

# **Reto de análisis**

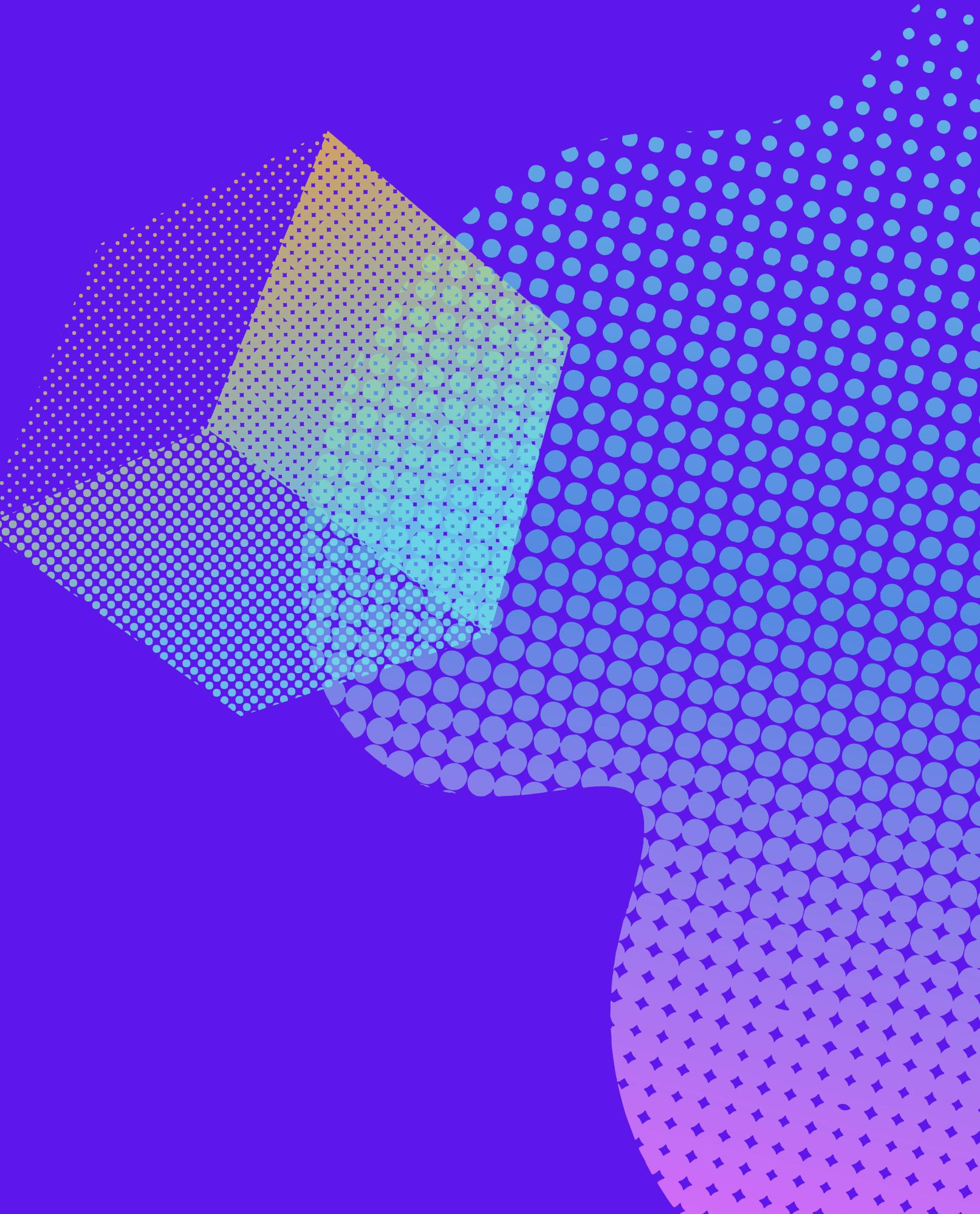
**Karen Paula Mayorga Guerrero**

**Duran Sanchez Pablo Ricardo A01663878**

**Andres Tavera Mihailide A01657332**

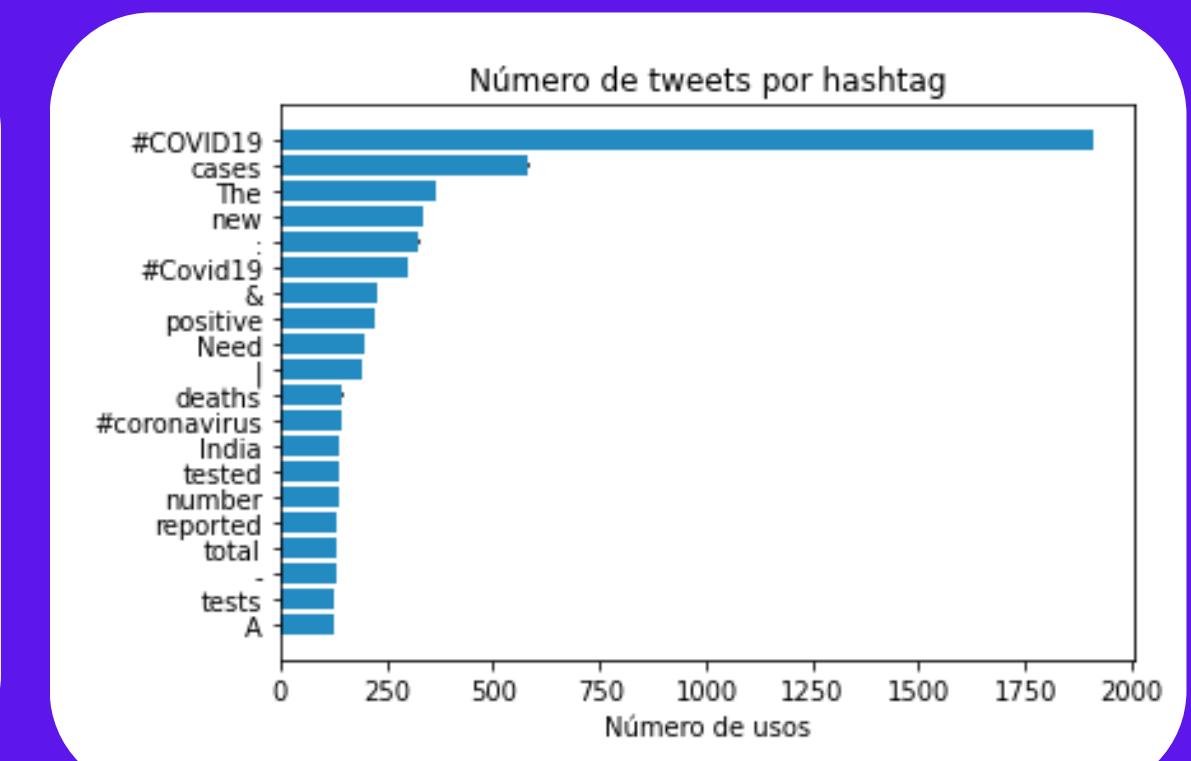
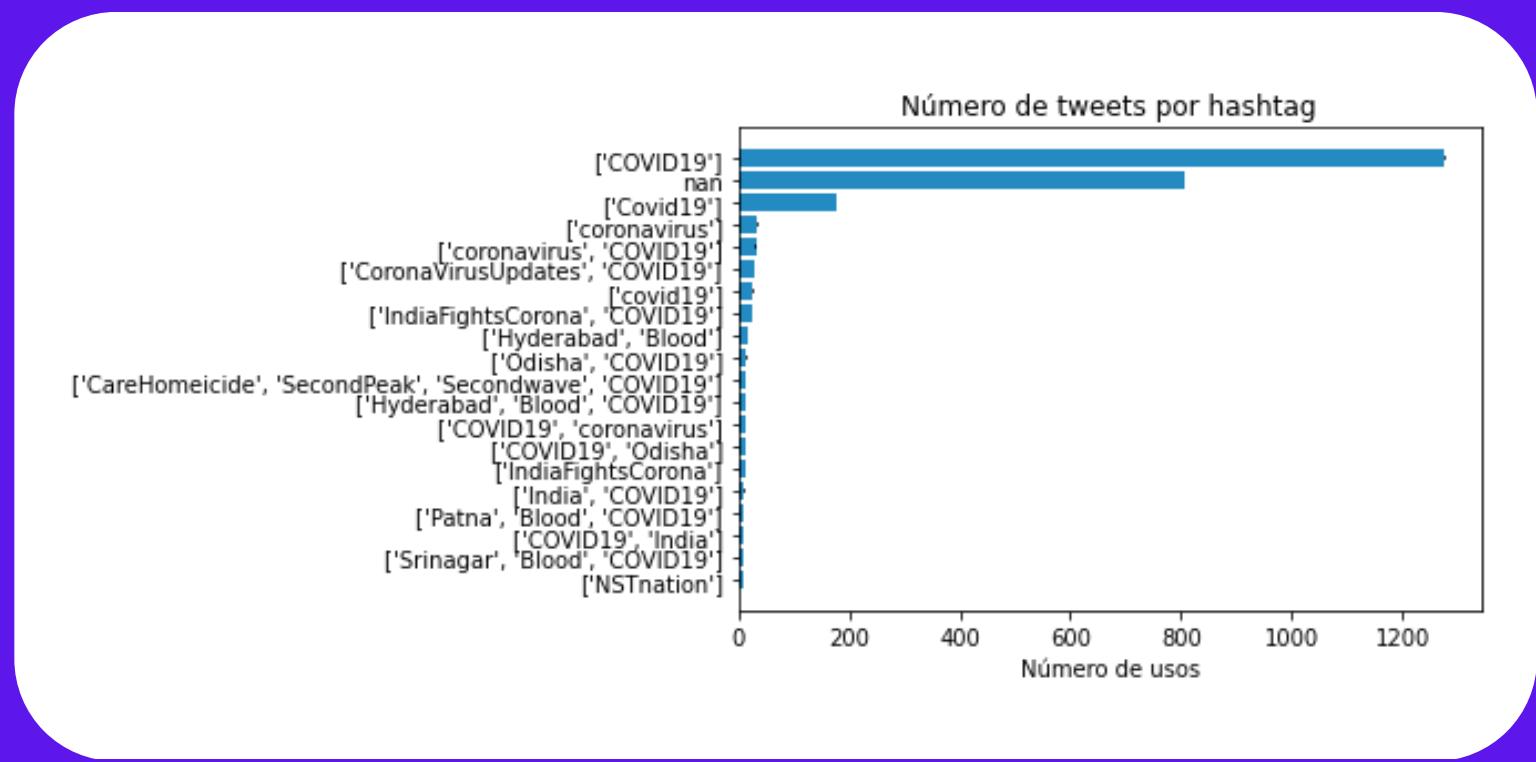
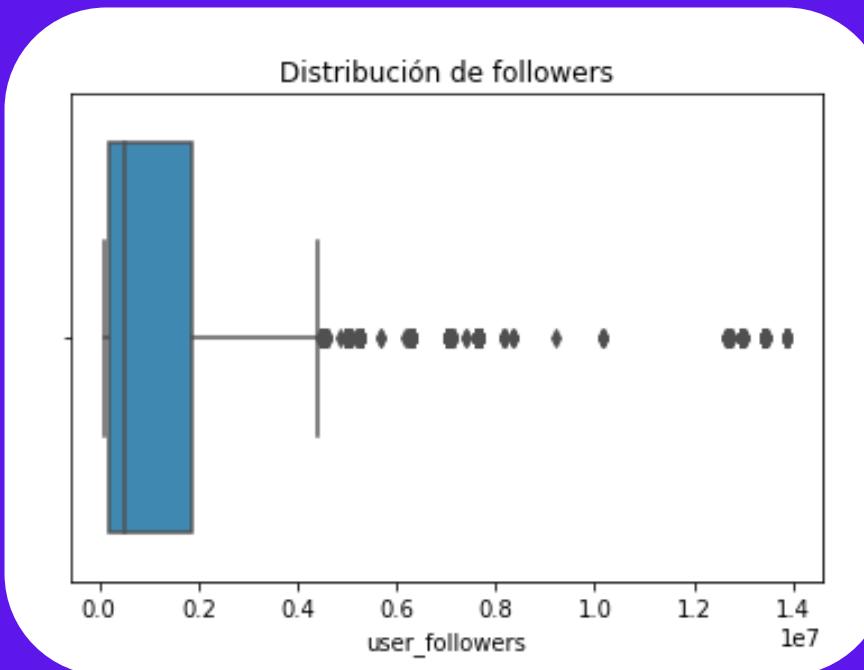
**José Antonio Pacheco Chargoy A01663815**

**Jorge Rodrigo Colín Rubio A1662960**



# Alejandro Mariacca Santin

Encontrar características entre usuarios con más de 100,000 seguidores a partir de los tweets que escriben



# Jorge Rodrigo Colín Rubio

Específico: Encontrar el máximo de número de personas en los países donde se utilizo el #Covid ademas del pais con mayor interracción con este hashtag,viendo desde que dispositivo se mando, si un iphone o un android

Medible:Conociendo el país de origen de las personas y los dispositivos desde donde se publico el tweet que estamos analizando, se determinó el país con mayor número activas en twitter durante la pandemia

Alcanzable:Aplicando los conocimientos de pandas y sus bibliotecas se realiza un análisis de la base de datos para poder alcanzar el objetivo anteriormente mencionado

Relevante:Se observa la población más activa de cada país así como el dispositivo que fue mas utilizado durante los tiempos covid

Tiempo: La meta debe de estar lista para antes de medio dia del 24 de marzo

```
In [6]: covid= dataframe.loc[dataframe.hashtags == "['COVID19']"]
covid
```

	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet
s	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	7254	False	2020-07-25 12:27:14	@diane3443 @wdunlap @realDonaldTrump Trump nev...	["COVID19"]	Twitter for Android	False
z	Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...	2019-03-07 01:45:06	197	987	1488	False	2020-07-25 12:27:10	@brookbanktv The one gift #COVID19 has give me...	["COVID19"]	Twitter for iPhone	False
6	Hotel living - various cities! Who needs a ho...	My ink "My Way...No Regrets"\nAlways Make Happ...	2010-03-25 21:16:20	0	8	32	False	2020-07-25 12:26:47	⚠️ @PattyHajdu @NavdeepSBains — no one will be ...	["COVID19"]	Twitter Web App	False
1	Africa	Official account of the Africa Youth Advisory ...	2019-05-13 06:27:47	830	254	3692	False	2020-07-25 12:26:47	Let's all protect ourselves from #COVID19.\nIt...	["COVID19"]	Twitter Web App	False
		Niko i ništa,						2020-	Holy water in times		Twitter	

```
In [7]: condition =(covid.source == "Twitter for Android") | (covid.source == "Twitter for iPhone")
```

```
In [8]: new = covid.loc[condition]
```

```
In [9]: new
```

```
Out[9]:
```

location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet
Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	7254	False	2020-07-25 12:27:14	@diane3443 @wdunlap @realDonaldTrump Trump nev...	["COVID19"]	Twitter for Android	False
Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...	2019-03-07 01:45:06	197	987	1488	False	2020-07-25 12:27:10	@brookbanktv The one gift #COVID19 has give me...	["COVID19"]	Twitter for iPhone	False
Graz	Niko i ništa, professor, so-called Balkan expe...	2009-06-18 09:46:10	18145	1389	13578	False	2020-07-25 12:26:28	Holy water in times of #COVID19 <a href="https://t.co/Y...">https://t.co/Y...</a>	["COVID19"]	Twitter for Android	False
Manhattan, NY	These days, I expose colonizers & exploits @t...	2008-02-23 19:02:29	86	259	9412	False	2020-07-25 12:26:26	I can imagine the same people profiting off th...	["COVID19"]	Twitter for Android	False
Top Heart	Inquisitive; Open minded but clear-	2014-02-03	6156	5463	193090	False	2020-07-25	@NicoleKowalski5 @chrissdameanor19	["COVID19"]	Twitter for	False

```
In [10]: local = new.user_location.dropna().value_counts()
```

```
In [11]: condicion = local > 41
```

```
In [12]: local
```

```
Out[12]:
```

United States	174
New Delhi, India	82
India	78
Los Angeles, CA	52
London	44

...  
\* - ⚑ || ⚒ - ⚓ || Asc - ⚖ 1

Hargeisa-Somaliland	1
Jamnagar, India	1
Odisha, India	1
Lost Wages, NV	1

```
Name: user_location, Length: 2829, dtype: int64
```

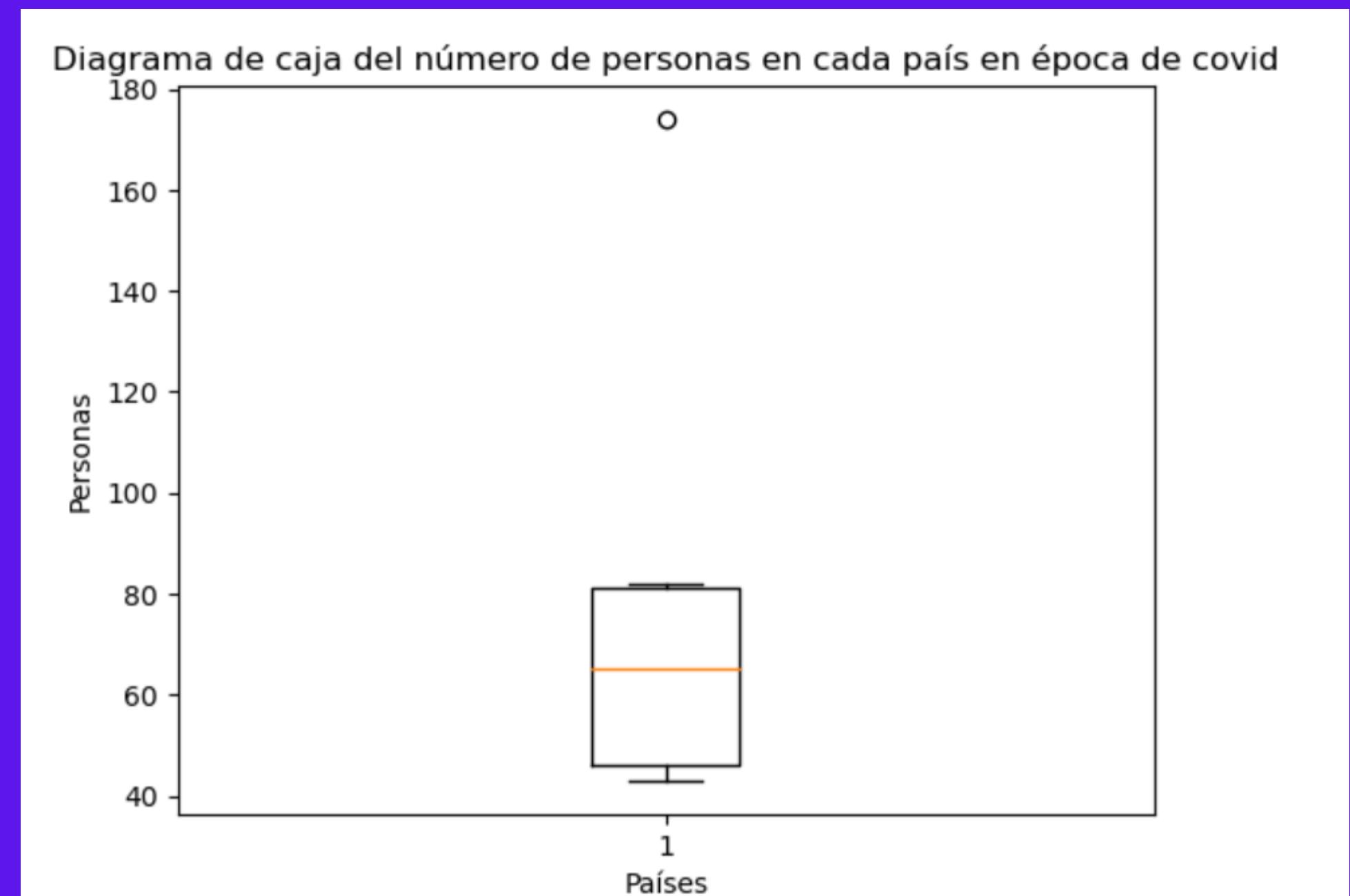
  

```
In [13]: conteo = local.loc[condicion]
```

```
In [14]: plt.boxplot(conteo)
plt.title('Diagrama de caja del número de personas en cada país')
plt.xlabel('Países')
plt.ylabel('Personas')
print()
```

**RESULTADOS**  
SE OBSERVÓ QUE ESTADOS UNIDOS  
FUE EL PAÍS CON MAYOR NÚMERO DE  
PERSONAS PUBLICANDO TWEETS  
ACERCA DEL COVID 19 MEDIANTE LOS  
DISPOSITIVOS IPHONE Y ANDROID



# Karen Paula Mayorga Guerrero

## OBJETIVO:

DETERMINAR EL IMPACTO EN LAS VARIABLES USER FOLLOWERS, USER FRIENDS Y USER FAVORITES PARA LOS USUARIOS DE TWITTER POR MEDIO DE DIFERENTES ANÁLISIS A TRAVÉS DE LA PRIMERA BASE Y ÚLTIMA BASE DE DATOS DEL AÑO 2020 Y GRACIAS A LAS HERRAMIENTAS PROPORCIONADAS POR LA MATERIA ESTA SEMANA ES QUE SE LOGRARÁ LLEVAR A CABO ESTA ACTIVIDAD. ADEMÁS, LO ANTERIOR SE LOGRARÁ TRABAJANDO EN UN LAPSO DE 8 HORAS EL DÍA 23 DE MARZO.



```

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
#inicio de la documentacion de tweets
data1 = pd.read_csv("covid19_tweets_inicio.csv",index_col = 'date')
data1.head(5)

```

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	text	hash
date										
2020-07-25 12:27:21	Vito↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	If I smelled the scent of hand sanitizers toda...	
2020-07-25 12:27:17	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	24	True	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	
2020-07-25 12:27:14	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	7254	False	@diane3443 @wdunlap @realDonaldTrump Trump nev...	['COV']
2020-07-25 12:27:10	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...	2019-03-07 01:45:06	197	987	1488	False	@brookbanktv The one gift #COVID19 has give me...	['COV']
2020-07-25 12:27:08	DIPR-J&K	Jammu and Kashmir	✓ Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	101009	168	101	False	25 July : Media Bulletin on Novel #CoronaVirus... ['CoronaVirusUpd']	['COV']

```

[17] #final de la documentacion de tweets
data2 = pd.read_csv("covid19_tweets_final.csv",index_col = 'date')
data2.head(5)

```

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	text	hash
date										
2020-07-25 12:27:21	Vito↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	If I smelled the scent of hand sanitizers toda...	
2020-07-25 12:27:17	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	24	True	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	
2020-07-25 12:27:14	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	7254	False	@diane3443 @wdunlap @realDonaldTrump Trump nev...	['COV']
2020-07-25 12:27:10	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...	2019-03-07 01:45:06	197	987	1488	False	@brookbanktv The one gift #COVID19 has give me...	['COV']
2020-07-25 12:27:08	DIPR-J&K	Jammu and Kashmir	✓ Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	101009	168	101	False	25 July : Media Bulletin on Novel #CoronaVirus... ['CoronaVirusUpd']	['COV']

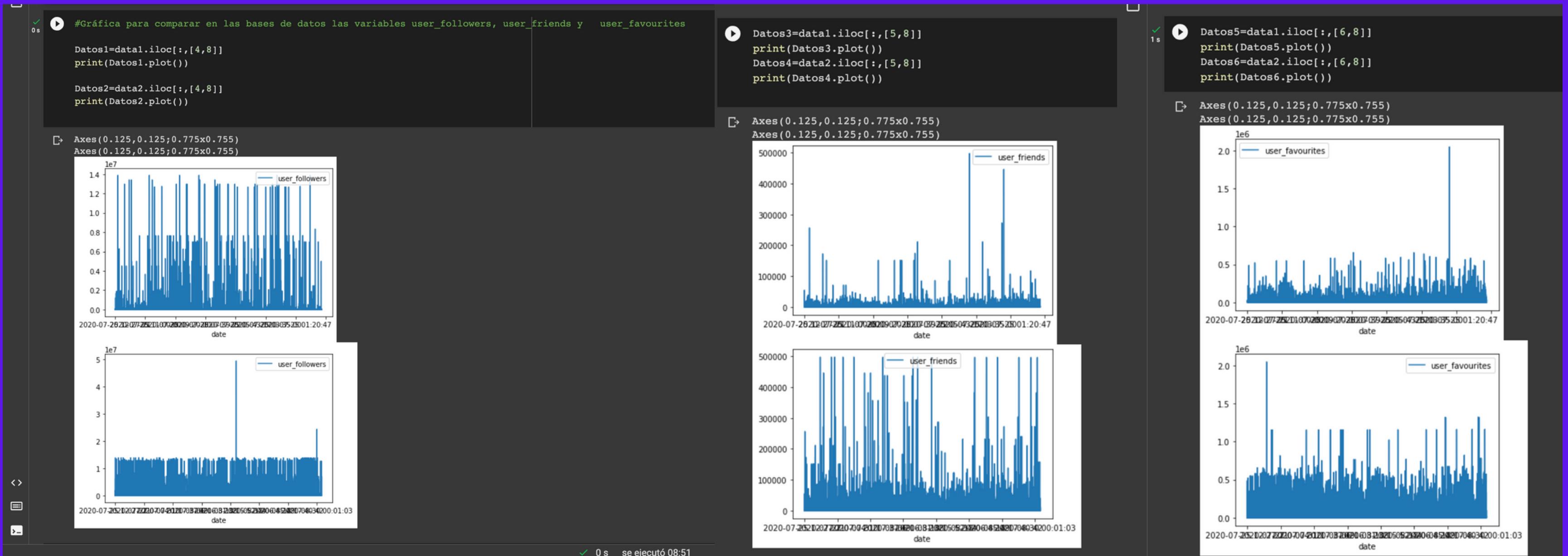
```

#Se crea una base de datos en base a la base inicial y la base final
superData=pd.merge(data1,data2, left_on="user_name", right_on="user_name")
superData.head(5)

```

	user_name	user_location_x	user_description_x	user_created_x	user_followers_x	user_friends_x	user_favourites_x	user_verified_x	text_x	hashtags_x
0	Vito↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	If I smelled the scent of hand sanitizers toda...	NaN
1	Vito↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	If I smelled the scent of hand sanitizers toda...	NaN
2	Vito↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	I miss isopropyl alcohol so much!!!! Ethanol i...	['COVID19']
3	Vito↑	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	I miss isopropyl alcohol so much!!!! Ethanol i...	['COVID19']
4	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	24	True	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	NaN

5 rows x 23 columns



```
#Análisis comparativo para las 3 variables
print(data1['user_followers'].describe())
print("\n")
print(data2['user_followers'].describe())
print("\n")
print(data1['user_friends'].describe())
print("\n")
print(data2['user_friends'].describe())
print("\n")
print(data1['user_favourites'].describe())
print("\n")
print(data2['user_favourites'].describe())
print("\n")
```

	user_followers	user_favourites
count	1.717600e+04	1.791080e+05
mean	1.213559e+05	1.090555e+05
std	8.839637e+05	8.414670e+05
min	0.000000e+00	0.000000e+00
25%	1.680000e+02	1.720000e+02
50%	9.510000e+02	9.920000e+02
75%	5.360000e+03	5.284000e+03
max	1.389284e+07	4.944256e+07
Name:	user_followers, dtype: float64	user_favourites, dtype: float64
count	1.717600e+04	1.791080e+05
mean	1.421896e+04	1.444411e+04
std	4.448230e+04	4.452270e+04
min	0.000000e+00	0.000000e+00
25%	1.760000e+02	2.060000e+02
50%	1.705000e+03	1.791000e+03
75%	9.378500e+03	9.388000e+03
max	2.047197e+06	2.047197e+06
Name:	user_favourites, dtype: float64	user_favourites, dtype: float64
count	17176.000000	179108.000000
mean	2083.050885	2121.701566
std	8750.942968	9162.553072
min	0.000000	0.000000
25%	133.750000	148.000000
50%	506.000000	542.000000
75%	1703.250000	1725.250000
max	497363.000000	497363.000000
Name:	user_friends, dtype: float64	user_friends, dtype: float64

```

[43] Datos11=superData.iloc[:,[4,15]]
print(Datos11.corr())
print("\n")

Datos11=superData.iloc[:,[4,16]]
print(Datos11.corr())
print("\n")

Datos11=superData.iloc[:,[4,17]]
print(Datos11.corr())
print("\n")

```

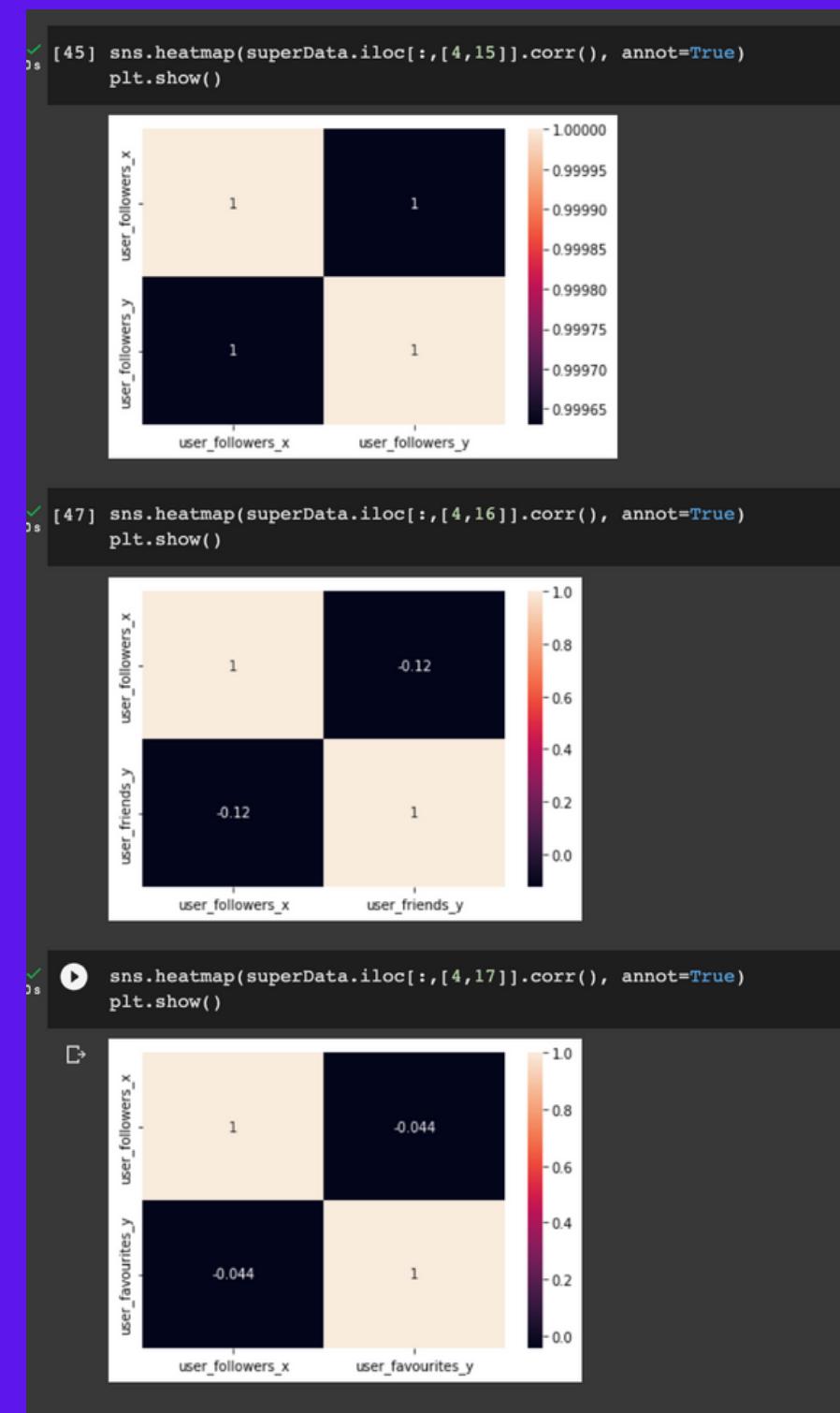
	user_followers_x	user_followers_y
user_followers_x	1.00000	0.99963
user_followers_y	0.99963	1.00000

	user_followers_x	user_friends_y
user_followers_x	1.000000	-0.124262
user_friends_y	-0.124262	1.000000

	user_followers_x	user_favourites_y
user_followers_x	1.000000	-0.043546
user_favourites_y	-0.043546	1.000000



# José Antonio Pacheco

## Objetivo SMART

### Especifico

DETERMINAR LA EVOLUCIÓN DE LOS TEMAS DE INTERÉS EN TWITTER (MEDIANTE HASHTAGS DE LOS TWEETS DE LOS USUARIOS) A LO LARGO DEL TIEMPO DURANTE LA PANDEMIA DE COVID-19.

### Relevante

SE BUSCA ALCANZAR LA SUBCOMPETENCIA "SING0202A ", PUESTO QUE SE BUSCA DETERMINAR LA INTERACCIÓN QUE EXISTE ENTRE DOS VARIABLES.

### Medible

SE DETERMINARÁN LOS TEMAS MÁS RELEVANTES EN EL TIEMPO DE ACUERDO A LA MAYOR CANTIDAD DE HASHTAGS USADOS POR LOS USUARIOS.

### A tiempo

SE ESTIMA QUE LA RESOLUCIÓN DE ESTE RETO TOMARÁ MENOS DE 8 HORAS EN TOTAL. LO CUAL ES TIEMPO SUFFICIENTE PARA PRESENTARLOS A LAS 12:00PM DEL DÍA SIGUIENTE.

### Accesible

SE ANALIZARÁ LA CORRELACIÓN ENTRE LOS HASHTAGS EMPLEADOS Y LA FECHA ESPECÍFICA EN LA QUE LOS USUARIOS DE TWITTER LOS USARON, EMPLEANDO HERRAMIENTAS DE ANÁLISIS DE DATOS APRENDIDAS EN EN CURSO.



# Creando un dataframe que solo contenga hashtags sobre Covid-19

```
# Obtenemos todos los tweets relacionados con el Covid19 = 42673

data = data.fillna("")
dataframecovid = data[data["hashtags"].str.contains("COVID19") | data["hashtags"].str.contains("covid19") | data["hashtags"].str.contains("coronavirus") | data["hashtags"].str.contains("CORONAVIRUS") | data["hashtags"].str.contains("Covid19") | data["hashtags"].str.contains("Corona")]
dataframecovid
```

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet
2	Time4fisticuffs	Pewee Valley, KY	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	9275	9525	7254	False	2020-07-25 12:27:14	@diane3443 @wdunlap @realDonaldTrump Trump ne...	['COVID19']	Twitter for Android	False
3	ethel mertz	Stuck in the Middle	#Browns #Indians #ClevelandProud #[] #Cavs ...	2019-03-07 01:45:06	197	987	1488	False	2020-07-25 12:27:10	@brookbanktv The one gift #COVID19 has give me...	['COVID19']	Twitter for iPhone	False
4	DIPR-J&K	Jammu and Kashmir	/ Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	101009	168	101	False	2020-07-25 12:27:08	25 July : Media Bulletin on Novel #CoronaVirus...	['CoronaVirusUpdates', 'COVID19']	Twitter for Android	False
5	Franz Schubert	Новороссия	🇪🇺 #Новороссия #Novorossiya #оставайсядома #S...	2018-03-19 16:29:52	1180	1071	1287	False	2020-07-25 12:27:06	#coronavirus #covid19 deaths continue to rise...	['coronavirus', 'covid19']	Twitter Web App	False
6	hr bartender	Gainesville, FL	Workplace tips and advice served up in a frien...	2008-08-12 18:19:49	79956	54810	3801	False	2020-07-25 12:27:03	How #COVID19 Will Change Work in General (and ...	['COVID19', 'Recruiting']	Buffer	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...
74430	Jimmy Slamat	Los Angeles, CA	I follow the market and all it's astounding ev...	2009-08-06 08:39:31	49	82	281	False	2020-08-04 03:13:29	So maybe Sweden did do it right, by doing noth...	['COVID19', 'coronavirus']	Twitter for iPhone	False
74431	Laura Wolfson	Lexington, KY	The only things I collect are memories.	2010-09-24 02:01:15	85	586	1902	False	2020-08-04 03:13:29	So far this summer I have filled up my lawn mo...	['COVID19', 'QuarantineLife']	Twitter for iPhone	False
74432	Professor Tonya M. Evans	👉 #stayathome	Law Prof @DickinsonLaw & Entrepreneur   Crypto...	2013-05-14 20:15:24	4289	1066	53569	False	2020-08-04 03:13:26	ICYMI: REPLAY: #TechIntersect ✨ #16: Isaiah "@B...	['TechIntersect', 'Bitcoin', 'COVID19']	Twitter Web App	False
74434	M0ser		Reagan conservative and attorney raised in the...	2014-02-18 03:46:28	2554	1733	129104	False	2020-08-04 03:13:19	If only we had a responsible media to warn us ...	['COVID19']	Twitter for iPhone	False
74435	Your Friend & Sabre ✌	Chicago, IL	My spectral decomposition has a significant da...	2016-12-19 19:55:00	310	1748	60133	False	2020-08-04 03:13:15	MAGA: #COVID19 is just a cold & it'd be go...	['COVID19', 'Hydroxychloroquine']	Twitter for Android	False

# Creando un dataframe que solo contenga hashtags sobre cubrebocas

[111] # Obtenemos todos los tweets relacionados con los cubrebocas/mascarillas = 1414														
	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags	source	is_retweet	
71	BestValueButtons	Denver, CO	We are a fabric outlet and specialty button Et...	2018-05-30 19:45:27	1267	2869	1204	False	2020-07-25 12:25:11	Check out Angel's beautiful mask and hair bow ...	['etsy', 'epiconetsy', 'masksforsale']	IFTTT	False	
103	Thomas Faires	Toronto, Ontario, Canada	...yup, still a husband, vegetarian, computer ...	2011-04-14 12:36:09	738	143	28004	False	2020-07-25 12:24:10	1.28% of the U.S. population is infected with ...	['COVID19', 'TrumpVirus', 'MaskItOrCasket']	Twitter for iPhone	False	
147	Illustrated London		Buy a print for your wall and look out onto an...	2020-07-25 12:04:11	1	13	0	False	2020-07-25 12:23:03	Mask Print #1 (Londoners Wear Masks) https://t...	['MaskUp', 'mask', 'Masks', 'MaskMoaners', 'CO...']	Twitter Web App	False	
169	Christine Murphy Estes, MM, MA, CCC-SLP	New York, NY	Speech-Language Pathologist, Voice Specialist,...	2018-10-07 02:31:06	622	1220	1121	False	2020-07-25 12:21:51	Your friendly weekend reminder that COVID didn...	['wearamask', 'covid', 'covid19']	Twitter for iPhone	False	
212	coley	Grays, England	26, Essex UK. bearded. long haired. engaged. p...	2015-12-11 22:20:21	33	188	108	False	2020-07-25 12:20:36	STOP FOGGY GLASSES WITH THIS HACK!!!! #COVID...	['COVID19', 'facecoverings', 'maskproblems', '...']	Twitter for iPhone	False	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
74289	A Voice For Choice Advocacy			2015-05-22 02:58:19	7409	658	2665	False	2020-08-04 03:19:00	I Read the #Mask Box WARNING to People at @Wal...	['Mask', 'Covid19', 'BitChute']	TweetDeck	False	
74364	The Bright Flame Official		The Bright Flame is a Bible-based, male-centri...	2020-06-10 21:29:20	2	0	0	False	2020-08-04 03:15:45	Email Sent! Masks Endanger The Childrens Futur...	['masks', 'covid']	Twitter for iPhone	False	
74385	Naya Z.	Everywhere	iSing•iWrite•iPromote love & cuss words•Stream...	2009-08-04 15:59:53	3111	126	538	False	2020-08-04 03:14:59	He's sick of y'all's sh*t!! \n#wearyodamnmask ...	['wearyodamnmask', 'covid19', 'thepandemicisnt...']	Instagram	False	
74394	Karen Cookson	California Baby	Computer person by trade, politics buff for fu...	2009-03-25 03:20:33	84	366	1150	False	2020-08-04 03:14:45	You think wearing a mask is uncomfortable try ...	['wearamask', 'covid19', 'ventilator']	Instagram	False	
74421	Tamela Jaco	Indiana	Servant of God, Domestic Violence Survivor, Lo...	2011-04-13 22:15:18	633	2195	2420	False	2020-08-04 03:13:55	Fabric Face Masks. Another satisfied customer...	['smallbusinessowners', 'covid19', 'facemasksf...']	Instagram	False	

1414 rows × 13 columns

# Creando un dataframe que solo contenga hashtags sobre Donald Trump

103	Thomas Faires	Toronto, Ontario, Canada	...yup, still a husband, vegetarian, computer ...	2011-04-14 12:36:09	738	143	28004	False	2020-07-25 12:24:10	1.28% of the U.S. population is infected with ...	[COVID19', 'TrumpVirus', 'MaskItOrCasket']	Twitter for iPhone	False			
234	BeautifulSkinYourIn		Daily Tips from our Nurses, Dietitians, Aesthe...	2016-11-17 14:30:31	2401	3053	69	False	2020-07-25 12:20:00	WORLD #NEWS #TODAY SHORTENED FOR BUSY PEOPLE\n...	['NEWS', 'TODAY', 'politics', 'trump', 'money']	Twitter for Advertisers	False			
412	marklaverdure	Los Angeles & Billings	I have passion for living life to it's fullest...	2011-03-04 15:17:55	10078	10316	19623	False	2020-07-25 12:13:53	@realDonaldTrump #trumpdeathtoll147K 4.2 Milli...	['trumpdeathtoll147K', 'COVID19', 'TrumpVirusC...']	Twitter Web App	False			
437	Mx. Joe~Anthony Sierra Let us pray #COVID19 away #StGeorge~#StatenIsland~#NYC		Hope to be insightful	2013-05-03 03:46:03	355	228	3685	False	2020-07-25 12:13:05	#Hope #USA #Trump vilifying #USPS & #Amazo...	['Hope', 'USA', 'Trump', 'USPS', 'Amazon', 'Po...']	Twitter Web App	False			
448	marklaverdure	Los Angeles & Billings	I have passion for living life to it's fullest...	2011-03-04 15:17:55	10078	10316	19623	False	2020-07-25 12:12:49	@realDonaldTrump Remember the #WorstPresident...	['WorstPresidentInHistory', 'trumpdeathtoll147...']	Twitter Web App	False			
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
74313	Mindcite	New York, NY	Beyond Memes... Firing up fascism-RESISTant cr...	2016-01-15 22:59:56	3349	3741	13133	False	2020-08-04 03:18:00	Official: The most nonpartisan Public Safety M...	[COVID19', 'TrumpLiesPeopleDie']	Twitter Web App	False			
74340	Crinklybum	United States	Genealogy, History, Traveler, USAF Veteran	2012-08-29 03:34:50	523	1019	127295	False	2020-08-04 03:16:52	@HillaryWarnedUs Where is the #TrumpGestapo to...	['TrumpGestapo', 'COVID19']	Twitter for iPhone	False			
74351	FDMillet	Maine, USA	You have to disturb the soil to grow a garden....	2011-09-15 12:20:19	2584	2155	6487	False	2020-08-04 03:16:30	#Trump and the #GOP's #coronavirus aid coronav...	['Trump', 'GOP', 'coronavirus']	Twitter Web App	False			
74357	Said Pulido	Mexico City	Periodista & Fotógrafo (Negocios, Responsabili...	2011-04-21 02:56:31	9923	4638	3996	False	2020-08-04 03:16:00	President Donald #Trump criticized Dr. Deborah...	['Trump', 'Birx', 'COVID19']	Twitter for iPhone	False			
74429	Left of Center	On a soapbox	I lean left of center on the political dial. T...	2015-07-29 00:45:47	79	91	137	False	2020-08-04 03:13:29	The #FDA and prominent doctors have said #hydr...	['FDA', 'hydroxychloroquine', 'covid19', 'Trump']	Twitter for iPhone	False			

# Creando data frames de cada tema para cada día

# Obtenemos los tweets sobre Covid19 escritos el 27 de julio de 2022												
dataframecovid_27jul = dataframecovid[dataframecovid["date"].str.contains("2020-07-27")] dataframecovid_27jul												
	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date			
24677	Bradley...	Hampshire	Creative geek, data-analyst, student of Zen, t...	2012-04-11 18:36:02	140	201	788	False	2020-07-27 07:17:19			
24678	Rajesh Kshitij	hyderabad	News 18 Network	2010-07-25 10:45:13	504	980	4613	False	2020-07-27 07:17:18			
24680	Rosebank Killarney	Johannesburg	All your local news supplied by the Rosebank K...	2010-08-31 08:55:43	5508	1313	348	False	2020-07-27 07:17:18			
24681	Arizton	Chicago, IL	A premium market research & consulting company...	2017-10-03 10:35:28	36	25	138	False	2020-07-27 07:17:16			
24683	Farhad TOOR	Paris	Conseiller citoyen   Porteur de #projet   équi...	2013-11-09 15:06:31	116	627	1959	False	2020-07-27 07:17:10			
...	...	...	...	...	...	...	...	...	...	...	...	...
32167	Pat Gannon	Chicago/Indiana		2010-09-15 17:33:36	312	358	515	False	2020-07-27 01:32:16			
32168	Andrew Miller Loves Bees 🐝🐝 & Greta Thunberg	Sydney	Believes in #renewables #ethicalfashion #susta...	2013-01-09 06:05:59	9455	9084	51238	False	2020-07-27 01:32:10			
32169	Republican Quotes	Right Coast USA	Sad to say but I am NOT a parody account. #TCO...	2010-11-19 20:51:55	22918	11529	130	False	2020-07-27 01:32:09			
32174	The Australian	Sydney, Australia	News from The Australian.	2007-10-22 23:57:33	728619	546	681	True	2020-07-27 01:32:01			
32175	DoingMyBest		Started the account to express my anger, but n...	2016-11-14 23:18:30	30	281	5113	False	2020-07-27 01:32:00			

4312 rows × 13 columns

# Obtenemos los tweets sobre cubrebocas/mascarillas escritos el 31 de julio de 2022												
dataframemask_31jul = dataframemask[dataframemask["date"].str.contains("2020-07-31")] dataframemask_31jul												
	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date			
44480	8min 46sec	Canberra	Twitter is my ONLY vice ... cartoons on @FB ...	2011-09-14 02:31:09	7021	2199	26025	False	2020-07-31 19:19:08			
44553	Josh Mumbrue	Wisconsin, USA	FVTC - Professional Communications, 2021. Land...	2015-07-23 17:59:10	57	416	1121	False	2020-07-31 19:17:41			
44604	8min 46sec	Canberra	Twitter is my ONLY vice ... cartoons on @FB ...	2011-09-14 02:31:09	7021	2199	26025	False	2020-07-31 19:16:37			
44605	Mamta Suri		Senior #softwaredevelopment Manager   ERG Lead...	2017-03-03 23:40:16	66	116	311	False	2020-07-31 19:16:37			
44759	Gail F 🇺🇸 ca		Just a simple girl trying to trudge through th...	2009-04-23 23:12:26	9	70	357	False	2020-07-31 19:13:49			
...	...	...	...	...	...	...	...	...	...	...	...	...
51690	PROMOrx	USA	Since 2000, we've been prescribing branded pro...	2009-03-10 23:31:31	1608	2058	369	False	2020-07-31 17:03:33			
51693	American Nurse		American Nurse, the official journal of the Am...	2009-03-17 13:51:07	43823	4503	3031	False	2020-07-31 17:03:23			
51700	B Wellard	Idaho, USA	Retired Engineer, Math & Science Teacher, Musi...	2014-01-26 16:02:58	6	297	18	False	2020-07-31 17:03:17			
51821	@cityofIndepmo	Independence, MO	Official account for the city of Independence, ...	2011-09-14 18:54:44	6949	366	2671	True	2020-07-31 17:01:16			
51831	Chadia Wannous		Passionate about public health, prevention, pr...	2017-01-10 10:43:13	100	202	296	False	2020-07-31 17:01:08			

137 rows × 13 columns

# Obtenemos los tweets sobre Donald Trump escritos el 01 de agosto de 2022												
dataframetrump_01ago = dataframetrump[dataframetrump["date"].str.contains("2020-08-01")] dataframetrump_01ago												
	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date			
51955	The Bad Angler		Angling writer and blogger at Fly Fishing and ...	2011-10-15 21:24:47	6265	6941	3516	False	2020-08-01 19:20:54			
51997	frannie o'grady 🇺🇸	Texas	Yeah, I'm a woman like that too. She/Her. I bl...	2010-02-27 20:46:02	1168	4676	14731	False	2020-08-01 19:19:33			
52008	Leesa	East Coast USA Planet Earth	Critical Thinker. Transformational Guru. Thou...	2012-01-01 04:03:03	1978	4489	10774	False	2020-08-01 19:18:57			
52031	Tommy Tirader 🏳️			2011-08-11 14:44:22	266	548	39986	False	2020-08-01 19:18:10			
52045	Thomas T. Thai		Helping communities be healthier and happier.	2018-02-23 07:51:16	3	1	8	False	2020-08-01 19:17:37			
...	...	...	...	...	...	...	...	...	...	...	...	...
59022	dan bolger	Santa Cruz, CA	Progressive before it had a name(1954).Never t...	2013-06-07 22:44:56	2745	2767	43559	False	2020-08-01 15:53:41			
59082	Wizastudios 🎨	Brooklyn, NY	Product Designer at @WorkMarket, #LatinxInTech...	2010-04-05 02:44:07	510	1568	2099	False	2020-08-01 15:51:47			
59211	roro	San Francisco, CA	Many people are afraid of the dark, but the re...	2014-05-31 00:25:30	312	586	6025	False	2020-08-01 15:47:38			
59241	Jasper #Ω.	Right behind you	#UnitedBlue https://t.co/xdhYM8QhAA PROUD Memb...	2012-03-09 18:55:38	14364	7726	225	False	2020-08-01 15:46:31			
59411	Sandrine	Los Gatos		2009-04-09 01:55:34	93	192	52	False	2020-08-01 15:42:12			

132 rows × 13 columns

# Usando matplotlib para graficar un histograma

```
# Creamos un arreglo de los días en los que se escribieron los Tweets (para ajustarlos a la gráfica)

dias = [" ", "Jul 25", "Jul 27", "Jul 29", "Jul 31", "Ago 02", "Ago 04"]

# Creamos un arreglo almacenando los valores obtenidos de la cantidad de tweets sobre el Covid19

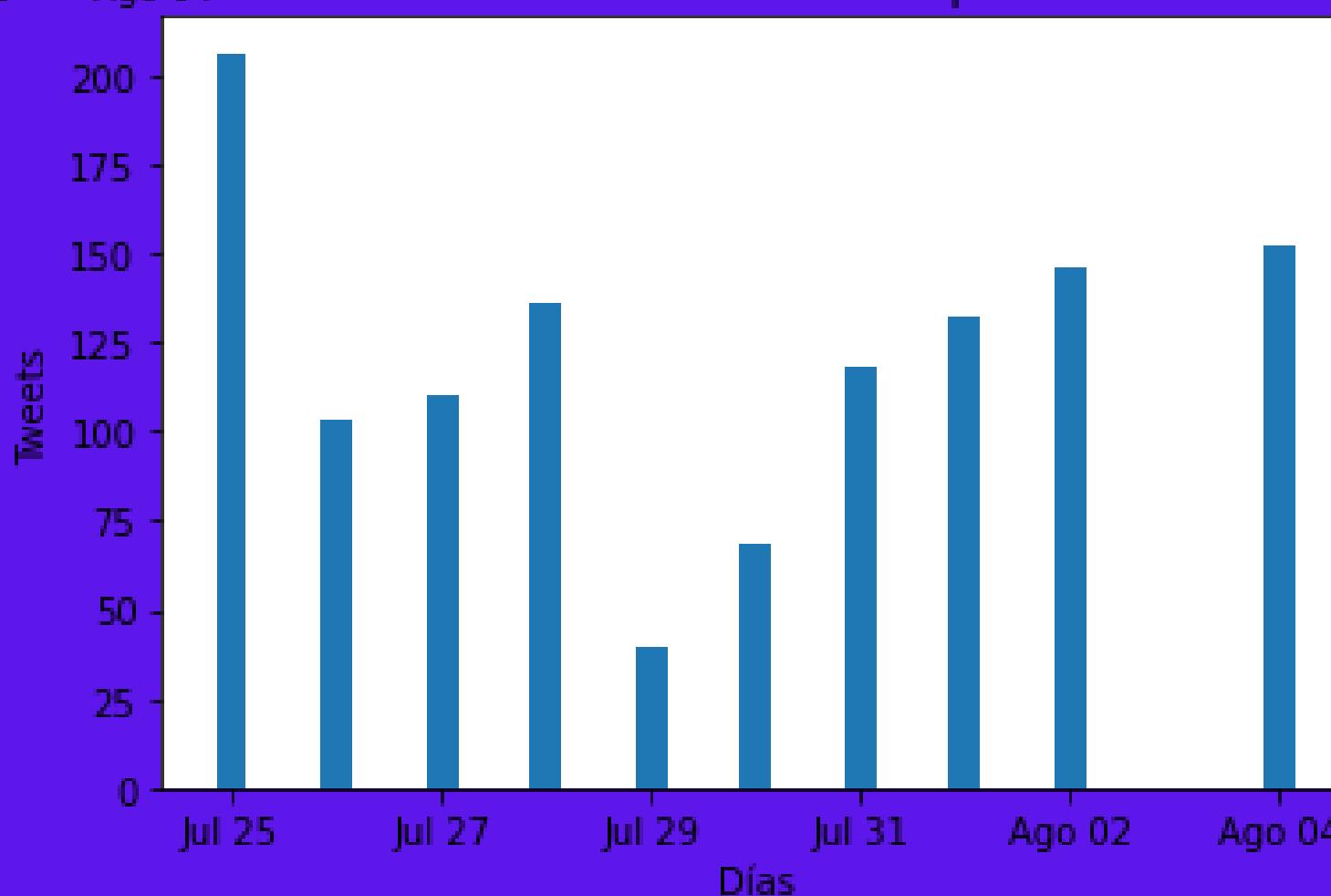
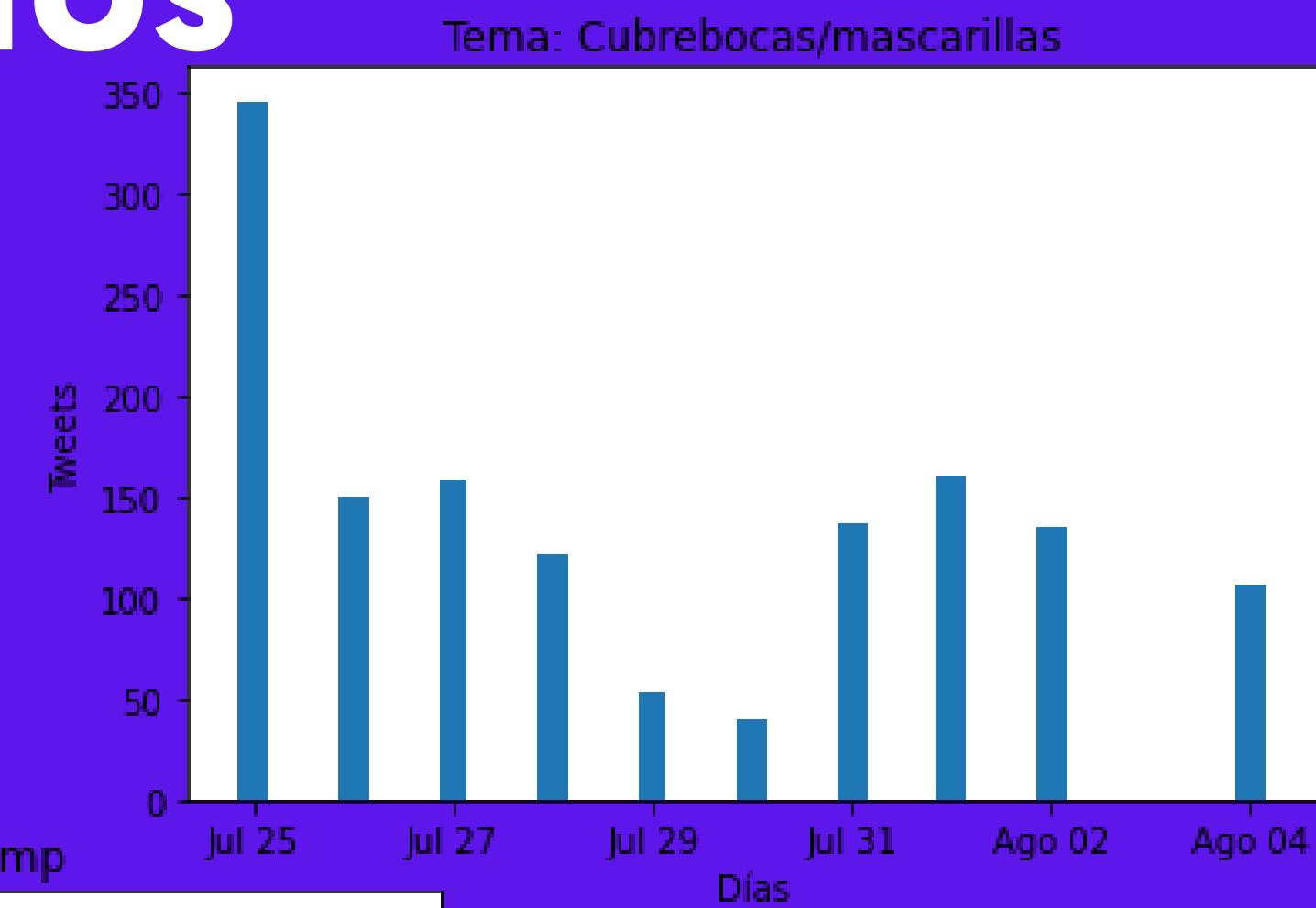
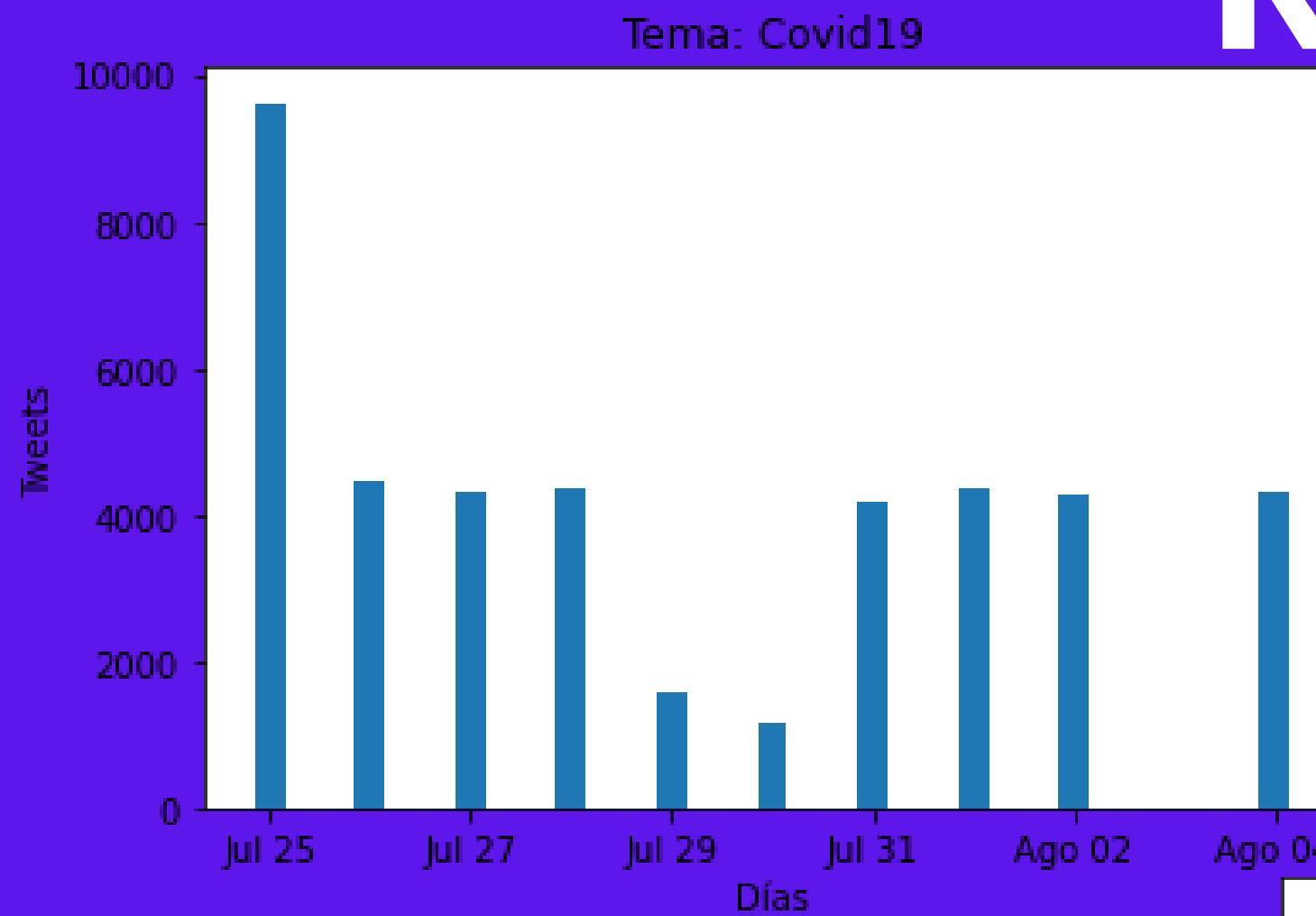
covid = [len(dataframecovid_25jul), len(dataframecovid_26jul), len(dataframecovid_27jul), len(dataframecovid_28jul), 1
print(covid)

# Hacemos un histograma de los valores obtenidos

tiempo = np.arange(len(covid))
esp = 0.3

fix, ax = plt.subplots()
ax.bar(tiempo, covid, esp)
ax.set_title("Tema: Covid19")
ax.set_xlabel("Días")
ax.set_ylabel("Tweets")
ax.set_xticklabels(dias)
```

# Resultados



# Objetivo SMART

## Especifico

SE BUSCA SABER LAS  
CARECTERISTICAS EN COMÚN  
QUE TIENEN LOS USUARIOS  
CON MÁS DE 500,000  
SEGUIDORES

## Accesible

CON LA BASE DE DATOS QUE  
TENEMOS Y DIFERENTES  
LIBRERIAS DE PYTHON,  
PODREMOS ESTUDIAR EL  
COMPORTAMIENTO DE LOS  
FOLLOWERS.

## Medible

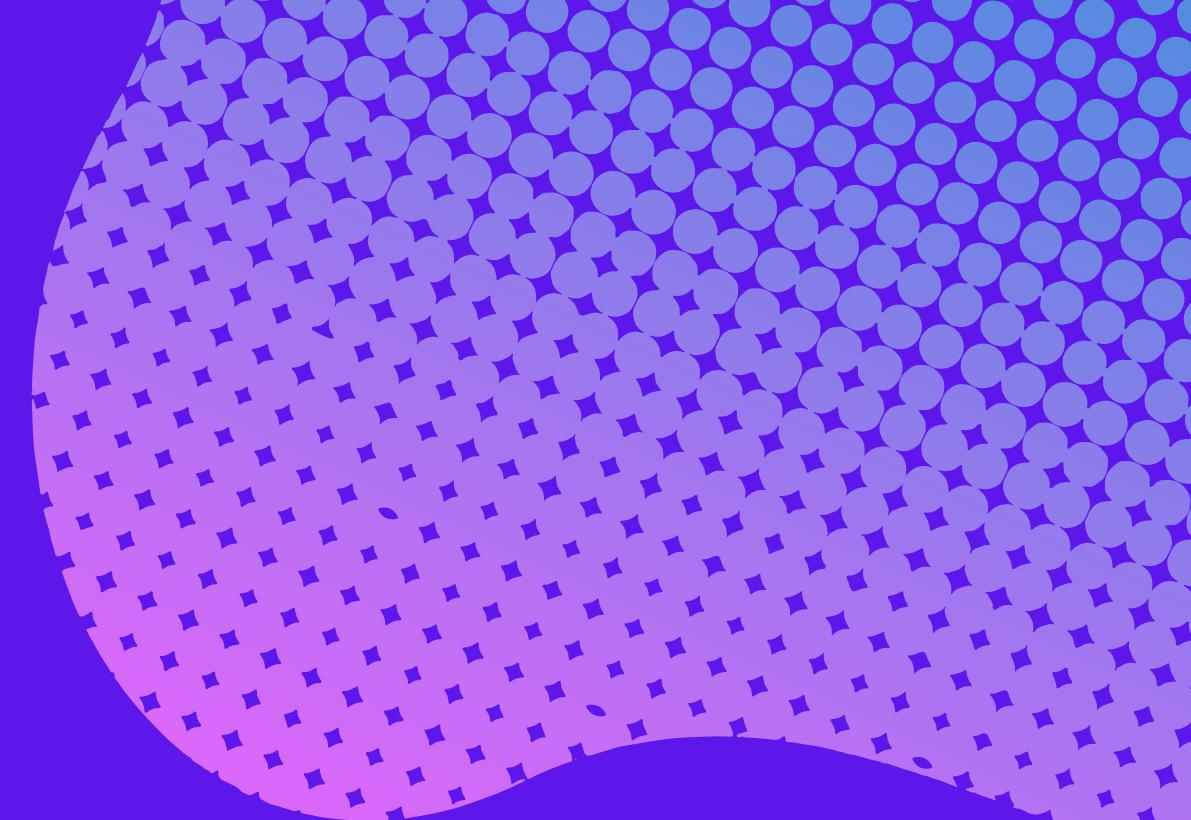
SE TIENE UNA BASE DE DATOS,  
A TRAVES DE ELLA PODEMOS  
ESTUDIAR DIFERENTES  
CAMPOS, COMO LO SON:  
SEGUIDORES, TWEETS,  
UBICACIÓN, ETC.

## Resultados

SI OBSERVAMOS QUE HAY DISTINTAS COSAS QUE  
CADA FAMOS TIENE EN COMÚN, PODEMOS  
CONLUIR QUE SI HAY CARACTERÍSTICAS QUE SE  
COMPARTEN

## Tiempo

DURACIÓN DE 8 HORAS PARA  
REALIZAR ESTE ESTUDIO. AS  
PUBLICACIONES



# ¿Qué es un famoso?

## Objetivo

Se busca saber las características en común que tienen los usuarios con más de 500,000 seguidores durante la pandemia

```
#Para eso debemos de filtrar la base de datos y trabajar solo con los que presenten más de 500,000 seguidores
df_nfamosos=df_famosos= df[df["user_followers"]<500000]
df_famosos= df[df["user_followers"]>500000]
df_famosos
```

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	hashtags
25	Blood Donors India	Mumbai, India	Focused on matching blood donors with those in...	2008-12-23 07:55:39	1215920	2047	19359	True	2020-07-25 12:26:26	#Kolar\nNeed #Blood Type : B-positive\nAt : J...	['Kolar', 'Blood Type', 'B-positive', 'At : J...', 'COVID19']
77	IMF	Washington, DC	Breaking news and alerts from the Internationa...	2009-03-18 16:13:51	1695520	444	182	True	2020-07-25 12:25:02	What economic measures are governments taking ...	['COVID19']
78	Oxfam International	NaN	Oxfam is a world-wide development organization...	2007-12-31 18:27:35	843670	1172	23652	True	2020-07-25 12:25:00	"#Covid19 is everyone's fight. I have covered ...	['Covid19']
87	New Straits Times	Malaysia	News, views and up-to-date reports from Malays...	2009-07-09 09:04:01	716342	411	874	True	2020-07-25 12:24:38	#NSTworld There will not be a #NationalDayRall...	['NSTworld', 'NationalDayRall', 'NDP']
120	Livemint	India	Breaking news and analyses of Indian and world...	2008-11-27 09:07:38	1902888	117	472	True	2020-07-25 12:23:44	Bihar witnesses biggest single-day spike of 2,...	['Covid19']
...	...	...	...	...	...	...	...	...	...	...	...
178823	Cognizant	Global (HQ: Teaneck, N.J.)	Cognizant (Nasdaq: CTSH) is dedicated to help...	2009-08-28 20:31:48	761451	2932	3543	True	2020-08-29 20:00:15	The remote work force has grown 44% — our 4-po...	['COVID19']
178845	AMA	Chicago/Washington D.C.	American Medical Association:	2009-03-31 17:50:31	734929	6915	3209	True	2020-08-29	@JournalofEthics speaks with members of an	['JournalofEthics']

```
[ ] df_famosos.describe().transpose()
```

	count	mean	std	min	25%	50%	75%	max	NaN
user_followers	5196.0	3.379612e+06	3.651779e+06	500070.0	761451.00	1878954.0	5007492.0	49442559.0	
user_friends	5196.0	3.190647e+03	2.822635e+04	0.0	126.00	335.0	1070.0	497363.0	
user_favourites	5196.0	4.185503e+03	1.070050e+04	0.0	143.75	874.0	2960.5	145938.0	

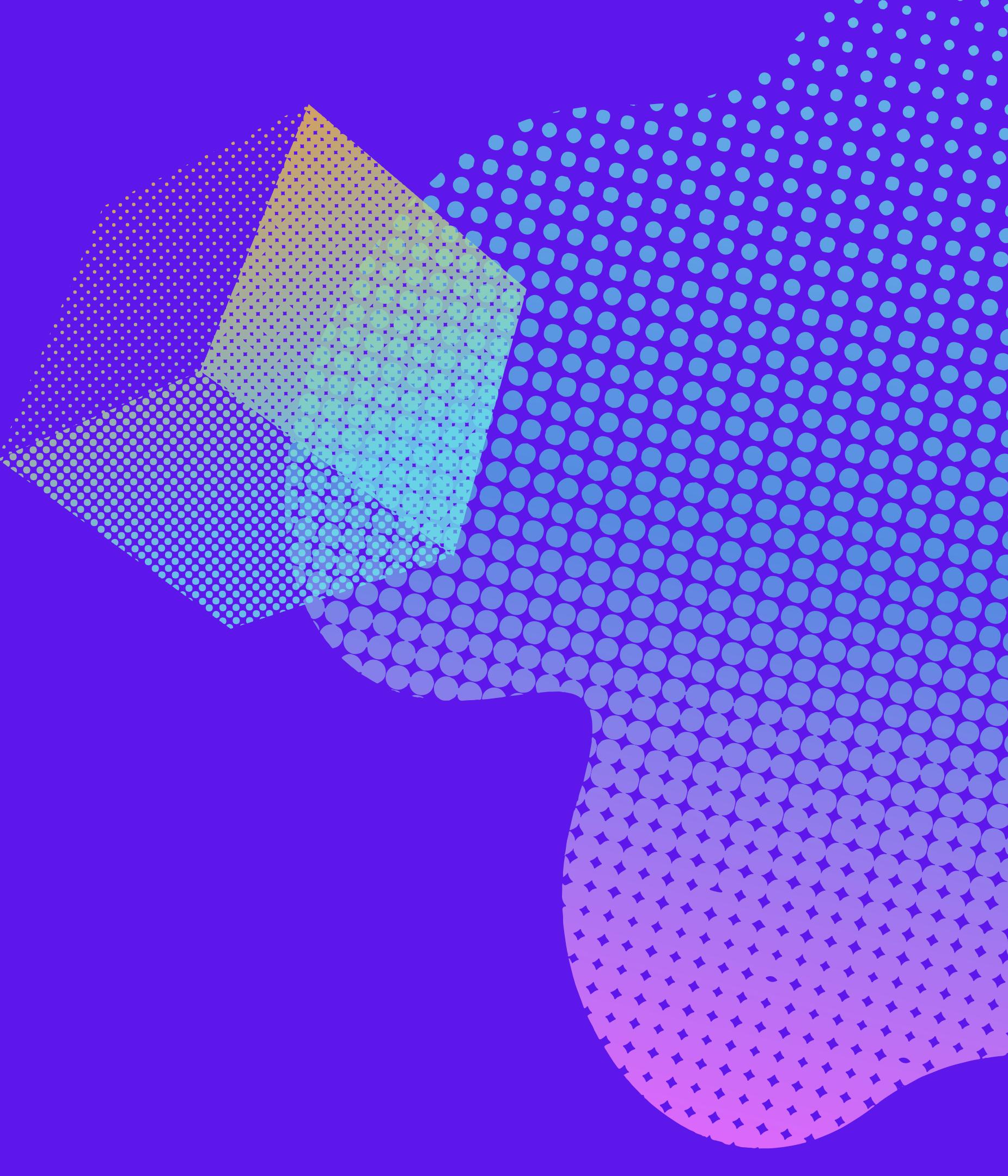
# ¿Preguntas a responder?

¿Cuántos tweets hace en promedio una persona famosa?

¿Están verificados?

¿Qué relación existe entre los seguidores y a cuantos sigues?

# Tweets en promedio



# Famosos

```
▶ tweets_per_username = df_famosos.groupby("user_name")
  ntweets_per_username = df_nfamosos.groupby("user_name")

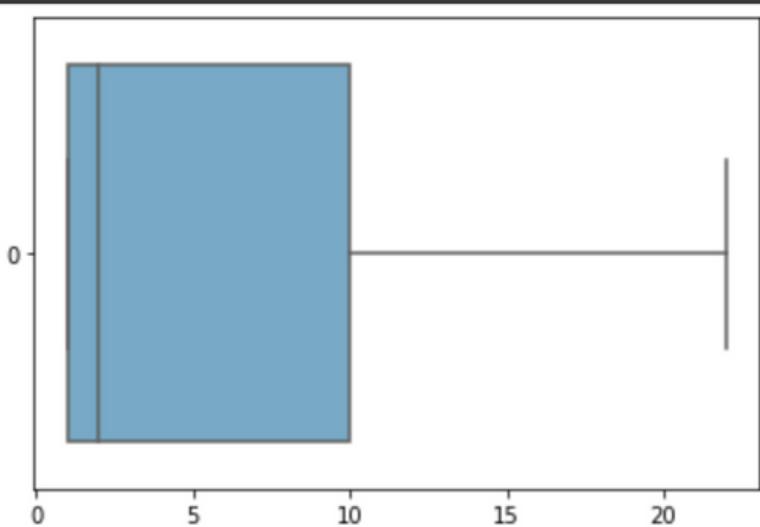
  #Promedio de tweets por persona
  df_famososbxplt=tweets_per_username["text"].count()

  df_nfamososbxplt=ntweets_per_username["text"].count()

  df_famososbxplt.mean()
```

```
↳ 14.761363636363637
```

```
▶ sns.boxplot(data=df_famososbxplt, orient="h", palette='Blues', showfliers=False)
  plt.show()
```



```
[ ] df_famososbxplt.median()
```

```
2.0
```

```
▶ print(df_famososbxplt.max())
  print(df_famososbxplt.min())
```

```
↳ 282
  1
  282
  2.0
```

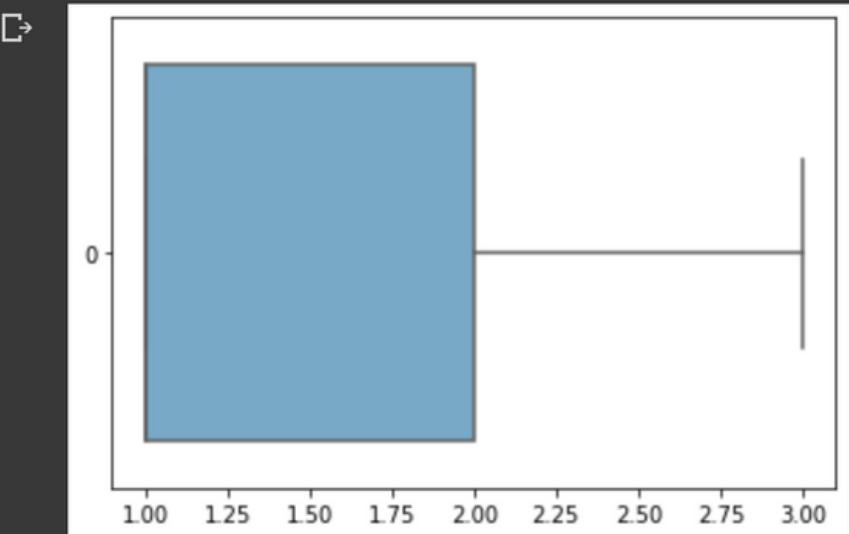
# No Famosos

Ahora veamos a los no famosos

```
[ ] print(df_nfamososbxplt.median())
print(df_nfamososbxplt.max())
print(df_nfamososbxplt.min())
print(df_nfamososbxplt.mean())
```

```
1.0
679
1
1.891725495741464
```

```
▶ sns.boxplot(data=df_nfamososbxplt, orient="h", palette='Blues', showfliers=False)
plt.show()
```



Aquí podemos apreciar como los usuarios no famosos varian, hay la mediana es de un tweet y su maxima alcanza los 679 twets. Aquí podemos encontrar tanto gente muy activa como muy poco activa. Aún así la media por usuario es de 1 tweet, es decir, son muy poco activos en su mayoría

# verificados



# Famosos



```
df_famosos.groupby("user_verified").count()
```



	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	date	text
user_verified									
False	59	55	59	59	59	59	59	59	59
True	5137	4813	5135	5137	5137	5137	5137	5137	5137



Podemos ver que en efecto estos usuarios, en su mayoría tienen la verificación. Esto se sospechaba desde un inicio, pero con lo que descubrimos previamente sobre la cantidad de tweets, teníamos que investigar.

# No Famosos

descubrimos previamente sobre la cantidad de tweets, tenemos que investigar.

```
▶ df_nfamosos.groupby("user_verified").count()
```

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	date	text
user_verified									
False	155954	121091		145679	155954	155954	155954	155954	155954
True	17958	16378		17949	17958	17958	17958	17958	17958



En cuanto a los no famosos, podemos observar que en su mayoría no estan verificados. Casi 18,000 de los casi 170,000 tweets fueron hechos por alguien no verificado.

Likes =  
seguidores  
?



# Famosos

```
print(df_famosos.iloc[:,[4,5]].corr())
```

	user_followers	user_friends
user_followers	1.0	-0.068686
user_friends	-0.068686	1.000000

```
sns.heatmap(df_famosos.iloc[:,[4,5]].corr(), annot=True)
plt.show()
```

The heatmap displays a 2x2 correlation matrix. The diagonal elements are 1.0. The off-diagonal elements are -0.068686. A color scale on the right indicates values from -1.0 (dark blue) to 1.0 (dark red).

Podemos ver que no hay ninguna similitud entre estos dos. Es decir no hay nada que parezca indicar que uno afecta a la otra.

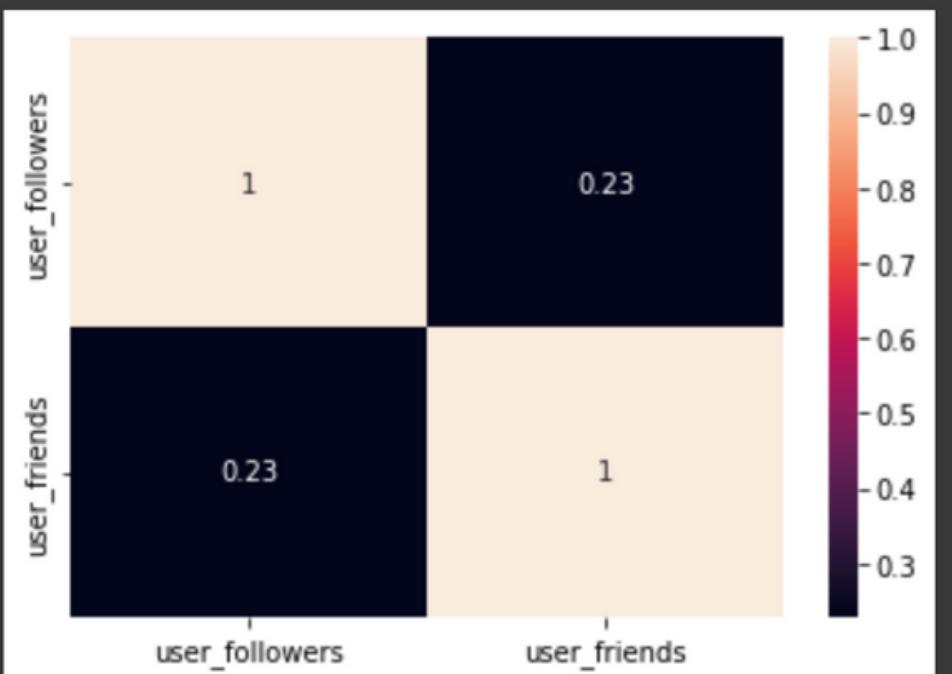
Hasta el momento no hemos visto nada que concluya que características tiene alguien famoso.

# No Famosos

```
[ ] print(df_nfamosos.iloc[:,[4,5]].corr())
```

	user_followers	user_friends
user_followers	1.000000	0.228873
user_friends	0.228873	1.000000

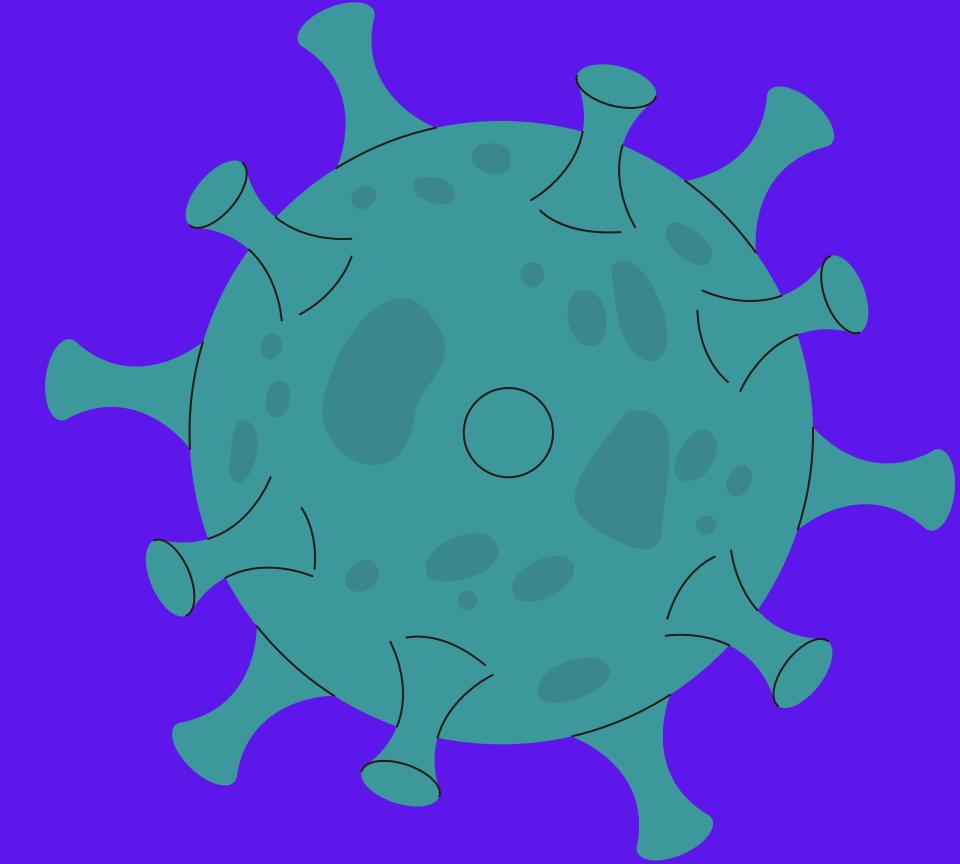
```
[ ] sns.heatmap(df_nfamosos.iloc[:,[4,5]].corr(), annot=True)  
plt.show()
```



Este si parece tener un poco más de correlación, es decir que en este rango si afecta un poco más la actividad que tengas en la aplicación.

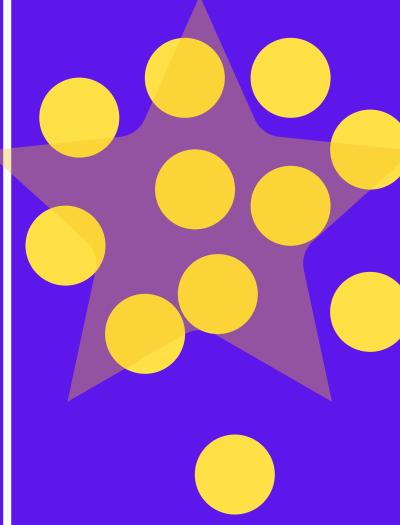
# Objetivo

**Identificar los hashtags más utilizados en las publicaciones relacionadas con COVID-19 en ciertas regiones geográficas a lo largo del tiempo y evaluar su impacto en el nivel de discusión sobre la pandemia.**

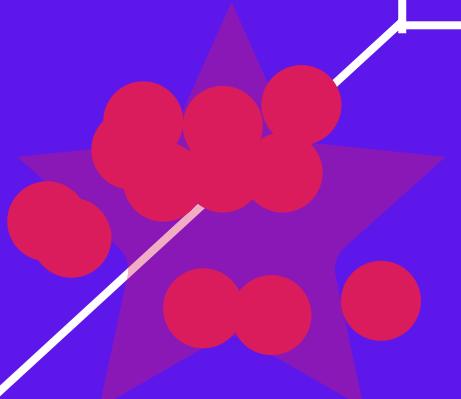
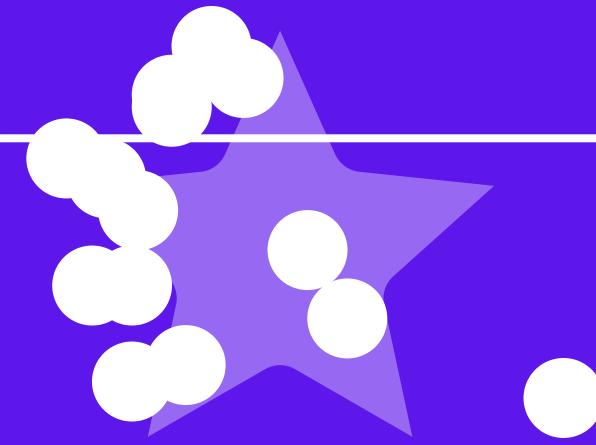


# Usando K-MEANS

HASHTAGS



Lugares

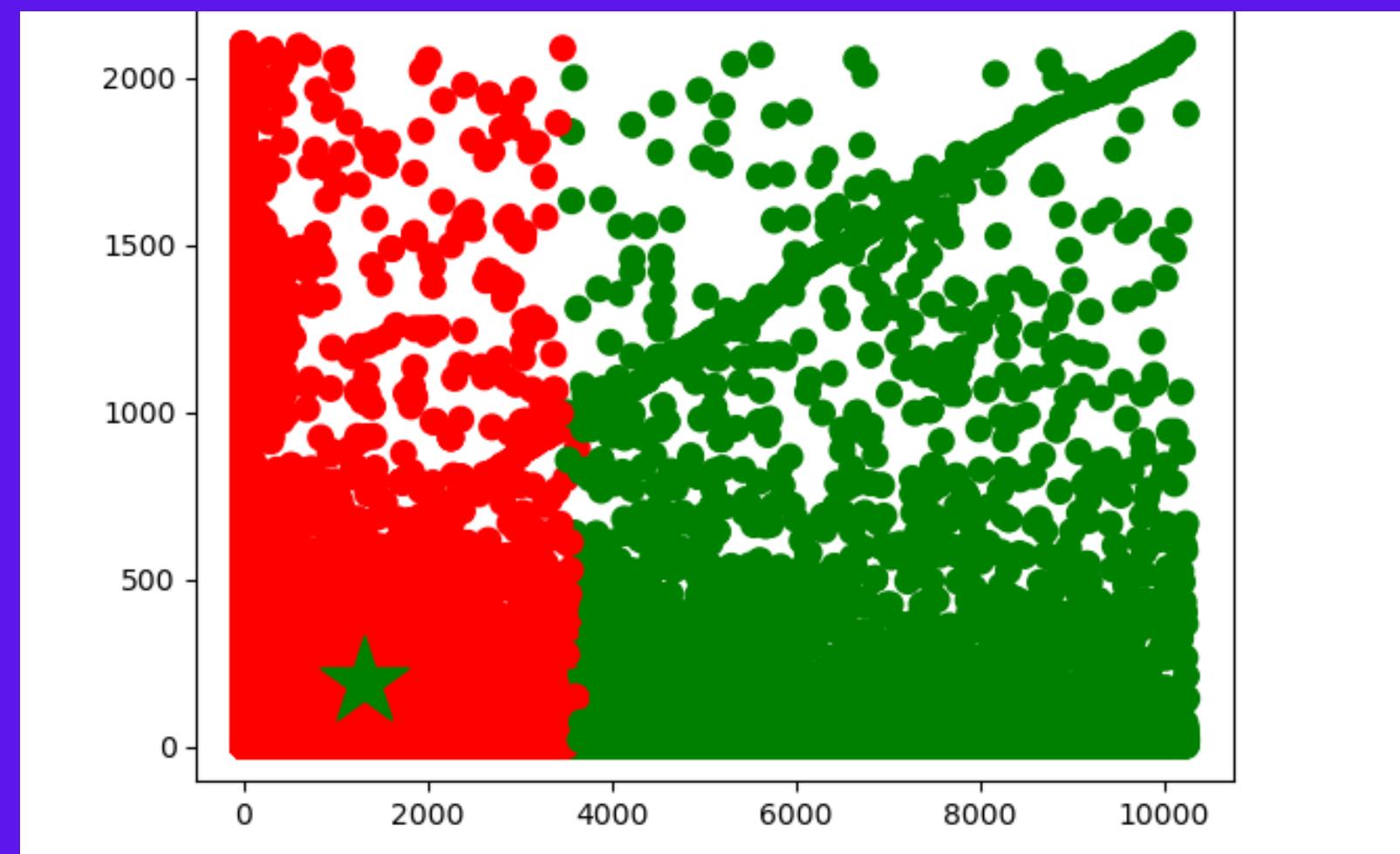
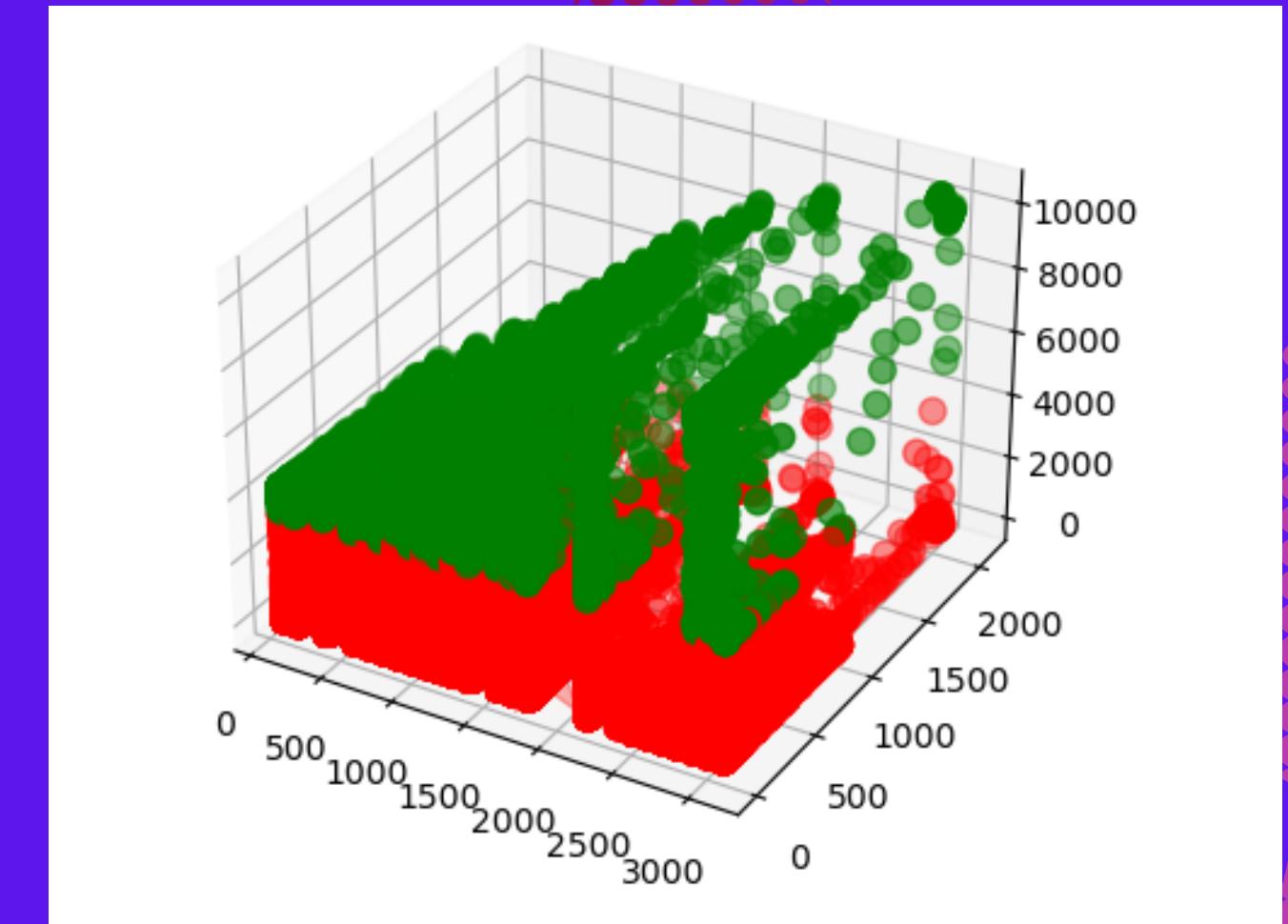
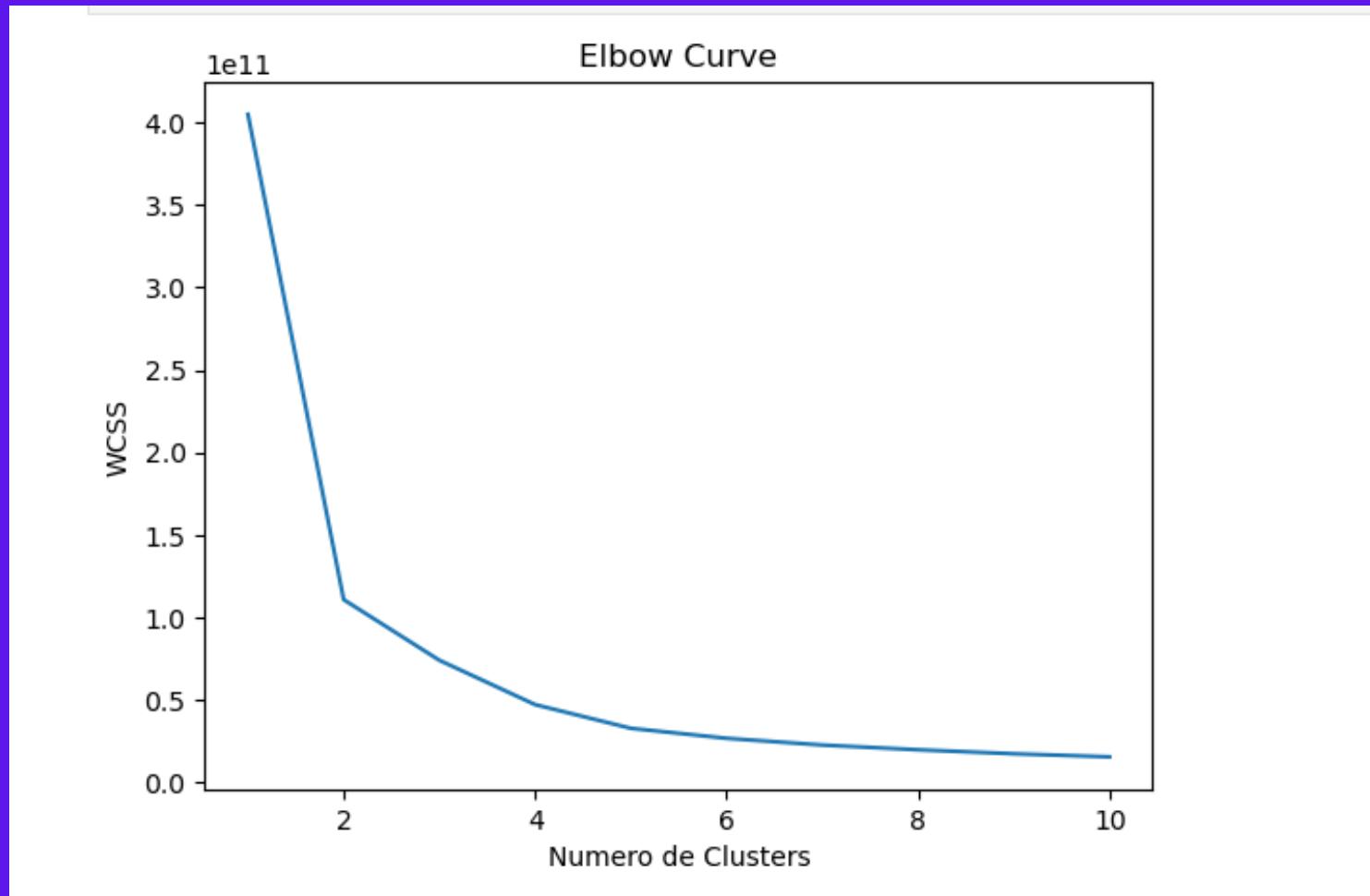


FECHAS DE 2020 DESDE 20/07 - 29/08

# A TENER EN CUENTA

	A	B	C	D	E	F
1	user_location	date_nogut	date_gut	date	hashtags	
2	astroworld	25/07	2507	7/25/2020 12:27		
3	New York, NY	25/07	2507	7/25/2020 12:27		
4	Pewee Valley, KY	25/07	2507	7/25/2020 12:27	['COVID19']	
5	Stuck in the Middle	25/07	2507	7/25/2020 12:27	['COVID19']	
6	Jammu and Kashmir	25/07	2507	7/25/2020 12:27	['CoronaVirusUpdates', 'CO	
7	ĐĐ¾Đ¾Ñ€Đ¾ìÑÑĐ,Ñ	25/07	2507	7/25/2020 12:27	['coronavirus', 'covid19']	
8	Gainesville, FL	25/07	2507	7/25/2020 12:27	['COVID19', 'Recruiting']	
9		25/07	2507	7/25/2020 12:27		
10		25/07	2507	7/25/2020 12:26	['covid19', 'covidPositive']	
11	ðÝ'±ðÝ»location at link belowðÝ'±ðÝ»	25/07	2507	7/25/2020 12:26	['HurricaneHanna', 'COVID1	
12		25/07	2507	7/25/2020 12:26		
13	Dhaka,Bangladesh	25/07	2507	7/25/2020 12:26	['logo', 'graphicdesigner', 'lo	
14	Hotel living - various cities! Who needs a ho	25/07	2507	7/25/2020 12:26	['COVID19']	
15	Africa	25/07	2507	7/25/2020 12:26	['COVID19']	
16	New Delhi	25/07	2507	7/25/2020 12:26		
17	Nagaland, India	25/07	2507	7/25/2020 12:26	['Covid19', 'keepsocialdistanc	
18		25/07	2507	7/25/2020 12:26	['COVID19', 'TamilNadu', 'ch	
19	Brussels	25/07	2507	7/25/2020 12:26	['COVID19', 'homework']	
20		25/07	2507	7/25/2020 12:26	['light']	

	A	B	C
1	country	date	hashtags
2	KY		2507 COVID19
3	FL		2507 COVID19
4	Bangladesh		2507 logo
5	India		2507 Covid19
6	India		2507 Kolar
7	NY		2507 COVID19
8	India		2507 Covid19
9	FL		2507 covid19
10	Canada		2507 COVID19
11	UK		2507 BorisJohnson
12	Maryland		2507 COVID19
13	Ireland		2507 PlayYourPart
14	South Africa		2507 COVID19
15	India		2507 Delhi
16	NY		2507 COVID19
17	LA		2507 covid19
18	GA		2507 COVID19
19	United Kingdom		2507 obesity
20	NY		2507 COVID19
21	CO		2507 covid19



The background features a solid purple color with abstract graphic elements. In the top left corner, there is a large, tilted hexagon filled with a blue halftone dot pattern. In the bottom right corner, there is a large, irregular shape filled with a blue halftone dot pattern.

Gracias