Managing Home Buyer Expectations

Evaluating the insightfulness of home-sale aggregator sites for consumers

Polina Minkovski
September 11, 2023

Problem at hand

Background

Explore Ames Housing Data to iteratively build a model that would predict Sale Prices of Homes

Question to be solved

How much can a home buyer really learn about home prices by way of a basic search on a Realtor Aggregator site?

Approach

- Keep it simple
- Add complexity
- Focus on the variables that most consumers pay attention to

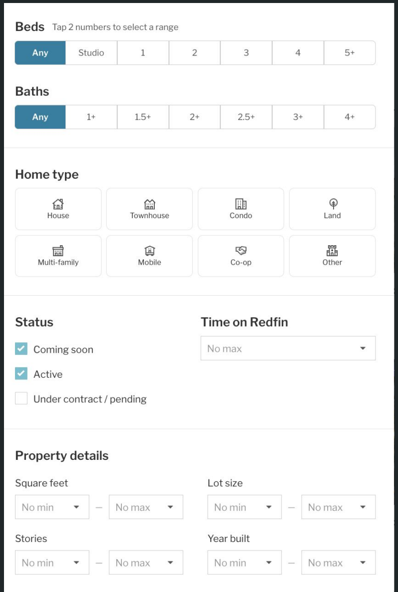
Dataset

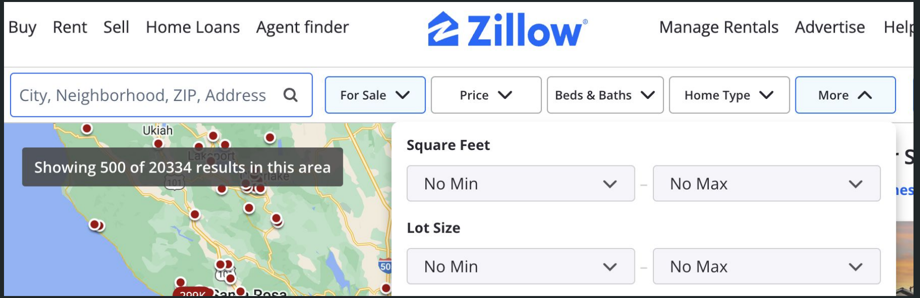
A dataset, encompassing 81 features of houses -- mostly single family suburban dwellings -- that were sold in Ames, Iowa in the period 2006-2010, which encompasses the housing crisis

More information about the AMES dataset can be found here

Baseline search criteria:

1. Neighborhood
2. Lot size
3. House size
4. # Bedrooms
5. # Bathrooms
6. Home Type
7. Cost/sqft
8. # Garage spots
9. Overall quality





Model performance

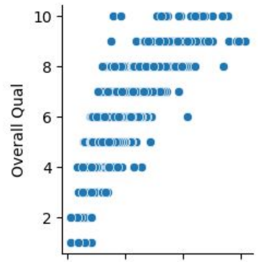
	Model	Summary	R^2	RMSE
1	Simple Linear Regression	No transformations/feature engineering. Variables with multicollinearity were not included. Variables include: Overall Quality, Total Basement SF, Gr Living Area, 1st Flr SF, Garage Cars	Train: 0.78; Test: 0.82	33980
2	Linear Regression	Feature engineered and scaled features. Variables include: Total sqft, Overall Condition, Overall Quality, Interaction Condition and Quality, Zone (Dummy), Neighborhood (Dummy)	Train: 0.81; Test: 0.85	31248
3	Same as Model 2, but Lasso Regularization	---	Train: 0.81; Test: 0.85	31043
4	Linear Regression	Feature engineered and scaled features. Variables include: Total sqft, Overall Condition, Overall Quality, Interaction Condition and Quality, Bathrooms, Bedrooms, Month sold, Neighborhood (Dummy)	Train: 0.79; Test: 0.80	34671
5	Linear Regression	Feature engineering and scaled features. Variables include: Total sqft, Lot Area, Overall Condition, Overall Quality, Interaction Condition and Quality, Bathrooms, Bedrooms, Neighborhood (Dummy), Garage Capacity, House Style (Dummy)	Train: 0.81; Test: 0.80	35175
6	Lasso Regression	Feature engineering and scaled features. Variables include: Overall Condition (log), Overall Quality, Interaction Condition and Quality, Bathrooms, Bedrooms, Total sqft House (log), Total sqft Lot Area (log)	Train: 0.77; Test: 0.80	34435
7	Linear Regression	Feature engineering and scaled features. Variables include: Overall Condition, Overall Quality, Interaction Condition and Quality, Bathrooms, Bedrooms, Total sqft Lot Area (log), Price per Sqft (per neighborhood)	Train: 0.82; Test: 0.84	30908
8	Lasso Regression	Feature engineering and scaled features. Variables include: Total sqft House (feature engineering), Overall Condition, Overall Quality, Interaction of Condition/Quality, Bathrooms, Bedrooms, Total sqft Lot Area (log), Price per Sqft (per neighborhood), Total sqft Lot Area (log), Price per Sqft (per neighborhood), Zone, Building Type	Train: 0.82; Test: 0.84	30800
9	Lasso Regression	Feature engineering and scaled features. Variables include: Neighborhood (dummy), Total sqft Lot Afea (log), Total sqft House (log), Bedrooms, Bathrooms, House Type (dummy), Bldg Type (dummy), Cost/Sqft (per neighborhood), Garage Capacity, Overall Quality	Train: 0.84; Test: 0.83	31773

What Consumers Look for First

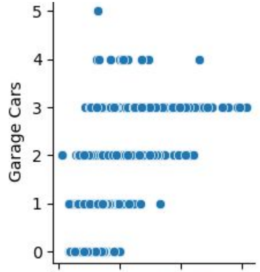
1. Sq Feet (House)
2. Sq Feet (Lot)
3. Bedrooms
4. Bathrooms
5. Garage Capacity
6. Neighborhood
7. House Type
8. Building Type
9. Quality

How will this impact their home purchase budget?

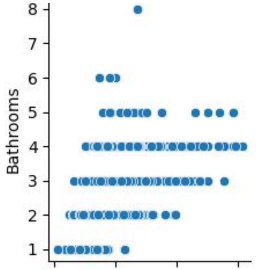
In general, House Prices are not going to prevent consumers from getting what they want in a house



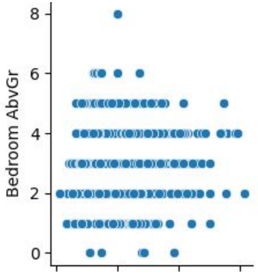
corr = 0.8



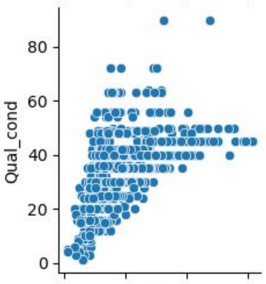
corr = 0.65



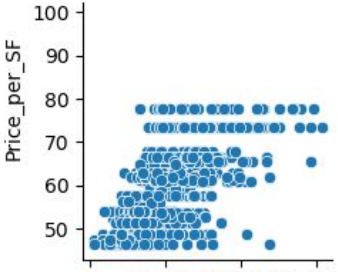
corr = 0.61



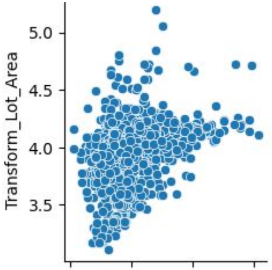
corr = 0.14



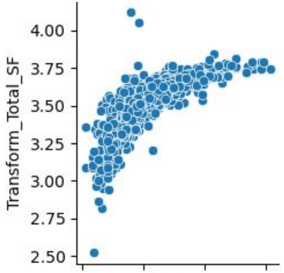
corr = 0.57



corr = 0.69



corr = 0.37



corr = 0.79

What Consumers Don't See

	Coefficient
Log(Sqft House)	20,679
Price per Sqft (Neighborhood)	18,182
Overall Quality	18,052
Number of Bathrooms	14,393
Single-Family Detached House Type	9,456
Log(Sqft Lot Area)	7,719
Quality and Condition (int)	7,364

- An extra sqft does not increase price in a uniform way
- Overall finish and materials used for the house matter
- Making a decision to go for a specific house type will change your expected sale price
- Number of bathrooms can make a difference to sale price

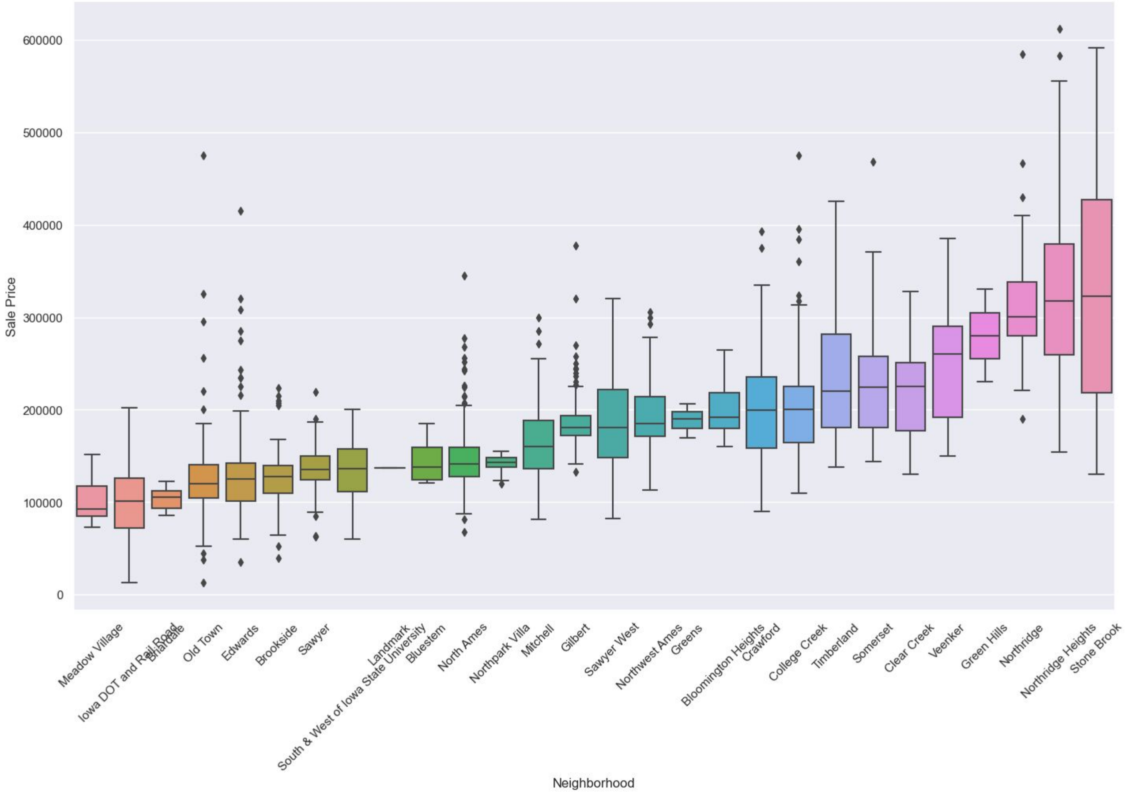
What Consumers Don't See

	Coefficient
Northwest Ames (relative to Sawyer)	-1,734
Two and one-half story - 2nd level unfinished (relative to Split Level)	-1,828
Bedrooms	-2,072
Northpark Villa (relative to Sawyer)	-2,332
Meadow Village (relative to Sawyer)	-2,491
Old Town (relative to Sawyer)	-3,404
Somerset (relative to Sawyer)	-5,154

- We excluded Sawyer neighborhood from our analysis in order to give us a point of comparison

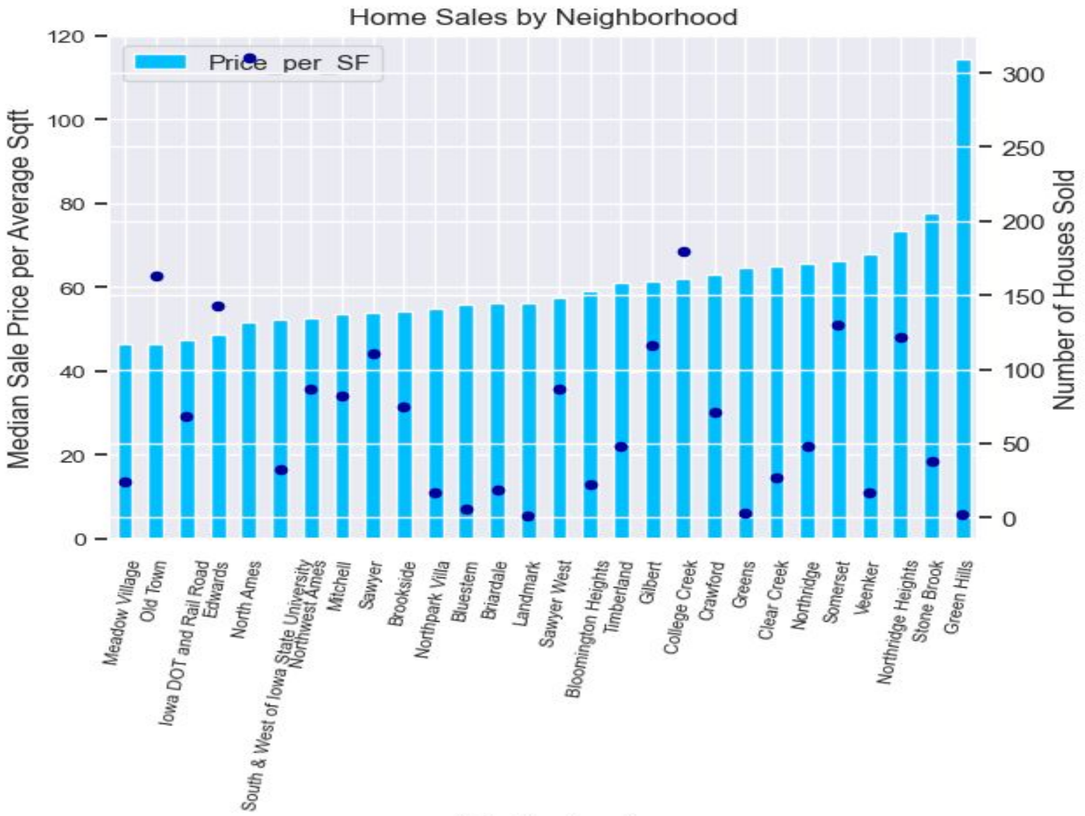
- We can see that home prices in some neighborhoods are likely to be significantly cheaper than Sawyer, all else being equal

Sale Price by Neighborhood



- Predicted house prices are expected to continue to reinforce the current neighborhood stack rank by median house price

Sale Price by Neighborhood



- However, sale price does not necessarily decrease when there are more houses sold in a given neighborhood

- At first glance, the most expensive neighborhoods (as evaluated by Median Sale Price/ Average Square Foot Sold) saw lower than average (73) houses sold

What does this surface level data tell us?

- Using Lasso Regression, our features are able to explain approximately 83% of variability in Sale Price (alpha = 147)
- The RMSE in the model is quite high ($31,773)
- Consumers need more information, in order to have more confidence with expected sale prices given their search criteria

What’s next?

1. Run the full analysis to determine whether scarcity value is a driver or outcome of Housing Market Activity
2. Conduct lead/lag analysis to determine how prices impact sale volumes by neighborhood
3. Make a shortlist of features that would be helpful for consumers to have on Realtor Aggregator websites, stack-ordered by impact on sale price
4. Provide insights to how much ‘on average’ a home price would change based on a consumer’s selection (either of additional features or within a specific feature)
5. Conduct additional analysis by year and season, to ensure time horizon and seasonality are included in recommendations
6. Look at by-neighborhood or by-neighborhood group behavior of various features, in case recommendations need to be adjusted based on where a consumer is looking to buy