

PARA EL FICHERO DE CSV

Yo tengo un fichero earthquake_in_turkey.csv para trabajar con el conjunto de datos de terremotos en Turquía. Para realizar un análisis completo de los datos de terremotos en Turquía, primero debemos realizar un análisis exploratorio de los datos en Python. Luego, podemos cargar los datos a Hive para realizar consultas más complejas.

Primero debemos copiarlos al sistema de archivos distribuido de Hadoop (HDFS). Luego, podemos lanzar Hive para comenzar a realizar consultas. Finalmente, creamos una tabla Hive para almacenar los datos.

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
[umaster@ibmuamdocker ~]$ cd Master_Platform
[umaster@ibmuamdocker Master_Platform]$ systemctl start docker
[umaster@ibmuamdocker Master_Platform]$ docker-compose start
Starting namenode           ... done
Starting datanode           ... done
Starting resourcemanager    ... done
Starting nodemanager        ... done
Starting historyserver      ... done
Starting zoo                ... done
Starting hbase-master       ... done
Starting hbase-region       ... done
Starting hive-server        ... done
Starting hive-metastore     ... done
Starting hive-metastore-postgresql ... done
Starting presto-coordinator ... done
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
[umaster@ibmuamdocker Master_Platform]$ docker cp earthquake_in_turkey.csv hive-server:.
Successfully copied 2.49MB to hive-server:.
[umaster@ibmuamdocker Master_Platform]$ docker exec -it hive-server bash
root@hive-server:/opt# hdfs dfs -put /earthquake_in_turkey.csv
root@hive-server:/opt# hdfs dfs -ls
Found 1 items
-rw-r--r--  3 root supergroup    2487346 2024-01-27 21:27 earthquake_in_turkey.csv
root@hive-server:/opt# hive
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:file:/opt/hive/lib/log4j-slf4j-impl-2.6.2.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:file:/opt/hadoop-2.7.4/share/hadoop/common/lib/slf4j-log4j12
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
k, tez) or using Hive 1.X releases.
hive> CREATE DATABASE terremotos;
OK
Time taken: 1.314 seconds
hive> USE terremotos;
OK
Time taken: 0.052 seconds
hive> CREATE TABLE terremotos (No INT, Earthquake Code INT, Latitude
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> CREATE TABLE terremotos (No INT, Earthquake_Code INT, Latitu
de FLOAT, Longitude FLOAT, Depth FLOAT, xM FLOAT, MD FLOAT, ML FL
OAT, Mw FLOAT, Ms FLOAT, Mb FLOAT, Type STRING, Location STRING, D
ate_Time TIMESTAMP) ROW FORMAT DELIMITED FIELDS TERMINATED BY ','
STORED AS TEXTFILE TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.639 seconds
hive> █
```

```
hive> LOAD DATA INPATH 'earthquake_in_turkey.csv' OVERWRITE INTO TABLE terremotos; █
```

H1) LOS TERREMOTOS SUELEN PRODUCIRSE CON MÁS FRECUENCIA EN UNA ESTACIÓN DETERMINADA

Utilizamos una sentencia CASE para clasificar los meses en estaciones.

A continuación, utilizamos GROUP BY para agrupar los terremoto según las estaciones calculadas.

Contamos el número de terremoto de cada estación mediante COUNT(*).

Por último, ordenamos los resultados por estaciones.

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> SELECT
>     CASE
>
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END AS Season,
>     COUNT(*) AS Earthquake_Count
> FROM
>     terremotos
> GROUP BY
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END
> ORDER BY
>     Season;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the futur
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
Total MapReduce CPU Time Spent: 5 seconds 100 msec
OK
Fall      4746
Spring    5215
Summer    4668
Winter    5439
Time taken: 79.944 seconds, Fetched: 4 row(s)
hive>
```

Hacemos la misma consulta para los terremotos superiores a 5

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
Time taken: 79.944 seconds, Fetched: 4 row(s)
hive> SELECT
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END AS Season,
>     COUNT(*) AS Earthquake_Count
> FROM
>     terremotos
> WHERE
>     xM > 5
> GROUP BY
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END
> ORDER BY
>     Season;
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
c HDFS Read: 5637 HDFS Write: 177 SUCCESS
Total MapReduce CPU Time Spent: 4 seconds 490 msec
OK
Fall      251
Spring    299
Summer    246
Winter    271
Time taken: 54.432 seconds, Fetched: 4 row(s)
hive> █
```

Según los resultados de nuestro análisis, los terremotos suelen producirse con mayor frecuencia en las estaciones de primavera e invierno, aunque la diferencia es muy pequeña. Por lo tanto, podemos confirmar la hipótesis de que los terremotos suelen producirse con más frecuencia en una estación determinada.

H2) EXISTE UNA RELACIÓN ENTRE LA MAGNITUD DEL TERREMOTO Y SU PROFUNDIDAD

Clasificamos las magnitudes de los terremotos en rangos en la columna xM con CASE.

Redondeamos la profundidad media a dos decimales utilizando ROUND(AVG(Profundidad), 2) con la función ROUND.

La función AVG calcula la profundidad media dentro de cada rango de magnitud.

Los resultados se agrupan por rangos de magnitud y se clasifican en orden ascendente en función de los rangos.

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> SELECT
>     CASE
>         WHEN xM >= 0 AND xM < 2 THEN '0-2'
>         WHEN xM >= 2 AND xM < 3.5 THEN '2-3.5'
>         WHEN xM >= 3.5 AND xM < 4 THEN '3.5-4'
>         WHEN xM >= 4 AND xM < 4.5 THEN '4-4.5'
>         WHEN xM >= 4.5 AND xM < 5 THEN '4.5-5'
>         WHEN xM >= 5 AND xM < 5.5 THEN '5-5.5'
>         WHEN xM >= 5.5 AND xM < 6 THEN '5.5-6'
>         WHEN xM >= 6 AND xM < 6.5 THEN '6-6.5'
>         WHEN xM >= 6.5 AND xM < 7 THEN '6.5-7'
>         WHEN xM >= 7 AND xM < 7.5 THEN '7-7.5'
>         WHEN xM >= 7.5 THEN '7.5+'
>     END AS Magnitude_Range,
>     ROUND(AVG(Depth), 2) AS Average_Depth
> FROM
>     terremotos
> GROUP BY
>     CASE
>         WHEN xM >= 0 AND xM < 2 THEN '0-2'
>         WHEN xM >= 2 AND xM < 3.5 THEN '2-3.5'
>         WHEN xM >= 3.5 AND xM < 4 THEN '3.5-4'
>         WHEN xM >= 4 AND xM < 4.5 THEN '4-4.5'
>         WHEN xM >= 4.5 AND xM < 5 THEN '4.5-5'
>         WHEN xM >= 5 AND xM < 5.5 THEN '5-5.5'
>         WHEN xM >= 5.5 AND xM < 6 THEN '5.5-6'
>         WHEN xM >= 6 AND xM < 6.5 THEN '6-6.5'
>         WHEN xM >= 6.5 AND xM < 7 THEN '6.5-7'
>         WHEN xM >= 7 AND xM < 7.5 THEN '7-7.5'
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
> FROM
>   terremotos
> GROUP BY
>   CASE
>     WHEN xM >= 0 AND xM < 2 THEN '0-2'
>     WHEN xM >= 2 AND xM < 3.5 THEN '2-3.5'
>     WHEN xM >= 3.5 AND xM < 4 THEN '3.5-4'
>     WHEN xM >= 4 AND xM < 4.5 THEN '4-4.5'
>     WHEN xM >= 4.5 AND xM < 5 THEN '4.5-5'
>     WHEN xM >= 5 AND xM < 5.5 THEN '5-5.5'
>     WHEN xM >= 5.5 AND xM < 6 THEN '5.5-6'
>     WHEN xM >= 6 AND xM < 6.5 THEN '6-6.5'
>     WHEN xM >= 6.5 AND xM < 7 THEN '6.5-7'
>     WHEN xM >= 7 AND xM < 7.5 THEN '7-7.5'
>     WHEN xM >= 7.5 THEN '7.5+'
>   END
> ORDER BY
>   Magnitude_Range;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available i
n the future versions. Consider using a different execution engine (i.
e. spark, tez) or using Hive 1.X releases.
Query ID = root_20240127231535_b453bbc8-0a74-48a5-b840-7da486f00b0d
Total jobs = 2
Launching Job 1 out of 2
Number of reduce tasks not specified. Estimated from input data size:
1
In order to change the average load for a reducer (in bytes):
  set hive.exec.reducers.bytes.per.reducer=<number>
In order to limit the maximum number of reducers:
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
Total MapReduce CPU Time Spent: 4 seconds 850 msec
OK
3.5-4    10.7
4-4.5    15.79
4.5-5    25.05
5-5.5    31.47
5.5-6    28.87
6-6.5    30.9
6.5-7    20.71
7-7.5    19.73
7.5+     35.08
Time taken: 50.876 seconds, Fetched: 9 row(s)
hive>
```

Vemos que hasta la magnitud 6.5, pequeños terremotos suelen producirse cerca de la superficie y los grandes terremotos suelen producirse a mayor profundidad. Por lo tanto, podemos decir que el tamaño del terremoto y su profundidad tiene una relacion pero después de magnitud 6.5 la profundidad no aumenta y esta relación no es lineal, ya que la profundidad no aumenta de manera constante con el aumento de la magnitud. Así que no hay relacion entre el tamaño del terremoto y su profundidad.

Tambien podemos ver la corelacion:

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> SELECT CORR(xM, Depth) AS Correlation
> FROM terremotos;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different ex
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
Total MapReduce CPU Time Spent: 3 seconds 620 msec
OK
0.29056857638848355
Time taken: 26.512 seconds, Fetched: 1 row(s)
hive> █
```

Como se puede observar, la correlación entre el tamaño del terremoto y su profundidad es de 0,29. Este valor indica que existe una relación entre las dos variables, pero no es una relación muy fuerte.

H3) LA MAYORÍA DE LOS GRANDES TERREMOTOS OCURREN DE NOCHE

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> SELECT
>     HOUR(Date_Time) AS HourOfDay,
>     COUNT(*) AS EarthquakeCount
> FROM
>     terremotos
> WHERE
>     xM > 6
> GROUP BY
>     HOUR(Date_Time)
> ORDER BY
>     HourOfDay;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution engine (i.e. spark, tez) or using Hive 1.X releases.
```

Utilizamos la función HOUR para extraer el componente horario de la columna Date_Time, que representa la marca de tiempo de cada terremoto.

Filtramos los terremotos utilizando WHERE para tener en cuenta únicamente los terremotos con una magnitud (xM) superior a 6.

Agrupamos los datos por hora del día mediante GROUP BY.

Se cuenta el número de terremotos de cada hora mediante la función COUNT.

Los resultados se ordenan por hora del día.

Según los resultados, podemos observar que hay más terremotos entre las 00:00 y las 12:00 que entre las 12:00 y las 24:00. Sin embargo, esta diferencia es pequeña. Cuando comparamos las horas 21:00 a 08:00 con las horas 08:00 a 20:00, podemos ver que el número de terremotos es aproximadamente igual. Por lo tanto, no podemos confirmar la hipótesis de que la mayoría de los grandes terremotos ocurren de noche.


```
umaster@ibmuamdocker:~/Master_Platform - □ x
File Edit View Search Terminal Help
seconds 460 msec
OK
0      7
1      6
2      5
3      4
4      1
5      4
6      5
7      5
8      3
9      3
10     5
11     2
12     5
13     3
14     6
15     2
16     5
17     5
18     3
19     4
20     4
21     3
22     4
23     2
Time taken: 48.648 seconds, Fetch
ed: 24 row(s)
hive>
```

H4) LOS TERREMOTOS DE CIERTA MAGNITUD SON MÁS FRECUENTES EN UNA REGIÓN DETERMINADA

Para realizar la hipótesis 4 de manera sencilla, podemos modificar el archivo earthquake_in_turkey.csv en Python para agregar una columna City_Name. (Se muestra en el archivo EarthQuakes_of_Turkey.ipynb). Luego, podemos crear la tabla nuevamente.

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> DROP TABLE terremotos;
OK
Time taken: 1.151 seconds
hive> CREATE TABLE terremotos (No INT, Earthquake_Code INT, Latitude FLOAT, Longitude FLOAT, Depth FLOAT, xM FLOAT, MD FLOAT, ML FLOAT, Mw FLOAT, Ms FLOAT, Mb FLOAT, Type STRING, Location STRING, Date_Time TIMESTAMP, City_Name STRING) ROW FORMAT DELIMITED FIELDS TERMINATED BY ',' STORED AS TEXTFILE TBLPROPERTIES ("skip.header.line.count"="1");
OK
Time taken: 0.573 seconds
hive> LOAD DATA INPATH 'earthquake_in_turkey.csv' OVERWRITE INTO TABLE terremotos;
Loading data to table terremotos.terremotos
OK
Time taken: 0.534 seconds
```

```
umaster@ibmuamdocker:~/Master_Platform
File Edit View Search Terminal Help
hive> SELECT
>     City_Name,
>     CASE
>         WHEN xM >= 5.5 THEN '5.5+'
>     END AS Magnitude_Range,
>     COUNT(*) AS Earthquake_Count
> FROM
>     terremotos
> WHERE
>     xM >= 5.5
> GROUP BY
>     City_Name,
>     CASE
>         WHEN xM >= 5.5 THEN '5.5+'
>     END
> ORDER BY
>     Earthquake_Count DESC;
WARNING: Hive-on-MR is deprecated in Hive 2 and may not be available in the future versions. Consider using a different execution
```

Seleccionamos el nombre de la ciudad, el rango de magnitud del terremoto ('5,5+') y el recuento de terremotos que cumplen la condición (magnitud mayor o igual a 5,5).

Filtramos los terremotos con una magnitud igual o superior a 5,5.

Agrupamos los terremotos por nombre de ciudad y rango de magnitud ('5,5+'), contando las ocurrencias de terremotos en cada grupo.

Ordenamos los resultados por recuento de terremotos en orden descendente.

OK				
AKDENIZ	5.5+	105		
EGE DENIZI	5.5+	20		
KUTAHYA	5.5+	15		
BINGOL	5.5+	14		
MALATYA	5.5+	13		
IZMIR	5.5+	12		
MANISA	5.5+	10		
VAN	5.5+	10		
BALIKESIR	5.5+	10		
SAKARYA	5.5+	10		
ERZURUM	5.5+	9		
DENIZLI	5.5+	9		
ERZINCAN	5.5+	9		
MUGLA	5.5+	8		
TURKIYE-IRAN SINIR BOLGESI	5.5+	7		
AFYONKARAHISAR	5.5+	7		
ELAZIG	5.5+	7		

- Podemos observar que la mayoría de los terremotos ocurren en las regiones del Mediterráneo (área "AKDENIZ") en el sur y el Egeo (área "EGE DENIZI") y oeste de Turquía.

- En particular, las ciudades de Izmir (12 terremotos), Manisa (10 terremotos), Denizli (9 terremotos) y Mugla (8 terremotos), que está en el oeste de Turquía, tienen una alta concentración de terremotos.

- En el noroeste del país, las ciudades de Balikesir y Sakarya también tienen una alta concentración de terremotos, con 10 terremotos cada una.

- En el centro de Turquía, la ciudad de Kutahya tiene 15 terremotos.

- En el este del país, las ciudades de Bingol (14 terremotos), Malatya (13 terremotos), Van (10 terremotos), Erzurum (9 terremotos) y Erzincan (9 terremotos) también tienen una alta concentración de terremotos.

- Finalmente, en el sureste del país, la zona fronteriza entre Turquía e Irán ("TURKIYE-IRAN SINIR BOLGESI") tiene 7 terremotos.

MUS	5.5+	6	
SIVAS	5.5+	6	
TUNCELI	5.5+	6	
BURSA	5.5+	6	
CANKIRI	5.5+	5	
CORUM	5.5+	5	
BURDUR	5.5+	5	
AYDIN	5.5+	5	
TEKIRDAG		5.5+	5
ADANA	5.5+	5	
DUZCE	5.5+	5	
KASTAMONU		5.5+	5
YOZGAT	5.5+	4	
TOKAT	5.5+	4	
KONYA	5.5+	4	
KOCAELI	5.5+	4	
KARS	5.5+	4	
KAHRAMANMARAS		5.5+	4
GURCISTAN		5.5+	4
AZERBAYCAN		5.5+	4

ESKISEHIR	5.5+	3	
MARMARA DENIZI	5.5+	3	
DIYARBAKIR	5.5+	3	
AGRI	5.5+	3	
CANAKKALE	5.5+	3	
GAZIANTEP	5.5+	3	
IRAK	5.5+	2	
HATAY	5.5+	2	
ADIYAMAN	5.5+	2	
KARABUK	5.5+	2	
OSMANIYE	5.5+	2	
BOLU	5.5+	2	
BITLIS	5.5+	2	
SANLIURFA	5.5+	2	
SIRNAK	5.5+	2	
ARDAHAN	5.5+	2	
ANTALYA	5.5+	2	
ANKARA	5.5+	2	
USAK	5.5+	2	

En el este de Turquía, también se observa una concentración de terremotos, en particular en las ciudades de Mus (6 terremotos), Tunceli (6 terremotos), Kars (4 terremotos), Kahramanmaras (4 terremotos), Diyarbakir (3 terremotos), Agri (3 terremotos), Gaziantep (3 terremotos), Bitlis (2 terremotos), Sanliurfa (2 terremotos) y Sirnak (2 terremotos). Además, los países vecinos de Azerbaiyán y Georgia también han registrado 4 terremotos cada uno.

En el centro de Turquía, las ciudades de Sivas, Çorum, Yozgat, Konya y Cankırı han registrado entre 4 y 6 terremotos de magnitud superior a 5,5. Eskişehir ha registrado 3 terremotos y Ankara, 2.

En el sur, las ciudades de Adana (5 terremotos), Antalya, Osmaniye, Adıyaman y Hatay (2 terremotos), han registrado terremotos. El país vecino de Irak (Iraq) (2 terremotos) también ha registrado terremotos.

En el norte, las ciudades de Düzce, Kastamonu y Tokat han registrado 5 terremotos cada una. Kocaeli ha registrado 4 terremotos. Karabük, Bolu y Ardahan han registrado 2 terremotos cada una.

En el oeste, la ciudad de Aydın ha registrado 5 terremotos.

En el noroeste del país, la ciudad de Tekirdağ ha registrado 5 terremotos y la región del Mar de Mármara (Marmara Denizi), 3.

En general, la región del este de Turquía es la más propensa a terremotos, seguida de las regiones del oeste y el norte. Las regiones del centro y el sur son las menos propensas a terremotos.


```

VAN GOLU      5.5+      1
IRAN      5.5+      1
AMASYA      5.5+      1
SAMSUN      5.5+      1
BARTIN      5.5+      1
AKDENİZ      5.5+      1
IGDIR      5.5+      1
ISPARTA      5.5+      1
ARDAHAN [South 1.4 km]  5.5+      1
HAKKARI      5.5+      1
KARADENİZ      5.5+      1
KIBRIS-GAZIMAGUSA      5.5+      1
KIRIKKALE      5.5+      1
ERZINCAN [North East 0.9 km]  5.5+      1
KIRSEHIR      5.5+      1
ISTANBUL      5.5+      1
MERSIN      5.5+      1
GİRESUN      5.5+      1
NIGDE      5.5+      1
Time taken: 38.12 seconds, Fetched: 75 row(s)
hive> █

```

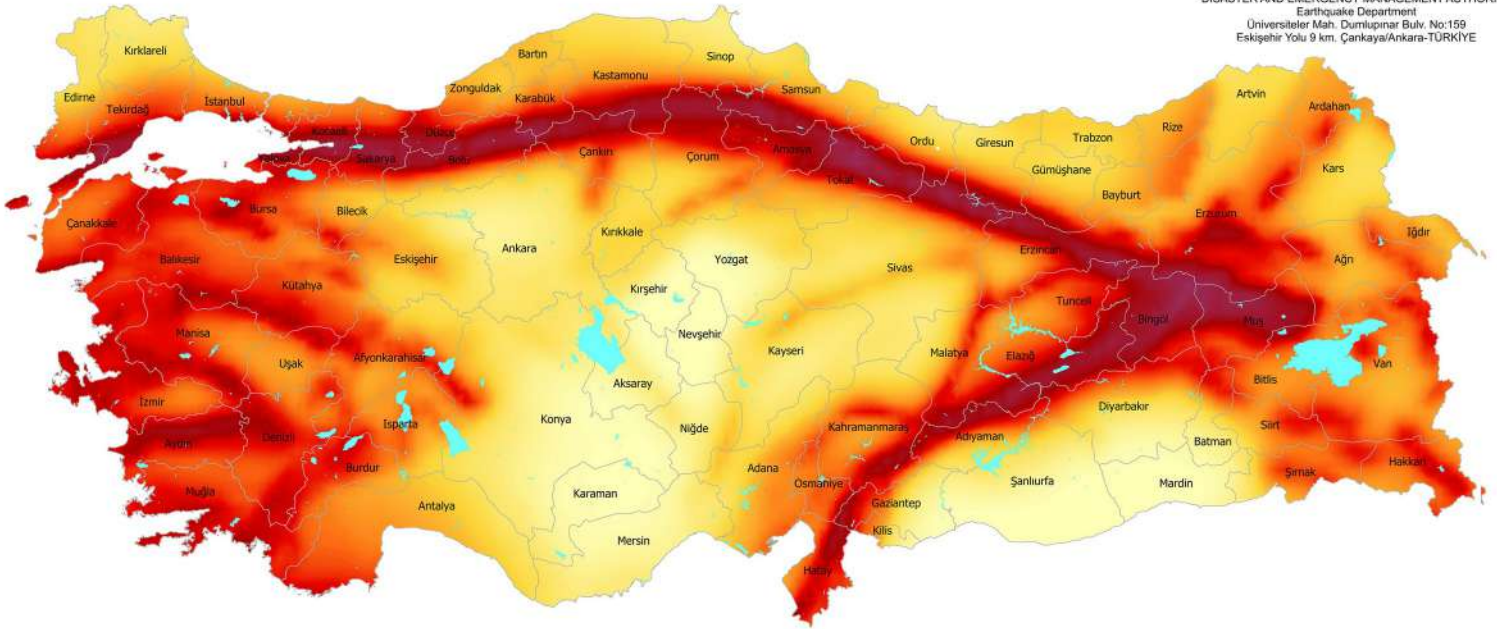
En estas ciudades, solo se ha registrado un terremoto en todas las regiones.

EARTHQUAKE HAZARD MAP OF TURKEY



afadbaskanlik

DISASTER AND EMERGENCY MANAGEMENT AUTHORITY
Earthquake Department
Universiteler Mah. Dumlupınar Bulv. No:159
Eskişehir Yolu 9 km. Çankaya/Ankara-TÜRKİYE



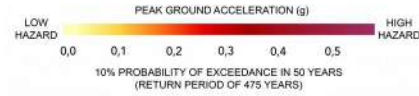
This map is a product of National Earthquake Research Fund supported R&D Project namely "Revision of Turkish Seismic Hazard Map"

This map is prepared considering soil condition (V_s)₃₀ = 760m/s and doesn't include the hazards caused by local soil conditions like liquefaction, ground amplification, subsidence, etc.

Referencing: AFAD, 2018. Earthquake Hazard Map of Turkey.

Copyright © 2018 by AFAD. All rights reserved.

EXPLANATIONS



Lake

Administrative Boundary

0 100 200 400 KM



Además, El Mapa de Peligrosidad Sísmica de Turquía confirma lo que hemos dicho.

PARA EL FICHERO DE PARQUET

Creamos la tabla

```
hive> CREATE TABLE terremotosparquet (No INT, Earthquake_Code BIGINT, Latitude DOUBLE, Longitude DOUBLE, Depth DOUBLE, xM DOUBLE, MD DOUBLE, ML DOUBLE, Mw DOUBLE, Ms DOUBLE, Mb DOUBLE, Type STRING, Location STRING, Date_Time TIMESTAMP, City_Name STRING) STORED AS PARQUET;
OK
Time taken: 0.625 seconds
hive> LOAD DATA INPATH 'earthquakes.parquet' INTO TABLE terremotosparquet;
Loading data to table default.terremotosparquet

OK

Time taken: 0.314 seconds
```

```
hive> SELECT * FROM terremotosparquet LIMIT 10;
OK
1      20231031165636 36.2608 31.6478 90.5 3.7 0.0 3.7 3.6 0.0 0.0 K
e      ANTALYA KORFEZI (AKDENIZ) 2023-10-31 15:56:36.22 AKDENIZ
2      20231030162547 38.0105 36.4315 5.5 4.2 0.0 4.2 4.2 0.0 0.0 K
e      ACIELMA-GOKSUN (KAHRAMANMARAS) [North East 1.7 km] 2023-10-30 15:25:47.72 KAHRAMANMARAS
3      20231028154903 38.9233 32.9713 4.7 3.9 0.0 3.9 3.9 0.0 0.0 K
e      TAVSANCALI-KULU (KONYA) [South West 3.4 km] 2023-10-28 13:49:03.65 KONYA
4      20231026171824 40.56 27.3095 5.0 3.7 0.0 3.7 3.5 0.0 0.0 K
e      KARABIGA ACIKLARI-CANAKKALE (MARMARA DENIZI) 2023-10-26 15:18:24.5 MARMARA DENIZI
5      20231026083211 36.3053 36.3693 7.0 3.6 0.0 3.5 3.6 0.0 0.0 K
e      KURTULUS-REYHANLI (HATAY) [South West 0.4 km] 2023-10-26 06:32:11.93 HATAY
6      20231026072423 35.2302 26.6357 5.0 4.1 0.0 4.1 4.1 0.0 0.0 K
e      GIRIT ADASI (AKDENIZ) 2023-10-26 05:24:23.16 AKDENIZ
7      20231026071843 35.2705 26.6185 5.0 4.1 0.0 4.1 4.0 0.0 0.0 K
e      GIRIT ADASI (AKDENIZ) 2023-10-26 05:18:43.22 AKDENIZ
8      20231025060947 35.4572 31.704 24.5 3.9 0.0 3.9 3.7 0.0 0.0 K
e      AKDENIZ 2023-10-25 04:09:47.97 AKDENIZ
9      20231024003931 38.1613 38.622 5.0 4.4 0.0 4.4 4.4 0.0 0.0 K
e      BASMEZRA-PUTURGE (MALATYA) [North West 3.8 km] 2023-10-23 22:39:31.78 MALATYA
10     20231022004846 38.0872 37.709 5.0 3.5 0.0 3.4 3.5 0.0 0.0 K
e      BEGRE-DOGANSEHIR (MALATYA) [North East 4.7 km] 2023-10-21 22:48:46.49 MALATYA
Time taken: 0.146 seconds, Fetched: 10 row(s)
```

H1) LOS TERREMOTOS SUELEN PRODUCIRSE CON MÁS FRECUENCIA EN UNA ESTACIÓN DETERMINADA

```
hive> SELECT
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END AS Season,
>     COUNT(*) AS Earthquake_Count
> FROM
>     terremotosparquet
> GROUP BY
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END
> ORDER BY
>     Season;
```

```
OK
Fall      4746
Spring    5214
Summer    4669
Winter    5439
Time taken: 48.913 seconds, Fetched: 4 row(s)
hive>
```

El análisis anterior tardó 80 segundos. Con parquet, tardó 49 segundos. Es decir, una reducción de tiempo del 50%.

Filtramos terremotosparquet con magnitud superior a 5

```
hive> SELECT
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END AS Season,
>     COUNT(*) AS Earthquake_Count
> FROM
>     terremotosparquet
> WHERE
>     xM > 5
> GROUP BY
>     CASE
>         WHEN MONTH(Date_Time) IN (12, 1, 2) THEN 'Winter'
>         WHEN MONTH(Date_Time) IN (3, 4, 5) THEN 'Spring'
>         WHEN MONTH(Date_Time) IN (6, 7, 8) THEN 'Summer'
>         WHEN MONTH(Date_Time) IN (9, 10, 11) THEN 'Fall'
>     END
> ORDER BY
>     Season;
```

```
OK
Fall      251
Spring    299
Summer    246
Winter    271
Time taken: 46.101 seconds, Fetched: 4 row(s)
hive> █
```

H2) EXISTE UNA RELACIÓN ENTRE LA MAGNITUD DEL TERREMOTO Y SU PROFUNDIDAD

```
hive> SELECT
>     CASE
>         WHEN xM >= 0 AND xM < 2 THEN '0-2'
>         WHEN xM >= 2 AND xM < 3.5 THEN '2-3.5'
>         WHEN xM >= 3.5 AND xM < 4 THEN '3.5-4'
>         WHEN xM >= 4 AND xM < 4.5 THEN '4-4.5'
>         WHEN xM >= 4.5 AND xM < 5 THEN '4.5-5'
>         WHEN xM >= 5 AND xM < 5.5 THEN '5-5.5'
>         WHEN xM >= 5.5 AND xM < 6 THEN '5.5-6'
>         WHEN xM >= 6 AND xM < 6.5 THEN '6-6.5'
>         WHEN xM >= 6.5 AND xM < 7 THEN '6.5-7'
>         WHEN xM >= 7 AND xM < 7.5 THEN '7-7.5'
>         WHEN xM >= 7.5 THEN '7.5+'
>     END AS Magnitude_Range,
>     ROUND(AVG(Depth), 2) AS Average_Depth
> FROM
>     terremotosparquet
> GROUP BY
>     CASE
>         WHEN xM >= 0 AND xM < 2 THEN '0-2'
>         WHEN xM >= 2 AND xM < 3.5 THEN '2-3.5'
>         WHEN xM >= 3.5 AND xM < 4 THEN '3.5-4'
>         WHEN xM >= 4 AND xM < 4.5 THEN '4-4.5'
```

```
>     terremotosparquet
> GROUP BY
>     CASE
>         WHEN xM >= 0 AND xM < 2 THEN '0-2'
>         WHEN xM >= 2 AND xM < 3.5 THEN '2-3.5'
>         WHEN xM >= 3.5 AND xM < 4 THEN '3.5-4'
>         WHEN xM >= 4 AND xM < 4.5 THEN '4-4.5'
>         WHEN xM >= 4.5 AND xM < 5 THEN '4.5-5'
>         WHEN xM >= 5 AND xM < 5.5 THEN '5-5.5'
>         WHEN xM >= 5.5 AND xM < 6 THEN '5.5-6'
>         WHEN xM >= 6 AND xM < 6.5 THEN '6-6.5'
>         WHEN xM >= 6.5 AND xM < 7 THEN '6.5-7'
>         WHEN xM >= 7 AND xM < 7.5 THEN '7-7.5'
>         WHEN xM >= 7.5 THEN '7.5+'
>     END
> ORDER BY
>     Magnitude_Range;
```

```
OK
3.5-4      10.7
4-4.5      15.79
4.5-5      25.05
5-5.5      31.47
5.5-6      28.87
6-6.5      30.9
6.5-7      20.71
7-7.5      19.73
7.5+       35.08
Time taken: 42.357 seconds, Fetched: 9 row(s)
hive>
```

Vemos las correlaciones

```
hive> SELECT CORR(xM, Depth) AS Correlation
> FROM terremotosparquet;
```

```
OK
0.29056857533836894
Time taken: 18.868 seconds, Fetched: 1 row(s)
hive> █
```

H3) LA MAYORÍA DE LOS GRANDES TERREMOTOS OCURREN DE NOCHE

```
hive> SELECT
>     HOUR(Date_Time) AS HourOfDay,
>     COUNT(*) AS EarthquakeCount
> FROM
>     terremotosparquet
> WHERE
>     xM > 6
> GROUP BY
>     HOUR(Date_Time)
> ORDER BY
>     HourOfDay;
```

```
OK
0      8
1      7
2      2
3      2
4      2
5      5
6      6
7      4
8      4
9      3
10     3
11     7
12     1
13     5
14     3
15     4
16     7
17     2
18     3
19     3
20     5
21     2
22     4
23     4
Time taken: 39.4 seconds, Fetched: 24 row(s)
hive>
```


H4) LOS TERREMOTOS DE CIERTA MAGNITUD SON MÁS FRECUENTES EN UNA REGIÓN DETERMINADA

```
hive> SELECT
>     City_Name,
>     CASE
>         WHEN xM >= 5.5 THEN '5.5+'
>     END AS Magnitude_Range,
>     COUNT(*) AS Earthquake_Count
> FROM
>     terremotosparquet
> WHERE
>     xM >= 5.5
> GROUP BY
>     City_Name,
>     CASE
>         WHEN xM >= 5.5 THEN '5.5+'
>     END
> ORDER BY
>     Earthquake_Count DESC;
```

AKDENIZ	5.5+	105	
EGE DENIZI	5.5+	20	
KUTAHYA	5.5+	15	
BINGOL	5.5+	14	
MALATYA	5.5+	13	
IZMIR	5.5+	12	
MANISA	5.5+	10	
VAN	5.5+	10	
BALIKESIR	5.5+	10	
SAKARYA	5.5+	10	
ERZURUM	5.5+	9	
DENIZLI	5.5+	9	
ERZINCAN	5.5+	9	
MUGLA	5.5+	8	
TURKIYE-IRAN SINIR BOLGESI	5.5+	7	
AFYONKARAHISAR	5.5+	7	
ELAZIG	5.5+	7	
MUS	5.5+	6	
SIVAS	5.5+	6	
TUNCELI	5.5+	6	
BURSA	5.5+	6	
CANKIRI	5.5+	5	
CORUM	5.5+	5	
BURDUR	5.5+	5	

AYDIN	5.5+	5	
TEKIRDAG		5.5+	5
ADANA	5.5+	5	
DUZCE	5.5+	5	
KASTAMONU		5.5+	5
YOZGAT	5.5+	4	
TOKAT	5.5+	4	
KONYA	5.5+	4	
KOCAELI	5.5+	4	
KARS	5.5+	4	
KAHRAMANMARAS		5.5+	4
GURCISTAN		5.5+	4
AZERBAYCAN		5.5+	4
ESKISEHIR		5.5+	3
MARMARA DENIZI		5.5+	3
DIYARBAKIR		5.5+	3
AGRI	5.5+	3	
CANAKKALE		5.5+	3
GAZIANTEP		5.5+	3

IRAK	5.5+	2	
HATAY	5.5+	2	
ADIYAMAN		5.5+	2
KARABUK	5.5+	2	
OSMANIYE		5.5+	2
BOLU	5.5+	2	
BITLIS	5.5+	2	
SANLIURFA		5.5+	2
SIRNAK	5.5+	2	
ARDAHAN	5.5+	2	
ANTALYA	5.5+	2	
ANKARA	5.5+	2	
USAK	5.5+	2	

```

VAN GOLU          5.5+      1
IRAN      5.5+      1
AMASYA    5.5+      1
SAMSUN    5.5+      1
BARTIN    5.5+      1
AKDENİZ    5.5+      1
IGDIR     5.5+      1
ISPARTA   5.5+      1
ARDAHAN [South 1.4 km] 5.5+      1
HAKKARI   5.5+      1
KARADENİZ    5.5+      1
KIBRIS-GAZIMAGUSA      5.5+      1
KIRIKKALE    5.5+      1
ERZINCAN [North East 0.9 km] 5.5+      1
KIRSEHIR     5.5+      1
ISTANBUL     5.5+      1
MERSIN      5.5+      1
GIRESUN      5.5+      1
NIGDE       5.5+      1
Time taken: 40.462 seconds, Fetched: 75 row(s)
hive> █

```

Para la hipótesis 1, el tiempo de ejecución de la consulta sin filtrar la columna xM, se redujo en un 50% al usar archivos Parquet. El tiempo de ejecución de la consulta filtrada se redujo de 54 segundos a 42 segundos, lo que sigue siendo un 25% más rápido que el archivo CSV.

En el caso de la hipótesis 2, el tiempo de ejecución se redujo de 50 a 42 segundos al examinar la relación entre la profundidad y la magnitud del terremoto. Al examinar las correlaciones, el tiempo de ejecución se redujo de 26 a 19 segundos.

Para la hipótesis 3, el tiempo de ejecución fue de 48 segundos en el archivo CSV y de 39 segundos en el archivo Parquet.


Para la hipótesis 4, el tiempo de ejecución fue similar para ambos formatos de archivo.

Según los resultados, el uso de archivos Parquet en Hive reduce el tiempo de ejecución de las consultas en un 50% en promedio.

EN SOLR

Solr Admin

localhost:8983/solr/#/~collections



Dashboard

Logging

Security

Cloud

Schema Designer

Collections

Java Properties

Thread Dump

Collection Sele... ▾

Core Selector ▾

+ Add Collection

gettingstarted

Terremotos1

Create Alias

Please select a collection or alias

Solr Admin

localhost:8983/solr/#/Terremotos1/query?q=*&q.op=OR&indent=true&wt=json&useParams=

Schema Designer

Collections

Java Properties

Thread Dump

Terremotos1

Overview

Analysis

Documents

Params

Files

Query

Stream

SQL

Schema

Core Selector

fq

sort

start, rows

010

df

paramset(s)

Select paramset(s)...

wt

json

☒ indent on

☐ debugQuery

defType

☐ hl

☐ facet

☐ spatial

☐ spellcheck

Raw Query Parameters

JSON Query

Execute Query

```
{
  "indent": "true",
  "q.op": "OR",
  "_forwardedCount": "1",
  "wt": "json",
  "useParams": "",
  "_": "1706192674836"
}

{
  "response": {
    "numFound": 20068,
    "start": 0,
    "numFoundExact": true,
    "docs": [
      {
        "No": [1],
        "Latitude": [36.2608],
        "Longitude": [31.6478],
        "xM": [3.7],
        "MD": [0.0],
        "ML": [3.7],
        "Mw": [3.6],
        "Ms": [0.0],
        "Mb": [0.0],
        "Type": ["Ke"],
        "Location": ["ANTALYA KORFEZI (AKDENIZ)"],
        "Date_Time": ["2023-10-31T16:56:36.220Z"],
        "id": ["5c0f3cf0-af3f-4902-8d28-79960e897184"],
        "Earthquake_Code": [20231031165636],
        "Depth_km": [90.5],
        "_version_": 1789072680830661248
      },
      {
        "No": [2],
        "Latitude": [38.0105],
        "Longitude": [36.4315],
        "xM": [4.2],
        "MD": [0.0],
        "ML": [4.2],
        "Mw": [4.2],
        "Ms": [0.0],
        "Mb": [0.0],
        "Type": ["Ke"],
        "Location": ["ACIELMA-GOKSUN (KAHRAMANMARAS) (North East 1.7 km)"],
        "Date_Time": ["2023-10-30T16:25:47.720Z"],
        "id": ["36-af3f-4902-8d28-79960e897184"],
        "Earthquake_Code": [20231030162547],
        "Depth_km": [90.5],
        "_version_": 1789072680830661248
      }
    ]
  }
}
```

Solr Admin

localhost:8983/solr/#/TerremotosL/query?q=*:*&q.op=OR&indent=true&wt=csv&useParams=

Solr

Dashboard

Logging

Security

Cloud

Schema Designer

Collections

Java Properties

Thread Dump

TerremotosL

Overview

Analysis

Documents

Params

Files

Query

Stream

SQL

Schema

Query Selector

Request Handler (q)

select

common

q

q.op

OR

fq

sort

start, rows

010

0

df

params

wt

csv

indent on

debugQuery

defType

hl

facet

URL http://localhost:8983/solr/#/TerremotosL/select?indent=true&q.op=OR&q=*:*%3A%3AuseParams=wt=csv

No,Ms,LaTtitude,Ma,Date,Time,Type,,M,MD,,version,,MD,Depth,km,,Id,Longitude,Ma,,Location,Earthquake,Code

1,0,0,38,2688,3.6,2023-10-31T26:56:36.220Z,Ka,3,7,0,0,1709072680036661248,0,0,5,5,5c0f3cfe-a33f-4002-8d28-7906b097184,31.6478,3.7,ANTALYA KÖRFEZİ [AKDENİZ],20231031145936

2,0,0,38,0185,4.2,2023-10-30T26:25:47.720Z,Ka,4,2,0,0,1709072680913207296,0,0,5,5,2fadf866-039b-4869-bacc-0932486b625b,36.4315,4.2,ACIELMA GÖKSUN (KAHRAMANMARAŞ) [North East 1.7 K

3,0,0,20,9233,2.9,2023-10-28T23:49:02.050Z,Ka,2,9,0,0,1709072680914255872,0,0,4,7,4502b345-c6c2-4009-86ae-b279e2193eff,32.9712,3.9,TAVSANCALI MÜLÜ (KONYA) [South West 3.4 km],2023

4,0,0,40,56,3.5,2023-10-26T17:18:24.550Z,Ka,3,7,0,0,1709072680915304448,0,0,5,0,5423b3b6-07c7-480c-853b-1a6fab0c3af7,27.3695,3.7,KARABİGA ACTIKLARI CANAKALE (MARMARA DENİZİ),202310

5,0,0,36,3053,3.6,2023-10-26T08:32:11.930Z,Ka,3,6,0,0,1709072680916353024,0,0,7,0,15069408-fa89-4276-bc72-b0b91608735a,36.3695,3.5,KURTULUS-REYHANLI İNCEKİ (South West 8.4 km),20

6,0,0,35,2382,4.1,2023-10-26T07:24:13.160Z,Ka,4,1,0,0,17090726809174091600,0,0,5,0,3032772a-e163-4526-bc61-ca0358b0a658,38.6357,4.1,GİDİR ADASI [AKDENİZ],20231026072423

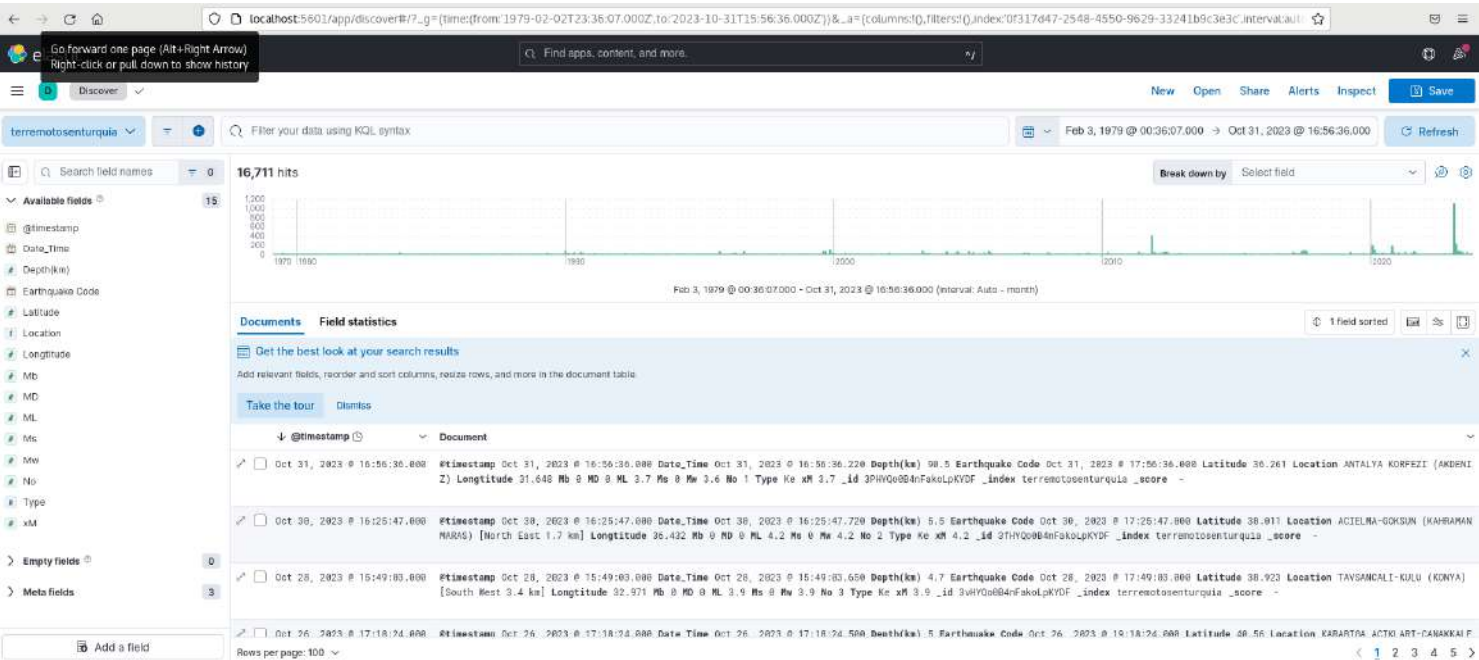
7,0,0,35,2785,4.0,2023-10-26T07:10:43.320Z,Ka,4,1,0,0,17090726809174091601,0,0,5,0,90318601-f676-4fcc-b433-78bac3c0b772,38.6185,4.1,GİDİR ADASI [AKDENİZ],20231026071843

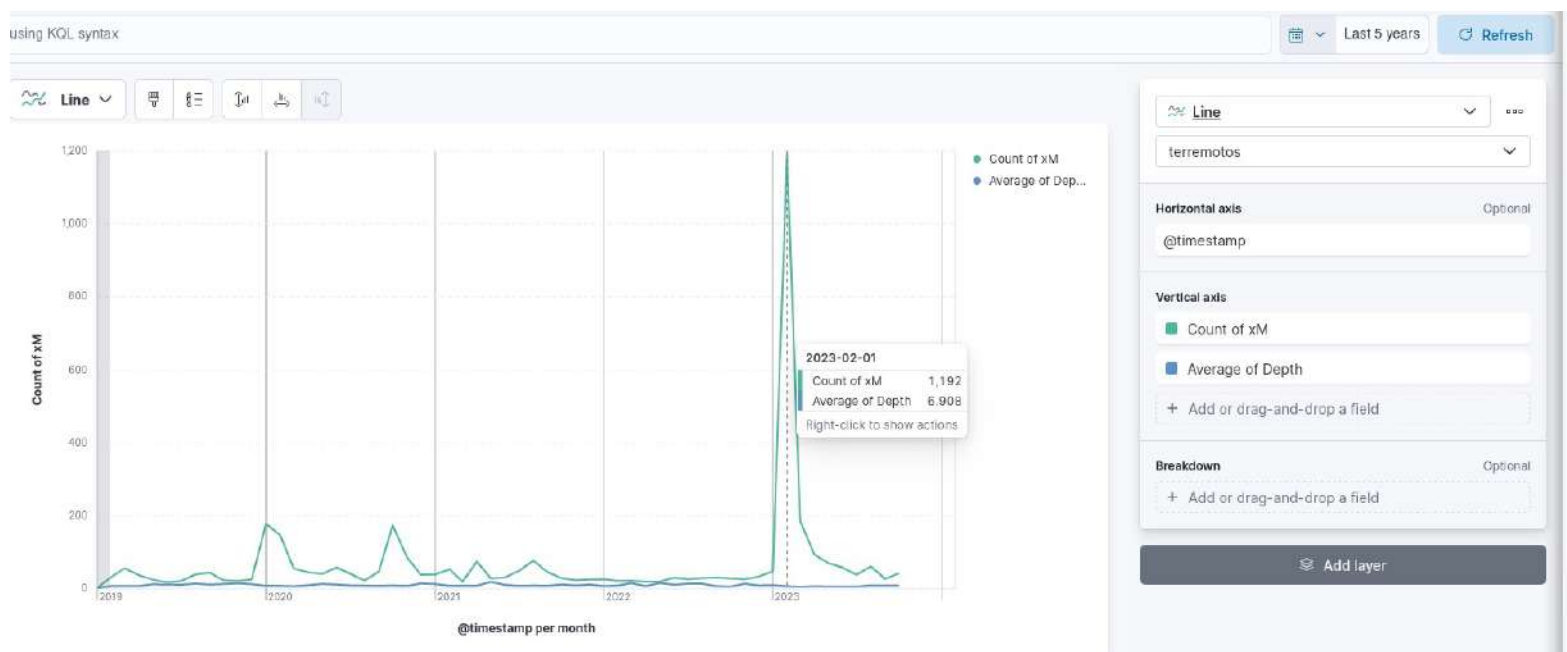
8,0,0,25,4272,3.7,2023-10-23T06:09:47.970Z,Ka,2,9,0,0,1709072680918450176,0,0,24,3,b14e032f-9202-404c-a7f5-2dc3ed1db07b,21.704,3.9,AKDENİZ,20231023060947

9,0,0,30,1613,4.4,2023-10-24T06:30:51.700Z,Ka,4,4,0,0,1709072680919400752,0,0,5,0,0f6ae0ff-7717-45eb-96cd-79e400fc076c,38.622,4.4,BAGMEZNA-PUYURGE (MALATYA) [North West 3.0 km],20

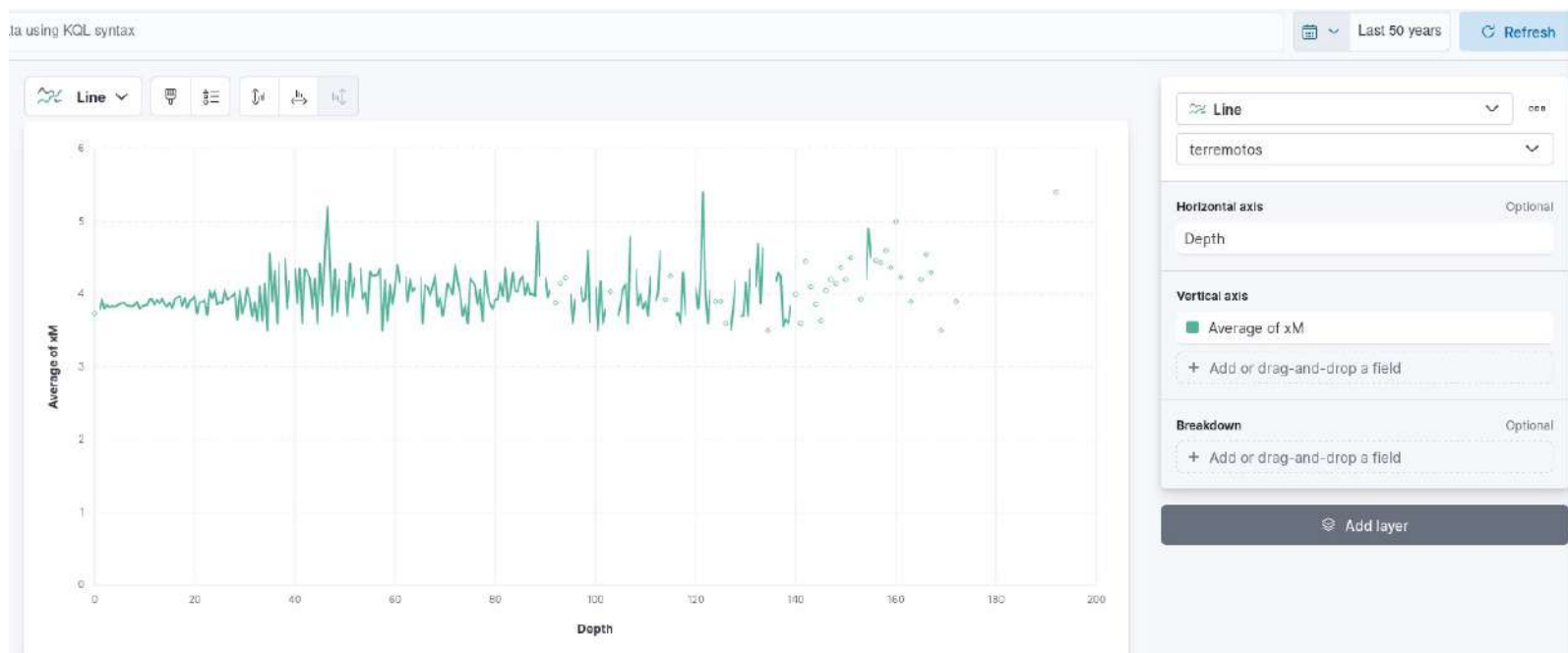
10,0,0,38,0072,3.5,2023-10-22T06:08:46.490Z,Ka,3,5,0,0,1709072680920547328,0,0,5,0,721f161b-cc4d-4674-9fcb-0648cb09ae77,37.709,3.0,BEĞRE-DÜĞANŞEHİR (MALATYA) [North East 4.7 km],2

EN KIBANA

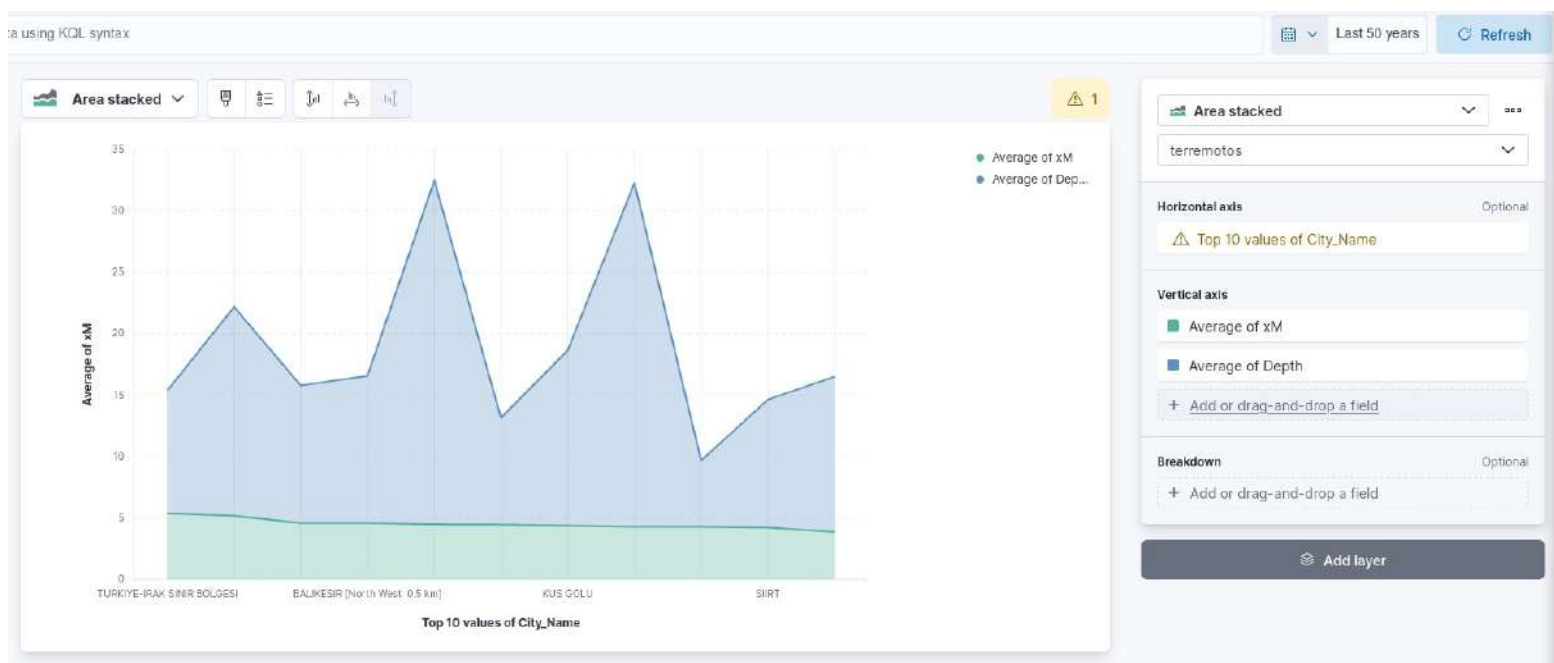




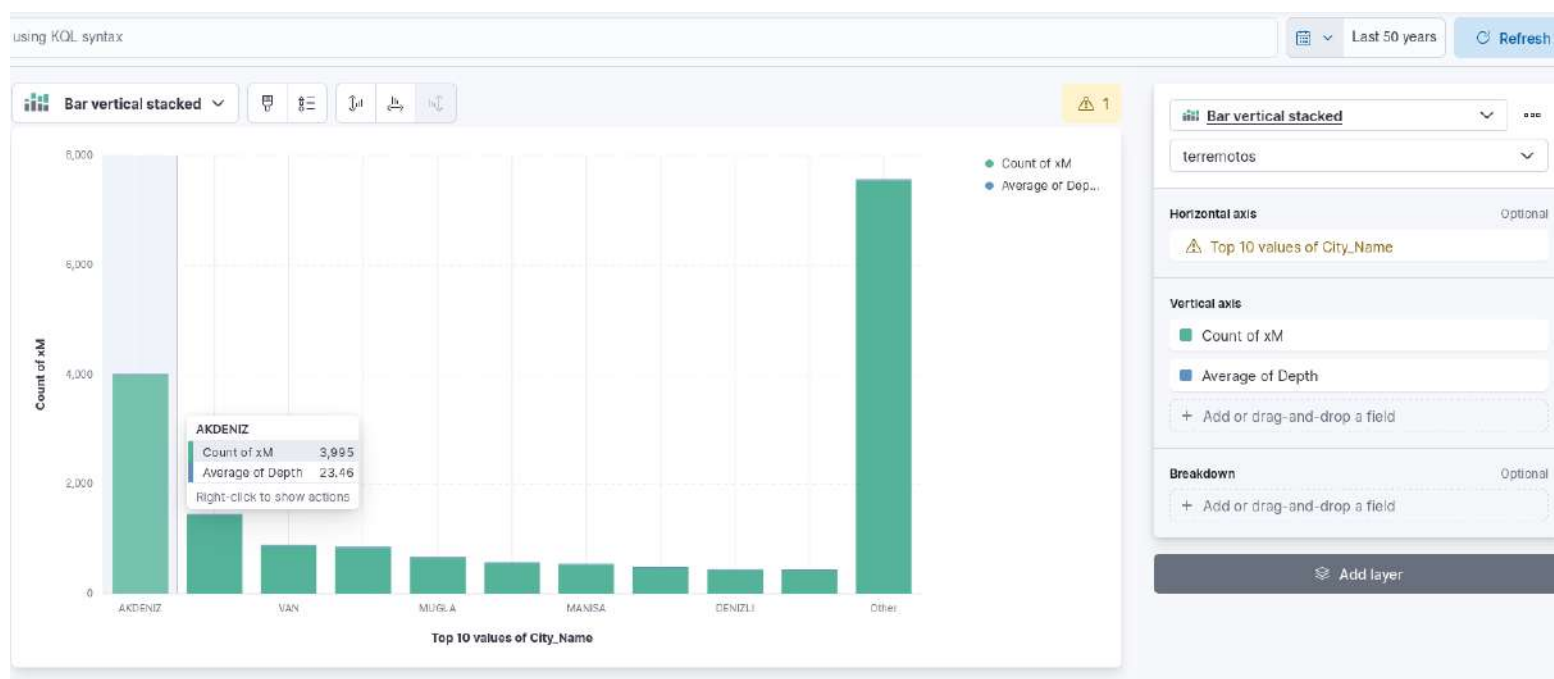
Aquí hemos filtrado por meses y podemos ver cuántos terremotos ha habido en los últimos 5 años y a qué profundidad de media.



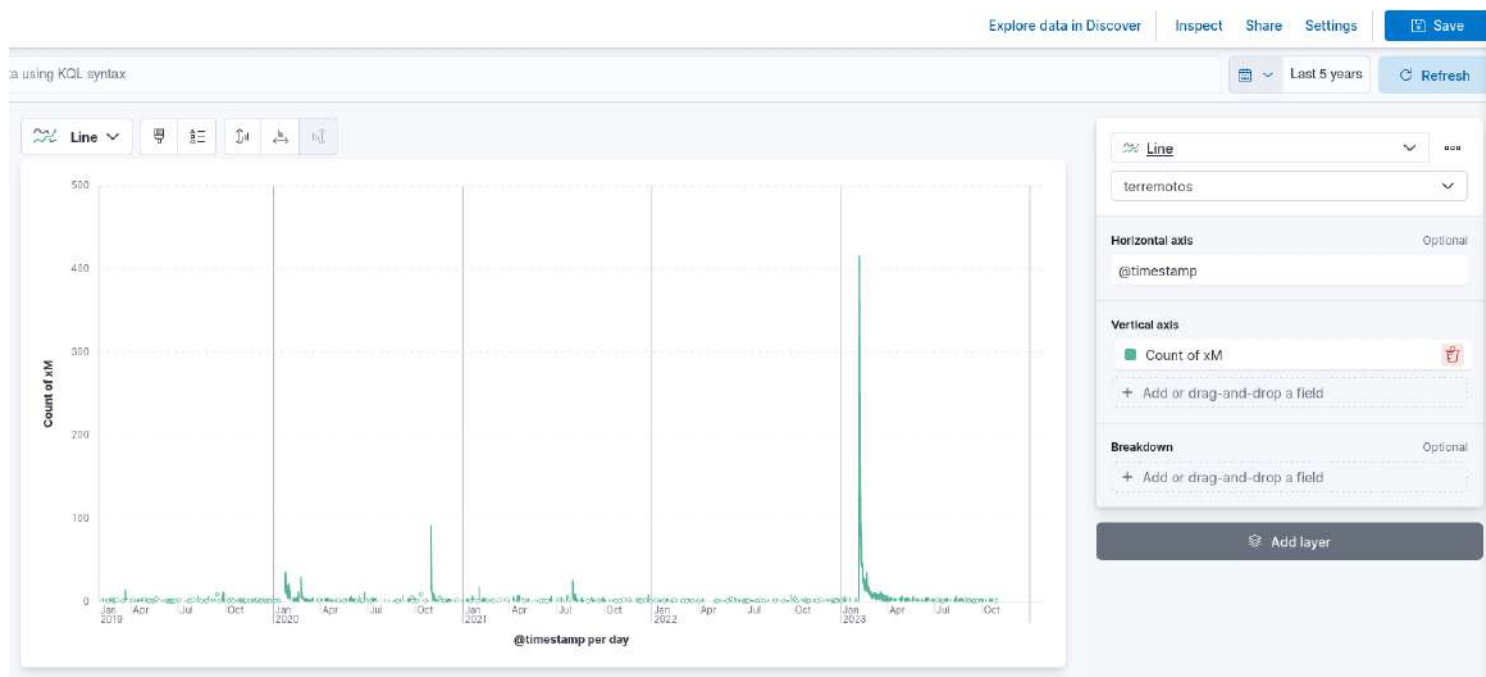
En este gráfico, muestra la relación entre la profundidad y la magnitud media de los terremotos de los últimos 50 años.



Este gráfico muestra la profundidad media y la magnitud media de los terremotos que han tenido lugar en los últimos 50 años, filtrados por ciudad.



Este gráfico muestra la distribución del número de terremotos y la profundidad media de los terremotos que han tenido lugar en los últimos 50 años.



De esta forma, podemos producir gráficos similares con datos de diferentes períodos de tiempo o regiones.

