



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

Correlaciones Entire home

	Variable 1	Variable 2	Correlación
0	host_acceptance_rate	host_response_rate	0.037175
1	review_scores_rating	calculated_host_listings_count	-0.105492
2	host_acceptance_rate	price	0.027415
3	availability_365	number_of_reviews	0.017335
4	host_acceptance_rate	number_of_reviews	-0.009731
5	reviews_per_month	review_scores_communication	-0.010648

Correlaciones Private Room

	Variable 1	Variable 2	Correlación
0	host_acceptance_rate	host_response_rate	0.016158
1	review_scores_rating	calculated_host_listings_count	-0.210086
2	host_acceptance_rate	price	0.039760
3	availability_365	number_of_reviews	0.070687
4	host_acceptance_rate	number_of_reviews	-0.030418
5	reviews_per_month	review_scores_communication	-0.083841

Correlaciones Hotel room

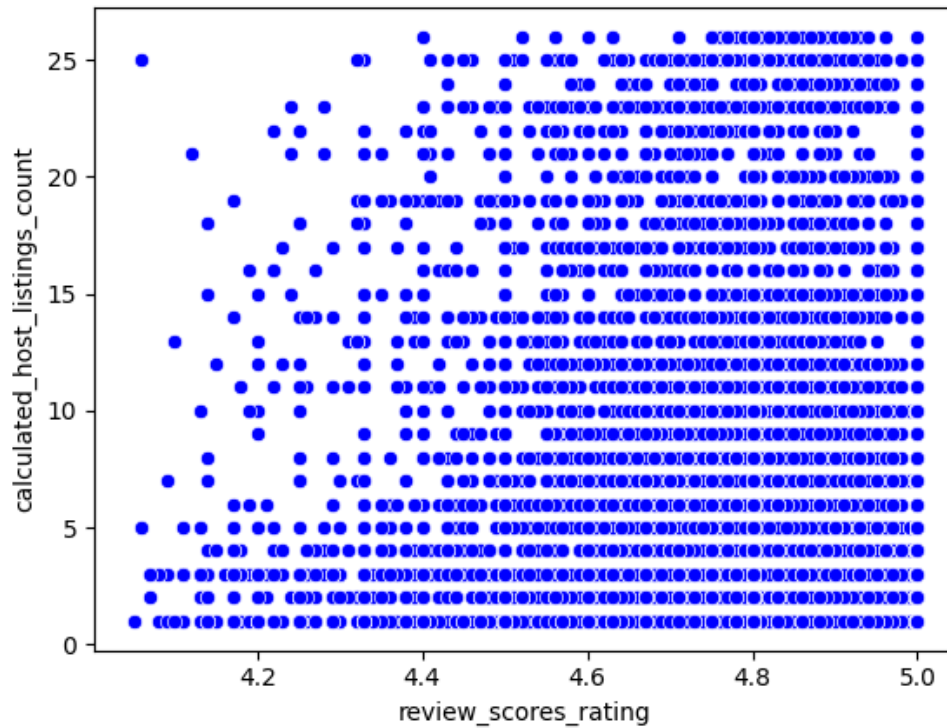
	Variable 1	Variable 2	Correlación
0	host_acceptance_rate	host_response_rate	0.451653
1	review_scores_rating	calculated_host_listings_count	-0.067108
2	host_acceptance_rate	price	-0.401594
3	availability_365	number_of_reviews	0.248788
4	host_acceptance_rate	number_of_reviews	-0.130408
5	reviews_per_month	review_scores_communication	-0.063587

Correlaciones Shared room

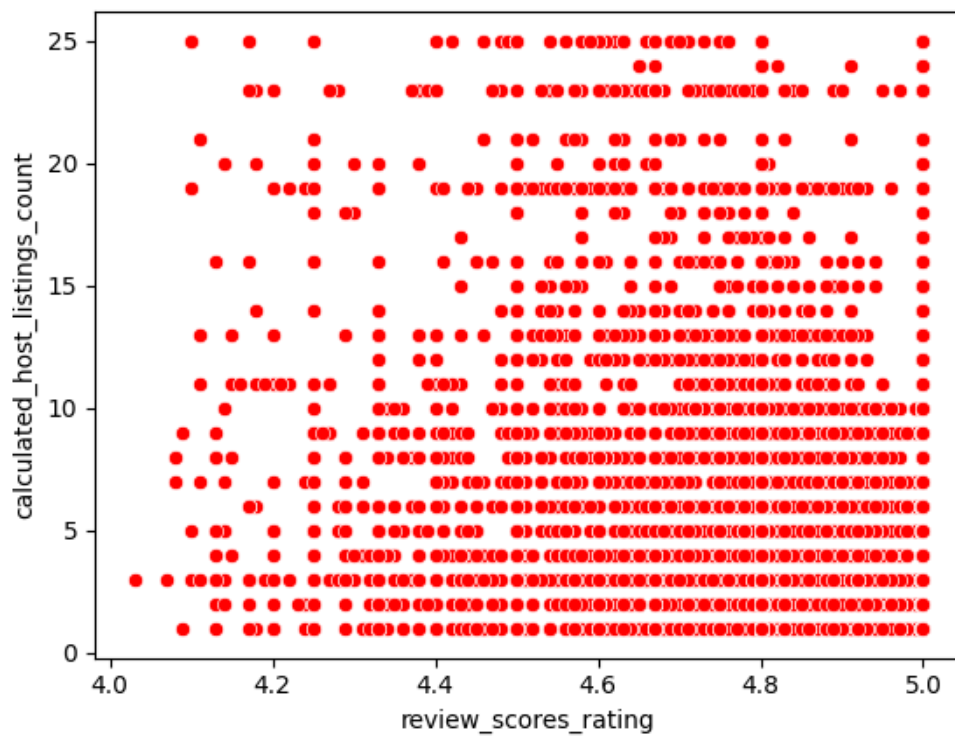
	Variable 1	Variable 2	Correlación
0	host_acceptance_rate	host_response_rate	-0.013882
1	review_scores_rating	calculated_host_listings_count	-0.065459
2	host_acceptance_rate	price	-0.036633
3	availability_365	number_of_reviews	-0.268849
4	host_acceptance_rate	number_of_reviews	-0.006667
5	reviews_per_month	review_scores_communication	-0.178008

Gráficos de variables con mayor correlación por tipo de habitación

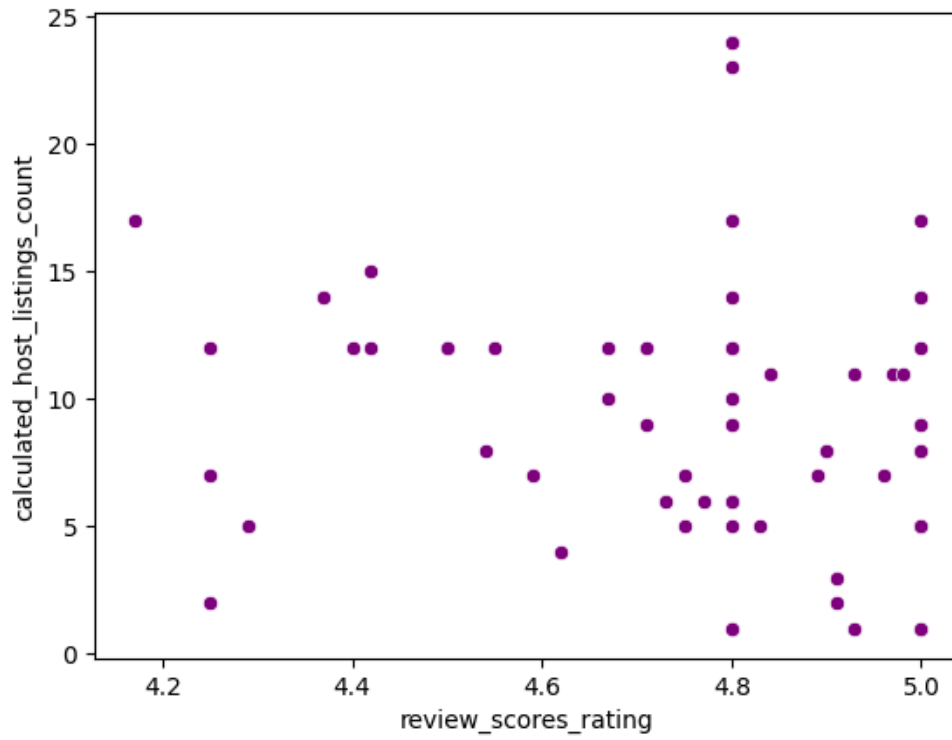
Entire home



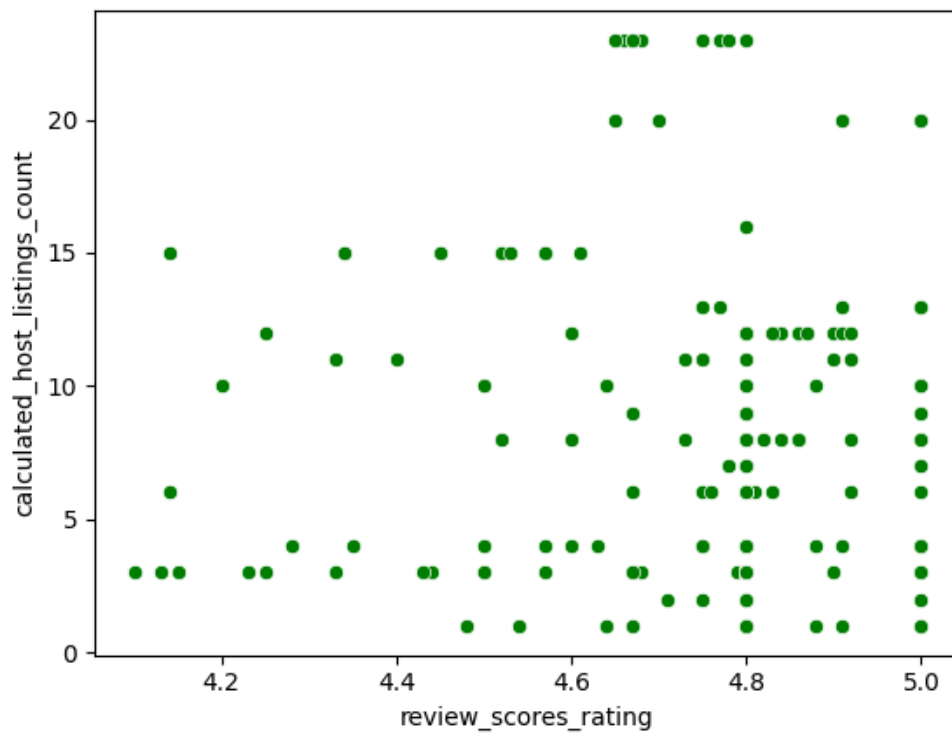
Hotel room



Shared room



Private Room



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

Al principio, las correlaciones más fuertes se deben a variables casi idénticas entre sí, como `minimum_nights` y `minimum_minimum_nights`. También destacan las 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
<code>estimated_occupancy_l365d</code>	<code>estimated_revenue_l365d</code>	0.934
<code>number_of_reviews_ltm</code>	<code>estimated_occupancy_l365d</code>	0.873
<code>number_of_reviews_ltm</code>	<code>estimated_revenue_l365d</code>	0.780
<code>number_of_reviews</code>	<code>estimated_occupancy_l365d</code>	0.748
<code>review_scores_rating</code>	<code>review_scores_communication</code>	0.743
<code>estimated_occupancy_l365d</code>	<code>reviews_per_month</code>	0.732
<code>review_scores_rating</code>	<code>review_scores_accuracy</code>	0.727
<code>number_of_reviews</code>	<code>estimated_revenue_l365d</code>	0.712
<code>id</code>	<code>months_since_first</code>	-0.705
<code>estimated_revenue_l365d</code>	<code>reviews_per_month</code>	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*

```
Top 3 correlaciones para: review_scores_rating
```

	Correlación
review_scores_accuracy	0.729541
review_scores_value	0.681655
review_scores_communication	0.619680

Resultado de modelo lineal múltiple

R²: 0.6589123820611507

R: 0.8117341818977137

Modelo de regresion lineal multiple *host_acceptance_rate*

Top 3 correlaciones para: host_acceptance_rate

	Correlación
minimum_nights	-0.060021
minimum_minimum_nights	-0.057223
maximum_maximum_nights	0.056526

Resultado de modelo lineal múltiple

R²: 0.006897364870583567

R: 0.08305037549935321

Modelo de regresion lineal multiple *host_is_superhost*

Top 3 correlaciones para: host_is_superhost_num

	Correlación
estimated_occupancy_1365d	0.351063
number_of_reviews_ltm	0.335287
number_of_reviews_ly	0.301968

Resultado de modelo lineal múltiple

R²: 0.13218365275982968

R: 0.3635706984340593

Modelo de regresion lineal multiple *host_total_listings_count*

Top 3 correlaciones para: host_total_listings_count

	Correlación
calculated_host_listings_count	0.621283
calculated_host_listings_count_entire_homes	0.353657
review_scores_rating	-0.133247

Resultado de modelo lineal múltiple

R²: 0.39807898991952884

R: 0.6309350124375163

Modelo de regresion lineal multiple *accommodates*

Top 3 correlaciones para: *accommodates*

	Correlación
<i>bedrooms</i>	0.649348
<i>beds</i>	0.551342
<i>bathrooms</i>	0.383753

Resultado de modelo lineal múltiple

R^2 : 0.52733810722506

R: 0.726180492181565

Modelo de regresion lineal multiple *bedrooms*

Top 3 correlaciones para: *bedrooms*

	Correlación
<i>accommodates</i>	0.649348
<i>bathrooms</i>	0.526317
<i>beds</i>	0.444764

Resultado de modelo lineal múltiple

R^2 : 0.5272266061761601

R: 0.726103715853431

Modelo de regresion lineal multiple *price*

Top 3 correlaciones para: *price*

	Correlación
<i>accommodates</i>	0.361541
<i>bedrooms</i>	0.354766
<i>bathrooms</i>	0.296714

Resultado de modelo lineal múltiple

R^2 : 0.16963182068750193

R: 0.4118638375573922

Modelo de regresión lineal múltiple *review_scores_value*

Top 3 correlaciones para: review_scores_value	
	Correlación
review_scores_rating	0.681655
review_scores_accuracy	0.611525
review_scores_cleanliness	0.536734

Resultado de modelo lineal múltiple

R^2 : 0.5080149847929951

R: 0.7127516992564767

Modelo de regresión lineal múltiple *bathrooms*

Top 3 correlaciones para: bathrooms	
	Correlación
bedrooms	0.526317
accommodates	0.383753
price	0.296714

Resultado de modelo lineal múltiple

R^2 : 0.2920132468962007

R: 0.540382500545864

Modelo de regresión lineal múltiple *reviews_per_month*

Top 3 correlaciones para: reviews_per_month	
	Correlación
number_of_reviews_ltm	0.775723
estimated_occupancy_1365d	0.751280
number_of_reviews_ly	0.587285

Resultado de modelo lineal múltiple

R^2 : 0.6321178025051799

R: 0.7950583642130808

Las variables de satisfacción de huéspedes como `review_scores_rating` se encuentran altamente correlacionadas con otras dimensiones de experiencia, como `accuracy`, `value` y `communication`, mostrando que la percepción del cliente se mueve de manera conjunta. El estatus de superhost guarda relación moderada con ocupación y volumen de reseñas, mientras que el número total de propiedades de un anfitrión se asocia casi exclusivamente con diferentes conteos de listados, sin impacto claro en la calidad. Las variables de capacidad (`accommodates`, `bedrooms`, `bathrooms`, `beds`) están fuertemente ligadas entre sí y también se relacionan con el precio, aunque este último depende de factores adicionales como ubicación o temporada. Finalmente, las métricas de demanda (`reviews_per_month`) se correlacionan de manera fuerte con reseñas acumuladas y con la ocupación, confirmando que más huéspedes generan más retroalimentación.

En general, se observa que las correlaciones más altas surgen entre variables que representan la misma dimensión (capacidad, satisfacción o demanda), lo cual explica los valores elevados de R^2 en esos modelos. También queda claro que el precio y la tasa de aceptación de anfitriones requieren información adicional para ser bien explicados, pues sus correlaciones simples son bajas. Un punto importante es que las correlaciones múltiples siempre superan a las simples, ya que la combinación de varias variables predice mejor que una sola por separado; sin embargo, esto también puede generar multicolinealidad cuando se incluyen predictores que miden casi lo mismo. En conclusión, los modelos lineales múltiples tienen mayor correlación que los simples.

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**.