Informe

Diego Da Rosa, Valentina Soldini, Nahuel Bizoso

2024-07-12

Table of contents

1 Introducción

El mercado inmobiliario ha sido de gran importancia durante muchos años, tanto para inversores como para personas que buscan una propiedad para alquilar. La variabilidad de los precios puede depender de diferentes factores, como es la ubicación, características de la propiedad, también la seguridad de la zona y algunos otros aspectos. Este estudio tiene como objetivo analizar como varían los precios en dólares de las propiedades teniendo en cuenta diferentes aspectos. Para realizar este análisis, planteamos las diferentes preguntas de investigación:

- ¿Cambia el precio con respecto al barrio en que se encuentra la propiedad?
- ¿En que municipio hay más alquileres?
- ¿Cuál es la variable que más influye en el precio?
- ¿Hay alguna variable que debería de eliminarse del modelo?

Estas preguntas guiarán nuestro análisis para comprender mejor los factores que afectan los precios en el mercado inmobiliario y proporciona una herramienta útil para los distintos agentes, ya sean compradores, vendedores, inversores. La variable de interés de este proyecto es *precio_m2*.

Deben revisar las faltas ortográficas cuando entregan un trabajo, no voy a marcarlas todas pero es importante.

2 Datos

Los datos utilizados en este análisis se han extraidos de una pagina web llamada **kaggle**, consiste en una plataforma web que reúne la comunidad Data Science, con más de 536 mil miembros activos en 194 países. Estos datos contienen variables que describen diversos aspectos de una propiedad que se encuentran disponibles para alquilar en el año 2023. La base de datos abarcan un total de 2361 registros de diferentes propiedades y 45 variables que incluyen desde el precio hasta diferentes características internas y externas de la propiedad.

Puedes encontrar el dataset en el siguiente enlace: Precio de alquiler de casas en Montevideo Dataset

3 Análisis exploratorio de los datos

Para responder adecuadamente las preguntas planteadas, se inició con una limpieza de los datos. Esta limpieza incluyó la eliminación de datos duplicados, valores atípicos y valores nulos. Además, se decidió eliminar las variables *Furnished...32*, *Furnished...42*, y *Floors*. Las dos primeras variables fueron eliminadas debido a la existencia de una tercera que contenía la misma información. La columna *Floors* se eliminó ya que no aportaba información relevante, contenía muchos datos faltantes y los datos presentes eran inconsistentes y poco fiables.

Otra modificación realizada fue la agrupación de las zonas de Montevideo en municipios. Se creó una variable adicional *Municipio* que contiene todas las zonas de Montevideo separadas por municipios, se puede observar en la Figura 1 que zonas corresponden a cada municipio. Adicionalmente, se convirtió la variable *precio* por *precio_m2* (precio por metros cuadrados) para facilitar el análisis de los datos.

Como se mencionó anteriormente, ciertas variables contenían valores nulos. Para abordar este problema, se emplearon diferentes métodos según las características de cada variable:

- **dirección**: Se asignó el nombre de la zona correspondiente a la propiedad.
- **precio_m2**: Se calculó la media de toda la columna de dicha variable y se coloco este valor en las filas donde había un dato faltante.
- **condición**: Se asignó la categoría más frecuente. Lo mismo se realizó con la variable *disposición*

Para las variables que se mencionó, se calculo la media de toda la columna y se coloco ese dato en las filas con datos faltantes. Se tomo la media ya que tiene una capacidad para reflejar la tendencia central del conjunto de datos en su totalidad, incluyendo la influencia de los valores extremos. Este enfoque proporciona una estimación consistente y efectiva para la preparación de los datos antes de su análisis detallado.

Además, se re-codificaron variables para mejorar su análisis y para evaluar cómo algunas de las variables del modelo afectaban lineal mente a la variable predictora, **precio_m2**

Por otro lado, a los datos duplicados se eliminaron teniendo en cuenta que las coincidencias únicas que tienen las filas, ya que no tenemos una fecha de snapshot de Mercado Libre.

Determinar que variables Influyen más en el Precio por metro cuadrado:

A continuación, se presentarán algunas visualizaciones para responder las preguntas planteadas.

hay que escribir en español, algunas veces se permite que el nombre de algunos métodos estadísticos estén en inglés ya que no se encuentran traducciones adecuadas pero el resto en español.

¿Cual es el municipio más costoso para alquilar?

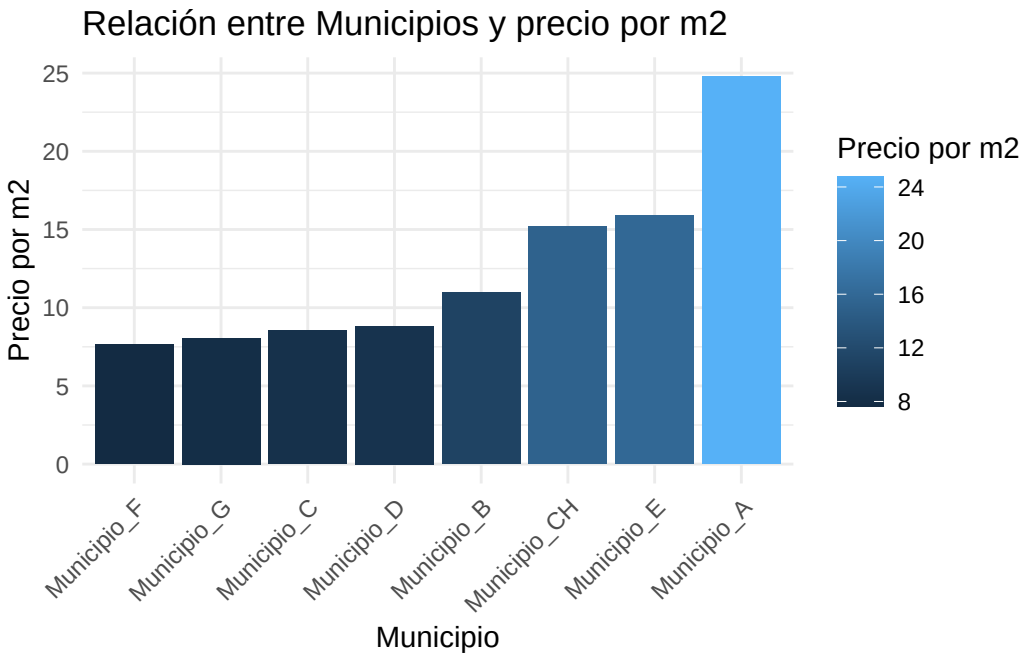


Figure 1: Relación entre el Precio por Metro cuadrado y Municipios

En este gráfico se puede apreciar como el Municipio A tiene los precios por metro cuadrado promedios mas altos, seguido por el Municipio E. Este gráfico es útil para ver que zonas de Montevideo son mas caras y cuáles son mas económicas, lo cual es valioso para los inquilinos al decidir en que área desean vivir según sus preferencias y presupuesto.

¿Cómo varía el precio por metro cuadrado según el año de construcción de la propiedad? no se ve claramente eso que dicen en el gráfico, tal vez sacando los atípicos se vería un poco mejo

A continuación, se presenta un gráfico que muestra la antigüedad de la propiedad y cómo varía el precio por metro cuadrado según el año de construcción. Se observa que a medida que las propiedades son más nuevas, el precio por metro cuadrado tiende a aumentar.

Para llevar a cabo este análisis, fue necesario acotar ciertos datos y eliminar algunos valores atípicos que distorsionaban la visualización.

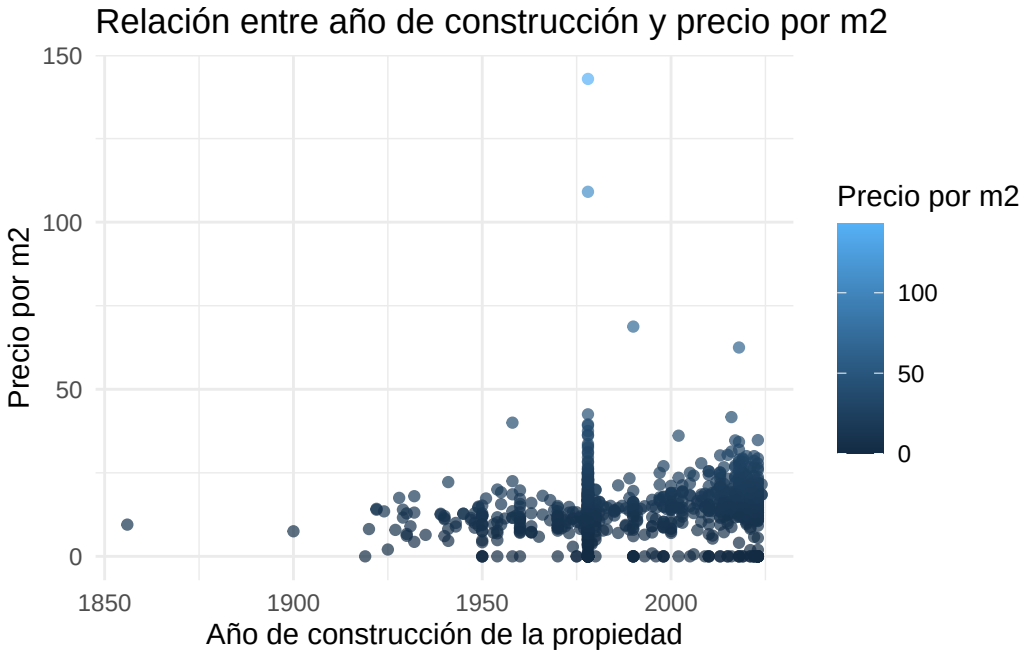


Figure 2: Relación entre año de construcción y precio por m2

Este patrón es consistente con la idea de que las propiedades más nuevas tienen características y otras comodidades, mejor infraestructura y probablemente esten ubicadas en lugares más deseables.

¿Que tipo de propiedad presenta los alquileres más elevados?

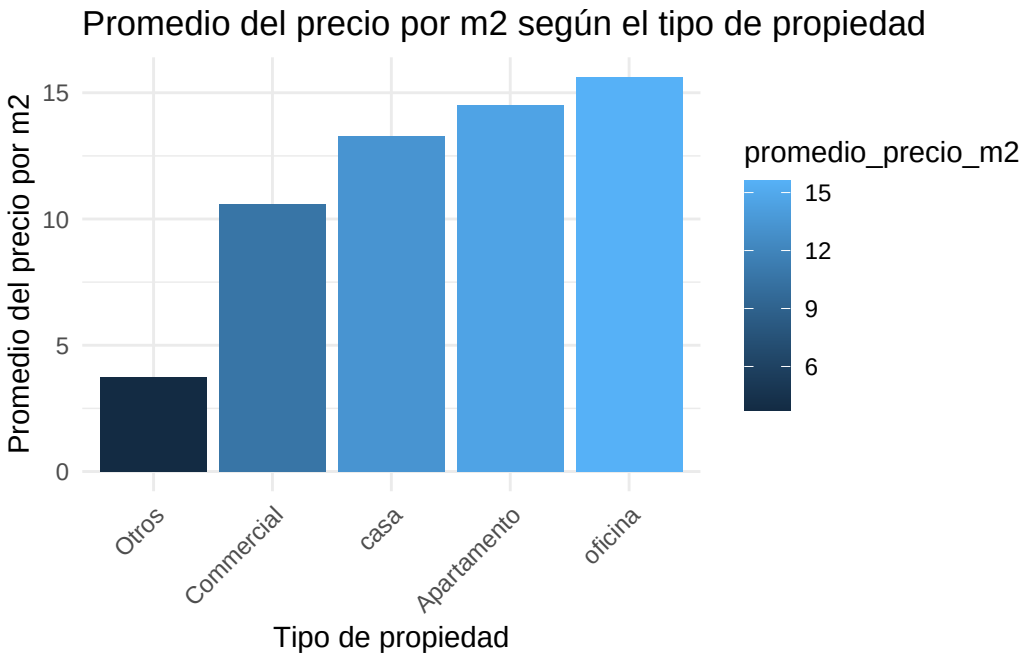


Figure 3: Relación entre el precio por m2 y el tipo de propiedad

Como se puede observar en promedio el tipo de propiedad que son mas costosas son las oficinas seguido por los apartamento. En este análisis también se re-codificaron ciertas observaciones para poder analizar mejor los datos.

¿Cómo es la distribución de los alquileres en función del precio promedio por metro cuadrado?

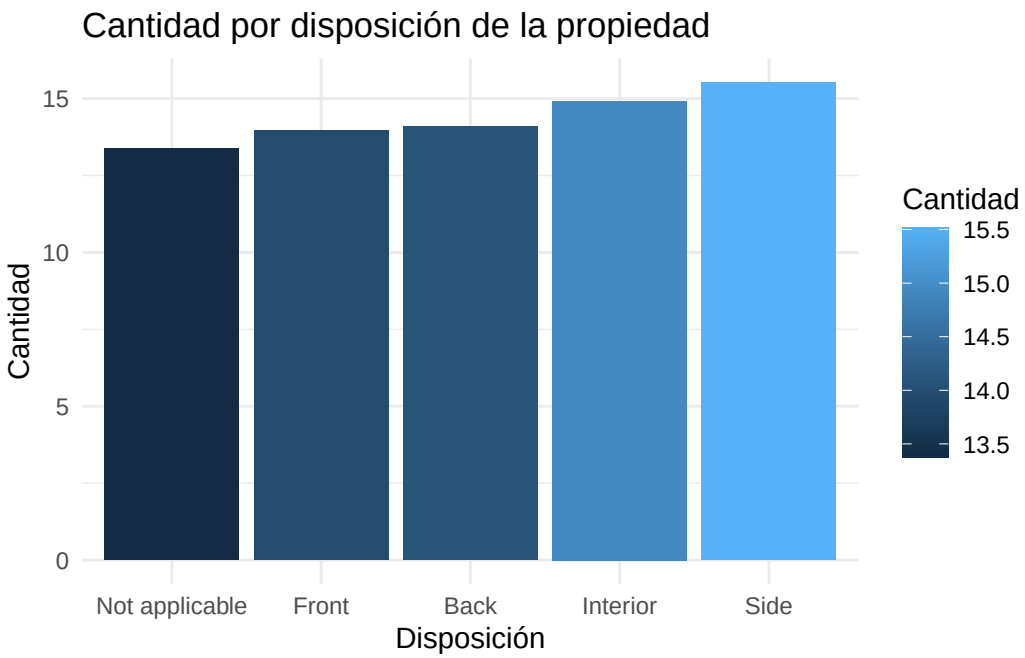


Figure 4: Relación entre el precio por m2 y la disposición de la propiedad

No hay una diferencia muy notoria con respecto al precio por metro cuadrado en cuanto a la dispocicion en que se encuentra las diferentes propiedades.

No tienen que poner dos títulos en cada gráfico, el título de los gráficos va abajo

¿Cómo se distribuye el precio promedio de los alquileres según las condiciones?

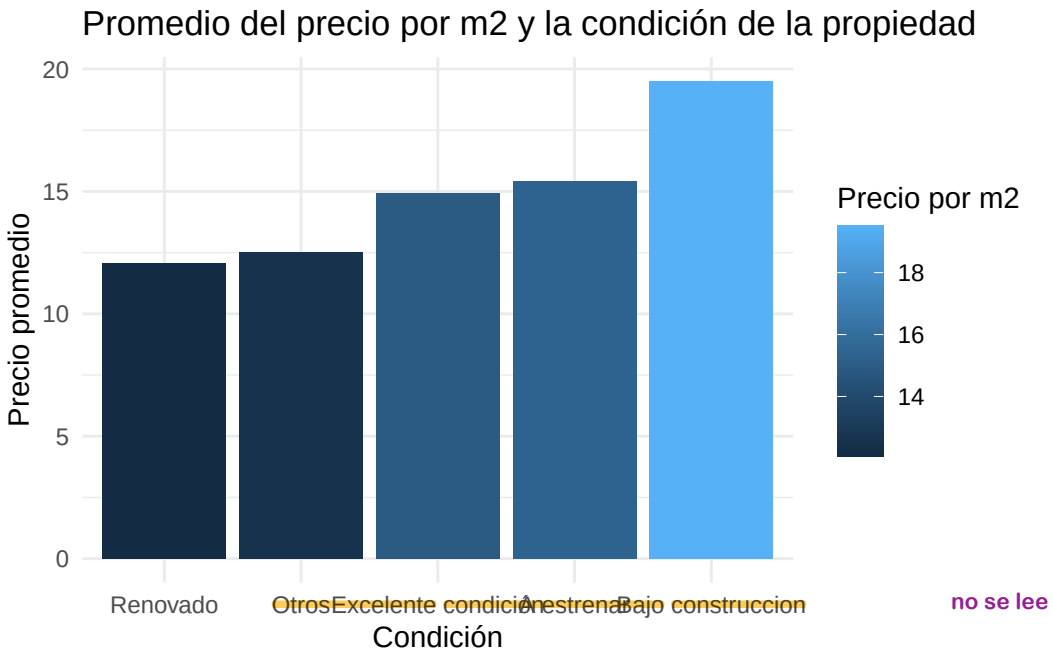
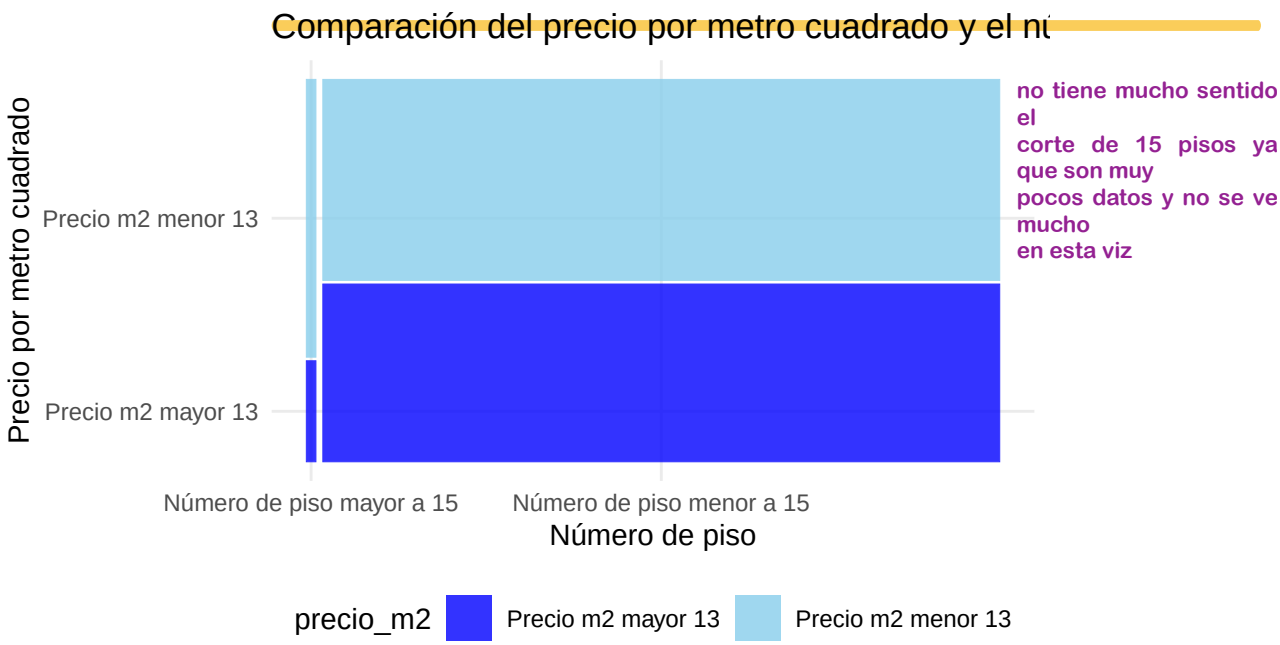


Figure 5: Relación entre el precio por m2 y la condición de la propiedad

Al considerar la condición de los alquileres, se puede observar que la mayoría de las propiedades se encuentran en buenas condiciones. Esto es una buena señal para inquilinos que directamente se quieren trasladar a esa propiedad. En términos de precio promedio por metro cuadrado, los alquileres “En Construcción”, “Para Estrenar” y en “Excelente Condición” son los más costosos.

¿Cómo varía el precio por metro cuadrado según la cantidad de pisos de la propiedad?

salió cortado



no tiene mucho sentido el corte de 15 pisos ya que son muy pocos datos y no se ve mucho en esta viz

Figure 6: Relación entre el precio por m2 y la nro. de piso de la propiedad

Este gráfico representa la relación entre el precio por metro cuadrado (*precio_m2*) y el número de piso (*nro_de_piso*) de las propiedades.

Re-caracterizando el precio promedio por metro cuadrado en dos grupos: - Menor o igual a 13 dólares y mayor a 13 dólares,

cómo seleccionaron el punto de corte 13?

También se re-categorizó el número de piso en dos grupos: - Menor o igual a 15 pisos y mayor a 15 pisos.

cómo seleccionaron el punto de corte 15?

Para los precios por metro cuadrado menor o igual a 13 dólares, se observa que son propiedades que tienen pisos bajos menores a 15. Para los precios por metro cuadrado mayores a 13 dólares, se observa que son propiedades que tienen pisos mayores a 15.

Por lo tanto, en pisos mas bajos, predominan los precios por metro cuadrado menores o iguales a 13 dólares. Por otro lado, para los pisos mas altos, los precios por metro cuadrado mayores a 13 dólares son los más comunes.

Esto sugiere que el número de piso podría tener una influencia en el precio por metro cuadrado, con pisos mas altos tendiendo a precios mayores por metro cuadrado en comparación con pisos más bajos.

están sobreinterpretando,
cuántos apartamentos con menos de 15 pisos tienen en los datps que analizan ?

¿Cómo influye en el precio del inmueble la cantidad de cuartos en una propiedad?

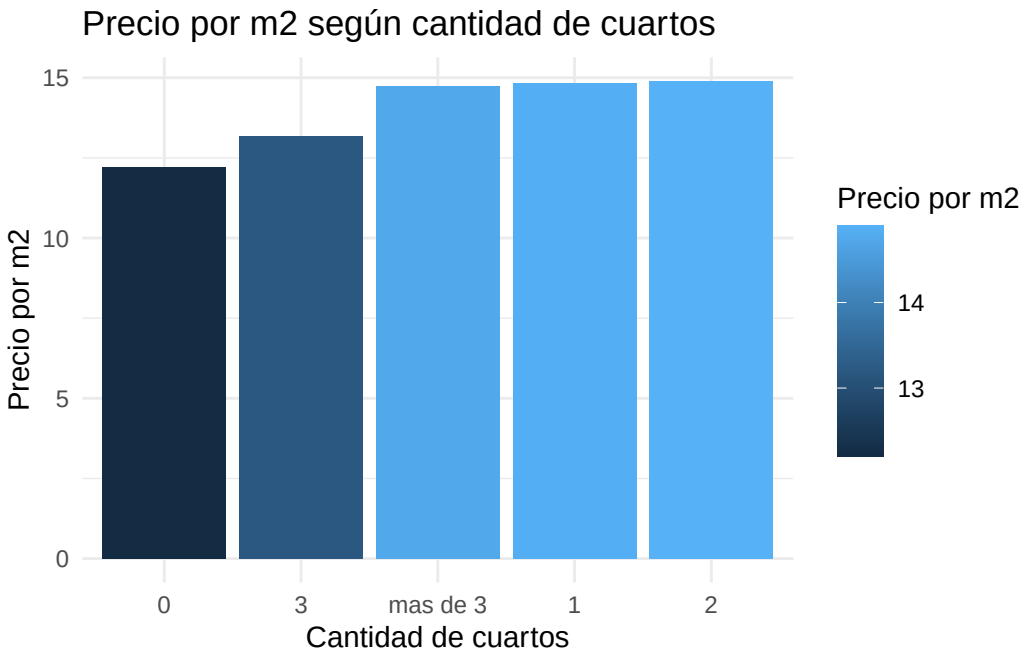


Figure 7: Relación entre el precio por m2 y la cantidad de cuertos de la propiedad

Como se puede observar, la cantidad de cuartos no afecta demasiado el cambio en el precio por metro cuadrado. Puede sonar un poco extraño ya que en la realidad cuantos mas cuartos es mas costosa la propiedad, por lo tanto es un poco contradictorio con la realidad.

este comentario es incorrecto ya que ustedes analizan el precio por metro cuadrado no el precio total

¿Como afecta el incremento del precio por metro cuadrado el aumentar en la cantidad de baños?

A continuación se muestra un gráfico que ilustra cómo la cantidad de baños afecta el precio por metro cuadrado. En este análisis, la variable baños se re-codifico en cinco grupos para facilitar su interpretación y análisis.

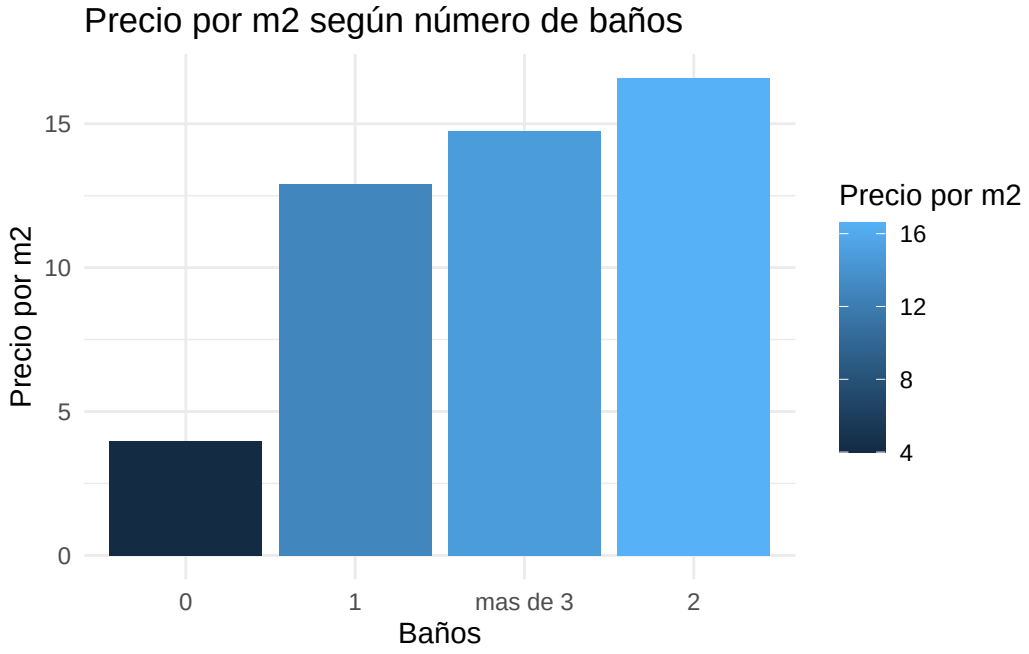


Figure 8: Relación entre el precio por m2 y el número de baños de la propiedad

Se observe que el precio se incrementa hasta la propiedad que tiene dos baños. Sin embargo, después de dos baños, el impacto en el precio es menor. Es decir, una propiedad con tres baños es mas cara que una con uno, pero es considerablemente más cara que una de dos baños. Parece que llega a un punto en el que la cantidad de baños no afecta el precio por metro cuadrado de una propiedad.

¿Como afecta al incremento del precio el aumento de la cantidad de estacionamientos?

Similar al anterior razonamiento, los valores de la variable *cant_de_est* se re-codificaron en cuatro grupos para una mejor interpretación.

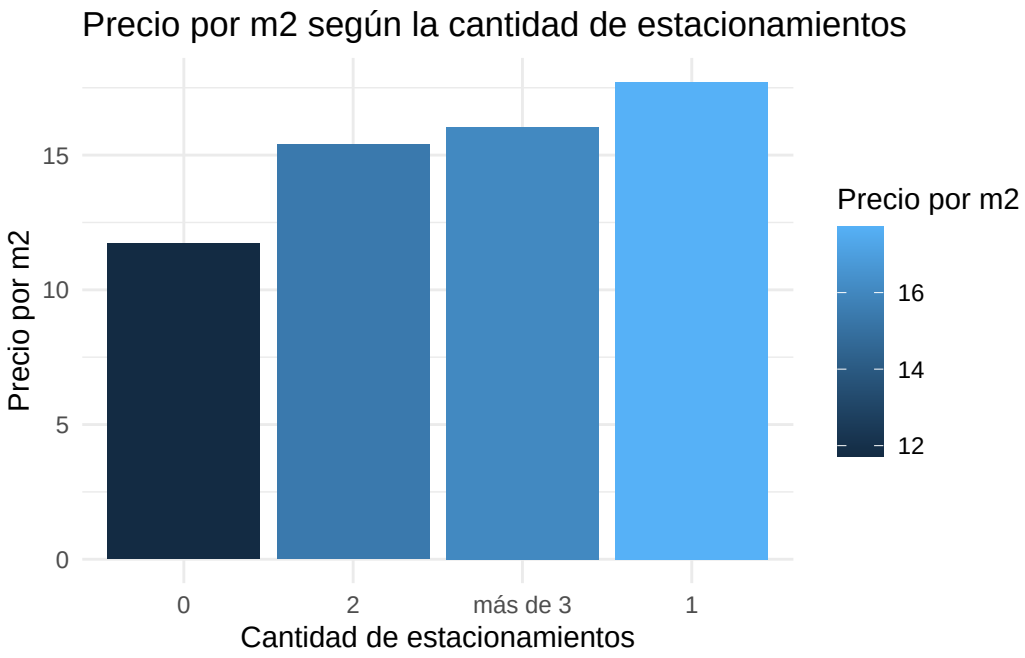


Figure 9: Relación entre el precio por m2 y la cantidad de estacionamientos de la propiedad

De manera similar a lo observado anteriormente, el precio por metro cuadrado de la propiedad es mayor cuando la propiedad tiene al menos un estacionamiento. Sin embargo, el incremento en el precio no tan relevare al pasar de uno a dos estacionamientos. Curiosamente, las propiedades con tres estacionamientos son más caras que las que tienen dos, aunque la diferencia no es tan marcada como entre tener cero y uno.

En los siguientes dos gráficos se re-codificó ahora la cantidad de cantidad de habitaciones y la cantidad de cuartos. Se observa que en general, el tener 2 cuartos o 2 habitaciones hace que el precio promedio por metro cuadrado sea mayor, esto también se cumple para los alquileres con 3 o más cuartos / habitaciones.

¿La cantidad de habitaciones que posee la propiedad modifica el precio por metro cuadrado?

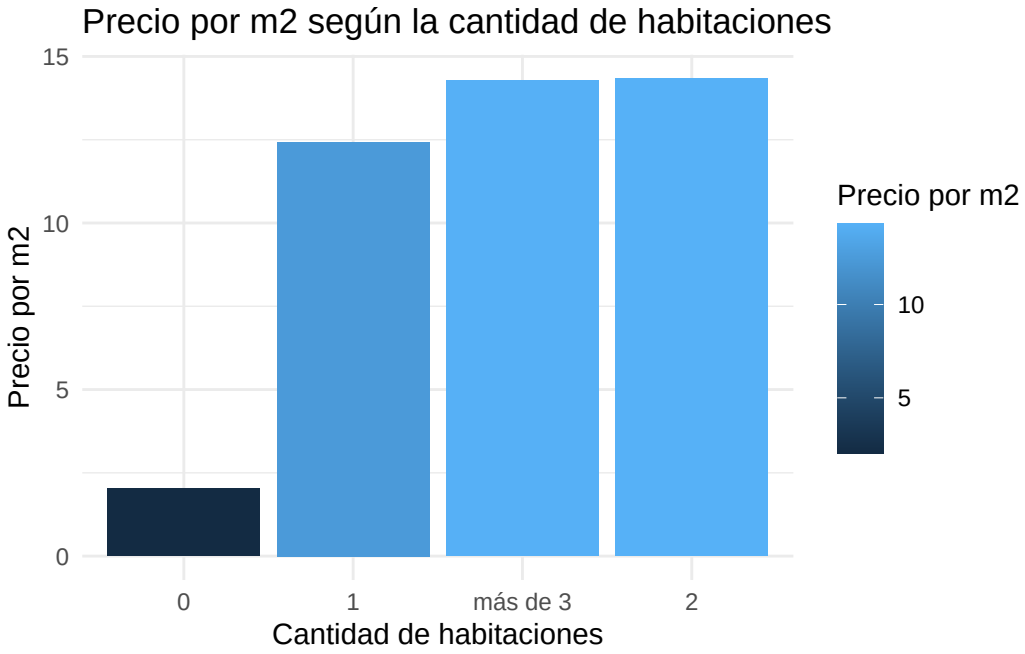


Figure 10: Relación entre el precio por m2 y la cantidad de habitaciones de la propiedad

¿El hecho de que las propiedades cuenten con calefacción impacta en el precio por metro cuadrado?

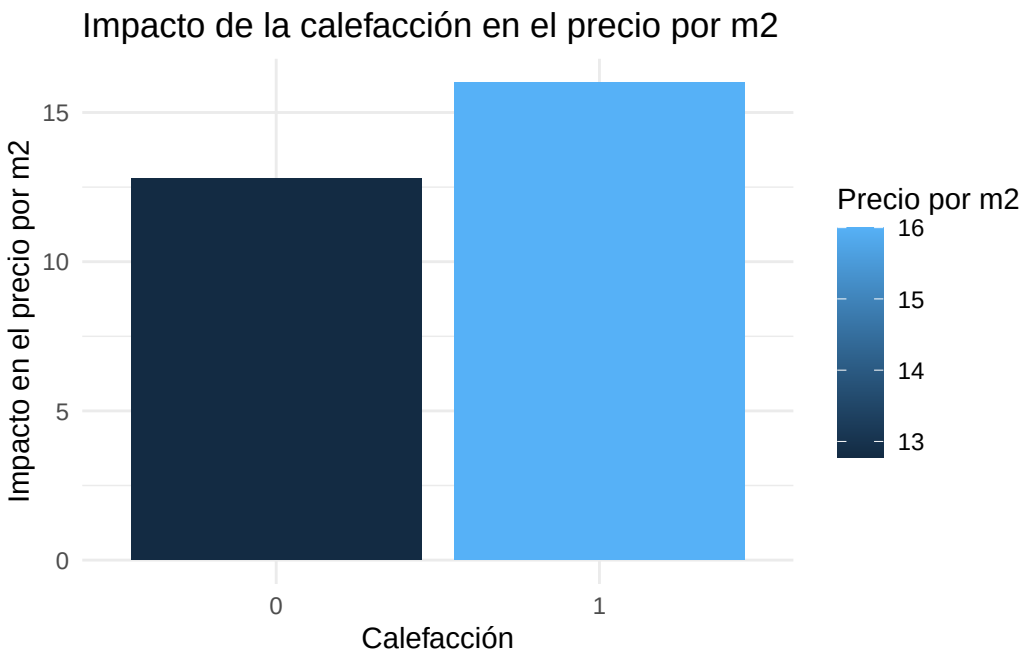


Figure 11: Relación entre el precio por m2 y la calefacción de la propiedad

¿El tener vista al mar influye en el precio por metro cuadrado de las propiedades?

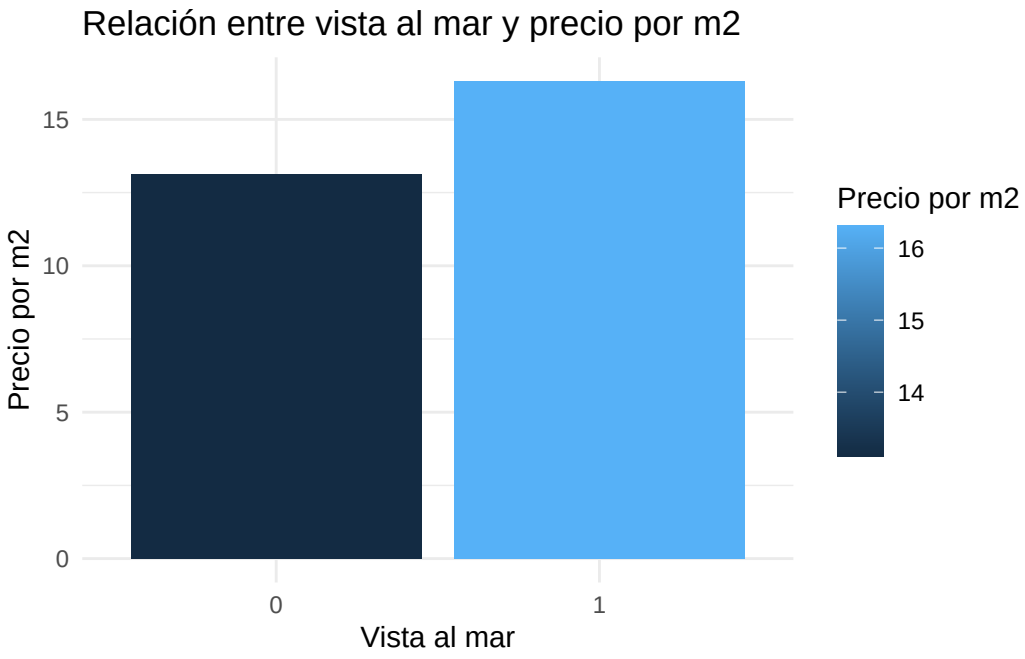


Figure 12: Relación entre el precio por m2 y la vista al mar de la propiedad

Analizando los anteriores dos gráficos se puede ver como las propiedades tienen un aumento del precio por metro cuadrado cuando tienen vista al mar y cuando tienen calefacción. Esto es importante a tener en cuenta cuando un inquilino esta interesado en alguna propiedad.

¿El incremento en los gastos comunes afecta negativamente el precio por metro cuadrado de la propiedad?

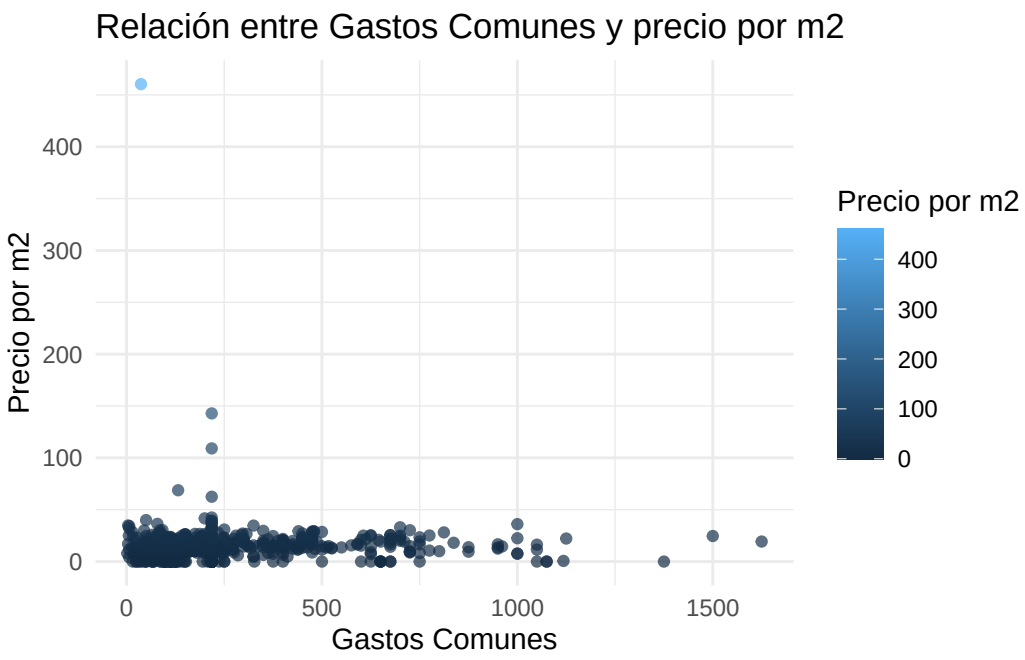


Figure 13: Relación entre el precio por m2 y los gatos comunes de la propiedad

En este gráfico, muestra una tendencia lineal, se puede decir que los gastos comunes son independientes con el precio por metro cuadrado.También se puede decir que son independientes con el tamaño de la propiedad, ya que generalmente las propiedades con mas metros cuadrados son mas costosas y no necesariamente pagan mas gastos comunes.

en la exploración fató que incorporaran visualizaciones que involucren más de una o dos variables casi todos sus gráficos son de barras.

4 Modelo Predictivo

supongo que es un gráfico de dispersión a lo que se refieren

A que le llaman gráfico lineal?

Con el objetivo de identificar si las variables del modelo explican correctamente la variable de respuesta, en este caso *precio_m2*, se decidió hacer una exploración con ciertas variables. Para ello, se elaboraron gráficos lineales que anteriormente están presentados que comparan distintas variables con el precio por metro cuadrado, permitiendo observar cómo cada una de estas afecta el precio de manera específica. Este análisis se ha realizado con casi todas las variables disponibles en los datos.

según el análisis exploratorio las que parecen tener una relación más fuerte

Una vez hecho este análisis, se paso a crear el modelo de regresión lineal. Este modelo se creo a partir de las variables que mas afectaba el cambio en el precio por metro cuadrado. Es así, que se eligieron las variables: *Municipio*, *condicion*, *baños*, *año_de_const*, *tipo_prop*, *vista_al_mar*, *cant_de_est*, *nro_de_piso*, a pesar de que se redujo la cantidad de variables comparando con todas las variables que hay en el set de datos, se interpretó que así se iba a poder ver un poco mejor que quieren mostrar las diferentes variables.

$$precio\_m2 = \beta 0 + Municipio\beta 1 + condicion\beta 2 + baños\beta 3 + año\_de\_const\beta 4 + tipo\_prop\beta 5 + vista\_al\_mar\beta 6 + cant\_de\_est\beta 7 + nro\_de\_piso\beta 8 + error$$

Una vez obtenido el modelo, se realizo el summary, y se obtuvieron estos resultados:

Table 1: Summary

Coeficientes	**Estimador**	**Error_Estandar**	**Valor_t**	**P_Valor**
(Intercept)	25.63	9.35	2.74	0.01
MunicipioMunicipio_B	-15.30	3.36	-4.55	0.00
MunicipioMunicipio_C	-15.14	4.26	-3.55	0.00
MunicipioMunicipio_CH	-12.61	3.33	-3.78	0.00
MunicipioMunicipio_D	-16.02	4.54	-3.53	0.00
MunicipioMunicipio_E	-11.96	3.49	-3.43	0.00
MunicipioMunicipio_F	-15.70	6.25	-2.51	0.01
MunicipioMunicipio_G	-16.93	5.81	-2.91	0.00
condicionBajo construccion	5.56	9.37	0.59	0.55
condicionExcelente condición	-1.84	1.45	-1.27	0.21
condicionOtros	-2.28	1.42	-1.61	0.11
condicionRenovado	-3.06	2.97	-1.03	0.30
baños1	2.09	8.53	0.25	0.81
baños2	3.69	8.60	0.43	0.67
bañosmas de 3	1.23	8.64	0.14	0.89
año_de_const	0.00	0.00	-1.08	0.28
tipo_propcasa	-1.14	1.71	-0.67	0.50
tipo_propCommercial	-1.57	2.06	-0.76	0.45
tipo_propoficina	2.49	2.27	1.09	0.27
tipo_propOtros	-8.87	3.74	-2.37	0.02

vista_al_mar1	3.80	1.79	2.12	0.03
cant_de_est1	4.24	1.28	3.32	0.00
cant_de_est2	2.47	2.18	1.13	0.26
cant_de_estmás de 3	2.85	3.41	0.84	0.40
nro_de_piso	0.00	0.01	-0.06	0.95

Por lo que se puede observar es que casi todos los coeficientes dan un p-valor mayor a 0.05 porque lo que significa que no son significativas al modelo. no son significativamente distintos de cero

Table 2: Raíz cuadrada de SE

Raíz cuadrada de SE
15.86975

La raíz cuadrada del error cuadrático medio (ECM) indica, en promedio, que las predicciones están desviadas por 15.90 dólares del valor real del precio por metro cuadrado, lo cual significa que las predicciones están bastante mal. si ya que es el valor promedi por metro cuadrado mas o menos no?

Luego por otro lado, se analizó el R^2, se puede interpretar que las variables explicativas no explican correctamente a la variable de respuesta ya que el R^2 dio muy bajo

Table 3: R2

R2
0.07

la variabilidad explicada por el modelo

Que la significancia no se cumpla es decir que las variables no sean significativas y que el R^2 sea tan chico que las variables explicativas no expliquen en absoluto a la variable de respuesta, es una mala señal.

Por ende, a continuación se van a evaluar los cuatros supuestos para identificar si un modelo predice bien o no.

5 Supuestos a cumplir

Para poder obtener conclusiones confiables, el modelo debe cumplir con determinados supuestos

- no multicolinealidad: exacta ni aproximada, para asegurar que la matriz X sea de rango completo (conformable).

- linealidad: la relación entre variables expicativas y la respuesta debe ser aproximadamente lineal.
- homoscedasticidad: la varianza de los errores no depende de ninguna de las variables explicativas.
- normalidad: los errores del modelo deben presentar una distribución normal.
- atípicos/influyentes: si bien no es un supuesto en si mismo, es recomendable identificar observaciones atípicas e influyentes al modelo.

5.1 Multicolinealidad:

La mayoria de las variables tienen valores de GVIF ajustados menores a 2, lo que generalmente se consudera aceptable por lo que no hay problemas graves de multicolinealidad en el modelo.

5.2 Homocedasticidad

ncvTest

Hipótesis:

- **Hipótesis Nula (H0)**: La varianza de los residuos es constante (Homocedasticidad).
- **Hipótesis Alternativa (H1)**: La varianza de los residuos no es constante (heterocedasticidad).

Procedimiento:

- El *ncvTest* examina la relación entre los valores ajustados (predicciones del modelo) y la varianza de los residuos.
- Se ajusta un modelo de regresión para predecir los residuos en función de los valores ajustados.
- Se calcula un estadístico de prueba basado en esta relación.
- El estadístico de prueba sigue una distribución chi-cuadrado.

  Un p-valor alto sugiere que no hay evidencia suficiente para rechazar la hipótesis nula, indicando Homocedasticidad.
- Un p-valor bajo indica heteroscedasticidad.

Table 4: p-valor

p__valor
0

Breusch-Pagan Test

Hipótesis:

- **Hipótesis Nula (H0)**: La varianza de los residuos es constante y no depende de las variables independientes (Homocedasticidad).
- **Hipótesis Alternativa (H1)**: La varianza de los residuos depende linealmente de las variables independientes (heteroscedasticidad).

Procedimiento:

- El test de *Breusch-Pagan* examina si la varianza de los residuos depende linealmente de las variables independientes del modelo original.
- Se realiza una regresión auxiliar de los residuos al cuadrado contra las variables independientes originales.
- Se calcula un estadístico de prueba basado en la regresión auxiliar.
- El estadístico de prueba sigue una distribución chi-cuadrado.

statistic	p.value	parameter	method	alternative
53.84243	0.0004474	24	Koenker (studentised)	greater

Ambos p-valores son super pequeños, por lo tanto, hay problema de Heterocedasticidad. Se puede concluir que se rechaza la hipótesis nula que no es nuestro objetivo

5.3 Normalidad

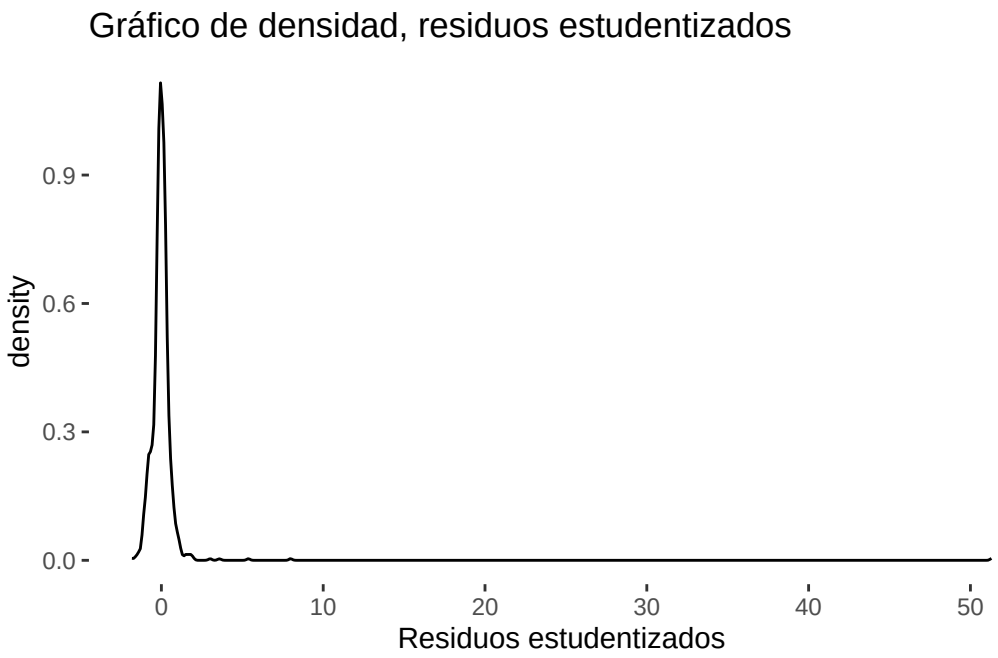


Figure 14: Normalidad de los residuos

La densidad de los residuos estudentizados no parecen comportarse como una normal. Por lo tanto nos indica que los residuos no se distribuyen normal.

Se va a corroborar lo visto en el gráfico mediante tres métodos:

- **Shaphiro-Wilk**: se basa en la comparación de los cuantiles empíricos y teóricos bajo el supuesto de normalidad.
- **Jarque-Bera**: se basa en la comparación de los estadísticos de asimetría y kurtosis bajo el supuesto de normalidad.
- **Kolmogorov-Smirnov**: se basa en la máxima discrepacnoa entre la función de distribución empírica y la teórica bajo el supuesto de normalidad.

Table 6: Normalidad

Método	p.valor
Shaphiro-Wilk	0
Jarque-Bera	0

Kolmogorov-Smirnov	0

Los test Shaphiro-Wilk, Jarque-Bera, Kolmogorov-Smirnov tiene un p-valor menor que 0.05, por lo que se rechaza la hipótesis nula, los residuos se distribuyen normal.

5.4 Lineanidad

A este se dio por hecho debido a que la forma en que se iba a mostrar era ilegible.

5.5 Datos Atípicos

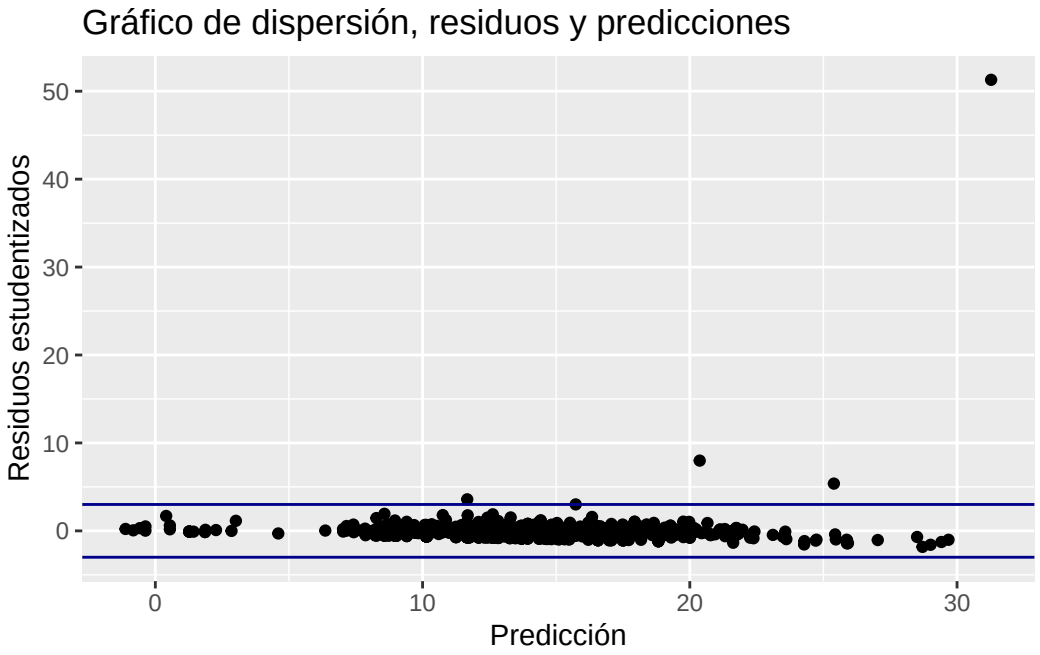


Figure 15: Presencia de Atípicos

Existen cuatro observación que podría tomarse como dato atípico.

Debido a que casi ningún supuesto se cumplió, se optó por utilizar otros métodos de prediccion, que son el método de árboles y bosques:

el modelo lineal no es apropiado, pero en este trabajo no profundizamos en como mejorarlo.
De todas formas para los requisitos del curso está bien el recorrido que
hicieron aunque no se pedían todos los diagnósticos que hicieron ya que no los cubrimos.

5.6 Árbol de decisión:

Los modelos basados en árboles consisten en una secuencia de particiones anidadas que dividen el espacio de los predictores donde en cada partición se usa un modelo para predecir la respuesta.

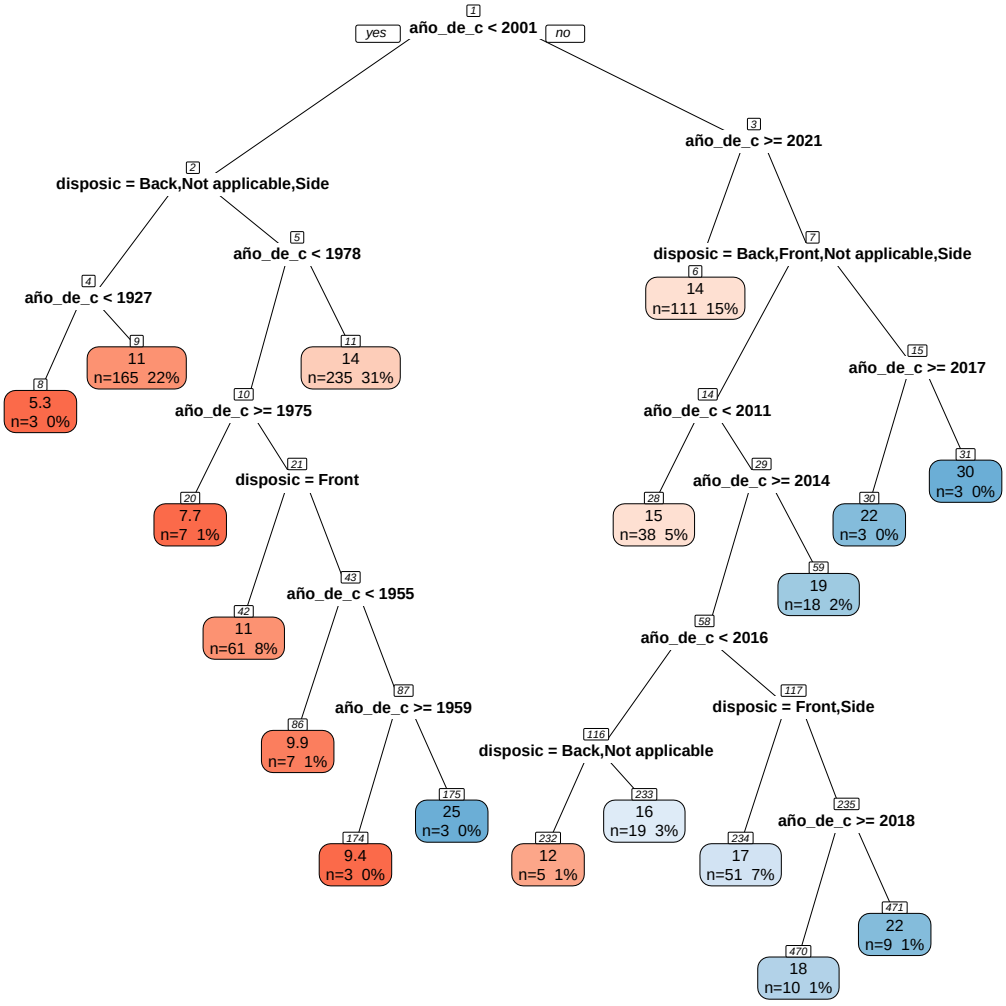


Figure 16: Árbol

El árbol de decisión muestra que el año de construcción y la disposición de las propiedades son factores importantes para predecir el precio por metro cuadrado. Las propiedades más recientes

tienden a tener un precio más alto, y la disposición también juega un papel importante, especialmente en las propiedades más nuevas.

No evaluaron el poder predictivo del método con los datos de testeo

5.7 Bosques:

Para hacer este análisis se quitaron las variables dirección y zona ya que daban problemas y por otro lado estaba municipio que lo que se evaluó en todo el trabajo.

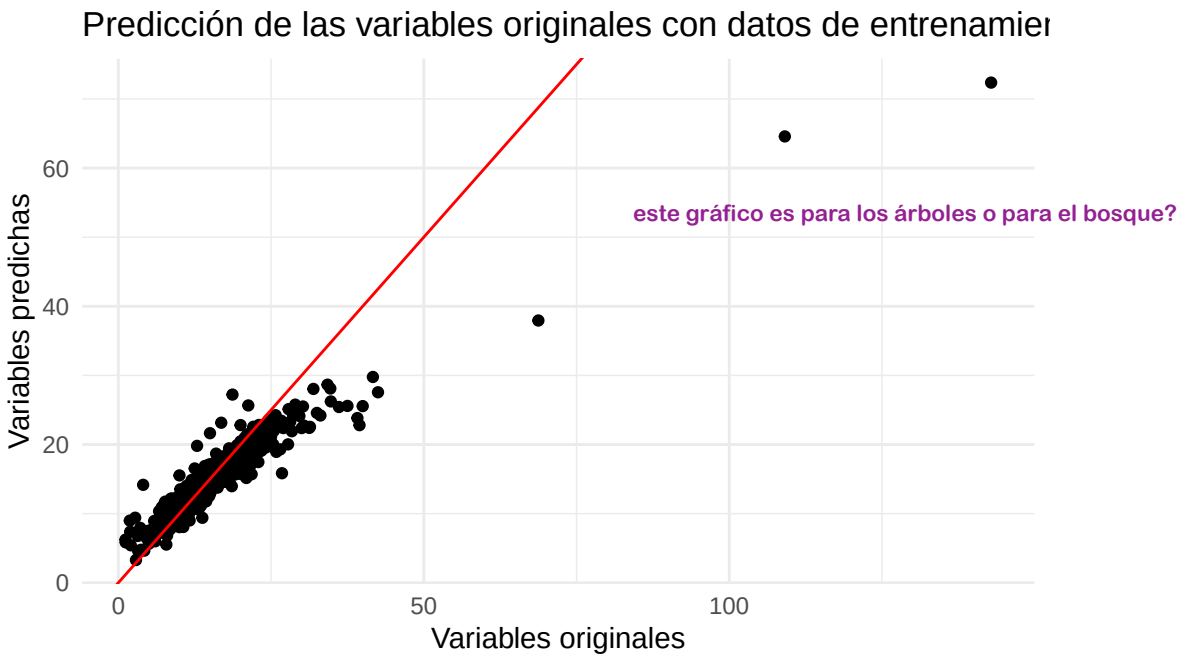


Figure 17: Conjunto de entrenamiento

En la gráfica, se muestra una comparación entre las variables originales y las variables predichas utilizando un modelo de random forest. Las predicciones serían perfectas si todos los puntos cayeran sobre la línea roja

La evaluacion del mdodelo se debe hacer con los datos de testeo

En este paso se utilizo el conjunto de entrenamiento en el que se esta utilizando únicamente el 30% de los datos.

es al revez se ajusta el modelo con el 70% de los datos y se testea con el 30%

Las observaciones dispersas alrededor de la línea roja indica que hay cierta variabilidad entre los valores originales y las predicciones. Esta predicción no es tan mala ya que la mayoría de los datos se encuentran bastante cerca de la línea roja. También se aprecia que las observaciones siguen la misma tendencia que dicha línea, esto significa que a medida que crecen las variables originales las predicciones también, por lo que es una buena señal.

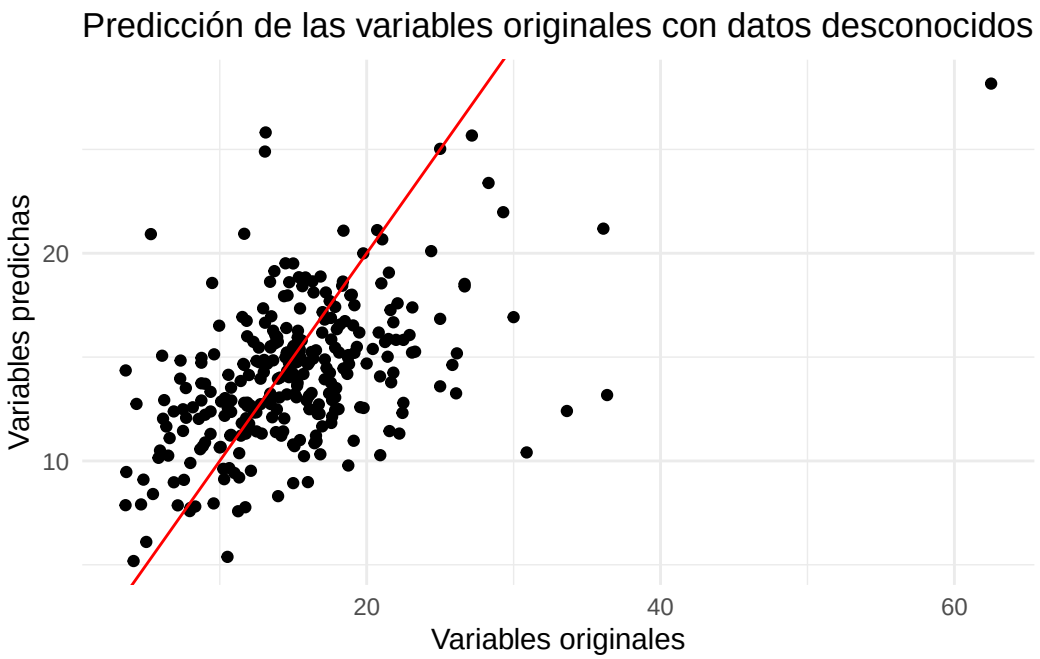


Figure 18: Predicción de variables originales

Como en el gráfico anterior se muestra la comparación entre las variables originales y las variables predichas usando el método de random forest, pero esta vez aplicándolo al conjunto de datos "testing" que son un 70% de los datos. lo que mencioné antes

Con respecto a la dispersión alrededor de la línea roja, en comparación al gráfico anterior, usando el conjunto de entrenamiento, se observa una mayor dispersión de las observaciones alrededor de la línea roja. Esto sugiere que el modelo tiene más dificultad para predecir correctamente.

La mayor dispersión y el alejamiento de algunos puntos de la línea roja indican que el modelo no está generalizado bien a datos nuevos y desconocidos. Este es un indicio de que el modelo podría estar sobre ajustando al conjunto de datos de entrenamiento.

Al final con cuantos datos ajustaron el modelo?
para el modelo de regresión lineal tambien lo hicieron con el 30% de datos de entrenamiento?
Ese sería un problema ya que tienen muchos parámetros a estimar con muy pocos
datos y por eso también da peor de lo que debería

6 Conclusión

Se puede concluir que, aunque el modelo de regresión lineal no fue completamente exitoso en predecir el precio por metro cuadrado debido a la falta de cumplimiento de algunos supuestos y la baja capacidad explicativa, los análisis realizados proporcionaron información valiosa sobre los factores que influyen en los precios inmobiliarios en Montevideo. Los métodos alternativos como los árboles de decisión y los bosques aleatorios muestran un camino prometedor para futuras investigaciones en el mercado inmobiliario. Igualmente parte del objetivo se cumplió, que era brindarle información a los individuos interesados en alquilar, cual eran las características de las propiedades que los propietarios tienen en cuenta para elevar el precio de las propiedades.

7 Metadatos

Table 7: Metadatos

Nombre de la variable	**Tipo de variable**	**Descripción**
direccion	Character	Ubicación donde se encuentra la propiedad
nro_de_piso	Character	Número de piso de la propiedad
tipo_prop	Character	Tipo de poropiedad, como apartamento, oficina, casa
Condicion	Character	Condición en la que esta la vivienda, es una variable que se separa en categorias (Excelente, Buena, Renovada
disposicion	Character	En que posición se encuentra la propiedad separada en categorías (interior, frente, fondo)
m2_terraza	Numérica	Metros cuadrado de la terraza
cant_cuartos	Numérica	Cantidad de dormitoros de la propiedad
vivienda_social	Dummy	Si la propiedad esta destinada para vivienda social, variable de si y no
Closet	Dummy	Espacio de guardaropas en la propiedad, variable de si y no
Garage	Dummy	Garage en la propiedad, variable de si y no
Playroom	Dummy	Sala de juegos en la propiedad, variable de si y no
WiFi	Dummy	Conección a wifi en la propiedad, variable de si y no
balcon	Dummy	Balcon en la propiedad, variable de si y no
jardin	Dummy	Jardin en la propiedad, variable de si y no
amueblado	Dummy	La propiedad viene con los muebles incluidos a la hora de ser alquilado, variable de si y no
Updated	Fecha	Última fecha que fue actualizado el listado
Apartments_per_Floor	Numérica	Número de apartamentos por piso en el edificio
zona	Character	En que zona de Montevideo se encuentra la propiedad
baños	Numérica	Cuántos cuartos hay en la propiedad

m2_totales	Numérica	Metros cuadrados de la propiedad, en dolares
cant_de_est	Numérica	Cantidad de estacionamientos que tiene disponible la propiedad
Barbecue	Dummy	Barbacoa en la propiedad, variable de si y no
Patio	Dummy	Patio en la propiedad, variable de si y no
Pet	Dummy	Si se aceptan perros o no
Duplex	Numérica	Si es un duplex o no, duplex es una propiedad que esta encima de otra y generalmente son dos
zona_barbacoa	Dummy	Si hay una zona de barbacoa en el edificio de la propiedad, variable de si y no
deposito	Dummy	Área de almacenaiento o depósito en la propiedad, variable variable de si y no
losa_radiante	Dummy	Losa radiante en la propiedad, variable de si y no
precio_m2	Numérica	El precio de la propiedad por metros cuadrados
Common_Expenses	Numérica	Gastos comunes asociados a la propiedad, en dolares
cant_de_piso	Numérica	Núemero total de piso en el edificio
cant_habitaciones	Numérica	Cantidad de habitaciones en la propiedad
año_de_const	Numérica	En que año fue construida la propiedad
vista_al_mar	Dummy	Propiedad con vista al mar o no, variable de si y no
Living_Room	Dummy	Propiedad con living o no
Pool	Numérica	Pool en la propiedad
coneccion_gas	Dummy	Propiedad cuenta con conección a gas directa o no, variable de si y no
calefaccion	Dummy	Propiedad con calefacción, variable de si y no
Gym	Dummy	Propiedad con gym, variable de si y no
Air_condicioning	Dummy	Propiedad con aire acondicionado, variable de si y no
Solarium	Dummy	Propiedad con solarium, variable de si y no

8 Biblografia

- Notas de la unidad curricular Ciencia de Datos con R
- Notas de la unidad curricular Modelos Lineales
- Auguie B (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. R package version 2.3, https://CRAN.R-project.org/package=gridExtra.
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686.
- Schloerke B, Cook D, Larmarange J, Briatte F, Marbach M, Thoen E, Elberg A, Crowley J (2024). *GGally: Extension to 'ggplot2'*. R package version 2.2.1, https://CRAN.R-

project.org/package=GGally.
- Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition. Sage, Thousand Oaks CA. https://socialsciences.mcmaster.ca/jfox/Books/Companion/.
- Farrar, Thomas J. (2024). skedastic: Heteroskedasticity Diagnostics for Linear Regression Models. R Package version 2.0.2. University of the Western Cape. Bellville, South Africa. https://github.com/tjfarrar/skedastic
- Martin Maechler, Peter Rousseeuw, Christophe Croux, Valentin Todorov, Andreas Ruckstuhl, Matias Salibian-Barrera, Tobias Verbeke, Manuel Koller, c("Eduardo", "L. T.") Conceicao and Maria Anna di Palma (2024). robustbase: Basic Robust Statistics R package version 0.99-2. URL http://CRAN.R-project.org/package=robustbase
- Trapletti A, Hornik K (2024). *tseries: Time Series Analysis and Computational Finance*. R package version 0.10-56, https://CRAN.R-project.org/package=tseries.
- Müller K (2020). *here: A Simpler Way to Find Your Files*. R package version 1.0.1, https://CRAN.R-project.org/package=here.
- Wickham H, Bryan J (2023). *readxl: Read Excel Files*. R package version 1.4.3, https://CRAN.R-project.org/package=readxl.
- H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York,2016.
- Wickham H (2023). *forcats: Tools for Working with Categorical Variables (Factors)*. R package version 1.0.0, https://CRAN.R-project.org/package=forcats.
- Iannone R, Cheng J, Schloerke B, Hughes E, Lauer A, Seo J (2024). *gt: Easily Create Presentation-Ready Display Tables*. R package version 0.10.1, https://CRAN.R-project.org/package=gt.
- Wickham H, François R, Henry L, Müller K, Vaughan D (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://CRAN.R-project.org/package=dplyr.
- A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18–22.
- Wickham H (2023). *modelr: Modelling Functions that Work with the Pipe*. R package version 0.1.11, https://CRAN.R-project.org/package=modelr.
- Xie Y (2023). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.45, https://yihui.org/knitr/.
- Therneau T, Atkinson B (2023). *rpart: Recursive Partitioning and Regression Trees*. R package version 4.1.23, https://CRAN.R-project.org/package=rpart.
- Williams, G. J. (2011), Data Mining with Rattle and R: The Art of Excavating Data for Knowledge Discovery, Use R!, Springer.
- Milborrow S (2024). *rpart.plot: Plot 'rpart' Models: An Enhanced Version of 'plot.rpart'*. R package version 3.1.2, https://CRAN.R-project.org/package=rpart.plot.
- Hvitfeldt E. (2021). paletteer: Comprehensive Collection of Color Palettes. version 1.3.0. https://github.com/EmilHvitfeldt/paletteer
- Neuwirth E (2022). *RColorBrewer: ColorBrewer Palettes*. R package version 1.1-3, https://CRAN.R-project.org/package=RColorBrewer.
- Jeppson H, Hofmann H, Cook D (2021). *ggmosaic: Mosaic Plots in the 'ggplot2' Frame-*

work. R package version 0.3.3, https://CRAN.R-project.org/package=ggmosaic.

**Faltó incluir en el documento la explicación
de la shiny app y como la pensaron y estructuraron.
Lograron que funcionara la reactividad
de la misma e incorporaron distintas visualizaciones
para explorar. Como les mencioné en la presentación hay
algún gráfico sin sentido pero hicieron un buen trabajo.**

**Como comentario general me gustó la presentación y el trabajo en equipo que hicieron. Deben revisar las faltas cuando escriben sus trabajos. Una debilidad del trabajo es el análisis exploratorio de datos, visualizaciones muy básicas y algunas sobreinterpretaciones. Además en el modelado la evaluación de los métodos al menos de árboles y bosques no es correcta y utizan menos datos para entrenar que para predecir. De todas formas estoy segura que van a seguir aprendiendo y profundizando a lo largo de los cursos de la licenciatura.
Total de puntos 85/100**