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?

The optimal value of alpha for ridge & lasso regression are 9 & 0.0001.
While running Ridge regression with the previous ridge (alpha=9.0)

```
R2 score (train) :  0.9162900520206345
R2 score (test)  :  0.8717043408883183
```

While running Lasso regression with the previous lasso (alpha=0.0001)

```
R2 score (train) :  0.9163346393184356
R2 score (test)  :  0.8717323298878759
```

As per the requirement when we double the value of alpha for both ridge & lasso, the new values becomes 18 & 0.0002.
While running Ridge regression with the new ridge (alpha=18.0)

```
R2 score (train) :  0.9161
R2 score (test)  :  0.872
```

While running Lasso regression with the new lasso (alpha=0.0002)

```
R2 score (train) :  0.9163
R2 score (test)  :  0.8722
```

From the above values,

In Ridge Regression metrics:
R2 score of train set decreased by 0.9163 to 0.9161
R2 score of test set remained same at 0.872

In Lasso metrics:
R2 score of train set remained same at 0.9163
R2 score of test set increased from 0.871 to 0.872

The most important predictor variables after we implement the changes are

- BsmtFinSF1
- LotArea
- BsmtUnfSF
- GarageArea
- KitchenQual

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?

It's actually dependent on the Use case.
If we have many variables & our objective is to do a feature selection, then we will go for Lasso Regression, while If we don't need to have large coefficients & here we our objective is reduction of coefficient magnitude, then we go for Ridge Regression.

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?

1. LotArea
2. BsmtFinSF1
3. SaleCondition_Normal
4. MSZoning_RL
5. Neighborhood_Somerst

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 can make sure the model is robust & generalizable by making the model simple, although the accuracy will be reduced. The simpler the model, the more the bias, but less variance. A robust & generalisable model will perform equally well on training & test data.

The model should be more complex in order to be robust & generalizable.
We need to find some balance between Model accuracy & complexity, which can be achieved by regularization techniques like Ridge Regression & Lasso.