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 values of alpha for Ridge and Lasso regression are 8 and 100, respectively. If we double these values to 16 for Ridge and 200 for Lasso, the models would become more conservative, shrinking coefficients towards zero even more. Despite this, the most important predictor variables would likely remain the same, albeit with smaller coefficients.

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:- Choosing between Ridge and Lasso regression depends on the problem's needs. Ridge is useful for handling multicollinearity and including all variables in the model, while Lasso is better for feature selection and producing a simpler, more interpretable 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:- Excluding the five most important predictor variables from the Lasso model would shift importance to the next most influential variables, resulting in them having larger coefficients in the updated model.

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 ensure a model is robust and generalizable, use techniques like

1.cross-validation

2. regularization

3. feature engineering

4. outlier detection

5. validation on external data.

Although this might slightly reduce accuracy on the training data, it improves performance on unseen data, which is more valuable for real-world use.