Quantifying NFL Quarterback Aggressiveness Using aDOT Over Expected (AYOE)

Amrit Vignesh[1*]

[1]Seminole High School, Sanford, FL

*Corresponding Author Email:
amritvignesh@gmail.com

Abstract

The quarterback of a football team is typically described to be the leader of an offense as their performance and playstyle can highly impact the characteristics of an offense in comparison to other positions. The aggressiveness of a quarterback in terms of their tendency to pass deep in any certain play can be determined by multiple factors related to situation (down, yard line, etc.) but also be affected by an inherent factor of trust in the system around them. Using an XGBoost model with a dataset split into training data from the 2014 to 2020 NFL season and test data from the 2021 to 2023 season for an approximate 70:30 split, the aggressiveness of a quarterback was quantified using the amount of air yards they generate per pass attempt compared to their expected amount influenced by the situational and system factors described previously. This helps transcend the use of just raw statistics related to air yards by accounting for various factors to give a more accurate statistical definition of aggressiveness.

Keywords: football, sports analytics, data science, statistics, machine learning

Introduction

Quarterbacks who are considered "elite" by the correspondent use of statistics and "eye test", or the process of evaluating players in a sport based on watching their film, may be very similar in terms of performance and team impact but highly different in playstyle. The quarterback position has evolved multiple times, especially in between the modern era from 1980 to the early 2000s, characterized by strong passers composed in the pocket with the need for accuracy, to the present-day game where multiple quarterbacks possess "dual-threat" ability as they balance their passing and running game using run-pass options to determine what is optimal for team success based on the situation of the game and recognition of defensive coverages [1]. In today's age, some quarterbacks play more aggressive than others and this can be determined by looking at their **aDOT** or average depth of target which is equivalent to the average amount of air yards. The yardage gained on plays is composed of the air yards generated by the quarterback's throw and the yards after catch generated by the receiver themselves. Offensive systems can be designed in different ways: the quarterback can intentionally be aggressive and act as a "gunslinger" by targeting deep receivers or act as a "game manager" by throwing checkdowns and short passes in order to build up momentum. Aggressiveness can be simply quantified by looking at the baseline aDOT statistic, however, using raw statistics with no respect to situation the quarterback is placed in and the system around him will not be a true measure of how aggressive a quarterback is as it will differ play-by-play based on situational factors. This calls for the creation of a new metric known as "aDOT Over Expected", abbreviated to AYOE as aDOT is just air yards per pass attempt. The expected amount of air yards generated on a certain play will be based on simple situational factors such as yard line, amount of seconds left in the half, etc. but also accounts for the system by looking at the performance of the targeted receiver before the game in which the target occurred. This will help show when quarterbacks exceed or fall behind the "necessary aggressiveness" on a certain play. We will also see whether this aggressiveness has any strong correlation to QB efficiency in order to tell whether there is preferred level of aggressiveness for maximum output.

Related Research

The average depth of target (aDOT) measures how deep a receiver is when targeted by the QB [2]. It has not only been used a statistic to describe quarterback Different forms of aDOT or Average Air Yards Over Expected have been created by other models using similar machine learning techniques but different inputs. In Early 2021, Tej Seth created his own form of AYOE where he exclusively includes situational factors (except accounting for defensive strength with defensive YPA). In this article, it is established that the AYOE metric created by the author and EPA (Expected Points Added) Per Pass are highly correlated [3]. The EPA metric has been popularized within the football analytics world as it represents a play-by-play version of what wins games according to Alok Pattani of ESPN who described the increasing use of the metric to describe efficiency and give an objective description of "best" in football [4]. Another AYOE model was created by Ajay Patel where he analyzes the progression of AYOE over time to see how aggressiveness has changed, possibly due to coaching changes which can lead to a system either focused on deep shots or game-managing plays [5]. Both of these models provide a great framework for the situational factors that can be included in the AYOE model in this research and provide ideas for interpretation by analyzing its correlation to QB passing efficiency and progression over the frame of the test data. The idea of creating "over expected" metrics popularized within football analytics is not new itself. These metrics in general are good at adjusting basic statistical values based on situations. Some "over expected" metrics already created are completion percentage over expected (CPOE), rushing yards over expected (RYOE), and yards after catch over expected (YACOE) [6]. The "expectation" aspect is generated by a machine learning model which includes a variety of situational factors. For example, for the CPOE metric, the depth of the throw and the location relative to the sideline are two primary factors influencing the creation of an xCP value (expected completion percentage). A high xCP pass is described to be a short pass towards the middle of the field compared to a low xCP pass being a long pass towards the sideline. In this case, completion percentage is just a binary variable between 0% and 100% or 0 and 1 for the real value for each passing attempt based on whether it is completed or not. The difference between the aggregate completion percentage and aggregated expected completion percentage is CPOE and it has proved to align positively with QB quality. The actual values of the metrics being described do not have to be binary like completion percentage on a passing attempt as it can describe a continuous value such as rushing yards in RYOE. Similar to how the validity and predictive power of these "over expected" metrics are described, this research aims to look at whether aggressiveness correlates to characteristics of a QB like efficiency.

Methods

In order to create an "expected air yards" model to reflect the expectation aspect of the metric, the XGBoost machine learning technique is utilized. XGBoost is an advanced technique which is said to solve "many data science problems in a fast and accurate way" [7]. Some features of XGBoost include the use of gradient boosting (the combination of multiple "weak" models which perform slightly better than random chance with each "weak" model correcting the errors of the previous in order to create one strong model), decision trees (utilizing simple decision rules in a tree-like structure like a hierarchical flow chart), and regularization (preventing overfitting through the use of multiple parameters in the XGBoost method).

Using nflverse data with the nflfastR library, the play by play data is extracted through the load_pbp function from 2014 to 2023 for both regular season and playoff games which means the data has a span of 10 years [8]. The throws are filtered so that they aren't a QB spike (which will intentionally have a negative amount of air yards) and it targets a certain receiver so that the system factor can be considered. Most of the situational factors are directly derived from the play by play data but the process to get the cumulative receiving EPA of the receiver in every game of that specific season before the game in which the respective pass attempt was completed was different. A grid was created with all unique receivers in each season from 2014 to 2023 with the maximum amount of weeks (17 weeks from 2014 to 2020 and 18 weeks from 2021 to 2023). Using the calculate_player_stats() function, the receiving EPA in each game was placed based on the correspondent season and week for each player and it is cumulatively added up to reflect the receiving EPA before the game. After performing this and attaching a value to each individual pass attempt with a targeted receiver, the following variables were selected for the XGBoost model: **season type (regular season or post season), yard line, amount of seconds remaining in the half, down, yards to go to get a first down or touchdown, shotgun formation, no huddle formation, quarterback dropback, expected points, win probability, score differential, and cumulative receiving EPA**. The **season type, down, shotgun formation, no huddle formation, and quarterback dropback** variables are categorical variables as they are either binary or are restricted to a set amount of categories. In order to train an XGBoost model with categorical variables, these variables can be converted into dummy variables with the creation of new columns all representing a certain category of the variable with a binary output dependent on if it fills that category or not. For example, for down, the new dataframe will use down.1, down.2, down.3, and down.4 instead as these are the categories of downs. The model is trained for seasons from 2014 to 2020 and tested on the data from 2021 to 2023. The following parameters were incorporated in the XGBoost model arbitrarily for maximum efficiency: **100 decision trees, the learning objective based on regression with squared loss, an early stopping rounds value to check if performance decreases every 10 rounds to reduce overfitting, maximum depth of a tree at 6, and a learning rate associated with prevent overfitting of 0.3** [9]. After generating air yards and predicted air yards on each pass attempt as applied to the train and test data, it is grouped by quarterback and year to get the total amount of air yards and predicted air yards. The amount of pass attempts is extracted from the initial play-by-play data to create columns for average air yards and average predicted air yards or aDOT and predicted aDOT. Subtracting these two values will give our AYOE value or aDOT over expected.

This value is visualized against the EPA (expected points added) per pass which is a measure of passing efficiency as a rate-based statistic to see whether there is a correlation between aggressiveness and efficiency as passers. Considering the test data is a range between 2021 and 2023, quarterbacks which played more than 1 season in this time range can be analyzed to see whether there are noticeable changes in their AYOE metric to draw conclusions regarding playstyle and offensive schemes.

Results

The most important factors within the AYOE model which utilizes train data for quarterbacks with at least 250 pass attempts from 2014 to 2020 are shown in the VIP plot below:

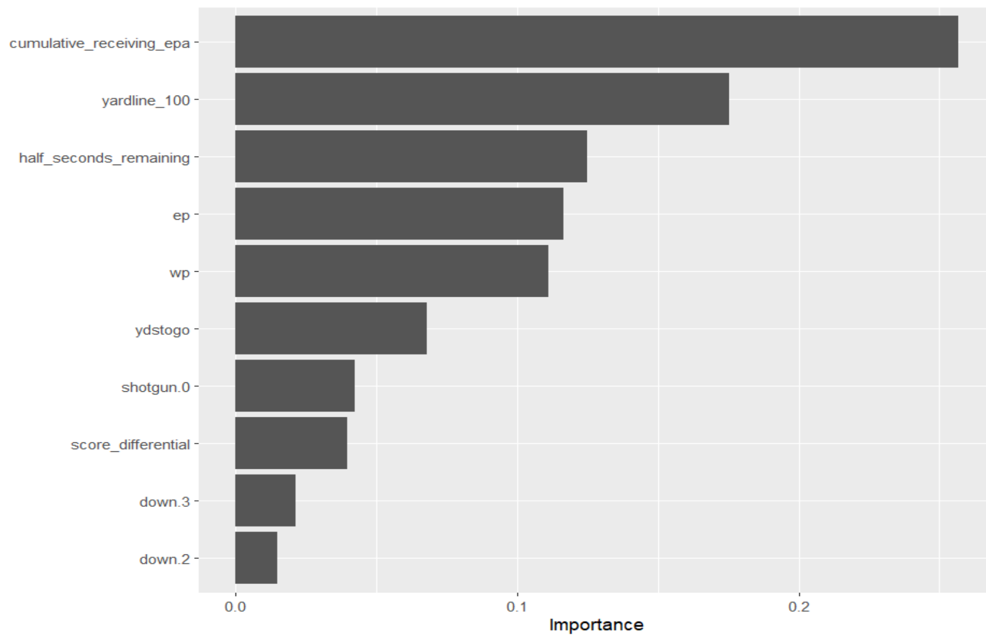
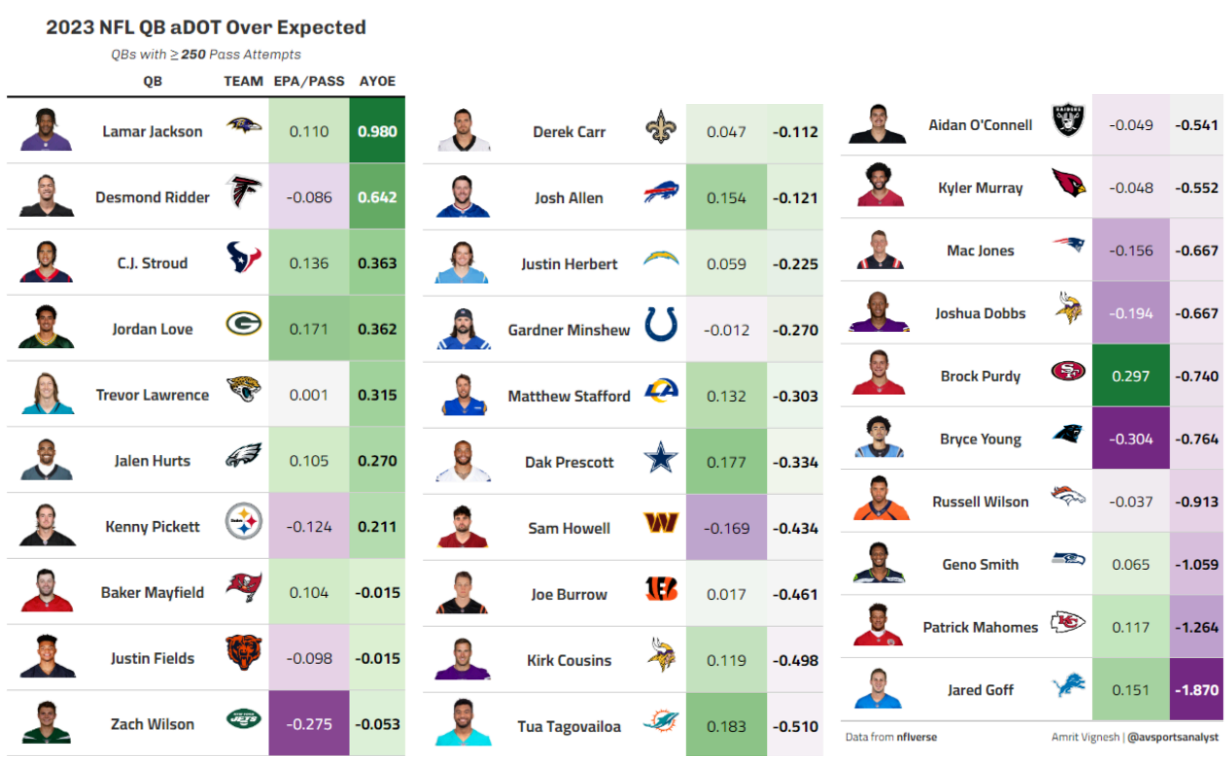


Figure 1: *Importance of Features in the AYOE Model*

After the creation of the model which creates an expected amount of air yards for every attempt, the aDOT over expected can be generated within the test data. For quarterbacks with at least 250 pass attempts, this is the leaderboard for aDOT over expected, arranged from highest to lowest:



2023 NFL QB aDOT Over Expected

*QBs with ≥ **250** Pass Attempts*

QB	TEAM	EPA/PASS	AYOE
Lamar Jackson		0.110	0.980
Desmond Ridder		-0.086	0.642
C.J. Stroud		0.136	0.363
Jordan Love		0.171	0.362
Trevor Lawrence		0.001	0.315
Jalen Hurts		0.105	0.270
Kenny Pickett		-0.124	0.211
Baker Mayfield		0.104	-0.015
Justin Fields		-0.098	-0.015
Zach Wilson		-0.275	-0.053
Derek Carr		0.047	-0.112
Josh Allen		0.154	-0.121
Justin Herbert		0.059	-0.225
Gardner Minshew		-0.012	-0.270
Matthew Stafford		0.132	-0.303
Dak Prescott		0.177	-0.334
Sam Howell		-0.169	-0.434
Joe Burrow		0.017	-0.461
Kirk Cousins		0.119	-0.498
Tua Tagovailoa		0.183	-0.510
Aidan O'Connell		-0.049	-0.541
Kyler Murray		-0.048	-0.552
Mac Jones		-0.156	-0.667
Joshua Dobbs		-0.194	-0.667
Brock Purdy		0.297	-0.740
Bryce Young		-0.304	-0.764
Russell Wilson		-0.037	-0.913
Geno Smith		0.065	-1.059
Patrick Mahomes		0.117	-1.264
Jared Goff		0.151	-1.870

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Figure 2: *aDOT over Expected Leaderboard With EPA/Pass and AYOE Displayed*

The aDOT over expected can be visualized against EPA per pass to see whether there is a correlation between two variables, and if there is, what direction it is in.

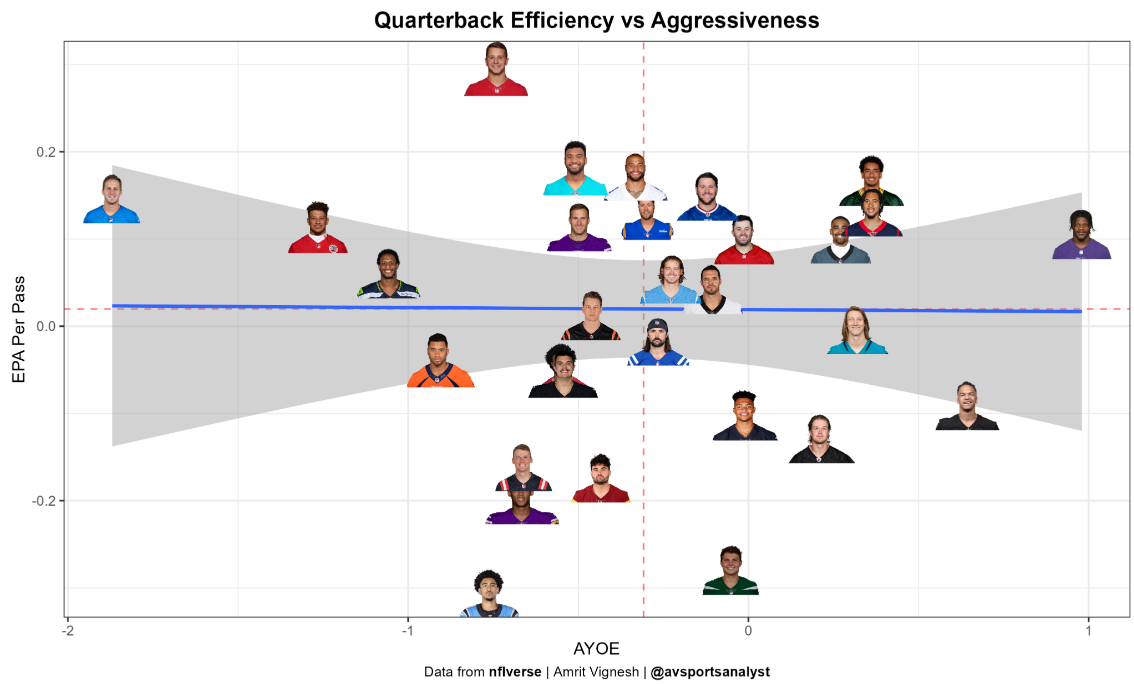


Figure 3: *Visualizing QB Efficiency vs Aggressiveness*

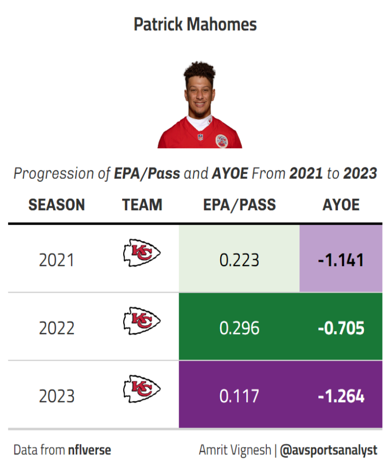
For quarterbacks who had at least 250 pass attempts each year from 2021 to 2023 (the test data range), their efficiency and aggressiveness progression can be analyzed. There are eighteen quarterbacks which satisfy this requirement:

Jared Goff

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		-0.058	-1.645
2022		0.172	-1.089
2023		0.151	-1.870

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**



Patrick Mahomes

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.223	-1.141
2022		0.296	-0.705
2023		0.117	-1.264

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Kirk Cousins

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.128	-1.281
2022		0.045	-0.784
2023		0.119	-0.498

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Justin Herbert

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.176	**-1.050**
2022		0.042	**-1.682**
2023		0.059	**-0.225**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Joe Burrow

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.151	**-0.445**
2022		0.144	**-1.275**
2023		0.017	**-0.461**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Kyler Murray

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.086	**-0.383**
2022		-0.047	**-0.286**
2023		-0.048	**-0.552**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Matthew Stafford

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.223	**0.151**
2022		-0.079	**-0.673**
2023		0.132	**-0.303**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Dak Prescott

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.149	**-0.951**
2022		0.129	**0.158**
2023		0.177	**-0.334**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

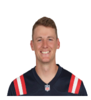
Trevor Lawrence

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		-0.100	**-0.043**
2022		0.116	**-1.375**
2023		0.001	**0.315**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Mac Jones



*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.097	**-0.526**
2022		-0.064	**0.339**
2023		-0.156	**-0.667**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

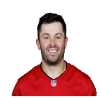
Derek Carr

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.059	**-0.170**
2022		0.058	**0.374**
2023		0.047	**-0.112**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Baker Mayfield



*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		-0.016	**0.497**
2022		-0.150	**-0.804**
2023		0.104	**-0.015**

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Josh Allen

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.171	0.248
2022		0.136	1.072
2023		0.154	-0.121

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Russell Wilson

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.041	1.107
2022		-0.116	0.493
2023		-0.037	-0.913

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Jalen Hurts

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.019	1.122
2022		0.165	0.050
2023		0.105	0.270

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Tua Tagovailoa

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.054	-0.476
2022		0.243	1.213
2023		0.183	-0.510

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Lamar Jackson

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		0.050	1.502
2022		0.029	1.001
2023		0.110	0.980

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Justin Fields

*Progression of **EPA/Pass** and **AYOE** From **2021** to **2023***

SEASON	TEAM	EPA/PASS	AYOE
2021		-0.240	2.201
2022		-0.207	1.331
2023		-0.098	-0.015

Data from **nflverse** Amrit Vignesh | **@avsportsanalyst**

Figure 4: *Efficiency and aggressiveness progression from 2021 to 2023*

The development of the aDOT over expected metric and its relation to efficiency can be analyzed further to reveal information about quarterback play styles and strategies.

The code for the project can be viewed in [10] in the references.

Discussion

The incorporation of the cumulative receiving EPA to represent receiver skill proved to be the most significant factor in terms of influencing the amount of air yards generated by each pass attempt. This exemplifies the importance of trust as a quarterback considering the offensive system around a quarterback initiates the strategy used, not necessarily just for the length of throws with the question of being a game manager or not, but also influencing the use of designed rushes versus passes at various rates. For example, Lamar Jackson does have the highest aDOT over expected but he also does not have as many pass attempts as other quarterbacks due to his tendency to act as a dual-threat quarterback and rush himself or create designed rushes. This also correlates with the use of RPOs (run-pass options) where the quarterbacks recognize the defensive formations and tactics in order to make a proper decision on what to do after the ball is snapped.

As shown by the quarterback efficiency vs aggressiveness plot, there is barely any non-zero correlation between the two factors as quarterbacks seem to be scattered throughout the plot with no clear relationship between the variables. As indicated by Tej Seth in a follow up article, the average depth of target carries less weight than it used to do due to its decreasing strength of correlation with efficiency, indicating that the lack of correlation between AYOE and EPA per pass in this model is supported by other literature as aDOT over expected is an adjustment to normal aDOT based on the situational factors and the receiver trust in this case [11]. However, other models could still carry a decent correlation especially if they do not account for factors like receiver trust.

For the individual quarterback progressions with the eighteen quarterbacks who passed at least 250 pass attempts from 2021 to 2023, the trend seems to be pretty random for most of the quarterbacks as reflected in the plot for quarterbacks in the 2023 season. There are few quarterbacks such as Patrick Mahomes with a direct correlation between EPA per pass and AYOE and Jalen Hurts with an inverse correlation between EPA per pass and AYOE. This could give insights on how the quarterbacks perform individually and whether it will be more valuable to be aggressive or be a game manager, but it still seems to have a pretty random correlation overall.

There are obvious limitations associated with the design process of the aDOT over expected model. Using cumulative receiving EPA to measure receiver's skill is not perfectly translatable, which applies to most sports metrics. Expected points added is inherently a team statistic as each play is assigned a certain value rather than each individual, which would require more advanced tracking data if each player wanted to have a certain value based on how much they contributed to the play. In addition, the cumulative receiving EPA is based on the games before the game in which the pass attempt occurred, which is not totally representative of trust because a late pass attempt in a game can be based on the trust developed throughout the game rather than the games before that game.

The parameters used for the XGBoost model are arbitrary but have been commonly used for the development of various models, which presents itself as a possible limitation. The features included in the XGBoost model may not have any level of causality in relation to the amount of air yards generated by the quarterback on a certain pass attempt as feature selection is subjectively based. However, XGBoost reliably handles feature selection by using different sets of variables for each iteration.

If more tracking data had widespread availability, further steps can be taken to measure receiver skill on a play-by-play basis as pass attempts can be attributed to the quarterbacks, receivers, and even the offensive linemen simultaneously. There are many unseen factors which can influence a quarterback's decision-making process which cannot be directly quantified.

Conclusion

Quantifying aggressiveness for quarterbacks accounting for various situational factors and trust helps go beyond the simple tracking statistic of average depth of target and expands it to an over-expected value with the expected value differing from other models with the inclusion of receiver trust as a factor. Even though no correlation was found between aggressiveness and efficiency, coaches can analyze the progression of individual quarterbacks with this metric and create designed plays based on what maximizes their efficiency with the quarterback initiating the team's offense and success.

Using over-expected metrics in general is pivotal in sports analytics in order to describe performance of players and teams in reference to situational factors and other features included in these types of models. It will help contribute to a growing revolution in sports analytics where information through data translates to informed decision-making.

References

[1] The Evolution of the Quarterback: How the Position has Changed Over Time, https://evolution-of-the-quarterback.netlify.app/.

[2] R. Greer, NFL Receiving Leaders, https://www.nfeloapp.com/nfl-receiving-leaders/.

[3] T. Seth, Creating Ayoe: A model to measure quarterback aggressiveness, https://mfootballanalytics.com/2021/01/30/creating-ayoe-a-model-to-measure-quarterback-aggressiveness/.

[4] A. Pattani, "Expected points and EPA explained," ESPN, https://www.espn.com/nfl/story/_/id/8379024/nfl-explaining-expected-points-metric.

[5] A. Patel, "Developing air yards over expected: A look at quarterback aggressiveness," Medium, https://ajaypatel18.medium.com/developing-air-yards-over-expected-a-look-at-quarterback-aggressiveness-6f7437da339c.

[6] R. Greer, Over Expected Metrics Explained -- What are CPOE, RYOE, and YACOE, https://www.nfeloapp.com/analysis/over-expected-explained-what-are-cpoe-ryoe-and-yacoe/.

[7] T. Chen, A scalable tree boosting system, https://dmlc.cs.washington.edu/xgboost.html#:~:text=XGBoost%20provides%20a%20parallel%20tree,problems%20beyond%20billions%20of%20examples.

[8] nflverse • Data and tools for NFL analytics, https://nflverse.nflverse.com/.

[9] XGBoost parameters, https://xgboost.readthedocs.io/en/stable/parameter.html.

[10] A. Vignesh, GitHub: NFL QB AYOE Final, https://github.com/amritvignesh/NFL-QB-AYOE-Final.

[11] T. Seth, Average Depth of Target Carries Less Weight Than it Used to, https://sumersports.com/the-zone/average-depth-of-target-carries-less-weight-than-it-used-to/.