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?

So for this case study i have build two Models the first one is using Recursive Feature Elimination (RFE) and then fetched to Ridge and Lasso Regression. The other one takes the data into Ridge and Lasso Regression directly and here are my findings:-

1st Model Using RFE -

The Optimal Value of Lambda for Ridge comes out to be = 0.1, the metrics it provided are as follows :-

R-squared on Training Set : 0.889548961870626
R-squared on Test Set : 0.8643301947971636
RSS on Training Set : 101.72540611715337
RSS on Test Set : 56.351218563353825
RMSE on Training Set : 0.1104510381293739
RMSE on Test Set : 0.14266131281861727

Doubling the Lambda Value of alpha for Ridge gave us these results :-

R-squared on Training Set : 0.889526447171181
R-squared on Test Set : 0.8643539155169226
RSS on Training Set : 101.74614215534228
RSS on Test Set : 56.34136602865298
RMSE on Training Set : 0.11047355282881899
RMSE on Test Set : 0.14263636969279234

The Optimal Value of Lambda for Lasso comes out to be = 0.0001, the metrics it provided are as follows :-

R-squared on Training Set : 0.8895343327184477
R-squared on Test Set : 0.8643590443363462
RSS on Training Set : 101.73887956630969
RSS on Test Set : 56.339235744586574
RMSE on Training Set : 0.11046566728155233
RMSE on Test Set : 0.1426309765685736

Doubling the Lambda Value of alpha for Lasso gave us these results :-

R-squared on Training Set : 0.8894631682851103
R-squared on Test Set : 0.8643976347502493
RSS on Training Set : 101.80442200941343
RSS on Test Set : 56.3232069985804
RMSE on Training Set : 0.11053683171488972
RMSE on Test Set : 0.14259039746476052

Top 3 most important predictor variables after doubling the alpha value are :-

For Ridge and Lasso both are same -

1) OverallQual_Excellent
2) OverallCond_Excellent
3) 2ndFlrSF

2nd Model Directly Putting the data into Ridge and Lasso -

The Optimal Value of Lambda for Ridge comes out to be = 6.0, the metrics it provided are as follows :-

R-squared on Training Set : 0.9291407978404908
R-squared on Test Set : 0.8861973785014026
RSS on Training Set : 65.26132518890805
RSS on Test Set : 47.26856051397958
RMSE on Training Set : 0.07085920215950928
RMSE on Test Set : 0.11966724180754323

Doubling the Lambda Value of alpha for Ridge gave us these results :-

R-squared on Training Set : 0.9253895742875182
R-squared on Test Set : 0.8833328272891623
RSS on Training Set : 68.71620208119575
RSS on Test Set : 48.4583680117167
RMSE on Training Set : 0.0746104257124818
RMSE on Test Set : 0.12267941268789039

The Optimal Value of Lambda for Lasso comes out to be = 0.001, the metrics it provided are as follows :-

R-squared on Training Set : 0.9276124350536974
R-squared on Test Set : 0.8888690336025521
RSS on Training Set : 66.66894731554467
RSS on Test Set : 46.15887350362608
RMSE on Training Set : 0.07238756494630258
RMSE on Test Set : 0.11685790760411667

Doubling the Lambda Value of alpha for Lasso gave us these results :-

R-squared on Training Set : 0.9210693839052689
R-squared on Test Set : 0.8849760361973702
RSS on Training Set : 72.69509742324732
RSS on Test Set : 47.77585192648145
RMSE on Training Set : 0.07893061609473108
RMSE on Test Set : 0.12095152386451

Top 3 most important predictor variables after doubling the alpha value are :-

For Ridge -

1) OverallQual_Excellent
2) Neighborhood_NridgHt
3) Neighborhood_Crawfor

and Lasso -

1) OverallQual_Excellent
2) Neighborhood_Crawfor
3) SaleType_New

So as we can see the optimal value lambda which is a hyperparameter helps the model to maintain the tradoff between bias and variance. Higher the value of lambda the coefficients will be penalised more.

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?

After building both the model and finding the optimal value of lambda for both Ridge and Lasso regression for my first model which is the model in which i have done used RFE for feature selection and then put it thorugh Ridge and Lasso i can go with Ridge Regression as i have already made reduced the number of feature using RFE and then manually removing High VIF and less significant variables thus we do not need extra feature selection from lasso and as we can see almost simmilar results for both. But for my second model where i have fed all the relevent indipendent variables in the Ridge and Lasso regression model directly i will choose Lasso although there might be a small difference in the score but the feature selection will make the model simpler thus more robust and generalizable. The value of all the scoring metrics have been posted in the last answer.

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?

After excluding the earlier top 5 most important predictor variables and rebulding the model we can see the new 5 most important predictor variable in our lasso model are :-

1st Model Using RFE -

1) GarageType_BuiltIn
2) BsmtFinType1_GLQ
3) SaleCondition_Partial
4) MasVnrType_Stone
5) FireplaceQu_TA

2nd Model Directly Putting the data into Ridge and Lasso -

1) GrLivArea
2) Neighborhood_NridgHt
3) SaleCondition_Partial
4) TotalBsmtSF
5) BsmtExposure_Gd

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?

To make a model robust we have to make it simpler and generalizable i.e we have to create a model with lesser number of dependent variables. Due to simplification of the model we might face the issue of underfitting and our model will not perform well .This will also increase the bias and reduce the variance of the model as well which reduces the overall accuracy of the model.The other side of the coin will be overfitting the model where we use a complex model and which will reduce the bias but will create high variance and the model will learn the training dataset and might not perform well on the out of time test dataset. To combat both of these issues we use Regularisation which help us in making a model that do not suffer from underfitting as well as overfitting and thus, making it a robust and generalisable model.