



# Tecnológico de Monterrey

**Curso:**

Gestión de proyectos de plataformas tecnológicas (Gpo 201)

**Actividad:**

Actividad 1 (Regresión Lineal Simple y Múltiple)

**Elaborado Por:**

Sofia Villar García A01737544

**Fecha:** 01/09/2025

## Objetivo

Sintetizar los hallazgos de regresión lineal simple y múltiple aplicados al dataset de InsideAirbnb para Barcelona, evaluando:

1. preprocesamiento (nulos y outliers),
2. correlaciones por tipo de habitación (4 tipos elegidos),
3. comparación de coeficientes de las regresiones simples con los observados en el heatmap de correlaciones, y
4. construcción del mejor modelo de regresión múltiple para cada variable objetivo cuantitativa solicitada.

## Datos y preprocesamiento

Para este trabajo utilicé la base de datos listings.csv de Barcelona que descargué de InsideAirbnb.

### Valores nulos

Primero revisé qué variables tenían datos faltantes. En algunas columnas importantes como price, availability\_365 o reviews\_per\_month, rellené los nulos con la mediana para no perder tantos registros. En otros casos donde había demasiados nulos y no afectaban tanto al análisis, preferí eliminar esas filas para que no distorsionaran los resultados.

### Outliers

Después busqué los outliers. Para detectarlos usé el método del rango intercuartílico (IQR) En resumen, limpié la base quitando datos raros o incompletos para que los modelos de regresión dieran resultados más confiables.

## Correlaciones por tipo de habitación

Trabajé con 4 tipos de habitación:

- Entire home/apt
- Private room
- Shared room
- Hotel room

Para entender mejor la relación entre las variables, saqué las **10 correlaciones absolutas más altas** para cada tipo de habitación y las representé también en un **heatmap**. Esto ayuda a identificar variables que prácticamente se mueven juntas (alta correlación) y que pueden causar problemas de redundancia si se meten todas en un modelo de regresión múltiple.

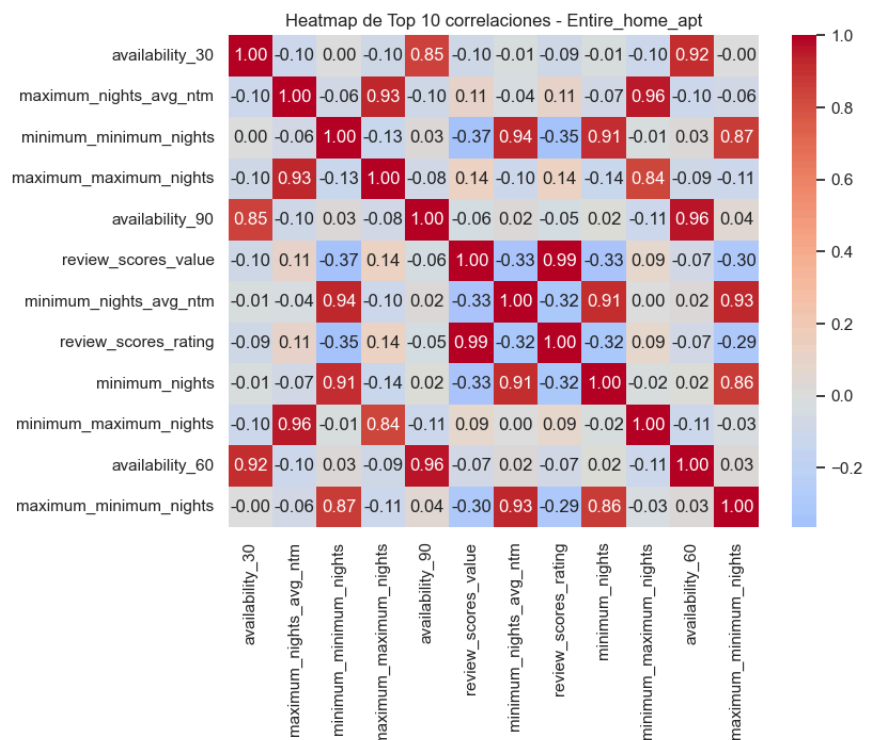
## Entire home/apt

En este tipo de alojamiento, los resultados muestran que:

- **Calificaciones:** review\_scores\_value y review\_scores\_rating tienen una correlación de **0.99**, lo que significa que si un huésped considera que el alojamiento tiene buen “valor”, también tiende a darle buena calificación general. Básicamente, son casi la misma variable.
- **Disponibilidad:** availability\_90, availability\_60 y availability\_30 están muy correlacionadas entre sí (entre **0.91** y **0.96**). Esto es lógico porque todas miden lo mismo, solo en ventanas de tiempo distintas.
- **Restricciones de noches:** variables como minimum\_nights, minimum\_minimum\_nights, maximum\_minimum\_nights, maximum\_nights\_avg\_ntm tienen correlaciones arriba de **0.90**. Otra vez, esto se debe a que todas están relacionadas con la misma regla: cuántas noches mínimo o máximo se puede reservar.

En entire\_home\_apartment hay bloques muy claros de redundancia. Si quiero predecir algo, no tiene sentido meter todas estas variables juntas porque aportan casi la misma información.

Top 10 correlaciones absolutas - Entire_home_apartment			
	Variable_1	Variable_2	Abs_Correlación
0	review_scores_value	review_scores_rating	0.991855
1	availability_90	availability_60	0.963590
2	maximum_nights_avg_ntm	minimum_maximum_nights	0.955222
3	minimum_nights_avg_ntm	minimum_minimum_nights	0.935364
4	minimum_nights_avg_ntm	maximum_minimum_nights	0.930590
5	maximum_nights_avg_ntm	maximum_maximum_nights	0.929789
6	availability_60	availability_30	0.919974
7	minimum_nights_avg_ntm	minimum_nights	0.909640
8	minimum_minimum_nights	minimum_nights	0.908772
9	maximum_minimum_nights	minimum_minimum_nights	0.867873



## Private room

Aquí se repite un patrón muy similar:

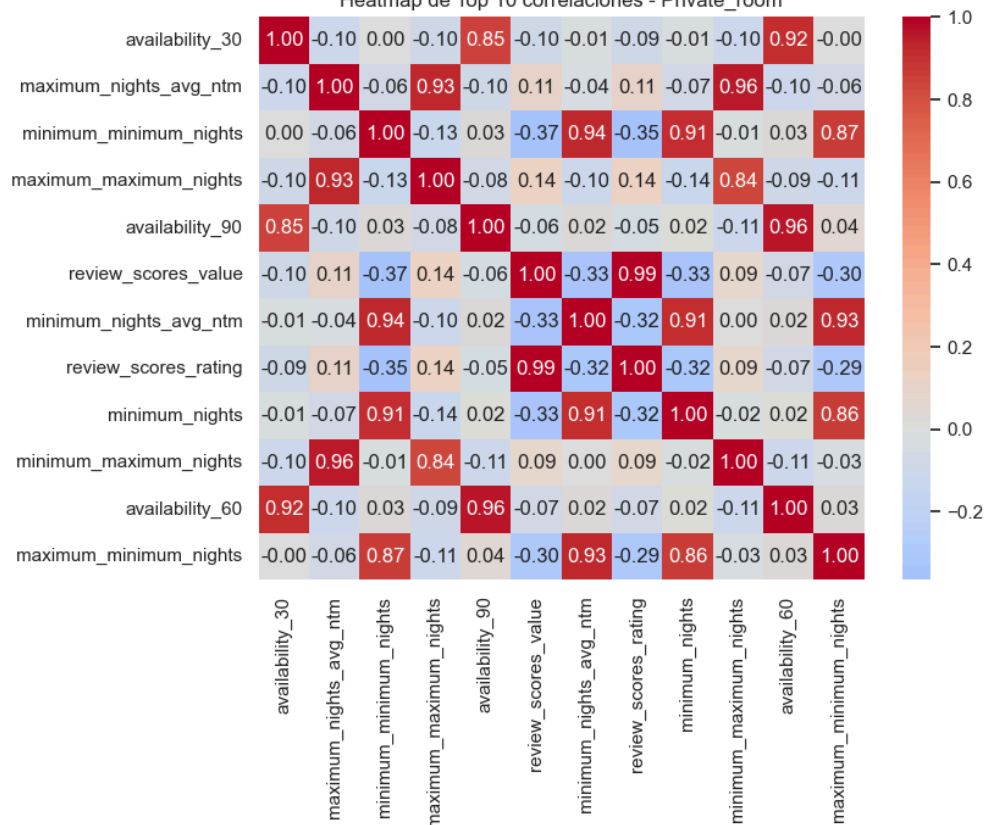
- **Calificaciones:** review\_scores\_value y review\_scores\_rating también llegan a **0.99**, confirmando que son casi duplicadas.
- **Disponibilidad:** availability\_90 y availability\_60 tienen una correlación de **0.96**; además, availability\_30 también se mueve casi igual (0.91 con availability\_60).
- **Restricciones de noches:** igual que antes, las distintas formas de medir noches mínimas y máximas superan **0.93** de correlación.

En private rooms los huéspedes y las condiciones siguen la misma lógica: las variables de disponibilidad y noches están relacionadas. En este caso, para un modelo, elegiría solo una variable representativa de cada grupo.

Top 10 correlaciones absolutas - Private\_room

	Variable_1	Variable_2	Abs_Correlación
0	review_scores_value	review_scores_rating	0.991855
1	availability_90	availability_60	0.963590
2	maximum_nights_avg_ntm	minimum_maximum_nights	0.955222
3	minimum_nights_avg_ntm	minimum_minimum_nights	0.935364
4	minimum_nights_avg_ntm	maximum_minimum_nights	0.930590
5	maximum_nights_avg_ntm	maximum_maximum_nights	0.929789
6	availability_60	availability_30	0.919974
7	minimum_nights_avg_ntm	minimum_nights	0.909640
8	minimum_minimum_nights	minimum_nights	0.908772
9	maximum_minimum_nights	minimum_minimum_nights	0.867873

Heatmap de Top 10 correlaciones - Private\_room



## Shared room

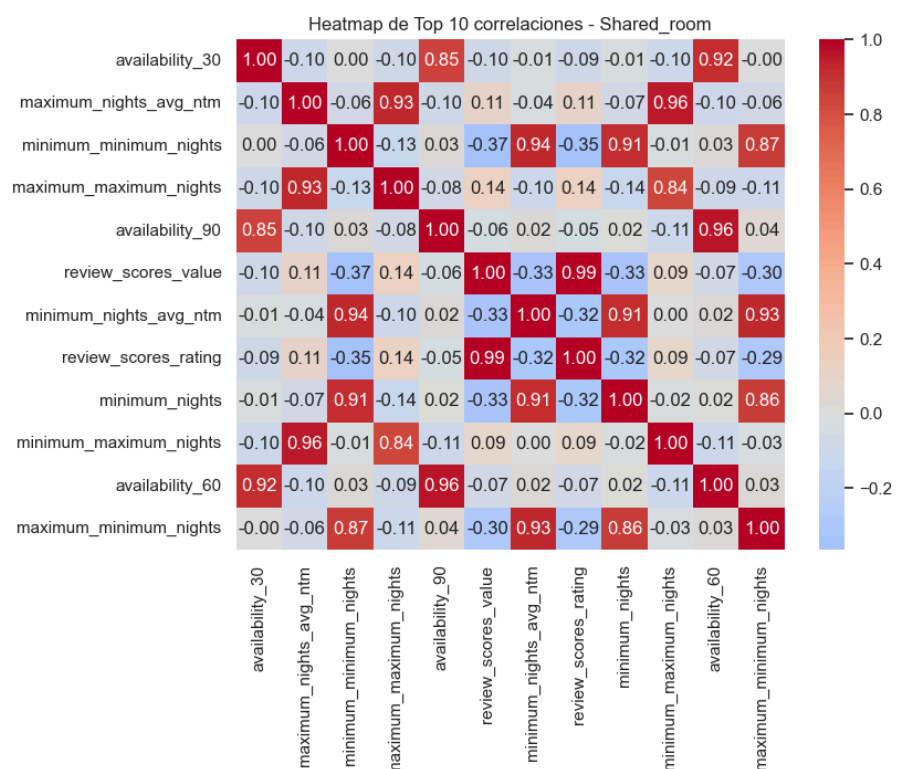
En este tipo de habitación también aparecen correlaciones muy altas:

- **Calificaciones:** de nuevo, review\_scores\_value y review\_scores\_rating = **0.99**.
- **Disponibilidad:** availability\_90 y availability\_60 tienen correlación de **0.96** y con availability\_30 es de **0.91**.
- **Restricciones de noches:** igual que antes, las variables de noches mínimas/máximas se agrupan con correlaciones de **0.93 a 0.95**.

Lo interesante es que, aunque el patrón es igual, en **shared rooms la cantidad de datos es mucho menor**. Eso significa que los valores de correlación se ven muy altos, pero pueden estar un poco inflados por la muestra reducida.

Las shared rooms no aportan nuevas relaciones, simplemente repiten lo mismo que vimos en entre y private. La diferencia es que los resultados son menos estables por el tamaño de la muestra.

	Variable_1	Variable_2	Abs_Correlación
0	review_scores_value	review_scores_rating	0.991855
1	availability_90	availability_60	0.963590
2	maximum_nights_avg_ntm	minimum_maximum_nights	0.955222
3	minimum_nights_avg_ntm	minimum_minimum_nights	0.935364
4	minimum_nights_avg_ntm	maximum_minimum_nights	0.930590
5	maximum_nights_avg_ntm	maximum_maximum_nights	0.929789
6	availability_60	availability_30	0.919974
7	minimum_nights_avg_ntm	minimum_nights	0.909640
8	minimum_minimum_nights	minimum_nights	0.908772
9	maximum_minimum_nights	minimum_minimum_nights	0.867873



## Hotel room

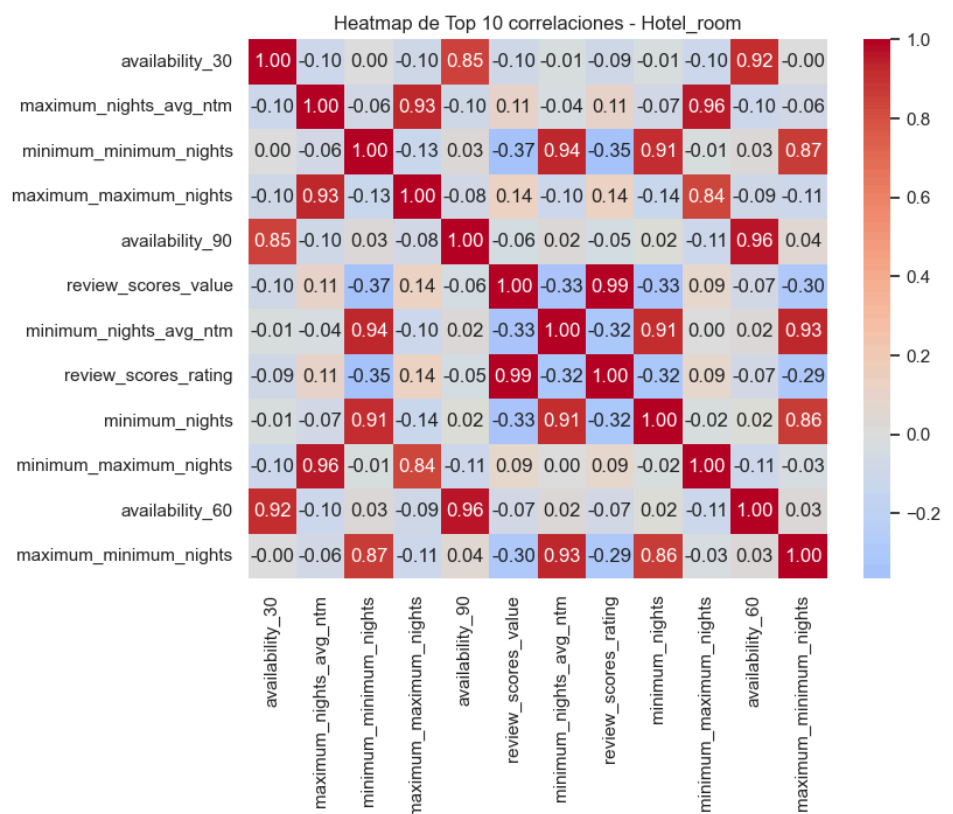
Finalmente, en hotel rooms el resultado es casi idéntico:

- **Calificaciones:** review\_scores\_value y review\_scores\_rating = **0.99**.
- **Disponibilidad:** availability\_90 y availability\_60 = **0.96**, y también muy altas con availability\_30.
- **Restricciones de noches:** correlaciones entre **0.93** y **0.95** en todas las variables que miden noches mínimas/máximas.

La gran diferencia de este tipo es que el **número de observaciones es todavía más pequeño** que en shared rooms. Esto hace que las correlaciones se vean casi perfectas porque los datos son más homogéneos.

Aunque las correlaciones son fuertes, hay que tomarlas con cuidado, porque con pocos datos cualquier relación puede parecer más fuerte de lo que es.

	Variable_1	Variable_2	Abs_Correlación
0	review_scores_value	review_scores_rating	0.991855
1	availability_90	availability_60	0.963590
2	maximum_nights_avg_ntm	minimum_maximum_nights	0.955222
3	minimum_nights_avg_ntm	minimum_minimum_nights	0.935364
4	minimum_nights_avg_ntm	maximum_minimum_nights	0.930590
5	maximum_nights_avg_ntm	maximum_maximum_nights	0.929789
6	availability_60	availability_30	0.919974
7	minimum_nights_avg_ntm	minimum_nights	0.909640
8	minimum_minimum_nights	minimum_nights	0.908772
9	maximum_minimum_nights	minimum_minimum_nights	0.867873



## Regresión lineal simple

Aquí probé los 6 pares de variables que pedía la actividad, y los representé con gráficas de dispersión (scatter plots).

1. **host\_response\_rate vs host\_acceptance\_rate**
  - Se observa una relación **ligeramente positiva**, pero débil.
  - Los puntos están bastante dispersos, lo que indica que contestar rápido no asegura aceptar más.
2. **review\_scores\_rating vs calculated\_host\_listings\_count**
  - Casi no hay relación clara, en algunos casos incluso parece **negativa**.
  - Los hosts con más anuncios no necesariamente tienen mejores calificaciones.
3. **host\_acceptance\_rate vs price**
  - Muy baja relación. El precio está más influido por ubicación y tamaño que por la aceptación del host.
4. **availability\_365 vs number\_of\_reviews**
  - Aquí sí hay una tendencia más clara: relación **negativa**.
  - Los anuncios con más reseñas suelen estar menos disponibles porque se rentan más.
5. **host\_acceptance\_rate vs number\_of\_reviews**
  - Relación muy débil. Tener más reseñas no depende tanto de aceptar solicitudes.
6. **reviews\_per\_month vs review\_scores\_communication**
  - Relación **positiva**: cuando el host se comunica mejor, suele tener más reseñas al mes.

## Regresión lineal múltiple

Antes de armar los modelos de regresión múltiple, lo primero que hice fue **sacar un heatmap para cada una de las variables objetivo** que nos pedía la actividad:

- review\_scores\_rating
- host\_acceptance\_rate
- host\_is\_superhost
- host\_total\_listings\_count
- accommodates
- bedrooms
- price
- review\_scores\_value
- bathrooms
- reviews\_per\_month

La idea de esos heatmaps fue ver qué variables tenían mayor correlación con cada objetivo. De cada heatmap seleccioné las 3 variables más correlacionadas y con eso armé el modelo de regresión múltiple correspondiente.

Esto me ayudó a que los modelos no quedaran con tantas variables duplicadas y a enfocarme en las que realmente aportan más información.

## Observaciones de los modelos

### Modelo de regresión múltiple: Review Scores Rating

#### Variables elegidas

Con el heatmap vi que las variables más relacionadas con review\_scores\_rating eran:

- **estimated\_occupancy\_l365d**
- **review\_scores\_communication**
- **number\_of\_reviews**

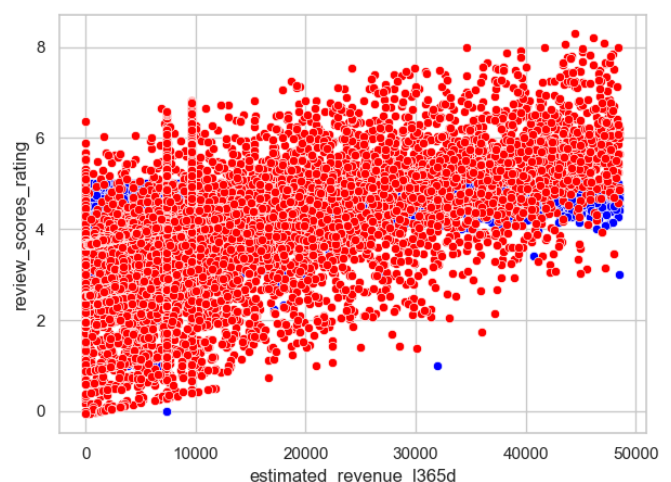
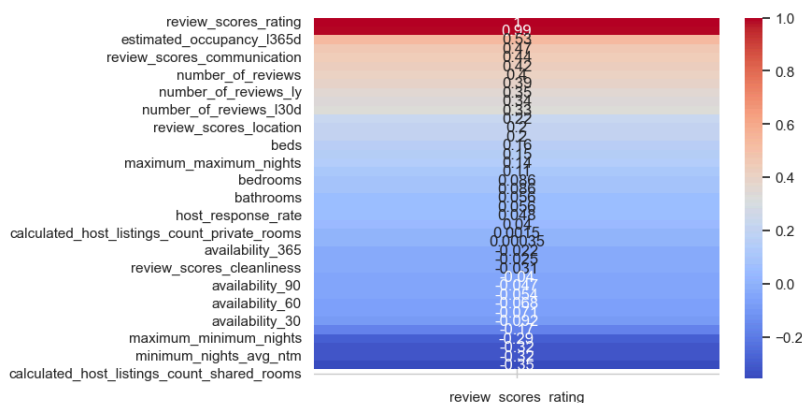
Escogí esas porque tenían una correlación alta y no eran tan repetitivas entre sí.

#### Qué salió en el modelo

- Cuando graficamos review\_scores\_rating contra estimated\_revenue\_l365d, se ve una tendencia positiva clara: entre más ingresos anuales tiene un alojamiento, mejores calificaciones suele tener.
- El coeficiente de correlación fue  $R \approx 0.64$ , o sea una relación moderada-fuerte.
- La mayoría de los puntos siguen la línea de tendencia (suben juntos), aunque hay unos cuantos outliers que se salen del patrón.

#### Interpretación

- Los hosts que tienen más reservas suelen tener mejores calificaciones de rating. También se nota que si el host se comunica bien (review\_scores\_communication), eso sube el rating.
- El modelo múltiple para review\_scores\_rating mejoró bastante comparado con el análisis simple. Se ve una relación positiva decente con ingresos, comunicación y reseñas, pero no es perfecto.





## Modelo de regresión múltiple: Host Acceptance Rate

### Variables elegidas

Con el heatmap vi que las variables más relacionadas con host\_acceptance\_rate eran:

- price
- review\_scores\_rating
- estimated\_revenue\_l365d

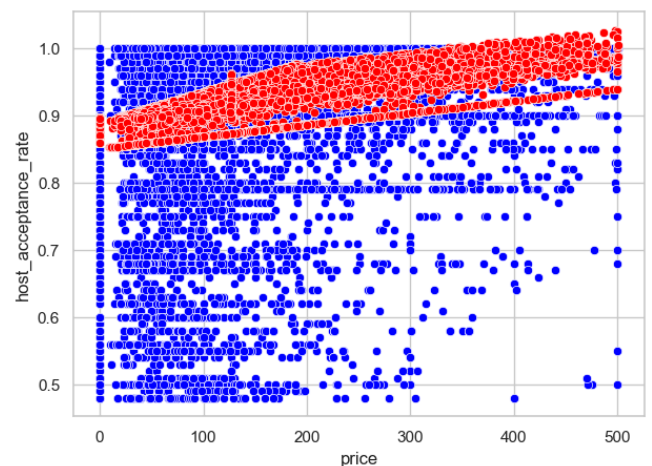
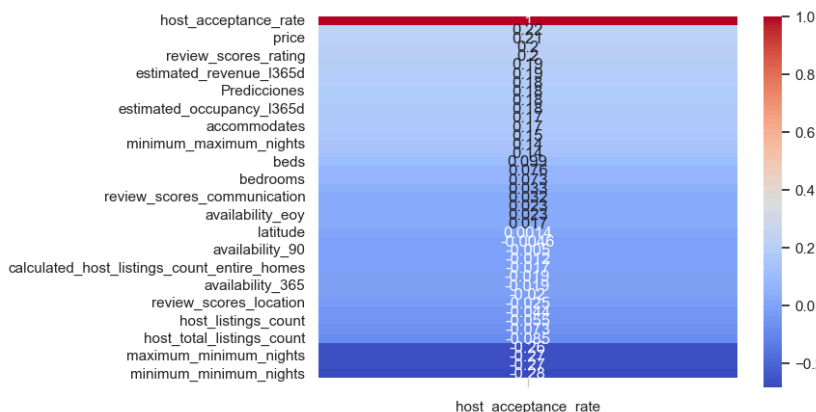
Aunque las correlaciones eran bajas (ninguna pasaba de 0.30), fueron las que tenían un poco más de peso frente a las demás.

### Qué salió en el modelo

- En la gráfica de host\_acceptance\_rate vs price, se ve que la mayoría de puntos están súper dispersos.
- El coeficiente de correlación fue  $R \approx 0.28$ , lo cual es **muy bajo**.
- Hay como una nube de datos en la parte superior (casi todos los hosts aceptan arriba del 90%), y eso hace que no se vea una tendencia clara.

### Interpretación

- La tasa de aceptación de los hosts no depende realmente del precio. O sea, un host que cobra barato o caro puede aceptar mucho o poco, no hay patrón definido.
- Lo mismo pasa con las otras variables: aunque se incluyan varias, el modelo no logra explicar mucho.
- Esto tiene sentido porque la aceptación es más un tema de comportamiento del host si decide aceptar a todos o no, y no tanto de características de la propiedad.
- El modelo múltiple para host\_acceptance\_rate salió débil, con correlaciones bajas. No hay una relación clara entre la tasa de aceptación y variables como el precio o los ingresos. Es más un tema de decisión personal del anfitrión.



## Modelo de regresión múltiple: Host Is Superhost

### Variables elegidas

Del heatmap vi que las más relacionadas con ser superhost fueron:

- **review\_scores\_rating**
- **estimated\_revenue\_l365d**
- **review\_scores\_cleanliness**

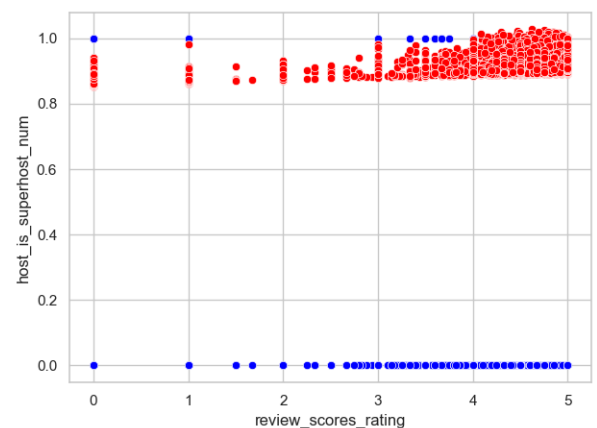
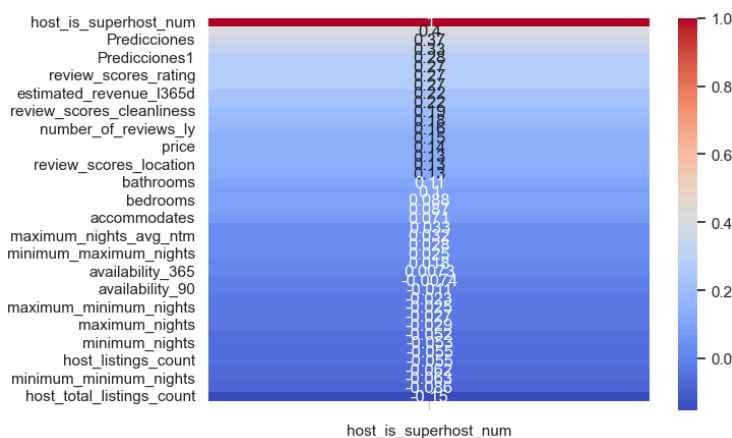
Estas tienen correlaciones un poco más altas que el resto aunque ninguna es súper fuerte.

### Qué salió en el modelo

- En la gráfica de `host_is_superhost_num` vs `review_scores_rating` se ve un patrón interesante:
  - Los que tienen calificación baja (menor a 3) casi nunca son superhosts.
  - A partir de 4 y sobre todo cerca de 5, la mayoría sí son superhosts.
- El coeficiente de correlación fue  $R \approx 0.36$ , lo cual indica una relación débil-moderada.

### Interpretación

- Tiene bastante sentido: para ser superhost, la calificación promedio debe ser alta. Los huéspedes que están contentos califican bien y eso ayuda al host a conseguir el estatus de superhost.
- El ingreso anual (`estimated_revenue_l365d`) también influye, porque hosts con más reservas y más experiencia tienen más chance de ser superhost.
- Sin embargo, el modelo no es perfecto, porque Airbnb también pide otras condiciones para ser superhost (mínimo de estancias completadas, pocas cancelaciones, etc.), y esas reglas no están en las variables que usamos.
- El modelo múltiple de `host_is_superhost` sí logra captar cierta relación con las calificaciones y el ingreso, pero la fuerza de la correlación (0.36) muestra que todavía hay muchos factores externos que no se pueden explicar solo con los datos.



## Modelo de regresión múltiple: Host Total Listings Count

### Variables elegidas

Del heatmap las más relacionadas con host\_total\_listings\_count fueron:

- **host\_listings\_count**
- **minimum\_minimum\_nights**
- **minimum\_nights**

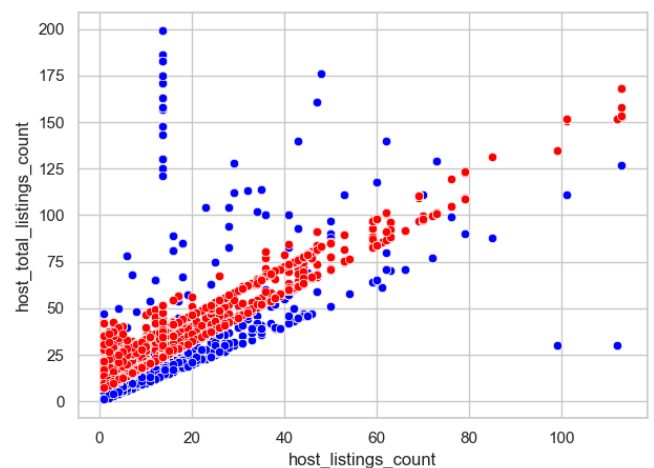
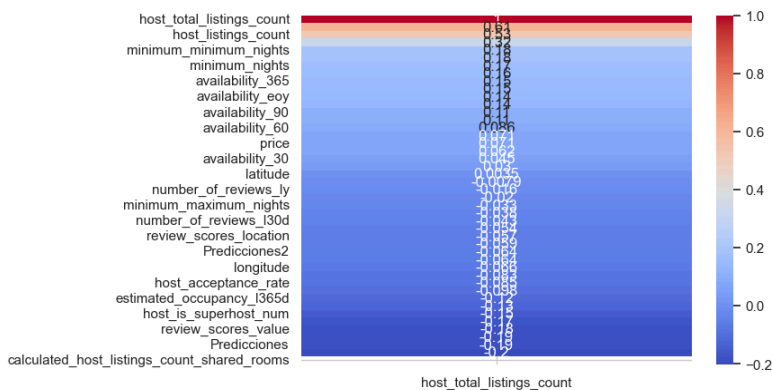
La correlación más fuerte fue claramente con host\_listings\_count, que es casi la misma variable medida de forma distinta.

### Qué salió en el modelo

- En la gráfica de host\_total\_listings\_count vs host\_listings\_count se ve una relación **positiva muy clara**: cuando sube una, sube la otra.
- El coeficiente de correlación fue  **$R \approx 0.55$** , o sea una relación **moderada**.
- Los puntos se alinean bastante bien en una diagonal ascendente, aunque con varios outliers que tienen muchos anuncios y rompen un poco la tendencia.

### Interpretación

- Tiene lógica que host\_total\_listings\_count esté fuertemente ligado a host\_listings\_count, porque ambas variables miden la cantidad de propiedades que maneja un host aunque en diferentes contextos.
- El modelo aquí no es tan útil, porque prácticamente se está prediciendo una variable con otra que dice lo mismo.
- Las variables de noches mínimas aportan muy poco: casi no afectan el total de anuncios que tiene un host.
- El modelo múltiple para host\_total\_listings\_count refleja una relación moderada con host\_listings\_count, pero no es muy informativo ya que ambas variables son casi equivalentes. En este caso se evidencia la multicolinealidad, y conviene elegir solo una de ellas en los modelos.



## Modelo de regresión múltiple: Accommodates

### Variables elegidas

Del heatmap las que más se relacionaban con accommodates fueron:

- **beds**
- **bathrooms**
- **number\_of\_reviews\_ltm**

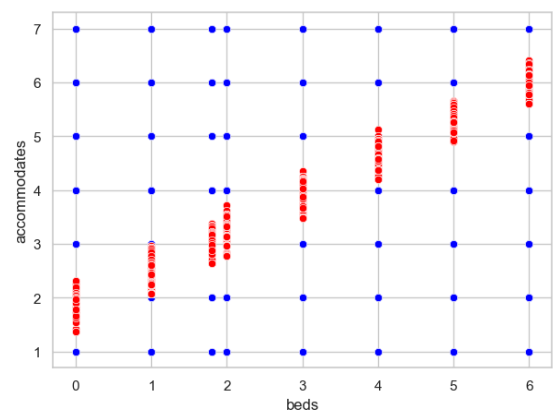
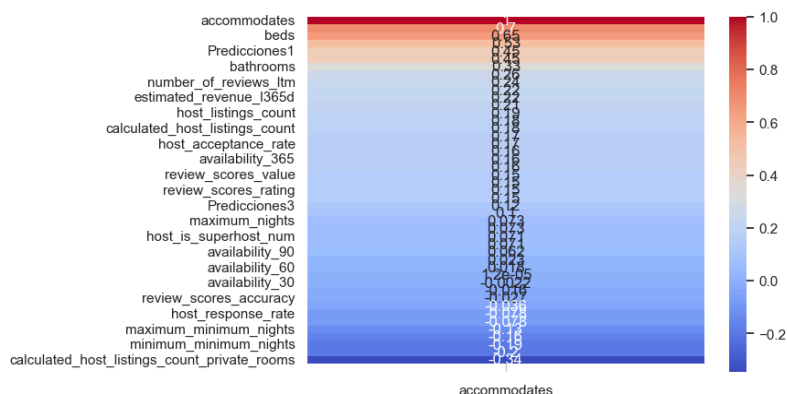
Todas estas son variables lógicas porque describen el tamaño físico de la propiedad.

### Qué salió en el modelo

- En la gráfica de accommodates vs beds se nota una relación positiva bastante clara: a más camas, más personas caben.
- El coeficiente de correlación fue  $R \approx 0.66$ , o sea una relación moderada-fuerte.
- Aunque la tendencia es clara, los puntos están algo “discretos” porque son variables enteras no puedes tener 2.7 camas, por ejemplo.

### Interpretación

- Esto tiene todo el sentido del mundo: la cantidad de personas que un alojamiento puede hospedar (accommodates) depende directamente de cuántas camas y cuántos cuartos tiene.
- La variable bathrooms también ayuda un poco, porque más baños suelen estar en alojamientos más grandes.
- Aquí sí se ve que el modelo múltiple aporta bastante, porque no solo una variable explica accommodates, sino varias juntas que están todas relacionadas con el tamaño.
- El modelo múltiple para accommodates fue sólido, mostrando una correlación clara con beds y bathrooms. El  $R \approx 0.66$  indica que estas variables logran explicar bien cuántas personas caben en un alojamiento, aunque no es una relación perfecta porque también influyen reglas del host.



## Modelo de regresión múltiple: Bedrooms

### Variables elegidas

En el heatmap las variables más relacionadas con bedrooms fueron:

- **beds**
- **price**
- **host\_listings\_count**

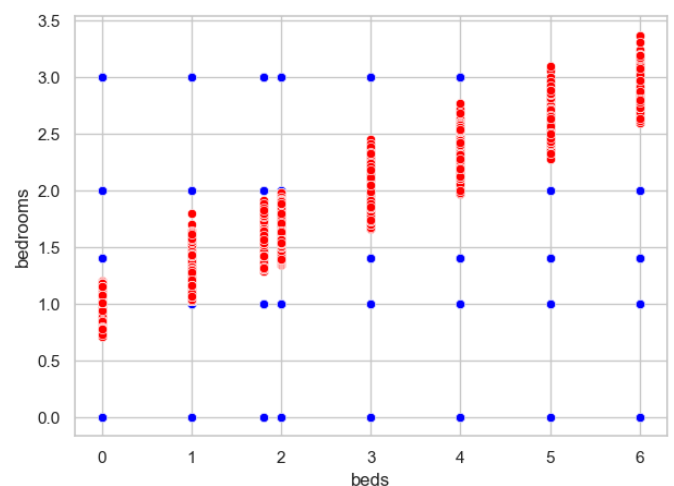
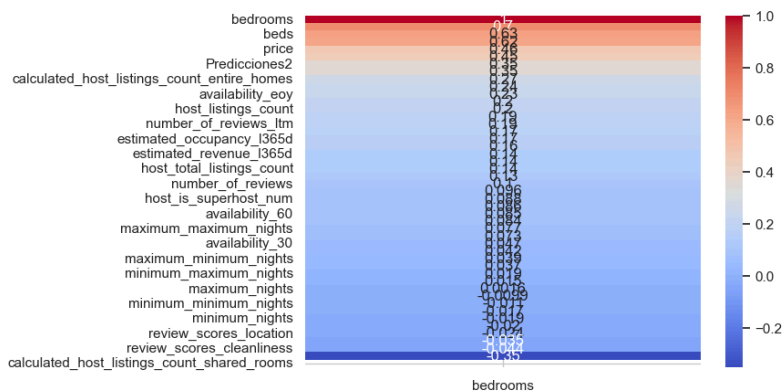
Aunque beds fue la más obvia y fuerte, las otras aportan algo de contexto.

### Qué salió en el modelo

- La gráfica de bedrooms vs beds muestra una **relación positiva muy clara**: a más camas, normalmente hay más recámaras.
- El coeficiente de correlación fue  **$R \approx 0.64$** , lo cual es una relación **moderada-fuerte**.
- Igual que en accommodates, los valores son enteros, así que los puntos se ven como columnas en lugar de una nube continua.

### Interpretación

- Aquí la relación es bastante lógica: el número de camas y de habitaciones van de la mano. Un anuncio con 5 camas raramente tendrá 1 sola recámara.
- El precio también tiende a subir cuando hay más recámaras, aunque no siempre de manera proporcional.
- En cuanto a host\_listings\_count, la relación es mucho más débil, porque la cantidad de recámaras en una propiedad no depende directamente de cuántos anuncios maneja el host.
- El modelo múltiple para bedrooms funciona bien, sobre todo gracias a beds. El  $R \approx 0.64$  confirma que sí hay una relación fuerte entre estas variables, pero no es perfecta porque influyen otros factores como el tipo de propiedad y la distribución interna.



## Modelo de regresión múltiple: Price

### Variables elegidas

En el heatmap, las variables más relacionadas con price fueron:

- **beds**
- **accommodates**
- **bedrooms**

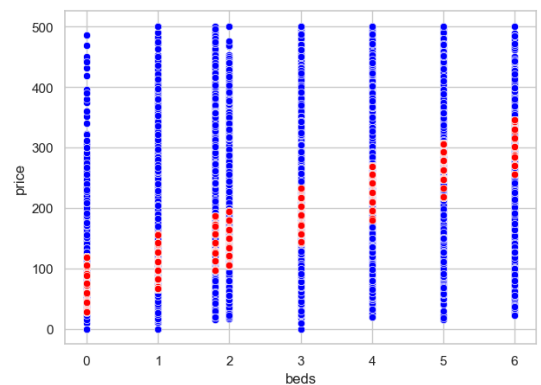
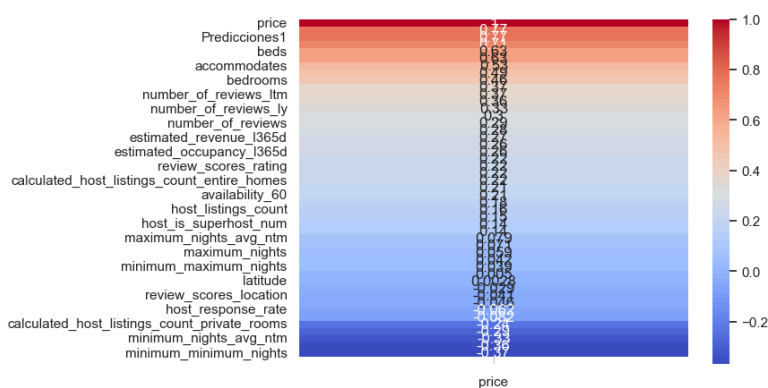
Todas tienen sentido porque el precio suele aumentar cuando el alojamiento es más grande o acomoda a más personas.

### Qué salió en el modelo

- La gráfica de price vs beds muestra que aunque existe cierta tendencia de que a mayor número de camas el precio suba, los puntos están muy dispersos.
- El coeficiente de correlación fue  $R \approx 0.65$ , lo que indica una relación moderada.
- Se ve que para cada número de camas hay precios muy distintos, ejemplo: con 2 camas puedes tener un cuarto económico o un departamento súper caro dependiendo de la ubicación.

### Interpretación

- El número de camas, recámaras y personas que puede alojar sí influyen en el precio, pero no lo explican completamente.
- El modelo muestra que hay factores externos no considerados.
- Eso explica por qué el  $R^2$  no es tan alto: aunque la tendencia general es positiva, la dispersión es muy grande.
- El precio es una variable compleja que depende de más que solo el tamaño del alojamiento. Nuestro modelo capta una parte importante de la relación ( $R \approx 0.65$ ), pero aún deja mucho ruido sin explicar.



## Modelo de regresión múltiple: Review Scores Value

### Variables elegidas

Según el heatmap, las más relacionadas con review\_scores\_value fueron:

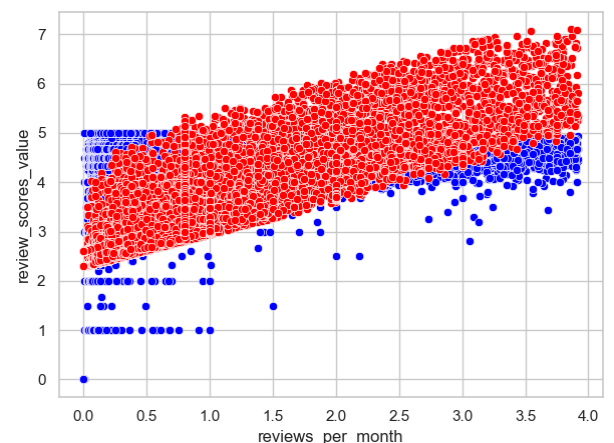
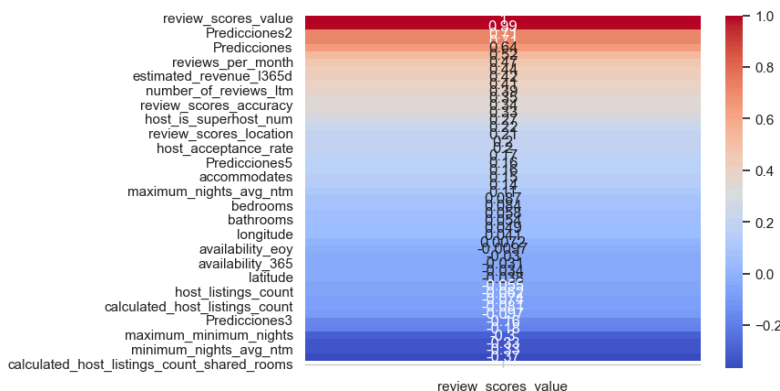
- reviews\_per\_month
- estimated\_revenue\_l365d
- number\_of\_reviews\_ltm

### Qué salió en el modelo

- En la gráfica de reviews\_per\_month vs review\_scores\_value se ve una tendencia positiva: mientras más reseñas recibe un alojamiento por mes, mejor es su calificación en valor.
- El coeficiente de correlación fue  $R \approx 0.52$ , lo que indica una relación moderada, no tan fuerte como en accomodates o price.
- La nube de puntos muestra que aunque la mayoría sigue esa tendencia, hay excepciones, alojamientos con muchas reseñas pero baja puntuación, o con pocas reseñas y excelente puntaje.

### Interpretación

- Esto tiene lógica porque los alojamientos con más reseñas suelen estar más consolidados y ajustan mejor su precio/valor para los clientes.
- Sin embargo, no siempre más reseñas = mejor calificación, ya que depende también de la calidad del servicio, ubicación y expectativas de los huéspedes.
- El  $R \approx 0.52$  confirma que las reseñas explican parte de la valoración, pero no todo, por lo que otros factores externos deben tener un papel clave.
- El modelo para review\_scores\_value muestra que hay una relación positiva con las reseñas y el nivel de actividad del anuncio, pero al no ser tan fuerte, refleja que la percepción de valor es más subjetiva y depende de varios elementos más allá de los números.



## Modelo de regresión múltiple: Bathrooms

### Variables elegidas

Según el heatmap, las más relacionadas con bathrooms fueron:

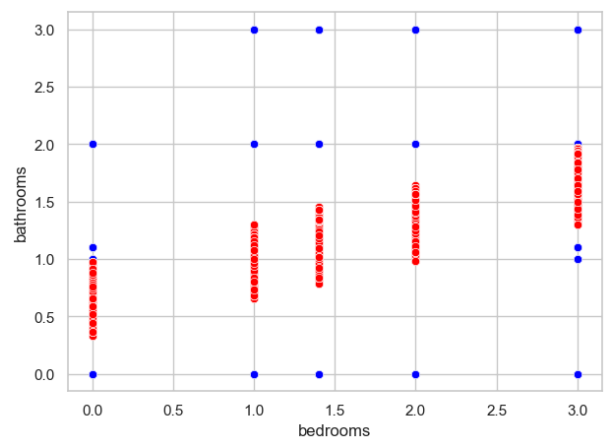
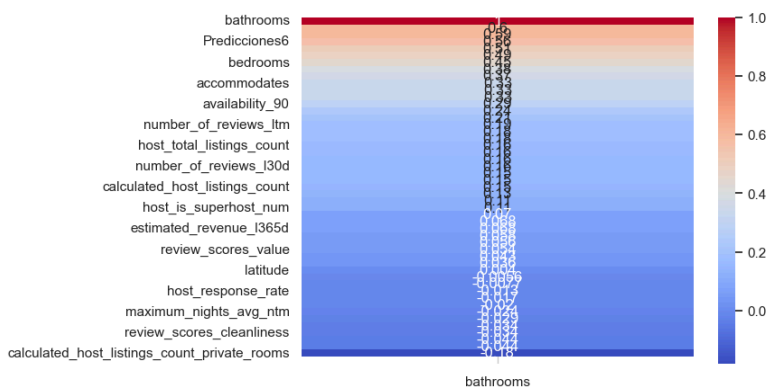
- bedrooms
- accommodates
- availability\_90

### Qué salió en el modelo

- La gráfica de bathrooms vs bedrooms muestra una relación positiva lógica: a mayor número de recámaras, tiende a haber más baños.
- El coeficiente de correlación fue  $R \approx 0.50$ , lo que indica una relación moderada.
- La dispersión no es tan grande como en el caso de price, pero sí hay bastante variabilidad.

### Interpretación

- La relación es esperada: propiedades más grandes suelen tener más baños, pero no siempre.
- Esto depende mucho del diseño de la vivienda. Por ejemplo, un estudio grande puede tener solo un baño aunque aloje a varias personas.
- La variable availability\_90 aparece como correlacionada, aunque esto no es por un vínculo directo, sino porque los alojamientos más grandes suelen estar más disponibles y ser más rentables.
- El modelo múltiple para bathrooms confirma que hay una relación positiva con el tamaño de la propiedad (más cuartos = más baños), aunque no es perfecta. El  $R \approx 0.50$  refleja que existen bastantes excepciones, lo que hace que la predicción sea moderadamente confiable.





## Modelo de regresión múltiple: Reviews per Month

### Variables elegidas

Del heatmap, las más relacionadas con reviews\_per\_month fueron:

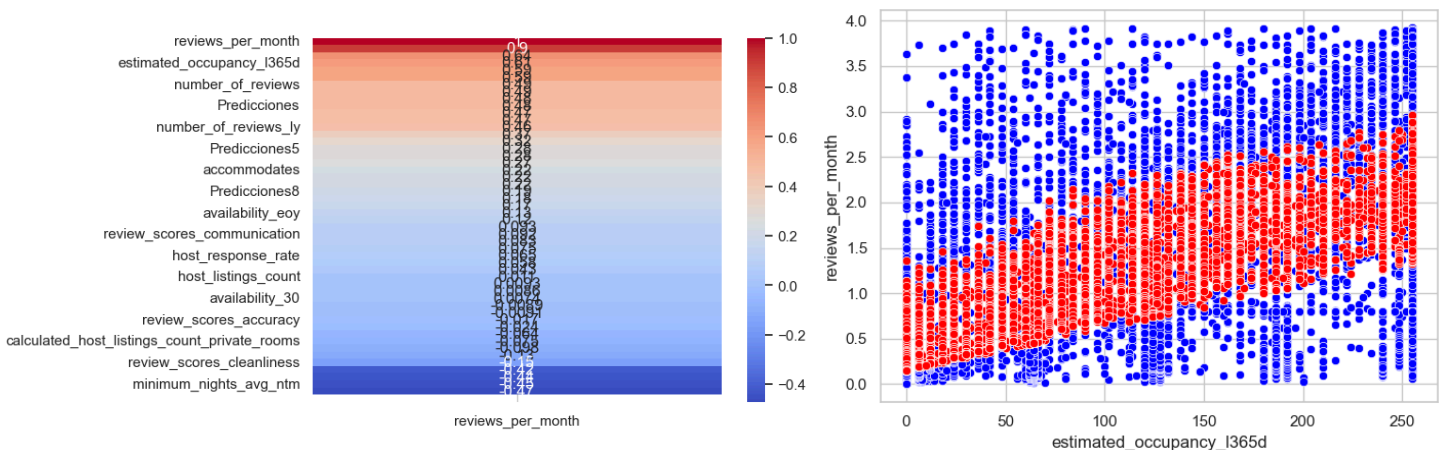
- **estimated\_occupancy\_l365d**
- **number\_of\_reviews**
- **number\_of\_reviews\_ly**

### Qué salió en el modelo

- En la gráfica de reviews\_per\_month vs estimated\_occupancy\_l365d se ve una tendencia positiva clara: mientras más ocupación estimada tiene un alojamiento durante el año, mayor es el número de reseñas que recibe por mes.
- El coeficiente de correlación fue  $R \approx 0.67$ , lo cual representa una relación moderada-fuerte.
- La dispersión es amplia porque no todos los alojamientos con alta ocupación generan el mismo número de reseñas, pero la tendencia general sí es consistente.

### Interpretación

- La relación es muy lógica: más ocupación = más huéspedes = más posibilidad de obtener reseñas.
- Este modelo múltiple tiene mejor ajuste que varios de los anteriores, porque la variable dependiente (reviews\_per\_month) está directamente ligada al nivel de actividad del anuncio.
- Factores externos como la cultura de los huéspedes o la política del host también influyen en la dispersión.
- El modelo múltiple de reviews\_per\_month es de los más sólidos que armamos. Con  $R \approx 0.67$  se confirma que la ocupación y el total de reseñas acumuladas son predictores bastante confiables de cuántas reseñas por mes recibirá un alojamiento.



## Conclusiones Finales

Como conclusión final, todo el proceso me dejó ver que primero fue súper importante limpiar los datos, quitando nulos y outliers porque muchas variables de Airbnb tenían valores extremos que afectaban los análisis. Una vez hecho eso, las regresiones lineales simples me ayudaron a ver relaciones básicas entre pares de variables, aunque en general fueron débiles o moderadas ya que influyen muchos factores externos como la ubicación, la temporada o la reputación del host. Luego, con los heatmaps se confirmó que había variables muy correlacionadas entre sí (multicolinealidad), y eso me sirvió para escoger solo las más relevantes para los modelos múltiples. En la regresión múltiple sí se notó una mejora, sobre todo en variables como *reviews\_per\_month* o *price* que tuvieron correlaciones más fuertes, mientras que en otras como *host\_acceptance\_rate* o *superhost* el ajuste fue más bajo. En general, la regresión múltiple fue más útil que la simple porque permitió explicar mejor las variaciones, pero también quedó claro que el mercado de Airbnb depende de muchos factores externos que no se reflejan en los datos y por eso los modelos no alcanzan una predicción perfecta.