Homework Two

Instruction: Consider the scenario from homework 1, where we explore the dataset ToyotaCorolla. It has been pointed out that there is an extreme value in the CC variable – 16,000. This error has been corrected to 1600. The categorical Fuel-type variable has been converted into dummy variables. The updated Toyota Corolla dataset, called PreToyotaCorolla.xls, will be used for this homework. Create a new Word document and save it as HW2Answers_X (where X is your team number). Where required, write your answers/paste screenshots into this Word document. Your response should not exceed 100 words for each below question. Write every member's full name and participation on the first page of the Word document as follows. You need to submit this Word document and XLMiner solution.

Participant	Complete the Assignment before the Meeting (Y/N)	Percentage of Contribution	Justification

Tasks:

1. Prepare the dataset (as factored into dummies) for data mining techniques of supervised learning by creating partitions using XLMiner's data partitioning utility. Select all the variables and use default values for the random seed and partitioning percentages for training (50%), validation (30%), and test (20%) sets. Describe the roles that these partitions will play in modeling.
2. Why should not every dummy variable be used to build regression models?
3. Run a multiple linear regression with the output variable Price and input variables Age_08_04, KM, Fuel_Type, HP, Automatic, Doors, Quarterly_Tax, Mfr_Guarantee, Guarantee_Period, Airco, Automatic_Airco, CD_Player, Powered_Windows, Sport_Model, and Tow_Bar. Report the model.
4. Repeat (3) using stepwise regression to reduce the number of predictors. Report the estimated model selected.
5. Compare the resulting best model to the one you obtained in (4) in terms of the predictors that are in the model(3). Compare the predictive accuracy of both models (3) and (4) using measures such as RMSE and average error and lift charts.

6. What appears to be the three or four most important car specifications for predicting the car's price?
7. Using model (4), predict the price on a car with the following characteristics: Age_08_04 = 20, KM=14000 kilometers, Fuel Type=Petrol, HP =110, Automatic=0, Doors =3, Quarterly_Tax = 85 Euros, Mfr_Guarantee =1, Guarantee_Period=3, Airco=1, Automatic_Airco =1, CD_Player=0, Powered_Windows=0, Sport_Model=0, and Tow_Bar=0.
8. Modify the worksheets' titles and names appropriately to reflect the contents.

Important submission instructions

Save your Word file and XLMiner solution. Use the link "Homework 1" to upload these files. **Due by 11.59 P.M. Feb. 24, 2019.**