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?

- Optimal Value of lambda for Ridge : 5
- Optimal Value of lambda for Lasso : 0.0001

After doubling the value of alpha for both ridge and lasso, below are the observations:

Ridge:

- R2 score of train set decreased from 0.94 to 0.93
- R2 score of test set remained same at 0.93

Lasso:

- R2 score of train set remained same at 0.94
- R2 score of test set remained same at 0.92

Most Important predictors before and after changing lambda values are shown below.

Ridge		Lasso	
Before	After	Before	After
SaleCondition_Alloca	GrLivArea	Condition2_PosA	SaleCondition_Alloca
OverallQual_9	OverallQual_9	SaleCondition_Alloca	OverallQual_9
GrLivArea	OverallQual_8	OverallQual_9	Condition2_PosA
OverallQual_8	SaleCondition_Alloca	OverallQual_8	OverallCond_9
OverallCond_9	Functional_Typ	OverallCond_9	OverallQual_8
Neighborhood_Crawfor	OverallCond_9	GrLivArea	GrLivArea
Functional_Typ	Neighborhood_Crawfor	SaleType_Oth	Neighborhood_Crawfor
Exterior1st_BrkFace	Exterior1st_BrkFace	RoofMatl_WdShngl	Functional_Typ
Neighborhood_StoneBr	CentralAir_Y	Neighborhood_Crawfor	Exterior1st_BrkFace
CentralAir_Y	TotalBsmtSF	MSZoning_FV	Neighborhood_StoneBr

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?

The model that we choose depends on the dataset.

Lasso:

- When we see too many features which are irrelevant and needs feature selection

Ridge:

- When the number of features are less and have high correlation.

Considering we have more features which irrelevant and needed feature selection. We choose Lasso Regression

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 dropping top 5 predictors, we got below 5 features as top predictors

- RoofMatl_WdShngl
- GrLivArea
- SaleType_Oth
- Functional_Typ
- Exterior1st_BrkFace

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?

- A model is considered robust and generalisable when the performance of the model does not change with the changing dataset.
- To make sure that a model is robust and generalisable, we need make sure that the model does not overfit the training data.
- The model should not the too complex.
- Apply Regularization to prevent overfitting and improve the generalisation
- The implications of ensuring a model for robustness and generalisation are
  - It will have an impact on accuracy since we are not trying to overfit the model on training data
  - We need to find the right balance between accuracy and generalisation so that the model performs well on unseen data and is reliable.