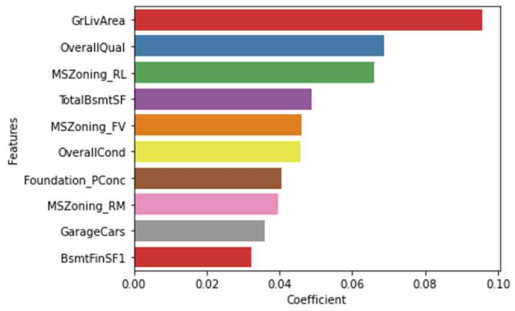
Question 1

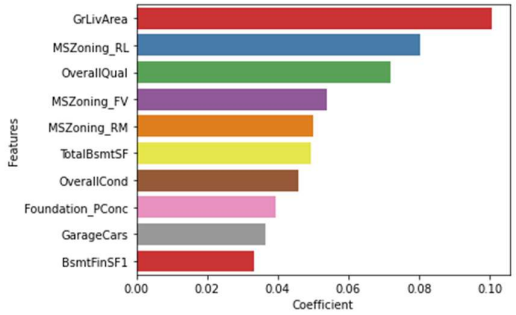
What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?

- ⇨ The optimal value of alpha for ridge and lasso as follows 10 and 0.0004.
- ⇨ Below image I have doubled the value of ridges alpha



And image clearly shows that GrLiveArea is most important feature in the model.

- ⇨ Below image I have doubled the value of lasso alpha



And image clearly shows that GrLiveArea is most important feature in the model.

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?

=> Lasso's Mean Squared Error is slightly lower than Ridge's.

Lasso also has an advantage over Ridge because it aids in feature reduction (when the coefficient value of one of the features becomes 0).

According to Lasso, the zoning classification, living area square feet, overall quality and condition of the house, and other factors all determine the price. The house's foundation type, The number of automobiles that can fit in the garage, the total basement space in square feet, and the finished square feet area of the basement

Question 3

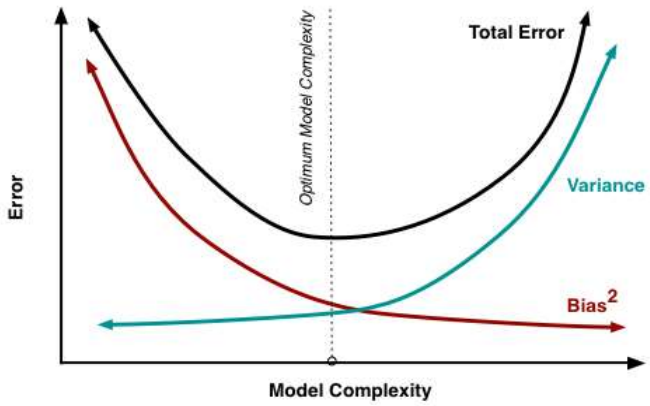
After building the model, you realised that the five most important predictor variables in the lasso model are not available in the incoming data. You will now have to create another model excluding the five most important predictor variables. Which are the five most important predictor variables now?

=>

1	GrLivArea	0.100763
2	MSZoning_RL	0.080436
3	OverallQual	0.072116
4	MSZoning_FV	0.053977
5	MSZoning_RM	0.049970

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?



=> The most important point in the model's making more generalizable is that we have to assure the variance and bias have to be in mid position.

So the image shows that when the bias increases, the model complexity will increase and the error term will go high.

And same with variance. If the model complexity gets high, then the model complexity will

increase, so we have to find a golden spot like shown in the image. The optimal model complexity line shows the model that is robust and easy.