Actividad Evaluable: Obtención de estadísticas descriptivas
Nombre: María Clarita Osorio Vergara A01654530

1. Carga los datos usando tu lector de csv o con pandas. Es recomendable hacerlo con pandas.

In [15]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np

data=pd.read_csv("semanaTec_Analitica/arte-de-analitica/covid19_tweets.csv")
data.head()
```

Out[15]:

	user_name	user_location	user_description	user_created	user_followers	user_friends	user_favourites	user_verified	date	text	
0	‘V’i☻ı¢†	astroworld	wednesday addams as a disney princess keepin i...	2017-05-26 05:46:42	624	950	18775	False	2020-07-25 12:27:21	If I smelled the scent of hand sanitizers toda...	
1	Tom Basile 🇺🇸	New York, NY	Husband, Father, Columnist & Commentator. Auth...	2009-04-16 20:06:23	2253	1677	24	True	2020-07-25 12:27:17	Hey @Yankees @YankeesPR and @MLB - wouldn't it...	
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 nev...	
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...	
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...	['CoronaVi

2. Verifica la cantidad de datos que tienes, las variables que contiene cada vector de datos e identifica el tipo de variables.

In [16]:
```
#Cantidad de usarios
print("Cantidad de usuarios:")
print(len(data.index))
#Variables
print("Variables:")
print(data.columns.values)
#Tipo de variables
print("Tipo de variables:")
data.dtypes
```

```
Cantidad de usuarios:
74436
Variables:
['user_name' 'user_location' 'user_description' 'user_created'
 'user_followers' 'user_friends' 'user_favourites' 'user_verified' 'date'
 'text' 'hashtags' 'source' 'is_retweet']
Tipo de variables:
```

Out[16]:
```
user_name           object
user_location       object
user_description    object
user_created        object
user_followers       int64
user_friends         int64
user_favourites      int64
user_verified         bool
date                object
text                object
hashtags            object
source              object
is_retweet            bool
dtype: object
```

3. Analiza las variables para saber qué representa cada una y en qué rangos se encuentran. Si la descripción del problema no te lo indica, utiliza el máximo y el mínimo para encontrarlo.

```
data_num.min()
```

```
user_followers      0
user_friends        0
user_favourites     0
dtype: int64
```

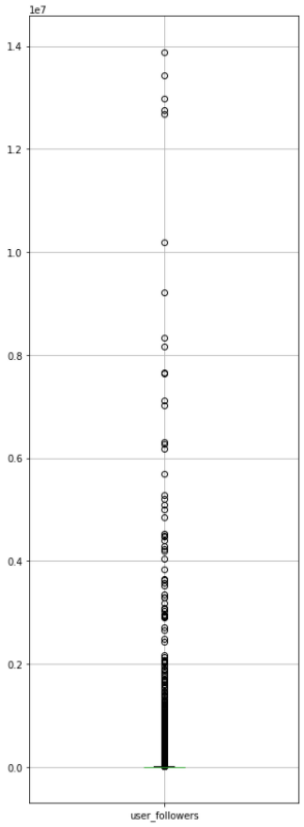
```
data_num.max()
```

```
user_followers      13892841
user_friends          497363
user_favourites      2047197
dtype: int64
```

In [76]:
```
d1=data.groupby(['user_name']).mean().sort_values(['user_followers'], ascending = False).groupby("user_name").head(10)
dU = d1[['user_followers']]
dU.head()
```

Out[76]:

user_name	user_followers
CGTN	1.389003e+07
NDTV	1.343905e+07
The Times Of India	1.298272e+07
United Nations	1.275416e+07
China Xinhua News	1.268052e+07



4. Basándose en la media, mediana y desviación estándar de cada variable, ¿Qué conclusiones puedes entregar de los datos?

```
data_num= data[["user_followers","user_friends","user_favourites"]]
data_num.head()
```

	user_followers	user_friends	user_favourites
0	624	950	18775
1	2253	1677	24
2	9275	9525	7254
3	197	987	1488
4	101009	168	101

```
data_num.describe()
```

	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

Conclusiones:

Después de visualizar la gráfica de la sección 3, me puedo percatar que la mayoría de los usuarios tiene pocos seguidores, por lo que infiero que estas cuentas pueden ser bots para generar polémicas relacionadas con los políticos, o bien, simplemente pueden ser personas con una opinión. Asimismo, los usuarios con muchos seguidores, solían publicar varios tweets con respecto al covid 19, esto me hace pensar que son noticieros o reporteros, cuyo propósito es informar a la población acerca de las actualizaciones de la situación que se vivía en ese entonces.

Lo anterior se puede confirmar con la desviación estándar de la variable user_followers, al ser un número sumamente grande, implica que los datos varían mucho, lo que quiere decir que hay cuentas con muy pocos seguidores y usuarios con una gran cantidad de seguidores.

En conclusión, existen varias cuentas con pocos seguidores, así como algunas cuentas con una enorme cantidad de seguidores.