

Analysis of Pfizer Vaccine Tweets

Group 15

Wendy Yu, Asfiya Baig, Wenchang Wang, Zhuo Chen

Motivation and Objective

- The breakout of Coronavirus disease(COVID-19)
- The Pfizer COVID-19 vaccine leads to much discussion on social media.
- Goal of this project: collect information and analyze the tweets people posted.

Dataset

- The Pfizer Vaccine Tweets dataset created by Gabriel Preda
- Contains 6281 tweets collected between 12/12/2020 and 02/22/2021.
- Also contains information about the user and the tweet, such as name and location of the user, hashtags, date of the tweet.
- Data cleaning:
 - Convert all words to lowercase
 - Remove punctuations, URL links, mentionings from tweets

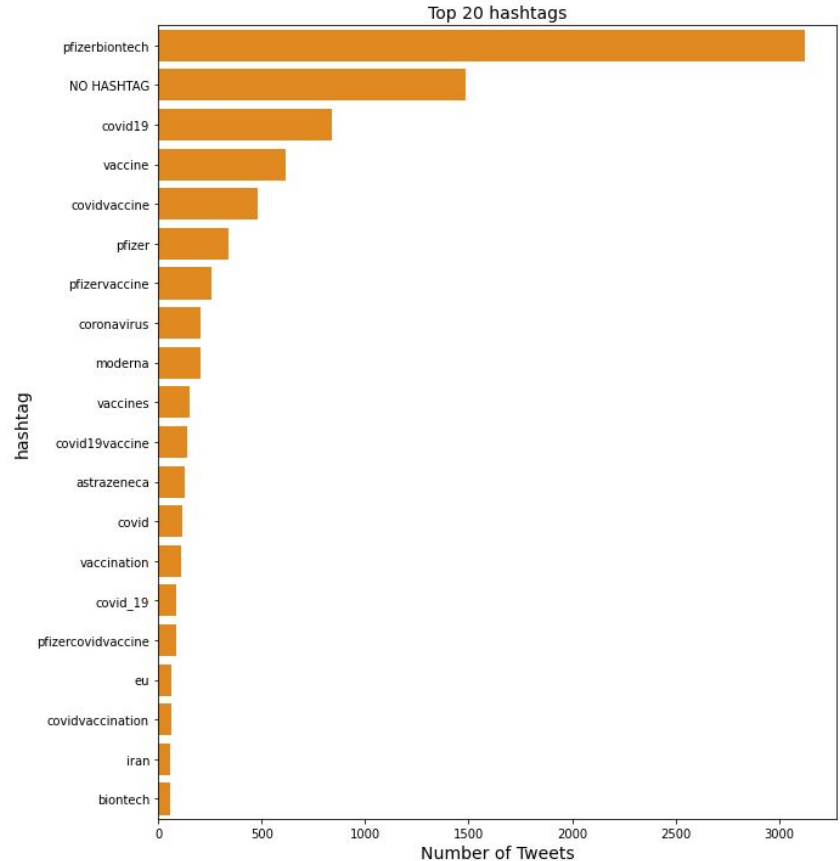
Methodology

- Exploratory data analysis
- Sentiment analysis
- Word embedding analysis

Exploratory Data Analysis

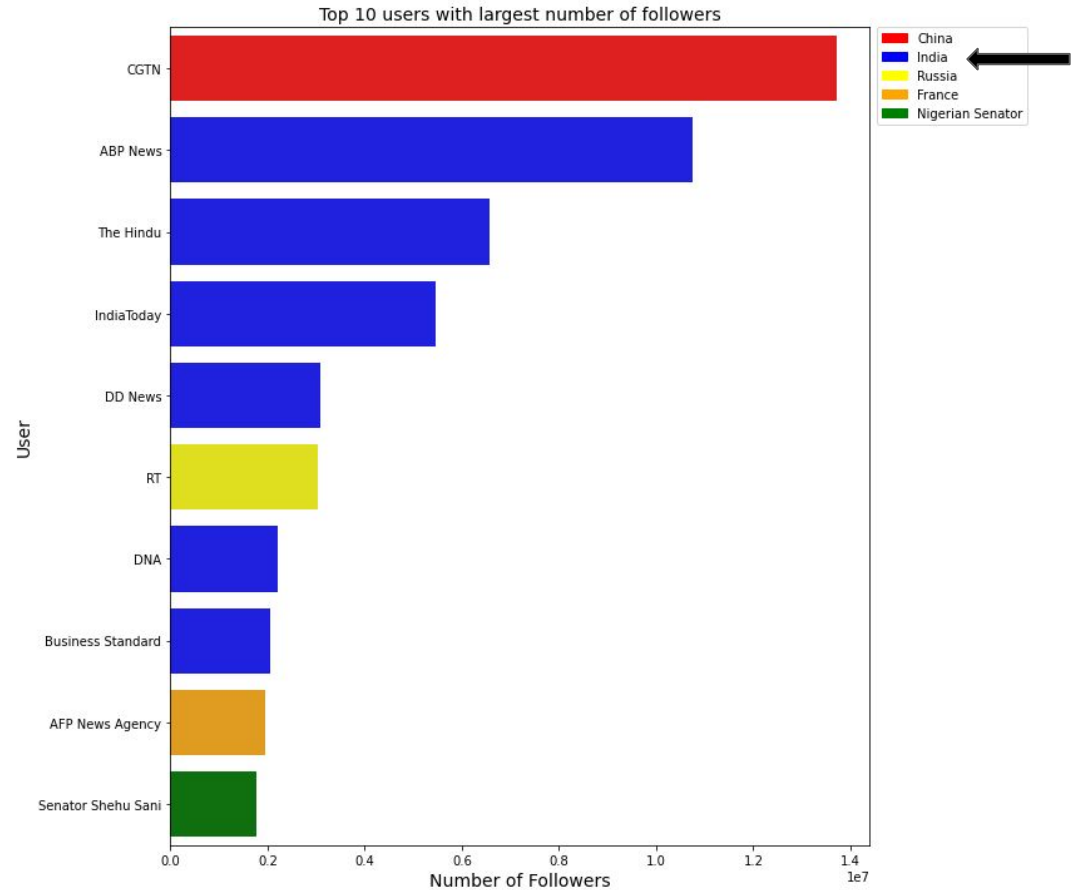
Most Common Hashtags

- Most used hashtag: pfizerbiontech



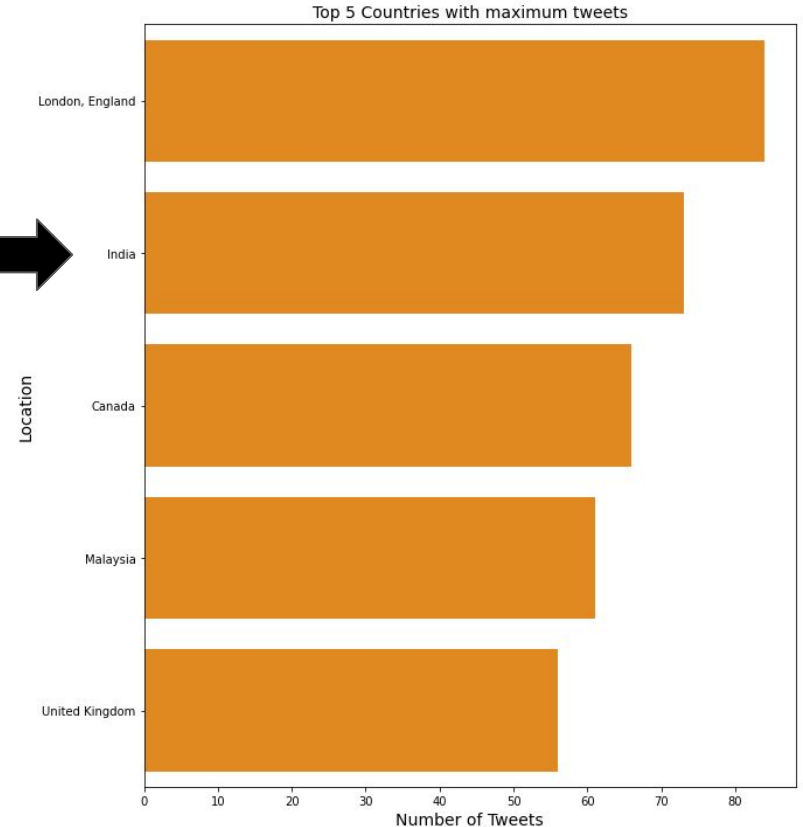
Users with Largest Number of Followers

- 9 of them are news channels
- 6 news channels from India

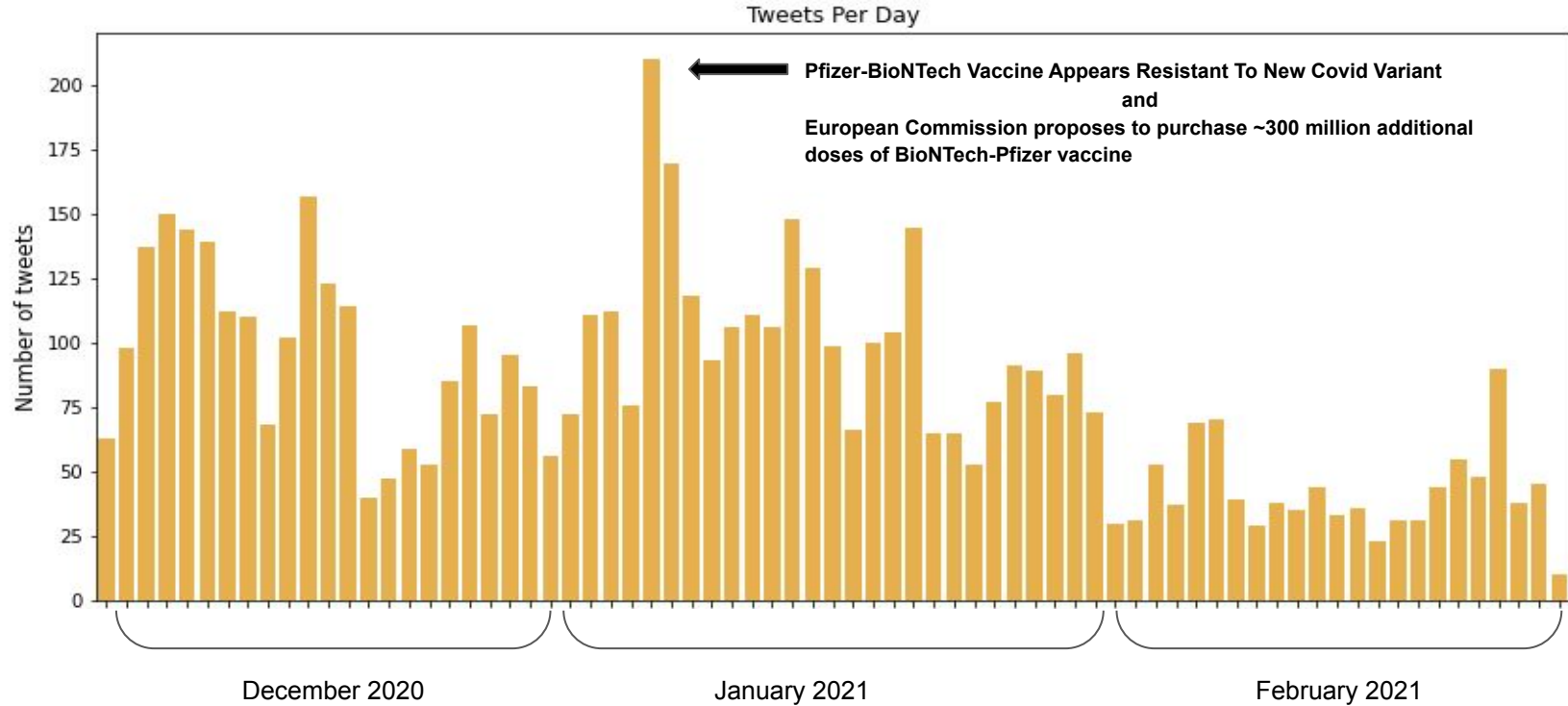


Locations with Largest Number of Tweets

- India: One of the top 5 countries



Tweets over Time



Sentiment Analysis

VADER sentiment

For example:

“There have not been many bright days in 2020 but here are some of the best”

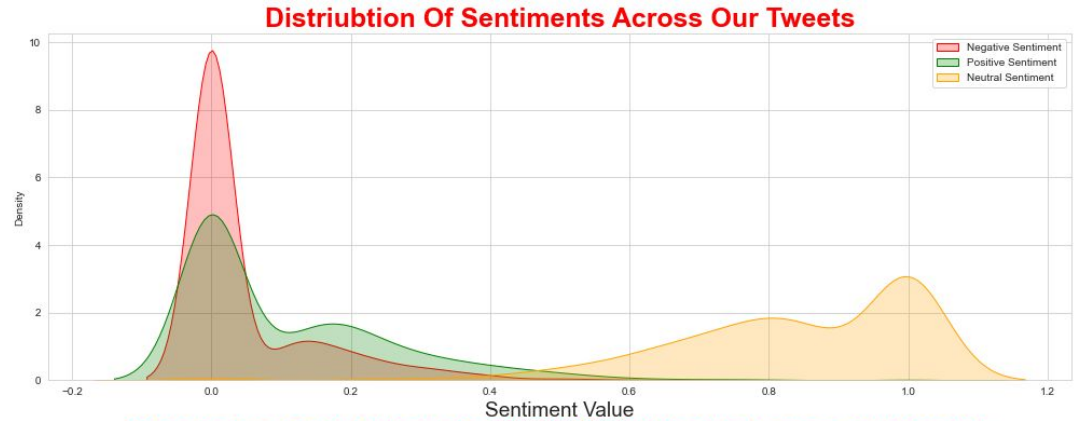
Positive : 39.8%

Negative : 53.6%

Neutral : 6.5%

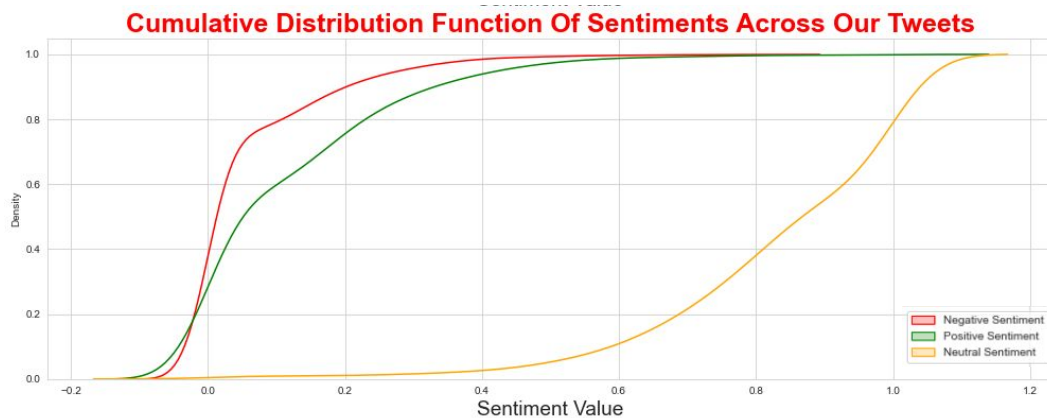
VADER sentiment

- Normal distribution
- Most tweets are neutral



VADER sentiment

- Normal distribution
- Most tweets are neutral



Textblob Sentiment Analysis

For example:

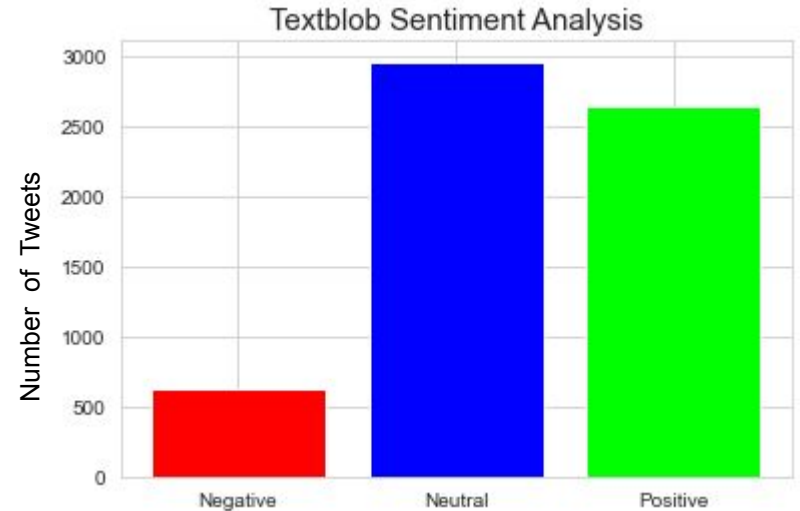
“There have not been many bright days in 2020 but here are some of the best”

Subjectivity : 0.5875

Polarity : 0.675

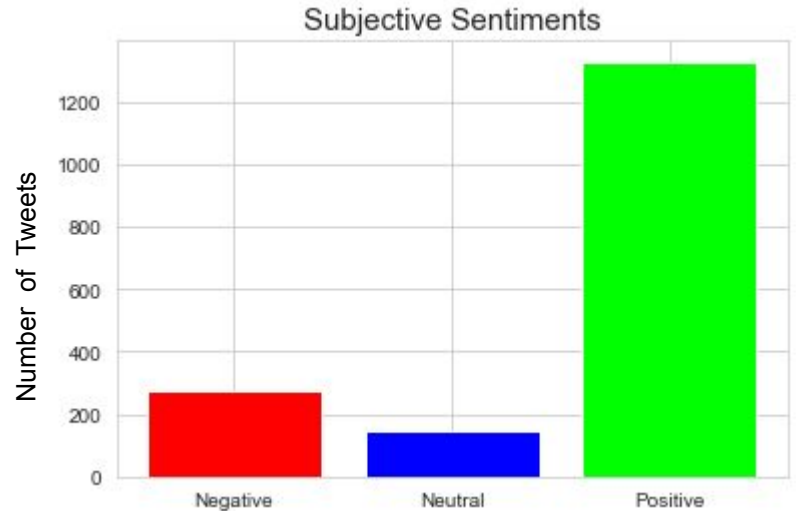
Sentiment Analysis

- Most tweets are neutral



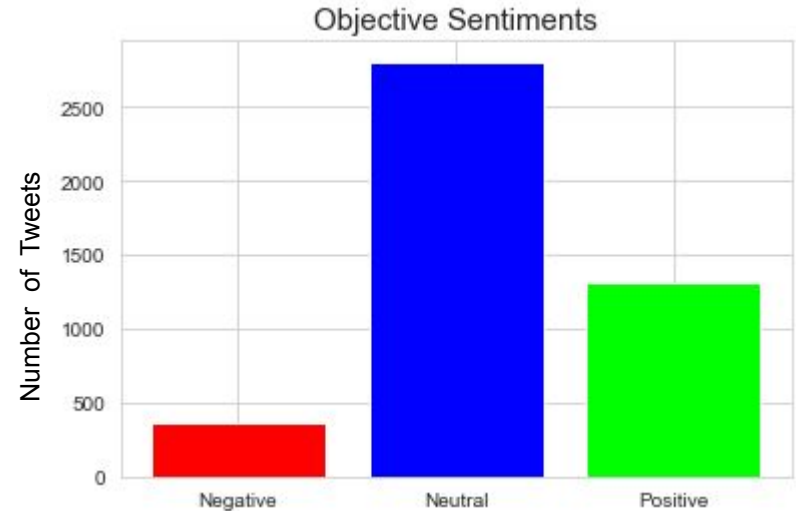
Sentiment Analysis: Subjective Tweets

- Subjective tweets tend to be positive



Sentiment Analysis: Objective Tweets

- Objective tweets tend to be neutral

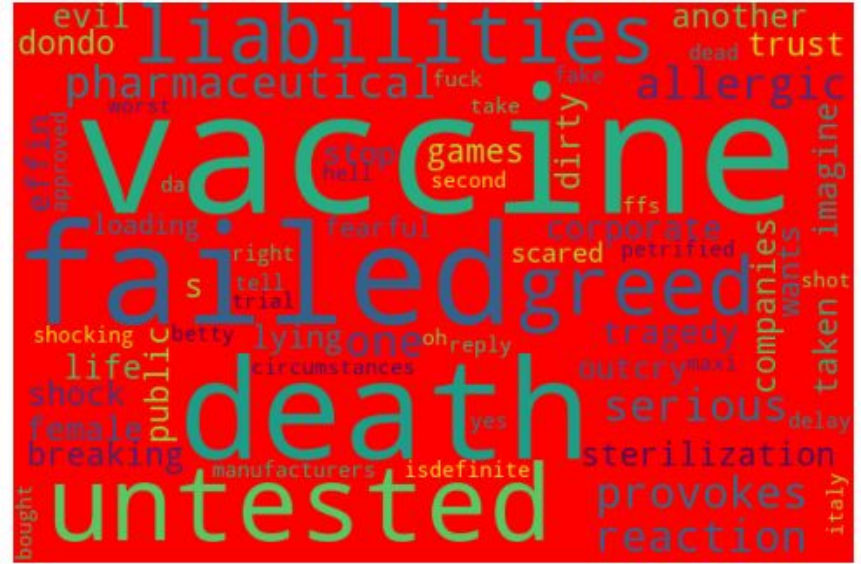


Wordcloud: Most Positive and Negative **TWEETS**

Common Words Among Most Positive Tweets



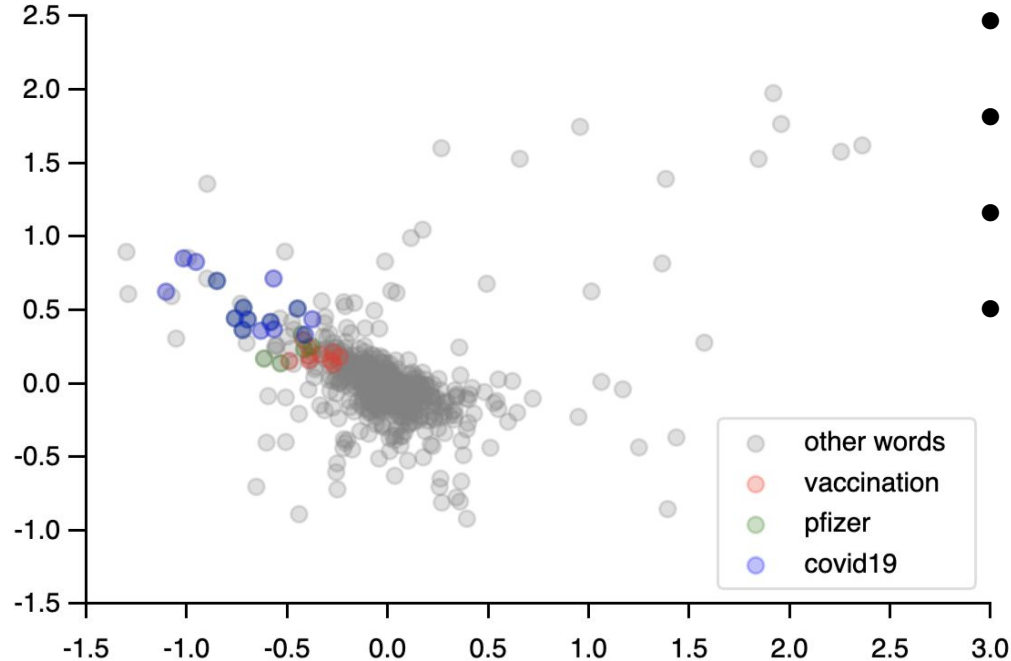
Common Words Among Most Negative Tweets



The word cloud features a variety of terms related to the COVID-19 pandemic and vaccine development. The most prominent words are 'PfizerBioNTech' in large blue letters, 'CovidVaccine' in large yellow letters, and 'COVID19' in large blue letters. Other visible words include 'vaccine', 'vaccination', 'coronavirus', 'PfizerVaccine', 'AstraZeneca', 'Moderna', 'IgotTheShot', 'VaccinesSaveLives', 'COVID19Vaccine', 'vaccine', 'vaccination', 'coronavirus', 'PfizerVaccine', 'AstraZeneca', 'Moderna', 'IgotTheShot', 'VaccinesSaveLives', 'COVID19Vaccine'.

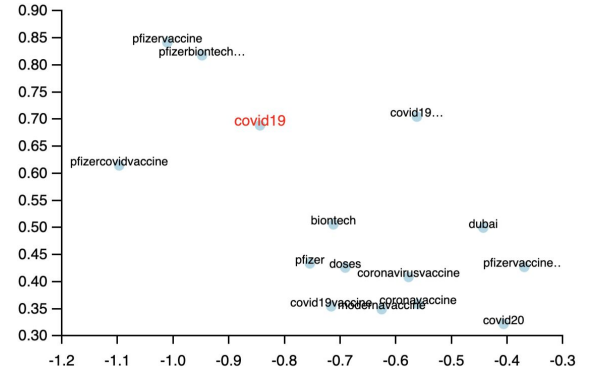
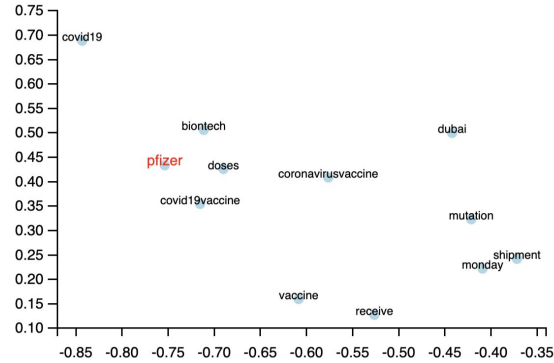
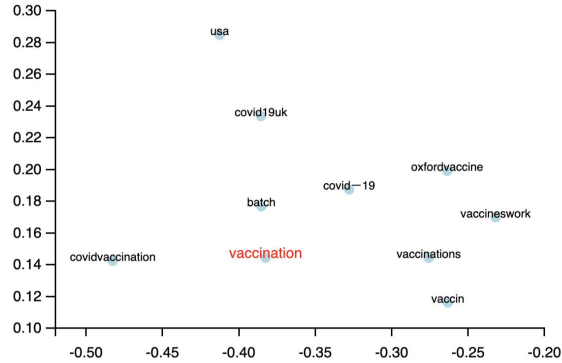
Word Embedding Analysis

Word Embeddings: PCA analysis of the Embedding values



- Word embedding is used for representing the word as a real-valued vector.
- The vector encodes the meaning of the words
- Train the Word2Vec word embedding model
- PCA: reduce from 100 to 2

Word Embeddings: PCA analysis of the Embedding values



- Vaccination: covid-19, usa, covidvaccination
- Pfizer: doses, covid19vaccine, biontech, receive, mutation, shipment
- Covid19: pfizercovidvaccine, biontech, pfizer, doses
- Tweets: abbreviation, emoji

Thank You