

**Instituto Tecnológico y de Estudios
Superiores de Monterrey**



**Tecnológico
de Monterrey**

Analítica de Datos y Herramientas de Inteligencia Artificial II

**Actividad 3.1:
Regresión no lineal**

Alumno:

Diego Sánchez Márquez

A01734778

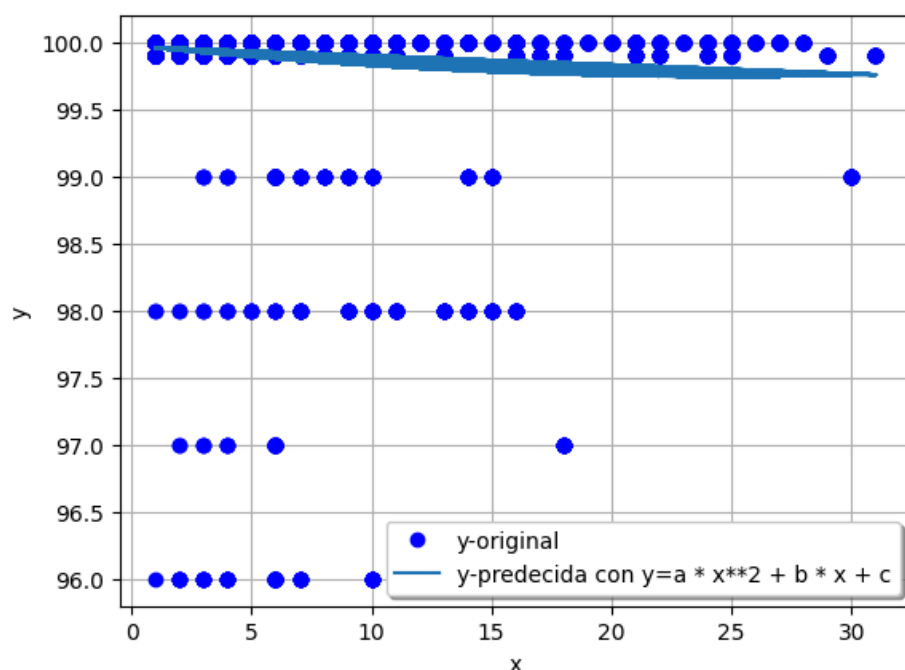
Grupo 501

13, Abril 2025

Para realizar un análisis no lineal más preciso, se eligieron algunas de las variables objetivo más relevantes de la base de datos de Airbnb de la ciudad de Porto y se compararon con las variables independientes que mostraron mayor correlación en los heatmap realizados para cada variable objetivo. Con esto, se aplicaron diferentes funciones no lineales como cuadrática, exponencial, logarítmica, tangencial, senoidal y valor absoluto, midiendo el coeficiente de determinación y el coeficiente de correlación para determinar su capacidad de predecir la variable objetivo de cada caso. El objetivo principal fue identificar qué tipo de modelo se ajusta mejor a los datos.

Para la variable objetivo `host_response_rate`, se utilizaron dos funciones con la variable independiente `calculated_host_listings_count`.

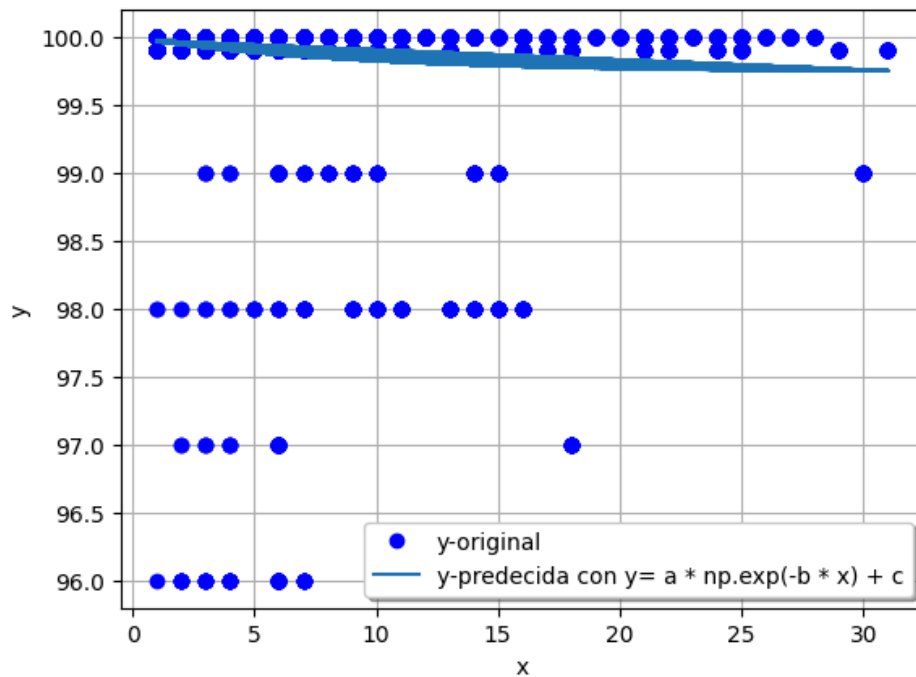
Modelo 1



- Función Cuadrática: $y = ax^2 + bx + c$
- Coeficiente de determinación: 0.017
- Coeficiente de correlación: 0.131

Este modelo muestra una precisión muy baja y no logra explicar la relación entre las variables.

Modelo 2

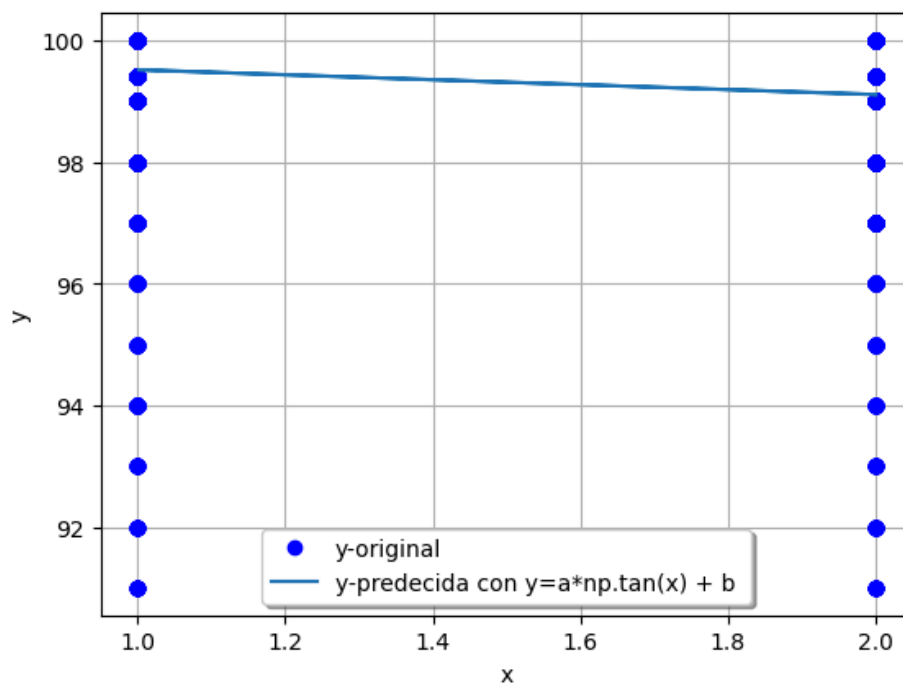


- Función Exponencial: $y = a \cdot \exp(-bx) + c$
- Coeficiente de determinación: 0.017
- Coeficiente de correlación: 0.133

El modelo exponencial tiene resultados prácticamente idénticos respecto al cuadrático, lo que indica una mala predicción de la variable objetivo.

Para la variable objetivo `host_acceptance_rate`, se utilizó la variable `instant_bookable`.

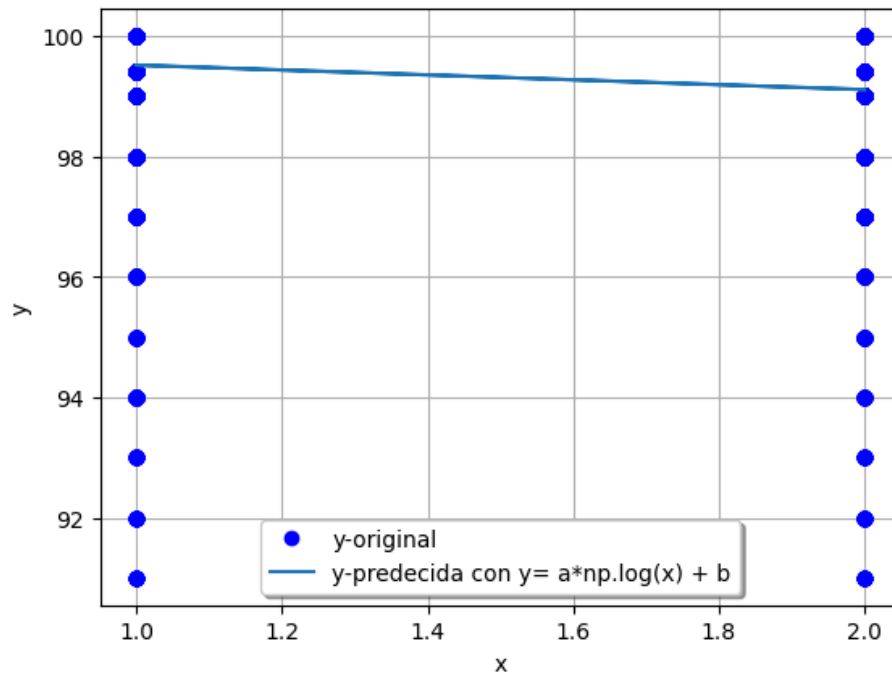
Modelo 1



- Función Tangencial: $y = a \cdot \tan(x) + b$
- Coeficiente de determinación: 0.021
- Coeficiente de correlación: 0.146

El modelo es poco efectivo, con una relación baja y poco explicativa.

Modelo 2

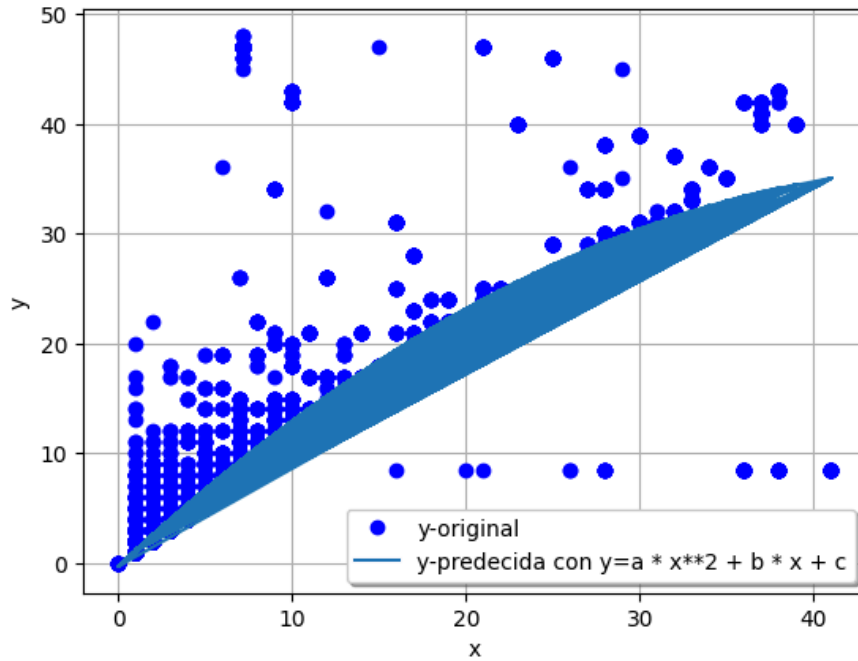


- Función Logarítmica: $y = a \cdot \log(x) + b$
- Coeficiente de determinación: 0.021
- Coeficiente de correlación: 0.145

La función logarítmica presenta resultados similares, con baja capacidad para explicar la variable objetivo.

Para la variable objetivo `host_total_listings_count`, se utilizó `host_listings_count` como variable independiente.

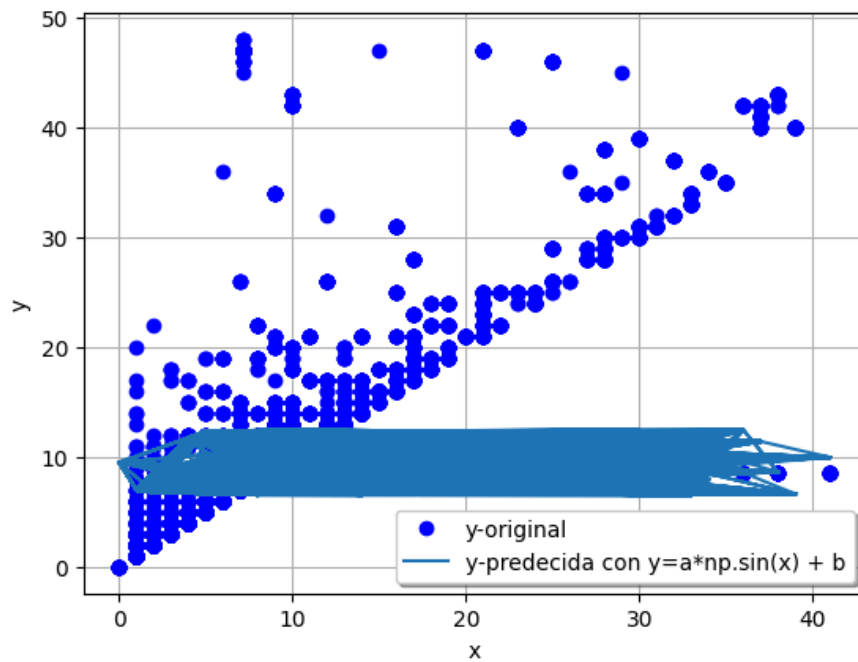
Modelo 1



- Función Cuadrática: $y = ax^2 + bx + c$
- Coeficiente de determinación: 0.716
- Coeficiente de correlación: 0.846

Este modelo tiene un rendimiento muy alto, explicando bien la variabilidad de la variable objetivo.

Modelo 2



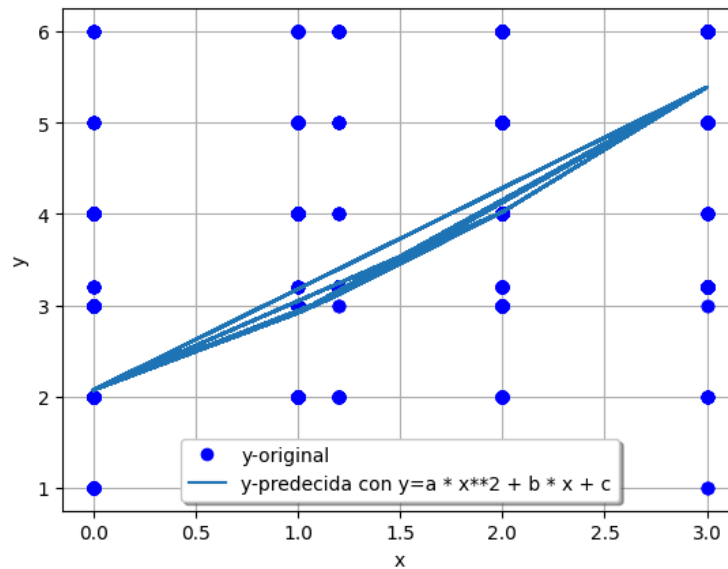
- Función Senoidal: $y = a \cdot \sin(x) + b$
- Coeficiente de determinación: 0.045

- Coeficiente de correlación: 0.213

La función senoidal no se adapta bien a la hora de realizar la predicción, presentando baja precisión.

Para la variable objetivo accommodates, se utilizó bedrooms como variable independiente.

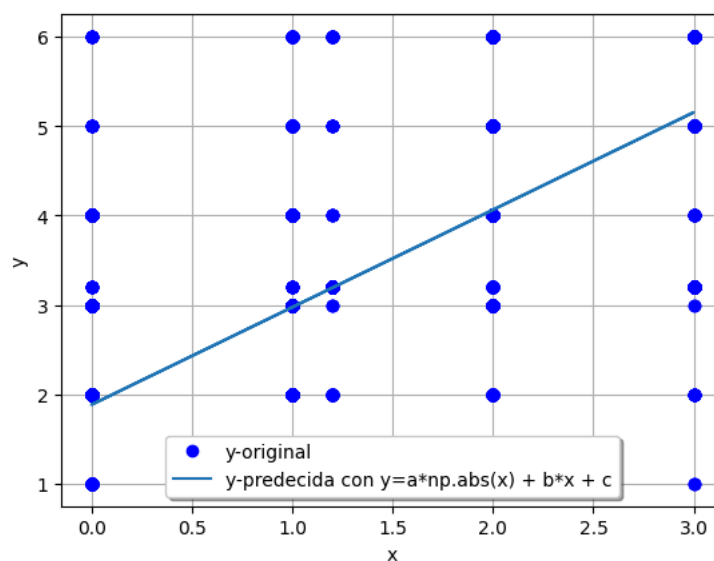
Modelo 1



- Función Cuadrática: $y = ax^2 + bx + c$
- Coeficiente de determinación: 0.401
- Coeficiente de correlación: 0.633

El modelo cuadrático representa moderadamente bien la relación.

Modelo 2

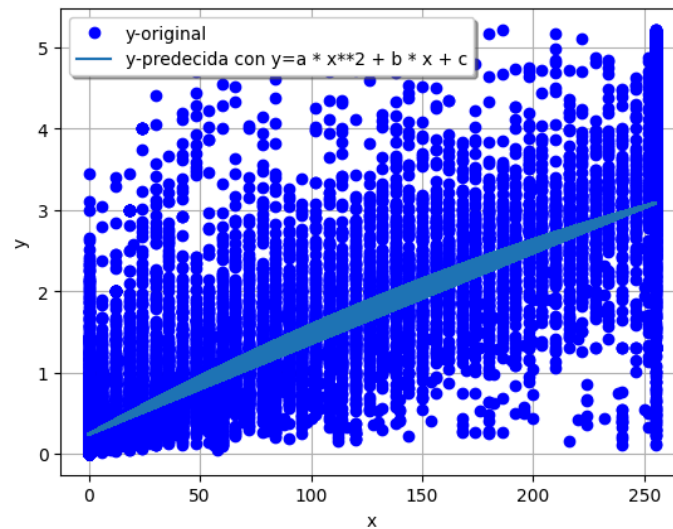


- Función Valor Absoluto: $y = a \cdot |x| + bx + c$
- Coeficiente de determinación: 0.394
- Coeficiente de correlación: 0.628

Este modelo tiene un desempeño similar al cuadrático, con una capacidad de predicción moderada.

Para la variable objetivo `reviews_per_month`, se utilizó `estimated_occupancy_l365d`.

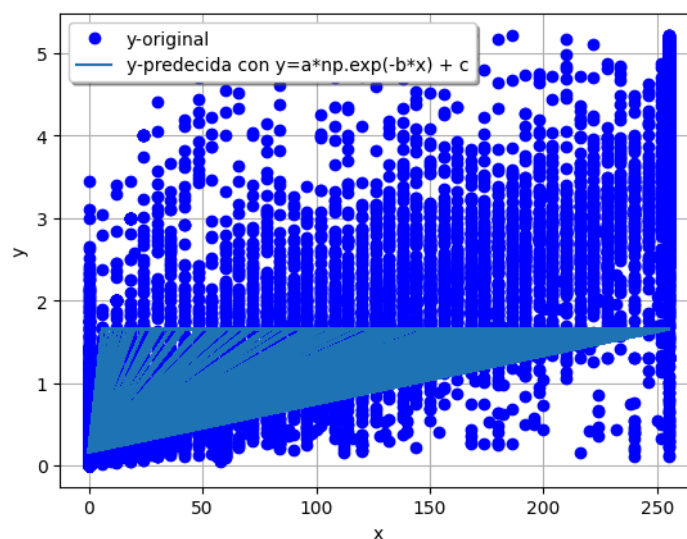
Modelo 1



- Función Cuadrática: $y = ax^2 + bx + c$
- Coeficiente de determinación: 0.669
- Coeficiente de correlación: 0.818

Este modelo se ajusta bien a las variables y muestra una relación fuerte.

Modelo 2

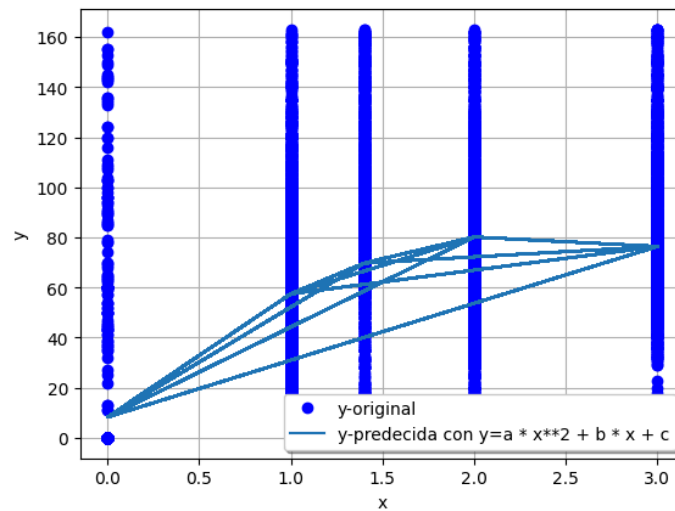


- Función Exponencial: $y = a \cdot \exp(-bx) + c$
- Coeficiente de determinación: 0.258
- Coeficiente de correlación: 0.508

El modelo exponencial tiene una relación moderada, pero menor que el cuadrático.

Para la variable objetivo price, se utilizó beds como variable independiente.

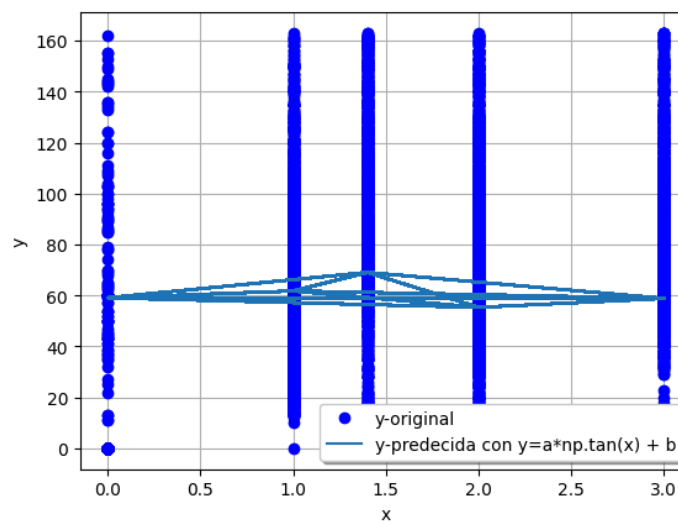
Modelo 1



- Función Cuadrática: $y = ax^2 + bx + c$
- Coeficiente de determinación: 0.406
- Coeficiente de correlación: 0.637

Se observa una relación moderada con buen poder explicativo.

Modelo 2



- Función Tangencial: $y = a \cdot \tan(x) + b$
- Coeficiente de determinación: 0.013

- Coeficiente de correlación: 0.114

Este modelo muestra un rendimiento bajo y no representa bien la relación entre las variables.

Conclusión

En general, los modelos con funciones cuadráticas demostraron ser los más efectivos para explicar la relación entre las variables objetivo e independientes, especialmente para las variables `host_total_listings_count` y `reviews_per_month`, con coeficientes de determinación superiores a 0.66 y correlaciones mayores a 0.81. Las funciones exponenciales, tangenciales y senoidales mostraron un rendimiento mucho más bajo.