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?

A1: The optimal alpha value chosen for ridge regression is 20. The optimal alpha value for lasso regression is 400. When the alpha values are doubled for both ridge and lasso models, the most important predictor variables remain the same and their rankings do not change. Although increasing alpha leads to more regularization, the effect is not strong enough to alter which predictors are most important in the models

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 would take using Ridge model, as I plot the metrics, Ridge has the highest R^2 value of 0.7718, indicating it explains the most variance in the target variable. Also it has the lowest MSE and RSS values.

Compared to lasso indicate ridge is more regularized model. Since prediction accuracy is highest, I would choose Ridge as the best approach for this problem.

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?

The new model of top 5 predictors variables are 'Foundation_CBlock', 'Condition2_Feedr', 'SaleType_WD', 'Condition2_RRNn' and 'Condition1_Norm' after excluding the original top 5 variables that are unavailable in the new incoming data.

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?

Tried to simpler models tend to be more robust and generalizable than overly complex or overfit models. Using regularization of biases the model like lasso and ridge. Also in feature selection, I have removes redundant or irrelevant features. This simplifies models and avoids overfitting of test data.