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?

Ans: The optimal value of alpha for ridge regression is 0.25 and for lasso regression it is 0.0001.

Changing the alpha from 0.25 to 0.5 with ridge regression changed the accuracy r2 score of training data from 0.9115 to 0.9108, r2 score of test data from 0.8921 to 0.8911. However, it is a negligible change in the accuracy of the model.

Changing the alpha from 0.0001 to 0.0002 with lasso regression changed the accuracy r2 score of training data from 0.9067 to 0.9026, r2 score of test data from 0.8931 to 0.8889. However, it is a negligible change in the accuracy of the model.

The most important predictor variables after the change are implemented is still GrLivArea: Ground Living Area

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?

Ans: I will choose 0.25 as the alpha value for ridge regression and 0.001 for lasso. Though I see doubling the alpha did not make much difference. Reason is with higher value of lambda the model will start to underfit and I want to avoid it. The model starts to become simpler and starts to fail to understand the pattern in the data. I would like to avoid Low bias and high variance problem, which will occur with higher alpha value.

Question 3

After building the model, you realized 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?

Ans: Below are the next top 5 features in order, we could recommend to customer to use while making decision about buying a house,

1. MSZoning_FV
2. GarageCars
3. BsmtQual
4. MSZoning_RL
5. MSZoning_RH

Question 4

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

Ans: To make sure that the model is more robust and generalizable, I would test the model with unseen data. Here, in building the house sale model, we have used all the date while finding the optimum value of alpha/lambda hyperparameter. I would keep aside a set of data from the data set to make sure that the model will not learn from that data set and then use it for testing. I would use outliers as well to test the robustness and generalizability. I would test the model with the SalePrice outlier I removed in the data set.

Implications or the inference that we could draw out of a robust and generalizable model with same accuracy with unknow data set is that it could be trusted to usable in reality, as its accuracy doesn't change much, if it is robust and generalizable.