Effects of Restaurant Neighborhood and Cuisine on Rating

Pradeep Gopinathan

Why Analyze Restaurant Ratings?

- Cites are expensive and competitive places to open restaurants.
- Location will determine clientele, which determines demand.
- Knowing demand may provide light into what restaurant might do well.

Can We Predict A Restaurant Rating?

Data

- Restaurants in and around Toronto

- Venues and venue ratings acquired through Foursquare API.
  - Only venues that Foursquare has a rating for.

- Treat prediction as classification problem.
  - Convert ratings into discrete classes: 1-10.

Data Observations

- Less than half of venues in Foursquare had ratings.
- Venue distribution heavily skewed in ratings, locations, and categories:

Rating	Count
8.0	271
7.0	260
6.0	258
9.0	83
5.0	42
4.0	2

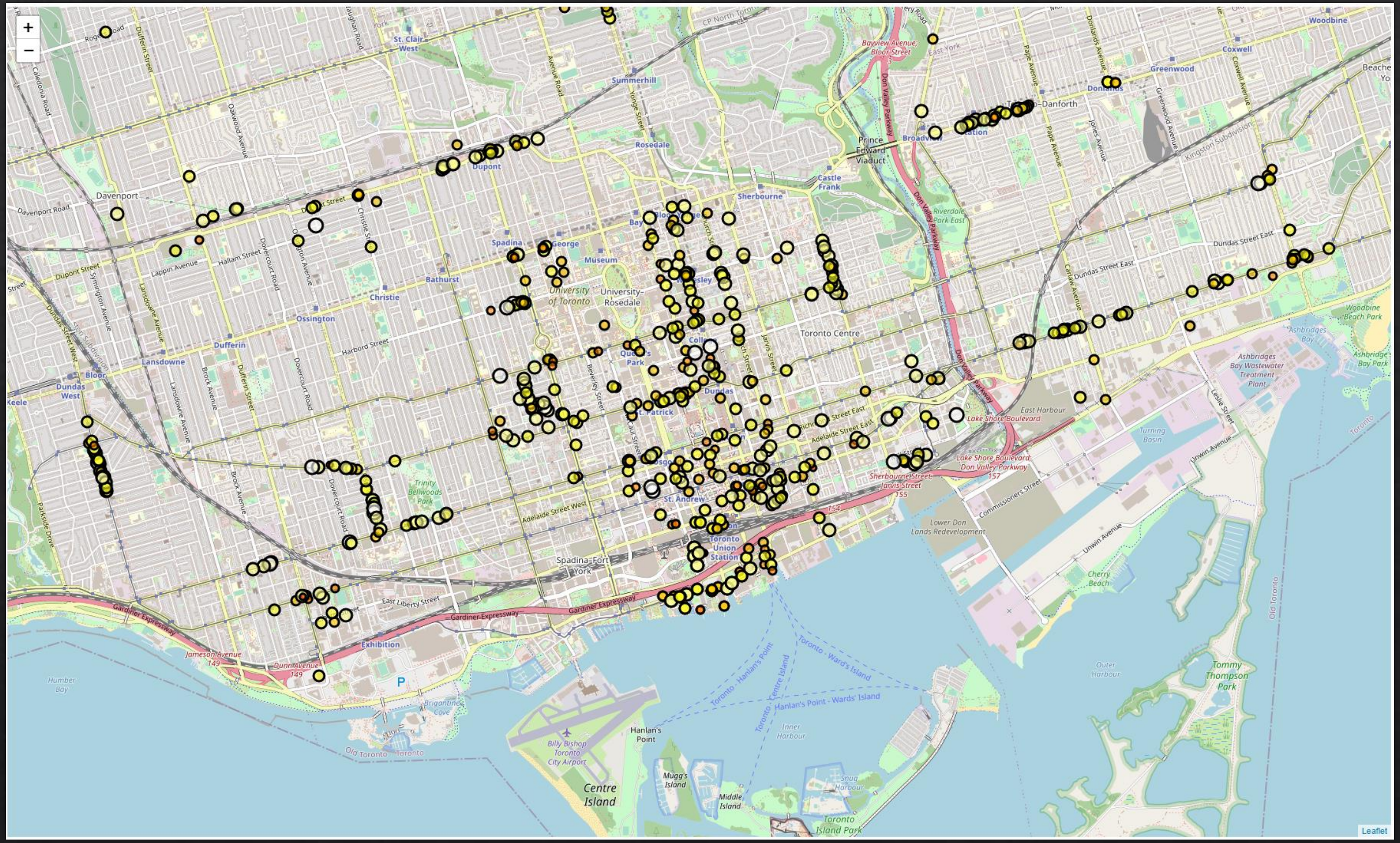
Venue Category	Count
Coffee Shop	156
Café	52
Sandwich Place	50
Restaurant	48
Fast Food Restaurant	46
...	...
Cajun / Creole Restaurant	1
Hakka Restaurant	1
Noodle House	1
Food & Drink Shop	1
Eastern European Restaurant	1

Neighborhood	Count
Harbourfront East, Union Station, Toronto Islands	58
St. James Town	35
Queen's Park, Ontario Provincial Government	34
Kensington Market, Chinatown, Grange Park	34
Garden District, Ryerson	34
...	...
The Kingsway, Montgomery Road, Old Mill North	1
Stn A PO Boxes	1
Islington Avenue, Humber Valley Village	1
Scarborough Village	1
Milliken, Agincourt North, Steeles East, L'Amoreaux East	1

Restaurants and Ratings:

Restaurants in data set plotted around Toronto.

The larger and darker circles are higher rated restaurants.



Predictive Modeling

- Two approaches:
  - Decision Trees
  - K-Nearest Neighbors

- Naïve approaches to measure against:
  - Guess 1-10 randomly
  - Guess 7 (average rating)
  - Guess 8 (most common rating)

Predictive Modelling Results

Approach	Accuracy	Optimal Parameters
Naïve Random (guess 1-10)	.10	N/A
Naïve Average (guess 7)	.283	N/A
Naïve Average (guess 8)	.296	N/A
Decision Tree	.363	Gini with max depth 2
k-Nearest Neighbors	.305	Uniform weight with 19 neighbors

Discussion

- Models have better accuracy, but only slightly.

- Many limitations in data:
  - Less than half of the restaurant had ratings
  - Not all ratings had exemplars in data
  - Few ratings for certain restaurant types and few ratings for certain neighborhoods

Conclusion

- Decision tree modelling appears to have some promise, but more research needs to be done.

- More data to be gathered:
  - Use alternate sources for restaurants/ratings: Yelp, Facebook, etc.
  - Replace neighborhood feature with quantitative features of neighborhood (age distribution, per-capita income, etc.)
    - Can introduce other cities into dataset
    - Can get better understanding of what effects a neighborhood might have on the demand for restaurants.