

# Descripción de la práctica

## 1. Contexto

En un contexto donde existe una gran demanda de vivienda en alquiler, estamos interesados en conocer la evolución de su renta. Para ello vamos a proceder a la recogida diaria de los datos relativos a determinados parámetros de interés que aparecen en los anuncios publicados en el portal inmobiliario **fotocasa**[1]. Establecemos como área piloto del estudio la ciudad de Barcelona, pudiendo ser elegida cualquier otra ciudad, región o país que pueda ser objeto de análisis. La extracción de los datos la vamos a realizar con técnicas de web scraping con la finalidad de construir un **dataset**, éste podrá ser utilizado para la posterior creación de un índice de precios coyuntural que tendrá una periodicidad de 3 meses. Este índice permitirá tener un conocimiento de cómo es la evolución de los precios de venta para una ciudad, región o país determinado (en nuestro caso la ciudad de Barcelona). El resto de datos del **dataset** se podrán utilizar para crear un perfil de vivienda tipo del área elegida.

[1] [www.fotocasa.es](http://www.fotocasa.es)

## 2. Título

HouseScraper

## 3. Descripción del dataset

Los datos que se van a recoger del anuncio son el identificador del anuncio, la fecha de consulta, el precio de venta de la vivienda, el número de habitaciones, el número de baños, los metros cuadrados, el tipo de vivienda, la dirección publicada en el anuncio, la descripción del anuncio y el teléfono del anunciante.

## 4. Representación gráfica

|    | A         | B                   | C      | D  | E   | F             | G           | H   | I      | J       | K      | L                      |
|----|-----------|---------------------|--------|--|---|---------------|-------------|---|--------|---------|--------|------------------------|
| 4  | 157502967 | 2020-10-12 18:26:11 | 173000 | Amplio piso recientemente reformado. Piso alto con vistas maravillosas, es ul  | Nou Barris  | Piso          | 932 512 467 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 3 hab. | 1 baño  | 95 m²  | con ascensor           |
| 5  | 157502862 | 2020-10-12 18:26:11 | 240000 | La vivienda agradable y sp   | Caner de Cantàbria, Sant Martí                    | Piso          | 935 434 518 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 3 hab. | 1 baño  | 73 m²  | con ascensor y terraza |
| 6  | 157501905 | 2020-10-12 18:26:11 | 260000 | Actualmente tiene puertas correderas. Este estupendo loft exterior tiene una p | Ciutat de Balaguer, Sarrià - Sant Gervasi         | Piso          | 910 750 918 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 1 hab. | 1 baño  | 60 m²  | con terraza            |
| 7  | 157501904 | 2020-10-12 18:26:11 | 275000 | Os presentamos este precioso piso en ESTUPENDO PISO EN VENTA, MUY              | Calle de la Mare de Déu de Port, Santa - Montjuïc | Piso          | 910 750 918 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 3 hab. | 1 baño  | 82 m²  | con ascensor y terraza |
| 8  | 157501834 | 2020-10-12 18:26:11 | 170000 | C/CADÍ - C/ ANETO. Duplex en venta t   | Cadi, 41, Nou Barris                              | Duplex        | 932 934 068 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 4 hab. | 2 baños | 80 m²  | con terraza            |
| 9  | 157501638 | 2020-10-12 18:26:11 | 245000 | Arqs 14, Casa unifamiliar en zona enri   | Arqs 14, Ciutat Vella                             | Casa o chalet |             | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 4 hab. | 2 baños | 133 m² |                        |
| 10 | 157500320 | 2020-10-12 18:26:11 | 315000 | Primero real con ascensor. Excelente z   | Travessera de Gràcia, Gràcia                      | Piso          | 932 939 517 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 4 hab. | 1 baño  | 85 m²  | con ascensor           |
| 11 | 157500240 | 2020-10-12 18:26:11 | 595000 | Tiene una superficie de aproximadamer  | Calle de Cartagena, Eixample                      | Piso          | 910 750 918 | <a href="https://www.fotocasa.es/es/comprar/vivienda/barcelona">https://www.fotocasa.es/es/comprar/vivienda/barcelona</a> | 3 hab. | 3 baños | 125 m² | con ascensor y terraza |

## 5. Contenido del dataset

Los datos se han obtenido haciendo peticiones GET a la url de la búsqueda de compra de viviendas, obteniendo el código HTML y recorriéndolo para buscar los datos de los parámetros de interés.

Cada entrada tiene los siguientes campos:

| Nombre            | Tipo    | Tamaño | Descripción   |
|-------------------|---------|--------|---|
| Identificador     | Integer |        | Identificador del anuncio                                 |
| Fecha de consulta | Date    |        | Fecha en la que se ha obtenido la información             |
| Precio            | Integer |        | Precio de la vivienda                                     |
| Descripción       | Text    | 250    | Descripción del anuncio                                   |
| Dirección         | Text    | 50     | Dirección de la vivienda anunciada                        |
| Tipo              | Text    | 10     | Tipo de vivienda (Piso, Casa, Chalet,...)                 |
| Télefono          | Text    | 9      | Teléfono del anunciante                                   |
| Url               | Text    | 100    | Enlace del anuncio  |
| Habitaciones      | Integer |        | Número de habitaciones                                    |
| Baños             | Integer |        | Número de baños   |
| Superficie        | Real    |        | Metros cuadrados  |
| Otros             | String  |        | Otras características como si tiene terraza, garaje, etc. |

Los datos se guardan con la fecha en la que se obtuvieron ya que van variando con el tiempo.

## 6. Agradecimientos

Los datos originales pertenecen a **Fotocasa**, portal inmobiliario líder de España fundado en 1999. Sin la existencia de la actual web no sería posible este proyecto.

## 7. Inspiración

Este conjunto de datos es interesante porque, como ya anunciamos en el apartado 1, nos permite tener un conocimiento detallado de cómo es la evolución de la renta de alquiler residencial en un área, ciudad o país determinado, creando índices y/o predicciones de esta evolución. Por la naturaleza de los datos recogidos, podemos establecer zonas o áreas tensionadas dentro de las grandes ciudades, ayudando de este modo, a la regulación del mercado residencial de alquiler. También este dataset nos puede servir para obtener un perfil de la vivienda tipo en un área determinada, así como crear modelos para predecir la renta de ésta.

## 8. Licencia

Debido a que los datos de origen se obtienen de una página con derechos de autor (licencia Copyright) y a que no está pensada para la descarga automática de sus datos, este dataset no puede tener una licencia que permita cualquier clase de modificación de los datos como las Creative Commons.

La licencia elegida sería **Unknown License** puesto que el objetivo es que un usuario a nivel particular pueda usar los datos para tener un índice de precios y ver la evolución, pero sin usar esos datos para lucrarse ya sea directamente o tratándolos para obtener otros.

## 9. Otras observaciones

La página de la que se obtiene la información original no dispone de API que permita consultarlos de forma sencilla y leerlos a través de un XML o JSON, así que es necesario el uso de web scraping a las páginas HTML.

Además, tampoco permite el uso de peticiones GET a través de medios externos al navegador como la librería **requests** de Python. El resultado que se obtiene variando el user-agent y otras cabeceras es un contenido que requiere de verificación humana a través de captchas, concretamente de seleccionar una serie de imágenes que estén relacionados con lo indicado.

Aunque existen algunas librerías gratuitas para tratar los captchas, sólo funcionan para versiones simples en las que hay que introducir las letras que se muestran en una imagen. Las versiones más complejas como en este caso requerirían un fuerte trabajo adicional que van más allá de la asignatura y la práctica.

Por ese motivo el medio elegido para capturar el contenido es Selenium.

## 10. Contribuciones

| Contribuciones              | Firma   |
|-----------------------------|---|
| Investigación previa        | Jose Manuel García Rodes, Jorge Miranda Álamo |
| Redacción de las respuestas | Jose Manuel García Rodes, Jorge Miranda Álamo |
| Desarrollo del código       | Jose Manuel García Rodes, Jorge Miranda Álamo |