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 for lasso is: **{'alpha': 0.0001}**

*for lasso: best alpha = **{'alpha': 0.0001}***
R2 score (train) : 0.9163346393184356
R2 score (test) : 0.8717323298878759
RMSE (train) : 0.11327486770421738
RMSE (test) : 0.15311964904715145

Optimal value of alpha for ridge is: **{'alpha': 9.0}**

*for ridge: best alpha = **{'alpha': 9.0}***
R2 score (train) : 0.9162900520206345
R2 score (test) : 0.8717043408883183
RMSE (train) : 0.11330504714465928
RMSE (test) : 0.15313635408287804

Double alpha for lasso & ridge to **{0.0002 & 18.0}**

*Model Evaluation : Lasso Regression, **alpha=0.0002***
R2 score (train) : 0.9163
R2 score (test) : 0.8722
RMSE (train) : 0.1133
RMSE (test) : 0.1528

*Model Evaluation : Ridge Regression, **alpha=18.0***
R2 score (train) : 0.9161
R2 score (test) : 0.872
RMSE (train) : 0.1134
RMSE (test) : 0.153

There is not significant change in R2 score and RMSE score for **Ridge regression** when value of alpha is doubled.

Similarly, there is not significant change in R2 score and RMSE score for Lasso regression when value of alpha is doubled.

Most important predictor variables after the change are implemented

a) 1stFlrSF
b) 2ndFlrSF
c) OverallQual
d) OverallCond
e) SaleCondition_Partial

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?

I will choose Lasso as its giving feature selection option also. It has removed unwanted features from model without affecting the model accuracy. Which makes are model generalized and simple and accurate.

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?

If the above occurs, we can select the five most important predictors.

1	*BsmtFinSF*
2	*LotArea*
3	*BsmtUnfSF*
4	*GarageArea*
5	*KitchenQual*

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?

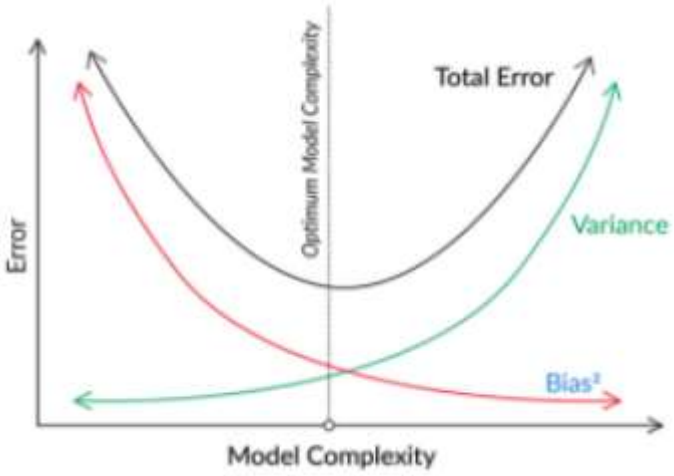
To make model robust and generalizable 3 features required:

1. **Model accuracy** should be > 70-75%: we got the accuracy of 91% (Train) and 87% (Test) which is correct.
2. **P-value** of all the features is < 0.05
3. **VIF** of all the features are < 5

Model should be generalized 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.

Bias-variance tradeoff - If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it's going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.

- *If the model is too complex, it will have low bias and high variance.* ***Overfitted***
- *If the model is too simple, it will have high bias and low variance* ***Underfit***
- *If we take the point in the bias variance trade off graph, where both intersect each other, that point will give perfect balance between bias-variance. It will ensure that model does not overfit while still having good variance*


Optimum Model Complexity
Total Error
Variance
Bias²
Error
Model Complexity