


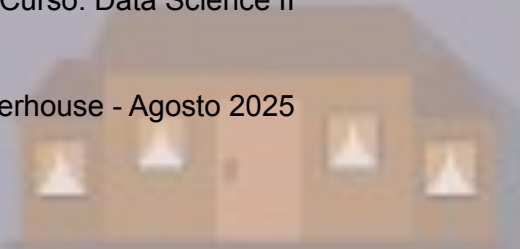
# Predicción de Precios de Casas con Machine Learning



Nombre : Danielle Baeza

Profesor : Joaquin Salas

Curso: Data Science II



Coderhouse - Agosto 2025



**CODERHOUSE**



**Objetivo:** predecir el precio de una vivienda usando características estructurales

- ¿Por qué este problema es relevante?

Las inmobiliarias constantemente ofrecen precios atractivos a través de sus campañas publicitarias, destacando beneficios como ubicación privilegiada, financiamiento accesible y alta plusvalía. Sin embargo, es fundamental analizar diversos factores antes de tomar una decisión, como la situación del mercado, las condiciones de compra y los costos asociados.

- ¿A quién le sirve esta predicción?

Este análisis puede ser útil para inmobiliarias, compradores o planificadores urbanos.

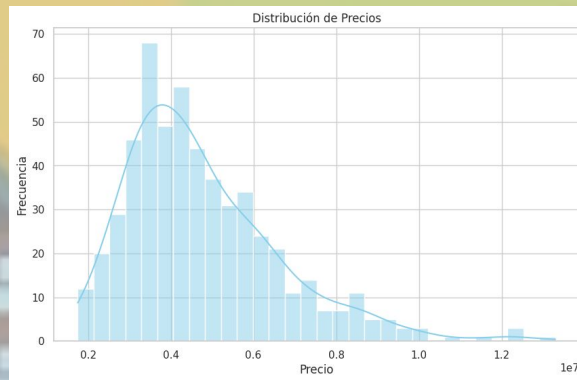


## Análisis Exploratorio de Datos (EDA)

### Distribución de precios

El histograma muestra que la mayoría de las viviendas tienen precios entre 2 millones y 5 millones

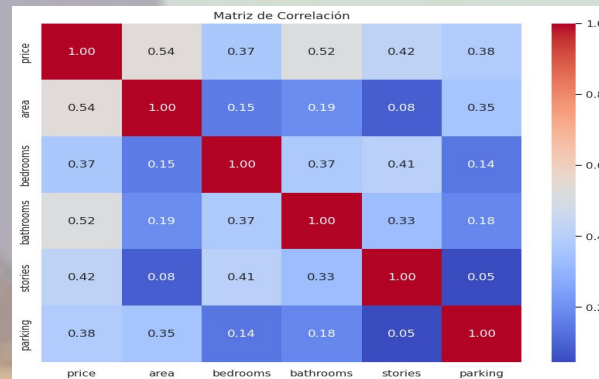
Hay una cola hacia la derecha, lo que indica una distribución asimétrica positiva (algunos precios muy altos).



### Correlaciones

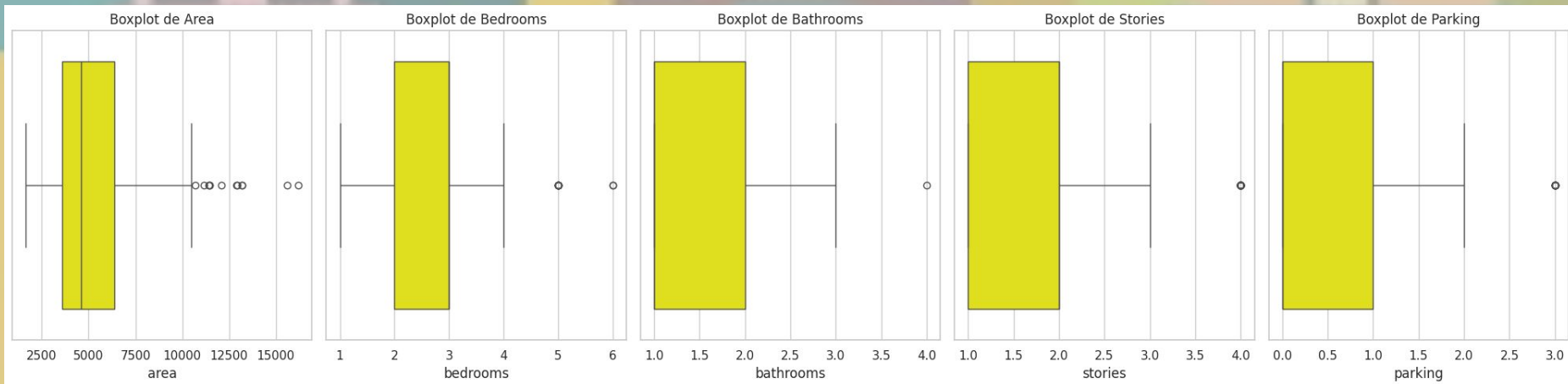
Las variables más correlacionadas con el price son:

1. área (0.53)
2. bathrooms (0.51)
3. bedrooms (0.47)
4. stories (0.41)
5. parking también presenta una correlación moderada.

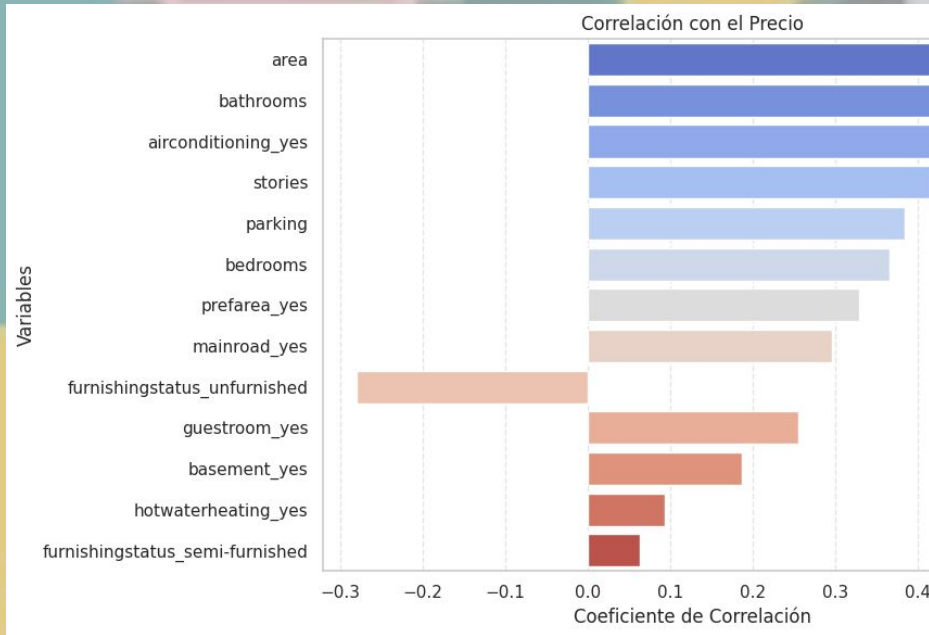


## Outliers por Variable

- área y bathrooms tienen valores extremos notables.
- parking tiene muchas observaciones en 0–1 y algunos outliers en 3–4.
- stories se concentra en 1 y 2, pero hay algunos valores altos (hasta 4).
- bedrooms tiene algunos casos con 6 o más dormitorios, lo que puede considerarse inusual



# Correlación precio



## Correlación favorable

- Área y baños se vinculan con el precio: propiedades más grandes y con más baños tienden a costar más.
- Otras variables positivas importantes: aire acondicionado, número de pisos (stories), estacionamiento, dormitorios, zona preferente (prefarea\_yes) y cercanía a la avenida principal (mainroad\_yes), lo que sugiere que comodidad, tamaño y ubicación elevan el precio.

## Correlaciones débiles/médias:

- Guestroom\_yes (~0.25) y basement\_yes (~0.18) también aportan, pero en menor medida.
- Hotwaterheating\_yes y furnishingstatus\_semi-furnished tienen efectos muy pequeños.

## Correlación negativa:

- Furnishingstatus\_unfurnished (~-0.27) indica que las propiedades sin amueblar tienden a tener precios más bajos respecto al baseline.



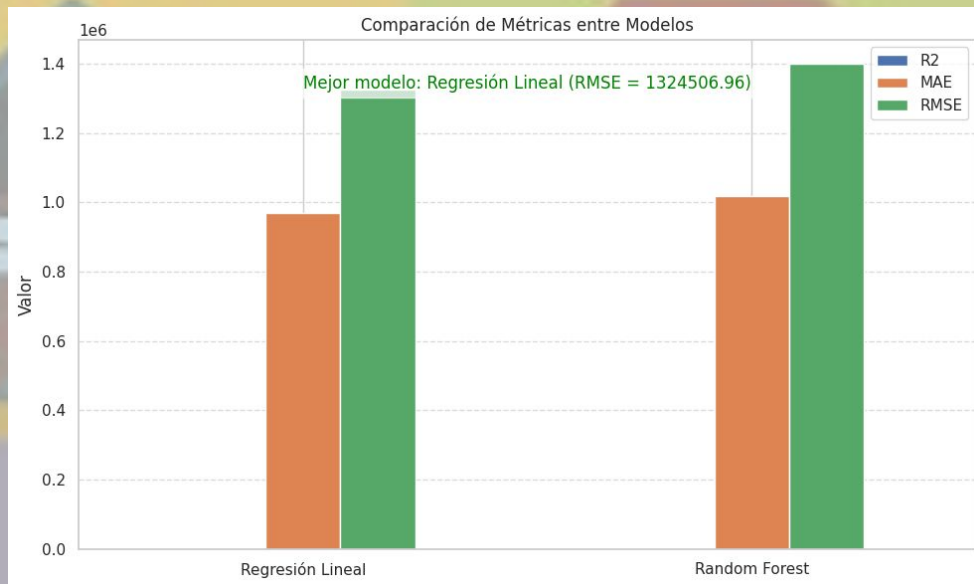
## Entrenamiento y Testeo de Modelos

1. Regresión Lineal
2. Random Forest

- $R^2$ : Aunque no se ve claramente en el gráfico, parece similar para ambos modelos.
- **MAE**: La Regresión Lineal tiene un MAE ligeramente menor, lo que indica que en promedio sus predicciones están más cerca del valor real.
- **RMSE**: La Regresión Lineal también tiene un RMSE más bajo (**1.324.506,96**), lo que sugiere mejor rendimiento al penalizar errores grandes.

Esto implica que, según la métrica RMSE, la Regresión Lineal supera al modelo de Random Forest para este conjunto de datos.

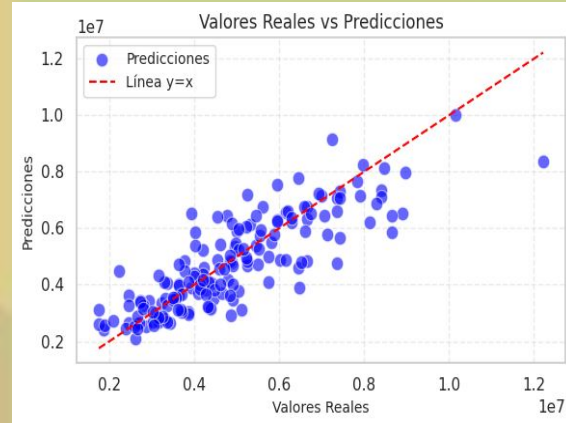
Mejor modelo: Regresión Lineal (RMSE = 1324506.96)



## Conclusiones

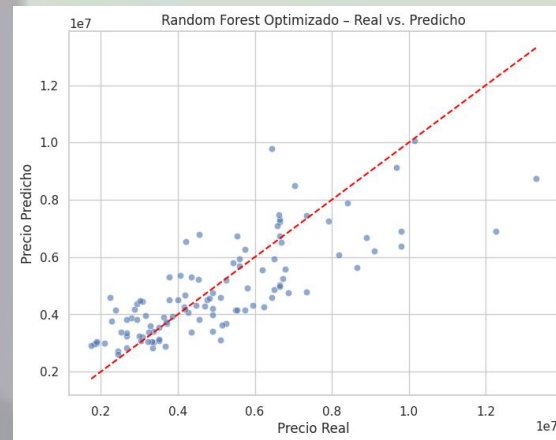
### Regresión Lineal

- Los puntos están más concentrados alrededor de la línea roja ( $y = x$ ).
- Esto sugiere que las predicciones están bastante alineadas con los valores reales.
- La dispersión es relativamente baja, lo que implica mayor precisión y menor error.
- Comportamiento más uniforme a lo largo de todo el rango de precios.
- Ofrece predicciones más cercanas al valor real, especialmente en todo el rango de precios.



### Random Forest

- Aunque sigue la tendencia general, muestra mayor dispersión, especialmente para precios más altos.
- Muchos puntos están por debajo de la línea ideal, lo que indica una subestimación sistemática del precio por el modelo.
- El modelo parece tener más dificultad para ajustar correctamente en los valores más extremos.
- Tiene mayor variabilidad y subestima con más frecuencia los precios altos.
- 



**Regresión Lineal** es más eficaz.