

Overview

This project explores whether we can use the sentiment of tweets to quantify if expectations were met, exceeded, or disappointed after a movie is released. As an executive at a movie studio, this will allow you to answer the questions; Did the film live up to the hype? Did we disappoint? Or are your viewers pleasantly surprised? This project tracked two movie hashtags daily over 1 month around their release dates and analyzed their content using sentiment analysis.

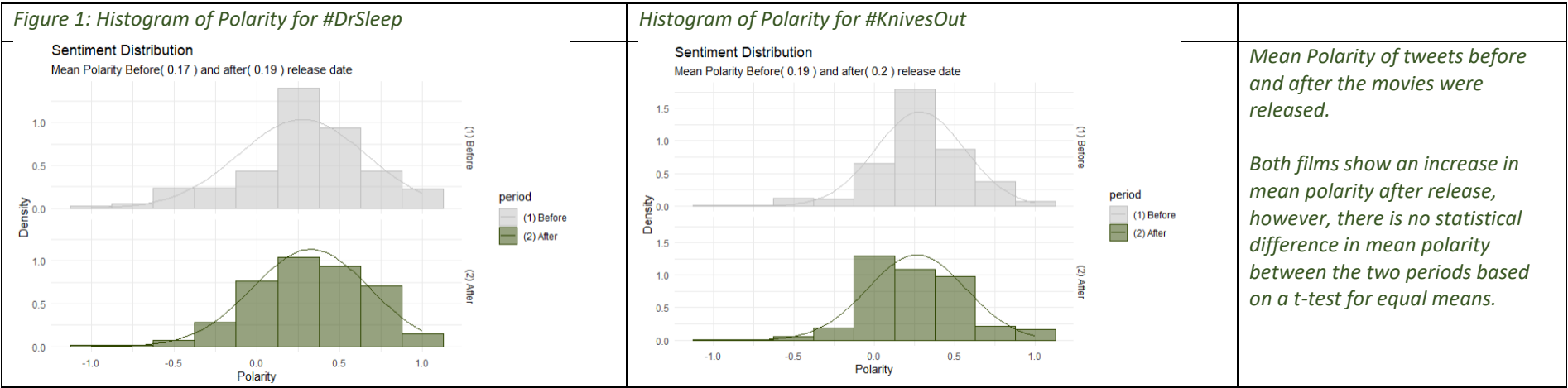
Assumptions

We can't know for certain whether the tweet author has seen the movie or not. For this analysis, we assume tweets occurring before the release date are in anticipation of seeing the movie and represent the prior expectations of moviegoers. Tweets created on or after the release date are assumed to be from people who have seen the movie and represent a movie review. Retweets have been cleaned to contain the original tweet text and included under the assumption that a retweet constitutes the same opinion as the original tweet.

Analysis

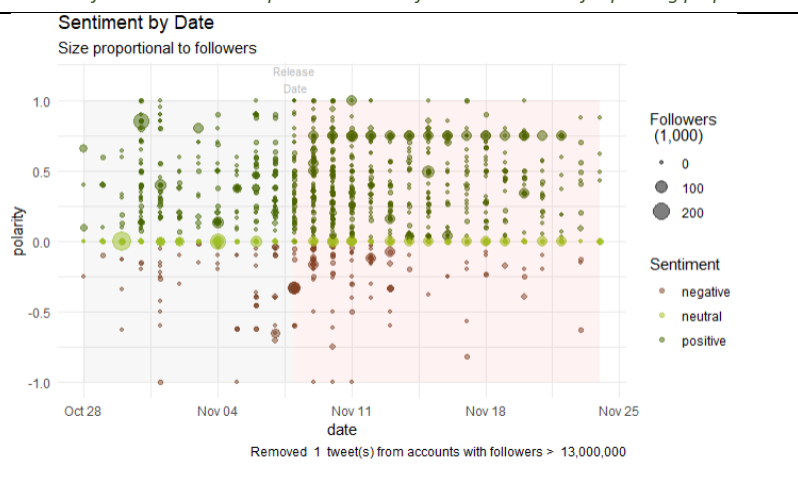
Approximately 2k tweets were collected for #DrSleep and 34k for #KnivesOut over 1 month leading up to and after each movie was released. Tweets were collected using the Tweepy REST API in extended mode to include text greater than the 144-character limit. Links were removed from each tweet, and in the case of retweets, the original text was collected instead of the retweeted text to exclude metadata not required for the analysis. Duplicate tweets and tweets from official movie accounts were removed by referencing the unique tweet id and user ids.

The python library TextBlob was used to determine the sentiment of each tweet, providing a polarity score from -1 (negative) to +1 (positive) with zero representing a neutral tweet. Figure 1 shows histograms of polarity before/after the release date. Figure 2 shows tweet polarity over time and is proportional to the number of followers the user has. To determine if there has been a statistically significant shift in sentiment, we perform a test of proportions which compares the percentage of positive/neutral/negative tweets before and after the release date.

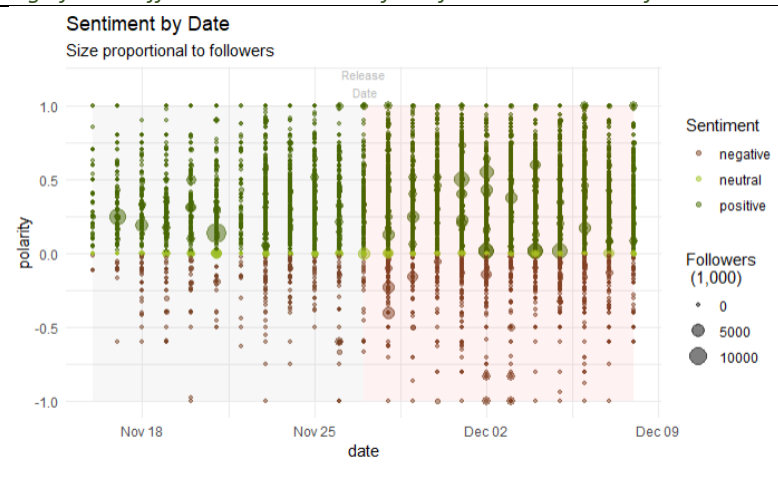


Both movies show a significant result for the test or proportions indicating the percentage of positive/neutral/negative has changed since the release of the movie – it would seem viewers' expectations were exceeded for #DrSleep and viewers were underwhelmed by #KnivesOut. Figure 3 shows the volume of tweets by sentiment over time as well as the rolling average polarity and percentage of positive tweets, showing an upward trend after the release date for #DrSleep and a slight negative trend for #KnivesOut.

Figure 2: Tweet Polarity by Date and Followers for #DrSleep
 *1 tweet from the associated press with >13M followers removed for plotting purposes



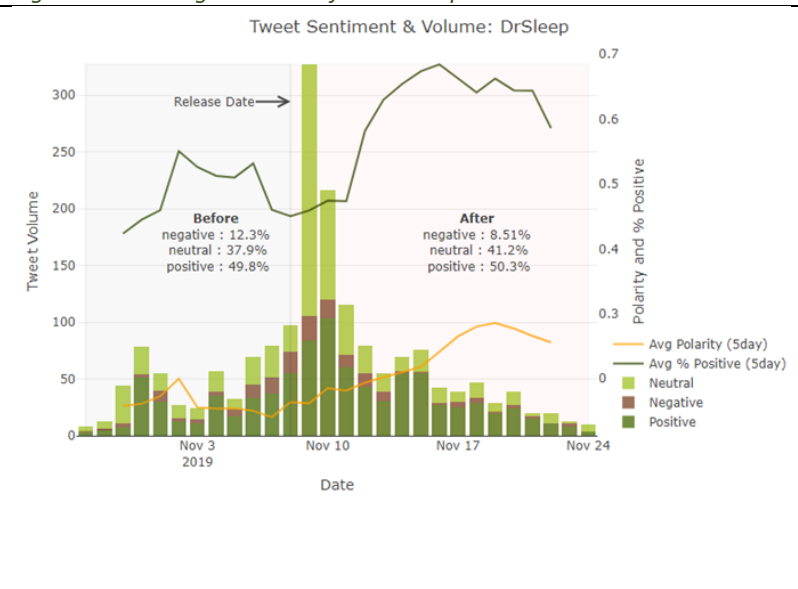
Tweet Polarity by Date and Followers for #KnivesOut
 Significant difference in sentiment for Influencers vs Non-Influencers



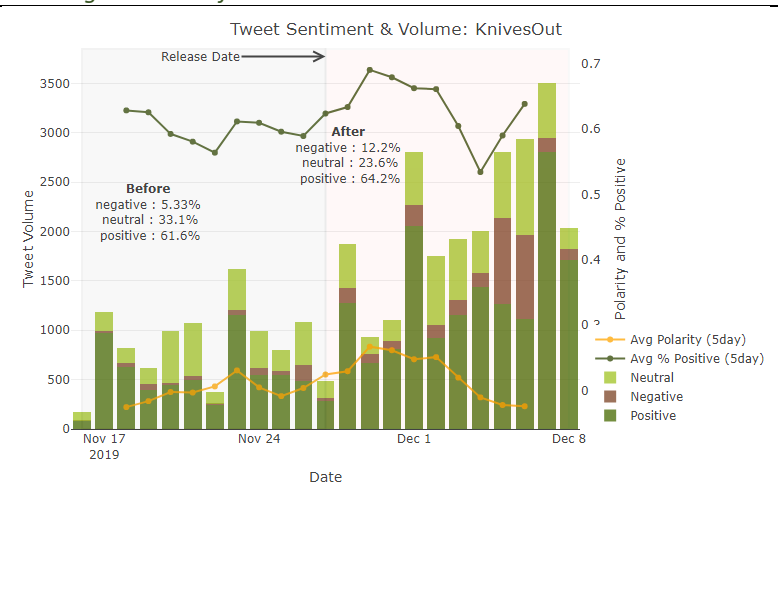
The more popular #KnivesOut has attracted the attention of a lot more tweets and influencers than #DrSleep.

There is no statistical difference in proportions between influencers (followers >= median) and non-influencers (followers < median followers) for DrSleep

Figure 3: Trending Sentiment for #DrSleep



Trending Sentiment for #KnivesOut



Views became more polarized after watching #KnivesOut, with the proportion of neutral tweets shifting mostly negative. It seems it didn't live up to the hype.

In contrast, #DrSleep gained more neutral & positive tweets after viewing.

The proportions were statistically different after viewing for both films.

Conclusion

Comparing the proportions of sentiment before and after a movie release is a simple method to quantify expectations and reactions to a movie. This principle can be extended to any product or event that has a launch or event date. It can also be used to track trending of sentiment over time to provide executives some indication of how their products are perceived over time.