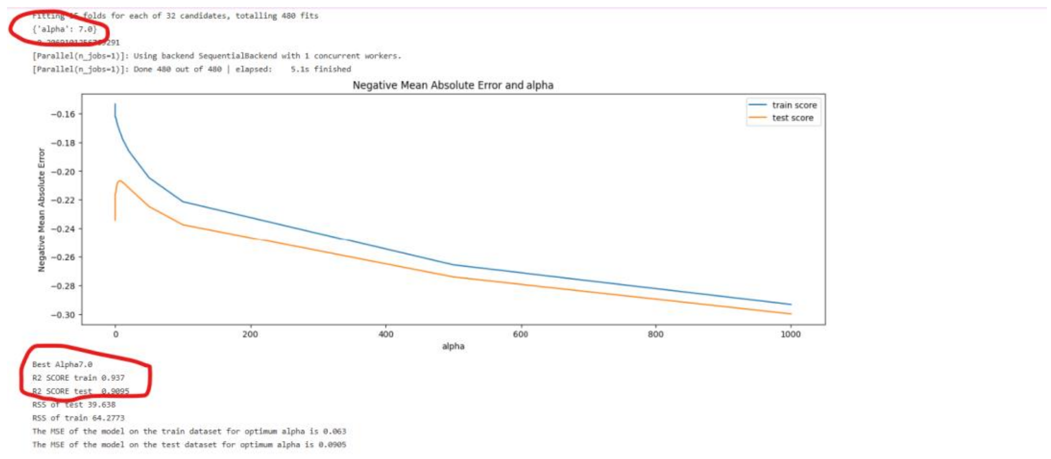
Question 1

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?

Optimal value of alpha(hyperparameter) for Ridge Regression is **11.5**

Optimal value of alpha(hyperparameter) for Lasso Regression is **0.001**

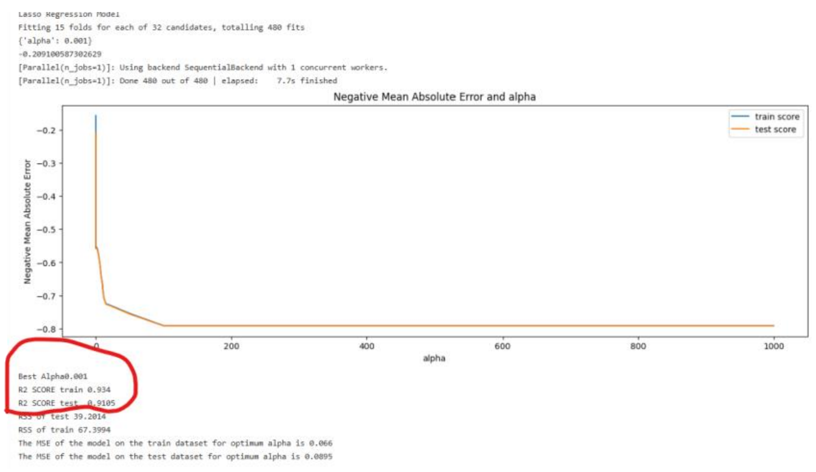


```
Fitting 15 folds for each of 32 candidates, totalling 480 fits
{'alpha': 7.0}
-0.2069[illegible]291
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 480 out of 480 | elapsed:    5.1s finished
```

Negative Mean Absolute Error and alpha
Negative Mean Absolute Error
alpha
train score
test score

```
Best Alpha7.0
R2 SCORE train 0.937
R2 SCORE test  0.9095
RSS of test 39.638
RSS of train 64.2773
The MSE of the model on the train dataset for optimum alpha is 0.063
The MSE of the model on the test dataset for optimum alpha is 0.0905
```

In ridge regression, increase in alpha reduces the negative mean absolute error terms



```
Lasso Regression Model
Fitting 15 folds for each of 32 candidates, totalling 480 fits
{'alpha': 0.001}
-0.209100587302629
[Parallel(n_jobs=1)]: Using backend SequentialBackend with 1 concurrent workers.
[Parallel(n_jobs=1)]: Done 480 out of 480 | elapsed:    7.7s finished
```

Negative Mean Absolute Error and alpha
Negative Mean Absolute Error
alpha
train score
test score

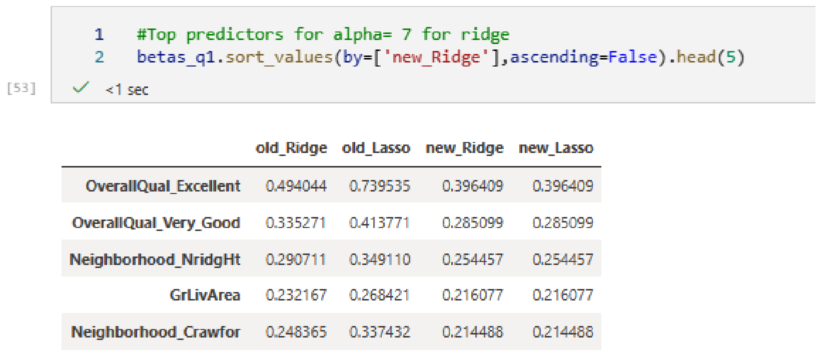
```
Best Alpha0.001
R2 SCORE train 0.934
R2 SCORE test  0.9105
RSS of test 39.2014
RSS of train 67.3994
The MSE of the model on the train dataset for optimum alpha is 0.066
The MSE of the model on the test dataset for optimum alpha is 0.0895
```

In Lasso regression, we must keep alpha less to reduce the error terms minimal.

When we double the values of the alpha (Ridge alpha=14 and Lasso alpha =0.002)

- We can see there is a slight reduction of R2 value when we double the alpha for ridge regression.
- We can see the MSE for Ridge has gone up for the ridge regression.
- We can see in both Ridge and Lasso the R2 scores on train reduced with double of alpha.
- We can see with an increase in alpha of the ridge regression, the test r2 score has slightly improved, which is a positive sign.
- We can see with the increase in the alpha for the lasso regression, the test r2 score has reduced a bit.
  - We can also observe that some of the features like SaleType_New, Screen Porch, Full Bath has completely been penalized with almost 0 coefficients.

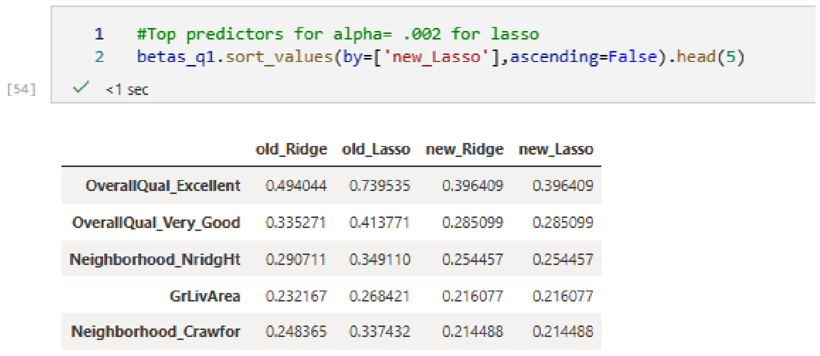
Top predictors for ridge and lasso



```
#Top predictors for alpha= 7 for ridge
betas_q1.sort_values(by=['new_Ridge'],ascending=False).head(5)
```

[53] ✓ <1 sec

	old_Ridge	old_Lasso	new_Ridge	new_Lasso
OverallQual_Excellent	0.494044	0.739535	0.396409	0.396409
OverallQual_Very_Good	0.335271	0.413771	0.285099	0.285099
Neighborhood_NridgHt	0.290711	0.349110	0.254457	0.254457
GrLivArea	0.232167	0.268421	0.216077	0.216077
Neighborhood_Crawfor	0.248365	0.337432	0.214488	0.214488



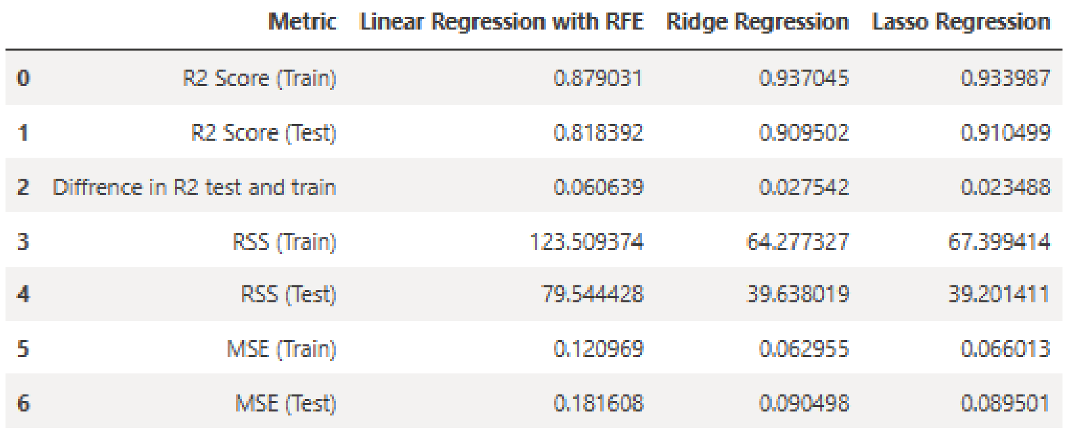
```
#Top predictors for alpha= .002 for lasso
betas_q1.sort_values(by=['new_Lasso'],ascending=False).head(5)
```

[54] ✓ <1 sec

	old_Ridge	old_Lasso	new_Ridge	new_Lasso
OverallQual_Excellent	0.494044	0.739535	0.396409	0.396409
OverallQual_Very_Good	0.335271	0.413771	0.285099	0.285099
Neighborhood_NridgHt	0.290711	0.349110	0.254457	0.254457
GrLivArea	0.232167	0.268421	0.216077	0.216077
Neighborhood_Crawfor	0.248365	0.337432	0.214488	0.214488

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?



	Metric	Linear Regression with RFE	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.879031	0.937045	0.933987
1	R2 Score (Test)	0.818392	0.909502	0.910499
2	Diffrence in R2 test and train	0.060639	0.027542	0.023488
3	RSS (Train)	123.509374	64.277327	67.399414
4	RSS (Test)	79.544428	39.638019	39.201411
5	MSE (Train)	0.120969	0.062955	0.066013
6	MSE (Test)	0.181608	0.090498	0.089501

```
- R2 score of lasso model is better
- MSE and RSS on lasso is lowest
```

Lasso model is better

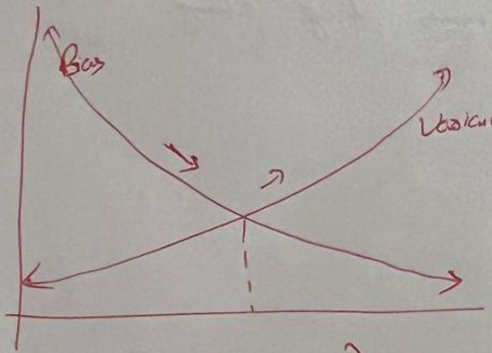
The only difference between the ridge and lasso is the way they penallise the error terms. Ridge includes all the features in the final model, while the lasso eliminates certain features by penalizing them to 0.Ridge tuning param lambda is square of magnitude of coefficients and the penalty is sum of squares hence coefficeients having greater value gets penalised.Ridge is not as simple or flexible.

Lasso also uses lambda as hyper parameter, but add absolute sum of coefficients as value of the coefficients increases from 0 lambda penalizes them making models decrease value. Lasso makes coefficients to 0 , hence we can use for feature selection when the features are too much and the problem of overfitting exists.

Ridge Regression.

For regularization, we add Penalty to Cost function

Trade Bias for reduction in variance



$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$Cost = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \underline{\lambda} \sum_{j=1}^{n} \beta_j^2$$

Lasso Regression

- Model Interpretation is difficult with ridge, as it does not do feature elimination

Lasso Regression Cost

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \sum_{j=1}^{n} |\beta_j|$$

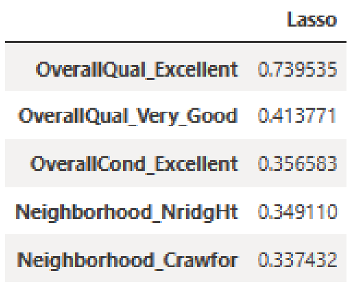
Question 3

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?

The top predictors of the final model on the ridge and lasso side are as follows

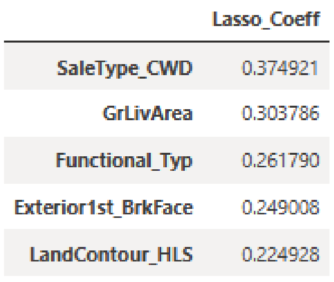
Lasso

Before removing top predictors



	Lasso
OverallQual_Excellent	0.739535
OverallQual_Very_Good	0.413771
OverallCond_Excellent	0.356583
Neighborhood_NridgHt	0.349110
Neighborhood_Crawfor	0.337432

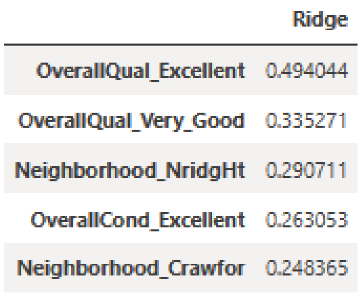
After removing top predictors and create new model



	Lasso_Coeff
SaleType_CWD	0.374921
GrLivArea	0.303786
Functional_Typ	0.261790
Exterior1st_BrkFace	0.249008
LandContour_HLS	0.224928

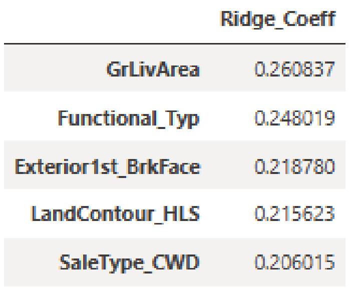
Ridge

Before removing top predictors



	Ridge
OverallQual_Excellent	0.494044
OverallQual_Very_Good	0.335271
Neighborhood_NridgHt	0.290711
OverallCond_Excellent	0.263053
Neighborhood_Crawfor	0.248365

After removing the top predictors and creating new model



	Ridge_Coeff
GrLivArea	0.260837
Functional_Typ	0.248019
Exterior1st_BrkFace	0.218780
LandContour_HLS	0.215623
SaleType_CWD	0.206015

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?

As per Occam’s Razor , Given 2 models having similar ‘performance’ in finite test and train data we should always pick simpler models

In-order to make a model is robust and generalisable we use techniques like regularization and grid search cv(cross-validation). The cross - validation splits the data into train, validation and test sets so that it doesnt sneak-peak into the test data while training. We also use back testing technique to evaluate the performance of the model in unseen data.

The implication of making a robust model and a generalizable model is that it may decrease in accuracy on the training set but will result in a increase of accuracy on the test set. The robust model is very unlikely to overfit on the train data and more likely to generalize on the new data.

we should choose simple models because they are generic (less variance) and needs less training data and easy to train. Simple models are robust while complex model change too much if there is a small change in the training data . Simple models tend to have low variance but high bias.

In case of linear regression problem, the simple model means selection of appropriate features and having lesser number of features. We can use RFE techniques combined with our business intuition to choose the right set of features and remove insignificant features.

Simple models can have high bias and low variance. Lower variance means the model accuracy for the test and train doesnt show significant difference hence very less overfitting can be observed. Also, extremely simple models are not advisable as model bias should be made less (accuracy should be increased).

Hence we can do a cautious call based on the bias-variance trade off.Simpler models are more likely to be correct because they capture the esssence of the problem without overcomplicating it.