

# Práctica 1: Web Scraping

Roberto Díaz Badra

## Título del dataset

El tiempo semanal en España.

## Subtítulo del dataset

Descripción de los factores temporales de los municipios de España a lo largo de una semana.

## Imagen Identificativa



## Contexto

Este conjunto de datos. Proporciona información obtenida de la página de la Aemet, e incorpora datos temporales, de factores como humedad y temperatura, para cada municipio de España a lo largo de una semana. Los datos no entran en gran detalle, ya que la Aemet solo proporciona datos detallados para los primeros 2 días de la semana, correspondientes al día actual y al siguiente.

## Contenido

Para cada municipio, se obtienen varios registros con los siguientes campos:

- **municipio:** El nombre del municipio de España.
- **provincia:** EL nombre de la provincia a la que pertenece el municipio.
- **fecha:** El día al que corresponde toda la información del municipio.

- **horario:** El intervalo de tiempo en el día con el que corresponden los datos, cuando aplica, ya que cuanto más se avanza en los días de la semana, estos campos dejan de detallarse.
- **temperatura\_actual:** La temperatura esperada en el horario especificado. Igual que con el **horario**, a medida que avanzamos en los días, estos datos dejan de detallarse.
- **probabilidad\_precipitacion:** Probabilidad de que llueva en un determinado municipio en un día y horario(cuando aplica) especificado.
- **temperatura\_minima:** Temperatura mínima esperada para ese día.
- **temperatura\_maxima:** Temperatura máxima esperada para ese día. Es necesario comentar, que para la creación de este conjunto de datos, tiene gran relevancia la hora del día a la que se toma, ya que la página de la AEMET cambia dependiendo de la hora. Así pues, este dataset fue obtenido en el intervalo de 18 - 24 horas, por lo que para evitar la existencia de errores, se recomienda ejecutar el código en este periodo de tiempo.

## Agradecimientos

Este dataset ha sido obtenido a través de la página de la AEMET, no obstante, posibilita la reutilización de la información con fines comerciales o no comerciales, como queda establecido en la Ley 18/2015.

## Inspiración

Este conjunto de datos es muy interesante, ya que tiene gran número de posibilidades. Obviamente se le podrían realizar técnicas de Aprendizaje Automático, para poder predecir las futuras condiciones climatológicas de una región.

Otra posibilidad, consiste en la utilización de técnicas de clustering, para observar qué municipios o regiones tienen climas similares o completamente dispares.

## Licencia

Para la publicación de este conjunto de datos, se ha optado por utilizar la licencia **CC0: Public Domain License**, ya que permite el máximo de grado de libertad a la hora de utilizarlo. Al utilizarlo, no solo no hace falta que los usuarios indiquen de dónde han sacado la información, sino que nos eximen de cualquier tipo de responsabilidad, por cualquier acto malicioso(no se me ocurre qué tipo de acción perjudicial se podría realizar con estos datos) que se pudiera llevar a cabo.

## Código y Dataset

Tanto el código como el conjunto de datos en formato csv, se proporcionan en el enlace github adjuntado a continuación.

## Enlace github

<https://github.com/rdbadra/WebScraping>