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: For ridge my model has option value of alpha as 12 (Score: 84.14%).

For lasso it is 2 (Score: 78.43).

If I double the score is getting reduced. Important predictor variables are:

For lasso:

['Id', 'LotFrontage', 'LotArea', 'OverallQual', 'OverallCond',
'YearBuilt', 'YearRemodAdd', 'MasVnrArea', 'BsmtUnfSF',
'TotalBsmtSF', '1stFlrSF', '2ndFlrSF', 'GrLivArea', 'BsmtFullBath',
'BedroomAbvGr', 'KitchenAbvGr', 'TotRmsAbvGrd', 'GarageCars',
'GarageArea', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch',
'3SsnPorch', 'ScreenPorch', 'PoolArea', 'MoSold', 'YrSold',
'MSZoning_FV', 'MSZoning_RH', 'MSZoning_RL', 'MSZoning_RM',
'Street_Pave', 'Alley_No Alley', 'Alley_Pave', 'BldgType_2fmCon',
'BldgType_Duplex', 'BldgType_Twnhs', 'BldgType_TwnhsE',
'RoofMatl_CompShg', 'RoofMatl_Membran', 'RoofMatl_Metal',
'RoofMatl_Roll', 'RoofMatl_Tar&Grv', 'RoofMatl_WdShake',
'RoofMatl_WdShngl', 'ExterQual_Fa', 'ExterQual_Gd', 'ExterQual_TA',
'BsmtQual_Fa', 'BsmtQual_Gd', 'BsmtQual_TA', 'CentralAir_Y',
'KitchenQual_Fa', 'KitchenQual_Gd', 'KitchenQual_TA',
'FireplaceQu_Fa', 'FireplaceQu_Gd', 'FireplaceQu_Po',
'FireplaceQu_TA', 'GarageFinish_RFn', 'GarageFinish_Unf',
'PoolQC_Fair', 'PoolQC_Good', 'PoolQC_No Pool', 'Fence_Good Wood',
'Fence_Min Privacy', 'Fence_Min wood wire', 'Fence_No Fence']

For ridge:

['Id', 'LotFrontage', 'LotArea', 'OverallQual', 'OverallCond',
'YearBuilt', 'YearRemodAdd', 'MasVnrArea', 'BsmtUnfSF',
'TotalBsmtSF', '1stFlrSF', '2ndFlrSF', 'GrLivArea', 'BsmtFullBath',
'BedroomAbvGr', 'KitchenAbvGr', 'TotRmsAbvGrd', 'GarageCars',
'GarageArea', 'WoodDeckSF', 'OpenPorchSF', 'EnclosedPorch',
'3SsnPorch', 'ScreenPorch', 'PoolArea', 'MoSold', 'YrSold',
'MSZoning_FV', 'MSZoning_RH', 'MSZoning_RL', 'MSZoning_RM',
'Street_Pave', 'Alley_No Alley', 'Alley_Pave', 'BldgType_2fmCon',
'BldgType_Duplex', 'BldgType_Twnhs', 'BldgType_TwnhsE',
'RoofMatl_CompShg', 'RoofMatl_Membran', 'RoofMatl_Metal',
'RoofMatl_Roll', 'RoofMatl_Tar&Grv', 'RoofMatl_WdShake',
'RoofMatl_WdShngl', 'ExterQual_Fa', 'ExterQual_Gd', 'ExterQual_TA',
'BsmtQual_Fa', 'BsmtQual_Gd', 'BsmtQual_TA', 'CentralAir_Y',
'KitchenQual_Fa', 'KitchenQual_Gd', 'KitchenQual_TA',
'FireplaceQu_Fa', 'FireplaceQu_Gd', 'FireplaceQu_Po',
'FireplaceQu_TA', 'GarageFinish_RFn', 'GarageFinish_Unf',
'PoolQC_Fair', 'PoolQC_Good', 'PoolQC_No Pool', 'Fence_Good Wood',
'Fence_Min Privacy', 'Fence_Min wood wire', 'Fence_No Fence

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: I would apply Ridge as it gives me a higher score. That is 84.14%

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?

Ans. Question is unclear

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 model robust and Generalisable outliers only relevant to the data must be treated instead of all. This would also increase the accuracy of the model.