Defining Greatness*

Mahomes' Path to Immortality vs. Brady's Unmatched Legacy

Alexander Guarasci

December 1, 2024

This paper compares National Football League (NFL) quarterbacks Tom Brady and Patrick Mahomes, focusing on their statistical performances. Analyzing their respective career data, this paper uses predictive modelling to forecast Patrick Mahomes' future. It finds that while both have impressive career statistics, ultimately, playoff victories, particularly Super Bowl wins, serve as a more definitive measure of greatness. By examining the performance of these players, the paper highlights how the NFL's dynamic context impacts player evaluation and legacy assessment.

1 Introduction

What defines greatness in professional sports? The question has captivated fans, analysts, and historians alike, sparking countless debates that go beyond numbers and records. In football, this discussion often converges on the legacies of two extraordinary quarterbacks: Tom Brady and Patrick Mahomes. Brady's career, spanning more than two decades, includes seven Super Bowl victories and an unmatched reputation for clutch performances. Mahomes, on the other hand, has quickly ascended as the face of a new generation, dazzling fans with his athleticism, creativity, and championship pedigree. This paper investigates the evolving narrative of greatness in the NFL through a detailed comparison of these two legends, exploring what their careers reveal about the broader concept of sports excellence.

While statistical achievements and accolades are important, they alone cannot encapsulate the essence of greatness. Both Brady and Mahomes boast extraordinary numbers, yet their legacies diverge in ways that reflect deeper, qualitative factors such as resilience, leadership, and the ability to deliver in high-pressure moments.

This paper ultimately aims to answer the question, can Patrick Mahomes usurp Tom Brady as the greatest football player of all time.

*Code and data are available at: https://github.com/AlexanderG123/nfl

To answer this question, this analysis draws on statistical data, examining past performance, and broader career trajectories. It evaluates what each quarterback has achieved to date and aims to project Patrick Mahomes' future.

The findings reveal that Brady's legacy is unequaled in its longevity and sheer accomplishment. While Mahomes' career trajectory suggests he does have the potential to surpass Brady, it is highly unlikely that he does from a statistical standpoint. These insights are significant not only for appreciating their contributions to football but also for understanding the broader implications of what society values in its icons.

The paper is structured as follows: Section 2 provides an explanation of the data used in this paper, what it shows and where it comes from. Section 3 discusses the model used to predict the future of Patrick Mahomes' career. Section 4 shows the results of the model. Section 5 discusses the implication of the model as well as provides more insight into the two players' legacies.

2 Data

2.1 Overview

The data used in this analysis originates from nflverse (Carl et al. 2023), a collection of packages tailored for NFL data. These packages aggregate statistics, player information and game results spanning several decades. For this paper, the focus was narrowed to quarterback (QB) performance data, specifically for Tom Brady and Patrick Mahomes.

The analysis was conducted in R (R Core Team 2022), leveraging tools from the Tidyverse (Wickham et al. 2019), Caret (Kuhn and Max 2008), Arrow (Richardson et al. 2024), Knitr (Xie 2014), and Testthat (Wickham 2011) packages. The code was styled using (Müller and Walthert 2024) The primary dataset included QB-specific data from every NFL game played since 2000, encompassing over 55 columns detailing various player performance metrics. Additional data sources were used for ESPN's Quarterback Rating (QBR), which encapsulates QB efficiency across multiple dimensions of gameplay.

The analysis focuses on comparing the careers of two of the greatest football players of all time: Tom Brady and Patrick Mahomes. Both players' careers are completely represented in the dataset, allowing for an in-depth comparison of Mahomes' trajectory against Brady's completed career. To achieve this, the dataset was filtered to include only these two players, with irrelevant data removed. For example, receiving yards were excluded since quarterbacks rarely make receptions. The refined dataset highlights key metrics such as completions, attempts, passing yards, interceptions, and a variety of quarterback rushing stats, which have become increasingly relevant in modern NFL play.

Additionally, a second dataset containing ESPN's Quarterback Rating (QBR) metric was incorporated to analyze playoff performance across their careers. This dataset, also filtered

to focus on Brady and Mahomes, provided season and postseason averages, though it lacked the granularity of the primary dataset. It served primarily during the exploratory phase to facilitate higher-level insights into what defines greatness in professional football.

2.2 Why This Dataset?

This dataset was selected for its granularity and scope, capturing raw metrics for NFL quarterbacks. While alternative datasets could have been used, nflverse was chosen for its open accessibility, ease of integration with R, and the breadth of historical data. The inclusion of QBR data, although less granular, provided a complimentary high-level view of player performance in both regular season and playoff contexts.

2.3 Measurement

The process of translating on-field football events into measurable data is fraught with complexity and subjectivity. A key example is the metric of completions, which records whether or not a quarterback's pass was caught. While this appears straightforward, the determination of a catch versus an incompletion is not always clear-cut. Referee rulings can occasionally be disputed, and the decision often hinges on nuanced interpretations of control, position, and timing. As a result, the recorded data reflects not just the physical outcome of the play but also the subjective judgment of officials, introducing an inherent layer of uncertainty into the dataset.

Another important element of football data collection lies in the placement of the ball after a play. Typically, a quarterback throws the ball, it is caught by a receiver, and the play concludes when the receiver either goes out of bounds or is tackled. At this point, referees must estimate the ball's precise location to mark the spot of forward progress or the boundary exit. While this measurement is essential for maintaining game flow and accuracy, it relies heavily on visual judgment and is thus more art than science. Although teams may challenge the spot in pivotal situations, most ball placements remain subjective, resulting in potential variability in the dataset.

These nuances highlight the challenges inherent in translating the dynamic, real-time events of a football game into static data entries. The dataset used in this analysis reflects not only the physical phenomena observed on the field but also the interpretive decisions made by officials, which can influence measures like completions and ball placement. This underscores the importance of considering the limitations and subjectivity embedded within the data when drawing conclusions about player performance and game outcomes.

2.4 Variables of Interest

Most of the variables used in this analysis are relatively straight forward. Passing yards is the vertical distance a passing play gains. For example, if a quarterback throws a pass and the receiver catches 10 yards down the field and runs for another 10 yards before being tackled or going out of bounds, that will count as 20 passing yards for the Quarterback. Rushing yards are the number of yards a player runs with the ball past the line of scrimmage - if the ball is caught as a pass, it does not count towards rushing yards. Passing touchdowns are simply a passing play that turns into a touchdown and rushing touchdowns are a rush that turns into a touchdown. Interceptions are passes made by a quarterback and caught by the other team. Carries are how many rush attempts the QB has.

The most complicated variable used in the analysis is ESPN QBR. This is a stat that aims explain the totality of how a quarterback is playing. It is a scale from 0-100 with 50 being average. ESPN says "it incorporates all of a quarterback's contributions to winning, including how he impacts the game on passes, rushes, turnovers and penalties. Also, since QBR is built from the play level, it accounts for a team's level of success or failure on every play to provide the proper context and then allocates credit to the quarterback and his teammate to produce a clearer measure of quarterback efficiency" (Katz and Burke 2016). Overall, it aims to give an overview of how well a quarterback played.

2.5 Data Visualization

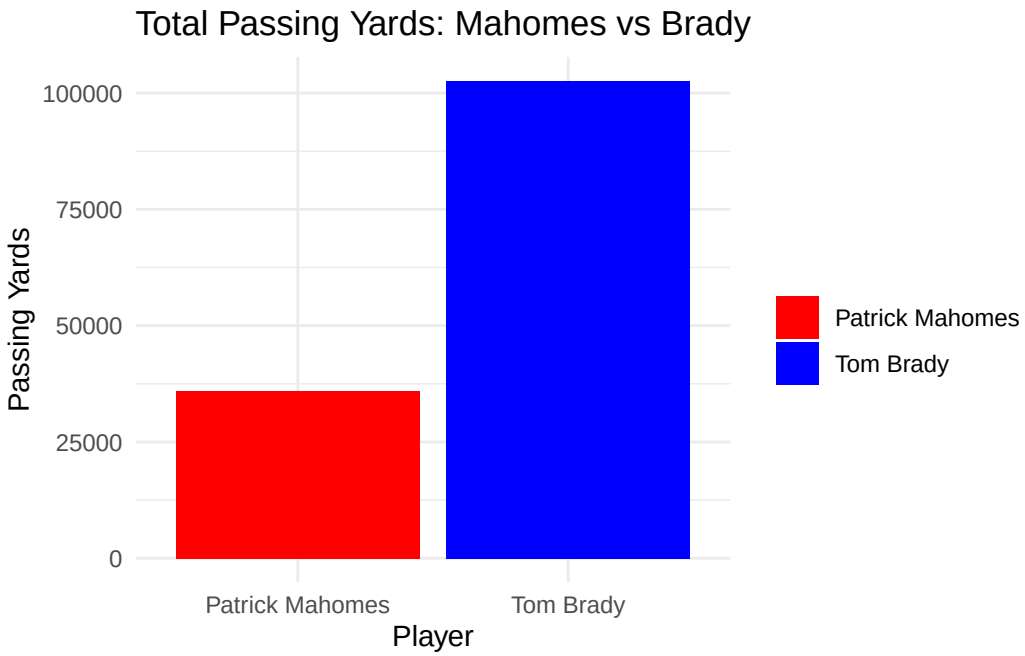

Figure 1: Total Passing Yards for Mahomes vs Brady

Figure 1 provides a comprehensive overview of the total passing yards achieved by Tom Brady and Patrick Mahomes. Tom Brady, with 102,614 total passing yards (including playoffs), significantly outpaces Patrick Mahomes, who has 35,963 yards as of November 21, 2024. This disparity is largely due to Brady's 381 career games compared to Mahomes' 124 games. What stands out is Brady's ability to maintain high passing numbers well into the later stages of his career, demonstrating longevity and consistency. On the other hand, Mahomes' current total is remarkable given the shorter duration of his career, highlighting his rapid ascent and dominance in the modern NFL.

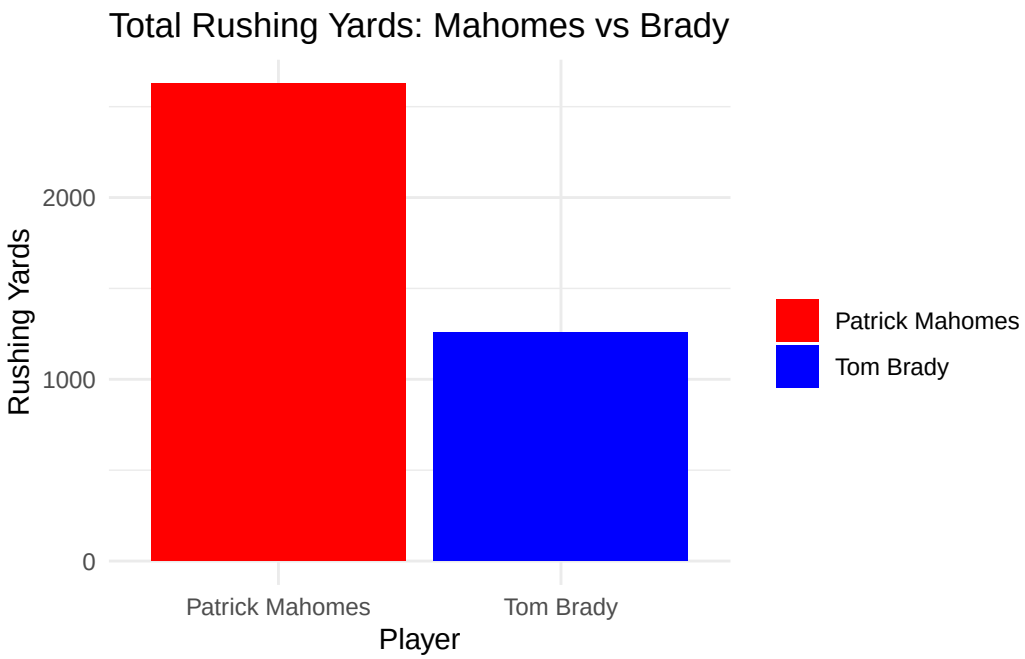

Figure 2: Total Rushing Yards for Mahomes vs Brady

Figure 2 shows an interesting contrast in rushing yards between the two quarterbacks. Despite playing over three times as many games as Mahomes, Tom Brady falls significantly short, with Mahomes amassing more than twice as many rushing yards. This reflects a broader evolution in the NFL, where modern quarterbacks are often required to be dual threats, capable of both passing and running effectively. Players like Mahomes exemplify this shift, adapting to the athletic demands of today's game, unlike Brady, who represents the traditional pocket-passing archetype of previous eras.

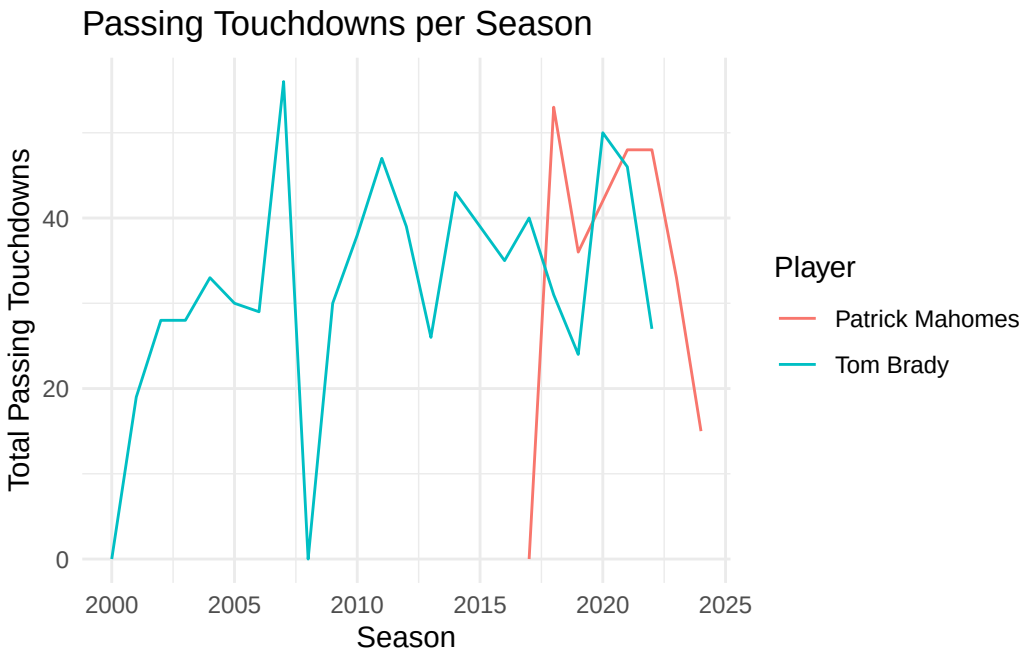

Figure 3: Passing Touchdowns per Season for Mahomes and Brady

Figure 3 illustrates the passing touchdowns per season for both players. Brady's peak in 2007 with the New England Patriots—arguably the most dominant offense in NFL history—stands out prominently. Mahomes' first season as a starter in 2018 is also noteworthy, with 53 touchdowns and just 12 interceptions. This extraordinary season set the stage for Mahomes' rise as one of the game's best quarterbacks, though it ultimately ended in a playoff loss to Brady. The graph underscores how both players have reached incredible heights in their respective careers.

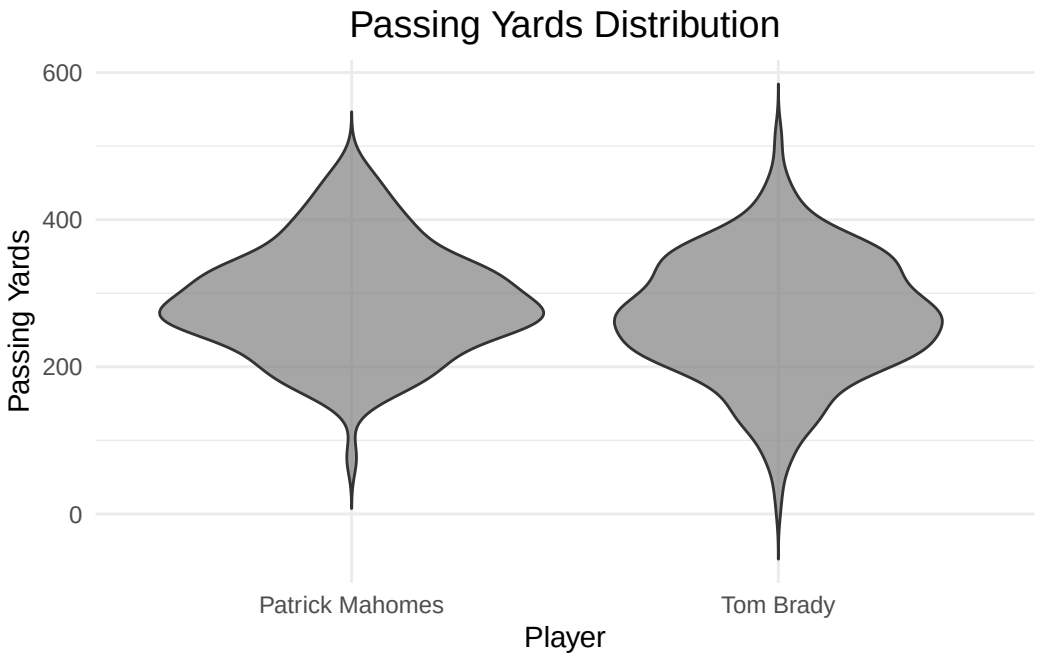

Figure 4: Passing Yard Distribution for Mahomes and Brady

Figure 4 examines the distribution of passing yards per game for each player. While both quarterbacks exhibit similar averages, Brady's larger sample size, due to his longer career, shows higher highs and lower lows. This graph captures the consistency of both players but also highlights the significance of longevity with respect to great and poor performances. Mahomes' distribution reflects his efficiency and performance within a shorter time frame.

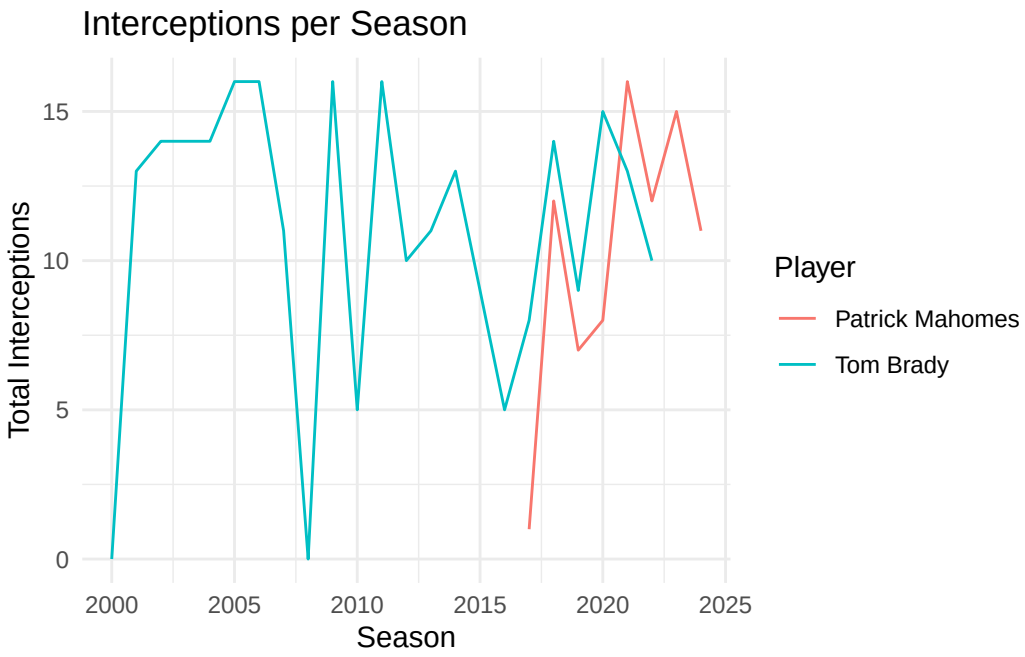

Figure 5: Interceptions per Season for Mahomes and Brady

Figure 5 highlights the number of interceptions thrown by each player per season. Interceptions are significant errors that often stem from poor decision-making or miscommunication. This graph allows us to see the players' decision-making abilities and their capacity to minimize costly turnovers.

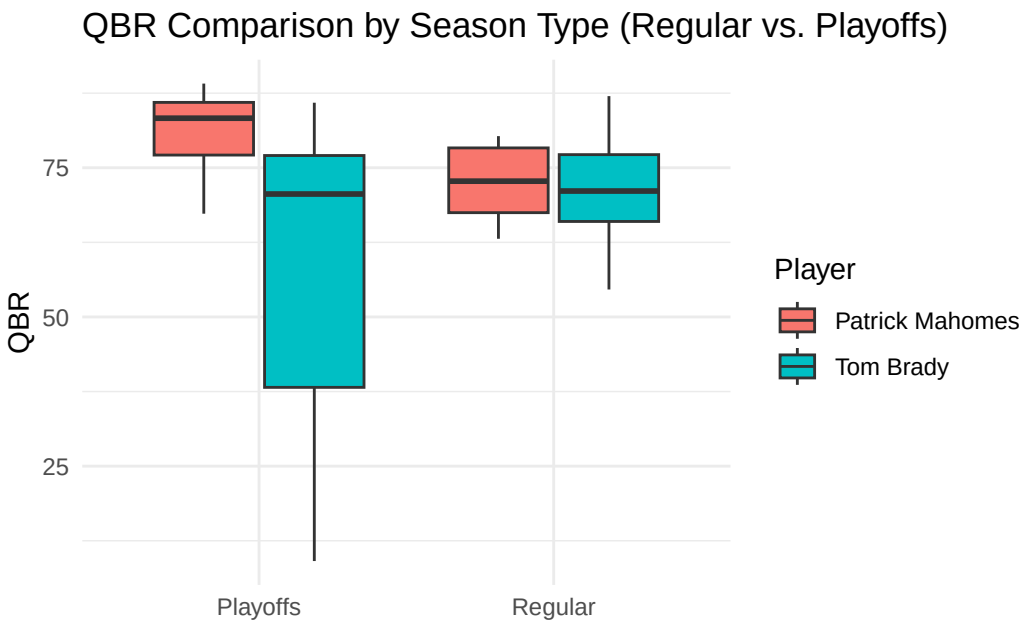

Figure 6: QBR Comparison by Season Type (Regular vs. Playoffs) for Brady and Mahomes

Although the previous graphs were useful in getting a picture of the two players' respective careers, what makes a player great in any sport is how they perform in a clutch situation. In other words, big players make big plays. Getting to the playoffs in the NFL is already very difficult, but when discussing greatness, winning is a prerequisite. Figure 6 showcases ESPN's Total Quarterback Rating (QBR) for each player, separated by regular season and playoffs. While both players demonstrate excellence in the regular season, Mahomes' significantly higher QBR in the playoffs is particularly striking. This trend underscores Mahomes' ability to elevate his game in high-pressure situations, a hallmark of greatness. Conversely, Brady's playoff QBR, though slightly lower than his regular season QBR, reflects his consistent ability to perform under postseason pressure throughout his career. This comparison underscores the different ways each player excels in their pursuit of greatness.

3 Model

3.1 Overview

The goal of this analysis is to predict Patrick Mahomes' lifetime career statistics over 200 additional games using a regression model. This model uses data from Tom Brady's career as

a proxy for player performance and is designed to forecast Mahomes' future performance across key metrics, such as completions, attempts, passing yards, rushing yards, and touchdowns.

3.2 Data

The dataset used in this analysis is a combination of game-level statistics for two players: Tom Brady and Patrick Mahomes. The data includes multiple variables that track each player's performance across different games. The key features used in this model are:

- **completions**: The number of successful passes thrown.
- **attempts**: The total number of pass attempts.
- **passing yards**: The total yards gained from passing attempts.
- **passing tds**: The total number of passing touchdowns.
- **interceptions**: The total number of interceptions thrown.
- **carries**: The number of rushing attempts.
- **rushing yards**: The total rushing yards gained.
- **rushing tds**: The total number of rushing touchdowns.

The model used in this analysis is a multiple linear regression model (Equation 1). We choose linear regression because it is a simple yet effective approach for modeling continuous outcomes and provides easy interpretability.

3.2.1 Equation for the Model

The model predicts the following variables for Patrick Mahomes:

- (completions, attempts, passing yards, passing tds, interceptions, carries, rushing yards, rushing tds)

The general form of the multiple linear regression model is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

Where:

- Y represents the outcome variable (e.g., completions, attempts, passing yards, etc.).
- $X_1, X_2, \ldots, X_n$ are the predictor variables (game statistics such as completions, attempts, etc.).
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients for each predictor.
- ϵ is the error term.

3.2.2 Training and Validation

We trained the model using data from Tom Brady's career, specifically using the following steps:

1. **Data Preparation**: We selected relevant features and ensured there were no missing values in the training dataset.
2. **Model Creation**: A multiple linear regression model was created using the `lm()` function in R. The model predicts the selected statistics for each game.
3. **Training/Test Split**: The dataset was split into training (80%) and testing (20%) sets to evaluate model performance.
4. **Prediction**: The trained model was used to predict Mahomes' statistics for the next 200 games.

We also calculated the Root Mean Squared Error (RMSE) for each predicted variable in the test set as an evaluation metric. RMSE (Equation 2) is a commonly used metric to measure the accuracy of regression models, and it is calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{2}$$

Where:

- y_i is the actual value of the statistic.
- $\hat{y}_i$ is the predicted value.
- n is the number of test samples.

The RMSE values for each variable in the test set are as follows:

completions: 6.01, attempts: 9.57, passing yards: 75.90, passing touchdowns: 1.28, interceptions: 0.80, carries: 1.72, rushing yards: 4.70 and rushing touchdowns: 0.15

These RMSE values indicate that the model has some level of predictive accuracy but also considerable error in certain variables, such as passing yards. The RMSE is higher than expected, which can be attributed to the limited dataset size used in this model. The dataset contains a relatively small number of games, which leads to higher variability and uncertainty in the predictions.

3.2.3 Alternative Models Considered

1. **Decision Trees**:
    - Strengths: Captures non-linear relationships.
    - Weaknesses: Tends to overfit without pruning; less interpretable than linear regression, and still had high error.
2. **Bayesian Regression**:
    - Strengths: Allows incorporation of priors, producing probabilistic predictions.
    - Weaknesses: Increased complexity and computational requirements and also had high error.
3. **Final Choice**:
    - Linear regression was chosen for its balance of simplicity, interpretability, and performance.

3.3 Assumptions and Limitations

The model makes several key assumptions that are important for interpreting the predictions:

1. **Similarity of Player Performance**: The model assumes that Patrick Mahomes' future performance will be similar to Tom Brady's past performance, using Brady's career as a proxy. This assumption may not hold as Mahomes' playing style and team dynamics significantly differ from Brady's.
2. **Linear Relationships**: The model assumes that the relationships between the predictors (game statistics such as completions, passing yards, etc.) and the outcomes are linear. This may not always be the case in practice, as player performance could be influenced by more complex, non-linear interactions.
3. **Constant Game Conditions**: The model assumes that game conditions, including factors such as opposing defenses and weather conditions, remain relatively consistent over time. However, these factors could vary and influence player performance in ways not captured by the model.
4. **Prediction for 200 Games**: The model assumes that Mahomes' performance will follow a similar trajectory over the next 200 games as it has in the past. This assumption may not account for potential career milestones, injuries, or other events that could disrupt this pattern.

5. **Uniform Dataset**: The model assumes that the dataset of Tom Brady's career is representative of typical game conditions and does not contain significant biases or outliers that would skew the predictions.

Ultimately, the model may be appropriate in scenarios where general trends in quarterback performance are needed for broad projections, such as for media analysis or high-level discussions of player comparisons. It can offer insights into how Mahomes might perform in a typical game based on historical data from a similar elite quarterback. This can be useful for estimating the overall trajectory of Mahomes' career and projecting his future career stats across a large sample of games.

However, the model may not be appropriate for precise, game-by-game predictions, especially when specific game conditions or player-specific factors (such as injury status or team dynamics) are likely to vary. For example, using this model to predict Mahomes' performance in a critical playoff game or during a season affected by significant roster changes could lead to inaccurate or misleading conclusions. Additionally, the linear nature of the model means it may fail to account for more complex interactions that influence a quarterback's performance in real-world scenarios.

In summary, while the model provides a reasonable estimation of Mahomes' future statistics based on Brady's historical data, it should be used cautiously in contexts where precision is critical or where Mahomes' unique playing style and career circumstances may diverge from Brady's past performance.

3.4 Conclusion

This model provides a reasonable estimate of Patrick Mahomes' lifetime career statistics over 200 additional games based on Tom Brady's historical data. Although the RMSE values suggest room for improvement, the model serves as a useful starting point for projecting Mahomes' future performance. Future work could include expanding the dataset, exploring non-linear modeling techniques, and accounting for player-specific factors that might differ between Brady and Mahomes.

4 Results

Table 1 compares Tom Brady's career statistics with Patrick Mahomes' projected lifetime statistics. Tom Brady, with 8,954 completions and 13,973 attempts, has significantly more career completions and attempts than Mahomes, who is projected to finish with 7,763 completions and 11,925 attempts. Brady's total passing yards of 102,616 far exceed Mahomes' projected 89,160, highlighting Brady's longevity and consistent performance. Similarly, Brady has thrown 738 passing touchdowns, whereas Mahomes is projected to finish with 654. Brady also has more interceptions (252 compared to Mahomes' 216). These differences underscore

Brady's long career and high volume of passing, which has contributed to his unrivaled career achievements.

Table 1: Career Totals: Brady vs Mahomes

Player	Comp	Att	Pass Yds	Pass TDs	INTs	Carries	Rush Yds	Rush TDs
Tom Brady	8954	13973	102616	738	252	807	1256	35
Patrick Mahomes	7763	11925	89160	654	216	928	3348	40

When comparing rushing statistics, Mahomes stands out as the more mobile quarterback. While Brady's rushing stats are modest, with 1,256 rushing yards and 35 rushing touchdowns, Mahomes is projected to amass 3,348 rushing yards and 40 rushing touchdowns, showing a clear advantage in rushing ability. This emphasizes Mahomes' dual-threat capability, which contributes to his versatility as a quarterback.

Despite Mahomes' impressive projected stats, Brady's career longevity and overall impact on the game have allowed him to achieve these remarkable milestones. Mahomes will need to maintain his performance over a long period to catch up to Brady's cumulative totals. Brady's consistency and durability throughout his career, as seen in his passing and rushing stats, have firmly established him as one of the greatest quarterbacks in NFL history. Mahomes, on the other hand, shows great potential and is already a future first ballot Hall of Famer, but he has yet to match Brady's career volume.

Table 2, shows the per game projected averages for Mahomes and compares it with Brady's stats.

Table 2: Career Averages: Brady vs Mahomes

Player (Per Game)	Comp	Att	Pass Yds	Pass TDs	INTs	Carries	Rush Yds	Rush TDs
Tom Brady	23.50	36.67	269.33	1.94	0.66	2.12	3.30	0.09
Patrick Mahomes	23.96	36.81	275.18	2.02	0.67	2.87	10.33	0.12

The analysis of per-game statistics shows a clear edge for Patrick Mahomes over Tom Brady across nearly every metric. Mahomes completes slightly more passes per game (23.96) than Brady (23.50), despite attempting a comparable number of throws (36.81 for Mahomes vs. 36.67 for Brady). This small difference results in a higher completion percentage (defined as the number of completions divided by the number of attempts) for Mahomes at 65.10%, compared to Brady's 64.08%, indicating a marginal but noteworthy advantage in passing efficiency. Mahomes also outpaces Brady in passing yards per game, averaging 275.18 to Brady's 269.33, further emphasizing his ability to generate more offense through the air.

In terms of scoring, Mahomes demonstrates a superior passing touchdown rate, averaging 2.02 touchdowns per game compared to Brady's 1.94. Both players maintain an impressive level of ball security, with nearly identical interception rates per game (0.66 for Brady to 0.67 for Mahomes).

The most significant disparity appears in rushing statistics, where Mahomes outshines Brady by a wide margin. Mahomes averages 10.33 rushing yards per game, more than triple Brady's 3.30, and his higher number of carries per game (2.87 vs. 2.12) underscores his ability to contribute as a dual-threat quarterback. Furthermore, Mahomes' rushing touchdown rate (0.12 per game) surpasses Brady's 0.09, further illustrating his dynamic impact on the ground.

These results suggest that Mahomes not only matches but surpasses Brady's per-game contributions in every key category, showcasing his all-around ability as a quarterback. The consistent superiority in passing, rushing, and overall efficiency highlights Mahomes' potential to redefine expectations for quarterback performance in the NFL.

5 Discussion

5.1 What is done in this paper?

This paper provides a rigorous statistical comparison of Tom Brady and Patrick Mahomes to evaluate whether Mahomes could eventually surpass Brady as the greatest quarterback of all time (GOAT). Through comprehensive data analysis, the study examines key performance metrics across both players' careers, incorporating traditional and more high-level statistics, including ESPN's Total Quarterback Rating (QBR). Additionally, the paper develops a predictive model to estimate Mahomes' career trajectory if he plays 200 more games. By combining historical data, exploratory visualization, and predictive modeling, the paper offers insights into the evolving standards of quarterback greatness in the NFL.

5.2 What is something that we learn about the world?

One significant thing that we learn from this paper is the impact of career longevity on perceptions of greatness. Tom Brady's career, spanning over two decades, showcases not only

impressive cumulative statistics but also a sustained level of excellence, particularly in high-stakes post-season environments. This longevity has set an unprecedented benchmark for future players. In contrast, Patrick Mahomes, despite a much shorter career, demonstrates a rapid accumulation of achievements, emphasizing efficiency and adaptability to the modern NFL. This analysis highlights how differing career arcs can define greatness and challenges the traditional emphasis on cumulative statistics alone. The reality of the situation is, you need to play a lot of football to win seven Super Bowls, and the nature of the sport makes it really difficult to assume a player will even be in the league five years from now and even more difficult to assume that they will have any form of continued success over this period.

5.3 What is something else that we learn about the world?

Greatness in sports, especially in the context of legendary quarterbacks like Tom Brady and Patrick Mahomes, cannot be boiled down to statistics alone. While both players have put up astonishing numbers throughout their careers, attempting to compare them purely through statistics is an oversimplification of their legacies. Statistics, while important, fail to account for the intangibles—the ability to perform under pressure, the knack for leading a team to victory in clutch moments, and the sheer will to win when it matters most.

When evaluating greatness, Super Bowl victories tend to stand as the ultimate benchmark, and here lies the defining gap between Brady and Mahomes. Tom Brady has set an unprecedented standard with seven Super Bowl wins, a number that remains a daunting goal for any quarterback to reach. Mahomes, as brilliant as he has been in his career so far, still has a long way to go in this regard. With three Super Bowl wins to his name, he would need four more championships just to tie Brady's record—an extraordinary feat that would require sustained excellence over many years.

Adding to Brady's case is his direct record against Mahomes in the playoffs. The two have faced off twice in the playoffs, and on both occasions, Brady emerged victorious. The first was in the 2018 AFC Championship game, a thrilling overtime battle that ended with Brady's Patriots edging out Mahomes' Chiefs. The second, and perhaps more significant, was in Super Bowl LV, where Brady's Buccaneers decisively defeated Mahomes and the Chiefs, cementing Brady's legacy as the ultimate competitor on the biggest stage.

However, Mahomes' career is still young, and his playoff track record so far is nothing short of remarkable. Aside from the two playoff losses to Brady, Mahomes has only lost one other postseason game. In fact, he has never been eliminated from contention before reaching the AFC Championship, a testament to his dominance and consistency. His playoff record of 15-3 is astounding and reflects a level of performance that no other quarterback in history has ever matched.

This also highlights another point about Mahomes: his ability to consistently elevate his game when the stakes are highest. Winning playoff games is no small feat, yet Mahomes has made it look routine, guiding his team deep into the postseason year after year. His style of

play—combining pinpoint accuracy, athleticism, and creativity—has redefined the quarterback position and captivated fans and analysts alike.

Ultimately, while Brady's resume sets the gold standard for greatness, Mahomes is building a legacy that could rival it in time. For now, Brady's longevity, unmatched playoff success, and head-to-head victories give him the edge in the GOAT debate. But Mahomes' trajectory, both statistically and in terms of team success, suggests that he could one day close the gap. What we learn about the world through their stories is that greatness is not about one-dimensional measures like statistics—it's about how athletes respond to challenges, seize opportunities, and etch their names in history through consistent excellence and defining moments.

5.4 What are some weaknesses of what was done?

While this paper offers some insights, there are several limitations worth acknowledging. First, the predictive model relies on the assumption that Mahomes' career trajectory will follow trends observed in Brady's career, which may not account for unforeseen injuries, team dynamics, or changes in the NFL landscape. Second, the analysis predominantly focuses on individual statistics without fully considering the impact of team success, coaching, or supporting cast, all of which significantly contribute to a quarterback's achievements. Lastly, ESPN's QBR as a metric introduces subjectivity, as its proprietary formula may not fully capture the nuances of on-field performance.

5.5 What is left to learn or how should we proceed in the future?

Future research should address the limitations identified above by incorporating a broader set of variables, such as coaching quality, team composition, and strength of schedule, to provide a more holistic assessment of quarterback performance. Additionally, as Mahomes' career progresses, longitudinal studies can reevaluate his trajectory using updated data. Advances in machine learning could also enable the development of more sophisticated models to predict player performance under varying conditions. Lastly, looking into how the definition of greatness evolves across sports can provide context for comparing athletes from different eras, enhancing the broader conversation about legacy and excellence.

Other than that, the only thing left to do is wait and see. The Mahomes led Chiefs are 9-1 as this is being written. They are coming off back-to-back Super Bowl victories, they are in the midst of a dynasty and all that is left to do is witness greatness unfold.

6 Appendix: Observational Data and Measurement Challenges in NFL Performance Metrics

The analysis of NFL quarterback performance relies entirely on observational data, as the events captured are recorded during live games rather than under controlled experimental conditions. While this provides an authentic reflection of player performance, it also introduces specific challenges associated with surveys, sampling, and observational methodologies. This appendix explores the implications of these challenges, the methods used to address them, and how these considerations influence the interpretation of results.

6.1 Observational Data in NFLverse

Observational data inherently captures events as they unfold in their natural context. In the case of the nflverse dataset, this includes game statistics, player-specific metrics, and play-by-play details recorded during live games. Unlike experimental data, there is no controlled randomization or intervention, meaning that any observed relationships are correlational rather than causal. For example, a quarterback's passing yards are affected not only by their skill but also by external factors such as the quality of their offensive line, the defensive strategies employed by opponents, and weather conditions.

Despite its observational nature, the nflverse dataset is highly granular, capturing over 55 variables per game, which mitigates some biases by enabling a comprehensive multivariate analysis. However, the dataset remains subject to inherent challenges, including subjective measurements and incomplete data points.

6.2 Sampling and Representativeness

The dataset used in this analysis includes quarterback-specific data from all NFL games since 2000, encompassing tens of thousands of observations. While this large sample size minimizes the impact of random noise, it does not eliminate systemic biases:

1. **Selection Bias**: The dataset excludes quarterbacks who did not play enough games to generate meaningful statistics. As a result, it inherently overrepresents high-performing players, such as Tom Brady and Patrick Mahomes, while underrepresenting backups or quarterbacks with shorter careers.
2. **Era Bias**: Comparing players across eras introduces biases due to evolving rules, strategies, and player conditioning. For instance, the NFL has shifted toward more pass-heavy offenses in recent years, which inflates metrics such as passing yards and touchdowns for modern quarterbacks compared to their predecessors.

To address these biases, the analysis normalized certain statistics, showing statistical averages per game, to account for differences in season length. Additionally, Mahomes' projected career totals were included to balance the comparison with Brady's completed career.

6.3 Measurement Challenges and Subjectivity

The NFLverse dataset depends on a combination of human and technological data collection methods. Human referees make decisions on plays, introducing subjectivity into measurements such as completions and ball placement:

- **Completions**: Determining whether a pass is caught often involves split-second judgments by referees, who must assess factors like possession, control, and boundary location. Video reviews mitigate errors but do not eliminate them entirely.
- **Ball Placement**: The spotting of the ball is another area of potential variability. Referees estimate the exact location based on visual assessment, which can result in small but impactful discrepancies in rushing and passing yard statistics.

Building on these frameworks, this analysis incorporated metrics like ESPN's QBR, which contextualizes a quarterback's contributions within the broader game environment. While QBR addresses some limitations of traditional statistics, it remains an observational metric and is thus subject to the same biases and variability.

6.4 Conclusion: Strengths and Limitations of Observational Data

The nflverse dataset represents one of the most comprehensive sources of quarterback performance data available, but its observational nature necessitates careful consideration of biases, subjectivity, and measurement variability. Ultimately this analysis seeks to provide a balanced comparison of Brady and Mahomes.

While observational data cannot establish causality, it offers insights into patterns and trends, particularly when analyzed with an awareness of its limitations. Future work could enhance these findings by incorporating experimental methods, such as simulations or controlled studies, to isolate causal factors influencing quarterback performance.

References

Carl, Sebastian, Ben Baldwin, Lee Sharpe, Tan Ho, and John Edwards. 2023. *Nflverse: Easily Install and Load the 'Nflverse'*. https://CRAN.R-project.org/package=nflverse.

Katz, S., and B. Burke. 2016. "How Is Total QBR Calculated? We Explain Our Quarterback Rating." https://www.espn.com/blog/statsinfo/post/_/id/123701/how-is-total-qbr-calculated-we-explain-our-quarterback-rating.

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Müller, Kirill, and Lorenz Walthert. 2024. *Styler: Non-Invasive Pretty Printing of r Code*. https://CRAN.R-project.org/package=styler.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2011. "Testthat: Get Started with Testing." *The R Journal* 3: 5–10. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Wickham.pdf.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.