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?

Ans: The optimal values of the alpha for ridge and lasso regression are 0.1 and 0.001,

After fine tuning , the optimum values choosen are 0.6 and 0.0012, then the r2 scores of the train and test sets are:

```
For Ridge:
train score for Ridge regression 0.9482917065149058
Test score for Ridge regression 0.8423517325209512
For Lasso:
r2_score on train data for lasso: 0.9065827275850737
r2_score on test data for Lasso: 0.8702887135691463
```

If we increase the Alpha value by double, the r2 scores are as follows:

```
For Ridge:
r2_score on train data using Ridge regression: 0.9418267618997298
r2_score on train data using Ridge regression: 0.8646246226630604

for Lasso:
r2_score on train data using Lasso regression: 0.8873051908321303
r2_score on test data using Lasso regression:: 0.8738075199509353
```

The most important predictor variables after the change is implemented are:

For Ridge:

	Feature	Coef
0	MSSubClass	11.053643
38	MSZoning_RL	0.278698
37	MSZoning_RH	0.264344
36	MSZoning_FV	0.261873
39	MSZoning_RM	0.246014
117	RoofMatl_WdShngl	0.221891
249	SaleType_ConLD	0.196025
90	Condition2_PosA	0.181873
111	RoofMatl_CompShg	0.159732
61	Neighborhood_Crawfor	0.153384

For Lasso

	Feature	Coef
0	MSSubClass	11.839185
15	BsmtFullBath	0.111414
61	Neighborhood_Crawfor	0.099636
76	Neighborhood_Somerst	0.093699
3	OverallCond	0.093477
71	Neighborhood_NridgHt	0.078829
81	Condition1_Norm	0.057439
5	YearRemodAdd	0.052484
38	MSZoning_RL	0.049480
120	Exterior1st_BrkFace	0.049403

If we increase the Alphavalue by double, the test score have increased by 0.02 i.e. with aplha = 0.6 the test score is 0.84 whereas with alpha=1.2, the test score is 0.86 and the most important predictor variables after the change is implemented is as mentioned above for ridge.

If we increase the Alphavalue by double, the test score have increased by 0.02 i.e. with aplha = 0.0012 the test score is 0.84 whereas with alpha=0.0024, the test score is 0.86 and the most important predictor variables after the change is implemented is as mentioned above for lasso.

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?

Ans: I'll choose Lasso regression as the best model because it is predicting well on both the training and test data sets, also we know from the theory that the Lasso regression makes the coefficients of many variables to 0 which will make our model simpler. As per Occam's Razor principle, the model with fewer parameters is a good model that can be explained well in business. Hence I'll use Lasso's optimal value i.e. 0.0012 as the Alpha value for the model given.

```
For Ridge:alpha value:0.6
train score for Ridge regression 0.9482917065149058
Test score for Ridge regression 0.8423517325209512
For Lasso:alpha value:0.0012
r2_score on train data for lasso: 0.9065827275850737
r2_score on test data for Lasso: 0.8702887135691463
```

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?

Ans: After creating the Lasso Regression model, the five most important variables are as below:

	Feature	Coef
0	MSSubClass	11.839185
15	BsmtFullBath	0.111414
61	Neighborhood_Crawfor	0.099636
76	Neighborhood_Somerst	0.093699
3	OverallCond	0.093477

Now, after excluding these predictor variables, the new important predictor variables are :

	Feaure	Coef
0	LotArea	11.428473
33	MSZoning_RH	0.354390
35	MSZoning_RM	0.351430
34	MSZoning_RL	0.337598
36	Street_Pave	0.314373
244	SaleType_ConLI	0.175740
204	FireplaceQu_Fa	0.112920
115	Exterior1st_CBlock	0.099121
11	LowQualFinSF	0.080601
10	2ndFlrSF	0.077684

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?

Ans: A model is called robust when the there is a very minute difference in the train and test scores of the model, i.e for example, if the training score is 0.90 and the testing score is in the range of 0.86-0.90 is , then it is called as a robust model, in other way , there should be very low bias and low variance, then that model is called as a robust model, this can achieved by making the learn to clean the data using Exploratory Data Analysis such as treating the null values, Nan values, outliers etc from the data.

Generalization refers to the model's capability to adapt to the new changes or new additions of data to the training data, i.e in real-world scenarios, data is constantly added to the model such that the model learns the new inputs from the new data and performs the regression and gives the best test scores on the unseen data, such model is called as robust and generalisable model.