

DATA SCIENCE MELBOURNE HOUSING DATA

Nombre: Francisca Paredes

Fecha: 02-08-2024

AGENDA

1. CONTEXTO
2. AUDIENCIA
3. HIPÓTESIS
4. METADATA
5. ANÁLISIS EXPLORATORIO
6. INSIGHTS

CONTEXTO

El conjunto de datos de propiedades de Melbourne proporciona una visión detallada de las ventas de propiedades en la región de Melbourne, Australia. Cada entrada en el conjunto de datos representa una venta de propiedad individual, ofreciendo una amplia gama de detalles sobre la venta y la propiedad en cuestión. Los datos se han recopilado meticulosamente y están disponibles públicamente en [<https://github.com/njtierney/melb-housing-data/tree/master/data-raw>], lo que permite su acceso y análisis por parte de investigadores, analistas y profesionales de datos interesados en comprender el mercado inmobiliario de Melbourne. Este conjunto de datos ofrece una valiosa oportunidad para estudiar tendencias, patrones y factores que influyen en las ventas de propiedades en esta vibrante ciudad australiana.

AUDIENCIA

1. Analistas de datos y científicos de datos: Interesados en explorar y analizar las tendencias del mercado inmobiliario de Melbourne, utilizando técnicas estadísticas y de aprendizaje automático para extraer información significativa y tomar decisiones informadas.
2. Investigadores académicos: Que desean estudiar el mercado inmobiliario de Melbourne y utilizar este conjunto de datos como parte de su investigación en economía, urbanismo, sociología u otros campos relacionados.
3. Profesionales del sector inmobiliario: Que buscan comprender mejor el mercado y las tendencias de ventas en Melbourne para informar sus decisiones comerciales, como fijación de precios, estrategias de comercialización y pronósticos.

HIPÓTESIS

¿Las propiedades ubicadas en ciertos suburbios tienen un precio promedio más alto que otras?

¿La proximidad al CBD (Central Business District) tiene algún impacto en el precio de la propiedad?

¿La presencia de ciertas características como el número de habitaciones, influye en el precio de la propiedad?

Filas ---> 34857
Columnas ---> 20
•

METADATA

Suburb: Nombre del suburbio donde se encuentra la propiedad.

Rooms: Número de habitaciones en la propiedad.

Type: Tipo de propiedad, con las siguientes categorías:

br: Dormitorio(s)

h: Casa, cabaña, villa, semi, terraza

u: Unidad, dúplex

t: Casa adosada

dev site: Sitio de desarrollo

o res: Otro residencial

Price: Precio de venta en dólares.

Method: Método de venta de la propiedad:

S: Propiedad vendida

SP: Propiedad vendida anteriormente

PI: Propiedad transmitida

PN: Vendido antes no divulgado

SN: Vendido no divulgado

NB: Ninguna oferta

VB: Oferta del proveedor

W: Retirado antes de la subasta

SA: Vendido después de la subasta

SS: Precio vendido después de la subasta no revelado

N/A: Precio o oferta más alta no disponible

SellerG: Agente de bienes raíces.

Date: Fecha de la venta.

Distance: Distancia desde el Distrito Central de Negocios (CBD). Centro de la ciudad.

Postcode: Código postal de la propiedad.

Bedroom2: Número de dormitorios extraído de una fuente diferente.

Bathroom: Número de baños en la propiedad.

Car: Número de plazas de aparcamiento.

Landsize: Tamaño del terreno en metros cuadrados.

BuildingArea: Tamaño del edificio en metros cuadrados.

YearBuilt: Año de construcción de la propiedad.

CouncilArea: Consejo rector del área.

Latitude: Latitud geográfica de la propiedad.

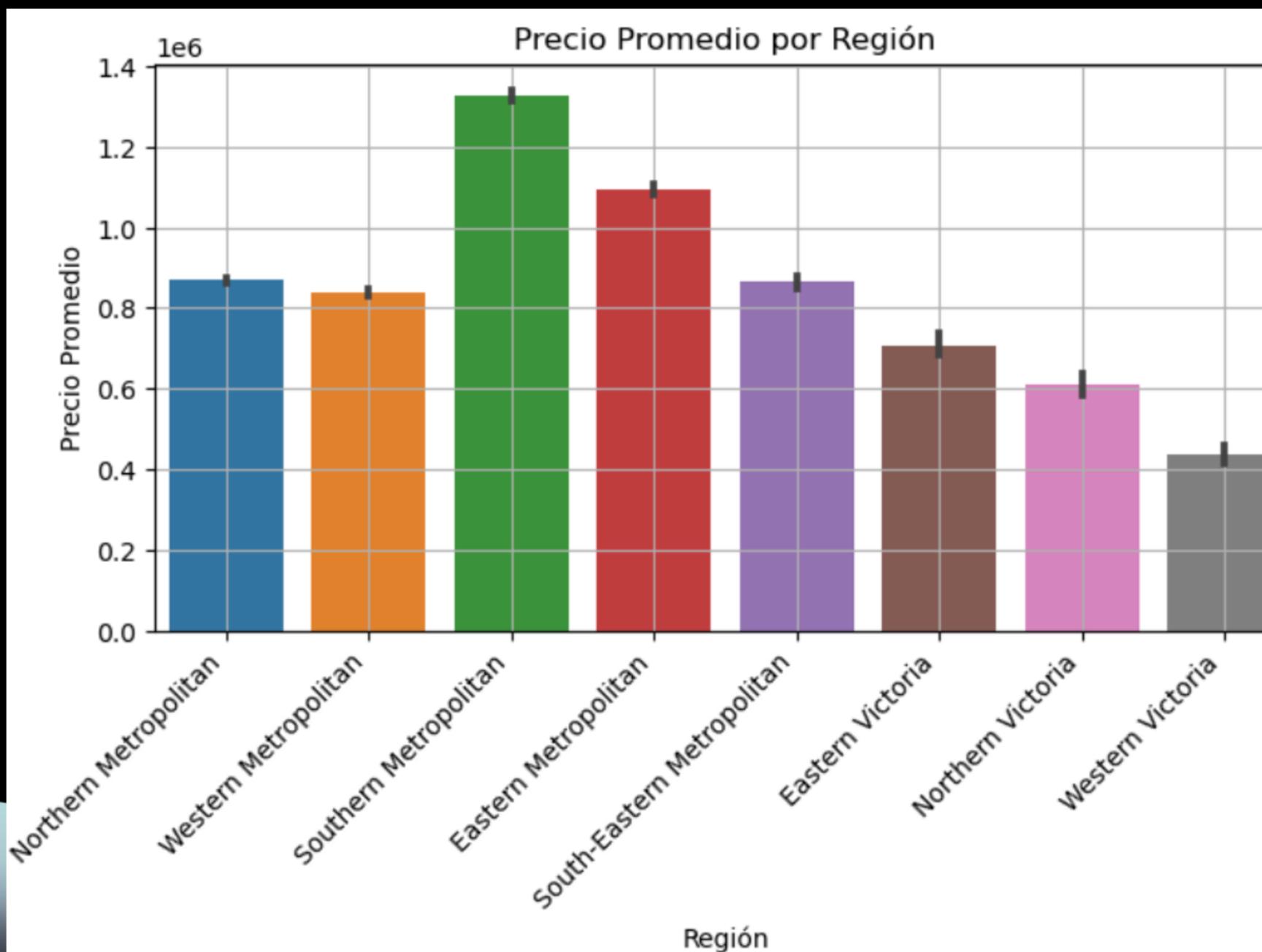
Longitude: Longitud geográfica de la propiedad.

Regionname: Región general (Oeste, Noroeste, Norte, Noreste, etc.).

Propertycount: Número de propiedades que existen en el suburbio.

EDA

¿Las propiedades ubicadas en ciertas regiones tienen un precio promedio más alto que otras?

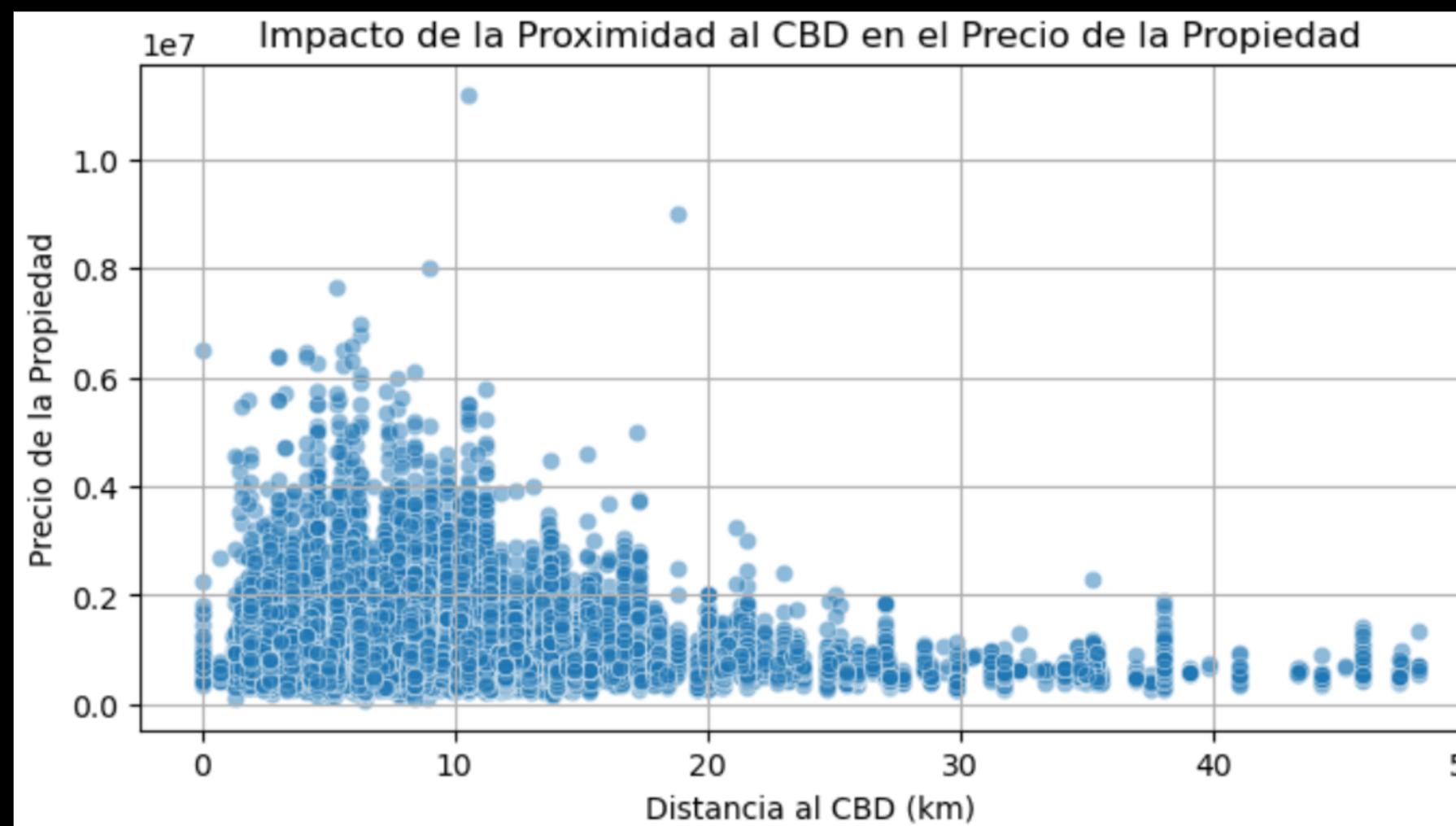


La regiones Southern Metropolitan y Easter Metropolitan tiene precios promedio mayores a los de las demás regiones.

Western Victoria y Northern Victoria tiene precios promedio menores al resto de las demás regiones.

EDA

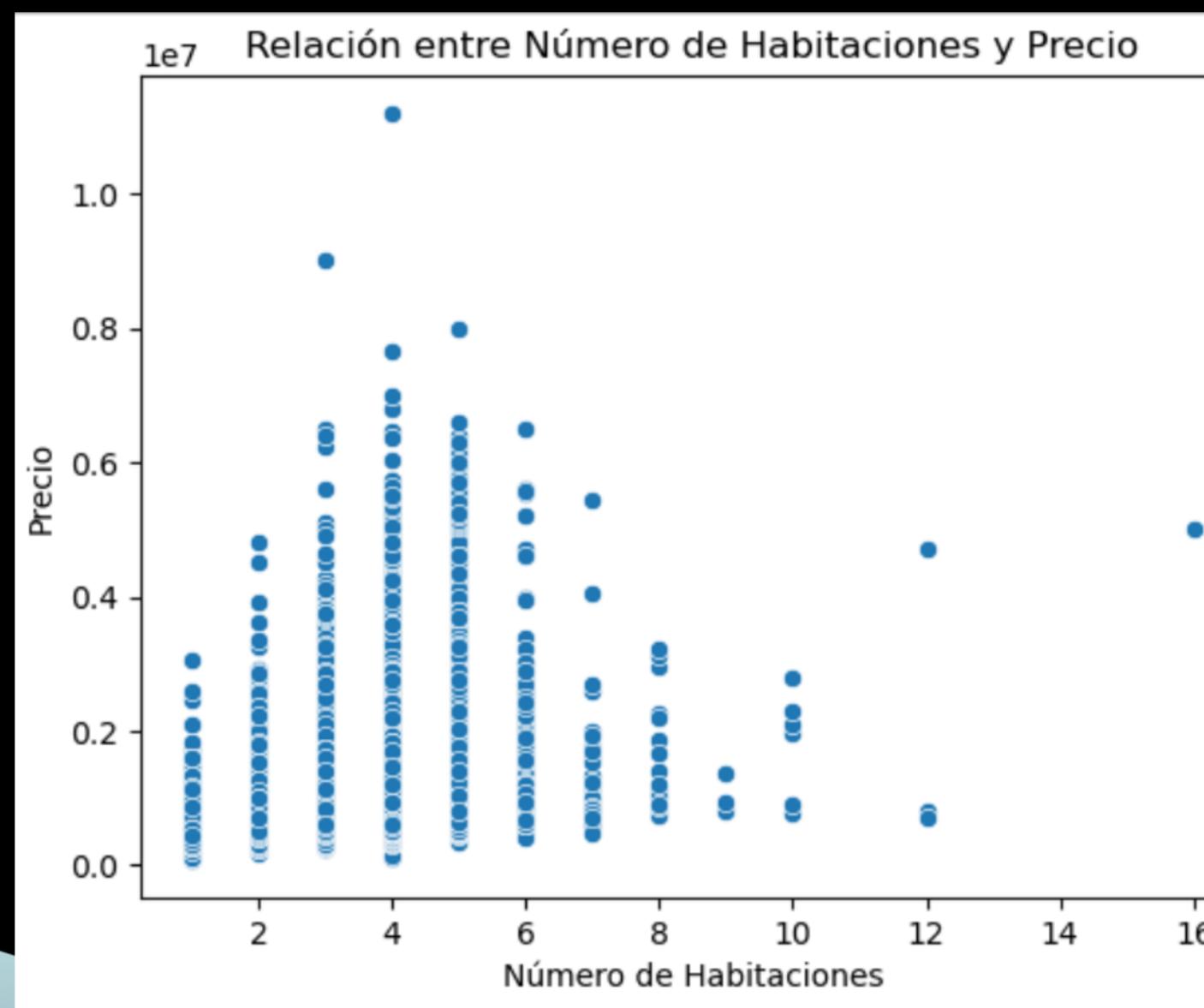
¿La proximidad al CBD (Central Business District) tiene algún impacto en el precio de la propiedad?



A menor distancia del centro CBD es mayor el precio de la propiedad, es decir en un rango de entre 0 a 11 kilómetros de distancia los valores de las propiedades son elevados, por el contrario los precios comienzan a descender significativamente desde los 12 km hasta los 50 km de distancia desde el CBD.

EDA

¿La presencia de ciertas características como el número de habitaciones, influye en el precio de la propiedad?



Podemos concluir que las propiedades que presentan entre 3, 4, 5, 6 habitaciones tiene precios mas elevados, siendo las propiedades que presentan 4 habitaciones las de mayor precio.

INSIGHT

A menor distancia del centro CBD es mayor el precio de la propiedad, es decir en un rango de entre 0 a 11 kilómetros de distancia los valores de las propiedades son elevados, por el contrario los precios comienzan a descender significativamente desde los 12 km hasta los 50 km de distancia desde el CBD.

La regiones Southern Metropolitan y Easter Metropolitan tiene precios promedio mayores a los de las demás regiones.

Western Victoria y Northerm Victoria tiene precios promedio menores al resto de las demás regiones.

Podemos concluir que las propiedades que presentan entre 3, 4, 5, 6 habitaciones tiene precios mas elevados, siendo las propiedades que presentan 4 habitaciones las de mayor precio.

PROYECTO FINAL: IMPLEMENTACIÓN MODELOS DE ML

Objetivos generales:

- Utilizando modelos de Machine Learning para resolver problemas de una industria o negocio.

Objetivos específicos:

- Modelar la situación como un problema de Machine Learning.
- Entrenar modelos de Machine Learning.
- Realizar ingeniería de atributos y normalización/estandarización de variables.
- Seleccionar el modelo con mejor performance.

Motivación:

- Ayudar a compradores y vendedores a tomar decisiones informadas, entender las tendencias del mercado, o mejorar las estrategias de marketing para agentes inmobiliarios.

Audiencia:

- Especifica que el análisis está dirigido a analistas de datos, científicos de datos, investigadores académicos, y profesionales del sector inmobiliario, quienes podrían usar esta información para análisis más profundos, investigaciones académicas o decisiones de negocio.

Preguntas/Problemas:

- ¿Tendrán los modelos de regresión lineal mayor precisión en comparación con los modelos de Random Forest al utilizar este conjunto de datos?

CONCLUSIONES

- Los valores mae, mse, y el rmse son demasiado grandes, lo que se interpreta como un modelo que con baja precisión para predecir precios de las propiedades. Esto puede deberse a varias razones, como modelo inapropiado, características insuficientes, normalización insuficiente, etc.
- Los valores de los errores son relativamente alto los que indica que se generan errores en las predicciones, mientras mas bajo esten o mas cercano a 1 es mejor, en este caso estan mas lejos. Es importante considerar que posiblemente el modelo no este capturando adecuadamente las características influyentes en los precios de las propiedades o que quizas tengan demasiada variabilidad, en este caso seria mejor utilizar un modelo mas complejo. En comparación al modelo de regresion lineal, tubo mejoras. Se podria inferir que el rendimiento de random forest para este conjunto de dato es mejor que el de regresion lineal, aunque seria conveniente utilizar modelos mas complejos.