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?

- According to my calculation considering alpha = 100, we see r-square results are good. Hence optimal value is 100 here.
- If we double the value of alpha for ridge then more coefficients will tend towards zero to make the model relatively simpler.
- If we double the value of alpha for Lasso then more coefficients will become zero to make the model relatively simpler.
- 'OverallQual' will be most important variable.

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?

- I will use Lasso over Ridge. Lasso will be relatively simple in nature which will have relatively less variance compromising marginally on bias part.

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?

- Next five important predictor variables would be:
  - Airconditioning
  - Hotwaterheating
  - Bathrooms
  - Prefarea
  - guestroom

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?

- Apply Lasso regression I've regularised the model. Hence by trading off on some bias my would have relatively low variance which will make the model more generic in nature.
- From a accuracy stand point it's bias will marginally increase but the variance will drop drastically.