

# PROYECTO ANÁLISIS DEL MERCADO INMOBILIARIO MEDIANTE REGRESIÓN LINEAL SIMPLE

*Sebastian Rincon,Alejandro Triana,Luis  
Bueno,Jose Pulido*



# DESCRIPCION DEL DATASET

contiene información sobre propiedades inmobiliarias listadas en la plataforma Properati. Cada fila del dataset representa una propiedad y contiene diversas características como fechas relevantes, ubicación geográfica, número de habitaciones, superficie, precio, tipo de propiedad, y más

	start_date	end_date	created_on	lat	lon	I1	I2	I3	rooms	bedrooms	bathrooms
0	2019-10-17	2019-12-23	2019-10-17	-34.605880	-58.384949	Argentina	Capital Federal	San Cristobal	7.0	7.0	2.0
1	2019-10-17	2019-11-21	2019-10-17	-34.624056	-58.412110	Argentina	Capital Federal	Boedo	2.0	1.0	2.0
2	2019-10-17	2019-11-01	2019-10-17	-34.593569	-58.427474	Argentina	Capital Federal	Palermo	2.0	1.0	1.0
3	2019-10-17	2019-12-23	2019-10-17	-34.581294	-58.436754	Argentina	Capital Federal	Palermo	2.0	1.0	1.0
4	2019-10-17	2020-03-11	2019-10-17	-34.914194	-57.938219	Argentina	Bs.As. G.B.A. Zona Sur	La Plata	2.0	1.0	1.0

# DESCRIPCION DEL DATASET



Fechas:

- start\_date: Fecha de inicio de la publicación.
- end\_date: Fecha de finalización de la publicación.
- created\_on: Fecha de creación de la publicación.

Ubicación:

- lat y lon: Latitud y longitud de la propiedad.
- l1, l2, l3: Localizaciones a distintos niveles (país, estado/provincia, ciudad/barrio).



# DESCRIPCION DEL DATASET

Propiedad:

- rooms: Número total de habitaciones.
- bedrooms: Número de dormitorios.
- bathrooms: Número de baños.
- surface\_total: Superficie total de la propiedad en metros cuadrados.
- surface\_covered: Superficie cubierta de la propiedad en metros cuadrados.
  - price: Precio de la propiedad.
  - currency: Moneda en la que está expresado el precio.
    - title: Título del anuncio.
    - description: Descripción de la propiedad.
- property\_type: Tipo de propiedad (casa, departamento, etc.).
- operation\_type: Tipo de operación (venta, alquiler, etc.).

# ANÁLISIS EXPLORATORIO DE DATOS (EDA)

El análisis exploratorio de datos tiene como objetivo comprender mejor el dataset, identificar patrones, valores atípicos, y posibles problemas con los datos

Para obtener un buen análisis se pueden usar los siguientes pasos

- 1 Carga y Visualización de los Datos
- 2 Verificación de Datos Faltantes
- 3 Estadísticas Descriptivas
- 4 Distribución Geográfica
- 5 Distribución del Precio y Superficie
- 6 Relaciones entre Variables
- 7 Análisis de Categorías
- 8 Análisis Temporal

# MODELO DE REGRESIÓN LINEAL

El objetivo principal del modelo de regresión lineal es predecir el precio de una propiedad inmobiliaria (price) en función del número de habitaciones (rooms). La regresión lineal es una técnica adecuada para este tipo de problema cuando se espera que la relación entre las variables sea aproximadamente lineal

Se implementó la función leerCSV para cargar los datos del archivo CSV en una estructura de Propiedad. Esto permitió acceder a las características de cada propiedad y utilizar estos datos para entrenar el modelo

# MODELO DE REGRESION LINEAL

Se implementa el modelo de regresión lineal simple utilizando la fórmula de mínimos cuadrados para ajustar una línea recta a los datos. Los coeficientes se calcularon de la siguiente manera:

- Pendiente ( $b_1$ ): Representa el cambio promedio en el precio por cada incremento en el número de habitaciones.
- Intersección ( $b_0$ ): Representa el precio predicho cuando el número de habitaciones es 0.

# MODELO DE REGRESIÓN LINEAL

Para evaluar el rendimiento del modelo, se calcularon dos métricas clave:

Error Cuadrático Medio (MSE): El MSE mide la diferencia promedio al cuadrado entre los valores observados y los valores predichos por el modelo. Es sensible a los valores atípicos, lo que significa que un MSE alto puede indicar la presencia de outliers significativos o un mal ajuste del modelo.

Coeficiente de Determinación ( $R^2$ ): El  $R^2$  indica qué proporción de la variabilidad en el precio de las propiedades es explicada por el modelo. Un  $R^2$  cercano a 1 indica que el modelo explica bien la variabilidad de los datos, mientras que un valor cercano a 0 sugiere que el modelo es ineficaz para predecir el precio basado en el número de habitaciones

# RESULTADOS DEL MODELO Y ANÁLISIS DE SU PRECISIÓN

El MSE mide el promedio de los errores al cuadrado entre los valores observados y los predichos. Un valor bajo de MSE indica que las predicciones están, en promedio, cerca de los valores reales

El R<sup>2</sup> mide la proporción de la variabilidad en los precios que puede ser explicada por el modelo

Run	R <sup>2</sup> : 0.00958265	Ask AI					
Run	Run	Ask AI					
Start Date	End Date	Created On	Lat	Lon	Location 1	Location 2	Location 3
"2019-10-17	2019-12-23	2019-10-17	-34.6059	-58.3849	Argentina	Capital Federal	San Cristobal
"2019-10-17	2019-11-21	2019-10-17	-34.6241	-58.4121	Argentina	Capital Federal	Boedo
"2019-10-17	2019-11-01	2019-10-17	-34.5936	-58.4275	Argentina	Capital Federal	Palermo
"2019-10-17	2019-12-23	2019-10-17	-34.5813	-58.4368	Argentina	Capital Federal	Palermo
"2019-10-17	2020-03-11	2019-10-17	-34.9142	-57.9382	Argentina	Bs.As. G.B.A. Zona Sur	La Plata
"2019-10-17	2019-11-01	2019-10-17	-34.5938	-58.4412	Argentina	Capital Federal	Villa Crespo
"2019-10-17	2019-10-31	2019-10-17	-34.601	-58.4331	Argentina	Capital Federal	Villa Crespo
"2019-10-17	2019-10-31	2019-10-17	-34.601	-58.4331	Argentina	Capital Federal	Villa Crespo
"2019-10-17	2019-12-23	2019-10-17	-34.6396	-58.4196	Argentina	Capital Federal	Parque Patricios
"2019-10-17	2019-12-23	2019-10-17	-34.6396	-58.4196	Argentina	Capital Federal	Parque Patricios
"2019-10-17	2019-12-07	2019-10-17	-34.5862	-58.4962	Argentina	Capital Federal	Villa Pueyrredón
"2019-10-17	2019-10-26	2019-10-17	-34.6037	-58.3816	Argentina	Capital Federal	Boedo
"2019-10-17	2019-12-19	2019-10-17	-34.6303	-58.4172	Argentina	Capital Federal	Boedo
"2019-10-17	2019-11-30	2019-10-17	-34.546	-58.8594	Argentina	Bs.As. G.B.A. Zona Norte	Pilar
"2019-10-17	2019-12-23	2019-10-17	-34.4529	-58.899	Argentina	Bs.As. G.B.A. Zona Norte	Pilar
"2019-10-17	2019-12-23	2019-10-17	-34.3954	-58.6499	Argentina	Bs.As. G.B.A. Zona Norte	Tigre
"2019-10-17	2019-12-23	2019-10-17	-34.4251	-58.5797	Argentina	Bs.As. G.B.A. Zona Norte	Tigre
"2019-10-17	2019-10-22	2019-10-17	0	0	Argentina	Bs.As. G.B.A. Zona Norte	Tigre
"2019-10-17	2019-10-18	2019-10-17	-34.423	-58.5734	Argentina	Bs.As. G.B.A. Zona Norte	Tigre
"2019-10-17	2019-12-23	2019-10-17	-34.3995	-58.6491	Argentina	Bs.As. G.B.A. Zona Norte	Tigre
"2019-10-17	2019-10-24	2019-10-17	-34.6082	-58.4443	Argentina	Capital Federal	Abasto
"2019-10-17	2019-10-28	2019-10-17	-34.6039	-58.411	Argentina	Capital Federal	Abasto
"2019-10-17	2019-11-13	2019-10-17	-34.6262	-58.4621	Argentina	Capital Federal	Flores
2019-10-17	2019-12-23	2019-10-17	-34.6256	-58.4624	Argentina	Capital Federal	Flores
"2019-10-17	2019-12-23	2019-10-17	-34.6367	-58.4712	Argentina	Capital Federal	Flores
"2019-10-17	2019-12-23	2019-10-17	-34.6339	-58.47	Argentina	Capital Federal	Flores
"2019-10-17	9999-12-31	2019-10-17	0	0	Argentina	Bs.As. G.B.A. Zona Oeste	Morón
2019-10-17	2019-12-23	2019-10-17	-34.5504	-58.4616	Argentina	Capital Federal	Núñez
"2019-10-17	2019-12-23	2019-10-17	-34.5448	-58.4716	Argentina	Capital Federal	Núñez
"2019-10-17	2019-11-01	2019-10-17	-34.6022	-58.4224	Argentina	Capital Federal	Almagro
"2019-10-17	2019-10-31	2019-10-17	-34.6035	-58.4146	Argentina	Capital Federal	Almagro
"2019-10-17	2019-10-27	2019-10-17	-34.6033	-58.428	Argentina	Capital Federal	Almagro
"2019-10-17	2019-11-01	2019-10-17	-34.604	-58.429	Argentina	Capital Federal	Almagro
"2019-10-17	2019-11-01	2019-10-17	-34.6072	-58.3707	Argentina	Capital Federal	Almagro
"2019-10-17	2019-10-31	2019-10-17	-34.59	-58.4238	Argentina	Capital Federal	Palermo
"2019-10-17	2019-12-23	2019-10-17	0	0	Argentina	Capital Federal	Palermo
2019-10-17	2019-10-27	2019-10-17	0	0	Argentina	Capital Federal	Palermo
"2019-10-17	2019-12-23	2019-10-17	-34.5844	-58.4198	Argentina	Capital Federal	Palermo
"2019-10-17	2019-10-22	2019-10-17	-34.5976	-58.4167	Argentina	Capital Federal	Palermo
"2019-10-17	2019-12-23	2019-10-17	-34.5833	-58.4262	Argentina	Capital Federal	Palermo
"2019-10-17	2019-10-22	2019-10-17	-34.5976	-58.4167	Argentina	Capital Federal	Palermo
"2019-10-17	2019-12-21	2019-10-17	-34.5753	-58.4289	Argentina	Capital Federal	Palermo

# RESULTADOS DEL MODELO Y ANÁLISIS DE SU PRECISIÓN

Correlación entre superficie total y precio: 0.0575466

Modelo de Regresión Lineal: Precio = 235432 + 7.8497 \* Superficie Total

MSE: 8.23428e+10

R<sup>2</sup>: 0.00331162

Evaluación del Modelo en el Conjunto de Prueba:

MSE: 7.86578e+10

R<sup>2</sup>: 0.00657485

UN VALOR DE MSE BAJO SUGIERE QUE EL MODELO TIENE UN BUEN RENDIMIENTO PREDICTIVO EN TÉRMINOS DE MINIMIZAR EL ERROR PROMEDIO AL CUADRADO

UN VALOR DE R<sup>2</sup> ALTO INDICA QUE EL MODELO CAPTURA BIEN LA RELACIÓN ENTRE EL NÚMERO DE HABITACIONES Y EL PRECIO

# CONCLUSION Y POSIBLES MEJORAS

LA PRECISION DE NUESTRO MODELO ES UN POCO LIMITADA YA QUE UN MODELO DE REGRESION LINEAL SIMPLE NO ES ADECUADO PARA MANEJAR DATOS TAN COMPLEJOS Y SERIA NECESARIO EXPLORAR MÁS VARIABLES, APLICAR TRANSFORMACIONES ADECUADAS, O CONSIDERAR MODELOS MÁS COMPLEJOS QUE PUEDAN CAPTURAR RELACIONES NO LINEALES O INTERACCIONES ENTRE MÚLTIPLES FACTORES



**Gracias  
por ver**