



CA-0305: Herramientas de Ciencia de Datos II

Predicción de Precios de Casas

Ashley Arrieta Padilla – C00753

Sofía Bocker Brenes – C11102

Bryan Campos Vega – C01654

Naydelin Hernández Vargas – C03795

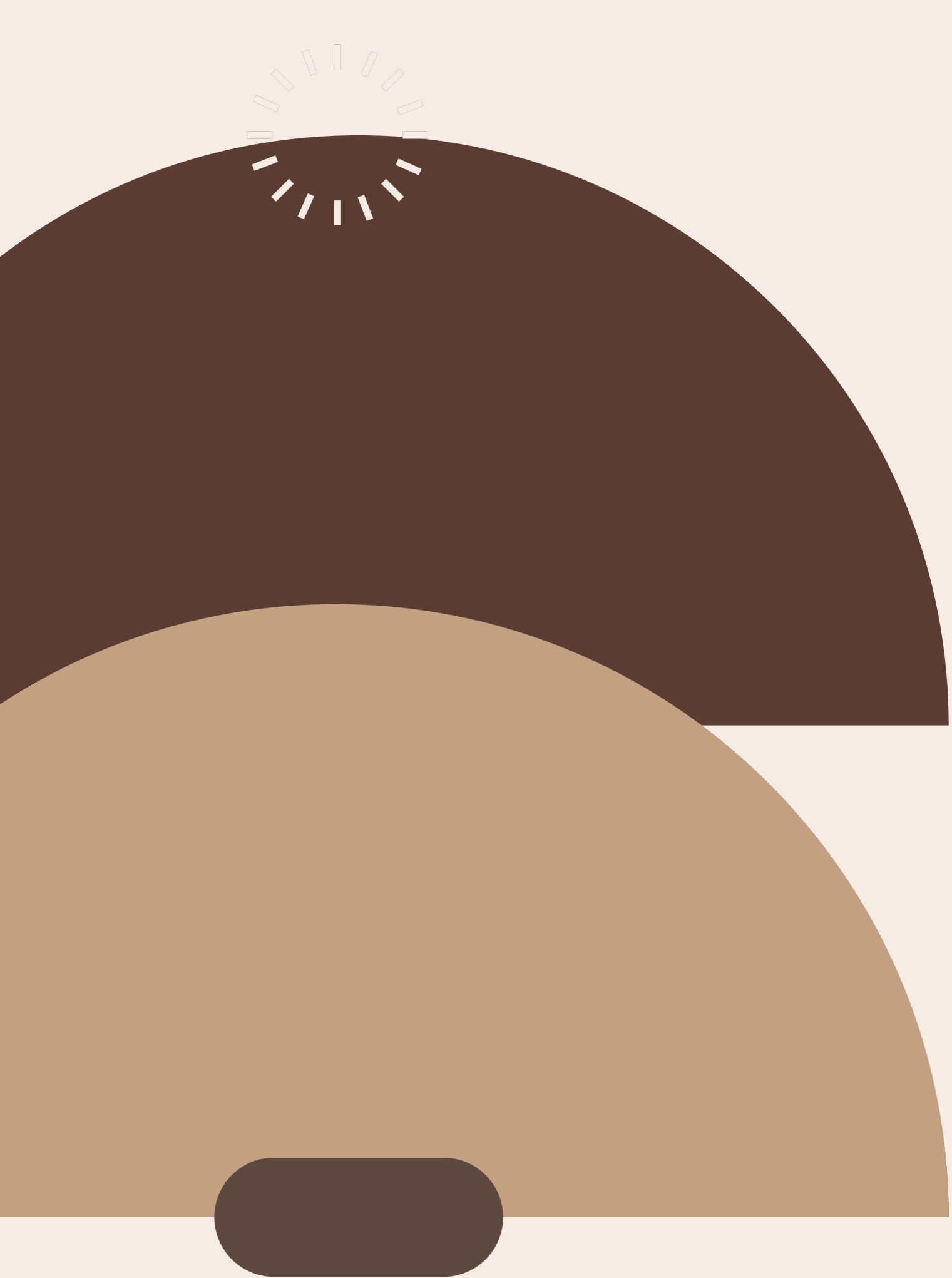
Dixon Montero Hernández – B99109

Contenido

- Tema Escogido
- Regresión Lineal Simple
- Regresión Polinomial
- Regresión Polinomial Ridge
- Regresión Elástica Net
- Regresión Lasso
- Resultados
- Dificultades
- Conclusiones y Recomendaciones

Tema Escogido

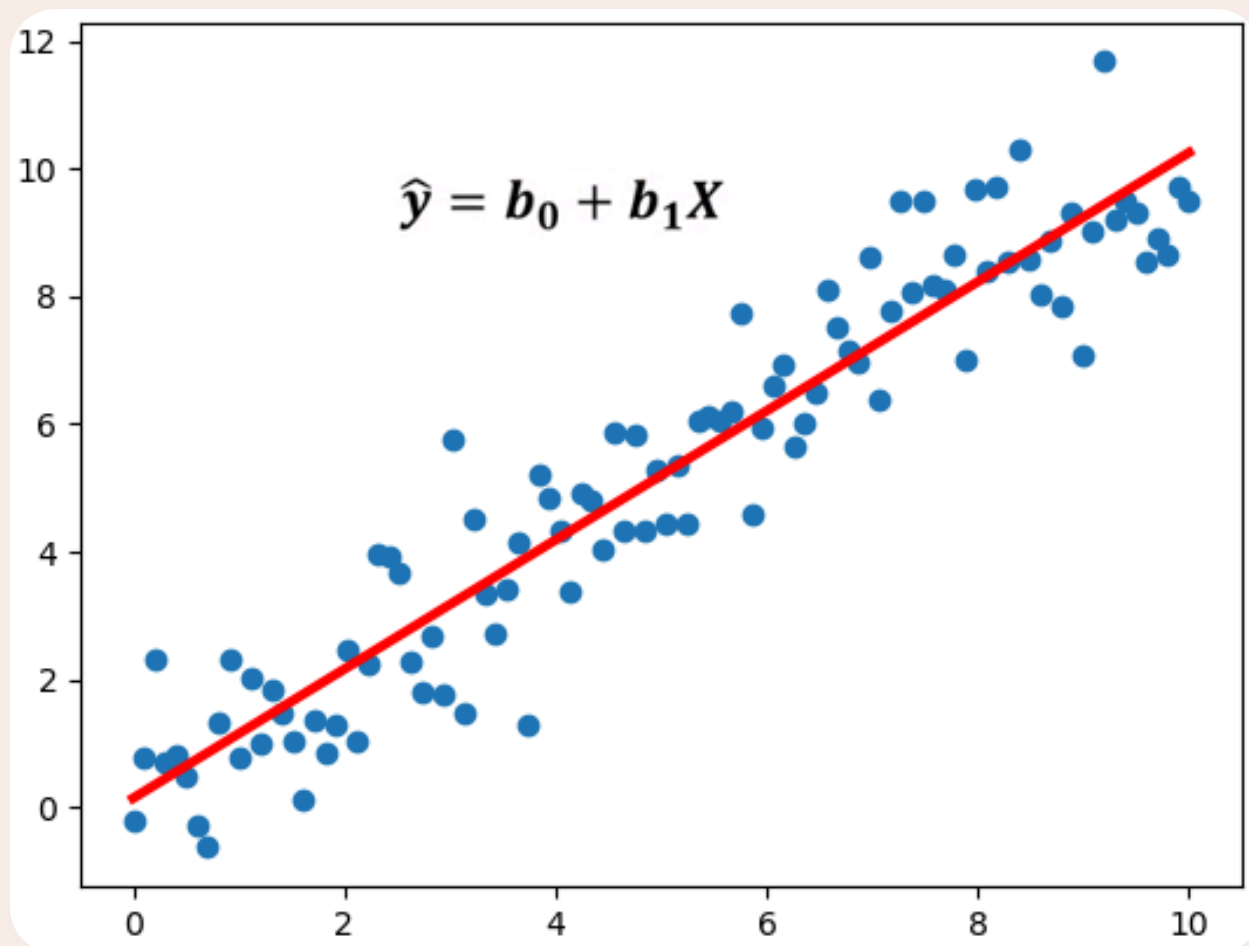
- La regresión busca relaciones entre variables.
- Por ejemplo, puede observar a varios empleados de alguna empresa e intentar comprender cómo sus salarios dependen de sus características.
- Del mismo modo, se puede intentar establecer la dependencia matemática de los precios de la vivienda con respecto a sus características.
- Generalmente, se considera algún fenómeno de interés y se tienen varias observaciones.



Tipos de regresiones



Regresión Lineal Simple



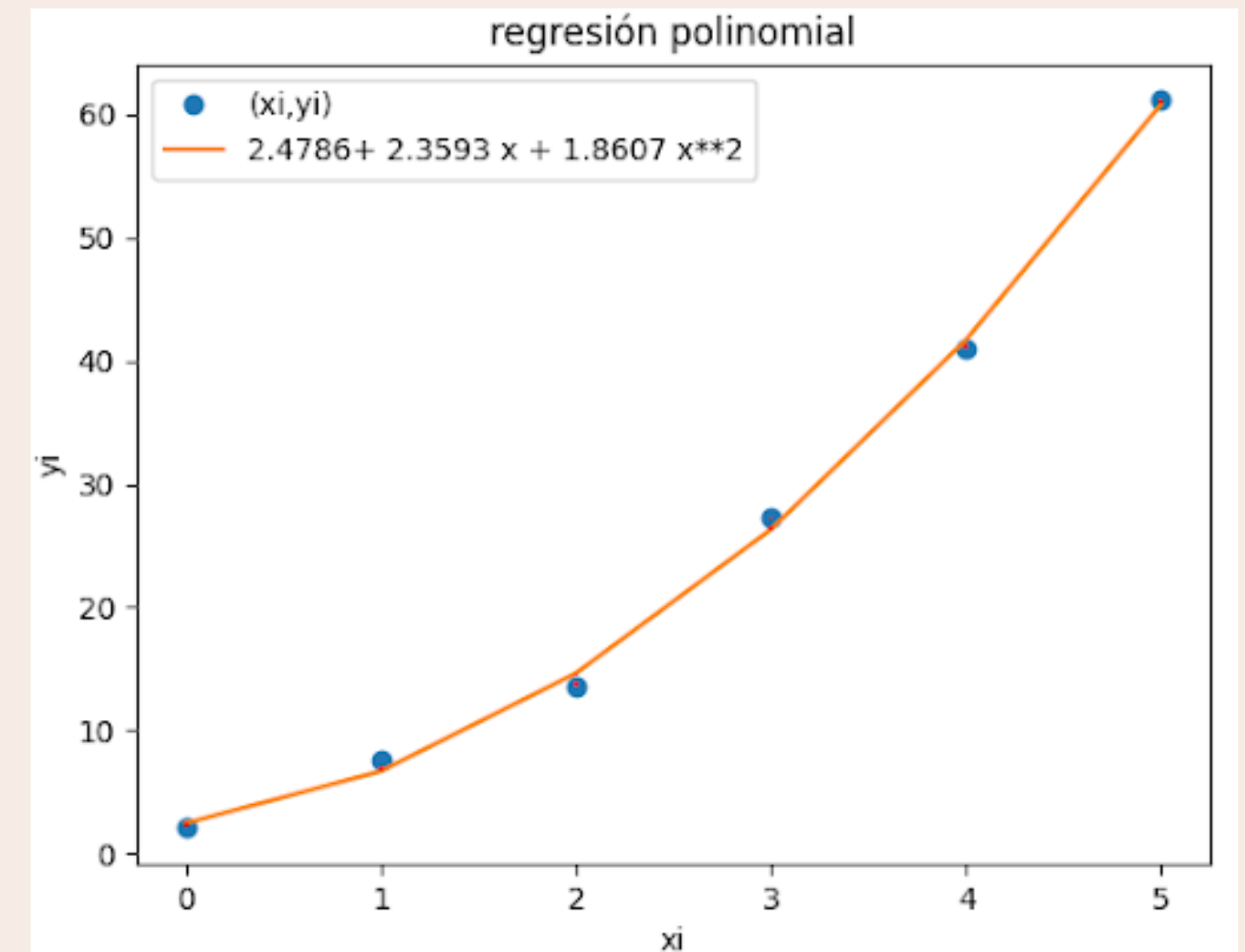
Modelo matemático que describe la relación entre varias variables. Debido a su capacidad para transformar datos, pueden utilizarse para simular una amplia gama de relaciones, y gracias a su forma, sus parámetros estadísticos se analizan y comparan con facilidad, lo que permite que se les extraiga información valiosa.

Regresión Polinomial

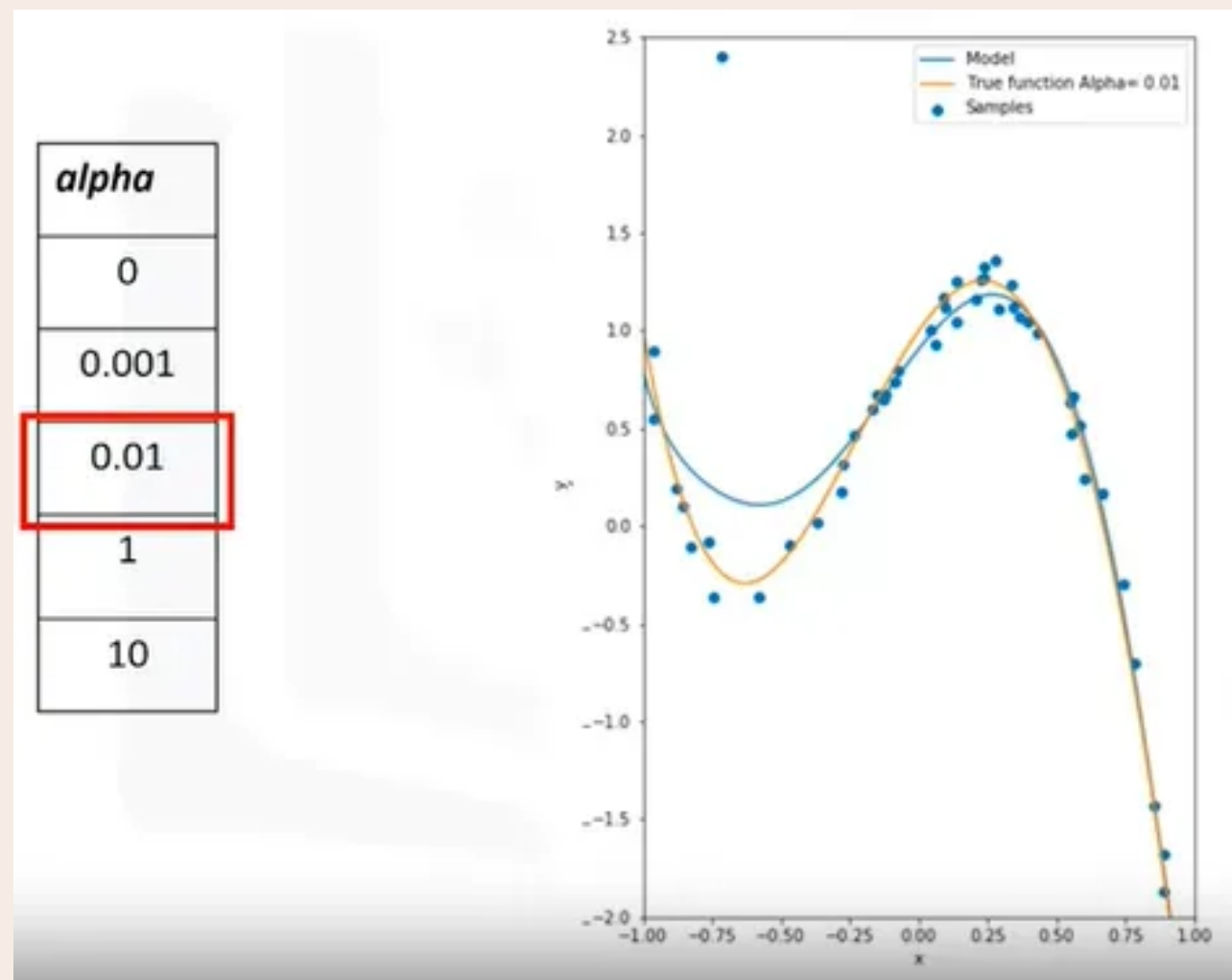
Es una forma sencilla de adaptar un modelo de regresión lineal a estructuras más complejas.

Para ello solo hace falta crear nuevos parámetros elevando al cuadrado, cubo o cualquier otra cifra las variables ya existentes.

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_n x^n + \varepsilon$$

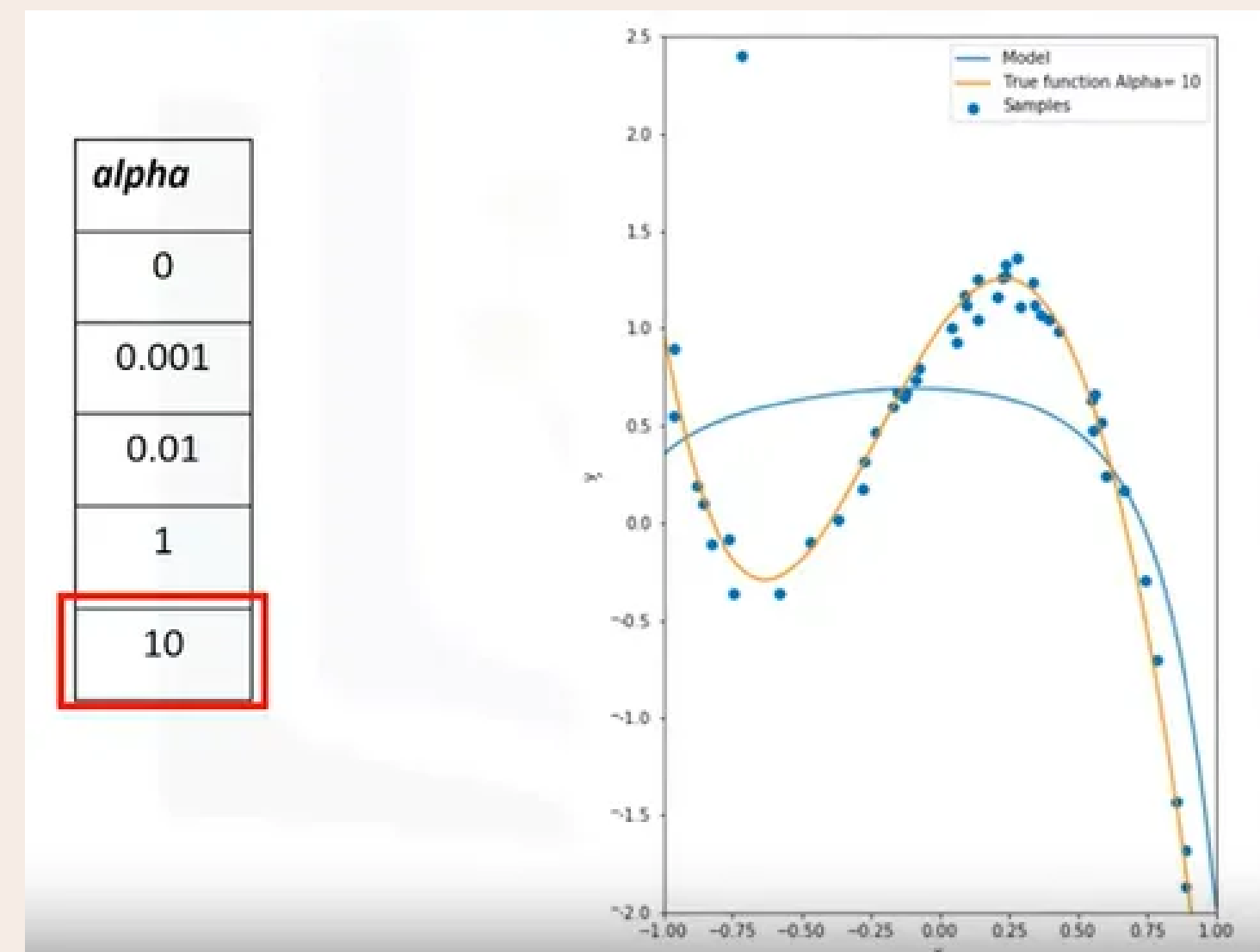
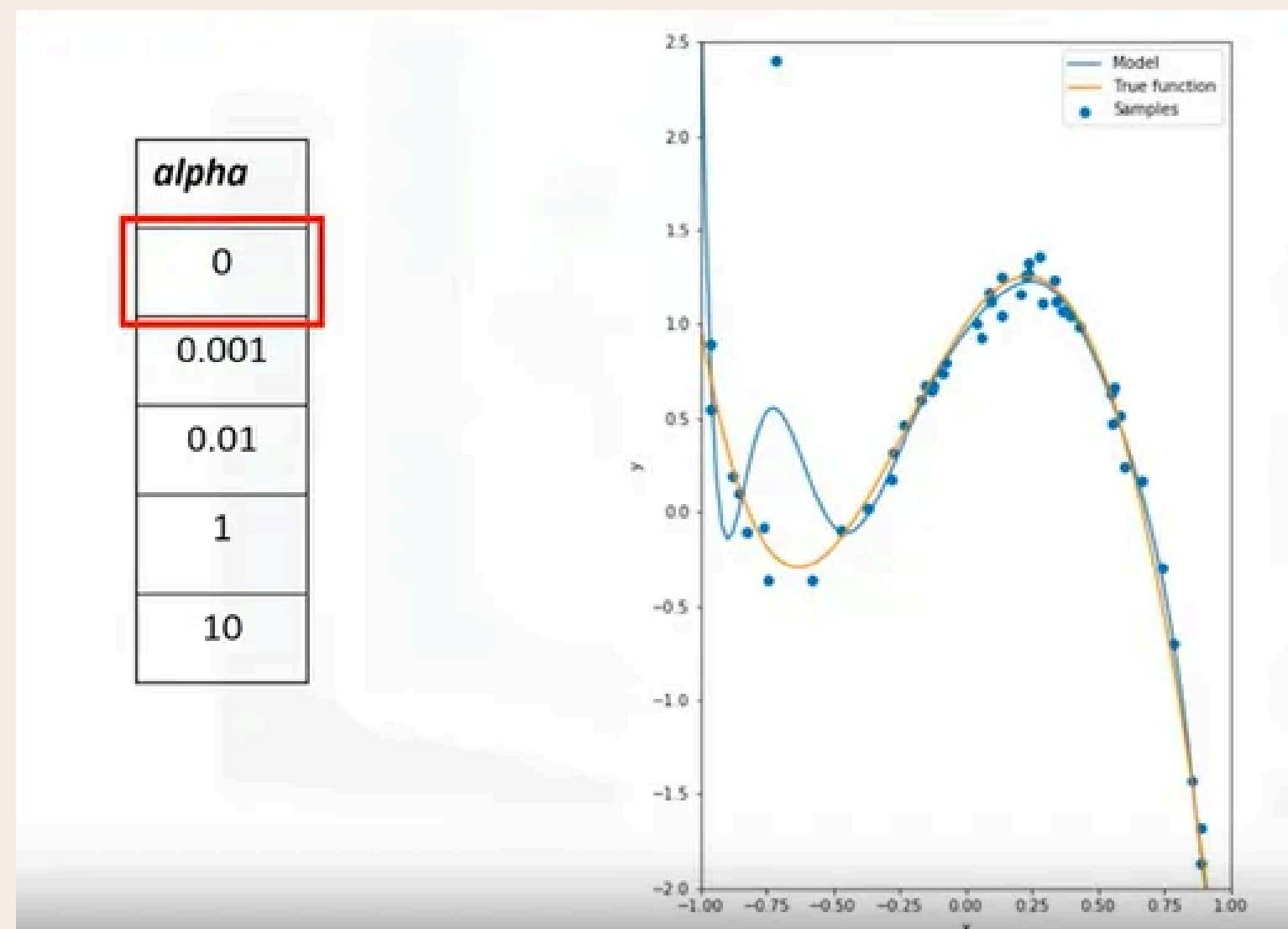


Regresión Polinomial Ridge



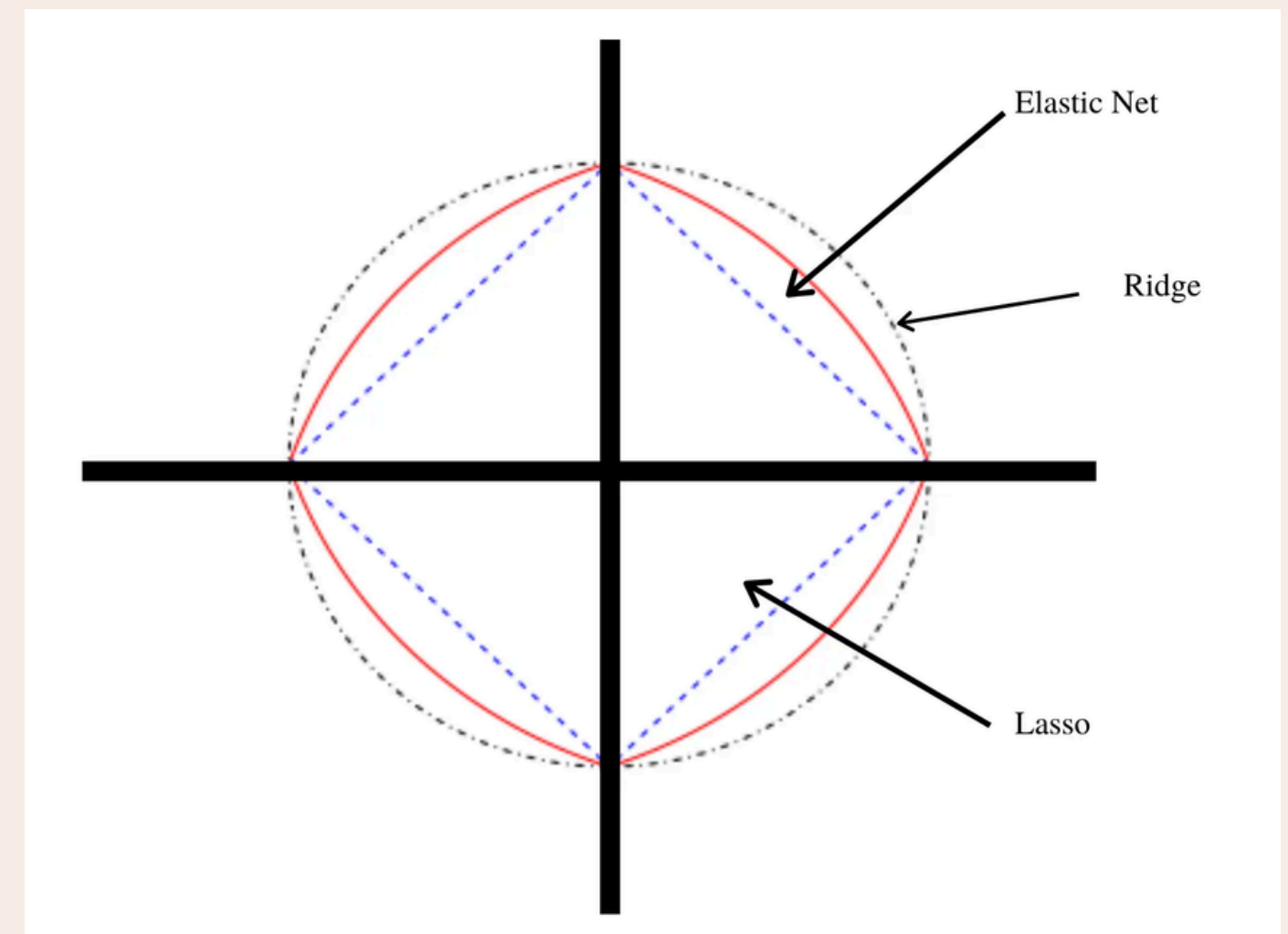
Modifica la función de coste para minimizar la complejidad del modelo. También es conocida como regularización L2, uno de varios tipos de regularización para modelos de regresión lineal. La regularización es un método estadístico para reducir los errores causados por el sobreajuste de los datos de entrenamiento.

Regresión Polinomial Ridge



Regresión Elástica Net

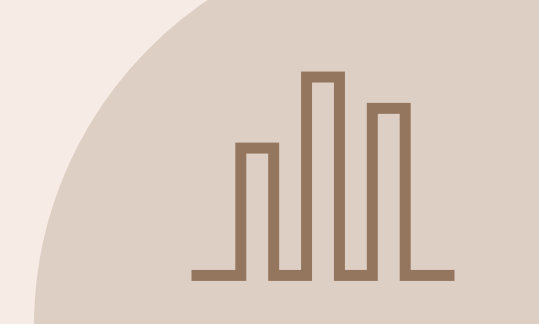
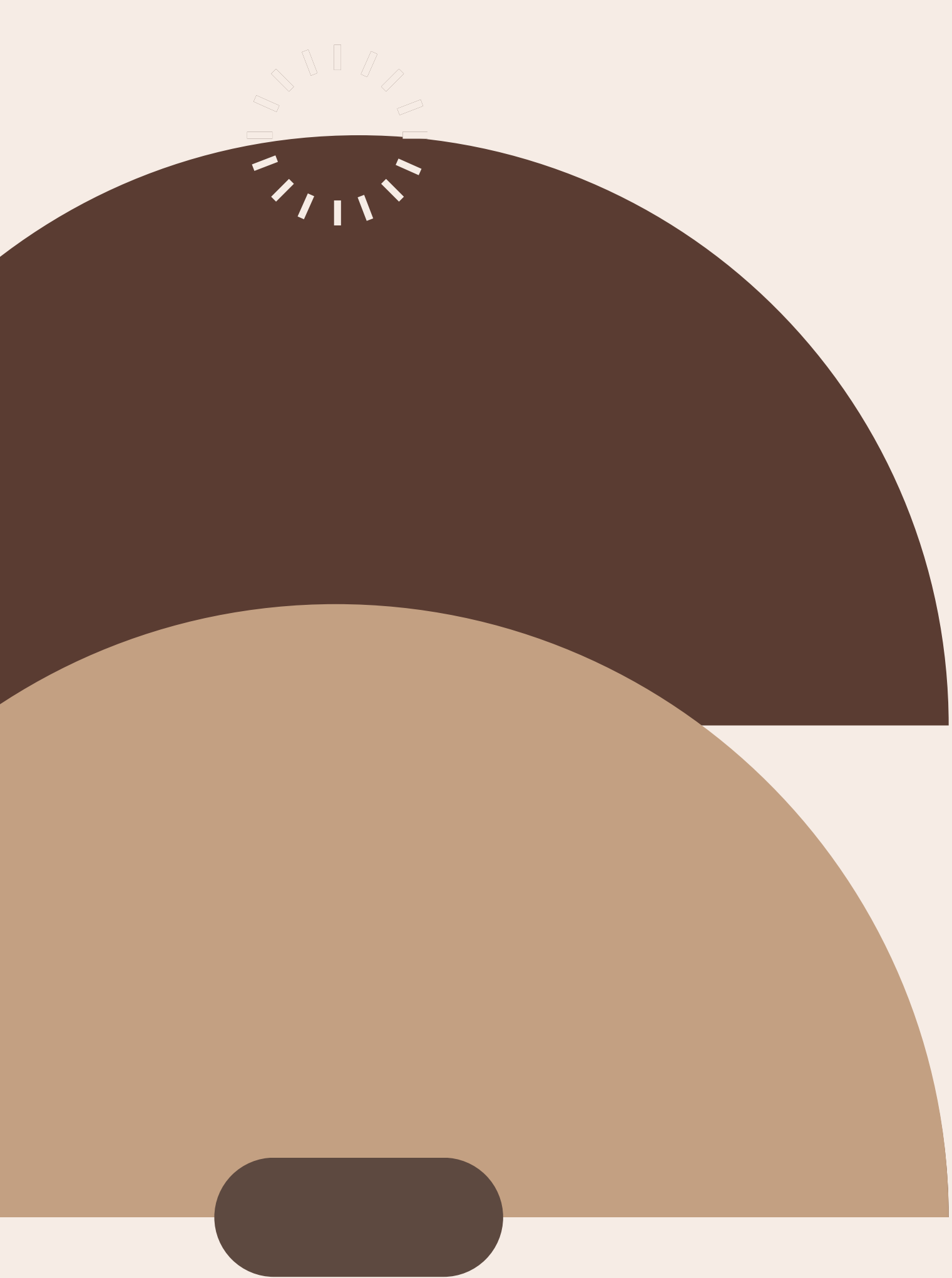
Combina las características de Lasso y Ridge Regression. Es una técnica de regresión regularizada que se utiliza para abordar los problemas de multicolinealidad y sobreajuste, que son comunes en conjuntos de datos de alta dimensión. L1 para selección de características y L2 para la reducción.



Regresión Lasso

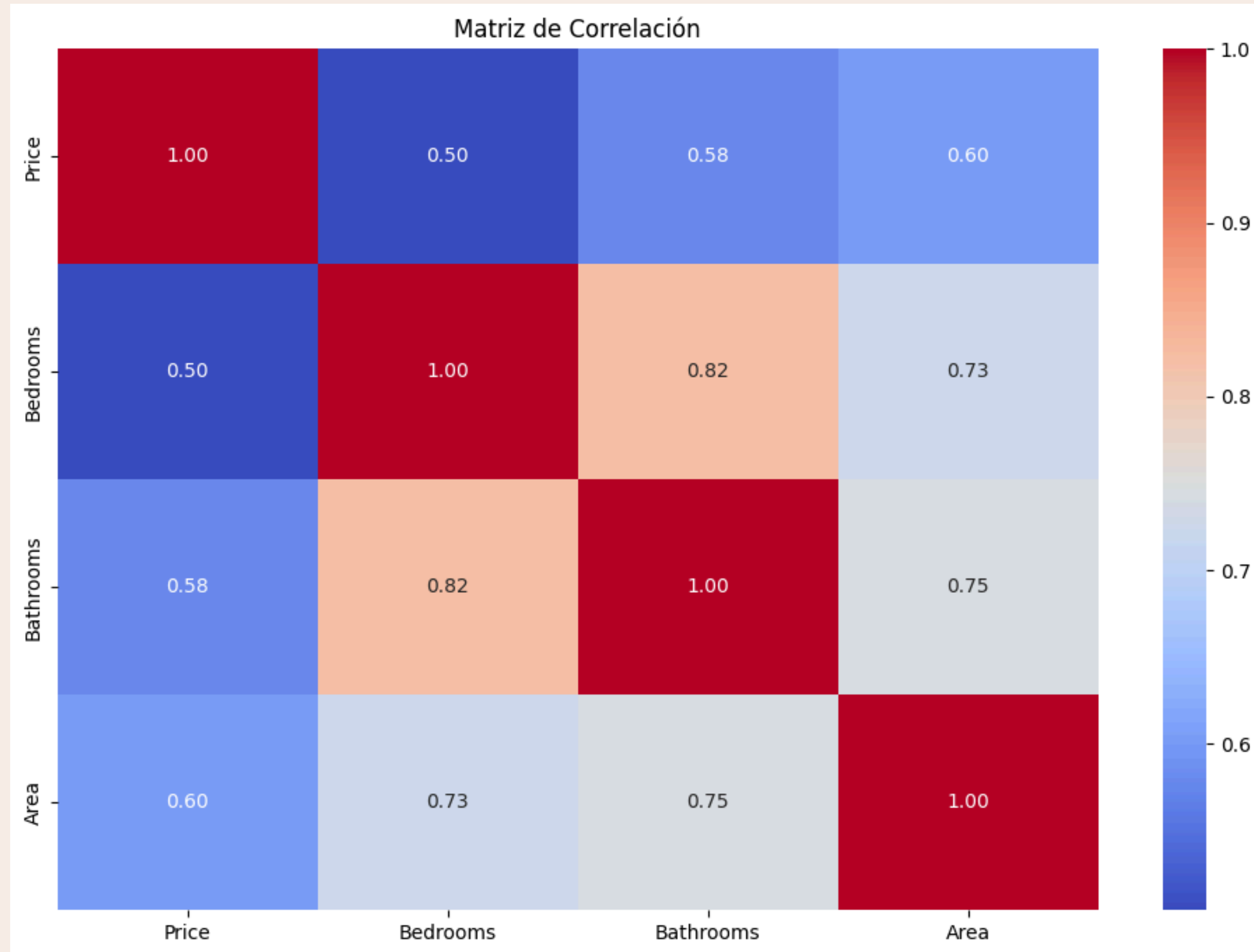


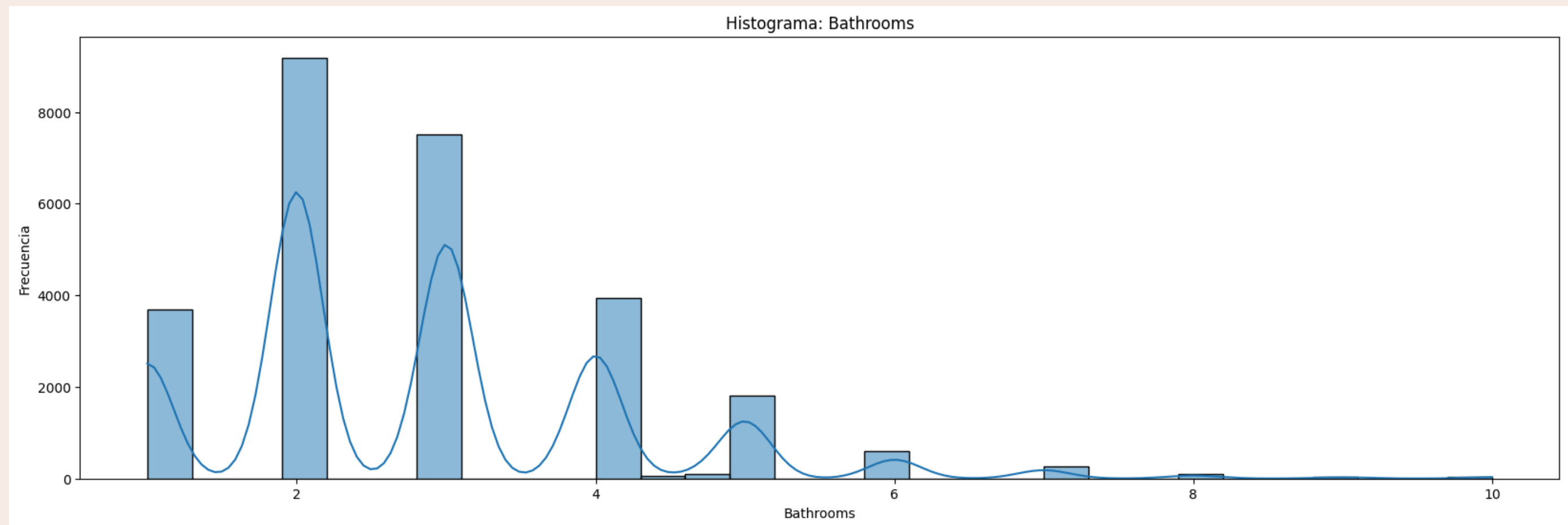
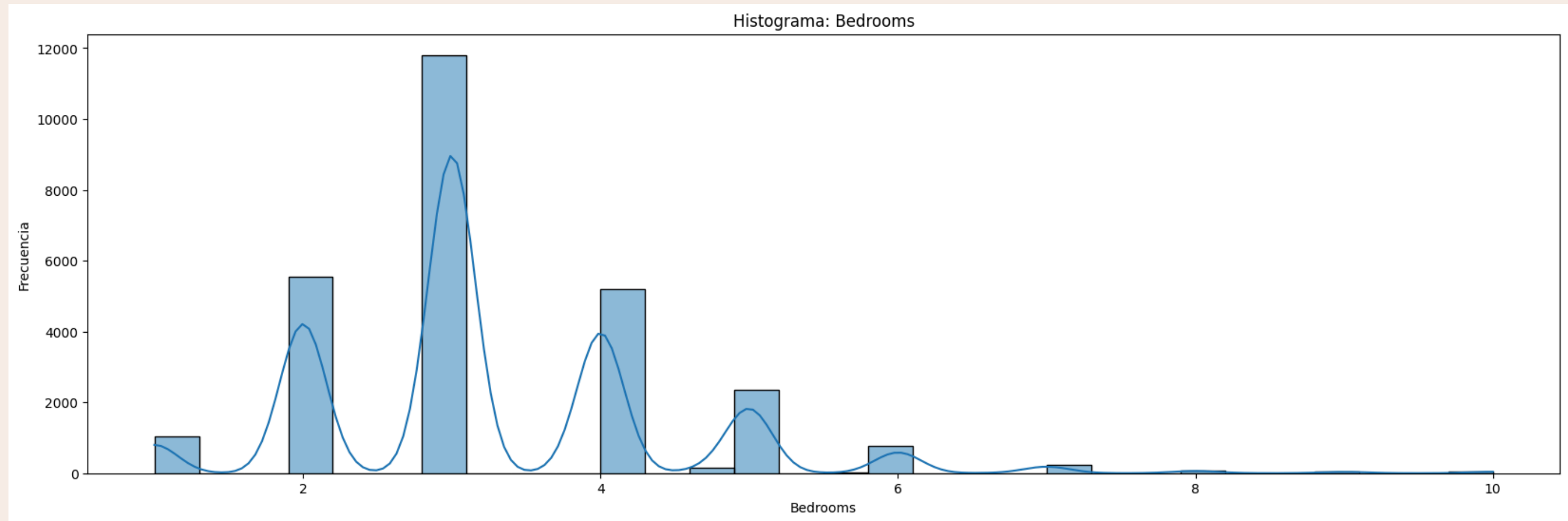
Tiende a generar “coeficientes dispersos”: vectores de coeficientes en los que la mayoría de ellos toman el valor cero. Esto quiere decir que el modelo va a ignorar algunas de las características predictivas, lo que puede ser considerado un tipo de selección automática de características. Al incluir menos variables suponemos un modelo más sencillo de interpretar.

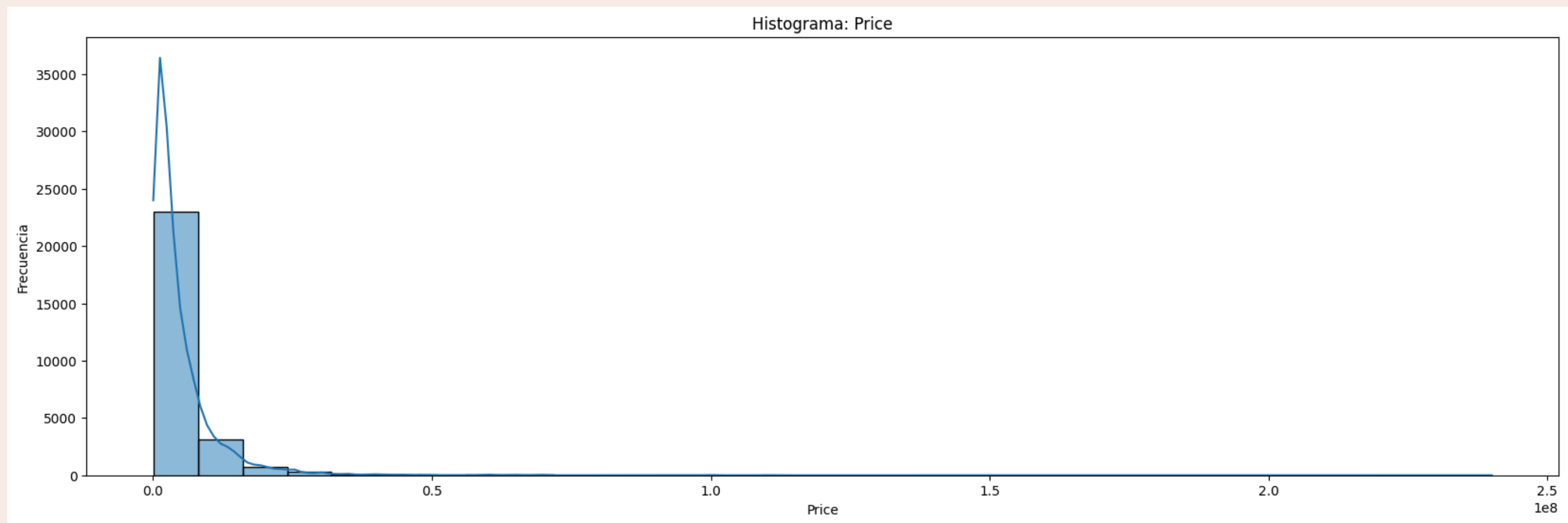
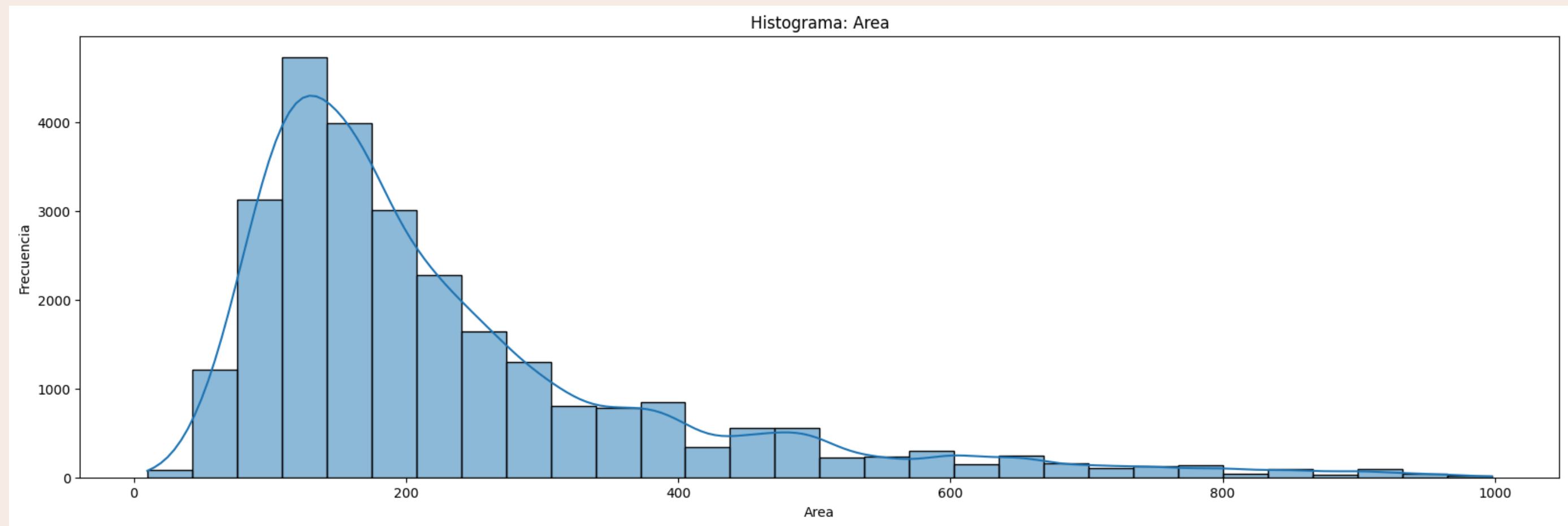


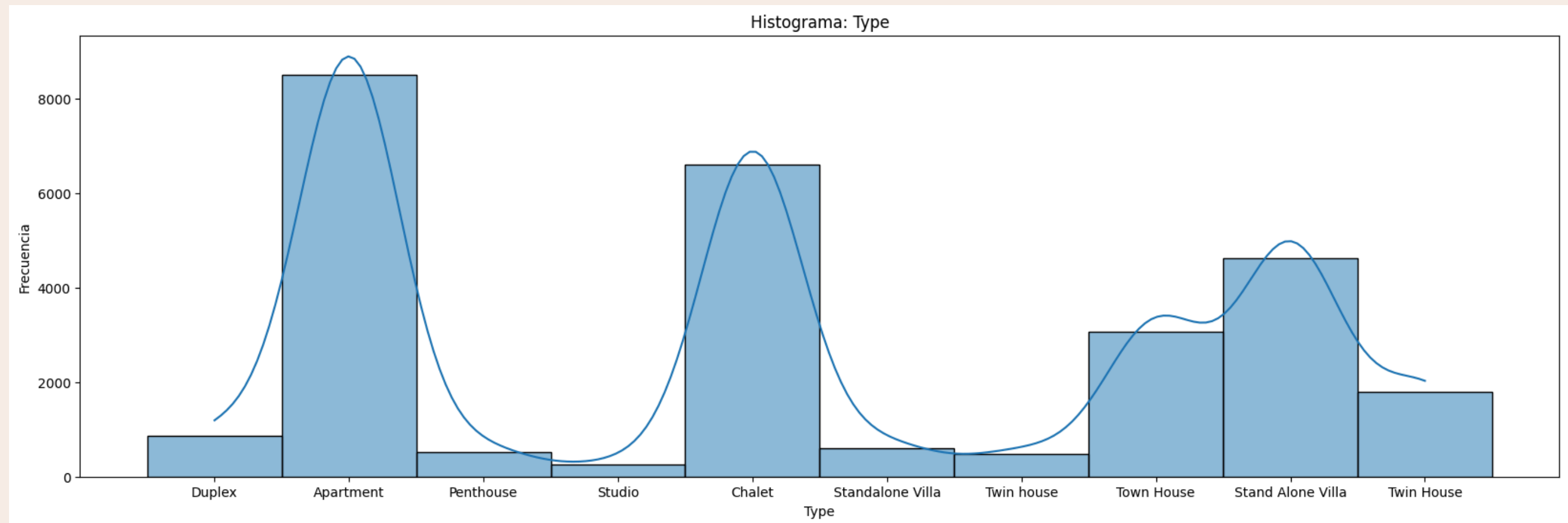
Resultados

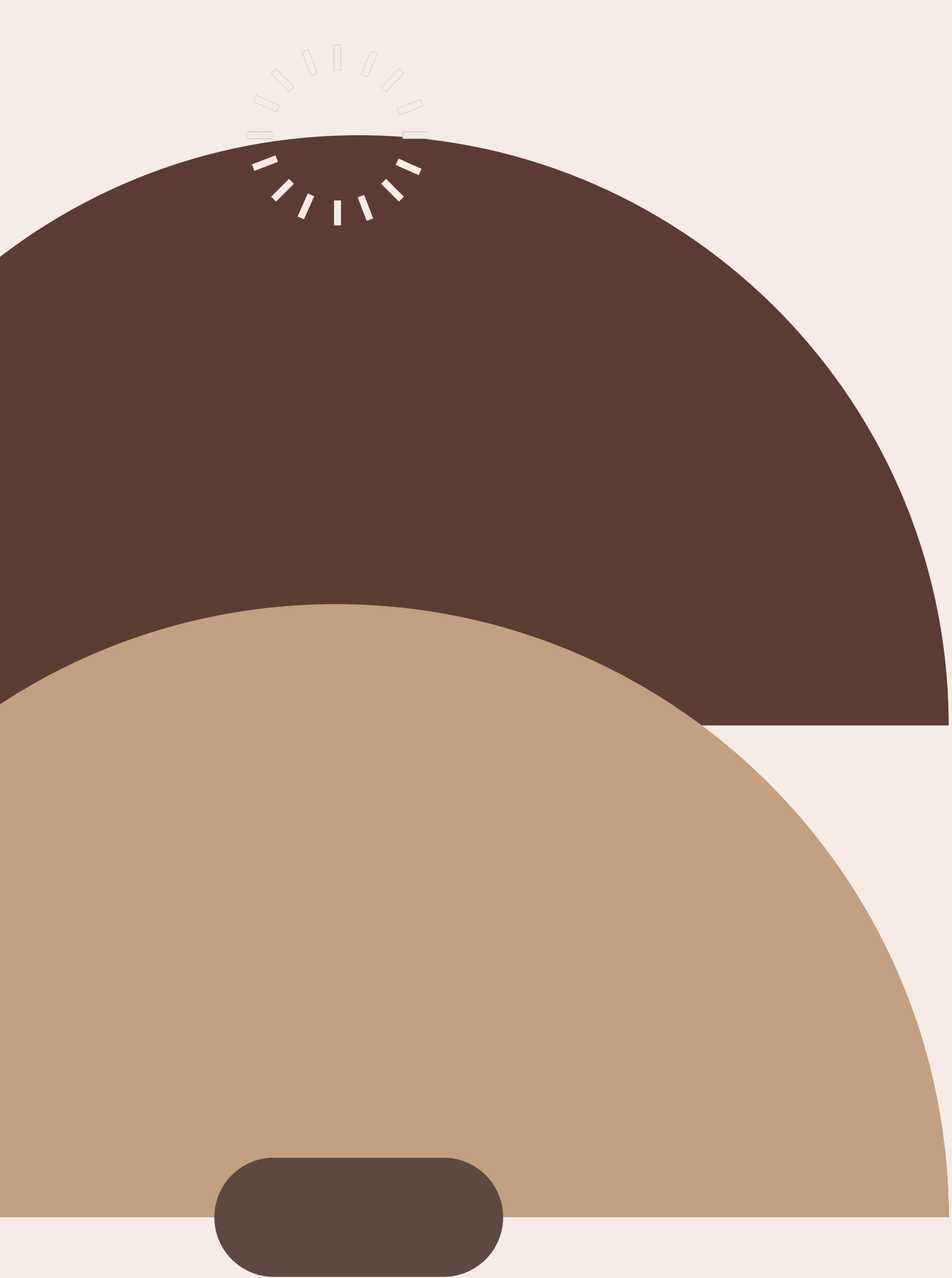












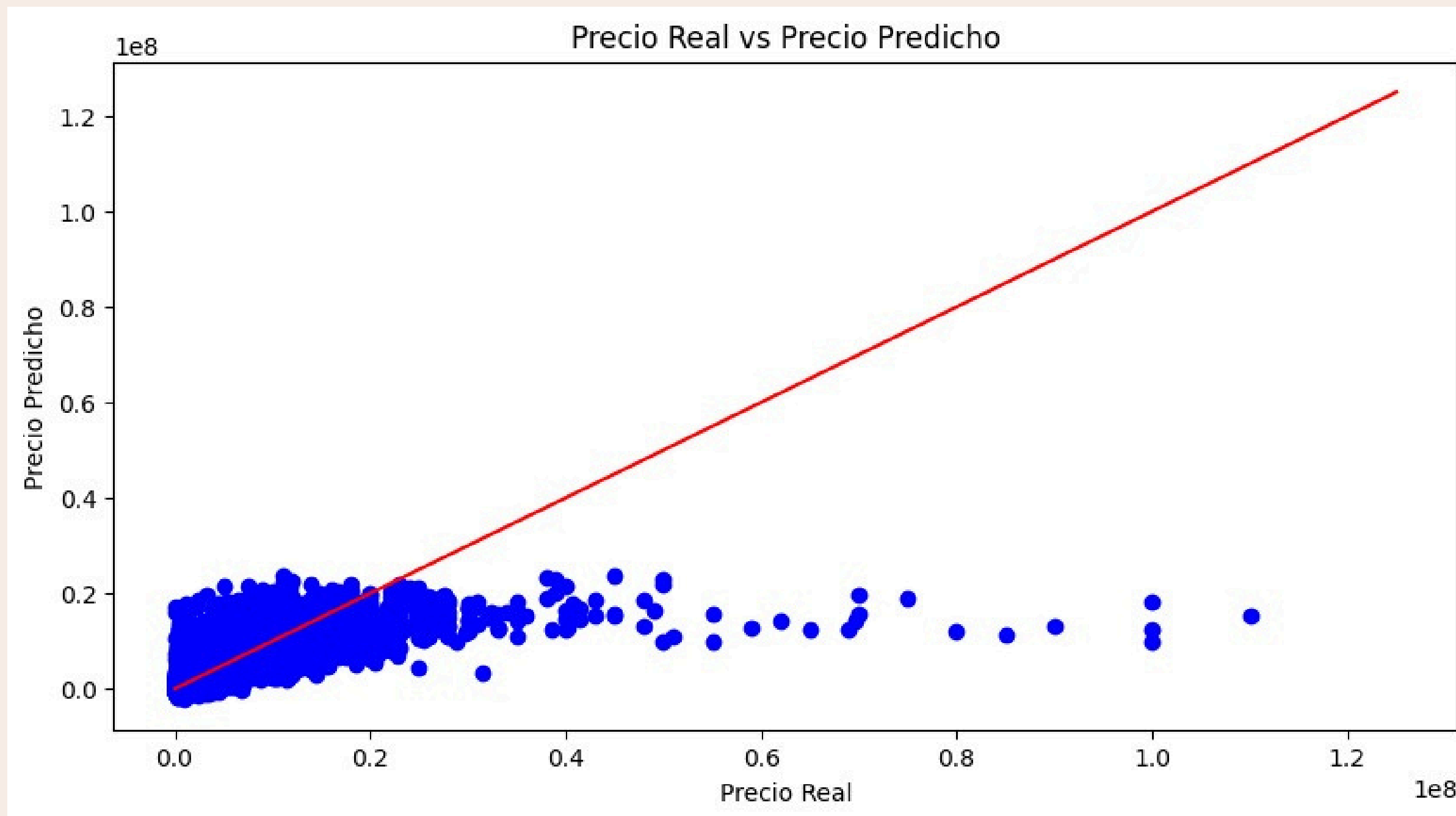
Regresiones

Regresión Lineal

Métrica	Valor
RMSE	4.795.077,77
R^2	0,4521
Error Absoluto Medio	2.624.418,06



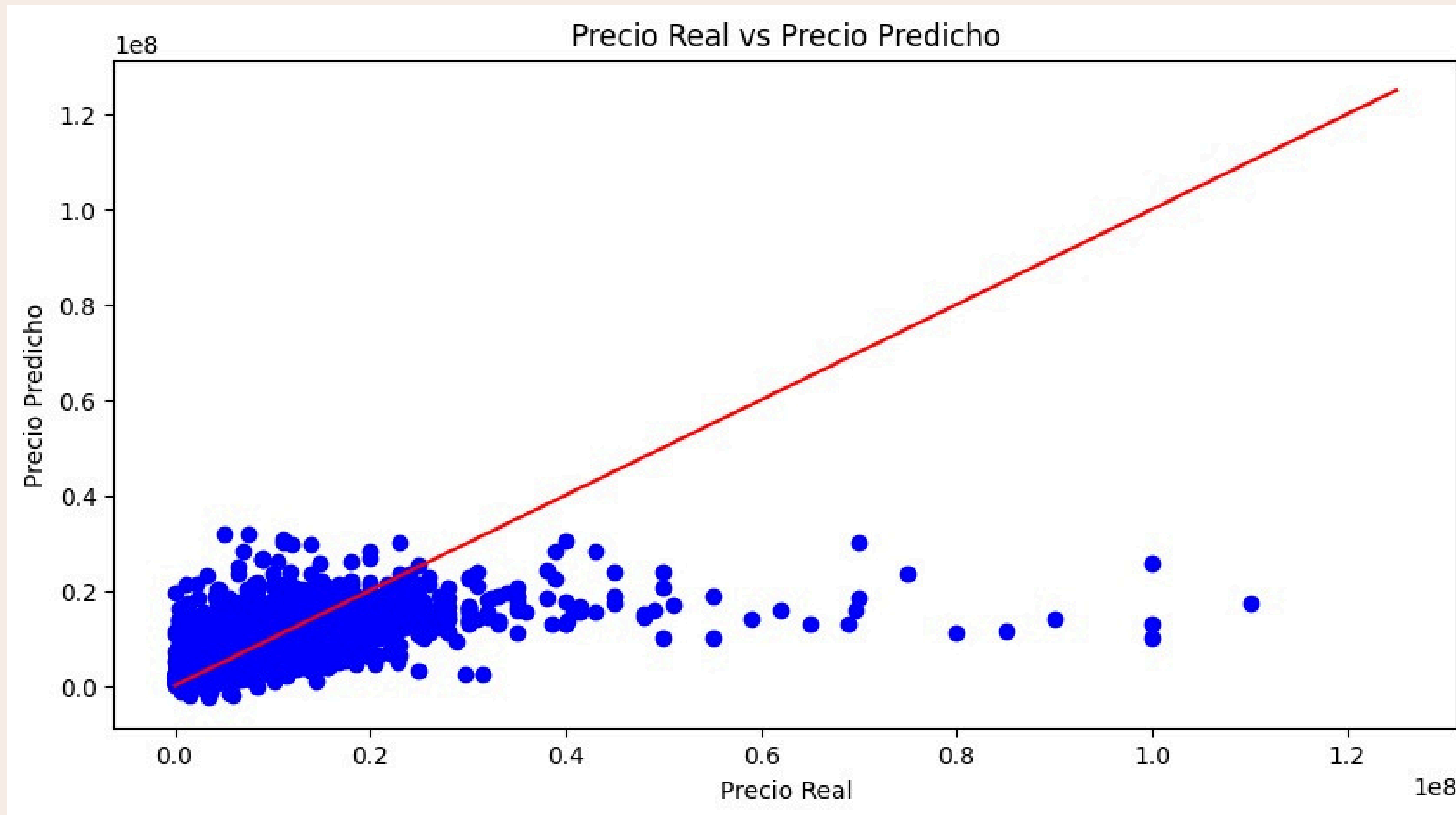
Regresión Lineal



Regresión Polinomial

Métrica	Valor
RMSE	4.672.520,38
R^2	0,4797
Error Absoluto Medio	2.448.893,99

Regresión Polinomial

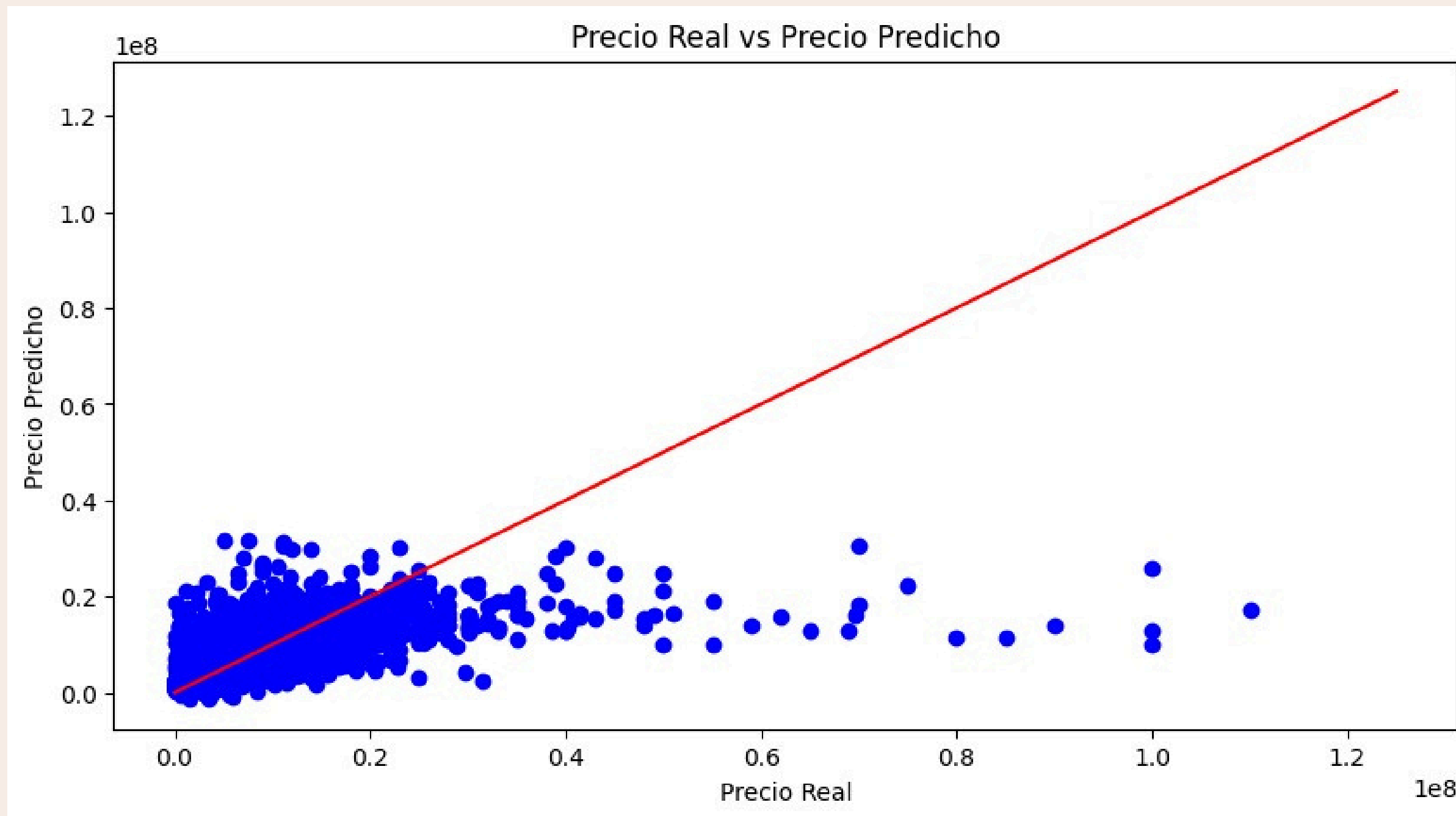


Regresión Ridge

Métrica	Valor
RMSE	4.670.647,66
R^2	0.4801
Error Absoluto Medio	2.447.526,15



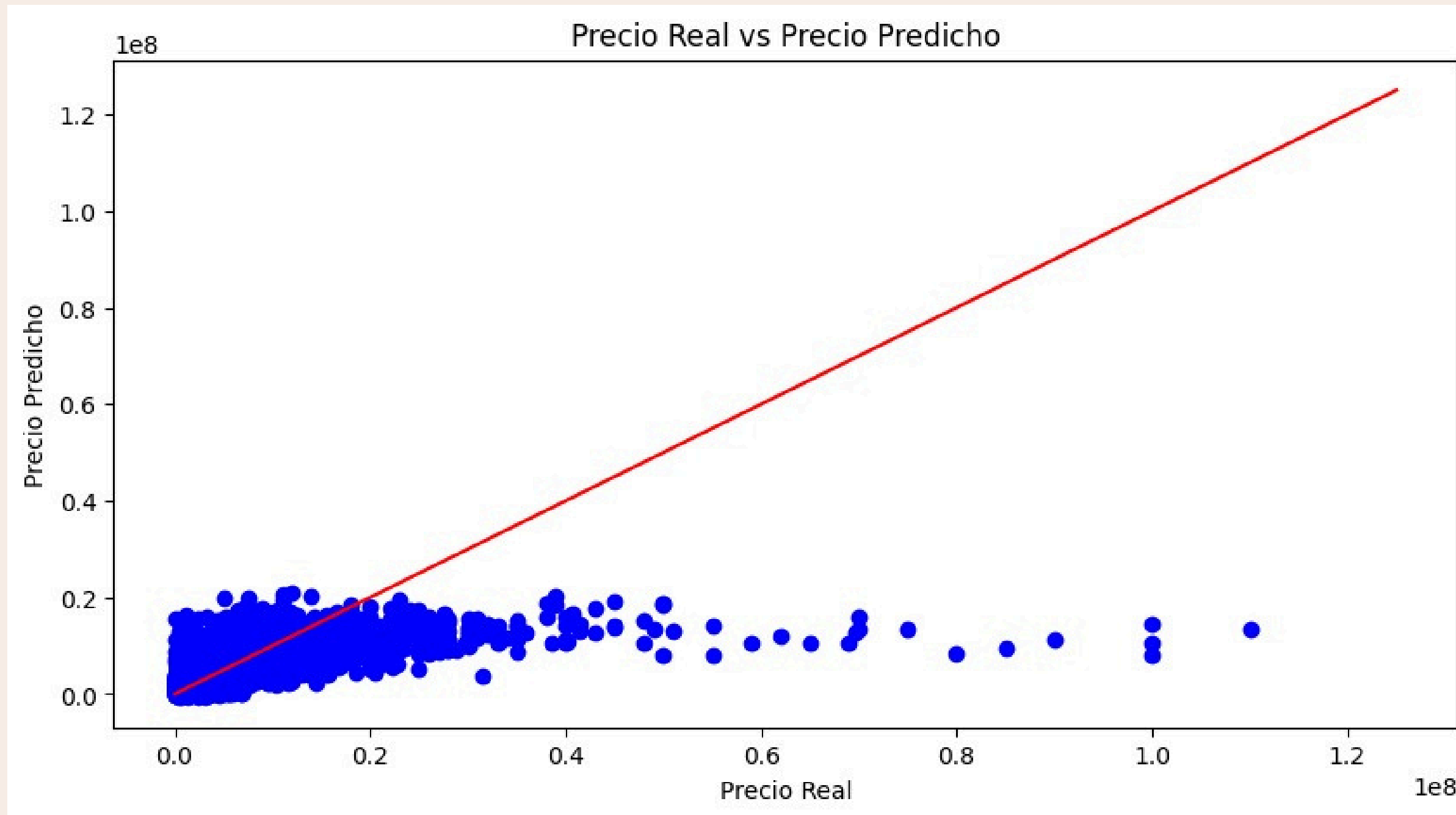
Regresión Ridge



Regresión Elastic Net

Métrica	Valor
RMSE	4.876.446.23
R^2	0,4333
Error Absoluto Medio	2.617.747,64

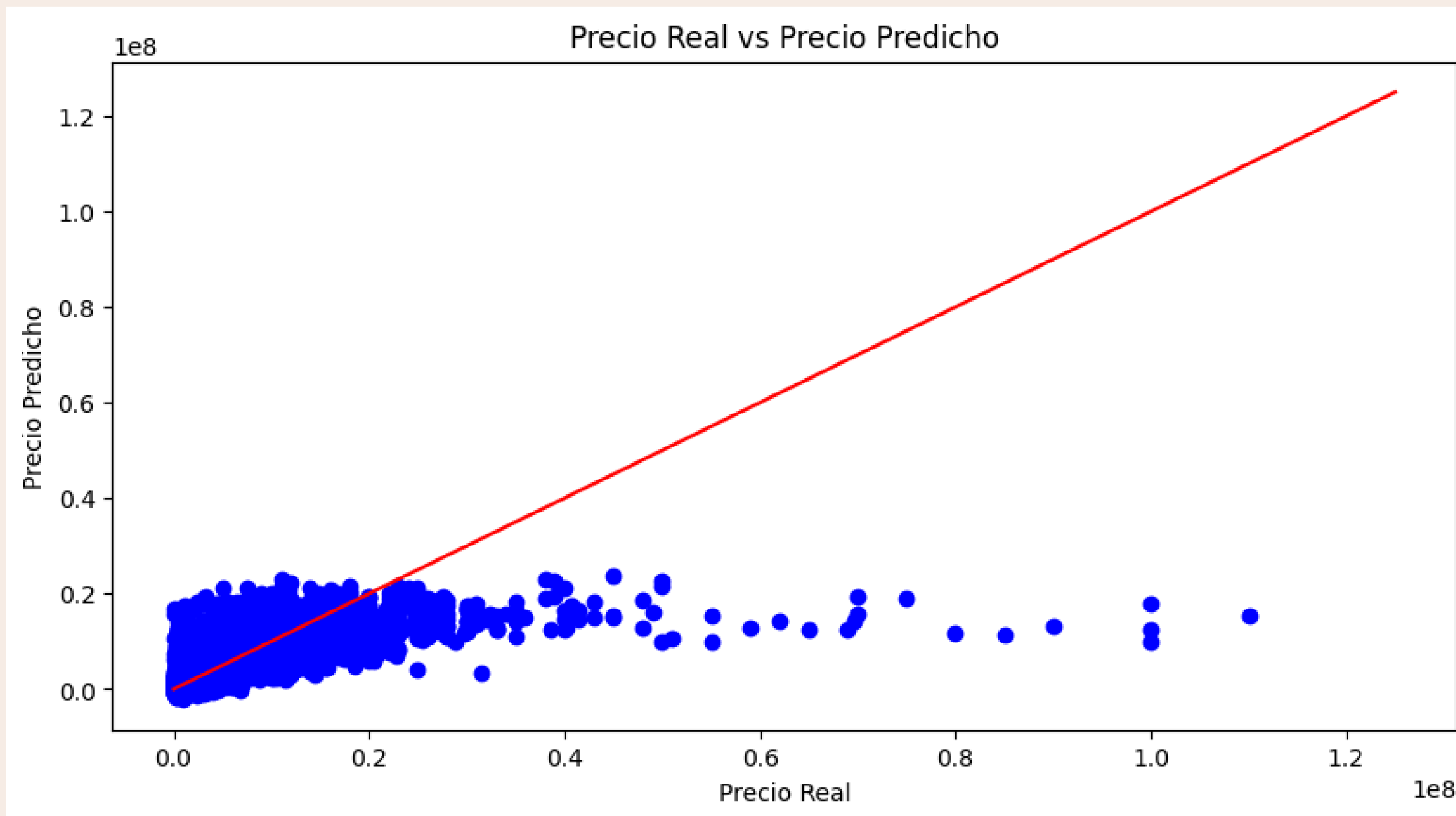
Regresión Elastic Net



Regresión Lasso

Métrica	Valor
RMSE	4.794.518,15
R^2	0,4521
Error Absoluto Medio	2.621.069,87

Regresión Lasso



Dificultades

**Encontrar un modelo que se ajuste
bien a la base**

Tener cuidado con los parámetros

Valores atípicos

Conclusiones y Recomendaciones

Métricas

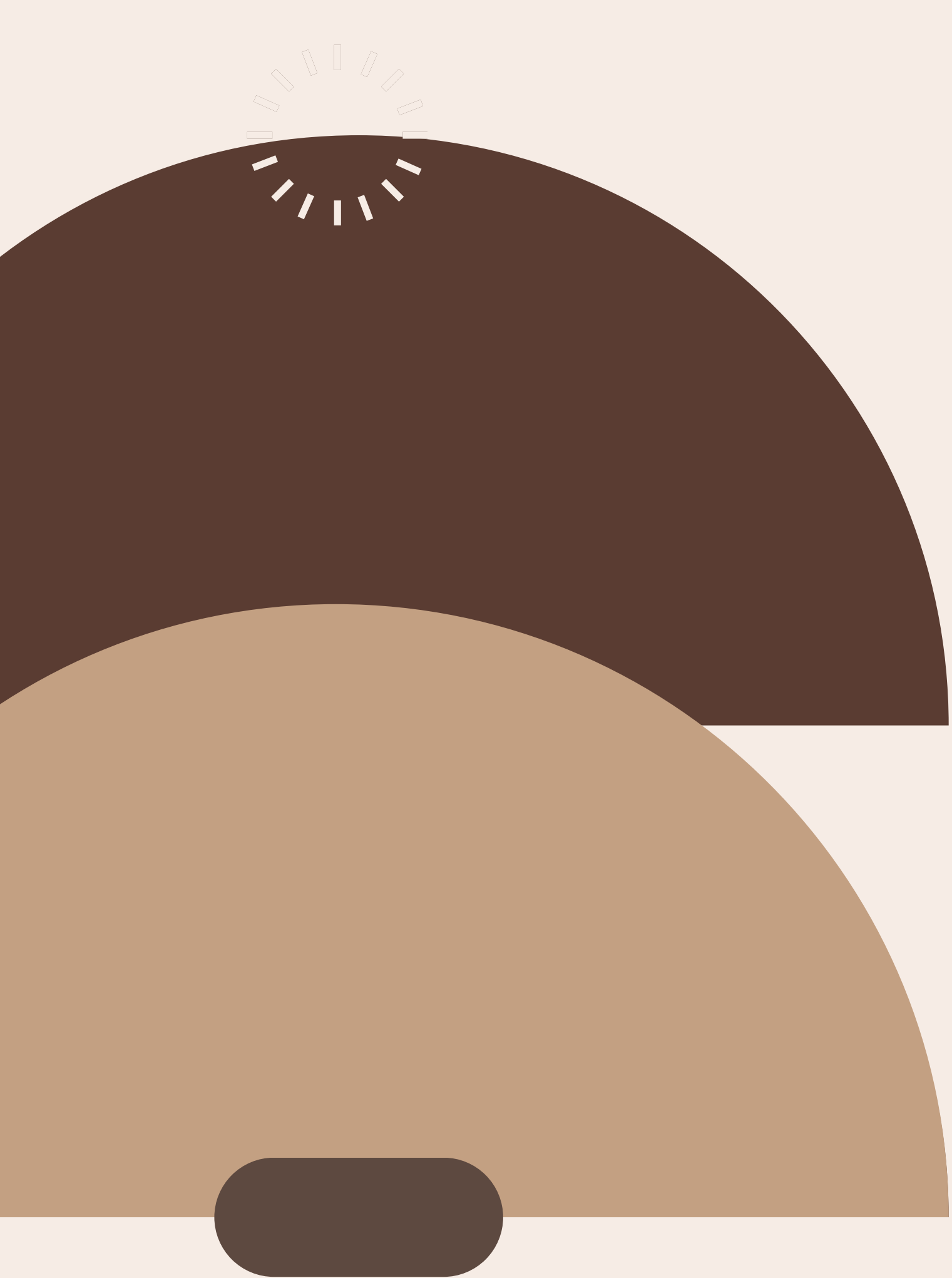
Los valores de R^2 para los modelos explican el 47% de las características. Mientras que los valores de los errores muestran una desviación significativa

Análisis de residuos

Para entender mejor dónde están fallando los modelos actuales y ajustar en consecuencia

Modelos no lineales

Como los modelos lineales y polinomiales no están entregando resultados positivos se recomienda intentar con otros modelos como árboles de decisión



¡Gracias!