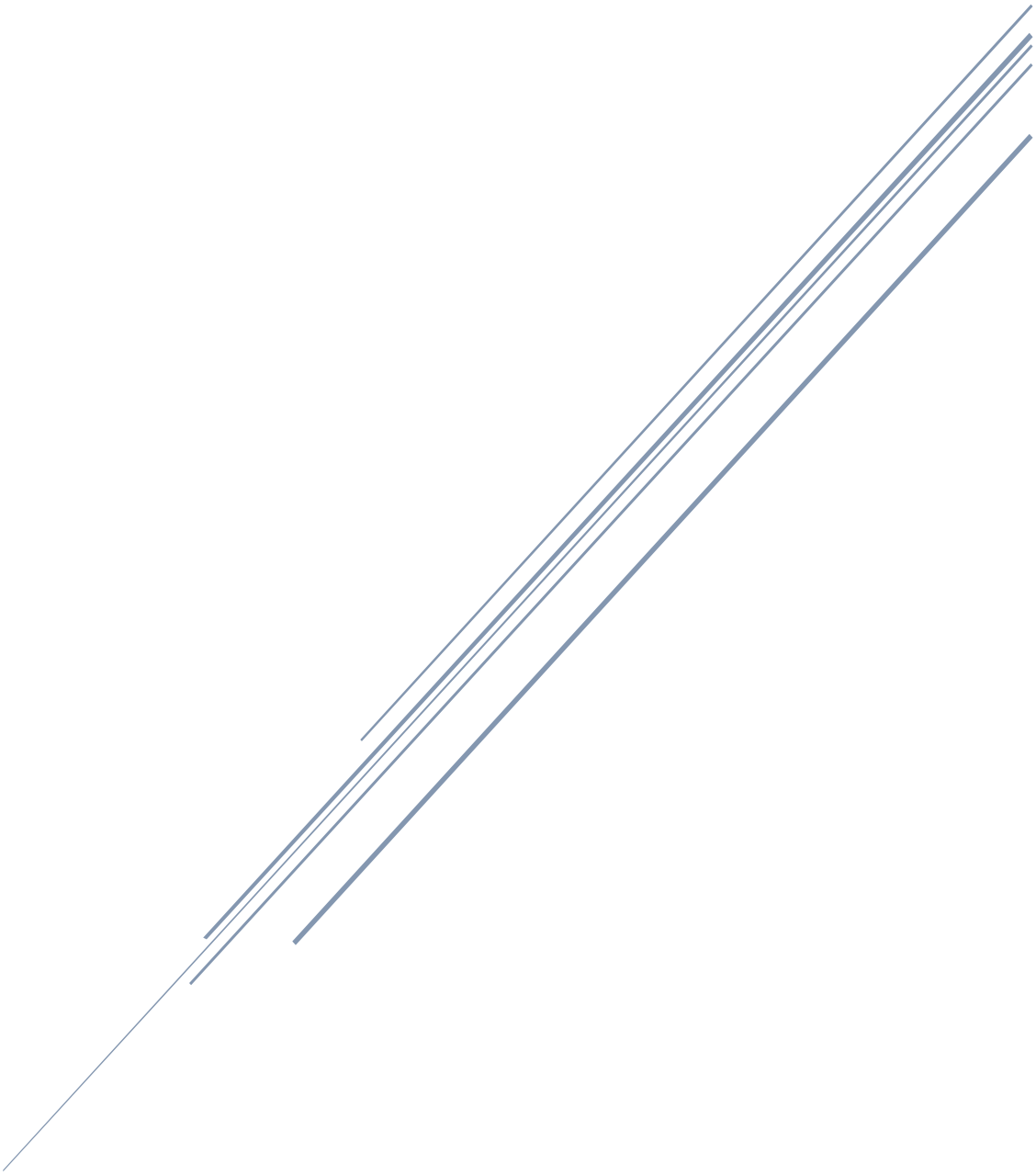
PRÁCTICA 1:

CICLO DE VIDA Y TIPOLOGÍA DE DATOS



Kevin Luna
Marc Nieto

1. Entregas

- Enlace a Github: https://github.com/kevinluna98/pec1_MarcKevin
- DOI Zenodo: https://zenodo.org/record/5597774#.YXb6uJ5BxhE
- Video: En progreso.

2. Contexto

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

El negocio inmobiliario ha sido uno de los principales focos en la especulación de España. Además, se trata de un tema de rabiosa actualidad [1] con la escalada de precios tanto en venta como alquiler en los últimos años, siendo un problema para los jóvenes conseguir acceder a una vivienda digna. Prueba de esto son las nuevas ayudas para los jóvenes que está poniendo en marcha el gobierno [2]. En esta práctica, pretendemos extraer información sobre los precios de las viviendas disponibles en Barcelona, para poder realizar un análisis de estos datos.

La página web elegida es www.pisos.com al ser uno de los portales inmobiliarios favoritos de España. Además, en la actualidad tiene listados más de 6.000 pisos en venta y 1.300 en alquiler en Barcelona. Por otro lado, la información es relevante y fácilmente extraíble.

3. Título

Definir un título que sea descriptivo para el dataset.

El nombre escogido para los archivos .csv que se irán generando del Web Scraping es el siguiente: barcelona_house_pricing_dd_MM_YY.csv el cual indica de una manera breve de que trata el dataset y la fecha de recogida.

El título del dataset [5] que hemos generado es pues: Barcelona House Pricing 24/10/2021

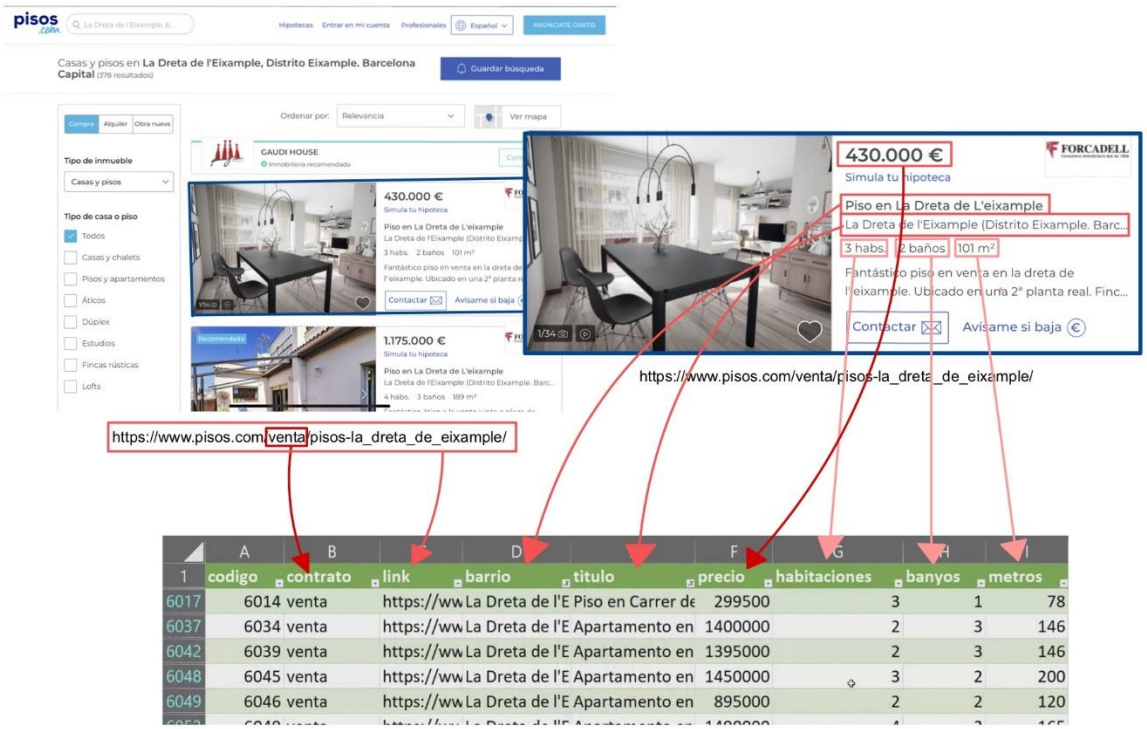
4. Descripción del dataset

Desarrollar una descripción breve del conjunto de datos que se ha extraído. Es necesario que esta descripción tenga sentido con el título elegido

El dataset creado proporciona los pisos en venta o alquiler y las características principales (número de habitaciones, baños, metros cuadrados, barrio y precio) de un piso en Barcelona, que estaban publicados a fecha 24/10/2021 en la web de pisos.com.

5. Representación gráfica

Dibujar un esquema o diagrama que identifique el dataset visualmente y el proyecto elegido.



	A	B	C	D	E	F	G	H	I
1	codigo	contrato	link	barrio	titulo	precio	habitaciones	banyos	metros
6017	6014	venta	https://ww	La Dreta de l'E	Piso en Carrer de	299500	3	1	78
6037	6034	venta	https://ww	La Dreta de l'E	Apartamento en	1400000	2	3	146
6042	6039	venta	https://ww	La Dreta de l'E	Apartamento en	1395000	2	3	146
6048	6045	venta	https://ww	La Dreta de l'E	Apartamento en	1450000	3	2	200
6049	6046	venta	https://ww	La Dreta de l'E	Apartamento en	895000	2	2	120

Figura 1. Esquema de la recolección de datos y almacenamiento en dataset.

6. Contenido

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

Los campos que incluye el dataset se explican a continuación:

- **codigo:** identifica de forma unívoca cada piso extraído.
- **link:** página web de la que se ha extraído, permite conocer su posición en la lista.
- **barrio:** barrio de Barcelona en el que se encuentra el piso.
- **descripcion:** pequeña descripción del piso.
- **precio:** precio del piso en cuestión.
- **habitaciones:** número de habitaciones que tiene el piso.
- **banyos:** número de baños del piso.
- **metros:** metros cuadrados del piso.

La extracción de los datos se ha realizado mediante técnicas de Web Scraping, en concreto usando el lenguaje Python y la librería Beautiful Soup.

En primer lugar, se ha analizado la página web de la que se va a scrapear, www.pisos.com. Aquí vemos que los links tienen la estructura www.pisos.com/{contract}/{neighborhood}/{page}, dónde {contract} puede ser venta o alquiler, en función de si se anuncia la venta o alquiler del

piso, {neighborhood} es el barrio y {page} la página. Por ejemplo, https://www.pisos.com/venta/pisos-la_dreta_de_eixample/2 indica que estamos viendo las viviendas en venta en el barrio de la Dreta de l'Eixample que hay en la página 2.

De esta forma, lo primero que necesitamos es una lista con todos los barrios de los que pretendemos sacar información, en nuestro caso el área de interés será la ciudad de Barcelona.

La implementación se basa en un triple bucle dónde primero iteramos el campo {contract} entre venta y alquiler, después la lista de barrios {neighborhood} y finalmente un bucle infinito que itera las distintas páginas {page} hasta que detecta que no hay más pisos y pasa al siguiente barrio.

Para cada página a la que se accede se extraen todos los objetos de anuncio (que se identifican con "div",class_="ad-preview__info") y como sus hijos se extrae la información del barrio ("p",class_="p-sm"), título (a",class_="ad-preview__title") y precio ("span",class_="ad-preview__price") del anuncio. Además, extraemos la información de número de habitaciones, baños y metros cuadrados, estos elementos no tienen ni tags de html ni clases únicas (se identifican todos como "p", class_="ad-preview__char p-sm"). Para diferenciarlos procesamos el texto ya que el dato del baño va acompañado del texto "baño" o "baños", en función de si es uno o más, análogamente para las habitaciones buscamos "habs." o "hab." y para los metros cuadrados "m²".

Finalmente, a los datos extraídos les añadimos un código único, el link del que provienen y si se trata de una venta o alquiler. Por supuesto todas las variables son inicializadas previamente a null, de forma que si en un anuncio falta algún dato se informará null.

La información se guarda en un .csv con la fecha de la extracción en el nombre. Para más detalle del funcionamiento se puede consultar el código.

7. Agradecimientos

Presentar al propietario del conjunto de datos. Es necesario incluir citas de análisis anteriores o, en caso de no haberlas, justificar esta búsqueda con análisis similares. Justificar qué pasos se han seguido para actuar de acuerdo con los principios éticos y legales en el contexto del proyecto.

Como se ha indicado previamente los datos han sido extraídos del portal inmobiliario www.pisos.com, lleva en funcionamiento desde el año 2009, con unas visitas mensuales de 6,18 millones [3].

El web scraping es un proceso de recolección de datos a través de internet, se usa en muchas operaciones legítimas de análisis de datos. Está técnica en si misma no es ilegal, pero puede llegar a serlo dependiendo del tipo de datos que recolectes, cómo decidas usar estos datos y, por último, la manera en que extraes los datos.

Respecto a la licencia de los datos y ética de uso de estos, en el apartado 4. Propiedad Intelectual e Industrial de la política de privacidad de pisos.com podemos leer lo siguiente:

> *4. Propiedad Intelectual e Industrial*
>
> *Los contenidos, elementos e información a los que el usuario pueda acceder a través de pisos.com están sujetos a derechos de propiedad industrial e intelectual, patentes marcas, copyright del titular de pisos.com o de terceros titulares de los mismos. En consecuencia, el acceso a estos contenidos o elementos no otorga al Usuario el derecho de alteración, modificación, explotación, reproducción, distribución o comunicación pública o cualquier otro derecho que corresponda al titular del derecho afectado.*
>
> *El Usuario se compromete a utilizar los contenidos y/o elementos a los que acceda a través de los Servicios de pisos.com para su propio uso y necesidades, y a no realizar en ningún caso una explotación comercial, directa o indirecta de los mismos.*

Este apartado es algo general, pero hay que tener en cuenta que los anuncios que estamos escrapeando se tratan de unos datos públicos y accesibles ya que la propia naturaleza de la plataforma es que la información de los pisos llegue lo más lejos posible para atraer a compradores y arrendadores. Además, el uso que estamos haciendo de los datos extraídos no se trata en ningún caso de una explotación comercial. Por lo que consideramos ético y correcto la extracción y análisis que vamos a realizar de esta información.

Como el dataset ha sido generado por nosotros mismos y a fecha reciente (24/10/2021) no existen, hasta dónde sabemos, análisis previos con estos datos. Si existen análisis similares con este tipo de datos, la mayoría hechos por los propios portales inmobiliarios. Por ejemplo, el mismo www.pisos.com tiene toda una sección [4] de su página web a análisis para hacer seguimiento de los precios de alquiler y compra de las viviendas.

8. Inspiración

Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Como se indica en la introducción en los últimos meses ha llegado hasta los medios noticias sobre el aumento de los precios de la vivienda y en concreto de los alquileres. También recientemente el gobierno ha lanzado ayudas al alquiler para los más jóvenes [1]. Todos estos sucesos, han hecho que la búsqueda de un alquiler o comprar de una vivienda, sobre todo en jóvenes, esté en alza. Por ello, nos ha parecido interesante recolectar una base de datos sobre los alquileres y ventas en Barcelona.

Las preguntas que pretendemos poder llegar a responder con esta base de datos son:

- Qué tipo de casa puede permitirse y en qué zona con el presupuesto que se posee.
- Barrios que son más asequibles.
- Barrios con las viviendas más grandes, por habitaciones o metros cuadrados.
- Barrios con la mejor relación de precio por metro cuadrado.

9. Licencia

Seleccionar una de estas licencias para el dataset resultante y justificar el motivo de su selección.

Se ha escogido la licencia **Creative Commons Attribution 4.0 International,** este tipo de licencia permite a lo usuarios distribuir, mezclar, adaptar y construir sobre el material en cualquier medio o formato.

10. Contribuciones

Contribuciones	Firma
Investigación previa	Kevin Luna, Marc Nieto
Redacción de las respuestas	Kevin Luna, Marc Nieto
Desarrollo del código	Kevin Luna, Marc Nieto

11. Bibliografía

[1] ABC Economía. Quiénes pueden solicitar la ayuda de 250 euros para el alquiler. https://www.abc.es/economia/abci-quienes-pueden-solicitar-ayuda-alquiler-250-euros-nsv-202110061249_noticia.html

[2] Ministerio de transportes, movilidad y agenda urbana. Programa de ayudas al alquiler de vivienda. https://www.mitma.gob.es/arquitectura-vivienda-y-suelo/programas-de-ayudas-a-la-vivienda/programa-de-ayudas-al-alquiler-de-vivienda

[3] HelpMyCash. Mejores portales inmobiliarios para vender piso en Expaña. https://www.helpmycash.com/cat/vender-piso/portales-inmobiliarios/

[4] Pisos.com. Sección Data. https://www.pisos.com/aldia/informes/alquiler/enero1970/1653008/

[5] Barcelona House Pricing 24/10/2021. https://doi.org/10.5281/zenodo.5597774