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?

>> Optimal value of alpha of lasso regression : 50

   Optimal value of alpha of ridge regression : 4

   Optimal value of alpha for ridge regression on variables selected by lasso regession : 1

DOUBLING Alpha for LASSO :

```
With Alpha = 50, R2 for train : 0.9372405328256925 R2 for test : 0.92546641
23086983
Double alpha in lass i.e. 100.
New  R2 for train : 0.9314089232702133
R2 for test : 0.9231590633923132
```

DOUBLING Alpha for RIDGE :

```
With alpha = 4, R2 for train and test :  0.9371096095852764
0.9253982765709685
With double alpha = 8 , R2 for train and test : 0.9254099264105871
0.9182570406142743
```

```
We observed that after doubling the alpha, R2 score has reduced a littl
e bit.
```

Most important predictor and coefficient : 'CentralAir'

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 see that there was consistency in the result of both the models. But **Lasso has better values of R2 and is more robust.**
- Also, **Lasso helps us in variable selection** as well in which it make variable coefficient to 0. This makes **is relatively simpler.** So, it can be used very well to have the better view of the important factors while choosing the variables are significant in predicting the price of a house from the reduced list.
- In our case study, we also used **Lasso for feature selection** and then used the reduced list of variable with Ridge Regression.   Also, the optimal value of lambda shifted to 1 in Ridge regression in this case.

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?

>> Removed the top 5 predictor variable from Lasso Model
   X.drop(['OverallCond','GarageCars','BsmtFinSF2','OverallQual','CentralAir'], axis=1)

Top 5 predictors after removing them - `OverallCond', 'house_age', 'TotalBsmtSF', 'OverallQual', 'GrLivArea'`

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?

We have seen that simple models are more robust and generalizable, but just choosing simple is not enough. Model should perform almost equally well on train and test sets in terms of the accuracy. If this is seen we should always go with simpler model.

Complex models are more sensitive and learn the peculiarity of training set. Due to which, these **become unstable** for changing data sets and **outliers i**n the training data. There for the accuracy goes on a toss here. Hence simpler models are better and are unlikely to change wildly.
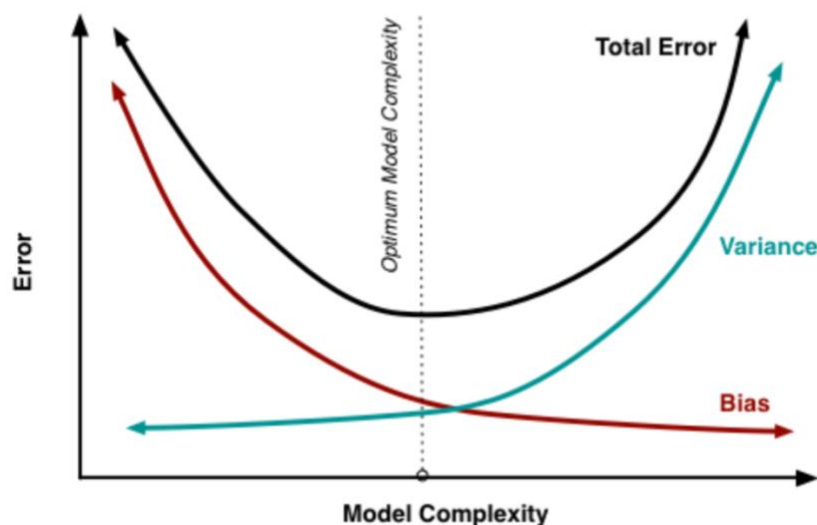
Also, there is also a trade-off between variance and bias to select optimal model. **The best model for a task is the one that balances both without compromising on accuracy.**