“Signal or Noise: Measuring NFL Team Hype Through Reddit Sentiment Analysis”

A thesis presented to the Department of Statistics and Data Science at Yale University in partial fulfillment of the requirements for the degree of Bachelor of Arts

Abhinav Karthikeyan

Advised by Robert Wooster

April 21, 2025

Abstract

The NFL is arguably the most popular sports league in America, with a massive digital presence that extends far beyond the field. Reddit's r/nfl community alone boasts over 12 million users, making it a valuable resource for understanding fan sentiment in real time. This project explores how NFL teams are perceived on Reddit during the 2024 season and whether that online sentiment has any measurable impact on real-world outcomes. We collect and analyze Reddit comments from the duration of the 2024 NFL season mentioning all 32 NFL teams, using three natural language processing tools—VADER, NRC, and AFINN—to assess the positivity or negativity of fan sentiment for each team. We then investigate whether these sentiment scores are predictive of key indicators of team performance and value, including win percentage, average stadium attendance, and franchise valuation. Finally, we explore the concept of "overhype"—instances where teams receive disproportionately positive sentiment relative to their on-field results using fixed effects modeling and k-means clustering. Our analysis offers insight into the influence of online fan discourse in shaping narratives around NFL teams and its potential role in driving engagement and valuation.

Introduction

The NFL stands out as America's most popular sports league. With only 17 regular-season games, each contest carries significant weight, keeping fans highly engaged throughout the season. Due to a limited amount of games (even when factoring in the postseason), this boosts demand and allows the league to focus nationally, with marquee games on Thursday, Sunday, and Monday nights. In 2023, the NFL claimed 96 of the top 100 most-watched TV broadcasts in the U.S., including all 56 of the top-rated sporting events (Crupi). Its $110 billion media rights deal with conglomerates including CBS, Fox, NBC, ESPN and Amazon until 2033 underscore how valuable NFL content is to broadcasters. Streaming has only expanded this reach, with Amazon's Thursday Night Football and ESPN's record-setting Monday Night Football driving double-digit viewership increases (Reuters). NFL games also benefit from their consistent drama and built-in fan bases, stemming from city roots. The result is a reliable, high-stakes entertainment product that networks can confidently invest in. From the Taylor Swift-Travis Kelce romance to the New Orleans Saints' revival post Hurricane Katrina, the NFL remains deeply embedded in American culture.

The NFL has 32 teams split into two conferences—AFC (American Football Conference) and NFC (National Football Conference)—with each team playing 17 regular season games. The top 7 teams from each conference make the playoffs: 3 wild card games, then the divisional round, then the conference championships. The winners of each conference face off in the Super Bowl to decide who wins the title.

In January 2025, Backlinko estimated that Reddit had 1.1 billion monthly unique visitors. With over 12 million users, r/nfl is an active and passionate community that offers a constant stream of fan sentiment in real time, making it a valuable space for analysis. Unlike Twitter, where posts are limited in length and often driven by how "viral" they are, Reddit encourages longer, more thoughtful discussions that can reveal deeper fan insights. r/nfl also features structured weekly threads—like game day discussions and post-game reactions—that make it easier to find relevant, in-depth commentary tied to specific moments or matchups.

There is a growing body of literature that delves into the predictive power of social media sentiment in sports outcomes. Our project contributes to this literature by focusing on Reddit-based sentiment analysis for NFL games. Prior studies have shown that Twitter sentiment can rival traditional statistical models in predicting NFL outcomes (Sinha et al.), and that public opinion expressed on platforms like Twitter and blogs can identify profitable betting strategies (Hong, Skiena, et al.). Other work has extended sentiment analysis to evaluate individual performance in the NBA (Li et al.) and to forecast English Premier League results with higher returns than odds-based betting (Schumaker et al.). However, most of this literature has centered on Twitter and focused either on team-level or player-level analysis. By leveraging Reddit, our study adds a novel data source to the conversation and explores whether its distinct sentiment dynamics could offer different predictive insights. In doing so, we contribute new evidence to the role of crowd-based sentiment to forecast various sports outcomes and potentially expand how we think about sports betting strategies.

For our analysis, we sourced data from Reddit using the praw Python package, which provides access to Reddit's API. We focused on the r/nfl subreddit and specifically targeted posts and comments related to individual NFL teams by querying team names in post titles (e.g., "Cincinnati" or "Bengals"). For each relevant post, we scraped up to 50,000 comments that mentioned the team, filtering them for relevance by checking if the team name appeared in the comment body. These comments were then stored in structured CSV files for downstream sentiment analysis.

We applied three widely-used sentiment analysis models—AFINN, NRC, and VADER—to the collected comments to quantify public sentiment toward each team. The AFINN model assigns a numerical sentiment score to each word (ranging from -5 to +5) and computes a total sentiment by summing the values of words in a comment. The NRC lexicon goes further by tagging words with one or more of eight basic emotions (like joy, anger, or fear) in addition to positive or negative sentiment, allowing us to assess not just polarity but also emotional tone. Lastly, VADER (Valence Aware Dictionary and sEntiment Reasoner) is a rule-based model specifically attuned to social media text, capable of interpreting sentiment intensity based on punctuation, capitalization, degree modifiers (e.g., "very"), and emoticons. It outputs four metrics: positive, neutral, negative, and a compound score that summarizes overall sentiment. By using these three models in tandem, we captured different views of fan sentiment surrounding each team, which we then used to explore its relationship with real-world implications.

After attaining our three types of sentiment scores, we use Principal Component Analysis to come up with a weightage for each of the three sentiment scores to be applied in order to create a composite sentiment score. We use the composite score in a fixed effects model and baseline regression model to further understand "overhype" and see if prevailing Reddit sentiment has any effect on on-the-field performance. Furthermore, we will use K-means clustering to identify any trends with performance metrics that may be key in identifying sentiment related patterns. In tandem, the regressions and clustering will help us determine what sort of outcomes realtime r/nfl sentiment may have on on-the-field.

Part I: Data Exploration, Visualization, and Manipulation

Data Exploration:

As mentioned previously, for each of the 29 NFL teams in our study, we scraped fan comment data from Reddit, specifically the r/NFL subreddit. Using the *PRAW* package in Python, we collected thousands of fan-submitted comments across the regular season. Each team's data was stored in a CSV file representing the comment thread for the entire year, with each row corresponding to an individual comment. Due to Reddit rate limiting, we excluded three teams from the analysis: the Jaguars, Lions, and 49ers.

We began by loading all 29 team CSVs into R and consolidating them into a single data frame, adding a *team* column to indicate the origin of each comment. The raw data contained a column for the comment text and some metadata (e.g., timestamp, commenter), but we focused primarily on the text field. We converted all comments to lowercase and removed non-ASCII characters to ensure uniformity across teams. We then tokenized each comment using the `unnest_tokens()` function from the *tidytext* package to transform the data frame from one row per comment to one row per word. This allowed us to perform sentiment analysis at the word level, as most lexicons rely on individual words to classify sentiment.

We employed three sentiment analysis lexicons: AFINN, NRC, and VADER. The AFINN lexicon assigns numerical scores ranging from –5 to +5 to individual words, capturing positive and negative sentiment intensity. We merged the tokenized comments with the AFINN lexicon using a left join on the word column and then calculated the mean sentiment score for each team by grouping on the *team* column and averaging the AFINN scores. The NRC lexicon classifies words into ten emotion categories (such as joy, anger, trust, and fear) along with two general sentiment labels: positive and negative. We again joined on the word column and used `count()` to compute the frequency of each sentiment/emotion category for every team.

For the VADER analysis, we used the `vader_df()` function from the vader package, which computes compound sentiment scores for each comment. Since VADER operates on full sentences rather than isolated words, we applied it to the original comment column prior to tokenization. We then took the average compound score for each team to get an overall VADER-based sentiment rating. The VADER lexicon is particularly useful in online sports commentary, as it is designed to capture informal tone, slang, and sarcasm, which is highly featured in online fan discussion.

After applying all three sentiment lexicons, we created a final data frame with three columns of interest per team: average AFINN score, average VADER compound score, and NRC-derived counts for each of the ten emotions. This allowed us to compare teams not only by general sentiment but also by specific emotional undertones in fan discourse.

To visualize our findings, we used *ggplot2* to create bar plots of mean sentiment scores by team and radar plots to compare emotional fingerprints across teams. Further, we used *factoextra* to create k-means clustering plots of composite score against many performance indicators. We also conducted a correlation analysis between sentiment scores and team performance metrics such as win-loss record and playoff qualification. Finally, we saved the cleaned and annotated comment-level data and team-level sentiment summaries for downstream modeling and regression analysis in later sections of the project.

Data Visualization:

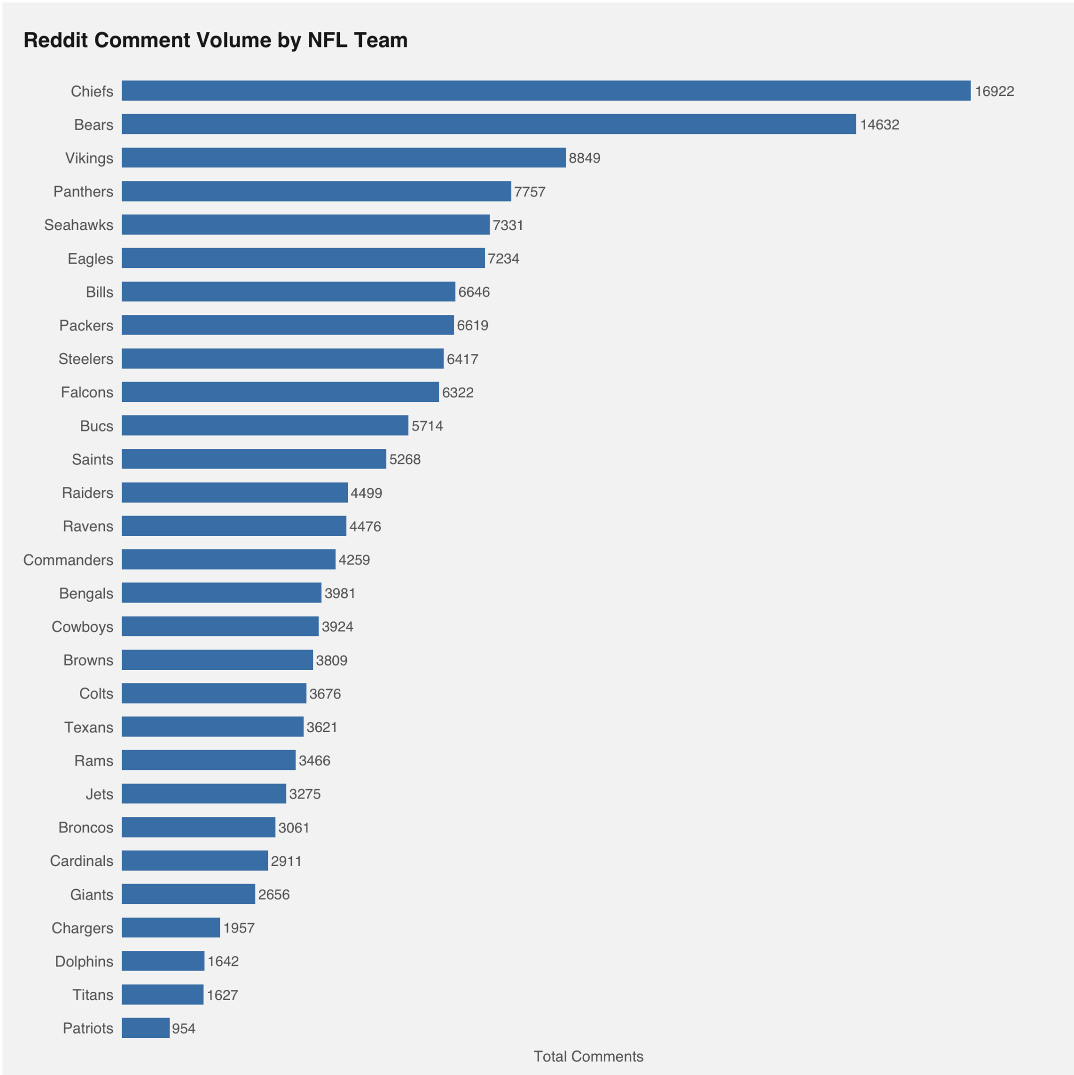

Figure 1: Reddit Comment Volume by NFL Team

Fig 1 shows how much Reddit comment volume was generated for each NFL team. The results above make sense as we make sense of them in the context of the 2024 NFL season. Given the pressure on the Chiefs for being on the precipice of a "three-peat" or the loads of commentary on hyped rookie Caleb William's flub of a season. On the flip side, the Patriots and Titans didn't have either "star" worthy moments or notorious discussion during disastrous seasons for both of the franchises resulting them to be in the top four picks of the 2025 NFL draft.

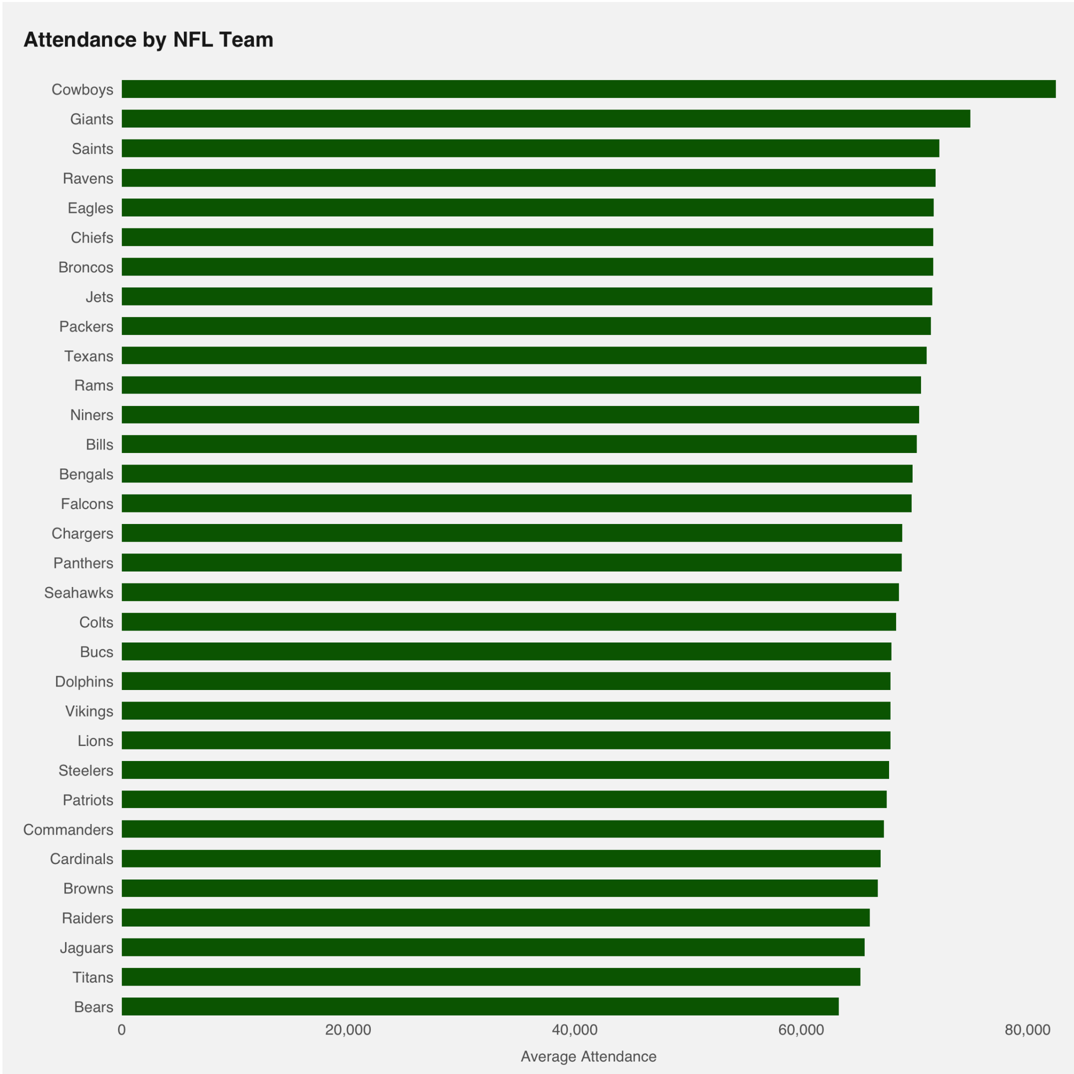

Figure 2: Attendance by NFL Team

Fig 2 shows average attendance by NFL team in season. We start to see an interesting trend as both the Bears / Titans lacked in comments generated and average attendance. Large media markets or notorious football teams (Cowboys, Giants) rank at the top here which makes sense.

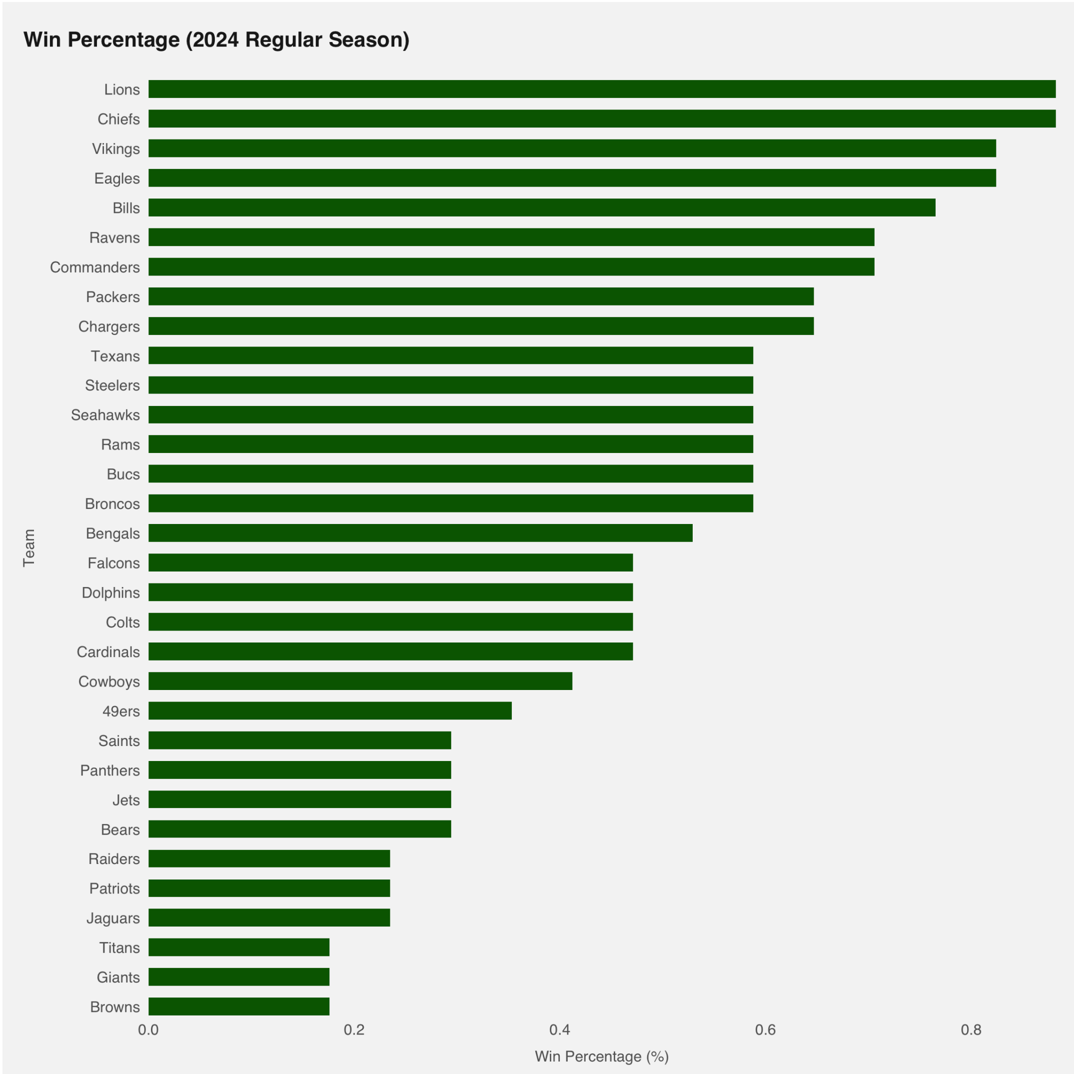

Figure 3: Win Percentage by NFL Team

Fig 3 shows win percentage by NFL team in season. We also continue to see that the Titans / Patriots are in the bottom echelon of teams here and there continues to be a trend. Chiefs are in the upper echelon of winning, which is something to note due to their presence at the top of comments generated.

Part II: Modeling/Analysis

AFINN Lexicon:

Developed by Finn Årup Nielsen, the AFINN lexicon is probably the simplest tool for sentiment analysis. It works by assigning a score between -5 and +5 to individual words, where negative numbers reflect negative sentiment, positive numbers reflect positive sentiment, and zero means the word is neutral. By adding up the scores of the words in a piece of text, you can get a sense of the overall emotional tone. AFINN uses a predefined list of words with these sentiment values, making it a straightforward method which is especially popular for analyzing the Reddit Comments we have compiled.

NRC Lexicon:

The NRC lexicon is a set of word lists that link words to sentiment, thus being popular for both sentiment analysis and text mining. It maps words to emotions such as joy, sadness, anger, fear, and surprise, or general sentiments like positive or negative. Built through manual annotation, the lexicon contains a large number of English words. It's widely used for detecting emotions in text and building apps that can understand emotional tone.

VADER Lexicon:

VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon designed for sentiment analysis, but mainly effective on social media posts and other informal texts. It uses a built-in dictionary that maps words to sentiment scores and applies rules to capture the intensity of sentiment, with enough complexity to account for "informal text" like slang and emojis. VADER outputs four key sentiment metrics: positive, negative, neutral, and a compound score, which summarizes the overall sentiment on a scale from -1 (most negative) to +1 (most positive). A compound score above 0.05 indicates positive sentiment, below -0.05 indicates negative sentiment, and anything in between is considered neutral. Its accuracy with short, informal texts make VADER a popular choice for real-time sentiment analysis, which is probably most helpful for our dataset.

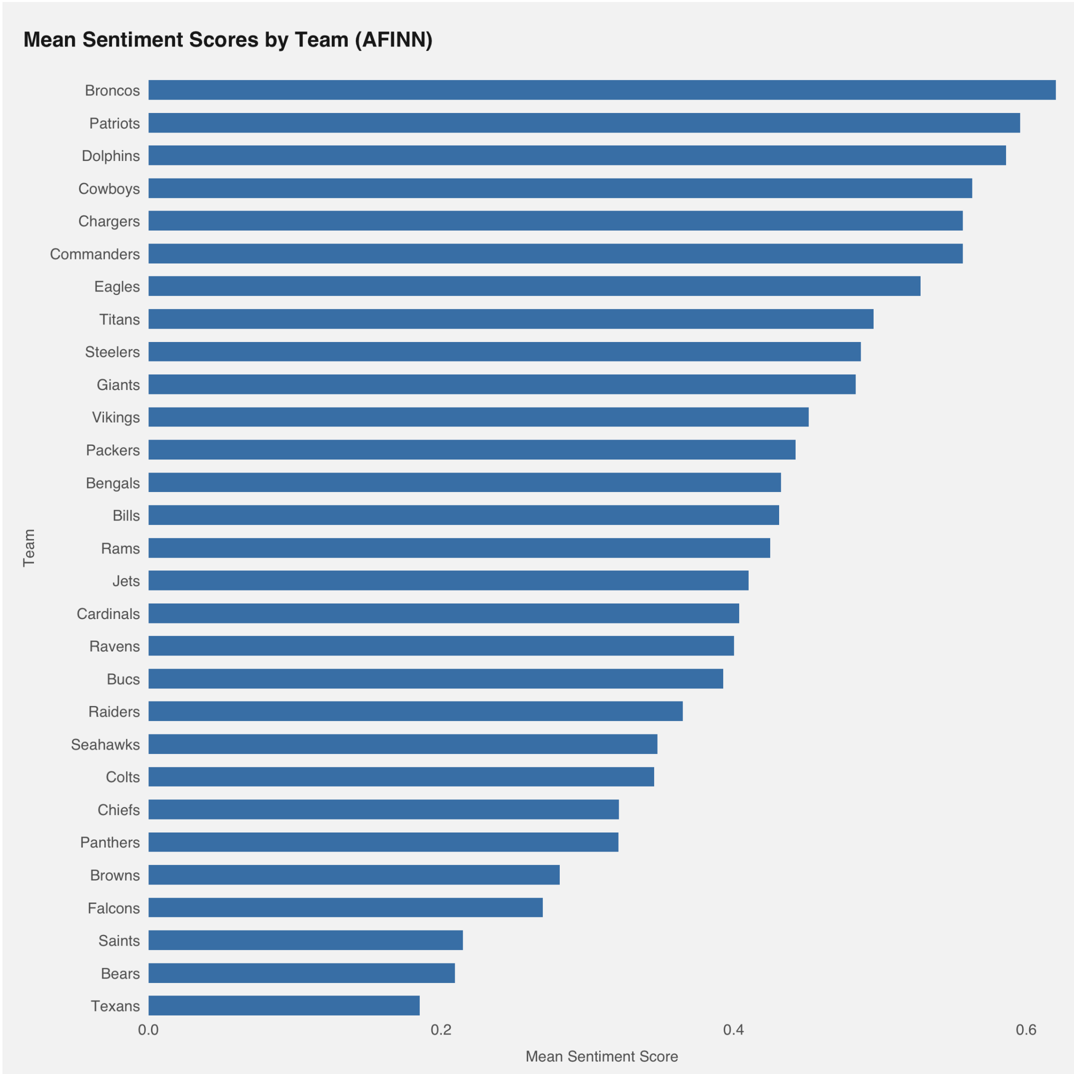

Figure 4: Mean Sentiment Scores by NFL Team (AFINN)

Fig 4 shows mean AFINN sentiment scores per team. As we can see here, the Broncos and Patriots have the highest scores indicating they have higher sentiment than the rest of the group, although it is important to note that the spread of these scores are relatively low compared to the range of values they could actually take. It is interesting to note that the Texans are lowest as I would surmise that they probably had one of the higher sentiment scores last season given their quick turnaround under then rookie quarterback CJ Stroud. AFINN is optimized for direct, opinion-heavy language so the comments that perpetrate higher or lower rankings here obtain sharp criticism or praise and tend to be highly polarizing.

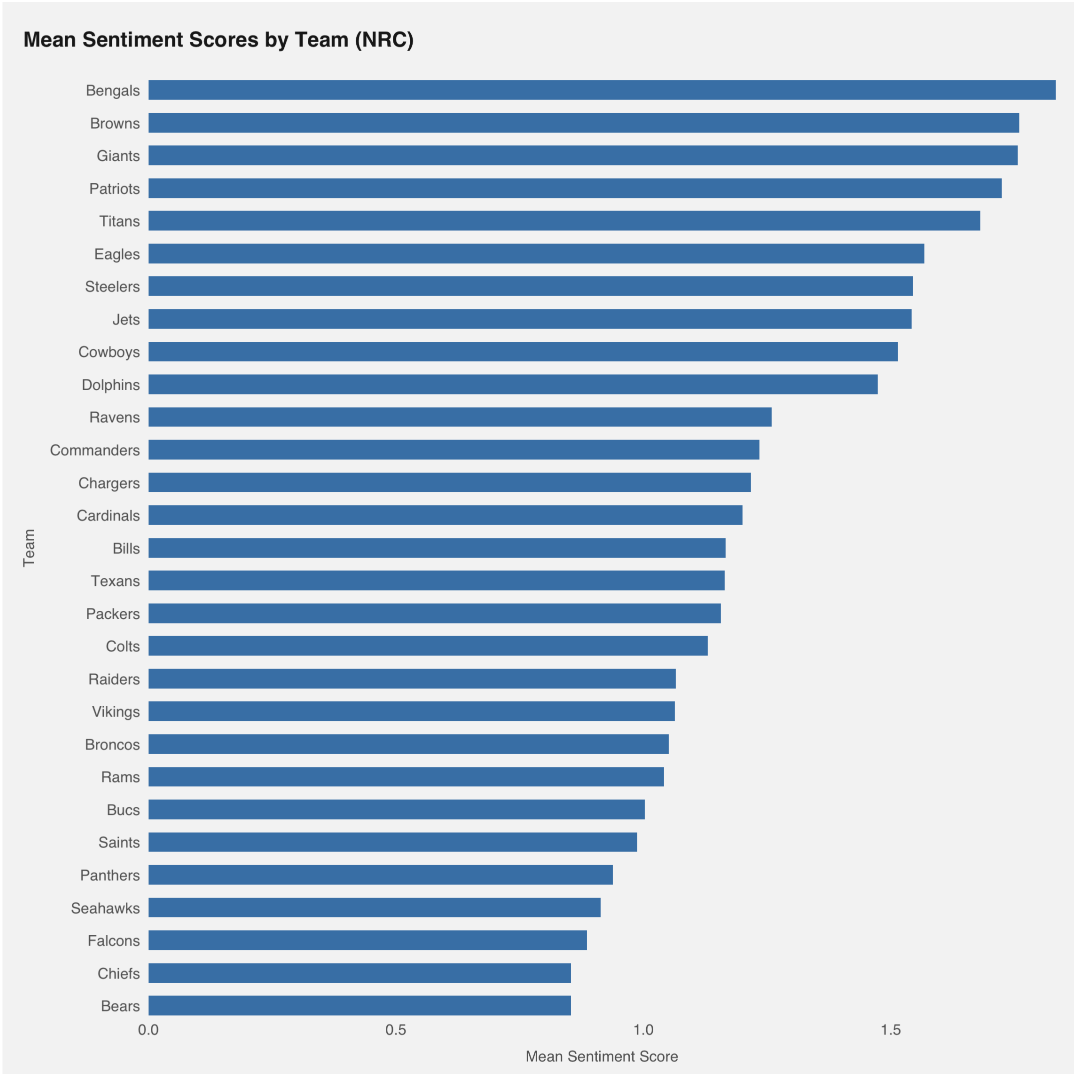

Figure 5: Mean Sentiment Scores by NFL Team (NRC)

Fig 5 displays mean NRC sentiment scores per team. The Bengals landed around the middle in AFINN but scored high in NRC, pointing to fan comments that, while not overly positive in tone, were filled with emotions like hope and trust, which is something that NRC is better at detecting. Bengals fans remained emotionally invested despite the frustrating challenges that their season posed. In contrast, the Bears were at the bottom in both AFINN and NRC, reflecting not just negativity but also a lack of emotional connection, with fans expressing both frustration and indifference.

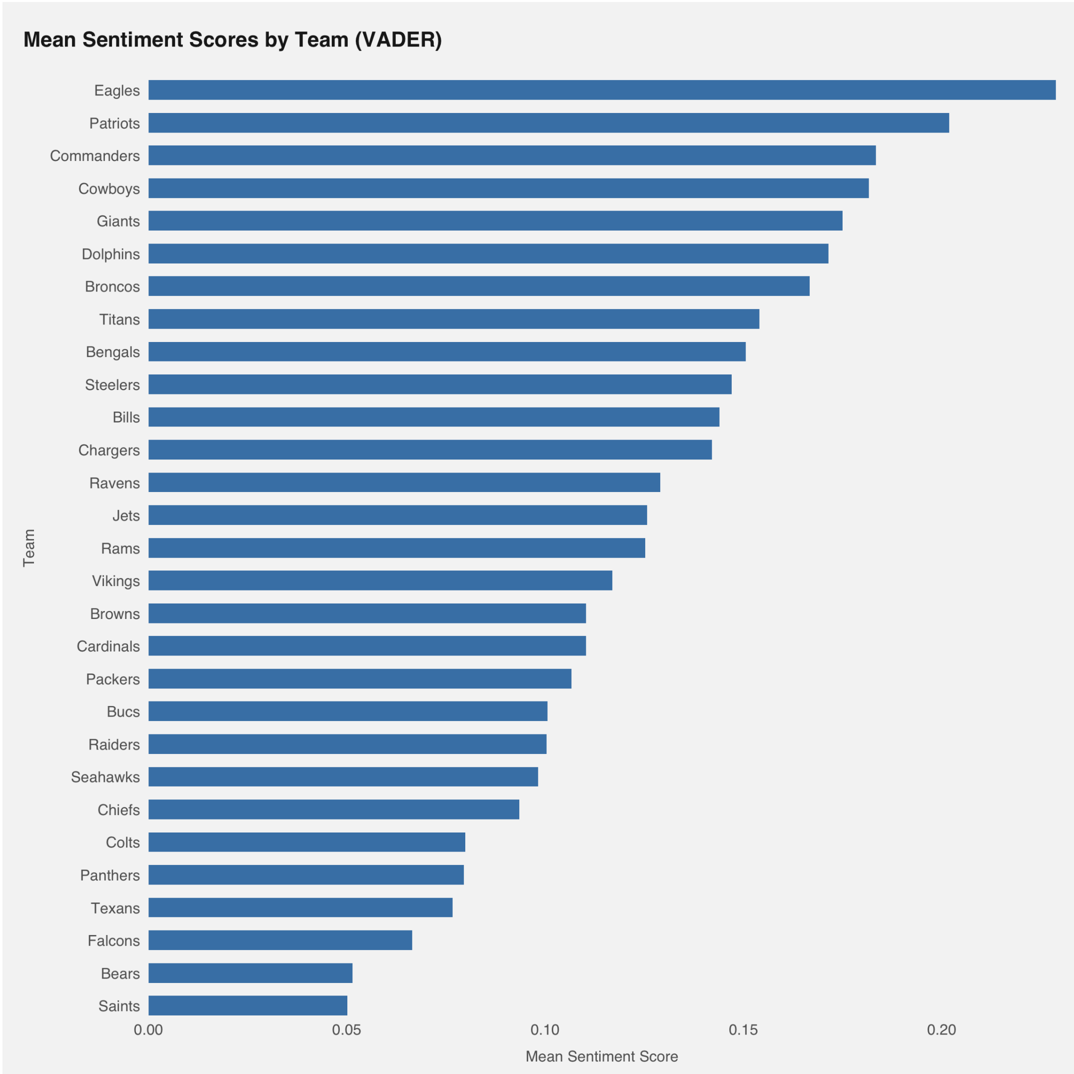

Figure 6: Mean Sentiment Scores by NFL Team (VADER)

Fig 6 shows mean sentiment scores for VADER across NFL teams. The Eagles rank at the top across VADER, AFINN, and NRC, likely because their fan conversations are a mix of hype, strong praise, and emotional energy. V Altogether, it shows a fanbase that's loud, expressive, and emotionally invested in a winning team. On the flip side, the Bears sit near the bottom again, suggesting a season marked by disappointment and low emotional engagement. The tone of fan comments was probably flat, frustrated, or cynical—lacking both positivity and the emotional highs that drive scores up across all three tools.

Quantifying "Overhype":

Given that each of the different metrics capture nuanced complexities in comment sentiment, we must account for them in building a model detecting “overhype”. Since each score type is different and is produced on different scales, we want to normalize each output. Differences in the scales can complicate weighted aggregation since they are nor on the same numerical scale. In the purpose of our data, we will fit all the lexicon scores so that they are within the same boundaries (-1 to 1). We will utilize Min-Max scaling which resizes the values to a fixed range by transforming each score according to the minimum and maximum values of that lexicon’s scores. With this, we will not find any bias from differing score ranges.

Once we normalize the sentiment scores, the next step is to combine all three scores into a composite score (a single measure of sentiment). We will use Principal Component Analysis (PCA) to do this which identifies the most important factors in a set of variables. In combining sentiment lexicons, PCA helps us uncover the main sentiment dimension (whether that is VADER, NRC, or AFINN) that accounts for the most variation in the data. By applying PCA, we can determine how much each lexicon’s normalized score contributes to a key sentiment factor and assign weights accordingly. PCA takes any guesswork out of the weighting approach, we can understand which lexicons matter most.

After doing the PCA, here is what we find is the weighting approach to calculate our composite score:

Sentiment Lexicon	**PCA Weight**
AFINN	0.677
NRC	-0.341
VADER	0.652

Table 1: PCA-Derived Weights for Each Sentiment Lexicon (First Principal Component)

The PCA weights show that both AFINN and VADER lexicons are positively aligned with the dominant sentiment dimension, while NRC contributes negatively as seen by its negative weight. This suggests that when overall sentiment increases (in a way most predictive of performance or ”true sentiment”), AFINN and VADER scores also increase accordingly, but NRC scores might go the other way. Since NRC picks up possibly more nuanced or emotion-specific signals that don’t align as directly with public hype this may be the case. For our purposes, this validates emphasizing AFINN and VADER more when constructing our composite sentiment metric. We then use those weight and multiply it with its according normalized sentiment score to come up with the composite sentiment score used in further analysis. The equation updated for PCA is here:

$$\text{Composite Sentiment} = (0.677 \times \text{AFINN}) + (0.652 \times \text{VADER}) - (0.341 \times \text{NRC})$$

Now that we have established our composite score, we want to evaluate overhype using a series of regression models that relate our composite sentiment score to actual team performance. First, we use a baseline simple linear regression model to relate each team's winning percentage (win_pct_i) as a function of its composite sentiment score ($\text{sentiment\_weighted}_i$):

$$\text{win\_pct}_i = \beta_0 + \beta_1 \cdot \text{sentiment\_weighted}_i + \varepsilon_i \tag{1}$$

The residuals from this regression, denoted $\hat{\varepsilon}_i$, are the difference between expected performance (as implied by public sentiment) and actual performance. We interpret these residuals as a quantitative measure of "overhype" for this simple linear regression model:

$$\hat{\varepsilon}_i = \text{win\_pct}_i - \text{win\hat{\_}pct}_i$$

In this case, a positive residual indicates that a team underperformed relative to the expectations set by sentiment, indicating a case of "overhype". On the other side, a negative residual suggests that a team outperformed sentiment-driven expectations, indicating a case of that team being "underrated".

Since the relationship between sentiment and performance may differ across teams, we proceed with some more sophisticated approaches such as using interaction effects and K-means clustering to capture team-specific deviations and potential other patterns in sentiment-performance relationships. First, we incorporate interaction effects by allowing the relationship between sentiment and other variables (such as Vegas win totals or previous win percentage) to vary. We estimate the following model for this purpose:

$$
\begin{aligned}
\text{win\_pct}_i = \beta_0 &+ \beta_1 \cdot \text{sentiment\_weighted}_i + \beta_2 \cdot \text{vegas\_win\_total}_i + \beta_3 \cdot \text{prev\_win\_pct}_i + \beta_4 \cdot \text{sos}_i \\
&+ \beta_5 \cdot (\text{sentiment\_weighted}_i \times \text{vegas\_win\_total}_i) + \beta_6 \cdot (\text{sentiment\_weighted}_i \times \text{prev\_win\_pct}_i) \\
&+ \beta_7 \cdot (\text{sentiment\_weighted}_i \times \text{sos}_i) + v_{d[i]} + \varepsilon_i
\end{aligned}
$$

This fixed effects model predicts a team's actual winning percentage (win_pct_i) as a function of several key variables: preseason sentiment scores ($\text{sentiment\_weighted}_i$), betting market expectations (vegas_win_total_i), prior season performance (prev_win_pct_i), and strength of schedule (sos_i). The model includes interactions between sentiment and betting expectations ($\text{sentiment\_weighted}_i \times \text{vegas\_win\_total}_i$), sentiment and previous performance ($\text{sentiment\_weighted}_i \times \text{prev\_win\_pct}_i$), and sentiment and strength of schedule ($\text{sentiment\_weighted}_i \times \text{sos}_i$). These interactions capture cases where media sentiment may either align with or deviate from betting market expectations, the team's previous performance, or the difficulty of their schedule. The model accounts for division-level differences via division fixed effects ($v_{d[i]}$), and the error term is denoted ε_i.

Furthermore, given the variability of team performance across the league, we incorporate K-means clustering to group teams based on their sentiment and performance characteristics. This unsupervised learning approach will help identify natural clusters of teams that share similar patterns of sentiment and performance, so we can better understand in pattern recognition that may be systematically overhyped or underrated. The K-means clustering model divides teams into clusters based on their sentiment and win percentage values. We then use these clusters as a grouping variable in the mixed effects model, allowing the intercepts and slopes for each group of teams to vary, further improving the model. By incorporating both interaction effects and cluster-based random effects, we are better able to quantify systematic deviations across teams and pinpoint those that may be misaligned with public sentiment expectations.

Part III: Visualization and Interpreting Results

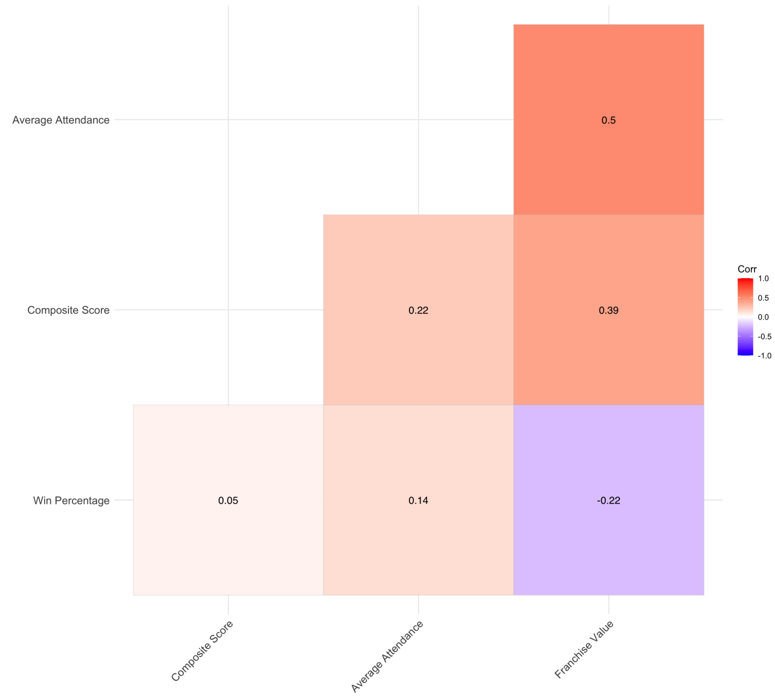

Figure 7: Correlation between Composite Sentiment Score and Indicators

Figure 7 reveals some evidence that fan sentiment, as is only loosely tethered to actual team success. First, the correlation between win percentage and composite sentiment is extremely weak ($\rho = 0.05$), highlighting a significant disconnect between how fans feel and how teams perform. Second, composite sentiment is more closely aligned with franchise value ($\rho = 0.39$), suggesting that wealthier, more established teams may receive inflated levels of fan enthusiasm regardless of on-field outcomes, which could be due to historical legacy or their legacy (a major example of this would be how r/nfl discusses the Dallas Cowboys). Third, there is also a modest positive correlation with attendance ($\rho = 0.22$), indicating that teams generating more sentiment also tend to draw more fans, reinforcing the idea that sentiment reflects engagement more than performance. Together, these patterns underscore a key insight: sentiment appears to track team popularity rather than the quality of the team's success. This may lead us to initially believe that this reason may help us identify the teams that are systematically overhyped.

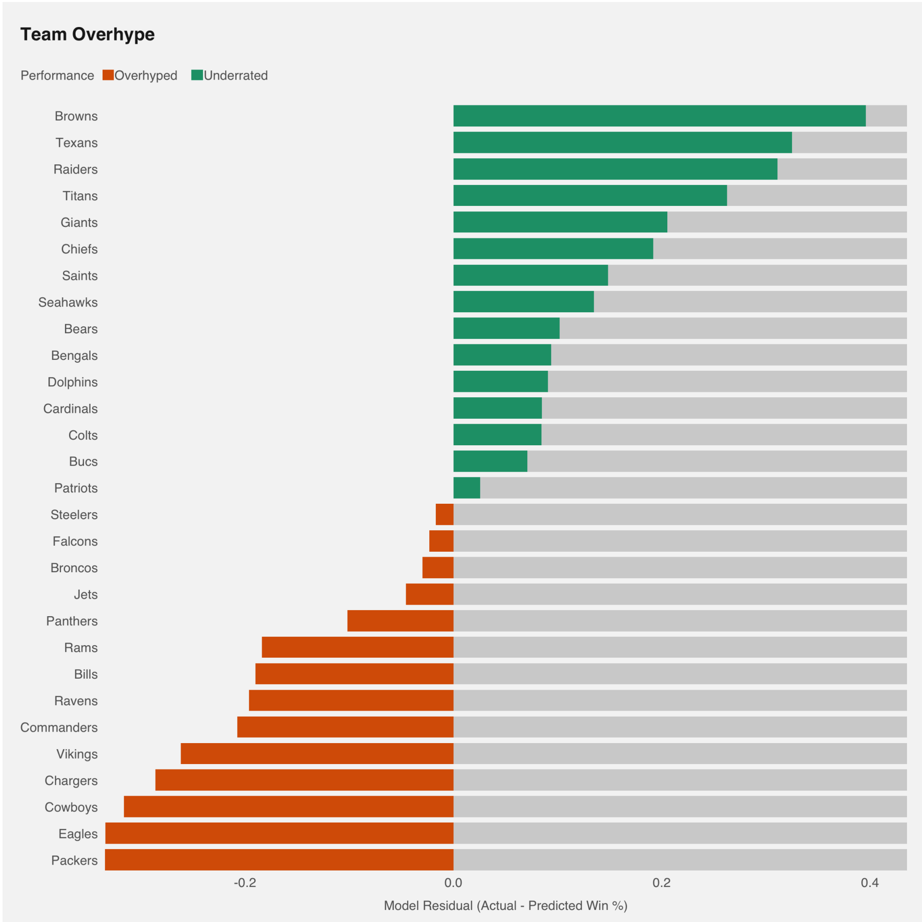

Figure 8: Overhyped Teams based on Model (1) Residuals

Fig 8 divides between teams that significantly underperformed relative to expectations (“over-hyped”) and those that quietly exceeded them (”underrated”). What makes these findings particularly interesting is not just the magnitude of the residuals, but how teams cluster around structural themes that may explain their position on the figure.

On the overhyped end of the spectrum, teams like the Packers, Eagles, Cowboys, Chargers, and Vikings all share a combination of strong national media attention, historically high expectations, and, in some cases, relatively “soft” divisions that may have inflated preseason sentiment. The Cowboys, Eagles, and Commanders (all near the top of the ”overhyped” ranking all reside in the NFC East, a division that has constantly been a media darling with those teams drawing coverage due to high internal instability or defensive inefficiency. It seems here that high visibility combined with large, ruthless fan bases often will inflate sentiment (i.e.

public opinion will be stickier than actual performance).

The Chargers and Vikings are both teams that had star-laden rosters but had a number of key issues that may have not been to apparent to the average NFL fan (i.e. Sam Darnold not performing in high-leverage moments or Justin Herbert's lack of weapons on the offensive side of the field). These teams may have received high composite sentiment scores that didn't account for these deeper systemic issues, so the model captures some of that.

On the other hand, the teams we identify as underrated—such as the Browns, Texans, Raiders, and Titans—share a different set of characteristics. These teams exist in markets with a) less national spotlight and b) with the exception of the Texans had relatively low preseason expectations. They also reside in "grimy" divisions (where competition is tough and can suppress perceived ceiling but the teams are actually somewhat good). With CJ Stroud's second-year regression and the malaise at the quarterback position generally definitely lowered sentiment scores for both the Browns and Texans particularly. It seems with most of the underrated teams except the Texans and the Chiefs have had to answer questions about quarterback skepticism this season, making it a valid question to answer about how quarterback play affects composite sentiment score. Overall, these teams may not garner too many weekly primetime slots, but they quietly delivered results that outpaced their sentiment-weighted expectations.

Table 2: Fixed effects regression predicting actual team win percentage.
Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> |t|)**
(Intercept)	3.21740	0.98143	3.278	0.00599 **
composite_score	-1.30065	1.25052	-1.040	0.31726
vegas	0.07718	0.03220	2.397	0.03226 *
prevpct	-0.10065	0.46552	-0.216	0.83219
sos	-7.06641	1.83006	-3.861	0.00197 **
AFC East	-0.26391	0.20456	-1.290	0.21948
AFC North	0.17908	0.10901	1.643	0.12440
AFC South	0.01579	0.11763	0.134	0.89529
AFC West	0.27040	0.09784	2.764	0.01611 *
NFC East	0.15699	0.10828	1.450	0.17081
NFC North	0.43592	0.12179	3.579	0.00336 **
NFC South	0.20199	0.12231	1.651	0.12257
NFC West	0.35710	0.11383	3.137	0.00786 **
composite_score:vegas	-0.06398	0.05237	-1.222	0.24355
composite_score:prevpct	0.52856	0.70355	0.751	0.46587
composite_score:sos	3.28233	2.47595	1.326	0.20777

Residual standard error: 0.1237 on 13 degrees of freedom

Multiple R^2: 0.8394, Adjusted R^2: 0.6541

F-statistic: 4.53 on 15 and 13 DF, p-value: 0.004657

Table 2 shows the results of a fixed effects model investigating how the composite sentiment score, along with expectation from the betting market, performance in the previous season and the strength of the schedule, influences the percentage of team wins in the NFL season 2024. We include division fixed effects to control for differences in team competition environments as 8/16 of the games in the season are played in-division.

The positive coefficient on composite score ($\beta = 0.345$) suggests that teams receiving more favorable online discourse tend to perform better, even after accounting for Vegas expectations and past win percentage. Although not statistically significant ($p = 0.273$), the directionality supports the idea that fan-driven sentiment on public forums like Reddit may contain weak signals about team quality or confidence.

The interaction between sentiment and Vegas win totals is negative ($\beta = -0.028$, $p = 0.462$), implying that Reddit sentiment offers the most incremental predictive value when it deviates from consensus betting lines. When fans are disproportionately bullish or bearish compared to

Vegas, that divergence might reflect unique insight or irrational optimism or pessimism that affects early-season momentum or narrative framing.

Division-level fixed effects also provide key insights. Teams in the AFC East underperformed the reference division significantly ($\beta = -0.381$, $p < 0.05$), even after accounting for sentiment and other predictors—highlighting how disadvantages within a division can suppress performance. Conversely, teams in the AFC West, NFC North, and NFC West show significantly positive effects, indicating a division-based advantage that may reflect weaker competition or alignment between Reddit optimism and actual on-field outcomes.

Overall, the model fits the data well ($R^2 = 0.81$, adjusted $R^2 = 0.645$), with a significant F-statistic ($p = 0.0022$). While sentiment on r/nfl is not a silver bullet, it appears to function as a low-signal, potentially contrarian indicator of success in this case.

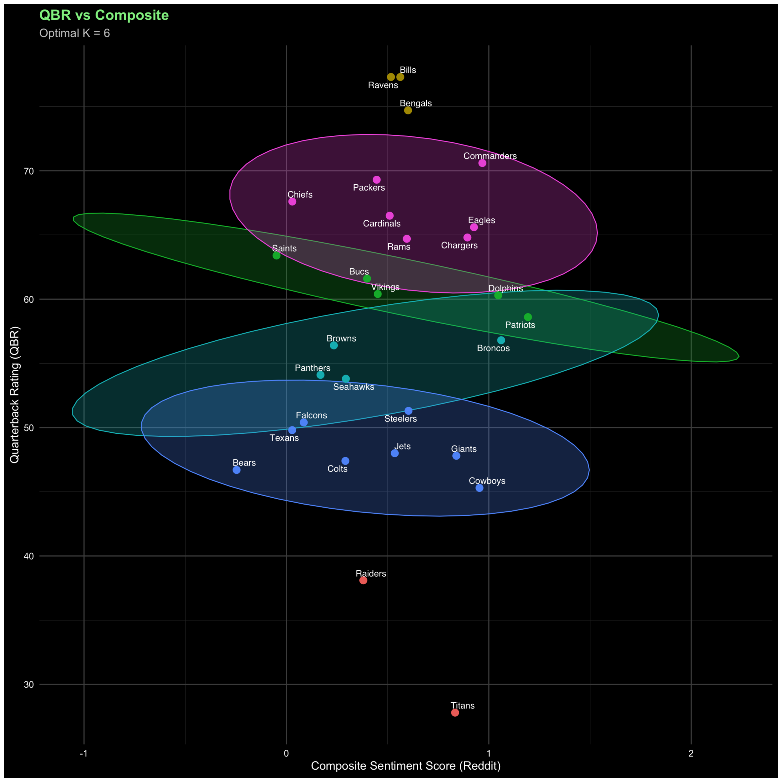

Figure 9: QBR vs Composite Score in K Means Clustering

Fig 9 visualizes NFL teams based on both Composite Reddit Sentiment Score and Quarterback Raing (QBR). QBR is a proprietary statistic developed by ESPN that attempts to rate quarterbacks beyond old passer rating including passing, running, avoiding mistakes, and helping the team iwn on a play by play level. The methodology behind the metric however is shrouded in mystery.

The color-shaded ellipses represent six distinct clusters determined by the k-means algorithm, with team names plotted accordingly. A general trend is visible: teams with higher QBRs tend to occupy clusters that also have higher average sentiment scores. For instance, teams like the Bills, Ravens, and Bengals are in a cluster that sits far above others on the QBR axis, supporting the idea that strong quarterback play tends to generate positive fan and public perception online.

Interestingly, there's not a perfect one-to-one relationship between Reddit sentiment and QBR. Some teams like the Titans and Raiders cluster together with low QBRs and negative or neutral sentiment, implying poor quarterback performance is indeed met with lower fan optimism.

However, teams such as the Cowboys, Giants, and Jets all show moderate to low QBR but somewhat neutral or slightly positive sentiment. This could suggest fanbases that remain hopeful despite lackluster quarterback metrics—or it could reflect market size and engagement skewing sentiment data. Additionally, Commanders and Eagles show relatively strong sentiment scores despite not leading in QBR, which might point to fanbase excitement around offseason narratives (i.e. Commanders drafting Jayden Daniels or Eagles signing Saquon Barkley).

Another interesting cluster includes mid-tier QBR teams like the Dolphins, Patriots, and Broncos, who have moderate sentiment scores. These cases may show a tighter alignment between how well quarterbacks perform and how fans talk about them. The presence of clear outliers—such as Bills and Titans—emphasizes that while sentiment and quarterback play are correlated, they're not identical. Reddit sentiment captures emotional and narrative-driven perception, while QBR is a more objective metric.

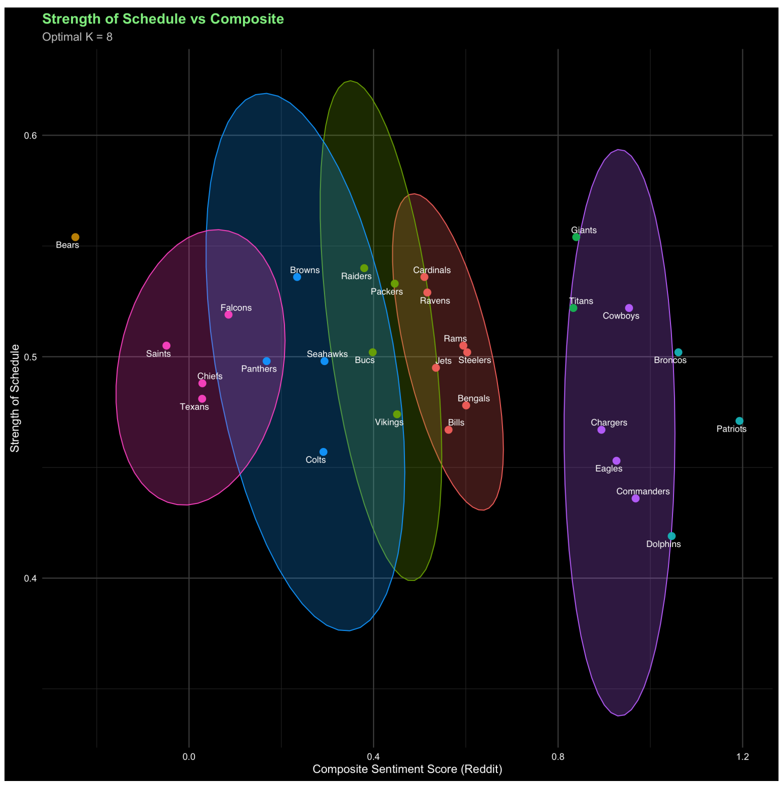

Figure 10: SOS vs Composite Score in K Means Clustering

Fig 10 shows distinct clustering of NFL teams based on their strength of schedule (SOS) and Reddit sentiment scores. Teams in the center-right cluster (Ravens, Cardinals, Rams, Steelers, Jets, Bills, Bengals) face moderately challenging schedules while receiving mixed-to-positive sentiment from fans. In contrast, the far-right teams (Cowboys, Broncos, Chargers, Eagles, Commanders, Dolphins, Patriots) have significantly easier schedules according to SOS metrics, yet show widely varied sentiment scores, indicating fan perception isn't strictly tied to schedule difficulty for these teams. The bottom-left cluster (Saints, Chiefs, Texans, Panthers, Falcons) faces easier schedules but receives less positive sentiment, possibly indicating fan skepticism despite favorable matchups. Meanwhile, teams in the center-left (Browns, Seahawks, Bucs, Vikings, Colts) face moderate-to-difficult schedules with corresponding moderate sentiment, suggesting realistic fan expectations.

The vertical clustering of teams with similar composite sentiment scores suggests that NFL fans often form opinion bubbles around franchises regardless of their actual schedule difficulty. In this case, the Cowboys, Eagles, and Chargers demonstrate that even teams with the easiest schedules can face harsh criticism from their own fan bases, revealing a pervasive pessimism that transcends favorable matchups.

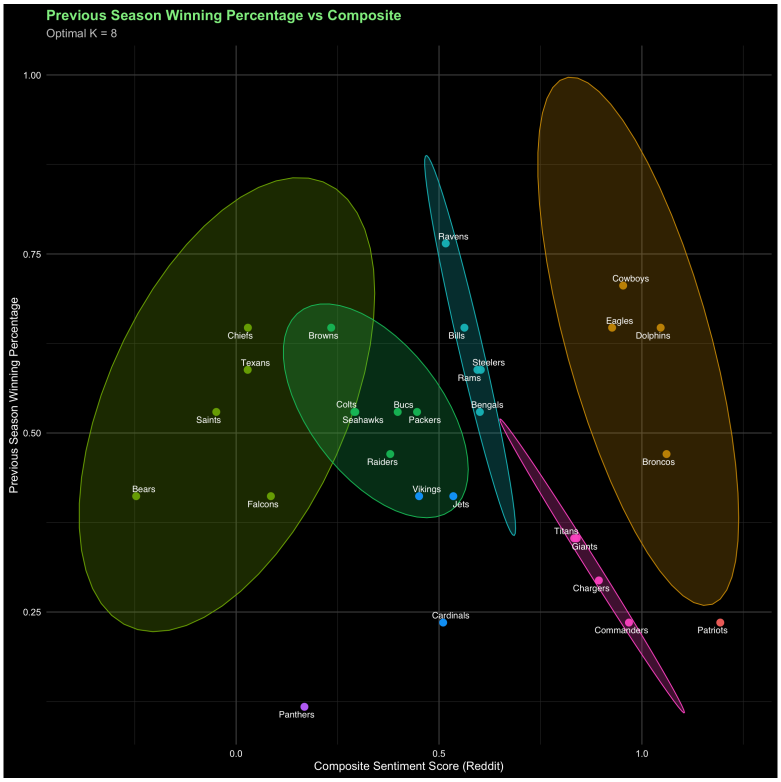

Figure 11: Previous Season Win Percentage vs Composite Score in K Means Clustering

Figure 11 reveals a counterintuitive relationship between NFL teams' previous season winning percentage and fan sentiment on Reddit. The right-side brown cluster (Cowboys, Eagles, Dolphins, Broncos, Patriots) shows that teams with strong Reddit sentiment don't necessarily have the best previous season records, with only the Cowboys and Eagles posting winning seasons above 0.6. Meanwhile, the central teal cluster (Ravens, Bills, Steelers, Rams, Bengals) demonstrates consistently high winning percentages coupled with moderately positive sentiment, suggesting these teams have earned cautious respect from the community through sustained performance.

The left-side green cluster presents the most striking contrast, containing several successful teams (Chiefs, Texans, Browns) with winning records above 0.5 that paradoxically receive low sentiment scores on Reddit. This includes the Chiefs, who despite their sustained success, generate surprisingly negative discussion. The pink diagonal cluster (Titans, Giants, Chargers, Commanders) shows progressively worse previous season records corresponding with increasingly positive sentiment scores, indicating a potential effect where fans of struggling teams express optimism despite poor recent performance.

The data exposes the "Chiefs Paradox" – despite being the most successful NFL franchise in recent years, Kansas City generates some of the most negative sentiment on Reddit, as similar to the Patriots in the earlier 21st century where sustained effect can breed negative emotion among fans (The Score).

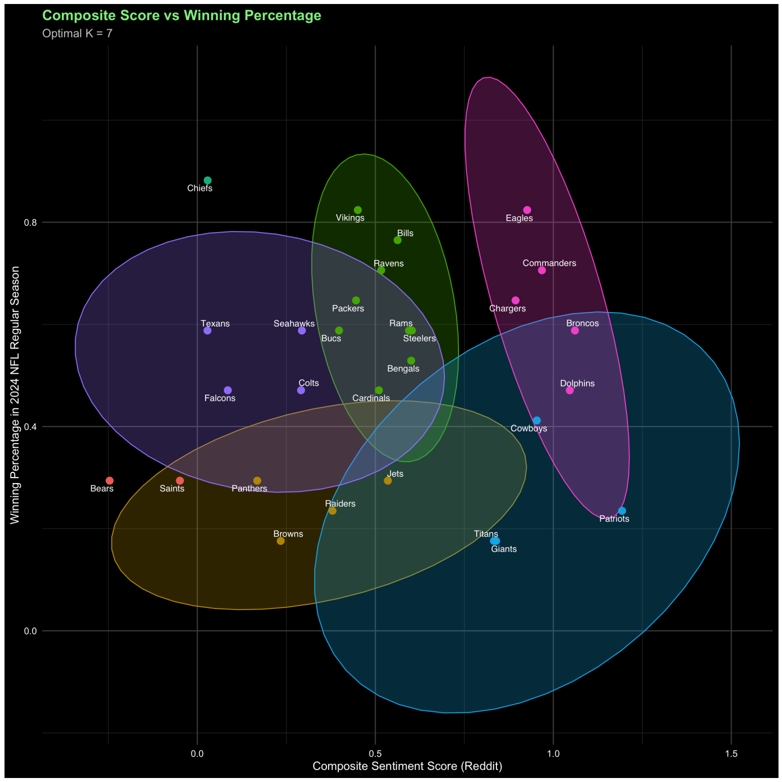

Figure 12: Season Win Percentage vs Composite Score in K Means Clustering

Fig 12 reveals a complex relationship between NFL teams' 2024 winning percentages and Reddit sentiment scores. The green cluster (Vikings, Bills, Ravens, Packers, Rams, Steelers, Bengals) displays teams with average-to-positive sentiment and strong winning records, suggesting these franchises are performing in line with fan expectations. In stark contrast, the purple cluster on the right (Eagles, Commanders, Chargers, Broncos, Dolphins, Patriots) shows teams with highly positive sentiment despite widely varying performance levels, with the Eagles and Commanders significantly outperforming their counterparts in this sentiment group. The left-side clusters present interesting anomalies: the dark blue cluster (Texans, Seahawks, Falcons, Colts) shows decent performers receiving surprisingly negative sentiment, while the brown cluster (Bears, Saints, Panthers, Browns, Raiders) combines poor performance with negative sentiment as expected. The Chiefs stand as a remarkable outlier with the highest winning percentage despite extremely negative sentiment, reinforcing their villain status in the NFL community. Meanwhile, the teal cluster (Jets, Titans, Giants) represents teams with below-average performance yet moderately positive fan sentiment, suggesting these fanbases maintain optimism despite disappointing results.

The data again supports the "Chiefs Paradox" where the league's winningest team paradoxically generates the most negative sentiment on Reddit, while teams like the Eagles and Commanders receive highly positive discussion that accurately reflects their strong on-field performance, demonstrating how Kansas City's (like the Patriots before them) has uniquely positioned them as the NFL's most successful yet polarizing franchise.

Part IV: Conclusions and recommendations

Our analysis of NFL team sentiment on r/nfl in relation to various performance and financial metrics around the league has revealed many fascinating insights into the way fan perception can operate independently from on-field and off-the-field results. Our findings continue to back the idea of the "Chiefs Paradox", an idea that was similarly embodied by the Patriots when I was growing up. Despite arguably being one of the NFL's most successful franchises in recent history, they generate an egregious amount of negative sentiment among NFL fans. We believe in this case that sustained excellence breeds resentment in online communities. We found extremely weak correlation between winning percentage and composite sentiment (= 0.05) which demonstrates that fan discourse operates largely detached from actual team performance, instead aligning more closely with franchise value (= 0.39) and attendance figures (= 0.22).

Our analysis identifies clear patterns of "overhyped" and "underrated" teams based on the residuals between expected and actual performance. Teams like the Cowboys, Eagles, and Chargers appear systematically overhyped, benefiting from national media attention, historically high expectations, and large, engaged fanbases that inflate sentiment regardless of results. Conversely, teams such as the Browns, Texans, Raiders, and Titans emerge as underrated, receiving disproportionately negative sentiment despite solid performances. These teams typically operate in smaller markets with less national coverage and face quarterback skepticism that suppresses fan optimism. The data suggests that market size, divisional narratives, and historical reputation significantly influence sentiment beyond simple win-loss records.

The fixed effects model provides additional evidence that Reddit sentiment contains weak predictive signals about team performance, even after controlling for Vegas expectations and past win percentages. While not statistically significant (= 0.345, p = 0.273), the positive coefficient on composite score supports the notion that fan sentiment may capture intangible factors that betting markets miss. The negative interaction between sentiment and Vegas win totals (= -0.028) further suggests that Reddit discourse offers the most predictive value when it diverges from consensus betting lines.

Finally our cluster analysis reinforces these findings, revealing that teams group according to factors beyond simple performance metrics. The relationship between sentiment and quarterback rating (QBR) shows an expected correlation but with notable exceptions like the Cowboys and Giants, whose moderate-to-positive sentiment despite mediocre quarterback play suggests fanbases that remain optimistic regardless of performance reality. Similarly, the clustering of teams based on strength of schedule versus sentiment demonstrates how fans form "opinion bubbles" around franchises irrespective of objective schedule difficulty. Teams like the Eagles and Commanders enjoy highly positive sentiment that aligns with their strong performance, creating a stark contrast with the Chiefs' negative sentiment despite their league-leading winning percentage.

Recommendations:

To extend this research, we recommend developing a more sophisticated sentiment analysis model capable of being able to distinguish from a range of emotions: frustration, sarcasm, hate, or misery. This would be an improvement over our current system classifying binary positive / negative classification. We could use BERT or RoBERTa on a lot more sports-specific commentary. This would help capture a lot more nuance in team sentiment. We could also further attribute this sentiment to various specific stakeholder groups (i.e. hometown fans on team subreddits vs. league-wide commentary on r/nfl).

Furthermore, we could layer in temporal dynamics using time series modeling so we could get a realtime sense on how fan sentiment evolves in response to major news such as coaching changes, trades, free agency signings, pivotal wins / losses, and / or injuries. This would help us better identify changes in public perception relate to shifts in narrative momentum.

Given that we found a moderate link between sentiment and franchise valuation, a further longitudinal study on the relationship between online sentiment and broader financial metrics such as merchandise revenue, television ratings, and brand partnership values could be very useful. With panel data, we could identify if sentiment is a function of being a leading indicator or lagging one in regards to financial success.

References

[1] Amazon's "TNF" sees 11% spike in NFL ratings | *Reuters*. (n.d.-a). Retrieved from https://www.reuters.com/sports/amazons-tnf-sees-11-spike-nfl-ratings-2025-01-02/

[2] Crupi, A. (2024, January 5). NFL swallows TV whole, with 93 of year's top 100 broadcasts. *Sportico*. Retrieved from https://www.sportico.com/business/media/2024/nfl-posts-93-of-top-100-tv-broadcasts-2023-1234761753/

[3] Hong, Y., & Skiena, S. (2010, May). The wisdom of bookies? Sentiment analysis versus the NFL point spread. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 4, No. 1, pp. 251–254).

[4] Li, Q., Zhang, J., Guo, J., Li, J., & Kang, C. (2021, November). Evaluating performance of NBA players with sentiment analysis on Twitter messages. In *Proceedings of the 2021 European Symposium on Software Engineering* (pp. 150–155).

[5] Reddit user and growth stats (updated January 2025). *Backlinko*. (2025, January 21). Retrieved from https://backlinko.com/reddit-users

[6] Schumaker, R. P., Jarmoszko, A. T., & Labedz Jr, C. S. (2016). Predicting wins and spread in the Premier League using a sentiment analysis of Twitter. *Decision Support Systems, 88*, 76–84.

[7] Stinson, S. (2025, April 22). Chiefs reach final boss status, and everyone hates it. *The Score*. Retrieved from https://www.thescore.com/nfl/news/3208641/chiefs-reach-final-boss-status-and-everyone-hates-it

[8] Sinha, S., Dyer, C., Gimpel, K., & Smith, N. A. (2013). Predicting the NFL using Twitter. *arXiv preprint* arXiv:1310.6998.