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 for ridge and lasso regression are:
Value of alpha for **Ridge** Regression : 10
Value of alpha for **Lasso** Regression : 0.001

If we choose to double the value of alpha for Ridge, it will lower the coefficients. If we choose to double the value of alpha for Lasso, it will make more feature's coefficient to zero.

The most important predictor variable after the change is implemented are those which are significant.

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:
Optimal Value of alpha for ridge and lasso regression are:
Value of alpha for **Ridge** Regression : 10
Value of alpha for **Lasso** Regression : 0.001

The scores for both models are good. We can go with Ridge Regression model as the R2 value is better for the Ridge and also the RSS value on the test is less for Ridge Regression than Lasso Regression.

Metric	Linear Regression	Ridge Regression	Lasso Regression
R2 Score(Train)	0.949801	0.915192	0.900741
R2 Score (Test)	0.862228	0.889950	0.886565
RSS (Train)	8.056537	13.610957	15.930252
RSS (Test)	9.929002	7.931112	8.175122
RMSE (Train)	0.088830	0.115460	0.124910
RMSE (Test)	0.150562	0.134564	0.136619

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?

Answer:
After removing the top5 significant feature, we have below feature for Ridge and Lasso:

Top5 most significant features in Ridge are:

- ('SaleType_Con', 0.054),
- ('SaleType_ConLD', 0.057),
- ('SaleType_ConLI', 0.058),
- ('SaleType_ConLw', 0.058),
- ('SaleType_New', 0.062)

Top5 most significant features in Lasso are:

- ((('SaleType_Con', 0.052),
- ('SaleType_ConLD', 0.056),
- ('SaleType_ConLI', 0.057),
- ('SaleType_ConLw', 0.059),
- ('SaleType_New', 0.06)

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:
In general Robustness of a model is the property that characterizes how effective our model is while being tested on the new independent dataset. So, Robust model is the one which has the testing error closer to the training error. If we see in our model we have low tried to minimize the RSS value to make the model Robust. The model is considered generalisable if it does not overfit the training data, and work well with the new independent data set. More accurate model might lead in overfitting of the model, so the robust and generalisable model will perform well on the test data maintaining the almost same accuracy for the model even when tested on the new independent dataset without changing much from training dataset.