José Luis Spíndola Gómez

Actividad 1: Regresión Lineal Simple y Múltiple

Para empezar trabajé con los datos nulos y outliers para que la base estuviera completa y limpia, lista para el análisis.

4)

Host acceptance rate vs host response rate

Entire home	0.53
Private room	0.54
Hotel	0.08
Shared room	0.43

Host acceptance rate vs price

Entire home	0.05
Private room	0.06
Hotel	0.04
Shared room	0.10

Host acceptance rate vs number of reviews

Entire home	0.19
Private room	0.12
Hotel	0.20
Shared room	0.15

Reviews score rating vs calculated host listings count

Entire home	0.04
Private room	0.02
Hotel	0.5

Shared room	0.16

Availability 365 vs number of reviews

Entire home	0.03
Private room	0.06
Hotel	0.12
Shared room	0.27

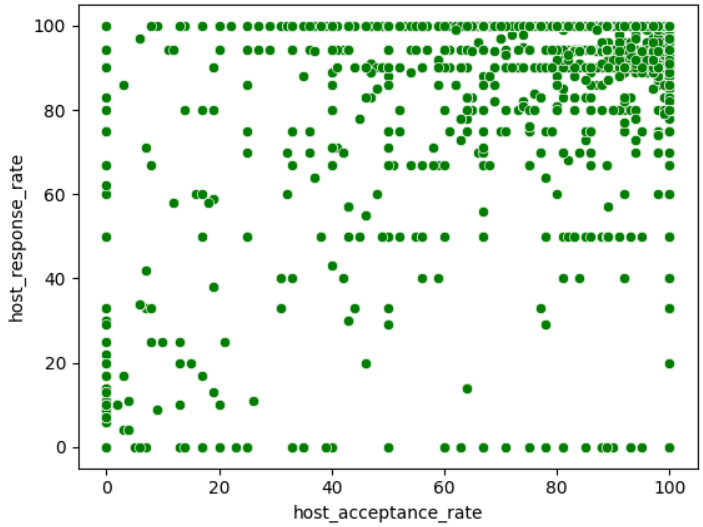
Reviews per month vs review scores communication

Entire home	0.96
Private room	0.97
Hotel	0.99
Shared room	0.97

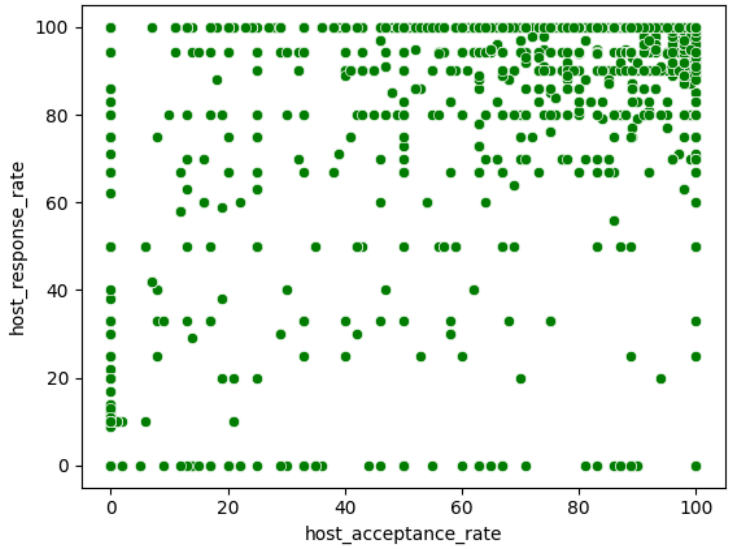
Decidí enfocar el análisis en availability 365 vs number of reviews ya que existe una gran variación que va desde 0.03 para entire home hasta 0.27 pta shared room.

En estos se puede apreciar que la tasa de aceptación se concentra en los porcentajes más elevados, aunque en entire home y private room hay más variación en general. En el caso de hotel y shared room la mayoría tienen buen host response y acceptance rate. Es importante considerar que los 0 son imputaciones hechas para los valores nulos.

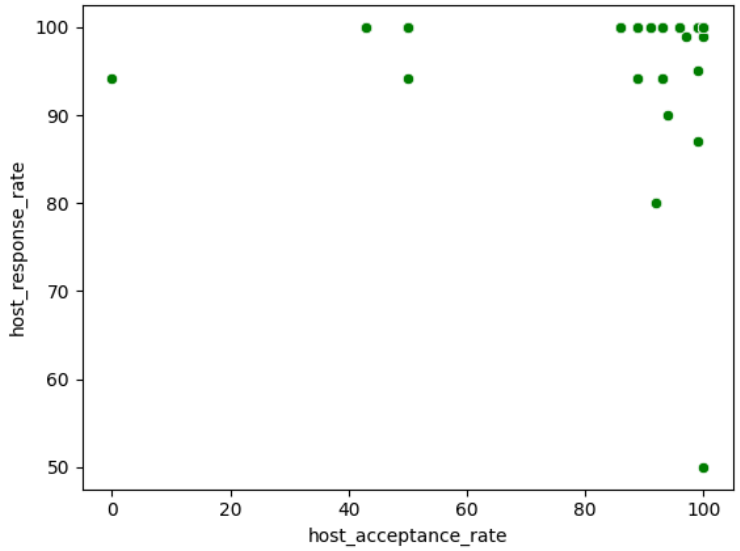
Entire home:



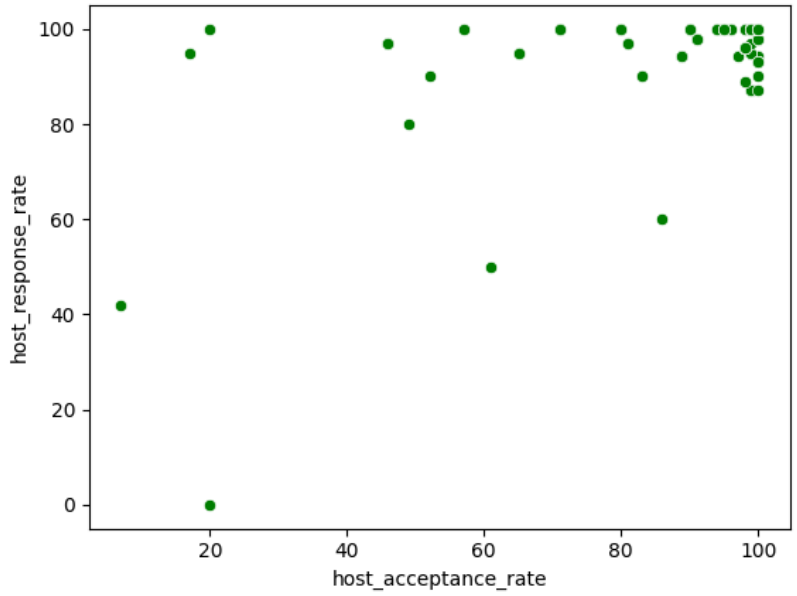
Private room:



Hotel:



Shared room:

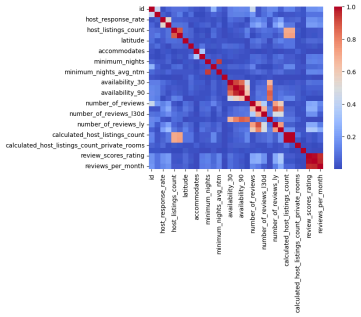


5)

Entire home:

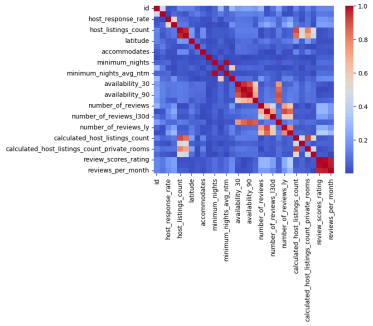
Calculated host listings count vs calculated host listings count entire homes	0.99
Review scores rating vs review score communication	0.98
Availability 60 vs availability 90	0.97

Review scores communication vs reviews per month	0.96
Review scores rating vs reviews per month	0.96
Minimum nights vs minimum nights avg ntm	0.93
Availability 30 vs availability 60	0.92
Number of reviews ltm vs number of reviews ly	0.90
Host listings count vs host total listings count	0.89
Availability 90 vs availability eoy	0.87



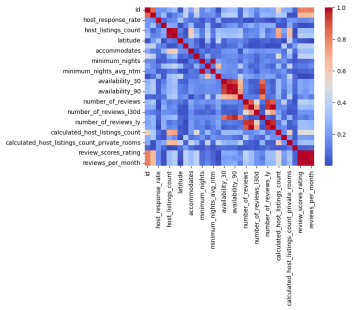
Private room

Minimum nights vs ,minimum nights atm	0.99
Review scores rating vs review scores communication	0.99
Availability 60 vs availability 90	0.97
Review scores communication vs reviews per month	0.97
Review scores rating vs reviews per month	0.96
Host listings count vs host total listings count	0.96
Availability 30 vs availability 60	0.94
Calculated host listings count vs calculated host listings count private rooms	0.92
Number of reviews ltm vs number of reviews ly	0.91
Host listings count vs calculated host listings count	0.90



Hotel:

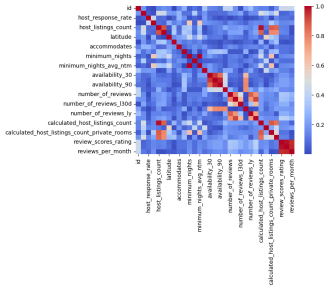
Number of reviews ltm vs estimated occupancy 365	1
Review scores rating vs review scores communication	0.99
Host listings count vs host total listings count	0.98
Minimum nights vs minimum nights avg ntm	0.98
Review scores communication vs reviews per month	0.98
Availability 60 vs availability 90	0.98
Availability 90vs availability eoy	0.96
Number of reviews ltm vs number of reviews ly	0.96
Number of reviews ly vs estimated occupancy l365	0.96



Shared room

Minimum nights vs minimum nights avg ntm	0.99

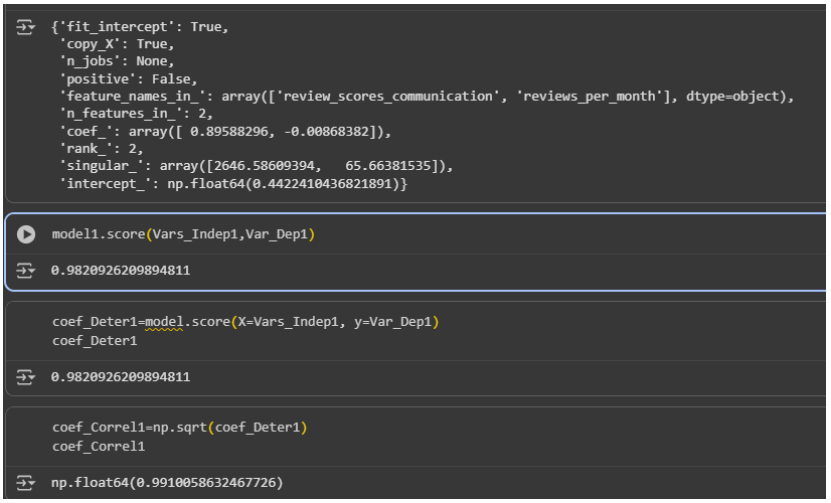
Host listings count vs calculated host listings count	0.98
Review scores rating vs review scores communication	0.98
Availability 60 vs availability 90	0.97
Review scores communication vs reviews per month	0.96
Review scores rating vs reviews per month	0.95
Availability 30 vs availability 60	0.93
Number of reviews ltm vs estimated occupancy l365	0.92
Host listings count vs host total listings count	0.90
Calculated host listings count vs calculated host listings count pr	0.86



De este paso puedo concluir que las variables con mayor correlación son similares en los 4 tipos de alojamientos. Algunas que me parecieron interesantes fueron reviews score rating vs review scores communication ya que muestra la importancia de la atención a l@s huéspedes en la calificación final, review scores communication vs reviews per month ya que nuevamente se muestra que cuando hay buen servicio se realizan más reservas y review scores rating vs reviews per month ya que indica que mientras más reservas el anfitrión mejora y tiene mejores reseñas.

6)

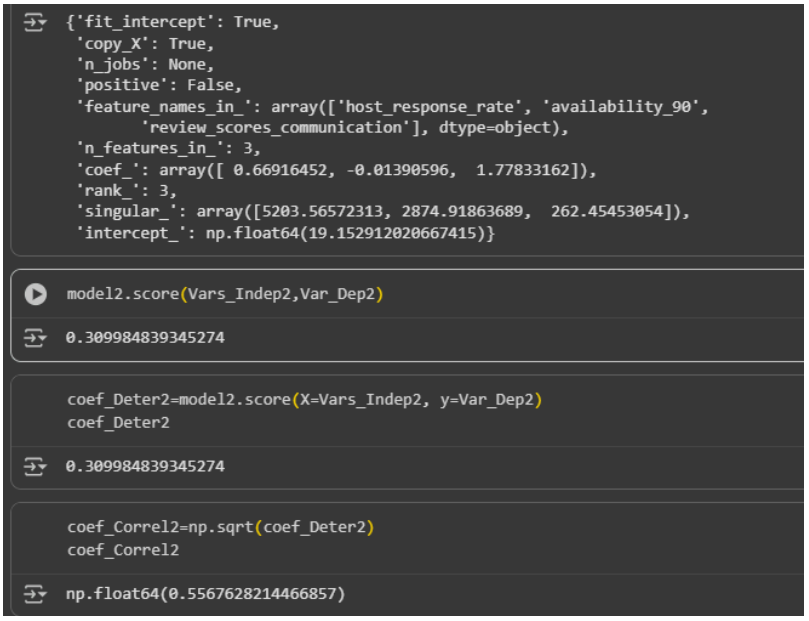
Review scores rating



Modelo: y = 0.89x1 -0.0006x2 + 0.44

Este modelo tiene un coeficiente de determinación de 0.98 y de correlación de 0.99, por lo que es muy confiable para predecir review scores rating, es decir la calificación más relevante y que resume el desempeño del anfitrión.

Host acceptance rate



Modelo: y = 0.66x1 -0.013x2 + 1.77x3

Este modelo no es muy preciso tal como lo dicen los coeficientes de determinación y correlación, de hecho este último mejoró muy poco respecto al análisis simple hecho para cada habitación en el paso 4.

Host total listings count

Modelo: y= 1.06x1 + 0.63x2 -2.34

Considero que esta es una variable difícil de predecir ya que aunque los coeficientes son altos, estos es debido a que estoy usando variables que habían tenido alta correlación pero que realmente son muy similares y no aportan valor, estas son host listings count y calculated host listings count.

Referencia del dataset:

Get the Data. (s.f.). Inside Airbnb. https://insideairbnb.com/get-the-data/