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

Question 1: Linear Regression with regularization [80 pts]

In this question, students will implement the ridge () regularized regression model to predict the real estate cost. **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 70% and 30% for the training and testing datasets. Please provide an explanation of the performance. Students can make visualizations if necessary.
2. [40 pts] Define a function model that takes the parameters features, labels, learning rate, lambda (regularization coefficient), epochs, and batch size. This model will perform mini-batch gradient descent. You can initialize random weights using NumPy. Report the mean squared error (MSE) at the end of every epoch. This model is going to return the learned weights. Make a visualization of MSE vs. epoch.
3. [10 pts] Use the weights obtained from the model and predict the output of the test set.
4. [5 pts] Calculate the mean squared error of the prediction.

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

In this question, students will use the Scikit-learn Ridge regression model to predict the real estate cost and compare the results found 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.

Note: If the problem looks daunting, please use office hours, and read through the slides. The solution is already present in the slides.