

Reporte Actividad 6.1 Regresión Lineal

Luis Alberto Mirón

A01735489

Se realizaron 48 modelos de regresión lineal entre 6 pares de variables para los datasets de Airbnb de México DF y California. A continuación se presenta un análisis de los resultados por ciudad y un análisis comparativo.

México DF/CDMX

Después de realizar 24 modelos de regresión lineal para los tipos de cuartos **Entire home/apt** y **private room** para los datos de la ciudad de México podemos observar que ninguna de las comparaciones muestra una relación lineal fuerte o un buen ajuste del modelo de regresión, ya que todos los valores de los coeficientes son menores a 0.5. La mayoría de las comparaciones tienen valores muy bajos o nulos, lo que sugiere que las variables no tienen una relación lineal clara o que hay otros factores que influyen en la variación de las variables. Es posible que existan relaciones más significativas entre otras variables del conjunto de datos o que existan modelos de otra naturaleza que sean más adecuados para nuestros datos.

La comparación que tiene el mayor coeficiente de correlación y determinación es la de `review_scores_location` vs `review_scores_cleanliness` para el tipo de cuarto **Private room**, con valores de 0.342 y 0.117 respectivamente. Esto indica que hay una correlación positiva moderada entre la puntuación de la ubicación y la puntuación de la limpieza para los cuartos privados en México. Sin embargo, el coeficiente de determinación indica que sólo el 11.7% de la variación en la puntuación de la limpieza puede explicarse por la variación en la puntuación de la ubicación, lo que implica que hay otros factores que afectan a la puntuación en limpieza.

California

Se llevaron a cabo 24 modelos de regresión lineal para los tipos de cuartos **Entire home/apt** y **private room** para los datos de Airbnb de la ciudad de California. Los valores obtenidos de coeficiente de correlación y determinación indican que ninguna de las comparaciones muestra una relación lineal fuerte o un buen ajuste del modelo de regresión, ya que todos los valores de los coeficientes son menores a 0.5. La mayoría de las comparaciones tienen valores muy bajos o nulos, lo que sugiere que las variables no tienen una relación lineal clara o que hay otros factores que influyen en la variación de las variables. La comparación que tiene el mayor coeficiente de correlación y determinación es la de `review_scores_location` vs `review_scores_cleanliness` para el tipo de cuarto **Entire home/apt**, con valores de 0.274 y 0.075 respectivamente. Esto indica que hay una correlación positiva moderada entre la puntuación de la ubicación y la puntuación de la limpieza para las casas/apartamentos enteros en California, es decir, que los apartamentos enteros con mejor ubicación tienden a tener mejor limpieza. Sin embargo, el coeficiente de determinación indica que sólo el 7.5% de la variación en la puntuación de la limpieza puede explicarse por la variación en la puntuación de la ubicación, lo que implica que hay otros factores que afectan la relación entre estas dos variables en California.