

Predicción de precio de vivienda de alquiler en barcelona

Aprendizaje supervisado

Reinel Mendez Reina

GifHub: reimenr



Objetivo General: Analizar y comprender, las variables que afectan el precio de la vivienda de alquiler en Barcelona.

Objetivos específicos:

1. Recomendar barrios según las preferencias de los usuarios.
2. Dar una predicción de la vivienda en Barcelona según las preferencias del usuario.
3. Entender que variables afectan el precio de la vivienda de alquiler
4. Entender como las dinámicas de la ciudad afectan el precio del alquiler de viviendas en Barcelona

Introducción:

El aprendizaje supervisado es un tipo de aprendizaje automático (machine learning) en el que se entrena un modelo para entender y aprender de la relación entre un conjunto de características (variables independientes) y una variable objetivo (variable dependiente) a partir de datos reales.

Regresión: En problemas de regresión, la variable objetivo es continua. El objetivo es predecir un valor numérico, como el precio de una casa, en función de ciertas características, como el número de habitaciones, el tamaño del terreno, etc.

**¿ Que variables influyen en el precio de la
vivienda de alquiler en Barcelona ?**

1.OBTENER DATOS

**¿ Donde encontrar datos
verídicos y confiables?**



<https://opendata-ajuntament.barcelona.cat/es>

2. LIMPIEZA DE DATOS INICIAL

	Any	Trimestre	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Lloguer_mitja	Preu
0	2023	1	1	Ciutat Vella	1	el Raval	Lloguer mitjà mensual (Euros/mes)	946.8
1	2023	1	1	Ciutat Vella	2	el Barri Gòtic	Lloguer mitjà mensual (Euros/mes)	1279.5
2	2023	1	1	Ciutat Vella	3	la Barceloneta	Lloguer mitjà mensual (Euros/mes)	978.7
3	2023	1	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera	Lloguer mitjà mensual (Euros/mes)	1065.8
4	2023	1	2	Eixample	5	el Fort Pienc	Lloguer mitjà mensual (Euros/mes)	1155.8
...
95	2023	1	5	Sarrià-Sant Gervasi	23	Sarrià	Lloguer mitjà per superfície (Euros/m ² mes)	18.0
96	2023	1	5	Sarrià-Sant Gervasi	24	les Tres Torres	Lloguer mitjà per superfície (Euros/m ² mes)	18.7
97	2023	1	5	Sarrià-Sant Gervasi	25	Sant Gervasi - la Bonanova	Lloguer mitjà per superfície (Euros/m ² mes)	17.5
98	2023	1	5	Sarrià-Sant Gervasi	26	Sant Gervasi - Galvany	Lloguer mitjà per superfície (Euros/m ² mes)	18.1
99	2023	1	5	Sarrià-Sant Gervasi	27	el Putxet i el Farró	Lloguer mitjà per superfície (Euros/m ² mes)	16.9

100 rows x 8 columns

anual

	Any	Codi_Districte	Nom_Districte	Codi_Barri	Nom_Barri	Lloguer_mitja	Precio anual
0	2023	1	Ciutat Vella	1	el Raval	Lloguer mitjà mensual (Euros/mes)	926.10
1	2023	1	Ciutat Vella	2	el Barri Gòtic	Lloguer mitjà mensual (Euros/mes)	1286.70
2	2023	1	Ciutat Vella	3	la Barceloneta	Lloguer mitjà mensual (Euros/mes)	990.45
3	2023	1	Ciutat Vella	4	Sant Pere, Santa Caterina i la Ribera	Lloguer mitjà mensual (Euros/mes)	1086.65
4	2023	2	Eixample	5	el Fort Pienc	Lloguer mitjà mensual (Euros/mes)	1170.60
...
68	2023	10	Sant Martí	69	Diagonal Mar i el Front Marítim del Poblenou	Lloguer mitjà mensual (Euros/mes)	1444.35
69	2023	10	Sant Martí	70	el Besòs i el Maresme	Lloguer mitjà mensual (Euros/mes)	777.25
70	2023	10	Sant Martí	71	Provençals del Poblenou	Lloguer mitjà mensual (Euros/mes)	1187.35
71	2023	10	Sant Martí	72	Sant Martí de Provençals	Lloguer mitjà mensual (Euros/mes)	943.10
72	2023	10	Sant Martí	73	la Verneda i la Pau	Lloguer mitjà mensual (Euros/mes)	1069.75

73 rows × 7 columns

TIPOS DE EQUIPAMIENTOS

Codi_Barri	Tipo	Cant_cultura_ocio
0	18 Parcs i jardins	20
1	18 Centres cívics	13
2	13 Centres cívics	11
3	13 Parcs i jardins	42
4	26 Centres cívics	11
...
237	35 Ocio nocturno	1
238	6 Restaurants	1
239	62 Ocio nocturno	3
240	35 Museos	1
241	15 Ocio nocturno	2

242 rows × 3 columns

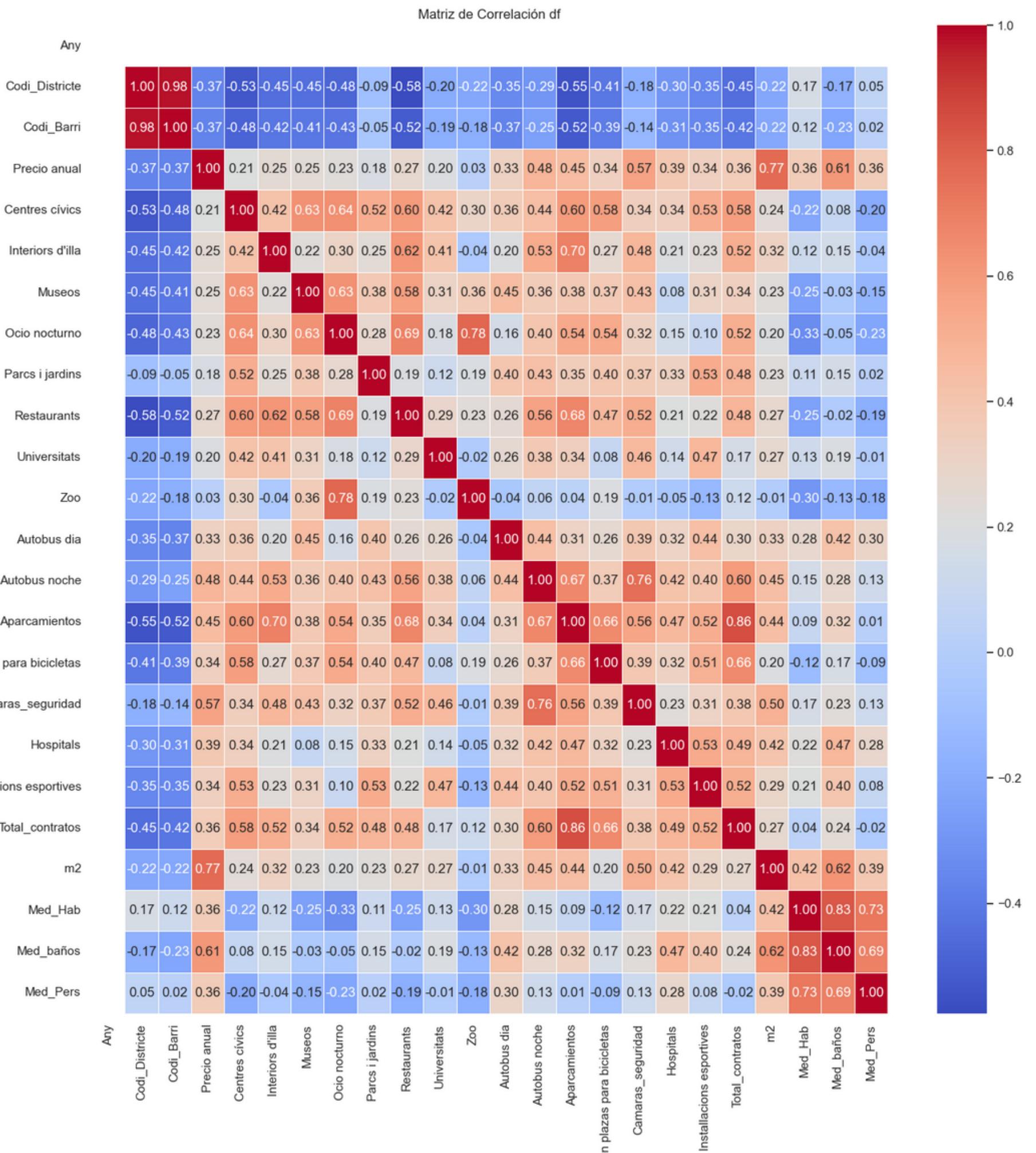
Agrupar y cuantificar

Convertir filas en columnas y la importancia de los ceros

Tipo_Cultural_Ocio	Codi_Barri	Centres cívics	Installacions esportives	Interior d'illa	Museos	Ocio nocturno	Parcs i jardins	Restaurants	Tibidabo	Universitats	Zoo
0	1	43	2	1	5	59	25	4	0	0	0
1	2	24	0	0	14	46	14	5	0	0	0
2	3	17	0	0	1	18	10	8	0	0	0
3	4	28	0	0	10	174	38	5	0	0	1
4	5	14	0	6	2	4	28	3	0	0	0
...
68	69	3	0	0	0	0	31	0	0	0	0
69	70	9	0	0	2	0	42	0	0	0	0
70	71	3	2	0	2	0	26	0	0	0	0
71	72	11	0	0	0	0	27	0	0	0	0
72	73	9	0	0	0	0	46	0	0	0	0

Exploración de datos (Análisis Exploratorio de Datos, EDA)

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 73 entries, 0 to 72
Data columns (total 27 columns):
 #   Column
 ---  -----
 0   Any
 1   Codi_Districte
 2   Nom_Districte
 3   Codi_Barri
 4   Nom_Barri
 5   Lloguer_mitja
 6   Precio anual
 7   Centres cívics
 8   Interiors d'illa
 9   Museos
 10  Ocio nocturno
 11  Parcs i jardins
 12  Restaurants
 13  Universitats
 14  Zoo
 15  Autobus dia
 16  Autobus noche
 17  Aparcamientos
 18  Aparcamientos con plazas para bicicle
 19  Camaras_seguridad
 20  Hospitals
 21  Installacions esportives
 22  Total_contratos
 23  m2
 24  Med_Hab
 25  Med_baños
 26  Med_Pers
dtypes: float64(10), int64(14), object(3)
memory usage: 16.0+ KB
```



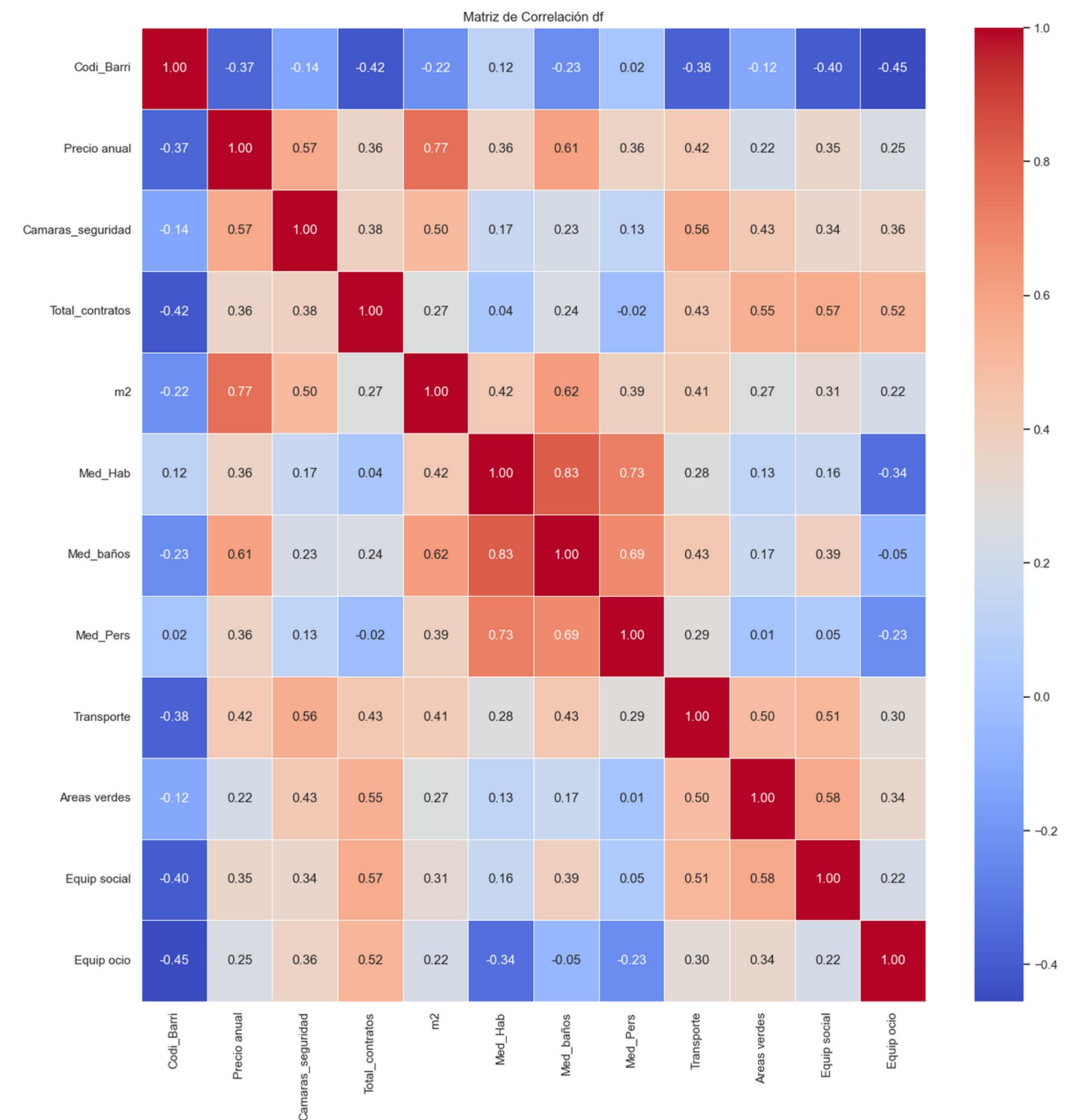
Ingeniería de características

Es el proceso de crear nuevas características o transformar las características existentes en un conjunto de datos para mejorar el rendimiento de un modelo de aprendizaje automático.

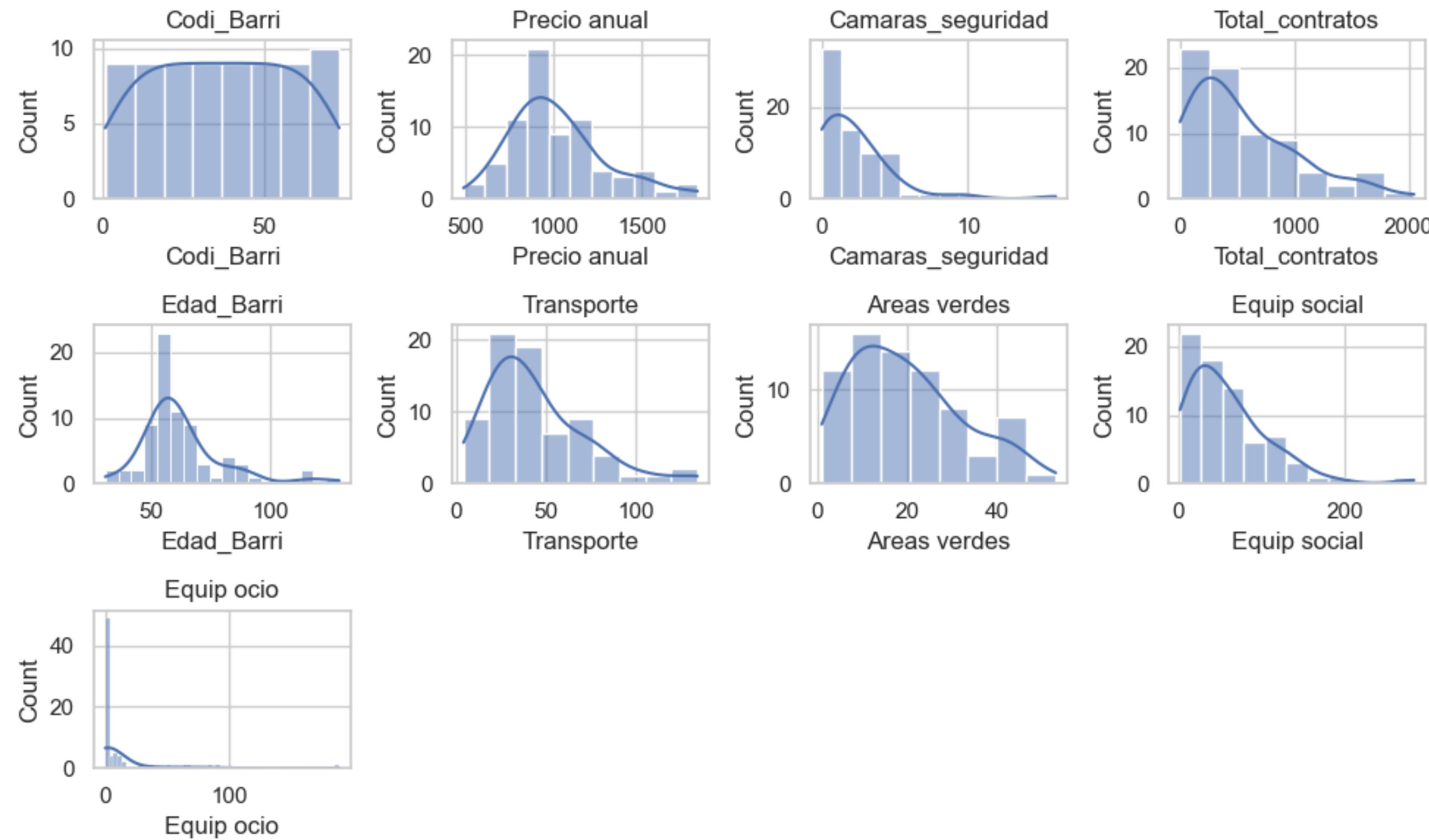
```

<class 'pandas.core.frame.DataFrame'>
Int64Index: 73 entries, 0 to 72
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Codi_Barri      73 non-null    int64  
 1   Precio_anual    73 non-null    float64 
 2   Camaras_seguridad 73 non-null  int32  
 3   Total_contratos 73 non-null    int64  
 4   m2               73 non-null    float64 
 5   Med_Hab          73 non-null    float64 
 6   Med_baños        73 non-null    float64 
 7   Med_Pers          73 non-null    float64 
 8   Transporte        73 non-null    int64  
 9   Areas_verdes    73 non-null    int64  
 10  Equip_social     73 non-null    int32  
 11  Equip_ocio       73 non-null    int64  
dtypes: float64(5), int32(2), int64(5)
memory usage: 6.8 KB

```

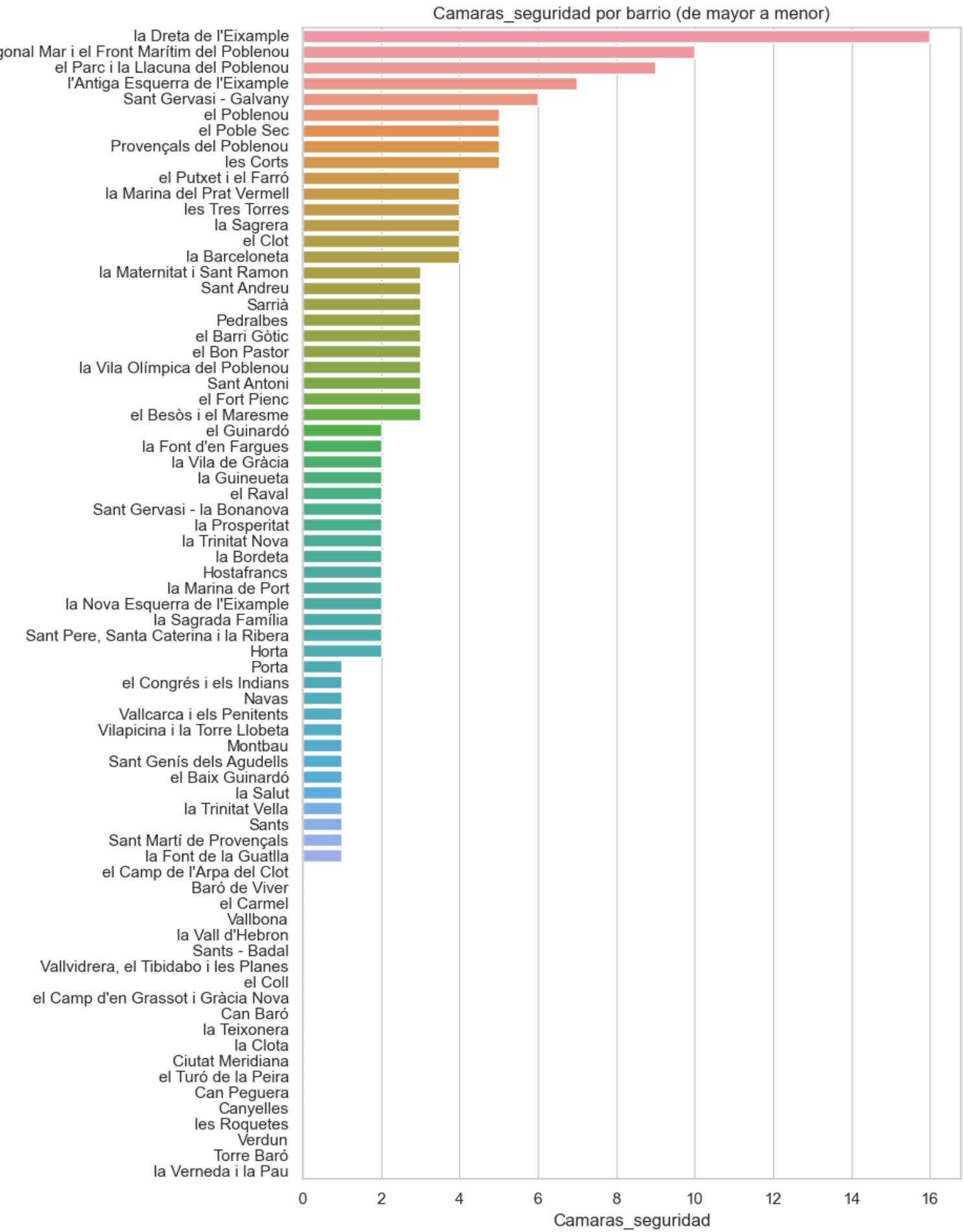
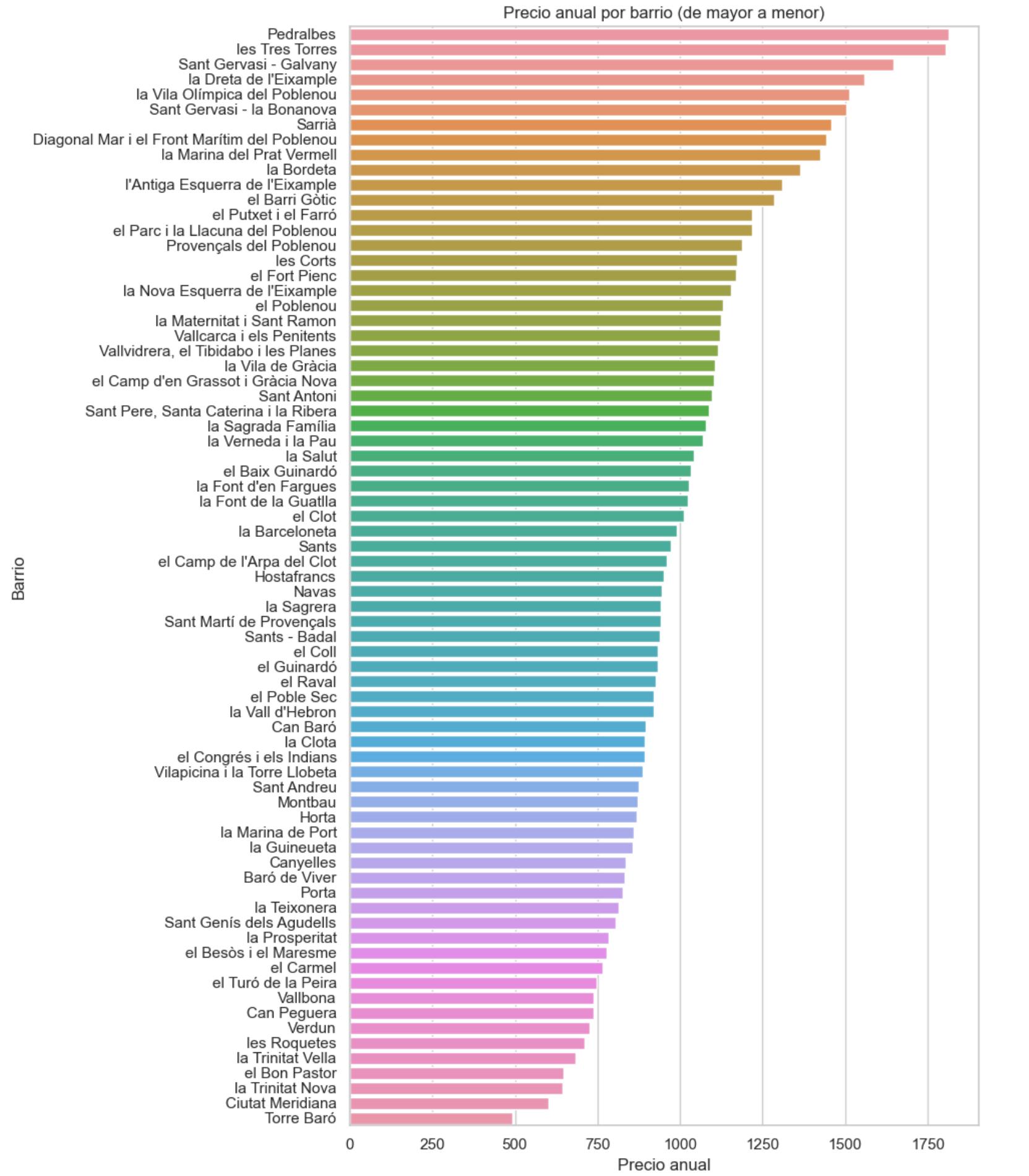


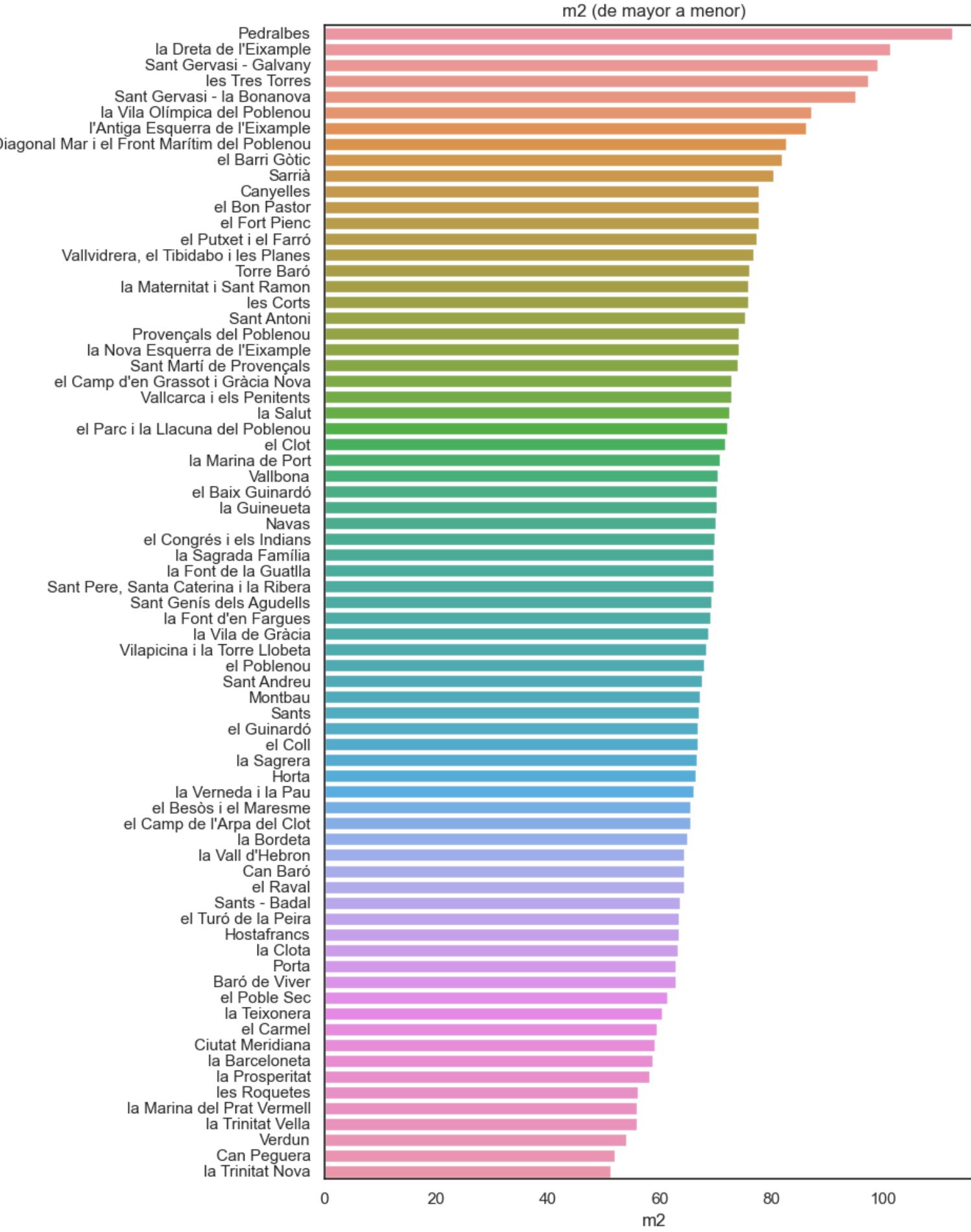
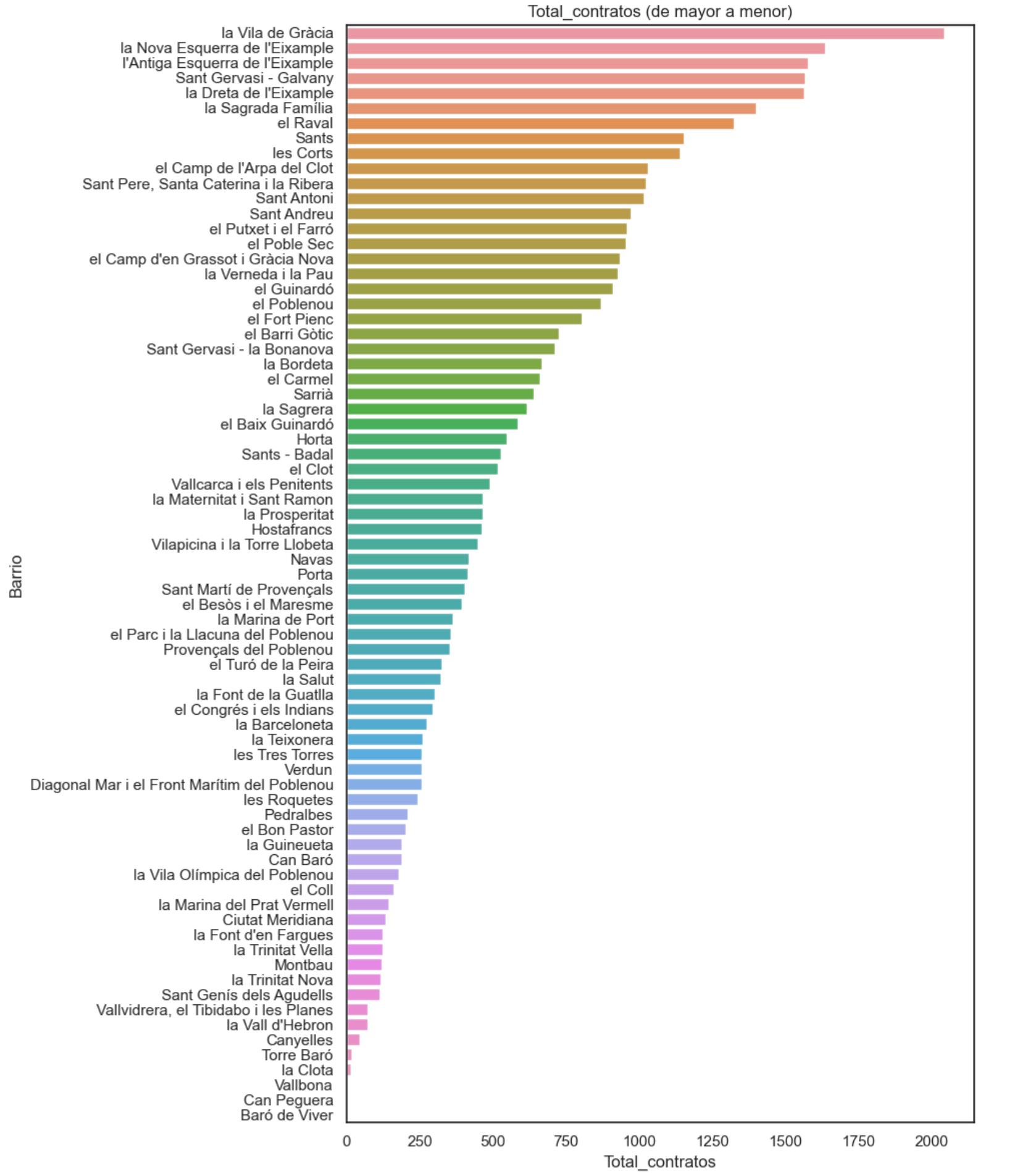
Histogramas de Variables

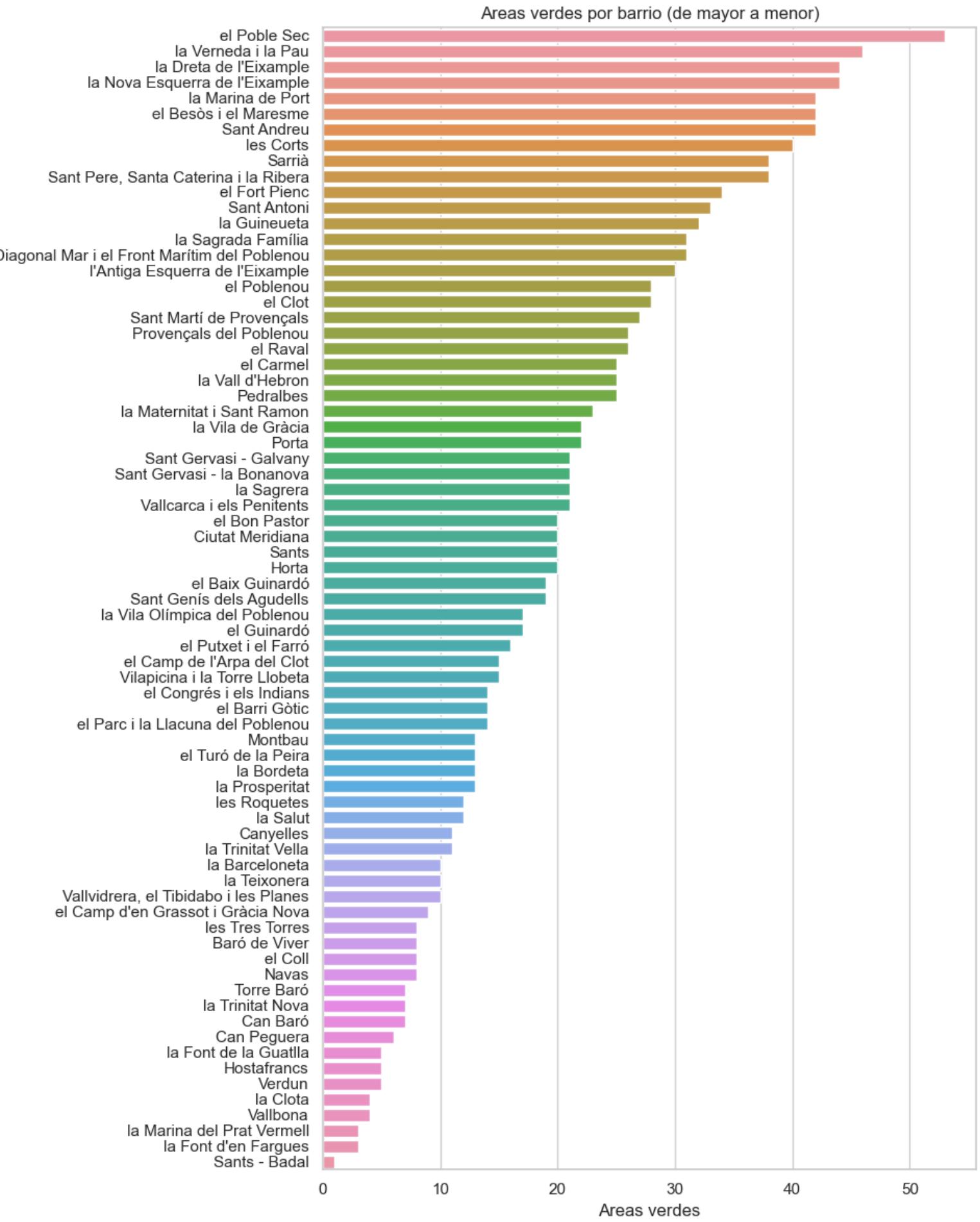
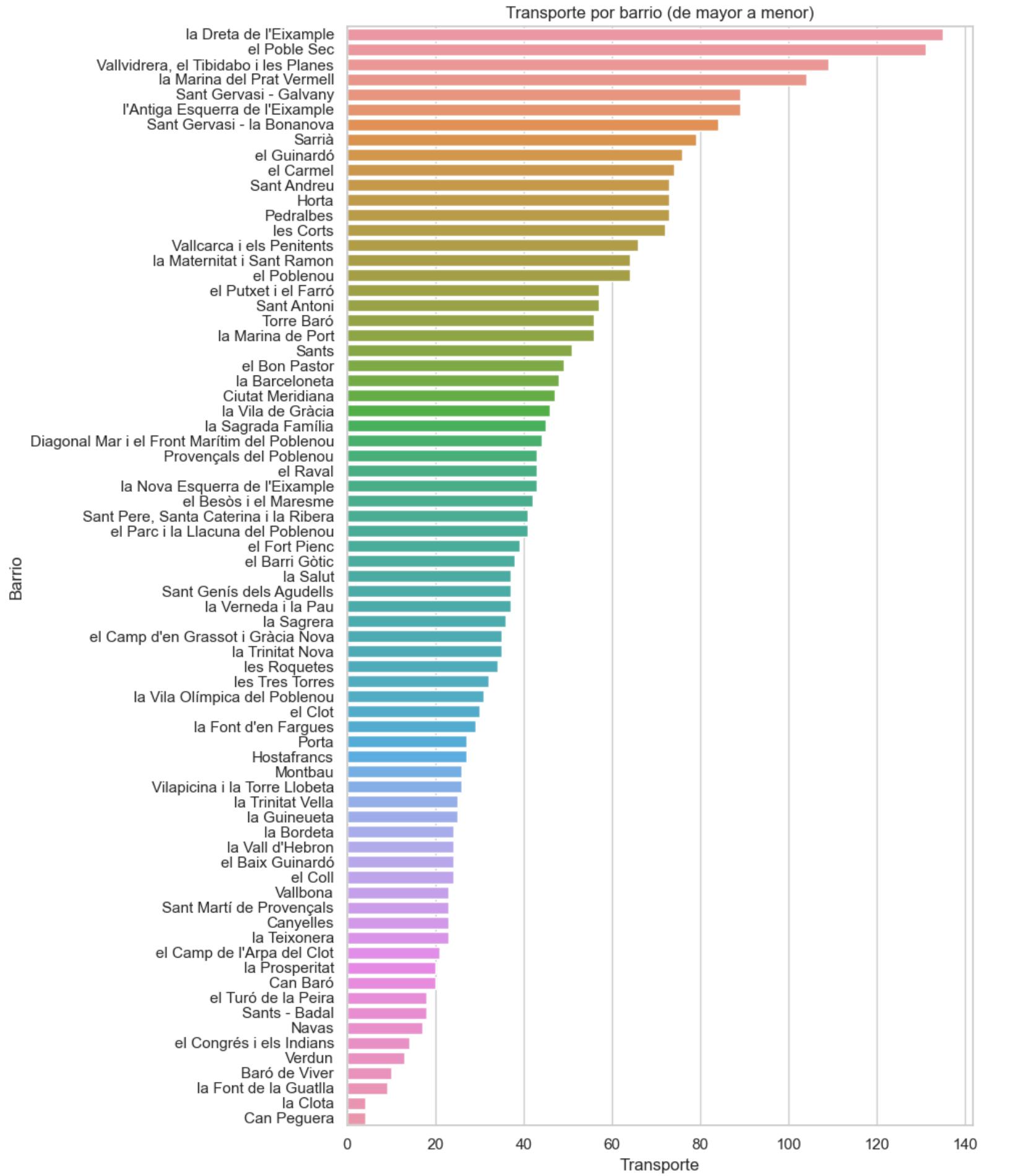


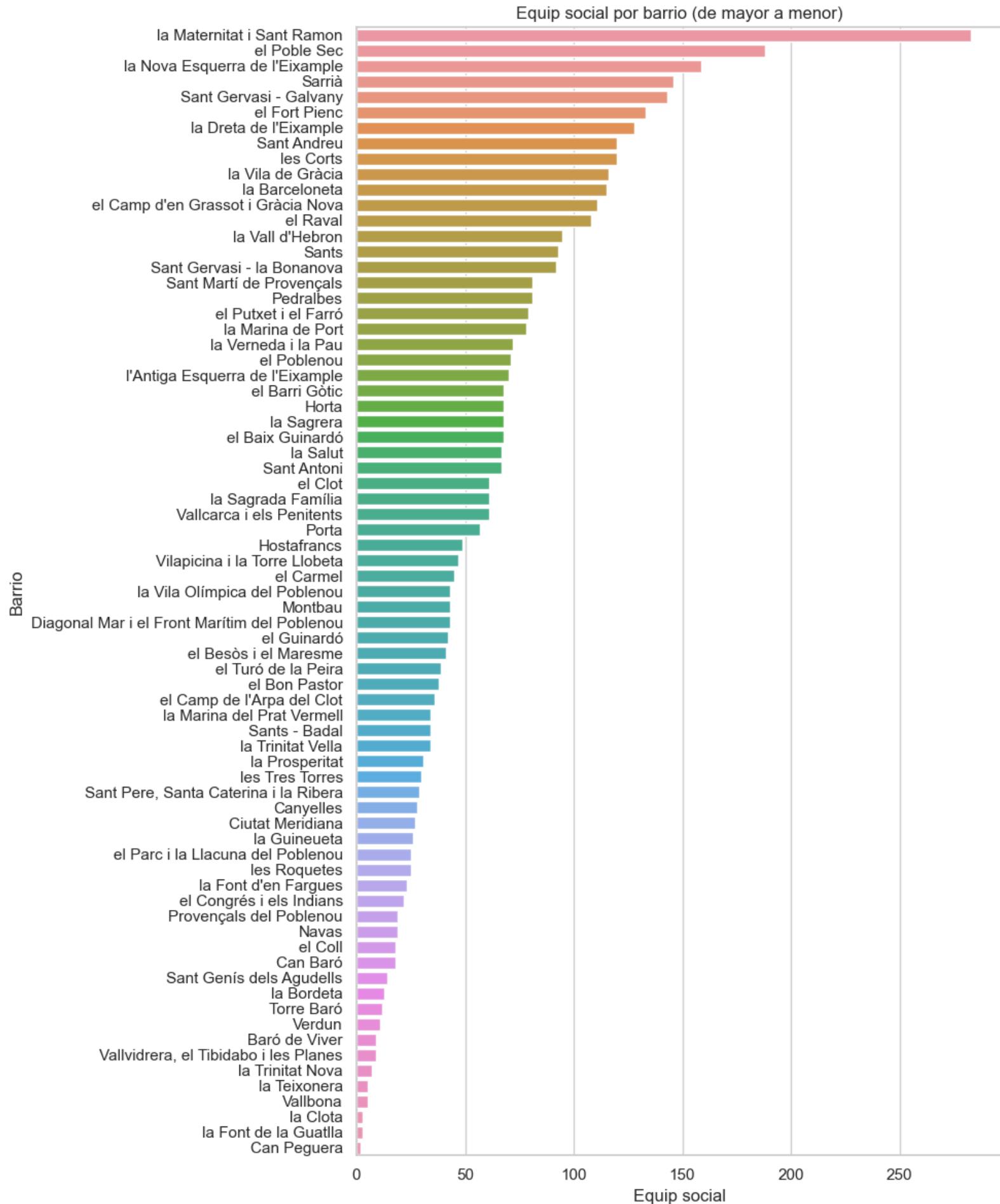
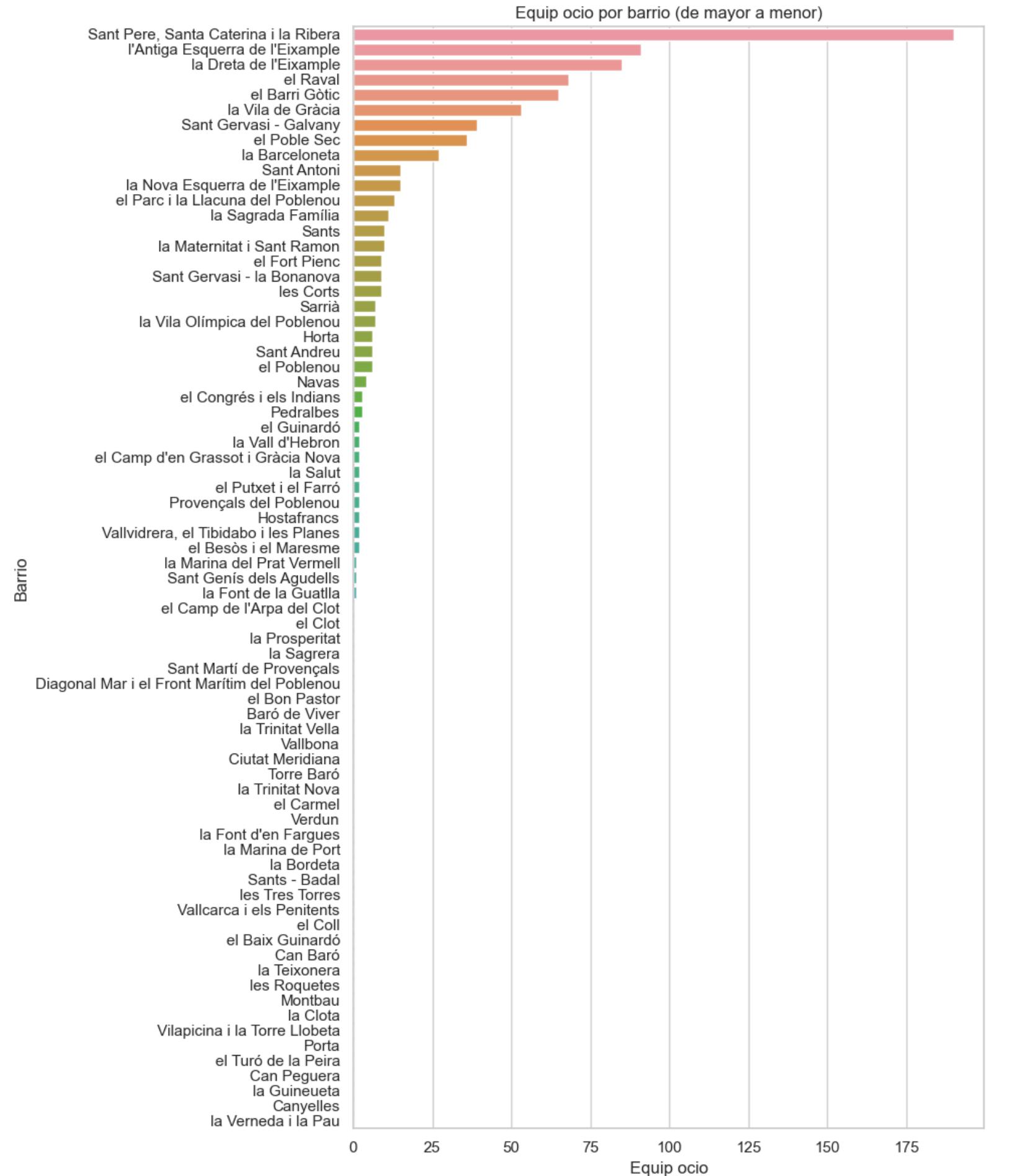
Análisis Estadístico

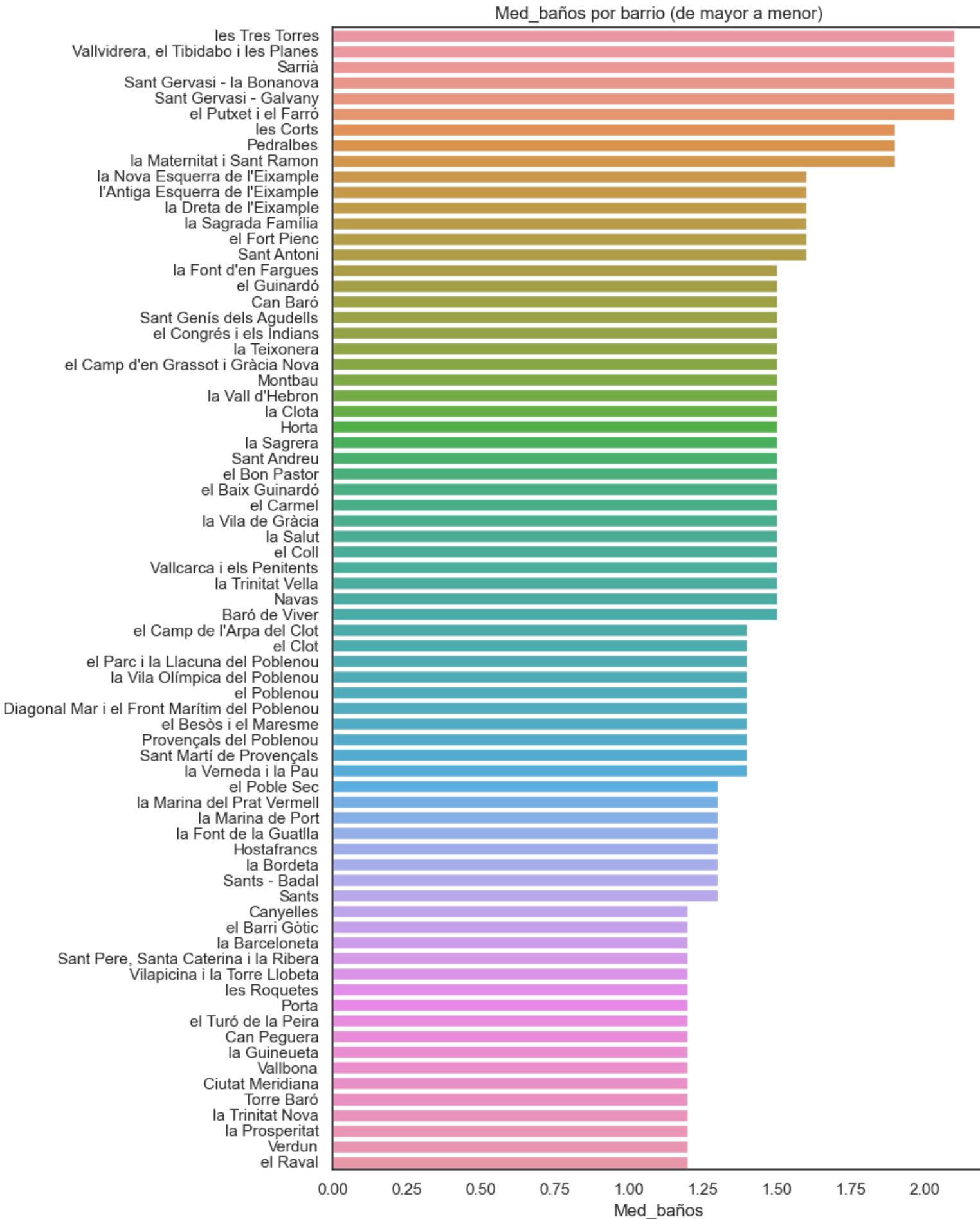
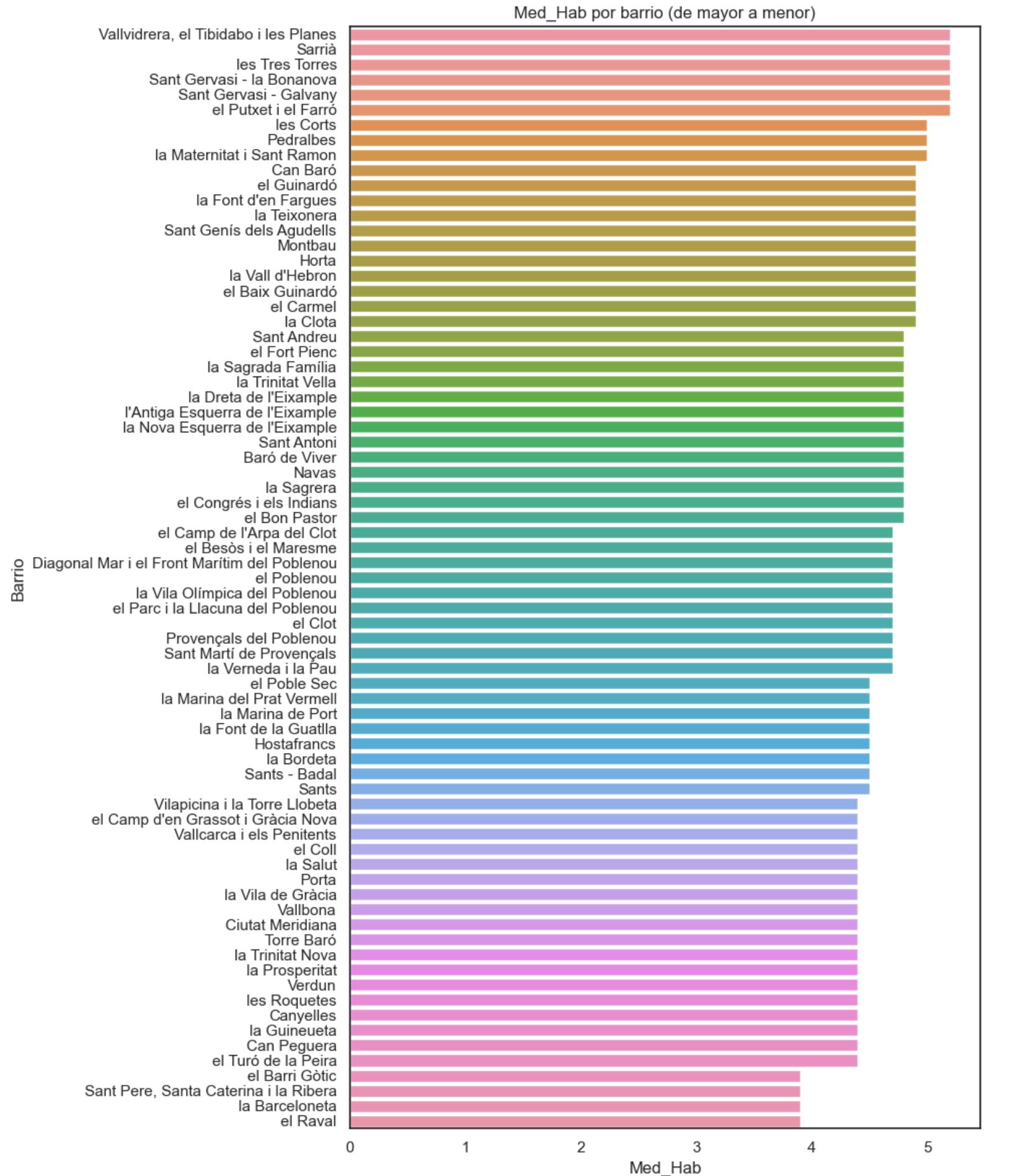
	Any	Codi_Districte	Codi_Barri	Precio_anual	Camaras_seguridad	Total_contratos	m2	Med_Hab	Med_baños	Med_Pers	Transporte	Areas_verdes	Equip_social	Equip_ocio
count	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0	73.0
mean	2023.0	6.0	37.0	1024.0	2.0	548.0	71.0	5.0	1.0	2.0	44.0	20.0	59.0	11.0
std	0.0	3.0	21.0	273.0	3.0	486.0	11.0	0.0	0.0	0.0	28.0	12.0	50.0	29.0
min	2023.0	1.0	1.0	492.0	0.0	2.0	51.0	4.0	1.0	2.0	4.0	1.0	2.0	0.0
25%	2023.0	4.0	19.0	857.0	0.0	188.0	64.0	4.0	1.0	2.0	24.0	10.0	25.0	0.0
50%	2023.0	7.0	37.0	951.0	2.0	414.0	69.0	5.0	2.0	2.0	37.0	19.0	43.0	1.0
75%	2023.0	8.0	55.0	1132.0	3.0	868.0	75.0	5.0	2.0	2.0	57.0	27.0	79.0	7.0
max	2023.0	10.0	73.0	1815.0	16.0	2042.0	112.0	5.0	2.0	3.0	135.0	53.0	283.0	190.0

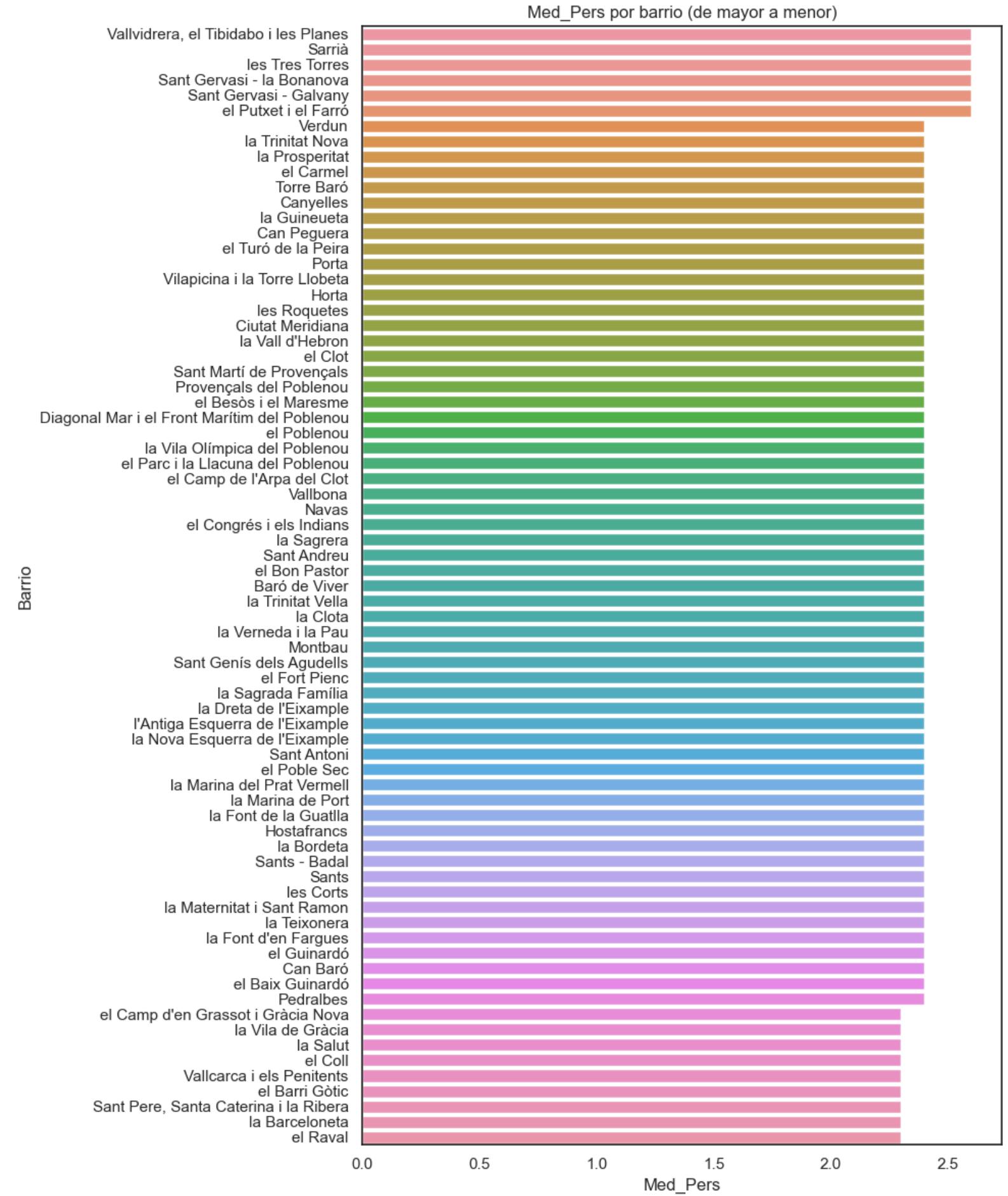








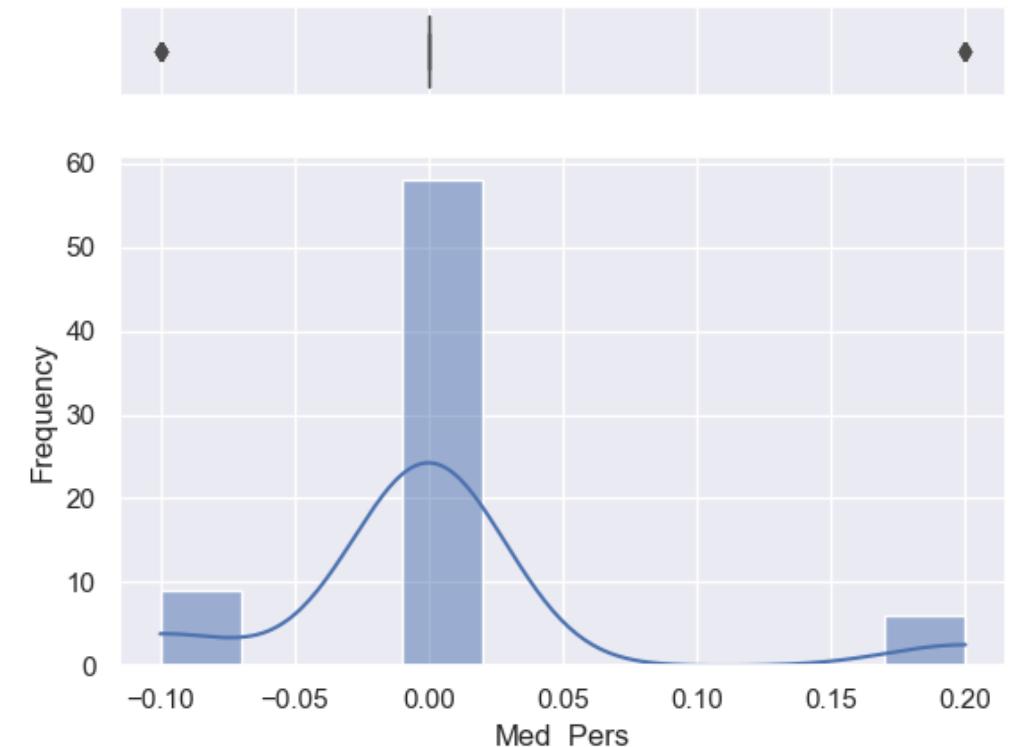
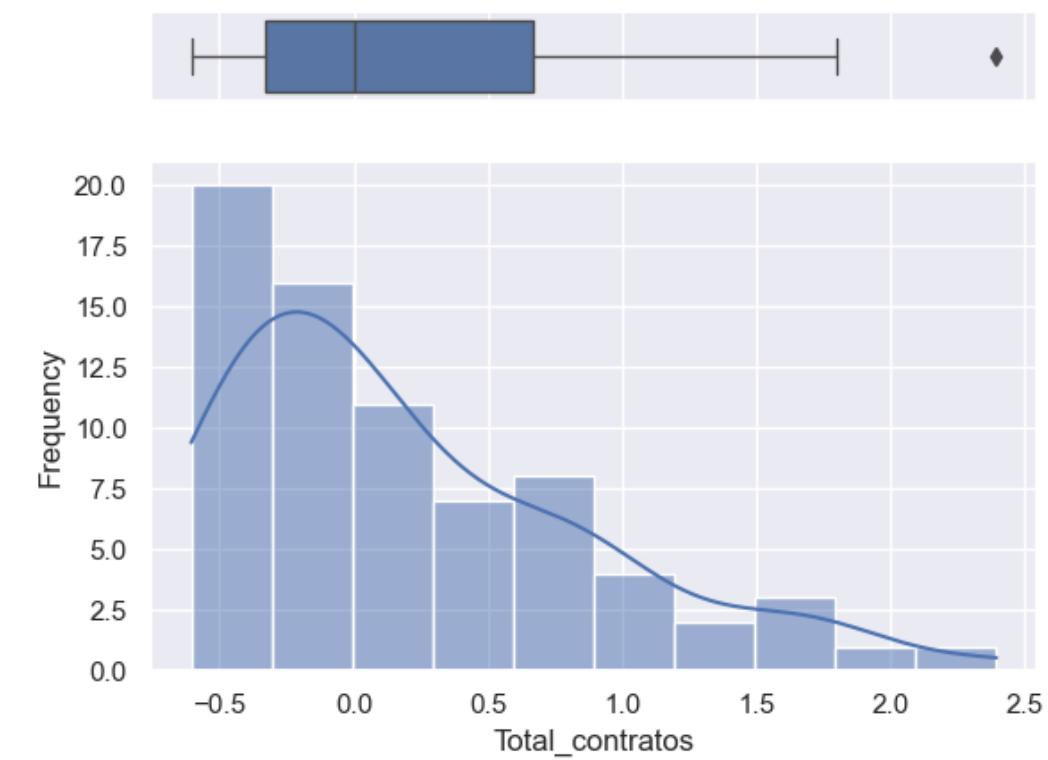
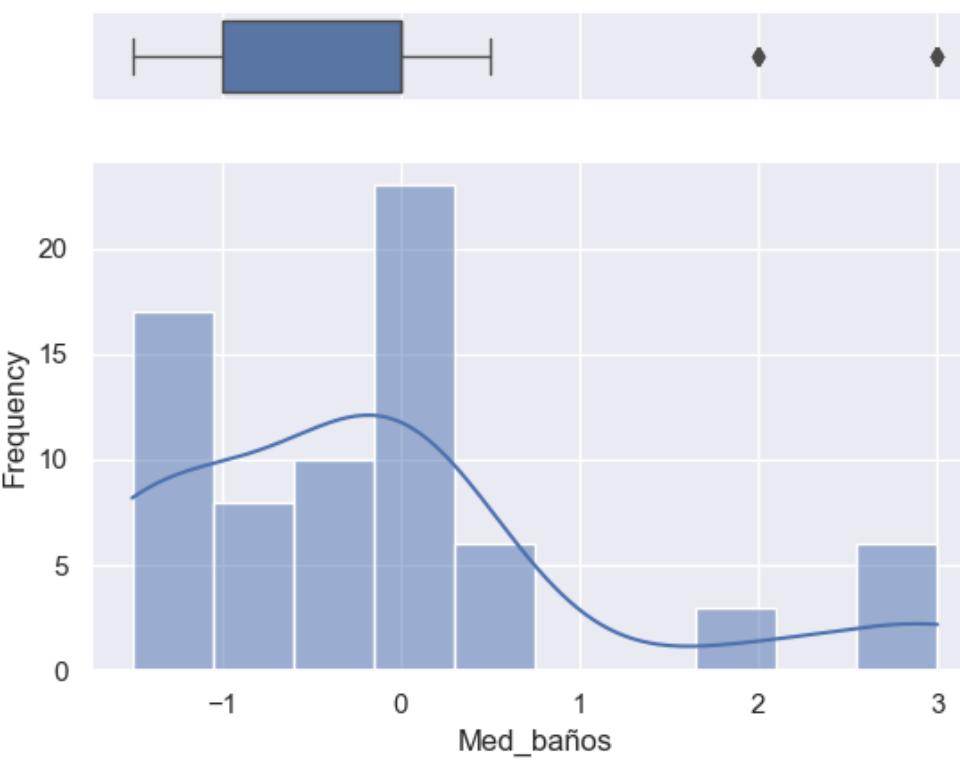
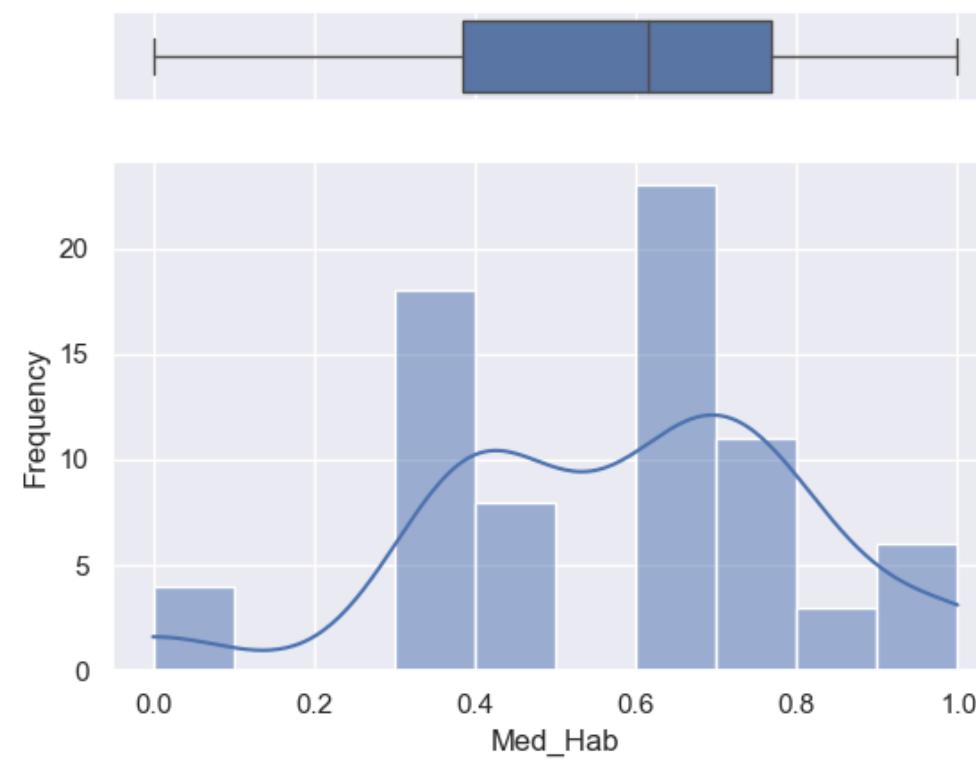


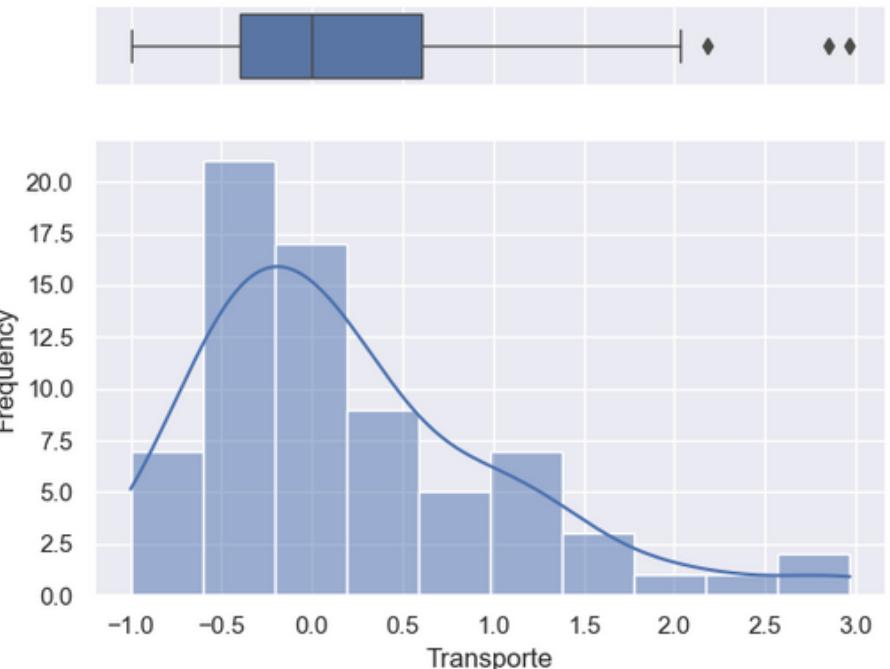
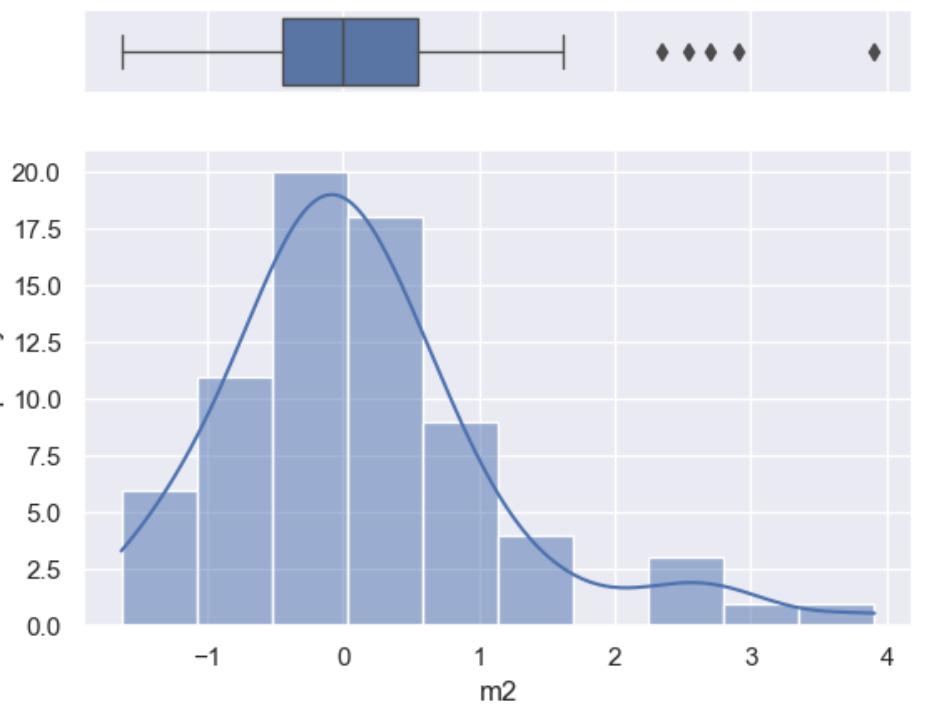
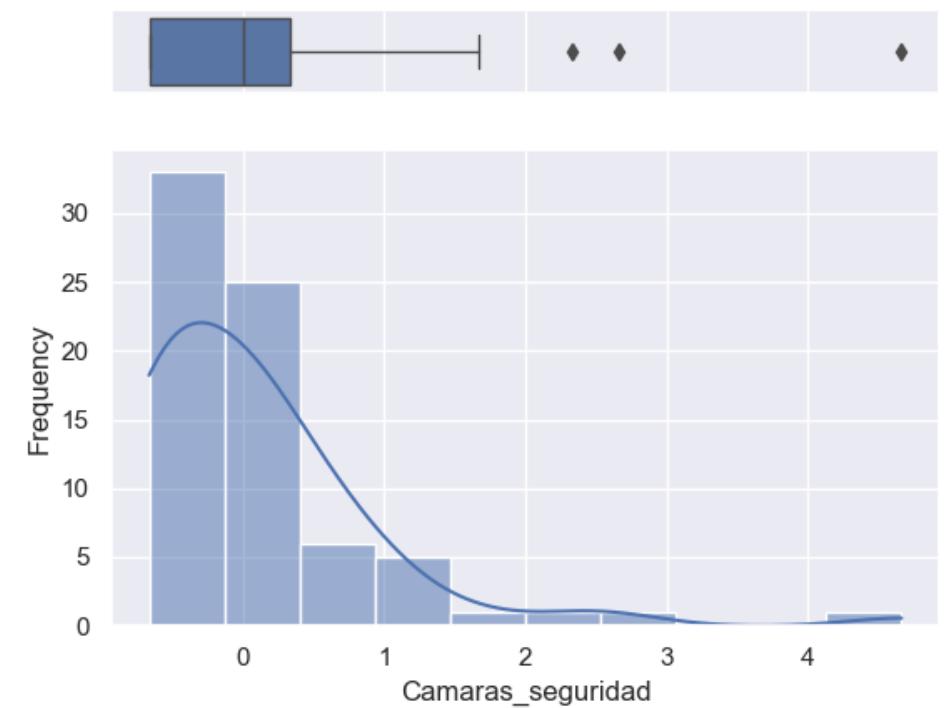
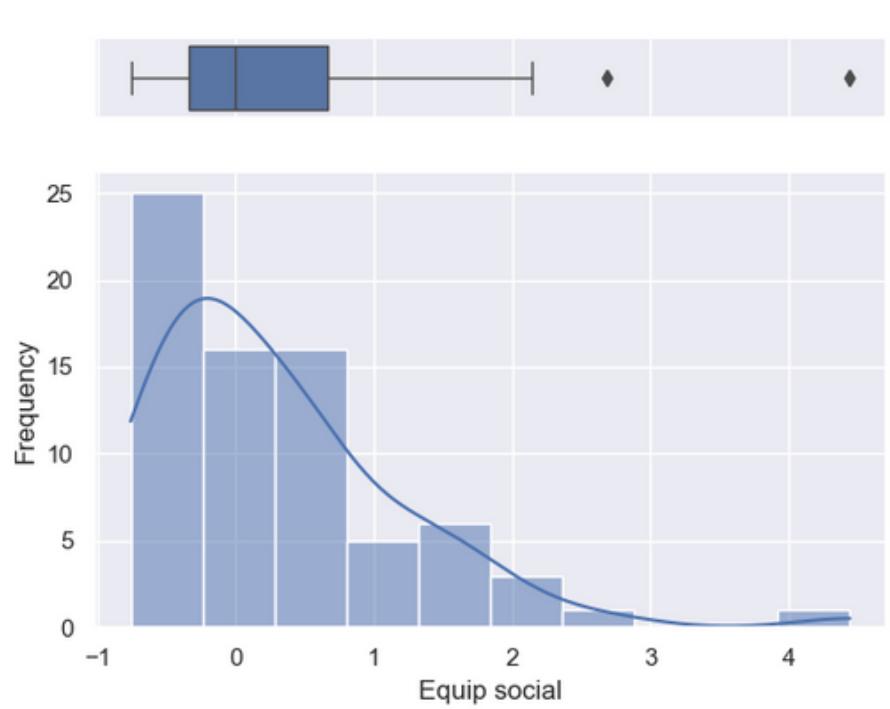
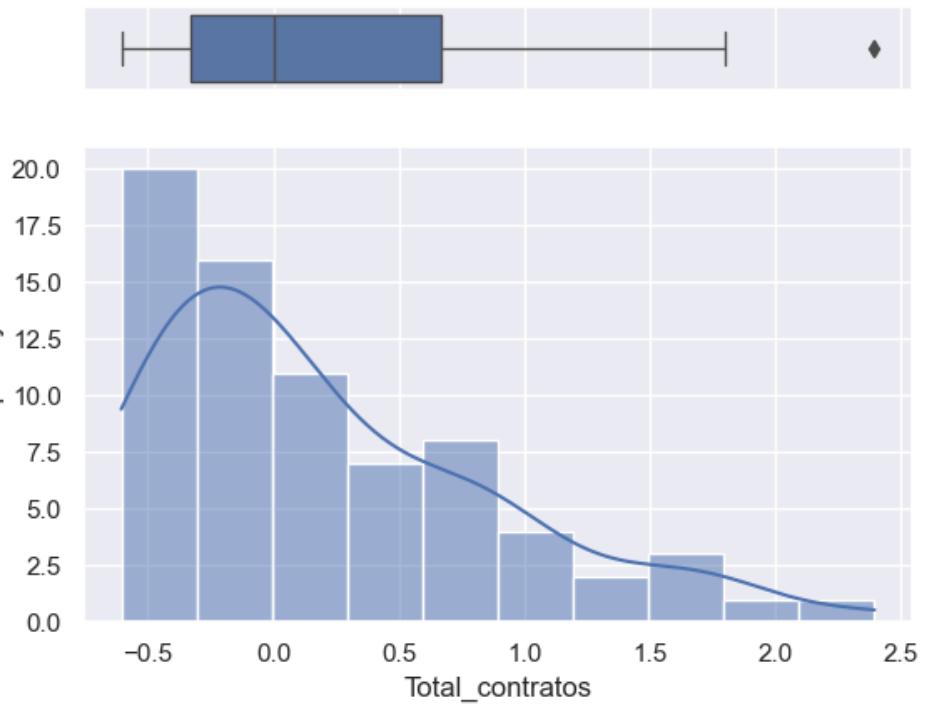
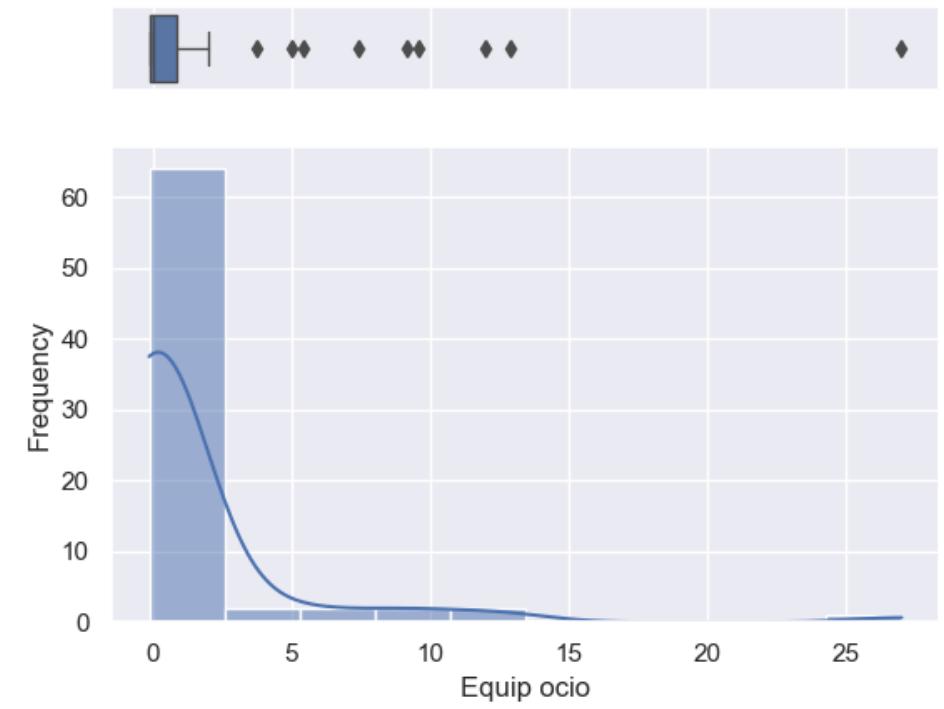


Se debe tener en cuenta el cambio constante de los datos en una ciudad como Barcelona

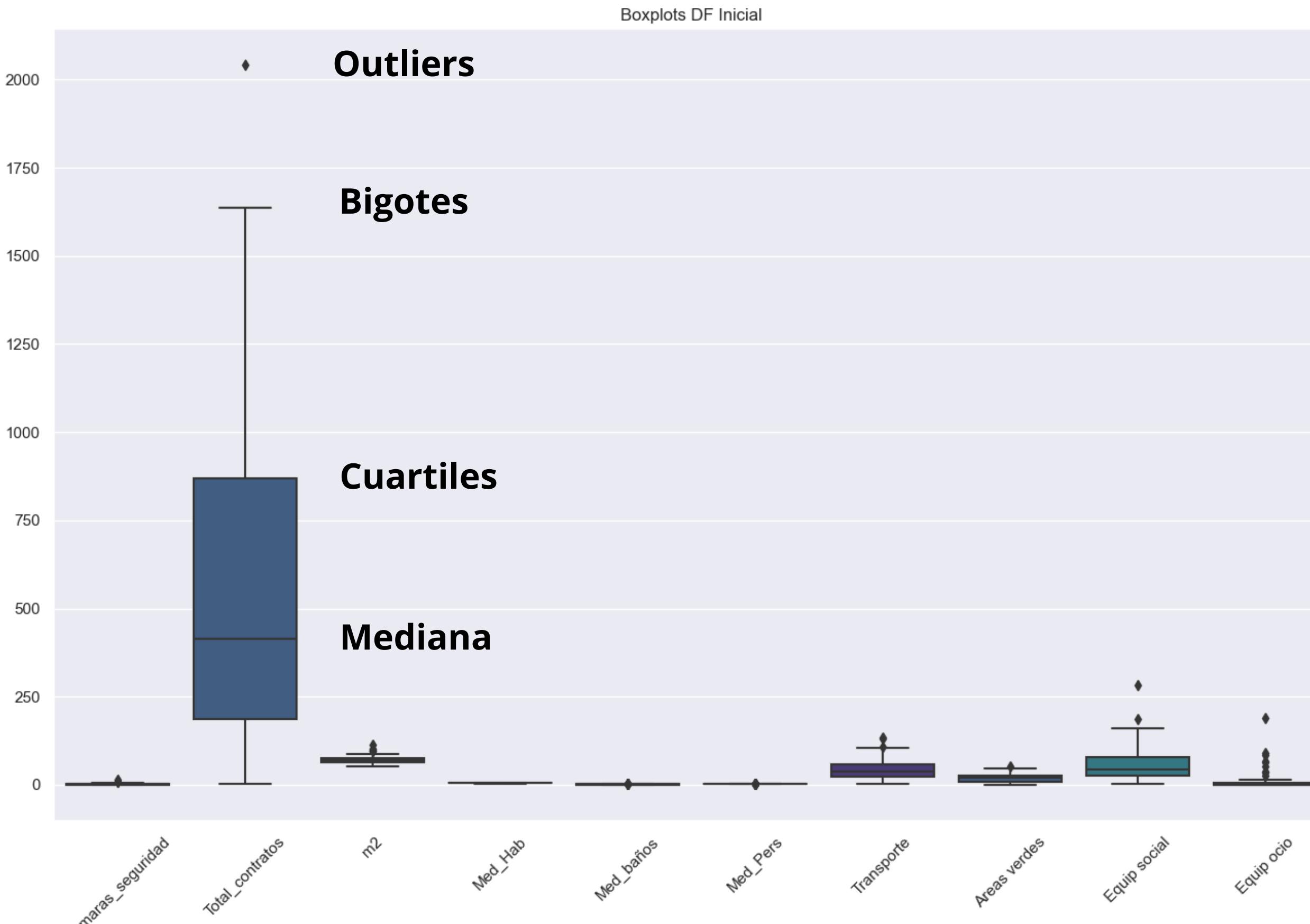
Preprocesamiento de datos:

Comprobamos normalidad y outliers basado en Test de Normalidad de Shapiro



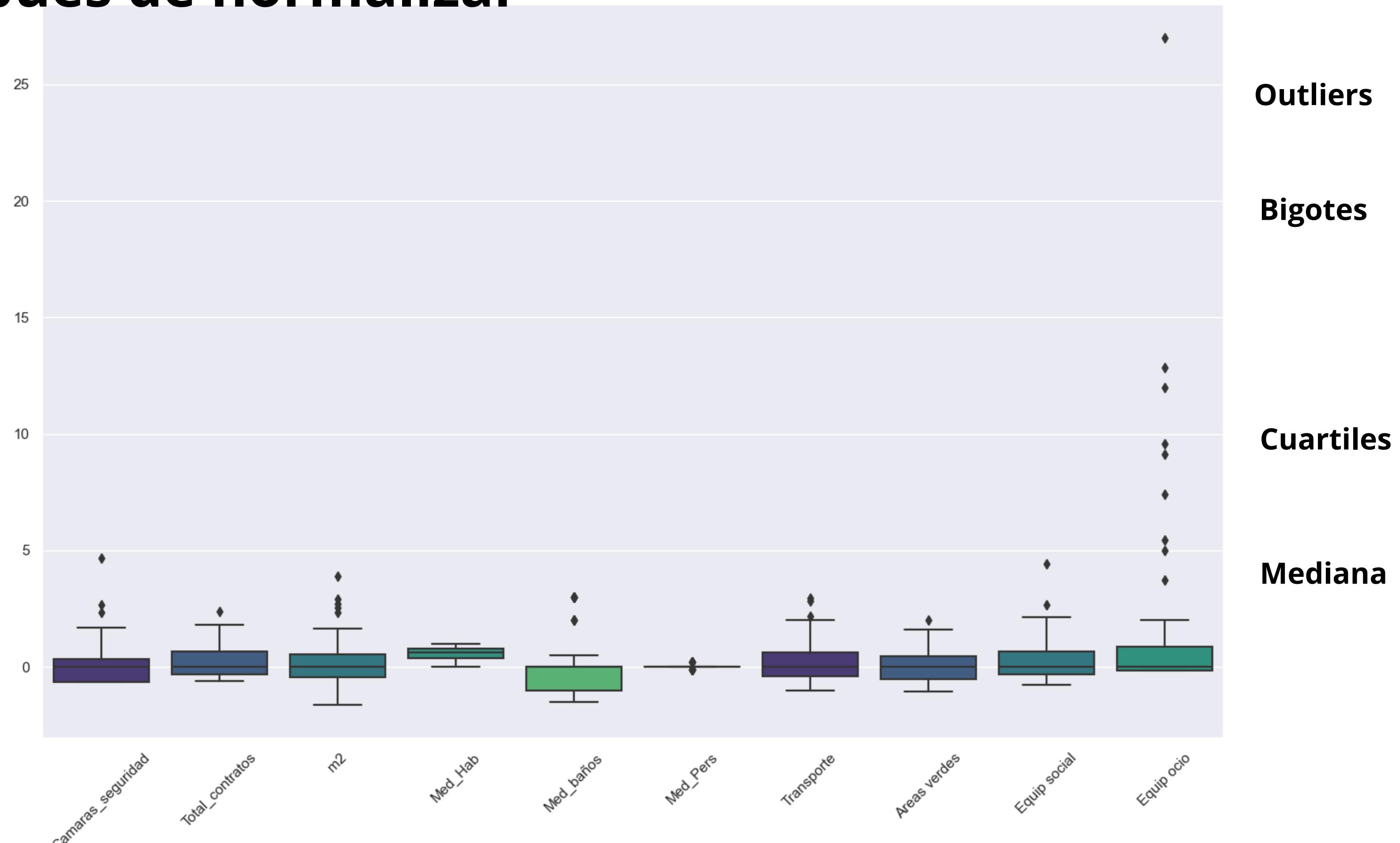


Antes de normalizar



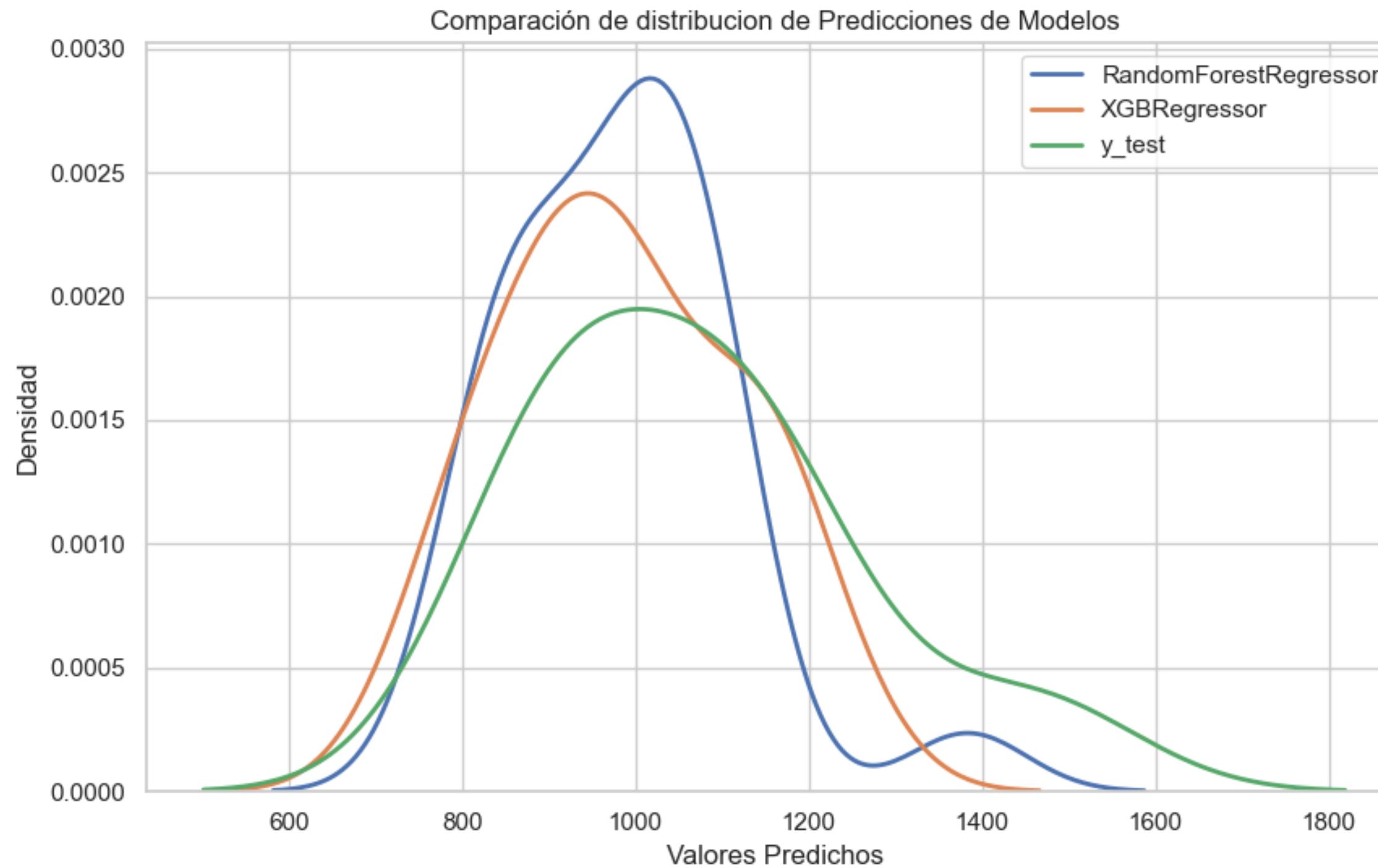
Después de normalizar

Boxplots con Normalizaciones



Modelos

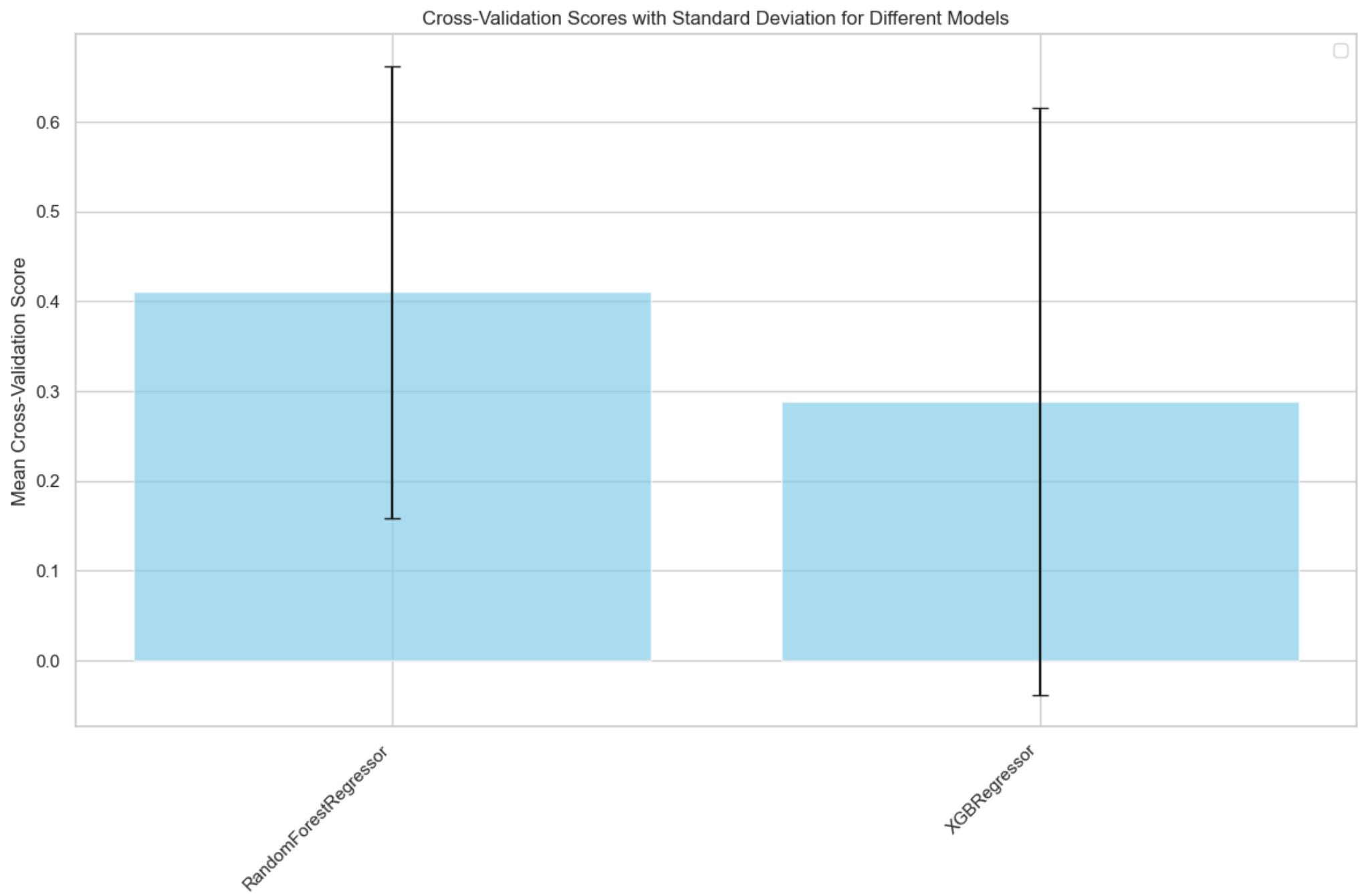
Entrenamos y generamos predicciones con cada modelo estándar



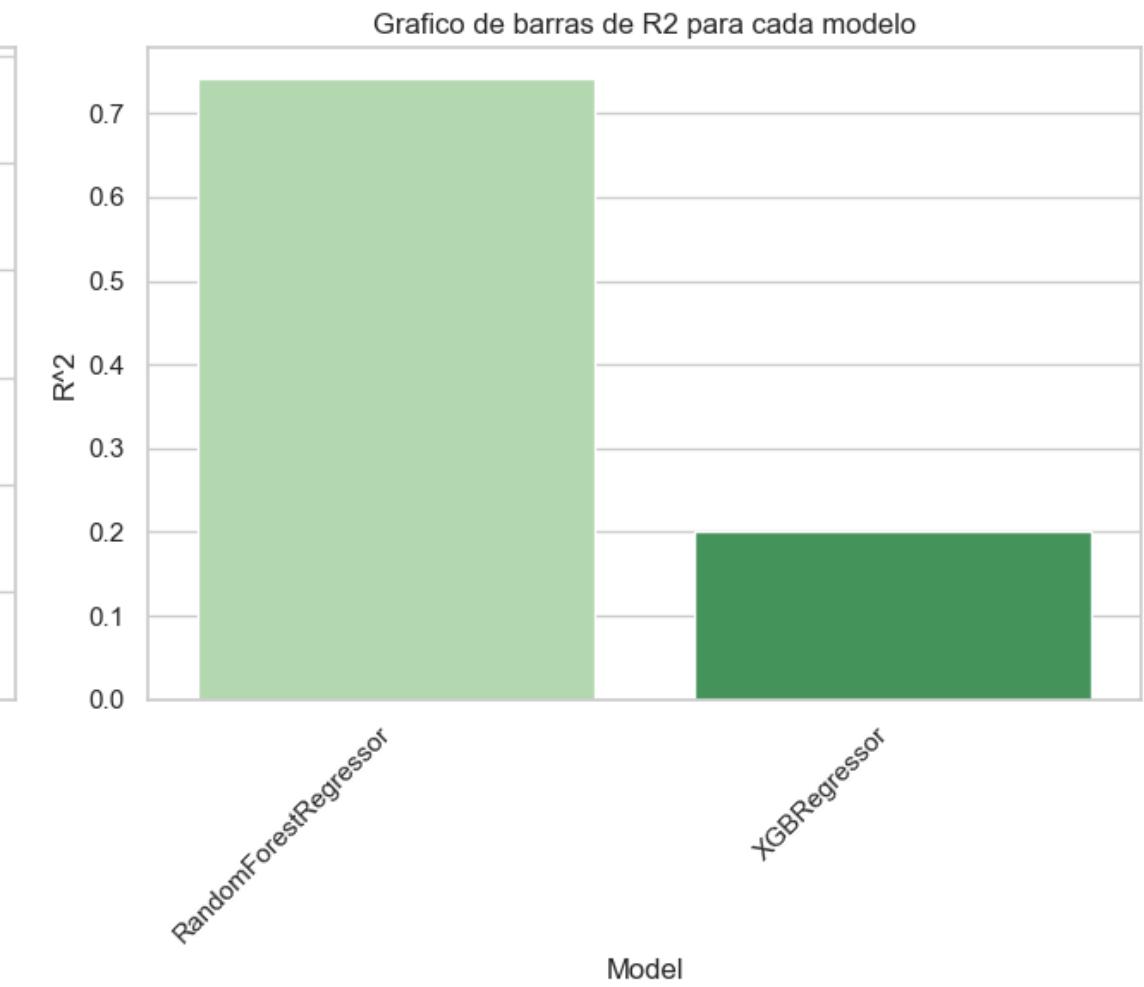
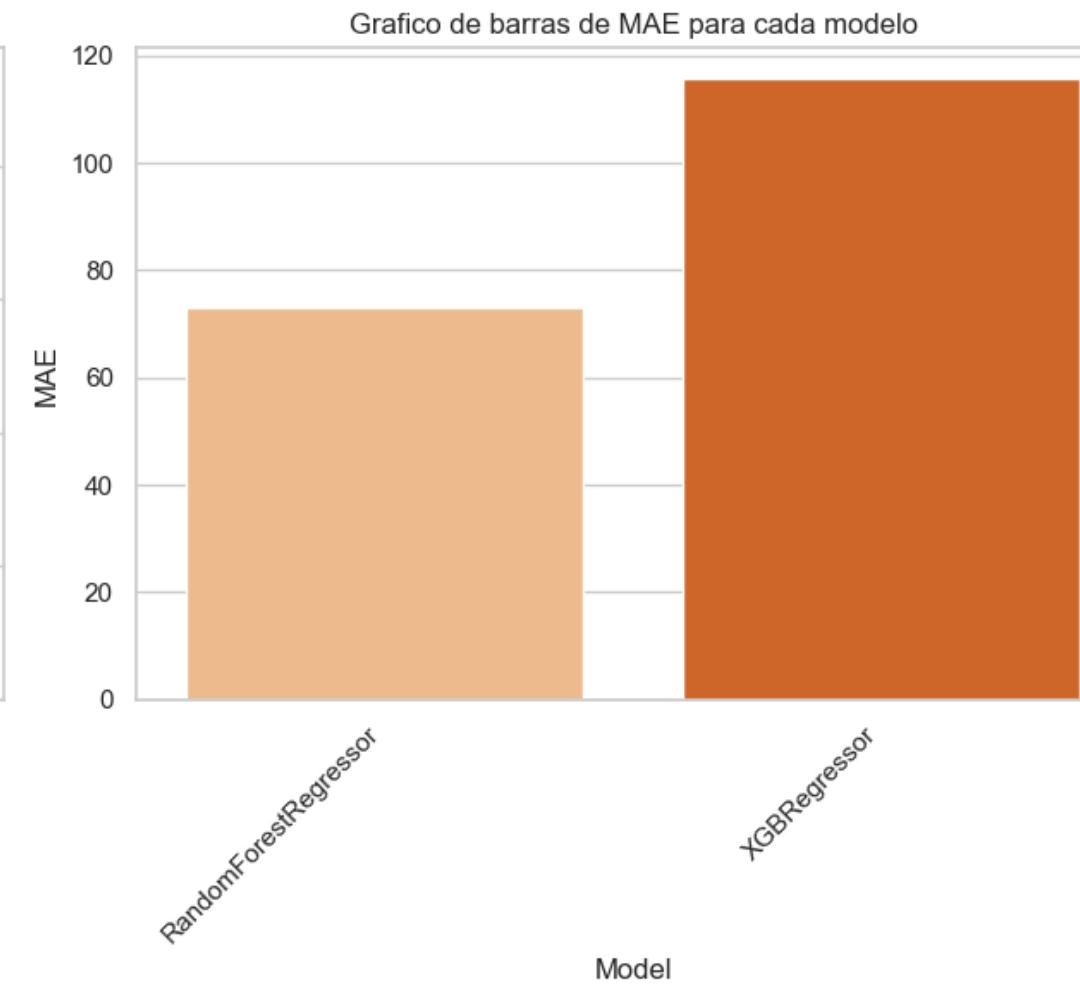
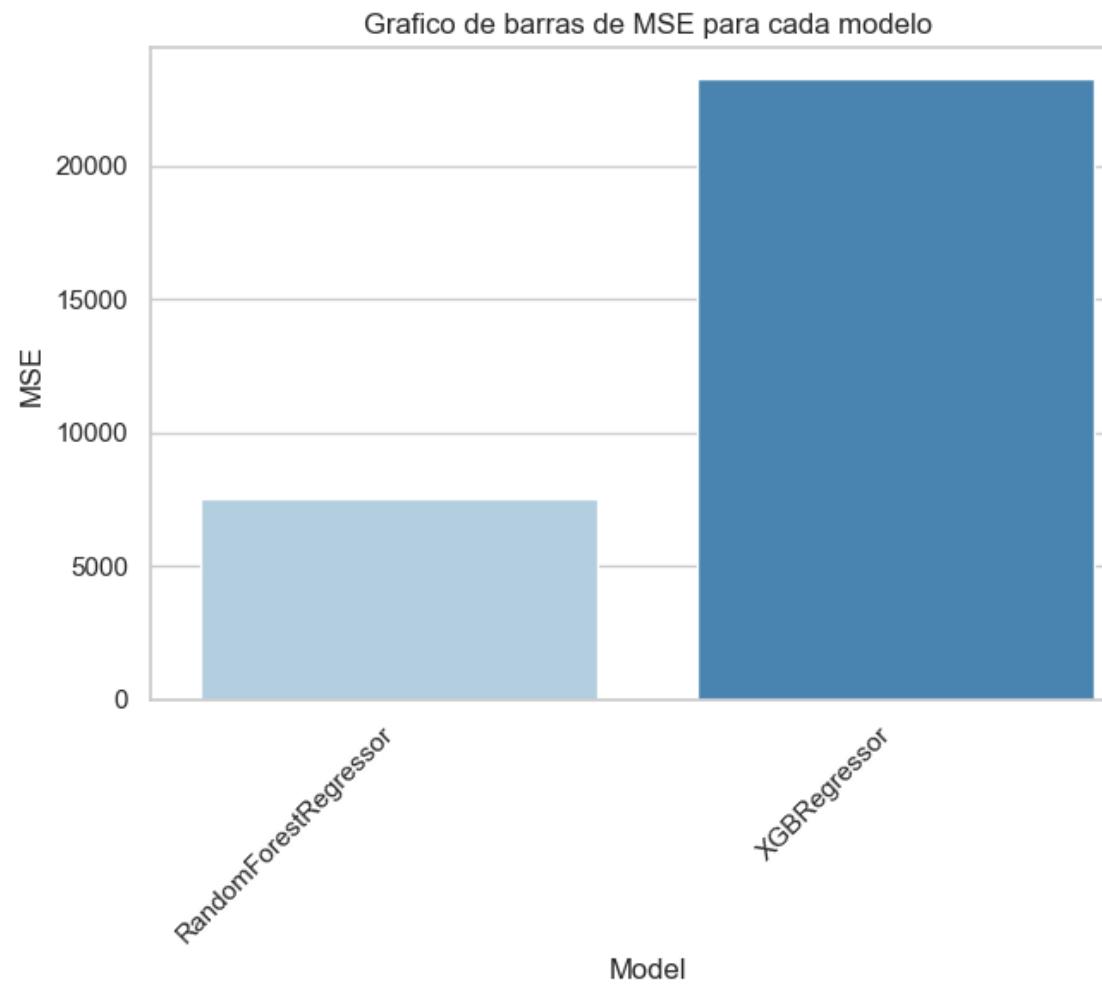
Cross - Validation de los modelos

	Modelo	Mean_R2_Score	Std_R2_Score
0	RandomForestRegressor	0.410520	0.252043
1	XGBRegressor	0.288246	0.326819

Grafico de barra de error

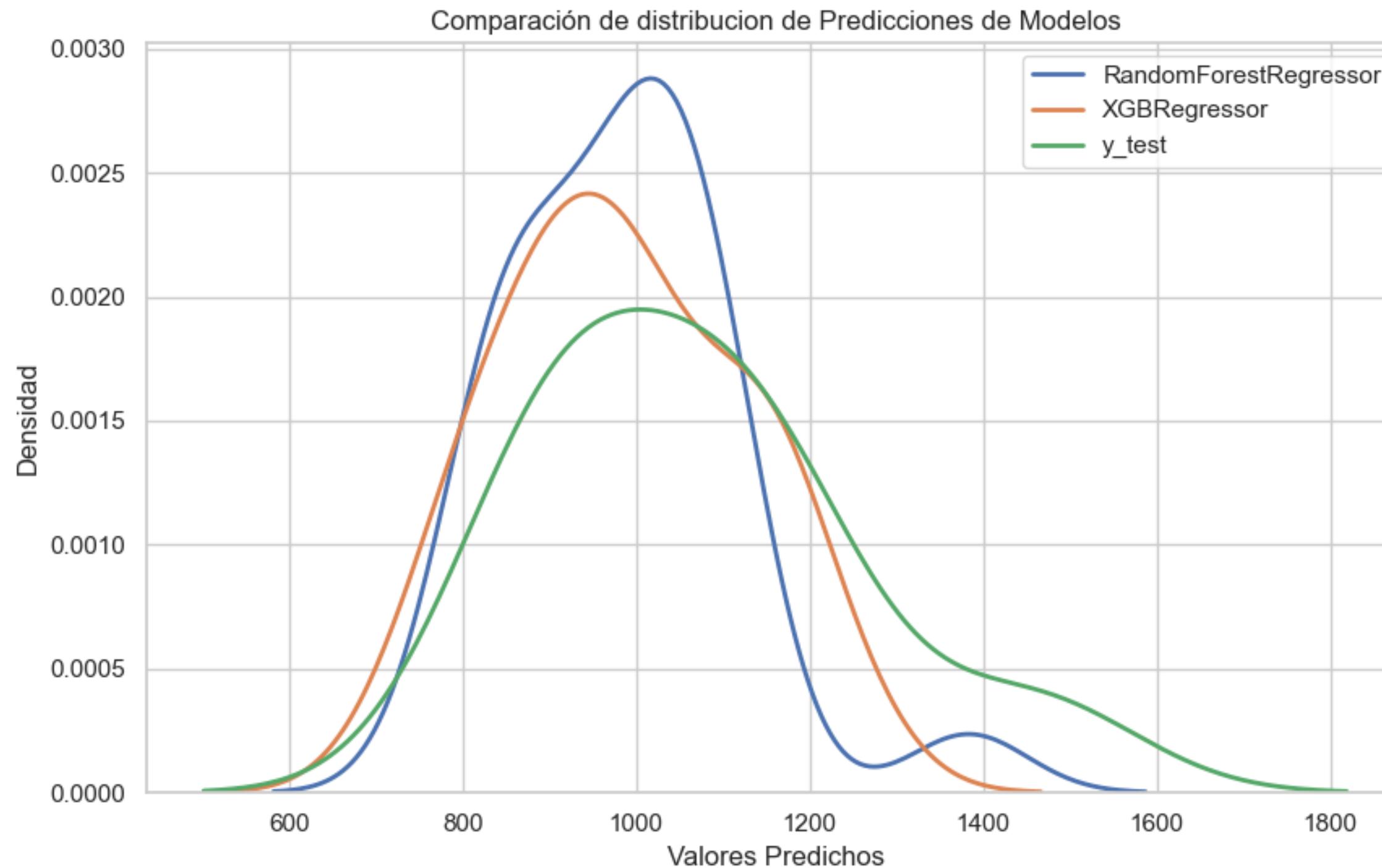


Métricas de evaluación



	Model	MSE	MAE	R^2
0	RandomForestRegressor	7535.871207	73.202660	0.741023
1	XGBRegressor	23261.976322	115.621855	0.200580

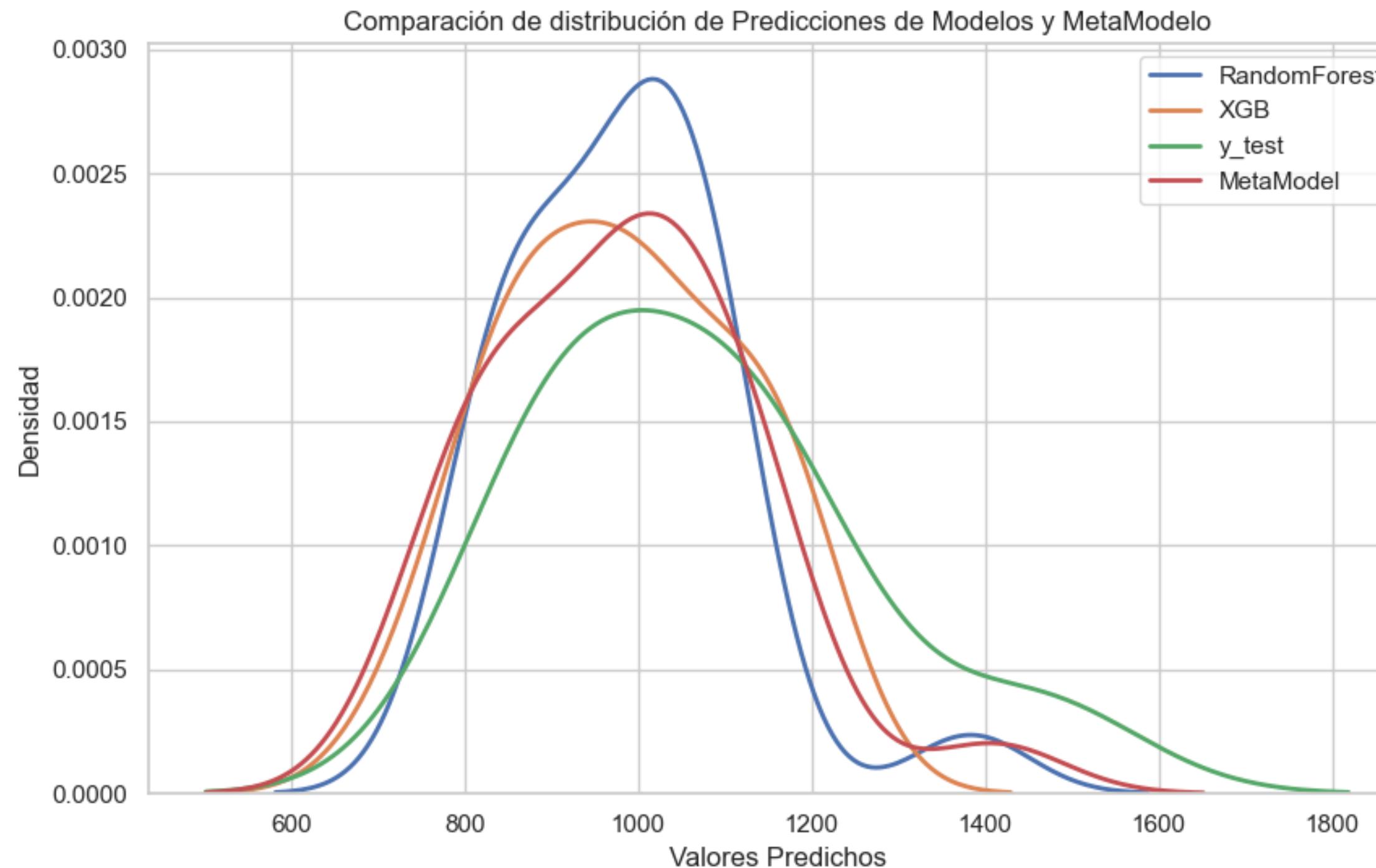
Entrenamos y generamos predicciones con cada modelo estándar



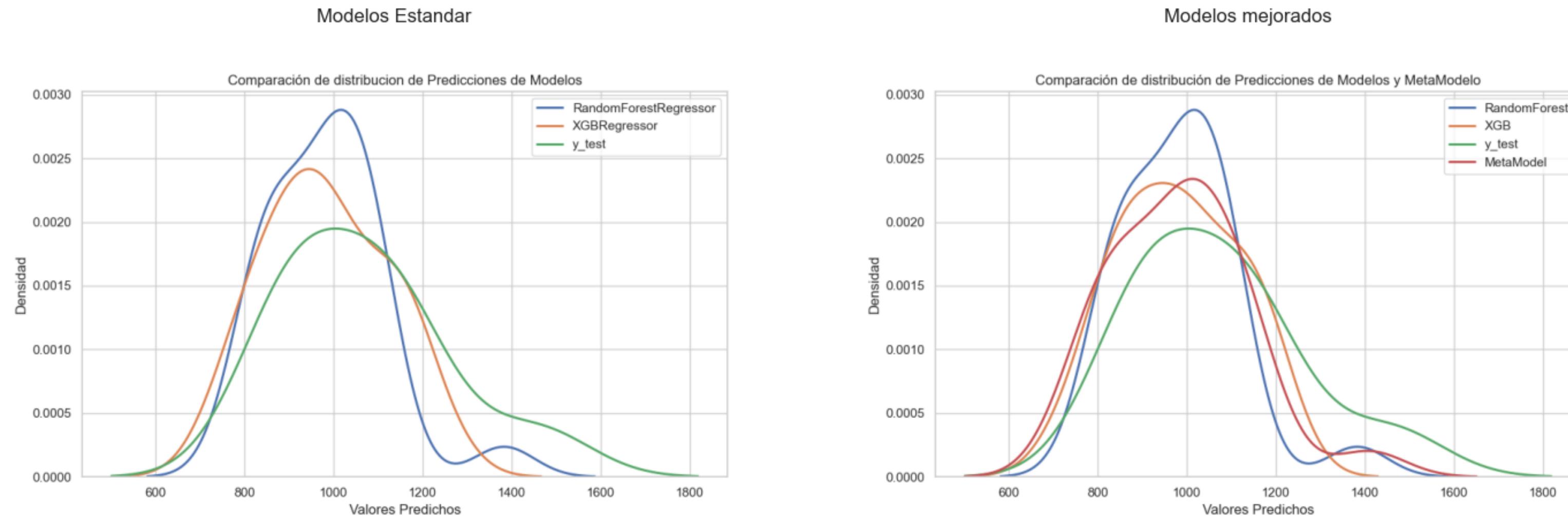
Meta modelo

Un meta modelo es un modelo que se entrena sobre las predicciones de otros modelos en lugar de los datos originales. En otras palabras, en lugar de entrenar directamente sobre las características de entrada y las etiquetas de salida, el meta modelo se entrena sobre las predicciones generadas por otros modelos (llamados modelos base) como características de entrada.

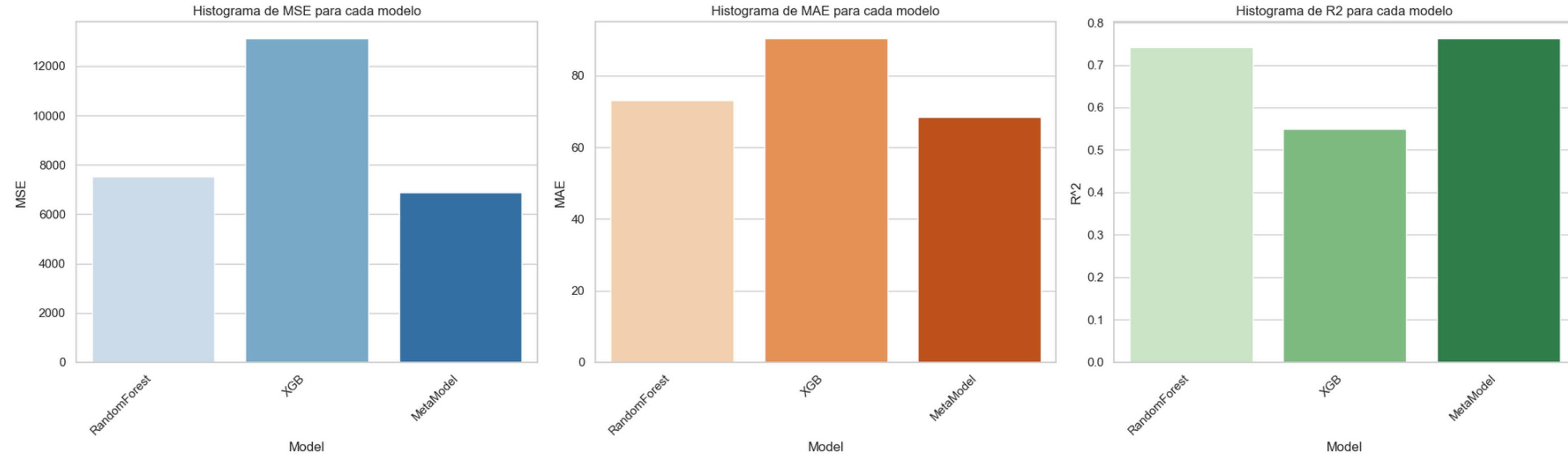
Entrenamos y generamos predicciones con cada modelo mejorado



Comparación de modelos, estándar - mejorados y Meta modelo.



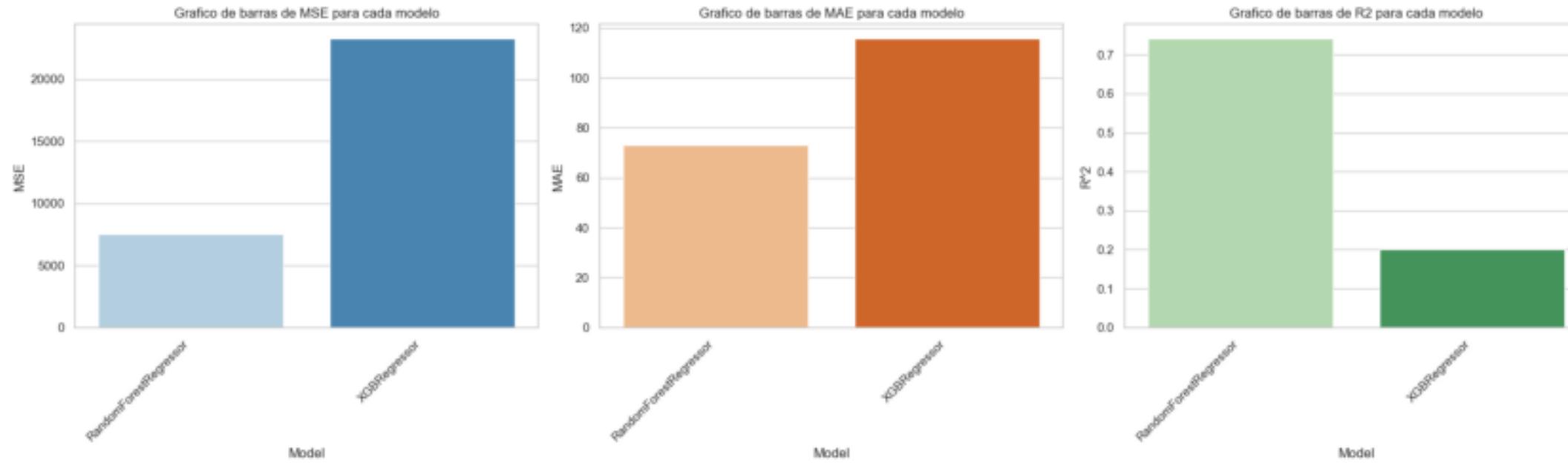
Métricas de evaluación



	Model	MSE	MAE	R ²
0	RandomForest	7525.299952	73.122240	0.741386
1	XGB	13131.708117	90.403769	0.548716
2	MetaModel	6895.917331	68.387411	0.763015

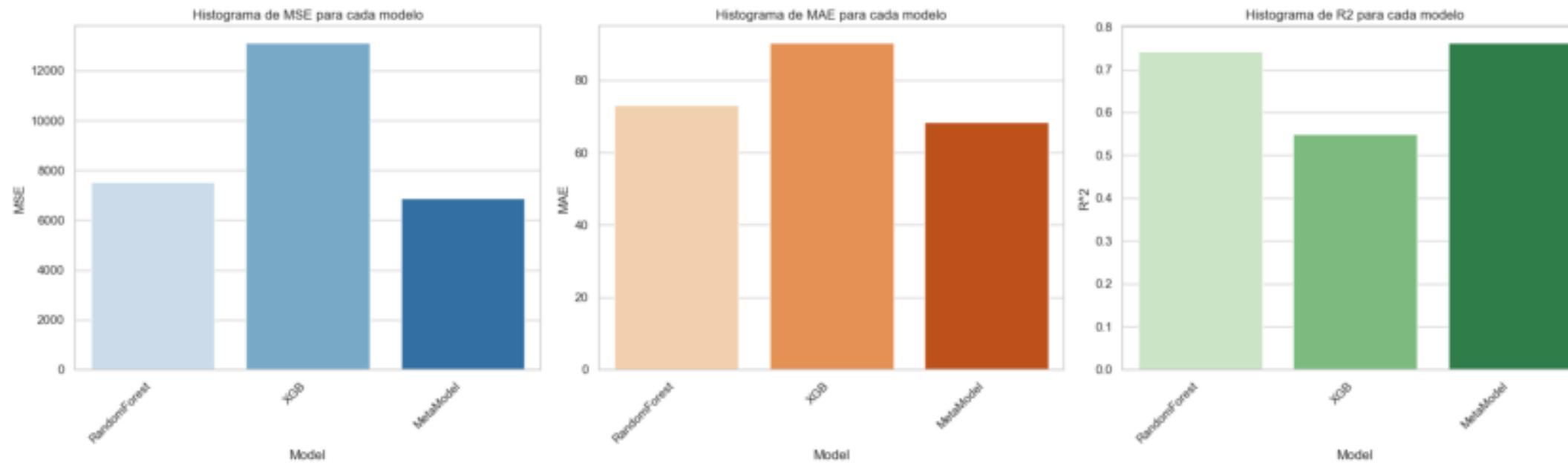
Comparación de Métricas de evaluación

Modelos Estandar



Model	MSE	MAE	R ²
RandomForestRegressor	7535.871207	73.202660	0.741023
XGBRegressor	23261.976322	115.621855	0.200580

Modelos mejorados



Model	MSE	MAE	R ²
RandomForest	7525.299952	73.122240	0.741386
XGB	13131.788117	90.403769	0.548716
MetaModel	6895.917331	68.387411	0.763015

Cross - Validation de los modelos mejorados

implementación de Meta modelo

RandomForest - Mean Cross-Validation Score: 0.4105201282198277

RandomForest - Standard Deviation of Cross-Validation Scores: 0.25204326699556673
[0.08341534 0.69671247 0.45143258]

XGB - Mean Cross-Validation Score: 0.32701274640762906

XGB - Standard Deviation of Cross-Validation Scores: 0.2781866692145293
[-0.06638674 0.52677309 0.52065189]

Meta-Modelo - Mean Cross-Validation Score: 0.5038334215079358

Meta-Modelo - Standard Deviation of Cross-Validation Scores: 0.24796426944913466
[0.23504868 0.83327833 0.44317326]

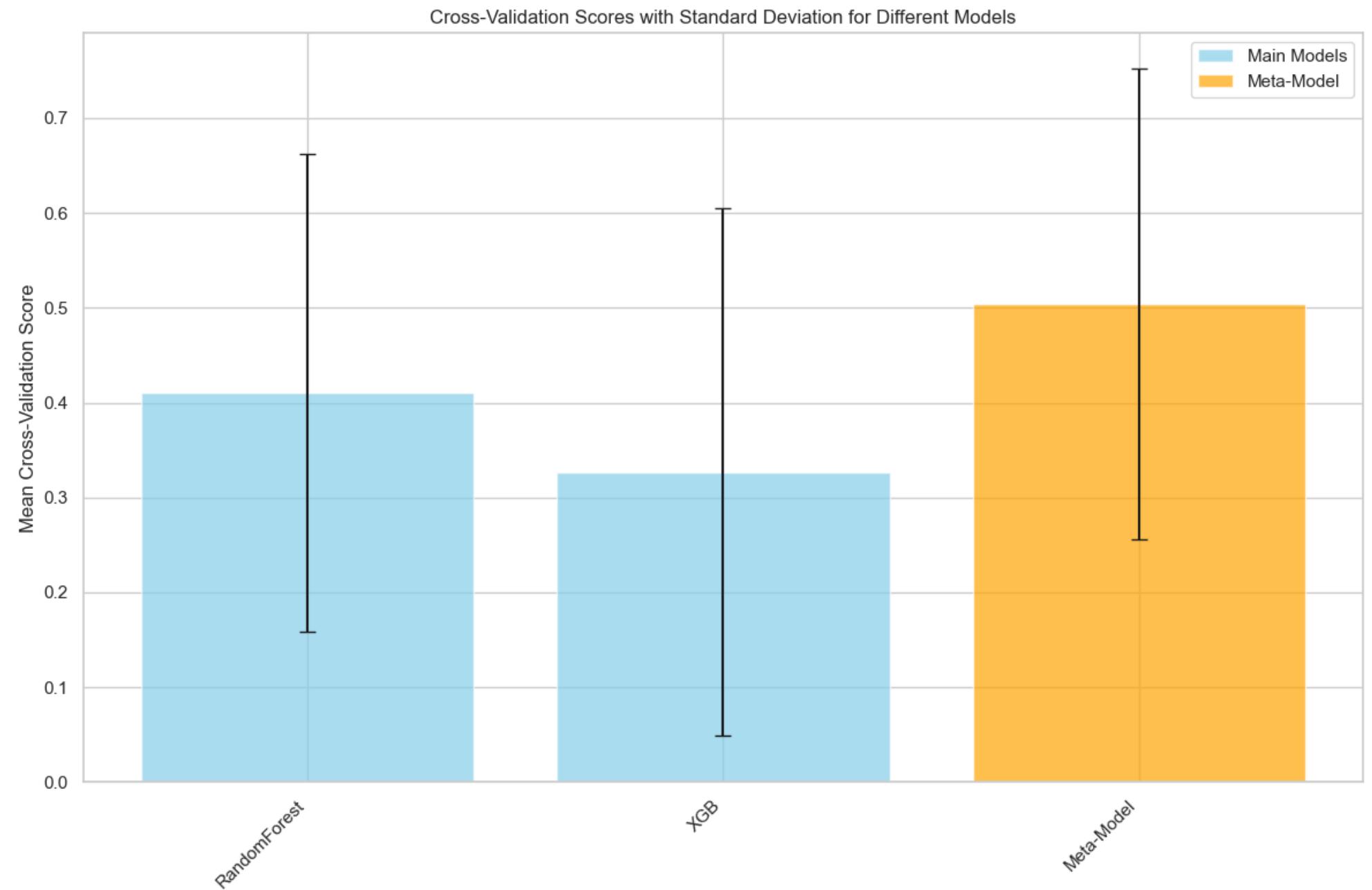
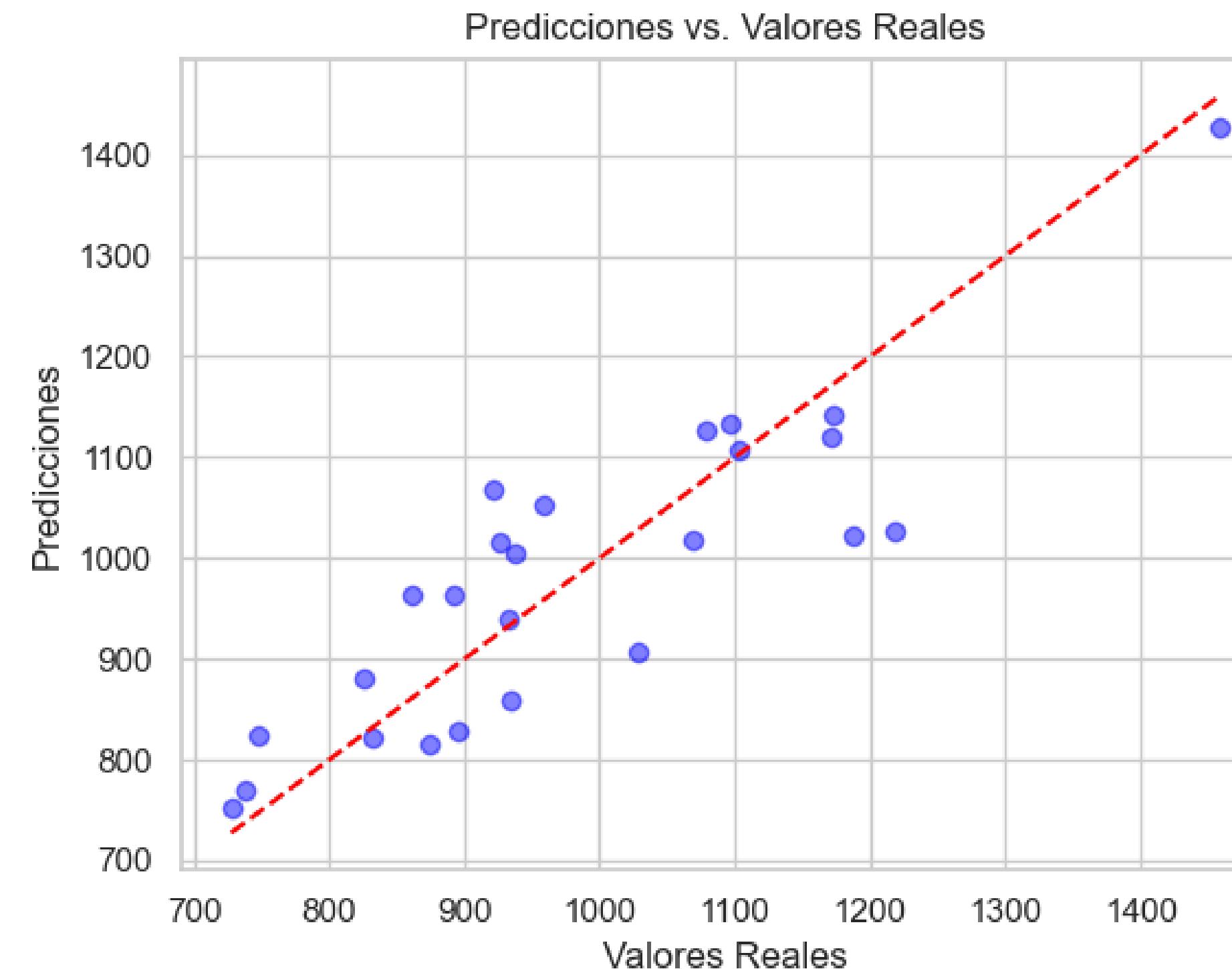
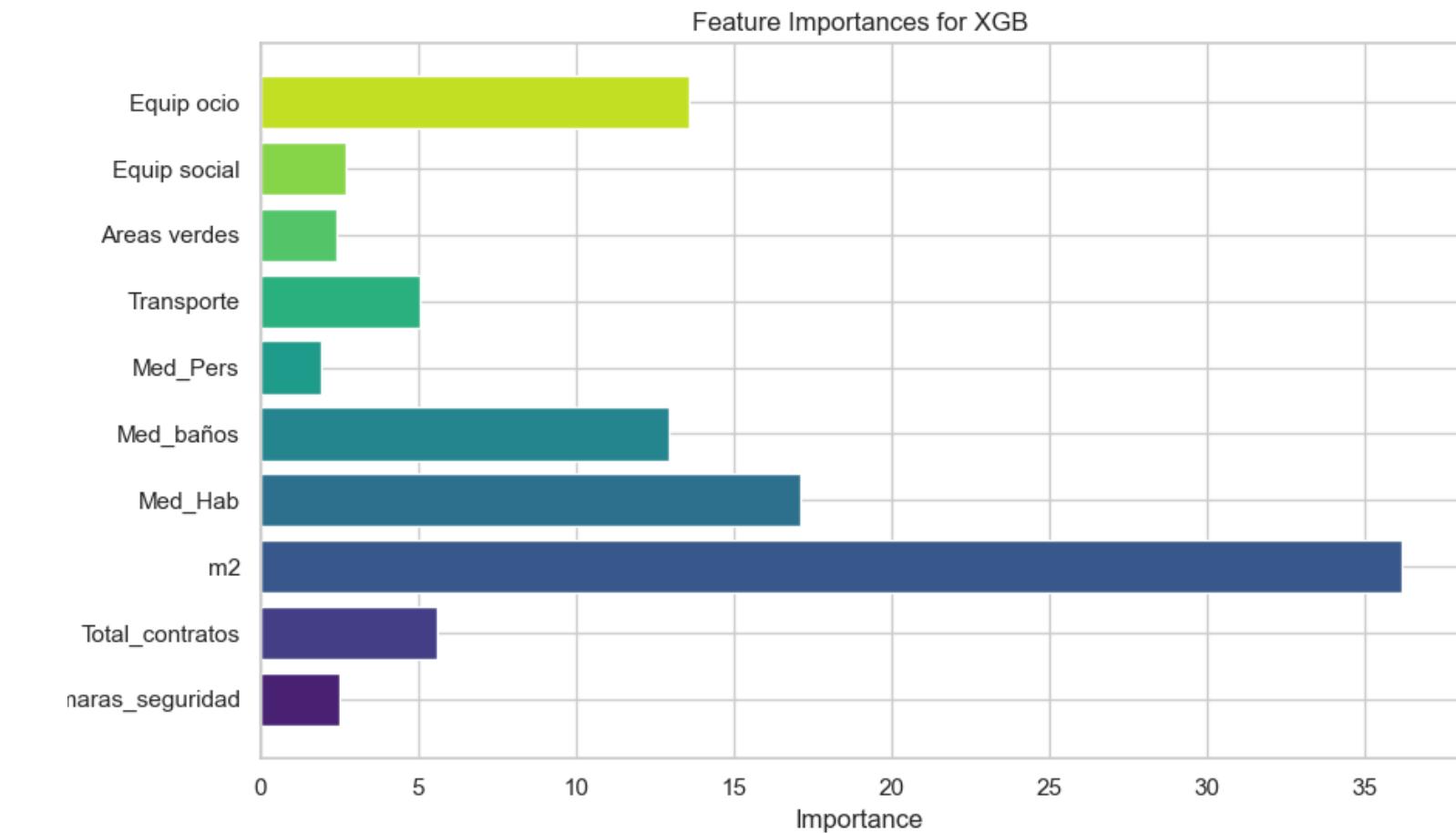
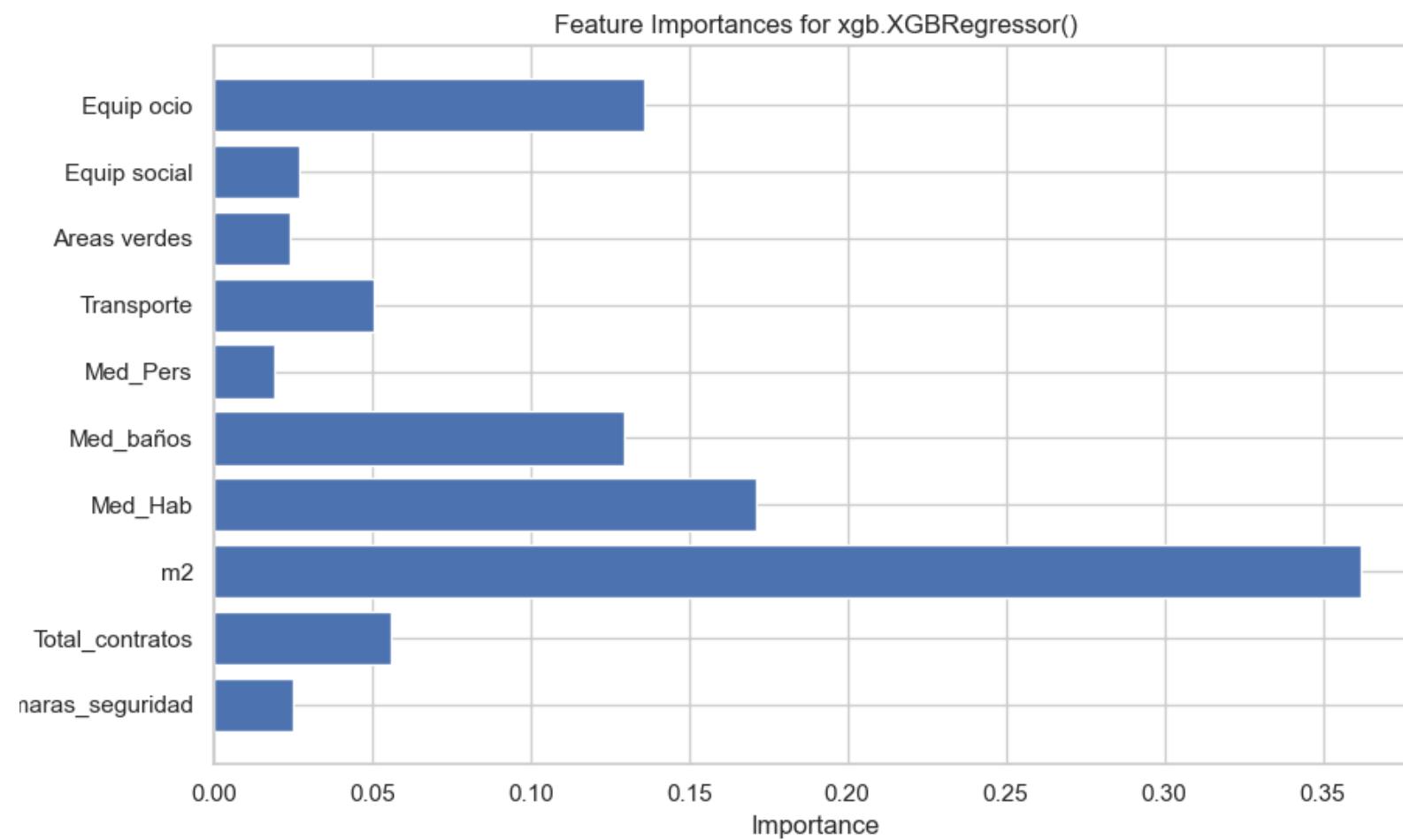


Grafico de barra de error

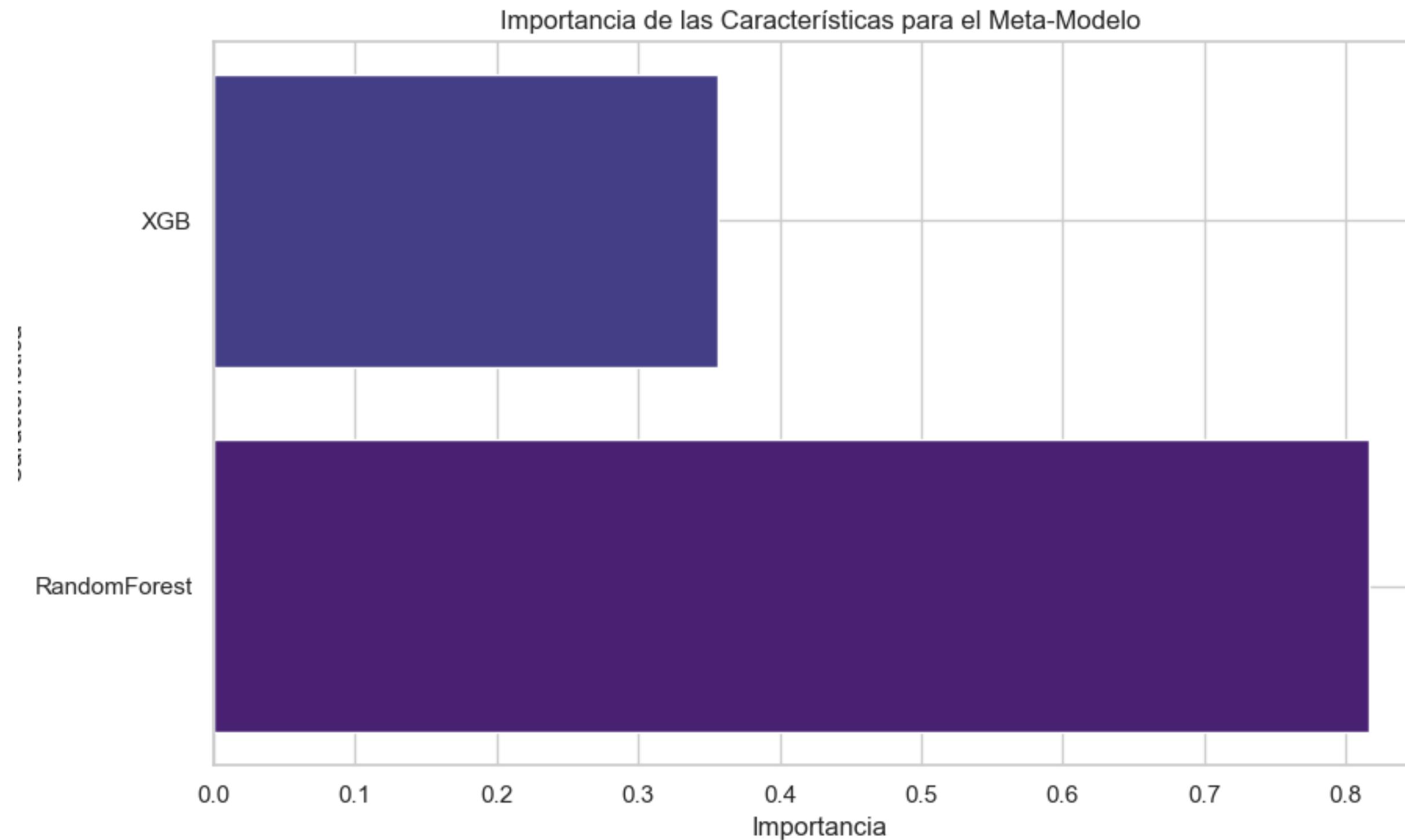
Como se ajustaron las predicciones a los valores reales?



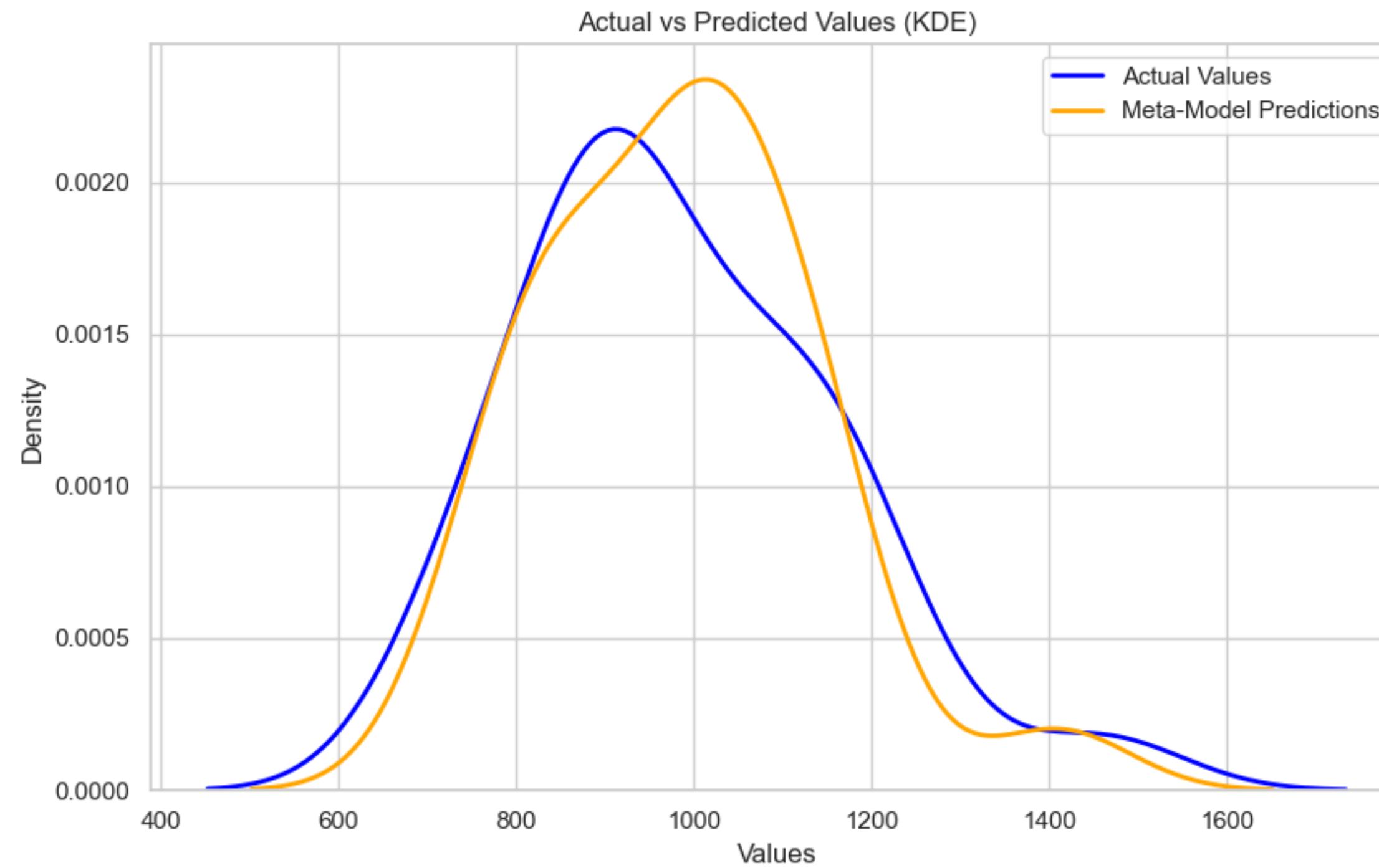
Importancia de las variables en cada modelo



Importancia de los modelos en el Meta - Modelo



Meta - Modelo vs Target



Conclusiónes:

- La implementación de un Meta - Modelo, para mejorar el ajuste de las predicciones a los datos reales, fue correcta y necesaria para mejorar la confianza en el proceso.
- los datos son la base de todo el proceso
- La correlación entre variables es un determinante a la hora de modelos de predicción, saber cuales son las variables que mas afectan a la variable target, puede ayudar a tomar decisiones a la hora de crear modelos.
- Es posible acercarse a la predicción del Precio de alquiler de vivienda en Barcelona. pero son necesarios mas esfuerzos gubernamentales por la recolección exacta y adecuada de los datos.
- Lo que hace a una ciudad equitativa, no es lo que tengamos que pagar por vivir mejor, sino la garantía de una calidad de vida para todos.