# Public Perception during the Spread of the Delta Variant: A Case of North America


**Supervisor:**
Dr. Ussama Yaqub

**Credit Hours:** 3

**Members:**

Omer Imran 2022-11-0024
Laibah Iqbal 2022-11-0006

## Table of Contents

## 1. Introduction

The COVID-19 pandemic has been deeply costly for human lives since it began to spread in late 2019. The World Health Organization declared it a pandemic on 11th March 2020 and since then it has taken 5.3 million lives. The human toll aside, it upheaved the lives of millions around the world. Economies spiralled, healthcare systems collapsed as the world came to grips with this disease. Some regions had it worse than others due the emergence of variants such as Delta and Lambda. Countries like India were reporting over 295,000 daily cases during its 2nd wave in April 2021 (Safi). The phenomenal speed with which it was ravaging the country was eye-opening for many experts around the world.

As more variants like Omicron emerge, more dangerous waves will sadly follow. It's more important than ever for governments to be vary of what their publics fear most. During these uncertain times, this could mean the difference between depleting the morale of a nation and surviving the pandemic. Studies like those of Behl et al. and Marcec and Likic have successfully done this before. As such, we propose using twitter tweets to analyze public sentiment and tag topics of concerns during the pandemic so that governments can be aware of online discourse and keep their citizens free of panic. This study aims to do the same; analyze sentiment and model topics from tweets in the North American region while diving into the following research questions:

**Research Question 1:** *What are common topics of discourse on Twitter in North American countries during the course of the pandemic?*

**Research Question 2:** *What was the sentiment of countries in North America on topics of concern that emerged during the course of the pandemic?*

## 2. Literature Review

To understand why social media platforms like Twitter are being used in research today, you have to realize the sheer volume of the information that is coming in. With millions of active users and tweets on a daily basis, the microblogging website is ripe with information that can be used to generate insights. One of the many ways this is possible, is using the power of sentiment analysis and topic modelling. This allows researchers to not only identify user perception on various topics but additionally, produce results that are then used in fields with the implications being limitless.

A prominent example of sentiment analysis in action is in the field of finance. In research conducted by Mittal and Goel, they found that they could accurately predict “market sentiment” i.e. predict stock market movement using sentiment analysis of twitter tweets. In a similar fashion, Karalevicius et al. were able to find a relationship between twitter sentiment and bitcoin. In fact, sentiment was considered such an effective indicator of stock price movement that Nguyen, Shirai, et al created a model that incorporated sentiment about various topics related to a company into an existing model to predict stocks. Interestingly, this model outperformed the model that only used historical prices in terms of accuracy by 6.07%. To further improve upon the accuracy of methods that are using explicit sentiment labels for comments, Deirakshan and Beigy introduced their “LDA-POS” method, showing that sentiment analysis is a powerful tool that can be even better utilized within this domain.

Another compelling application of tweets is in the realm of politics. With citizens frequently taking their political discussions online, one can imagine predicting elections could be a possibility. Researchers Budiharto and Meiliana, successfully produced predictions similar to reliable survey institutes in Indonesia to predict who would be the next president. Similarly, Yaqub et al were able to find out how sentiment of Hillary Clinton and Donald Trump supporters varied during the 2016 US presidential campaign. In another research Yaqub, Sharma, et al were able to confirm twitter sentiment during election campaigns corroborated with the actual results in both of the case studies in the research.

Another domain in which sentiment analysis proves relevant is disaster relief. Understanding the sentiment around disaster in different regions can allow authorities to identify gaps in resource distribution and better cater to affected populations. Ragini, Anand, et al's paper details their classification method which specifically deals with the challenge of the sheer amount of data generated on social media during disasters. Similarly, Behl, Rao, et al used a particular setting of multilayer perceptron layers and an optimizer trained on public data from the Nepal earthquake and Italy earthquake to achieve an accuracy of 83% on their COVID-19 test data.

As COVID-19 emerged, researchers made use of the plethora of information flowing in for a wide-variety of applications. Hussain et al used deep learning techniques to categorize topics of discussion as well calculate sentiment around vaccines in the UK and US. Another study by Marcec and Likik used lexicon-based sentiment analysis to conclude that the overall sentiment towards the AstraZeneca and Oxford vaccines was negative, and could threaten the success of the worldwide vaccination campaign. On the other hand, studies by Monselise, Chang, et al as well as Hussain, Tahir, et al found a mix of positive and negative sentiment with most concerns regarding vaccines being about supply, distribution and economic viability. Similarly, Yousefinaghani et al used over 43 million tweets to identify sentiment of users around vaccine hesitancy, which helped them ascertain that most of the negative sentiment was coming from twitter bots. This in turn showcases how twitter has been used as a medium of spreading misinformation during the pandemic – researchers have explored this avenue too. Shahi et al. studied such misinformed tweets, diving into how they spread, where they come from and how to identify them. Other researchers have gone on to claim this serious problem as an infodemic (Jamison et al.). A study by Chakraborty, Bhatia et al to analyze the spread of information during the pandemic found that most retweeted tweets related to COVID-19 contained no useful words implying that perhaps other channels would be better at providing accurate information relating to guidelines to citizens during emergencies and that fact checkers are a necessity during crises. Sentiment analysis has also been used to investigate the socioeconomic effects of the pandemic. For example, Nguyen, Criss, et al found that anti-Asian sentiment rose by 68.4% in the period November 2019 to June 2020 compared to the sentiment regarding other minorities which remained relatively stable. Costola, Nofer, et al studied the effects of COVID-19 sentiment on the stock market and found a significant, positive relationship between COVID-19 news sentiment score and market returns. Moreover, a number of studies have tried to identify the sentiment of citizens during the spread of COVID-19. As an example, research found that during March of 2020, as COVID spread across the world, most of the countries were taking a hopeful and positive sentiment to the pandemic (Dubey). Additionally, some researchers like Garcia and Berton, used tweets in both English and Portuguese to analyze sentiment of users in the USA and Brazil – negative sentiment came out to be dominant in their study. Likewise, Pokharel used the same analysis tools to study sentiment in Nepal.

Nonetheless, we could not find any literature that analyzed sentiment and modelled topics of discussion during the deadly wave of the delta variant of COVID-19. Moreover, no study looked at the difference across the countries in a continent where one of the countries had developed into a hotspot for the variant such as India in Asia or UK in Europe. This is where we believe our knowledge can add to the pool of research.

## 3. Research Questions

The COVID-19 Delta variant was first detected in India in late 2020, approximately one year into the pandemic when most of the worldwide population was coming to terms with the public health crisis. Given the toll of the variant on the already strained public health services and the ensuing aggravation of the psychological, socioeconomic and political effects of the pandemic on various countries, we wanted to study what were the most frequently discussed topics during the spread of the delta variant and the sentiment regarding these topics.

**Research Question 1:** *What are common topics of discourse on Twitter in North American countries during the course of the pandemic?*

As the delta variant emerged in India, it started to spread in other parts of the world too. Soon enough, countries like Russia, USA, UK, and more began to face the deadly variant in their homelands. As aforementioned hotspots developed, neighbouring countries in said hotspots prepared to defend themselves against this variant as well. This brings up a compelling case for investigating how sentiment varied around modelled topics in countries neighbouring those which were facing the variant. We limit this investigation to continents where the delta variant swiftly spread to. We propose the following question,

**Research Question 2:** *What was the sentiment in North America on topics of concern that emerged during the course of the pandemic?*

## 4. Research Methodology

**Data**

For this study, we downloaded a dataset of tweet IDs by Banda et al. This is a large-scale curated dataset of over 1.3 billion tweets related to COVID-19 chatter, collected from January 1st, 2020 onwards. Moreover, the dataset additionally collected tweets from all languages, with retweets included too. It is updated, with more tweets added to the dataset, on a bi-weekly basis. From this dataset, we chose to exclusively consider tweets in English from Canada, Mexico and USA created during three time periods that represent significant events from the pandemic in the region. Our first time period ranged from June $30^{th}$ to July $18^{th}$ 2020. During this time period, the US and Mexico were facing their first major peak in COVID-19 cases. The second time period, ranging from December $30^{th}$ 2020 to January $15^{th}$ 2021 coincided with the start of wide scale vaccinations. Lastly, the third time period, from July 30th to August $15^{th}$ 2021 corresponds to the greatest rise in COVID-19 cases in the region caused mostly by the spread of the Delta variant. We ensured that these tweet IDs did not contain any retweets and hydrated them through the Twitter API to acquire the rest of the information.

**Data Preprocessing**

After hydrating tweets, unnecessary metadata was added by the Twitter API. Our study only required looking at location, text, and time of the tweet. As such, we dropped all other columns. Moreover, some of these tweets did not contain location. Those tweets were dropped too giving us a total of approximately 216,000 tweets. These tweets were then prepared for the rest of the analysis processes. Firstly, the text field of our dataframe was converted to lowercase. This was to ensure that words like “The” and “the” are not considered different. Secondly, we removed stop words. Stop words consist of articles, prepositions, and conjunctions which are abundant but add no information for natural language processing. Lastly, the text was lemmatized, which converts words to present tense and first person.

**Topic Modelling**

For this we used the Latent Dirichlet Allocation algorithm commonly known as LDA. A common text mining technique, this unsupervised algorithm is able to identify common topics in large documents, which in our case are twitter tweets. We utilized the Python's Gensim library created by Rehurek and Sojka to tag the topics from our collection of tweets. After the aforementioned pre-processing, we calculated the coherence scores for each set of tweets of each time frame and used the number of topics with the highest scores for all of them. Using this, we chose 8 topics for this study.

**Sentiment Analysis**

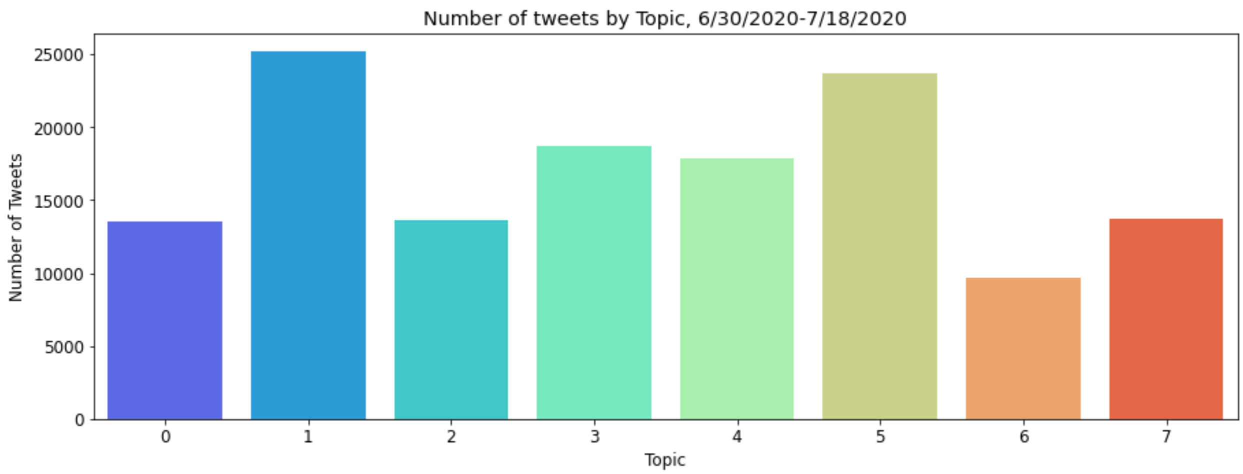
To understand the sentiment of the tweets under study, we used the Python's VADER library, a lexicon and rule-based sentiment analysis tool which works particularly well for social media texts (Hutto and Gilbert). VADER does not require any training data to be set up since it is constructed from a generalizable, valence-based, human-curated gold standard sentiment lexicon (Pandey). Moreover, one of the advantages of using VADER for sentiment analysis is that it does not require data preprocessing for it to run. Additionally, rather than labeling a tweet as "positive" or "negative", the algorithm outputs scores, ranging from -0.05 to +0.05. Scores of -0.05 – 0.00 indicate negative tweets while a score of 0.00 to +0.05 indicate positive tweets. These scores are then normalized on a -1 to +1 scale giving an overall compound score for ease of interpretation.

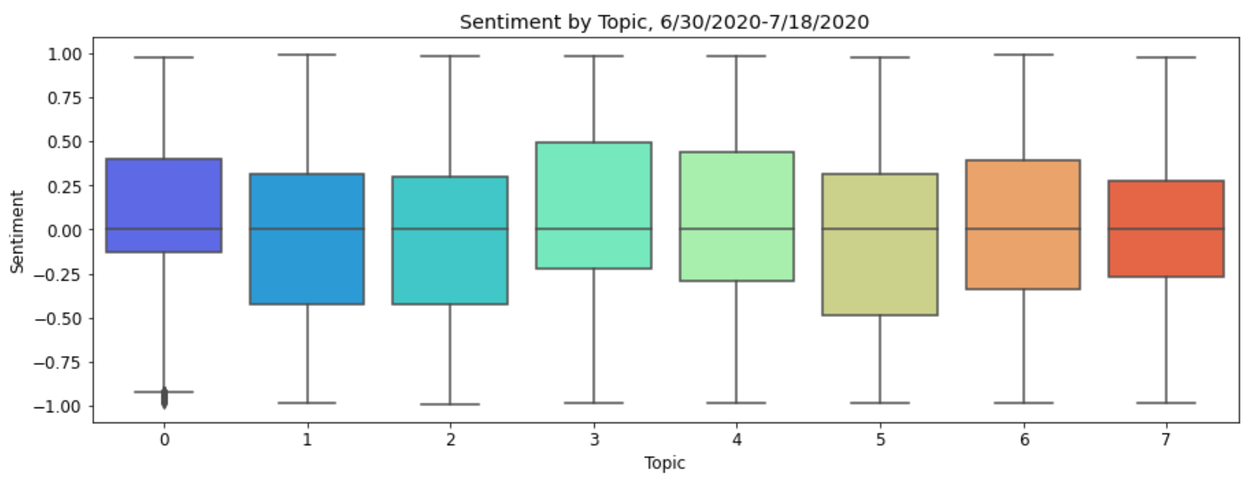
## 5. Results

### Topic Modelling

For all three time periods, we achieved the best coherence score using 8 topics. We extracted the topics for each time period, measured the number of tweets in each topic and the variation in sentiment across each topic.

**1st Wave of the Pandemic** (Jun 30, 2020 - Jul 18, 2020)

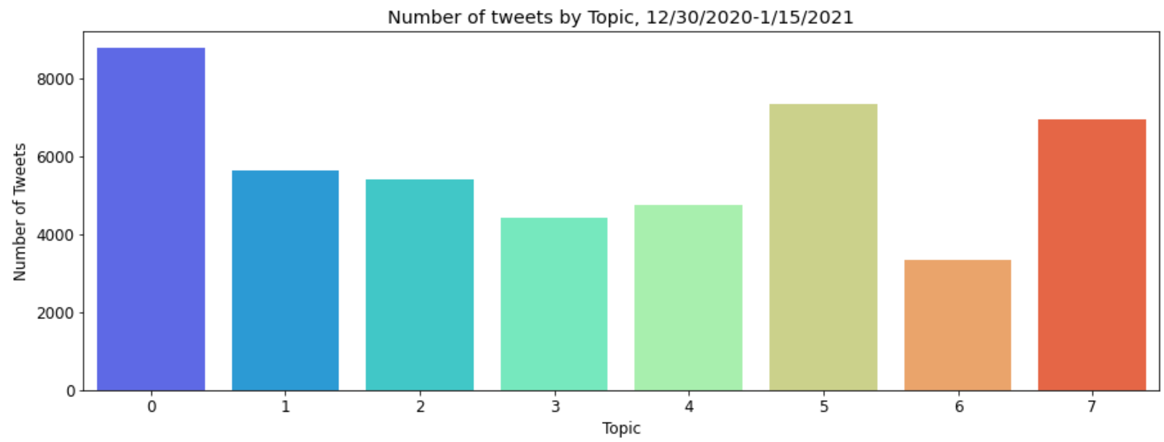


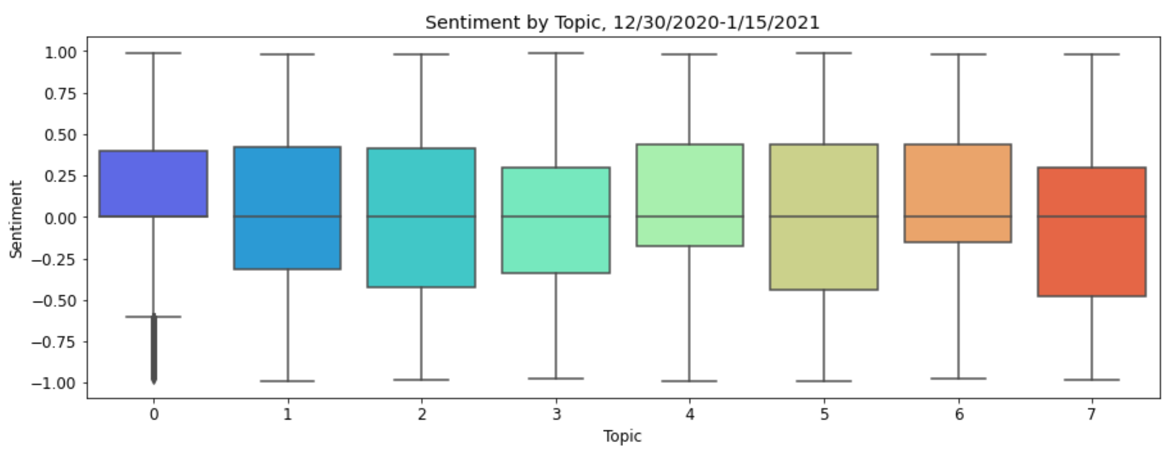



```
[(0,
  '0.020*"tested" + 0.019*"tested_positive" + 0.017*"positive" + '
  '0.017*"infant_tested" + 0.015*"school" + 0.011*"testing" + '
  '0.009*"open_school" + 0.009*"coronavirus" + 0.006*"relief_package" + '
  '0.006*"free"'),
 (1,
  '0.018*"people" + 0.016*"wearing_mask" + 0.015*"mask" + 0.011*"coronavirus" '
  '+ 0.009*"make" + 0.008*"time" + 0.007*"wearing" + 0.007*"think" + '
  '0.007*"virus" + 0.006*"would"'),
 (2,
  '0.027*"school" + 0.026*"trump" + 0.015*"kid" + 0.015*"white_house" + '
  '0.014*"coronavirus" + 0.011*"child" + 0.009*"house" + 0.008*"president" + '
  '0.008*"white" + 0.008*"back_school"'),
 (3,
  '0.018*"coronavirus" + 0.011*"good" + 0.011*"news" + 0.010*"great" + '
  '0.009*"like" + 0.009*"pandemic" + 0.008*"united_state" + 0.007*"state" + '
  '0.005*"people" + 0.005*"stay_safe"'),
 (4,
  '0.039*"mask" + 0.039*"wear_mask" + 0.021*"wear" + 0.013*"please" + '
  '0.010*"home" + 0.009*"people" + 0.009*"fighting_stigma" + 0.008*"stay_home" '
  '+ 0.007*"coronavirus" + 0.006*"stay"'),
 (5,
  '0.019*"trump" + 0.019*"coronavirus" + 0.015*"trump_administration" + '
  '0.010*"like" + 0.010*"people" + 0.009*"realdonaldtrump" + 0.009*"country" + '
  '0.009*"administration" + 0.008*"american" + 0.007*"look"'),
 (6,
  '0.022*"public_health" + 0.015*"test" + 0.012*"coronavirus" + 0.011*"public" '
  '+ 0.011*"health" + 0.010*"test_positive" + 0.006*"line" + 0.006*"covid" + '
  '0.006*"positive" + 0.006*"week"'),
 (7,
  '0.044*"case" + 0.028*"coronavirus" + 0.026*"death" + 0.018*"county" + '
  '0.018*"florida" + 0.016*"state" + 0.015*"number" + 0.013*"texas" + '
  '0.011*"contact_tracing" + 0.010*"data"')]
```

Topic 1, 3 and 5 appear to be the most popular topics in this time frame. Topic 1 had relatively neutral tweets as illustrated by the box plot, although deep diving into the LDA output indicates caution and fear among the general population. In addition to words expressing the need for social distancing in the word cloud, the LDA output shows that key words for this topic included *mask, wearing_mask* and *wearing.* Topic 5, consisting of over 20,000 tweets, was more political and negative in nature, with keywords mostly concerning the Trump administration. Topic 3 appears to be slightly more hopeful with keywords like *good, news, great,* and *stay_safe* as well as an overall positive sentiment.

**Vaccine Distribution Begins in North America** (Dec 30, 2020 - Jan 15, 2021)

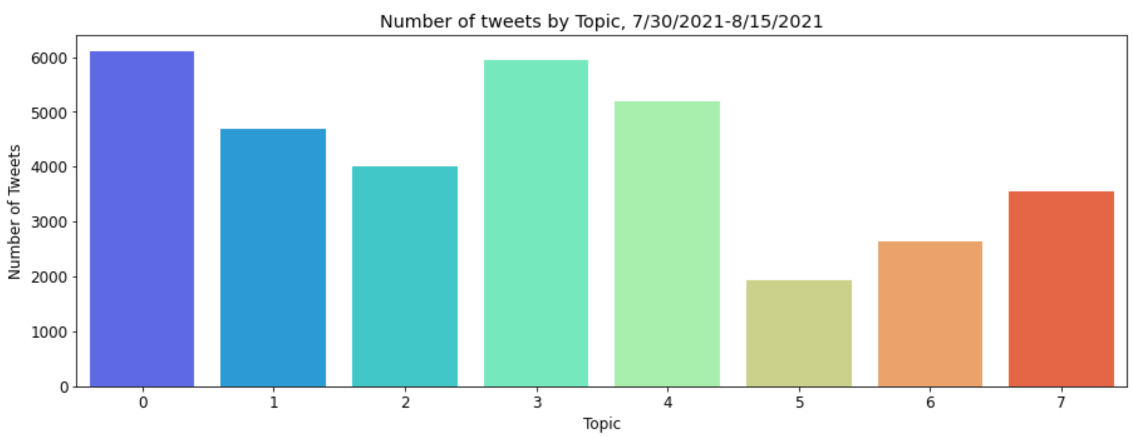


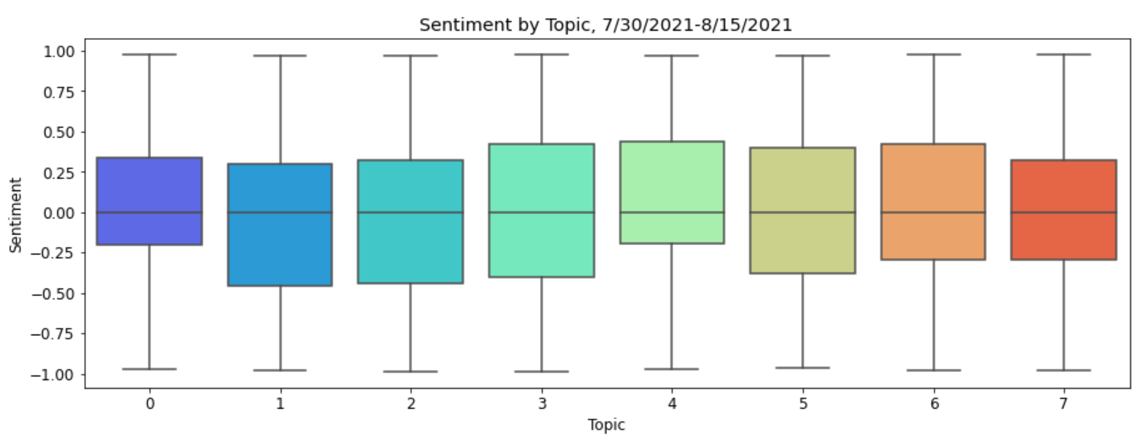




```
[(0,
  '0.051*"vaccine" + 0.014*"first" + 0.012*"vaccination" + 0.011*"today" + '
  '0.010*"coronavirus" + 0.007*"dose" + 0.007*"county" + 0.007*"state" + '
  '0.007*"shot" + 0.006*"plan"'),
 (1,
  '0.023*"health" + 0.013*"tested_positive" + 0.011*"public_health" + '
  '0.011*"fighting_stigma" + 0.010*"vaccine" + 0.009*"coronavirus" + '
  '0.008*"vaccine_rollout" + 0.007*"last_week" + 0.007*"case" + '
  '0.007*"tested"'),
 (2,
  '0.014*"like" + 0.012*"wearing_mask" + 0.011*"people" + 0.011*"pandemic" + '
  '0.010*"small_business" + 0.009*"know" + 0.009*"look_like" + '
  '0.008*"coronavirus" + 0.008*"look" + 0.007*"life"'),
 (3,
  '0.026*"death" + 0.021*"case" + 0.015*"next_week" + 0.014*"state" + '
  '0.013*"united_state" + 0.011*"week" + 0.009*"coronavirus" + 0.009*"covid" + '
  '0.009*"game" + 0.009*"case_death"'),
 (4,
  '0.017*"vaccine" + 0.016*"stay_home" + 0.013*"worker" + 0.012*"home" + '
  '0.009*"test" + 0.009*"site" + 0.009*"thank" + 0.008*"healthcare_worker" + '
  '0.007*"test_positive" + 0.007*"testing"'),
 (5,
  '0.016*"people" + 0.015*"trump" + 0.012*"year" + 0.009*"american" + '
  '0.008*"great" + 0.008*"biden" + 0.007*"right" + '
  '0.007*"trump_administration" + 0.006*"need" + 0.006*"many"'),
 (6,
  '0.025*"vaccine" + 0.019*"stay_safe" + 0.012*"please" + 0.012*"stay" + '
  '0.012*"safe" + 0.012*"friend" + 0.010*"good" + 0.009*"getting" + '
  '0.009*"appointment" + 0.008*"first_responder"'),
 (7,
  '0.023*"mask" + 0.023*"wear_mask" + 0.013*"wear" + 0.012*"would" + '
  '0.011*"people" + 0.010*"trump" + 0.010*"coronavirus" + 0.009*"make_sure" + '
  '0.009*"like" + 0.007*"time"')]
```

This time period was particularly interesting since there was greater variation in sentiment across topics. Topics 0, 5 and 7 were the most discussed, with topic 0, consisting of over 8,000 tweets being overwhelmingly positive. Keywords from this topic included *vaccine, vaccination, first_dose, today* and *vaccinated* showing the rising hope in the public due to the introduction of vaccines.  Topic 5 showed a relatively even spread in sentiment, with both Trump and Biden frequently mentioned in the tweets. Lastly, topic 7 leans towards the negative side with keywords like *kill, panic, pessimistic, make_sure* and *wear_mask.* Reasonably, we can assume that this topic represents people's growing frustration towards people who refused to wear masks and saw them as an infringement upon their rights.

**Delta Variant Spreads in North America** (Jul 30, 2021 - Aug 15, 2021)



```
[(0,
  '0.041*"mask" + 0.034*"school" + 0.023*"public_health" + 0.016*"wear_mask" + '
  '0.015*"mask_mandate" + 0.014*"child" + 0.011*"student" + 0.011*"public" + '
  '0.010*"health" + 0.010*"parent"'),
 (1,
  '0.032*"vaccinated" + 0.027*"people" + 0.023*"delta_variant" + '
  '0.021*"getting_vaccinated" + 0.020*"covid" + 0.018*"next_month" + '
  '0.013*"virus" + 0.013*"delta" + 0.013*"vaccine" + 0.013*"getting"'),
 (2,
  '0.020*"require_mask" + 0.016*"canada" + 0.010*"freedom" + 0.010*"sense" + '
  '0.009*"lie" + 0.008*"would" + 0.008*"coronavirus" + 0.008*"instead" + '
  '0.007*"sunday" + 0.007*"president"'),
 (3,
  '0.015*"taste_smell" + 0.014*"good" + 0.014*"would" + 0.012*"people" + '
  '0.012*"year" + 0.010*"give" + 0.010*"like" + 0.009*"little" + 0.009*"covid" '
  '+ 0.009*"think"'),
 (4,
  '0.027*"vaccine" + 0.025*"people" + 0.017*"family_friend" + '
  '0.016*"vaccinated" + 0.015*"fully_vaccinated" + 0.012*"friend" + '
  '0.012*"fully" + 0.011*"hospital" + 0.010*"shot" + 0.008*"family"'),
 (5,
  '0.014*"side" + 0.014*"stock" + 0.014*"twitter" + 0.013*"health_care" + '
  '0.013*"american" + 0.012*"democracy" + 0.012*"account" + 0.012*"republican" '
  '+ 0.012*"biden" + 0.011*"interest"'),
 (6,
  '0.039*"federal_provincial" + 0.031*"vaccine_passport" + '
  '0.026*"necessary_canada" + 0.026*"stop_sign" + 0.026*"petition_cdnchange" + '
  '0.022*"vaccine" + 0.022*"sign" + 0.018*"fully_vaxxed" + 0.018*"stop" + '
  '0.015*"passport"'),
 (7,
  '0.034*"case" + 0.019*"past_week" + 0.019*"hospital_bed" + '
  '0.019*"delta_variant" + 0.016*"coronavirus" + 0.014*"week" + 0.014*"death" '
  '+ 0.012*"screw" + 0.011*"bed" + 0.011*"unvaccinated"')]
```

The countplot shows that the most widely discussed topic was topic 0 with the main areas of discussion being *public health, mask mandate, public* and *student.* Over a year into the pandemic with a newer deadlier strain, we can see that desperation for normalcy has increased among the public with more people discussing the need for mask mandates. Topic 3 was almost as popular as topic 0 and seems to revolve around symptoms of the virus with keywords like *taste smell* and *covid.* Lastly, topic 4 with a largely positive sentiment shows that the communities of Twitter users were getting vaccinated as shown by words like *family, family friend, friend, vaccinated* and *fully vaccinated,* possibly due to people realizing the necessity for vaccinations with widespread infections because of the Delta variant during the time. Other keywords from this time period include *death, delta variant, hospital bed,* and *getting vaccinated,* representing the damage to lives and public health.

## 6. Discussion

To answer our first research question, we closely examined the output from our LDA model. We found that topics over the three time periods did not vary much. Throughout the course of the pandemic, the main recurring topics were those regarding the government (with several keywords referencing the Trump administration and other politicians like Joe Biden), several topics concerning masks and social distancing with keywords such as *mask mandates, wear masks, stay home, masks,* and *stay safe,* topics regarding healthcare such as *healthcare worker, testing, positive tests,* and *hospital bed.* Unsurprisingly, the introduction of vaccines in late December 2020 lead to them becoming a widely discussed topic with keywords such as *vaccine, dose,* and *vaccination* occurring in several topics in the last two time periods. Furthermore, keywords like *fighting stigma, friend, family,* and *fully vaccinated* showed a promising view of the future and people's outlook regarding vaccines. *Delta variant, death, hospital bed* and *case* became more popular with the spread of the deadly variant. To paint a more complete picture of our results and answer our next question, we further analyzed these topics along with their sentiments and how they varied over days during our chosen time periods.

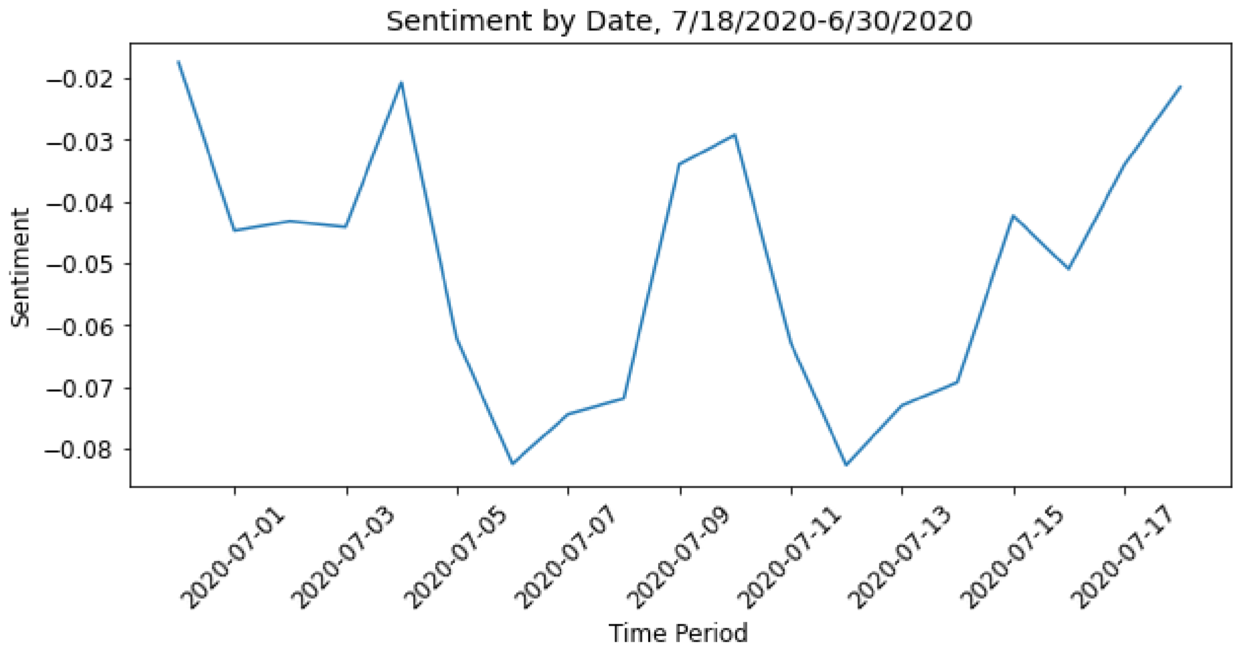



*Figure 1: Average user sentiment from 30th June to 18th July 2020*

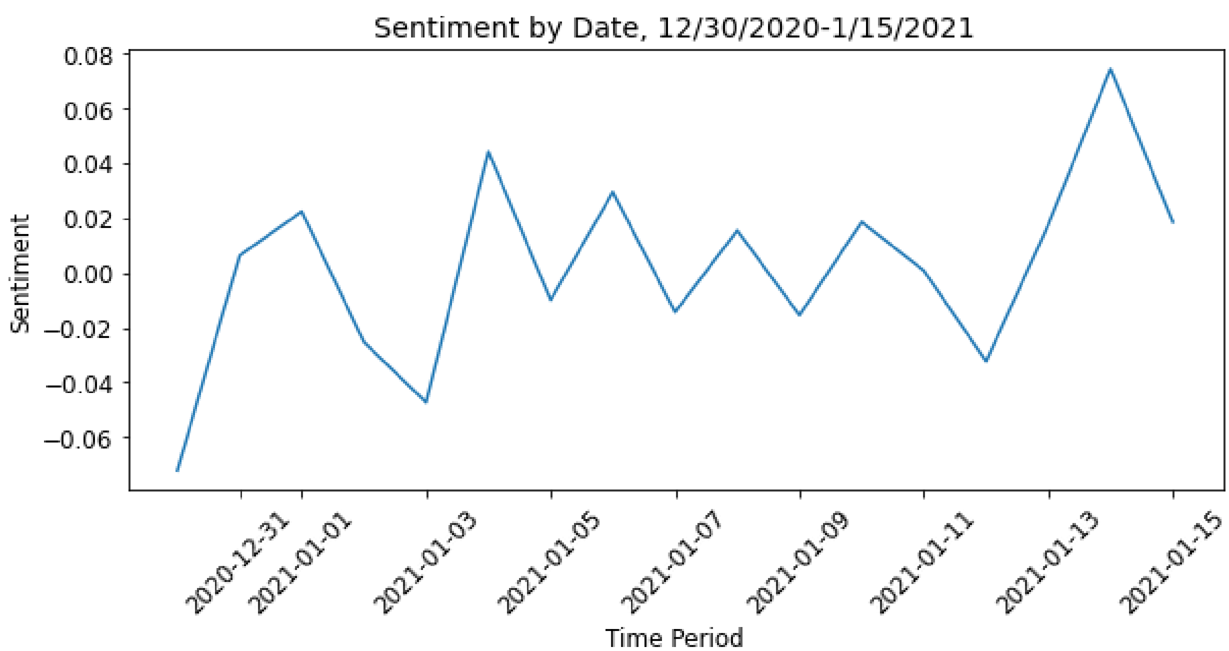



*Figure 2: Average user sentiment from 30th December 2020 to 15th January 2021*

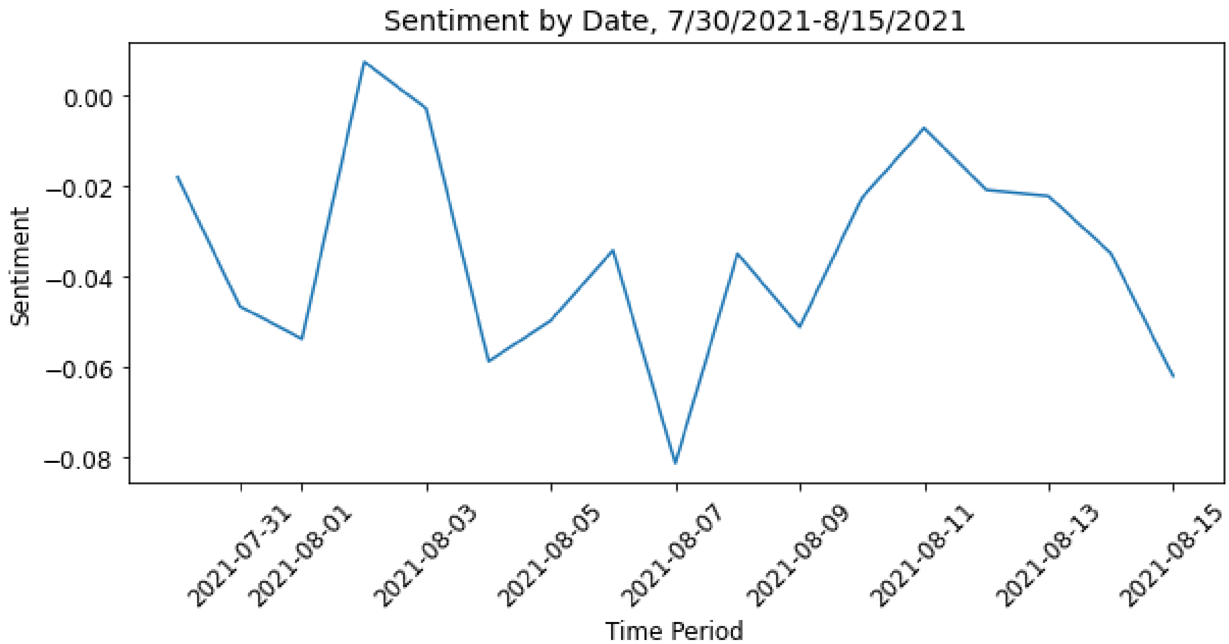



*Figure 3: Average user sentiment from 30th July to 15th August 2021*

Our second research question was to investigate the sentiment in the North American region as the pandemic spread. Our hypotheses were that during the waves of the pandemic, especially when the Delta variant hit the US, the sentiment would be negative but progressively improve as people would begin to adapt and solutions emerge. Moreover, we believed the sentiment would improve much more as vaccines began to roll out in the beginning of 2021. Our results indicate the same. As the world came to grips with quarantines and lockdowns in the first major wave of the pandemic, the people were much more afraid as we can see in Figure 1 above. The daily average sentiment of each day in that timeframe is in the negative range; no positive sentiment whatsoever. Fast forward to late 2020/early 2021, the vaccine distribution campaign in North America commences. If you observe in Figure 2, then you can see that even though daily sentiment is fluctuating, it is moving in an upward trend. While some would still be hesitant about the vaccine's efficacy, the general public sentiment grew more positive since it was the first glimmer of hope in those trying times. Subsequently, during the spread of the delta variant in August of 2021, we see that the sentiment then begins to wane and falls below into the negative category again. Looking at Figure 1 and 3, we can also notice that the sentiment during the delta variant was slightly more positive than the first wave in the region.

## 7. Conclusion

In this paper, we used COVID-19 tweets to study public sentiment and topics of interest for the population in North American countries over the course of the pandemic. We found that overall, the public was concerned about safety with several topics encouraging others to wear masks and social distance. Also, contradictory to initial fearmongering and conspiracy theories regarding vaccines, the overall sentiment regarding vaccines was positive and pushed others to get vaccinated, get their second dose, and fight stigma. Sentiment also improved over January 2021 as the worldwide vaccination campaign took off. On the other hand, we saw how public sentiment declined during the spread of the Delta variant, with Twitter acting as a digital archive of the deaths and suffering faced by the population over the course of the pandemic with keywords like *death, panic, hospital bed, case,* and *delta variant.* We believe this study and others like it provide valuable insight for authorities to understand the population's concerns, opinions, and sentiments during such uncertain times.

## 8. Limitations and Future Study

Our study investigated tweets from the North American continent. The tweets extracted were not balanced in terms of number of tweets for each country. A significant majority of our tweets were from the United States, followed by Canada, and a very small number of tweets from Mexico. This can be attributed to the population differences across the countries and to the fact that other languages like French and Spanish were left out which are spoken in Canada and Mexico. Moreover, we did not filter out tweets that were posted by bots. This could potentially have an impact on the sentiments of your tweets. Lastly, since we did not possess substantial computing power, we had to limit our time horizon to

15 days. An interesting future study could look at sentiment analysis and topic modelling for regions with languages like Spanish and French. Additionally, having a balanced amount of tweets in the region of study would create more compelling results too.

## 9. References


Behl, Shivam, et al. “Twitter for Disaster Relief through Sentiment Analysis for COVID-19 and Natural Hazard Crises.” *International Journal of Disaster Risk Reduction*, 2021.

Budiharto, Widodo, and Meiliana Meiliana. “Prediction and Analysis of Indonesia Presidential Election from Twitter Using Sentiment Analysis.” *Journal of Big Data*, vol. 5, no. 1, 2018. doi:10.1186/s40537-018-0164-1.

Banda, Juan M., et al. “A Large-Scale COVID-19 Twitter Chatter Dataset for Open Scientific Research—An International Collaboration.” *Epidemiologia*, vol. 2, no. 3, 2021, pp. 315–24, doi:10.3390/epidemiologia2030024.

Chakraborty, Koyel, et al. “Sentiment Analysis of COVID-19 Tweets by Deep Learning Classifiers—A Study to Show How Popularity Is Affecting Accuracy in Social Media.” *Applied Soft Computing*, 2020.

Costola, Michele, et al. “Machine Learning Sentiment Analysis, Covid-19 News and Stock Market Reactions.” *Publikationen Von Forscherinnen Und Forschern Des Leibniz-Instituts Für Finanzmarktforschung SAFE*, 2020.

Derakhshan, Ali, and Hamid Beigy. “Sentiment Analysis on Stock Social Media for Stock Price Movement Prediction.” *Engineering Applications of Artificial Intelligence*, vol. 85, Oct. 2019, pp. 569–578.

Dubey, Akash Dutt. “Twitter Sentiment Analysis during COVID19 Outbreak.” *SSRN Electronic Journal*, 2020. doi:10.2139/ssrn.3572023.

Garcia, Klaifer, and Lilian Berton. “Topic Detection and Sentiment Analysis in Twitter Content Related to COVID-19 from Brazil and the USA.” *Applied Soft Computing*, vol. 101, 2021, p. 107057. doi:10.1016/j.asoc.2020.107057.

Hussain, Amir, et al. “Artificial Intelligence–Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study.” *Journal of Medical Internet Research*, vol. 23, no. 4, 2021, p. e26627. doi:10.2196/26627.

Hutto, C.J and Eric Gilbert. " VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text." Eighth International Conference on Weblogs and Social Media (ICWSM-14), Ann Arbor, MI, June 2014 https://github.com/cjhutto/vaderSentiment

Jamison, Amelia M., et al. “Not Just Conspiracy Theories: Vaccine Opponents and Proponents Add to the COVID-19 ‘Infodemic’ on Twitter.” *Harvard Kennedy School Misinformation Review*, 2020. doi:10.37016/mr-2020-38.

Karalevicius, Vytautas, et al. “Using Sentiment Analysis to Predict Interday Bitcoin Price Movements.” *The Journal of Risk Finance*, vol. 19, no. 1, 2018, pp. 56–75. doi:10.1108/jrf-06-2017-0092.

Marcec, Robert, and Robert Likic. “Using Twitter for Sentiment Analysis towards AstraZeneca/Oxford, Pfizer/BioNTech and Moderna COVID-19 Vaccines.” 2021.

Mittal, Anshul, and Arpit Goel. “Stock Prediction Using Twitter Sentiment Analysis.” *Semantic Scholar*, 2011, https://www.semanticscholar.org/paper/Stock-Prediction-Using-Twitter-Sentiment-Analysis-Mittal/4ecc55e1c3ff1cee41f21e5b0a3b22c58d04c9d6. Accessed 2 Dec. 2021.

Monselise, Michal, and Chia-Hsuan Chang. “Topics and Sentiments of Public Concerns Regarding COVID-19 Vaccines: Social Media Trend Analysis.” *Journal of Medical Internet Research*, 2021, https://www.jmir.org/2021/10/e30765/. Accessed 4 Dec. 2021.

Nguyen, Thien Hai, and Kiyoaki Shirai. “Topic Modeling Based Sentiment Analysis on Social Media for Stock Market Prediction.” 2015, https://www.researchgate.net/publication/280609514_Topic_Modeling_based_Sentiment_Analysis_on_Social_Media_for_Stock_Market_Prediction. Accessed 2 Dec. 2021.

Nguyen, Thu T, et al. “Exploring U.S. Shifts in Anti-Asian Sentiment with the Emergence of COVID-19.” *Big Data for Public Health Research and Practice*, 2020.

Pandey, Parul. “Simplifying Sentiment Analysis Using VADER in Python (on Social Media Text).” *Medium*, 11 Sept. 2020, medium.com/analytics-vidhya/simplifying-social-media-sentiment-analysis-using-vader-in-python-f9e6ec6fc52f.

Pokharel, Bishwo Prakash. “Twitter Sentiment Analysis During Covid-19 Outbreak in Nepal.” *SSRN Electronic Journal*, 2020. doi:10.2139/ssrn.3624719.

Ragini, J. Rexiline, et al. “Big Data Analytics for Disaster Response and Recovery through Sentiment Analysis.” *International Journal of Information Management*, vol. 42, 2018, pp. 13–24.

Rehurek, Radim & Petr Sojka. “Software Framework for Topic Modelling with Large Corpora”. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks,* ELRA, 22 May, 2010, 45-50. 10.13140/2.1.2393.1847.

Safi, Michael. “India’s Shocking Surge in Covid Cases Follows Baffling Decline.” *The Guardian*, 21 Apr. 2021, www.theguardian.com/world/2021/apr/21/india-shocking-surge-in-covid-cases-follows-baffling-decline.

Shahi, Gautam Kishore, et al. “An Exploratory Study of COVID-19 Misinformation on Twitter.” *Online Social Networks and Media*, vol. 22, 2021, p. 100104. doi:10.1016/j.osnem.2020.100104.

Yaqub, Ussama, Soon Ae Chun, et al. “Analysis of Political Discourse on Twitter in the Context of the 2016 US Presidential Elections.” *Government Information Quarterly*, vol. 34, no. 4, 2017, pp. 613–26. *Crossref*, doi:10.1016/j.giq.2017.11.001.

Yaqub, Ussama, Nitesh Sharma, et al. “Location-Based Sentiment Analyses and Visualization of Twitter Election Data.” *Digital Government: Research and Practice*, vol. 1, no. 2, 2020, pp. 1–19. doi:10.1145/3339909.

Yousefinaghani, Samira, et al. “An Analysis of COVID-19 Vaccine Sentiments and Opinions on Twitter.” *International Journal of Infectious Diseases*, vol. 108, 2021, pp. 256–62. doi:10.1016/j.ijid.2021.05.059.