Analyzing NFL Quarterback Versatility

by Emeka Amadi, and Matt Bolaños

Motivation

In the realm of sports, analytics have emerged as a crucial element of decision-making over the past decade. Both teams and sports analytics companies have actively pursued insights derived from data for various purposes including player improvement, roster optimization, and player evaluation.

Our project is fueled by our deep-rooted passion for sports, particularly American football, and our unwavering interest in delving into the intricacies of the game. As a team comprising of individuals already involved in the sports industry or aspiring to make a mark in this field, we firmly believe that this project will not only align with our personal interests but also contribute significantly to the broader understanding of quarterback performance.

While there have been notable NFL versatility studies published in articles, such as the one featured in The Touchdown, we recognize the need for a more comprehensive analysis with enhanced data granularity. By leveraging advanced tracking data and play-by-play data, we aim to bridge this gap and provide deeper insights into the factors that contribute to quarterback success.

Background and Questions

By leveraging player-tracking data and play-by-play data, our project aims to shed light on the effectiveness of quarterbacks when throwing from different positions, angles, and distances as well as their overall performance in various areas of the field. Through this analysis, we seek to identify which QBs exhibit the most versatility.

Our main question we want to answer is, “Which NFL quarterbacks are the most adept at completing passes from different angles and directions?” Our goal is to bring quantitative measures of versatility into the discourse. Future refined studies on this topic could be leveraged in player evaluation frameworks at both the amateur and professional levels to improve roster decision-making.

Data Sources

[dataset_creation.ipynb]

Data Sources

In this project, we analyzed data from the first eight weeks of the 2021 NFL season. The data set is sourced from Kaggle and is an official data set from the NFL that was utilized for the NFL Data Bowl competition. It comprises various files including play-by-play data, player data, game data, tracking data, and scouting data. For our project, we will primarily concentrate on the play-by-play and tracking data.

The tracking data provides x-y coordinates for all players on the field, including the football, at a rate of 10 frames per second. This information will enable us to determine the locations from which quarterbacks throw passes and the positions of the receivers they target. Additionally, we will utilize the play-by-play to provide context to the tracking data. The play-by-play data includes relevant features such as pass completion status, the intended receiver, the offensive formation/drop-back type, and the number of yards gained. By leveraging these data sources, we aim to gain insights into quarterback performance and versatility.

Data Breakdown

The tracking data consists of eight files (./data/week1.csv to ./data/week8.csv) with over 1 million rows and 16 columns each. To focus our analysis, we filtered the data in dataset_creation.ipynb, extracting only QB, WR, RB, and TE data points during pass-forward or pass-arrived events. This streamlined approach ensured that we were working with the specific data we needed for a more effective analysis, and cooperated with GitHub's file size limit for our shared repository.

01_dataset_creation.ipynb

Preparing the Data

[dataset_creation.ipynb & functions.py]

Tracking Data Anomalies

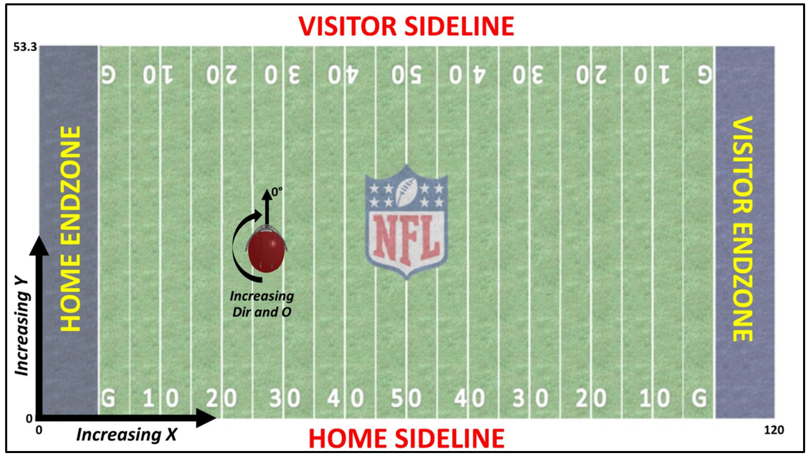
Given the volume of data points produced on each play, there were bound to be instances of unrealistic x-y coordinates. To account for these, caps were applied to values that exceeded the maximum or minimum values for the given dimensional coordinate using the .loc method in pandas. Values below zero would be set to zero, and an x value greater than 120 would be set to 120 (see the field diagram to get a better sense of the coordinate system), for example.

Adding Receiver IDs

The dataset did not provide who the QB intended to throw the ball to on a given play. This roadblock prompted us to parse the play-by-play data with regular expressions. The **playDescription** column of the play-by-play denoted the intended receiver (when applicable) in a pattern of “to {FirstInital}.{LastName}”, so a regular expression of **"to\s[A-Z]\.[A-Za-z]+"** was mapped to each play’s description. The player’s names in the player lookup were transformed to the same {FirstInital}.{LastName} format to join each receiver **nflId**. We were then able to join this information to the tracking dataset on **nflId, gameId,** and **playId**.

QB -> Receiver Angles

Much of our analysis centered around the ability of quarterbacks to complete passes from multiple angles. To derive this, we used the **atan2** function for the math library to calculate the arc tangent of the difference between the receiver and QB’s x-y coordinates before using the **degrees** function from the same library to convert from radians to degrees. The resulting column had values spanning from -180 to 180, which were subsequently divided into buckets of 45 degrees using the **cut** function from pandas. The negative angle values could re-mapped by adding 360 degrees to them in order to make the value range 0 to 360 degrees, but our analysis only cared about the distribution between these values rather than the magnitude of the values themselves.



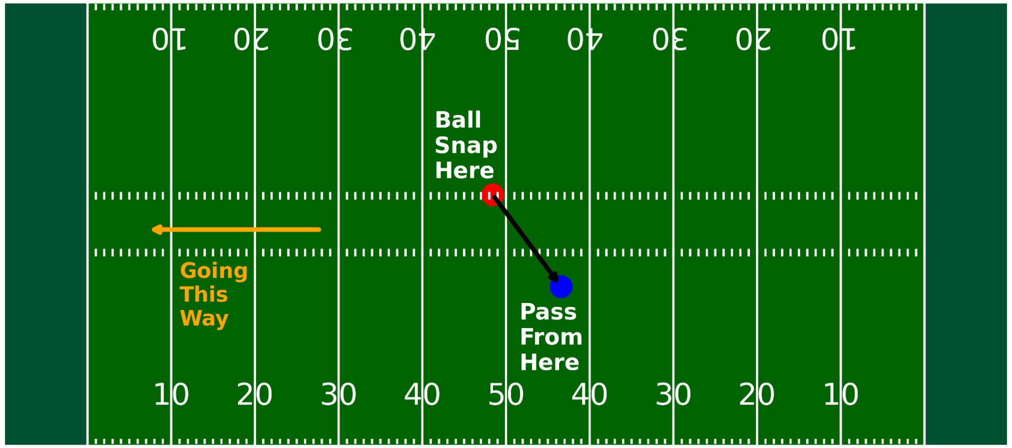
Preparing the Data (cont.)

[dataset_creation.ipynb & functions.py]

Quarterback Movement Directions

To identify if a quarterback was moving to their left or right when throwing the ball, we filtered for rows that corresponded to the snap of the ball and the frame of the pass being thrown, according to the **event** column in the tracking data. We then joined these coordinates back to the tracking data on the **nflId**, **gameId**, and **playId** columns. The difference between these y coordinates in tandem with the **playDirection** column allows us to determine if the QB threw while moving left or right (e.g. if the QB's y position when the ball was snapped is greater than their y position when the pass was thrown while their team is driving towards the left endzone, then the QB is moving to their left).

Example of a Quarterback Moving to Left to Pass



Gini Coefficient

Gini coefficients are a way to measure statistical dispersion, or "equality" (will use balance going forward). A value of 0 represents perfect equality (e.g. an array of 4 values where all the values are .25), and a value of 1 represented perfect inequality (e.g. an array of 4 values where the values are 1, 0, 0, 0). Using the angle buckets that we derived for the arctangent between the quarterback and the receiver, we grouped by each quarterback's **nflId** and calculated the proportion of their passes that were thrown in each bucket. After pivoting the dataframe wider, we treated each angle bucket column as a value in an array and applied our Gini coefficient function to all rows via a **lambda** function. The same process was repeated for completed passes.

One challenge that popped up immediately was small sample size players skewing the data. A player who only throws one pass will be assigned a Gini coefficient of 1, which is not really helpful in our goal of measuring quarterback versatility. To combat this, a minimum of 25 completed passes and 50 total passes was required to be included in the analysis.

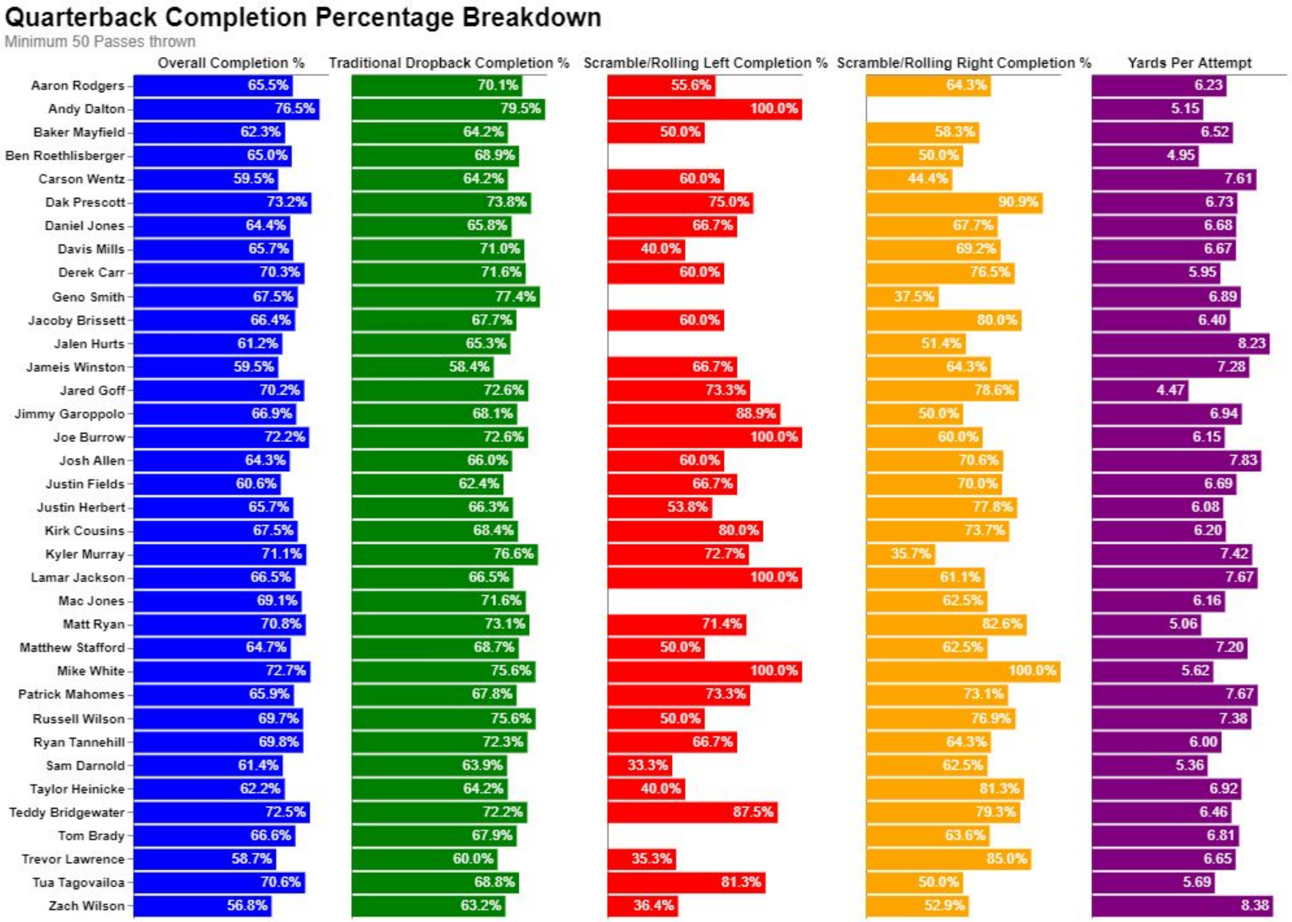
Data Analysis and Visualizations

[qb_breakdown.ipynb]

Visualization Process

The visualization on the right displays overall completion percentage, percentage during traditional dropbacks going left and right, and the distance between QBs and receivers at the time of the pass. While completion percentages alone does not provide a comprehensive evaluation of quarterback talent, it serves as a proficiency indicator. Successful QBs often maintain a high completion percentage while also achieving high yards per attempt.

The charts suggests that the majority of quarterbacks demonstrate higher completion percentages on traditional dropbacks. Interestingly all QBs, save the left-handed Tua Tagovailoa, exhibited higher attempts and completion percentages when rolling right. This finding aligns with our belief that QBs predominantly roll towards their dominant hand. Additionally, the graphic highlights an inverse relationship between completion percentage and yards per attempt. The ultimate goal for QBs is to maintain a high completion percentage while increasing their yards per attempt, all while demonstrating versatility in their passing abilities.



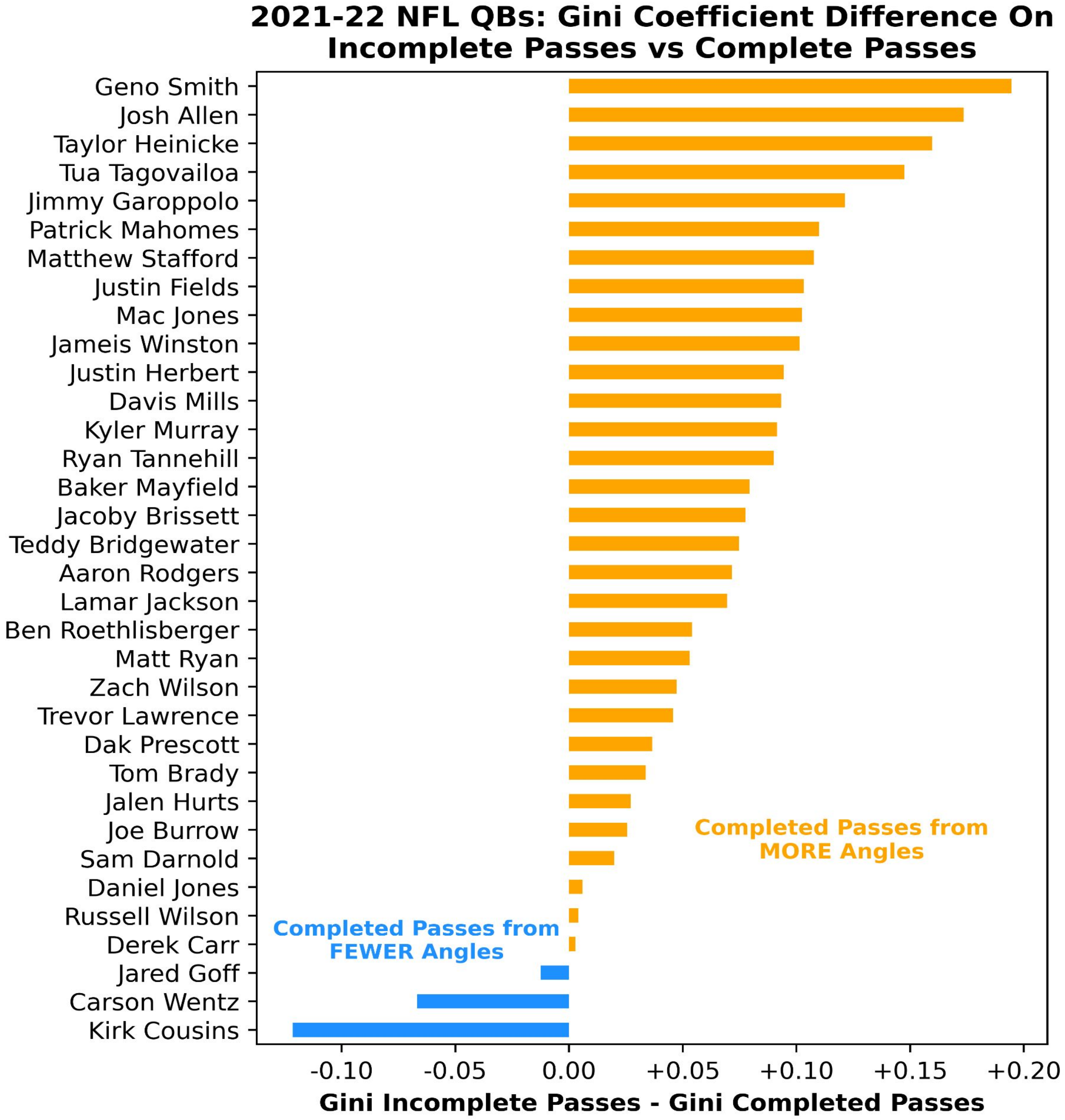
Incomplete Passes vs Completed Passes

[gini_analysis.ipynb]

Passing Profiles - Multiple Angles of Attack

A data point we wanted to investigate was how the variety of a quarterback's passing angles shifted on incomplete passes they threw versus the ones they completed successfully. Taking the difference between a QB's Gini coefficient on their incomplete passes and their Gini coefficient for complete passes gives us a snapshot of how balanced their passing portfolio is on completed passes. Interestingly, most all quarterbacks completed passes from a wider variety of angles than on their incomplete passes, with only three quarterbacks attempted a wider variety of angles on their incomplete passes. This could suggest that the angles in which the quarterbacks struggle is much more narrow than those they excel in, which makes logical sense given these are potentially the best 30 or so quarterbacks in the world.

Some of the players towards the top of the chart match what fans would expect from a metric measuring how many different types of passes a quarterback completes, as Patrick Mahomes is renowned as an improvisational passer. Overall, we can take away from the chart that the vast majority of QBs in our sample had a more balanced distribution of passing angles in their completions versus their incompletions.



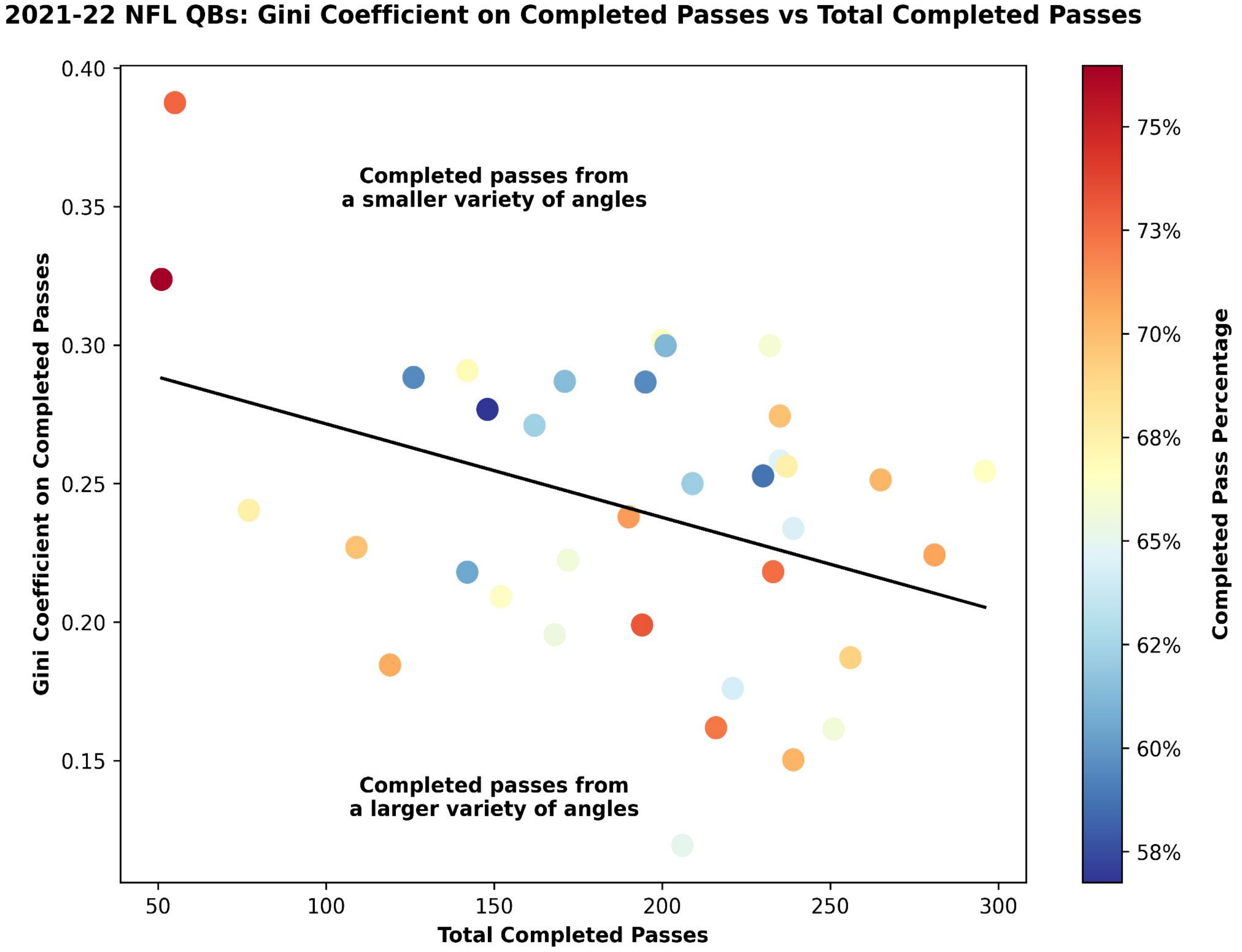
Completion Versatility

[gini_analysis.ipynb]

Benefits of Adaptability

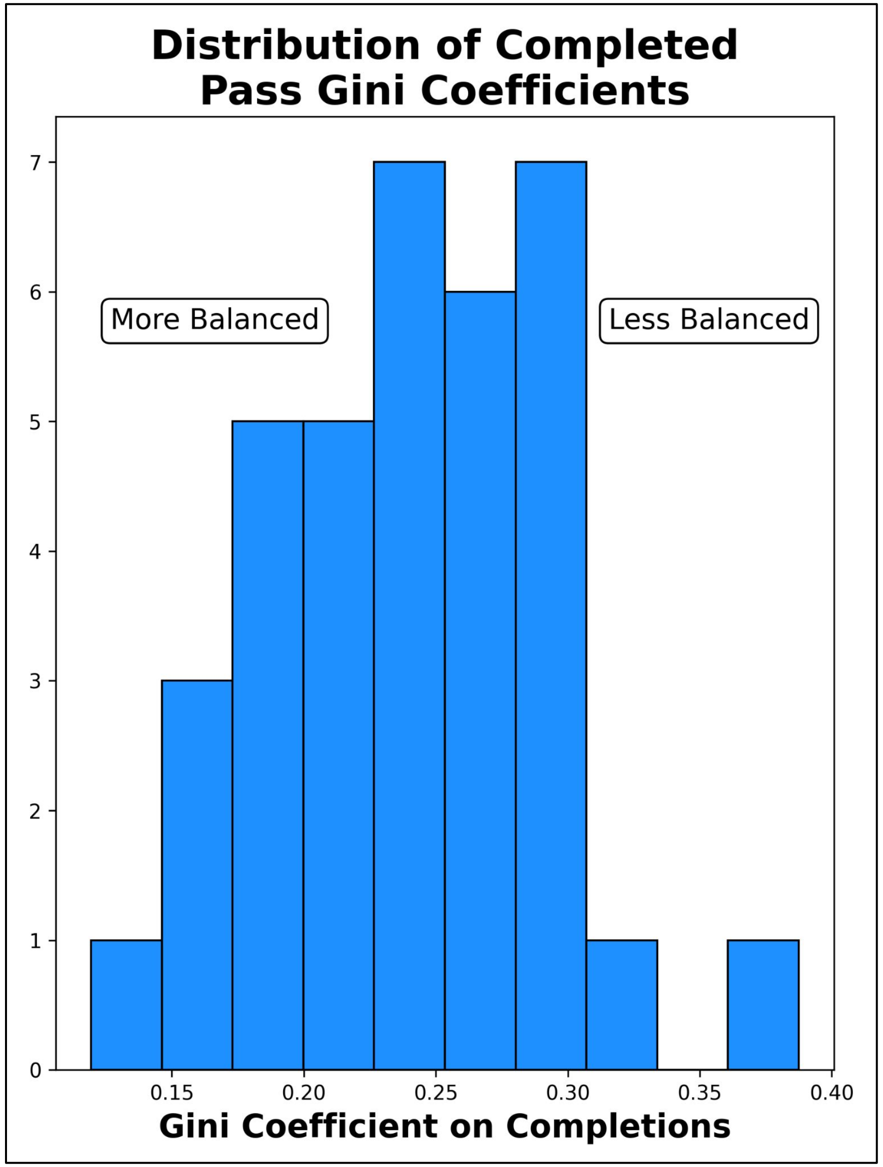
To take a glance at how common NFL stats like completion percentage and total completed passes interact with a quarterback's Gini coefficient on said completed passes, we plotted the two and created a line of best fit. One takeaway that jumps out right away is the general trend of a higher number of completed passes leading to a smaller Gini score. This suggests that more proficient QBs on average have more malleable methods of completing passes than the rest of the pack.

A more complex insight is how among the quarterbacks with a sample of 200+ completed passes, all of players with a completion rate above 73% are below the line of best fit. This furthers the idea that being able to sling the football from a healthy balance of different angles can be a factor in overall quarterback success. It would be interesting to re-create this graphic with multiple seasons of data to see if the trends hold up over a larger sample. It may be the case that the best quarterbacks are inherently more versatile, so those with higher completion marks are predisposed to having a lower Gini coefficient.



Gini Coefficient Wrap-Up

[gini_analysis.ipynb]



Most Balanced NFL QBs

Players with lowest Gini coefficients on completed passing angles
2021-22 Regular Season Through Week 8
A Gini coeff. of 0 denotes perfect balance between angles

Player	Completed Passes	Gini Coefficient
Ben Roethlisberger	206	0.12
Derek Carr	239	0.15
Justin Herbert	251	0.16
Joe Burrow	216	0.16
Josh Allen	221	0.18
Tua Tagovailoa	119	0.18
Mac Jones	256	0.19
Aaron Rodgers	168	0.20
Dak Prescott	194	0.20
Jacoby Brissett	152	0.21

Least Balanced NFL QBs

Players with highest Gini coefficients on completed passing angles
2021-22 Regular Season Through Week 8
A Gini coeff. of 0 denotes perfect balance between angles

Player	Completed Passes	Gini Coefficient
Lamar Jackson	200	0.30
Patrick Mahomes	232	0.30
Jalen Hurts	201	0.30
Jimmy Garoppolo	142	0.29
Jameis Winston	126	0.29
Sam Darnold	171	0.29
Carson Wentz	195	0.29
Zach Wilson	148	0.28
Ryan Tannehill	235	0.27
Baker Mayfield	162	0.27

Summary

Takeaways

Overall, our team was able to find several insights that helped answer our overall question of who the most versatile quarterbacks are in the NFL. First, we concluded that QBs on average are more balanced than unbalanced when it comes to their passing repertoire. All QBs with at least 25 completed passes had a Gini coefficient that was under 0.5 as shown by the histogram and scatter plot on previous slides.

Second, we were able to identify the most balanced ("versatile") quarterbacks with the leaderboard medium using Gini coefficients as a proxy. The findings from the leaderboard slightly differed from the diverging bar findings, suggesting that there is more nuance to a quarterback's success than how well distributed their passing portfolio is. External factors such as roster composition and coaching style heavily influence a QB's holistic on-field performance, so it is possible that some players have better chances of completing passes from angles.

Future Studies and Limitations

Replicating our analysis with additional seasons of data would be a logical follow-up plan to investigate if the trends we discovered hold up over a larger sample and period of time. It is very possible that what we found only holds true for the eight weeks of data we had access to, which is merely half of a single NFL season. Future work could also make use of the speed and acceleration features of the dataset, as we hoped to dig into how increased quarterback speed can affect completion percentages if we had enough time. Deeper analyses could also attempt to incorporate the talent of a quarterback's offensive linemen at protecting the quarterback, as the time that a QB has to make a decision is a key factor on each play.

The sample of passes we could derive receivers for was also limited, so it could prove useful to join an official dataset detailing the intended receiver on each play. This dataset may not be readily available to the public however, and the tracking data is almost certainly not available in full to non-team personnel.

References

References

- https://thetouchdown.co.uk/christian-mccaffrey-a-study-in-versatility
- https://www.kaggle.com/competitions/nfl-big-data-bowl-2023/data
- https://github.com/cooperdff/nfl_data_py
- https://www.sonofacorner.com/whos-attempting-the-most-crosses-in-the-premier-league/