

Jalen Hurts Player Analysis

Noah McMahon



Outline



Executive
Summary



Introduction



Methodology



Results



Conclusion



Appendix

Executive Summary

Methodology

Data Collection through downloaded CSVs

Exploratory Data Analysis

- Data wrangling, visualization, SQL

Comparative Analysis

Machine Learning Prediction



Summary of Results

EDA found most impactful features on game results

Comparative Analysis found key differences between subject and his peers

Machine Learning found the best model for predicting game result

Introduction



Objective of Analysis

Evaluate Jalen Hurts's probability of winning based on his unique metrics



Questions

Which element of Hurts's game is most successful? Least successful?

How does he compare to his peers?

What is the best model for predicting his success?

METHODOLOGY



Data Collection

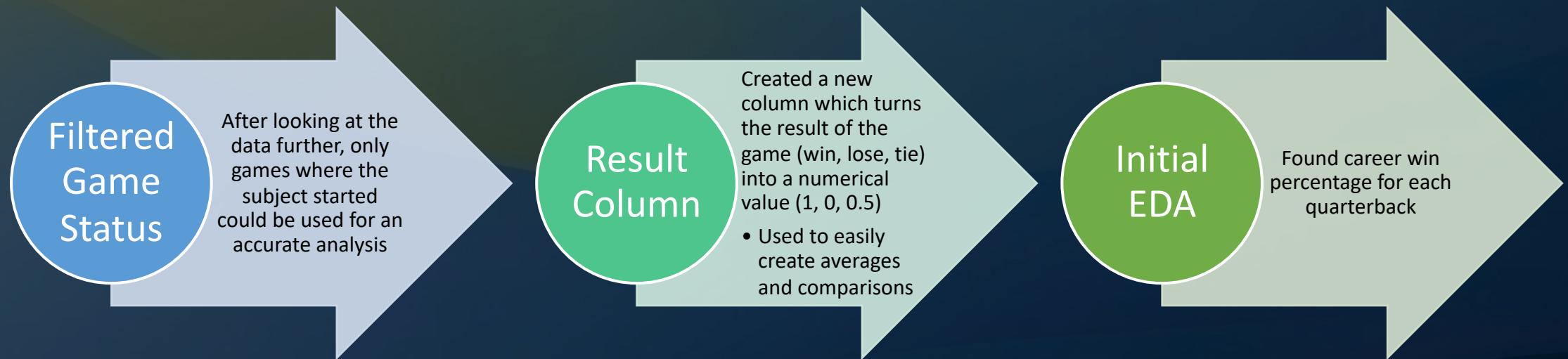
- Data sources
 - Jalen Hurts Data -
<https://www.profootballreference.com/players/H/HurtJa00/gamelog/>
 - Patrick Mahomes Data -
<https://www.pro-football-reference.com/players/M/MahoPa00/gamelog/>
 - Joe Burrow Data -
<https://www.pro-football-reference.com/players/B/BurrJo01/gamelog/>
 - Derek Carr Data -
<https://www.pro-football-reference.com/players/C/CarrDe02/gamelog/>

Process of Using CSVs

Search	Found required information in CSV format from Pro Football Reference
Transformation	Turned the CSV file into a Dataframe
Filtering	Filtered the data into only significant metrics <ul style="list-style-type: none">• Such as Passing Yards, Touchdowns, etc.
Handling	Dealt with missing values and data types

Steps were performed for all 4 data sources

Data Wrangling



EDA With Visualization

Initial

- Gathered base information
 - Averages, Medians, and Range

Graphing

- Created scatterplots, bar plots, line graphs, and box plots to understand relationships between various statistical points

Step was performed solely for Jalen Hurts

EDA With SQL

- Queries used
 - Ranges
 - Passing Yards, Completion Percentage, Sacks, Rushing Yards
 - Opponent Based Statistics
 - Total Passing Yards, Average Passing Yards
 - Completion Percentage
 - Average Rushing Yards
 - Average Interceptions, Total Interceptions
 - Average Sacks, Total Sacks
 - Win/Lose Counts by Home/Away Status
 - Rushing Yards and Passing Yards Counts based on Win/Lose

Step was performed solely for Jalen Hurts

Comparative Analysis

- Comparisons Made
 - Total Passing Yards,
 - Total Rushing Yards
 - Per Attempt
 - Passing Yards
 - Rushing Yards
- By Year
 - Completion Percentage
 - Win Percentage
 - Interceptions
 - Sacks

Steps were performed for all 4 data sources

Machine Learning

Preprocessed, Fit, and Transformed Data

Split data into Training and Test Splits

Types of Machine Learning Models used

- Logistic Regression Model
- Support Vector Machine (SVM)
- Decision Tree Classifier
- K-Nearest Neighbors

Steps performed for all Models

- Found best parameters
- Fit the training data to the model
- Found accuracy scores for the model and the test data
- Created confusion matrix for visualization

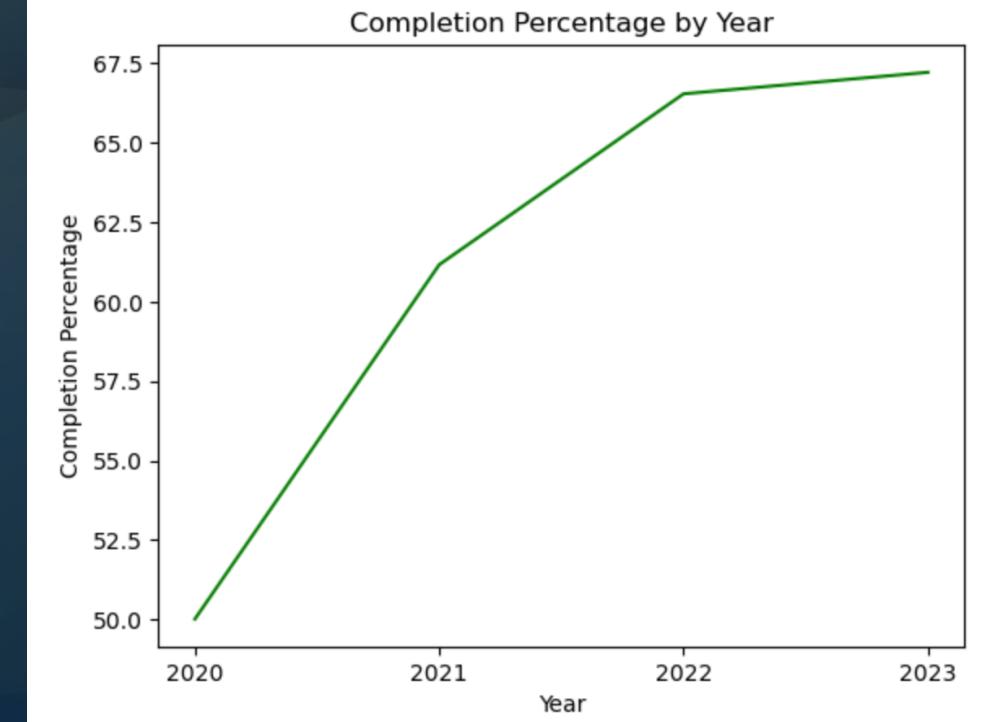
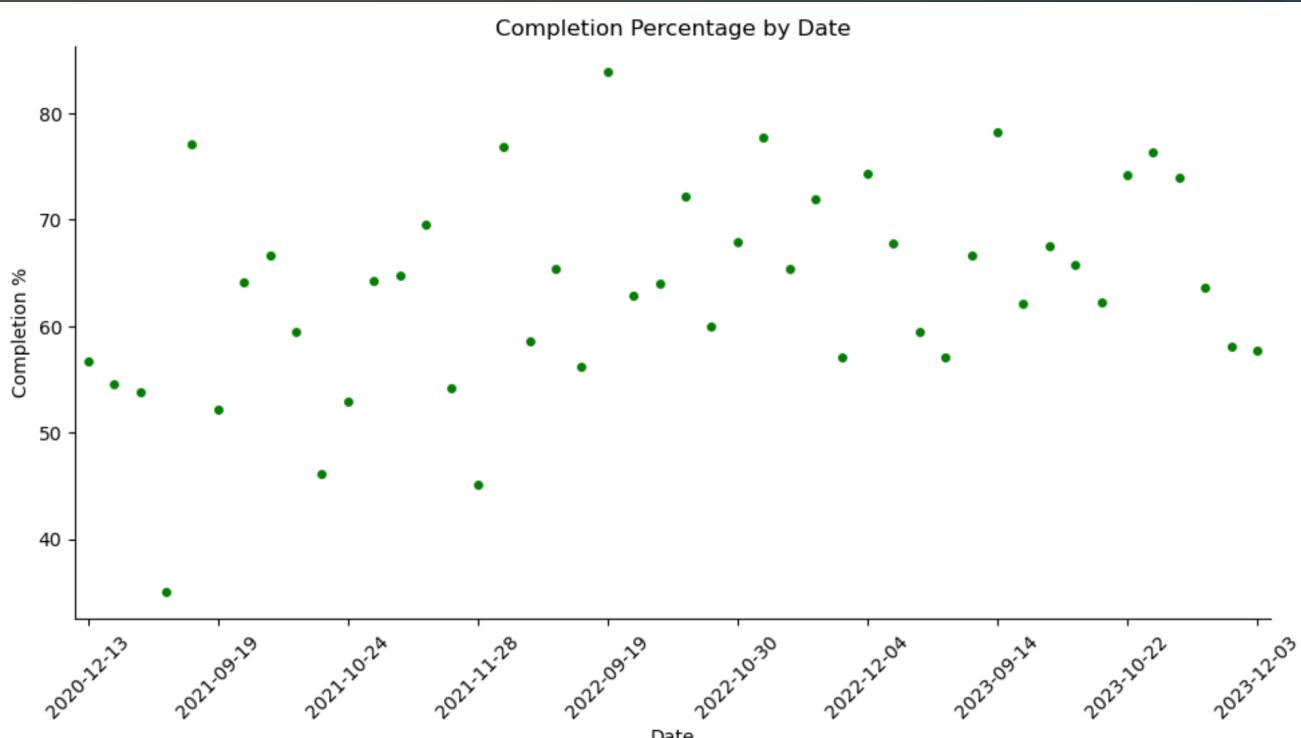
Accuracy of all models were visualized in a bar plot

Step was performed solely for Jalen Hurts

EXPLORATORY DATA ANALYSIS



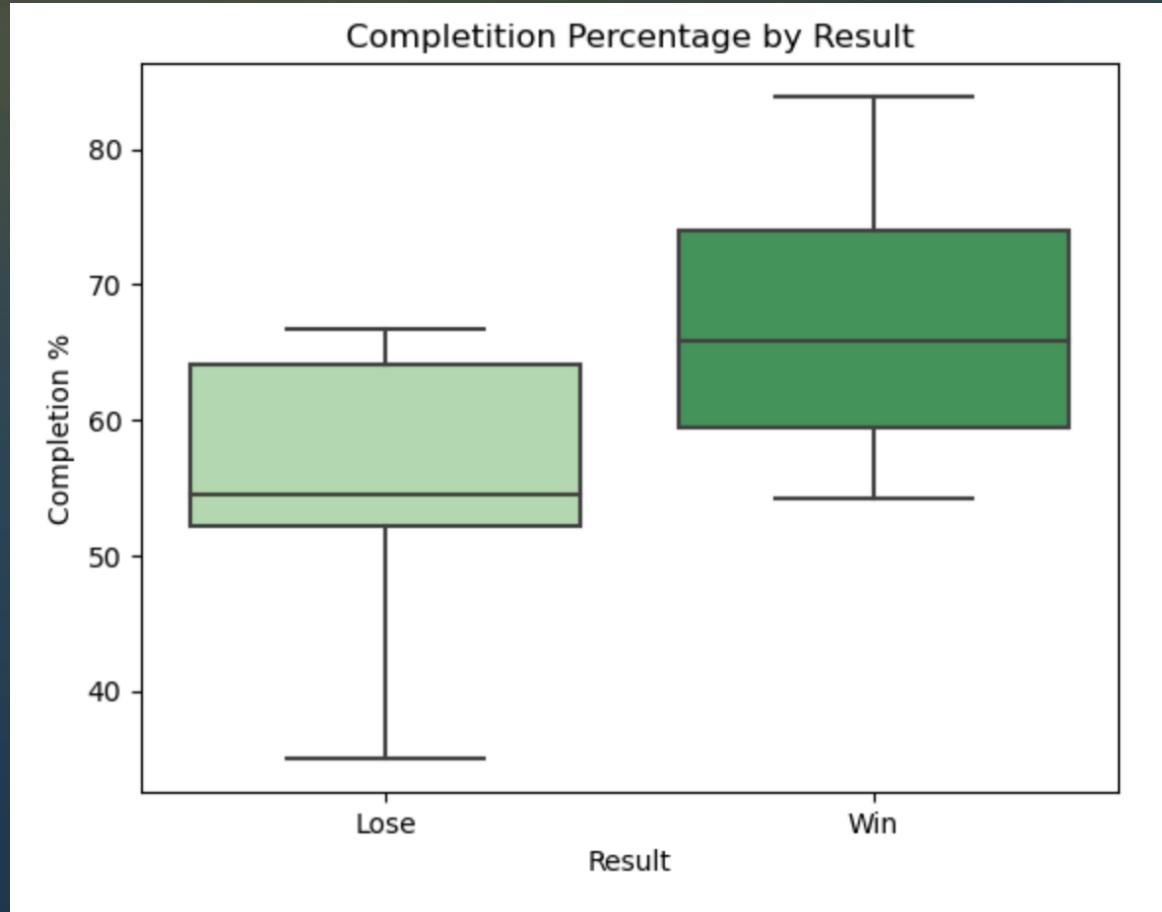
Completion Percentage



From December 2020 to December 2023, his average completion percentage has increased

1. 50% in 2020
2. 61% in 2021
3. 67% in 2022
4. 66% in 2023

Completion Percentage

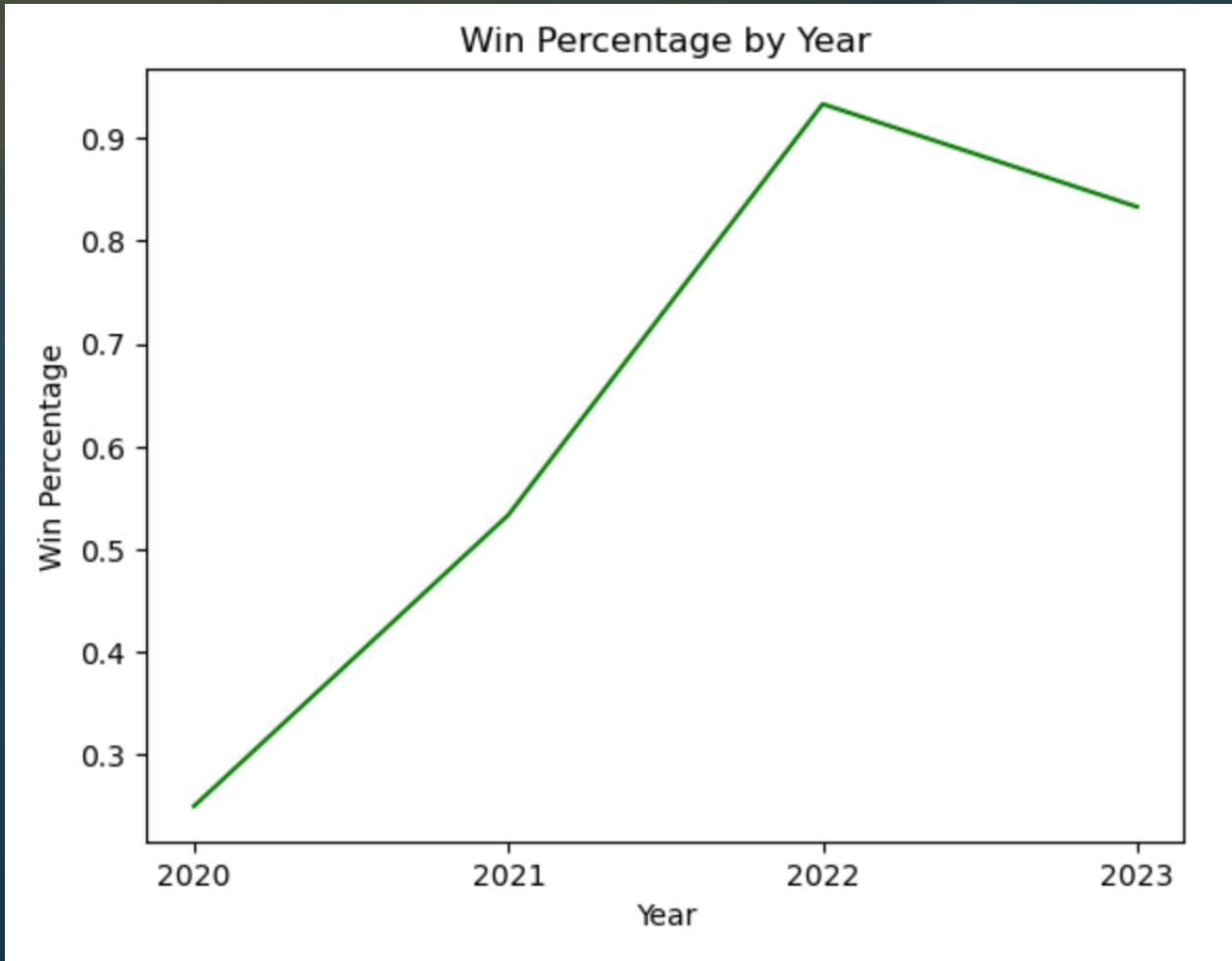


By win/lose result, for wins he averages around 66-68% and for loses he averages 55%.

Higher completion percentage leads to more wins

With his average increasing each year, we see an increase in win percentage

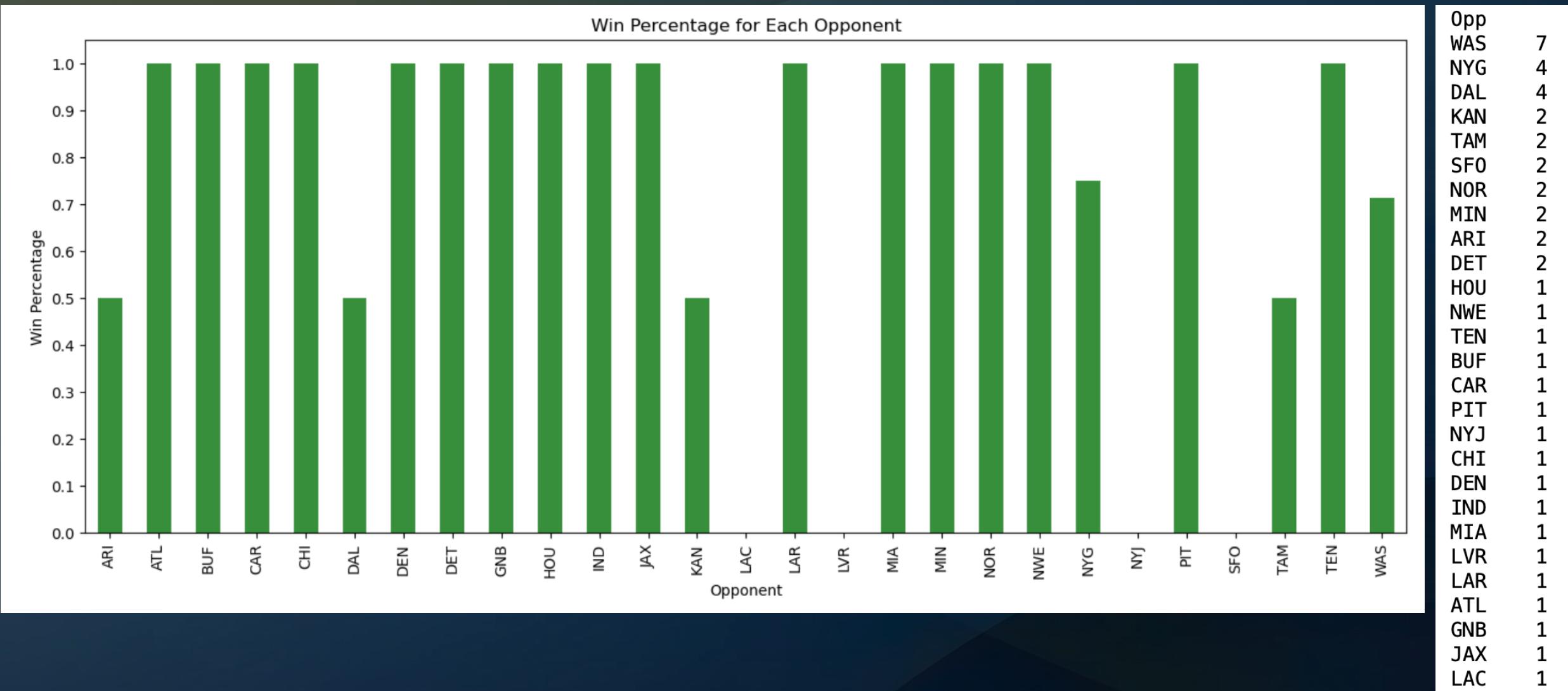
Win Percentage



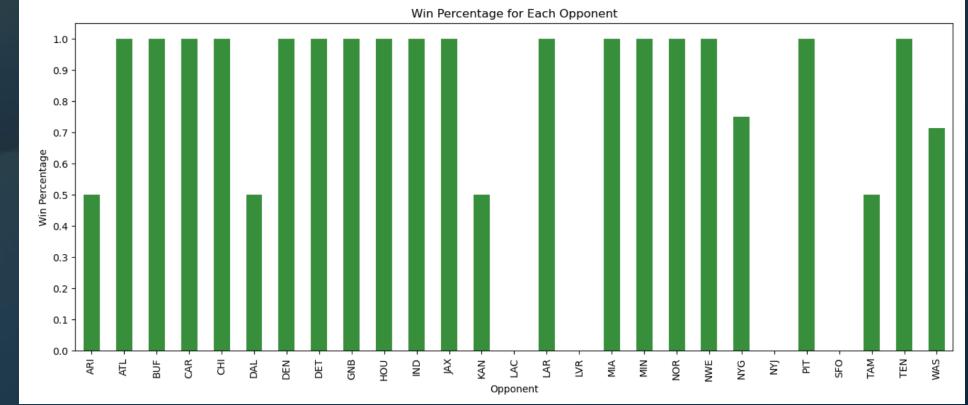
From December 2020 to December 2023, his win percentage has increased but slightly dropped in 2023 from 2020

1. Sub 30% in 2020
2. 50% in 2021
3. 90%+ in 2022
4. 82% in 2023

Win Percentage



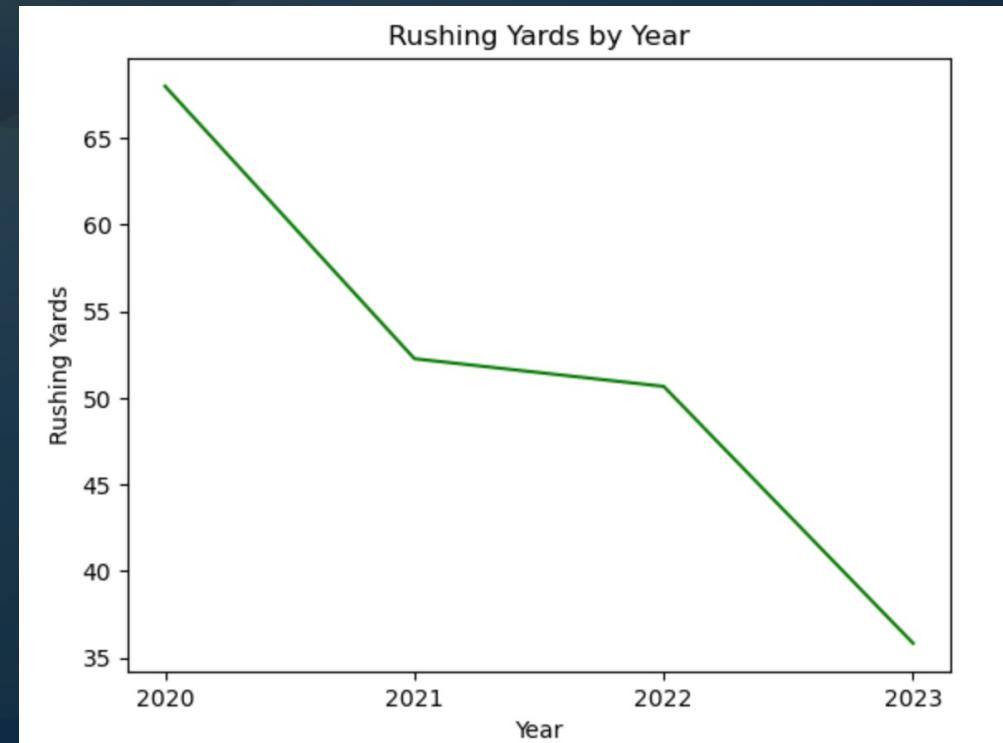
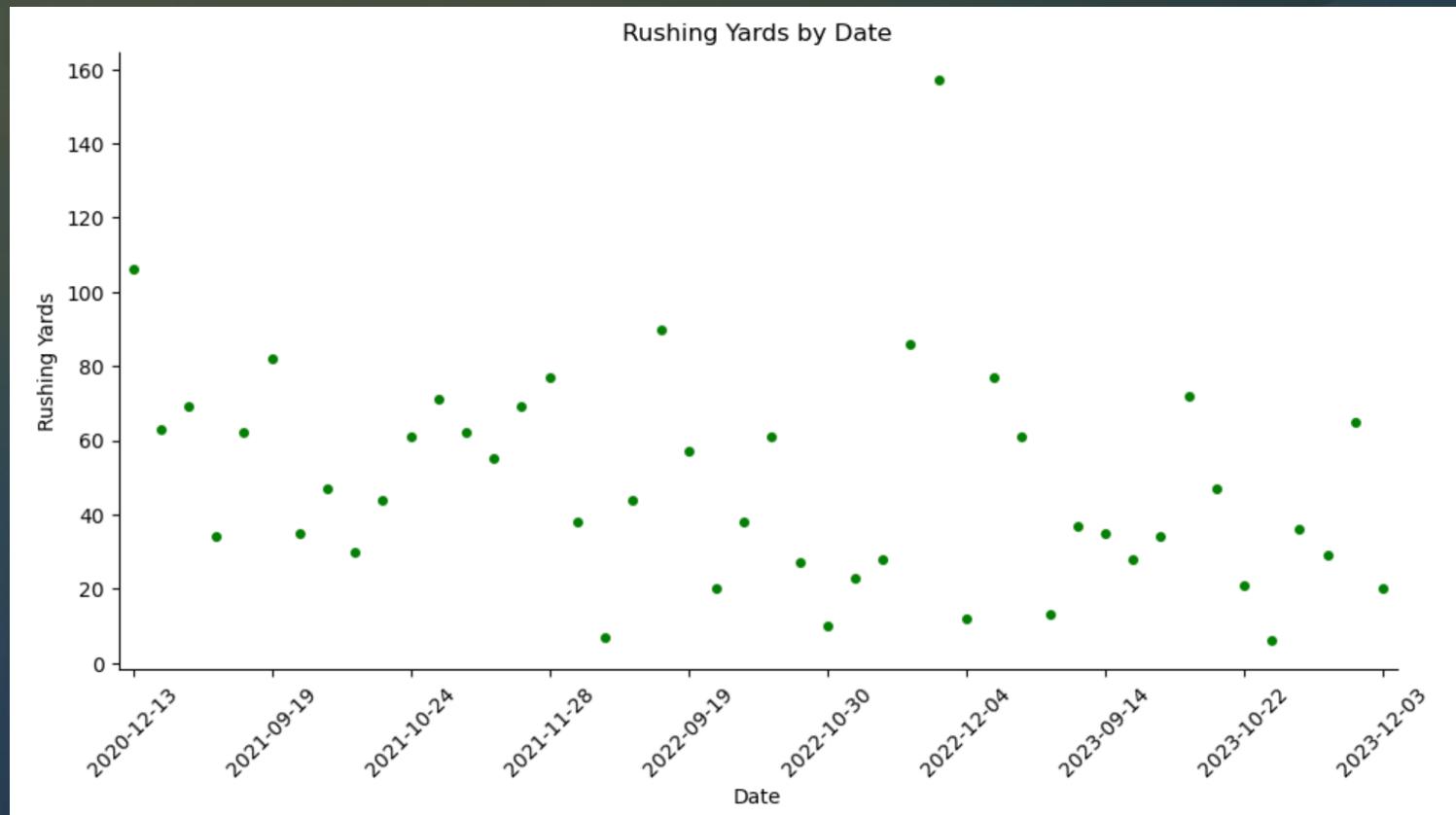
Win Percentage



By opponent is somewhat difficult to show as he has only played some teams once which may not result in accuracy in long term performance, but we can get a decent view at performance from some of the opponents' games.

1. Washington: 71% win rate from 7 games.
2. New York Giants: 75% win rate from 4 games
3. Dallas: 50% win rate from 4 games
4. San Francisco: 0% win rate from 2 games
5. Minnesota: 100% win rate from 2 games

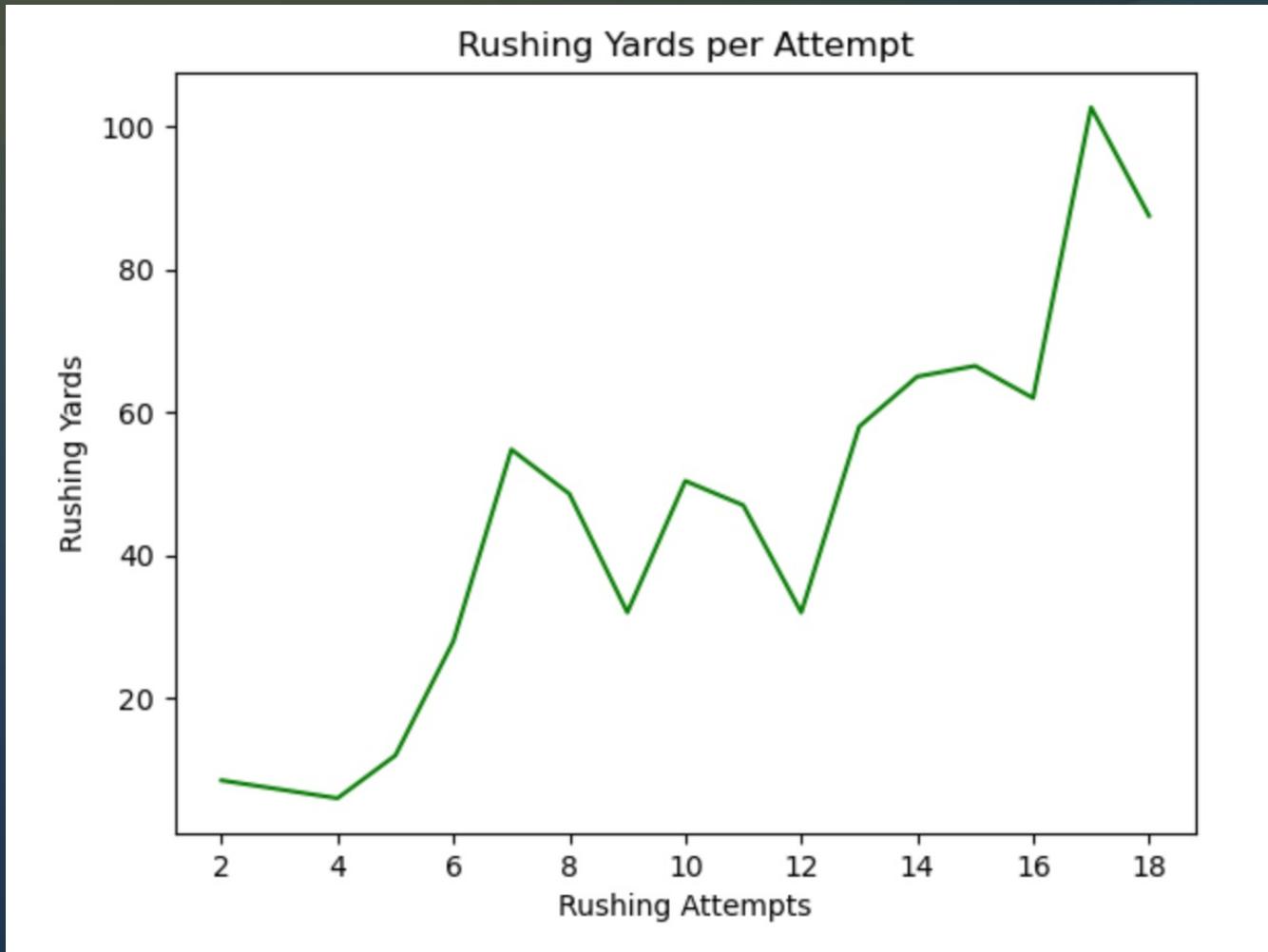
Rushing



From December 2020 to December 2023, his average rushing yards have been in decline

1. 70 yards in 2020
2. 51 yards in 2021
3. 50 yards in 2022
4. 36 yards in 2023

Rushing

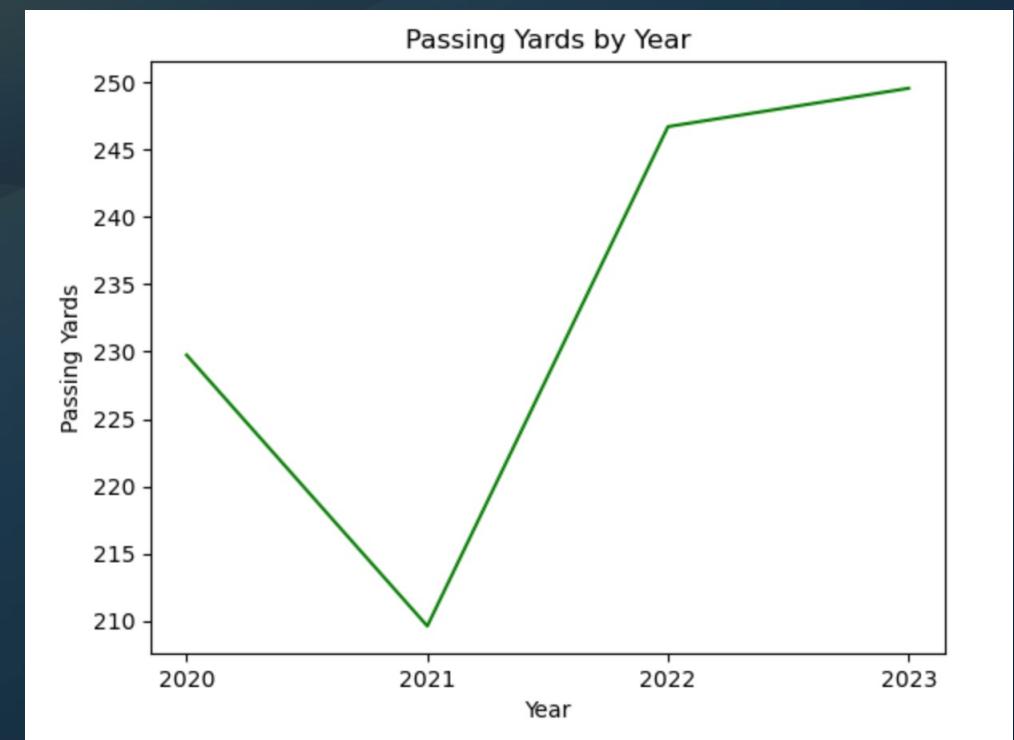
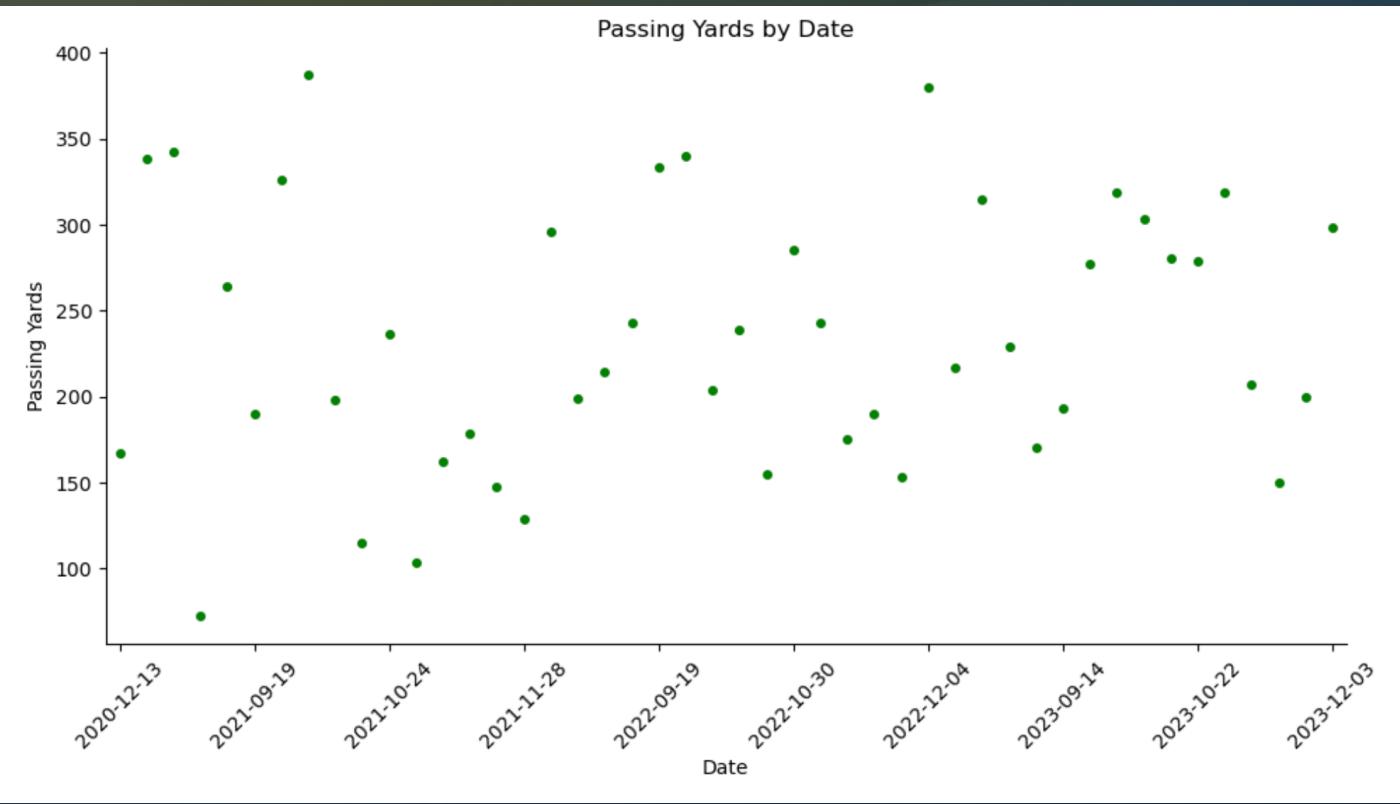


When he has a higher number of rushing attempts, he has more rushing yards.

The decline in his rushing yards would suggest a decline in number of rushes.

He could be shifting his game away from the run and into a better passing game.

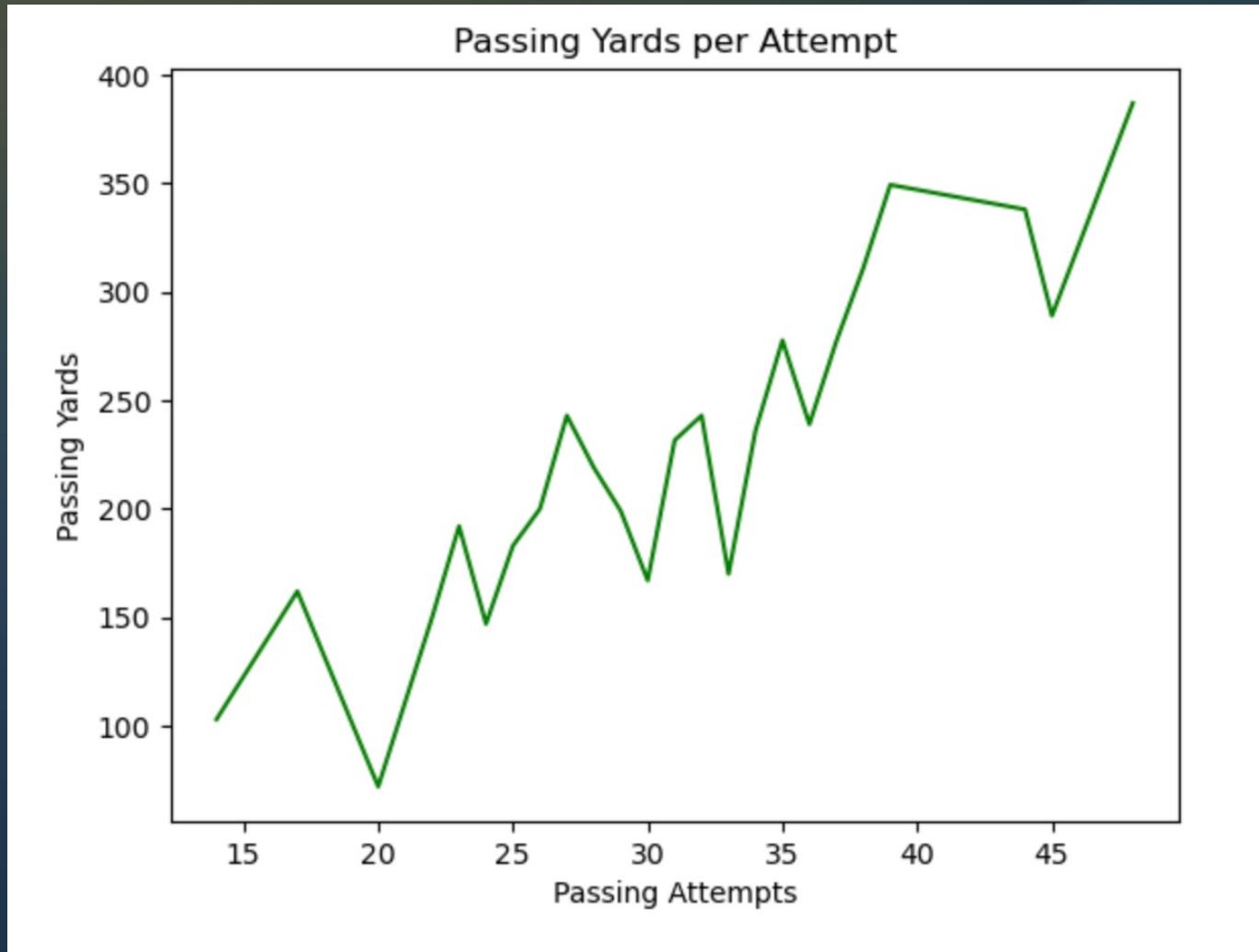
Passing



From December 2020 to December 2023, his average passing yards has increased outside of a dip from 2020 to 2021

1. 230 yards in 2020
2. 210 yards in 2021
3. 247 yards in 2022
4. 250 yards in 2023

Passing



The more passing attempts he makes, the higher the yardage he averages. It is clear Hurts has increased the number of passes he is throwing.

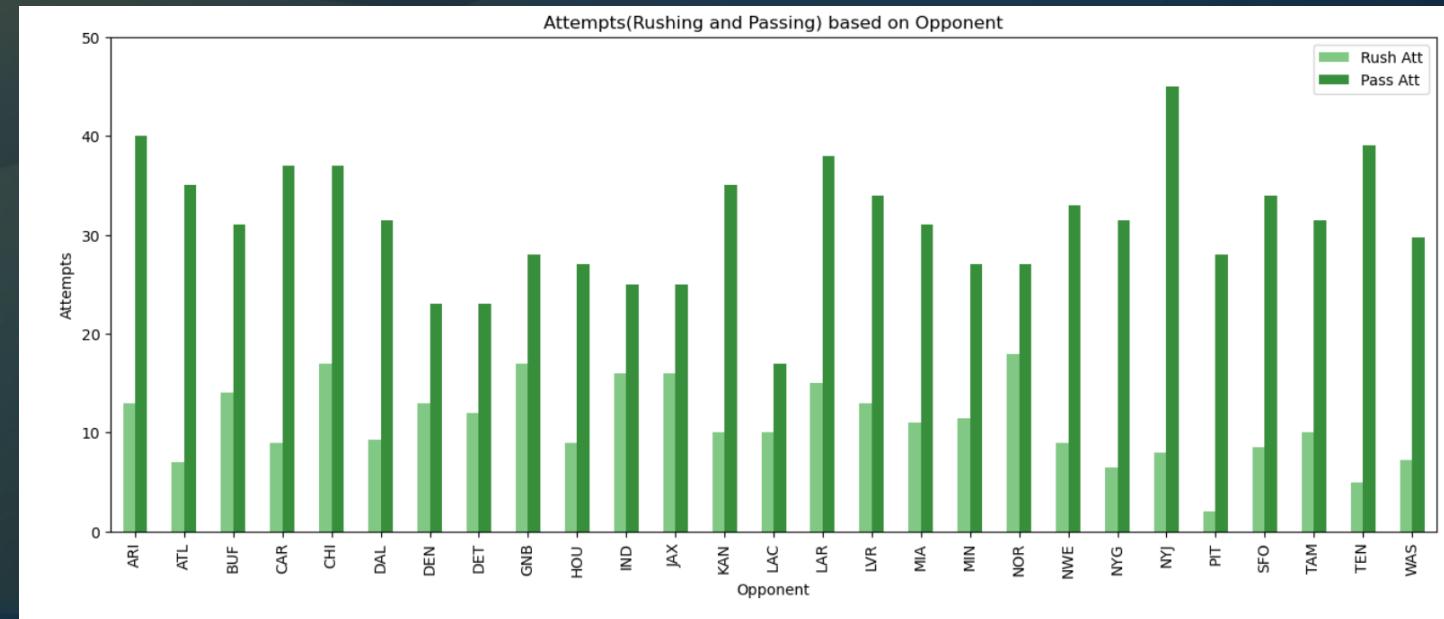
This confirms the hypothesis about rushing less and focusing on a better passing game.

After shifting his game from rush heavy to pass heavy he has won a larger portion of games as well, pushing forward the idea that this was the correct move.

Opponent

Using the same opponents from the Win Percentage section, we will see how the average passing attempts and rushing attempts compare for those teams

1. Washington: passing: 30 attempts , rushing: 7 attempts
2. New York Giants: passing: 32 attempts , rushing: 7 attempts
3. Dallas: passing: 32 attempts , rushing: 9 attempts
4. San Francisco: passing: 34 attempts , rushing: 9 attempts
5. Minnesota: passing: 27 attempts , rushing: 12 attempts



Trying to hybrid his play too much shows a conclusion of struggling to achieve a positive win percentage. Either have a passing heavy game or create a great rushing threat but both tend to stretch him too thin.

Details to this conclusion can be found on Analysis Insights document.

Opponent

Average Passing Yards by Opponents

1. Washington: 248 yards
2. New York Giants: 194 yards
3. Dallas: 258 yards
4. San Francisco: 244 yards
5. Minnesota: 263 yards
6. Tennessee: 380 yards (highest)
7. Green bay: 153 yards (lowest)

Average Rushing Yards by Opponent

1. Washington: 29 yards
2. New York Giants: 44 yards
3. Dallas: 42 yards
4. San Francisco: 51 yards
5. Minnesota: 46 yards
6. Green Bay: 157 yards (highest)
7. Pittsburgh: 10 (lowest)

Opp	Avg Pass Yards
TEN	380.0
CHI	315.0
LAR	303.0
ARI	289.0
PIT	285.0
NYJ	280.0
MIA	279.0
KAN	269.0
ATL	264.0
MIN	263.0
DAL	258.0
WAS	248.0
SFO	244.0
HOU	243.0
LVR	236.0
JAX	204.0
BUF	200.0
CAR	198.0
TAM	196.0
NYG	194.0
IND	190.0
DEN	178.0
DET	173.0
NWE	170.0
LAC	162.0
NOR	157.0
GNB	153.0

Opp	Avg Rush Yards
GNB	157.0
NOR	88.0
IND	86.0
DET	81.0
LAR	72.0
BUF	65.0
LAC	62.0
ATL	62.0
ARI	62.0
LVR	61.0
CHI	61.0
DEN	55.0
SFO	51.0
NYJ	47.0
MIN	46.0
NYG	44.0
DAL	42.0
KAN	38.0
JAX	38.0
NWE	37.0
TAM	36.0
CAR	30.0
WAS	29.0
HOU	23.0
MIA	21.0
TEN	12.0
PIT	10.0

HYPOTHESIS



Hypothesis

From the EDA, I propose a hypothesis to the success of Jalen Hurts

- For games in which he passes for more yards, he rushes for less and vice versa
 - Example: Green Bay from SQL Analysis
- If this is the case in some instances, what is the most successful combination of rushing and passing yards
 - Above average passing yards and Below average rushing yards
 - Above average passing yards and Above average rushing yards
 - Below average passing yards and Above average rushing yards
 - Below average passing yards and below average rushing yards

Hypothesis

Over/Under AVG	Games Count
302.0	22
172.0	24

Out of the 46 games that we are using, 22 of those games he has thrown above his average in passing yards and 24 of those games he has thrown under his average in passing yards.



Over/Under AVG	Games Count
66.0	9
26.0	13

Out of those 22 games where he threw over his average, 13 of them he has run under his average in rushing yards. This shows that when he is able to throw for more yards, he tends to run the ball less. Or when he can't run the ball very well in game, he tends to throw the ball more.

Over/Under AVG	Games Count
82.0	11
31.0	13

When hurts passing yards are below his average, we can see that the hypothesis stays true. His rushing averages both went up. This means that the more passing yards he is getting, the less rushing yards he is inclined to produce and vice versa.

But which is better? Which produces more wins? Rushing more and passing less or passing more and rushing less?

Hypothesis

Result	Total
0	4
1	9

Above average passing yards and below average rushing yards.

Hurts wins 69.2% of the time in this scenario.

Result	Total
0	3
1	8

Below average passing yards and above average rushing yards.

Hurts wins 72.7% of the time in this scenario.

Result	Total
0	3
1	6

Above average passing yards and above average rushing yards.

Hurts wins 66.7% of the time in this scenario.

Result	Total
0	3
1	10

Below average passing yards and below average rushing yards.

Hurts wins 76.9% of the time in this scenario.

Hypothesis Conclusion

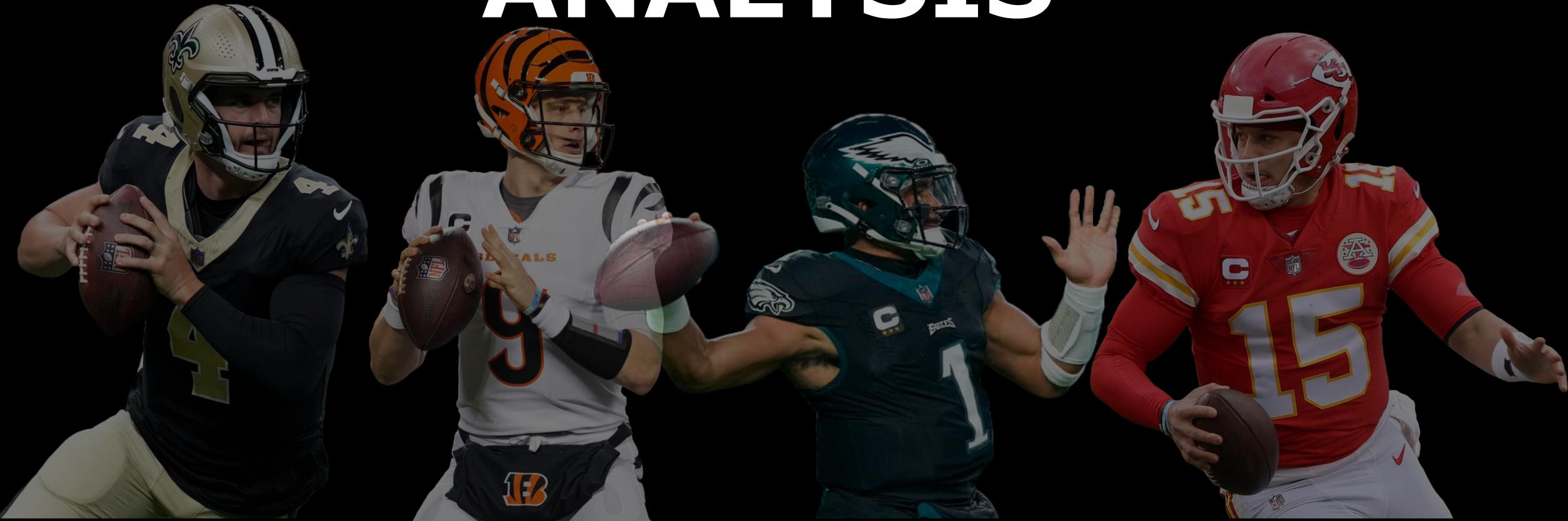
From these findings, it can be concluded that less passing yards seems to be the most productive play style for Jalen Hurts based off of the end goal of winning games.

Rushing yards has some say in the matter in that when he rushes less he tends to win more.

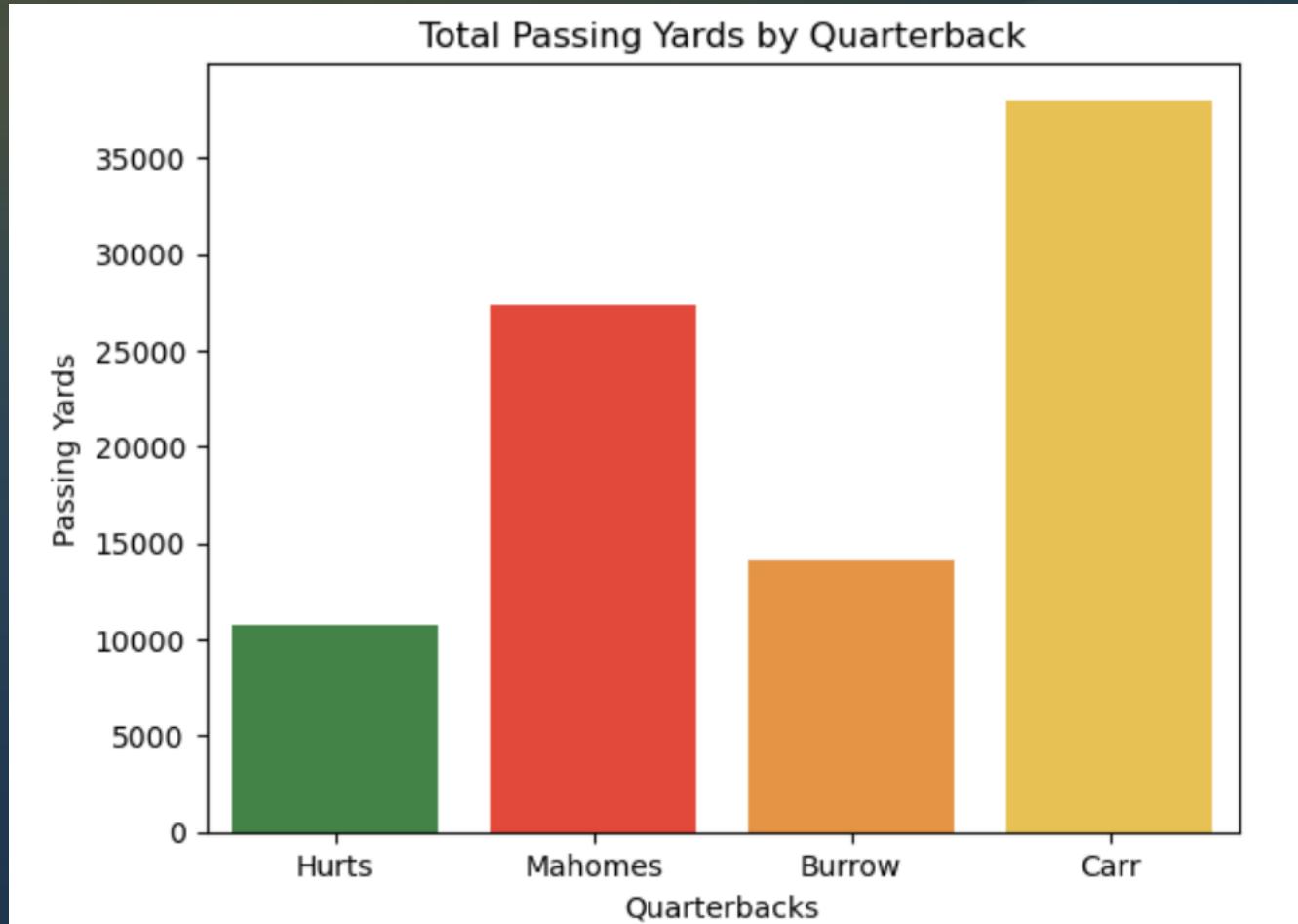
Hurts loses most frequently when he is both passing for above average yards and rushing for above average yards.

The conclusion here relates to the Anti-Hybrid Play Gameplan set a few slides back. The lowest win percentage came from games with both high rushing and passing yards.

COMPARATIVE ANALYSIS



Passing



Carr has the highest passing yards

Hurts has the lowest

Stat is swayed by time in the league. But Burrow has 400+ more passing yards than Hurts while being in the league for the same amount of time.

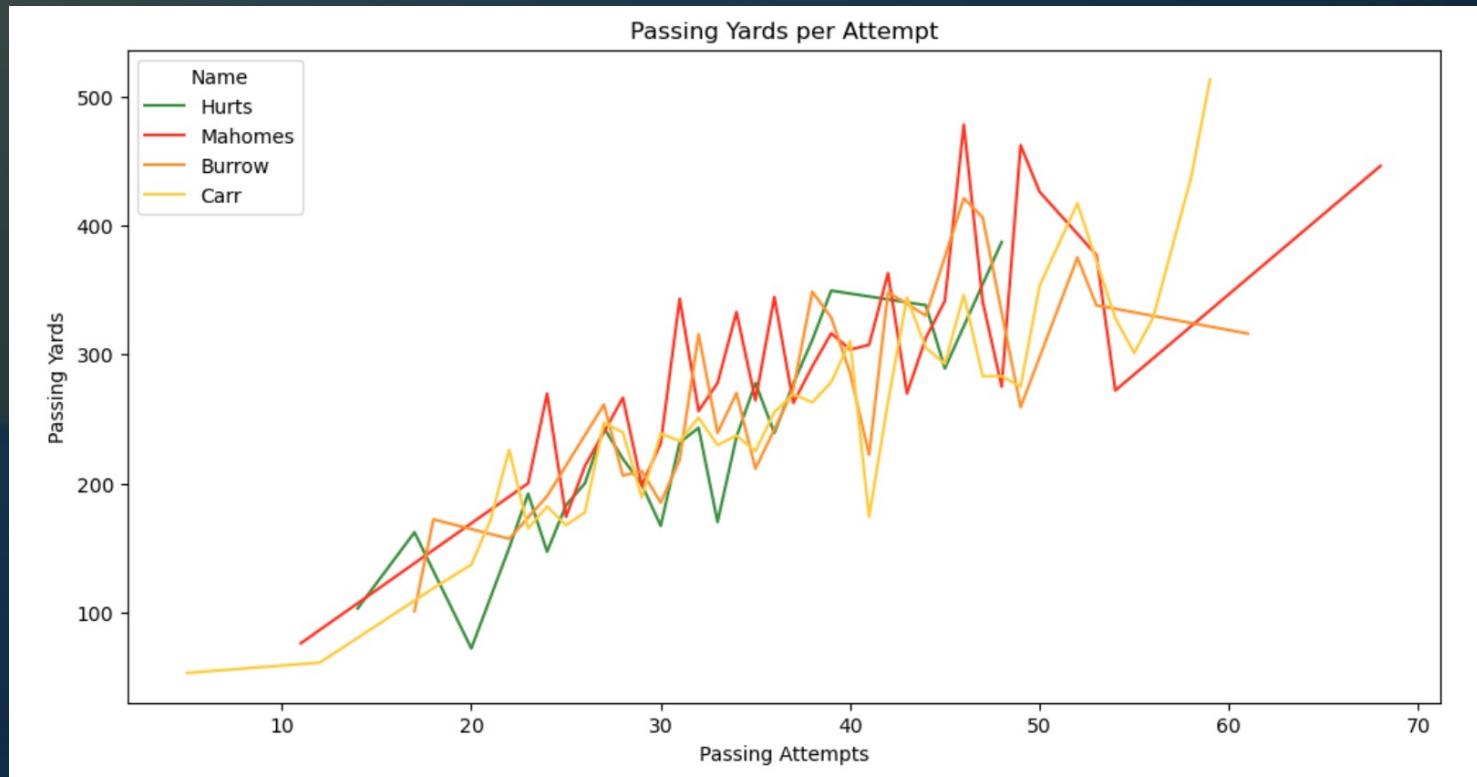
Passing

While Mahomes, Burrow, and Carr have all had at least one game where they threw 60+ attempts, Hurts has never broken the 50 attempts mark.

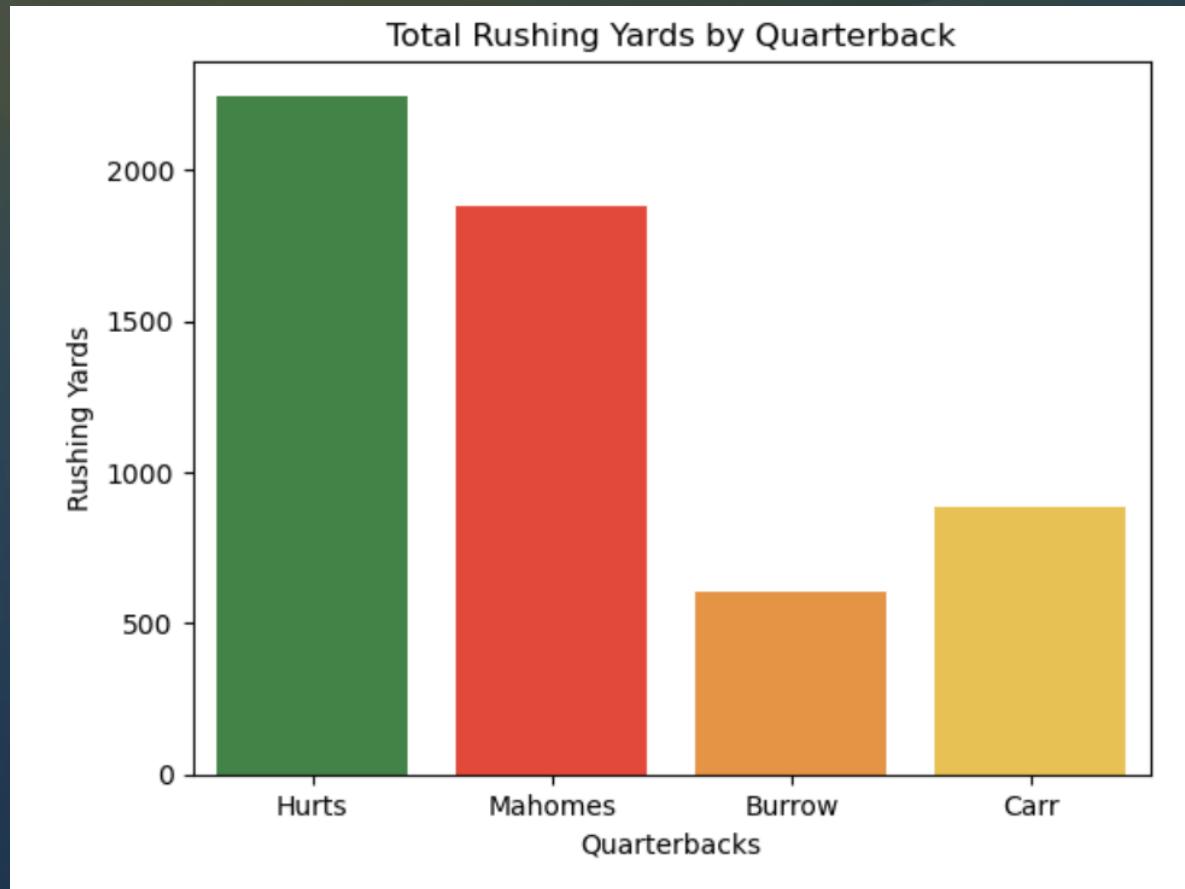
He is not as pass heavy as some of his elite peers.

By attempt, Hurts's passing yards stay equivalent to his peers

The reason for lack of high yardage is due to attempts rather than lack of passing skill.



Rushing



Hurts has the highest rushing yards

Burrow has the lowest

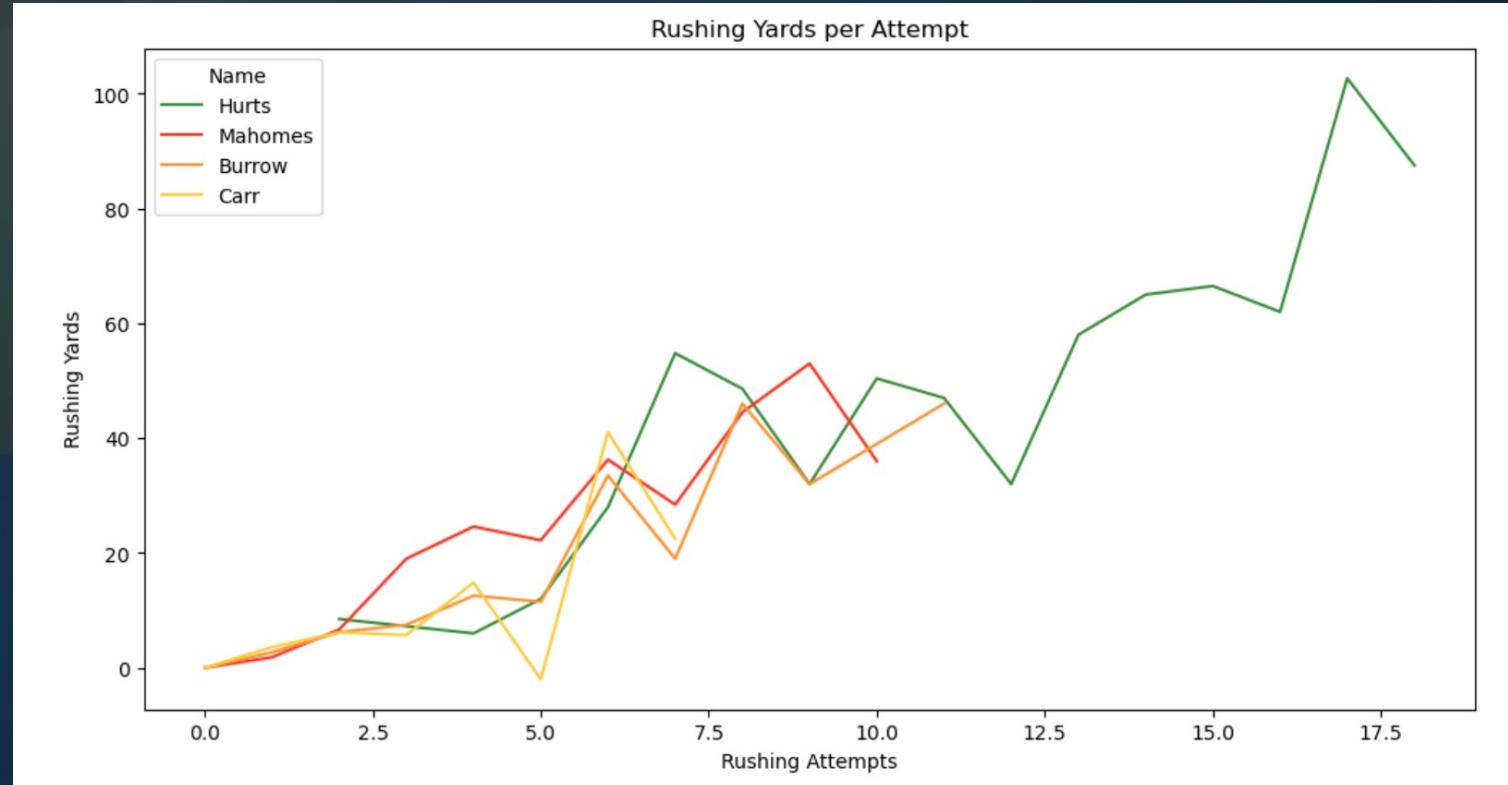
While stat is swayed with time in the league, it demonstrates an interesting conclusion. Hurts has rushed more than the other 3 quarterbacks while only being in the league for a fraction of the time.

Hurts is one of the highest rushing quarterbacks.

Rushing

Mahomes, Burrow, and Carr have never rushed more than 12 times in a game, while Hurts has rushed a maximum of 18 times.

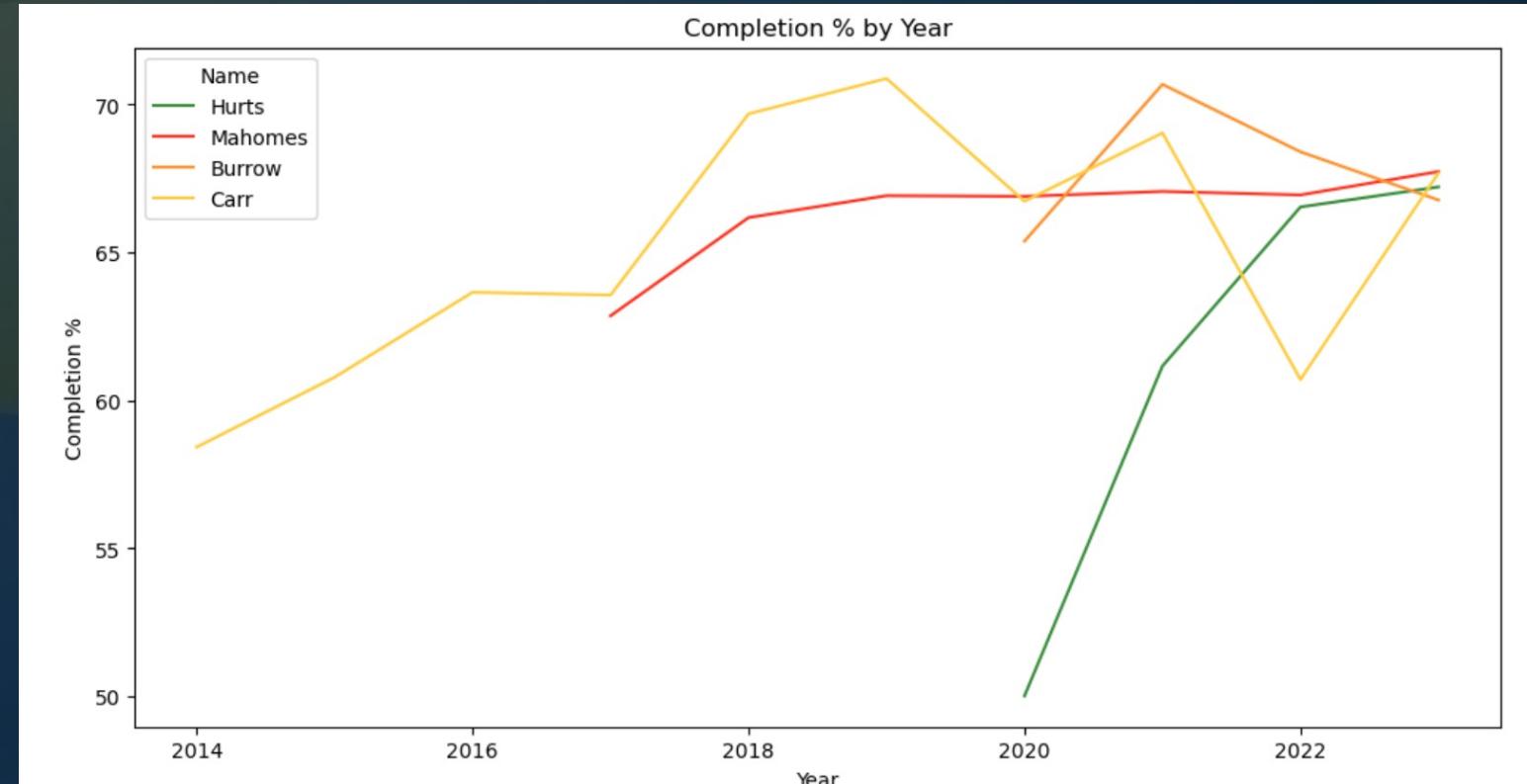
While he may not pass as much as the other quarterbacks, he outruns them significantly.



Completion Percentage

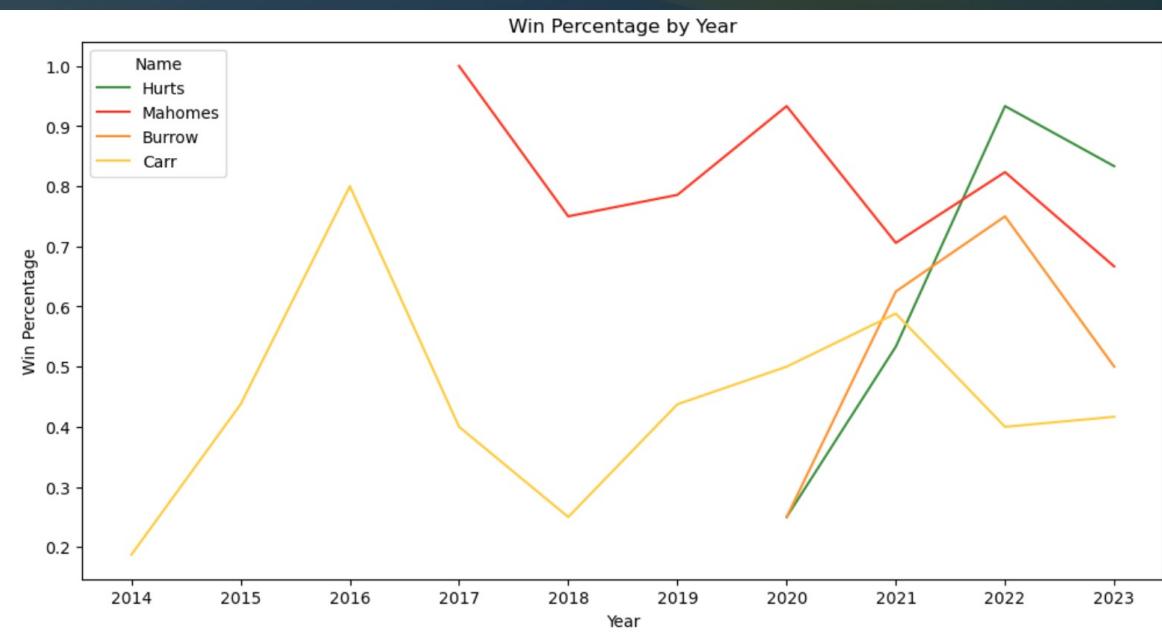
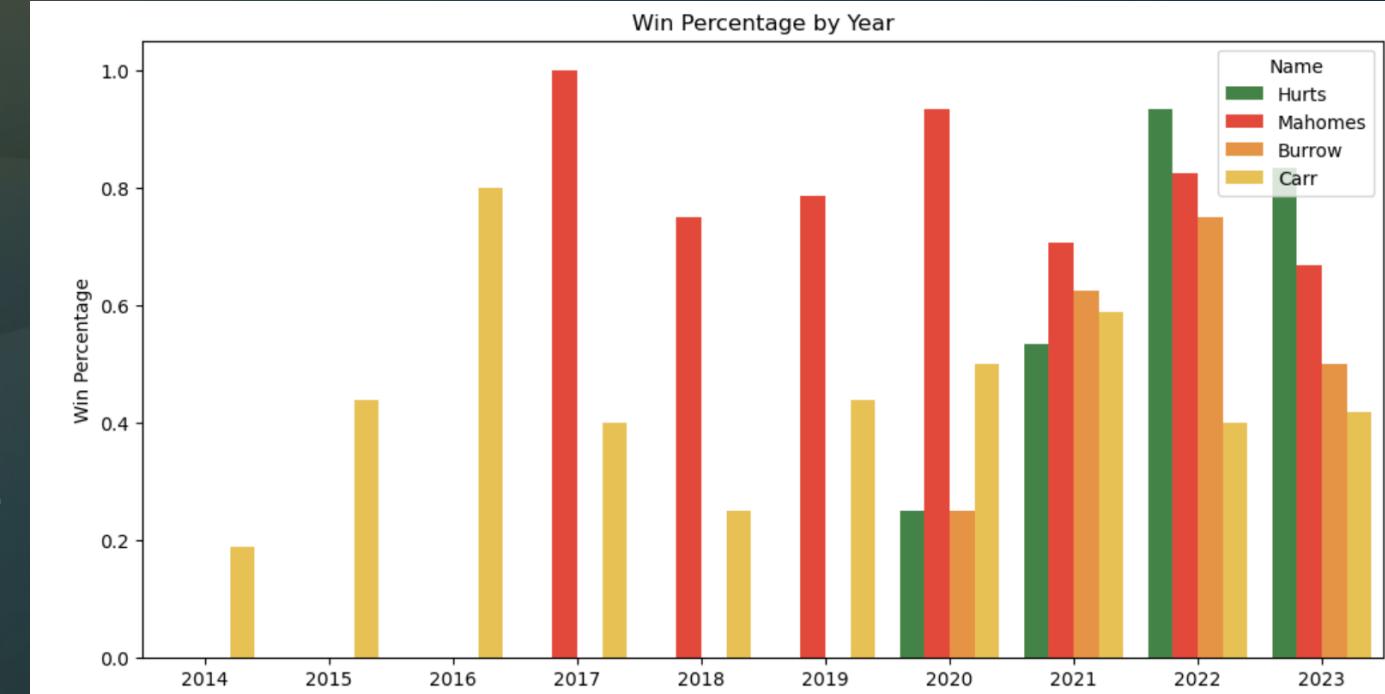
Hurts had the lowest completion percentage to start his career but has improved every year. Now having an equal percentage with the other quarterbacks.

1. Mahomes has had the most consistent completion percentage
2. Carr has fluctuated heavily
3. Burrow has dropped since 2021



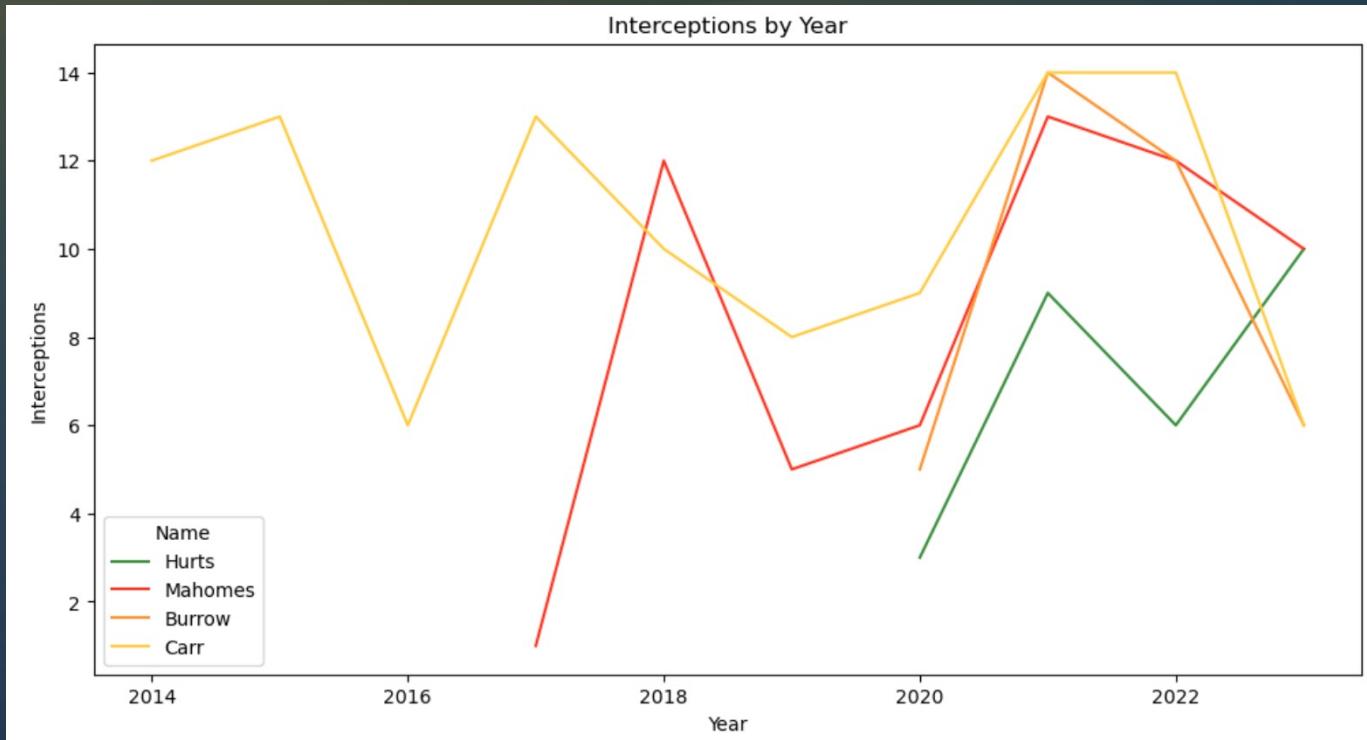
Win Percentage

In the past two years, Hurts has had the highest win percentage of any of the other quarterbacks



Every quarterback's win percentage has decreased in the past year. Hurts has had the fastest win percentage increase in the past 3 years.

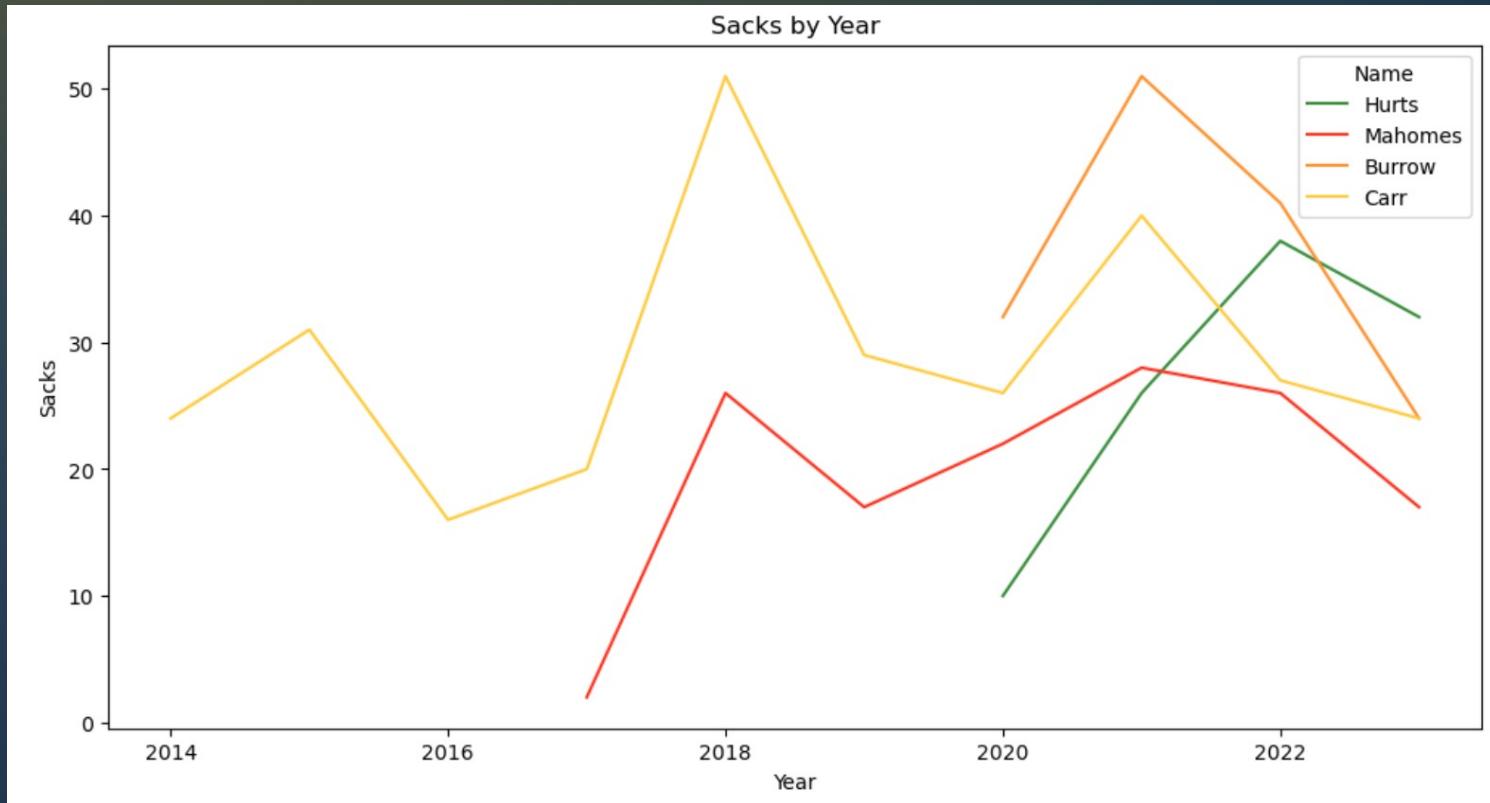
Interceptions



In the past year, every quarterback except for Hurts has decreased the number of interceptions thrown.

Hurts, who has thrown the least number of attempted passes, is leading in interceptions. Far too many turnovers.

Sacks



Hurts leads in sacks.

Needs to get rid of the ball or make a play quicker.

MACHINE LEARNING



Confusion Matrix

From the test data, all 4 models gave us the same predictions.

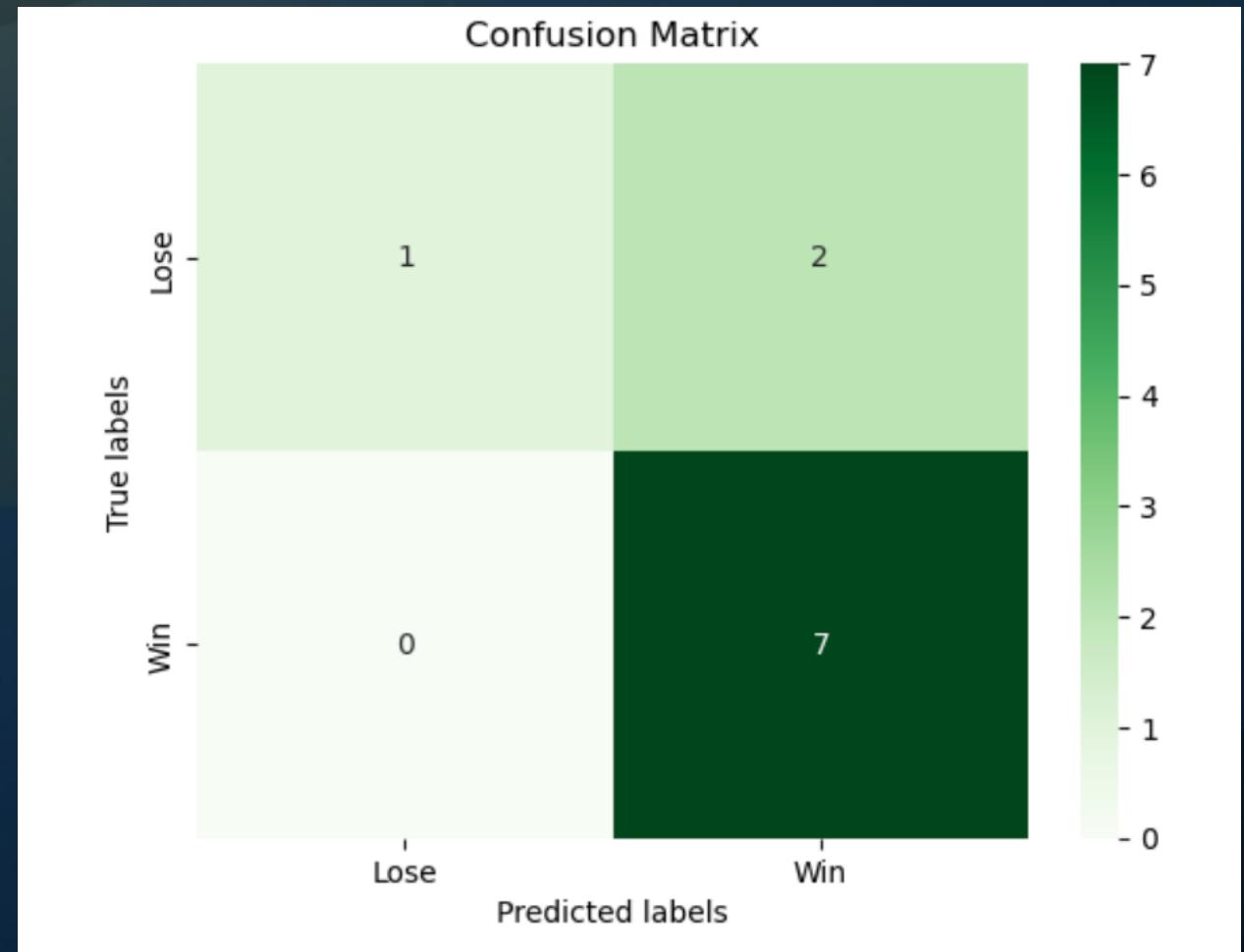
Here we can see that when Jalen wins a game, the prediction model is always correct (7/7).

But when Jalen loses a game, the prediction made is wrong 66% of the time.

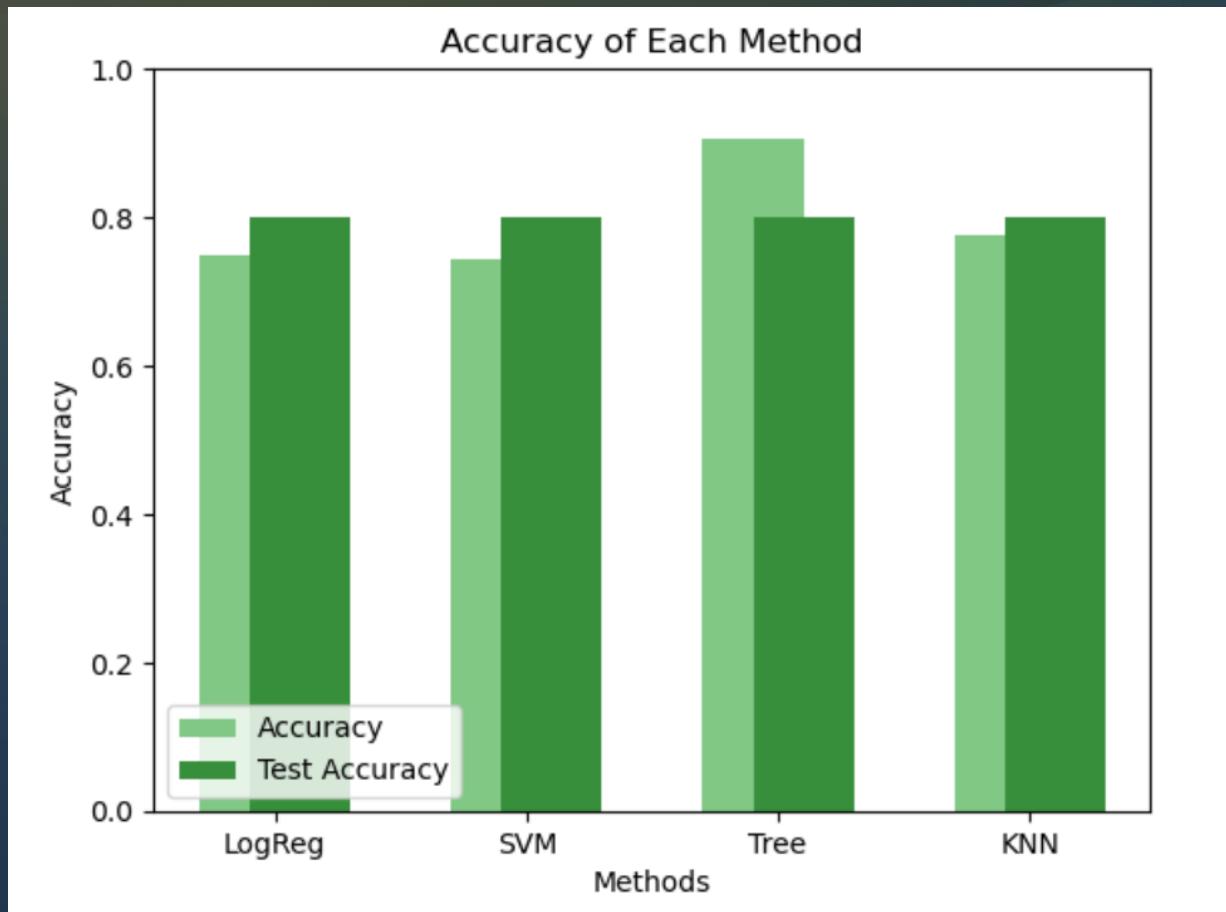
Only one loss was correctly predicted.

In total 8/10 predictions were accurate and 2/10 were not.

This is a problem of false positives.



Accuracy of Models



All models had a test data accuracy of 0.8 or 80%

Overall Accuracies:

Logistic Regression Model – 75%
Support Vector Machine – 74.375%
Decision Tree Classifier – 90.625%
K-Nearest Neighbors – 77.5%

RESULTS



Exploratory Data Analysis



Attempts at playing a hybrid game backfire greatest

Needs to choose either pass heavy or rush heavy per game

While he has one a greater percentage of games as he has shifted from an extremely rush heavy game to incorporating more passing, he needs to keep his passing numbers at a minimum

Pass heavy, at first thought to be a good thing, was shown to be negative in SQL analysis (Hypothesis Section)

Comparative Analysis

Positives

- By attempt, Hurts's passing yards stay equivalent to his peers, but he is not nearly as pass heavy as them.
- Hurts's rushing game is far superior to his peers.
- Being a double threat in this manner has set him apart from the other 3 quarterbacks.
- He has shown the greatest improvement from his time in the league in completion percentage
- He has had the highest winning percentage of any of the quarterbacks for the past 2 years.

Negatives

- In the past year, every quarterback except for Hurts has decreased the number of interceptions thrown.
- Hurts, who has thrown the least number of attempted passes, is leading in interceptions. Far too many turnovers.
- Hurts leads in sacks.
- Needs to get rid of the ball or make a play quicker. Must find a solution to this problem.

Machine Learning

Since every model recorded the same accuracy for test data and the same false positives, I am inclined to believe these 2 losses are outliers and should have been wins based on the understood data.

Other factors most likely made the team lose such as lack of defensive presence, missed scoring opportunities from other players, etc.

Decision Tree Classifier was the most accurate model and best model created.



Conclusion

Best strategy is to limit passing

- Due to the number of turnovers in interceptions and his high sack count, he must develop as a passer before implementing it to a greater degree.

He is unique compared to his peers

- Despite his short time in the league, he has rushed for more than any quarterback on this list. This is his biggest threat and advantage

Using a Decision Tree Classifier is the best model for predicting success

- With other factors added such as opponent statistics, defensive statistics, and total offensive statistics, the model would become increasingly more accurate

Thank you

