



Tecnológico de Monterrey

Actividad 2 (Regresión No Lineal)

Erick Antonio Nava Melchy | A01424422

Profesor: Alfredo García Suárez

Co-Titular: Elena García y García

Co-Titular: Maria Luisa Gomez Barrios

**Gestión de proyectos de plataformas tecnológicas
Grupo 201**

07/10/2025

Tabla de contenido

Resumen general.....	2
Hallazgos específicos	2
Conclusiones	9
Nota técnica: linealidad en los parámetros vs no linealidad en la forma.....	9

Resumen general

Este análisis se realizó utilizando una base de datos de Airbnb, con el propósito de evaluar distintas relaciones entre variables relevantes de los anuncios y de los anfitriones. Se aplicaron tres modelos diferentes: lineal, cuadrático y MODELO 2 (una forma no lineal sin intercepto), comparando su poder explicativo a través del coeficiente de determinación R^2 .

Las variables objetivo fueron:

- host_response_rate
- host_acceptance_rate
- host_total_listings_count
- accommodates
- reviews_per_month
- price

A partir del heatmap de correlaciones, se observó que las variables host_response_rate y host_acceptance_rate, aunque ambas son variables objetivo, fueron las más correlacionadas entre sí.

Esto indica que existe una relación operativa estrecha entre el hecho de que un anfitrión responda rápidamente a las consultas y la probabilidad de que acepte las solicitudes, lo cual tiene sentido dentro de la dinámica del negocio de Airbnb.

En general:

- El **modelo lineal** es el más interpretable y resulta útil para relaciones directas y proporcionales.
- El **modelo cuadrático** capta mejor las **curvaturas suaves** y mejora ligeramente el ajuste en varios casos.
- El **MODELO 2** ofrece **ajustes muy altos** en variables donde las relaciones presentan límites naturales o comportamientos escalonados, aunque debe interpretarse con cuidado porque puede exagerar el ajuste en algunos casos.

Hallazgos específicos

1) host_response_rate (y) - host_acceptance_rate

- **Correlación base:** 0.355
- **R^2 :** Lineal = 0.126 | Cuadrático = 0.135 | MODELO 2 = 0.988

Ambas variables reflejan el comportamiento operativo del anfitrión. Un host que acepta más solicitudes suele también responder con mayor rapidez, lo que puede deberse al uso de herramientas automáticas, buena organización, o experiencia en la gestión de anuncios. El MODELO 2 captura la curvatura natural que surge porque estas tasas están acotadas entre 0 y 1, pero su R^2 casi perfecto puede responder más a esa estructura limitada que a una relación causal directa.

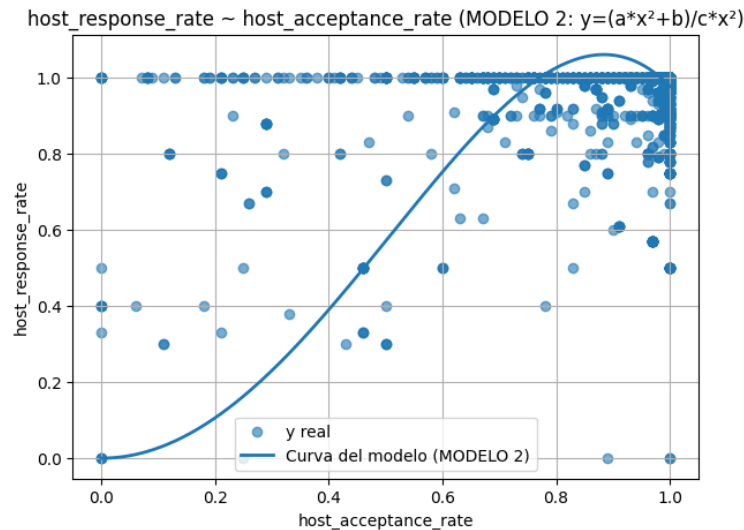


Figura 1. *host_response_rate - host_acceptance_rate (MODELO 2).*

♦ Top 10 correlaciones con 'host_response_rate':

	Variable	Correlación
0	host_acceptance_rate	0.354989
1	minimum_nights	-0.200014
2	review_scores_communication	0.188588
3	review_scores_checkin	0.178584
4	review_scores_accuracy	0.176297
5	review_scores_value	0.151244
6	review_scores_rating	0.141376
7	review_scores_location	0.140631
8	review_scores_cleanliness	0.138737
9	estimated_occupancy_l365d	0.134676

Tabla 1. *Top 10 correlaciones de host_response_rate.*

2) host_acceptance_rate (y) - host_response_rate

- **Correlación base:** 0.355
- **R²:** Lineal = 0.126 | Cuadrático = 0.130 | MODELO 2 = 0.987

Este caso refleja prácticamente la misma relación, pero en sentido inverso. Los anfitriones que responden con rapidez suelen tener una actitud más abierta hacia las reservas. Esto se asocia con prácticas de servicio proactivo y estrategias que buscan maximizar la ocupación y la satisfacción del huésped. Nuevamente, el MODELO 2 encaja de forma casi perfecta, pero debe interpretarse como una curvatura estructural más que como una dependencia exacta.

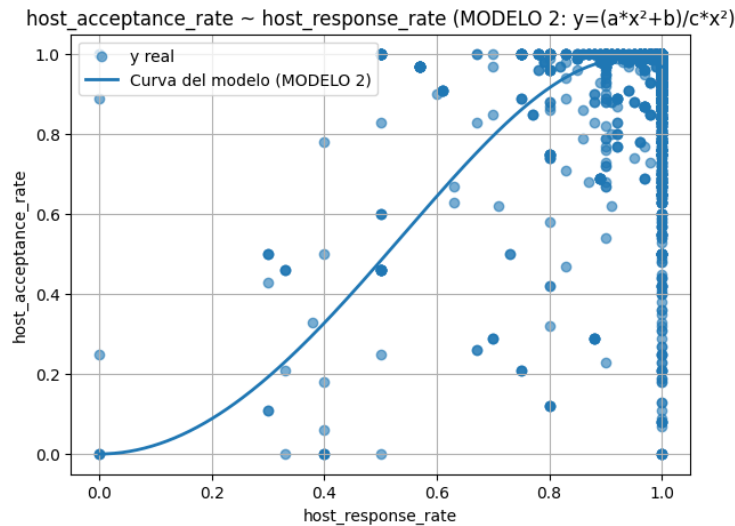


Figura 2. *host_acceptance_rate* - *host_response_rate* (MODELO 2).

♦ Top 10 correlaciones con '*host_acceptance_rate*':

	Variable	Correlación
0	<i>host_response_rate</i>	0.354989
1	<i>minimum_nights</i>	-0.247931
2	<i>maximum_minimum_nights</i>	-0.221910
3	<i>reviews_per_month</i>	0.193829
4	<i>estimated_occupancy_l365d</i>	0.190021
5	<i>number_of_reviews_ltm</i>	0.184232
6	<i>number_of_reviews_ly</i>	0.169090
7	<i>estimated_revenue_l365d</i>	0.165454
8	<i>number_of_reviews_l30d</i>	0.164831
9	<i>number_of_reviews</i>	0.120020

Tabla 2. Top 10 correlaciones de *host_acceptance_rate*.

3) *host_total_listings_count* (y) - *host_listings_count*

- **Correlación base:** 0.96
- **R²:** Lineal = 0.923 | Cuadrático = 0.933 | MODELO 2 = 0.901

En el contexto de Airbnb, ambas variables miden el tamaño del portafolio de propiedades que maneja un anfitrión. El modelo cuadrático ofrece un ajuste ligeramente mejor, lo que sugiere una curvatura leve en la relación entre el número total de listados y los listados principales. Esto puede reflejar que, conforme un host crece, el aumento en listados totales no siempre es perfectamente proporcional al número de unidades principales, ya que puede haber coanfitriones, anuncios compartidos o agrupados. En resumen, el modelo cuadrático describe mejor esa pequeña no linealidad, mientras que el lineal sigue siendo útil y sencillo.

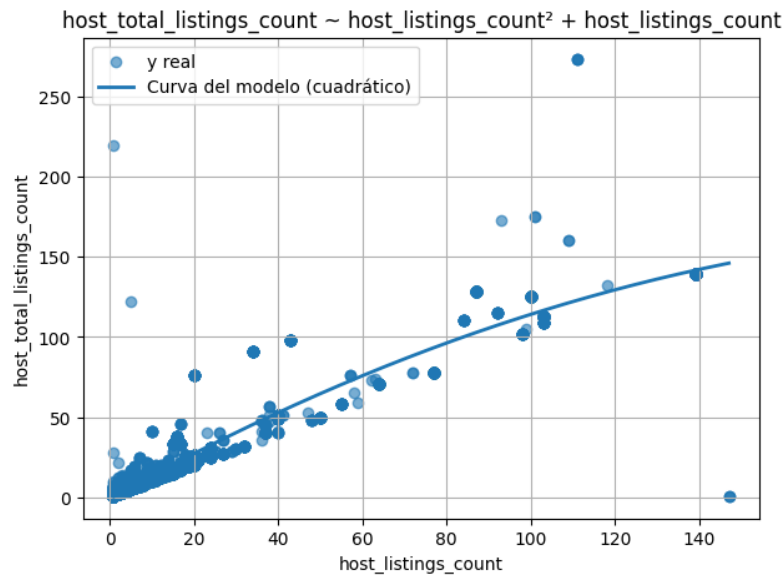


Figura 3. *host_total_listings_count - host_listings_count (Modelo cuadrático).*

♦ Top 10 correlaciones con 'host_total_listings_count':

	Variable	Correlación
0	host_listings_count	0.960967
1	calculated_host_listings_count	0.616905
2	calculated_host_listings_count_entire_homes	0.490141
3	maximum_nights	0.342403
4	maximum_maximum_nights	0.208160
5	price	0.191560
6	number_of_reviews	-0.173788
7	review_scores_value	-0.170080
8	reviews_per_month	-0.169615
9	review_scores_communication	-0.159518

Tabla 3. *Top 10 correlaciones de host_total_listings_count.*

4) accommodates (y) - bedrooms

- **Correlación base:** 0.705
- **R²:** Lineal = 0.497 | Cuadrático = 0.501 | MODELO 2 = 0.842

A mayor número de recámaras, generalmente aumenta la capacidad de alojamiento, pero no de forma estrictamente lineal. Intervienen factores como la distribución interna, sofás cama, literas o el uso del espacio común. El MODELO 2 logra un ajuste mucho más alto, reflejando cómo el aumento de habitaciones impacta de manera más fuerte en los primeros incrementos (pasar de un estudio a un departamento) y se estabiliza en rangos altos, donde sumar más cuartos ya no cambia tanto la capacidad.

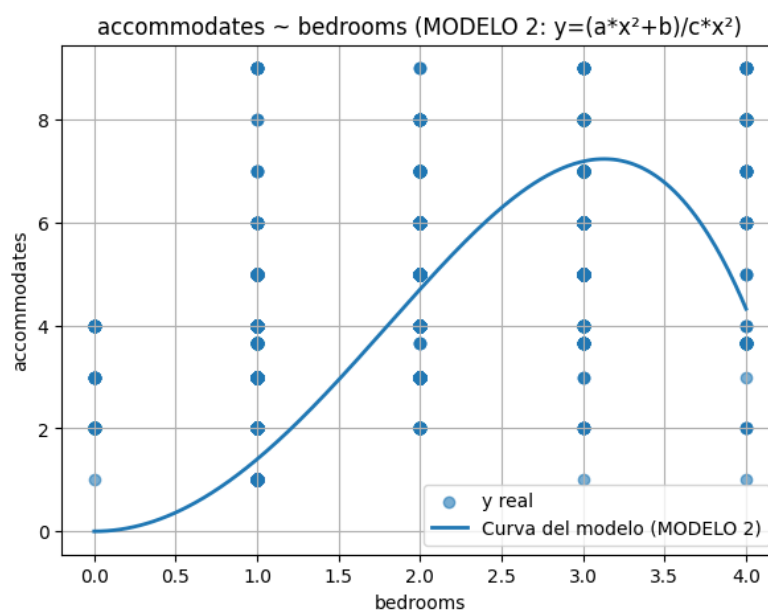


Figura 4. accommodates - bedrooms (MODELO 2).

◆ Top 10 correlaciones con 'accommodates':

	Variable	Correlación
0	bedrooms	0.705102
1	beds	0.657001
2	bathrooms	0.414471
3	price	0.328168
4	calculated_host_listings_count_private_rooms	-0.264054
5	estimated_revenue_l365d	0.168038
6	availability_60	-0.142042
7	host_total_listings_count	0.139642
8	host_listings_count	0.130722
9	availability_30	-0.120185

Tabla 4. Top 10 correlaciones de accommodates.

5) reviews_per_month (y) - number_of_reviews_ltm

- Correlación base: 0.796
- R^2 : Lineal = 0.633 | Cuadrático = 0.635 | MODELO 2 = 0.714

Las propiedades con más reseñas acumuladas suelen recibir más interacción reciente. Esto se relaciona con efectos de visibilidad, reputación y posicionamiento dentro de la plataforma. El MODELO 2 mejora el ajuste, indicando que hay una saturación natural: cuando un anuncio

ya tiene muchas reseñas, las nuevas llegan con menor frecuencia. En otras palabras, el crecimiento en reseñas recientes no sigue siendo constante conforme aumenta el total histórico.

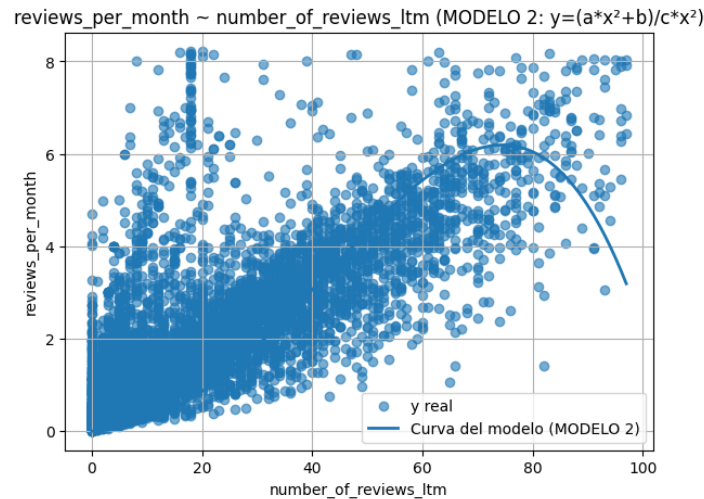


Figura 5. reviews_per_month - number_of_reviews_ltm (MODELO 2).

♦ Top 10 correlaciones con 'reviews_per_month':

	Variable	Correlación
0	number_of_reviews_ltm	0.795793
1	estimated_occupancy_l365d	0.745389
2	number_of_reviews_ly	0.715451
3	number_of_reviews_l30d	0.696444
4	number_of_reviews	0.521699
5	estimated_revenue_l365d	0.506014
6	maximum_minimum_nights	-0.219759
7	review_scores_value	0.210227
8	minimum_nights	-0.198995
9	host_acceptance_rate	0.193829

Tabla 5. Top 10 correlaciones de reviews_per_month.

6) price (y) - bedrooms

- **Correlación base:** 0.378
- **R²:** Lineal = 0.143 | Cuadrático = 0.156 | MODELO 2 = 0.607

El precio de los alojamientos no crece linealmente con el número de recámaras. En Airbnb, el salto de una habitación a dos suele representar un cambio de categoría (de pareja a familia, por ejemplo), lo que incrementa el valor más que en pasos posteriores (de tres a cuatro habitaciones). El MODELO 2 capta bien esta estructura escalonada del precio, aunque es evidente que influyen muchos otros factores, como ubicación, amenidades, vista, demanda y temporada.

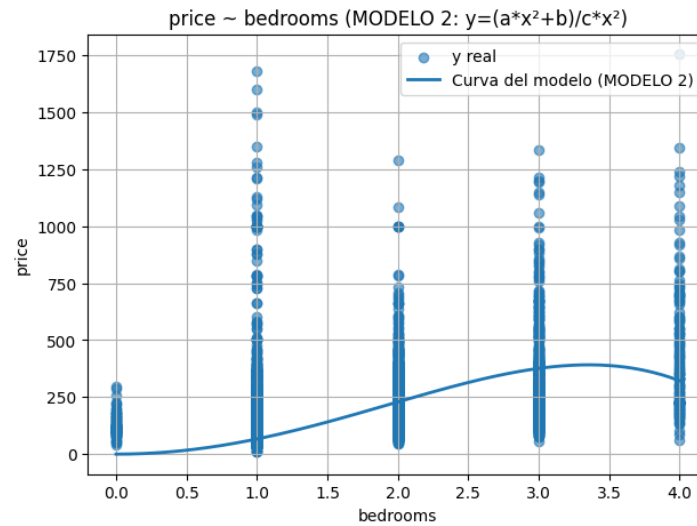


Figura 6. price - bedrooms (MODELO 2).

♦ Top 10 correlaciones con 'price':

	Variable	Correlación
0	bedrooms	0.377667
1	bathrooms	0.351666
2	accommodates	0.328168
3	longitude	0.244822
4	latitude	-0.243606
5	beds	0.227832
6	estimated_revenue_l365d	0.227036
7	host_total_listings_count	0.191560
8	host_listings_count	0.171359
9	calculated_host_listings_count	0.162297

Tabla 6. Top 10 correlaciones de price.

La siguiente tabla resume los principales resultados de los tres modelos aplicados (lineal, cuadrático y MODELO 2) para cada variable objetivo y su predictor más correlacionado dentro del contexto de Airbnb. Se observan diferencias claras en el poder explicativo (R^2) de cada modelo: mientras el modelo lineal describe adecuadamente las relaciones directas, los modelos no lineales captan mejor las curvaturas naturales y los efectos de saturación presentes en ciertas variables, como las tasas o el precio.

target	predictor	corr_y_x	r2_lineal	r2_cuadratico	r2_mod2
host_response_rate	host_acceptance_rate	0.354989	0.126017	0.134648	0.987728
host_acceptance_rate	host_response_rate	0.354989	0.126017	0.130351	0.987107
host_total_listings_count	host_listings_count	0.960967	0.923457	0.932471	0.901395
accommodates	bedrooms	0.705102	0.497169	0.500625	0.841647
reviews_per_month	number_of_reviews_ltm	0.795793	0.633286	0.634657	0.714235
price	bedrooms	0.377667	0.142633	0.155962	0.606712

Tabla 7. Comparativa de correlaciones y coeficientes de determinación (R^2) obtenidos para cada modelo y variable objetivo.

Conclusiones

- Las tasas de anfitrión (host_response_rate y host_acceptance_rate) muestran una dependencia mutua: quienes responden más rápido también aceptan más solicitudes. Esto puede deberse a automatización, experiencia o mayor profesionalismo en la gestión de anuncios.
- Las relaciones casi proporcionales, como entre host_total_listings_count y host_listings_count, se explican bien con modelos simples.
- Las variables de capacidad y precio muestran comportamientos no lineales reales, reflejando la dinámica del mercado y las características físicas de las propiedades.
- En términos generales, el MODELO 2 destaca en relaciones con curvaturas pronunciadas o efectos de saturación, mientras que el modelo lineal basta para relaciones directas y proporcionales.
- El modelo cuadrático sirve como punto intermedio: capta pequeñas desviaciones de linealidad sin perder interpretabilidad.

Nota técnica: linealidad en los parámetros vs no linealidad en la forma

Aunque los tres modelos (lineal, cuadrático y MODELO 2) se ajustan mediante regresión OLS (Ordinary Least Squares), es importante distinguir dos conceptos diferentes:

1. Linealidad en los parámetros del modelo

En todos los casos, el modelo estima coeficientes β_i de forma lineal, es decir, se ajustan combinaciones como

$$\hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots$$

Esto implica que la estimación sigue siendo “regresión lineal” en el sentido estadístico, lo que permite usar técnicas estándar para analizar bondad de ajuste, significancia, residuos, r^2 , etc.

2. No linealidad en la relación yvs x

Aunque el método de estimación es lineal, la forma de la relación entre y y x puede ser no lineal porque usamos variables derivadas (como x^2 , x^4 , etc.) como predictores.

- En el **modelo lineal**, usamos directamente x .
- En el **modelo cuadrático**, usamos x^2 y x .

- En el **MODELO 2**, usamos x^4 y x^2 , sin intercepto, para capturar curvaturas más marcadas o efectos de saturación. Por ello, aunque el algoritmo es OLS, la relación entre y y x no tiene que ser una recta, sino que puede tener forma de parábola o de curva más compleja.

Referencias

- IBM SPSS Statistics. (2025, June 10). <https://www.ibm.com/docs/es/spss-statistics/31.0.0?topic=regression-nonlinear>
- Frost, J. (2025, May 1). *The Difference between Linear and Nonlinear Regression Models*. Statistics by Jim. <https://statisticsbyjim.com/regression/difference-between-linear-nonlinear-regression-models/>
- Reyes, G. (n.d.). *ENSAYO ESTADISTICO*. Scribd. <https://es.scribd.com/document/422852597/ENSAYO-ESTADISTICO>
- *Regresión no lineal como utilizar la regresión no lineal para la previsión de inversiones* - FasterCapital. (n.d.). FasterCapital. <https://fastercapital.com/es/contenido/Regresion-no-lineal--como-utilizar-la-regresion-no-lineal-para-la-prevision-de-inversiones.html>
- *Regresión no lineal: Tipos, Ejemplos* | StudySmarter. (n.d.). StudySmarter ES. <https://www.studysmarter.es/resumenes/ingenieria/matematicas-de-la-ingenieria/regresion-no-lineal/>