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?

**Ans:** Optimal value for ridge is 1.0 and lasso is 0.001. There is not much change in r2 score if we choose double the value of alpha for both ridge and lasso

The most important predictor variables after the change is implemented

* FullBath
* BsmtFinSF2
* ExterQual
* MSSubClass
* GarageQual

1. 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?

**Ans:** r2\_scores are almost same for both ridge and lasso but as lasso will penalize more on the dataset and can also help in feature elimination thus I will choose Lasso as final model.

1. 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?

**Ans:** Following are most important predictor variable, excluding the five most important predictor variables are:

* BedroomAbvGr
* BsmtUnfSF
* BsmtFinType2
* EnclosedPorch
* LotFrontage

1. 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?

**Ans:** A model needs to be made robust and generalizable so that they are not impacted by outliers in the training data. The model should also be generalisable so that the test accuracy is not lesser than the training score. The model should be accurate for datasets other than the ones which were used during training. Too much weightage should not give to the outliers so that the accuracy predicted by the model is high. Those outliers which it does not make sense to keep must be removed from the dataset. This would help increase the accuracy of the predictions made by the model. This would help standardize the predictions made by the mode. If the model is not robust, it cannot be trusted for predictive analysis.