Task 1: apply different normalization methods to the data and create different normalized data sets

Task 2: divide the normalized data sets to training and validation sets

Task 3: use all of the divided data sets to train different regression models (use various cost functions)

Task 4: compare the performance of different models according to their respective validation sets

First of all, we want to use cleaned data for our project. Since I already cleaned the data in the last project.

Task 1&2:

In these tasks, we need to normalize and split the data. It's important to decide which step to perform first. From my research, I found that it's best to split the data first, and then apply the normalization.

I split the data with 80% for training and 20% for validation. After that, I applied four different normalization methods to the training data:

a. Min-Max scaling
b. Standard scaling
c. Robust scaling
d. Max-Abs scaling

Task 3:

In Task 3, we need to train the data using linear regression. However, this must be done with each of the normalization techniques mentioned earlier, and we are also required to use two loss functions: MSE (Mean Squared Error) and R^2 (coefficient of determination).

To achieve this, I wrote a method that performs the following:

1. It initializes the model and fits the data.
2. It defines y_pred to calculate the loss functions.
3. Finally, the method returns both the MSE and R^2 values. This method can take data normalized by any of the four normalization techniques, train it using a linear regression model, and return the loss values.

Task 4:

In this task, we are required to compare the results and display them in a graph. I accomplished this using matplotlib to visualize the comparisons.