

The hidden effects of the Covid vaccine in the real world

Programming challenge and mini research project

Arefeh Yavary

Table of Content

- Main Idea and Approach
- Research Questions
- Data Collection Process
- Statistics and Exploration of Data
- Methods for Analyzing the Dataset
- Auxiliary Dataset
- Main Results and Findings
- Comparisons with baselines
- Future Research Directions
- Limitations and Issues
- References

Main Idea and Approaches

There are several effects of Covid and so Covid Vaccine on the living of people all over the world. We want to using social media analyses methods in order to study the hidden effects of covid vaccine.

The main ideas and approach to this study the hidden effects of Covid Vaccine are a follows:

- Find the main topic of discussion in about Covid, Covid Vaccine and Anti-Vaccine subject of speak.
- Figure out the problems of people because of Covid using topic modeling in tweets.
- Location based issues about Covid, Covid Vaccine and Anti-Vaccine subject of speak.
- Language based issues about Covid, Covid Vaccine and Anti-Vaccine subject of speak.
- Study different Aspects of Covid, Covid Vaccine and Anti-Vaccine during time. (If people be negative about vaccine the government will face to a problem.)

Main Idea and Approaches

- Sentiment about vaccinate/anti-vaccine through time, location and language.
- Study depression because of Covid
- Correlation with statistics (Deaths, Cases, Recovered) of Coronavirus
- Fake news in Covid. (Because I worked on information verification in my MSc thesis, I think that It was repetitive to I work on misinformation on Covid.)
- Detect the fake news and rumor that are impressive on people during Covid and preventing them from spreading.
- Study the reason of the grew or fell of different business types during Covid.
- The result of news on the the sentiment of people about vaccine or anti vaccine. Study the things that are effectual on the sentiment of people about Covid.

Main Idea and Approaches (cont.)

- Predict feedback and sentiment of people about Covid vaccine.
- Predict statistics (Deaths, Cases, Recovered) of Coronavirus Disease using people sentiment on Covid vaccine.
- The effect of sentiment of people about Covid vaccine on different financial markets.
- Predict financial markets using people sentiment on Covid vaccine.
- The Correlation of Covid and Vaccine Sentiment with Financial markets.
- Detection of False statistics (Deaths, Cases, Recovered) of Coronavirus Disease in News Media using Social Media Analysis. (Seems that Location and Language base approach is more helpful). This is helpful because of the fact that some news agencies state that some of the statistics about Covid are fake.
- Detect when the news agencies report fake statistics about Coronavirus Disease.

Research Questions

These are the Research Questions which are depending to the main ideas and approaches of this mini-project.

- What are the main topic of discussion on social media about Covid, Covid Vaccine and Anti-Vaccine?
- What are the main problems of people during Covid?
- What are the location based issues about Covid, Covid Vaccine and Anti-Vaccine subject of speak?
- What are the language based issues about Covid, Covid Vaccine and Anti-Vaccine subject of speak?
- How different aspects of Covid, Covid Vaccine and Anti-Vaccine are changing during time?

Research Questions

- What are the sentiment about vaccinate/anti-vaccine through time, location and language?
- Does depression is decreased or increased during the Covid pendemic?
- Is the any correlation with statistics (Deaths, Cases, Recovered) of Coronavirus?
- What are the fake news about Covid?
- How and which fake news are effectual on people during Covid? How to prevent them from spreading?
- Why some businesses are felling down during Covid and why some of them business are growing? How to help felling down business to growing up in Covid situation?
- How news are effectual on the sentiment of people about vaccine? What are the things that are impressive on the centime of people about vaccine?

Research Questions

- Can we predict the feedback of the people about Covid vaccine?
- Can we predict statistics (Deaths, Cases, Recovered) of Coronavirus Disease using people sentiment on Covid vaccine?
- What are the effect of sentiment of people about Covid vaccine on different financial markets?
- Can we predict financial markets using people sentiment on Covid vaccine? Or whether sentiment of the people about Covid is useful for the prediction of financial markets during Covid?
- Is there any Correlation of Covid and Vaccine Sentiment with Financial markets?
- Can we detect the statistics (Deaths, Cases, Recovered) of Coronavirus Disease in News Media using Social Media Analysis?
- Are the predictions of Social Media Analysis in Coronavirus disease cases are matched with real statistics? When the reported statistics in news agencies are

Data-Collection Process

- We collected data using Tweepy API.
- I got access to the key from Tweeter Free Api before about 3 years ago with some limitation, I do not know may be the newer access for other country would be better but as I have the concern that my key for api will be invalid, I did not send any new request for APi.
- Reddit and Reddit API for getting data is inaccessible completely and I do not know any way to can access to it.
- I collected the tweets using Keywords and Date parameters.
- The Key words was: #Covid, #Vaccine, #antiVax
- The Selection of these keywords are because the covid, vaccine and anti-vaccine topics are the depending topic and hashtags for our studying.

Data-Collection Process

- These hashtags could be studied and selected using several methods like as finding the number of occurrence of the hashtags each day. But because of the limitation of time and sources I selected the keywords for collecting data using the most trivial and probable keywords.
- I am interested to collecting data with other keywords and hashtags that are collated in tweets with the above hashtags, but because of the limitations in data collection using the available api I could not add any other new keywords.
- The start of the day for collection data is: Dec 1 2020 (This time is near to the of anniversary of Covid up to what news agencies stated.)
- The start of the day for collection data is: Jan 10 2021
- I selected this long time in order to can find long time effects in data.

Data-Collection Process

- We collected every day between the start and end date. Each day we collected 200 tweets. I am interested and it would be very effectual if I add more tweets, but it was not possible because of limitation in time and data collection sources.
- There are about 10 days that some data are missing and are less than 200 tweet pre day. I recollect some of them again. But because of time, some of them missed again.
- The data for each keyword are collected separately.
- Besides by the fact that Covid get harder and so there is more discussion about it in about 6 months ago, I decided to collect data of #Covid keyword for about 2500 tweet pre day in the previous last 6 months.
- I collected the tweets using “#Fake AND #Covid” query for finding the misinformation about Covid in order to filter fake information about Covid, before processing of tweets for other purposes.
- The tweets that are collected using “#Fake AND #Covid” query are gathered exactly in the same start and date time of the previous queries.
- We collected 20 tweets per day with “#Fake AND #Covid” query. This number is set because of the limitation in data collecting processing that I mentioned before.
- Data Collected in English Language

Data-Collection Process

- The features that each collected tweets have, are:
 - Text
 - Created date
 - Location
 - Number of followers
 - Number of friend
 - Number of following
 - Followers (That is useful for graph-mining approaches)

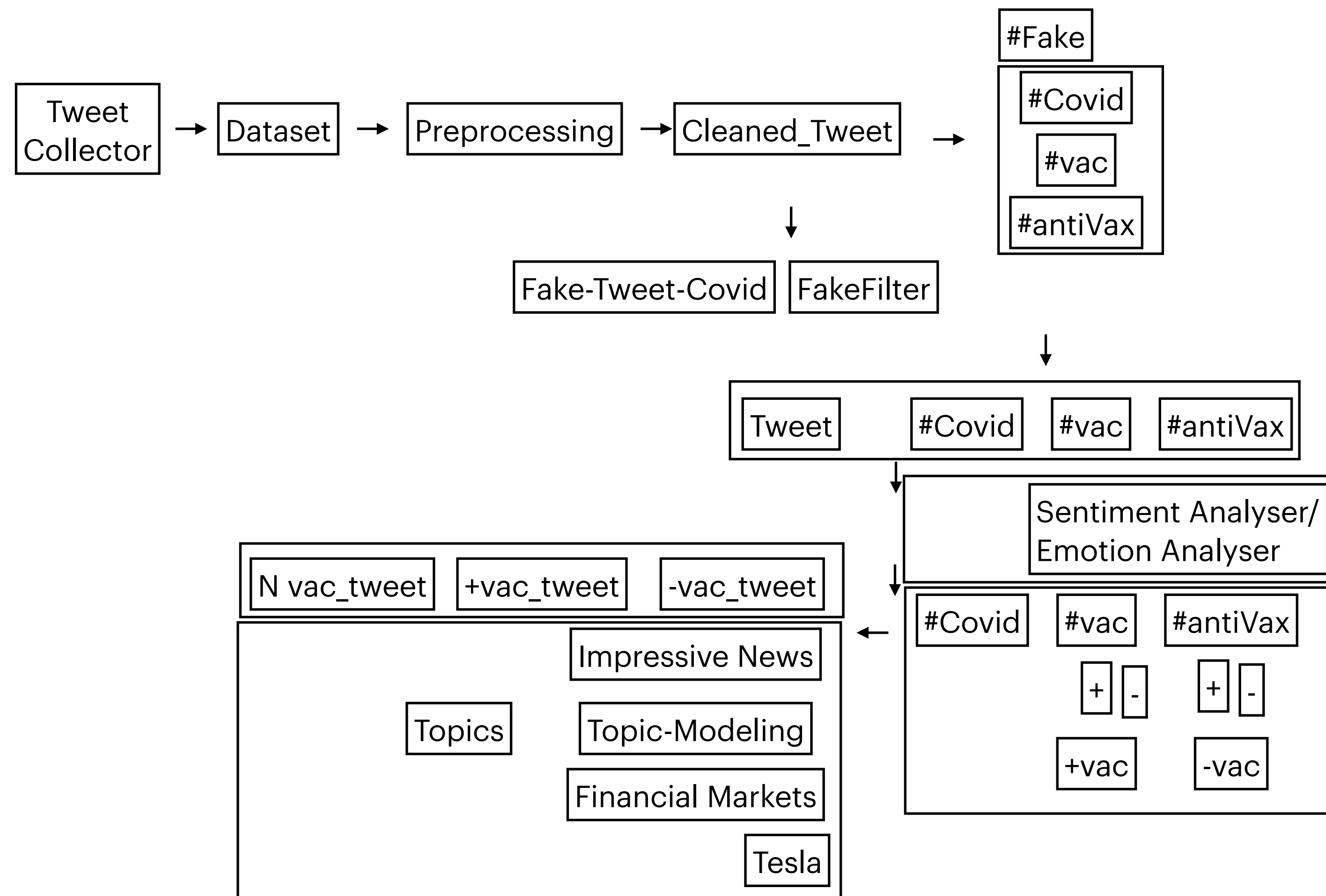
Statistics and Exploration of Collected Data

- Location in different queries.
- Number of Retweets different queries.
- Start Date: 10 - 01 - 2020
- End Date: 10 - 01 - 2021
- About 1000 tweet per day in the queries with :
 - “anti-vax”, “vaccine”, “covid”
- About 20 tweet per day in the queries:
 - for #fake AND #Covid.

Methods for Analyzing the Dataset

- In order to answer the questions to that mentioned in this research, I must first find the ratio of the person who are in agree with vaccinate and the person who are not agree with sentiment.
- In the next slide I illustrate the structure of the methods what are used for this purpose.
- Each of the sub-methods are describes in the following slides.

The Architecture of our Method



Tweet Collector

- Tweet Collector is explained in the Data-Collection Process section.
- I would state the data is collected in this research is a static one, which would be a dynamic and streaming process in the real and online applications.
- The online process of the proposed method would have some small changes.

PreProcessing Tweets

- We used the library with some changes.
- Also I code some other things that not considered in this library.
- This is an important section that is effectual in preprocessing Tweets.

FakeTweet Filter

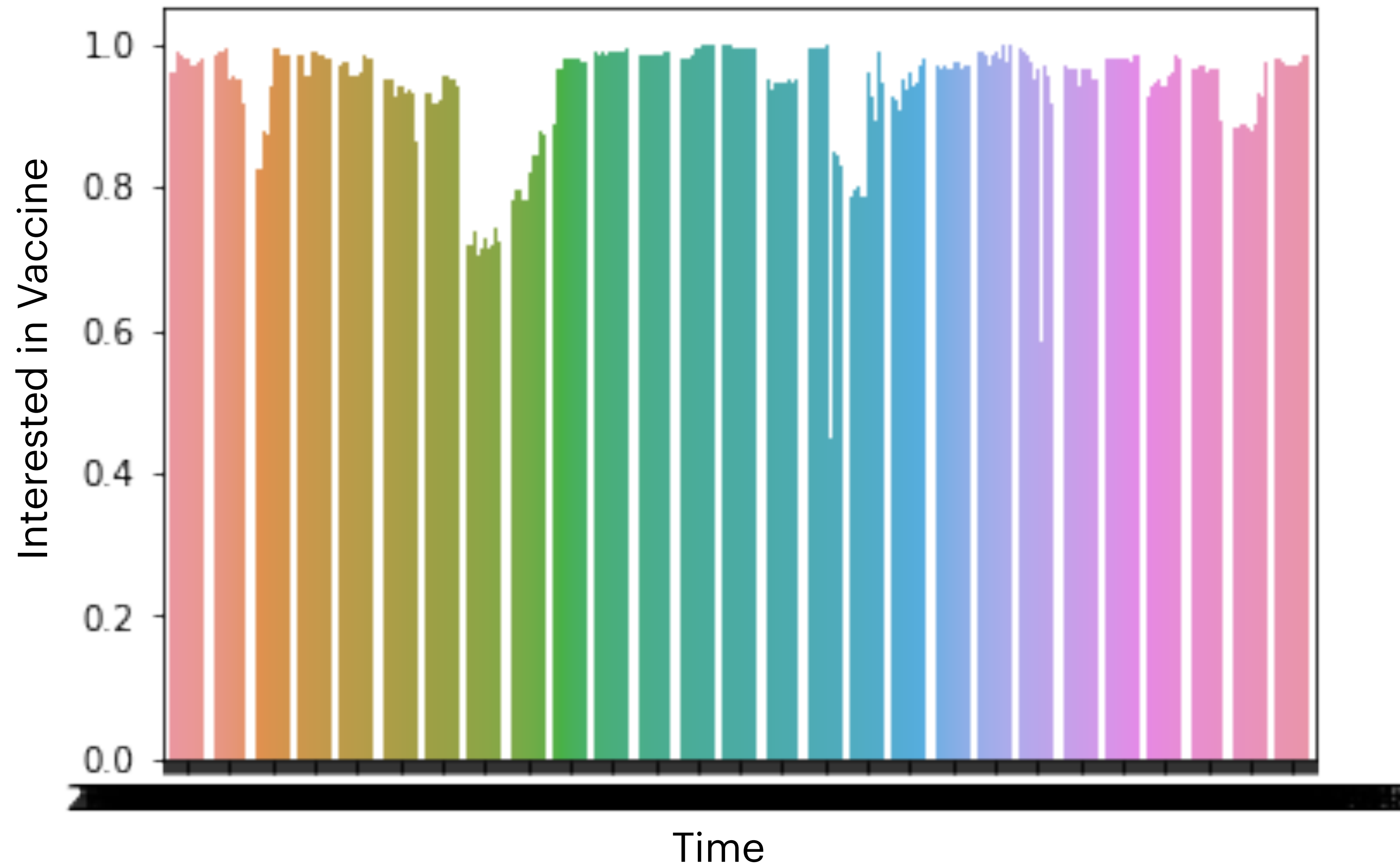
- This method using the tweets that has the hashtags of #fake and #covid.
- Then the #covid, #anti-vax and #vaccine tweets are filtered using the FakeTweet Filter.
- The FakeTweet Filter is working at follows:
- We find the most similar tweets to FakeTweet Filter.
- The similarity measure is the wordvetor of tweets.
- Then we omit them from the dataset.
- There are several other methods for finding fake tweets and also similar tweets to fake tweets. But I selected the most fast ones because of the lake of time and resources.
- This module is essential because there are a lot of rumor about the Covid.

Covid and Depression

- Emotion analysis methods on each covid tweet on each day.
- Also there are a lot of external news like as US election, Christmas or Easter which can effect the overall emotion of people. So if we want to study the Covid and depression we must evaluate it with the data of one year before.
- But as we do not have enough time and sources for collecting data, We just studied whether the depression has correlation with Covid statistics, vaccine sentiments. Moreover we studied how the depression is changes over time.
- The depression is analyzed using the similarity of tweet_wordvec with the words that have different emotion.

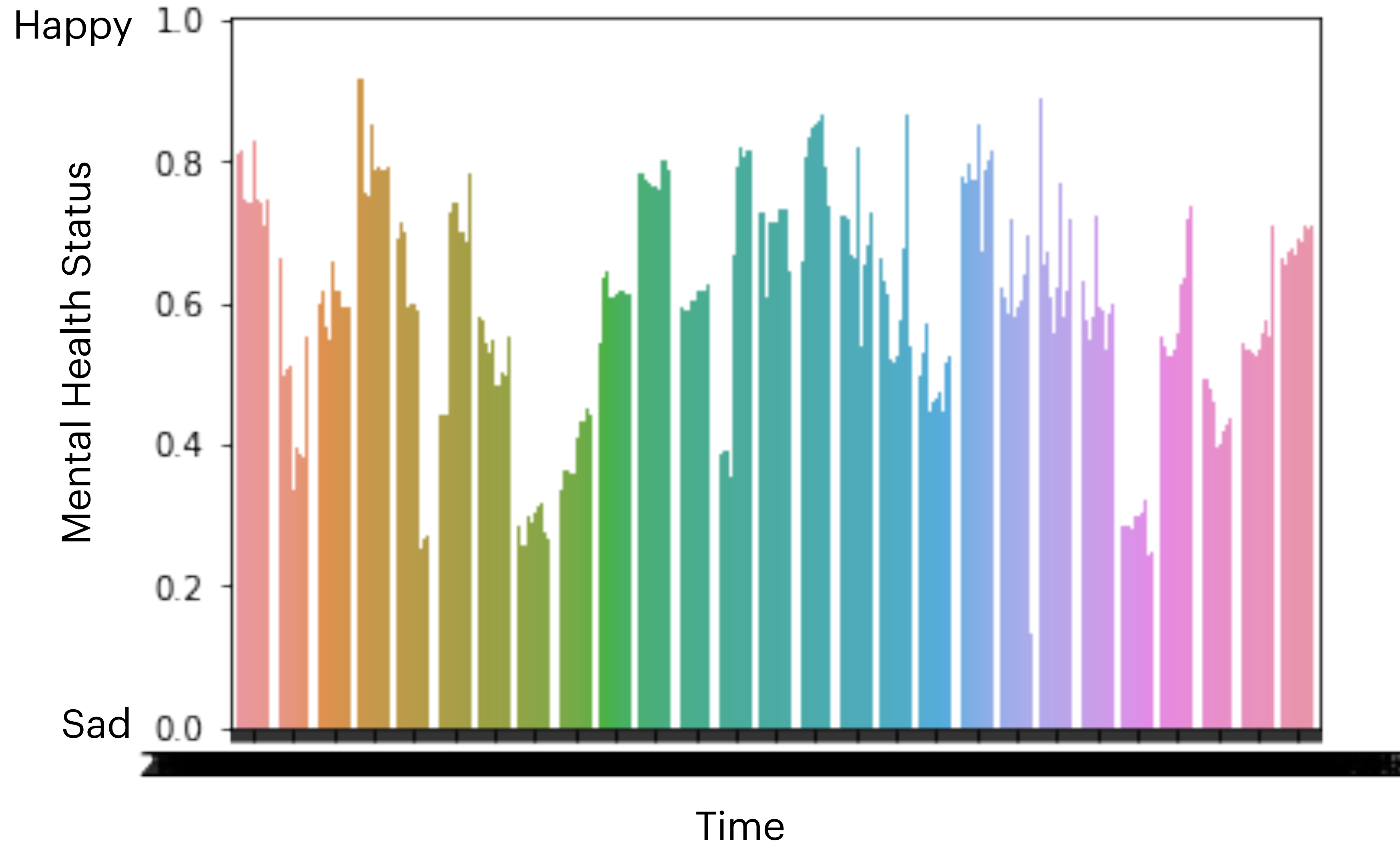
Sentiment Analyser

The feedback of people to Covid vaccine in Twitter.



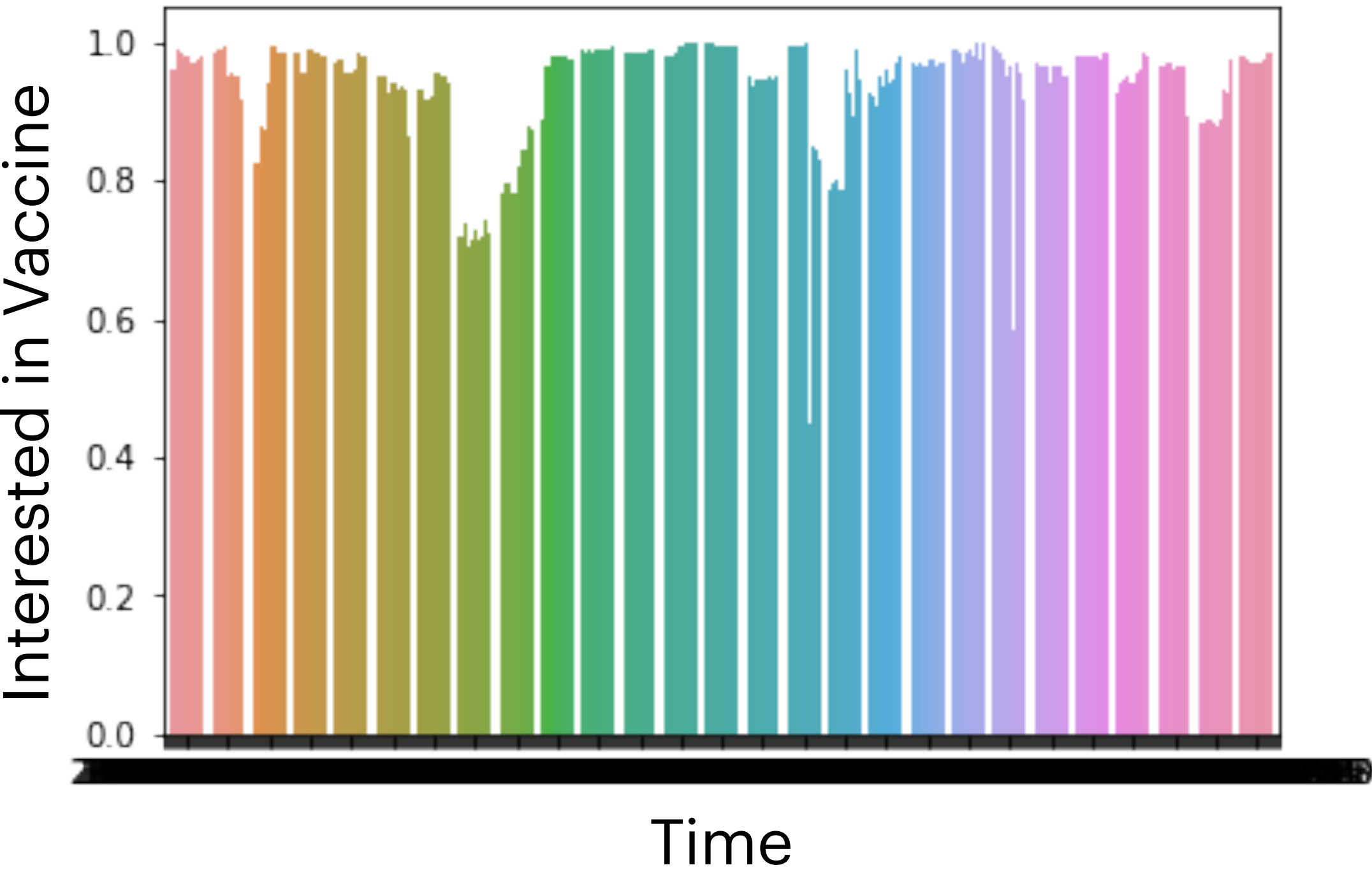
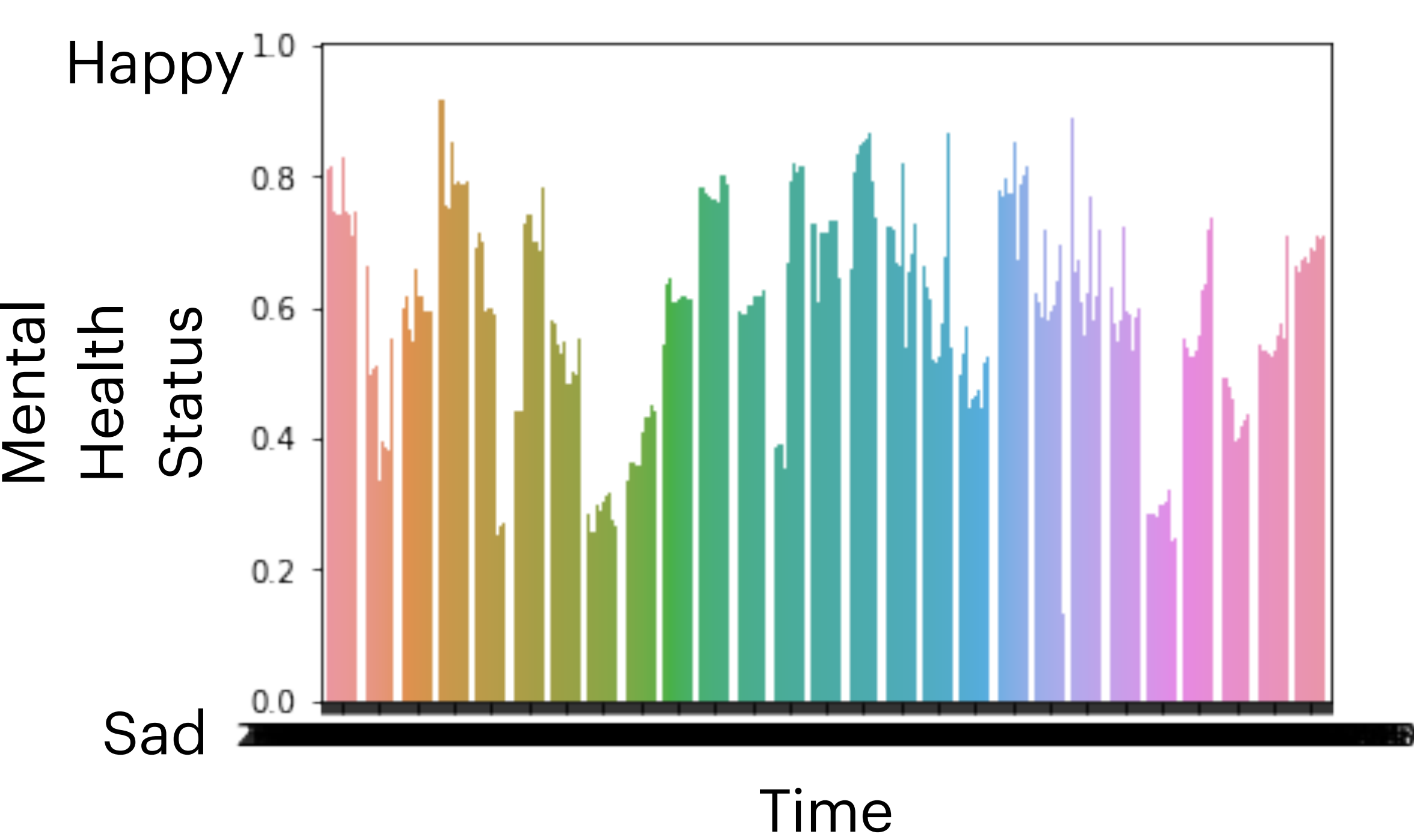
Mental Health

The feedback of people to Covid vaccine in Twitter.



Mental Health and Correlation with Vaccine Feedback

Mental Health
Correlation: 0.370

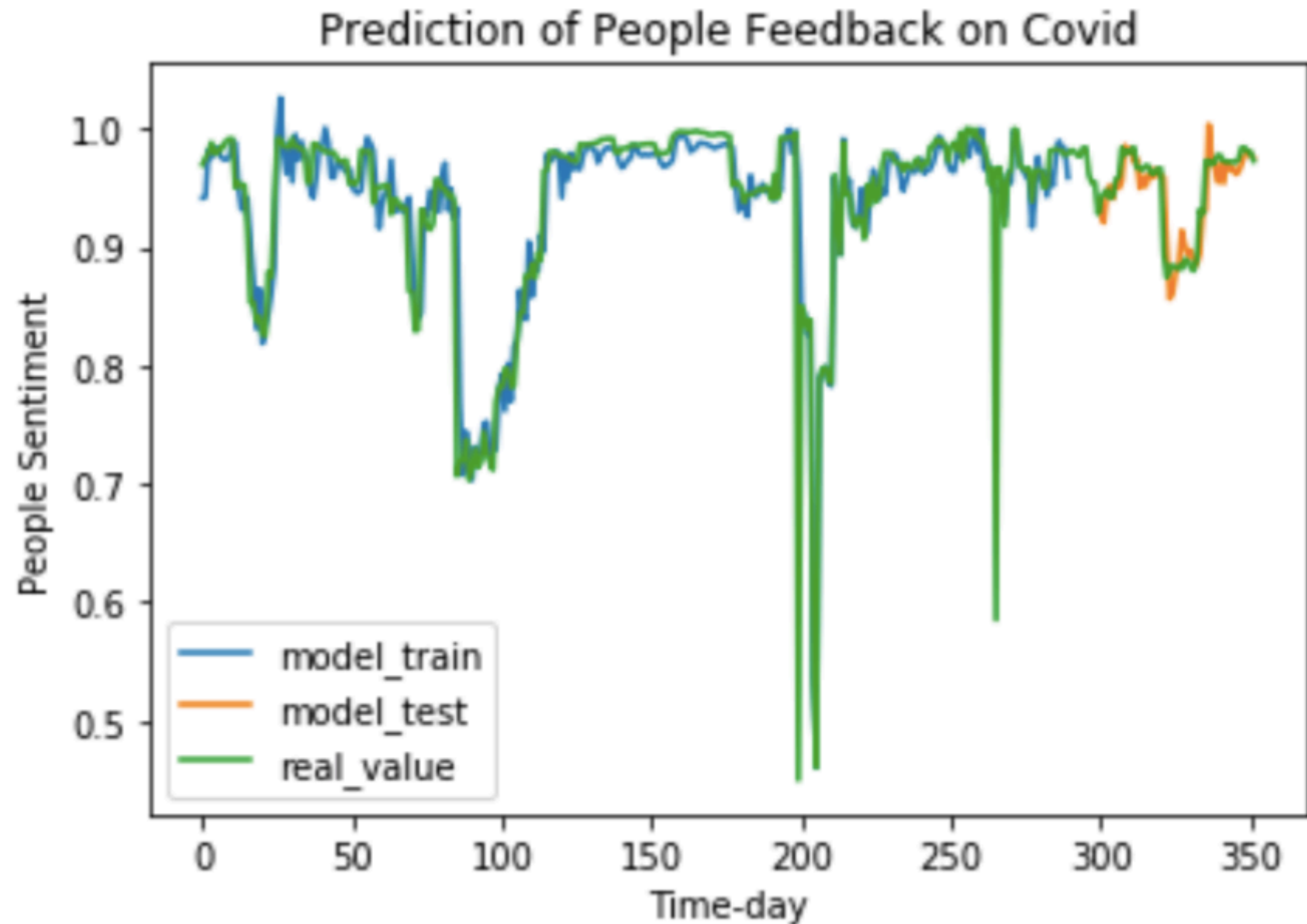


Sentiment Prediction

- Using Sequence Modeling with Multi-nominal Regression

r_{sq} , train: 76%

r_{sq} , test: 77%



Impressive News

2020-04-08 - Local Minima in Vaccine Sentiment

(#BBC OR #CNN) AND “Covid”

- The news correspond to local minima
- the trump administration now pimping media to claim they never had the stock pile of covid 19
- recent news has hit the media that infers some leniency in the travel conditions of temporary visa holders currently
- update north carolina is reporting new cases of coronavirus in the last hours.
- the united states department of state says researchers in wuhan lab were experimenting with a virus genetically

Impressive News

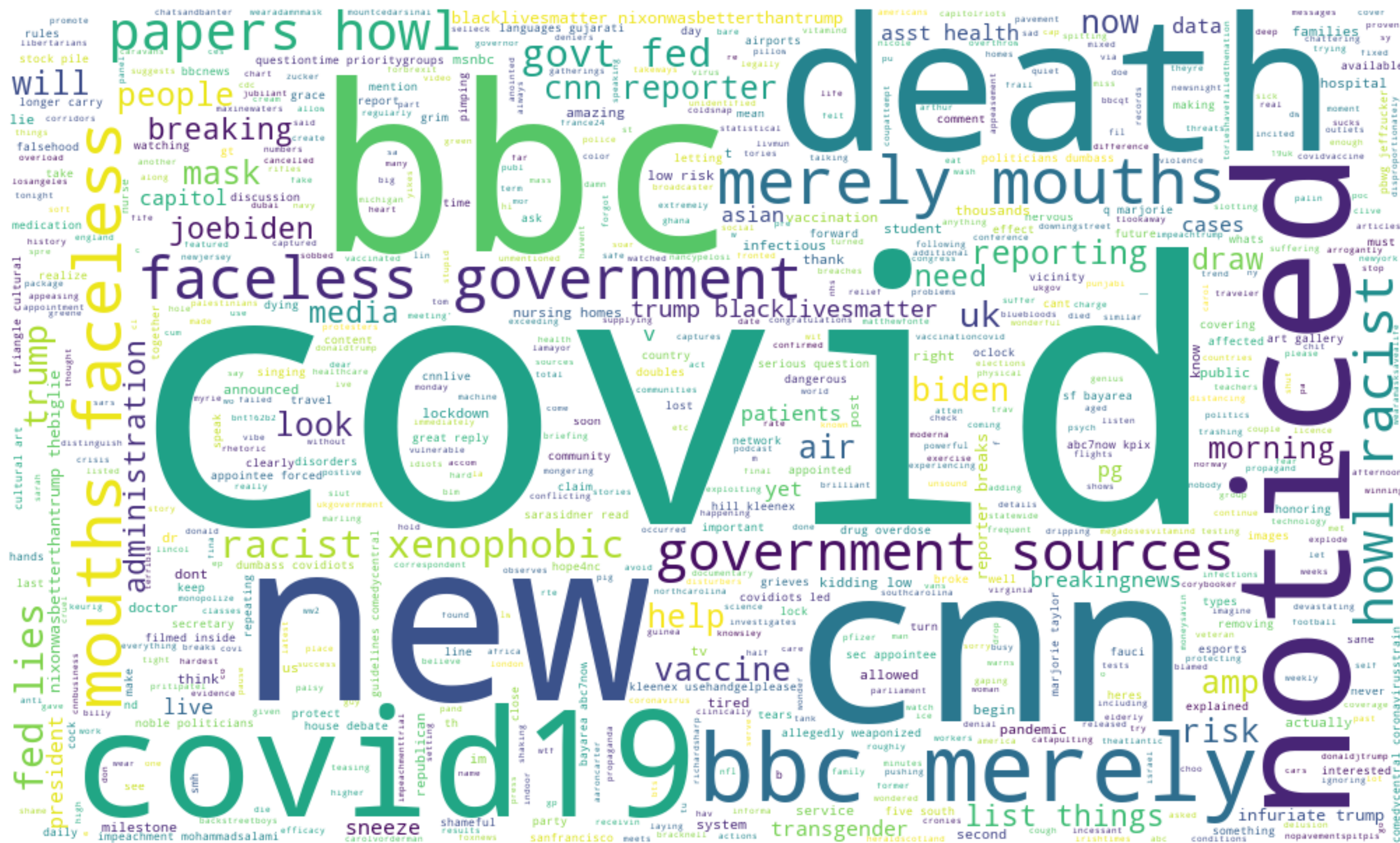
2020-02-24 Local Minima in Mental Health

(#BBC OR #CNN) AND “Covid”

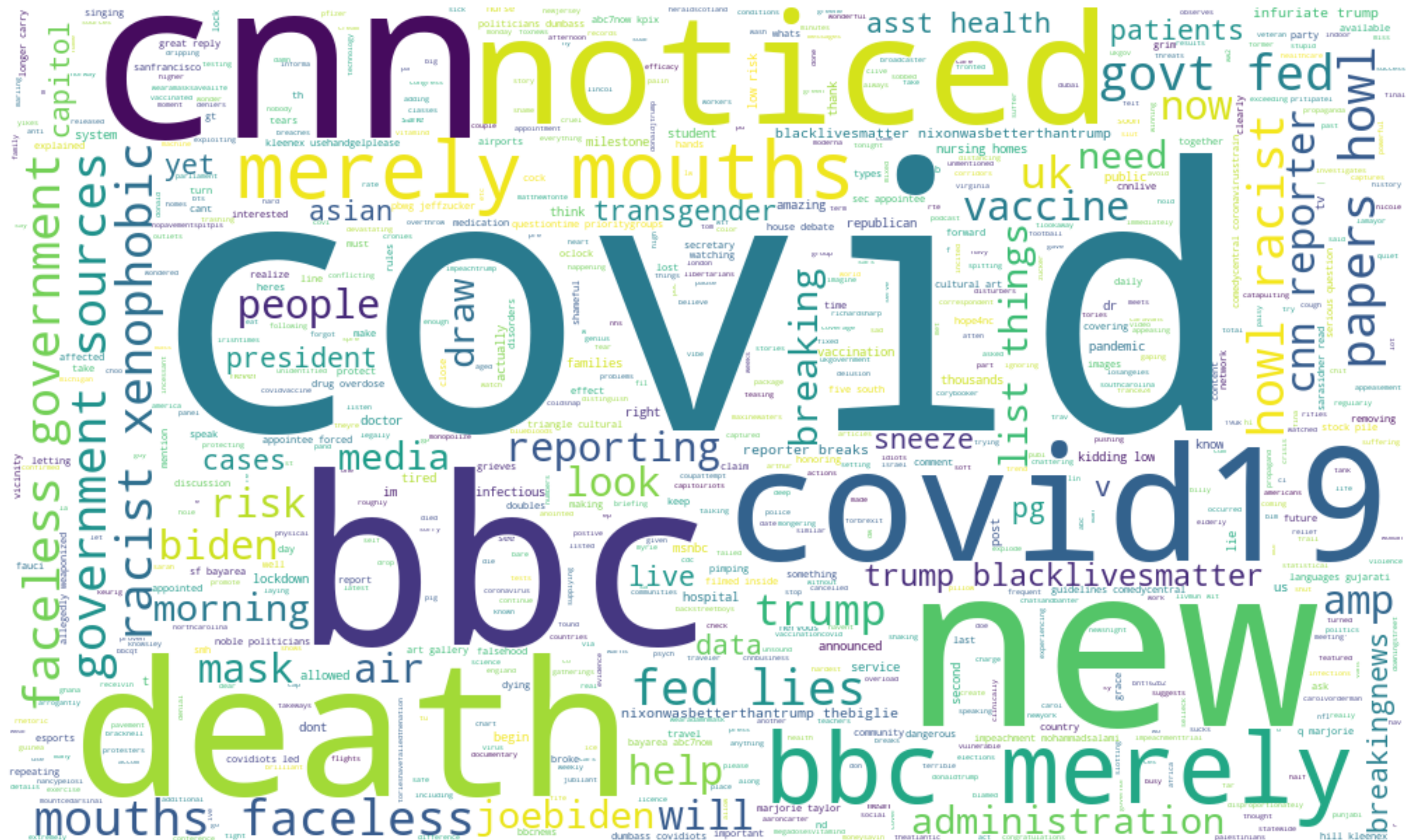
- The news correspond to local minima
- highest death rate today. very sad. teachers need vaccines. education news covid vaccine wednesday
- update north carolina is reporting new cases of coronavirus and new deaths in the last hours.
- cnn reports this morning that half of all the uk's covid infections have occurred in the past weeks.
- pbwg jeffzucker announced that airports will no longer carry cnn.

Topic Modeling

2020-02-24 Local Minima in Mental Health



2020-04-08 - Local Minima in Vaccine Sentiment



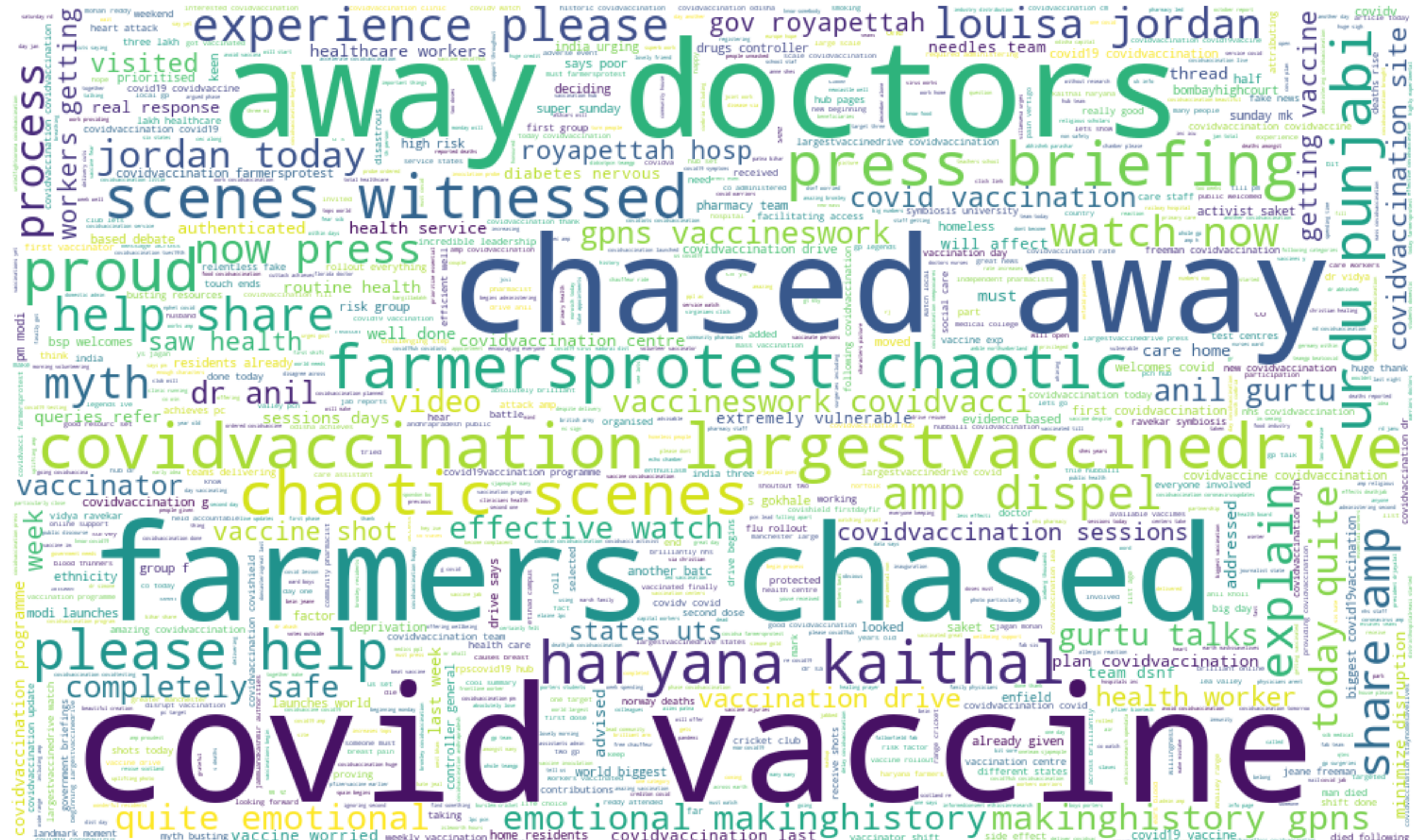
Topic Modeling

#Covid



Topic Modeling

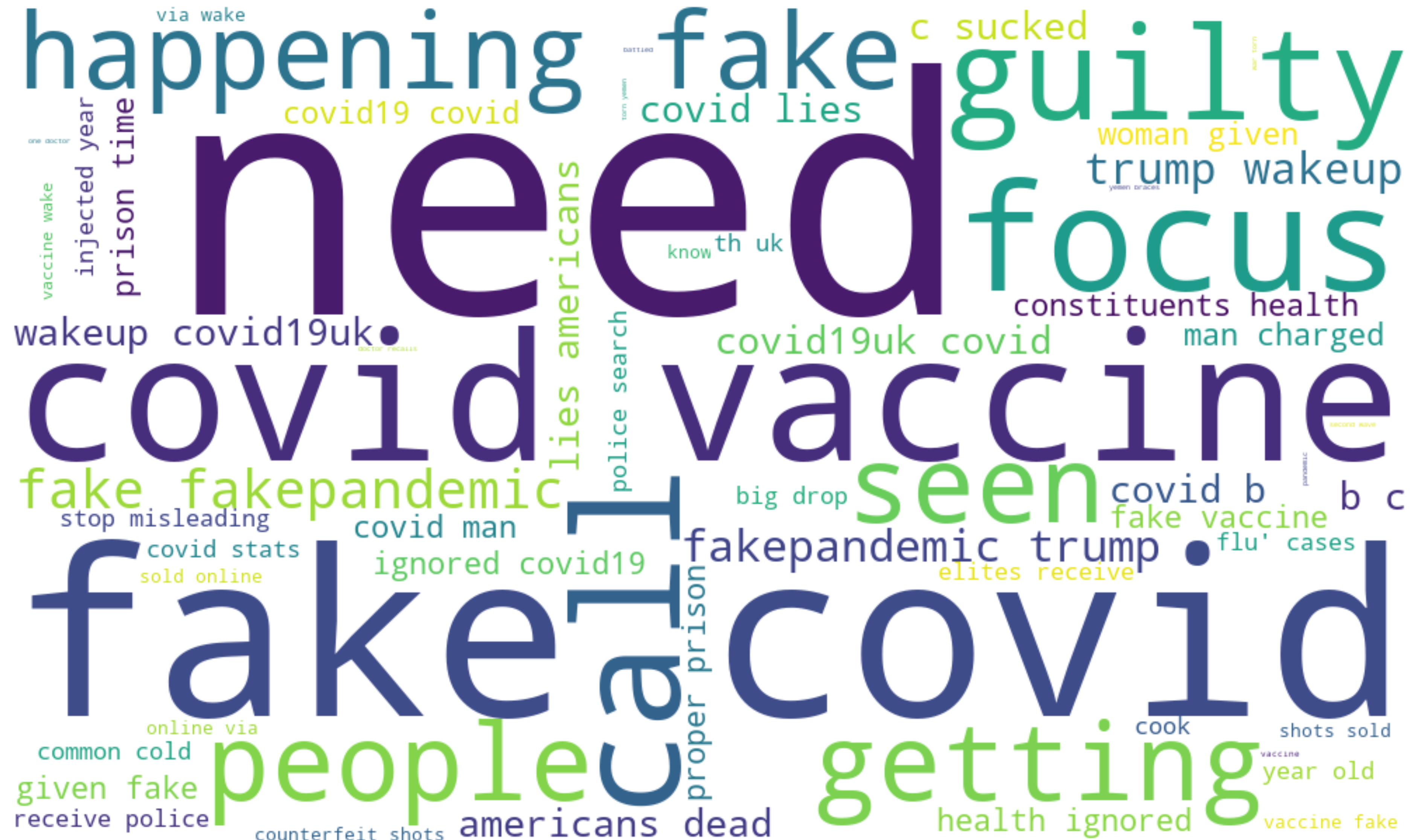
#CovidVaccination



#anit-Vax

Topic Modeling

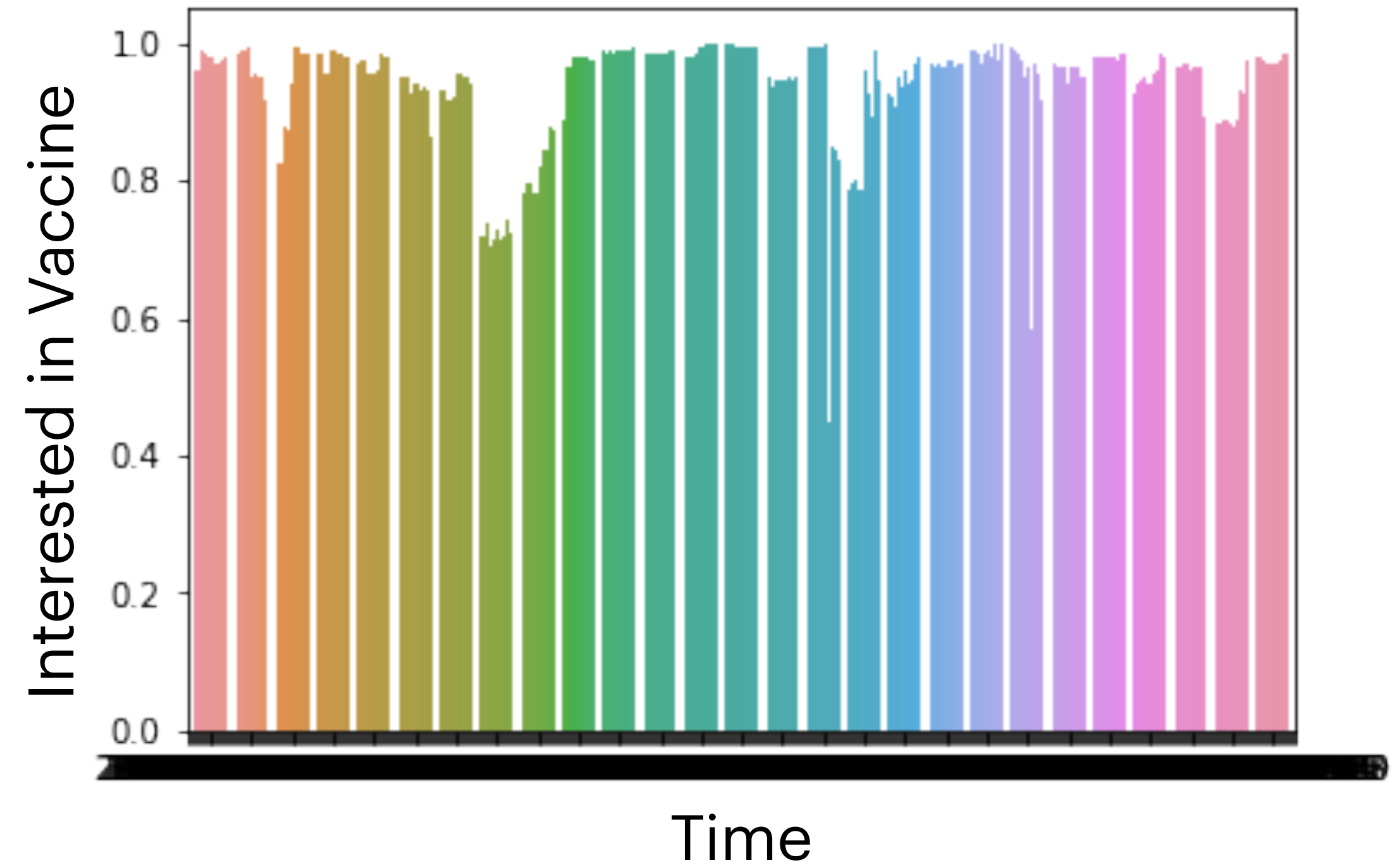
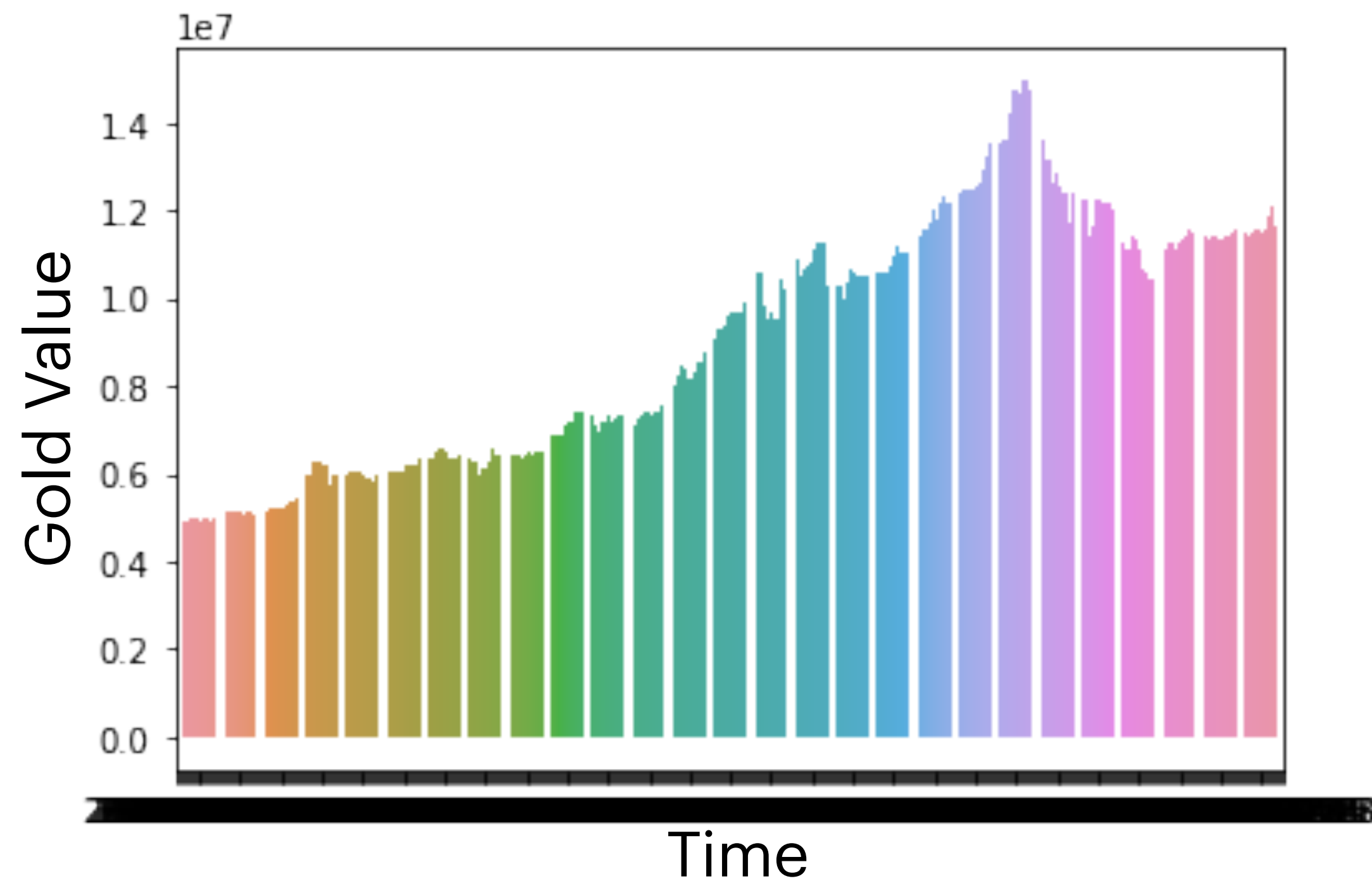
#Covid AND #Fake



Market Analyzer and Correlation with Sentiment

Gold Price

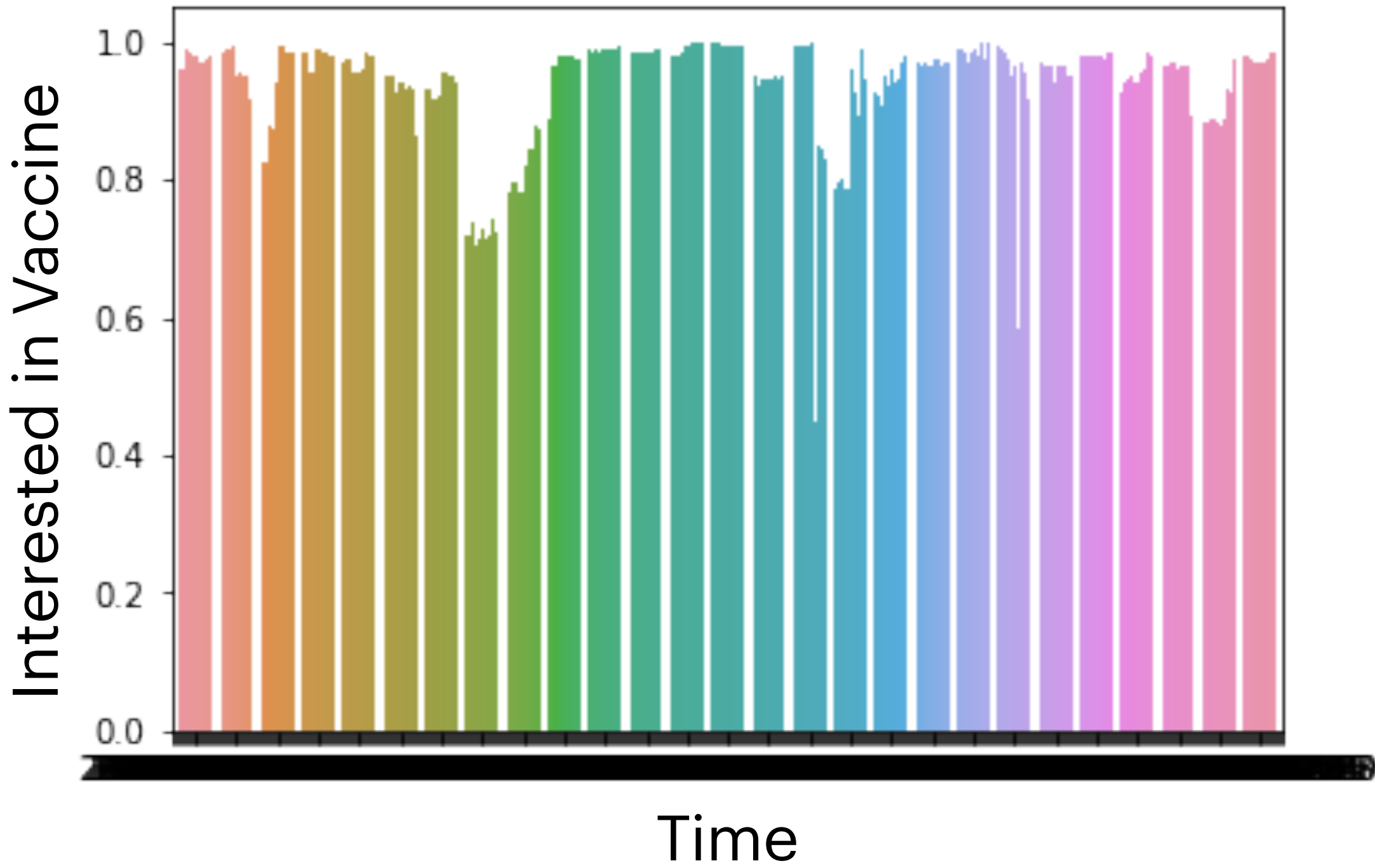
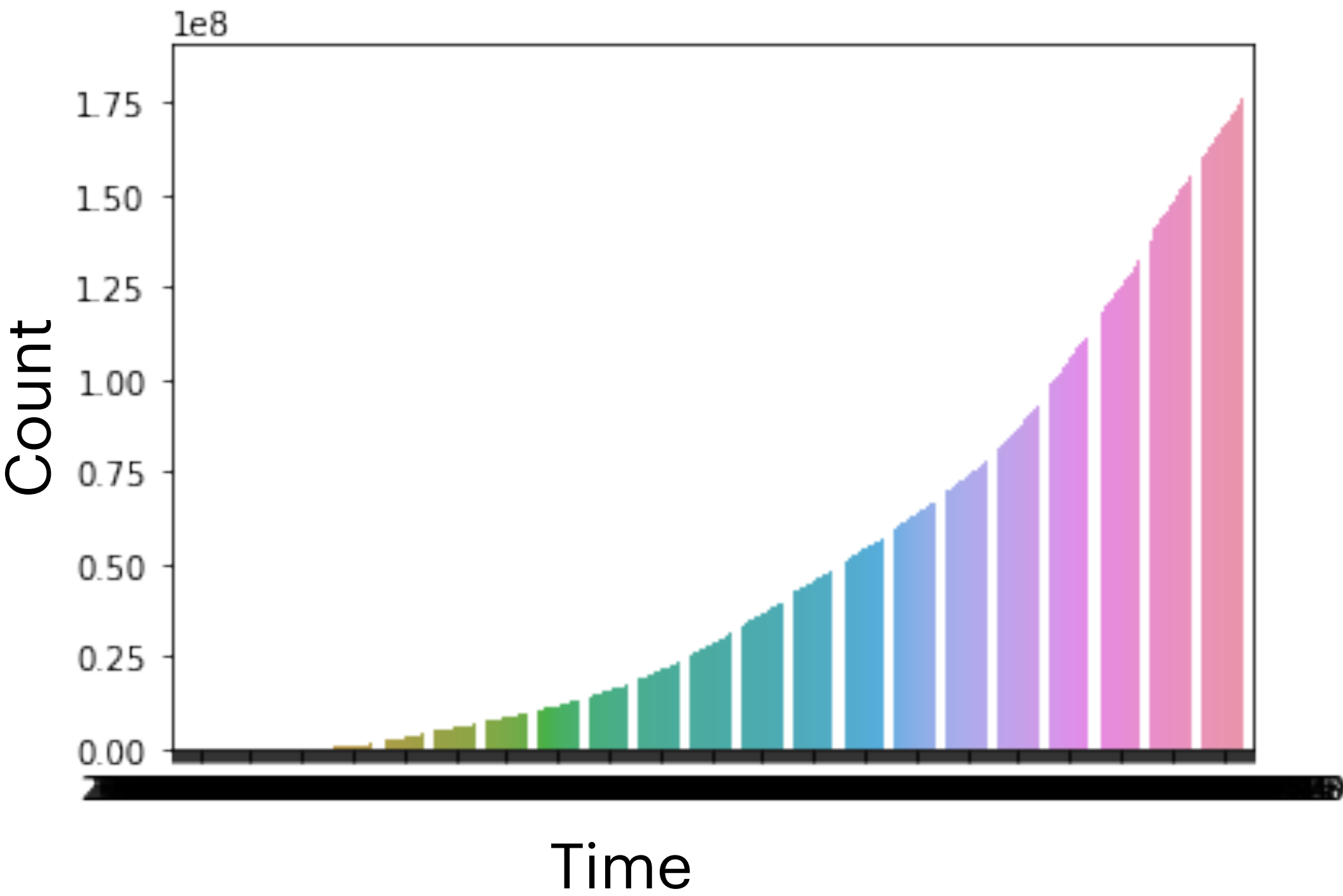
Correlation: 0.138



Covid Statistics and Correlation with Sentiment

Total Cases

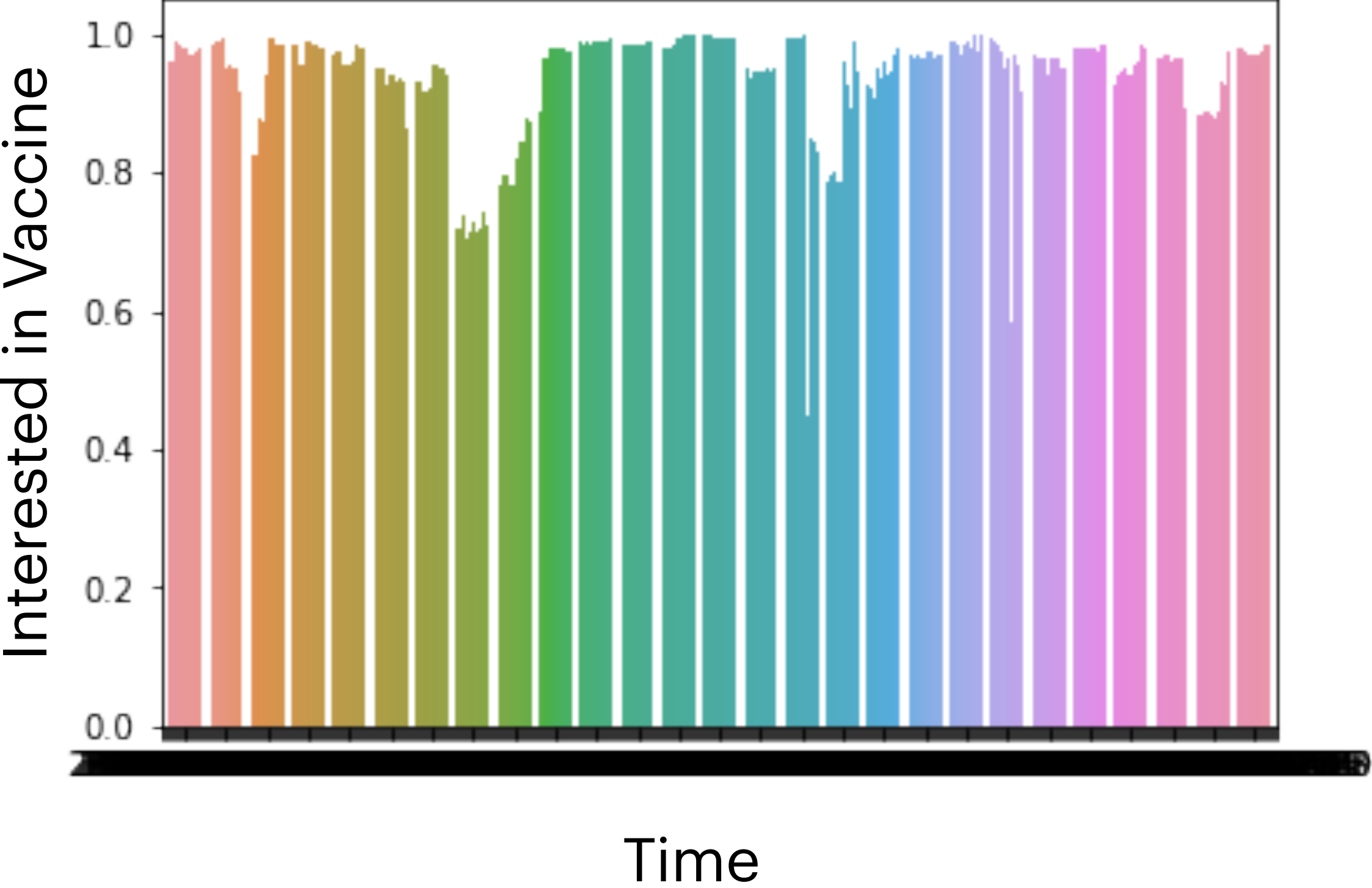
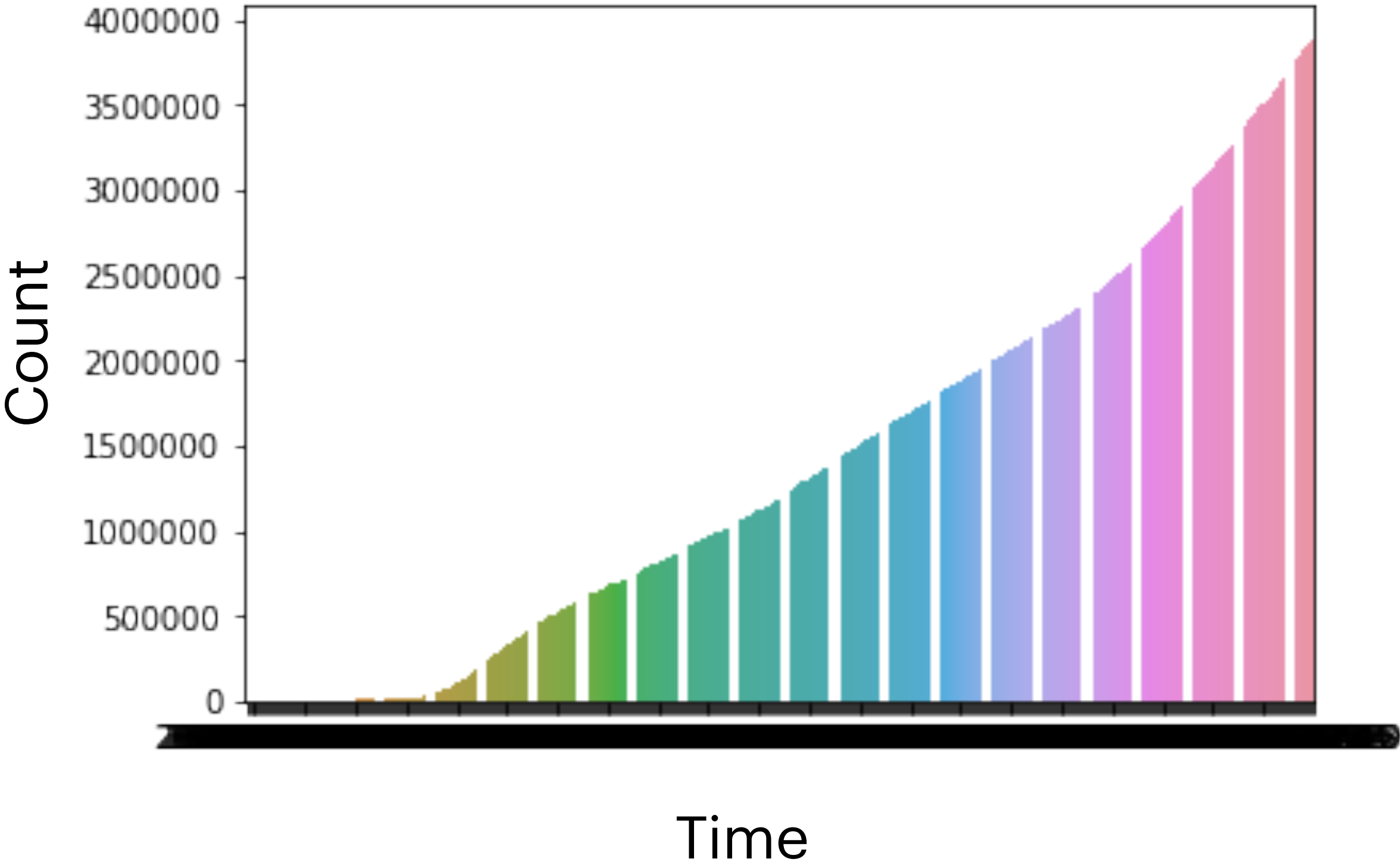
Correlation: 0.124



Covid Statistics and Correlation with Sentiment

Total Death

Correlation: 0.122



Comparisons with Baselines

- In order to comparison with baseline and up to the literatures about Covid and Social media that I studied, there are two group of approaches:
- Misinformation [1,2]
- Some statistics about the effect of Covid and the people feedback on Vaccine. [3-6]
- The main questions that we studied in this research are not well-aligned with the available researches up to my knowledge. So I did not find any baseline for comparison.
- Furthermore, as our datasets are not the same with these literatures, and their dataset are not publicly available, I can not find comparison the section of our research that has intersection with them.

Main Results and Findings

- We can predict the sentiment of people about Covid, Vaccine and Anti-vaccine.
- The correlation of Financial Market and Covid Vaccine is positive.
- The correlation of Covid statistics like as new cases and death numbers are positive.
- The correlation of Covid statistics like as new cases and death numbers with vaccine sentiments are positive.
- The correlation of mental health of people with vaccine sentiments are positive.
- Find the Impressive news from #BBC and #CNN based on the local minima of Mental Health Status and Vaccine Feedback of people.

Future Research Directions

- This is a very vast topic and a lot of variant thins can be considered. So, I just mention the important ones.
- In main ideas and questions of research
 - I am mostly interested to study the effect of Covid and Vaccinate on business using linked in.
 - There are a lot of other ideas
 - The effect of quarantine
 - The rate of the scholars
 - Scholars in covid per different cities and countries
 - The rate of the scholars are increased or not
 - Business, #Tesla
 - Low variant of data for location and language
 - Retweets and study them
 - Study user conversational tree

Future Research Directions

- Data Collection process:
 - Collecting data in Reddit which has more technical and professional data
 - Collect more data
 - Compute dependent keywords for searching query using word to vector.
 - Language based collection
 - Location based data collection
 - Retweets and study them
 - Study user conversational tree
- Methods
 - Graph Mining
 - Graph of Scholars
 - Result of scholars

Limitations and Issues

- Twitter and the Twitter-API for getting is filtered and inaccessible in Iran. Also because of problem in internet connection in Iran it takes me about 6 days to collect data. By these problems I can not collect the volume of data that I needs. This is very effectual on the results, and so
- The google Colab is not accessible in Iran. We can use it with hardship but because of the internet connection it is somehow useless.
- When I do some research in our university we are accessing to computational servers, but my personal laptop has ram 4 and Corei5, because of that I had some problem in computation too.

References

Literature Review of Research

1. Singh, Lisa, Shweta Bansal, Leticia Bode, Ceren Budak, Guangqing Chi, Kornraphop Kawintiranon, Colton Padden, Rebecca Vanarsdall, Emily Vraga, and Yanchen Wang. "A first look at COVID-19 information and misinformation sharing on Twitter." *arXiv preprint arXiv:2003.13907* (2020).
2. Yang, Kai-Cheng, Christopher Torres-Lugo, and Filippo Menczer. "Prevalence of low-credibility information on twitter during the covid-19 outbreak." *arXiv preprint arXiv:2004.14484* (2020).
3. Bonnevie, Erika, Allison Gallegos-Jeffrey, Jaclyn Goldbarg, Brian Byrd, and Joseph Smyser. "Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic." *Journal of Communication in Healthcare* (2020): 1-8.
4. Jahanbin, Kia, and Vahid Rahmanian. "Using Twitter and web news mining to predict COVID-19 outbreak." *Asian Pacific Journal of Tropical Medicine* 13 (2020).
5. Chou, Wen-Ying Sylvia, and Alexandra Budenz. "Considering Emotion in COVID-19 vaccine communication: addressing vaccine hesitancy and fostering vaccine confidence." *Health communication* 35, no. 14 (2020): 1718-1722.
6. Alshaabi, Thayer, Michael V. Arnold, Joshua R. Minot, Jane Lydia Adams, David Rushing Dewhurst, Andrew J. Reagan, Roby Muhamad, Christopher M. Danforth, and Peter Sheridan Dodds. "How the world's collective attention is being paid to a pandemic: COVID-19 related n-gram time series for 24 languages on Twitter." *PloS one* 16, no. 1 (2021): e0244476.