

## 1.) Crear una aplicación en Twitter Developer, para obtener los tokens y keys de acceso

Una vez creada la cuenta en Twitter Developer he generado los siguientes valores:

For security, this will be the last time we'll display these. If something happens, you can always regenerate them. [Learn more](#)

### API Key

A7c89OZEucTIWxeff2ZJChGh

Copy

### API Key Secret

qMuUh6uZajedu8mhN2M5lqMjCk4KMNfBIZf4kyd95XGBPdkc

Copy

### Bearer Token

AAAAAAAAAAAAAAAAAAFDzUwEAAAAKpyVRU4RVWBzgQbUP  
oPq1EkRhWw%3DZ0naB70kA4pqbmMcW0VYtD6kyDBcAdN9L9LOCTK  
gtujtotbXPO

Copy

ACCESS TOKEN 1450729853116825603-qQY3Kfp2imGen2MK5tR81g6VvGG6VF

Access Token Secret jNuDewNo6jDqGj9wbZ6ey9A4JIGfkJyQPX3jHjOx267oa

- 2.) Utilizando el script de Python facilitado por el profesor, realizar los cambios necesarios para:
- Elegir las cuentas de Twitter de donde extraer los datos (csv).

```
TwitterMongo01MBID.py accountsEMBS.csv
1 Unique_ID,org_name,org_url,Twitter_URL,Twitter_handle,earliest_tweet_in_db,number_of_tweets_in_db
2 1,'Netflix',https://twitter.com/netflix,netflix,,
3 2,'Disneyplus',https://twitter.com/disneyplus,disneyplus,,
4 3,'Marvel',https://twitter.com/Marvel,Marvel,,
5 4,'PrimeVideo',https://twitter.com/PrimeVideo,PrimeVideo,,
6 5,'amazon',https://twitter.com/amazon,amazon,,
7 6,'paramountplus',https://twitter.com/paramountplus,paramountplus,,
```

En la imagen anterior podemos apreciar las cuentas de Twitter que he seleccionado.

```
File Edit View Navigate Code Refactor Run Tools VCS Window Help pythonProject3 - ...\OneDrive\Documents\Maestria\01MBID\TwitterMongo01MBID.py
C:\Users\Grant\OneDrive\Documents\Maestria\01MBID\TwitterMongo01MBID.py
Project
71
72
73 ##### PARTE 4: Configurar la base de datos Mongo y las colecciones #####
74
75 #Establecimiento Conexión a MongoDB Atlas
76
77
78 # Datos de ejemplo, necesario modificar por vuestra instancia en MongoDB Atlas o local
79 dbStringConnection = "mongodb://JuanSebastian:Mecanica20@clusterviu-shard-00-00.ik7er.mongodb.net:27017,clusterviu-shard-00-01.ik7er.mongodb.net:27017?ssl=true&replicaSet=atlas01shard00-00&authSource=admin"
80
81
82 dbName = '01MBIDJuanDuenez'
83 dbCollectionA = 'TwitterActCuentas'
84 dbCollectionT = 'TweetsAct'
85
86 client = pymongo.MongoClient(dbStringConnection)
87
88 # Definición de la base de datos MongoDB
89 db = client[dbName]
90
91 # Crea la colección accounts in la base de datos para los detalles de las cuentas de twitter
92 accounts = db[dbCollectionA]
93
```

En la imagen anterior se puede apreciar que el código en dbStringConnection ha sido editado para conectar al Cluster que se ha creado previamente en MongoDB Atlas llamado clusterviu, también el código para el nombre de la base de datos se ha editado a '01MBIDJuanDuenez' y el nombre de las colecciones es: 1. dbCollectionA= 'TwitterActCuentas' 2. dbCollectionT='TweetsAct'

Añadir un pequeño procesamiento en cada tweet, Agregarle la Fecha y hora y que usuario lo ha hecho.

En la siguiente imagen extraída del script de python el código ha sido modificado para agregar estos dos parámetros mediante la librería time, la variable `now = datetime.now()` nos da el momento hora y fecha cuando se está realizando el procesamiento este debe ser agregado al diccionario donde los tweets han sido cargados usando el método `update` se puede agregar estos dos nuevos parámetros de la siguiente manera:

```
Loaded_entry.update({"Fecha y hora": now, 'Usuario': 'Juan Duenez'})
```

```
##### Escribimos los datos en MongoDB -- iteramos sobre cada tweet
for entry in d:

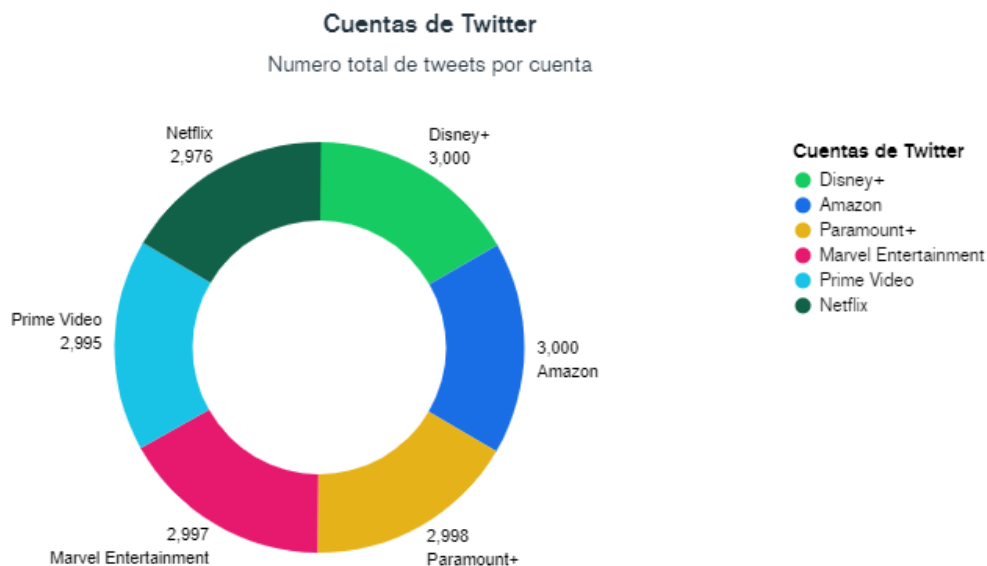
    #Convertimos los datos de twitter para insertarlos en Mongo
    t = json.dumps(entry)

    #print 'type(t)', type(t)                                #<type 'str'>
    loaded_entry = json.loads(t)
    now = datetime.now()
    loaded_entry.update({'Fecha y hora': now, 'Usuario': 'Juan Duenez'})
    #print type(loaded_entry) , loaded_entry                #<type 'dict'>

    #Insertamos el tweet en la base de datos -- A menos que este ya existe
    try:
        tweets.insert_one(loaded_entry)
        # current date and time
```

Extracción de las cuentas elegidas los últimos Tweets, hasta el máximo permitido para cada una de ellas.

La base de Datos ha cargado 17966 Tweets en total y en la siguiente grafica podemos apreciar el máximo permitido por cada una de ellas



Consultas en MongoDB usando la plataforma de MongoDB Atlas en la sección de Aggregation se pueden realizar las consultas necesarias.

- a) Numero Total de Tweets: mediante el método count, cuenta todos los valores en la BD

```
1 [{
2   $count: 'Total'
3 }]
```

Total: 17966

- b) Número Total de Tweets de cada cuenta: usando el método group e identificando el id con el user.name la suma da el valor total de tweets por cuenta

```
1 [{
2   $group: {
3     _id: "$user.name",
4     Total: {
5       $sum: 1
6     }
7   }
8 }]
```

\_id: "Marvel Entertainment" Total: 2997  
 \_id: "Netflix" Total: 2976  
 \_id: "Amazon" Total: 3000  
 \_id: "Prime Video" Total: 2995  
 \_id: "Paramount+" Total: 2998  
 \_id: "Disney+" Total: 3000

- c.) El ranking de los primeros cinco idiomas en los que se ha escrito el tweet: nuevamente se usa el método group pero ahora por lenguaje luego se ordena mediante sort y al ingresar el valor de -1 ordena de mayor a menor luego el método limit para limitar los primeros 5 valores

```
1 [{
2   $group: {
3     _id: "$lang",
4     Top: {
5       $sum: 1
6     }
7   }
8 }, {
9   $sort: {
10    Top: -1
11  }
12 }, {
13   $limit: 5
14 }]
```

Total: 17075 \_id: "und"  
 \_id: "en" Total: 729  
 \_id: "es" Total: 40  
 \_id: "fr" Total: 22  
 \_id: "in" Total: 19

- d.) Ordenar las cuentas de mayor influencia a menor: usando el método group se agrupa por user.name y se busca el valor de los seguidores que se encuentra en user.followers\_count y el orden seria Netflix, MarvelEntertainment, Amazon, Disney+, PrimeVideo y Paramount+

```
1 [{
2   $group: {
3     _id: "$user.name",
4     Total: {
5       $max: "$user.followers_count"
6     }
7   }
8 }]
```

\_id: "Netflix" Total: 13770281  
 \_id: "Marvel Entertainment" Total: 12601966  
 \_id: "Amazon" Total: 3986450  
 \_id: "Disney+" Total: 2689357  
 \_id: "Prime Video" Total: 2090253  
 \_id: "Paramount+" Total: 172921

e.) Realizar una consulta para listar los 20 hashtags más utilizados: primero se usa el método unwind en entities.hashtags ya que es un diccionario, luego con el método Project seleccionamos el texto y de la entidad hashtag luego lo ordenamos de mayor a menor y se mediante el método grupo se hace la sumatoria se ordena de mayor a menor y luego se limita a los primeros 20

My Pipeline:

```
1 ▾ [{
2   $unwind: {
3     path: "$entities.hashtags",
4   }
5 }, {
6 }, {
7   $project: {
8     id: 1,
9     text: 1,
10    "entities.hashtags": 1
11  }, {
12 }, {
13   $sort: {
14     id: 1
15   }, {
16 }, {
17   $group: {
18     _id: "$entities.hashtags.text",
19     Total: {
20       $sum: 1
21     }
22   }, {
23 }, {
24   $sort: {
25     Total: -1
26   }, {
27 }, {
28   $limit: 20
29 }]
```

My Pipeline:

```
12 ▾ }, {
13   $sort: {
14     id: 1
15   }, {
16 }, {
17   $group: {
18     _id: "$entities.hashtags.text",
19     Total: {
20       $sum: 1
21     }
22   }, {
23 }, {
24   $sort: {
25     Total: -1
26   }, {
27 }, {
28   $limit: 20
29 }]
```

A continuación se muestran algunos de los Hashtags

<code>_id: "BundleUpForASweepstakes"</code> Total: 610	<code>_id: "ParamountPlus"</code> Total: 296	<code>_id: "Allstars6"</code> Total: 275
<code>_id: "DisneyPlus"</code> Total: 263	<code>_id: "RuOClock"</code> Total: 240	<code>_id: "iCarly"</code> Total: 240

### 3.) Explorar las funcionalidades de MongoDB Charts

#### Cuentas de Twitter Asociadas a la BD

Seleccione cuentas que tiene servicios de streaming

Nombre de las Cuentas
'Disneyplus'
'Marvel'
'Netflix'
'PrimeVideo'
'amazon'
'paramountplus'

#### Tweets Base de Datos

El numero total de tweets cargado a la Base de Datos

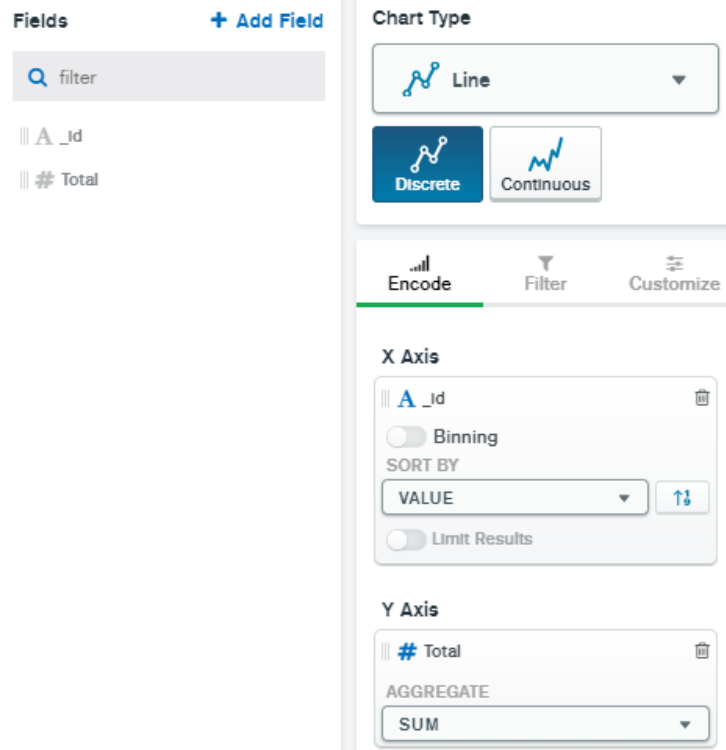
17,966

El tipo de grafico usado en los dos gráficos anteriores ha sido texto, y se ha podido crear una tabla con las cuentas que han sido seleccionadas para usar los tweets de ellas y la otra grafica es el número total de tweets que se han subido a la base de Datos.

El siguiente grafico es de tipo Lineal y muestra los 20 Hashtags más usados en la base de Datos en este grafico en particular se usó el código para encontrar los Hashtags más usados y MongoDB atlas tiene una función que permite ingresar el código y genera los campos que serán usadas para crearlo.

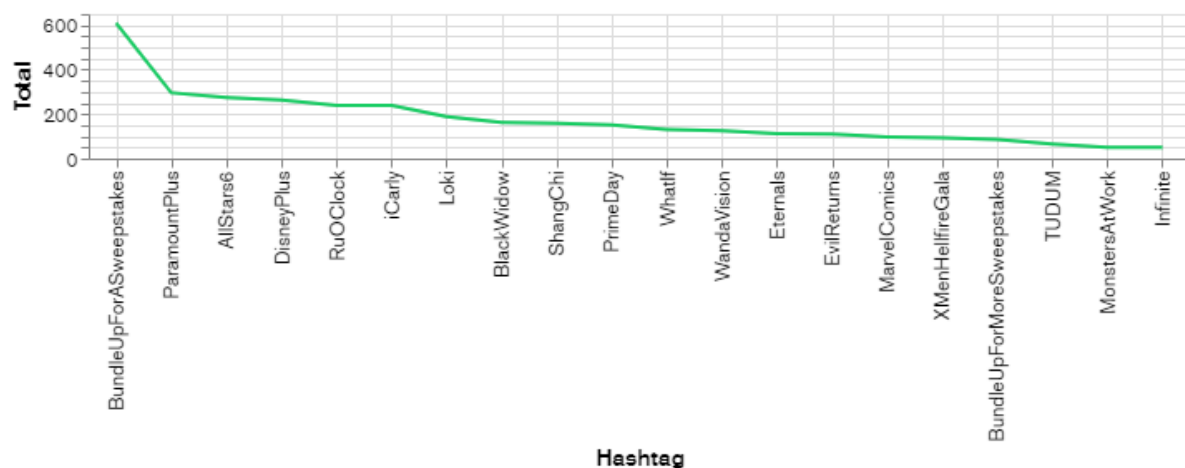
Query 

```
[{$unwind: { path: "$entities.hashtags", }}, {$project: { id:1, text:1, "entities.hashtags":1}},
{$sort: { id:1}}, {$group: { _id: "$entities.hashtags.text", Total: { $sum : 1 } }}, {$sort: { Total: -1}}]
```

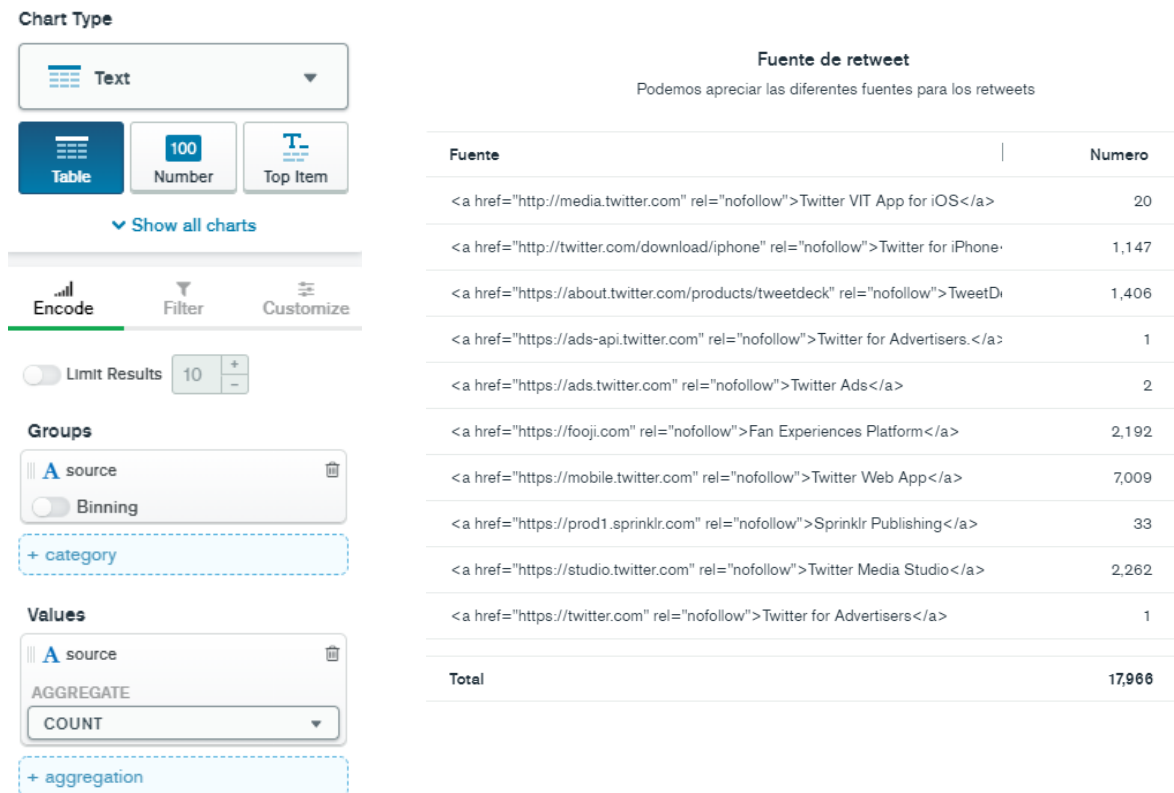


**Grafica Lineal de los 20 Hashtags mas usados**

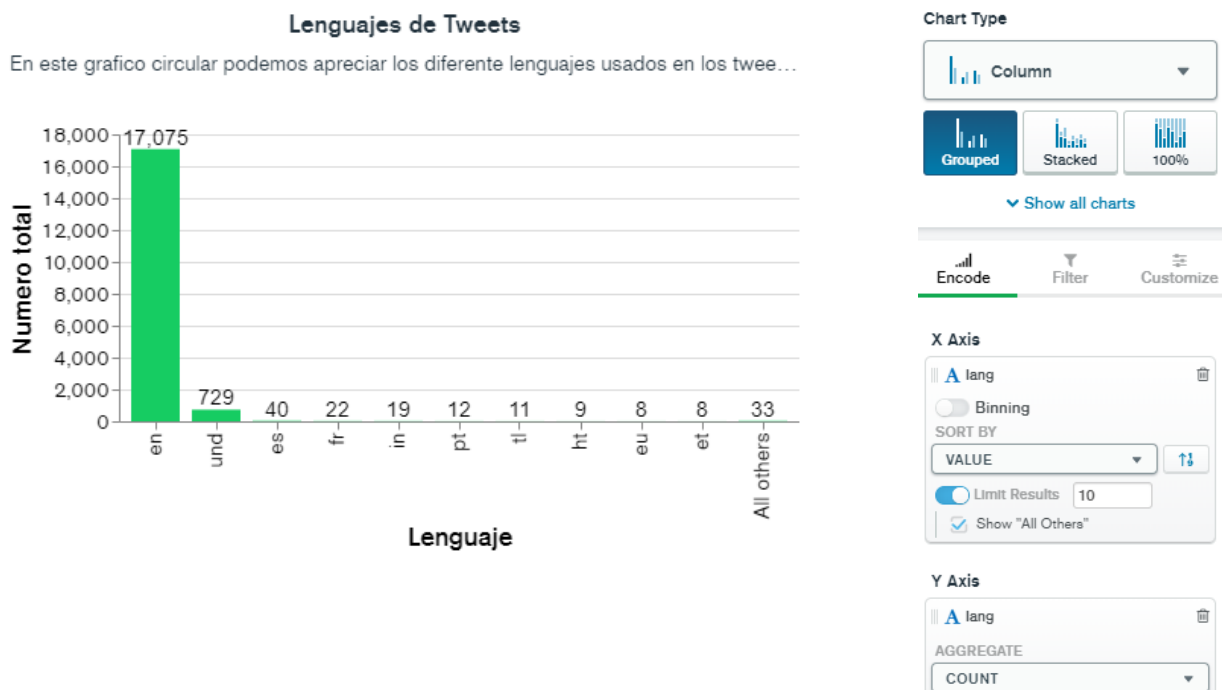
En la grafica podemos apreciar el valor de cada Hashtag que ha sido mas usado en nuestra base de Datos



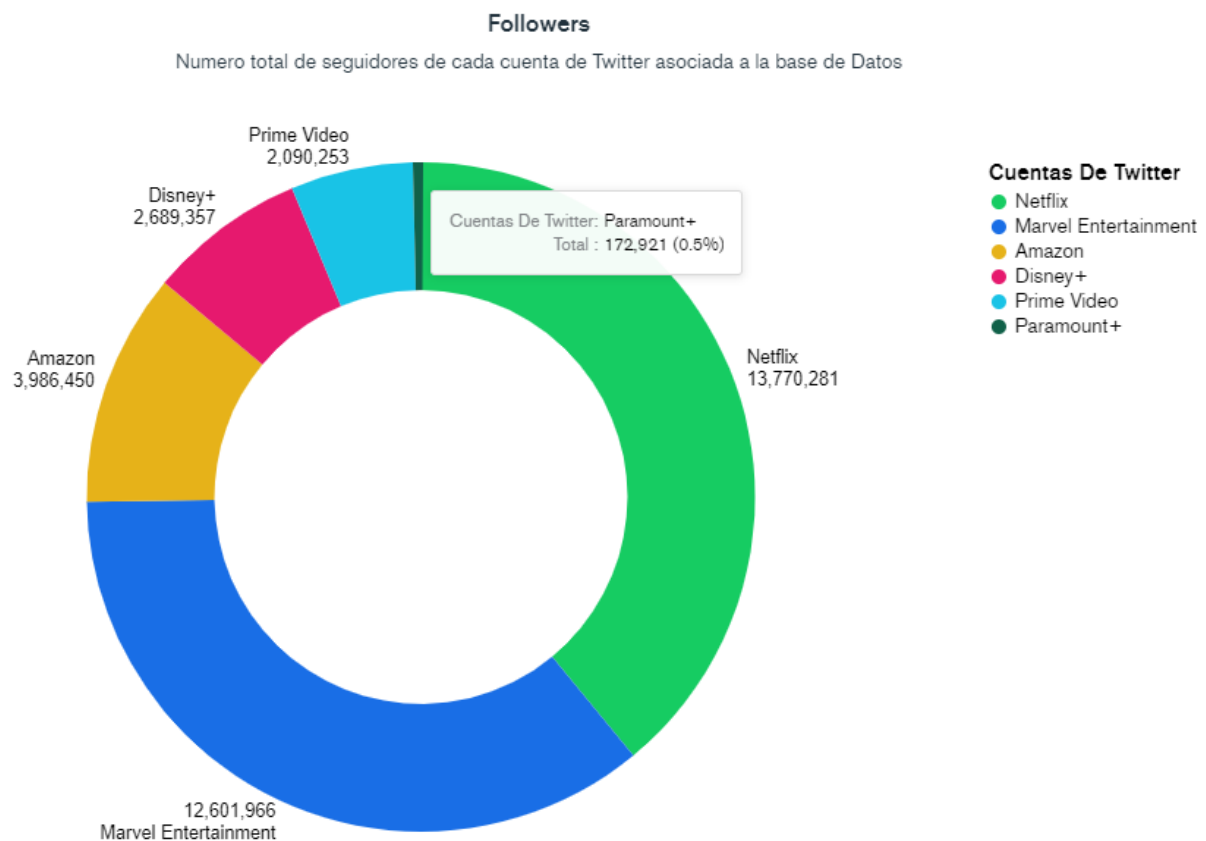
El siguiente grafico es un gráfico de tipo Texto los campos usados han sido source en las siguientes imágenes se puede apreciar la configuración y la tabla generada para el grafico



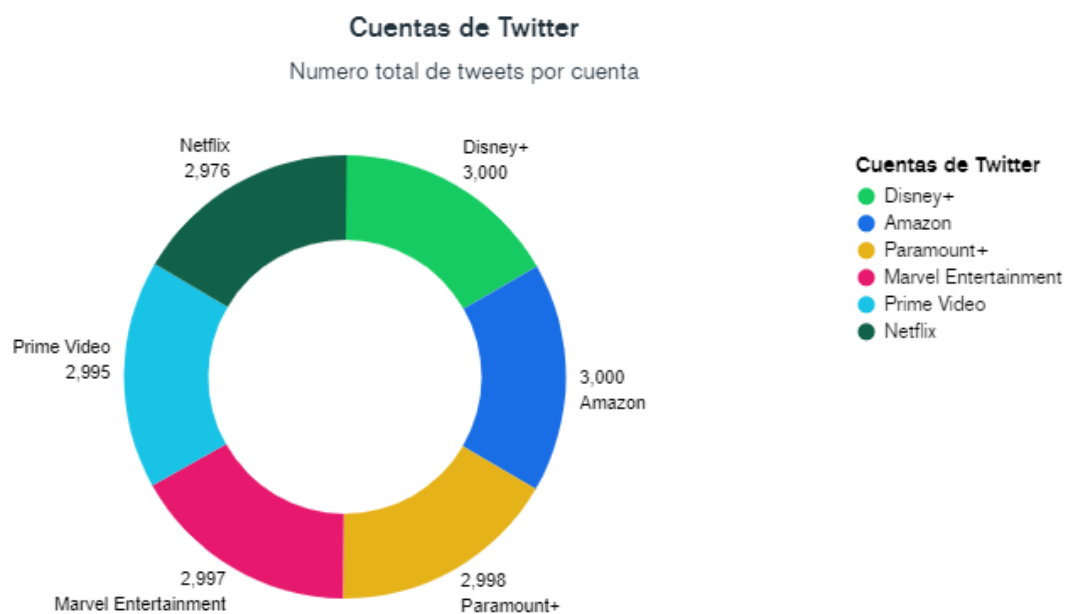
El siguiente Grafico de columnas muestra los lenguajes más usados en los tweets nueva



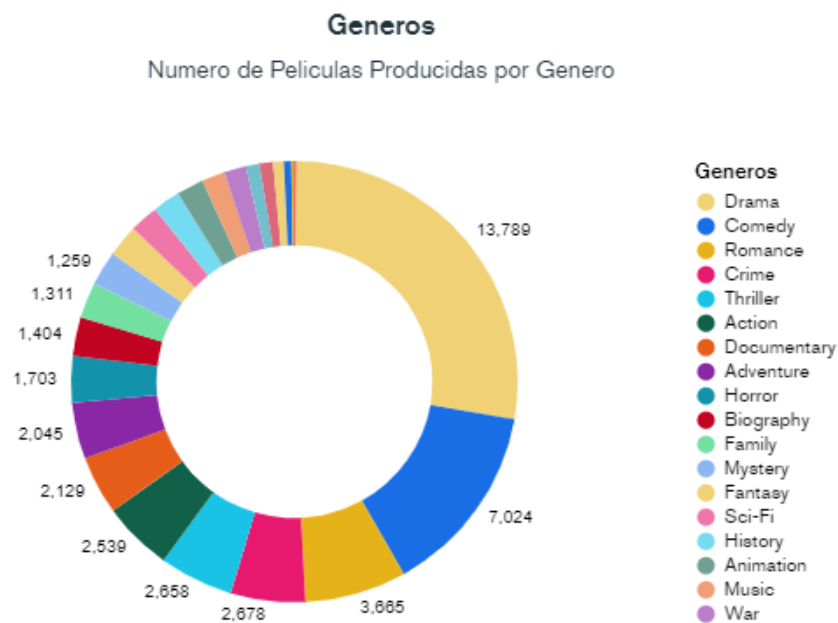
Nuestros siguientes tres gráficos son de tipo circular: primer grafico número de seguidores por cuenta



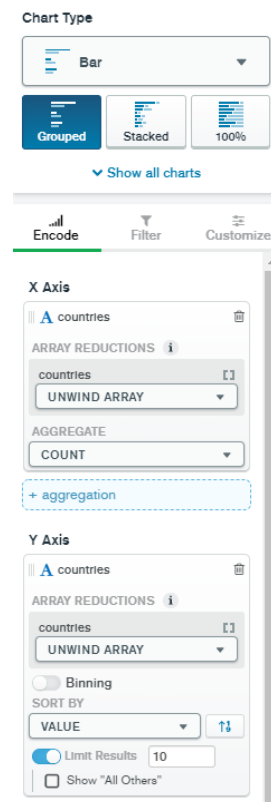
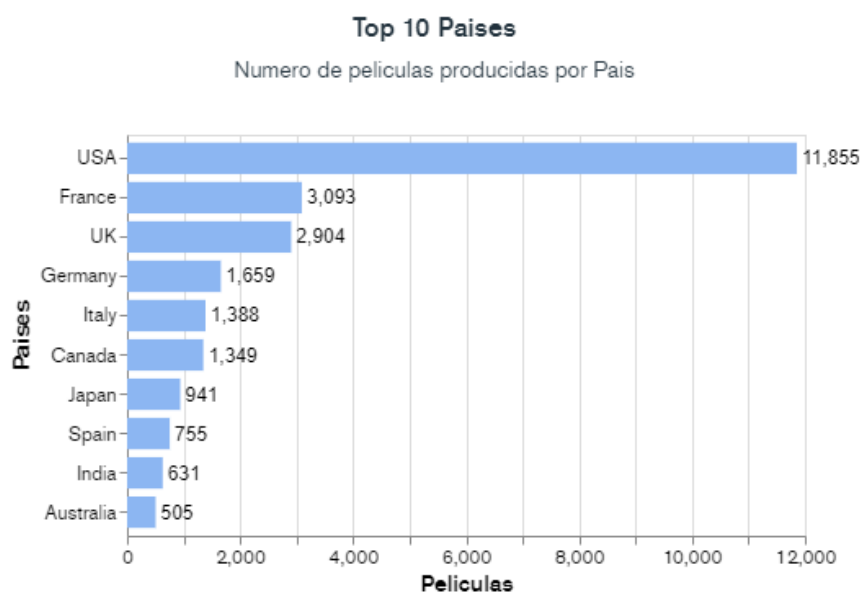
Segundo grafico número total de tweets subidos a la base de datos



El tercer grafico es un gráfico generado con la base de datos que tiene como por ejemplo MongoDB atlas llamada movie y se puede apreciar los diferentes géneros y las películas que se han producido en cada género



Los siguientes dos gráficos son de tipo barra y tienen la siguiente información: el primero el número de películas producidas por País





El segundo grafico es los 10 primeros idiomas en los cuales se han producido las películas

