

Análisis Real State Madrid 2022

Sora Phi Rodríguez Gutiérrez



Índice

1

Presentación
del Dataset

2

Resumen
Estadístico

3

Conclusiones

4

Clasificación

Presentación

Inicial

Análisis Inicial

Información General del Dataset:

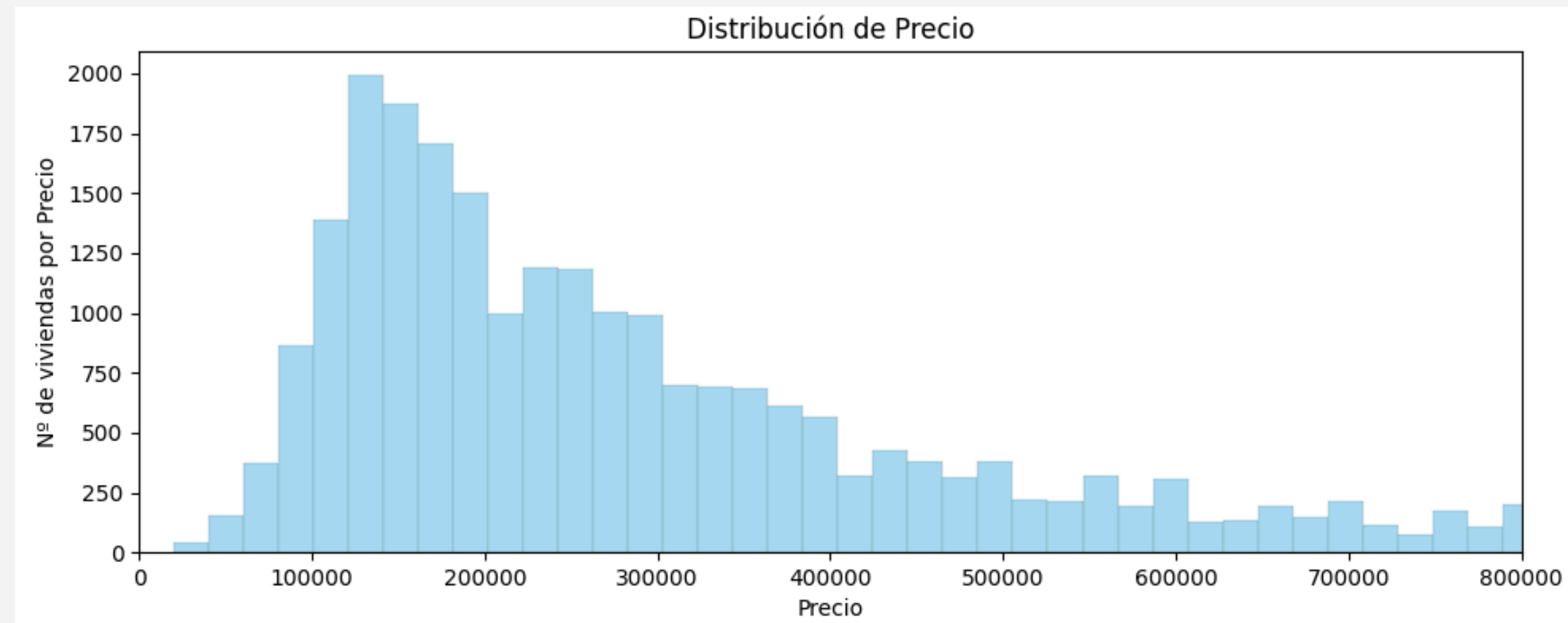
- El DataFrame contiene 26026 filas y 74 columnas
- Hay varias columnas con valores nulos como Anunciante, Distrito, etc, pero no parecen estar relacionadas con el precio de la vivienda

Valores Nulos:

- Algunas columnas importantes tienen muchos valores nulos, como Calle, que tiene 11966, conformando casi la mitad de registros
- Otras columnas, como Superficie o Habitaciones, no tienen ningún valor nulo

Resumen Estadístico:

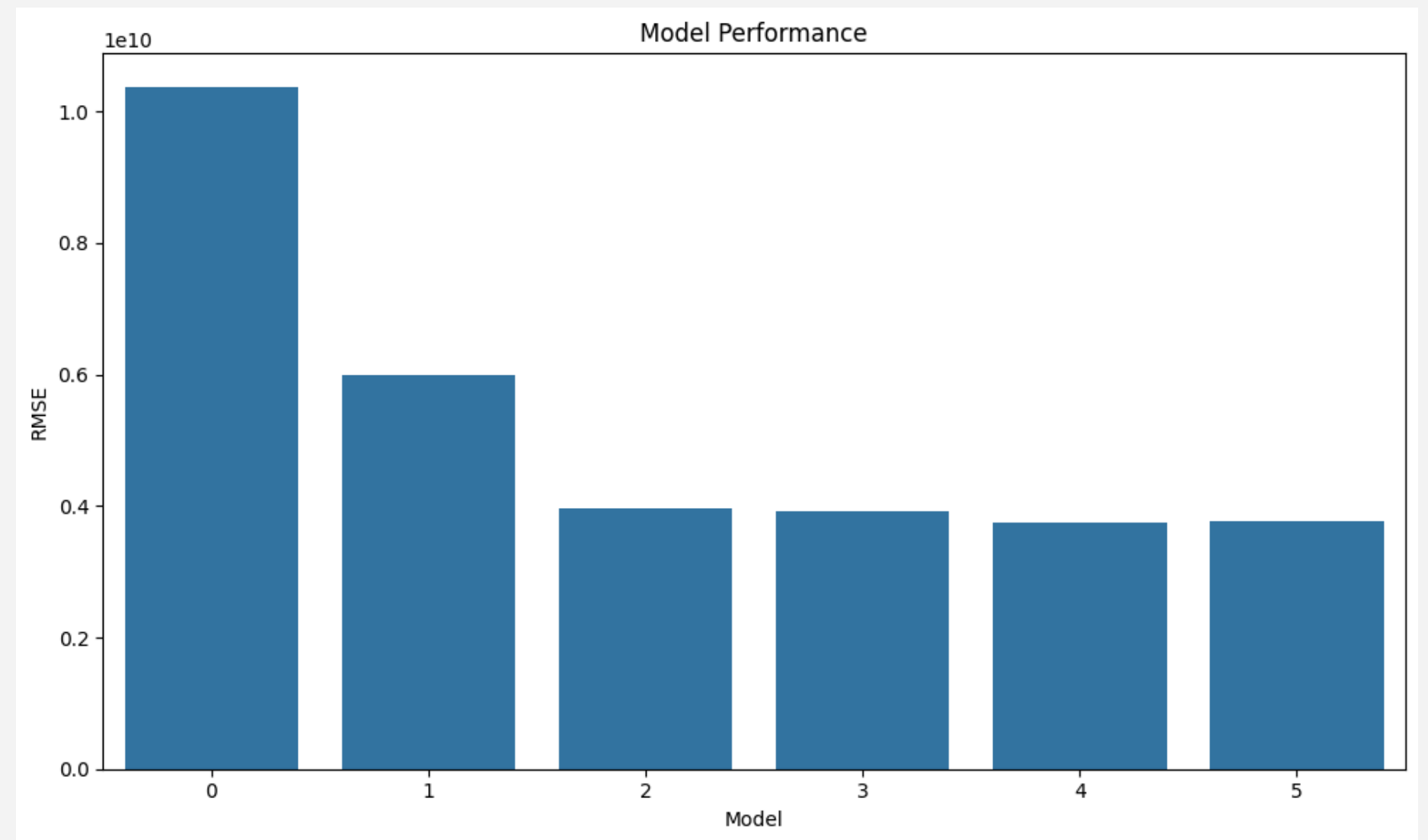
- El Precio mínimo es de 20000€, mientras que el máximo es de 17000000€, con una media de 438973€



Conclusiones

Como se puede observar, algunos modelos consiguen una precisión bastante más acertada que otros

(De izquierda a derecha: Linear Regression, Decission Tree, Random Forests, XGboost, LightGBM y Catboost)



Clasificación

Al observar cómo se comportarían los modelos clasificando por rangos de precio, pude llegar a algunas conclusiones

- Con 5 variables, sólo Random Forest fue capaz de superar el 70% de Accuracy
- Decision Tree se acercó bastante con un 68%
- KNN fue el que peor resultados dio, con un 59%

