

# **Tipología y Ciclo de Vida de los Datos**

## **PRÁCTICA 1**

Eva García Ocaña  
Carmen Nieves Ojeda Guerra

## Índice

1. Contexto .....	3
2. Título.....	3
3. Descripción del <i>dataset</i> .....	3
4. Representación gráfica .....	3
5. Contenido .....	4
6. Agradecimientos .....	7
7. Inspiración .....	7
8. Licencia.....	8
9. Código .....	8
10. Dataset .....	8
Referencias .....	8
Contribución .....	9

## 1. Contexto

El contexto del trabajo se enmarca en la situación actual que está viviendo la isla canaria de La Palma con la erupción del volcán de Cumbre Vieja y la gran sismicidad medida en la zona. La información se ha recolectado desde la página web del [Instituto Geográfico Nacional](https://www.ign.es) o [IGN](https://www.ign.es), que se encarga de la planificación y gestión de sistemas de detección y comunicación a las instituciones de los movimientos sísmicos ocurridos en territorio nacional.

## 2. Título

Terremotos próximos sentidos de los días previos y posteriores a la erupción del volcán de Cumbre Vieja en la isla de La Palma, Islas Canarias, España.

## 3. Descripción del *dataset*

El conjunto de datos representa los terremotos en España de los 4 días previos y posteriores a la erupción del volcán de Cumbre Vieja, de intensidad entre II y IV (sentidos), magnitud en el rango entre 2 y 5 y una profundidad entre 0 y 30 km. La información está sujeta a modificaciones como consecuencia de la continua revisión del análisis sísmico.

## 4. Representación gráfica

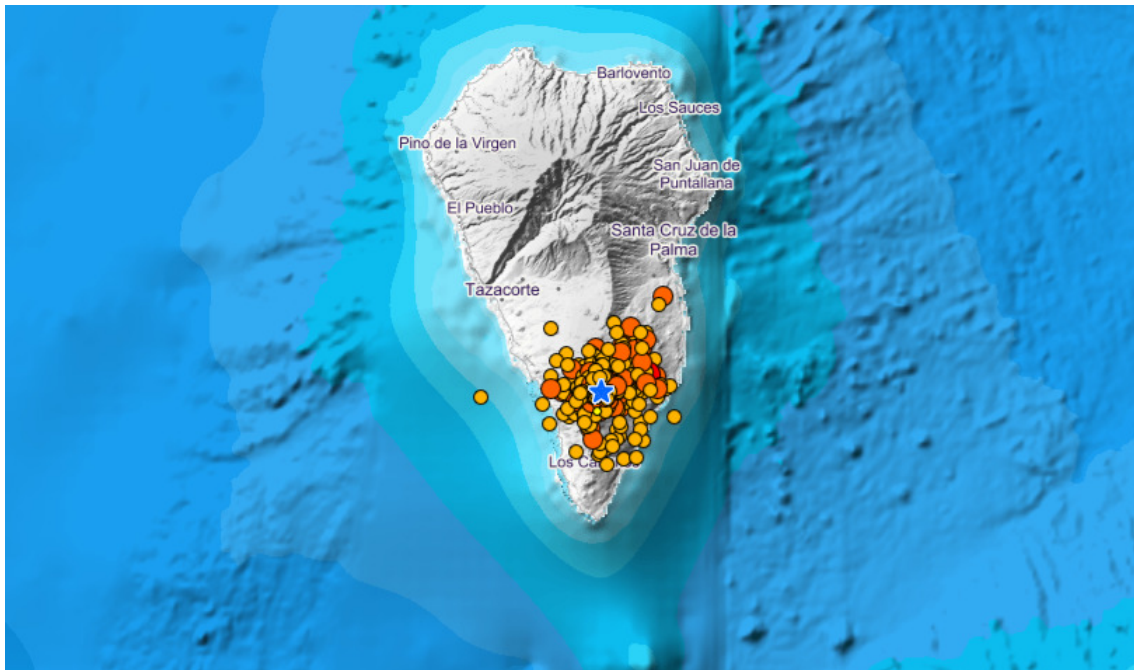


Figura 1: Sismos de los últimos 10 días (Fuente: IGN)

En la Figura 1 se puede apreciar los sismos de los últimos 10 días en la Isla de La Palma (Islas Canarias-España) donde se utilizan colores y tamaños de los círculos para indicar la magnitud e intensidad. Asimismo, el último terremoto producido se marca con una estrella azul.

Los terremotos se sitúan en la (*latitud, longitud*) en la que se produjeron. Esta información es la representación gráfica de la que se presenta en la tabla recuperada mediante *Web Scraping*.

## 5. Contenido

Previo al *Web Scraping* se analizó el archivo *robots.txt* de la web del IGN (<http://www.ign.es/web/robots.txt>), con el siguiente resultado:

```
User-Agent: *  
Disallow:
```

Con lo que se permite un acceso completo a todos los robots y ha sido posible utilizar este *dataset* para el trabajo pedido.

Asimismo, el sitio web del *Instituto Geográfico Nacional* no dispone de mapa del sitio, aunque es fácilmente navegable (Figura 2). Concretamente, en la página web <https://www.ign.es/web/ign/portal/sis-catalogo-terremotos>, está el formulario para filtrar la información general y obtener aquella que se ajusta a los criterios establecidos en la descripción del *dataset* (Figura 3).



Figura 2: Acceso al formulario para filtrar los datos (Fuente: IGN)

Figura 3: Formulario para selección de los datos de la descripción (Fuente: IGN)

A través de la página del formulario (Figura 3), se accede a la información relevante de los sismos producidos. Inspeccionando la estructura de la página web (Figura 4), se recupera la información que formará parte del *dataset* y que corresponde con los siguientes campos:

**Evento:** indica el nombre asignado al sismo por parte del IGN (*String*).

**Fecha:** fecha en la que se produjo el sismo (Fecha).

**Hora UTC:** hora UTC en la que se produjo el sismo (Hora).

**Hora Local:** hora local en la que se produjo el sismo (Hora).

**Latitud:** latitud geográfica en la que se produjo el sismo (Real).

**Longitud:** longitud geográfica en la que se produjo el sismo (Real).

**Profundidad:** profundidad en kilómetros en la que se produjo el sismo (Real).

**Magnitud:** magnitud del sismo (Real).

**Tipo:** tipo de magnitud (*String*).

**Intensidad:** intensidad máxima del sismo (*String*).

**Localizacion:** nombre de la ubicación geográfica del sismo (*String*).

Figura 4: Inspección de la página de análisis (Fuente: IGN)

Asimismo, también se ha analizado el tamaño del sitio, invocando a [site:www.ign.es](http://www.ign.es), obteniéndose alrededor 111.000 resultados. En el caso que nos ocupa, las páginas a las que se quiere acceder son unas concretas, que ya se indicaron anteriormente, por lo que la falta de eficiencia no es un punto a tener en cuenta por el tamaño del sitio.

- **Analytics and Tracking:** Google Analytics
- **ecommerce:** Cart Functionality
- **font-scripts:** Google Font API, Font Awesome
- **web-servers:** Apache, Apache Tomcat Coyote

[illegible]

Los datos recogidos por las estaciones pueden tener ciertas modificaciones debido a la continua revisión del análisis sísmico que se realizan sobre ellos. Poco tiempo después se convierten en datos históricos sin modificación posterior. En el *dataset* generado a partir del *web scraping*, los datos se almacenarán como *String*, que es como se recuperan, dejándose para un tratamiento posterior la conversión de cada campo a su tipo adecuado.



## 6. Agradecimientos

Los datos se han recolectado desde el sitio web del *Instituto Geográfico Nacional* que es el **propietario** de los datos. Las autoras de este trabajo agradecen al *IGN* la disponibilidad de su información, sin ninguna restricción, y se comprometen a un uso responsable de los mismos.

Así, según la especificación del trabajo, se solicita que el *dataset* generado se publicite a través de Zenodo. Al hacer este punto, las autoras de este trabajo utilizarán la licencia indicada por el *IGN* para sus productos. Asimismo, al hacer el análisis con el *scraper*, las autoras de este trabajo se comprometen a acceder, exclusivamente, a la información que se indica en este documento, además de hacer público el código del *scraper* a través del enlace a *github* que se indica en otro apartado de esta memoria.

Respecto a análisis previos de estos datos, las autoras de este trabajo no tienen constancia de que existan análisis similares sobre estos datos en concreto, sin embargo, se han encontrado varios trabajos sobre el mismo tema, con otros *dataset* sobre terremotos y volcanes. Por ejemplo, en la biblioteca **Kaggle** tenemos los *dataset* indicados en los enlaces presentados<sup>1,2,3</sup>.

Por otro lado, hay trabajos que analizan la clasificación de las ciudades más y menos propensas a tener terremotos<sup>4</sup> para ayudar a las personas a decidir dónde vivir. En otros casos, se analizan y visualizan los parámetros sísmicos y las correlaciones entre ellos para la detección de patrones<sup>5,6</sup> y ayudar a la detección temprana de sismos que puedan ser sentidos. En “*Data mining and volcanic eruption forecasting*”<sup>7</sup> se presenta una tesis doctoral sobre la predicción de erupciones volcánicas en relación con patrones precursores de sismicidad o deformación.

Como puede observarse, existe preocupación por estos temas ya que, las zonas sísmicas más activas del planeta, coinciden también con zonas ampliamente pobladas y el uso de la inteligencia artificial puede ayudar a evitar pérdidas humanas.

## 7. Inspiración

En el caso que nos ocupa, el *dataset* elaborado puede utilizarse a título informativo, para saber, por ejemplo, los lugares en los que se han producido los últimos sismos y, en general, las características de estos.

---

<sup>1</sup> <https://www.kaggle.com/c/predict-volcanic-eruptions-ingv-oe>

<sup>2</sup> <https://www.kaggle.com/jessemotipak/volcano-eruptions>

<sup>3</sup> <https://www.kaggle.com/c/LANL-Earthquake-Prediction>

<sup>4</sup> <https://towardsdatascience.com/earthquake-popular-venues-data-analysis-of-turkey-28f8d02cb0d>

<sup>5</sup> <https://www.frontiersin.org/articles/10.3389/fphy.2021.656310/full>

<sup>6</sup> [https://rpubs.com/ykashou92/eq\\_wmap](https://rpubs.com/ykashou92/eq_wmap)

<sup>7</sup> [https://www.researchgate.net/publication/280318797 Data mining and volcanic eruption forecasting](https://www.researchgate.net/publication/280318797_Data_mining_and_volcanic_eruption_forecasting)

Asimismo, se puede utilizar para elaborar modelos que ayuden a predecir dónde se producirá el próximo terremoto, a qué posible profundidad y de qué magnitud. Asimismo, sería interesante combinarlo con datos referidos a la actividad volcánica y ver qué relación existen entre los sismos y los volcanes.

Como se puede observar, la necesidad del análisis del *dataset* es similar a los realizados por otros autores.

## 8. Licencia

La licencia elegida para el conjunto de datos es CC BY 4.0, que permite compartir y modificar los datos como se quiera mientras se reconozca la autoría de los mismos. Se ha elegido esta licencia puesto que es la [licencia](#) con la que el *Instituto Geográfico Nacional* comparte sus datos. Con esta licencia se permite:

- Compartir, copiar y redistribuir el material en cualquier medio o formato.
- Adaptar, remezclar, transformar y construir sobre el material para cualquier propósito, incluso comercial.

Sin embargo, analizando la licencia y los productos a los cuales se puede acceder bajo la misma, no aparecen los datos sismológicos sino la información geográfica (cartografía en general). Aun así, se aplicará la misma licencia al producto usado por coherencia.

## 9. Código

El código puede ser accesible a través del enlace público de **github** siguiente: [https://github.com/cnojedaUOC/PRA1\\_TCVD.git](https://github.com/cnojedaUOC/PRA1_TCVD.git).

## 10. Dataset

El *dataset* puede ser accesible a través del siguiente DOI de Zenodo: **10.5281/zenodo.5643138**.


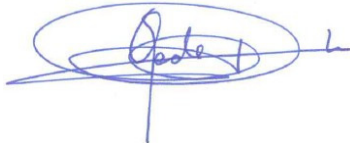


## Referencias

- [1] L. Subirats, M. Calvo, “Web Scraping”, Editorial UOC, 2019.
- [2] R. Lawson, “Web Scraping with Python”, Packt Publishing Ltd. Chapter 2. Scraping the Data, 2015.
- [3] V. Alcántara, “Cómo codificar un scraping Bot con Selenium y Python”, 2021. [En línea]. Disponible en: <https://www.freecodecamp.org/espanol/news/como-codificar-un-scraping-bot-con-selenium-y-python/>. [Accedido: 01-nov-2021]



[4] Sitio web de Selenium, "The Selenium Browser Automation Project", 2021. [En línea]. Disponible en: <https://www.selenium.dev/documentation/>. [Accedido: 01-nov-2021]

## Contribución

Contribuciones	Firma	
Investigación previa	Eva García y Carmen N. Ojeda	
		
Desarrollo del código y generación del archivo csv	Eva García y Carmen N. Ojeda	
		
Generación de la documentación escrita y visual	Eva García y Carmen N. Ojeda	
	