

---

# Trabajo Final de Estructuras de Datos y su Almacenamiento

---

Sebastián Jaimes  
Leonard Kennedy  
Juan Borquez  
Juan Camilo Bohórquez Pérez

La Salle Barcelona  
Universitat Ramon Llull  
2024

# ÍNDICE

<b>1. Introducción</b>	<b>2</b>
1.1. Contexto General	2
1.2. Importancia del Análisis de Datos Sísmicos	2
1.3. Metodología de Investigación	2
<b>2. Extracción de Datos</b>	<b>3</b>
2.1. Acceso a la Información Sísmica	3
2.2. Selección Preliminar de Datos	3
2.3. Datos Recolectados	3
2.4. Proceso de Extracción de Datos	3
<b>3. Transformación de Datos</b>	<b>3</b>
3.1. Limpieza y Preparación	3
3.2. Primera Fase de Transformación de Datos	4
3.2.1. Reorganización de Coordenadas Geográficas	4
3.2.2. Análisis de Centros de Observación	4
3.2.3. Corrección de la Localización	4
3.2.4. Enriquecimiento del Conjunto de Datos	4
3.2.5. Visualización de Datos	4
3.3. Segunda Fase de Transformación de Datos	6
3.3.1. Preparación de Datos para Visualización	6
3.3.2. Mapas de Dispersión Geográfica	6
3.3.3. Histogramas y Gráficos de Distribución	6
3.3.4. Análisis de la Profundidad vs Magnitud de Terremotos	7
3.3.5. Correlación entre la Actividad Sísmica y la Ubicación de las Placas Tectónicas	8
3.3.6. Series Temporales de Magnitud	9
<b>4. Load</b>	<b>11</b>
4.1. Establecimiento de Conexión con MongoDB	11
4.2. Inserción de Datos en la Colección	11
<b>5. Visualización de Datos</b>	<b>12</b>
5.1. Interfaz de Usuario	12
5.2. Presentación de la Información	12
<b>6. Conclusiones</b>	<b>13</b>

# 1. INTRODUCCIÓN

## 1.1. CONTEXTO GENERAL

Los terremotos son fenómenos naturales que ocurren de manera impredecible y tienen el potencial de causar impactos significativos en las sociedades humanas. Este trabajo se enmarca en el estudio de estos eventos con el objetivo de aportar conocimientos que puedan contribuir a la prevención de riesgos y a la mejora de las condiciones de vida en las áreas donde suceden con frecuencia.

## 1.2. IMPORTANCIA DEL ANÁLISIS DE DATOS SÍSMICOS

El análisis detallado de la actividad sísmica no solo es de interés para la comunidad científica, sino también para el diseño de políticas de construcción, planes de emergencia y sistemas de alerta temprana que pueden salvar vidas. Por tanto, la representación y comprensión de los datos sobre terremotos es un paso crucial para avanzar en estas áreas.

## 1.3. METODOLOGÍA DE INVESTIGACIÓN

En esta investigación, se adopta una metodología de ETL, siglas en inglés para Extracción, Transformación y Carga, con el fin de procesar y analizar los datos de terremotos proporcionados por la API del Servicio Geológico de Estados Unidos (USGS). Este método nos permite organizar y preparar los datos para su posterior análisis y comprensión.

El siguiente diagrama de flujo muestra el movimiento del dato desde su obtención de la API, hasta la página web que permite al usuario visualizarlo en contexto. El enlace a la página web se encuentra adjunto en paquete de entregables.

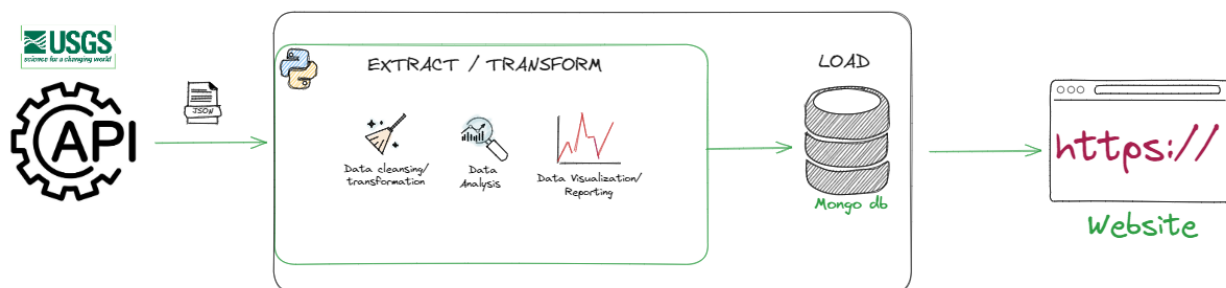


Figura 1: Diagrama de flujo del proceso.

## 2. EXTRACCIÓN DE DATOS

### 2.1. ACCESO A LA INFORMACIÓN SÍSMICA

La fase inicial de nuestro trabajo consiste en recopilar datos de la API pública del USGS. Para ello, utilizamos la librería de programación Python requests, que nos permite interactuar con la API y obtener la información necesaria. Los datos recopilados incluyen una amplia gama de variables que describen los terremotos, desde su ubicación geográfica hasta detalles técnicos como la magnitud y la profundidad.

### 2.2. SELECCIÓN PRELIMINAR DE DATOS

En la gran cantidad de datos recopilados, se hace una selección inicial basada en la magnitud de los terremotos. Se establece un umbral de magnitud 3.0 en la escala de Richter, ya que eventos con una magnitud inferior suelen ser imperceptibles y de menor interés para los objetivos de este estudio. Este filtro preliminar asegura que el análisis se centre en eventos con potencial de tener un impacto más significativo.

### 2.3. DATOS RECOLECTADOS

El enlace específico utilizado para la extracción es <https://earthquake.usgs.gov/fdsnws/event/1/query?format=geojson>, que nos brinda información actualizada sobre los terremotos más recientes. Cada registro de terremoto viene con información detallada, que va desde la latitud y longitud del evento, hasta el nombre del lugar o país afectado, pasando por la posibilidad de que haya generado un tsunami, e incluye enlaces a descripciones y datos más detallados.

### 2.4. PROCESO DE EXTRACCIÓN DE DATOS

El proceso de extracción se realiza mediante comandos específicos que permiten filtrar y descargar los datos relevantes. Este proceso se explica con más detalle en las secciones técnicas siguientes, donde también se abordará cómo se transforman y se preparan los datos para su análisis.

## 3. TRANSFORMACIÓN DE DATOS

### 3.1. LIMPIEZA Y PREPARACIÓN

Una vez extraídos los datos, el siguiente paso es limpiarlos y prepararlos para el análisis. Esto incluye eliminar registros duplicados, corregir errores y completar información faltante, para asegurarnos de que el análisis se realice sobre una base sólida y confiable.

## **3.2. PRIMERA FASE DE TRANSFORMACIÓN DE DATOS**

### **3.2.1. REORGANIZACIÓN DE COORDENADAS GEOGRÁFICAS**

Inicialmente, los datos recibidos de la API ubicaban las coordenadas de latitud y longitud en una única columna. Para facilitar el análisis, estas han sido separadas, asignando a cada una su propia columna. Esta reorganización es fundamental para simplificar los pasos subsiguientes del proceso de transformación.

### **3.2.2. ANÁLISIS DE CENTROS DE OBSERVACIÓN**

La información proporcionada por la API incluye el lugar de registro de cada terremoto. Hemos compilado estos datos en una tabla de frecuencia para identificar los centros de observación más activos del último mes. Además, esta tabla nos ofrece un valor agregado: la magnitud promedio registrada en cada centro, lo cual es esencial para entender la intensidad de la actividad sísmica por ubicación.

### **3.2.3. CORRECCIÓN DE LA LOCALIZACIÓN**

En el campo "place" de la API, se ofrece una descripción del lugar del terremoto, incluyendo el país. No obstante, se ha detectado que la asignación de países puede ser inexacta, especialmente en los datos referentes a Estados Unidos, donde en lugar de aparecer el país, se menciona el estado. Para solucionar esto y obtener la localización precisa de cada terremoto, hemos empleado los datos de latitud y longitud con la ayuda de la librería 'reverse geocode'. Así, transformamos las coordenadas en el nombre del país correspondiente.

### **3.2.4. ENRIQUECIMIENTO DEL CONJUNTO DE DATOS**

Con la adición de la nueva columna 'country', ahora es posible elaborar una tabla de frecuencia mejorada que muestra los países con la mayor cantidad de terremotos en el último mes, junto con la magnitud promedio de estos eventos. Esta información es vital para identificar patrones de actividad sísmica a nivel mundial.

### **3.2.5. VISUALIZACIÓN DE DATOS**

Finalmente, se ha generado un histograma que ilustra los 15 países con la actividad sísmica más alta en el último mes. Este tipo de visualización permite una rápida comprensión de los datos y facilita la identificación de áreas de alto riesgo.

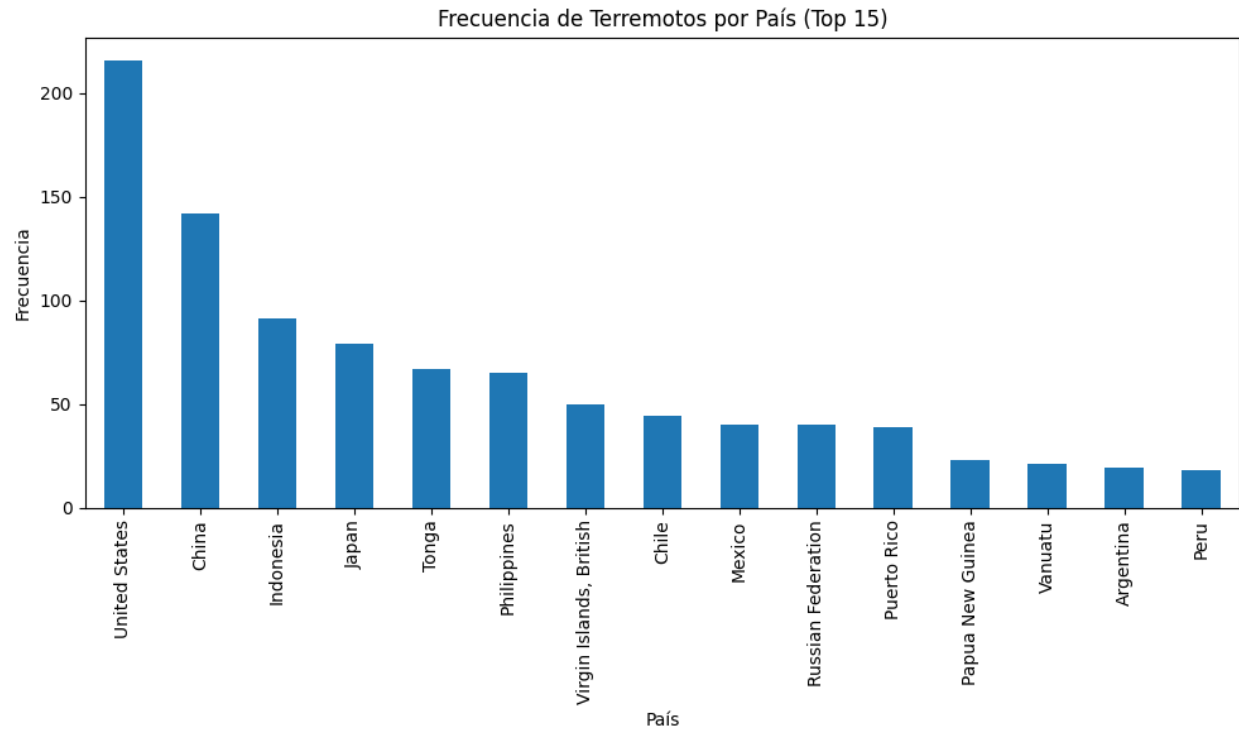


Figura 2: Histograma mensual de países con mayor actividad sísmica.

### 3.3. SEGUNDA FASE DE TRANSFORMACIÓN DE DATOS

Las visualizaciones de los datos sísmicos son un paso crucial en nuestro proceso de ETL, ya que permiten una comprensión más profunda y accesible de los patrones y tendencias subyacentes. En esta sección, se describe cómo se utilizaron las bibliotecas de Python, específicamente Plotly, para transformar los datos de terremotos en representaciones visuales interactivas.

#### 3.3.1. PREPARACIÓN DE DATOS PARA VISUALIZACIÓN

Los datos extraídos en la etapa anterior se procesaron y limpiaron para garantizar la calidad y la coherencia necesarias para la visualización efectiva. Este proceso incluyó la normalización de las magnitudes de los terremotos y la conversión de los timestamps UNIX a fechas legibles.

#### 3.3.2. MAPAS DE DISPERSIÓN GEOGRÁFICA

Utilizando Plotly, se crearon mapas de dispersión geográfica que muestran la ubicación y magnitud de los terremotos. Estos mapas proporcionan una vista intuitiva de la distribución global de la actividad sísmica y resaltan las áreas de mayor riesgo.

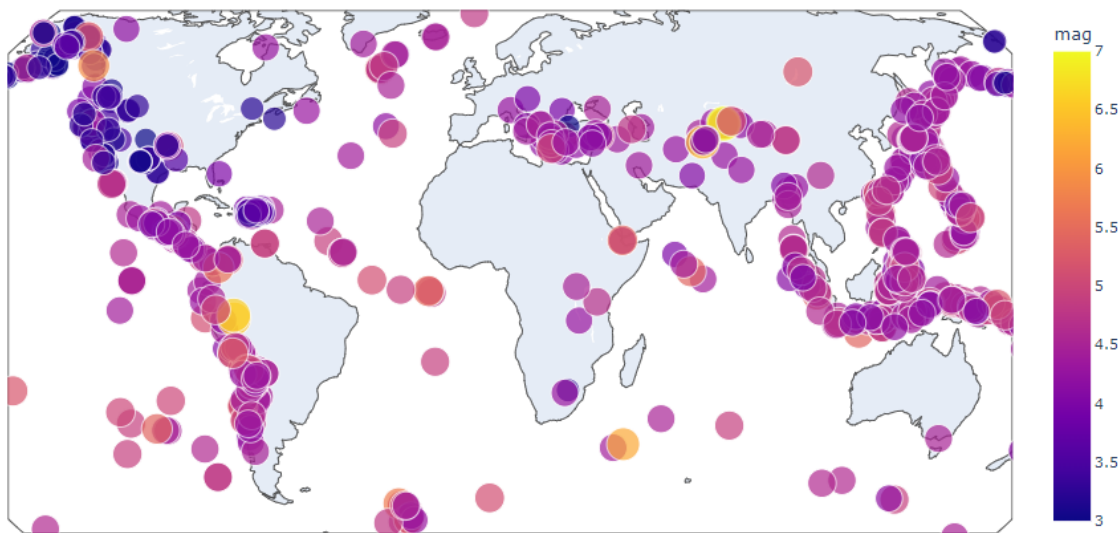


Figura 3: Mapa de dispersión de los terremotos alrededor del mundo.

#### 3.3.3. HISTOGRAMAS Y GRÁFICOS DE DISTRIBUCIÓN

Se generaron histogramas para visualizar la frecuencia de los terremotos por magnitud, lo que permite identificar rápidamente los rangos de magnitud más comunes y los eventos sísmicos más significativos.

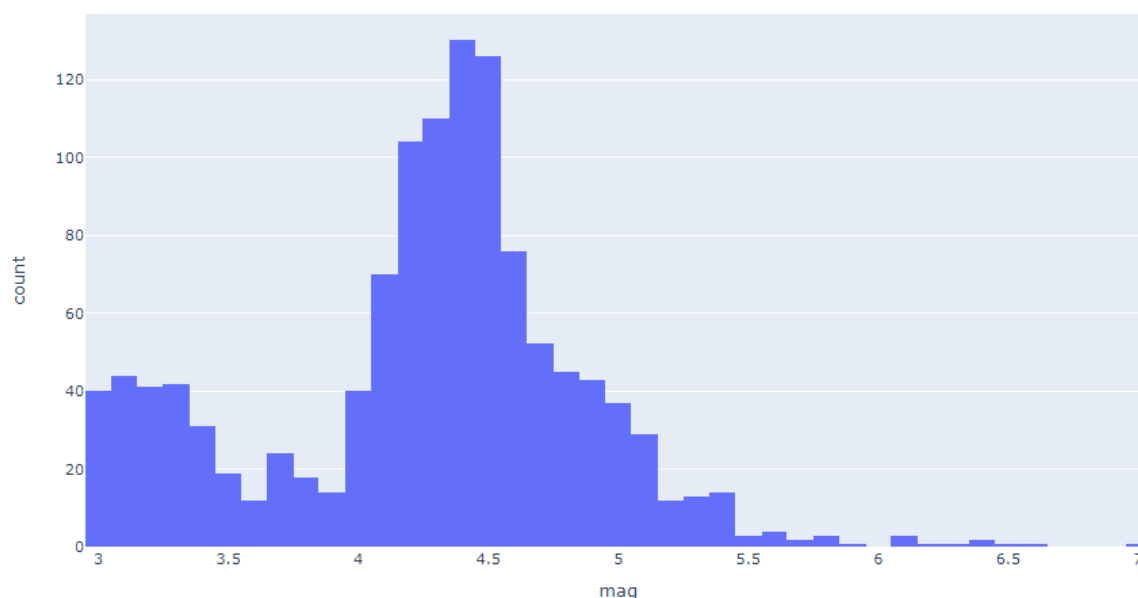


Figura 4: Histograma de la distribución de magnitudes de terremotos.

#### 3.3.4. ANÁLISIS DE LA PROFUNDIDAD VS MAGNITUD DE TERREMOTOS

La relación entre la profundidad y la magnitud de los eventos sísmicos se visualiza mediante un gráfico de dispersión, que nos ayuda a comprender si existe alguna tendencia entre la profundidad a la que ocurren los terremotos y su magnitud. En nuestro análisis, observamos una gran cantidad de terremotos que ocurren a profundidades menores a 100 kilómetros y con magnitudes que oscilan principalmente entre 3 y 6 en la escala Richter. Este rango sugiere que la mayoría de los terremotos registrados son moderadamente potentes y ocurren cerca de la superficie terrestre, lo cual es consistente con el comportamiento típico de la actividad sísmica global.

La visualización no muestra una correlación lineal clara entre la profundidad y la magnitud, indicando que factores adicionales podrían influir en la magnitud de un terremoto más allá de la profundidad a la que ocurre. Los datos se presentan bastante dispersos en el gráfico, sin seguir un patrón que indique que una variable aumenta o disminuye consistentemente con la otra.

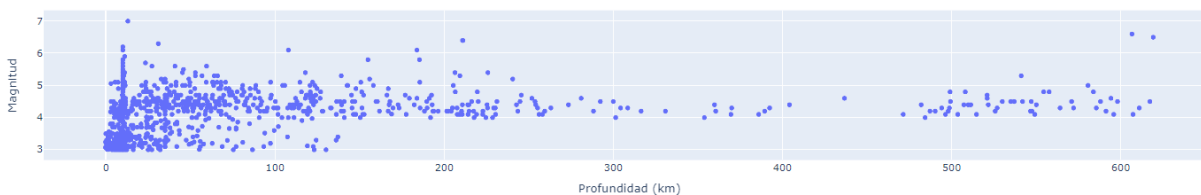


Figura 5: Gráfico de dispersión que muestra la relación entre la profundidad y la magnitud de los terremotos.



En el texto descriptivo bajo la figura, se resumen dos observaciones clave del gráfico:

- **Concentración de Datos:** La mayoría de los terremotos se concentran en profundidades menores a 100 km y con magnitudes moderadas, lo que sugiere que son fenómenos relativamente superficiales y de magnitud moderada.
- **Correlación:** La ausencia de una correlación clara entre la profundidad y la magnitud sugiere que la magnitud de un terremoto no está directamente relacionada con la profundidad a la que ocurre.

### 3.3.5. CORRELACIÓN ENTRE LA ACTIVIDAD SÍSMICA Y LA UBICACIÓN DE LAS PLACAS TECTÓNICAS

La distribución de la actividad sísmica registrada entre el 10 de enero de 2024 y el 9 de febrero de 2024 muestra una alineación significativa con las fronteras de las placas tectónicas del mundo. Esto se evidencia al comparar los patrones de localización de terremotos con un mapa de las placas tectónicas. Los terremotos tienden a concentrarse en áreas donde las placas tectónicas se encuentran o se deslizan unas contra otras, lo cual es coherente con el conocimiento actual de que los movimientos y colisiones de estas placas son a menudo el origen de los sismos.

Esta correlación se visualiza claramente al superponer los datos sísmicos sobre el mapa de placas tectónicas. Los contornos de alta actividad sísmica, marcados por los terremotos con magnitudes más significativas, siguen las líneas de las placas tectónicas, particularmente a lo largo de la Placa del Pacífico y los límites de la Placa Euroasiática.



Figura 6: Mapa de las placas tectónicas con la superposición de la actividad sísmica registrada, destacando la correlación entre terremotos y placas tectónicas.

Los datos sugieren fuertemente que las áreas de alta actividad sísmica están directamente asociadas con las interacciones de las placas tectónicas, y que las regiones de mayor riesgo sísmico son aquellas en los bordes de las placas. Esto subraya la importancia de monitorear estas zonas para la prevención de desastres y la preparación en caso de terremotos.

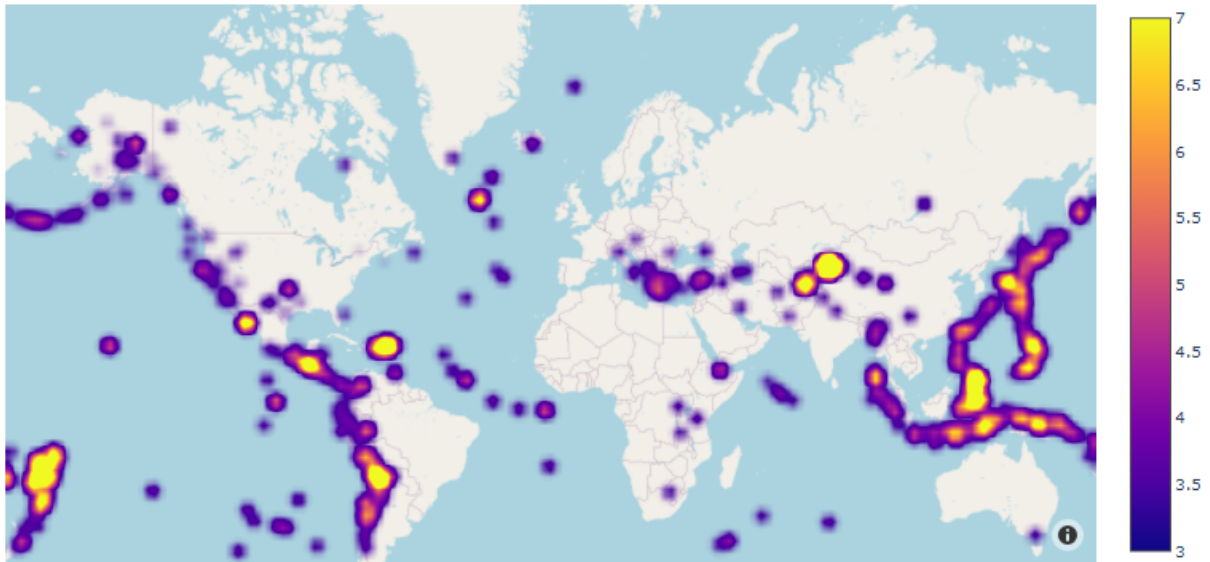


Figura 7: Gráfico que compara el patrón de terremotos con la ubicación de las placas tectónicas, ilustrando la correlación entre la actividad sísmica y las fronteras de las placas.

La consistencia de estos patrones fortalece la comprensión de que la tectónica de placas es un factor determinante en la ocurrencia y distribución de la actividad sísmica en la Tierra.

### 3.3.6. SERIES TEMPORALES DE MAGNITUD

Los gráficos de línea de tiempo muestran cómo la magnitud de los terremotos cambia a lo largo del tiempo, ofreciendo una perspectiva dinámica sobre la evolución de la actividad sísmica.

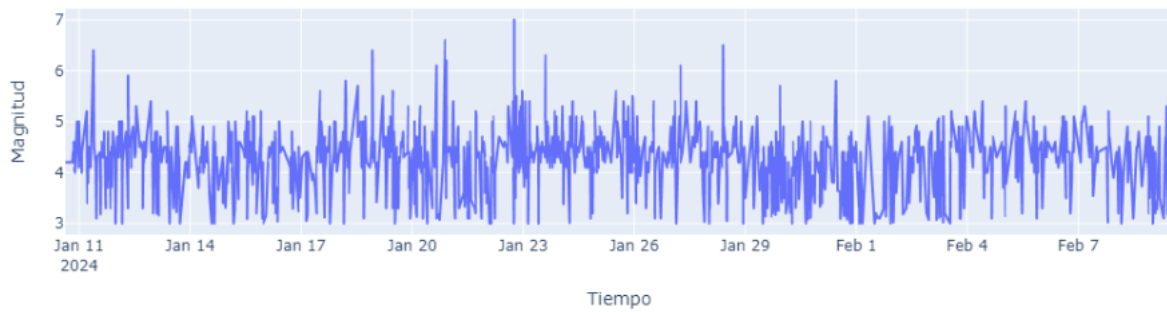


Figura 8: Serie temporal de magnitudes de terremotos.

Cada una de estas visualizaciones desempeña un papel fundamental en nuestra comprensión de los datos y en la comunicación de nuestros hallazgos a un público más amplio. En las etapas subsiguientes del proceso ETL, estos gráficos se integrarán en una base de datos MongoDB y se presentarán en un sitio web para acceso público.

## 4. LOAD

### 4.1. ESTABLECIMIENTO DE CONEXIÓN CON MONGODB

La primera etapa del proceso de carga de datos implica establecer una conexión efectiva con la base de datos MongoDB. El siguiente fragmento de código Python demuestra cómo se realiza esta conexión con la base de datos local 'bd terremotos', que opera en el puerto predeterminado. Se elige la colección 'terremotos' para las operaciones de manipulación de datos. Es crucial mencionar que este ejemplo prescinde de la autenticación por motivos de simplicidad, ya que la conexión se efectúa en un entorno de desarrollo local.

```
# Conectar a la base de datos MongoDB
client = MongoClient('mongodb://localhost:27017/')
db = client['bd_terremotos']
collection = db['terremotos']
```

Figura 9: Proceso de conexión a la base de datos MongoDB.

### 4.2. INSERCIÓN DE DATOS EN LA COLECCIÓN

Tras establecer la conexión, se procede con la inserción de datos. En nuestro caso, los datos corresponden a registros de actividad sísmica, cada uno representado como un documento en formato de diccionario de Python. Estos documentos son agregados a la colección 'terremotos', lo que permite que la información sea accesible para futuras consultas y análisis.

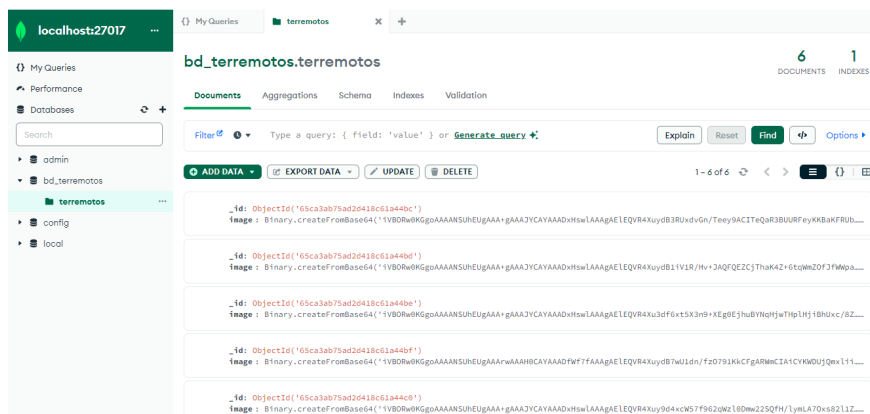


Figura 10: Ejemplo de la inserción de datos en la colección 'terremotos'.

## 5. VISUALIZACIÓN DE DATOS

### 5.1. INTERFAZ DE USUARIO

La interfaz de usuario juega un papel crucial en la visualización de los datos almacenados. Se ha desarrollado una página web que sirve como interfaz, permitiendo a los usuarios visualizar la información almacenada en la base de datos de MongoDB. Esta página es accesible desde cualquier navegador web estándar y está diseñada para ser intuitiva y fácil de usar.

### 5.2. PRESENTACIÓN DE LA INFORMACIÓN

La interfaz muestra los datos de manera estructurada y permite interacciones como el zoom y la manipulación de imágenes. La siguiente figura ilustra cómo se presenta la información relacionada con los terremotos en la página web.

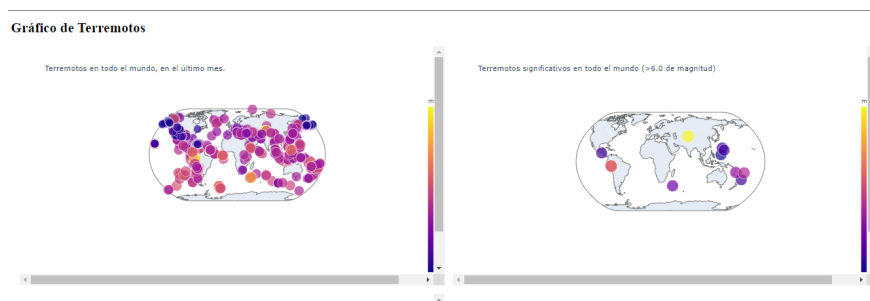


Figura 11: Interfaz de usuario para la visualización de datos de terremotos.

## 6. CONCLUSIONES

Siguiendo los conocimientos adquiridos en la clase de Estructura de Datos y su Almacenamiento, hemos decidido utilizar una base de datos NoSQL, específicamente MongoDB, para almacenar la información de forma local en nuestro proyecto. Esta elección nos permite mantener un repositorio actualizado cada vez que realicemos cambios o actualizaciones en nuestros datos.

En nuestro proyecto, hemos almacenado y creado una visualización de las imágenes que fueron otorgadas durante el análisis y la transformación de los datos. Estas imágenes representan información relevante para nuestro estudio y son un elemento fundamental en nuestro proceso de análisis de datos. La información que se visualiza no se consulta a la base de datos debido a que la conexión es local, y nuestro objetivo es que la persona pueda visualizar la información más reciente cargada.

Al elegir MongoDB como nuestro sistema de gestión de base de datos, obtenemos la flexibilidad necesaria para almacenar datos no estructurados, como imágenes, de manera eficiente y escalable. Esto nos permite organizar y acceder a nuestra información de manera efectiva, facilitando así el desarrollo de nuestro proyecto y la realización de análisis posteriores.

Además, al mantener la información localmente, tenemos un mayor control sobre nuestros datos y podemos trabajar en un entorno familiar y seguro. Esto nos brinda la confianza necesaria para realizar experimentos y pruebas sin preocuparnos por la integridad o la disponibilidad de nuestros datos. Hemos optado por MongoDB como solución de almacenamiento de datos debido a su capacidad para manejar datos no estructurados, su escalabilidad y su facilidad de uso.