Analyzing Twitter Trends On COVID-19 Vaccinations

A quantitative study comprising discussions and thematic analysis

Highlights

- 2.1 million tweets from individual users and 0.59 million tweets from organizations between January 2021 and March 2021 were utilized for the Twitter text analytics

- 34.7% people had positive sentiments for the COVID-19 vaccines as compared to 24.6% of negative sentiments

- The number of tweets containing sentiments of people reduced significantly moving from first to second week of February, with steady increase in March

- As the conversations around side-effects of vaccines took a dip of 30% in March (w.r.t. previous week), there is a 97% rise in the number of tweets by people rooting for COVID-19 vaccines

Background

- Five months into 2021, the normalcy post COVID-19 hasn't returned yet, and many parts of the world are still struggling with COVID-19

- Vaccinations have proved to be a light at the end of the tunnel, diminishing the scare of pandemic

- Growing human rights concerns, vaccine movements, and skepticism towards the vaccines, its effects and efficacy have resulted in a multitude of conversations on social media and the process of vaccination becoming a complicated task

- No major studies have been conducted to analyze people's perception of COVID-19 vaccines on social media for the year 2021

Project Goal

- To extract information from tweets (between January and March) related to COVID vaccine where opinions are highly unstructured, heterogeneous and are either positive or negative, or neutral in some cases

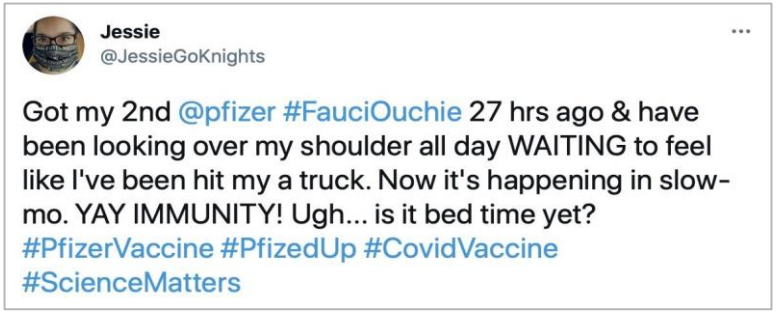
- To explore conversations and abstract "topics" that occur in the collected tweets using topic modeling and text analytics

- To visualize the trends in sentiments and popularity associated with the discovered topics

COVID Vaccine Conversations on Twitter


CNN Breaking News
@cnnbrk
About 1 in 6 US residents are fully vaccinated against Covid-19, CDC data shows
6:03 AM · Mar 31, 2021 · Twitter Web App


Bee
@izziewithaY
All the conspiracy theorists will start going on again about the microchips being used in the Covid vaccines.
12:02 AM · Feb 10, 2021 · TweetDeck


Jessie
@JessieGoKnights
Got my 2nd @pfizer #FauciOuchie 27 hrs ago & have been looking over my shoulder all day WAITING to feel like I've been hit my a truck. Now it's happening in slow-mo. YAY IMMUNITY! Ugh... is it bed time yet? #PfizerVaccine #PfizedUp #CovidVaccine #ScienceMatters


Louisville Scoundrel
@DoctorColby
Covid vaccine 24 hour update: alternating fever and chills last night, foggyheadedness due to lack of sleep. Not awful, but if you're fortunate enough to be able to take off work the next day, I recommend it. Head and limbs feel heavy, but no body aches yet.
8:40 AM · Feb 10, 2021 · Twitter Web App
Louisville Scoundrel
@DoctorColby
48 hours after 2nd covid vaccine: I slept well and feel pretty good. Arm is sore but a bit better.
9:34 AM · Feb 11, 2021 · Twitter for iPhone

Questions

- RQ 1:

  What are the sentiments of people towards COVID-19 vaccine? How does it differ across three months of 2021?

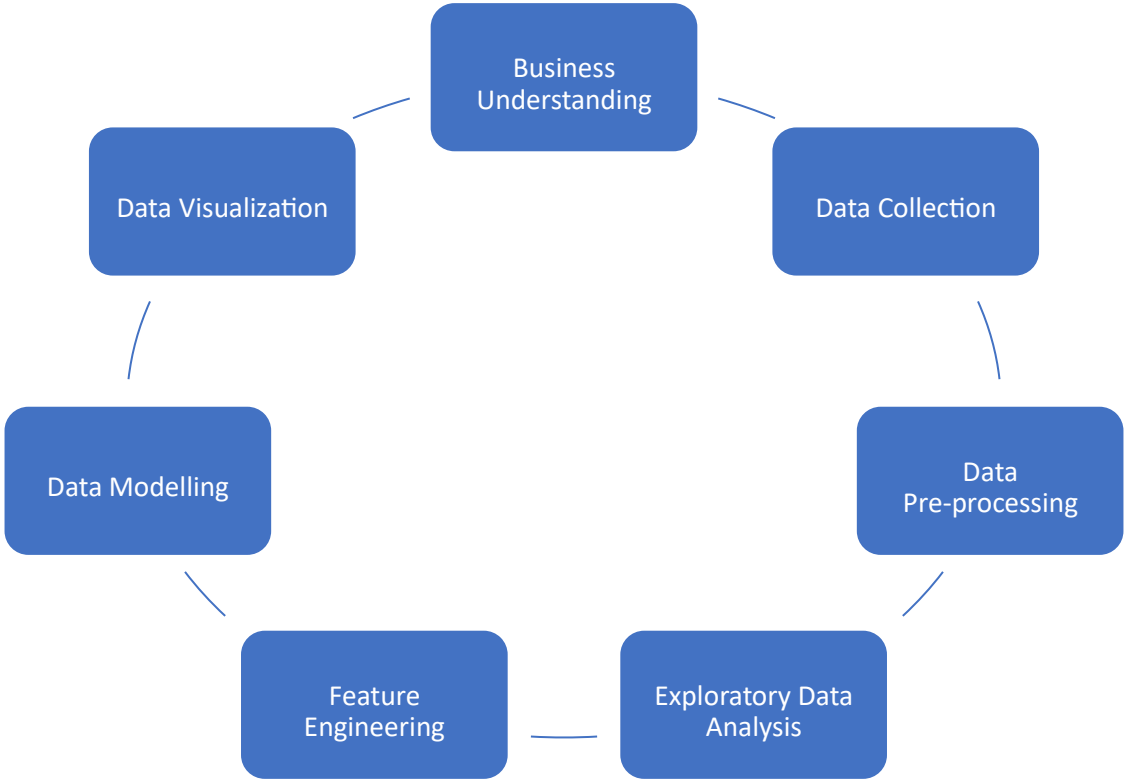
- RQ 2:

  What are the conversations around COVID-19 vaccines on Twitter and how has it changed over time?

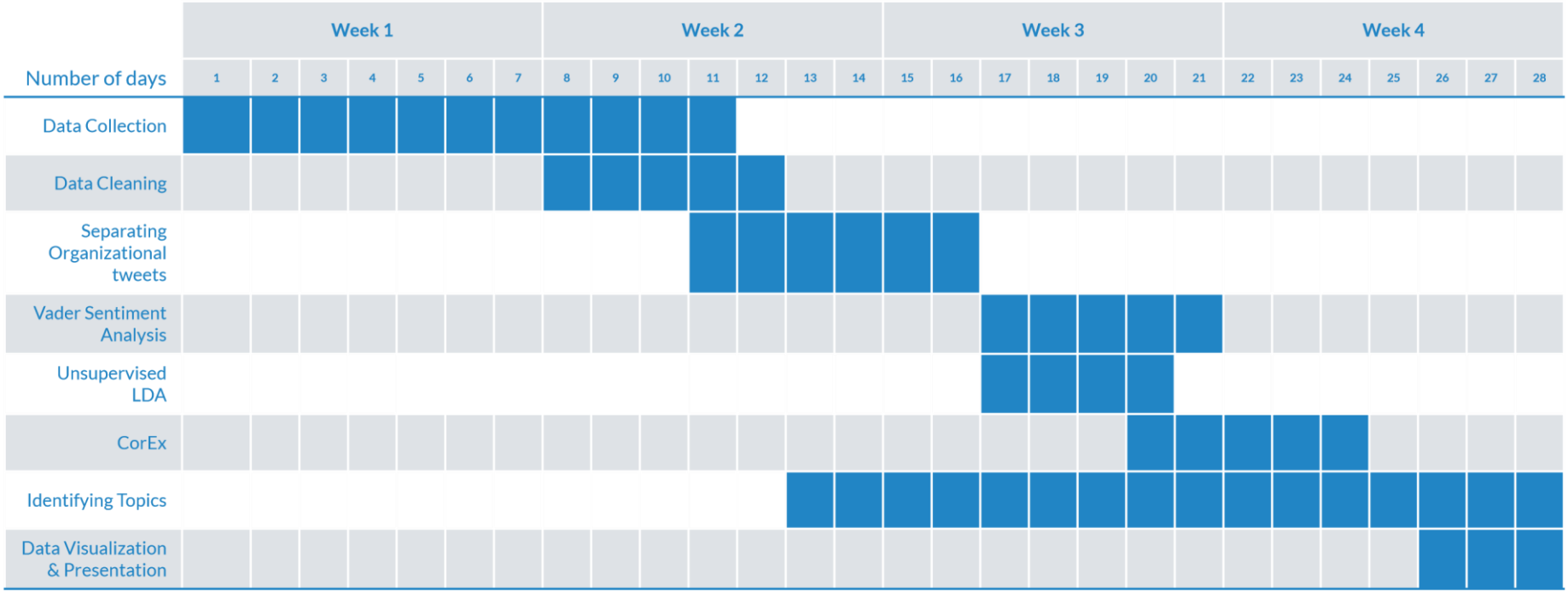
- RQ 3:

  What are the most popular implicit topics in the tweets?

Approach


Business Understanding
Data Collection
Data Pre-processing
Exploratory Data Analysis
Feature Engineering
Data Modelling
Data Visualization

Gantt chart


Week 1
Week 2
Week 3
Week 4
Number of days
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Data Collection
Data Cleaning
Separating Organizational tweets
Vader Sentiment Analysis
Unsupervised LDA
CorEx
Identifying Topics
Data Visualization & Presentation

Data Collection

- Package used: `snscrape`

Method	Notes
Tweepy	3200 tweets; no historical data
GetOldTweets3	Twitter has removed the endpoint the GetOldTweets3 uses
TWINT	Twitter throws a more strict device + IP-ban after a certain amount of queries.
snscrape	Scrapped 100K tweets - 96,641 English tweets
Octoparse	Very time consuming with the event loop

- Language: English
- Keywords: `covid vaccine`
- Timeframe:
  January 1, 2021 to March 31, 2021
- Number of tweets collected =
  2.74 million
- No null values identified

About the Data

Total number of tweets	2.7 million

Individuals

Organizations

Number of tweets by individuals	2.10 million
Unique accounts	916,716
Tweets per User	2.30

Number of organization tweets	0.59 million
Unique accounts	88,584
Tweets per Organization	6.74

Personal Tweets Stats about COVID Vaccine

Total number of individual users	916,716	
Count of Users with exactly 1 Tweet	622,296	67.88%
Count of Users with more than 1 Tweets	294,419	32.12%

Count of Users with 2-9 Tweets	261,715	88.90%
Count of Users with 10-19 Tweets	21,356	7.25%
Count of Users with 20-49 Tweets	8,391	2.85%
Count of Users with more than 50-99 Tweets	2,221	0.75%
Count of Users with more than 100 Tweets	736	0.25%

Week-wise Distribution of tweets

1	165860
2	159618
3	167140
4	203272
5	222731
6	124551
7	122847
8	123921
9	151928
10	196856
11	203533
12	267170

Top 10 Users	Count
AndyVermaut	2336
barasajoel2	1803
dev_discourse	1760
mlnangalama	1446
sarang143u	1173
inquirerdotnet	1028
lokeshjaral1	975
AJBlackston	958
isearch247	846
AdamAda13410588	842

Data Pre-Processing

Data Cleaning

a. Removed punctuation using `remove_punct` function with library `re`

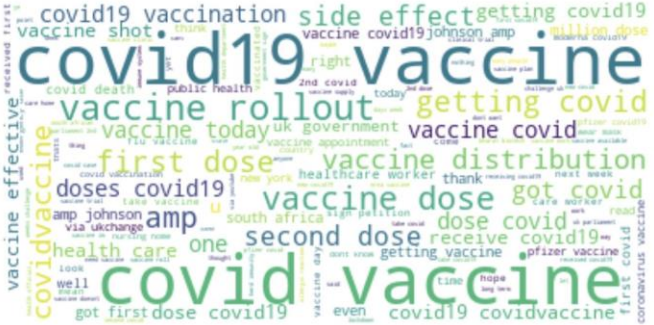
b. Removed URLs and emojis in Tokenization using library `re`

c. Removed stopwords using `nltk`

d. Lemmatization of Tweets using `nltk.WordNetLemmatizer()` **Individual vs Organizational Tweets**

a. Created a Bag-of-Words with ~175 keywords to filter on Display Names

b. Removed 22% of the data



c. 2,109,427 tweets remain after removing organizational accounts

d. Assigned week numbers (1 to 12) to the dataset

Sentiment Analysis

sentiment **score**

- Replaced bigrams indicating Covid positive as cpos and Covid negative as cneg

Negative	-0.411274
Neutral	0.000210
Overly Negative	-0.840563
Overly Positive	0.837640
Positive	0.422623

Positive	732395
Neutral	579439
Negative	525866
Overly Positive	154470
Overly Negative	117257

	frequency	bigram/trigram
0	30	working cyprus
1	30	walmart china
2	30	vechainofficial technology
3	30	vaccine distribution
4	30	technology immutable
5	30	supply chain
6	30	several orher
7	30	secures vaccine
8	30	save vechainofficial
9	30	prtnrs please

- Assigned polarity scores as cpos = -3 and cneg = 3 - Defined sentiments based on intensity scores:

Overly Positive : >0.75

Positive : Between 0.05 & 0.75

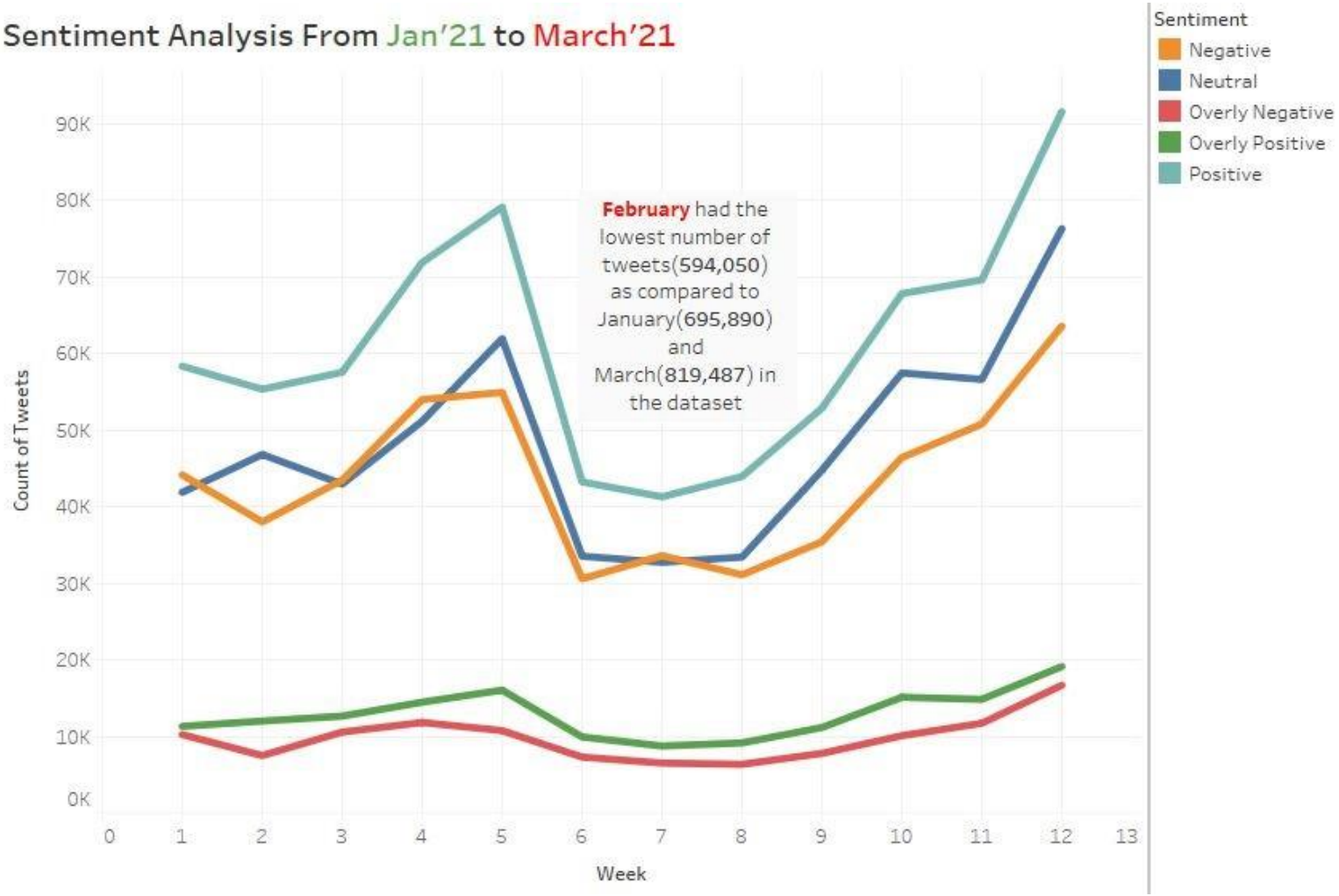
Overly Negative : <-0.75

Negative : Between -0.05 & -0.75

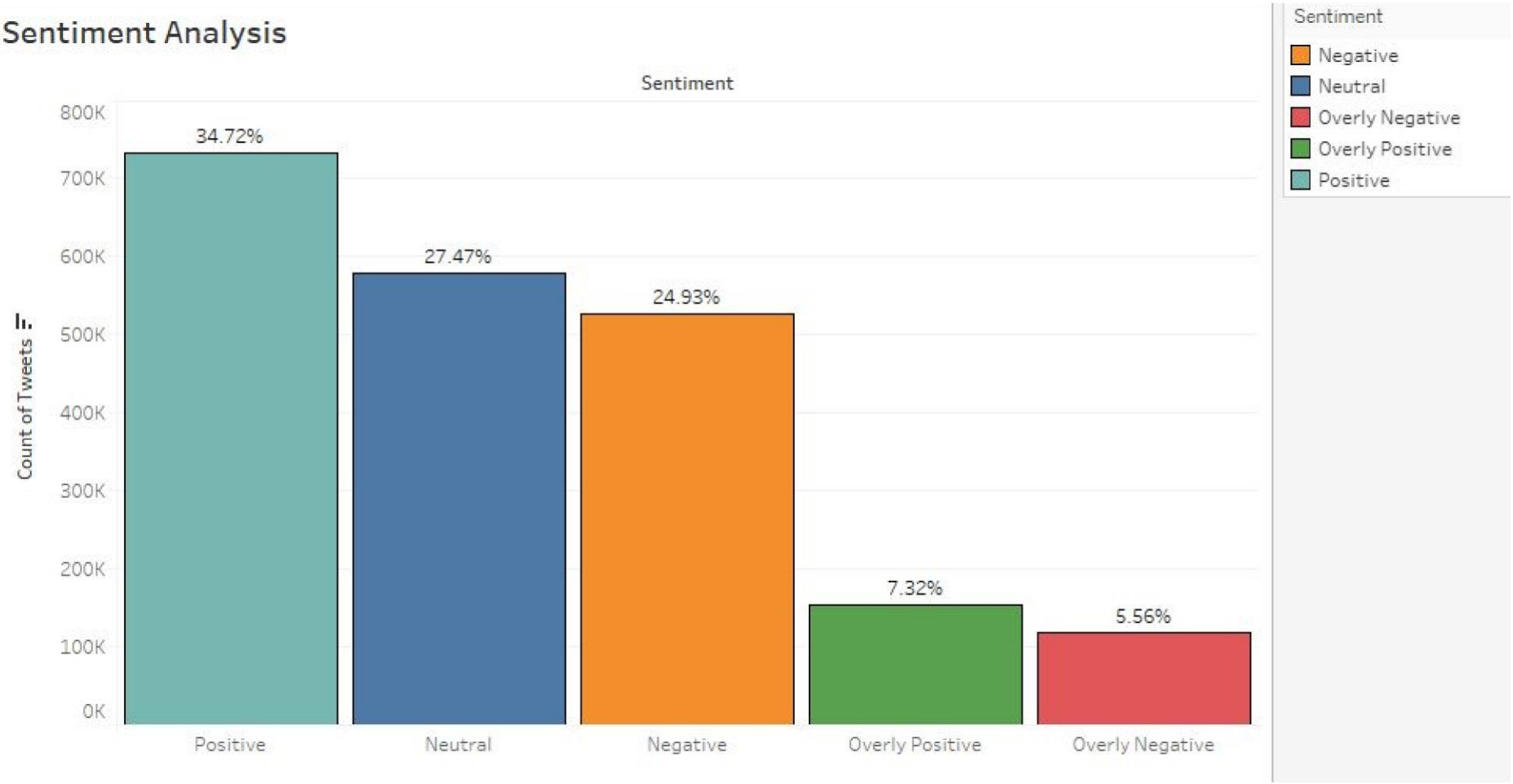
Neutral : -0.05 to 0.05

sentiment	frequency

Sentiment Analysis From Jan'21 to March'21

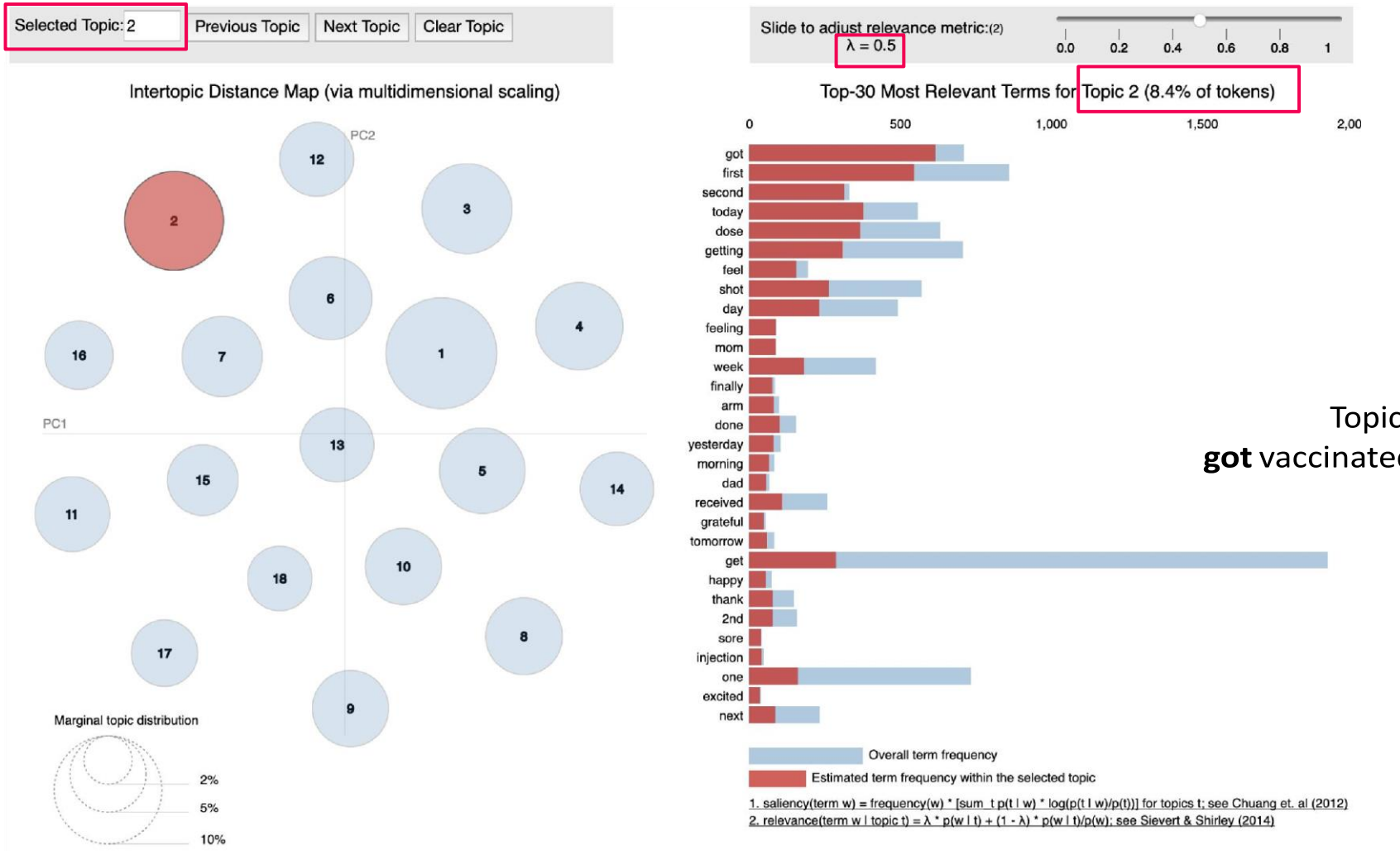


Sentiment Analysis



Data Modelling - Unsupervised LDA

- Removed 73 stopwords from the tweets
- CountVectorizer:
    - All terms occur over 75% times in our document corpus. We say in this case that the terms occurring more than this threshold are not significant, most of them are stopwords - All the terms occur at least twice in the entire corpus.
- Computed total number of words and unique keywords across 12 weeks
- From the Document-Term matrix, we built our LDA to extract 18 topics from the underlined texts
- Declared number of iterations as 30 (default value is 10) and found top 20 words per topic
- We computed a probability score of how likely a tweet belongs to each of the 18 topics
- Package used to visualize intertopic distance map: `pyLDAvis.sklearn`



Topic:
got vaccinated

Topics Identified

Vaccination Eligibility	65, department, eligible, hospitalized, 75, local,
Vaccine Registration, Scheduling, Appointment	booked, book, schedule, call, website, email
Vaccine Approval	approved, approval, clinical, trial, approves, rollout
Healthcare Workers	healthcare, workers, doctor, dr, nurse, employees
Advocacy towards Vaccine	getvaccinated, readytovaccinate, goandgetvaccinate, vaccinated, vaccineswork, vaccinessavelives
Hesitancy towards Vaccine	antivaccine, vaccinekills, novaccination, stopvaccines, nomandatoryvaccine, forced vaccination
Vaccination Experience	mother, dad, kid, yesterday, morning, today
Side effects of vaccine	symptom, fever, chills, body, pain, paralysis
Vaccination with Pandemic Management Measures	variant, pandemic, lockdown, mask, distancing, transmission

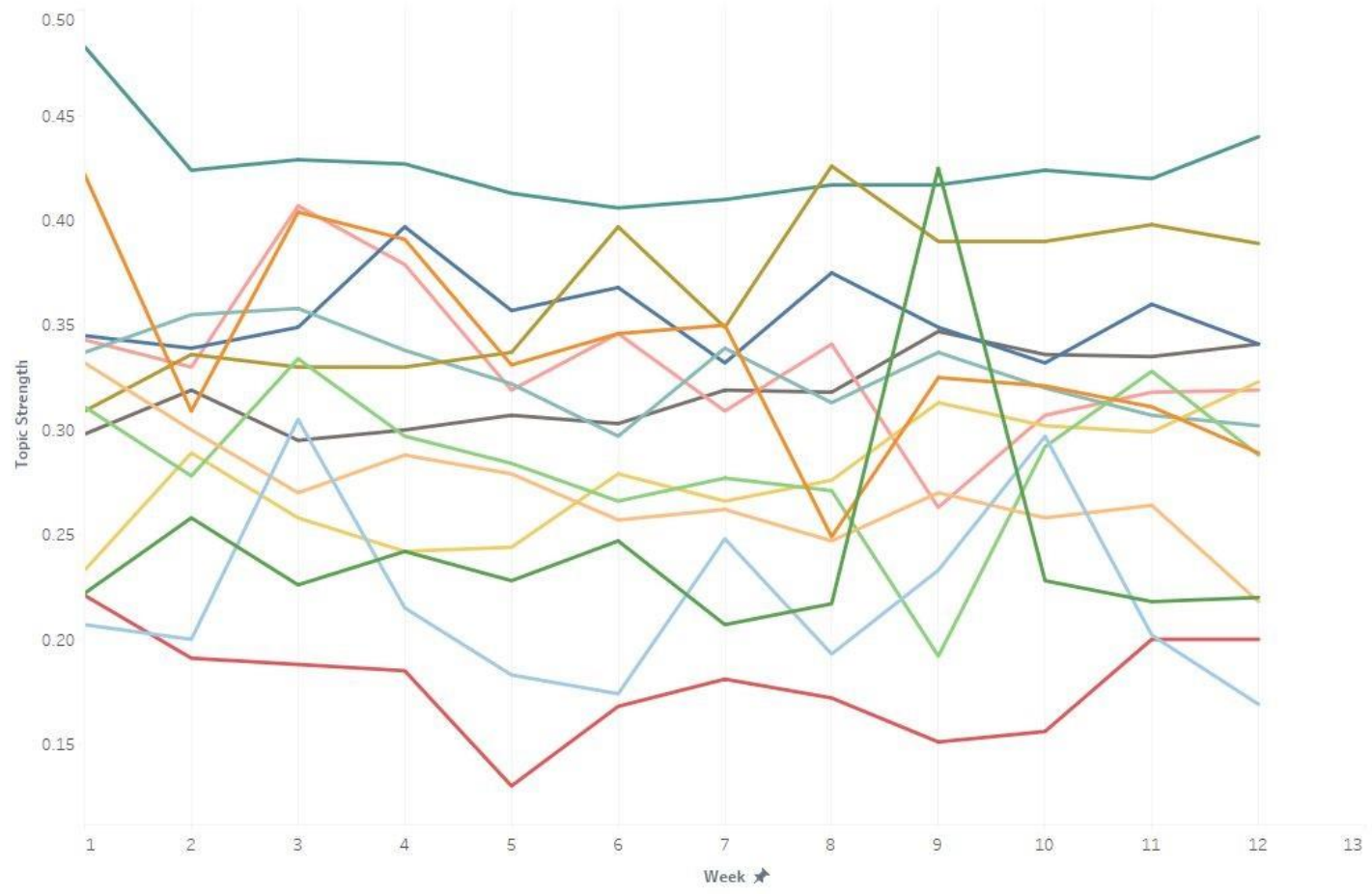
Data Modelling - CorEx

- CountVectorizer: Transformed data into a sparse matrix (n_docs X m_words)

- **n_hidden** = 14 (number of topics as in hidden latent topics)
- **anchor_strength** = 2 (how much weight CorEx puts towards maximizing mutual information between anchor words and their respective topics)
- Normalized Topic Correlation: represent the correlations within an individual document explained by a particular topic. Measure how "surprising" documents are with respect to given topics

```
0: 65, resident, old, eligible, age, local, senior, elderly, department, 75
1: available, appointment, line, open, list, schedule, online, website, register, demand
2: rollout, distribution, million, plan, trial, effective, phase, supply, testing, official
3: trump, biden, modi, fauci, joe, cuomo, boris, desantis, trudeau, namo
4: country, risk, access, travel, passport, ban, border, iran, airport, flight
5: new, pandemic, mask, test, spread, variant, lockdown, social, safety, prevent
6: government, uk, india, china, israel, president, canada, minister, africa, eu
7: worker, care, home, patient, staff, healthcare, medical, working, family, school
8: work, vaccinated, ready, save, antibody, vaccineswork, vaccinessavelives, getvaccinated, life, thisisourshot
9: stop, mandatory, warning, anti, chip, danger, sideeffects, occupying, international, ignoring
10: mom, woman, child, parent, dad, kid, mother, mum, wife, husband
11: got, today, day, week, second, received, dose, hour, tomorrow, yesterday
12: death, ill, died, positive, infection, symptom, reaction, body, sore, injection
13: state, hospital, county, community, site, 10, pharmacy, city, clinic, black
```

Vaccine Conversation Trend from Jan'21 to Mar'21



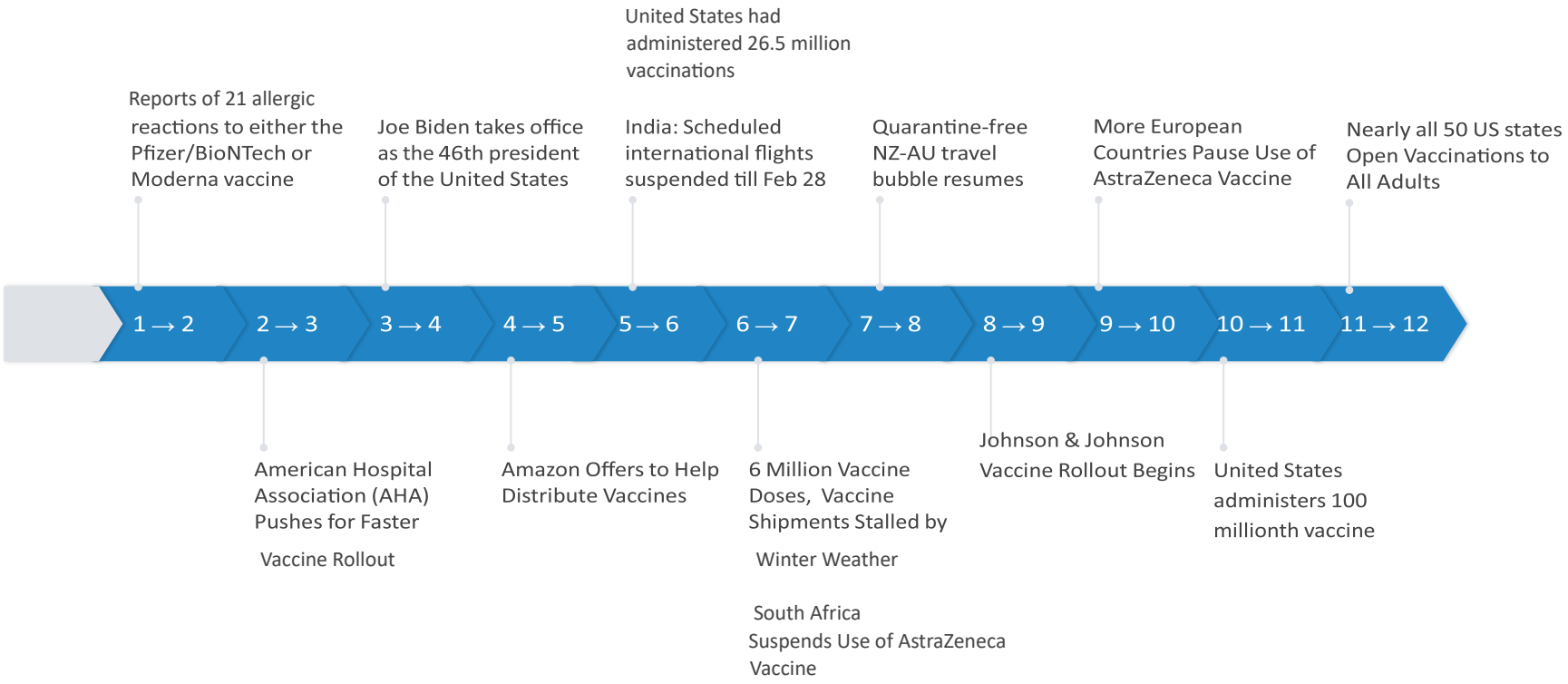
Topics

- Advocacy For Vaccine
- Celebrity Appeal and Endorsements
- Geography
- Healthcare workers
- Hesitancy for Vaccine
- Side effects of vaccine
- Travel
- Vaccination Eligibility
- Vaccination Experience
- Vaccination Sites
- Vaccination with Pandemic Management Measures
- Vaccine Approval
- Vaccine Registration, Scheduling, Appointment

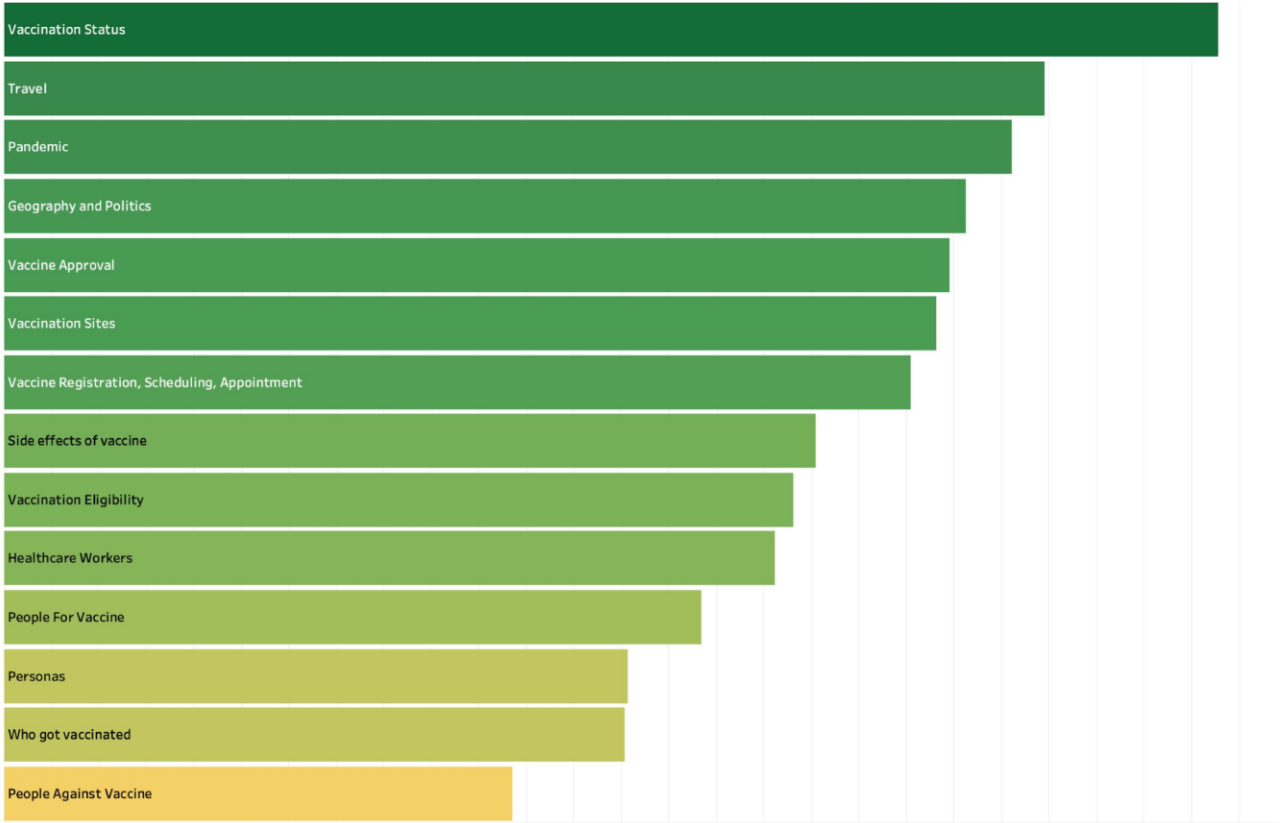
Week

(All)

Weekly Timeline of Major Events



Popular Topics between January 1 - March 31 (2021)



Challenges and Future Scope

Challenges

- Online team collaboration and integration
    - Identifying package and recognizing limitations for extracting tweets
- Large execution times and runtime errors
- Memory limitation for running bigram analysis on entire dataset
- Creating exhaustive list of stopwords and keywords

Future Scope

- Low impact insights from VADER Sentiment Analysis opens up a scope for deep dive into topics independently like People For/Against vaccines
- Explore conversations and sentiments in organizational tweets
- Number of active COVID cases, recoveries and deaths for the three months

Conclusion

- Results obtained are based on collected data over 12 weeks

- Positive sentiment contributed the most in overall sentiment (732,395), followed by neutral (579,493) and negative (525,866) sentiments

- People were discussing most about topics like:

  Vaccination status, Travel, Pandemic, Politics, Vaccine Approval

- Topics that remained underrepresented:

  People Against Vaccine, Political and COVID leaders, Who Got Vaccinated

- These findings can help understand specific topics and human sentiments creating greater traction

- Governments can develop communications strategies that combat misinformation and the most pressing public concerns, thereby being better able to strategize and plan for the future

Thank You!