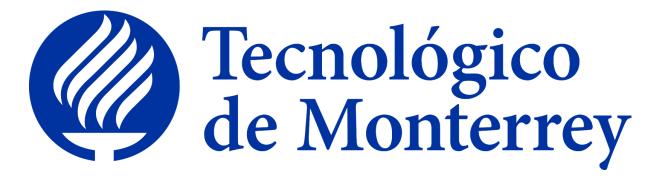
Tecnológico de Monterrey Campus Puebla



Analítica de datos y herramientas de inteligencia artificial I (Gpo 101)

Actividad 2.1 (Regresión Lineal)

Ximena Italya Jiménez Huerta A01277667 Tomando en cuenta los resultados obtenidos de los 4 códigos, pude darme cuenta que en realidad hay muchas variables que tenen correlación con muchas variables del DataFrame, pero hay otras que no tienen para nada de correlación con bastantes, esto depende del filtro que usemos, en este caso fue de room_type que sólo cuenta con 4 categorías, rellené cada una de las columnas que tenían datos 'object' o string en todo el archivo, después pude sacar la correlación con cada una de las columnas dependiendo del filtro de room_type, e cuanto a:

"host_acceptance_rate vs host_response_rate"

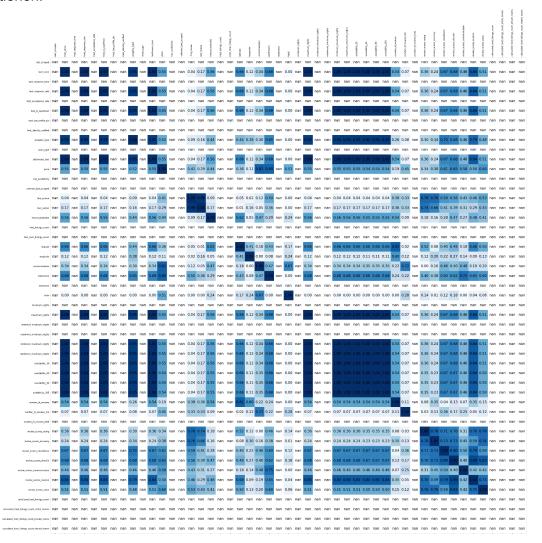
"review_scores_location vs review_scores_cleanliness"

"host_acceptance_rate vs price"

"availability_365 vs number_of_reviews"

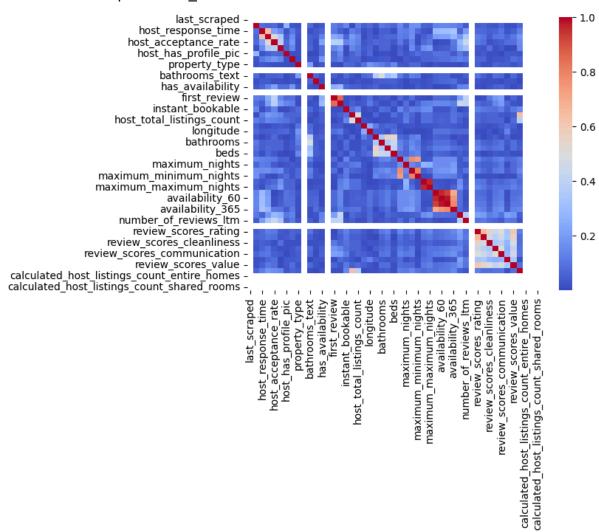
"host_acceptance_rate vs number_of_reviews" "reviews_per_month vs review_scores_communication"

podemos notar que dependiendo de la variable filtrada será el tipo de correlación que tienen.

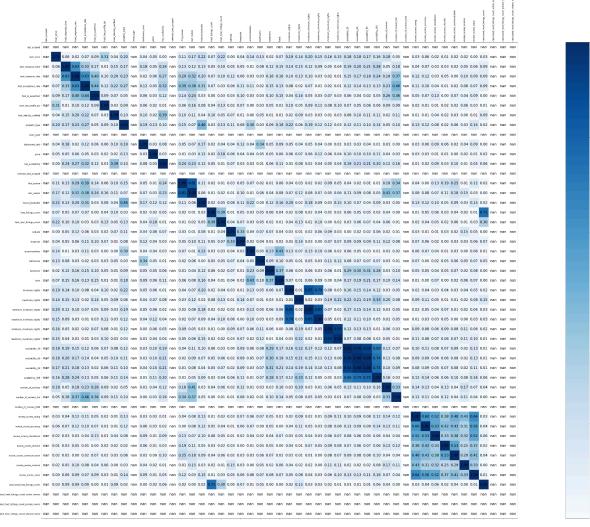


Esta imagen mostrada pertenece a 'Hotel_room' podemos observar que tiene 0 relación con 'host_aceptance_rate' y 'host_response_rate' tiene alta correlación con review_scores_location.

Este es el Heatmap de Entire home:

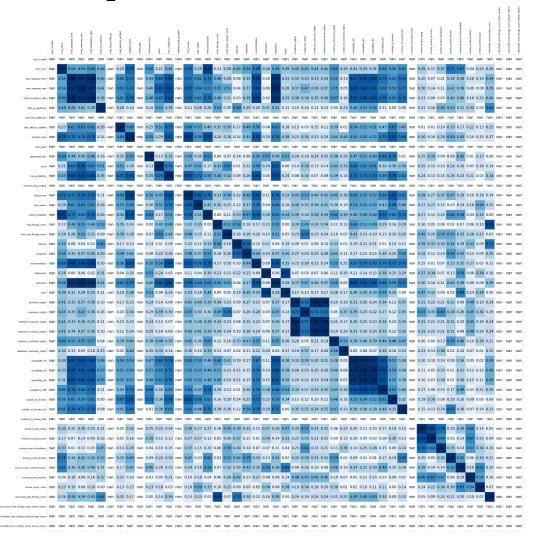


A simple vista observamos que no tiene tantos valores nulos como Hotel room.



Esta es Heatmap de 'Private_room' que no tiene tanta correlación con las demás variables en general.

Y al final 'Shared_room':



Notamos que tiene una correlación media con las variables.

En conclusión, podemos decir que todas los mapas cuentan con muchos valores nulos, tendría que haber una comparación con Heatmaps que hayan eliminado las columnas categóricas para analizar que tanto afecta el dejarlas con eliminarlas, pero en general no hay mucha correlación.