

PREDICTING HOUSING PRICES WITH REGRESSION

The Ames Housing Data Set
Regression Challenge Project





WHAT SHOULD WE DO?

We need to make as accurate model as possible to predict **housing sale prices**.

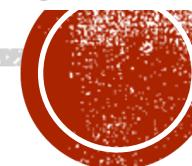
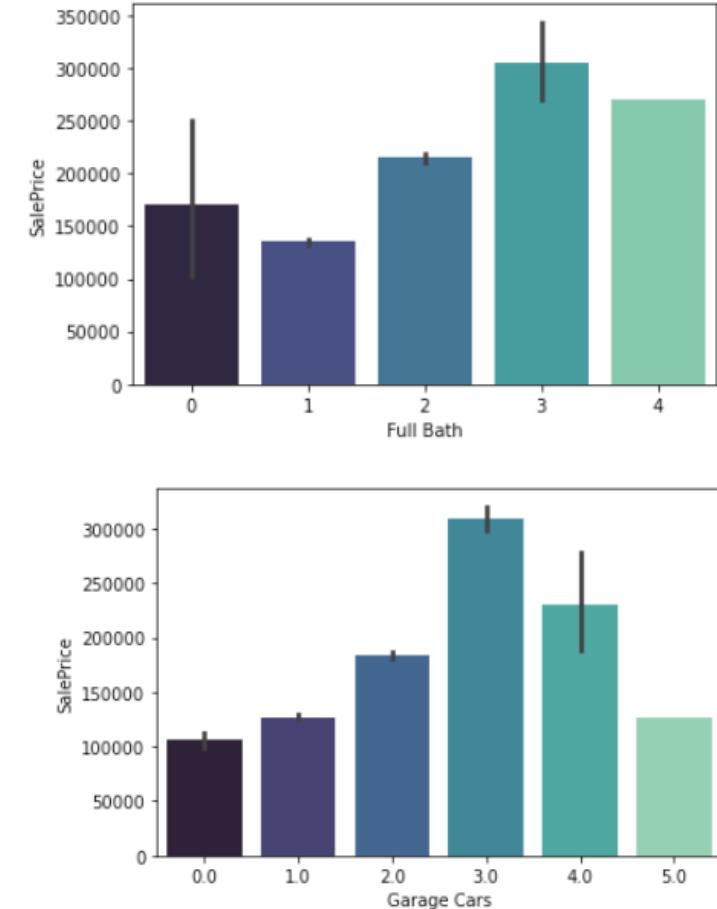
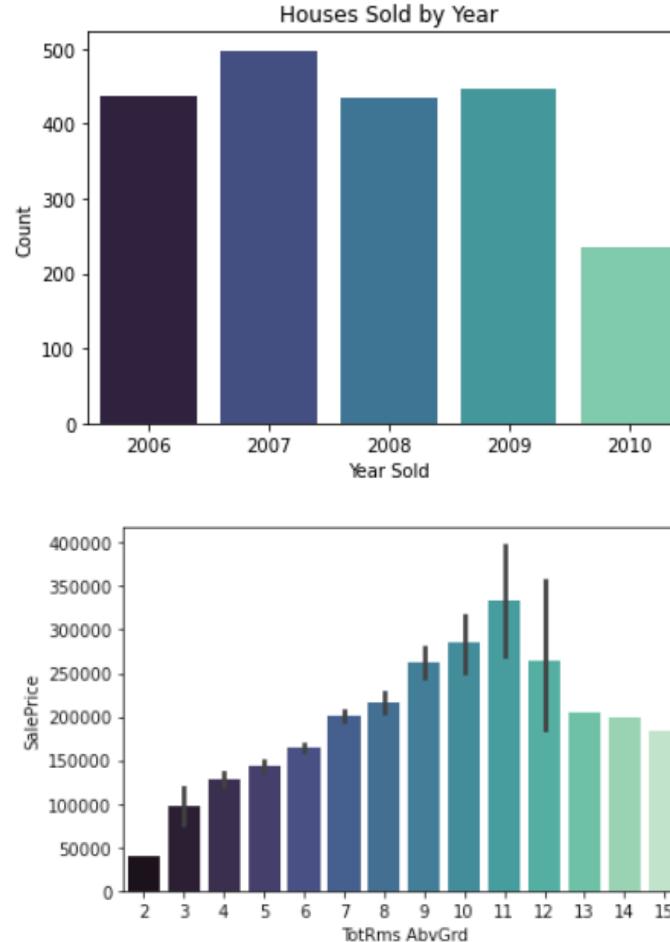
PREDICTING HOUSING SALE PRICES.

WHAT DO WE HAVE?

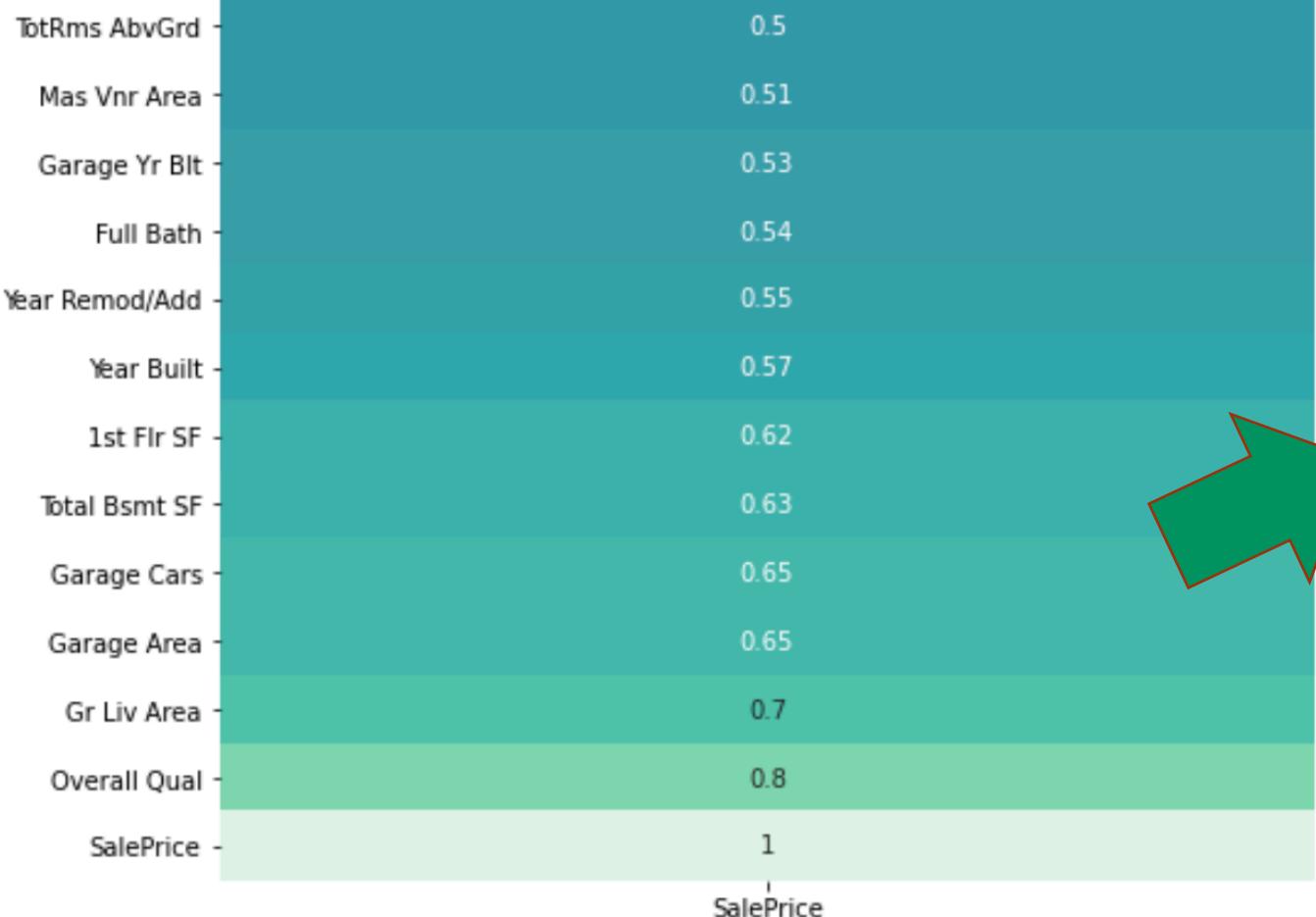
- Summarize Ames, Iowa sale prices
- Data set contains more than 2900 properties sold from 2006- 2010!
- 75 descriptive variables
 - 1- Categorical
 - * Nominal , e.g.: Neighborhood
 - * Ordinal, e.g. : Overall Quality
 - 2- Numerical
 - * Discrete, e.g.: Year Built
 - * Continuous, e.g. : Lot Area



DISTRIBUTION OF FEATURES

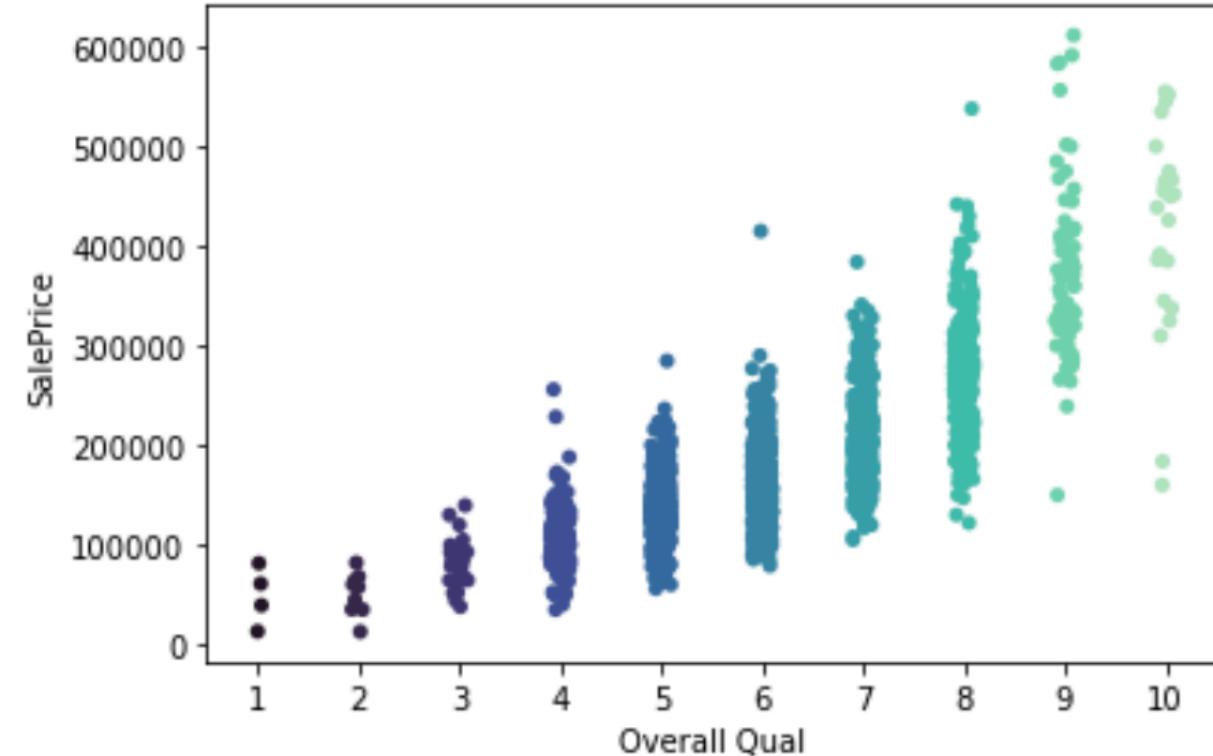


MOST HIGHLY CORRELATED HOME FEATURES WITH SALE PRICE

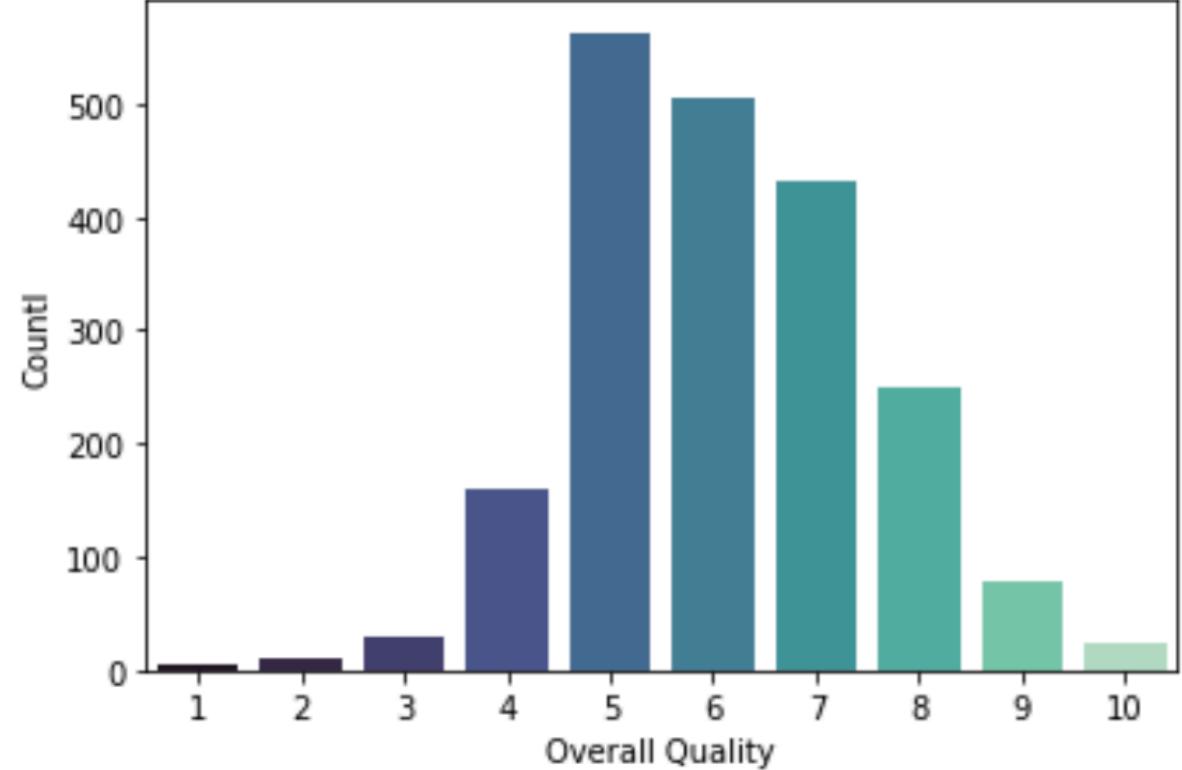


Feature	Correlation
Overall Quality	0.8
Ground Living Area	0.7
Garage Area	0.65
Year Built	0.57



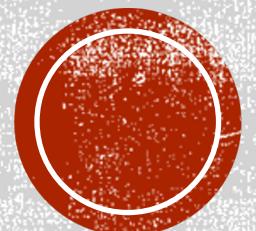


Houses Sale Price by Overall Quality

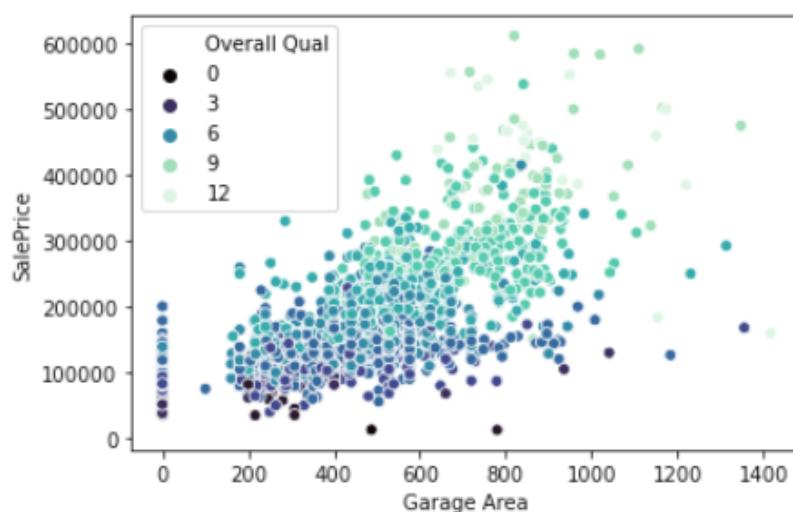


Distribution of Houses overall quality

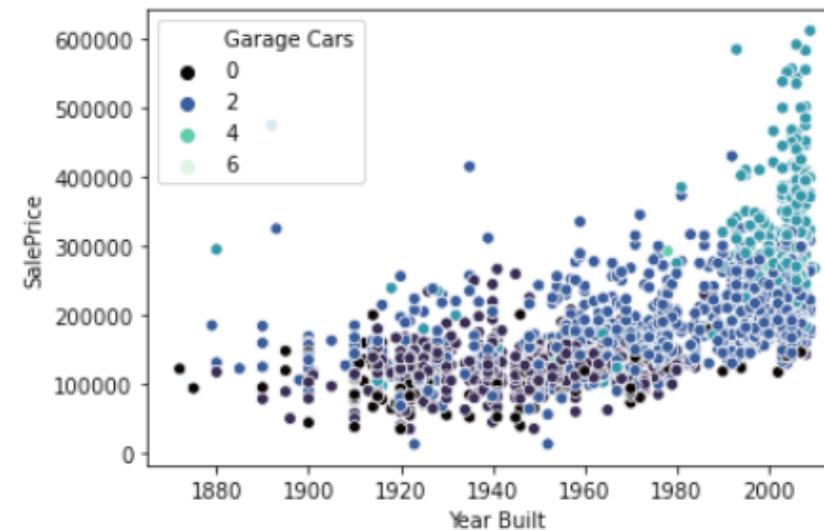
HIGHLY CORRELATED WITH HOUSE PRICE



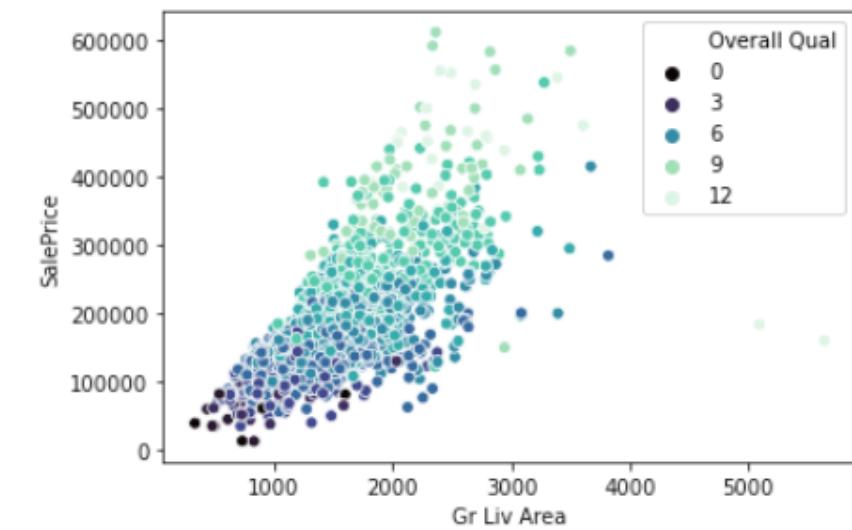
House Sale Price by Garage Area
and Overall Quality



House Sale Price by Year Built and
Garage Cars



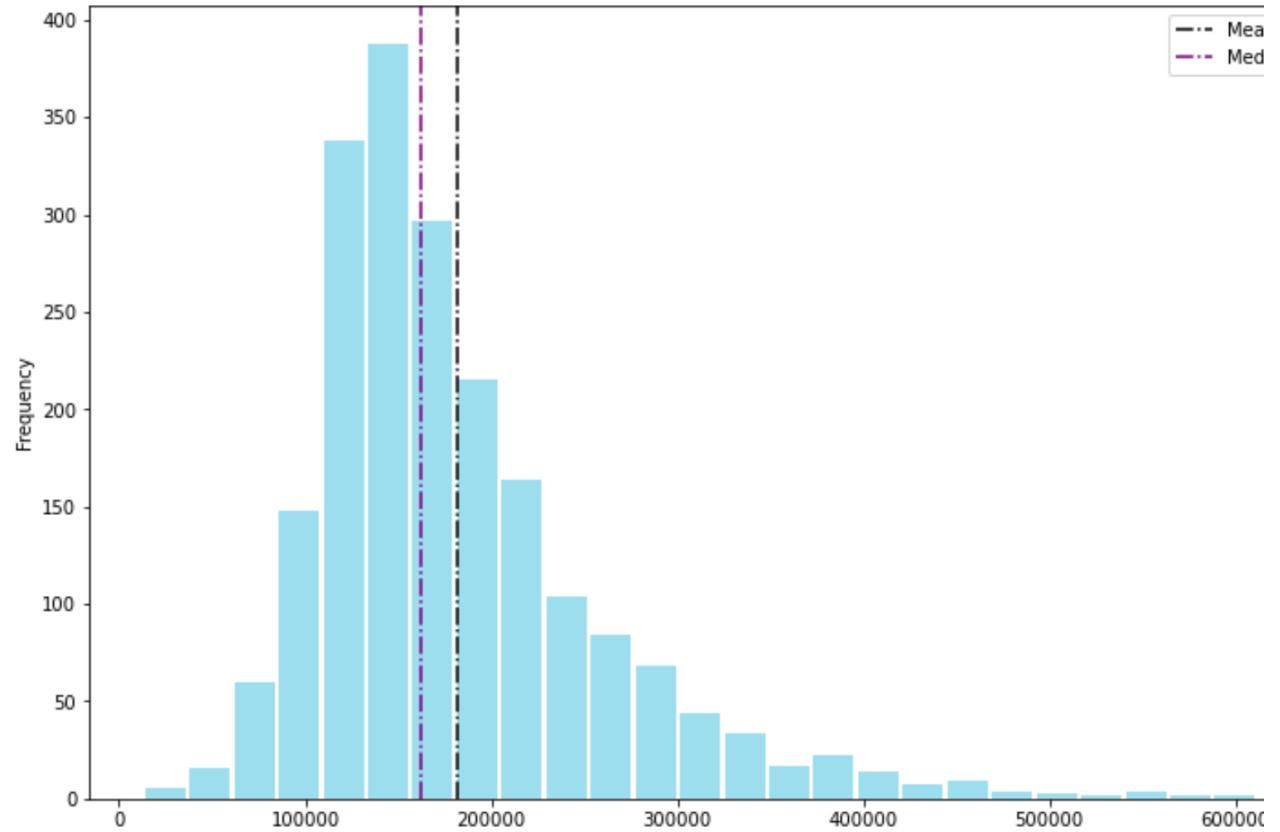
House Sale Price by Living Area SF
and Overall Quality



HOUSE SALE PRICE AND FEATURES CORRELATIONS



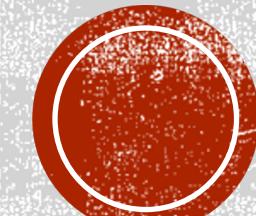
House Sale Price Distribution



Log House Sale Price Distribution

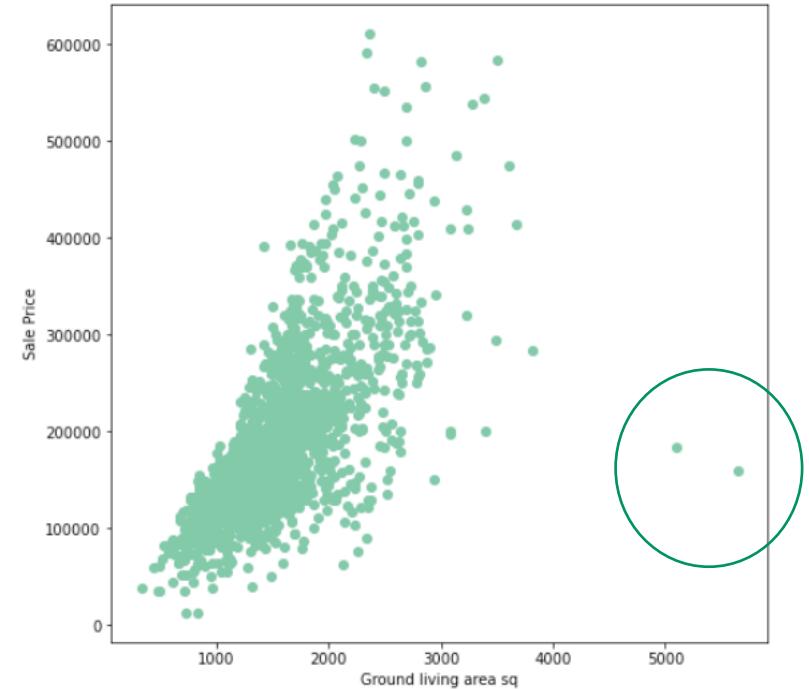


**SKEWED DISTRIBUTION
HOUSE SALE PRICE**



IMPLEMENTATION

- Removing outliers



- Feature Elimination

- 1- Low Qual Fin SF
- 2- Misc Val
- 3- 1st Flr SF
- 4- 2nd Flr SF
- 5- Low Qual

- Scaling the data set (Standard Scaler)

Coefficients	
Gr Liv Area	0.120238
Overall Qual	0.076007
Year Built	0.049491
Overall Cond	0.040683
Total Bsmt SF	0.038644
Lot Area	0.027297
BsmtFin Type 1	0.021056
Garage Area	0.017533
Bsmt Full Bath	0.017099
Year Remod/Add	0.015338
Fireplaces	0.015169
Neighborhood_Crawfor	0.014752
Screen Porch	0.014571
Paved Drive	0.014064
Heating QC	0.013994
Lot Frontage	0.013922
Kitchen Qual	0.012454
Bsmt Exposure	0.012436
Condition 1_Norm	0.011946
Garage Cars	0.011103

FINAL MODEL!

- param_gs22 = {
- 'alpha': [.01,.1,.25,.3,0.31,.35,.5,1],
- 'l1_ratio': np.arange(0.0, 1.0, 0.1)
- }

- gs22 = GridSearchCV(ElasticNet(), param_gs22 , cv=5, n_jobs=-1)

- gs22.fit(X_train_sc,y_train_lg)

Predictions vs. actual sale price

