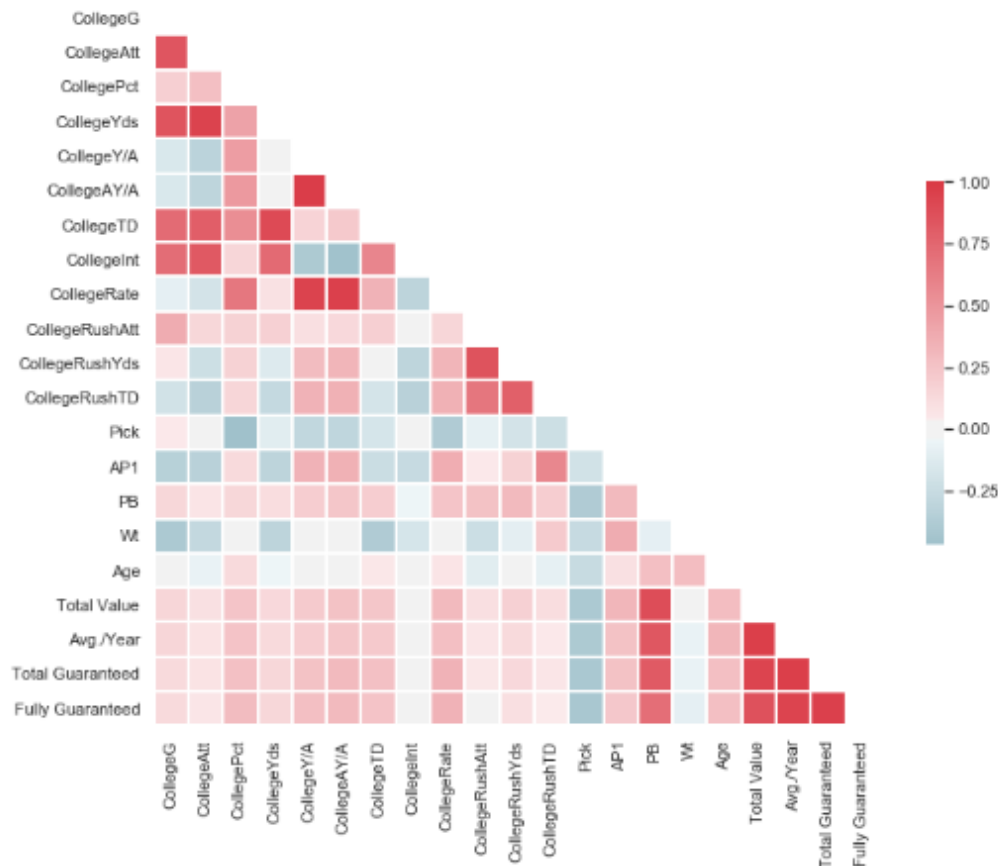


Summary: Research on Predictive Analytics in the NFL and an Application on Quarterbacks Prospects

Overview: For my research, I analyzed college football statistics to find if any predictive power exists with a player's performance in college to correlate with future NFL success. To do so, I relied on Python, a programming language, to run a regression analysis. Regression looks at the degree in which certain variables can help determine the likelihood of another variable occurring—the causal relationship between the two. This is a project I began my Junior Year of College and I recently looked at again after developing strong data analytical skills through my MBA program at Santa Clara University.

Methodology: Initially, I broke down the data by position group to look at the pertinent individual factors for each group. I then gathered data from 2005-2015 for each group while focusing on the players' high school recruiting profile, college career statistics, and combine data. After cleaning up the data, I used display packages to give an overview for each position group (see below for example).¹ This display package indicates the correlation for which key statistics a Yee & Dubin representative can focus in on. This insight reveals which factors increase the odds of signing successful athletes. Further, the team can develop a value strategy where the organization can focus on players projected in later rounds who still thrive in important categories.

Correlation Heat Map for Quarterbacks:



Application: To demonstrate the capability of this analysis, I will indicate how a Yee & Dubin can assess the Quarterback Position Group for 2020 as an example. After assessing the Correlation Chart above, we can see that certain college statistics bode well with NFL success. The primary categories for NFL success I relied on are “AP1” (First Team All Pro Selections), “Total Value” (A Sports Reference Metric for Career Value produced) and “Avg./Year” (what the QB makes per year). I think these are quality hallmarks for a successful QB—the more all pro appearances, money they get paid, and value of their career would be the type of player Yee & Dubin would want to target.

From here, I found the strongest correlations for these categories were College Adjusted Years/Attempt (College AYA), College QB Rating (CollegeRate), and College Completion Percentage (CollegePct). On the comparison, categories

¹ Analysis for each position group is ready and available upon further request.

like College Starts, Pass Attempts, and Touchdowns thrown have negative correlations with NFL success. This helps frame why certain prospects may be more likely to succeed in the NFL.

To determine the strength of the predictive abilities, an analyst can use regression models and look at the R-squared value. R-squared “is the proportion of variance explained, meaning the proportion of variance in the observed data that is explained by the model, or the reduction in error.” R-squared is “between 0 and 1, and higher is better because it means more of the variance is explained by the model.”² When using these tools, we can produce an adjusted R squared which range from 25% to as high as 40% with the QB Class. While this concludes we can’t explain 100% of NFL success, the agency can at least equip itself with tools to increase the odds of success by focusing on statistically significant results.

Next, I looked at the Class of 2020 prospects to see which quarterbacks thrive at key metrics.

NFL Class of 2020 Prospects: Sorted by Statistically Significant Results						
Player	School	Conf	Pct	AY/A	Rate	Projected Draft Selection
Tua Tagovailoa*	Alabama	SEC	69	12.8	199.4	1
McKenzie Milton*	UCF	American	67.1	11.1	179.3	UDFA
Jake Fromm*	Georgia	SEC	67.4	10.1	171.3	10
Jordan Love*	Utah State	MWC	64	9.4	158.3	UDFA
James Morgan*	Florida International	CUSA	65.3	9	157.6	UDFA
Nathan Rourke*	Ohio	MAC	59.9	9.2	156.3	UDFA
K.J. Costello*	Stanford	Pac-12	65.1	8.8	155	54
Ian Book*	Notre Dame	Ind	68.2	8.6	154	200
Zac Thomas*	Appalachian State	Sun Belt	62.6	8.6	152.6	UDFA
Andrew Ford	Massachusetts	Ind	63.2	9	151.1	UDFA
Brady White*	Memphis	American	62.8	8.7	150.7	UDFA
Jalen Hurts*	Alabama	SEC	60.4	9.3	150.2	138
Khalil Tate	Arizona	Pac-12	56.3	8.9	149.8	UDFA
Shea Patterson*	Michigan	Big Ten	64.6	8.4	149.8	85
Mason Fine*	North Texas	CUSA	64.6	8.8	149.4	UDFA
Alex Hornibrook*	Florida State	Big Ten	62.3	7.8	148.6	UDFA
Bryce Perkins*	Virginia	ACC	64.5	8	147.5	UDFA
Jack Abraham	Southern Mississippi	CUSA	73.1	7.2	147.4	UDFA

3

Analysis: I believe this exercise helps provide insight into where Yee & Dubin can find great value. As I discussed with Carter Crow, the firm must find mid-round value players to avoid the upfront costs of signing 1st round draft picks. While every agency would want Tua Tagovailoa as a client, the expenses to sign him are difficult to justify. However, a player like McKenzie Milton might be an excellent prospect for the firm based on his profile.

Specifically, Milton’s Sophomore year campaign saw him finish in the 99th percentile for Adjusted Passing Yards/Attempt, 98th percentile for QB Rating, and 90th percentile for College Percentage. Further, he backed his best season with a strong 3rd year where he showed strong AY/A (82nd) and QB Rating (87th). Despite this, he is currently not being drafted, likely because of his leg injury suffered at the end of his Junior season. Yet, this possibility opens up an opportunity for Yee & Dubin to sign a player for cheaper while his statistical background offers the upside of a NFL Starter, creating a surplus of value for the Agency.

Conclusion: This logic can reveal insight into how to use statistics to guide decision making for a Sports Agency even if Milton per se may not be the perfect prospect. The goal for a mid-level firm ought to be to find prospects with strong significant statistical backgrounds that will be selected in Rounds 3-7. This allows the firm to save money on the more expensive prospects while increasing their odds of finding successful NFL players.

Next Steps: As of Mid-June, I can repeat this same exercise for all NFL Position Groups besides Offensive Lineman. I have already successfully completed the historical analysis—step one: finding the most predictive statistics—and would simply need to apply these categories to prospects in the Class of 2020.

² https://nbviewer.jupyter.org/github/justmarkham/DAT4/blob/master/notebooks/08_linear_regression.ipynb

³ Draft Selection Numbers from Draft Site. <https://www.draftsite.com/nfl/position/qb/14/>