

Han Thar

Loureen Vilorio

Final Project – Social Networks

## Analyzing Twitter Users' Sentiment about COVID-19 Vaccines

### Introduction

COVID-19, the pandemic that took over the world since early 2020 and is still taking over some countries worldwide. As of today, December 9, 2021, there are a total of 268 million cases of COVID-19 worldwide. After being quarantined for months, adapting to various mandates set by the government to keep us healthy and safe, many people became frantic and voiced their concerns about the pandemic in social media.

Earlier this year, Pfizer released their vaccine to the masses, giving members of our community protection against the virus. Other pharmaceutical companies like Johnson & Johnson and Moderna also released their vaccines later of this year. With everyone having access to vaccines against COVID, many still voiced their skepticism about the vaccine with Twitter being one of their outlets to voice their opinions.

Because of this divide, my partner and I decided to analyze tweets regarding the user's opinions about the vaccines to find the overall sentiment about COVID-19 vaccines.

The dataset we used for this analysis was the COVID-19 All Vaccines Tweets by Gabriel Preda (<https://www.kaggle.com/gpreda/all-covid19-vaccines-tweets>). This dataset was last updated on November 23, 2021, and currently have 228,207 data from various Twitter users collected from December 12, 2020, until November 15, 2021. The dataset contains the following:

- |                    |                   |              |
|--------------------|-------------------|--------------|
| • Id               | • User_friends    | • Source     |
| • User_name        | • User_favourites | • Retweets   |
| • User_location    | • User_verified   | • Favorites  |
| • User_description | • Date            | • Is_retweet |
| • User_created     | • Text            |              |
| • User_followers   | • Hashtags        |              |

For this analysis, we are only using the user\_name, user\_location, date, text, hashtags, and source as these are the data that can provide quantitative data, we need to provide an in-depth analysis about the user's sentiment about the vaccines.

### Methods of Implementation

Our group used Jupyter Notebook to create this program as Jupyter is a Python IDE that is easy to use for data science. We installed and updated the necessary dependencies and imported libraries we think are useful to create and get the data we need for this project.

```
#imports
from nltk.corpus import stopwords
from wordcloud import WordCloud
from textblob import TextBlob
import matplotlib.pyplot as plt
import neattext as ntx
import pandas as pd
from collections import Counter
import numpy as np
import string
import nltk
import re
nltk.download('stopwords')
nltk.download('wordnet')
```

We implemented a handful of functions in the program to make it easier for us to remove unnecessary texts, columns and punctuations as well as getting data we need to make a graph. Overall, we implemented nine functions for this program.

```
def remove_punctuation(text): #removing punctuations in text
def stopwords_removal(tweet): #removing unnecessary stop words
def text_blob(text): #function used to assign polarity and sentiment
def get_max_token_words(tweets, num = 50): #function to get the most used words
def data_figs(x, sent): #function to create graphs for tweet sentiments
def word_cloud(tweet, sent): #creates a wordcloud figure
def sent_percent(x,sent): #calculates the percentage of each sentiment
def vaccine_reference(tweet, reference): #counts how many vaccine references
def get_stats(pf, bi, sp, a, mo, si): #calculates the polarity and subjectivity's
mean, max, min, and median of each vaccine
```

## Results and Discussions

In this section, we will talk about all the results our team gathered from our program by reading and analyzing the contents of vaccination\_all\_tweets.csv based on their description, type, polarity, and sentiments. The vaccines that we focused on this analysis were the following:

- Pfizer
- Sinovac
- Sputnik V
- Covaxin
- Moderna
- Sinopharm
- AstraZeneca

Below are the results and in-depth discussion of the output we gathered for this project.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
id	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	retweets	favorites	is_retweet
1	1.34E+18 Rachel Roh	La Crescenta-Montrose, CA	Aggregator of Asian American n	4/8/2009 17:52	405	1692	3247	FALSE	#####	Same folk /PzizerBk Twitter fo			0	0	FALSE
2	1.34E+18 Albert Fong	San Francisco, CA	Marketing dude, tech geek, hea	9/21/2009 15:27	834	666	178	FALSE	#####	While the world has Twitter W			1	1	FALSE
3	1.34E+18 elidYtZdY+GyTgDvHfCYC	Your Bed	heil, hydra Vy-ä"n	6/25/2020 23:30	10	88	155	FALSE	#####	#coronavi /coronavi Twitter fo			0	0	FALSE

Figure 1. Snippet of vaccinations\_all\_tweets.csv

## Location

The first data we gathered was the locations of the users. Based on the graphs shown below, most Twitter users were from India, with Canada and USA coming in second and third.

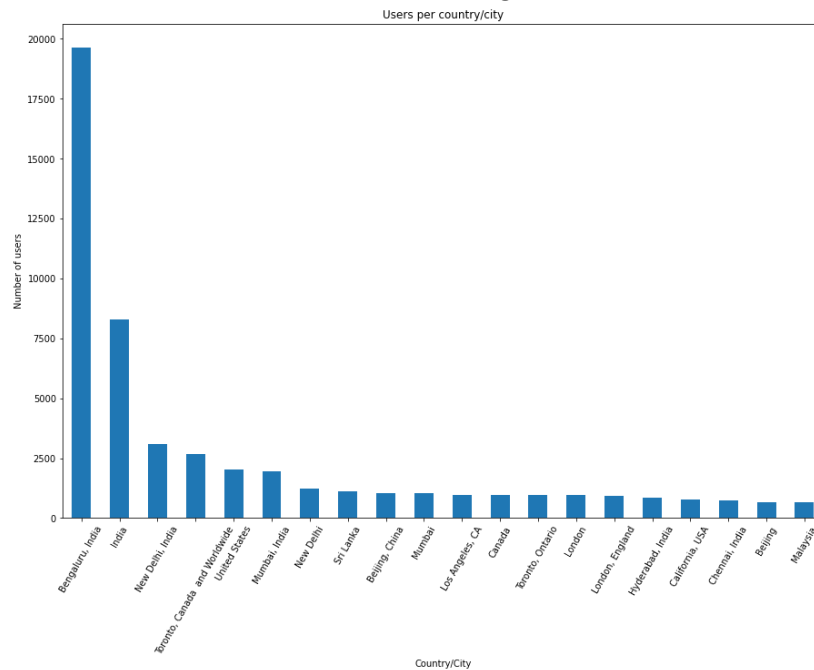


Figure 2. Users per country/city (bar graph)

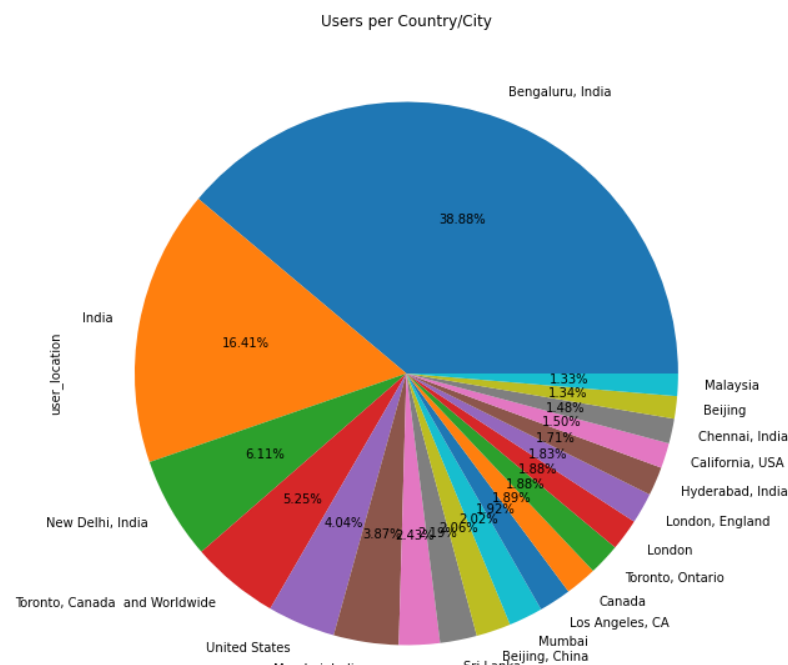


Figure 3. Users per country/city (pie chart)

## Source of tweet

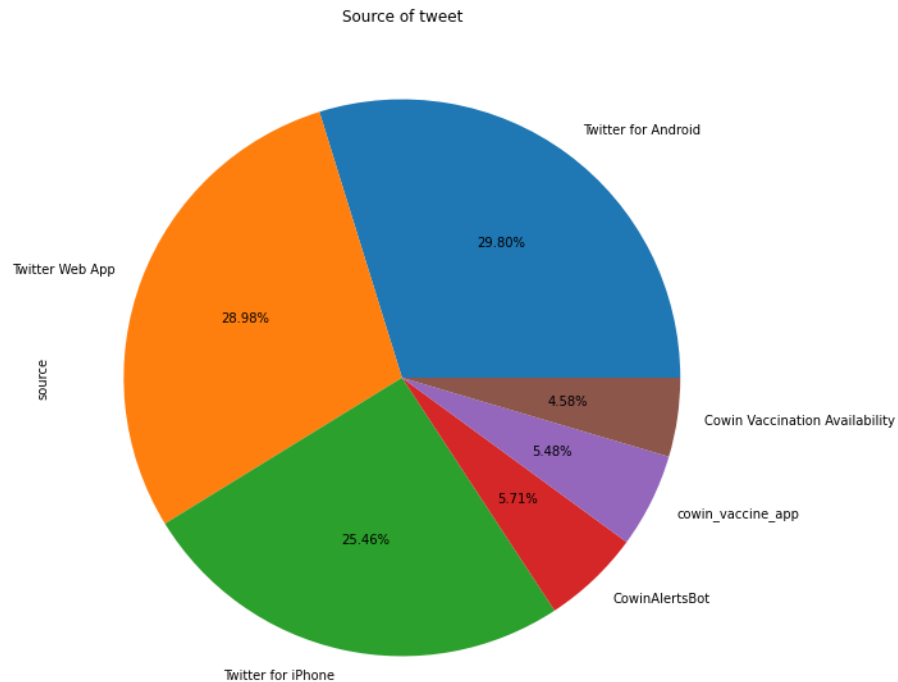


Figure 4. Source of tweet

The next data that we pulled was the source of Tweet, what app or device did the users used to send a tweet. As shown above, most users send their tweets by using Twitter's application through the top three devices most consumers use: Android, iPhone, and the Web(browser). The rest of the tweets came from a COVID app based in India.

## Sentiments

After getting the locations and source of tweet, our team "cleaned" the text column by removing stop words, urls, user handles, multiple spaces, and special characters on each tweet. After removing the unnecessary words and symbols on each tweet. We then assigned the polarity, subjectivity, and sentiment of each tweet by creating a function using text\_sentiment function of TextBlob library. After that, we separated each sentiment to their own series and gathered the maximum token words to get the most used words that relay different sentiments.

```
#assigning polarity and subjectivity to the tweets
def text_blob(text):
    text_sentiment = TextBlob(text)
    sentiment_pol = text_sentiment.sentiment.polarity
    sentiment_sub = text_sentiment.sentiment.subjectivity

    if sentiment_pol > 0:
        r = 'Positive'
```

```

elif sentiment_pol < 0:
    r = 'Negative'
elif sentiment_pol == 0:
    r = 'Neutral'

sentiment_result = {'polarity': sentiment_pol, 'subjectivity': sentiment_sub,
'sentiment': r}

return sentiment_result

```

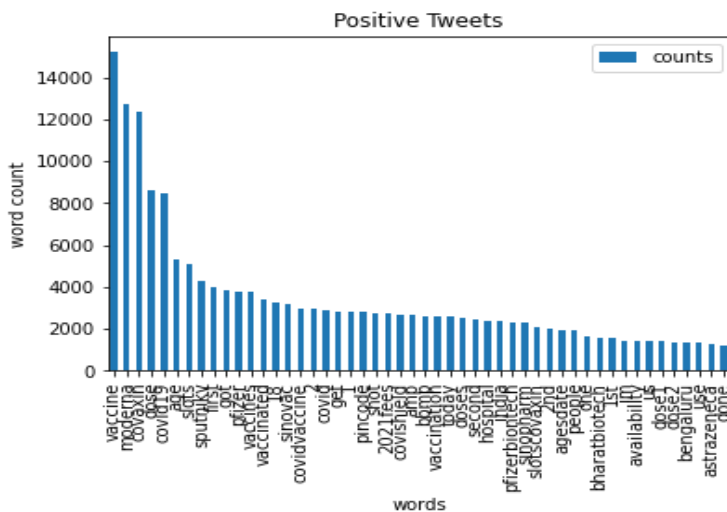


Figure 5. Most used words that were used in a positive tweet

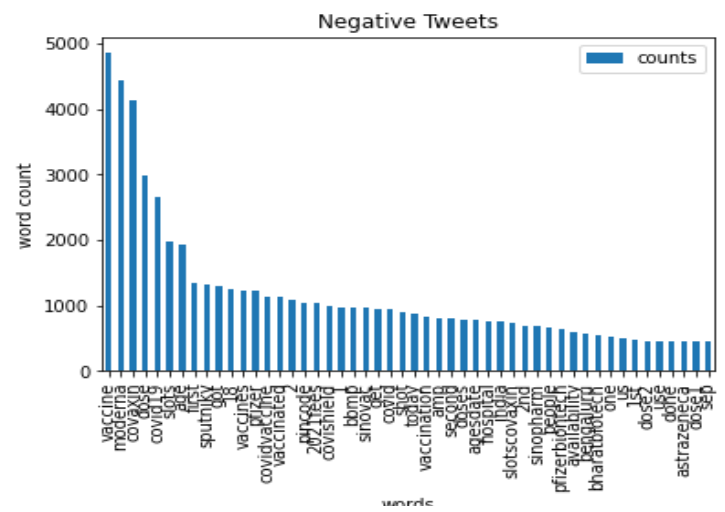


Figure 6. Most used words that were used in a negative tweet

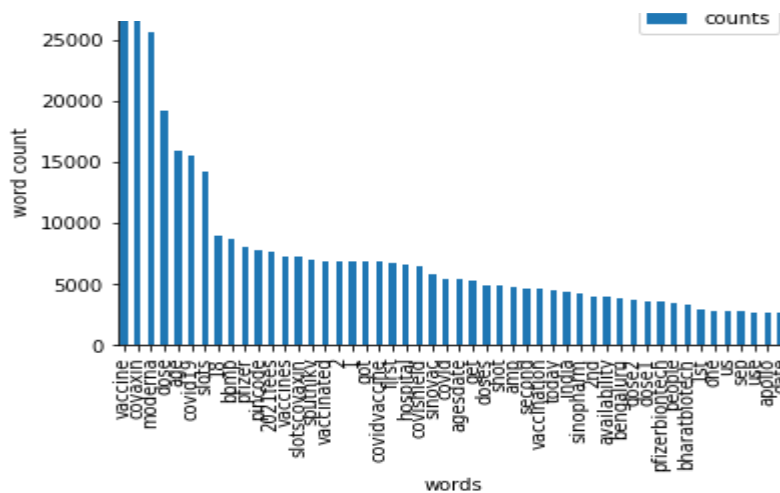


Figure 7. Most used words that were used in a neutral tweet

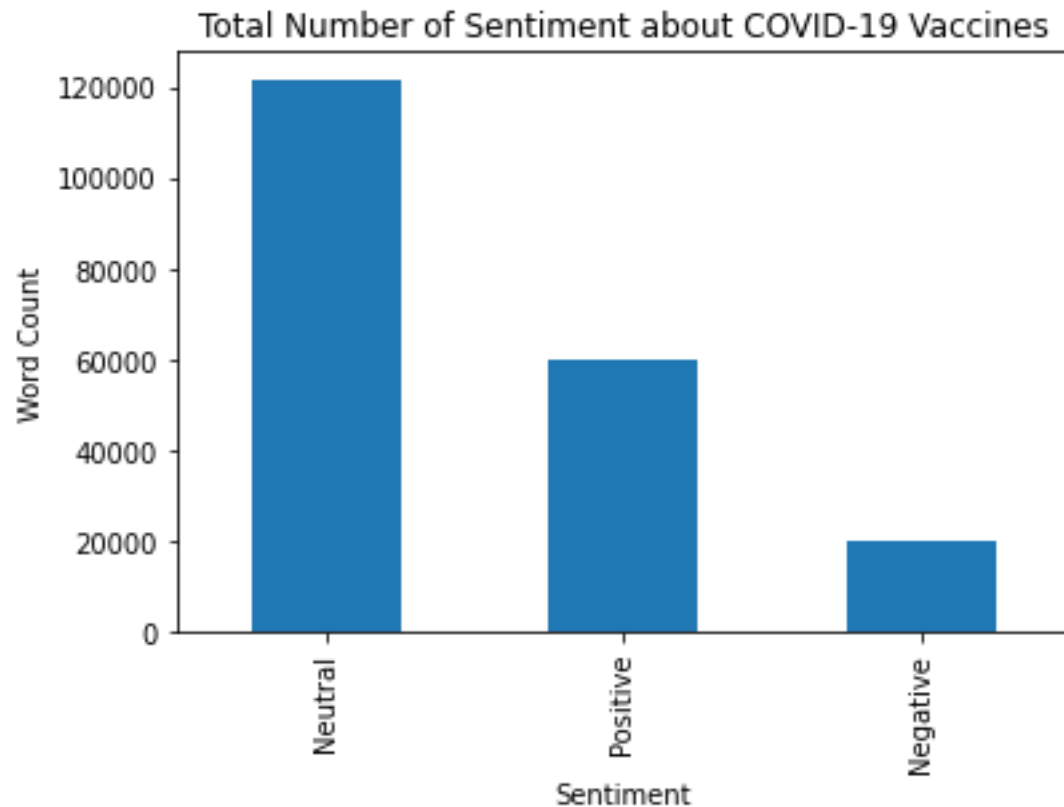


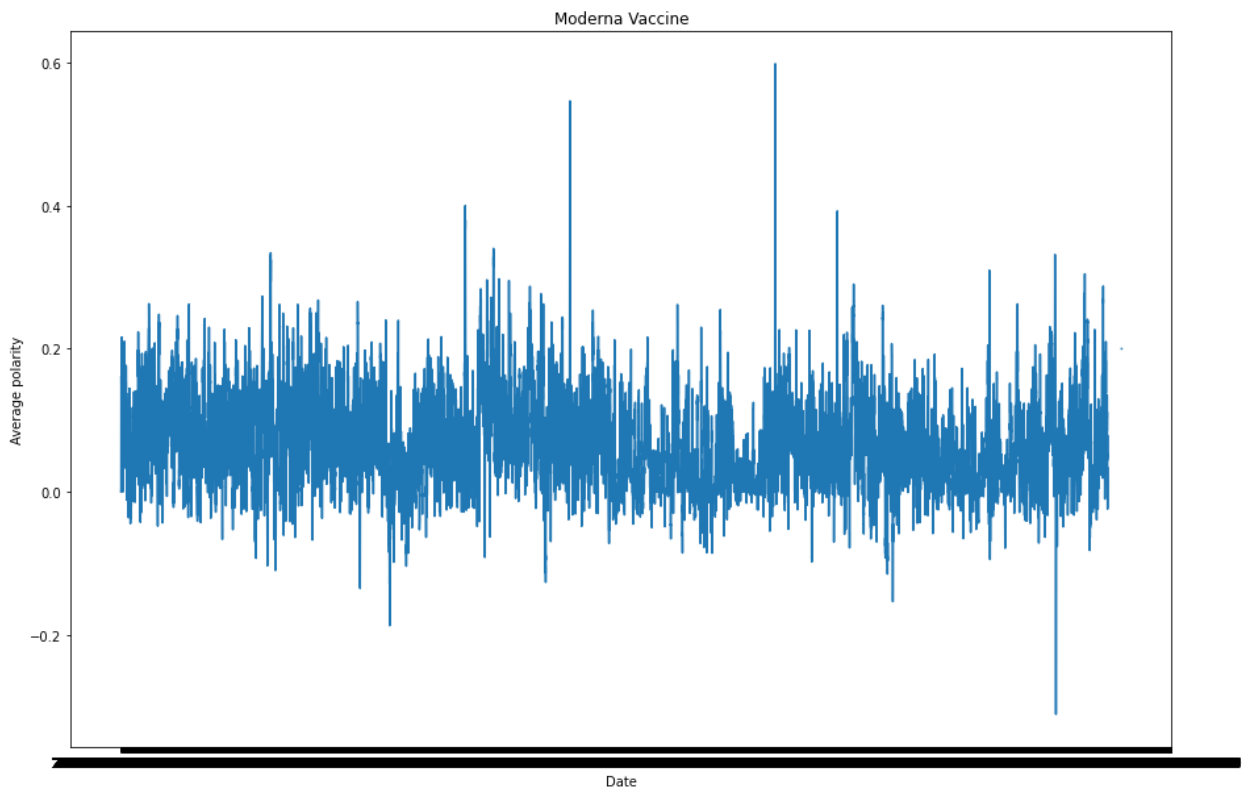
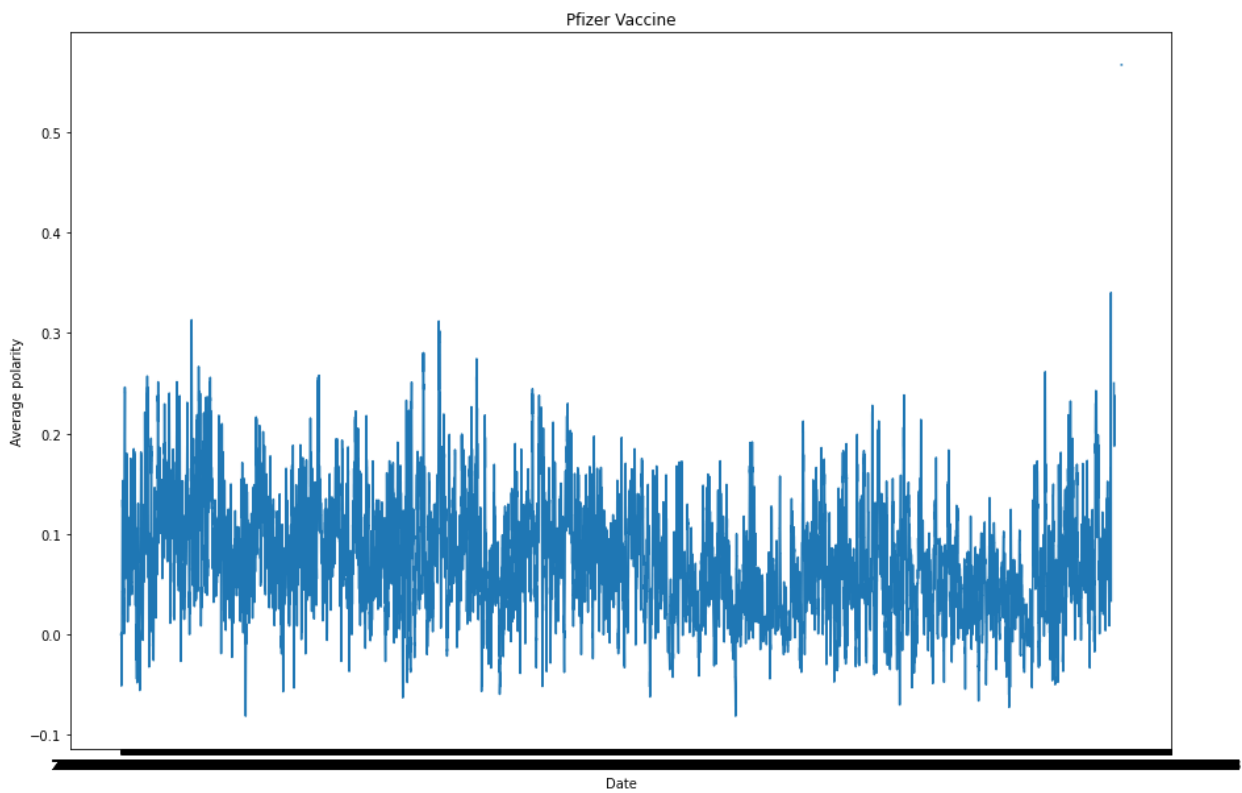
Figure 8. Value of each sentiment

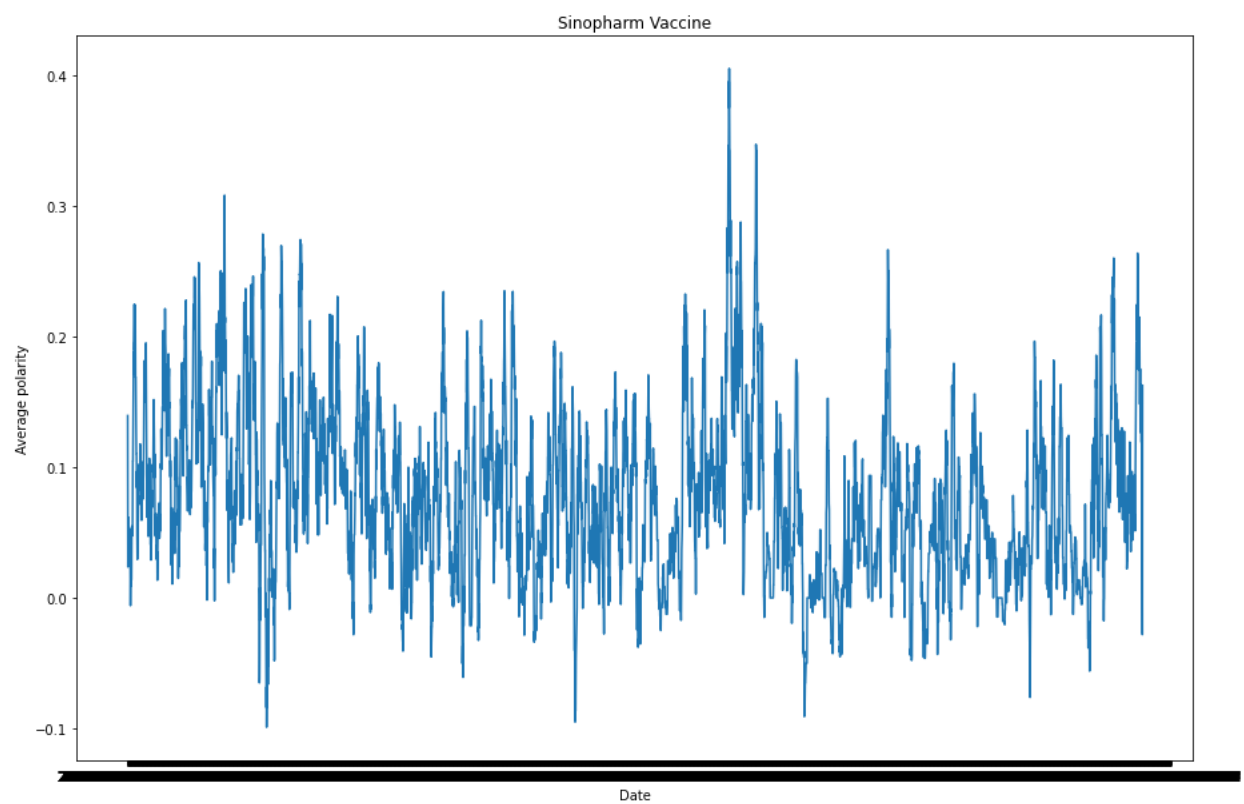
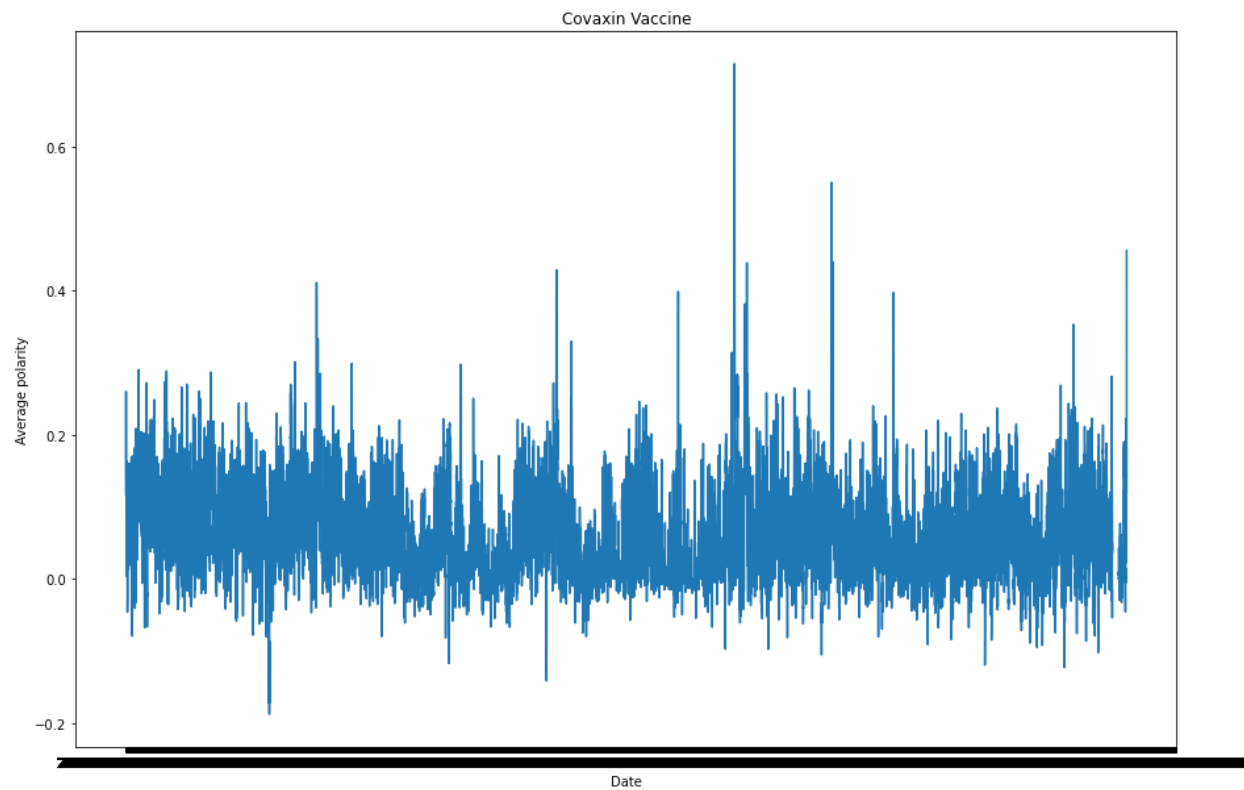
Based on the data we gathered from getting sentiments on each tweet, most people feel neutral about the vaccines. When calculating the percentage of each sentiment, it shows that most users are 56.75% feel neutral about the vaccines, while 27.94% of the users feel positive and 9.28% of users feel negative about the vaccine.

Looking at individual sentiment graphs, vaccines are the most used word next are Covaxin and Moderna. After these words, it varies what were the other words used by the users that made them feel positive, negative, or neutral about the vaccine.

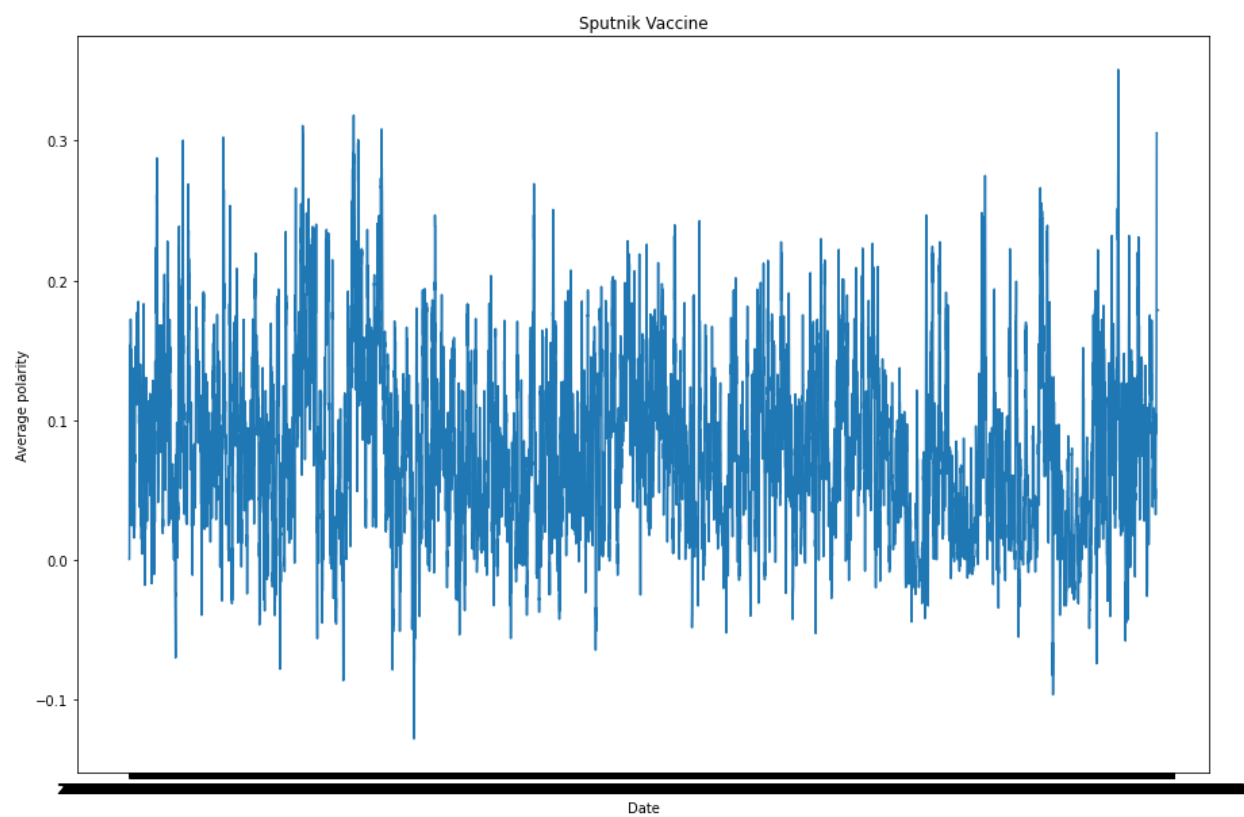
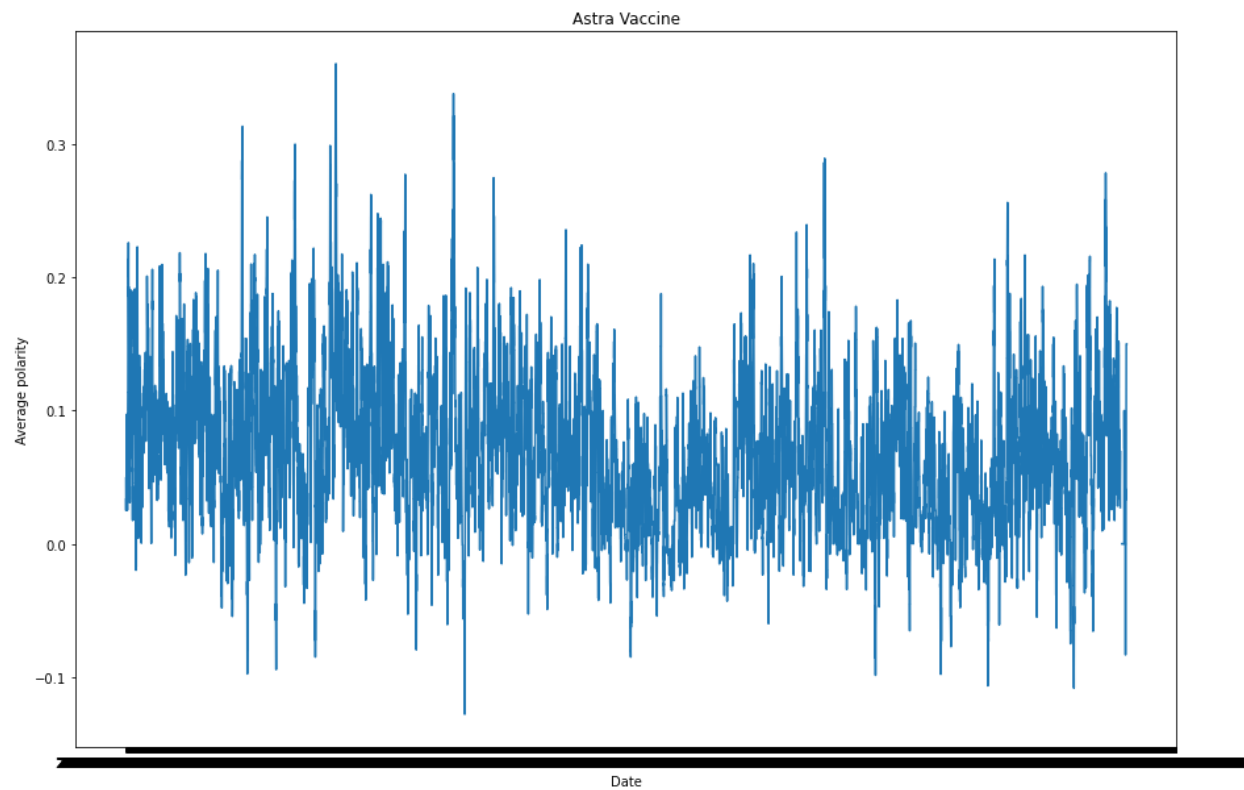
For the positive sentiment, one can assume that the users were happy about the vaccine as the most at risk people are getting it already and overtime more vaccines became available worldwide. While the users with negative sentiment about the vaccine were possibly scared on how fast it was researched and distributed and possibly scared about the dose and the age range of people who can get vaccinated. Lastly, the people who felt neutral about the vaccine can be a combination of both sentiments but still unsure on where they stand.

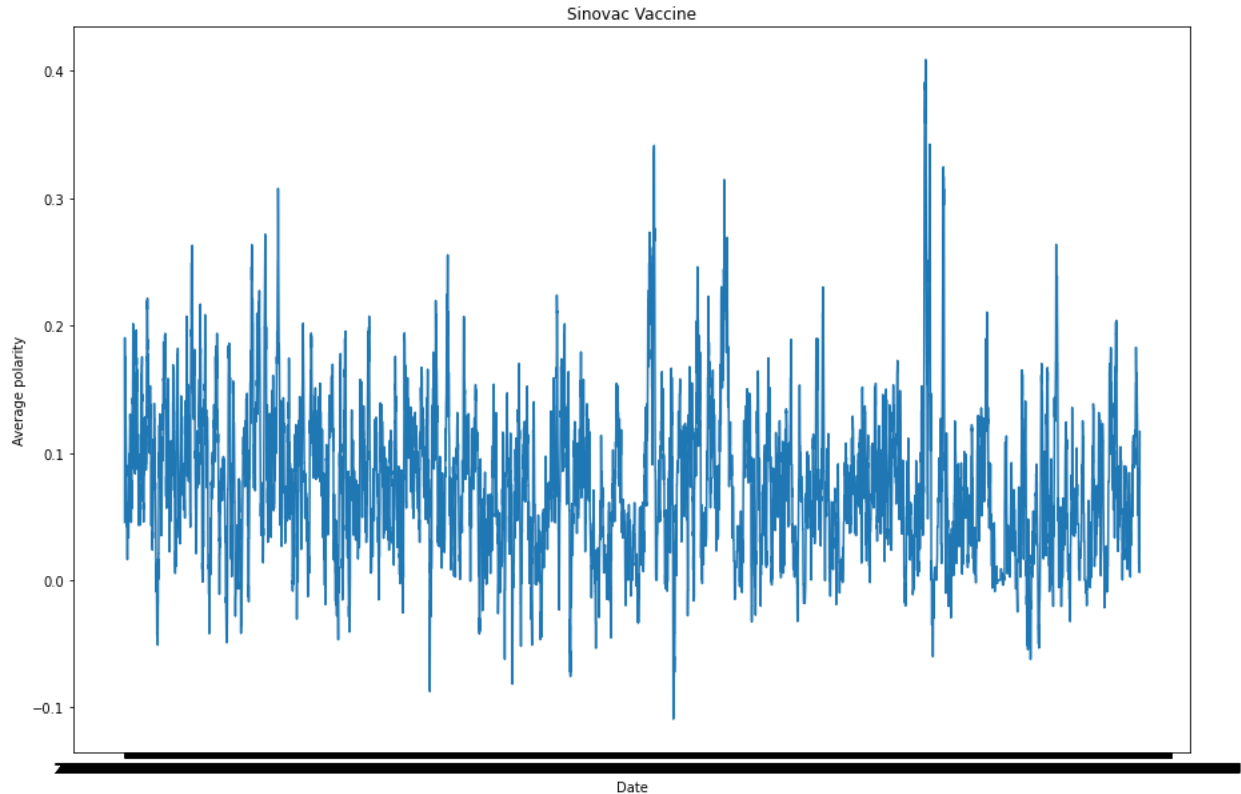
**Polarity of each vaccine**











Based on the graphs above, it shows that most users feel neutral about all vaccines as none of the vaccines got to 1, which represents as the most positive. We can see that the polarity spikes up and down as time goes by, but never fully achieved positive sentiments from Twitter users. The vaccines that were above neutral (0.5) were Covaxin and Moderna. These vaccines received above average sentiment from the users against the other brand of vaccines.

## Conclusion

Although the vaccines are being distributed worldwide, 56.75% of users from Twitter still feel neutral about the vaccines. 27.94% of users feel positive about the vaccine and already had their 2<sup>nd</sup> or booster shot. One possibility on why most users feel neutral about vaccines is because of COVID-19 variants that started spreading worldwide later this year. With the new variants, Twitter users voiced their concerns if the vaccines will still work or if drug companies need to research a new vaccine against this variant.

## Areas of improvement

If my team are the ones who will gather data from Twitter about COVID-19 vaccines, I would do the following:

1. Create a program that collect data from Twitter by using a Python library called Tweepy.
2. Focus on a specific country or continent to make the data analysis more concrete and concise.
3. Instead of using the location of the user that is listed in their profile, I would use the user's location when the tweet was posted.
4. Research more about various libraries that could make data analysis better, such as creating a graph that shows monthly intervals instead of daily.