

RITDataBI

Sentiment Mining Project

Ishan Shah

1. Introduction:

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique used to determine the sentiment or opinion expressed in a piece of text. The goal is to identify whether the text expresses a positive, negative, or neutral sentiment. Sentiment analysis has various applications, including social media monitoring, customer feedback analysis, and product reviews analysis.

The intersection of gaming communities and popular franchises often gives rise to dynamic and diverse discussions. In this context, the PlayStation subreddit serves as a hub for enthusiasts to share opinions and experiences. This report delves into the sentiments expressed within comments containing the keyword "Spiderman" in the PlayStation subreddit. The goal is to unravel the community's emotional response and provide valuable insights into the reception of Spiderman-related content. We will use R, a powerful and versatile programming language for statistical computing and data analysis, offers several libraries and tools for sentiment analysis.

2. Data collection: (Source file: RedditExtractor.R)

R was used for data collection. The R code utilized the RedditExtractoR library to access the Reddit API, collecting comments related to "spiderman" in the PlayStation subreddit. After filtering for the subreddit, the code retrieved and compiled the comment data to be stored in structured dataset, named "ps5_comments.csv," that facilitated subsequent sentiment analysis.

3. Sentiment Analysis: (Source file: sentiment analysis.R)

After importing the comment data from 'ps5_comments.csv,' sentiment scores were computed using different methods—'syuzhet,' 'afinn,' 'bing,' and 'nrc.' The sentiment scores were then appended to the original dataset, creating 'withsentiment.csv.' This comprehensive approach allowed for a nuanced understanding of sentiment variations in Spiderman-related discussions, enabling a more informed exploration of community attitudes and preferences within the PlayStation subreddit.

As shown by figure 1, the correlation between all 4 libraries is high.

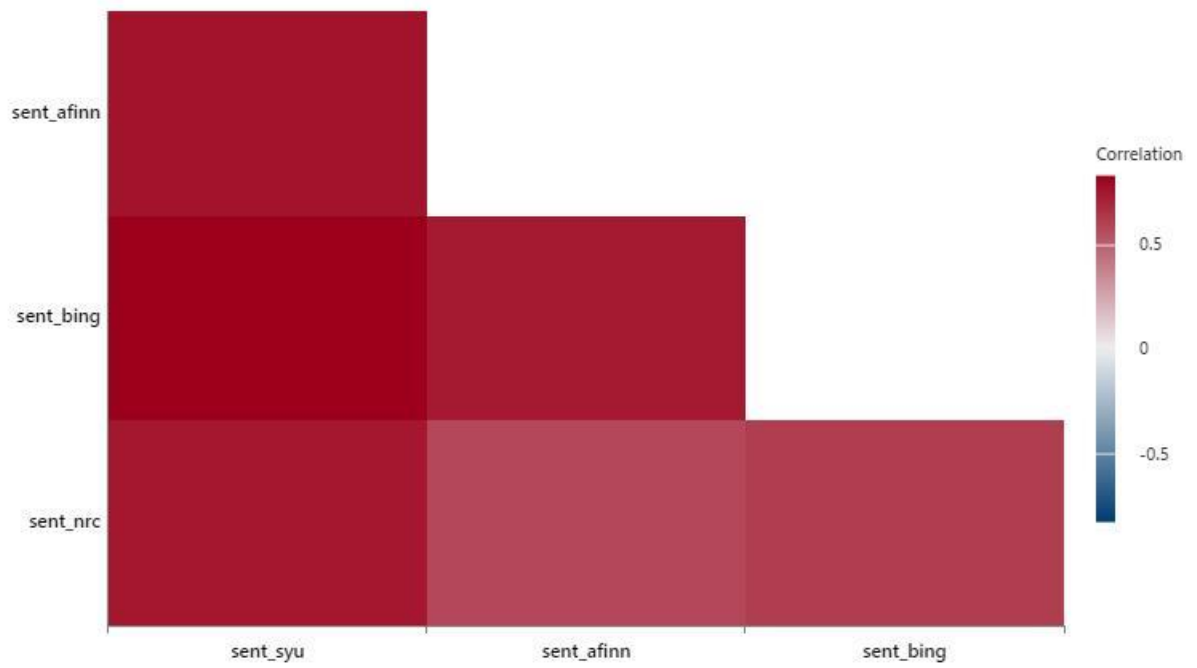


Figure 1: Correlation between libraries

Sentiment Analysis v/s Score

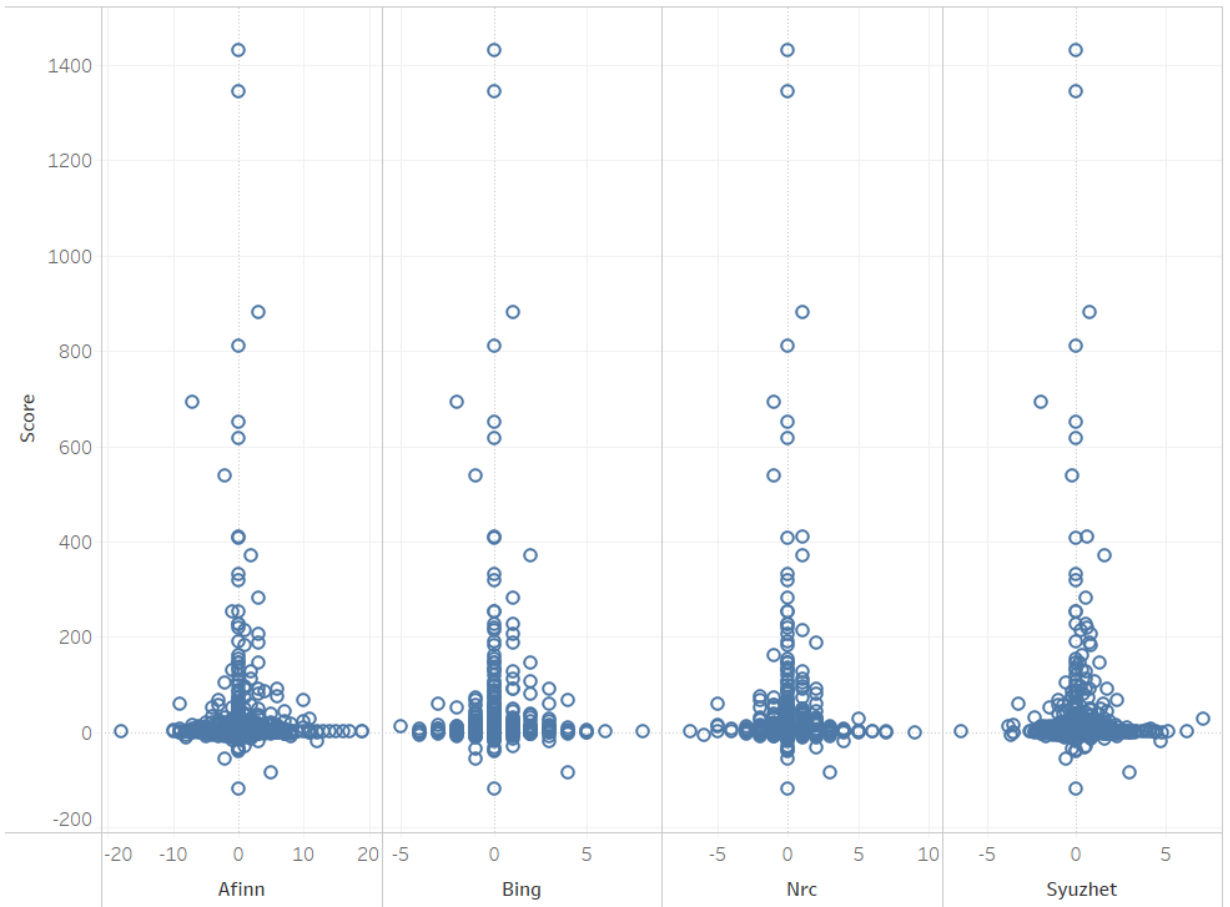


Figure 2: Sentiment Analysis v/s Score

Using figure 2, I observed a trend where higher scores correlate with more neutral sentiments. Here are a few potential explanations:

1. **Popularity and Agreement:** Highly scored comments may reflect content that is widely agreed upon or appreciated by the community. These tend to be more neutral and less controversial, leading to higher scores.
2. **Subject Matter:** Discussions related to Spiderman on the PlayStation subreddit may involve sharing information about the game, discussing gameplay mechanics, or expressing excitement about upcoming releases. In such cases, sentiments might naturally lean toward neutral or positive, especially in well-received posts.
3. **Community Engagement:** Content creators and community members may actively engage with discussions that are more neutral and informative. This engagement can contribute to higher scores for those comments.
4. **Avoidance of Controversy:** Higher-scoring comments may intentionally avoid taking strong stances or being overly negative to appeal to a broader audience. Users may choose to express opinions in a more neutral manner to prevent potential disagreements or downvotes.

It's important to interpret these findings in the context of the playstation subreddit and the nature of discussions around Spiderman.

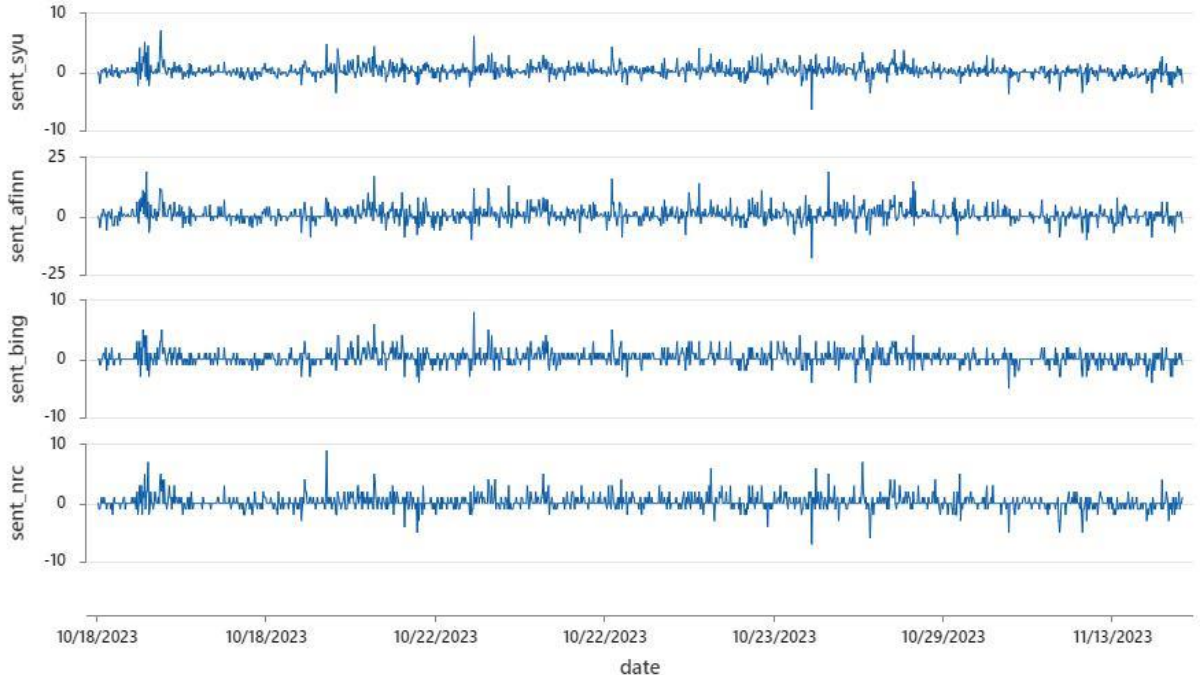


Figure 3: Time series plot of sentiment scores

As observed in time series plot,

1. Positive Sentiments at Game Release:
 - The initial positive sentiments around the time of the new Spiderman 2 game release (19th October, 2023) are likely indicative of excitement, anticipation, and positive reactions from the community. Players and fans often express their enthusiasm when a highly anticipated game is launched.
2. Neutral Sentiments in the Following Days:
 - The shift to neutral sentiments in the days immediately following the game release could be related to a period of assessment and exploration by players. After the initial excitement, users might be evaluating the game, sharing their thoughts, or discussing gameplay experiences, leading to a more balanced sentiment distribution.
3. Huge Spikes of Positivity:
 - The subsequent spikes of positivity in the following month may suggest ongoing positive experiences with the game. These spikes could be associated with community interactions related to the Spiderman 2 game. This could also align with the release of updates, patches, or additional content for the game, enhancing the overall gaming experience.
4. Random Bursts of Negativity:
 - The intermittent bursts of negativity could be attributed to various factors. These might include bugs or technical issues with the game, negative reviews or

critiques, discussions on controversial topics related to the game, or fluctuations in community dynamics. It's common for online communities to experience both positive and negative sentiments, and these fluctuations could be driven by a variety of factors.

4. Keyword Analysis: (Source file: textMining.R)

Analyzing terms positively associated with "Spiderman" based on correlation scores reveals what resonates well with the audience. It helps me understand specific themes and attributes that drive positive sentiments. This insight informs content creation and community engagement, allowing me to tailor messages and enhance the overall positive experience for Spiderman enthusiasts.

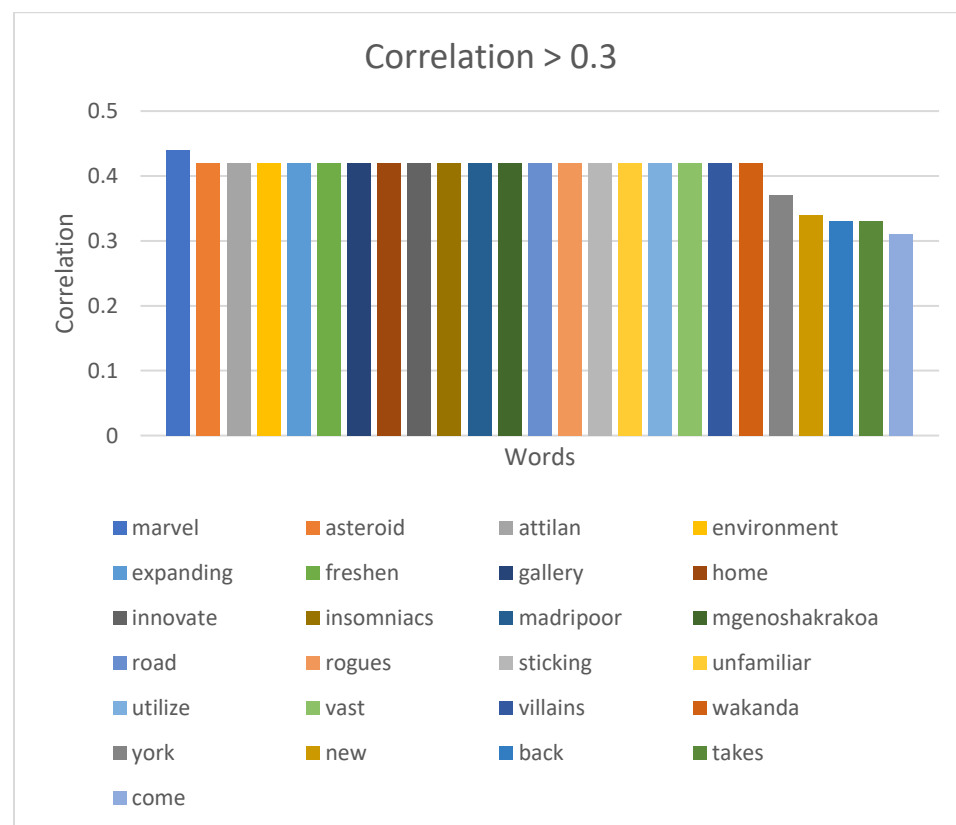


Figure 4: High Correlation

Here are some observations from Figure 4:

1. High Associations (Positive):
 - Terms like "marvel," "asteroid," "attilan," "environment," "expanding," "freshen," and others have relatively high positive associations with "spiderman".
2. Moderate Associations (Positive):

- Terms like "gallery," "home," "innovate," "insomniacs," "madripor," "mgenoshakrakoa," "road," "rogues," "sticking," "unfamiliar," "utilize," "vast," "villains," "wakanda," and others have moderate positive associations with "spiderman" (correlation scores around 0.42).
3. Lower Associations (Positive):
 - Terms like "york," "new," "back," "takes," and "come" have somewhat lower positive associations with "spiderman" (correlation scores between 0.34 and 0.31).

Interpretation:

1. Broad Range of Topics: The expanded list includes a diverse range of terms, indicating that discussions related to Spiderman in the PlayStation subreddit cover various topics. From specific terms like "2004" and "2014" to broader concepts like "games," "city," and "article," the associations suggest a broad and multifaceted discussion.
2. Continued High Associations: Terms such as "marvel," "asteroid," "attilan," "environment," "expanding," "freshen," "gallery," "home," "innovate," "insomniacs," "madripor," "road," "rogues," "sticking," "unfamiliar," "utilize," "vast," "villains," and "wakanda" continue to have high positive associations. This consistency may reflect recurring themes or closely related concepts in Spiderman-related discussions.
3. Associations with Gameplay and Excitement: Terms like "games," "city," "will," "article," "game," "spidey," "fresh," "excited," "puts," and "stock" suggest a connection to the gaming experience. The association with terms like "excited" and "game" indicates that discussions involve not only information, but also emotional and experiential aspects related to Spiderman games.
4. Temporal References: Terms such as "2004" and "2014" may indicate discussions related to specific Spiderman games or movies (Spiderman 2 and The Amazing Spiderman 2 respectively) released in those years. This temporal reference adds a historical context to the discussions and suggests that the community may be revisiting or discussing older game or movie releases.
5. Character and Story Elements: The emergence of terms like "wolverine," "miles," "shadows," "happened," "another," "dimensions," "shattered," "make," "part," "things," "break," and "fun" suggests discussions around various character and story elements within the Spiderman universe. The use of terms like "shadows," "dimensions," and "shattered" might hint at story developments or in-game events.

5. Popular words: (Source file: textMining.R)

Analyzing the most frequent words alongside sentiment scores provides valuable insights into recurring topics and key subjects within textual data. It allows for a nuanced understanding of sentiment by correlating specific words with positivity, negativity, or neutrality. This approach aids in tracking trends over time and adapting strategies accordingly. The analysis of popular words serves as a practical tool for quality assurance, content strategy, and informed decision-making in community management.

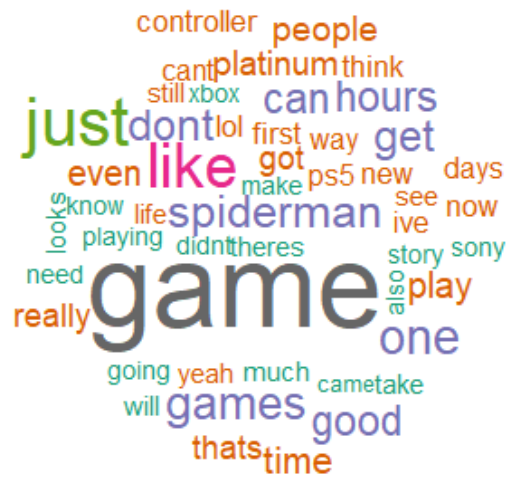


Figure 5: Popular words

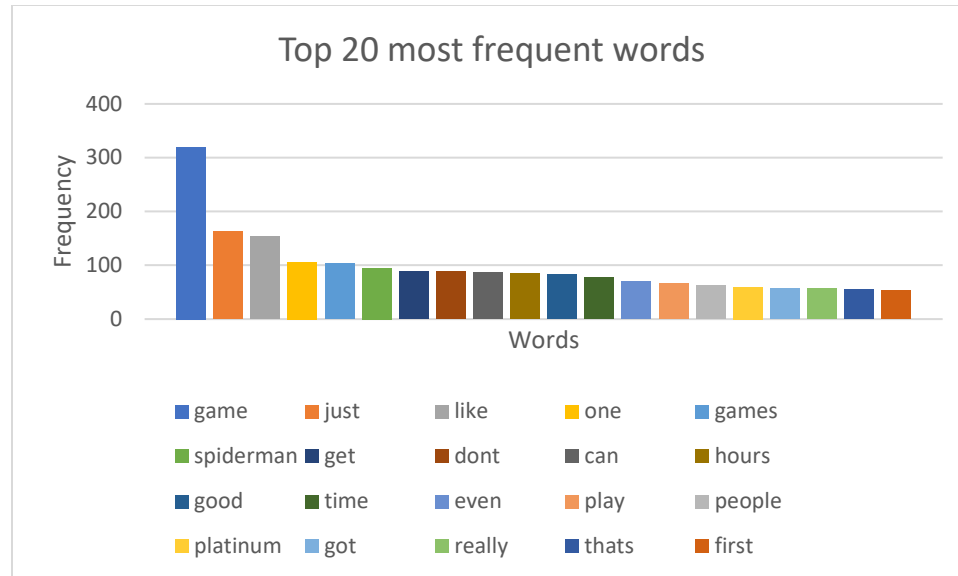


Figure 6: Top 20 most frequent words

An analysis of the most frequent terms:

1. Game: The most frequent term, "game," appears 320 times. This is a broad term and could refer to discussions about Spiderman games in general or specific titles.
2. Just and Like: "Just" and "like" are common words used in various contexts. They may be used for emphasis, comparisons, or general expressions, and their high frequency suggests they play a role in the overall tone and style of the discussions.
3. One: "One" appears 106 times. Without further context, it's a general term, but it could be related to discussions about a particular Spiderman game.
4. Games and Spiderman: These terms are directly related to the Spiderman theme and gaming. "Games" (66 occurrences) likely refers to discussions about the games, while "Spiderman" (95 occurrences) is specific to the superhero character and the franchise.
5. Get and Don't: "Get" (89 occurrences) and "don't" (89 occurrences) are action-oriented terms. Discussions using "get" may involve tips, strategies, or experiences related to achieving specific goals in the games. "Don't" could be used in expressions of caution, or negative experiences.
6. Hours and Time: "Hours" (85 occurrences) and "time" (77 occurrences) suggest discussions related to the time investment in playing Spiderman games. This could include playtime recommendations, discussions about game length, or time-related experiences.
7. Even and Play: "Even" (70 occurrences) and "play" (66 occurrences) are versatile terms that could be used in various contexts. "Even" may indicate surprise, emphasis, or balance, while "play" is central to discussions about gaming experiences.
8. People and Platinum: "People" (62 occurrences) suggests discussions involving the community or player base, while "platinum" (60 occurrences) likely relates to achieving a platinum trophy, a significant accomplishment in PlayStation games.

The analysis of prevalent words in Spiderman-related discussions on PlayStation platforms reveals a strong focus on gaming experiences, strategies, and player interactions. Terms like "game," "Spiderman," and action-oriented words highlight the community's engagement with gameplay dynamics. The emphasis on time-related terms presents opportunities to enhance features related to gameplay duration and experiences. The frequent mentions of "platinum" suggest a notable achievement within the gaming community, providing opportunities for recognition and celebration. These insights can inform targeted marketing strategies, community engagement initiatives, and feature enhancements to elevate the overall PlayStation and Spiderman gaming experience.

6. Conclusion:

As the Spiderman game community evolves, ongoing sentiment analysis remains a valuable tool for understanding player experiences and community dynamics. Future analyses could explore the impact of ongoing updates on community sentiments. In essence, this sentiment analysis serves as a strategic guide for community management, offering actionable insights to enhance player experiences and maintain a positive and vibrant Spiderman game community.