New York Eateries: Wisdom of the Crowds?

Jenna Green and Jacob Mundt
{jennargreen, jake.mundt} @ gmail.com
Data Overview – Yelp Reviews

- 100,000s of reviews, 5 years of history
- 1 to 5 stars, plus text review
- User-submitted data
  - Price range/attire
  - Category
  - Alcohol service
  - Ambiance
- Limit analysis Columbia neighborhood
Data Overview – Yelp Reviews

Yelp Rating by Date

![Graph showing Yelp ratings over time](image-url)
Experts vs. the Crowds

• Crowdsourcing: CHEAP, but how objective?
• Actual food is too subjective—even paid critics disagree.

• Gold-standard evaluation metric. Must be:
  • Objective
  • Evaluated by expert
  • Something also perceptible to the average customer
Experts vs. the Crowds

Evaluation metric
The Health Department uses regular (1-3 times a year) restaurant inspections to make sure businesses are maintaining food safety standards.

In addition to demographic info, this data includes:

- Date/time of inspection
- Specific violation descriptions
- Total violation score (between 0 and 80, low is good)
# Data Overview – Most common violations

<table>
<thead>
<tr>
<th>Violation</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facility not vermin proof.</td>
<td>903</td>
<td>10%</td>
</tr>
<tr>
<td>Cold food item held above 41º F</td>
<td>718</td>
<td>8%</td>
</tr>
<tr>
<td>Plumbing not properly installed or maintained...equipment or floor not properly drained; sewage disposal system in disrepair or not functioning properly.</td>
<td>508</td>
<td>6%</td>
</tr>
<tr>
<td>Food service operation occurring in room used as living or sleeping quarters.</td>
<td>475</td>
<td>5%</td>
</tr>
<tr>
<td>Non-food contact surface improperly constructed.</td>
<td>474</td>
<td>5%</td>
</tr>
<tr>
<td>Unacceptable material used.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Overview – Violations by Cuisine

- Bakery
- Sandwich
- Ice Cream
- Hamburger
- Donuts
- Other
- Soul Food
- Vegetarian
- Greek
- Spanish
- Mediterranean
- Cafe/Co
- Delicate
- American
- Bagels/Pizza
- Seafood
- Korean
- Tex-Mex
- Indian
- Asian
- Mexican
- French
- Japanese
- Eastern
- Italian
- African
- Thai
- Chinese
Since July 2010, have also required restaurants to very prominently display an overall Grade “A”, “B” or “C” to customers.

Can restaurant attributes (parking, coat check, neighborhood) predict its inspection violation score?
No.

• The range of the response variable, Score, is 0 to 80 (score and grade are the same info)
• In trying to fit a simple linear regression, the best $r^2$ value was 0.2208
• Failure to fit a line was easily foreseeable based on some of the exploratory data analysis:
Neighborhood

Full Bar, Beer/Wine, No Bar

Price: $, $$, $$$
Variance in Score within Business

- Ollie’s:

<table>
<thead>
<tr>
<th>Date Of Inspection</th>
<th>Total Inspection Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/4/2009</td>
<td>36</td>
</tr>
<tr>
<td>2/26/2009</td>
<td>12</td>
</tr>
<tr>
<td>6/22/2009</td>
<td>22</td>
</tr>
<tr>
<td>4/5/2010</td>
<td>74</td>
</tr>
<tr>
<td>4/28/2010</td>
<td>54</td>
</tr>
<tr>
<td>5/20/2010</td>
<td>30</td>
</tr>
<tr>
<td>6/4/2010</td>
<td>15</td>
</tr>
<tr>
<td>2/14/2011</td>
<td>36</td>
</tr>
<tr>
<td>4/12/2011</td>
<td>51</td>
</tr>
<tr>
<td>9/6/2011</td>
<td>11</td>
</tr>
</tbody>
</table>
A Model for predicting Yelp ratings

Base model: predict the median (4 stars)
• Precision/Recall: .124/.352
• Correctly classifies 35% of instances

Decision tree with all business attributes
• Precision/Recall: .378/.386 (10-fold cross validation)
• Correctly classifies 39% of instances
• Adding recent inspection violation count as prediction variable does NOT improve accuracy
The worst of the worst
Continuing/Future work

• Restaurant inspection ratings don’t predict yelp reviews, unless ratings are posted
• What about other direction? Predict future inspection violations based on mining text from yelp reviews
  • Inspections are only a few times a year—just like the Google Flu trends project, this might help uncover dangerous health situations before they affect population at large
• Does this model generalize to the rest of NYC? other cities? California also requires posting of grades.