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:

Optimal value:

```
optimal value for alpha for Ridge : {'alpha': 20}
optimal value for alpha for Lasso : {'alpha': 0.001}
Before change model:
For the above alpha value of 20 for Ridge model score is as below:
Ridge r2 train score : 0.9119947994302648
Ridge r2 test score : 0.8927993758613273

For the above alpha value of 0.001 for lasso model score is as
below:
Lasso r2 train score : 0.9013056802607013
Lasso r2 test score : 0.8930870166728498
After alpha change model:
After doubling the alpha value(Ridge :40, Lasso: 0.002) impact on
the model is as below:

Ridge r2 train score : 0.9043653163514741
Ridge r2 test score : 0.8931468599628377

Lasso r2 train score : 0.887489277878646
Lasso r2 test score : 0.8889050776244498

Hence we can say that after doubling the alpha value r2 score
reduced on the train set for both model also in case of lasso r2
on test set also reduced and in case of lasso difference in train
set & test set score reduced which is good.
Predictor change after alpha double:
After doubling the alpha in case of lasso predictor variable
Foundation_PConc and MasVnrArea becomes important however
Neighborhood_StoneBr and BsmtExposure_Gd becomes less important.
```

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:

```
After comparison of Ridge and Lasso model performance we can say
that although there is not much difference in the r2 score, the
performance on train set and test set is better in case of lasso
model and also it provides the feature elimination. Hence I choose
to apply Lasso model.
```

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:

```
Dropping below 5 most important predictor
MSSubClass,OverallCond,BsmtFullBath,GarageArea,Neighborhood_Crawfo
r.

Lasso model r2 performance now becomes as below,
Lasso r2 train score : 0.8764116245262235
Lasso r2 test score : 0.8744833091893206

5 most important predictor now are:
```

	features	coef
0	LotArea	11,944866
3	MasVnrArea	0,115529
13	BsmtHalfBath	0,083468
21	PavedDrive	0,056770
117	Exterior1st_BrkFace	0,053630

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:

The model should be as simple as possible. Also in terms of number of feature it should be minimal to keep model simple.
It can be understood using bias-variance tradeoff as simpler model will have more bias and less variance on training set however it will generalise better and so better performance on test set.

Robust and generalized model can perform equally good on train and test set by avoiding overfitting with less features.