

REPORTE DIRECTIVO

Análisis del precio en la vivienda en California- Hallazgos

Autor: Juan Manuel Alvarado Sandoval

Fecha: 17 de febrero de 2026

0. Índice

1. Resumen Ejecutivo	2
2. El Problema que Buscamos Resolver	2
3. Lo que los Datos nos Contaron	3
3.1. El dinero atrae al dinero: el ingreso lo explica casi todo	3
3.2. California tiene dos mercados, no uno	3
3.3. El tamaño no lo es todo	3
3.4. Una alerta que no puede ignorarse: los datos tienen un techo	4
4. El Modelo Predictivo: Qué tan Confiable es	4
4.1. Dos modelos, un ganador claro	4
4.2. ¿Qué significa esto en la práctica?	4
5. Riesgos que Deben Conocerse Antes de Decidir	5
6. Recomendaciones y Próximos Pasos	5
6.1. Lo que se recomienda hacer ahora	5
6.2. Lo que se recomienda construir a mediano plazo	6
6.3. Lo que no debe hacerse	6

1. Resumen Ejecutivo

Entender qué hace que una vivienda valga lo que vale no es trivial. Detrás de cada precio hay una combinación de factores: quiénes viven en la zona, dónde está ubicada, cuánto tiempo llevan en pie esas construcciones. Para una organización que toma decisiones sobre propiedades —ya sea para invertir, valorar, financiar o planificar— contar con una herramienta que descifre esa lógica tiene un valor estratégico concreto.

Este análisis examinó más de **20,000 bloques residenciales del estado de California** y llegó a tres conclusiones centrales:

- **El nivel de ingresos de una comunidad es, con diferencia, el factor que más explica el valor de sus viviendas.** Más que el tamaño, más que la antigüedad, más que la densidad poblacional: donde vive gente con mayores ingresos, las viviendas valen más.
- **La ubicación geográfica traza fronteras de valor muy claras.** Las zonas costeras concentran las propiedades más valiosas. El interior del estado se comporta como un mercado distinto, con valores significativamente menores.
- **Se construyó un modelo predictivo con capacidad demostrada para estimar el valor de una vivienda.** El modelo seleccionado, tras comparar dos alternativas, mostró un comportamiento estable y confiable: sus estimaciones sobre datos que nunca había visto fueron incluso mejores que sus propias proyecciones internas, lo que indica solidez real, no un buen desempeño fabricado sobre datos conocidos.

Recomendación: incorporar este modelo como herramienta de apoyo en procesos de valoración, segmentación de mercado y análisis de riesgo, con las precauciones que se detallan en este documento.

2. El Problema que Buscamos Resolver

Valorar una propiedad de manera objetiva es más difícil de lo que parece. Dos casas similares en papel —mismo tamaño, misma antigüedad— pueden tener precios radicalmente distintos dependiendo de dónde están ubicadas y quiénes son sus vecinos. Este fenómeno hace que la valoración manual sea costosa, lenta y susceptible a criterios subjetivos que varían de un evaluador a otro.

La pregunta que guió este análisis fue directa: **¿es posible construir un modelo que, a partir de información objetiva y verificable sobre un bloque residencial, estime con precisión razonable el valor de las viviendas en ese bloque?**

La respuesta, como muestra este reporte, es que sí. Y los hallazgos del camino hacia esa respuesta son en sí mismos información estratégica valiosa.

3. Lo que los Datos nos Contaron

3.1. El dinero atrae al dinero: el ingreso lo explica casi todo

De todas las características analizadas, el nivel de ingresos de los residentes de un bloque resultó ser, con amplia ventaja, el indicador más fuerte del valor de sus viviendas. Esto no es sorprendente en sí mismo, pero la magnitud de esa relación sí lo es: **conocer el ingreso mediano de una zona permite anticipar su nivel de precios inmobiliarios con una precisión que ninguna otra variable logra por sí sola.**

La implicación práctica es clara: en cualquier ejercicio de valoración, segmentación o análisis de riesgo, el perfil socioeconómico del entorno no puede ser ignorado.

Figura 1: A mayor ingreso mediano de la zona, mayor valor de las viviendas. La relación es la más clara y consistente del análisis.

3.2. California tiene dos mercados, no uno

El análisis geográfico reveló algo que los datos confirman con claridad: **California no tiene un mercado inmobiliario único, sino al menos dos mundos distintos.**

Las zonas costeras —especialmente las bahías y las áreas a menos de una hora del océano— concentran los valores más altos, la mayor densidad poblacional y la mayor variabilidad de precios. El interior del estado opera bajo una lógica diferente: precios más bajos, distribución más homogénea, menor presión de demanda.

Esta segmentación tiene implicaciones directas para cualquier estrategia que involucre propiedades en California: **aplicar criterios uniformes a zonas geográficamente distintas generará estimaciones sesgadas y decisiones equivocadas.**

Figura 2: Distribución geográfica del valor mediano de las viviendas. El gradiente entre costa e interior es estructural y consistente en toda la extensión del estado.

3.3. El tamaño no lo es todo

Podría esperarse que los bloques con más habitaciones, más hogares o más población tuvieran también las viviendas más valiosas. El análisis mostró que **esa relación no existe de forma directa.** El tamaño de un bloque residencial, medido en número de cuartos o de personas, dice muy poco sobre el valor de sus propiedades si no se considera en relación con cuántos hogares comparten ese espacio.

En otras palabras, la densidad relativa importa más que el volumen absoluto. Un bloque con muchas habitaciones pero también muchos hogares no es necesariamente más caro que uno más pequeño con menos familias.

3.4. Una alerta que no puede ignorarse: los datos tienen un techo

Durante el análisis se identificó que los registros originales fueron cortados artificialmente en un valor máximo fijo. Esto significa que **todas las propiedades que en realidad valían más que ese techo aparecen registradas con el mismo valor límite**, independientemente de su precio real.

Las consecuencias son concretas: el modelo aprende que los precios nunca superan ese umbral y, por lo tanto, **subestimará sistemáticamente el valor de propiedades de alto perfil**. Cualquier aplicación del modelo en el segmento de vivienda de lujo o alta gama debe considerar esta limitación de manera explícita.

4. El Modelo Predictivo: Qué tan Confiable es

4.1. Dos modelos, un ganador claro

Se desarrollaron y compararon dos modelos predictivos. Ambos pertenecen a una familia de algoritmos reconocida por su solidez en este tipo de problemas: construyen su estimación combinando cientos de reglas de decisión simples, aprendiendo de sus propios errores de forma iterativa.

El modelo ganador fue **LightGBM**, que superó a su competidor en la prueba más importante: estimar correctamente el valor de propiedades que nunca había visto durante su entrenamiento.

Cuadro 1: Desempeño comparativo de los modelos evaluados

Modelo	Error en validación	Error en datos nuevos	Interpretación
XGBoost	46,450	46,509	Estable. Desempeño consistente entre ambas pruebas.
LightGBM	46,104	45,057	Mejora al enfrentarse a datos nuevos. Mayor solidez.

Los valores representan el error promedio de estimación en las unidades de la variable objetivo. Un número menor indica mayor precisión.

4.2. ¿Qué significa esto en la práctica?

Lo más relevante del resultado no es el número en sí, sino lo que revela sobre el comportamiento del modelo. **LightGBM fue más preciso ante datos nuevos que ante los datos con los que fue entrenado**. Esto es poco frecuente y altamente deseable: indica que el modelo no memorizó patrones específicos del pasado, sino que aprendió reglas genuinas que aplican más allá de los datos conocidos.

En contraste, un modelo que se comporta muy bien en los datos de entrenamiento pero falla en datos nuevos —fenómeno conocido como sobreajuste— ofrece una falsa sensación de

seguridad. Ninguno de los dos modelos evaluados cayó en ese problema, lo que valida la solidez del proceso seguido.

5. Riesgos que Deben Conocerse Antes de Decidir

Todo modelo tiene límites. Usarlo sin conocerlos es el mayor riesgo que existe. Los siguientes puntos no son argumentos en contra del modelo, sino condiciones bajo las cuales sus resultados deben interpretarse con precaución adicional:

Las estimaciones son zonales, no individuales. El modelo trabaja con promedios por bloque residencial, no con propiedades específicas. Usarlo para valorar un inmueble concreto introduce un margen de error adicional que no está capturado en los resultados reportados.

El segmento de alta gama está subrepresentado. Como se señaló anteriormente, los datos tienen un techo artificial que distorsiona el aprendizaje del modelo en el extremo superior del mercado. Las estimaciones para propiedades de alto valor deben tratarse como referencias orientativas, no como cifras definitivas.

Zonas poco representadas generan estimaciones menos confiables. El análisis identificó que ciertas zonas geográficas —particularmente las áreas insulares— cuentan con muy pocos registros en los datos. El modelo tiene menos información sobre esas zonas y, en consecuencia, sus estimaciones allí son menos precisas.

El modelo apoya decisiones; no las toma. Ningún algoritmo, por sofisticado que sea, sustituye el juicio experto en transacciones de alto impacto. Su función es reducir la incertidumbre y objetivar el análisis preliminar, no automatizar decisiones que requieren contexto, experiencia y responsabilidad humana.

6. Recomendaciones y Próximos Pasos

6.1. Lo que se recomienda hacer ahora

Adoptar LightGBM como modelo de referencia para valoración asistida. Su desempeño superior y su comportamiento estable ante datos nuevos lo convierten en la opción más confiable de las dos evaluadas. Está listo para ser usado como primera referencia objetiva en análisis de propiedades dentro del alcance documentado.

Definir el contexto de uso antes de implementar. Antes de cualquier despliegue operativo, es fundamental delimitar con precisión para qué decisiones se usará el modelo y para cuáles no. Esto protege a la organización de un uso inadecuado que genere errores costosos.

6.2. Lo que se recomienda construir a mediano plazo

Actualizar el modelo con datos del mercado actual. Esta es la acción de mayor impacto para incrementar la utilidad práctica. Un modelo entrenado con información reciente tendría relevancia operativa directa, no solo analítica.

Enriquecer la información disponible. Incorporar datos sobre proximidad a servicios, calidad educativa del entorno, conectividad vial y dinámicas de inversión recientes aumentaría significativamente la capacidad predictiva del modelo y reduciría su margen de error.

Diseñar un mecanismo de monitoreo continuo. Los modelos envejecen junto con los mercados que describen. Establecer indicadores de alerta que detecten cuándo el modelo comienza a perder precisión permite actuar antes de que sus estimaciones se vuelvan poco confiables.

6.3. Lo que no debe hacerse

No utilizar este modelo para valorar propiedades individuales de alta gama sin ajustes específicos para ese segmento. No extrapolar sus estimaciones a mercados fuera de California sin una validación previa. No reemplazar con él el criterio de tasadores o analistas especializados en transacciones complejas.

*Para consultar la metodología técnica detallada, las pruebas estadísticas realizadas y la configuración de los modelos, referirse al **Reporte Operativo Técnico** complementario a este documento.*