

INFORME DE ANÁLISIS DE DATOS DEL PROYECTO DE BIENES RAÍCES EN GUATEMALA

RESUMEN Y EVALUACIÓN DEL MERCADO
DE VENTAS DE PROPIEDADES EN 2024

**GUATEMALA
REAL ESTATE**

ING. DANIEL PEDRO PABLO
ALVAREZ VIELMAN



Contenido

1. Introducción
2. Objetivos
3. Contexto
4. Methodología
5. Resultados e Interpretaciones
6. Conclusiones

Introducción

El proyecto "Predicción de Precios de Casas y Apartamentos" ofrece un análisis detallado de los precios de bienes raíces en el contexto de municipios específicos de Guatemala (17). A partir de datos extraídos de un sitio web de bienes raíces (www.encuentra24.com), este estudio se enfoca en propiedades dentro del rango de precios entre Q400,000 y Q2,000,000, examinando tanto casas como apartamentos. El proyecto sirve como una demostración práctica de habilidades de análisis de datos, con el objetivo de equipar a compradores potenciales, inversores y profesionales de la industria con valiosos conocimientos del mercado y la capacidad de predecir los precios de las propiedades.

Inicialmente, el proyecto implicó fases meticulosas de extracción y limpieza de datos. Los datos se filtraron y normalizaron para abordar las variabilidades en los listados, asegurando que solo se incluyeran propiedades dentro del rango de precios especificado. A través de un análisis exhaustivo, el municipio y la zona se determinaron con precisión a partir de los títulos de las listas que contienen los anuncios de casas y apartamentos.

En la fase exploratoria, profundizamos en las distribuciones de precios y las correlaciones de características, utilizando métodos estadísticos como matrices de correlación y pruebas ANOVA. Estos análisis destacan los factores más influyentes en el precio de las propiedades, enfatizando el importante papel de la ubicación geográfica.

El diverso panorama municipal de Guatemala informa este estudio, mostrando variaciones en la dinámica del mercado inmobiliario en diferentes áreas. En última instancia, el proyecto culminará en un modelo predictivo de precios de propiedades y visualizaciones en un tablero en Power BI.

Objetivos

1. Extracción y limpieza de datos:

- Extraer datos inmobiliarios desde una plataforma en línea, enfocándose en propiedades dentro del rango de precios entre Q400,000 a Q2,000,000.
- Limpiar y normalizar el conjunto de datos, asegurando la precisión en la representación de los detalles de la propiedad.

2. Análisis Geográfico:

- Determinar el municipio y la zona para cada listado de propiedad, aprovechando la información del título y mejorando los conocimientos de la localidad.

3. Análisis Descriptivo:

- Explorar distribuciones de precios e identificar características clave: metros cuadrados, número de habitaciones, estacionamiento y baños que influyen en los precios de las propiedades.

4. Correlación de características y pruebas estadísticas:

- Realizar análisis de correlación y pruebas ANOVA para evaluar el impacto de diferentes municipios y zonas en los precios de las propiedades.

5. Modelado predictivo:

- Utilizar diferentes modelos para predecir con precisión los precios de las propiedades, utilizando los conocimientos adquiridos en la fase de análisis exploratorio.

6. Visualizaciones del panel:

- Crear un Dashboard en Power BI para proporcionar un análisis interactivo de visualización de bienes inmuebles.

7. Perspectivas del mercado:

- Proporcionar información valiosa sobre la dinámica del mercado inmobiliario guatemalteco para guiar a compradores, inversores y profesionales de la industria.

Contexto

Este proyecto se enfoca en el mercado inmobiliario dentro de los municipios de Guatemala (17). Utilizando datos obtenidos de un sitio web de bienes raíces, exploramos listados de propiedades filtrados para reflejar el rango de precios entre Q400,000 y Q2,000,000. El conjunto de datos incluye casas y apartamentos, lo que ofrece una instantánea de la dinámica del mercado en diversas regiones.

Los municipios de Guatemala exhiben características variadas, lo que influye significativamente en las valoraciones de las propiedades. Al comprender estas diferencias regionales, nuestro objetivo es arrojar luz sobre los factores que impulsan los precios de los bienes raíces y brindar a las partes interesadas información útil.

El análisis considera propiedades de municipios clave, incluyendo Ciudad de Guatemala, Mixco, Villa Nueva, Santa Catarina Pinula, Fraijanes y otros, brindando una visión en profundidad de cómo se comparan estas áreas en términos de asequibilidad y demanda de propiedades. Este contexto prepara el escenario para desarrollar un modelo predictivo de precios, destinado a mejorar la toma de decisiones estratégicas en el mercado inmobiliario.

Metodología

Este proyecto emplea una metodología integral utilizando Python, el cual integra varias herramientas y técnicas en diferentes fases de análisis:

1. Web Scraping:

Los datos se extrajeron de un sitio web inmobiliario utilizando **Selenium** para la automatización del navegador y **BeautifulSoup** para el análisis de HTML. Las bibliotecas clave utilizadas incluyen:

- **Selenium.webdriver**: para automatizar la interacción del navegador web y administrar los controladores de Firefox.
- **WebDriver_manager**: para agilizar la configuración del controlador.
- **BeautifulSoup**: para analizar contenido HTML y extraer datos.

2. Limpieza de datos: el conjunto de datos se limpió y preparó utilizando **Pandas** para la manipulación de datos, con expresiones regulares (**re**) y normalización de texto (**unidecode**) para garantizar la coherencia de los datos.

3. Determinación geográfica y del tipo de propiedad:

- **Fuzzywuzzy**: se utilizó para determinar con precisión el municipio y la zona a partir de los títulos de propiedad y para categorizar los tipos de propiedad como casa o apartamento.

4. Análisis Descriptivo:

- Se llevaron a cabo análisis matemáticos y estadísticos utilizando **Pandas**, **Numpy** y **Scipy** para obtener información sobre los datos numéricos.
- La visualización de datos se realizó con **Matplotlib** y **Seaborn**, lo que ayudó en el examen de distribuciones de precios, análisis de correlación y variaciones de precios regionales.

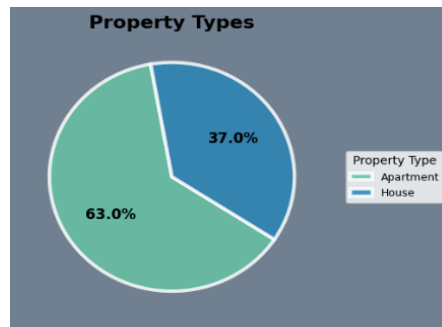
5. Modelado predictivo:

- Planificado para su futura implementación, el modelado predictivo implicará la utilización de técnicas de regresión y otros algoritmos para pronosticar los precios de las propiedades en función de las características clave identificadas.

Estas metodologías en conjunto facilitan una exploración y un análisis exhaustivos del mercado inmobiliario, respaldando la toma de decisiones informadas a través de datos sólidos.

Resultados e Interpretaciones

1. Distribución del tipo de propiedad: El conjunto de datos se compone de 37% casas y 63% apartamentos. Esta distribución sugiere una mayor disponibilidad o demanda en el mercado de apartamentos dentro de un rango de precios específico.



2. Estadísticas descriptivas: Las casas tienden a ser más grandes, con más habitaciones y espacios de estacionamiento en comparación con los apartamentos, y ambos tipos de propiedad exhiben rangos de precios similares. Las casas tienen una superficie media de 183,03 m² y 3,30 habitaciones, mientras que los apartamentos tienen una superficie media de 84,34 m² y 2,10 habitaciones. Es probable que estas diferencias influyan en las predicciones de precios y sugieran factores clave a considerar en futuras mejoras del modelo.

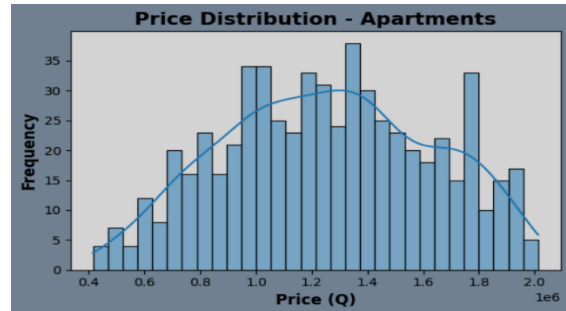
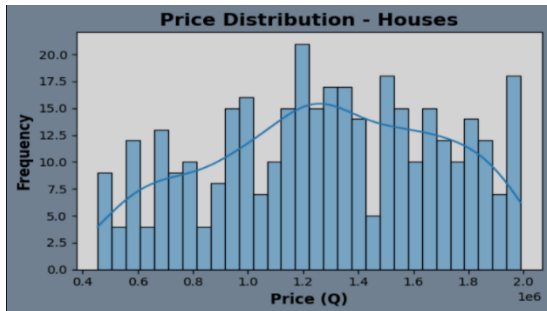
House

	mean	std	min	25%	50%	75%	max
Price (Q.)	1,282,560.38	411,895.80	455,000.00	975,000.00	1,300,000.00	1,618,700.00	1,989,000.00
Square meters	183.03	86.09	37	120	175.5	225.25	544
Rooms	3.3	1.3	1	3	3	3	13
Parking	2.23	1.01	0	2	2	2	6
Baths	2.3	0.95	1	2	2	3	7

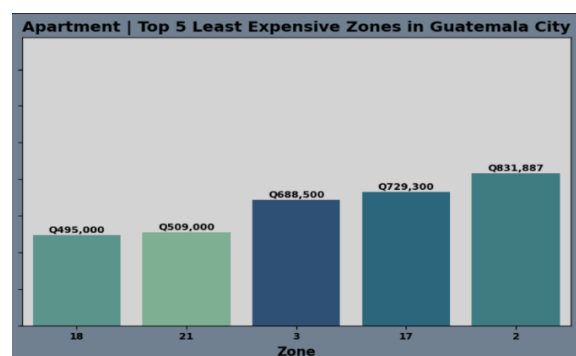
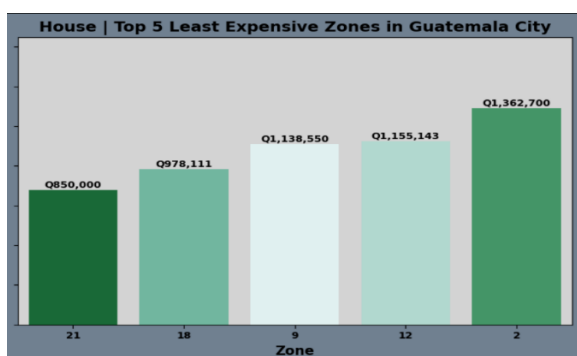
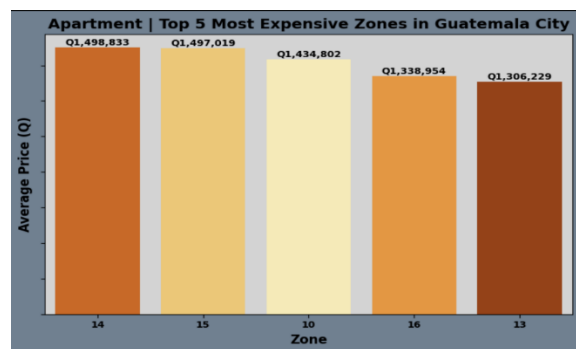
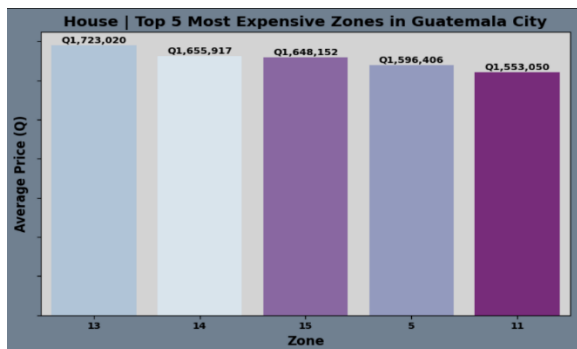
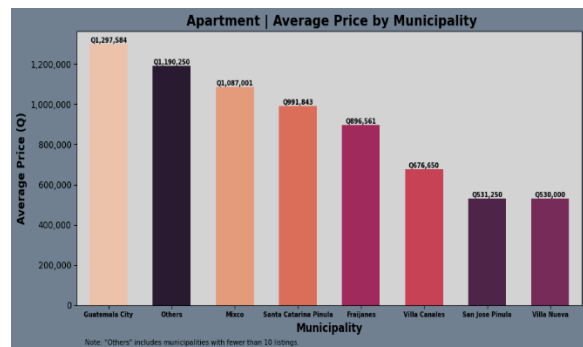
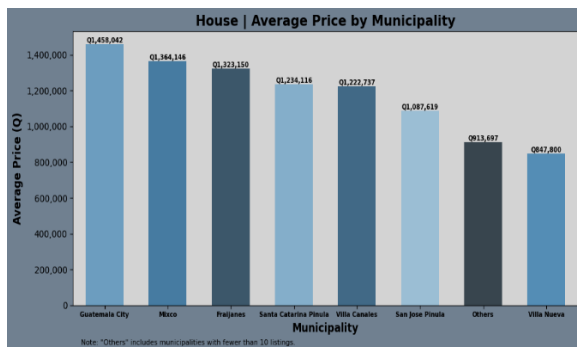
Apartment

	mean	std	min	25%	50%	75%	max
Price (Q.)	1,259,094.10	373,304.17	415,000.00	975,000.00	1,250,000.00	1,550,250.00	2,012,400.00
Square meters	84.34	34.21	19	62.25	79	100	400
Rooms	2.1	0.8	1	2	2	3	10
Parking	1.61	0.58	0	1	2	2	3
Baths	1.6	0.6	0	1	2	2	6

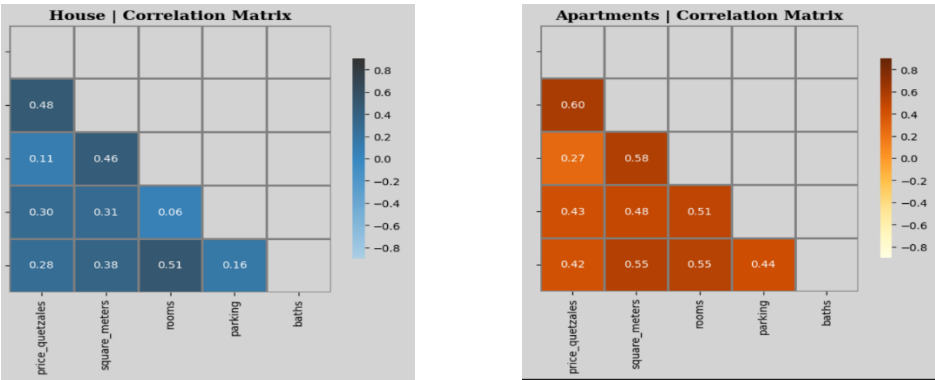
3. Distribución de precios: La distribución de precios de las casas muestra una ligera bimodalidad, mientras que los apartamentos muestran una distribución más simétrica.



4. Análisis de precios por municipio y zona: La Ciudad de Guatemala muestra los precios promedio más altos tanto para casas como para apartamentos, destacando el importante papel de la ubicación en la valoración de propiedades.



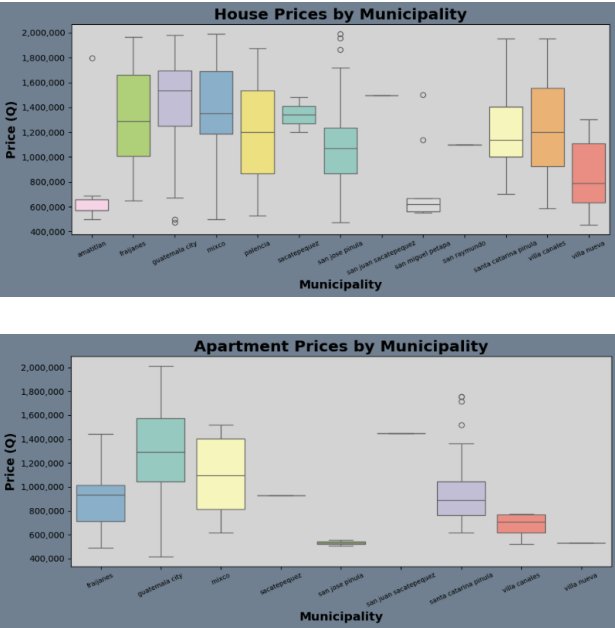
5. Perspectivas de correlación: Se observaron fuertes correlaciones entre el precio y atributos como los metros cuadrados y el número de habitaciones, lo que indica que estos factores son determinantes esenciales de los precios de las propiedades.



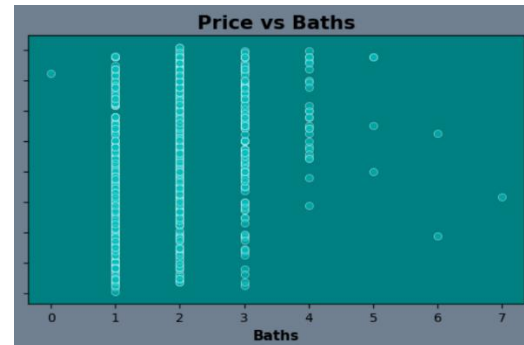
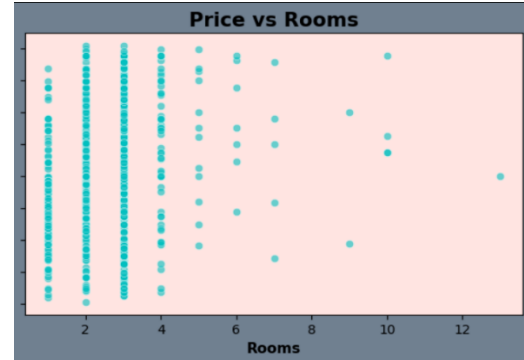
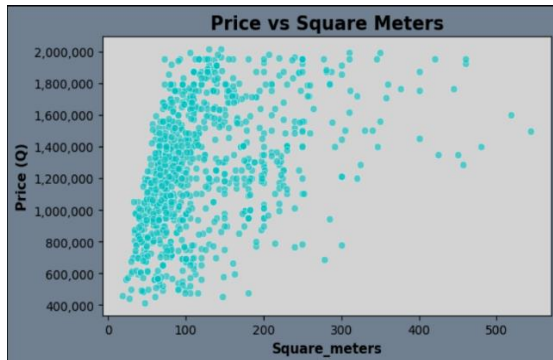
6. Análisis ANOVA: Tanto las zonas como los municipios tienen un impacto significativo en los precios de las propiedades, y las zonas muestran un efecto particularmente fuerte en las valoraciones de los apartamentos.

Type	Category	F-Statistic	p-Value
House	Zone	4.558079	2.81E-09
	Municipality	7.797261	7.36E-13
Apartment	Zone	18.826713	3.53E-45
	Municipality	8.409038	7.40E-11

7. Diagramas de caja por municipio: los diagramas de caja resaltan las distribuciones de precios entre municipios, revelando variabilidad y valores atípicos en los precios.



8. Precio frente a características clave: los diagramas de dispersión indican una correlación positiva entre el precio y características como los metros cuadrados y el número de habitaciones. Sin embargo, la variabilidad sugiere factores de influencia adicionales.



9. Resultados del modelado predictivo: aplicamos modelos de regresión múltiple para predecir los precios de las propiedades tanto para casas como para apartamentos. Los resultados se resumen a continuación:

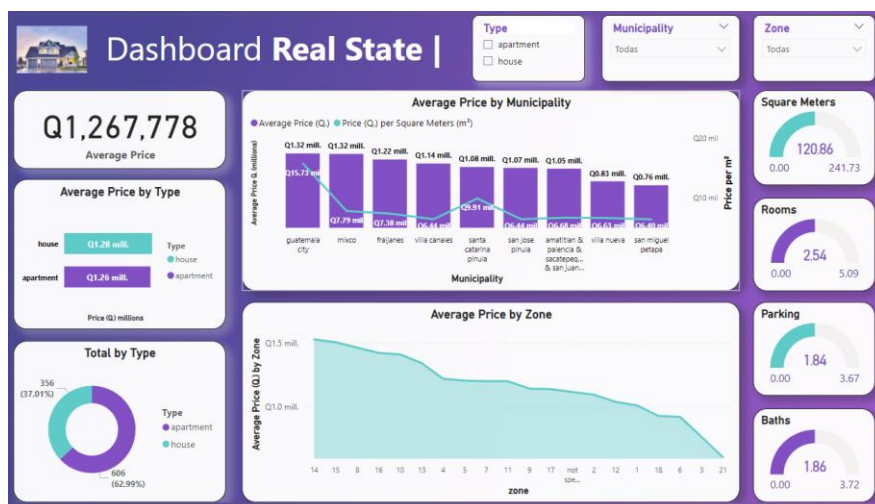
- **Casas:** Los modelos de conjunto como Random Forest y Gradient Boosting mostraron un rendimiento superior, capturando patrones complejos mejor que los modelos lineales básicos.
- **Apartamentos:** los modelos lineales funcionaron bien, pero los métodos de conjunto ofrecieron una precisión ligeramente mejor, con el aumento de gradiente a la cabeza.
- El conjunto de datos de casas es reducido lo que puede haber limitado el rendimiento del modelo, destacando el beneficio potencial de adquirir más datos.
- Los precios de los apartamentos son más predecibles con las características actuales, lo que indica una relación más directa entre las variables.

Model	Houses MSE	Houses R2	Apartments MSE	Apartments R2
Linear Regression	1.14E+11	0.308819	3.76E+10	0.676422
Lasso Regression	1.14E+11	0.308812	3.76E+10	0.67642
Decision Tree	1.14E+11	0.312328	5.47E+10	0.52928
Random Forest	9.05E+10	0.453939	3.82E+10	0.671704
Gradient Boosting	9.35E+10	0.4358	3.64E+10	0.68664

10. Power BI: El panel de visualización de Power BI complementa los resultados analíticos ofreciendo una interacción visual con los datos. Los conocimientos clave se representan a través de varias visualizaciones:

1. Distribución del tipo de propiedad: gráfico de anillos que ilustra la proporción de casas y apartamentos.
2. Análisis de precios por tipo: gráfico de barras apiladas que muestra comparaciones de precios promedio entre casas y apartamentos.
3. Tendencias geográficas de precios: gráfico combinado de columnas y líneas que revela las diferencias de precios y el precio por metro cuadrado entre municipios. Gráfico de áreas que muestra precios promedio por zona, destacando las variaciones regionales.
4. Descripción general de funciones: indicadores que brindan instantáneas clave de las características de la propiedad, como metros cuadrados, habitaciones, estacionamiento y baños.

Estas herramientas visuales respaldan los hallazgos del análisis de datos exploratorios, reforzando las conclusiones extraídas de los métodos estadísticos utilizados a lo largo del proyecto.



Conclusiones

- **Distribución del tipo de propiedad:** el conjunto de datos se compone predominantemente de apartamentos (63%), lo que puede indicar una mayor disponibilidad o demanda dentro del rango de precios analizado.
- **Estadísticas Descriptivas:** Las casas son generalmente más grandes y más caras que los apartamentos. Sin embargo, ambos tipos de propiedades exhiben un rango de precios similar, lo que sugiere variabilidad en otros factores que influyen.
- **Análisis de Precios por Municipio y Zona:** La Ciudad de Guatemala presenta consistentemente los precios promedio más altos tanto para casas como para departamentos. Se observan importantes disparidades de precios entre diferentes zonas y municipios, lo que enfatiza el papel de la ubicación en la valoración de la propiedad.
- **Perspectivas de correlación:** Las fuertes correlaciones entre ciertas características (por ejemplo, precio y metros cuadrados) sugieren que los atributos físicos son determinantes cruciales de los precios de las propiedades, particularmente para los apartamentos.
- **Hallazgos de ANOVA:** Tanto las zonas como los municipios influyen significativamente en los precios de las propiedades, observándose un efecto notablemente más fuerte en los apartamentos. Esto subraya la necesidad de incorporar factores geográficos en los modelos de predicción de precios.
- **Dinámica general del mercado:** el análisis demuestra que, si bien los atributos físicos y la ubicación influyen en gran medida en los precios de las propiedades, existe una variabilidad sustancial, lo que indica un papel potencial de otros factores no capturados en este conjunto de datos.
- **Información sobre datos:** el análisis enfatiza el papel crucial de la ubicación y los atributos físicos de la propiedad en la determinación de los precios inmobiliarios.
- **Eficacia del modelo:** Los conjuntos de modelos proporcionaron mejores predicciones tanto para casas como para apartamentos, lo que indica su capacidad para capturar relaciones complejas dentro de los datos.
- **Desafíos y limitaciones:** los datos limitados de casas afectaron la precisión del modelo. Un conjunto de datos más grande podría mejorar los hallazgos.
- **Direcciones futuras:** se recomienda mejorar la ingeniería de funciones y la recopilación de datos adicionales para refinar aún más las predicciones del modelo y obtener una visión más profunda del mercado inmobiliario.