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 is 7.0 and for Lasso is 500.
- For Ridge regression, the coefficients values are almost getting half. For Lasso, the number of significant features got reduced to 21 from 28.
- 'GrLivArea', 'GarageCars_3', 'Fireplaces_other', 'FullBath_other' are top four predictors, however 'BsmtFinType1_GLQ' , replaces 'KitchenQual_other' at fifth position, when alpha is doubled for Lasso.

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?

- We will choose to apply lasso regression as it is reduced the betas of less –significant features to zero, hence helping in feature selection.

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?

- TotalBsmtSF, GarageArea, MasVnrArea, BsmtQual_other, ExterQual_other are most important predictor now, post dropping the top five predictor in first iteration.

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?

- By running cross-validation, where we will get an optimized Hyperparameter to get a robust and generalized model. A robust and generalized model will give low error, with less bias and variance.