LING211

Yi-Chyun W., Akram A., Hayden B., Bashir A., Najib A.

Group Project Report

Professor Osborne

Social Sentiments to Football Standings:
How Social Sentiments Translate to EPL Standings

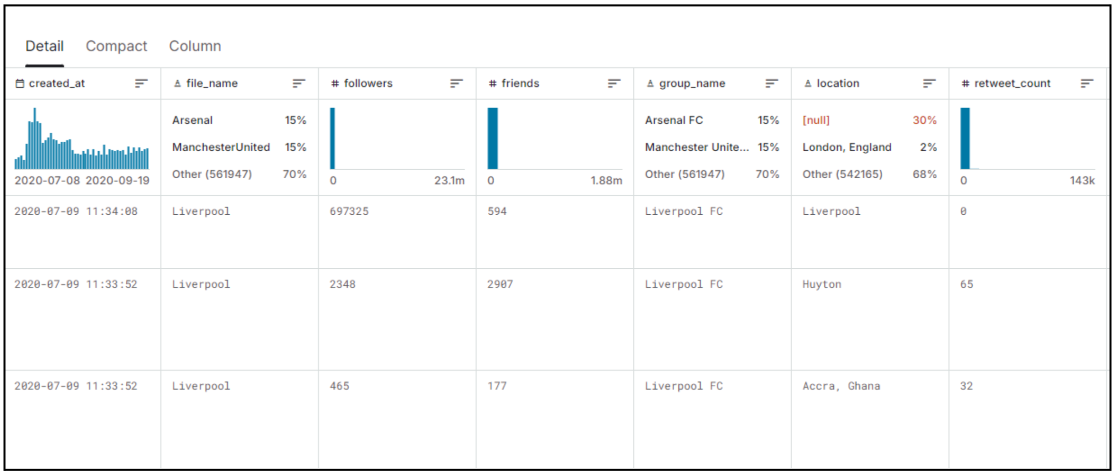
Introduction

Soccer fans are highly active on Twitter, constantly sharing their emotions, opinions, and reactions during and after matches. This vibrant activity creates a vast and dynamic dataset that captures fan sentiment, offering valuable insights into public perceptions of teams and players. By analyzing Twitter data, this project aims to identify trends in fan reactions and correlate them with team performance throughout the season. These insights go beyond just understanding fan sentiment, our goal is to use this information to predict the final standings of the Premier League. This approach connects the emotions of fans with real-world outcomes on the field, offering a data-driven perspective on the sport. Using advanced sentiment analysis techniques, such as a modified Vader model, we uncover patterns that reflect the relationship between digital engagement and on-field performance. By leveraging these insights, we aim to demonstrate how social media sentiment can serve as a predictive tool for understanding and forecasting Premier League results.

Data

For this analysis, we utilized the Kaggle dataset titled EPL Teams - Twitter Sentiment Dataset. This dataset includes two subsets of tweets about English Premier League teams, covering two periods: July 2020 to September 2020 and September 2020 to October 2020. It offers a comprehensive collection of fan reactions, providing a rich source of data for sentiment analysis and team performance evaluation. The dataset includes key fields such as team hashtags, which identify the team being discussed (for example, #Chelsea, #Arsenal), and the full text of the tweets, which serve as the raw material for sentiment scoring. Sentiment polarity is calculated using the Vader model, categorizing tweets as positive, neutral, or negative based on their content. Additionally, each tweet includes a timestamp, allowing for temporal analysis of sentiment trends throughout the season. By combining these elements, the dataset enables a nuanced exploration of fan sentiment and its relationship to team performance. The presence of sentiment scores and detailed timestamps allows us to analyze not only the volume of fan interactions but also the tone and timing of these reactions. This dataset forms the backbone of our study, providing the necessary information to uncover trends and correlations between Twitter sentiment and Premier League standings.

Kaggle Dataset:



created_at	file_name	# followers	# friends	group_name	location	# retweet_count
2020-07-08 2020-09-19	Arsenal 15% ManchesterUnited 15% Other (561947) 70%	0 23.1m	0 1.88m	Arsenal FC 15% Manchester Unite... 15% Other (561947) 70%	[null] 30% London, England 2% Other (542165) 68%	0 143k
2020-07-09 11:34:08	Liverpool	697325	594	Liverpool FC	Liverpool	0
2020-07-09 11:33:52	Liverpool	2348	2907	Liverpool FC	Huyton	65
2020-07-09 11:33:52	Liverpool	465	177	Liverpool FC	Accra, Ghana	32

Methods

SAS Viya Concept Rules

We tried to conduct sentiment analysis by finding matches in tweets that could be categorised as positive and negative in the context of football. We also included matches that could be categorised as “winning” and “losing,” which are terms that relate to winning and losing, also in the context of football.

Positive and Negative Terms and Positive and Negative Concept Rules

We curated a list of positive and negative terms in the context of football. The two lists are paired, which means each term in the positive or negative terms list has a counterpart in the other list. For example, in the positive terms list, we have the term “clean sheet,” which means not conceding a goal in a match. In the negative terms list, the foil to “clean sheet” is “conceded.” The two lists have roughly the same amount of terms (35 for positive, and 34 for negative). The reason behind curating our own list is because these terms are more reflective on the context of football, and maintaining a similar amount of terms on both lists would prevent over-representing one list over the other, which was an issue we had when we used the list provided on the course site.

We compiled the two lists into two SAS concepts: nlpPositiveTerms and nlpNegativeTerms. Below are the two concepts:

nlpPositiveTerms	nlpNegativeTerms
CLASSIFIER: Impressive	CLASSIFIER: Disappointing
CLASSIFIER: Brilliant	CLASSIFIER: Dreadful
CLASSIFIER: Amazing	CLASSIFIER: Terrible
CLASSIFIER: Prolific	CLASSIFIER: Wasteful
CLASSIFIER: Skillful	CLASSIFIER: Clumsy
CLASSIFIER: Incredible	CLASSIFIER: Mediocre
CLASSIFIER: Ambitious	CLASSIFIER: Complacent
CLASSIFIER: Authentic	CLASSIFIER: Fake
CLASSIFIER: Accomplished	CLASSIFIER: Amateur
CLASSIFIER: Adaptable	CLASSIFIER: Rigid
CLASSIFIER: Assertive	CLASSIFIER: Passive
CLASSIFIER: Audacious	CLASSIFIER: Timid
CLASSIFIER: Assured	CLASSIFIER: Uncertain
CLASSIFIER: Attentive	CLASSIFIER: Careless
CLASSIFIER: Commanding	CLASSIFIER: Weak
CLASSIFIER: Clean sheet	CLASSIFIER: Conceded
CLASSIFIER: Victory	CLASSIFIER: Defeat
CLASSIFIER: Champion	CLASSIFIER: Relegated
CLASSIFIER: Top scorer	CLASSIFIER: Goal drought
CLASSIFIER: Assist	CLASSIFIER: Turnover
CLASSIFIER: Clinical	CLASSIFIER: Wasteful
CLASSIFIER: Dominant	CLASSIFIER: Dominated
CLASSIFIER: Promotion	CLASSIFIER: Relegation
CLASSIFIER: Winner	CLASSIFIER: Loser
CLASSIFIER: Leader	CLASSIFIER: Struggler
CLASSIFIER: Masterclass	CLASSIFIER: Disaster
CLASSIFIER: Unstoppable	CLASSIFIER: Ineffective
CLASSIFIER: Solid	CLASSIFIER: Shaky
CLASSIFIER: World-class	CLASSIFIER: Poor
CLASSIFIER: Elite	CLASSIFIER: Unprofessional
CLASSIFIER: Professional	CLASSIFIER: Nervous
CLASSIFIER: Composed	CLASSIFIER: Erratic
CLASSIFIER: Consistent	CLASSIFIER: Static
CLASSIFIER: Dynamic	CLASSIFIER: Lethargic

CLASSIFIER: Energetic	

To make sure that the matches have not been negated and therefore representing the other sentiment, we remove the negated form of it in the final concept. The negated concepts are called nlpNegatedPositive and nlpNegatedNegative , which are shown below.

nlpNegatedPositive	nlpNegatedNegative
C_CONCEPT:nlpNegation _c{nlpPositiveTerms}	C_CONCEPT:nlpNegation _c{nlpNegativeTerms}

nlpNegation is a concept provided in the course site with a list of negations such as “didn’t,” and “cannot.”

The final concepts used are called nlpPositive and nlpNegative, and are shown below.

nlpPositive	nlpNegative
CONCEPT:nlpPositiveTerms REMOVE_ITEM:(ALIGNED, "nlpNegatedPositive","_c{nlpPositive}")	CONCEPT:nlpNegativeTerms REMOVE_ITEM:(ALIGNED, "nlpNegatedNegative","_c{nlpNegative}")

Win and Lose Terms and Win and Lose Concept Rules

Similar to positive and negative terms, we also created a list of win and lose terms. The lists are much shorter than the positive and negative terms. The two lists are also compiled into concepts: nlpWinTerms and nlpLoseTerms.

nlpWinTerms	nlpLoseTerms
CONCEPT::win@ CONCEPT:victory@ CONCEPT:victorious CONCEPT:dominate@	CONCEPT:lose@ CONCEPT:loss@ CONCEPT:lost@ CONCEPT:loser@ CONCEPT:defeat@

We again check for negations before compiling them into the final concepts: nlpWin and nlpLose. The intermediate concepts for negations are called nlpWinNegated and nlpLoseNegated.

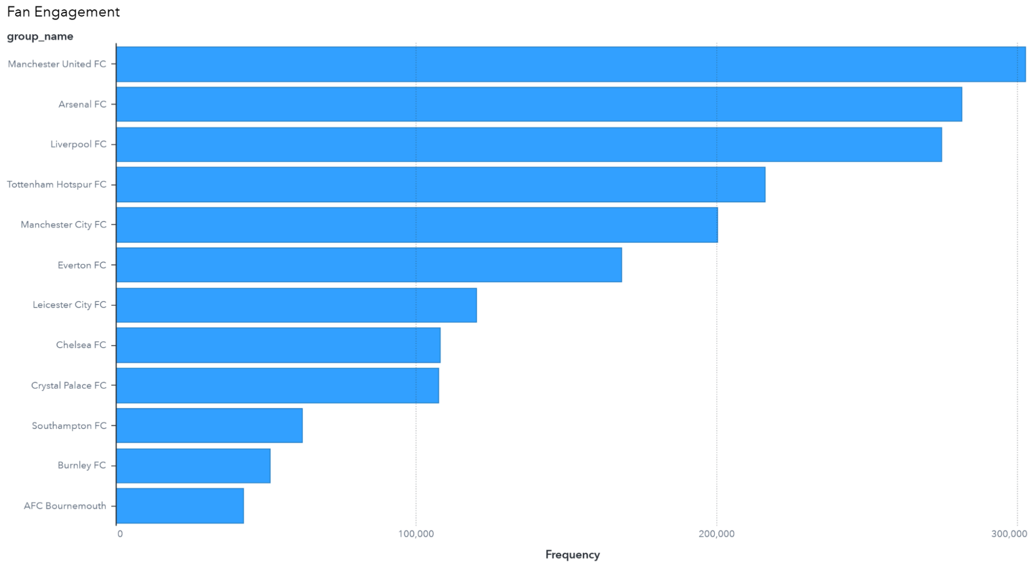
nlpWinNegated	nlpLoseNegated
C_CONCEPT:nlpNegation _c{nlpWinTerms}	C_CONCEPT:nlpNegation _c{nlpLoseTerms}

nlpWin	nlpLose
CONCEPT:nlpWinTerms REMOVE_ITEM:(ALIGNED, "nlpWinNegated","_c{nlpWin}")	CONCEPT:nlpLoseTerms REMOVE_ITEM:(ALIGNED, "nlpLoseNegated","_c{nlpLose}")

Retrieving Results from SAS Viya

1. Fan Engagement

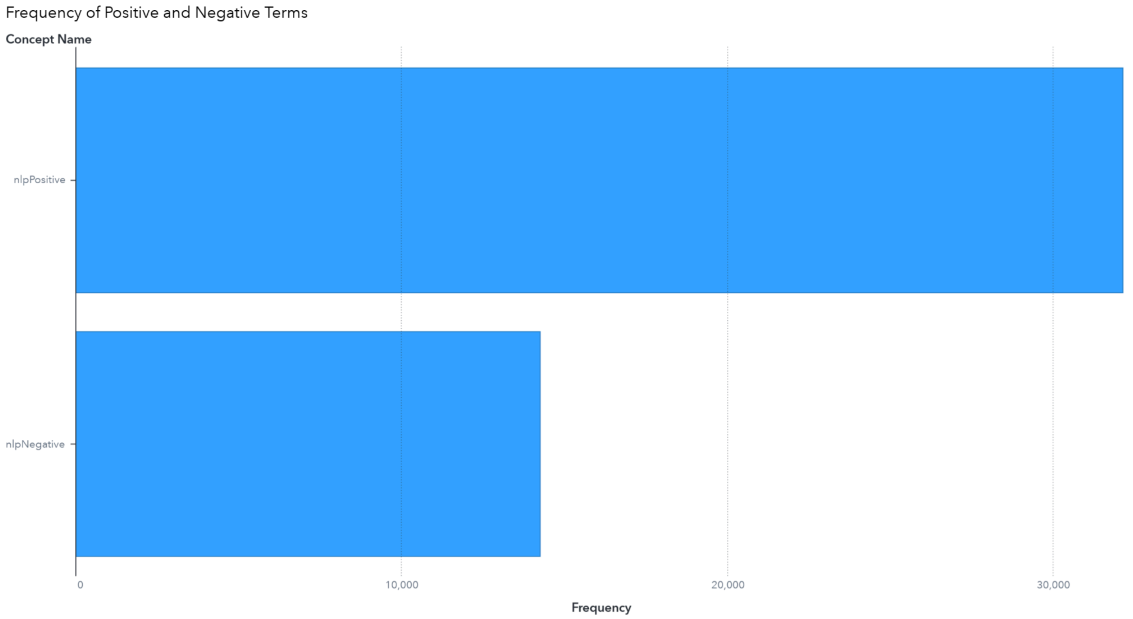
We retrieved the results of the matches from the results of the concept node on SAS Viya. First, we retrieved the results of fan engagement, which is used to measure the activity of fans from each team. Fan engagement is measured by the number of tweets attributed to each team. Below is a bar chart comparing fan engagement of different teams.



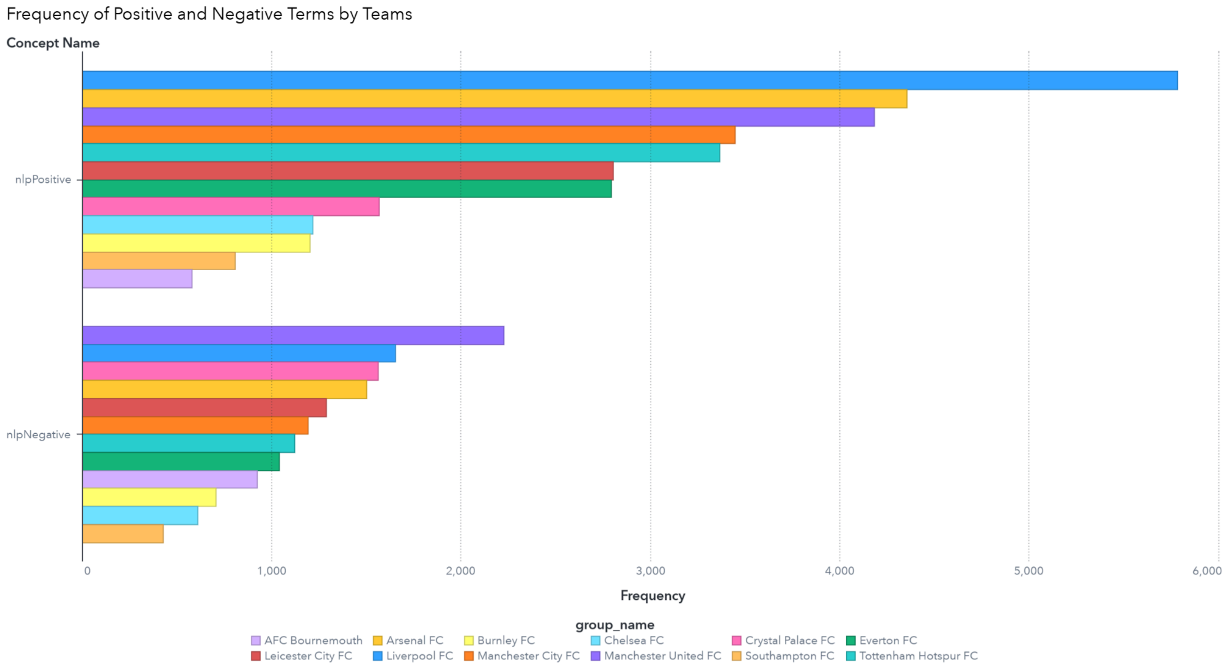
As we can see, teams that are very well-known and popular globally such as Manchester United, Arsenal, and Liverpool have high fan engagement, whereas smaller teams like Bournemouth and Burnley have lower fan engagement.

2. Frequency of Positive and Negative Terms

The frequency of positive and negative terms are retrieved from the matches of the tweets from nlpPositive and nlpNegative. From the bar chart below, we can observe that there are more positive than negative matches. This is likely because fans use social media platforms to share the positive moments of their teams, for example, the individual brilliances of a player during a game.

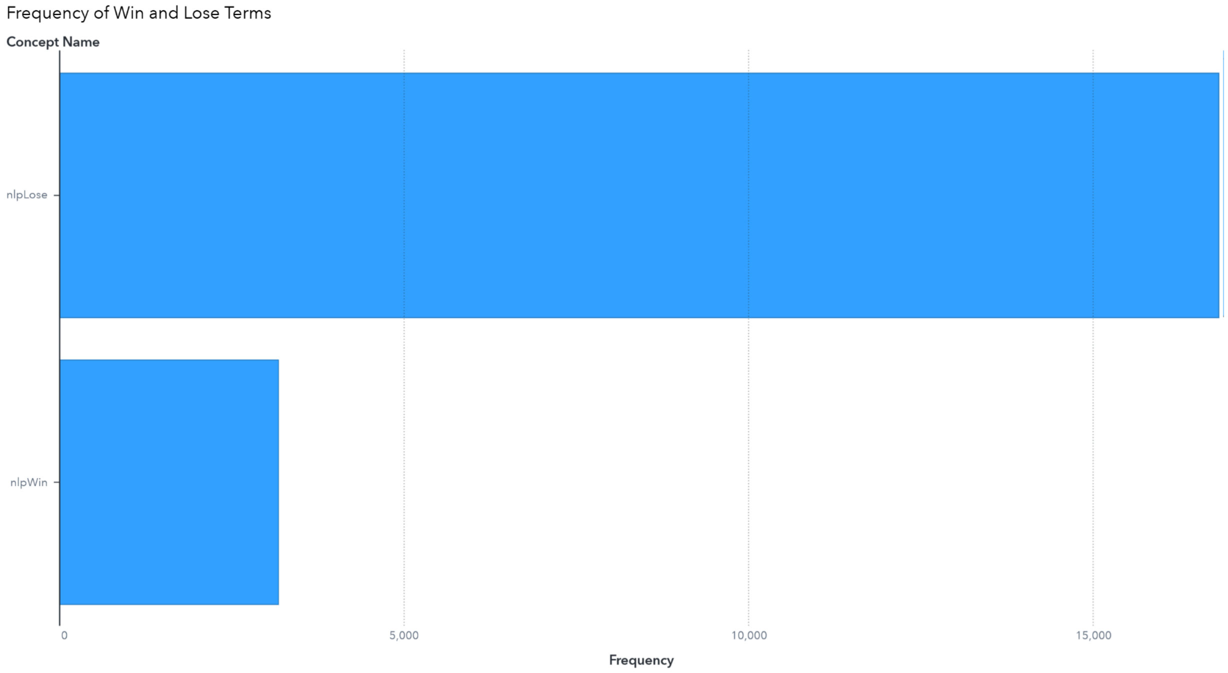


We can also see the frequency of positive and negative terms of each team below. Liverpool, which had a great start to the season, has the most positive tweets. Manchester United, on the other hand, struggled in the beginning of the season, and therefore has the most negative tweets. However, we must acknowledge that the number of positive and negative tweets is also affected by the overall fan engagement, so whilst one team may have more negative tweets than another, it may be because the team has a much larger fanbase than the team with less negative tweets.

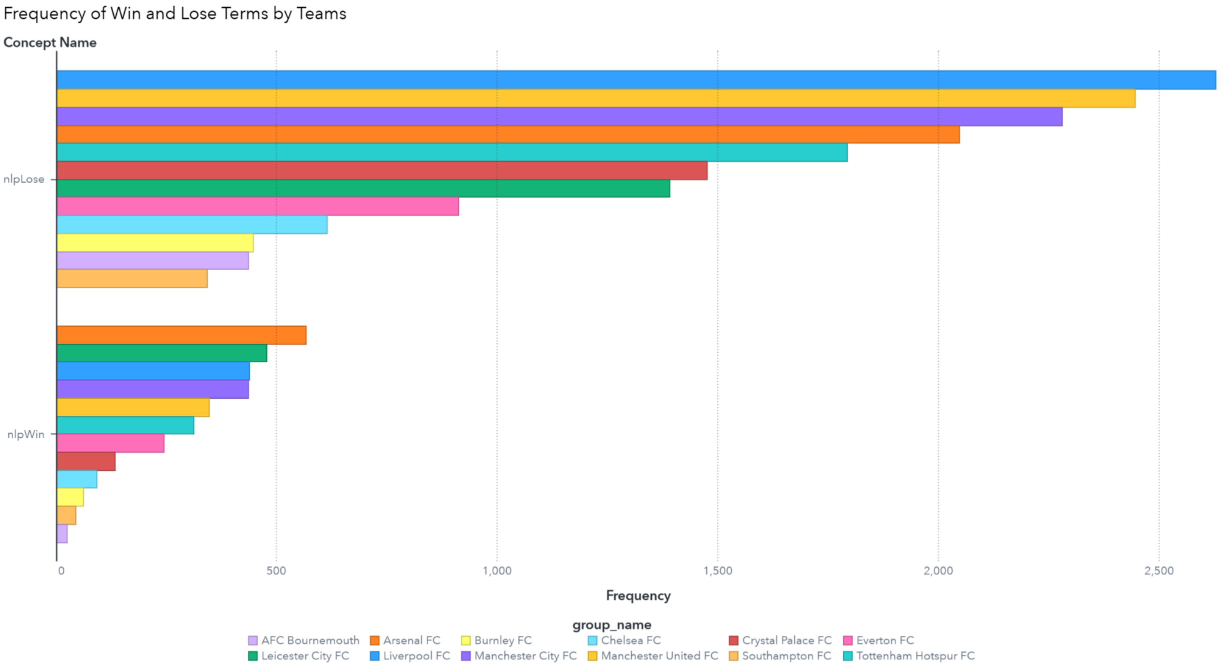


3. Frequency of Win and Lose Terms

The frequency of win and lose terms are retrieved from the matches of the tweets from nlpWin and nlpLose. We can see from the bar chart below that there are a lot more lose matches than win matches. This is likely because even though fans share more of the positives of their teams, they also react more enthusiastically to a loss than to a win.



Looking at the frequency of win and lose terms for each team, we can see that Liverpool fans reacted heatedly to their losses, despite their good start to the season. Manchester United, who had a slow start to the season, have a lot of lose matches. Everton, on the other hand, had won all their games during the period when the tweets were collected. They rank 8th on the lose matches, and 7th on the win matches, which suggests that they have a better win-to-lose tweet ratio than other teams.



Encoding Data

After retrieving all the data, we encoded the data for each team into Python objects called PremTeam. Each PremTeam object has a team name, engagement, frequency of positive terms, frequency of negative terms, frequency of win terms, frequency of lose terms, ranking as of matchday 4 (matchday of the season when the data collection concluded), and ranking at the end of the season. Below is the format of PremTeam for the team Liverpool.

```
liverpool = PremTeam(team_name="Liverpool FC", engagement=274707,
frq_pos=5784, frq_neg=1654, frq_win=439, frq_lose=2628, rank_md4=4,
rank_eos=3)
```

We can then test out different ranking criteria based on the different fields of data we have and see which ranking criteria best matches the rankings as of matchday 4 or at the end of the season.

Ranking Criteria

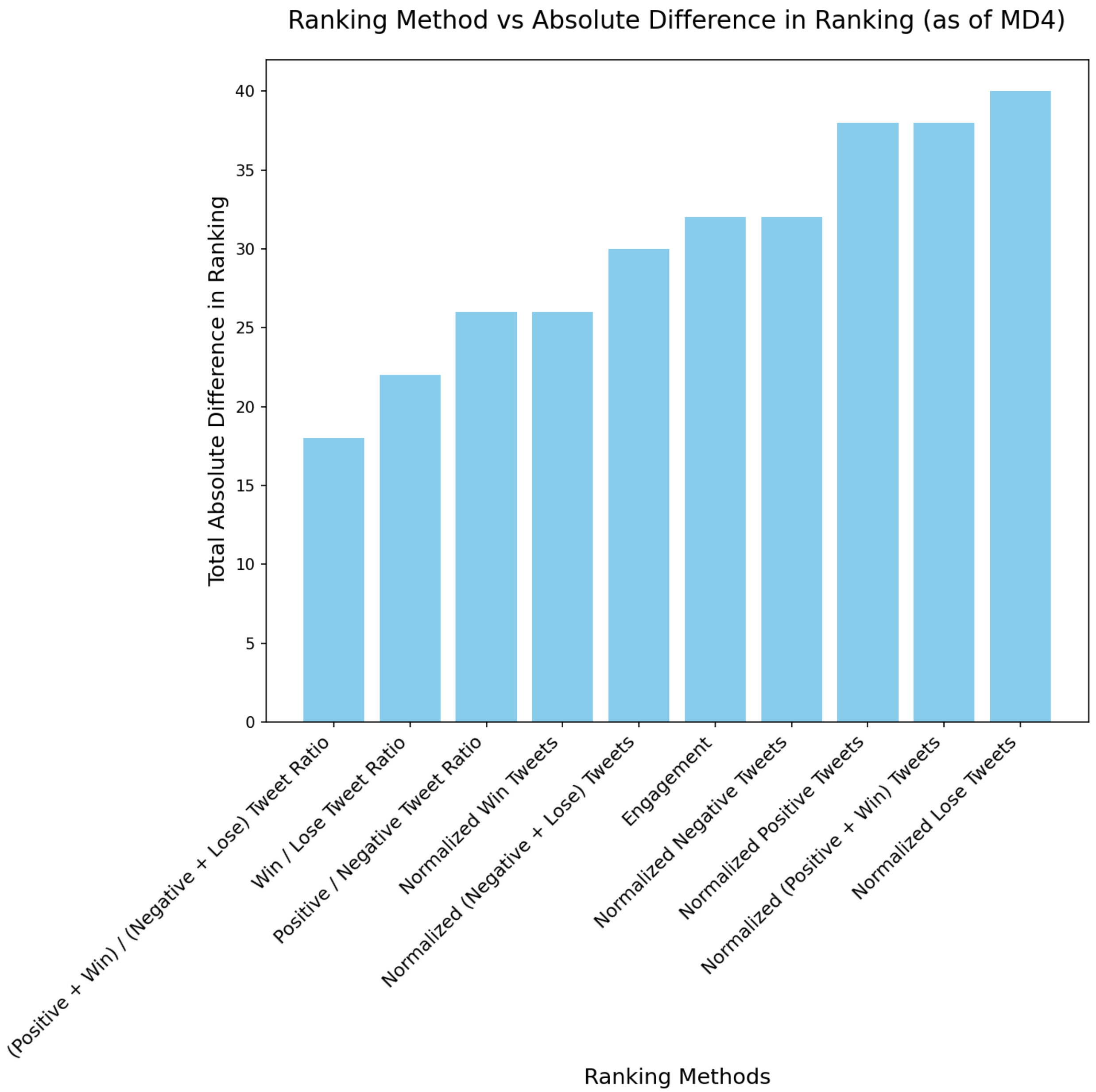
We propose two types of ranking criteria: engagement-based and ratio-based.
With engagement-based criteria, the total fan engagement for each team is factored in. The most basic one is Engagement, which ranks teams by the number of tweets attributed to each team. Then, we have the Normalised Positive Tweets, which ranks the teams by the number of positive tweets divided by the total number of tweets. Normalised Negative Tweets works similarly to Normalised Positive Tweets, except that the ranking is in ascending order, meaning the smaller the value, the higher the team ranks. In similar fashion, we have Normalised Win Tweets, Normalised Lose Tweets (ascending order), Normalised (Positive + Win) Tweets, and Normalised (Negative + Lose) Tweets (ascending order).

With ratio-based criteria, we rank the teams by the ratio of Positive and/or Win and Negative and/or Lose tweets. Positive / Negative Tweet Ratio takes the ratio of positive to negative tweets; Win / Lose Tweet Ratio takes the ratio of win to lose tweets; and (Positive + Win) / (Negative + Lose) Tweet Ratio takes the ratio of positive and win tweets to negative and lose tweets.

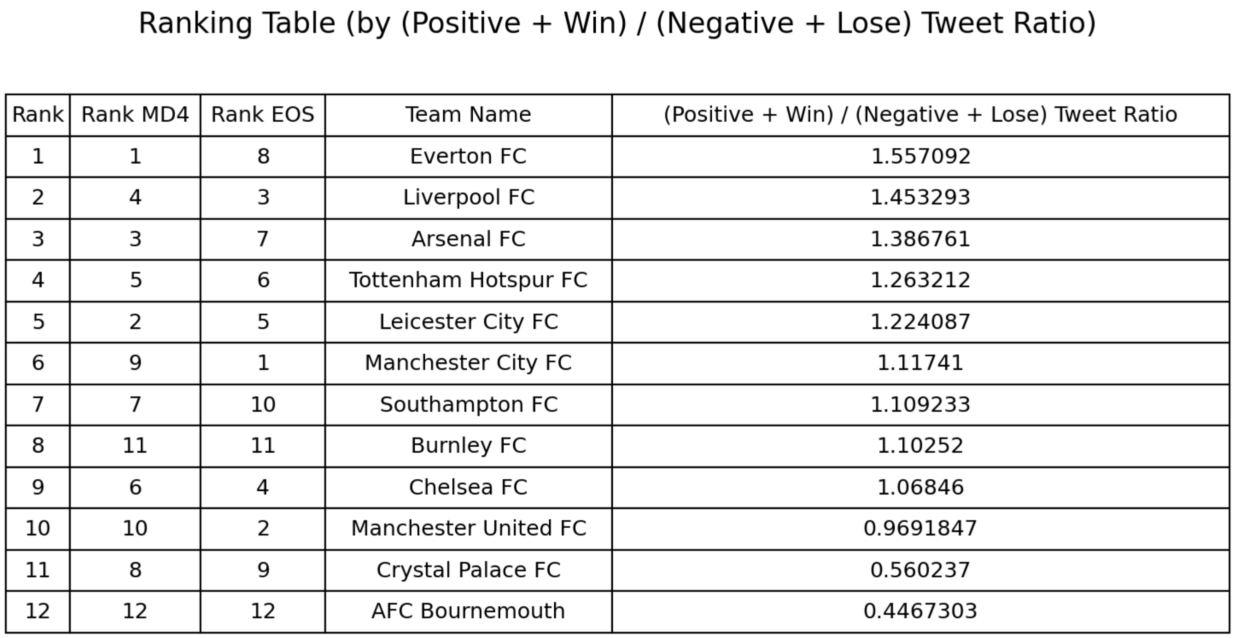
Results

Rankings as of Matchday 4

To evaluate the rankings we proposed above, we summed the absolute difference between the actual standings and the rankings each criterion yielded. Below is the result of the ranking criteria evaluated up to matchday 4 (MD4).



As we can see from the graph, the best criterion to predict standings of the league at a given point of time during the season is the ratio-based (Positive + Win) / (Negative + Lose) Tweet Ratio criterion. The table below illustrates how each team is ranked.



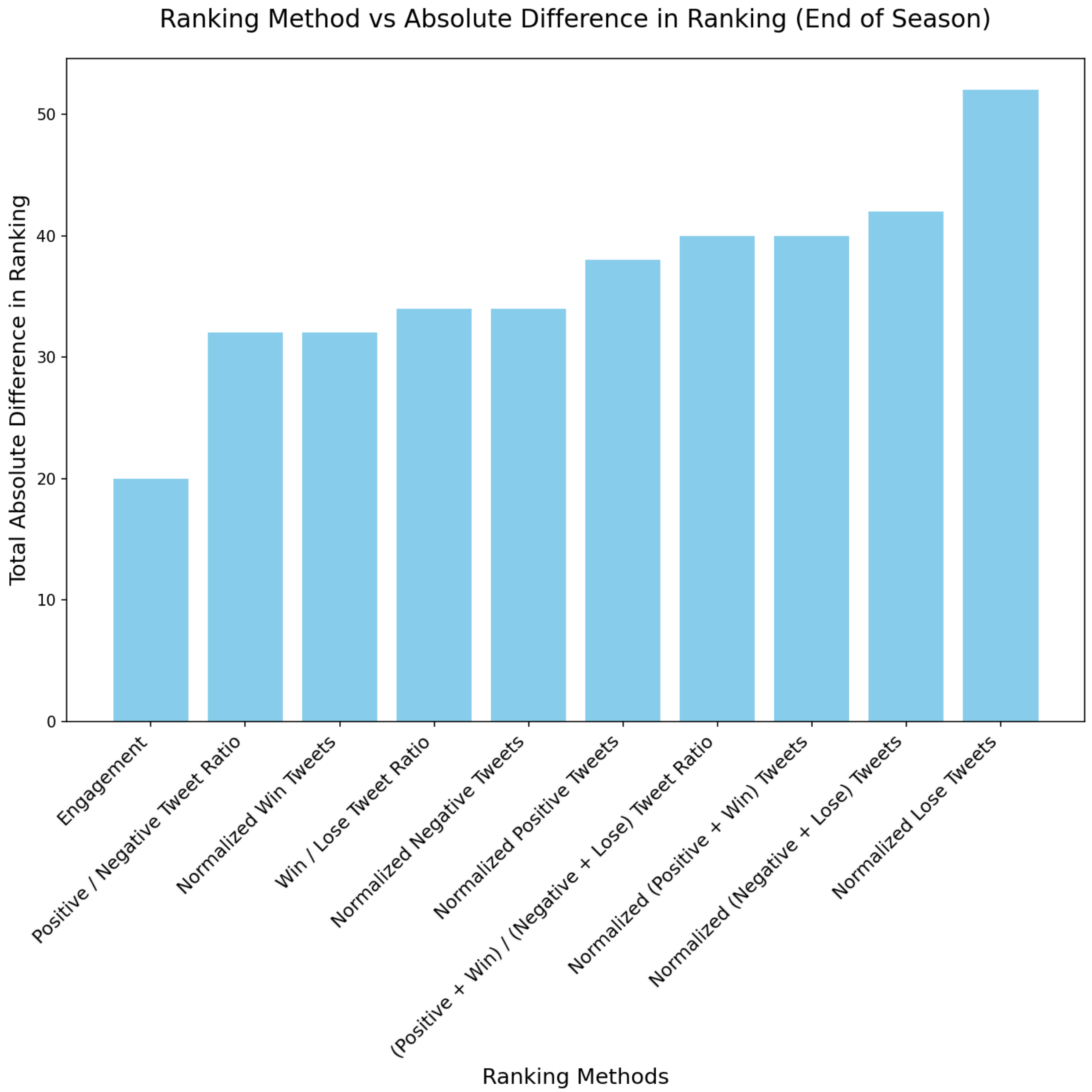
Ranking Table (by (Positive + Win) / (Negative + Lose) Tweet Ratio)

Rank	Rank MD4	Rank EOS	Team Name	(Positive + Win) / (Negative + Lose) Tweet Ratio
1	1	8	Everton FC	1.557092
2	4	3	Liverpool FC	1.453293
3	3	7	Arsenal FC	1.386761
4	5	6	Tottenham Hotspur FC	1.263212
5	2	5	Leicester City FC	1.224087
6	9	1	Manchester City FC	1.11741
7	7	10	Southampton FC	1.109233
8	11	11	Burnley FC	1.10252
9	6	4	Chelsea FC	1.06846
10	10	2	Manchester United FC	0.9691847
11	8	9	Crystal Palace FC	0.560237
12	12	12	AFC Bournemouth	0.4467303

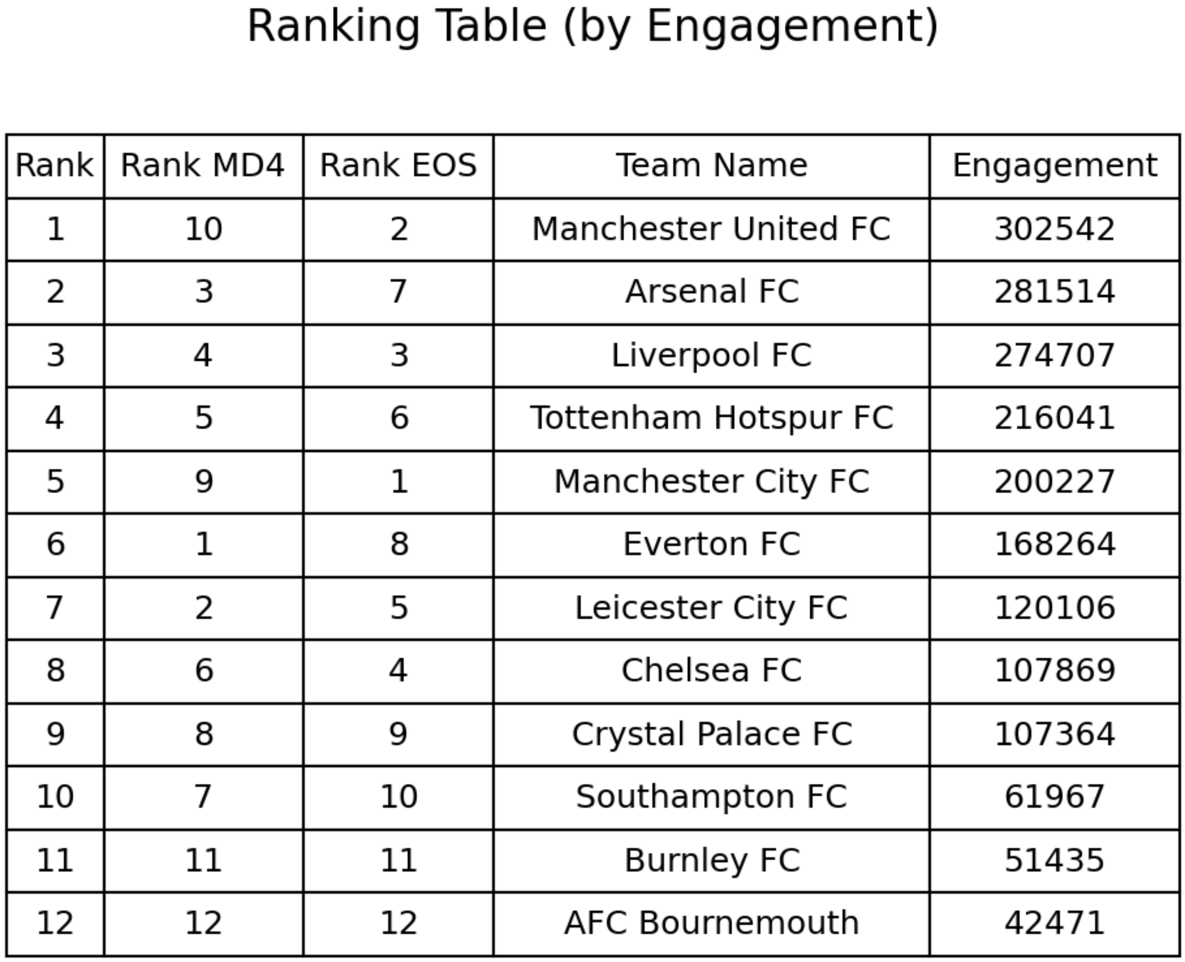
We can see that Everton, who has the highest ratio, is ranked first in this metric. This matches their standings as of matchday 4. Manchester United, which struggled early on in the season, were ranked 10th as of matchday 4. Their rank according to the ratio also put them low amongst the other teams. This ranking criterion makes sense for the task of predicting the rankings as of a current point in the season as the good and bad tweets often reflect the current form of the team.

Ranking at the End of the Season

Below is the result of the ranking criteria evaluated for the end of the season.



As we can see in the graph, it was in fact fan engagement that has the best performance out of all the criteria that we test for the rankings at the end of the season. The table below shows how the teams are ranked according to fan engagement.



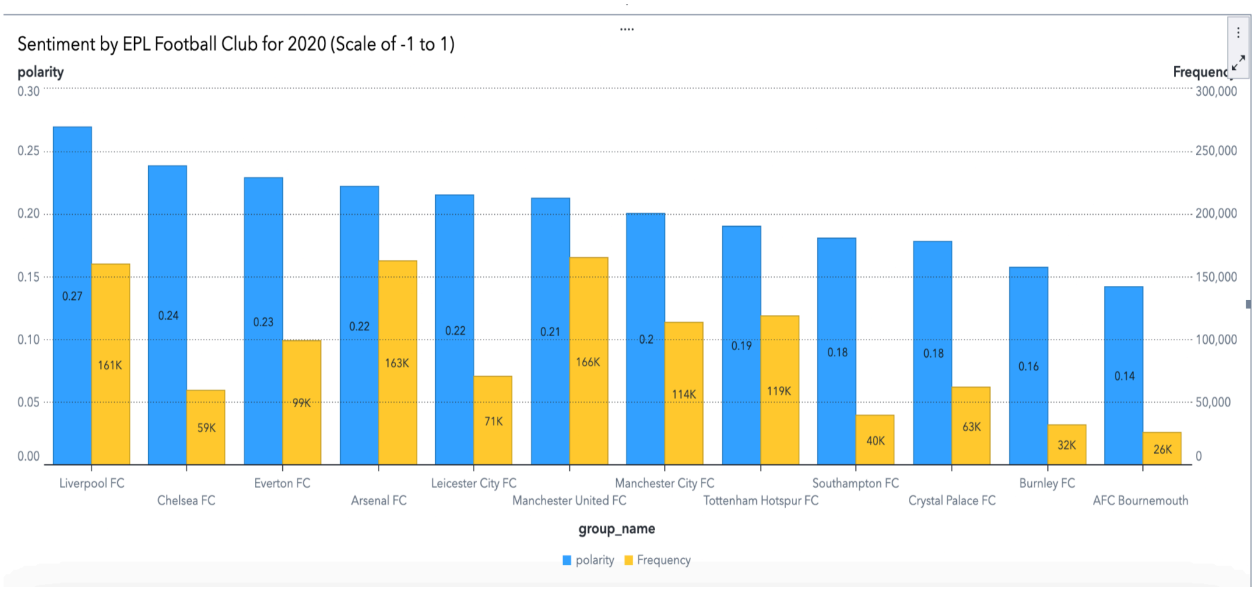
Ranking Table (by Engagement)

Rank	Rank MD4	Rank EOS	Team Name	Engagement
1	10	2	Manchester United FC	302542
2	3	7	Arsenal FC	281514
3	4	3	Liverpool FC	274707
4	5	6	Tottenham Hotspur FC	216041
5	9	1	Manchester City FC	200227
6	1	8	Everton FC	168264
7	2	5	Leicester City FC	120106
8	6	4	Chelsea FC	107869
9	8	9	Crystal Palace FC	107364
10	7	10	Southampton FC	61967
11	11	11	Burnley FC	51435
12	12	12	AFC Bournemouth	42471

We can see that the clubs that are more well-known across the world such as Manchester United, Arsenal, and Liverpool, are ranked high in engagement, which we have previously discussed before. Fan engagement being the best criterion to estimate the standings at the end of the season is a little surprising as it is the simplest metric of the ones tested, yet it does make sense. One of the reasons for the teams' high fan engagement can be attributed to the teams' past success. The same reason correlates to the teams' ranking at the end of the season, as traditionally strong teams are more likely to finish higher in the league as the season concludes. This shows the wisdom in this commonly used phrase in football: "Form is temporary but class is permanent."

Vader Model Sentiment Analysis

In order to accurately measure the sentiment for EPL (English Premier League) teams on twitter, we use a modified Vader model to get polarity scores between -1 and 1. Basically, we take into account the negative words, positive words, and some other more complex modifiers to make a holistic assessment of each tweet. Each tweet is assigned a score, and we can then use these scores to analyze each team's sentiment score. Before graphing our data, we have to categorize each tweet by team. We look for instances of each team as well as a lack of mention of other teams, as well as what team each tweet is sent in response to. Once all tweets have been categorized, we run our pipeline in SAS, and generate a Sentiment report. See below for a graph with each team-related tweet on the x-axis and the sentiment represented as polarity on the y-axis.



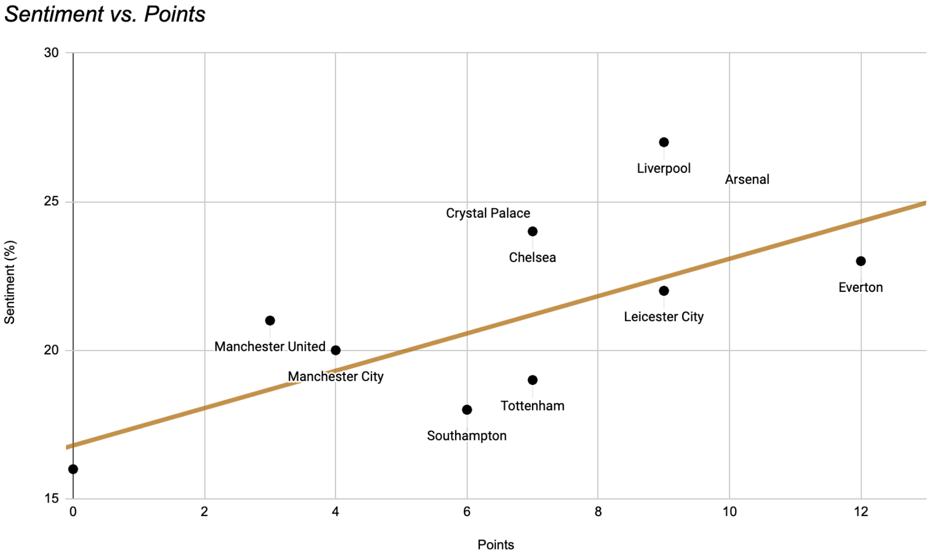
As you can see from the graph, Liverpool has by far the most positive sentiment, with AFC Bournemouth and Burnley having the least positive sentiment. On the other part of the bar chart, highlighted in yellow, we see the frequency of each team, or how many tweets each team has. The most popular teams are Liverpool, Arsenal, and Manchester United, each with over 160k tweets. The least popular teams are Southampton, Burnley, and AFC Bournemouth, each with less than 50k tweets.

In order to compare how match results affected the sentiment of each team, we pulled the match results for all 12 teams listed during the time-frame of the tweets. As can be seen from the table below, one team is conspicuously absent: AFC Bournemouth. More on them later.

Place	**Team**	**W**	**Tie**	**L**	**Points**	**Sentiment (%)**	**Place**	**Team**	**W**	**Tie**	**L**	**Points**	**Sentiment (%)**

1	Everton	4	0	0	12	23
2	Leicester City	3	0	1	9	22
3	Arsenal	3	0	1	9	22
4	Liverpool	3	0	1	9	27
5	Tottenham	2	1	1	7	19
6	Chelsea	2	1	1	7	24
7	Southampton	2	0	2	6	18
8	Crystal Palace	2	0	2	6	18
9	Manchester City	1	1	1	4	20
10	Manchester United	1	0	2	3	21
11	Burnley	0	0	3	0	16

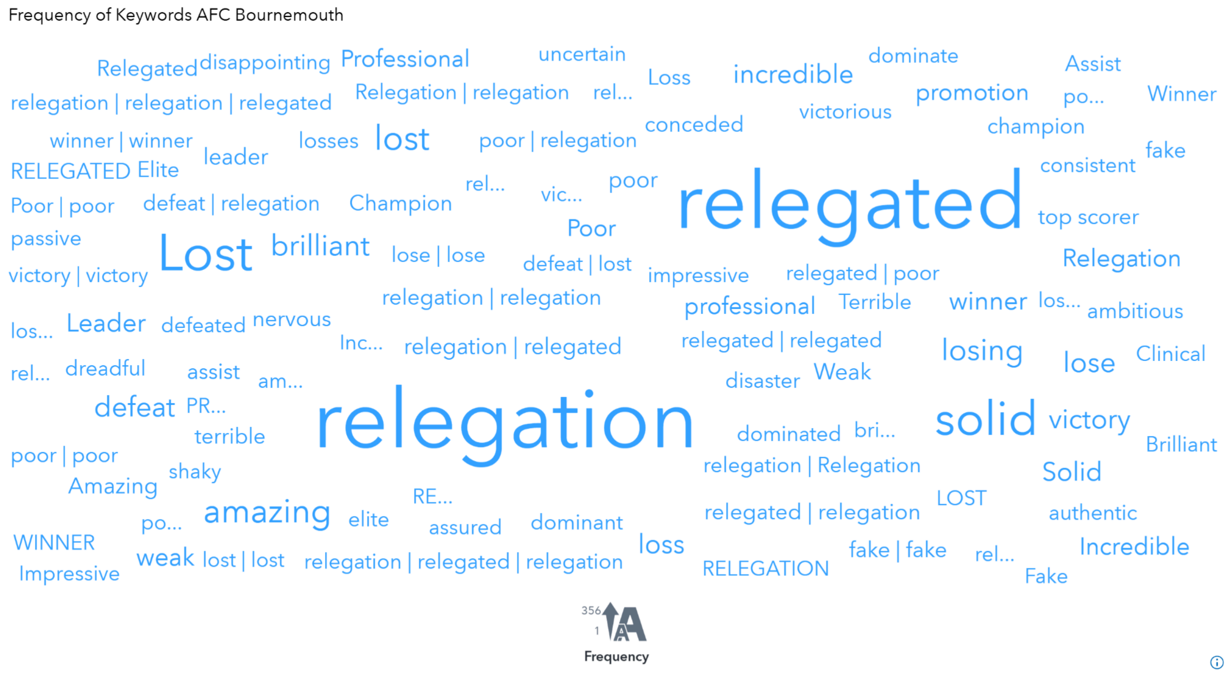
Everton is off to a hot start here, as the only team that won all their games during this time-frame. Leicester City, Arsenal, and Liverpool are hot on their heels, each with three wins. Towards the bottom, we have Manchester City with one win, tie, and loss, and Manchester United with one win and two losses. At the bottom, Burnley has three losses and zero wins. In order to visualize the match results versus sentiment for each team, we can graph the points and the sentiment on a scatter plot.



This chart plots the sentiment of each team on the y-axis, and the total points in the time-frame on the x-axis. We then can draw a regression line to see approximately what kind of correlation we are looking at here. The line has a slightly positive slope, which implies that there may be some correlation between match results and team sentiment. Most of the teams follow this line of best fit, with Everton as one of the most positive team sentiments, and Burnley with one of the least positive team sentiments. Liverpool is a notable outlier here, as they have by far the most positive team sentiment, but are not leading the league during that time frame. According to the regression line, their sentiment is much higher than expected.

Bournemouth Analysis

Bournemouth Word Cloud

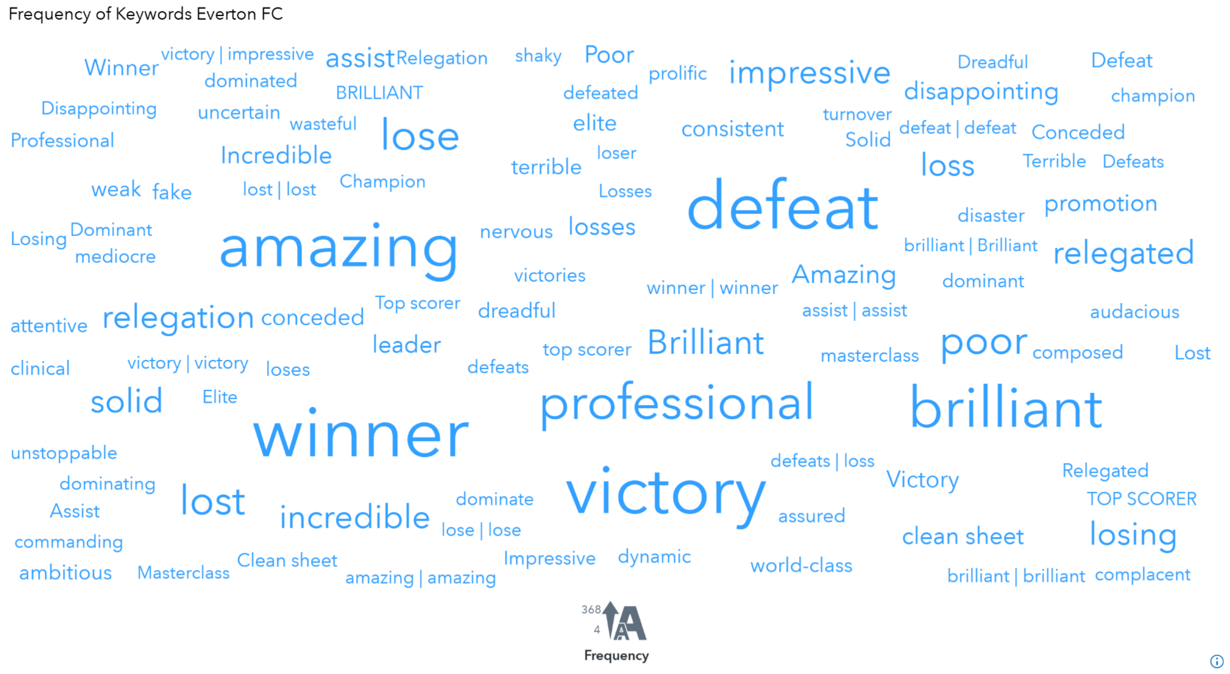


Referring back to AFC Bournemouth, why weren't they included in the match results table we had above? The simple answer? They weren't in the league at the time. After the previous season Bournemouth had been relegated, which means that they were in the bottom three teams in the standings at the end of last season, and were kicked or "relegated" down to a lower league. Interestingly, however, they were still included in the data set we received. Whether it was a mistake or not, we can use the data to analyze fan sentiment for a relegated team. Unsurprisingly, the sentiment for AFC Bournemouth was the least positive out of all the analyzed teams which reveals several interesting trends. Firstly, their relegation likely contributed to the overwhelmingly negative sentiment associated with their team. Fans may have expressed disappointment, frustration, and anger over their team's performance, leading to Bournemouth's sentiment score being the lowest among the analyzed teams. This is consistent with the theory that fan sentiment is heavily influenced by a team’s performance on the field, and relegation represents the ultimate underperformance.

Furthermore, Bournemouth’s inclusion in the dataset despite their relegation provides a lens to study the resilience and engagement of fan bases for lower-tier teams. While their total volume of tweets was among the lowest in the dataset, the persistent negativity suggests that fans were still vocal about their dissatisfaction even after the team’s relegation. Bournemouth’s case highlights potential limitations in sentiment analysis for sports data. Since Bournemouth was not active in the Premier League during the dataset’s timeframe, much of the negativity in their sentiment score may stem from lingering disappointment from the previous season or frustration with other off-field factors. For future research, a more nuanced approach could account for these dynamics, perhaps by analyzing tweet contexts or comparing sentiment trends across relegated teams to identify common patterns.

Everton FC Analysis

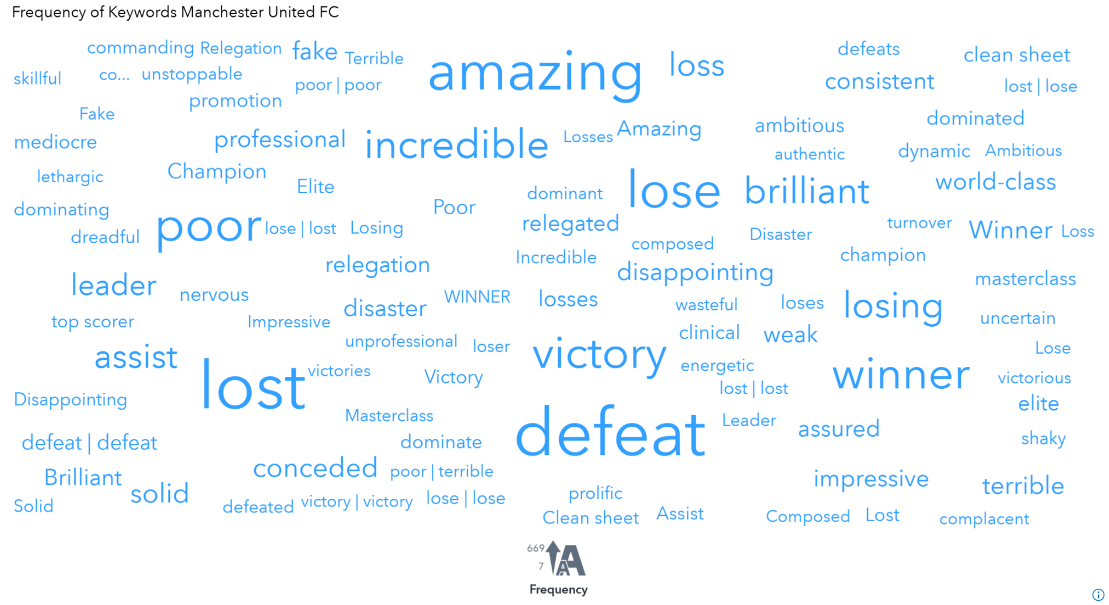
Everton FC Word Cloud



Everton FC was leading after Matchday 4 (MD4) as the league leaders with a flawless record of four wins in four games. Despite their on-field success, our analysis reveals a mixed sentiment picture on Twitter. The word cloud for Everton highlights predominantly positive keywords such as “amazing,” “winner,” “victory,” and “brilliant.” These terms showcase a fan base elated with the team's performance at the top of the table. However, the presence of negative terms like “defeat” and “lose” suggests pessimism among supporters, or perhaps lingering feelings from previous seasons. Everton's fan engagement on Twitter ranks only sixth, trailing teams with less successful starts to the season. This discrepancy raises important questions about the factors driving fan engagement, which may not solely depend on match results. Everton’s case highlights the dynamic nature of fan sentiments, where performance on the field does not always translate into proportional online enthusiasm. While their points tally aligned with their sentiment score during this timeframe, their performance dipped towards the end of the season, which may explain the tempered engagement levels observed on Twitter. Everton’s dominance at MD4 reinforces the strong correlation between team performance and positive sentiment, as seen in their 23% sentiment score. Yet, their drop in engagement compared to their table-topping performance hints at a more complex relationship between fan sentiment, team history, and engagement metrics. This complexity emphasizes the need for further studies into how external factor, such as historical team rivalries or specific fan base dynamics, influence social media trends in the Premier League.

Manchester United Analysis

Manchester United Word Cloud

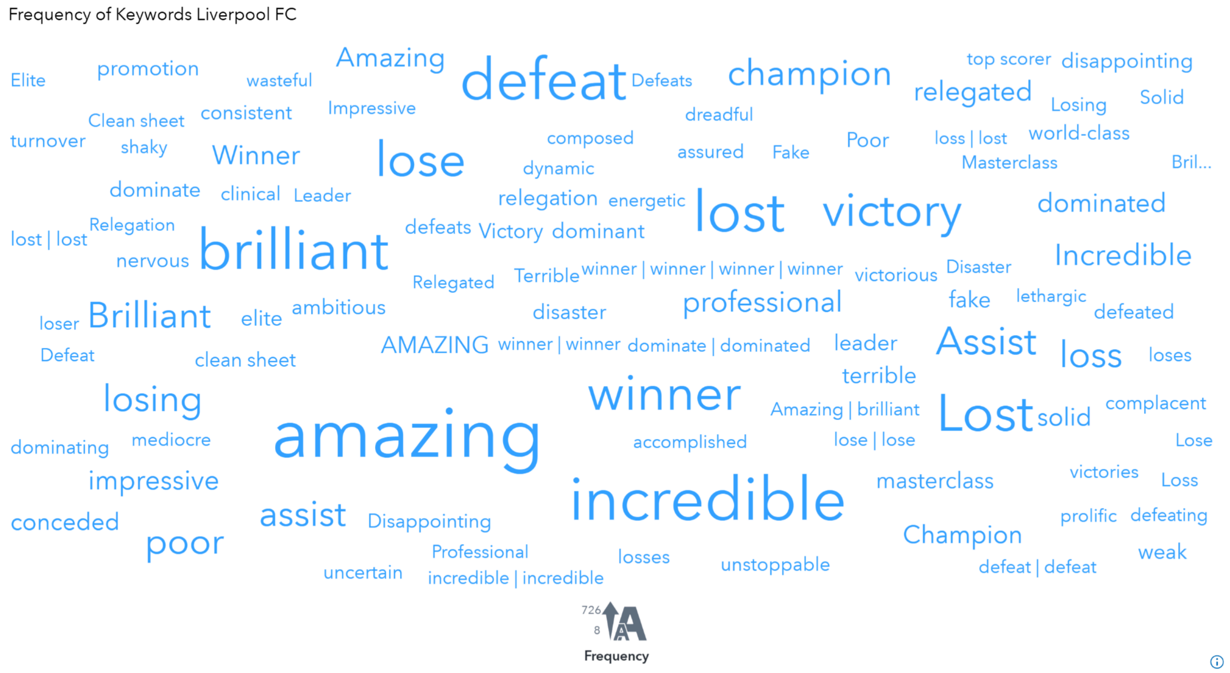


Unlike Everton, Manchester United’s start to the 20/21 season was defined by significant struggles, as shown by their position at 16th in the league after Matchday 4 (MD4). Their early-season form led to a strikingly negative sentiment among fans on Twitter, with keywords such as “lost,” “losing,” “defeat,” and “poor” dominating the word cloud. These terms illustrate widespread frustration and disappointment, reflecting the fans’ dissatisfaction with their team’s underwhelming performances on the pitch. Despite this negativity, Manchester United's engagement on social media remained notably high. The sheer volume of tweets discussing their performance highlights the passionate and vocal nature of their fan base, even in moments of crisis. This high engagement may stem from the club's long history and global reach, where any result generates significant conversation, whether positive or negative. The presence of positive terms such as “amazing” and “brilliant” in the word cloud, although in smaller frequency, suggests that some fans or rival supporters may have been highlighting isolated moments of brilliance or sarcastically referencing the team's struggles. This dual nature of sentiment underscores the polarized reactions that Manchester United often evoke, further amplified by their status as one of the most scrutinized clubs in the league.

From a broader perspective, Manchester United’s poor start and negative sentiment provide an important counterpoint to Everton’s early success and positive sentiment. Where Everton’s on-field performances translated into optimism and positivity, Manchester United’s underperformance fueled widespread discontent. However they eventually resurged to finish second at the end of the season. This analysis highlights the importance of analyzing sentiment within the broader context of a team’s reputation, history, and fan expectations.

Liverpool FC Analysis

Liverpool FC Word Cloud



Liverpool FC, starting the season in decent form, was ranked 5th in the league after Matchday 4 (MD4). Their performance was marked by three wins and one loss, reflected in the diverse sentiment captured in their word cloud. Positive keywords such as “brilliant,” “amazing,” “winner,” “incredible,” and “victory” are prominent, showcasing fan optimism and appreciation for the team’s achievements. However, negative terms like “lose,” “defeat,” and “lost” highlight moments of frustration likely tied to their single loss during this period.

Liverpool’s high fan engagement on Twitter stands out as a significant factor, consistent with their global reputation and passionate fan base. Their sentiment trends emphasize the complex relationship between results on the field and the reactions of their supporters. Despite their position outside the top four during MD4, Liverpool maintained a sentiment score of 27%, one of the highest among all teams. This positivity aligns with the fans' long-term faith in the team's ability to recover and succeed, ultimately validated by their 3rd-place finish at the end of the season. Interestingly, the balance of positive and negative sentiments in the word cloud suggests that Liverpool fans are vocal in both victory and defeat. Positive terms dominate the discussion, but the presence of loss-related words underscores that even a single misstep can spark strong reactions from the fan base. This duality highlights the intensity of Liverpool's following, where every result is met with high engagement and passionate commentary. Liverpool's sentiment profile demonstrates the enduring connection between on-field performance and fan perception. It also underscores how historical success and strong global branding can sustain fan enthusiasm and positive sentiment, even when results momentarily dip. Their resilience and consistency make them a compelling case study in understanding how social media sentiment reflects and influences team dynamics in the Premier League.

Limitations and Future Work

This study demonstrates the potential of social media sentiment analysis in understanding trends in sports engagement and performance. However, it is limited by the reliance on Twitter data, which may not capture the full range of fan interactions across other platforms such as Instagram, Reddit, or Facebook. The demographic skew of Twitter users also limits the generalizability of the findings. Additionally, the sentiment analysis model, while effective, struggles with nuances like sarcasm, irony, or exaggeration, which are prevalent in sports-related language, leading to possible misclassification of sentiments. Another limitation is the snapshot nature of the analysis, which focuses on specific timeframes and does not account for real-time shifts in sentiment caused by events like match results, player injuries, or managerial changes. This restricts the ability to capture the dynamic and evolving nature of fan sentiment.

Future work could expand the scope of data sources to include other social media platforms, forums, and traditional media, offering a more holistic view of fan sentiment. Integrating match statistics, player performance data, and financial metrics would provide richer context and strengthen the analysis. Refining sentiment models with advanced natural language processing techniques, could improve accuracy, particularly for sarcasm and complex language. Multilingual support would enable broader applicability across diverse fan bases. Additionally, exploring the predictive power of sentiment analysis could open new avenues. For instance, analyzing whether shifts in sentiment correlate with team performance, ticket sales, or fan engagement trends could provide actionable insights for sports organizations. These directions would transform sentiment analysis from an observational tool into a strategic resource for decision-making in the sports industry. In conclusion, while challenges remain, sentiment analysis offers a unique perspective on fan dynamics and broader trends in sports culture. As tools and methodologies improve, this approach has the potential to significantly deepen our understanding of public sentiment and its implications across various domains.

References

Kaggle, EPL Teams - Twitter Sentiment Dataset. Retrieved from https://www.kaggle.com/datasets/wjia26/epl-teams-twitter-sentiment-dataset