Subjective Questions

Question 1: **What is the optimal value of alpha for ridge and lasso regression? What will be the changes in the model if you choose to double the value of alpha for both ridge and lasso? What will be the most important predictor variables after the change is implemented?**

Answer: The optimal value for Ridge Regression is 1 and for the Lasso Regression it is 0.00009. After doubling the alpha values there is not much difference in r2 scores and slight difference in the coefficient values both in ridge and lasso regression.

The following are the top results before and after changes

Ridge Regression

Before

	Ridge Co-Efficient
GrLivArea	0.050729
OverallQual	0.046780
BsmtExposure_Gd	0.044848
Neighborhood_NridgHt	0.042378
Neighborhood_Crawfor	0.039044
SaleType_CWD	0.036295
Neighborhood_StoneBr	0.034124
SaleCondition_Partial	0.032772
BsmtCond_Gd	-0.000382
KitchenQual_TA	-0.030739
Neighborhood_Blueste	-0.031650
RoofMatl_WdShake	-0.044786
BldgType_Twnhs	-0.050624
KitchenQual_Fa	-0.060561
Neighborhood_OldTown	-0.062382
Neighborhood_MeadowV	-0.067624
Neighborhood_BrDale	-0.075234

After

	Ridge Doubled Alpha Co-Efficient
GrLivArea	0.050733
OverallQual	0.047428
BsmtExposure_Gd	0.043952
Neighborhood_NridgHt	0.041123
Neighborhood_Crawfor	0.037102
SaleCondition_Partial	0.032310
Neighborhood_StoneBr	0.031689
SaleType_CWD	0.027306
BsmtCond_Gd	0.000060
Neighborhood_Blueste	-0.021085
RoofMatl_WdShake	-0.029908
KitchenQual_TA	-0.030418
BldgType_Twnhs	-0.051192
KitchenQual_Fa	-0.053415
Neighborhood_OldTown	-0.059859
Neighborhood_MeadowV	-0.062542
Neighborhood_BrDale	-0.068918

Lasso Regression

Before

	Lasso Co-Efficient
GrLivArea	0.050578
OverallQual	0.047210
BsmtExposure_Gd	0.044367
Neighborhood_NridgHt	0.041452
Neighborhood_Crawfor	0.038132
SaleCondition_Partial	0.032279
Neighborhood_StoneBr	0.031600
SaleType_CWD	0.020867
BsmtCond_Gd	-0.000000
Neighborhood_Blueste	-0.000000
RoofMatl_WdShake	-0.022466
KitchenQual_TA	-0.030356
BldgType_Twnhs	-0.049690
KitchenQual_Fa	-0.059154
Neighborhood_OldTown	-0.062237
Neighborhood_MeadowV	-0.067918
Neighborhood_BrDale	-0.076943

After

	Lasso Doubled Alpha Co-Efficient
GrLivArea	0.050733
OverallQual	0.047428
BsmtExposure_Gd	0.043952
Neighborhood_NridgHt	0.041123
Neighborhood_Crawfor	0.037102
SaleCondition_Partial	0.032310
Neighborhood_StoneBr	0.031689
SaleType_CWD	0.027306
BsmtCond_Gd	0.000060
Neighborhood_Blueste	-0.021085
RoofMatl_WdShake	-0.029908
KitchenQual_TA	-0.030418
BldgType_Twnhs	-0.051192
KitchenQual_Fa	-0.053415
Neighborhood_OldTown	-0.059859
Neighborhood_MeadowV	-0.062542
Neighborhood_BrDale	-0.068918

***Question 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?**

Answer: The optimum lambda value in case of Ridge is 1

The optimum lambda value in case of Lasso is 0.00009

The Mean Squared Error in case of Ridge and Lasso are 0.0032632780613532845 and 0.0032619715193470656. The Mean Squared Error of both the models are almost same.

Lasso model has more near zero coefficients than ridge. So, I prefer Lasso over Ridge even though validation scores are almost equal.

Question 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?

Answer: The following are the top -5 features of current Lasso model.

1. 'GrLivArea',
2. 'OverallQual',
3. 'BsmtExposure_Gd',
4. 'Neighborhood_NridgHt'
5. 'Neighborhood_Crawfor'

After removing these features from dataset and rebuild the lasso model. Following are the top-5 features:

	Lasso Co-Efficient
SaleType_CWD	0.131506
Neighborhood_StoneBr	0.064718
SaleCondition_Partial	0.045327
BsmtCond_Gd	0.030206
Neighborhood_Blueste	-0.000000

Question 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?**

Answer: According to, Occam's Razor-given two models that show comparable 'execution' in the limited preparation or test information, we ought to pick the one that makes less on the test information because of understanding reasons: -

- Simpler models are normally more 'conventional' and are all the more generally pertinent
- Simpler models require less preparation tests for viable preparation than the more complicated ones and henceforth are more straightforward to prepare.
- Simpler models are stronger. o Complex models will generally change fiercely with changes in the preparation informational index o Simple models have low difference, high inclination and complex models have low predisposition, high fluctuation o Simpler models make more mistakes in the preparation set. Complex models lead to overfitting - they function admirably for the preparation tests, bomb hopelessly when applied to other test tests Therefore, to make the model stronger and more generalizable, simplify the model however not more straightforward which won't be of any utilization. Regularization can be utilized to simplify the model. Regularization assists with finding some kind of harmony between keeping the model straightforward and not making it too gullible to even consider being of any utilization. For relapse, regularization includes adding a regularization term to the expense that includes the outright qualities or the squares of the boundaries of the model. Likewise, Making a model straightforward prompts Bias-Variance Trade-

off: • A perplexing model should change for every last change in the dataset and thus is truly temperamental and incredibly delicate to any progressions in the preparation information.

- A less difficult model that modified works out some example followed by the information focuses given is probably not going to change ridiculously regardless of whether more focuses are added or taken out. Inclination evaluates how exact is the model liable to be on test information. An intricate model can make a precise showing expectation gave there is sufficient preparation information. Models that are too guileless, for e.g., one that offers same response to all test sources of info and makes no separation at all has an exceptionally enormous inclination as its normal blunder across all test inputs are extremely high. Fluctuation alludes to the level of changes in the actual model concerning changes in the preparation information. Hence exactness of the model can be kept up with by keeping the harmony among Bias and Variance as it limits the complete mistake as displayed in the beneath diagram