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:-

Optimal Value of alpha for ridge regression :- **270**

Optimal Value of alpha for lasso regression :- **0.004**

Changes After doubling value of alpha

For Ridge

Before Change	**After Change**
Train Score **0.905926363041003** Test Score **0.8817831974348409**	Train Score **0.8969057157083914** Test Score **0.8764750821983084**

Top 10 features & its coefficient

	Ridge_Features	Ridge_Coeff	AbsRidge_Coeff
4	OverallQual	0.05	0.05
22	GrLivArea	0.04	0.04
5	OverallCond	0.03	0.03
32	GarageCars	0.03	0.03
19	1stFlrSF	0.03	0.03
59	Neighborhood_Crawfor	0.02	0.02
30	TotRmsAbvGrd	0.02	0.02
31	Fireplaces	0.02	0.02
25	FullBath	0.02	0.02
23	BsmtFullBath	0.02	0.02

Top 10 features & its coefficient

	Ridge_Features	Ridge_Coeff	AbsRidge_Coeff
4	OverallQual	0.04	0.04
22	GrLivArea	0.03	0.03
23	BsmtFullBath	0.02	0.02
30	TotRmsAbvGrd	0.02	0.02
69	Neighborhood_NridgHt	0.02	0.02
31	Fireplaces	0.02	0.02
32	GarageCars	0.02	0.02
25	FullBath	0.02	0.02
33	GarageArea	0.02	0.02
68	Neighborhood_NoRidge	0.02	0.02

For Lasso

Before Change	After Change
Train Score **0.9046794291135135** Test Score **0.886180836982245**	Train Score **0.8930659286789758** Test Score **0.879412980506506**

Top 10 features & its coefficient

	lasso_Features	lasso_Coeff	Abslasso_Coeff
22	GrLivArea	0.10	0.10
4	OverallQual	0.09	0.09
32	GarageCars	0.04	0.04
5	OverallCond	0.04	0.04
42	Age	-0.03	0.03
69	Neighborhood_NridgHt	0.03	0.03
6	YearRemodAdd	0.02	0.02
59	Neighborhood_Crawfor	0.02	0.02
31	Fireplaces	0.02	0.02
29	KitchenQual	0.02	0.02

Top 10 features & its coefficient

	lasso_Features	lasso_Coeff	Abslasso_Coeff
22	GrLivArea	0.10	0.10
4	OverallQual	0.10	0.10
32	GarageCars	0.04	0.04
42	Age	-0.04	0.04
5	OverallCond	0.03	0.03
10	BsmtQual	0.02	0.02
29	KitchenQual	0.02	0.02
12	BsmtFinType1	0.02	0.02
45	MSZoning_RL	0.02	0.02
74	Neighborhood_Somerst	0.02	0.02

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:-

Value of lambda for lasso(0.004) is very less compared to lambda value for ridge so model simplicity is there and still model performance for test data for lasso is better than ridge.

So will choose lambda for lasso.

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:-

	lasso_Features	lasso_Coeff	Abslasso_Coeff
18	2ndFlrSF	0.08	0.08
17	1stFlrSF	0.08	0.08
4	YearRemodAdd	0.04	0.04
8	BsmtQual	0.04	0.04
29	GarageArea	0.04	0.04

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:-

By selecting the optimal value of lambda & By checking the model performance for both train and test & if it is high for both data then we can say that model is neither underfitting nor overfitting and it is robust and generalized.

If the value of lambda is too high then model will be simpler and bias will be high so chances of error will be high and model performance will not be good.

If the value of lambda is too small then complexity of model will be higher and so variance will be high and it will overfit and performance for test data will not be good.

So we need to tune the hyperparameter (lambda) to optimal such that model is generalizable and its performance/accuracy is good for both test & train data.