Market Research Report on Twitter Sentiment Dynamics in Covid-19 Pandemic

Executive Summary

Understanding the emotional states of the population is critical for business owners and policymakers when facing a crisis. Data from twitter provides us with frequency measures to track various aspects of people's mood, thoughts, and experiences. In this study, we used sentiment analysis, descriptive analysis, topic modeling and Word2Vec to analyze the sentiment pattern around the world, compare the mood difference between the United States and other countries, explore the trending topics related to the pandemic, and shed light on the most frequently discussed concerns. With all information on hands, we provide management suggestions to help business owners navigate through challenging times. These suggestions include firstly the implementation of post-pandemic wellness programs for employees to increase the work satisfaction and stability; Secondly, companies need to collaborate closely with the state and local health department to obtain timely and appropriate responses. And lastly, we suggest the adoption of a cloud-based, word-embedding system that tracks the vaccination records and eases up daily work routine for work-at-home employees.

Motivation

Covid-19 has caused a significant crisis around the world, and the lingering issues continue to affect every corner of our lives. In fact, many people not only went through financial crisis or unemployment, but also the fear, anger, and stress triggered by major life changes. During this time, social media's involvement and interaction have increased dramatically as more people share their emotions and viewpoints on pandemic related topics. That is the reason we chose one of the representative social platforms, twitter, to analyze the public's thoughts and sentiment on concerns, awareness, health issues.

There are many related research on Facebook, twitter and reddit breaking down text information to find the trending topics and sentiment over a period. And some of research specifically focuses on the prediction of sentiment using machine learning or deep learning method. With the help of previous studies and research, we scrutinize the user's sentiment from March 2020 to

March 2021 in different countries to assist in finding more trending topics around this time. We believe the analysis on this topic will open us to more business management intervention strategies and help us to design effective tools and campaigns based on public perceptions. Meantime, there are also challenges on this topic since the covid tweet dataset contains a large amount of self-report data, so any causal inferences from the dataset need further investigation. In summary, despite the challenges on the topic, this study is designed to leverage twitter data to understand public sentiment and to provide business-related insights based on public perceptions.

Description of the Dataset

We chose the Covid-tweet dataset from class session, and then randomly selected 1 percent of data from each of the 346 separate excel files between the date March 22nd, 2020, and March 3rd, 2021. In this way, we could effectively reduce the size of the data size while still having a representative sampling set for the analysis. After data prepossessing and data cleaning, there are a total of 17,363 rows and 5 columns in the dataset. The reason we deleted user_friends_count, user_follower_count, retweet_count and user_favorite_count is that the data is not a real time reflection of the user's profile. Despite country_code is a more accurate representation of the location than the user location column, there is an over 98% of missing values in the country code column. We ended up deleting the country_code as well.

User_id: The unique id for each twitter user

Lang: The language used in the tweets

Created_at: the date when the twitter was posted

Text: the text message sent by twitter user

User_location: the location of the user

Method

We used sentiment analysis, basic descriptive analysis – word cloud and frequency of hashtags, Topic modeling and Word2Vec for the research. Details are listed as follows.

(1) The sentiment analysis

To analyze the sentiment on text, we adopted three lexicon packages: Bing, Afinn, NRC. From the Afinn lexicon, we found that the sentiment intensity clustered the most at score 2, indicating a slightly positive score for the overall text information (see Graph 1). Moreover, NRC model split the emotion into 10 categories, and we discovered that the sentiment "trust" is the most frequent sentiment in the tweets (see Graph 2). Moreover, if we take a closer look at the top mentioned words under each major category in graph 5, we find interesting topics for real world application. The virus, pandemic and death rates have triggered emotion like "sadness", "fear" and "disgust", while the invention of vaccines provides positive feelings and hope for the community. Trust between people is the most prevalent emotion as mentioned before, and we found that the trust to the hospitals, the president and the school are the connections that hold people together.

In terms of the difference of sentiment between countries, we took the United States and the United Kingdom as the examples since they are the countries with the highest number of tweets. The results show that in comparison to the United Kingdom, U.S showed higher level of fear and sadness in tweets, with less degree of anticipation (see Graph 6). The overall pessimistic feelings in tweets might affect U.S business conditions since more business owners or employees are dwelling on negative thoughts. Moreover, both countries demonstrate less anger in tweets, which is interesting, as anger is often associated with sin, fight or conspiracy theories. We also find more cultural difference when we include non-English speaking countries like Mexico into the picture (see Graph 7). In comparison to U.S, Mexico shows the highest level of anger and anticipation indicating a mixed feeling to extreme degree. There is room for interpretation since people with diverse cultural backgrounds tend to express and experience their emotions differently. Overall, results shown in this section indicate that people are more common to have mixed emotions rather than purely negative ones. With proper emotional support from the employers and community, we could expect a faster recovery from the pandemic.

(2) Basic Descriptive Analysis

We conducted basic descriptive analysis, including word cloud and frequency of hashtags, to understand twitter users' main concerns concerning coronavirus.

Before creating word cloud and the list of hashtags, we preprocessed the text data. Besides some general steps, we remove some subject-specific stop words, such as covid19, coronavirus, virus to help us extract keywords related to covid rather than covid itself. In addition, we remove some predicate and verbs because these words might require context to understand, and nouns will deliver the message more directly in the word cloud.

From the word cloud output (see Graph 3), it turns out that twitter users often mention vaccine, lockdown, pandemic, job, and business. People seek for the development of vaccine. Moreover, judging from the word combination of lockdown, job, and business, covid19 has a significant impact on people's work. Due to lockdown policy, many people had to work from home and even faced unemployment. Many industries, such as catering services, airline business, have been in hard times.

We also draw a plot illustrating the frequency of hashtags in tweets. Hashtags are used to highlight the main topic of a tweet, so it is of importance for us to analyze. Judging from the bar plot (see Graph 4), it is like the result of word cloud. The keyword vaccine still ranks high among the words.

(3) Topic Models

The LDA topic model was utilized to further extract insights. Latent Dirichlet Allocation (LDA) is an example of topic modeling and is used to classify text in a document to a particular topic. It builds a topic per document model and words per topic model, modeled as Dirichlet distributions. LDA from Gensim package was utilized to generate the top 9 topics, within which 10 most frequent words were selected as representatives (see Graph 8). For interpretation purposes, only tweets in English were selected in the topic modeling process.

From topic 1, 4, 7, and 8, netizens showed concern and anxiety about the outbreak. They closely track positive cases and deaths figures. Topic about vaccines is also highly discussed, showing the intense willingness of epidemic prevention.

The outbreak not only affected netizens' emotions but also their daily lives. As shown in topic 0 and 2, netizens complain about the restriction of movement during the outbreak. It forced them to change the venue of study and work to home. Moreover, a significant amount of workers are facing layoffs and financial difficulties. With workers absent from office and the depressed economy, companies need to seek new solutions for employee management.

(4) Word2Vec

Our team further extracts actionable insights regarding business applications using Word2Vec models. Word2vec takes as its input a large corpus of text and produces a vector space, typically of several hundred dimensions, with each unique word in the corpus being assigned a corresponding vector in the space. Word vectors are positioned in the vector space such that words that share common contexts in the corpus are located close to one another in the space.

As the code shown in graph 9, our model uses the processed text in the English tweets data frame that will focus only on the interpretation of the English tweets. "Most similar" function is employed to shed light on what people care about the specific realms of keywords we choose, in macroeconomics, in business operations, in microeconomics, from the firm to individual employee or consumers during this tough time.

Insights

By global standards, the U.S economy has bounced back rapidly starting from the beginning of 2021, but as we have seen from the sentiment analysis, people's emotional well-being is still challenged in every aspect. Instead of having employees continue to experience mixed emotions and a high level of uncertainty (see Graph 6), we think it is time for companies to offer or improve post-pandemic wellness plans and online training for coping strategies. If you are a startup company looking for long term survival in a certain field, a transparent wellness policy is necessary to increase stability and work satisfaction with remote working.

Secondly, public concerns over the vaccination, health issues and work restriction are the mostly discussed topics in tweets (see Graph 3 & Graph 8). Therefore, business and employers need to play a key role in preventing and slowing the spread of the pandemic in the workplace. Moreover, business and employers should coordinate with state and the local officials to make timely responses.

As for the real-world application, since we found employees are practicing the ideas of remote working or hybrid working (see Graph 15), we suggest launching a comprehensive management system that can help to address the impact of the COVID-19 pandemic on business and create a better work from home environment for employees.

The suggested cloud-based system might include the following functions:

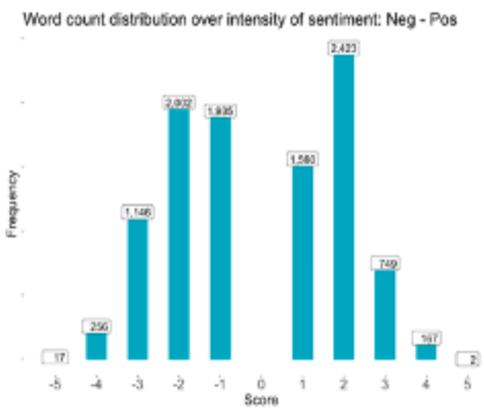
- Update employees' personal information, track employee health, especially vaccination records
- Perform employee/manager check-ins, view work schedules and time-off, calculate payroll and taxes by automatically tracking employee time
- Announce state and company regulations about COVID-19
- Communicate with employees in real time, such as scheduling meetings, submitting requests, reporting incidents, and answering questions
- Allow employee/manager to complete use e-signature on documents

Conclusion

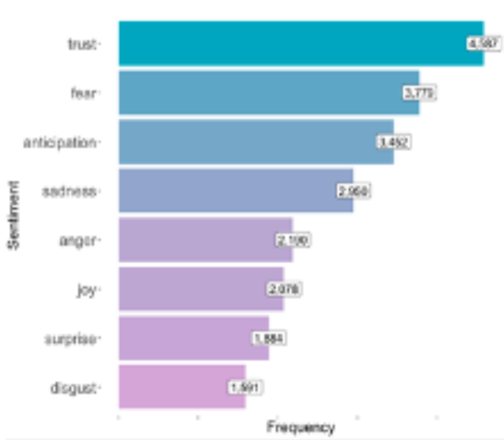
In this research, we have applied sentiment analysis, descriptive analysis, topic modeling and Word2Vec on twitter data through the pandemic. Based on what we found about the public perception, thoughts, and concerns, we believe business owners and employers should prepare, respond, and control workplace and management factors to provide safe work environments for their employees. If all companies adapt well to the post-pandemic changes, the road to recovery and beyond will be much smoother than we expected.

Appendices

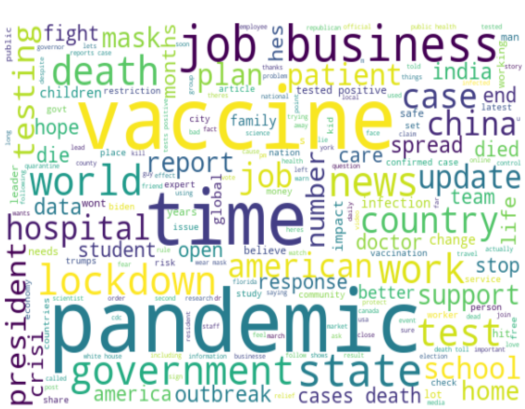
Graph 1



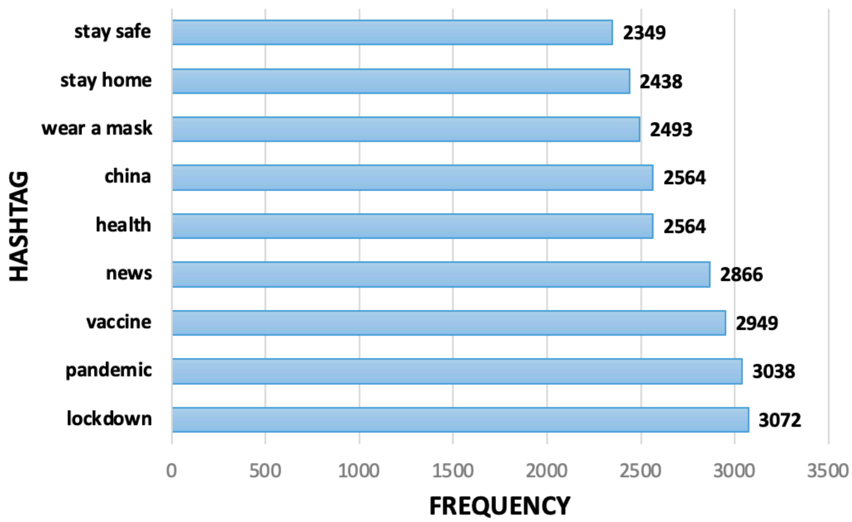
Graph 2



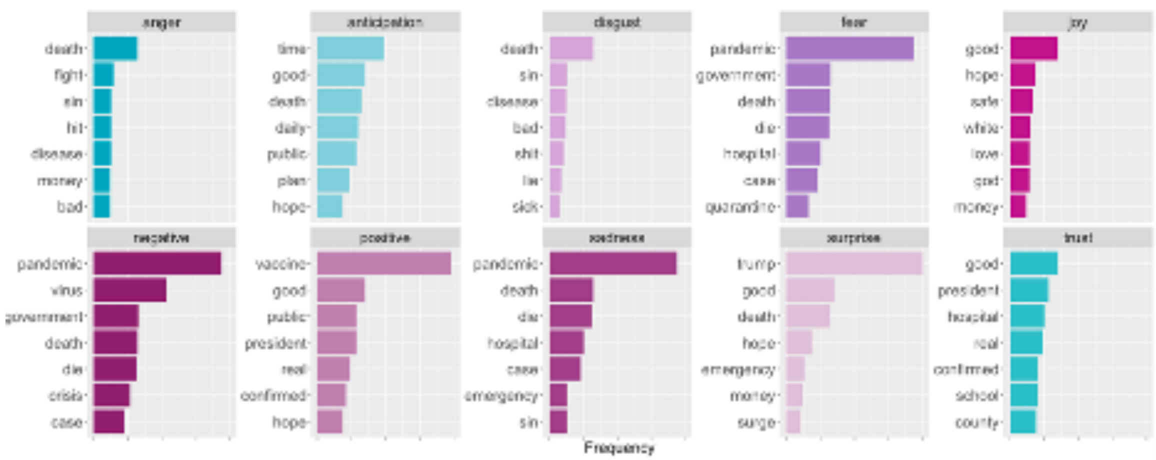
Graph 3



Graph 4

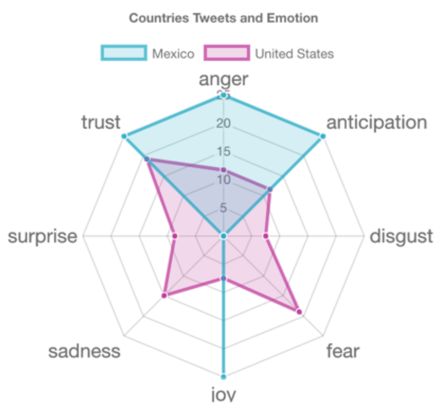


Graph 5



Graph 6

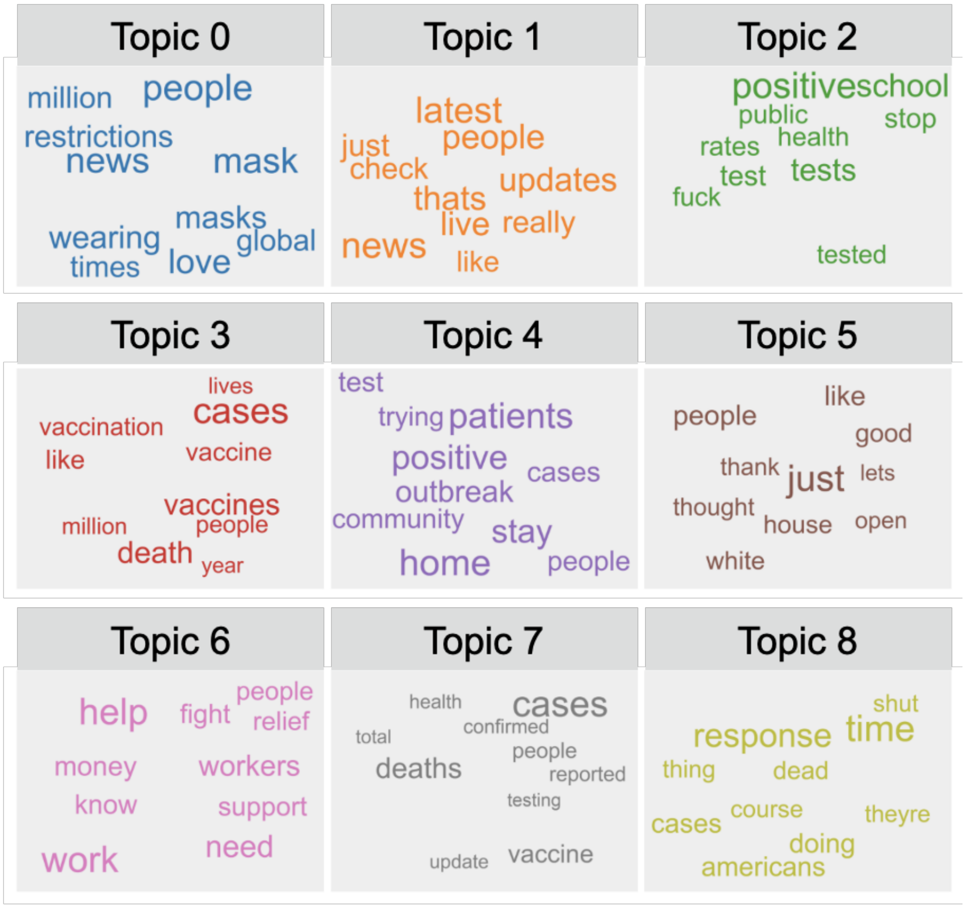
Graph 7



Topic Modeling

Graph 8

Top 9 topics (10 words of highest frequency)



Word2Vec Graphs

Graph 9: Model

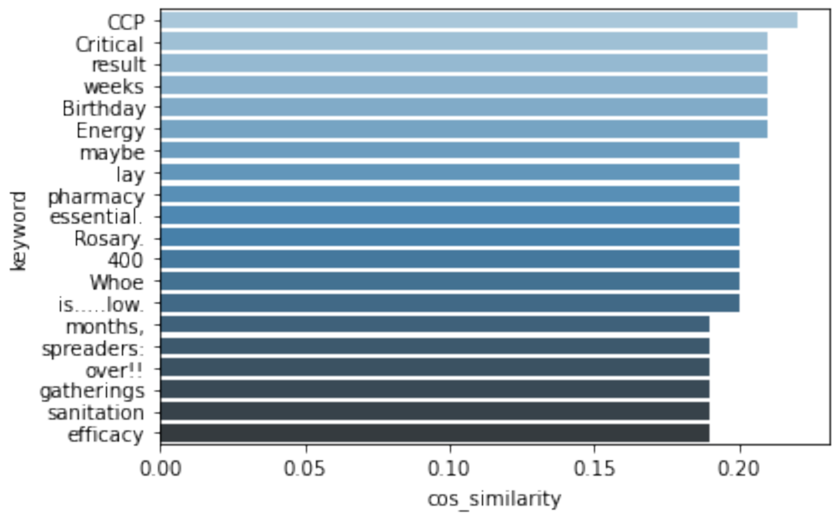
```
model = Word2Vec(sentences=df_english['text_processed'].tolist(), size=256, min_count=1, window=9,workers=-1,seed=12
```

```
vocab = model.wv.index2word
```

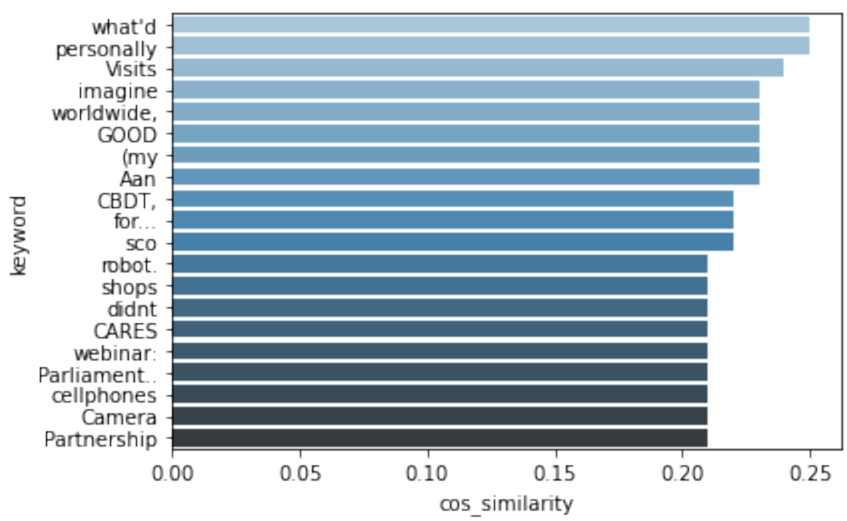
```
len(vocab)
```

```
26772
```

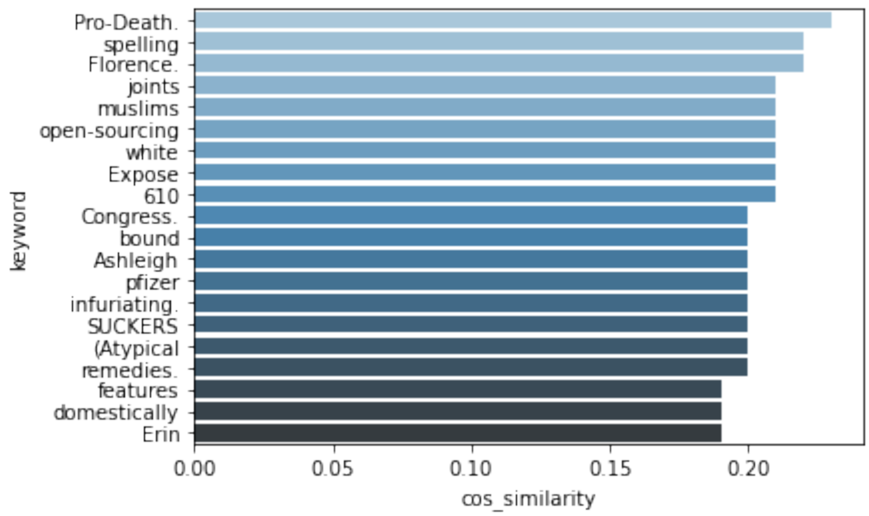
Graph 10: Lockdown



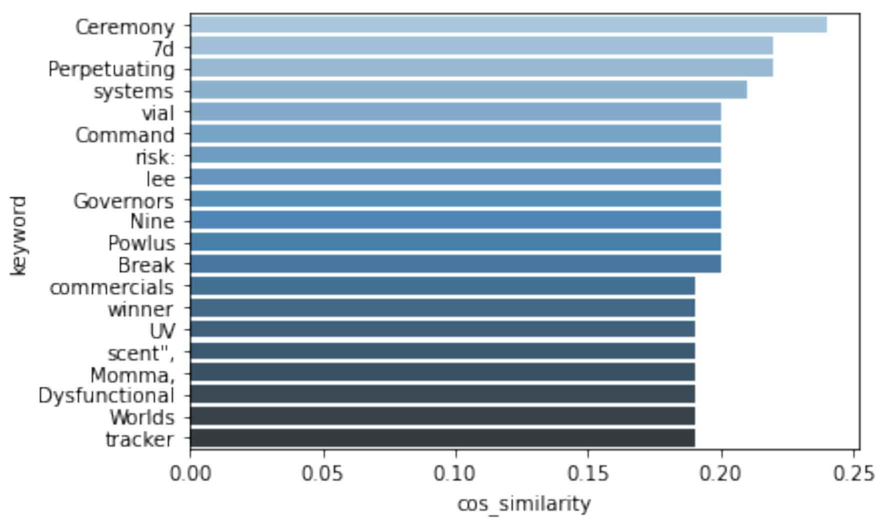
Graph 11: Work



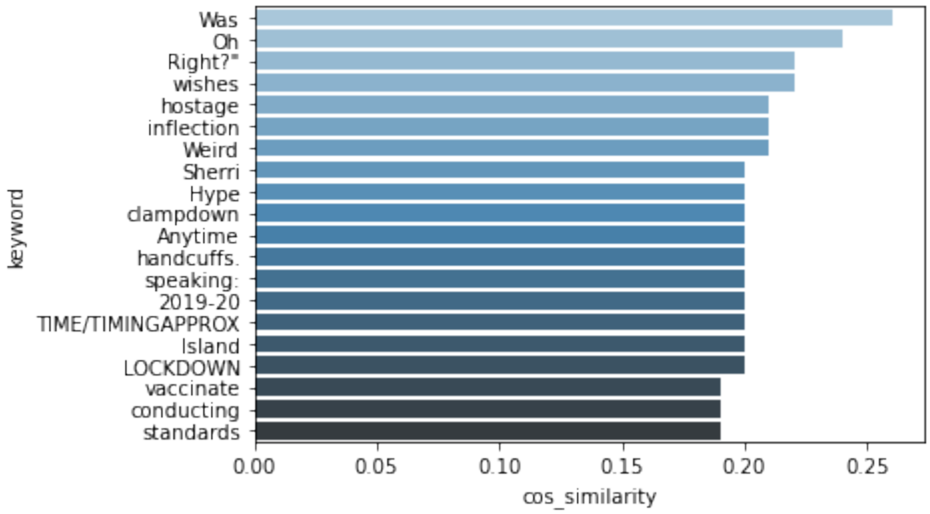
Graph 12: Self-employed



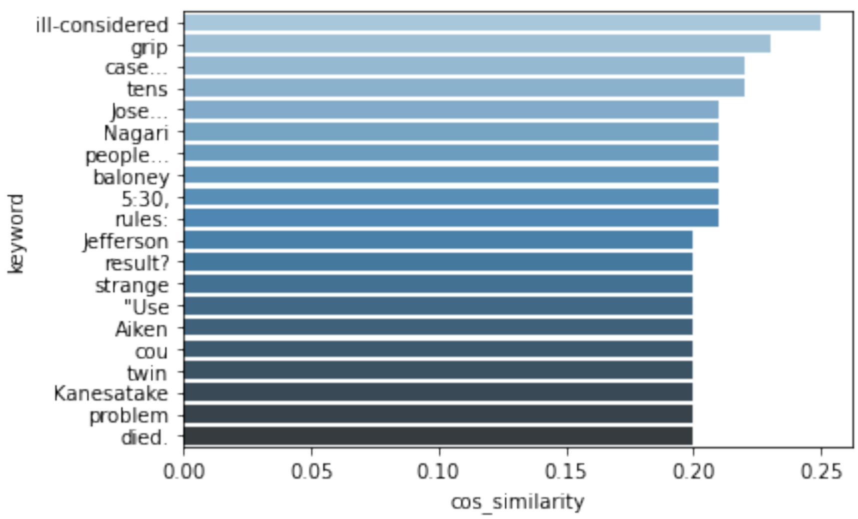
Graph 13: Management



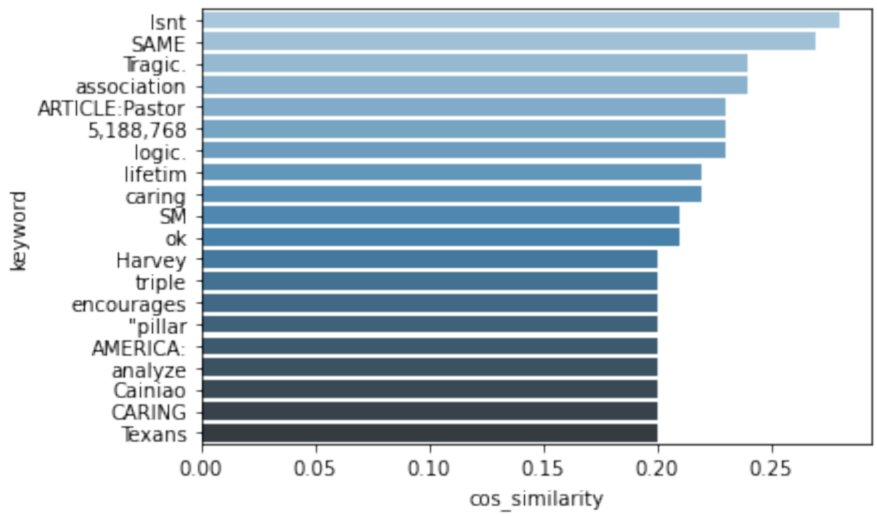
Graph 14: Employee



Graph 15: Remote



Graph 16: Home



Graph 17: Payroll