



# PROYECTO FINAL

Base de Datos Multidimensionales

Integrantes:

Alevar Andres, Chinacalle Karla, Salgado Luis, Vargas Erick



# PROYECTO FINAL

## OBJETIVO GENERAL

Diseñar una arquitectura Data Lake que contenga el insumo de datos de al menos las siguientes fuentes:

Dos bases de datos SQL

Dos bases de datos NoSQL

Dos fuentes de Internet (facebook, twitter, webscrapping, etc)

## OBJETIVOS ESPECÍFICOS

Diseñar la arquitectura de solución Data lake.

Crear cronograma de actividades.

Instalar herramientas a utilizar.

## OBJETIVOS ESPECÍFICOS

Designar actividades a cada integrante.

Conectar herramientas para la abstracción de datos.

Buscar los temas de interés establecidos.

## OBJETIVOS ESPECÍFICOS

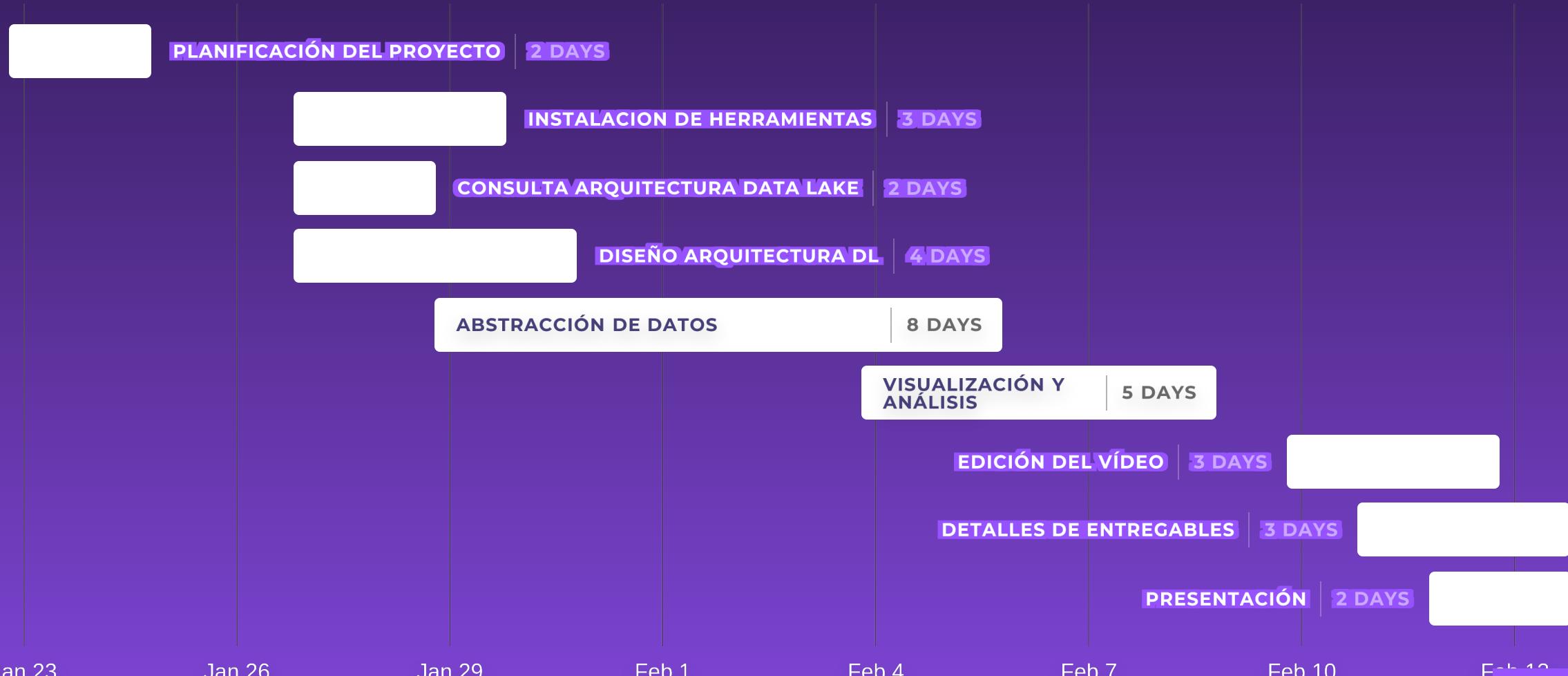
Abstraer datos.

Analizar la información obtenida.

# PLANIFICACIÓN DE ACTIVIDADES



# CRONOGRAMA DE ACTIVIDADES



# HERRAMIENTAS UTILIZADAS



## BASES DE DATOS SQL

- \* SQL Server
- \* MySQL



## BASES DE DATOS NO SQL

- \* CouchDB
- \* Couchbase



## FUENTES DE INTERNET

- \* Facebook
- \* Twitter



## OTRAS HERRAMIENTAS

- \* Facebook API
- \* Elasticsearch
- \* Twitter API
- \* Bounding Box
- \* Excel
- \* CNE
- \* INEC

# DIVISIÓN DE TEMÁTICAS

- **FACEBOOK - COUCHBASE**

- \* Actividades y Hobbies.
- \* Post de cada uno de los integrantes.
- \* Top 10 twitteros en 5 ciudades de Ecuador.

- **TWITTER - COUCHDB**

- \* Restaurantes y sitios de esparcimiento.
- \* Eventos o noticias mundiales.
- \* Top 10 quejas en Ecuador.

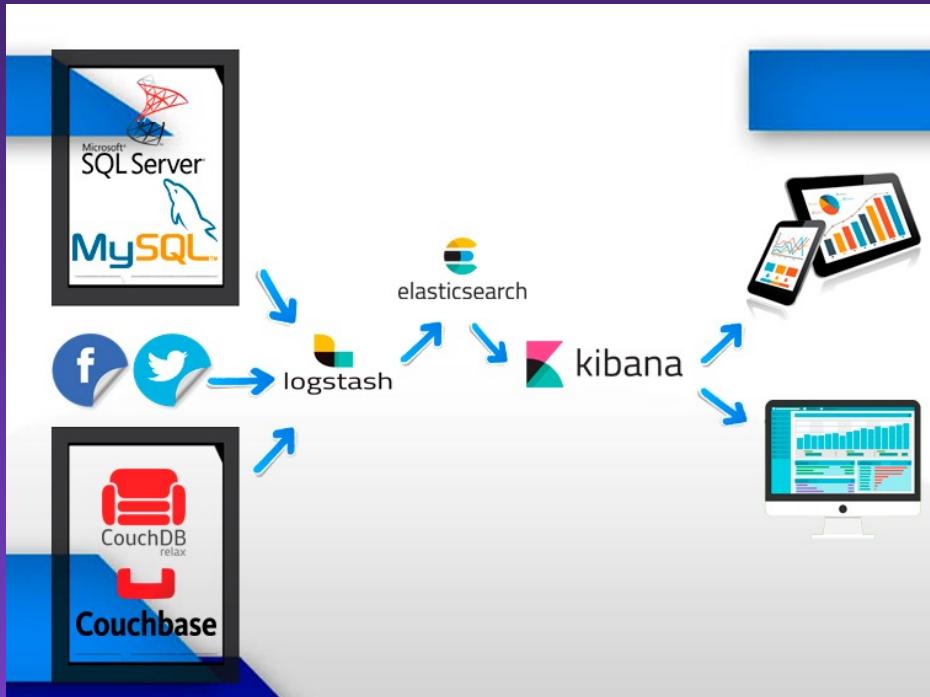
- **EXCEL - SQLSERVER**

- \* Tráfico vehicular en las 5 principales ciudades de Ecuador.
- \* Pulso político en 5 ciudades de Ecuador.

- **EXCEL - MYSQL**

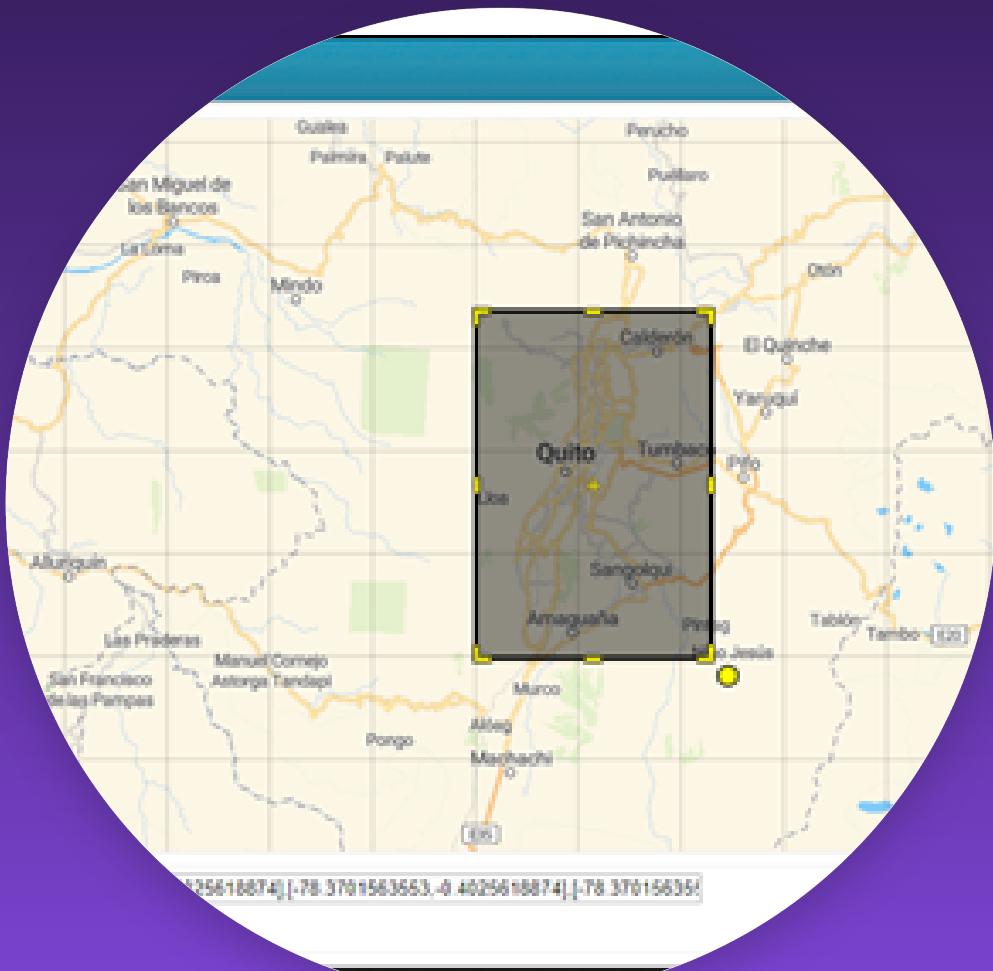
- \* Conciertos y Eventos públicos.
- \* Eventos deportivos en los principales estadios de Ecuador.

# ARQUITECTURA DE LA SOLUCIÓN



- **CONCENTRADOR DE DATOS:**  
ElasticSearch.

# EXTRACCIÓN DE DATOS

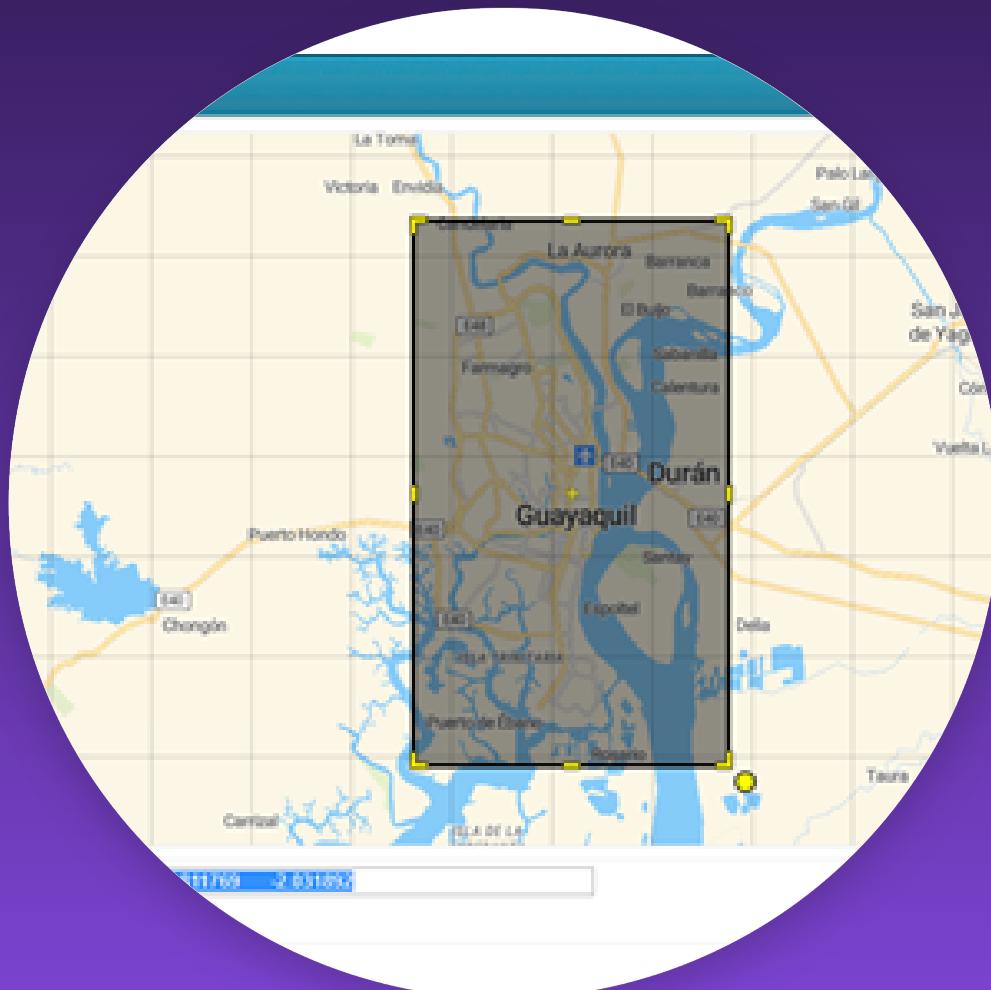


## TOP 10 TWITTEROS - QUITO

Utilizamos la herramienta Bounding Box que nos permite, por medio de un mapa, filtrar las ciudades para obtener sus coordenadas.

```
twitterStream.filter(locations=[-78.596,-0.402562,-78.370156,-0.064736])
```

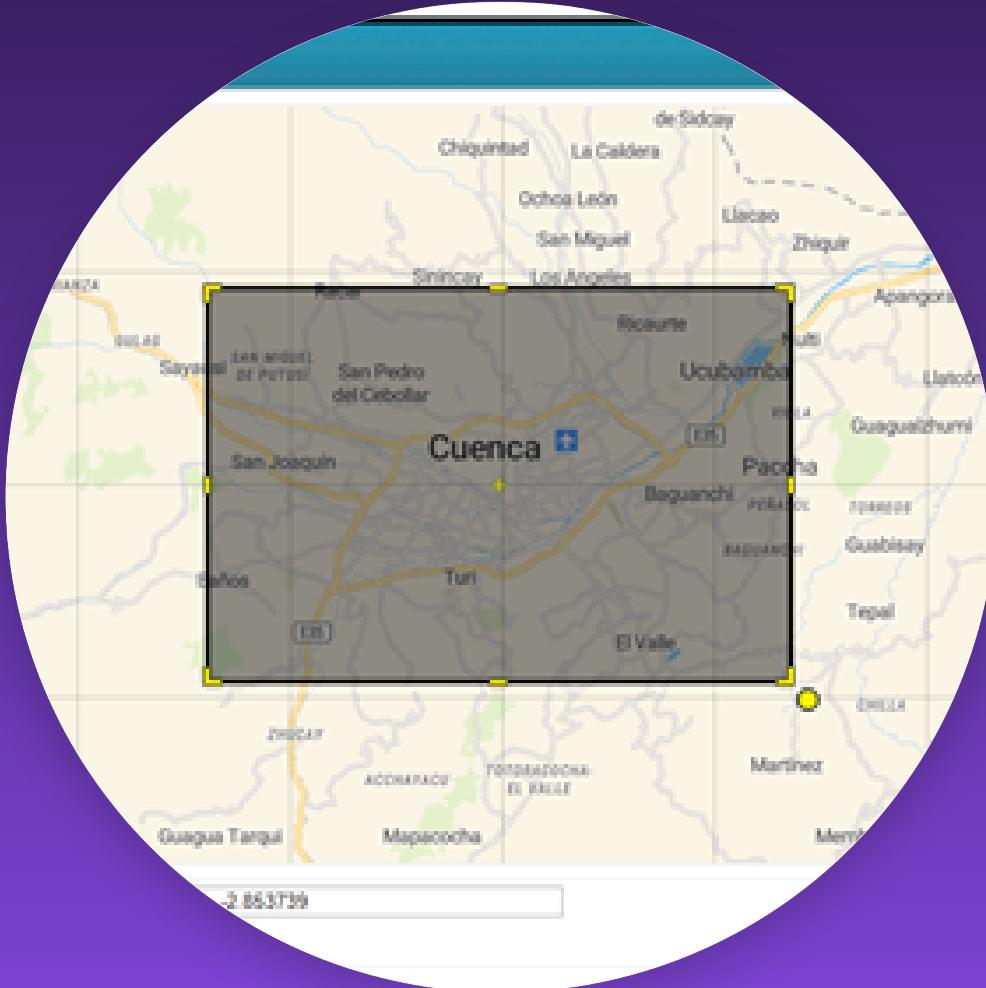
# EXTRACCIÓN DE DATOS



## GUAYAQUIL

```
twitterStream.filter(locations=  
[-79.961,-2.304295,-79.811769,-2.031892])
```

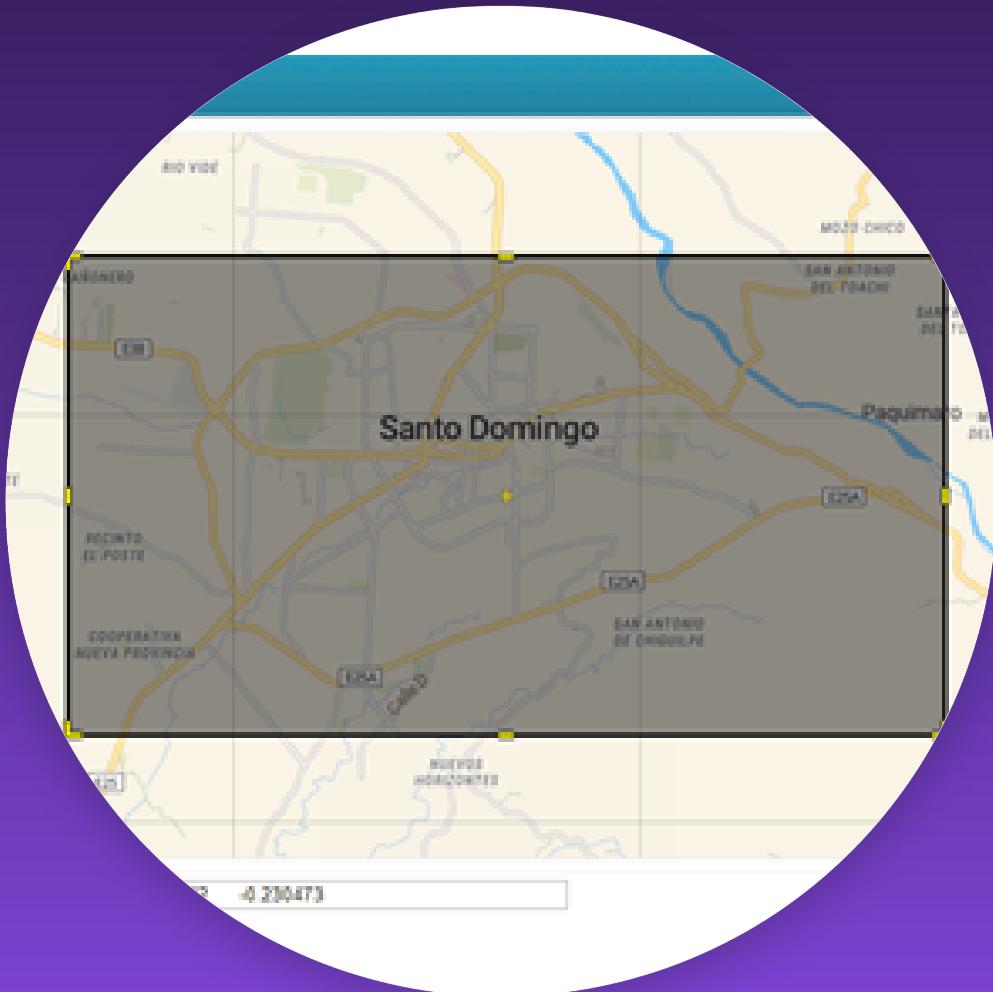
# EXTRACCIÓN DE DATOS



## CUENCA

```
twitterStream.filter(locations=  
[-79.96901,-2.304,-79.811769,-2.031892])
```

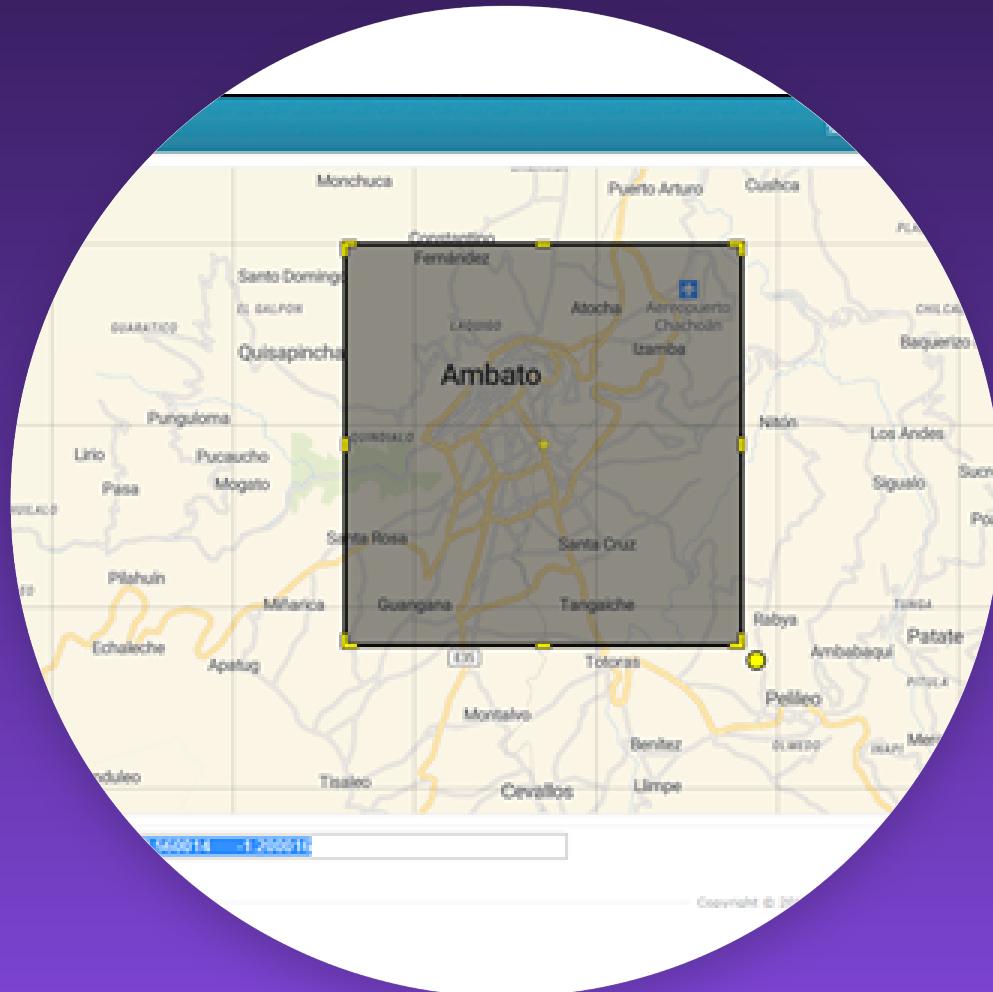
# EXTRACCIÓN DE DATOS



## SANTO DOMINGO

```
twitterStream.filter(locations=[-79.220346,-0.289352,-79.112542,-0.230473])
```

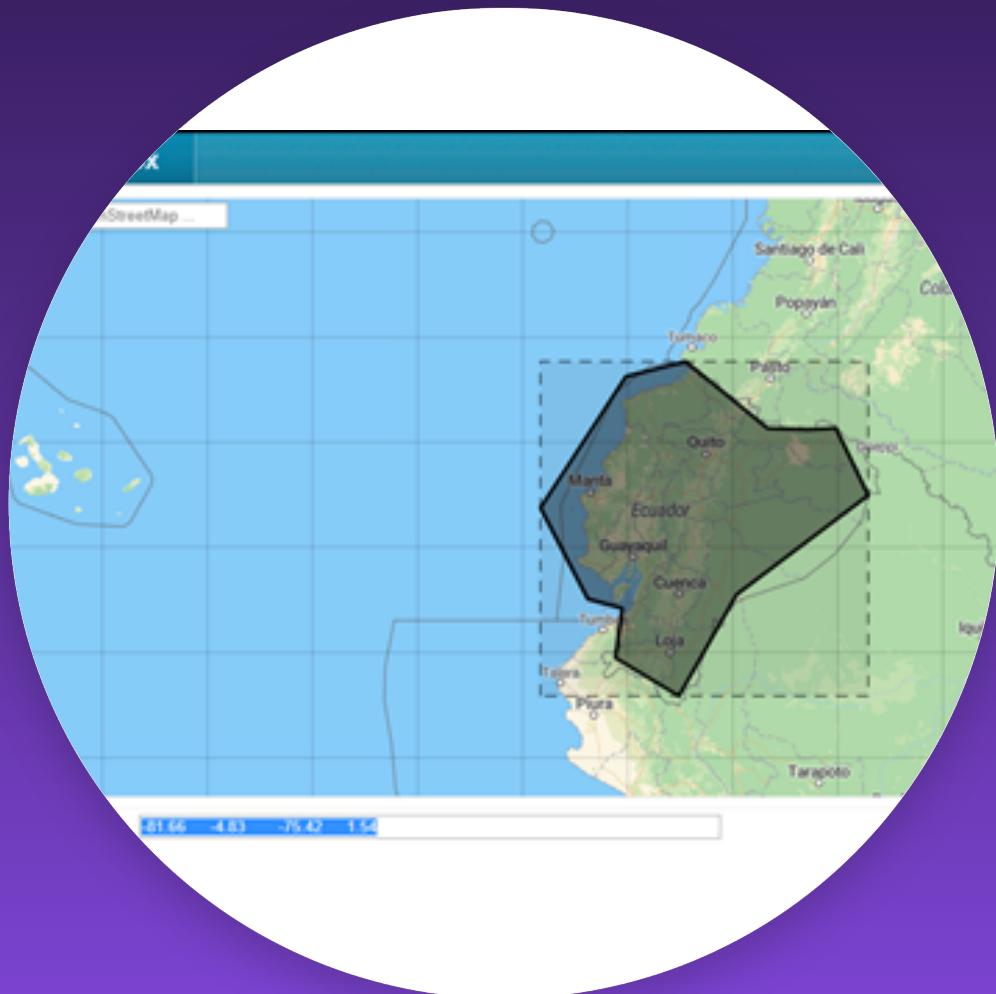
# EXTRACCIÓN DE DATOS



## AMBATO

```
twitterStream.filter(locations=  
[-78.96911,-2.304295,-79.811769,-2.031892])
```

# EXTRACCIÓN DE DATOS

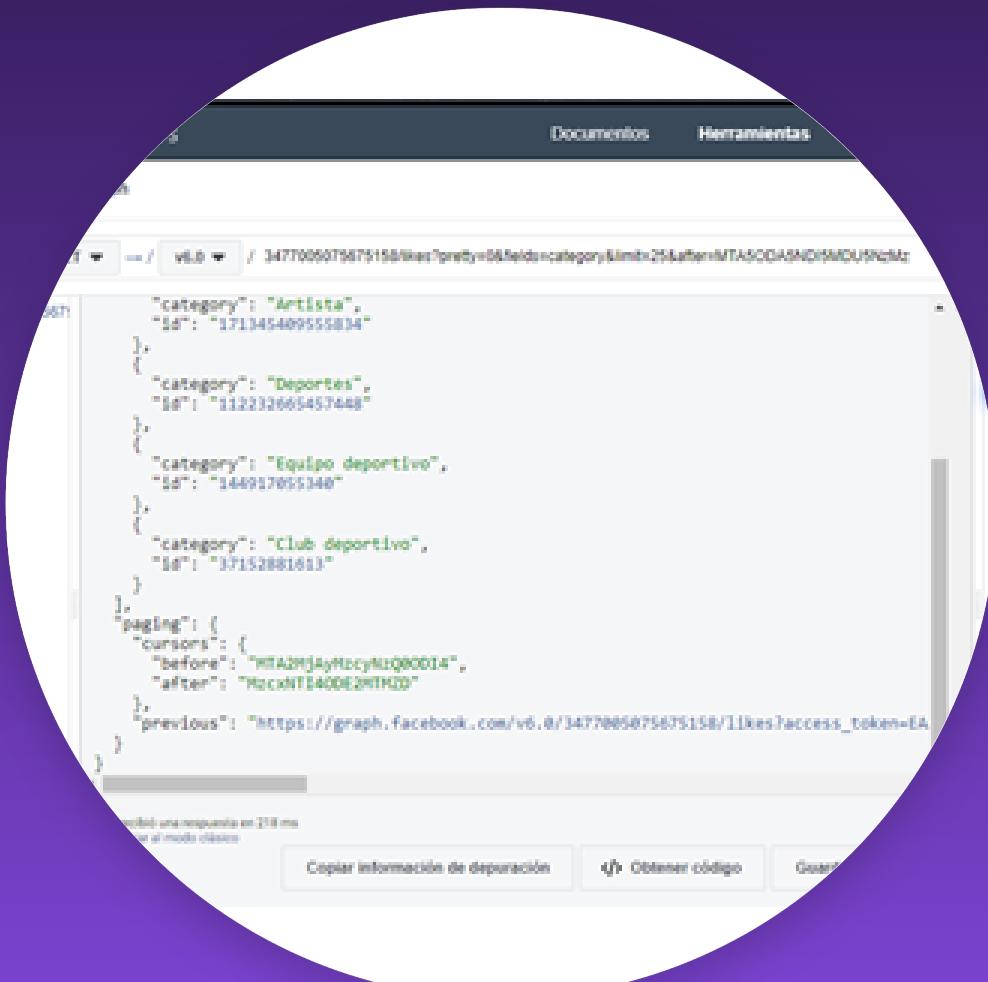


## QUEJAS EN ECUADOR

Ahora seleccionamos como límite todo el país, obtenemos las palabras más buscadas en quejas del siguiente link:

[https://www.eltelegrafo.com.ec/  
noticias/sociedad/6/525-reclamos-de-  
consumidores-se-receptaron-en-2014-  
la-mayoria-por-productos-infografia](https://www.eltelegrafo.com.ec/noticias/sociedad/6/525-reclamos-de-consumidores-se-receptaron-en-2014-la-mayoria-por-productos-infografia)

# EXTRACCIÓN DE DATOS



A screenshot of a browser's developer tools Network tab, specifically the 'Responses' section. The URL in the address bar is `/v6.0/3477005075675158/likes?pretty=1&fields=category,limit=25&filter=INTERESTS&access_token=EA...`. The response is a JSON object containing an array of likes. Each like object has a 'category' field (e.g., 'Artista', 'Deportes') and an 'id' field (e.g., '171345489555834', '112223665457448'). There are five likes listed. Below the array is a 'paging' object with 'cursors' for 'before' and 'after'. At the bottom of the JSON object is a 'previous' link to `https://graph.facebook.com/v6.0/3477005075675158/likes?access_token=EA...`.

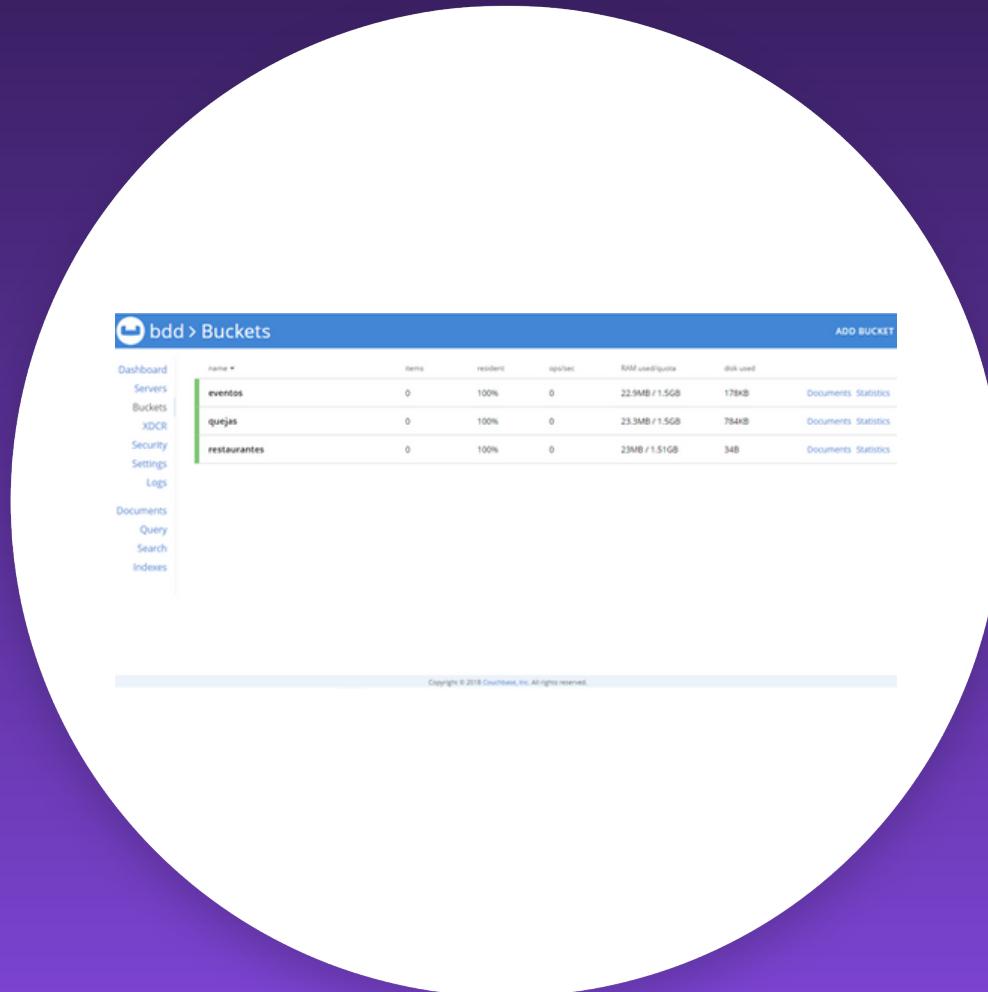
```
[{"category": "Artista", "id": "171345489555834"}, {"category": "Deportes", "id": "112223665457448"}, {"category": "Equipo deportivo", "id": "144917655348"}, {"category": "Club deportivo", "id": "37152881613"}, {"paging": { "cursors": { "before": "MTA2MjAyMzYwNzUyQzQ0ODI4", "after": "MTA2MjAyMzYwNzUyQzQ0ODI0" } }, "previous": "https://graph.facebook.com/v6.0/3477005075675158/likes?access_token=EA..."}
```

Copy debug info   Get code   Save

## ACTIVIDADES Y HOBBIES

Usando API Facebook obtenemos los post de cada uno de los miembros del grupo y la información correspondiente si todavía sigue vigente la publicación.

# EXTRACCIÓN DE DATOS

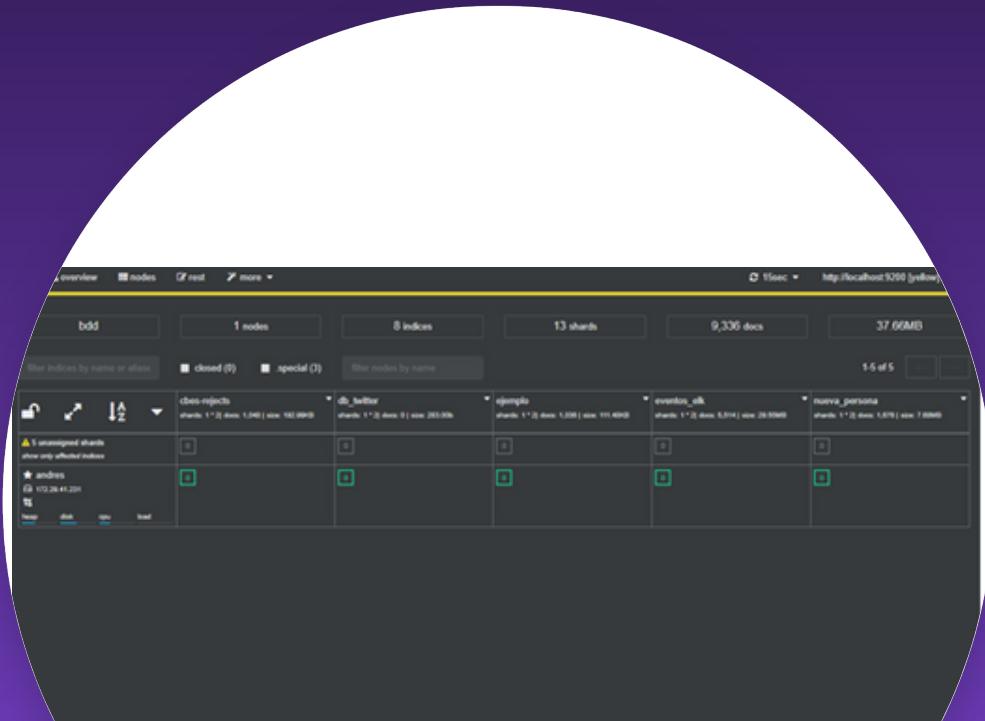


## EVENTOS Y NOTICIAS MUNDIALES

Primero vamos a crear los buckets donde irán almacenados nuestros datos con la siguiente línea de comando los filtramos desde Twitter.

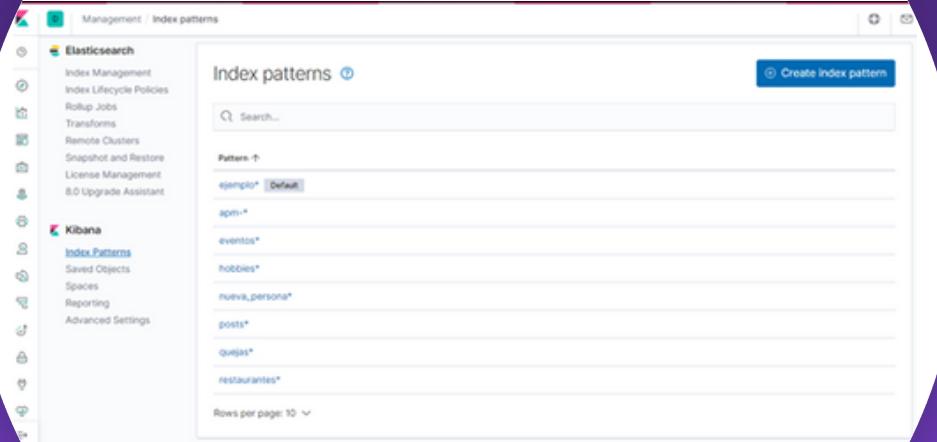
```
twitterStream.filter(track=["evento","noticia","mundial"])
```

# EXTRACCIÓN DE DATOS



Pasamos los datos recolectados en CouchBase por medio de elastic a cerebro, para esto realizaremos configuraciones en el archivo "cbes" dentro de la carpeta bin de cBeC-4.2.1, que es una aplicación que nos permite pasar los datos desde couchBas a elasticSearch.

# EXTRACCIÓN DE DATOS



Una vez hayamos repetido el mismo proceso para pasar todas nuestras bases desde Couch a ElasticSearch las guardaremos en kibana, este proceso se realiza automáticamente.

Ahora que tenemos nuestras bases en Kibana podemos ver una breve descripción de los documentos y la memoria que ocupan.

# EXTRACCIÓN DE DATOS

Provincia	total	Estado	Alquiler	Particular	Otros
AZUAY	154.697	1.941	10.447	142.291	18
BOLÍVAR	25.796	459	2.861	22.475	1
CAÑAR	44.62	407	3.946	40.265	2
CARCHI	23.666	331	3.417	19.917	1
COTOPAXI	70.996	2.099	7.554	61.338	5
CHIMBORAZC	68.074	1.696	6.97	59.405	3
EL ORO	116.063	1.163	9.747	105.143	10
ESMERALDAS	45.369	251	4.37	40.744	4
GUAYAS	529.603	4.546	26.72	498.222	115
IMBABURA	58.724	906	5.997	51.82	1
LOJA	69.424	949	5.748	62.725	2
LOS RÍOS	131.764	449	6.576	124.736	3
MANABÍ	207.42	1.388	13.421	192.593	18
MORONA SAN	14.934	524	1.518	12.89	2
NAPO	9.379	322	1.114	7.943	-
PASTAZA	11.663	426	1.131	10.104	2
PICHINCHA	540.827	12.791	38.492	488.819	725
TUNGURAHU.	107.224	1.48	10.969	94.771	4
ZAMORA CHÍ	9.989	463	1.395	8.131	-
GALÁPAGOS	2.952	117	366	2.469	-
SUCUMBÍOS	29.725	592	2.616	26.515	2
ORELLANA	18.582	517	2.034	16.03	1
SANTO DOMI	81.253	777	7.137	73.335	4
ANTA ELENA	30.907	681	3.639	26.582	5

## TRÁFICO VEHÍCULAR EN LAS PRINCIPALES CIUDADES DE ECUADOR

# EXTRACCIÓN DE DATOS

partidoPolitic sexo		
Esmeraldas	UNIDAD POP F	118
Esmeraldas	MOVIMIENTO F	213
Esmeraldas	PARTIDO POL F	46
Esmeraldas	MOVIMIENTO F	66
Esmeraldas	PARTIDO SOC F	755
Esmeraldas	MOVIMIENTO F	17
Esmeraldas	MOVIMIENTO F	50
Esmeraldas	FUERZA.EC F	34
Esmeraldas	UNION ECUA F	47
Esmeraldas	PARTIDO ADE F	57
Esmeraldas	MOVIMIENTO F	41
Esmeraldas	ALIANZA PRO F	2930
Esmeraldas	UNIDAD POP M	3349
Esmeraldas	MOVIMIENTO M	116
Esmeraldas	MOVIMIENTO M	153
Esmeraldas	PARTIDO POL M	67
Esmeraldas	MOVIMIENTO M	68
Esmeraldas	PARTIDO SOC M	655
Esmeraldas	MOVIMIENTO M	15
Esmeraldas	MOVIMIENTO M	65
Esmeraldas	FUERZA.EC M	51
Esmeraldas	UNION ECUA M	39
Esmeraldas	PARTIDO ADE M	47
Esmeraldas	MOVIMIENTO M	28
Esmeraldas	ALIANZA PRO M	2795
Esmeraldas	UNIDAD POP F	7
Esmeraldas	MOVIMIENTO F	7

## PULSO POLÍTICO EN ECUADOR

# EXTRACCIÓN DE DATOS

Local	Visitante	Fecha	Ciudad	Torneo
Christian Beni Guayaquil City	Macara	08/02/2019	Guayaquil	Liga Pro
Rodrigo Paz LDU Quito	olmedo	09/02/2019	Quito	Liga Pro
Alejandro Seri D. Cuenca	Ind. Del Valle	10/02/2019	Cuenca	Liga Pro
9 de Mayo Fuerza Amaril Delfin		09/02/2019	Machala	Liga Pro
Monumental Barcelona SC	El Nacional	09/02/2019	Guayaquil	Liga Pro
Chillogallo Aucas	Emelec	10/02/2019	Quito	Liga Pro
olimpico Atah America de Q	Mushuc Runa	11/02/2019	Quito	Liga Pro
Bellavista Tecnico Unive U. Catolica		10/02/2019	Ambato	Liga Pro
olimpico Atah El Nacional	Aucas	15/02/2019	Quito	Liga Pro
olimpico Atah U. Catolica	LDU Quito	16/02/2019	Quito	Liga Pro
Bellavista Mushuc Runa D. Cuenca		16/02/2019	Ambato	Liga Pro
Jocay de Man Delfin	Guayaquil City	16/02/2019	Manta	Liga Pro
olimpico de R olmedo	America de Q	17/02/2019	Riobamba	Liga Pro
Bellavista Macara	Barcelona SC	17/02/2019	Ambato	Liga Pro
Capwell Emelec	Fuerza Amaril	17/02/2019	Guayaquil	Liga Pro
Estadio de San Ind. Del Valle	Tecnico Unive	18/02/2019	Sangolqui	Liga Pro
Alejandro Seri D. Cuenca	Emelec	22/02/2019	Cuenca	Liga Pro
olimpico Atah America de Q	LDU Quito	23/02/2019	Quito	Liga Pro
Chillogallo Aucas	U. Catolica	23/02/2019	Quito	Liga Pro
olimpico Atah El Nacional	Delfin	23/02/2019	Quito	Liga Pro
9 de Mayo Fuerza Amaril olmedo		24/02/2019	Machala	Liga Pro
Bellavista Tecnico Unive	Macara	24/02/2019	Ambato	Liga Pro
Monumental Barcelona SC	Mushuc Runa	24/02/2019	Guayaquil	Liga Pro
Christian Beni Guayaquil City	Ind. Del Valle	25/02/2019	Guayaquil	Liga Pro
Alejandro Seri D. Cuenca	America de Q	01/03/2019	Cuenca	Liga Pro
Rodrigo Paz LDU Quito	Guayaquil City	02/03/2019	Quito	Liga Pro
Bellavista Mushuc Runa	El Nacional	02/03/2019	Ambato	Liga Pro
olimpico de R olmedo	Macara	02/03/2019	Riobamba	Liga Pro

EVENTOS  
REALIZADOS EN LOS  
PRINCIPALES  
ESTADIOS DE  
ECUADOR

# EXTRACCIÓN DE DATOS

		Fecha	Nombre
	Quito	01/01/2019	The first dan
	Quito	01/01/2019	Fin de año pa fiesta
	Centro del I Guayaquil	11/01/2019	Lolabum Guaya concierto
	bassick telefe Quito	11/01/2019	sasha conferencia
	sala Alfonso C Cuenca	17/01/2019	El estupendo teatro
	republica sur Cuenca	17/01/2019	La madre tira teatro
	hotel Dann Ca Quito	19/01/2019	women's lab conferencia
	sabai beer Quito	19/01/2019	Fiesta de nuev fiesta
	museo rock Quito	19/01/2019	Amover el cul fiesta
	selina Cuenca	19/01/2019	Hit me baby o fiesta
	bassick telefe Quito	25/01/2019	Robag Wruhn concierto
	la ideal Quito	01/02/2019	Nicola cruz sil fiesta
	HangOver Cuenca	01/02/2019	La noche vudu concierto
	selina Quito	08/02/2019	El principio de concierto
	ventura sport Cuenca	08/02/2019	Lineup 19 fiesta
	sky pool Montañita	09/02/2019	Corona sunse fiesta
	molinos del B Cuenca	14/02/2019	El dia del amo conferencia
	hosteria caba Cuenca	15/02/2019	Silvestre Dang concierto
	bassick telefe Quito	15/02/2019	Audiojack concierto
	el utero Quito	16/02/2019	El destape fiesta
	selina Cuenca	21/02/2019	Huaira concierto
	la guarida Cuenca	22/02/2019	Las rosas de l teatro
	Casa Morning Santa Elena	23/02/2019	Casa Corona fiesta
	ndo juvenil Quito	23/02/2019	Influencer pro conferencia
	Casa cu Cuenca	28/02/2019	Sal y Mileto conciert
	Quito	01/03/2019	Anja Schneide concierto
		07/03/2019	Erlend over
		10/03/2019	

CONCIERTOS Y  
EVENTOS PÚBLICOS  
EN ECUADOR

# BASES CARGADAS

The screenshot displays a Microsoft SQL Server Management Studio (SSMS) interface with several windows open:

- Object Explorer:** Shows the database structure of "DESKTOP-Q8GEPTN (SQL Server 14.0.1000.169 - karla)". It includes nodes for Databases, System Databases, Database Snapshots, baseEventos, bdd, Condominio, ejemplo, and Intento.
- SQL Query1.sql - DES...ntento (karla (53)):** A query window containing the command: `select * from intento`. The results grid shows data for various provinces in Ecuador, including their names, total population, state, rental status, and other metrics.
- Query 1:** Another query window showing the results of the command: `SELECT * FROM prueba2.futbol`. The results grid displays a list of football matches with columns: Estadio, Local, Visitante, Fecha, Ciudad, and Torneo.
- File Explorer:** Shows the file system navigation path: "C:\Users\karla\Documents\GitHub\prueba2\prueba2\prueba2\bin\Debug\prueba2.exe".
- Properties:** A properties window for the executable file.
- Output:** A window showing the execution history of the query: "1 13:03:41 SELECT \* FROM prueba2.futbol LIMIT 0, 1000" and the message "102 row(s) returned".
- File:** A file menu with options like File, Open, Save, and Exit.
- Help:** A help menu with options like Help, About, and Contents.

# ELASTICSEARCH

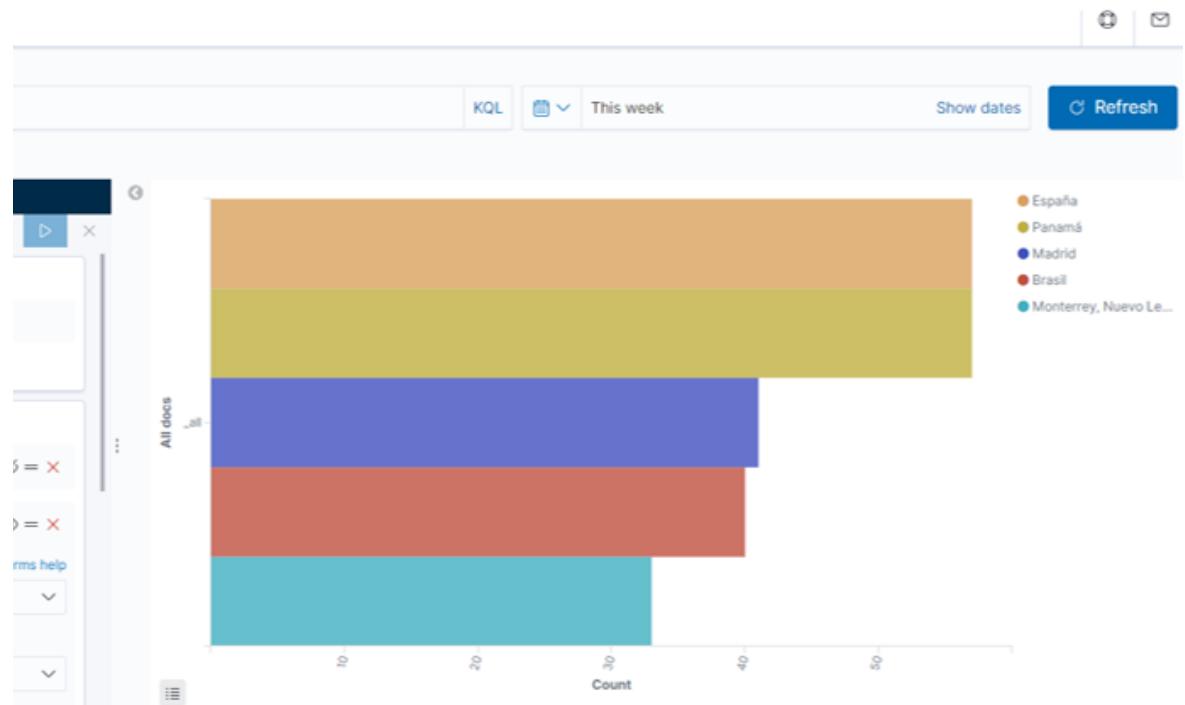
The screenshot shows the Elasticsearch dashboard interface. At the top, there are navigation links: overview, nodes, red, more, a search bar (1sec), and a URL (http://localhost:9200 [yellow]). Below the header, there are summary statistics: 0 bad, 1 nodes, 13 indices, 23 shards, 73,546 docs, and 336.72MB.

Filtering options include "filter indices by name or alias" (closed: 0, special: 0) and "filter nodes by name". A page indicator shows 6-10 of 10. On the left, there's a sidebar with icons for locks, filters, and a dropdown menu, and a message about unassigned shards. The main area displays a grid of indices: eventlog\_elk, hobbies\_elk, nuova\_persons, queue\_elk, and restaurants\_elk. Each index card includes shard counts, sizes, and a small icon. Below the grid, there are sections for "addresses" and "tags".

Index	Shards	Size
eventlog_elk	1 * 2	6.67GB   size: 26.60GB
hobbies_elk	1 * 2	94   size: 180.42GB
nuova_persons	1 * 2	1.67GB   size: 2.66GB
queue_elk	1 * 2	80.28GB   size: 233.27GB
restaurants_elk	1 * 2	34.47GB   size: 98.32GB

# VISUALIZACIONES

Restaurantes y Sitios de esparcimiento - Top 5 países donde más tiwttean acerca de restaurantes



MADE WITH

**beautiful.ai**

# VISUALIZACIONES

Restaurantes y Sitios de esparcimiento - Top 20 personas que más tiwttean acerca de restaurantes.

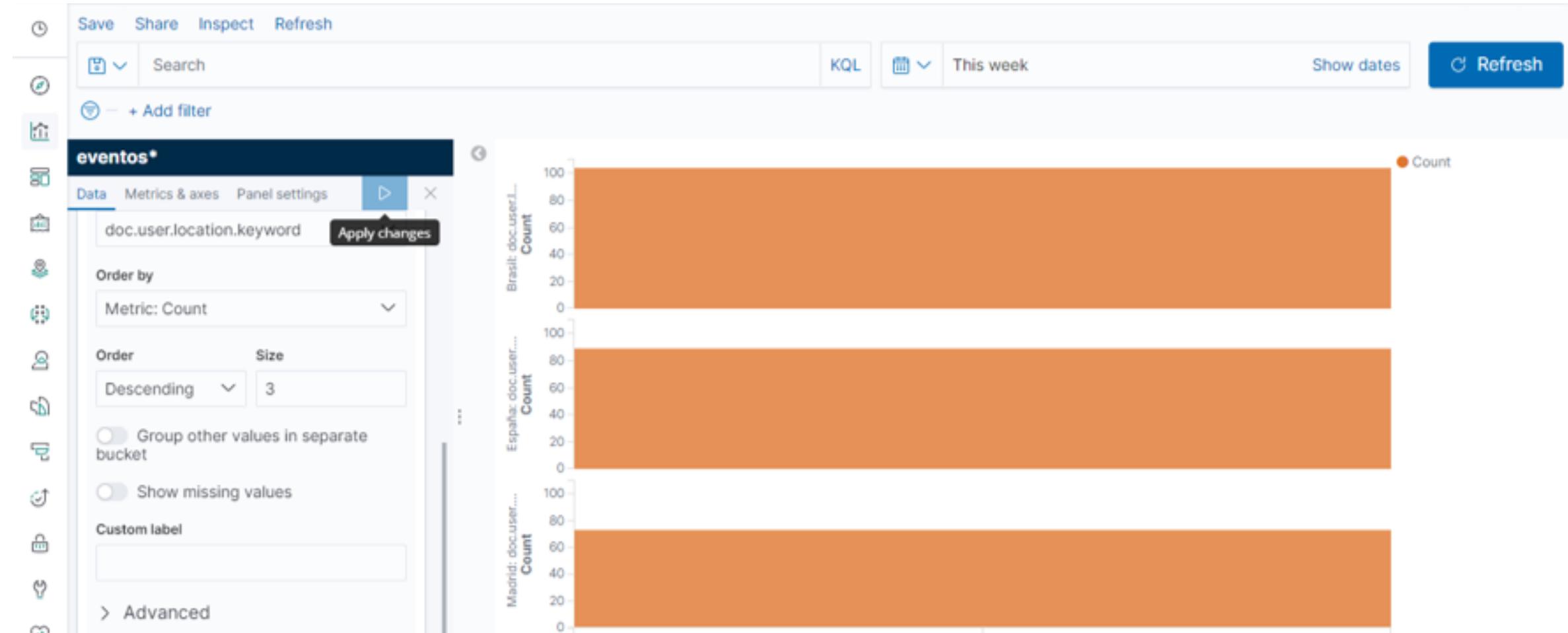
The screenshot shows a data visualization interface with the following elements:

- Top Bar:** Save, Share, Inspect, Refresh, Search input, KQL button, This week button, Show dates button, Refresh button.
- Left Sidebar:** A vertical list of icons for Save, Share, Inspect, Refresh, Search, Add filter, and a list of entities.
- Entity Panel:** A modal for "restaurante\*s" with tabs for Data (selected) and Options, and buttons for Metrics, Buckets, Apply changes, and Add.
- Metrics Panel:** Shows a list of metrics: Metric Count, Add.
- Buckets Panel:** Shows a list of buckets: Split rows doc.entities.user\_m... (with a red X), Add.
- Main Content Area:** A table showing the top 20 tweeters about restaurants, ordered by count (descending).

	Count
Andrea Peraldi	589
Jair M. Bolsonaro	496
Convivencia Clívica	161
Sash 🐶 🎉 🎉	101
Patricio Jijon Larco	57
Enrique Aristeguieta	41
Cona	37
Raúl Conte	23
Gabriel Nemer	20
ABC Digital	19
- Bottom Buttons:** Export: Raw (CSV icon), Formatted (HTML icon).

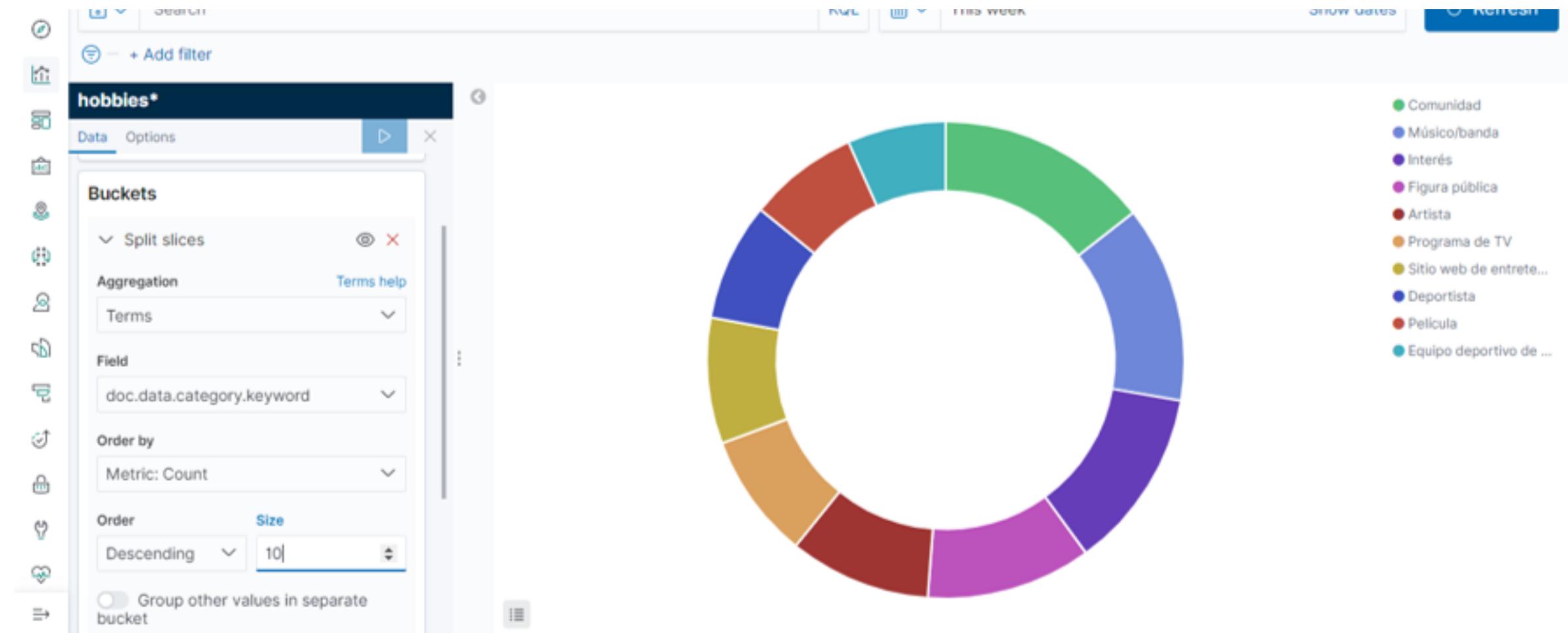
# VISUALIZACIONES

Eventos - Los tres países con más eventos realizados.



# VISUALIZACIONES

## Hobbies - Top 10 hobbies más populares



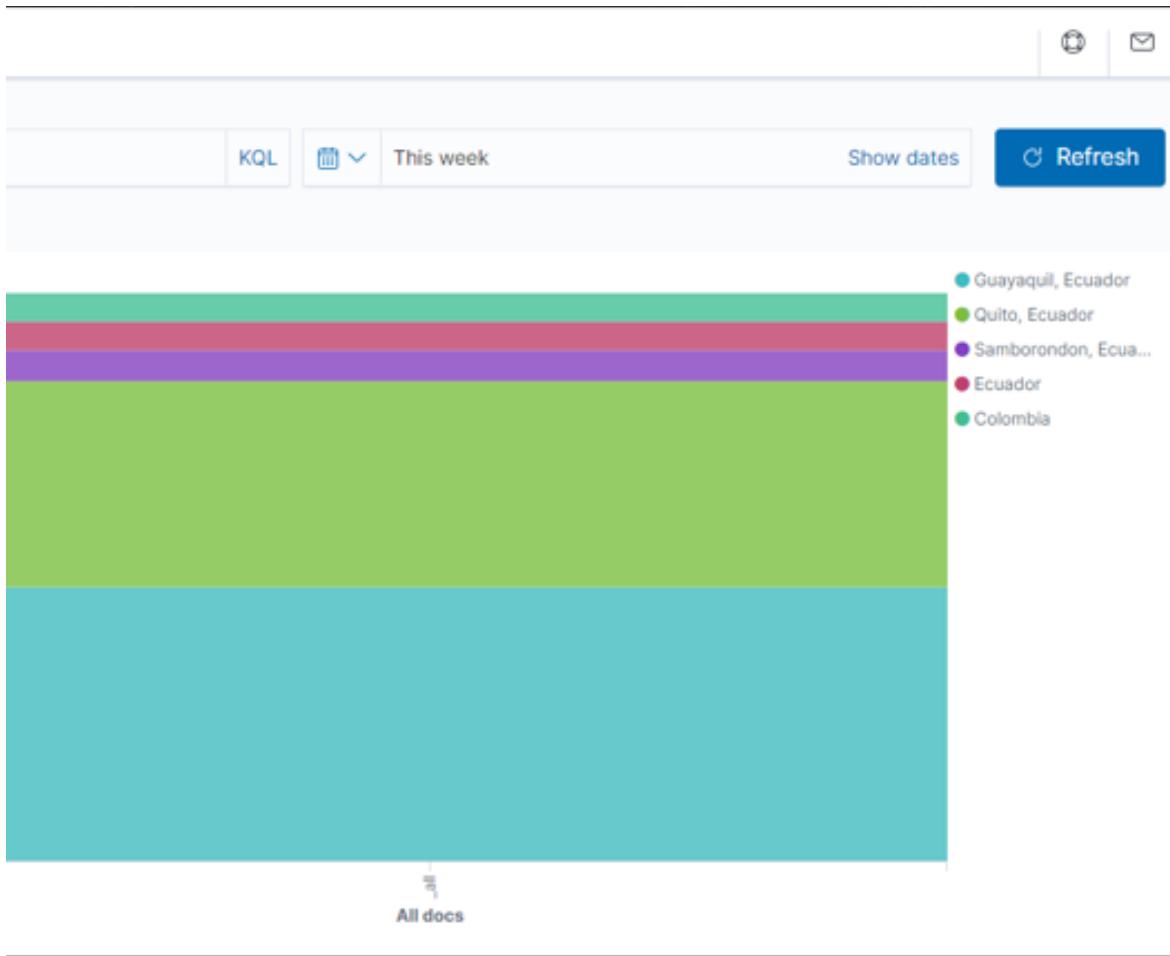
# VISUALIZACIONES

Quejas - País con más quejas.



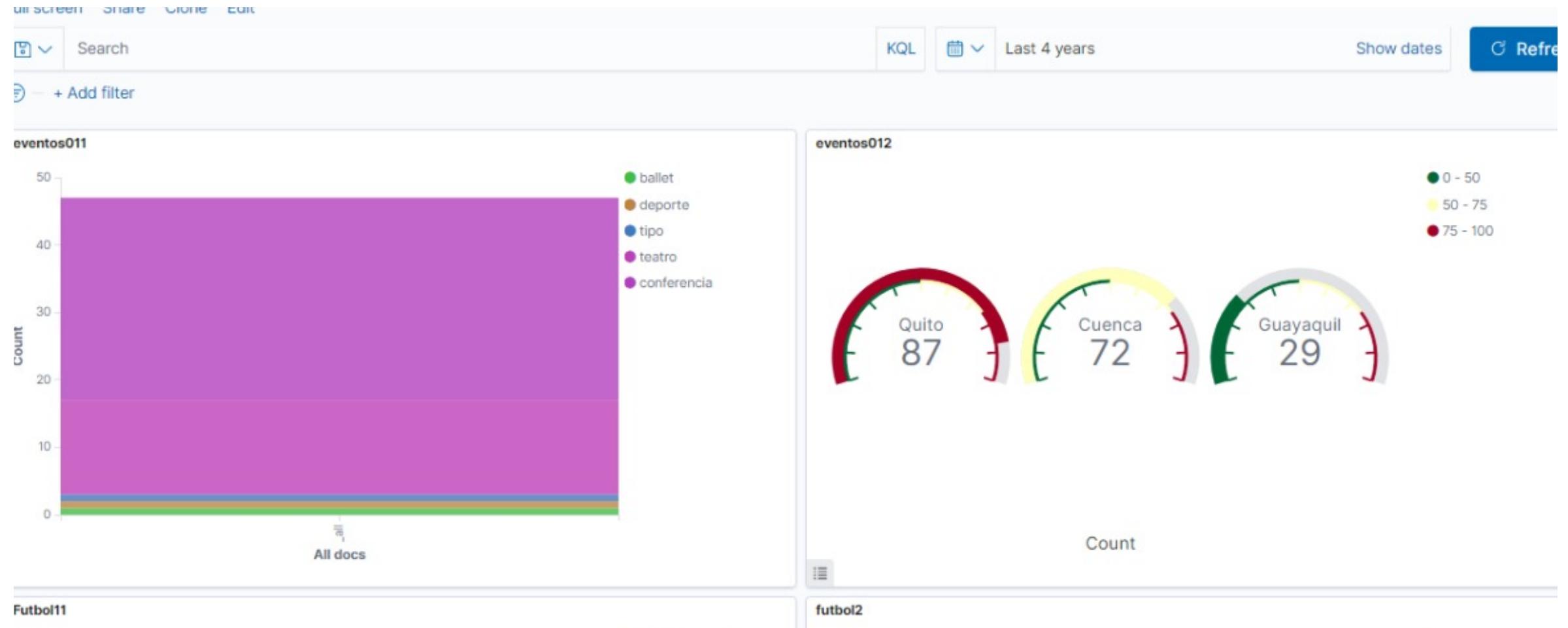
# VISUALIZACIONES

Top twitteros - Lugares que más twittean.



# VISUALIZACIONES

Eventos públicos y culturales



# VISUALIZACIONES

Eventos realizados en los principales estadios del país



restaurantesofix.xlsx

Mostrar todo

MADE WITH

beautiful.ai

# CONCLUSIONES Y RECOMENDACIONES

## CONCLUSIONES

Se reunen los datos de diferentes fuentes en Elasticsearch y se realiza las visualizaciones en su misma herramienta que es el Kibana.

## CONCLUSIONES

Al momento de instalar las herramientas verificar que estén en la misma versión para que exista armonía entre ellas.

## RECOMENDACIONES

Para poder realizar un buen análisis de cualquiera de las temáticas se deben recoger un mínimo de 2000 registros, así tendrás mejores gráficas y podrás analizar de mejor manera tus datos.

## RECOMENDACIONES

Seguir la documentación oficial de las herramientas que se utilizarán para abstraer datos.

# DESAFÍOS Y PROBLEMAS ENCONTRADOS

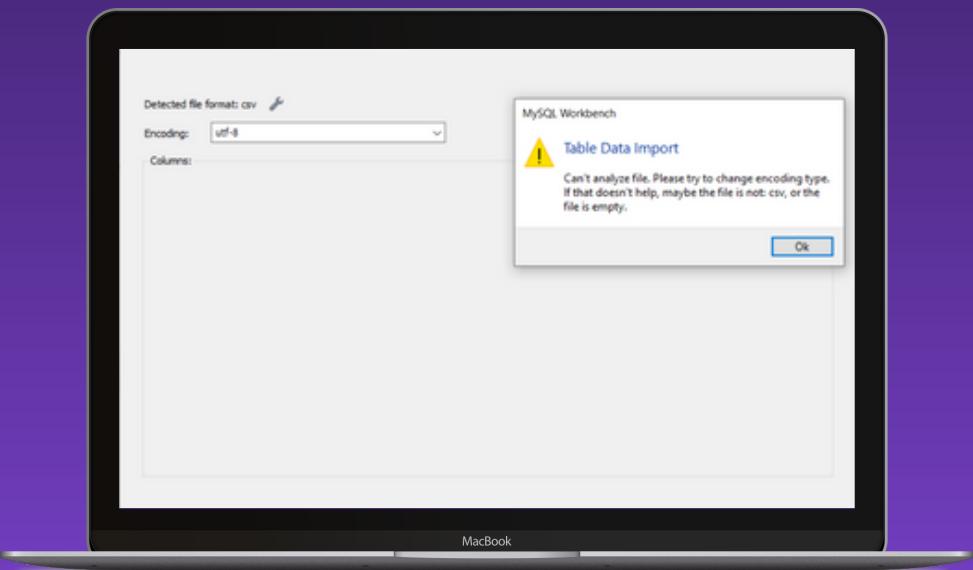
## DATOS NO ENCONTRADOS

Muchas de las temáticas no tenían una base de datos exacta, nuestra solución fue recopilar nosotros los datos y armar nuestra propia base.

## ERROR AL IMPORTAR EN ELASTICSEARCH

Se buscaron diferentes formas de importar los datos debido a que nuestra IP y puerto no funcionó.

1



2

3

## ERROR AL IMPORTAR DATOS DE EXCEL

Para poder importar nuestra base a MySQL teníamos errores debido al formato con el que creamos nuestra base en excel, esto se solucionó cambiando el formato de la fecha y eliminando todas las tildes, que era por lo que teníamos el error en nuestro caso.

# REFERENCIAS

- **COUCH BASE**

<https://pypi.org/project/couchbase/>

- **COUCH BASE & PYTHON**

<https://docs.couchbase.com/python-sdk/current/start-using-sdk.html>

- **SCRIPT PARA FACEBOOK**

<https://drive.google.com/file/d/14H7il0uxKIQLJ3FERReQc5tD1JfYaq8o/view>

- **CONECTOR ELASTIC & COUCH BASE**

<https://docs.couchbase.com/elasticsearch-connector/4.2/configuration.html>

- **TWITTER DEVELOPER**

<https://developer.twitter.com/>

- **FACEBOOK FOR DEVELOPERS**

<https://developers.facebook.com/>

- **BOUDING BOX**

<https://boundingbox.klokantech.com/>

- **KIBANA**

<https://www.elastic.co/guide/en/beats/packetbeat/current/kibana-queries-filters.html>

- **ELASTIC SEARCH**

<https://www.elastic.com/elasticsearch>

**GRACIAS  
POR TU  
ATENCIÓN**