

A7: Selección del conjunto de datos

Asignatura: M2.859 - Visualización de datos aula 2

Autor: Ander Estebanez Centeno (aestebanezc)

Para esta práctica se presenta un conjunto de datos extraído de la web de subastas del BOE¹. La extracción se diseñó a lo largo de la asignatura *Ciclo de vida de los datos*, en 2019. El código ha tenido que ser actualizado con tal de adaptarlo a la nueva estructura de la página web.

La web contiene todas las subastas judiciales realizadas en España desde 2016 hasta ahora. Cada subasta contiene:

- Información general de la subasta como su código, fecha de inicio, tipo de subasta, cantidad reclamada, tasación, puja mínima, ...
- Información del administrador de la subasta.
- Listado de acreedores: nombre, NIF, dirección, localidad, ...
- Listado de lotes: Hay pujas con más de un lote, para cada uno se recoge las características del bien subastado. En caso de una vivienda será la referencia catastral, dirección, ... En el caso de vehículo será la matrícula, la marca, ...
- Listado de pujas: Aunque rara vez está informado recoge el código de la subasta, el código del lote y el importe de la puja.

Adicionalmente se ha descargado de Idescat información sociodemográfica para cada uno de los municipios de Cataluña.

El proyecto de extracción de datos junto con los datos extraídos y la descripción de la BD está disponible en el [GitHub](#)².

Preguntas:

- 1) *La relevancia del conjunto de datos en su contexto. ¿Son datos actuales? ¿Tratan un tema importante para algún colectivo concreto? ¿Se ha tenido en cuenta la perspectiva de género? (15%)*

Los datos contienen información desde 2016 hasta la actualidad y pueden ayudar a entender mejor los embargos: la clase de bienes que son embargados, el perfil de acreedor para cada tipo de bien, donde se concentran los bienes embargados, los importes, ...

Cabe destacar el potencial de la BBDD para identificar geográficamente donde se encuentran los bienes embargados y comparar esta frecuencia con las características sociodemográficas. Lo que equivale, en el caso de las viviendas, a conocer donde se dan los desahucios.

Debido a que la web no tiene datos relativos a los dueños de los bienes la perspectiva de género está limitada a la información descargada por municipio del Idescat, por ejemplo, el paro de hombre y mujeres en un municipio.

- 2) *La complejidad (tamaño, variables disponibles, tipos de datos, etc.). ¿Tiene cientos o miles de registros? ¿Tiene decenas de variables? ¿Combina datos*

¹ <https://subastas.boe.es/>

² <https://github.com/anderestebanez/Practica-1-Web-scraping>

categoricos y cuantitativos? Nota: Evita los conjuntos excesivamente simples. (30%)

La BD recoge un total de 24.171 subastas, lo que supone un total de 46.655 bienes. Lo que supone una tasación total de 3.096 millones de € (solo aquellos bienes con tasación disponible) y una cantidad reclamada total igual a 13.396 millones de €.

Respecto a la tipología de variables disponibles, existen tanto numéricas como categorías:

Tabla	Cuantitativos	Categoricos
General	10	24
Acreedores	0	5
Lotes	9	21
Pujas	1	0
Datos sociodemográficos	30	1

El contenido de cada uno de las tablas está recogido en el Excel disponible en Github ([./pdf/DataDictionary](#)).

- 3) *La originalidad. No repitas los conjuntos de datos clásicos; puedes, sin embargo, mejorar visualizaciones existentes. ¿ Hay otras visualizaciones basadas en este conjunto de datos? ¿Es una evolución o actualización de un conjunto anterior? ¿Has enriquecido un conjunto de datos ya existente? (20%)*

No conozco ninguna visualización que trabaje con estos datos.

En caso de considerarlo oportuno la BD se enriquecerá mediante la API del catastro, con el objetivo de asignar a cada bien inmueble, con referencia catastral informada, su coordenadas. Esto permitiría ver los bienes de forma individual y no agrupados por zona geográfica.

- 4) *Las cuestiones que responderás con la visualización de datos: ¿ Tienen en cuenta los puntos anteriores? ¿Están bien planteadas? ¿Son adecuadas para el conjunto de datos elegido? (35%)*

Mediante la visualización se intentará responder a:

- ¿Dónde está localizados los bienes que se subastan?
- ¿Existen correlaciones entre los datos sociodemográficos de cada municipio y la cantidad de bienes subastados?
- ¿Quiénes son los acreedores principales? ¿Existe diferencias en el perfil de bien/subasta según el acreedor?
- Cambiaría la respuesta a las preguntas anteriores en caso de filtrar los bienes y las subastas por importe, tipología, tipo de acreedor, ...