

Clasificación de viviendas okupadas

JOEL FERNÁNDEZ ROMÁN

Adquisición y preparación de los datos

Fuente de los datos

pisos.com pisocompartido pisosblog Hipotecas Simulador de luz solar Profesionales [Publica gratis](#) [Entrar](#)


Tu nueva vida a un clic

Comprar Alquilar Obra nueva

Casas y pisos

Casas y pisos en venta en Montilivi-Palau, Girona Capital [Buscar](#)



Publicar tu anuncio ¡no cuesta nada! 

Navega aún más rápido con nuestra App 

Activar Windows
Ve a Configuración para activar Windows.

Scraping incremental

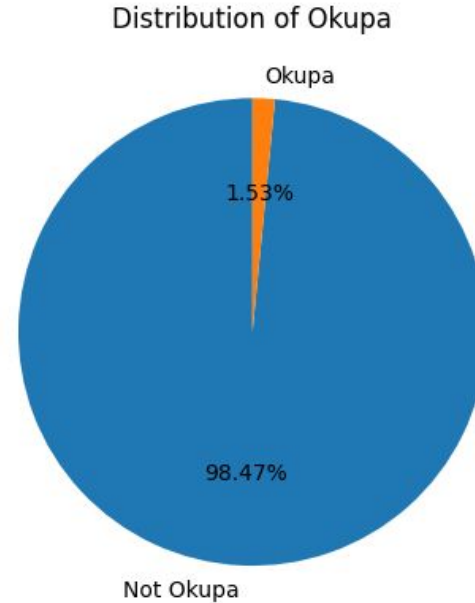


Variable objetivo

Esta propiedad está **ocupada** por terceros. **No se puede visitar**. No apta para compradores que requieran financiación hipotecaria. El comprador será responsable de los trámites y gastos relacionados con la **desocupación** del inmueble.

Oportunidad de inversión: inmueble ocupado. Consulte con nuestro Agente Comercial la información adicional sobre el inmueble y las condiciones especiales de compraventa.

Datos desbalanceados



Dataframe inicial

	area_usable	area_built	rooms	bathrooms	preservation	floor	age	latitude	longitude
0	<NA>	70 m²	2	2.0	<NA>	<NA>	<NA>	36.556251	-4.61515
1	<NA>	120 m²	2	2.0	<NA>	<NA>	<NA>	36.625604	-4.495205
2	850 m²	850 m²	6	8.0	<NA>	<NA>	Entre 20 y 30 años	36.521003	-4.919171
3	<NA>	878 m²	1	1.0	<NA>	<NA>	<NA>	36.7856	-4.36842
4	80 m²	92 m²	3	2.0	<NA>	1ª	<NA>	37.203513	-3.61871
5	220 m²	326 m²	3	5.0	<NA>	<NA>	Menos de 5 años	36.562832	-4.60567
6	54 m²	62 m²	2	1.0	En buen estado	1ª	Entre 10 y 20 años	36.730139	-6.428031
7	83 m²	91 m²	3	2.0	<NA>	5ª	Entre 20 y 30 años	36.545659	-4.642789
8	125 m²	172 m²	3	2.0	<NA>	<NA>	Entre 30 y 50 años	37.218471	-3.604242
9	140 m²	160 m²	4	2.0	<NA>	<NA>	<NA>	37.676938	-3.566194

	count
string[python]	67
boolean	54
Int64	7
Float64	4
object	1
combined_df.shape	
(209284, 134)	

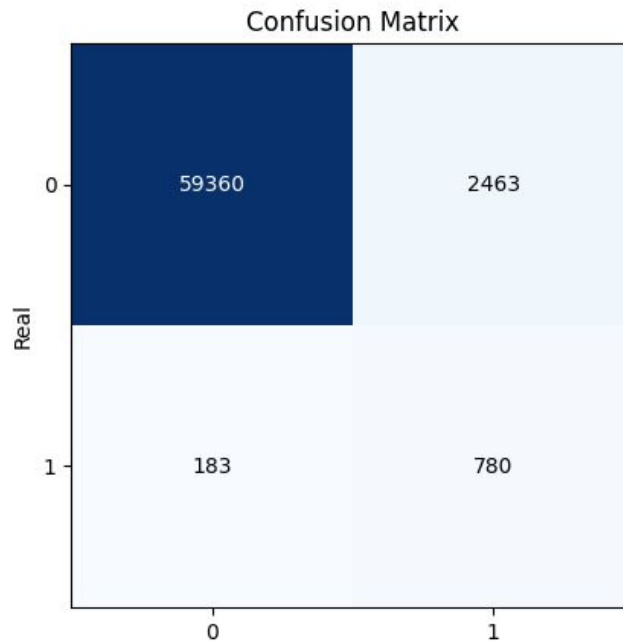
Train sin tratamiento previo

Accuracy: 95.79%

Precision: 24.05%

Recall: 81.00%

F1-score: 37.09%



Data cleaning - iteració 1

- Normalizació de los datos, ex: 60 m² => 60
- Eliminar columnas sin ninguna relación aparente
- Eliminar columnas con muchos NaN

F1-score: 37.09% => F1-score: 39.87%

Data cleaning - iteración 2

- Eliminar columnas constantes
- Eliminar filas duplicadas
- Eliminar columnas con muchos valores únicos
- Eliminar filas con muchos NaN
- Label encoding de Age y Floor

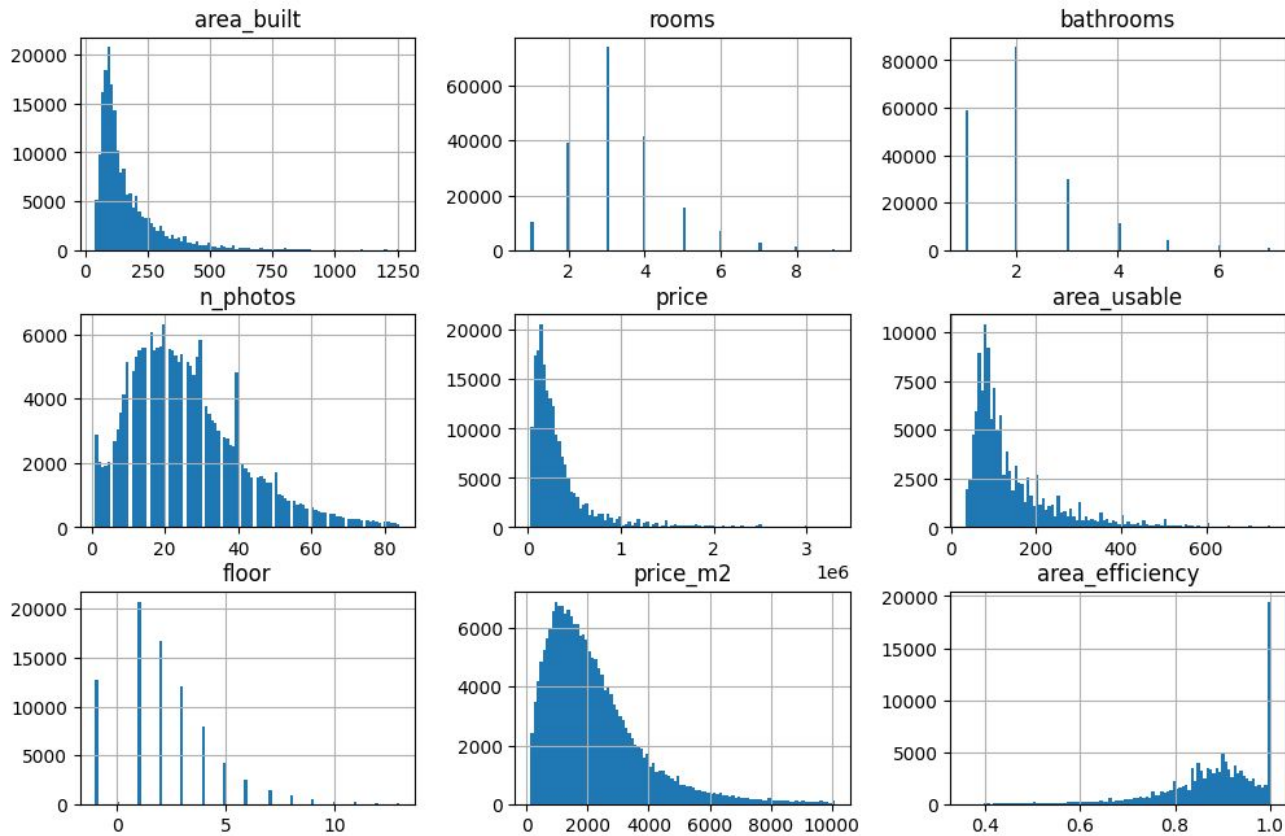
age		floor	
	count		count
Más de 50 años	26364	1ª	20731
Entre 30 y 50 años	21636	2ª	16655
Entre 10 y 20 años	18741	Bajo	12365
Entre 20 y 30 años	15880	3ª	12052
Menos de 5 años	9402	4ª	7929
Entre 5 y 10 años	2021	5ª	4204
		6ª	2472
		7ª	1465
		8ª	895
		9ª	557
		Más de 20	499
		Principal	336
		10ª	331

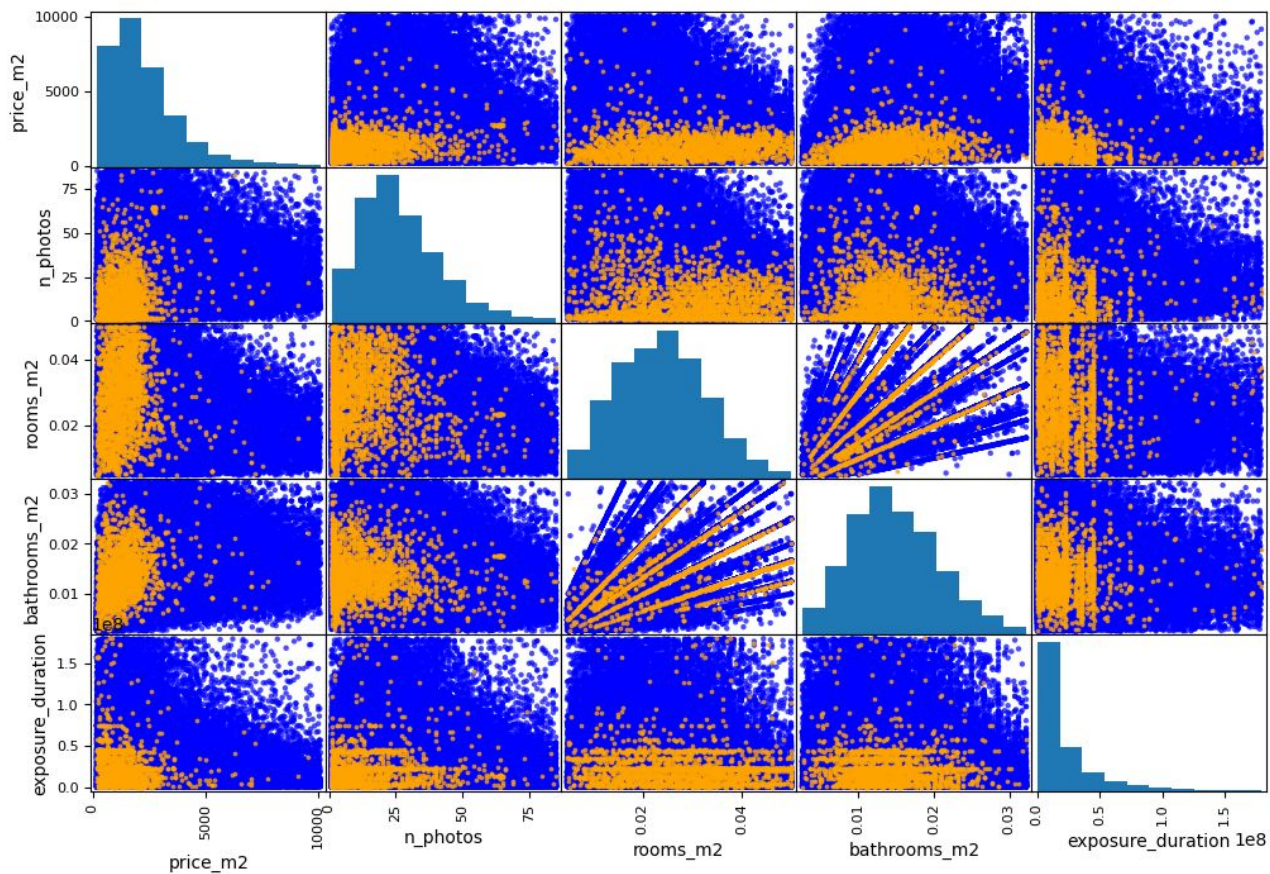
F1-score: 39.87% => F1-score: 54.00%

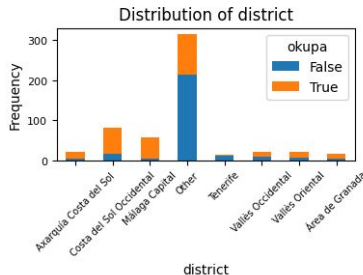
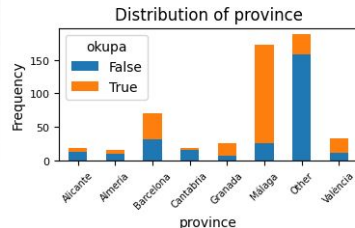
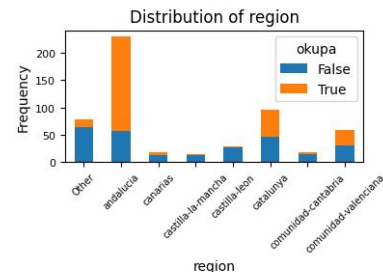
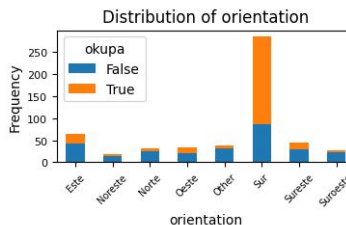
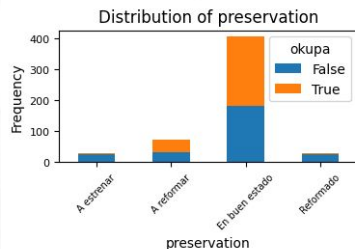
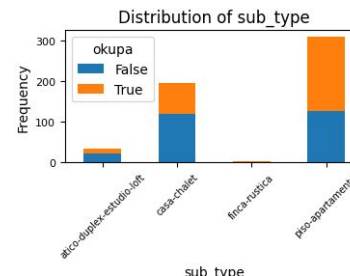
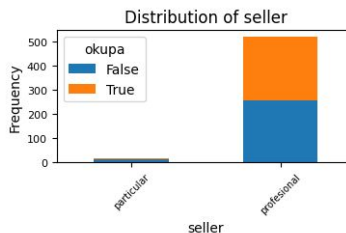
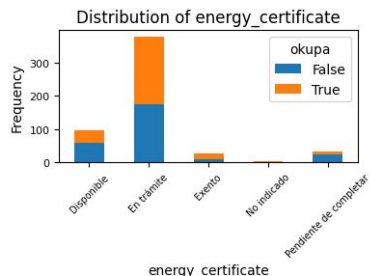
Feature engineering (pre EDA)

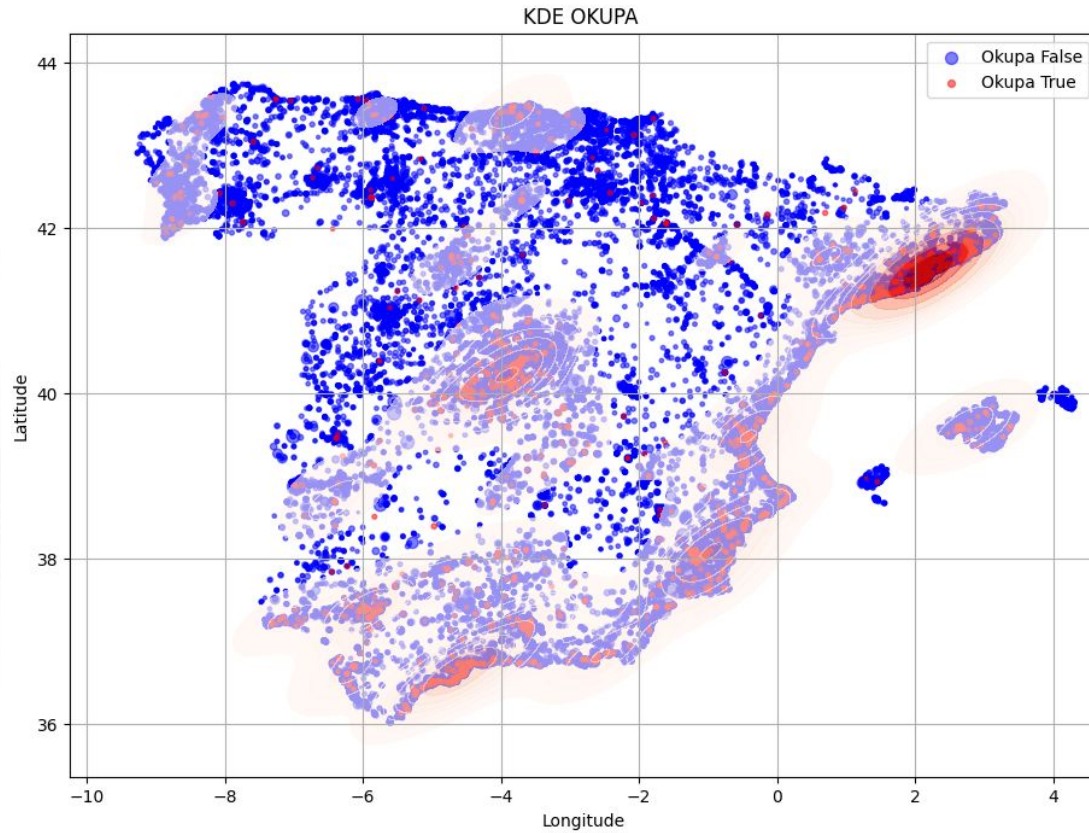
- $\text{price_m2} = \text{price} / \text{area_built}$
- $\text{area_efficiency} = \text{area_usable} / \text{area_built}$
- $\text{rooms_m2} = \text{rooms} / \text{area_built}$
- $\text{bathrooms_m2} = \text{bathrooms} / \text{area_built}$
- $\text{exposure_duration} = \text{fetch_day} - \text{first_publication}$
- ...

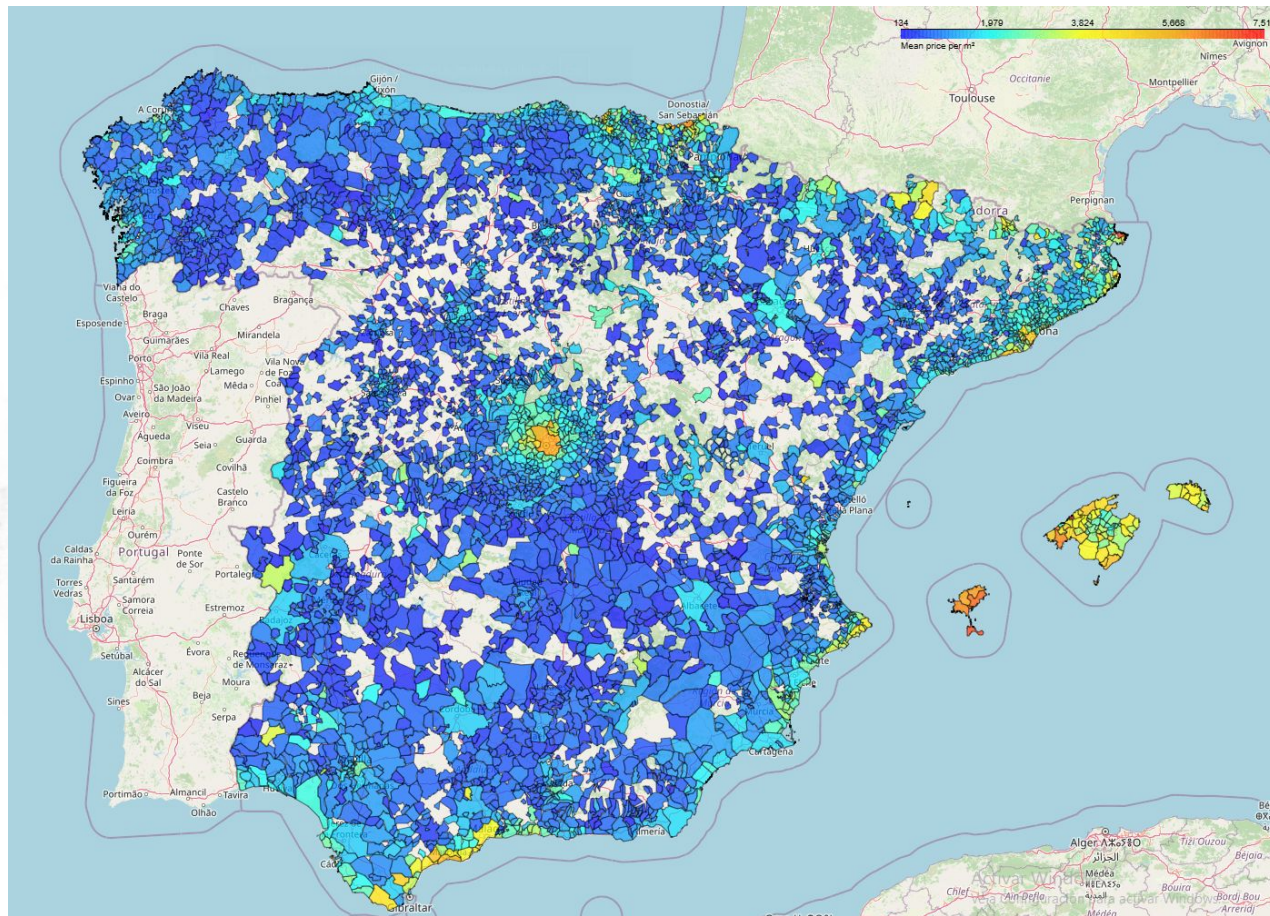
F1-score: 54.00% => F1-score: 54.54%







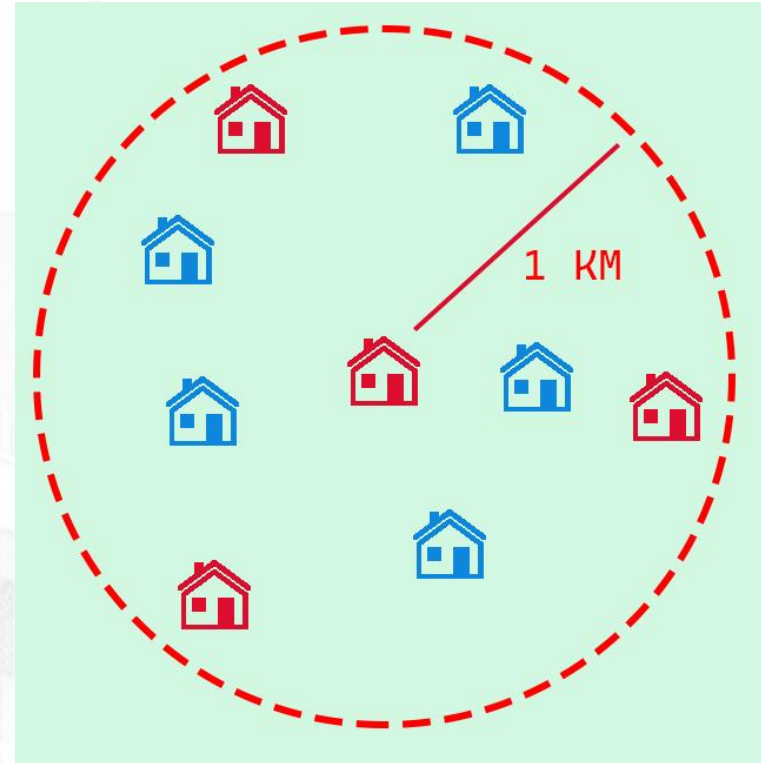




Feature engineering - fase 2

- Densidad de okupación de la zona
- Porcentaje de okupació de la zona
- *okupa_proximity* de la zona
- Precio medio por m² de la zona

F1-score: 54.54% => F1-score: 55.84%



Eliminación de errores y Imputación de datos

- Eliminar filas con errores geoespaciales
- Eliminar columnas no entrenables, ej: “description”, “url”...
- Imputar datos faltantes (Iterative Imputer)

F1-score: 55.84% => F1-score: 58.13%

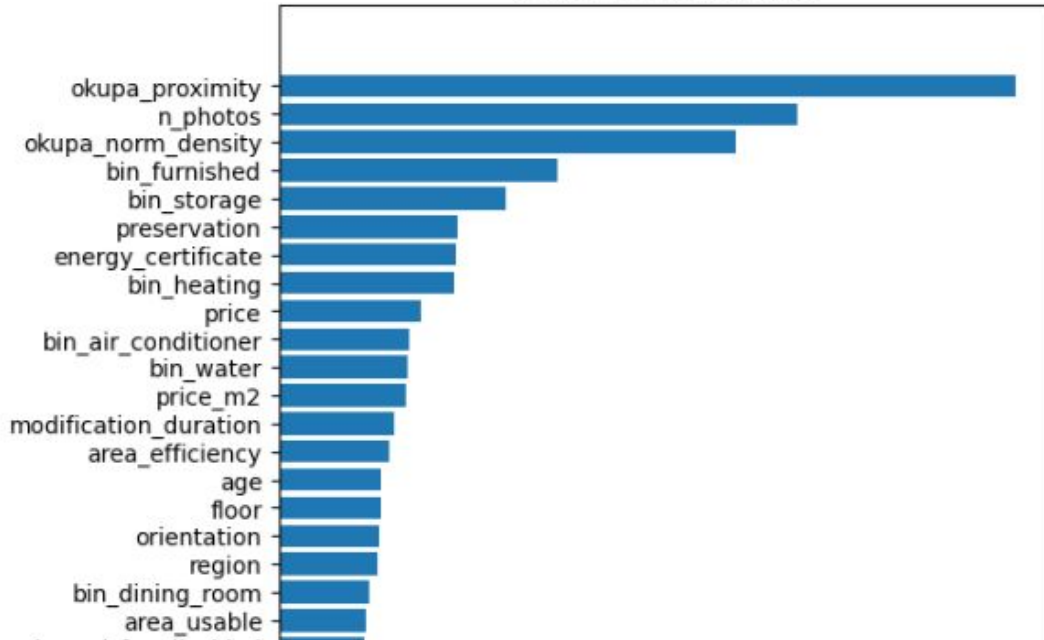
Category binning y label encoding

- 20 valores únicos de orientation
- 30 valores únicos de district
- Label encoding de la variables categóricas

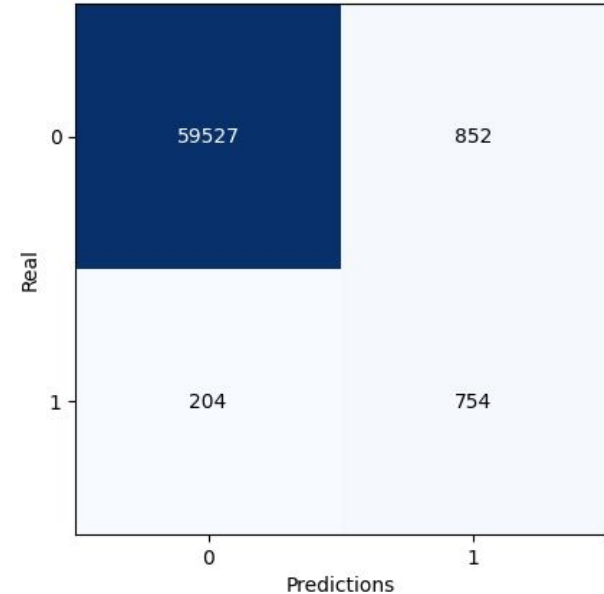
F1-score: 58.13% => F1-score: 58.81%

Resultados finales

Features importances



Confusion Matrix



Resultados finales (f1-score)

