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?

Answer: The optimal value of alpha for ridge regression is 10 & for lasso regression is 0.001.

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.935201	0.914017
R2 Score (Test)	0.920952	0.919032
RSS (Train)	8.149486	10.813744
RSS (Test)	2.729937	2.796244
MSE (Train)	0.006977	0.009258
MSE (Test)	0.009349	0.009576
RMSE (Train)	0.083530	0.096220
RMSE (Test)	0.096691	0.097858

Figure 1: with alpha 10 & 0.001

Metric	Ridge Regression	Lasso Regression
R2 Score (Train)	0.929321	0.897839
R2 Score (Test)	0.922358	0.905190
RSS (Train)	8.889102	12.848390
RSS (Test)	2.681357	3.274278
MSE (Train)	0.007611	0.011000
MSE (Test)	0.009183	0.011213
RMSE (Train)	0.087238	0.104882
RMSE (Test)	0.095827	0.105893

Figure 2: with double values of alpha

Changes in Ridge Regression metrics:

- R2 score of train set decreased from 0.93 to 0.92
- R2 score of test set remained same at 0.92

Changes in Lasso metrics:

- R2 score of train set decreased from 0.91 to 0.89
- R2 score of test set decreased from 0.91 to 0.90.

So, the most important predictor variables after we double the alpha values are:-

- GrLivArea
- OverallQual_8
- OverallQual_9

- Functional_Typ
- Neighborhood_Crawfor
- Exterior1st_BrkFace
- TotalBsmtSF
- CentralAir_Y

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?

Answer: The model we will choose to apply will depend on the use case. If we have too many variables and one of our primary goals is feature selection, then we will use Lasso. If we don't want to get too large coefficients and reduction of coefficient magnitude is one of our prime goals, then we will use Ridge 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?

Answer: Here, we will drop the top 5 features in Lasso model and build the model again.

Top 5 Lasso predictors were:

Functional_Typ, GrLivArea, OverallQual_8,Neighborhood_Crawfor & Exterior1st_BrkFace.

After dropping our top 5 lasso predictors, we get the following new top 5 predictors:-

2ndFlrSF, Functional_Typ, 1stFlrSF, MSSubClass_70 & Neighborhood_Somerst.

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?

Answer:

- When a model is resilient, any fluctuation in the data has little effect on its performance.
- A generalizable model may adapt to additional, previously unknown data obtained from the same distribution as the one used to generate the model.
- To ensure that a model is resilient and generalizable, we must ensure that it does not overfit. This is because an overfitting model has a very high variance, and even the smallest change in data has a large impact on model prediction.
- In general, we must establish a happy medium between model correctness and complexity. Regularisation techniques such as Ridge Regression and Lasso can help with this.