

“AÑO DE LA RECUPERACIÓN Y CONSOLIDACIÓN
DE LA ECONOMÍA PERUANA”.



ESCUELA PROFESIONAL DE CIENCIA DE LA
COMPUTACIÓN
TÓPICOS CIENCIA DE DATOS

Informe Análisis Exploratorio de Datos

Estudiante:

Andrea del Rosario Lopez Condori

Docente :

ANA MARIA CUADROS
VALDIVIA



Hipótesis iniciales

Motivación

El análisis de eventos sísmicos permite estudiar fenómenos naturales y antropogénicos que generan vibraciones en la corteza terrestre. El catálogo del USGS incluye tanto sismos naturales como eventos artificiales (explosiones, pruebas, etc.). Las hipótesis surgieron al observar la presencia de eventos clasificados por tipo (`type`) y analizar variables físicas relacionadas como magnitud (`mag`) y profundidad (`depth`).

Hipótesis 1

Pregunta: ¿Existen diferencias estadísticamente significativas entre los sismos naturales y los eventos sísmicos artificiales en términos de magnitud, profundidad y distribución geográfica?

Hipótesis: Los eventos sísmicos de origen artificial tienden a presentar magnitudes menores, ocurren a profundidades más superficiales y se localizan en regiones geográficas específicas y repetitivas, en contraste con los sismos naturales que muestran mayor variabilidad en estas características.

Hipótesis 2

Pregunta: ¿Las regiones del Cinturón de Fuego del Pacífico concentran una mayor proporción de eventos sísmicos con magnitud mayor o igual a 6.0 en comparación con otras zonas geográficas?

Hipótesis: Los sismos de magnitud ≥ 6.0 se concentran predominantemente en las regiones del Cinturón de Fuego del Pacífico, incluyendo países como Japón, Chile, Indonesia y Alaska, debido a la alta actividad tectónica de estas zonas.

Hipótesis 3

Pregunta: ¿Se ha incrementado la frecuencia de eventos sísmicos con magnitud igual o superior a 5.0 durante la última década en comparación con periodos anteriores?

Hipótesis: La frecuencia de sismos con magnitud ≥ 5.0 ha mostrado un aumento leve en los últimos 10 años, atribuible tanto a mejoras en la sensibilidad y cobertura instrumental como a variaciones en la actividad tectónica global.

Plan de análisis

1. Exploración inicial del dataset: tipos de datos, valores faltantes, distribución.
2. Limpieza: eliminación de duplicados, tratamiento de valores nulos.
3. Transformaciones: creación de columnas `year` y `hour`, normalización de texto.
4. Visualizaciones exploratorias por tipo de evento, ubicación y magnitud.
5. Comparación entre eventos naturales y artificiales usando boxplots y mapas.

6. Análisis temporal y espacial para tendencias de magnitud y frecuencia.

Fuente de datos

Fuente: U.S. Geological Survey (USGS)

URL: <https://earthquake.usgs.gov/earthquakes/search/>

Fecha de obtención: Mayo 2025

Responsables: USGS (agencia científica del gobierno de EE.UU.)

Técnica de recolección: Redes sísmicas globales e instrumentación automatizada

Área del conocimiento: Geofísica, Ciencia de Datos

Problema computacional: Diferenciar eventos naturales y artificiales con base en datos físicos

Variables importantes: magnitud, profundidad, tipo de evento, localización

Referencias:

- USGS Earthquake Catalog: <https://earthquake.usgs.gov/data/comcat/>
- Kim, W.-Y. (2003). "The discrimination of nuclear explosions from earthquakes using seismic data" *Bulletin of the Seismological Society of America (BSSA)*.

Descripción del conjunto de datos

A nivel de atributos

- **Periodo cubierto:** 1990–2024
- **Tamaño:** 489,126 registros, 22 columnas
- **Principales atributos:**
 - **time** (datetime): Fecha/hora del evento. No nulos.
 - **latitude, longitude** (float): Coordenadas del epicentro. Rango latitud: -84.4 a 87.3, longitud: -179.9 a 180.
 - **depth** (float): Profundidad en km. Rango: -3.7 a 735.8
 - **mag** (float): Magnitud. Rango: 3.38 a 9.1
 - **magType** (string): Tipo de medición (mb, ml, md). 25 valores únicos.
 - **type** (string): Tipo de evento (earthquake, explosion, blast). Variable categórica nominal.
 - **nst, magNst** (int): Número de estaciones. Hasta 954. Muchos valores nulos.
 - **gap, dmin, rms**: Parámetros técnicos con dispersión en sus valores.

A nivel de registros

Cada fila representa un evento sísmico único. No hay etiquetas de clasificación, pero se distingue por tipo (**type**). Nivel de granularidad: evento individual.

Relación entre atributos

Se detectan correlaciones débiles entre magnitud y profundidad. Algunas variables muestran correlación según el tipo de evento. Variables categóricas como `type` permiten diferenciar eventos por su comportamiento físico.

Terminología especial

- **magType**: tipo de magnitud (mb, ml, mw)
- **rms**: error cuadrático medio
- **gap**: cobertura angular entre estaciones

Cuadro resumen de atributos: (ver tablas de resumen en anexos del análisis estadístico)

Formato

El dataset original se encontraba en formato CSV descargado directamente desde el portal del USGS.

Transformaciones

- Conversión de `time` a tipo `datetime`.
- Extracción de columnas `year` y `hour` para análisis temporal.
- Normalización del texto en `place` (minúsculas, eliminación de espacios).

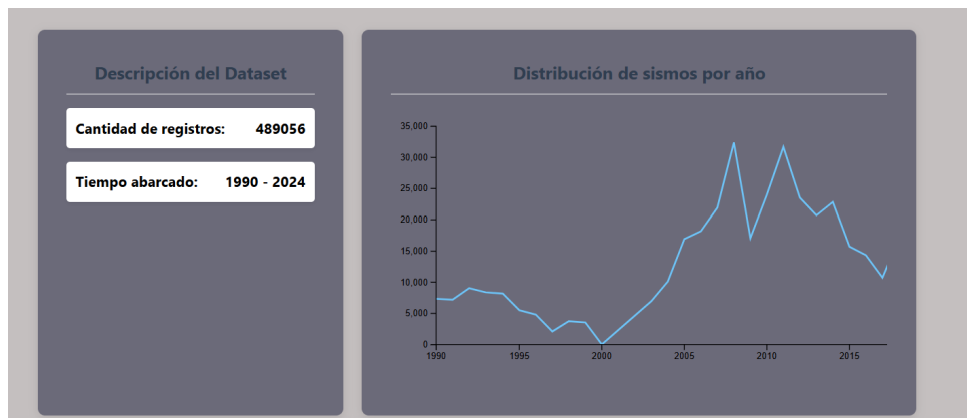
Limpieza de datos

- Eliminación de más de 114,000 registros duplicados.
- Eliminación de registros con valores nulos en columnas críticas.
- Conversión de tipos de datos y validación de rangos válidos.

Exploración

Se realizaron más de 10 visualizaciones, entre ellas:

1. Histograma de magnitudes.



2. Boxplot de profundidad por tipo de evento.

Boxplot de Profundidades

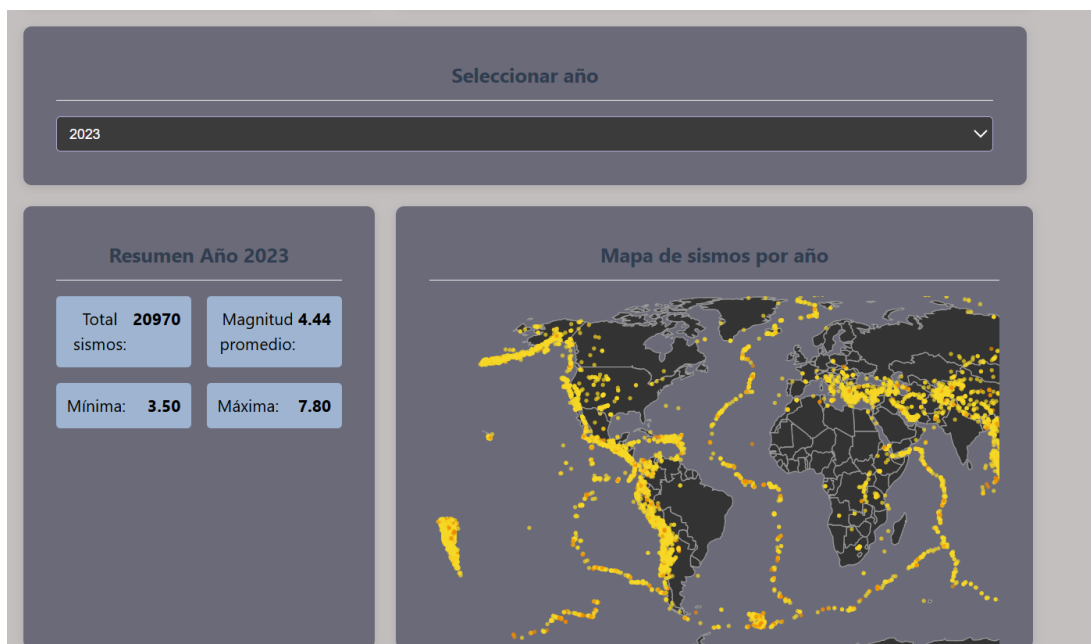


3. Boxplot de magnitud por tipo de evento.

Boxplot de Magnitudes

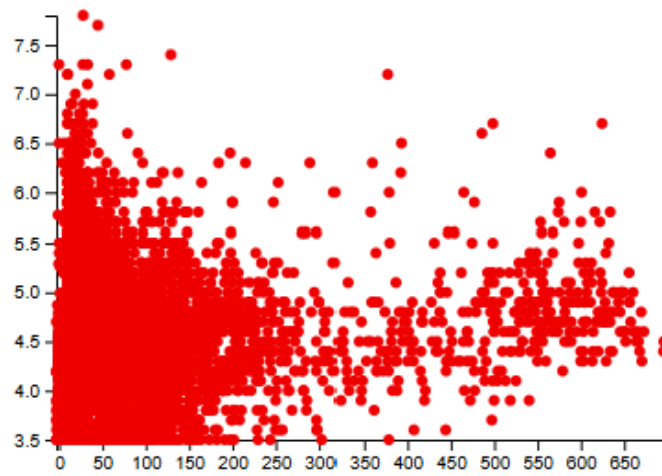


4. Mapa geográfico de eventos sísmicos.

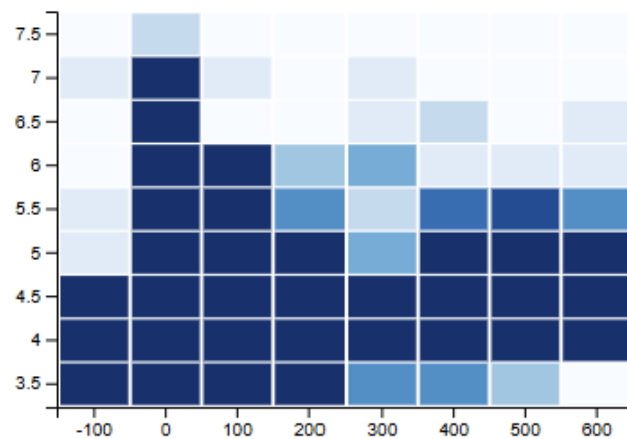


5. Frecuencia anual de sismos con magnitud $\geq 5,0$.

6. Diagrama de dispersión profundidad vs magnitud.



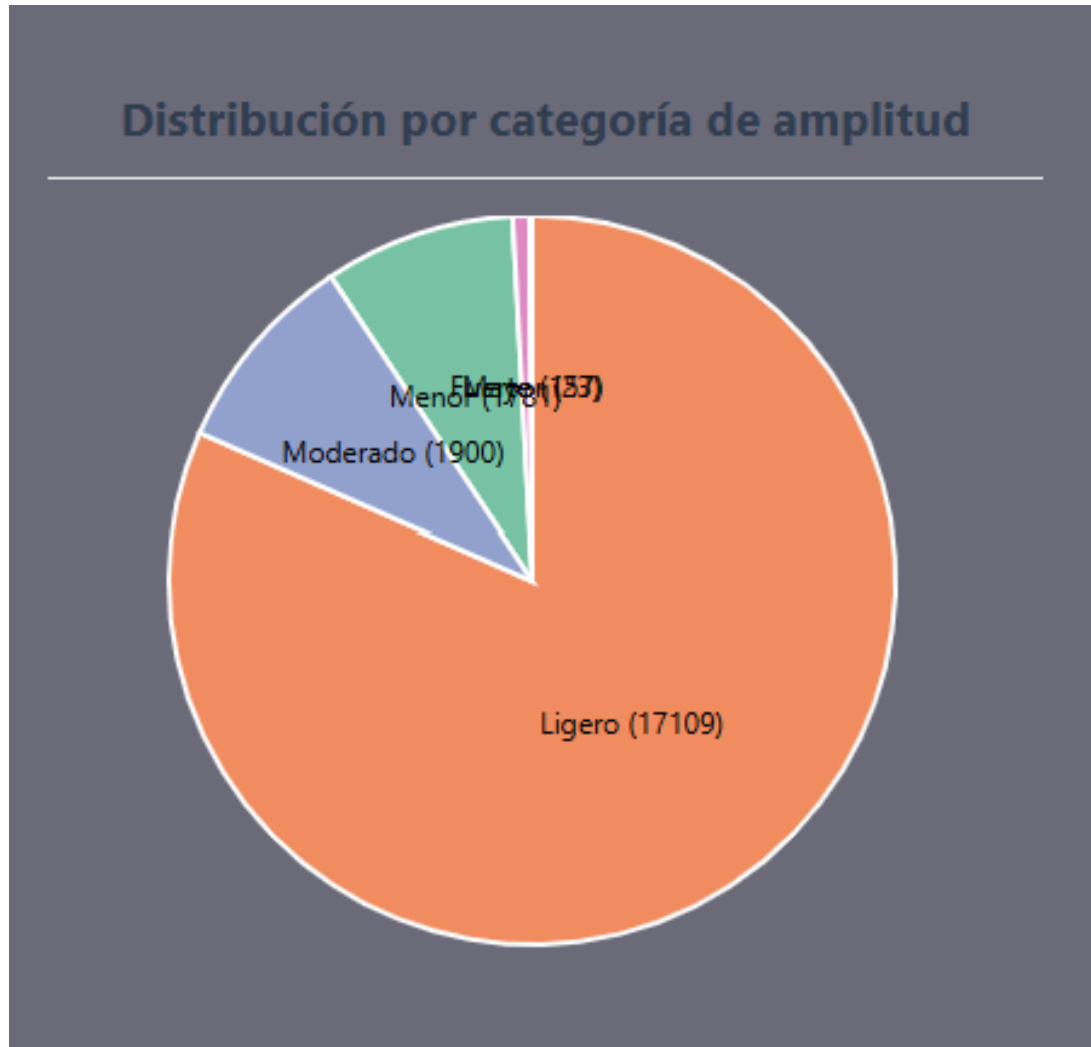
7. Matriz de correlación (heatmap).



8. Barras de frecuencia por meses.



9. Distribución de eventos por hora del día.
10. Distribución de tipos de magnitud (magType).



Conclusión

Conocimiento obtenido:

- Los sismos artificiales se diferencian por presentar magnitudes menores y ocurrir a profundidades más superficiales.
- Las zonas más sísmicas corresponden al Cinturón de Fuego del Pacífico, confirmando alta concentración de eventos fuertes.
- No se observa una tendencia clara y sostenida de incremento en la frecuencia de eventos sísmicos en la última década.

Conclusiones por hipótesis:

- **Hipótesis 1:** Parcialmente confirmada. Se evidencian diferencias físicas entre eventos naturales y artificiales.

- **Hipótesis 2:** Confirmada. Los sismos de magnitud mayor o igual a 6.0 se concentran geográficamente en el Cinturón de Fuego.
- **Hipótesis 3:** No confirmada definitivamente. La frecuencia de eventos no muestra un aumento claro y sostenido.

Anexos

- **PDF Colab:** <https://colab.research.google.com/drive/1hydS3hKnqoiVWV28IR318u7oBnCAusp=sharing>
- **Etapas del ciclo de vida de Ciencia de Datos:** Andrea del Rosario López Condori
- **Dashboard:** <https://andrealopezco20.github.io/DASHBOARDOF/>