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 Values obtained from the model are:
For Ridge: 3
For Lasso: 0.0001

If we further double the value of alpha for both ridge and lasso, the coefficients attained for this further increases.
The values of Ridge Coefficients before change:

```
array([ 0.00834711,  0.01785953,  0.07468984,  0.03662499,  0.01549659,
       -0.0104559 ,  0.04685621,  0.11349083,  0.00874511,  0.03621555,
        0.01098445,  0.01098919, -0.01994957, -0.02959229,  0.01218282,
        0.01426202,  0.011921  ,  0.00893938,  0.0169886 ,  0.01514262,
        0.01444668,  0.01399868, -0.01862705,  0.01724551,  0.07223607,
        0.02780612,  0.1193712 ,  0.08916066,  0.01931235, -0.01071828,
       -0.01268185,  0.02212435, -0.02151298,  0.01241693, -0.00048968,
       -0.00285983,  0.01170888, -0.01282511,  0.01133163, -0.03963872,
       -0.02571184,  0.01628899, -0.00892485,  0.03024845,  0.01714139,
        0.01994835,  0.03959117,  0.0117096 ,  0.00910327, -0.01193149])
```

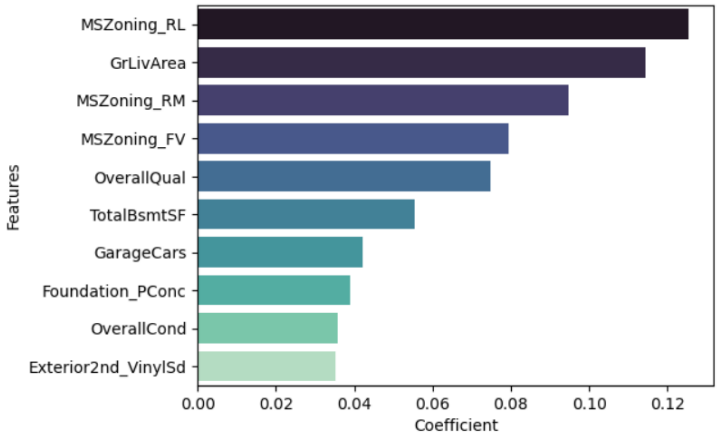
The values of Ridge Coefficients after change:

```
array([ 0.00847862,  0.01776393,  0.07524297,  0.03689817,  0.01619217,
       -0.00987397,  0.0464668 ,  0.11294708,  0.00863576,  0.03581785,
        0.01101509,  0.01116841, -0.02037084, -0.02983547,  0.01184629,
        0.01407448,  0.01179607,  0.0090347 ,  0.0169761 ,  0.01540506,
        0.01482239,  0.01392441, -0.01821299,  0.01787901,  0.06316516,
        0.0232786 ,  0.101801  ,  0.07311829,  0.01924485, -0.01045213,
       -0.01267281,  0.02178172, -0.02113064,  0.0121498 , -0.00031492,
       -0.00287917,  0.01180402, -0.01045963,  0.01134889, -0.03523871,
       -0.02485626,  0.01422905, -0.00886656,  0.02625351,  0.01634869,
        0.0187054 ,  0.03839566,  0.01145391,  0.00768097, -0.01216442])
```

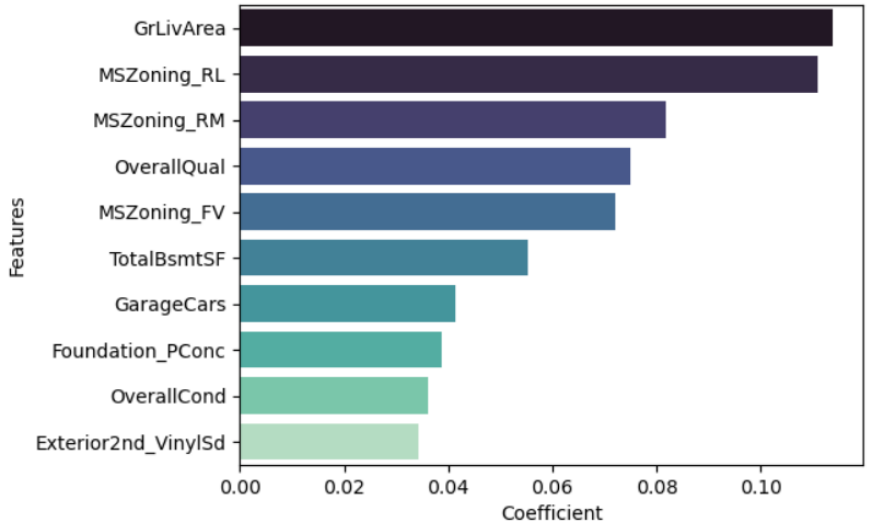
Likewise lasso has The MSE values will be greater values compared to prior case.

However, Lasso has significantly less value than Ridge.Ridge regression is utilised instead of Lasso. In order to select the most important variables for house price prediction, the variables predicted by Lasso can be used.

The most important predictor variables before the change is implemented:



The most important predictor variables after the change is implemented:



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:

When comparing Ridge and Lasso, we also consider other measures. Lasso is worth a little less than ridge. Ridge regression is utilized instead of Lasso.

In order to select the most important variables for house price prediction, the variables predicted by Lasso can be used. Additionally, Lasso aids in feature selection and produces model parameters such that the coefficients of the less significant features become zero.

Additionally, Lasso has the advantage of being computationally intensive.

Question 3:

After building the model, you realized 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:
The five most important predictor variables excluding the five most important predictor variables, after dropping the before 5 most predictor variables
-> 2ndFlrSF
-> 1stFlrSF

-> TotalBsmtSF
-> OverallCond
=> Foundation_PConc

The five most important predictor variables based on the coefficients

-> MSZoning_RL
-> GrLivArea
-> MSZoning_RM
-> MSZoning_FV
-> OverallQual

Question 4:

How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Answer:

Even though the two models perform similarly on finite data, you should choose the one that doesn't overfit the test data for the following reasons: -

1) Compared to Complex models, which have a tendency to vary as the training data set changes, Simpler models are typically more "generic" and are more widely applicable.

2) Less complex models are easier to train because they require fewer training samples for effective training than do simpler models. More reliable models are simpler ones.

3) Simple models have high bias and low variance, while complicated models have high bias and low variance.

4) Complex models result in overfitting, whereas simpler models produce more errors in the training set.

Therefore, simplify the model without making it vain in order to increase its robustness and generalizability. The model can be made easier by regularization. Regularization aids in striking the difficult balance between keeping the model straightforward and preventing it from being overly simplistic and useless. Regression regularization entails multiplying the squares or absolute values of the model's parameters by a regularization term, which is added to the cost.

Additionally, a model's bias-variance trade-off results from simplification.

A complicated model is exceedingly unstable and highly sensitive to any changes in the training data because it must be changed for any tiny change in the dataset.

Thus, by maintaining the proper balance between Bias and Variance, as demonstrated in the graph below, the model's accuracy may be preserved while also minimizing overall error.