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?**

=> optimal value of lambda for ridge : 10
=> optimal value of lambda for lasso : 0.001

When alpha is double for ridge regression: So here our alpha for ridge regression is 20
Now we will calculate the R-Squared, RSS, MSE and RMSE for train and test model

R-Squared value for (Train) = 0.93
R-Squared value (Test) = 0.93
RSS value (Train) = 9.37
RSS value (Test) = 2.82
MSE value (Train) = 0.01
MSE value (Test) = 0.01
RMSE value(Train) = 0.09
RMSE value (Test) = 0.10

When alpha is double for laso regression: So here our alpha for lasso regression is 0.002

R-Squared value for (Train) = 0.91
R-Squared value for (Test) = 0.91
RSS value for (Train) = 13.49
RSS value for (Test) = 3.45
MSE value for (Train) = 0.01
MSE value for (Test) = 0.01
RMSE value for (Train) = 0.11
RMSE value for (Test) = 0.11

Changes in Ridge regression stats when alpha is doubled then R-Squared for train model decreased from 0.94 to 0.93
R-Squared for test model did not changed it’s as it is i.e. 0.93

Changes in Lasso regression stats when alpha is doubled then R-Squared for train model decreased from 0.92 to 0.91
R-Squared for test model decreased from 0.93 to 0.91

Now I have calculated top 10 features when alpha is double for Ridge and Lasso regression

Top features are : GrLivArea, OverallQual_8, OverallQual_9, Neighborhood_Crawfor, Functional_Typ, Exterior1st_BrkFace OverallCond_9, TotalBsmtSF, CentralAir_Y,OverallCond_7, YearRemodAdd, Condition1_Norm

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?**

Ridge regression: When you have lots of data and many predictors that could matter, and they might be related, choosing Ridge regression is a smart move. It's like finding a good balance (lambda around 10) to avoid the model from getting too excited and making predictions that don't make sense.

Lasso regression: On the flip side, if you think only a few predictors really matter, and you want to keep things simple and clear, Lasso regression is the way to go. It helps to set less useful predictors to zero (lambda around 0.001), like tidying up and focusing only on the really important stuff.
To wrap it up, deciding between Ridge and Lasso depends on how many predictors are important and how you want your model to be. Lasso is like a strict cleaner, throwing out what's not needed, great when only a few things matter a lot. Ridge is more inclusive, keeping all the potential factors in mind, useful when many things are important in a similar way. The choice depends on what fits best with your dataset and what you want to achieve with your analysis.

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 Lasso predictors are : GrLivArea', 'OverallQual_9', 'OverallQual_8', 'Neighborhood_Crawfor', 'Exterior1st_BrkFace'

We are dropping above 5 variable and creating another model after dropping above 5 variables:
Now we get optimum value of lambda for lasso is 0.001
we will calculate the R-Squared, RSS, MSE and RMSE for train and test model 1

R-Squared value for (Train) = 0.91
R-Squared value for (Test) = 0.92
RSS value for(Train) = 12.75
RSS value for (Test) = 3.02
MSE value for (Train) = 0.01
MSE value for (Test) = 0.01
RMSE value for (Train) = 0.10
RMSE value for (Test) = 0.10

Now we will see coefficient values after applying regularization 1
After excluding top 5 feature, we are going to get top 5 feature for new Lasso model These are new top 5 predictors after dropping 5 top features 2ndFlrSF, Functional_Typ, 1stFlrSF, MSSubClass_70, Neighborhood_Somerst

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?

To ensure a model is robust and generalizable, we follow a few essential practices:
Use Sufficient Data: Having enough data helps the model learn patterns effectively. More data allows the model to capture a broader range of scenarios, enhancing its generalization.

Split Data for Training and Testing: We divide our dataset into two parts: one for training the model and another for testing its performance. This helps assess how well the model generalizes to unseen data.

Cross-Validation: In addition to a simple train-test split, we use techniques like cross-validation. This involves splitting the data into multiple subsets, training the model on different combinations, and evaluating its performance. It ensures a more robust assessment of the model's generalization.

Feature Selection: Choose relevant features that genuinely affect the outcome. Too many irrelevant features can hinder generalization. Techniques like Lasso and Ridge regression aid in feature selection by emphasizing important features and reducing the impact of less relevant ones.

Regularization: Regularization methods like Ridge and Lasso help control the model's complexity. They add penalties to the model parameters, preventing overfitting and promoting a more generalized model.

Avoid Overfitting: Overfitting occurs when the model learns the training data too well but struggles with new data. By using regularization and suitable model evaluation techniques, we can mitigate overfitting and create a more robust, generalizable model.

Optimize Hyperparameters: Tuning hyperparameters, such as the alpha in Ridge and Lasso regression, ensures the model is well-adjusted. The right hyperparameters lead to a more accurate and generalizable model.