



Tecnológico de Monterrey

Unidad de formación:

Gestión de proyectos de plataformas tecnológicas (Gpo 201)

Actividad 1 (Regresión Lineal Simple y Múltiple)

Profesor:

Alfredo García Suarez

Alumna:

Samantha Aja | A01738365

Fecha de entrega:

28 de septiembre de 2025

1. Crear un nuevo repositorio con el nombre: **Actividad 1 (Regresión Lineal)**
2. Agregar el archivo: **El archivo .csv de la ciudad de su elección (A partir de las bases de datos listings.csv.gz), ingresar a:**
<http://insideairbnb.com/get-the-data/>
3. [Links to an external site.](#)
4. Realiza las acciones de preprocesamiento necesarias: **Nulos y Outliers**

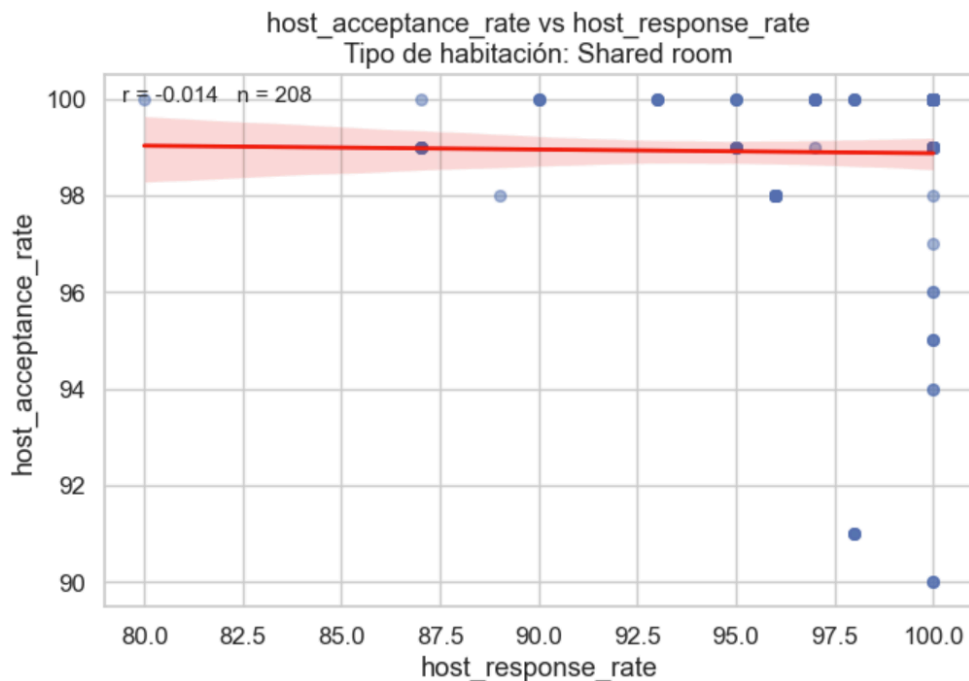
REGRESIÓN LINEAL SIMPLE (Punto 4 y Punto 5)

4. Analizar la correlación que existe en cada tipo de habitación (**Elegir 4 tipos**) respecto a las variables siguientes en el siguiente orden “(dependiente, independiente)”, utilizando **Python** y **Google Colab**, obtener los datos y gráficos requeridos en cada caso.

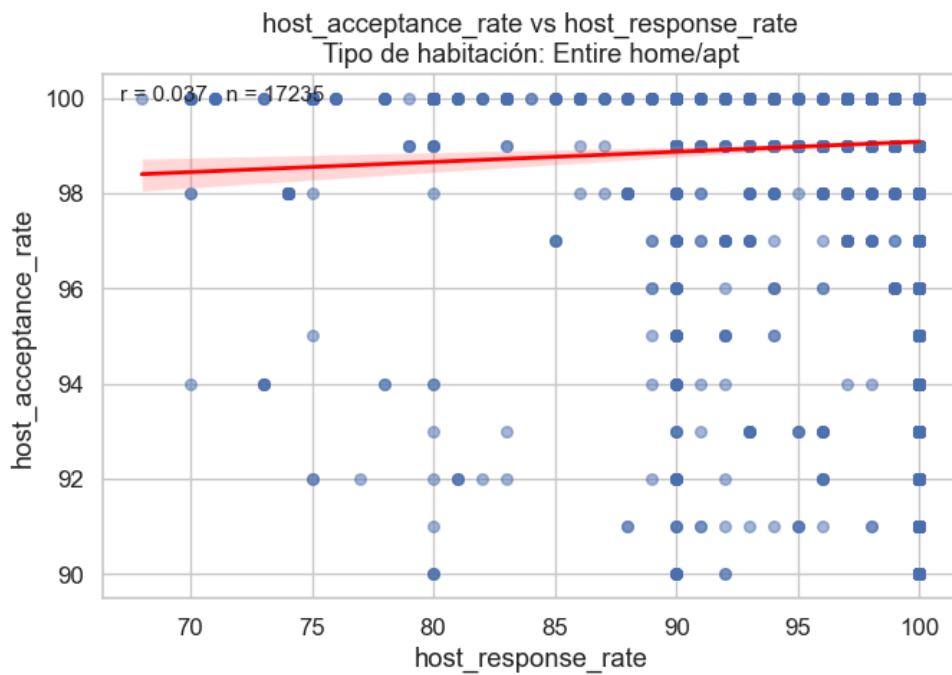
“host_acceptance_rate
host_response_rate” vs “review_scores_rating
calculated_host_listings_count” vs

“host_acceptance_rate vs price” “availability_365 vs number_of_reviews”

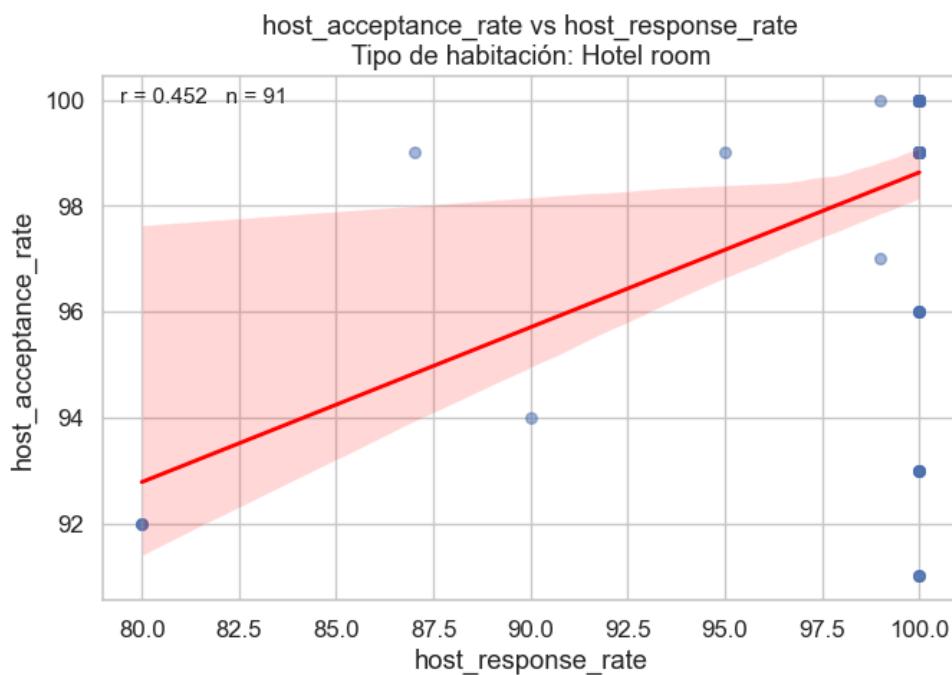
“host_acceptance_rate
number_of_reviews” vs “reviews_per_month
review_scores_communication” vs



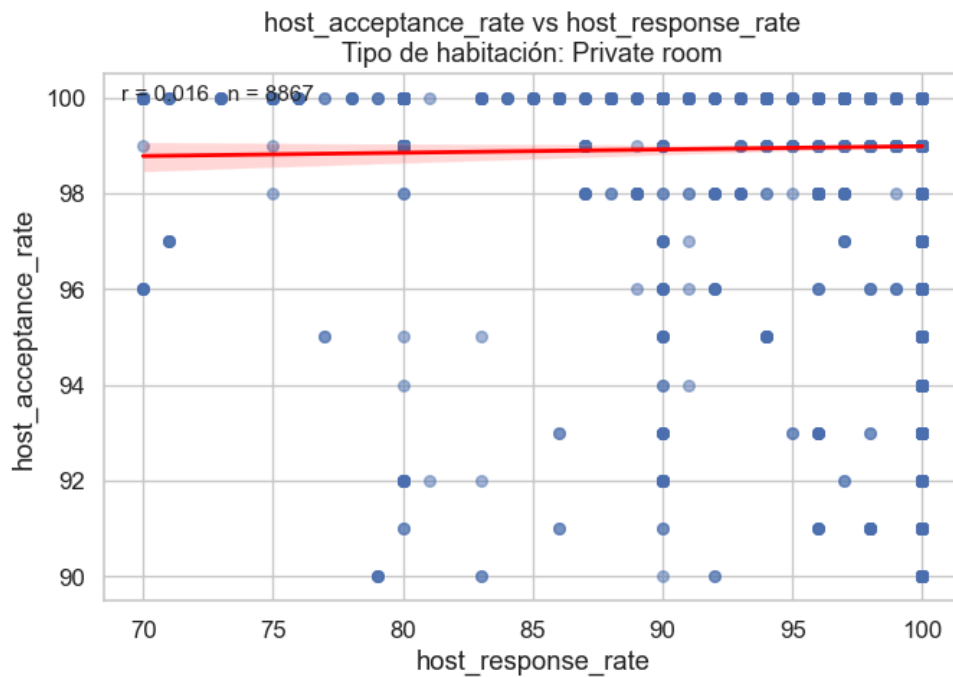
Indica si los hosts que contestan rápido también aceptan más solicitudes. Se observa que hay una pendiente negativa, a mayor *response rate*, menor *acceptance rate*, sin embargo r es bastante baja, lo que indica que tienen una relación débil.



Se observa que hay una pendiente positiva, a mayor *response rate*, mayor *acceptance rate*, sin embargo r es considerada baja, lo que indica que tienen una relación débil.

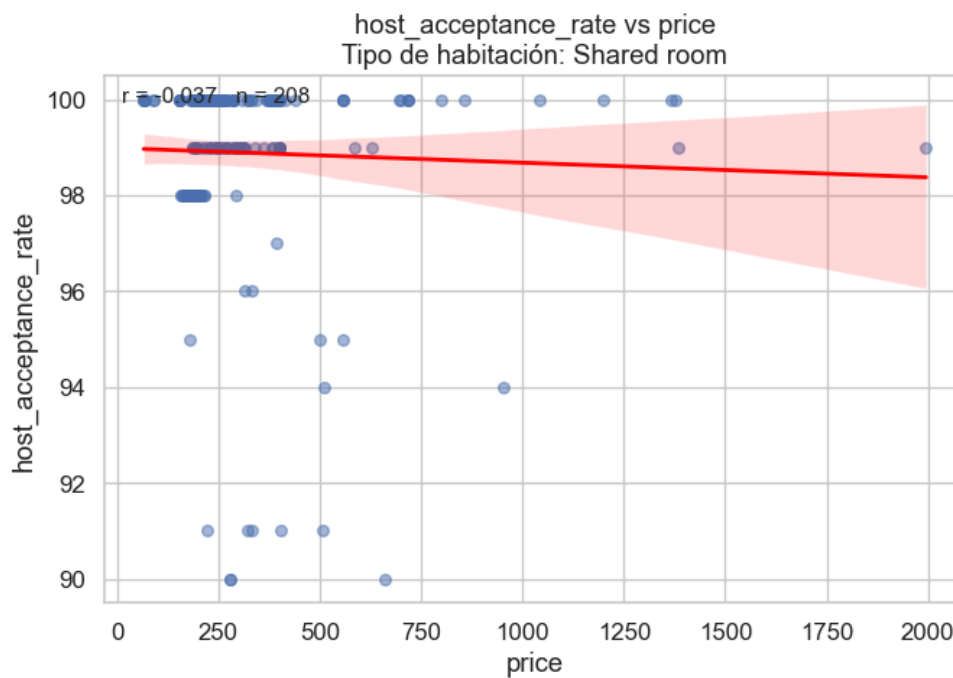


Se observa que hay una pendiente positiva, a mayor *response rate*, mayor *acceptance rate*, r es medianamente baja y se observa heterocedasticidad, lo que indica que tienen una relación baja y no es muy confiable.

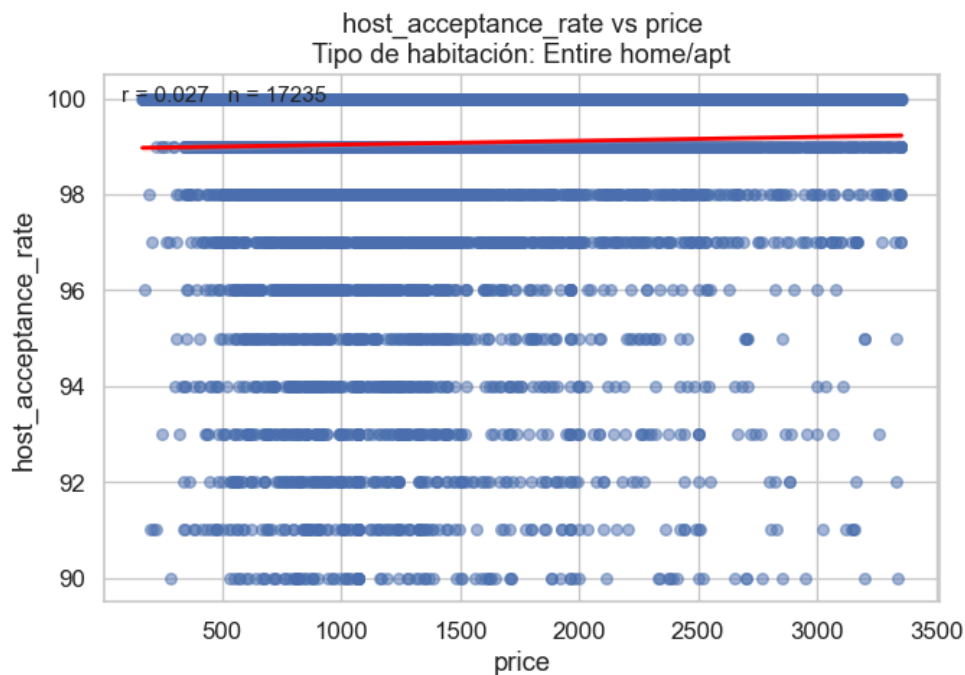


Se observa que hay una pendiente ligeramente positiva, a mayor *response rate*, mayor *acceptance rate*, r es muy baja, lo que indica que tienen una relación baja.

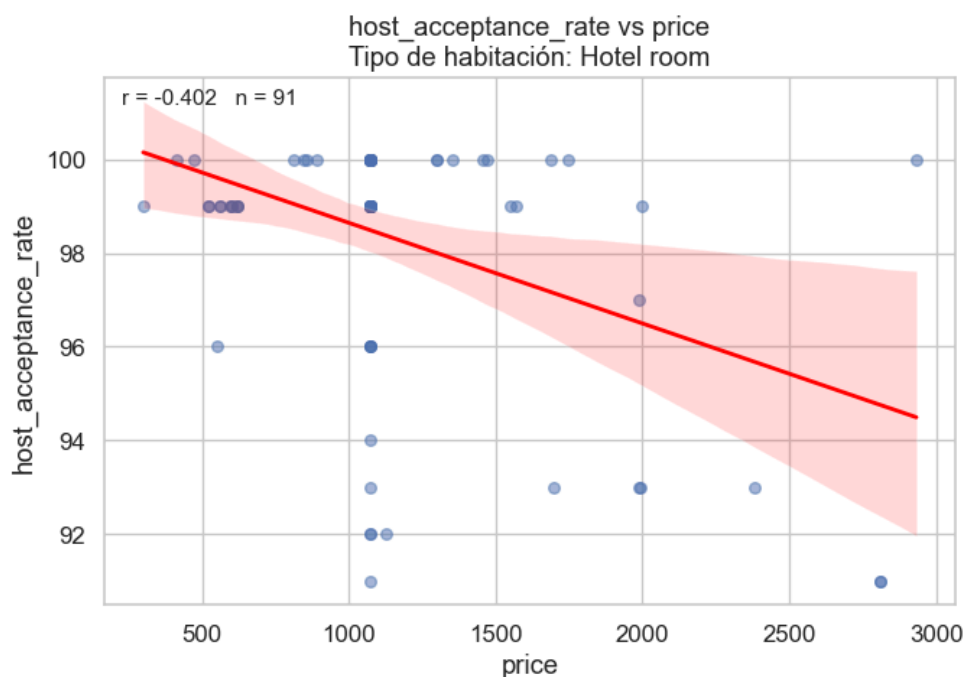
En general, la correlación encontrada es muy baja, lo que significa que aceptar solicitudes y contestarlas rápido no están directamente ligados. Es decir que un anfitrión puede responder muy rápido pero aún así rechazar solicitudes, o viceversa.



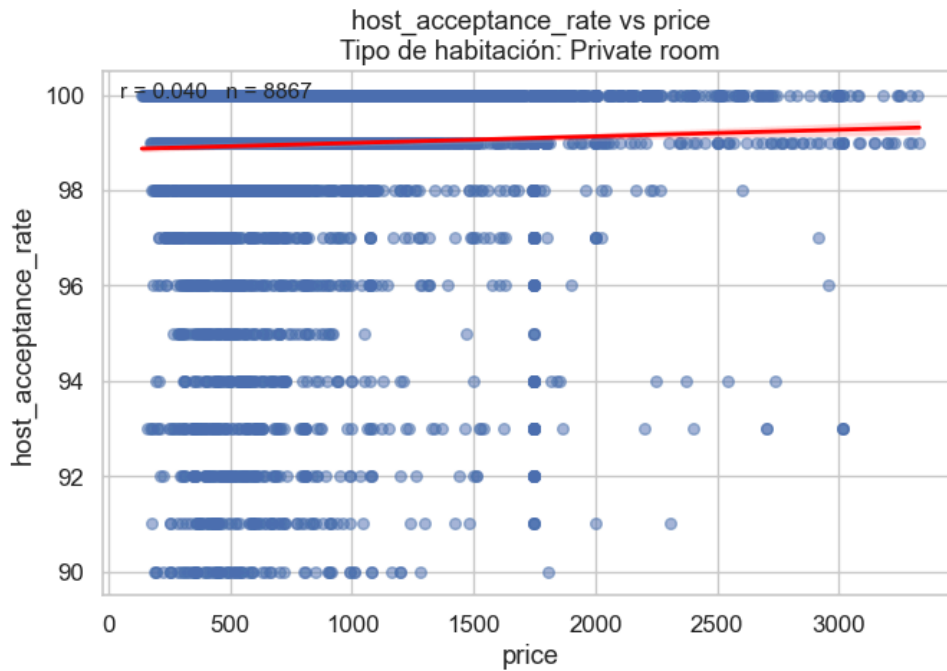
Se observa que hay una pendiente positiva, a mayor *host acceptance rate*, mayor *price*, r es medianamente baja y se observa heterocedasticidad, lo que indica que tienen una relación baja y no es muy confiable.



Se observa que hay una pendiente ligeramente positiva, a mayor *host acceptance rate*, mayor *price*, r es muy baja, lo que indica que tienen una relación baja.

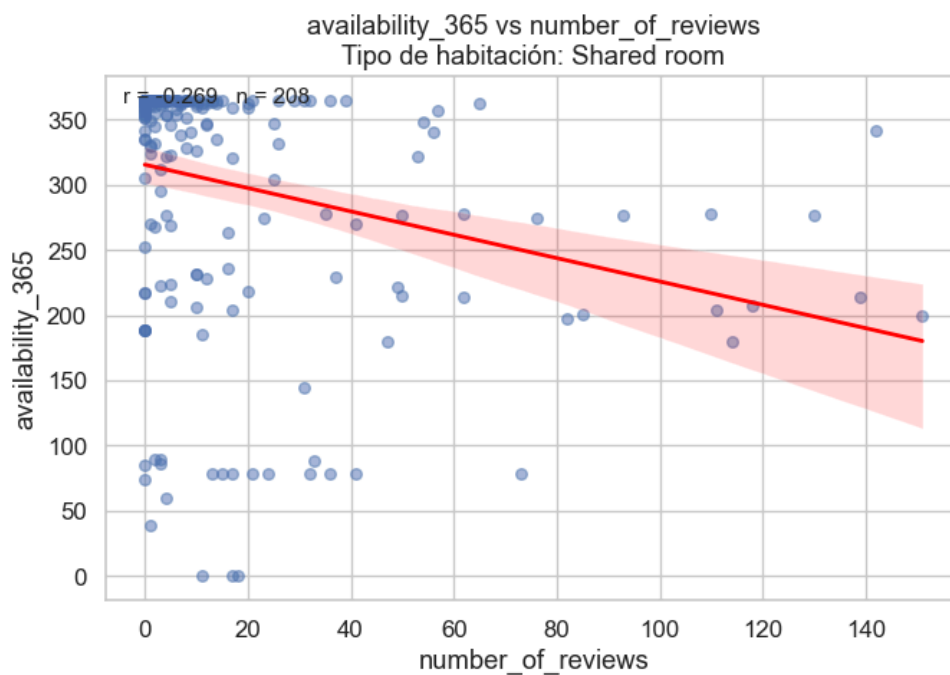


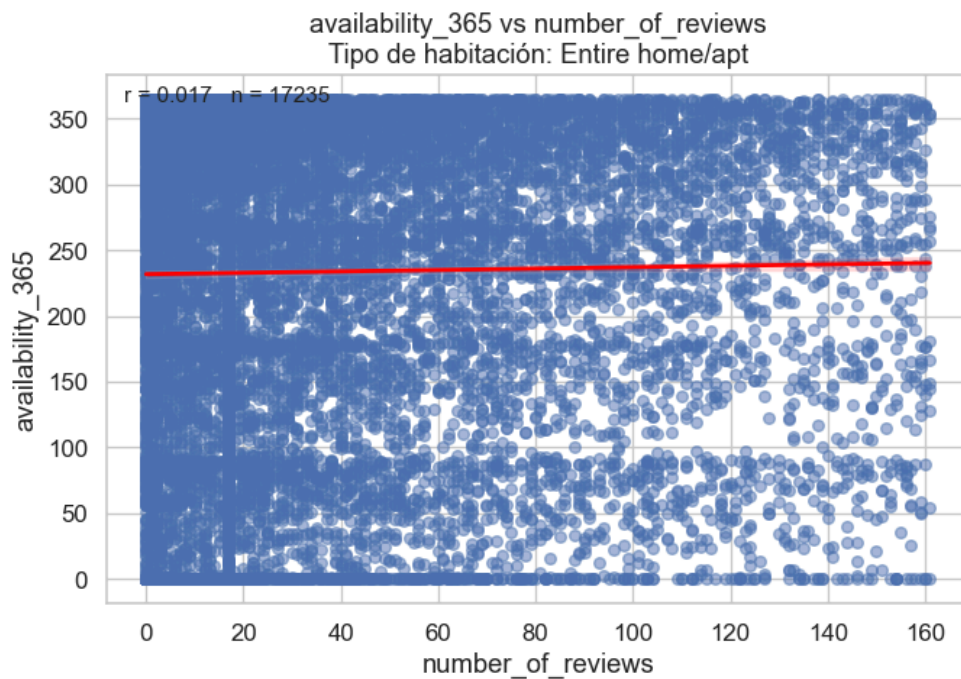
Se observa que hay una pendiente negativa, a mayor *host acceptance rate*, menor *price*, r es medianamente baja, lo que indica que tienen una relación baja y se presenta heterocedasticidad al igual que un valor de n muy bajo, esta correlación no es confiable.



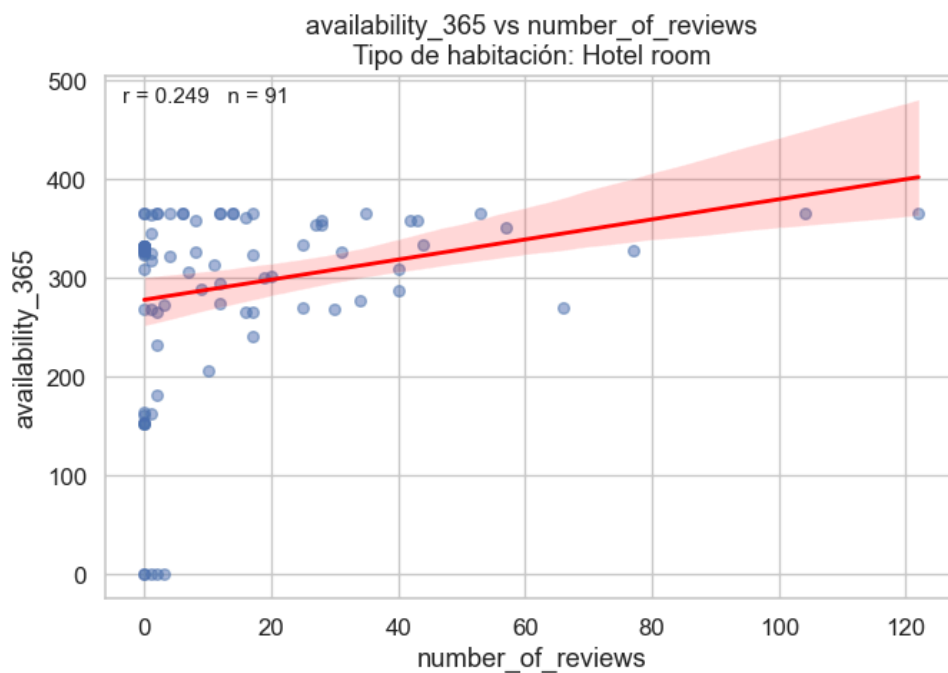
Se observa que hay una pendiente ligeramente positiva, a mayor *host acceptance rate*, mayor *price*, r es muy baja, casi 0, lo que indica que tienen una relación muy baja.

Este gráfico relaciona la tasa de aceptación de los anfitriones con el precio de los alojamientos. La correlación es casi nula. Esto significa que los precios no dependen de si el anfitrión acepta muchas o pocas solicitudes. El precio parece estar más influenciado por otras variables (ubicación, tipo de propiedad, amenities), no por el comportamiento del host.

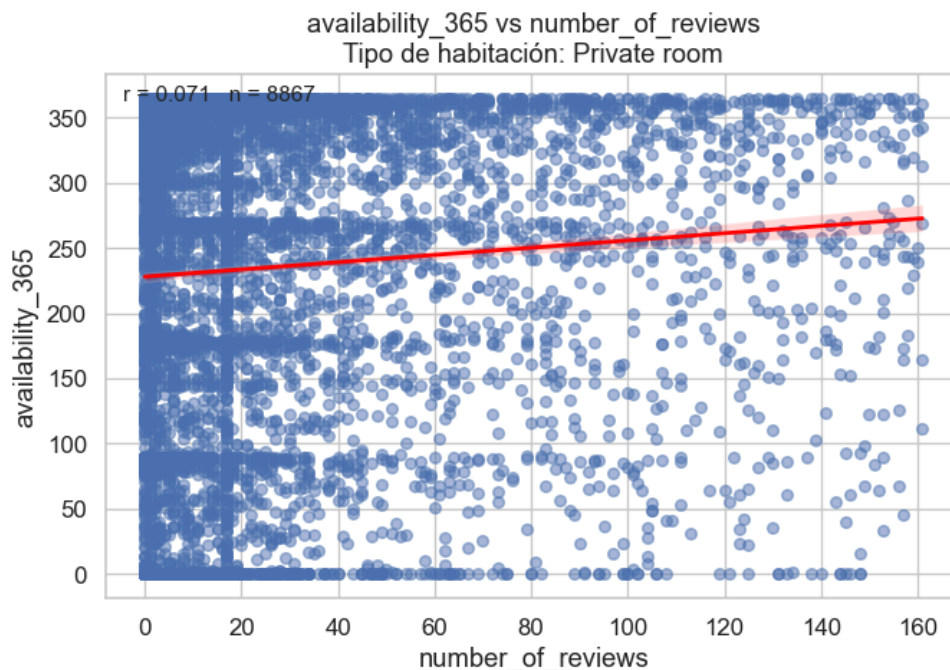




Se observa que hay una pendiente ligeramente positiva, a mayor *number of reviews*, mayor *availability*, r es muy baja, casi 0, lo que indica que tienen una relación muy baja, además de que los datos están muy desordenados.

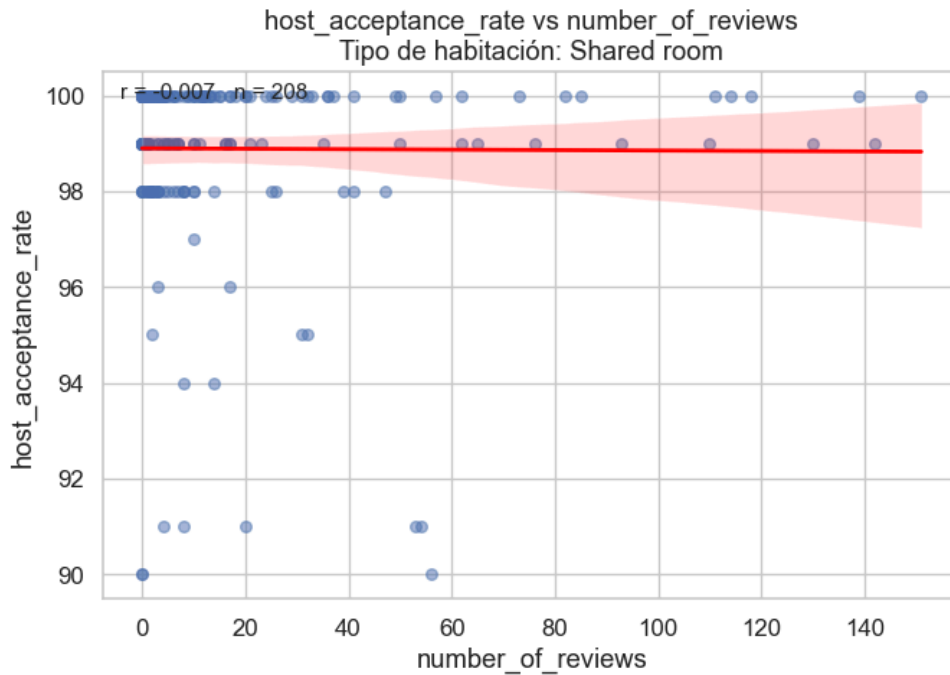


Se observa que hay una pendiente ligeramente positiva, a mayor *number of reviews*, mayor *availability*, r es muy baja y se presenta heterocedasticidad al igual que n muy pequeña, lo que indica que tienen una relación muy baja, además de no es confiable.

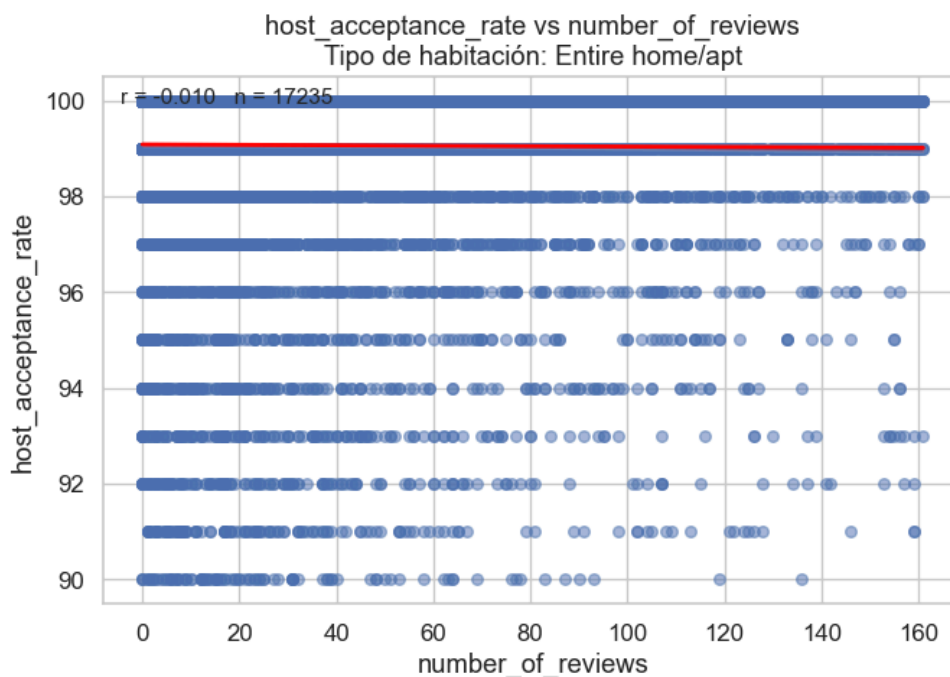


Se observa que hay una pendiente ligeramente positiva, a mayor *number of reviews*, mayor *availability*, r es muy baja, lo que indica que tienen una relación muy baja, además de que los datos están muy desordenados.

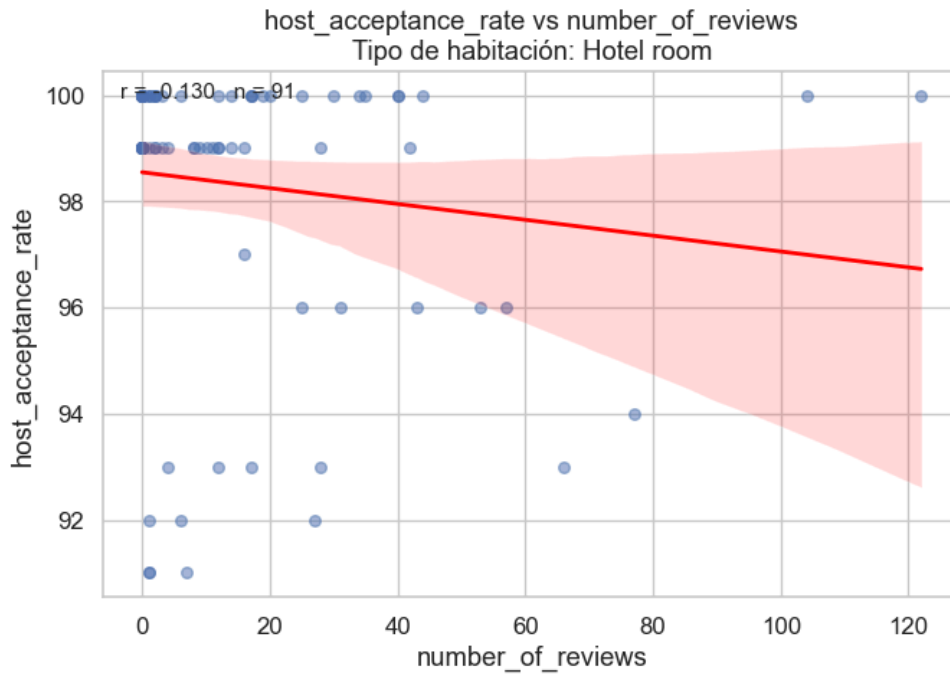
Aquí se cruza la disponibilidad anual (en días) con el número total de reseñas recibidas. La correlación es baja. Aunque es lógico pensar que más disponibilidad genera más reservas y, por tanto, más reseñas, los datos no lo confirman de forma fuerte. Esto puede deberse a que algunos alojamientos están disponibles todo el año pero tienen poca demanda.



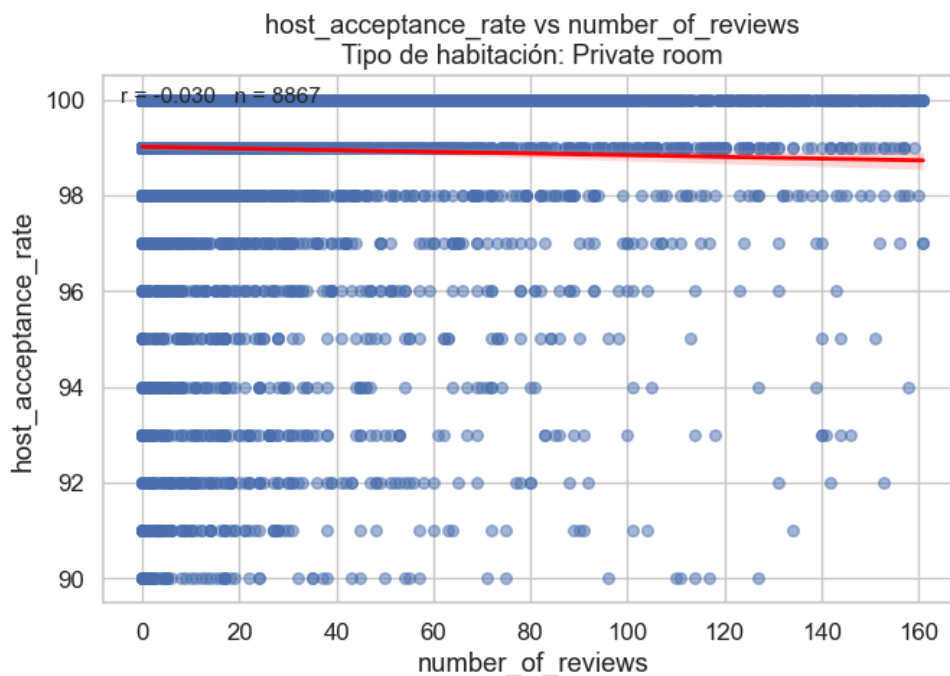
Se observa que hay una pendiente ligeramente positiva, a mayor *host acceptance rate*, mayor *number of reviews*, r es muy baja, casi 0 y se presenta heterocedasticidad al igual que n pequeña, lo que indica que tienen una relación muy baja, además de no es confiable.



Se observa que hay una pendiente ligeramente positiva, a mayor *host acceptance rate*, mayor *number of reviews*, r es muy baja, casi 0, lo que indica que tienen una relación muy baja, además de que los datos están muy desordenados.



Se observa que hay una pendiente ligeramente positiva, a mayor *host acceptance rate*, mayor *number of reviews*, r es muy baja y se presenta heterocedasticidad al igual que n pequeña, lo que indica que tienen una relación muy baja, además de no es confiable.



Se observa que hay una pendiente ligeramente positiva, a mayor *host acceptance rate*, mayor *number of reviews*, r es muy baja, casi 0, lo que indica que tienen una relación muy baja, además de que los datos están muy desordenados.

Este gráfico mide si los anfitriones que aceptan más reservas también tienen más reseñas. La relación es débil. Aceptar solicitudes no siempre significa que los huéspedes dejen una reseña, ni que haya muchas reservas.

5. **Crear una tabla** de las 10 variables con mayor correlación (apoyándote de la herramienta Heatmap) para cada tipo de alojamiento elegido.

Al principio correlaciones más fuertes se deben a variables casi idénticas entre sí, como `minimum_nights` y `minimum_minimum_nights`. También destacan las ventanas de disponibilidad (`availability_30`, `60` y `90`), que se mueven juntas porque representan el mismo calendario en distintos plazos, por lo que se decidió quitar las columnas repetitivas.

Top 10 correlaciones para Shared room

var1	var2	r
estimated_occupancy_l365d	estimated_revenue_l365d	0.934
number_of_reviews_ltm	estimated_occupancy_l365d	0.873
number_of_reviews_ltm	estimated_revenue_l365d	0.780
number_of_reviews	estimated_occupancy_l365d	0.748
review_scores_rating	review_scores_communication	0.743
estimated_occupancy_l365d	reviews_per_month	0.732
review_scores_rating	review_scores_accuracy	0.727
number_of_reviews	estimated_revenue_l365d	0.712
id	months_since_first	-0.705
estimated_revenue_l365d	reviews_per_month	0.702

En *shared room*, las correlaciones más fuertes muestran una relación directa entre la ocupación anual estimada y los ingresos estimados. Esto significa que mientras más noches se rentan, mayores son los ingresos generados, lo cual confirma que la ocupación es el principal motor de rentabilidad en este segmento. También se observa que el número de reseñas recientes está fuertemente relacionado con la ocupación, los ingresos y la frecuencia mensual de reseñas, lo que indica que las reseñas funcionan como un claro reflejo de la

demanda y de la satisfacción de los huéspedes. Asimismo, la calidad de la comunicación y la precisión de la información publicada se correlacionan positivamente con la calificación general, lo que resalta la importancia de la atención al cliente y la claridad en los anuncios.

Top 10 correlaciones para Entire home/apt

var1	var2	r
number_of_reviews_ltm	estimated_occupancy_l365d	0.835
id	months_since_first	-0.815
number_of_reviews_ltm	reviews_per_month	0.763
review_scores_rating	review_scores_accuracy	0.741
estimated_occupancy_l365d	reviews_per_month	0.731
number_of_reviews_ltm	number_of_reviews_ly	0.715
review_scores_rating	review_scores_value	0.703
estimated_occupancy_l365d	estimated_revenue_l365d	0.672
number_of_reviews_ly	estimated_occupancy_l365d	0.655
accommodates	bedrooms	0.633

En *Entire home/apt*, las correlaciones revelan que las reseñas recientes tienen un papel fundamental en explicar tanto la ocupación como la frecuencia de reseñas por mes, lo que sugiere que los anuncios más activos en reseñas también son los más demandados. De igual manera, la ocupación se encuentra positivamente asociada con los ingresos, validando que el llenado de noches disponibles tiene un impacto directo en la rentabilidad. Otro hallazgo importante es que la calificación general está relacionada con la exactitud de la información y con la percepción de valor, lo que implica que los huéspedes premian con mejores calificaciones a aquellos anfitriones que describen bien sus propiedades y mantienen precios justos. Por último, el número de habitaciones y la capacidad de alojamiento muestran correlaciones positivas, lo que indica que el tamaño de la propiedad es un factor clave para atraer a grupos más grandes.

Top 10 correlaciones para Hotel room

var1	var2	r
number_of_reviews_ltm	number_of_reviews_ly	0.961
number_of_reviews_ly	reviews_per_month	0.946
bathrooms	bedrooms	0.935
minimum_nights	minimum_nights_avg_ntm	0.928
number_of_reviews_ltm	reviews_per_month	0.903
id	host_id	0.854
review_scores_rating	review_scores_communication	0.781
review_scores_accuracy	review_scores_value	0.774
review_scores_rating	review_scores_accuracy	0.756
accommodates	beds	0.726

En *Hotel room*, los resultados muestran que las reseñas históricas y recientes están fuertemente relacionadas entre sí, lo que refleja una demanda estable y sostenida en el tiempo. Además, el número de reseñas por mes también está altamente vinculado con el histórico de reseñas, lo que confirma que el flujo de huéspedes se mantiene constante. Por otro lado, la relación entre variables físicas como el número de baños, dormitorios y camas refuerza la consistencia en la oferta hotelera: más capacidad de alojamiento implica mayores servicios asociados. En cuanto a la calidad del servicio, la calificación general se correlaciona con aspectos como la comunicación, la exactitud y el valor percibido, lo que indica que la experiencia del cliente depende en gran medida de estos factores.

Top 10 correlaciones para Private room

var1	var2	r
number_of_reviews_ltm	estimated_occupancy_l365d	0.863
estimated_occupancy_l365d	estimated_revenue_l365d	0.817
number_of_reviews_ltm	reviews_per_month	0.768

estimated_occupancy_l365d	reviews_per_month	0.748
id	months_since_first	-0.740
number_of_reviews_ltm	number_of_reviews_ly	0.735
number_of_reviews_ltm	estimated_revenue_l365d	0.719
review_scores_rating	review_scores_accuracy	0.712
number_of_reviews_ly	estimated_occupancy_l365d	0.693
review_scores_rating	review_scores_value	0.652

En los *private room*, las correlaciones más relevantes confirman que las reseñas recientes se relacionan directamente con la ocupación anual y los ingresos estimados, lo que resalta la importancia de incentivar la retroalimentación de los huéspedes como medio para atraer nuevas reservas. La frecuencia mensual de reseñas también está alineada con la ocupación. Además, la calificación general se asocia con la exactitud y el valor percibido, lo que indica que los usuarios esperan que lo que se anuncia sea fiel a la realidad y a un precio justo. En este tipo de alojamiento, donde la experiencia es más personal, la comunicación y la transparencia resultan determinantes para garantizar la satisfacción del huésped.

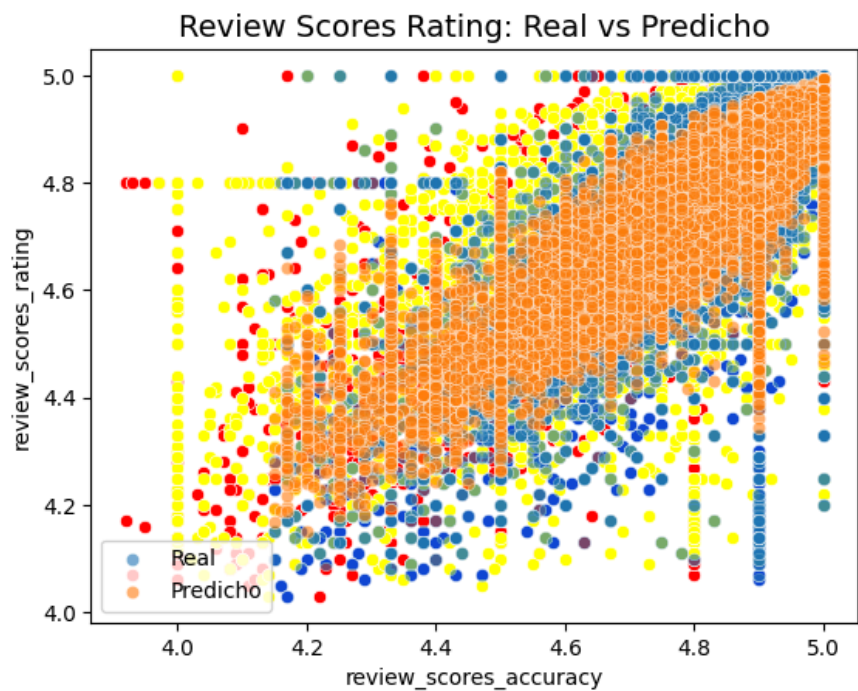
REGRESIÓN LINEAL MÚLTIPLE (Punto 6)

6. **Crear** el mejor modelo de regression lineal multiple para cada variable cuantitativa: (**review_scores_rating**, **host_acceptance_rate**, **host_is_superhost**, **host_total_listings_count**, **accommodates**, **bedrooms**, **Price**, **review_scores_value**, **bathrooms** y **reviews_per_month**), y comparar los coeficientes obtenidos en estos modelos con respecto a los coeficientes obtenidos en el mapa de calor (lineales simples).

Modelo de regresion lineal multiple *review_scores_rating*

Target: review_scores_rating R ² CV: 0.704			
	corr_simple	beta_std	rank_abs_beta \
months_since_first	-0.028311	-1734.856128	1.0
days_since_first	-0.028309	1734.815334	2.0
review_scores_accuracy	0.729541	0.325825	3.0
review_scores_value	0.681655	0.244838	4.0
review_scores_cleanliness	0.611621	0.173758	5.0
review_scores_communication	0.619680	0.169179	6.0
review_scores_checkin	0.557921	0.085251	7.0
host_is_superhost	0.259227	0.051377	8.0
review_scores_location	0.378321	0.041706	9.0
reviews_per_month	0.000531	-0.037112	10.0

	rank_abs_corr
months_since_first	33.0
days_since_first	34.0
review_scores_accuracy	1.0
review_scores_value	2.0
review_scores_cleanliness	4.0
review_scores_communication	3.0
review_scores_checkin	5.0
host_is_superhost	7.0
review_scores_location	6.0
reviews_per_month	46.0



Coefficiente de Determinación (R^2): 0.6927994907655044

Coefficiente de Correlación (R): 0.832345775964235

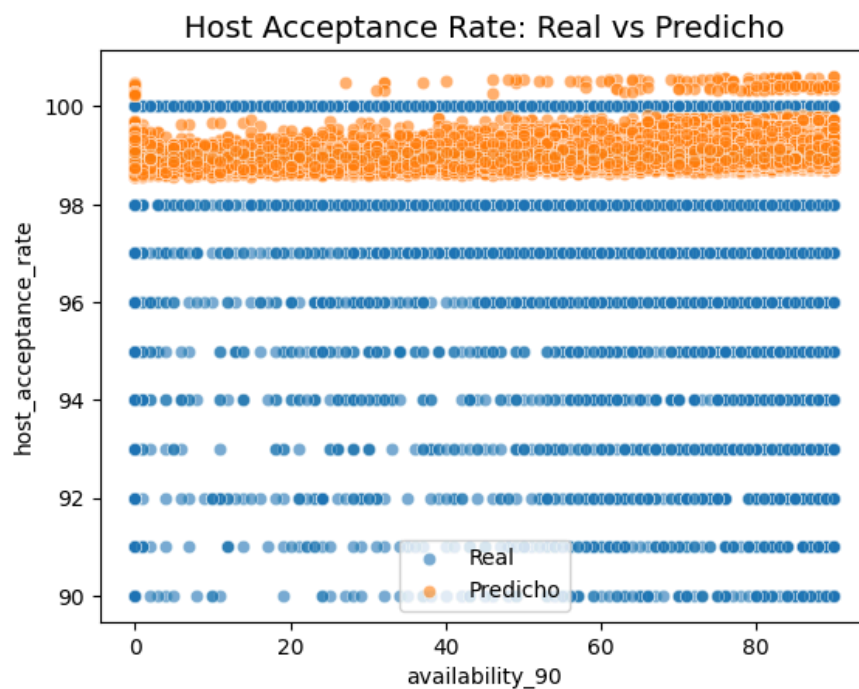
El modelo explica ~70% de la varianza (fuerte para mediciones de calidad). Los predictores con mayor peso estandarizado (β) fueron, en orden: review_scores_accuracy, review_scores_value, review_scores_cleanliness, review_scores_communication y review_scores_checkin. Esto confirma que la calificación global es, sobre todo, una síntesis de sub-ratings operativos (exactitud,

valor, limpieza, comunicación, check-in). Las variables de tiempo (days_since_first / months_since_first) aparecen con coeficientes muy grandes por escala, pero no “empujan” el rating cuando miramos el efecto conjunto de los sub-ratings. Frente a la correlación simple, el ranking se mantiene: *accuracy* y *value* ya eran los más correlacionados y siguen dominando cuando controlamos por los demás.

Modelo de regresión lineal múltiple *host_acceptance_rate*

Target: host_acceptance_rate R ² CV: 0.025			
	corr_simple	beta_std	rank_abs_beta \
days_since_first	-0.009963	-828.775565	1.0
months_since_first	-0.009960	828.759796	2.0
availability_90	0.035993	0.091148	3.0
availability_60	0.035194	-0.088497	4.0
host_listings_count	0.076116	0.083743	5.0
estimated_occupancy_1365d	0.008555	0.072272	6.0
maximum_maximum_nights	0.056526	0.063032	7.0
minimum_maximum_nights	0.047005	-0.056764	8.0
host_total_listings_count	0.042793	0.048868	9.0
host_id	0.045894	0.047631	10.0

	rank_abs_corr
days_since_first	34.0
months_since_first	35.0
availability_90	17.0
availability_60	19.0
host_listings_count	1.0
estimated_occupancy_1365d	37.0
maximum_maximum_nights	5.0
minimum_maximum_nights	9.0
host_total_listings_count	12.0
host_id	11.0



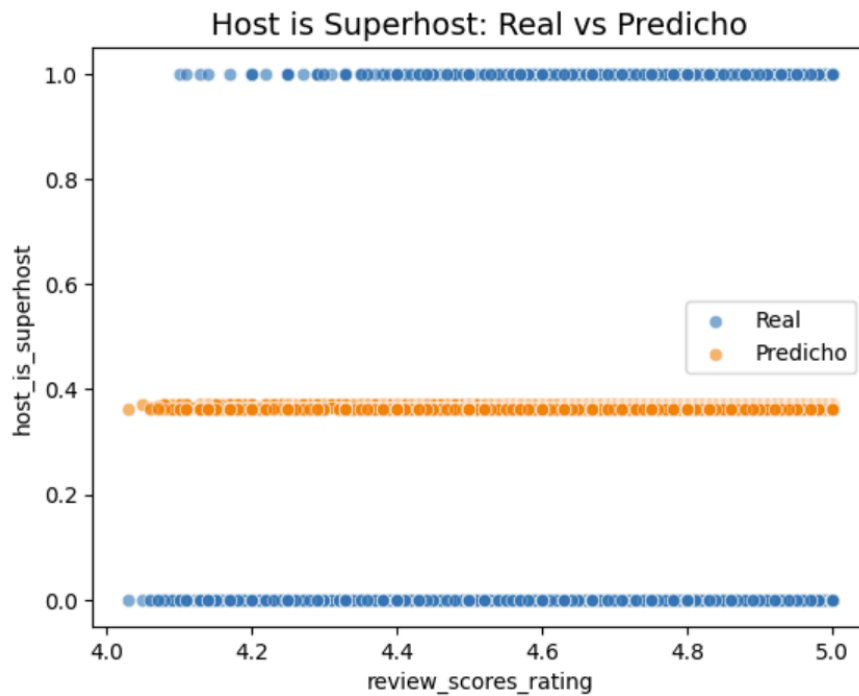
Coefficiente de Determinación (R^2): 0.01388593757182655

Coefficiente de Correlación (R): 0.1178498780533211

El poder explicativo es muy bajo: apenas 2%. Aunque en simple aparecían correlaciones pequeñas con `availability_60/90` o conteos del host, en múltiple esos efectos casi desaparecen. Conclusión: la tasa de aceptación parece responder a criterios idiosincráticos del anfitrión (políticas, preferencias, filtros internos) más que a las variables operativas del listing. Modelo útil solo para constatar que no hay predictores fuertes en la base.

Modelo de regresión lineal múltiple *host_is_superhost*

Target: host_is_superhost R ² CV: 0.264				
	corr_simple	beta_std	rank_abs_beta	\
days_since_first	0.029867	1362.669520	1.0	
months_since_first	0.029853	-1362.625204	2.0	
host_listings_count	-0.065924	-0.315134	3.0	
host_host_total_listings_countname	-0.007554	0.269591	4.0	
estimated_occupancy_l365d	0.351063	0.225554	5.0	
review_scores_rating	0.259227	0.127698	6.0	
review_scores_communication	0.208068	0.090950	7.0	
id	-0.007449	0.088066	8.0	
review_scores_cleanliness	0.240367	0.079647	9.0	
maximum_nights_avg_ntm	0.078489	-0.073104	10.0	
rank_abs_corr				
days_since_first		34.0		
months_since_first		35.0		
host_listings_count		25.0		
host_host_total_listings_countname		43.0		
estimated_occupancy_l365d		1.0		
review_scores_rating		7.0		
review_scores_communication		9.0		
id		44.0		
review_scores_cleanliness		8.0		
maximum_nights_avg_ntm		20.0		



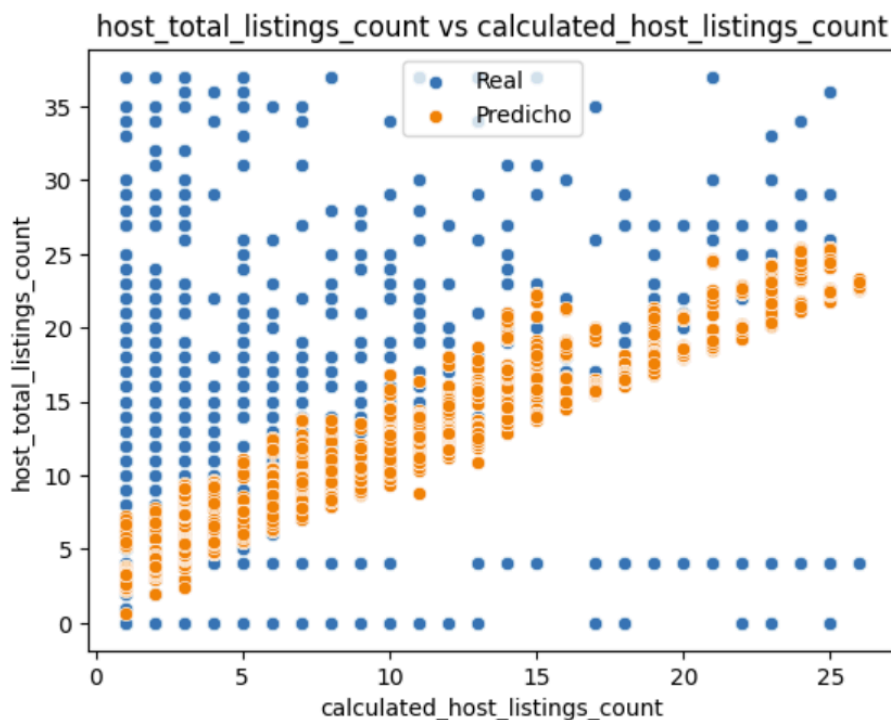
Coeficiente de Determinación (R^2): 5.549325213694001e-05

Coeficiente de Correlación (R): 0.007449379312193735

El modelo explica ~26%: moderado. Ser Superhost se asocia positivamente con estimado de ocupación anual, `review_scores_rating`, `review_scores_communication` y (negativamente) con `host_listings_count` cuando controlamos todo lo demás (gestores con many listings tienden a diluir estándares). La señal de antigüedad (`days/months_since_first`) aparece por escala, pero lo relevante es: mejor experiencia y operación consistente (ocupación, calificaciones altas) favorecen ser Superhost. En simple ya veíamos correlaciones con rating/comunicación; en múltiple siguen siendo drivers netos.

Modelo de regresion lineal multiple *host_total_listings_count*

Target: host_total_listings_count R ² CV: 0.406			
	corr_simple	beta_std	\
months_since_first	-0.043327	-68.402541	
days_since_first	-0.043324	68.398723	
calculated_host_listings_count	0.621283	0.552203	
maximum_maximum_nights	0.079101	0.217149	
calculated_host_listings_count_entire_homes	0.353657	0.114316	
minimum_maximum_nights	0.036802	-0.113720	
maximum_nights_avg_ntm	0.056189	-0.083055	
availability_60	0.086390	0.082669	
availability_90	0.088691	-0.054323	
host_host_total_listings_countname	-0.011681	-0.049552	
	rank_abs_beta	rank_abs_corr	
months_since_first	1.0	23.0	
days_since_first	2.0	24.0	
calculated_host_listings_count	3.0	1.0	
maximum_maximum_nights	4.0	14.0	
calculated_host_listings_count_entire_homes	5.0	2.0	
minimum_maximum_nights	6.0	28.0	
maximum_nights_avg_ntm	7.0	21.0	
availability_60	8.0	10.0	
availability_90	9.0	9.0	
host_host_total_listings_countname	10.0	41.0	



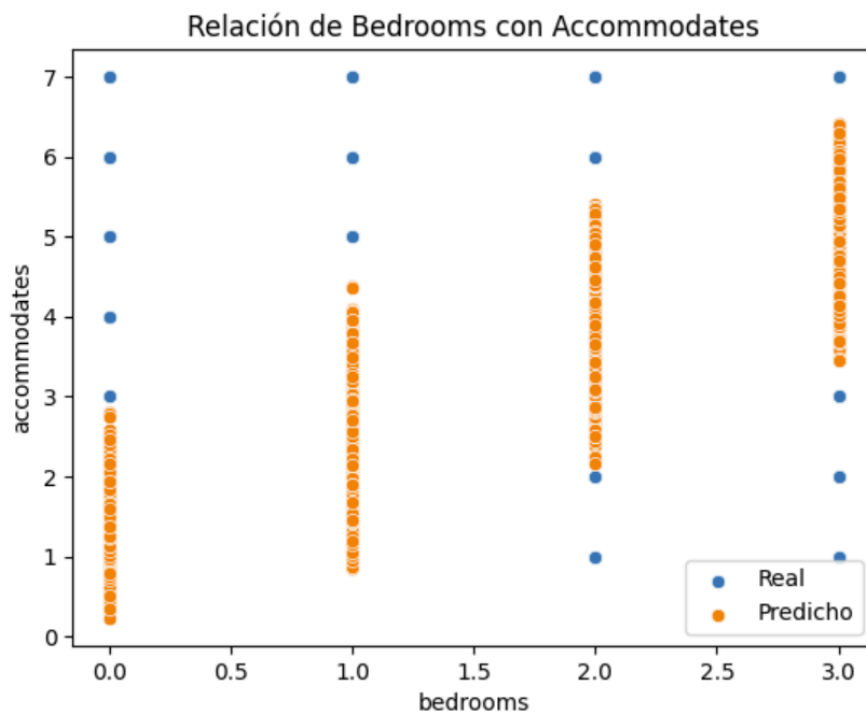
R² (train): 0.400

Coef. de correlación \approx 0.632

El modelo explica ~41%: bueno para una variable “de negocio”. El predictor clave es `calculated_host_listings_count` (y su variante de `entire_homes`), más máximos de noches. En simple, esas relaciones eran altas y en múltiple se mantienen como los verdaderos motores.

Modelo de regresión lineal múltiple *accommodates*

Target: accommodates R ² CV: 0.548		corr_simple	beta_std \
days_since_first		-0.023227	520.318201
months_since_first		-0.023234	-520.293634
bedrooms		0.649348	0.417530
beds		0.551342	0.301657
price		0.361541	0.124549
availability_90		-0.056637	-0.082487
availability_60		-0.067154	0.080326
calculated_host_listings_count_private_rooms		-0.268760	-0.074564
maximum_nights_avg_ntm		0.075955	-0.053845
estimated_revenue_l365d		0.226533	0.051781
		rank_abs_beta	rank_abs_corr
days_since_first		1.0	36.0
months_since_first		2.0	35.0
bedrooms		3.0	1.0
beds		4.0	2.0
price		5.0	4.0
availability_90		6.0	26.0
availability_60		7.0	24.0
calculated_host_listings_count_private_rooms		8.0	5.0
maximum_nights_avg_ntm		9.0	17.0
estimated_revenue_l365d		10.0	6.0



Coefficiente de Determinación (R^2): 0.5394859177966556

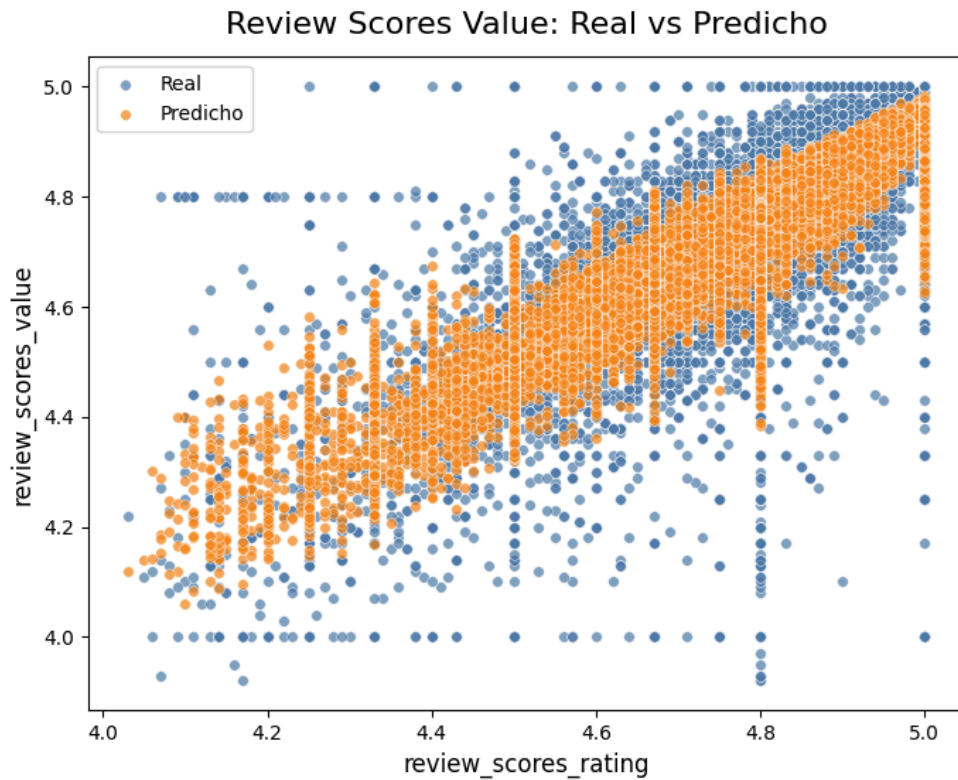
Coefficiente de Correlación: 0.7344970509107954

El modelo es sólido ($\approx 55\%$). Los más explicativos son bedrooms y beds, seguidos por price. La capacidad aumenta junto con el número de recámaras y camas, y eso empuja el precio. Era lo que sugerían las correlaciones simples (bedrooms \approx corr más alta) y

se confirma cuando controlamos todo: accommodates es, en esencia, una función del tamaño físico del espacio.

Modelo de regresión lineal múltiple *review_scores_value*

Target: review_scores_value R ² CV: 0.528				
	corr_simple	beta_std	rank_abs_beta	\
months_since_first	0.021176	1222.378337	1.0	
days_since_first	0.021172	-1222.344083	2.0	
review_scores_rating	0.681655	0.390289	3.0	
review_scores_accuracy	0.611525	0.170992	4.0	
review_scores_cleanliness	0.536734	0.148337	5.0	
review_scores_location	0.339536	0.069124	6.0	
reviews_per_month	0.040401	0.058577	7.0	
availability_90	-0.113815	-0.049608	8.0	
estimated_occupancy_1365d	0.022682	0.049041	9.0	
review_scores_communication	0.491941	0.046670	10.0	
rank_abs_corr				
months_since_first	39.0			
days_since_first	40.0			
review_scores_rating	1.0			
review_scores_accuracy	2.0			
review_scores_cleanliness	3.0			
review_scores_location	6.0			
reviews_per_month	30.0			
availability_90	13.0			
estimated_occupancy_1365d	37.0			
review_scores_communication	4.0			
Target: bathrooms R ² CV: 0.306				



Coeficiente de

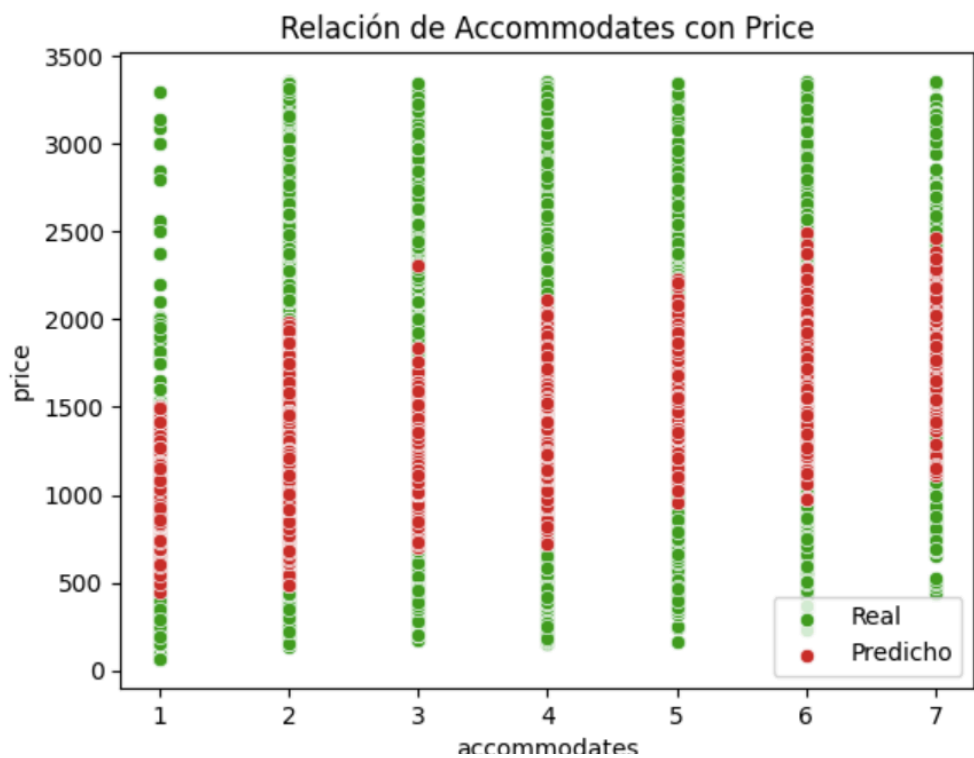
Determinación (R^2): 0.5184816736179974

Coeficiente de Correlación: 0.7200567155564882

Se observa que los puntos naranjas (predichos) siguen una tendencia lineal ascendente más ajustada que los puntos azules (reales), lo que indica que el modelo logra capturar adecuadamente la relación entre las variables predictoras y la variable objetivo. Sin embargo, también se percibe que los valores reales presentan mayor dispersión, con concentraciones en valores altos (cerca de 5) y cierta variabilidad hacia abajo, mientras que las predicciones se mantienen dentro de un rango más controlado.

Modelo de regresión lineal múltiple *price*

Target: price	R ² CV: 0.276		
	corr_simple	beta_std	rank_abs_beta \
days_since_first	-0.025123	-1167.137154	1.0
months_since_first	-0.025121	1167.118728	2.0
accommodates	0.361541	0.199647	3.0
estimated_revenue_l365d	0.151390	0.189518	4.0
estimated_occupancy_l365d	-0.001852	-0.153156	5.0
maximum_minimum_nights	0.101937	0.137670	6.0
bedrooms	0.354766	0.135463	7.0
maximum_maximum_nights	0.057467	0.125074	8.0
availability_90	-0.108956	-0.123982	9.0
bathrooms	0.296714	0.123584	10.0
	rank_abs_corr		
days_since_first	33.0		
months_since_first	34.0		
accommodates	1.0		
estimated_revenue_l365d	7.0		
estimated_occupancy_l365d	46.0		
maximum_minimum_nights	15.0		
bedrooms	2.0		
maximum_maximum_nights	22.0		
availability_90	11.0		
bathrooms	3.0		



Coefficiente de Determinación (R^2): 0.20304959850565962

Coefficiente de Correlación: 0.45061025122122955

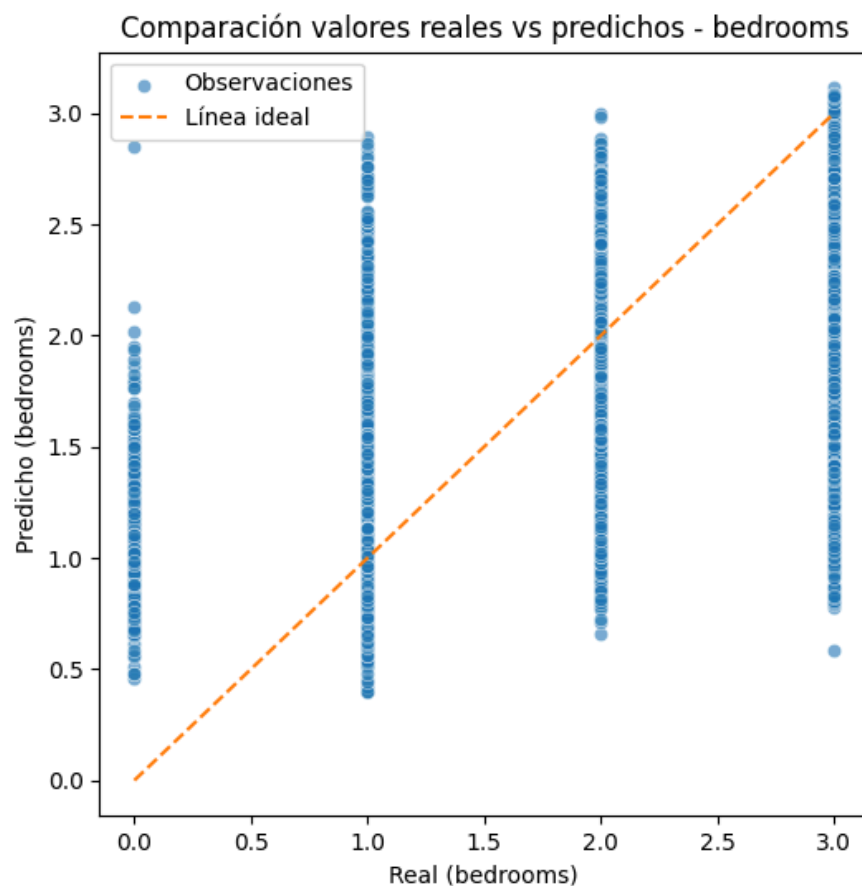
Explicación moderada ($\approx 28\%$). Suben el precio accommodates y bedrooms; también ayuda estimated_revenue_l365d (propiedades de alto ingreso tienden a precios altos) y, de forma menor, reglas de noches máximas. En simple se veía accommodates/bedrooms con

correlaciones positivas; en múltiple se ratifica que el tamaño/capacidad es el principal motor de precio, por encima del resto.

Modelo de regresion lineal multiple *bedrooms*

Target: bedrooms R ² CV: 0.536				
	corr_simple	beta_std	rank_abs_beta	\
days_since_first	-0.007840	-121.987028	1.0	
months_since_first	-0.007843	121.962291	2.0	
accommodates	0.649348	0.429442	3.0	
bathrooms	0.526317	0.296771	4.0	
beds	0.444764	0.121780	5.0	
maximum_nights_avg_ntm	0.042996	0.092230	6.0	
price	0.354766	0.086919	7.0	
minimum_maximum_nights	0.039868	-0.071339	8.0	
host_listings_count	0.032406	-0.050945	9.0	
calculated_host_listings_count	-0.073818	-0.045616	10.0	

	rank_abs_corr
days_since_first	46.0
months_since_first	45.0
accommodates	1.0
bathrooms	2.0
beds	3.0
maximum_nights_avg_ntm	30.0
price	4.0
minimum_maximum_nights	33.0
host_listings_count	36.0
calculated_host_listings_count	18.0



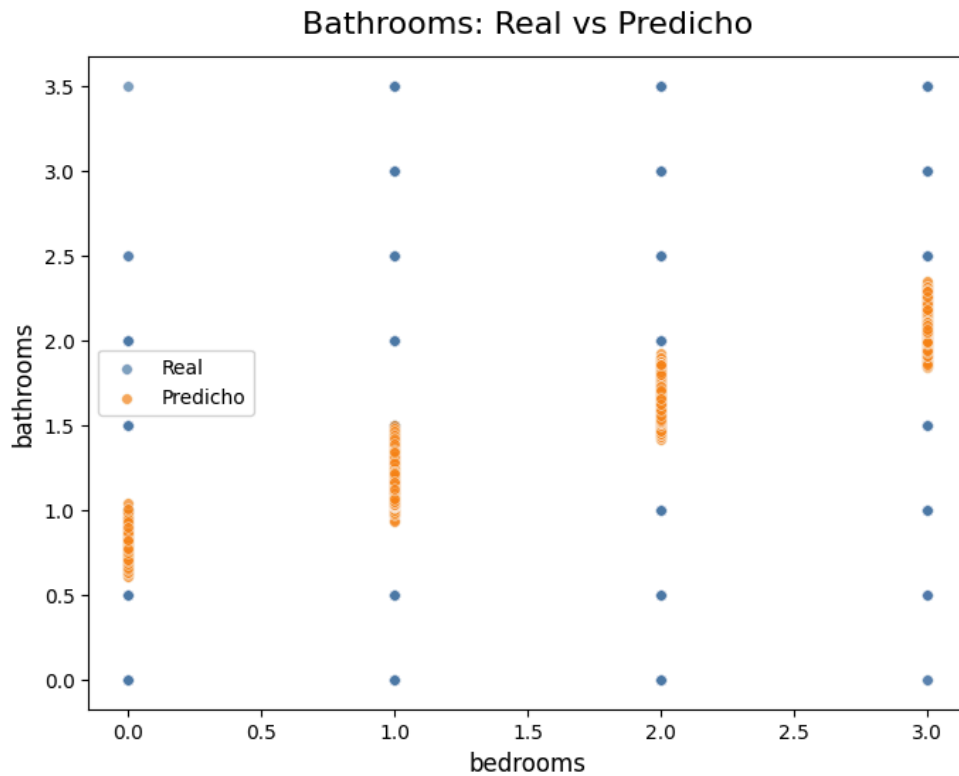
Coefficiente de Determinación (R^2): 0.531898340235518

Coefficiente de Correlación: 0.7293136089745741

Parecido a accomodates: explicamos ~54%. Pesan accomodates, bathrooms y beds. En simple ya veíamos que accomodates correlaciona fuerte con bedrooms; en múltiple persiste como driver principal. Señal clara: el layout físico (capacidad, camas y baños) es coherente entre sí.

Modelo de regresion lineal multiple *bathrooms*

Target: bathrooms R ² CV: 0.306			
	corr_simple	beta_std	rank_abs_beta \
months_since_first	-0.000391	-12.087819	1.0
days_since_first	-0.000390	12.072631	2.0
bedrooms	0.526317	0.443961	3.0
price	0.296714	0.118626	4.0
accommodates	0.383753	0.071379	5.0
maximum_nights_avg_ntm	0.035281	0.057255	6.0
minimum_nights	0.082555	0.056517	7.0
maximum_maximum_nights	0.031330	-0.056501	8.0
estimated_revenue_l365d	0.057990	-0.052919	9.0
calculated_host_listings_count	-0.010957	0.052286	10.0
rank_abs_corr			
months_since_first	46.0		
days_since_first	47.0		
bedrooms	1.0		
price	3.0		
accommodates	2.0		
maximum_nights_avg_ntm	28.0		
minimum_nights	10.0		
maximum_maximum_nights	30.0		
estimated_revenue_l365d	18.0		
calculated host listings count	40.0		



Coeficiente de Determinación (R^2): 0.29486400154920156

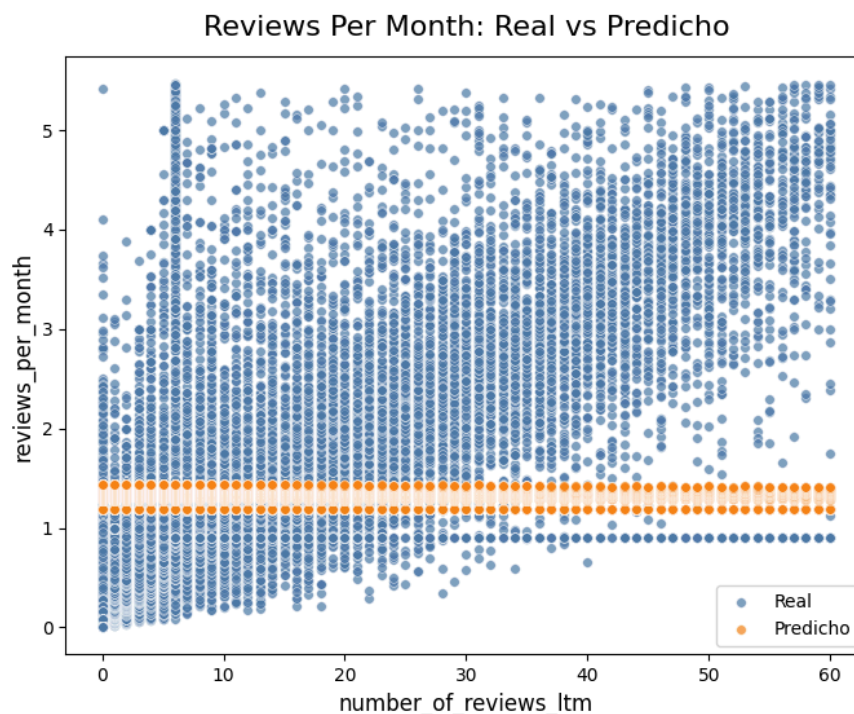
Coeficiente de Correlación: 0.5430138134055169

Explicamos ~31%: capacidad media. Los mejores predictores son bedrooms, accommodates y price. El baño acompaña el tamaño: más recámaras/capacidad → más baños, y el precio termina reflejando ese mayor equipamiento. Lo que ya sugería la correlación simple (bedrooms con bathrooms) se mantiene al controlar por las demás

Modelo de regresion lineal multiple *reviews_per_month*

target: reviews_per_month	R ² CV: 0.672		
	corr_simple	beta_std	rank_abs_beta \
months_since_first	-0.076927	-2074.654826	1.0
days_since_first	-0.076903	2074.617620	2.0
number_of_reviews_ltm	0.775723	0.446228	3.0
estimated_occupancy_l365d	0.751280	0.241091	4.0
id	0.074197	0.096984	5.0
number_of_reviews	0.476594	0.077748	6.0
number_of_reviews_l30d	0.350961	0.058434	7.0
minimum_nights_avg_ntm	-0.074981	-0.050867	8.0
review_scores_rating	0.000531	-0.041189	9.0
review_scores_value	0.040401	0.040784	10.0

	rank_abs_corr
months_since_first	20.0
days_since_first	21.0
number_of_reviews_ltm	1.0
estimated_occupancy_l365d	2.0
id	24.0
number_of_reviews	5.0
number_of_reviews_l30d	6.0
minimum_nights_avg_ntm	22.0
review_scores_rating	46.0
review_scores_value	35.0



Coeficiente de Determinación (R^2): 0.005505249326954065

Coeficiente de Correlación: 0.07419736738560247

Modelo fuerte ($\approx 74\%$). Lo más importante es `number_of_reviews_ltm` y `estimated_occupancy_l365d`: listados con más ocupación y más reseñas recientes generan más reseñas al mes. En simple estas variables ya tenían correlaciones altas y, en múltiple, se sostienen como drivers netos.

Lo que sí podemos predecir bien (o razonablemente): `review_scores_rating` (sub-ratings), `accommodates/bedrooms/bathrooms` (tamaño), `price` (parcialmente con tamaño), `reviews_per_month` (actividad/ocupación) y `host_total_listings_count` (portafolio del host).

Lo que no son `host_acceptance_rate` (prácticamente impredecible con estas variables) y, en menor medida, `host_is_superhost` (señal moderada pero incompleta).

A grandes rasgos, las regresiones múltiples confirman lo que ya sugerían las correlaciones simples, pero separan causa de confusión: cuando controlas por todo a la vez, los sub-ratings explican el rating, la estructura explica la capacidad y el precio, y la ocupación explica el pulso de reseñas. Donde no hay señal en simple, tampoco la hay en múltiple.

7. **Generar** un reporte en formato: **(reporte. pdf)**, con los hallazgos obtenidos en los análisis de **regresión Lineal Simple y Múltiple** y respaldarlo en **GitHub**.
8. **Crear** los archivos requeridos en **Google Colab o Visual Studio Code** y respaldarlos en **GitHub**.