

TFM - DATA SCIENCE
KSCHOOL

HOUSE PRICE PREDICTOR

CARMEN MUÑOZ

DAR RESPUESTA A

- ¿Cuáles son las viviendas más rentables de Churriana?
- ¿Qué precio estimado debería tener mi vivienda si quiero ponerla a la venta?

¿CÓMO?

Modelo predictivo para predecir el precio de la vivienda a partir de técnicas de **aprendizaje supervisado**.

Variable objetivo:
Precio de las viviendas

Zona de estudio: Churriana,
Málaga.

“

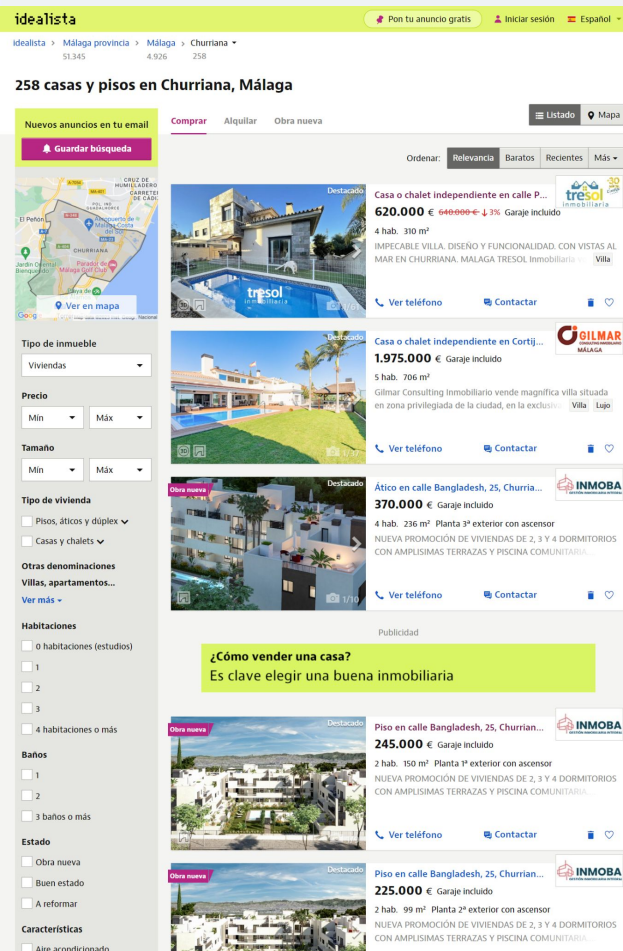
PASOS LLEVADOS A CABO

- Extracción de datos
- Análisis EDA y Limpieza de datos
- Preparación de datos para el modelo
- Evaluación de modelos
- Aplicación del modelo final
- Visualización Front-end
- Conclusiones y mejoras

EXTRACCIÓN DE DATOS

Portal Idealista
Web Scraping

Zona: Churriana



idealista

¡Pon tu anuncio gratis

Inicio sesión

Español

idealista > Málaga provincia > Málaga > Churriana

51.345 4.926 258

258 casas y pisos en Churriana, Málaga

Nuevos anuncios en tu email

Guardar búsqueda

Comprar Alquilar Obra nueva

Ordenar: Relevancia Baratos Recientes Más

Ver en mapa

Tipo de inmueble

Viviendas

Precio

Min Máx

Tamaño

Min Máx

Tipo de vivienda

☐ Pisos, áticos y dúplex

☐ Casas y chalets

Otras denominaciones

Villas, apartamentos...

Ver más

Habitaciones

☐ 0 habitaciones (estudios)

☐ 1

☐ 2

☐ 3

☐ 4 habitaciones o más

Baños

☐ 1

☐ 2

☐ 3 baños o más

Estado

☐ Obra nueva

☐ Buen estado

☐ A reformar

Características

☐ Aire acondicionado

Destacado

Casa o chalet independiente en calle P...
620.000 € 640.000 € ↓ 3% Garaje incluido
4 hab. 310 m²
IMPECABLE VILLA. DISEÑO Y FUNCIONALIDAD. CON VISTAS AL MAR EN CHURRIANA, MALAGA TRESOL Inmobiliaria Villa

Ver teléfono Contactar

Destacado

Casa o chalet independiente en Cortij...
1.975.000 € Garaje incluido
5 hab. 706 m²
Gilmar Consulting Inmobiliario vende magnífica villa situada en zona privilegiada de la ciudad, en la exclusiva Villa Lugo

Ver teléfono Contactar

Destacado

Ático en calle Bangladesh, 25, Churria...
370.000 € Garaje incluido
4 hab. 236 m² Planta 3ª exterior con ascensor
NUEVA PROMOCIÓN DE VIVIENDAS DE 2, 3 Y 4 DORMITORIOS CON AMPLÍSIMAS TERRAZAS Y PISCINA COMUNITARIA.

Ver teléfono Contactar

Publicidad

¿Cómo vender una casa?
Es clave elegir una buena inmobiliaria

Destacado

Piso en calle Bangladesh, 25, Churrian...
245.000 € Garaje incluido
2 hab. 150 m² Planta 1ª exterior con ascensor
NUEVA PROMOCIÓN DE VIVIENDAS DE 2, 3 Y 4 DORMITORIOS CON AMPLÍSIMAS TERRAZAS Y PISCINA COMUNITARIA.

Ver teléfono Contactar

Destacado

Piso en calle Bangladesh, 25, Churrian...
225.000 € Garaje incluido
2 hab. 99 m² Planta 2ª exterior con ascensor
NUEVA PROMOCIÓN DE VIVIENDAS DE 2, 3 Y 4 DORMITORIOS CON AMPLÍSIMAS TERRAZAS Y PISCINA COMUNITARIA.

Características Extraídas

- Título
- Localización
- Latitud
- Longitud
- Price (variable a predecir)
- Certificado energético
- Número de habitaciones
- Número de baños
- Si dispone o no de jardín
- Si dispone o no de terraza
- Si dispone o no de parking
- Si dispone o no de piscina
- Si dispone o no de ascensor
- Metros cuadrados construidos
- Si es de nueva construcción
- Si necesita ser renovado
- Si está en buenas condiciones

De un total de 342 viviendas localizadas se consigue extraer información de 186 viviendas con 17 variables. Un 55% del total.

Análisis EDA y Limpieza de datos

➔ Pasamos los Nan a valor 0

```
Nan_numbers = properties_churriana.isnull().sum(axis = 0).sum()
```

```
print(f"Se encuentran {Nan_numbers} Nan numbers")
```

Se encuentran 193 Nan numbers

has_garden	23
has_terrace	17
has_parking	18
has_swimmingpool	29
has_lift	106

➔ Eliminamos columnas innecesarias.

➔ Análisis estadístico y descriptivo de los datos

Grandes diferencias entre los valores mínimos y máximos de algunas de las características del dataset.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 187 entries, 0 to 186
Data columns (total 18 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   titulo                187 non-null   object  
1   localizacion          187 non-null   object  
2   latitude              187 non-null   float64  
3   longitude             187 non-null   float64  
4   price                 187 non-null   int64  
5   energy_certification  187 non-null   object  
6   basic_characteristics 187 non-null   object  
7   room_number           187 non-null   int64  
8   bath_number           187 non-null   int64  
9   has_garden            164 non-null   float64  
10  has_terrace           170 non-null   float64  
11  has_parking           169 non-null   float64  
12  has_swimmingpool      158 non-null   float64  
13  has_lift              81 non-null    float64  
14  constructed_area      187 non-null   int64  
15  is_new_development    187 non-null   int64  
16  is_needs_renovating   187 non-null   int64  
17  is_goog_condition     187 non-null   int64  
dtypes: float64(7), int64(7), object(4)
memory usage: 27.8+ KB
```

	count	mean	std	min	25%	50%	75%	max
latitude	187.0	36.692094	0.113276	36.610730	36.659499	36.663815	36.671115	3.714505e+01
longitude	187.0	-4.449519	0.202287	-4.686328	-4.506868	-4.501821	-4.490153	-3.641632e+00
price	187.0	482993.689840	512781.940307	54900.000000	249900.000000	299900.000000	567000.000000	5.000000e+06
room_number	187.0	3.625668	1.629362	1.000000	3.000000	3.000000	4.000000	1.400000e+01
bath_number	187.0	2.556150	1.395385	1.000000	2.000000	2.000000	3.000000	8.000000e+00
has_garden	164.0	0.695122	0.461766	0.000000	0.000000	1.000000	1.000000	1.000000e+00
has_terrace	170.0	0.947059	0.224578	0.000000	1.000000	1.000000	1.000000	1.000000e+00
has_parking	169.0	0.899408	0.301681	0.000000	1.000000	1.000000	1.000000	1.000000e+00
has_swimmingpool	158.0	0.867089	0.340558	0.000000	1.000000	1.000000	1.000000	1.000000e+00
has_lift	81.0	0.975309	0.156150	0.000000	1.000000	1.000000	1.000000	1.000000e+00
constructed_area	187.0	243.433155	270.325987	50.000000	108.500000	152.000000	256.000000	2.500000e+03
is_new_development	187.0	0.229947	0.421928	0.000000	0.000000	0.000000	0.000000	1.000000e+00
is_needs_renovating	187.0	0.053476	0.225585	0.000000	0.000000	0.000000	0.000000	1.000000e+00
is_goog_condition	187.0	0.716578	0.451870	0.000000	0.000000	1.000000	1.000000	1.000000e+00

HeatMap

Correlaciones entre las variables

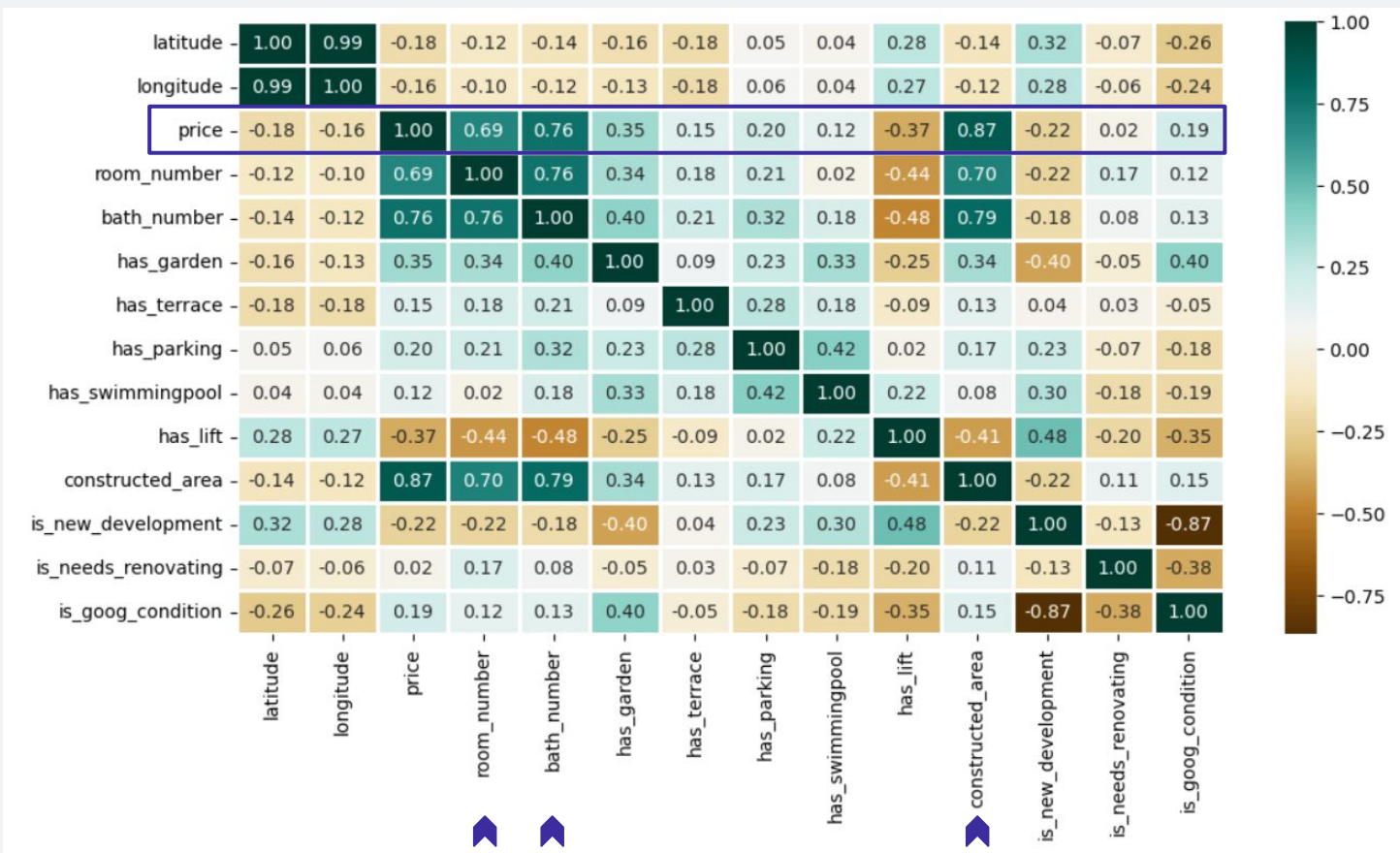
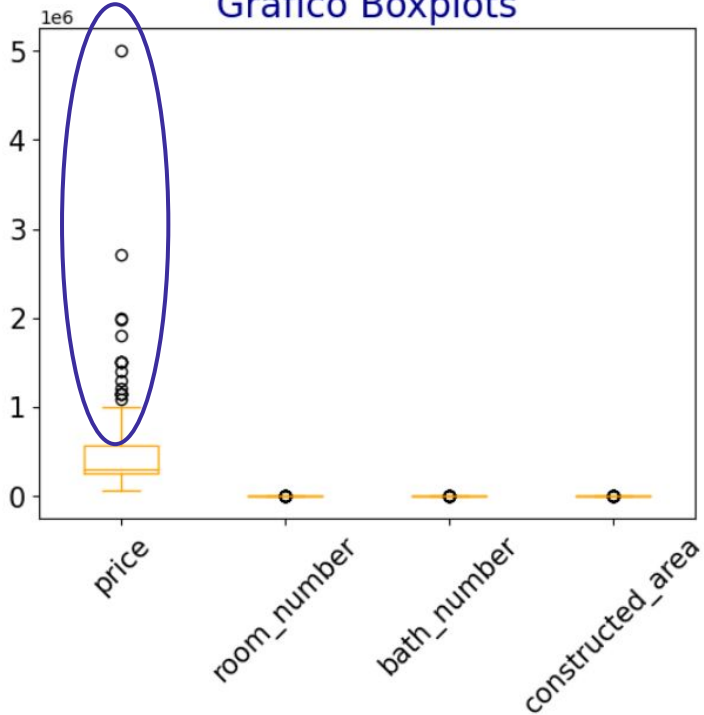
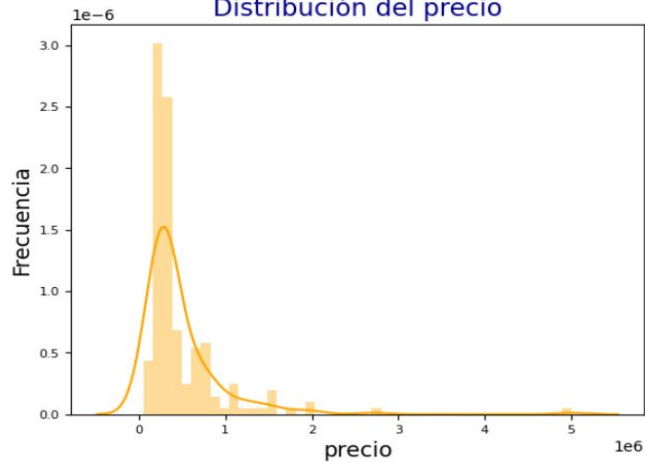


Gráfico Boxplots

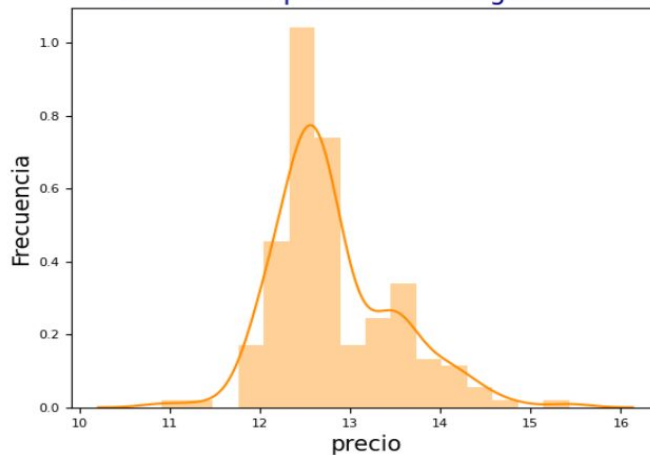


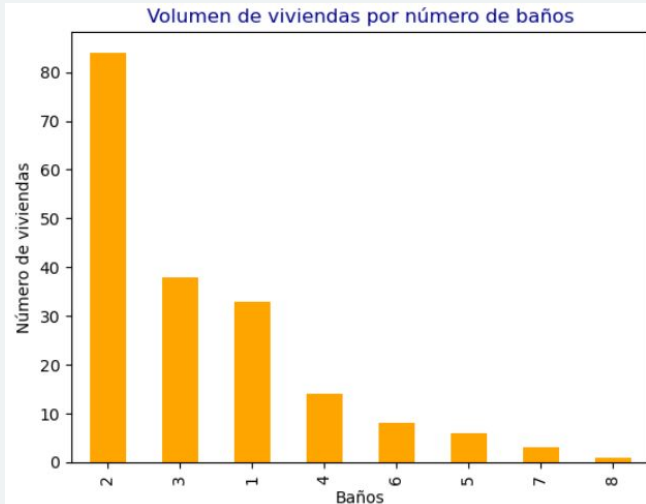
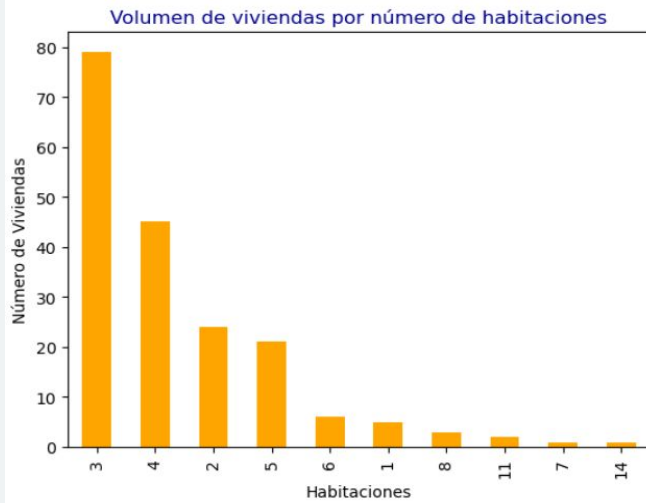
Distribución del precio



Curva sesgada a la derecha, con skew positivo.

Distribución precio escala logarítmica





Se afirma la existencia de Outliers, pero no corresponden con valores atípicos sino que corresponde con valores extremos ya que son viviendas mucho más grandes en m2 y con mayor volumen de habitaciones (podría corresponder con una mansión o cortijo).

Exclusión de Outliers

Se calcula el valor intercuartílico **IQR** para poder delimitar el **extremo inferior** y **superior** en función de la variable precio.

```
q1 = np.percentile(data_clean["price"], 25)
q3 = np.percentile(data_clean["price"], 75)
```

```
iqr = q3 - q1
```

```
# calculamos los límites inferior y superior del precio
```

```
upper_limit = q3 + 1.5 * iqr
```

```
lower_limit = q1 - 1.5 * iqr
```

```
print(f"Todas las viviendas con precios superiores a: {upper_limit}€ serán excluidas")
```

```
print(f"Todas las viviendas con precios inferiores a: {lower_limit}€ serán excluidas")
```

```
Todas las viviendas con precios superiores a: 1042650.0€ serán excluidas
```

```
Todas las viviendas con precios inferiores a: -225750.0€ serán excluidas
```

```
data_outliers = data_clean[data_clean["price"] > upper_limit].shape[0]
```

```
print(f"Excluiríamos {data_outliers} viviendas")
```

```
Excluiríamos 17 viviendas
```

Preparación datos para el modelo

Transformación de variables categóricas a numéricas usando **LabelEncoder()**

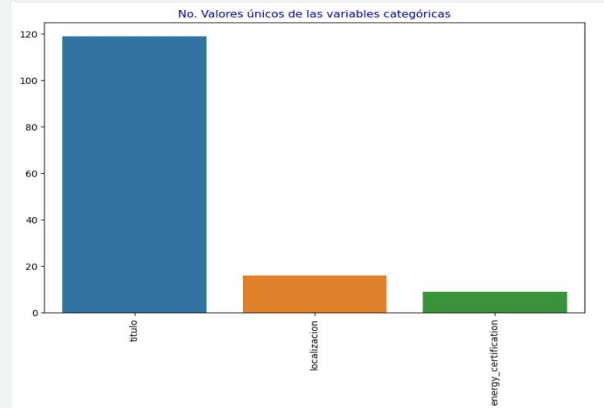
#tipo de variables

```
obj = (data_clean.dtypes == 'object')
object_cols = list(obj[obj].index)
print("Categorical variables:", len(object_cols))
```

```
int_ = (data_clean.dtypes == 'int')
num_cols = list(int_[int_].index)
print("Integer variables:", len(num_cols))
```

```
fl = (data_clean.dtypes == 'float')
fl_cols = list(fl[fl].index)
print("Float variables:", len(fl_cols))
```

```
Categorical variables: 3
Integer variables: 7
Float variables: 7
```



Evaluación de modelos

Datos sin procesar y escalados

- Regresión Lineal
- Regresión Bridge
- SVM-Support Vector Machine
- XgBoost Regresion
- Gradient Boosting Regressor
- Arbol de Regresión

Datos completos

	Modelo	Detalles	R2 train	R2 test	MAE train	MAE test
16	Árbol de regresión	Sin procesar	0.5774	0.7596	95455.722670	172703.538059
4	SVM linear	Sin procesar	0.7639	0.6973	107439.937805	138592.268673
1	Regresión Lineal	Datos escalados	0.8226	0.5991	111651.180321	241667.780677
17	Árbol de regresión	Datos sscalados	0.5774	0.5314	95455.722670	261597.541502
14	Gradient Boosting Regressor	Sin procesar	0.9958	0.5268	21729.779228	167876.985202

Datos Sin Outliers

	Modelo	Detalles	R2 train	R2 test	MAE train	MAE test
0	Regresión Lineal	Sin procesar/Sin outliers	0.7928	0.7354	61043.839656	75237.052916
12	xgboost Regresion	Sin procesar/Sin outliers	1.0000	0.7013	218.786305	68433.148438
2	Regresión Bridge	Sin procesar/Sin outliers	0.7402	0.7010	69056.248992	75928.876317
14	Gradient Boosting Regressor	Sin procesar/Sin outliers	0.9925	0.6748	12069.541356	66372.415119
4	SVM linear	Sin procesar/Sin outliers	0.6527	0.6331	74947.345882	84178.377877

Modelo aplicado

Nuevos datos de entrada

Aplicamos el modelo para nuevos datos de entrada. En este caso usamos los datos completos de Churriana y le aplicamos el modelo entrenado.

Creamos las siguientes columnas:

- Precio predicho.
- Variación entre precio predicho y real.
- % de rentabilidad.

Además, se usa `pd.concat` para volver a **incluir en el dataset las columnas título y localización**.

Finalmente tenemos una **BBDD con 169 viviendas** a la venta de Churriana con un total de **20 variables**.

```
data_final.info()
```

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 170 entries, 0 to 169
```

```
Data columns (total 20 columns):
```

#	Column	Non-Null Count	Dtype
0	titulo	170 non-null	object
1	localizacion	170 non-null	object
2	latitude	170 non-null	float64
3	longitude	170 non-null	float64
4	price	170 non-null	int64
5	energy_certification	170 non-null	int64
6	room_number	170 non-null	int64
7	bath_number	170 non-null	int64
8	has_garden	170 non-null	float64
9	has_terrace	170 non-null	float64
10	has_parking	170 non-null	float64
11	has_swimmingpool	170 non-null	float64
12	has_lift	170 non-null	float64
13	constructed_area	170 non-null	int64
14	is_new_development	170 non-null	int64
15	is_needs_renovating	170 non-null	int64
16	is_goog_condition	170 non-null	int64
17	Price_predict	170 non-null	float64
18	Var_Prices	170 non-null	float64
19	%_rentabilidad	170 non-null	float64

```
dtypes: float64(10), int64(8), object(2)
```

```
memory usage: 26.7+ KB
```

Visualización Front end

- Buscador de viviendas rentables
- ¿Cuál es el precio de mi vivienda?



BUSCADOR DE VIVIENDAS CON RENTABILIDAD

BARRIO DE CHURRIANA, MÁLAGA

Buscador de viviendas rentables ¿Cuál es el precio de mi vivienda?

BUSCADOR DE VIVIENDAS CON RENTABILIDAD

BARRIO DE CHURRIANA, MÁLAGA

Buscador de viviendas rentables ¿Cuál es el precio de mi vivienda?

Aplicar Filtros

Rentabilidad mínima

-92.48

-92.48

81.95

Precio Máximo

54999

995999

Selecciona el número de habitaciones

1

8

Selecciona el número de baños

1

5

Mínimo Metros cuadrados

50

782

- ☐ Piscina
- ☐ Jardín
- ☐ Ascensor
- ☐ parking
- ☐ Terraza
- ☐ Nueva Construcción
- ☐ Necesita Reforma
- ☐ En Buena Condición

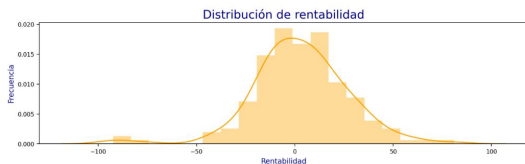
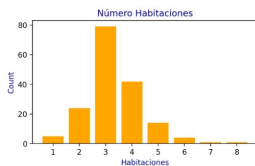
Rentabilidad positiva: 95 Viviendas en Churriana

Número de viviendas: 170

Precio medio: 360,341 €

Area construida media: 181 m2

	localizacion	latitud	longitud	price	enr
Piso en venta en calle Miami, 7	La Carlusela - Los Nidos, Torremolinos	36.6107	-4.5038	175000	
Chalet adosado en venta en RIBERA	San Francisco - El Chorrillo, Las Gubias	37.1329	-3.6706	179999	
Chalet adosado en venta en calle Arempace, 30	Cortijo de Maza Finca Monsalvez El Olivar, Málaga	36.6487	-4.5018	182100	
Casa o chalet independiente en venta en Churriana-El Pizarillo- La Noria-Guadalupe	Churriana, Málaga	36.6656	-4.4723	410000	
Casa o chalet independiente en venta en calle Decano Juan Antonio Rando	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6689	-4.4762	215000	
Piso en venta en calle Decano Joaquín Wuestler, 3	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6659	-4.4731	215000	
Chalet adosado en venta en Churriana-El Pizarillo-La Noria-Guadalupe	Churriana, Málaga	36.6634	-4.4986	205000	
Casa o chalet en venta en Churriana-El Pizarillo-La Noria-Guadalupe	Churriana, Málaga	36.6633	-4.5029	204500	
Chalet adosado en venta en Churriana-El Pizarillo-La Noria-Guadalupe	Churriana, Málaga	36.6634	-4.4989	299990	
Casa o chalet independiente en venta en carretera de Coín, 108	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6618	-4.5131	422500	



Ejemplo seleccionando mostrar las viviendas en venta que necesitan reforma y tienen rentabilidad positiva.

BUSCADOR DE VIVIENDAS CON RENTABILIDAD

BARRIO DE CHURRIANA, MÁLAGA

Buscador de viviendas rentables ¿Cuál es el precio de mi vivienda?

Aplicar Filtros

Rentabilidad mínima

-92.48

81.95

Precio Máximo

54999

995999

Selecciona el número de habitaciones

1

8

Selecciona el número de baños

1

5

Mínimo Metros cuadrados

50

782

- ☐ Piscina
- ☐ Jardín
- ☐ Ascensor
- ☐ parking
- ☐ Terraza
- ☐ Nueva Construcción
- ☒ Necesita Reforma
- ☐ En Buena Condición

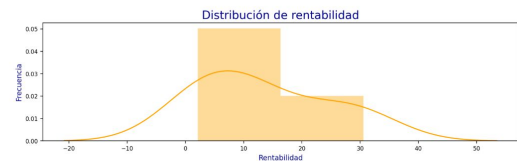
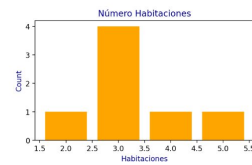
Rentabilidad positiva: 95 Viviendas en Churriana

Número de viviendas: 7

Precio medio: 366,942 €

Area construida media: 255 m2

	localizacion	latitud	longitud	price	enr
Chalet adosado en venta en calle Echol Blanco, 50	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6670	-4.5011	200000	
Chalet adosado en venta en Cortijo de Maza Finca Monsalvez El Olivar	Churriana, Málaga	36.6489	-4.5018	182100	
Casa o chalet en venta en Cortijo de Maza Finca Monsalvez El Olivar	Churriana, Málaga	36.6520	-4.4940	795000	
Chalet pasado en venta en CL ESTACION, 20	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6679	-4.5027	85000	
Chalet adosado en venta en calle Maestro Vicente Livé, 10	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6643	-4.5000	205000	
Casa o chalet independiente en venta en avenida Marquesa Casa Loring, 7	Cortijo de Maza Finca Monsalvez El Olivar, Málaga	36.6508	-4.4907	790000	
Chalet adosado en venta en C. Jovella Morales, 39	Churriana-El Pizarillo-La Noria-Guadalupe, Málaga	36.6713	-4.5088	250500	





BUSCADOR DE VIVIENDAS CON RENTABILIDAD

BARRIO DE CHURRIANA, MÁLAGA

Buscador de viviendas rentables ¿Cuál es el precio de mi vivienda?

Latitud	Longitud	Cert. energética
0.00000 - +	0.00000 - +	desconocido ▼
Habitaciones	Baños	Área construida
0 - +	0 - +	0 - +
<input type="checkbox"/> Terraza	<input type="checkbox"/> Parking	<input type="checkbox"/> Piscina
<input type="checkbox"/> Jardín	<input type="checkbox"/> Nueva construcción	<input type="checkbox"/> Necesita reformas
<input type="checkbox"/> Ascensor		
<input type="checkbox"/> Buenas condiciones		
<button>Calcular</button>		

Vídeo

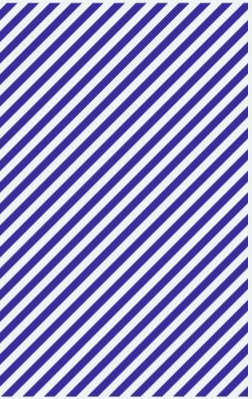
Ejemplo Calculado

Latitud	Longitud	Cert. energética
36.66970 - +	-4.50340 - +	a ▼
Habitaciones	Baños	Área construida
4 - +	3 - +	120 - +
<input checked="" type="checkbox"/> Terraza	<input checked="" type="checkbox"/> Parking	<input checked="" type="checkbox"/> Piscina
<input checked="" type="checkbox"/> Jardín	<input type="checkbox"/> Nueva construcción	<input type="checkbox"/> Necesita reformas
		<input type="checkbox"/> Ascensor
		<input checked="" type="checkbox"/> Buenas condiciones
<button>Calcular</button>		

El precio estimado de tu vivienda es

334,652 €

“



Conclusiones y mejoras

Conclusiones:

- Modelo de uso para el proyecto pero no aplicable a la vida real. Falta de datos, datos insuficientes para un buen entrenamiento de los modelos. Resultados poco precisos y realistas.
- Bloqueo de las consultas en Idealista que paraba la extracción de información.
- A pesar de que los resultados obtenidos no son lo mejor que se podría, estoy muy satisfecha y contenta con el trabajo realizado.

Mejoras:

- Mayor volumen de datos probando otras técnicas de scraping como Selenium.
- Más variables y características de la vivienda.
- Extraer datos de forma automática a diario.
- Sería interesante incluir datos externos de la zona para nutrir el estudio y la predicción.
- Aplicar el modelo a otras zonas para poder hacer comparativas de precios y rentabilidad.
- Automatizar con funciones ciertas partes del proyecto para facilitar su aplicación a nuevos datos y análisis de zonas.
- Incluir las urls de los inmuebles y los datos de contacto del propietario en la parte de visualización para mejorar la usabilidad.
- Incluir un buscador en la app para que el usuario pudiera elegir la zona deseada y obtener datos en tiempo real.

The background is a solid blue color. It features decorative elements: a square in the top-left corner, a square in the bottom-left corner, and a square in the top-right corner, all filled with white diagonal stripes. The remaining areas are solid blue.

¡Muchas Gracias!