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 value for ridge and lasso regression as follows :

- Ridge(alpha=20)
- Lasso{'alpha': 0.001}

Metrics for Rigde :

```
R-Squared (Train) = 0.91
R-Squared (Test) = 0.81
RSS (Train) = 14.56
RSS (Test) = 13.14
MSE (Train) = 0.01
MSE (Test) = 0.03
RMSE (Train) = 0.12
RMSE (Test) = 0.17
```

Metrics for Lasso :

```
R-Squared (Train) = 0.91
R-Squared (Test) = 0.80
RSS (Train) = 15.11
RSS (Test) = 13.56
MSE (Train) = 0.01
MSE (Test) = 0.03
RMSE (Train) = 0.12
RMSE (Test) = 0.18
```

Coefficients for lasso model :

```
Column
constant              11.975
GarageFinish_Unf       0.106
GarageFinish_RFn       0.096
GarageFinish_NA        0.094
GarageType_NA          0.089
GarageType_Detchd      0.079
```

Coefficients for Ridge model :

```
Column
constant              12.02607
GarageFinish_Unf       0.08494
GarageFinish_RFn       0.08092
GarageFinish_NA        0.07081
GarageType_NA          0.05634
GarageType_Detchd      0.05214
```

If we double the values of alpha, it will produce the following results.
Ridge :

```
R-Squared (Train) = 0.91
R-Squared (Test) = 0.81
RSS (Train) = 15.25
RSS (Test) = 13.23
MSE (Train) = 0.01
MSE (Test) = 0.03
RMSE (Train) = 0.12
RMSE (Test) = 0.17
```

Lasso:

```
R-Squared (Train) = 0.90
R-Squared (Test) = 0.80
RSS (Train) = 16.44
RSS (Test) = 13.86
MSE (Train) = 0.02
MSE (Test) = 0.03
RMSE (Train) = 0.13
RMSE (Test) = 0.18
```

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 : After a careful consideration of the regression, we will have to choose based on the following parameters.

- If the coefficients are not very important we can consider using Ridge regression.
- If we have the RFE included we can go with Lasso.

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?

Top 5 predictors for Lasso :

```
Column
constant              11.975
GarageFinish_Unf       0.106
GarageFinish_RFn       0.096
GarageFinish_NA        0.094
GarageType_NA          0.089
GarageType_Detchd      0.079
```

Top 5 predictors for Ridge :

```
Column
constant              12.02607
GarageFinish_Unf       0.08494
GarageFinish_RFn       0.08092
GarageFinish_NA        0.07081
GarageType_NA          0.05634
GarageType_Detchd      0.05214
```

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 make sure model is robust and generalisable, we have to make sure the data is diverse and avoids overfitting as well. Make sure the scaling is done using the right methods and analyse the training dataset based on the requirement.

Implications on accuracy :

- Training data accuracy doesn't guarantee the accuracy on the test data or the real world data. Overfitting will be an issue when it comes to training the data.
- If the data is underfitting again it will take toll on the accuracy of test data.
- Regularization helps in getting the right balance between generalizations and robustness.