

INFORME PROYECTO FINAL

Conjunto de datos:

Los datos están recopilados de Kaggle

<https://www.kaggle.com/datasets/noaa/seismic-waves>

Hay dos dataset: sources y waves.

sources: Se refiere a los eventos que desencadenan la generación de tsunamis, como terremotos, deslizamientos de tierra submarinos, erupciones volcánicas, impactos de meteoritos u otras perturbaciones bajo el agua que causan un movimiento significativo del lecho marino.

waves: Son las ondas de agua masivas que se propagan a través del océano como resultado de un evento de tsunami. Estas olas pueden viajar grandes distancias desde su fuente original y pueden tener un impacto devastador en las áreas costeras cuando llegan a tierra firme.

Dado que "sources" se refiere a los eventos que generan tsunamis, y "waves" hace referencia a las olas de tsunami resultantes he decidido que trabajaré con el df de "sources" ya que para mi proyecto solo necesitaré ese conjunto de datos.

Características generales:

EL autor de este dataset recopiló sus datos de NOAA/WDS (National Oceanic and Atmospheric Administration) Administración Nacional Oceánica y Atmosférica.

Tal y como se describe en la página, la base de datos de tsunamis de la NOAA/WDS es un listado de eventos históricos de origen de tsunamis y ubicaciones de llegada en todo el mundo desde el 2000 a.C. hasta la actualidad.

Los Centros Nacionales de Información Ambiental (NCEI) de la NOAA y el Servicio Mundial de Datos para Geofísica recopilaron y publicaron esta base de datos de tsunamis para los centros de alerta de tsunamis, ingenieros, oceanógrafos, sismólogos y el público en general

Definición de las variables:

El dataset inicial contiene un total de 45 columnas y 2582 filas. Sin embargo después de revisar los datos y teniendo en cuenta el objetivo de mi proyecto utilizaré las siguientes columnas:

- SOURCE_ID: identificador único para cada registro

- YEAR: año en el que ocurrió el tsunami
- MONTH: mes del evento
- DAY: día del tsunami
- CAUSE: Código de Causa del Tsunami. Valores válidos: 0 a 11. La fuente del tsunami:
 - 0: Desconocido
 - 1: Terremoto
 - 2: Terremoto cuestionable
 - 3: Terremoto y deslizamiento de tierra
 - 4: Volcán y terremoto
 - 5: Volcán, terremoto y deslizamiento de tierra
 - 6: Volcán
 - 7: Volcán y deslizamiento de tierra
 - 8: Deslizamiento de tierra
 - 9: Meteorológico
 - 10: Explosión
 - 11: Marea astronómica
- VALIDITY: Valores válidos: -1 a 4. La validez del tsunami real se indica mediante una calificación numérica de los informes de ese evento:
 - -1: Entrada errónea
 - 0: Evento que solo causó un seiche o perturbación en un río o lago interior
 - 1: Tsunami muy dudoso
 - 2: Tsunami cuestionable
 - 3: Tsunami probable
 - 4: Tsunami definitivo
- REGION_CODE: Regional boundaries defined as follows:
 - 87 - Alaska (including Aleutian Islands)
 - 40 - Black Sea and Caspian Sea
 - 74 - Caribbean Sea
 - 78 - Central Africa
 - 84 - China, North and South Korea, Philippines, Taiwan
 - 81 - E Coast Australia, New Zealand, South Pacific Is.
 - 75 - East Coast USA and Canada, St Pierre and Miquelon
 - 76 - Gulf of Mexico
 - 80 - Hawaii, Johnston Atoll, Midway I
 - 60 - Indian Ocean (including west coast of Australia)
 - 83 - Indonesia (Pacific Ocean) and Malaysia
 - 85 - Japan
 - 86 - Kamchatka and Kuril Islands
 - 50 - Mediterranean Sea
 - 82 - New Caledonia, New Guinea, Solomon Is., Vanuatu
 - 73 - Northeast Atlantic Ocean
 - 72 - Northwest Atlantic Ocean
 - 30 - Red Sea and Persian Gulf
 - 70 - Southeast Atlantic Ocean
 - 71 - Southwest Atlantic Ocean
 - 77 - West Coast of Africa

- 88 - West Coast of North and Central America
- 89 - West Coast of South America
- COUNTRY: El país donde ocurrió la fuente del tsunami
- LOCATION: El país, estado, provincia o isla donde ocurrió la fuente del tsunami (por ejemplo, ingresa: Japón o Honshu). Esta es solo una ubicación geográfica aproximada. Los eventos anteriores a 1900 no fueron ubicados instrumentalmente, por lo tanto, la ubicación proporcionada se basa en la latitud y longitud de la ciudad donde ocurrieron los efectos máximos. Si hay diferentes formas de escribir el nombre de una ciudad, los nombres adicionales están entre paréntesis.

.Objetivo:

El objetivo general del proyecto es crear un modelo que permita predecir el número de tsunamis por año