Question 1

What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

- ⇨ The optimal value of alpha for Ridge is 0.02 and for Lasso it is 0.001. With these alphas the R2 of the model was approximately 0.855.
  And after doubling the alpha values in the Ridge and Lasso, the prediction accuracy remains around 0.85 but there is a small change in the co-efficient values. The new model is created and demonstrated in the Jupiter notebook.
  Since the alpha values are very minimal, we cannot see any huge change in the model after doubling the alpha.

  The most implemented variables after changes are:

  - RoofMatl_Membran
  - RoofMatl_WdShngl
  - MSZoning_RM
  - Condition2_PosA
  - MSZoning_FV
  - MSZoning_RL
  - MSSubClass

Question 2

You have determined the optimal value of lambda for ridge and lasso regression during the assignment. Now, which one will you choose to apply and why?

- ⇨ In the model prepared the Mean Squared Error of both the models are very similar. Since Lasso helps in feature reduction (as the coefficient value of some of the features become zero), Lasso value has a better edge over Ridge value and can be used as the final model.

Question 3

After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

- ⇨ The five most important predictor variables in the current lasso model are:
  - RoofMatl_Membran
  - RoofMatl_WdShngl
  - MSZoning_RM
  - Condition2_PosA
  - MSZoning_FV
  - MSZoning_RL
  - MSSubClass
- ⇨ The another model after predictor variales are:

	Feaure	Coef
0	MSSubClass	10.803302
114	RoofMatl_Membran	0.434518
39	MSZoning_FV	0.402129
41	MSZoning_RL	0.401668
40	MSZoning_RH	0.396884
42	MSZoning_RM	0.354961
116	RoofMatl_Roll	0.308425
115	RoofMatl_Metal	0.303484
119	RoofMatl_WdShngl	0.294146
238	PoolQC_NA	0.248107

Question 4

How can you make sure that a model is robust and generalisable? What are the implications of the same for the accuracy of the model and why?

⇨ A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data. The model should also be generalisable so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much weightage should not give to the outliers so that the accuracy predicted by the model is high.

To ensure that this is not the case, the outlier analysis needs to be done and only those which are relevant to the dataset need to be retained. Those outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model. Confidence intervals can be used (typically 3-5 standard deviations). This would help standardize the predictions made by the model. If the model is not robust , it cannot be trusted for predictive analysis.