

Actividad 6.1 (Regresión Lineal Simple)

En este programa hicimos un modelo matemático para tener la correlación de diferentes variables usando bases de datos filtradas por un tipo de cuarto en específico. Esto fue uno a uno para tener las variables que más relación se tienen unas con otras.

Primeramente descargamos todas las bases de datos en las cuales tenemos diferentes bases de datos con Airbnb (DF, Girona y California) de la cual limpiamos las bases de datos en cuanto a datos nulos o variables sin ningún tipo de registro. Hacemos el análisis de eliminación de datos extremos con el rango intercuartílico así como eliminar simbologías en las columnas que estudiaran para que se pueda hacer todo el análisis numérico que se tiene.

Una vez que tenemos listas las bases de datos, lo que hacemos es crear una base de datos filtrada por el tipo de cuarto para así poder hacer el estudio de cada una de las variables que nos piden. Una vez que se tienen estas bases de datos definimos cuales son nuestras variables dependientes e independientes acorde a los 6 parámetros que se nos dio en los cuales podemos ver los diversos comportamientos de las variables en cada una de las bases de datos.

México: Como podemos observar en el caso de México el que tiene mayor correlación en los dos tipos de habitación con la aceptación del anfitrión o la respuesta del mismo, considerando que aún y cuando la correlación es baja es positiva y si pueden llegar a ser relevantes en el modelo pero aún así todas las variables que se tienen en modelo tienen un nivel de R^2 muy bajo por lo cual la relación no es lineal en la mayor parte de los casos. Tenemos variables con niveles tan bajos casi llegando al 0 que podemos decir que esas correlaciones son nulas a un 100%. Vemos variables que realmente no tienen ningún tipo de relación como el precio con la aceptación del anfitrión, disponibilidad 365 con el número de reseñas por parte de los huéspedes con las reseñas por comunicación.

	Determinación	Correlación
0	0.121342	0.348342
1	0.085970	0.293207
2	0.000022	0.004733
3	0.003813	0.061750
4	0.020555	0.143369
5	0.008707	0.093310
6	0.125335	0.354027
7	0.174424	0.417641
8	0.000014	0.003762
9	0.005180	0.071970
10	0.022724	0.150744
11	0.071941	0.268219

(Del 0 al 5 son de "Entire Room" y los del 6 al 11 son de "Private Room". El orden de 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, reviews per month vs review scores communication)

Renata Emilia Chávez Martínez
A01351716

California: Vemos un caso muy parecido en California al de México, aún y cuando tenemos coeficientes levemente superiores aún así ninguna de las variables son realmente representativas como para que la variable independiente pueda determinar el comportamiento de la variable independiente en todos los casos en los que vemos. Tenemos un comportamiento similar para todas las variables en las cuales no podemos decir que una representante más que otra.

	Determinación	Correlación
0	0.037843	0.194533
1	0.091715	0.302844
2	0.000611	0.024720
3	0.010922	0.104509
4	0.007167	0.084660
5	0.053915	0.232196
6	0.013741	0.117220
7	0.058813	0.242513
8	0.001609	0.040116
9	0.011497	0.107223
10	0.015799	0.125694
11	0.049061	0.221496

(Del 0 al 5 son de "Entire Room" y los del 6 al 11 son de "Private Room". El orden de 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, reviews per month vs review scores communication)

Girona: Finalmente en la última base de datos tenemos un comportamiento casi igual que en las pasadas considerando se compararon las mismas variables y tengamos en cuenta que fueron variables preseleccionadas las cuales queríamos ver su comportamiento si se hacía la comparación.

	Determinación	Correlación
0	0.082443	0.287128
1	0.084508	0.290703
2	0.000130	0.011382
3	0.002316	0.048120
4	0.000466	0.021577
5	0.021485	0.146578
6	0.126501	0.355669
7	0.150319	0.387710
8	0.000633	0.025166
9	0.000002	0.001428
10	0.003598	0.059986
11	0.068295	0.261332

(Del 0 al 5 son de "Entire Room" y los del 6 al 11 son de "Private Room". El orden de 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, reviews per month vs review scores communication)

Finalmente podemos ver que los coeficientes de evaluación son bajos sin embargo si son positivamente correlacionados esto nos da entender que si bien no son las variables más correlacionadas o perfectamente correlacionadas podríamos escoger otras variables que muestran un mejor comportamiento que los análisis.