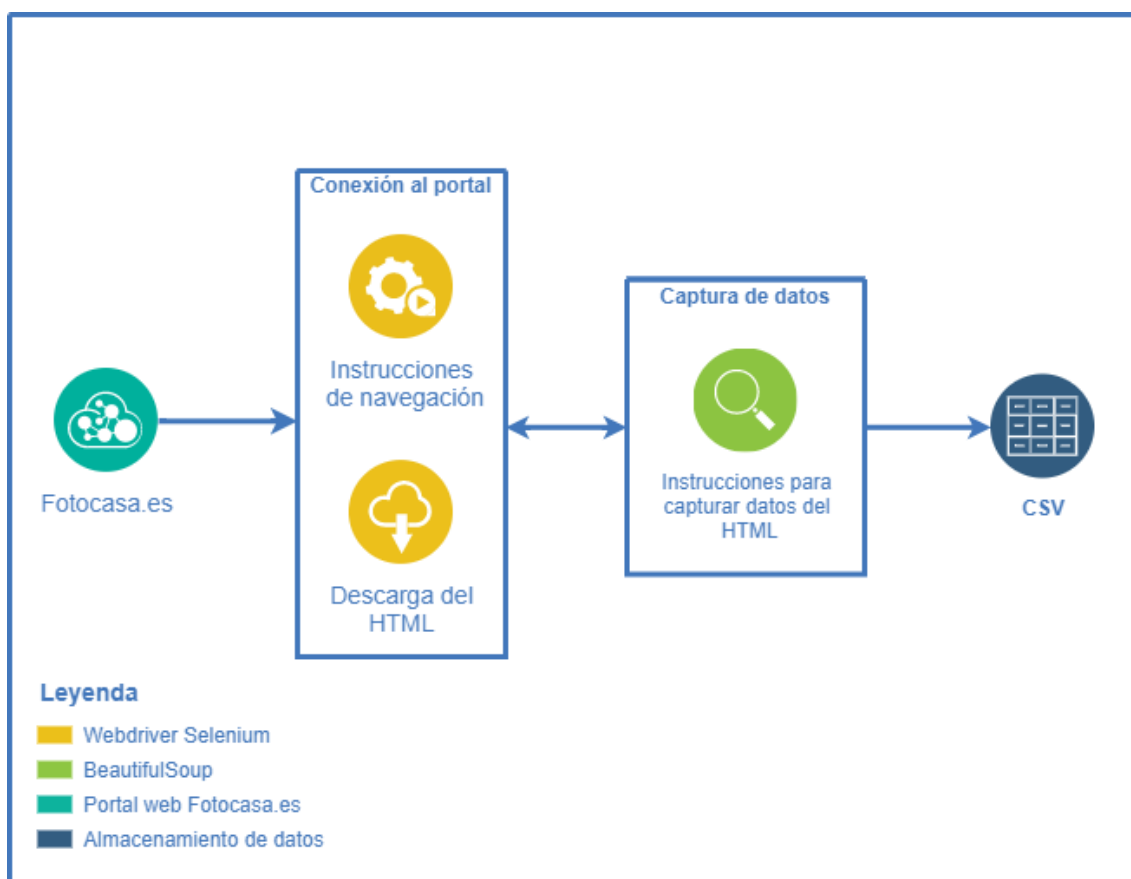




PRÁCTICA 1: WEB SCRAPING



Titulo

Conjunto de anuncios de viviendas publicados en Gran Canaria.

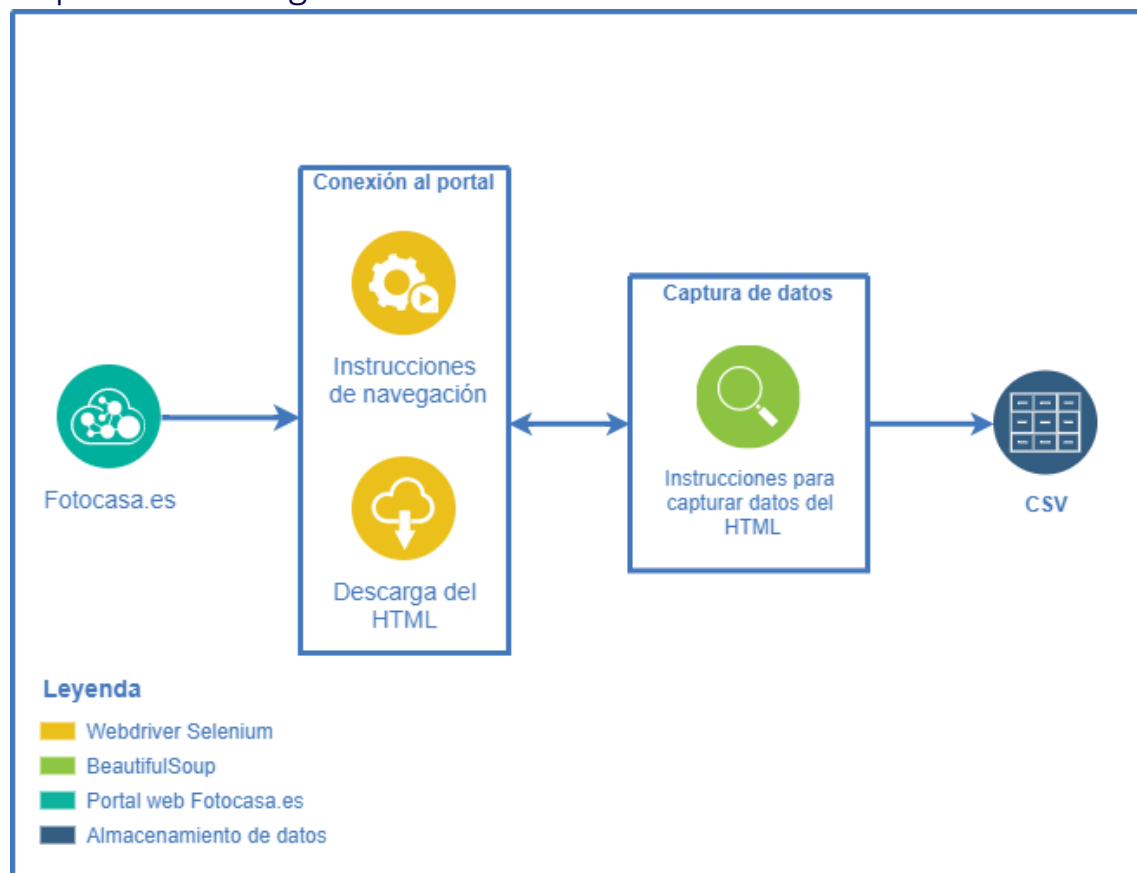
Contexto

La información se ha recolectado para conocer la situación inmobiliaria de viviendas en la isla de Gran Canaria, España. En futuras revisiones se podrían capturar otras regiones de España. Se ha elegido Fotocasa como fuente de datos porque es un portal web muy reconocido donde se anuncian inmuebles para alquiler y compraventa en toda España.

Descripción

El dataset contiene la información básica de los anuncios publicados en el portal Fotocasa.es para el territorio de la isla de Gran Canaria. Esta información es la que aparece en las “tarjetas” de cada anuncio en el listado de anuncios que vemos después de seleccionar la región.

Representación gráfica



Contenido

El dataset, almacenado en un fichero CSV separado por “;” contiene 8 campos:

- **href:** Contiene el enlace parcial al anuncio completo del inmueble. Esto es justo la parte que haya a continuación del dominio <https://www.fotocasa.es>.
- **title:** Título del anuncio
- **price:** Precio al que se oferta el inmueble, tanto el precio de compra como la mensualidad del alquiler según corresponda. Es un campo de tipo texto con formato “#.### €” o “#.### €/mes”

- **features:** Son las características principales con las que cuenta el inmueble. Formato tipo lista de Python ["feature 1", "feature 2", "feature N"].
- **propertyType:** Es la categoría que tiene inmueble, por ejemplo, "chalet o casa", "finca rústica", etc.
- **shortLocation:** Es una localización de referencia que no obedece a ningún criterio concreto. A veces es el municipio, otras veces la calle y el municipio, etc.
- **adType:** Es el campo que especifica si es un anuncio de compra, alquiler o de obra nueva. Obra nueva puede ser un alquiler también.
- **date:** Es la fecha de cuando se captura el dato en formato YYYY/MM/DD.

Los datos se han obtenido haciendo web scraping usando un script escrito en lenguaje Python. Se realizó partiendo de 3 enlaces diferentes según el tipo de anuncio (comprar, alquiler, obra nueva) y siempre para Gran Canaria. Cada enlace lleva a la página primera de una lista de anuncios del mismo tipo. He establecido un número máximo de paginación de 100.

Agradecimientos

Los datos se han obtenido del portal web Fotocasa.es que publica anuncios de compraventa y alquiler de inmuebles, especialmente de viviendas.

Así mismo, Fotocasa.es también ha publicado análisis de la evolución del precio medio de la vivienda en España como el que cito a continuación:

<https://www.fotocasa.es/blog/compraventa/conoce-el-precio-de-la-vivienda-en-venta-este-mes/>

Inspiración

Este conjunto de datos me resulta de interés para hacer un análisis inmobiliario en la isla de Gran Canaria, donde opera la empresa en la que trabajo. Se pretende hacer un seguimiento de la evolución de los precios de los inmuebles en la isla. También responder a las preguntas: ¿Qué características poseen los inmuebles que se ofertan? y ¿Qué influencia tienen dichas características en el precio?

Para hacer un análisis de evolución se necesitaría realizar el scraping web mensualmente. Además, veo interesante realizar una segunda versión que amplíe el contenido navegando al interior de cada anuncio para extraer más características y la descripción completa del inmueble.

Licencia

Se escoge la licencia **CC BY-SA 4.0 License** para publicar los datos. Considero que es la licencia adecuada por las siguientes cláusulas:

- Hay que dar el nombre del creador del conjunto de datos generados, indicando también los cambios que se han hecho. Se reconoce así el trabajo ajeno y qué aportaciones se han realizado respecto al trabajo original.
- Se permite el uso comercial, dando pie a las empresas a utilizar los datos generados que pueden propiciar cierto reconocimiento al autor original.
- Todas las contribuciones realizadas posteriormente sobre el trabajo publicado con esta licencia deben distribuirse con la misma licencia. De este modo, el trabajo del autor original continuará distribuyéndose bajo los mismos términos que él planteó.

Código

Se adjunta el código en un script de Python.

Dataset

Se adjunta un dataset con formato CSV con los campos separados por “;”.

Se ha publicado en zenodo con el DOI **10.5281/zenodo.4679925** .