

# Evaluable 2

Alejandro Hernández De la Torre

In [19]:

```
import pandas as pd
import matplotlib.pyplot as plt

import seaborn as sns; sns.set_theme()
df = pd.read_csv("covid19_tweets.csv")
print (df)
```

	user_name	user_location	
0	Violence	astroworld	
1	Tom Basile US	New York, NY	
2	Time4fisticuffs	Pewee Valley, KY	
3	ethel mertz	Stuck in the Middle	
4	DIPR-J&K	Jammu and Kashmir	
...	...	...	
74431	Laura Wolfrom	Lexington, KY	
74432	Professor Tonya M. Evans	🏠 #stayathome	
74433	People's Daily app	北京, 中华人民共和国	
74434	M0ser	NaN	
74435	Your Friend & Sabre ✂	Chicago, IL	

	user_description	user_created	
0	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	
1	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	
2	#Christian #Catholic #Conservative #Reagan #Re...	2009-02-28 18:57:41	
3	#Browns #Indians #ClevelandProud #[]_[] #Cavs ...	2019-03-07 01:45:06	
4	🖋 Official Twitter handle of Department of Inf...	2017-02-12 06:45:15	
...	...	...	
74431	The only things I collect are memories.	2010-09-24 02:01:15	
74432	Law Prof @DickinsonLaw & Entrepreneur   Crypto...	2013-05-14 20:15:24	
74433	Our mission is to provide news and perspective...	2018-02-04 12:36:42	
74434	Reagan conservative and attorney raised in the...	2014-02-18 03:46:28	
74435	My spectral decomposition has a significant da...	2016-12-19 19:55:00	

	user_followers	user_friends	user_favourites	user_verified	
0	624	950	18775	False	
1	2253	1677	24	True	
2	9275	9525	7254	False	
3	197	987	1488	False	
4	101009	168	101	False	
...	...	...	...	...	
74431	85	586	1902	False	
74432	4289	1066	53569	False	
74433	1413	102	16	False	
74434	2554	1733	129104	False	
74435	310	1748	60133	False	

	date	text	
0	2020-07-25 12:27:21	If I smelled the scent of hand sanitizers toda...	
1	2020-07-25 12:27:17	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	
2	2020-07-25 12:27:14	@diane3443 @wdunlap @realDonaldTrump Trump nev...	
3	2020-07-25 12:27:10	@brookbanktv The one gift #COVID19 has give me...	
4	2020-07-25 12:27:08	25 July : Media Bulletin on Novel #CoronaVirus...	
...	...	...	
74431	2020-08-04 03:13:29	So far this summer I have filled up my lawn mo...	
74432	2020-08-04 03:13:26	ICYMI: REPLAY: #TechIntersect 🌀 #16: Isaiah "@B...	
74433	2020-08-04 03:13:22	Community workers in Tianshan District of Urum...	

```

74434 2020-08-04 03:13:19 If only we had a responsible media to warn us ...
74435 2020-08-04 03:13:15 MAGA: #COVID19 is just a cold & it'd be go...

```

```

                                hashtags                                source \
0                                NaN    Twitter for iPhone
1                                NaN    Twitter for Android
2                                ['COVID19']    Twitter for Android
3                                ['COVID19']    Twitter for iPhone
4                                ['CoronaVirusUpdates', 'COVID19']    Twitter for Android
...
74431                                ['COVID19', 'QuarantineLife']    Twitter for iPhone
74432    ['TechIntersect', 'Bitcoin', 'COVID19']    Twitter Web App
74433                                ['China', 'Xinjiang']    Twitter Web App
74434                                ['COVID19']    Twitter for iPhone
74435                                ['COVID19', 'Hydroxychloroquine']    Twitter for Android

                                is_retweet
0                                False
1                                False
2                                False
3                                False
4                                False
...
74431                                False
74432                                False
74433                                False
74434                                False
74435                                False

```

```
[74436 rows x 13 columns]
```

```
In [20]: df.describe()
```

```

Out[20]:

```

	user_followers	user_friends	user_favourites
count	7.443600e+04	74436.000000	7.443600e+04
mean	1.059513e+05	2154.721170	1.529747e+04
std	8.222900e+05	9365.587474	4.668971e+04
min	0.000000e+00	0.000000	0.000000e+00
25%	1.660000e+02	153.000000	2.200000e+02
50%	9.600000e+02	552.000000	1.927000e+03
75%	5.148000e+03	1780.250000	1.014800e+04
max	1.389284e+07	497363.000000	2.047197e+06

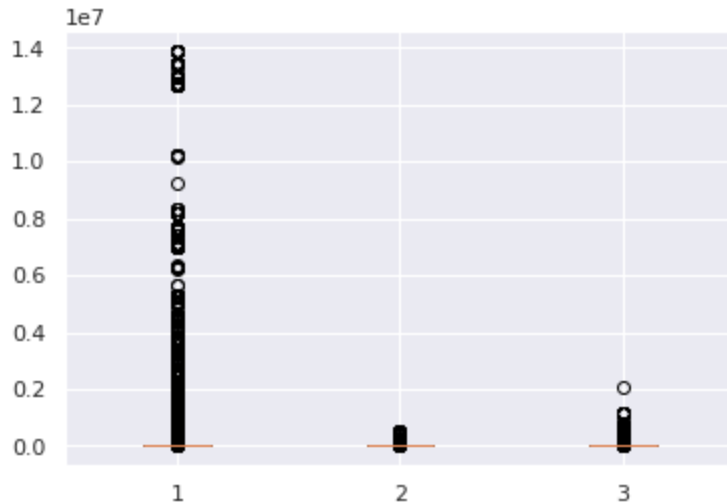
## Interpretación Describe

Los siguientes datos nos marcan las estadísticas de las personas que tweetearon acerca del COVID-19. En periodos de tiempo consisos. Decidimos enfocarnos en los seguidores, usuarios y favoritos. Por ejemplo con la función "describe" nos dimos cuenta de varias cosas interesantes, por ejemplo con la función count nos dimos cuenta que los user\_followers es casi igual a los user\_favorites. Es decir que ambas cifras son parecidas en cantidad. Por lo que se podría decir que los usuarios ganan followers si otros le da un "favorite" o estan de acuerdo con lo que piensan o dicen. Pero cuando observamos la estandarización podemos ver que los followers se desvian muy fuerte en comparación de los favoritos, haciendo que un fenomeno de fluctuación suceda en comparación al punto anterior,esta fluctuación tambien la podemos ver en los cuartiles los cuales estan

indicados con numeros y porcentajes. Lo interesante se puede ver en la maxima y minima. Ya que hay usuarios que no tienen seguidores, ni amigos ni favoritos. Por lo cual podriamos decir que hubo bots envueltos en esta toma de datos, pero en la maxima vemos numeros muy grandes lo cual tambien puede significar que alguna celebridad tambien fue parte de este rastreo.

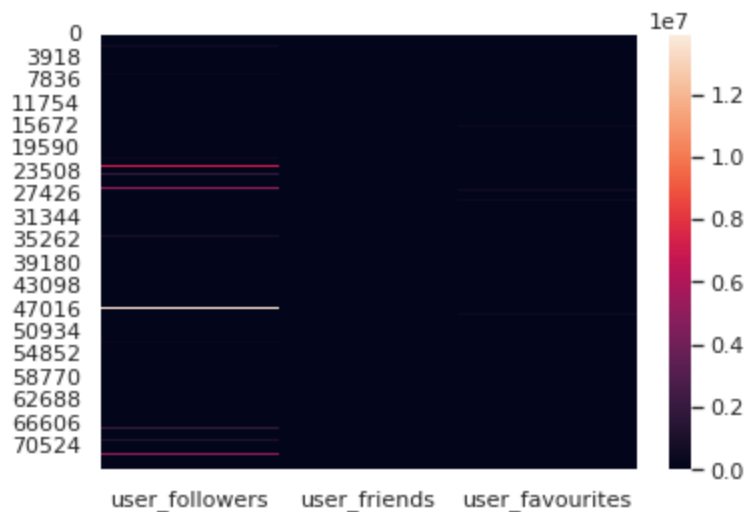
In [21]:

```
plt.boxplot(df[["user_followers","user_friends","user_favourites"]])  
  
plt.show()
```



In [24]:

```
ax = sns.heatmap(df[["user_followers","user_friends","user_favourites"]])
```



## Interpretación Bigote y Heatmap

En estos mapas podemos ver que tan cegados estan los datos y como estan las medias de los datos. En estas graficas no se alcanza a ver los numeros ni las relaciones correctamente, por lo cual podemos decir que los datos simplemente son muy diferentes los unos a los otros, tanto que ningún heatmap o caja de bigote en los incisos de "user\_followers","user\_friends","user\_favourites"; puede encontrar una relación específica con los datos y sus medias tambien por eso en la función describe pudimos ver cegos gigantes comparando "mean" con los cuartiles.

In [25]:

```
df[["user_followers","user_friends","user_favourites","user_verified"]].corr(method='kendalltau')
```

Out[25]:

	<b>user_followers</b>	<b>user_friends</b>	<b>user_favourites</b>	<b>user_verified</b>
<b>user_followers</b>	1.000000	0.410663	0.242534	0.415364
<b>user_friends</b>	0.410663	1.000000	0.439099	-0.003114
<b>user_favourites</b>	0.242534	0.439099	1.000000	-0.032220
<b>user_verified</b>	0.415364	-0.003114	-0.032220	1.000000

## Interpretación Correlación

En estos datos podemos ver que tanto tienen que ver unos datos con otros, por eso en las columnas y filas que tienen nombre similar se puede ver un 1, ya que son la misma por lo tanto tienen que ver mucho una con otra por que son la misma. Pero lo interesante ocurre en los demas rubros por ejemplo en user\_followers, se puede ver que tiene un .41 con los amigos, podiamos intuir que un poco menos de la mitad de los followers tienden a ser amigos de los usuarios y la mitad de esos son favoritos, pero menos de la mitad de los usuarios tienden a seguir a cuentas verificadas. Tambien podemos ver que las cuentas verificadas tienen muy poco amigos verificados, lo cual es logico, ya que no existen tantas cuentas verificadas.