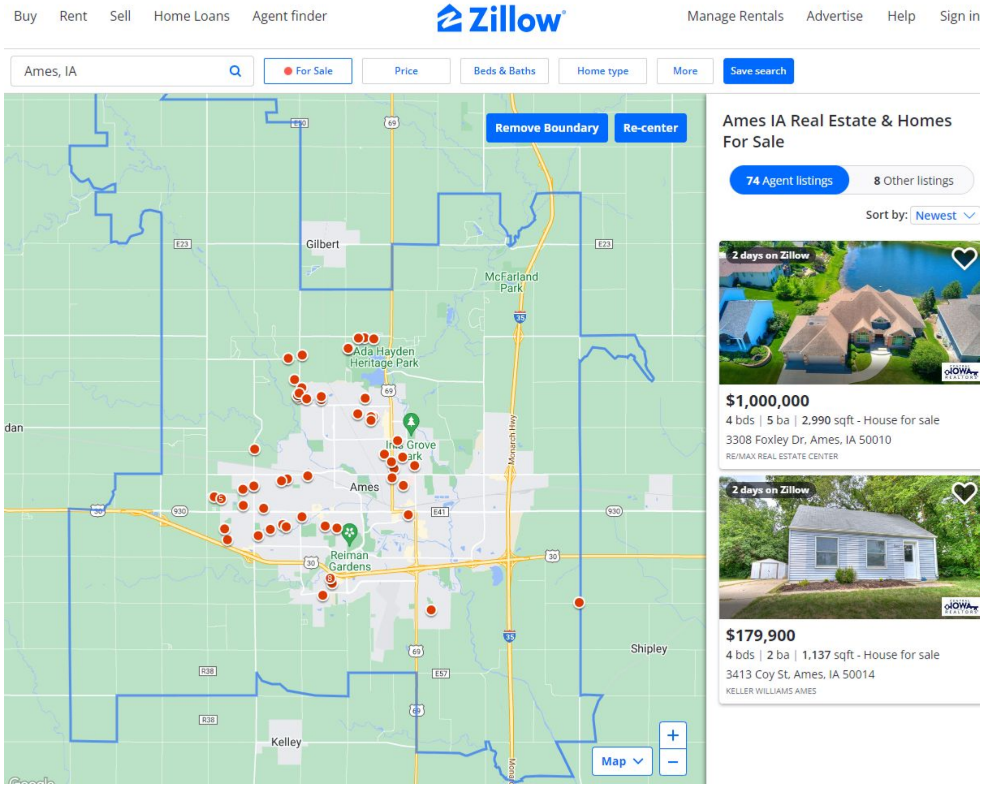
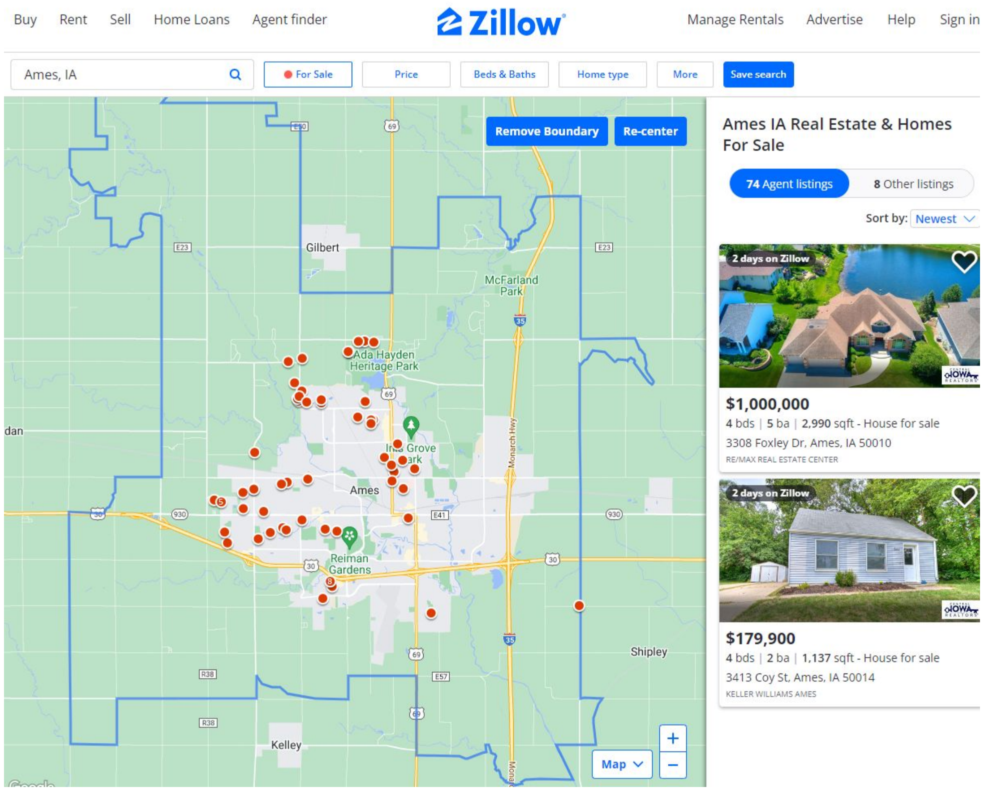
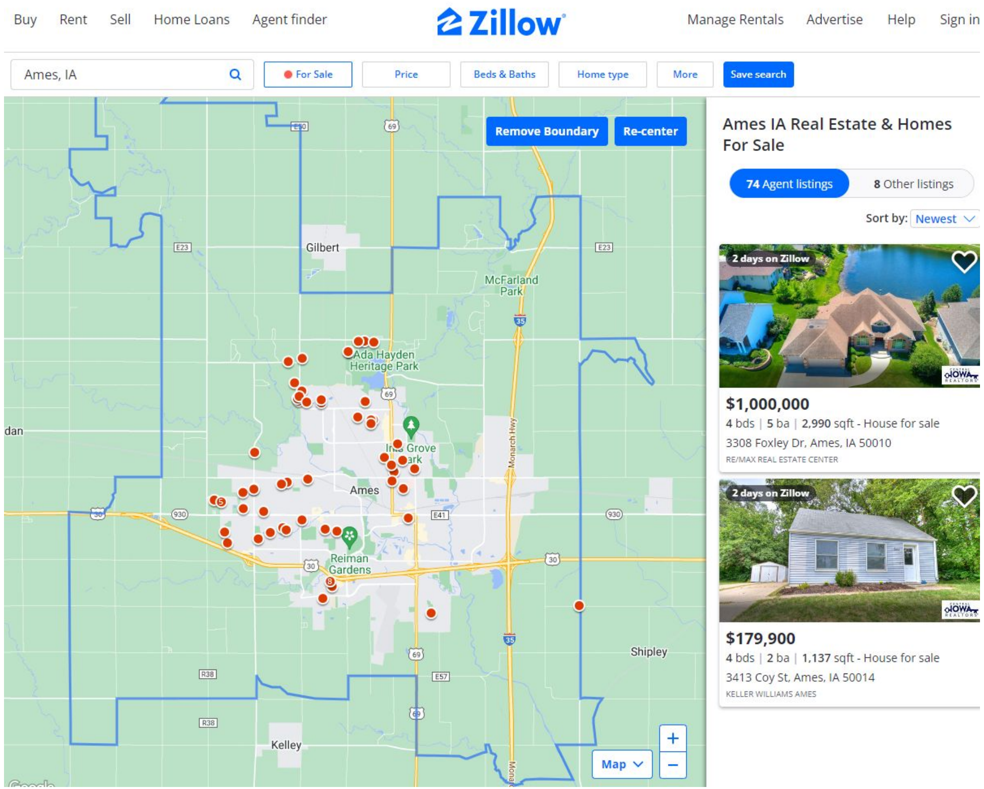
Predicting Home Prices in Ames, Iowa

Zillow

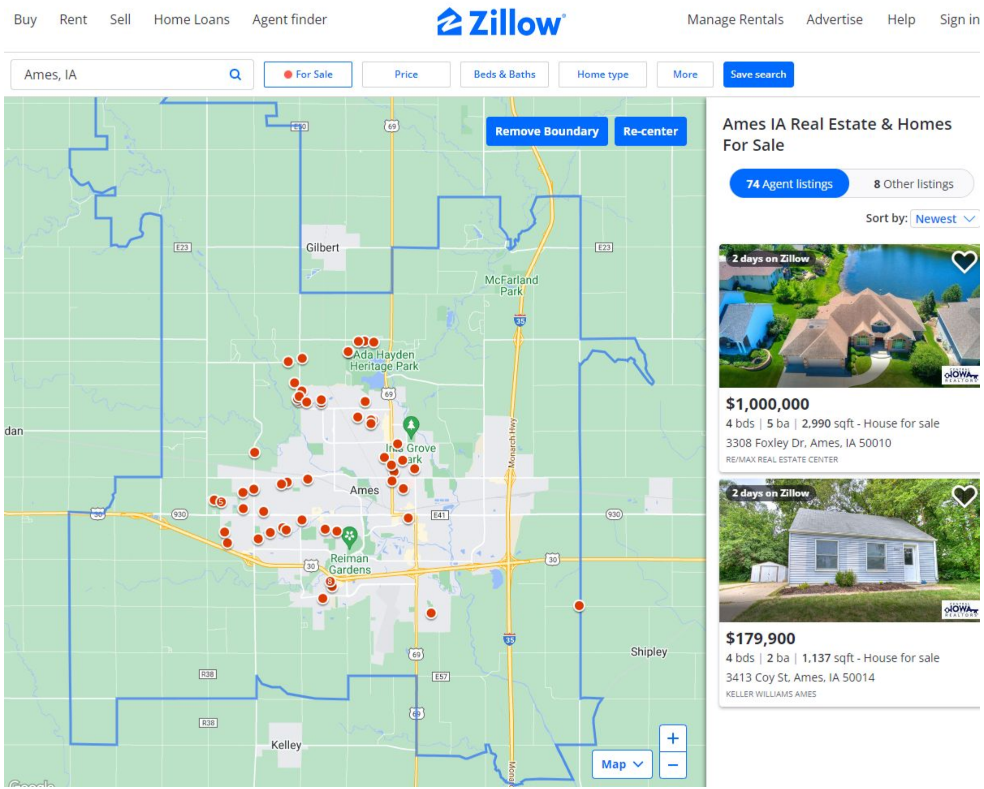

Buy
Rent
Sell
Home Loans
Agent finder
Zillow
Manage Rentals
Advertise
Help
Sign in
Ames, IA
For Sale
Price
Beds & Baths
Home type
More
Save search
Remove Boundary
Re-center
Gilbert
McFarland Park
Ada Hayden Heritage Park
Ames
Reiman Gardens
Shipley
Kelley
Map
Ames IA Real Estate & Homes For Sale
74 Agent listings
8 Other listings
Sort by: Newest
2 days on Zillow
$1,000,000
4 bds | 5 ba | 2,990 sqft - House for sale
3308 Foxley Dr, Ames, IA 50010
RE/MAX REAL ESTATE CENTER
2 days on Zillow
$179,900
4 bds | 2 ba | 1,137 sqft - House for sale
3413 Coy St, Ames, IA 50014
KELLER WILLIAMS AMES



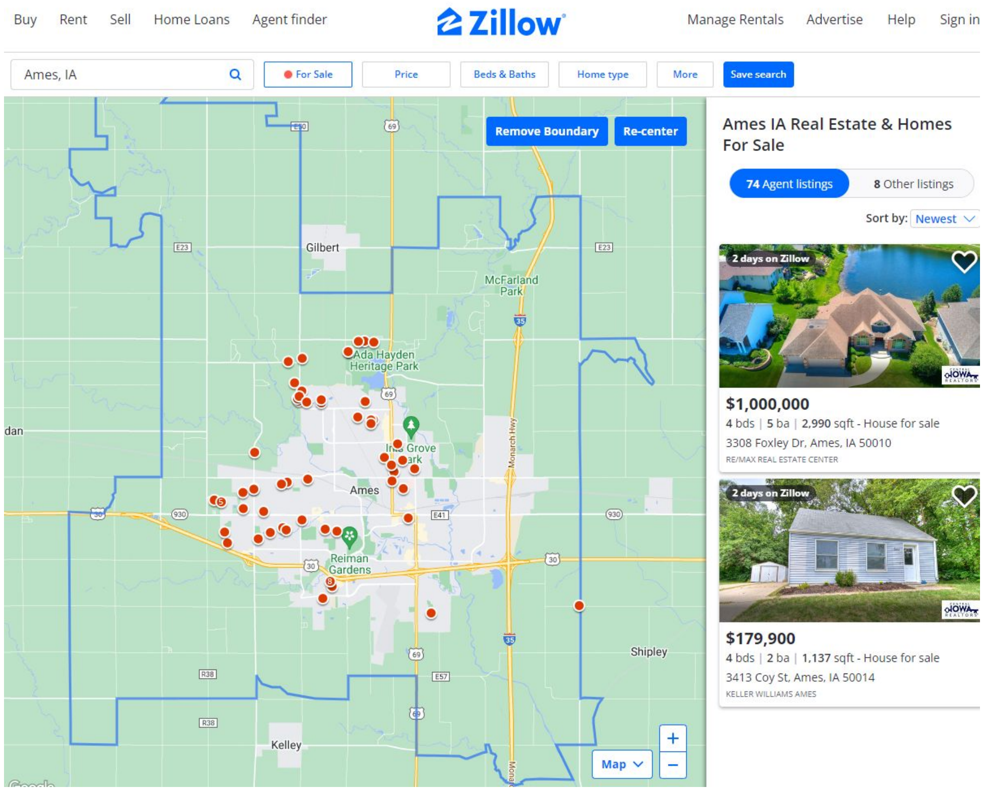
- Advertise home and rental units



- Advertise home and rental units
- Ads serve as primary source of revenue



- Advertise home and rental units
- Ads serve as primary source of revenue
- Widely known for sharing estimated value for many homes



- Advertise home and rental units
- Ads serve as primary source of revenue
- Widely known for sharing estimated value for many homes - **FREE**

Before we begin…

The **more reliable our predictions are** the:

Before we begin...

The **more reliable our predictions are** the:

- better our brand is

Before we begin...

The **more reliable our predictions are** the:

- better our brand is

- More people will trust our service

Before we begin...

The **more reliable our predictions are** the:

- better our brand is
- More people will trust our service
- **More people will be exposed to and interested in our advertisement service**

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

- Square feet

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

- Square feet
- Number of bathrooms

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

- Square feet
- Number of bathrooms
- Neighborhood

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

- Square feet
- Number of bathrooms
- Neighborhood
- Year built

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

- Square feet
- Number of bathrooms
- Neighborhood
- Year built
- Overall quality and condition

Our Goal

Build a linear regression model that can estimate the price of a home in Ames, Iowa from information like:

- Square feet
- Number of bathrooms
- Neighborhood
- Year built
- Overall quality and condition
- etc

The Process

- Clean our data

The Process

- Clean our data
- Build new data from old

The Process

- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**

The Process

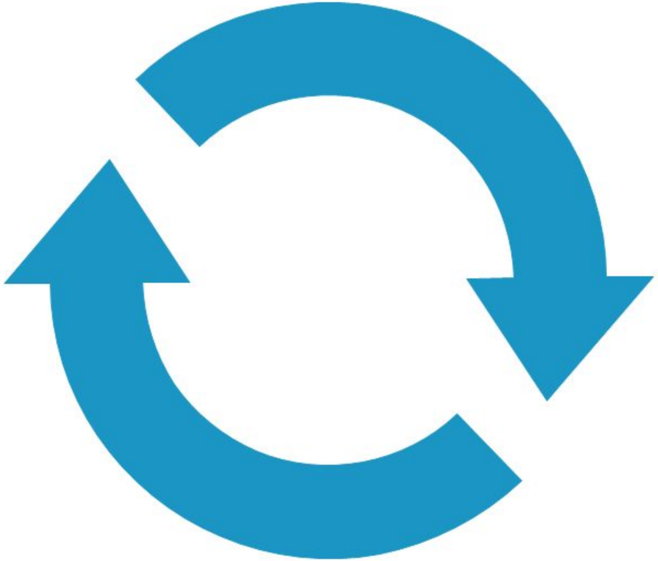
- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**
- Build the model

The Process

- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**
- Build the model
- Score the model

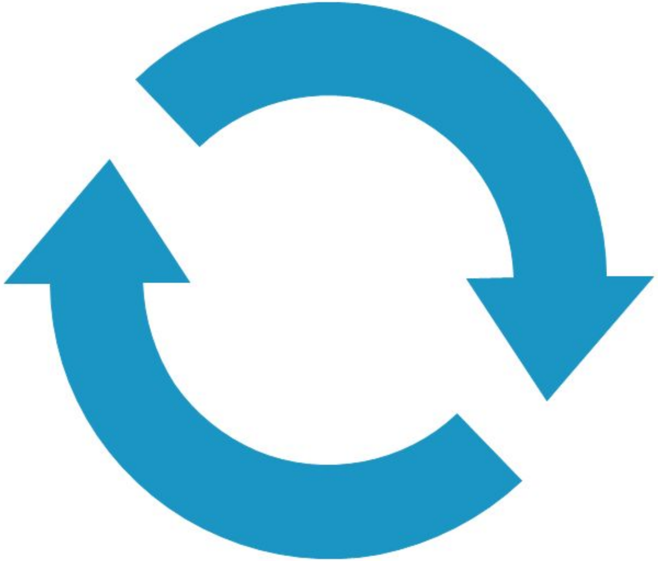
The Process

- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**
- Build the model
- Score the model
- Choose new **features**



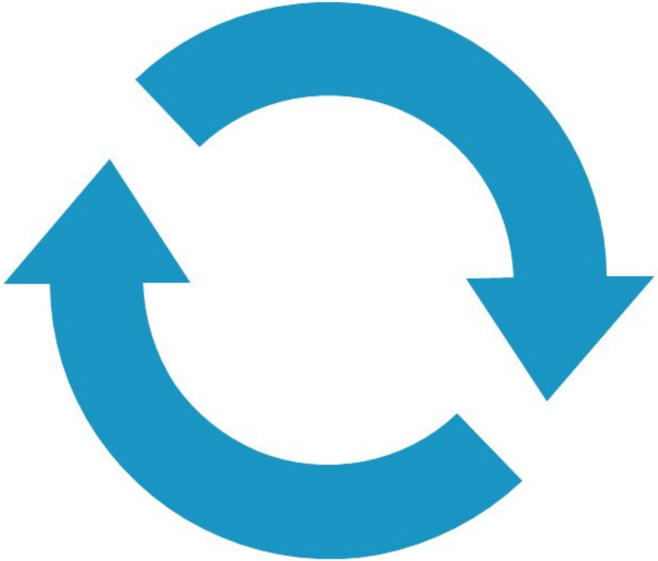
The Process

- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**
- Build the model
- Score the model
- Choose new **features**
- Build the model



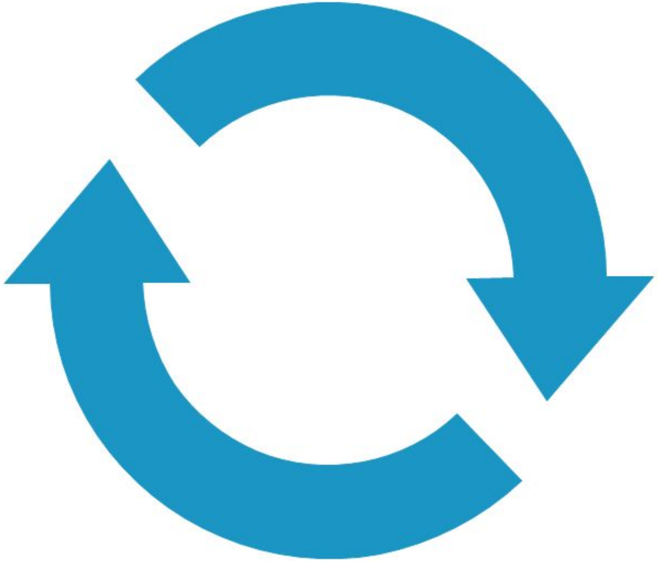
The Process

- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**
- Build the model
- Score the model
- Choose new **features**
- Build the model
- Score the model…



The Process

- Clean our data
- Build new data from old
- Choose which house information (**features**) we think is most relevant to predicting **sale price**
- Build the model
- Score the model
- Choose new **features**
- Build the model
- Score the model…



Assessing Our Model

One common metric used for scoring a linear model is the **R squared score**

Assessing Our Model

One common metric used for scoring a linear model is the **R squared score**

- It's a percentage

Assessing Our Model

One common metric used for scoring a linear model is the **R squared score**

- It’s a percentage
- Describes how close our line is to the data points

Assessing Our Model

One common metric used for scoring a linear model is the **R squared score**

- It's a percentage
- Describes how close our line is to the data points
- In other words - how well the **features** we choose predict the **sale price**

Assessing Our Model

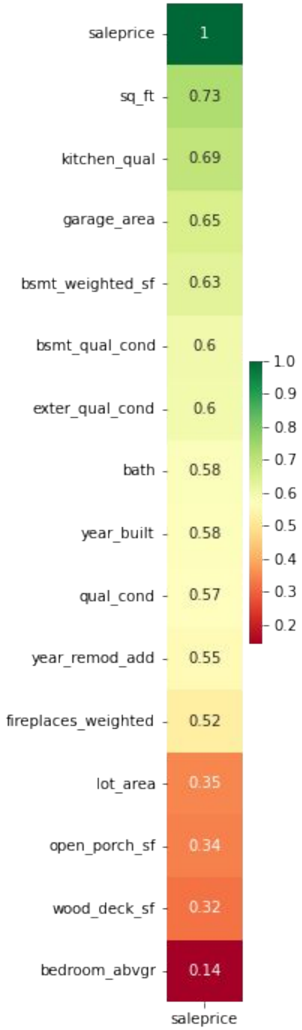
One common metric used for scoring a linear model is the **R squared score**

- It's a percentage
- Describes how close our line is to the data points
- In other words - how well the **features** we choose predict the **sale price**

Success: getting an R squared score greater than 75%

Not All Features Are Created Equal

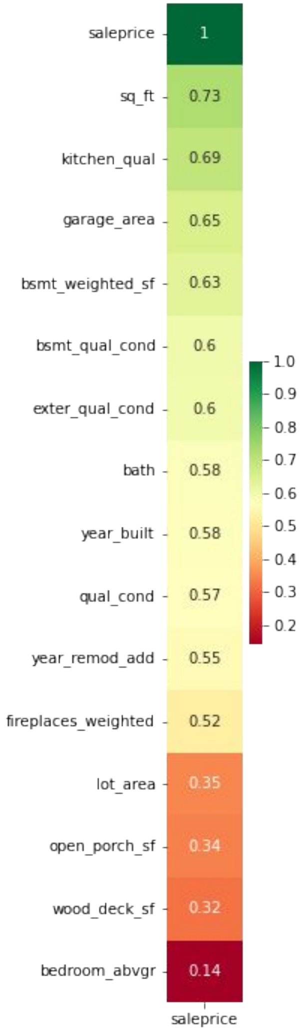
After cleaning the data and building some new data from old:



Not All Features Are Created Equal

After cleaning the data and building some new data from old:

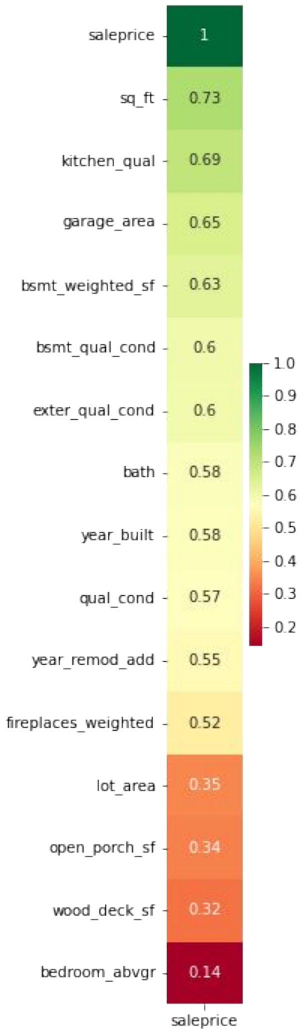
- we picked **numerical** features by determining if any were **correlated** to **sale price**



Not All Features Are Created Equal

After cleaning the data and building some new data from old:

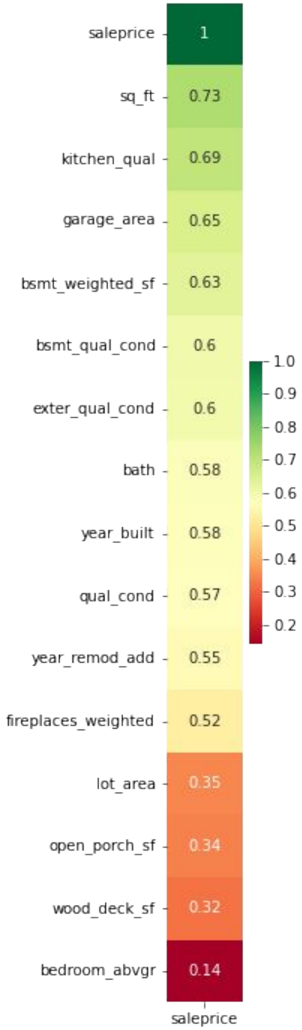
- we picked **numerical** features by determining if any were **correlated** to **sale price**
- And excluded **numerical** features that were **correlated** to **each other**



Not All Features Are Created Equal

After cleaning the data and building some new data from old:

- we picked **numerical** features by determining if any were **correlated** to **sale price**
- And excluded **numerical** features that were **correlated** to **each other**
- Used **categorical** features that were not overwhelmingly a single value

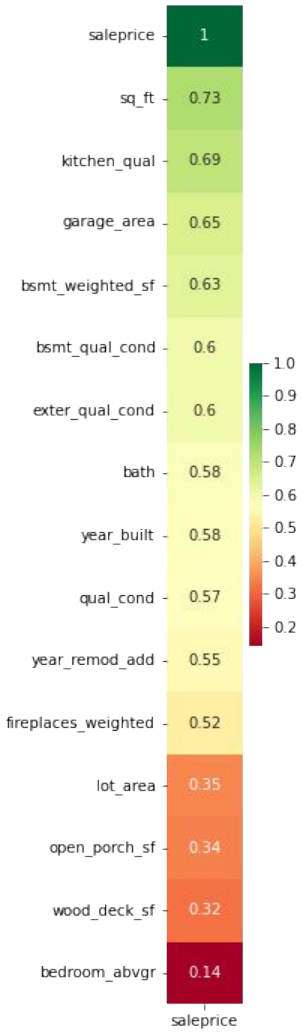


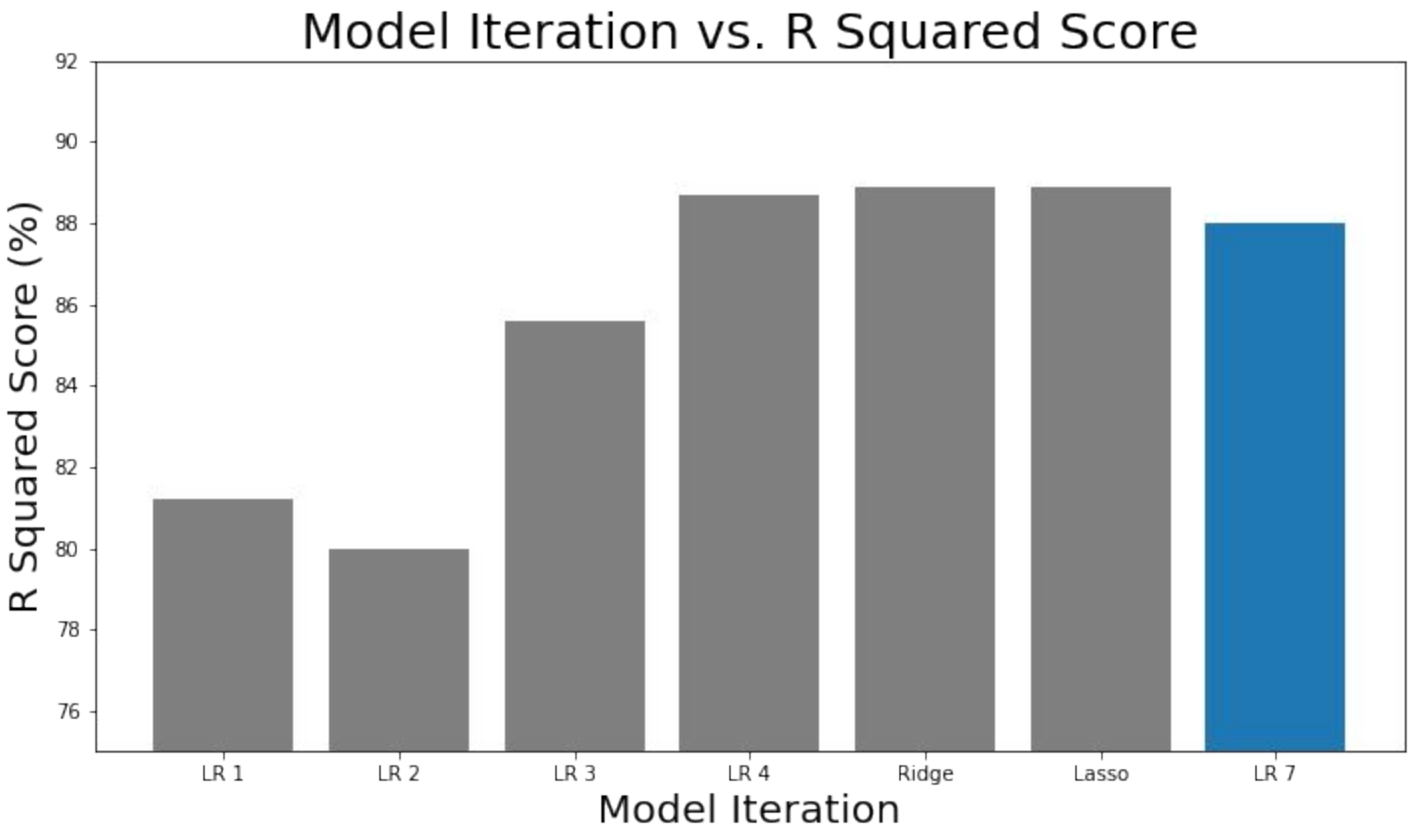
Not All Features Are Created Equal

After cleaning the data and building some new data from old:

- we picked **numerical** features by determining if any were **correlated** to **sale price**
- And excluded **numerical** features that were **correlated** to **each other**
- Used **categorical** features that were not overwhelmingly a single value

First model features: 1st and 2nd floor square feet, basement square feet, neighborhood, and house style





Model Iteration vs. R Squared Score
R Squared Score (%)
92
90
88
86
84
82
80
78
76
LR 1
LR 2
LR 3
LR 4
Ridge
Lasso
LR 7
Model Iteration

Interpretability Gained At Small Prediction Cost

Under the right conditions, linear models also include information about how much the output (**price**) changes when we adjust the input (**features**)

Neighborhood	Price Above	Neighborhood	Price Below
Green Hills	$82,623	Greens	-$19,130
Stone Brook	$41,322	Edwards	-$20,379
Veenker	$27,030	Brookside	-$20,652
North Ridge	$21,249	Meadow Village	-$26,773
North Ridge Heights	$18,848	Old Town	-$34,586

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%
- The production model we created scored an 88% on unseen data - **Success!**

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%
- The production model we created scored an 88% on unseen data - **Success!**
- Improving our predictions improve our brand, improve ad revenue

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%
- The production model we created scored an 88% on unseen data - **Success!**
- Improving our predictions improve our brand, improve ad revenue
- A small sacrifice in prediction score gives us interpretability

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%
- The production model we created scored an 88% on unseen data - **Success!**
- Improving our predictions improve our brand, improve ad revenue
- A small sacrifice in prediction score gives us interpretability
- Share insights from interpretability to public to promote our brand

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%
- The production model we created scored an 88% on unseen data - **Success!**
- Improving our predictions improve our brand, improve ad revenue
- A small sacrifice in prediction score gives us interpretability
- Share insights from interpretability to public to promote our brand

- To improve, try including homogenous data

Conclusion

- Our goal was to create a model to predict house prices with an R squared score larger than 75%
- The production model we created scored an 88% on unseen data - **Success!**
- Improving our predictions improve our brand, improve ad revenue
- A small sacrifice in prediction score gives us interpretability
- Share insights from interpretability to public to promote our brand

- To improve, try including homogenous data
- I started with features that stood out to me, try features that did not