All the models below were submitted to the competition. The Benchmark score is 0.40890.

1) LinearRegression()
    a. Model1 – Score: 0.19883

- Used all columns: Categorical, numeric and the ones created during visual analysis;
- To use the categorical data, we first added the data from the "test.csv" file to the training data, then we generated the dummy columns. Later we removed the test data and trained the model. This way, we guarantee that the columns on both training and testing dataset are the same. **This strategy will be used in all models below;**

- The column "LotFrontage" had the missing values filled by an equation determined on Excel;

- No normalization done to the data

    b. Model2 - Score: 1.55232

- Data normalized

    c. Model3 - Score: 0.19921

- Column "LotFrontage" calculated with LinearRegression();

2) LassoCV() – Model4 - Score: 0.15103

    Alpha = 100

3) SVC – didn't work (predicted all rows the same values)

4) LinearSVC() – Model5 - Score: 0.40867

5) ElasticNet() – Model6 – Score: 0.15345

- Pipeline with StandardScaler and GridSearchCV

    Ps: ElasticNet without the StandardScaler didn't run

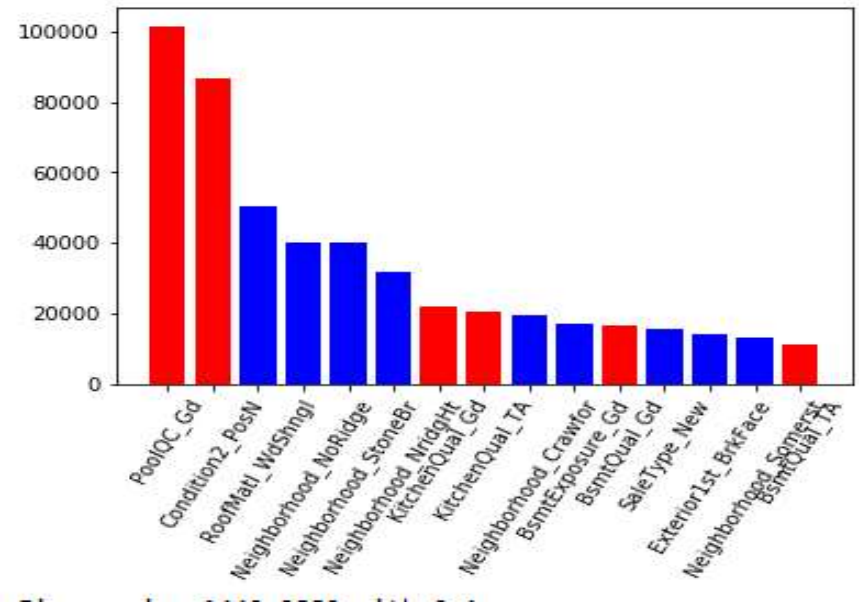
6) RandomForest
    a. Model7 – Score: 0.32349
        i. Random forest →max_depth = 3
        ii. GridSearchCV →cv = 3

    b. Model8 – Score: 0.30830
        i. Random forest →max_depth = 4
        ii. GridSearchCV → cv = 3

c. Model9 - Score: 0.30828
   i. Random forest →max_depth = 4 (found by gridsearch (best_params_)
   ii. GridSearchCV → cv = 5

d. Model10 – Score: 0.25614
   i. 'RF__max_depth': 7,
   ii. 'RF__n_estimators': 110
   iii. GridSearchCV → cv = 5

Using the best model (LassoCV), these are the most relevant features:



The model is good for predictions, but not so good to suggest a homeowner to upgrade their kitchen, because the data is separated in different columns.

For our project, it is relevant to keep the order of the materials according to their cost (1 = cheap, 10 = expensive), as well as keeping the ratings as numbers. This way, a regression makes more sense.

TO be continued…