1. After running the cells, add code to this notebook to achieve the following: Split the data for training and testing, to use 80 percent as training data.  use 21 as your randomization seed (so you achieve same results for us to grade).  <https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html>  random\_state = 21  train\_size = .8 In the following steps use training set to fit the model, and test set to evaluate it.

The accuracy of the training scores 0.8127547098528082

The accuracy of the training RMSE 20999.866601912054

1. 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)

These are the sorted top 10 highest scores [('OverallQual', 0.625651892462118), ('GrLivArea', 0.5021486502718042), ('ExterQual\_Coded', 0.4659963342490333), ('KitchenQual\_Coded', 0.4350717915853295), ('GarageCars', 0.41012393993308693), ('GarageArea', 0.388666759031819), ('TotalBsmtSF', 0.37648109325168655), ('1stFlrSF', 0.36705686969596585), ('BsmtQual\_Coded', 0.34246746596334543), ('FullBath', 0.31434385485924576)]

These are the sorted top 10 highest RMSE scores [('Electrical\_Mix', 197583.5712198401), ('Heating\_Floor', 197583.5712007741), ('Exterior1st\_AsphShn', 197583.57110544422), ('Exterior1st\_CBlock', 197583.57108811152), ('Exterior2nd\_CBlock', 197583.57108811152), ('Condition2\_RRAn', 197583.57097751155), ('RoofMatl\_Roll', 197583.57097718221), ('Utilities\_NoSeWa', 197583.57097544894), ('Exterior1st\_BrkComm', 197583.57095985126), ('RoofMatl\_ClyTile', 197583.57089745178)]

1. 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 and GrLivArea score is 0.35165781206918495 RMSE is 1597.8929645090116

OverallQual and ExterQual\_Coded score is 0.5274804461466409 RMSE is 3.8475218852573465

OverallQual and KitchenQual\_Coded score is 0.4533743408949562 RMSE is 3.7400095230997357

OverallQual and GarageCars score is 0.3608053097696037 RMSE is 4.471753050255096

OverallQual and GarageArea score is 0.3158684549059213 RMSE is 513.1556013850737

OverallQual and TotalBsmtSF score is 0.2892379811797141 RMSE is 1138.8482544111582

OverallQual and 1stFlrSF score is 0.2267891353818806 RMSE is 1219.178564129503

OverallQual and BsmtQual\_Coded score is 0.396118069597755 RMSE is 2.822730631707192

OverallQual and FullBath score is 0.3031600399564058 RMSE is 4.683547420316801

GrLivArea and ExterQual\_Coded score is 0.1900838435071074 RMSE is 1601.5779930456772

GrLivArea and KitchenQual\_Coded score is 0.17687302400791638 RMSE is 1601.4592365406525

GrLivArea and GarageCars score is 0.21832015037076447 RMSE is 1602.1397128805718

GrLivArea and GarageArea score is 0.21995863373530566 RMSE is 1141.526124602799

GrLivArea and TotalBsmtSF score is 0.20690508168916 RMSE is 684.4615231788626

GrLivArea and 1stFlrSF score is 0.3203831334097772 RMSE is 565.6594527083907

GrLivArea and BsmtQual\_Coded score is 0.1075548927269564 RMSE is 1600.5332700791125

GrLivArea and FullBath score is 0.3969146744120402 RMSE is 1602.3312229202816

ExterQual\_Coded and KitchenQual\_Coded score is 0.5128309988924686 RMSE is 0.48751536681084456

ExterQual\_Coded and GarageCars score is 0.2770866227741511 RMSE is 0.9118699739080504

ExterQual\_Coded and GarageArea score is 0.2457954483382392 RMSE is 516.7295223954874

ExterQual\_Coded and TotalBsmtSF score is 0.22144511894201813 RMSE is 1142.4490057862179

ExterQual\_Coded and 1stFlrSF score is 0.15827298549084334 RMSE is 1222.8274940223523

ExterQual\_Coded and BsmtQual\_Coded score is 0.3198559576847898 RMSE is 1.3127405193841153

ExterQual\_Coded and FullBath score is 0.2342070372898496 RMSE is 1.0085253034362762

KitchenQual\_Coded and GarageCars score is 0.25990583662950373 RMSE is 1.0233573539935121

KitchenQual\_Coded and GarageArea score is 0.2397136279552221 RMSE is 516.607562829468

KitchenQual\_Coded and TotalBsmtSF score is 0.1871683878851187 RMSE is 1142.3356660686923

KitchenQual\_Coded and 1stFlrSF score is 0.1497901209252963 RMSE is 1222.7087310949776

KitchenQual\_Coded and BsmtQual\_Coded score is 0.26105768114497985 RMSE is 1.252668384765809

KitchenQual\_Coded and FullBath score is 0.18891889728649355 RMSE is 1.149746845274723

GarageCars and GarageArea score is 0.7787628568112382 RMSE is 517.147612137701

GarageCars and TotalBsmtSF score is 0.18886397819633505 RMSE is 1143.0088738135144

GarageCars and 1stFlrSF score is 0.19299925778311156 RMSE is 1223.3926572591552

GarageCars and BsmtQual\_Coded score is 0.20177564107508228 RMSE is 1.9242504322534002

GarageCars and FullBath score is 0.22059182822718781 RMSE is 0.7181597487032633

GarageArea and TotalBsmtSF score is 0.23684327363016422 RMSE is 698.8271867824701

GarageArea and 1stFlrSF score is 0.2398860686970381 RMSE is 767.9398556643666

GarageArea and BsmtQual\_Coded score is 0.16492456880887918 RMSE is 515.7051051516283

GarageArea and FullBath score is 0.16455695945728321 RMSE is 517.51151929473

TotalBsmtSF and 1stFlrSF score is 0.6716293799317503 RMSE is 273.7766688079202

TotalBsmtSF and BsmtQual\_Coded score is 0.31843310682238224 RMSE is 1141.3535952771406

TotalBsmtSF and FullBath score is 0.1047962010713148 RMSE is 1143.251531546563

1stFlrSF and BsmtQual\_Coded score is 0.08619979417928536 RMSE is 1221.781403502773

1stFlrSF and FullBath score is 0.144884902579707 RMSE is 1223.6217295333968

BsmtQual\_Coded and FullBath score is 0.13782133485939774 RMSE is 2.1010434459201064

1. 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.

The accuracy of the score using all the data: 0.9184728890914108

The accuracy of the RMSE using al the data: 22675.410745448393

As of the scores for all features is high percentage 91.85% and RMSE for all feature is also high. It has a better accuracy

As for the top10 with 2 models, the scores and RMSE is unpredictable.

1. Use the 5NN and 10NN regressor with all features, and list the RMSE and score for these 2 models <https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsRegressor.html>  observe if the results are better than linear regression? Which regressor is better for inference? (edited)

Comparing with the 2 Models from Top 10 and the all features for 5NN and 10 NN. Its prediction from 5NN to 10NN gets to bigger number. However, the better predictor is for the all features and with a higher NN such as 10NN

Using the score and RMSE would be a better regressor for inference than using KNN because scores has a range on or within 1. RMSE has a range close to mean.