# ASSIGNMENT

**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 1**

Optimal value for ridge is alpha=10 and for lasso alpha=100.
If we choose double the value of alpha, the model will be strongly regularized and model coefficients will be driven more closely to zero. Model will become less complex with low variance and high bias. Model will lead to underfitting.
Most important variable after changes are implemented are:
For ridge = GrLivArea
For lasso = GrLivArea

**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 2**

I will chose lasso regression as :

| | Ridge | Lasso |
|---|---|---|
| R2_train | 0.943 | 0.945 |
| R2_test | 0.881 | 0.887 |

The difference between R2_train and R2_test is less for lasso than ridge with almost equal scores for both test and train. Also, lasso provides feature selection, which makes it easier to interpret models

**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?

**Answer 3**

After creating new model, excluding five most important predictor variables, the five most important predictor variables now are:

- 1stFlrSF
- 2ndFlrSF
- RoofMatl_WdShngl
- PoolArea
- MasVnrArea

**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?

**Answer**

There are several ways to make sure that a model is robust and generalizable:

1. Splitting the data into training, validation, and test sets: This ensures that the model has not overfitted to the training data and is able to generalise to new data.
2. Use cross-validation: This helps to ensure that the model is robust and generalizable, as it has been trained and evaluated on different subsets of the data.
3. Regularize the model: This encourages the model to use only the most important features and reduces the complexity of the model.

In terms of the accuracy of the model, making sure that the model is robust and generalizable can have a positive impact on its accuracy. A model that is overfitted to the training data will have poor accuracy on new data, as it has not learned to generalize to new situations. On the other hand, a model that is robust and generalizable will have better accuracy on new data, as it has learned to handle a wide range of scenarios and conditions.

.