DATA MINING 4

Zadanie 1

A. Pobierz dane ze zbioru **Boston Housing** opisanego na stronie WWW https://www.kaggle.com/c/boston-housing i zawartego w pliku **Boston.csv** dostępnym w folderze **Lab_4**.
B. Załaduj dane do **Colab**.
C. Odpowiedz na pytania:
    a. Jak liczny jest zbiór danych?
    b. Iloma atrybutami jest opisany?
    c. Jakiego typu są wartości poszczególnych atrybutów?
    d. Czy w zbiorze są brakujące dane (pola NULL)?

    Przetestuj działanie metody `info()`:
    https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.info.html

Zadanie 2

Dla każdego atrybutu (kolumny) znajdź:

A. Wartość min
B. Wartość max
C. Wartość średnią
D. Odchylenie standardowe - korzystając z metody **`std()`**:
    https://numpy.org/doc/stable/reference/generated/numpy.std.html

Zadanie 3

A. Narysuj wykres rozrzutu dla atrybutów **rm** i **medv**. Wykorzystaj metodę **`plt.scatter()`**.
B. Wykorzystując metodę najmniejszych kwadratów znajdź **prostą aproksymującą** punkty na wykresie.

Zadanie 4

Powtórz **zadanie 4** dla dwóch innych par atrybutów (kolumn). W wyborze par atrybutów uwzględnij wartość korelacji między atrybutami.