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 1) The optimal value of Alpha in the ridge and lasso regression are as follows :-

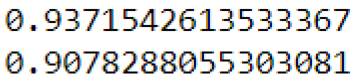
Lasso Alpha= 0.0001

Ridge Alpha= 2

For Lasso:-

The R2 score when alpha is 0.0001 is 0.9371 for y_train and 0.9078 for y_test.

Kindly refer to the Screenshot of the Output for your reference.

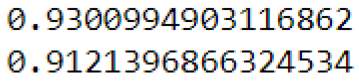


```
0.9371542613533367
0.9078288055303081
```

For Ridge:-

The R2 score when alpha is 2 is 0.9300 for y_train and 0.9121 for y_test.

Kindly refer to the Screenshot of the Output for your reference.



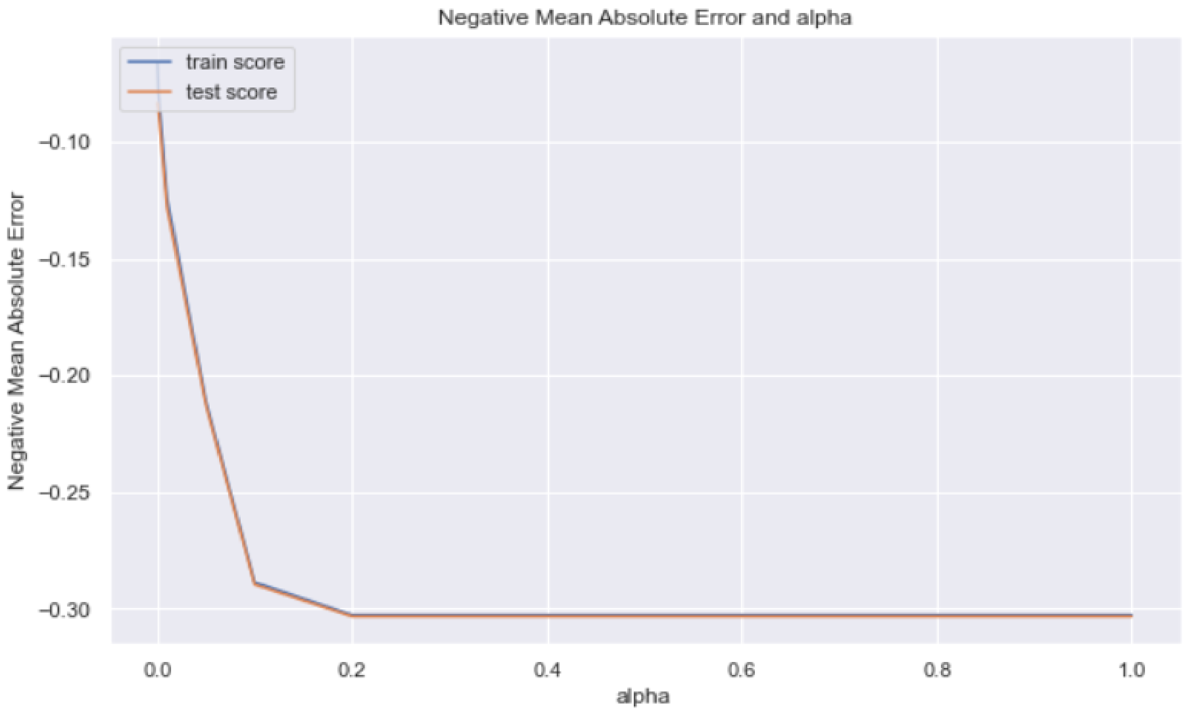
```
0.9300994903116862
0.9121396866324534
```

Once we double the alpha values in Lasso and Ridge they will become.

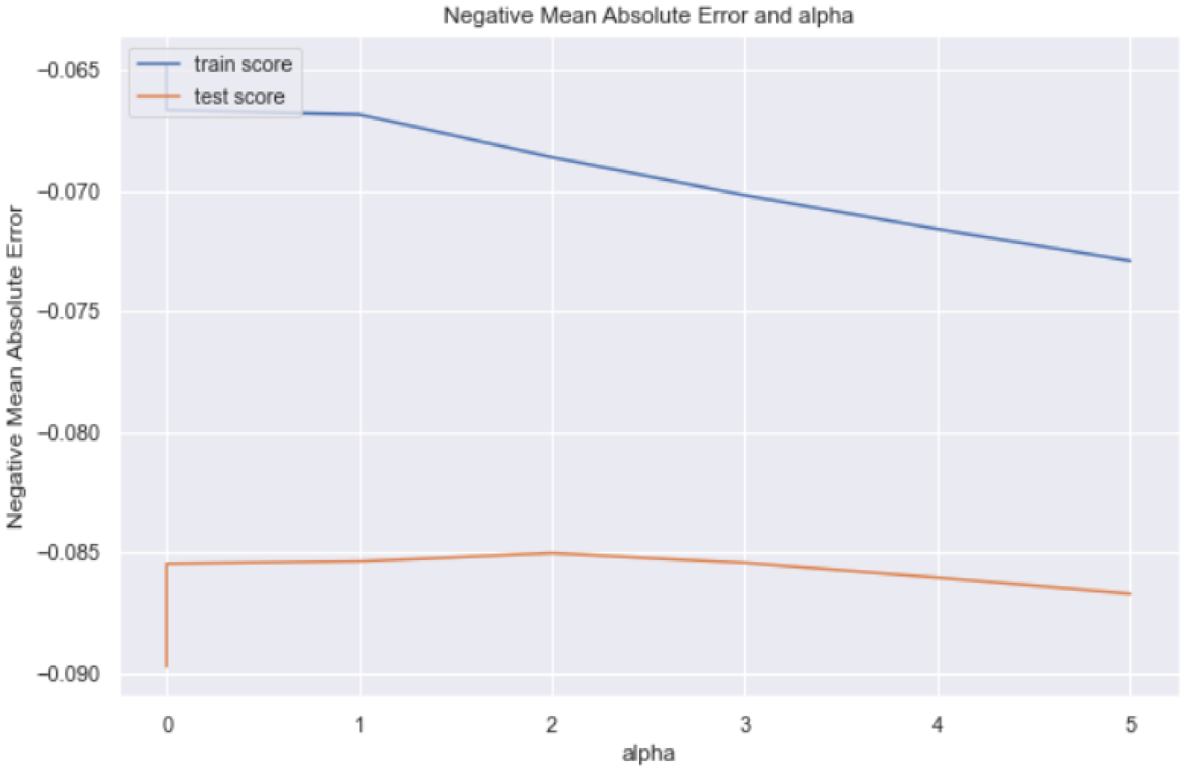
Lasso Alpha= 0.0002

Ridge Alpha= 4

The Graph representing Negative Mean Absolute Error with respect to alpha in Lasso.



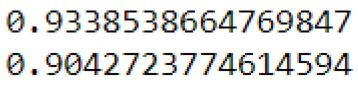
The Graph representing Negative Mean Absolute Error with respect to alpha in Ridge.



For Lasso:-

The R2 score when alpha is 0.0002 is 0.9338 for y_train and 0.9042 for y_test.

Kindly refer to the Screenshot of the Output for your reference.

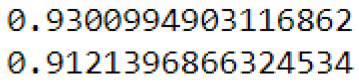


```
0.9338538664769847
0.9042723774614594
```

For Ridge:-

The R2 score when alpha is 4 is 0.9338 for y_train and 0.9042 for y_test.

Kindly refer to the Screenshot of the Output for your reference.

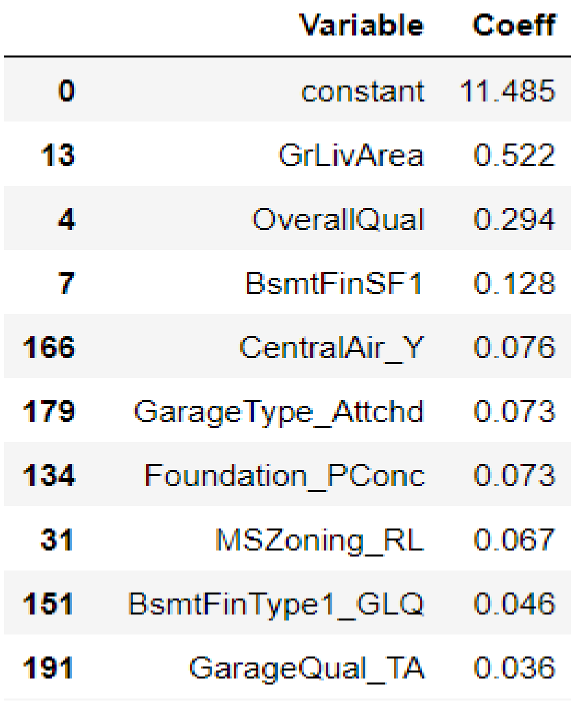


```
0.9300994903116862
0.9121396866324534
```

So, once we double the value of Alpha for the Ridge regression, the new alpha value will become 4 and it will apply more penalty on the curve, and it will try to make it more generalized. This can be seen in the above-mentioned graph where the train and test lines are far-away from each other, and hence it has more error.

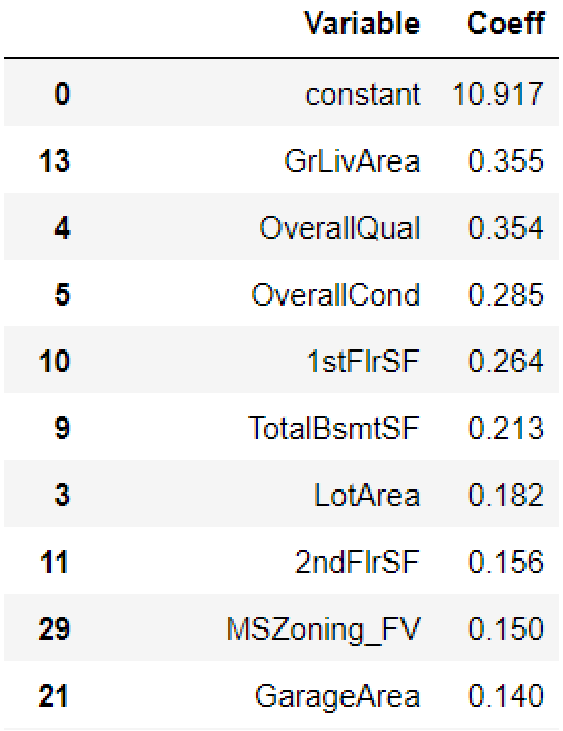
Similarly, if we double the value of Alpha for Lasso Regression we will penalize more in our model and hence it will try to reduce more Coefficients to 0. As we increase the Alpha value from 0.0001 to 0.0002 the R2 score decreases.

The most important Factors post the change in Lasso are:-



	Variable	Coeff
0	constant	11.485
13	GrLivArea	0.522
4	OverallQual	0.294
7	BsmtFinSF1	0.128
166	CentralAir_Y	0.076
179	GarageType_Attchd	0.073
134	Foundation_PConc	0.073
31	MSZoning_RL	0.067
151	BsmtFinType1_GLQ	0.046
191	GarageQual_TA	0.036

The most important Factors post the change in Ridge are:-



	Variable	Coeff
0	constant	10.917
13	GrLivArea	0.355
4	OverallQual	0.354
5	OverallCond	0.285
10	1stFlrSF	0.264
9	TotalBsmtSF	0.213
3	LotArea	0.182
11	2ndFlrSF	0.156
29	MSZoning_FV	0.150
21	GarageArea	0.140

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 2) The optimal value of lambda for Ridge and Lasso are as follows: -

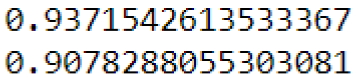
Lasso Alpha= 0.0001

Ridge Alpha= 2

For Lasso:-

The R2 score when alpha is 0.0001 is 0.9371 for y_train and 0.9078 for y_test.

Kindly refer to the Screenshot of the Output for your reference.



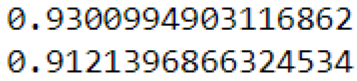
```
0.9371542613533367
0.9078288055303081
```

The RMSE value for lasso is 0.114

For Ridge:-

The R2 score when alpha is 2 is 0.9300 for y_train and 0.9121 for y_test.

Kindly refer to the Screenshot of the Output for your reference.



```
0.9300994903116862
0.9121396866324534
```

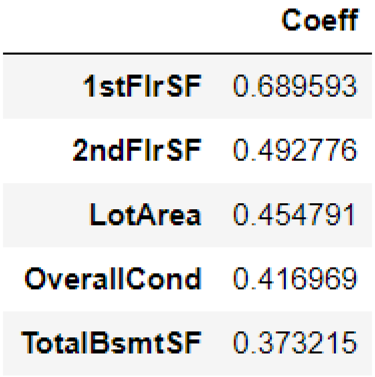
The RMSE value for Ridge is 0.112

The **Lasso model is better** to interpret the dataset, yet according to the R2 score the Ridge performs better than Lasso by a small difference. As the difference is not huge we will prefer the Lasso model as in terms of interpretability it is more generalized and it assigns 0 to the insignificant attributes making the prediction easier.

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 3) The Five most important predictor variables post removing the earlier top five predictors from the dataset in Lasso model are as follows:-



	Coeff
1stFlrSF	0.689593
2ndFlrSF	0.492776
LotArea	0.454791
OverallCond	0.416969
TotalBsmtSF	0.373215

The predictor variable mentioned above define the following: -

1stFlrSF - 1st Floor Square Feet.

2ndFlrSF- 2nd Floor Square Feet.

LotArea- Lot size in squares.

OverallCond- Rates the Overall Condition of the house.

TotalBsmtSF- Total Square Feet of the basement area.

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 4) To make sure that a model is Robust and Generalisable, we need to understand what they mean and how they can be corelated.

The term generalization is used to describe how good the model is at predicting the new instances which it didn't see before. Ideally speaking the model should generalize the relationship as same as during the training phase.

Whereas on the other hand the term regularization is referred to as a process that enhances the generalization capabilities of the model. Regularization is the process of shrinking or regularizing the coefficients towards zero in machine learning. To put it another way, regularization prevents overfitting by discouraging the learning of a more complicated or flexible model.

In most scenarios, we generally focused on the training set or tried to optimize the performance on the training set. But this is not the correct way of model building, there is a lot of uncertainty (such as noise) in the unseen data which is taken from the same distribution as the training set. So in such cases, we also should aim for our model that can generalize well on those unseen data.

Fortunately, there is a simple method for assessing a model's generalization performance. Simply put, we divide our data into three subsets.

- A training set is a collection of training examples on which the network is trained.

- A validation set is used to fine-tune hyperparameters like the number of hidden units and the learning rate.
- A test set designed to evaluate generalization performance.

The losses on these subsets are referred to as training, validation, and test loss, in that order. It should be evident why we need distinct training and test sets: if we train on test data, we have no notion if the model is correctly generalizing or merely memorizing the training examples.

There are other variations on this basic method, including what is known as cross-validation. These options are typically employed in cases with tiny datasets, i.e. less than a few thousand examples. Most advanced machine learning applications include datasets large enough to be divided into training, validation, and test sets.

Apart from all these techniques, there is one called Regularization. Regularization has no effect on the algorithm's performance on the data set used to learn the model parameters (feature weights). It can, however, increase generalization performance, i.e., performance on new, previously unknown data, which is exactly what we want.

So in order for a model to be robust the output dependent variable should be consistently accurate, even if one or more of the input independent attributes or features are drastically changes due to any unforeseen circumstances. In order for a model to be better generalised we need to fit the model with dataset which is not biased and the model should be trained in such a way that it avoids any kind of overfitting and underfitting.

The poor generalization is due to problems like overfitting, underfitting, and bias-variance issues this in turn effects the accuracy of the model. We can easily interpret the generalization status of our model by just observing the training and validation accuracy or testing accuracy scores. Whenever there is a poor generalization, we apply one of the regularization techniques like Lasso and Ridge. It penalizes the magnitude of coefficients which is responsible for generalization towards zero.