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

```
import seaborn, matplotlib
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import math

# Cargando datos
data = pd.read_csv("covid19_tweets.csv")
```

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

```
# Cantidad de datos que se tienen
datosVariable = data.count()
datosTotales = sum(data.count())

print("\n ---> Cantidad de datos por variable")
print(datosVariable)
print("\n ---> Cantidad de datos que se tienen = ",datosTotales)
```

```
 ---> Cantidad de datos por variable
user_name           74436
user_location       59218
user_description    70079
user_created        74436
user_followers      74436
user_friends        74436
user_favourites     74436
user_verified       74436
date                74436
text                74436
hashtags            53002
source              74424
is_retweet          74436
dtype: int64

 ---> Cantidad de datos que se tienen =  926647
```

```
# Variables que contiene cada vector
def variablesVector(matriz):
    variables = data.columns.values
    for i in range(0, len(variables)):
        print(variables[i])

print(" \n ---> Variables que contiene cada vector")
variablesVector(data)
```

```
 ---> Variables que contiene cada vector
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
tiposVariables = data.dtypes

print("\n --->Tipos de variables")
print(tiposVariables)
```

```
 --->Tipos de variables
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
```

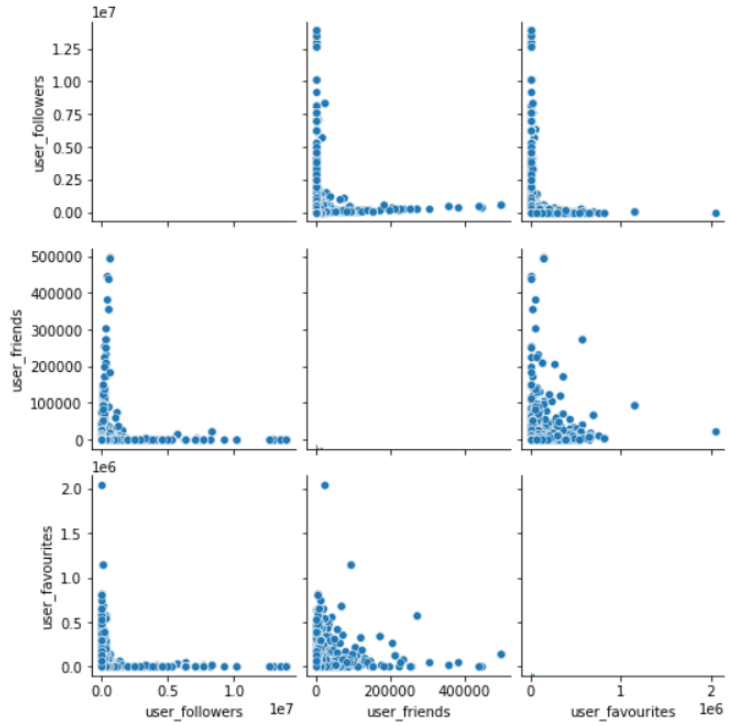
3. Analiza las variables para saber qué representa cada una y en qué rangos se encuentran. Si la descripción del problema no te lo indica, utiliza el máximo y el mínimo para encontrarlo.

Haciendo un análisis de las variables se puede observar que la primera “user_name” representa y brinda información sobre el nombre de usuario que la persona tiene para acceder a la plataforma de Twitter. La segunda variable “user_location” representación el lugar registrado como locación que proporciono el usuario al crear su cuenta, la tercera variable “user_description” hace referencia a la información que muestra el usuario dentro de su presentación de perfil. La cuarta variable “user_created” muestra la fecha en la que la cuenta fue creada. La quinta variable “user_followers” se refiere a la cantidad de cuentas de usuario que sigue la persona, la sexta variable “user_friends” representa el número de amigos que tiene el usuario agregados a su cuenta y la septima variable “user_favourites” muestra la cantidad de Tweets marcados como favoritos por parte del usuario. La variable “user_verified” hace referencia a un valor booleano que da a entener si el perfil es de una cuenta verificada o no. La variable “date” muestra la fecha y hora en la que el Tweet fue publicado, “text” el contenido que tiene dicha publicación, “hashtags” los que uso el usuario dentro de su publicacion y la variable “source” indica el dispositivo desde el cual fue realizado el Tweet y el cual usa el usuario. Por último, la variable “is_retweet” muestra si la publicación realizada viene por parte del mismo usuario o se tomo de otro perfil.

Dado que las únicas variables de tipo de dato numérico son user_followers, user_friends y user_favourites, son las variables a las que se les obtendrá el rango en el que se encuentran. Al igual que de las variables booleanas “user_verified” y “is_retweet” dado que van a mostrar si hay mas cuentas verificadas o no, al igual que nos reRetweet realizados.

```
#Correlacion
new_data = data[['user_followers','user_friends','user_favourites']]
sns.pairplot(new_data)
plt.show()
```

```
# Rangos en los que se encuentran las variables
print("\n ---> Rango de user_followers = de ",data["user_followers"].min(), "hasta",data["user_followers"].max())
print("---> Rango de user_friends = de ",data["user_friends"].min(), "hasta",data["user_friends"].max())
print("---> Rango de user_favourites = de ",data["user_favourites"].min(), "hasta",data["user_favourites"].max())
print("---> Rango de user_verified  = de minimo",data["user_verified"].min(), "hasta maximo",data["user_verified"].max())
print("---> Rango de is_retweet = de minimo",data["is_retweet"].min(), "hasta maximo",data["is_retweet"].max())
```




```
 ---> Rango de user_followers = de  0 hasta 13892841
---> Rango de user_friends = de  0 hasta 497363
---> Rango de user_favourites = de  0 hasta 2047197
---> Rango de user_verified  = de minimo False hasta maximo True
---> Rango de is_retweet = de minimo False hasta maximo False
```

Dada la gráfica anterior, se puede observar que la mayoría de los usuarios tienen pocos amigos pero siguen a muchas personas, esto puede dar a entender que puede que las personas que sigan sean personas famosas que no es normal que regresen el seguimiento. A su vez, existen pocos usuarios que tienen como favoritos pocas publicaciones pero cuentan con muchos amigos, esto puede dar a entender que navegan por el foro de publicaciones sin necesidad de interactuar con sus amigos. Por otro, está la parte viceversa donde hay pocos usuarios que tienen muchos seguidores pero pocos amigos. Sin embargo, en la mayoría de los casos los usuarios tienen menos amigos que seguidores.

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

```
# Obtención de media, mediana y desviación estándar
estadistica = data.describe()
print("\n --->Media, Mediana y Desviación estándar")
print(estadistica)
```

```
 --->Media, Mediana y Desviación estándar
       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
```

Media → `mean  1.059513e+05  2154.721170  1.529747e+04`

Mediana →

50%	9.600000e+02	552.000000	1.927000e+03

Desviación estándar →

std	8.222900e+05	9365.587474	4.668971e+04

Dados los datos anteriores, comenzando por la media y la desviación estándar se puede concluir que los usuarios de Twitter durante el brote de la variante Covid19 tuvieron un aproximado de seguidores en sus cuentas de 105,951, cantidad que varía en un estimado de 822,290 seguidores por cuenta de usuario. A su vez, tenían agregados alrededor de 2,154 amigos, los cuales se alteraban en un aproximado de 552 amigos por cuenta a partir del promedio y por último se tenía señalado dentro de sus favoritos una cantidad de 15,297 tweets los cuales variaron alrededor de 1927 favoritos.

En cuanto a la mediana, se puede determinar que la mitad de los usuarios de Twitter tenía una cantidad de seguidores durante este periodo de pandemia igual o por debajo de los 960 y la otra mitad igual o por encima de esta cantidad. A su vez, una parte de la misma división de usuarios contaba con una cantidad de amigos agregados menor o igual a 522 y la otra parte mayor o igual a dicha cantidad. Por otro lado, en cuanto a Tweets favoritos se puede determinar que la mitad de los usuarios destacó una cantidad igual o menor a 1927 y la otra mitad una cantidad igual o mayor a la mencionada.