

**Instituto Tecnológico y de Estudios  
Superiores de Monterrey**



**Tecnológico  
de Monterrey**

**Analítica de Datos y Herramientas de Inteligencia Artificial II**

**Actividad 2.1:  
Regresión lineal**

**Alumno:**

Diego Sánchez Márquez

**A01734778**

**Grupo 501**

**06, Abril 2025**

## 1. Hallazgos de los heatmaps por tipo de cuarto

### Entire Home:

Factores de correlación de los conjuntos de variables dependientes e independientes necesarios (Entire Home)

- “host\_acceptance\_rate vs host\_response\_rate” = 0.02
- “review\_scores\_location vs review\_scores\_cleanliness” = 0.30
- “host\_acceptance\_rate vs price” = 0.01
- “availability\_365 vs number\_of\_reviews” = 0.01
- “host\_acceptance\_rate vs number\_of\_reviews ” = 0.05
- “reviews\_per\_month vs review\_scores\_communication” = 0.07

### Private Room:

Factores de correlación de los conjuntos de variables dependientes e independientes necesarios (Private Room)

- “host\_acceptance\_rate vs host\_response\_rate” = 0.03
- “review\_scores\_location vs review\_scores\_cleanliness” = 0.31
- “host\_acceptance\_rate vs price” = 0.03
- “availability\_365 vs number\_of\_reviews” = 0.03
- “host\_acceptance\_rate vs number\_of\_reviews ” = 0.0
- “reviews\_per\_month vs review\_scores\_communication” = 0.15

### Hotel Room:

Factores de correlación de los conjuntos de variables dependientes e independientes necesarios (Hotel Room)

- “host\_acceptance\_rate vs host\_response\_rate” = 0.08
- “review\_scores\_location vs review\_scores\_cleanliness” = 0.37
- “host\_acceptance\_rate vs price” = 0.31
- “availability\_365 vs number\_of\_reviews” = 0.10
- “host\_acceptance\_rate vs number\_of\_reviews ” = 0.07
- “reviews\_per\_month vs review\_scores\_communication” = 0.32

### Shared Room:

Factores de correlación de los conjuntos de variables dependientes e independientes necesarios (Shared Room)

- “host\_acceptance\_rate vs host\_response\_rate” = 0.02
- “review\_scores\_location vs review\_scores\_cleanliness” = 0.31
- “host\_acceptance\_rate vs price” = 0.08
- “availability\_365 vs number\_of\_reviews” = 0.15
- “host\_acceptance\_rate vs number\_of\_reviews ” = 0.10
- “reviews\_per\_month vs review\_scores\_communication” = 0.14

## 2. Modelos de regresión simple por tipo de cuarto

### Entire Home:

Variable dependiente: number\_of\_reviews

Variable independiente: reviews\_per\_month

- El modelo explica el 29% de la variabilidad en la variable dependiente.
- El coeficiente de correlación (0.54) indica una correlación positiva moderada.

Modelo matemático:  $y = 20.84042845x + 17.08358051899713$

### Private Room:

Variable dependiente: reviews\_per\_month

Variable independiente: number\_of\_reviews

- El modelo explica el 35% de la variabilidad en la variable dependiente.
- El coeficiente de correlación (0.59) indica una correlación positiva alta.

Modelo matemático:  $y = 0.01598904x + 0.36692179809869035$

### Hotel Room:

Variable dependiente: reviews\_per\_month

Variable independiente: number\_of\_reviews

- El modelo explica el 56% de la variabilidad en la variable dependiente.
- El coeficiente de correlación (0.75) indica una correlación positiva muy fuerte.

Modelo matemático:  $y = 0.01484072x + 0.01638343271422278$

### Shared Room:

Variable dependiente: reviews\_per\_month

Variable independiente: number\_of\_reviews

- El modelo explica el 46% de la variabilidad en la variable dependiente.
- El coeficiente de correlación (0.68) indica una correlación fuerte.

Modelo matemático:  $y = 0.01754182x + 0.2291584361264406$

## 3. Modelos de regresión múltiple

### host\_id:

Variables independientes: review\_scores\_communication, review\_scores\_checkin, number\_of\_reviews\_l30d, review\_scores\_accuracy y review\_scores\_cleanliness

- El modelo explica el 1.16% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.1078, lo que nos indica una correlación muy baja.

#### **host\_acceptance\_rate:**

Variables independientes: host\_response\_rate, instant\_bookable, host\_identity\_verified, review\_scores\_rating y property\_type

- El modelo explica el 2.27% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.15, lo que nos indica una correlación positiva muy baja.

#### **host\_is\_superhost:**

Variables independientes: host\_response\_rate, host\_has\_profile\_pic, host\_identity\_verified, instant\_bookable, review\_scores\_accuracy

- El modelo explica el 11.64% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.34, lo que nos indica una correlación positiva baja.

#### **host\_total\_listings\_count:**

Variables independientes: host\_acceptance\_rate, property\_type, room\_type, instant\_bookable, accommodates

- El modelo explica el 3.82% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.19, lo que nos indica una correlación muy baja.

#### **room\_type:**

Variables independientes: property\_type, accommodates, price, instant\_bookable, review\_scores\_value

- El modelo explica el 32.8% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.57, lo que nos indica una correlación alta entre las variables.

#### **accommodates:**

Variables independientes: bedrooms, price, property\_type, room\_type, review\_scores\_value

- El modelo explica el 46.03% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.67, lo que nos indica una correlación positiva alta entre las variables.

#### **bedrooms:**

Variables independientes: accommodates, price, review\_scores\_rating, review\_scores\_cleanliness, room\_type

- El modelo explica el 40.86% de la variabilidad en la variable dependiente.

- El coeficiente de correlación es 0.63, lo que nos indica una correlación alta entre las variables..

#### **price:**

Variables independientes: accommodates, bedrooms, room\_type, property\_type, review\_scores\_value

- El modelo explica el 14.41% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.37, lo que nos indica una correlación baja entre las variables del modelo.

#### **review\_scores\_value:**

Variables independientes: review\_scores\_cleanliness, review\_scores\_communication, review\_scores\_location, review\_scores\_accuracy, price

- El modelo explica el 57.21% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.75, lo que nos indica una correlación muy alta entre las variables del modelo.

#### **reviews\_per\_month:**

Variables independientes: number\_of\_reviews\_l30d, review\_scores\_value, review\_scores\_cleanliness, review\_scores\_accuracy, room\_type

- El modelo explica el 19.68% de la variabilidad en la variable dependiente.
- El coeficiente de correlación es 0.44, lo que nos indica una correlación moderada entre las variables de este modelo.

### **Conclusiones generales**

- Las variables de review\_scores\_cleanliness y review\_scores\_location muestran una alta correlación entre sí, especialmente al filtrar los tipos de cuarto en Shared Room y Hotel Room.
- reviews\_per\_month y number\_of\_reviews presentan una fuerte correlación en todos los tipos de habitaciones.
- La combinación de variables de review\_scores\_location y review\_scores\_cleanliness mostraron ser el par de variables sugeridos que tienen la correlación más alta en todos los tipos de cuarto.
- La variable price se ve influida principalmente por accommodates, bedrooms y el tipo de habitación.
- room\_type influye en la forma en que otras variables se relacionan entre sí, siendo de las variables que más impacto tuvo en la mayoría de los modelos.
- También se pudo observar que host\_identity\_verified e instant\_bookable son relevantes a la hora de predecir modelos relacionados con el comportamiento del host.
- En general, la mayoría de los modelos resultaron ser no estadísticamente relevantes según sus muy bajos coeficientes de correlación y determinación.
- Sin embargo, destacó el modelo de review\_scores\_value, ya que arrojó los coeficientes de determinación (57.21%) y correlación (0.75) más altos de todo el análisis. Esto nos indica que las variables de review\_scores\_cleanliness,

review\_scores\_communication, review\_scores\_location, review\_scores\_accuracy y price tienen relevancia a la hora de predecir la variable objetivo.

- En conclusión, debido a los resultados tan bajos arrojados por los modelos, será necesario realizar un análisis más profundo para poder explicar y predecir las variables de mi base de datos.