Hype or Insight? Exploring the Link Between Reddit Reactions and NBA Performance

Vikram Kadyan
Rutgers University
vk414@scarletmail.rutgers.edu

Oways Jaffer
Rutgers University
omj9@scarletmail.rutgers.edu

1 Introduction

In this project, we investigate the relationship between NBA fan sentiment and player in-game performance using unsupervised learning techniques. Our primary goal is to explore and characterize the connections between key performance metrics -such as points scored, rebounds, assists, and minutes played - and sentiment scores derived from Reddit posts and comments regarding games and players. The central question we are addressing is: to what extent does online fan sentiment align with, and potentially reflect underlying patterns in, present player performance, and are there discernible associations that might suggest how sentiment co-evolves with performance narratives over time?

We have chosen to focus on a small group of NBA stars, selected for their consistent media attention, to ensure a rich dataset of fan commentary for sentiment analysis. We collect game-by-game performance data and pair it with aggregated Reddit sentiment captured around the time of each game. This project is undertaken with an unsupervised approach to analyze how sentiment trends correspond with player performance trends and to identify if any distinct groups or clusters emerge from the combined data, revealing characteristic game-sentiment profiles.

This approach is tied closely to concepts and papers covered in the course, particularly unsupervised methods such as clustering and dimensionality reduction, which help to reveal structure in complex datasets without predefined labels. The project also incorporates aspects of natural language processing (NLP), as it requires extracting and quantifying sentiment from unstructured text data.

2 Motivation

In the last couple of decades, the intersection of sports analytics and social media has become increasingly prominent. Millions of fans each day discuss players and games online, using platforms like Reddit and X. This gives us a unique opportunity in history to tap into collective fan sentiment and analyze how public opinion forms around athletic performance.

This project is driven by curiosity about whether online sentiment reflects objective performance, anticipates it, reacts to it, or is barely correlated. Understanding this relationship could offer insights into fan psychology, narrative formation, and potentially even performance forecasting.

We are particularly excited about this project because it allows us to combine natural language processing and unsupervised learning in a culturally relevant real-world context: NBA basketball. Fan communities on Reddit are active, opinionated, and data-rich, making them ideal for this kind of analysis. Additionally, the project blends qualitative social signals (text sentiment) with quantitative game metrics, allowing for a multi-dimensional view of performance.

Sentiment analysis is not new, but it is underused in analyzing the sports and social media space. This analysis could be valuable for sports media firms, fantasy sports users, sports gamblers, sports gambling companies, social media companies, and professional sports psychologists in understanding how fan perception does or doesn't align with reality.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

3 Methodology

3.1 Dataset Description and Data Preparation

Our dataset combined two sources: Official NBA player performance statistics and sentiment scores of content produced by Reddit users.

Player performance data was gathered using the official NBA API [1], which holds the official box scores and statistics for players and teams in all NBA games. It is structured in a tabular format, with each row being a specific game. Columns include the date of the game, teams playing one another, player name, and statistical metrics for that player in this game. The statistical metrics we focused on were the following: Points, assists, rebounds, steals, blocks, plus-minus, and game outcome (win or loss). We took data from the 2022-2023, 2023-2024, and 2024-2025 NBA regular seasons. The players chosen for this study were LeBron James, Stephen Curry, Jalen Brunson, Giannis Antetokounmpo, Luka Doncic, Shai Gilgeous-Alexander, Anthony Edwards, and Donovan Mitchell.

The sentiment data was collected from the r/NBA subreddit on Reddit [2]. Using the Python Reddit API Wrapper (PRAW) and Reddit's webapp API, we scraped posts and comments mentioning selected NBA players around and on the dates of their games. Each Reddit comment was timestamped and associated with the selected player based on direct mention of their name or known nickname. Sentiment analysis was performed using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool from NLTK, which returns scores for positive, neutral, negative, and compound sentiment for each text entry. These scores were calculated separately for post titles, bodies, and comment samples, and then appended to the dataset to provide a structured representation of fan sentiment surrounding each game.

The final dataset merges the player's stats for a game with average fan sentiment from Reddit surrounding that player for that game.

3.2 Modeling Approach / Analysis

We performed an exploratory analysis using unsupervised techniques to compare Reddit Sentiment and NBA player performance across games. Our goal was to find patterns across different players' data.

We visualized univariate distributions of key performance metrics (such as points, assists, shooting percentage) and sentiment features (such as mean sentiment and post volume). We then calculated correlation matrices for these relationships. We did this to identify linear relationships between these categories, such as between player sentiment and points scored.

Evaluation was based on consistency and clarity of observed patterns across players. We considered the analysis successful if sentiment metrics showed clear variation when game outcome and statistics changed, as this would show that there is a meaningful connection between fan sentiment on Reddit and game outcome.

4 Results

Figure 1 shows the relative lack of correlation between fan sentiment online and whether a player won or lost their recent game. Mean sentiment increased a tiny amount- about 0.15- between a win and a loss. This shows fans were slightly but not greatly affected by whether the players they watched won or lost. Since our data was only from the regular season, where game outcome is not generally thought to be as important as in the playoffs, this raises the question of whether this sentiment trend would change if playoff games were the games analyzed.

Figure 2 shows a deeper view of the variability of fan sentiment across game outcomes. While median sentiment is higher for wins, the spread is far wider and contains many lower outliers. This shows the reaction to wins is far more varied than the reactions to losses. Again, this raises questions of whether these results would remain the same for playoff games as opposed to the regular season games being analyzed.

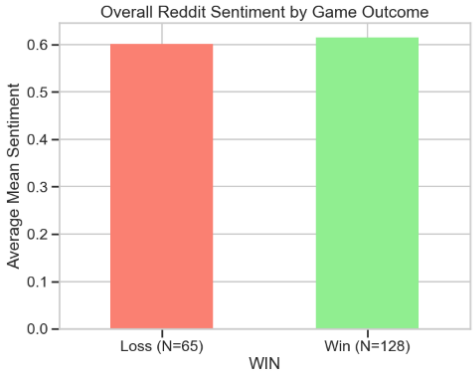

Figure 1: Compares average Reddit sentiment about star players after their team has lost or won a game

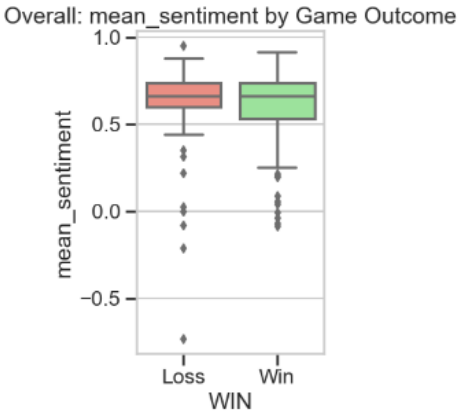

Figure 2: Displays the distribution of Reddit sentiment scores for star players following wins and losses. Sentiment tends to be higher and less variable after wins, and more mixed after losses.

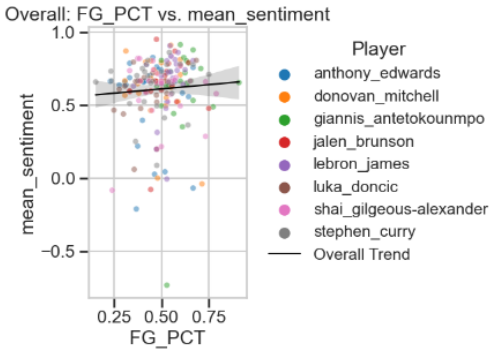

Figure 3: Shows the relationship between a player's field goal percentage in a game and the average sentiment of Reddit posts discussing that performance

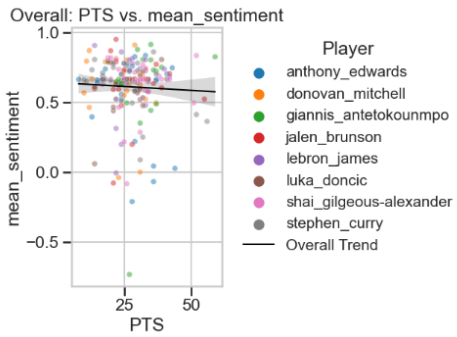

Figure 4: Plots player point totals against Reddit sentiment to evaluate whether higher scoring games correlate with more positive fan reactions

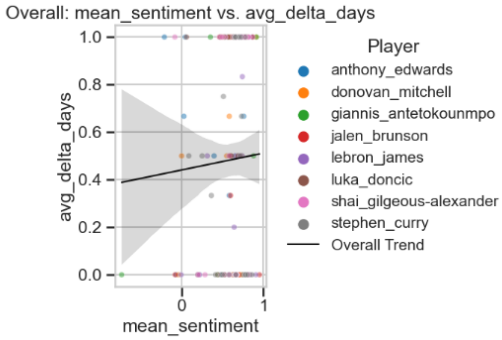

Figure 5: Examines how the average delay between a game and related Reddit comments correlates with sentiment, exploring possible recency or cooling-off effects

Figure 3 shows that there is a weak positive correlation between a player's field goal percentage and their mean sentiment score for a game. The spread is broad, but players who shoot more efficiently for a game get noticeably higher sentiment on average. In basketball, shooting efficiency is an essential indicator of skill and offensive capability. Highly efficient, high-volume scorers are incredibly valuable players to have on a team. This raises questions around fan awareness of a player's field goal percentage for a game, as this is not a stat that is always immediately available about a player's performance.

Figure 4 shows a slight negative correlation between points scored and mean sentiment. As players scored more, the mean sentiment tended to go down a little bit. However, the spread was very wide here, like it was with the rest of the data. Higher-scoring games (>35) were relatively rare in the dataset, and so it is also likely that there could be outliers in the very high point totals skewing the results. This raises an obvious question- why are points not positively correlated with sentiment?

Here, Figure 5 shows a positive correlation between avg_delta_days (days between the game happening and a post about the game being made) and mean sentiment about the game. There is still a pretty wide spread, and the correlation does not seem to be especially meaningful. However, it suggests fan sentiment may get slightly more positive the longer they have been between posting about a game on Reddit and watching the game live.

Plus-minus is a basketball stat that measures the point differential of a team while a specific player is on the court. There is no meaningful trend of sentiment compared to plus-minus. Plus-minus is considered to be a meaningful basketball statistic when measuring a player's impact on the game, with a more positive plus-minus being better.

No meaningful statistical correlations were found between sentiment metrics and other basketball metrics, such as assists or blocks.

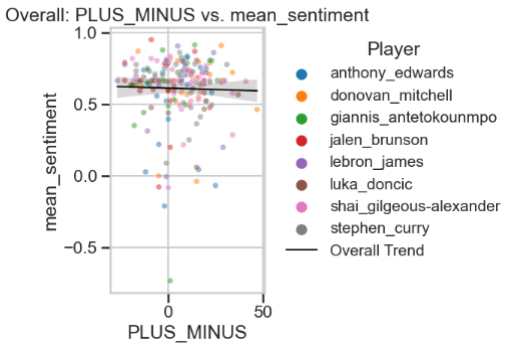

Figure 6: This shows the correlation between player plus-minus score for a game and mean sentiment

5 Discussion

5.1 Expected Outcome

We expected our results to show a positive correlation between sentiment and generally positive player statistics. The expectation was for the correlation to be pretty linear, allowing for slight variance between games simply due to player characteristics or other variables that cannot be quantified through the NBA API. We also expected general sentiment to carry momentum between games, as previous performances can skew expectations for future performance.

5.2 Analysis and Discussion of Results

Our analysis revealed some relatively weak connections, although none were as strong as we had hoped when setting out to do this project. Some connections were also not in the direction that we predicted in the expected outcome section.

Our analyses revealed a nuanced and unexpected relationship between many aspects of player performance and fan sentiment. The results were far more complex and dynamic than positive fan sentiment, correlating well with good player performance.

Figures 1 and 2 show a very mild amount of responsiveness of fan sentiment to game outcome. Although players on average receive slightly higher sentiment after a win, the difference is almost negligible. The higher range of sentiment for wins slightly contradicts our prediction, as there were many low-sentiment scores given to players who had won games.

It is interesting to note that for this aspect of the analysis, we had considerably more won games than lost games in our dataset (128 wins to 65 losses). This is because we only looked at eight superstar players, all of whom are on good teams in the league, meaning they win more than they lose. This could have an odd effect on the data, as more games that were won could allow for more variance in those games. Since fans would also be more used to seeing these players win, their sentiment would be less positively affected by watching them win, as this is expected. Maybe players on worse teams would have higher sentiment when they won games, as wins were not expected but a welcome surprise.

We observed a weak positive correlation between field goal percentage and sentiment as shown in Figure 3, which aligns with our expectations. It makes sense that this would be a weak correlation, as a player's field goal percentage for a game is not always very well-known. Field goal percentage often ends up being a statistic only a few devoted fan analysts track game to game.

Surprisingly, Figure 4 shows an unexpected negative trend between points scored and sentiment. This contradicts the intuitive assumption that higher-scoring games would generate more praise. There are several factors that could account for this. High-scoring games in losses may make fans feel as if a player is stat-padding or ball hogging. Additionally, since we are just looking at star players, high fan expectations may be causing fans not to value high-point total performances, especially. Sentiment may be far more influenced by harder-to-measure things like game context, clutch moments, and efficiency than by raw scoring.

Figure 5 introduces an interesting angle: the sentiment becomes slightly more positive the further a comment is removed from the date of the game. This suggests a sort of cool-off effect, where fans become more generous when given more time to think about a game. This is a very interesting insight and sets the stage for more research into how sentiment changes when fans are given more time to think and distance themselves from a game.

An area of our study that may have fallen short is in its ability to account for context. Factors such as game importance, performance expectations, opponent quality, and memorable in-game moments are all likely to shape sentiment significantly but were not measured in our dataset. Additionally, our sentiment analysis was limited to direct name and nickname mentions, meaning we may miss other relevant discussions, such as replies, memes, or more subtle commentary.

These contextual factors could be what is separating our measured results from our expected outcome. Maybe sentiment doesn't closely align with in-game performance because these factors play a much bigger role in creating narratives that appeal to fans, thus changing fan sentiment. Future research would have to be done on this topic.

A potential takeaway is that fan sentiment may not be driven by just performance metrics, but by more abstract or subjective factors such as narrative momentum, media coverage, or individual player popularity. This suggests that modeling sentiment in sports may require mixing structured performance data with unstructured data such as social and cultural context. Our report underscores the limitations of using only quantitative data to understand and predict qualitative social reactions such as fan sentiment.

Based on how success was defined earlier in this paper, our project was partially a success. We were able to establish some weak correlations between some performance metrics and fan sentiment. However, since the correlations were weak and the data exhibited wide variance, the project fell short of fully achieving its goal. We were unable to identify strong in-game performance indicators that consistently aligned with positive fan sentiment, which was the central aim of our analysis.

5.3 Difference from Original Plan and Future Possibilities

Our final project differed from our original plan mainly in that we used Reddit posts and comments instead of tweets from X.com (formerly Twitter). This change was necessary due to recent API changes—querying X now requires payment, while Reddit's API remains free and accessible. This allowed us to stay within our $0 project budget. Despite the platform change, our sentiment analysis methodology and overall project scope remained largely the same.

There are several ways future versions of this project, or similar projects, could be improved in terms of data quality and analytical depth. First, expanding beyond r/NBA would significantly enhance sentiment diversity. We were limited to a single subreddit due to the time-intensive nature of scraping Reddit data. With more time and computing resources, future work could incorporate additional NBA-related subreddits.

With additional budget, future projects could also access the X.com API to pull short-form, real-time reactions. The X community is highly active during games, and comparing sentiment patterns across Reddit and X could reveal platform-specific trends or predictive differences.

Adding more contextual data could also strengthen the analysis. Factors like opponent strength, game location, rest days, or general game hype might influence both performance and fan sentiment. Incorporating this context could help isolate when sentiment truly reflects performance versus external circumstances.

All of these improvements would contribute to a more robust, nuanced, and scalable analysis and could help set a foundation for future research at the intersection of sports analytics and online sentiment. These improvements may also cause results more aligned with our expected outcome for this subject.

References

[1] National Basketball Association. NBA.com/stats. `https://www.nba.com/stats`. Accessed: 2025-04-14.

[2] Reddit. r/nba. https://www.reddit.com/r/nba/. Accessed: 2025-04-15.