



Pipeline de Cs. de Datos

Andrea López Condori





DATASET

El United States Geological Survey (USGS) es una agencia científica del gobierno de Estados Unidos dedicada al estudio de los recursos naturales, los peligros geológicos y la cartografía. Entre sus áreas clave, el USGS se especializa en la vigilancia sísmica a través del Advanced National Seismic System (ANSS), proporcionando datos públicos y en tiempo real sobre eventos sísmicos globales.

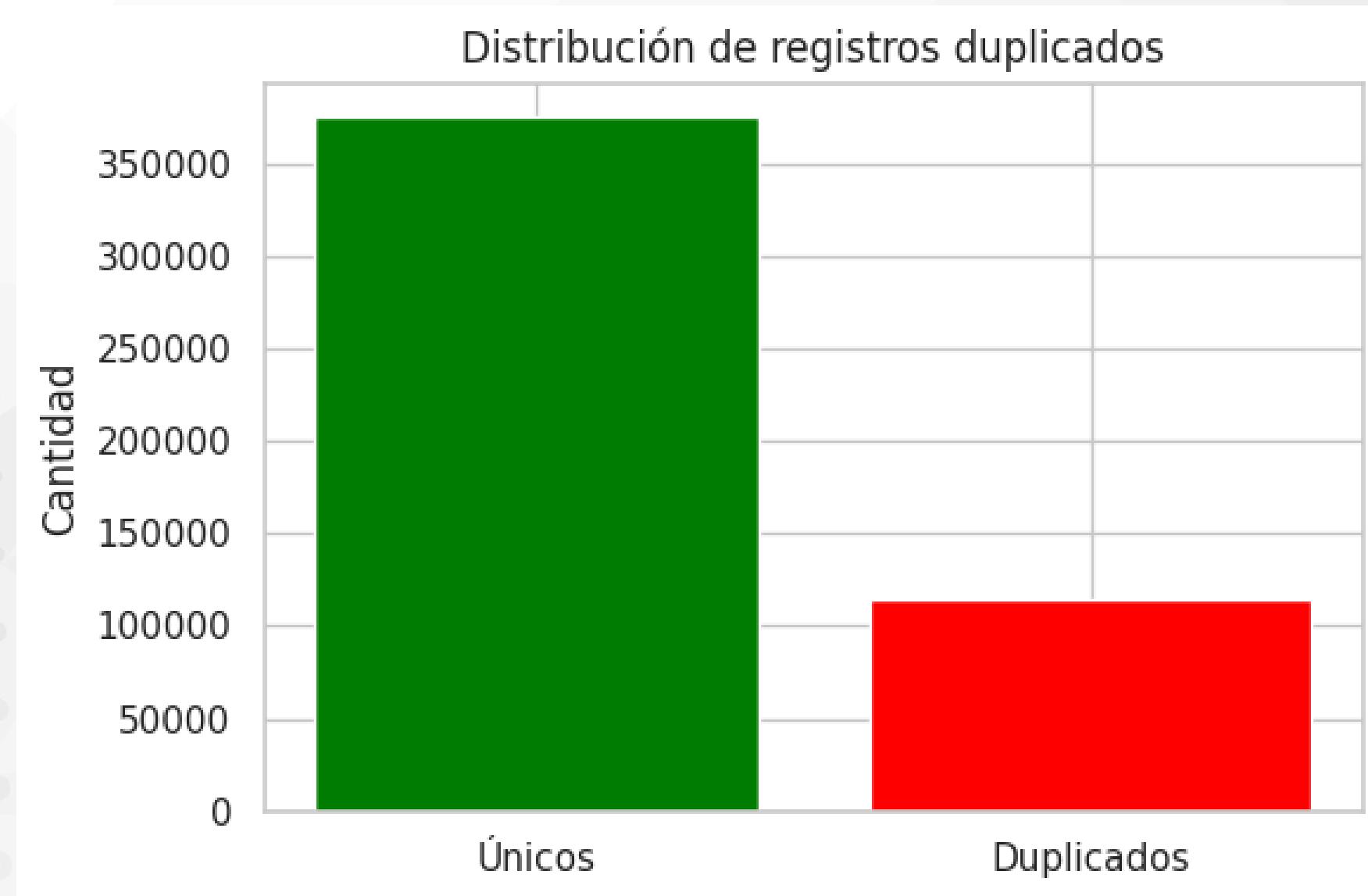
Los datos proporcionados por el USGS son ampliamente utilizados por la comunidad científica para estudios de predicción sísmica, modelamiento geológico, evaluación de riesgos y, más recientemente, para aplicaciones de ciencia de datos como la clasificación de eventos naturales frente a eventos inducidos por el ser humano.

¿Qué problema identifican en el dataset?

Durante la revisión inicial del conjunto de datos sísmicos proporcionado por el USGS, se detectaron múltiples problemas que comprometen la calidad del análisis posterior. Estos problemas no solo afectan la integridad del dataset, sino también la validez de cualquier conclusión que pudiera extraerse de forma directa sin aplicar un proceso riguroso de limpieza. Los principales problemas identificados fueron:

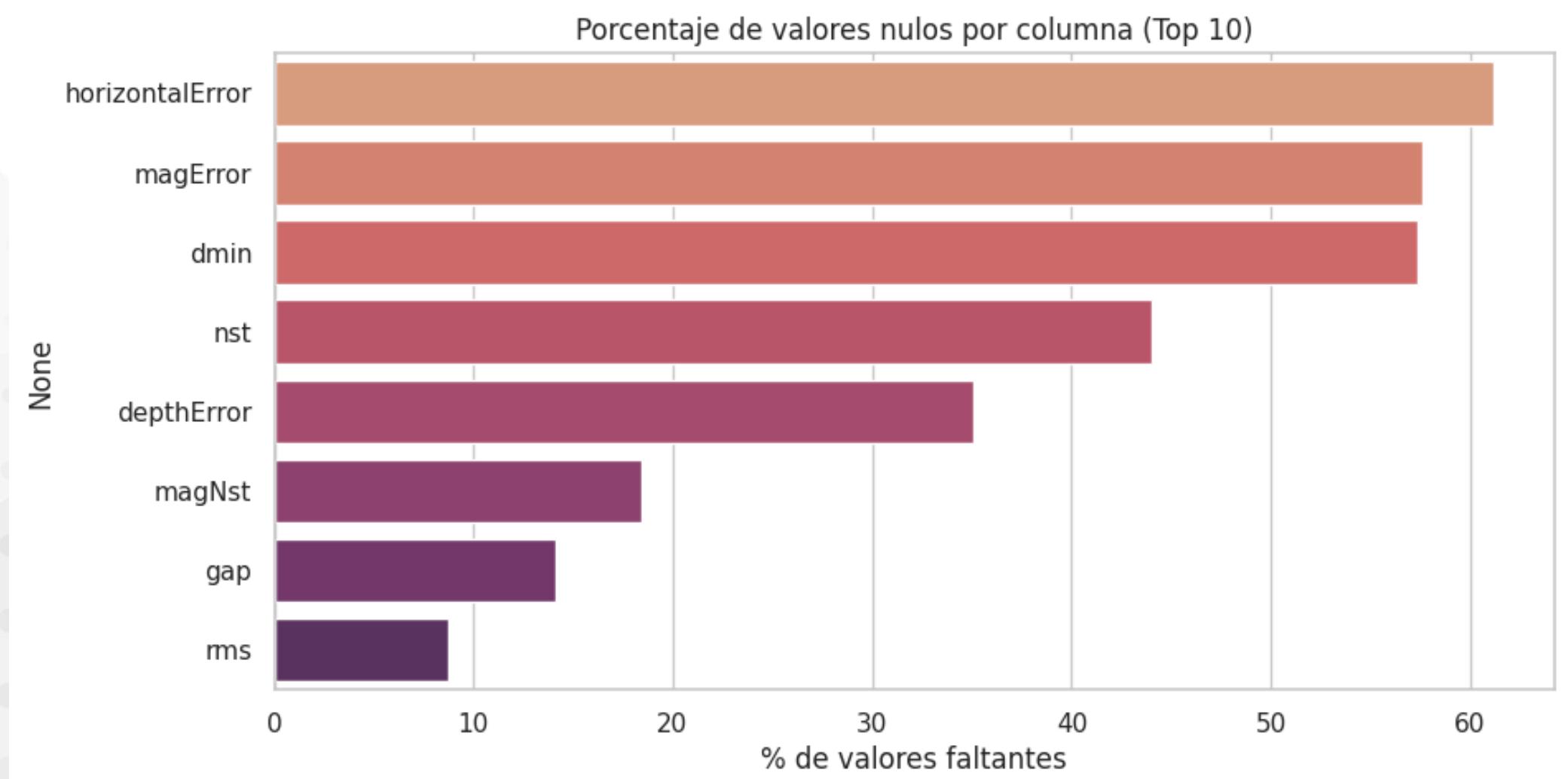
Registros duplicados:

Se encontraron más de 114,000 registros idénticos, lo cual representa aproximadamente el 23% del total. Si no se eliminan, estos inflan la frecuencia de eventos y distorsionan los análisis estadísticos.



Valores faltantes:

Varias columnas presentan un porcentaje elevado de valores nulos, especialmente:





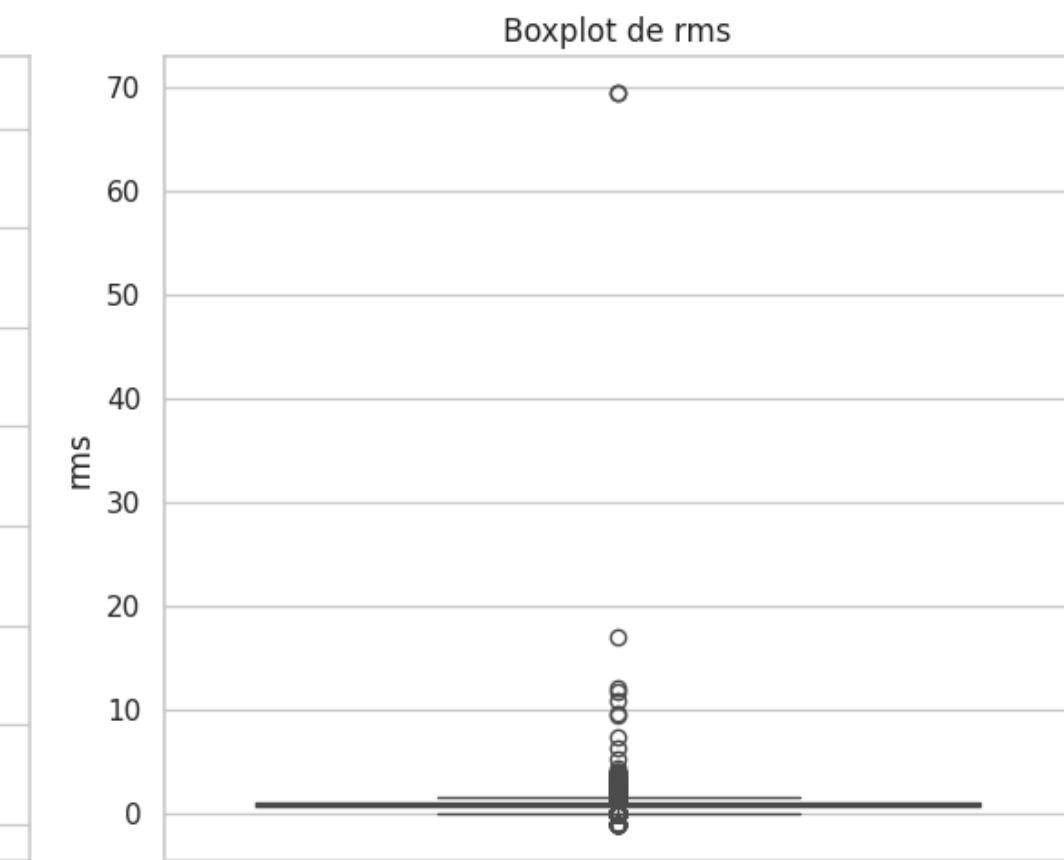
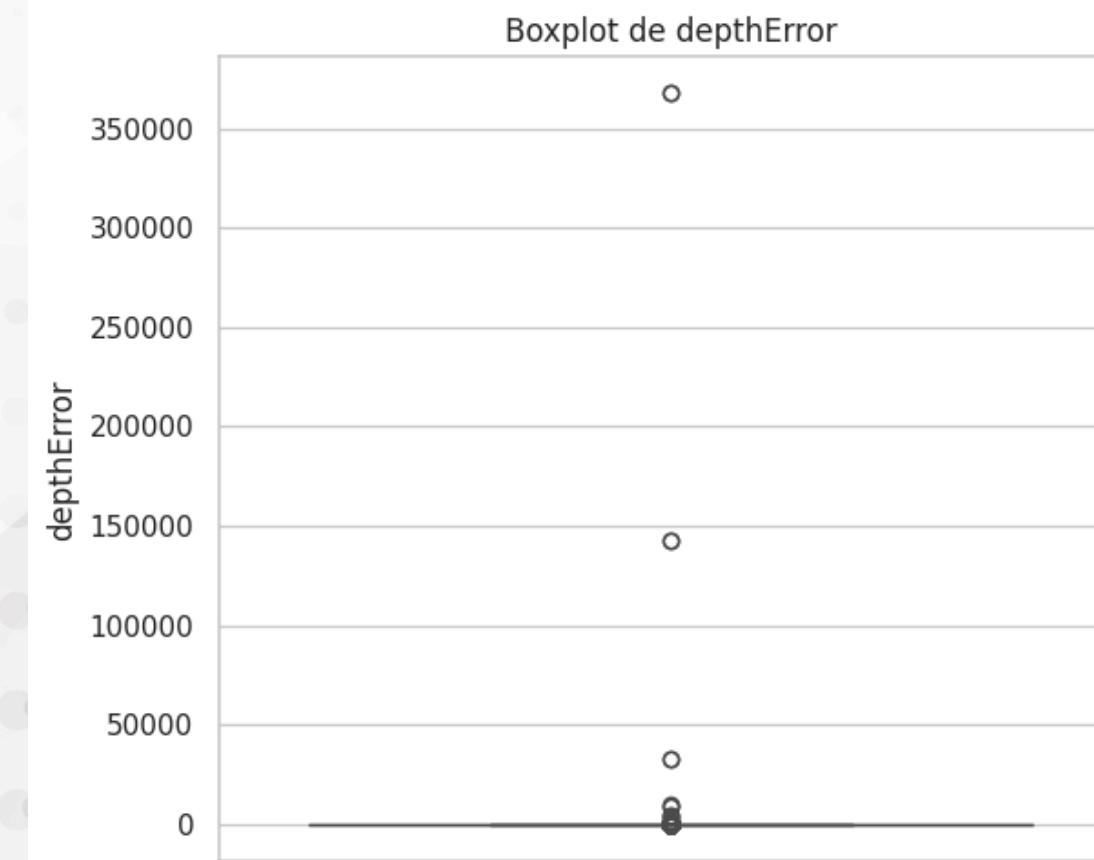
Problemas de estandarización textual:

La columna *place*, que indica la ubicación legible del sismo, presentaba formatos diversos, inconsistencias de mayúsculas/minúsculas, y espacios adicionales. Esto hacía difícil agrupar correctamente por regiones o países.



Presencia de valores extremos:

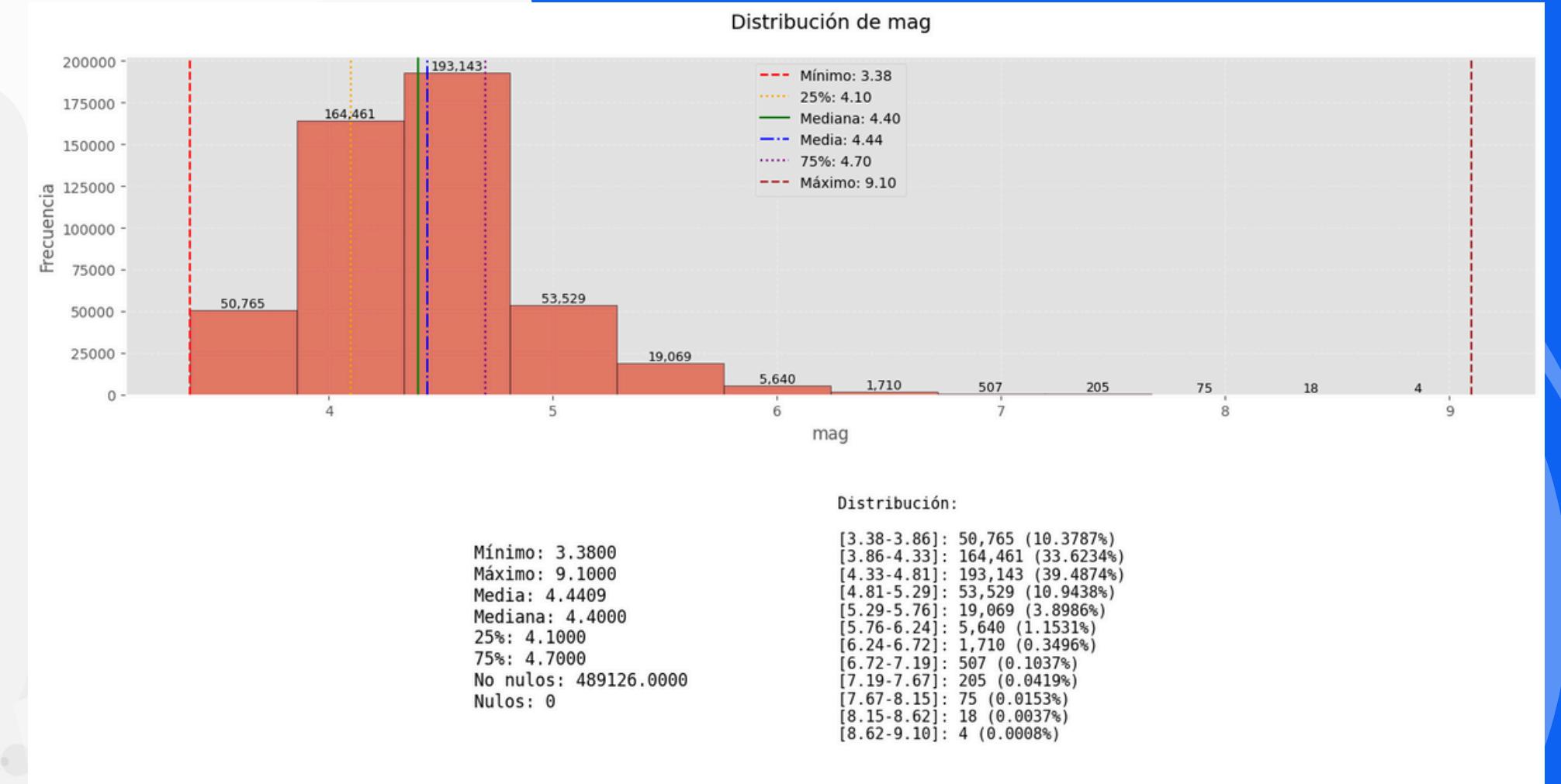
Algunas variables como `depthError` y `rms` contenían valores que superaban ampliamente los rangos esperados. Aunque no todos los valores extremos son errores, su presencia exige validación adicional.



¿Qué descubrieron al analizar los datos?

Grafico de la columna Magnitud

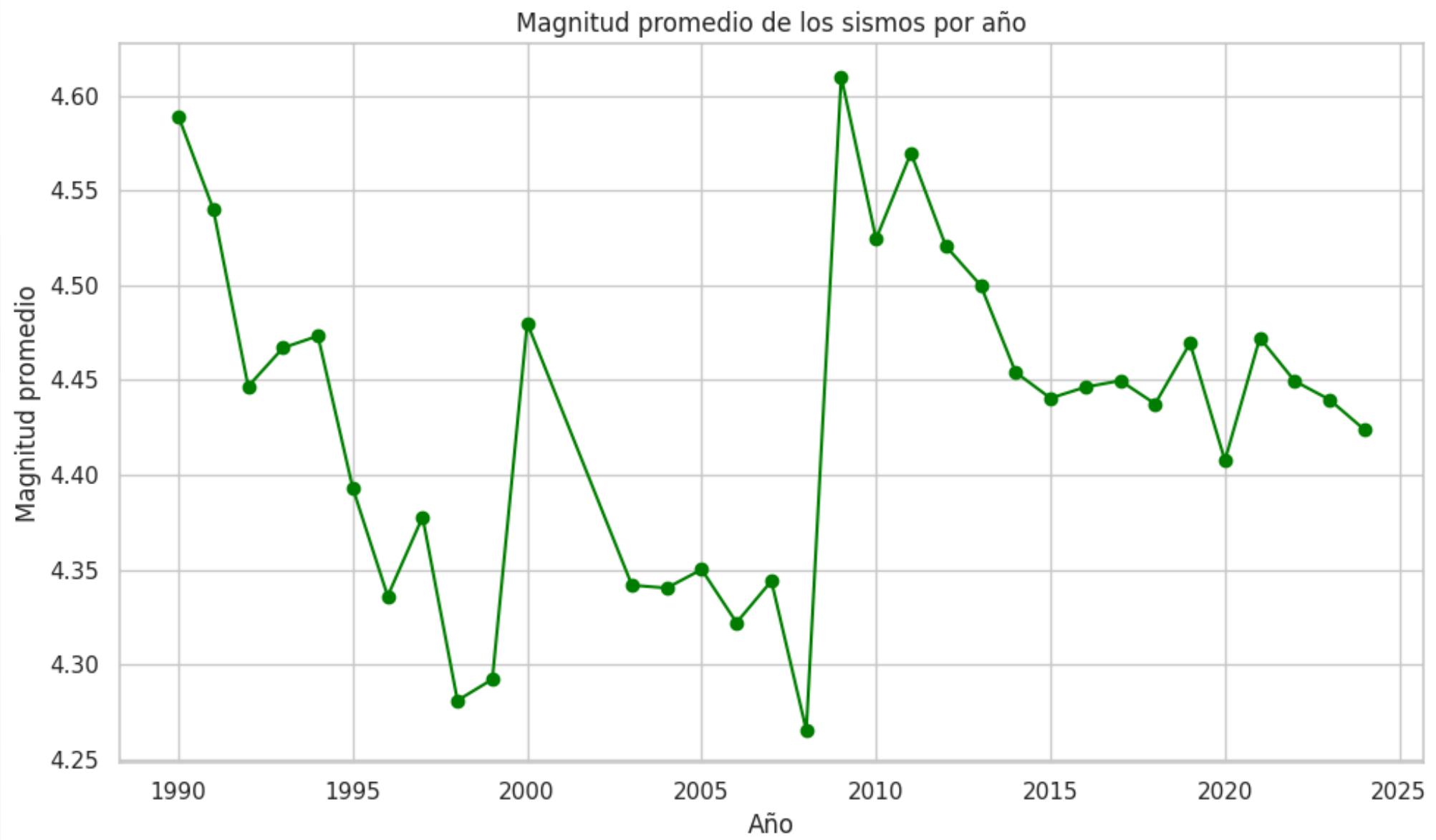
La mayoría de los eventos tienen magnitudes entre 4.1 y 4.7.



¿Qué descubrieron al analizar los datos?

Gráfico de líneas de magnitud a lo largo del tiempo:

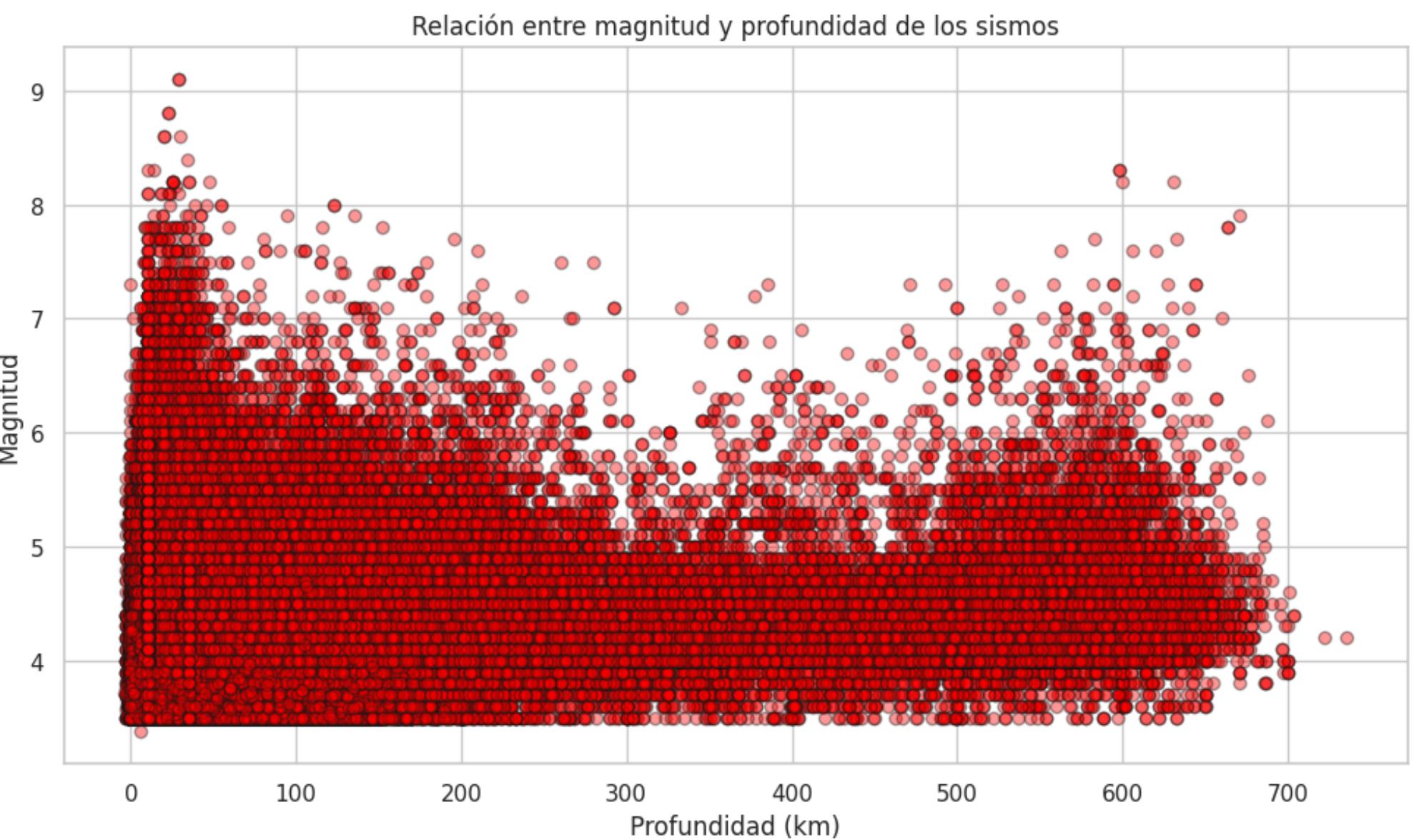
Esta visualización permitió observar cómo ha variado la magnitud de los eventos sísmicos entre los años 2022 y 2025. Se identifican períodos con mayor concentración de eventos de alta magnitud



¿Qué descubrieron al analizar los datos?

Gráfico de dispersión de magnitud vs. profundidad:

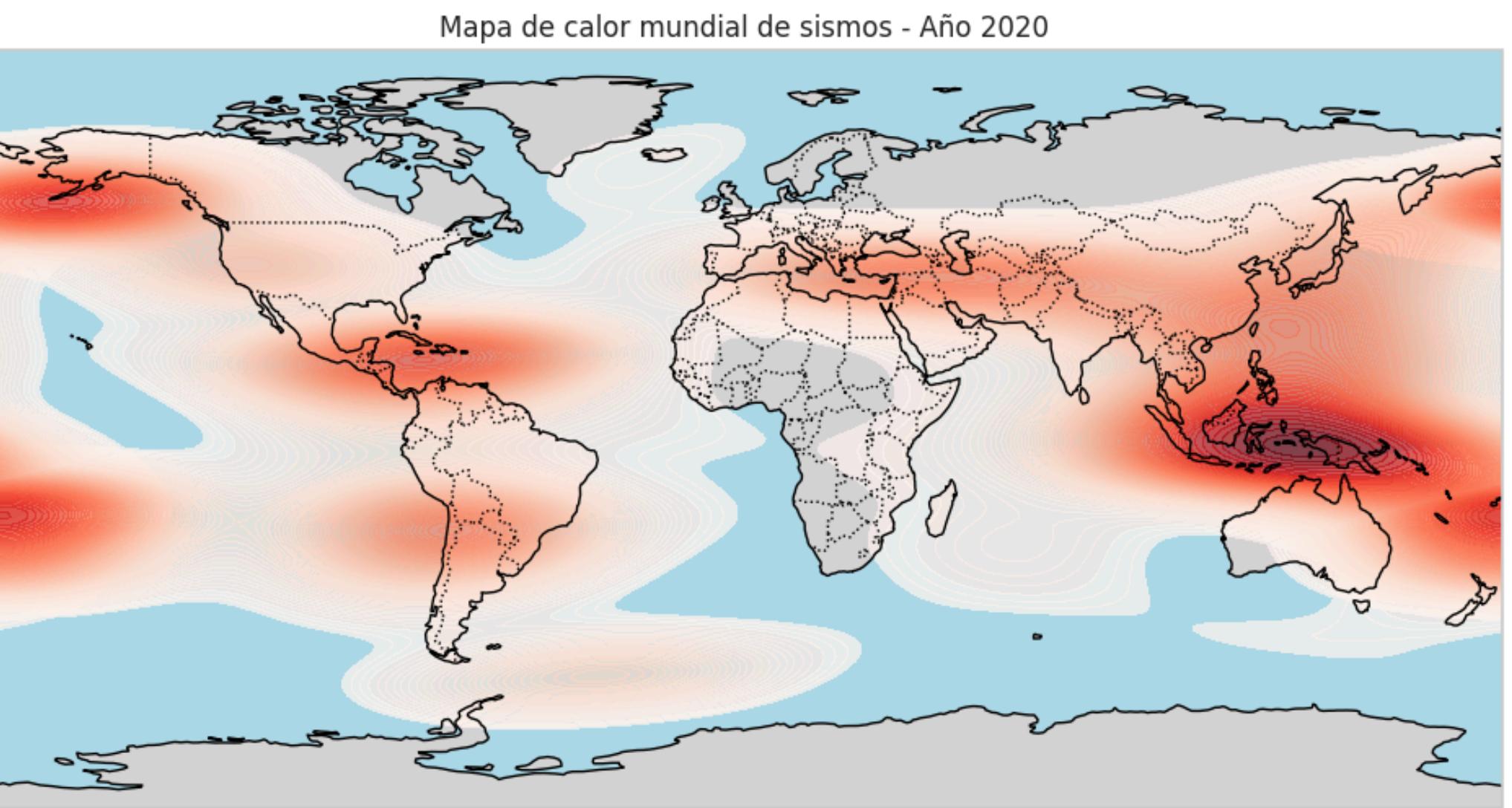
Este gráfico ayudó a visualizar la relación entre la magnitud y la profundidad de los eventos. Se observó que muchos sismos de alta magnitud ocurren a profundidades intermedias, mientras que los eventos más superficiales tienden a tener una magnitud moderada o baja.



¿Qué descubrieron al analizar los datos?

Mapa de calor de la ubicación de eventos:

Se generó una visualización de densidad geográfica utilizando las coordenadas de latitud y longitud. Esto permitió identificar regiones con alta frecuencia de actividad sísmica, revelando zonas geográficamente activas.



¿Qué reflejan los patrones de tendencia?

El análisis de los patrones de tendencia permitió identificar comportamientos recurrentes y relaciones importantes en los datos sísmicos.

Evolución anual del número de eventos sísmicos

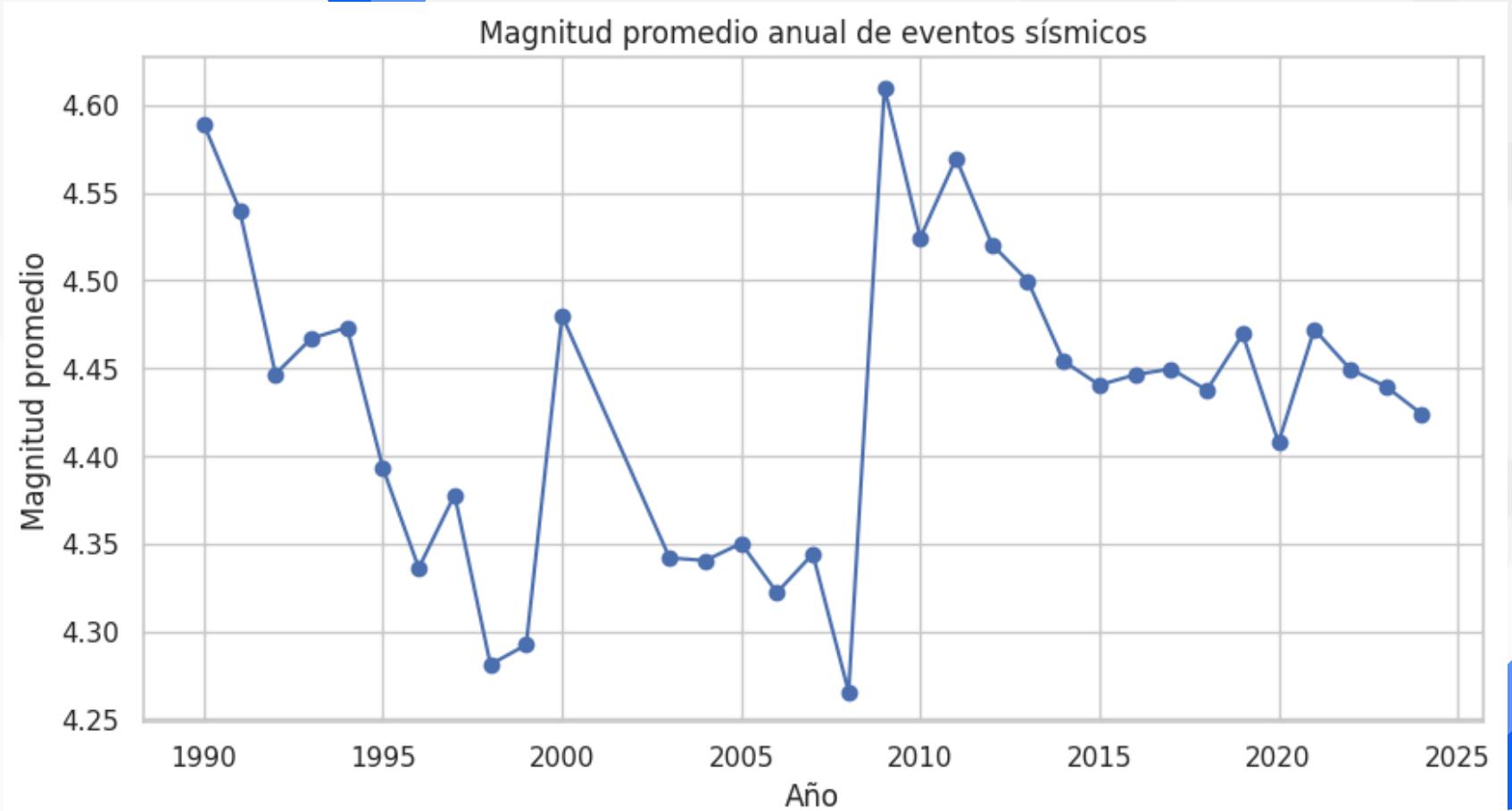
Se observa que la cantidad de eventos sísmicos ha aumentado en ciertos períodos, lo cual podría estar relacionado con un mayor monitoreo o actividad tectónica.



¿Qué reflejan los patrones de tendencia?

Tendencias en la magnitud promedio de los sismos

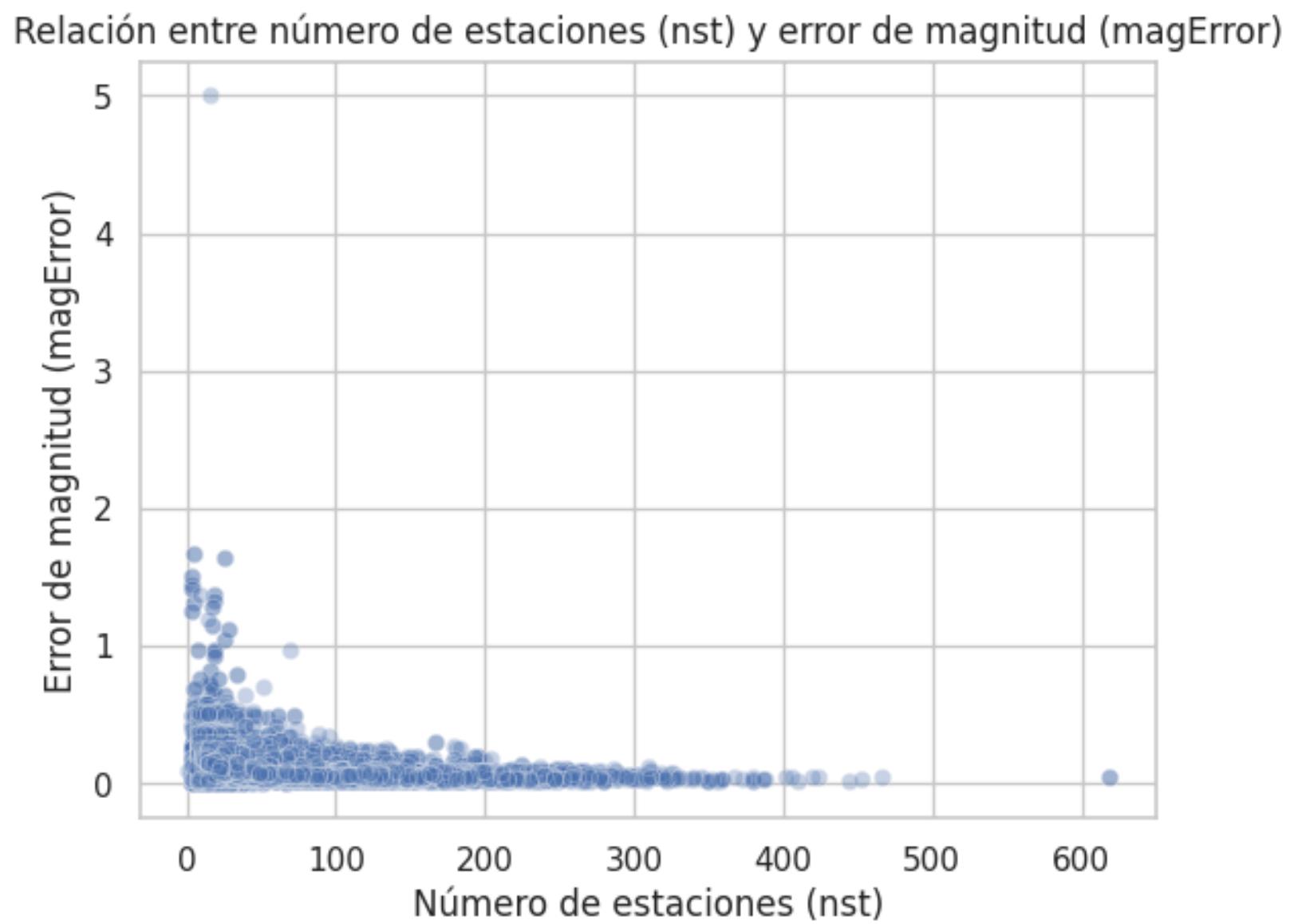
La magnitud promedio se mantiene relativamente constante, aunque se identifican años con valores ligeramente más altos, lo cual puede reflejar eventos significativos.



¿Existe relación entre el número de estaciones y el error en la magnitud?

A mayor número de estaciones que detectan un sismo nst, menor será el error en la estimación de su magnitud

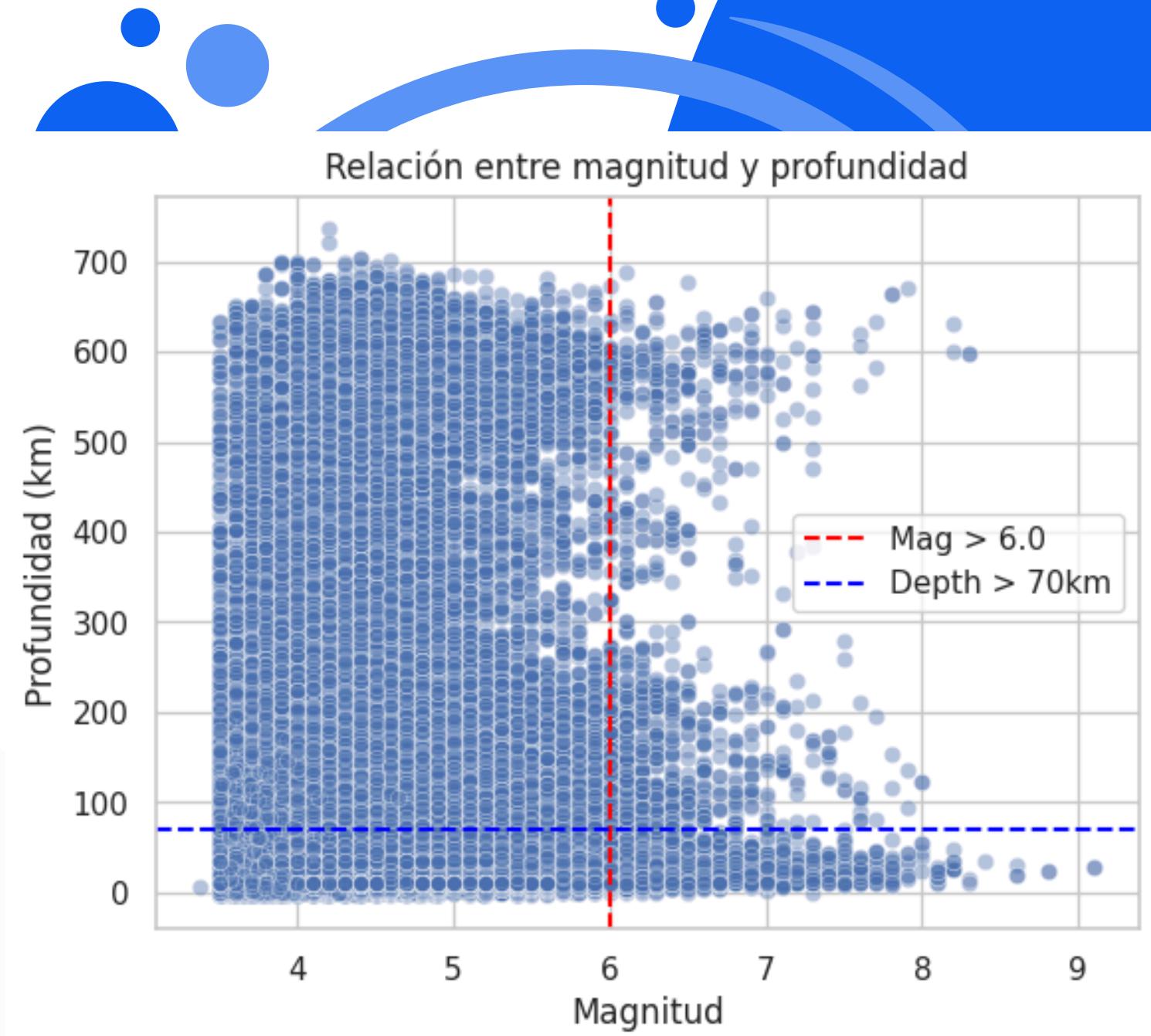
Esta hipótesis parte del supuesto de que contar con más estaciones sismológicas mejora la precisión del cálculo. Para evaluar esta relación, se elaboró un gráfico de dispersión donde se observa cómo varía el error de magnitud según el número de estaciones involucradas en la detección. Se espera una tendencia descendente si la hipótesis es válida.



¿Los sismos de mayor magnitud ocurren a mayor profundidad?

Los sismos con magnitudes superiores a 6.0 tienden a ocurrir a profundidades mayores a 70 km.

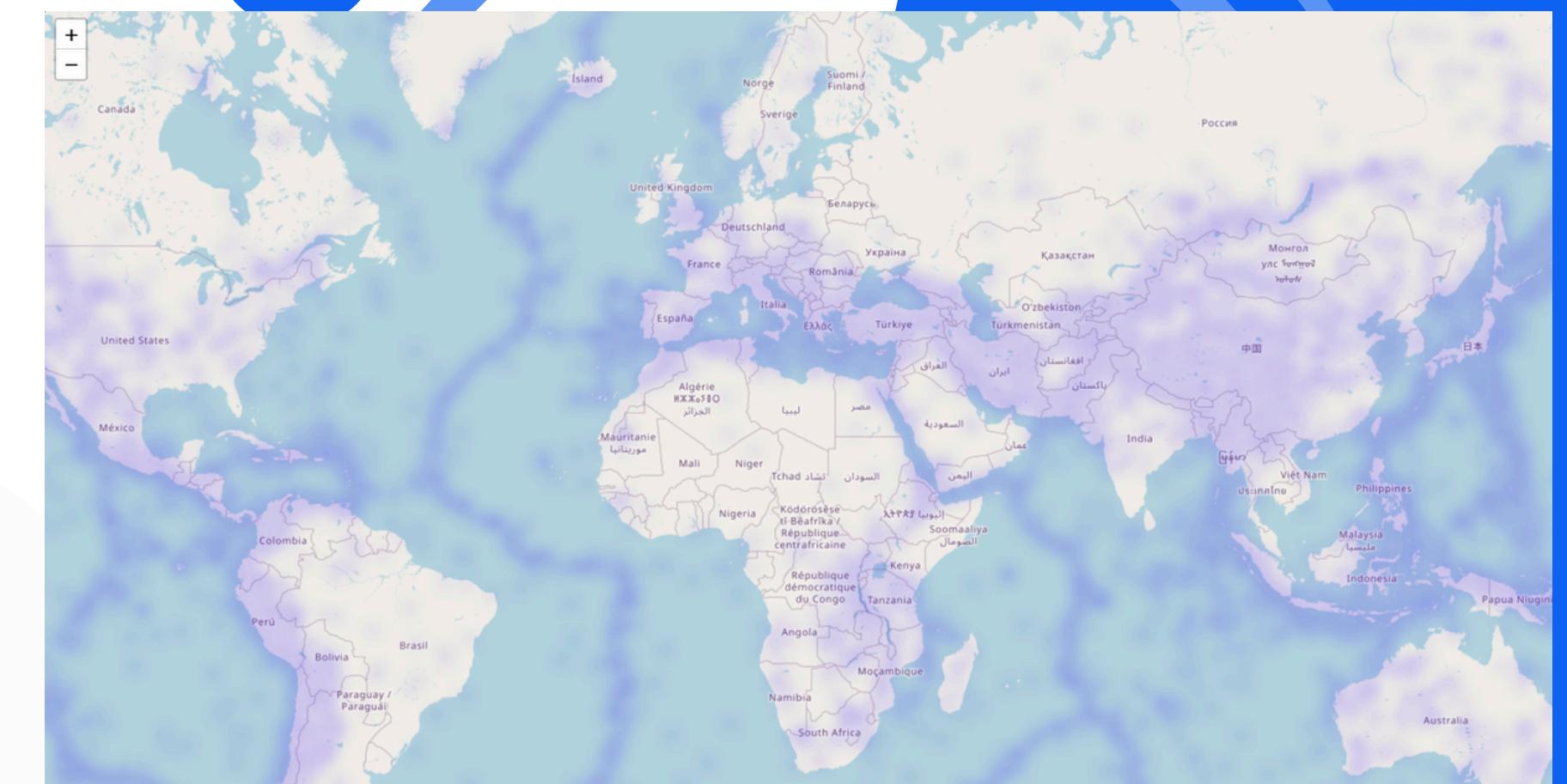
Esta hipótesis busca explorar la relación entre la energía liberada por un sismo y su ubicación en el subsuelo. Un gráfico de dispersión entre magnitud y profundidad permite observar si hay una agrupación de eventos fuertes ($\text{mag} > 6.0$) en zonas profundas ($\text{depth} > 70 \text{ km}$), lo cual podría indicar un patrón geológico relevante.



¿La mayoría de los eventos sísmicos ocurren en zonas del Cinturón de Fuego del Pacífico?

Existe una alta concentración de eventos sísmicos en regiones que forman parte del Cinturón de Fuego del Pacífico, como Japón, Fiji, Chile y Alaska.

El Cinturón de Fuego del Pacífico es conocido por su intensa actividad sísmica y volcánica. Para validar esta hipótesis, se generó un mapa de calor global basado en las coordenadas de los eventos registrados. Se espera observar concentraciones significativas en las zonas del cinturón, confirmando su actividad predominante.



GRACIAS