Fernández Montes Mariel Quetzali

Torres Pacheco Samuel

Actividad 6.1 (Regresión Lineal Simple)

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

Se realizó 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 simple 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
* review\_scores\_location vs review\_scores\_cleanliness
* host\_acceptance\_rate vs price
* availability\_365 vs number\_of\_reviews
* host\_acceptance\_rate vs number\_of\_reviews
* reviews\_per\_month vs review\_scores\_communication

Realizando la comparación de los resultados obtenidos de ambos documentos, podemos concluir que de forma general la correlación entre las variables analizadas fue mínima, ninguna sobrepasando el 70% de eficiencia del modelo, obteniendo por consecuencia una correlación mínima.

Las variables con mayor correlación fueron host\_acceptance\_rate vs host\_response\_rate para ambos documentos analizados específicamente en el tipo de residencia “Entire home/apt”, obteniendo un coeficiente de determinación mayor al 45% en ambos casos, un coeficiente de correlación mayor a .6. Para tener claros los términos anteriores se explican a continuación:

* 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: 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.

Por último, las variables que tuvieron un menor puntaje entre su correlación fueron el conjunto de availability\_365 vs number\_of\_reviews obteniendo menos del 1% en su coeficiente de determinación y menos de un .1 del coeficiente de correlación.