

PRÁCTICA 1

TIPOLOGÍA Y CICLO DE VIDA DE LOS DATOS

Marta Rodríguez y Pedro Féliz

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

Las emisiones, durante el periodo de confinamiento, causado por el COVID-19 se han reducido drásticamente. Esta reducción se ha apreciado tanto en los satélites, como en los datos de contaminación que se recopilan en las grandes ciudades.

Se han realizado distintos estudios sobre la reducción de los niveles de contaminación. Uno de ellos, es el que ha llevado a cabo el periódico El País, que tomando los datos de la Agencia Europea del Medio Ambiente (AEMA), ha analizado la reducción de dióxido de nitrógeno, como se puede ver en el siguiente artículo:

<https://elpais.com/sociedad/2020-04-11/el-confinamiento-reduce-a-la-mitad-la-contaminacion-en-las-80-ciudades-mas-pobladas-de-espana.html>

Siendo conscientes de la reducción de la contaminación, queremos comparar el posible cambio meteorológico de este año con los anteriores. Con el objetivo de averiguar si ha habido un impacto medioambiental debido a la reducción de las emisiones causada por la pandemia.

Para ello, hemos considerado que la información sobre temperaturas y precipitaciones, nos podía proporcionar información suficiente sobre el posible cambio climatológico. A la hora de recopilar información, tras estudiar distintos sitios web, consideramos que Aemet nos podía proporcionar información fiable y suficiente para nuestro estudio. Ya que dispone distintas estaciones meteorológicas repartidas por toda España, que recopilan información meteorológica y posteriormente Aemet pone a disposición pública a través de su página web.

2. Definir un título para el dataset. Elegir un título que sea descriptivo.

El título para el dataset es; Temperaturas y precipitación mensuales en España desde 2015.

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título elegido).

La información recopilada, contiene datos meteorológicos mensuales de temperaturas medias, y precipitaciones desde el año 2015 para cada una de las estaciones meteorológicas de Aemet.

Únicamente se han seleccionado datos desde el 2015, ya que nuestro estudio pretende captar variaciones en el clima en un periodo muy concreto. Por lo tanto, si se hubiera

seleccionado un periodo más amplio, el cambio meteorológico debido al cambio climático interferiría en nuestro estudio.

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente

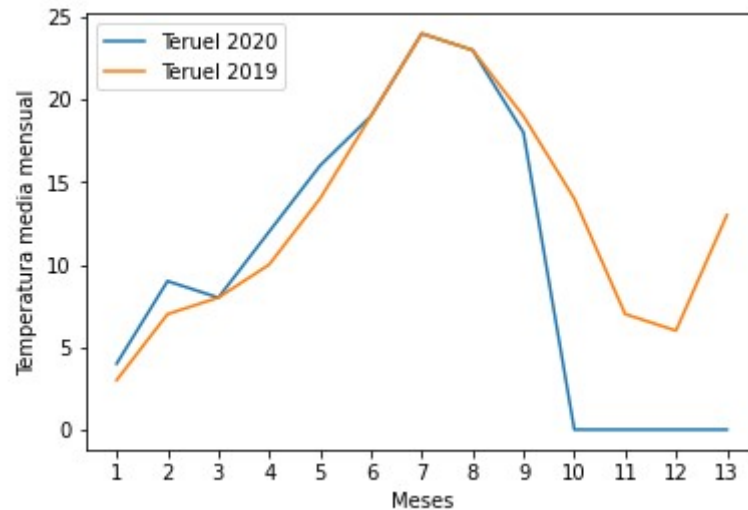
Por cada una de las estaciones meteorológica, se ha recogido la información de Ciudad, año-mes, temperatura media mensual, temperatura media de las máximas, de las mínimas, número de días de lluvia, precipitación total. De forma que para cada uno de los meses tendríamos la información que se puede ver en la siguiente imagen.



Para tener una visión general del uso del dataset, hemos realizado un ejemplo utilizando la temperatura media mensual en Teruel en los años 2019 y 2020.

Como se puede observar, los valores de la temperatura media para los meses de octubre, noviembre y diciembre de 2020 cae a 0, ya que no hay valores registrados en estos casos.

Se observa durante los meses de confinamiento domiciliario, la temperatura media en 2020 fue superior a la registrada en el año 2019. Al suprimir este confinamiento, los meses de verano (junio, julio y agosto) las temperaturas se mantuvieron idénticas al año anterior. En el mes de septiembre, que fue el mes de la vuelta a la normalidad, se aprecia una disminución en la temperatura media comparado con el año anterior.



5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.

El dataset contiene información mensual del tiempo desde el año 2015 en distintos puntos de España.

Para recopilar la información se ha realizado web scraping a la página de Aemet, con el objetivo de obtener las distintas estaciones meteorológicas de España. Una vez se tiene la información de las distintas estaciones meteorológicas, con la API de Aemet y una API Key (que hemos tenido que solicitar por correo electrónico a Aemet) se realizan las distintas consultas. De este modo, obtenemos los datos meteorológicos por fecha y ubicación.

De toda la información proporciona Aemet, hemos realizado una selección sobre los atributos más relevantes, como son las distintas temperaturas medias, la precipitación total y el número de días que ha llovido.

El dataset generado consta de siete columnas que se describen a continuación:

- ciudad: Ciudad en la que se han recogido los datos de meteorología.
- anio: Variable con el formato YYYY-MM que indica el mes al que corresponde la información. Donde los meses del 1 al 12 corresponden con la información mensual, mientras que el mes 13 corresponde con la información anual.
- temperaturaMediaMensual: La temperatura media que hizo en la ciudad en el mes correspondiente.
- temperaturaMediaMensual_maximas: La temperatura media de las máximas del mes.
- temperaturaMediaMensual_minimas: Temperatura media de las mínimas del mes.
- numDiasLluvia: Número de días que ha llovido en el mes correspondiente.
- PrecipitaciónTotalMensual: Precipitación total en el mes correspondiente, medida en mm.

A la hora de recopilar información, nos hemos encontrado con distintos errores que nos devolvía el servidor de Aemet. Uno de ellos era debido a que ciertas estaciones meteorológicas no contenían la información para alguno de los años. Esto provocaba un

error que hemos tenido que controlar haciendo que el código pasara al siguiente año del bucle.

Por otro lado, también nos hemos encontrado con una limitación, Aemet solo permite realizar un número muy limitado de consultas por minuto, y en caso de que lo superes, devuelve un estado de error 429; para evitar saturar el servidor con consultas, si éste devuelve error, hemos incorporado un tiempo de espera.

Para ciertas estaciones no se encuentra toda la información deseada. Por lo tanto, a la hora de recoger dicha información se realiza un control de nulos y si no existe el valor deseado se completa con “-”.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).

Gracias a la API proporcionada por Aemet, hemos podido recopilar el conjunto de datos que nos interesaba de una forma ordenada, clara y fiable. Además, permite publicar los datos obtenidos y utilizarlos con la única condición que se reconozca su origen.

Al realizar una búsqueda de un sitio web que proporcionara información meteorológica, encontramos distintas páginas web, que disponían de dicha información. Algunas de ellas facilitaban la recopilación de dicha información a través de una API, ya fuera de pago o con una versión gratuita.

Antes de encontrar el sitio web de Aemet, investigamos estas otras páginas, que finalmente descartamos por distintos motivos:

- <https://openweathermap.org/api>
- <https://www.el-tiempo.net/>

La página web openweathermap.org la descartamos, ya que la versión gratuita no disponía de información sobre datos históricos.

Por otro lado, la API de el-tiempo.net, aunque es gratuita y dispone de información de datos históricos, la información proporcionada era escasa para nuestro objetivo.

7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Como se ha explicado en la primera pregunta, debido a la reducción de las emisiones contaminantes, se desea estudiar los efectos que causa la contaminación a corto plazo. Para ello, se ha querido sacar información sobre la evolución de la temperatura en los últimos años, y compararla con la temperatura del año actual. Con el objetivo de comprobar si ha tenido algún efecto la pandemia sobre la temperatura y las precipitaciones.

Una cuestión que nos resulta interesante, es averiguar la gravedad de la contaminación y si su reducción en un corto periodo de tiempo tiene efecto en el clima. De este modo, podríamos averiguar si el efecto climático es reversible, es decir, si estamos a tiempo de revertirlo o ya es demasiado tarde. Es decir, las preguntas a las que se quiere dar respuesta serían las siguientes:

- ¿El ser humano con su actividad diaria interfiere en el cambio climático?
- ¿La reducción de actividad tan drástica de estos meses ha provocado un cambio en el tiempo?

- ¿Cuál es la rapidez con la que se producen los cambios en el tiempo a causa de la acción humana? ¿Es posible corregirlo a corto plazo?

8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

La licencia aplicada al dataset obtenido es una licencia Released Under CC BY-SA 4.0 License, por dos razones. La primera es que Aemet permite la reproducción de los datos obtenidos siempre que se mencione a © AEMET como autora de la información. La segunda, es que se permite el uso de los datos obtenidos con fines comerciales, al contrario de lo que ocurre con la licencia Released Under CC BY-NC-SA 4.0 License.

9. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

El dataset se ha publicado en formato CSV en Zenodo y aquí está su enlace:

<https://doi.org/10.5281/zenodo.4252794>

Contribuciones	Firma
Investigación previa	Marta, Pedro
Redacción de las respuestas	Marta, Pedro
Desarrollo código	Marta, Pedro