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 :- optimal value of alpha in **lasso regression is 0.001** and optimal value of alpha in **ridge regression is 0.5,** If we double the alpha the R2 scores will have more difference meaning more chances of overfitting, Initially for Lasso the R2 score was 76 and 73 for train and test respectively after doubling the alpha value it changed to 69 and 67 for train and test respectively. And for ridge initial was 83 and 77 after doubling alpha its 82 and 77, the predictor variables stay as it is, The most important predictor variables are OverallCond_Below Average, OverallQual_Very Good, OverallQual_Very Excellent, BedroomAbvGr, MSZoning_FV.

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 :- We can use **Lasso if there are many features** because **Lasso helps in feature elimination**, in our case as there are more number of features, I will choose to apply lasso regression because in this model the difference between the R2 score of train and test is 3 which is less compared to ridge regression, and it means that lasso regression model is able to generalize the learning on test or unseen data in a better way then ridge regression, where the difference between the R2 score is more than 5.

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 :- OverallCond_Fair,BedroomAbvGr,MSZoning_FV,2ndFlrSF,Condition2_PosN

Question 4

How can you make sure that a model is robust and generalizable ? What are the implications of the same for the accuracy of the model and why?

Answer :- We can make the model **robust** when the **variation** on the unseen data **does not affect the bot performance**, we can make the model **generalizable by solving the issue like overfitting** using the techniques like ridge and lasso regression. In terms of **accuracy** , more complex model will be more accurate but can't perform well even if there is slight variation in data , so in order to build an accurate model we need to **balance between complex and variation .**