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?

OverallQual, OverallCond, MSZoning_RH, PoolArea, 1stFlrSF, LotArea, LotFrontage are some of the variables which describes dependent variable in Lasso regression.
OverallQual, OverallCond, YearBuilt, MasVnrArea, 1stFlrSF, 2ndFlrSF ,GrLivArea, ExterQual_TA, SaleType_New are some of the variables which describes dependent variable in Lasso regression
Alpha for lasso is 620.99
Alpha for ridge is 4.070
Lasso score for train data and test data is 0.847 and 0.842 respectively
Ridge score for train data and test data is respectively 0.855 and 0.866 respectively
After doubling the alpha

Lasso score for train data and test data is 0.839 and 0.869 respectively
Ridge score for train data and test data is respectively 0.854 and 0.868 respectively
There is not a drastic change in Lasso and Ridge if we double the alphas.
LotFrontage, LotArea, OverallQual, OverallCond, 1stFlrSF, LowQualFinSF, PoolArea, LowQualFinSF are some of the variables which describes dependent variable in Lasso regression.
OverallQual, OverallCond, YearBuilt, MasVnrArea, 1stFlrSF, 2ndFlrSF , GrLivArea, ExterQual_TA, SaleType_New are some of the variables which describes dependent variable in Lasso regression
So there is again not a major change in coeffiecients of Lasso and Ridge if we double the alphas.

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?

I will use Ridge regression since the scores are slightly better when compared to Lasso. Ridge performs better on the test data set, as the Rsquare value in test data is higher than train data.

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?

Mentioned in notebook :
OverallCond',
'LotArea',
'MSZoning_RH',

'MiscVal',
'2ndFlrSF',

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?

Model is robust and generalizable can be checked by :
- Having a test data set which a model hasn't seen to validate results
- Having a validation dataset helps in fine tuning hyperparameters and ruling out overfitting
- Efficient EDA and feature engineering
- Apply regularization to not overfit and have no mulitcolinearity
- Handle outliers
- Using cross validation methods to assess model

The main goal is to build a model which doesn't have overfitting and can predict well on unseen data. So a model should be simple. Techniques like EDA, Feature Engineering, Regularization , Cross Validation helps in model accuracy since the key idea is to not only have the model perform well on train data, but to perform well on unseen test data.