



Tecnológico de Monterrey

Actividad 3.1 Regresión No Lineal

Lucero Jannete López García A01736938

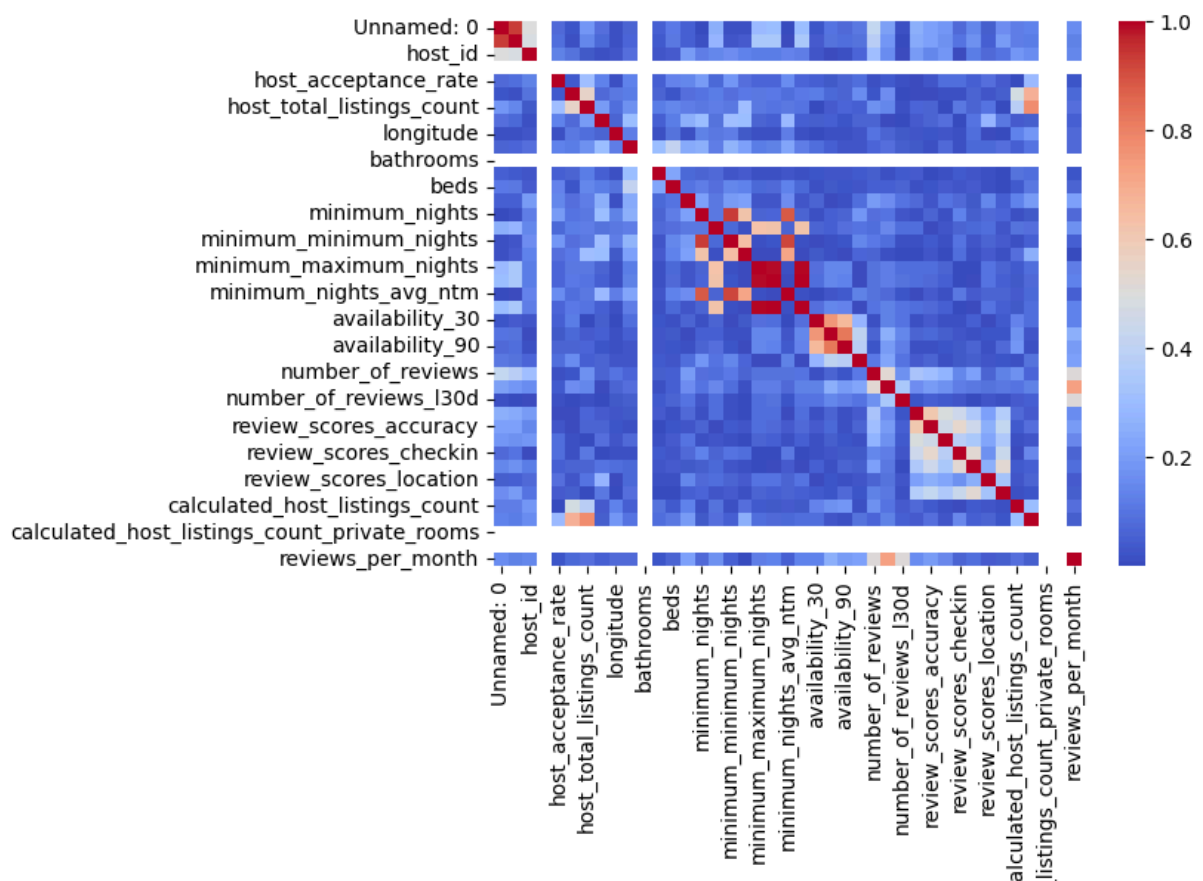
**Analítica de datos y herramientas de inteligencia artificial
I (Gpo 101)**

Fecha
08/04/2025

Para poder generar la comparativa entre variables de los dos modelos No lineales contra el modelo lineal, hay que recordar el coeficiente de correlación lineal simple de las variables cuantitativas de la anterior actividad. Un claro ejemplo es que de las variables. `host_response_rate` (variable objetivo), `host_acceptance_rate` (variable objetivo), `host_total_listings_count` (variable objetivo), `accommodates` (variable objetivo), `reviews_per_month` (variable objetivo) y `price`(variable objetivo), **la variable `host_response_rate` en el modelo lineal simple cuenta con valores nulos en toda la columna**, por otra parte, las demás columnas algunas cuentan con valores nulos, son pocos pero existen.

A continuación se presenta nuestro mapa de calor y tablas comparativas entre los coeficientes de determinación y de correlación de cada variables objeto por modelo, se encuentran separadas ya que, para el segundo modelo de cada variable objeto se ocuparon diferentes según dieran mejores resultados.

Para el modelo de regresión lineal el cálculo de su coeficiente de determinación y correlación son de la actividad anterior.



Host_response_rate / review_per_month

Coefficiente	Modelo NL 1 Función Cuadrática	Modelo NL 2 Función Senoidal	Modelo lineal de regresión simple	Observaciones
Determinación	NAN	NAN	NAN	En estos modelos todos dan NAN debido a que no hay correlación, desde el modelo lineal se observó este comportamiento.
Correlación	NAN	NAN	NAN	

Al intentar analizar la relación entre Host_response_rate y review_per_month usando distintos modelos (lineal, cuadrático y senoidal), todos arrojaron NaN en los resultados. Esto me llevó a revisar los datos y descubrí que Host_response_rate es una variable constante: todos los registros tienen el mismo valor.

Al ser constante, no hay forma de calcular una variación ni ajustar un modelo, ya que no existe variación que permita encontrar una relación con la otra variable. Incluso al aplicar funciones más complejas, como la cuadrática o senoidal, el resultado sigue siendo el mismo porque el valor de entrada no cambia.

Este comportamiento me permitió concluir que Host_response_rate, al menos en este conjunto de datos, no es útil para explicar el número de reseñas por mes. Fue importante detectarlo, ya que también ayuda a identificar qué variables realmente aportan al análisis.

.

host_acceptance_rate / host_total_listings_count

Coefficiente	Modelo NL 1 Función Cuadrática	Modelo NL 2 Función Logarítmica	Modelo lineal de regresión simple	Observaciones
Determinación	0.015	0.013	0.014	Se escogió el modelo cuadrático y logaritmo porque son los que brindan correlaciones similares al modelo lineal
Correlación	0.122	0.116	0.120	

Analicé la relación entre host_acceptance_rate y host_total_listings_count utilizando tres modelos: lineal, cuadrático y logarítmico. Los tres modelos mostraron valores de

determinación (R^2) muy bajos, alrededor de 0.013 a 0.015, lo que indica que la capacidad de estos modelos para explicar la variación en la variable dependiente es limitada.

La estimación en los tres casos fue también muy baja, entre 0.116 y 0.122, lo que confirma que no hay una relación fuerte entre ambas variables. Aun así, se eligieron los modelos cuadráticos y logarítmicos porque ofrecieron resultados similares al modelo lineal y permitieron observar si alguna transformación podía mejorar mínimamente el ajuste.

Aunque hay una leve relación positiva, es muy débil, por lo que esta variable por sí sola no es un buen predictor. Aun así, sirve para explorar posibles patrones o complementar otros análisis con variables adicionales.

host_total_listings_count / host_listings_count

Coefficiente	Modelo NL 1 Función Cuadrática	Modelo NL 2 Función cuadrática inversa	Modelo lineal de regresión simple	Observaciones
Determinación	0.73	-0.8	0.69	La relación entre host_total_listings_count y host_listings_count es fuerte y positiva, lo que se refleja tanto en la correlación alta como en el buen ajuste del modelo.
Correlación	0.85	NAN	0.8346	

La relación entre host_total_listings_count y host_listings_count. Los resultados fueron bastante distintos a los casos anteriores. El modelo lineal y el modelo cuadrático mostraron un buen ajuste, con valores de determinación de 0.69 y 0.73 respectivamente, lo que indica una relación bastante fuerte entre las dos variables. Además, la calificación fue alta en ambos modelos (0.8346 y 0.85), lo que refuerza esta relación positiva.

Por otro lado, el modelo de función cuadrática inversa no fue adecuado para este conjunto de datos, ya que arrojó un valor de R^2 negativo (-0.8), lo que significa que el modelo es peor que simplemente usar el promedio. Además, la evaluación no se pudo calcular (dio NaN), probablemente por problemas con la forma de la función y los valores involucrados. Sin embargo, fue el que más se acercó a 0, debido a que otros modelos no lineales daban correlaciones de -44.

Tanto el modelo lineal como el cuadrático fueron útiles y mostraron una relación fuerte entre estas variables. Esto tiene sentido, ya que ambos miden aspectos similares del número de propiedades que gestiona un host.

accommodates / bedrooms

Coeficiente	Modelo NL 1 Función Cuadrática	Modelo NL 2 Función Logarítmica	Modelo lineal de regresión simple	Observaciones
Determinación	0.41	0.40	0.3963	Los tres modelos presentaron un desempeño similar, con una correlación moderada que sugiere cierta relación entre las variables.
Correlación	0.64	0.63	0.629	

La relación entre las dos variables usando tres modelos: el lineal, el cuadrático y el logarítmico. Todos mostraron un comportamiento bastante similar, con coeficientes de determinación (R^2) entre 0.396 y 0.41. Esto indica que hay una relación moderada entre las variables: no es débil, pero tampoco lo suficientemente fuerte como para explicar completamente la variabilidad.

La calificación también se mantiene en un rango muy parecido en los tres modelos, entre 0,629 y 0,64. Por eso, se eligieron tanto el modelo cuadrático como el logarítmico como opciones válidas, ya que muestran un ajuste muy similar al modelo lineal. Además de que los demás modelos resultaban en correlaciones más bajas o errores.

reviews_per_month / number_of_reviews_ltm

Coeficiente	Modelo NL 1 Función Cuadrática	Modelo NL 2 Función cociente entre polinomios	Modelo lineal de regresión simple	Observaciones
Determinación	0.58	0.54	0.540	Los modelos cuadrático y no lineal también ofrecieron buenos resultados, por lo que cualquiera de los tres es válido para representar esta relación.
Correlación	0.76	0.73	0.73	

En este análisis evalué la relación entre `reviews_per_month` y `number_of_reviews_ltm` utilizando tres modelos: lineal, cuadrático y uno basado en cociente entre polinomios. Los tres modelos mostraron resultados bastante consistentes, con valores de R^2 entre 0.54 y 0.58, lo que indica que hay una relación significativa entre estas dos variables.

La correlación también fue alta en todos los casos, especialmente en el modelo lineal (0.76), lo que confirma que hay una conexión fuerte y positiva. El modelo cuadrático y el modelo no lineal por cociente también reflejan bien esta tendencia, aunque con ligeras variaciones.

Esto tiene sentido, ya que el número de reseñas en los últimos doce meses está directamente relacionado con el promedio mensual. Es una relación esperada, pero útil de confirmar con datos y modelos.

price / accommodates

Coefficiente	Modelo NL 1 Función Cuadrática	Modelo NL 2 Función Valor absoluto	Modelo lineal de regresión simple	Observaciones
Determinación	0.076	0.075	0.075	Todos los modelos evaluados dieron resultados similares, por lo que ninguna mejora de manera significativa el ajuste.
Correlación	0.275	0.274	0.274	

Price y accommodates, al uso de tres modelos: un modelo lineal, un modelo cuadrático y un modelo de valor absoluto. Los tres modelos mostraron resultados muy similares, con un R^2 de alrededor de 0,075 y un rango de aproximadamente 0,274. Esto indica que la relación entre estas dos variables es muy débil.

Aunque se podría pensar que una mayor capacidad debería implicar un mayor precio, los datos muestran que esta relación no es tan clara. Es probable que el precio se vea influenciado por otros factores más relevantes, como la ubicación, el tipo de propiedad y los servicios incluidos, entre otros.

Por lo tanto, aunque existe una ligera tendencia positiva, no es suficiente predecir el precio basándose únicamente en el número de personas que puede alojar un alojamiento.

Conclusión

Tras analizar las diferentes relaciones entre variables mediante modelos lineales y no lineales, me di cuenta de que el modelo de regresión lineal simple era el más consistente y útil en la mayoría de los casos. Aunque probé modelos cuadráticos, logarítmicos y otros modelos no

lineales, generalmente no mejoraron mucho los resultados y, en algunos casos, incluso arrojaron resultados NaN o negativos.

El modelo lineal, por el contrario, funcionó bien, fue fácil de interpretar y dio resultados bastante sólidos, especialmente en relaciones como `reviews_per_month` con `number_of_reviews_ltm` o `host_total_listings_count` con `host_listings_count`, donde la compensación fue fuerte y clara.

En otros casos, como precio y acomodación, la relación era muy débil, pero aún así el modelo lineal mostró un comportamiento estable, similar al de los no lineales, sin complicar el análisis.

Aunque no todos los modelos mostraron relaciones sólidas, el modelo lineal fue el más adecuado debido a su simplicidad, estabilidad y utilidad. En el futuro, si queremos mejorar el análisis o realizar predicciones más precisas, sería beneficioso incluir más variables o utilizar modelos más complejos; sin embargo, en este caso, la regresión lineal fue suficiente para comprender el comportamiento general.

De igual manera, debemos comprender el tipo de comportamiento de los datos mediante diagramas de dispersión para observar su variación y poder decidir cuál es el mejor modelo en función de lo que queremos analizar y el comportamiento de las variables.