

Práctica 1 (25% nota final)

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

El contexto en el que se recolectan los datos es para la estimación de la variación del agua dulce en embalses y presas de España cruzando con datos de predicciones meteorológicas. Los datos de la cantidad de agua que hay en los embalses se obtienen de la web <https://www.embalses.net/> y la predicción meteorológica empleando el API que pone a nuestra disposición <https://opendata.aemet.es/centrodedescargas/inicio> .

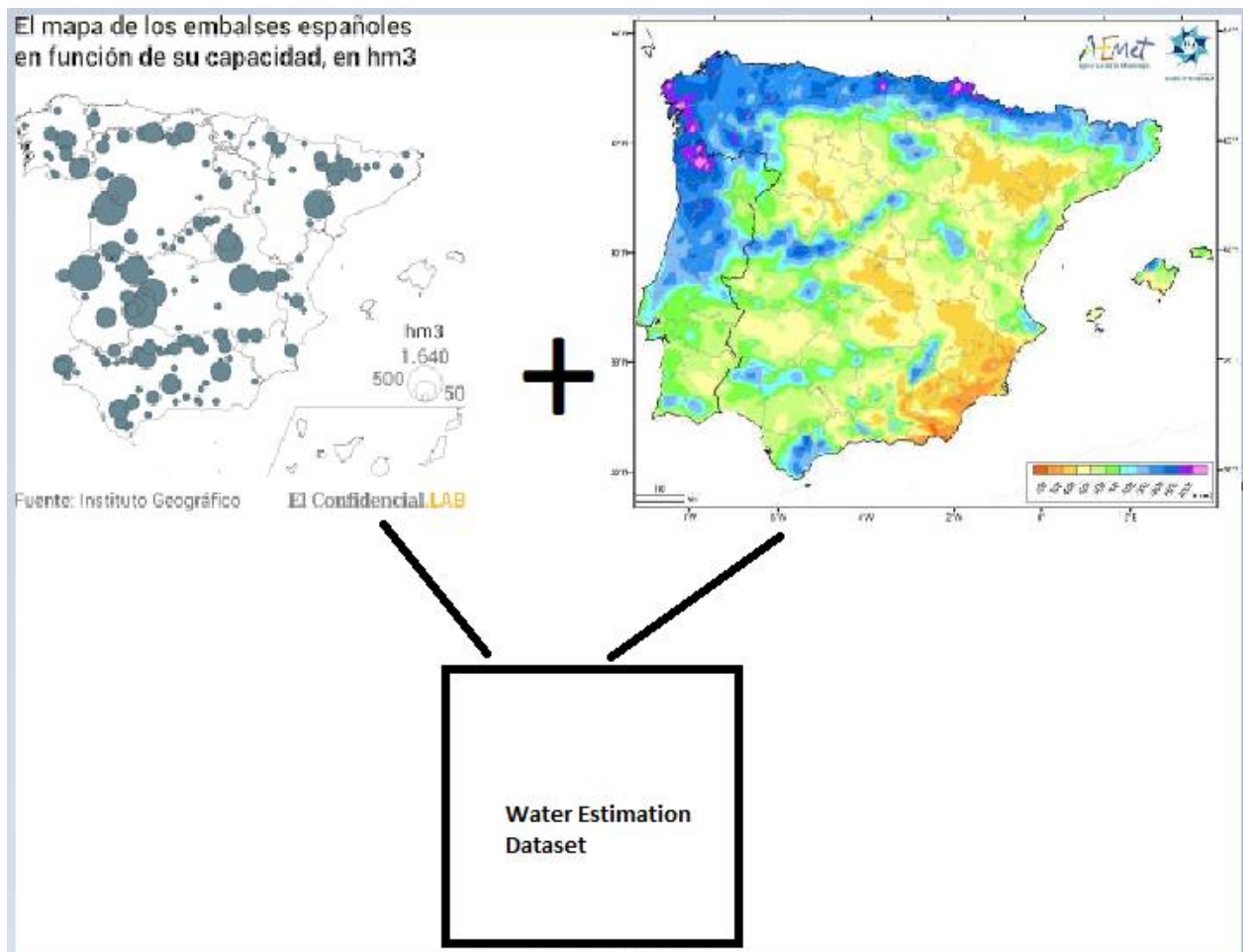
2. **Título.** Definir un título que sea descriptivo para el dataset.

El dataset se llama: Water-Estimation-Dataset

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.

El conjunto de datos identifica los diferentes puntos en los que disponemos de embalses y presas, contiene variables que codifican el nivel actual de los diferentes embalses junto con datos meteorológicos para poder estimar la evolución del agua dulce en España y anticipar posibles sequías.

4. **Representación gráfica.** Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.



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

- Lat: Latitud del embalse
- Lon: (coordenadas) Longitud del embalse
- Embalse: Nombre del embalse
- Nivel actual (hm³): Nivel del embalse en el momento de la medición.
- Diff: Diferencia del nivel del embalse respecto de la medición anterior.
- lon_x : longitud de la estación AEMET más cercana.
- lat_x= latitud de la estación AEMET más cercana
- Fecha de medición de precipitaciones
- Precipitaciones registradas

6. **Agradecimientos.** Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares. Justificar qué pasos se han seguido para actuar de acuerdo a los principios éticos y legales en el contexto del proyecto.

Los propietarios de la página de embalses.net son la empresa "ENOM , INC." :
{"registrar": "ENOM, INC.", "whois_server": "WHOIS.ENOM.COM"}
En cuanto a los datos meteorológicos proceden del portal de datos abiertos de la Agencia meteorológica (AEMET).

En cuanto al portal de AEMET como he podido comprobar que había una API disponible, he leído en un documento de preguntas frecuentes que se pueden realizar hasta 50 consultas por segundo, lo cual es muy superior a las necesidades del proyecto he optado por esta vía para lo que he tenido que obtener un API-key poniendo mi correo y ya he podido trabajar.
Por otra parte embalses.net no disponía de dicha API por lo que he tenido que consultar el fichero robots.txt en el que me he encontrado la siguiente recomendación:

```
robots = get_page("https://www.embalses.net/robots.txt")
robots
```

```
User-Agent: *
Disallow: /contenidos/
```

Por tanto, únicamente no deberé acceder a la carpeta contenidos y el resto lo puedo obtener los datos tranquilamente.

7. **Inspiración.** Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Este conjunto de datos es muy interesante porque puede ayudarnos a anticipar sequias problemas de riego en agricultura e incluso de abastecimiento de agua en ciertas regiones. Ya que las personas requerimos de agua dulce para subsistir y mantener cultivos y ganado es especialmente relevante tener un control de las zonas que requieren de transvases como el conocido Tajo-Segura.

8. **Licencia.** Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección:
- Released Under CC0: Public Domain License.
 - Released Under CC BY-NC-SA 4.0 License.
 - Released Under CC BY-SA 4.0 License.
 - Database released under Open Database License, individual contents under Database Contents License.
 - Other (specified above).
 - Unknown License.

Elijo “*Released Under CC0: Public Domain License*” porque permite a los científicos profesores y otros interesados aumentar el interese en este dataset ya que tienen todos los derechos de reusar y construir sobre este dataset soluciones que permitan ayudar a mejorar el problema de la sequía que lamentablemente por el cambio climático cada vez va a ser algo más importante.

9. **Código.** Adjuntar en el repositorio Git el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El repositorio en el que se encuentra el código:

<https://github.com/MAProjects/Water-Estimation-Dataset>

10. **Dataset.** Publicar el dataset obtenido (*) en formato CSV en Zenodo con unabreve descripción. Obtener y adjuntar el enlace del DOI.

El enlace del DOI es el siguiente:

<https://zenodo.org/badge/DOI/10.5281/zenodo.5652032.svg>

Contribuciones	Firma
Investigación previa	M.A
Redacción de las respuestas	M.A
Desarrollo del código	M.A