Analyzing Online Discourse on COVID-19 Vaccines: Evidence from Twitter

Garda Ramadhito, Jianing Wang, Leo Liu, Zhengze Hou

Data and Methodology

Data:

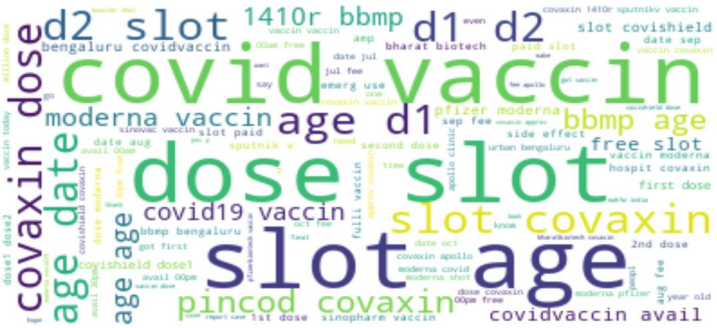
- Twitter data collected through the Twitter API by Gabriel Preda (Kaggle)
- No. of observations: 228,207 tweets
- Date: 12/20/2020 to 11/23/2021
- Vaccines include: Pfizer/BioNTech, Sinopharm, Sinovac, Moderna, Oxford/Astrazeneca, Covaxin and Sputnik V

Methodology:

- Sentiment Analysis: TextBlob and VADER
- Word Clouds, TF-IDF
- Time series

How do discussions about the vaccines differ by different kinds of COVID-19 Vaccines?

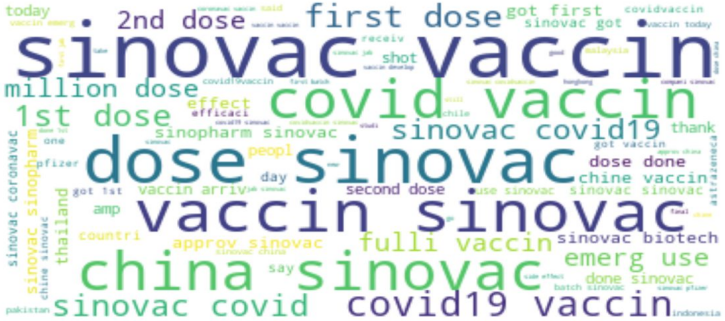
The word cloud for different kinds of vaccines



Graph 1:All kinds of vaccines

Graph 2: Pfizer vaccines

Graph 3: Russian vaccines



Graph 4: Chinese vaccines

Total	Amount	Percentage
neutral	12314	56.28
positive	6284	28.72
negative	3280	14.99

Table 1: The sentiment about Pfizer vaccine

Total	Amount	Percentage
neutral	11938	66.86
positive	4051	22.69
negative	1866	10.45

Table 2: The sentiment about Russian vaccine

The word cloud for user_location of different kinds of vaccines

Graph 1: AZ vaccine

Graph 2: Pfizer vaccine

Graph 3: Chinese vaccine

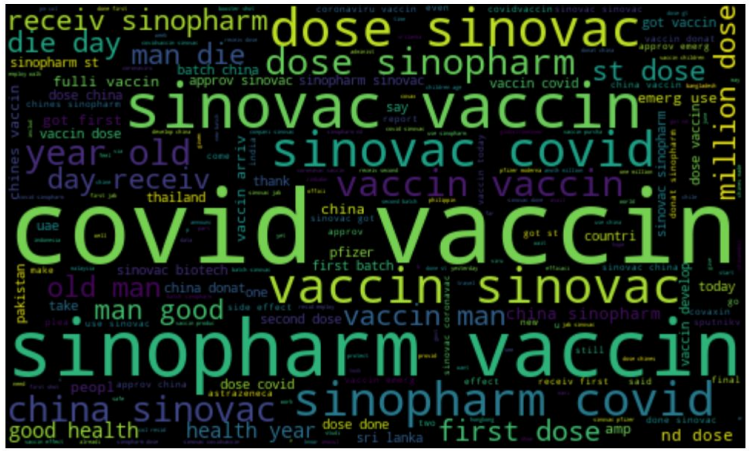
Graph 4: Russian vaccine

What are the sentiment differences between Chinese Vaccines and non-Chinese Vaccines?

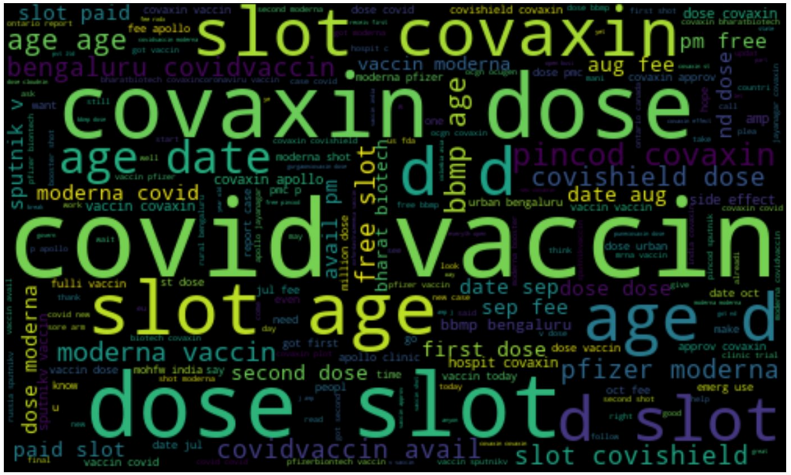
Chinese vaccines include: Sinopharm, Sinovac
Non-Chinese Vaccines generally include: Pfizer/BioNTech, Sputnik, Astra, Moderna and Covaxin.

The word cloud for
Chinese Vaccines vs. All other vaccines

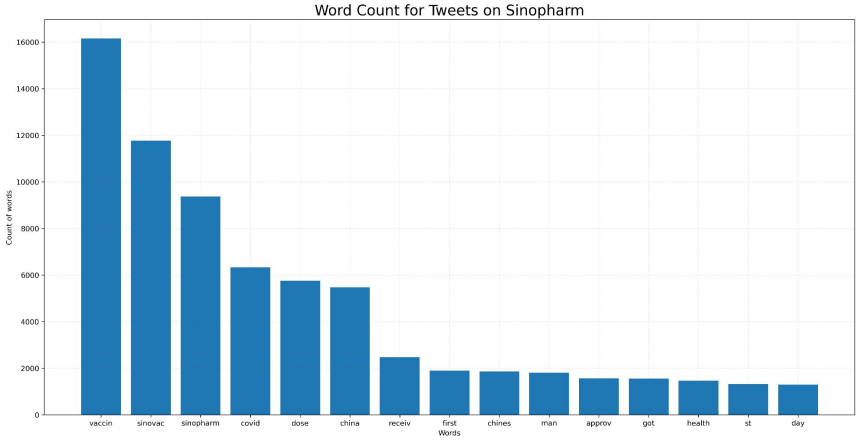
The word cloud for Sinopharm

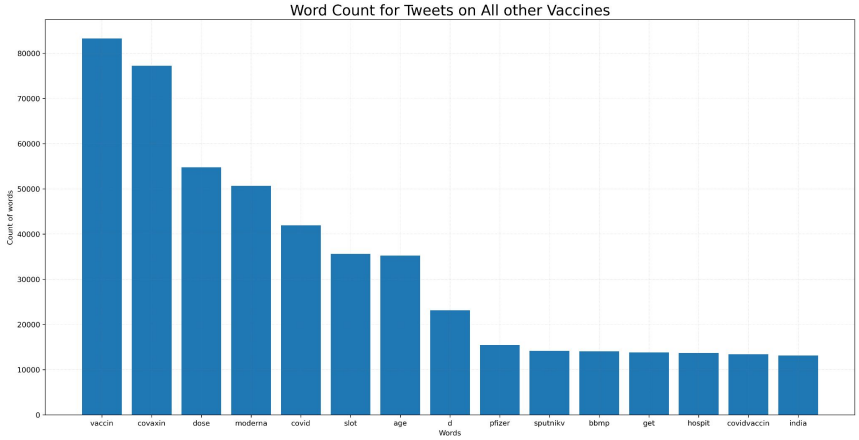


The word cloud for All other Vaccines

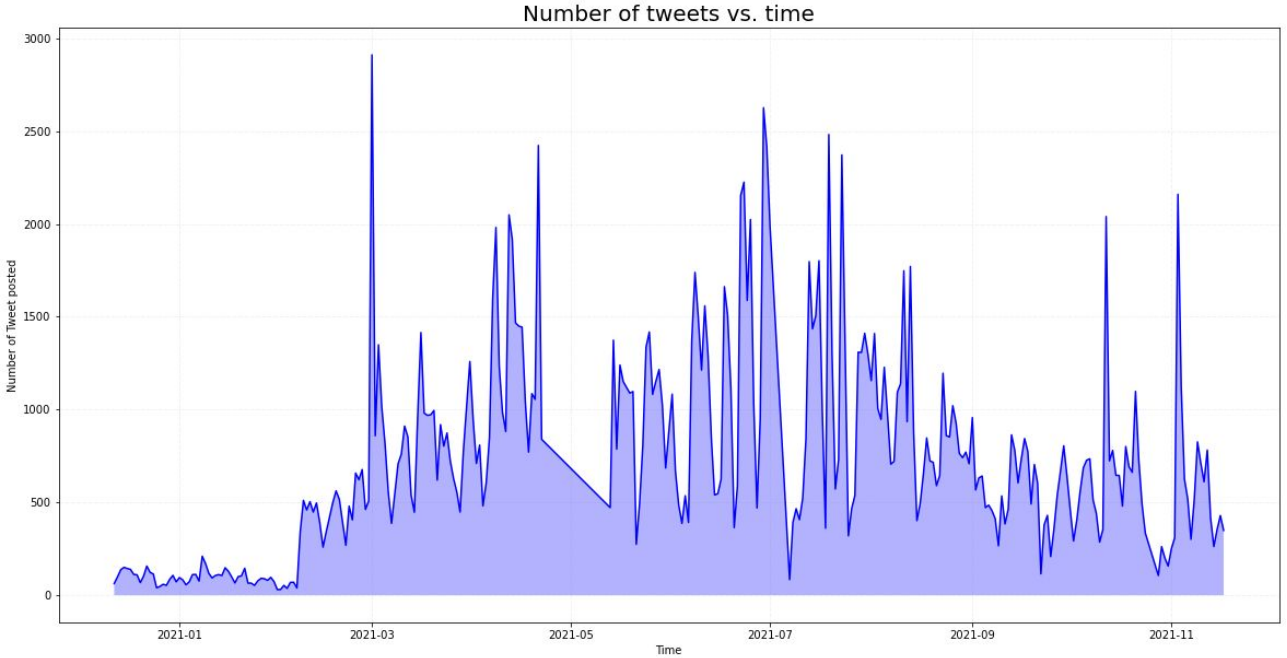


The word count for
Chinese Vaccines vs. All other vaccines



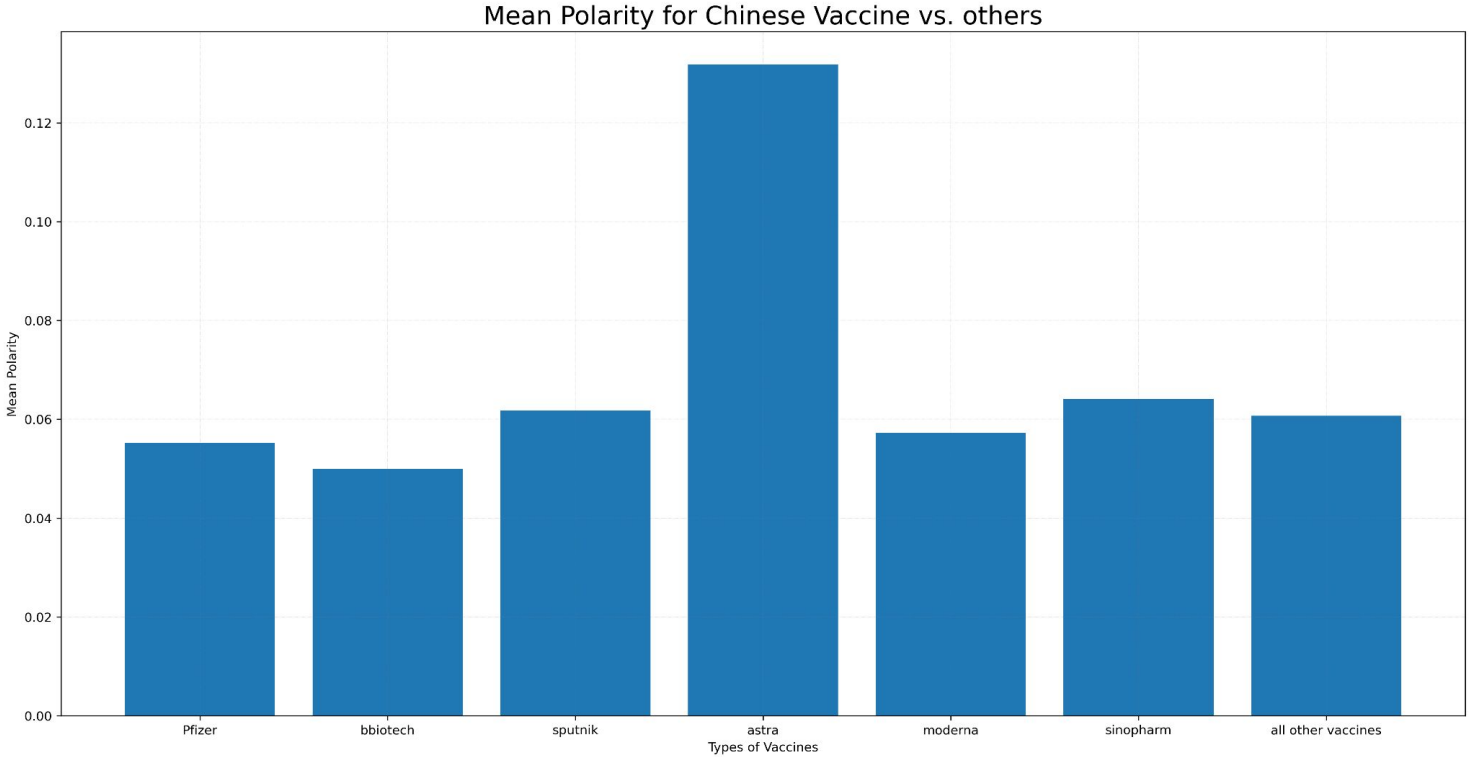


Frequency of tweets plotted against time

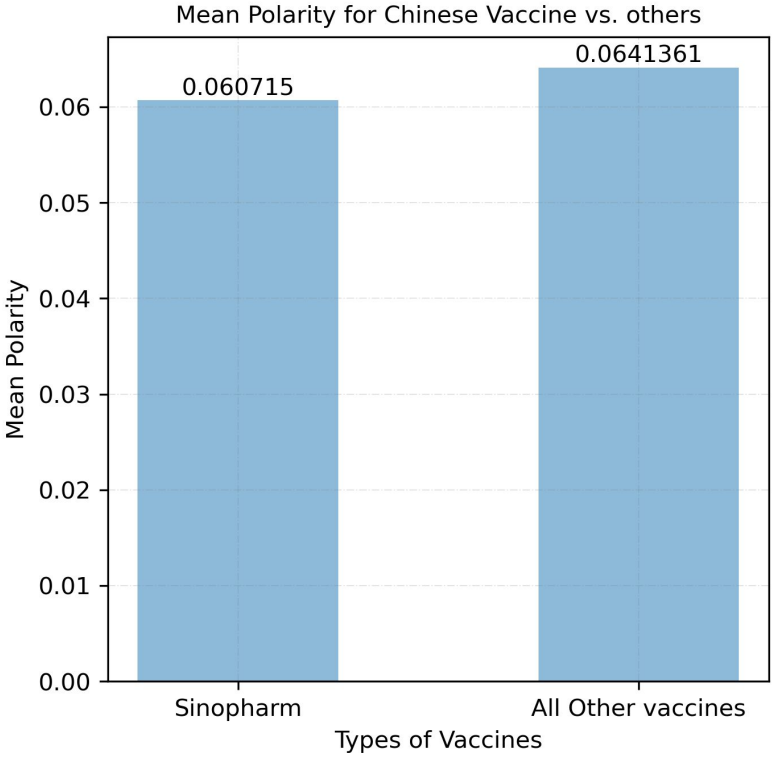


The announcement made by J&J on March 1, 2021, saying that the first batch of vaccines would be sent to states and pharmacies and distributed to the public.

Comparing mean polarity for different types of vaccines



Mean polarity for Chinese Vaccine vs. all other vaccines



We can see that there is a certain degree of bias against Chinese vaccines.

Do tweets about mRNA vaccines differ from those about non-mRNA vaccines?

mRNA: Pfizer/BioNTech, Moderna
non-MRNA: Sinopharm, Sinovac, Astrazeneca, Covaxin, Sputnik V

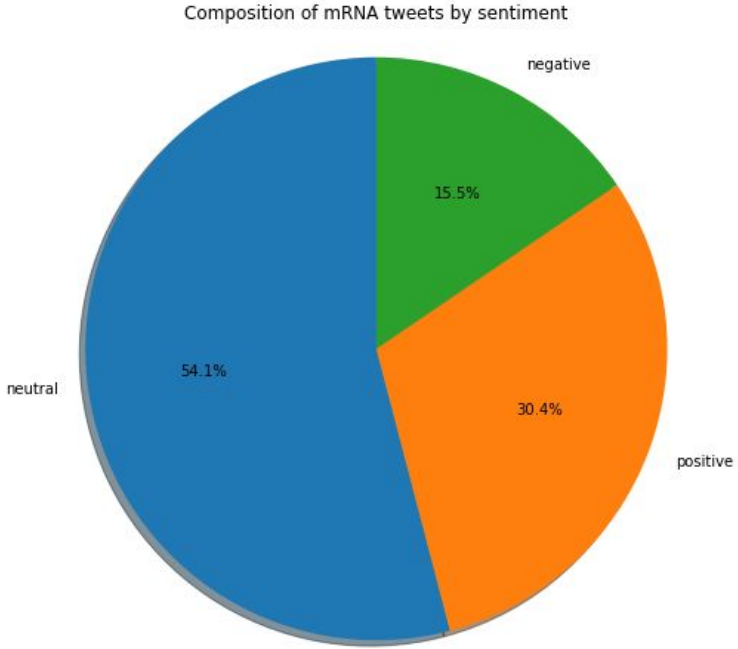
Hypotheses

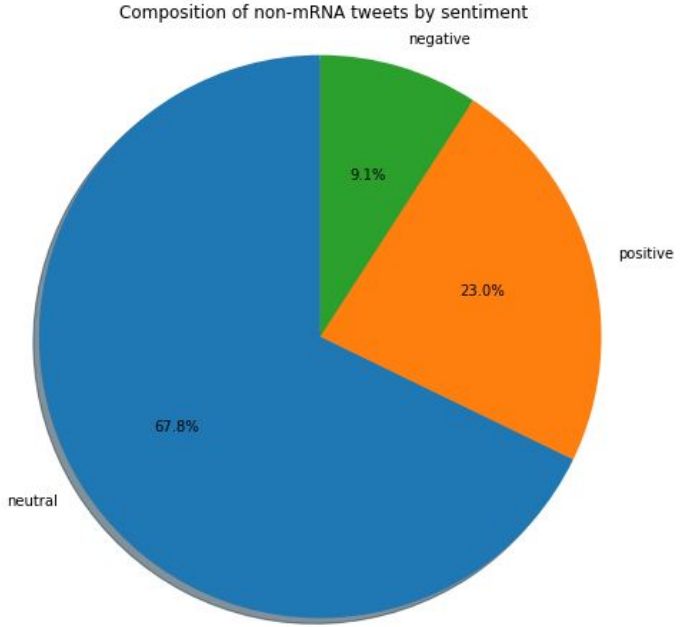
H_0: No substantive difference

H_1: mRNA more likely to cause vaccine hesitancy -> Lower sentiment score for mRNA vaccines

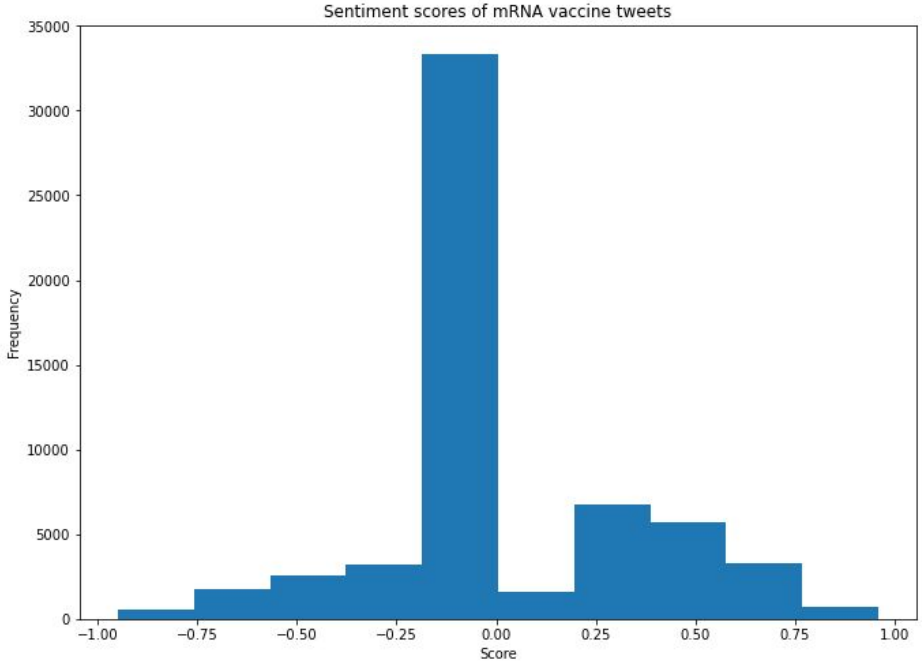
H_2: mRNA more novel -> mRNA causes more excitement -> Higher sentiment score for mRNA vaccines

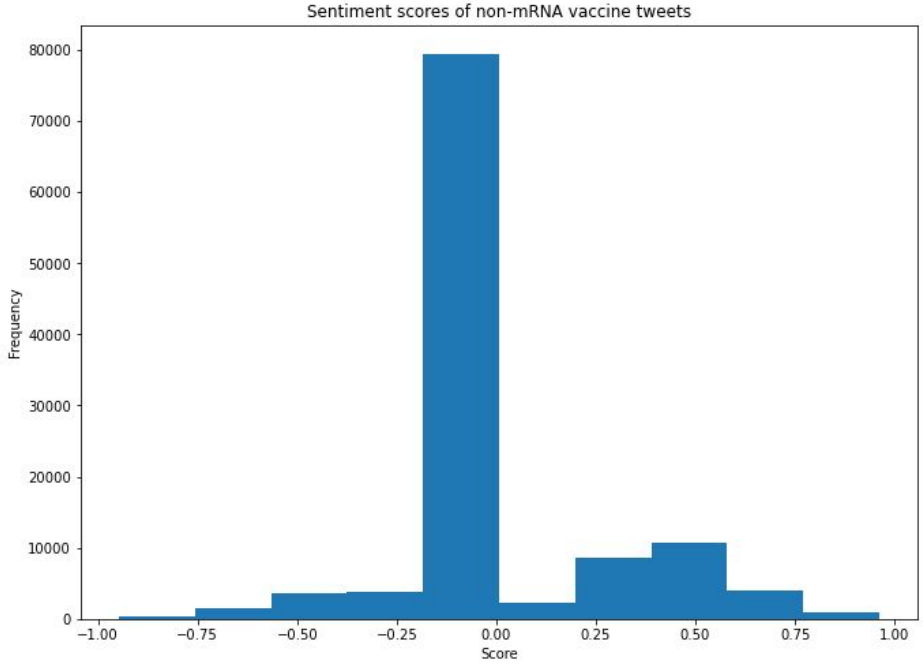
Sentiment Analysis: VADER



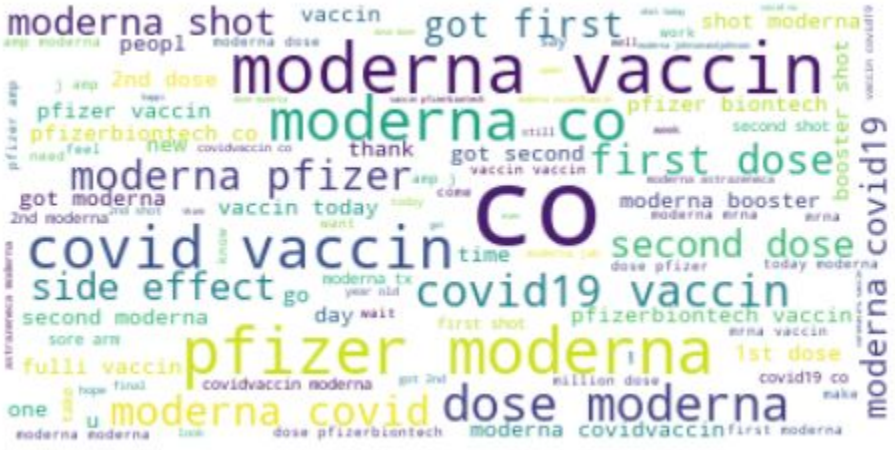


Sentiment Analysis: VADER

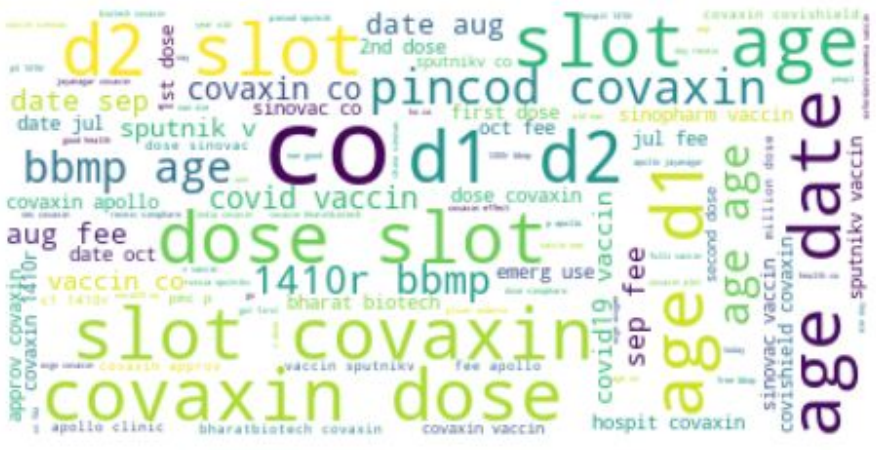




Word Clouds

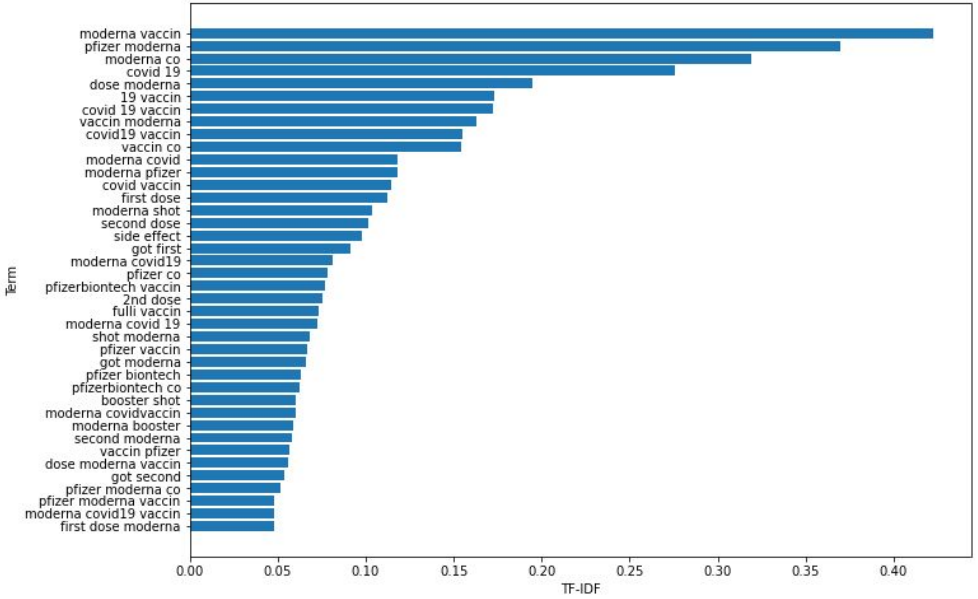


mRNA

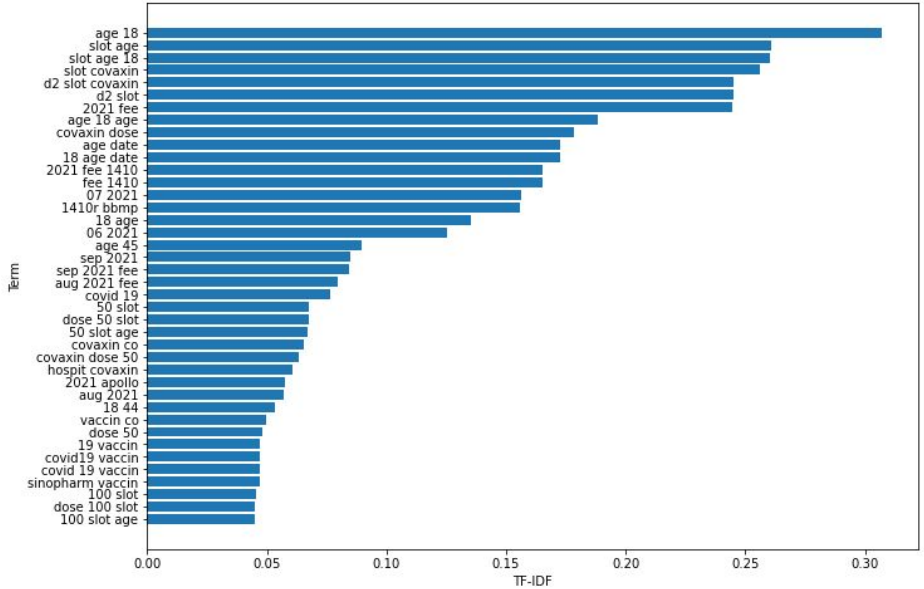


Non-mRNA

TF-IDF

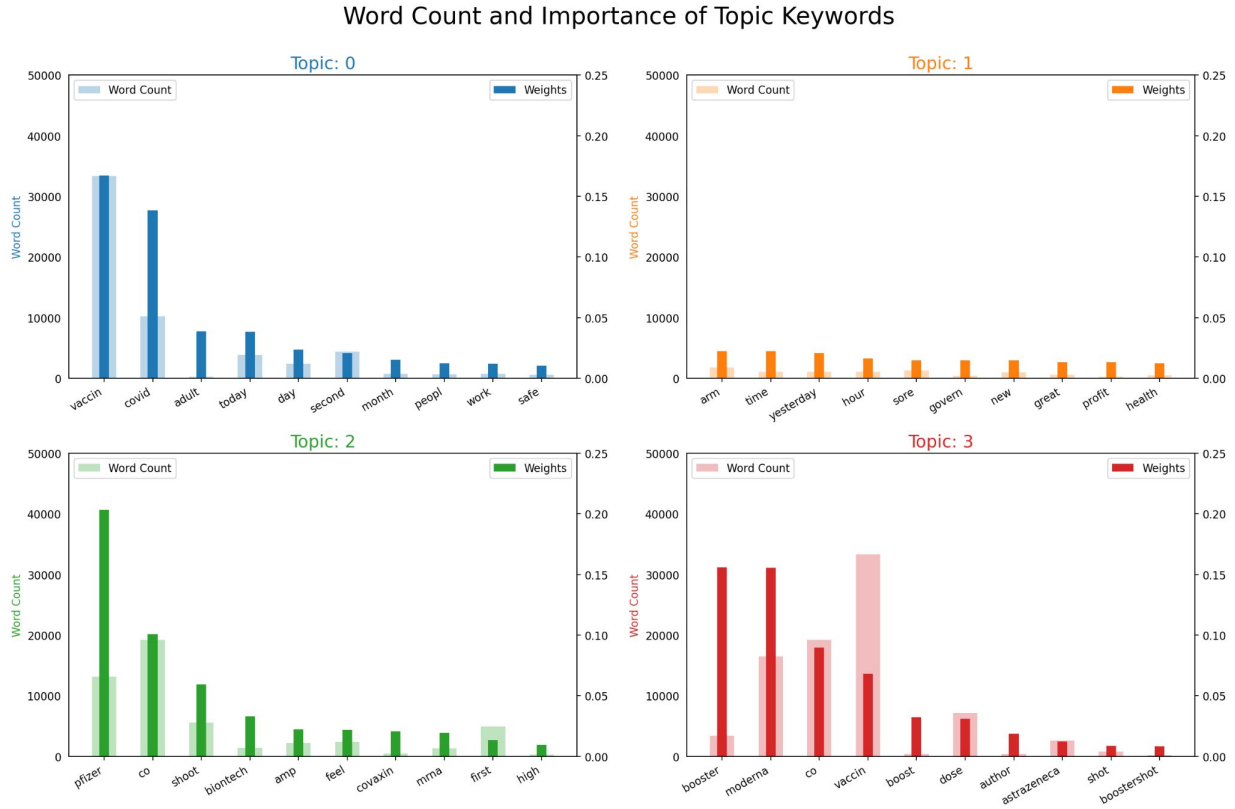


mRNA

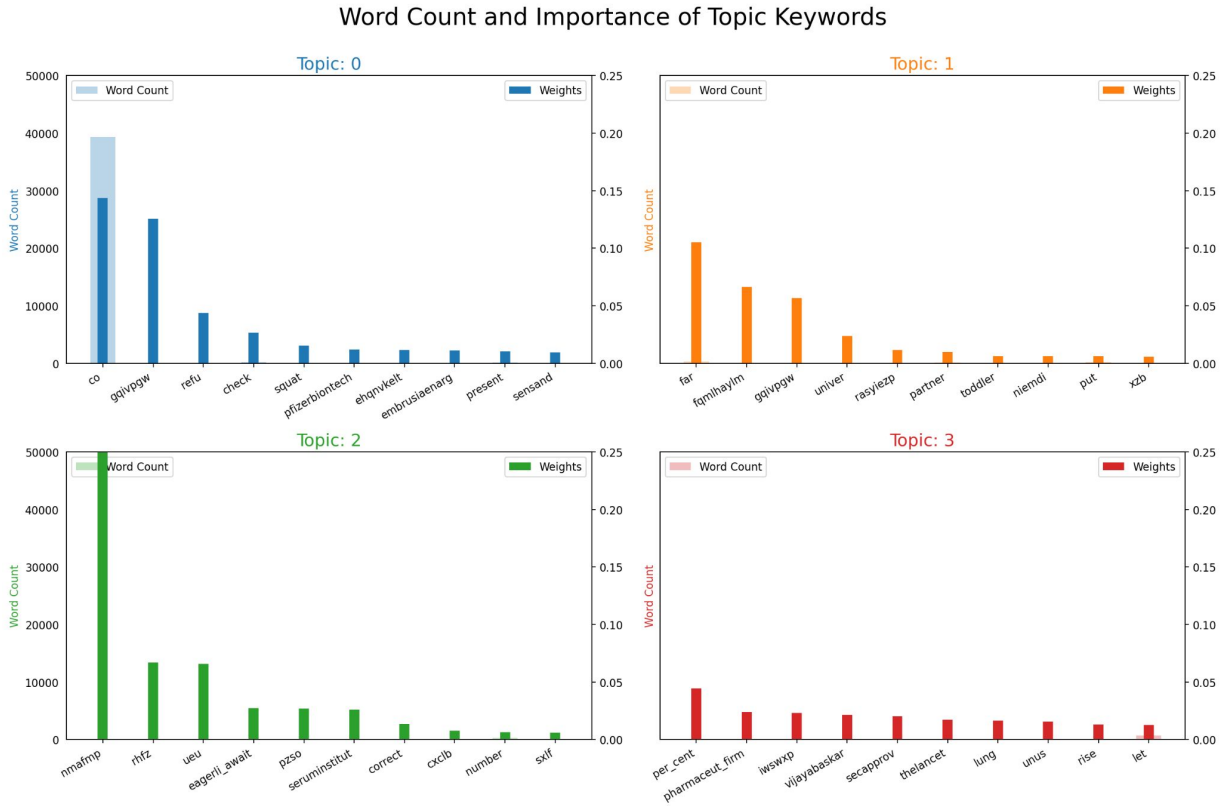


Non-mRNA

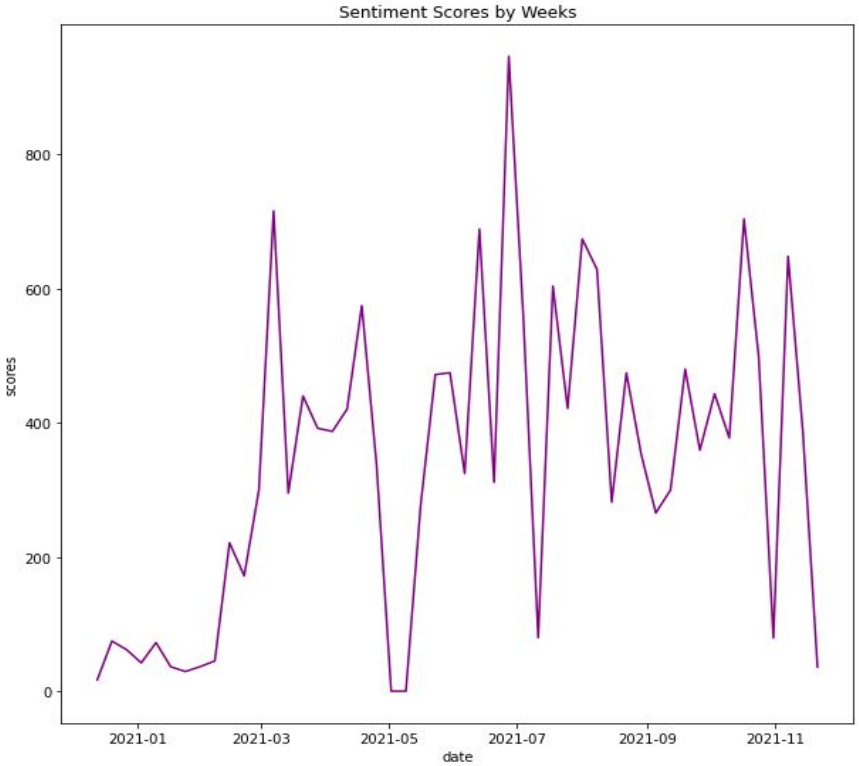
Topic Modeling (LDA): mRNA tweets

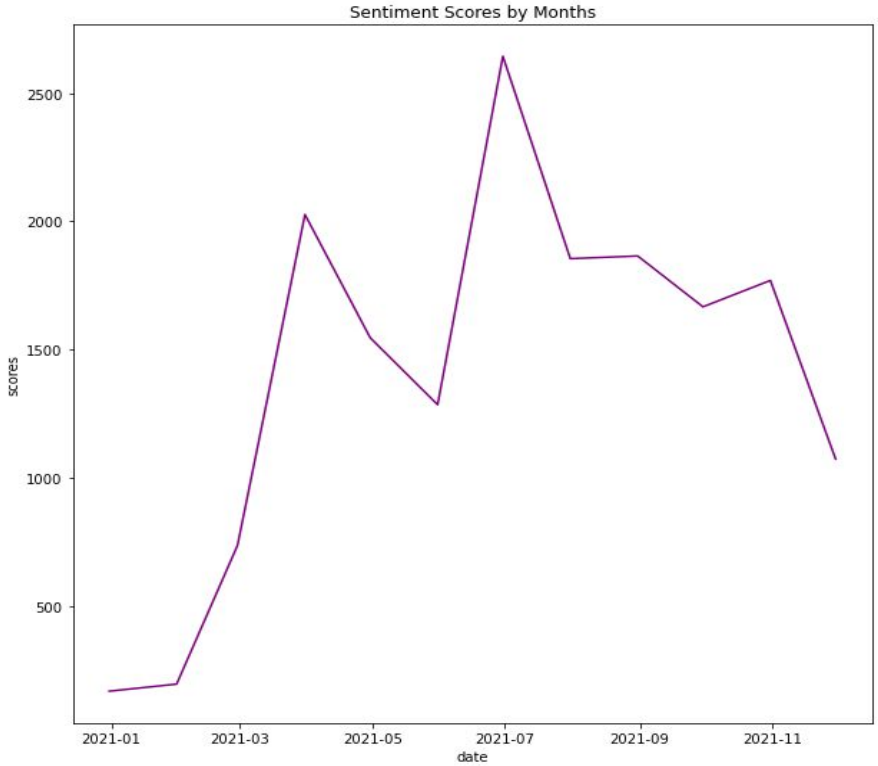


Topic Modeling (LDA): non-mRNA tweets



Sentiment Changes in Time





Conclusion

1. There is indeed some difference; H_0 likely to be rejected

2. Difference in sentiment scores are likely significant

3. A lot of noise still remains. Ways to improve analysis:
    a. More precise way of subsetting the sample
    b. Removing spam/tweets by bots
    c. Using specific dictionary (e.g. emotions)