

Práctica 2 [PRA 2]

Marcos Iturbe Marengo

1. Descripción del dataset

El *dataset* escogido es “**AlquiESP(v0.04/2023)** - Pisos de alquiler en España disponibles en Pisos.com” que es el generado en la práctica pasada. El alto costo de los alquileres es un problema que afecta a muchas personas y comunidades. Los inquilinos a menudo tienen dificultades para encontrar un lugar adecuado para vivir a precios razonables, lo que puede tener un impacto significativo en su calidad de vida. La [Nueva Ley de la Vivienda](#) en España, parece ser la medida que calme esta situación. Con *datasets* como este se pueden observar las tendencias de los precios y ver que medidas son efectivas y cuales no.

El *dataset* incluye las siguientes columnas:

- **título** - Título del anuncio de la vivienda en alquiler
 - Ejemplo: *Piso en Carretera de Ciudad de Cádiz*
- **url** - Link al anuncio concreto
 - Ejemplo:
https://www.pisos.com/alquilar/piso-urbanizacion-de-roquetas-las-marinas04740-33340257518_100500/
- **zona** - Provincia a la que pertenece el anuncio
 - Ejemplo: *Alquilar en Almería (382)*
- **dirección** - Dirección de la vivienda en cuestión
 - Ejemplo: *Urbanización de Roquetas-Las Marinas (Roquetas de Mar)*
- **precio** - Precio del alquiler de la vivienda
 - Ejemplo: *1.000 €/sem*
- **detalles** - Lista de detalles. Dependiendo de la vivienda puede no contenerlos todos pero son los siguientes:
 - habitaciones
 - baños
 - m²
 - planta
 - Ejemplo: ['2 habs.', '1 baño', '60 m²', 'Bajo']
- **descripción** - Descripción de la vivienda
 - Ejemplo: *Este piso se encuentra en avenida ciudad de cádiz, 04740, roquetas de mar, almería, situado en el distrito de roquetas de mar, en ...*
- **fecha extracción** - Fecha y hora de extracción del dato
 - Ejemplo: *2023-04-25 11:42:59*

2. Integración y selección

Antes de realizar esta parte se ha hecho una limpieza previa de los datos. Sin embargo para mantener el orden del enunciado se presentará esta parte primero.

En cuanto a la **selección** se han obtenido todos los pisos excepto los que pertenecen a 'Cerdanya Francesa' ya que no se encuentra dentro de España. También se han seleccionado únicamente los pisos que tienen un precio mensual.

Por otro lado se han **integrado** los datos de las coordenadas geográficas lo que permite mostrar en un mapa las viviendas. Para se ha usado Nominatim que es un servicio de geocodificación y búsqueda geográfica proporcionado por OpenStreetMap (OSM), un proyecto colaborativo de mapeo abierto. Nominatim se utiliza para convertir direcciones o nombres de lugares en coordenadas geográficas (latitud y longitud) y viceversa.

<https://github.com/osm-search/Nominatim>

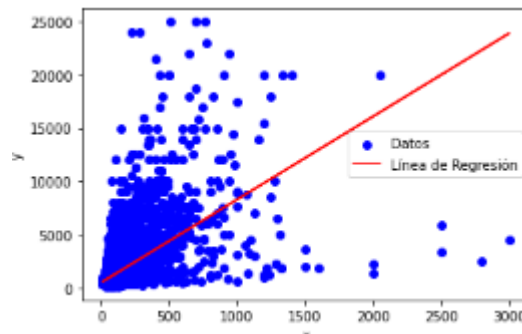
3. Limpieza de los datos

Se han realizado varios procesos de limpieza y procesado de los datos.

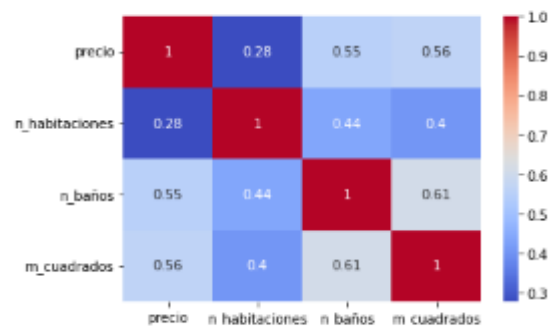
- Se ha limpiado la columna Zona. Esta columna había sido generada a partir de los links del pie de página, por lo tanto tenían el siguiente formato: "*Alquilar en Almería (321)*". Con una expresión regular se ha obtenido la provincia.
- Se ha limpiado la columna Precio. Para quitarle la parte de €/mes y poder así obtener un entero.
- Se ha procesado la columna de Detalles para separarla en los campos correspondientes: 'n_habitaciones', 'n_baños', 'm_cuadrados', 'planta'.
- Para poder comparar los precios en el mapa se hace un filtrado de los datos. De esta forma se puede distinguir las zonas más baratas y caras.

4. Análisis de los datos

- Se han seleccionado las siguientes columnas para el análisis: 'n_habitaciones', 'n_baños', 'm_cuadrados', 'precio'.
- Se verifica que siguen una distribución normal con el test de Shapiro-Wilk.
- Se hace una regresión lineal entre los metros cuadrados de una vivienda y su precio.



- Se calcula la matriz de correlación. Hay una correlación del 0.56 entre los metros cuadrados de una vivienda y su precio.

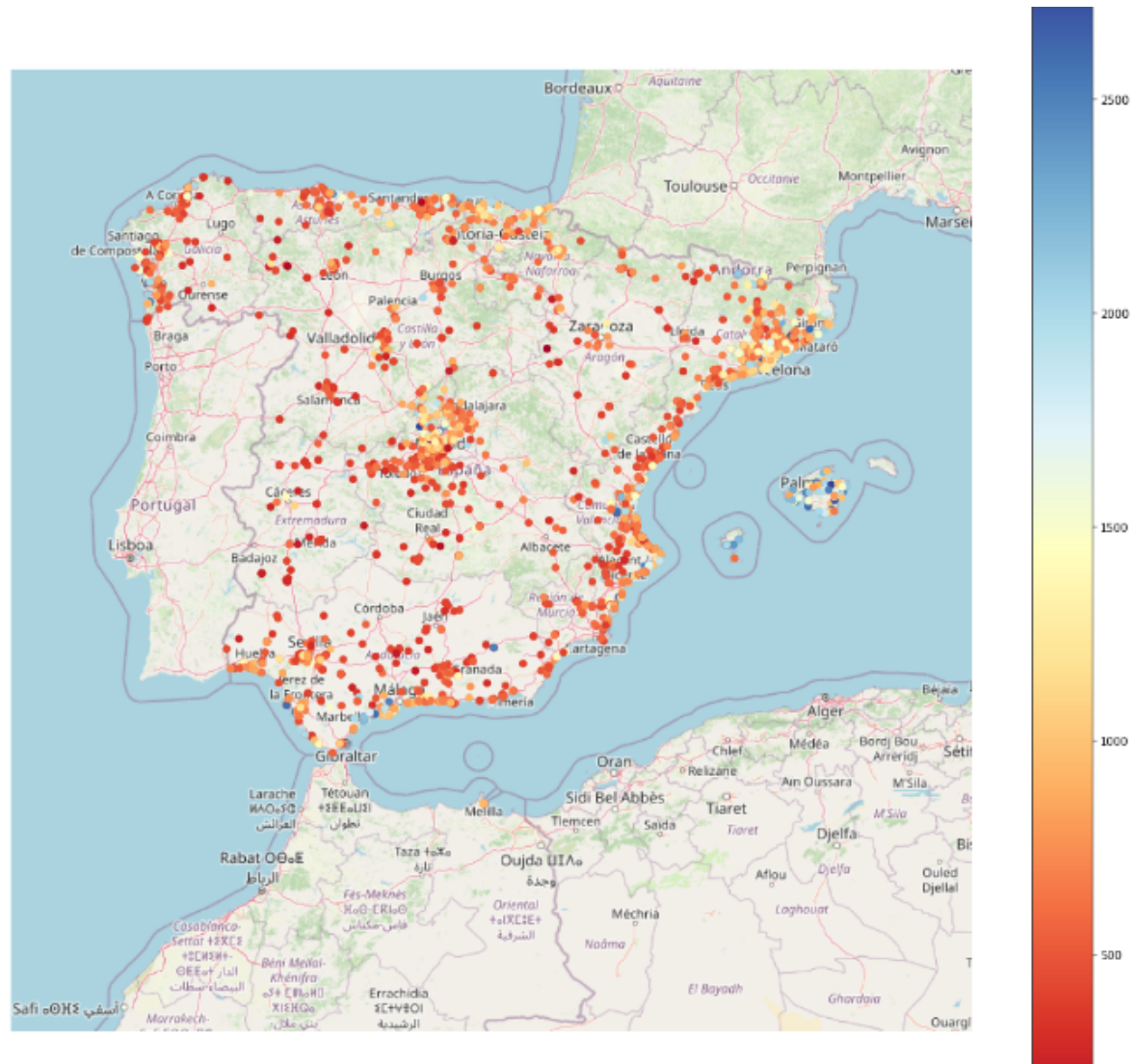


- Se hace un contraste de hipótesis para ver si en una provincia las casas son más caras que en otra con una confianza del 0.05.

Se obtiene un Valor p: 0.00025201564818417963 por lo que se rechaza la hipótesis nula. Hay evidencia estadística de que las casas de Vizcaya son más caras que las de Álava.

5. Representación de los resultados

Se ha realizado una visualización gráfica en un mapa para visualizar los apartamentos en alquiler en España.



Por un lado se puede observar que zonas tienen más oferta y por otro se puede averiguar a simple vista que zonas son más caras.

6. Resolución del problema

Una limpieza y análisis de datos del precio de la vivienda en España, donde se grafican en un mapa los alquileres disponibles con colores que indican su precio puede resolver varios problemas y proporcionar información valiosa para diferentes partes interesadas.

También es útil para detectar zonas donde la vivienda es más baratas. Este análisis de datos puede revelar patrones geográficos en los precios de alquiler de viviendas en diferentes regiones de España. Esto puede ayudar a identificar áreas donde los alquileres son más altos o más bajos, lo que puede ser útil para inquilinos, propietarios y agentes inmobiliarios.

En un futuro cuando se añadan datos históricos, al analizar los datos de precios de alquiler, se pueden identificar tendencias a lo largo del tiempo. Esto puede ayudar a predecir cambios futuros en los precios de alquiler y proporcionar información valiosa.

7. Código

El código está disponible en el siguiente repositorio:
<https://github.com/markositu/realEstateWebScrapper>

8. Vídeo

Esta disponible en el siguiente enlace:
<https://drive.google.com/file/d/1okJBOFirofegRTMpILW6qS4yJDJ0Of8e/view?usp=sharing>