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?

Answer - The optimal value for the ridge is: 2 and Optimal value for lasso is: 100 regressions respectively

And if we try to double the values for ridge and lasso i.e. from α =2 to α =4 for ridge and for lasso α =100 to α =200 we will find that the r2 value for them changes as shown below

Normally decrease in r2 score in train and test

	Ridge (α =2)	Lasso (α =100)	Ridge (α =4)	Lasso (α =200)
R2 Train	91.74	91.07	91.19	89.99
R2 Test	90.25	90.43	90.01	89.54

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?

Answer -

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	9.228228e-01	9.174924e-01	9.107017e-01
1	R2 Score (Test)	-4.010215e+20	9.025976e-01	9.043741e-01

We can find for Lasso the test data is working pretty good than ridge so we will try to apply Lasso due to good fitting of data with respect to ridge,

Since lasso work better with model which have high number of feature which we can find useful here.

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?

Answer -

	Linear	Ridge	Lasso
GrLivArea	2.693952e+16	82876.459632	2.401447e+05
OverallQual	6.831562e+04	63297.363054	7.522377e+04
TotalBsmtSF	1.068428e+05	54417.697764	4.146809e+04
MasVnrArea	3.481250e+04	39095.544602	3.918365e+04
OverallCond	4.868520e+04	40429.424383	3.780737e+04

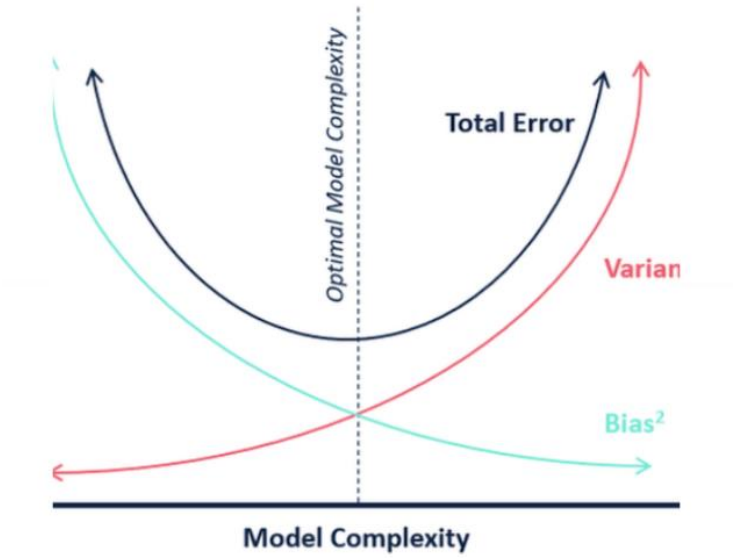
These are current top 5 feature if we we donot have these variable in out data set than the left out predictor variable are :

	Linear	Ridge	Lasso
BsmtFinSF1	1.915971e+04	35474.233666	3.281707e+04
ExterQual_2	2.631326e+04	26684.376414	2.567377e+04
KitchenQual_2	2.255173e+04	24591.189049	2.271597e+04
GarageCars	2.542062e+04	22547.765151	2.208081e+04
KitchenQual_0	2.816684e+04	24399.129266	2.141715e+04

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?

Answer – In order to make model more robust and generalisable we have to make our model simpler which mean we have to check with bias variance trade off



We need simple model that mean we should have less variance and more bias in a model to make our model more robust .

When we compare with respect to accuracy we want to see over fitting , underfitting and good fitting if we have training score more than test score that mean over fitting and of we have test score more than training score that mean underfitting and if we have both the training and test score are almost close to each other which almost overcome the fitting we can say model is good with respect to accuracy.