



Pre-Entrega de proyecto final: Obtención de insight a partir de visualizaciones

Maximiliano Ramirez



Abstracto

El proyecto busca generar un modelo de interpretación de datos de terremotos, estos entendidos como movimientos sísmicos de magnitudes significativas y que podrían provocar daños. La motivación radica en la búsqueda de patrones o tendencias que permitan establecer niveles de riesgo geológicos y temporales. Estos niveles de riesgo pueden ayudar a programas de prevención y educación sísmica, e incluso a organismos de emergencia, tanto para actividades preventivas como formativas.

El interés por los terremotos es atemporal, su carácter y naturaleza violenta han mantenido ocupada a la ciencia, ingeniería, economía, y muchas otras disciplinas. Aún existe gran interés en comprender las causas y variables de interés que se relacionan con los terremotos, y las tecnologías empleadas para registrar cada movimiento sísmico nos permiten obtener datos desde cientos de sensores al unísono, con ello obteniendo un registro de gran calidad de cada evento, que incluye incluso metadatos de las fuentes y errores.



Resumen de metadatos

La data es obtenida a través de la plataforma kaggle, y se titula Significant Earthquakes 1965-2016, y que incluye movimientos de carácter significativo.

Link: <https://www.kaggle.com/datasets/usgs/earthquake-database>

#	Column	Non-Null Count	Dtype
0	Date	23412 non-null	object
1	Time	23412 non-null	object
2	Latitude	23412 non-null	float64
3	Longitude	23412 non-null	float64
4	Type	23412 non-null	object
5	Depth	23412 non-null	float64
6	Depth Error	4461 non-null	float64
7	Depth Seismic Stations	7097 non-null	float64
8	Magnitude	23412 non-null	float64
9	Magnitude Type	23409 non-null	object
10	Magnitude Error	327 non-null	float64
11	Magnitude Seismic Stations	2564 non-null	float64
12	Azimuthal Gap	7299 non-null	float64
13	Horizontal Distance	1604 non-null	float64
14	Horizontal Error	1156 non-null	float64
15	Root Mean Square	17352 non-null	float64
16	ID	23412 non-null	object
17	Source	23412 non-null	object
18	Location Source	23412 non-null	object
19	Magnitude Source	23412 non-null	object
20	Status	23412 non-null	object

dtypes: float64(12), object(9)
memory usage: 3.8+ MB



Resumen de metadatos

Descripción de variables tipo "object" del modelo:

```
Date: ['01/02/1965' '01/04/1965' '01/05/1965' ... '12/28/2016' '12/29/2016'
'12/30/2016']
Time: ['13:44:18' '11:29:49' '18:05:58' ... '12:38:51' '22:30:19' '20:08:28']
Type: ['Earthquake' 'Nuclear Explosion' 'Explosion' 'Rock Burst']
Magnitude Type: ['MW' 'ML' 'MH' 'MS' 'MB' 'MWC' 'MD' nan 'MWB' 'MWW' 'MWR']
ID: ['ISCGEM860706' 'ISCGEM860737' 'ISCGEM860762' ... 'US10007NAF'
'US10007NL0' 'US10007NTD']
Source: ['ISCGEM' 'ISCGEMSUP' 'OFFICIAL' 'CI' 'US' 'NC' 'GCMT' 'UW' 'ATLAS' 'NN'
'SE' 'AK' 'PR']
Location Source: ['ISCGEM' 'CI' 'US' 'H' 'U' 'G' 'NC' 'B' 'GCMT' 'AG' 'UW' 'SPE' 'HVO'
'BRK' 'ATLAS' 'AGS' 'PGC' 'BOU' 'SLC' 'OTT' 'AEI' 'AEIC' 'CASC' 'ISK'
'ATH' 'THE' 'ROM' 'MDD' 'WEL' 'GUC' 'UNM' 'CSEM' 'RSPR' 'JMA' 'NN' 'CAR'
'SJA' 'TEH' 'BEO' 'UCR' 'SE' 'TUL' 'TAP' 'THR' 'LIM' 'US_WEL' 'AK' 'PR']
Magnitude Source: ['ISCGEM' 'OFFICIAL' 'CI' 'US' '1020' 'BRK' 'NC' '1000' 'GCMT' '1009' 'UW'
'1023' 'ATLAS' 'HRV' 'PAR' 'NIED' 'NN' 'SE' 'PGC' 'US_GCMT' 'US_PGC' 'AK'
'PR' 'GUC']
Status: ['Automatic' 'Reviewed']
```



Preguntas e hipótesis

El objetivo del análisis es comprender las posibles relaciones entre las variables físicas, geográficas y temporales relacionadas con los terremotos. Y para ello se plantean tres preguntas tras las cuales podrían haber relaciones hipotéticas:

Primera pregunta: ¿Existe una relación entre la magnitud y la ubicación?

Segunda pregunta: ¿Existe una relación entre la magnitud y la fecha?

Tercera Pregunta: ¿Existe una relación entre la magnitud y la profundidad?

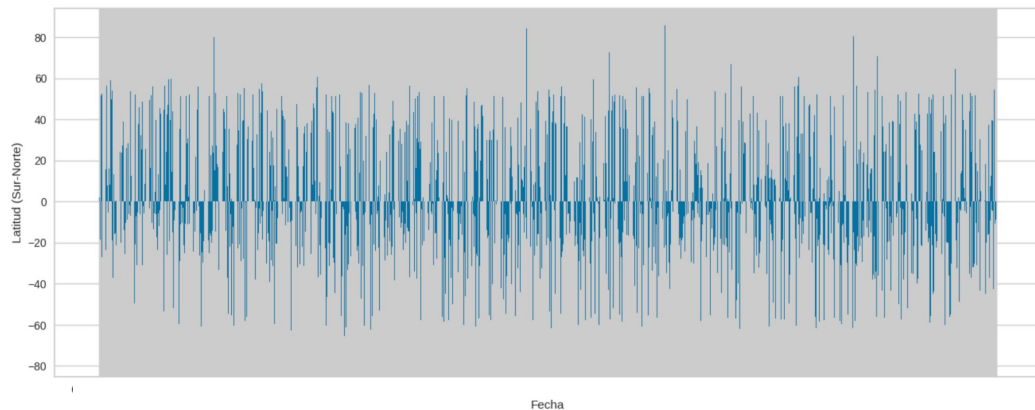
¿Relación entre la magnitud y la ubicación?

Al observar la data graficada sobre un mapa, obtendremos una imagen familiar, donde las zonas de alta actividad sísmica se hacen notar fuertemente (véase el cinturón de fuego del pacifico). La ubicación de los terremotos está fuertemente asociada a las zonas de confluencia tectónica, caracterizadas por su alta actividad. Fuera de estas zonas la actividad decae fuertemente.



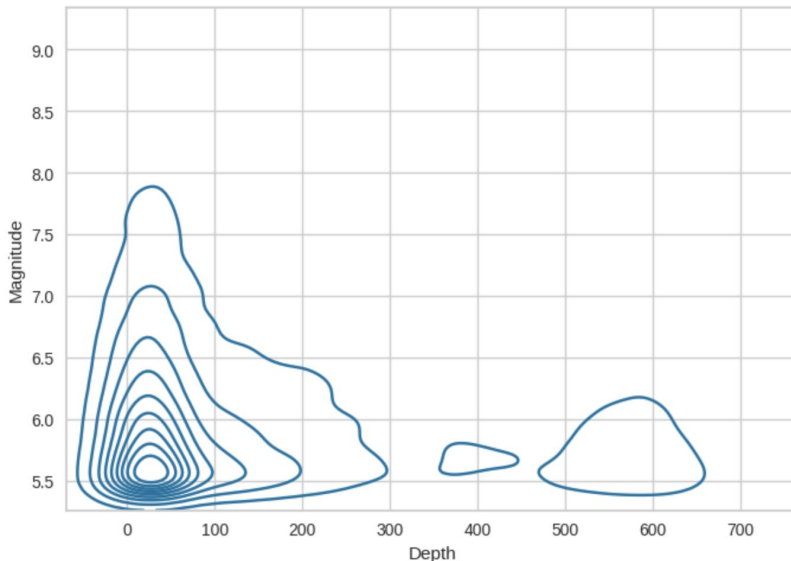
¿Relación entre la magnitud y la fecha?

Al observar la data a través del tiempo no obtendremos una relación evidente, e incluso si la separamos en hemisferios norte y sur. Sin embargo hoy contamos con modelos entrenados para reconocer patrones característicos de sismos de determinadas magnitudes, por lo que si existiría una relación tiempo/magnitud, la cual radica en la distribución de las ondas percibidas.



¿Relación entre la profundidad y la magnitud?

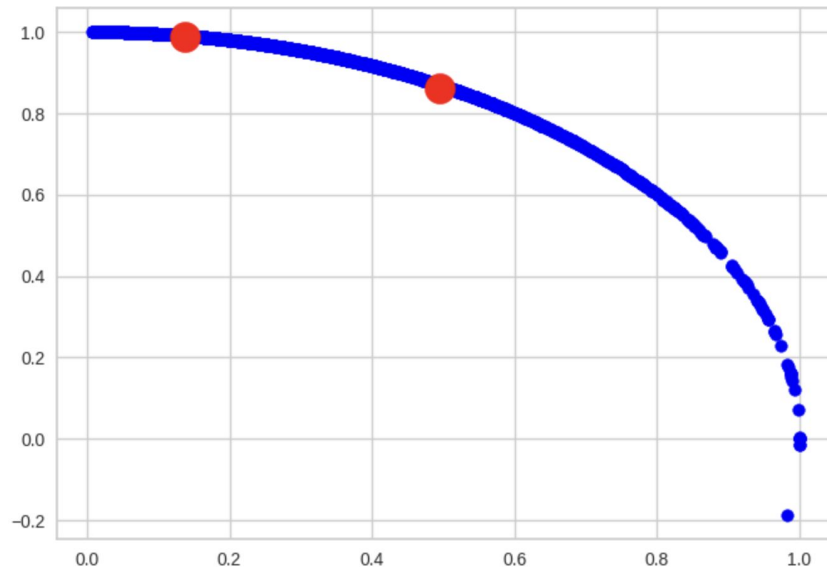
En esta pregunta nos detendremos a analizar los datos con mayor detención, ya que a simple vista no se ve una relación evidente, pero a la vez también se observan zonas con alta densidad eventos, lo que sugiere la existencia de clusters ocultos dentro de los datos.



¿Relación entre la profundidad y la magnitud?

Para identificar los clusters emplearemos k-means, un modelo de aprendizaje no supervisado.

De este proceso se obtendrán dos clusters, los cuales representan a su vez dos zonas de alta densidad de terremotos agrupados en magnitud y profundidad.



Dentro del dataset existen otras variables las cuales entregan información acerca del error en la medición del movimiento sísmico, de estas también es posible encontrar relaciones de interés que convendría analizar.

