

```
In [1]: import tweepy
from textblob import TextBlob
from wordcloud import WordCloud
import pandas as pd
import numpy as np
import re
import matplotlib.pyplot as plt
plt.style.use('fivethirtyeight')
```

```
In [2]: consumerKey = ""
consumerSecret = ""
accessToken = ""
accessTokenSecret = ""
```

```
In [3]: authenticate = tweepy.OAuthHandler(consumerKey, consumerSecret)
```

```
In [4]: authenticate.set_access_token(accessToken, accessTokenSecret)
```

```
In [5]: api = tweepy.API(authenticate, wait_on_rate_limit = True)
```

```
In [6]: # %%time
# tweets_df = pd.DataFrame()
# for index in range(0,900):
#     print(index)
#     posts = api.search_tweets(q="vaccination"+" -filter:retweets", lang="en", count=1000)
#     df = pd.DataFrame([(tweet.text, tweet.created_at, tweet.entities['hashtags']) for tweet in posts], columns=['tweet', 'date',
#     tweets_df = tweets_df.append(df, ignore_index=False)
```

```
In [7]: hastags = ['vaccination', 'vaccine', 'mrna', 'boostershot', 'covid19', 'Antibody', 'Allergy',
'Antigens', 'Antiviral', 'pandemic', 'efficacy', 'immunity', 'infectious', 'outbreak',
'placebo', 'quarantine', 'side effect', 'strain', 'pfizer', 'moderna', 'j&j', 'johnson&johnson',
'covaccine', 'covishield', 'biontech', 'janssen', 'zycov-d', 'astrazeneca', 'sputnik', 'convidecia',
'sinopharm', 'coronavac', 'novavax', 'epivaccorona', 'CancelCovid', 'HealingStartsHere', 'CovidVaccine',
'ImVaccinated', 'IGotVaccinated', 'VaccinesSaveLives', 'VaccinesWork', 'TheTruthAboutCovid', 'CovidTruths',
'CovidMyths', 'LifeInAPandemic', 'VaccinesSaveLives', 'trustscience', 'vaccinated', 'frontlineworkers',
'cdc', 'hospital', 'doctor', 'HealthForAll', 'Antivax', 'longtermcare', 'nurse', 'vax', 'deadvirus', 'virus',
'Coronavirus', 'mandatevaccine', 'omicron', 'deltavariant', 'spread', 'walgreens', 'cvs', 'pharmacy', 'Antivaxxer',
'forcedvaccination', 'adultvaccination', 'kidsvaccination', 'vaccinationrate', 'mutation', 'maskmandate', 'n95',
'flu', 'delta', 'ppe', 'ventilator', 'incubation', 'communityspread', 'Asymptomatic', 'Presymptomatic',
'NovelStrain', 'Pathogen', 'Socialdistancing', 'SelfIsolation', 'SelfQuarantine', 'SuperSpreader', 'Epidemic',
'SurgicalMask', 'HerdImmunity', 'SARS', 'SARS-CoV-2', 'covid-19', 'Cluster', 'Transmission', 'IncubationPeriod',
'DropletTransmission', 'FlattenTheCurve', 'DriveThruTesting', 'RTPCR', 'rapidtest',
'faceshield', 'covidtesting', 'lockdown', 'reopen', 'deltaplus']
```

```
In [8]: %%time
tweets_df = pd.DataFrame()

for index in range(0,10):
    for search in hastags:
        posts = api.search_tweets(q=search+" -filter:retweets", lang="en", count=1000)
        df = pd.DataFrame([(tweet.text, tweet.created_at, tweet.entities['hashtags']) for tweet in posts], columns=['tweet', 'date',
        tweets_df = tweets_df.append(df, ignore_index=False)
```

Wall time: 5.58 s

```
In [9]: tweets_df['date'] = tweets_df['date'].dt.tz_localize(None).dt.floor('d')
tweets_df.head(10)
```

```
Out[9]:
```

	tweet	date	hashtags
0	.@NFLPA should demand that @NFL345 investigate...	2021-12-08	[]
1	@mybmc My second dose of Sputnik V vaccination...	2021-12-08	[]
2	@arulscaria At the moment, however, it is not ...	2021-12-08	[]
3	Gov. Sanwo-Olu asks Lagosians to present COVID...	2021-12-08	[]
4	perhaps #omicron is a hint that mix-and-match...	2021-12-08	[{'text': 'omicron', 'indices': [8, 17]}, {'t...
5	Gov. Sanwo-Olu asks Lagosians to present COVID...	2021-12-08	[]
6	"Vaccine-resistant mutations"\n\nYou can't esc...	2021-12-08	[]
7	@stefanyvonne @jacobqvirin @bwfmedia I agree w...	2021-12-08	[]
8	@rpwhite3 @Breaking911 How about this article ...	2021-12-08	[]
9	The Orange Door network provides family violen...	2021-12-08	[]

```
In [10]: tweets_df.to_csv('vaccine_tweets_data.csv')
```