



Análisis y Predicción de Precios Inmobiliarios en Beijing

Isidro Perasso
Nazareno Ohanian
Andres Glauberman

Contexto y Oportunidad de Negocio



Objetivo del Análisis



Objetivo General

Que las propiedades se publiquen con un **precio justo** y que Lianjia pueda ajustar ese precio **dinámicamente**



Preguntas Clave

¿Qué **factores** explican mejor el **precio**?
¿Cómo **interactúan** los factores?
¿Se puede predecir el precio con un **modelo confiable**?



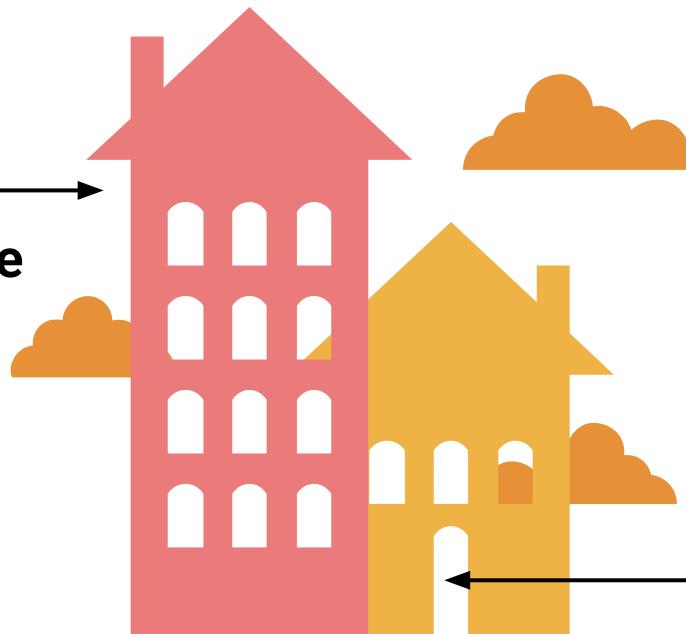
Relevancia

- Precisión
- Transparencia
- Adaptación

Metodología y Datos Utilizados

Dataset: Housing price in Beijing (Kaggle)

- Variables estructurales
- Variables de localización
- Variables de mercado



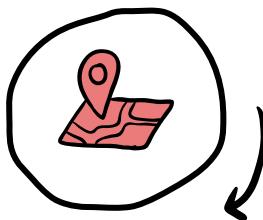
Transformaciones

- Limpieza
 - Tratamiento de Outliers
 - Manejo de Faltantes
 - Ingeniería de Variables
-

Hallazgos e insights principales



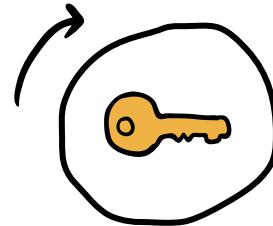
Segmentación
por tamaño



Ubicación y
accesibilidad



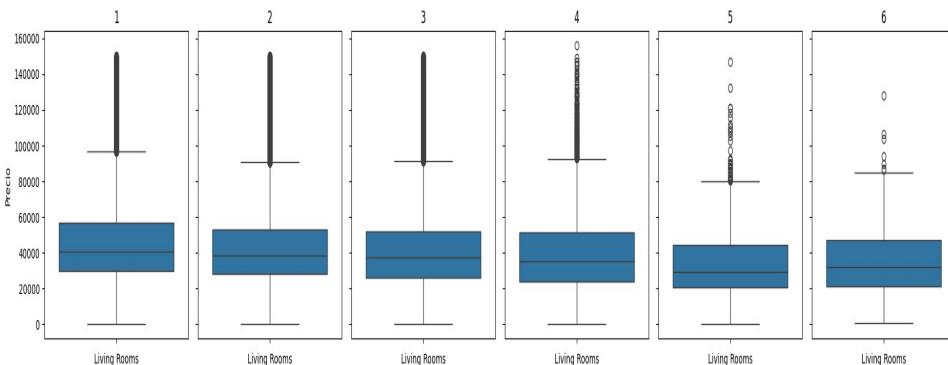
Calidad y tipo de
construcción



Popularidad

Hipótesis 1

“Los departamentos de menos ambientes presentan un precio por m² más alto que los de mayor tamaño”



El precio por m² es mayor para departamentos con menos ambientes, salvo cuando se comparan propiedades de 5 contra 6 ambientes

Hipótesis 2

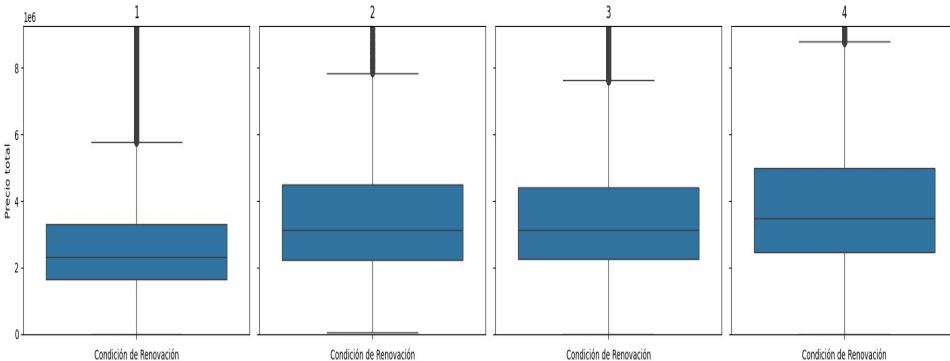
“La superficie es el factor principal del precio”

correlacion con el precio

square	0.575849
bathRoom	0.442436
livingRoom	0.434316
communityAverage	0.421396
tradeTime	0.400807
drawingRoom	0.322709
followers	0.152645
constructionTime	0.123158
kitchen	0.086426
Lat	0.019941
ladderRatio	-0.001064
Lng	-0.070036
distancia_km_a_subte_mas_cercano	-0.103786

Hipótesis 3

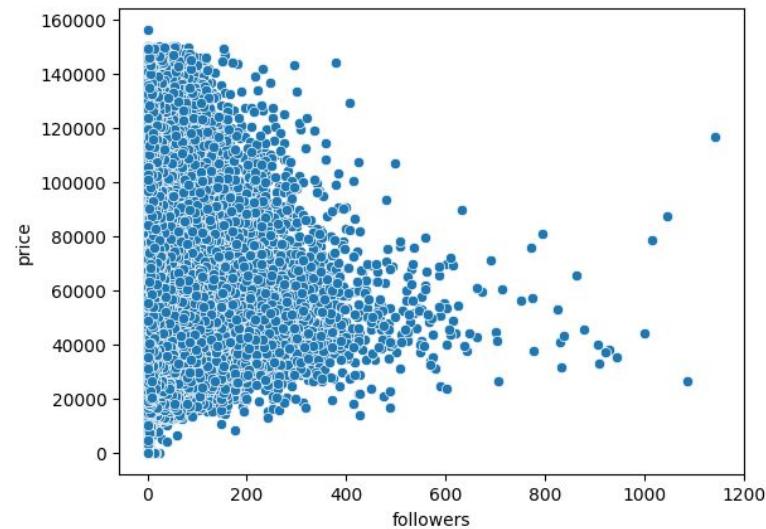
“La condición de renovación tiene un gran impacto en el precio”



Se puede ver que hay diferencias en algunos de los precios dependiendo de la condición de renovación

Hipótesis 4

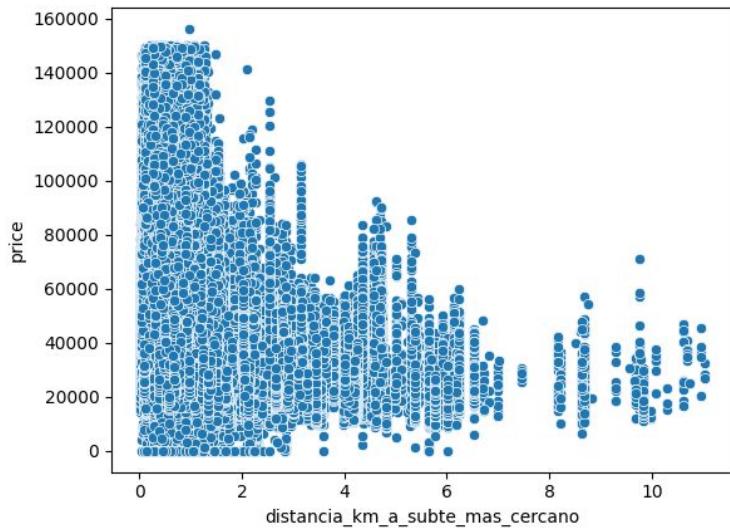
“La cantidad de seguidores tiene una alta correlación con el precio por m²”



Coeficiente de correlación de Pearson entre los seguidores y el precio: 0.2571

Hipótesis 5

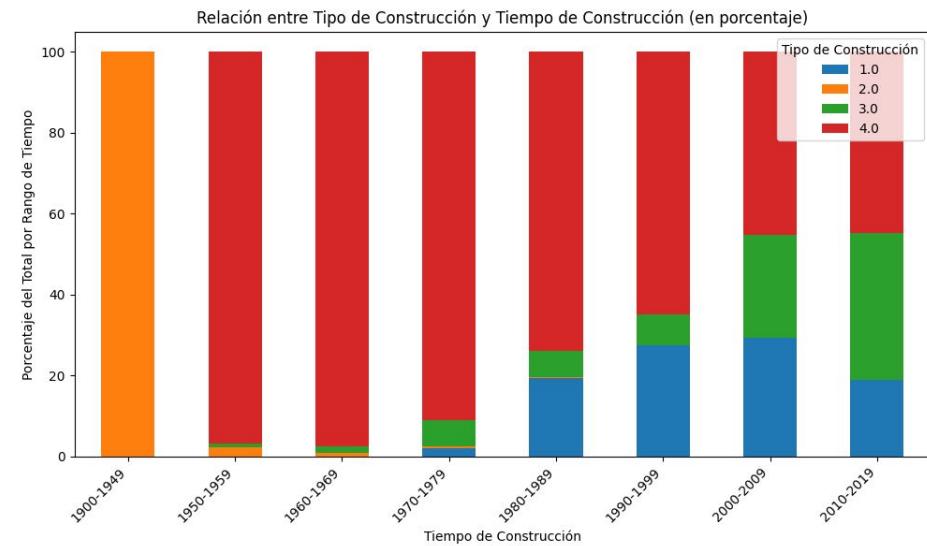
“Hay un impacto en el precio por m² por la cercanía al metro”



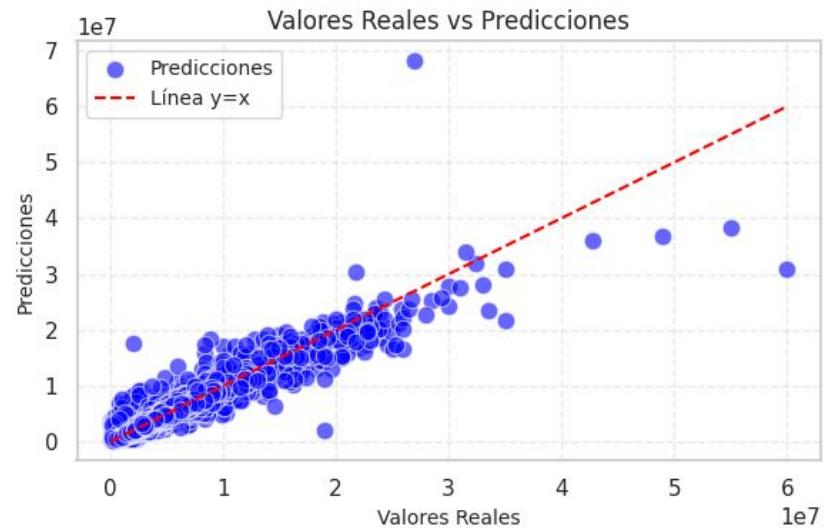
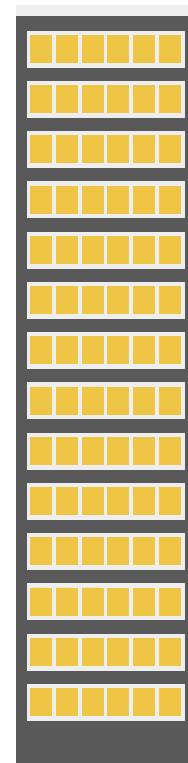
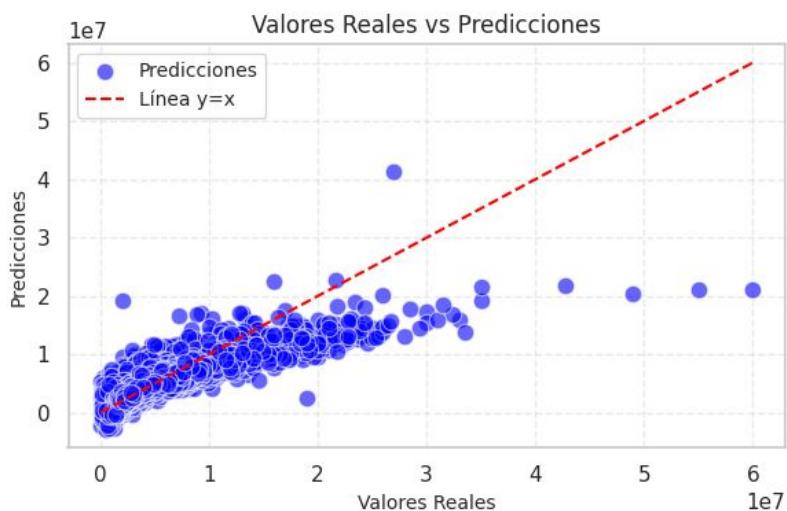
Coeficiente de correlación de Pearson entre la distancia al subte y el precio: -0.2375

Hipótesis 6

“Hay determinados tipos de construcciones/edificios que solo se relacionan con ciertos rangos de construcción de los edificios”



Modelo predictivo: Random Forest

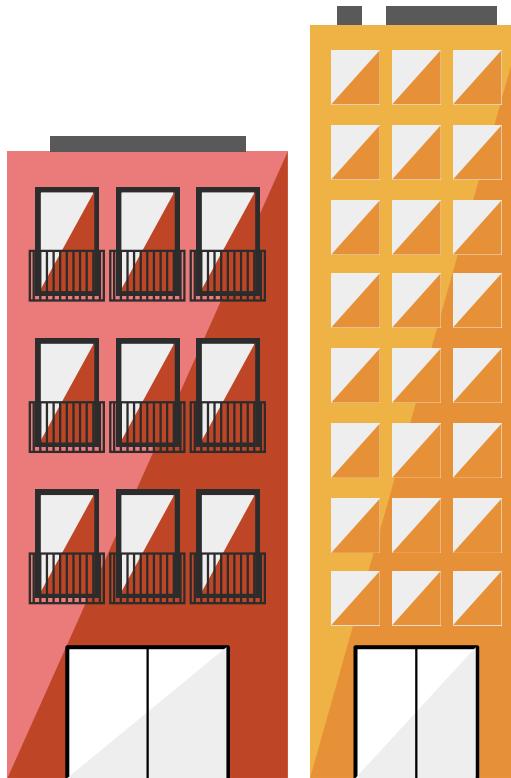


Modelo predictivo: Random Forest

94%

Es la variabilidad total que explica el modelo (R^2)

- Modelo regresión simple: 77.9%



15.7%

Es el % que representa el error promedio (RMSE) del modelo respecto de la media de valores de venta.

- Promedio de propiedad: ¥3.5 millones
- RMSE: ¥551k. Eso es ¥79k menos que el óptimo anterior.

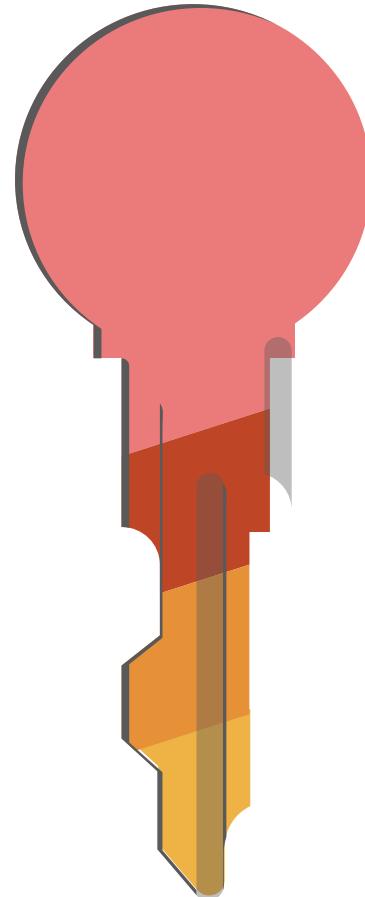
Puesta en producción



Estático



Dinámico



Próximos pasos y líneas futuras de trabajo

