Respuestas **PRA1**

Jesica Piñón Rodríguez



Dataset: Últimos terremotos ocurridos en la Península Ibérica y Canarias

Contexto

Esta práctica se desarrolla como parte de la asignatura de "Tipología y ciclo de vida de los datos" del máster de Ciencia de Datos de la Universitat Oberta de Catalunya. La práctica consiste en aplicar técnicas de "web scrapping" para extraer información de alguna página web y generar un fichero con esos datos. Esta práctica se ha realizado usando el lenguaje de programación de Python.

La información del conjunto de datos se ha recogido de la web "https://www.ign.es/web/sis-area-sismicidad" (https://www.ign.es/web/sis-area-sismicidad" (https://www.ign.es/web/sis-area-sismicidad" (https://www.ign.es/web/ign/portal/ultimos-terremotos/) que contiene la información relativa a los fenómenos sísmicos de España.

Título

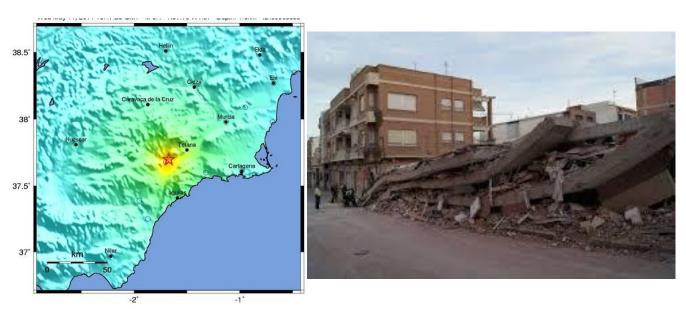
Como se indica en la parte superior, el título del conjunto de datos sería el siguiente: "Últimos terremotos ocurridos en el área de la Península Ibérica y Canarias"

Descripción

El conjunto de datos consta de una serie de campos con información sobre los últimos terremotos, de hace 10 días, 30 días o el último año, que han tenido lugar en la Península Ibérica y Canarias.

Algunas de las variables que se recogen en el conjunto de datos son la fecha, la hora, la localización y la magnitud del terremoto.

Representación gráfica



Contenido

El conjunto de datos que se obtiene contiene los siguientes campos:

- ID: Identificador de la columna.
- IDterremoto: Identificador del terremoto, código alfanumérico.
- Fecha %d%m%Y: Fecha en la que tuvo lugar el terremoto
- Hora inicio UTC
- Hora inicio local
- Latitud
- Longitud
- Profundidad (km)
- Magnitud
- Unidades: Unidades de medida del a magnitud del terremoto (tipo de magnitud)
- Intensidad
- Localización: Nombre del lugar donde ha tenido lugar el terremoto.

Agradecimientos

Los datos han sido recolectados desde la base de datos online del Instituto Geográfico Nacional. Para ello, se ha hecho uso del lenguaje de programación Python y de técnicas de Web Scraping para extraer la información alojada en las páginas HTML.

Inspiración

Considero que este conjunto de datos es útil para diversas áreas de ingeniería y medio ambiente. Analizando los terremotos ocurridos en la Península Ibérica y Canarias, se pueden llegar a aplicar técnicas de machine learning para poder predecir futuros terremotos (considerando otras fuentes de datos además del histórico de terremotos). A pesar de que las páginas utilizadas sólo permiten ver hasta el último año, lo ideal sería automatizar este proceso (ejecutar el script cada X días) para ir actualizando la información sin perder los datos históricos.

Entre otras cosas, además de sacar el csv, se podría generar un mapa automático para ayudar, de forma visual, a obtener la información de dónde y cuándo ha habido terremotos.

Licencia

La licencia escogida para la publicación de este conjunto de datos ha sido **CC BY-SA 4.0 License** con el fin de poder compartir la información sin ningún tipo de problema al reconocerse los autores iniciales (el origen de la fuente de datos).

Código fuente

El código se llama "main.py". Dentro del código, básicamente se define una función que permite utilizar diferentes periodos de tiempo. En la web se puede consultar la información para obtener los datos de los terremotos de los últimos 10 días, 30 días o el último año.

Al final del código se usa la función con los tres periodos. Esto es a modo de ejemplo para el ejercicio.

Dataset

Podrán encontrarse los tres ficheros correspondientes a los diferentes periodos de extracción de datos posible.