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?

A) The optimal values of Alpha for Ridge and Lasso are as below
   Ridge – 0.1
   Lasso – 0.0001

With increasing the alpha by doubling it, the regularization increases. Both Ridge and Lasso will push the coefficients towards 0, and Lasso might eliminate more variables.

```
Before changing Alpha

Train:-
R-squared (Train): 0.9532827103844842
Mean Squared Error (Train): 0.0073434538515147925
Root Mean Squared Error (Train): 0.08569395457974147

Test:-
R-squared (Test): 0.7253041479703011
Mean Squared Error (Test): 0.045197883444007284
Root Mean Squared Error (Test): 0.21259793847544073

Compare with below

Train:-
R-squared (Train): 0.9580291624903569
Mean Squared Error (Train): 0.006597362794333152
Root Mean Squared Error (Train): 0.0812241515457881

Test:-
R-squared (Test): 0.6813925840690762
Mean Squared Error (Test): 0.05242300072330667
Root Mean Squared Error (Test): 0.228960696896447
```

R-squared value has slightly increased for train, but decreased for test set.
MSE and RMSE has decreased for train, but increased for test.

The model with the doubled alpha performs better on the training set, but there is a trade-off as it performs slightly worse on the test set. This indicates a potential overfitting issue with the higher regularization strength

	Ridge
GrLivArea	0.542
1stFlrSF	0.524
LotArea	0.450
MSZoning_FV	0.368

MSZoning_RL 0.358
Exterior1st_BrkComm -0.190
Functional_Sev -0.282
OverallQual_2 -0.310
Condition2_PosN -1.012
PoolQC_Gd -1.521

PoolQC_Gd will be most important predictor after the change in 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?

A) As we already know that the efficiency of the model depends on the score from the Test set. We see that the Test from Ridge has R-square of 0.76 and Test from Lasso has R-square of 0.68, so I'll choose Ridge over Lasso here. But if the requirement is to keep the model as simple as possible we'll choose Lasso, as it has feature elimination and removed around 87 variables/features, it's simpler than Ridge.

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?

A) After removing top 5 features
PoolQC_Gd, Condition2_PosN, GrLivArea, 1stFlrSF,LotArea

And building a new model, we get below features as top 5 predictors
Condition2_PosN,PoolQC_Gd,MSZoning_FV,MSZoning_RL,MSZoning_RH

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) A model can be Robust, when it's varience is low, and it can generalizable when it can adapt very well to new unseen data. It can be achieved by making the system not too complex, and also not too simple. It's like trading off varience for achieving less bias. As the model gets complex it get high variance, and as it gets simple it gets high bias. So with help of Regularization techniques like Ridge, and Lasso we can make sure that the model is robust and generalizable.

Same applies for the Accuracy, for a model to be more accurate it has to be very complex, and we know that to make a complex to simplex, we have to compromise on bias(accuracy) to make it generalizable and reduce the variance. Same is done with Ridge, Lasso or any regularization techniques.