Predicting Housing Prices in Ames, Iowa

Presentation by Joey Navarro

Photo by Tom Rumble on Unsplash

TABLE OF CONTENTS

01 Executive Summary

02 Key Decisions Made

03 Attribute Relationship to Sale Price

04 Model Results

05 Conclusion & Next Steps

01

Executive Summary

Photo by Hunters Race on Unsplash

Executive Summary

This project was was to predict housing prices for Ames, Iowa, given a dataset from the **Ames, Iowa Assessor's Office** containing information on houses sold between **2006 and 2010**.

My regression analysis will, at its core, will examine the influence of 80 given home features (**independent variables**) on the home sale price (**dependent variable**).

This model would benefit a company or an individual who wants to know the value of a house based on that home's at tributes.

Photos from City of Ames and Google search

Why and Who Cares?

Brokers looking to **increased sales** through accurate housing market price predictions.

Buyers looking for a **home within their price range** based off of features wanted.

Photo by Javier Allegue Barros on Unsplash

Key Decisions Made

02

Key Decisions Made

Model Comparison

Data Analysis Work	Features Model 1	Features Model 2
Null Values	Replaced with 0	Replaced with 0
Outliers	Ground Living Area < 3700 Lot Area < 47000	Ground Living Area < 3700 Lot Area < 47000
Columns	Kept 21 feature columns to decided by Sale Price hypothesis test with p-value < 0.05	Kept only numeric feature columns and added dummy columns for categorical features
Normalizing Features	Logarithmic transformation of first floor square foot feature column Logarithmic transformation of Sale Price column	Logarithmic transformation of first floor square foot feature column
Transforming Features	Used StandardScaler()	Used StandardScaler() and PolynomialFeatures()
Model Pipeline	Did Not Use	Did use

03

Feature Relationship to Price

Photo by vivek kumar on Unsplash

Top 10 Features for Features Model 1

Attribute	coefficient
Fireplaces	5932.3
Total Rooms Above Ground	5210.1
Basement Full Bath	3293.1
Car Garage	523.1
Year Remodeled	501.2
Year Built	459.7
Month Sold	169.0
Ground Living Area	150.78
Screen Porch	58.05
Square Footage of Finished Basement	23.3

Top 10 Features for Features Model 2

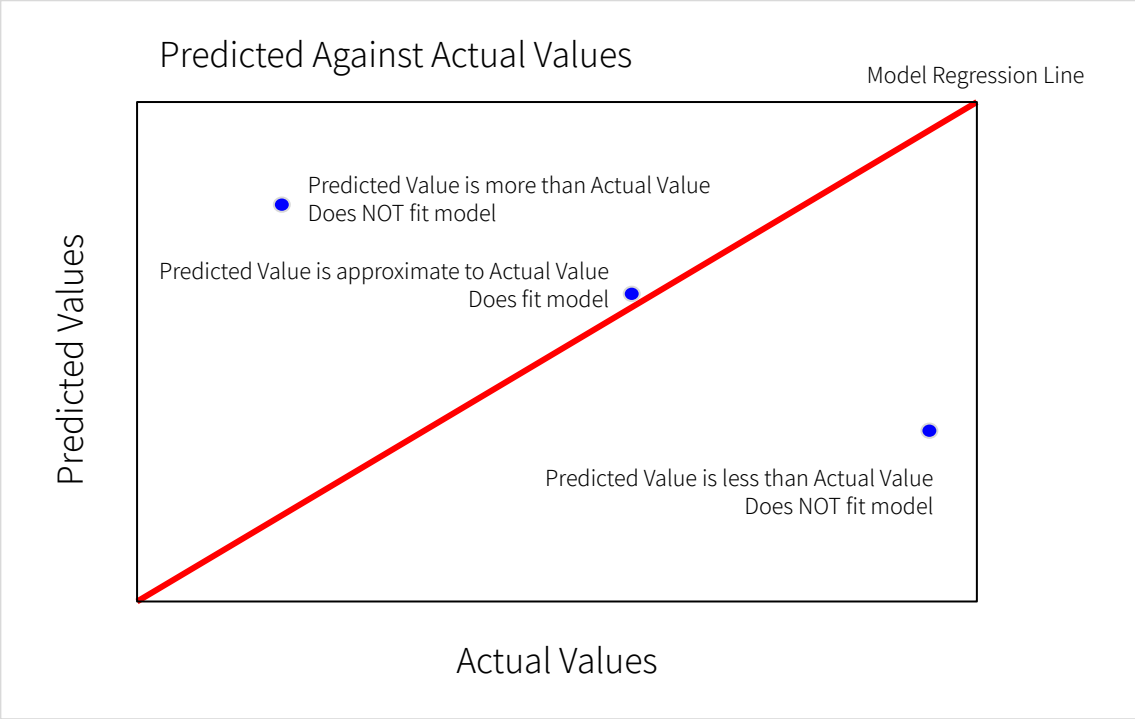
Attribute	Coefficient
Total Basement Square Footage * Ground Living Area	11376.60
Overall Condition * Ground Living Area	10336.17
Year Built * Year Remodeled / Added	7578.17
Year Built * Year Built	6434.28
Total Basement Square Footage * Overall Quality 9	5650.46
Ground Living Area * Garage Area	5248.69
Lot Area * Overall Quality 8	4578.47
Total Basement Square Footage * Overall Quality 10	4240.92
Sale Type New * Neighborhood Stone Br	4063.99
Fireplaces * Overall Quality 9	3313.03

Photo by Annie Spratt on Unsplash

Model Results

04

Quick Note on Regression Visualization Interpretation

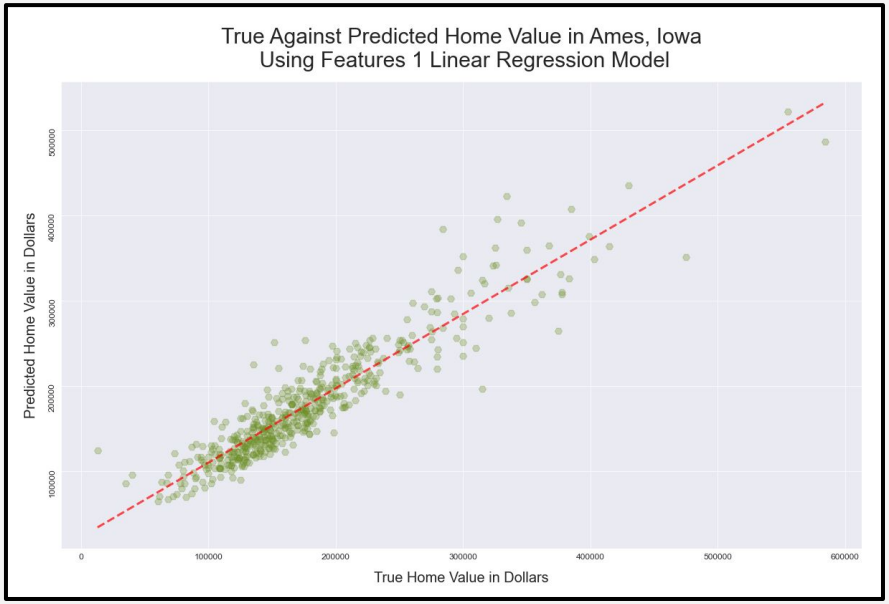


My regression analysis will, at its core, will examine the influence home features (**independent variables**) on the home sale price (**dependent variable**).

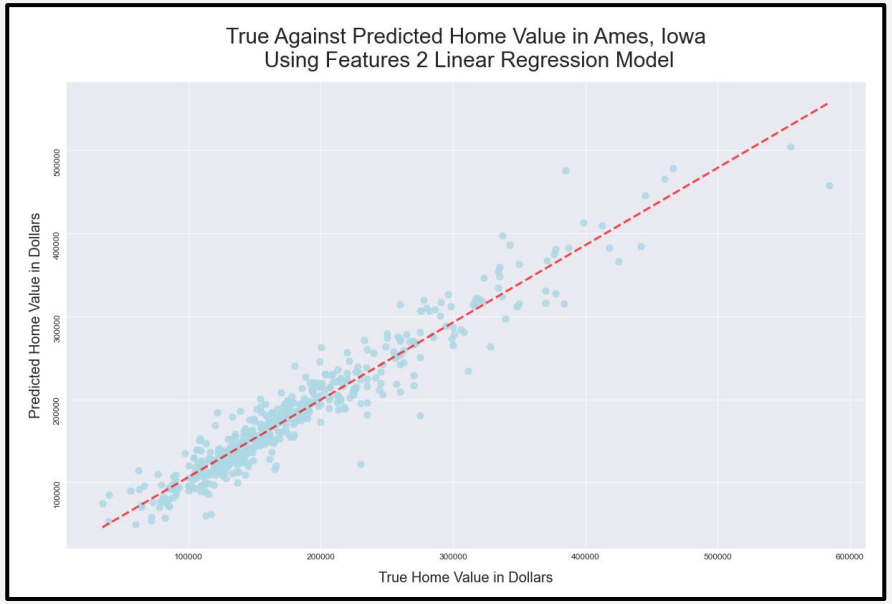
Quick Note on Model Statistics Interpretation

R2	R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s); **accuracy**. **100% means that** all movements of the dependent variable (**Features selected**) are **completely explained by** movements in the independent variable (**Sale Price**).
MSE	Describes **how precise** the mean of the sample is as an estimate of the true mean of the population. The **lower the value the better** and **0 means the model is perfect**.
RMSE	Is a measure of how spread out the prediction errors are. In other words, it tells you how concentrated the data is around the line of best fit. **Lower values of indicate better fit**.

Linear Regression Model Results

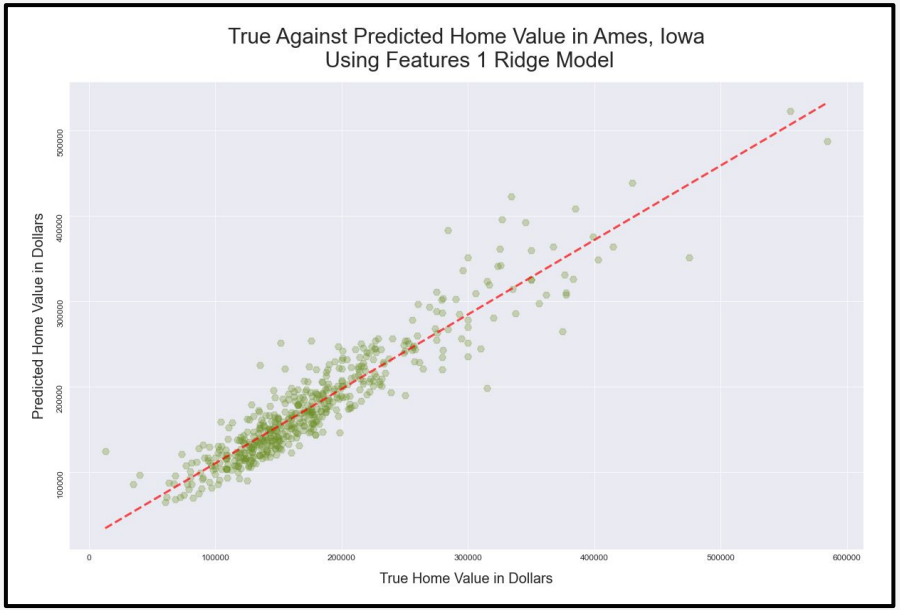


R2	87 %
MSE	686366622.78
RMSE	26198.6

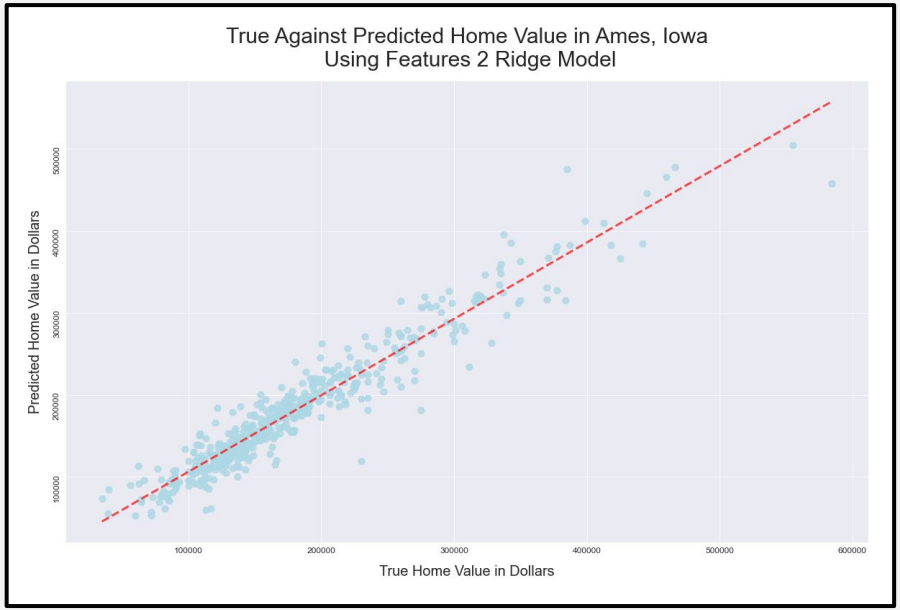


R2	92 %
MSE	499113738.49
RMSE	22340.85

Regression Model Using Ridge Regularization Results

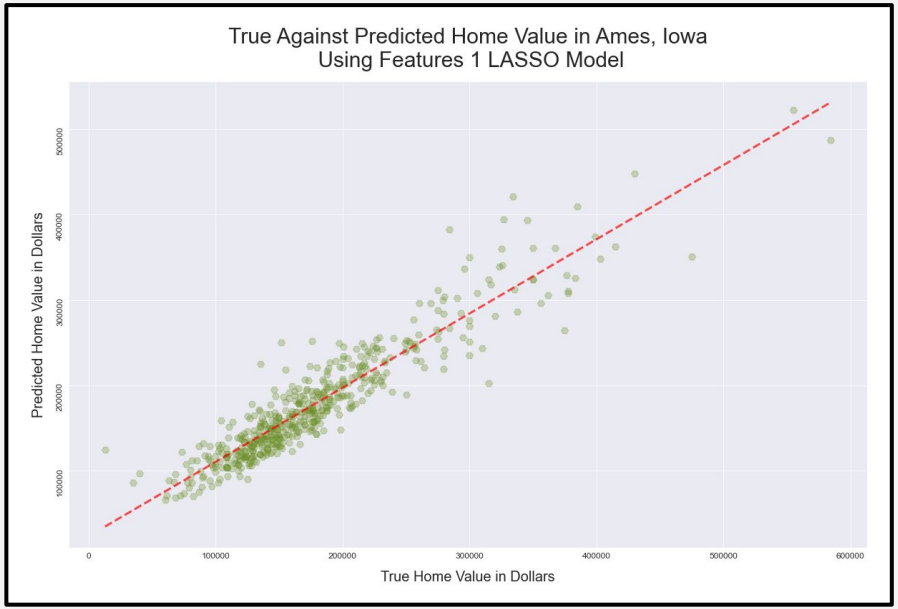


R2	87 %
MSE	684714215.79
RMSE	26167.04

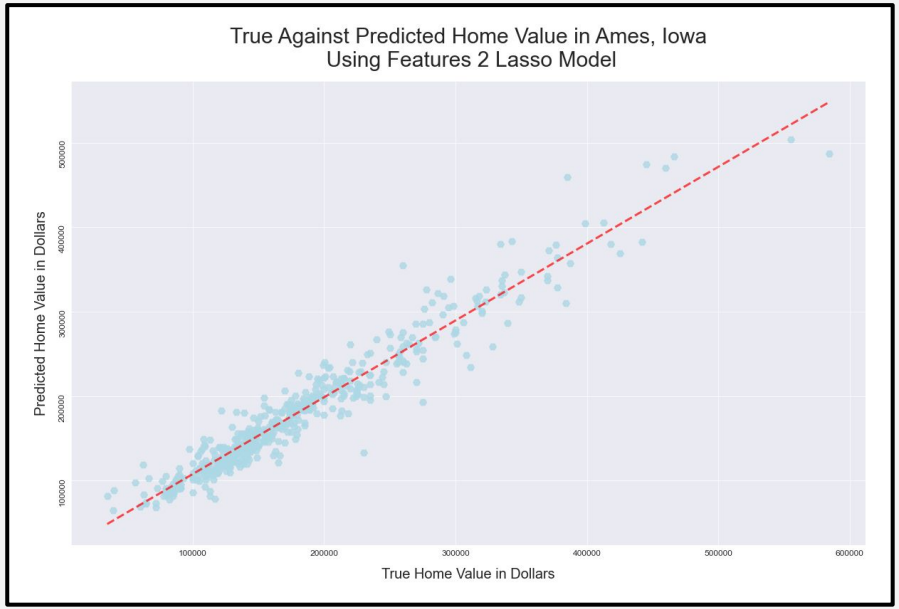


R2	92 %
MSE	499301271.01
RMSE	22345.05

Regression Model Using LASSO Regularization Results



R2	87 %
MSE	682891161.14
RMSE	26132.17



R2	93 %
MSE	435141709.78
RMSE	20860.05

Conclusions & Next Steps

Using a linear regression with LASSO regularization along with a few data transformations, removal of extreme outliers, and the creation of dummy variables allowed me to fit features model 2 that performed with an R2 value of 93 %. The next steps are to attempt to predict the exact prices, I may want to try more advanced modeling methods or use other classifying methods for variable selection.