

```
In [1]: #Libreria de Twitter
import twint
import nest_asyncio
nest_asyncio.apply()
```

```
In [2]: # Configuracion y Busqueda en Twitter
```

```
conf = twint.Config()
conf.Search = "covid_19"
conf.Since = "2021-05-01 00:00:00"
conf.Until = "2021-05-12 00:00:00"
conf.Limit = 1000
conf.Store_csv = True
conf.Output = "covid_19.csv"
conf.Pandas = True
conf.Lang = "es"
twint.run.Search(conf)
```

a @naHadaDeNada Se ve que no se dieron nunca una vacuna contra la gripe que dice muy claro coronavirus... es más, en algún momento del año pasado se habló de que esas vacunas reconocían el corona del covid19 y las atacaban. Siento que estamos como en abril de 2020.

1391905645566337025 2021-05-10 18:59:00 -0500 <ghaffar1466> Hakeem Abdul Ghaffar Kasur Pk: Health News COVID-19 هلتھ نیو COOL WATER... <https://t.co/1s2AaJaNmN> (https://t.co/1s2AaJaNmN) راج ویلفیئر مطب (کلینک) حکیم عبد الغفار قصور طبیبہ شازیہ غفار

1391905645293711360 2021-05-10 18:59:00 -0500 <covid19brbot> Há 5.570 (100,00%) municípios com pelo menos um caso, e 5.490 (98,56%) com pelo menos um óbito. 59,93% dos casos e 48,95% dos óbitos foram no interior do Brasil. Mapa por município, gráficos, dados completos e atualizações em <https://t.co/nIg9dOss5b> (https://t.co/nIg9dOss5b)

1391905645243342850 2021-05-10 18:59:00 -0500 <colmedptarenas> Atención con el calendario de vacunación Covid-19 de esta semana. Se incorporan crónicos entre 30 y 34 años, además de esperar a los rezagados <https://t.co/cwvAZHrR7i> (https://t.co/cwvAZHrR7i)

1391905642739343362 2021-05-10 18:59:00 -0500 <WOWK13News> Lottery officials say anyone over the age of 18 who gets their first or second COVID-19 vaccine dose at a participating location will receive a coupon for a free Kentucky

In [2]: *# Leer el archivo CSV con Pandas y filtrado en Español*

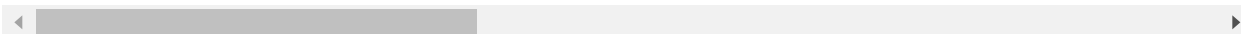
```
import pandas as pd
data = pd.read_csv("covid_19.csv")
espData = data[data['language'] == 'es']
espData
```

Out[2]:

	id	conversation_id	created_at	date	time	timezone
8	1391955828345819137	1391955828345819137	2021-05-10 22:18:25 Hora de verano central (Mé...	2021-05-10	22:18:25	-500 110775743
20	1391955815779741698	1391955815779741698	2021-05-10 22:18:22 Hora de verano central (Mé...	2021-05-10	22:18:22	-500 135892419
38	1391955789443645443	1391955789443645443	2021-05-10 22:18:16 Hora de verano central (Mé...	2021-05-10	22:18:16	-500 83268625
52	1391955770560892933	1391955770560892933	2021-05-10 22:18:11 Hora de verano central (Mé...	2021-05-10	22:18:11	-500
53	1391955769264848896	1391950408994762755	2021-05-10 22:18:11 Hora de verano central (Mé...	2021-05-10	22:18:11	-500
...
4319	1391902935223832584	1391902935223832584	2021-05-10 18:48:14 Hora de verano central (Mé...	2021-05-10	18:48:14	-500
4322	1391902924461248524	1391894801805623297	2021-05-10 18:48:12 Hora de verano central (Mé...	2021-05-10	18:48:12	-500

	id	conversation_id	created_at	date	time	timezone	
4323	1391902913153404929	1391902913153404929	2021-05-10 18:48:09 Hora de verano central (Mé...	2021-05-10	18:48:09	-500	
4325	1391902908304826373	1391902908304826373	2021-05-10 18:48:08 Hora de verano central (Mé...	2021-05-10	18:48:08	-500	
4327	1391902901942071296	1391902873961779205	2021-05-10 18:48:06 Hora de verano central (Mé...	2021-05-10	18:48:06	-500	126153682

772 rows × 36 columns



```
In [3]: # Extraer Twits
twits = pd.DataFrame(espData, columns=['tweet'])['tweet']
twits
```

```
Out[3]: 8      Parte del Dpto. Lima sigue a la baja y otra su...
20      12 M "Dia Internacional de la Enfermeria" Apl...
38      Arequipa: Mamá abandona el hospital bailando t...
52      El ciclismo está de duelo: A los 30 años Crist...
53      @Black_AngelxXV Todo bien en la COMARCA @Lucas...
...
4319     Porque los noticieros no informan el porcentaj...
4322     @CaracolDeportes Jajajaj la alcaldesa le parec...
4323     EE. UU.: hombre usa préstamos de ayuda por COV...
4325     Macri rompe promesa y se vacuna contra la COVI...
4327     Los países de la región lo venían haciendo bie...
Name: tweet, Length: 772, dtype: object
```

```
In [4]: # Filtrado y repeticiones de palabras
from sklearn.feature_extraction.text import CountVectorizer
count_vectorize = CountVectorizer(stop_words='spanish')
count_vectorize = CountVectorizer()
matriz = count_vectorize.fit_transform(twits)

repeticiones = pd.DataFrame(matriz.todense(), columns= count_vectorize.get_feature_names())
repeticiones
```

```
Out[4]:
```

	00	000	003	015	02	025	028	02wxm161i9	030	05	...	única	único	úsalo	DEMAGO
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
...
767	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
768	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
769	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
770	0	0	0	0	0	0	0	0	0	0	...	0	0	0	
771	0	0	0	0	0	0	0	0	0	0	...	0	0	0	

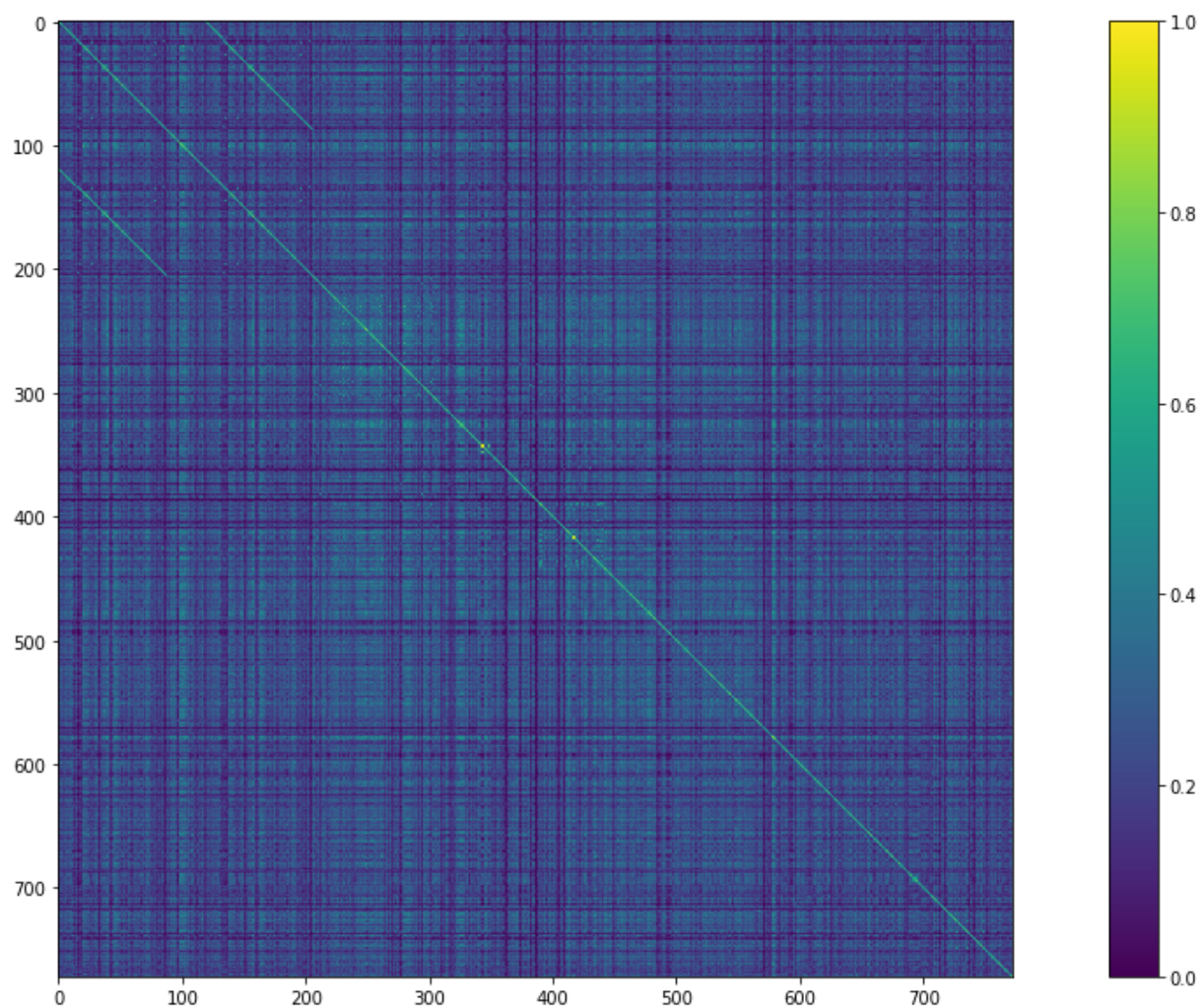
772 rows × 5093 columns

```
In [5]: # Similitud de cosenos
from sklearn.metrics.pairwise import cosine_similarity
similitudCosenos = cosine_similarity(repeticiones, repeticiones)
similitudCosenos
```

```
Out[5]: array([[1.          , 0.05311607, 0.05884899, ..., 0.05056499, 0.04494666,
0.08989331],
[0.05311607, 1.          , 0.28653413, ..., 0.22978626, 0.21884405,
0.33556088],
[0.05884899, 0.28653413, 1.          , ..., 0.24549513, 0.14547859,
0.14547859],
...,
[0.05056499, 0.22978626, 0.24549513, ..., 1.          , 0.08333333,
0.22222222],
[0.04494666, 0.21884405, 0.14547859, ..., 0.08333333, 1.          ,
0.22222222],
[0.08989331, 0.33556088, 0.14547859, ..., 0.22222222, 0.22222222,
1.          ]])
```

```
In [6]: # Graficar La Matriz de Confucion
import matplotlib.pyplot as plt
plt.figure(figsize=(20,10))
plt.imshow(similitudCosenos)
plt.colorbar()
```

Out[6]: <matplotlib.colorbar.Colorbar at 0x22c3b9f2700>



```
In [14]: from wordcloud import WordCloud, STOPWORDS
from stop_words import get_stop_words

stopwords = get_stop_words('spanish')
stopwords.append('https')
usernames = espData['username'].to_list()
for usr in usernames:
    stopwords.append(usr)

words = ''
for val in twits:
    val = str(val)
    tokens = val.split()
    for i in range(len(tokens)):
        tokens[i] = tokens[i].lower()
    words += " ".join(tokens)+" "

wordcloud = WordCloud(width = 1000,
                      height = 1000,
                      background_color='white',
                      stopwords = stopwords,
                      min_font_size = 5). generate(words)
plt.figure(figsize=(20,20))
plt.imshow(wordcloud)
plt.axis("off")
plt.show()
```



Análisis de Twists a través de la Matriz de Confusión

- En cuanto a la matriz de confusión, podemos observar principalmente que además de la línea principal, se genera otra línea secundaria entre los twitts 125 y 100, por lo que podemos decir que existe un tema en común en el tiempo o en twitts secundarios.
- También se puede observar que en la línea principal por el twitt 375, hay una bolita de color amarillo por lo que se puede decir que en estos puntos los twitt siguientes y posteriores hablaron de un tema común.
- También podemos ver que entre el twitt 375 y 450 hay un pequeño cubo que se empieza a generar de color amarillo por lo que se puede ver que estos tienen la misma estructura y podrían ser boots.
- Finalmente podemos decir que el fondo de la matriz, aunque es en su mayoría azul o morada, existen cruces de color amarillo por lo que se puede deducir que se hablo de un tema en común, luego pasaron a varios para seguir con otro tema en común, además que debes en cuando existen puntos amarillos entre twitts por lo que tienen una gran coincidencia.

Análisis de Twits a través de la nube de palabras

En este podemos observar que principalmente en el mes de mayo, que podemos ver que los principales temas son acerca de la vacuna contra covid y la pandemia.

In []: