



PREDICCIÓN DE PRECIOS DE VIVIENDA EN BOGOTÁ

Usando Machine Learning

Presentado por: Esteban Hincapie - Alexander Ossa - Maryory - Oscar Medina - Aydee López

Tabla de contenido

1. Objetivo del problema y explicación del proyecto
2. Introducción
3. Análisis y preparación de los datos
4. Justificación
5. Presentación del modelo
6. Resultados obtenidos y conclusiones
7. Link GitHub
8. Referencias



1. Explicación del problema

Objetivo:

diseñar un modelo de regresión lineal, que ayude a predecir cuanto puede costar una vivienda en la ciudad de Bogotá, mediante técnicas de Machine Learning.

Explicación del problema:

2. Introducción

La compra de una vivienda es una decisión sin precedentes y sin comparación para cada comprador o en sentido contrario para los vendedores.



3. Análisis y preparación de los datos

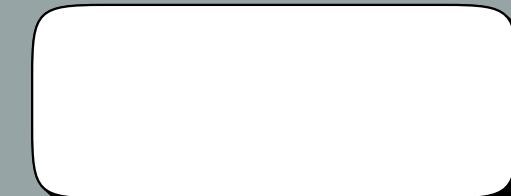
diseñar un modelo de regresión lineal, que ayude a predecir cuanto puede costar una vivienda en la ciudad de Bogotá, mediante técnicas de Machine Learning.



3. Análisis y preparación de los datos

Tabla 1. Distribución de los datos

Total

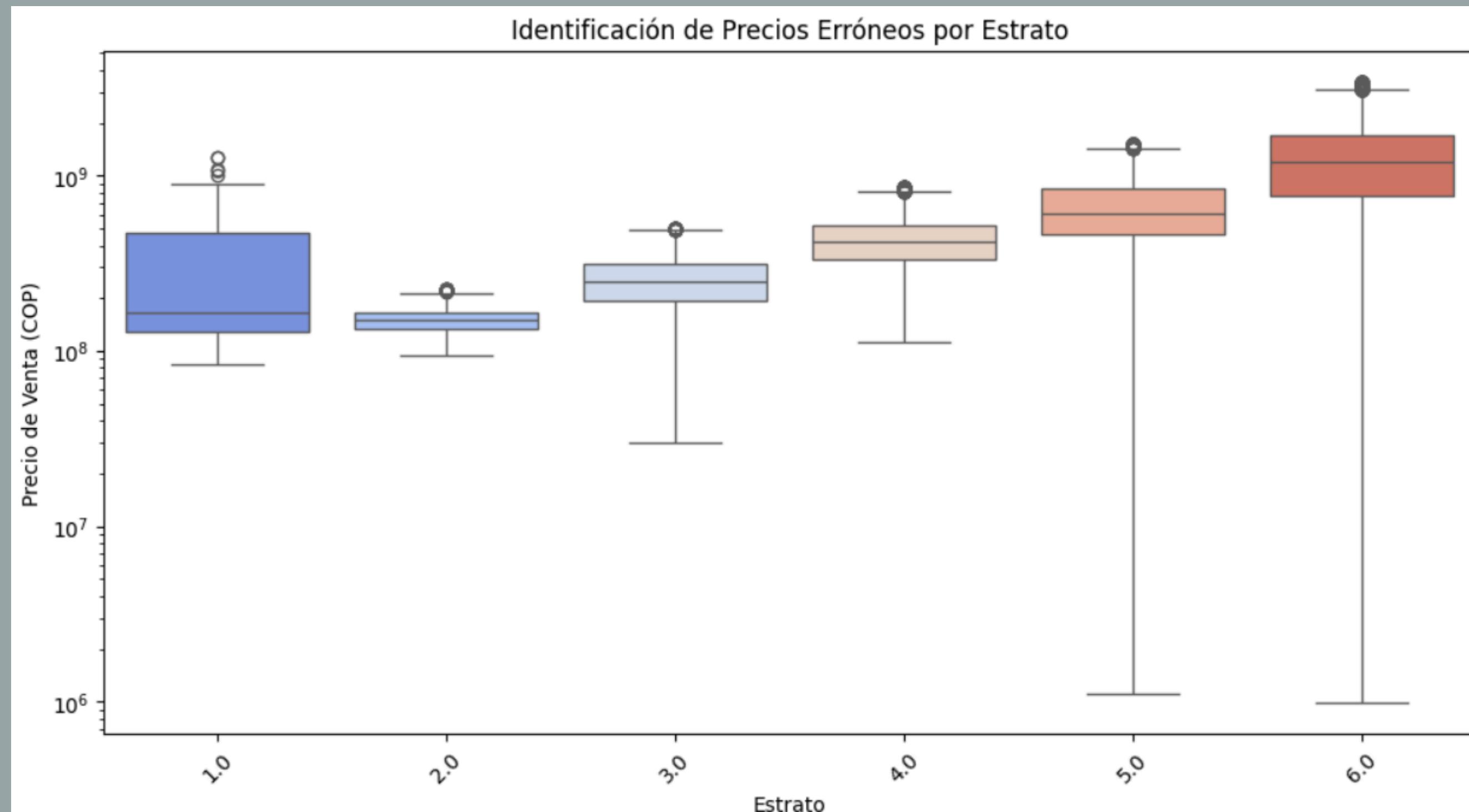


3. Análisis y preparación de los datos

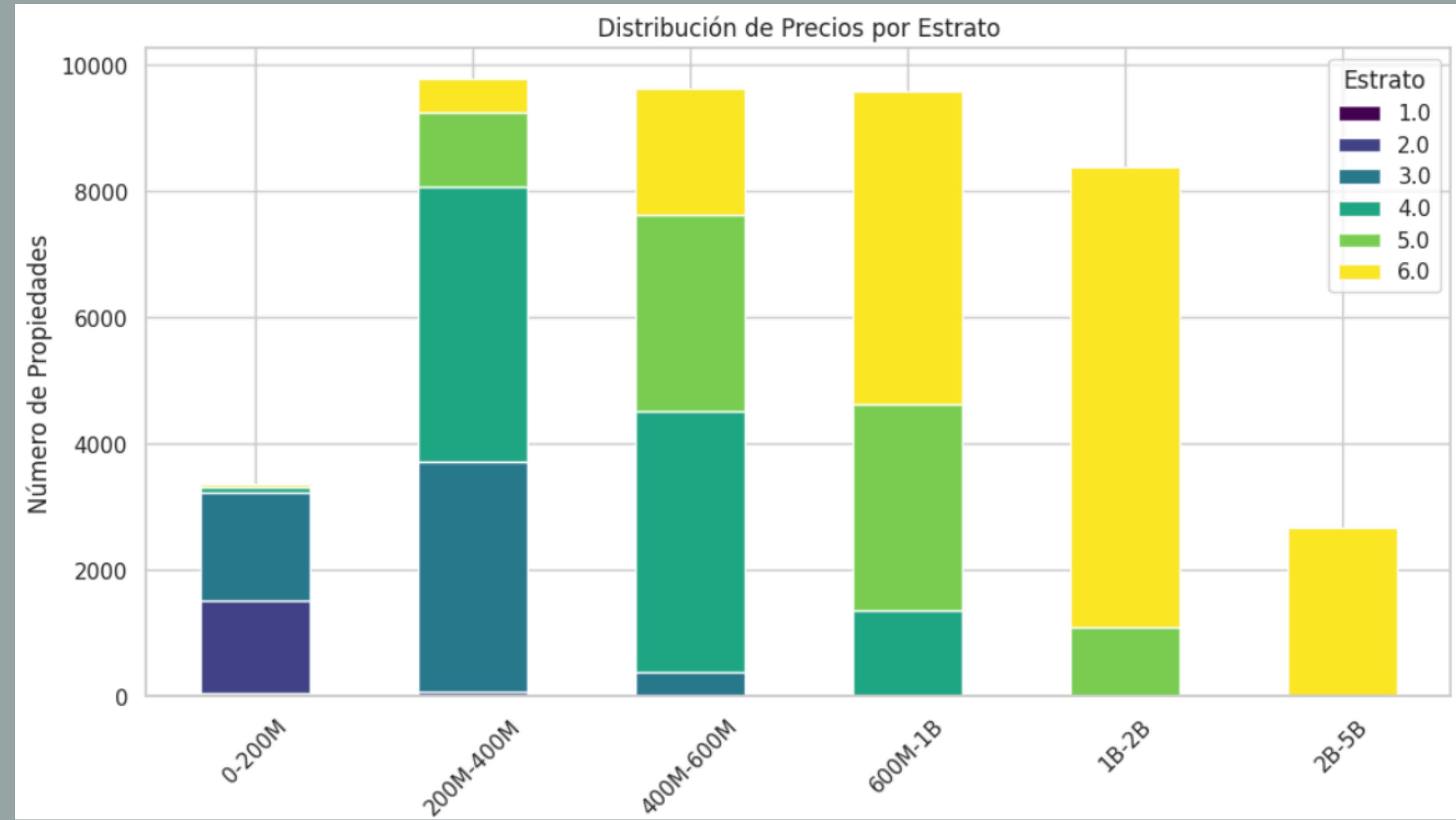
Información del DataSet

```
<ipython-input-79-982ba5e75df6>:2: DtypeWarning: Columns (7,13,14) have mixed types. Specify dtype option on import or set low_memory=False.  
raw_data = pd.read_csv("/content/builder_raw_data.csv")  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 68973 entries, 0 to 68972  
Data columns (total 16 columns):  
 #   Column      Non-Null Count  Dtype     
---  --          --          --  
 0   codigo       68973 non-null   object    
 1   tipo_propiedad 68973 non-null   object    
 2   tipo_operacion 68973 non-null   object    
 3   precio_venta  48138 non-null   float64   
 4   area          68971 non-null   float64   
 5   habitaciones  68965 non-null   float64   
 6   banos         68967 non-null   float64   
 7   administracion 58679 non-null   object    
 8   parqueaderos  68965 non-null   float64   
 9   sector         63998 non-null   object    
 10  estrato        68953 non-null   float64   
 11  antiguedad    68945 non-null   object    
 12  estado         67692 non-null   object    
 13  longitud       68164 non-null   object    
 14  latitud         68164 non-null   object    
 15  datetime       68973 non-null   object    
dtypes: float64(6), object(10)  
memory usage: 8.4+ MB  
 codigo  tipo_propiedad  tipo_operacion  precio_venta  area  habitaciones  banos  administracion  parqueaderos  sector  estrato  antiguedad  estado  longitud  latitud
```

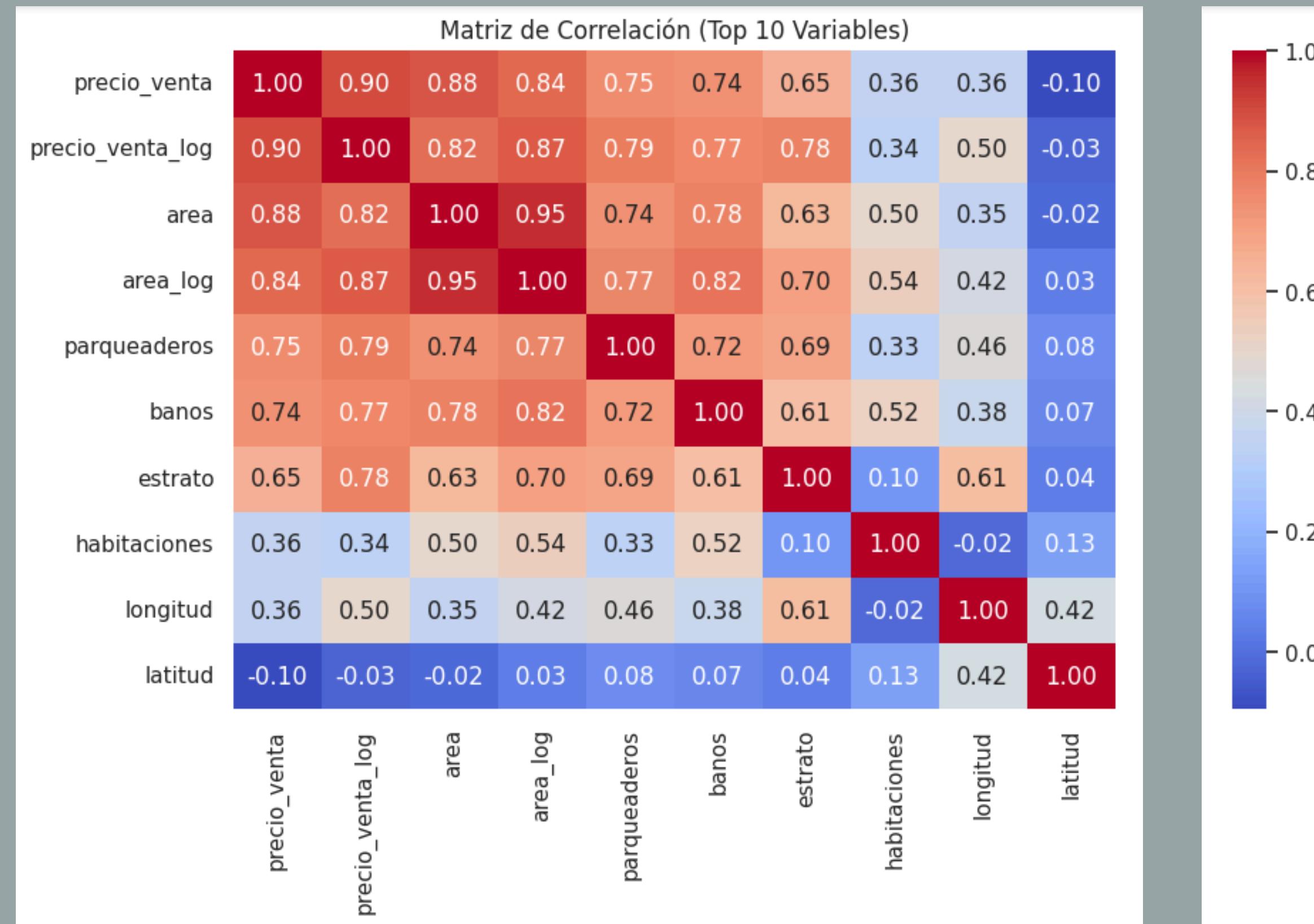
Identificación de precios erróneos por estrato



Distribución de precios por estrato



Matriz de correlación



4. Justificación

Resultados modelos de entrenamiento

Algoritmo	R2	MAE	MSE
LinearRegression	0.8394	163498508.3267	64920761547310752.00 00
DecisionTreeRegressor	0.85300	116624783.70899697	5.943080859298668e+1 6
KNeighborsRegressor	0.84075	0.117529	0.091542
XGBRegressor	-0.121656	45376.503696706284	112893755729.40189
Random Forest	0.8658	146019227.1574	54240268255285272.00 00

Regresión como herramienta de análisis:

- Los precios de las casas dependen de múltiples factores (ubicación, tamaño, antigüedad, estrato, número cuartos, entre otros).
- Permite modelar la relación entre las variables independientes (factores de la vivienda) y el precio

5. Presentación del modelo

Modelo

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit. Quisque
non elit mauris. Cras euismod, metus a**

Preparación de datos

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit. Quisque
non elit mauris. Cras euismod, metus a**

Predicciones iniciales

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit. Quisque
non elit mauris. Cras euismod, metus a**

Evaluación final del modelo

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit. Quisque
non elit mauris. Cras euismod, metus a**

Resultados obtenidos y conclusiones

```
→ <class 'pandas.core.frame.DataFrame'>
RangeIndex: 43404 entries, 0 to 43403
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   precio_venta    43404 non-null   float64
 1   area              43404 non-null   float64
 2   habitaciones      43404 non-null   float64
 3   banos             43404 non-null   float64
 4   administracion   43404 non-null   float64
 5   parqueaderos      43404 non-null   float64
 6   estrato           43404 non-null   float64
 7   antiguedad        43404 non-null   int64  
 8   longitud          43404 non-null   float64
 9   latitud            43404 non-null   float64
 10  area_log          43404 non-null   float64
 11  precio_venta_log 43404 non-null   float64
dtypes: float64(11), int64(1)
memory usage: 4.0 MB


- MAE (Error Absoluto Medio): 2978557.30
- MSE (Error Cuadrático Medio): 65133832698170.23
- RMSE (Raíz del Error Cuadrático Medio): 8070553.43


R2: 0.9998388982234899
```

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit. Quisque
non elit mauris. Cras euismod,
metus a**

**Lorem ipsum dolor sit amet,
consectetur adipiscing elit. Quisque
non elit mauris. Cras euismod,
metus a**

Resultados obtenidos y conclusiones

7. Link de GitHub

Link del perfil

Link del proyecto

8. Referencias

**Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque non elit
 mauris. Cras euismod, metus a**

The background image shows a panoramic view of the Bogotá city skyline during sunset. The sky is filled with large, billowing clouds colored in shades of orange, yellow, and grey. In the foreground, numerous buildings of various heights are visible, their windows glowing with warm light. A prominent skyscraper on the left side of the frame reaches towards the top of the image. The overall atmosphere is one of a bustling urban environment at dusk.

Gracias