

Práctica 1: Idealista scraping

1. Título del dataset. Poned un título que sea descriptivo.

Inmuebles en venta dentro de Barcelona capital.

2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo.

Información sobre los inmuebles en venta ubicados en Barcelona capital, útil para realizar un análisis descriptivo y poder tomar decisiones en base a ello.

3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente.



4. Contexto. ¿Cuál es la materia del conjunto de datos?

El conjunto de datos explica las ofertas de venta de inmuebles en Barcelona capital. Podemos encontrar inmuebles adosados y pisos tanto de obra nueva como de segunda mano. Así mismo, estos pueden estar publicados directamente por un particular o por inmobiliaria (siendo estos últimos mayoritarios).

5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

En el conjunto de datos podemos encontrar los siguientes campos:

- Date: fecha en la que se descarga la publicación.
- Floor: planta del piso.
- Garage: indica si el inmueble tiene garaje.
- Link: enlace a la publicación.
- M2: metros cuadrados del inmueble.
- Off: descuento en el precio.
- Price: precio del inmueble.
- Rooms: número de habitaciones del inmueble.
- Title: título de la publicación (se puede encontrar la zona o la calle).

El periodo de tiempo recogido depende de la frecuencia de actualización/publicación de anuncios en el idealista. El crawler recoge, por defecto, las 50 primeras páginas de publicaciones.

6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Incluid citas de investigación o análisis anteriores.

El propietario de los datos es el sitio web idealista.com. Agradecemos mucho su trabajo, así como la accesibilidad al procesado de estos por parte de agentes externos (crawlers).

7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?

El conjunto de datos es interesante porque te permite realizar un análisis del mercado inmobiliario en Barcelona ciudad a partir de datos clave en el contexto.

A partir de un buen trabajo de estos datos se podría llegar a saber si está realizando una buena compra/venta, estimar el valor de un inmueble, detectar “chollos”, etc.

8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado:

- 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

La licencia escogida ha sido la **GNU General Public License v3.0** particularmente para que exista una coherencia entre el proyecto en Github y este documento. Esta licencia comparte características con la **CC BY-SA 4.0 License**. Por lo que, los motivos que han llevado a la elección de esta licencia son:

- Se permite tanto un uso privado como comercial, por lo que una evolución del software puede ayudar a mejorar la toma de decisiones de muchas personas en un tema tan importante como puede ser la vivienda. Y, por ende, se permite la modificación, evolución y distribución del software.
- Se debe citar la creación original del software, así se puede reconocer el trabajo propio y conocer las evoluciones sobre el mismo ya que debe de existir un estado de cambios.
- No existen responsabilidades legales ni garantía reconocida sobre el software en cuestión.

9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset.

Se adjunta el código en Python donde usándose la librería scrapy se genera el dataset.

10. Dataset. Dataset en formato CSV.

Se adjunta el dataset separado por comas.