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:

Optimum value for alpha for Ridge is 0.01 and Optimum value for alpha for Lass is 0.001, which are as below:

```
#### Ridge Regression:
    - Train Score : 0.897965123607899
    - Test Score : 0.8738708902765001
    - Optimum value of alpha : {'alpha': 0.01}

#### Lasso Regression:
    - Train Score : 0.9444867153088153
    - Test Score : 0.879317310572314
    - Optimum value of alpha : {'alpha': 0.001}
```

Top 2 feature from Ridge Regression:

- OverallCond
- BsmtFullBath

Top 2 feature from Lasso Regression:

- Exterior1st_AsphShn
- RoofMatl_Tar

If we double the of alpha, the value of coefficients would decrease

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:

The optimal value of alpha for ridge regression is 10 and for Lasso regression is 0.00006. With this value of alpha, the test score are as below:

```
#### Ridge Regression:
    - Train Score : 0.897965123607899
    - Test Score : 0.8738708902765001
    - Optimum value of alpha : {'alpha': 0.01}

#### Lasso Regression:
    - Train Score : 0.9444867153088153
    - Test Score : 0.879317310572314
    - Optimum value of alpha : {'alpha': 0.001}
```

Though I see almost similar score among dataset with Ridge Regression and Lasso Regression, Though I prefer to use Lasso Regression, because there are some features with absolute zero coefficient.

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:

The important predictor variables after removing the earlier important variables

> Ridge Regression (alpha: 10)

```
('MasVnrType_Stone', 0.132),
('Foundation_CBlock', 0.143),
('Foundation_Others', 0.146),
('Foundation_PConc', 0.146),
('GarageType_BuiltIn', 0.16),
```

> Lasso Regression (alpha: 0.00006)

```
('MasVnrType_Stone', 0.139),
('Foundation_CBlock', 0.147),
('Foundation_Others', 0.155),
('Foundation_PConc', 0.16),
('GarageType_BuiltIn', 0.181),
```

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:

The following operations are done on the model that makes it more robust and generalisable.

1. The skewed columns are removed and processed to make normal

2. Null or empty values are replacing with median for columns having outliers and mean for other continuous variables.

3. The columns having less correlation to target variable have been removed

4. The best model out of 5-fold cross validated model using Ridge and Lasso regularization has been chosen as final model. This shows the best model out of the available training set.

5. On plotting mean test and train scores with alpha we could observe similar curve for both train and test data. As a result of the above points, I could observe a decent score for Ridge and Lasso Regression which are as follows