

# EC 401

04 May 2022

# An analytic study of evolving public opinion towards the COVID-19 vaccines on online social network (Twitter)

• **Ajit Singh**  
**1801010**

• **Ankit Kumar**  
**1801021**

**Supervisor: Dr. Surajit Panja**



**Dept. of Electronics and Communication Engineering**  
**Indian Institute of Information Technology Guwahati**

# Problem Statement

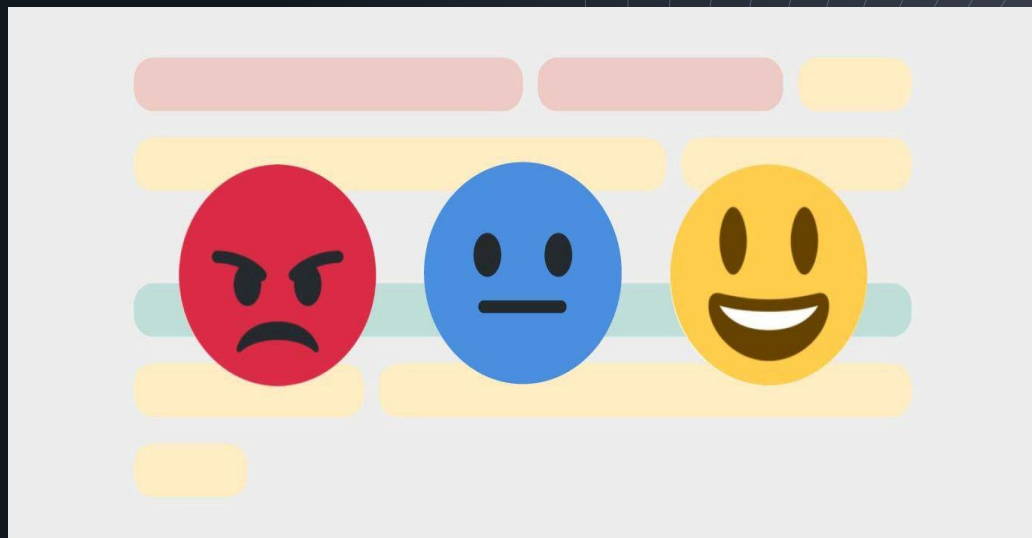
- World Wide Covid 19-Vaccination Programs
- Discussions and debates surrounding vaccines/vaccination programs on Online Social Networks and Discussion Forums (eg: Twitter)
- Polarization of society into two factions-one wherein people have reservation about being vaccinated and voice this on online social media, and the one wherein people advocate being vaccinated.

# Objectives

- Analyzing the people's reaction to covid vaccination over two succinctly different time periods (pre vaccine rollout and pre booster rollout) on the basis of twitter data.
- Examining the Indian public sentiment related to the COVID-19 vaccine in India.
- Analyzing the most frequent words people use in their conversations online (pro and anti Vaccination) to help understand certain communities and their concerns.
- Analyzing the changing public emotions towards covid vaccination over time.
- Understanding the prevalent topics in positive and negative tweets surrounding COVID-19 vaccines.

# Sentiment Analysis

The main task in sentiment analysis is classifying the polarity of a given text at the document, sentence, or feature level whether the expressed opinion in a document, a sentence or a feature is positive, negative, or neutral.



# Ngrams (Bigrams) Analysis

N-grams are basically a set of successive co-occurring words within a given window. The set of character n-grams of a document comprises the substrings of length n of the original text.



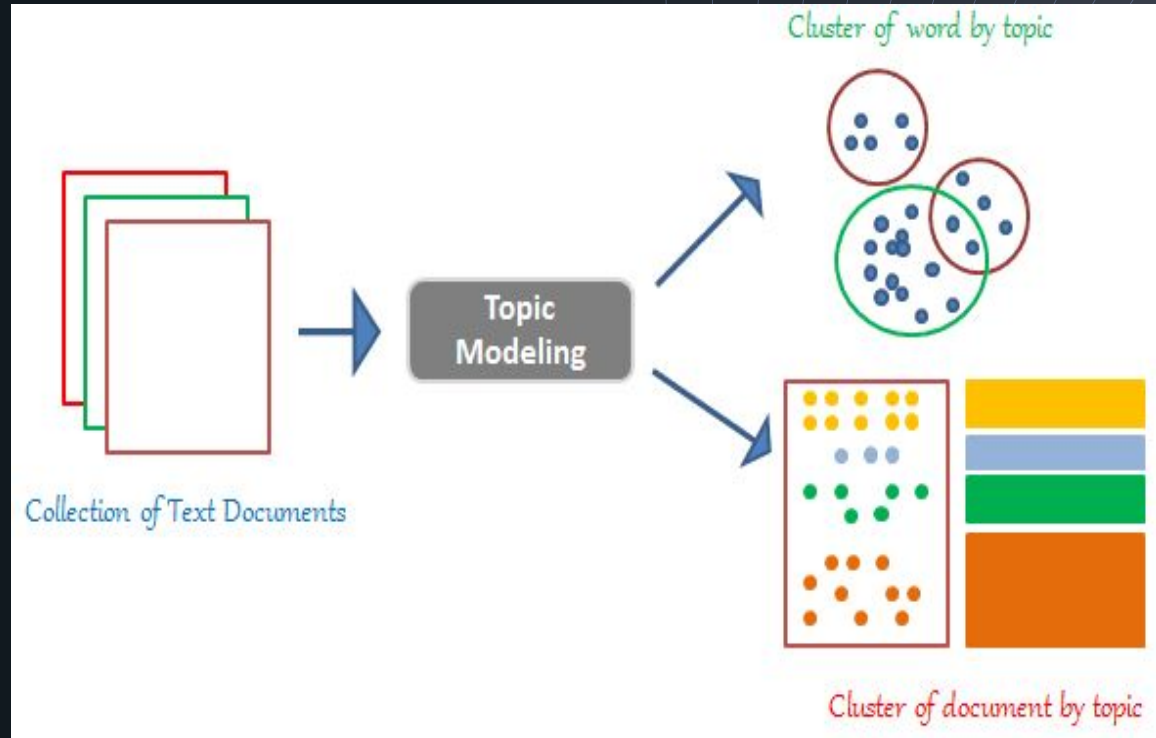


# Emotion Analysis

Emotion analysis is an evolved form of sentiment analysis. It detects and analyzes the well known emotions like Fear, Joy, Surprise, Anger, Sad etc. from any textual data.

# Topic Modeling

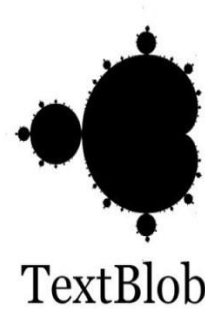
Topic Modeling is a technique to extract the hidden topics from large volumes of text.





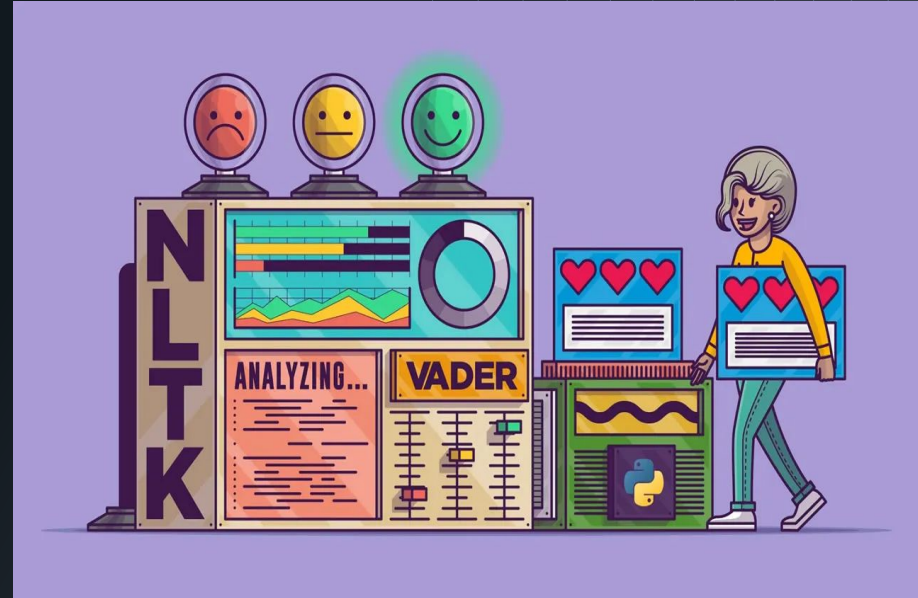
# TextBlob

TextBlob is a lexicon based sentiment analysis method. It is itself based on the NLTK (Natural Language Toolkit) platform, and provides a very good general overview of the data.



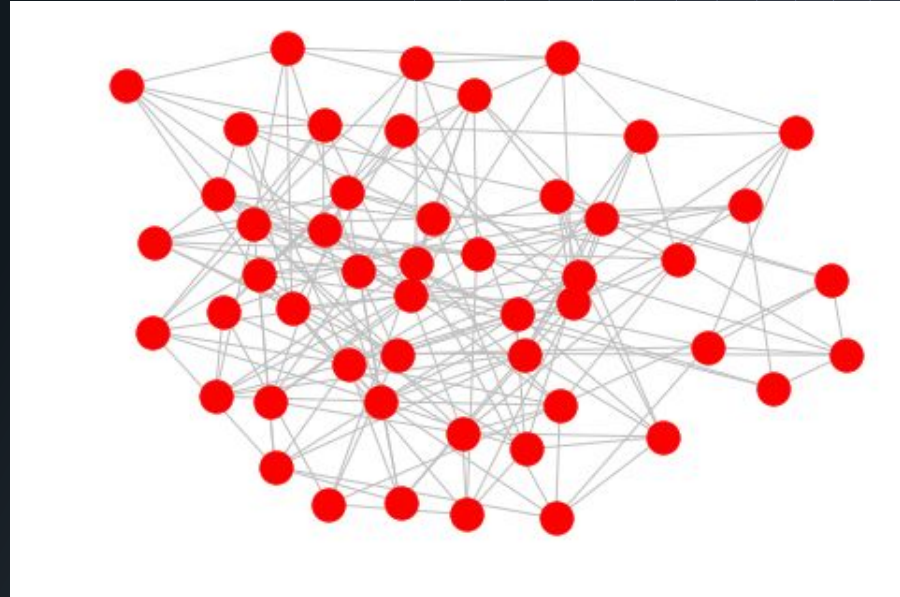
# VADER

VADER( Valence Aware Dictionary and sEntiment Reasoner ) is also a lexicon based sentiment analysis method. VADER has been quite successful when dealing with social media texts, product reviews etc.



# NetworkX

NetworkX is a python library for studying graphs and networks. With NetworkX we can load and store networks in standard and nonstandard data formats, generate many types of random and classic networks, analyze network structure etc



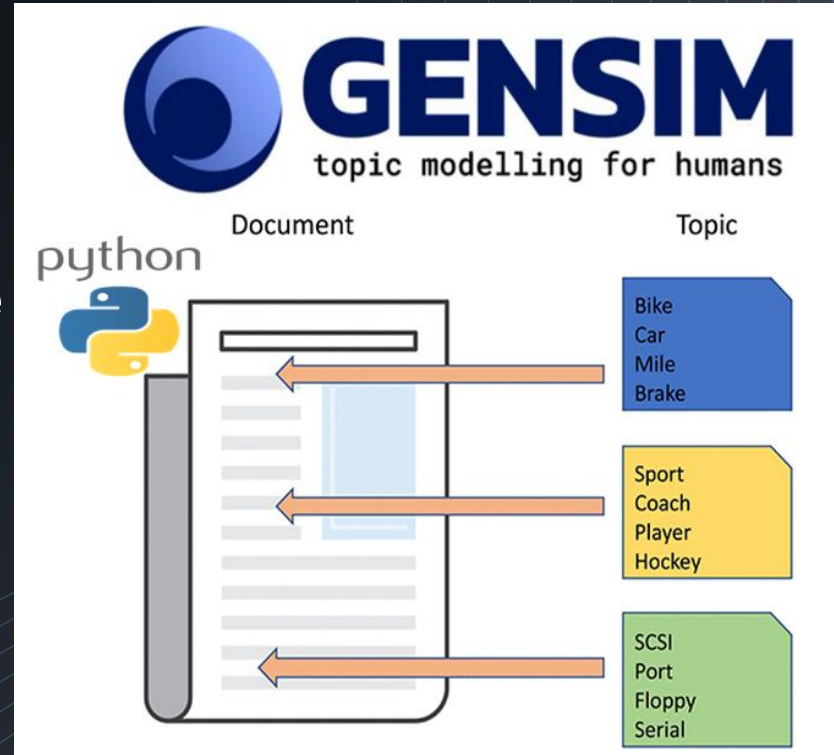
# NRCLex

NRCLex will measure emotional effects from a body of text. It basically finds all eight well known basic emotion categories such as Anger, Fear, Trust, Surprise etc. and two sentiments like Positive and Negative.



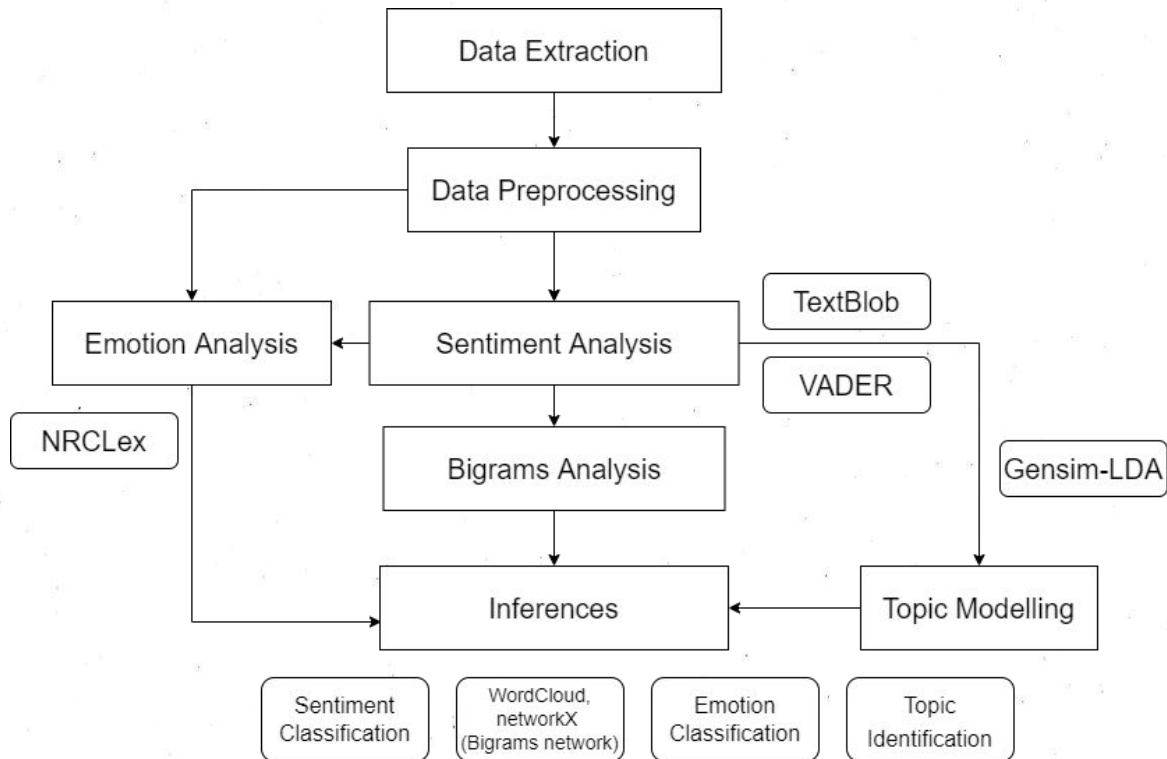
# Topic Modelling using Gensim LDA

LDA discovers the semantic structure of documents by examining statistical co-occurrence patterns within a corpus of training documents.



# Methodology

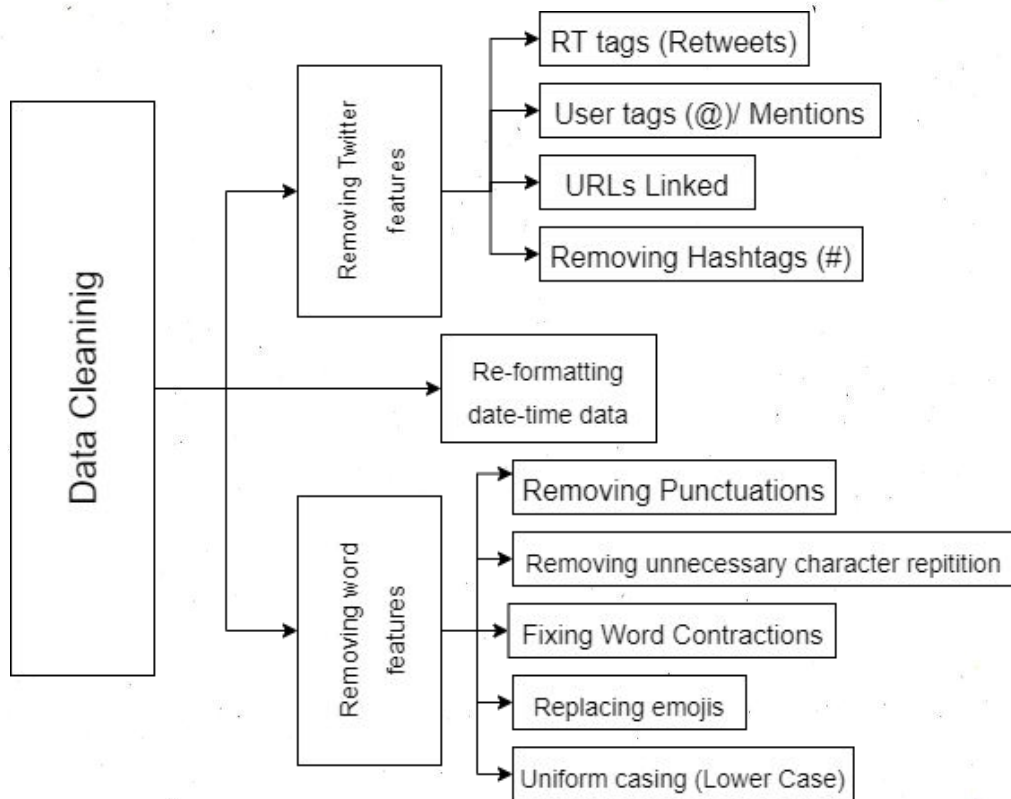
Overall workflow:





# Data Pre-Processing

Preprocessing  
Workflow:



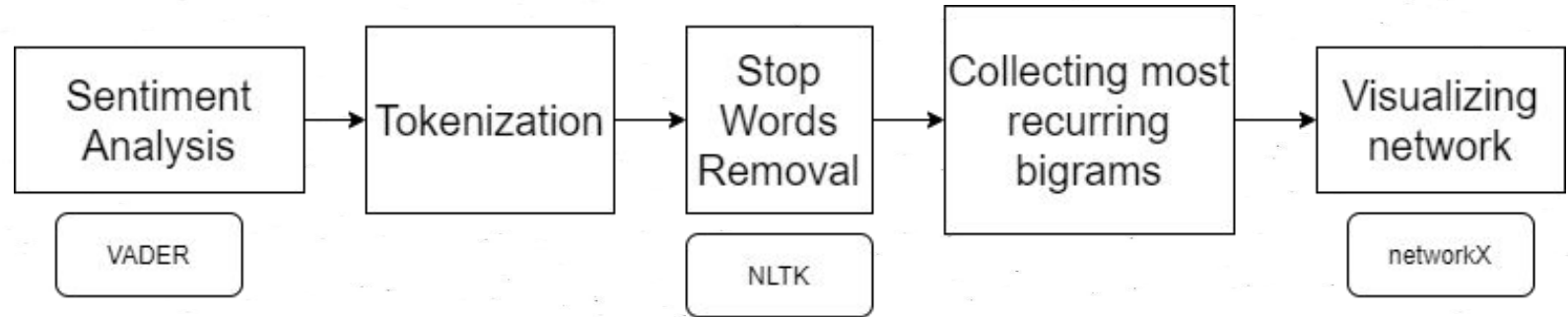
# Classifying Sentiments

Score Range	Sentiment Category
$-0.05 < \text{score} < 0.05$	Neutral
$\text{score} < -0.05$	Negative
$\text{score} > 0.05$	Positive

Table describing our rules to categorize sentiments based on Polarity Score (TextBlob) and Compound Score (VADER)

# N-grams Analysis

Workflow:



```
['This is a list']
```

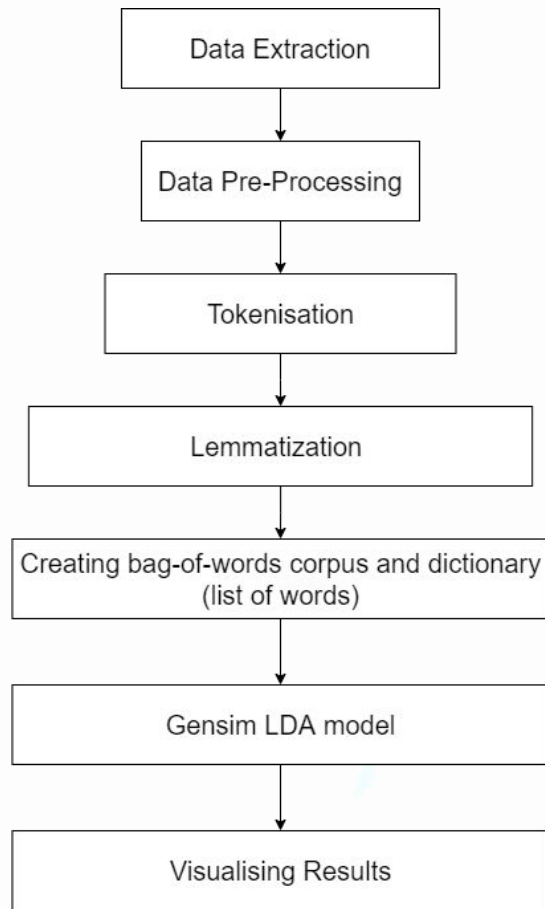
A list containing a full sentence

```
['This', 'is', 'a', 'list', 'of', 'words']
```

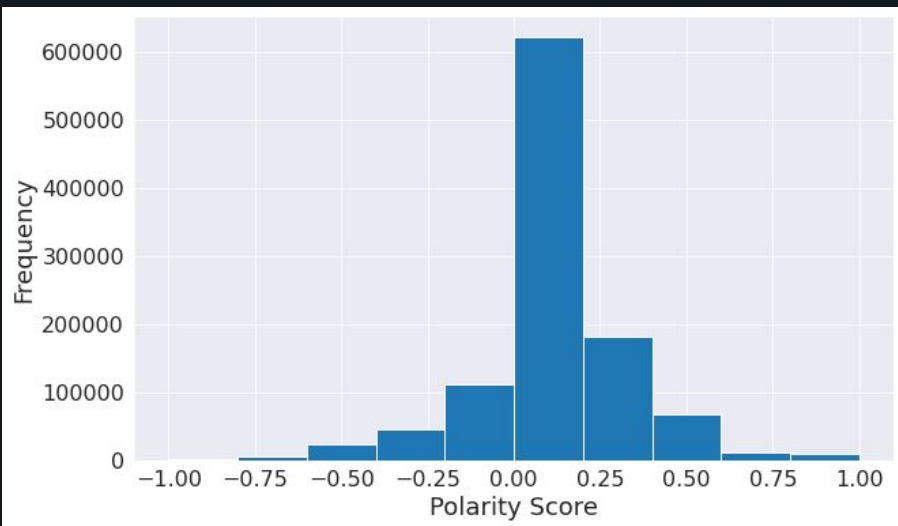
A list containing a set of tokens

# Topic Modelling:

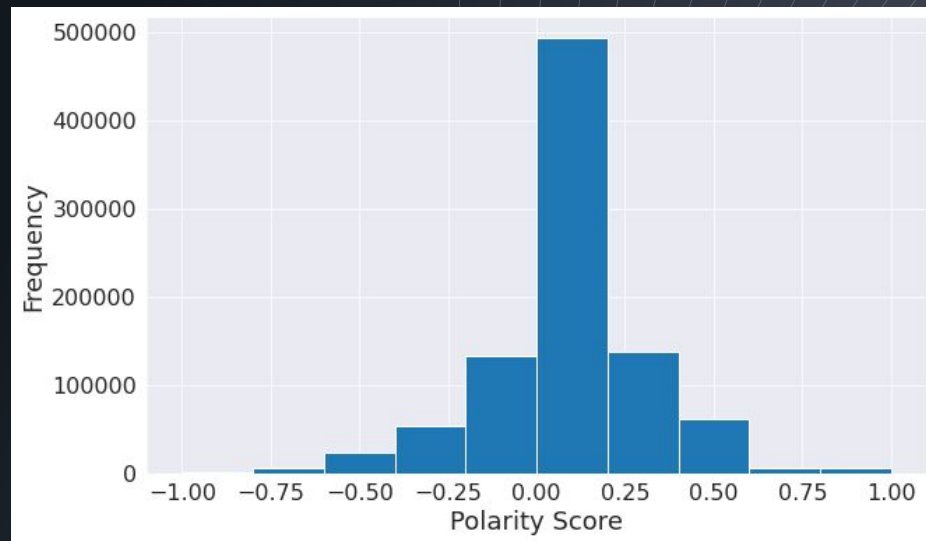
Selected number of topics: **4**



# Evaluation/Observations: Sentiment Analysis

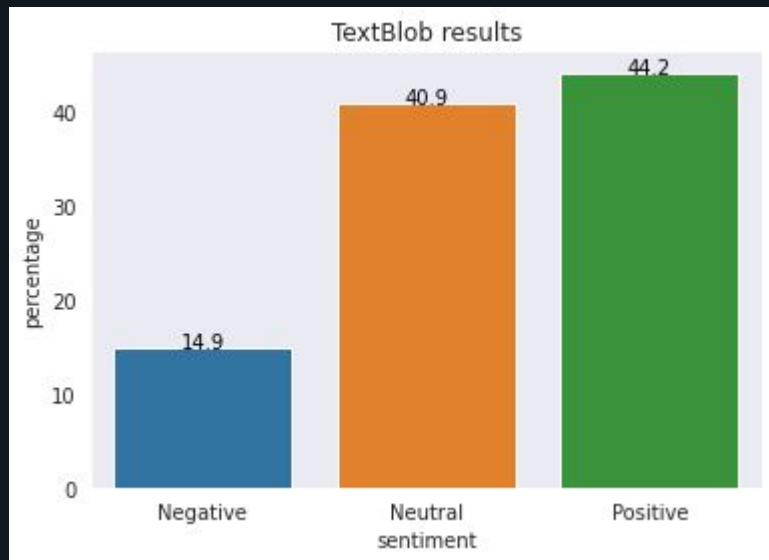


Distribution of Polarity Scores -  
December 2020 (TextBlob)

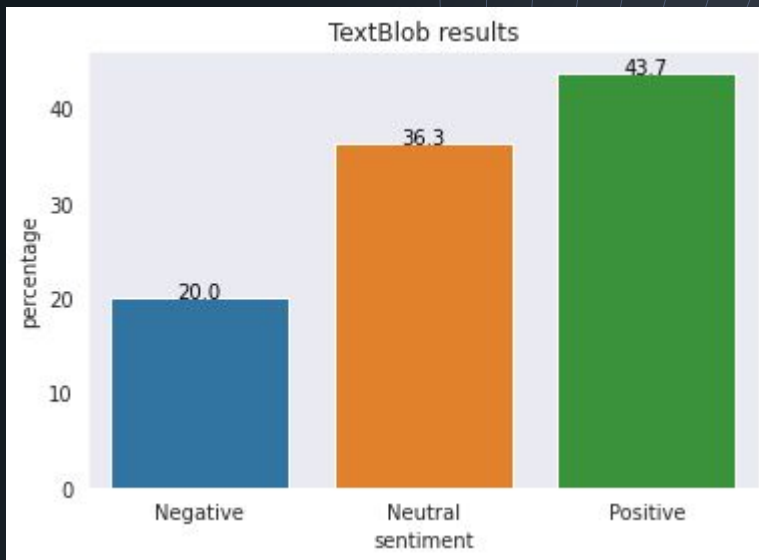


Distribution of Polarity Scores -  
September 2021 (TextBlob)

# Sentiment Analysis



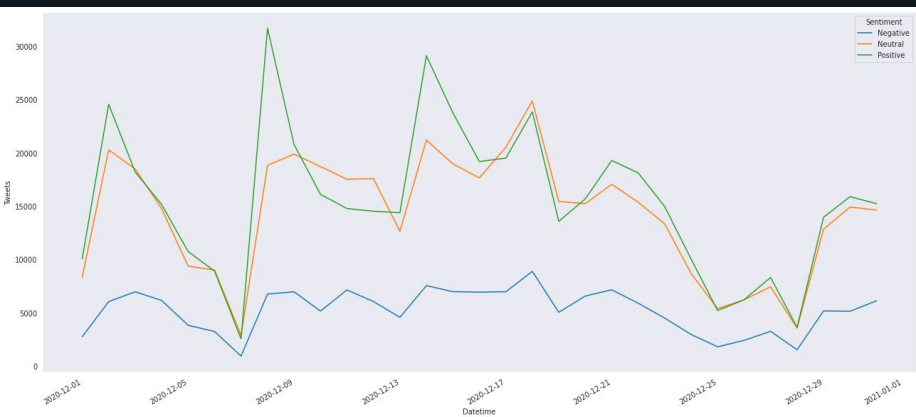
Percentage of Tweets across sentiment categories - December 2020 (TextBlob)



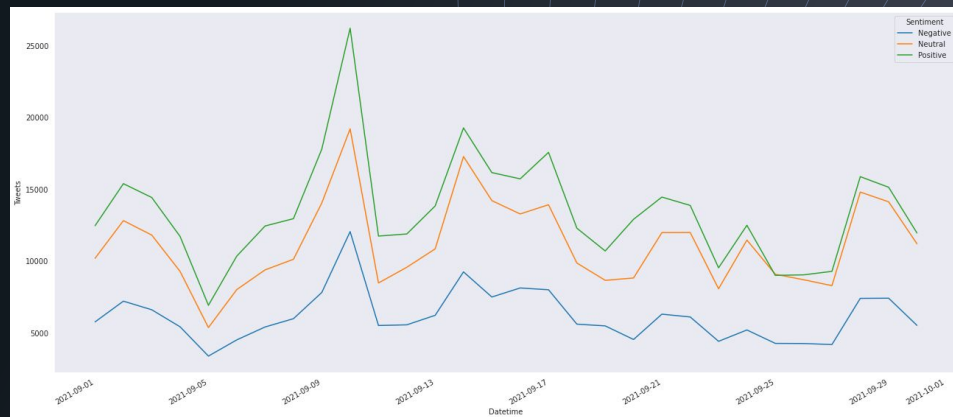
Percentage of Tweets across sentiment categories - September 2021 (TextBlob)



# Sentiment Analysis

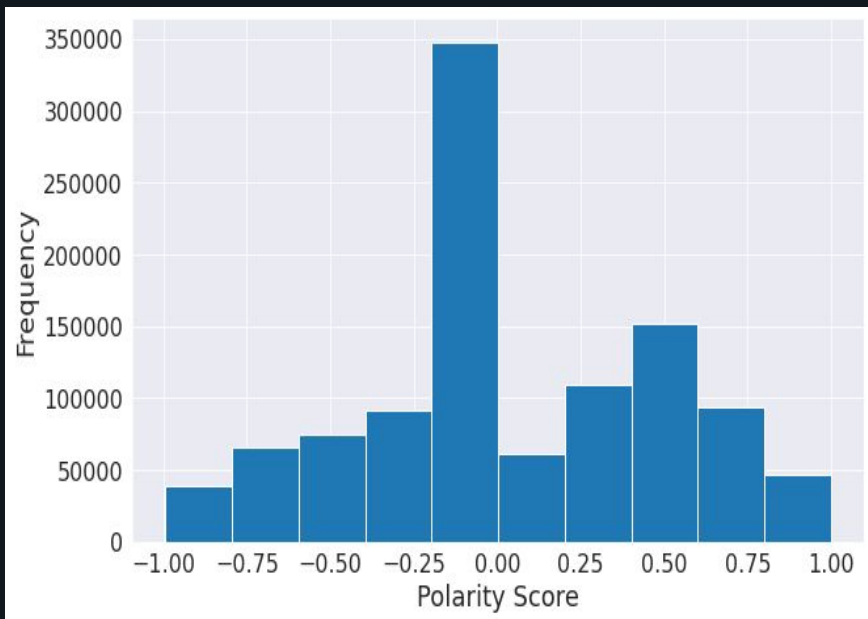


Time Series plot of tweets across the sentiment categories over the month of December (TextBlob)

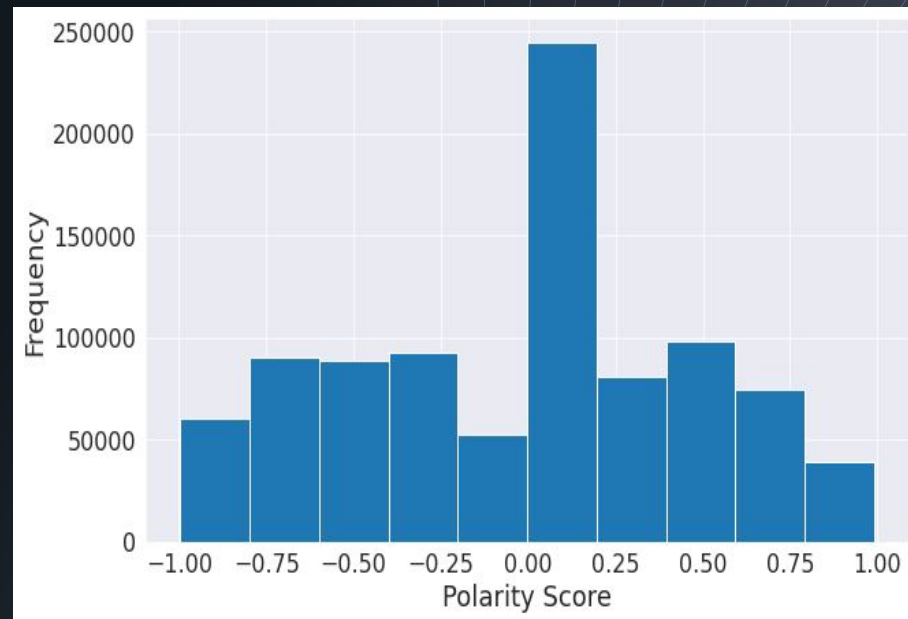


Time Series plot of tweets across the sentiment categories over the month of September (TextBlob)

# Sentiment Analysis

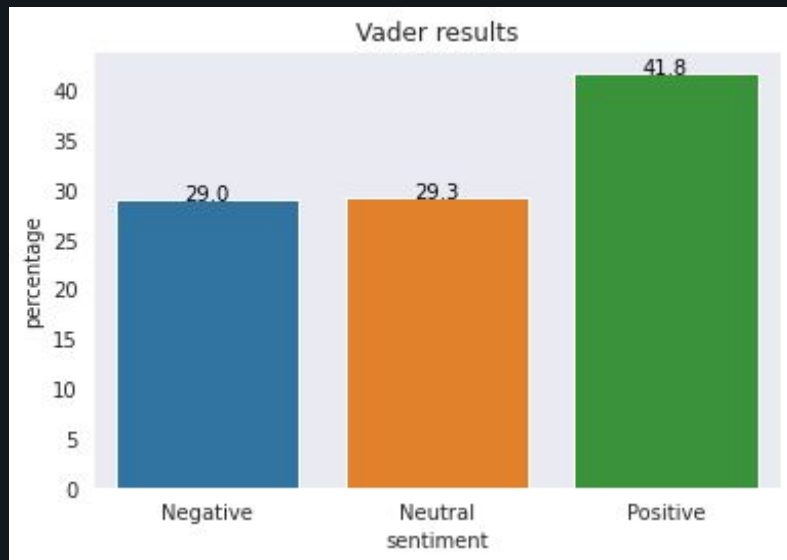


Polarity scores' distribution - December Data (VADER)

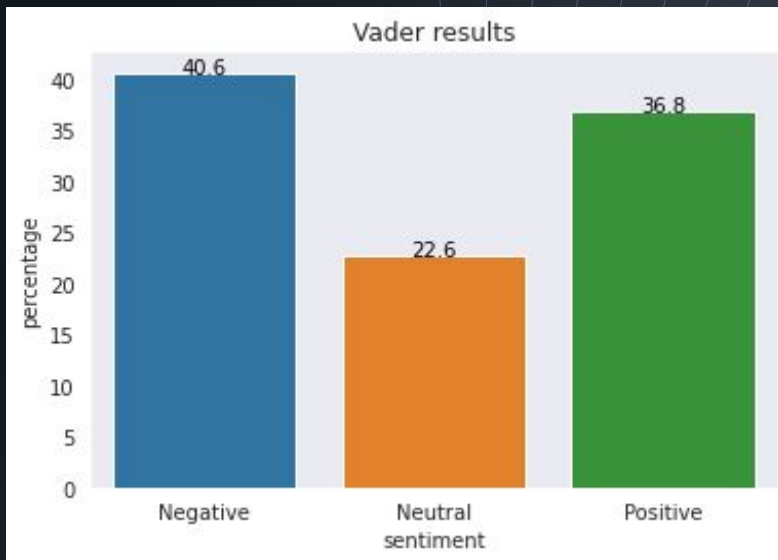


Polarity scores' distribution - September Data (VADER)

# Sentiment Analysis



Percentage of Tweets across sentiment categories - December 2020 (VADER)

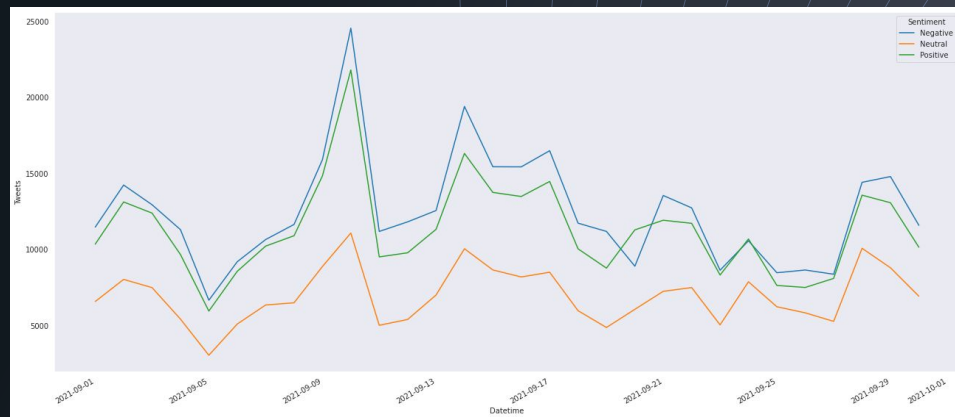


Percentage of Tweets across sentiment categories - September 2021 (VADER)

# Sentiment Analysis



Time Series plot of tweets across the sentiment categories over the month of December (VADER)

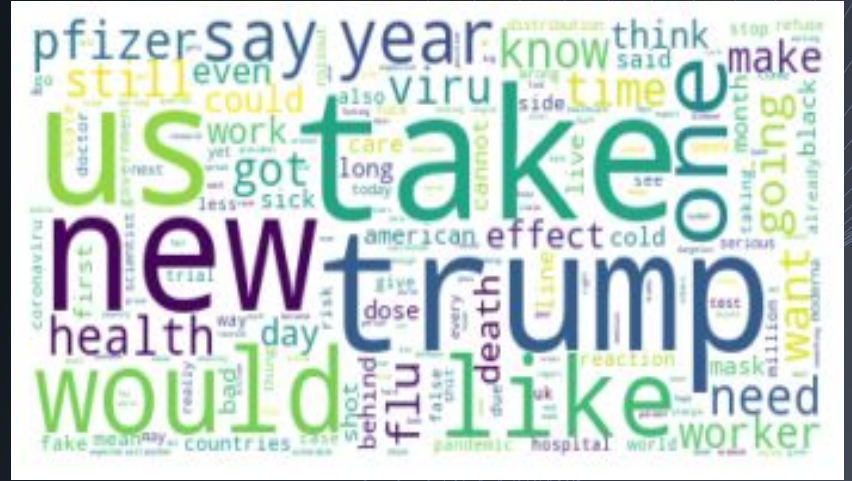


Time Series plot of tweets across the sentiment categories over the month of September (VADER)

# Sentiment Analysis-Word Clouds



Word cloud for positive sentiments  
(December Data, TextBlob)



Word cloud for negative sentiments  
(December Data, TextBlob)



# Sentiment Analysis



### Word cloud for positive sentiments (September Data, TextBlob)



### Word cloud for negative sentiments (September Data, TextBlob)



# Sentiment Analysis



Word cloud for positive sentiments  
(December Data, VADER)



Word cloud for negative sentiments  
(December Data, VADER)

# Sentiment Analysis



Word cloud for positive sentiments  
(September Data, VADER)



Word cloud for negative sentiments  
(September Data, VADER)

# Sentiment Analysis

0	@thehill There are still a lot of uncertainties about Vaccines. The best bet to fight against COVID is to maintain your immune system and follow the safety protocols so you won't get infected even without getting the vaccine.	1.0	bellasirenay
1	@Laird_Admiral @westernsussex @CNOEngland @WorthingJournal @Iloveworthing @CoolTownCrier Awesome Malcolm - see you on the #jabbingline 🍑 #CovidVaccine	1.0	maggiedavies
2	😞 Hmm. Sounds like the vaccine is the perfect solution to population control, eh Moderna's Bill Gates???\n\nABC Fertility Report: 'Men May Want to Consider Freezing Sperm' Before COVID Vaccine <a href="https://t.co/k5zjaF1WMa">https://t.co/k5zjaF1WMa</a> via @RaheemKassam	1.0	egatimra3
3	The best! \nStella Parton, singer and sister of Dolly, lashes out at politicians in viral tweet over COVID-19 vaccine rollout <a href="https://t.co/mUJqNLn5YU">https://t.co/mUJqNLn5YU</a>	1.0	415holgate
4	Checkout @DanLarremore in the @DenverChannel on which COVID vaccine distribution strategies work best:\n\n <a href="https://t.co/LR7aowwk9d">https://t.co/LR7aowwk9d</a>	1.0	TheRhoten
5	Dr. Anthony Fauci has received the Moderna COVID-19 vaccine, 'the best hope against the pandemic' <a href="https://t.co/pjHzi83wKL">https://t.co/pjHzi83wKL</a>	1.0	USATODAY
6	@g1 the president of ur country also trivializes science and says'I'm not going to get this vaccine bcs I got the best vaccine,the 'virus' without side effects' #covid +	1.0	MariaSilviaPP_
7	Today has turned out to being probably the best day of 2020! 1) Biden accepts the Presidential nomination & is addressing the people in 30 minutes on tv.\n2) Barr has resigned. 3) COVID-19 vaccine has started to be given\n\n#Biden2020	1.0	Momma_Carmen
8	Excellent thread, explaining Covid vaccine virology with lots of swearing. <a href="https://t.co/hP9w1hsVGB">https://t.co/hP9w1hsVGB</a>	1.0	knotgirl1
9	Awesome poster by @sairamaliha #CovidVaccine <a href="https://t.co/aaLK9T210c">https://t.co/aaLK9T210c</a>	1.0	christinakouts3



# Sentiment Analysis

	Text	TBScore	Username
0	'Truly, Truly Evil': Tucker Carlson Slammed For Sowing Doubt About COVID-19 Vaccines - <a href="https://t.co/UWrP1NCtTm">https://t.co/UWrP1NCtTm</a> <a href="https://t.co/HPHlB5PCqw">https://t.co/HPHlB5PCqw</a>	-1.0	southasiansnews
1	Y'all gone feel real stupid when @rihanna drop a covid vaccine called "R9".	-1.0	PourJorden
2	god watching everyone in this terrible, murderous administration get the covid vaccine before everyone else feels fucking terrible	-1.0	zoenone0none
3	#ruportmurdoch gets vaccine in #uk but was against #covid measures. His papers views are disgusting. @rupertmurdoch <a href="https://t.co/GjonHnGXe7">https://t.co/GjonHnGXe7</a>	-1.0	xy798
4	SHOCKING! TRUTH COVID-19 VACCINE.SCIENTISTS WARN HUMANITY. MSM BANNED! <a href="https://t.co/iMOZGeCJ6S">https://t.co/iMOZGeCJ6S</a>	-1.0	ian_
5	@marcorubio #FireLiddleMarco \n\nGov. Sununu Blasts Congress Members Who Got COVID-19 Vaccine: 'Irresponsible and Insulting' <a href="https://t.co/DOKothnpiH">https://t.co/DOKothnpiH</a>	-1.0	sfk3226
6	@guypbenson What's pathetic is you working for a network that just promoted people not taking the covid vaccine.	-1.0	xtreme_cent
7	#Tory MP says forcing people to get Covid-19 vaccine would set a 'terrible precedent' THIS IS TODAY'S NEWS	-1.0	vijay72554845
8	@SenJoniErnst You're disgusting you hypocrite #ernst #covid #vaccine #denier <a href="https://t.co/zv3qMmgB4o">https://t.co/zv3qMmgB4o</a>	-1.0	bigtrumploser
9	@adrianbowyer The way to reduce the likelihood of mutations is to reduce the number of infections by rapidly distributing the vaccines, the worst way would be herd immunity, of yeah that was proposed by Covid Cummins.	-1.0	Capt_Kremin

Vaccine	Positive Words
Pfizer/BioNTech	effective safe heart grateful well better happy protect please best ready hope approves share
Covaxin	safe best better like thanks effective great immune proud please approved thank free hope top
Moderna	feeling ready thanks grateful ease good effective happy number free want approved well excited
Oxford/AstraZeneca	proud feeling effective safety delighted approved thanks well grateful want fine amazing please
Sputnik	supreme number approved well help ready want approval great top launched approves please agreed

Vaccine	Negative Words
Pfizer/BioNTech	no emergency risk warning death refused
Covaxin	severe strain emergency no shortage
Moderna	no ill sore emergency pain
Oxford/AstraZeneca	no stop risk suspend sore rejected ill
Sputnik	no emergency demand death fight

# Bigrams Analysis

0	(health, care)	9892
1	(next, week)	8230
2	(care, workers)	6140
3	(approves, pfizerbiontech)	5116
4	(approved, use)	5079
5	(side, effects)	5038
6	(let, us)	4859
7	(uk, approves)	4828
8	(emergency, use)	4126
9	(use, uk)	3833
10	(first, dose)	3719
11	(united, states)	3714
12	(safe, effective)	3686
13	(healthcare, workers)	3667

Most recurring Bigrams - Tweets with Positive Sentiments (December 2020)

0	(emergency, use)	7756
1	(side, effects)	6972
2	(allergic, reactions)	5082
3	(allergic, reaction)	3578
4	(new, strain)	3241
5	(healthcare, workers)	2949
6	(long, term)	2796
7	(use, authorization)	2794
8	(flu, shot)	2390
9	(wear, mask)	2335
10	(let, us)	2306
11	(front, line)	2251
12	(severe, allergic)	1932
13	(people, get)	1931

Most recurring Bigrams - Tweets with Negative Sentiments (December 2020)



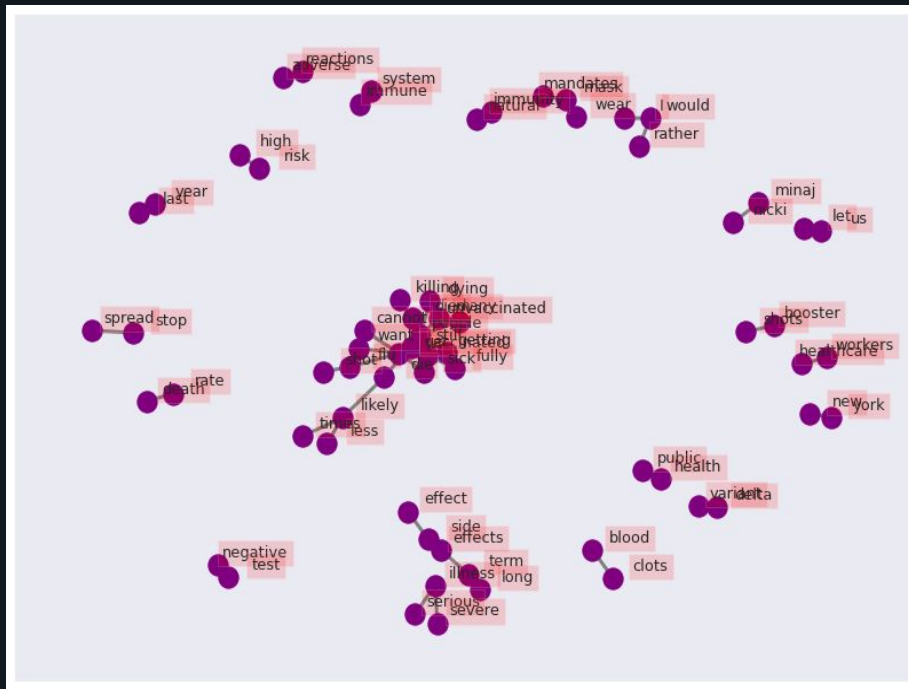


0	(side, effects)	9049
1	(get, vaccinated)	6343
2	(people, died)	5297
3	(long, term)	5205
4	(still, get)	4513
5	(flu, shot)	4362
6	(fully, vaccinated)	4219
7	(vaccinated, people)	4168
8	(many, people)	3734
9	(people, get)	3681
10	(delta, variant)	3209
11	(let, us)	3033
12	(natural, immunity)	3010
13	(get, sick)	2704

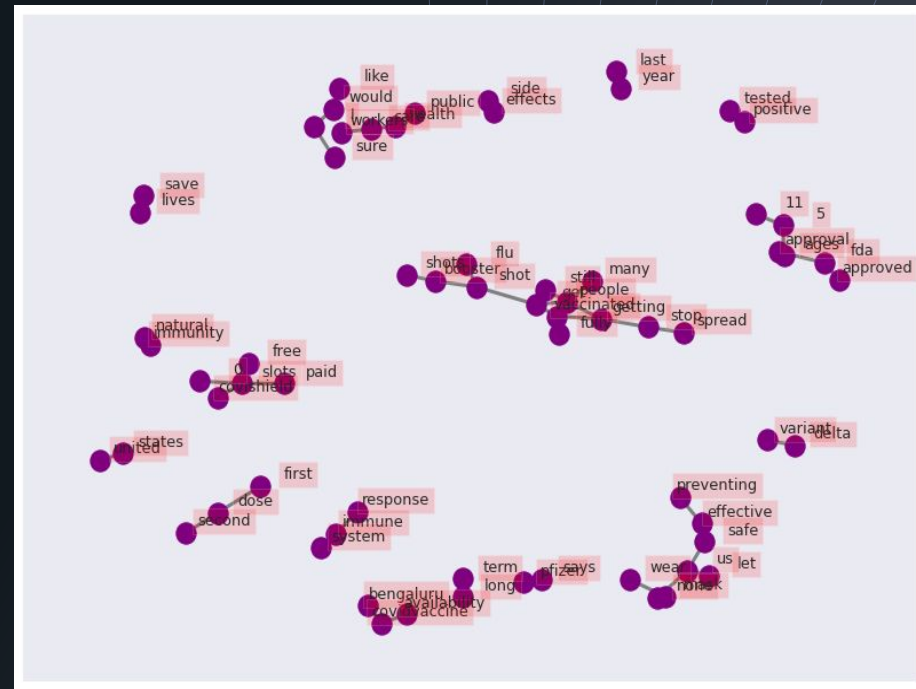
## Negative Sentiments - September 2021

0	(natural, immunity)	11248
1	(get, vaccinated)	6972
2	(side, effects)	5053
3	(safe, effective)	4605
4	(fully, vaccinated)	3877
5	(immune, system)	3838
6	(let, us)	3440
7	(long, term)	3393
8	(still, get)	3235
9	(health, care)	3226
10	(vaccinated, people)	2979
11	(people, get)	2961
12	(delta, variant)	2734
13	(booster, shots)	2610

## Positive Sentiments - September 2021



Negative Sentiments -  
September 2021



Positive Sentiments -  
September 2021

Alleged 'Covid-19' deaths more than 3,000% higher than this time **last year** and 80% of those dying had the fake vaccine according to the latest Public Health data (the fake vaccine is killing them)

<https://t.co/yykWaRljAy>

@EliseRCook COVID vaccines in the **last year** have killed more people than all other vaccines combined since 1990, and that's just the deaths that are getting reported. <https://t.co/OUyRtLoNxS>

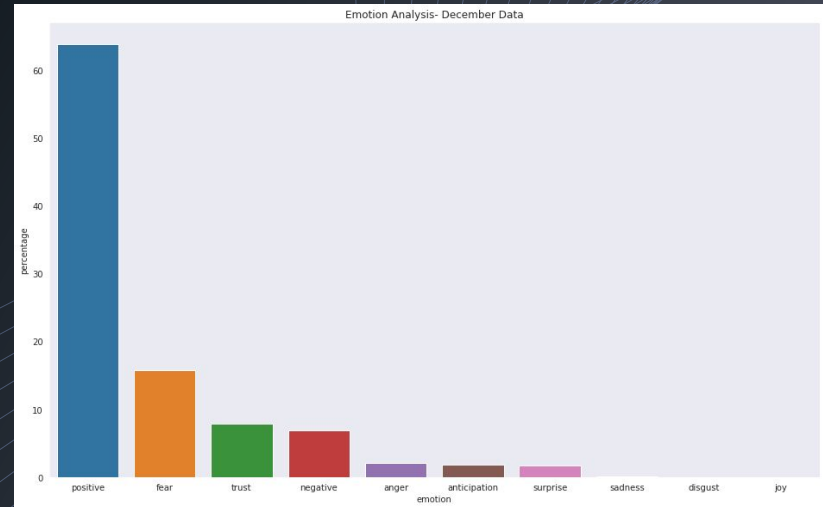
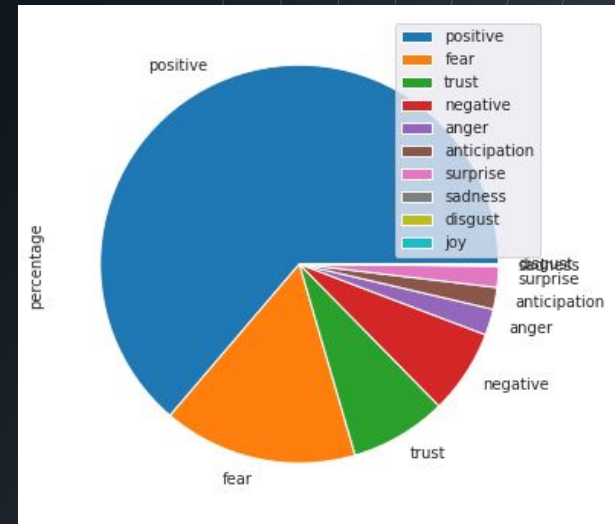
@HeyNurseKat @FourWinns298 Says Kat who has a vaccine in her body for **last years** covid. Keep that same energy around your 8th or 9th shot. Can you still catch covid once vaccinated?? Hmmm.  
\*\*\*\* your science.

An example of a set of tweets containing these bigrams described above - This bigram was mostly used in anti-vaccine tweets.

# Emotion Analysis-Dec

	emotion	counts	percentage
0	positive	689037	63.76
1	fear	170433	15.77
2	trust	84801	7.85
3	negative	73994	6.85
4	anger	23063	2.13
5	anticipation	19137	1.77
6	surprise	18454	1.71
7	sadness	1468	0.14
8	disgust	345	0.03
9	joy	21	0.00

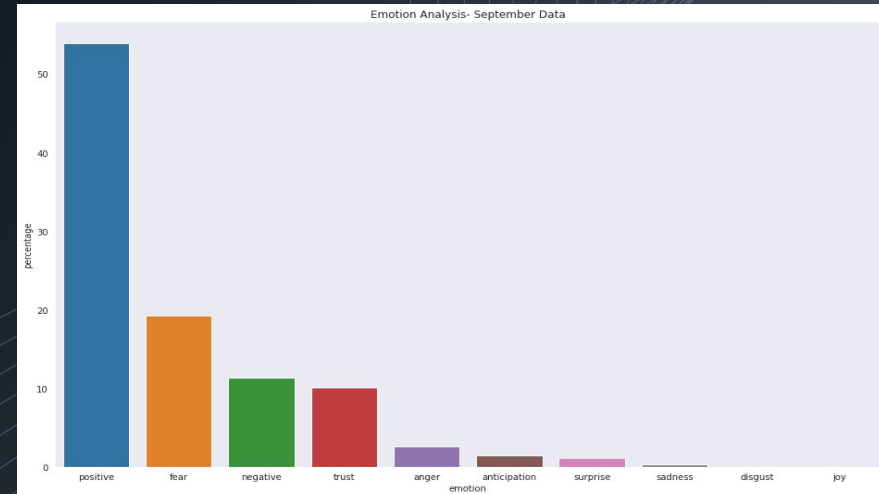
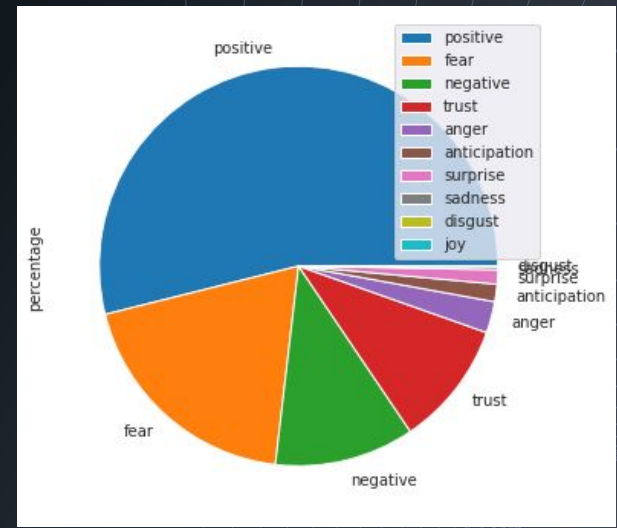
Table depicting results from emotion analysis performed using NRClex on covid vaccine (December 2020) data



# Emotion Analysis-Sept

	emotion	counts	percentage
0	positive	496427	53.89
1	fear	177300	19.25
2	negative	104455	11.34
3	trust	93271	10.12
4	anger	23436	2.54
5	anticipation	13003	1.41
6	surprise	10656	1.16
7	sadness	2413	0.26
8	disgust	248	0.03
9	joy	32	0.00

Table depicting results from emotion analysis performed using NRCLEX on covid vaccine (September 2021) data

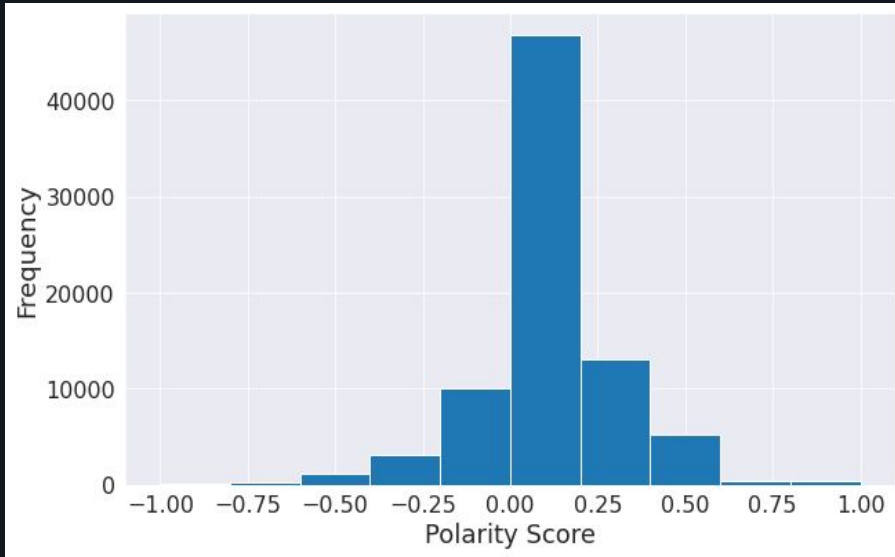




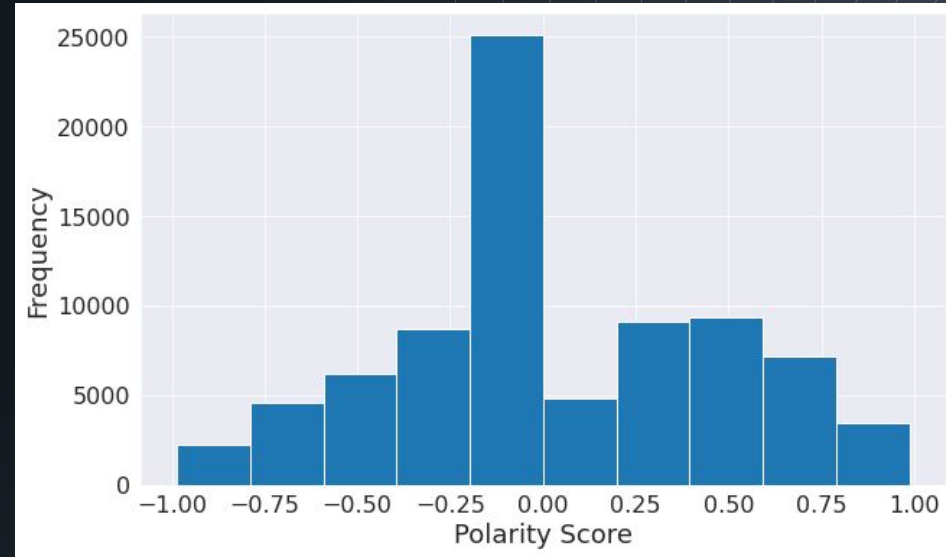
People's opinion changed over this time period ( Dec 2020 to Sep 2021 ):

Emotions	Changed over this time period
Positivity	<b>-9.87 %</b>
Fear	<b>+3.48 %</b>
Negativity	<b>+4.49 %</b>
Trust	<b>+2.27 %</b>
Anger	<b>+0.41 %</b>
Anticipation/Joy/Surprise/Sadness	Not Significant

# Sentiment Analysis - Vaccine Booster Data



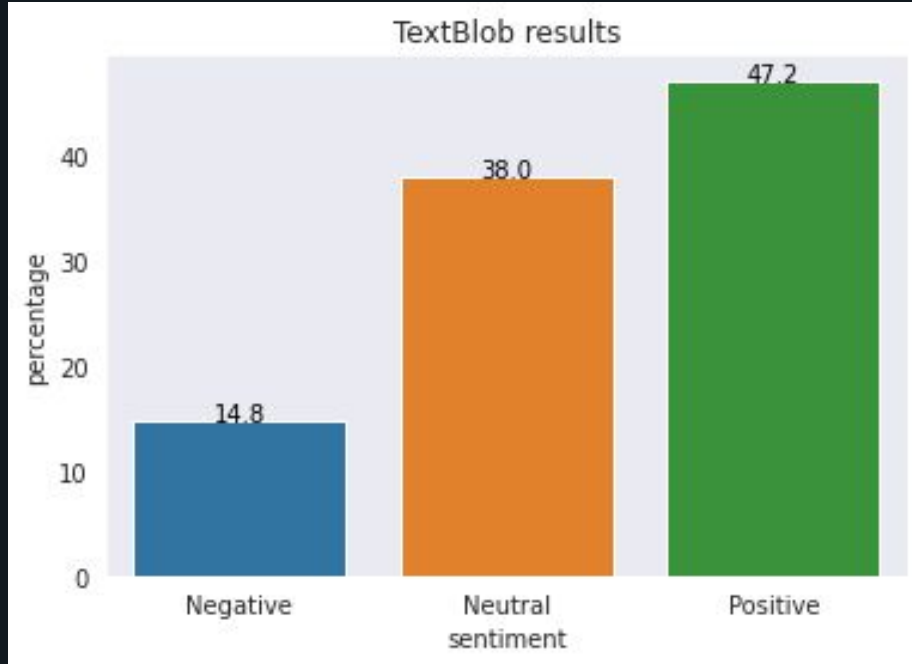
Polarity scores' distribution -  
Vaccine booster Data  
(TextBlob)



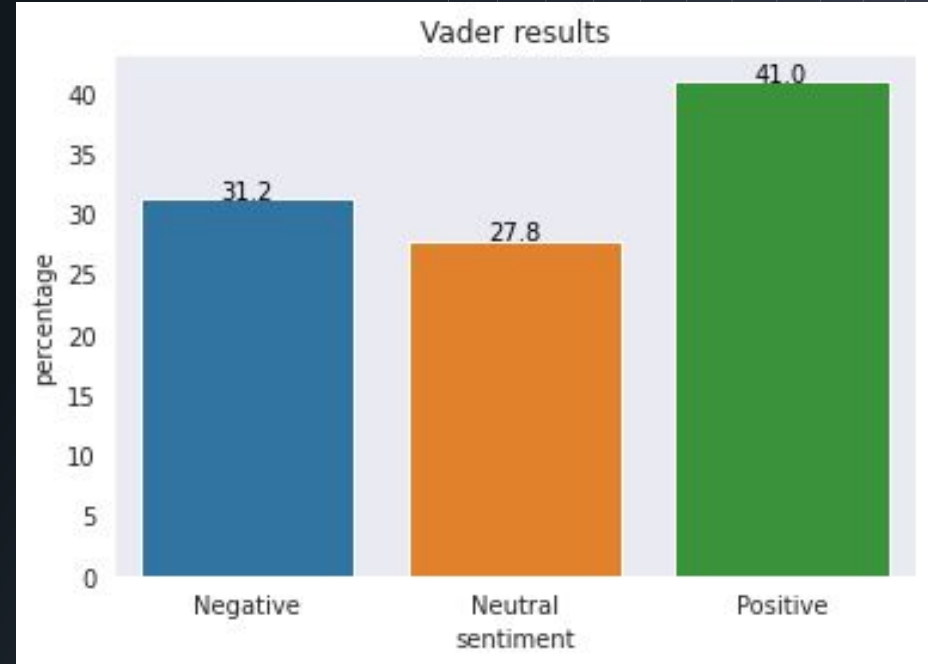
Polarity scores' distribution -  
Vaccine booster Data  
(VADER)



# Sentiment Analysis - Vaccine Booster Data



Percentage of Tweets across sentiment categories - Vaccine booster Data (TextBlob)



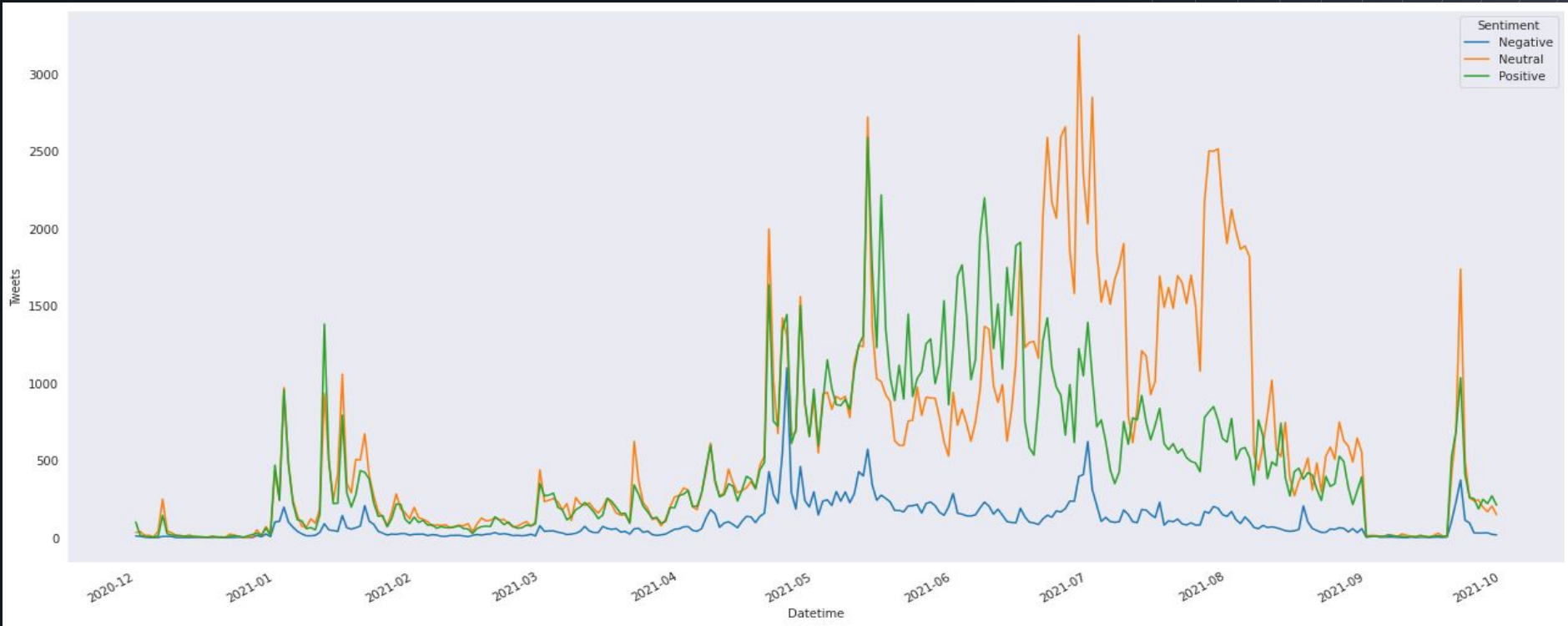
Percentage of Tweets across sentiment categories -Vaccine Booster Data (VADER)

Sentiment	Classifier	December 2020	Vaccine Booster
Negative	TextBlob	<b>14.9</b>	<b>14.8</b>
Positive	TextBlob	<b>44.2</b>	<b>47.2</b>
Neutral	TextBlob	<b>40.9</b>	<b>38.0</b>

Sentiment	Classifier	December 2020	Vaccine Booster
Negative	VADER	<b>29.0</b>	<b>31.2</b>
Positive	VADER	<b>41.8</b>	<b>41.0</b>
Neutral	VADER	<b>29.3</b>	<b>27.8</b>

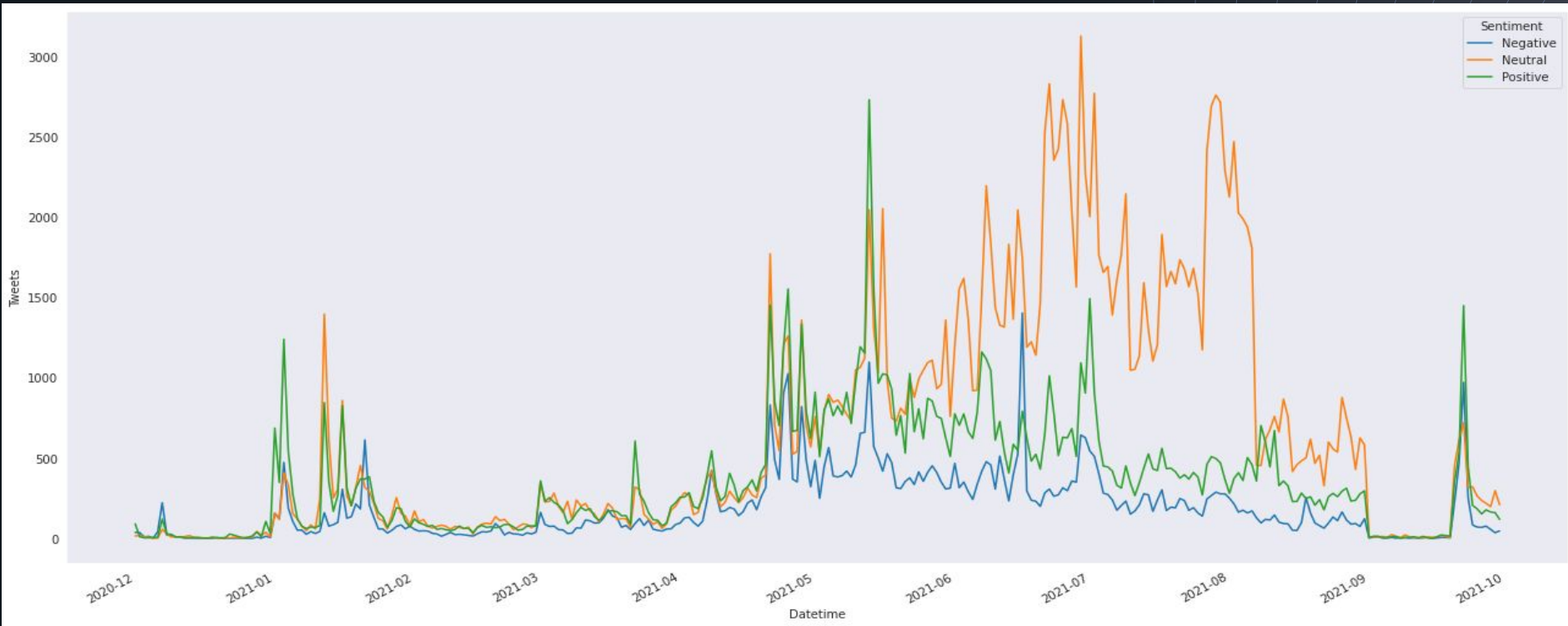
Tables comparing sentiment classification results for pre-vaccine rollout and pre-booster rollout.

# Time-Series analysis on Covishield Data



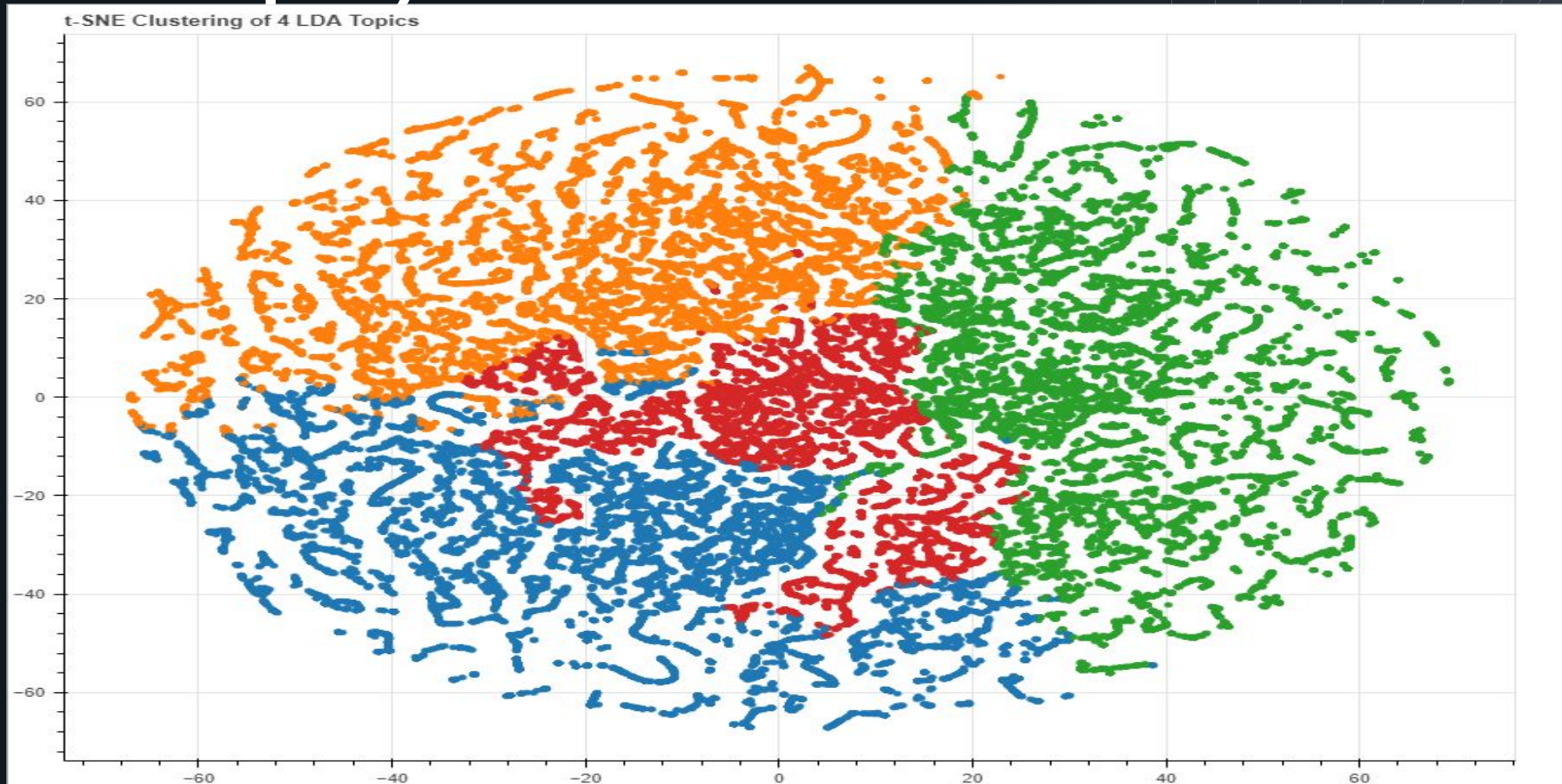
Evolving public sentiments for the tweets containing the keyword 'covishield' -  
Classified using TextBlob

# Time-Series analysis on Covishield Data



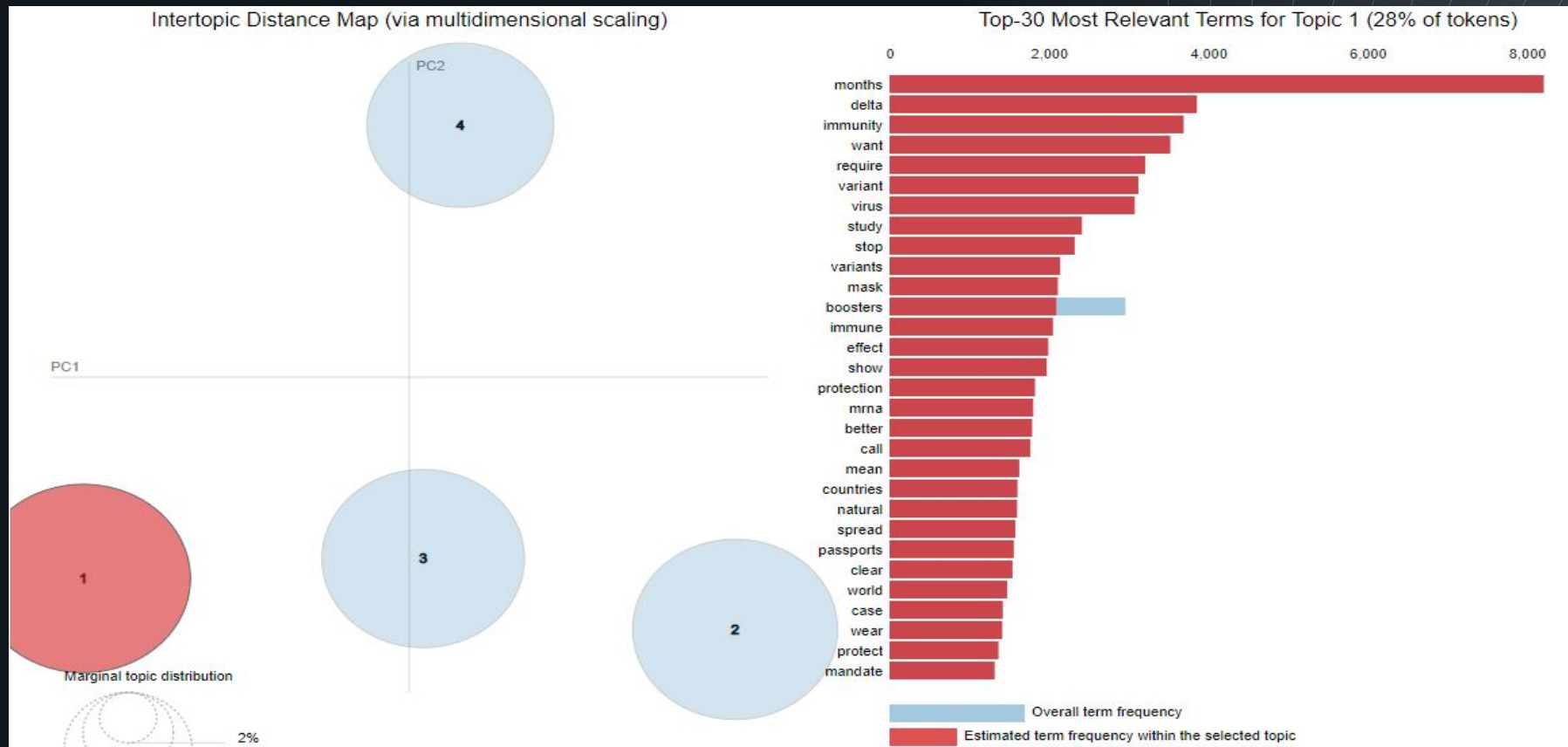
Evolving public sentiments for the tweets containing the keyword 'covishield' -  
Classified using VADER

# Topic Modeling Results (t-SNE clustering of 4 LDA topics)



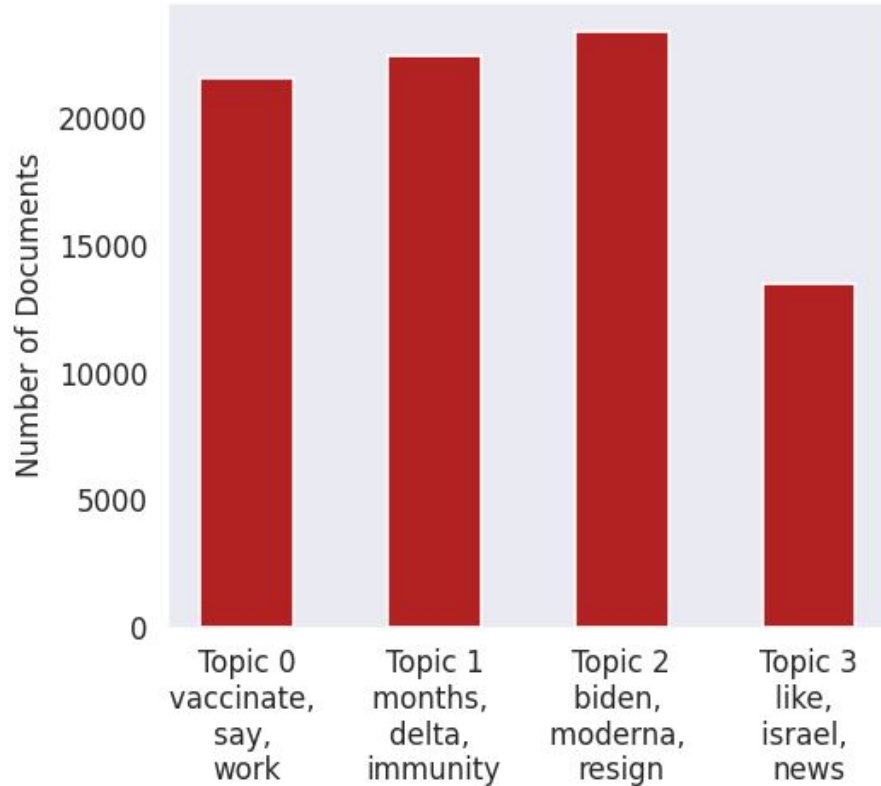


# Result (pyLDAvis visualisation)

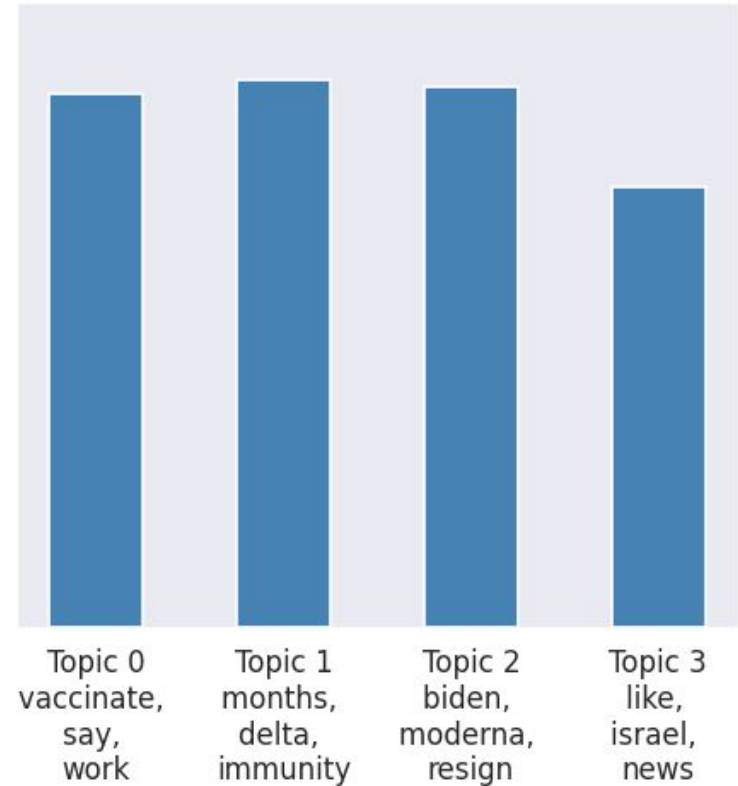


# Topics Identified - Booster Data

Number of Documents by Dominant Topic



Number of Documents by Topic Weightage



Topic 1	Topic 2	Topic 3	Topic 4
vaccine	feel	thank	go
say	today	shot	people
well	day	arm	make
able	good	do	see
friend	week	hit	safe
come	time	ready	work

Most discussed topics in positive tweets

Topic 1	Topic 2	Topic 3	Topic 4
vaccine	arm	be	get
first	dose	shot	weak
day	sore	go	ill
feel	second	amp	report
still	shoot	make	find
pain	yesterday	may	kill

Most discussed topics in negative tweets



# References:

- Bonta, Venkateswarlu & Kumares, Nandhini & Janardhan, N.. (2019). A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer Science and Technology. 8. 1-6. 10.51983/ajcst-2019.8.S2.2037.
- Cinelli, Matteo & Morales, Gianmarco & Galeazzi, Alessandro & Quattrociocchi, Walter & Starnini, Michele. (2021). The echo chamber effect on social media. Proceedings of the National Academy of Sciences. 118. e2023301118. 10.1073/pnas.2023301118.
- COVID-19 Echo Chambers: Examining the Impact of Conservative and Liberal News Sources on Risk Perception and Response. Kenneth A. Lachlan, Emily Hutter, and Christine Gilbert. HealthSecurity. Feb2021.21-30. <http://doi.org/10.1089/hs.2020.0176>.
- Loomba, S., de Figueiredo, A., Piatek, S.J. et al. Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. Nat Hum Behav 5, 337–348 (2021). <https://doi.org/10.1038/s41562-021-01056-1>.
- <https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/>
- <https://www.fda.gov/news-events/press-announcements/fda-authorizes-booster-dose-pfizer-biontech-covid-19-vaccine-certain-populations>
- <https://economictimes.indiatimes.com/topic/Indias-Covid-19-Vaccine/2>
- <https://www.reuters.com/world/us/top-us-health-officials-plan-narrower-covid-19-booster-launch-source-2021-09-03/>
- <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/>
- <https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html>
- <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17>
- <https://www.thehindu.com/news/national/coronavirus-live-june-22-2021-updates/article34898077.ece>
- Sarker, I.H. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT. SCI. 2, 160 (2021). <https://doi.org/10.1007/s42979-021-00592-x>

# References:

- D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet allocation,” Journal of Machine Learning Research, vol. 3, pp. 993–1022, 2003.
- Boon-Itt S, Skunkan Y, “Public Perception of the COVID-19 Pandemic on Twitter: Sentiment Analysis and Topic Modeling Study”, JMIR Public Health Surveill 2020;6(4):e21978
- <https://radimrehurek.com/gensim/models/ldamodel.html> (Gensim official docs)
- Mendon, S., Dutta, P., Behl, A., & Lessmann, S. (2021). A Hybrid Approach of Machine Learning and Lexicons to Sentiment Analysis: Enhanced Insights from Twitter Data of Natural Disasters. Information Systems Frontiers. doi:10.1007/s10796-021-10107-x

**Thank You!**

The background is a dark blue gradient. It features a complex pattern of thin, light blue lines that intersect to create a perspective effect, resembling a grid or a series of converging lines that create a sense of depth and movement.