AMES Housing Prices

DSI-7 : Project 2

Arielle Miro
'Data Scientist'

DATA SCIENCE PROBLEM

Grow Zillow's user base & advertising revenue .

How? ***By increasing confidence in our product.***

- Build a model that best predicts home prices in Ames, Iowa
- More accurate depictions of home prices = more confidence in the Zillow platform

METHODOLOGY

- Data Cleaning
  - Ordinal Columns
  - Feature Engineering
- Scaling & Tuning
  - Standard Scaler
  - Alpha
- Regularization
  - Ridge
  - Lasso

OBSTACLES

- Data Cleaning
  - The cleaner the data the better the model
- Null Values
  - Regression models do not accept NA values
- Outliers
  - Skew the data
  - Must account for future outliers

MODEL PERFORMANCE

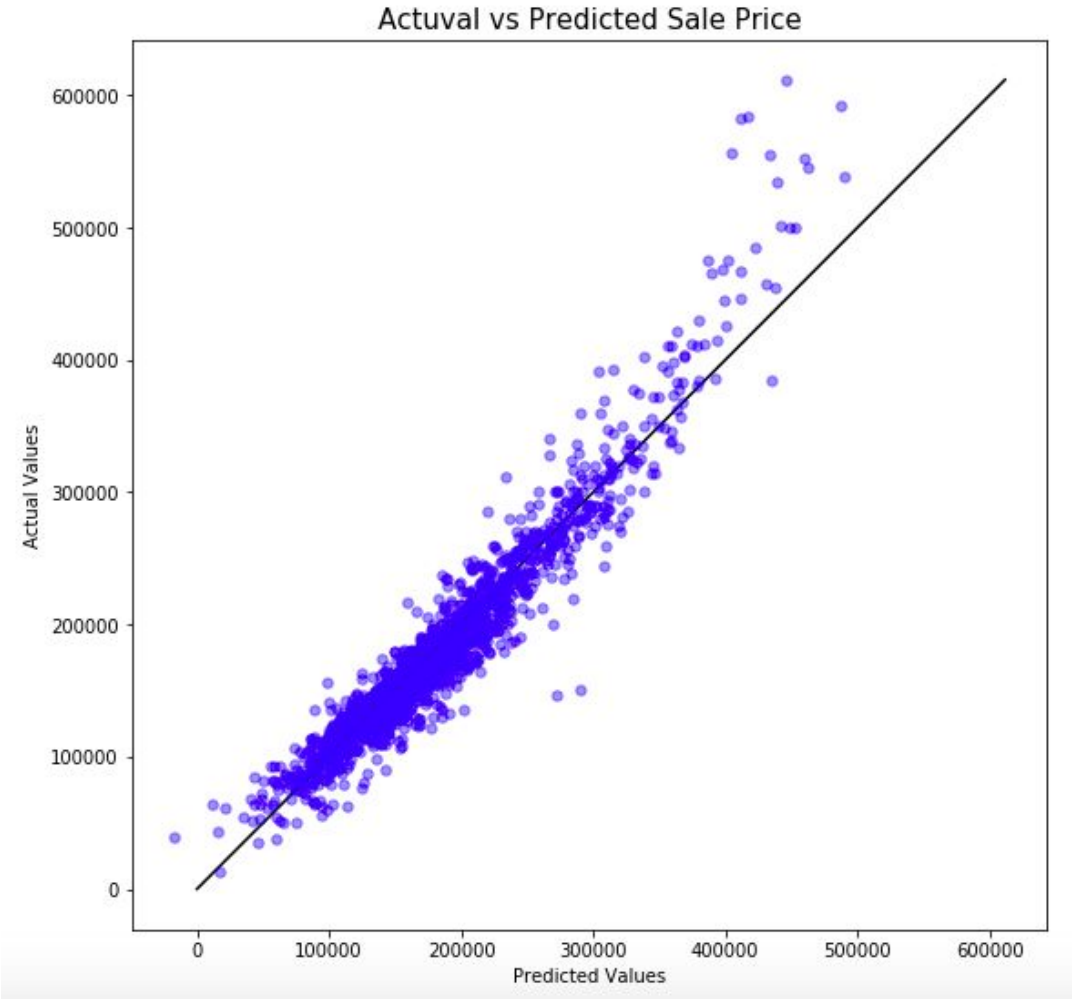
Regression Model:	Train Score	Test Score	Difference
Linear	0.9341302295764 98	0.898445759567386	0.035684470009112
Ridge	0.9340385167485 19	0.902954393504624	0.031084123243894
Lasso	0.9190104490374 38	0.910722602018539	0.008287847018897

*adjusted alpha

FINAL MODEL

Lasso Regression

- ➢ Closest train and test scores
- ➢ Tuning with alpha
- ➢ Scaled coefficients

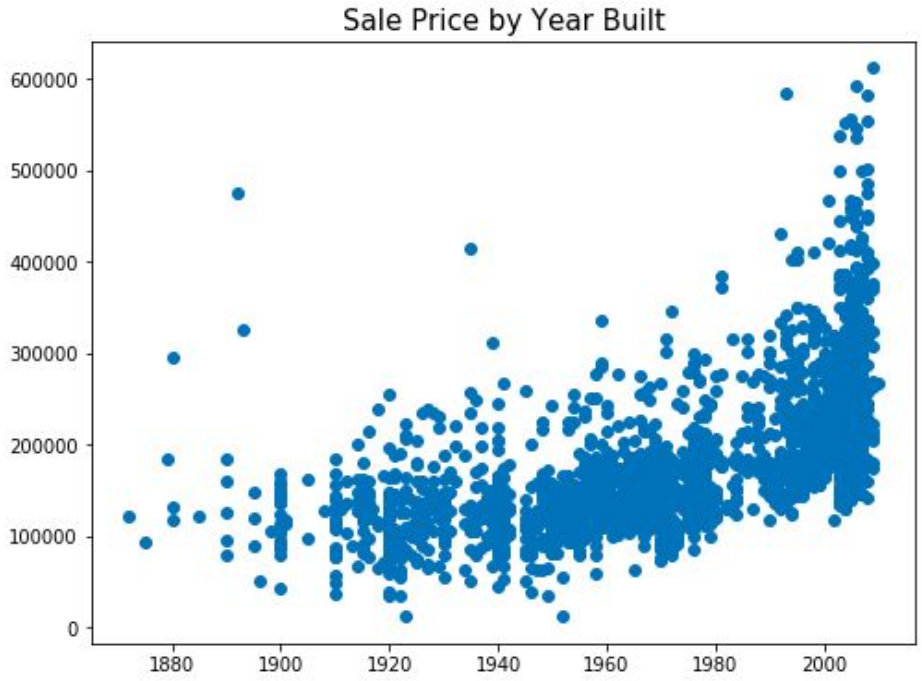


INFORMATIVE FEATURES

- Homes with larger 1st floors are more expensive
- Homes with basements are more expensive

INFORMATIVE FEATURES

- Newer homes are more expensive



TAKEAWAYS

As the overall quality ranking of the house increases by one unit, the sale price of the house increases by:

$12,818

RECOMMENDATIONS

Next Steps:

- Test the Lasso Model on recent data (2010 -2018)
- Refine accordingly
- Explore more features (less intuitive)
- Roll out model on the website to test for 30 days