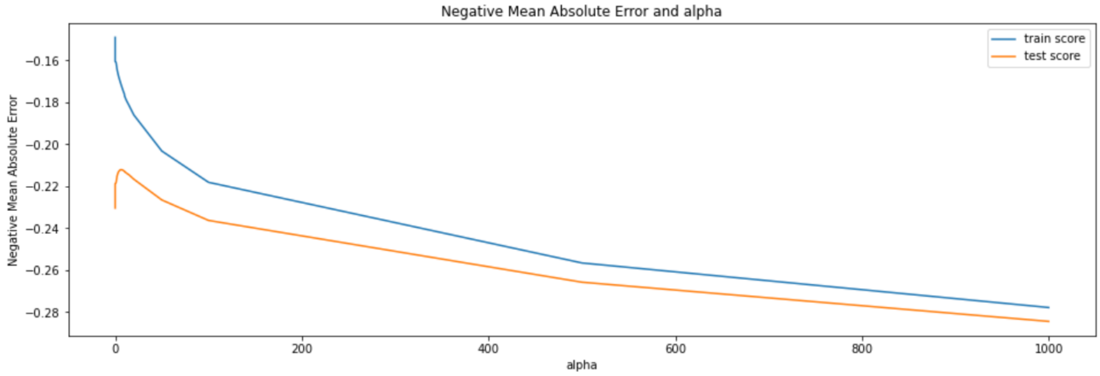
Subjective Questions

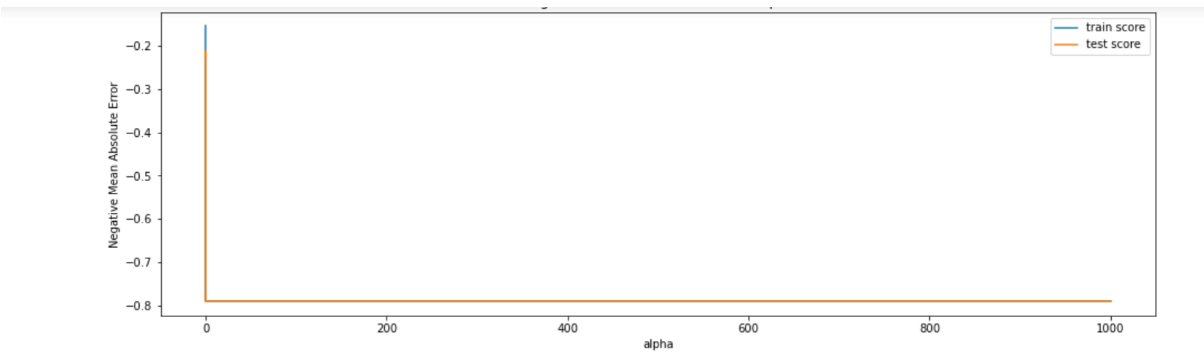
1. **What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?**

Ans:

Optimal Value:
Ridge: 6.0
Lasso : 0.001
In Ridge Alpha vs Neg mean absolute error score can be seen that with increase in alpha Mean error score reduces.
Fig below shows Ridge Graph



But in Lasso we can see alpha value should be less to make error close to zero. With increase in alpha mean error is increasing and greater value of mean score almost becomes 1 which is maximum value of error score to be reached. Hence in lasso we mustn't increase alpha but its better we decrease alpha to get better score.
Fig below shows Lasso Graph



Top predictor value changed in Ridge :

```
In [333]: betas_question['Ridge_12'].sort_values(ascending=False).head()

Out[333]: GrLivArea                0.276946
          Functional_Typ           0.235331
          OverallQual_Excellent    0.232694
          Neighborhood_Crawfor     0.219314
          OverallQual_Very_Good    0.204171
          Name: Ridge_12, dtype: float64
```

Top predictor value changed in Lasso :

```
In [334]: betas_question['Lasso_0.002'].sort_values(ascending=False).head()

Out[334]: OverallQual_Excellent    0.366540
          GrLivArea                0.363908
          Functional_Typ           0.262494
          Neighborhood_Crawfor     0.256675
          OverallQual_Very_Good    0.239217
          Name: Lasso_0.002, dtype: float64
```

Detailed Analysis(Refer Notebook as well)

```
In [784]: new_ridge=ModelLassoandRidgeChanged(15,True,X_train,y_train,X_test,y_test,12.0)

0.9327314998726124
0.9160001686640548
The MSE of the model on the test dataset for optimum alpha is 0.08399983133594523
```

```
In [785]: new_lasso=ModelLassoandRidgeChanged(15,False,X_train,y_train,X_test,y_test,0.002)

0.92412681053262
0.9146610105670694
The MSE of the model on the test dataset for optimum alpha is 0.0853389894329306
```

For better reference please check notebook

Here model was trained on Ridge and Lasso for double value .The optimum value obtained was Ridge alpha =6 and Lasso alpha =0.001 .

After doubling (Ridge alpha=12 and Lasso alpha =0.002)

- We can see improvement in R2 test score and reduction of MSE in ridge
- We can see the MSE for Lasso has gone up but very less variation can be seen.But yes increase of MSE indicates model can be better
- We can see in both Ridge and Lasso the R2 scores on train reduced with double of alpha.
- We can clearly see that feature 'SaleCondition_Alloca' and 'RoofMatl_WdShngl' is penalised and coeffiecent is very very less .And lasso has removed both of them

Out[788]:

	Ridge_12	Lasso_0.002	Ridge	Lasso
GrLivArea	0.276946	0.363908	0.311575	0.359464
OverallQual_Excellent	0.232694	0.366540	0.265588	0.337086
Functional_Typ	0.235331	0.262494	0.259892	0.299671
Neighborhood_Crawfor	0.219314	0.256675	0.248996	0.286721
OverallCond_Excellent	0.173502	0.192925	0.217277	0.251027
Neighborhood_StoneBr	0.184651	0.198187	0.226731	0.249646
Exterior1st_BrkFace	0.171830	0.217248	0.193461	0.236472
OverallQual_Very_Good	0.204171	0.239217	0.221034	0.234442
SaleCondition_Alloca	0.120085	0.000000	0.191020	0.193150
RoofMatl_WdShngl	0.111270	0.000000	0.177895	0.183107

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?

Ans:

Out[67]:

	Metric	Linear Regression with RFE	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.871161	0.937838	0.932899
1	R2 Score (Test)	0.813077	0.914665	0.915403
2	Diffrence in R2 test and train	0.058084	0.023173	0.017497
3	RSS (Train)	131.545116	63.467463	68.509785
4	RSS (Test)	81.872408	37.376853	37.053583
5	MSE (Train)	0.128839	0.062162	0.067101
6	MSE (Test)	0.186923	0.085335	0.084597

Lasso is better model

Reason

- r2 score of train and test of Lasso is 1.74 but ridge is 2.32.Hence the difference between the r2 score of train and test should be less
- r2 score for test of Lasso is better than ridge. In regression we consider r2 score test to be better
- MSE on test data for lasso is lower than MSE on test data for ridge. Lower MSE indicates lower error in prediction

In general ,

Regularization indicates improving accuracy also makes model more general.

Ridge and Lasso does same only difference is way they penalise is different. Ridge tuning param lambda is square of magnitude of coefficients. Penalty is sum of squares hence coefficients having greater value gets penalised. Ridge includes all the variables in final model unlike Lasso. Ridge can only reduce complexity of model but not advisable for feature selection as final model includes all .Hence Ridge model is always not flexible or in other words never simple as per Occam Razors

Lasso also uses lambda as finetuning parameter but add absolute sum of coefficients .As value of coefficients increases from 0 lambda penalises them making models decrease value. Lasso makes coefficients to 0 hence ideal for feature selection.

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?**

   Ans:

   The top predictors were

	Lasso
GrLivArea	0.359464
OverallQual_Excellent	0.337086
Functional_Typ	0.299671
Neighborhood_Crawfor	0.286721
OverallCond_Excellent	0.251027

We remove these and train and create a new model.
X_train and X_test dropped these columns shown in notebook and Lasso model with alpha=0.001 was used to build it again.
This time new set of coefficients have come in model.Among the top predictors were

	Lasso Co-Efficient
RoofMatl_WdShngl	0.331958
Exterior1st_BrkFace	0.275331
2ndFlrSF	0.258616
Neighborhood_StoneBr	0.244307
SaleCondition_Alloca	0.212651

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?

Ans:
As per Occam's Razor ,
Given 2 models having similar 'performance' in finite test and train data we should always pick simpler models
Why should we have simple models ?

- Simple models are generic ie; less variance
- Need less training data ,easy to train
- Simple models are robust
  - Complex model change too much if there is small change in training data
  - Simple are low variance and high bias but complex models are opposite

How to make model simple?
Use appropriate feature selection mechanism , domain know-how to decrease on insignificant features. Feature selection is important key to unlock simple model.
But as mentioned earlier simple model can have high bias and lower variance
Hence bias is higher means train accuracy is lower also test lower .Lower variance means the model accuracy for test and train doesn't show significant difference hence very less overfitting can be seen.

That also means extremely simple models are not advisable as model bias should always be made less ie; accuracy of train and test should be aimed to increase. Hence a cautious call need to taken with respect to and considering bias-variance trade off.

Ideally a model should have low bias and low variance ie good accuracy along with difference in accuracy seen on test and train is negligible.