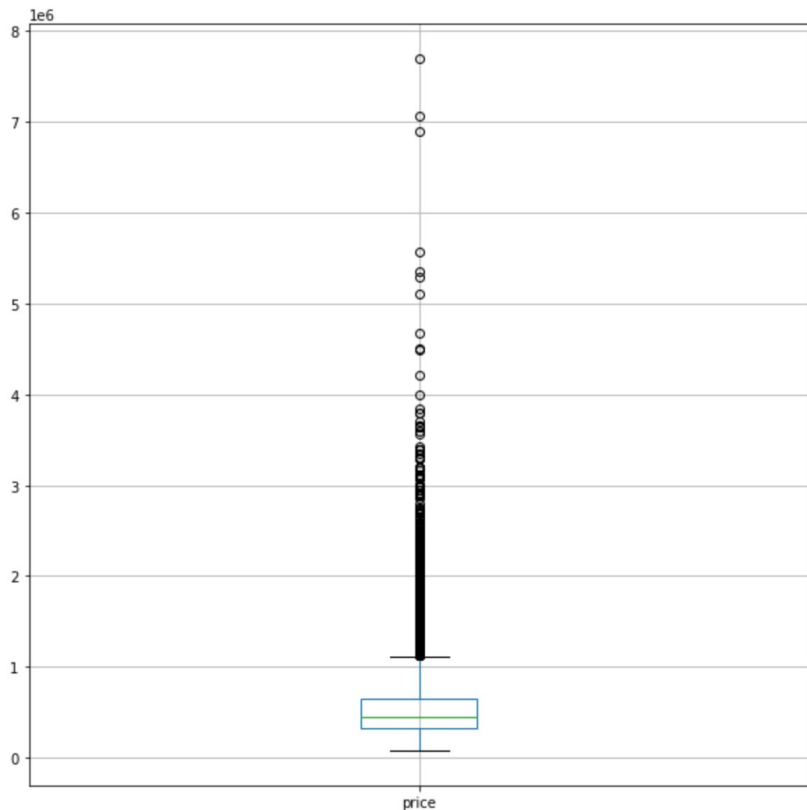


# PROYECTO DE REGRESIÓN

María Ramón García  
Andrea Casas Saez

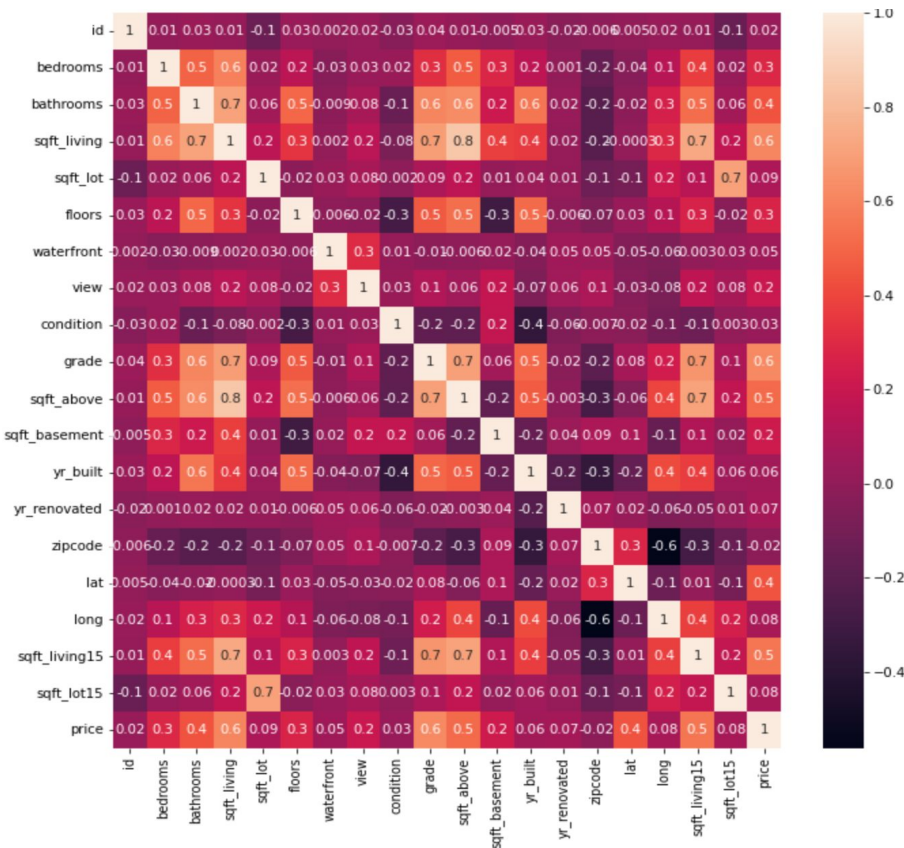
# DIAGRAMA DE CAJA



Se han reunido una cantidad definida de datos que representan, de manera simulada, los precios de unas casas.

El primer paso antes de generar un modelo de regresión es observar la relación entre las variables y a continuación representamos los valores atípicos en este caso serán las casas con precios elevados al resto.

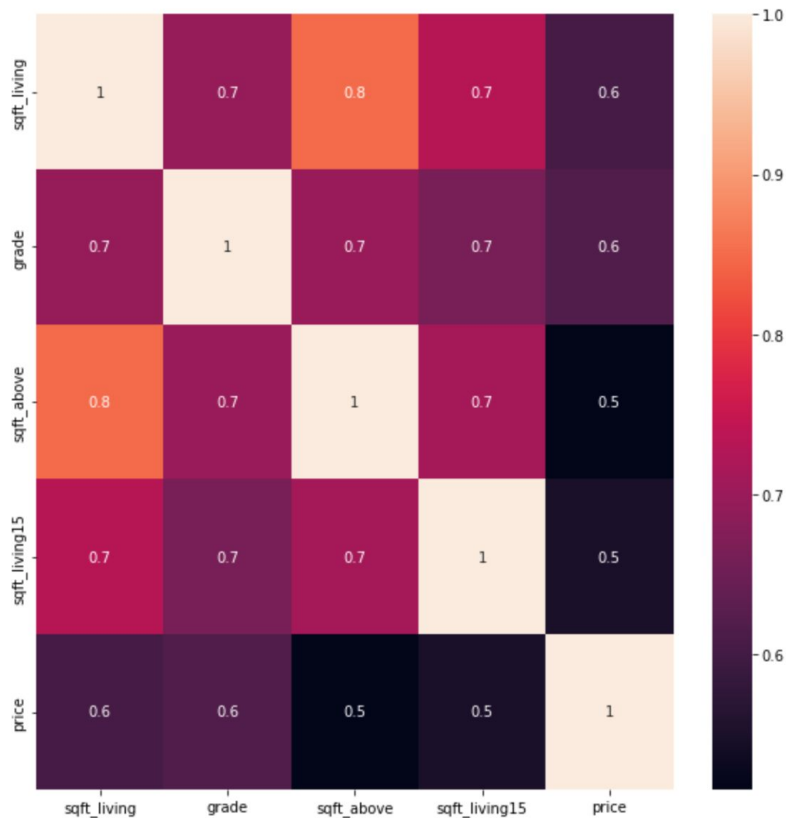
# ANÁLISIS DE CORRELACIÓN: heatmap matriz de correlación



Cuando se dispone de múltiples variables numéricas, es conveniente estudiar el grado de correlación entre las variables disponibles. Una forma de hacerlo es mediante matrices de correlación, en las que se muestra el coeficiente de correlación para cada par de variables.

Los coeficientes de correlación a lo largo de la diagonal de la tabla son todos iguales a 1 porque cada variable está perfectamente correlacionada consigo misma.

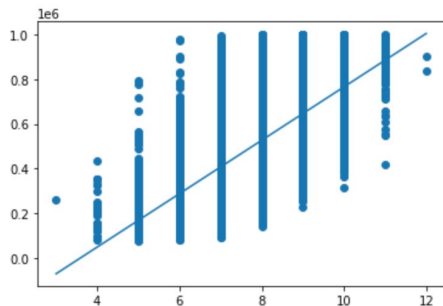
# ANÁLISIS DE CORRELACIÓN FUERTE



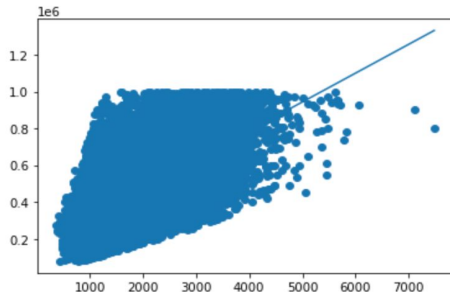
Selección de pares de correlación fuerte, es decir, magnitud superior a 0,5.  
Por lo tanto, elegimos las características fuertemente relacionadas, filtrando aquellos pares de características cuyos valores de coeficiente de correlación sean mayores que 0,5.

# DISPERSIÓN DE LAS VARIABLES CON RESPECTO A PRICE

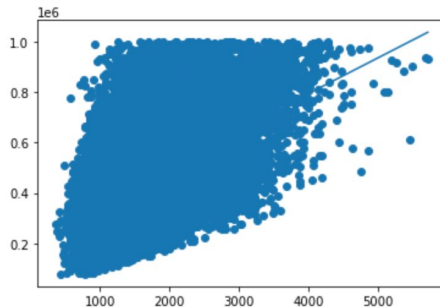
Coefficients:  
[119291.89239156]  
Mean squared error: 23662862977.80  
Coefficient of determination: 0.38



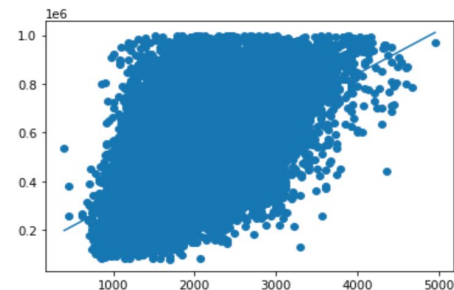
Coefficients:  
[156.51627288]  
Mean squared error: 24261234713.44  
Coefficient of determination: 0.37



Coefficients:  
[141.79261745]  
Mean squared error: 28056671354.48  
Coefficient of determination: 0.27



Coefficients:  
[178.86705869]  
Mean squared error: 26685963561.04  
Coefficient of determination: 0.30



Se observa la relación lineal entre las variables independientes con respecto al precio, de tal forma que es posible estimar visualmente la relación que tienen.