

Johana Rátiva

Juan Manuel Rivera

## **Ciencia de datos 2025-2**

### **Proyecto 2: predicción de valores de inmuebles**

HabitAlpes es una empresa inmobiliaria que requiere una herramienta para el cálculo automático de avalúos de apartamentos en venta. Se espera que esta herramienta disminuya el tiempo que toma realizar un avalúo de 6 horas a 1 hora. Estimamos que, en caso de usarse, el modelo generaría un ahorro de \$9.880 pesos por cada avalúo realizado.

En este informe se mostrarán los resultados del entrenamiento de dicha herramienta.

## Caracterización de los datos

Para el desarrollo de la herramienta se contaban con avalúos previos de inmuebles en Bogotá, con un total de 43.013 avalúos. Los datos fueron filtrados dejando únicamente los avalúos de apartamentos en venta. Además, se decidió trabajar únicamente con apartamentos usados.

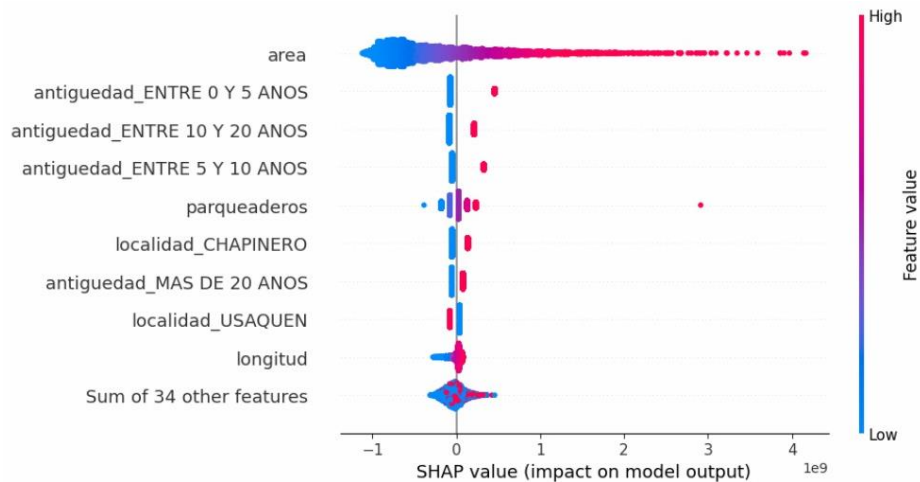
La mayor parte de apartamentos corresponde a estratos 4, 5 y 6, ubicados las localidades de Usaquén, Teusaquillo y Suba. Además, la mayor parte tiene más de 10 años.

Dado que había apartamentos con precios inusuales, se ajustaron los precios de algunos predios acorde con la moda del barrio en el que estaban ubicados.

## Modelo propuesto

Se utilizó una regresión lineal para estimar el precio de venta de los apartamentos.

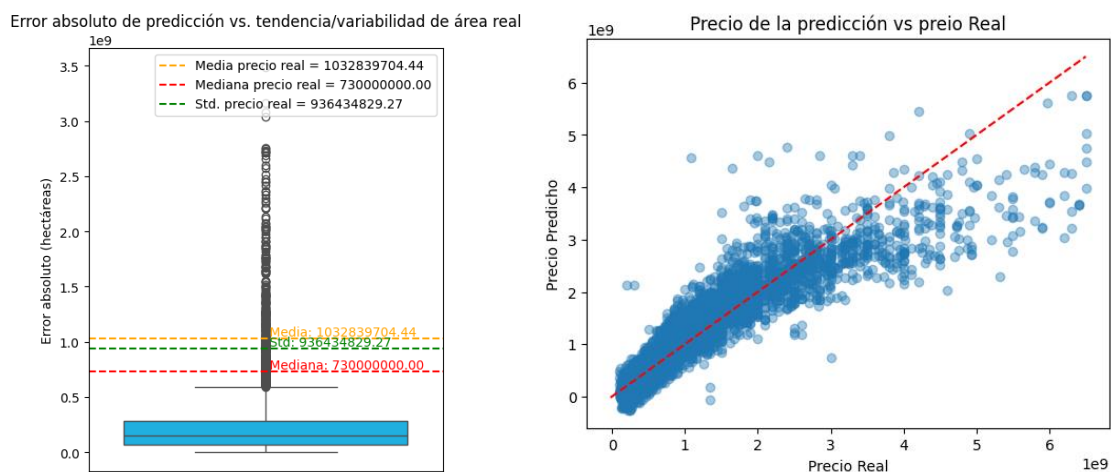
El modelo elegido calcula un mayor valor para apartamentos de mayor área que tengan parqueaderos ubicados en el oriente, especialmente en la localidad de Chapinero (ver figura 1). Además, entre menor tiempo de construido tenga, será más costoso.



*Figura 1. Efecto de las variables con mayor impacto en el modelo*

El modelo obtenido tuvo un error promedio de aproximadamente 380 millones de pesos. Teniendo en cuenta que el conjunto de datos tiene una distribución de precios dispersa (desviación estándar = 936 millones), el error de las predicciones que realiza el modelo se encuentra dentro del rango que tienen los datos originalmente (ver figura 2).

Si bien el modelo tiende a sobreestimar los precios de algunos inmuebles, el 56% de las predicciones no subestiman el precio por más de 20 millones. Adicionalmente, para inmuebles de más de 3.000 millones el valor tiende a subestimarse en más del 90% de los casos.



*Figura 2. Comparación de los errores con la mediana, la media y la desviación estándar de los datos (izquierda) y comparación del precio predicho vs el precio real (derecha). Las predicciones por debajo de la línea roja requieren un avalúo presencial.*

## Valor para el negocio

Dado que el modelo puede ahorrar 5 horas por avalúo, esto supone un ahorro de \$47.500 pesos por cada avalúo donde se use.

Cuando la predicción sea menor a 20 millones del valor del predio se prevé que se solicitará un avalúo presencial, este avalúo supondría un costo adicional de 9 horas para el evaluador (\$85.500 pesos).

Sin embargo, teniendo en cuenta que este avalúo se pediría en el 44% de los avalúos, se estima que habría un ahorro de \$9.880 pesos por avalúo. Se espera que en 5 meses se recupere la inversión del modelo.

## Conclusiones e *insights*

El modelo subestima el costo de los apartamentos en el 44% de los casos. Esto supondría un aumento en el número de avalúos presenciales, pero dado que reduce el tiempo que toma hacer cada avalúo en un 84%, a largo plazo generará un beneficio.

Aunque haya un potencial beneficio, las estimaciones tienen un margen de error considerable, por lo que se recomienda continuar con la experimentación antes de llevarlo a producción.

Se recomienda limitar el uso del modelo para apartamentos de menos de 3.000 millones de pesos, pues tiende a subestimar más estos apartamentos.

Se recomienda aumentar la muestra de apartamentos de estratos 2 y 3 ubicados hacia el sur de la ciudad, con el objetivo de disminuir el sesgo que presenta el modelo hacia dichos apartamentos.