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:

Property	Ridge				Lasso			
Optimal value of alpha	1.0				10.0			
Changes if value of alpha is doubled	Property	Optimal value	Double Value		Property	Optimal value	Double Value	
	Train r-squared	0.872	0.869		Train r-squared	0.869	0.866	
	Test r-squared	0.816	0.817		Test r-squared	0.818	0.819	
	Parameters	96	86		Parameters	100	100	
Most important predictor after change	**RemodAge**				**RemodAge**			

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: For Ridge I choose 1.0 and for Lasso I choose 10.0. Reason being that the predictive power of the model is highest at these values. If the value is changed then the predictive power of model will either decrease of will not change.

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?

Ans: After taking the mod of all variables. And arranging them in ascending order. I removed the top predictors. Then by building the model again, top 5 predictors found are: Age, MSSubClass_180, ScreenPorch, 1stFlrSF, MSSubClass_90

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?

Ans: To make sure that the model is robust and generalizable we have used the K-fold method to get the most optimal value of alpha and also checked if r-squared value of train data and test data does not have too much difference.

If the difference in the test and train r-squared value is low. It implies that the accuracy of model is good. But we should make sure that the value of r-square should not be too low as then model would loose it's predictive power.