Exploring the Effects of COVID-19 Vaccines Post-Vaccination: A Twitter Analysis Based on Dose and Vaccine Type

Anagha. Saksham. Purva. Rachana.

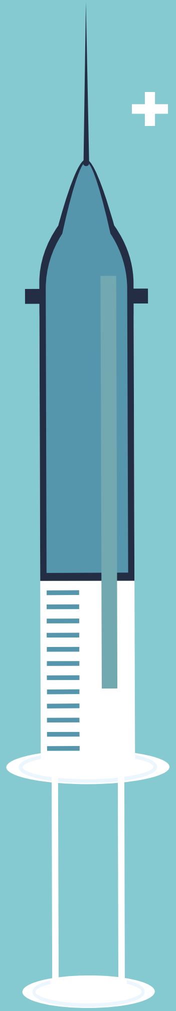
TABLE OF CONTENTS

01 INTRODUCTION

02 GOALS

03 DATA COLLECTION AND DATASET

04 DATA PREPROCESSING



05 EXAMPLE SCREENSHOTS

06 DATA ANALYSIS

07 FINDINGS / RESULTS

08 SUMMARY AND LIMITATIONS

INTRODUCTION

- COVID-19 caused death and disruption worldwide
- Vaccines offer hope, but side effects and disclosure are concerns
- Twitter is popular for sharing COVID-19 vaccine experiences
- Project analyzes Twitter data on vaccine side-effects disclosure

GOALS

To ascertain the varied effects of different vaccines with the same dosage

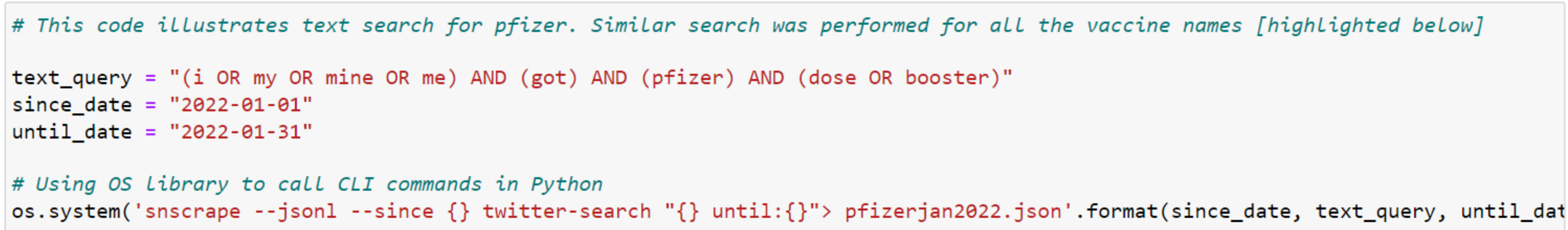
Determine whether there are differences in the types of side effects reported on Twitter for different dosages of the COVID-19 vaccine (e.g., dose 1, dose 2, booster).

To investigate whether there are any significant variations in the common topics and side-effects that people tweet about after receiving COVID-19 vaccines, based on the specific type of vaccine and the number of doses taken.

DATA COLLECTION

- In order to scrape data from Twitter we had 2 option.
  1. Use Twitter API's and Python library like Twarc and Tweepy
  2. Use open-source Python library like Snscrape
- We chose to use Snscrape to get all the tweets from twitter.

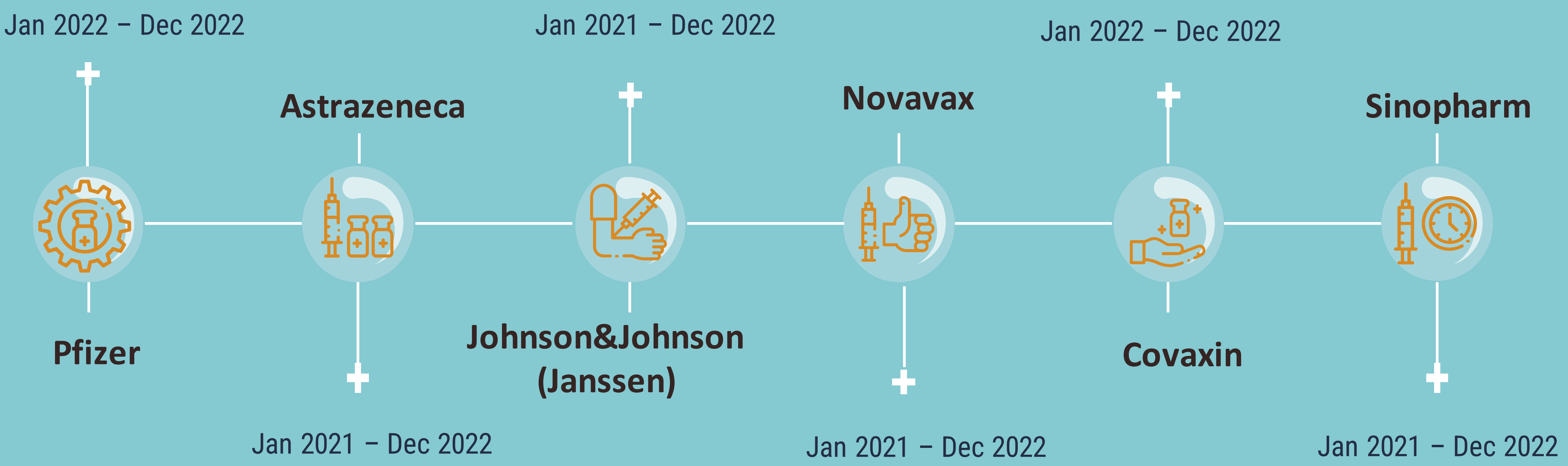
Example screenshot of snscrape:



```
# This code illustrates text search for pfizer. Similar search was performed for all the vaccine names [highlighted below]

text_query = "(i OR my OR mine OR me) AND (got) AND (pfizer) AND (dose OR booster)"
since_date = "2022-01-01"
until_date = "2022-01-31"

# Using OS library to call CLI commands in Python
os.system('snscrape --jsonl --since {} twitter-search "{} until:{}"> pfizerjan2022.json'.format(since_date, text_query, until_dat
```

- The text query includes keywords like 'I', 'me', 'mine', or 'my' for personalized individual tweets, followed by the name of the vaccine and then what kind of 'dose' or 'booster'.
- This is followed by a date range.

DATASET



Pfizer

Jan 2022 – Dec 2022

Astrazeneca

Jan 2021 – Dec 2022

Johnson&Johnson (Janssen)

Jan 2021 – Dec 2022

Novavax

Jan 2021 – Dec 2022

Covaxin

Jan 2022 – Dec 2022

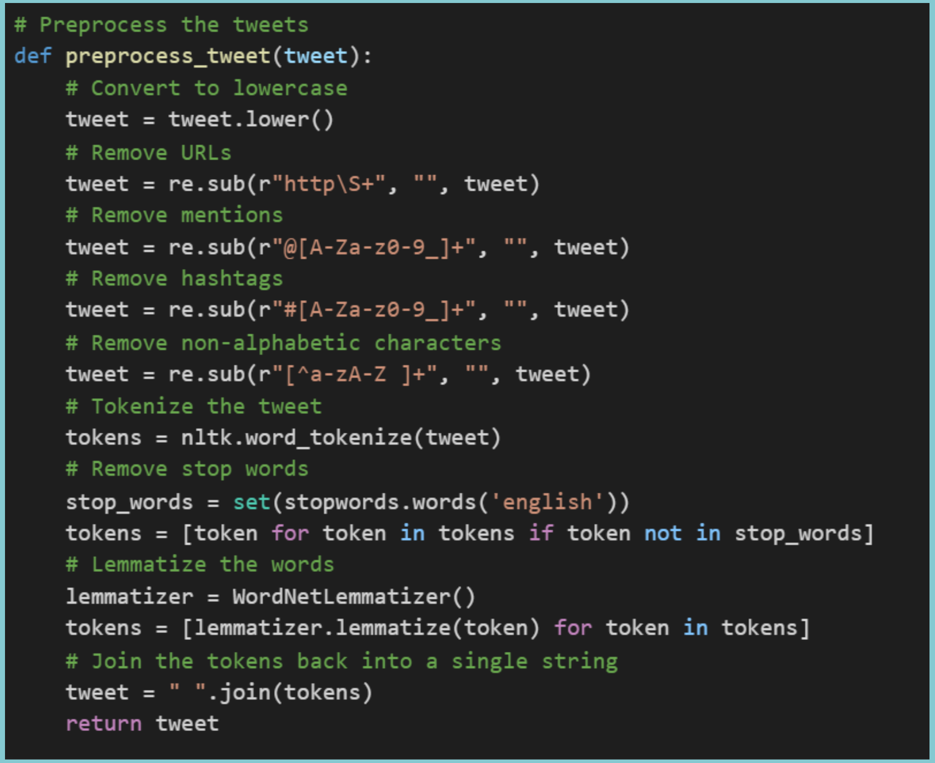
Sinopharm

Jan 2021 – Dec 2022

- The total number of tweets collected from the above dataset is **32,092** and these tweets are stored in .csv files.
- We initially considered 6 vaccines but due to computation problems, we dropped Sinopharm and Covaxin.
- For Pfizer, we could only scrape 2022 tweets.

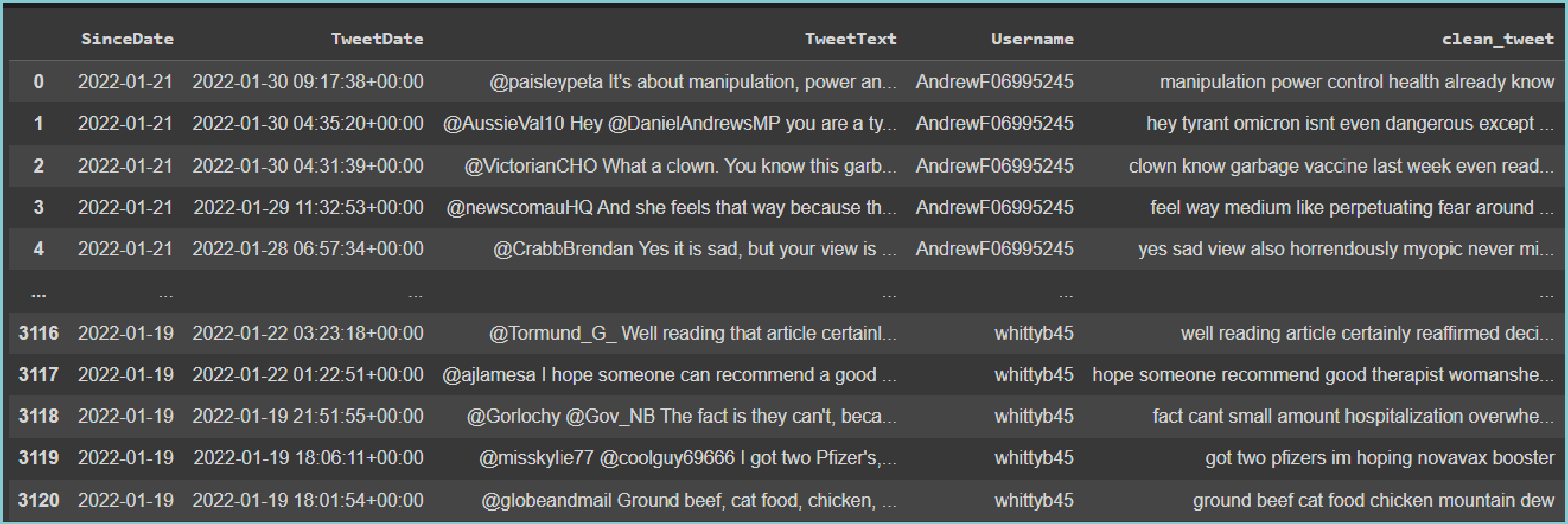
DATA PRE-PROCESSING

- For each user, we collected all the tweets they made in the 10 days following the initial tweet. These included health and non-health-related tweets.
- The total number of tweets at this stage was: 1,904,551
- The next step was to clean the tweets. We used NLTK to clean all the tweets. This included 2 parts:
  - Removal
    - Stop words
    - Hashtags
    - Mentions
    - URLs
    - Non-alphabetic characters.
  - Conversion
    - Tweet to lowercase
    - Tokenized the tweets
    - Lemmatized the words
    - Joined the token back into a single string



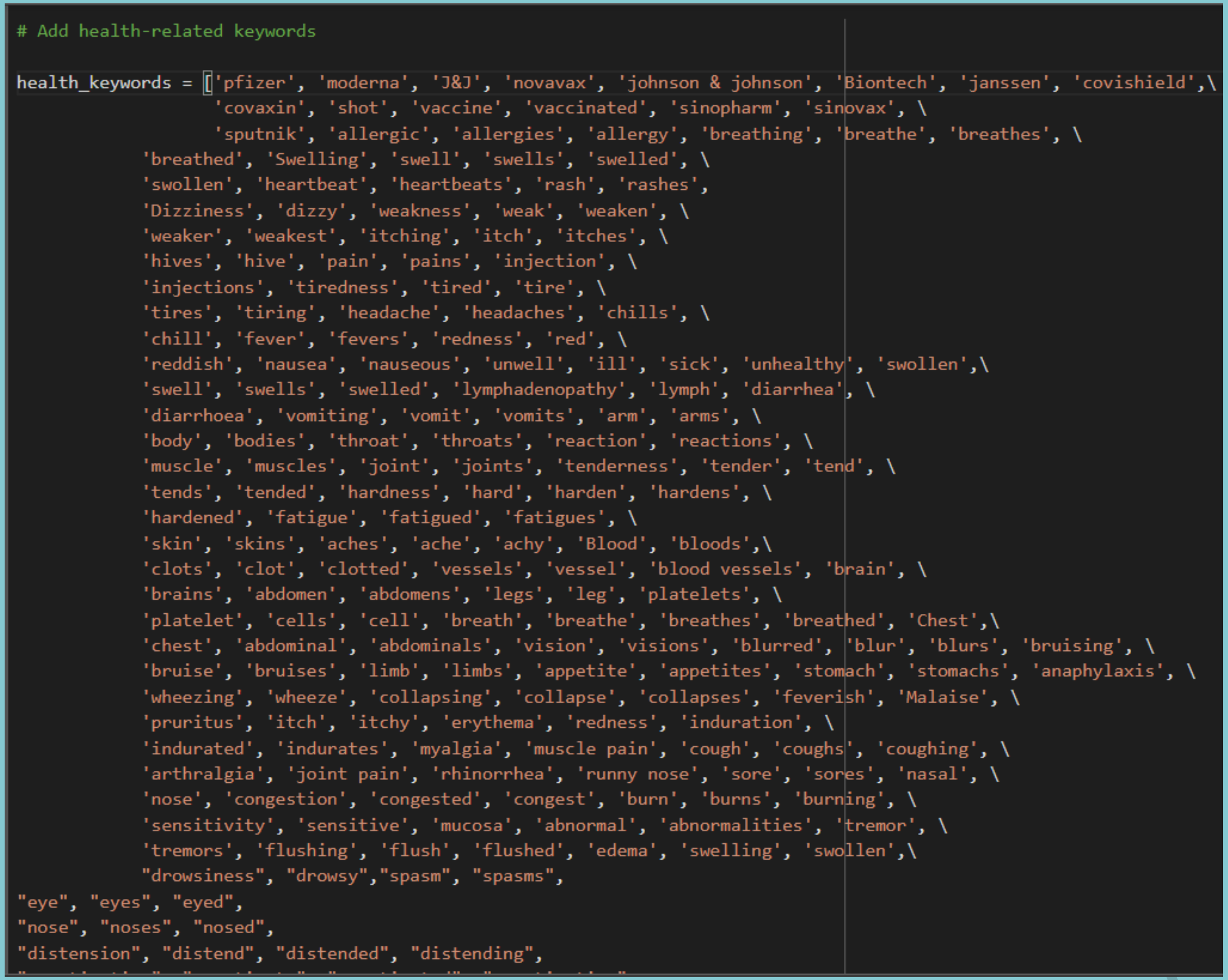
```
# Preprocess the tweets
def preprocess_tweet(tweet):
    # Convert to lowercase
    tweet = tweet.lower()
    # Remove URLs
    tweet = re.sub(r"http\S+", "", tweet)
    # Remove mentions
    tweet = re.sub(r"@[A-Za-z0-9_]+", "", tweet)
    # Remove hashtags
    tweet = re.sub(r"#[A-Za-z0-9_]+", "", tweet)
    # Remove non-alphabetic characters
    tweet = re.sub(r"[^a-zA-Z ]+", "", tweet)
    # Tokenize the tweet
    tokens = nltk.word_tokenize(tweet)
    # Remove stop words
    stop_words = set(stopwords.words('english'))
    tokens = [token for token in tokens if token not in stop_words]
    # Lemmatize the words
    lemmatizer = WordNetLemmatizer()
    tokens = [lemmatizer.lemmatize(token) for token in tokens]
    # Join the tokens back into a single string
    tweet = " ".join(tokens)
    return tweet
```

EXAMPLE OF CLEANED TWEETS



	SinceDate	TweetDate	TweetText	Username	clean_tweet
0	2022-01-21	2022-01-30 09:17:38+00:00	@paisleypeta It's about manipulation, power an...	AndrewF06995245	manipulation power control health already know
1	2022-01-21	2022-01-30 04:35:20+00:00	@AussieVal10 Hey @DanielAndrewsMP you are a ty...	AndrewF06995245	hey tyrant omicron isnt even dangerous except ...
2	2022-01-21	2022-01-30 04:31:39+00:00	@VictorianCHO What a clown. You know this garb...	AndrewF06995245	clown know garbage vaccine last week even read...
3	2022-01-21	2022-01-29 11:32:53+00:00	@newscomauHQ And she feels that way because th...	AndrewF06995245	feel way medium like perpetuating fear around ...
4	2022-01-21	2022-01-28 06:57:34+00:00	@CrabbBrendan Yes it is sad, but your view is ...	AndrewF06995245	yes sad view also horrendously myopic never mi...
...	...	...	...	...	...
3116	2022-01-19	2022-01-22 03:23:18+00:00	@Tormund_G_ Well reading that article certainl...	whittyb45	well reading article certainly reaffirmed deci...
3117	2022-01-19	2022-01-22 01:22:51+00:00	@ajlamesa I hope someone can recommend a good ...	whittyb45	hope someone recommend good therapist womanshe...
3118	2022-01-19	2022-01-19 21:51:55+00:00	@Gorlochy @Gov_NB The fact is they can't, beca...	whittyb45	fact cant small amount hospitalization overwhe...
3119	2022-01-19	2022-01-19 18:06:11+00:00	@misskylie77 @coolguy69666 I got two Pfizer's,...	whittyb45	got two pfizers im hoping novavax booster
3120	2022-01-19	2022-01-19 18:01:54+00:00	@globeandmail Ground beef, cat food, chicken, ...	whittyb45	ground beef cat food chicken mountain dew

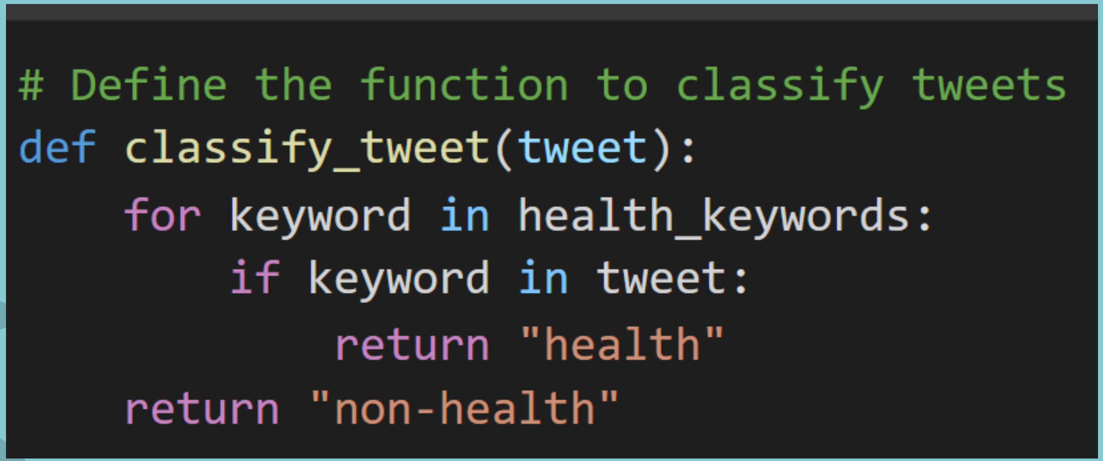
PREPROCESSING



```python
# Add health-related keywords

health_keywords = ['pfizer', 'moderna', 'J&J', 'novavax', 'johnson & johnson', 'Biontech', 'janssen', 'covishield',\
                   'covaxin', 'shot', 'vaccine', 'vaccinated', 'sinopharm', 'sinovax', \
                   'sputnik', 'allergic', 'allergies', 'allergy', 'breathing', 'breathe', 'breathes', \
            'breathed', 'Swelling', 'swell', 'swells', 'swelled', \
            'swollen', 'heartbeat', 'heartbeats', 'rash', 'rashes',
            'Dizziness', 'dizzy', 'weakness', 'weak', 'weaken', \
            'weaker', 'weakest', 'itching', 'itch', 'itches', \
            'hives', 'hive', 'pain', 'pains', 'injection', \
            'injections', 'tiredness', 'tired', 'tire', \
            'tires', 'tiring', 'headache', 'headaches', 'chills', \
            'chill', 'fever', 'fevers', 'redness', 'red', \
            'reddish', 'nausea', 'nauseous', 'unwell', 'ill', 'sick', 'unhealthy', 'swollen',\
            'swell', 'swells', 'swelled', 'lymphadenopathy', 'lymph', 'diarrhea', \
            'diarrhoea', 'vomiting', 'vomit', 'vomits', 'arm', 'arms', \
            'body', 'bodies', 'throat', 'throats', 'reaction', 'reactions', \
            'muscle', 'muscles', 'joint', 'joints', 'tenderness', 'tender', 'tend', \
            'tends', 'tended', 'hardness', 'hard', 'harden', 'hardens', \
            'hardened', 'fatigue', 'fatigued', 'fatigues', \
            'skin', 'skins', 'aches', 'ache', 'achy', 'Blood', 'bloods',\
            'clots', 'clot', 'clotted', 'vessels', 'vessel', 'blood vessels', 'brain', \
            'brains', 'abdomen', 'abdomens', 'legs', 'leg', 'platelets', \
            'platelet', 'cells', 'cell', 'breath', 'breathe', 'breathes', 'breathed', 'Chest',\
            'chest', 'abdominal', 'abdominals', 'vision', 'visions', 'blurred', 'blur', 'blurs', 'bruising', \
            'bruise', 'bruises', 'limb', 'limbs', 'appetite', 'appetites', 'stomach', 'stomachs', 'anaphylaxis', \
            'wheezing', 'wheeze', 'collapsing', 'collapse', 'collapses', 'feverish', 'Malaise', \
            'pruritus', 'itch', 'itchy', 'erythema', 'redness', 'induration', \
            'indurated', 'indurates', 'myalgia', 'muscle pain', 'cough', 'coughs', 'coughing', \
            'arthralgia', 'joint pain', 'rhinorrhea', 'runny nose', 'sore', 'sores', 'nasal', \
            'nose', 'congestion', 'congested', 'congest', 'burn', 'burns', 'burning', \
            'sensitivity', 'sensitive', 'mucosa', 'abnormal', 'abnormalities', 'tremor', \
            'tremors', 'flushing', 'flush', 'flushed', 'edema', 'swelling', 'swollen',\
            "drowsiness", "drowsy","spasm", "spasms",
"eye", "eyes", "eyed",
"nose", "noses", "nosed",
"distension", "distend", "distended", "distending",
```

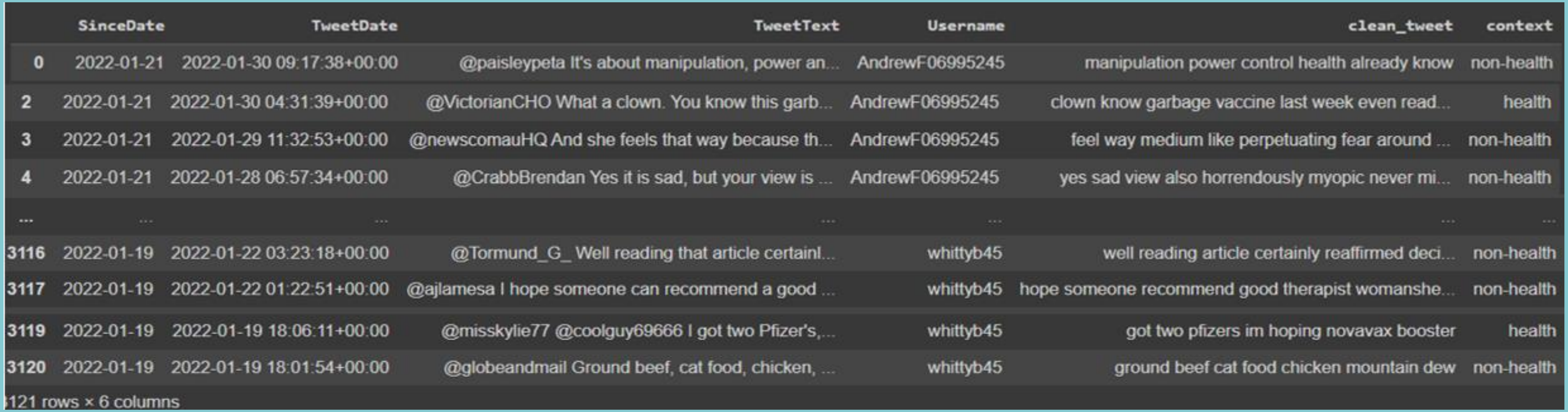
Other approaches which could be considered are: Corex and BERT Cosine Similarity.

1. After cleaning, the tweets were categorized into Health related and Non-health related tweets.
2. This was done using a bag of words. We used ChatGPT to create a list of more than 300 health-related words.
3. Although we lemmatized the tweets, the list contains different forms of the same word (eg: swell, swelling, swollen, swelled)
4. Every tweet was parsed through a function that checked if even a single health-related word was there. If yes, then the tweet was categorized as a health tweet else non-health.



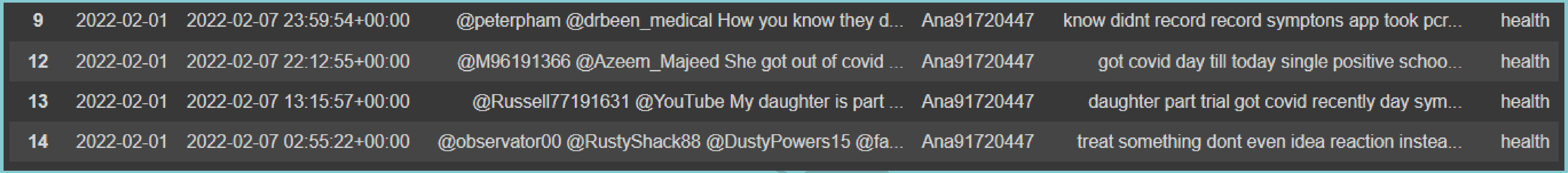
```python
# Define the function to classify tweets
def classify_tweet(tweet):
    for keyword in health_keywords:
        if keyword in tweet:
            return "health"
    return "non-health"
```

TWEETS CATEGORIZED AS “HEALTH” AND “NON-HEALTH”



	SinceDate	TweetDate	TweetText	Username	clean_tweet	context
0	2022-01-21	2022-01-30 09:17:38+00:00	@paisleypeta It's about manipulation, power an...	AndrewF06995245	manipulation power control health already know	non-health
2	2022-01-21	2022-01-30 04:31:39+00:00	@VictorianCHO What a clown. You know this garb...	AndrewF06995245	clown know garbage vaccine last week even read...	health
3	2022-01-21	2022-01-29 11:32:53+00:00	@newscomauHQ And she feels that way because th...	AndrewF06995245	feel way medium like perpetuating fear around ...	non-health
4	2022-01-21	2022-01-28 06:57:34+00:00	@CrabbBrendan Yes it is sad, but your view is ...	AndrewF06995245	yes sad view also horrendously myopic never mi...	non-health
...	...	...	...	...	...	...
3116	2022-01-19	2022-01-22 03:23:18+00:00	@Tormund_G_ Well reading that article certainl...	whittyb45	well reading article certainly reaffirmed deci...	non-health
3117	2022-01-19	2022-01-22 01:22:51+00:00	@ajlamesa I hope someone can recommend a good ...	whittyb45	hope someone recommend good therapist womanshe...	non-health
3119	2022-01-19	2022-01-19 18:06:11+00:00	@misskylie77 @coolguy69666 I got two Pfizer's,...	whittyb45	got two pfizers im hoping novavax booster	health
3120	2022-01-19	2022-01-19 18:01:54+00:00	@globeandmail Ground beef, cat food, chicken, ...	whittyb45	ground beef cat food chicken mountain dew	non-health

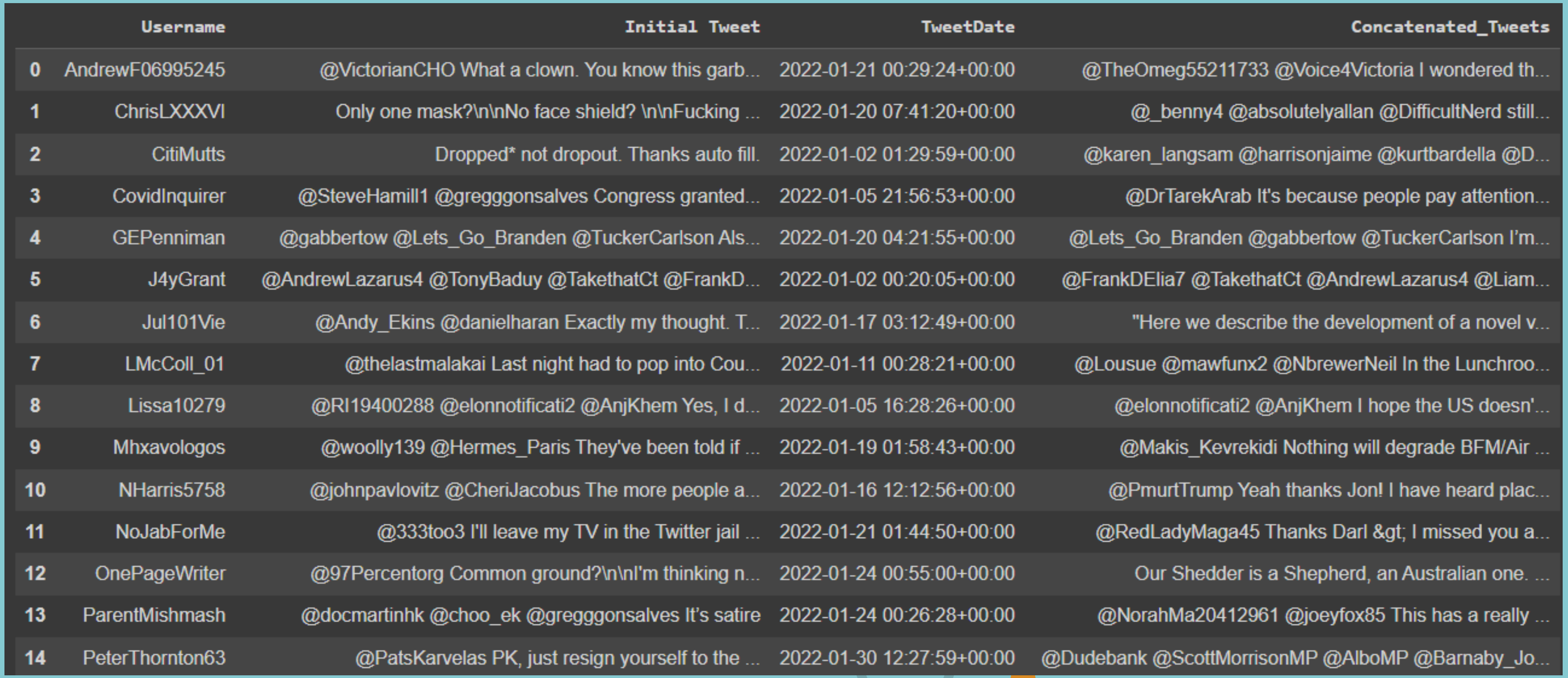
3121 rows × 6 columns



9	2022-02-01	2022-02-07 23:59:54+00:00	@peterpham @drbeen_medical How you know they d...	Ana91720447	know didnt record record symptons app took pcr...	health
12	2022-02-01	2022-02-07 22:12:55+00:00	@M96191366 @Azeem_Majeed She got out of covid ...	Ana91720447	got covid day till today single positive schoo...	health
13	2022-02-01	2022-02-07 13:15:57+00:00	@Russell77191631 @YouTube My daughter is part ...	Ana91720447	daughter part trial got covid recently day sym...	health
14	2022-02-01	2022-02-07 02:55:22+00:00	@observator00 @RustyShack88 @DustyPowers15 @fa...	Ana91720447	treat something dont even idea reaction instea...	health

CONCATENATION OF TWEETS

- After categorization, we dropped all the non-health tweets.
- Next, we concatenated all the Health tweets made by a user. This gave us a flat-file.



	Username	Initial Tweet	TweetDate	Concatenated_Tweets
0	AndrewF06995245	@VictorianCHO What a clown. You know this garb...	2022-01-21 00:29:24+00:00	@TheOmeg55211733 @Voice4Victoria I wondered th...
1	ChrisLXXXVI	Only one mask?\n\nNo face shield? \n\nFucking ...	2022-01-20 07:41:20+00:00	@_benny4 @absolutelyallan @DifficultNerd still...
2	CitiMutts	Dropped* not dropout. Thanks auto fill.	2022-01-02 01:29:59+00:00	@karen_langsam @harrisonjaime @kurtbardella @D...
3	CovidInquirer	@SteveHamill1 @gregggonsalves Congress granted...	2022-01-05 21:56:53+00:00	@DrTarekArab It's because people pay attention...
4	GEPenniman	@gabbertow @Lets_Go_Branden @TuckerCarlson Als...	2022-01-20 04:21:55+00:00	@Lets_Go_Branden @gabbertow @TuckerCarlson I'm...
5	J4yGrant	@AndrewLazarus4 @TonyBaduy @TakethatCt @FrankD...	2022-01-02 00:20:05+00:00	@FrankDElia7 @TakethatCt @AndrewLazarus4 @Liam...
6	Jul101Vie	@Andy_Ekins @danielharan Exactly my thought. T...	2022-01-17 03:12:49+00:00	"Here we describe the development of a novel v...
7	LMcColl_01	@thelastmalakai Last night had to pop into Cou...	2022-01-11 00:28:21+00:00	@Lousue @mawfunx2 @NbrewerNeil In the Lunchroo...
8	Lissa10279	@RI19400288 @elonnotificati2 @AnjKhem Yes, I d...	2022-01-05 16:28:26+00:00	@elonnotificati2 @AnjKhem I hope the US doesn'...
9	Mhxavologos	@woolly139 @Hermes_Paris They've been told if ...	2022-01-19 01:58:43+00:00	@Makis_Kevrekidi Nothing will degrade BFM/Air ...
10	NHarris5758	@johnpavlovitz @CheriJacobus The more people a...	2022-01-16 12:12:56+00:00	@PmurtTrump Yeah thanks Jon! I have heard plac...
11	NoJabForMe	@333too3 I'll leave my TV in the Twitter jail ...	2022-01-21 01:44:50+00:00	@RedLadyMaga45 Thanks Darl &gt; I missed you a...
12	OnePageWriter	@97Percentorg Common ground?\n\nI'm thinking n...	2022-01-24 00:55:00+00:00	Our Shedder is a Shepherd, an Australian one. ...
13	ParentMishmash	@docmartinhk @choo_ek @gregggonsalves It's satire	2022-01-24 00:26:28+00:00	@NorahMa20412961 @joeyfox85 This has a really ...
14	PeterThornton63	@PatsKarvelas PK, just resign yourself to the ...	2022-01-30 12:27:59+00:00	@Dudebank @ScottMorrisonMP @AlboMP @Barnaby_Jo...

WORD COUNT TO FIND DOSE 1, DOSE 2 OR BOOSTER

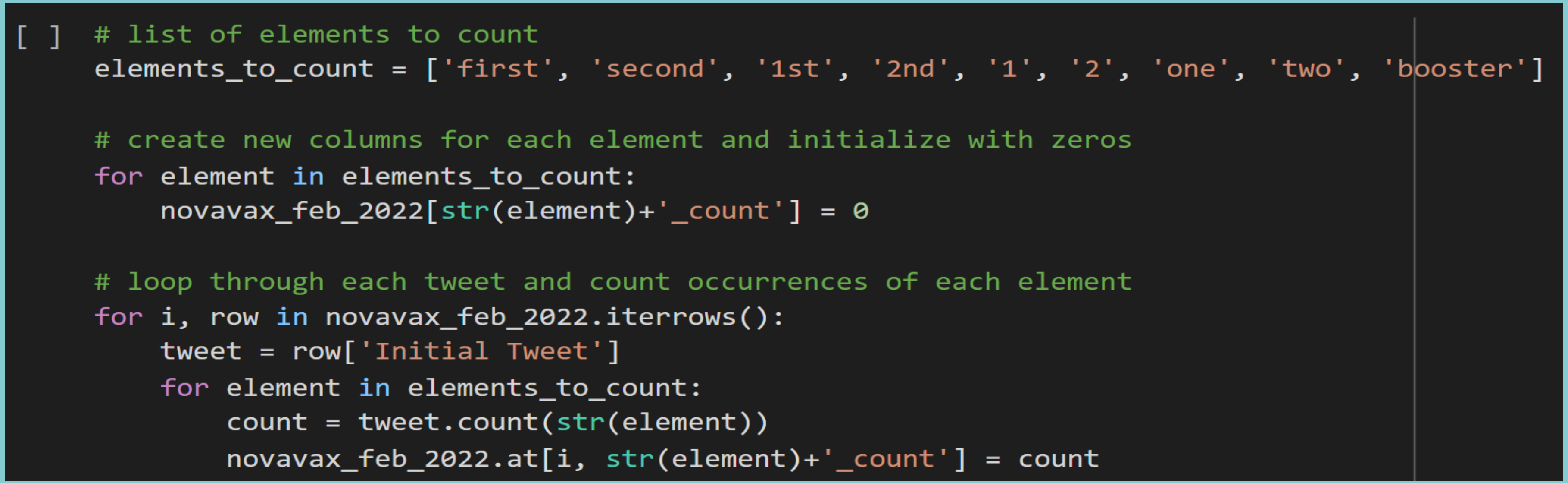
- In order to determine which dose, the user talking about, we did a word count and counted the number of times the following elements were in a tweet:
  ‘First', ’Second', '1st', '2nd', '1', '2', 'one', 'two', 'booster’.

- If a tweet did not contain any of these words, we dropped that tweet.

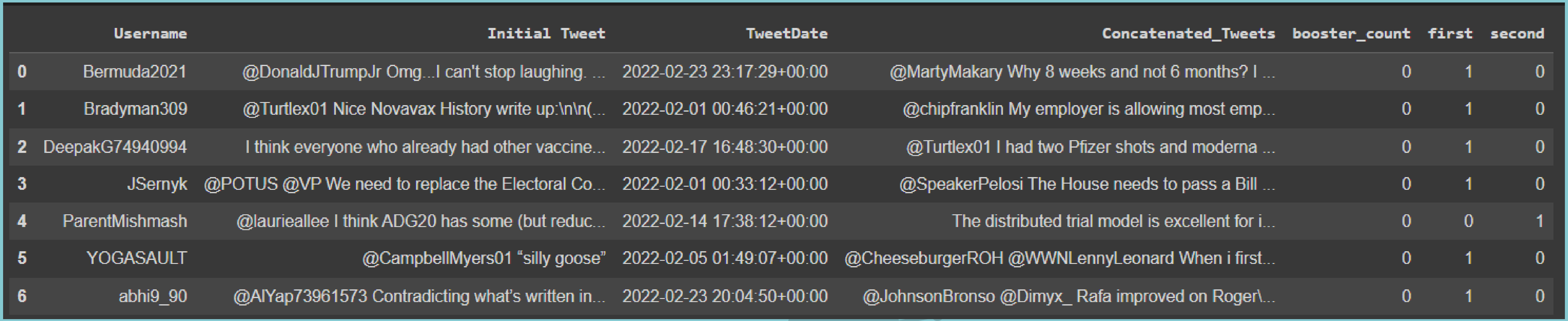
- Count for First, 1st, one, and 1 for each tweet was added and stored in a new column. The same was done for Second, 2nd, two, and 2.

- Count of the booster was kept as it was.

DETERMINING 1ST DOSE, 2ND DOSE OR BOOSTER



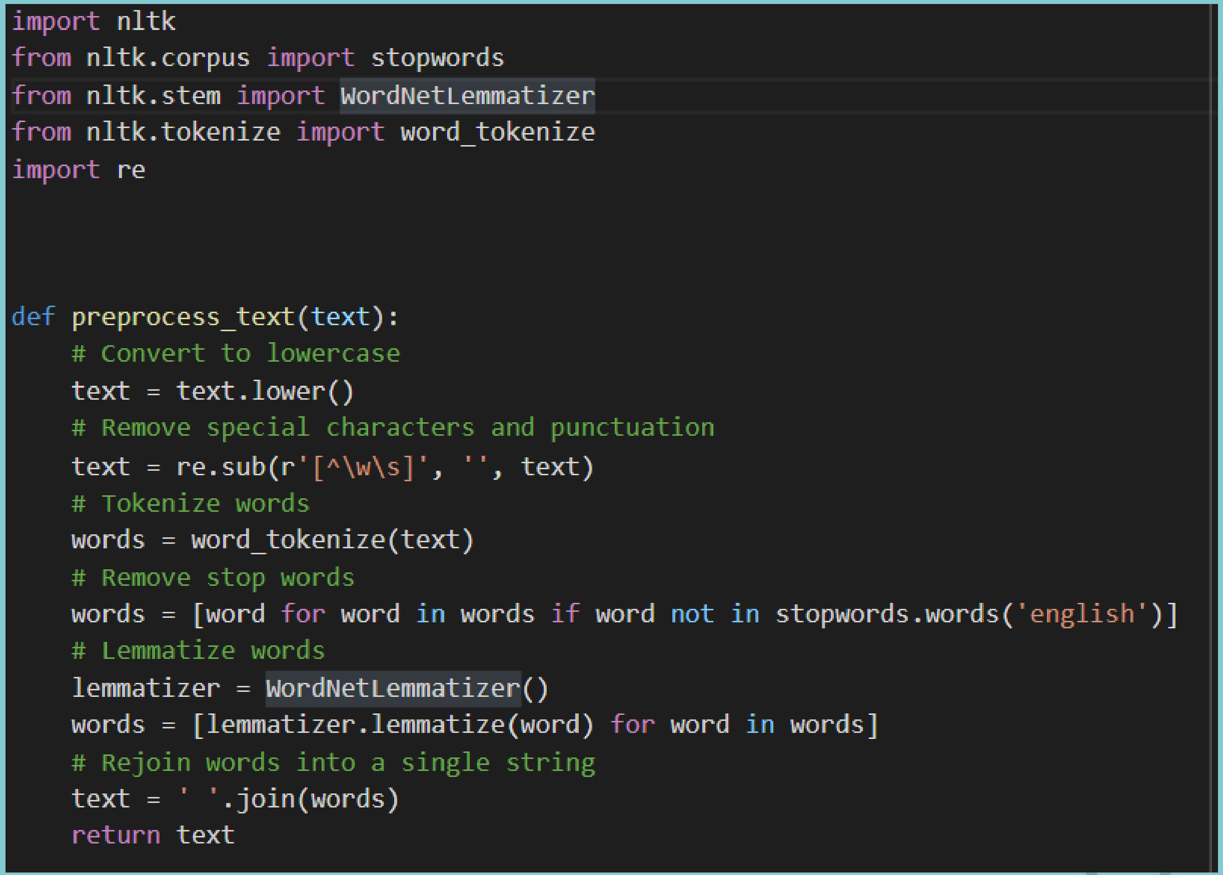
```
[ ]  # list of elements to count
     elements_to_count = ['first', 'second', '1st', '2nd', '1', '2', 'one', 'two', 'booster']

     # create new columns for each element and initialize with zeros
     for element in elements_to_count:
         novavax_feb_2022[str(element)+'_count'] = 0

     # loop through each tweet and count occurrences of each element
     for i, row in novavax_feb_2022.iterrows():
         tweet = row['Initial Tweet']
         for element in elements_to_count:
             count = tweet.count(str(element))
             novavax_feb_2022.at[i, str(element)+'_count'] = count
```



	Username	Initial Tweet	TweetDate	Concatenated_Tweets	booster_count	first	second
0	Bermuda2021	@DonaldJTrumpJr Omg...I can't stop laughing. ...	2022-02-23 23:17:29+00:00	@MartyMakary Why 8 weeks and not 6 months? I ...	0	1	0
1	Bradyman309	@Turtlex01 Nice Novavax History write up:\n\n(...	2022-02-01 00:46:21+00:00	@chipfranklin My employer is allowing most emp...	0	1	0
2	DeepakG74940994	I think everyone who already had other vaccine...	2022-02-17 16:48:30+00:00	@Turtlex01 I had two Pfizer shots and moderna ...	0	1	0
3	JSernyk	@POTUS @VP We need to replace the Electoral Co...	2022-02-01 00:33:12+00:00	@SpeakerPelosi The House needs to pass a Bill ...	0	1	0
4	ParentMishmash	@laurieallee I think ADG20 has some (but reduc...	2022-02-14 17:38:12+00:00	The distributed trial model is excellent for i...	0	0	1
5	YOGASAULT	@CampbellMyers01 “silly goose”	2022-02-05 01:49:07+00:00	@CheeseburgerROH @WWNLennyLeonard When i first...	0	1	0
6	abhi9_90	@AIYap73961573 Contradicting what’s written in...	2022-02-23 20:04:50+00:00	@JohnsonBronso @Dimyx_ Rafa improved on Roger\...	0	1	0

PREPROCESSING - 2

The concatenated tweets were cleaned again using NLTK.



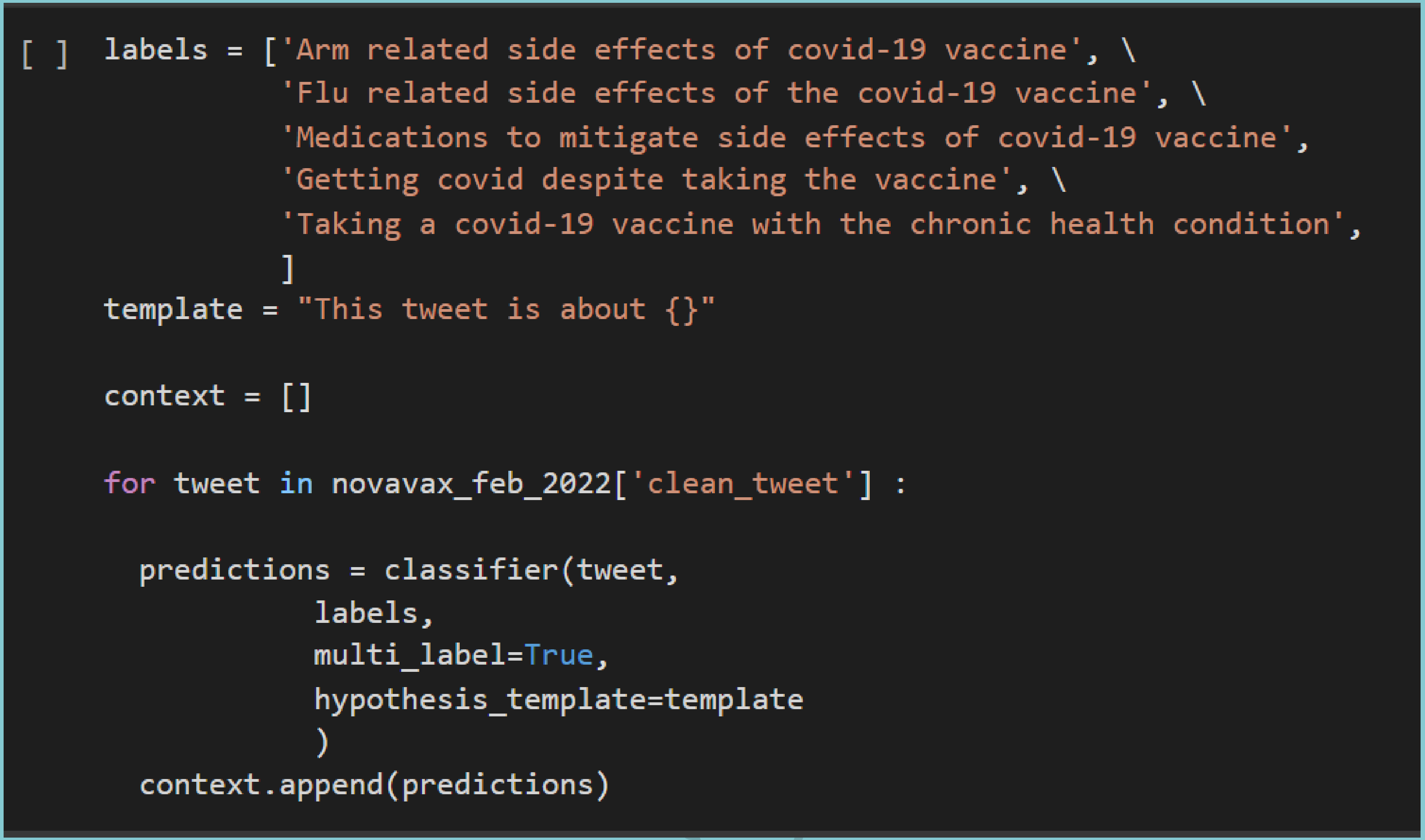
```python
import nltk
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from nltk.tokenize import word_tokenize
import re



def preprocess_text(text):
    # Convert to lowercase
    text = text.lower()
    # Remove special characters and punctuation
    text = re.sub(r'[^\w\s]', '', text)
    # Tokenize words
    words = word_tokenize(text)
    # Remove stop words
    words = [word for word in words if word not in stopwords.words('english')]
    # Lemmatize words
    lemmatizer = WordNetLemmatizer()
    words = [lemmatizer.lemmatize(word) for word in words]
    # Rejoin words into a single string
    text = ' '.join(words)
    return text
```

No of Distinct Users at this stage were **4388**

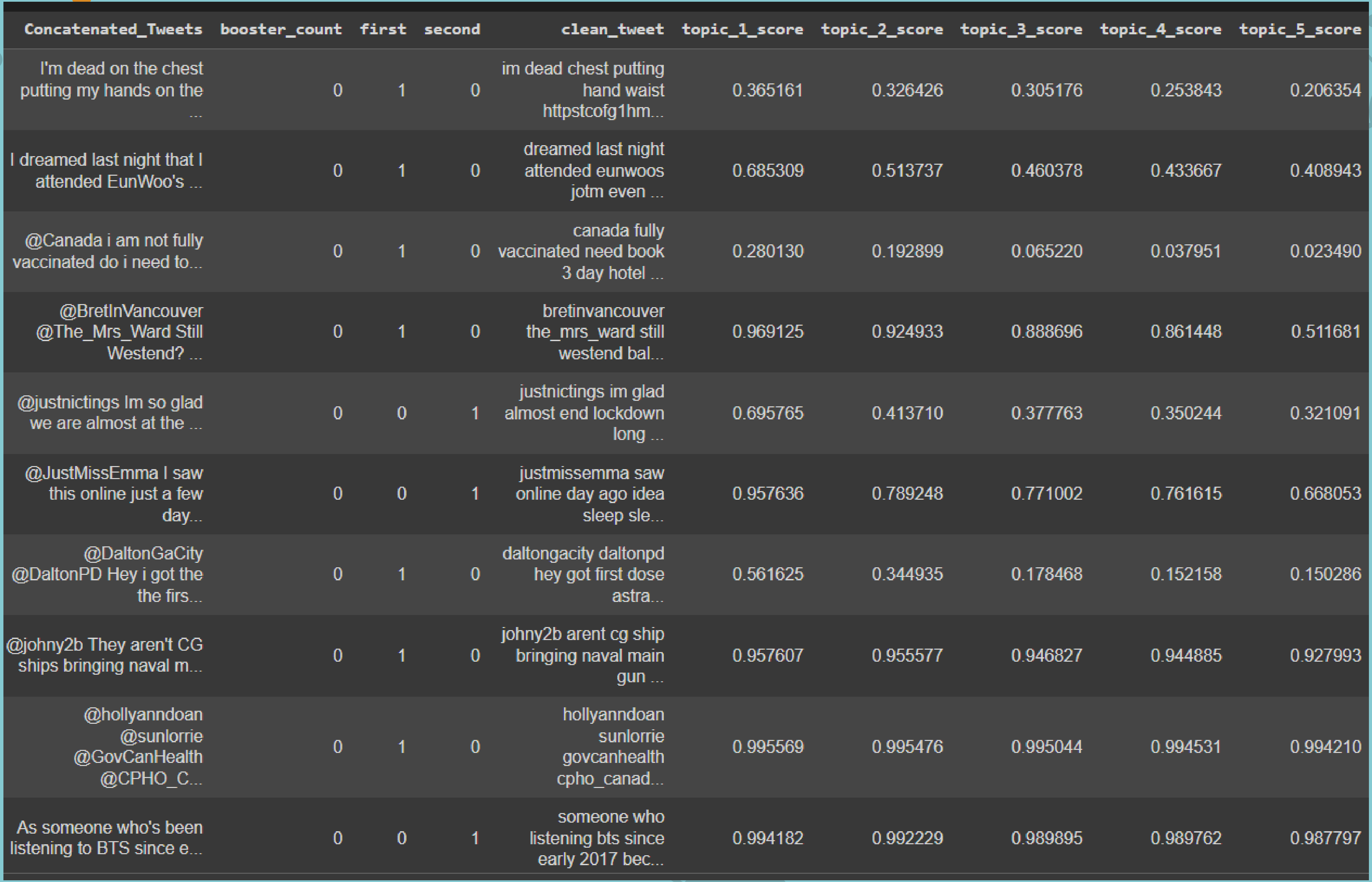
ZERO-SHOT CLASSIFIER

- To implement the zero-shot classifier, we defined five categories:
  - Arm pain related
  - Flu-cold related
  - Underlying chronic condition related
  - Any medication consumed after taking the vaccine
  - Getting COVID despite taking the vaccine or booster.

- Each tweet was then analyzed by the classifier, and a score was assigned for each category based on how well the tweet fit into that category.



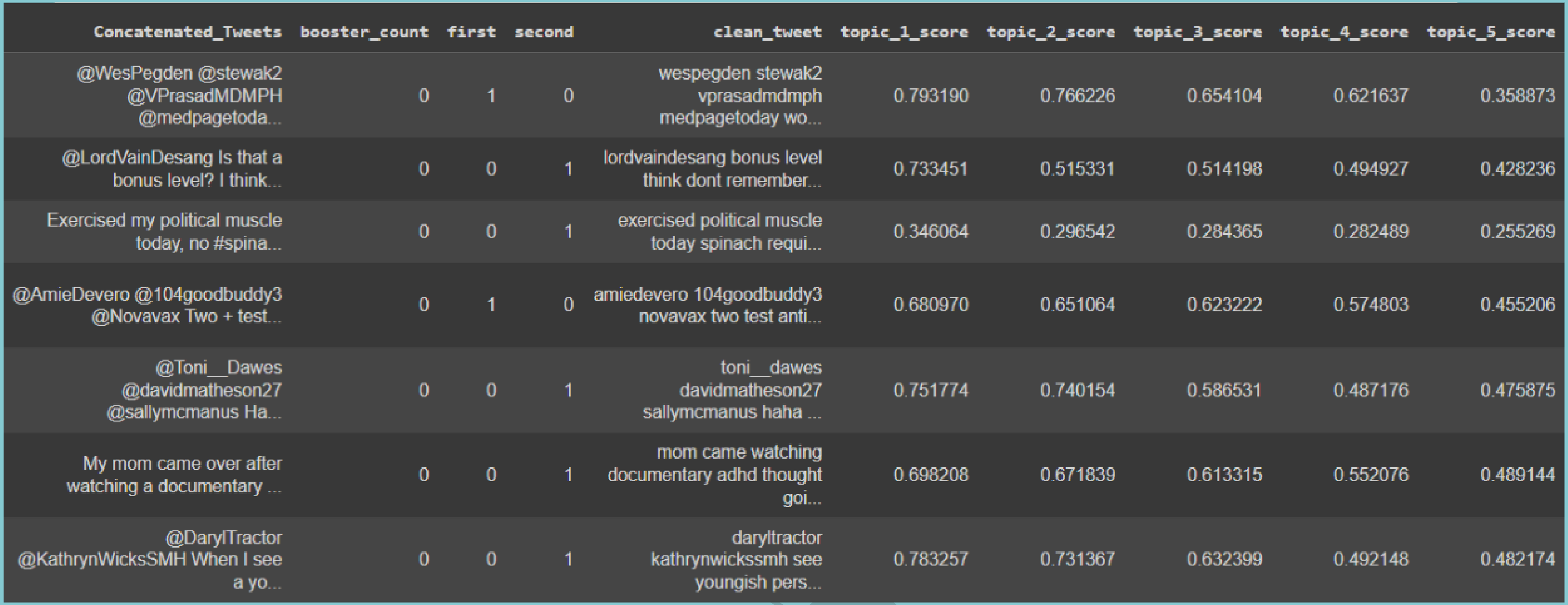
```
labels = ['Arm related side effects of covid-19 vaccine', \
          'Flu related side effects of the covid-19 vaccine', \
          'Medications to mitigate side effects of covid-19 vaccine',
          'Getting covid despite taking the vaccine', \
          'Taking a covid-19 vaccine with the chronic health condition',
          ]
template = "This tweet is about {}"

context = []

for tweet in novavax_feb_2022['clean_tweet'] :

  predictions = classifier(tweet,
            labels,
            multi_label=True,
            hypothesis_template=template
            )
  context.append(predictions)
```



Concatenated_Tweets	booster_count	first	second	clean_tweet	topic_1_score	topic_2_score	topic_3_score	topic_4_score	topic_5_score
I'm dead on the chest putting my hands on the ...	0	1	0	im dead chest putting hand waist httpstcofg1hm...	0.365161	0.326426	0.305176	0.253843	0.206354
I dreamed last night that I attended EunWoo's ...	0	1	0	dreamed last night attended eunwoos jotm even ...	0.685309	0.513737	0.460378	0.433667	0.408943
@Canada i am not fully vaccinated do i need to...	0	1	0	canada fully vaccinated need book 3 day hotel ...	0.280130	0.192899	0.065220	0.037951	0.023490
@BretInVancouver @The_Mrs_Ward Still Westend? ...	0	1	0	bretinvancouver the_mrs_ward still westend bal...	0.969125	0.924933	0.888696	0.861448	0.511681
@justnictings Im so glad we are almost at the ...	0	0	1	justnictings im glad almost end lockdown long ...	0.695765	0.413710	0.377763	0.350244	0.321091
@JustMissEmma I saw this online just a few day...	0	0	1	justmissemma saw online day ago idea sleep sle...	0.957636	0.789248	0.771002	0.761615	0.668053
@DaltonGaCity @DaltonPD Hey i got the the firs...	0	1	0	daltongacity daltonpd hey got first dose astra...	0.561625	0.344935	0.178468	0.152158	0.150286
@johny2b They aren't CG ships bringing naval m...	0	1	0	johny2b arent cg ship bringing naval main gun ...	0.957607	0.955577	0.946827	0.944885	0.927993
@hollyanndoan @sunlorrie @GovCanHealth @CPHO_C...	0	1	0	hollyanndoan sunlorrie govcanhealth cpho_canad...	0.995569	0.995476	0.995044	0.994531	0.994210
As someone who's been listening to BTS since e...	0	0	1	someone who listening bts since early 2017 bec...	0.994182	0.992229	0.989895	0.989762	0.987797

DROPPING IRRELEVANT TWEETS

We dropped those tweets where all 5 topic scores were either greater than 0.8 or less than 0.2



Concatenated_Tweets	booster_count	first	second	clean_tweet	topic_1_score	topic_2_score	topic_3_score	topic_4_score	topic_5_score
@WesPegden @stewak2 @VPrasadMDMPH @medpagetoda...	0	1	0	wespegden stewak2 vprasadmdmph medpagetoday wo...	0.793190	0.766226	0.654104	0.621637	0.358873
@LordVainDesang Is that a bonus level? I think...	0	0	1	lordvaindesang bonus level think dont remember...	0.733451	0.515331	0.514198	0.494927	0.428236
Exercised my political muscle today, no #spina...	0	0	1	exercised political muscle today spinach requi...	0.346064	0.296542	0.284365	0.282489	0.255269
@AmieDevero @104goodbuddy3 @Novavax Two + test...	0	1	0	amiedevero 104goodbuddy3 novavax two test anti...	0.680970	0.651064	0.623222	0.574803	0.455206
@Toni__Dawes @davidmatheson27 @sallymcmanus Ha...	0	0	1	toni__dawes davidmatheson27 sallymcmanus haha ...	0.751774	0.740154	0.586531	0.487176	0.475875
My mom came over after watching a documentary ...	0	0	1	mom came watching documentary adhd thought goi...	0.698208	0.671839	0.613315	0.552076	0.489144
@DarylTractor @KathrynWicksSMH When I see a yo...	0	0	1	daryltractor kathrynwickssmh see youngish pers...	0.783257	0.731367	0.632399	0.492148	0.482174

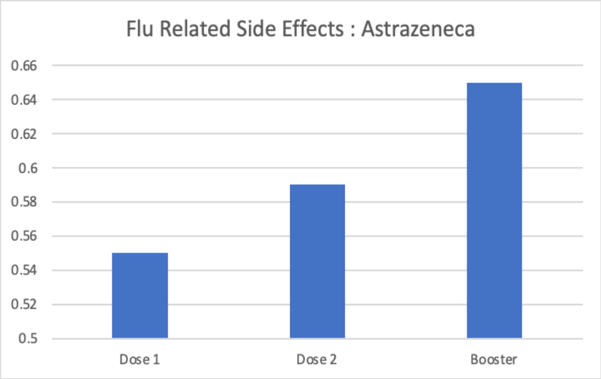
RESULTS

	Pfizer Data : 2022			Astrazeneca Data : 2021 & 2022			Johnson & Johnson Data : 2021 & 2022			Novavax Data : 2021 & 2022		
	Dose 1 (876)	Dose 2 (402)	Booster (331)	Dose 1 (911)	Dose 2 (563)	Booster (187)	Dose 1 (442)	Dose 2 (307)	Booster (129)	Dose 1 (204)	Dose 2 (128)	Booster (0)
Arm Related Side Effects	0.67	0.69	0.72	0.64	0.69	0.70	0.68	0.72	0.71	0.75	0.68	-
Flu Related Side Effects	0.67	0.61	0.69	0.55	0.59	0.65	0.61	0.58	0.67	0.70	0.62	-
Medications to Mitigate Side Effects	0.49	0.53	0.59	0.48	0.51	0.55	0.55	0.54	0.53	0.60	0.57	-
Getting Covid-19 After Taking the Vaccine	0.40	0.44	0.44	0.41	0.42	0.43	0.50	0.49	0.49	0.56	0.49	-
Taking a Covid-19 Vaccine with a Pre-Existing Chronic Health Condition	0.32	0.34	0.36	0.32	0.31	0.30	0.36	0.39	0.36	0.37	0.30	-

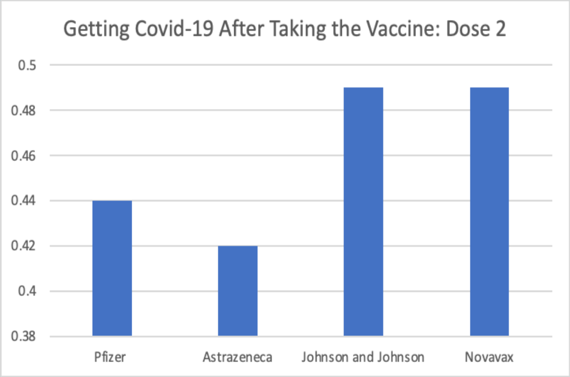
FINDINGS

1. Arm-related side effects appear to be higher for Novavax for dose 1 and dose 2.
2. Pfizer had significantly higher topic scores for flu-related side effects as compared to other vaccines.
3. The booster dose had higher topic scores for flu-related side effects as opposed to dose 1 and dose 2 across all vaccines.
4. Novavax had a significantly higher topic score for "Getting covid after taking the vaccine" as opposed to other vaccines.
5. The topic scores for "Taking a Covid-19 Vaccine with a Pre-Existing Chronic Health Condition" remained almost comparable across all vaccines and doses.
6. An ANOVA was also conducted to check if these differences were statistically significant. ANOVA proved that these differences are not statistically significant and hence, further tests and analysis may be needed to affirm the same.

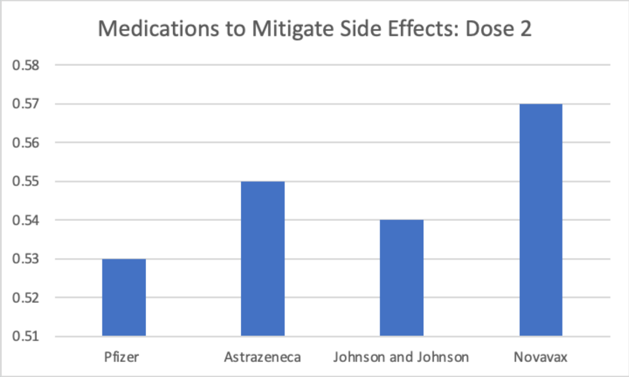
INTERESTING INSIGHTS


Flu Related Side Effects : Astrazeneca

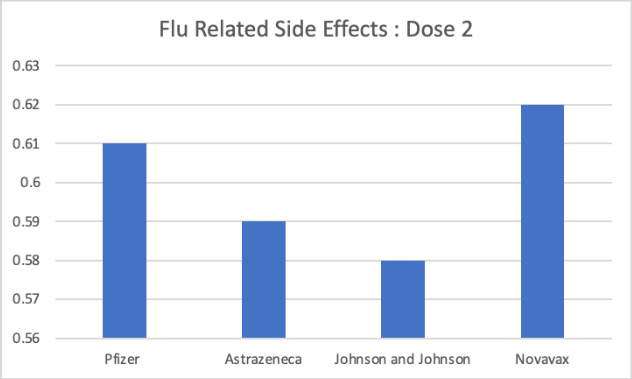
Graph 1


Getting Covid-19 After Taking the Vaccine: Dose 2

Graph 2


Medications to Mitigate Side Effects: Dose 2

Graph 3


Flu Related Side Effects : Dose 2

Graph 4

Graph 1: AstraZeneca Booster reporting High flu-related symptoms while in other vaccines Dose 1 had the higher value.
Graph 2: Novavax and J&J show a higher probability of getting covid after taking the 2nd dose.
Graph 3: People who AstraZeneca dose 2 are more likely to take medication to mitigate sight effects.
Graph 4: Novavax and Pfizer show higher flu-related side effects.

SUMMARY

- The COVID-19 vaccine has set a record for being the fastest vaccine ever developed and brought to market. While researchers may take some time to fully comprehend the vaccine's side effects in different populations due to the rapid testing conducted across the three phases, analyzing user responses via tweets is currently the most effective way to study the vaccine's side effects and related topics during the early phases of vaccination.

- This study serves as a valuable starting point for identifying the side effects of the vaccine based on vaccine types and doses. However, it could be further optimized with time and expertise to explore the side effects by age, race, and chronic conditions and gain deeper insights from people's responses.

LIMITATIONS

SCRAPING

Twitter does not allow easy scraping of data. It blocked us multiple times. Twitter API's don't always work and getting access to developer account might take a lot of time

TWEETS CLASSIFICATION

Bag of words might not be the best approach to classify the tweets into health and non health tweets.

COMPUTING POWER

High computational power is required to scrape and analyze the data. We could not scrape data for Moderna and Pfizer(2021)

TIME

Zero-shot classification consumes a lot of time for each vaccine.