

# What people concern about covid-19 vaccines ?

## Table of content:

1. Introduction
2. Data inspection and cleaning
3. Overall thoughts of vaccine
4. Top 20 popular questions that people asked the most
5. Vaccine type investigation
6. References

## 1. Introduction

Covid-19 is one of the most catastrophic pandemics in human history, which causes millions of deaths worldwide and has a detrimental impact on many aspects of life such as economy, physical/mental health, transportation, etc. Therefore, vaccination became extremely urgent during the peak of the pandemic. During a period of two years (2020-2021), a variety of vaccines were produced, which made a rapid improvement to the disease situation. However, the process of vaccine development and clinical trials in a short period of time has raised many doubts and even objections from many people. In this report, we analyse questions taken from The COVID-19 Vaccine Intent Expressions dataset, to gain insights about what people concern about covid-19 vaccines.

## 2. Data inspection and cleaning

### 2.1 Data understanding

The **COVID-19 Vaccine Intent Expressions** dataset contains 7,990 varying expressions for 181 common questions about COVID-19 vaccines. This means that each expression is annotated with one of 181 labels. Besides, 324/7,990 expressions from the dataset are taken from a chatbot from Johns Hopkins that answers COVID-19 vaccine questions.

### 2.2. Preliminary data inspection

First of all, we go through the dataset to what it looks like

#### [2.2.1] Load all the packages

In [1]:

```
# Import all packages
import pandas as pd
import numpy as np
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
import string
import re
import matplotlib.pyplot as plt
import math
```

```

from textblob import TextBlob
from textblob import Word

import nltk
from collections import Counter
from nltk.corpus import stopwords
from nltk.corpus import wordnet
from nltk.tokenize import word_tokenize
from nltk.stem.wordnet import WordNetLemmatizer
nltk.download('punkt')

import gensim.downloader as api

from sklearn.feature_extraction.text import TfidfVectorizer

from wordcloud import WordCloud

import warnings
warnings.simplefilter(action='ignore', category=FutureWarning)

import plotly.io as pio
pio.renderers.default='notebook'

```

```

[nltk_data] Downloading package punkt to
[nltk_data]     /Users/nguyenthao/nltk_data...
[nltk_data]     Package punkt is already up-to-date!

```

**[2.2.2]** Because the dataset is split into 3 files *dev*, *test*, and *train* we need to combine them.

```

In [2]: # import data from csv file
dev_23 = pd.read_csv('./AT1B_Datasets/IBM_Debater_(R)_Intent_expressions/dev_23.csv')
test_23 = pd.read_csv('./AT1B_Datasets/IBM_Debater_(R)_Intent_expressions/test_23.csv')
train_23 = pd.read_csv('./AT1B_Datasets/IBM_Debater_(R)_Intent_expressions/train_23.csv')

```

```

In [3]: # concatenate 3 datasets
df = pd.concat([dev_23, test_23, train_23], ignore_index = True)

```

```

In [4]: # create a copy to clean
df_cleaned = df.copy()

```

**[2.2.3]** What are 181 common question labels ?

```

In [5]: df_cleaned['label'].unique()

```

```

Out[5]: array(['Are booster shot side effects worse than those from the second shot?',
   'Are regular safety measures enough to stay healthy?',
   'Are some vaccines safer for younger children than others?',
   'Are the side effects worse for the second shot',
   'Are there medical contraindications to the vaccines?',
   'Are women more likely to get worse side effects than men?',
   'COVID vaccines can be worse than the disease itself',
   'COVID-19 is not as dangerous as they say',
   'COVID-19 is over, why should I get the vaccine?',
   'COVID-19 vaccines cause brain inflammation',
   'Can I get COVID-19 from the vaccine?',
   'Can I get COVID-19 twice?',
   'Can I get a second dose even after a COVID exposure?',
   'Can I get other vaccines at the same time?',
   'Can I get swollen lymph nodes from the vaccine?',
   'Can I get the vaccine if I have Multiple Sclerosis?',
   'Can I get the vaccine if I have allergies?',
   'Can I get the vaccine if I have had allergic reactions to vaccines before?',
   'Can I have the vaccine as a Catholic?',
   'Can I have the vaccine if I'm allergic to penicillin?',
```

"Can I meet in groups after I'm vaccinated?",  
'Can I still get COVID even after being vaccinated?',  
'Can I take a pain reliever when I get vaccinated?',  
'Can children get the vaccine?',  
'Can my kids go back to school without a vaccine?',  
"Can my newborn become immune to COVID-19 if I'm vaccinated?",  
'Can other vaccines protect me from COVID-19?',  
'Can we change the vaccine quickly if the virus mutates?',  
'Can we choose which vaccine we want?',  
'Can you mix the vaccines?',  
'Did a volunteer in the Oxford trial die?',  
'Did one woman die after getting the J&J vaccine?',  
'Do I need the vaccine?',  
'Do I need to change my masking and social distancing practices depending on which COVID-19 vaccine I got?',  
'Do I need to continue safety measures after getting the vaccine?',  
'Do I qualify for the vaccine?',  
'Do children receive the same dose of Pfizer as adults?',  
'Do people become magnetic after getting vaccinated?',  
"Do the COVID-19 vaccines cause Bell's palsy?",  
'Do the mRNA vaccines contain preservatives, like thimerosal?',  
'Do the vaccines work in obese people?',  
'Do vaccinated people need to quarantine if exposed to COVID-19?',  
'Do vaccines work against the mutated strains of COVID-19?',  
'Do you have to be tested for COVID before you vaccinated?',  
'Do you need a social security number to get a COVID-19 vaccine?',  
'Do you need to be a U.S. citizen to get a COVID-19 vaccine?',  
'Does the COVID-19 vaccine cause autism?',  
'Does the Johnson and Johnson vaccine cause Rare Nerve Syndrome?',  
'Does the Pfizer vaccine cause heart problems?',  
'Does the Pfizer vaccine cause myocarditis?',  
'Does the vaccine cause impotence?',  
'Does the vaccine contain animal products?',  
'Does the vaccine contain eggs?',  
'Does the vaccine contain live COVID virus?',  
'Does the vaccine impact pregnancy?',  
'Does the vaccine prevent transmission?',  
'Does the vaccine work if I do not experience any side effects?',  
'How can I get the vaccine?',  
'How can I stay safe until I'm vaccinated?',  
'How common are vaccine side effects?',  
'How do I convince my family and friends to get the COVID-19 vaccine?',  
"How do I know I'm getting a legitimate, authorized vaccine?",  
'How do I protect myself indoors?',  
'How do I report an adverse reaction or side-effect',  
'How effective is the vaccine against the Omicron variant?',  
'How is the COVID-19 vaccine different than others?',  
'How long am I immune from COVID-19 if I had the virus?',  
'How long do I have to wait between doses?',  
'How long does the immunity from the vaccine last?',  
'How long until I will be protected after taking the vaccine?',  
'How many doses do I need?',  
'How many people already got the vaccine?',  
'How many people died from COVID-19?',  
'How much will I have to pay for the vaccine',  
"How soon after I've had COVID-19 can I get the vaccination?",  
'How was the vaccine tested?',  
'I am afraid the vaccine will change my DNA',  
'I am concerned I will be a guinea pig',  
'I am concerned about getting the vaccine because of my medications.',  
'I am concerned getting the vaccine because I have a pre-existing condition',  
'I am not sure if I can trust the government',  
'I am worried about blood clots as a result of the vaccine',  
"I am young and healthy so I don't think I should vaccinate",  
'I distrust this vaccine',  
"I don't think the vaccine is necessary",

"I don't trust the companies producing the vaccines",  
"I don't trust vaccines if they're from China or Russia",  
"I don't want my children to get the vaccine",  
"I don't want the v-safe app monitoring or tracking me",  
"I don't want to share my personal information",  
'I think the vaccine was not tested on my community',  
"I'm concerned the vaccine will make me sick.",  
"I'm going to get vaccinated",  
"I'm not sure it is effective enough",  
"I'm still experiencing COVID symptoms even after testing negative - should I still take the vaccine?",  
"I'm waiting to see how it affects others",

"If I live with an immuno-compromised individual, do I still need to wear a mask outdoors if I'm vaccinated?",

'Is breastfeeding safe with the vaccine',  
'Is getting vaccinated painful?',  
"Is it okay for me to travel internationally if I'm vaccinated?",  
'Is it safe for my baby to get the vaccine?',  
'Is it safe for my teen to get the vaccine?',  
"Is it safe to go to the gym indoors if I'm vaccinated?",  
'Is the Delta variant more dangerous for kids?',  
'Is the Johnson & Johnson vaccine less effective than the others?',  
'Is the Pfizer vaccine safe for young men?',  
'Is the booster shot dangerous?',  
'Is the booster the same as the original vaccine?',  
'Is the vaccine FDA approved?', 'Is the vaccine Kosher?',  
'Is the vaccine halal?', 'Is there vaccine safety monitoring?',  
'Is this Pfizer vaccine equally effective in kids as it is in adults?',  
'Long term side-effects were not researched enough',  
'Other vaccines have caused long-term health problems',  
'Should I get the COVID-19 vaccine if I am immunocompromised',  
"Should I get the vaccine if I've tested positive for antibodies?",  
'Should people that had COVID get the vaccine?',  
'Side effects and adverse reactions worry me',  
'Tell me about the vaccine', 'The COVID vaccine is not safe',  
'The mortality rate of COVID-19 is low, why should I get the vaccine?',  
'The vaccine includes fetal tissue or abortion by-products',  
'The vaccine should not be mandatory', 'The vaccine was rushed',  
"There are many reports of severe side effects or deaths from the vaccine",  
'They will put a chip/microchip to manipulate me',  
'Vaccine side effects are not getting reported',  
'Were the COVID-19 vaccines tested on animals?',  
'What about reports of abnormal periods due to the vaccine?',  
'What are the effects of long COVID?',  
'What are the side effect of the vaccine?',  
'What are the side effects of booster shots?',  
'What are the side effects of the vaccine in children?',  
'What can this chatbot do?',  
'What can you tell me about COVID-19 vaccines?',  
'What do I do if I lose my COVID-19 vaccination card?',  
'What does vaccine efficacy mean?',  
"What happens if there is a COVID-19 case at my child's school?",  
'What if I still get infected even after receiving the vaccine?',  
"What if I've been treated with convalescent plasma?",  
"What if I've been treated with monoclonal antibodies?",  
'What is Ivermectin?', 'What is in the vaccine?', 'What is mRNA?',  
'What is the J&J vaccine?', 'What is the Moderna vaccine?',  
'What is the Omicron variant?', 'What is the Pfizer vaccine?',  
'What is the delta variant?',  
'What is the difference between mRNA and viral vector vaccines?',  
'What is the difference between quarantine and isolation?',  
'What is the difference between the third shot and a booster shot?',  
'When can I go back to normal life?',  
'Where are we required to wear masks now?',  
'Which one of the vaccines should I take?',  
'Which vaccines are available?',

```
'Who can I talk to about COVID-19 in person?',
'Who can get the Pfizer vaccine?', 'Who developed the vaccine?',
'Who is required to get vaccinated under the federal vaccine mandate?',
'Why are COVID-19 vaccination rates slowing in the U.S.?',
'Why are there different vaccines?',
"Why do I need the COVID vaccine if I don't get immunized for flu",
"Why do my kids need a vaccine if they're unlikely to get sick with COVID-19?",
"Why do vaccinated people need to wear a mask indoors?",
"Why do we need the vaccine if we can wait for herd immunity?",
'Why get vaccinated if I can still transmit the virus?',
'Why is AstraZeneca not approved in the USA?',
'Why should I trust you?', 'Will 1 dose of vaccine protect me?',
'Will I need a booster shot?',
'Will I test positive after getting the vaccine?',
'Will my child miss school when they get vaccinated?',
'Will my child need my permission to get vaccinated?',
'Will the US reach herd immunity?', 'Will the vaccine benefit me?',
'Will the vaccine make me sterile or infertile?',
'Will there be a booster shot for J&J and Moderna?',
'Will vaccination lead to more dangerous variants?',
'what is covid?'], dtype=object)
```

#### [2.2.4] Information about each column and their non-null values.

There are 3 columns: sentence, label and label\_idx.

There is no missing values in this dataset.

```
In [6]: # Generate information: index dtype and columns, non-null values and memory usage
df_cleaned.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7972 entries, 0 to 7971
Data columns (total 3 columns):
 #   Column      Non-Null Count  Dtype  
---  --  
 0   sentence    7972 non-null   object  
 1   label       7972 non-null   object  
 2   label_idx   7972 non-null   int64  
dtypes: int64(1), object(2)
memory usage: 187.0+ KB
```

#### [2.2.5] Some sample rows of the dataset.

```
In [7]: # Look at the first n-rows
df_cleaned.head()
```

	sentence	label	label_idx
0	If I apply the third dose, will I have stronge...	Are booster shot side effects worse than those...	175
1	What effects do these reinforcements have?	Are booster shot side effects worse than those...	175
2	i dont think the vaccine is neccessary the bas...	Are regular safety measures enough to stay hea...	20
3	If I comply with the security measures, will I...	Are regular safety measures enough to stay hea...	20
4	This is like the flu, being cautious is enough.	Are regular safety measures enough to stay hea...	20

```
In [8]: # Look at random n-rows
df_cleaned.sample(100)
```

	sentence	label	label_idx
3755	I want to get vaccinated but do I have to wait...	Do I qualify for the vaccine?	32

6474	They did not test the vaccine enough.	The vaccine was rushed	70
4002	If I am not a resident, can I get vaccinated?	Do you need to be a U.S. citizen to get a COVI...	106
7097	I don't know what is the omciron variant	What is the Omicron variant?	179
4960	I am afraid of being used for tests	I am concerned I will be a guinea pig	6
...	...	...	...
5667	Are babies harmed by the vaccine if it goes th...	Is breastfeeding safe with the vaccine	61
6349	vacinne is safe?	The COVID vaccine is not safe	23
1513	What is the efficacy of the Covid vaccine on o...	Do the vaccines work in obese people?	46
2899	currently have the vaccines presented any cont...	Are there medical contraindications to the vac...	117
6432	People should be able to choose for themselves...	The vaccine should not be mandatory	24

100 rows x 3 columns

**[2.2.6]** There are 7972 rows and 3 columns

In [9]: `df_cleaned.shape`

Out[9]: `(7972, 3)`

## 2.3 Data cleaning

**[2.3.1]** Create a stopword list

In [10]: `stop_words=stopwords.words("english")  
stop_words.extend(["a", "the", "i'm", "i've", "i'll", 'us'])`

**[2.3.2]** Replace punctuations (except for apostrophes, hyphens, ampersand) from sentences and labels

In [11]: `# replace punctuations for sentences  
df_cleaned['sentence'] = df_cleaned['sentence'].apply(lambda s: re.sub(r'[^w\s\-\'\&]',`

In [12]: `# replace punctuations for labels  
df_cleaned['label'] = df_cleaned['label'].apply(lambda s: re.sub(r'[^w\s\-\'\&]', ' ', s))`

**[2.3.3]** Lowercase all words (except for apostrophes, hyphens, ampersand) in sentences and labels

In [13]: `# convert all words to lowercase (sentence)  
df_cleaned['sentence'] = df_cleaned['sentence'].apply(lambda s: ' '.join([w.lower() for`

In [14]: `# convert all words to lowercase (label)  
df_cleaned['label'] = df_cleaned['label'].apply(lambda s: ' '.join([w.lower() for w in s`

**[2.3.4]** This step look for all approved vaccine types in the datasets. These vaccines types are *Pfizer*, *Moderna*, *Novavax*, *Johnson & Johnson/J&J*, *Astra Zeneca*, *Sinopharm*, *CoronaVac*, *Covaxin*, *Sputnik V*. Any vaccine found will be corrected to the right and consistent format. For example, *Johnson and Johnson*, *Johnson & Johnson*, *J & J* are refering to the same thing, so they are all converted to *johsonjohnson*.

```
In [15]: # identify all vaccine names (Pfizer, Moderna, Novavax, Johnson & Johnson/J&J, Astra Zeneca)
question_text = ' '.join(df_cleaned['sentence'])
vaccine_matches = re.findall(r'\b(pfizer|moderna|novavax|johnson\s?and\s?johnson|johnson|johnson&johnson)\b')
set(vaccine_matches)

Out[15]: {'astrazeneca',
 'astrazeneca',
 'j & j',
 'j&j',
 'johnson',
 'johnson & johnson',
 'johnson and johnson',
 'johnson& johnson',
 'moderna',
 'pfizer'}
```

```
In [16]: # identify and correct covid-19 terms (covid 19, astra zeneca)
def identify_and_correct_vaccine_names(text):
    text = re.sub(r'\bastra\szzeneca\b', 'astrazeneca', text, flags=re.IGNORECASE)
    text = re.sub(r'\b(johnson\s?and\s?johnson|johnson\s?&\s?johnson|j\s?&\s?j|johnson)\b', 'johnson', text, flags=re.IGNORECASE)
    text = re.sub(r'covid\s?19', 'covid-19', text, flags=re.IGNORECASE)
    return text
```

```
In [17]: # apply identify_and_correct_vaccine_names to sentences
df_cleaned['sentence'] = df_cleaned['sentence'].apply(identify_and_correct_vaccine_names)
```

```
In [18]: # apply identify_and_correct_vaccine_names to labels
df_cleaned['label'] = df_cleaned['label'].apply(identify_and_correct_vaccine_names)
```

```
In [19]: # remove ampersand from sentences
df_cleaned['sentence'] = df_cleaned['sentence'].apply(lambda s: re.sub(r'&', '', s))
```

```
In [20]: # remove ampersand from labels
df_cleaned['label'] = df_cleaned['label'].apply(lambda s: re.sub(r'&', '', s))
```

### 3. Overall thoughts about vaccine

In this section, we use the entire dataset to look into:

- Attitude toward vaccine
- Biggest concerns through keyword
- Concerns over adverse events
- Concerns over side effect

#### 3.1 What does sentence word count tell us about attitude towards vaccine ?

The word count of sentence helps to gain insight into the level of its complexity. It could be seen from Figure 1 that 75% of expressions have word counts  $\leq 14$ . It suggests that most of questions might be simple enquiries about vaccine information such as "What are vaccine side effects ?". Some negative are also expressed such as "How far is it dangerous to take the delta variant for a kid ?"\_, but with a mild attitude. This can be illustrated more by Figure 3.

However, Figure 2 shows that there are still very long expressions. It is concluded from expressions with more than 20 words (Figure 4) that the long expressions reflect a higher level of concern over

vaccination. These concerns are about side effects, pregnancy, kids, etc. which are expressed with writer's own experiences, opinions, and knowledge. For example, a writer gave a news about dna altered by vaccine to express her/his worry: "*the fear that a vaccine will somehow change your dna is one we've seen aired regularly on social media the bbc asked three independent scientists about this they said that the coronavirus vaccine would not alter human dna some of the newly created vaccines including the one now approved in the uk developed by pfizerbiontech use a fragment of the virus's genetic*".

According to the word count, the entire dataset suggests that most of people want to inquiry basic information about vaccine. However, there are still significant number of people who are highly concerned about vaccine because of negative news, personal experiences, etc.

```
In [21]: # function to return word count of a text
def word_count(text):
    words = text.split()
    return len(words)
```

```
In [22]: # add word_count column to store word count of a sentence
df_cleaned['sentence_word_count'] = df_cleaned['sentence'].apply(word_count)
```

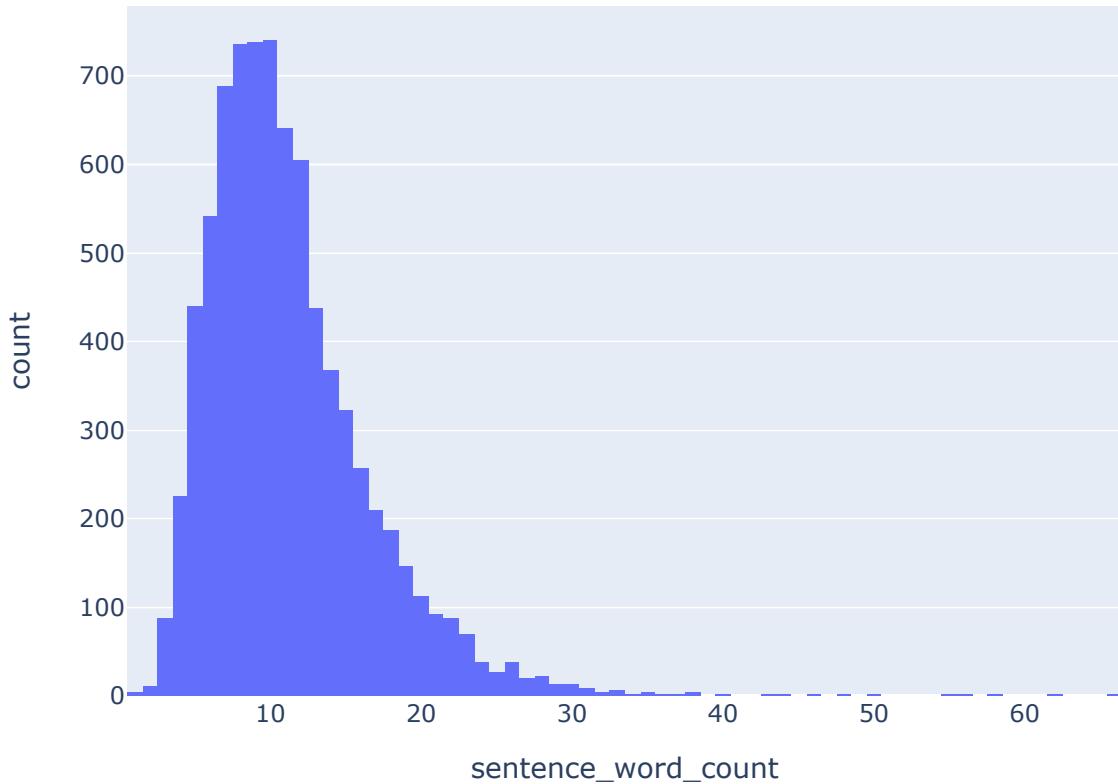
```
In [23]: # plot descriptive statistics of sentence word count
fig1 = go.Figure(data=[
go.Table(
    header=dict(
        values=["Count", "Mean", "Std",
                "Min", "25%", "50%",
                "75%", "Max"],
        align="left"
    ),
    cells=dict(
        values=list(df_cleaned['sentence_word_count'].describe()),
        align = "left")
)])
fig1.update_layout(title='Figure 1: Descriptive statistics of sentence word count')
fig1.show()
```

Figure 1: Descriptive statistics of sentence word count

Count	Mean	Std	Min	25%	50%	75%	Max
7972	11.19894	5.358464	1	7	10	14	66

```
In [24]: # create a histogram to show distribution of sentence word count
fig2 = px.histogram(df_cleaned, x = 'sentence_word_count', title='Figure 2: Distribution'
fig2.show()
```

Figure 2: Distribution of sentence word count



```
In [25]: # Sentences with word count less than 14
shortest_sentences = df_cleaned[df_cleaned['sentence_word_count'] < 14].sort_values(by='
fig3 = go.Figure(data=[
go.Table(
    header=dict(
        values=["Shortest sentences"],
        align="left"
    ),
    cells=dict(
        values=[list(shortest_sentences.values)],
        align = "left"
    )
])
fig3.update_layout(title='Figure 3: Sentences with word count less than 14')
fig3.show()
```

Figure 3: Sentences with word count less than 14

### Shortest sentences

how far is it dangerous to take the delta variant for a kid  
i already had covid so i'm already immune and don't need the shot  
how long is there between the first and second doses of the vaccine  
currently there is a decrease in the vaccination rate which can be alarming  
what could be worse than the disease itself i just wanna be safe  
my goal is to exercise i am fully vaccinated can i do it  
do i need to be born in the united states to get vaccinated  
i've been treated with convalescent plasma do i still need to get vaccinated  
do you need to be a us citizen to get a covid-19 vaccine  
since i have a a vaccine reaction before will this one be worse  
i want the pfizer vaccine how many times do i get a shot  
when should masks be worn by the general public during the covid-19 pandemic  
whats the point in getting the vaccine if i can still get covid  
i had an allergic reaction to a vaccine before will i be safe

In [26]:

```
# Sentences with word count more than 20
longest_sentences = df_cleaned[df_cleaned['sentence_word_count'] > 20].sort_values(by='s
fig4 = go.Figure(data=[
go.Table(
    header=dict(
        values=["Longest sentences"],
        align="left"
    ),
    cells=dict(
        values=[list(longest_sentences)],
        align = "left")
)])
fig4.update_layout(title='Figure 4: Sentences with word count more than 20')
fig4.show()
```

Figure 4: Sentences with word count more than 20

### Longest sentences

make sure you understand and comply with all the requirements of the airline and your destination country regarding travel screening or quarantine which may differ from the requirements of the united states if you do not meet the requirements of your duty station you may be denied entry and must return to the united states check the current situation of covid-19 in your country of destination

the fear that a vaccine will somehow change your dna is one we've seen aired regularly on social media the bbc asked three independent scientists about this they said that the coronavirus vaccine would not alter human dna some of the newly created vaccines including the one now approved in the uk developed by pfizerbiontech use a fragment of the virus's genetic

vaccination is one of the most effective public health interventions in the world for saving lives and promoting good health only clean water which is considered to be a basic human right performs better despite this uptake of vaccines has reduced in some countries and this is thought to be partly caused by misguided concerns over vaccine safety

these companies could be out to make a profit out of the whole pandemic i

### 3.2 What are common keywords among all expressions ?

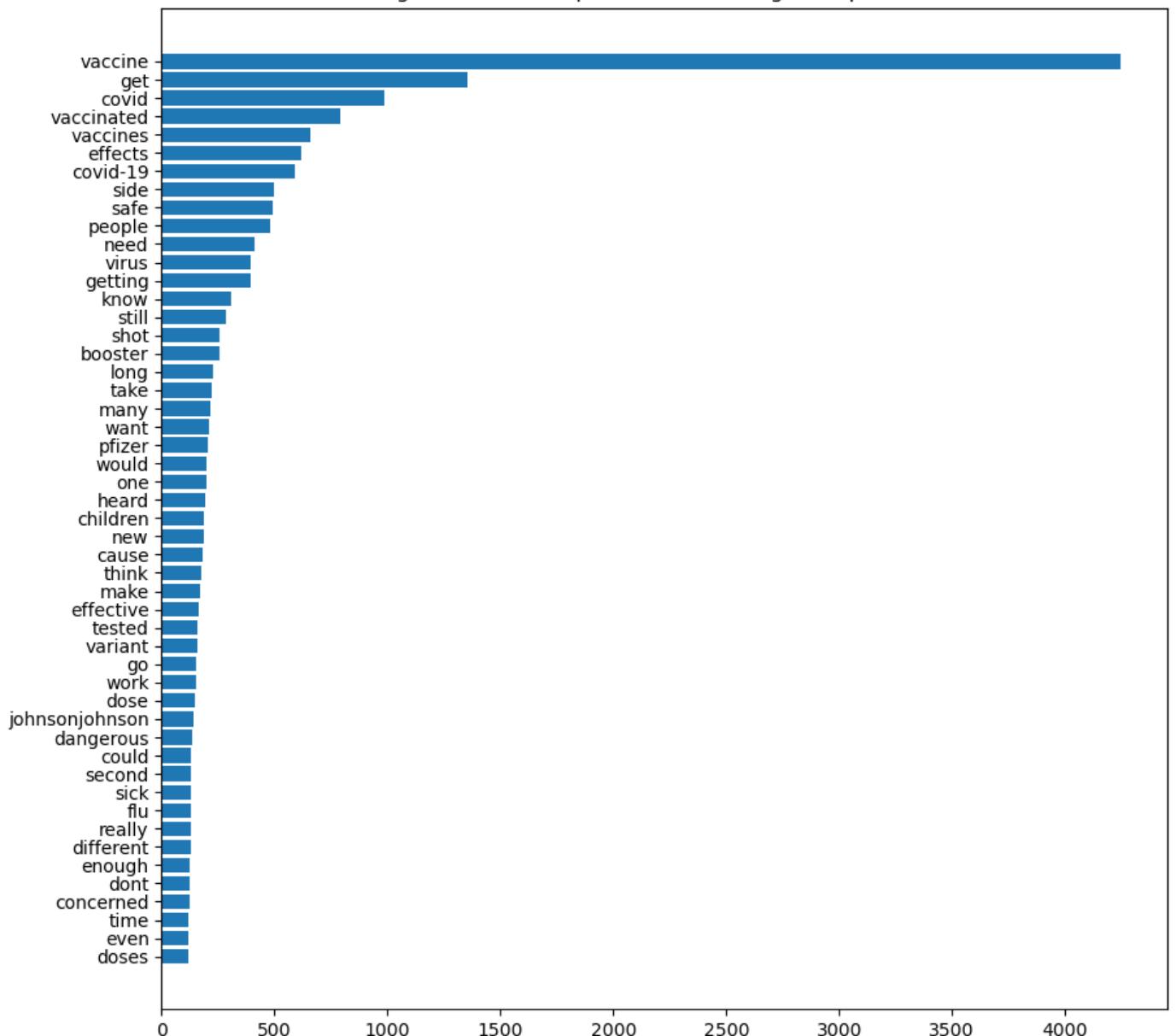
Figure 5 helps us to see top 50 words mentioned by all expressions. From these key words, we could easily guess what people care about most. Below are concerns re-written by using key words (bold) in Figure 5, ordered by descending word frequency:

1. what are **side effects** of vaccines ?
2. is covid-19 vaccine **safe**?
3. is **booster shot/second doses** safe/neccessary ?
4. **pfizer**
5. **johson & johnson**
6. is vaccine safe for **children** ?
7. is vaccine **effective**?
8. do I need to have **test** before getting vaccine ?
9. do vaccines work with new **variants** ?
10. am I **sick** after vaccination ?
11. i don't get **flu** vaccines, then why should I not get covid vaccines ?

```
In [27]: def most_common_words_all_questions (df,column,title, top_n):  
    # get all words  
    all_words = [w for s in df[column] for w in s.split()]  
  
    # count occurence of each word (i.e 'if': 108)  
    word_count_dict = dict(Counter(all_words))  
  
    # sort words by their occurences (desc)  
    popular_words = sorted(word_count_dict, key = word_count_dict.get, reverse = True)  
  
    # remove words that are in stopwords list  
    popular_words_nonstop = [w for w in popular_words if w not in stop_words]  
  
    # get 50 popular_words_nonstop and reverse their order to get values.  
    # y coordinates of the bars, labels  
    plt.barh(range(top_n), [word_count_dict[w] for w in reversed(popular_words_nonstop[0:top_n])])  
  
    # Get or set the current tick locations and labels of the y-axis.  
    plt.yticks([x for x in range(top_n)], reversed(popular_words_nonstop[0:top_n]))  
    plt.title(title)  
    plt.show()
```

```
In [28]: plt.figure(figsize=(10,10))  
most_common_words_all_questions(df_cleaned,'sentence',"Figure 5: Most frequent words amo
```

Figure 5: Most frequent words among all expressions



### 3.3 Which adverse events are people concerned about ?

According to Centers for Disease Control and Prevention (2023), these are common adverse events:

- Anaphylaxis: a severe type of allergic reaction 4
- Thrombosis - TTS: blood clots or issues with clotting
- Myocarditis: inflammation of the heart muscle
- Pericarditis: inflammation of the outer lining of the heart
- Guillain-Barré Syndrome - GBS: rare disorder where the body's immune system damages nerve cells
- Death

It can be seen from Figure 6 that all adverse events listed by CDC are mentioned in this dataset. *Death* is the biggest concern, which is proved by 300 times mentioned. *Anaphylaxis* and *Thrombosis* are also common, with 64 and 54 times respectively. *Myocarditis*, *Pericarditis* and *GBS* are unpopular with less than 20 times mentioned. This result is quite interesting because it reflects what actually happens according reports collected by the CDC. Regardless of age and gender, *death*, *Anaphylaxis* and *Thrombosis* occurred at a rate of 0.0029%, 0.0005% and 0.0004% respectively. Meanwhile,

*Myocarditis, Pericarditis and Guillain-Barré Syndrome were reported by males of specific age group only.*

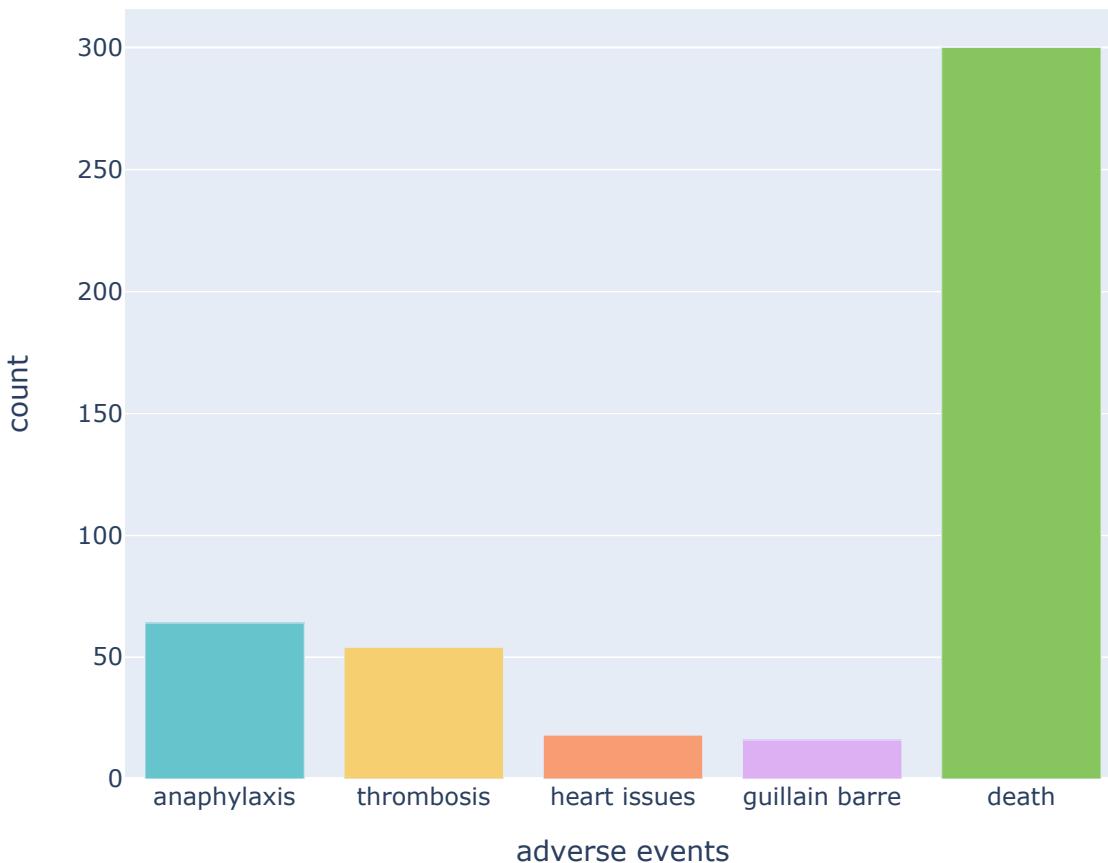
```
In [29]: # identify all adverse events by their names and definitions
# then count occurrences of adverse event and store in a dict
def count_adverse_event(text):
    adverse_event_count = {}

    adverse_event_count['anaphylaxis'] = len(re.findall(r'anaphylaxis|allergic(?: reacti
adverse_event_count['thrombosis'] = len(re.findall(r'thrombosis|blood clot', text))
adverse_event_count['heart issues'] = len(re.findall(r'myocarditis|pericarditis|hear
adverse_event_count['guillain barre'] = len(re.findall(r'guillain-barre|guillain(?: 
adverse_event_count['death'] = len(re.findall(r'death|die|dying', text))

    return adverse_event_count
```

```
In [30]: # create a bar chart to show occurrences of adverse events in the entire dataset
adverse_count = count_adverse_event(question_text)
adverse_count_df = pd.DataFrame({'adverse events': adverse_count.keys(), 'count': advers
fig6 = px.bar(adverse_count_df, x='adverse events', y='count', color='adverse events', c
fig6.update_layout(title='Figure 6: Occurrences of adverse events', showlegend=False)
fig6.show()
```

Figure 6: Occurrences of adverse events



### 3.4 Which side effects are people concerned about ?

According to Centers for Disease Control and Prevention (2023), these are reported side effects of vaccine:

- pain
- lymph nodes
- irritability
- sleepiness
- swelling
- tiredness
- headache
- chills
- fever

Figure 7 indicates that the number of times pain mentioned outnumbers other side effects (90 times mentioned). People are also afraid of Lymph nodes (25 times mentioned), swelling (9 times mentioned) and fever (5 times mentioned) associated with covid-19 vaccines. Other side effects are barely mentioned.

```
In [31]: # side effects
side_effects = ['pain', 'lymph nodes', 'irritability', 'sleepiness', 'swelling', 'tiredn
```

```
In [32]: # NOTE: it takes a while to load this model.
model = api.load('glove-wiki-gigaword-300')

def remove_underscore(word):
    return word.replace('_', ' ')

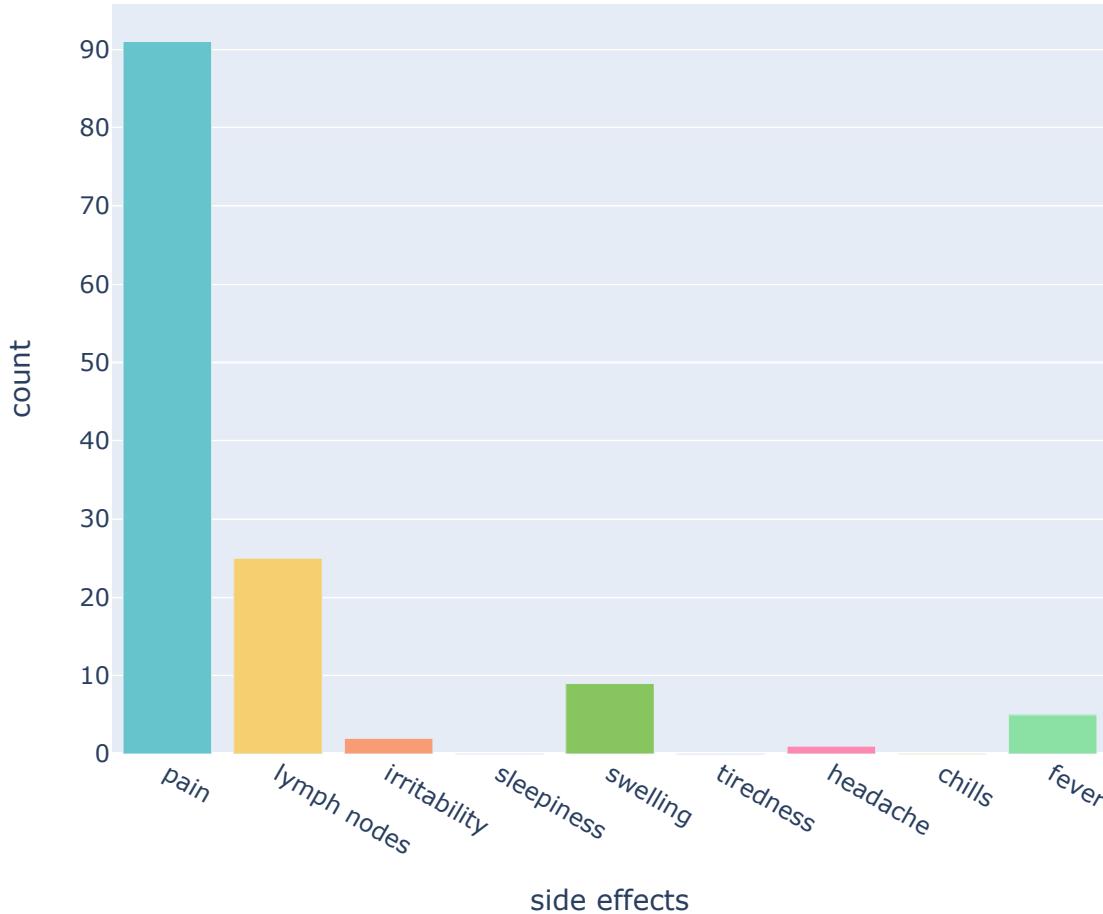
def find_synonyms_all_forms(word):
    """
    Find all synonyms of a word regardless of form. For example, gorgeous has synonyms s
    """
    synonyms_all_forms = []
    synonyms_all_forms.append(word)
    for synset in wordnet.synsets(word):
        for lemma in synset.lemmas():
            try:
                # only use lemma with similarity > 0.3
                if model.similarity(word, lemma.name()) > 0.3:
                    synonyms_all_forms.append(remove_underscore(lemma.name()))
            except KeyError:
                pass
    # traverse through words derived from the lemma. For example, happy -> happy
    for related_form in lemma.derivationally_related_forms():
        w = remove_underscore(related_form.name())
        if w != None:
            w = remove_underscore(w)
            try:
                # only use derived word with similarity > 0.3
                if model.similarity(word, w) > 0.1:
                    synonyms_all_forms.append(w)
            except KeyError:
                pass
    return list(set(synonyms_all_forms))
```

```
In [33]: def generate_regex_from_word_list(words):
    reg = ""
    for w in words:
        reg += w
        reg += '|'
    reg = reg[:-1]
    return re.compile(reg)
```

```
In [34]: def count_side_effect(text):
    side_effect_count = {}
    for side_effect in side_effects:
        all_syn = find_synonyms_all_forms(side_effect)
        if side_effect == 'headache':
            # the result of find_synonyms_all_forms() for 'headache' contains 'concern',
            # remove these words because they are not right in the context
            all_syn.remove('concern')
            all_syn.remove('worry')
        reg = generate_regex_from_word_list(all_syn)
        # count number of time a side_effect is found and store in a dict (k: side effect, v: count)
        side_effect_count[side_effect] = len(reg.findall(text))
    return side_effect_count
```

```
In [35]: # create bar chart to show occurrences of side effects in the dataset
side_effect_count = count_side_effect(question_text)
side_effect_count_df = pd.DataFrame({'side effects': side_effect_count.keys(), 'count': side_effect_count.values()})
fig7 = px.bar(side_effect_count_df, x='side effects', y='count', color='side effects', color_continuous_scale=px.colors.sequential.Plasma)
fig7.update_layout(title='Figure 7: Occurrences of side effects in the dataset', showlegend=False)
fig7.show()
```

Figure 7: Occurrences of side effects in the dataset



## 4. Top 20 popular questions that people asked the most

In this section, we look into top 20 popular labels.

### 4.1 What are 20 commonly asked questions ?

Figure 8 displays top 20 most popular questions according to the number of expressions. Questions in this table are quite similar with questions written based on most frequent keywords in section 3.2.

In [36]:

```
# groupby label and count their occurrences
df_label_count = df_cleaned.groupby('label').size().reset_index(name = 'count')
# sort labels by their occurrences by desc
df_label_count = df_label_count.sort_values('count', ascending = False)
# get top 20 labels
top20_label_count = df_label_count[0:20]
# create a table to show top 20 questions
fig8 = go.Figure(data=[go.Table(
    header=dict(
        values=["Question", "Count"],
        align="left"
    ),
    cells=dict(
        values=[top20_label_count['label'], top20_label_count['count']],
        align = "left"
    )
)])
fig8.update_layout(title='Figure 8: Top 20 most asked questions')
fig8.show()
```

Figure 8: Top 20 most asked questions

Question	Count
how effective is the vaccine against the omicron variant	101
how much will i have to pay for the vaccine	101
i don't trust vaccines if they're from china or russia	100
do i need to continue safety measures after getting the vaccine	100
what is the omicron variant	100
the covid vaccine is not safe	99
can i still get covid even after being vaccinated	97
how many doses do i need	93
i'm not sure it is effective enough	91

## 4.2 How negative are these 20 questions ?

According Figure 9, these are 10 questions whose negative expressions are more than positive ones:

- can I still get covid even after being vaccinated ?
- do I qualify for the vaccine

- how can I get the vaccine
- how long does the immunity from the vaccine last
- how much will I have to pay for the vaccine
- I am concerned the vaccine will make me sick
- side effects and adverse reactions worry me
- what is in the vaccine
- what is the omicron variant
- what is covid

However, this number could be more than 10. Sentiment analysis of TextBlob might have not captured the negativity of these questions which are quite "obviously negative":

- I don't trust vaccines if they are from China or Russia
- Is the booster shot dangerous?
- The covid vaccine is not safe
- There are many reports of severe side effects or deaths from the vaccine
- They will put the microchip to manipulate me

In conclusion, among 20 commonly asked, there are up to 15 negative questions. This number is a reliable indicator of people's worry about covid-19 vaccines.

```
In [37]: # get records of these 20 labels from original dataset
df_top20_label = df_cleaned[df_cleaned['label'].isin(top20_label_count['label'])].copy()
```

```
In [38]: def get_wordnet_pos(treebank_tag):
    if treebank_tag.startswith('J'):
        return wordnet.ADJ
    elif treebank_tag.startswith('V'):
        return wordnet.VERB
    elif treebank_tag.startswith('N'):
        return wordnet.NOUN
    elif treebank_tag.startswith('R'):
        return wordnet.ADV
    else:
        return wordnet.NOUN
```

```
In [39]: # clean a text before calculating its sentiment score
def clean_text(text):
    # create word tokens
    all_tokens = text.split()
    # pos tag tokens
    tagged_tokens = nltk.pos_tag(all_tokens)
    # TextBlob does not automatically perform lemmatization so I lemmatize first
    lemmatizer = WordNetLemmatizer()
    lemmatized_tokens = [lemmatizer.lemmatize(token, pos=get_wordnet_pos(tag)) for token in tagged_tokens]
    # filter out stopwords
    lemmatized_tokens_nonstop = [token for token in lemmatized_tokens if token not in stop_words]

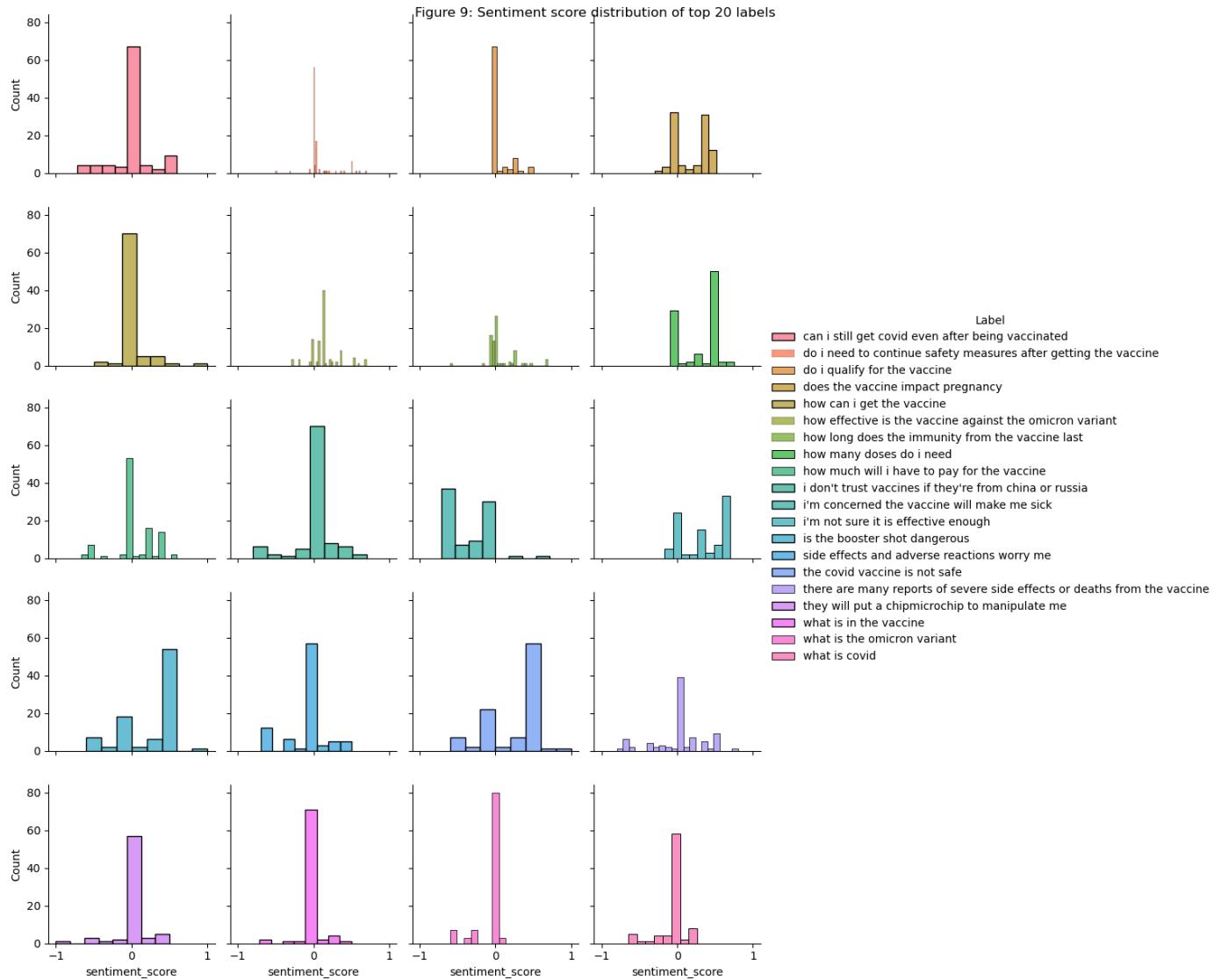
    return " ".join(lemmatized_tokens_nonstop)
```

```
In [40]: # calculate sentiment score function
def sentiment_analysis(text):
    return TextBlob(clean_text(text)).sentiment.polarity
```

```
In [41]: # calculate sentiment for each sentence
df_top20_label['sentiment_score'] = df_top20_label['sentence'].apply(lambda s: sentiment_analysis(s).polarity)
```

```
In [42]: # sentiment distribution
hist_top20_label_sentiment = sns.FacetGrid(df_top20_label,col = 'label', hue="label", height=6)
hist_top20_label_sentiment.map(sns.histplot, "sentiment_score")
hist_top20_label_sentiment.set_titles("")
hist_top20_label_sentiment.fig.suptitle("Figure 9: Sentiment score distribution of top 20 labels")
hist_top20_label_sentiment.add_legend(title="Label", ncol=1, bbox_to_anchor=(1, 0.5))
```

Out[42]: <seaborn.axisgrid.FacetGrid at 0x7f8567497b80>



## 4.3 What are people talking about in these 20 questions ?

In this section, we find out keywords of each questions using tf-idf. The WordCloud makes it easy to capture what people want to ask in each question. Below are interesting findings from Figure 10:

**Do I need to continue safety measures after getting the vaccine ?** People want to know if they will have to: (1) wear mask, (2) keep social distancing from family members/friends.

**Do I qualify for the vaccine ?** People want to know: (1) information about priority group, (2) if they can get vaccinated if they hold work visa, (3) at what age they can get vaccinated, (4) if they have to wait their turns to get vaccinated, (4) why teachers belong to priority group, (5) if cancer/diabetic patients can get vaccines

**Does the vaccine impact pregnancy ?** People want to know: (1) if vaccine causes miscarriage, (2) if vaccine is safe for fetuses

**How can I get the vaccine ?** People want to know: (1) where to get vaccinated, (2) requirements to get vaccine, (3) if they need to make appointment, (4) if they need to visit doctor

**Is the booster shot dangerous?** People want to know: (1) if emphysema patients can get booster shot, (2) if booster shot have been proved to be safe ?

**The COVID vaccine is not safe** People wonder if (1) the vucuna is not completely safe

**There are many reports of severe side effects or deaths from the vaccine** People are worried about: (1) blood clots, (2) side effects on children and the elderly

**What is in the vaccine** People wonder if vaccine: (1) contains microchips, (2) is toxic, (3) trigger allergic reactions

```
In [43]: # group expressions/setences by label and merge them into a string
df_label_merged_sentence = df_cleaned.groupby('label')['sentence'].agg(lambda s: ' '.join(s))
```

```
In [44]: # calcuate tf-idf
vectorizer = TfidfVectorizer(stop_words = stop_words)
tf_idf_matrix = vectorizer.fit_transform(df_label_merged_sentence['sentence'])
feature_names = vectorizer.get_feature_names()
tf_idf_df = pd.DataFrame(tf_idf_matrix.toarray(), columns=feature_names)
tf_idf_df = pd.merge(df_label_merged_sentence, tf_idf_df, left_index=True, right_index=True)
```

```
In [45]: top20_label_tf_idf_df = tf_idf_df[tf_idf_df['label'].isin(top20_label_count['label'])]
top20_label_tf_idf_df = top20_label_tf_idf_df.drop('sentence', axis=1)
top20_label_tf_idf_df.set_index(['label'], inplace = True)
```

```
In [46]: import matplotlib.image as mpimg

# NOTE: this code takes a while to execute due to intensive process of images

# create wordclouds to show important words for each of the 20 labels
fig, axs = plt.subplots(10, 2, figsize = (50, 100))
i = 1

for label in top20_label_tf_idf_df.index:
    word_tf_idf = {}
    document = top20_label_tf_idf_df.loc[label]
    sorted_tf_idf = document.sort_values(ascending=False)
    top50_tf_idf = sorted_tf_idf[0:50]

    for word in top50_tf_idf.index:
        word_tf_idf.update({word: top50_tf_idf[word]})

    # display word cloud
    col = 1 if i % 2 == 0 else 0
    row = math.ceil(i / 2) - 1
    axs[row, col].plot()
    axs[row, col].axis('off')
    word_cloud = WordCloud(mode='RGBA', background_color='white', max_words=50, width=150, height=150)
    word_cloud.to_file('wordcloud'+str(i)+'.png')
    wordcloud_img = mpimg.imread('wordcloud'+str(i)+'.png')
    axs[row, col].imshow(wordcloud_img, interpolation='bilinear', extent=[0, 300, 0, 200])
    axs[row, col].set_title(label, fontsize=30, fontweight="bold")
    i += 1
```



longer & contract SICK  
hear vaccines  
totally never catch guarantee  
vaccinated infected virus percentage

**COVID**

do i qualify for the vaccine  
**get groups**  
frontline workers covid information  
group roll out diabetic life  
old till want age turn  
qualifications priorities go  
priority qualified level cancer 19-31  
vaccinated listed know  
vaccine state requirements  
queue teachers

how can i get the vaccine  
**vaccinated**  
gofind get near  
steps take make qualify  
one need locally vaccination  
covidissed doctor tried  
appointment contact obtain authority  
requirements places downtown  
vaccine available closer venue  
available hassle localeligible

how long does the immunity from the vaccine last  
**days 90**  
need effective often lasts many dose  
vaccines true flu know heard life  
protective like booster protected  
vaccine effectiveness antibody protection  
last year yearly seems  
long immunity months good covid  
months longer works work vaccinated

how much will i have to pay for the vaccine  
**free insurance**  
paying anything buy needs expensive  
relaxed everyone expanded medicare  
much far vaccinated  
struggling worth way cash gaps  
get pay insured costs  
coverage charging paid without  
die price value hundreds dont  
poor destined health covid 19  
bourgeois kind uninsured vaccin

i'm concerned the vaccine will make me sick  
**sick**  
maybe affects reactions get  
don't affect claims another concern  
contrary may contradictory long negative  
afraid something kill  
concerned people ill symptoms  
make causes secret getting term remedy

is the booster shot dangerous  
**booster**  
goes scientifically well think use  
effective booster 100 guaranteed  
dangerous think emphysema  
established wht mom  
comes what proven yes  
really covid shots know  
shots dont ineffectiveness compiled

the covid vaccine is not safe  
**vaccine**  
dangerous completely well  
watched vaccinated enough know people  
vaccine probability 19 assures proven  
vaccine what health effects  
vaccine may

safety vaccination follow without  
continue even  
vaccine getting talon masks back people  
vaccinated friends using family  
required

does the vaccine impact pregnancy  
**pregnancy**  
unborn problems serious heard  
recommended months take  
affect birth safe  
miscarriages woman born time  
woman fetus gain jab  
pregnant risk harm 19  
likely vaccine high  
enough fetuses concerns baby tested  
impact wait group concerned  
miscarry illnes

how effective is the vaccine against the omicron variant  
**omicron**  
multiple recent serve  
virus new scientists covid current  
omicron effective help  
variant stop counter  
african supportive  
south vaccine appearing works

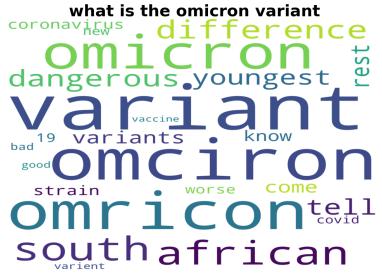
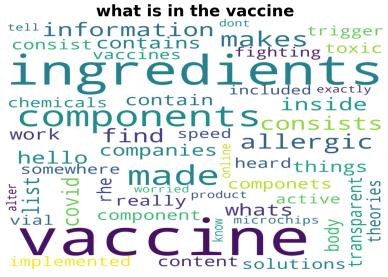
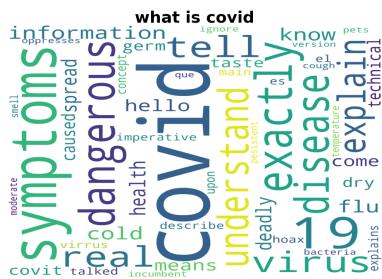
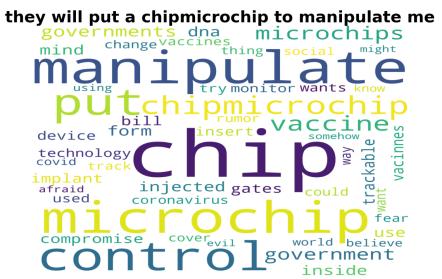
how many doses do i need  
**doses**  
immunized vaccinated waste  
enough two vaccine hello shot  
placed get many take better  
comply proper necessary single amount  
1st confused different dosis  
vaccines dosage ensure protected  
start different dosis  
recommended vary different boosters  
require diary effective depending  
whole number administer receive

i don't trust vaccines if they're from china or russia  
**china**  
countries russian  
reliable vaccines  
trust unreliable  
china origin  
russia origins  
made vaccine

i'm not sure it is effective enough  
**effective**  
virus 19 gonna think perhaps  
say thorough efficacy really  
say veracity heard good effective 100  
satisfied maybe opinion  
complete burning strand north evidence  
insufficient leaves modern creator  
vaccine enough work strongly  
little people sure offers  
improve specific effectiveness  
emerging make success

side effects and adverse reactions worry me  
**side effects**  
adverse reactions concern disease  
without stress worried  
worry bring secondary  
cause reaction events  
reaction risk things known  
reaction happen im  
ill covid negative dont begin  
bad know vaccine suffer yet  
side worrisome concerned

there are many reports of severe side effects or deaths from the vaccine  
**vaccine**  
died really dying  
statistics astraZeneca  
lives read bad  
causes serious  
someone true  
tryng fatality  
blood



## 5. Vaccine Investigation

### 5.1 How do people feel about each vaccine type ?

After the data cleaning step 2.3.4, there are only four vaccine types mentioned in the dataset: Astra Zeneca, Johnson & Johnson, Moderna, and Pfizer. In this section, sentiment analysis is performed on each vaccine type. The method used includes these steps:

- Find all expressions that contains a vaccine type
- Group these expressions by vaccine types
- Calculate the sentiment analysis score of each expression
- Create a table to show count/percentage of positive/neutral/negative expressions of each vaccines (0 - neutral, > 0: positive, < 0: negative)

Of all vaccines, neutral expressions account for the most proportion, followed by positive and negative. For Astra Zeneca, positive rate and negative rate are nearly equal, 21.43% and 17.86% respectively. Its negative rate is also the largest among all vaccines. For the other vaccines, positive rates are significantly higher than negative rate. Among all vaccines, Johnson and Johnson has the highest positive rate (43.36%).

```
In [47]: # all vaccine names based on 2.3.4
vaccines = ['astrazeneca', 'johnsonjohnson', 'moderna', 'pfizer']
```

```
In [48]: # create a dataframe to store vaccine, sentence, and label
df_vaccine_sentence = pd.DataFrame(columns=['vaccine', 'sentence', 'label'])
# traverse through each rows in df_cleaned
for i in range(len(df_cleaned)):
    # traverse through all vaccines
    for v in vaccines:
        # if the sentence/expression mentions the vaccine
        if v in df_cleaned.iloc[i]['sentence']:
            df_vaccine_sentence = df_vaccine_sentence.append({'vaccine': v, 'sentence':
```

```
In [49]: # calculate sentiment score for all sentences
df_vaccine_sentence['sentiment_score'] = df_vaccine_sentence['sentence'].apply(lambda s:
```

```
In [50]: positives = []
pct_positives = []
neutrals = []
pct_neutrals = []
negatives = []
pct_negatives = []
totals = []

for v in vaccines:
    vaccine_rows = df_vaccine_sentence[df_vaccine_sentence['vaccine'] == v]
    total = len(vaccine_rows)
    positive = len(vaccine_rows[vaccine_rows['sentiment_score'] > 0])
    pct_positive = round(positive / total * 100, 2)
    neutral = len(vaccine_rows[vaccine_rows['sentiment_score'] == 0])
    pct_neutral = round(neutral / total * 100, 2)
    negative = len(vaccine_rows[vaccine_rows['sentiment_score'] < 0])
    pct_negative = round(negative / total * 100, 2)

    positives.append(positive)
    pct_positives.append(pct_positive)
    neutrals.append(neutral)
    pct_neutrals.append(pct_neutral)
    negatives.append(negative)
    pct_negatives.append(pct_negative)
    totals.append(total)

sentiment_stats = [vaccines, positives, pct_positives, neutrals, pct_neutrals, negatives]
```

```
In [51]: # create a table to show count/percentage of positive/neutral/negative expressions vs to
fig12 = go.Figure(data=[
go.Table(
    header=dict(
        values=["Vaccine", "Positive (count)", "Positive (%)", "Neutral (count)", "N
                "Negative (%)", "Total (count)"],
        align="left"
    ),
    cells=dict(
        values=sentiment_stats,
        align = "left")
)])
fig12.update_layout(title='Figure 12: Sentiment analysis of vaccines')
fig12.show()
```

Figure 12: Sentiment analysis of vaccines

Vaccine	Positive (count)	Positive (%)	Neutral (count)	Neutral (%)	Negative (count)	Negative (%)	Total (count)
astrazeneca	6	21.43	17	60.71	5	17.86	28
johnsonjohnson	62	43.36	77	53.85	4	2.8	143
moderna	7	15.22	38	82.61	1	2.17	46
pfizer	84	39.81	119	56.4	8	3.79	211

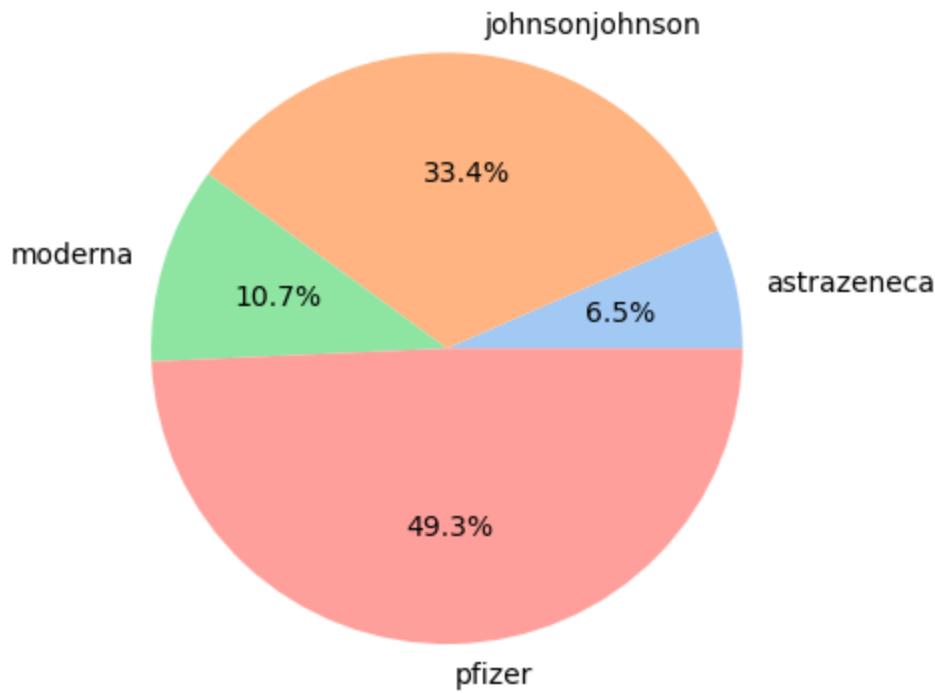
## 5.2 How popular is a vaccine type ?

According to Figure 13, Pfizer is the most popular vaccine in this dataset, which accounts for 49.3% of expressions that mention vaccine names. Followed by Pfizer is Johnson & Johnson (33.4%) and Moderna (10.7). Astra Zeneca is the least mentioned vaccine among expressions. These numbers could be partly explained by Covid 19 Vaccine Tracker (2023): Pfizer is approved in 149 countries, Johnson & Johnson is approved in 113 countries, and Moderna is approved in 88 countries. AstraZeneca is the only vaccine that does not follow this pattern. Although it is not mentioned much in this dataset, it is very popular around the world, being used by 149 countries.

In [52]:

```
# vaccine vs questions
df_vaccine_sentence_count = df_vaccine_sentence.groupby('vaccine')['sentence'].count().r
plt.pie(df_vaccine_sentence_count['count'], labels=df_vaccine_sentence_count['vaccine'],
plt.title('Figure 13: Vaccine popularity')
plt.show()
```

Figure 13: Vaccine popularity



## 5.3 What are people curious about each vaccine type ?

In [53]:

```
# top 10 most frequently asked questions of each vaccine
```

```

def top10_vaccine_label(vaccine):
    data = df_vaccine_sentence[df_vaccine_sentence['vaccine'] == vaccine]
    data = data.groupby('label').size().reset_index(name = 'count')
    return data.sort_values('count', ascending=False).head(10)

```

## Moderna

Figure 14 and Figure 15 show what people question about Moderna. It is concluded that people are most curious about booster shot of Moderna. One special thing is whether Moderna can be mixed with other vaccines.

```

In [54]: # top 10 questions/labels and their occurrences
moderna_expression_count = top10_vaccine_label('moderna')
fig14 = go.Figure(data=[go.Table(
    header=dict(
        values=["Question", "Count"],
        align="left"
    ),
    cells=dict(
        values=[list(moderna_expression_count['label']), list(moderna_expression_count['count'])],
        align = "left"
    )
)])
fig14.update_layout(title='Figure 14: Moderna questions')
fig14.show()

```

Figure 14: Moderna questions

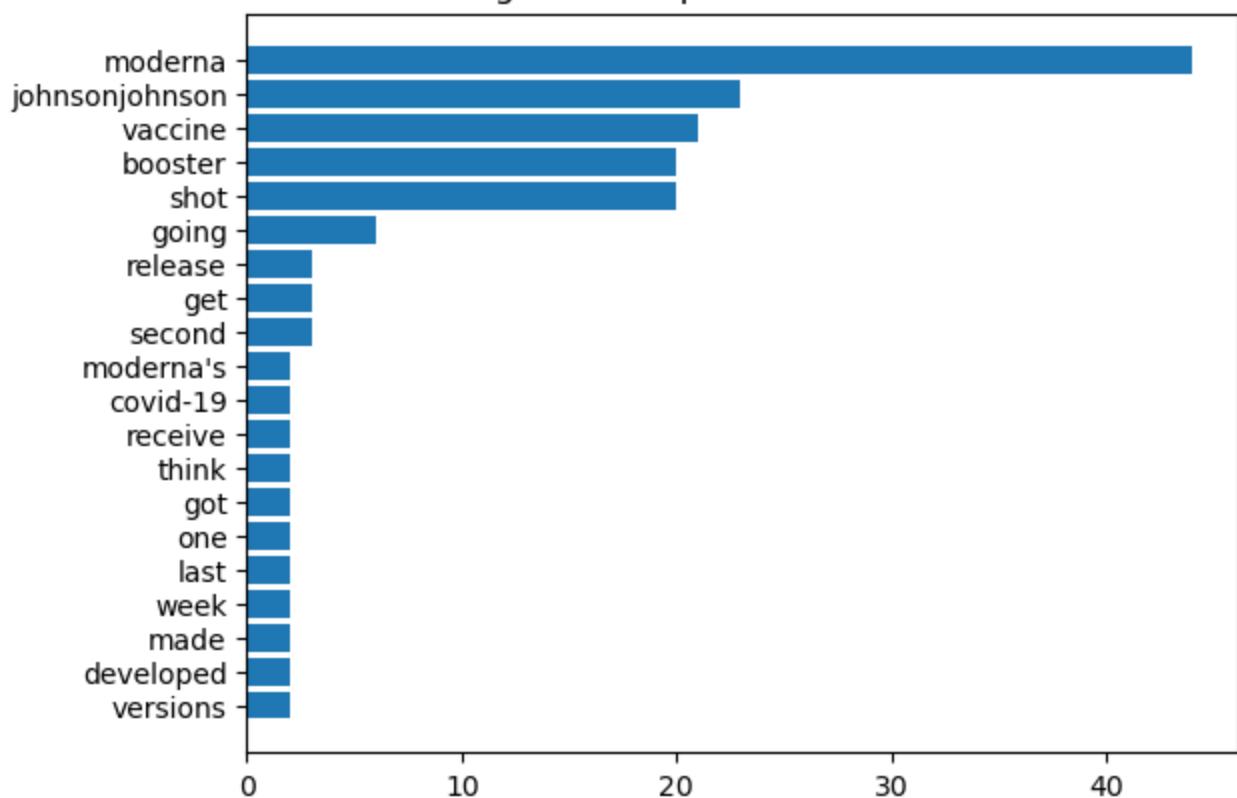
Question	Count
will there be a booster shot for johnsonjohnson and moderna	23
what is the moderna vaccine	20
can you mix the vaccines	1
how long do i have to wait between doses	1
which vaccines are available	1

```

In [55]: # top 20 frequent words
most_common_words_all_questions(df_vaccine_sentence[df_vaccine_sentence['vaccine'] == 'm'])

```

Figure 15: Top 20 common words



## Pfizer

According to Centers for Disease Control and Prevention (2023), most Myocarditis cases are reported by male in the age groups 12-24 after they took the second dose of mRNA vaccine. This explain why in both Figure 16 and 17, people most frequently ask questions about *booster shot*, *young men*, *myocarditis*, *heart problems*.

```
In [56]: # top 10 questions/labels and their occurrences
pfizer_expression_count = top10_vaccine_label('pfizer')
fig14 = go.Figure(data=[go.Table(
    header=dict(
        values=["Question", "Count"],
        align="left"
    ),
    cells=dict(
        values=[list(pfizer_expression_count['label']), list(pfizer_expression_count['count']),
                align = "left")]
)))
fig14.update_layout(title='Figure 16: Pfizer questions')
fig14.show()
```

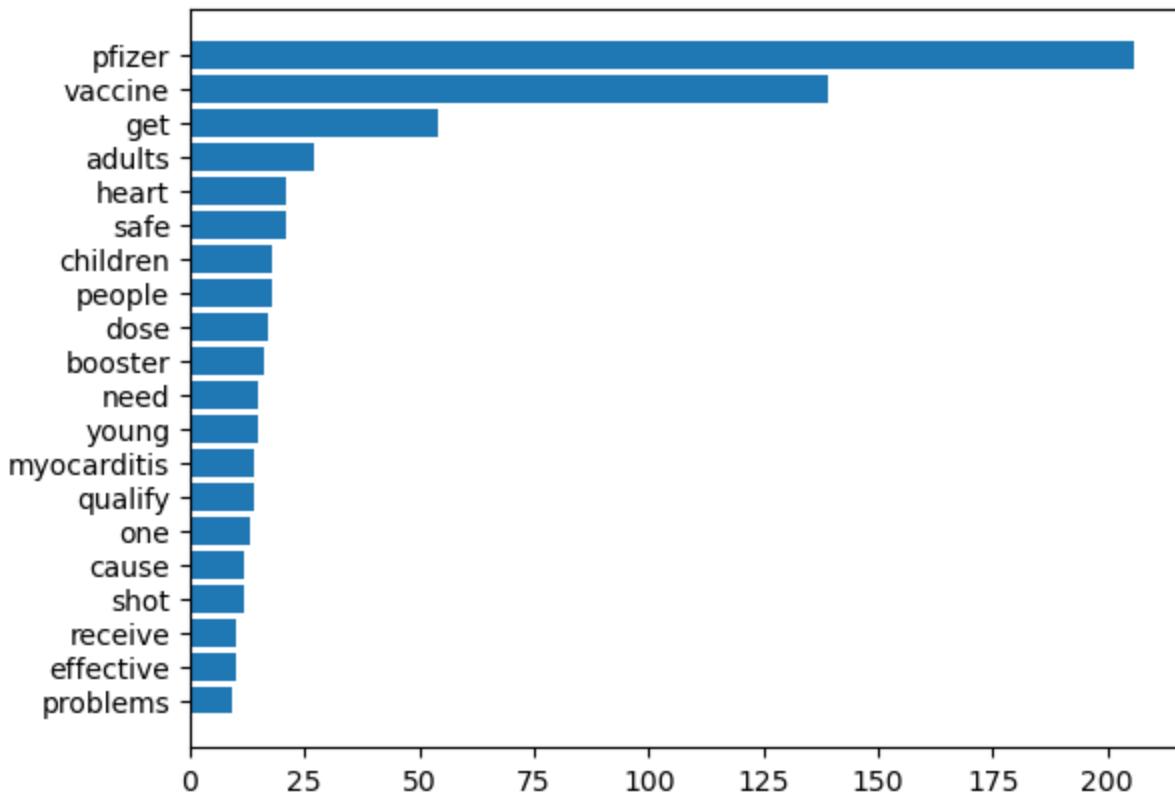
Figure 16: Pfizer questions

Question	Count
who can get the pfizer vaccine	46
will i need a booster shot	25
is the pfizer vaccine safe for young men	23
is the pfizer vaccine effective	22

what is the pfizer vaccine	23
does the pfizer vaccine cause heart problems	22
does the pfizer vaccine cause myocarditis	18
is this pfizer vaccine equally effective in kids as it is in adults	17
do children receive the same dose of pfizer as adults	14
i don't trust the companies producing	4

```
In [57]: # top 20 frequent words
most_common_words_all_questions(df_vaccine_sentence[df_vaccine_sentence['vaccine'] == 'p'])
```

Figure 17: Top 20 common words



## Johnson and Johnson

Katella (2023) states that Johnson & Johnson "showed 67% efficacy in preventing moderate to severe/critical disease by 14 days after vaccination, and 66% effective 28 days after vaccination". This could explain why the most popular question about Johnson & Johnson vaccine in the dataset is whether it is less effective than other vaccines (Figure 18). According to Centers for Disease Control and Prevention (2023), Guillain-Barré Syndrome (GBS) is a rare disorder found in some people receiving Johnson & Johnson. That is reason why in both Figure 18 and 19, we can see questions about rare nerve syndrome. People are also worried about Johnson & Johnson side effects such as mentioned *death*, and *blood clots*. Moreover, like Moderna, people are also curious about its ability to mix with other vaccines.

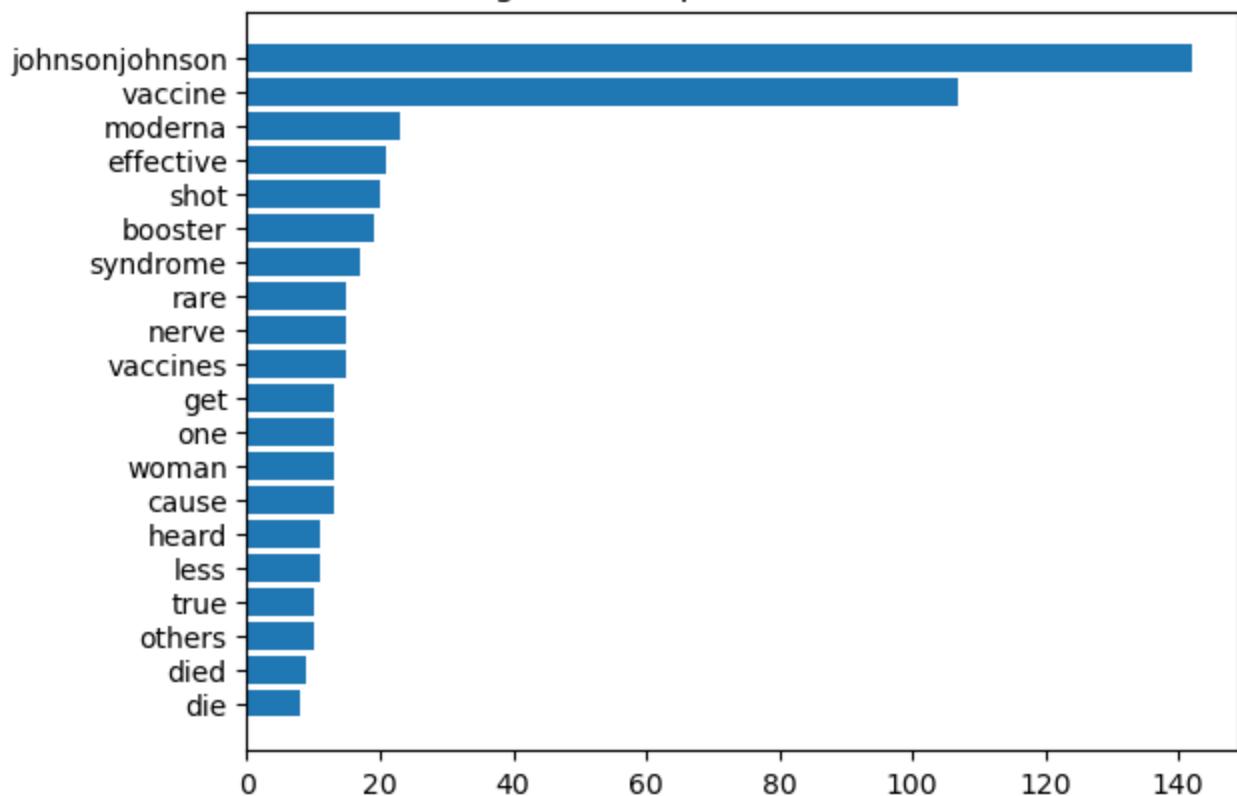
```
In [58]: # top 10 questions/labels and their occurences
pfizer_expression_count = top10_vaccine_label('johnsonjohnson')
fig17 = go.Figure(data=[
go.Table(
    header=dict(
        values=["Question", "Count"],
        align="left"
    ),
    cells=dict(
        values=[list(pfizer_expression_count['label']), list(pfizer_expression_count['count'])],
        align = "left"
    )
])
fig17.update_layout(title='Figure 18: Johnson & Johnson questions')
fig17.show()
```

Figure 18: Johnson & Johnson questions

Question	Count
is the johnsonjohnson vaccine less effective than the others	43
will there be a booster shot for johnsonjohnson and moderna	24
did one woman die after getting the johnsonjohnson vaccine	23
does the johnsonjohnson vaccine cause rare nerve syndrome	21
what is the johnsonjohnson vaccine	21
can i have the vaccine as a catholic	2
can we choose which vaccine we want	2
can you mix the vaccines	2
i am worried about blood clots as a	2

```
In [59]: # top 20 frequent words
most_common_words_all_questions(df_vaccine_sentence[df_vaccine_sentence['vaccine'] == 'j']
```

Figure 19: Top 20 common words



## Astra Zeneca

Figure 20 shows that people are most curious about why Astra Zeneca is not approved by USA. Besides, like Johnson & Johnson, there is a concern about its side effects such as *death* and *blood clots*. Moreover, people also want to know if it is possible to mix Astra Zeneca with other vaccines.

In [60]:

```
# top 10 questions/labels and their occurrences
astrazeneca_expression_count = top10_vaccine_label('astrazeneca')
fig17 = go.Figure(data=[go.Table(
    header=dict(
        values=["Question", "Count"],
        align="left"
    ),
    cells=dict(
        values=[list(astrazeneca_expression_count['label']), list(astrazeneca_express
        align = "left")
    )
])
fig17.update_layout(title='Figure 20: Astra Zeneca questions')
fig17.show()
```

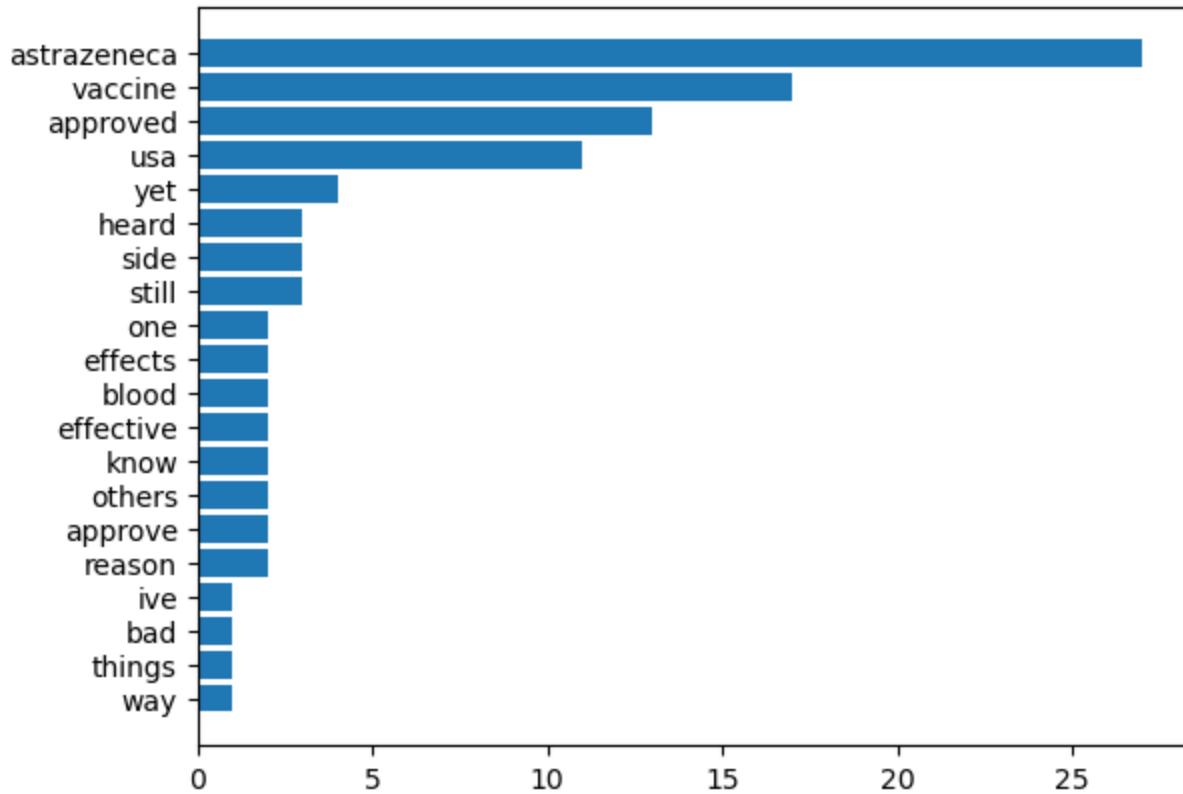
Figure 20: Astra Zeneca questions

Question	Count
why is astrazeneca not approved in the usa	19
there are many reports of severe side effects or deaths from the vaccine	2

which one of the vaccines should i take	2
can we choose which vaccine we want	1
can you mix the vaccines	1
did a volunteer in the oxford trial die	1
do i need to change my masking and social distancing practices depending on which covid-19 vaccine i got	1
	1

```
In [61]: # top 20 frequent words
most_common_words_all_questions(df_vaccine_sentence[df_vaccine_sentence['vaccine'] == 'a
```

Figure 21; Top 20 common words



## 6. References

Centers for Disease Control and Prevention. (2023, March). Safety of COVID-19 vaccines.  
<https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/safety-of-vaccines.html>

Centers for Disease Control and Prevention. (2023, March). Selected adverse events reported after COVID-19 vaccination. <https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/adverse-events.html>

Covid 19 Vaccine Tracker. (2023, March). Approved vaccines.  
<https://covid19.trackvaccines.org/vaccines/approved/>

Freeman, D., Lambe, S., Yu, L.-M., Freeman, J., Chadwick, A., Vaccari, C., Waite, F., Rosebrock, L., Petit, A., Vanderslott, S., Lewandowsky, S., Larkin, M., Innocenti, S., McShane, H., Pollard, A. J., & Loe, B. S. (2023). Injection fears and COVID-19 vaccine hesitancy. *Psychological Medicine*, 53(4), 1185-1195. <https://doi.org/10.1017/S0033291721002609>

Katella, K. (2023, January 30). Comparing the COVID-19 vaccines: How are they different? Yale Medicine. <https://www.yalemedicine.org/news/covid-19-vaccine-comparison>

Ruiz, J. B., & Bell, R. A. (2021). Predictors of intention to vaccinate against COVID-19: Results of a nationwide survey. *Vaccine*, 39(7), 1080-1086. <https://doi.org/10.1016/j.vaccine.2021.01.010>