**TWEET ARCHIVER:**

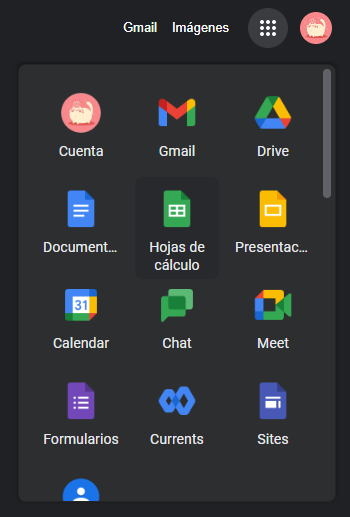
**Recopilación de Datos de la red Social Twitter.**

Luis Felipe Narvaez Gomez. E-mail: luis.narvaez@usantoto.edu.co. Cod: 2312660. Facultad de Ingeniería de Sistemas.

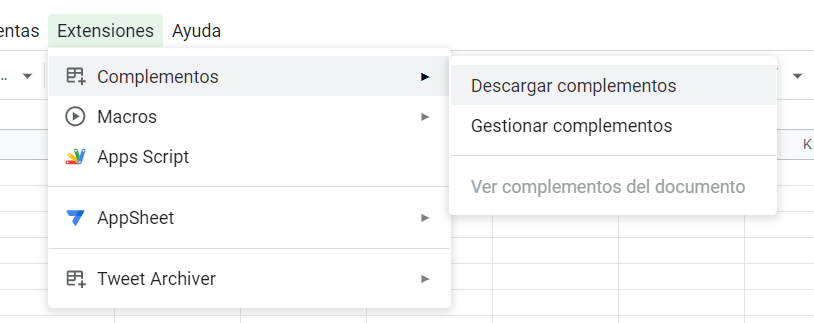
**GUÍA DE USO BÁSICO - TWEET ARCHIVER**

Las hojas de cálculo de las cuales disponemos en Google al crearnos una cuenta de GMAIL poseen varios tipos de exenciones especiales que nos pueden servir para la extracción, recopilación y clasificación de data en bruto de sitios web de interés,, una de estas herramientas es TWEET ARCHIVER la cual nos permite extraer comentarios publicados en la red social de twitter sobre determinados temas específicos, temas en tendencia y hagstags importantes que nosotros necesitemos, a manera de Query que se ira armando según las especificaciones que nosotros le introduzcamos.

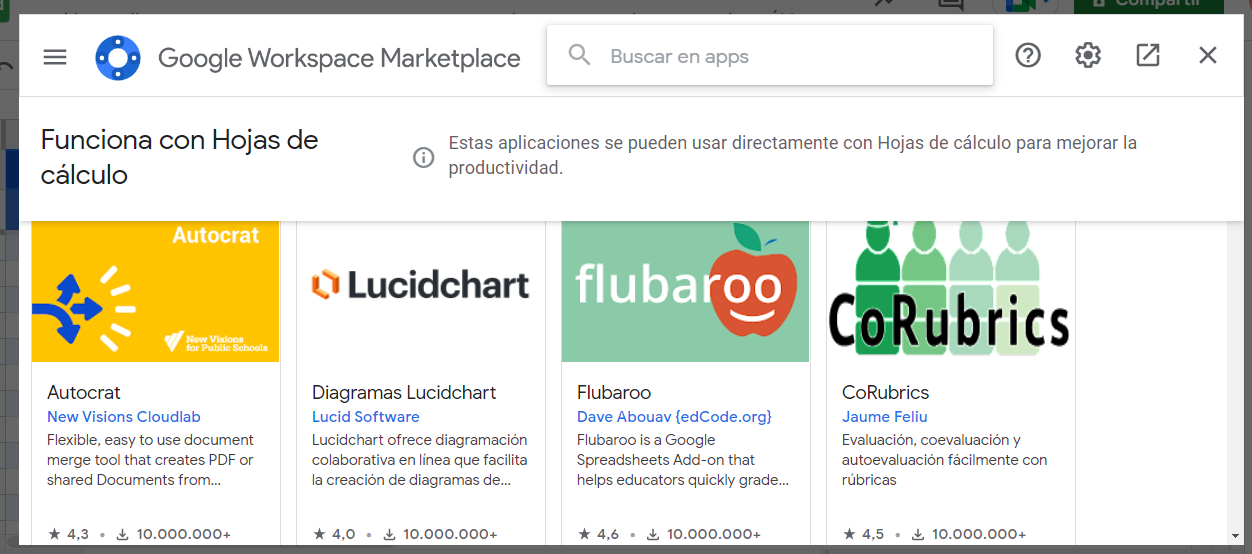
Para acceder a esta herramienta primero debemos loguearnos en GMAIL, donde se recomienda estar en el navegador nativo de Google, Google Chrome. En la parte superior derecha del navegador encontraremos el menú de despliegue de herramientas, en el encontraremos las hojas de cálculo fácilmente localizable por su icono de color verde, este lo seleccionaremos.



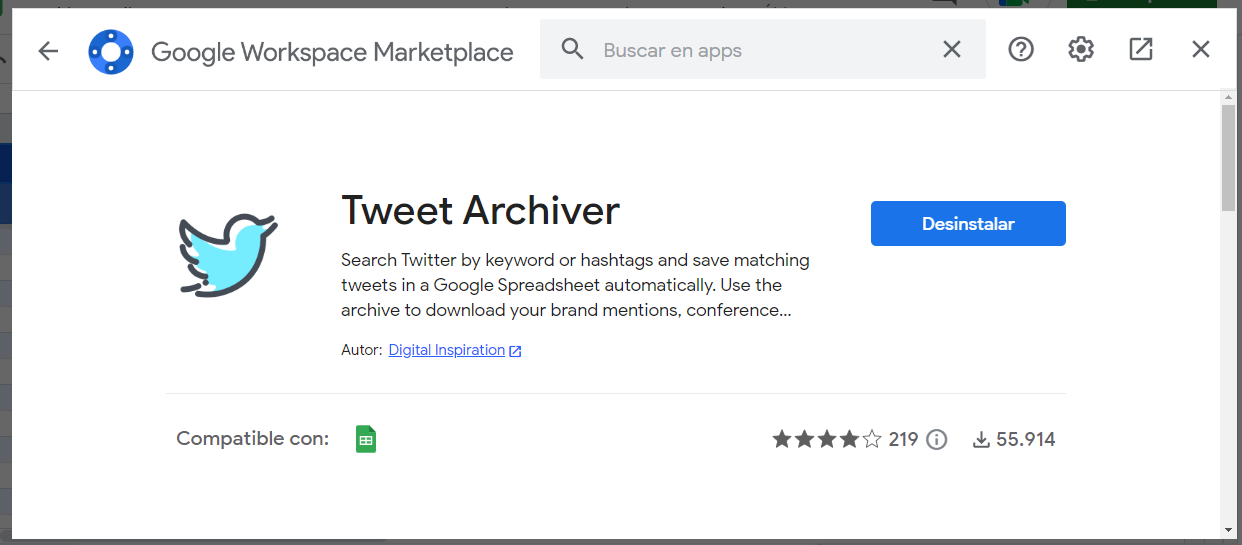
Una vez dentro de la hoja de cálculo nos dirigiremos a la barra de menús en el apartado de Extensiones, en el seleccionamos la herramienta de Complementos y dentro de ella optamos por descargar complementos.



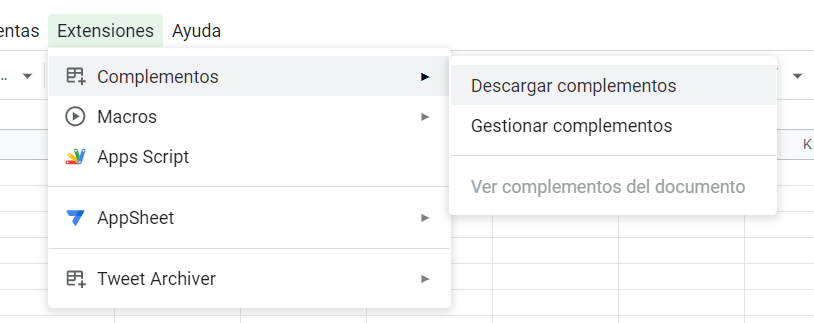
Una vez seleccionamos esta herramienta se nos desplegara una ventana emergente donde podremos buscar la herramienta que necesitemos, en nuestro caso TWEET ARCHIVER.



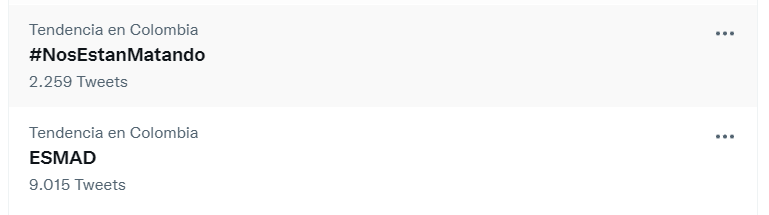
En la barra de búsqueda buscamos la herramienta que necesitamos por su nombre directo, al encontrarla solo la descargamos. Debemos tener en cuenta que la herramienta se ligará a la hoja de calculo de la cuenta de GMAIL del que estemos logeados, a su vez se recomienda iniciar sesión en TWITTER en el mismo navegador pues a continuación se nos pedirá que vinculemos la herramienta a esta cuenta para poder extractar la información.



Ahora para poder utilizarla solo vamos al mismo menú de Extensiones y la encontraremos instalada, solo queda abrirla.

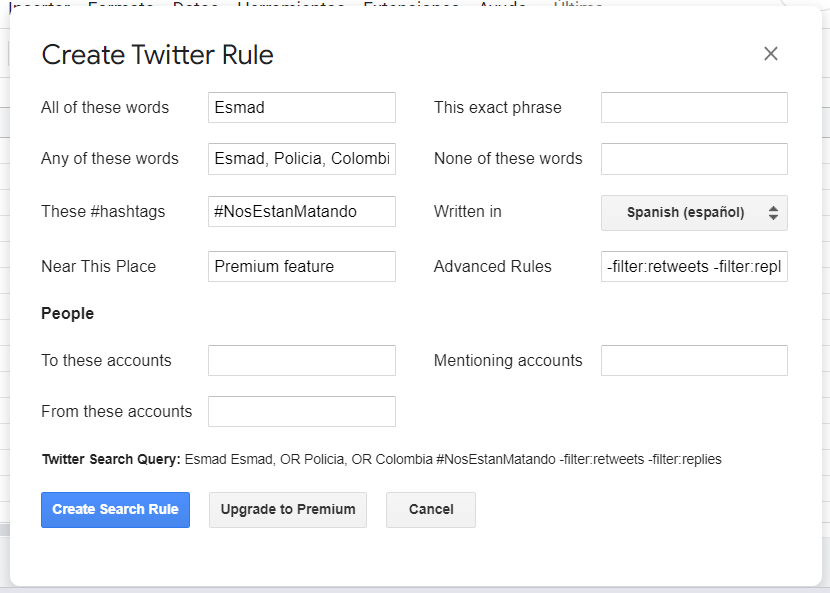


Paralelamente en nuestra cuenta de Twitter podemos buscar alguna tendencia o tema de nuestro interés del cual querramos rescatar los datos de las publicaciones en mi caso seleccione el siguiente

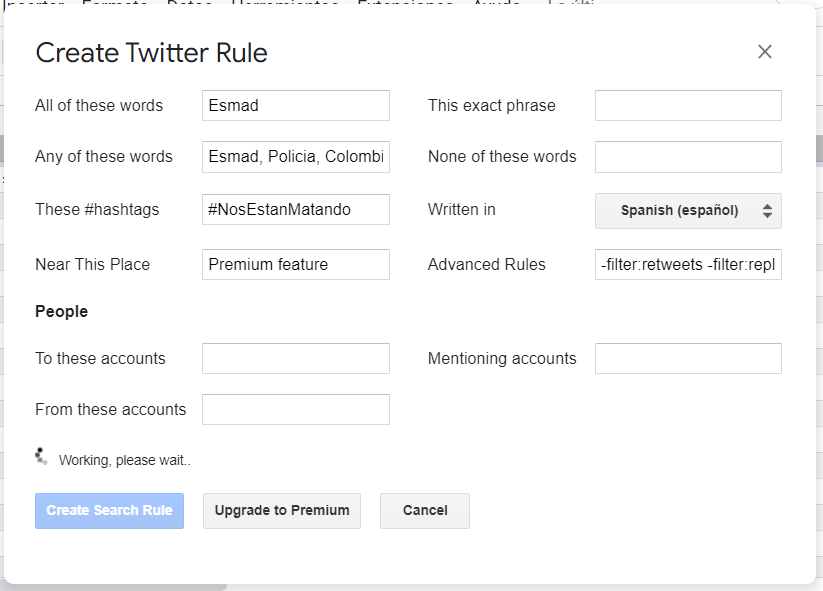


Ahora bien, dentro de nuestra hoja de calculo y con la herramienta de TWEET ARCHIVER abierta, daremos en la opción de crear una nueva regla “créate rule”.

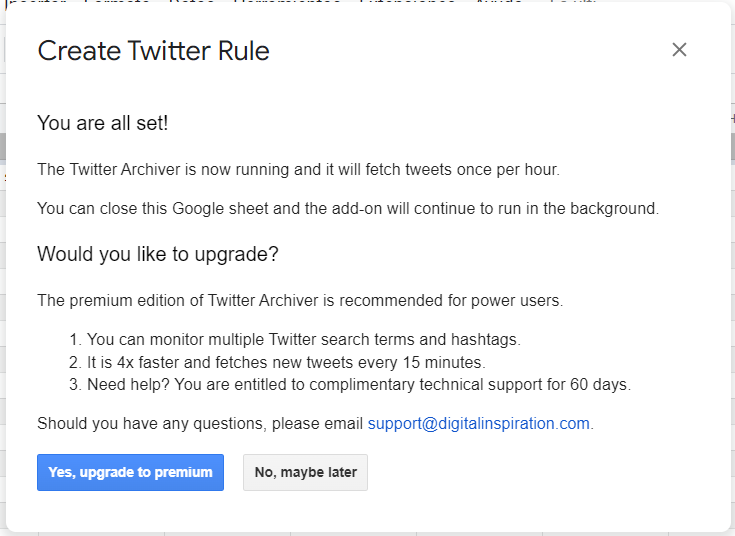
Allí daremos las indicaciones de que tan especifica queremos nuestra query para la búsqueda en la red social. Mientras más específica sea la consulta, como en toda base de datos, menores serán los resultados obtenidos, pero estos serán más de acuerdo a lo que necesitamos, esto podemos observarlo en la configuración de la consulta que hicimos. Pues a pesar de que el hashtag tenga 2.259 Tweets y la palabra ESMAD tenga otros 9.015, realmente rescataremos muchísimos menos tweets propios de la especificidad de la consulta.

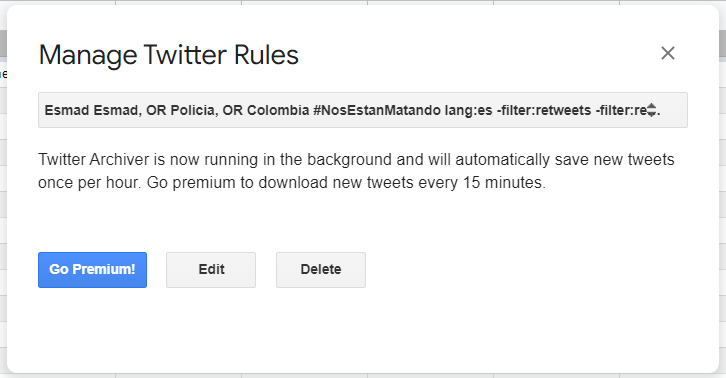


Una vez puestos los datos podemos dar en el botón azul que tal como especifica, creara la regla de búsqueda o el Query y empezara la búsqueda en la red social a través de nuestra cuenta personal. Podemos observar la consulta en la parte inferior, antes de la sección de botones como “Twitter Search Query” la cual al iniciar la búsqueda se reemplazar con un mensaje de “Working, please wait”.

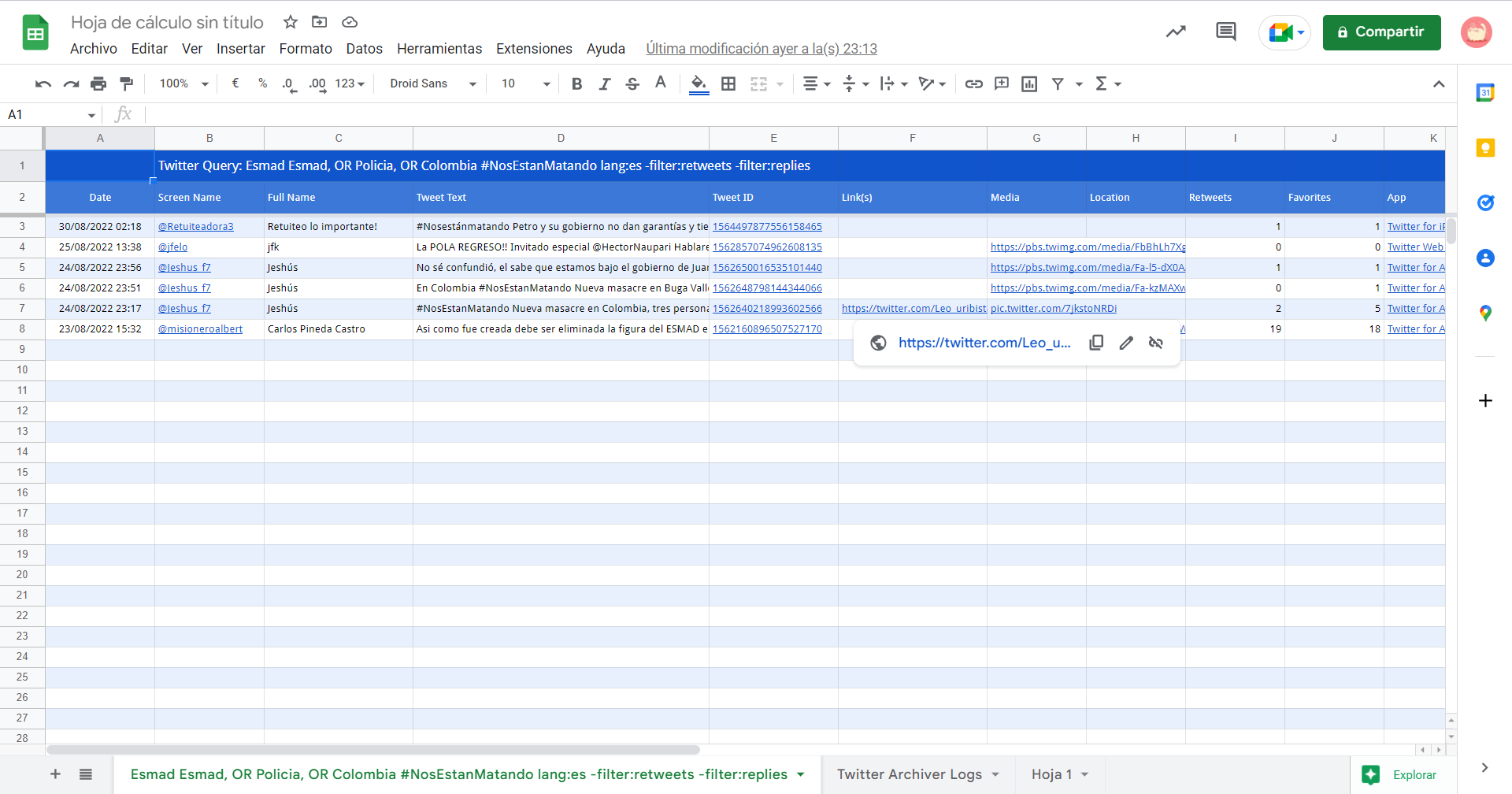


Luego de esto, obtendremos en nuestra hoja de calculo los datos de los tweets que hayamos rescatado. Sin embargo, aun tendremos unas ventanas emergentes de nuestra herramienta que nos recordaran que la misma tiene una opción de uso gratuito y su capa de pago. Así como el mostrarnos que reglas tenemos creadas hasta el momento, en caso de modificar o crear una nueva.



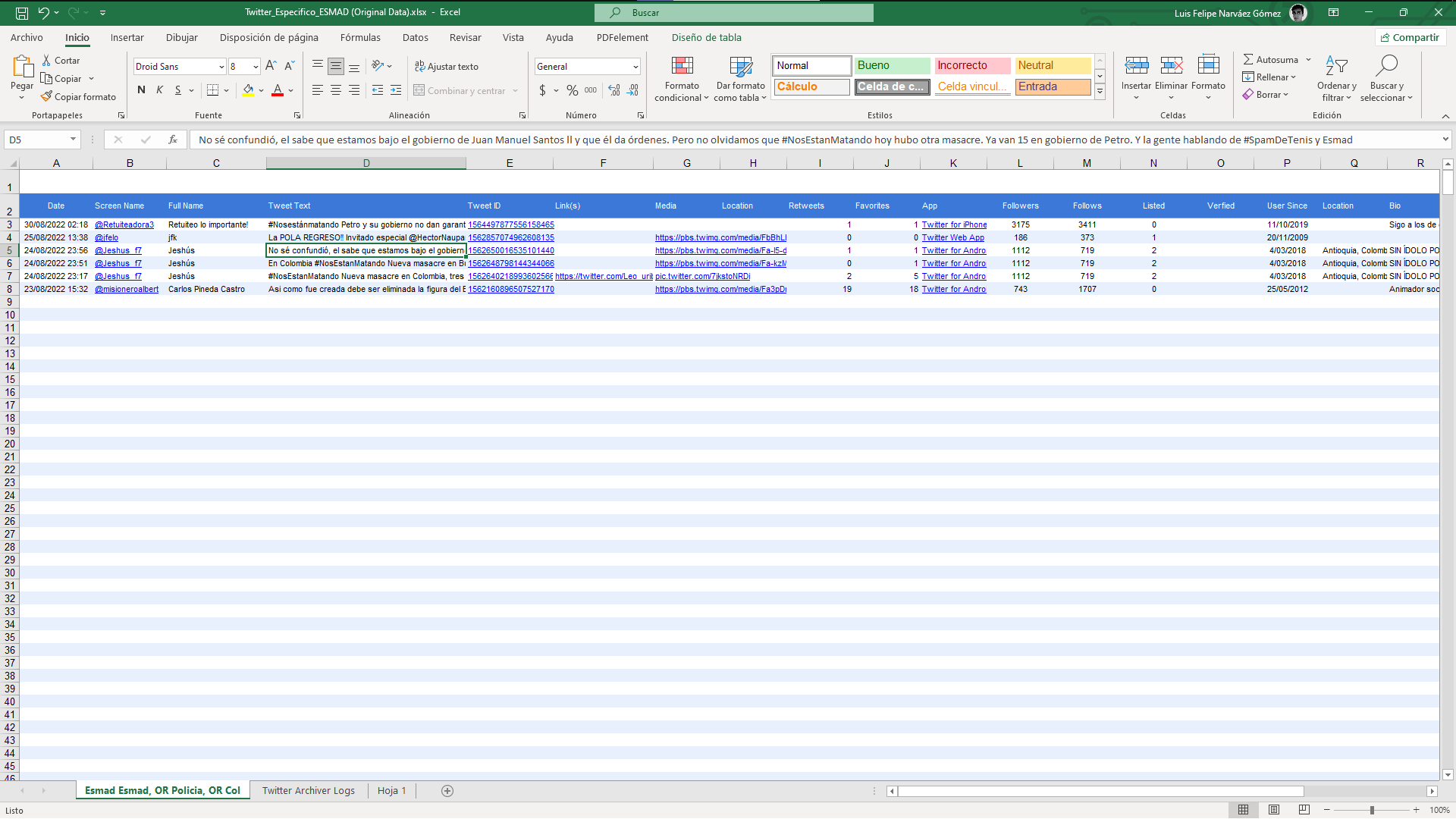


En nuestra hoja de calculo tendremos algo similar a lo siguiente:

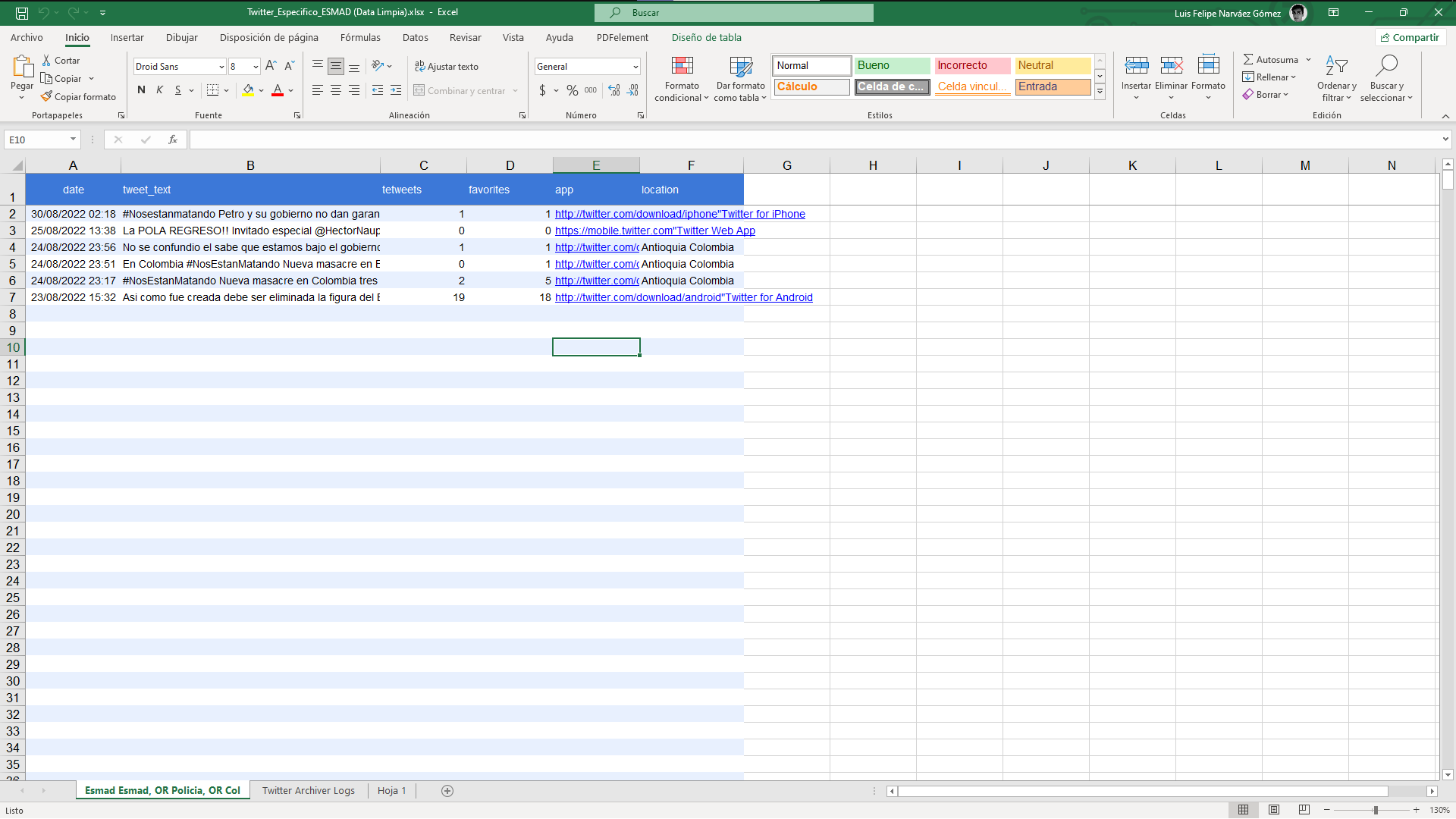


Solo queda descargar el archivo como un archivo de Excel, no directamente como CSV pues aun necesitamos limpiar la data que tenemos. Cuando nos referimos a limpiar la data hablamos de eliminar todos aquellos caracteres especiales que no necesitamos o pueden presentarse un conflicto al momento de subirlos en la base de datos por su codificación. A su vez también hablamos de eliminar datos que por norma y/o especificación de lo que necesitamos vamos a eliminar las columnas.

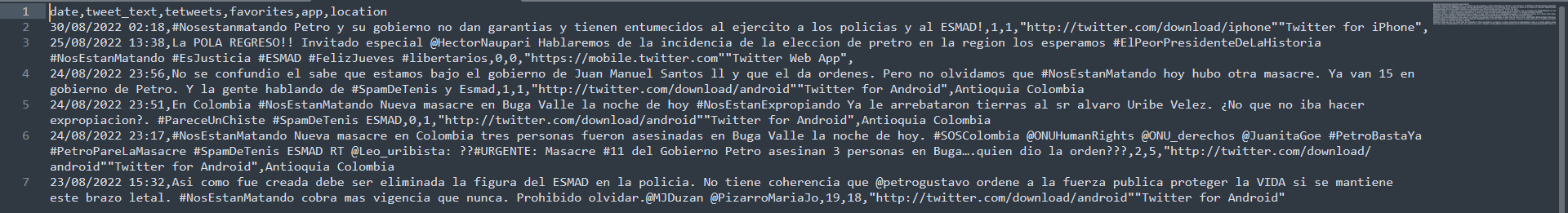
Así luce la data original:



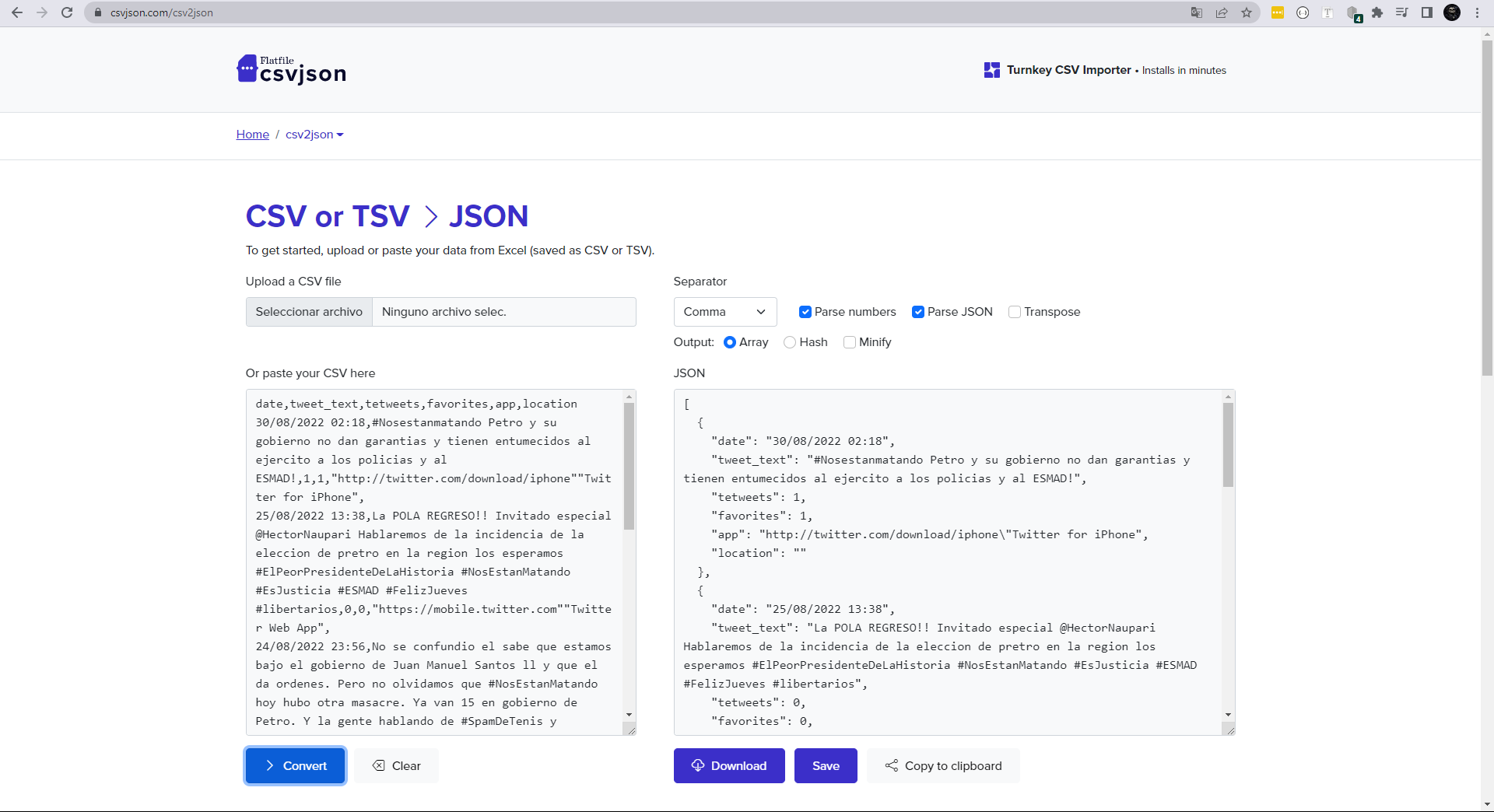
Ahora así luce la data en limpio.



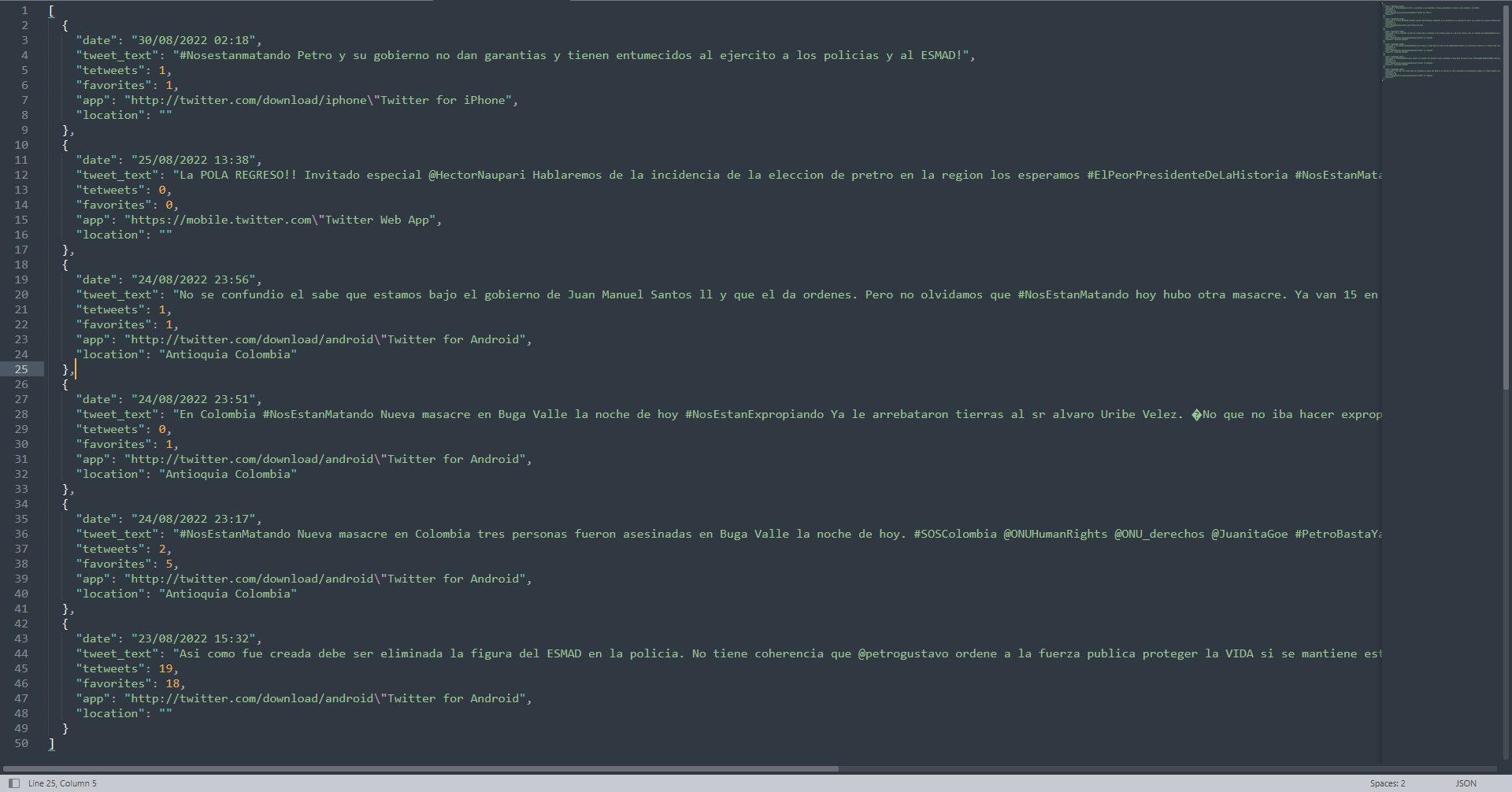
Una vez tenemos los datos limpios, podemos guardar la hoja como un CSV delimitado por comas, el mismo lo abriremos con un editor de texto básico como Bloc de Notas, esto con la finalidad de corroborar que cada dato de las filas este separado por comas y no por otro carácter, si este ultimo es el caso podemos reemplazar el mismo por la coma.



Una vez tenemos nuestra data visitamos el sitio web <https://csvjson.com/csv2json> para convertir nuestro archivo CSV a JSON.



Al descaragar este archivo obtendremos un archivo de diccionarios Llave/Valor de a siguiente manera:



Solo queda subir nuestros datos a una colección de Mongo DB. Las instrucciones de inserción de un archivo JSON en MongoDB como su confirmación con la búsqueda de información y el conteo de datos insertados lo podemos ver a continuación:

