INSIDE THE RENT ZESTIMATES

UseR! 2016
June 28, 2016

Yeng Bun
yengb@zillowgroup.com
Flipping Houses

Flip

Buy

Sell
Flipping Houses

Flop

Cost
- Tax
- Interest
- Commission
- Fixup
- Purchase

Sell

Zillow Group Rentals
Land load
What is the rental price?

An Iterative Method

<table>
<thead>
<tr>
<th>Percent higher than Rent Zestimate</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>• 3 applications on the first day of listing</td>
</tr>
<tr>
<td>5%</td>
<td>• 2 applications within 2 weeks</td>
</tr>
<tr>
<td>10%</td>
<td>• 1 application within 5 weeks</td>
</tr>
<tr>
<td>7%</td>
<td>• 1 application within 3 weeks</td>
</tr>
</tbody>
</table>
My Rental Estimate = rz[t] * (1 + \varepsilon)

Where \( rz[t] \) is Rental Zestimate for a property at time \( t \)
and \( |\varepsilon| \ll 1 \)
EX: \( \varepsilon = \text{median}(0.05, 0.10, 0.07) = 0.07 \)
Rent Zestimate Engine

Data from multiple sources:
- Physical attributes
- Rental listings
- User updates
- Home values (Zestimates)

CLEANING
- Reconciling property attributes

TRAINING
- Models trained with recent rental listings

SCORING
- Models applied to all homes every day

**RENT ZESTIMATE:**
- Value: $1,450/mo
- Range: $1,300 - $1,600K
Map Reduce vs. Zillow Parallel Library (ZPL) for R

MapReduce

- **Input**
- **Head Node**
- **Map**
- **Reduce**
- **Worker Node 1**
- **Worker Node 2**
- **Worker Node N**
- **Output**

ZPL

- **Input**
- **Head Node**
- **zplOnCompute()**
- **Update Node**
- **zplOnUpdate()**
- **zplOnUpdate()**
- **Worker Node 1**
- **Worker Node 2**
- **Worker Node N**
- **Output**

Hadoop

RDBMS
Data Partitioning and Parallel Computing

Head Node

Parallel R Worker Nodes

Update Node

Task Queue

Input Database

Combine Data

Output Database
Accuracy Measurement

Hold-out 30% of training data

If \( rz \) are the Rent Zestimates for homes in the hold-out dataset, then the percent estimated errors are

\[
e = 100 \times \frac{(rz - r)}{r},
\]

where \( r \) are the actual rental listing prices.

Key metrics:

Median (\(|e|\))

Percent of Rent Zestimates within \( x \)% of rent price

\[
\frac{100}{N} \sum_{i=1}^{N} ( |e_i| < x ),
\]

where \( x \% : 5\%, 10\% \) and 20\% and \( N \) is number properties in the hold-out set.
Commit to transparency


<table>
<thead>
<tr>
<th>Location</th>
<th>Rent Zestimate Accuracy</th>
<th>Homes on Zillow</th>
<th>Homes with Rent Estimates</th>
<th>Within 5% of Rent Price</th>
<th>Within 10% of Rent Price</th>
<th>Within 20% of Rent Price</th>
<th>Median Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portland, OR</td>
<td>★★★</td>
<td>1.6M</td>
<td>1.3M</td>
<td>39.3%</td>
<td>64.0%</td>
<td>83.4%</td>
<td>7.0%</td>
</tr>
<tr>
<td>Riverside, CA</td>
<td>★★★</td>
<td>791.0K</td>
<td>708.9K</td>
<td>39.6%</td>
<td>65.7%</td>
<td>86.8%</td>
<td>6.5%</td>
</tr>
<tr>
<td>Sacramento, CA</td>
<td>★★★</td>
<td>827.7K</td>
<td>746.3K</td>
<td>40.2%</td>
<td>64.4%</td>
<td>85.3%</td>
<td>6.8%</td>
</tr>
<tr>
<td>San Antonio, TX</td>
<td>★★★</td>
<td>906.1K</td>
<td>874.4K</td>
<td>35.0%</td>
<td>58.4%</td>
<td>80.8%</td>
<td>7.9%</td>
</tr>
<tr>
<td>San Diego, CA</td>
<td>★★</td>
<td>1.3M</td>
<td>1.3M</td>
<td>31.0%</td>
<td>55.2%</td>
<td>78.2%</td>
<td>8.7%</td>
</tr>
<tr>
<td>San Francisco, CA</td>
<td>★★</td>
<td>1.3M</td>
<td>1.2M</td>
<td>34.0%</td>
<td>58.4%</td>
<td>83.2%</td>
<td>8.1%</td>
</tr>
<tr>
<td>Seattle, WA</td>
<td>★★</td>
<td>1.2M</td>
<td>1.1M</td>
<td>33.7%</td>
<td>56.1%</td>
<td>82.5%</td>
<td>8.3%</td>
</tr>
<tr>
<td>St. Louis, MO</td>
<td>★★</td>
<td>1.2M</td>
<td>1.1M</td>
<td>39.3%</td>
<td>63.3%</td>
<td>85.9%</td>
<td>7.0%</td>
</tr>
<tr>
<td>Tampa, FL</td>
<td>★★</td>
<td>1.3M</td>
<td>1.2M</td>
<td>42.4%</td>
<td>67.7%</td>
<td>87.7%</td>
<td>6.2%</td>
</tr>
<tr>
<td>Washington, DC</td>
<td>★★</td>
<td>1.9M</td>
<td>1.9M</td>
<td>42.4%</td>
<td>67.7%</td>
<td>87.7%</td>
<td>6.2%</td>
</tr>
</tbody>
</table>

**Definitions**

- **ZESTIMATE ACCURACY/STAR RATING:**
  - This rating is tied to the Median Error in an area.
  - The ratings are as follows:
    - ★★★★ = Best Rent Zestimate
    - ★★★ = Good Rent Zestimate
    - ★★ = Fair Rent Zestimate
    - ★ = Unable to compute Rent Zestimate accuracy
    - 0 stars = No valuation
Monthly Accuracy: box plot of |Pct Error| by month
Systematic Error: box plot of Pct Error by month.
Key Takeaways

• **Rent Zestimate**
  – A starting point in determining the monthly rental price for a specific property.
  – Best tool for landlords, tenants, investors,…

• **Serious production software in R ?**
  – Yes
We are hiring
http://www.zillow.com/jobs/