

1. Contexto

La información se ha extraído de la web www.embalses.net. Esta web contiene información actualizada semanalmente del estado de las reservas hídricas en cada uno de los embalses del territorio nacional de España. Esta web recoge la información del Ministerio de Transición Ecológica a través de su boletín hidrológico.

2. Título del dataset

Hemos decidido llamarlo Reservoirs/Embalses pues consideramos que es el nombre más adecuado habida cuenta de que contiene información de las reservas hídricas de los embalses así como otra información relevante a los mismos.

3. Descripción del dataset

El dataset contiene información acerca de las reservas hídricas de los diferentes embalses del territorio nacional español.

- Datos cuantitativos:

Estos datos contienen la cantidad de agua embalsada media en hectómetros cúbicos y su porcentaje durante la semana del año en la que nos encontramos. Además, como referencia, también contiene la cantidad embalsada y su porcentaje respecto a la misma semana del año anterior, y la misma semana de hace diez años.

Adicionalmente, contiene la capacidad total del embalse.

- Datos cualitativos:

Son datos referentes a la ubicación del embalse (Cuenca hidrográfica, provincia y municipio), río que abastece de agua al embalse, tipo de construcción y año de construcción.

- Información de uso:

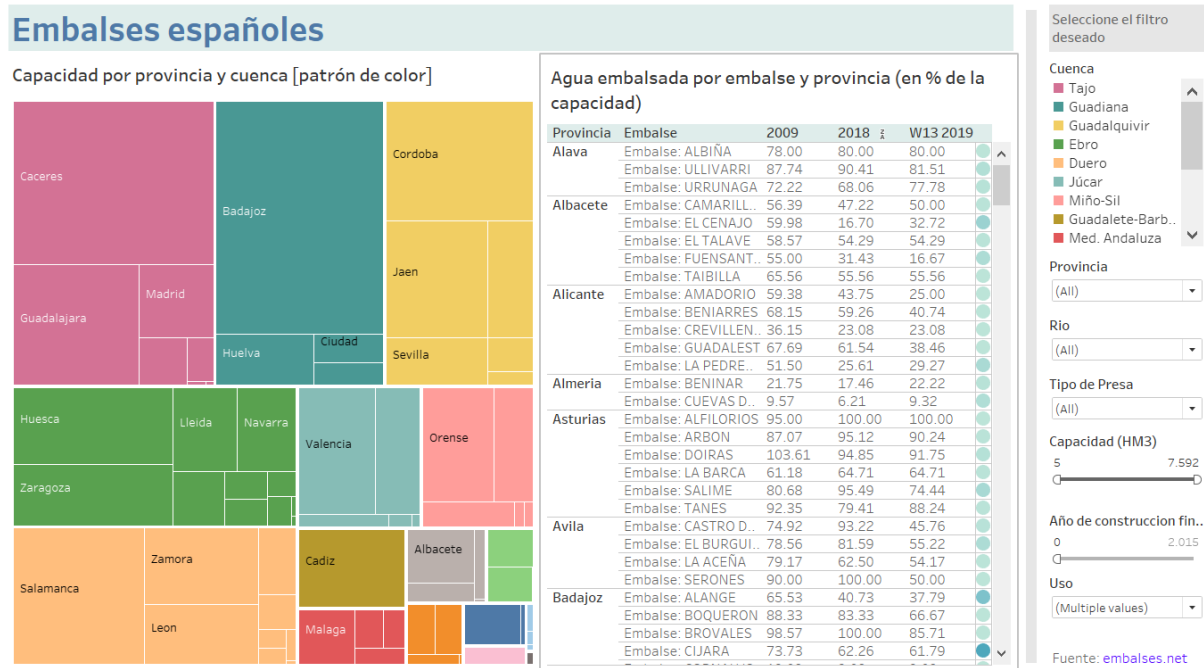
Estos datos indican cuales son los principales usos del embalse (abastecimiento, industrial, navegación, baño, pesca, etc.). Esta información se guarda en un valor booleano para cada una de las categorías.

4. Representación Gráfica

Si se tiene en consideración el volumen de datos del dataset, se puede representar de diversas formas.

En nuestro caso hemos decidido crear gráficos interactivos en Tableau que permiten al usuario seleccionar diferentes filtros como por ejemplo: provincia, cuenca, río, tipo de presa o uso del embalse.

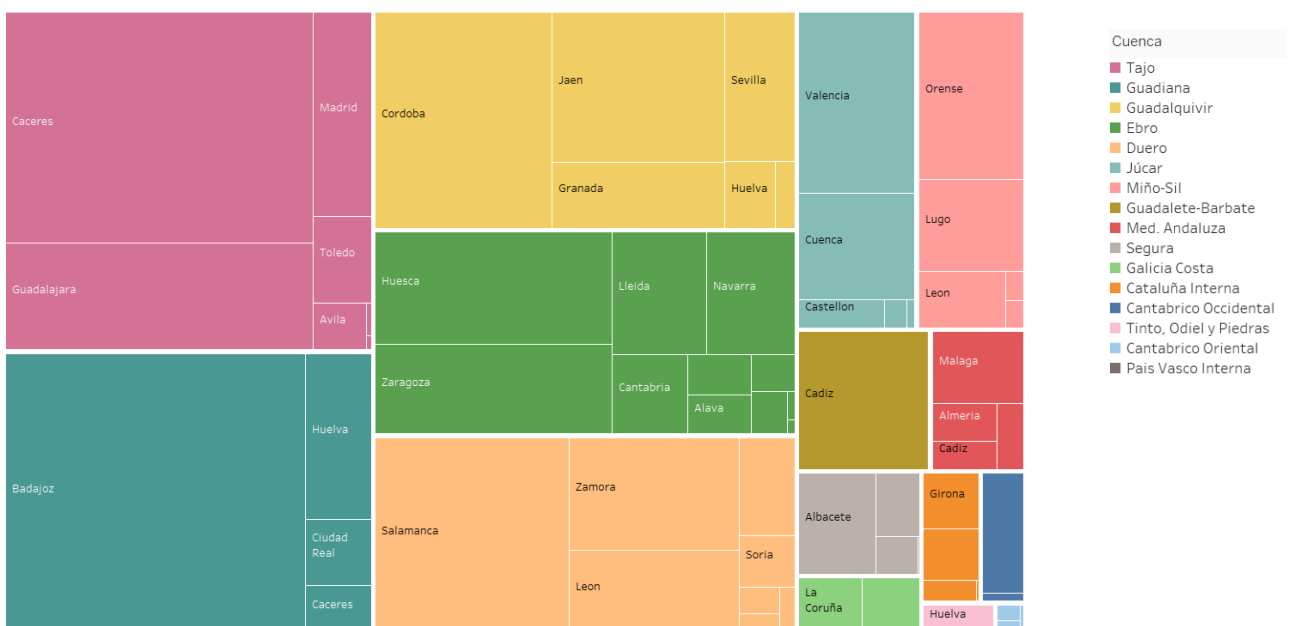
La publicación en Tableau se puede encontrar bajo este enlace:
<https://public.tableau.com/profile/aida.c3420#!/vizhome/EmbalsesReservoirsSpain/Maindashboard?publish=yes>



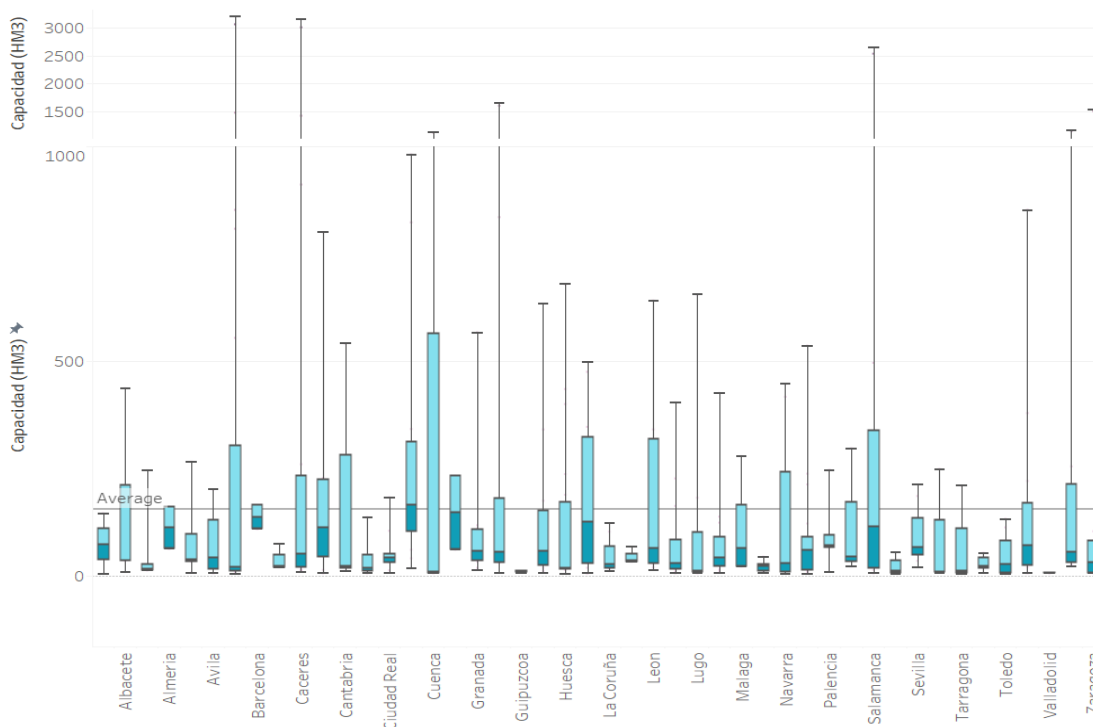
Una manera de empezar a explorar los datos es mediante los filtros a la derecha. Un gráfico adecuado es el diagrama de árbol o treemap, que nos muestra los datos estructurados del dataset en forma de árbol, por jerarquía, como un conjunto de rectángulos anidados.

Al existir una correlación entre el color (que representa cada cuenca) y el tamaño (que representa la capacidad en HM3), el diagrama ayuda a identificar patrones que, de otro modo, serían difíciles de ver.

Capacidad por provincia y cuenca [patrón de color]



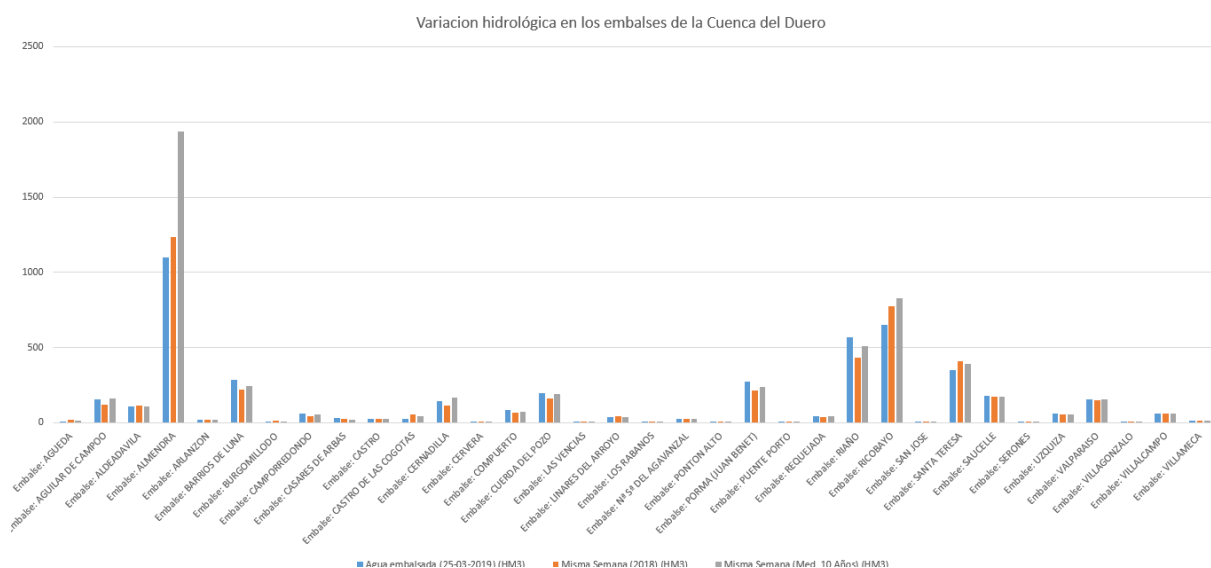
También podemos visualizar la capacidad por provincia, y ver la distribución de valores:



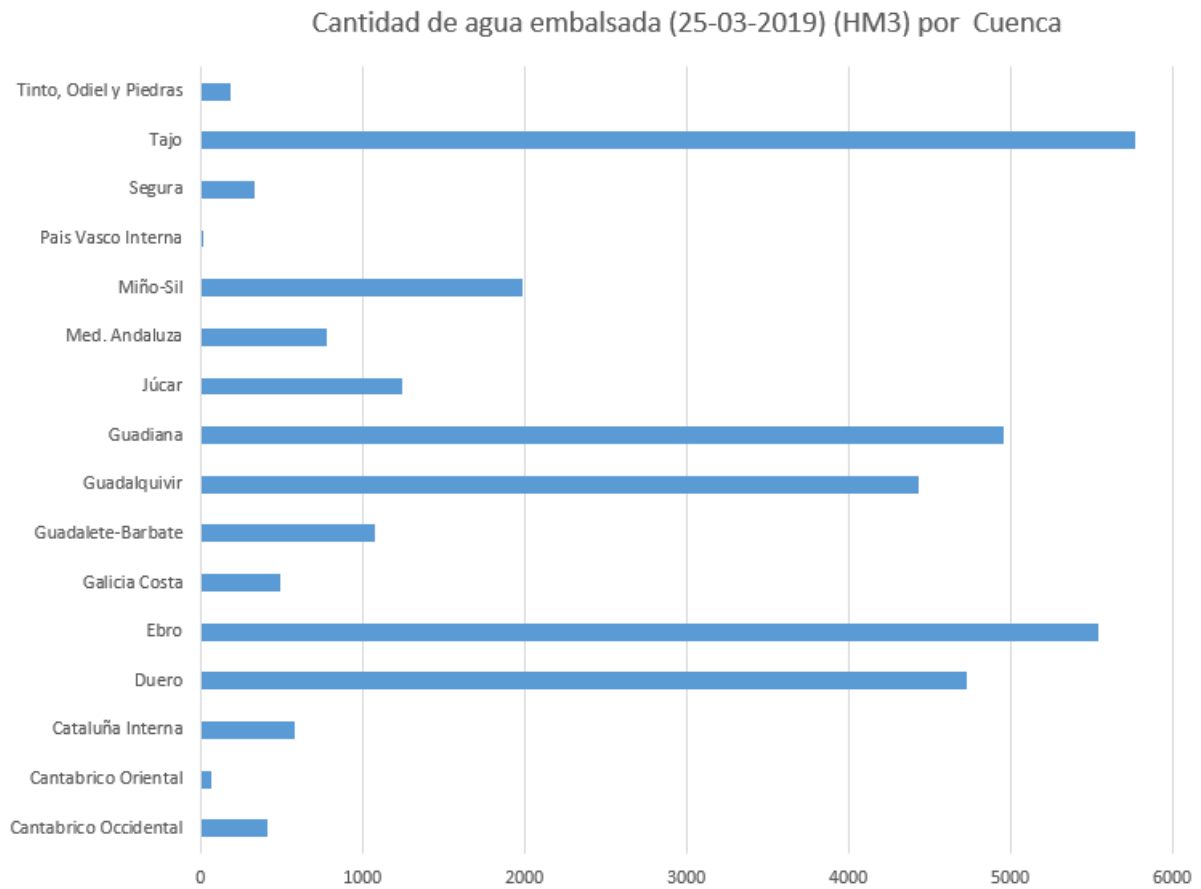
La media se sitúa en torno a los 150HM3, con provincias como Cuenca o Cantabria por encima de la media.

Igualmente, podemos representar los datos en Excel. Una manera de representarlos es mediante un diagrama de barras que muestre el agua embalsada con respecto al mismo dato la semana anterior de hace un año y respecto a hace 10 años. Esta información puede servir como referencia para indicar la tendencia.

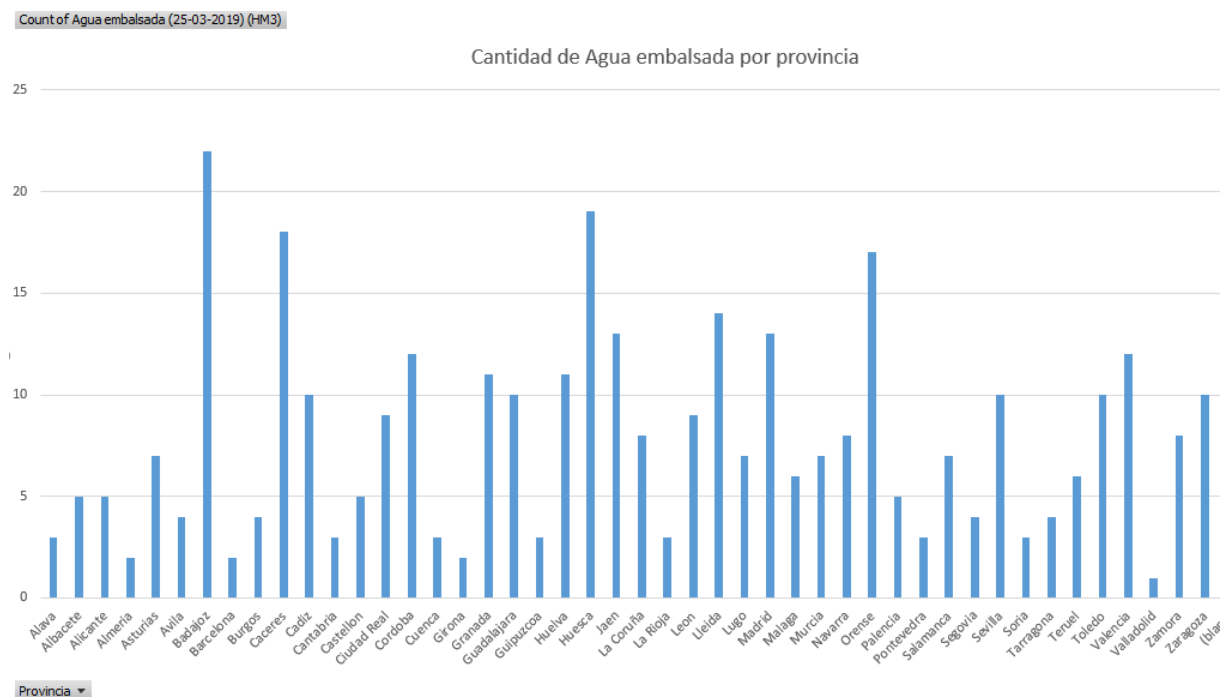
Debido al gran número de embalses, en el diagrama que se muestra debajo únicamente se han incluido los embalses de la cuenca del Duero.



Pero también se puede agrupar de otras formas para, por ejemplo, obtener la cantidad de agua total embalsada por cuenca hidrográfica:



O por provincia



5. Contenido del dataset

Los campos contenidos en el dataset son los siguientes:

- **Nombre del embalse:** Atributo que distingue unívocamente el embalse.
- **Agua embalsada (HM3):** Cantidad de agua media embalsada durante la última semana en hectómetros cúbicos.
- **Porcentaje:** Porcentaje de agua embalsada durante la última semana con respecto a su capacidad total.
- **Variación respecto a semana anterior (HM3):** Diferencia de agua embalsada con respecto a la semana anterior en hectómetros cúbicos.
- **Porcentaje Capacidad:** Mismo dato que el anterior, pero en porcentaje con respecto a la capacidad total
- **Capacidad:** Capacidad total del embalse en hectómetros cúbicos.
- **Misma Semana (año anterior) (HM3):** Cantidad de agua embalsada en la misma semana del año anterior, en hectómetros cúbicos.
- **Porcentaje Cuenca :** Misma información que en el campo anterior, mostrada en porcentaje con respecto a la capacidad total.
- **Misma Semana (Med. 10 Años) (HM3):** Cantidad de agua embalsada en la misma semana de hace 10 años, en hectómetros cúbicos.
- **Porcentaje:** Misma información que en el campo anterior, mostrada en porcentaje con respecto a la capacidad total
- **Cuenca:** Cuenca hidrográfica a la que pertenece el embalse.
- **Provincia:** Provincia en la que se encuentra el embalse.
- **Municipio:** Municipio al que pertenece
- **Río:** Río que abastece de agua al embalse.
- **Tipo de Presa:** Tipo de construcción de la presa (gravedad, bóveda, materiales sueltos, etc...)
- **Año de construcción:** Año en el que la presa fue construida.
- **Superficie:** Superficie en hectáreas.

Adicionalmente, se guarda información del uso del embalse. Se incluye un valor booleano para indicar si el embalse tiene alguno (o varios) de los siguientes usos:

- **Abastecimiento**
- **Riego**
- **Electricidad**
- **Industrial**
- **Pesca**
- **Navegación**
- **Baño**
- **Pic-nic**
- **Restaurantes**

El periodo de tiempo representa de una semana, ya que la web de la que se han extraído los datos actualiza la información semanalmente (cada domingo). Los datos presentados pertenecen a la semana del 25 al 31 de Marzo de 2019.

Para obtener los datos, el script de scraping, se conecta a la página principal de la web www.embalses.net, busca la página que contiene todas las cuencas y a partir de éstas, abre la página perteneciente a cada uno de los embalses y extrae la información.

6. Agradecimientos:

La información se ha obtenido de la web www.embalses.net que obtiene la información del Ministerio de Transición Ecológica. Dado que no hay ninguna referencia al propietario de la web embalses.net, nuestro agradecimiento va dedicado al Ministerio de Transición Ecológica y a su boletín hidrológico.

7. Inspiración

Consideramos esta dataset interesante porque puede ser usado en diversas áreas.

Por ejemplo, puede ser usado por investigadores para demostrar el efecto del cambio climático en las reservas hídricas en España. Si se guarda un histórico de datos durante un intervalo de tiempo, se podría observar la evolución de forma muy precisa.

También puede tener uso empresarial con el objetivo de encontrar nuevas oportunidades de negocio: considerando los campos relativos al uso de los embalses, se puede observar que hay varios de ellos dedicados al baño, pesca y/o navegación que carecen de áreas de pic-nic o restauración.

8. Licencia

La licencia que seleccionaríamos sería: Released Under CC0: Public Domain License (Fuente: <https://creativecommons.org/publicdomain/zero/1.0/deed.es>)

El Ministerio de Transición Ecológica ofrece acceso a estos datos al dominio público, por ello suponemos que renuncia a todos los derechos a la creación de ellos. Intuimos por ello que es permitido copiar, modificar, distribuir e interpretar la obra, incluso para propósitos comerciales, sin pedir permiso.

9. Código

El código se encuentra bajo:

https://github.com/Acent1/UOC_Web scraping_reservoirs/src

10. Dataset

El data set se encuentra bajo:

https://github.com/Acent1/UOC_Web scraping_reservoirs/data

Contribuciones

Contribuciones	Firma
Investigación Previa	AC, GC
Redacción de las respuestas	AC,GC
Deasarrollo del código	AC,GC