

Para trabajar con los datos sin importar el país lo primero que se hizo fue realizar la limpieza de la base de datos y la separación, inicialmente se borraron las columnas del DataFrame que no nos importaban. Después se filtraron los datos y se comenzó con el tratamiento de los datos faltantes, sustituyéndolos por los promedio y los outliers, nuevamente sustituyéndolos por los promedios. Ahora si los datos están listos para comenzar con los análisis de correlación para realizar los modelos de regresión ya sea lineal o multiple.

MEX

para realizar el modelo de regresión lineal simple, se eligió la variable independiente que tuviera más correlación con la variable dependiente, en este caso las correlaciones no eran muy altas, sin embargo para el caso de entire home fue `review_scores_rating` con una correlación o una r de $.0.015$. Una vez realizado este mismo procedimiento para todos los tipos de cuarto se obtuvieron se agregó una columna donde se estaban colocando las predicciones para poder comparar. Además, se sacó el r cuadrada o el coeficiente de determinación de 0.0009 lo cual indica porcentualmente que tan representativo es el modelo a comparación con la realidad en este caso el modelo no va a dar predicciones correctos por lo que lo recomendable sería descartarlo y encontrar otro tipo de tratamiento para los datos u otro tipo de modelo que se acerque más a los datos. También es importante considerar que existen muchos otros factores que posiblemente no se estén contabilizando tales como la época, los vecinos, de donde son normalmente los clientes. etce.

OTROS PAÍSES

Para el caso de Brasil y Swiza se hicieron análisis similares sin embargo la regresión en este caso se hicieron múltiples en vez de regresiones lineales, el criterio de selección fueron las mismas. Para no tener un modelo tan complejo se eligieron únicamente las variables con correlaciones relativamente altas a comparación de las demás para así tener buena representación. En estos casos y para comparar el mismo tipo de cuarto en los tres países se obtuvieron los siguientes coeficientes de determinación (r^2) y de correlación (r) para PrivateHome en:

Brasil:

Coeficiente de Determinación: 0.09929660110831229
Coeficiente de Correlación: 0.31511363205725057

Swiza:

Coeficiente de Determinación: 0.04691614677506184
Coeficiente de Correlación: 0.21660135450883458

GENERAL

Se puede observar que no es necesariamente un tema cultural, y que para analizar datos y proponer modelos es super importante que los datos lleven el tratamiento

adecuado previo a comenzar la manipulación con los modelos. Se tiene tambien que tomar en consideración todos los factores posibles para tener un buen entendimiento de los datos y poder tener modelos representativos que sean importantes y/o controlables para poder determinar ciertos aspectos.

Entire Home México

Análisis de correlación (r)

	Predicciones_EntireHome	host_acceptance_rate	review_scores_rating	price	review_scores_cleanliness	availability_365	review_scores_communication	room_type	number_of_reviews
Predicciones_EntireHome	1.000000	0.030007	-0.120710	-0.058859	-0.207912	-0.072600	-0.155502	NaN	1.000000
host_acceptance_rate	0.030007	1.000000	-0.010317	0.030287	0.020066	0.064165	-0.016215	NaN	0.030007
review_scores_rating	-0.120710	-0.010317	1.000000	0.123884	0.491470	-0.087712	0.538377	NaN	-0.120710
price	-0.058859	0.030287	0.123884	1.000000	0.114736	0.041327	0.059271	NaN	-0.058859
review_scores_cleanliness	-0.207912	0.020066	0.491470	0.114736	1.000000	-0.005057	0.343191	NaN	-0.207912
availability_365	-0.072600	0.064165	-0.087712	0.041327	-0.005057	1.000000	-0.090068	NaN	-0.072600
review_scores_communication	-0.155502	-0.016215	0.538377	0.059271	0.343191	-0.090068	1.000000	NaN	-0.155502
room_type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
number_of_reviews	1.000000	0.030007	-0.120710	-0.058859	-0.207912	-0.072600	-0.155502	NaN	1.000000

Análisis de determinación (r~2)

	Predicciones_EntireHome	host_acceptance_rate	review_scores_rating	price	review_scores_cleanliness	availability_365	review_scores_communication	room_type	number_of_reviews
Predicciones_EntireHome	1.000000	0.000900	0.014571	0.003464	0.043227	0.005271	0.024181	NaN	1.000000
host_acceptance_rate	0.000900	1.000000	0.000106	0.000917	0.000403	0.004117	0.000263	NaN	0.000900
review_scores_rating	0.014571	0.000106	1.000000	0.015347	0.241543	0.007693	0.289850	NaN	0.014571
price	0.003464	0.000917	0.015347	1.000000	0.013164	0.001708	0.003513	NaN	0.003464
review_scores_cleanliness	0.043227	0.000403	0.241543	0.013164	1.000000	0.000026	0.117780	NaN	0.043227
availability_365	0.005271	0.004117	0.007693	0.001708	0.000026	1.000000	0.008112	NaN	0.005271
review_scores_communication	0.024181	0.000263	0.289850	0.003513	0.117780	0.008112	1.000000	NaN	0.024181
room_type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
number_of_reviews	1.000000	0.000900	0.014571	0.003464	0.043227	0.005271	0.024181	NaN	1.000000

Entire Home Brasil

Análisis de correlación (r)

	Predicciones_EntireHome	host_acceptance_rate	review_scores_rating	price	review_scores_cleanliness	availability_365	review_scores_communication	room_type	number_of_reviews
Predicciones_EntireHome	1.000000	0.472598	-0.755915	-0.201467	-0.449710	-0.543707	-0.370022	NaN	0.315114
host_acceptance_rate	0.472598	1.000000	-0.026202	-0.101639	0.005280	-0.115269	-0.084198	NaN	0.148922
review_scores_rating	-0.755915	-0.026202	1.000000	0.117466	0.523362	0.040126	0.401879	NaN	-0.238199
price	-0.201467	-0.101639	0.117466	1.000000	0.045982	0.163409	0.145417	NaN	-0.209116
review_scores_cleanliness	-0.449710	0.005280	0.523362	0.045982	1.000000	-0.001216	0.252543	NaN	-0.141710
availability_365	-0.543707	-0.115269	0.040126	0.163409	-0.001216	1.000000	0.085163	NaN	-0.171329
review_scores_communication	-0.370022	-0.084198	0.401879	0.145417	0.252543	0.085163	1.000000	NaN	-0.368459
room_type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
number_of_reviews	0.315114	0.148922	-0.238199	-0.209116	-0.141710	-0.171329	-0.368459	NaN	1.000000

Análisis de determinación (r^2)

	Predicciones_EntireHome	host_acceptance_rate	review_scores_rating	price	review_scores_cleanliness	availability_365	review_scores_communication	room_type	number_of_reviews
Predicciones_EntireHome	1.000000	0.223349	0.571407	0.040589	0.202239	0.295617	0.136916	NaN	0.099297
host_acceptance_rate	0.223349	1.000000	0.000687	0.010330	0.000028	0.013287	0.007089	NaN	0.022178
review_scores_rating	0.571407	0.000687	1.000000	0.013798	0.273908	0.001610	0.161507	NaN	0.056739
price	0.040589	0.010330	0.013798	1.000000	0.002114	0.026703	0.021146	NaN	0.043730
review_scores_cleanliness	0.202239	0.000028	0.273908	0.002114	1.000000	0.000001	0.063778	NaN	0.020082
availability_365	0.295617	0.013287	0.001610	0.026703	0.000001	1.000000	0.007253	NaN	0.029354
review_scores_communication	0.136916	0.007089	0.161507	0.021146	0.063778	0.007253	1.000000	NaN	0.135762
room_type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
number_of_reviews	0.099297	0.022178	0.056739	0.043730	0.020082	0.029354	0.135762	NaN	1.000000

Entire Home Swiza

Análisis de correlación (r)

	Predicciones_EntireHome	host_acceptance_rate	review_scores_rating	price	review_scores_cleanliness	availability_365	review_scores_communication	room_type	number_of_reviews
Predicciones_EntireHome	1.000000	0.073027	-0.992471	-0.079887	-0.688802	0.084881	-0.536688	NaN	0.216601
host_acceptance_rate	0.073027	1.000000	-0.081236	-0.016178	0.001535	-0.020611	-0.048468	NaN	0.132134
review_scores_rating	-0.992471	-0.081236	1.000000	0.077637	0.594822	-0.088048	0.527860	NaN	-0.214970
price	-0.079887	-0.016178	0.077637	1.000000	0.064782	0.019127	0.038552	NaN	0.015187
review_scores_cleanliness	-0.688802	0.001535	0.594822	0.064782	1.000000	-0.035940	0.398008	NaN	-0.149195
availability_365	0.084881	-0.020611	-0.088048	0.019127	-0.035940	1.000000	-0.072671	NaN	-0.009161
review_scores_communication	-0.536688	-0.048468	0.527860	0.038552	0.398008	-0.072671	1.000000	NaN	-0.194260
room_type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
number_of_reviews	0.216601	0.132134	-0.214970	0.015187	-0.149195	-0.009161	-0.194260	NaN	1.000000

Análisis de determinación (r~2)

	Predicciones_EntireHome	host_acceptance_rate	review_scores_rating	price	review_scores_cleanliness	availability_365	review_scores_communication	room_type	number_of_reviews
Predicciones_EntireHome	1.000000	0.005333	0.984998	0.006382	0.474448	0.007205	0.288034	NaN	0.046916
host_acceptance_rate	0.005333	1.000000	0.006599	0.000262	0.000002	0.000425	0.002349	NaN	0.017459
review_scores_rating	0.984998	0.006599	1.000000	0.006027	0.353813	0.007752	0.278636	NaN	0.046212
price	0.006382	0.000262	0.006027	1.000000	0.004197	0.000366	0.001486	NaN	0.000231
review_scores_cleanliness	0.474448	0.000002	0.353813	0.004197	1.000000	0.001292	0.158411	NaN	0.022259
availability_365	0.007205	0.000425	0.007752	0.000366	0.001292	1.000000	0.005281	NaN	0.000084
review_scores_communication	0.288034	0.002349	0.278636	0.001486	0.158411	0.005281	1.000000	NaN	0.037737
room_type	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
number_of_reviews	0.046916	0.017459	0.046212	0.000231	0.022259	0.000084	0.037737	NaN	1.000000