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?

	Ridge (alpha = 100)	Lasso (alpha = 0.001)	Ridge (alpha = 200)	Lasso (alpha = 0.002)
R2 on train	.8822	.8852	.8778	.8817
R2 on test	.8827	.8818	.8818	.8810
RSS on train	18.65	18.19	19.35	18.74
RSS on test	8.72	8.78	8.78	8.84
MSE on train	0.018	0.018	0.019	0.018
MSE on test	0.020	0.020	0.020	0.020
Top Predictor	OverallQual	GarageCars	OverallQual	GarageCars
Top 5 Predictors	OverallQual GarageCars BsmtQual OverallCond BsmtFullBath	GarageCars OverallQual CentralAir BsmtQual BsmtFullBath	OverallQual GarageCars OverallCond BsmtQual BsmtFullBath	GarageCars OverallQual BsmtQual BsmtFullBath OverallCond

When we double the alpha values, we see slight decrease in the model performance for both Ridge and Lasso.

As we can see above that the 5 most important predictors in Ridge regression are not changing much. But, there is a slight change in Lasso as CentralAir has moved out of top 5 list when we doubled the value of alpha.

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?

After determining the optimal value of lambda for ridge and lasso regression, I will choose Lasso for below reasons:

1) Lasso helps in feature selection and removes the features are not very significant
2) Model performance(R2) in Lasso regression is slightly better than the Ridge regression as shown below:

	Ridge (alpha = 100)	Lasso (alpha = 0.001)
R2 on train	.8822	.8852
R2 on test	.8827	.8818
MSE on train	0.018	0.018
MSE on test	0.020	0.020
Top 5 Predictors	OverallQual GarageCars BsmtQual OverallCond BsmtFullBath	GarageCars OverallQual CentralAir BsmtQual BsmtFullBath

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?

After removing the top 5 most important predictor variables from the data set, below changes have been observed:

1) The optimal value of alpha is 0.01
2) R2 on train set is .8311
3) Top 5 most important predictors are now:
    - OverallCond
    - FireplaceQu
    - KitchenQual
    - Functional
    - BsmtExposure

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?

Model is robust and generalisable because of below facts:

1) The model is not overfitted. To avoid the overfitting, we have used regularisation technique here which uses the bias and variance trade off to make the model complexity acceptable and stable.
2) As we can see that the R2 score on train and test data set are comparable. The performance of the model is almost same on train and test data sets.
3) The model evaluation has been done using the cross-validation technique which makes the model more regressive on unseen data set. It makes the model performance more accurate as the model is not evaluated once but 5 times (folds)