

Process Book

Abstract

For this project, we explored two datasets -- tweets about the Covid-19 vaccine and the reported side effects of the vaccine. Firstly, we looked into what people are discussing when they tweet about the Covid-19 vaccine and explored their feelings about it. This would give us some insight into people's general attitudes towards it. Secondly, we focused on the adverse reactions reported from 2020-12-01 to 2021-3-31. The visualization aims to provide an insight into who are the people reporting side effects and how do they compare to the general population; the most common reported symptoms etc. Altogether, our project aims to provide insights to guide more appropriate actions in promoting Covid-19 vaccination, as well as more effective actions in informing and relieving side effects.

PART I

1. # Overview of Tweet Content

What Do People Tweet about Covid-19 Vaccines?

Data: Text content of all the tweets about Covid-19 vaccines.

Process: Clean and Lemmatization text data of tweets about vaccines, and use the hunspell package to get stems as complete words. Make a word cloud to present most commonly mentioned keywords.

Conclusion: This word cloud includes the popular keywords (appeared more than 600 times) used in tweeting about Covid-19 vaccine.

The most commonly mentioned word is of course "vaccine", followed by "moderna" and "covid", while "pfizer" and "pfizerbiontech" are much smaller.

We can also see some common keywords seemingly describing experiences "dose", "receive", "today", suggesting many of these tweets may be recording people's vaccination experiences.

There's some discussion about China ("china" and "chinese"), because Chinese also manufacture and hand out Covid-19 vaccine.

It's worth noting that "sore" and "side" also appears a lot, so maybe a couple of people suffering side effects.

A Co-occurrence Network of Keywords

Data: Keywords extracted from last step.

Process: Keep only the top 1000 keywords in terms of occurrence and calculate the times of co-occurrence for each pair of words. Keep only the ties indicating more than 100 co-occurrences, and make a network to exhibit the connections between those commonly mentioned keywords. Walk trap algorithm is used to detect word clusters.

Conclusion: There are four major clusters detected – one is the major group with two central points – “vaccine” and “covid”; another is one around “moderna” the manufacturer, which may come from tweets reporting new progresses of moderna vaccine; another one is more dispersed with three centers – “today”, “russia” and “antario”, which may come from those focus on vaccine exportation news; the other one at the intercept is more dispersed and doesn’t have a central term. The clusters are interwoven together, but can offer some hints on different popular topics. Readers can freely explore the network and look for the relevant words they are interested in.

2. #Sentiment Analysis

Sentiment State Distribution of Tweets

Process: Clean and Lemmatization text data of tweets about vaccines. Vader sentiment analysis.

Conclusion: 1. Most tweets about vaccines are neutral one or positive one. Negative sentiment is not widely available.

2. Trends over time of numbers of tweets posted of three sentiment types are similar.

3. After seeing the common words of positive, neutral and negative tweets, we find people share their happiness about the arrival of vaccines and give positive feedback after receiving a shot in positive tweets; neutral tweets are just objective statements of vaccines news or information; people worry about the side effect of vaccines and whether vaccines will work in negative tweets.

Portrait of Popular Tweets and Users

To study the possible influence of sentiment attribute of popular tweets and users on twitter, we made these portraits.

Data: Base on classification result of sentiment analysis

Conclusion: 1. We can see sentiment attributes of popular tweets (based on favorite times). Most of the top 15 popular tweets are neutral one or positive one, which means that people didn't show a preference for negative tweets.

2. We can see the sentiment attribute of popular tweets (based on retweeted times). Most of the top 15 popular tweets still are neutral one or positive one, which means that people didn't show a preference for retweeting negative tweets and maybe kept a positive attitude towards the effect of vaccines.

3. We also check tweets of popular users (based on number of followers they have), because they have great influence among the public. Most of these users mainly post objective statements of vaccines news or information and they show more positive sentiment than negative sentiment.

PART II

1. # Who are the people reporting adverse reactions?

AGE AND GENDER

Do elders suffer more from side effects? Not exactly

Conclusion: Women and Younger people tend to report more cases. Is it possible that fewer elders got vaccinated thus fewer reports?

We decided to dive deeper into who got vaccinated by looking at different age groups.

Vaccinated Rate by Different Age Group

Figure: (by age group) Percentage of People that Have Received at Least One Dose of Cov Vaccine, by Mar 31, 2021

Type: plotly interactive line chart

source : <https://covid.cdc.gov/covid-data-tracker/#datatracker-home>

Report rate by different age group (animated bar? to show the changes by time)

Pre-illness

most common illness (wordcloud) (allergy is actually a big category, what are the common allergies mentioned?)

common allergies (there are all kinds of allergies containing other allergy, eg: food allergy, nut allergy or words in reverse order, eg: penicillin allergy, allergy penicillin. Don't really know how to handle it .give up △)

2.#When did side effects kick in ?

(bar chart, taking average value for different age group and sex)

(tried line charts, but it took extreme values into consideration, the displaying is not very nice. I guess average makes more sense, instead of showing value for every age)

3.#What are the side effects symptoms?

top 10 common symptoms (bar chart)

##emotional words in symptoms description(how do people feel) (facet)

4.# Where are the reports from?

allocation rate by state

This part uses a bar chart and map to show the allocation of vaccines of different brands in different states before 2021-03-31.

Source:

<https://data.cdc.gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/saz5-9hgg>

<https://data.cdc.gov/Vaccinations/COVID-19-Vaccine-Distribution-Allocations-by-Juris/b7pe-5nws>

Conclusion:

The number of vaccine allocations in each state does not have a brand tendency. In every state, the number of vaccine allocations for the two brands is basically the same.

report rate by state

This part uses a bar chart and map to show the side effect case report rate (from 2020-12-14 to 2021-03-31) in 50 states and the District of Columbia and the relationship between report rate and 2020 election result.

The report rate was calculated by dividing the number of cases in the VARES data set by the number of people vaccinated in the daily administered data set.

Source:

<https://covid.cdc.gov/covid-data-tracker/#datatracker-home>

VARES data set:

<https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions>

The results of the state elections come from the data provided in the Week 5 lecture.

Conclusion:

The report rate in the northeast and northwest regions is higher than other regions. The reporting rate of vaccine side effects in each state does not seem to be significantly related to the party's victory in the 2020 election. But New York has the highest reporting rate, more than double that of Montana, the second highest.

Cases by manufacturer

This part uses bar chart and map to show the number of reported side effect cases of different brands of vaccines (from 2020-12-14 to 2021-03-31) in 50 states and the District of Columbia

Source:

VARES data set:

<https://www.kaggle.com/ayushggarg/covid19-vaccine-adverse-reactions>

Conclusion:

According to previous visualizations, there is no significant difference in the number of vaccine allocations between Moderna and Pfizer in each state. However, Pfizer's vaccine has more reported cases of side effects.