Home Sale Price Estimator

Daniel Rossetti
Data Science Consultant
for Zillow

Problem Statement & Requirements

Create a prototype home sale price estimator for Zillow

- Provided data from Ames, Iowa
- Intended Users: Home buyers and sellers
- Goal: On-average, ***< $30k*** sale price error
- Model focus: Prediction

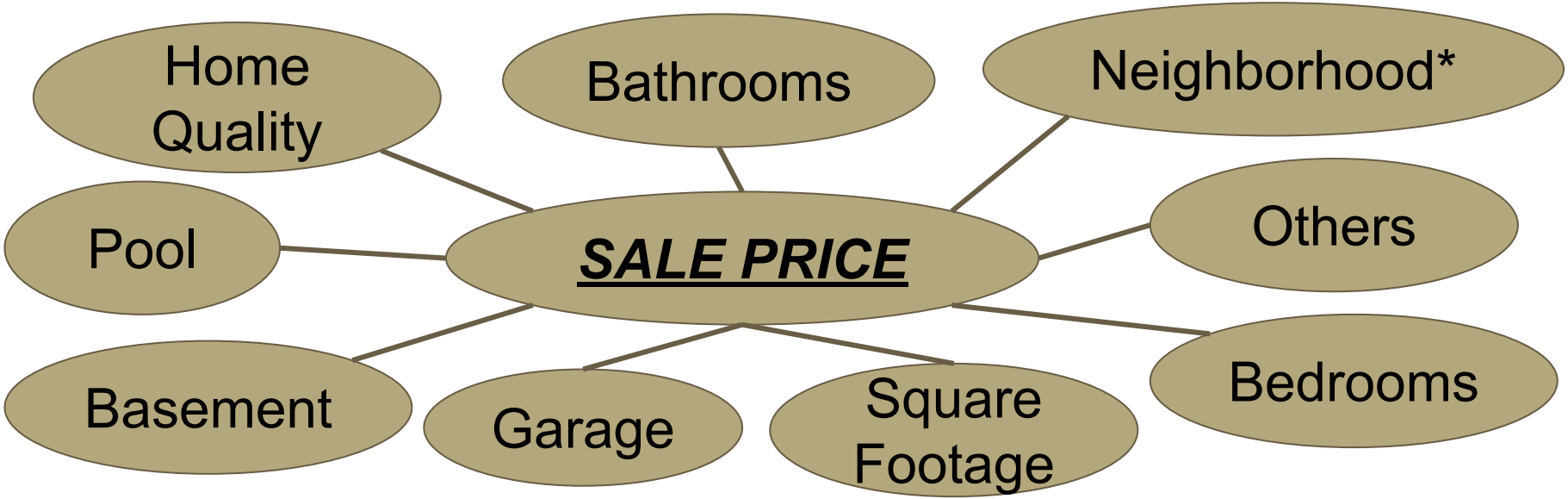
Bottom Line Upfront

Prototype model predicts sale prices within $30K on-average

Required user/model inputs:

- Basement square footage
- Garage square footage
- Overall home quality (1 - 10 scale)
- Remod/Addition/Built Year
- Total square footage
- Square footage of masonry veneer
- Number of fireplaces
- Neighborhood

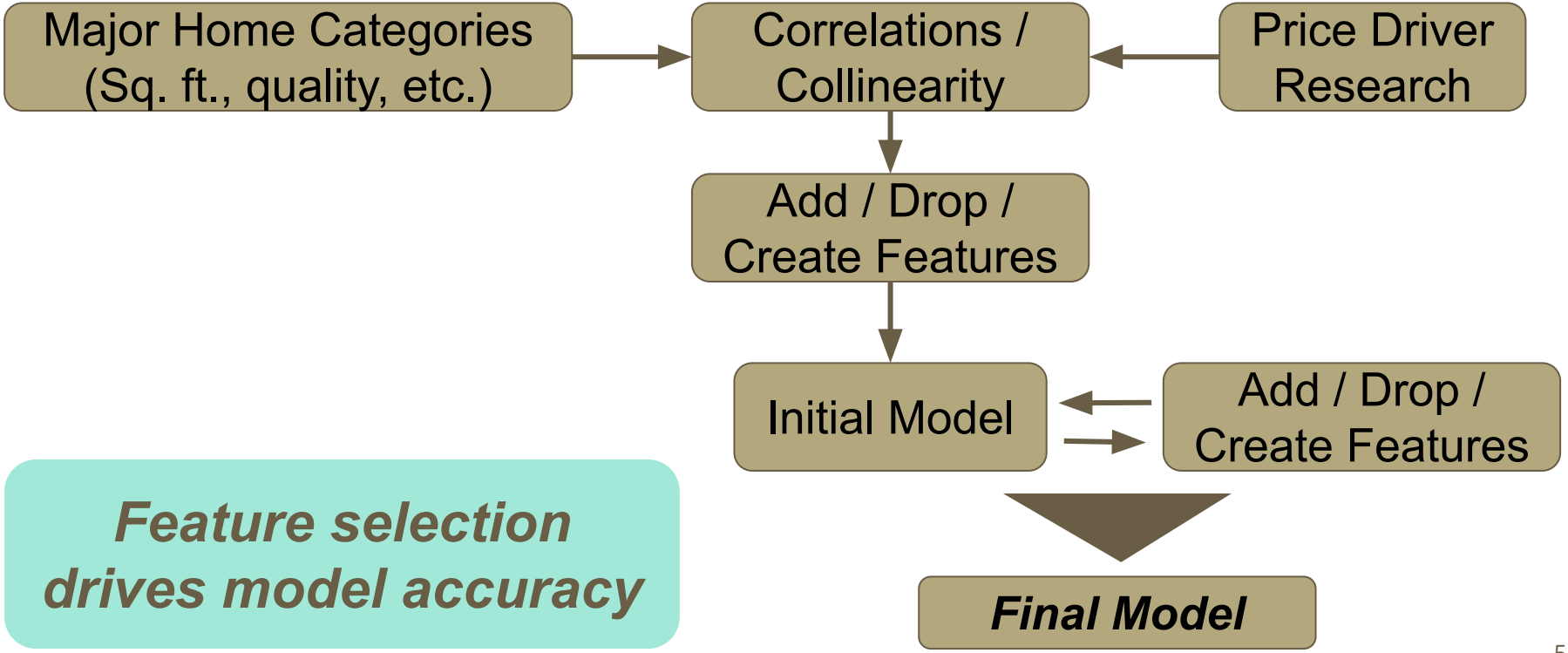
Provided Data - Perceived Value Adders



80 Home Characteristics Provided
⇒ Focus on Major Home Categories ⇐

Source [1]

Modeling Process



Feature selection drives model accuracy

Model Performance

- 83.8% of predictions within $30k*
  - Average error: $18,643

- Good performance on new data
  - Model bias and variance balanced

* On training and test data

Next Steps - Model Improvement

Outliers / Flanks

▼

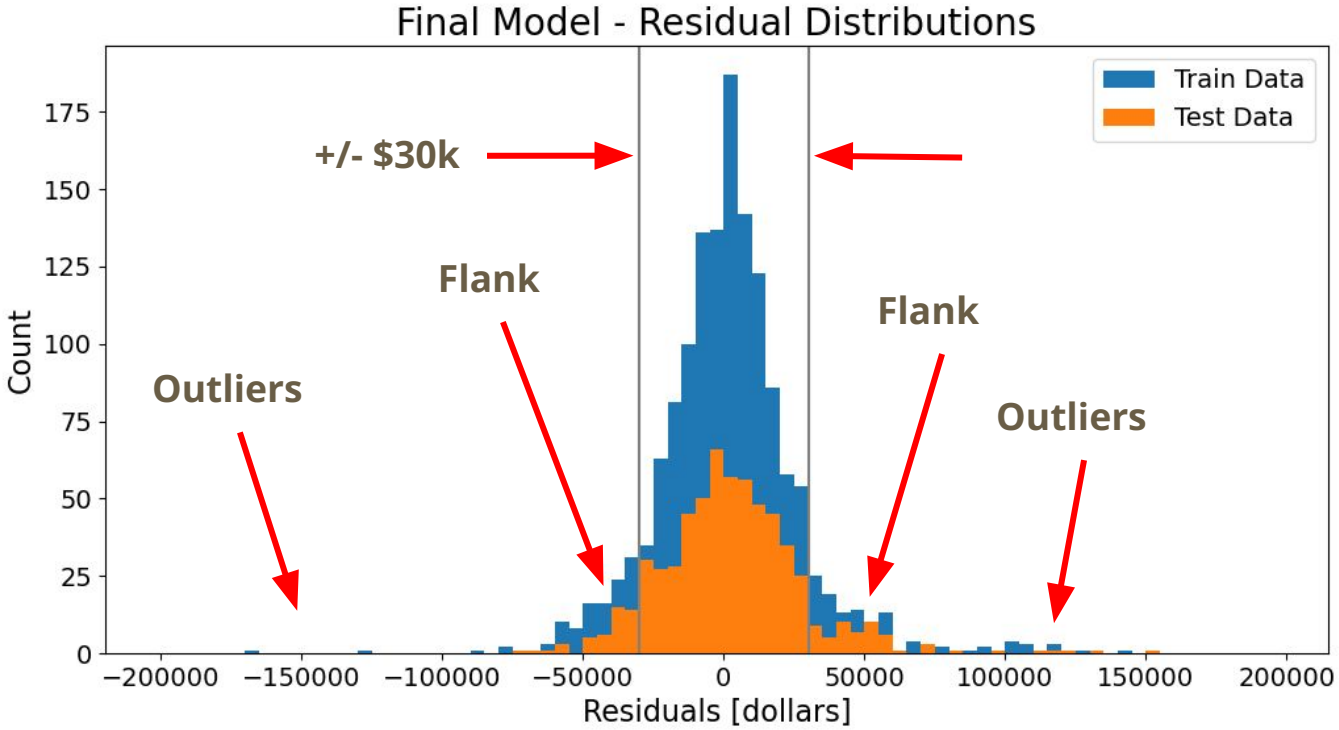
Trends

▼

Improvement

▼

Opportunity



Next Steps - Model Improvement

Higher Sale Price

▼

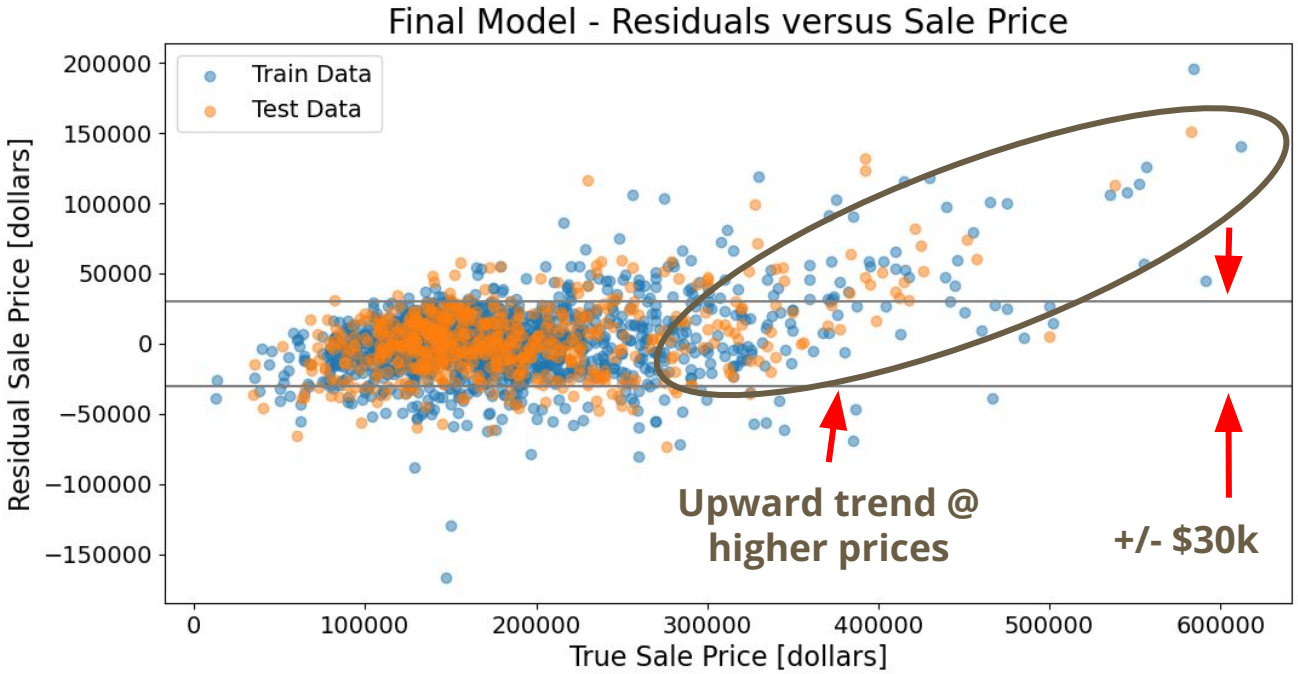
Higher Error

▼

Characterize

▼

Improve



Summary

- Prototype home sale price estimator created
- Met $30k mean error goal
- Robust performance on new data
- Continue to examine pitfalls and improve model

Sources

1. *8 critical factors that influence a home's value*: https://www.opendoor.com/articles/factors-that-influence-home-value
2. Kaggle Competition Data: https://www.kaggle.com/competitions/dsir-320-project-2-regression-challenge/data
3. Other sources provided in code notebook

Detailed Model Performance Parameters

Performance Parameter	Train Data	Test Data
R-Squared	0.8946	0.8921
Root Mean Squared Error (RMSE)	$25,789	$25,887
Mean Absolute Error (MAE)	$17,692	$18,643
Null Model RMSE	$79,420	$78,862
Null Model R-Squared	0	0