

Reporte Actividad 6.2 Regresión Lineal Múltiple

Luis Alberto Mirón

A01735489

En este reporte se describen los resultados de los 24 modelos de regresión lineal múltiple elaborados para los datos de la plataforma Airbnb de 3 ciudades (México DF, California, Girona).

México DF/CDMX

Para la Ciudad de México se elaboraron 8 modelos de regresión lineal múltiple para 2 tipos de habitaciones: (Entire home/apt y Private room). En estos modelos se exploró la relación de ciertas variables seleccionadas. De manera general podemos decir que los coeficientes de determinación de los modelos realizados para ambos tipos de habitaciones fueron bajos. El coeficiente de determinación más alto fue obtenido en el modelo que tomaba a la variable **number_of_reviews** como dependiente y las variables `review_scores_checkin`, `review_scores_communication`, `review_scores_cleanliness` `review_scores_rating` como independientes al analizar los cuartos privados. El coeficiente obtenido en este modelo fue de 0.116. Esto quiere decir que el 11.7% de la variación en el número de reseñas puede explicarse por la variación de las variables `review_scores_checkin`, `review_scores_communication`, `review_scores_cleanliness` `review_scores_rating`.

El siguiente modelo con el coeficiente de correlación más alto fue aquel en el que se tomó la variable **price** como dependiente y como variables independientes `availability_365`, `host_response_rate`, `number_of_reviews`, `accommodates`. El valor para el coeficiente fue de 0.1137.

California

Si hablamos de la ciudad de California podemos observar un comportamiento similar que en México DF. Para los 8 modelos de regresión lineal múltiple realizados para 2 tipos de habitaciones: (Entire home/apt y Private room) se obtuvieron resultados poco significativos en los valores de los coeficientes de correlación. El modelo con el coeficiente más alto obtenido fue en el que se tomó la variable **price** como la variable dependiente y las variables `availability_365`, `host_response_rate`, `number_of_reviews`, `accommodates` como independientes. El valor del coeficiente de correlación para este modelo fue de 0.19 De tal manera que el 19% de la variación del precio puede explicarse por la variación de `availability_365`, `host_response_rate`, `number_of_reviews`, `accommodates`.

El siguiente modelo más significativo fue el que tomaba a la variable **number_of_reviews** como variable dependiente y las variables `review_scores_checkin`, `review_scores_communication`, `review_scores_value` y `review_scores_accuracy` como variables independientes para los datos de Casas/apartamentos completos. El coeficiente de correlación de este modelo fue de 0.1463. Esto quiere decir que únicamente el 14.63% de la variación en el número de reseñas puede explicarse por la variación de las variables `review_scores_checkin`, `review_scores_communication`, `review_scores_value` y `review_scores_accuracy`.

Girona

Finalmente, para la ciudad de Girona se elaboraron 8 modelos de regresión lineal múltiple para dos tipos de habitaciones (Entire home/apt y Private room). Al igual que los otros modelos, podemos decir que los valores de los coeficientes de determinación son poco significativos. El modelo con el coeficiente de determinación más alto fue aquel en el que se utilizó la variable **number_of_reviews** como variable dependiente y las variables `review_scores_accuracy`, `review_scores_value`, `review_scores_location` y `review_scores_rating` como variables independientes para el tipo de cuarto privado. El coeficiente de determinación obtenido para esta variable fue de 0.2101. Esto quiere decir que el 21.01% de la variación en el número de reseñas puede ser explicada por la variación de `review_scores_accuracy`, `review_scores_value`, `review_scores_location` y `review_scores_rating`. El siguiente modelo más significativo fue en el que se propuso a la variable **price** como variable dependiente y a las variables `availability_365`, `host_response_rate`, `number_of_reviews`, `accommodates` como variables independientes. El coeficiente de correlación para este modelo fue de 0.167.