long-term draft value for different teams. Additionally, I examined whether certain elite physical and statistical metrics in different categories played a role in helping quarterbacks and linemen earn high draft selections.

**Purpose**

Although scouting in other sports may be easier, the NFL Draft is regarded as a very tough setting for professional general managers to hit the jackpot on a player and earn great value. It’s widely-debated what predictors of "success,” “value,” and even “high draft status” are for different positions. In fact, even defining those terms for different NFL draftees is a difficult thing to do. Examining a combination of datasets in a number of different areas – rather than just a few statistics in one area of the player’s identity – allows for the possibility to develop a greater understanding of each player's profile: their pre-draft background and their career outcome.

In doing so, the combination of these datasets – as detailed in the appendix – allows for users to explore 30-year and 36-year historical ranges: 1985 to 2015 and 1985 to 2021. With a large historical base consisting of thousands of draftees, exploring this data can lead to greater progress toward defining and developing a critical understanding of the terms mentioned above.

**Major Results**

Striving towards an understanding of the terms above, this EDA looks at three areas: (1) the driving factors behind NFL career success, and how one’s draft spot relates to that success, (2) factors that potentially indicate high draft status (getting selected earlier and higher ESPN prospect grades) for quarterbacks and offensive/defensive linemen, and (3) the factors behind defining a long-term “high value” draft pick, identifying them, and identifying which teams have succeeded the most in doing so.

Firstly, I found that a high approximate value for a player, or AV, correlates with career success. It shows a positive relationship with accolades like Pro Bowl appearances and First Team All-Pro Selections. However, a very small number of draftees record even a marginally-high AV, and the chance of selecting an AV player begins to fall dramatically after Round 1 of the draft. The median AV for each pick steadily decreases by pick in the first round, but afterward it remains relatively constant.

Secondly, I found that for the most elite quarterbacks, a combination of historic college estimated points added (EPA) and historically-high quarterback rating (QBR) tend to lead to a very high draft status. The relationship isn’t as strong when considering only one variable, but QBR has a stronger relationship with draft status than EPA does. When considering all quarterback prospects, height slightly plays a role, but it isn’t an end-all, be-all regarding pick number – it plays a greater role in influencing a QB’s ESPN prospect grade. For linemen, height and weight are marginal when it comes to its influence on draft status; however, teams in the 2020s are beginning to select heavier offensive tackles and guards higher, while high-ranked defensive ends are a little lighter than they were in the 2000s.

Finally, I developed a new metric: AV over expected – to gauge each draftee’s value depending on their career success relative to others selected near their draft spot. I found that Tom Brady is the highest-value draft pick in the datasets used, and that the Green Bay Packers and Pittsburgh Steelers have fared the best when it comes to identifying high-value players in later rounds (after Round 2) since 1985. Both of those teams are also the two best teams when adding up the teams’ entire accumulated approximate value over expected. That metric has shown to increase over time since most draftees have low career expectations to begin with. Ultimately, these are the biggest conclusions for each exploration, helping us equate success with approximate value, “value” with AV relative to similar draftees, and “high draft status” for quarterbacks and linemen as a product of a number of factors, namely EPA and QBR for the elite of the elite quarterbacks.