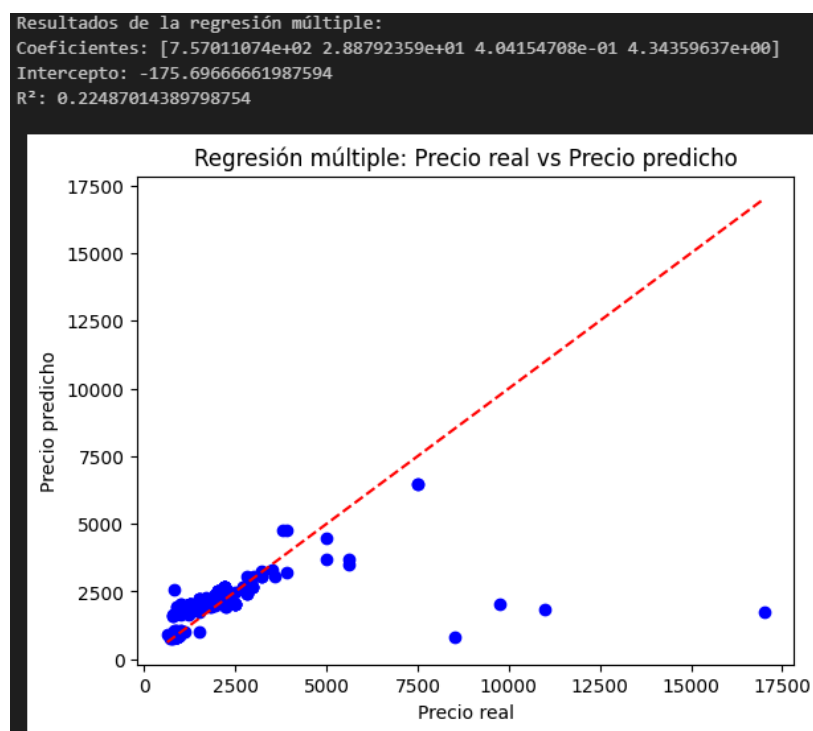


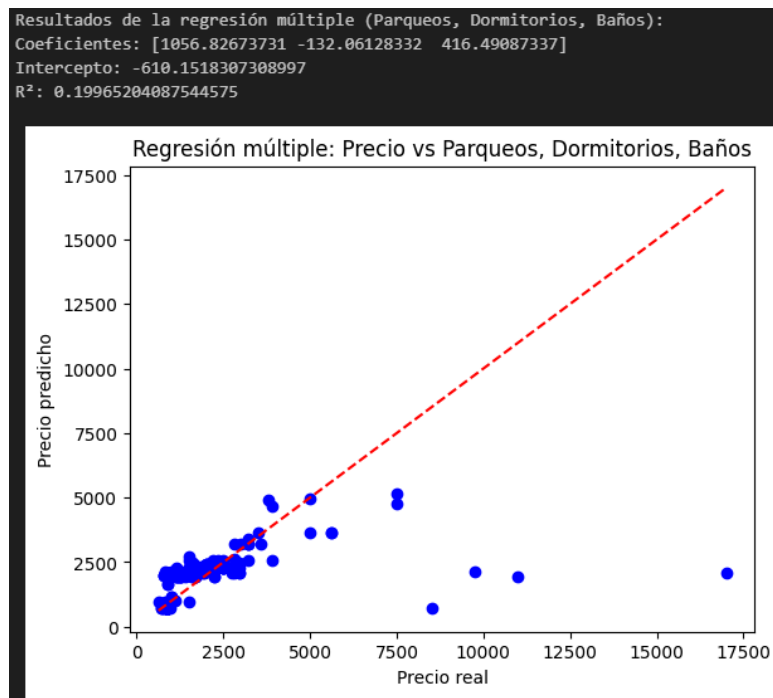
Regresión Lineal Múltiple

Previamente se realizó un Web Scraping de la página Mapainmueble, específicamente de la sección de apartamentos en alquiler de la Zona 14. El Scraping se realizó para las primeras 25 páginas de la sección, resultando en un total de 214 propiedades. De estas se rescataron las siguientes características: precio, número de parqueos, número de dormitorios, número de baños, ID de la propiedad, tamaño, y el enlace. Para la parte que se desarrollará en este documento, las características importantes son el precio, número de dormitorios, número de baños, tamaño y número de parqueos. Luego, se realizaron cuatro distintas regresiones lineales simples, comparando el precio con cada una de esas variables.

Ahora, el objetivo de esta parte era realizar regresiones lineales múltiples. Por lo tanto, en este caso se usaron más de una variable al momento de realizar el modelo. Primero, se utilizaron las cuatro variables para crear un modelo *baseline*. Esto significa que se realizó una regresión lineal múltiple comparando el precio con el número de parqueos, dormitorios, baños, y el tamaño. El cual se puede ver en la siguiente imagen y tuvo como resultado un R^2 de 0.22487. Esto último significa que aproximadamente el 22% de la varianza está explicada por el modelo.



Luego, se decidió realizar otros 2 modelos más, donde por lo menos uno mostrará un mejor rendimiento que el *baseline*. Para el primero, se utilizaron como variables únicamente: el número de parqueos, dormitorios y baños. Este modelo se mostró con un R^2 inferior siendo 0.19965, por lo tanto podemos decir que su rendimiento fue inferior.



Para el tercer modelo, se utilizaron únicamente los parqueos, dormitorios y el tamaño. Se escogió este después de realizar comparaciones de R^2 en otras combinaciones y los coeficientes. Esto porque no se logró encontrar un modelo que mostrará un R^2 mayor al *baseline*, debido a que es el que tiene la mayor cantidad de variables. Sin embargo, esta combinación obtuvo casi el mismo R^2 , siendo 0.22478. Entonces, podemos concluir que este modelo tiene mejor rendimiento que el *baseline*, porque teniendo una variable menos, tiene prácticamente el mismo R^2 .

