Analyzing Mental Health and Loneliness During COVID-19

Simone Rittenhouse, Ashley Brill, and Evan Lau
New York University

Contributions

Simone was responsible for the coding and Methods section of the report. Ashley was responsible for the Abstract and Results section of the report, as well as contributing to the Discussion section. Evan was responsible for the Background and Limitations section of the report, as well as contributing to the Discussion section.

Abstract

The COVID-19 pandemic posed many unique and unprecedented challenges for individuals across the world. Many individuals were forced to quarantine by themselves as a result of nationwide lockdowns in many different countries. The uniqueness of these circumstances come not only from the global health crisis, but also because individuals had access to advanced technology unlike any other time in history. People had access to social media and were able to express their thoughts, feelings, and experiences throughout their time in isolation. These posts to social media now serve as a record of how peoples' mental health was impacted by the pandemic. We can analyze these posts, specifically Tweets, to see how topics regarding health and mental health changed throughout the course of the pandemic and how peoples' attitudes also changed.

We hypothesized that many of the topics being discussed during this period revolved around physical health and emotional well-being as a result of people being stationary and lack of mental health resources being available during lockdown. Additionally, we expected sentiment of the Tweets to become more negative over time. To test these ideas, we scraped Tweets from May 1st to July 1st 2020 and used Topic Modeling techniques, such as Latent Dirichlet Allocation, and dictionary-based sentiment analysis on them. The topics that appeared most frequently in relation to the pandemic were loneliness, isolation, and health. Additionally, the intensity of negative sentiment was consistently higher than positive sentiment throughout the two month period.

Background

Throughout 2020, the world was faced with unprecedented health challenges - both physical and mental - due to the COVID-19 pandemic. Entire countries entered periods of lockdown, asking citizens to remain in their homes for weeks on end. Even without mandated lockdowns, the rapid spread of the virus and the rising death toll kept many sheltering-in-place. Now in 2022, as COVID-19 restrictions are being relaxed and the world moves towards a new normal, it is important to study mental health changes in the pandemic's initial wave in an effort to understand the lingering impacts that individuals carry into the present-day.

One study conducted in the early part of the virus' spread within the United States found that, in April and May of 2020, "Combined prevalence estimates of current depression, initiating or increasing substance use, and suicidal thoughts/ideation were 28.6%, 18.2%, and 8.4%, respectively" (McKnight-Eily et al., 2021). This shows a dramatic rise from the United States' 2019 National Health Interview Survey estimate of depression in those above the age of 18, which was 7.0% (Villarroel & Terlizzi, 2020, p. 5). The rising rates of depression and suicidal ideation in the early parts of the pandemic hint at the increased social isolation that the world experienced during this time.

Increased negative mental health outcomes throughout 2020 may have been symptoms of the underlying feelings of loneliness caused by lockdowns and social distancing. This is discussed in another study which administered a survey to measure loneliness during the third week of lockdown in the United States. Their results show that, "Loneliness was elevated, with 43% of respondents scoring above published cutoffs, and was strongly associated with greater depression and suicidal ideation" (Killgore et al., 2020). Taken as a whole, research shows that the initial stages of the pandemic were accompanied by new mental health challenges stemming from isolation due to social distancing measures. Yet, though people may have been more physically distant from one another, online social media platforms offered individuals a lifeline to communicate with and support each other. It is therefore necessary to examine the role of social media during the pandemic.

To answer the question of how individuals dealt with the social distancing that accompanied the early stages of the COVID-19 pandemic, we attempted to replicate and expand on a study done by Jing Xuan Koh and Tau Ming Liew on how loneliness was discussed on social media during this time. Their dataset included 4,492 total Tweets posted by individuals (not organizations) with keyword identifiers such as "loneliness" and "COVID-19" contained in the Tweets. The study used Topic Modeling to identify underlying topics hidden in the speech patterns of Twitter users from May 1st to July 1st 2020. They identified 3 unique topics that they interpreted as "(1) Community impact of loneliness during COVID-19; (2) Social distancing during COVID-19 and its effects on loneliness; and (3) Mental health effects of loneliness during COVID-19" (2022, p. 317).

Our replication and expansion on the study include Topic Modeling of our own and sentiment analysis over the same 2 month timeframe. Although similar methods were deployed, we expect there to be differences in our results because we used a different method for our data scraping. Additionally, our dataset was not limited to Tweets posted by individuals, which will likely also contribute to differences in the topics that emerge. Prior to analyzing the data, we developed the following hypotheses around what topics may emerge and how sentiment trends may change over time:

> **<u>H1</u>**: Conversations around isolation during COVID-19 may be split into two general topics, one relating to physical health due to being stationary in lockdown (e.g. discussions of the symptoms of COVID-19 and its impact) and another relating to emotional well-being due to limited access to resources (e.g. discussions of mental health).
>
> **<u>H2</u>**: The intensity of negative sentiment increased and positive sentiment decreased as lockdowns continued due to feelings of hopelessness, fatigue, and the rising number of COVID-19 cases.

We hypothesize that only two distinct topics will emerge instead of the three found by Koh and Liew because of the inclusion of Tweets posted by organizations. It seems likely that organizations, particularly news outlets, would focus more on the global impacts of COVID-19 and less on specific and individual-level mental health challenges. Conversely, individuals may be more likely to have discussions along the same themes found by Koh and Liew, primarily relating to emotional well-being and the isolation experienced during social distancing.

Methods

In an effort to follow the methods outlined by Koh and Liew, Twitter data was collected between May 1st and July 1st 2020. Twint, a Twitter scraping tool that does not require API access, was used to pull Tweets from this time period with the constraints that each Tweet contained the key terms "COVID-19" and "loneliness." An additional requirement was placed to ensure that each Tweet was written in English. In total, 1,127 Tweets were scraped from days ranging from May 1st to June 29th 2020. Several days were omitted from the originally-specified range as Twint returned that there was no data to scrape from these days. It is unlikely that no Tweets exist containing the specified keywords on these omitted days; however, this will be further discussed in the limitations section. The size of this corpus is about one-fourth of the size of that used by Koh and Liew, who used 4,492 Twitter feeds in their analyses. This difference in sample size is another reason why we expect our results to differ from Koh and Liew's findings.

The collected data was then pre-processed for analysis. Capitalization was removed to ensure that all words were treated identically independently of whether or not they appeared at the beginning of a sentence. All punctuation was removed except for '!', which was kept in the corpus under the assumption that this character may affect the intensity of sentiment expressed and was therefore relevant to our analyses. Python's NLTK package was used to obtain a list of stop words to remove from the corpus. In addition to the words specified by NLTK, digits, URLs, and non-ASCII characters such as emojis were dropped. The final text that remained after these steps was tokenized under the Bag of Words assumption using NLTK and normalized with the WordNetLemmatizer to remove uninformative prefixes and suffixes like the 's' in plural words. This tokenized data was then transformed into numerical data to use as input for Topic Modeling through tf-idf vectorization. This method was chosen over word count as tf-idf takes into account the relevance of a word in a document weighted by its rarity in the corpus as a whole - meaning that tokens found in high quantities throughout the corpus like the obligatory keywords 'covid' and 'loneliness' were downweighted. This was preferable since these words are not necessarily informative to the topics being discussed around COVID-19 and loneliness.

To test our first hypothesis, the resulting tf-idf matrix was used to conduct Topic Modeling and assess the kinds of conversations surrounding isolation during this period of the pandemic. Latent Dirichlet Allocation was used to find keyword distributions within the specified number of topics. The optimal number of topics was set to be three, as this was the number of unique themes found by Koh and Liew (2022), and the maximum number of iterations over the data was set to be 20. This is twice the number of iterations as the default value in scikit-learn, the package used to conduct Topic Modeling, which we found appropriate as the corpus being modeled is relatively small and therefore may be more affected by random noise and require extra training time for a meaningful output. For each of the three topics specified, the top twenty words ranked by weight (roughly corresponding to the number of times a word was assigned to a particular topic) were outputted to interpret what types of discussions were present in the corpus.

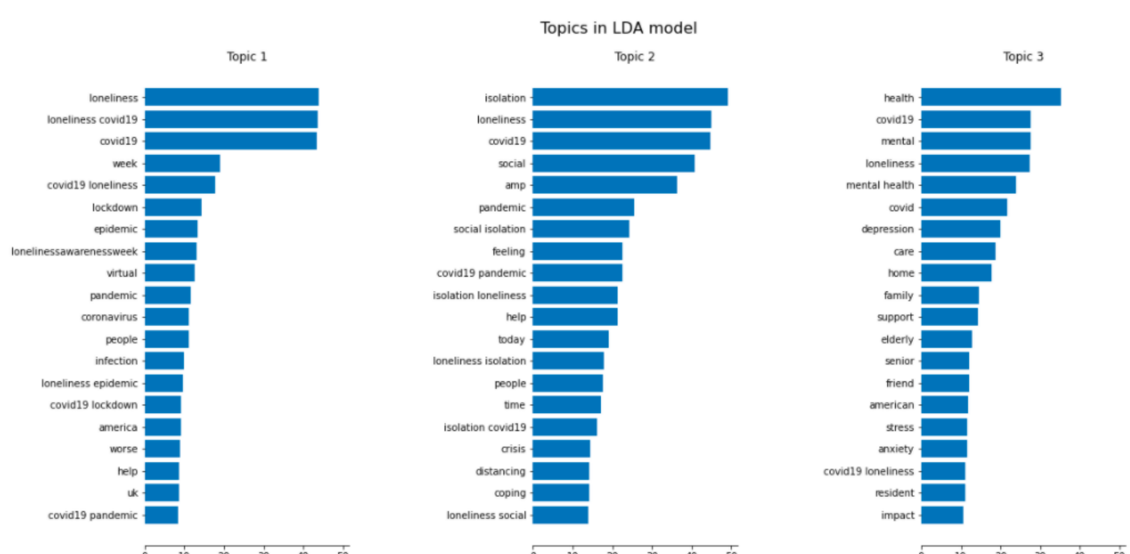
An additional analysis was conducted to explore whether or not the frequency of conversations around these topics were stable over time. This was first done by removing any words that appeared more than once in the top twenty words per topic. The remaining words - which were unique to each topic - were then treated as keywords to assess the frequency of discussions in a particular topic. The counts of these keywords were computed for each day in the corpus such that the count of a topic per day was the summation of the counts of each of this topic's unique keywords. After finding the raw counts of each topic per day, rolling averages were taken with a window size of three in order to more easily interpret the overall trends in frequency over time. Because certain dates in the corpus had a higher overall word count than others, the rolling average of total word count per day was also computed to examine the extent to which the frequency of a topic was moderated by overall word count. The relationships between each topic and overall word count - as well as with the frequency of all other topics - was found using a correlation matrix.

To test our second hypothesis, dictionary-based sentiment analysis was conducted using several pre-built and readily available sentiment dictionaries. Dictionary-based analysis was chosen over classification as the corpus was unlabeled, making it difficult to use to train a classifier. The AFINN dictionary was used as a baseline, which was compared to NLTK's

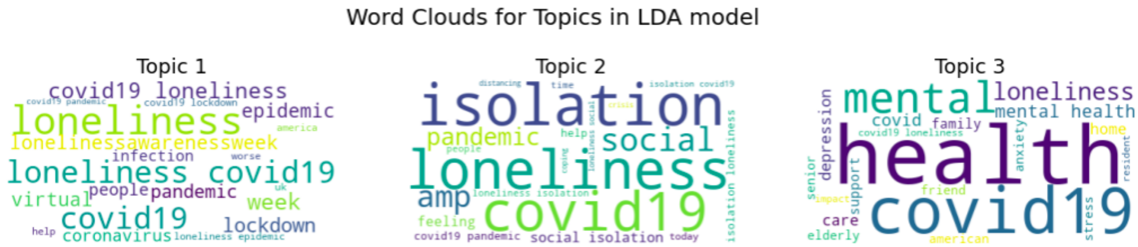
Opinion Lexicon and NLTK's VADER Sentiment Intensity Analyzer. For each dictionary, sentiment scores were calculated for every Tweet in the pre-processed corpus. Then, to assess the intensity of positive and negative sentiment over time, positive sentiment was defined as Tweets with a sentiment score greater than zero, while negative sentiment was defined as Tweets with a score less than zero. The intensity of these quantities per day was defined as the sum of the absolute value of the sentiment scores of the 'positive' or 'negative' Tweets divided by the count of 'positive' or 'negative' Tweets on that day. The total sentiment was normalized by the number of Tweets corresponding to a particular sentiment to avoid days with more collected Tweets having the illusion of more intense sentiment. Rolling averages were taken of the normalized positive and negative intensities using a window size of two in order to see the overlying trends in positive and negative sentiment over time. In addition to computing these quantities for each dictionary, the normalized positive and negative intensity was found using the same method on the average sentiment of each Tweet across the three dictionaries tested.

Results

As shown in **Figure 1**, there were three distinct topics that arose as a result of the LDA analysis. Topic 1 includes terms related to the western world and growing concerns over government handling, like "uk," "america," and "lockdown." Topic 2 features words related to actual feelings of loneliness in relation to quarantine very frequently. Finally, Topic 3 revolves around stress and anxiety about being away from family and loved ones. Unlike the first two topics, Topic 3 includes words like "elderly," "support," and "family," indicating the topic mainly concerns the impact that being away from family, and specifically elders, had on peoples' mental well-being.



a.



b.

Figure 1. (a) The top 20 words per topic ranked by weight. **(b)** Word clouds of the top words per topic.

Topics 2 and 3 were highly correlated with the word count of a particular day. **Figure 2** shows that over time, the two often overlapped. This indicates that the two were likely discussed at a similar rate depending on how many Tweets were recovered for any given day. Additionally, **Figure 3** shows that Topics 2 and 3 had a correlation of 0.606, therefore the two were relatively correlated. This could indicate overlapping conversations between the two subjects, or that the two were discussed in tandem.

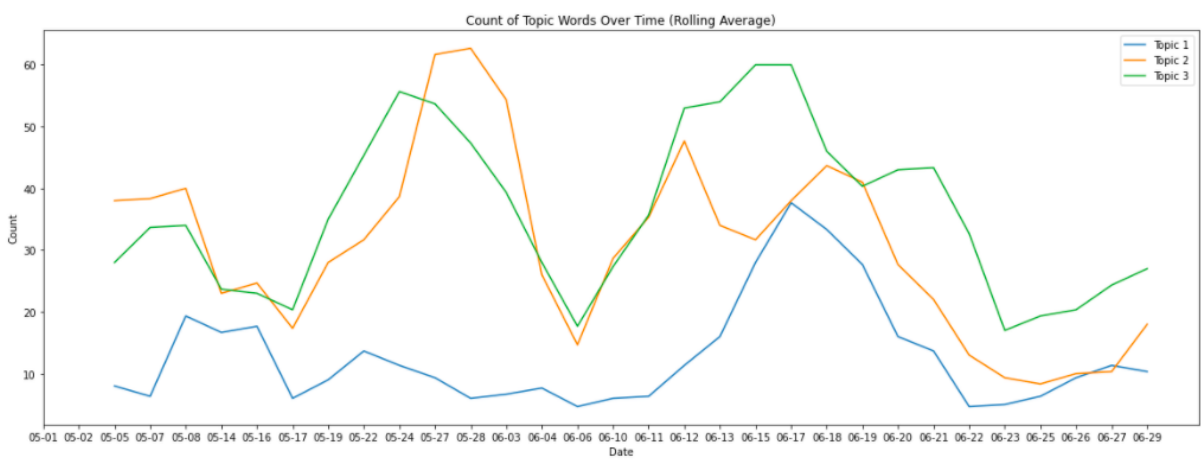
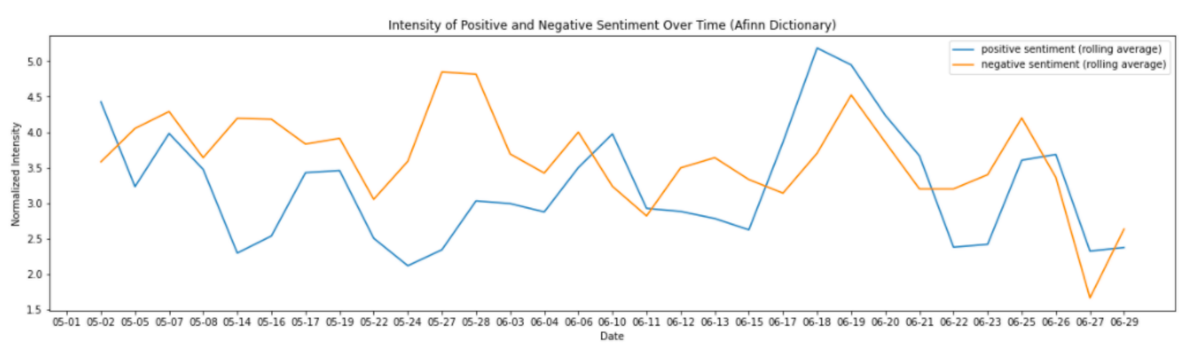


Figure 2. Counts for topic-specific words over time.

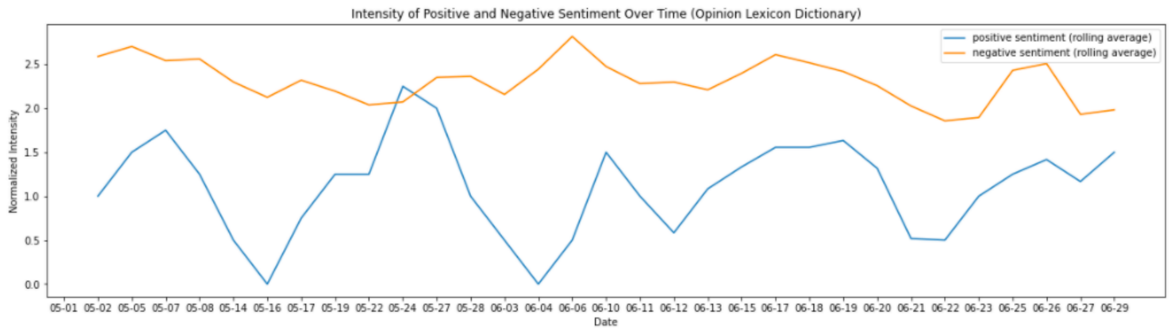
	Topic 1	Topic 2	Topic 3	Full Word Count
Topic 1	1.00000	0.251600	0.453200	0.556890
Topic 2	0.25160	1.000000	0.605841	0.765244
Topic 3	0.45320	0.605841	1.000000	0.902834
Full Word Count	0.55689	0.765244	0.902834	1.000000

Figure 3. Correlations between topic-specific word counts and overall word count per day.

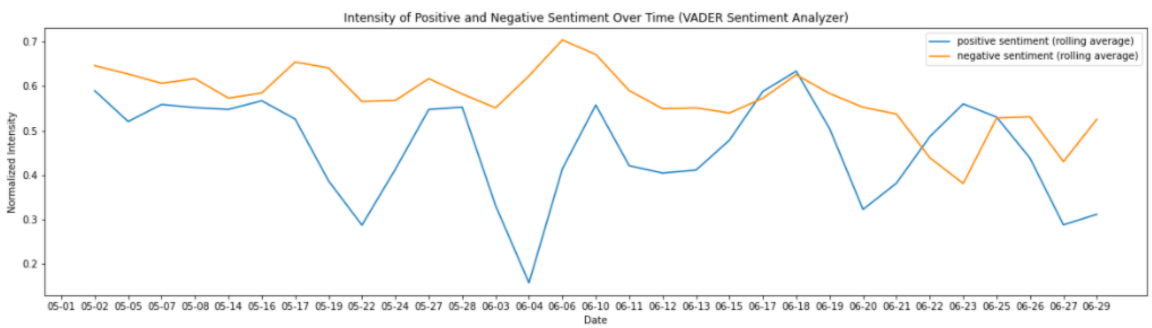
Figure 4 displays the results of the different methods of dictionary based sentiment analysis. **Figure 4a** shows the results from the AFINN dictionary based analysis. In this graph, we can see that the positive sentiment was more intense for the majority of the time period, however, the intensity of the negative sentiment increased drastically following 6/13/2020 and peaked on 6/20/2020, where it was more intense than the positive sentiment. The intensity of positive sentiment has a sharp decrease at the end of the timeframe. Similar to the AFINN dictionary-based method, the opinion lexicon dictionary analysis showed the intensity of positive sentiment was consistently higher than that of the negative sentiment through this period. Unlike the AFINN method, there was no drastic peak of the negative sentiment where it was more intense than the positive sentiment. In the VADER model, **Figure 4c**, the positive sentiment has a relatively consistent intensity throughout the two month period. The negative sentiment experiences more changes in intensity throughout the time, however, it is only briefly more intense than the positive sentiment. Finally, **Figure 4d**, the averaging dictionaries, once again shows a consistent intensity for the positive sentiment and many changes to the negative sentiment with only brief periods where the negative sentiment was more intense than the positive sentiment.



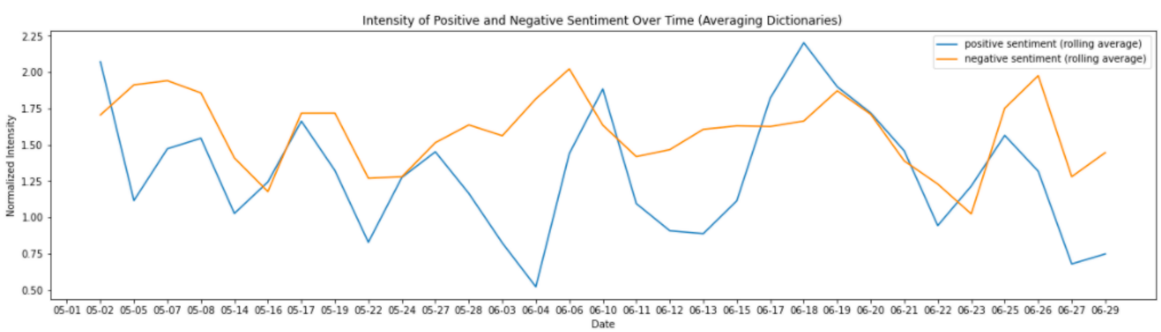
a.



b.



c.



d.

Figure 4. Average positive and negative sentiment intensity per day using **(a)** the AFINN dictionary, **(b)** the Opinion Lexicon dictionary, **(c)** VADER Sentiment Analyzer, and **(d)** an average of the three dictionaries.

Using the average sentiment scores across the three dictionaries, the Tweet with the highest positive sentiment was found to be:

> "‼️ Great news that @TNLComFund have awarded @GVATweets over £99k to help them continue their hard work helping our community combat Covid-19 & loneliness this #LonelinessAwarenessWeek. 👏 Thanks to all the brilliant volunteers & staff for their hard work."

While the Tweet with the highest averaged negative sentiment was:

> "This COVID-19 sh*t sucks. The symptoms are bad but worst of all the depression and loneliness it hits you with is way worse…"

These findings suggest that positive sentiment with respect to loneliness in the pandemic may have been related to social support and community involvement while negative sentiment was based more in personal struggle and a feeling of hopelessness amidst isolation.

Discussion

Referring back to our initial hypotheses, we predicted that the topics that would emerge from topic modeling would relate to COVID-19's toll on physical health and emotional well-being. While we were correct in the sense that topics surrounding mental well-being did emerge, the conversations concerning mental health included topics that we did not initially hypothesize. For example, the topics that arose were related to rising fears around the government's handling of the pandemic and growing worries about being away from loved ones. These topics do have to do with mental health and well-being, however, in relation to themes that were not a part of our original hypothesis. None of the topics included keywords that may indicate conversations suggesting concerns over physical health from being lonely or isolated. Additionally, the topics that emerged were somewhat in line with those found by Koh and Liew, as several of the topic keywords related to the third topic they found: "Mental health effects of loneliness during COVID-19" (2022, p. 317).

For our sentiment analysis, we hypothesized that the intensity of negative sentiments would increase and intensity of positive sentiments would decrease over the two-month period. This was not necessarily correct, as our results across the averaged dictionaries suggested that positive sentiment intensity consistently vacillated with no clear downward or upward trend. Similarly, negative sentiment intensity remained consistently higher than positive sentiment; however, this did not drastically change over time. It is possible that negative sentiment was relatively stable over time as rates of depression had already increased and stabilized. This possibility is supported by McKnight-Eily et al., who found elevated rates of depression in April and May, which overlapped with the first half of our corpus' time frame. Therefore, it's possible that, rather than becoming more hopeless over time, individuals were already facing higher levels of negative mental health outcomes - even at the start of our study. Though the results for each dictionary were relatively similar, the AFINN dictionary results provided us with key insights. Specifically, the dramatic increase in negative sentiment intensity that occured on 6/13/20 might be related to the Trump Administration reversing health protections for transgender people (Simmons-Duffin, 2020). Additionally, the peak in negative sentiment intensity on 6/20/2020 may have been due to the UK's mishandling of the COVID-19 pandemic ("Nearly 1,500 deaths in one day: UK ministers accused of downplaying covid-19 peak", 2020). These major events in the western world seem to have had a visible impact on our results and are reflected in the rises in negative sentiment on those particular days.

Limitations

Though we tried to replicate the study as closely as we could, there were limitations of the data that we were working with. Since we didn't use a traditional Twitter API to scrape Tweets, we resorted to using Twint, which only yielded a fraction of the number of Tweets that was used in the study. Since Topic Modeling usually requires a lot of data for emerging topics to become interpretable, we may not have gotten obvious differences in topics, which was shown in the overlap of words in topics 2 and 3. Additionally, we found that the correlation between topics 2 and 3 was 0.6, suggesting that there was a decent amount of overlap in their frequencies.

Another issue that came up during the data scraping process was reasonable doubt that the Tweets were from a random sample. Some days in between the 2 month window did not yield any Tweets due to inconsistencies in Twint's performance and output. If the scraping tool only yielded Tweets for some days and not others in a systematic way, it may not capture the overall sentiment of Twitter users. Moreover, if the data was not truly randomly sampled, then we may see biases in our results caused by selection bias.

One separate limitation of our analyses was that we used dictionary-based sentiment analysis. This was chosen over classification since the corpus was unlabeled; however, this left our results somewhat unstable and highly dependent on the dictionary used to compute sentiment scores. Each dictionary included different terms weighted with differing levels of positive and negative sentiment. As shown in our results, using three different dictionaries on the same corpus yielded three different outcomes in terms of the corpus' overall positive and negative sentiment - which also led to slightly different trends in sentiment over time. Though measuring sentiment can be difficult and somewhat subjective, ideally each analysis should yield similar results since the dictionaries were used on the same data. Because this was not the case, our sentiment analysis results appear to be highly dependent on the dictionary used, which should be taken into consideration when discussing the generalizability and conclusivity of our findings.

A final limitation is that the data used was completely in English, which may include western biases in the topic models and sentiment analysis. Our analysis explicitly takes into account only the sentiment of English speakers, which may disregard the impact of the pandemic on mental health in non-English-speaking countries. It is important to acknowledge this since different governments had different lockdown policies. For example, Topic 1 focuses a lot on the western world's views on the government's (mis)handling of the pandemic by downplaying its severity. Furthermore, many media outlets in the United States, as well as influential media figures, downplayed the pandemic so much as to call it a "hoax," which generated much controversy on social media. These trends in western society are likely to be reflected in the English Tweets, and thus we can only conclude that the topic models and analysis represent western sentiments on COVID-19.

Future research could focus on a more global view of mental health in the early stages of the pandemic by including non-English Tweets or by using geolocation to sample Tweets from a variety of regions. An additional future direction could widen the scope of our research by including data from earlier in the pandemic. For example, including data as far back as February or March of 2020 could provide more insight as to how sentiment shifted in the first stage of the pandemic, as this time period is when many countries observed their first COVID-19 outbreaks.

References

(1) Guardian News and Media. (2020, June 19). *Nearly 1,500 deaths in one day: UK ministers accused of downplaying covid-19 peak*. The Guardian. Retrieved May 6, 2022, from https://www.theguardian.com/world/2020/jun/19/over-1000-deaths-day-uk-minister-accused-downplaying-covid-19-peak

(2) Killgore, W. D. S., Cloonan, S. A., Taylor, E. C., & Dailey, N. S. (2020). Loneliness: A signature mental health concern in the era of COVID-19. Psychiatry Research, 290, 113117. https://doi.org/10.1016/j.psychres.2020.113117

(3) Koh, J. X., & Liew, T. M. (2022). How loneliness is talked about in social media during COVID-19 pandemic: Text mining of 4,492 twitter feeds. *Journal of Psychiatric Research*, *145*, 317–324. https://doi.org/10.1016/j.jpsychires.2020.11.015

(4) McKnight-Eily, L. R., Okoro, C. A., Strine, T. W., Verlenden, J., Hollis, N. D., Njai, R., Mitchell, E. W., Board, A., Puddy, R., & Thomas, C. (2021). Racial and Ethnic Disparities in the Prevalence of Stress and Worry, Mental Health Conditions, and Increased Substance Use Among Adults During the COVID-19 Pandemic - United States, April and May 2020. MMWR. Morbidity and mortality weekly report, 70(5), 162–166. https://doi.org/10.15585/mmwr.mm7005a3

(5) Simmons-Duffin, S. (2020, June 12). *Transgender Health Protections reversed by Trump administration*. NPR. Retrieved May 6, 2022, from https://www.npr.org/sections/health-shots/2020/06/12/868073068/transgender-health-protections-reversed-by-trump-administration

(6) Villarroel MA, Terlizzi EP. Symptoms of depression among adults: United States, 2019. NCHS data brief, no. 379. Hyattsville, MD: US Department of Health and Human Services, CDC, National Center for Health Statistics; 2020.