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:
The optimal value of lambda for ridge regression is 0.2 and lasso regression is 0.005.
If we double the value of alpha for both ridge and lasso it might lead to underfitting the model and the coefficients for the variables will change.
After the change the most important predictor variables are –
'OverallQual', 'GarageArea', 'BsmtQual_TA', 'KitchenQual_TA'

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 lambda for ridge regression is 0.2 and lasso regression is 0.005.
I will choose to perform Lasso Regression with alpha 0.005 as Ridge regression still overfits the model and we see a huge r2 score drop when the model is tested with test data, whereas, the Lasso Regression with alpha 0.005 doesn't overfit the data and performs well when tested with test data as well, with just 3.2% of r2 score drop.

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 five most important predictor variable now are –
'1stFlrSF', '2ndFlrSF', 'TotalBsmtSF', 'KitchenQual_Fa', 'BsmtQual_TA'

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:
We can make sure that a model is robust and generalisable by finding an optimal value of alpha which makes the model perform well on train dataset as well as on test data with an r2 score change of not more than 5%.
Here we see an r2 score of 90.93% on train dataset and 60.72% on test dataset when we perform Linear Regression, which implies that the model overfits.
When we apply an optimal value of alpha i.e., 0.005 and perform Lasso Regression, we see an r2 score of 72.17% on train dataset and 68.98% on test dataset. Which implies the model has not overfit and is robust and generalisable.