In this assignment, the main goal is to predict the cost of real estate.

Question 1: NumPy Linear Regression [80 pts]

In this question, students will implement the linear regression model via stochastic gradient descent (SGD) to predict the real estate cost using the provided data. **Students must use only NumPy, Stat, Pandas, and Matplotlib in this question.**

1. [25 pts] Perform pre-processing and EDA on the data and split it into training and testing datasets. Have a split ratio of 75% and 25% for the training and testing datasets, respectively. Please explain the performance. Students can make visualizations if necessary.
2. [35 pts] Define a function linear regression model that takes the learning rate = 0.05 and iteration = 1000. This model will perform SGD. You can initialize random weights using NumPy. Report the mean squared error (MSE). This model is going to return the learned weights.
3. [15 pts] Use different learning rate values to retrain the model.
   1. Learning rates = [5E-3, 1E-2, 1E-1, 5E-1]
   2. Report the MSE value for each learning rate.
4. [5 pts] Using the final model obtained from c), generalize the model using the test set. Report the mean squared error value.

Question 2: Scikit-learn linear regression. [20 pts]

In this question, students will use the Scikit-learn linear regression model to predict the real estate cost and compare the results in Question 1.

1. [10 pts] Now import linear regression from Scikit-learn and train the model.
2. [5 pts] Get the prediction on the test set.
3. [5 pts] Compare the result of your model to that of the Scikit-learn model.