



ESCUELA POLITÉCNICA NACIONAL
ESCUELA DE FORMACIÓN DE TECNÓLOGOS



BASE DE DATOS MULTIDIMENSIONAL

ASIGNATURA:

Base de datos multidimensional

PROFESOR:

Ing. Zaldumbide Juan

PERÍODO ACADÉMICO:

Oct. 2018 - Mar. 2019

PROYECTO FINAL

TÍTULO:

MINERIA DE DATOS

ESTUDIANTE

Iza Jorge
Coro Andres

FECHA DE REALIZACIÓN:

FECHA DE ENTREGA: 05 de febrero de 2019

CALIFICACIÓN OBTENIDA:

FIRMA DEL PROFESOR:

Contenido

1	Definición del caso de estudio	2
2	Objetivos Generales y específicos	2
3	Descripción del equipo de trabajo y actividades realizadas por cada uno	2
4	Cronograma de actividades (diagrama de Gantt).....	1
5	Recursos y herramientas utilizadas	1
6	Arquitectura de la solución	1
7	Extracción de datos	1
8	Análisis de información	5
9	Visualización de información	16
10	Resultados obtenidos.....	18
11	Conclusiones y Recomendaciones	20
12	Desafíos y problemas encontrados	20

1 Definición del caso de estudio

El presente proyecto tiene como finalidad realizar la cosecha o recolección de datos de Twitter, analizar los datos obtenidos, y visualizarlos mediante dashboard en la web, con respecto al pulso político ecuatoriano actual, el cual nos permita tener una idea de cuales podrían ser los resultados reales en las votaciones electorales para alcaldes en la ciudad de Quito, Cuenca y Guayaquil que se celebrarán el día 24 de marzo de 2019. Para ello lo que se considera es la popularidad que tiene cada uno de los alcaldes según el número de veces que son nombrados en tweets. De la misma forma se pretende realizar una cosecha, extracción y visualización de los obtenidos de Twitter relacionados con los sismos que se han suscitado recientemente en el mundo, aplicando todos los conocimientos obtenidos a lo largo de la clase de base de datos Multidimensional y utilizando las herramientas de ELK (“Elasticsearch, Logstash y Kibana”).

2 Objetivos Generales y específicos

OBJETIVO PRINCIPAL

El objetivo general del proyecto es la extracción, análisis y visualización de la información social accesible a través de Twitter relacionada con pulso político ecuatoriano actual mediante la utilización de herramientas ELK

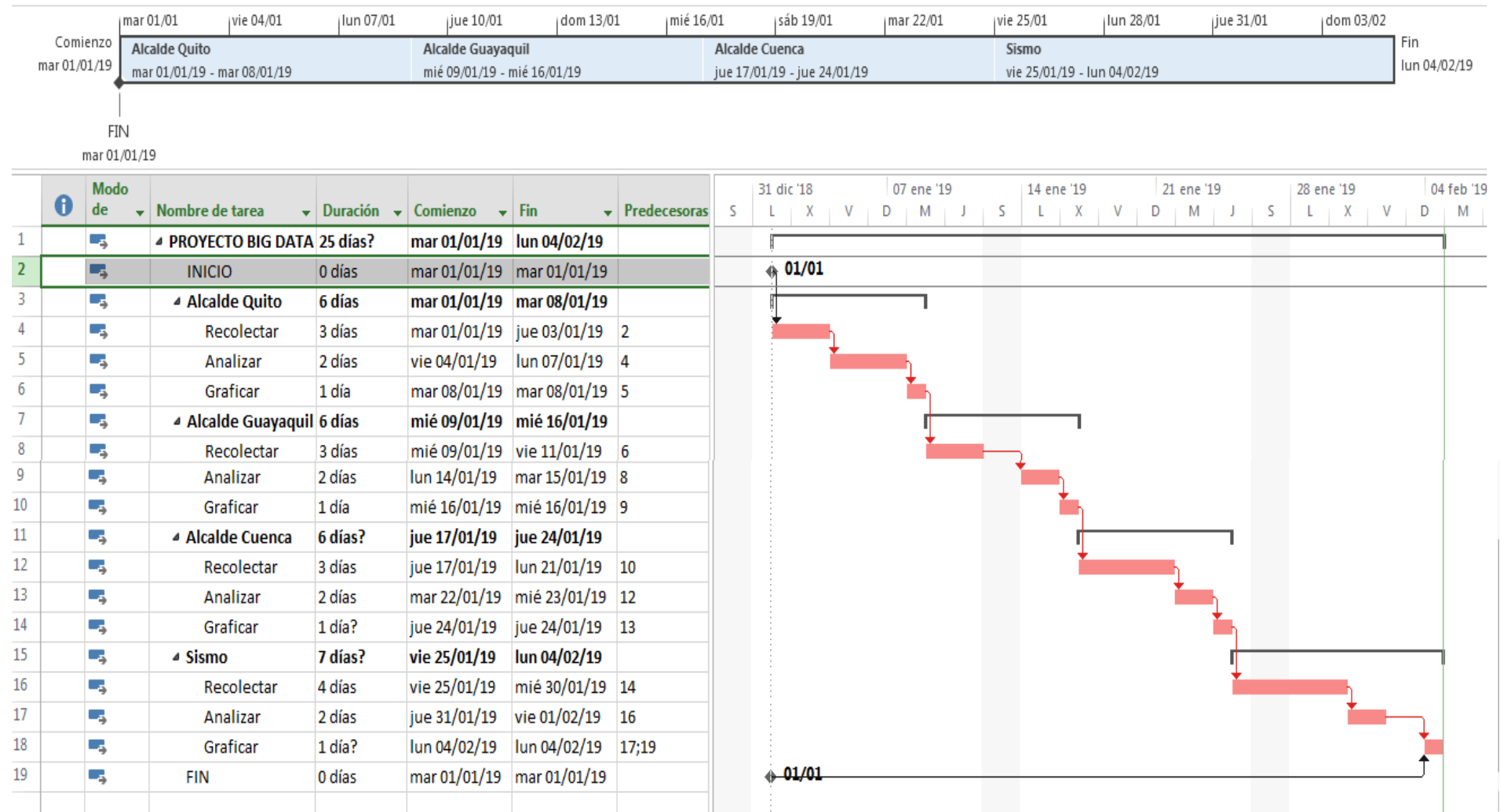
OBJETIVO ESPECIFICOS:

- Recopilar datos de Twitter relacionados con los candidatos a la alcaldía de Quito, Guayaquil y Cuenca usando la API de Twitter, Python y la librería Tweepy.
- Recopilar datos de Twitter relacionados con los sismos que se han suscitado actualmente usando la API de Twitter, Python y la librería Tweepy.
- Realizar un mapeo correcto de los datos obtenidos de twitter en especial las que son de tipo date y geo_point.
- Crear una arquitectura de datos que nos permita agrupar distintos índices del pulso político ecuatoriano en una sola base de datos de couch y posterior unificarla en elasticshare
- Crear graficas en Kibana que me permitan analizar y visualizar los datos obtenidos mediante la cosecha de tweets.
- Visualizar las gráficas mediante dashboard en la web

3 Descripción de equipo de trabajo y actividades realizadas por cada uno

	Minería de ciudad Quito	Minería de ciudad Guayaquil	Minería de ciudad Cuenca	Minería de sismos
Coro Andrés	X	X		
Iza Jorge			X	X

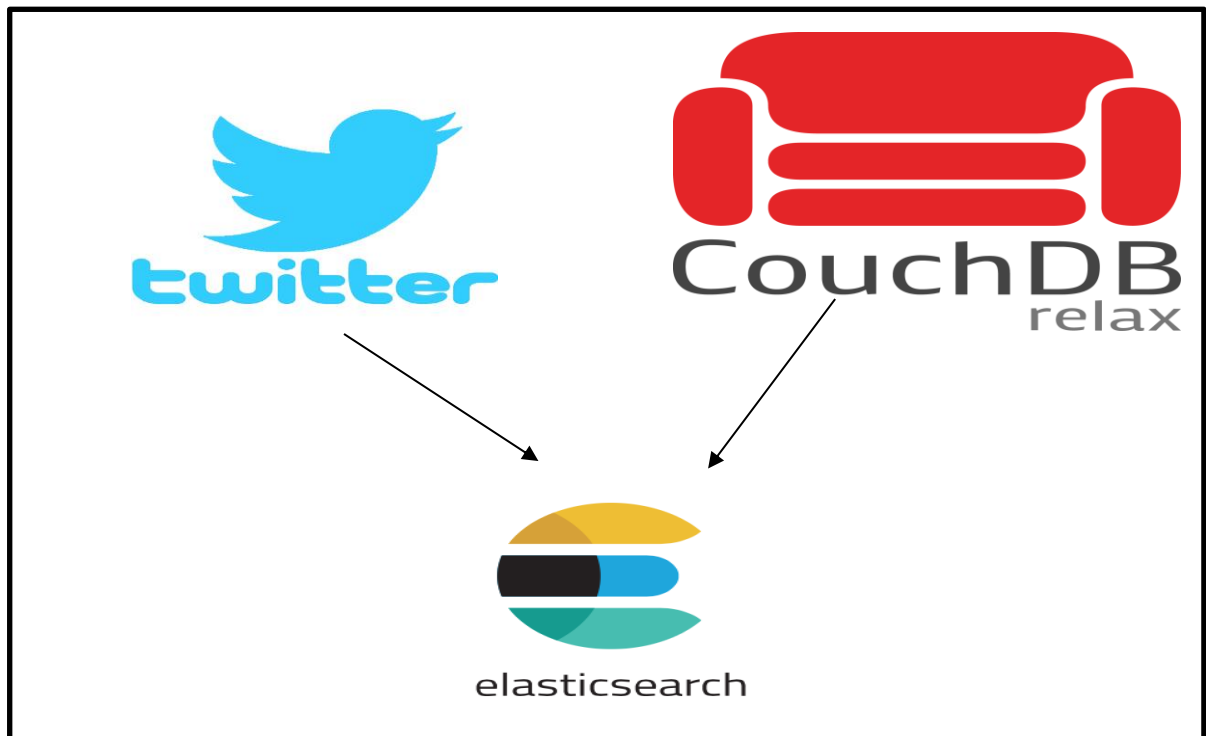
4 Cronograma de actividades (diagrama de Gantt)



5 Recursos y herramientas utilizadas

- Subline
- Python 5.0 64 bits
- notePad ++
- elasticsearch 6.5.4
- kibana 6.5.4
- logstash 6.5.4
- cerebro 0.8.1

6 Arquitectura de la solución



7 Extracción de datos

En este punto importamos las librerías de Twitter que nos permitirán recolectar datos de la appi de Twitter

```
1 import json
2 from tweepy.streaming import StreamListener
3 from tweepy import OAuthHandler
4 from tweepy import Stream
5 from textblob import TextBlob
6 from elasticsearch import Elasticsearch
```

En esta clase TweetStreamListener procedemos a leer los datos que estamos recolectando de Twitter, y guardamos en una variable llamada “dict_data” el cual nos permitirá acceder a la estructura que tiene cada dato recolectado.

```
class TweetStreamListener(StreamListener):

    # on success
    def on_data(self, data):

        lon = 0
        lat = 0

        # decode json
        dict_data = json.loads(data)
```

Una vez obtenido la estructura de los datos que estamos recolectando procedemos a buscar dentro de nuestra “dict_data” todos los valores que estén dentro de esta variable llamada “created_at” que es la fecha de cuando se a recolectado el dato.

```
try:
    fecha = dict_data["created_at"]
    print(obtenerFecha(fecha))
    formatoFecha = obtenerFecha(fecha)
    geo = dict_data["coordinates"]["coordinates"]
    print(geo)
    lon = geo[0]
    lat = geo[1]
```

Para dar formato al dato fecha procedemos a usar la función “Obtener fecha” en el cual mediante su código nos permite cambiar el formato de mes que se encuentra en número “(01)” a “Jan” que es el formato que nos permite subir elelasticshare las fechas. De la misma forma se procede a dar el formato a los datos que son de tipo geo_point obteniendo son latitud y longitud.

A continuación, se muestra el formato que nos permite ingresar los datos de tipo geo_point

```
PUT my_index
{
  "mappings": {
    "_doc": {
      "properties": {
        "location": {
          "type": "geo_point"
        }
      }
    }
  }
}

PUT my_index/_doc/4
{
  "text": "Geo-point as an array",
  "location": [ -71.34, 41.12 ] ④
}
```

continuación, se muestra el formato de fecha que nos acepta elasticshare.

```

PUT my_index
{
  "mappings": {
    "_doc": {
      "properties": {
        "date": {
          "type": "date" ❶
        }
      }
    }
  }
}

PUT my_index/_doc/1
{ "date": "2015-01-01" } ❷

```

```

def obtenerFecha (fecha):
    listaFecha = fecha.split(' ')
    #print(listaFecha[1])
    mes = " "
    if listaFecha[1] == "Jan":
        mes = "01"
    elif listaFecha[1] == "Feb":
        mes = "02"
    elif listaFecha[1] == "Mar":
        mes = "03"
    elif listaFecha[1] == "Apr":
        mes = "04"
    elif listaFecha[1] == "May":
        mes = "05"
    elif listaFecha[1] == "Jun":
        mes = "06"
    elif listaFecha[1] == "Jul":
        mes = "07"
    elif listaFecha[1] == "Aug":
        mes = "08"
    elif listaFecha[1] == "Sep":
        mes = "09"
    elif listaFecha[1] == "Oct":
        mes = "10"
    elif listaFecha[1] == "Nov":
        mes = "11"
    else:
        mes = "12"

    año=listaFecha[5]
    mes = mes
    día = listaFecha[2]
    nuevaFecha = año+"-"+mes+"-"+día

    return nuevaFecha

```

Después procedemos a crear el mapeo de los datos en el cual indicamos los nombres de las variables que tendrá nuestra estructura de datos recolectados como son el “autor”, la fecha “date”, el mensaje, la locación “location”. ctr

```

es.index(index="sentiment",
        doc_type="test-type",
        body={"author": dict_data["user"]["screen_name"],
              "date": formatoFecha,
              "message": dict_data["text"],
              "location": [lon,lat],
              "geo": dict_data["user"]["location"],
              "polarity": tweet.sentiment.polarity,
              "subjectivity": tweet.sentiment.subjectivity,
              "sentiment": sentiment})

return True

```

Filtrado de palabras para los candidatos de la alcaldía de Quito, Guayaquil, Cuenca

QUITO

```
# search twitter for "Ecuador" keyword
stream.filter(track=['LoroHomer0', 'PacoMoncayo', 'CesarMontufar51', 'PaolaVintimilla', 'juancaholguin'])
#stream.filter(locations=[-78.586922,-0.395161,-78.274155,0.021973])
```

CUENCA

```
# search twitter for "Ecuador" keyword
stream.filter(track=['jeffersonperezq', 'MarceloHCabrera', 'PaulCarrascoC', 'fcorderoc', 'EstebanBernalB'])
#stream.filter(locations=[-78.586922,-0.395161,-78.274155,0.021973])
```

GUAYAQUIL

```
# search twitter for "Ecuador" keyword
stream.filter(track=['CynthiaViteri6', 'jimmyjairala', 'Simon_BolivarEC', 'abdalabucaran', 'fcojimenez21'])
#stream.filter(locations=[-78.586922,-0.395161,-78.274155,0.021973])
```

ARCHIVO DE MAPEO

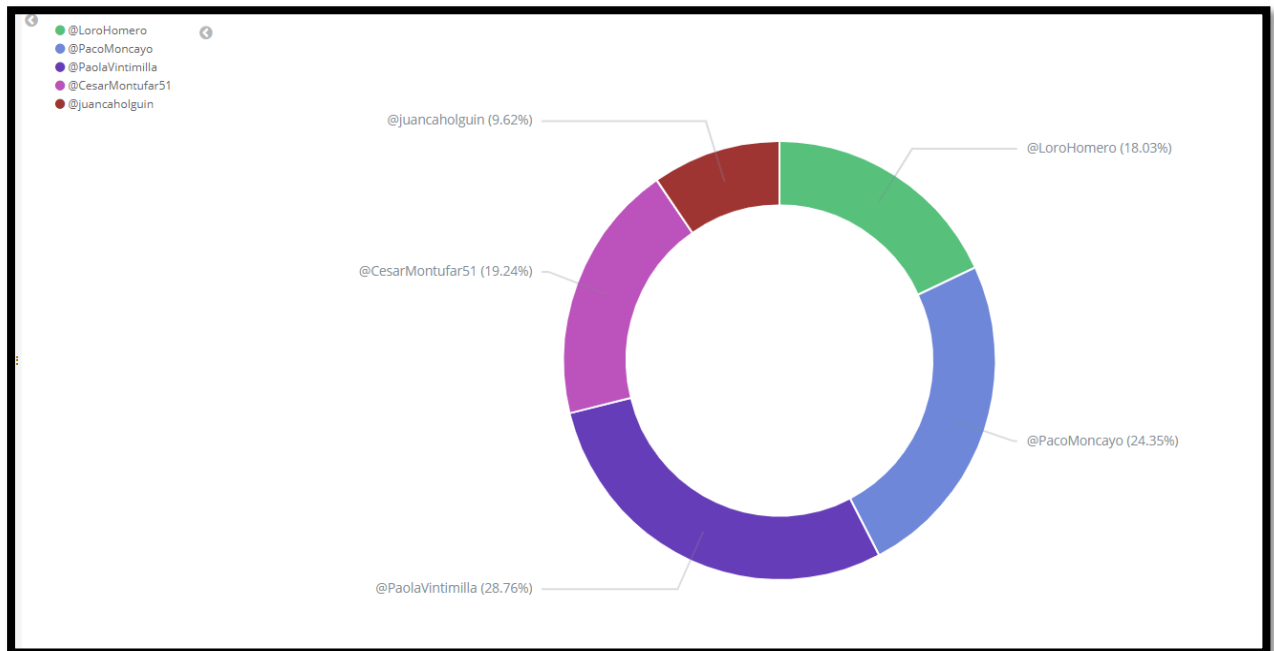
```
{
  "mappings": {
    "test-type": {
      "properties": {
        "author": {
          "type": "text",
          "fields": {
            "keyword": {
              "type": "keyword",
              "ignore_above": 256
            }
          }
        },
        "date": {
          "type": "date"
        },
        "location": {
          "type": "geo_point"
        },
        "message": {
          "type": "text",
          "fields": {
            "keyword": {
              "type": "keyword",
              "ignore_above": 256
            }
          }
        },
        "polarity": {
          "type": "float"
        },
        "sentiment": {
          "type": "text",
          "fields": {
            "keyword": {
              "type": "keyword",
              "ignore_above": 256
            }
          }
        },
        "subjectivity": {
          "type": "float"
        }
      }
    }
  }
}
```


8 Análisis de información

1. Pulso político en 5 ciudades de Ecuador próximas votaciones

1.1. Pulso político en la ciudad de Quito próximas votaciones.

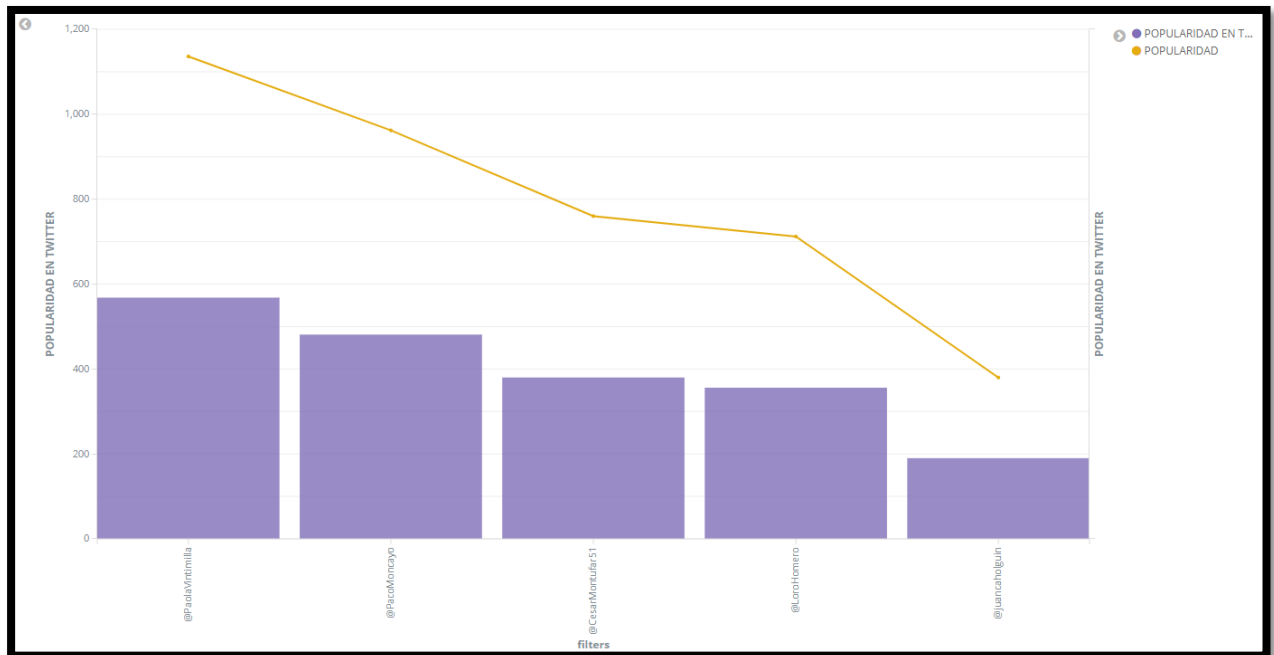
1.1.1. Grafica de pastel del nivel de popularidad de los candidatitos a la alcaldía de Quito:



Conclusión: Mediante los datos obtenidos en la gráfica de pastel llamada “Grafica de pastel del nivel de popularidad de los candidatitos a la alcaldía de Quito: ” nos podemos dar una idea de cual es el nivel de popularidad de los 5 candidatos que se encuentra postulando para la alcaldía de Quito referente al número de veces que un usuario de Twitter posteo un Tweet en el cual se nombrara a uno de los candidatos o ha varios, de la cual se obtiene los siguientes datos:

- Jorge Yunda “@LoroHomero” tiene un nivel de popularidad del 18.03% Twitter
- Paco Moncayo “@PacoMoncayo” tiene un nivel de popularidad del 24.35% en Twitter
- Paola Veintimilla “@PaolaVeintimilla” tiene un nivel de popularidad del 28.76%, en Twitter
- Cesar Montufar “@CesarMontufar51” tiene un nivel de popularidad del 19.24% en Twitter
- Juan Carlos Holgin “@juancaholguin” tiene un nivel de popularidad de 9.62% en Twitter

1.1.2. Grafica en barras del nivel de popularidad de los candidatitos a la alcaldía en Quito



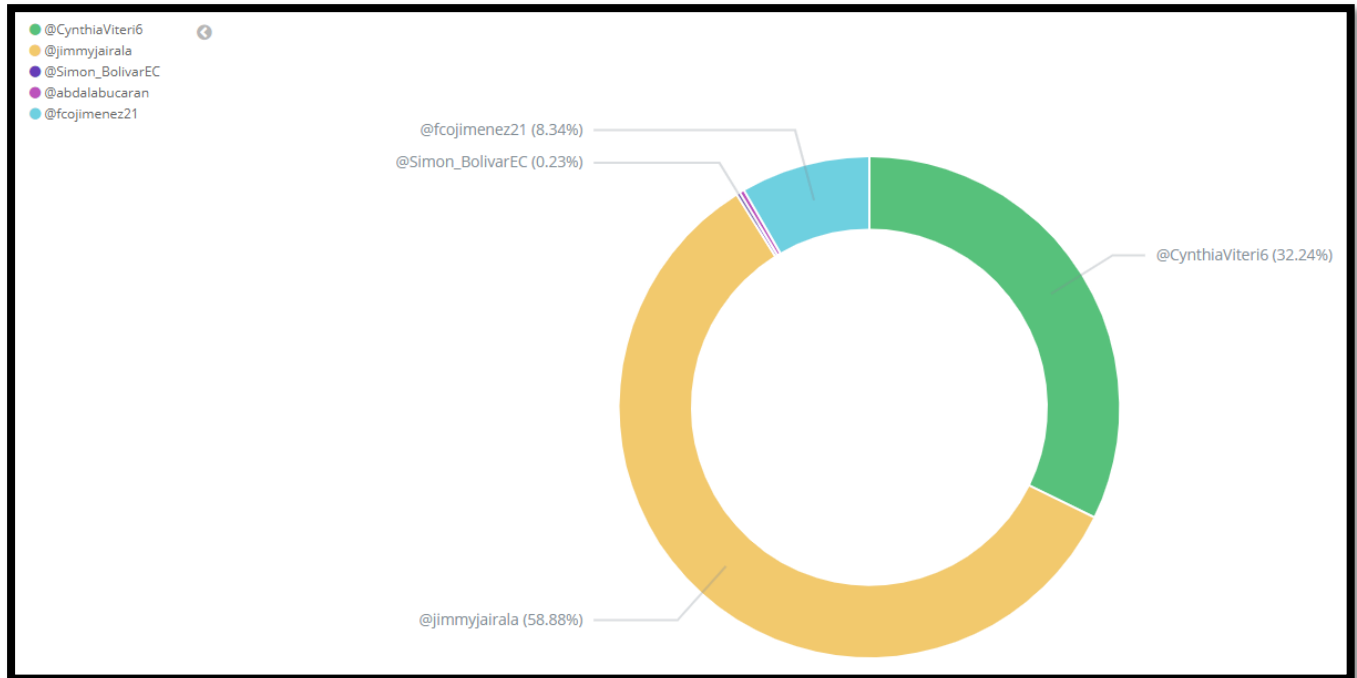
Conclusión: Mediante la gráfica en barras llamada “Grafica en barras del nivel de popularidad de los candidatitos a la alcaldía en Quito” se puede observar claramente de manera ordenada (mayor a menor) el nivel de popularidad que tienen los candidatos a la alcaldía de Quito en la red social de Twitter. Sabiendo que los datos obtenidos están basados en el número de veces que un usuario de Twitter escribió algún tweet nombrando a alguno o varios de los candidatos.

Observando los datos de esta forma nosotros podemos darnos una idea de cual podrían ser los resultados reales en las elecciones para la alcaldía de Quito el 24 de marzo de 2019, o de la misma forma apreciamos cuales son los 2 primeros posibles candidatos que estarán peleando a la par por el puesto de alcalde como por ejemplo en la gráfica nos muestra que esta entre los candidatos Paola Veintimilla y Paco Moncayo.

- Paola Veintimilla “@PaolaVeintimilla” con 578 tweets
- Paco Moncayo “@PacoMoncayo” con 481 tweets
- Cesar Montufar “@CesarMontufar51” con 380 tweets
- Jorge Yunda “@LoroHormero” con 356 tweets
- Juan Carlos Holguin “@juancaholguin” con 190 tweets

1.2. Pulso político en la ciudad de Guayaquil próximas votaciones.

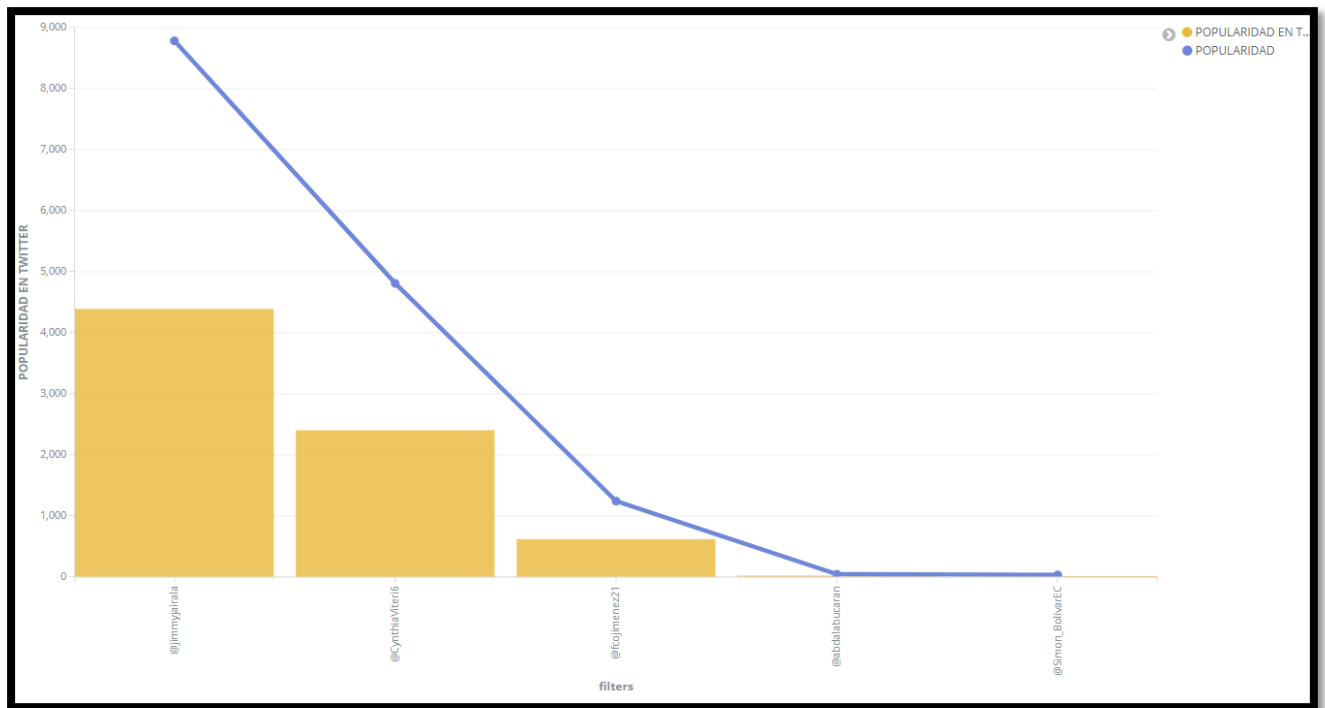
1.2.1. Grafica de pastel del nivel de popularidad de los candidatitos a la alcaldía de Guayaquil:



Conclusión: Mediante los datos obtenidos en la gráfica de pastel llamada “Grafica de pastel del nivel de popularidad de los candidatitos a la alcaldía de Guayaquil: ” nos podemos dar una idea de cuál es el nivel de popularidad de los 5 candidatos que se encuentra postulando para la alcaldía de Guayaquil referente al número de veces que un usuario de Twitter posteo un Tweet en el cual se nombrara a uno de los candidatos o a varios, de la cual se obtiene los siguientes datos:

- Cynthia Viteri “@CynthiaViteri6” tiene un nivel de popularidad del 32.24% en Twitter
- Jimmy Jairala “@jimmyjairala” tiene un nivel de popularidad del 58.88% en Twitter
- Simon Bolivar “@Simon_BolivarEC” tiene un nivel de popularidad del 0.23%, en Twitter
- Abdala Bucaran “@abdalabucaran” tiene un nivel de popularidad del 0.31% en Twitter
- Francisco Jimenez “@fcojimenez21” tiene un nivel de popularidad de 8.34% en Twitter

1.2.2. Grafica en barras del nivel de popularidad de los candidatitos a la alcaldía en Guayaquil



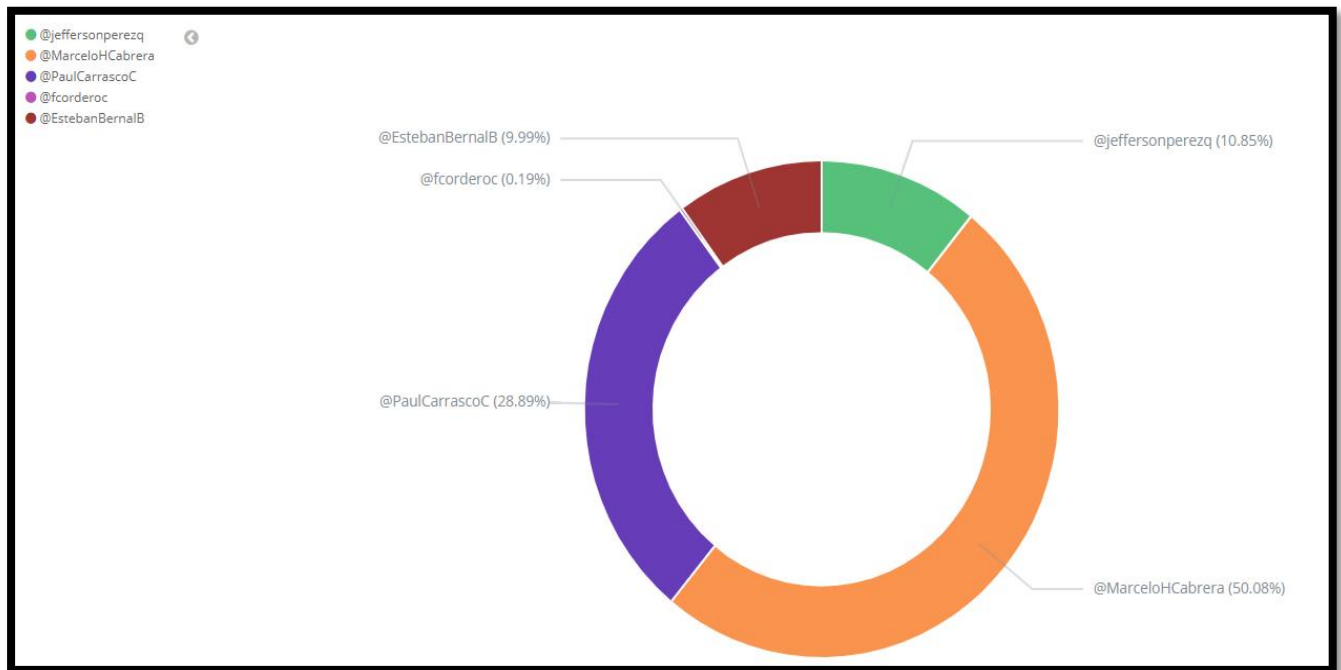
Conclusión: Mediante la gráfica en barras llamada “Grafica en barras del nivel de popularidad de los candidatitos a la alcaldía en Guayaquil” se puede observar claramente de manera ordenada (mayor a menor) el nivel de popularidad que tienen los candidatos a la alcaldía de Guayaquil en la red social de Twitter. Sabiendo que los datos obtenidos están basados en el número de veces que un usuario de Twitter escribió algún tweet nombrando a alguno o varios de los candidatos.

Observando los datos de esta forma nosotros podemos darnos una idea de cual prodrian ser los resultados reales en las elecciones para la alcaldía de Guayaquil el 24 de marzo de 2019, o de la misma forma apreciamos cuales son los 2 primeros posible candidaros que estaran peleando a la par por el puesto de alcalde como por ejemplo en la grafica nos muestra que esta entre los candidatos Jimmy Jairala y Chyntia Viteri.

- Jimmy Jairala “@jimmyjairala” con 4392 tweets
- Cynthia Viteri “@CynthiaViteri6” con 2405 tweets
- Francisco Jimenez “@fcojimenez21” con 622 tweets
- Abdala Bucaran “@abdalabucaran” con 22 tweets
- Simon Bolivar “@Simon_BolivarEC” con 17 tweets

1.3. Pulso político en la ciudad de Guayaquil próximas votaciones.

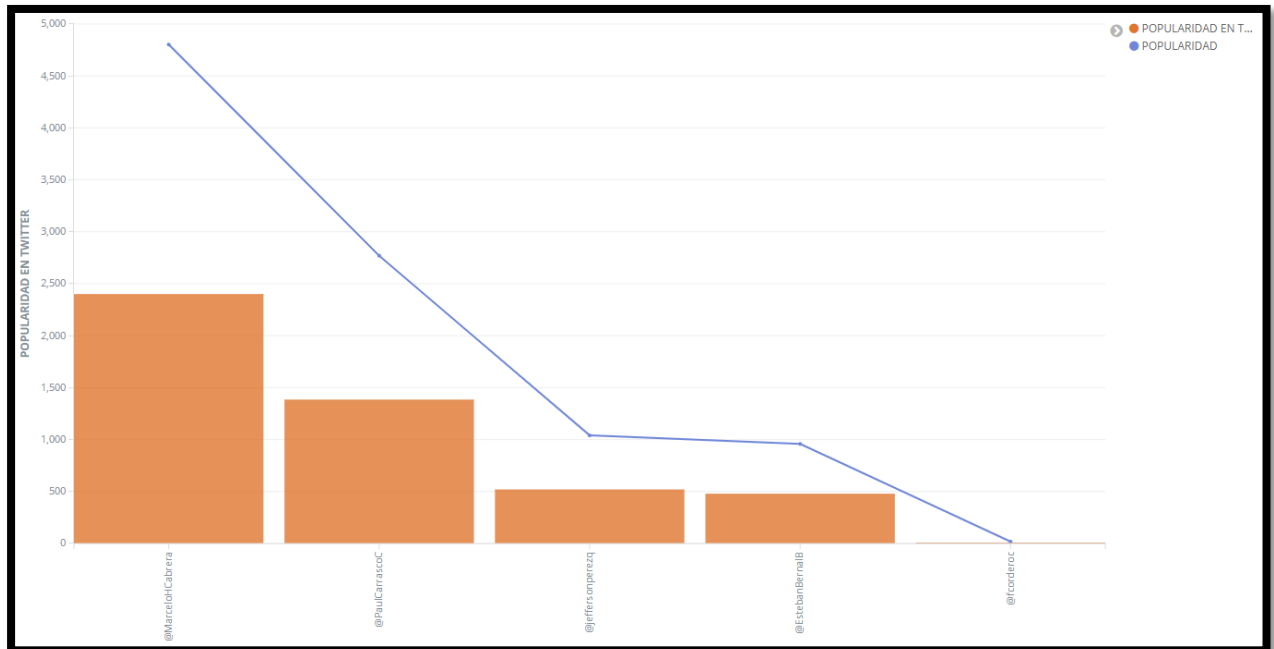
1.3.1. Grafica de pastel del nivel de popularidad de los candidatitos a la alcaldía de Cuenca:



Conclusión: Mediante los datos obtenidos en la gráfica de pastel llamada “Grafica de pastel del nivel de popularidad de los candidatitos a la alcaldía de Cuenca:” nos podemos dar una idea de cuál es el nivel de popularidad de los 5 candidatos que se encuentra postulando para la alcaldía de Cuenca referente al número de veces que un usuario de Twitter posteo un Tweet en el cual se nombrara a uno de los candidatos o a varios, de la cual se obtiene los siguientes datos:

- Jefferson Perez “@jeffersonperezq” tiene un nivel de popularidad del 10.85% en Twitter
- Marcelo Cabrera “@MarceloHCabrera” tiene un nivel de popularidad del 50.08% en Twitter
- Paul Carrasco “@PaulCarrascoC” tiene un nivel de popularidad del 28.89%, en Twitter
- Fernando Cordero “@fcorderoc” tiene un nivel de popularidad del 0.19% en Twitter
- Esteban Bernal “@EstebanBernalB” tiene un nivel de popularidad de 9.99% en Twitter

1.3.2. Grafica en barras del nivel de popularidad de los candidatitos a la alcaldía en Cuenca



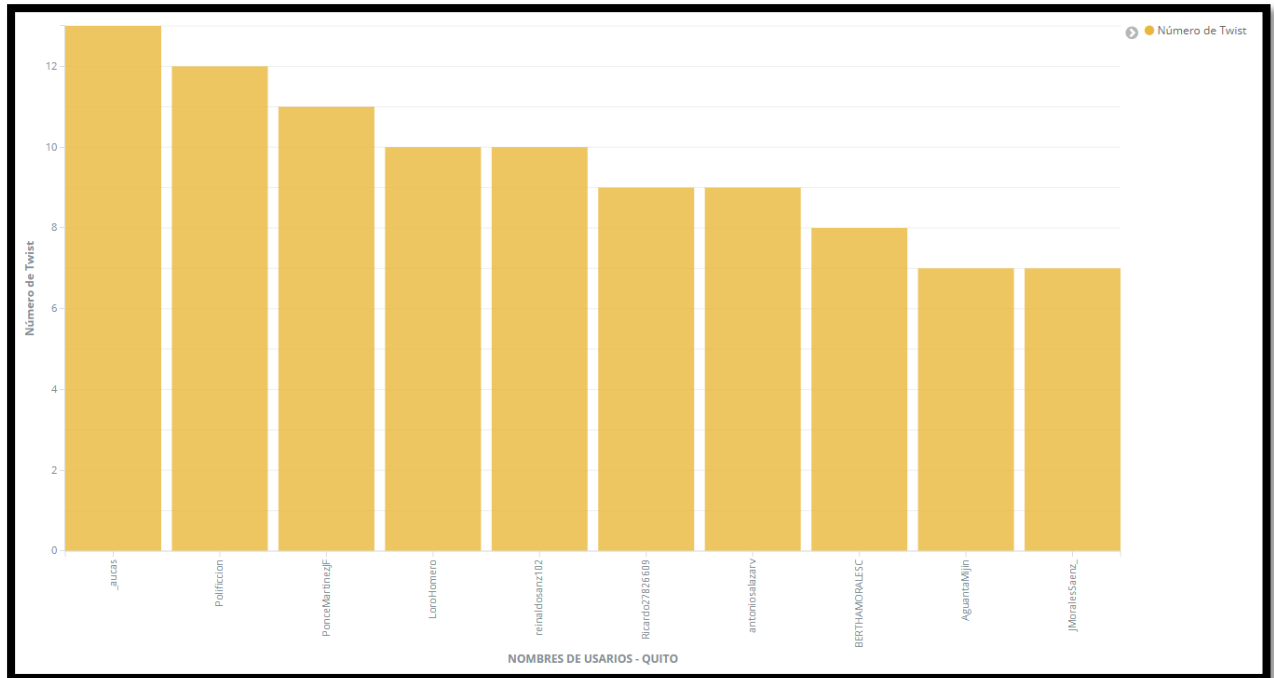
Conclusión: Mediante la gráfica en barras llamada “Grafica en barras del nivel de popularidad de los candidatitos a la alcaldía en Cuenca” se puede observar claramente de manera ordenada (mayor a menor) el nivel de popularidad que tienen los candidatos a la alcaldía de Cuenca en la red social de Twitter. Sabiendo que los datos obtenidos están basados en el número de veces que un usuario de Twitter escribió algún tweet nombrando a alguno o varios de los candidatos.

Observando los datos de esta forma nosotros podemos darnos una idea de cual prodrian ser los resultados reales en las elecciones para la alcaldía de Cuenca el 24 de marzo de 2019, o de la misma forma apreciamos cuales son los 2 primeros posible candidatos que estaran peleando a la par por el puesto de alcalde como por ejemplo en la grafica nos muestra que esta entre los candidatos Marcelo Cabrera y Paul Carrasco.

- Marcelo Cabrera “@MarceloHCabrera” con 2401 tweets
- Paul Carrasco “@PaulCarrascoC” con 1385 tweets
- Jefferson Perez “@jeffersonperezq” con 520 tweets
- Esteban Bernal “@EstebanBernalB” con 479 tweets
- Fernando Cordero “@fcorderoc” con 9 tweets

2. Top 10 twitteros en 5 ciudades de Ecuador.

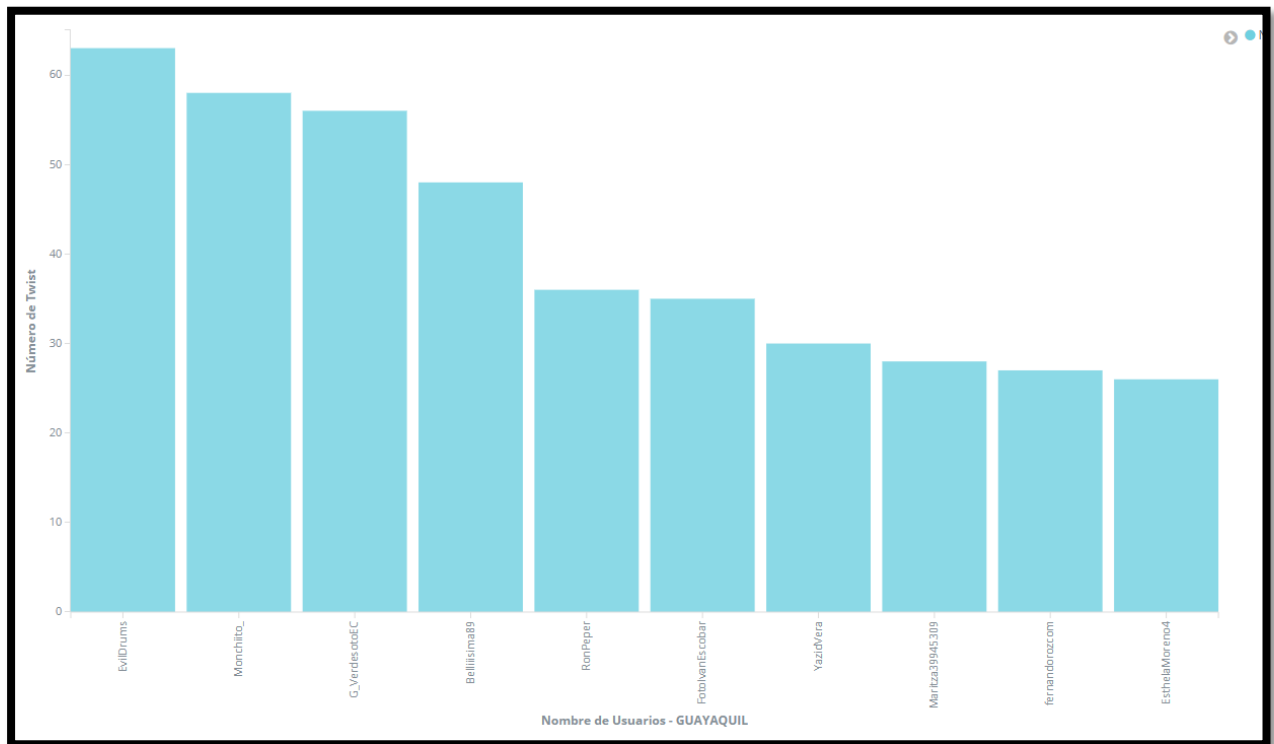
2.1. Top 10 twitteros en la ciudad de Quito



Mediante la gráfica de barras llamada “Top 10 twitteros en la ciudad de Quito” se puede observar en orden descendente el top de 10 twitteros en la ciudad de Quito, referente al número de veces que a twitteado un tweet con algún nombre de los candidatos a la alcaldía de quito nombrados anteriormente. Mediante esta tabla obtenemos los siguientes datos:

1. _aucas con 13 tweets
2. Polificacion con 12 tweets
3. PonceMartinezJF con 10 tweets
4. LoroHomero con 10 tweets
5. ReinaldoSaenz102 con 10 tweets
6. Ricardo278266 con 9 tweets
7. Antoniosalazarv con 9 tweets
8. BERTHAMORALESC con 8 tweets
9. AguantaMijin con 7 tweets
10. JMoralesSaenz con 7 tweets

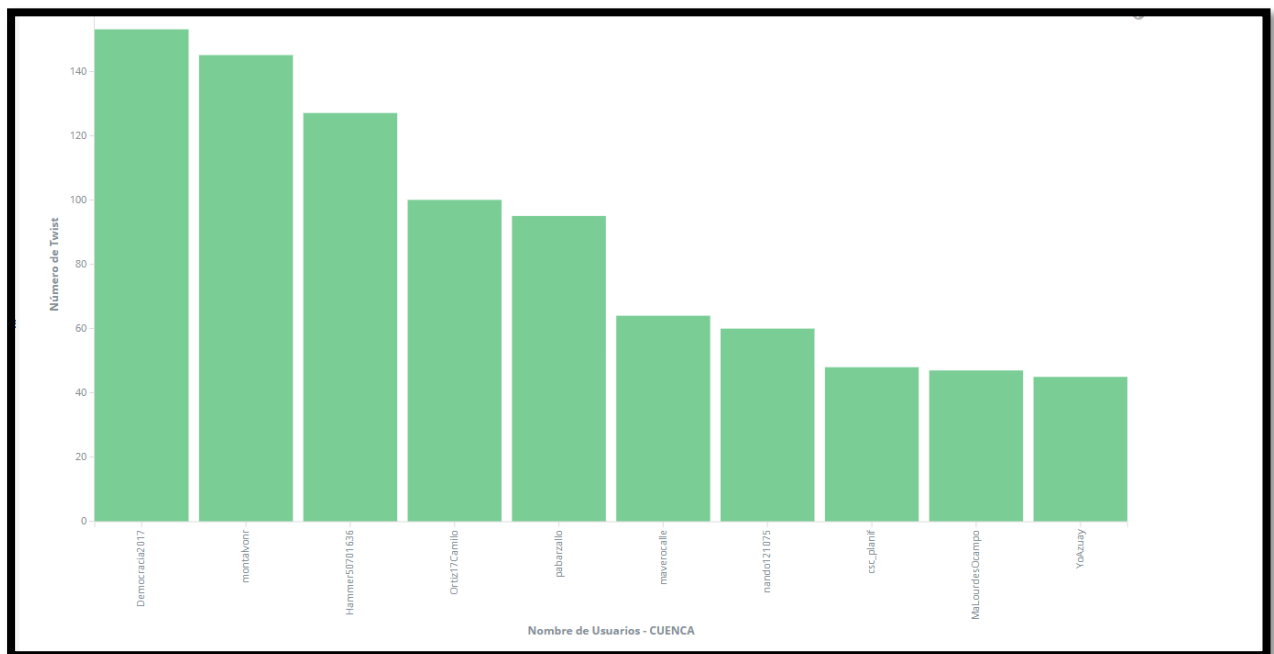
2.2.Top 10 twitteros en la ciudad de Guayaquil



Mediante la gráfica de barras llamada “Top 10 twitteros en la ciudad de Guayaquil” se puede observar en orden descendente el top de 10 twitteros en la ciudad de Guayaquil, referente al número de veces que a twitteado un tweet con algún nombre de los candidatos a la alcaldía de quito nombrados anteriormente. Mediante esta tabla obtenemos los siguientes datos:

1. EvilDrums con 63 tweets
2. Monchiito_ con 58 tweets
3. G_VerdesotoEC con 56 tweets
4. Belliisima89con 48 tweets
5. RonPeper con 36 tweets
6. FotoIvanEscobar con 35 tweets
7. YazidVera con 30 tweets
8. Maritza39945309 con 28 tweets
9. Fernandorozcom con 27 tweets
10. EsthelaMoreno4 con 26 tweets

2.3.Top 10 twitteros en la ciudad de Cuenca

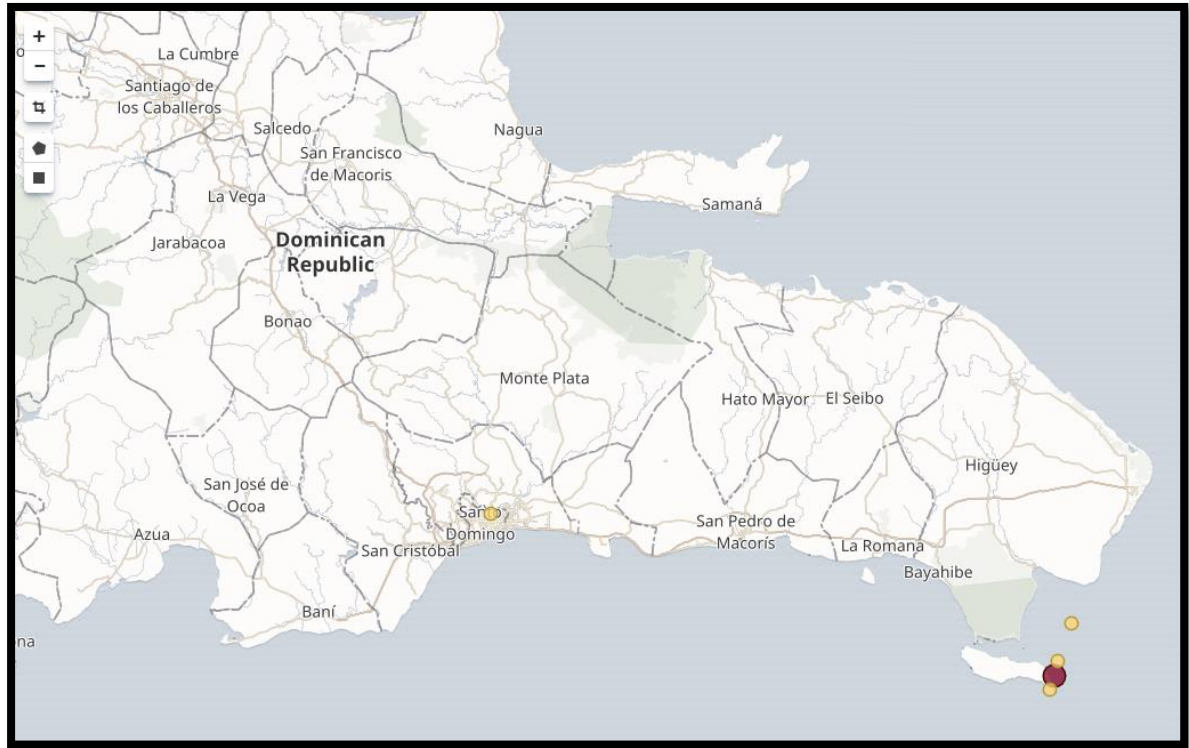


Mediante la gráfica de barras llamada “Top 10 twitteros en la ciudad de Guayaquil” se puede observar en orden descendente el top de 10 twitteros en la ciudad de Guayaquil, referente al número de veces que a twitteado un tweet con algún nombre de los candidatos a la alcaldía de quito nombrados anteriormente. Mediante esta tabla obtenemos los siguientes datos:

1. Democracia2017 con 153 tweets
2. Montalvonr con 145 tweets
3. Hammer50701636 con 127 tweets
4. Ortiz17Camilo con 100 tweets
5. Pabarzallo con 95 tweets
6. Maverocalle con 64 tweets
7. nando121075 con 60 tweets
8. csc_planif con 48 tweets
9. MaLourdesOcampo con 47 tweets
10. YoAzuay con 45 tweets

3. Eventos o noticias mundiales

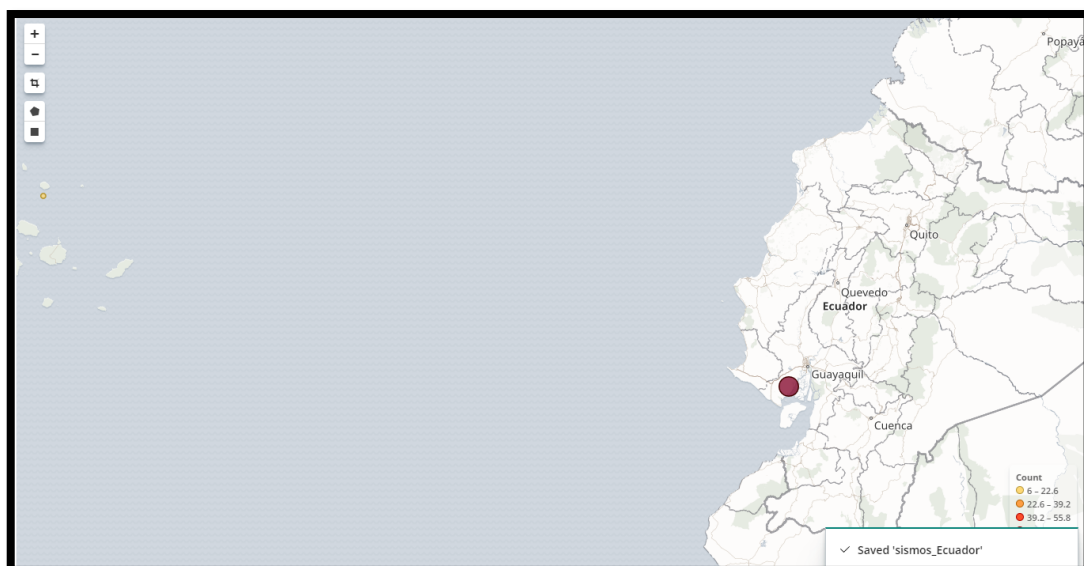
3.1 Grafica Cordinate Map de los sismos ocurridos en Republica Dominicana



Mediante la gráfica podemos observar los datos que se ha recolectado de Twitter referente al sismo ocurrido en Republica Dominicana reportado por “AnTrendsRepDon”, “CesarEcand”. Este sismo tuvo una magnitud de 5.3 grados en la localidad al este de la isla Saona a las 10:34 de esta mañana. El día 04-02-2019

location	message	author
-69.917, 18.501	#sismo es ahora una tendencia en Dominican Republic https://t.co/heYMIIFQVs https://t.co/6rwizE7n9W	TrendsRepDom
-80.1453, -2.4627	RT @elcomerciocom: Los empleados de entidades públicas y privadas evacuaron los edificios en el centro de Guayaquil a causa del sismo https...	MarianaS2223
-80.1453, -2.4627	📢 OTRO #TEMBLOR pero ahora en Rep. Dominicana, EL MUNDO TIEMBLA...	
-80.1453, -2.4627	#ATENCIÓN Un sismo de magnitud 5.3 grados sacude a la República Dominicana » https://t.co/7igaCRKn81 https://t.co/5sg12w7hMA	edurobayo7
-80.1453, -2.4627	RT @elcomerciocom: #ATENCIÓN Un sismo de magnitud 5.3 grados sacude a la República Dominicana » https://t.co/FBYSYKQsd8 https://t.co/erGUL	lucybayasc
-80.1453, -2.4627	RT @elcomerciocom: #ATENCIÓN Un sismo de magnitud 5.3 grados sacude a la República Dominicana » https://t.co/FBYSYKQsd8 https://t.co/erGUL	CesarECand
-80.1453, -2.4627	RT @elcomerciocom: #ATENCIÓN Un sismo de magnitud 5.3 grados sacude a la República Dominicana » https://t.co/FBYSYKQsd8 https://t.co/erGUL	fxalvear

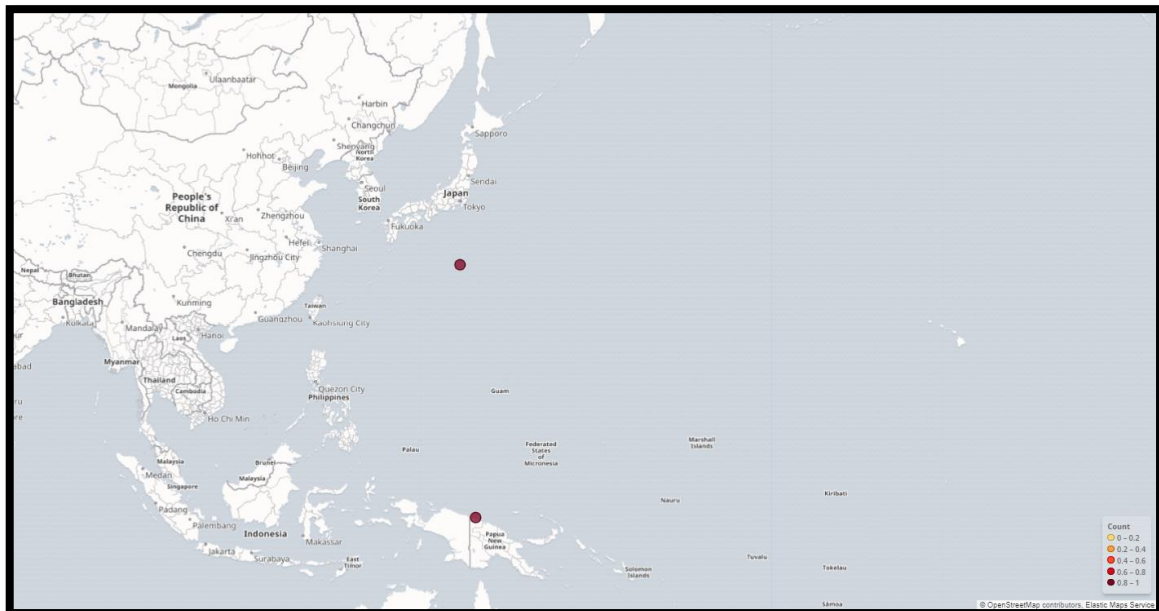
3.1 Grafica Cordinate Map de los sismos ocurridos en Ecuador



Mediante la gráfica podemos observar los datos que se ha recolectado de Twitter referente al sismo ocurrido en Ecuador-Guayaquil reportado por “Andrea_dal”, “RodrigoMoreno1”. Este sismo tuvo una magnitud de 6.16 en la localidad de PlayasGuayas el día 04-02-2019

location	message	author
-80.1453, -2.4627	RT @ecuavisa: #ÚltimaHora Se registra un potente sismo en #Guayaquil #Ecuador. #sismoEcuador Según @IGecuador fue de magnitud 6.13 con epic...	RodrigoMoreno1
-80.1453, -2.4627	RT @jimyjairala: Hacemos un llamado a la calma. Apparently no hay daños importantes reportados en Guayaquil donde el sismo se sintió co...	Andrea_dal
-80.1453, -2.4627	RT @elcomerciocom: Los empleados de entidades públicas y privadas evacuaron los edificios en el centro de Guayaquil a causa del sismo https...	MarianaS2223

3.1 Grafica Cordinate Map de los sismos ocurridos en Indonesia

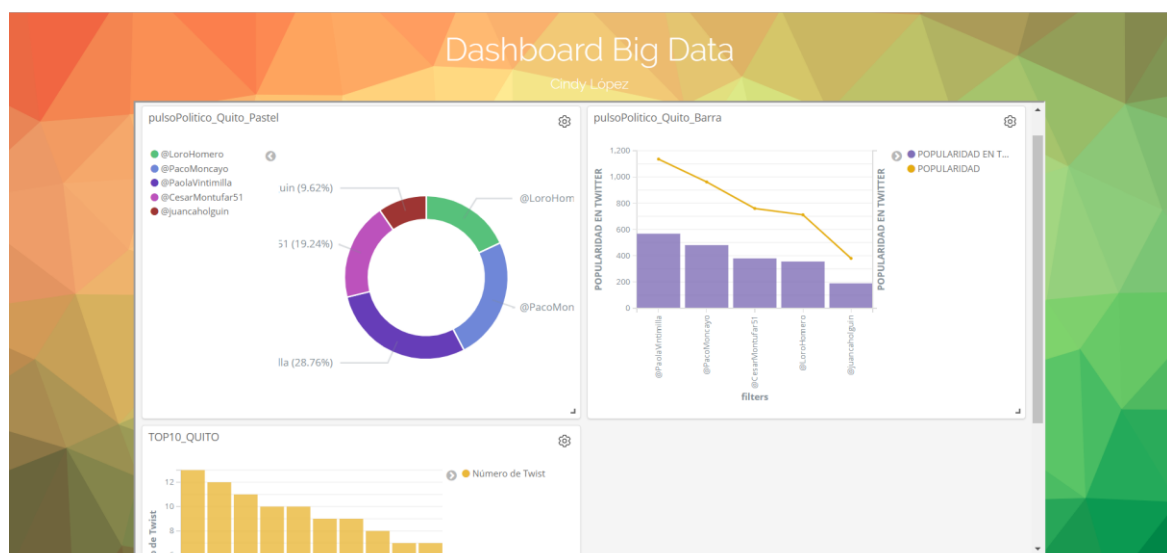


Mediante la gráfica podemos observar los datos que se a recolectado de Twitter referente a uno de los sismos ocurridos en Indonesia cerca a Japón que ha reportado el instituto “eq_map_es” en la región de las Islas Bonin el 04/02/2019 a la 18.08 a 454.3km de Japón

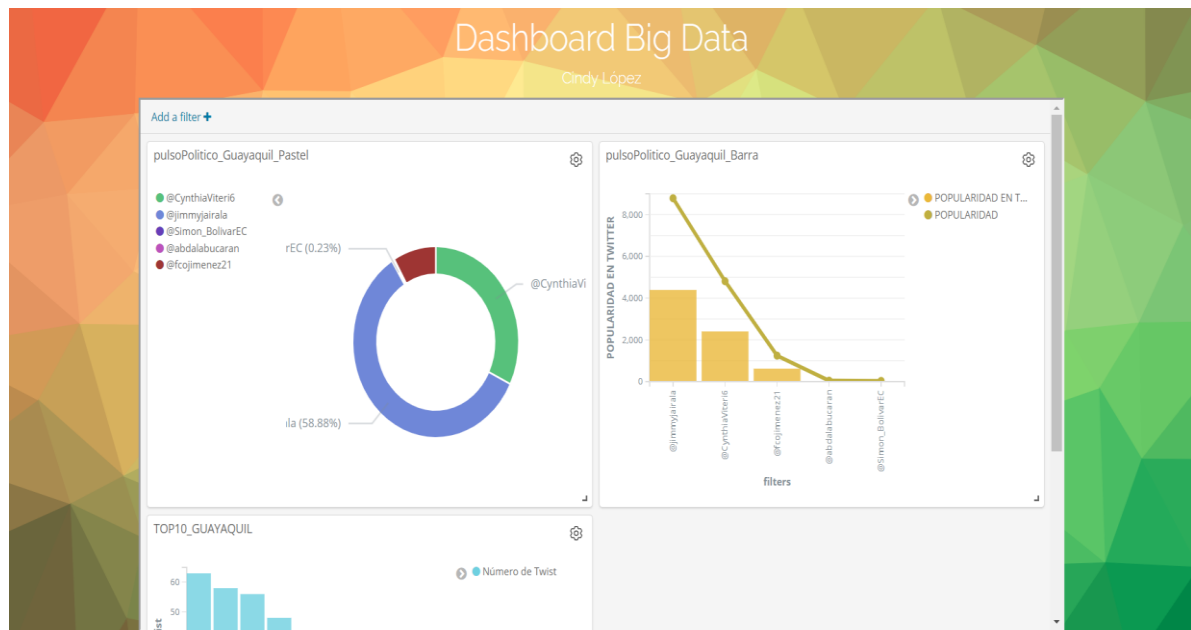
location	message	author
139.7039, 28.6924	<p>【M4.3】 REGION DE LAS ISLAS BONIN, JAPON 454.3km 04/02/2019 18:08:19 UTC #Sismo</p> <p>(G)https://t.co/IHuupGor69 (USGS)https://t.co/LASBaepbC4</p>	eq_map_es

9 Visualización de información

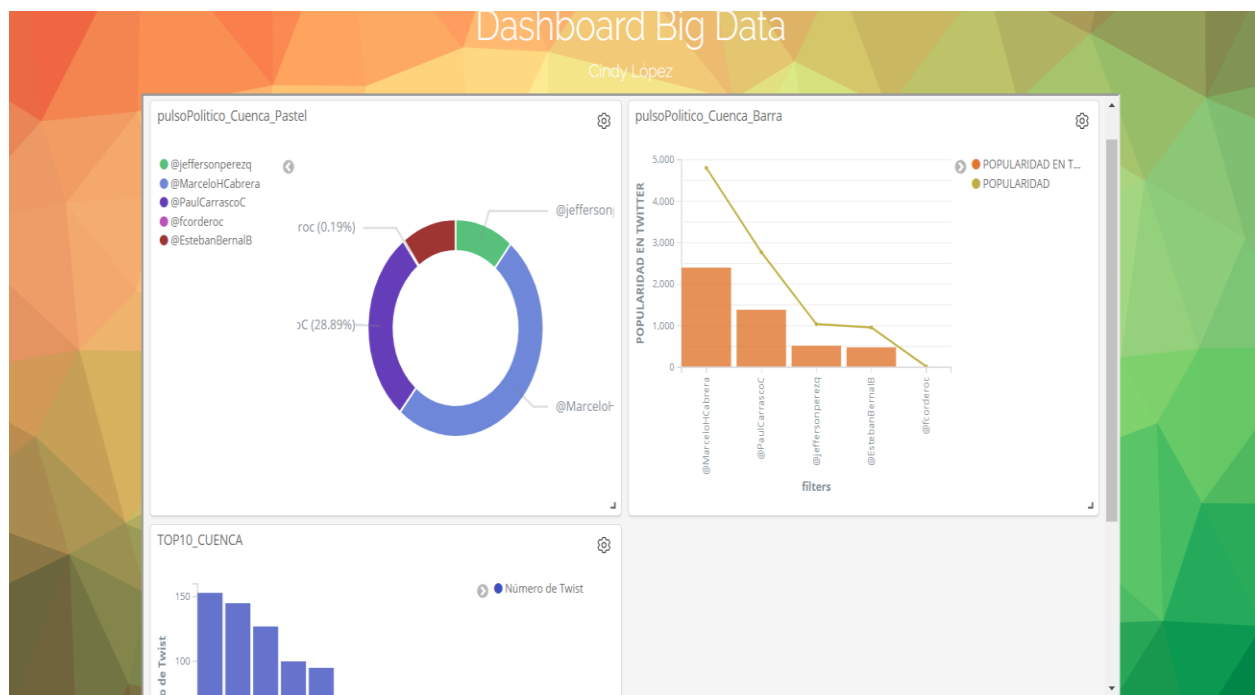
9.1. Visualización de Pulso político para la alcaldía de Quito y el Top 10 de los twitteadores en Quito



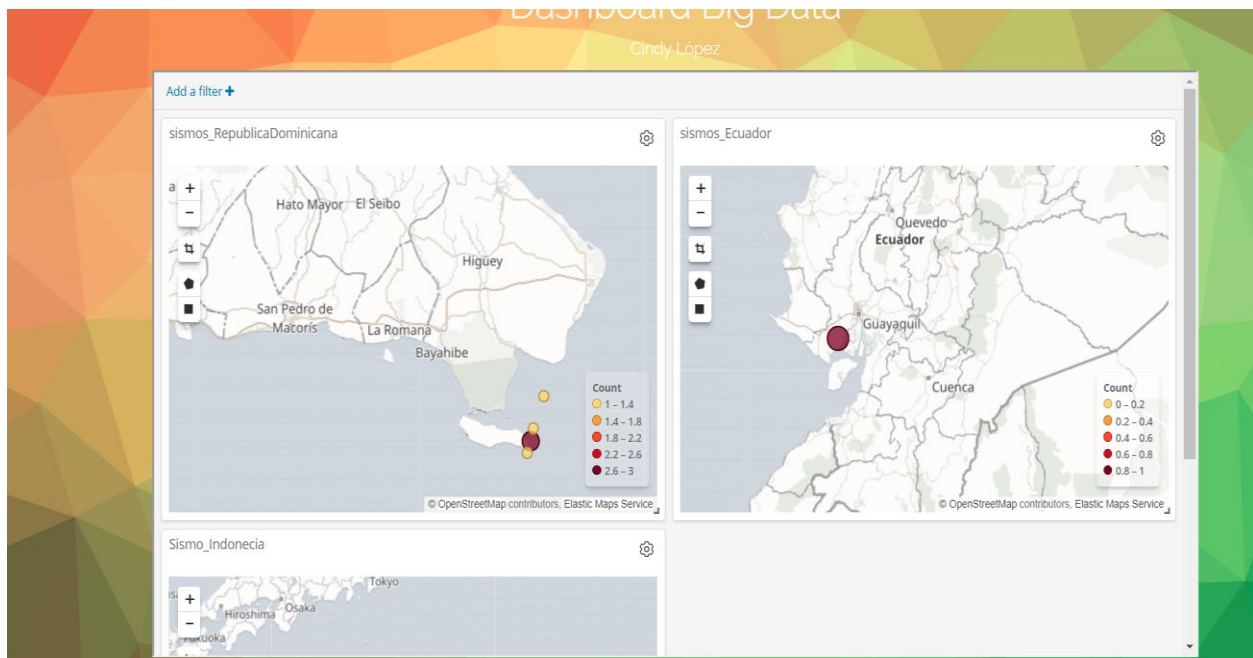
9.2. Visualización de Pulso político para la alcaldía de Guayaquil y el Top 10 de los twitteadores en Guayaquil



9.2. Visualización de Pulso político para la alcaldía de Cuenca y el Top 10 de los twitteadores en Cuenca



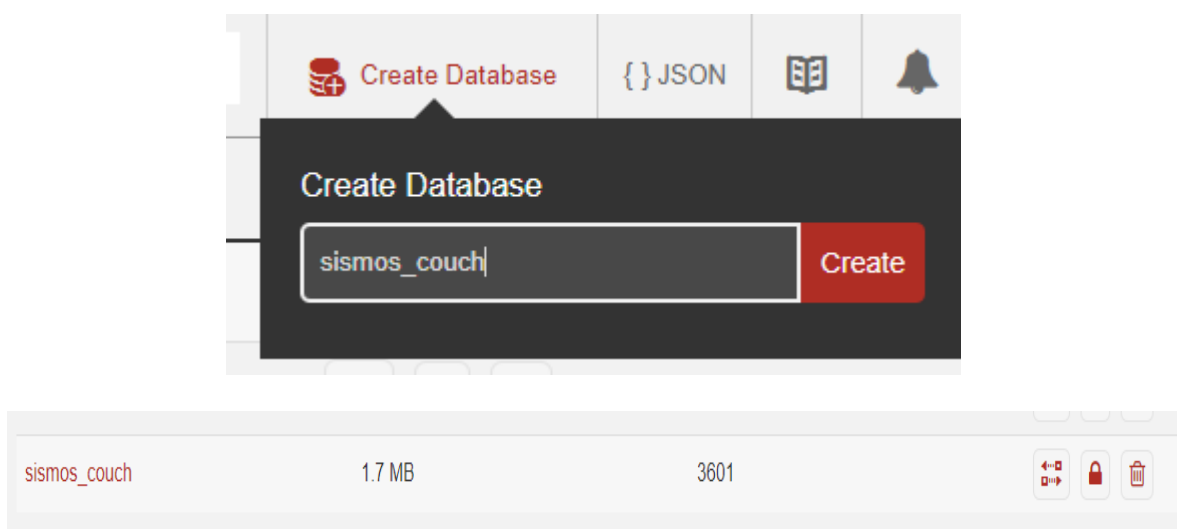
9.3. Visualización de los Sismo ocurridos en Costa Rica, Ecuador y Indonesia el 4 de febrero del 2019

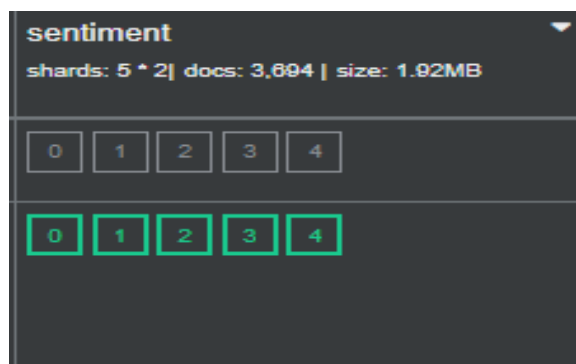


En estas visualizaciones podemos ver los sismos registrados la primera nos muestra varios sismos registrados en República Dominicana a diferencia de la grafica de la izquierda que es de sismos Ecuador que se registra uno en Guayas y la otra grafica nos indica el lugar donde ocurrió el sismo de Indonesia el 4 de febrero del 2019.

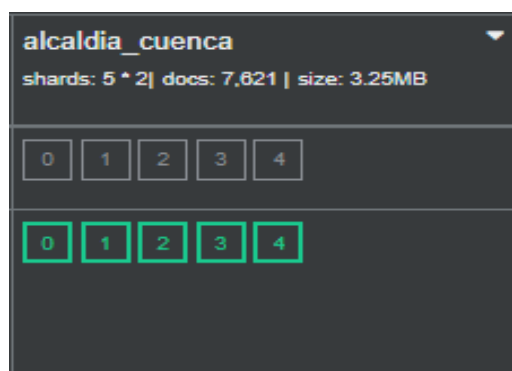
10 Resultados obtenidos

➤ Datos CouchDB

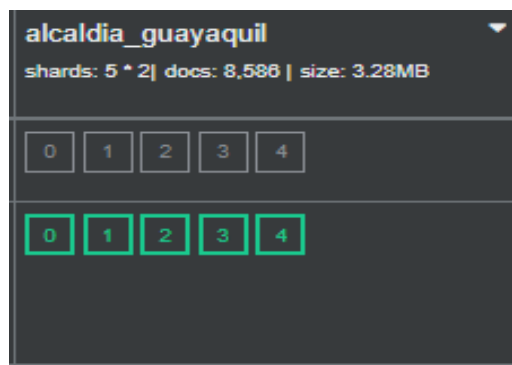




➤ Datos Elasticsearch alcaldes_Cuenca



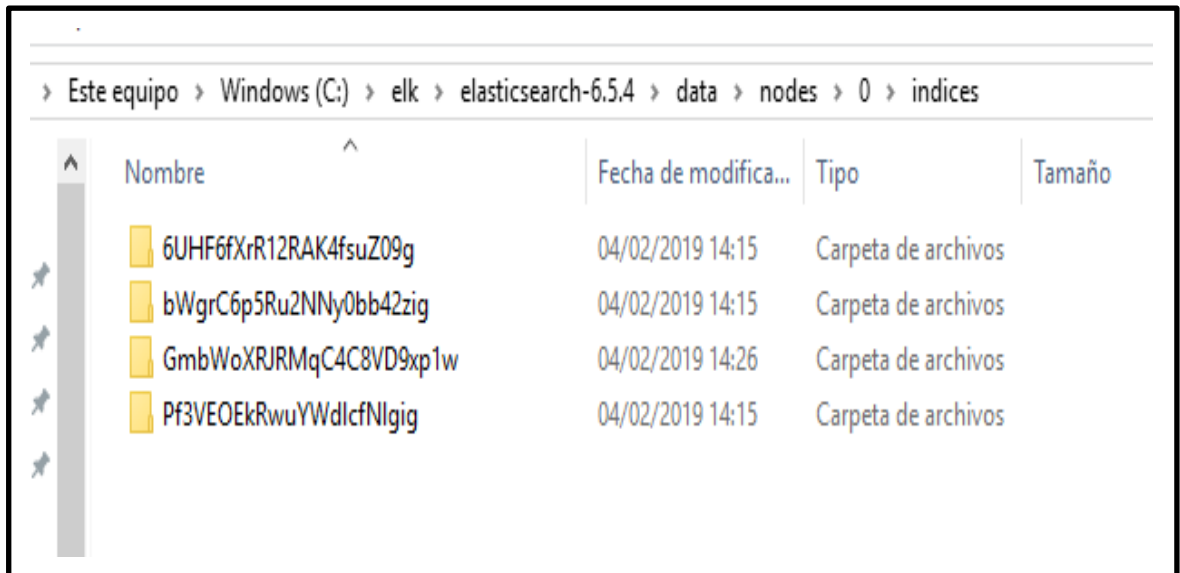
➤ Datos alcaldía_guayaquil



➤ Datos alcaldía_quito



➤ Archivos Data



Nombre	Fecha de modifica...	Tipo	Tamaño
6UHF6fXrR12RAK4fsuZ09g	04/02/2019 14:15	Carpeta de archivos	
bWgrC6p5Ru2NNy0bb42zig	04/02/2019 14:15	Carpeta de archivos	
GmbWoXRJRMqC4C8VD9xp1w	04/02/2019 14:26	Carpeta de archivos	
Pf3VEOEKrwuYWdlcfNIgig	04/02/2019 14:15	Carpeta de archivos	

11 Conclusiones y Recomendaciones

- Mediante el desarrollo del proyecto pudimos darnos cuenta que un correcto mapeado de datos recolectados de Twitter nos permite al momento de realizar graficas o visualizaciones obtener los recursos para poder realizar operaciones, filtrado o búsquedas de datos de una manera mas sencilla.
- Mediante la realización del proyecto en cada una de sus etapas ha permitido conocer cual el pulso político actual y que podría darse en realmente las elecciones de estas votaciones tal como se observo se pudo determinar que Paola Veintimilla y Paco Mocayo. Jimmy Jairala y Chyntia Viteri. Al igual que Marcelo Cabrera y Paul Carrasco. Son los dos maximos contrincantes que posiblemente podrian ganar las alcaldía de Quito, Guayaquil y Cuenca respectivamente.
- Usar Elasticsearch hace que implementen una gran cantidad de funciones, tales como la división personalizada de texto en palabras (Mapeo), derivación personalizada, búsqueda personalizada, etc. Además, Una búsqueda difusa es buena para los errores de ortografía. Al usar Elasticsearch te permite encontrar lo que estás buscando, aunque tenga un error ortográfico. Elasticsearch está orientado a documentos. Almacenas entidades complejas del mundo real como documentos JSON estructurados e indexa todos los campos de forma predeterminada, con un resultado de rendimiento superior.

12 Desafíos y problemas encontrados

- **ERROR 410:** al ejecutar Python este error ocurre cuando las credenciales de Twitter no son las correctas. La solución fue utilizar unas credenciales facilitadas por el instructor

- **ERROR 420:** Este error se produjo cuando se cosechaba datos desde Twitter de tres índices diferentes, al querer realizar otra cosecha nos saltaba ese error que investigando supimos que se produce cuando se están realizando demasiadas consultas o cosas de datos a Twitter, ya que tan solo esta permite un cierto número limitado de cosechas de tweets. Por ende la solución fue abrir en una máquina virtual y cosechar en dos máquinas diferentes.
- **PASAR DATOS A COUCH:** los más comunes errores fue que el documento JSON no está con las etiquetas de inicio “{“docs”:

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Versión 10.0.17134.523]
(c) 2018 Microsoft Corporation. Todos los derechos reservados.

C:\Users\usuario\Desktop\MINERIA DE DATOS>curl -7.62.0-win64-mingw\bin>curl -d @"C:\Users\usuario\Desktop\base_sismo.json" -H "Content-type: application/json" -X POST http://127.0.0.1:5984/simos_couch/_bulk_docs
```

- Otro de los problemas que tuvimos al realizar la cosecha de datos de Twitter utilizando código de Python y tweepy es que si se bloqueaba la pantalla del computador o si se perdía la conexión a internet este saltaba un error indicando que la conexión en el equipo remoto se ha producido un error. Por ende, se soluciona dejando en el equipo que está recolectando un video que se reproduzca el cual no permita que este bloquee su pantalla.

[illegible]