

Predicción del Precio Real de Productos en Amazon

Aplicación de Machine Learning para mejorar
estrategias de pricing

POR ESTEFANY AMESTY

Problema de Negocio

Amazon necesita fijar precios reales coherentes y competitivos para sus productos. Actualmente, se observan inconsistencias en el precio base (sin descuento), lo que puede afectar la confianza del cliente y la estrategia comercial. Muchas veces los precios parecen inflados para mostrar descuentos llamativos.

Problema: No hay un sistema automatizado que permita estimar si el precio real publicado es correcto comparandolo con las características del producto.

Solución: Desarrollar un modelo de regresión que prediga el "precio real" estimado de un producto usando Machine Learning.

Explorando la idea

DATASET

El dataset contiene 1.465 productos con las siguientes columnas relevantes:

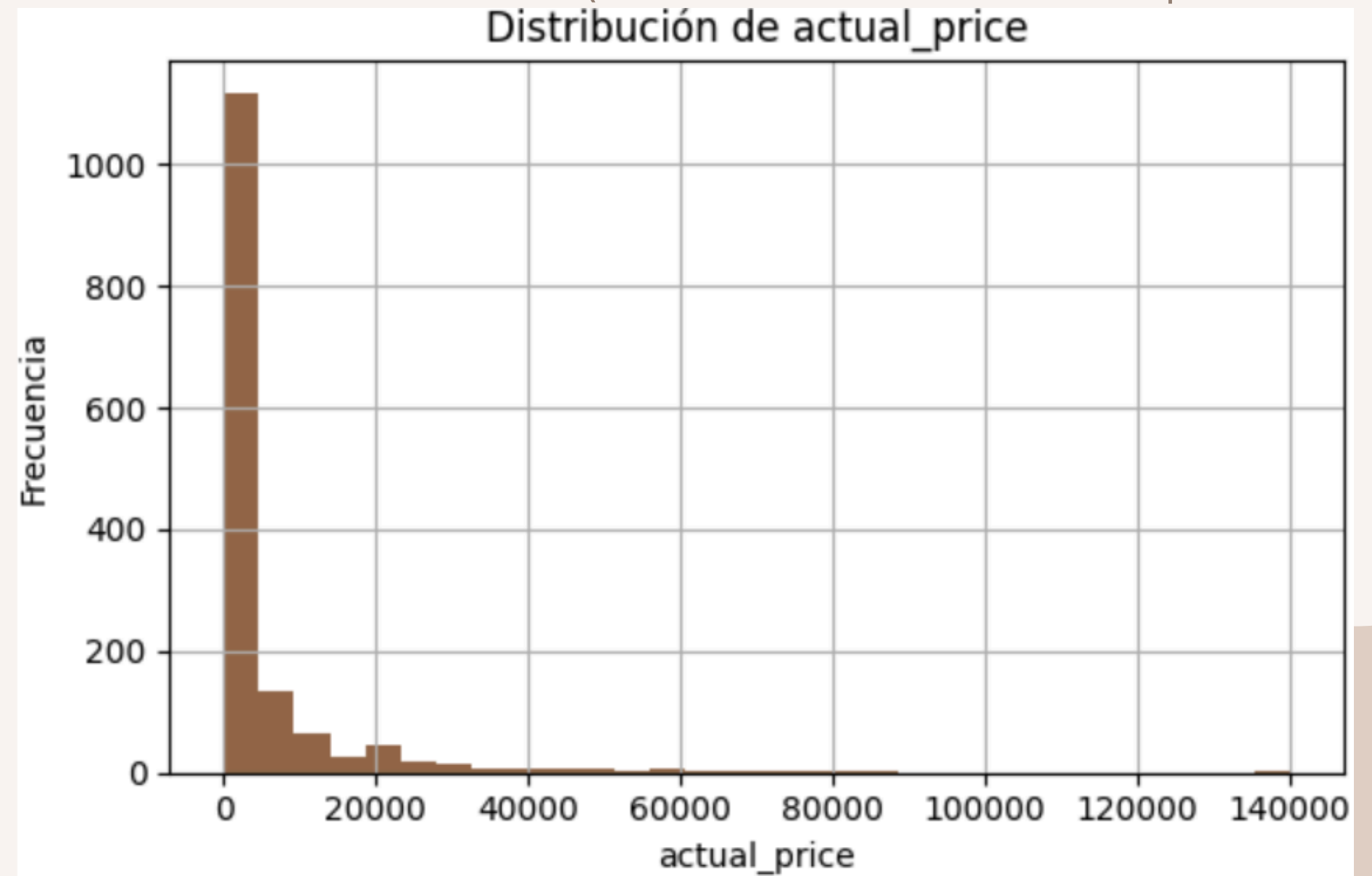
- discounted_price: precio con descuento
- actual_price: precio real (objetivo a predecir)
- discount_percentage: porcentaje de descuento
- rating: puntuación promedio del producto
- rating_count: número de valoraciones
- category: categoría jerárquica del producto (ej: "Electronics|TV")

FASE 1: MINI EDA

Se realizó un análisis exploratorio inicial para:

- Ver tipos de variables y valores nulos
- Convertir precios y ratings a formato numérico
- Explorar distribuciones (rating, precio, descuentos)
- Detectar columnas con muchos valores únicos o irrelevantes (como IDs, links, descripciones largas)

Conclusión: actual_price es una variable continua ideal para regresión.



FASE 2: PREPARACIÓN DE DATOS

SELECCIÓN DE COLUMNAS ÚTILES:

- Se eliminaron columnas de texto libre, IDs y URLs.
- Se conservaron: discounted_price, discount_percentage, rating, rating_count, category

IMPUTACIÓN DE VALORES NULOS:

- rating: rellenado con la media
- rating_count: rellenado con la mediana

CODIFICACIÓN DE VARIABLES CATEGÓRICAS:

- Se extrajo la categoría principal desde la columna category
- Se aplicó One-Hot Encoding

DIVISIÓN DEL DATASET:

- train: 80%
- test: 20%

FASE 3: MODELADO

Se probaron tres modelos base:

**Linear
Regression**

**RandomForest
Regressor**

**GradientBoosting
Regressor**

Se utilizó validación cruzada para comparar los modelos según su RMSE (raíz del error cuadrático medio).

Resultado:

Gradient Boosting fue el mejor (RMSE promedio: 2009.9)

FASE 4: OPTIMIZACIÓN DE HIPERPARAMETROS

Se usó GridSearchCV para ajustar:

N_ESTIMATORS: [100, 200, 300]

LEARNING_RATE: [0.01, 0.05, 0.1]

MAX_DEPTH: [3, 4, 5]

Mejor combinación encontrada:

`{'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 300}`

FASE 5: EVALUACIÓN DEL MODELO FINAL

Modelo: GradientBoostingRegressor entrenado con los mejores hiperparámetros.

Resultados en test:

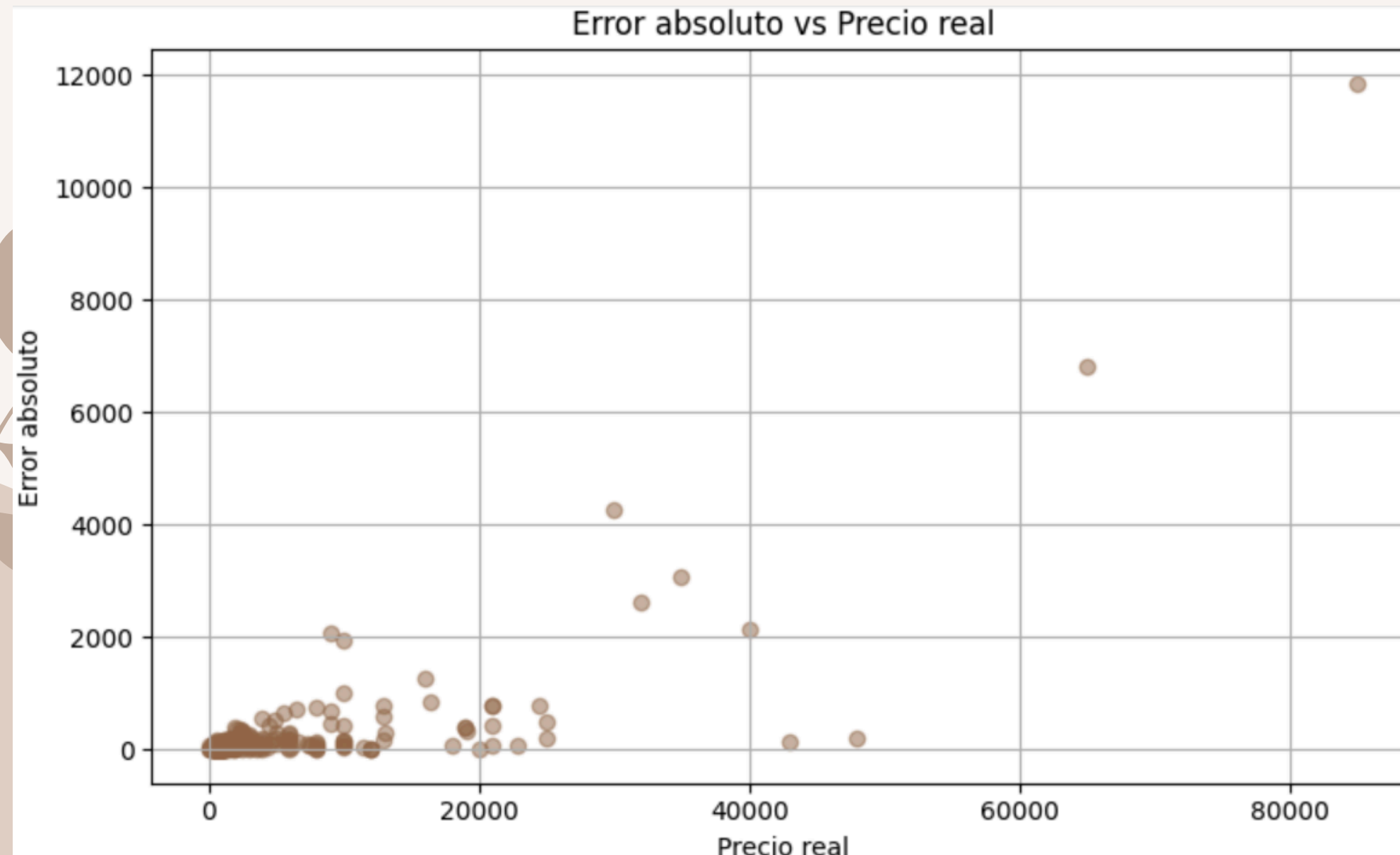
MAE: ₹234.49

RMSE: ₹918.79

R^2 : 0.9899 → El modelo explica el 99% de la variabilidad del precio real.

FASE 6: ANÁLISIS DE ERRORES

Se estudiaron los errores por:

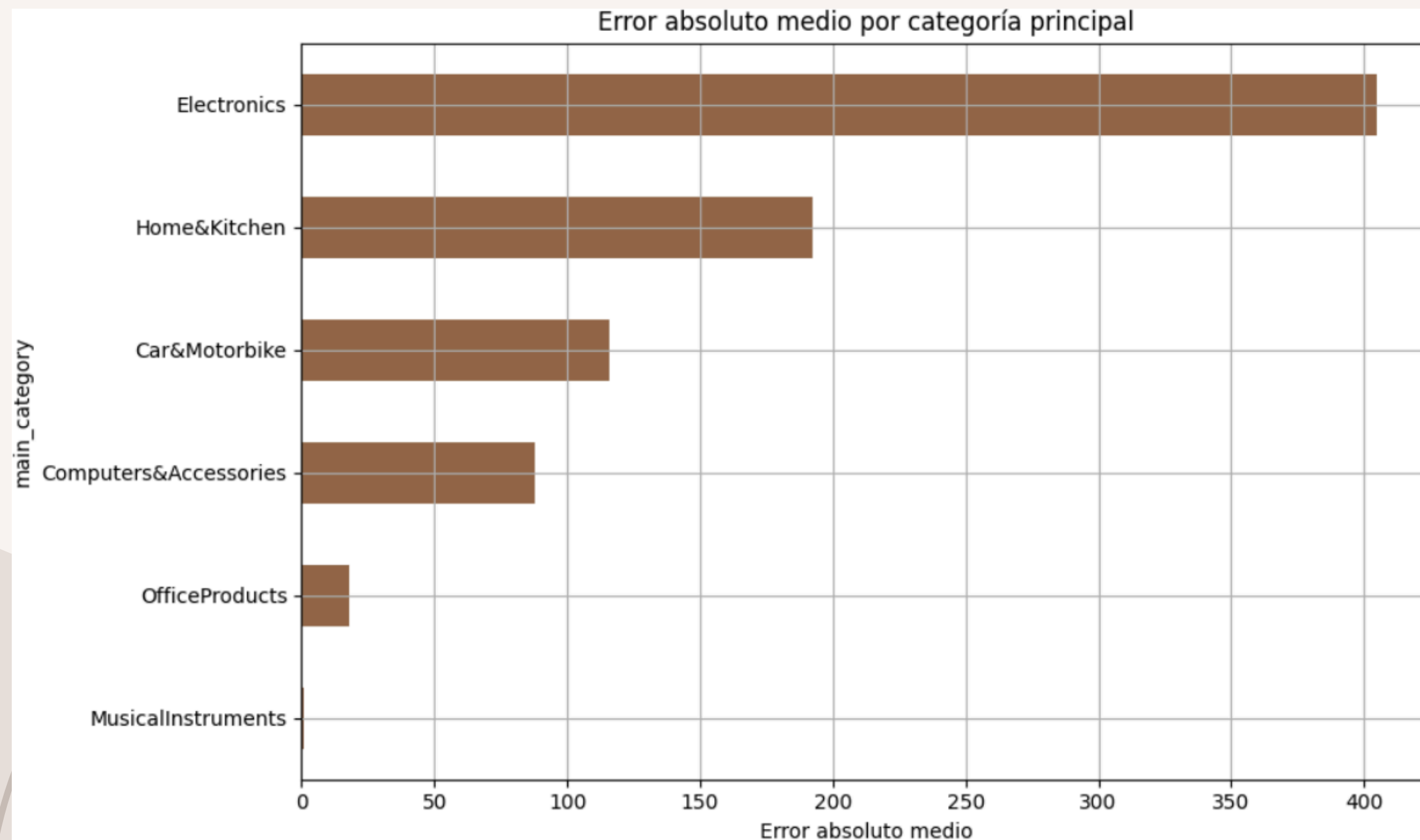


Precio real: el modelo comete errores más altos en productos muy caros

Categoría: algunas como "Electronics" tienen error absoluto alto pero bajo error relativo

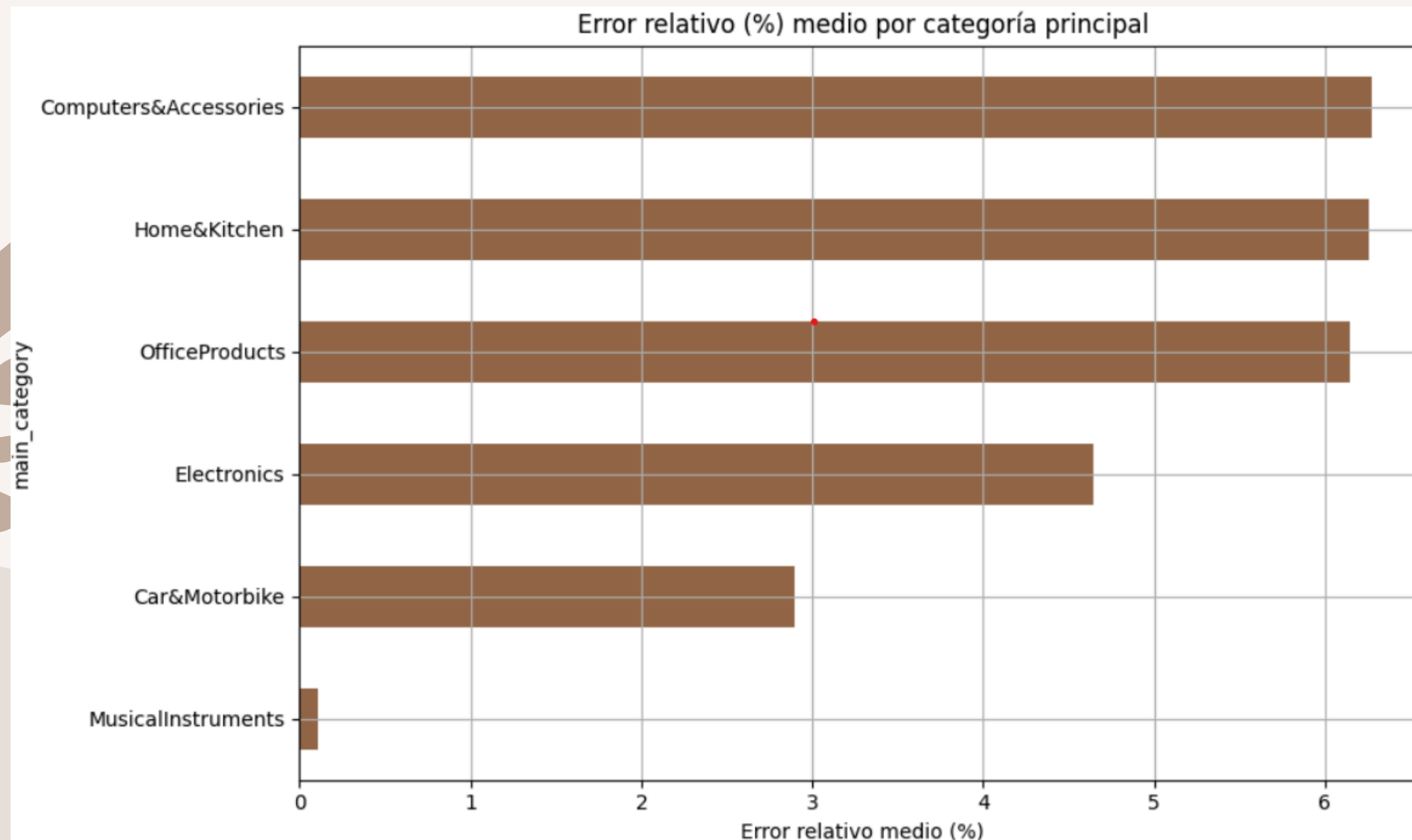
Error relativo por categoría: las más problemáticas fueron "MusicalInstruments" y "HomeImprovement"

ERROR ABSOLUTO POR CATEGORÍA



Electronics, HomeTheater, Computers tienen más error absoluto porque son productos caros

ERROR RELATIVO % POR CATEGORÍA



MusicalInstruments y HomeImprovement tienen errores porcentuales altos. Esto indica que en productos baratos, el modelo puede sobre o subestimar más en proporción.

Impacto de la marca

CONCLUSIÓN FINAL

El modelo de regresión entrenado permite predecir con alta precisión el precio real de productos a partir de sus características. Esto podría integrarse en un sistema de soporte para fijación de precios, ayudando a evitar incoherencias, mejorar el margen comercial, y ofrecer descuentos más honestos y competitivos.

The background is a light beige color. It features several decorative elements: a brown abstract shape in the top-left corner with white dots; a grey abstract shape in the top-right corner with a thin brown line; a brown leafy branch on the right side; a brown leafy branch on the left side; a grey abstract shape in the bottom-left corner with a thin brown line; and a brown abstract shape in the bottom-right corner with white dots.

¡MUCHAS GRACIAS!