

Dataset: Información de las estaciones meteorológicas de España.

Contiene información de temperatura, precipitación, viento y permite descargar mapas de temperatura del agua del mar.

Imagen descriptiva



Agencia Estatal de Meteorología

Contexto

La materia del conjunto de datos son las observaciones meteorológicas de todas las estaciones de medida que tiene AEMET repartidas por España. Es importante señalar que el dataset contiene las observaciones y no las predicciones meteorológicas que se puedan hacer con dichas observaciones. Esto implica que el código que se proporciona para descargar los datos deberá correrse al día siguiente de los datos que se deseen obtener. Hay que señalar que AEMET publica en su página web las observaciones de los 7 días previos al de consulta. Adicionalmente, el código permite descargar el mapa de la temperatura del agua del mar alrededor de la península, Baleares y Canarias. Este mapa no forma parte del Dataset, pero complementa la información de las estaciones.

Contenido

Hay dos carpetas con datos. Por un lado, la carpeta imágenes contiene los mapas de temperatura del agua del mar. Hay una imagen por cada día que se ha utilizado el programa.

Por otro lado, en la carpeta csv se encuentra el conjunto de datos en formato csv. Hay un archivo csv por cada día que se ha realizado la descarga. Los campos que contienen todos los csv son los mismos, aunque para algunos campos AEMET a veces introduce la forma completa y en otras la reducida. Los campos son los siguientes:

- **Estación/Est.:** Nombre de la estación meteorológica (generalmente coincide con municipio)
- **Provincia/Prov.:** Provincia en la que se encuentra la estación meteorológica
- **Temperatura máxima (°C)/T.max.** Temperatura máxima del día consultado en la estación

- **Temperatura mínima (°C)/T.min.** Temperatura mínima del día consultado en la estación
- **Temperatura media (°C)/T.media** Temperatura media del día consultado en la estación
- **Racha (km/h)/Racha** Velocidad de la racha de viento más fuerte del día consultado en la estación
- **Velocidad máxima (km/h)/V.max.** Velocidad sostenida máxima del viento en la estación en el día consultado.
- **Precipitación 00-24h (mm)/Pr. 00-24h** Precipitación total en la estación en el día consultado
- **Precipitación 00-06h (mm)/ Pr. 00-06h** Precipitación entre las 0 y las 6 en la estación
- **Precipitación 06-12h (mm)/ Pr. 06-12h** Precipitación entre las 6 y las 12 en la estación
- **Precipitación 12-18h (mm)/ Pr. 12-18h** Precipitación entre las 12 y las 18 en la estación
- **Precipitación 18-24h (mm)/ Pr. 18-24h** Precipitación entre las 18 y las 24 en la estación
- **Fecha** Día en los que las observaciones se registraron
- **Temperatura máxima (°C)(hora)/T.max.(hora)** Hora de la temperatura máxima
- **Temperatura mínima (°C) (hora)/T.min.(hora)** Hora de la temperatura mínima
- **Racha (km/h) (hora)/Racha(hora)** Hora de la racha de viento
- **Velocidad máxima (km/h)(hora)/V.max(hora)** Hora de la velocidad máxima de viento

La página web de AEMET permite consultar sin cargo las observaciones meteorológicas de los siete días previos a la fecha de consulta.

Agradecimientos

Los datos provienen de la página web de la Agencia Estatal de Meteorología (AEMET). Para extraerlos, se ha realizado un código que hace *Web Scraping* en su web. El lenguaje de programación utilizado es python, utilizando las librerías pandas, numpy, urllib y bs4.

Inspiración

El conjunto de datos puede ser de gran utilidad en una amplia variedad de campos, entre otros motivos porque, si se extraen datos de forma diaria durante varios meses, se tendría el dataset meteorológico gratuito con mayor granularidad del país. Varios ejemplos de aplicación en el campo del *Big Data* podrían ser los siguientes:

- Al tener un alto detalle geográfico (hay una estación por municipio), podría utilizarse para crear un modelo que permita determinar la posible generación renovable para un determinado día. De este modo, si una empresa de generación de energía renovable utilizase este modelo podría ajustar su oferta en el mercado diario para asegurar que siempre vende su energía sin comprometer necesariamente sus ingresos.

- Del mismo modo que antes, al tener un alto detalle geográfico, si se cruzan los datos con la demanda eléctrica o la demanda de gas natural se podría obtener un modelo que, en función de la localidad y el tiempo previsto, pudiera prever el consumo eléctrico o de gas.
- Cruzando este dataset con los precios medios diarios del mercado eléctrico español, se podría generar un modelo que permita predecir, en función de las predicciones meteorológicas para cada municipio, el precio medio de electricidad en el mercado diario. Esto podría tener aplicaciones para la compraventa a corto plazo de energía en el mercado, por ejemplo.

Licencia

Se ha elegido publicar el conjunto de datos y el código bajo la licencia CC BY-SA 4.0 License. Los motivos por los que se ha elegido esta licencia son los siguientes:

- Con esta licencia, se debe reconocer la autoría del dataset y señalar las modificaciones que se vayan realizado.
- Se mantiene la misma licencia en las posibles adaptaciones o contribuciones que se puedan hacer sobre el trabajo original
- Se permite un uso comercial, lo que puede ayudar a la difusión en el ámbito empresarial y reconocer el trabajo original y su autor.

Código y archivos

Tanto el código como los archivos que descarga pueden encontrarse en el siguiente link:

<https://github.com/dcastejonm/Practica1-Web-Scraping>

El código se encuentra en dos archivos:

- **main.py**: Corre el código principal. Es el archivo a correr para realizar la descarga de los datos.
- **functions.py**: Define todas las funciones que se usan en el archivo main.py.

Para ejecutar, se corre el archivo main.py. Posteriormente, en la consola se preguntará al usuario cuantos días quiere descargar y qué método prefiere usar para realizar la descarga de los datos:

1. La opción 1 busca el botón de los csv en la web de AEMET y descarga la información desde dicho link. Luego agrupa todos los datos en un único archivo y realiza una limpieza preliminar de los datos.
2. La opción 2 busca en el html de la página web toda la información necesaria. Agrupa toda la infomación en un único archivo y realiza una limpieza preliminar de los datos.

Finalmente, preguntará al usuario si desea descargar el mapa de temperatura del agua del mar

Recursos

- Lawson, R. (2015). Web Scraping with Python. Packt Publishing Ltd. Chapter 2. Scraping the Data.
- Subirats, L., Calvo, M. (2018). Web Scraping. Editorial UOC