

NFL fan bases' general sentiment throughout the 2024-2025 season

[GitHub Link](#)

[Video Demo](#)

Introduction

Online sports discussion forums are often criticized for being wildly reactionary and fickle. Even more so in the National Football League (NFL), adopting a "what have you done for me lately" mentality, where aging and/or injury-ridden players are quickly moved on despite their past accomplishments. This can lead to fans quickly changing their opinion of their team depending on wins/losses, trades, signings, etc. The converse is also true, as NFL team management and owners often want to please their fans, who often only care about current team success. This could lead to short-sighted decisions being made, such as trading away future draft capital for an aging star player, which may be detrimental to the team overall. Thus, it is important to have some way to track how fans' opinions change over time with certain events. This project will specifically track the NFL fans' sentiment toward their team after every game throughout the 2024-2025 season.

Since we want NFL fans' opinions right after the games, we look for an online forum where fans post their opinions after their team's games. The perfect candidate for this is Reddit's post-game threads on Reddit on individual teams' subreddits (i.e., r/steelers). So we will scrape the comments from these Reddit threads and put the comments through VADER (Valence Aware Dictionary and sEntiment Reasoner) to determine whether the comments are positive or negative. Then we will calculate the ratio of positive and negative comments in each thread. This will be one part of our dataset. The other parts will come from the NFL games themselves. This includes the team, opponents, week, whether the team won or not, etc. On this dataset, we will use hierarchical clustering, data visualization, as well as other methods to identify features that determine fan sentiment.

The project will heavily involve the use of several techniques we learned in the course such as data visualization, web scraping, hierarchical clustering and text processing.

Motivation

Since the advent of social media, there has been a growing number of people whose main source of information is online forums such as Reddit. Thus, the overall sentiment of these forums can heavily shape a person's view of the outside world. By investigating only the sports forums, we have a unique opportunity to analyze how well online sentiments map to reality and also how these sentiments, thus public opinion, can change due to outside influences. An exciting and existing question this project can answer is the existence of "moral victory" and "Pyrrhic victory." These occur when a team loses a game but the fanbase is optimistic, and

vice versa. For example, a fanbase can be revitalized when their rookie quarterback nearly wins a game against a Super Bowl favorite. It is just one phenomenon we may see that goes against our preconceived notion of fan sentiment. John Naujoks, a data scientist, did a [similar project](#) to measure the Cleveland Browns fans' sentiment on Twitter day by day. This project will be more post-game focused and will heavily use Reddit comments as well as information about each game.

Method

Building the Dataset

The file "NFL 2024-25 Subreddit - Sheet1.csv" contains the post-game thread URL in each subreddit. For each URL, the comments were scraped following a similar method to [Scrapfly](#)'s tutorial "How to Scrape Reddit Posts, Subreddits and Profiles". HTTPX was used to request and retrieve the HTML from the subreddit. Parsel was used to parse the comments. Loguru displayed the progress of the data scrape. Asyncio improved the speed of scraping through asynchronous code. These retrieved comments are sorted by the comments that had the most upvotes. Through the comment, we recursively scraped each reply as its own comments. Once we have retrieved the comments, we will put each comment through VADER, which scores the comment's positive, negative, and neutral sentiment. The overall compound score will be used, which adds positive sentiment and subtracts the value of negative sentiments. This results in a score of a value between -1 and 1, where a positive value represents positive sentiment and a negative value represents negative sentiment. Any comments with a compound score less than -0.05 will be marked as a negative comment, and comments with a compound score of more than 0.05 will be marked as a positive comment. This is to remove neutral comments from the dataset as they are often unreliable. Then, from this set, we will calculate the ratio of positive and negative comments for each post. The "team_week_sentiment_long.csv" contains the dataset of the sentiment analysis for each team's games with columns of "PositiveRatio", "NegativeRatio", and "TotalComments", which represent the total number of comments scraped. Then the dataset was merged with other CSVs containing information on each NFL game that were built manually from information on [ESPN](#). This information contains booleans such as whether the team won or lost, home games, playoff games, division games, and whether the game was broadcast on Prime time. This also contains numerical data such as the "Margin of Victory", which is the many points the team won or lost by. The code to build this dataset is in "FetchandBuildData.ipynb".

Preprocessing

To reduce the dimensionality, the "NegativeRatio" and "TotalComments" were dropped. "NegativeRatio" was dropped because it is inversely related to "PositiveRatio." "TotalComments" was dropped because we noticed that, due to how

Reddit's API and its site work, we were not able to scrape all the comments from the threads. We also could not tell how the code chose which comments to scrape or not. Thus, we felt it was not a reliable data point to use.

For the boolean data point, we used one-hot encoding to convert the data to 0s and 1s. "Yes" or "W" was converted to "1" while "No" or "L" was converted to "0".

Data Analysis and Techniques

To get a sense of factors that contribute to fan sentiment, we will graph fan sentiment and analyze the 2024-25 season of the "Philadelphia Eagles." We will also graph a heatmap of fan sentiment for the entire 2024-25 NFL season to get a general view of the dataset.

To view the important features of our data, we will use hierarchical clustering and visualize the cluster using a dendrogram. We will use Ward's Method of clustering, which computes clusters to minimize variances. A correlation matrix was also used to see how our features relate to each other.

Lastly, we will filter and look through the data to look for any unexpected results that don't match with the rest of the data. This could be a "moral victory" or a "Pyrrhic victory," where a loss leads to positive fan sentiment or a win leads to negative fan sentiment.

Results and Discussion

The Philadelphia Eagles 2024-2025 season

Figure 1 shows how Eagles fan sentiment changed throughout the NFL season. The Eagles were expected to be Super Bowl contenders after being in the Super Bowl two years ago in 2024. After a disappointing 2023-2024 season, where the Eagles lost six of the last 7 games after starting the season 11-1, fans felt impatience and felt the team needed to show good results or a change was needed. Thus, we predict that wins were very important to fan sentiment. This is clearly shown in the dramatic drop in fan positive sentiment after the team lost in Week 2 and Week 4.

In Week 4, following a 17-point loss to the Buccaneers, one of the most upvoted comments stated: *"This game was a microcosm of every game we've watched the last two seasons. It's obnoxious. It's frustrating. And I don't think it'll change unless a new Head Coach is brought in (hope I'm wrong though)."* These comments achieved a compound score of -0.7269, showing a case where VADER correctly identifies negative sentiment. However, other comments did show weaknesses in VADER sentiment evaluation, especially in comments that require outside information that VADER was not trained on. For example, on the same thread, a comment reads, *"This season is just a continuation of last season,"* which scored a 0 in the compound score, meaning that VADER did not evaluate the comment, deeming it as fully neutral. However, with the context of the Eagles' last season,

that comment indicates negative sentiment. Further analysis of the comments does show cases where VADER incorrectly evaluation sentiment from the comments, but overall the performance is good enough to show trends in the data. We can see that wins and losses have a significant effect on fan sentiment, as a win in Week 5 significantly increases fan sentiment.

In Week 9, we see a decrease in sentiment despite a 5-point win against the Jacksonville Jaguars. This was largely due to complaints about the referees and the view that the Jaguars were an inferior team that shouldn't have won nearly. This shows another weakness in VADER, as negative sentiment in a thread may be negative sentiment directed towards the team. Week 14 mirrored Week 9, with the Eagles narrowly defeating the perceived weaker Carolina Panthers by 6 points. Despite the win, much of the negative sentiment was directed toward star quarterback Jalen Hurts. These two data points highlight that fan sentiment is influenced by more than just wins, including factors like the quality of the opponent, margin of victory, and external elements such as quarterback performance and officiating. This could be advantageous for a future, more sophisticated dataset.

The Philadelphia Eagles ignored fans' calls for change after their drop in fan sentiment and stayed course. They were rewarded with winning the 2025 Super Bowl in a blowout victory over the Kansas City Chiefs. This gives credence to the idea that organizations shouldn't give in to their fans' opinions as long-term plans and decisions pay off.

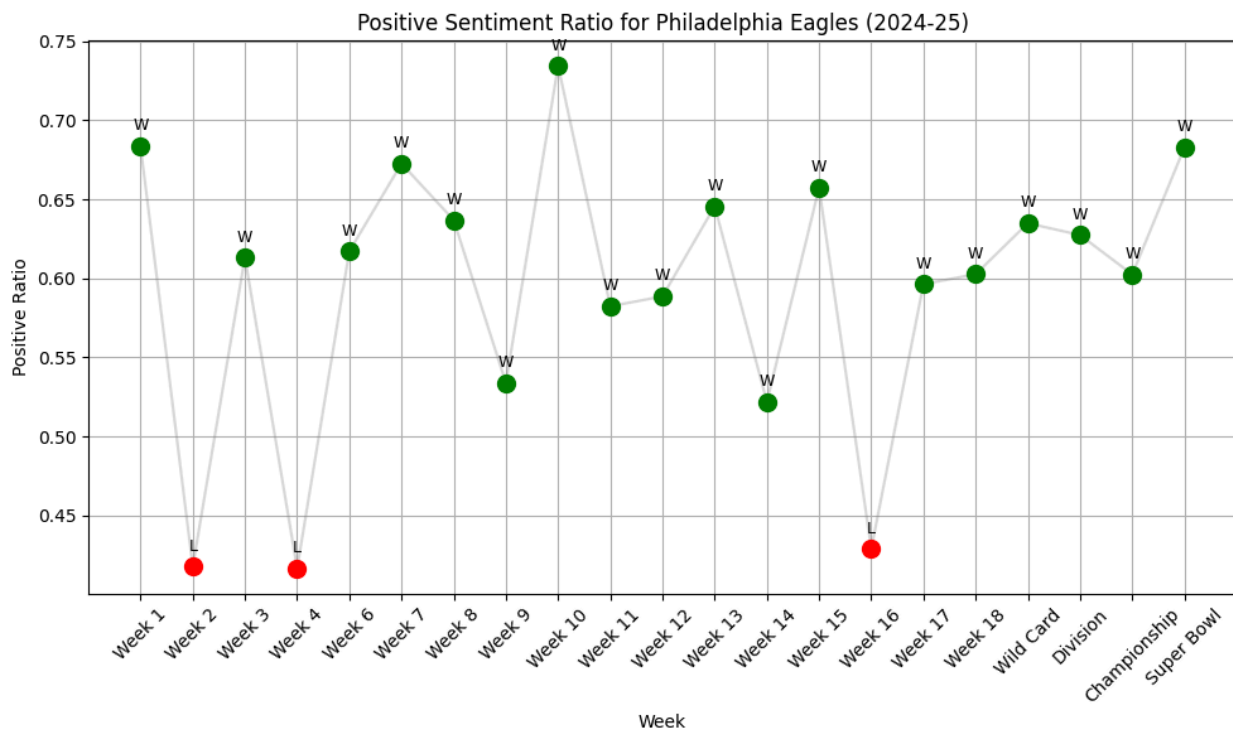


Figure 1: Positive Sentiment Ratio for the Philadelphia Eagles throughout the 2024-25 Season

Heatmap of the NFL 2024-25 season

Figure 2 shows fan sentiment for each NFL team throughout the 2024-25 season. White areas indicate missing data, usually because the team did not play due to a 'bye' (a scheduled week off with no game) or because they did not qualify for that round of the playoffs. The exception to this is Week 13, when r/Patriots deleted the post-game thread following a last-second loss to the Colts, resulting in no data being collected for that game. Overall, the mean fan sentiment is higher than 0.5, which could indicate a bias in VADER for positive sentiment or, in general, fans are happy with their NFL teams.

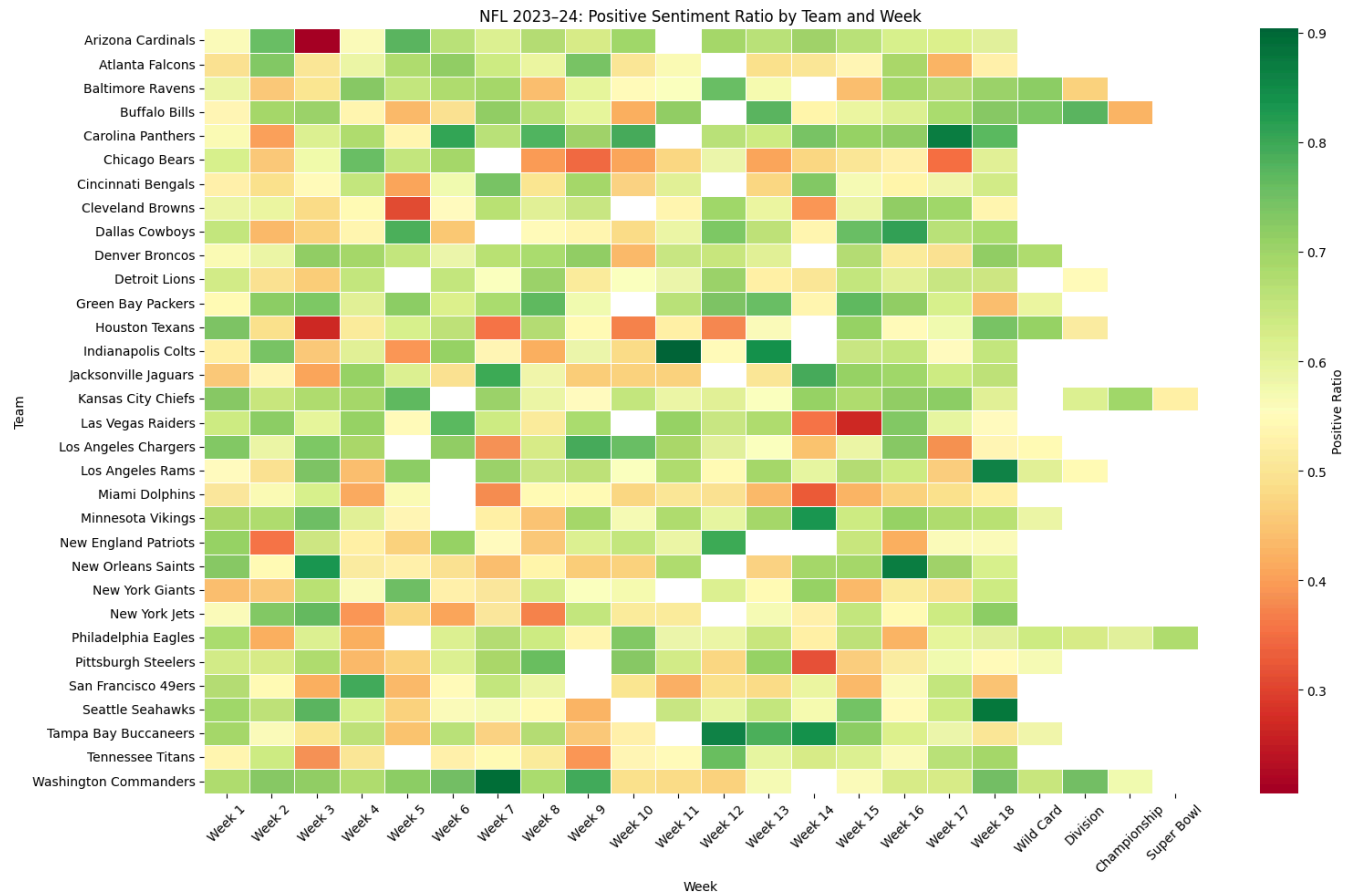


Figure 2: Heatmap Visualizing Positive Fan Sentiment by Team and Week Throughout the 2024–25 NFL Season

Correlation Analysis through Hierarchical Clustering

The clustering Dendrogram is shown in Figure 3. We can see 3 distinct clusters of data points. We can average out the values in each cluster, which is shown in Figure 4. Cluster 2 includes all the playoff games, which also happen to be prime time matchups (games broadcast individually without competing games airing simultaneously). This is not surprising it separated two of the features in the dataset (“Playoffs” and “Primetime”). Also, playoff games are usually different from the rest of the datasets. One reason is that some teams have young stars like the Denver Broncos in Bo Nix and are happy to be in the playoffs despite losing in the Wild Card round. This led to fan sentiments after a playoff loss to be relatively positive and similar to a playoff win. Another reason is that playoff games often match similar-strength teams, so blowouts are rare, which leads to small margins of victory. Cluster 1 contains all of the losses in the regular season, and Cluster 3 contains all of the wins. This illustrates that the cluster picked up on how playoff games were different, and also how wins and losses affected the dataset. From this clustering, we predict that wins/losses as well as the margin of victory heavily affect

fan sentiment. Margin of victory is, of course, related to wins/losses as a positive margin of victory will indicate a win, and a negative one will indicate a loss.

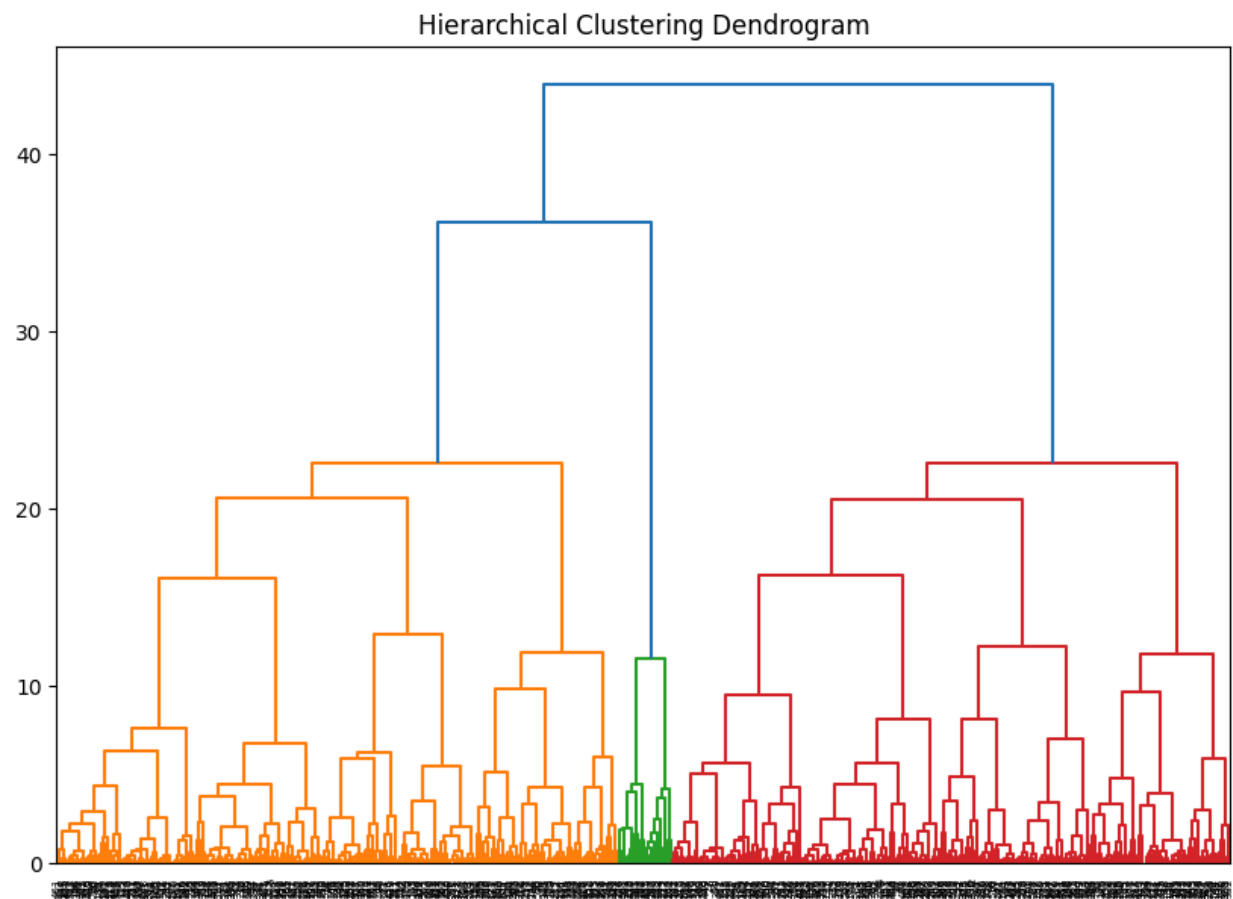


Figure 3: Dendrogram of Team Sentiment Patterns and Game Features Across the Season

	PositiveRatio	Div	Home	Playoff	Prime Time	Win-Lose	\
Cluster							
1	0.546189	0.350554	0.483395	0.0	0.254613	0.0	
2	0.613547	0.153846	0.461538	1.0	1.000000	0.5	
3	0.649463	0.356618	0.533088	0.0	0.253676	1.0	
	Margin of Victory						
Cluster							
1	-11.236162						
2	0.000000						
3	11.194853						

Figure 4: Means of Cluster Groups

Correlation Analysis through Correlation Matrix

Figure 5 shows the correlation matrix between the features of our dataset. As expected margin of victory correlates with wins. Both of these parameters correlate moderately with positive fan sentiments, with wins being the most important factor. Other factors may correlate more strongly with fan sentiment than just wins that the dataset does not cover. It may also be possible that other factors contribute differently to the dataset depending on whether the team won or lost the game. For example, fan sentiment following a win may remain high regardless of the opponent's strength. In contrast, a loss to a weaker team tends to generate significantly more negative sentiment than a loss to a strong opponent. Interestingly, there is a small correlation between home games and wins/margin of victory. This contributes to the NFL narrative of home-field advantage. This is attributed to favorable referees and fans making noises to attempt to disrupt the away team's offense from running their plays correctly.

Unexpected Data Points

Figure 6 shows candidates for “moral victories,” games where fan sentiment is positive after a loss. Examples of this are shown in the Panthers and Saints games, where fans know their team isn't very good and are relatively happy after a loss. It also could be that Panthers and Saints subreddits are less popular and thus attract more dedicated fans to the threads, who will support their team through losses. They could also be happy that their losing season is about to be over. The exception to this is the Los Angeles Rams, who were Super Bowl contenders. The high fan sentiment in Week 18 is largely attributed to the Rams resting their starters, aiming for a more favorable playoff matchup against the Minnesota Vikings rather than the Washington Commanders. Fans also expressed amusement at their division rival for barely securing a win against a weakened lineup.

Figure 7 shows a possible “Pyrrhic victory” where teams won, but fan sentiment is low. This is true with the Bills, Titans, and Dolphins games, where fans were not happy about their team's performance despite the win. The Steelers' game appears to have drawn negative sentiment not toward their own team, but rather against their rivals, the Browns. Many of the comments focused on Browns star defensive end Myles Garrett, particularly criticizing him for shushing the crowd and then losing. The sentiment with the Chargers' game is confusing, as many of the comments on the thread are positive. It could be that the data-scraper malfunctioned and took a bad sample of comments, which influenced the sentiment. This corroborates with the fact that only 52 comments were scraped and used.

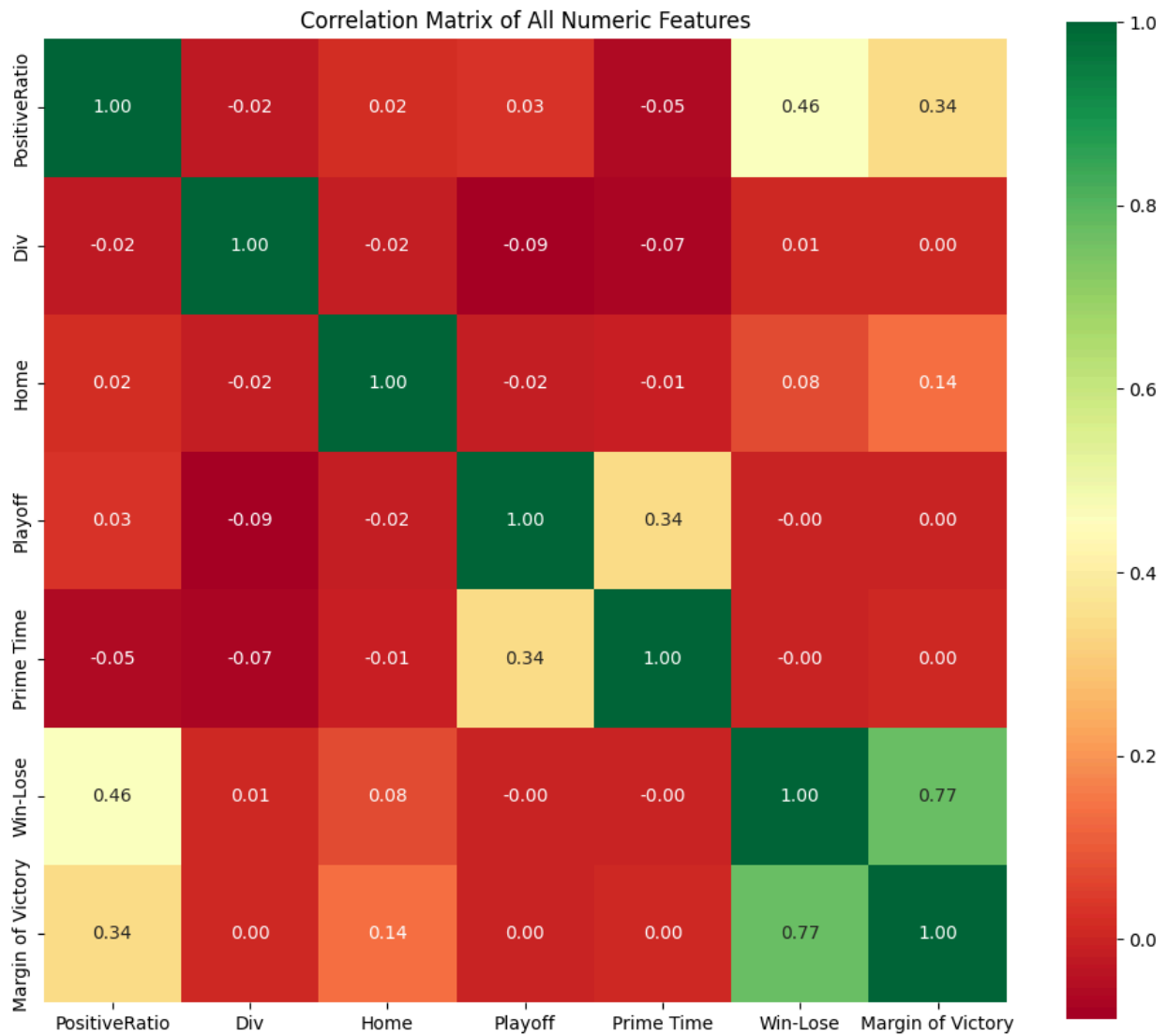


Figure 5: Correlation Matrix of the Dataset's Features

Top Losses by Positive Sentiment							
	Team		Week	PositiveRatio	NegativeRatio	TotalComments	
242	Carolina Panthers		Week 17	0.867995	0.132005	803	
418	New Orleans Saints		Week 16	0.866352	0.133648	636	
563	Los Angeles Rams		Week 18	0.861214	0.138786	807	
306	New Orleans Saints		Week 3	0.831407	0.168593	2076	
98	Carolina Panthers		Week 6	0.807560	0.192440	291	
	Div	Home	Margin of Victory		Opponent	Playoff	Prime Time \
242	Yes	No	-34.0		Tampa Bay Buccaneers	No	No
418	No	No	-34.0		Green Bay Packers	No	Yes
563	Yes	Yes	-5.0		Seattle Seahawks	No	No
306	No	Yes	-3.0		Philadelphia Eagles	No	No
98	Yes	Yes	-18.0		Atlanta Falcons	No	No

Figure 6: Top 5 Most Positive Fan Sentiment after a loss

Bottom Wins by Positive Sentiment							
	Team		Week	PositiveRatio	NegativeRatio	\	
165	Buffalo Bills		Week 10	0.420086	0.579914		
515	Tennessee Titans		Week 9	0.391061	0.608939		
255	Los Angeles Chargers		Week 17	0.383019	0.616981		
217	Miami Dolphins		Week 14	0.326923	0.673077		
221	Pittsburgh Steelers		Week 14	0.315399	0.684601		
	TotalComments		Div	Home	Margin of Victory		Opponent Playoff \
165	926		No	No	10.0		Indianapolis Colts No
515	895		No	Yes	3.0		New England Patriots No
255	530		No	No	33.0		New England Patriots No
217	52		Yes	Yes	8.0		New York Jets No
221	1078		Yes	Yes	13.0		Cleveland Browns No
	Prime Time		Win-Lose				
165	No		W				
515	No		W				
255	Yes		W				
217	No		W				
221	No		W				

Figure 7: Top 5 Most Negative Fan Sentiment after a Win

Conclusion

This project was overall successful in building and analyzing a dataset with features to correlate fans' sentiment with their NFL team over time. We found that wins and margin of victory correlated with positive fan sentiments. Surprisingly, we

found evidence that suggested a home-field advantage in our correlation matrix. We also found data points that match the characteristics of “moral victories” and “Pyrrhic victory.”

Possible improvements and areas of future studies include using a better method to scrape data, as not all comments were scraped from each thread. This could mean possible integration with Reddit’s API. Improvements could be made to VADER to analyze for NFL-specific comments. We could also build more sophisticated datasets that include the strength of opponents and the team's overall performance on offense and defense. Improvements in these areas would greatly improve our analysis capabilities and draw more accurate conclusions from the data.