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?**

* Ridge regression optimal alpha value is 0.01 and for Lasso regression alpha value is 0.0001.
* After double value of alpha there was no noticeable difference in accuracy Lasso model but in Ridge model accuracy decrease and it became 0.0035.
* Most important predictor for Lasso model are **'GrLivArea', 'YearBuilt', 'GarageCars', 'OverallQual\_9', 'BsmtFinSF1', 'OverallQual\_8', 'LotArea', 'OverallCond\_3', 'Fireplaces', 'OverallQual\_3'.**
* Most important predictor for Lasso model are **'OverallQual\_5', 'LotArea', 'YearBuilt', 'OverallQual\_3', 'OverallQual\_4', 'OverallQual\_6', 'OverallQual\_7', 'OverallQual\_8', 'OverallQual\_9', 'OverallQual\_10'.**

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?**

* By looking accuracy Lasso and Ridge both done well on Train and Test data.
* I would choose Lasso as my final model since it reduced 5 co-efficient to 0. So, it also worked as feature selection.
* If we change alpha value then there is no drastic change in model accuracy and predictor variable.

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?**

* Five most important predictors after remove top 5 predictors from previous model are **'1stFlrSF', 'TotalBsmtSF', 'OverallQual\_3', 'OverallQual\_4', 'KitchenAbvGr'.**

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?**

* A **model** is considered to be **robust** and **generalised** if its output dependent variable is consistently accurate even if one or more of the input independent variables or assumptions are drastically changed due to unforeseen circumstances.
* Here, our model perform well on train data as well as test data. So, we can say that our model is not overfitted. Train accuracy of data is high So, model is not underfitted.