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?

The optimal value of alpha for Ridge Regression is 2.

The optimal value of alpha for Lasso Regression is 0.0001.

The below table shows the comparison of model performance for Ridge Regression after doubling the alpha value.

	Metrics	Ridge Regression (alpha = 2.0)	Ridge Regression (alpha = 4.0)
0	R2 Score (Train)	0.923499	0.917262
1	R2 Score (Test)	0.904511	0.899900
2	RSS (Train)	2.174511	2.351781
3	RSS (Test)	1.237231	1.296975
4	MSE (Train)	0.046632	0.048495
5	MSE (Test)	0.053703	0.054984
6	Number of features	240.000000	240.000000

The below table shows the comparison of model performance for Lasso Regression after doubling the alpha value.

	Metrics	Lasso Regression (alpha = 0.0001)	Lasso Regression (alpha = 0.0002)
0	R2 Score (Train)	0.920582	0.912227
1	R2 Score (Test)	0.901871	0.897187
2	RSS (Train)	2.257414	2.494918
3	RSS (Test)	1.271439	1.332121
4	MSE (Train)	0.047512	0.049949
5	MSE (Test)	0.054440	0.055724
6	Number of features	131.000000	95.000000

The most important variable after doubling the value of alpha for Ridge Regression will be GrLivArea.

	Parameters	Coefficient Value
13	GrLivArea	0.224385
3	OverallQual	0.137813
10	TotalBsmtSF	0.111725
20	GarageArea	0.095014
34	TotFullBath	0.084400

The most important variable after doubling the value of alpha for Lasso Regression will be GrLivArea.

	Parameters	Coefficient Value
13	GrLivArea	0.345644
3	OverallQual	0.169870
10	TotalBsmtSF	0.127469
20	GarageArea	0.085895
89	Neighborhood_StoneBr	0.078157

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?

The below table shows the comparison between multiple linear regression (OLS), ridge regression and lasso regression.

	Metrics	Multiple Linear Regression	Ridge Regression (Alpha = 2.0)	Lasso Regression (Alpha = 0.0001)
0	R2 Score (Train)	0.932029	0.923499	0.920582
1	R2 Score (Test)	0.887668	0.904511	0.901871
2	RSS (Train)	1.932058	2.174511	2.257414
3	RSS (Test)	1.455454	1.237231	1.271439
4	MSE (Train)	0.043955	0.046632	0.047512
5	MSE (Test)	0.058247	0.053703	0.054440
6	Numbe of features	251.000000	240.000000	131.000000

From the above summary table it can be seen that model performance wise ridge regression performs marginally better than the lasso regression model, but I'll choose the lasso regression model as lasso regression as indirectly performed feature selection and has given a model which has a test score only marginally less in comparison to the ridge regression.

Also in general the ridge regression won't help with feature selection but lasso regression will help with feature selection and hence I'll choose lasso regression over ridge regression.

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?

The five most important variables for the original lasso model are as below:

GrLivArea, OverallQual, TotalBsmtSF, Neighborhood_StoneBr and GarageArea.

The five most important variables after removing the top 5 variables of the original lasso model are as below:

TotFullBath, LotArea, Conditons_Artery & PasA, Neighborhood_Gilbert and MasVnrArea.

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?

A model which has smaller number of features and whose coefficients are simpler with less number of digits after the decimal point are generally termed to be simple models. Simple models are generalizable to larger quantity of data points in comparison to models which have huge number of parameters, as the complex models will be very sensitive to the training data, if the training data changes there will be a fluctuation in the parameter values. In terms of accuracy a more robust model will have similar train and test accuracies but the values may be low as a simpler model will have a very high bias but low variance, thus for a model to be generalizable it must have a small number of features and the parameter values should also be simple rather than having many digits after the decimal point.