**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?

Ans: The Optimal value of alpha for Ridge regression=1 and the optimal value of alpha for Lasso regression=0.001
When the value of alpha is doubled for Ridge regression there isn't much change in neither the accuracy nor the important features. There is a slight change of the value of coefficients though and only a total of 4 features in the top 10 of both negative and positive features had changed. The most important predictor variable remains unchanged in Ridge regression.
When the value of alpha is doubled for Lasso regression there is a drop in the accuracy by 2 percent and the number of significant variables also decreases from 72 to 53. However there isn't much change in the important predictor variables although their coefficients change slightly. The most important predictor variable still remains the same in Lasso regression as well.

**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?

Ans: The r2_score for Ridge regression model is 0.905, whereas for Lasso regression model is 0.91. It can be seen that Lasso model returns a better accuracy on the test data and also it reduces the number of significant features to 72 which makes it easier to interpret and also it is simpler and robust model.

**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?

Ans: If the 5 most important variables are absent then the new variables along with their coefficients are:

```
GarageArea             0.197013
BsmtFinSF1             0.102318
Neighborhood_Crawfor   0.090041
BsmtExposure_Gd        0.085325
Fireplaces             0.077566
```

**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?

Ans: To make a model robust and generalisable we need to have a proper trade-off between bias and variance. A complex model with too many features is often overfitted and it performs very bad on testing data while a simple model with less features underfits, giving less accuracy on training as well as testing. This trade-off can be achieved by applying model selection methods like GridSearchCV and Ridge or Lasso Regression.

By taking care of overfitting we reduce the accuracy on training data, as the model now focuses more on learning the underlying trends rather than memorizing the data, but however the testing accuracy increases.