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?**

   I have built multiple models. The optimal value we get is dependent on the model. Taking one of the model for the explanation

   Before doubling

   Alpha Ridge 4.0
   Alpha Lasso 0.001

	Metric	Ridge Regression	Lasso Regression
0	R2 Score	0.877926	0.877311
1	-ve MSE	-0.019702	-0.019327
2	RMSE	0.140364	0.139023
3	Poisson_deviance	-0.820643	-0.881213
4	Gamma_deviance	-0.001595	-0.001606
5	Abs pt error	-0.000133	-0.000133
6	Abs err score	-0.007898	-0.007632
7	Explained variance	0.878754	0.878071

   When the alpha value is doubled, the way the model predicts changes because the co-eff values change. The observation is that the predictions now start underfitting. This can be observed in the decrease in R2 score.

   8.0
   0.002

	Metric	Ridge Regression	Lasso Regression
0	R2 Score	0.876061	0.864500
1	-ve MSE	-0.020094	-0.021481
2	RMSE	0.141754	0.146565
3	Poisson_deviance	-0.785573	-0.815449
4	Gamma_deviance	-0.001618	-0.001778
5	Abs pt error	-0.000135	-0.000148
6	Abs err score	-0.008013	-0.008286
7	Explained variance	0.876804	0.865268

   The most important predictor variable is the one with high co-eff value. After the change is implemented it is 'GrLivArea'

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 got the cross_val_score with various scoring metrics and the results are here -

	Metric	Ridge Regression	Lasso Regression
0	R2 Score	0.877926	0.877311
1	-ve MSE	-0.019702	-0.019327
2	RMSE	0.140364	0.139023

3	Poisson_deviance	-0.820643	-0.881213
4	Gamma_deviance	-0.001595	-0.001606
5	Abs pt error	-0.000133	-0.000133
6	Abs err score	-0.007898	-0.007632
7	Explained variance	0.878754	0.878071

Among Lasso and Ridge you can choose based on the requirement. If accuracy is more important you can go for Ridge, but the model will be complex. If you want a lesser number of predictors then go for Lasso. In my case, though Ridge has a better R2-score, I will go for Lasso because it has a lesser predictor variable.

3. **After building the model, you realized 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 removing the top 5 predictors from the incoming data, the top 5 predictors are
0	Neighborhood_Crawfor	1.099682
1	MSZoning_FV	1.376475
2	BsmtQual_Ex	1.776683
3	YearRemodAdd	4.977824
4	KitchenQual_Ex	1.663053

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?**

More simple model is robust and generalizable. So always try to select the model which is more simple and has better explainability.
By doing so, we may compromise little on the accuracy, but it will do better on unseen data.