Subjective Questions:

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 Value of alpha for Ridge Regression is 0.6.

Model performance metrics using the alpha as 0.6 are shown below:

	Train	Test
Ridge_MSE	0.018349	0.025502
Ridge_RMSE	0.135457	0.159695
Ridge_R2Score	0.883237	0.845266

The Optimal Value of alpha for Lasso Regression is 0.0001.

Model performance metrics using the alpha as 0.0001 are shown below:

	Train	Test
Lasso_MSE	0.018335	0.025248
Lasso_RMSE	0.135406	0.158896
Lasso_R2Score	0.883325	0.846811

Generally, if we increase the alpha value the model becomes underfitting and if we decrease the alpha value the model becomes overfitting.

Doubling the Optimal Value for Ridge Regression:

The new optimal value of alpha is 1.2.

Model performance metrics using the alpha as 1.2 are shown below:

	Train	Test
Ridge_MSE	0.018628	0.025113
Ridge_RMSE	0.136483	0.158470
Ridge_R2Score	0.881461	0.847631

Doubling the Optimal Value for Lasso Regression:

The new optimal value of alpha is 0.0002.

Model performance metrics using the alpha as 0.0002 are shown below:

	Train	Test
Lasso_MSE	0.018135	0.026484
Lasso_RMSE	0.134666	0.162738
Lasso_R2Score	0.884597	0.839312

Most Important Predictor Variables after doubling the optimal values are:

Ridge (alpha = 1.2)

Features	Coefficient
MSZoning_RL	0.335134
MSZoning_RH	0.262377
MSZoning_FV	0.225064
MSZoning_RM	0.223024
ExterCond_Fa	-0.172527
Neighborhood_Somerst	0.171598
LandContour_Low	0.166616
Foundation_Stone	0.160927
Exterior2nd_VinylSd	0.158551
Neighborhood_Veenker	0.153071

Lasso(alpha=0.0002)

Features	Coefficient
MSZoning_RL	0.538465
MSZoning_RH	0.490033
MSZoning_FV	0.440788
MSZoning_RM	0.430366
ExterCond_Fa	-0.381862
Foundation_Stone	0.275779
Exterior1st_BrkComm	-0.264903
ExterCond_TA	-0.263576
ExterCond_Gd	-0.251936
GarageQual_Po	-0.214004

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:

- The Optimal lambda Values of Ridge and Lasso are:
  - Ridge ----- 0.6
  - Lasso ------ 0.0001

- The MSE values of Ridge and Lasso are:
  - Ridge (MSE Train) ------ 0.018349
  - Ridge (MSE Test) -------- 0.025502
  - Lasso (MSE Train) ------- 0.018335
  - Lasso (MSE TEST) ------ 0.025248

- R2 Score of Ridge and Lasso are:
  - Ridge (R2 Score Train) ------ 0.883237
  - Ridge (R2 Score Test) ------- 0.845266
  - Lasso (R2 Score Train) ------ 0.883325
  - Lasso (R2 Score Test) -------- 0.846811

We can see that the difference in between lasso and ridge is very minute and we can see the lasso has little bit of good performance than ridge.

And also, if we compare the coefficients, we can see that lasso coefficients are near to 0 than ridge so, the variables predicted by lasso can be used in predicting the house price.

Also Lasso helps in Feature Elimination making it as simpler and more robust.

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:

We have selected Lasso Model so we are dropping top 5 coefficients from it.

Alpha value is 0.0001.

Model Performance metrics:

	Train	Test
Lasso_MSE	0.018335	0.025248
Lasso_RMSE	0.135406	0.158896
Lasso_R2Score	0.883325	0.846811

Dropped Features:

Features	Coefficient
MSZoning_RL	0.538465
MSZoning_RH	0.490033
MSZoning_FV	0.440788
MSZoning_RM	0.430366
ExterCond_Fa	-0.381862

After dropping the above 5 features now after building the model the top features are: Alpha Value is 0.0001

Features	Coefficient
LandContour_Low	0.184309
Exterior2nd_VinylSd	0.179995
Foundation_Stone	0.179255
Neighborhood_Veenker	0.174444
BldgType_Twnhs	-0.167209

Model Performance Metrics:

	Train	Test
Lasso_MSE	0.021108	0.028367
Lasso_RMSE	0.145285	0.168424
Lasso_R2Score	0.865680	0.827887

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:

- We say a model is robust and generalisable:
  - when it is simple but not complex.
  - The model is generalisable when the test accuracy is not lesser than the training accuracy.
  - The model should not be impacted from outliers.
  - It should perform similar to the training set on any test set.
- The accuracy of the model can be maintained by balancing the bias and variance as it minimizes the error.