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?
Answer:
The optimal values of alpha for ridge and lasso regression are 2 and 0.01 respectively. The reason to choose alpha 2 for ridge regression is that test error is minimum when we draw the plot between negative mean absolute error and alpha. We can observe from the plot that the alpha increasing from 0 test error is decreasing and train error is increasing. At alpha = 2, the test error is minimum. From the plot between Negative mean absolu, te error and alpha, negative mean error is minimum at alpha = 0.4 and stabilises later For lasso regression, I choose alpha as 0.04. For this alpha a few coefficients become zero as alpha increases and R2 values of test and train or not getting matched. So I chose alpha = 0.01 to balance the trade-off between Bias-Variance.
If double the alpha a few coefficients become 0 as alpha increases

The important predictor variables for Lasso are:
GrLivArea, OverallQual, OverallCond, TotalBsmtSF, BsmtFinSF1, GarageArea, Fireplaces, LotArea, LotFrontage, and BsmtFullBath

The important predictor variables for Lasso are:
MSZoning_FV, MSZoning_FV, Neighborhood_Crawfor, MSZoning_RH, MSZoning_RM, SaleCondition_Partial, Neighborhood_StoneBr, GrLivArea, SaleCondition_Normal, and Exterior1st_BrkFace

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?
Answer:
I have chosen lasso regression to build the model because it makes the model simple with fewer variables. As the lambda increases the lasso makes the coefficients zero and it makes variables exactly equal to 0. Which in turn makes the model simple. As we increase the lamda, the variance in the model decreases and bias remains constant. Ridge regression includes all variables in the model which makes the model complex.

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?
Answer:
GarageArea, Fireplaces, LotArea, LotFrontage, and BsmtFullBath

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?
Answer:
A model is considered to be robust and generalisable iif the accuracy of the output is consistent with drastically change in input variables. To make sure that a model is robust and generalisable, outlier analysis needs to be done on the data set. Only the data relevant to the dataset needs to be retained. The model should be accurate for datasets which were used during the training.