



# Análisis de Datos

## INTEGRANTES:

JOHAN QUINATO A

JHOSEF REA

BRYAN QUISAGUANO

ANTONY SANCHEZ

# OBJETIVOS

- Diseñar una arquitectura Data lake.
- Recopilar datos de fuentes de internet referente a los temas dados.
- Almacenar en bases de datos SQL y NoSQL la información recopilada.
- Utilizar elasticsearch como concentrador de datos.
- Realizar un caso de estudio para cada temática.



# INTRODUCCIÓN

# DEFINICION DEL CASO DE ESTUDIO

- Juegos en línea. No cabe duda que en época de pandemia los videojuegos han sido para muchas personas un medio de distracción y relajación dentro del confinamiento, por lo cual los videojuegos se han convertido aún más en el principal medio de entretenimiento, con esto en mente se pretende analizar los datos que se obtengan de países, para así saber: los juegos que son más populares dentro de una serie de países. Además, se obtendrá el país que tiene una mayor producción de videojuegos exitosos.

# DEFINICIÓN DEL CASO DE ESTUDIO

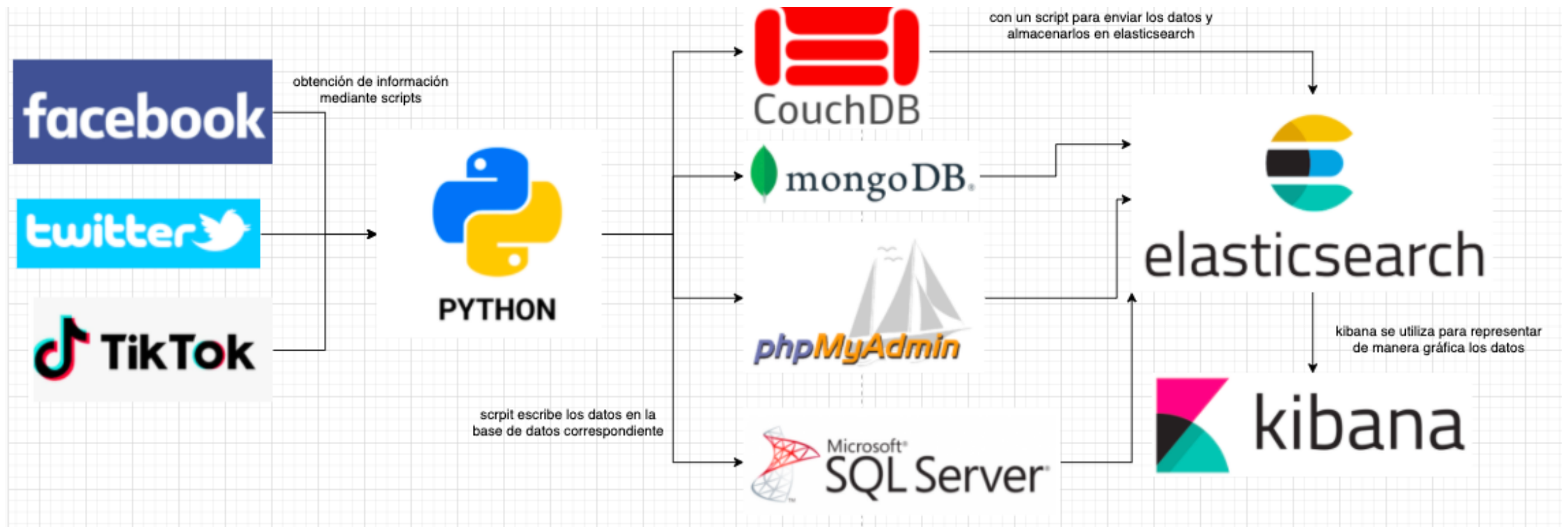
- Pulso político en 20 ciudades del país, se pretende recolectar toda la información posible de los temas políticos que más han sonado en este año, en especial luego de las elecciones presidenciales y de delegados de asambleístas, esto servirá para calcular el nivel de aceptación o conformidad que tienen los habitantes por sus nuevos gobernantes.
- Pulso político en provincias del Ecuador. Debido a que en este año sucedieron las elecciones tanto presidenciales como de representantes asambleístas, los datos recopilados y analizados, permiten evidenciar cuán diferente fue la decisión de los habitantes al momento de acudir a las votaciones Y el candidato o lista mayormente votadas por provincias.

# DEFINICIÓN DEL CASO DE ESTUDIO

- Python. Ya que hoy en día el lenguaje de programación Python es uno de los más populares en el mundo, se pretende obtener datos para así saber que tan grande es la comunidad de América Latina y que tanto este tema es mencionado dentro de las redes sociales.
- Noticias mundiales. Las noticias que causan revuelo internacional son el pan de cada día en emisoras tanto de radio como de televisión, empleando distintas herramientas para pretender saber qué tanto revuelo pueden causar, y que esas diversas noticias sean catalogadas como internacionales dentro de internet específicamente redes sociales.


# RECURSOS Y HERRAMIENTAS UTILIZADAS





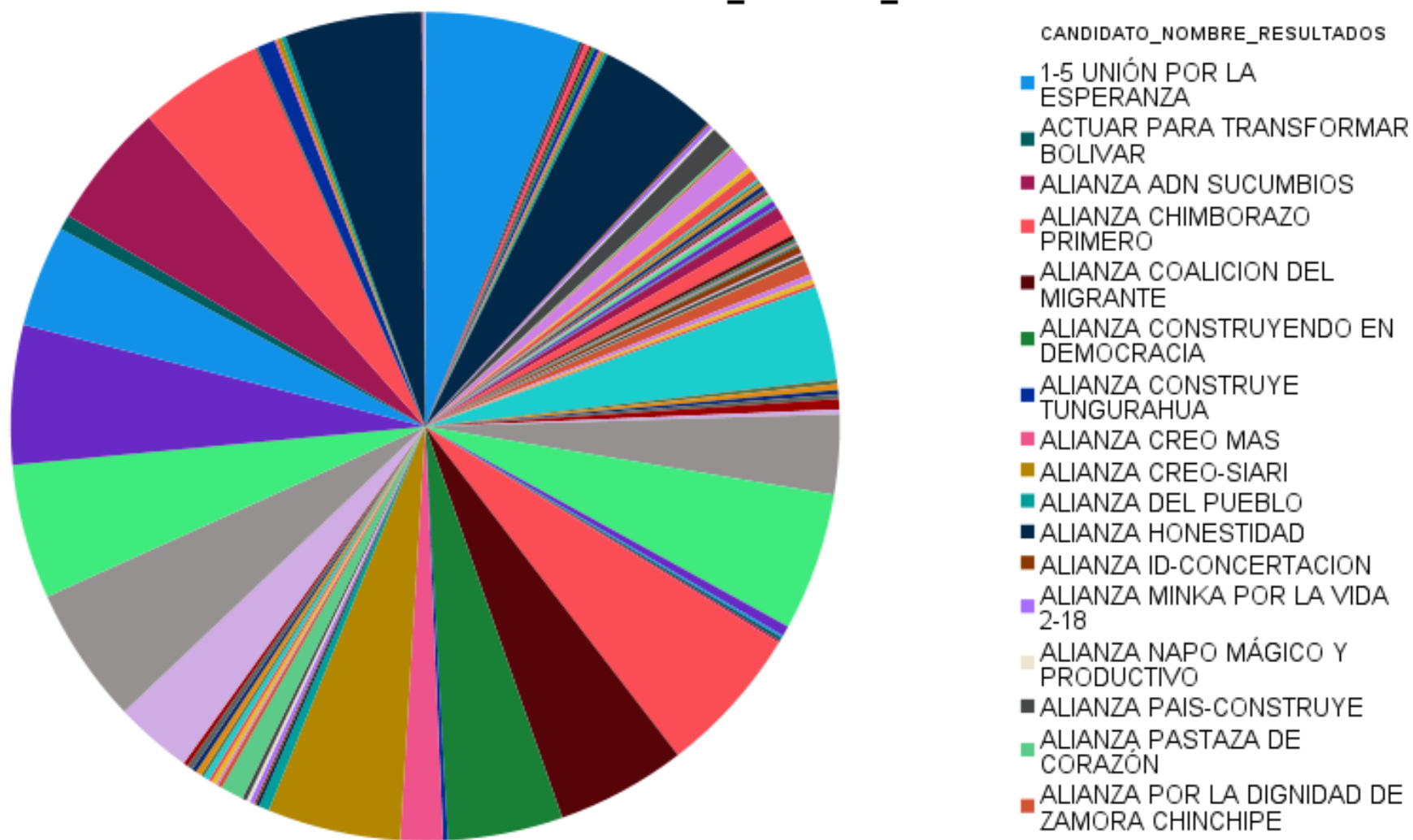
# ARQUITECTURA DE LA SOLUCIÓN



The background of the image is a photograph of the Shanghai skyline, featuring the Oriental Pearl Tower and the Shanghai Tower. A semi-transparent teal rectangle is overlaid on the left side of the image, containing the title text in white.

# PULSO POLITICO EN 20 CIUDADES DEL ECUADOR

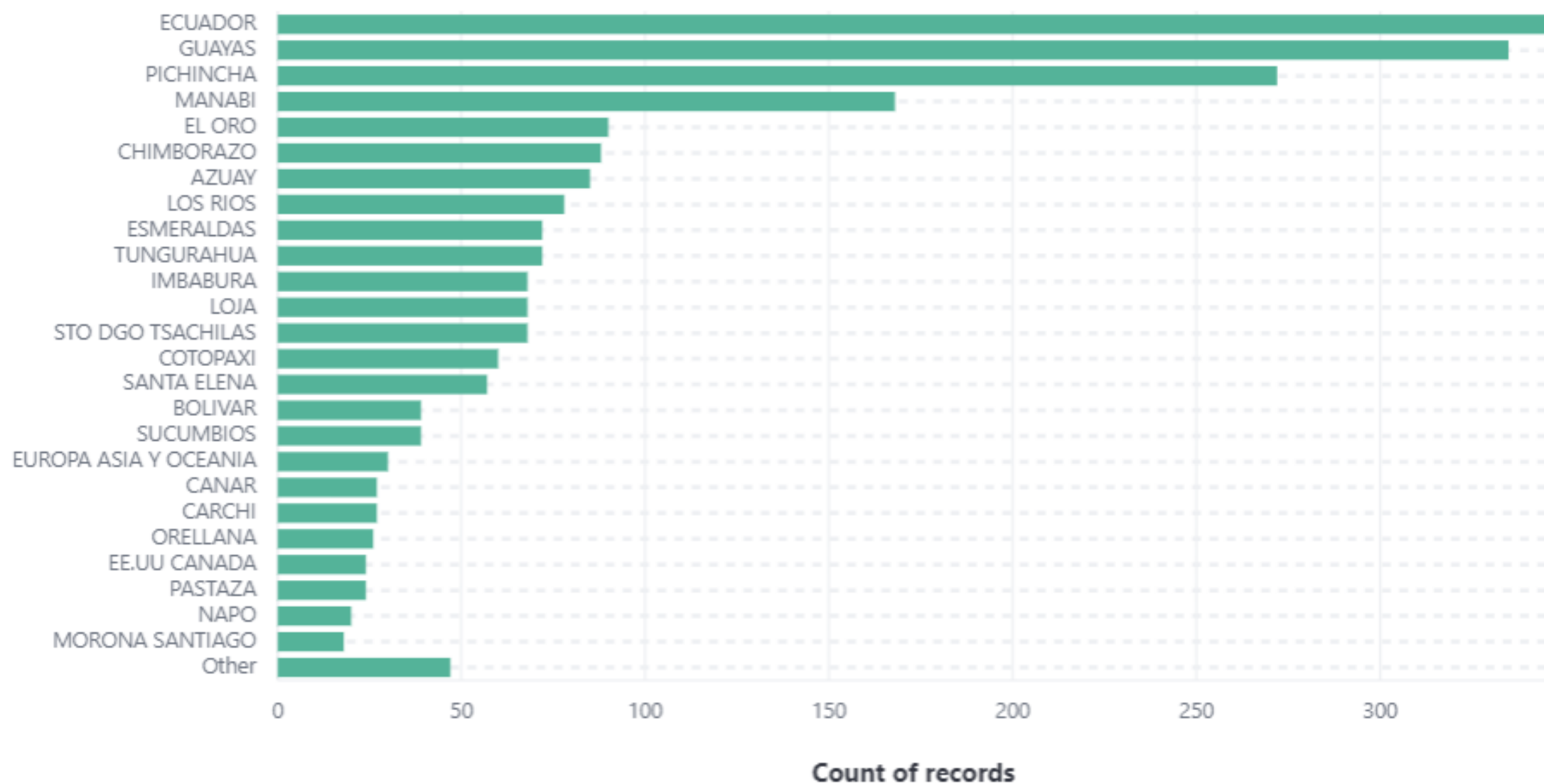
Gráfico circular Recuento de CANDIDATO\_NOMBRE\_RESULTADOS



# PULSO POLITICO POR PROVINCIAS



Top values of candidato\_provincia\_nombre.keyword

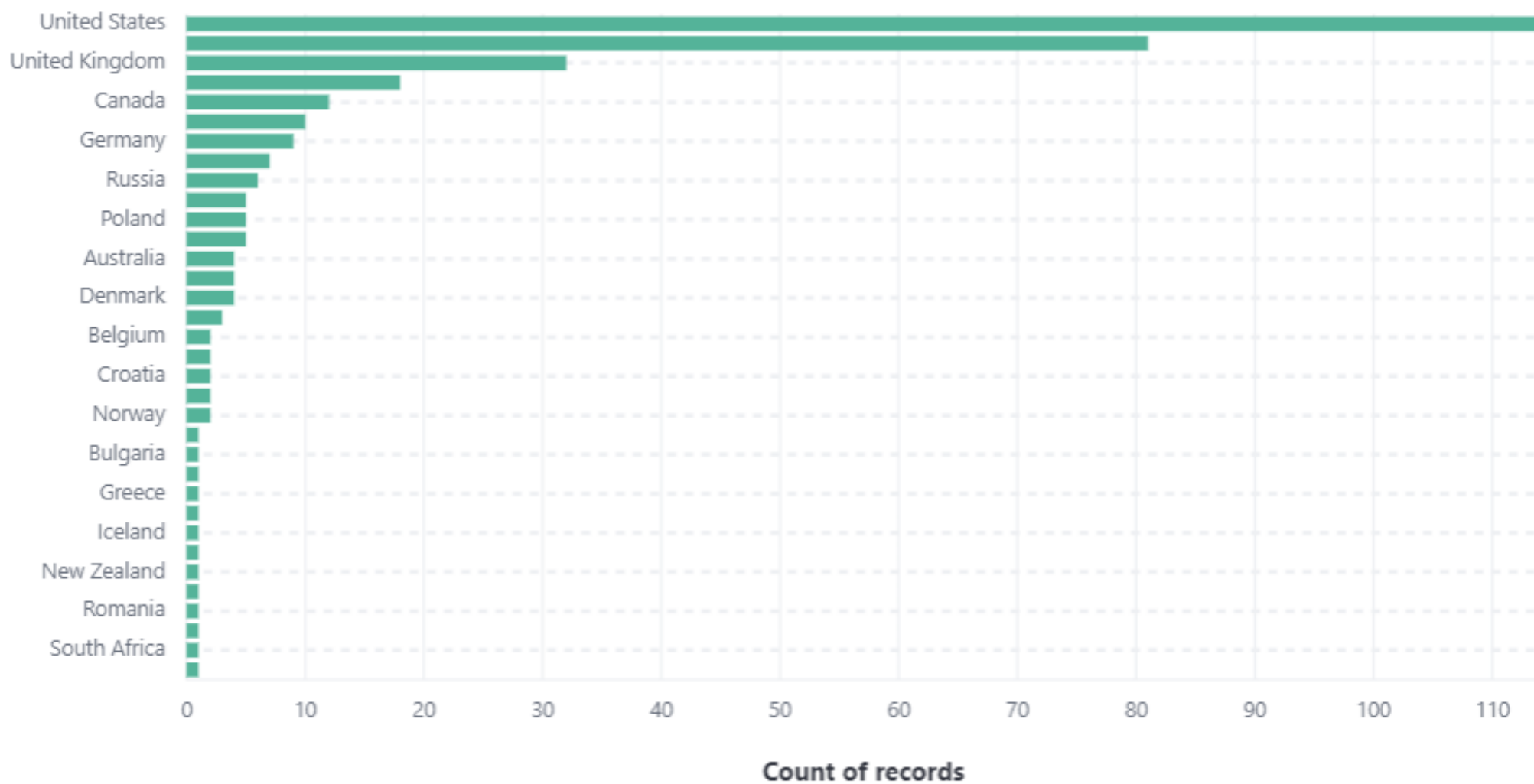




# Juegos en línea por países



Top values of country.keyword



# Eventos o noticias mundiales



+ Add filter

couchnoticias\*

Search field names

Field filters 0

user.is\_translator.keyword

user.lang.keyword

user.listed\_count.keyword

user.location.keyword

user.name.keyword

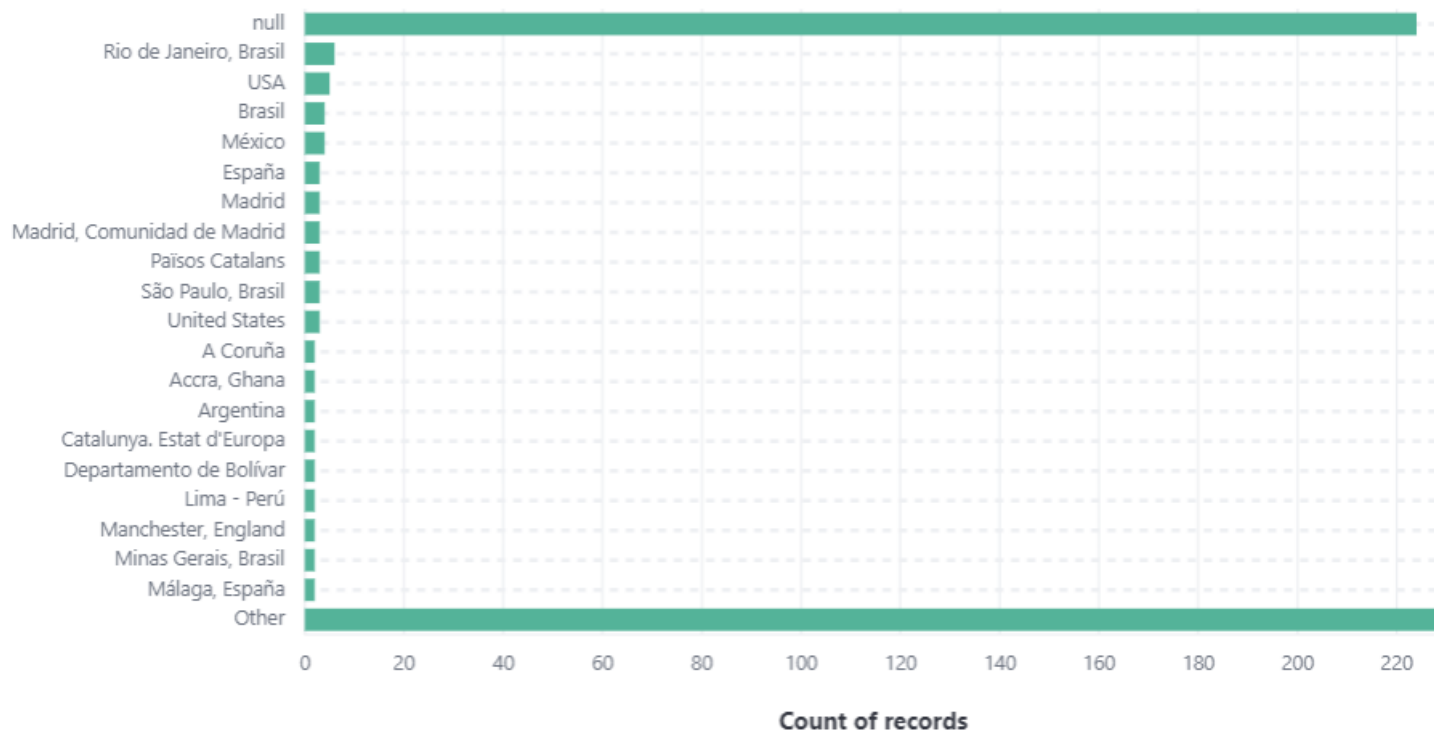
user.notifications.keyword

user.  
profile\_background\_color.  
keyword

user.  
profile\_background\_image

Bar horizontal

Top values of user.location.keyword





# Python

```
for mirror_object in mirror_objects:  
    operation = "MIRROR_X":  
    mirror_mod.use_x = True  
    mirror_mod.use_y = False  
    mirror_mod.use_z = False  
    operation = "MIRROR_Y":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = True  
    mirror_mod.use_z = False  
    operation = "MIRROR_Z":  
    mirror_mod.use_x = False  
    mirror_mod.use_y = False  
    mirror_mod.use_z = True
```

```
selection at the end -add  
mirror_ob.select=1  
context.scene.objects.active = mirror_ob  
("Selected" + str(modifier_name))  
mirror_ob.select = 0  
bpy.context.selected_objects = [mirror_ob]  
data.objects[mirror_ob.name].select = 1  
print("please select exactly one mirror")
```

```
--- OPERATOR CLASSES ---
```

```
types.Operator):  
    X mirror to the selected  
    object.mirror_mirror_x"
```

mongodb\_python\* ▾



111 hits

Sep 16, 2021 @ 14:50:37.229 - Sep 16, 2021 @ 16:50:37.229

Auto ▾

Hide chart

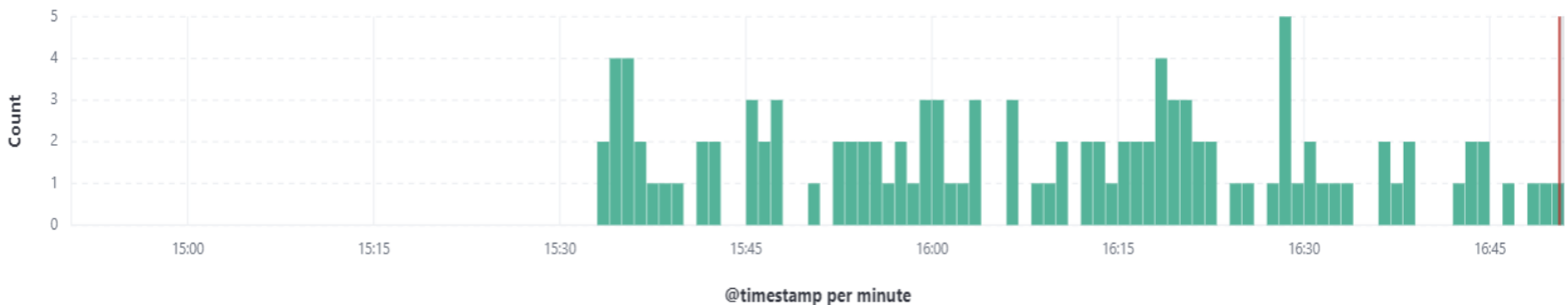
Search field names

Filter by type 0 ▾

Available fields

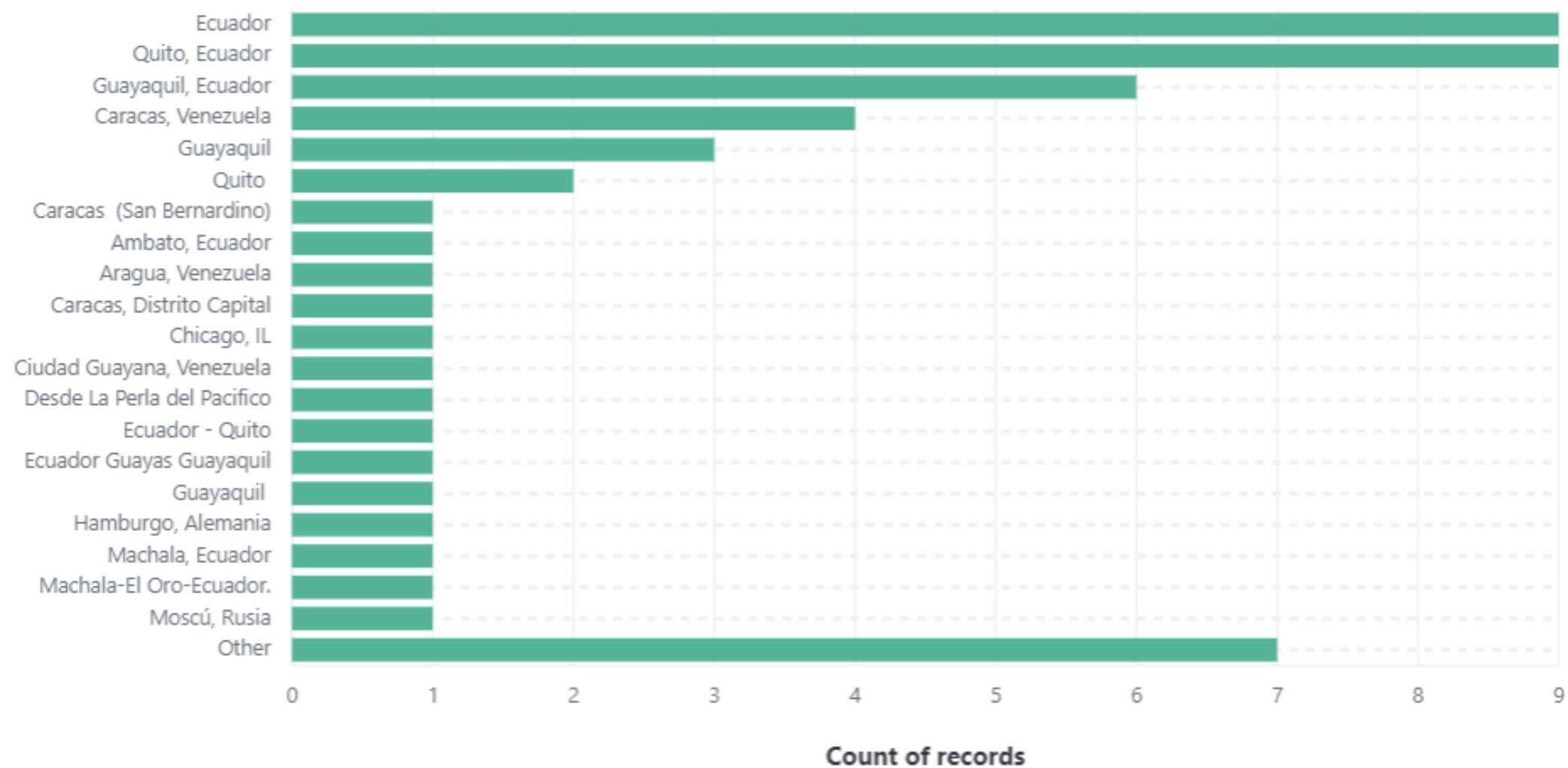
567

- \_id
- \_index
- \_score
- \_type
- @timestamp
- @version
- created\_at
- display\_text\_range
- entities.hashtags.indices
- entities.hashtags.text
- entities.media.display\_url
- entities.media.expanded\_url
- entities.media.id



Time ▾	Document
> Sep 16, 2021 @ 16:50:12.000	<pre>@timestamp: Sep 16, 2021 @ 16:50:12.000 @version: 1 @version.keyword: 1 created_at: Thu Sep 16 21:50:12 +0000 2021 created_at.keyword: Thu Sep 16 21:50:12 +0000 2021 entities.hashtags.indices: 71, 85 entities.hashtags.text: RespondeLasso entities.hashtags.text.keyword: RespondeLasso entities.user_mentions.id: 167,955,344 entities.user_mentions.id_str: 167955344 entities.user_mentions.id_str.keyword: 167955344 entities.user_mentions.indices: 3, 16 entities.user_mentions.name: Gabriel Layana</pre>
> Sep 16, 2021 @ 16:49:54.000	<pre>@timestamp: Sep 16, 2021 @ 16:49:54.000 @version: 1 @version.keyword: 1 created_at: Thu Sep 16 21:49:54 +0000 2021 created_at.keyword: Thu Sep 16 21:49:54 +0000 2021 entities.hashtags.indices: 61, 74, 75, 89 entities.hashtags.text: ParoNacional, RespondeLasso entities.hashtags.text.keyword: ParoNacional, RespondeLasso entities.media.display_url: pic.twitter.com/LG1DGAs11B entities.media.display_url.keyword: pic.twitter.com/LG1DGAs11B entities.media.expanded_url: https://twitter.com/W611Mars1asDw1s/status/1420566705007050870/photo/1</pre>

Top values of user.location.keyword



# DESAFÍOS ENCONTRADOS

- Falta de conocimiento en cuanto a DataLake
- Empleo de API's
- Herramientas necesarias para ELK
- Encontrar métricas adecuadas.

## CONCLUSIONES

- Búsquedas de data
- Con Elasticsearch se obtuvo un manejo de los modelos de los datasets.
- Mediante este análisis de datos se puede obtener diferentes interpretaciones dependiendo de las métricas que se empleen.
- En cuanto a resultados obtenidos, podemos mencionar que en el caso de la política aun habiendo pasado las elecciones las personas siguen hablando de ellos, ya sea a favor o en contra.



- Como experiencia del proyecto, se recomienda utilizar Elasticsearch en la nube; pues generar los procesos de logstash, kibana y cerebro exigen recursos a la máquina.
- Se recomienda que, Elasticsearch no se debería optarla como alternativa a una base de datos relacional.

## RECOMENDACIONES