



PROYECTO FINAL

ANALISIS DE DATOS

NOMBRES:

 CRISTIAN PAREDES

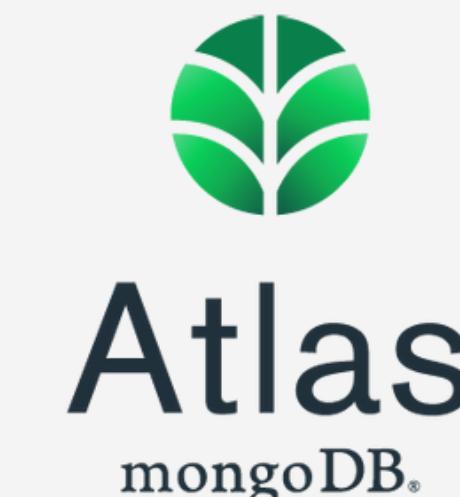
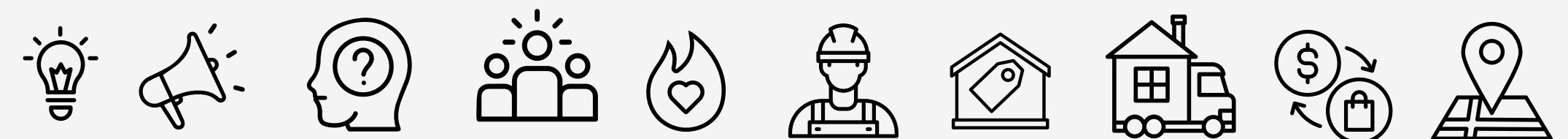
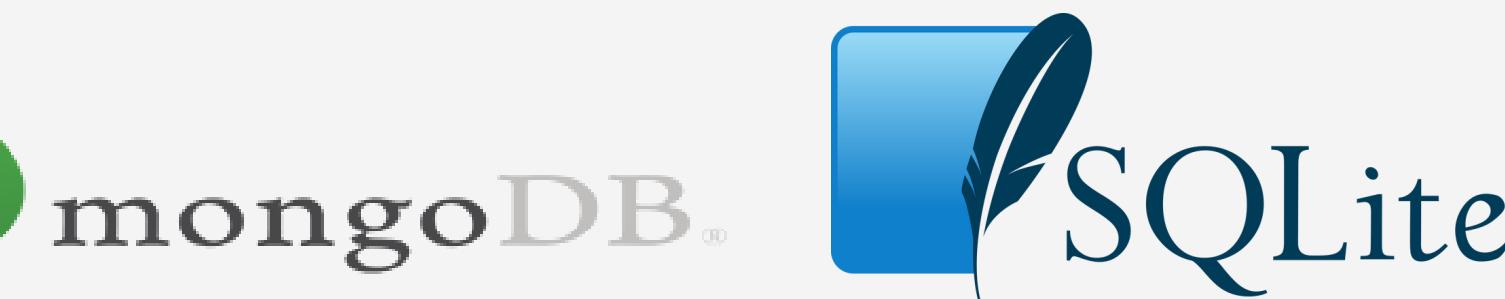
 MELANI BARRERA

 JHORDY AGUAS

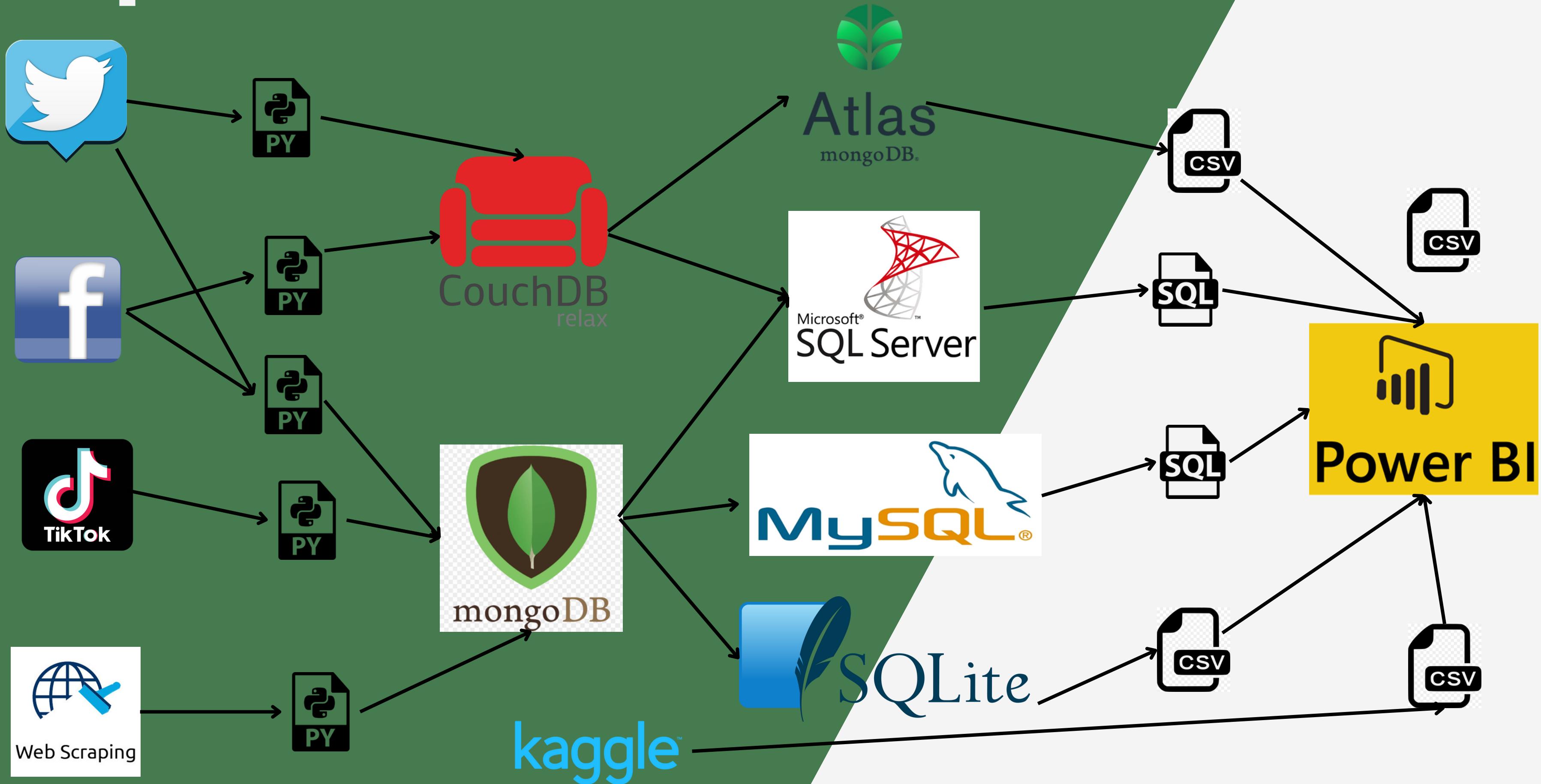
Fuentes de extraccion de datos



Base de Datos



Arquitectura





DEFINICION DE LOS CASOS DE ESTUDIO

01

Pulso político en 20
ciudades principales de
Ecuador

02

Pulso político por
provincias en
Ecuador

03

Eventos o noticias
mundiales.

05

Juegos en línea
por países

04

Tema definido por
el estudiante

CÓDIGO PARA LA LIMPIEZA/FILTRACIÓN DE DATOS

Función para limpiar los caracteres no deseados de una cadena y convertirlo de csv a json def limpiar_cadena(cadena):

Eliminamos las comas, puntos y comas, y espacios vacíos

```
cadena_limpia = "join(caracter for caracter in cadena if caracter not in [',', '.', ' '])  
return cadena_limpia
```

Función para convertir un archivo CSV a JSON con la limpieza realizada previamente def csv_to_json(csvFilePath, jsonFilePath):

```
jsonArray = []  
  
csvFilePath = r'accidentes.csv'  
  
jsonFilePath = r'accidentes.json'  
  
csv_to_json(csvFilePath, jsonFilePath)  
  
with open(csvFilePath, encoding='utf-8') as csvf:
```

```
def json_to_csv(jsonFilePath, csvFilePath):  
  
    jsonFilePath = 'eventslimpio.json'  
  
    csvFilePath = 'events3.csv'  
  
    json_to_csv(jsonFilePath, csvFilePath)  
  
    jsonData = []  
  
    # Cargar el archivo JSON  
  
    with open(jsonFilePath, 'r', encoding='utf-8') as jsonf:  
  
        jsonData = json.load(jsonf)  
  
        # Crear una lista de encabezados  
  
        if jsonData:  
  
            headers = list(jsonData[0].keys())  
  
            # Abrir el archivo CSV y escribir los encabezados  
  
            with open(csvFilePath, 'w', encoding='utf-8', newline='') as csvf:  
  
                csvWriter = csv.DictWriter(csvf, fieldnames=headers)  
  
                csvWriter.writeheader()  
  
                # Escribir los datos limpios en el archivo CSV  
  
                for data in jsonData:
```

Pulso político en 20 ciudades principales de Ecuador

Recopilación de los datos de Twitter

```
import couchdb
from tweepy import Stream
from tweepy import OAuthHandler
from tweepy.streaming import StreamListener
import json
💡

###API #####
ckey = '#####'
csecret = '#####'
atoken = '1763995873191624704-#####'
asecret = '#####'

✓ 0.0s
```



```
class listener(StreamListener):

    def on_data(self, data):
        dictTweet = json.loads(data)
        try:
            dictTweet["_id"] = str(dictTweet['id'])
            doc = db.save(dictTweet)
            print ("SAVED" + str(doc) + "=>" + str(data))
        except:
            print ("Already exists")
            pass
        return True

    def on_error(self, status):
        print (status)
```

```
auth = OAuthHandler(ckey, csecret)
auth.set_access_token(atoken, asecret)
twitterStream = Stream(auth, listener())

''' =====couchdb ===== '''
server = couchdb.Server('http://Cristian:#####@localhost:5984/')
try:
    db = server.create('politico')
except:
    db = server['politico']
```

← ⌂ ⓘ localhost:5984/_utils/#login

Log In to CouchDB

Enter your username and password.

Cristian

.....

Log In



```

#twitterStream.filter(locations=[ -0.225219,-78.5248])
#CRITERIO DE FILTRACIÓN
#/twitterStream.filter(track=['Ecuador','Quito','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Guayaquil','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Manta','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Babahoyo','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Cuenca','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Loja','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Chone','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Esmeraldas','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Los Ríos','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Tulcán','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Machala','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Ambato','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Tena','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Salinas','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Santo Domingo','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Villamil','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Otavalo','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
#/twitterStream.filter(track=['Ecuador','Latacunga','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])
twitterStream.filter(track=['Ecuador','Guaranda','Pulso Politico','Elecciones','Candidatos','Presidencia','Diputados'])

```

Databases					
	Name	Size	# of Docs	Partitioned	Actions
	_replicator	10.4 KB	14	No	  
	ambato	43.4 MB	15903	No	  
	chone	10.6 MB	3842	No	  
	cuenca	43.4 MB	15903	No	  
	esmeraldas	11.6 MB	4178	No	  
	guaranda	26.9 MB	9809	No	  
	guayaquil	43.4 MB	15903	No	  

SQLEXPRESS (SQL Server 16.0.1000 - Cristian\crist)

Databases

- System Databases
- Database Snapshots
- amigos
- BANCO
- BANCOS
- Biblioteca
- ejemplo
- EXAMEN**
- FACT
- PEDID
- PRUE

Security

Server Objects

Replication

Management

XEvent Profiling

Tasks

- Detach...
- Take Offline
- Bring Online
- Encrypt Columns...
- Azure Data Studio...
- Data Discovery and Classification...
- Reports...
- Shrink...
- Rename...
- Delete...
- Refresh...
- Properties...
- Generate Scripts...
- Extract Data-tier Application...
- Deploy Database to Microsoft Azure SQL Database...
- Export Data-tier Application...
- Upgrade Data-tier Application...
- Import Flat File...
- Import Data...**
- Export Data...

SQLQuery1.sql - CRI...RISTIAN\crist (69)*

```
CREATE DATABASE EXAMEN
```

Base de datos SQL Server

Servidor (1)

CRISTIAN\SQLEXPRESS

Base de datos (opcional)

EXAMEN

Modo Conectividad de datos (1)

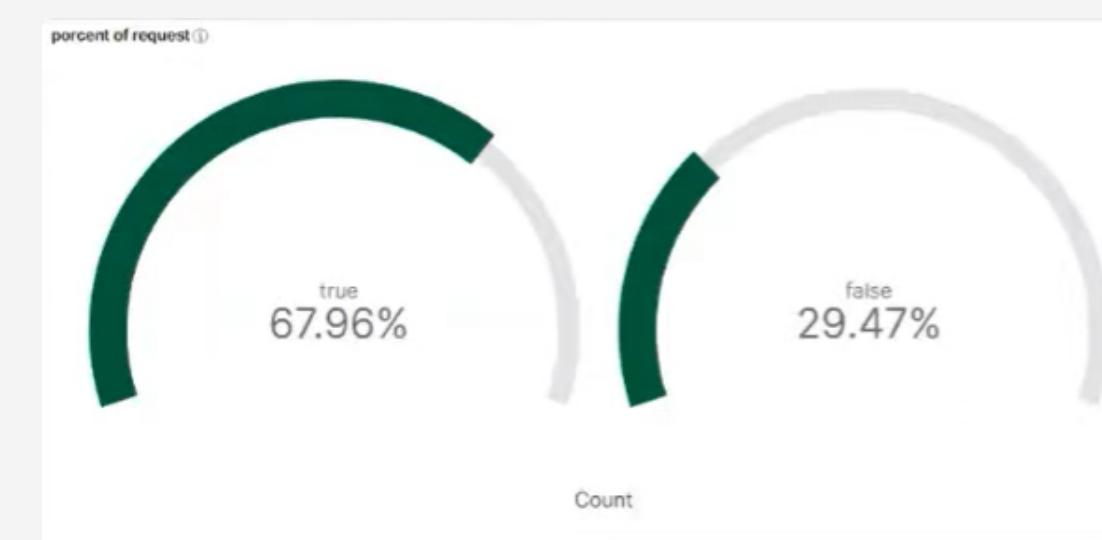
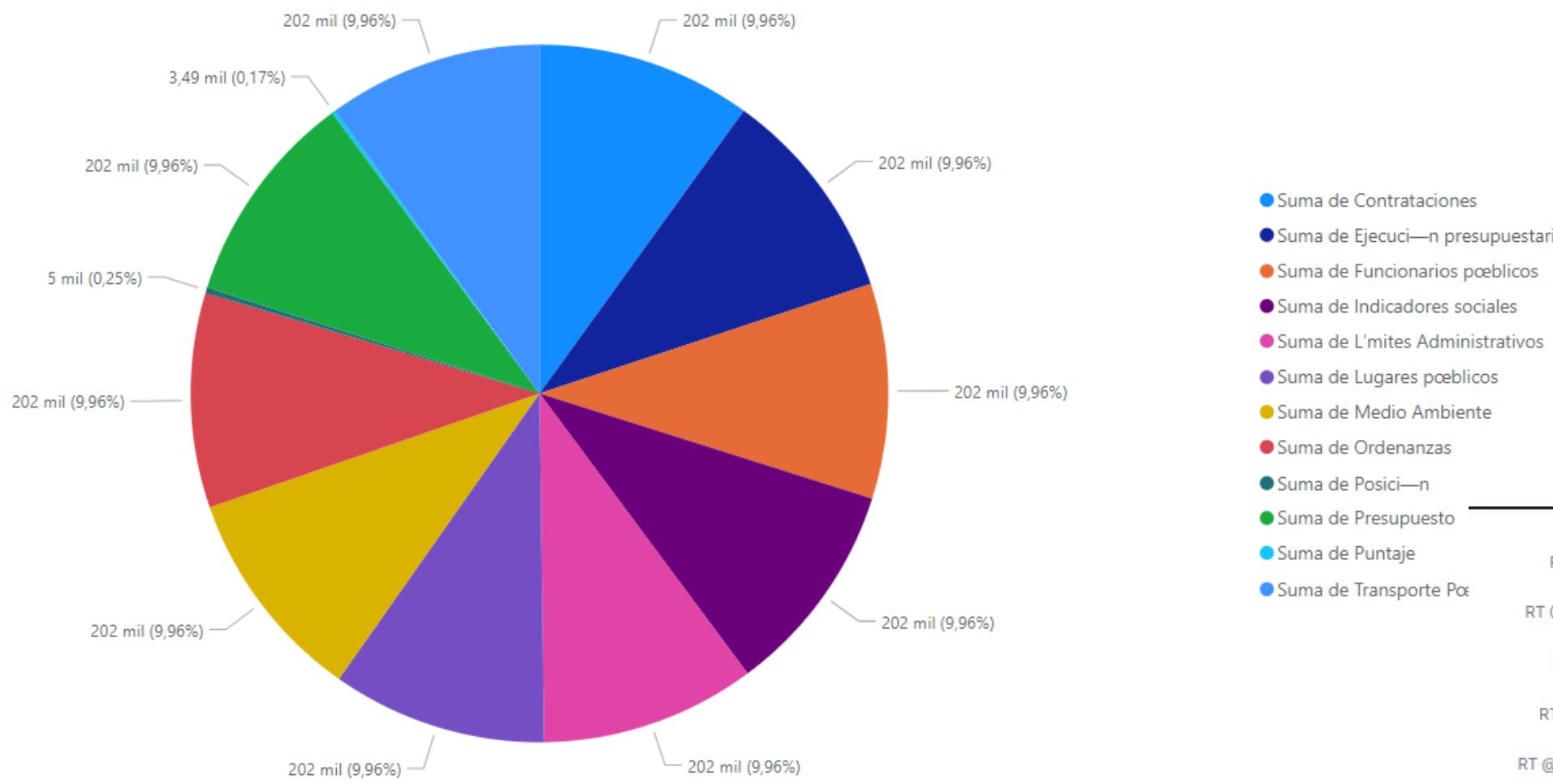
Importar

DirectQuery

► Opciones avanzadas



Suma de Contrataciones, Suma de Ejecuci—n presupuestaria, Suma de Funcionarios pœblicos, Suma de Indicadores sociales, Suma de Lmites Administrativos, ...
Lugares pœblicos, Suma de Medio Ambiente, Suma de Ordenanzas, Suma de Posici—n, Suma de Presupuesto, Suma de Puntaje y Suma de Transporte Pœblico



- RT @btschartdata: "Dynamite" is now #1 in 20 countries #1 Brazil #1 Chile #1 Colombia #1 Costa Rica #1 Ecuador #1 Estonia #1 Finland #1 Ka...
- RT @Oscarqugo: Rafael López Aliaga confesó en @willaxtv que a nombre propio ha invertido 200 mil dólares, pero que le da dinero a sus candi...
- RT @lavecinarubia: Hasta el 24 de abril se puede solicitar el voto por correo para las elecciones de Madrid en este enlace. Es importante q...
- RT @mariobryceofic: Se nota la diferencia con otros candidatos. Es Rafael López Aliaga quién inteligentemente maneja la entrevista con argu...
- RT @julioguzmanperu: Falta un mes para las elecciones y mi esperanza es que el gran porcentaje de peruanos que aún no sabe por quién va a v...
- RT @parodirevoredo: Rafael López Aliaga reconoce en @willaxtv haber realizado donaciones a sus candidatos al congreso, luego se da cuenta q...
- RT @AntonioMautor: Ya sabéis amiguitos/as, hacer elecciones en Catalunya era una locura y una imprudencia temeraria. Hacer elecciones en Ma...
- RT @brozoxmiswebs: El PRD en la Cámara de Diputados presentó una denuncia de hechos ante la FGR en contra de Hugo López-Gatell, por poner e...
- RT @pvallin: URGENTE: Pablo Iglesias disputará la presidencia de Madrid a Isabel Díaz Ayuso. <https://t.co/E3vMvgDNfu>
- RT @SuperRoStar: Si se celebran elecciones en Catalunya, es que ponen en peligro la salud de los ciudadanos. Si se celebran en Madrid, es...
- RT @julioguzmanperu: Hace 5 años me excluyeron y eso cambió el resultado final. En un mes todo puede cambiar. Mi esperanza es que la gente...
- RT @gallifantes: De las elecciones de Madrid espero algunos vean cómo el fascismo preocupa a la izquierda española bastante menos que el i...
- RT @DiaDatv: #EncuestaElecciones2021 Si las elecciones fueran mañana ¿A qué candidato escogería?
- RT @3porciento_3: Las elecciones en Madrid son una irresponsabilidad, lo que sí es muy responsable es cambiar el Ministro de Sanidad en ple...
- RT @PabloEchenique: Ayuso ha convocado elecciones un martes (día laborable) y no ha sido sin intención. Por eso es vital que difundas este...

All docs

Pulso político por provincias en Ecuador

```
File: CouchProv.py
1 import couchdb
2 from tweepy import Stream
3 from tweepy import OAuthHandler
4 from tweepy.streaming import StreamListener
5 import json
6
7
8 #####API SE OCULTA POR SEGURIDAD
9 ckey = "#####"
10 csecret = "#####"
11 atoken = "#####-kyuyyls0BxT7I4m0MlQjS3wUhNvbotn"
12 asecret = "#####"
13 #####
14
15
16 class listener(StreamListener):
17
18     def on_data(self, data):
19         dictTweet = json.loads(data)
20         try:
21
22             dictTweet["_id"] = str(dictTweet['id'])
23             doc = db.save(dictTweet)
24             print ("SAVED" + str(doc) + "=>" + str(data))
25         except:
26             print ("Already exists")
27             pass
28         return True
29
30     def on_error(self, status):
31         print (status)
32
33 auth = OAuthHandler(ckey, csecret)
34 auth.set_access_token(atoken, asecret)
35 twitterStream = Stream(auth, listener())
36
37 #####couchdb#####
38 server = couchdb.Server('http://Cristian:123456@localhost:5984/')
39 try:
40     db = server.create('tweets')
41 
```

```
envioAMongoAtlas.py ×  () bolivaraas.json
envioAMongoAtlas.py > ...
1 import os
2 import json
3 from pymongo import MongoClient
4
5 def import_json_files_to_mongodb(folder_path, db_name):
6     # Conexión a la base de datos MongoDB
7     client = MongoClient("mongodb+srv://fjhordy1:kWngm3uFRzApZvJD@cluster0.rrfo8ne.mongodb.net/?retryWrites=true&w=majority&appName=Cluster0")
8     db = client[db_name]
9
10    # Obtener lista de archivos JSON en la carpeta
11    json_files = [f for f in os.listdir(folder_path) if f.endswith('.json')]
12
13    # Iterar sobre cada archivo JSON
14    for file_name in json_files:
15        # Nombre de la colección será el nombre del archivo sin la extensión .json
16        collection_name = os.path.splitext(file_name)[0]
17        collection = db[collection_name]
18
19        file_path = os.path.join(folder_path, file_name)
20        with open(file_path, 'r') as f:
21            try:
22                # Cargar datos desde el archivo JSON
23                data = json.load(f)
24                if 'rows' in data:
25                    # Insertar cada tweet en la colección de MongoDB
26                    for tweet in data['rows']:
27                        collection.insert_one(tweet['doc'])
28                        print(f"Datos insertados en la colección '{collection_name}' correctamente.")
29                else:
230                    print(f"El archivo '{file_name}' no contiene datos válidos de tweets.")
211            except Exception as e:
222                print(f"Error al cargar datos desde el archivo '{file_name}': {e}")
23
24 if __name__ == "__main__":
25     folder_path = r"D:\2023\Analisis\Proyecto\pulso_politico" # Ruta de la carpeta en Windows
26     db_name = "pulso_politico"
27
28     import_json_files_to_mongodb(folder_path, db_name)
```

DATOS

Datos insertados en la colección 'cañaraaa' correctamente.
Datos insertados en la colección 'chimborazoaa' correctamente.
Datos insertados en la colección 'chimborazogl' correctamente.
Datos insertados en la colección 'cotopaxiaa' correctamente.
Datos insertados en la colección 'cotopaxigl' correctamente.
Datos insertados en la colección 'eloroaa' correctamente.
Datos insertados en la colección 'elorogl' correctamente.
Datos insertados en la colección 'esmeraldasaa' correctamente.
Datos insertados en la colección 'esmeraldasgl' correctamente.

localhost:5984/_utils/#/_all_dbs

Databases

Name	Size	# of Docs	Partitioned	Actions
_replicator	0 bytes	0	No	
_users	2.3 KB	1	No	
azuay	0.6 MB	1	No	
bolivar	76.6 KB	1	No	
cotopaxi	0.6 MB	1	No	
esmeraldas	88.6 KB	1	No	
galapagos	8.2 KB	1	No	
guayaquil	267.0 KB	1		
imbabura	0.9 MB	1		
politico	77.1 KB	1		

MongoDB Compass - local/provincias

Connect Edit View Help

local

- + Create collection
- Refresh
- View
- Sort by Collection Name

Bolívar

Storage size: 4.10 kB Documents: 50 Avg. document size: 56.00 B Indexes: 1 Total index size: 4.10 kB

Cañar

Storage size: 4.10 kB Documents: 49 Avg. document size: 55.00 B Indexes: 1 Total index size: 4.10 kB

Carchi

Storage size: 4.10 kB Documents: 49 Avg. document size: 55.00 B Indexes: 1 Total index size: 4.10 kB

Chimborazo

Storage size: 4.10 kB Documents: 49 Avg. document size: 55.00 B Indexes: 1 Total index size: 4.10 kB

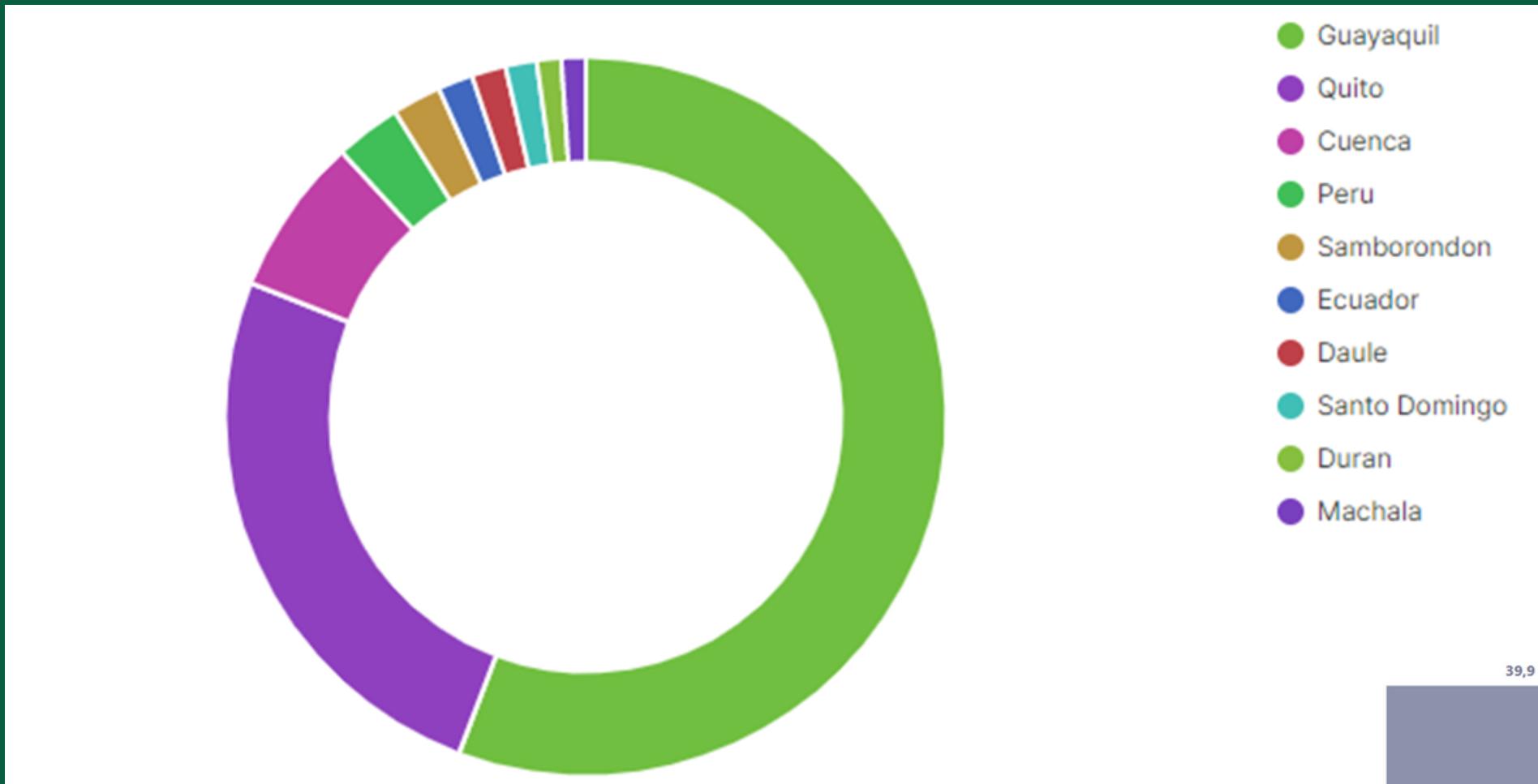
Pichincha

Storage size: 4.10 kB Documents: 50 Avg. document size: 56.00 B Indexes: 1 Total index size: 4.10 kB

Local Databases:

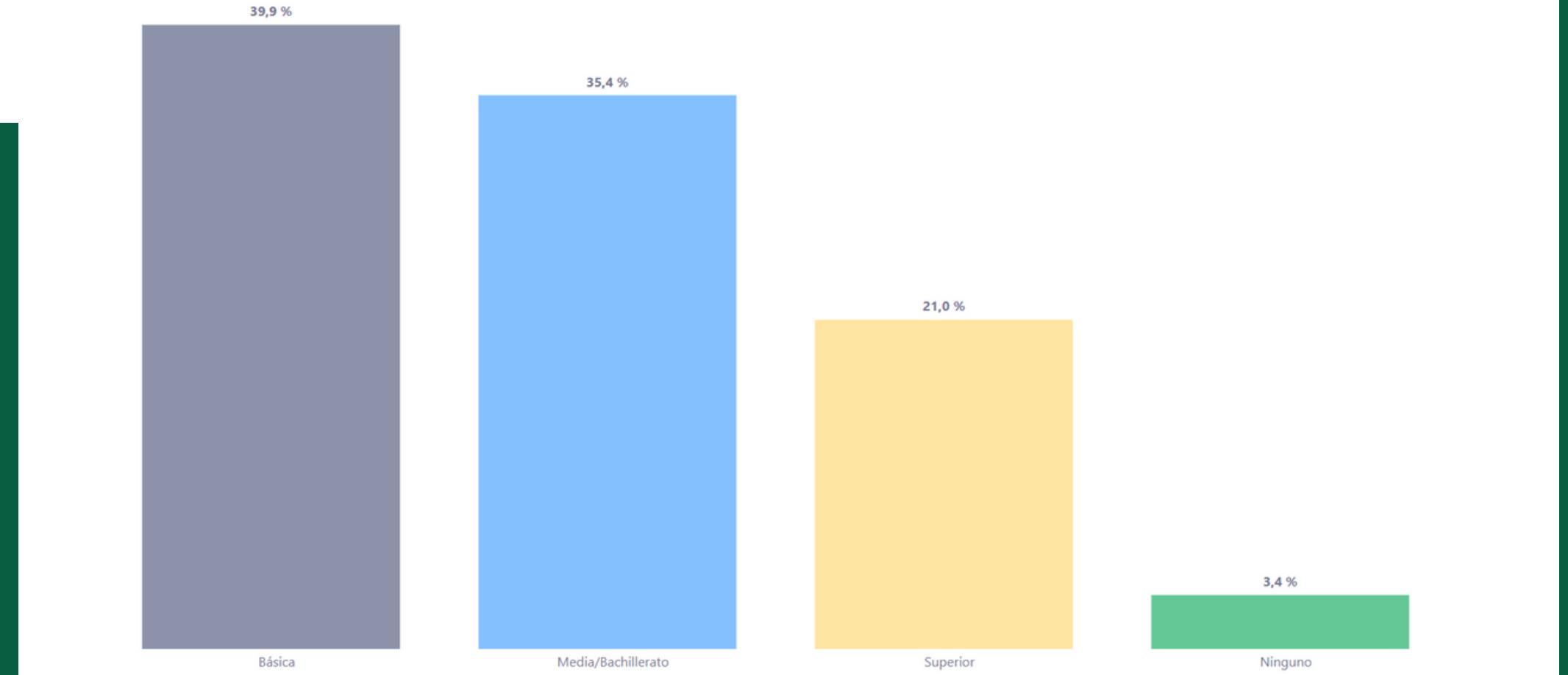
- Netflixt
- admin
- ciudades
 - ambato
 - azuay
 - bolivar
 - guayaquil
 - imbabura
- config
- examen
- local
- provincias
 - Bolívar
 - Carchi
 - Cañar
 - Chimborazo
 - Pichincha

RESULTADOS



**La población de estudio
correspondiente a la: “Formacion
Institucional” del numero de
personas que participan en
eventos politicos**

**Visualización de las provincias
del Ecuador donde se realiza
twists relacionaos con partidos
Politicos**



RESULTADOS



Usados en twitter que los usuarios postearon dentro del ambito politico



Candidatos con mas respuesta del publico



Eventos o noticias mundiales

Uso de las plataformas de kaggle e inec

https://www.ecuadorencifras.gob.ec/vacunacion-covid-19/

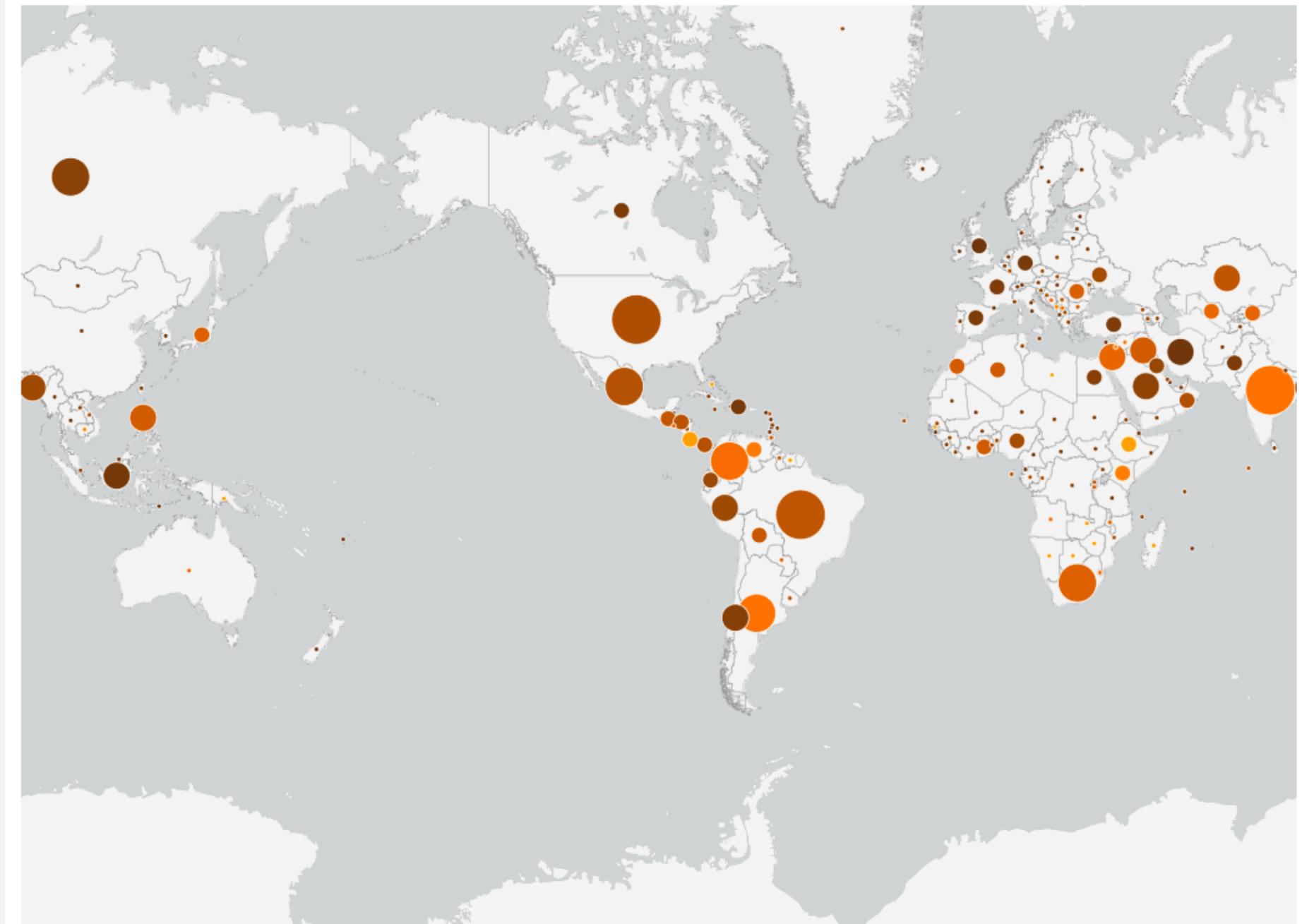
The screenshot shows a web browser displaying the INEC (Instituto Nacional de Estadística y Censos) website. The URL in the address bar is https://www.ecuadorencifras.gob.ec/vacunacion-covid-19/. The page features the INEC logo and a navigation menu with links like 'Estadísticas por tema', 'Estadísticas por fuente', 'Geografía Estadística', 'Banco de Datos', 'Consultas Especializadas', and 'Comunicamos'. A search bar is also present. Below the menu, there's a section titled 'Vacunación – COVID 19' with a sub-section 'Salud' containing links to 'Vacunación – COVID 19', 'Actividad Física y Sedentarismo', 'Actividades y Recursos de Salud', 'Camas y Egresos Hospitalarios', 'Salud, Salud Reproductiva y Nutrición', and 'Tecnologías de la Información y la Comunicación'. To the right of this section is a large image of a medical professional in blue gloves performing a vaccination. On the far right, a SQLite database viewer window titled 'covid_mundial.db' is open, showing a table with data from 1 to 38 countries. The columns include Country/Region, Confirmed cases, Deaths, Recovered, Active cases, New cases, New deaths, New recovered, Deaths / 100 Cases, Recovered / 100 Cases, and Deaths / Recovered.

Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / Recovered
Afghanistan	36263	1269	25198	9796	106	18	3.5	69.49	5.84	
Albania	4880	144	2745	1991	117	6	2.95	56.25	5.25	
Algeria	27973	1163	18837	7973	616	8	4.16	67.34	6.17	
Andorra	987	52	803	52	10	0	5.73	88.53	6.48	
Angola	950	41	242	667	18	1	4.32	25.47	16.94	
Antigua and Barbuda	86	3	65	18	4	0	3.49	75.58	4.62	
Argentina	167416	3859	72575	91782	4890	120	1.83	43.35	4.21	
Armenia	37390	711	26665	10014	73	6	1.9	71.32	2.67	
Australia	15303	167	9311	5825	368	6	1.09	60.84	1.79	
Austria	20558	713	18246	1599	86	1	3.47	88.75	3.91	
Azerbaijan	30446	423	23242	6781	396	6	1.39	76.34	1.82	
Bahamas	382	11	91	280	40	0	2.88	23.82	12.09	
Bahrain	39482	141	36110	3231	351	1	421	91.46	0.39	
Bangladesh	226225	2965	125683	97577	2772	37	1801	1.31	55.56	2.36
Barbados	110	7	94	9	0	0	6.36	85.45	7.45	
Belarus	67251	538	60492	6221	119	4	67	0.8	89.95	0.89
Belgium	66428	9822	17452	39154	402	1	14	14.79	26.27	56.28
Belize	48	2	26	28	0	0	0	4.17	54.17	7.69
Benin	1770	35	1036	699	0	0	0	1.98	58.53	3.38
Bhutan	99	0	86	13	4	0	1	0.0	86.87	0.0
Bolivia	71181	2647	21478	47056	1752	64	309	3.72	30.17	12.32
Bosnia and Herzegovina	10498	294	4930	5274	731	14	375	2.8	46.96	5.96
Botswana	739	2	63	674	53	1	11	0.27	8.53	3.17
Brazil	2442375	87618	1846641	508116	23284	614	33728	3.59	75.61	4.74
Brunei	141	3	138	0	0	0	0	2.13	97.87	2.17
Bulgaria	10621	347	5585	4689	194	7	230	3.27	52.58	6.21
Burkina Faso	1100	53	926	121	14	0	6	4.82	84.18	5.72
Burma	350	6	292	52	0	0	2	1.71	83.43	2.85
Burundi	378	1	301	76	17	0	22	0.26	79.63	0.33
Cabo Verde	2328	22	1550	756	21	0	103	0.95	66.58	1.42
Cambodia	226	0	147	79	1	0	4	0.0	65.04	0.0
Cameroon	17110	391	14539	2180	402	6	0	2.29	84.97	2.69
Canada	116458	8944	0	107514	682	11	0	7.68	0.0	Infinity
Central African Republic	4599	59	1546	2994	0	0	0	1.28	33.62	3.82
Chad	922	75	810	37	7	0	0	8.13	87.85	9.26
Chile	347923	9187	319954	18782	2133	75	1859	2.64	91.96	2.87
China	86783	4656	78869	3258	213	4	7	5.37	90.88	5.9
Colombia	257101	8777	131161	117163	16306	508	11494	3.41	51.02	6.69

RESULTADOS

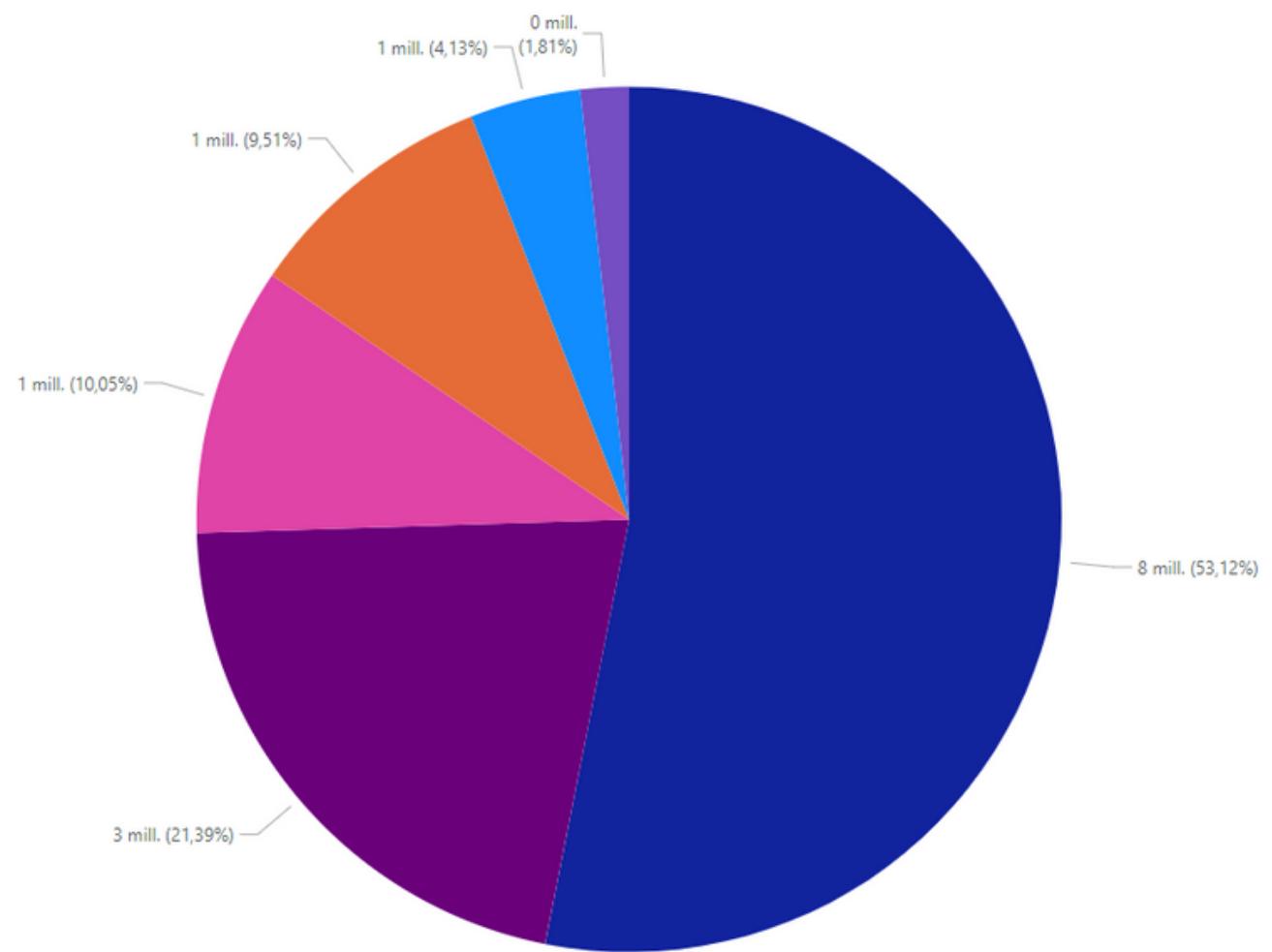
Suma de ciudad 11 mil M	Suma de conglomerado 24 mil M	Suma de hogar 92 mil
Suma de area 116 mil	Suma de panelm 2 mill.	Suma de vivienda 415 mil

Contaminacion de la poblacion en ecuador durante el periodo de covid 19



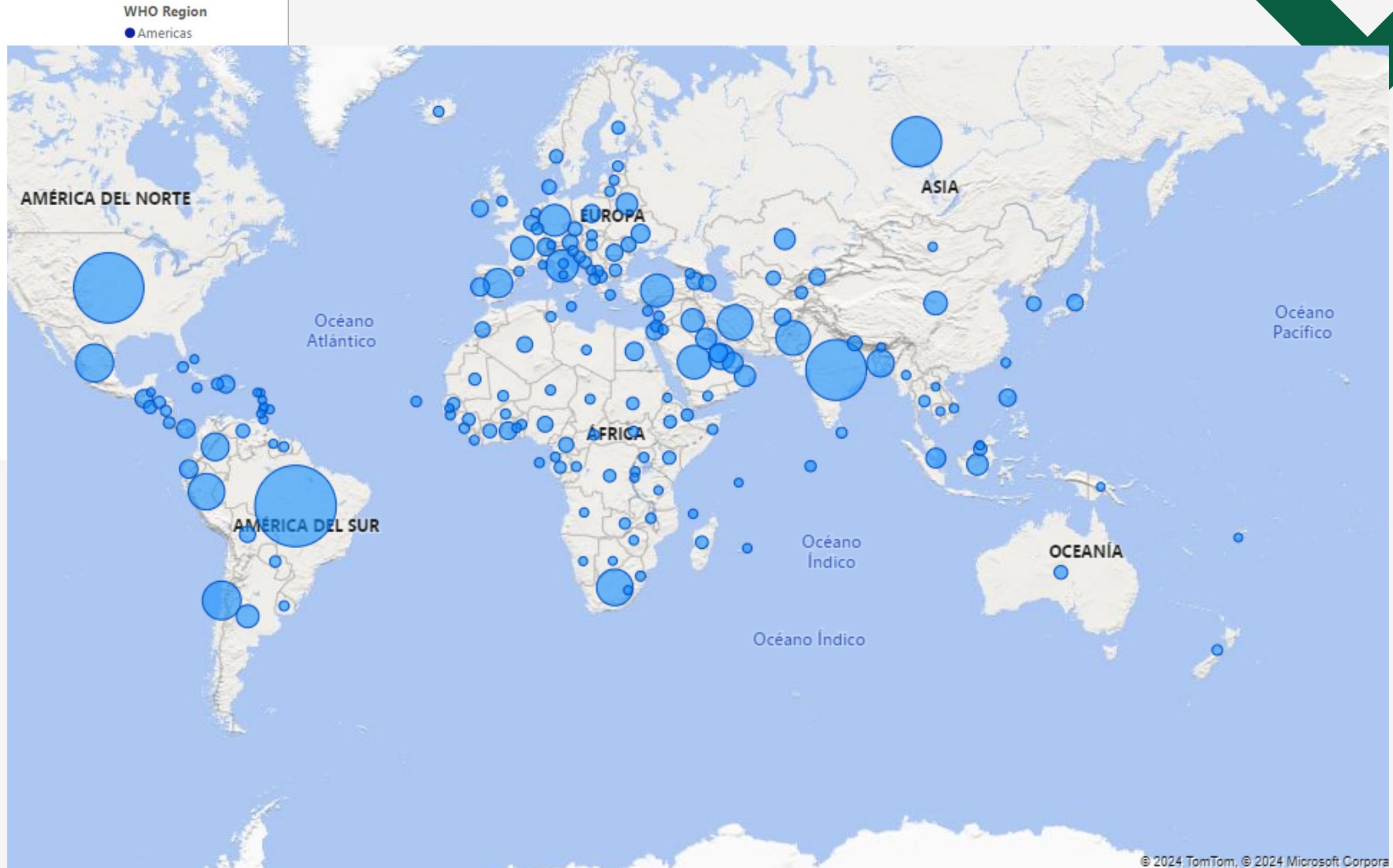
Numero de muertes confirmadas dentro del periodo de contaminacion de covid a nivel mundial

RESULTADOS



**Suma de contaminación por semana
de los continentes en periodo de
COVID**

**Datos de los nuevos casos registrados
de infección de la población mundial**



Juegos en línea por países



The image shows a desktop interface with three main windows:

- MongoDB Compass:** A window titled "videojuegos.juegos" showing a collection with 0 documents and 1 index. It has a sidebar with databases like "cluster0.rrfo8ne..." and "videojuegos". A modal dialog titled "Import" is open, showing the file "vgsales_final.csv" selected for import into the "videojuegos.juegos" collection.
- Microsoft Excel:** A window showing a file named "vgsales_final.csv" with a size of 1,323 KB. The ribbon tabs include Inicio, Insertar, Modelado, Ver, Optimizar, and Ayuda. The "Datos" tab is selected, showing options like "Obtener Libro de Centro de datos de OneLake", "Excel", "SQL Server", and "Introducir DataView".
- Power BI Data Explorer:** A window titled "Obtener datos" showing a list of data sources. The "Base de datos" section is expanded, listing various database types: SQL Server, Access, SQL Server Analysis Services, Oracle Database, IBM Db2, Informix, Netezza, MySQL, PostgreSQL, Sybase, Teradata, SAP HANA, SAP Business Warehouse, SAP Business Warehouse messaging, Amazon Redshift, and Impala. The "Conectar" button is at the bottom right.

Sin título - Power BI Desktop

Inicio Insertar Modelado Ver Optimizar Ayuda

Cortar Copiar Copiar formato Portapapeles

Obtener Libro de Centro de datos de SQL Server OneLake Introducir Dataverse Origenes recientes Consultas Datos Transformar Actualizar datos Nuevo objeto visual Cuadro de texto más objetos visuales Insertar Cálculos Confidencialidad Publicar Compartir

Base de datos MySQL

b9803rmmh50qgoti14mu-mysql.services.clever-cloud.com

Nombre de usuario: un4uy9xmv6mykylr
Contraseña:
Seleccionar en qué nivel hay que aplicar esta configuración: b9803rmmh50qgoti14mu-mysql.services.clever-cloud.com

Atrás Conectar Cancelar

Importar datos

Obtener datos de otro origen →

Visualizaciones

Filtros

Obtener datos

Buscar

Base de datos

Todo

Archivo

Base de datos

Microsoft Fabric

Power Platform

Azure

Servicios en línea

Otras

Cubos de AtScale

BI Connector

Data Virtuality LDW

Denodo

Dremio Software

Dremio Cloud

Exasol

Indexima

InterSystems IRIS (beta)

Jethro (beta)

Kyligence

Bases de datos PICK / MultiValue con Linkar (beta)

MariaDB

MarkLogic

MongoDB Atlas SQL (beta)

TIBCO(R) Data Virtualization

Conectores certificados Aplicaciones de plantilla Conectar Cancelar

The image shows a split-screen view of the Power BI Desktop application. On the left, a modal dialog box titled 'Base de datos MySQL' is displayed, prompting for connection details: 'Nombre de usuario' (User name) set to 'un4uy9xmv6mykylr', 'Contraseña' (Password) masked, and 'Seleccionar en qué nivel hay que aplicar esta configuración' (Select at which level this configuration should be applied) set to 'b9803rmmh50qgoti14mu-mysql.services.clever-cloud.com'. On the right, the main Power BI interface shows the 'Obtener datos' (Get Data) dialog box with the 'Base de datos' (Database) category selected. A list of available data sources is visible, including AtScale, BI Connector, Data Virtuality LDW, Denodo, Dremio Software, Dremio Cloud, Exasol, Indexima, InterSystems IRIS (beta), Jethro (beta), Kyligence, Bases de datos PICK / MultiValue con Linkar (beta), MariaDB, MarkLogic, MongoDB Atlas SQL (beta), and TIBCO(R) Data Virtualization. The 'MongoDB Atlas SQL (beta)' option is currently highlighted.

Database Deployments

Find a database deployment... Edit Config + Create

Cluster0 Connect View Monitoring Browse Collections ...

Visualize Your Data FREE SHARED

R: 0 W: 0 Last 2 hours

Connections: 10.0 Last 2 hours

In: 22.3 KB/s Out: 1.1 KB/s Last 2 hours

Data Size: 959.8 KB / 512.0 MB (0%) 512.0 MB

Dismiss Explore Charts

VERSION REGION CLUSTER TIER TYPE BACKUPS LINKED APP SERVICES ATLAS SQL ATLAS SEARCH

6.0.14 AWS / Sao Paulo (sa-east-1) M0 Sandbox (General) Replica Set - 3 nodes Inactive Application-0 Connect Create Index

+ Add Tag

MongoDB Atlas SQL

MongoDB URI: `b8-enxol.a.query.mongodb.net/pulso_politico?ssl=true&authSource=admin`

Database: `pulso_politico`

Aceptar Cancelar

ODBC Data Source Administrator (64-bit)

User DSN System DSN File DSN Drivers Tracing Connection Pooling About

User Data Sources:

Name	Platform	Driver
Excel Files	64-bit	Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlbs)
mongodb	64-bit	MongoDB Atlas SQL ODBC Driver
MS Access Database	64-bit	Microsoft Access Driver (*.mdb, *.accdb)

Add... Remove Configure...

An ODBC User data source stores information about how to connect to the indicated data provider. A User data source is only visible to you and can only be used on this computer.

OK Cancel Apply Help



Navegador

Opciones de presentación

- mongodb://atlas-sql-6584fe006496ae0cfa4e51...
- análisis
- examen
- pulso_político
- test
- videojuegos [1]
- juegos

juegos

1	2006	29	3	41	8
3	2008	{"02":"12.88"}	{"77":"3.79"}	{"49":"15.85"}	{"4
4	2009	{"02":"11.01"}	{"77":"3.28"}	{"49":"15.75"}	{"4
6	1989	{"02":"2.26"}	{"77":"4.22"}	{"49":"23.2"}	{"4
12	2005	{"02":"7.57"}	{"77":"4.13"}	{"49":"9.81"}	{"4
19	1990	{"02":"3.75"}	{"77":"3.54"}	{"49":"12.78"}	{"4
20	2005	{"02":"9.26"}	{"77":"4.16"}	{"49":"4.75"}	{"4
25	2002	{"02":"5.49"}	{"77":"0.47"}	{"49":"8.41"}	{"4
33	2013	{"02":"4.05"}	{"77":"4.34"}	{"49":"5.17"}	{"4
34	2015	{"02":"5.81"}	{"77":"0.35"}	{"49":"5.77"}	{"4
38	2011	{"02":"5.82"}	{"77":"0.49"}	{"49":"5.54"}	{"4
40	2008	{"02":"2.61"}	{"77":"2.66"}	{"49":"6.75"}	{"4
41	2010	{"02":"4.44"}	{"77":"0.48"}	{"49":"5.98"}	{"4
46	2009	{"02":"2.77"}	{"77":"3.96"}	{"49":"4.4"}	{"4
49	2007	{"02":"3.4"}	{"77":"1.2"}	{"49":"6.16"}	{"4
59	2004	{"02":"2.65"}	{"77":"3.15"}	{"49":"4.34"}	{"4
65	2012	{"02":"3.07"}	{"77":"2.47"}	{"49":"3.66"}	{"4
66	2012	{"02":"2.36"}	{"77":"0.04"}	{"49":"6.63"}	{"4
70	1999	{"02":"3.42"}	{"77":"1.69"}	{"49":"3.88"}	{"4
74	2012	{"02":"2.32"}	{"77":"4.36"}	{"49":"2.01"}	{"4
77	1992	{"02":"1.24"}	{"77":"3.81"}	{"49":"3.54"}	{"4
83	2012	{"02":"5.05"}	{"77":"0.13"}	{"49":"1.06"}	{"4
85	1997	{"02":"2.01"}	{"77":"0.13"}	{"49":"5.8"}	{"4
90	1982	{"02":"0.45"}	{"77":"0"}	{"49":"7.28"}	{"4

Cargar

Transformar datos Cancelar

Agre

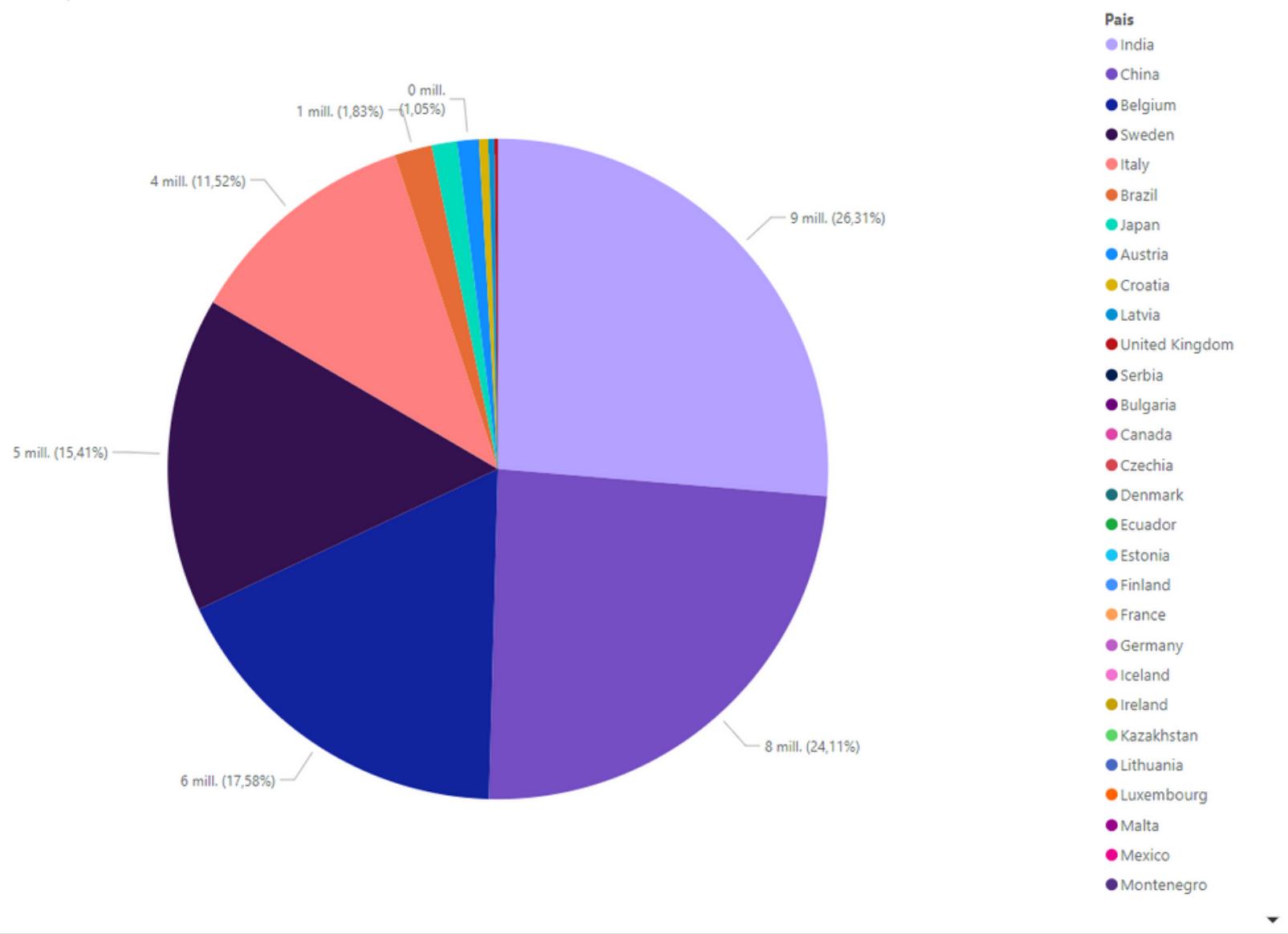
a vez cargados lo

Importar datos de SQL

Continuar Cancelar

Obtener datos de otro origen →

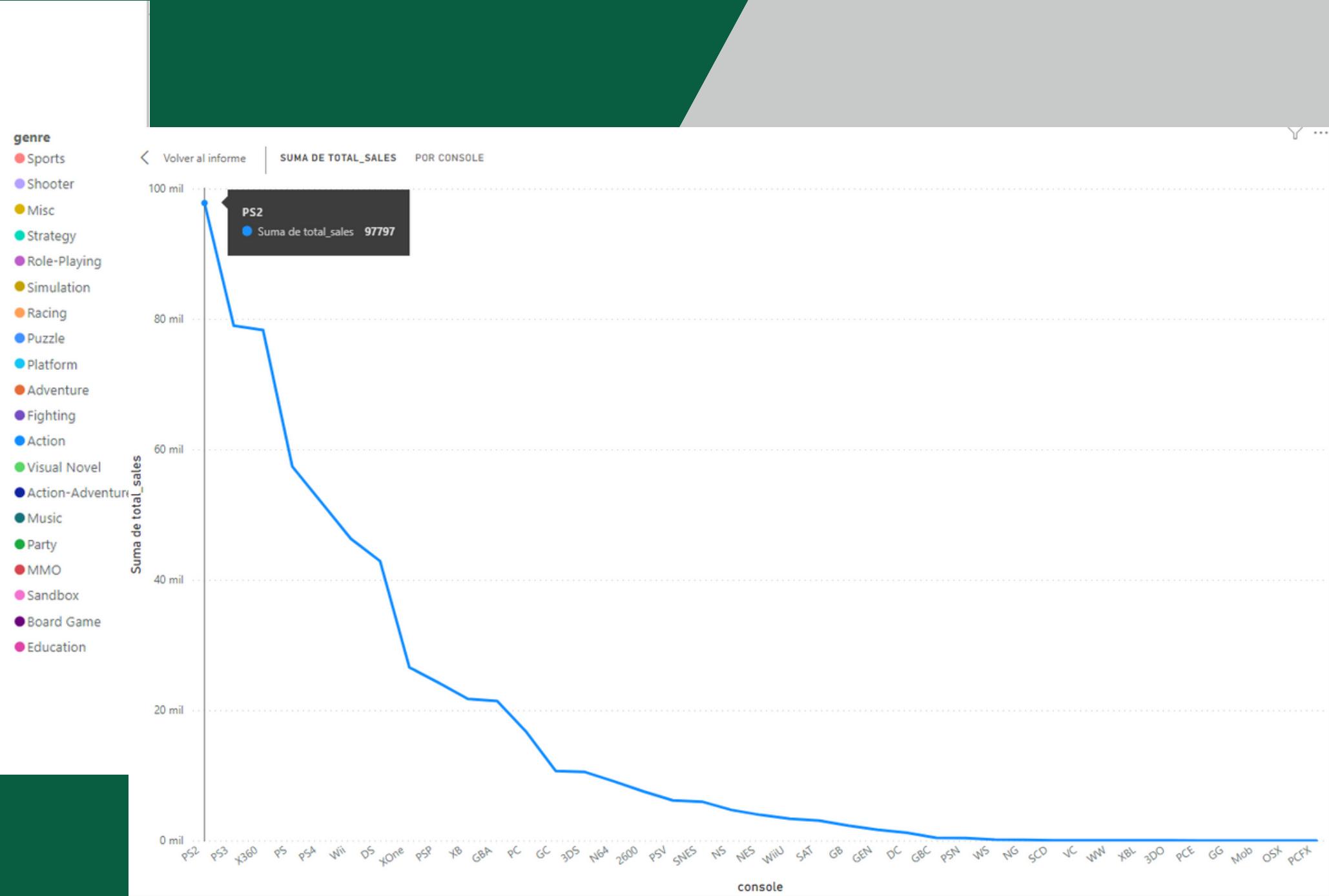
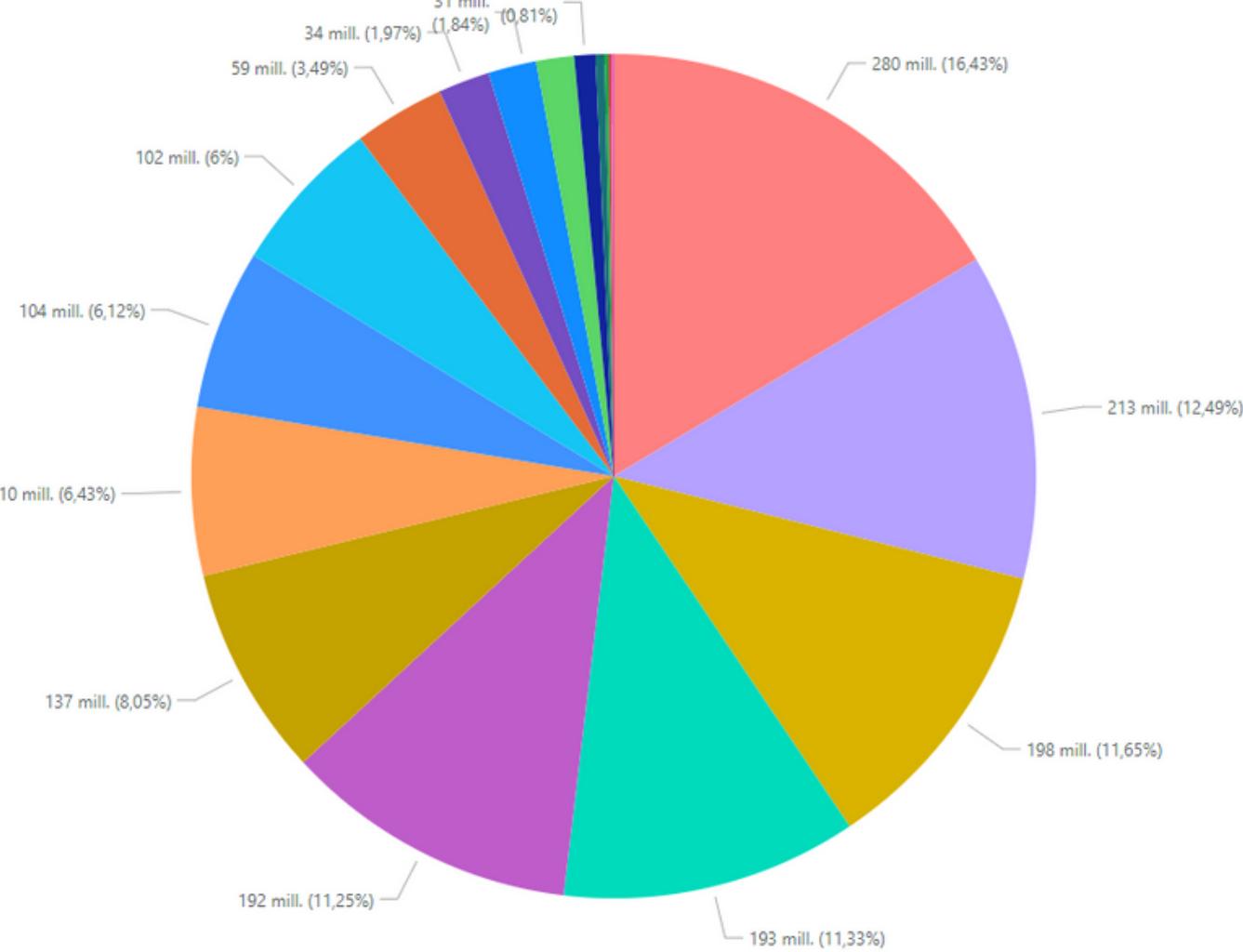
RESULTADOS



RESULTADOS



Categoría de los videojuegos con mayor numero de ventas



Mayor numero de ventas a nivel mundial en las diferentes consolas

Artistas mas escuchados de Spotify

De web

Básico Uso avanzado

Dirección URL

`https://libretilla.com/cantantes-mas-seguidores-spotify/`

Aceptar

MongoDB Compass - localhost:27017/Artistas.Spotify

localhost:27017

Artistas.Spotify

Documents Aggregations Schema Indexes Validation

Type a query: { field: 'value' } or [generate query](#).

60 DOCUMENTS 1 INDEXES

Explain Reset Find Options

1 - 20 of 60

Document 1: _id: ObjectId('65e67dbd6c0fd291f0b974ec') Column1: 1 Nombre: "Ed Sheeran" Seguidores (en millones): "114,87"

Document 2: _id: ObjectId('65e67dbd6c0fd291f0b974ed') Column1: 2 Nombre: "Ariana Grande" Seguidores (en millones): "93,56"

Document 3: _id: ObjectId('65e67dbd6c0fd291f0b974ee') Column1: 3 Nombre: "Arijit Singh" Seguidores (en millones): "88,54"

Document 4: _id: ObjectId('65e67dbd6c0fd291f0b974ef') Column1: 4 Nombre: "Billie Eilish" Seguidores (en millones): "85,93"

Import completed. 60 documents imported.

```

import pymongo
import pyodbc

# Conexión a MongoDB
mongo_client = pymongo.MongoClient("mongodb://localhost:27017/")
mongo_db = mongo_client["Artistas"]
mongo_collection = mongo_db["spotify"]

# Conexión a SQL Server
sql_conn = pyodbc.connect('Driver={SQL Server};'
                          'Server=root;'
                          'Database=spotify;'
                          'Trusted_Connection=yes;')

sql_cursor = sql_conn.cursor()

# Obtener datos de MongoDB
mongo_data = mongo_collection.find()

# Iterar sobre los documentos y guardarlos en SQL Server
for doc in mongo_data:

    sql_query = "INSERT INTO tabla_sql_server (columna1, columna2, columna3) VALUES (?, ?, ?)"
    sql_cursor.execute(sql_query, (doc["campo1"], doc["campo2"], doc["campo3"]))
    sql_conn.commit()

# Cerrar conexiones
sql_cursor.close()
sql_conn.close()
mongo_client.close()

```

Navigator SQL File 5* SQL File 6* estudiante curso estudiante SQL File 7* ×

SCHEMAS

Filter objects

- file2
- matricula
- netflix_2
- spotify**
 - Tables artistas
 - Views
 - Stored Procedures
 - Functions
- sys
- test

Administration Schemas Information

Schema: spotify

1
2 • USE spotify;
3 • SELECT *from artistas;
4

Result Grid | Filter Rows: Export: Wrap Cell Content:

Column1	2	Nombre	Seguidores (en millones)
1		Ed Sheeran	114,87
2		Ariana Grande	93,56
3		Arijit Singh	88,54
4		Billie Eilish	85,93
5		Taylor Swift	85,48
6		Drake	79,46
7		Eminem	76,29
8		Justin Bieber	74,11
9		Bad Bunny	73,9
10		The Weeknd	71,12
11		BTS	69,27
12		Rihanna	57,7
13		Adele	51,18
14		Bruno Mars	50,79
15		Imagine Drag...	50,13
16		Queen	48,26
17		Coldplay	47,06
18		Blackpink	45,2
19		Selena Gomez	44,98
20		XXXTentacion	42,97
21		Shawn Mendes	42,44
22		Neha Kakkar	42,37
23		Post Malone	41,99
24		Dua Lipa	41,46
25		Maroon 5	41,44

artistas 1 ×

Navegador

Opciones de presentación 🔍

- Tablas HTML [1]
 - Tabla 1
- Tablas sugeridas [5]
 - Tabla 2
 - Tabla 3**
 - Tabla 4
 - Tabla 5
 - Tabla 6
- Texto [2]
 - Código HTML
 - Texto mostrado

Agregar tabla mediante ejemplos

Vista de tabla Vista web

Tabla 3

Column1	Column2	Column3	Column4	Column5	Column6
null	Nombre	Seguidores (en millones)		1 Ed Sheeran	114,87
2	Ariana Grande	93,56		3 Arijit Singh	88,54
4	Billie Eilish	85,93		5 Taylor Swift	85,48
6	Drake	79,46		7 Eminem	76,29
8	Justin Bieber	74,11		9 Bad Bunny	73,9
10	The Weeknd	71,12		11 BTS	69,27
12	Rihanna	57,70		13 Adele	51,18
14	Bruno Mars	50,79		15 Imagine Dragons	50,13
16	Queen	48,26		17 Coldplay	47,06
18	Blackpink	45,20		19 Selena Gomez	44,98
20	XXXTentacion	42,97		21 Shawn Mendes	42,44
22	Neha Kakkar	42,37		23 Post Malone	41,99
24	Dua Lipa	41,46		25 Maroon 5	41,44
26	Alan Walker	39,31		27 Karol G	39,23
28	A.R. Rahman	38,04		29 Ozuna	38,01
30	J Balvin	37,87		31 Beyoncé	35,85
32	Marshmello	35,30		33 Maluma	34,5
34	Daddy Yankee	33,83		35 One Direction	33,08
36	Marília Mendonça	32,84		37 Anuel AA	32,16
38	Camila Cabello	31,48		39 Shakira	31,27
40	Juice WRLD	30,57		41 Guns N' Roses	29,91
42	Nicki Minaj	29,46		43 Harry Styles	29,23
44	Katy Perry	29,20		45 Michael Jackson	28,65
46	Lady Gaga	27,86		47 Olivia Rodrigo	27,72
48	Lana Del Rey	27,60		49 AC/DC	27,3
50	Sia	26,91		51 David Guetta	26,6
52	Doja Cat	26,19		53 The Beatles	26,12

Cargar Transformar datos

Cargar

Tabla 1

Evaluando...

Tabla 3

Evaluando...

Tabla 4

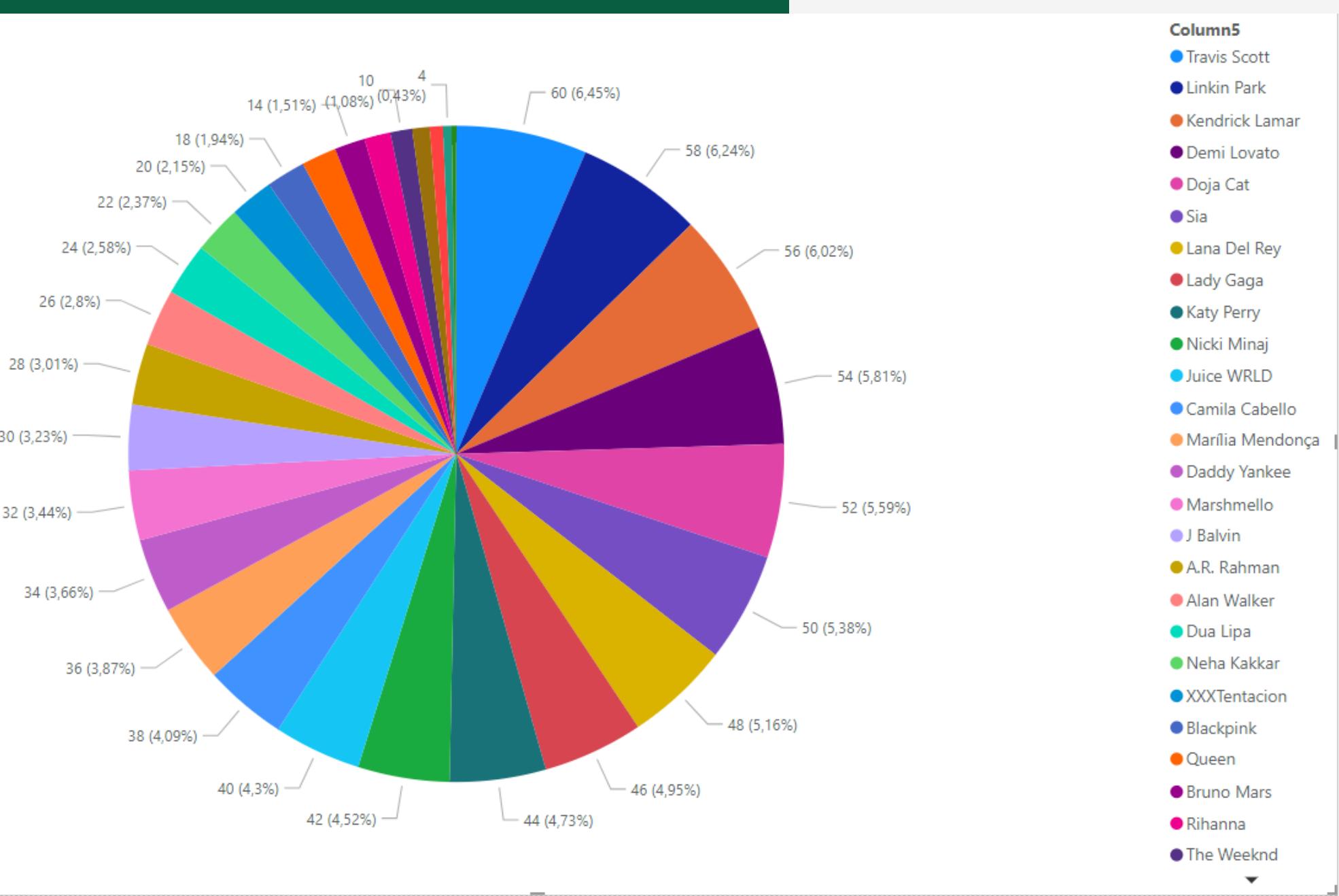
Evaluando...

Tabla 5

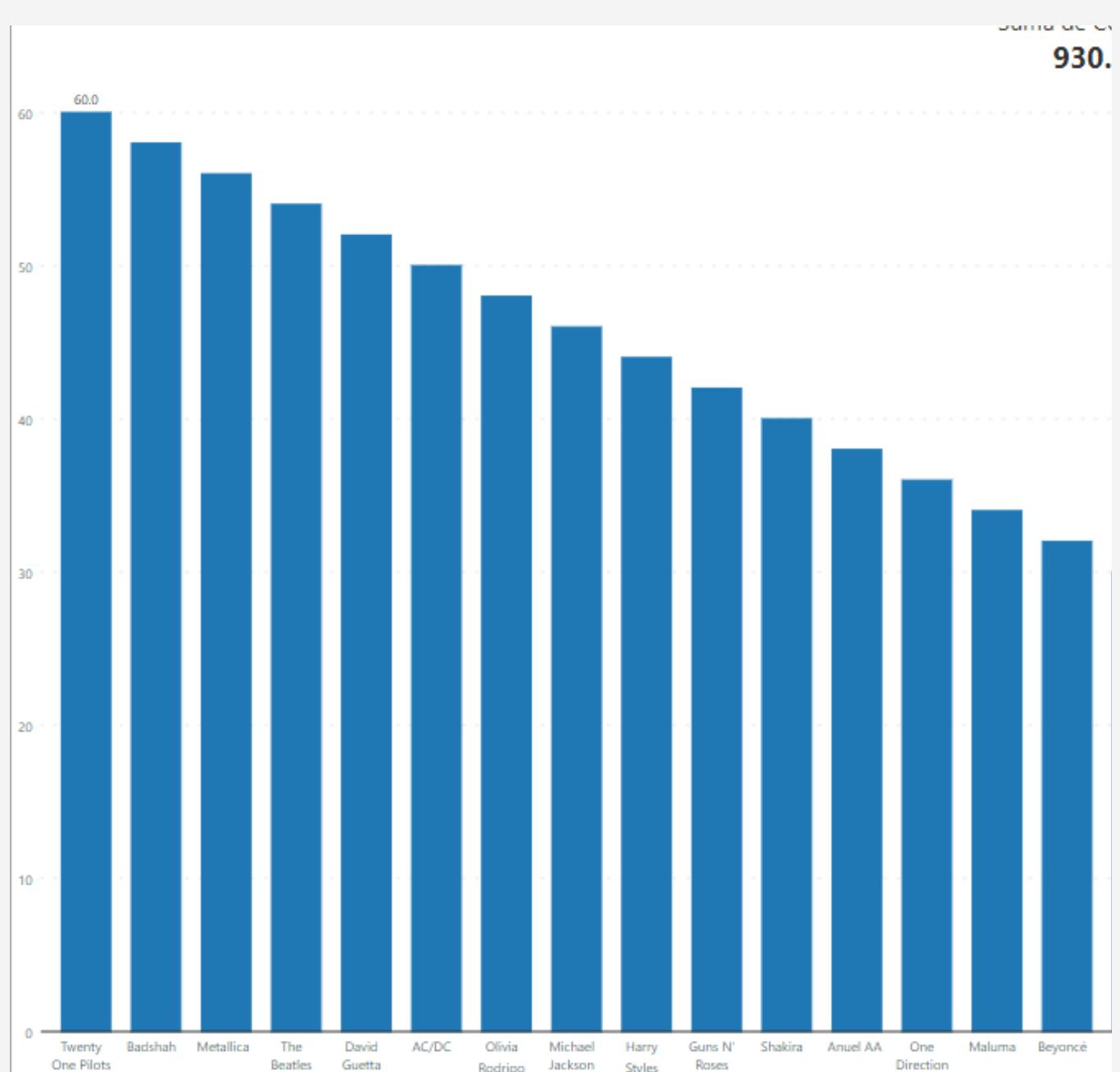
Evaluando...

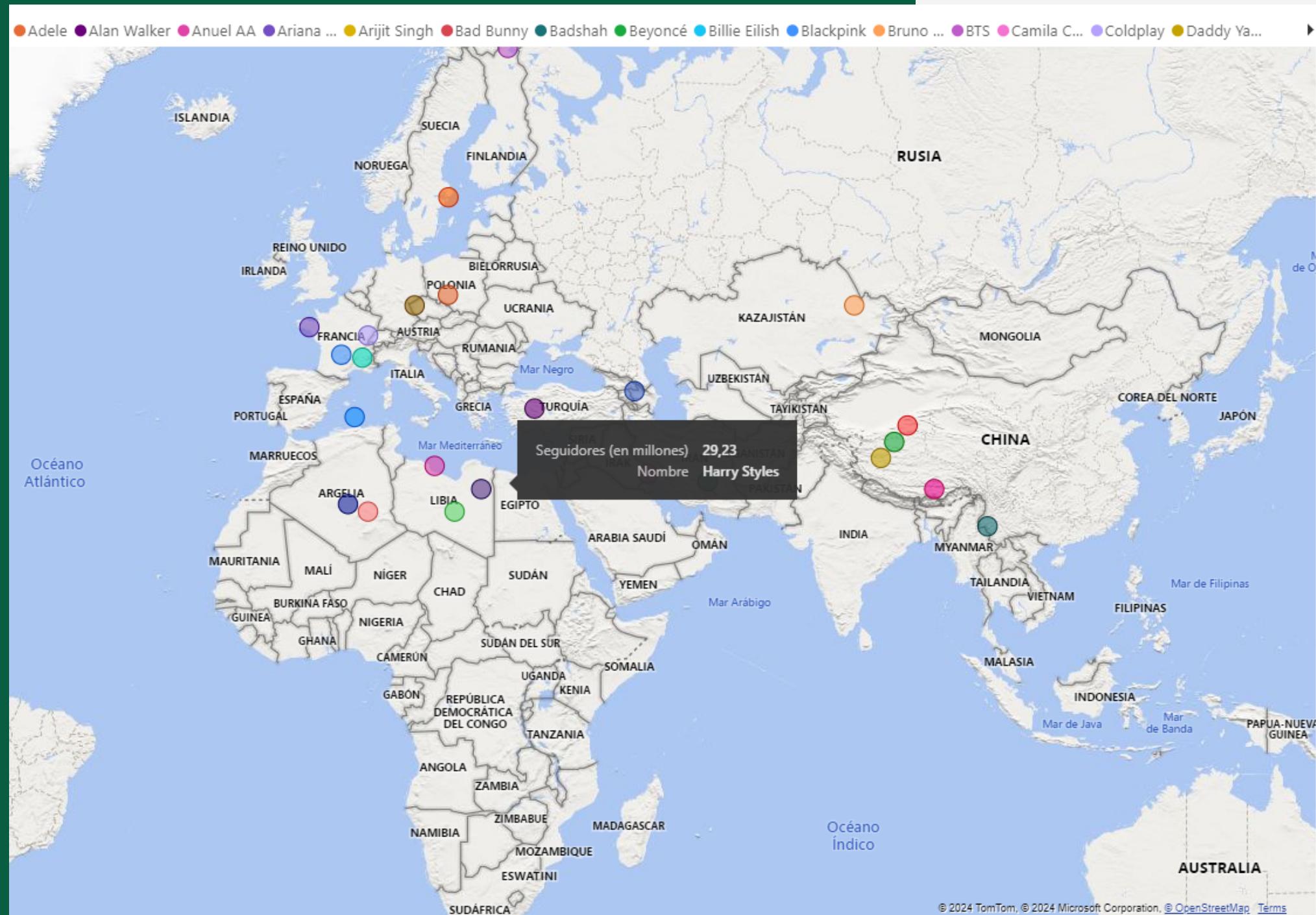
Cancelar

Artistas mas escuchados mensualmente en el genero de ingles



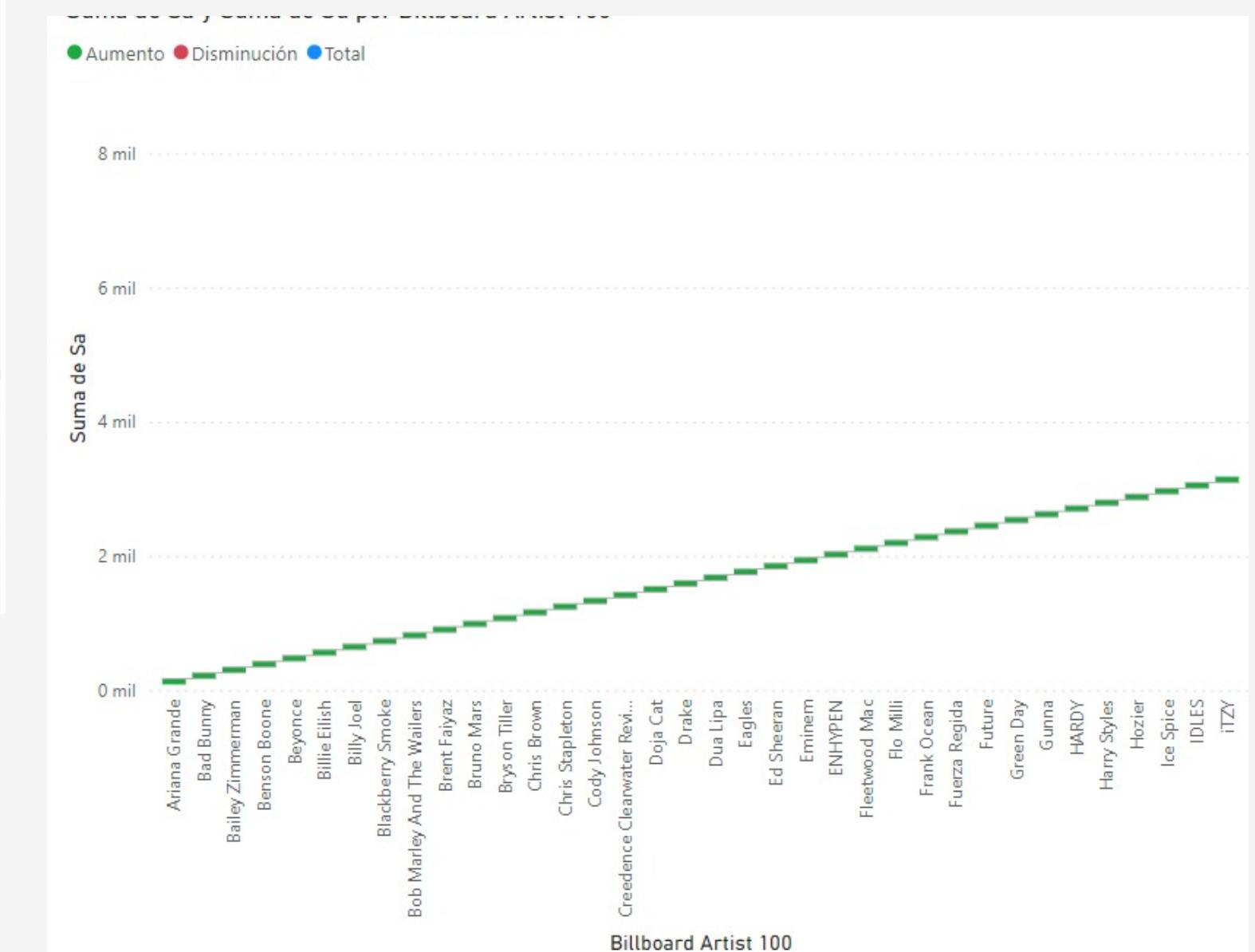
Numero de reproducciones a nivel mundial de los últimos meses dentro de la plataforma





Numero de reproducciones de los diferentes artistas con mayor escuchas dentro de los países

Artistas dentro de una lista de reproducción de los billboard según sus reproducciones



CONCLUSIONES

- * La integración de múltiples fuentes de datos, incluyendo redes sociales, webscraping y archivos estáticos, permite obtener una visión holística y completa de los temas analizados.
- * El diseño de una arquitectura de Data Lake con múltiples bases de datos NoSQL y relacionales permite gestionar eficientemente la gran cantidad y variedad de datos recopilados, aunque también implica desafíos en términos de mantenimiento y escalabilidad.
- * La conexión entre el concentrador de datos y PowerBI facilita la creación de dashboards interactivos y visualizaciones dinámicas, lo que permite explorar y comunicar de manera efectiva los hallazgos obtenidos a partir del análisis de datos.
- * Los resultados obtenidos del análisis de datos ofrecen insights significativos sobre el pulso político en diferentes regiones de Ecuador, la popularidad de los juegos en línea en distintos países y otros temas de interés, lo que puede ser de utilidad para la toma de decisiones en diversos ámbitos.

RECOMENDACIONES

Optimización de la Arquitectura de Datos, continuar evaluando y refinando la arquitectura de Data Lake para mejorar la eficiencia y la escalabilidad, considerando la posibilidad de implementar herramientas de orquestación y automatización para simplificar los procesos de recopilación, almacenamiento y procesamiento de datos.

Actualización Periódica de los Dashboards y Visualizaciones, mantener los dashboards y visualizaciones actualizados de forma regular con nuevos datos y análisis, con el fin de garantizar que la información presentada sea relevante y útil para la toma de decisiones en tiempo real.

Muchas Gracias

