Will Nguyen

NFL Quarterback Statistical Analysis

Introduction

It is common for sports fans to debate who the best player is. And when it comes to the NFL, no debate is more widely discussed than who is the best quarterback. For the better part of this century, Tom Brady has been at the center of this discussion. With Patrick Mahomes recent rise to the top, it's no surprise he has often been compared to Brady. The motivation behind this project is to attempt to use statistics to compare their careers and performances to both each other and the other top quarterbacks each season.

This project will use Python and statistical analysis techniques on data pulled from Pro Football Reference to analyze the two quarterbacks performances since 2006 (the last 17 seasons). The following Python libraries were used:
- **pandas**: data manipulation and analysis library
- **NumPy**: universal numerical data library
- **Matplotlib**: vector graphics/plotting library
- **Seaborn**: data visualization library
- **Beautiful Soup**: HTML parsing library
- **Scikit-learn**: algorithms library
- **SciPy**: algorithms library

As Brady and Mahomes are often considered to be top quarterbacks, it is only natural to compare them to their peers; the other top quarterbacks in the NFL in any given year. It is important to keep in mind that football is a team sport, and while the quarterback position is arguably the most valuable, there may be other factors impacting performance. Injuries to the quarterback themselves and their supporting cast are also not factored in this analysis.

The main challenge when comparing quarterbacks is deciding what metric to use. Fans most often use who threw the most yards or touchdowns to support their claims. Or to argue a point, they will cite how often a given quarterback throws an interception or simply how many games they have lost (or won). For the purpose of this analysis, it will be assumed a quarterback's win rate is the most important statistic, as afterall, the true objective when competing is to win.

Research Questions
- Which quarterback statistics have the strongest correlation to winning?
- How do Tom Brady and Patrick Mahomes compare statistically to other top quarterbacks in the NFL in any given year?
- Is there evidence that Patrick Mahomes is the "next" Tom Brady?

Methodology and Framework

It is important to consider what the research questions are asking and how they will guide answering the problem. This project is about the top quarterbacks. The NFL currently consists of 32 teams and most teams will have at least one backup on their roster. This means that there are at least 32 quarterbacks who may have data that may skew the results. To account for this, the data will have to be filtered to contain what is assumed to be each team's starting quarterback.

Another factor to consider is that even though a quarterback may be the starter, that does not mean that they will all play the same number of games. Starting players may miss games due injury, illness, or suspension. Or in the case of the top players on the top teams, if a team has already guaranteed their position in the playoff standings, then they may be rested and miss a game. To account for some of these discrepancies, all of the statistics will need to be converted to a per game average as opposed to the season total.

This project will utilize the following techniques and models:
- **Correlation Analysis** via scatter plots and linear regression
- **Descriptive Statistics** via box plots
- **Time Series Analysis** via line charts
- **Test Statistics** (student t-tests)

The high-level framework of this project is as follows
1. Conduct a correlation analysis to determine which passing statistics have the strongest correlation to winning and to identify how Brady and Mahomes fare amongst their peers
2. Utilize summary statistics to compare Brady and Mahomes' performance in the metrics with the highest correlation to winning amongst their peers on a yearly basis
3. Compare Brady and Mahomes head to head for the years in which both were active for those same metrics
4. Conduct t-tests to identify if there are statistically significant differences in Brady and Mahomes performances in those same metrics

Data

The data was scraped from Pro Football Reference using Python. Pro Football Reference organizes their data by position and year. This means that in order to collect the data from the 2006 season to the 2022 season, 16 pages will have to be parsed. The Requests and BeautifulSoup libraries made this easy. The pandas library will be used to store all of the data in an easily manipulatable object (a DataFrame). The 2006 season was selected as the starting point as that was when QBR, one of the key passing metrics, began being calculated and recorded. The 2006 season was also the earliest season in which at least one of Brady or Mahomes was an active quarterback in the NFL. Otherwise the data collected would have begun when either started their NFL career.

Note: Patrick Mahomes began his career in 2018. The 2018 season was not chosen as the starting point because there would be less data to analyze. Starting with the 2006 season gives an additional 12 years of data that would have been omitted.

Because each page had to be scraped separately, this resulted in 16 tables (one for each season). To make comparing and calculations quicker, the tables were combined into one table and an additional column was added to indicate the year in which the stats were recorded.

While Pro Football Reference made collecting the data relatively easy, the data was not ready to go for analysis. Some of the data had to be converted to a different type, some rows and columns had to be removed, and some numbers had to be recalculated. Some examples include removing extra column headers, unused columns such as team, converting "70%" to "0.7", or adjusting "total touchdowns" to "touchdowns per game". As previously mentioned, the data was also filtered to only include the top 32 quarterbacks for each season in order to remove outlier data from back-ups and other non-quarterbacks being included in the passer data.
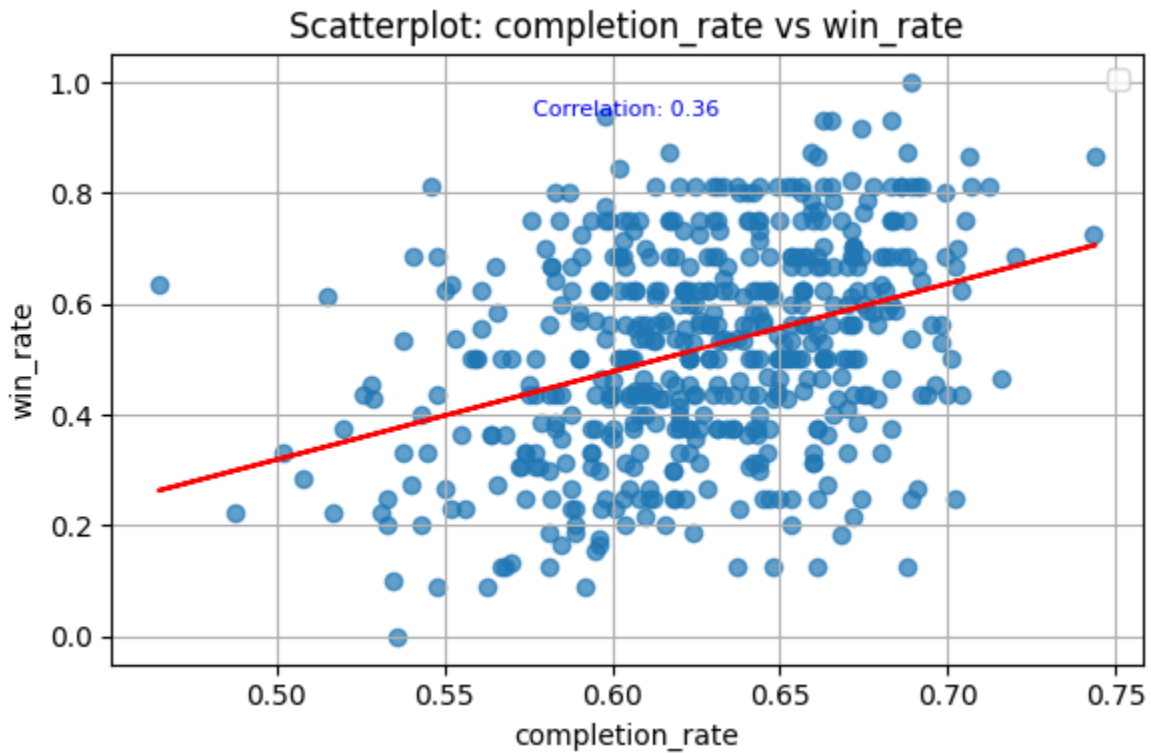
Other steps included filtering out non-quarterbacks (non-quarterbacks may occasionally make passes during trick plays), renaming columns, splitting combined stats (such as win loss record), and stripping unnecessary characters which might result in duplication errors. An example of this would be if a player made the Pro Bowl or was selected as an All-Pro. Their name would be annotated with an "*" or an "+" (ex: Tom Brady*+). If the same player made either one year, but not the next, they would be counted as two separate instances and might skew some of the calculations.
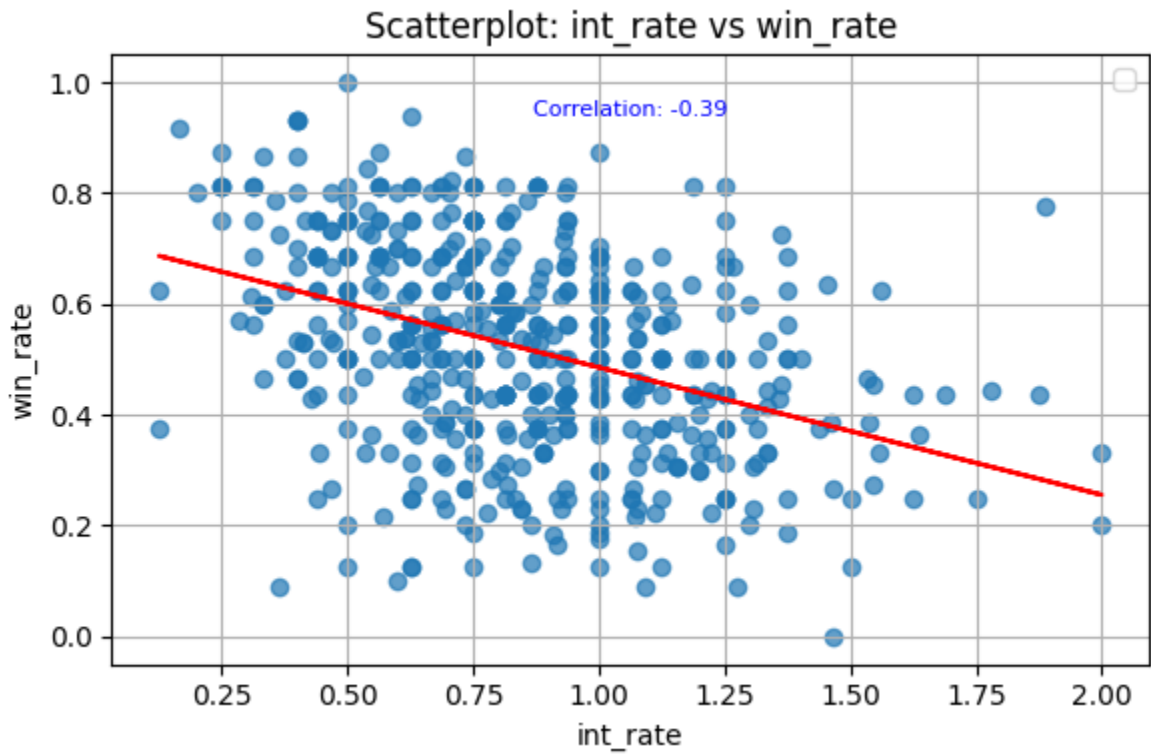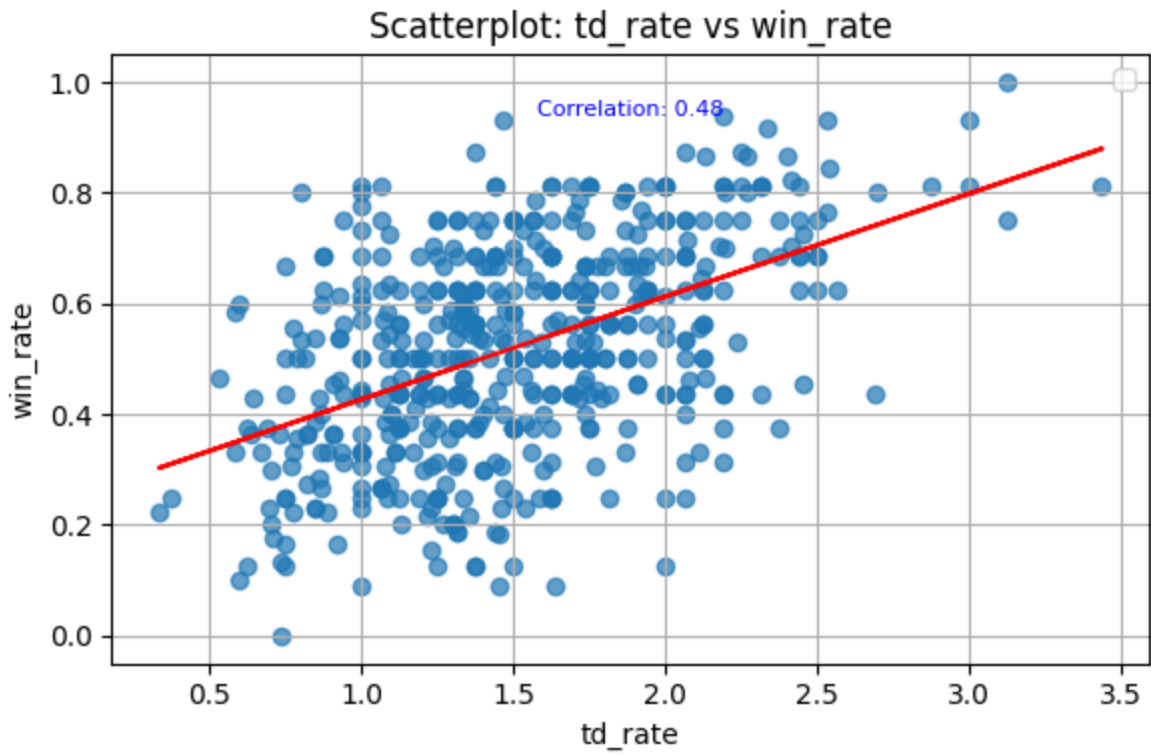
Summary of Variables
- **Player**: the player's name
- **completion_rate**: the percentage of throws caught per game
- **td_rate**: the number of passes caught for a touchdown per game
- **int_rate**: the number of passes thrown for an interception per game
- **ANY/A**: adjusted net yards per attempt; a passing metric that considers total passing yards, but rewards touchdowns and penalizes interceptions and sacks.
- **QBR**: quarterback rating; a passing metric that incorporates passes, rushes, turnovers, and penalties.
- **sack_rate**: the number of times sacked per game
- **win_rate**: the number of games won divided by games played
- **td_int**: the number of touchdowns to interceptions thrown
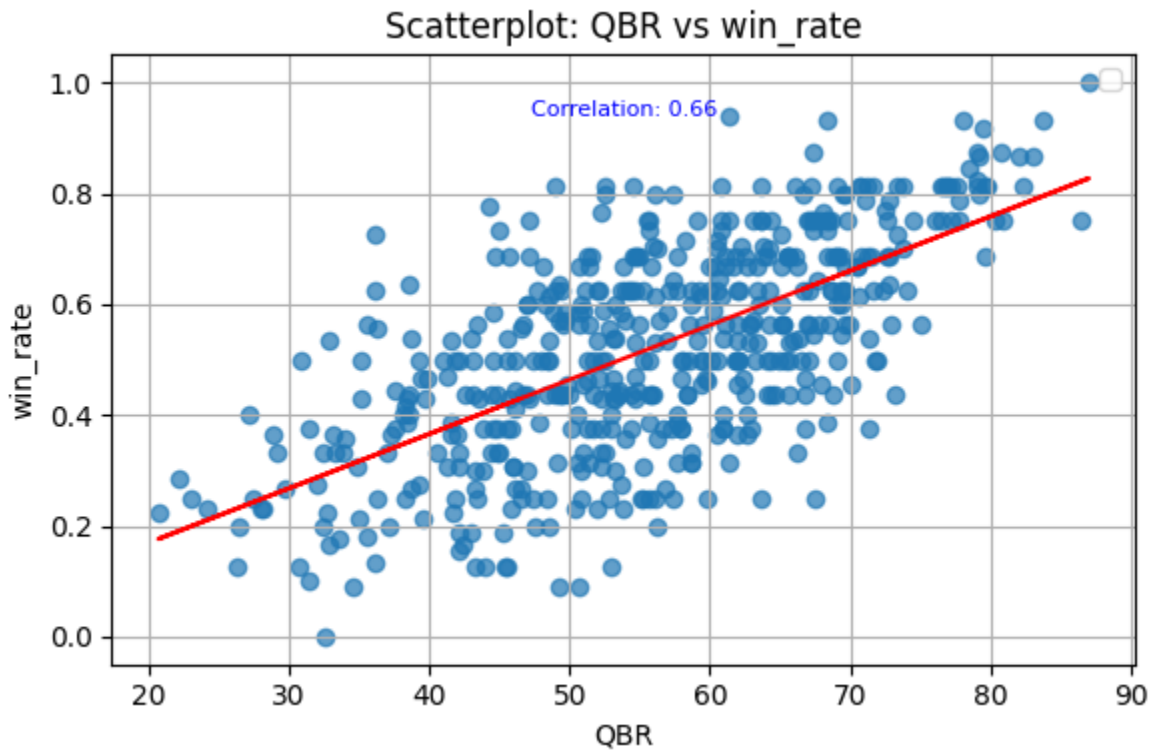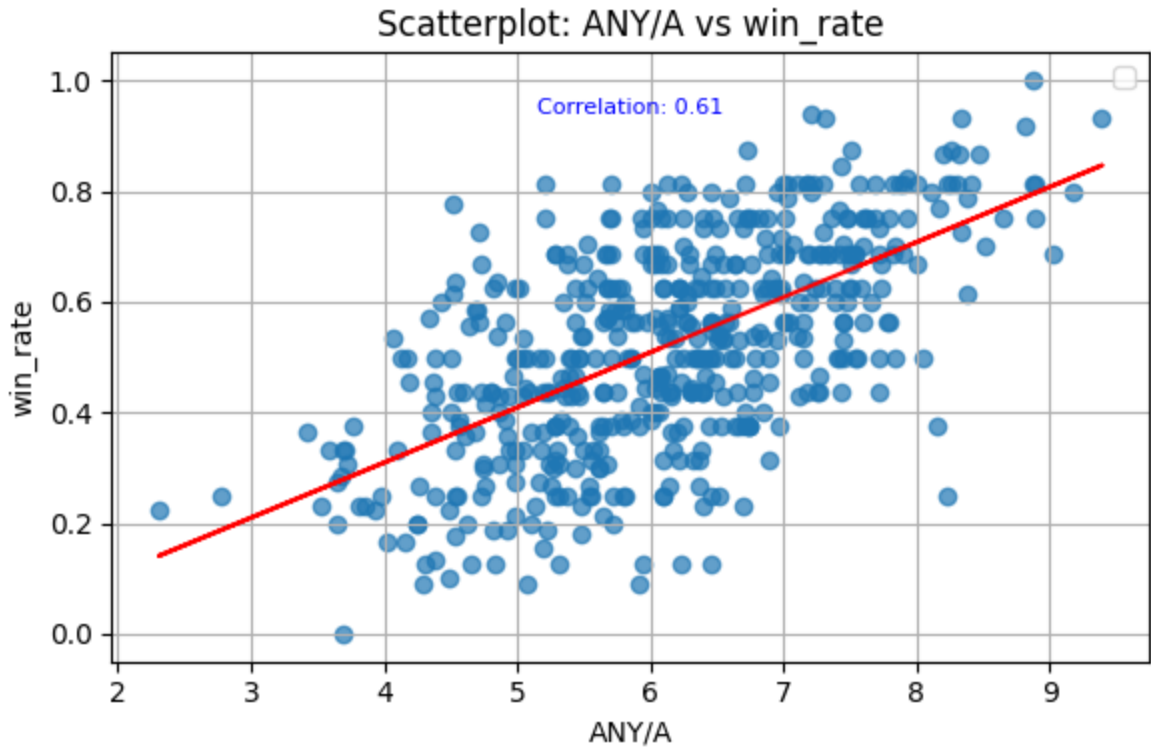- **year**: the year the statistic was recorded

Results

As previously stated, the first objective is to identify which passing metrics have the strongest correlation to winning (win_rate). To do this, each of the stats were plotted against win_rate in a scatter plot. Linear regression was used to calculate a line of best fit for each. The results are as follows.
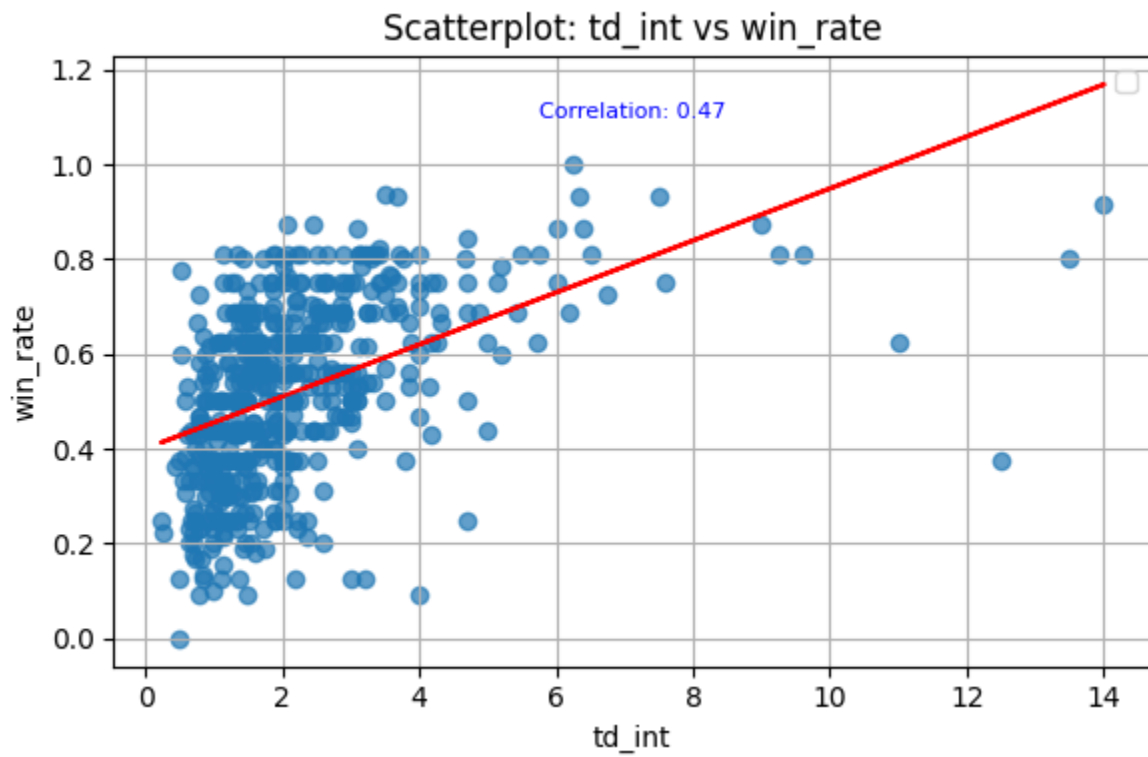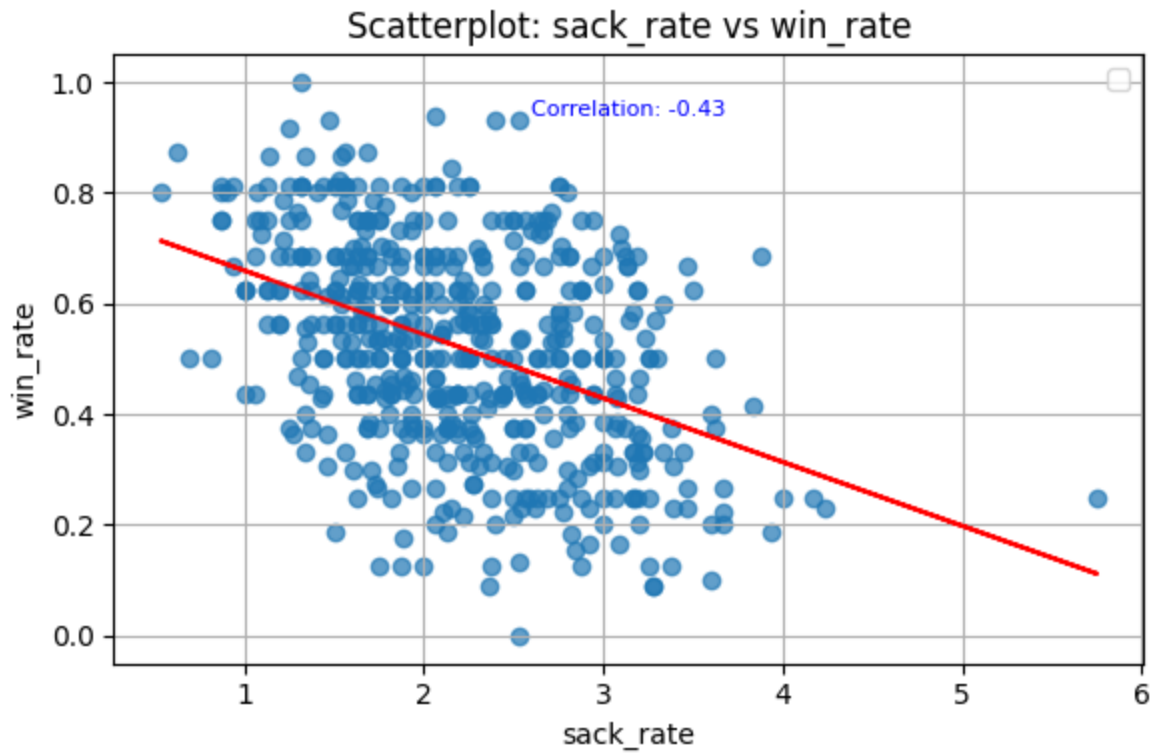

Scatterplot: completion_rate vs win_rate

Scatterplot: td_rate vs win_rate

Scatterplot: int_rate vs win_rate

Scatterplot: ANY/A vs win_rate

Correlation: 0.61

win_rate vs ANY/A

Scatterplot: QBR vs win_rate

Correlation: 0.66

win_rate vs QBR

Scatterplot: sack_rate vs win_rate

Correlation: -0.43

Scatterplot: td_int vs win_rate

Correlation: 0.47

And here are the correlation coefficients in tabular form:

Passing Metric	Correlation Coefficient
completion_rate	0.36
td_rate	0.48
int_rate	-0.39
ANY/A	**0.61**
QBR	**0.66**
sack_rate	-0.43
td_int	0.47

It is no surprise that QBR and ANY/A are the two metrics that have the highest correlations to win_rate, as the two metrics are designed to evaluate quarterback performance.

The next step is to compare Tom Brady and Patrick Mahomes in those key passing metrics to the rest of the NFL during each season. This will be done using correlation analysis and scatter plots once again, as well as descriptive statistics and boxplots.
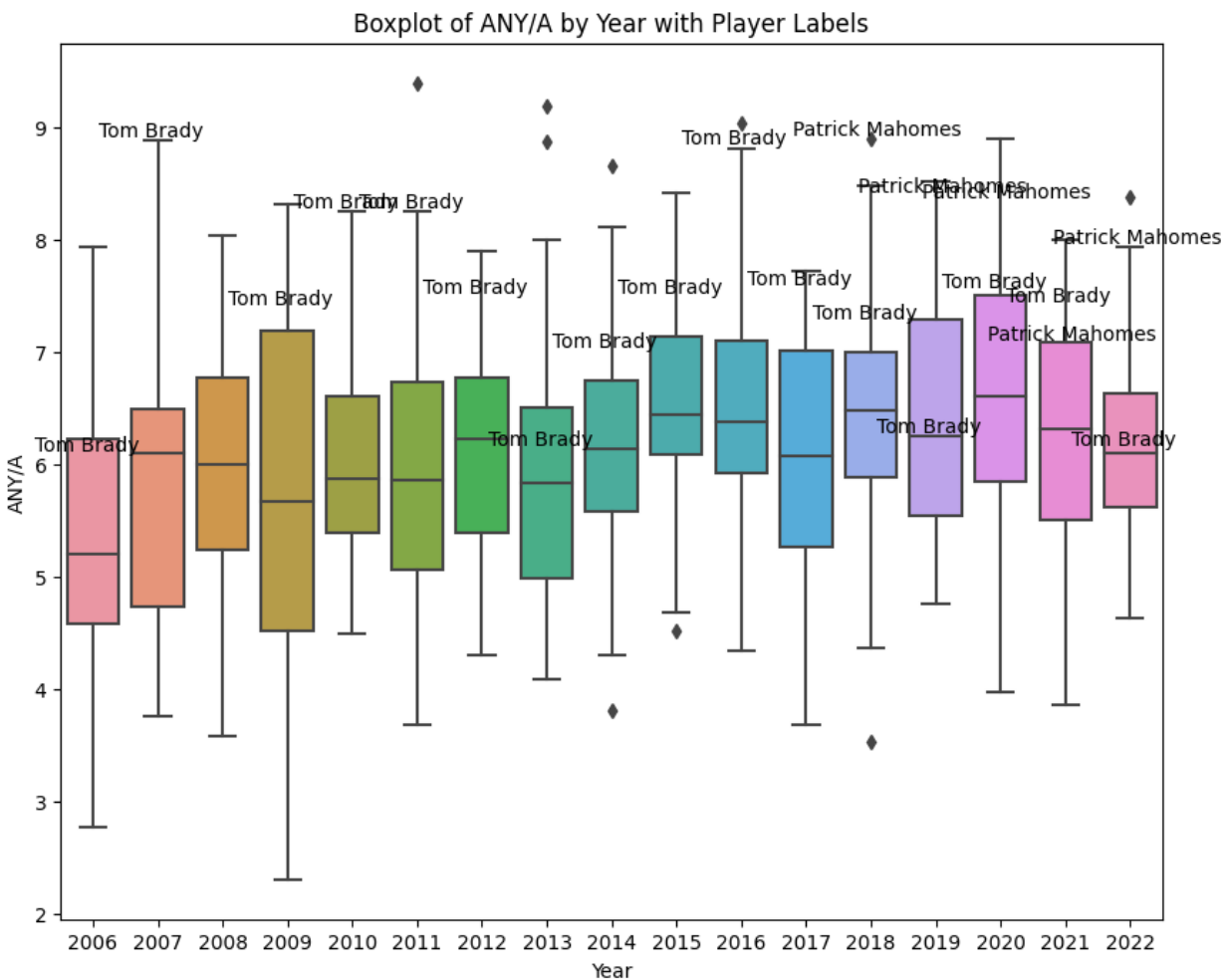
NFL Quarterback Statistical Analysis

ANY/A vs win rate

QBR vs win rate
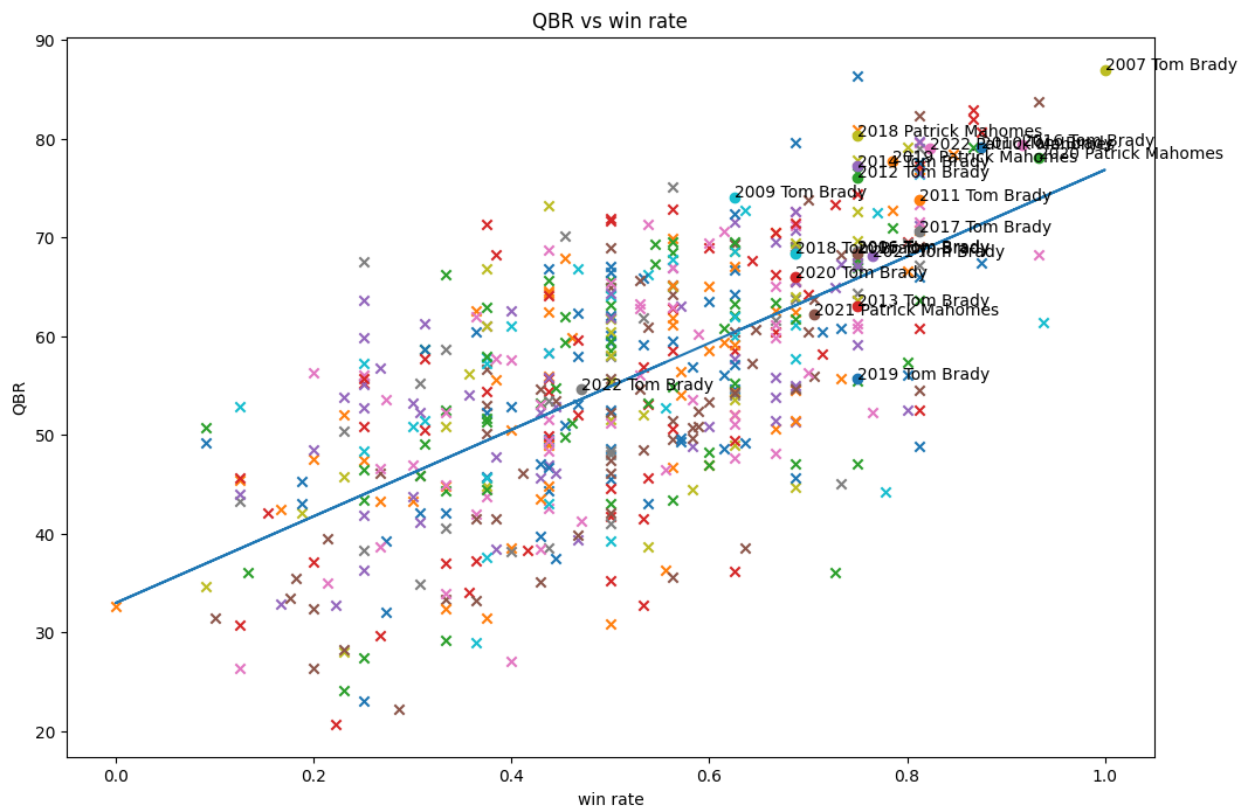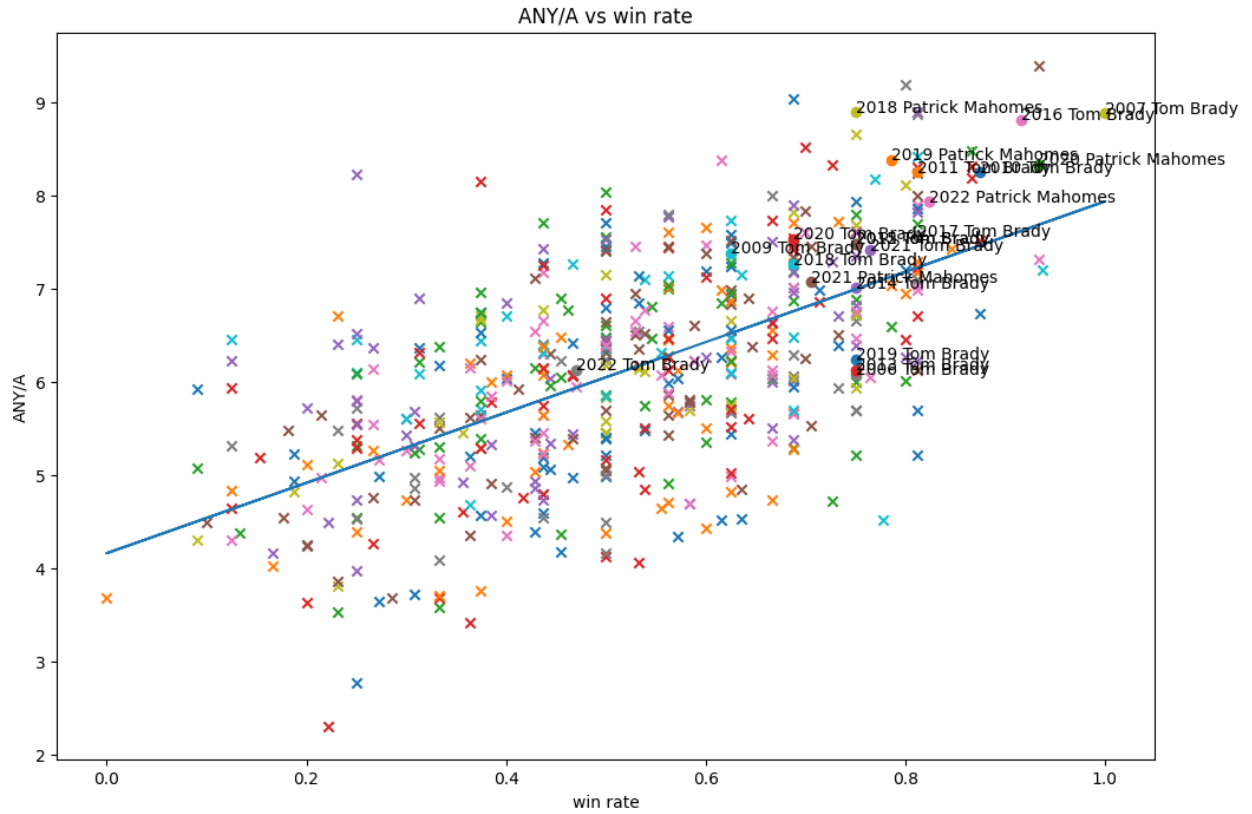
Based on the scatter plots, it is clear that Tom Brady and Patrick Mahomes have been at least better than the NFL average over the past 16 seasons with a few exceptions. Those exceptions being for ANY/A 2006, 2013, 2019 Brady and for QBR 2013, 2019 Brady and 2021 Mahomes. However, outside of 2022 Brady, both quarterbacks are still above the average during that time span.

Let's use box plots to get a closer look at each quarterback's relative performance year by year.

Note: Patrick Mahomes didn't become an NFL starter until the 2018 season, which is why he isn't included before then.

Note: Tom Brady has been a starter in the NFL since the 2001 season. This analysis is using the last 16 seasons worth of data (starting with the 2006 season).
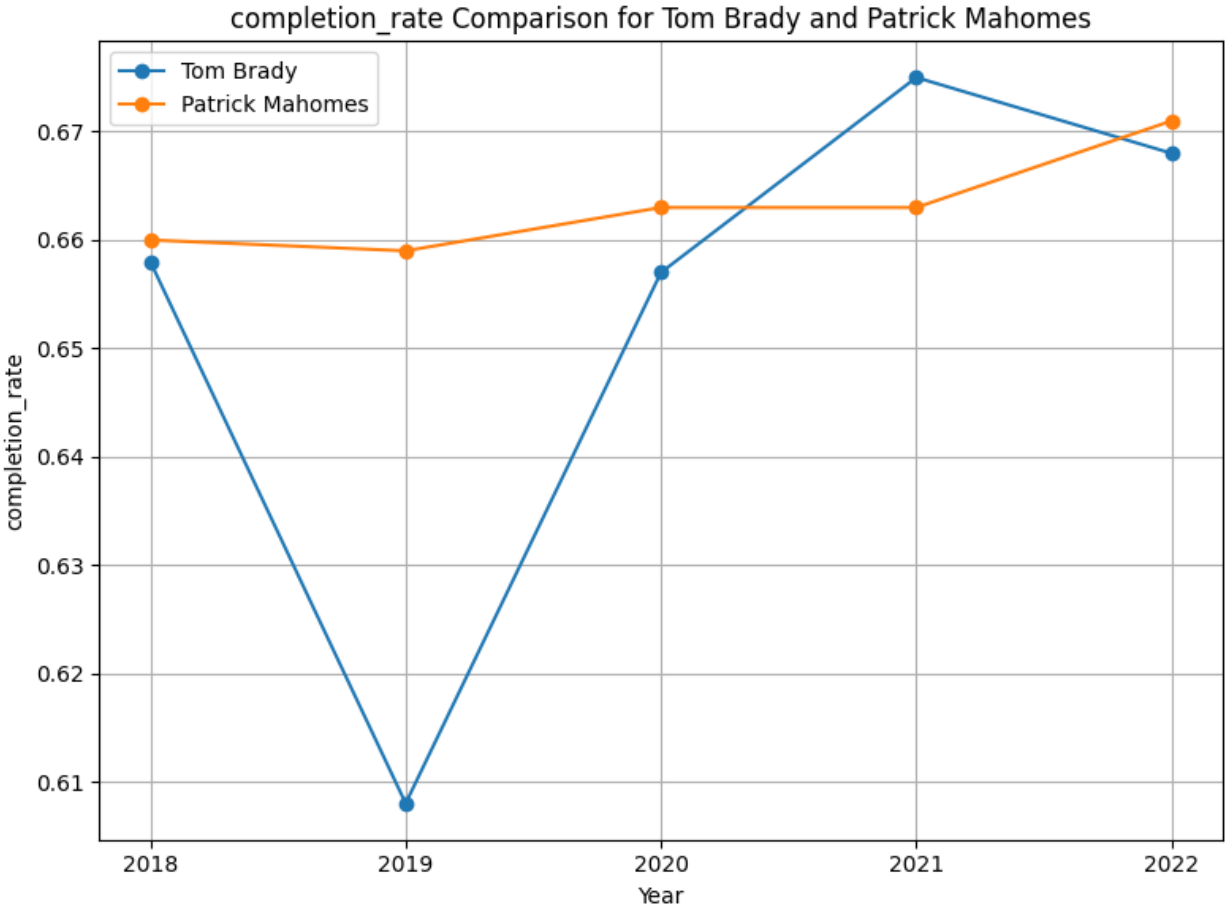

Boxplot of ANY/A by Year with Player Labels
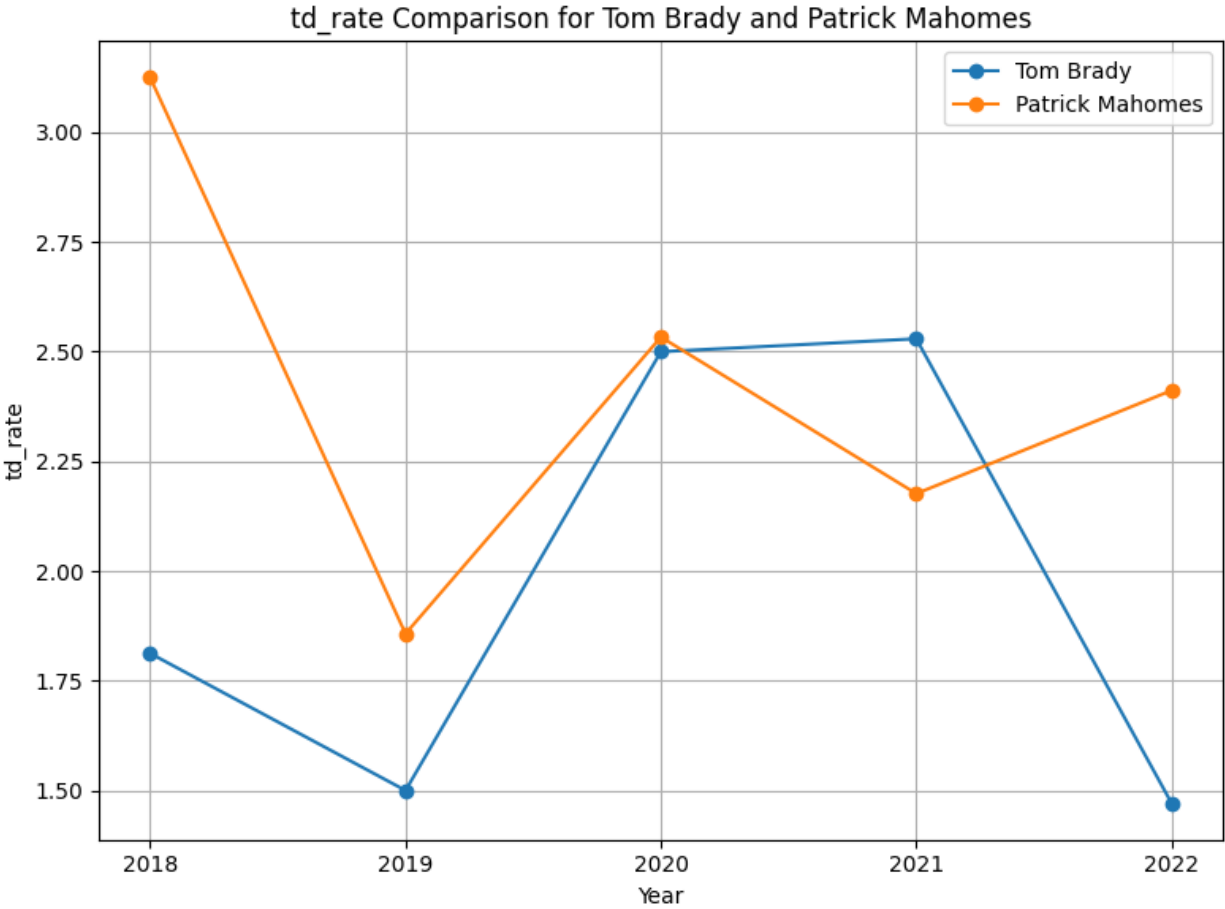
Boxplot of QBR by Year with Player Labels
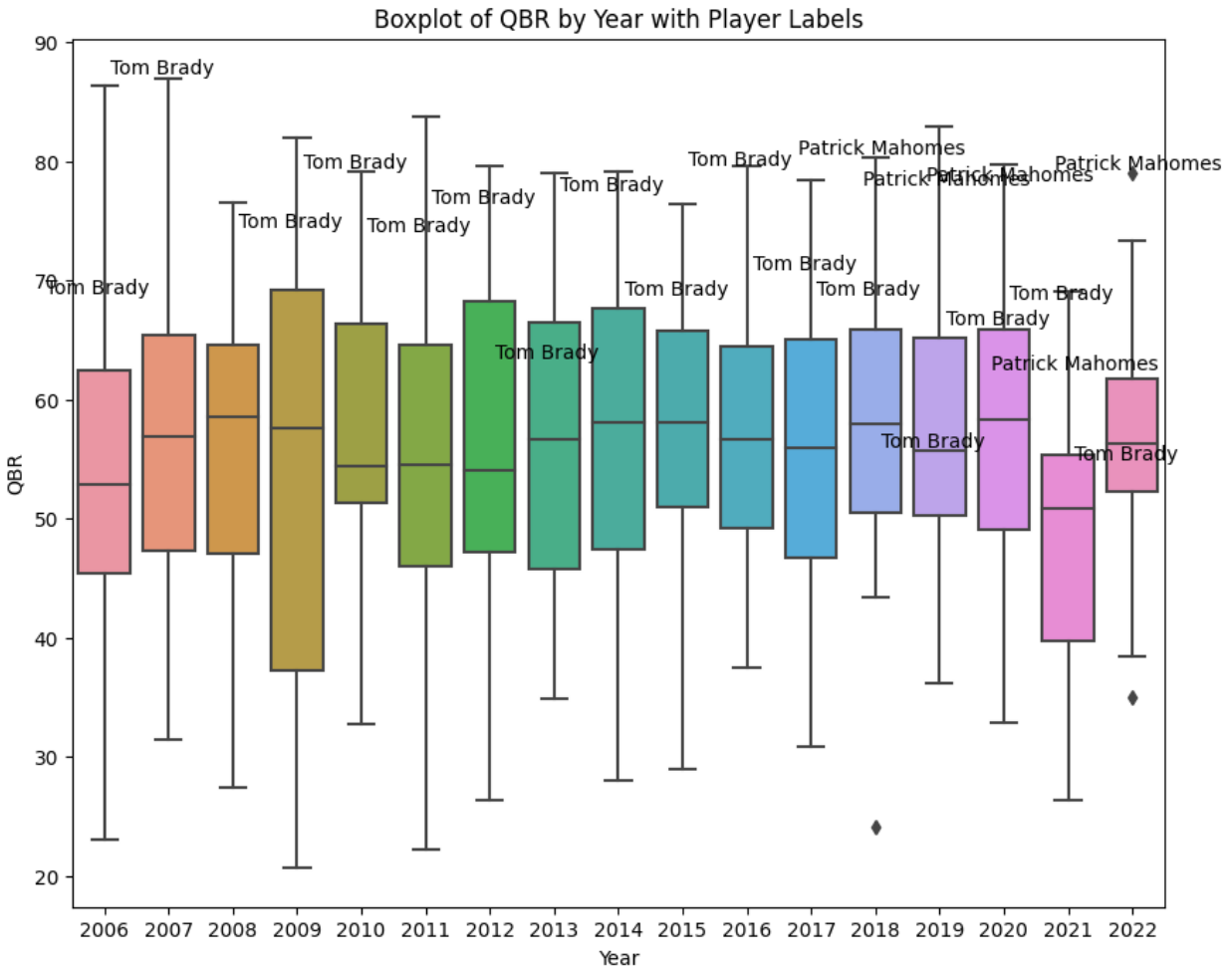
Now after breaking up the data year by year, it reveals a different story. When looking at each year individually, only in the 2022 season was Brady's QBR just below that year's QBR average. Otherwise, both quarterbacks performed in the top half for both ANY/A and QBR. In fact, for the majority of the 16 seasons, both quarterbacks performed in the top 25% for both ANY/A and QBR (17 out 21 combined chances).

The last step is to compare the two head to head and determine if the differences, if any, are statistically significant. For this, all of the passing metrics will be looked at, but again, the focus will be on ANY/A and QBR. A time series analysis will be performed to compare Brady and Mahomes while both were active in the NFL, starting from the 2018 season until the 2022 season (a total of 5 years).

completion_rate Comparison for Tom Brady and Patrick Mahomes

td_rate Comparison for Tom Brady and Patrick Mahomes

int_rate Comparison for Tom Brady and Patrick Mahomes

ANY/A Comparison for Tom Brady and Patrick Mahomes

QBR Comparison for Tom Brady and Patrick Mahomes

sack_rate Comparison for Tom Brady and Patrick Mahomes

win_rate Comparison for Tom Brady and Patrick Mahomes
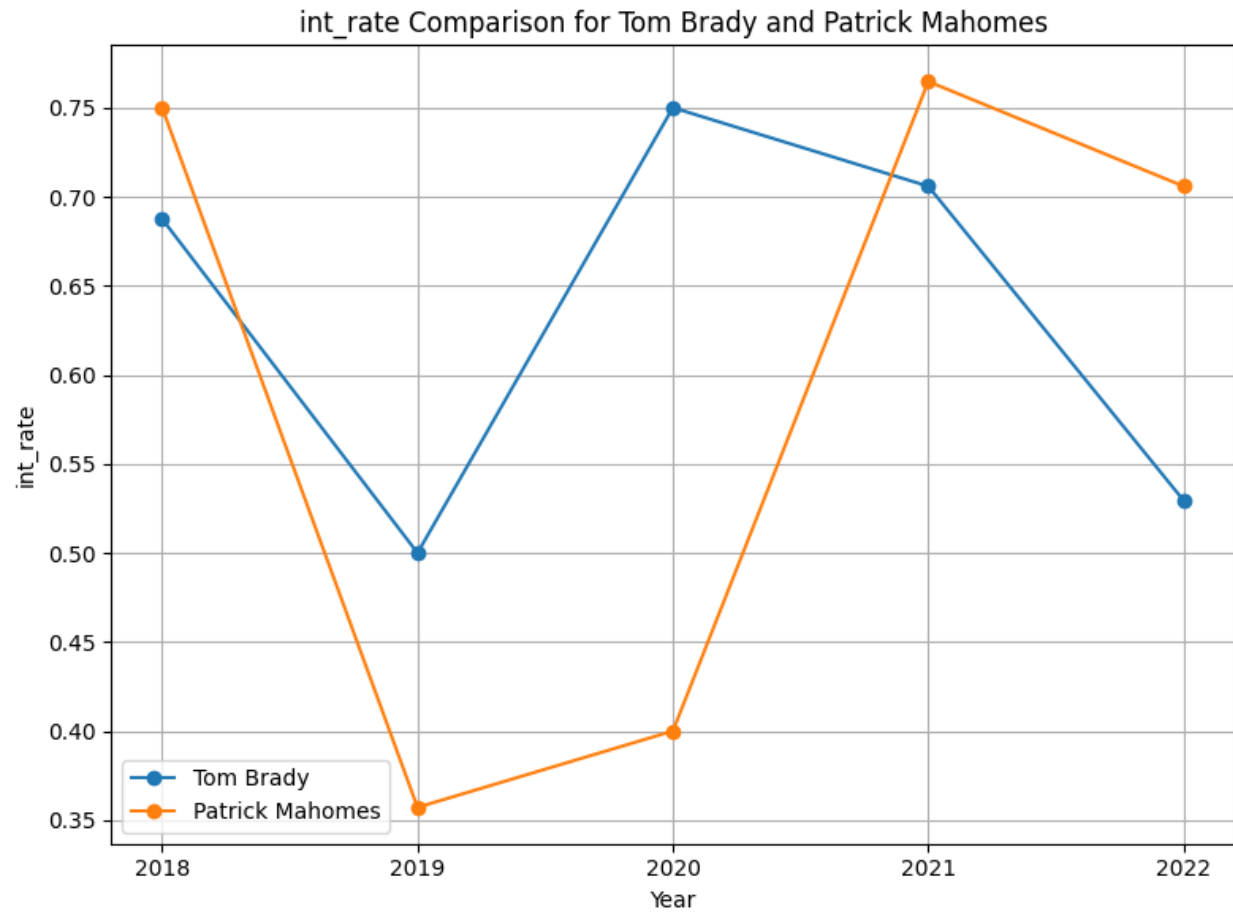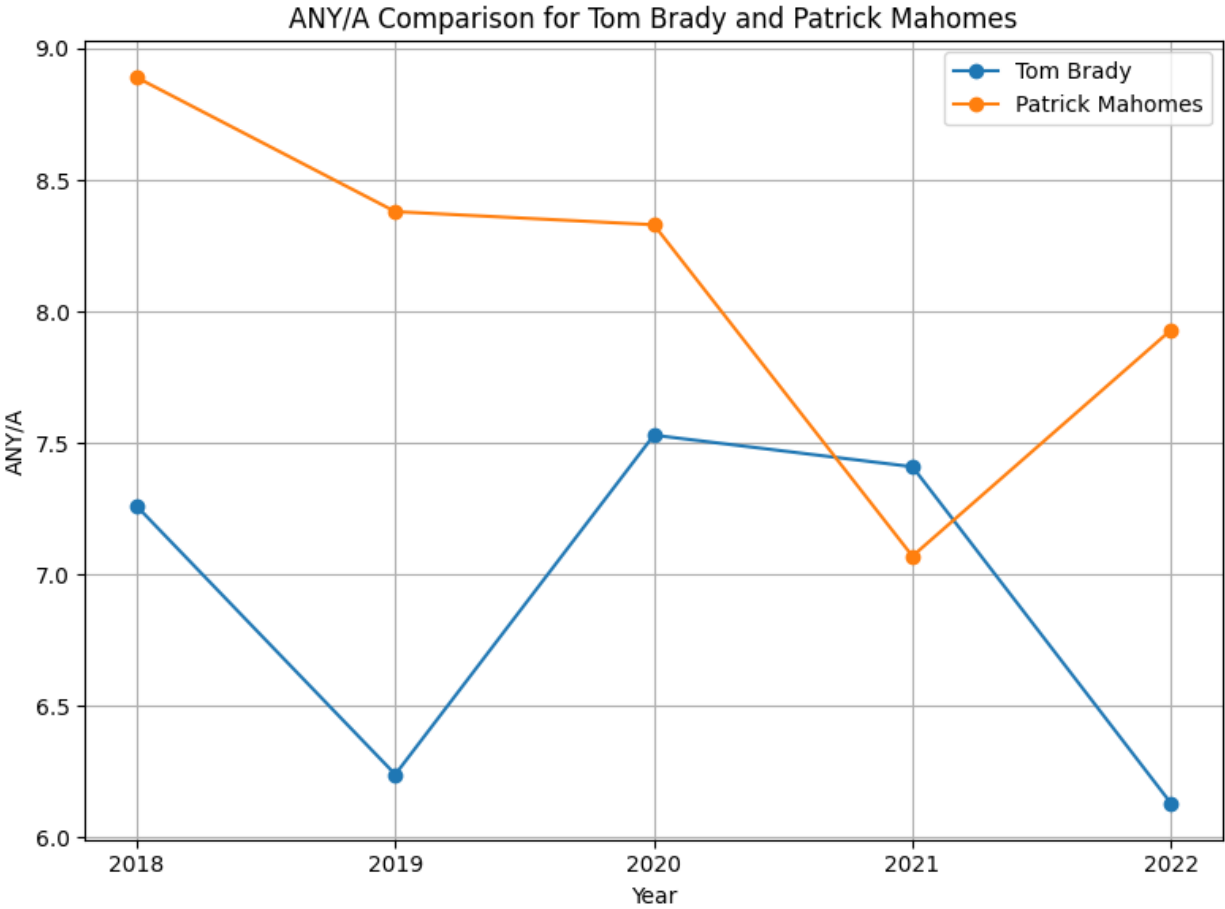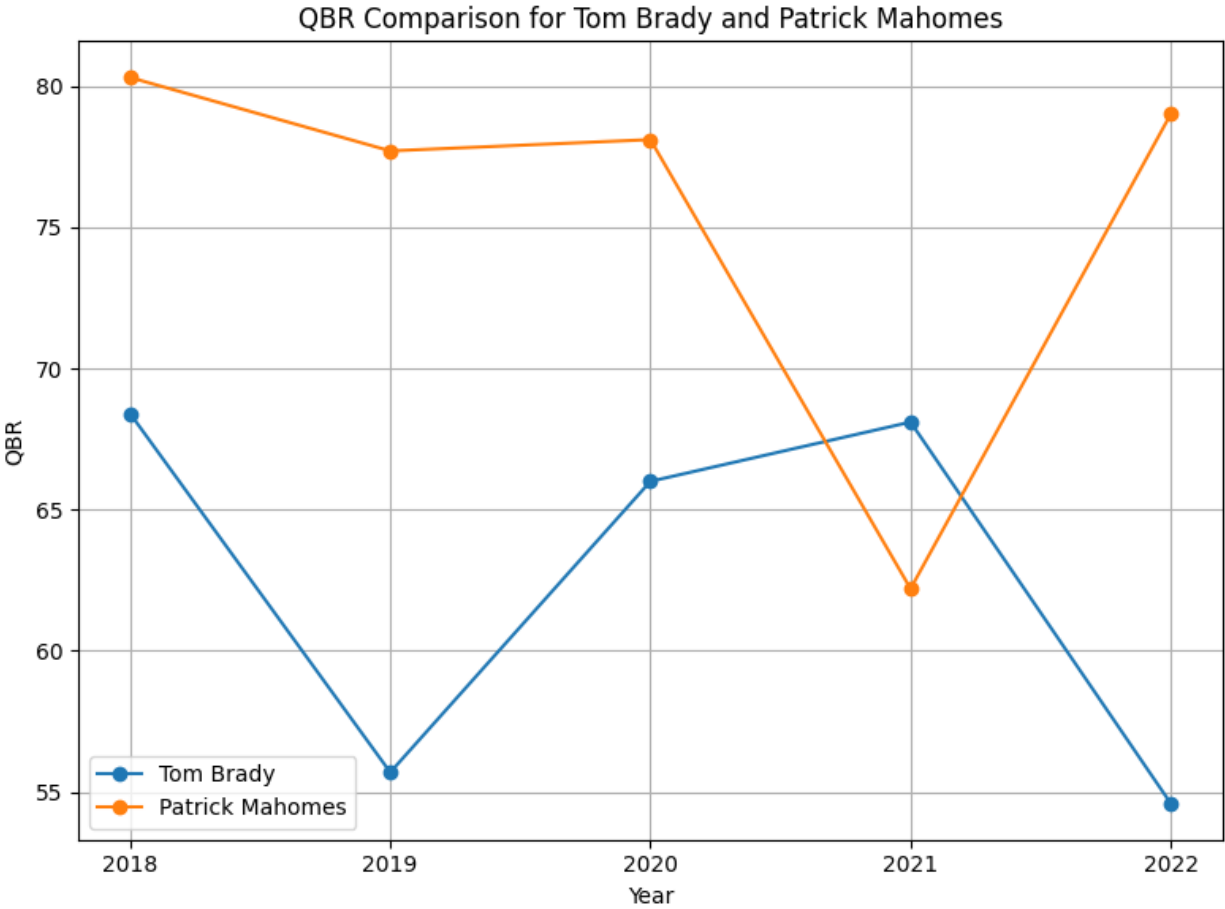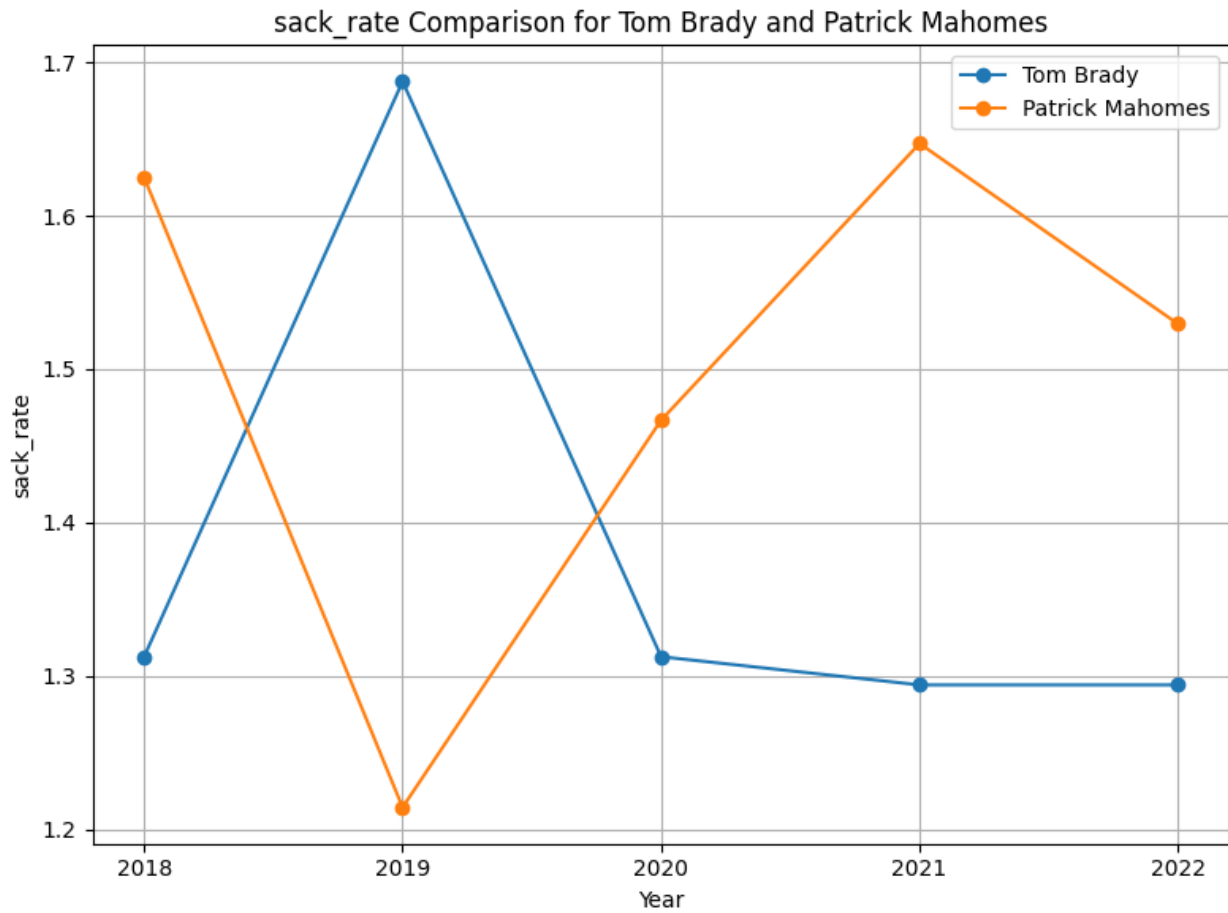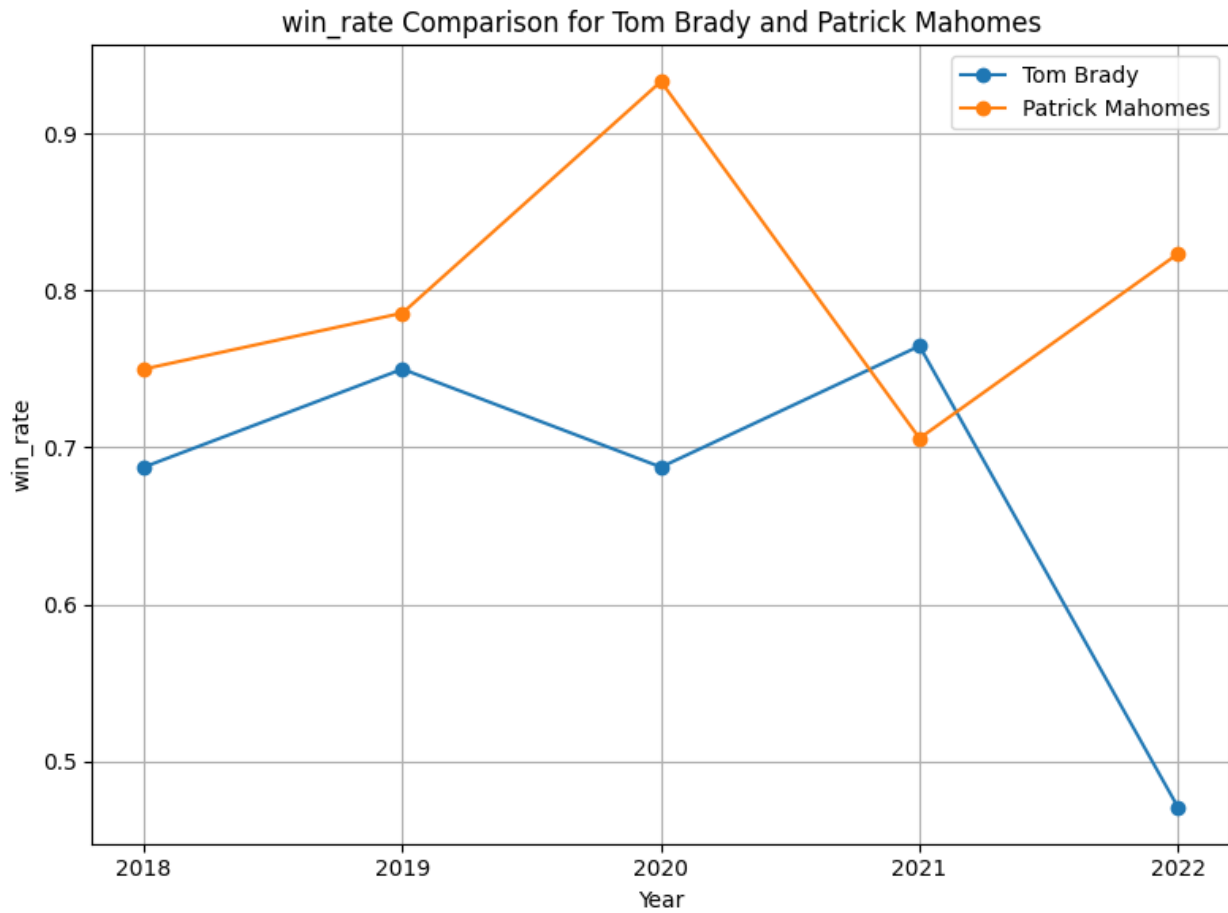
td_int Comparison for Tom Brady and Patrick Mahomes
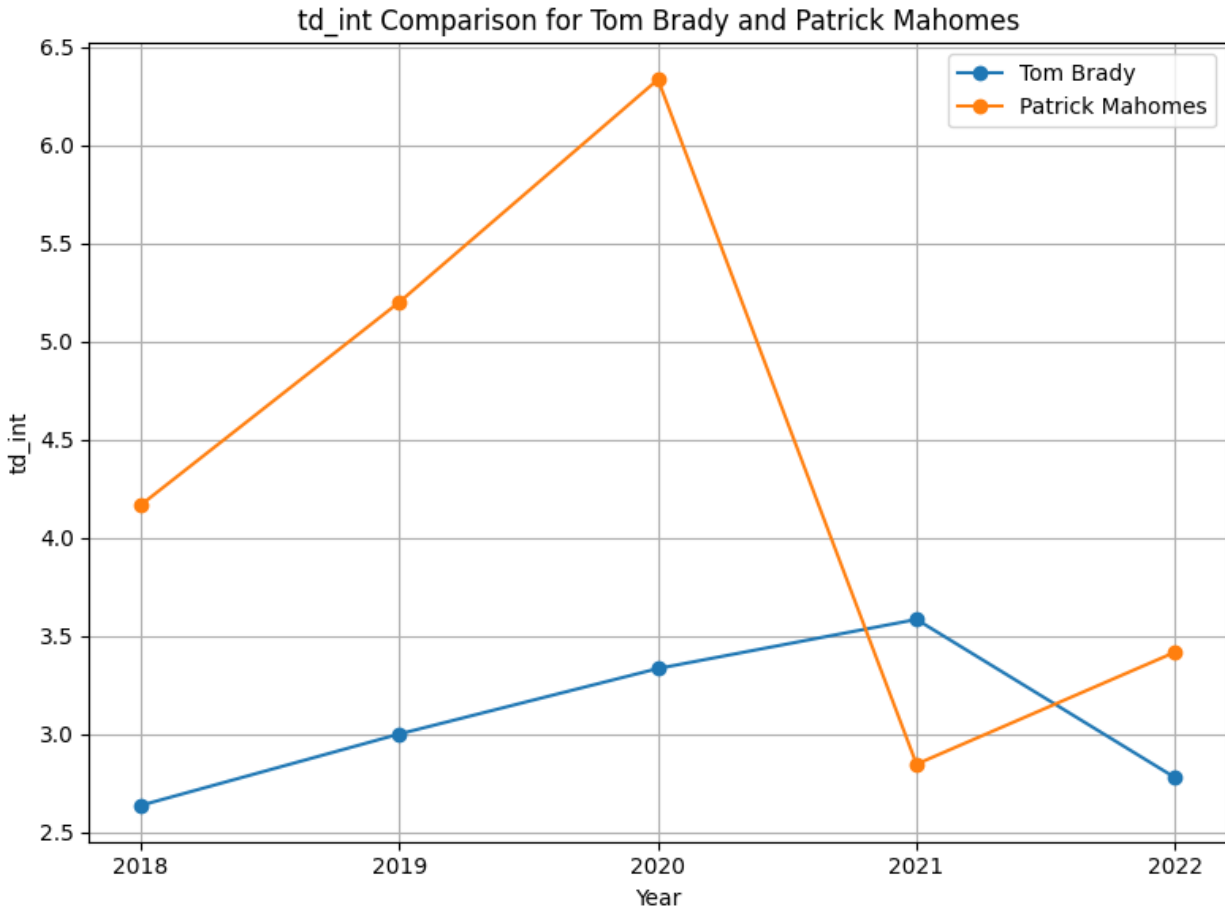
From his debut, Patrick Mahomes appears to perform better in every passing statistic for the majority of the time period. The 2021 season is notable for being the season in which it appears Tom Brady outperforms Mahomes. Additionally, sack_rate is the only metric in which Brady consistently outperforms Mahomes, with the 2020 season being the outlier there.

To determine if the differences are statistically significant, the student's t-test can be used. In this case, between Tom Brady and Patrick Mahomes' ANY/A and QBR. Running the tests produced the following p-values.

Metric	P-Value
ANY/A	0.10325
QBR	0.27179

The conventional threshold or tolerance level is 0.05. As both p-values for ANY/A and QBR were greater than the significance level of .05, this indicates that there might not be a

statistical significant difference between Tom Brady and Patrick Mahomes in either ANY/A or QBR.

Discussion

As a reminder, the research questions were to determine which passing metric has the highest correlation to winning, how good are Tom Brady and Patrick Mahomes, and how do they compare head to head.

Based on the results of the correlation analysis conducted on the top 32 quarterbacks in the NFL of the 16 most recent seasons, the two passing statistics with the highest correlation to winning were QBR and ANY/A (0.66 and 0.61 respectively). Based on descriptive statistics, Tom Brady and Patrick Mahomes were very, very good.

In regards to QBR, both quarterbacks performed exceptionally well. Brady performed in the top 25% of quarterbacks in 13 out of 15 (86%) eligible seasons (Brady was injured during the 2008 season and did not play). Mahomes performed in the top 25% in terms of QBR in every season of his pro career so far (5 for 5, or 100%). Not only that, but Brady was a top 5 quarterback in terms of QBR in 11 out of 15 eligible seasons (69%). Mahomes was a top 5 quarterback in terms of QBR in 5 out of 5 (100%) eligible seasons.

In regards to ANY/A, both quarterbacks were once again exceptional. Brady was in the top 25 percentile in 12 out of 15 (80%) eligible seasons. Mahomes, once again 5 for 5 (100%). In terms of ANY/A, Brady was a top 5 quarterback in 8 out of 15 (53%) seasons. Mahomes, 4 out of 5 (80%).

Seasons in which Tom Brady and Mahomes were top 5 in either QBR or ANY/A

QBR	ANY/A
2006 Brady (2nd)	2007 Brady (1st)
2007 Brady (1st)	2010 Brady (1st)
2009 Brady (5th)	2011 Brady (2nd)
2010 Brady (1st)	2012 Brady (2nd)
2011 Brady (3rd)	2015 Brady (4th)
2012 Brady (2nd)	2016 Brady (2nd)
2014 Brady (3rd)	2017 Brady (5th)

2015 Brady (5th)	2018 Mahomes (1st)
2016 Brady (2nd)	2019 Mahomes (2nd)
2017 Brady (3rd)	2020 Mahomes (2nd)
2018 Mahomes (1st)	2021 Brady (5th)
2019 Mahomes (2nd)	2022 Mahomes (2nd)
2020 Mahomes (2nd)	
2021 Brady (2nd)	
2021 Mahomes (5th)	
2022 Mahomes (1st)	

That might seem impressive, but what about other top quarterbacks? Surely there have been other quarterbacks who have been in the top 5 for either QBR or ANY/A multiple times. Let's look at how frequently other quarterbacks are top 5 in either passing category.

Quarterbacks with multiple top 5s in QBR or ANY/A and
how many seasons they've played in the same time frame

QBR (Player, top 5s / total seasons)	ANY/A (Player, top 5s / total seasons
Tom Brady, 11 / 15 (73%)	Drew Brees, 11 / 15 (73%)
Drew Brees, 10 / 15 (66%)	**Tom Brady, 8 / 15 (53%)**
Peyton Manning 8 / 9 (88%)	Aaron Rodgers, 7 / 14 (50%)
Matt Ryan, 6 / 15 (40%)	Peyton Manning, 6 / 9 (66%)
Aaron Rodgers, 6 / 14 (42%)	Philip Rivers, 6 / 15 (40%)
Patrick Mahomes, 5 / 5 (100%)	**Patrick Mahomes, 4 / 5 (80%)**

This shows just dominant Brady and Mahomes are. Brady has the most top 5 QBRs in the last 16 seasons. Only Peyton Manning and Mahomes beat Brady top 5 QBRs when taking into account seasons played. In ANY/A, Drew Brees does have more top 5s in the same amount of seasons, and Manning once again has a higher rate of being top 5, but Mahomes is once again on top of the pack when considering he's only in his 5th year in the NFL.

Based on the time series analysis of the two quarterbacks head to head, since Mahomes entered the league, he has outperformed Brady in almost every passing metric. Completion rate? Brady bested him only in 2021. Touchdown rate? In 2020, the two were neck and neck (2.53 vs 2.50), but only during the 2021 season was Brady's touchdown rate higher than Mahomes. Brady's interception rate was better than Mahomes in 3 out of 5 seasons (2018, 2021, and 2022). Sack rate is another category in which Brady outperformed Mahomes, only in 2019 was Brady's higher. Win rate and touchdown to interception ratio? The only year Brady bested Mahomes in these categories was once again 2021.

What about the key metrics that have the highest correlation to winning, QBR and ANY/A? 2021 was once again a relatively great year for Tom Brady as that is the only year in which his QBR and ANY/A was higher than Mahomes. In fact, 2021 was the only year in which Brady was better than Mahomes in every passing statistic while both quarterbacks were in the NFL.

Tom Brady, Patrick Mahomes head to head

	Tom Brady	**Patrick Mahomes**
Total Seasons	15	5
Average completion_rate	0.6501	**0.6632**
Average td_rate	2.0755	**2.4207**
Average int_rate	**0.5681**	0.5955
Average ANY/A	7.3675	**8.1200**
Average QBR	70.6375	**75.46**
Average sack_rate	1.6071	**1.4965**
Average td_int	**4.4573**	4.3926
Best completion_rate	**0.689**	0.671
Best td_rate	**3.125**	**3.125**
Best int_rate	**.1667**	.3571
Best ANY/A	8.88	**8.89**

Best QBR	**87.0**	80.3
Best sack_rate	**1.0000**	1.2143
Best td_int	**14.000**	6.3333

Does that mean that Patrick Mahomes will surely be as successful as Tom Brady was as a quarterback in the NFL? No, of course not. Part of what makes the NFL so entertaining is the fact that it is near impossible to predict what's going to happen, let alone if someone were to match or even surpass Tom Brady's record 7 Super Bowl victories. But the data does suggest that Mahomes is on the right track to having a successful career.

References

Sports Reference LLC. Pro-Football-Reference.com - Pro Football Statistics and History. https://www.pro-football-reference.com/. 8 Dec. 2023.