Ames House Price Predictions

Linear Regression

Reid Nelson | Data Scientist

The Setting

- Working on a real estate investment advisory team in Ames Iowa in late 2010
- Tasked with developing a model help brokers, agents, and investors better predict sale prices after the 2008 housing market crash.
- The team already has a working model, but one that is not robust enough, given the recent market volatility.

Overview

The Problem

Predicting house prices has become difficult given the volatility in the wake of the 2008 market crash.

The current sale price prediction model is too error prone for current market.

Our Task

The task is two-fold:

1) Develop a model that minimizes prediction error as much as possible.
2) Share actionable insights about how to use this new model to minimize risk in the current market.

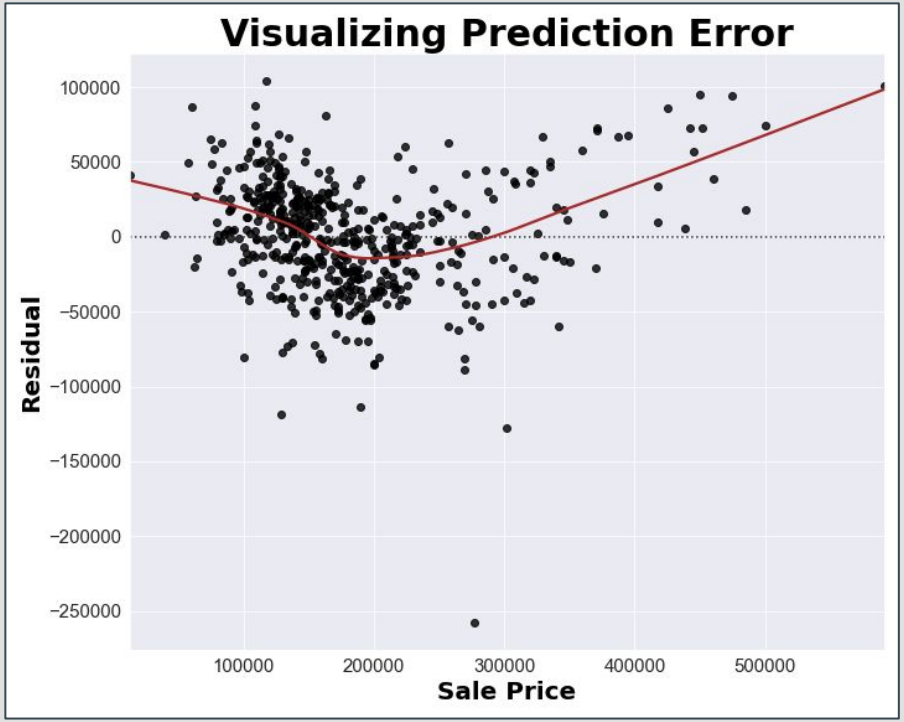
Current Model:

Performance:

- R2: 0.69
- MAE: 30986

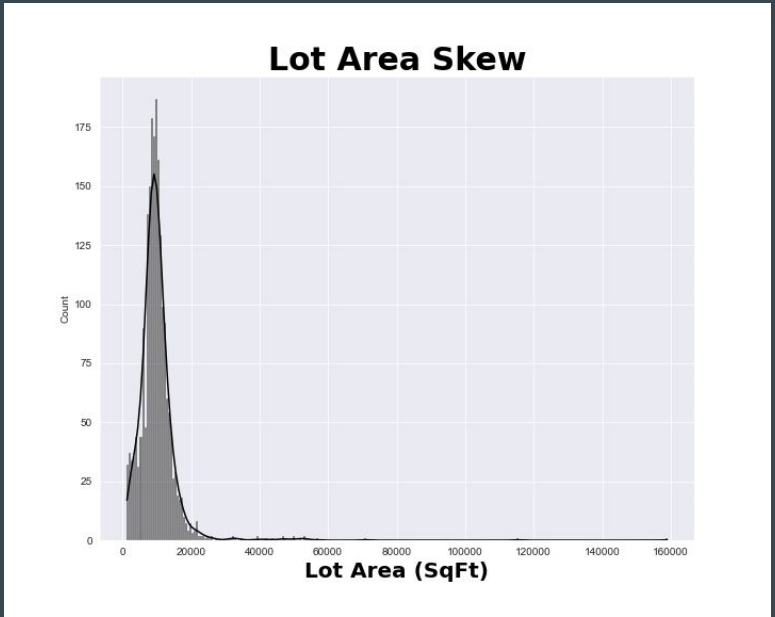
Features:

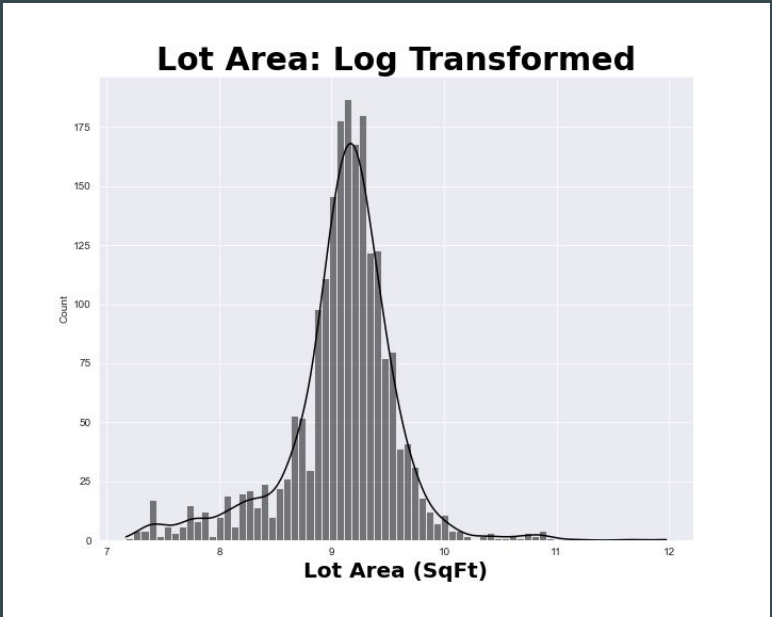
- Square Footage
- Lot Size
- Overall Condition
- Year Built



Normalizing Target and Feature

Normalizing Target and Feature





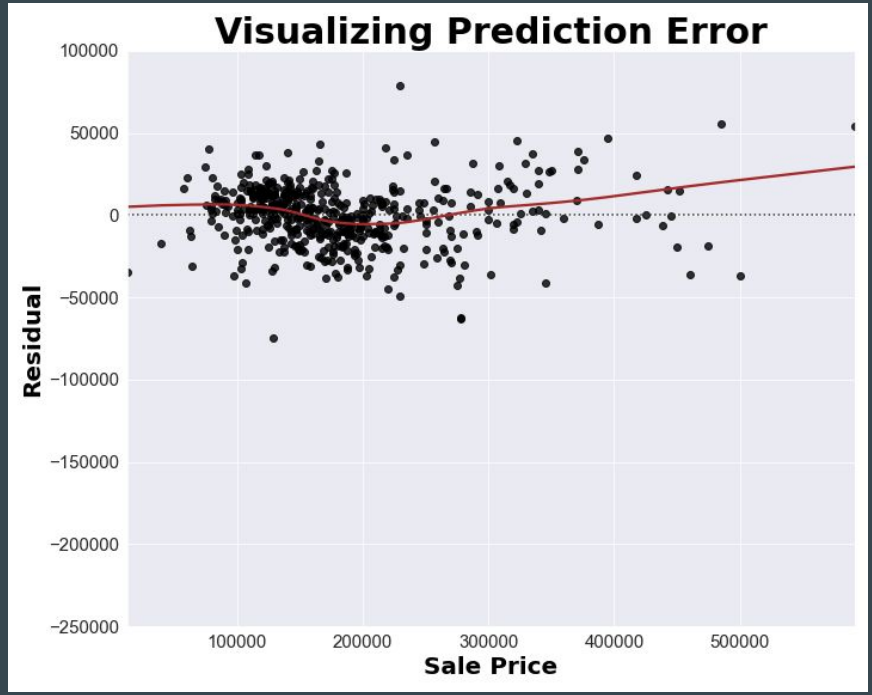
Proposed Model

Performance:

- R2: 0.89
- MAE: 14931

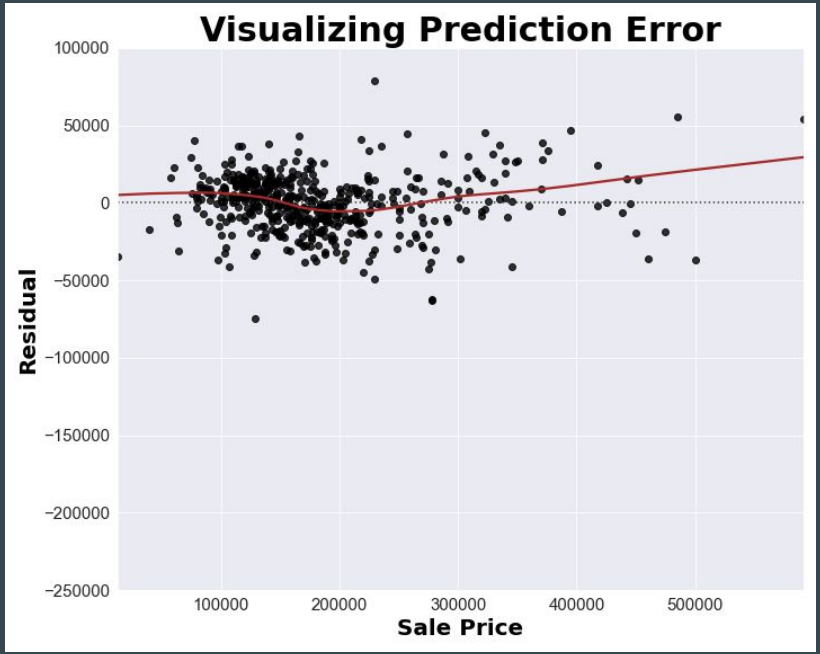
Features:

- 152 Features



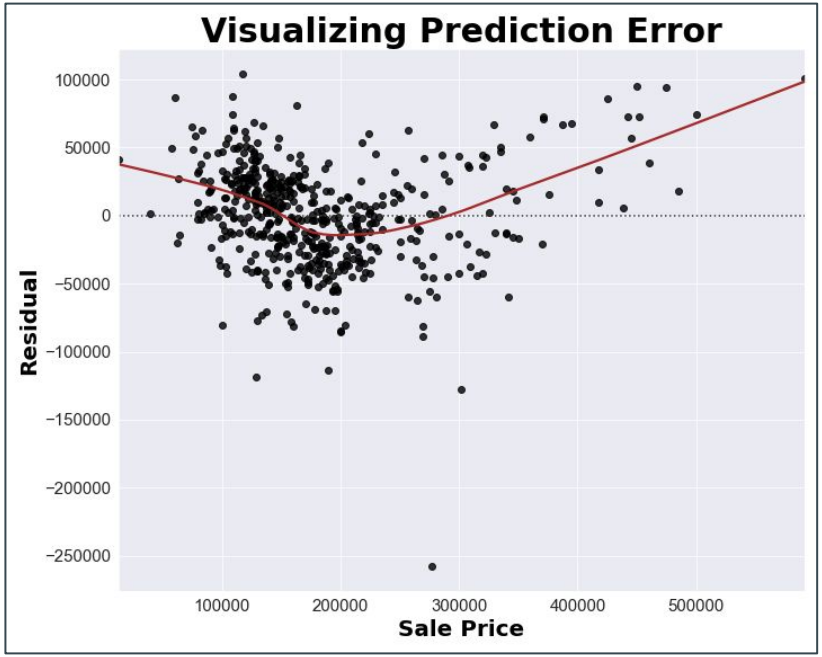
Proposed Model

MAE: $14,931
R2: 0.89



MAE: $30,986
R2: 0.69

Current Model



Take-Aways

Performance

The model delivered reduces prediction error by more than 50%.

The model performs better across a wider range of prices, and we expect it to generalize better to unseen data.

Limitations

This model is not designed to make inferences about how individual features contribute to sale price.

The model will perform most reliably where we have the most data–between $100,000 and $250,000