

Avalúo automático de apartamentos para HabitAlpes

Ciencia de Datos Aplicada 2025-2

Juan David Valencia - 201728857

Juan Esteban Cuellar - 202014258

Contexto y objetivo

HabitAlpes debe valorar cientos de apartamentos al mes en Bogotá.

Hoy cada avalúo requiere ~6 horas de trabajo de un experto.

Si el precio sugerido se aleja mucho del “precio mental” del cliente, se generan fricciones y retrabajo.

El proceso es lento, costoso y poco escalable.



Objetivo del proyecto

Objetivo general:

Diseñar un modelo de datos que estime el valor de un apartamento en Bogotá y apoye al perito en su decisión.

KPIs principales:

- Reducir tiempo de valoración de 6 horas a 1 hora por apartamento.
- Lograr un $R^2 \geq 0,85$ en el conjunto de validación.
- Mantener un $MAE \leq \$160M$ COP en precio de venta estimado.
- Alcanzar que al menos 15–20% de las valoraciones queden dentro de $\pm \$20M$ COP del precio real.



Datos utilizados

43.013 inmuebles registrados en Bogotá con 46 variables.

Subconjunto con precio de venta: ~27.000 apartamentos.

Variables:

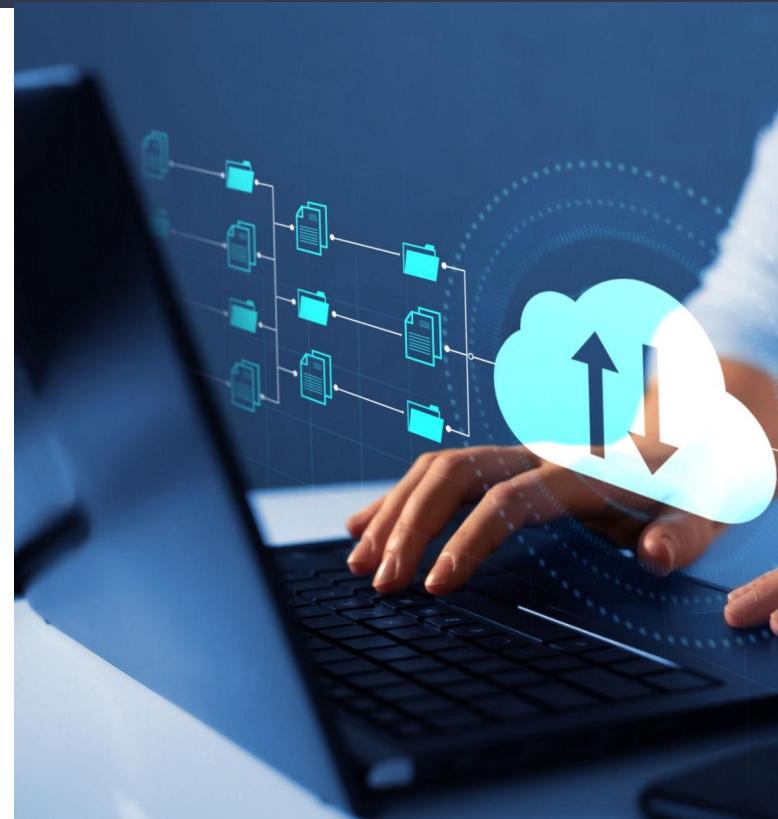
Ubicación: localidad, barrio, coordenadas.

Características físicas: área, habitaciones, baños, parqueaderos, antigüedad.

Socioeconómicas: estrato.

Amenidades: ascensor, piscina, gimnasio, vigilancia, etc.

Negocio: precio de venta, precio de arriendo, administración

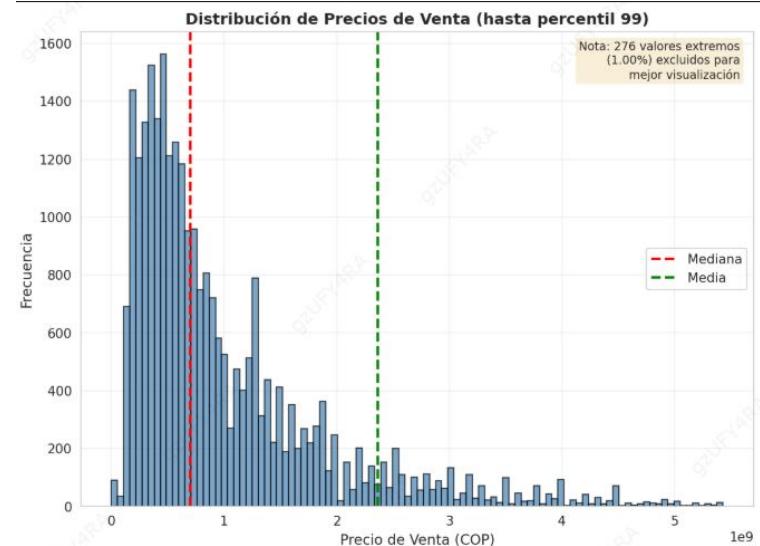


Exploración del mercado inmobiliario

Los precios de venta están fuertemente sesgados hacia la derecha.

Hay pocos apartamentos muy caros que “jalan” la cola de la distribución.

El estrato está fuertemente asociado al nivel de precio.

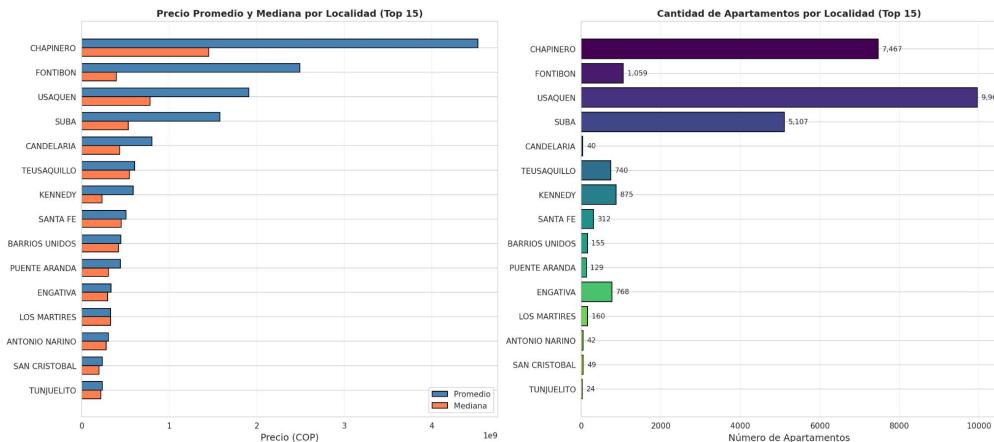


Variación geográfica del precio

Zonas del norte como Usaquén y Chapinero concentran precios más altos.

Localidades del sur y occidente tienden a valores más bajos.

El patrón espacial es coherente con la intuición del mercado inmobiliario



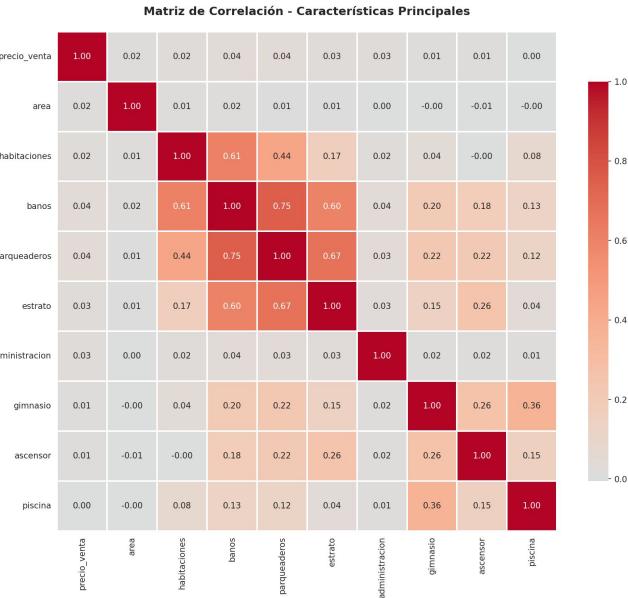
Enfoque de modelado

Filtramos inmuebles con precio de venta válido y sin outliers extremos.

Dividimos en: 60% entrenamiento, 20% prueba, 20% validación.

Probamos varios algoritmos de regresión (lineales y basados en árboles).

Seleccionamos el mejor modelo según MAE, RMSE y R².



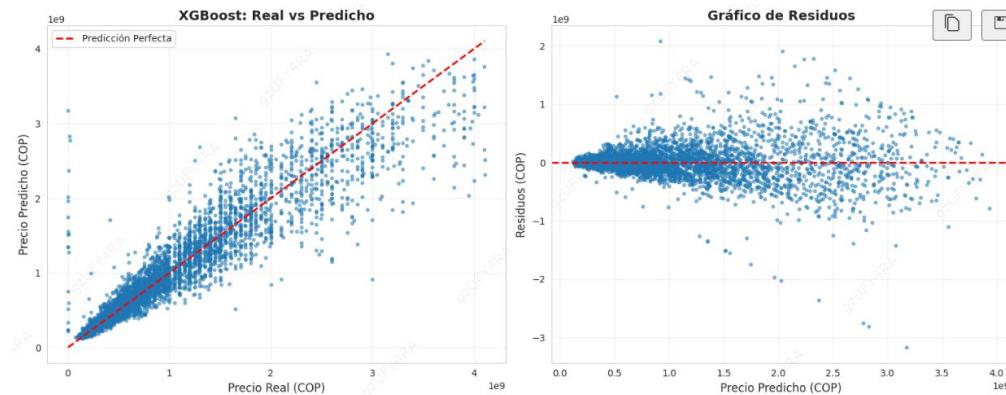
Resultados del mejor modelo

Mejor desempeño: XGBoost Regressor.

En el conjunto de validación:

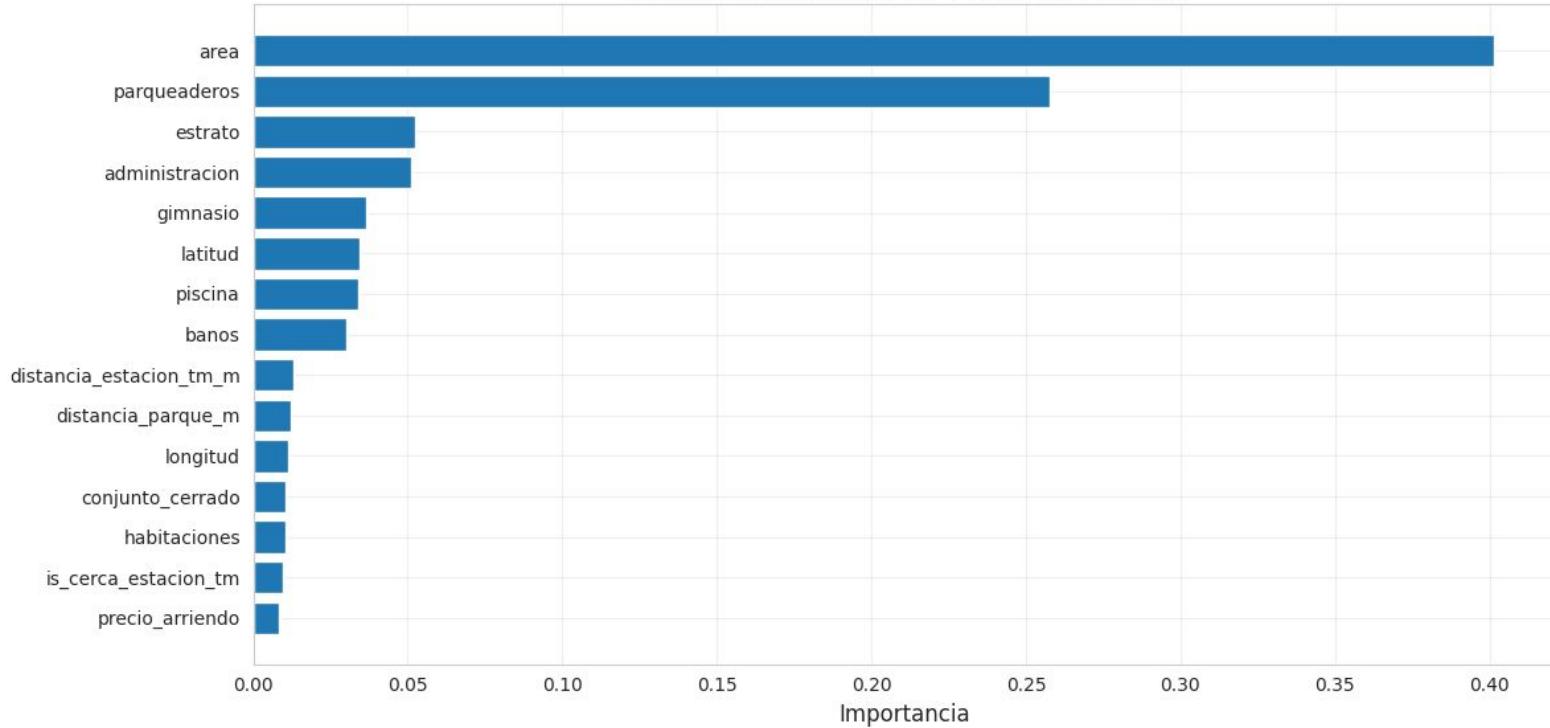
Métricas de XGBoost (Validación):

MAE (Error Absoluto Medio):	\$153,476,108
RMSE (Raíz Error Cuadrático):	\$266,454,454
R ² (Coef. Determinación):	0.8875
MAPE (Error % Absoluto Medio):	131.00%

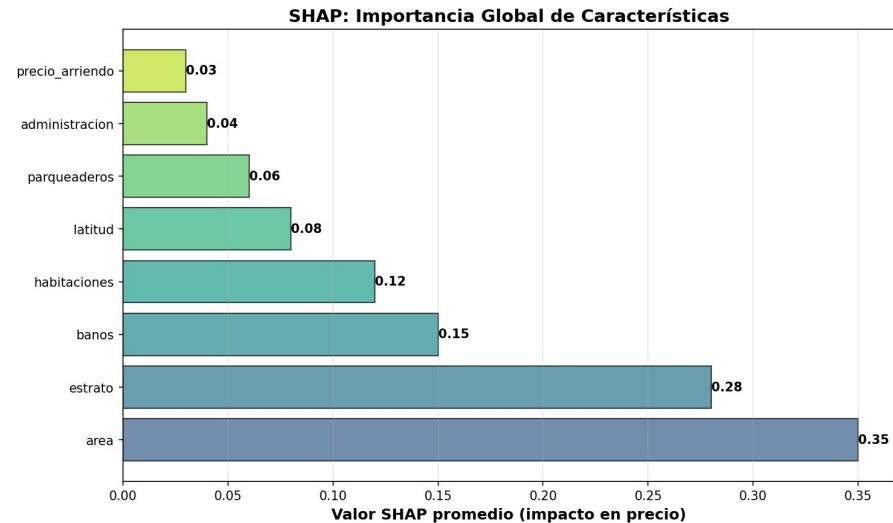
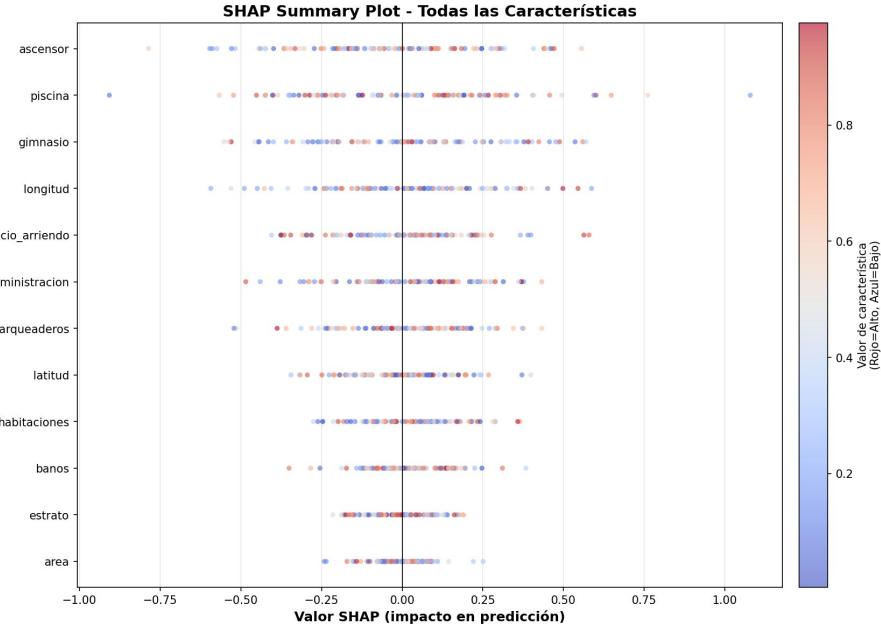


¿Qué variables mandan?

Top 15 Características Más Importantes



¿Qué variables mandan?

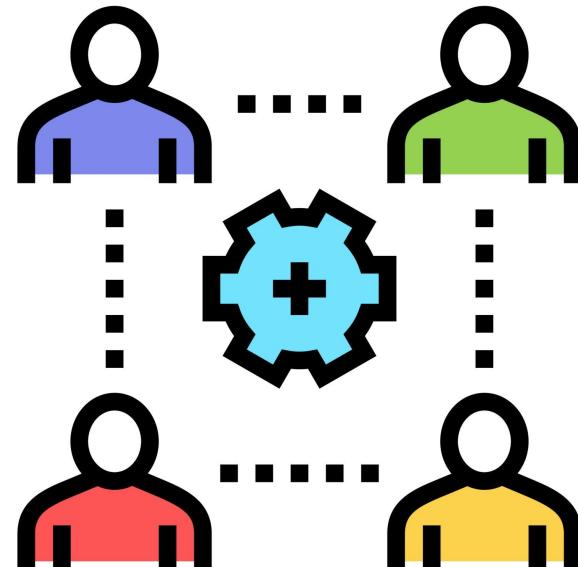


¿Qué variables mandan? Key Insights

- El área del apartamento es el factor que más pesa: a mayor m², mayor precio.
- La ubicación (localidad y estrato) es el segundo gran driver: zonas “premium” suben el valor en decenas de millones.
- El número de baños suele ser más relevante que el de habitaciones, porque refleja nivel de comodidad y “lujo”.
- Las amenidades (piscina, gimnasio, ascensor, etc.) suman valor, especialmente en estratos altos, aunque su efecto es más moderado.
- La proximidad a transporte y parques también influye: estar cerca de TransMilenio o zonas verdes tiende a aumentar el precio, con impacto que cambia según el sector de la ciudad.

Ejemplo de explicación individual

- Seleccionamos un apartamento de alto valor como ejemplo.
- El modelo parte de un valor base y suma/resta aportes de cada característica.
- Factores que aumentan el precio: mayor área, estrato alto, más baños y parqueaderos.
- Factores que reducen el precio: mayor antigüedad, piso bajo sin vista/terraza, etc.



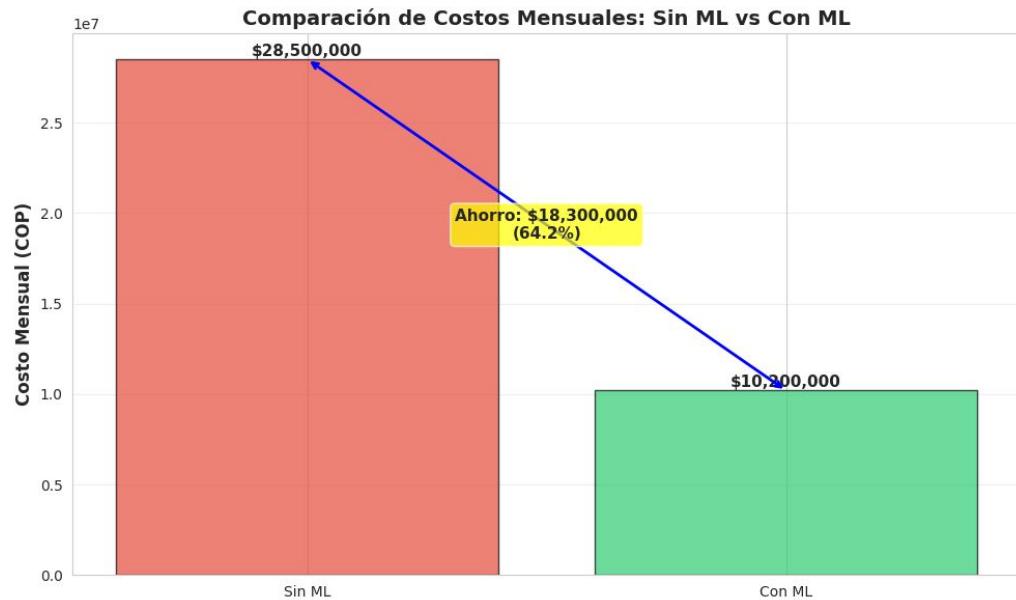
Impacto en costos: sin modelo vs con modelo

Escenario actual (sin modelo):

- 6 horas de experto por avalúo × \$9.500 COP/h.

Con el modelo:

- 1 hora de experto + uso del modelo.
- Con 500 apartamentos/mes y 80% de adopción:
- Ahorro neto mensual ≈ \$18,3M COP.



ROI y punto de equilibrio

- Inversión inicial estimada: \$4M COP de desarrollo + costos operativos.
- Con el ahorro mensual proyectado, se alcanza el punto de equilibrio en X meses (usa el valor que te da el notebook).
- A 3 años, el ROI estimado supera el 2000%.



Conclusiones y próximos pasos

- El modelo de avalúo automático logra buena precisión global ($R^2 \approx 0,89$) en un mercado muy heterogéneo.
- La herramienta es explicable gracias a SHAP/LIME, lo que facilita su adopción por parte de los peritos.
- El impacto económico proyectado es alto (ahorros mensuales y ROI muy atractivo).

Próximos pasos:

- Integrar el modelo en una herramienta interna para el perito.
- Ajustar el modelo por segmentos (estrato, zona, tipo de inmueble).
- Medir satisfacción de los usuarios internos y de los clientes finales.

Cierre

Este análisis sienta las bases para un producto de datos capaz de predecir el crecimiento y optimizar la inversión en retención.

Gracias por su atención.