



Tecnológico de Monterrey Campus Puebla

Materia:

Gestión de proyectos de plataformas tecnológicas (Gpo 202)

Entregable:

Actividad 2: Análisis Exploratorio

Alumnos:

María Fernanda Ledesma Martínez

Rodrigo Ruiz Teodoro

Elizabeth Cobos Hernández

Fecha de entrega:

30-09-2022

Actividad 2: Análisis Exploratorio

Esta actividad tiene como objetivo desarrollar un modelo matemático que logre describir el comportamiento de precios de las diferentes habitaciones de AIRBNB en México y compararlo con otras 2 ciudades (Buenos Aires y Atenas)

En primera instancia se descargaron 3 bases de datos de las ciudades antes mencionadas (México, Atenas y Buenos Aires); Posterior a eso, se realizó una limpieza de valores nulos y outliers para obtener una base limpia con la que se pudiera realizar los análisis correspondientes.

Nuestra variable predictora será 'price' y se buscará analizar su comportamiento con cada tipo de cuarto de las 3 ciudades seleccionadas. De igual forma, se calcularán los coeficientes de determinación y correlación que tiene con las siguientes variables: 'minimum_nights', 'number_of_reviews', 'reviews_per_month', 'calculated_host_listings_count', 'availability_365' y 'number_of_reviews_ltm'

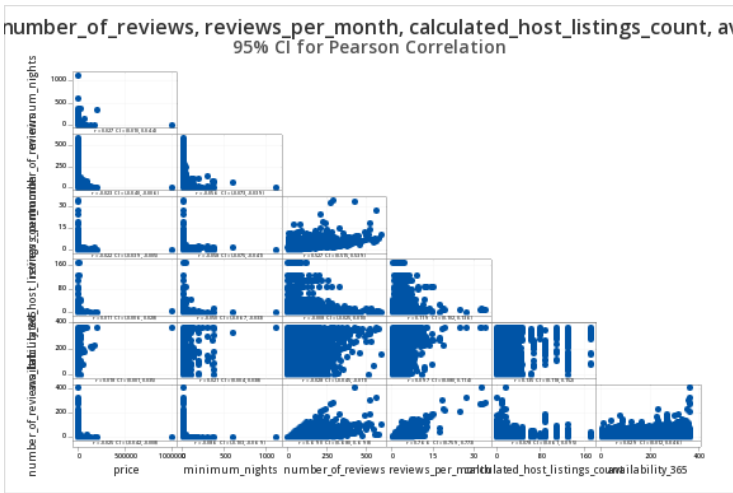
México

Price	Enteri Home/apt		Hotel Room		Private Room		Shared Room	
	Coef Deter	Coef Corre	Coef Deter	Coef Corre	Coef Deter	Coef Corre	Coef Deter	Coef Corre
minimum_nights	0.07%	0.027	0	0.004	0	-0.004	14.28%	0.378
number_of_reviews	0.05%	-0.023	0.35%	-0.059	0.06%	-0.024	0.43%	-0.66
reviews_per_month	0.05%	-0.022	0.11%	-0.034	0	0	0.79%	0.089
calculated_host_listings_count	0.01%	0.011	1.28%	-0.113	1.21%	0.110	1.80%	-0.134

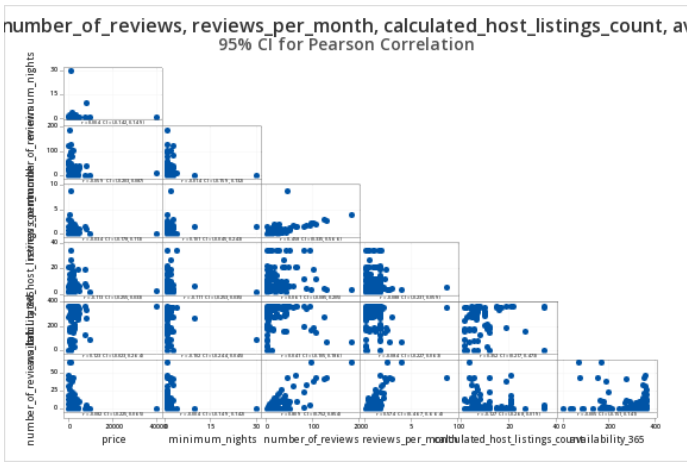
availability_365	0.03%	0.018	1.51%	0.123	0.07%	0.027	0.51%	0.071
number_of_reviews_ltm	0.06%	-0.025	0.67%	-0.082	0.03%	-0.017	0.47%	-0.069

$$price = \text{minimnights}(23.30) + 561$$

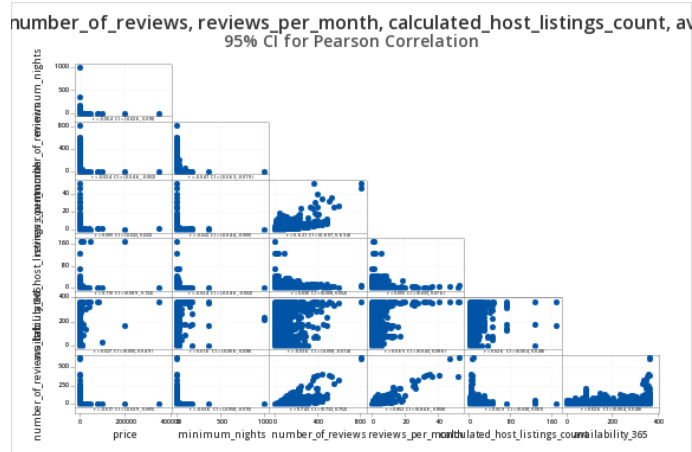
Entire Home/Apt



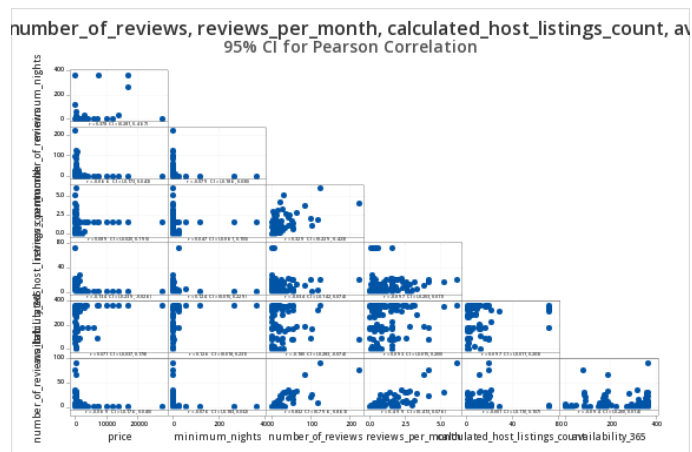
Hotel Room



Private Room



Shared Room

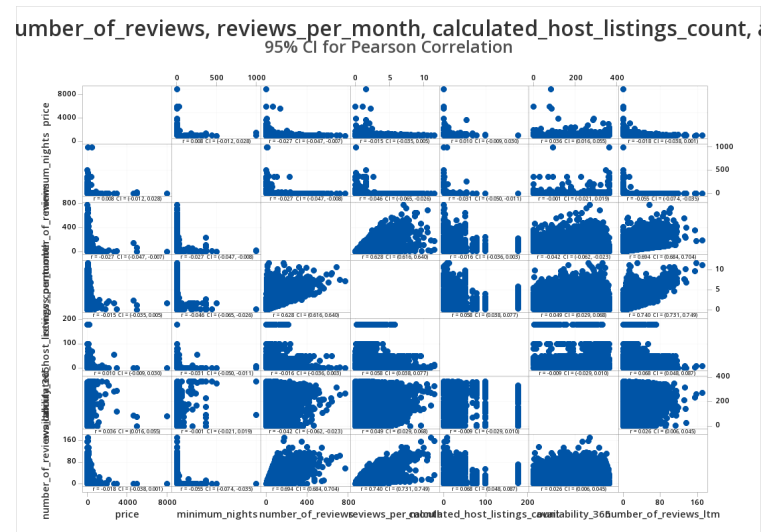


Teniendo en cuenta lo presentado anteriormente, hemos de decir que, ninguno de los modelos propuestos para este caso de estudio es ideal para predecir la variabilidad del precio, considerando que, el mejor modelo es: price y minimum_nights en el subconjunto de “Shared room”, con un coeficiente de determinación que explica únicamente el 14.28% de las variaciones del precio con respecto al mínimo de noches.

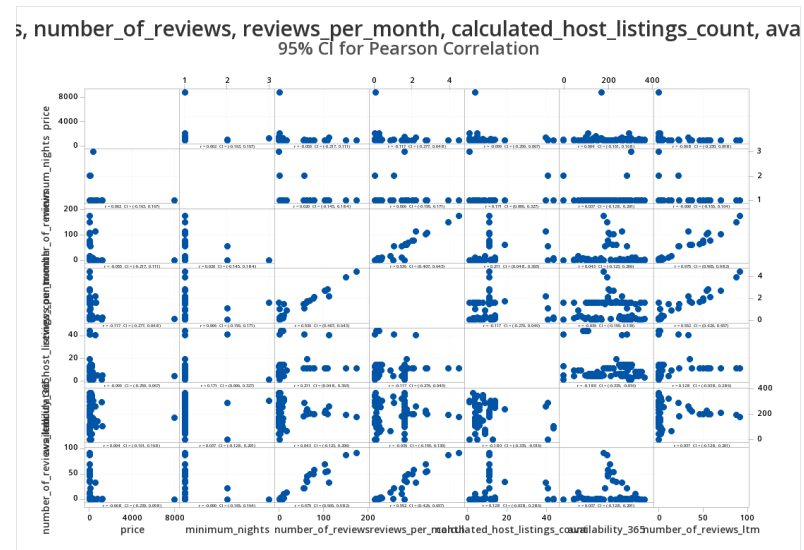
Atenas

Price	Enteri Home/apt		Hotel Room		Private Room		Shared Room	
	Coef Deter	Coef Corre	Coef Deter	Coef Corre	Coef Deter	Coef Corre	Coef Deter	Coef Corre
minimum_nights	0.00006	0.008	0.000004	0.002	0.039601	0.199	0.0059	-0.077
number_of_reviews	0.00073	-0.027	0.003025	-0.055	0.003136	-0.056	0.1475	0.384
reviews_per_month	0.00023	-0.015	0.013689	-0.117	0.000036	0.006	0.0110	-0.105
calculated_host_listings_count	0.00010	0.010	0.009801	-0.099	0.000676	-0.026	0.0188	0.137
availability_365	0.00130	0.036	0.000016	0.004	0.000049	0.007	0.0433	0.208
number_of_reviews_ltm	0.00032	-0.018	0.004624	-0.068	0.003025	-0.055	0.0289	-0.170

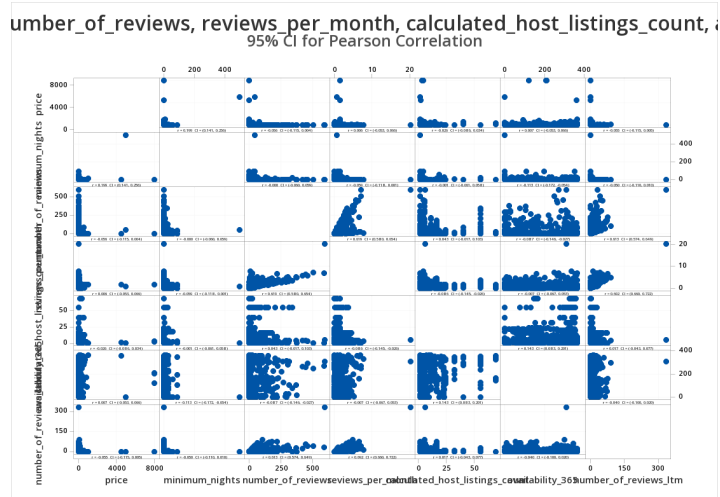
Enteri Home/apt



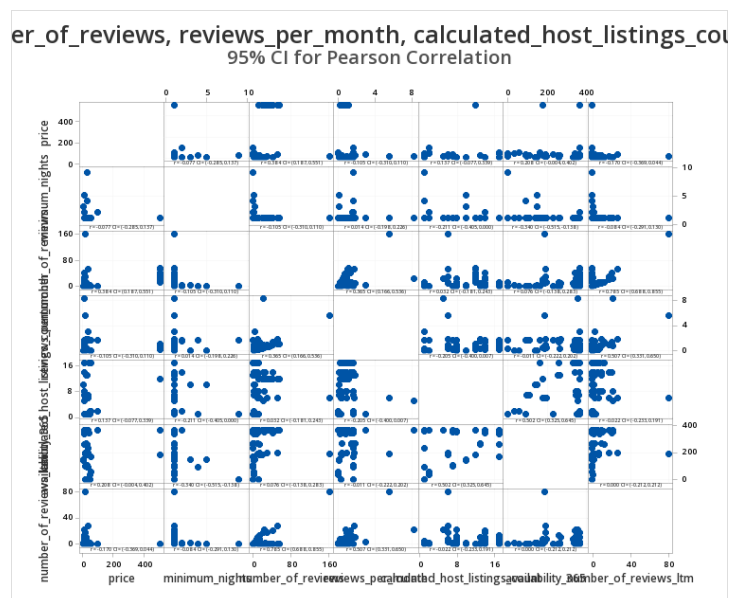
Hotel Room



Private Room



Shared Room



Como se puede observar en la tabla, existe un nivel muy bajo de correlación entre ‘price’ y las demás variables independientes respecto a los diferentes tipos de cuartos en la ciudad de Atenas. No obstante, podemos proponer el siguiente modelo con base a las 2 correlaciones más elevadas obtenidas en ‘Shared Room’ que son number_of_reviews con (0.384) y availability_365 con (0.208).

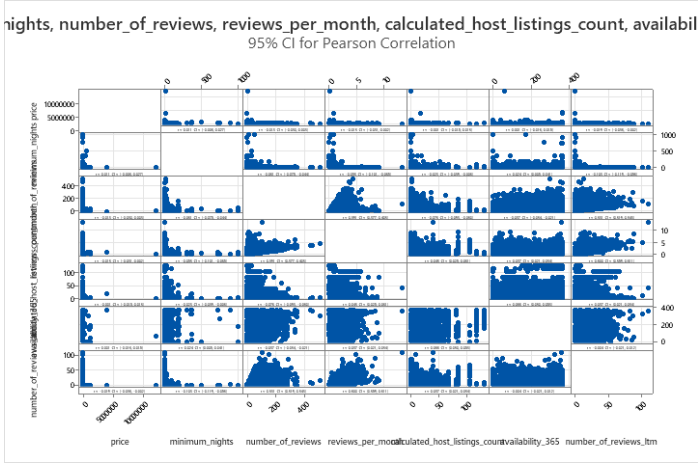
El modelo propuesto queda de la siguiente forma:

$$price = number_of_reviews(-0.110677) + availability_365(0.036876) + 89.754842$$

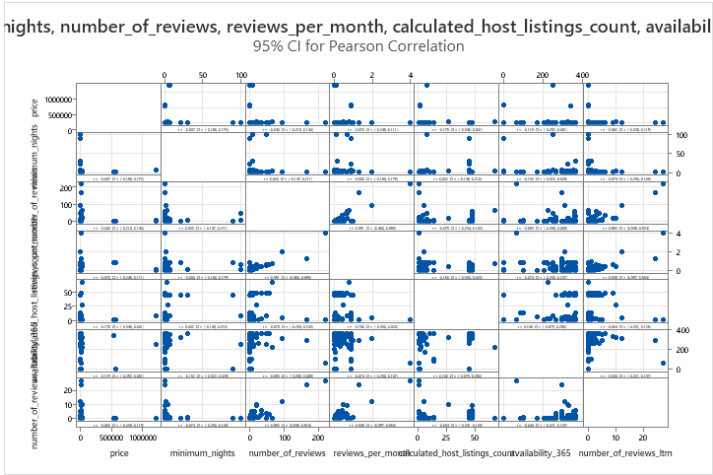
Buenos Aires

Price	Enteri Home/apt		Hotel Room		Private Room		Shared Room	
	Coef Deter	Coef Corre	Coef Deter	Coef Corre	Coef Deter	Coef Corre	Coef Deter	Coef Corre
minimum_nights	0.01%	0.011	0.004%	-0.007	0.04%	0.021	0.05%	-0.022
number_of_reviews	0.02%	-0.013	0.16%	-0.040	0.04%	-0.020	0.15%	-0.038
reviews_per_month	0.02%	-0.015	0.49%	-0.070	0.01%	0.010	0.15%	0.039
calculated_host_listings_count	0.0001%	-0.001	3.18%	-0.178	0.02%	0.015	0.43%	-0.066
availability_365	0.0001%	0.001	1.42%	-0.119	0.10%	0.031	0.72%	0.085
number_of_reviews_ltm	0.04%	-0.019	0.38%	0.062	0.03%	-0.018	0.15%	-0.038

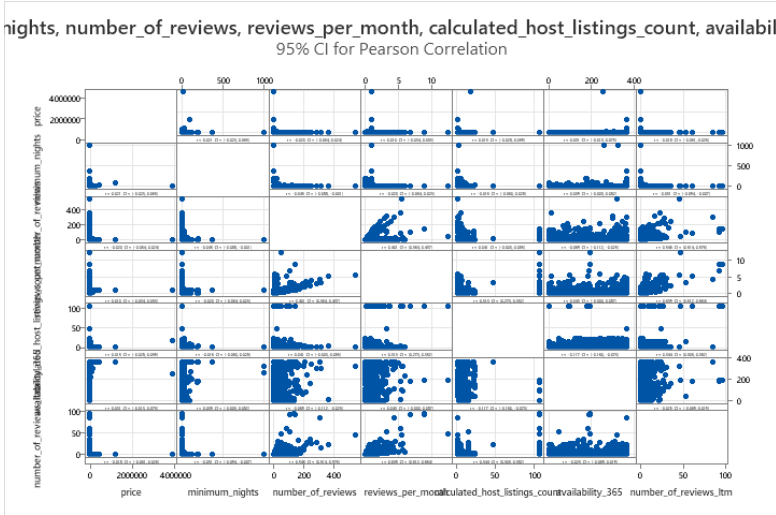
Entire Home



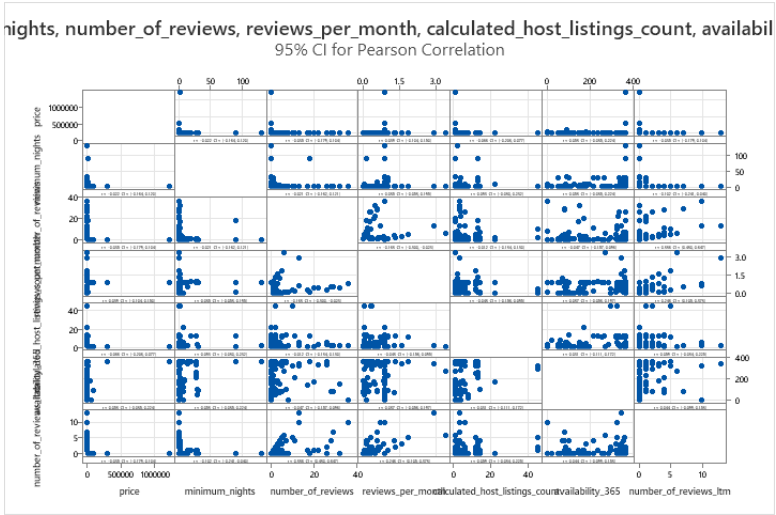
Hotel room



Private Room



Share Room



En la ciudad de Buenos Aires, Argentina a diferencia de las otras dos ciudades en diferentes países, tenemos mejores resultados dentro de Hotel Room y no tanto en Shared Room. Sin embargo, las coeficientes de determinación y correlación siguen siendo demasiado bajos como para poder explicar la variabilidad del precio de estas habitaciones. Y de la misma manera, nuestros variables más fuertes son diferentes a las otras ciudades y estas son “availability_365” con correlación de 12% y determinación del 1.42% y “calculated_host_listings_count” con correlación con price de 18% y determinación de 3.18%.

Como hemos mencionado, estas variables sí tienen cierta relación con nuestra variable respuesta; pero, no explican su variación ni la del modelo. Con cualquier modelo propuesto obtenemos el mismo resultado. Sin embargo, consideramos que el mejor es el que tenga estas dos variables mencionadas anteriormente y para el tipo de habitación de hotel. Quedando de la siguiente manera:

$$price = 148947 - 1670 \text{ calculated_host_listings_count} - 219 \text{ availability_365}$$