HOUSING PRICE PREDICTION

An Nguyen

Advisors: Chris Fernandes, Nick Webb & Harlan Holt
0. BACKGROUND

Features:
- Beds
- Baths
- Size

34 Shadow Wood Ln, Thomaston, GA 30286

Note: This property is not currently for sale or for rent. The description below may be from a previous listing.

A rare GEM located just minutes outside of town, this 4-sided brick ranch has it all! All interior finishes recently updated as well as a remodeled kitchen & master bath. Features include an open floor plan, hardwood floors, and a fully finished basement. 

Sold on 01/12/18

Zestimate: $184,777

Sold Price: $160,000
0. BACKGROUND

- **Zillow**
  - 26.2% market share
  - 110 million houses
  - Zestimate

- **Trulia**
  - 9.6% market share

- **Redfin**
  - 3.5% market share
1. PROBLEMS

- Zillow correctly estimates \(~50\%\) of their houses within \(5\%\) range of the actual sold price
2. QUESTIONS

- Can I get close to or beat the Zestimate?
1. PROBLEMS

- Zillow tends to overestimate their properties
2. QUESTIONS

- Can my models get rid of the overestimation problem?
1. PROBLEMS

- Do we need a lot of attributes to have a good prediction for house price?
2. QUESTIONS

- Most important attributes?
2. QUESTIONS

- Can I get close to or beat the Zestimate?
- Can my models get rid of the overestimation problem?
- Most important attributes?
2. DATA COLLECTION

Cowlitz, WA
29.3%

Montgomery, IL
8.7%

Cayuga, NY
16.7%

Hunt, TX
19.8%

Upson, GA
10.7%
2. DATA COLLECTION

- Cowlitz, WA: 354 houses
- Montgomery, IL: 209 houses
- Cayuga, NY: 399 houses
- Hunt, TX: 195 houses
- Upson, GA: 310 houses
2. DATA COLLECTION

- Sources: Zillow, Trulia, and Redfin
- Tools: Python, Selenium, and VBA
- Attributes:
  - Internal Factors: Beds, Baths, Size, Appliances, Garage, etc.
  - External Factors: Tax Info, School Info, Walkability, Nearby Lifestyle Amenities, Comparable Houses’ Sold Prices
215 P G Sweet Rd
Kelso, WA 98626

Redfin:
- Status: Sold
- Last Sold Price: $325,000
- Sold on 10/03/17

Zillow:
- Status: Off Market - Sold on 10/03/17
- Sold Price: $233,427

Trulia:
- Status: Sold
- Last Sold Price: $325,000

215 P G Sweet Rd
Kelso, WA 98626

3 beds • 1.5 baths • 1,375 sqft • 0.7 acres lot size • Single-Family Home

$233,427

Redfin

$325,000

Zillow

$235,000

Trulia

$233,427
3. MODELS

● **Linear Regression** (Baseline model):
  ○ Frequently used in Economics paper

● **Support Vector Regression** (**SVR**):
  ○ Good at finding signals and ignoring noises

● **Random Forest** (**RF**):
  ○ Good for datasets with missing values
3. RESULTS

Percentage Of Houses In 5% Range

- **My Data's Baseline**
- **SVR - Different Attributes**
- **Zillow's Countywide Baseline**

**Cayuga**:
- My Data's Baseline: 28.6
- SVR - Different Attributes: 17.3
- Zillow's Countywide Baseline: 16.7

**Cowlitz**:
- My Data's Baseline: 27.7
- SVR - Different Attributes: 23.5
- Zillow's Countywide Baseline: 29.3

**Hunt**:
- My Data's Baseline: 39.5
- SVR - Different Attributes: 42.7
- Zillow's Countywide Baseline: 19.8

**Montgomery**:
- My Data's Baseline: 21.1
- SVR - Different Attributes: 13.4
- Zillow's Countywide Baseline: 8.7

**Upson**:
- My Data's Baseline: 13.9
- SVR - Different Attributes: 13.2
- Zillow's Countywide Baseline: 10.7
### 3. RESULTS

<table>
<thead>
<tr>
<th>UPSON</th>
<th>COWLITZ</th>
<th>MONTGOMERY</th>
<th>HUNT</th>
<th>CAYUGA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bed</strong></td>
<td><strong>Bed</strong></td>
<td><strong>Bed</strong></td>
<td><strong>Size</strong></td>
<td><strong>Size</strong></td>
</tr>
<tr>
<td>Bath</td>
<td>Bath</td>
<td>Bath</td>
<td>Date Built</td>
<td>Tax Amount</td>
</tr>
<tr>
<td>Dishwasher</td>
<td>Asphalt Roof</td>
<td>Dishwasher</td>
<td>Tax Amount</td>
<td>Dishwasher</td>
</tr>
<tr>
<td>Assessment</td>
<td>Assessment</td>
<td>Assessment</td>
<td>Assessment</td>
<td>Assessment</td>
</tr>
<tr>
<td>Lot</td>
<td>Elementary-School Score</td>
<td>Size</td>
<td>Hardwood-Floor</td>
<td>Last Remodel-Year</td>
</tr>
<tr>
<td>Date Built</td>
<td>Size</td>
<td>Date Built</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walk Score</td>
<td>Walk Score</td>
<td>Walk Score</td>
<td>Walk Score</td>
<td>Walk Score</td>
</tr>
</tbody>
</table>

...
3. RESULTS

SAME ATTRIBUTE:

1. Bed
2. Bath
3. Dishwasher
4. Size
5. Tax Amount
6. Walk Score
7. Price Listed
8. Date Built
9. Assessment
10. Comparables’ Sold Price
3. RESULTS

Different Attributes vs Same Attributes

- My Data's Baseline
- SVR - Different Attributes
- SVR - Same Attributes

Cayuga: 28.6
Cowlitz: 27.7
Hunt: 39.5
Montgomery: 21.1
Upson: 13.9

SVR - Different Attributes

Cayuga: 17.3
Cowlitz: 23.5
Hunt: 42.7
Montgomery: 13.4
Upson: 13.2

SVR - Same Attributes

Cayuga: 15.5
Cowlitz: 23.7
Hunt: 54.6
Montgomery: 13.9
Upson: 10.6
3. RESULTS

Overestimated : Underestimated Ratio = 3:2  
Zillow

Overestimated : Underestimated Ratio = 1:1  
My Predictor
3. RESULTS

- Weights of Important Attributes Across 5 Counties:
  - $1 increase in Tax Assessment increases Sold Price by 54 cents
  - $1 increase in Comparables’ Sold Price increases Sold Price by 34 cents
  - $1 increase in Price Listed increases Sold Price by 38 cents
  - 1 more Bathroom increases Sold Price by $15,787
4. CONCLUSION

- Can I get close to or beat the Zestimate?
  - Beat Hunt’s accuracy score and come close to Cowlitz’s and Upson’s.
- Can my models get rid of the overestimation problem?
  - Reduce the overestimated to underestimated ratio from 3:2 to 1:1
- Most important attributes?
  - Tax Assessment, Comparables’ Sold Price, Price Listed, and Num of Bathrooms.
THANK YOU FOR LISTENING!