Fernández Montes Mariel Quetzali / A01707797

Torres Pacheco José Samuel /

Actividad 6.2 (Regresión Lineal Múltiple)

Se realizó el análisis de regresión lineal múltiple de los documentos “California”, “CDMX” y “España” con datos ya limpios, sin valores nulos ni outliers.

Utilizamos el filtro de 2 tipos de habitaciones “Entire home/apt” y “Private room” para ambos documentos y en ambos casos se ejecutó el modelo de regresión lineal múltiple con el fin de determinar la correlación más afianzada que pueda existir entre un conjunto de variables. Las variables analizadas fueron las siguientes:

* host\_acceptance\_rate vs host\_response\_rate, price, number\_of\_reviews, review\_scores\_rating
* price vs availability\_365, host\_response\_rate, number\_of\_reviews, accommodates
* host\_acceptance\_rate vs number\_of\_reviews, review\_scores\_cleanliness, nreview\_scores\_communication

Mencionaremos los principales hallazgos de cada documento tomando en cuenta los coeficientes de correlación y de correlación, para que la explicación sea más clara detallamos el significado de cada término anteriormente mencionado:

* Coeficiente de determinación: Nos ayuda a medir qué tan bueno es el modelo. Proporción de varianza representada por el modelo y se representa en %, por lo que, entre más cerca estemos de 100 más varianza de la variable dependiente conseguimos explicar con el modelo y será un modelo mejor.
* Coeficiente de correlación: Nos dice que tan fuerte es una correlación lineal entre 2 variables. Que tan conveniente es hacer un ajuste lineal para explicar una variable a partir de la otra. Entre más se acerque a 1 el resultado, la relación entre las variables es mayor, entre más se acerque a 0, la relación entre las variables es menor.

Al analizar los distintos escenarios entre los modelos de cada ciudad, tuvimos varios descubrimientos:

1. En general, el coeficiente de determinación para todas las combinaciones de variables en Ciudad de México son más altos que en las otras dos ciudades. Esto sugiere que las variables independientes seleccionadas pueden tener una relación más fuerte con la variable dependiente en Ciudad de México que en las otras dos ciudades.

2. En Ciudad de México, el coeficiente de correlación r es más alto para las variables de tipo Ej1 (host\_acceptance\_rate vs host\_response\_rate, Price, number\_of\_reviews , review\_scores\_rating) y Ej3 (host\_acceptance\_rate vs number\_of\_reviews,review\_scores\_cleanliness,review\_scores\_communication) en comparación con la ciudad de California y Girona. Esto indica que estas variables tienen una relación más lineal con la variable dependiente en Ciudad de México

3. Por otro lado, en California, el coeficiente de determinación para la variable Ej1 es más alto que en Ciudad de México y Girona. Esto sugiere que estas variables independientes pueden tener una relación más fuerte con la variable dependiente en California.

Estas son solo algunas observaciones generales basadas en los resultados generados hasta el momento. Es importante tener en cuenta que estos resultados son preliminares y que se requiere un análisis más detallado para comprender mejor las relaciones entre las variables en cada ciudad, incluyendo por ejemplo resultados de regresión y otras pruebas estadísticas que nos ayuden a comprender que tanto influye cada variable.