Q1. 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: Optimal value of alpha for ridge and lasso regression is the value which provide us the balance between bias and variance i.e., to reduce model complexity and prevent overfitting as well as predict target value with good accuracy.

In this case,

Ridge alpha value=0.05

Lasso alpha value=0.0001

If we double the value of alpha, it reduces the coefficient values of features hence reducing the complexity further but with a decrease in bias i.e., model will predict the dependent feature with less accuracy.

Most important predictor: **GrLivArea**

Q2. 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 use Lasso Regression as it reduces some unnecessary features to coefficient 0 and others to very minimal (->0). Also, we can conclude from the output we got from both regularization technique that Lasso is performing slightly better on test dataset than Ridge regression (lower values of RSS and MSE and higher value of R2).

	Metric	Linear Regression	Ridge Regression	Lasso Regression
0	R2 Score (Train)	0.902888	0.902877	0.900786
1	R2 Score (Test)	0.894225	0.894315	0.895090
2	RSS (Train)	2.531568	2.531848	2.586361
3	RSS (Test)	0.668211	0.667643	0.662744
4	MSE (Train)	0.003309	0.057529	0.058145
5	MSE (Test)	0.003480	0.058969	0.058752

Q3. 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?

Ans:

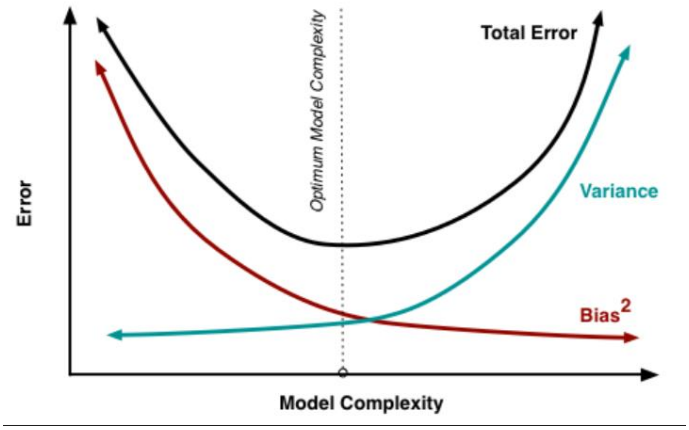
1. OverallCond_8
2. OverallQual_8
3. KitchenQual_Fa
4. OverallCond_7
5. KitchenQual_TA

	Linear	Ridge	Lasso
GrLivArea	0.361134	0.357112	3.622347e-01
YearBuilt	0.289283	0.287874	2.904408e-01
TotalBsmtSF	0.230316	0.231540	2.220681e-01
OverallCond_9	0.171161	0.167486	1.295733e-01
Neighborhood_Crawfor	0.131204	0.130905	1.264123e-01
OverallCond_8	0.158564	0.155647	1.230475e-01
OverallQual_8	0.117528	0.117761	1.207420e-01
KitchenQual_Fa	-0.127048	-0.126841	-1.201400e-01
OverallCond_7	0.151489	0.148749	1.172244e-01
KitchenQual_TA	-0.116782	-0.116494	-1.083047e-01
MSSubClass_160	-0.110792	-0.110447	-1.057402e-01
OverallQual_9	0.148939	0.142636	8.544029e-02
KitchenQual_Gd	-0.093645	-0.093041	-8.275073e-02
OverallCond_6	0.112669	0.110033	7.939951e-02
BsmtUnfSF	-0.074969	-0.074830	-7.340870e-02

Q4. How can you make sure that a model is robust and generalizable? What are the implications of the same for the accuracy of the model and why?

Ans: Robust and generalizable model is the model where its dependent variable is consistently accurate with unseen data (on changing the independent variable drastically).

To make a model more robust and generalizable, we need to minimize the total error of the model that is created due to variance and bias.



For achieving it, we need to do the EDA very thoroughly like removing **outliers**, handling **missing values** Properly (using mean, median, mode) then check **correlations** and **multicollinearity** between independent variables and properly treat them. Then at last use proper regularization to reduce the complexity and overfitting of model hence reducing the **Total Error** of the model by balancing **variance** and **Bias**.

More the robustness of the model more accuracy it will provide with the unknown data.