
# Máster Universitario en Big Data y Ciencia de Datos

ASIGNATURA: 01MBID *Fundamentos de la Tecnología Big data*

*Actividad 1 – Tareas portafolio*

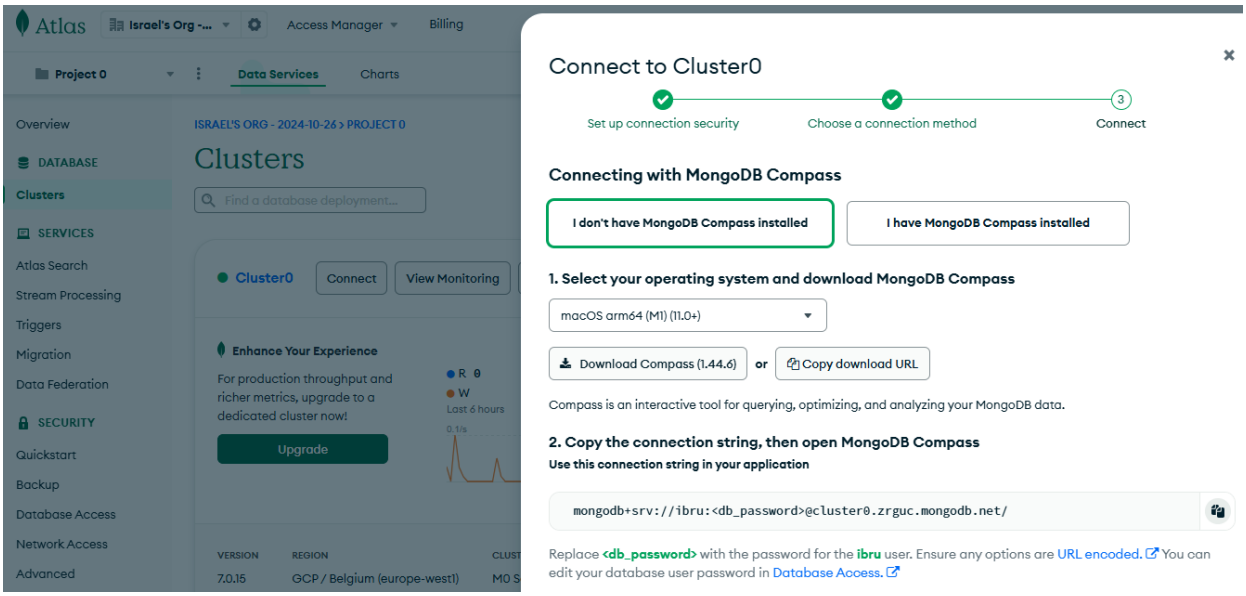
Alumno: **Bru Montes, Israel**

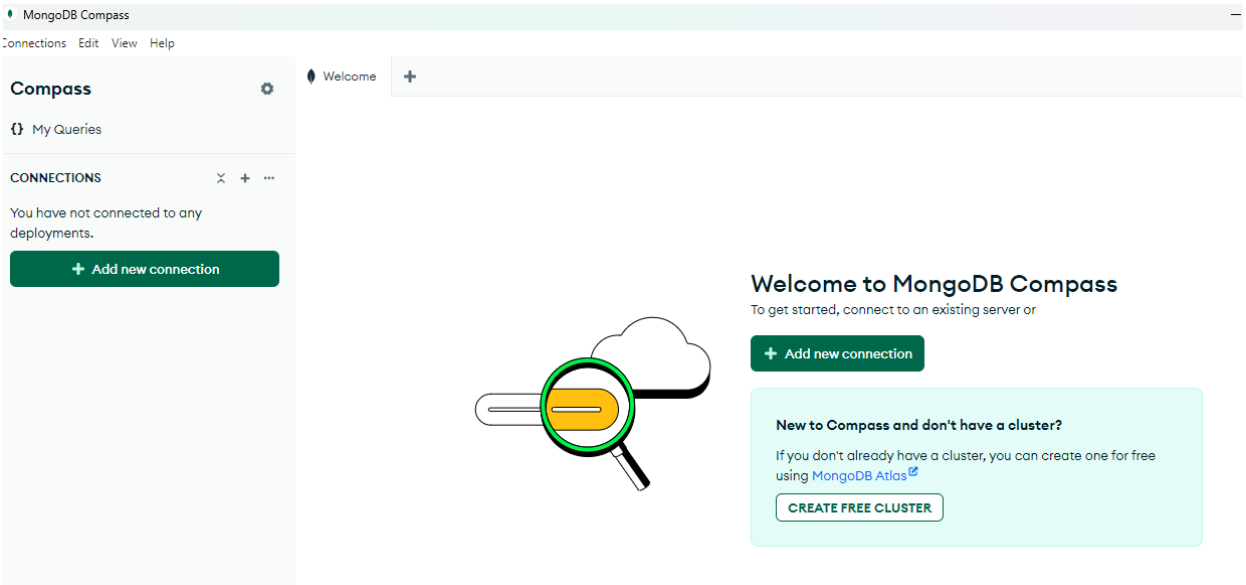
Edición **Octubre 2024 – Grupo A** a 8/11/2024

# 1. Tareas

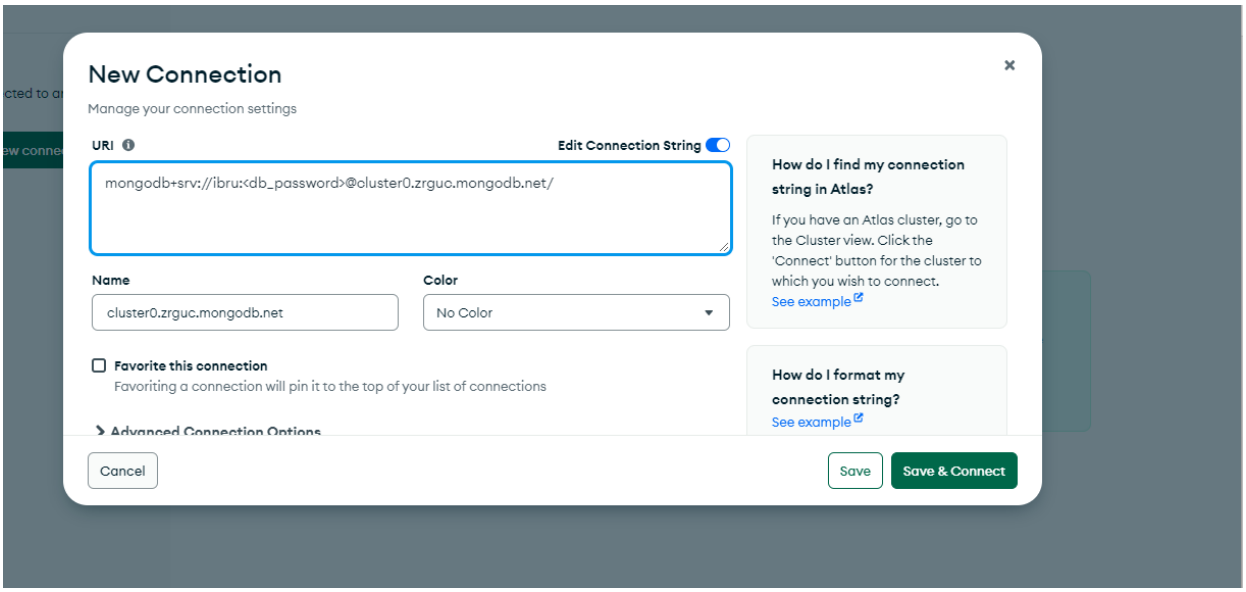
## 1.- Crear una base de datos MongoDB en la nube usando MongoDB Atlas (https://www.mongodb.com/)

Nos vamos a descargar el GUI compass para poder administrar la base de datos desde una aplicación de escritorio, para ello cogemos la url de conexión y cambiamos el parámetro <password>

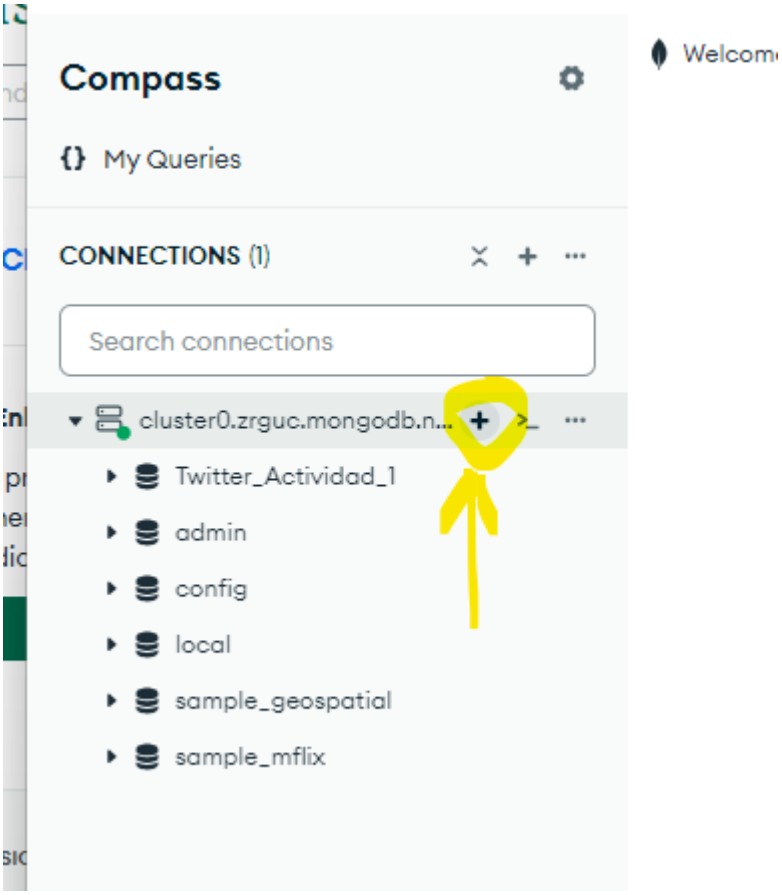


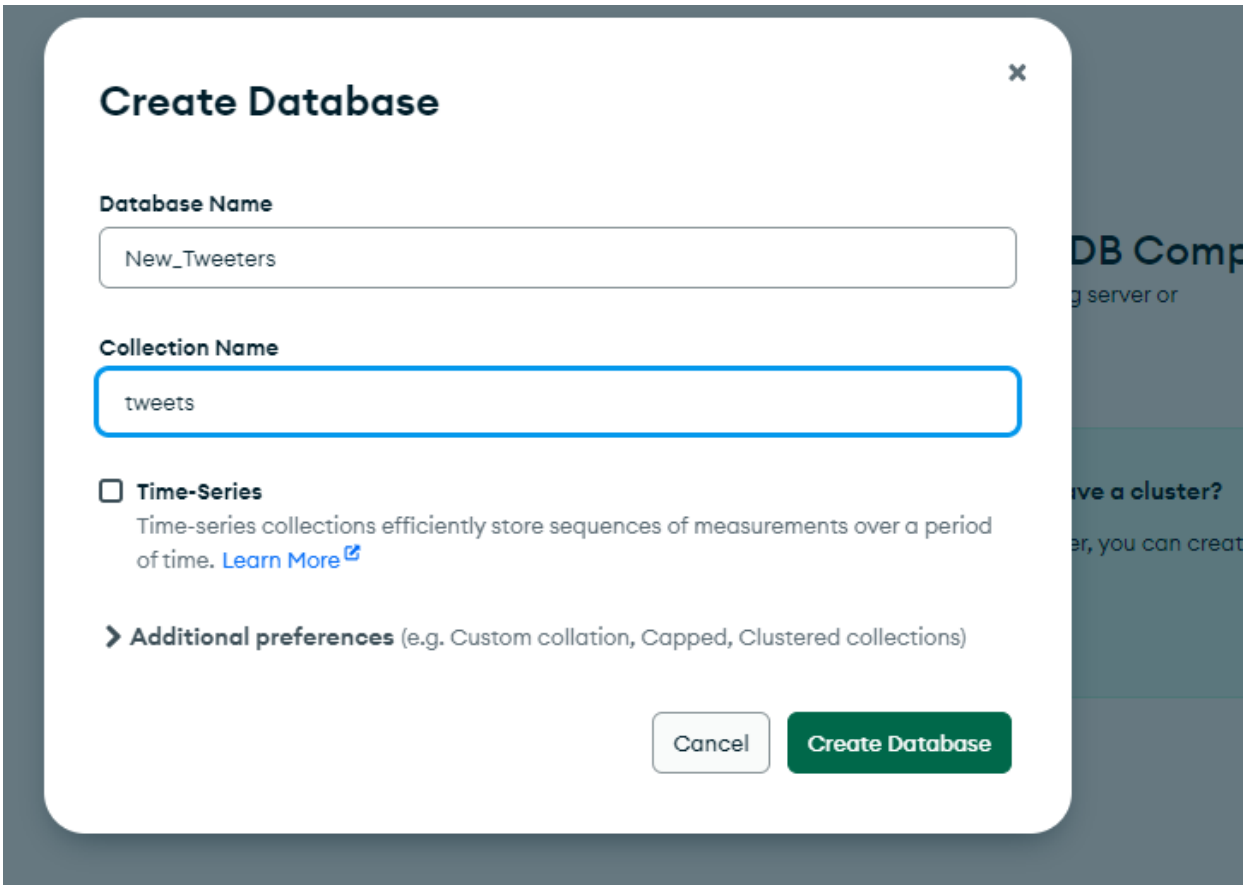




Pulsamos sobre “add new connection”

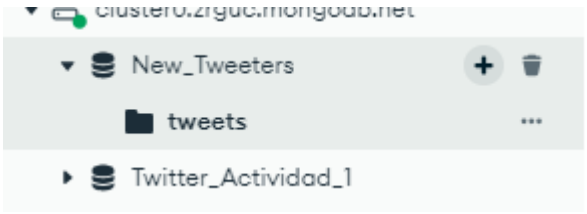


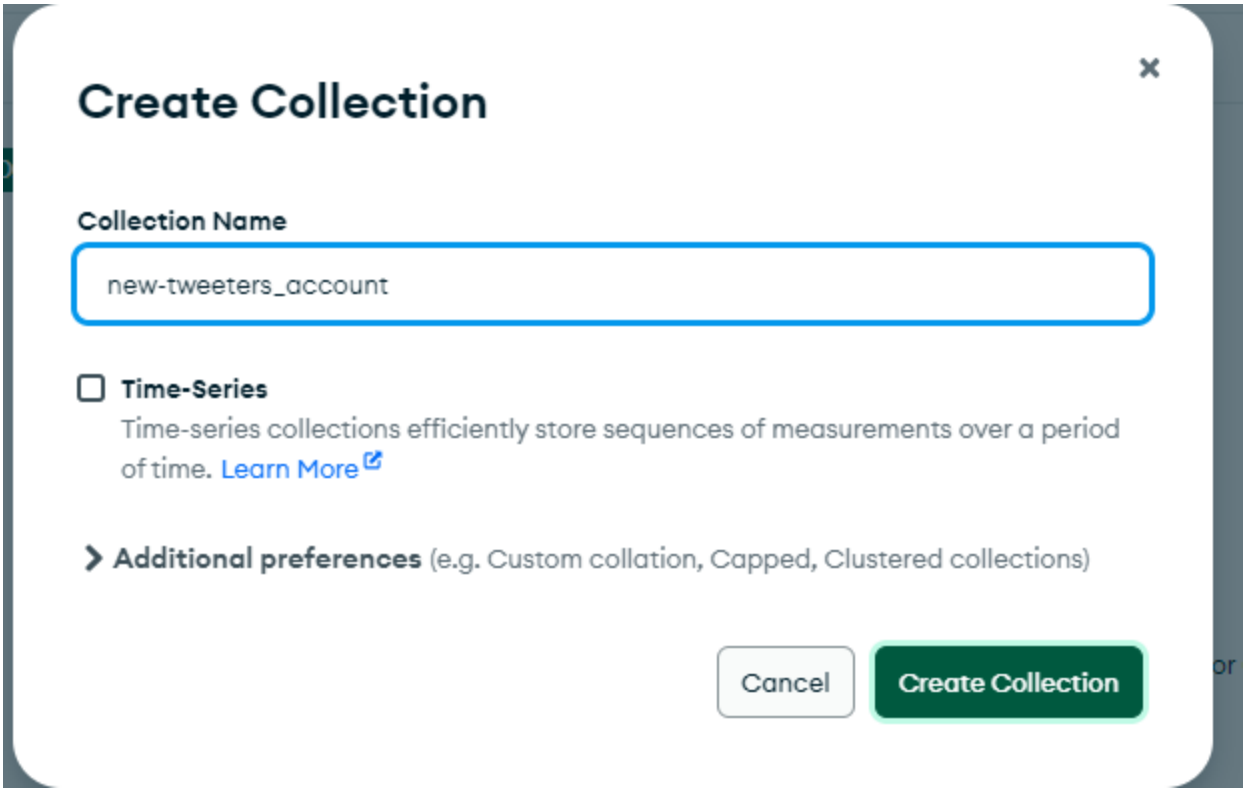

A través del botón “+”

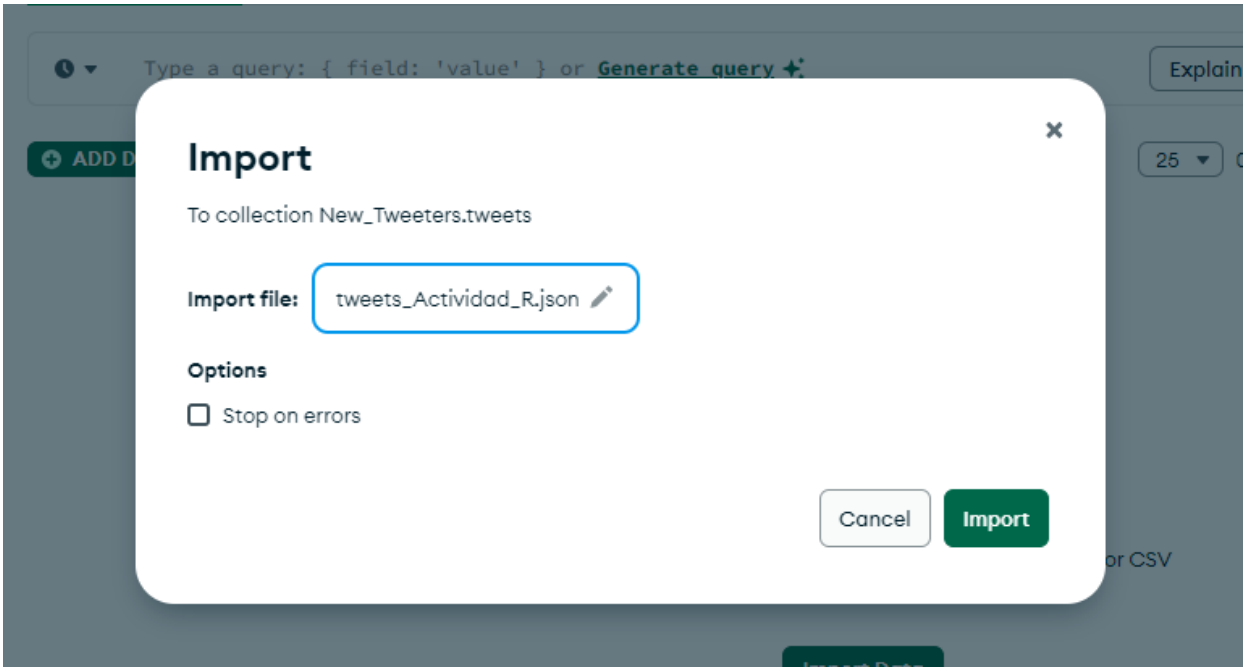


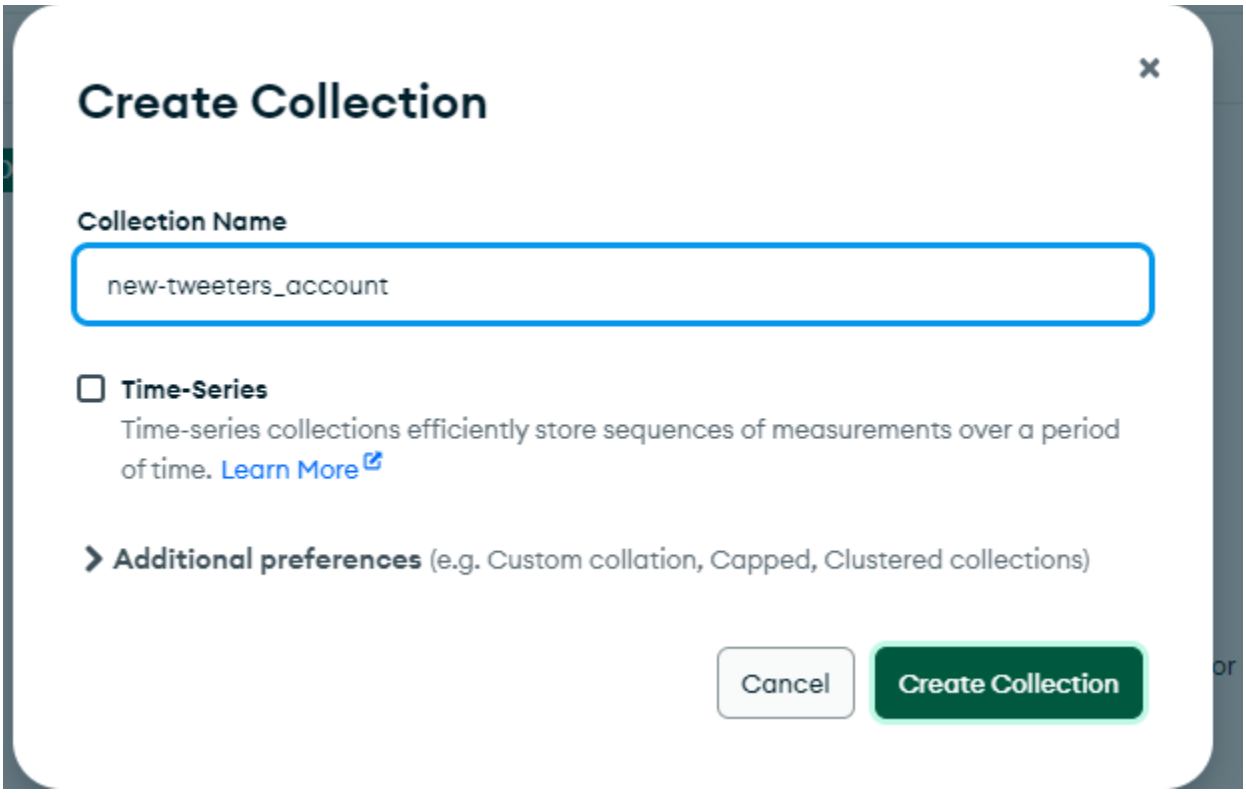


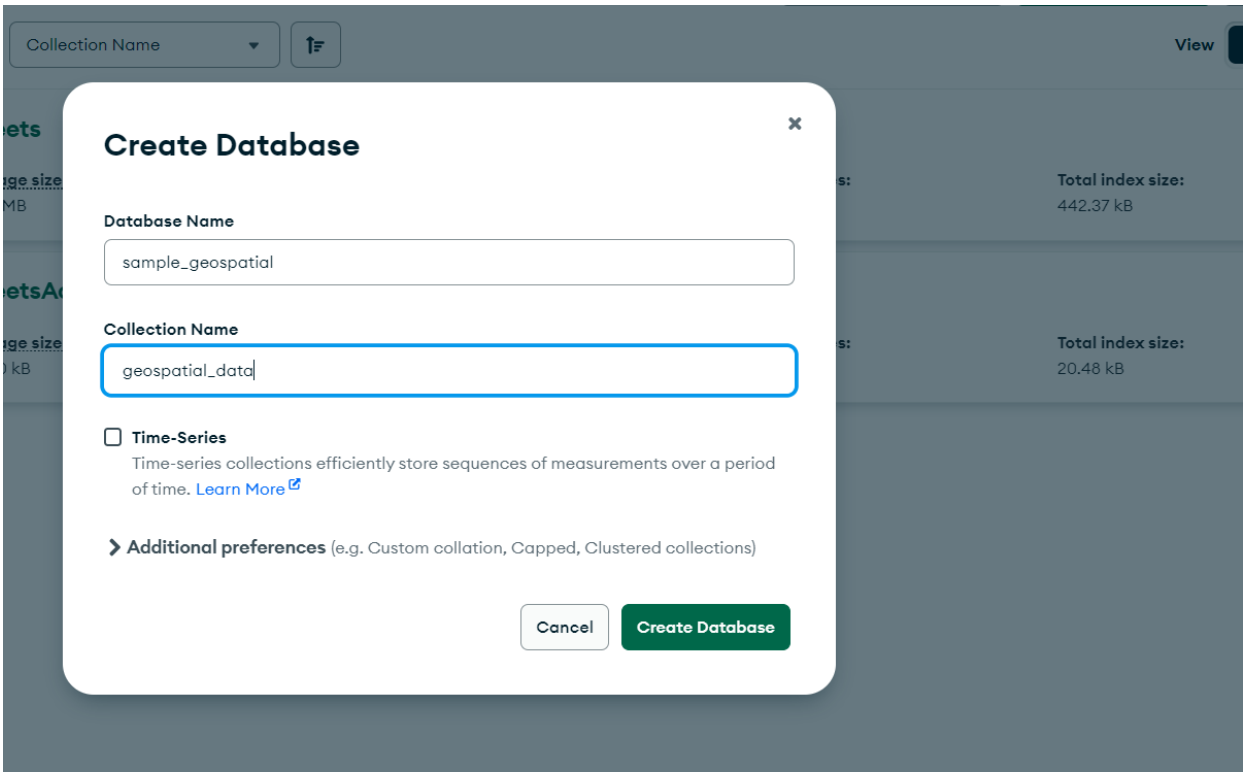

## 2.- Crear las colecciones



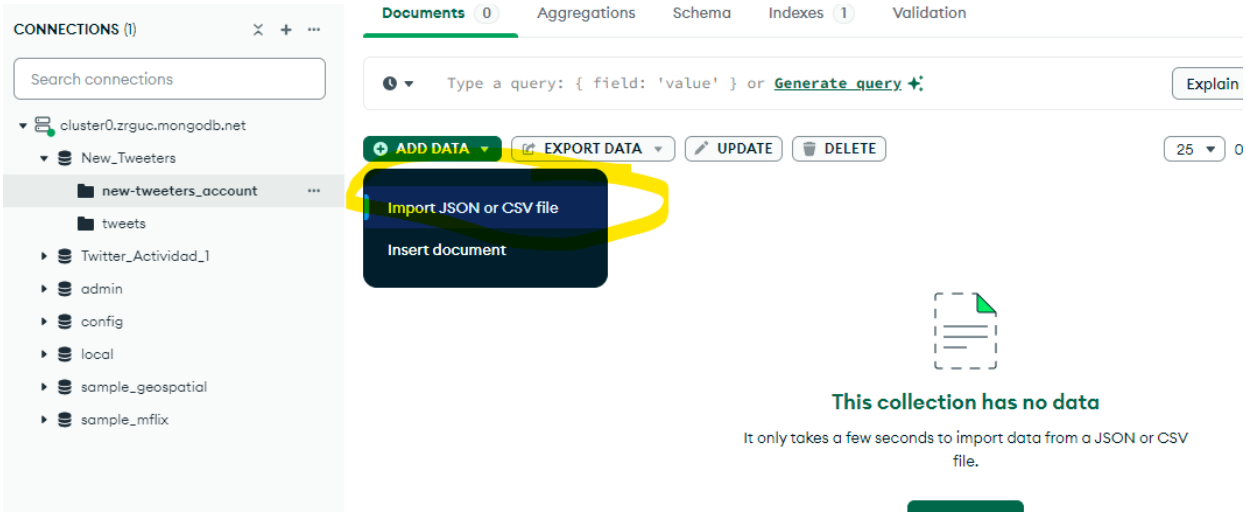

Create Collection
Collection Name
new-tweeters_account
Time-Series
Time-series collections efficiently store sequences of measurements over a period of time. Learn More
Additional preferences (e.g. Custom collation, Capped, Clustered collections)
Cancel
Create Collection


Type a query: { field: 'value' } or Generate query
Explain
ADD D
25
Import
To collection New_Tweeters.tweets
Import file:
tweets_Actividad_R.json
Options
Stop on errors
Cancel
Import
or CSV

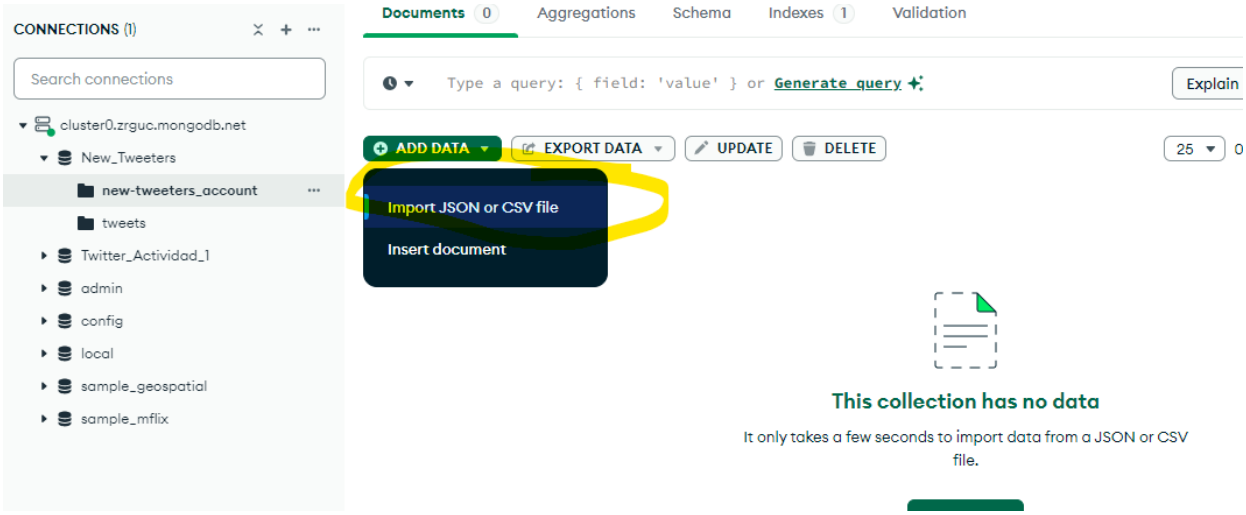


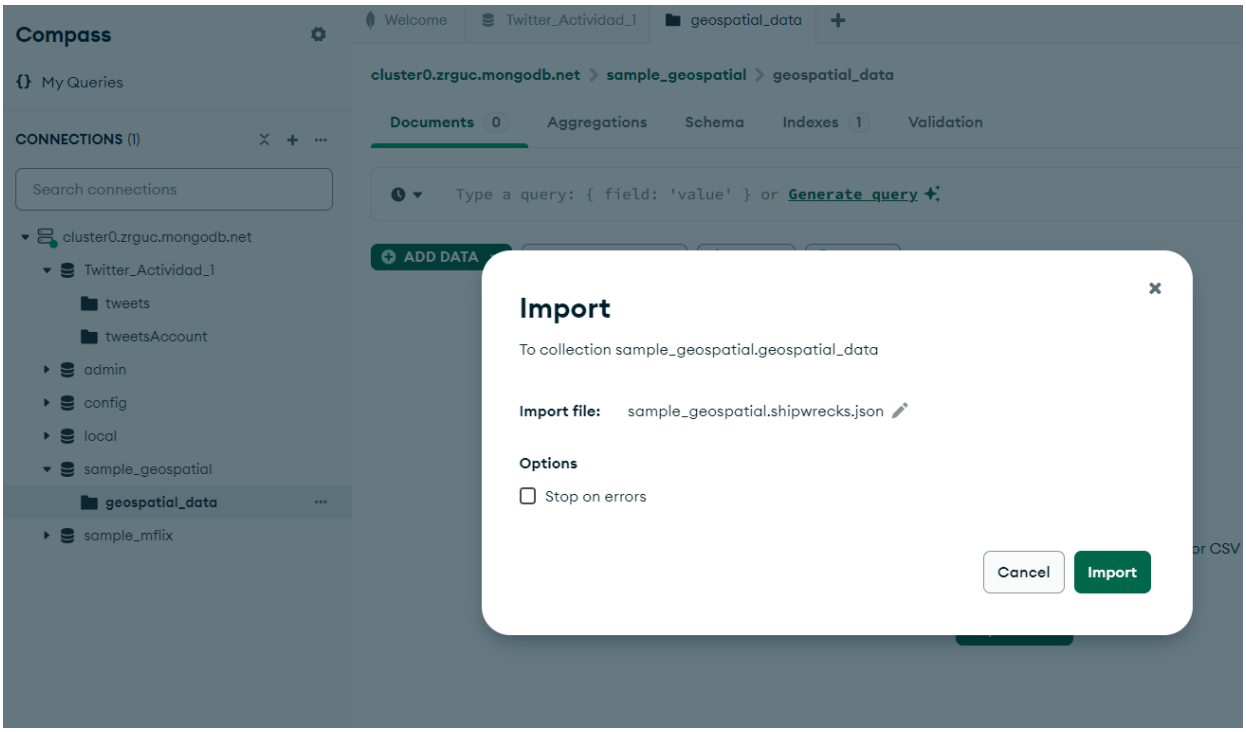




## 2.1.- Cargar los datasets de cuentas de twitter y tweets proporcionados por el profesor.



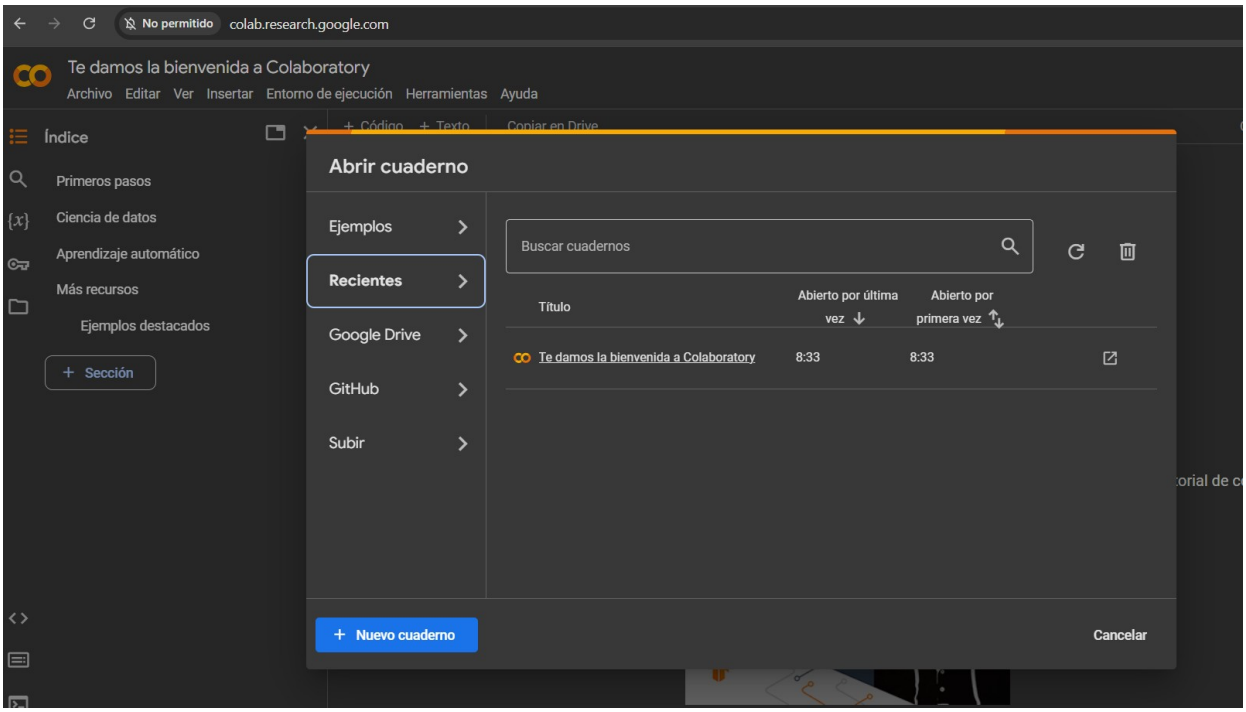

## 2.2.- Cargar la colección de ejemplo geolocalizada de MongoDB Atlas (sample_geospatial -> shipwrecks).



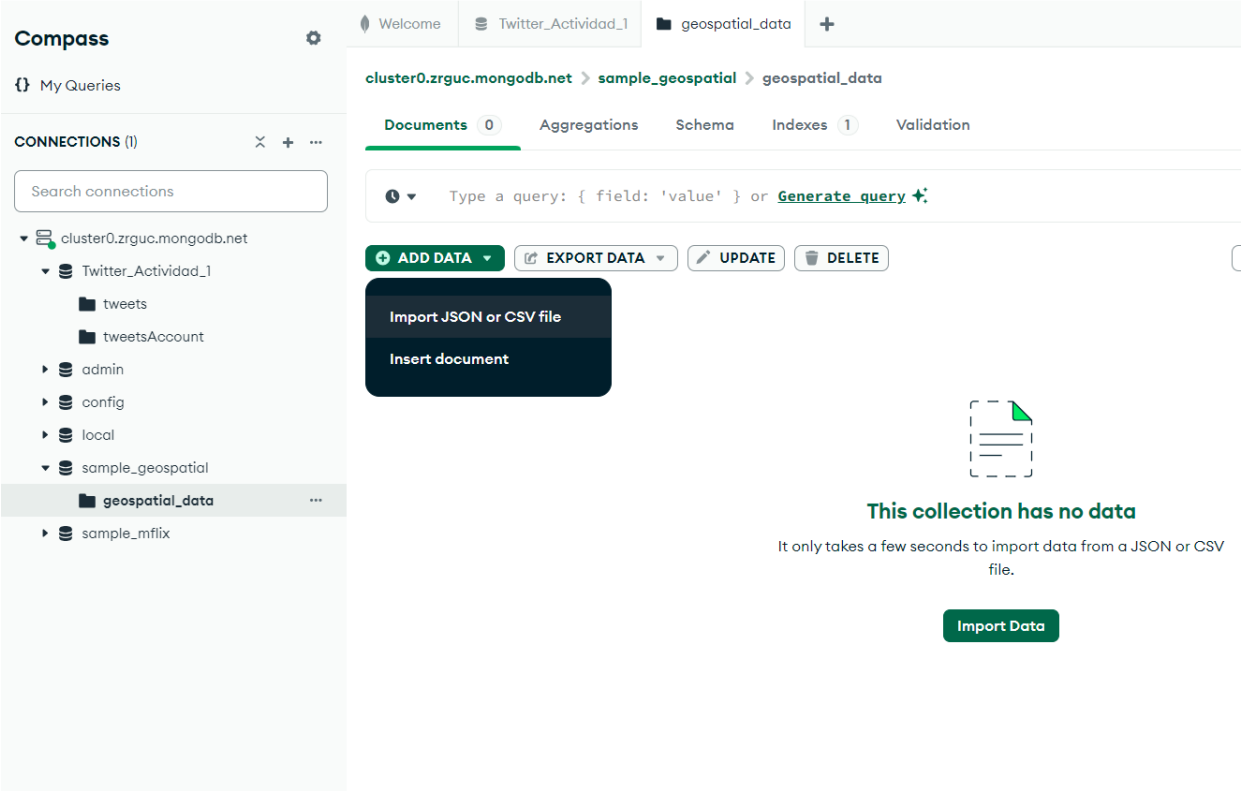



## 3.- Crear una cuenta en Google Colab.

Aunque con una cuenta de gmail es supersencillo, te das de alta directamente y cargas el notebook.



4.- Utilizando como base el script en Python profesor, realizar los cambios necesarios para:




4.1.- En la colección de cuentas de twitter, tener los campos amigos y tweets enviados, cargar los datos correspondientes mediante consulta mongodb + código python.

En el archivo phyton.

4.2.- En la colección de tweets, calcular la antigüedad para cada tweet en función de la fecha actual considerando antigüedad 0 el día de hoy y sumando +1 por cada día transcurrido. Incluir en el mismo documento del tweet un nuevo campo que se llamará antiguedad_dias con esa antigüedad calculada.

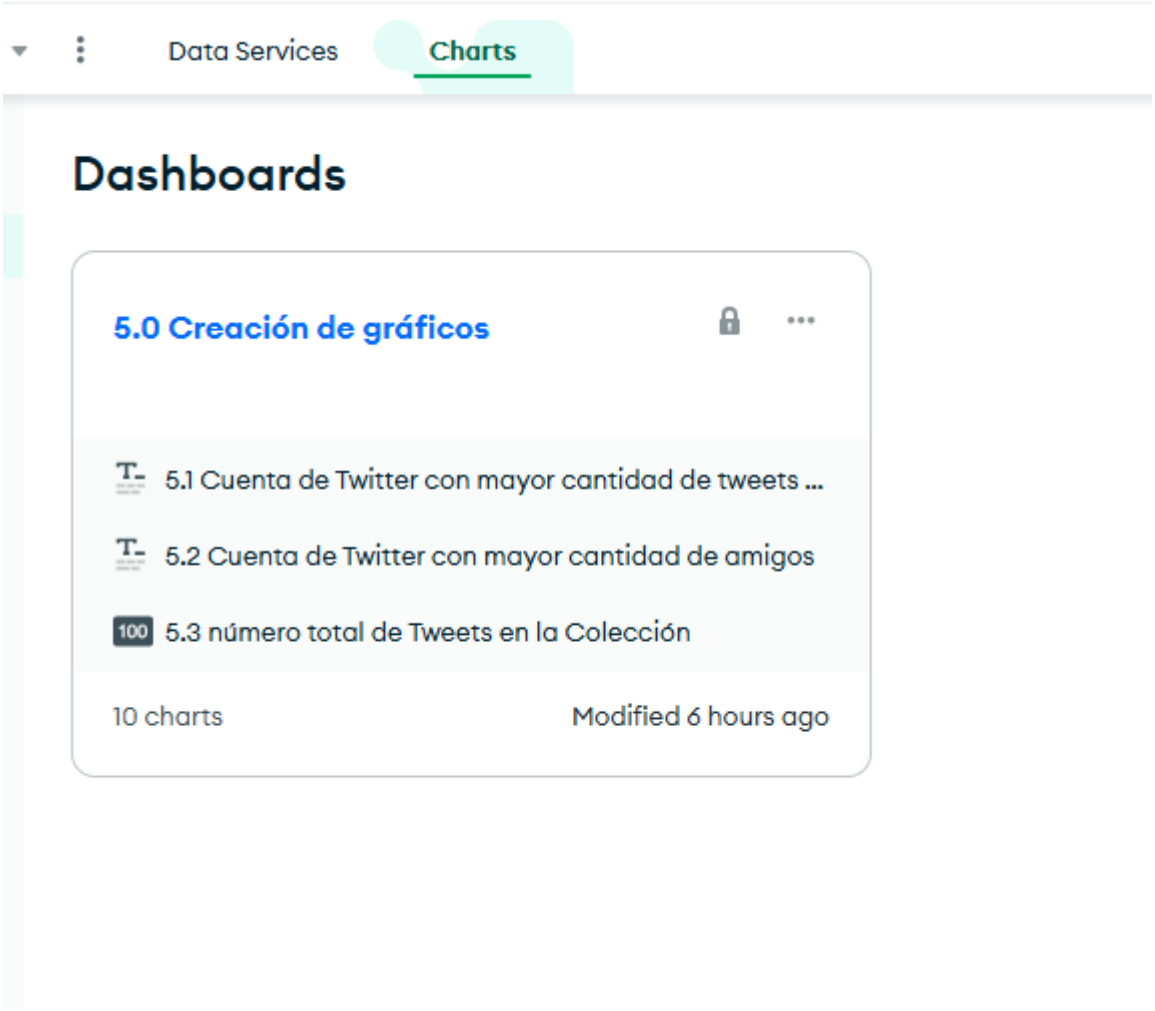
En el archivo phyton.

4.3.- En la colección de tweets, calcular la antigüedad de cada tweet relativa con la fecha de creación de la cuenta más antigua de la colección actual de datos. Considerando antigüedad 0 si fue enviado el mismo día de creación y sumando +1 por cada día transcurrido desde entonces en función de la fecha

del tweet. Incluir en el mismo documento del tweet un nuevo campo se llamará frescura_relativa_dias.

En el archivo phyton.

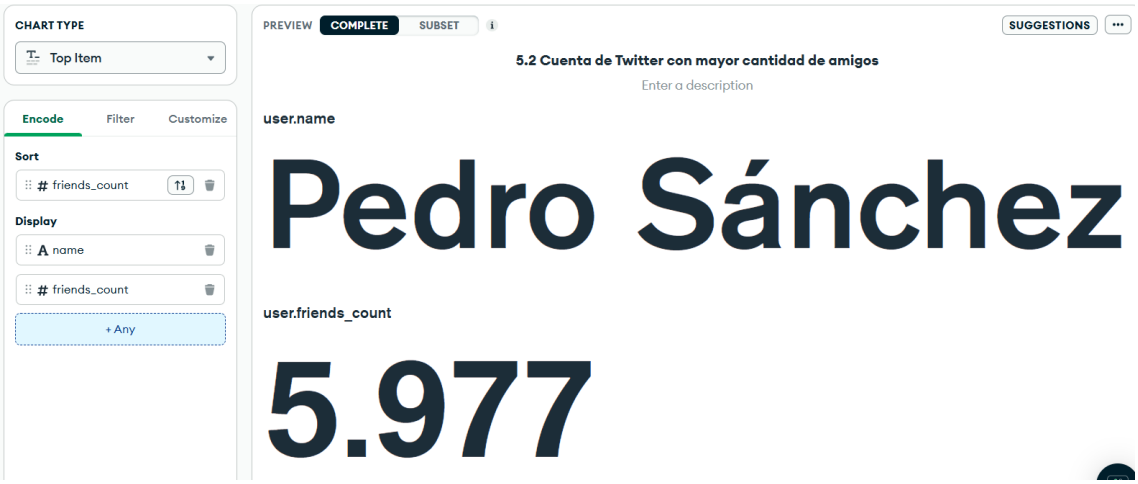
## 5.- Cuadro de mandos (MongoDB Charts)




## 5.1.- Crear un chart que muestre la cuenta de Twitter con mayor cantidad de tweets enviados.

He elegido “top item” porque permite mediante el campo “sort” ordenar el campus agregado “statuses_count” y mostrarlo.

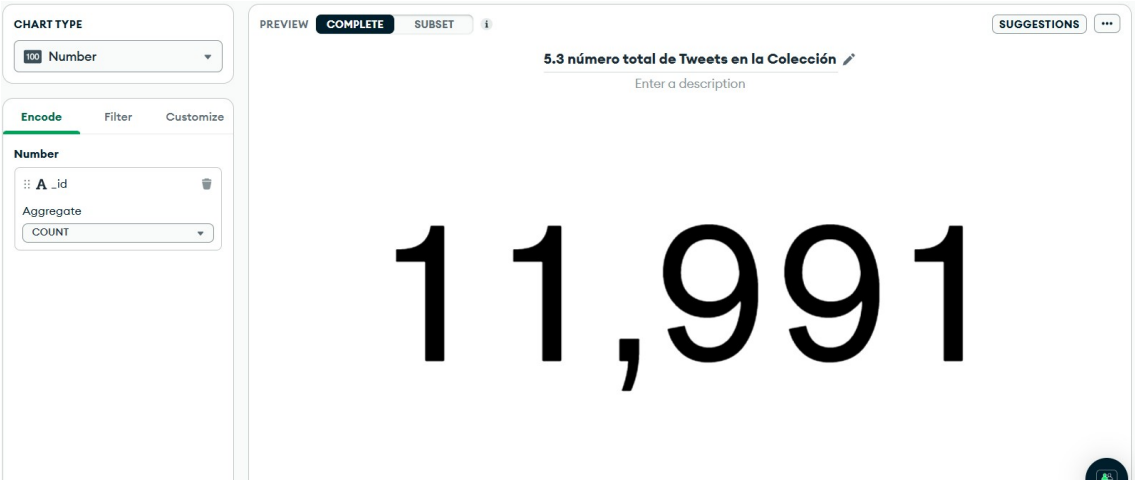
## 5.2.- Crear un chart que muestre la cuenta de Twitter con mayor cantidad de amigos (cuentas a las que sigue el usuario).



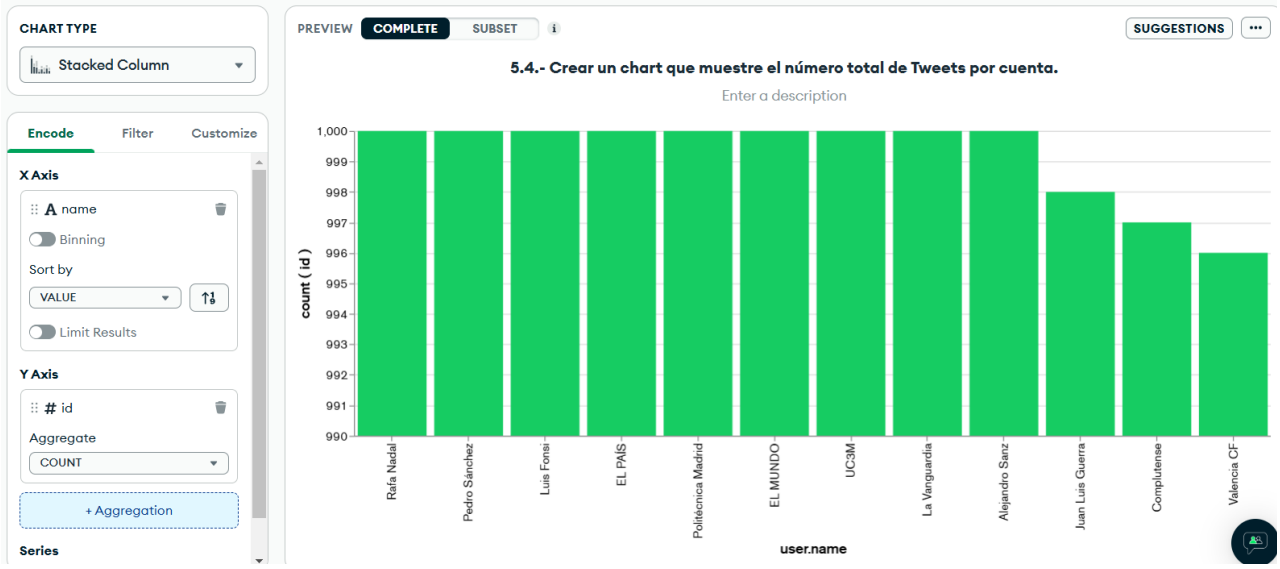

Como el anterior, con el tipo de gráfico “top item” nos permite crear una tarjeta con un valor agregado y seleccionar tanto el máximo como el mínimo, para nuestro caso seleccionamos el máximo.

## 5.3.- Crear un chart que muestre el número total de Tweets en la Colección.

Con el gráfico “Number” podemos obtener de forma agregada, para nuestro caso hacemos un count, todos los tweets que existen en nuestra colección.



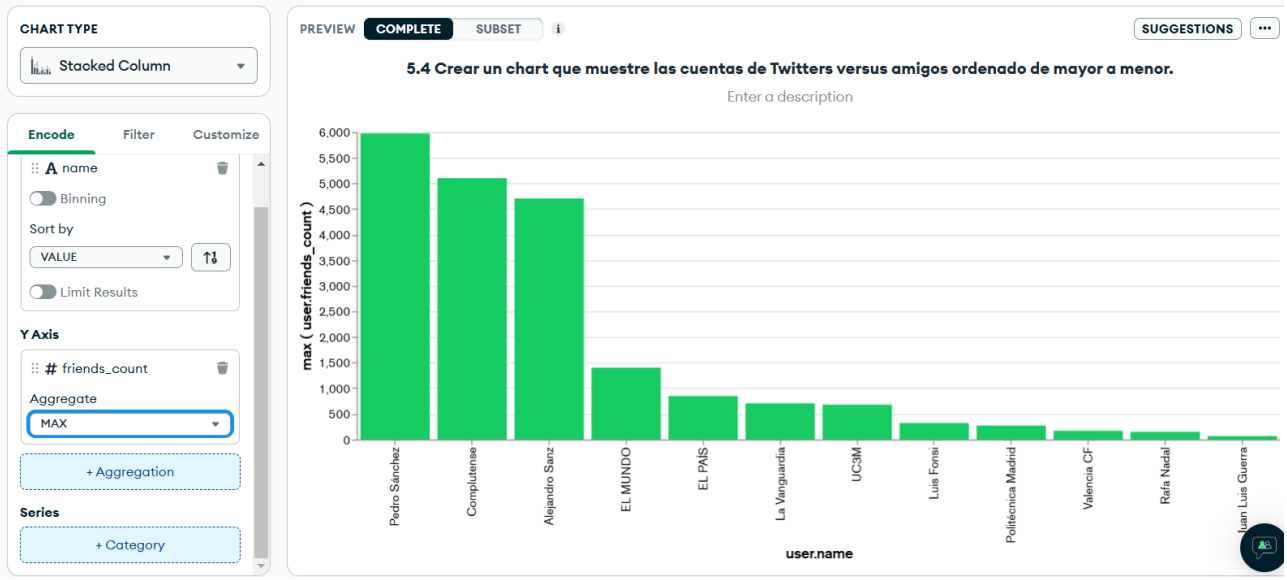

## 5.4.- Crear un chart que muestre el número total de Tweets por cuenta.



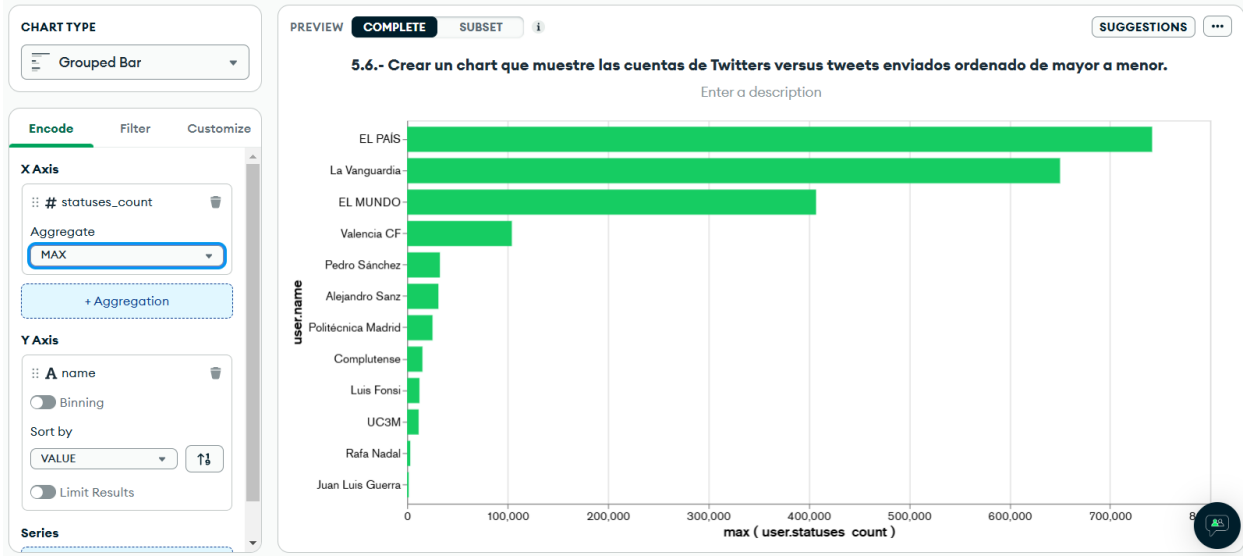

Utilizando el gráfico de barras por columnas “Stacked column”, hemos ordenado en modo escendente el número de tweets (count(id)) de cada cuenta de Tweet (name); hemos tenido que tilizar la “customizacion” para poder escalar el mínimo a 990, ya que no se apreciaba bien el ráfico.

## 5.5.- Crear un chart que muestre las cuentas de Twitters versus amigos ordenado de mayor a menor.

Elegimos el tipo de gráfico de barras “Stacked Column”, para mostrar la cuenta de Twitter haciendo un “max(friends.count” y ordenarlo de forma descendente.

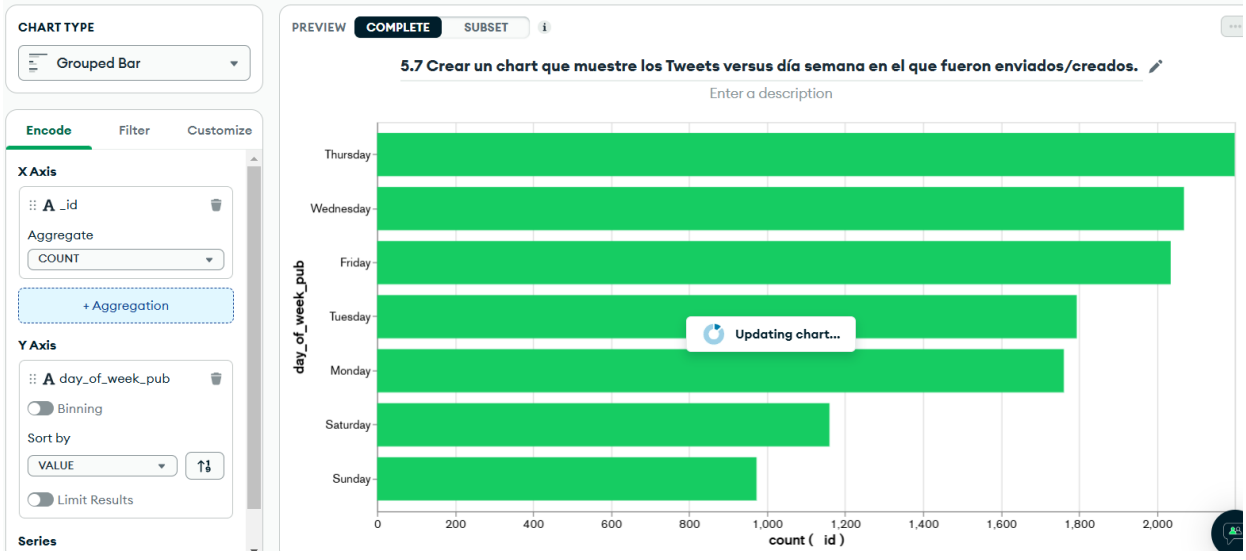


## 5.6.- Crear un chart que muestre las cuentas de Twitters versus tweets enviados ordenado de mayor a menor.




Aquí hemos utilizado el gráfico “grouped Bar” y hemos obtenido el máximo de los tweets enviados por cuenta, ordenandolo de forma descendente (el que tiene mayor número el primero).

## 5.7.- Crear un chart que muestre los Tweets versus día semana en el que fueron enviados/creados.

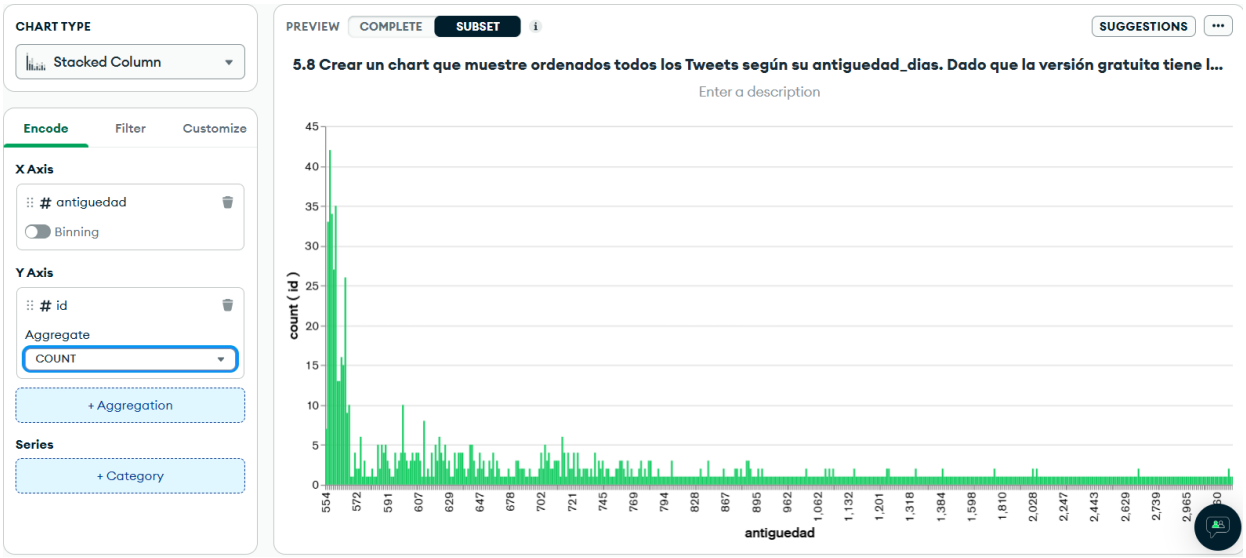



He creado desde la actividad 4.2 la columna fecha_pub y day_of_week_pub para indicar el día de la semana que se creo

## 5.8.- Crear un chart que muestre ordenados todos los Tweets según su antiguedad_dias. Dado que la versión gratuita tiene limitaciones de la cantidad

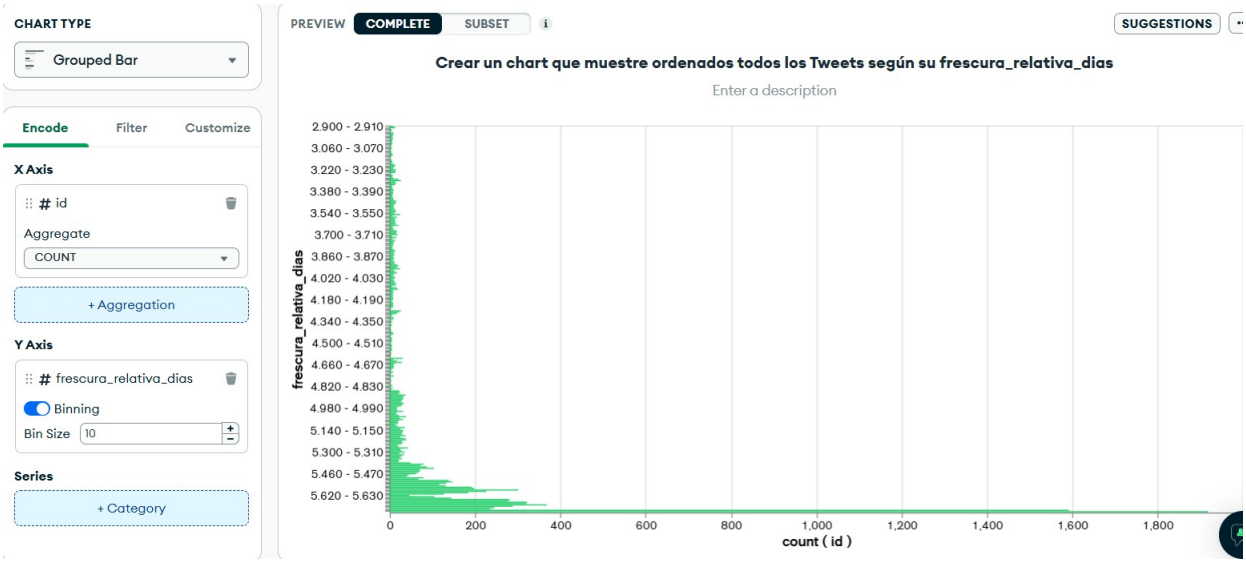
de datos a mostrar, mostrar dos charts: uno ordenado de mayor a menor y otro de menor a mayor.

Mostrando el count de los tweets agrupados sobre la antigüedad de estos.




## 5.9.- Crear un chart que muestre ordenados todos los Tweets según su frescura_relativa_dias. Dado que la versión gratuita tiene limitaciones de la cantidad de datos a mostrar, mostrar dos charts: uno ordenado de mayor a menor y otro de menor a mayor.

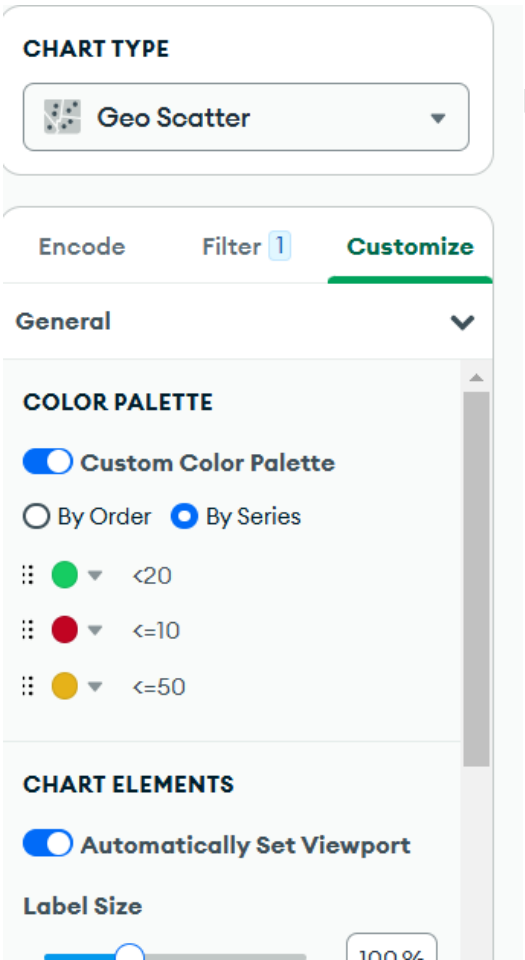
Con el group Bar mostramos los datos de la frescura relativa de dias respecto de los tweets.



## 5.10.- Crear un chart para visualizar en el mapa mundial todos los pecios hundidos hasta 50 metros de profundidad. Pintando en verde los hundidos hasta los 10 metros, amarillo hasta los 20 metros y en rojo hasta los 50 metros.

Creando una nueva columna “depth_group” que agrupamos los resultados y después filtramos aquellos que son ‘+50’ para quitarlos:

```
{
        $cond: { if: {$lte: ['$depth', 10]}, then: '<=10',
                else: {$cond: { if: {$gte: ['$depth', 50]}, then: '+50',
                        else: {$cond: { if: {$lte: ['$depth', 20]}, then: '<20',
                                else: '<=50' }}
                        }
                }
        }
}
```




podemos crear una serie nueva:

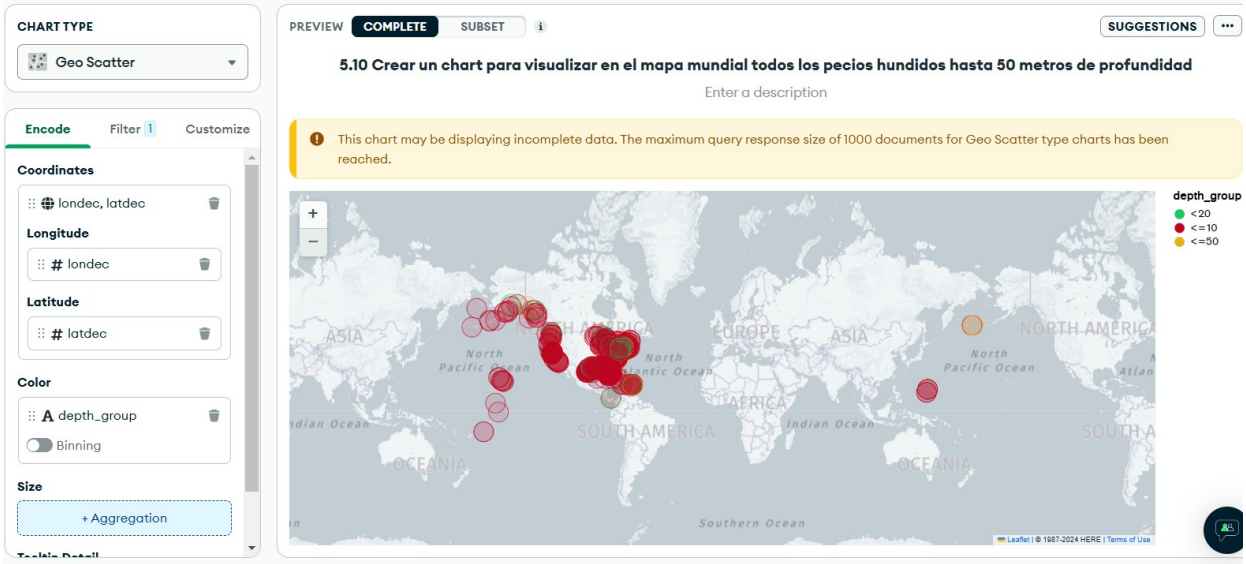


CHART TYPE
Geo Scatter
Encode
Filter 1
Customize
Coordinates
londec, latdec
Longitude
londec
Latitude
latdec
Color
depth_group
Binning
Size
+ Aggregation
PREVIEW
COMPLETE
SUBSET
SUGGESTIONS
5.10 Crear un chart para visualizar en el mapa mundial todos los pecios hundidos hasta 50 metros de profundidad
Enter a description
This chart may be displaying incomplete data. The maximum query response size of 1000 documents for Geo Scatter type charts has been reached.
depth_group
<20
<=10
<=50
ASIA
EUROPE
NORTH AMERICA
SOUTH AMERICA
OCEANIA
Southern Ocean
Leaflet | © 1987-2024 HERE | Terms of Use