

## Actividad Evaluable 2: Obtención de Estadísticas Descriptivas

Análisis del documento *covid19\_tweets* usando la librería *Pandas*

Cantidad de Datos: 926,647

(Usando `shape` encontramos un tamaño de  $74,436 * 13 = 967,668$  y usando la función `isna()` se encontraron 41,021 valores indefinidos dando un total de  $967,668 - 41,021 = 926,647$ )

### **user\_name**

Tipo de Variable: String / cadena de caracteres

Método de Examinación: Imposible

Inferencia: Ninguna

### **user\_location**

Tipo de Variable: String / cadena de caracteres / Ubicación

Método de Examinación: Usando `value_counts()`

Inferencia: Ya que la mayoría de los tweets son de cuenta provenientes de India y EE. UU. inferimos que estos países tienen más interés en el tema, tienen mayor acceso a redes sociales o una combinación de ambas

### **user\_description**

Tipo de Variable: String / cadena de caracteres

Método de Examinación: Imposible

Inferencia: Ninguna

### **user\_created**

Tipo de Variable: String / cadena de caracteres con formato de Fecha

Método de Examinación: Usando `.year` sabemos el año en que fueron creadas

Inferencia: Al observar que las cuentas hayan sido creadas antes del tema podemos asegurarnos de que no son cuentas falsas promoviendo desinformación

### **user\_followers**

Tipo de Variable: Int / Número

Método de Examinación: Usando `.sort_values()`

Inferencia: Podemos, usando la función anterior, observar las cuentas con más followers que han opinado del tema, en otras palabras, las cuentas más influyentes

### **user\_friends**

Tipo de Variable: Int / Número

Método de Examinación: Imposible

Inferencia: Ninguna

### **user\_favourites**

Tipo de Variable: Int / Número

Método de Examinación: Imposible

Inferencia: Ninguna

### **user\_verified**

Tipo de Variable: boolean

Método de Examinación: Usando value\_counts()

Inferencia: Al observar que las cuentas hayan sido podemos asegurarnos de que no son cuentas falsas promoviendo desinformación

### **date**

Tipo de Variable: String / cadena de caracteres con formato de Fecha

Método de Examinación: Usando .sort\_values()

Inferencia: Organizando los tweets podemos inferir las fechas en las que este tema cobro relevancia

### **text**

Tipo de Variable: String / cadena de caracteres

Método de Examinación: Imposible

Inferencia: Ninguna, deberíamos observar lo que dice cada una lo cual es imposible

### **hashtags**

Tipo de Variable: String / cadena de caracteres

Método de Examinación: Imposible

Inferencia: Ninguna

### **source**

Tipo de Variable: String / cadena de caracteres

Método de Examinación: Usando value\_counts()

Inferencia: Ninguna

## **is\_retweet**

Tipo de Variable: String / cadena de caracteres

Método de Examinación: Usando value\_counts()

Inferencia: Ninguna es retweet, todas son opiniones originales