**Práctica 1: Web scraping**

**Asignatura: *M2.851 - Tipología y ciclo de vida de los datos aula 1***

**Autor: Ander Estebanez Centeno (aestebanezc)**

Contenido

[1. Contexto 2](#_Toc6176496)

[2. Definir un título para el *dataset* 2](#_Toc6176497)

[3. Descripción del *dataset* 2](#_Toc6176498)

[4. Representación gráfica 3](#_Toc6176499)

[5. Contenido 5](#_Toc6176500)

[6. Agradecimientos 9](#_Toc6176501)

[7. Inspiración 9](#_Toc6176502)

[8. Licencia. 9](#_Toc6176503)

[9. Tabla de contribuciones 10](#_Toc6176504)

## Contexto

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

Después de que en 2008 estallara la burbuja inmobiliaria en España dejando en evidencia el gran endeudamiento de familias y empresas. Además, las entidades bancarias acaban adjudicándose multitud de propiedades que acaban quedando vacías a la espera que llegue algún comprador. Por ello he considero interesante analizar por un lado el comportamiento de la morosidad en Cataluña, pero también saber que propiedades se han adjudicado los bancos y conocer el funcionamiento de las subastas.

¿Cuándo se produce la subasta? Si un acreedor impaga un préstamo de forma reiterada el deudor podrá recurrir a la justicia para recobrar el importe. La justicia podrá llevar a subasta los bienes que garantizaban el préstamo con el fin de que el acreedor salde su deuda. Los bienes subastados más conocidos son las viviendas, pero también pueden ser vehículos y otros bienes muebles.

Se espera que el *dataset* permitasaber dónde se concentran los morosos, analizar las entidades acreedoras, la procedencia de la subasta, los precios con los que salen, como se cierran, ...

El sitio web elegido contiene todas las subastas desde 2016 hechas en España con información sobre el acreedor, la autoridad gestora, información general de la subasta, los importes de apertura y condiciones de puja, la puja final, ... En definitiva, todo lo necesario para analizar las ejecuciones de garantías mediante procedimiento judicial o administrativo en España.

## Definir un título para el *dataset*

Elegir un título que sea descriptivo.

Subastas judiciales y administrativas

## 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).

Listado de todas las subastas judiciales y administrativas que ha habido en Catalunya desde 2016. Cada subasta contiene información sobre las normas que aplican a dicha subasta, los sujetos que participan y los bienes subastados.

## Representación gráfica

Presentar una imagen o esquema que identifique el dataset visualmente

La web tiene la siguiente estructura:

Información general

Acreedor

Administrador concursal

Interesados

Autoridad gestora

Bienes

Lote

Pujas

**Búsqueda de subastas**

Las subastas se muestran en grupos de 200

**Información básica**

Siempre se muestra y excepto con las pujas en caso de lote, la información es única por subasta.

**Información implicados**

No se muestra siempre y recoge información de sujetos implicados en la subasta.

**Información bienes**

Información de los bienes subastados. Los lotes son conjuntos de bienes que se subastan por separado, debiéndose extraer información de cada uno de ellos.

Siempre se muestra uno de los dos. Pero son excluyentes.

La información que contiene cada una de las subastas es diferente por lo que se debe ser abrir el código a todas las casuísticas.

A nivel gráfico hay que destacar las siguientes ventanas:

* *Resultado búsqueda:*



* *Estructura de la información de cada subasta*

**

## Contenido

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

Se ha descargado información de todas las subastas judiciales y administrativas desde 2016.

La información se almacena con un formato de diccionario de Python (*20190414\_listSubastas\_scrape*), que es el formato más adecuado porque:

* Permite que cada registro tenga diferentes campos.
* Permite que algunos sean listas de otros diccionarios. Esto es útil para guardar, por ejemplo, dentro del registro de una subasta todos los lotes que tiene en el campo ‘*lotes*’.
* Al estar toda la información contenida en un único registro no es necesario hacer cruces de tablas.

Debido a la estructura de la información la extracción no se puede hacer a un solo csv. La solución propuesta es generar diferentes *csv* que se relacionan entre sí a través de la clave *codSubasta*.

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

* *infoSubastas.csv*: Contiene la información que es única para una subasta e información descriptiva del resto de csv, como el número de interesados, número de lotes, …

|  |  |  |  |
| --- | --- | --- | --- |
| **Campo** | **Descripcion** | **Type** | **Ejemplo** |
| codSubasta | código de la subasta. | str | SUB-JA-2016-3971 |
| aut\_Fax | Fax de la autoridad gestora de la subasta. | str | 972219577 |
| aut\_Correo electrónico | Email de la autoridad gestora de la subasta. | str | - |
| Tipo de subasta | Tipo de subasta. | str | JUDICIAL EN VIA DE APREMIO |
| aut\_Dirección | Dirección de la autoridad gestora de la subasta. | str | AV RAMON FOLCH 4-6 4-6 , 17001 GIRONA |
| aut\_Código | Código de la autoridad gestora de la subasta. | int | 1707942003 |
| descripcion | Descripción de la subasta. | str | C. Doctor Fleming 9 2º 3ª DE Banyole |
| adm\_NIF | NIF de la autoridad gestora de la subasta. | str | B55125827 |
| Identificador | Identificador de la subasta. Es como el codSubasta. | str | SUB-JA-2016-39710 |
| Expediente |  | str | 1578/13 |
| aut\_Teléfono | Teléfono de la autoridad gestora. | str | 972181718 |
| Forma adjudicación | Indica si es una subasta por lotes. | str | Separada para cada lote |
| Valor subasta | Valor total de la subasta sumando todos los lotes existentes. | currency | 308.183,21 € |
| Lotes | Número de lotes que tiene la subasta. | int | Sin lotes |
| Tramos entre pujas | Diferencia que tiene que haber entre una puja y otra. En caso de lotes, la media de los lotes. | currency | Sin tramos |
| Fecha de conclusión | Fecha en la que concluye la subasta. | Datetime | 29-04-2019 18:00:00 CET (ISO: 2019-04-29T18:00:00+02:00) |
| adm\_Nombre | Nombre del administrador. | str | FRADE GOBEO S.L.P |
| Tasación | Valor de tasación de todos los bienes de los que conste la subasta. | currency | No consta |
| adm\_Localidad | Localidad del administrador judicial. | str | 17001 Girona |
| Estado | Estado de la subasta. | str | Celebrándose |
| Autoridad | El juzgado encargado de gestionar la subasta. | str | JUZGADO 1ª INSTANCIA 3 - GIRONA |
| Importe del depósito | Importe del depósito necesario para la subasta. En caso de más de un lote, la media de los lotes. | currency | 15.409,16 € |
| adm\_Provincia | Provincia del administrador. | str | Gerona |
| Puja mínima | Puja mínima de la subasta. En caso de más de un lote, la media de los lotes. | currency | Sin puja mínima |
| Scrap | Si la subasta se ha podido scrapear en su totalidad. | bool | True |
| Anuncio BOE | BOE en el que se ha publicado el anuncio. | str | BOE-B-2019-14859 |
| Fecha de inicio | Fecha de inicio de la subasta. | Datetime | 09-04-2019 18:00:00 CET (ISO: 2019-04-09T18:00:00+02:00) |
| aut\_Descripción | Descripción de la autoridad gestora de la subasta. | str | JUZGADO 1ª INSTANCIA 3 |
| adm\_Dirección | Dirección del administrador. | str | Plaza Josep Pla 4 1 6 |
| Cantidad reclamada | Cantidad total reclamada al demandado. | currency | 185.889,15 € |
| Link | Link con toda la información de la subasta. | str | https://subastas.boe.es/... |
| numInteresados | Número de personas que están listadas en la pestaña de interesados. | int | 5 |
| numAcreedores | Número de acreedores de la pestaña acreedores. | int | 1 |
| totPuja | Total del importe de las pujas hechas. | float | 156969.67 |
| acreedorPrivilegiado | Indica si la puja tiene un acreedor privilegiado. | bool | False |

* *interesadosSubastas.csv*: Información de todos los sujetos clasificados como interesados.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Campo** | **Descripcion** | **Type** | **Formato** | **Ejemplo** |
| codSubasta | Código de la subasta | str |  | SUB-NV-2019-365699 |
| NIF | NIF del interesado | str |  | B55125827 |
| Localidad | Localidad del interesado | str |  | 17001 Girona |
| Nombre | Nombre del interesado | str |  | FRADE GOBEO SLP |
| País | País del interesado | str |  |  |
| Provincia | Provincia del interesado | str |  | Gerona |
| Dirección | Dirección del interesado | str |  | PLAÇA JOSEP PLA Nº 4 |

* *acreedoresSubastas.csv*: Información de todos los acreedores.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Campo** | **Descripcion** | **Type** | **Formato** | **Ejemplo** |
| codSubasta | Código de la subasta | str |  | SUB-JA-2019-123056 |
| NIF | NIF del acreedor | str |  | B62563960 |
| Localidad | Localidad del acreedor | str |  | 08302 Mataró |
| Nombre | Nombre del acreedor | str |  | ARQUITECTUS REGA, S.L. |
| País | País del acreedor | str |  | España |
| Provincia | Provincia del acreedor | str |  | Barcelona |
| tipoPrivilegiado | Acreedor del tipo privilegiado | bool |  | False |
| Dirección | Dirección del acreedor | str |  | C. Biada, 60 1º 1ª |

* *lotesSubastas.csv*: Información de cada uno de los lotes de las subastas y en los casos en los que solo hay un bien (>99% de los casos), información del bien subastado.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Campo** | **Descripcion** | **Type** | **Formato** | **Ejemplo** |
| codSubasta | Código de la subasta | str |  | SUB-JA-2019-123877 |
| lote | Código del lote | str |  | Lote 1 |
| Número de bastidor | Número de bastidor del vehículo | str |  | U5YHN511AFL193544 |
| Localidad | Localidad del bien | str |  | L'HOSPITALET DE LLOBREGAT |
| Valor Subasta | Valor de subasta del lote | float |  | 83.485,43 € |
| Puja mínima | Púja mínima del lote | float |  | Sin puja mínima |
| Referencia Registral | Referencia catastral del bien inmueble | str |  | 20160001228 |
| Cuota |  | int |  | 66 |
| Información adicional |  | str |  | INSCRITA en el Registro de la Propiedad de Gavá, al tomo 1102, libro 707, folio 21 |
| IDUFIR | Código IDUFIR del edificio | str |  | 8086000327226 |
| Título jurídico | Tipo jurídico de propiedad | str |  | PLENO DOMINIO |
| Descripción |  | str |  | marca RENAULT MEGANE. matrícula B-1620-UL. |
| Modelo | Modelo de vehículo | str |  | MEGANE |
| Fecha de adquisición | Fecha de adquisición del vehículo | date |  | 02/09/2015 |
| Fecha de matriculación | Fecha de matriculación del vehículo | date |  | 22/07/1998 |
| Inscripción registral |  | str |  | FINCA 12.843, INSCRITA EN EL REGISTRO DE LA PROPIEDAD Nº DE L'HOSPITALET DE LLOBREGAT |
| Vivienda habitual | Si es la vivienda habitual del actual inquilino | bool |  | No |
| Situación posesoria | Forma en la que posee el uso | str |  | Ocupantes con derecho de permanencia |
| Matrícula | Matrícula del vehículo | str |  | B-1620-UL |
| Marca | Marca del vehículo | str |  | RENAULT |
| Importe del depósito | Importe del depósito del lote | float |  | 4.174,27 € |
| Referencia catastral | Referencia catastral del inmueble | str |  | 8,083E+12 |
| Cargas | Si el bien tiene alguna otra carga. |  |  | A DISPOSICIÓN EN LA SECRETARIA DE ESTE JUZGADO |
| Visitable | Si se puede visitar el bien |  |  | No consta |
| Dirección | Dirección del bien |  |  | MOSSEN JAUME BUSQUETS 37. PLANTA PRINCIPAL PUERTA 1ª |
| Valor de tasación | Valor de tasación |  |  |  |
| numBienes | Número total de bienes que tiene el lote |  |  | 1 |
| Tramos entre pujas | Tramos mínimos entre las pujas del lote |  |  | 1.669,70 € |
| Provincia | Provincia del bien |  |  | Barcelona |
| Código Postal | Código postal del bien |  |  | 8903 |
| Depósito | Depósito para poder pujar |  |  |  |
| Cantidad reclamada | Cantidad total reclamada. |  |  |  |

* *pujasSubastas.csv*: Las pujas que se han hecho para cada una de las subastas.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Campo** | **Descripcion** | **Type** | **Formato** | **Ejemplo** |
| codSubasta | Código de la subasta | str |  | SUB-JA-2019-123877 |
| lote | Código del lote | int |  | 1 |
| Puja | Importe de la puja o indicado que no tiene puja | str |  | Sin puja |

## Agradecimientos

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

Agradecimiento a la *Agencia Estatal Boletín Oficial del Estado (AEBOE)* por poner al alcance de todos los ciudadanos información de las subastas, quitando opacidad al proceso.

También agradecer a kennethreitz (@kennethreitz) por haber creado la librería *requests* y a Leonard Richardson por la librería *Beautiful Soup*. Ambas librerías son imprescindibles en el código.

## Inspiración

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

El *DECRET LLEI 1/2015, de 24 de març, de mesures extraordinàries i urgents per a la mobilització dels habitatges provinents de processos d'execució hipotecària* establece el derecho de tanteo y retracto para la Generalitat, ayuntamientos y promotores sociales.

Actualmente los promotores sociales tienen un rol pasivo en el uso de este derecho. Solo esperan a que les llegue un comunicado, y en ese momento deberán tomar una decisión en solo 5 días, luego de forma apresurada y sin un análisis suficiente.

La razón por la que me ha interesado este *dataset* es que creo que puede permitir a las entidades de promoción social saber con antelación dónde hay un inmuble adjudicado por una entidad bancaria y decidir antes incluso de recibir el comunicado si les interesaría comprarlo o no. También daría la posibilidad a las entidades de acudir directamente a la entidad bancaria para llegar a un acuerdo que beneficie a ambas partes.

Además de lo anterior me parece que esta información puede resultar muy útil para analizar la morosidad del país.

## Licencia.

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

Para el *dataset* aplicaría la licencia *ODC Open Database License (ODbL)*, permitiendo a otras personas analizarlo y sacar conclusiones. Eso sí, creo que todo lo que se derive del *dataset* debería seguir siendo abierto.

Respecto al código de Python, el cual permite extraer el *dataset*, el tipo de licencia creo que debería ser diferente. Le aplicaría un licencia *Released Under CC BY-NC-SA 4.0 License*, debido a que permitiría:

* El uso libre del código.
* Mejorar y adaptar el código a las necesidades de uno manteniendo la licencia.
* Mantener la autoría.
* Evitar que otras personas comercialicen el trabajo hecho de forma voluntaria.

## Tabla de contribuciones

|  |  |
| --- | --- |
| **Contribuciones** | **Firma** |
| Investigación previa | AE |
| Redacción de las respuestas | AE |
| Desarrollo código | AE |