

# Actividad 2: Regresión no lineal

Tecnológico de Monterrey, Campus Puebla

Gestión de Proyectos de Plataformas Tecnológicas

Karyme Pérez Chatú A01174367

---

## Objetivo:

Analizar los hallazgos obtenidos por medio de los modelos de regresión no lineal aplicados a la base de datos de Airbnb México.

## Índice:

1. Repositorio
2. Introducción
3. Análisis: Correlación de modelos
  - 2.1. HOST RESPONSE RATE
  - 2.2. HOST ACCEPTANCE RATE
  - 2.3. HOST LISTINGS COUNT
  - 2.4. ACCOMMODATES
  - 2.5. REVIEWS PER MONTH
  - 2.6. PRICE
4. Conclusión

## 1. Repositorio

[https://github.com/KarymeCh/Actividad2\\_RegresionLineal.git](https://github.com/KarymeCh/Actividad2_RegresionLineal.git)

---

## 2. Introducción

Los modelos no lineales son técnicas estadísticas utilizadas para analizar la relación entre una variable dependiente y una o más variables independientes cuando dicha relación no sigue un patrón lineal. A diferencia de la regresión lineal, en los modelos no lineales la relación entre las variables puede representarse mediante funciones exponenciales, logarítmicas, polinomiales u otras formas que capturan comportamientos curvos o complejos.

Estas herramientas se aplican ampliamente en áreas como la biología, la economía, la ingeniería y las ciencias sociales, permitiendo realizar predicciones más precisas, evaluar el impacto de múltiples factores y tomar decisiones basadas en datos. En este caso, se aplican para analizar el comportamiento y la relación de las variables cuantitativas presentes en la base de datos de Airbnb México.

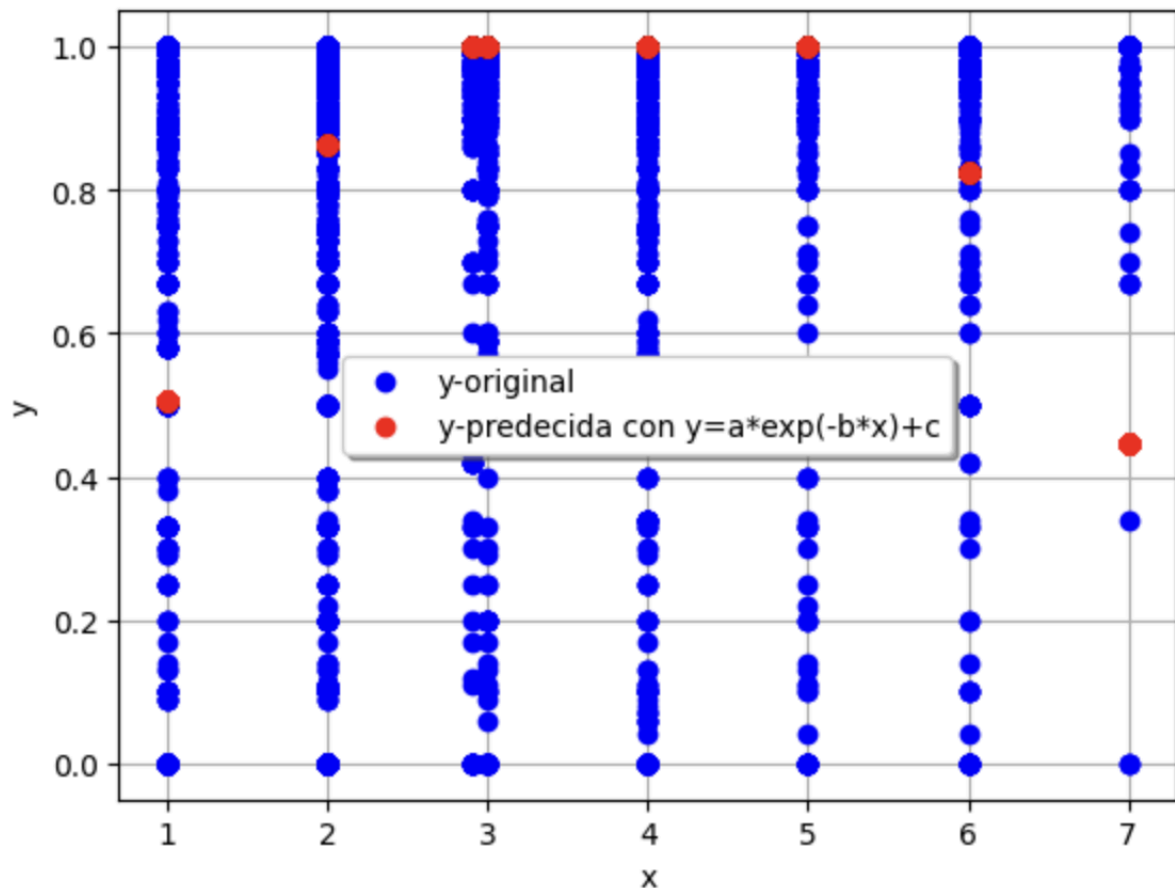
---

## 3. Análisis: Correlación de modelos

### 2.1. HOST RESPONSE RATE

	Modelo	R2	Correlación (r)
0	Host Response Rate: Lineal	0.067059	0.258957
1	Host Response Rate: Función Polinomial Inversa	0.824240	0.907877
2	Host Response Rate: Logarítmica	0.005464	0.073918

Para el análisis de la primera variable, observamos que si se realiza un modelo de regresión lineal para predecir la tasa de respuesta del host, con base en la cantidad de baños con los que un alojamiento cuenta, se obtendría un coeficiente de correlación bastante bajo, siendo del 25%, con el cual no se lograría capturar de manera significativa el comportamiento de Y con base en la independiente. Sin embargo, al ajustarla a una función polinomial inversa, el ajuste se dispara de manera excelente a un 90% de correlación.

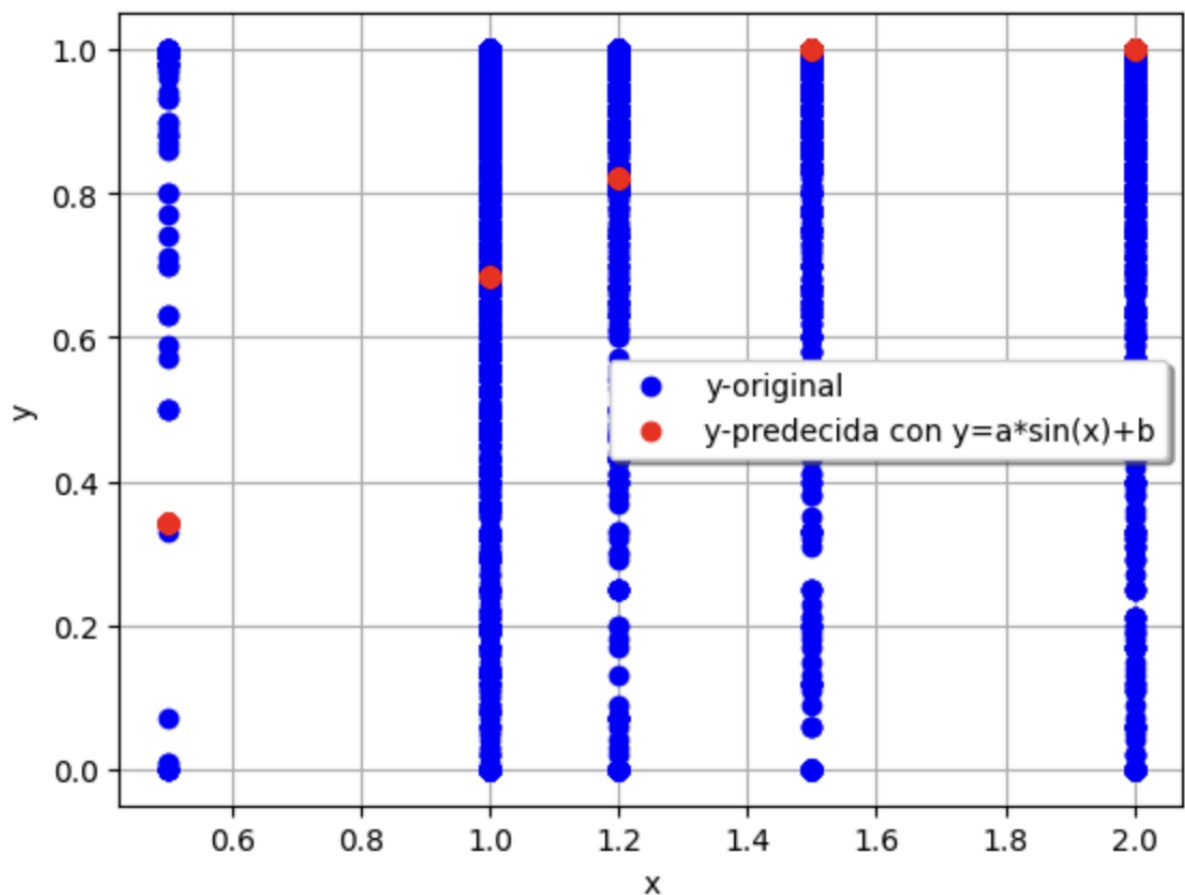


Con este modelo no lineal, se puede interpretar que el efecto de la cantidad de baños (x) sobre la tasa de respuesta de un host (y) es más fuerte al inicio y disminuye conforme x crece, y la relación se aplana hacia un límite horizontal. Es útil cuando cambios iniciales en x tienen gran impacto, pero aumentos posteriores casi no cambian y.

## 2.2. HOST ACCEPTANCE RATE

	Modelo	R2	Correlación (r)
0	Host Acceptance Rate: Lineal	0.015441	0.124260
1	Host Acceptance Rate: Función Polinómica de Se...	0.001701	0.041240
2	Host Acceptance Rate: Función exponencial	0.672003	0.819758

Se observa una correlación baja (12.44%) del modelo lineal para predecir la tasa de aceptación de un host, con base en la cantidad de cuartos con los que un alojamiento cuenta. Al probar otros modelos no lineales, se pudo concluir que una función exponencial permitía capturar de mejor manera el comportamiento de las variables, resultando en un coeficiente de correlación de 81.97%.

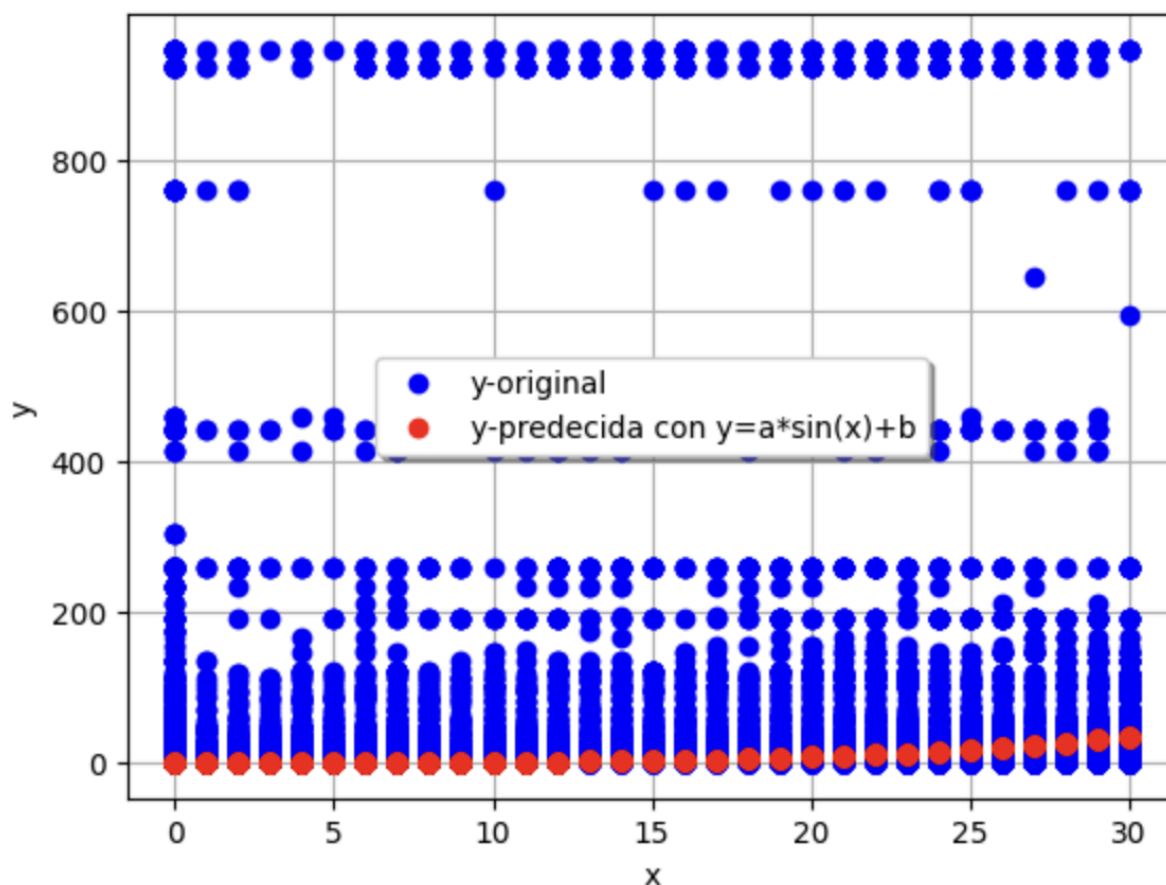


Esto nos indica que los cambios en la tasa de aceptación del host dependen del valor actual de la misma variable, y que la relación con los cuartos de cada alojamiento crece o decrece de forma proporcional, no constante. La pendiente no es lineal; el efecto de x se amplifica o disminuye según la posición en la curva.

### 2.3. HOST LISTINGS COUNT

	Modelo	R2	Correlación (r)
0	Host Total Listings Count: Lineal	0.001525	0.039054
1	Host Total Listings Count: Función Inversa	0.028060	0.167510
2	Host Total Listings Count: Polinomio de grado 4	0.056018	0.236681

Para poder predecir el conteo de propiedades con las que un host cuenta (y) con base en la disponibilidad en un rango de 30 días para rentar una propiedad (x) , el modelo que logra capturar de mejor manera la relación en un polinomio de 4 grados. Aunque el coeficiente de correlación es considerablemente bajo (23.66%), se puede observar como este mejora en comparación de un modelo lineal, lo cual indica que su comportamiento no se trata de un crecimiento constante sobre el tiempo.

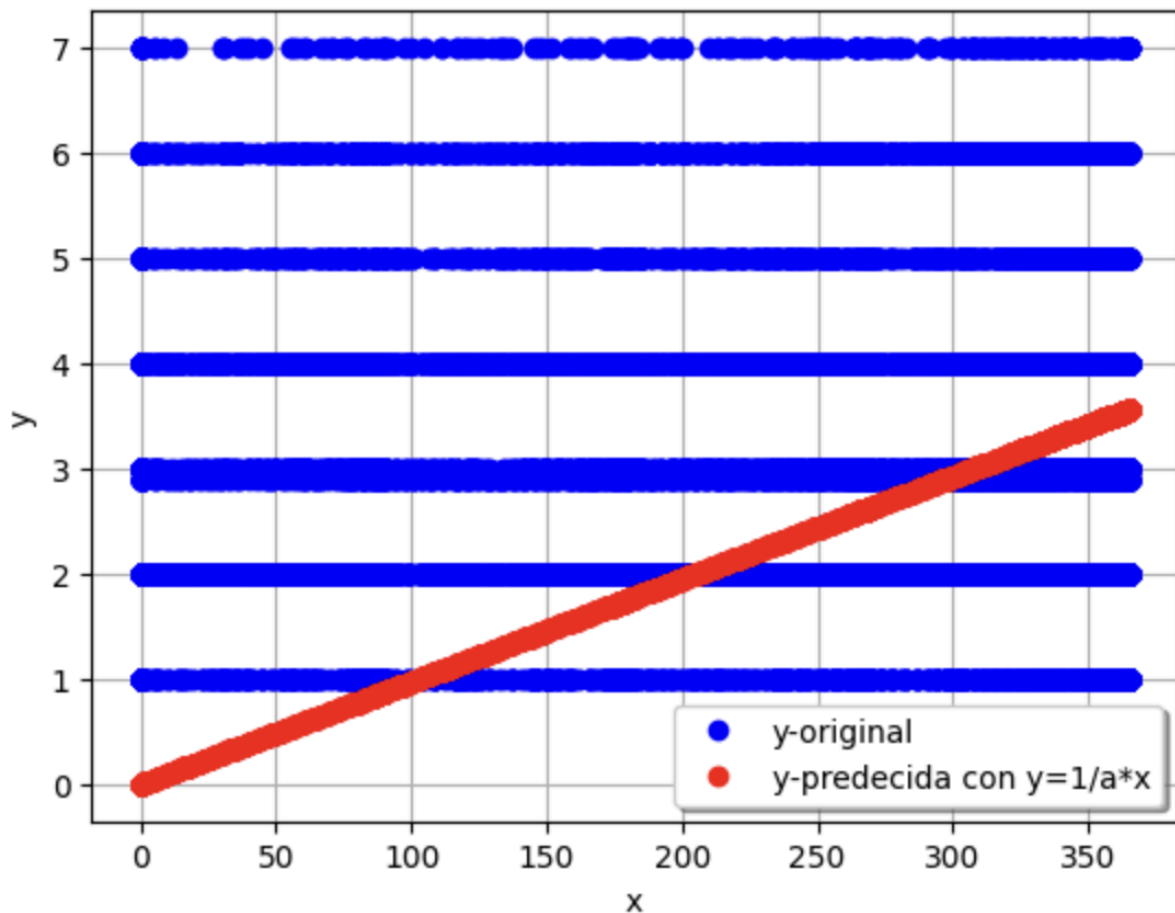


A través de un modelo polinomial de grado 4, se logra capturar curvaturas en el comportamiento de los datos, que es algo complejo de lograr cuando se tiene mucha variabilidad en la distribución de los datos.

#### 2.4. ACCOMMODATES

	Modelo	R2	Correlación (r)
0	Accommodates: Lineal	0.019904	0.141080
1	Accommodates: Función Inversa	0.958316	0.978936
2	Accommodates: Función Cuadrática Inversa	1.481825	1.217303

En el caso de querer predecir la variable 'accommodates', la cual hace referencia a la capacidad de personas que un alojamiento permite, con base en la relación que existe con la disponibilidad del lugar durante todo el año, se observa que al modelar con una ecuación lineal, no se logra capturar el movimiento de los datos con éxito, pues la relación tan solo es del 14.10%. Sin embargo, se puede lograr por medio de una ecuación de función inversa, con la que se obtiene una correlación muy alta del 97.89%.



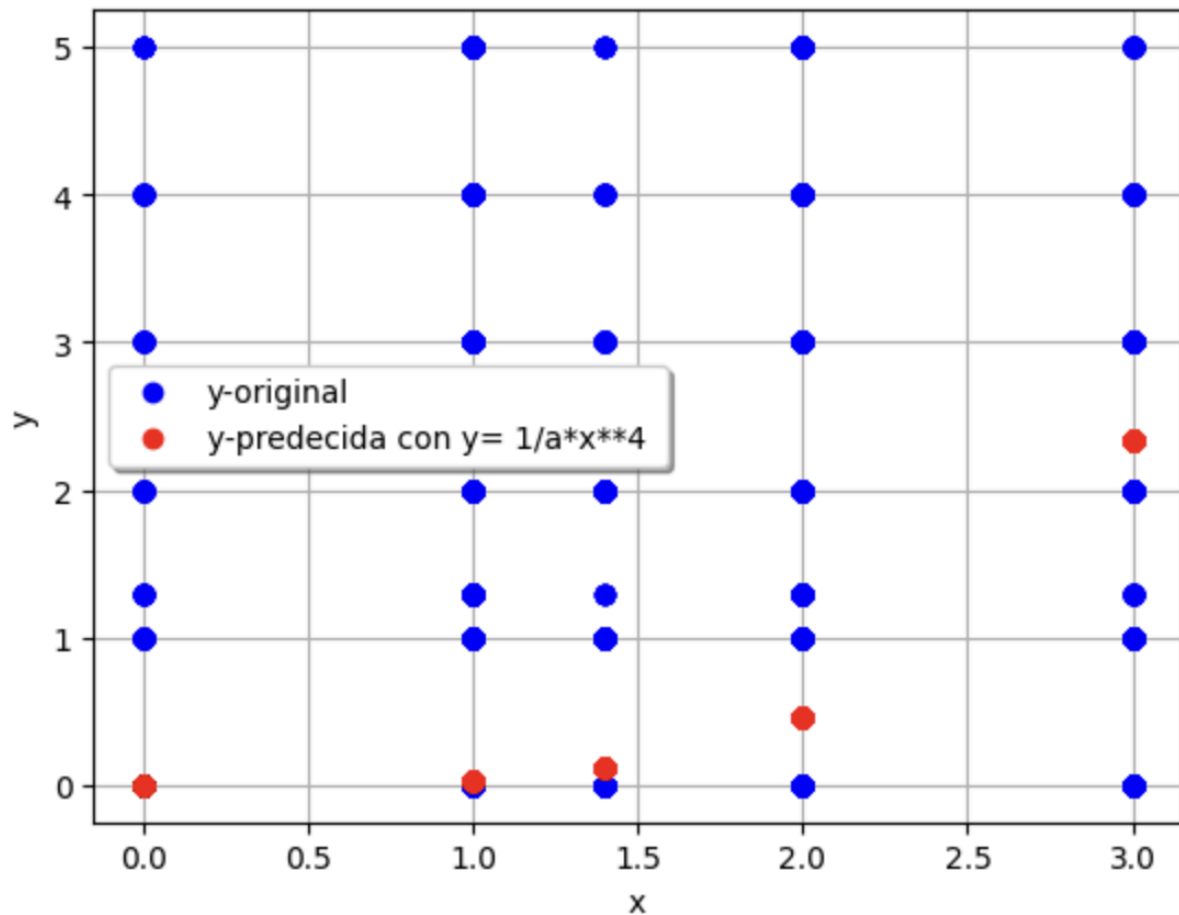
En un modelo inverso, los cambios en x tienen mayor efecto cuando x es pequeño y menos efecto cuando x es grande, y Y se aproxima a un valor límite horizontal b. Es ideal para relaciones donde los aumentos iniciales son críticos, pero más allá de cierto punto, el efecto se estabiliza. Esto se puede interpretar como que, entre menos disponibilidad tenga una alojamiento durante todo el año, será de mayor capacidad dicho alojamiento.

## 2.5. REVIEWS PER MONTH

	Modelo	R2	Correlación (r)
0	Reviews per month: Lineal	0.077408	0.278223
1	Reviews per month: Función Polinomial Inversa	0.065255	0.255451
2	Reviews per month: Función Polinómica de Grado 4	0.671263	0.819307

Para lograr predecir las reviews mensuales con base en el número de cuartos que tienen los alojamientos rentados, se observa un coeficiente de correlación lineal bajo; sin embargo al ser del 27% podríamos considerar que si hay una relación creciente que no es capturada de la mejor manera si se piensa como una línea sin curvas que solo crece. Por lo tanto, al utilizar una función polinomial de grado 4 observamos como el coeficiente R aumenta a

81.93%, indicando que por medio de este modelo se puede considerar el verdadero movimiento de los datos.



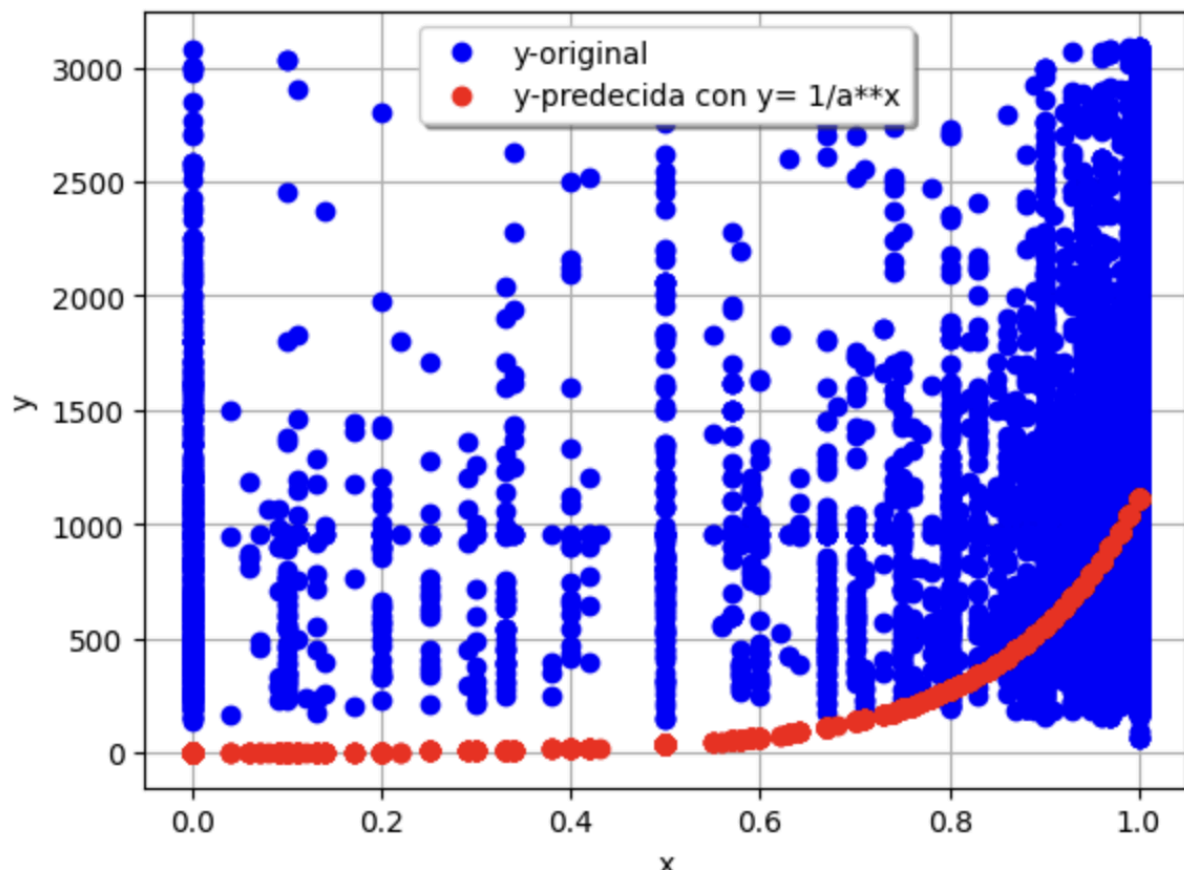
Un polinomio de grado 4 permite analizar **relaciones no muy lineales con múltiples cambios de dirección** entre x e y.

El modelo no lineal ajustado con un polinomio de grado 4 muestra que la relación entre los cuartos de un alojamiento y los reviews mensuales es compleja: se estabiliza inicialmente, luego aumenta en valores medios y se exponencia con posibles caídas futuras. Esto sugiere que la influencia de los cuartos sobre la calificación de los hosts y su servicio depende del tamaño del alojamiento y del rango de cuartos.

## 2.6. PRICE

	Modelo	R2	Correlación (r)
0	Price: Lineal	0.068980	0.262641
1	Price: Función Cociente entre Polinomios Grado 6	0.092528	0.304184
2	Price: Función Exponencial Decreciente	0.169383	0.411561

En el caso de  $y = \text{precio}$  y  $x = \text{tasa de respuesta del host}$ , se observa cómo se logra mejorar la correlación entre diversos modelos ajustados. Inicialmente se nota una correlación de 26.26% en un modelo lineal, lo que indica que si hay crecimiento en la relación, pero la distribución de los datos no son solo crecientes. Es así como, al llegar a un modelo de función exponencial, el coeficiente aumenta a 41.15%, indicando mayor fiabilidad para predecir con este tipo de ecuación.



Al observar el gráfico se puede comprender porqué existe una relación lineal no de 0, pero a su vez, porqué al ocupar un modelo exponencial se logra capturar de mucha mejor manera la relación de las variables para predecir con mayor fiabilidad. A través de este modelo, comprendemos que el precio del alojamiento aumenta (o disminuye) de manera proporcional y creciente con la tasa de respuesta del anfitrión. Cada incremento en la tasa de respuesta tiene un efecto mayor cuando el precio ya es alto.

---

#### 4. Conclusión

Los modelos no lineales son esenciales cuando las relaciones entre variables no siguen una tendencia recta, como suele ocurrir en muchas bases de datos, por ejemplo Airbnb México.



Este tipo de modelos permiten capturar patrones más realistas y dinámicos, como efectos que crecen o decrecen de forma proporcional, curvas con inflexiones, y relaciones que se estabilizan o se amplifican según el rango de valores.

Su importancia radica en que:

- Flexibilidad: Pueden ajustarse a relaciones complejas que los modelos lineales no detectan.
- Predicción más precisa: Mejoran la capacidad de estimar valores de  $y$  en escenarios donde la dependencia de  $x$  es variable o curvada.
- Análisis detallado: Permiten identificar rangos de la variable independiente donde los cambios tienen mayor o menor efecto sobre la dependiente.

En resumen, los modelos no lineales son herramientas poderosas para predecir y analizar datos con relaciones complejas, ofreciendo un entendimiento más realista de cómo las variables interactúan en situaciones prácticas.