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?

Solution:
By comparing the metrics table of double the alpha and the optimal alpha, we can see that R-squared value has gone down by 2%. The most important predictor variables after doubling the alpha are –
['GrLivArea', 'BsmtFullBath', 'KitchenAbvGr', 'GarageCars', 'MSSubClass_20', 'MSSubClass_50', 'MSSubClass_60', 'MSSubClass_70', 'MSSubClass_80', 'MSSubClass_90']

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?

Solution:
Compare two models, we have to choose by using accuracy (R2 score, as per regression) or Mean Squared Error or F-statistic/ANOVA test.
The model having least mean squared error is considered to be the good fit.
Similarly, higher the R-squared value, good model it is. Hence, “Lasso Regression" is a best model as it has least MSE and high R2score/R-squared value
among all regression models.

As Lasso regression that helps in feature elimination and the model is more robust, now. The main advantage of Lasso regression is it can produce many solutions to the same problem.

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?

Solution:
The top 5 most predictor variables are –
['MSZoning_FV',
'MSZoning_RL',
'Neighborhood_NoRidge',
'MSZoning_RH',
'Neighborhood_NridgHt']

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?

Solution:

The best model is said to be robust and generalizable is when outliers doesn't impact model. We have to make sure that the model is generalisable so that the train accuracy is always greater than test score. As the accuracy of the model depends on the outliers present in the data. So, it’s better to make a generalised analysis on outlier and scale accordingly. Few scaling techniques like robust scaler, standard scaler make the model more generalised to scale the outlier and thus, increases the accuracy on the prediction made by the model. At times Confidence Intervals within 3 Standard Deviation is useful to make model robust.