

# Home Sale Price Predictions

A Story about Linear Regression...

# ...set in the city of Ames

Iowa Heartland Inspiration. World-Class Innovation\*

\*tagline proposed in 2011 by the city's Visioning Steering Committee

# SAFE... EASY... IMPORTANT

*Census 2020*

READ MORE



W E L C O M E T O A M E S , I O W A

[www.cityofames.org](http://www.cityofames.org)

# And rather affordable.

With a Median Home Price of \$162,500\*

From our somewhat dated data circa 2006 - 2010

# Lessons from predicting home prices in Ames

So Zillow can get more ad revenue

# Statistical Significance in Feature Selection

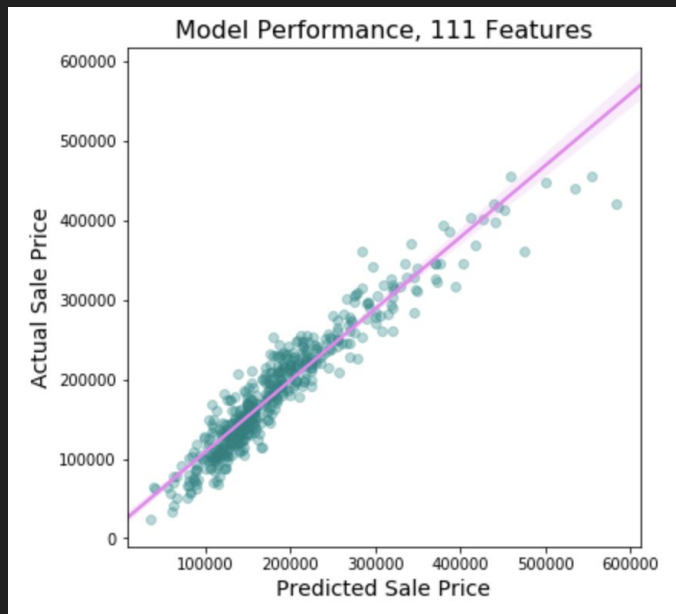
Let's Marie Kondo that. (does that feature bring you joy?)

## 111 Features

Mean Absolute Error: **\$18,256**

R2 Train: **.911**

R2 Test: **.907**

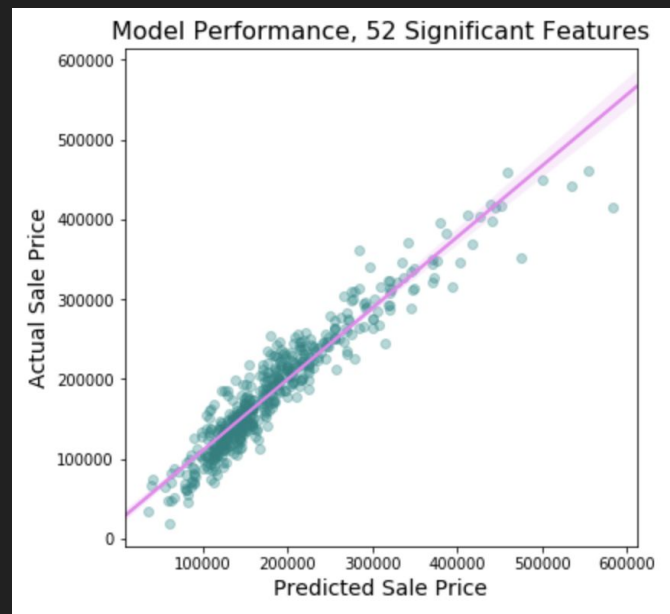


## 52 Features (P-value < .05)

Mean Absolute Error: **\$18,300**

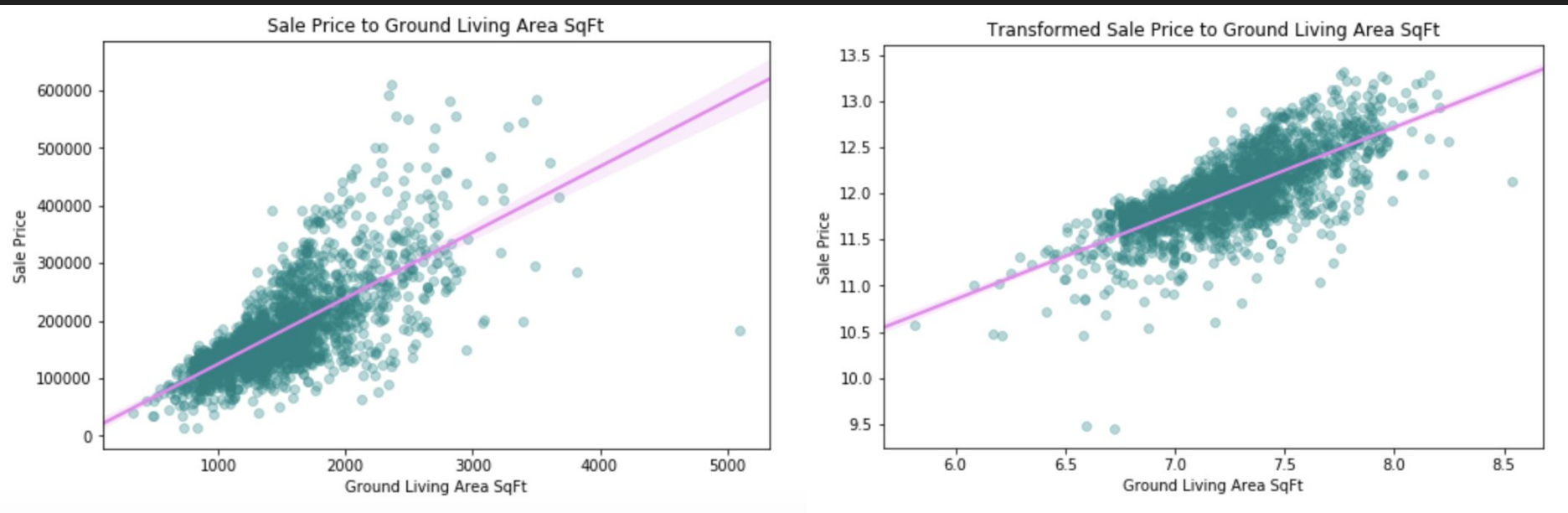
R2 Train: **.906**

R2 Test: **.905**



# Transforming data can enable more accurate models!

## Ground Floor Sqft to Sale Price Correlation



The old me.

The new me. (Log transformed)

# Model Performance with Log Transforming

**52 Features** (P-value < .05)

Mean Absolute Error: **\$18,300**

R2 Train: **.906**

R2 Test: **.905**



The improved model can predict home sale prices with an average of \$4,000 less error, a 20% improvement from the previous one!

**52 Features** (P-value < .05)

Mean Absolute Error: **\$14,460**

R2 Train: **.926**

R2 Test: **.923**





# Fancy Feature Selection with LASSO

With the Least Absolute Shrinkage and Selection Operator the computer can select features for you, and for little effort, give you a pretty good model!

To compare what LASSO did with the same feature as the last model:

**Last Model:**

**52 Features**

MAE: \$14,460

R2 Train: .926

R2 Test: .923

**LASSO:**

**45 Features**

MAE: \$14,060

R2 Train: .921

R2 Test: .926



Just a little something to crash that new Macbook:

With Polynomial Feature Creation our 52 features can become 1431 features!

This is going to be great.

# Lasso Selects the Best 100 Features out of 1431, and....

For all this extra computing power, we get only \$400 closer in our predictions of home sale prices, and an over fit model.

ಽ\_(ツ)\_/

Features: 100  
Mean Absolute Error: \$13,930  
R2 Train: .937  
R2 Test: .926



# For future improvements/some business ideas:

- Create a 'good enough' simple model/App with minimal features that can be used on-site to get a quick rough estimate
- Refine the model: explore features and their relationships, exclude collinear or unhelpful features
- Make trade-offs for data collection to save costs/server space, and potentially increase likelihood of collecting complete data
- Look at trends to choose which neighborhoods to promote or invest in to attract people to Ames



# THERE'S NO PLACE LIKE HOME

*Search property records*

READ MORE

W E L C O M E   T O   A M E S ,   I O W A

[www.cityofames.org](http://www.cityofames.org)

# Top 10 Correlated Features that Affect Sale Price

Overall Quality

Ground Living Area Sqft (log)

Exterior Quality

Kitchen Quality

Garage Cars

Garage Area

Total Basement SqFt

Year Built

First Floor Sqft (log)

Remodel/Addition Year

