

Informe Ejecutivo

- Objetivo
- Entendimiento de los datos
- Modelamiento
- Interpretabilidad del modelo
- Generación de valor
- Insights/ Estrategias recomendadas

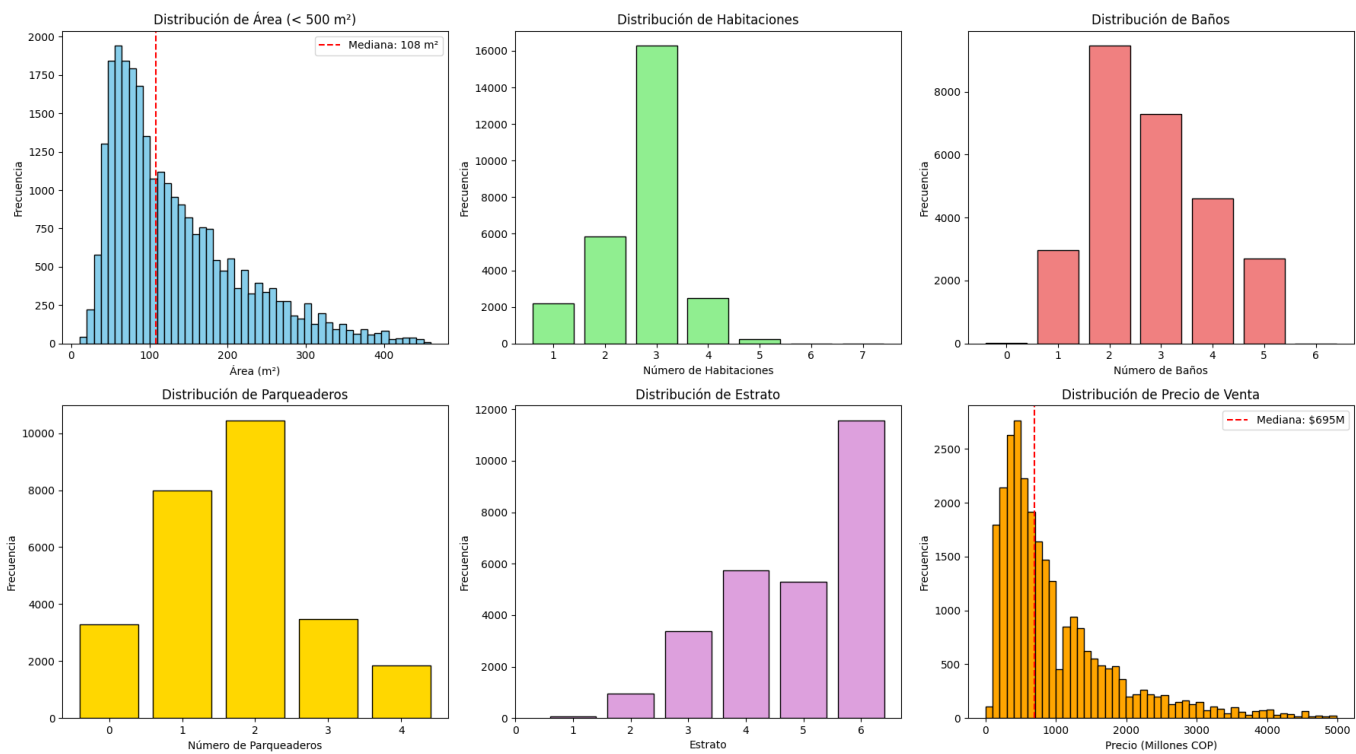
Objetivo

Este informe tiene como propósito ofrecer a HabitAlpes mun conjunto de estrategias que los ayuden en el negocio inmobiliario. Además, de mostrar los hallazgos e insights más importantes en términos del modelo y los resultados.

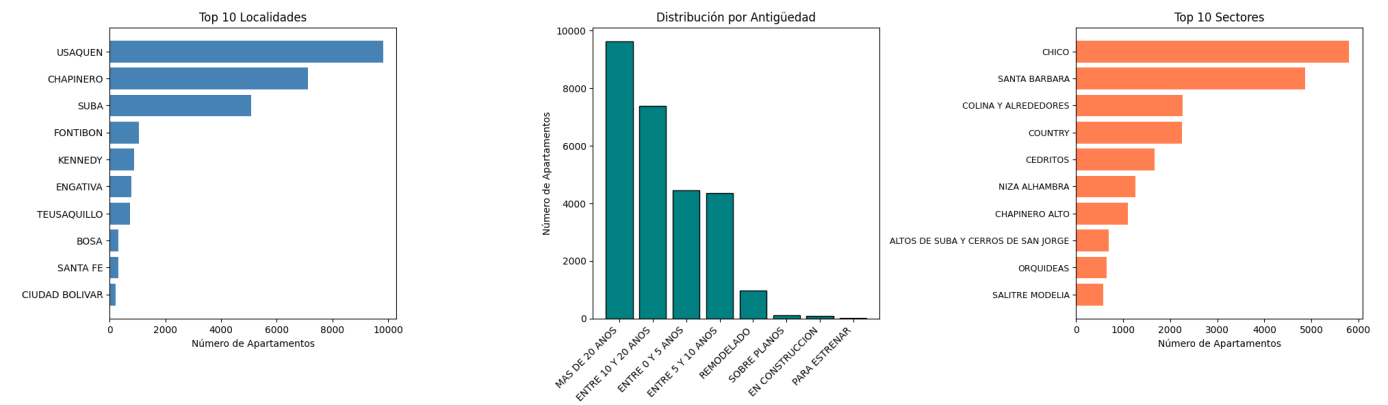
Entendimiento de los datos

Para el desarrollo del modelo predictivo de precios de venta de apartamentos, se contó con un conjunto de datos que incluye variables estructurales y de ubicación.

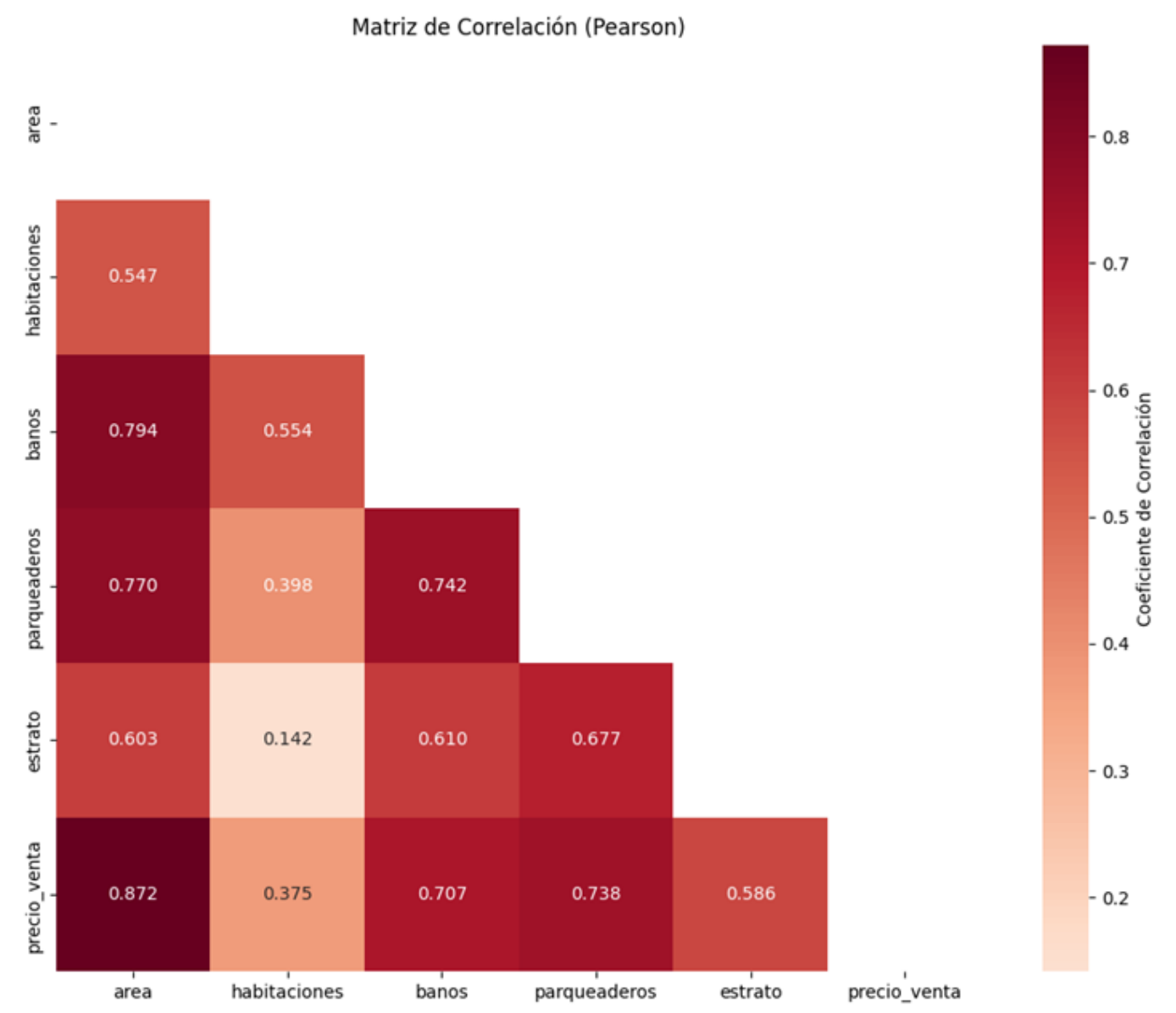
Las variables estructurales consideradas fueron: área (en metros cuadrados), número de habitaciones, número de baños, número de parqueaderos y estrato socioeconómico.



En cuanto a la ubicación, se incluyeron variables categóricas que representan la localidad y el sector donde se encuentra cada apartamento. Adicionalmente, se consideró la variable categorica de antigüedad del inmueble.



La matriz de correlación de Pearson entre las variables estructurales del apartamento y el precio de venta muestra que todas las relaciones son positivas, destacándose una correlación muy fuerte entre **área y precio_venta** (0,872), lo que confirma que los apartamentos más grandes tienden a ser más costosos. También se evidencian correlaciones altas entre el precio y el número de **parqueaderos** (0,738), así como con el número de **baños** (0,707) y el **estrato** (0,586). La variable **habitaciones** presenta una correlación moderada con el precio (0,375), menor que la del área, lo que sugiere que el metraje total pesa más que el simple conteo de cuartos. Adicionalmente, se observa una fuerte asociación entre las propias variables físicas, como área con baños (0,794) y parqueaderos (0,770), lo que refleja que los apartamentos más amplios suelen ofrecer más servicios y parqueaderos.



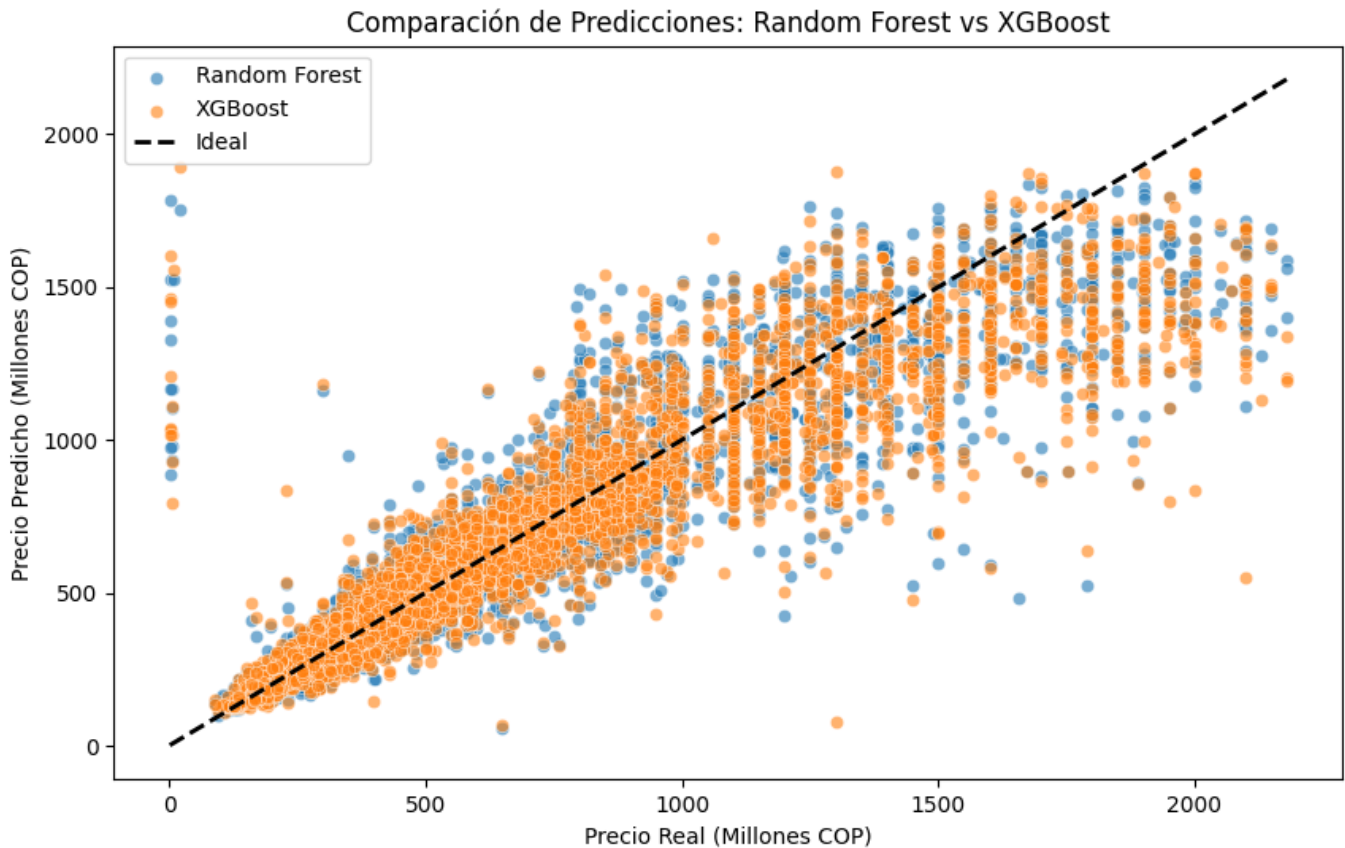
Modelamiento

Se entrenaron y evaluaron dos modelos de regresión para estimar el precio de venta de apartamentos en Bogotá: Random Forest y XGBoost. En el conjunto de prueba se obtuvieron las siguientes métricas:

Aquí está la tabla en Markdown:

Métrica	Random Forest	XGBoost
MAE	117,873,838	120,763,679
RMSE	193,264,685	202,673,275
R²	0.835	0.8185

Se seleccionó Random Forest debido a que tiene menores errores absolutos y mejor capacidad explicativa (menor MAE y RMSE, mayor R².)



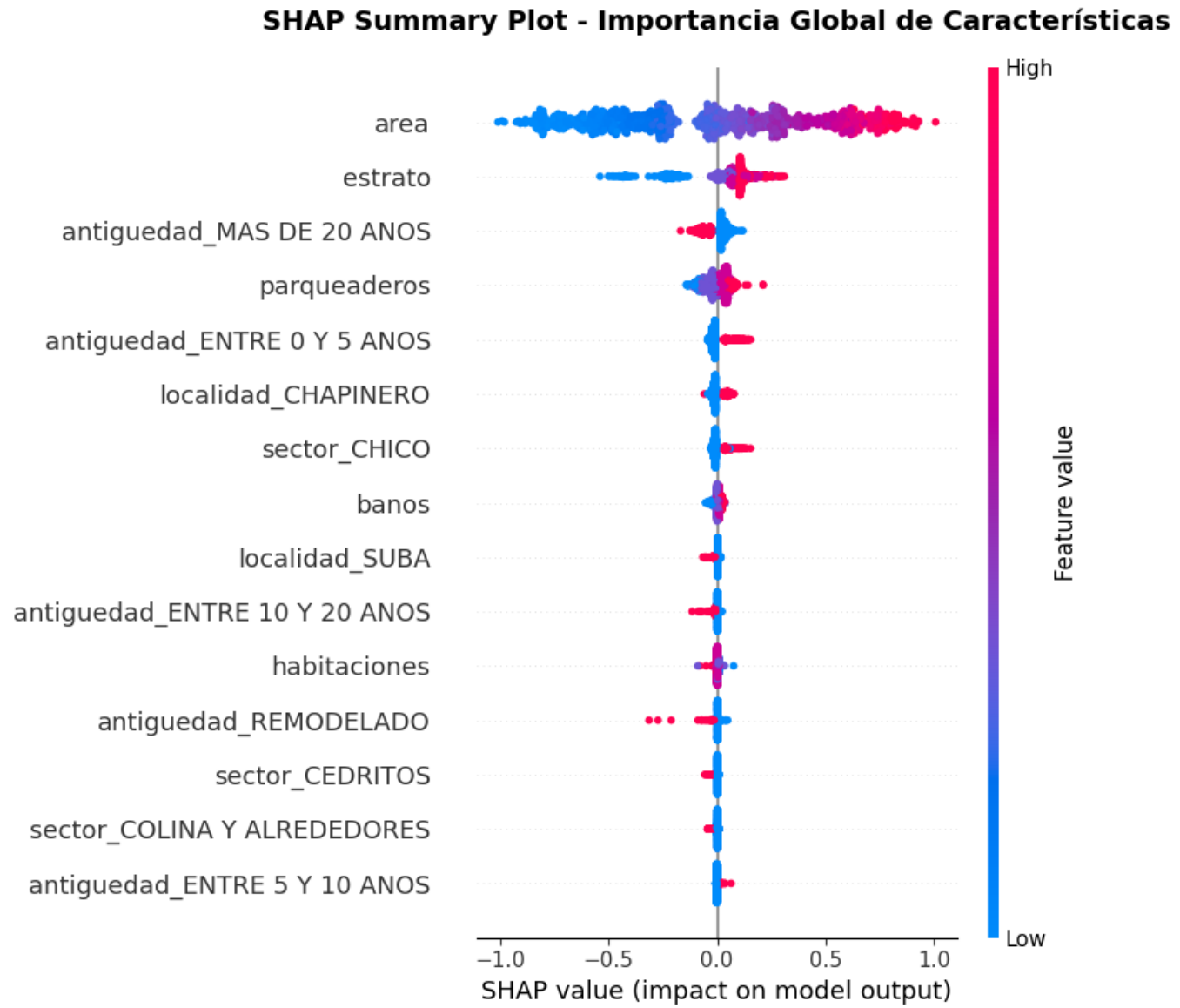
La grafica anterior muestra que la mayoría de los puntos se concentran alrededor de la línea ideal, lo que indica que los modelos capturan adecuadamente la relación entre las variables explicativas y el precio. Se observa, sin embargo, una mayor dispersión a precios muy altos o muy bajos, donde los modelos tienden a subestimar o sobreestimar algunos apartamentos, lo cual es esperable en segmentos con menor cantidad de datos y mayor heterogeneidad.

Interpretabilidad del modelo

Se utilizó SHAP (SHapley Additive exPlanations) y LIME (Local Interpretable Model-agnostic Explanations) para interpretar las predicciones del modelo Random Forest.

El SHAP Summary Plot detalla cómo varía el impacto en la predicción según el valor específico de cada característica. En este gráfico, cada punto representa un apartamento y el eje X corresponde al valor SHAP (positivo cuando la característica aumenta el precio predicho y negativo cuando lo disminuye), mientras que el color indica si el valor de la característica es bajo (azul) o alto (rojo). Se observa que áreas grandes y estratos altos se asocian sistemáticamente con valores SHAP positivos, incrementando el precio estimado, mientras que áreas pequeñas, estratos bajos y antigüedad mayor a 20 años tienden a generar valores SHAP negativos, reduciendo el valor predicho. Estos resultados confirman que el modelo fundamenta sus decisiones principalmente en el tamaño del apartamento, el estrato y la antigüedad, complementados por el número de

parqueaderos y la zona de la ciudad.



Generación de valor

La implementación del modelo de estimación automática de precios permite cuantificar de forma directa la generación de valor para HabitAlpes. En el conjunto de prueba se identificaron 1.623 apartamentos subvaluados por más de 20 millones de COP, lo que corresponde al 39,12 % de los casos. Bajo los supuestos de costo de tiempo del perito y de los avalúos presenciales, el retorno esperado por cada estimación realizada con el modelo es de **\$1.572.847 COP**. Frente a una inversión total en el proyecto de **\$17.876.000 COP**, el punto de equilibrio se alcanza tras aproximadamente **662 estimaciones**, es decir, en cerca de **1,3 meses** de operación, por lo que el modelo se recupera rápidamente y empieza a generar dividendos netos en el muy corto plazo.

Al analizar el desempeño por segmento de precio, se observa que los apartamentos baratos representan el 24,51 % del portafolio, con un precio promedio de **\$246,6 millones**. En este grupo el modelo presenta un MAPE extremadamente alto (618,38 %) y un MAE de **\$53,3 millones**, lo que indica predicciones poco fiables; además, el error asociado a sobreestimaciones (**\$43,9 millones**) es mucho mayor que el asociado a subestimaciones (**\$9,3 millones**), evidenciando una tendencia a inflar el precio de los inmuebles más económicos y, por tanto, riesgo de ofrecer valores poco competitivos. En el segmento medio, que concentra

el 49,82 % de los apartamentos con un precio promedio de **\$631,4 millones**, el rendimiento del modelo es sensiblemente mejor: el MAPE es de 14,31 % y el MAE de **\$91,2 millones**, con errores de subestimación (**\$40,8 millones**) y sobreestimación (**\$50,4 millones**) relativamente balanceados, lo que sugiere un comportamiento más estable y simétrico.

Finalmente, el segmento de apartamentos caros, que representa el 25,67 % del total y tiene un precio promedio de **\$1.429,5 millones**, también presenta un MAPE moderado (15,67 %) pero con un patrón de error distinto: el MAE asciende a **\$231,3 millones** y el error agregado por subestimaciones (**\$187,8 millones**) es muy superior al de sobreestimaciones (**\$43,5 millones**). Esto indica que, aunque el error relativo es aceptable, el modelo tiende a subvalorar de manera importante los inmuebles de mayor valor, lo que en la práctica se traduce en un número elevado de casos que probablemente requerirán avalúo presencial y, por ende, en un incremento de los costos operativos en este segmento. En conjunto, los resultados muestran que el modelo ya genera valor económico neto a nivel global, pero también señalan la necesidad de priorizar mejoras en los extremos de la distribución de precios, especialmente en los apartamentos baratos y en la corrección de la subvaloración de los inmuebles caros.

Insights/ Estrategias recomendadas

Con base en los hallazgos del análisis y el desempeño del modelo, se proponen las siguientes estrategias para maximizar el valor generado por HabitAlpes:

- **Priorizar despliegue del modelo en apartamentos de precio medio**, donde el error es bajo y el volumen es mayor, para maximizar ahorro de tiempo y retorno.
- **Adoptar un flujo híbrido modelo–perito**: automatizar decisiones en casos de alta confianza y enviar a revisión manual los de baja confianza o en segmentos extremos.
- **Establecer revisión obligatoria del perito en apartamentos baratos**, debido al MAPE extremadamente alto y al riesgo de sobreprecio.
- **Definir alertas automáticas para subestimaciones en apartamentos caros**, enviando estos casos a avalúo experto antes de publicar precios.
- **Concentrar campañas comerciales y de captación en unidades de estrato medio–alto, áreas amplias y antigüedad moderada**, donde el modelo es más preciso y el margen potencial es mayor.
- **Integrar explicaciones SHAP en la interfaz interna**, para que los peritos comprendan por qué el modelo sugiere cierto precio y puedan corregir sesgos.
- **Explorar nuevas líneas de negocio ligadas a inmuebles antiguos**, como servicios de remodelación, aprovechando que la antigüedad es un factor clave en la caída de precio.