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 alpha for Ridge is 20 and Lasso is 0.001

If we double the value of alpha for Ridge, the RSS will quite increase and the R2 score will be slightly affected along with lowering the coefficients.
If we double the value of alpha for Lasso, it will make more coefficients of features towards 0 , i.e more feature variables having coefficients equal 0.

After the change is implemented,
For Ridge the top 3 variables are = 'OverallQual' , ' Neighborhood_Crawfor' and 'GrLivArea'
For Lasso the top 3 variables are = 'OverallQual' , 'GrLivArea' and 'GarageCars'

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 application/usage of the model, depends on the business goals. Since the company is expecting which predictor variables will be responsible, the Lasso regression is a great choice. Lasso regression not only gave good scores but also performed feature engineering and provided significant variables, as result making coefficients of less significant variables to 0.

Question 3
After building the model, you realized 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 excluding the top 5 predictor variables in Lasso model,
The below are predictor variables now:
'Exterior1st_BrkFace', 'GarageCars', 'Condition1_Norm', 'Functional_Typ', 'CentralAir_Y'

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 model is robust and generalisable using techniques such as:

- Performing Data Pre-Processing and generating high quality data
- Using Cross validation
- Using Feature engineering
- Using Regularization
- Using Model building and comparison

Such techniques implicate higher accuracy, lower bias, consistent performance w.r.t generalization and adaptive to unseen data since we understand the data and modify/process/compare them to get utmost very good results and make the model reliable and robust.