ANALYZING THE AMES HOUSING MARKET

Problem Statement

Create and curate a set of synergistic features that gives insight in how to effectively increase a house's value in Ames, Iowa

Linear Regression Modeling

Optimize for R^2 Score, but watch squared errors to find best fitting variables

Throw out anything that negatively impacts score no matter traditional advice

One Percent = Thousands of dollars

Method One

Bare Bones

No data transformation

Only use features ready for modeling in original data

Look at multiple modeling methods (OLS, Ridge, LASSO)

Method One ⟶ Method Two

Bare Bones

No data transformation

Only use features ready for modeling in original data

Look at multiple modeling methods (OLS, Ridge, LASSO)

Kitchen Sink

Transform Sale Price to normal distribution and clean up outliers

Create as many relevant features possible and use all with relevant correlation

Look at multiple modeling methods (OLS, Ridge, LASSO)

Method One ⟶ Method Two ⟶ Method ∞

Bare Bones

No data transformation

Only use features ready for modeling in original data

Look at multiple modeling methods (OLS, Ridge, LASSO)

Kitchen Sink

Transform Sale Price to normal distribution and clean up outliers

Create as many relevant features possible and use all with relevant correlation

Look at multiple modeling methods (OLS, Ridge, LASSO)

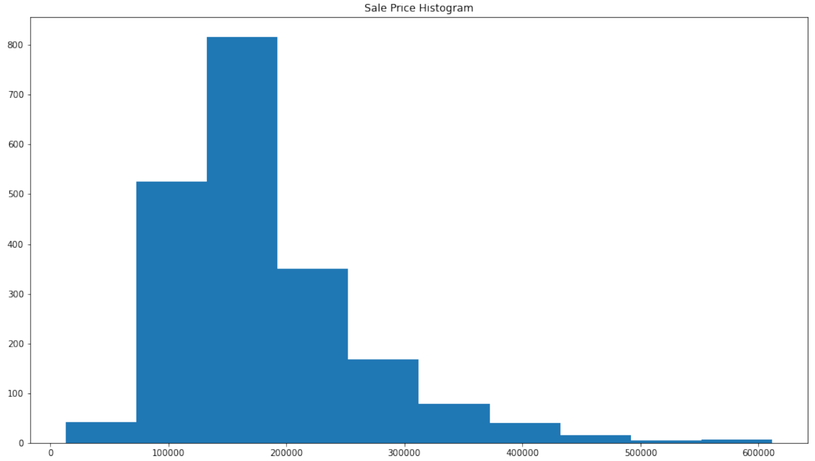
Iterate, Iterate, Iterate

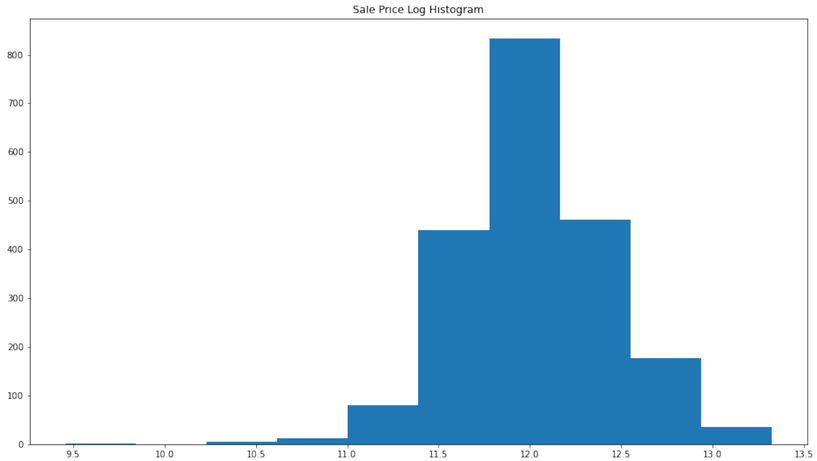
Compare untouched vs transformed and clean up outliers

Create a feature list that optimizes R^2 Score

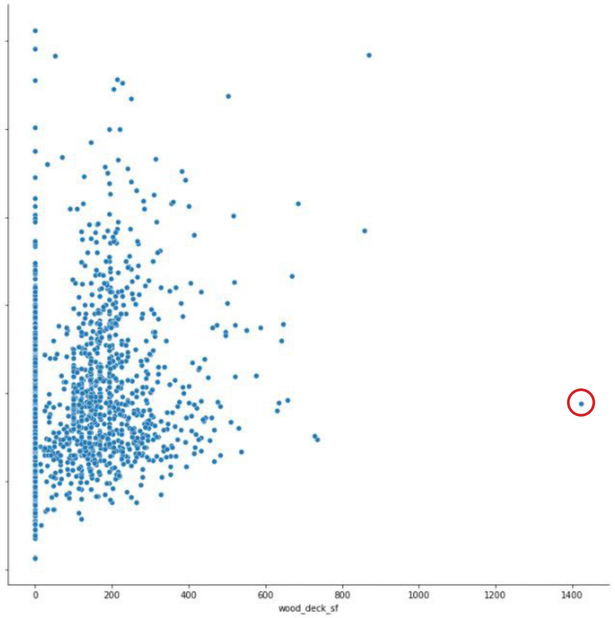
Look at multiple modeling methods (OLS, Ridge, LASSO)

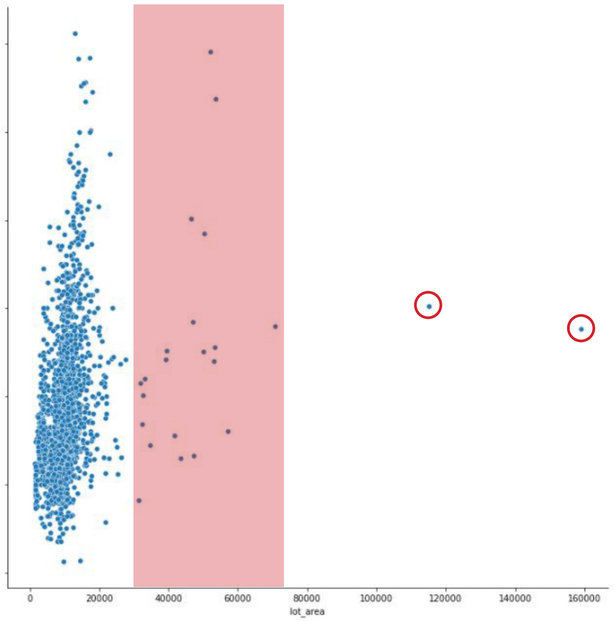
Log Transform



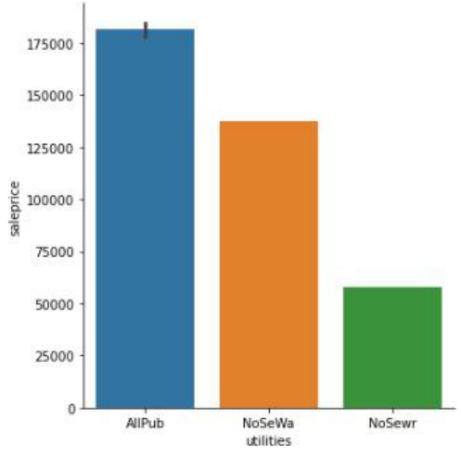


Scrub Outliers



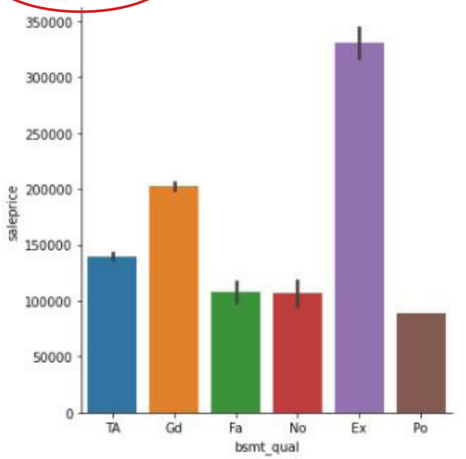


New Features




```
df_train['utilities'].value_counts()
```

```
AllPub    2047
NoSewr       1
NoSeWa       1
```




```
df_train['bsmt_qual'].value_counts()
```

```
TA    887
Gd    864
Ex    182
Fa     60
No     55
Po      1
```

Feature Methodology

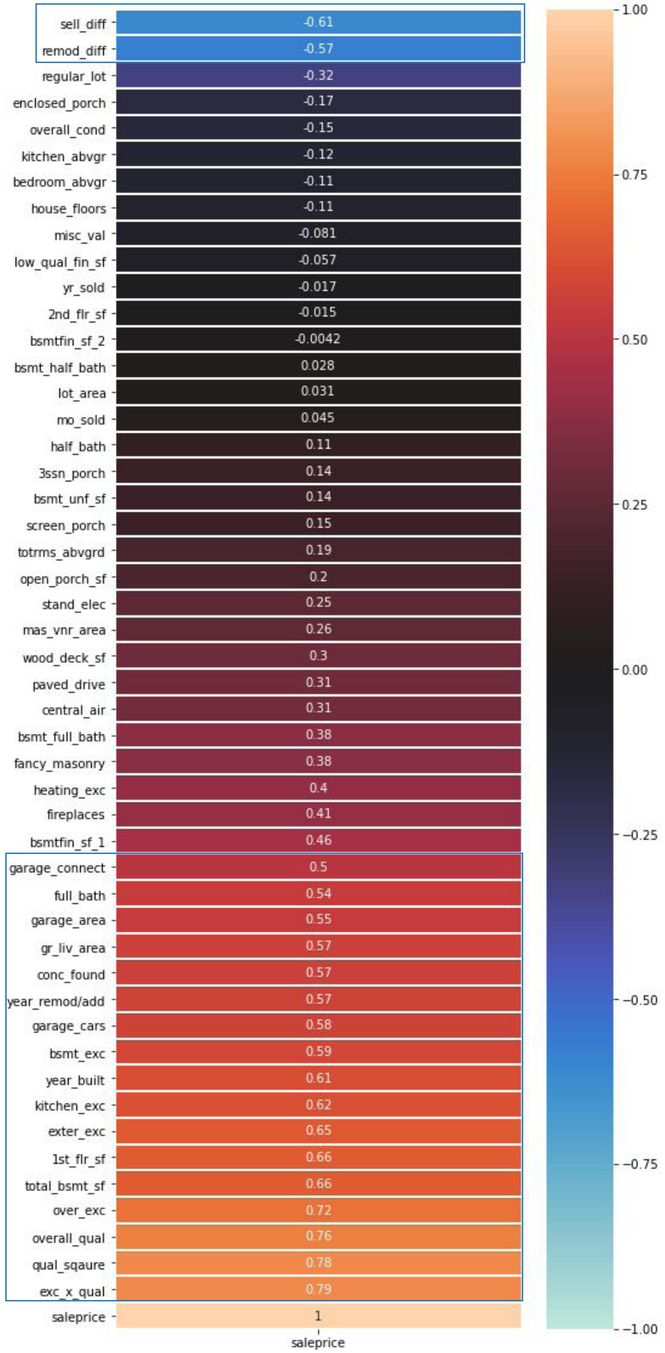
Absolute correlation value of 0.5 or higher

At least $5,000 differential between values

No value greater than 90% of total value

Forgettable Features: Regular lot?, Fancy masonry?, Paved driveway?,
Heating exceed average?, Central air? Standard heating?

Bad Feature: Amount of floors



Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method One ⟶ Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

Year Built
Year Last Remodeled/Rebuilt
~~Overall Quality~~
Above Grade SF
Above Grade Rooms
Garage SF
~~Garage Capacity~~
Basement SF
First Floor SF
Number of Bathrooms
~~Masonry Veneer SF~~
~~House Age at Sell~~
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
~~Kitchen Exceed Average?~~
~~Exterior Exceed Average?~~
Features over Average
(Features over Average)*(Overall Quality)

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

Year Built
Year Last Remodeled/Rebuilt
~~Overall Quality~~
~~Above Grade SF~~
~~Above Grade Rooms~~
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
~~Time Since Last Remodel~~
~~Neighborhood~~
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
~~Exterior Exceed Average?~~
Features over Average
~~(Features over Average)*(Overall Quality)~~

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

Year Built
~~Year Last Remodeled/Rebuilt~~
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
~~First Floor SF~~
~~Number of Bathrooms~~
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
~~Neighborhood~~
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
~~Features over Average~~
(Features over Average)*(Overall Quality)

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

Year Built
~~Year Last Remodeled/Rebuilt~~
~~Overall Quality~~
~~Above Grade SF~~
~~Above Grade Rooms~~
Garage SF
Garage Capacity
Basement SF
~~First Floor SF~~
~~Number of Bathrooms~~
~~Masonry Veneer SF~~
~~House Age at Sell~~
~~Time Since Last Remodel~~
~~Neighborhood~~
Garage Connected to House?
Concrete Foundation?
~~Basement Exceed Average?~~
~~Kitchen Exceed Average?~~
~~Exterior Exceed Average?~~
~~Features over Average~~
~~(Features over Average)*(Overall Quality)~~

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

~~Year Built~~
~~Year Last Remodeled/Rebuilt~~
~~Overall Quality~~
~~Above Grade SF~~
~~Above Grade Rooms~~
~~Garage SF~~
~~Garage Capacity~~
Basement SF
First Floor SF
~~Number of Bathrooms~~
~~Masonry Veneer SF~~
~~House Age at Sell~~
Time Since Last Remodel
~~Neighborhood~~
Garage Connected to House?
~~Concrete Foundation?~~
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

~~Year Built~~
~~Year Last Remodeled/Rebuilt~~
~~Overall Quality~~
Above Grade SF
Above Grade Rooms
Garage SF
~~Garage Capacity~~
Basement SF
First Floor SF
~~Number of Bathrooms~~
~~Masonry Veneer SF~~
~~House Age at Sell~~
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method One

Method Two

Method ∞

	Ridge Scores
Cross Variable R^2	0.824
Mean Squared Error	546,945,723
Root Mean Squared Error	23,386

Method One

	Ridge Scores
Cross Variable R^2	0.824
Mean Squared Error	546,945,723
Root Mean Squared Error	23,386

Method Two

	Ridge Scores
Cross Variable R^2	0.860
Mean Squared Error	458,006,241
Root Mean Squared Error	21,401

Method ∞

Method One

	Ridge Scores
Cross Variable R^2	0.824
Mean Squared Error	546,945,723
Root Mean Squared Error	23,386

Method Two

	Ridge Scores
Cross Variable R^2	0.860
Mean Squared Error	458,006,241
Root Mean Squared Error	21,401

Method ∞

	OLS Scores
Cross Variable R^2	0.873
Mean Squared Error	333,578,729
Root Mean Squared Error	18,264

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

~~Year Built~~
~~Year Last Remodeled/Rebuilt~~
~~Overall Quality~~
Above Grade SF
Above Grade Rooms
Garage SF
~~Garage Capacity~~
Basement SF
First Floor SF
~~Number of Bathrooms~~
~~Masonry Veneer SF~~
~~House Age at Sell~~
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method One

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF

	Coefficient	Effectiveness
Kitchen	8,111	11.73x
Basement	691	1.00x
Exterior	-950	-1.37x

Method Two

Year Built
Year Last Remodeled/Rebuilt
Overall Quality
Above Grade SF
Above Grade Rooms
Garage SF
Garage Capacity
Basement SF
First Floor SF
Number of Bathrooms
Masonry Veneer SF
House Age at Sell
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

Method ∞

~~Year Built~~
~~Year Last Remodeled/Rebuilt~~
~~Overall Quality~~
Above Grade SF
Above Grade Rooms
Garage SF
~~Garage Capacity~~
Basement SF
First Floor SF
~~Number of Bathrooms~~
~~Masonry Veneer SF~~
~~House Age at Sell~~
Time Since Last Remodel
Neighborhood
Garage Connected to House?
Concrete Foundation?
Basement Exceed Average?
Kitchen Exceed Average?
Exterior Exceed Average?
Features over Average
(Features over Average)*(Overall Quality)

In Conclusion

Marked improvement when transforming the data

Binary variables with skewed splits (90/10) less effective for modeling

Remodeling efforts have great impact

Still heavily reliant on experts' opinion

Who sold the house?, Staging?, Time on the market?

Find out subjective scoring methodology

Are there better evaluators out there?

Fit more house area imporovements to find out best areas to improve