Pricing aims for Ames

A MODEL TO BEST PREDICT HOME VALUE IN AMES, IOWA

One problem, one solution

We are looking to best predict home values in Ames, Iowa

- This information is beneficial to both buyers and sellers

The information gained in this home modeling process tells us more than just price

- What can we do to INCREASE sale price? Stay tuned.

We need a way to do it. I created it.

- Given the high linear correlation of many house attributes, we can predict home value using linear regression

A note on our data set

- 2051 entries of home sales in Ames, Iowa
- 81 home features including information on space, condition, quality, location and other attributes
- Sales ranging from 2006 until 2010
- Data set is available through link on my project readme

First data problems and their solutions

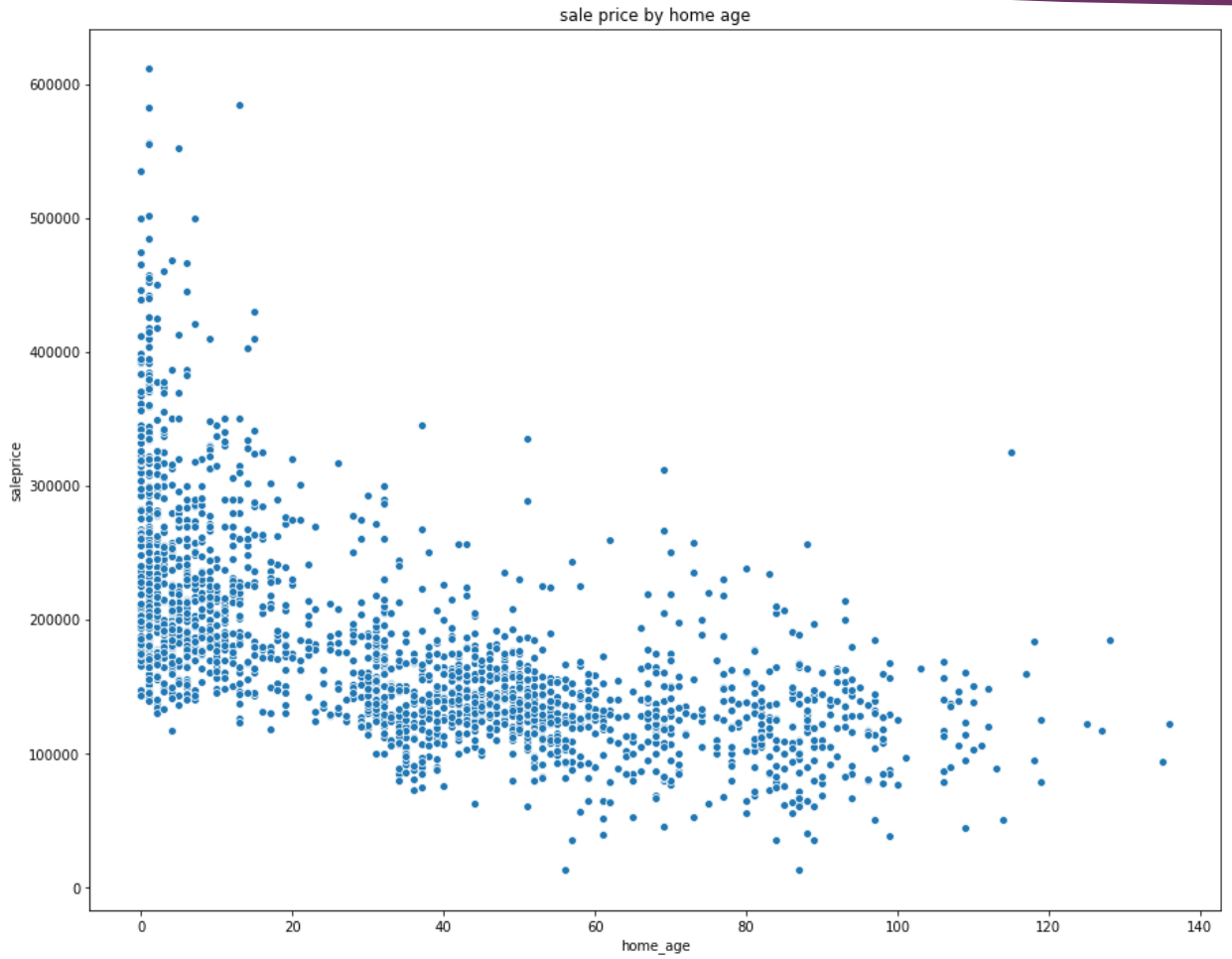
- Lots of null values, most indicating the absence of things such as fences, basements, etc.
- Categorical variables converted to dummy variables, null values dropped if indicating absence of something, if not, first categories dropped

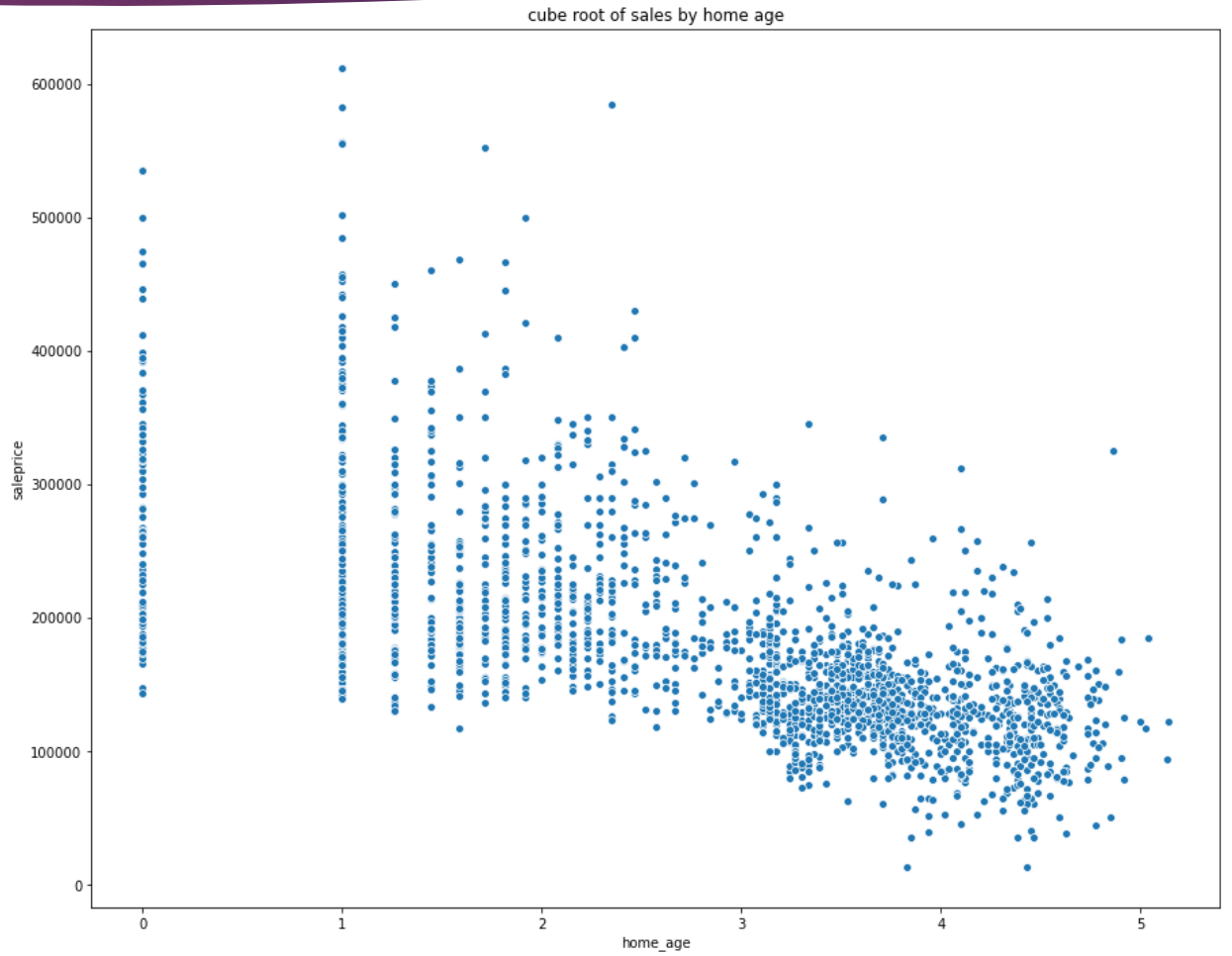
	Id	PID	MS SubClass	MS Zoning	Lot Frontage	Lot Area	Street	Alley	Lot Shape	Land Contour	...	Screen Porch	Pool Area	Pool QC	Fence	Misc Feature	Misc Val	Mo Sold	Yr Sold	Sale Type
0	109	533352170	60	RL	NaN	13517	Pave	NaN	IR1	Lvl	...	0	0	NaN	NaN	NaN	0	3	2010	WD
1	544	531379050	60	RL	43.0	11492	Pave	NaN	IR1	Lvl	...	0	0	NaN	NaN	NaN	0	4	2009	WD
2	153	535304180	20	RL	68.0	7922	Pave	NaN	Reg	Lvl	...	0	0	NaN	NaN	NaN	0	1	2010	WD
3	318	916386060	60	RL	73.0	9802	Pave	NaN	Reg	Lvl	...	0	0	NaN	NaN	NaN	0	4	2010	WD
4	255	906425045	50	RL	82.0	14235	Pave	NaN	IR1	Lvl	...	0	0	NaN	NaN	NaN	0	3	2010	WD

Selecting the best features, engineering the rest

- Created polynomial features and interaction terms, new features to describe combined area
  - Total bathrooms, total bathrooms adjusted, inside and outside square feet
- The combined features I created did not have as much predictive power as the individual, more specific variables
- Cube root home age only one with enough predictive power to keep

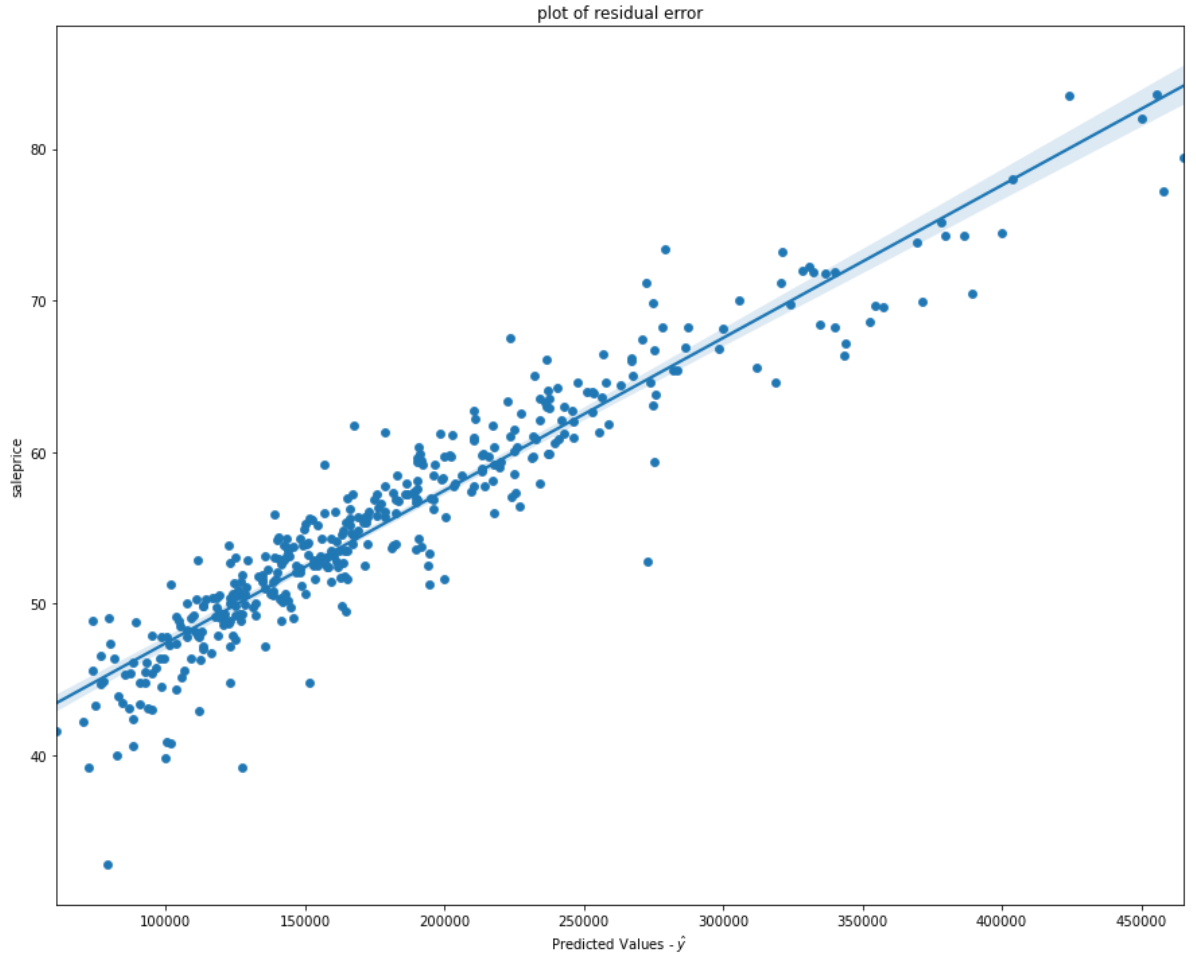
Cube root transformation





What my model is based on

- Total basement square footage
- Above ground square footage
- Kitchen quality
- Exterior quality
- Neighborhood
- Basement quality
- Foundation type
- Basement finish state
- Garage type
- Garage area
- Fireplaces
- Cube root home age



About the final model

- Ridge CV regularized model
- Alpha = 1
- Train RMSE: $21,506.73
- Test RMSE: $26,343.08
- Train r^2: 0.904
- Test r^2: 0.901

Location, location, location…and quality

Top positive coefficients:

Green Hill Neighborhood: **7.183**

Kitchen Quality 5/5: **5.051**

Exterior Quality 5/5: **4.680**

Stonebrook Neighborhood: **3.194**

Top negative coefficients:

Meadowvale Neighborhood: -**3.308**

Iowa DOT and Rail Road: -**1.846**

Briardale Neighborhood: **-1.419**

2-type Garage: **-1.347**

MLR model comparison to RidgeCV

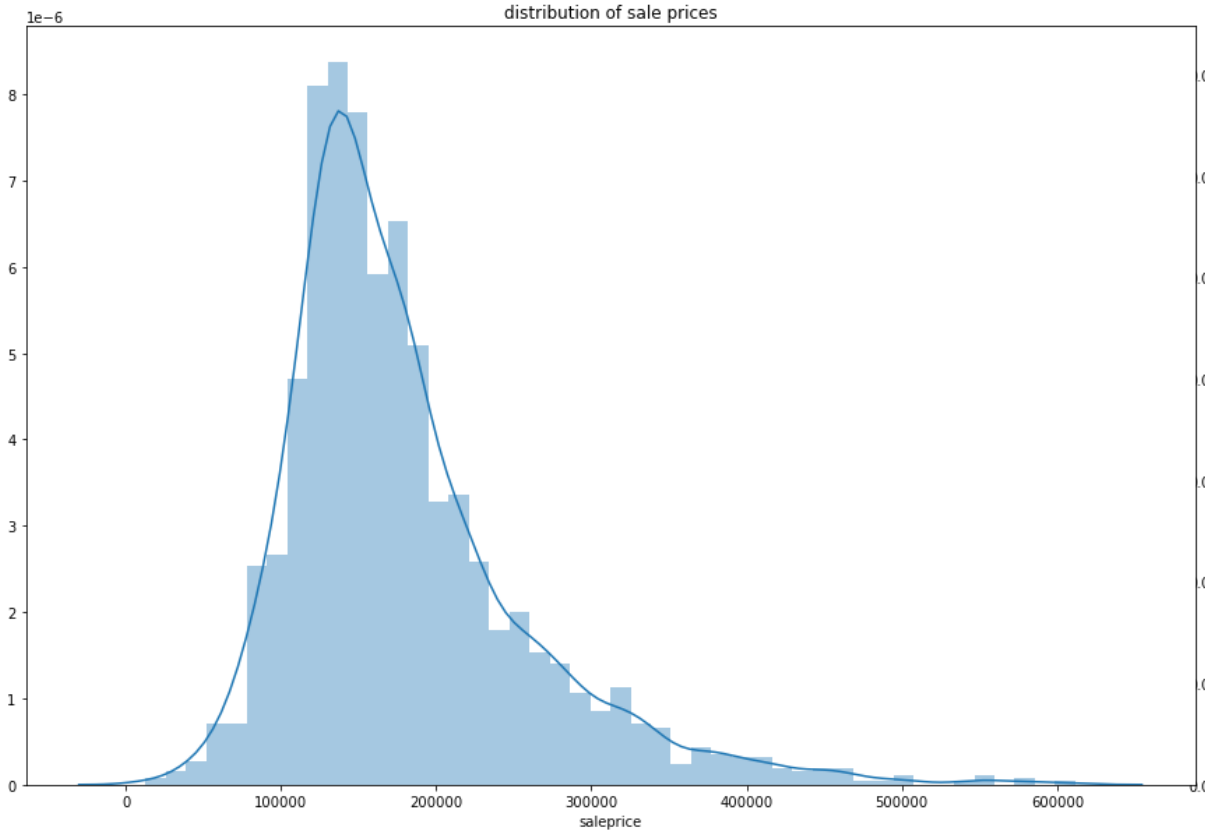
Top positive coefficients:

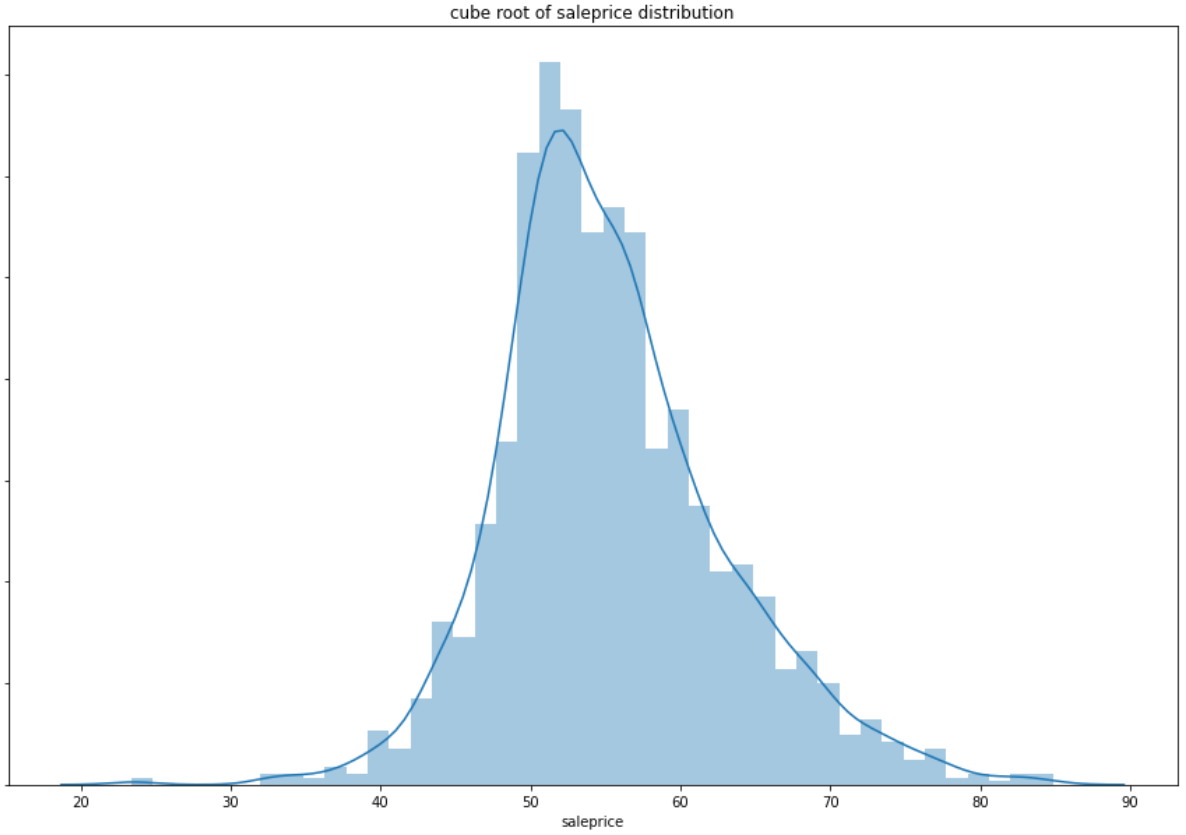
Green Hill Neighborhood: $125,520.43	Median home value: $280,000 (2 entries)
Kitchen Quality 5/5: $48,610.25	
Exterior Quality 5/5: $51,614.45	
Stonebrook Neighborhood: $62,366.19	Median home value: $322,450

Top negative coefficients:

Meadowvale Neighborhood: -$8,506.18	Median home value: $92,450
Iowa DOT and Rail Road: $6,449.16	Median home value: $101,500
Briardale Neighborhood: $10,037.92	Median home value: $105,500
2-type Garage: -$20,322.51	

Sale price distribution and cube root transform





The effects of RidgeCV model with cube root transformation

Multiple linear regression model without ridge or cube root transformation

- Train RMSE: $23,224.33
- Test RMSE: $28,599.53
- Train r^2 score: 0.906
- Test r^2 score: 0.888

RidgeCV model with root transformation results:

- Train RMSE: $21,506.73
- Test RMSE: $26,343.08
- Train r^2 score: 0.904
- Test r^2 score: 0.901

Increase the value of your home

We know from the data that a few things negatively affect home value in Ames, Iowa:

- Home age and how long since a remodel—so remodel!
- Poor kitchen and exterior quality
- No masonry veneer

To increase price for sale, remodel now and focus on the areas associated with highest value increase per square foot:

- Upgrade kitchen, outside of home, and above ground living area
- Consider adding stone veneer for added value

Strengths of the model

- 90% of variability explained in both training and testing sets
- No real overfitting: model should respond well when given housing prices from 2006-2010 in Ames, Iowa.
- Relatively little noise even before regularization
- Model best at predicting home values without extremely high or low values.

Room for improvement? Always.

- Model does a great job for the time period, but that was a decade ago.
- Evident from residuals there are likely more outliers to account for.
- Typically seen error off by as much as $26,000 give or take (16% of median home value in Ames, Iowa)
- Account for home age in a different way?
- Area of exploration: what is the relationship between a two-type garage and lower home value?
- Further research the Green Hill neighborhood. Two entries may not give us the full picture of that neighborhood.


FOR
SOLD