

Universitat
Oberta
de Catalunya

TIPOLOGÍA Y CICLO DE VIDA DE LOS DATOS

PRA1: WEB SCRAPING

CREACIÓN DE UN BOT PARA ANALIZAR PISOS DE ALQUILER EN FOTOCASA

Autores:

Javier Guimerans Alonso

y

Gerson Villalba Arana

ABRIL 2022

Índice

1	Contexto	3
2	Título	4
3	Descripción del dataset	5
4	Representación gráfica	6
5	Contenido	7
6	Agradecimientos	8
7	Inspiración	9
8	Licencia	10
9	Código	11
10	Dataset	12

1. Contexto

Los datos que se quieren recolectar son sobre precios de alquileres de viviendas en distintas ciudades Españolas. Por lo tanto, se plantean diferentes alternativas de webs que puedan proporcionar dicha información. Hay dos portales principales en el mercado español de viviendas que pueden proporcionar los datos:

- **Idealista**
- **Fotocasa**

Para la recolección de datos, se elige el segundo de ellos porque dispone de unas restricciones menores a la hora de hacer web scraping. Existen más portales, pero el resto tienen una cantidad de viviendas sustancialmente menor que los dos mencionados, y por ello se descartan.

2. Título

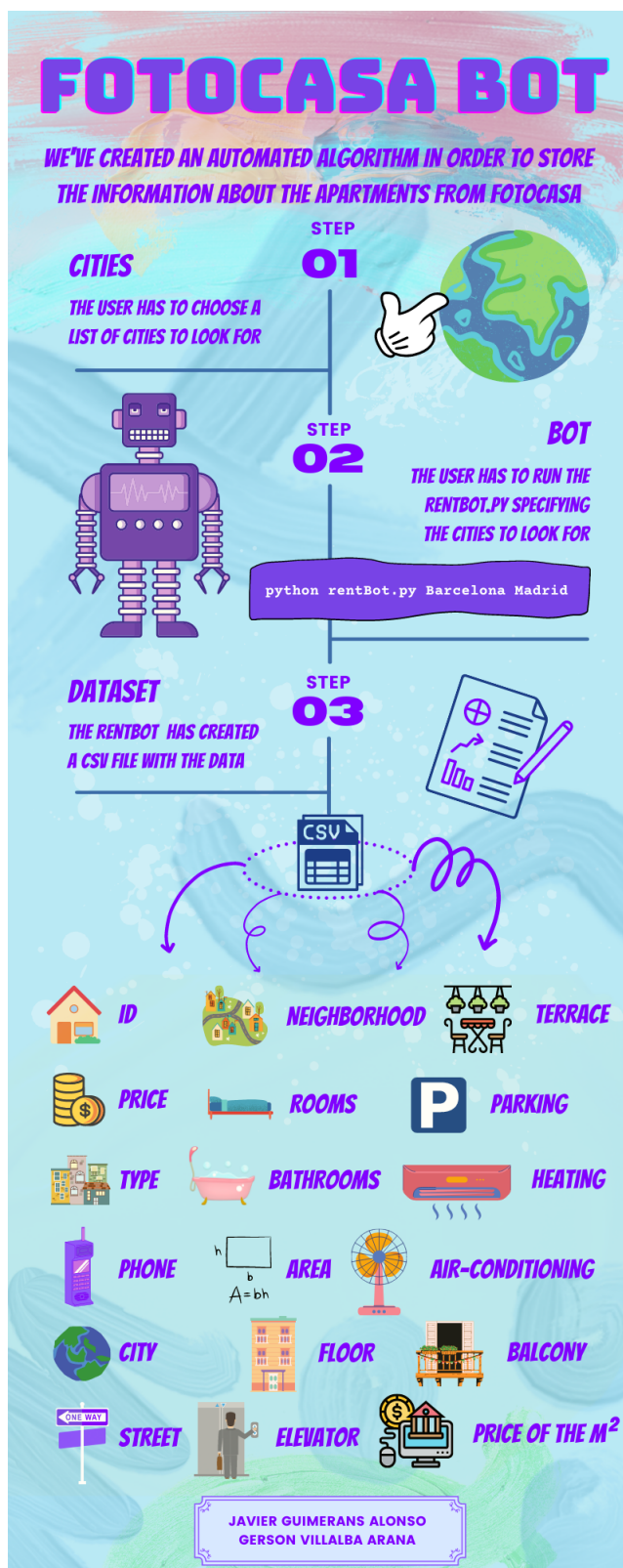
3. Descripción del dataset

El dataset recolectado a partir del web scraping sobre la web de Fotocasa contiene las características disponibles en el anuncio sobre la vivienda en cuestión ofertada, además del precio mensual del alquiler y número de teléfono del anunciante.

Estos datos obtenidos pueden tener gran utilidad para gran cantidad de aplicaciones. Algunas de estas pueden ser las siguientes:

- Análisis y comparativa de precios de alquiler en distintas ciudades españolas.
- Obtención y filtrado de viviendas según distintos parámetros.
- Predicción de precios de viviendas según características (regresión) con distintos algoritmos de machine learning.
- Agrupamiento de los pisos según características similares, para, por ejemplo, recomendar a un cliente viviendas similares a la que está buscando. Para ellos se podría utilizar un algoritmo no supervisado de clustering.

4. Representación gráfica



5. Contenido

El dataset obtenido a partir del web scraping tiene los siguientes campos:

- **ID:** Número identificativo de la vivienda.
- **Precio (€/mes):** Precio de la vivienda en euros mensuales.
- **Tipo:** Tipo de vivienda.
- **Teléfono:** Número de teléfono del anunciante.
- **Ciudad:** Ciudad donde se encuentra el inmueble.
- **Dirección:** Calle donde se encuentra el inmueble.
- **Barrio:** Barrio donde se encuentra la vivienda.
- **Habitaciones:** Número de habitaciones.
- **Baños:** Número de baños.
- **Superficie (m2):** Superficie de la vivienda en metros cuadrados.
- **Planta:** Planta donde se encuentra la vivienda.
- **Ascensor:** La vivienda dispone de ascensor.
- **Terraza:** La vivienda dispone de terraza.
- **Parking:** La vivienda dispone de parking.
- **Calefacción:** La vivienda dispone de calefacción.
- **Aire:** La vivienda dispone de aire acondicionado.
- **Balcón:** La vivienda dispone de balcón.
- **Precio del m2 (€/m2):** Relación entre el precio y la superficie de la vivienda en euros por metro cuadrado.

6. Agradecimientos

7. Inspiración

8. Licencia

El dataset se publica bajo una licencia CC BY-NC-SA 4.0.

9. Código

El código fuente completo que se ha utilizado para la obtención del dataset se encuentra en los siguientes repositorios:

<https://github.com/JavierGuimerans/rentBot>

<https://github.com/gvillalba86/rentBot>

10. Dataset

El dataset puede encontrarse en el mismo repositorio que el código fuente. Por otro lado, se ha publicado también en Zenodo en el siguiente enlace:

[Zenodo](#)