Opendoor - Price Predictions

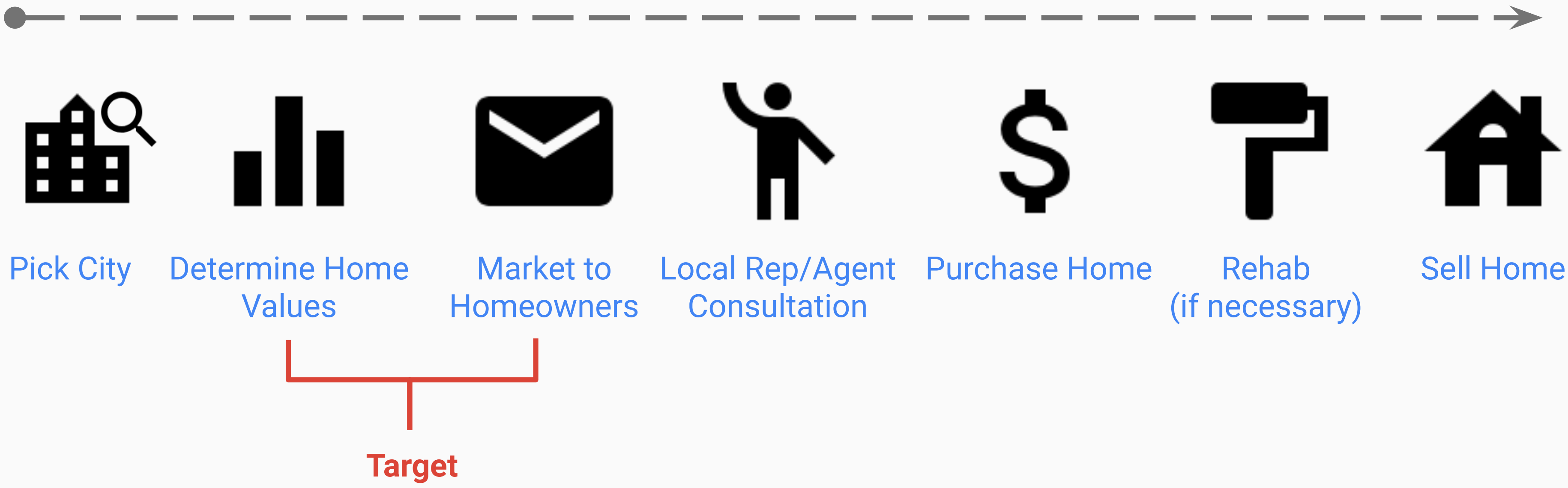
Scott Rosengrants
Project 2 - DEN - Flex - 10

Mission Statement

"The revolutionary way to buy and sell your home"

Opendoor: https://www.opendoor.com/

How it works


Pick City
Determine Home Values
Market to Homeowners
Local Rep/Agent Consultation
Purchase Home
Rehab (if necessary)
Sell Home
Target

How it works

Acquisition Offer Made:	$300,000	$300,000
Service Charge	6 - 14%	$18,000 - $42,000
Avg. Service Charge	6.4%	$19,200
Avg. Repairs needed	$10,000 (from Seller proceeds)	$10,000 (from Seller proceeds)
Resale Price (target)	$320,000	$320,000
Sale Co-Op Fee	3% of resale value	$9,600
Avg. Profit	Avg.Profit	**$29,600**

What we need

In order for Opendoor to be successful we need to know the **acquisition price** and **sale price**. To get an accurate sale price we need specific home information and a predictive model.

- Home Condition
- Year Built
- Total Square Footage
- Bedroom Count
- Bathroom Count
- Kitchen Quality
- Type of Neighborhood
- Garage Features
- Type of Home (layout)
- Etc.

The Case Study

Location: Ames Iowa

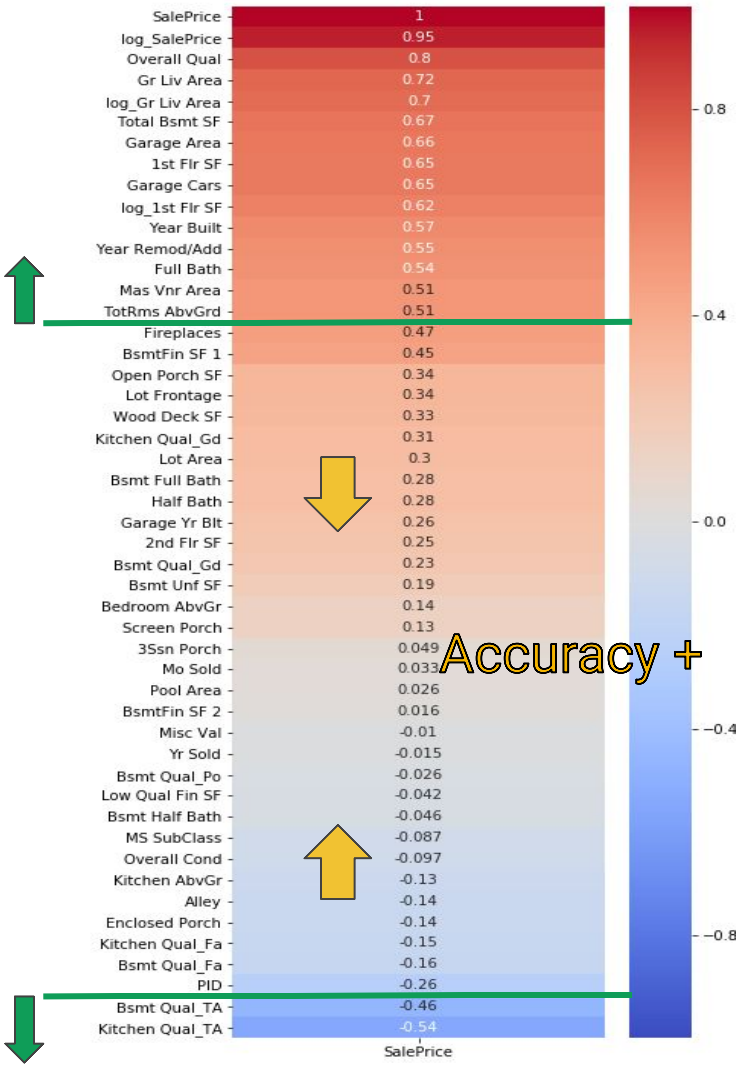
The data set contained information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames, IA from 2006 to 2010.

Hypothesis: A predictive price model can be built based on assessor data to predict the value and subsequent sale price of a home. The model will be accurate enough to be used in Open Door's business model. The model will work in new markets.

EDA

Through extensive investigation of the data the following observations were made:

1. Overall Linear relationships
2. High impact features to price:
    - Overall Quality
    - Total Sqft
    - Basement Sqft
    - Year Built
    - Year of remodel
    - Kitchen Quality
3. The addition of features (data) continued to improve the prediction accuracy
4. The maximum feature count was assumed to be 45



Predictive Model

Several linear regression models were applied in the case study:

- OLS
- LASSO
- Elastic Net

OLS Linear Regression: was the most accurate and yielded the least amount of error in the prediction of the true home price.

OLS Regression Model

How did the model do?

The optimized model predicts home prices within

± $15,600

Some technical jargon...

MAE: 15,608.65 (average price difference from the predicted price and actual price)

RMSE: 22,894.78 (average price difference from predicted price and actual price, includes larger misses)

R^2: 0.902 (goodness of the model fit)

CV: 0.893 (the goodness of fit over multiple samples)

Overall the model is performing well!

$$\frac{\text{Model Error}}{\text{Average Home Price}} = \frac{\$15{,}600}{\$180{,}190} = 8.6\%$$

On average, the model's price prediction is only ± 8.6% off from the true price.

Note: the predictions do vary more as the price surpasses $350,000, this is largely due to the limited data on homes above that price

Model Improvements

What improved the model?

The model improved with increase features and specific features (data such as # of full bathrooms)

Adding features in this fashion has a drawback, the model becomes less interpretable. It is harder to make statements about exactly what is determining the home price. Is the bedroom count or total square footage more impactful? It is tough to tell…

However, at Opendoor we care more about **accuracy than interpretability**. We need to know what to buy and sell the home for, the local sales agent will help sell the rest of the home's features.

Model Checks and Balances

What kind of system checks were put in place? Are these results reproducible?

Split Tests: The model went through 5 randomized samplings, this helps ensure the results of the sample are not due to random chance

Unknown Data: The model was evaluated against completely unknown data provided by Kaggle.com. The RMSE score (discussed earlier) was comparable.

Internal RMSE score: 22,894.78
Kaggle RMSE score: 19,213.10

What do we still need?

What would make the model better?

- Continuing to add data will improve the accuracy of the model.
- Try out different modeling types outside of the Linear Regression family

What additional data could we utilize if it was accessible

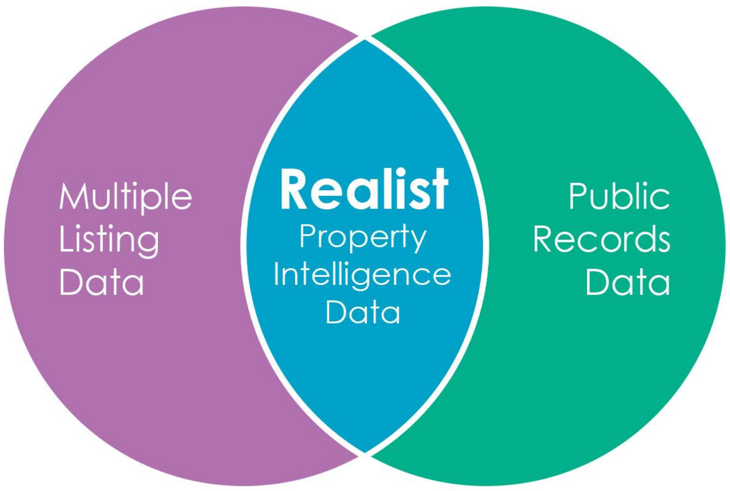
- More feature data pertaining to the quality of the home (these are high determiners of value)
- Local sales cycle data (this would add a time dimension to the model)

What data is already out there?

Realist Tax - data aggregator

- 4.5B + records spanning 50+ years
- > 99.9% of U.S. property records
- Assessor & tax data
- Building Permits

Standardizes the format for all States & Counties



Realist Tax a Corelogic product: https://www.corelogic.com/products/realist.aspx

Hypothesis Revisited

Part 1: A predictive price model can be built based on assessor data to predict the value and subsequent sale price of a home.

True - an OLS regression model gives a home price range of ± $15,600

Part 2: The model will be accurate enough to be used in Opendoor's business model.

True - homeowners are provided a range and the quality of the home is validated by a local agent

Part 3: The model will work in new markets.

True - using resources such as Realist Tax a similar model can be created for new markets

Further Research

A dataset from Realist Tax should be analyzed to ensure similar results are received.

Different modeling types (outside of Linear Regression) should be tested to see if they outperform the current model.

Research should be conducted in the new target city for days on market and overall market conditions, as these can greatly influence the ability to resale a home.

Questions