Identifying NFL Pro Bowl Success from College Physical Attributes and Rankings

Abstract: The growing presence of the National Football League as a business and as America's biggest and most pressurized scene has increased the pressure of teams needing to have higher success. As such, they need to get better players, and that starts by drafting more superior college players. Previous research has mostly looked at what rank players were drafted at and used that to make predictions on the amount of pro bowls. A clear limitation of this research is that it does not take into account college stats and attributes, merely their ranking in the NFL draft once they declared for the draft. However, our research also includes the datasets of both college and NFL combined which include characteristics of how they got their rankings, such as their age, height, and statistics and college. We used numerous datasets from GITHUB and Kaggle and performed our machine learning and data code to understand the most important statistics when evaluating college athletes and how they will translate to the NFL level in order to maximize the potential number of pro bowlers. We looked at every football position and some positions actually have data suggesting that playing those positions actually decreases your chances of making a Pro Bowl, it's actually detrimental for safeties and receivers and beneficial for special teams and quarterbacks. Contrary to the more prevalent opinion, it turned out that having better or worse physical tools, college stats, and rankings had no measurable impact on predicting pro bowls. In fact, the most crucial category turned out to be age, as we discovered that drafting players between the ages of 21-23 is the single best thing that a team can do to maximize their chances of drafting a pro bowler.

The National Football League. So long it has been the single most exciting pastime in the United States and is currently what Americans spend the most time doing with their time. We all make time for the Super Bowl, and the NFL playoffs, but just how difficult and unlikely is it to actually succeed in the NFL? Every year, the NFL draft attempts to pick the best players from college and teams line up in order to pick them and hopefully draft a superstar to help their team succeed. But drafting superstars is a lot harder than it sounds, as teams spend millions of dollars on scouts, and countless hundreds of hours scouting players all the way from high school into college just to feel certain that they will be great for their respective team only for it to fail. You might be surprised as to just how often this actually happens, as every year, as the majority of first-round picks end up being “busts”. In short, drafting superstars is a lot easier said than done, and yet the benefits are enormous for a franchise’s success, both economically and for winning Super Bowls. For this research paper, I will be using Pro Bowls as a level of success, where the Pro Bowl is an event that is fan-voted and helps identify some of the best players at their respective positions every year.

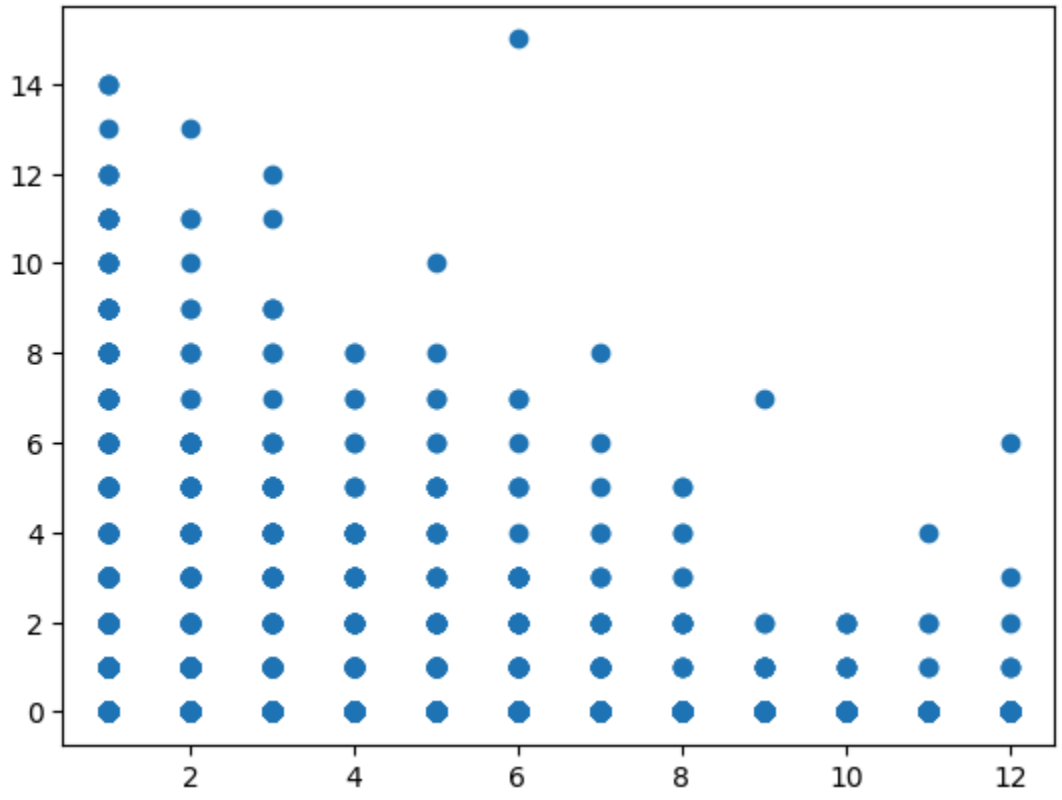
First of all, let’s examine a few successful cases of teams drafting absolute stars, and what that actually did for the franchise. (Brady for the Patriots). When the Patriots drafted Tom Brady in 2001, they were one of the poorest and least successful franchises of any team and only had a net worth of $571 million. In 2002, that was one of the lowest ones. Fast forward to the current day, and it couldn’t be more different. Now, the franchise has a value of $7 billion, which is the exact opposite of the spectrum, as now they are ranked all the way at number 2 in the US, and crazily, number 3 IN THE ENTIRE WORLD FOR ALL FRANCHISES. They literally went from one of the worst in their own sport to number 3 in the entire globe for all years combined after Tom

Brady, an absolute icon was drafted to them and helped them win 6 Super Bowls. So, in short, drafting good players helps win Super Bowls which consequently then helps increase net valuations, which is obviously what owners want. That's why this data is important because owners really do have a motivation to draft the best players possible in order to gain the most success and profit.

RANK ^	TEAM	STATE	CURRENT VALUE	CHANGE IN VALUE (%)	DEBT/ VALUE (%)	REVENUE	OPERATING INCOME
1	Dallas Cowboys	TX	9 B	13	3	1.14 B	504 M
2	New England Patriots	MA	7 B	9	4	684 M	206 M

Second of all, let's examine why we chose pro bowls. These are the 6 teams that have drafted the most pro bowlers since 1990, and there is a clear pattern here. The teams are: the Chiefs with 27, the Cowboys with 26, the Packers with 26, the Ravens with 25, and the Steelers with 23. It is not a coincidence that these are some of the most respected and historical franchises during that time frame. The Packers, cowboys, and Steelers have the fourth, third, and second most Super Bowls of any time, while the Bears and Ravens are not far behind them. As such, it is reasonable to assume that drafting pro bowlers is definitely beneficial to a franchise.

Every year, college players are evaluated, in order and told where they will end up being drafte



d. Well, just how do the NFL teams decide who to pick though? There are a couple of important variables here to consider that teams value with extreme importance including general statistics like age, height, weight, school, and conference, but also position-specific statistics including passing yards and touchdowns(for quarterbacks), sacks(for defensive lineman), and interceptions(for defensive backs). We will consider all of these stats as well as where they were drafted as that clearly helps establish a set criteria for the standard of players.

Parenthetical citation: We obtained Pro Bowl information from NFLverse (Carl et al., 2023).

Inline citation: Carl et al. (2023) included features for age and number of pro bowls.

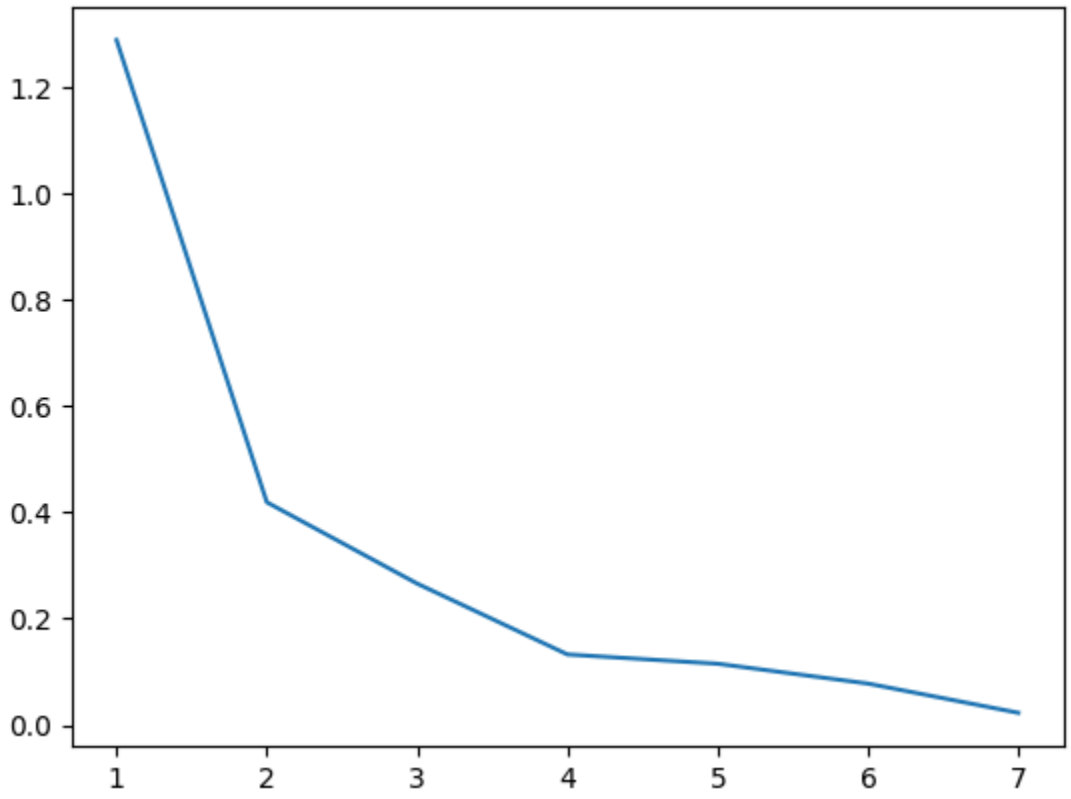
When we combined the data for all of these variables for college stats, it made sense as to how they were ranked, and a clear pattern emerged: the ones with the best physical tools, not necessarily the best stats, ended up becoming the highest ranked prospects coming out of college. Our goal is to determine what were the most important characteristics in determining NFL success.

We started by taking the following dataset
https://github.com/nflverse/nflverse-data/releases/download/draft_picks/draft_picks.csv
, which told us every single player drafted since 1980, which team drafted them, their position, all of their stats, and whether they made the Pro Bowl or hof or not.

We filtered that same dataset by draft round and the number of pro bowls and constructed the following scatterplot data tabletable
The x-axis is the round, and the y-axis is the number of pro bowlers. Please understand that pre-1997, there were 12 draft rounds, and after that, it was shortened to 7, as this dataset starts at 1980.

Before we proceed, it is crucial to understand one crucial thing about the draft. Higher-ranked players are supposed to do better than lower-ranked ones. Much, much better. That's why they were ranked as higher prospects in college, as they were expected to do better in the NFL based on their statistics and physical profile. So, in theory, the first round is supposed to be only for superstars, and the last round is only for "bad players"? Right?

Well, not exactly.



We developed the following correlation table scatterplot between the draft round and a number of pro bowls. This is between rounds and problems. This actually makes sense, because it helps support the age-old idea that higher-ranked players in college will do better in the NFL. As you can see, there is a steep drop off in the amount of pro bowls starting from round 1 and ending at round 7, an almost linear downward regression.

That's what it looks like at least.

However, we must remember to actually consider whether it is actually significant. To do this, we had to do it for all of the previous attributes mentioned above.

Because we now have both the college stats the clearly display all the characteristics of each individual player before the draft, and we now have all the information we need, like having

NFL stats that display the football stats of those same players once they got the NFL, we can then use the players name that correspond to each other in both sets in order to draw our conclusions. When we merged the datasets, by using a merge function with the player names and draft year on the left side, and the same player names and number of seasons on the right side. Finally, we took the correlation of the category pro bowls of this new merged dataset and these were the results we received.

```
draft_year          -0.061015
player_id           -0.074319
pick_x              -0.107696
overall             -0.296992
round_x             -0.296368
weight               0.031358
height               0.039447
pos_rk              -0.173052
ovr_rk              -0.275168
grade                0.278531
season              -0.061015
round_y             -0.300705
pick_y              -0.301102
hof                  0.224771
age                 -0.144693
to                   0.192580
allpro               0.718231
probowls             1.000000
seasons_started      0.585908
w_av                 0.719839
car_av                    NaN
dr_av                0.714595
games                0.488600
pass_completions     0.218380
pass_attempts        0.214325
pass_yards           0.221876
pass_tds             0.237306
pass_ints            0.187693
rush_atts            0.201545
rush_yards           0.207762
rush_tds             0.220432
receptions           0.292338
rec_yards            0.289848
rec_tds              0.287018
def_solo_tackles     0.312676
```

We continued by grouping the positions to simplify the criteria. The data we received came out in 20 unique positions, so we grouped them by positional similarity. We assigned a return type of “O-Line” for Centers, Offensive Guard, and Offensive Tackle, we returned the position for fullacks, running backs, and quarterbacks, “Secondary” was assigned for Cornerbacks, Safeties, and defensive backs, “DLine” was assigned for defensive end and defensive tackle, “Linebacker” was assigned for inside and outside linebackers as well as another subset group called linebacker, “Special teams” was assigned for punter, place kicker, kick returner, and long snapper, and finally, tight ends and wide receivers were assigned to catchers.

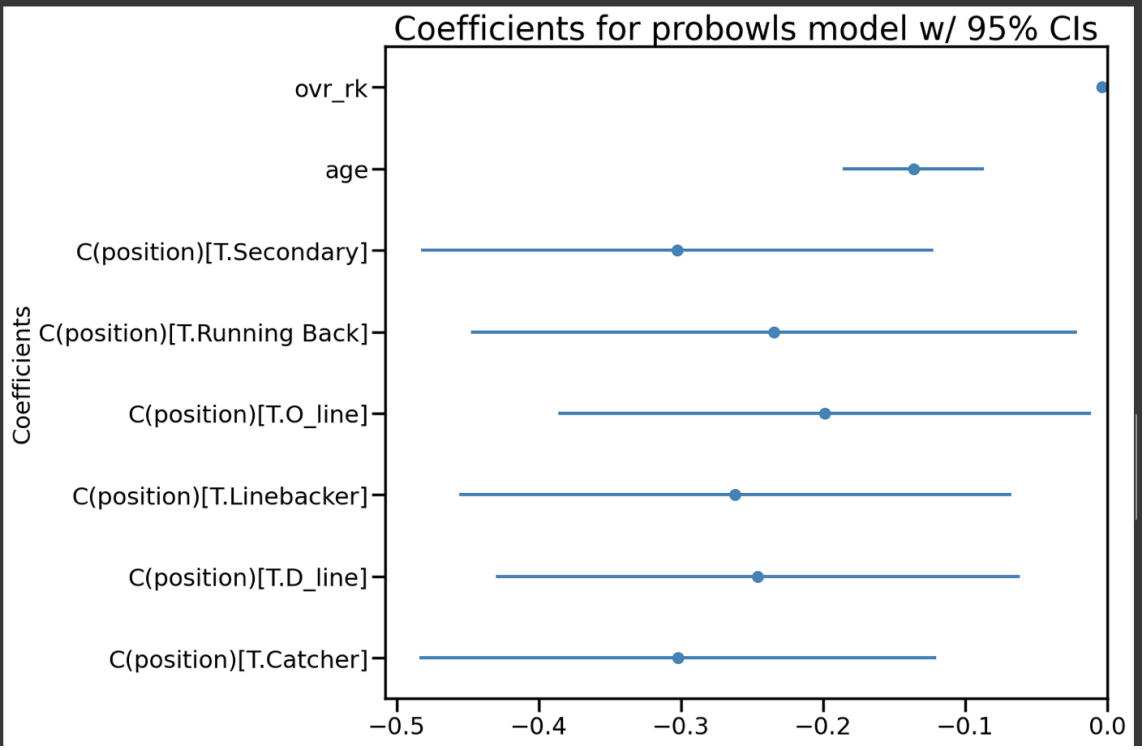
```
def simplify_position(position):
  if position == 'Center': return "O_line"
  if position == 'Cornerback': return "Secondary"
  if position == 'Defensive Back': return "Secondary"
  if position == 'Defensive End' : return "D_line"
  if position == 'Defensive Tackle': return "D_line"
  #if position == 'Fullback' : return position
  if position == 'Inside Linebacker': return "Linebacker"
  #if position == 'Kick Returner': return "Special_Teams"
  if position == 'Linebacker': return "Linebacker"
  #if position == 'Long Snapper': return "Special_Teams"
  if position == 'Offensive Guard' : return "O_line"
  if position == 'Offensive Tackle': return "O_line"
  if position == 'Outside Linebacker': return "Linebacker"
  #if position == 'Place Kicker': return "Special_Teams"
  #if position == 'Punter' : return "Special_Teams"
  #if position == 'Quarterback': return position
  if position == 'Running Back': return position
  if position == 'Safety': return "Secondary"
  if position == 'Tight End': return "Catcher"
  if position == 'Wide Receiver': return "Catcher"
  return "AA_Other"
```

Furthermore, a 95% confidence interval was constructed for the coefficients with pro bowl being the depvariable as “module 1”. Using this, we developed the following metrics: and a correlation coefficient chart.

```
[ ]  metrics(mod1)

     (8572.0, -4279.0, 0.086, 0.644, 43.0)
```

```
[ ]  coefplot(mod1, depvar)
```

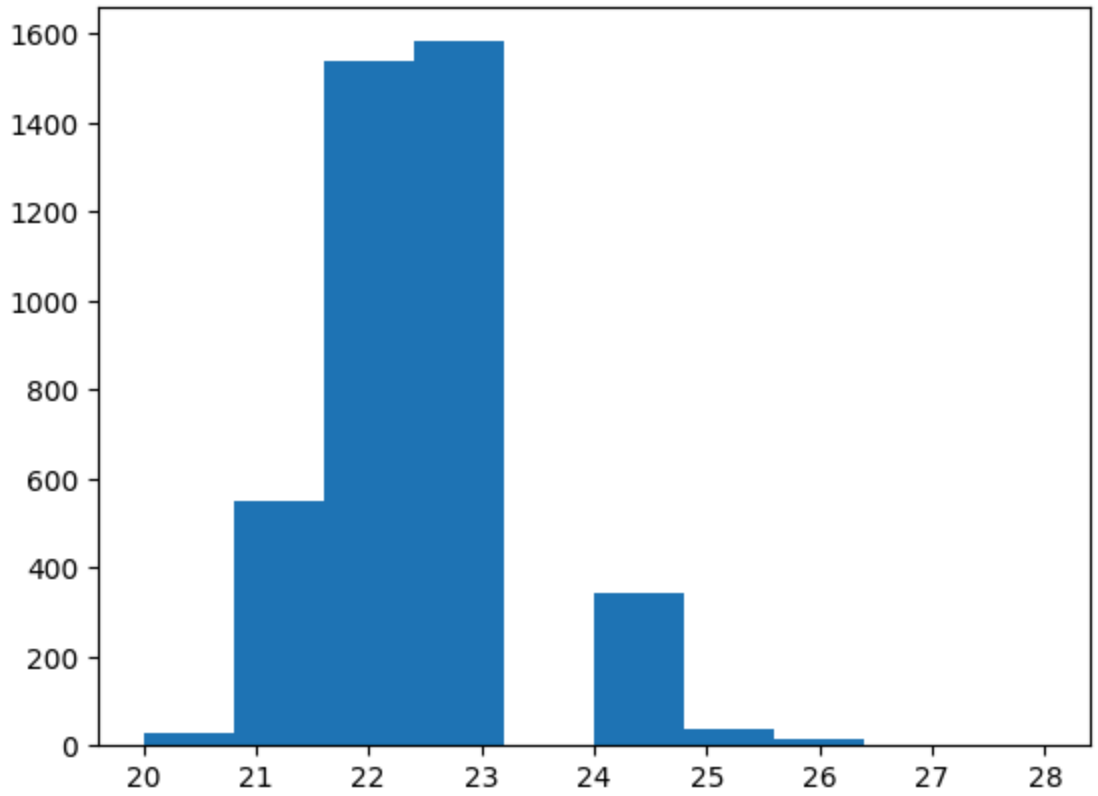


What you can clearly see here is that age is the only value that is not touching the “0” mark, meaning that it's the only significant value for all the positions when evaluating whether or not they will make a Pro Bowl. The rest of the variables had too much variance on the lines. The lines represent a possible range of values. As such, it looks like being a catcher or secondary player actually seems like the biggest indicators of not making a Pro Bowl, and most

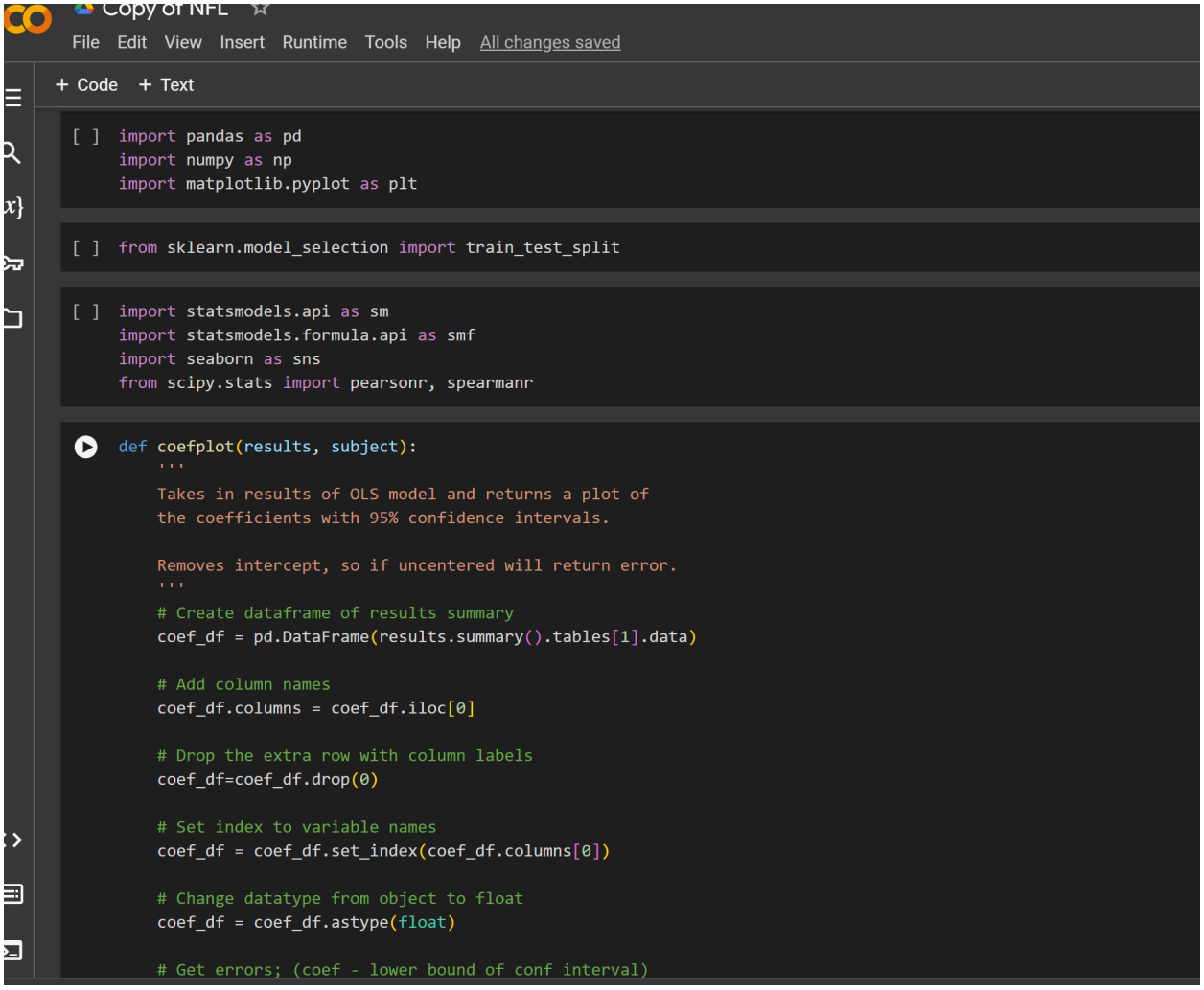
interestingly, overall rank is flat at 0 meaning there is no correlation whatsoever. It does look like age may be the most important factor when determining pro bowls.

There are a couple of possible reasons as to why age could be very relevant when making pro bowls. For one thing, if you enter the league younger, typically you would have a higher chance of making more pro bowls simply because you would have more time to advance your career and would have more years to potentially make one, plus the added bonus of having more time before their body begins to break down on them which usually happens in most athlete's mid-thirties. It is common knowledge that the younger you are as an athlete, the more elusive, quick, and physical you are able to be rather than when you are older. Another potential reason is that if they enter the league younger, they are able to be coached longer by NFL coaches, whereas in contrast spending more time in college vs the NFL understandably may yield more college success than NFL, but wouldn't necessarily allow for the same kind of progression as would being coached by an NFL Coach, since the pro bowl is an NFL award. Therefore, it is reasonable to assume that, If you only have a limited amount of time until you retire, being coached by an NFL coach rather than a college coach for longer definitely helps your case for making one, and that can be achieved by prioritizing younger players.

Finally, this was the last graph we had where we looked at in order to see that the ages between 21-23, when drafted, have the highest chances of being selected to the Pro Bowl and it may be a good idea to prioritize them over age groups.



More pictures of the code and what we did:



Citing Sources:

https://www.yardbarker.com/nfl/articles/six_teams_that_have_drafted_the_most_pro_bowlers_since_1996/s1_13132_31792388

https://www.statista.com/statistics/194482/franchise-value-of-the-new-england-patriots-since-2006/

https://www.statista.com/statistics/194482/franchise-value-of-the-new-england-patriots-since-2006/

https://www.forbes.com/lists/nfl-valuations/?sh=5a4eedc17386

https://www.forbes.com/sites/mikeozanian/2023/09/08/the-worlds-50-most-valuable-sports-teams-2023/?sh=78f5b44b2b44

Carl S, Baldwin B, Sharpe L, Ho T, Edwards J (2023). nflverse: Easily Install and Load the 'nflverse'. https://nflverse.nflverse.com/, https://github.com/nflverse/nflverse.

A

Authors:

Clayton Greenberg and Rishabh Raman