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?

Regression Technique	Alpha
Ridge	6
Lasso	0.0001

Below are the observations after building model by doubling alpha

Regression Technique	Alpha	R2 score (Train)	R2 score (Test)	Observation
Ridge	6	0.8878	0.8718	R2 Score of Train and Test is decreased slightly by doubling alpha along with change in co-efficient values
	12	0.8751	0.8623	
Lasso	0.0001	0.8999	0.8675	R2 Score of Train is decreased but Test score is improved slightly by doubling alpha along with change in co-efficient values
	0.0002	0.8867	0.8703	

Most important predictor variables after the change is implemented:

Feature
BsmtFullBath
OverallCond
GarageArea
Neighborhood_NridgHt
Neighborhood_OldTown
MasVnrArea
Neighborhood_Timber
BsmtHalfBath
Fireplaces
BsmtExposure_Mn

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?

- Both ridge and lasso regression is assessed by checking r2 score of training and testing data.
- R2 value of both models are almost giving same results
- For current data set we have relatively higher number of features to assess the model.
- Hence Lasso model is preferred over Ridge since model contains large set of features With help of lasso we are able to set certain coefficients to exactly 0 helping us with variable(feature) selection. This variable selection results in models that are easier to interpret.

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?

Features
BsmtHalfBath
MSSubClass
Exterior1st_BrkComm
WoodDeckSF
Neighborhood_NridgHt

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?

Most important factor to consider for making model more robust and general is to keep the model as simple as possible. We also need to consider the fact that model should not become too simple as well. In order to handle this bias-variance trade off, we can apply Regularization (Lasso and Ridge) methods to arrive at optimal value of hyper parameter which helps to keep model more simple and robust. If model is built with cross-validation and Regularization techniques then we see good results on both train and testing data

A robust and generalizable model may have a lower accuracy on training data but a higher accuracy on unobserved data, which has implications for the model's accuracy. This is due to the fact that a model that is overly complicated and overfits the training set will underperform on unobserved data. In conclusion, it's critical to make a model robust and generalizable in order to guarantee that it will work well with unobserved data. This may result in slightly less accuracy on the training data, but it is frequently worthwhile to make this trade-off in order to improve performance on fresh, untested data.