

1. En este ejercicio usaremos la base de datos de Boston\_housing\_dataset (<https://www.kaggle.com/apratim87/housingdata>) la cual consta de 13 variables independientes y la variable dependiente. Son 506 datos de casas cuyo objetivo es la predicción de su costo (MEDV).
  - a. Obtener el modelo de Regresión Lineal Múltiple (RLM) del modelo y con base a dicho modelo contestar los siguientes incisos:
    - i. Interpretar el valor del coeficiente de determinación ajustado  $\bar{R}^2$ .
    - ii. Determina qué variables no resultaron significativas considerando un  $valor-p = 0.01$ .
    - iii. Obtener un segundo modelo de RLM omitiendo las variables que no resultaron significativas en el inciso anterior. Comparar su coeficiente de determinación ajustado con base con respecto al primer modelo.
    - iv. Obtener la matriz de correlación de Pearson de las 14 variables. Con base a dicha matriz identifica los pares de variables independientes que tengan un coeficiente de correlación mayor o igual a 0.7 en valor absoluto. Obtener un tercer modelo de RLM y compara la significancia de los coeficientes y su coeficiente de determinación ajustado con respecto a los dos primeros modelos.
    - v. Con base a la matriz de correlación con las 14 variables, identifica ahora aquellas variables independientes que tengan una correlación de pearson mayor o igual a 0.5 con respecto a la variable dependiente MEDV. Obtener un cuarto modelo RLM y compararlo con los tres anteriores.
    - vi. Usando únicamente la matriz  $X$  con las 3 variables independientes del inciso anterior y la variable de salida MEDV, obtener la matriz de Pearson  $X^+ = (X^T X)^{-1} X^T$ . Posteriormente obtener los coeficientes  $\beta$  del modelo RLM resultante mediante la expresión:  $\beta = X^+ Y$ . Compáralo con el inciso anterior y escribe tus conclusiones.
    - vii. Finalmente escribe tus conclusiones de los resultados obtenidos.