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 of Ridge : 10

The optimal value of Lasso : 100

There are no changes in the r2 score for the ridge and lasso regression after doubling of the alpha values but the coefficient value does change. We don't see huge change in the model even after doubling the alpha values.

```
Neighborhood_NoRidge is the most important predictor after the change i
s implemented.

MSZoning', 'LotArea', 'LandContour', 'LotConfig', 'LandSlope', 'Neighbo
rhood', 'Condition1', 'BldgType', 'HouseStyle', 'OverallQual', 'Overall
Cond', 'YearBuilt', 'YearRemodAdd', 'RoofMatl', 'Exterior1st', 'Exterio
r2nd', 'MasVnrType', 'MasVnrArea', 'Foundation', 'BsmtCond', 'BsmtExpos
ure', 'BsmtFinType1', 'BsmtFinType2', 'BsmtFinSF2', 'BsmtUnfSF', 'Total
BsmtSF', '2ndFlrSF', 'GrLivArea', 'BsmtFullBath', 'BsmtHalfBath', 'Full
Bath', 'HalfBath', 'TotRmsAbvGrd', 'Functional', 'Fireplaces', 'GarageT
ype', 'GarageCars', 'WoodDeckSF', 'EnclosedPorch', '3SsnPorch', 'Screen
Porch', 'SaleType'

This are positive parameters which are most 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?

```
- r2_score in train dataset:
   - r2_score for ridge: 0.89235125467758
   - r2_score for lasso: 0.8999225441409708

- r2 score in test dataset:
   - r2_score for ridge: 0.88
   - r2_score for lasso: 0.87
```

Ridge regression because it has good r2 score for train as well as test dataset as compared to that of Lasso Regression

	Metric	Ridge Regression	Lasso Regression
0	R2 Score (Train)	8.923513e-01	8.999225e-01
1	R2 Score (Test)	8.769232e-01	8.670007e-01
2	RSS (Train)	7.285403e+11	6.772997e+11
3	RSS (Test)	3.002826e+11	3.244916e+11
4	MSE (Train)	2.579406e+04	2.487043e+04
5	MSE (Test)	2.868261e+04	2.981641e+04

By comparing Ridge is better than that of Lasso Regression but there is no major differences between them.

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?

```
('Neighborhood_Crawfor', 19246.625), ('RoofMatl_WdShngl', 21208.91), ('
GrLivArea', 24131.973), ('Neighborhood_NridgHt', 27323.348), ('Neighbor
hood_NoRidge', 37730.022)
```

These are the top 5 most important variables now

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?

We can say model is generic when it require fewer training samples for effective training than the more complex and which are easier to train as well.

Models are robust when it does not change wildly with changes in training data set.

Simpler models has low variance , high bias
Complex models have low bias , high variance

Simpler models make errors in training dataset whereas complex model leads to overfitting
To make model robust and generic, me make model simple but not so simple that it underfits as well.

To make models simpler we use the Regularization which we have applied in this case study.