2)    Analyze which feature alone would give the best prediction, list the scores and RMSE errors achieved by the top 10 predictors by score.

(create a loop to train a linear model for each feature and save the scores and RMSE and sort by score)

OverallQual score: 0.6454631197278845 rmse: 49018.435788124974

ExterQual\_Coded score: 0.49917534712671574 rmse: 58260.14865535065

GrLivArea score: 0.4442896667814059 rmse: 61369.55515927839

KitchenQual\_Coded score: 0.43970293723342646 rmse: 61622.30113827634

TotalBsmtSF score: 0.42491266111423953 rmse: 62430.33167581401

1stFlrSF score: 0.4166679048480356 rmse: 62876.2562386846

GarageCars score: 0.41193116659665785 rmse: 63131.022100901646

GarageArea score: 0.4073263621599067 rmse: 63377.71018105261

BsmtQual\_Coded score: 0.3409700148262621 rmse: 66831.5181312807

GarageFinish\_Coded score: 0.3162849544224503 rmse: 68071.65468394083

as we can see, 'OverallQual score' would give us the best prediction.

3)    Select all possible 2 pairs of these top 10 predictors, and train 45 linear models, list the scores and RMSE errors achieved by the top 10 predictors by score.

OverallQual, 1stFlrSF score: 0.7173882583039418 rmse: 43764.68831784155

OverallQual, TotalBsmtSF score: 0.7067911081524372 rmse: 44577.66409611637

OverallQual, GrLivArea score: 0.6921986779040231 rmse: 45673.467009316504

OverallQual, GarageArea score: 0.6813946935673895 rmse: 46468.13490082074

OverallQual, KitchenQual\_Coded score: 0.6730869536074612 rmse: 47070.07230442363

OverallQual, GarageCars score: 0.6723247870365776 rmse: 47124.910051666164

OverallQual, ExterQual\_Coded score: 0.668014734352153 rmse: 47433.824630568364

OverallQual, BsmtQual\_Coded score: 0.6566660974647347 rmse: 48237.75463112186

OverallQual, GarageFinish\_Coded score: 0.651902046585978 rmse: 48571.27166300007

ExterQual\_Coded, 1stFlrSF score: 0.6364054863411375 rmse: 49640.64317074208

4)    Train a single model using all features. Calculate RMSE and score. Observe how much of the prediction power was in the 2 pairs, vs all features.

'All Features Score: 0.8127547098527926 rmse: 35623.30345000717'

based on the results for all the features and the 2 feature prediction, we can conclude that all feature prediction had better prediction in overall.

5)    Use the 5NN and 10NN regressor with all features, and list the RMSE and score for these 2 models

'5NN Score: 0.6015421069563244 rmse: 51966.07983368139'

'10NN Score: 0.5949812006956909 rmse: 52392.16317256475'

observe if the results are better than linear regression?

we can see that the results of the linear regression is better than KNN algorithm

Which regressor is better for inference?

also linear regression is better for inference