



Tecnológico de Monterrey

Actividad 1. (Regresión Lineal Simple y Múltiple)

Sara Michelle Menocal Ramírez A01340128

28 de Septiembre, 2025

Gestión de proyectos de plataformas tecnológicas

Group 201

I. Limpieza de los datos

La base de datos contiene 79 columnas y 26,401 registros de la cual se eliminaron tres columnas vacías ('neighbourhood_group_cleansed', 'calendar_updated', 'license'), quedando con 76 columnas y los mismos registros. Para la limpieza de los datos se realizó:

- Valores Nulos

Para el tratamiento de los valores nulos se utilizaron diferentes métodos para rellenar los registros vacíos, se analizó debidamente cada columna y se definió qué método es el que se ajusta mejor a la necesidad de la variable. Imputando con la mediana para aquellas variables que necesitaran un número entero que no desequilibró la descripción estadística, así como la media para aquellos valores que necesitaran decimales y especificando cuántos dependiendo de la columna. Se rellenó con valores específicos en algunas dependiendo de lo que indicaba, como urls específicos, 0, no information provided, etc., de la misma manera se rellenó con la respuesta más popular, o con f para variables categóricas.

- Outliers:

Para tratar los outliers se realizó un

Tipos de habitación

El análisis se enfocará en comparar por tipos de habitación de la columna "room type", la cual cuenta con 4 tipos de cuarto. Para conocer cómo algunas variables impactan dependiendo del tipo de cuarto y si nos muestra información relevante.

Tabla 1.1

Características de los sub data frames

Tipo de cuarto	% del df	Registros
Entire home/apt	65%	17, 235

Private room	33%	8, 867
Shared room	0.76%	208
Hotel room	0.34%	91

El data frame se dividió en 4, mostrando que tanto abarca cada uno de los registros totales, siendo Entire home/apt el tipo de cuarto más popular con un 65% (17, 235 registros) de todos los tipos de cuarto. Private room es el segundo tipo de cuarto más popular con el 33% (8,867 registros) y Shared room y Hotel room con menos del 1% de los datos totales.

II. Correlaciones por tipo de habitación

1. La correlación entre la tasa de aceptación de reservas por parte del anfitrión y la tasa de respuesta por tipo de cuarto es:

Tabla 2.1

*Coeficientes de correlación en **host_acceptance_rate** vs **host_response_rate** (y, x) por tipo de cuarto*

Tipo de cuarto	Correlación
Entire home/apt	0.527
Private room	0.447
Shared room	0.213
Hotel room	0.206

La tabla nos muestra como los coeficientes de correlación son positivos en todos los tipos de habitación, aunque sólo uno muestra débilmente una correlación significativa estadística de Entire home/apt, lo que podría indicar que los anfitriones que responden rápido también aceptan más reservas. Private room muestra una correlación moderada, nada sobresaliente, al igual Shared room y Hotel room no muestra correlaciones altas, lo que podría tratarse de procesos con menor relación entre la respuesta y aceptación del host.

2. La correlación entre la calificación promedio de los huéspedes con la cantidad de propiedades que administra un anfitrión por tipo de cuarto es:

Tabla 2.2

*Coeficientes de correlación en **review_scores_rating** vs **calculated_host_listings_count** (y, x) por tipo de cuarto*

Tipo de cuarto	Correlación
Entire home/apt	-0.086
Private room	-0.105
Shared room	-0.155
Hotel room	-0.261

De acuerdo con las correlaciones, se puede identificar que la relación entre la calificación promedio con la cantidad de propiedades es negativa en todos los tipos de habitación, sugiriendo una ligera tendencia a que hay menor calificación cuando el anfitrión cuenta con más propiedades, lo que podría significar que un host con más propiedades no cuenta con el mismo

servicio que uno con menos propiedades. La correlación más pronunciada se encuentra en el tipo de cuarto que es Hotel room con -0.261, pudiéndose tratar de experiencias más impersonales o estandarizadas. Los otros tipos de habitaciones también tienen correlaciones algo significativas. El tipo de cuarto Entire home/apt muestra que es el tipo menos afectado por esto.

3. La correlación entre la tasa de aceptación del anfitrión y el precio de alojamiento por tipo de cuarto es:

Tabla 2.3

*Coeficientes de correlación en **host_acceptance_rate vs price** (y, x) por tipo de cuarto*

Tipo de cuarto	Correlación
Entire home/apt	0.028
Private room	0.076
Shared room	-0.097
Hotel room	-0.162

Los coeficientes se muestran un poco variados, ya que en los tipos de cuarto Entire home/apt hay una correlación casi nula, lo cual no indica mucho y en Private room existe una correlación leve positiva, que podría indicar que los anfitriones que aceptan más podrían tener precios un poco más elevados. En cambio con Hotel room si existe un efecto más pronunciado negativo, a mayor aceptación menor precio promedio, en el tipo Shared room existe también una correlación negativa pero muy baja, lo que podría asociar que a mayor aceptación los precios podrían ser menores.

4. La correlación entre la disponibilidad en el año y número de reseñas por tipo de cuarto es:

Tabla 2.4

*Coeficientes de correlación en **availability_365** vs **number_of_reviews** (y, x) por tipo de cuarto*

Tipo de cuarto	Correlación
Entire home/apt	0.017
Private room	0.046
Shared room	-0.269
Hotel room	0.249

Esta correlación presenta un poco de variabilidad ya que muestra una correlación significativa en Hotel room con 0.249 indicando que la disponibilidad si influye en la cantidad de reseñas, entre mayor disponibilidad más reseñas, aunque en el tipo de habitación Shared room muestra una correlación negativa con -0.269 lo que podría significar que aunque exista una disponibilidad durante todo el año, este no influye en más reservas ni reseñas. Con los otros dos tipos de habitación restante existe una correlación casi nula, lo que podría sugerir que hay otros factores que podrían influir más en la cantidad de reseñas que la disponibilidad.

5. La correlación entre la tasa de aceptación por host y el número de reseñas por tipo de cuarto es:

Tabla 2.5

*Coeficientes de correlación en **host_acceptance_rate** vs **number_of_reviews** (y, x) por tipo de cuarto*

Tipo de cuarto	Correlación
Entire home/apt	0.108
Private room	0.071
Shared room	-0.01
Hotel room	0.291

Las correlación muestra ser todas positivas a excepción en Shared room con una correlación negativa casi nula con -0.01, señalando que no existe una relación clara en la tasa de aceptación y el número de reseñas en los tipos de habitación compartida. El tipo de cuarto que muestra una correlación más alta es en Hotel room con 0.291 marcando que a mayor aceptación, más reseñas. Con Entire home/apt y Private room existe una relación leve, en donde también se podría decir que la aceptación influye poco en la cantidad de reseñas.

6. La correlación entre las reseñas por mes y la calificación en comunicación por tipo de cuarto es:

Tabla 2.6

Coeficientes de correlación en reviews_per_month vs review_scores_communication (y, x) por tipo de cuarto

Tipo de cuarto	Correlación
Entire home/apt	-0.125
Private room	0.148

Shared room	-0.336
Hotel room	-0.441

La tabla nos muestra que la correlación entre estas variables es negativa en casi todos los tipos de habitación, con una influencia significativa mayormente en Hotel room y Shared room, indicando que alta rotación de huéspedes podría implicar una comunicación más impersonal y automatizada. Con Private room la correlación es positiva en donde a mayor frecuencia de reseñas, mejor percepción de comunicación.

III. Correlaciones entre todas las variables

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

Entire home/apt

Tabla 3.1

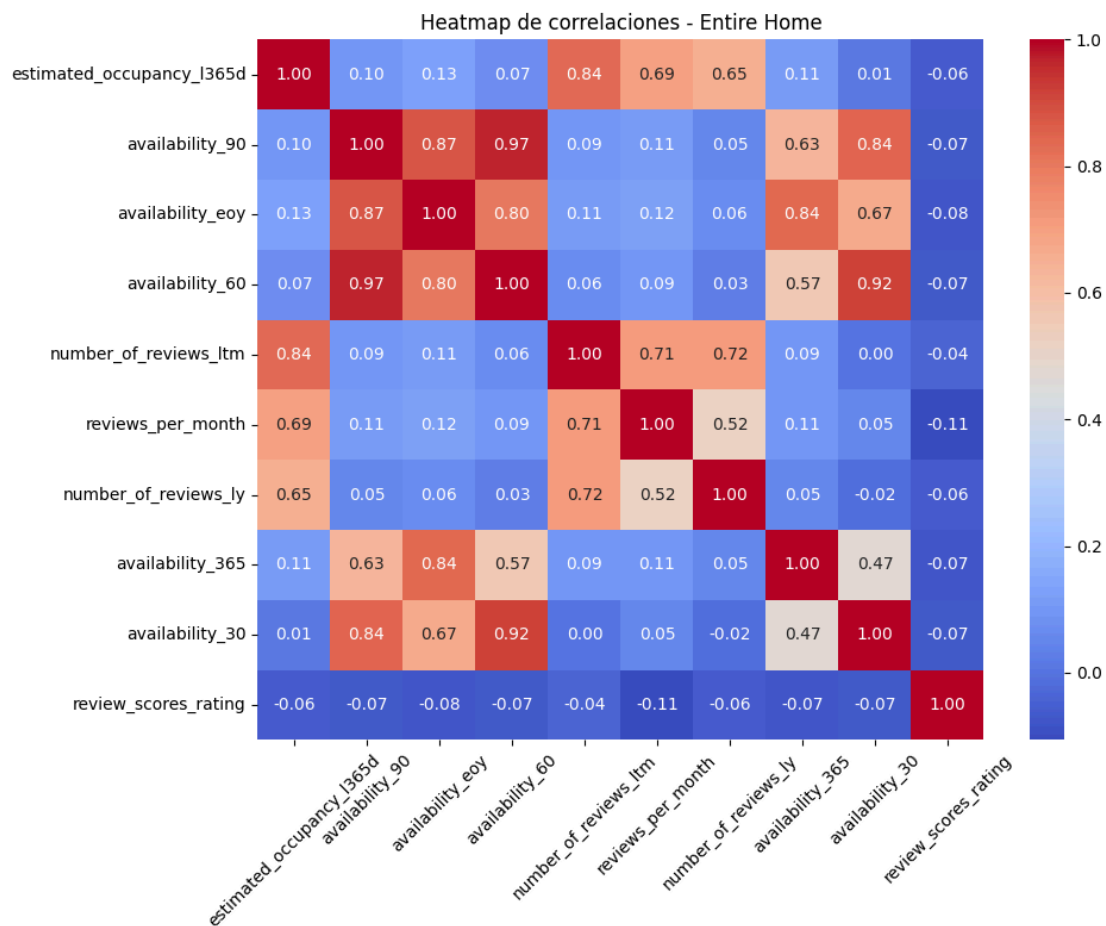
Top 10 variables más correlacionadas entire home/apt

Variable A	Variable B	Correlación
availability_60	availability_90	0.9677
availability_60	availability_30	0.9181
availability_90	availability_eoy	0.8747
availability_365	availability_eoy	0.8448

number_of_reviews_ltm	estimated_occupancy_l365d	0.8354
availability_30	availability_90	0.8448
availability_60	availability_eoy	0.8029
number_of_reviews_tlm	reviews_per_mont	0.7090
number_of_reviews_ltm	number_of_reviews_ly	0.7152
reviews_per_month	estimated_occupancy_l365d	0.6932

Gráfico 1

heatmap top correlaciones entre home/apt



Private room

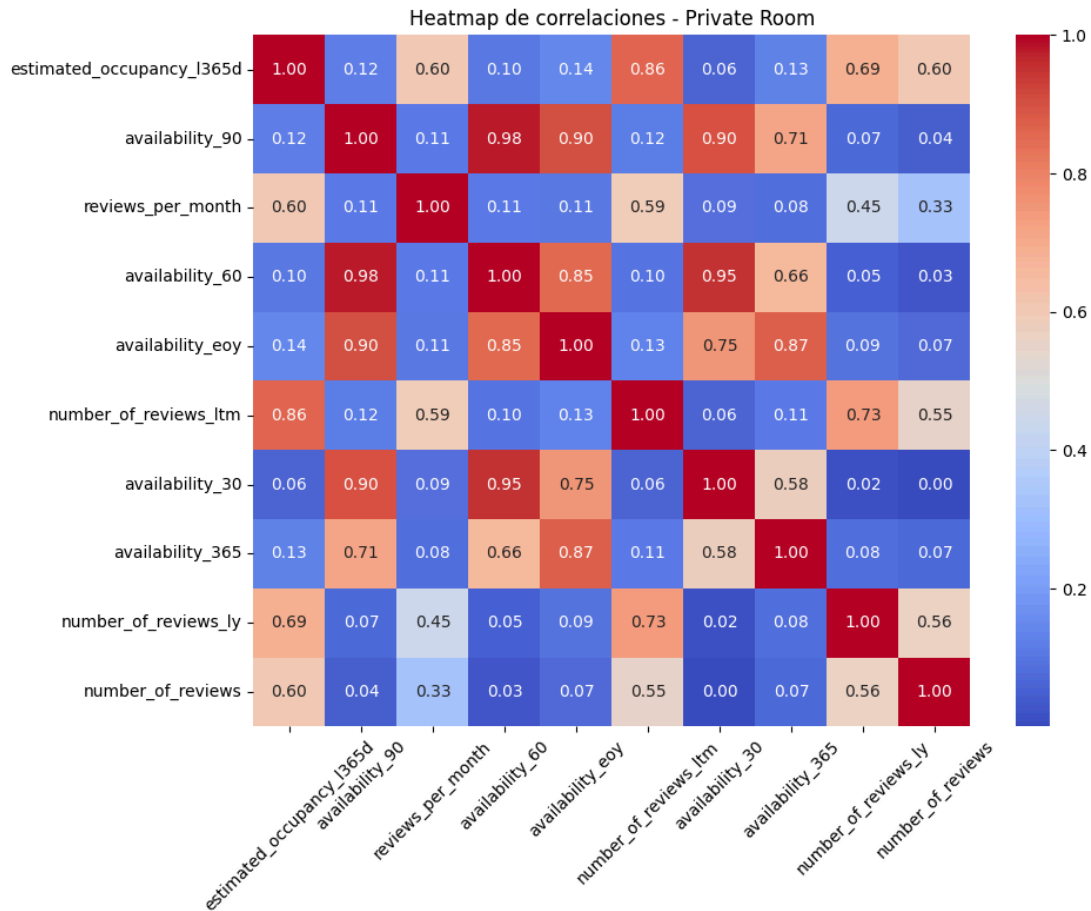
Tabla 3.2

Top 10 variables más correlacionadas private room

Variable A	Variable B	Correlación
availability_60	availability_90	0.9798
availability_60	availability_30	0.9479
availability_90	availability_eoy	0.9029
availability_365	availability_eoy	0.8737
number_of_reviews_ltm	estimated_occupancy_l365d	0.8628
availability_30	availability_eoy	0.7545
number_of_reviews_ltm	number_of_reviews_ly	0.7350
availability_365	availability_90	0.7108
reviews_per_month	number_of_reviews_ltm	0.5930
number_of_reviews_ltm	estimated_occupancy_l365d	0.5987

Gráfico 2

heatmap top correlaciones private room



Shared room

Tabla 3.3

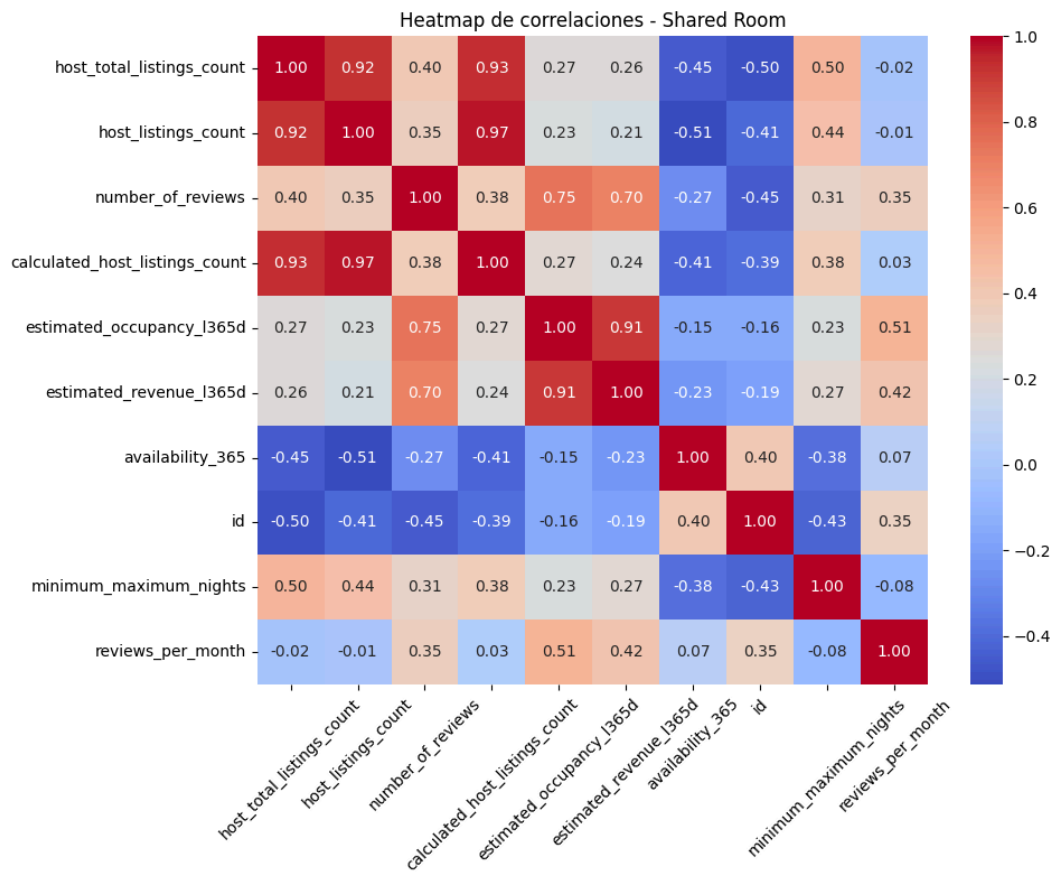
Top 10 variables más correlacionadas shared room

Variable A	Variable B	Correlación
host_listings_count	calculated_host_listings_count	0.9713
estimated_occupancy_l365d	estimated_revenue_l365d	0.9065
host_total_listings_count	calculated_host_listings_count	0.9319
host_total_listings_count	host_listings_count	0.9242

number_of_reviews	estimated_occupancy_l365d	0.7477
number_of_reviews	estimated_revenue_l365d	0.6982
minimum_maximum_nights	host_total_listings_count	0.4963
reviews_per_month	estimated_occupancy_l365d	0.5124
number_of_reviews	reviews_per_month	0.3545
id	host_total_listings_count	-0.4976

Gráfico 3

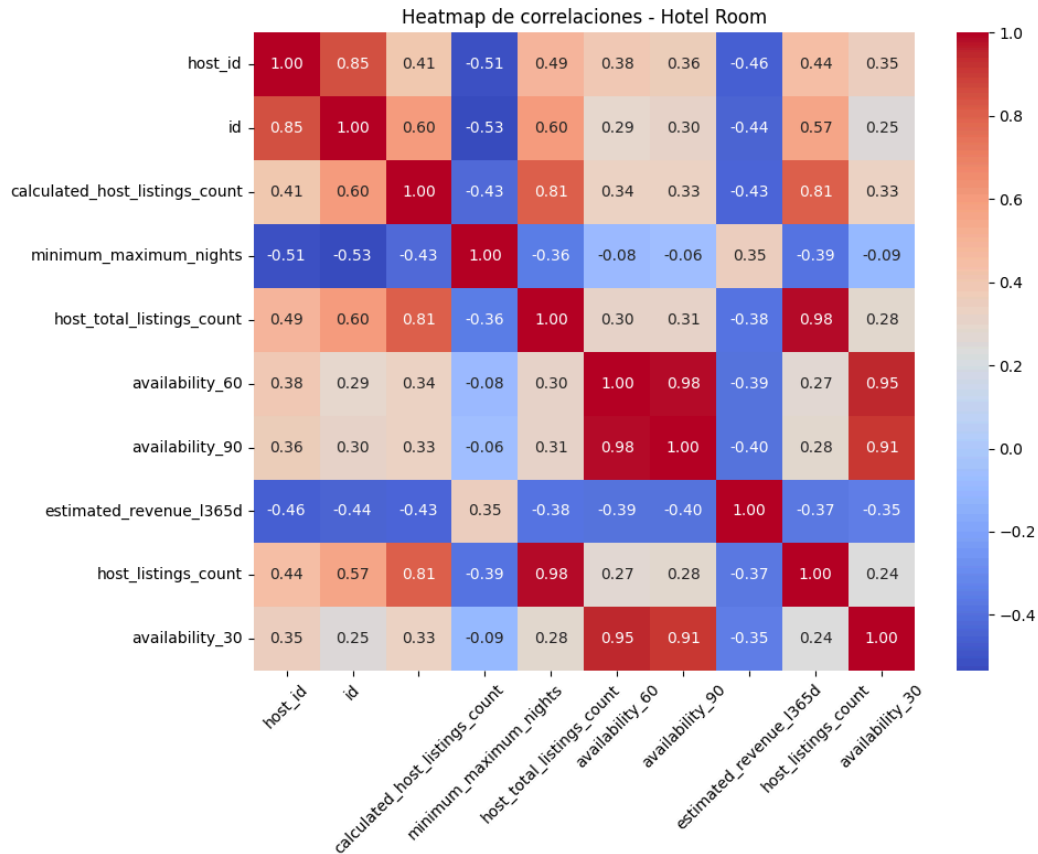
heatmap top correlaciones shared room



Hotel room
Tabla 3.4
Top 10 variables más correlacionadas hotel room

Variable A	Variable B	Correlación
availability_60	availability_90	0.9838
availability_60	availability_30	0.9494
availability_90	availability_30	0.9150
host_total_listings_count	host_listings_count	0.9844
calculated_host_listings_count	host_total_listings_count	0.8102
calculated_host_listings_count	host_listings_count	0.8091
host_id	id	0.8542
estimated_occupancy_1365d	estimated_revenue_1365d	0.9065
id	host_listings_count	0.5666
minimum_maximum_nights	id	-0.5347

Gráfico 4
heatmap top correlaciones hotel room



IV. Modelo de regresión lineal

Crear el mejor modelo de regresión lineal múltiple 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).

Calificación general de los huéspedes ('review scores rating')

Las variables con mayor correlación con review scores rating son:

Tabla 4.1.1*Variables que muestran correlaciones significativas en comparación a todas*

Variable	Correlación
review_scores_value	0.6591
review_scores_accuracy	0.6475
review_scores_cleanliness	0.5536
review_scores_communication	0.5299
review_scores_checkin	0.4924
review_scores_location	0.3136
host_is_superhost	0.1821

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 7 .

De acuerdo a los coeficientes se eligió sólo 5 variables ('review_scores_accuracy', 'review_scores_value', 'review_scores_cleanliness', 'review_scores_communication', 'review_scores_checkin') para ver que tanto poder predictivo tiene sobre aportan sobre nuestra y.

Los resultados de coeficientes por variable del modelo fueron:

- **review scores accuracy = 0.3041**, refiriendo a que la precisión del anuncio tiene una influencia significativa en la calificación general.
- **review scores value = 0.2787**, la percepción del valor si influye notoriamente.

- **review scores communication = 0.1807**, una buena comunicación aporta positivamente a una buena calificación
- **review scores cleanliness = 0.1588**, significando que la limpieza impacta de manera moderada .
- **Review scores checkin = 0.1366**, la facilidad del check in u otras razones mejoran el modelo.

Tabla 4.1.2*Modelo de regresión lineal múltiple para review scores rating*

Métrica	Coefficiente
R ² (coef. determinación)	0.5823
Correlación	0.7631

El modelo explica el 58.23% de la variabilidad de la calificación general de los huéspedes, la precisión del anuncio, el valor percibido, la limpieza, la buena comunicación y la facilidad de check in influyen significativamente en cómo califican los huéspedes a los hosts. Un valor de 0.7631 marca una correlación fuerte y positiva, a medida que mejoran estos aspectos la calificación también.

El modelo demuestra robustez y una buena explicación, la calificación que de un huésped es influenciado significativamente por buenos aspectos en la experiencia general del alojamiento, buena comunicación, limpieza, salida fácil, siendo elementos de los cuales los hosts se pueden basar para mejorar sus calificaciones.

Tasa de aceptación del host ('host acceptance rate')

Las variables con mayor correlación con host acceptance rate son:

Tabla 4.2.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
hots_response_rate	0.4968
estimated_occupancy_l365d	0.2175
number_of_reviews_ltm	0.1885
availability_365	0.1662

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 4 .

Se intentó el modelo con las 4 variables con una correlación significativa pero number_of_reviews_ltm y availability_365 no demostraron impactos valiosos a este modelo por lo que el modelo final quedó con solo 2 variables, siendo host_response_rate y estimated_occupancy_l365d. El modelo dio unos coeficientes de:

- **host response rate = 0.5324**, es la variable que más influye en la tasa de aceptación, a mayor respuesta mayor probabilidad de aceptación.
- **estimated occupancy l365d = 0.0076**, aunque tiene un efecto marginal pero positivo, aumentó la predicción del modelo, ya que a mayor ocupación mayor aceptación.

Tabla 4.2.2

Modelo de regresión lineal múltiple para review scores rating

Métrica	Coefficiente
R ² (coef. determinación)	0.2623
Correlación	0.5121

El modelo es simple, con solo dos variables explica el 26.23% de la variabilidad en la tasa de aceptación del host, demostrando una moderada predicción, su correlación también demuestra un desempeño moderado entre las predictoras y la dependiente.

A mayor tasa de respuesta, existe mayor probabilidad que se acepte la reserva, significando que la tasa de aceptación depende mayoritariamente por sus capacidad de respuesta, la ocupación durante todo el año también influye en menor medida.

Super anfitrión ('host is superhost')

Las variables con mayor correlación con host acceptance rate son:

Tabla 4.3.1*Variables que muestran correlaciones significativas en comparación a todas*

Variable	Correlación
estimated_occupancy_l365d	0.3511
number_of_reviews_ltm	0.3353
number_of_reviews_ly	0.3020
number_of_reviews	0.2635

estimated_revenue_l365d	0.2627
reviews_per_month	0.2172
host_response_rate	0.2166

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 7

Se intentó un modelo incluyendo las 7 variables pero se rechazó la inclusión de number of reviews ly, number of reviews, estimated revenue l365d y reviews per month. El modelo final para host is superhost quedo con:

- **host response rate = 0.0153**, la tasa de respuesta es la mayor predictora para saber si un anfitrión es superhost, a mayor tasa, mayor probabilidad de ser superhost.
- **number of reviews ltm = 0.0042**, las reseñas recientes incrementan la probabilidad de ser considerado superhost.
- **estimated occupancy l365d = 0.001**, a mayor ocupación anual, mayor probabilidad aunque tiene un efecto menor.
-

Tabla 4.3.2

Modelo de regresión lineal múltiple para host is superhost

Métrica	Coefficiente
R ² (coef. determinación)	0.1510
Correlación	0.3886

El modelo demuestra un desempeño moderado explicando un 15.10% de la variabilidad de si es un super anfitrión o no, la correlación moderadamente baja entre los predictores y la dependiente, aunque no captura la complejidad del si es o no, permite identificar perfiles con alto potencial. Los anfitriones con alta tasa de respuesta, buena ocupación y bastantes reseñas tienen mayor probabilidad de ser considerado un súper anfitrión.

Número total de propiedades listadas por el anfitrión ('host total listings count')

Las variables con mayor correlación con host total listings count son:

Tabla 4.4.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
host_listings_count	0.7927
calculated_host_listings_count	0.6373
calculated_host_listings_count_entire_homes	0.3605
host_acceptance_rate	0.1317
host_response_rate	0.1053

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 5.

Para este modelo solo se utilizaron dos variables como predictoras ya que las demás no presentaban aportes significativos al modelo, los coeficientes resultaron de:

- **host listings count = 0.8538**, es la predictora con mayor influencia, a mayor número de anuncios registrados, mayor será el total atribuido.
- **calculated host listings count = 0.1455**, aportando moderadamente, la estimación interna de airbnb, puede incluir co-anfitriones o propiedades gestionadas indirectamente.

Tabla 4.4.2

Modelo de regresión lineal múltiple para host total listings count

Métrica	Coefficiente
R ² (coef. determinación)	0.6332
Correlación	0.7957

El modelo demuestra un buen desempeño en cuanto a host total listings count, refleja que tanto el número de anuncios registrados directamente (la mayor predictora) y co-hosting, listados ocultos o propiedades no declaradas (con una influencia moderada) estiman el número de anuncios a un anfitrión. La correlación es alta significando que el modelo tiene una relación fuerte y positiva, concluyendo que en mayor host listings count y calculated listings count, habrá mayor anuncios atribuidos a un host.

Capacidad de alojamiento ('accommodates')

Las variables con mayor correlación con accommodates son:

Tabla 4.5.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
----------	-------------

bedrooms	0.6493
beds	0.5513
price	0.3978
bathrooms	0.3791
estamatided_revenue_1365d	0.2163

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 5.

Para el modelo se excluyó estimated revenue 1365d ya que no tenía efecto predictor, utilizando solo 4 variables para estimar la capacidad de alojamiento, los coeficientes son:

- **bedrooms = 0.9854**, el número de habitaciones aporta casi un huésped más, siendo la predictora más fuerte, entre más habitaciones, mayor espacio.
- **beds = 0.6429**, entre más camas, más huéspedes, influye directamente.
- **price = 0.0004**, aunque con un efecto muy marginal, tiene un poder predictivo en el modelo.
- **bathrooms = 0.0910**, los baños aportan 0.09 huéspedes adicionales, siguiendo el patrón entre más espacio mayor capacidad de alojamiento.

Tabla 4.5.2

Modelo de regresión lineal múltiple para accommodates

Métrica	Coefficiente
---------	--------------

R ² (coef. determinación)	0.5314
Correlación	0.7289

El modelo muestra una base lógica, entre mayor espacio hay mayor capacidad de alojamiento, explica el 53.14% de la variabilidad de accommodates, captando la influencia de las habitaciones, el número de camas, el precio y los baños. La correlación de 0.72 demuestra que en efecto las predictoras tienen una relación fuerte y positiva con la dependiente, capturando la tendencia general.

Habitaciones ('bedrooms')

Las variables con mayor correlación con bedrooms son:

Tabla 4.6.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
accommodates	0.6493
bathrooms	0.5238
beds	0.4448
price	0.3489

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 4.

Para el modelo se excluyó price ya no que representaba ninguna aportación al modelo, creando un modelo teniendo 3 variables predictoras, teniendo coeficientes de:

- **accommodates = 0.2059**, la capacidad de alojamiento tiene una relación directa con el número de habitaciones, más recámaras permitiría mayores huéspedes.
- **bathrooms = 0.3627**, es la que mayor tiene correlación, dado que los lugares con más baños sugieren que habrás más habitaciones.
- **beds = 0.1072**, el número de camas influirá en el número de habitaciones, así como para también conocer el espacio de las habitaciones.

Tabla 4.6.2

Modelo de regresión lineal múltiple para bedrooms

Métrica	Coefficiente
R ² (coef. determinación)	0.5210
Correlación	0.7218

El modelo presenta una explicación moderada, explicando el 52% de la variabilidad en el número de cuartos, al igual que el anterior modelo, se basa en algo lógico, en mayor espacio hay mayor probabilidad de que haya más habitaciones, entre más camas, más baños y mayor capacidad de alojamiento, el lugar tendrá más recámaras. Su correlación demuestra una relación fuerte, corroborando que capta la tendencia general, que las x si influyen en y.

Precio ('price')

Las variables con mayor correlación con price s son:

Tabla 4.7.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
accommodates	0.3978
bedrooms	0.3489
bathrooms	0.2750
beds	0.2479
estimated revenue l365d	0.2243
calculated host listings count entire homes	0.2046

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 6.

Para el modelo se mantuvieron sólo 4 variables, ya que beds y estimated revenue no aportan valor al modelo, los coeficientes obtenidos son:

- **accommodates = 102.50**, la capacidad de alojamiento es el mayor influenciador con el aumento del precio, cada huésped adicional son \$102 más.
- **bedrooms = 95.70**, el número de habitaciones aumenta \$96 más.
- **bathrooms = 113.00**, la cantidad de baños de un alojamiento influye fuertemente en el precio, cada baño adicional son \$113 más.

- **calculated hots listings count entire hombres = 28.04**, aunque la variable aporta muy poco, agrega ajuste al modelo, puede indicar profesionalización del host.

Tabla 4.7.2*Modelo de regresión lineal múltiple para price*

Métrica	Coefficiente
R ² (coef. determinación)	0.2055
Correlación	0.4533

El modelo captura débilmente la variabilidad del precio por capacidad de alojamiento, habitaciones, baños y el número de propiedades tipo entire home, con un 20.5% de poder explicativo. Su correlación es moderadamente baja entre las predictoras. Este modelo se basa en la estimación del precio por características físicas del alojamiento y su correlación no es tan significativa lo que indica que existen factores que son estas siendo capturados adecuadamente.

Calificación del valor percibido ('review scores value')

Las variables con mayor correlación con review scores value son:

Tabla 4.8.1*Variables que muestran correlaciones significativas en comparación a todas*

Variable	Correlación
review scores rating	0.6591
review scores accuracy	0.6044

review scores cleanliness	0.5098
review scores communication	0.4884
review scores checkin	0.4597
review scores location	0.395

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 6.

Para el modelo solo se utilizaron 3 variables que aportan significativamente a en la calificación del alojamiento, basándose en:

- **review scores rating = 0.4484**, La calificación general influye fuertemente en la percepción del valor, siendo la variable con mayor influencia.
- **review scores accuracy = 0.3335**, la veracidad de la información del lugar influye en la percepción del valor.
- **Review scores communication = 0.2076**, la buena comunicación influye positivamente en la percepción del valor.

Tabla 4.8.2

Modelo de regresión lineal múltiple para review scores value

Métrica	Coefficiente
R ² (coef. determinación)	0.5020
Correlación	0.7085

El moldeo captura bien la variabilidad de la percepción del valor en base una buena calificación general, buena descripción acertada y buena comunicación, entre mejor desempeño en estas áreas mayor percepción se tendrá del valor del alojamiento, explica moderadamente los datos con 50.2% de poder. Su correlación es significativa lo que corrobora que estos elementos influyen en la predicción del valor percibido.

Baños ('bathrooms')

Las variables con mayor correlación con bathrooms son:

Tabla 4.9.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
bedrooms	0.5238
accommodates	0.3791
price	0.2750
beds	0.2276
maximum minimum nights	0.1169

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 5.

Para el modelo solo se utilizaron 2 variables, de las cuales obtuvimos los coeficientes:

- **bedrooms = 0.4587**, el número de habitaciones influye significativamente en el número de baños que tendrá un alojamiento, entre más cuartos mayor probabilidad de tener ,más baños.

- **maximum minimum nights = 0.0283**, la mayor duración mínima permitida influye moderadamente en el poder del modelo, lo que podría indicar alojamientos más grandes.

Tabla 4.9.2

Modelo de regresión lineal múltiple para accommodates

Métrica	Coefficiente
R ² (coef. determinación)	0.2773
Correlación	0.5266

El modelo presenta un poder predictivo débil, la dos variables capturan débilmente los baños que tendrá un alojamiento, cuenta con 27.73% de explicación de la variabilidad, se podría analizar si hay factores que no se estén capturando, por lo que se podría en empezar en otros elementos que podrían ayudar a capturar la tendencia del número de baños, la correlación es significativa pero podría ser mayormente a la variable bedrooms.

Reseñas por mes ('reviews per month')

Las variables con mayor correlación con bathrooms son:

Tabla 4.10.1

Variables que muestran correlaciones significativas en comparación a todas

Variable	Correlación
number of reviews ltm	0.6882
estimated occupancy l365d	0.6784

number of reviews ly	0.5148
estimated revenue l365d	0.4619
number of reviews	0.3753

Nota. En este caso el número de variables que demostraron valores diferenciados en comparación con todas fueron 5.

Para el modelo sólo se consideraron 2 variables ya que las otras no tienen importancia alguna al modelo, sus coeficientes son de:

- **number of reviews ltm = 0.0307**, las reseñas de los últimos 12 meses se asocia con un aumento de 0.03 en reseñas por mes, lo que indica actividad constante.
- **estimated occupancy l365d = 0.0042**, la mayor ocupación estimada se asocia con mayores reseñas mensuales.

Tabla 4.10.2

Modelo de regresión lineal múltiple para reviews per month

Métrica	Coefficiente
R ² (coef. determinación)	0.5049
Correlación	0.7105

El modelo presenta una significancia en la explicación de la variabilidad en las reseñas mensuales con un 50% de poder explicativo, así mismo, tiene una correlación fuerte con las predictoras y las reseñas por mes. Lo que nos permite estimar la frecuencia de las reseñas que

tendrán los alojamientos mensuales a partir de si tienen una actividad constante en los anuncios y su nivel de ocupación.