

Predicting Ames House Sales

By: Joshua Mizraji

Problem: A Poor Home Buying Experience

Seller

- Undersell
- Long Sale Cycle
- Poor Targeted Marketing
- Invest In Low-Value Features

Buyer

- Overpay
- Missed Opportunities Due to Price



[Image Source](#)

Objective

By better understanding what influences house sale prices, **create a predictive model that gives us the expected sale price**, so that we can create a better home buying experience for buyers and sellers.



[Image Source](#)

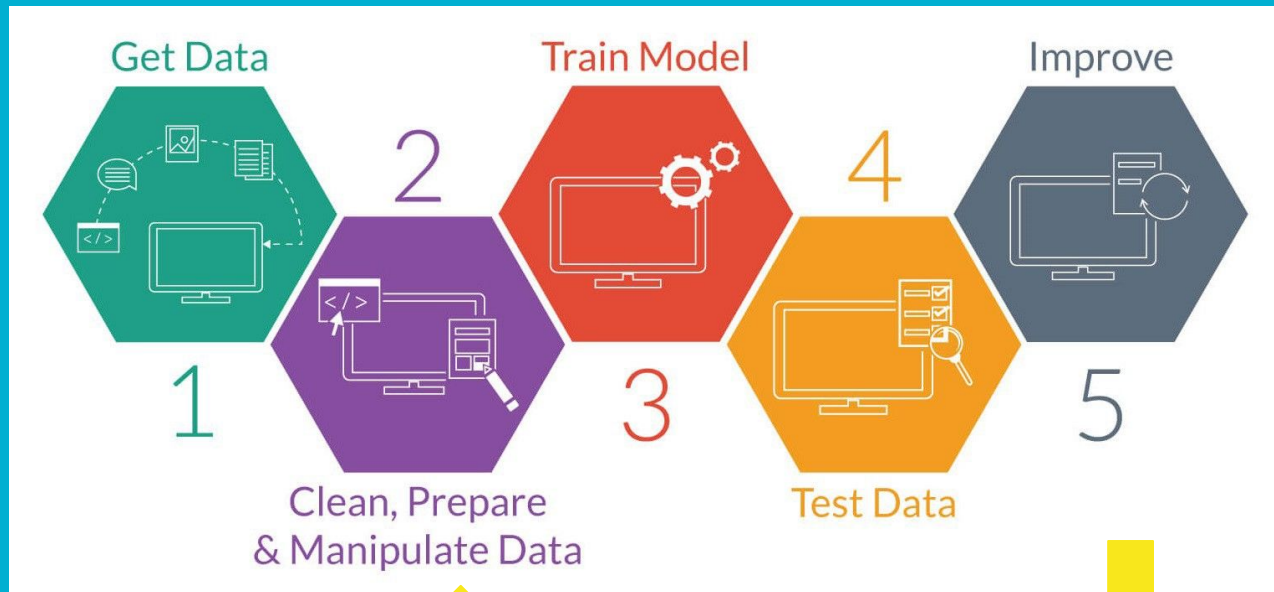
Understanding The Data & Process

Training Set

- Rows: 2051
- Columns: 81

Testing Set

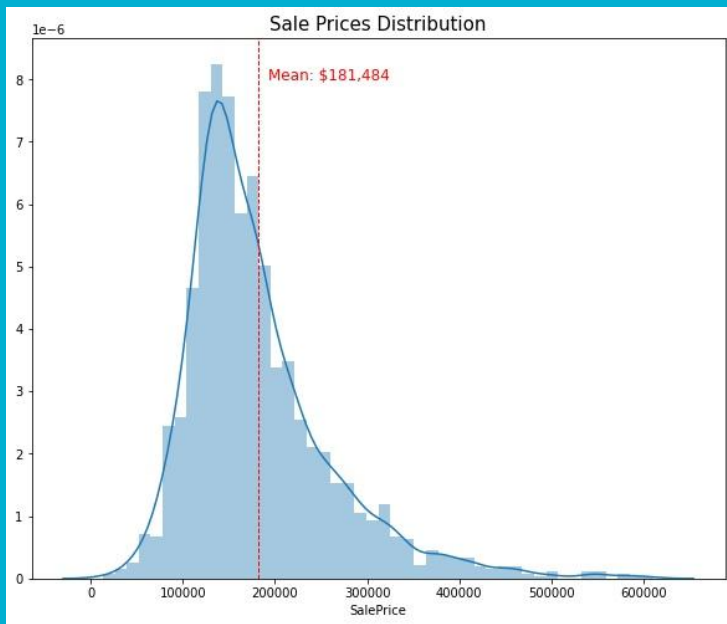
- Rows: 878
- Columns: 80



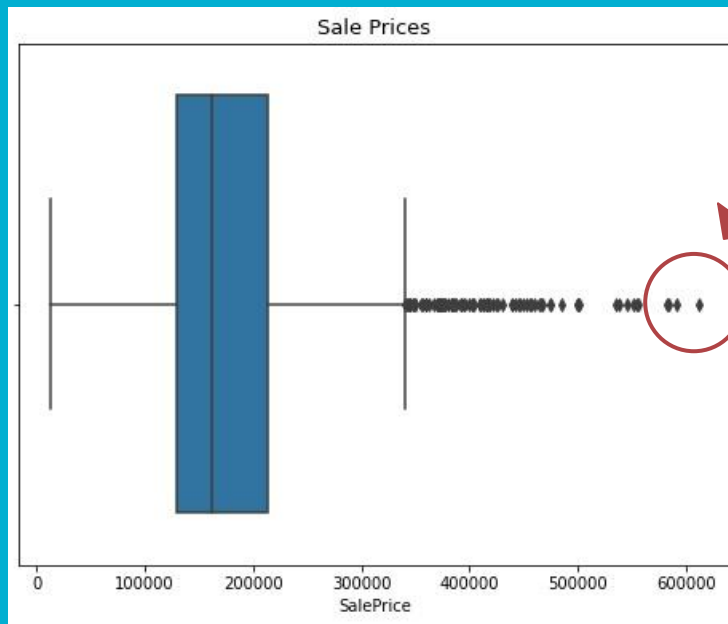
Data Exploration

What Are The Sales Prices Of The Homes?

How are sales distributed?



Are there any outliers?



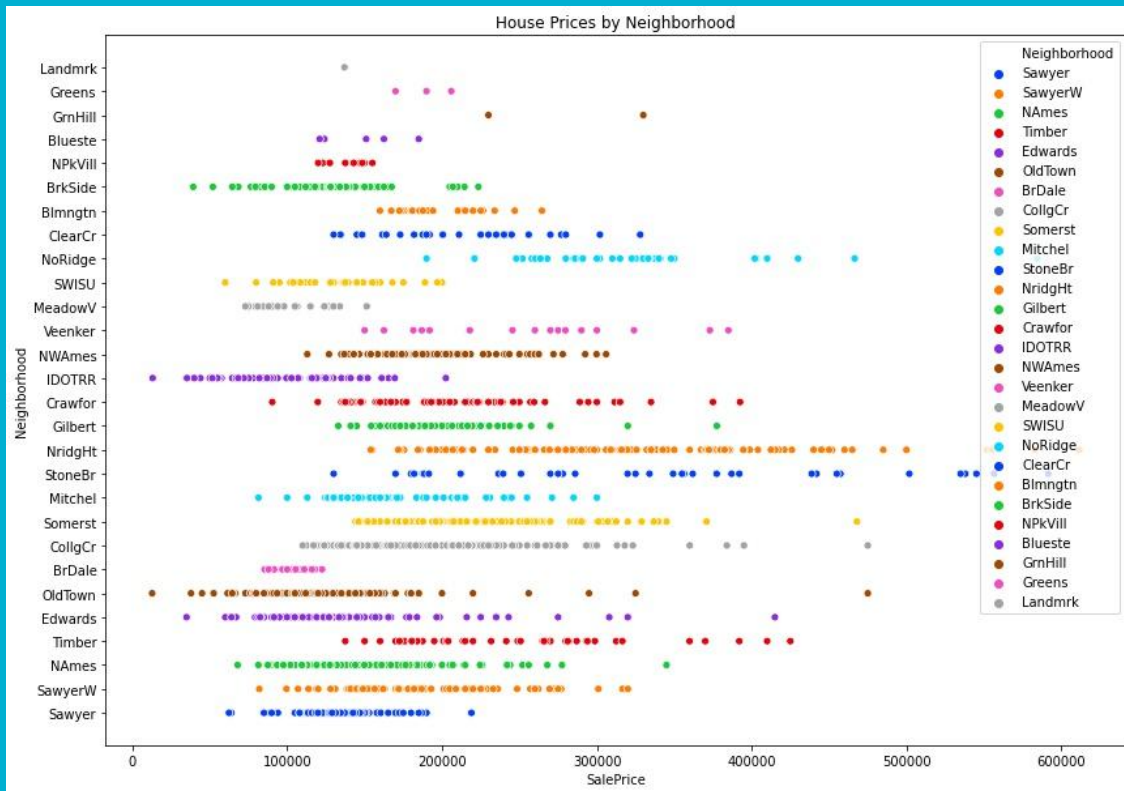
What Neighborhoods Are Most Expensive?

Most Expensive

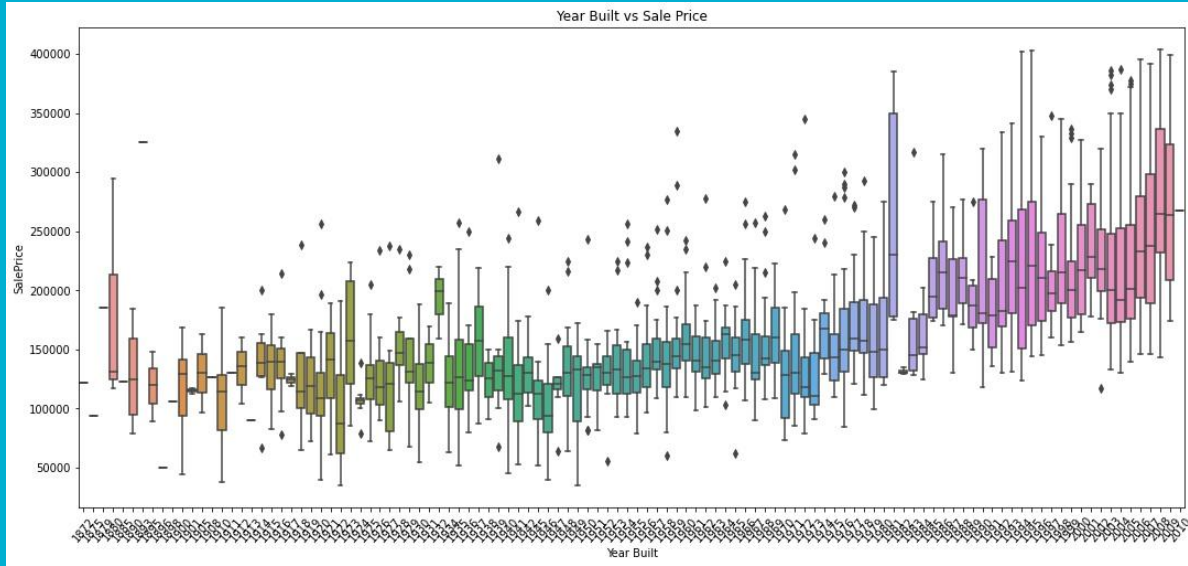
- StoneBr
- NridgHt

Least Expensive

- MeadowV
- IDOTRR



How Does Time Affect House Prices?

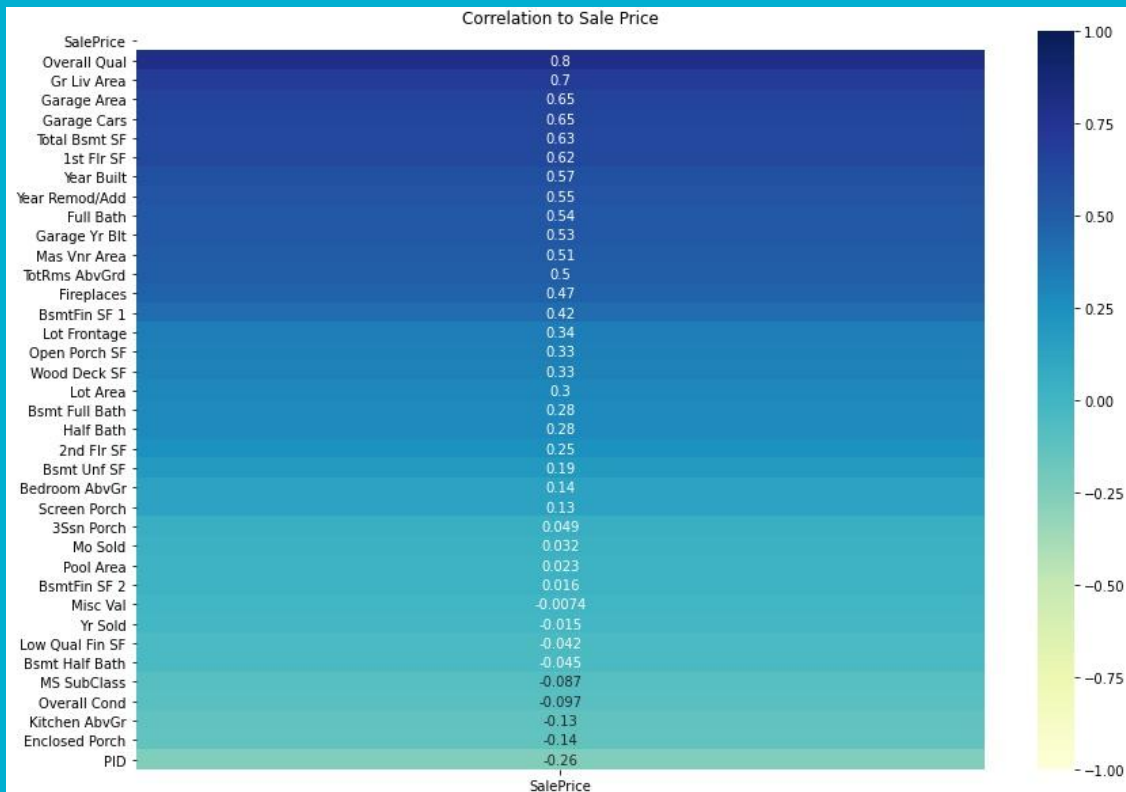


Takeaways: Newer houses generally sell for more, prices have remained relatively steady from 2006 - 2010.

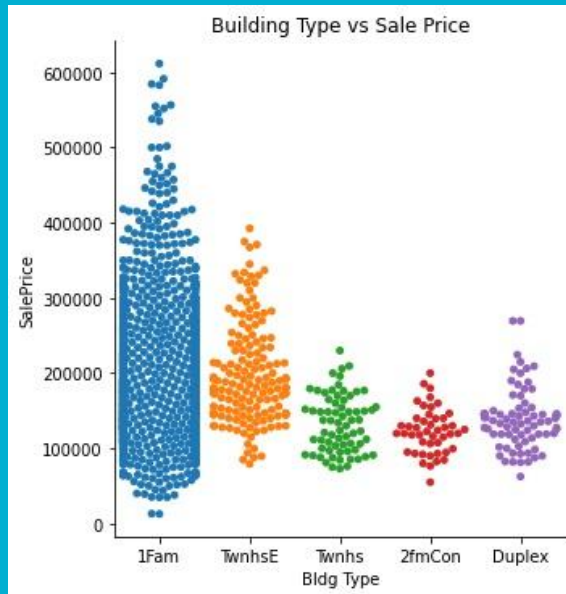
What Features Are Most Correlated to Price?

Most Correlated

- Overall Qual
- Gr Liv Area
- Garage Area
- Garage Cars
- Total Bsmt SF
- 1st Flr SF
- Year Built
- Year Remod/Add



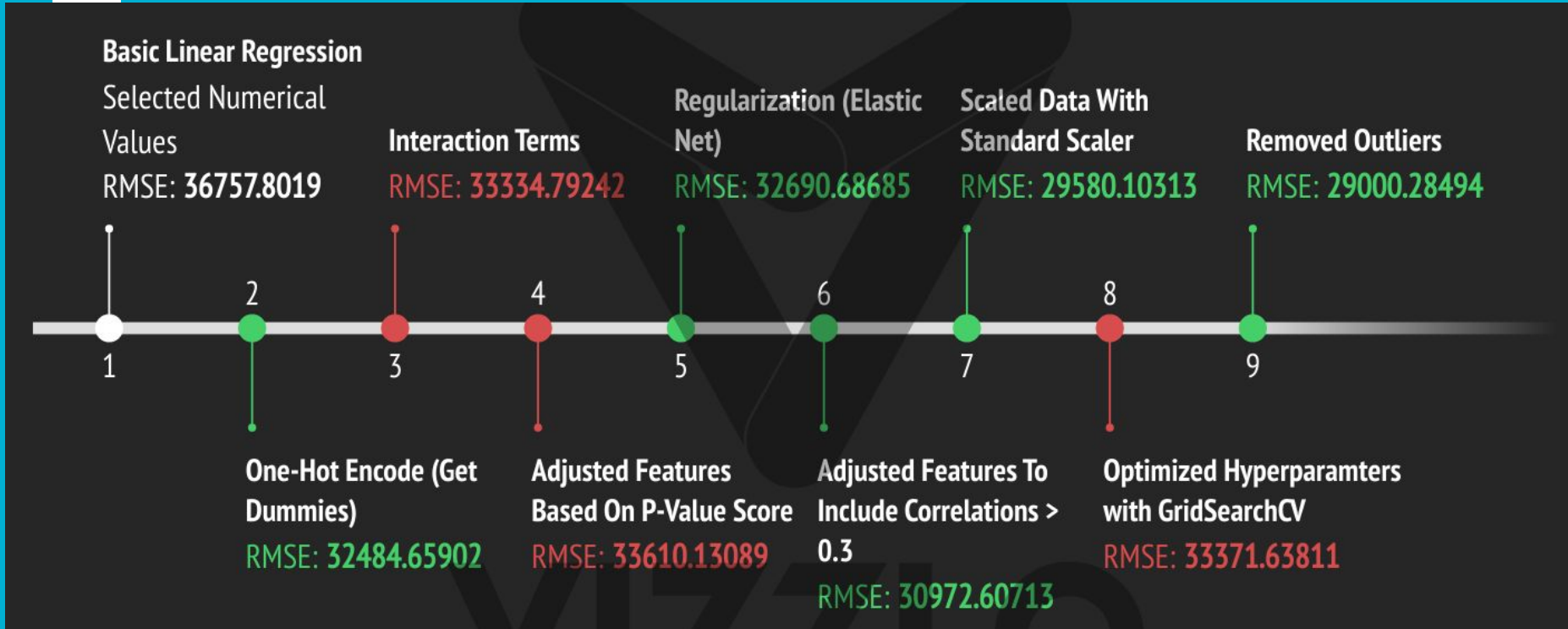
How Are Categorical Features Correlated?



Takeaways: Ensure your right to sell or build new, target 1 family homes, and invest in your kitchen.

Modeling

A Journey Through The Modeling Process



Model Results

Train, Test, Split:

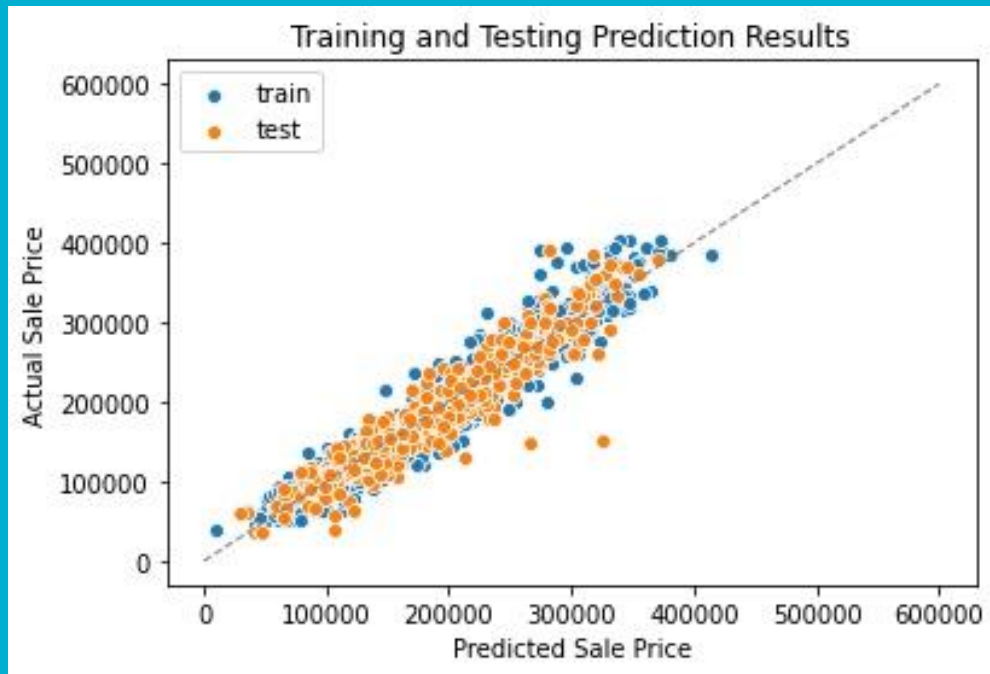
- Train: 0.917859994162
- Test: 0.8833326064739

Mean Cross Validation Score:

- 0.8972232611308646

Kaggle Submission (RMSE):

- 29000.28494



Recommendations For Buyers/Sellers

- Invest in quality homes with quality materials
- Keep the kitchen quality or remodel the kitchen
- Target 1 family homes for a competitive price
- Invest in a mid-priced house in either StoneBr or NridgHt, as house prices could continue to stay high as the neighborhood grows
- Build a new home to sell at the highest price
- Older homes generally sell at lower prices, but are not a dealbreaker

Future Analysis

- Deeper explore feature relationships and fill out null values so I can include them in my model
- Additional optimization of features based on coefficients and p-values

Thank You!