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 Alpha value for Ridge and Lasso Regressions:

Regression Type	**Feature Set**	**Optimal Alpha value**
Ridge Regression	All features (260)	657.93
Ridge Regression	Top features (10)	84.975
Lasso Regression	All features (260)	10.00
Lasso Regression	Top features (10)	10.00

Doubled alpha values:

Regression Type	**Feature Set**	**Doubled Alpha value**
Ridge Regression	All features (260)	1315.86
Ridge Regression	Top features (10)	169.96
Lasso Regression	All features (260)	20
Lasso Regression	Top features (10)	20

Changes in the Ridge Regression model when the alpha value is doubled:
1. Test R-square decreases
2. Coefficients shrink more
3. Overfitting problem is reduced

Changes in the Lasso Regression model when the alpha value is doubled:
1. More features are eliminated
2. Test R-square decreases
3. Model becomes simpler since feature selection is stricter by eliminating weak features
4. RMSE increases

Most important predictor variables after doubling alpha:

Top feature aspect	**Ridge with doubled alpha**	**Lasso with doubled alpha**
1st top feature	OverallQual	poolQC_Name
2nd top feature	GrLivArea	GarageQual_TA
3rd top feature	TotalBsmtSF	garageCond_TA

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:
Ridge Regression (Alpha = 657.93) can be chosen for the following reasons:
1. Housing dataset modelling requires all the features
2. Ridge RMSE is 30,600 which shows more precise prediction than Lasso's RMSE 31,800.
3. Housing features are highly correlated and Ridge Regression handles by shrinking the related coefficients

Additionally, can also go with the Ridge Regression model with top 10 features, for the below reasons:

1. Uses the most correlated features (top 10)
2. Alpha value is 84.98
3. Model becomes simpler
4. But, accuracy might be sacrificed

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:
Based on my analysis, the following are the five most important predictor variables in Lasso model based on correlation:

1. OverallQual (coefficient: 19,612.65)
2. GarageCars (coefficient: 11,011.19)
3. FullBath (coefficient: -6,923.09)
4. TotRmsAbvGrd (coefficient: 408.45)
5. YearRemodAdd (coefficient: 314.75)

The most important predictor variables after removing the above variables are as follows:

1. MasVnrArea
2. Neighborhood_StoneBr
3. Neighborhood_NridgHt
4. Fireplaces
5. GarageArea

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:

1. The model can be robust and general if we can solve the overfitting problem by considering the Regularization techniques such as Ridge and Lasso Regressions.
2. When doing generalisation, the accuracy might decrease but if it fits well with both train and test data then such a model becomes more robust.
3. In the assignment, as per my analysis, the Ridge model with all features achieves both generalisation and good accuracy, thus preventing overfitting problem.