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:

I have got the optimum value of alpha for:

1. Ridge Regression: 7
2. Lasso Regression: 0.001

When the value of alpha is doubled for both Ridge and Lasso Regression i.e., now the value for alpha for:

1. Ridge Regression: 14
2. Lasso Regression: 0.002

Here, I've observed that the model performance remained the same in both the cases.

For Ridge Regression:

Alpha = 7:

```
RMSE Train 0.10981730795053707
R2 Score Train 0.9203536737103156
RMSE Test 0.13366186155982007
R2 Score Test 0.8877697530412173
```

Alpha = 14:

```
RMSE Train when alpha is doubled 0.11255110375379085
R2 Score Train when alpha is doubled 0.9163388791349765
RMSE Test when alpha is doubled 0.13433726515816904
R2 Score Test when alpha is doubled 0.8866326713763528
```

For Lasso Regression:

Alpha = 0.001:

```
RMSE Train 0.11814050735346564
R2 Score Train 0.907823157933937
RMSE Test 0.13433726515816904
R2 Score Test 0.8866326713763528
```

Alpha = 0.002:

```
RMSE Train when alpha is doubled 0.1249792530902665
R2 Score Train when alpha is doubled 0.896842689233902
RMSE Test when alpha is doubled 0.13433726515816904
R2 Score Test when alpha is doubled 0.8866326713763528
```

As we can observe both Ridge & Lasso model has similar RMSE Train & Test value and R2 Score Train & Test value for optimal alpha and when alpha is doubled. Which indicates that performance is almost similar for both models.

After the changes is implemented the most important predictor variables are:

1. **Ridge Regression:**

	Features	Coeff_val
19	Neighborhood_Crawfor	0.085638
18	OverallQual	0.075176
17	Exterior1st_BrkFace	0.072708
16	Total_Bathroom_sq_feet	0.060493
15	Neighborhood_StoneBr	0.059963
0	Neighborhood_Gilbert	0.056849
1	KitchenQual_TA	0.053332
2	BsmtQual_Fa	0.052927
14	Neighborhood_NoRidge	0.050000
13	TotRmsAbvGrd	0.048013

2. **Lasso Regression:**

	Features	Coeff_val
19	OverallQual	0.089765
18	Neighborhood_Crawfor	0.075012
17	Exterior1st_BrkFace	0.071534
16	Total_Bathroom_sq_feet	0.058764
15	TotRmsAbvGrd	0.046335
14	Foundation_PConc	0.045973
13	Fireplaces	0.045860
0	BsmtFinType1_Unf	0.036153
12	OverallCond	0.034608
1	KitchenQual_TA	0.033989

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:

Both the models gave very similar results. Ridge Regression model was slightly better than Lasso Regression model by 1%. But still, I will prefer Lasso Regression model because as it helps in some feature elimination.

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:

After building model excluding 5 most important predictor variables, the 5 most important predictor variables now are:

	Features	Coeff_val
0	BsmtFinType2_No	0.130508
1	Neighborhood_MeadowV	0.124776
2	BsmtQual_Fa	0.091905
3	KitchenQual_TA	0.090065
19	Exterior2nd_BrkFace	0.087162

Question-3:

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:

As per, Occam's Razor – given two models that show similar 'performance' in the finite training or test data, we should pick the one that makes fewer on the test data due to following reasons:

1. Simpler models are usually more 'generic' and are more widely applicable.
2. Simpler models require fewer training samples for effective training than the more complex ones and hence are easier to train.
3. Simpler models are more robust.
    a. Complex models tend to change wildly with change in the training data set.
    b. Simple models have low variance, high bias and complex models have low bias, high variance.

c. Simpler models make more errors in the training set. Complex models lead to overfitting, they work very well for the training samples, fail miserably when applied to other test samples.

Therefore, to make the model more robust and generalizable, make the model simple but not simpler which will not be of any use.

Regularization can be used to make model simpler. Regularization helps to strike the delicate balance between keeping the model simple and not making it too naïve to be of any use. For regression, regularization involves adding a regularization term to the cost that adds up to the absolute values or the squares of the parameters of the model.

Also, making a model simple lead to Bias-Variance Trade-Off:

- Complex model will need to change for every little change in the dataset and hence is very unstable and extremely sensitive to any changes in the training data.
- A simpler model that abstracts out some pattern followed by the data points given is unlikely to change wildly even if more points are added or removed.

Bias quantifies how accurate is the model likely to be on test data. A complex model can do an accurate job prediction provided there is enough training data. Models that are too naive, for e.g., one that gives same answer to all test inputs and makes no discrimination whatsoever has a very large bias as it's expected error across all test inputs are very high.

Variance refers to the degree of changes in the model itself w.r.t changes in the training data. Thus, accuracy of the model can be maintained by keeping the balance between Bias and Variance as it minimizes the total error.