Netflix Analysis

Executive Summary

Introduction

Following recent news of Netflix potentially introducing commercials and eliminating password sharing, there was over a 200,000 loss in subscribers and a drop in stock price. Through this project, our goal is to analyze recent customer sentiment that would provide business insights and help to determine possible solutions Netflix could implement.

Process/Methods

The process we will take to conduct this analysis involves utilizing an API scraper to collect comments and threads from relevant subreddits on Reddit. The methods that will be used to analyze text data are sentiment analysis and topic modeling.

Significance

After examining our results, we learned that, overall, there is more negative customer sentiment in circulation toward Netflix’s commercials and the various topics that were discussed in the data we collected. Lastly, we were able to develop a potential business strategy Netflix could apply to their service and marketing that responds to the customer sentiment we observed.

| **Group Members** | **Emails** | **Roles/Responsibilities** |
| --- | --- | --- |
| Giorgio Pisano | gapisano@calpoly.edu | Insights |
| Ben Stephan | bestepha@calpoly.edu | Everything, Results |
| Yongxin Lin | ylin46@calpoly.edu | Methods |
| Joann To | jto13@calpoly.edu | Problem, Insights, Conclusion |
| Zane Mazor-Brown | zmazorbr@calpoly.edu | Dataset/Collection, Methods, Editor |

**Problem**

In April of 2022, news broke that Netflix lost 200,000 global subscribers in the first quarter of the year. This was the first time the company lost subscribers in over 10 years. As a result, the company’s stock price dropped 35%. This occurred due to an announcement that the platform plans on launching a cheaper, ad-supported plan sometime this year as well as cracking down on password sharing. These two factors are likely behind the recent subscriber loss and poor stock performance.

In this project, we wanted to conduct an analysis of Netflix’s customer sentiment following these events. Because Netflix relies heavily on its streaming customer base for revenues, customer sentiment is incredibly influential to the company’s success, and analysis of this sentiment can provide significant, actionable intelligence relevant to the future of the company. Based on the recent stock crash and the discussion of advertisements coming to the platform, we initially predicted that customer sentiment will have decreased significantly in recent weeks.

The goal of our analysis is to determine how customers are reacting to and feeling about these changes that Netflix plans on making and what factors or reasons are behind their sentiment. With that information, we will then develop ideas that Netflix can execute to either improve or maintain current sentiment. This will ultimately increase the number of subscribers and profits for the company.

**Dataset and Collection Methods**

Our dataset was sourced by scraping the “Technology” and “Stocks” subreddits on Reddit through an API. Querying for “netflix'' and “nflx,” we obtained 13,197 total threads and comments from these subreddits. We then created a SQL database using pandas to store our records, as well as creating a csv file for the threads table and one for the comments table. For the threads table, we included these attributes: thread\_id, title, author, url, created\_utc, num\_comments, and score. For the comments table, we used several other attributes: thread\_id, body, author, created\_utc, and score. In our analysis, we only utilized the thread\_id and title from the threads table and the body text from the comments table; however, future analysis could make use of the score from the comments table to prioritize comments. This dataset allowed us to perform multiple analyses on the text data.

**Methods**

Prior to the standardized preprocessing and tokenization of the text data, we filtered our data for relevant records on a specific topic. For this analysis, we chose to search for threads and comments relating to ads, advertisements, or commercials using regular expressions. We also included comments with the same thread\_id as a thread relating to this topic. Filtering for these specific criteria narrowed our text data from 13,192 records to 1,198 records. It should be noted that our filtered dataset was only utilized in the sentiment analysis portion of our study. When performing topic modeling, we utilized the full dataset.

To best prepare our text data for analysis, we implemented extensive pre-processing using a variety of methods. First, we used regular expressions to remove special characters and manipulate various expressions for cleaner, more digestible text data. Next, we expanded contractions to their long-hand form for more consistent word frequency. After converting all words to lowercase, we tokenized the text, removed stopwords (including other special characters we discovered), removed words with character lengths less than 1 or greater than or equal to 20, and finally obtained the root of each word by lemmatizing the data.

For our project, we utilized two different text mining methods: sentiment analysis with supervised learning and topic modeling. First, we created a support vector machine (SVM) regression model using sentiment analysis to predict customer sentiment surrounding the discussion of Netflix potentially adding advertisements and/or commercials to their platform. Following this, we conducted topic modeling to gain insight into what Netflix customers truly care about.

For our SVM regression model, we first used doc2vec to create vectors for each document in our dataset, which allowed us to analyze our data quantitatively. We found that a vector size of 200 words with a window of 5 produced optimal results for our SVM model.

Our sentiment analysis involved the use of natural language processing (NLP) techniques to quantify and study affective states and subjective information. To best identify the sentiment of each text document, we used two popular sentiment analyzers, Vader and TextBlob. Running our text data through these tools, each labeled and categorized the documents as positive or negative sentiments. When TextBlob and Vader categorized a review as the same sentiment (ex. Vader: positive, TextBlob: positive), we labeled that review as a “true” positive or negative sentiment to be used in our model.

Using our list of true sentiments, we split the text data into two sets for supervised learning, the training set (20% partition of the data) and the testing set (80% partition of the data). Finally, we used the two sets of data to derive our SVM regression model and the best prediction results. After testing, we decided that the poly kernel produced the best results for our SVM model.

The second text mining technique we employed was topic modeling. Topic modeling is a machine learning technique that enables us to scan a series of documents, detect words, and find patterns. By investigating the patterns in the text of the dataset and the distance between words, we are able to find the words that are frequently utilized to describe customer sentiment. First, we conducted the topic coherence measures, which score the topics depending on the semantic similarity of the high-scoring words in each topic. By selecting the highest coherence score, we determined the optimal number of topics for our text data. We found that 5 topics were optimal for our data at a coherence score of about 0.62. After inputting 5 as the number of topics desired for topic modeling, we calculated the top-30 most relevant terms for the topic.

**Results**

Sentiment Analysis:

Our sentiment analysis results can be seen in Figure 1. Textblob and Vader identified 856 true sentiments out of 1,198 records, or about 71% of the dataset. Out of these, 366 were labeled positive and 490 were labeled negative, already indicating that the current overall sentiment towards advertisements and commercials on Netflix is more negative leaning than positive leaning. Our classification report highlights that our precision was higher for the negative sentiments than the positive sentiments (0.73 > 0.55), and the recall was higher for the positive sentiments over the negative sentiments (0.72 > 0.56). In our case, we care about both the precision and the recall for different reasons. We care about precision because we want to understand how people feel about the possibility of Netflix adding advertisements immediately after hearing about it. We care about recall because we want to be able to detect and predict future negative sentiments and plan company strategy and action accordingly. Additionally, our SVM model obtained an f1-score of 0.64 for predicting negative sentiments, and an f1-score of 0.63 for predicting positive sentiments. Although these f1-scores are not particularly impressive, they do indicate that the model predicts positive and negative sentiments consistently with each other, which is more valuable than very polarized f1-scores for each sentiment.

Topic Modeling:

Our topic modeling results can be seen in Figures 2-7. Our coherence test is highlighted in Figure 2, where the highest coherence score is around 5 topics. Figure 3 illustrates our results for topic 1. Words such as show, watch, content, and stream highlight that, above all else, customers care about the content on the Netflix platform. Figure 4 highlights the results for topic 2. Some of the most frequent words were stock, go, market, buy, invest, and sell, indicating that the consumer base we analyzed was also discussing stock trading and advice. This is expected as we obtained part of our dataset from the stocks subreddit following the historic 35% drop in Netflix share price. Analyzing this topic in more depth in the future may indicate more accurate customer sentiment towards Netflix stock, which also has a huge impact on the success of the company. Topic 3 discusses streaming options with words like buy, stream, company, get, content, service, perhaps indicative of the competitive landscape of the online streaming industry. Netflix must constantly be attentive to the competitive landscape in order to stay ahead of the stiff competition. Topic 4 is similar to Topic 2, but emphasizes company growth rather than stock trading with words such as market, stock, year, and company. Topic 5 discusses other important aspects of streaming platforms for users, highlighting words like price, subscription, and account but mixed with negative words like bad, cancel, and various curse words. This may indicate that customers feel negatively towards Netflix’s prices, subscriptions, and accounts.

**Insights**

From the sentiment analysis results, we can view the customer sentiment towards Netflix after the news of possible changes to their subscriptions by adding commercials and advertisements to the current plans. With the majority of the sentiment being negative, we found that adding commercials and ads to Netflix could possibly hurt the public perception of the company (Fig. 1). In reference to our findings, if advertisements are implemented on Netflix’s platform, we recommend they not be a mandatory inclusion for all of Netflix’s customer base. Reduced-price advertisement and commercial plans might change some negative sentiment towards these changes if these customers prefer to have a more affordable service plan.

With external research, we found the range of sentiments was a result of various consumer responses. Some customers were interested in paying less and watching Netflix advertisements in return for the lower price. Others said they specifically signed up for Netflix to avoid such nuisances. Therefore, by proposing optional and inclusive ad-supporting plans, the company can satisfy both subsets of their customers.

From the topic modeling results, we can see frequently used words within the dataset. Netflix could use words that aim to have a positive change in the messaging toward customers. These commonly used words might have an impact on how those in the public, with a negative sentiment, view commercials and ads in subscription services. For example, words that are found within the topic modeling results are watch, show, and content (Fig. 3). Netflix can use some of these words to add to their commercial advertising campaigns as it may give those with a negative sentiment a different way to view the changes being made. A marketing campaign can incorporate commonly used words such as, “With the new Netflix advertisement plan, customers get to watch the best content at a lower price”. As suggested, the words found within the dataset were highlighted in the marketing phrase.

Also found in the stock topic in the topic modeling results, gathered in the past few weeks, are the words growth and buy (Fig. 4). These words could be used to represent the possible outlook on Netflix’s stock from public perception, as this data was mined from public Reddit’s technology and stock subreddits. Although the data may show a negative sentiment towards making changes to Netflix's short-term plans, the data may represent that there are long-term bullish investors using the words “buy” and “growth” when also mentioning Netflix.

**Conclusion**

Prior to conducting the analysis, we originally predicted Netflix’s customer sentiment would be dramatically low in light of recent events, and, while our results indicate there is a more negative sentiment out there than positive, we don’t believe the discrepancy quite constitutes a ‘dramatic’ drop in customer sentiment. Alternatively, there are customers who are open to the potential cheaper service that Netflix could provide. Ultimately, we suggest the introduction of an optional advertisements plan and the incorporation of a more targeted marketing campaign.

As our topic is a recent development and the discussion is still emerging, the data records we were able to collect were limited and customer sentiment is still evolving. However, with the models we’ve created, we would be able to use this analysis to predict future sentiment based on reddit threads and comments in the next few months. This would provide even more insights about the business and customers’ outlook of Netflix.

Moving forward, as Netflix continues to carry out new plans or change their current services, we can pull data from new Reddit threads and evaluate the public’s opinions that would help the company appropriately respond and create value.

**Appendix**

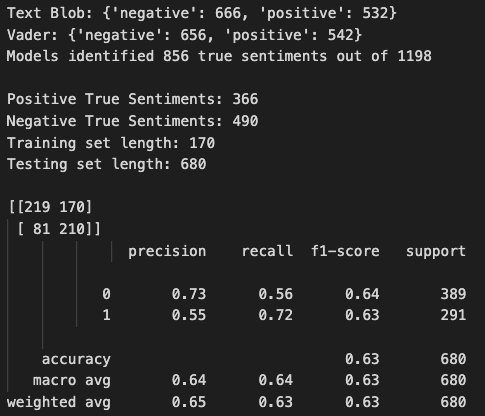


Figure 1

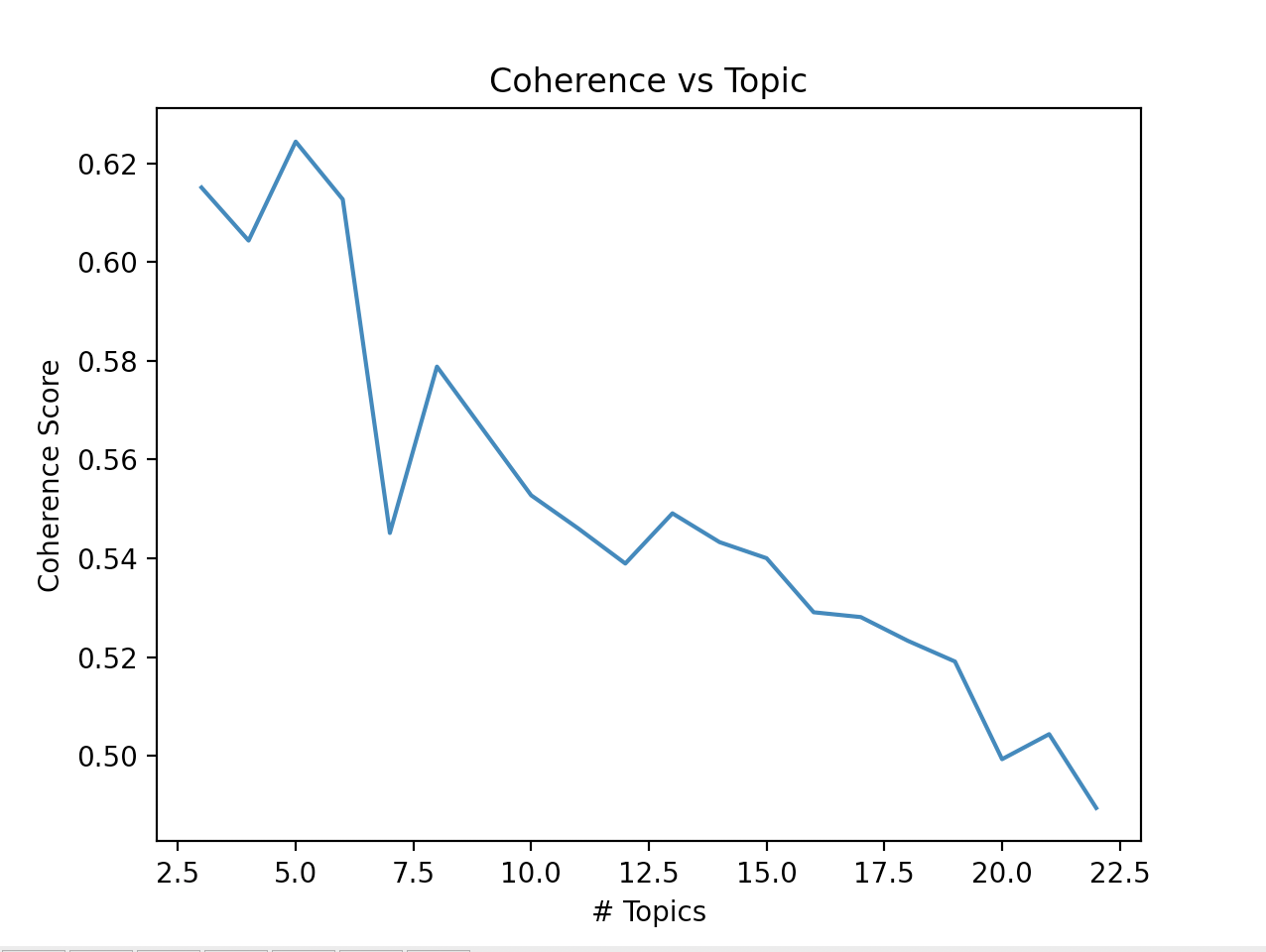


Figure 2

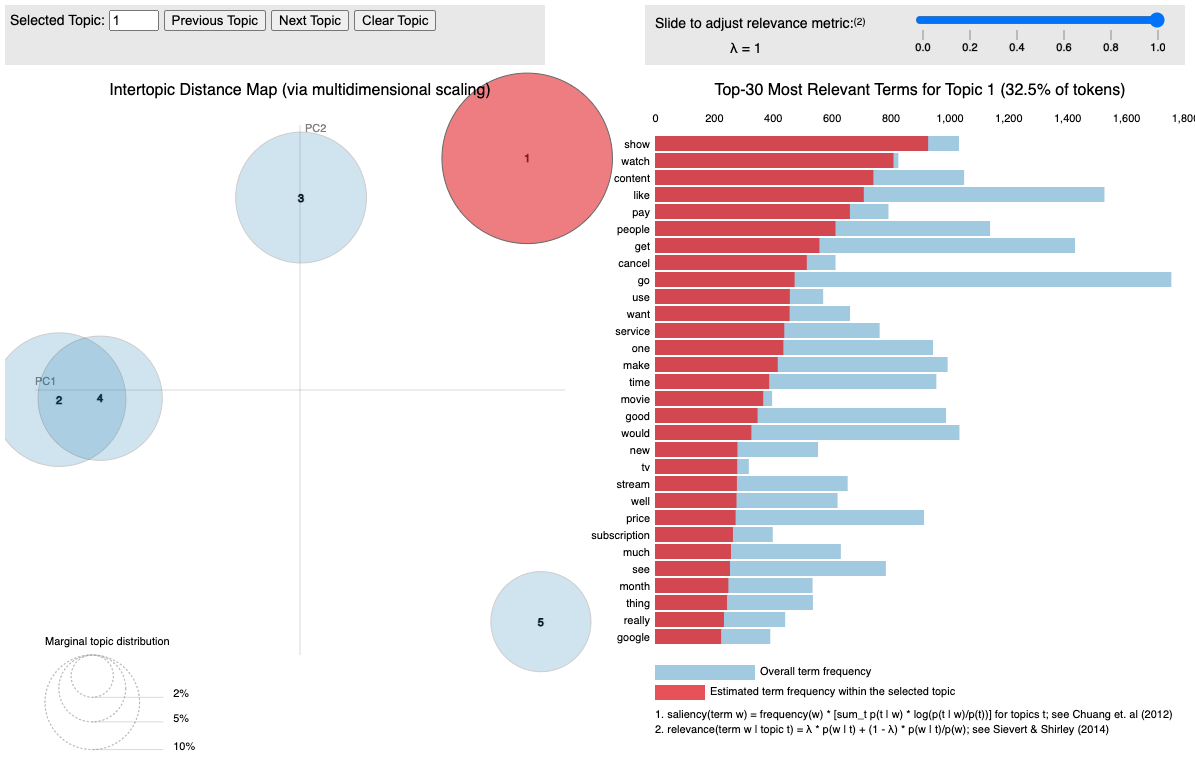


Figure 3

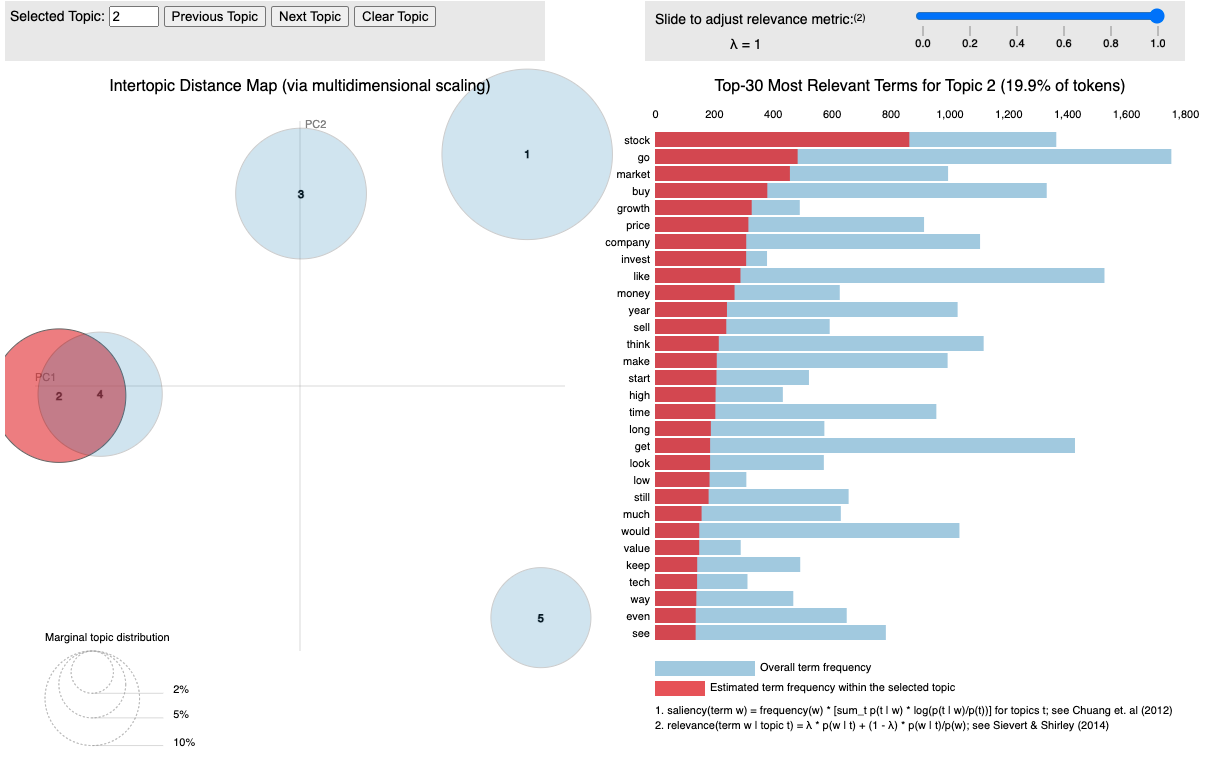


Figure 4

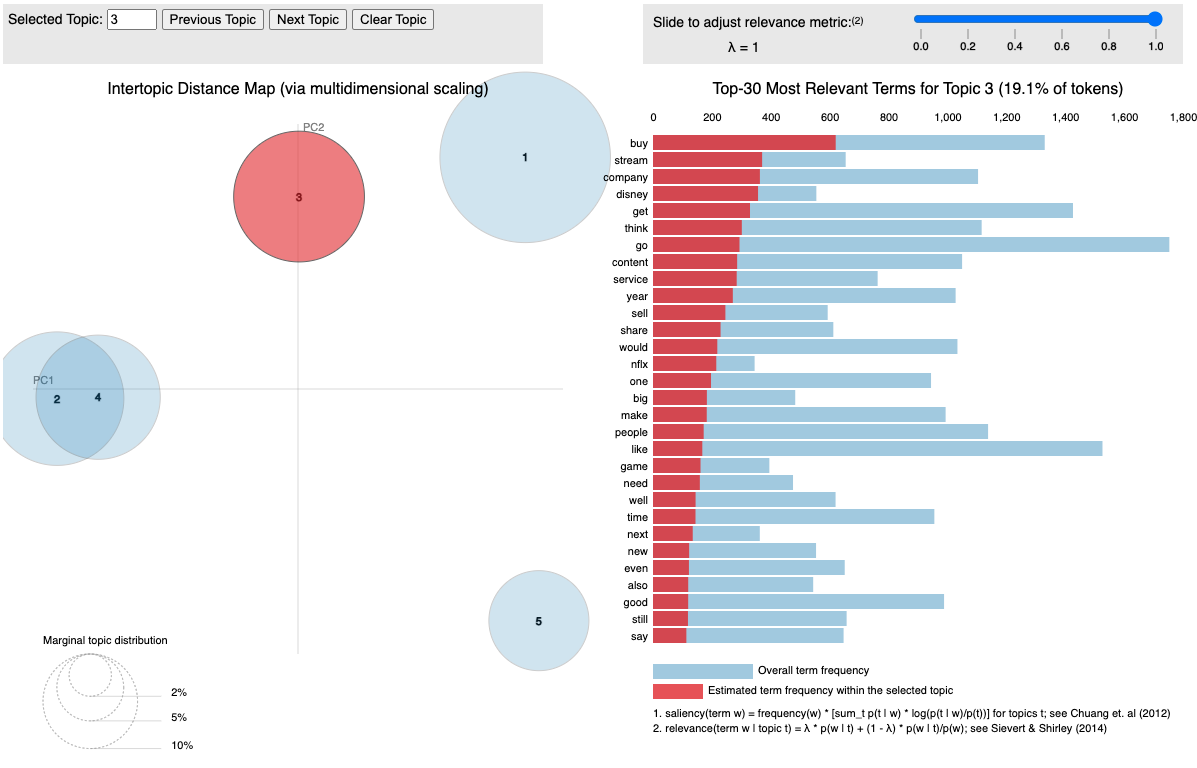


Figure 5

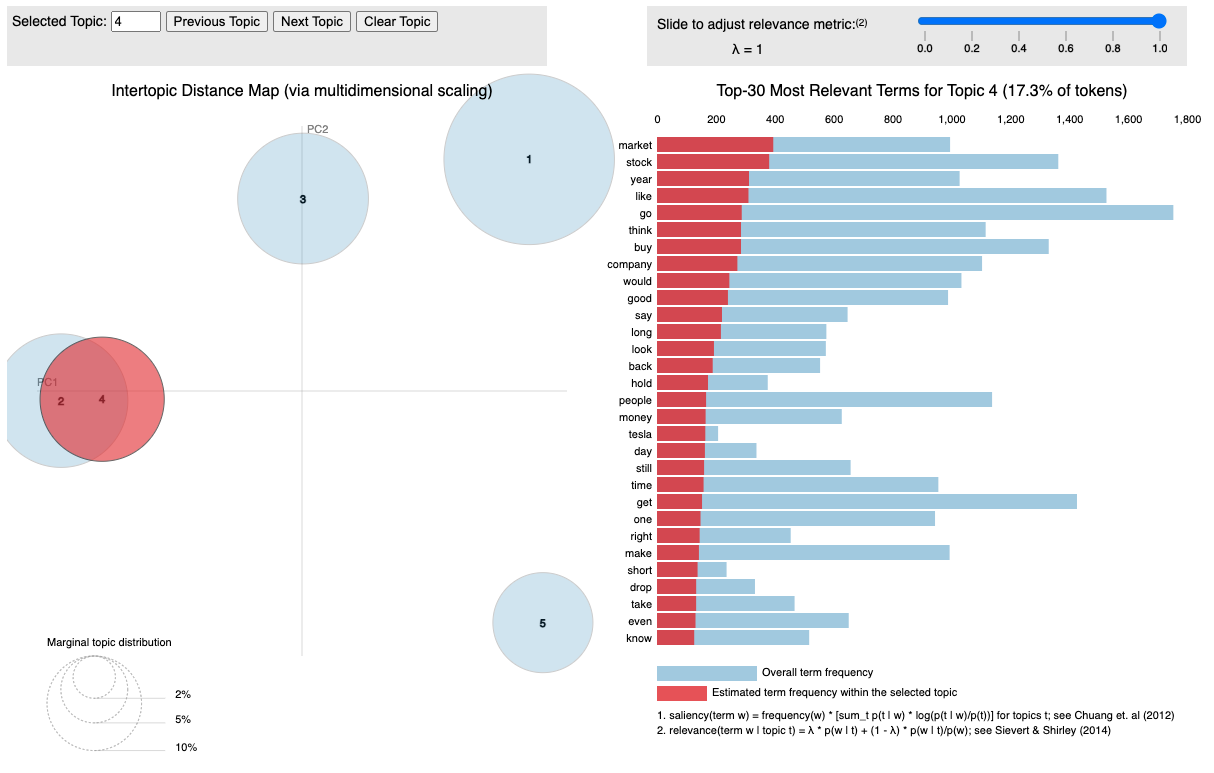


Figure 6

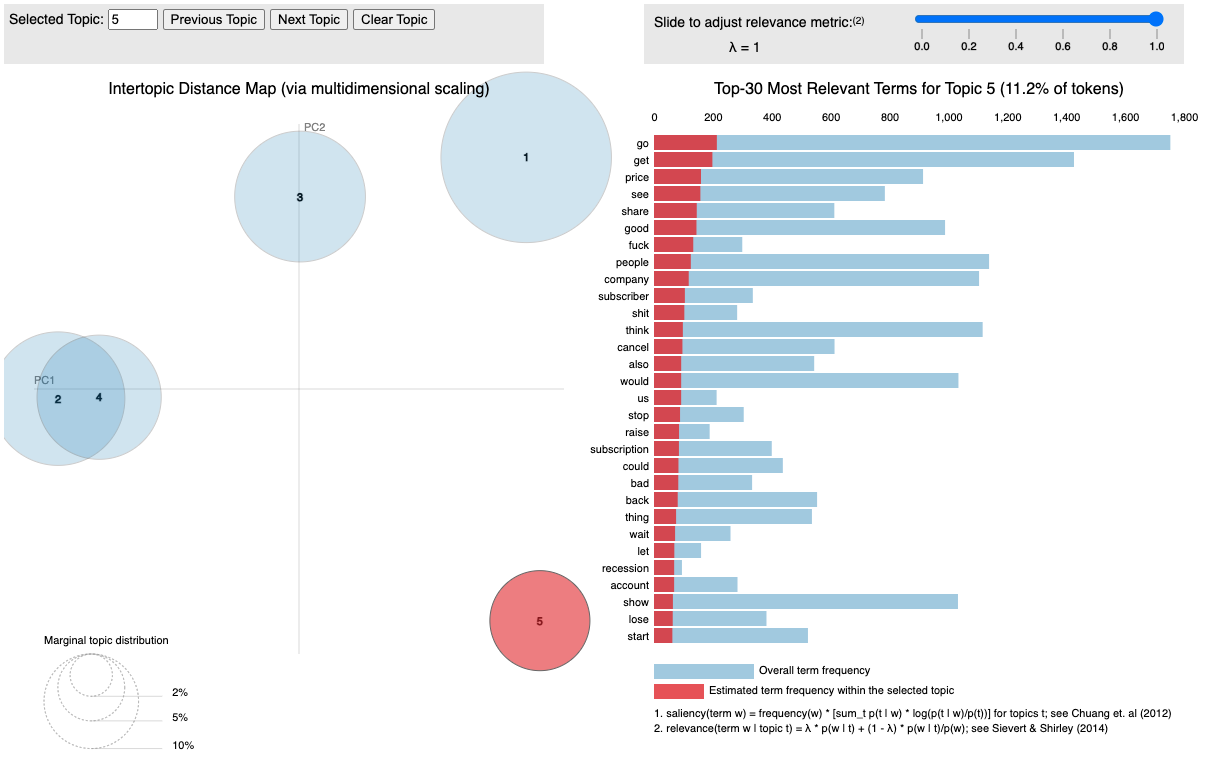


Figure 7