DATA-445 IN-CLASS ASSIGNMENT SEPTEMBER 20, 2023

Consider the `CarPrice_Assignment.csv` data file. This data is public available on the Kaggle website, and has information on cars (characteristics related to car dimensions, engine and more). The goal is to use car information to predict the price of the car. **In Python**, answer the following:

1. (5 points) Using the pandas library, read the csv data file and create a data-frame called `car_price`.
2. (5 points) Using the `wheelbase`, `enginesize`, `compressionratio`, `horsepower`, `peakrpm`, `citympg`, and `highwaympg` as the predictor variables, and `price` is the target variable, split the data into train (80%) and test (20%).
3. (5 points) Using the train dataset, perform LASSO as a variable selector.
4. (5 points) Using the train dataset, build a linear regression model with the selected features from part (3). Make sure to standardize the input features with `StandardScaler` and `Pipeline`. After that, use this model to predict on the test dataset. Report the MSE of this model.
5. (5 points) Using the train dataset, build a ridge regression model with the selected features from part (3). Make sure to standardize the input features with `StandardScaler` and `Pipeline`. After that, use this model to predict on the test dataset. Report the MSE of this model.
6. (3 points) Using the results from parts (4) and (5), what model would you use to predict car prices? Explain.