Categorizing Neighborhood Restaurant Data in Toronto

By: Alex Mason

Why Categorizing Neighborhood Data is Valuable

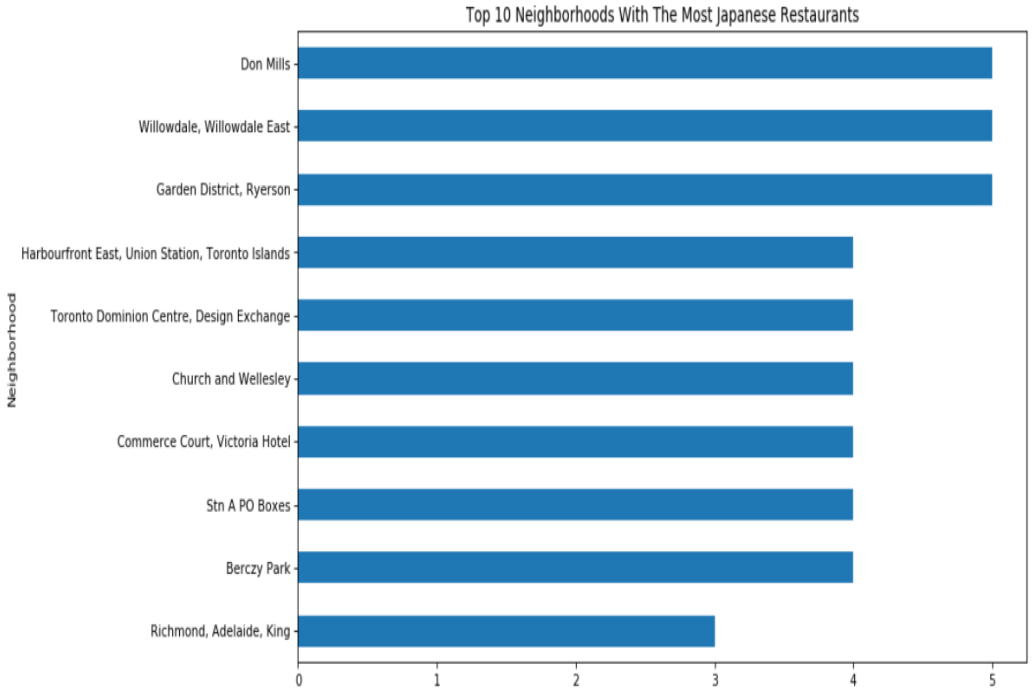
- Restaurants are a very appealing market for cities especially as tourists attractions and as a display of Culture
- Categorizing preferences of different areas with a city can be useful for home renters and buyers deciding on potential neighborhoods to live in.
- Understanding consumer preferences helps restaurant owners and potential owners decide where to strategically place their restaurants and the preferences for restaurant menus and dishes to better help their businesses thrive and grow.

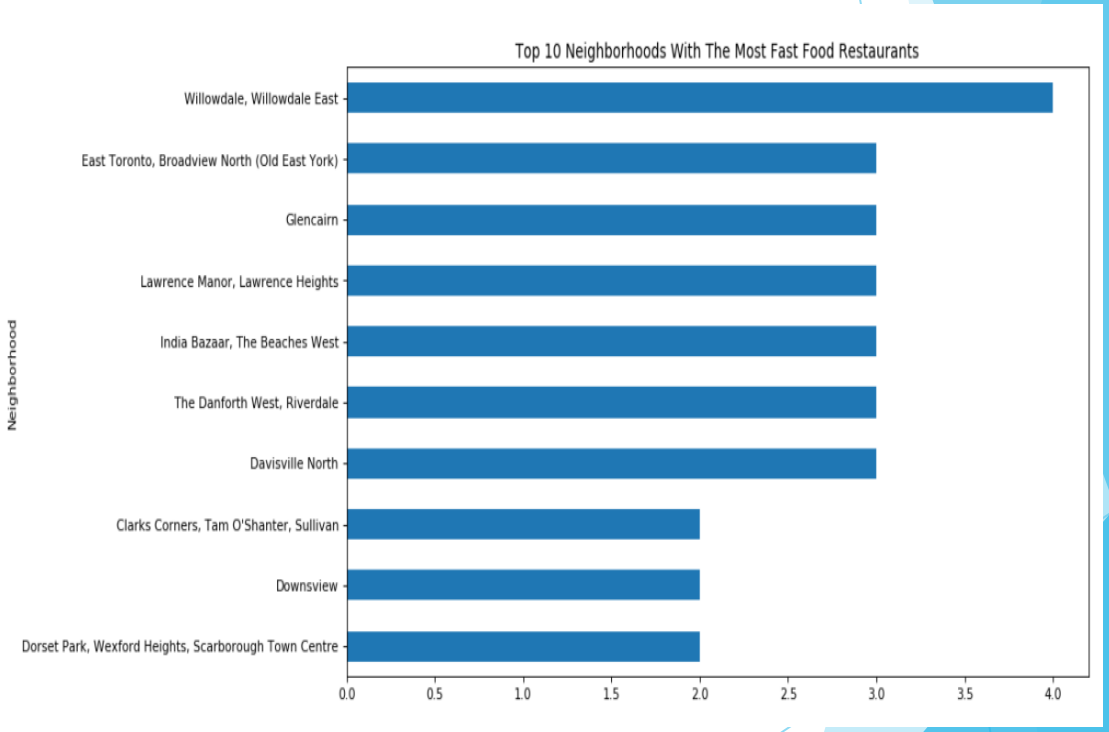
Data Resources and Cleansing

- Wikipedia Toronto Postal Code Data: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- IBM Geospatial Toronto Location Data: http://cocl.us/Geospatial_data
- Foursquare Venue Categories: https://developer.foursquare.com/docs/build-with-foursquare/categories/

- Columns with ‘Not Assigned’ boroughs or neighborhoods were dropped from the dataframe.
- Toronto postal code data and geospatial datasets merged into one dataframe for better easy of access and to ensure data matches.

Data Visualization for Some of the Top Restaurant Categories in Each Neighborhood

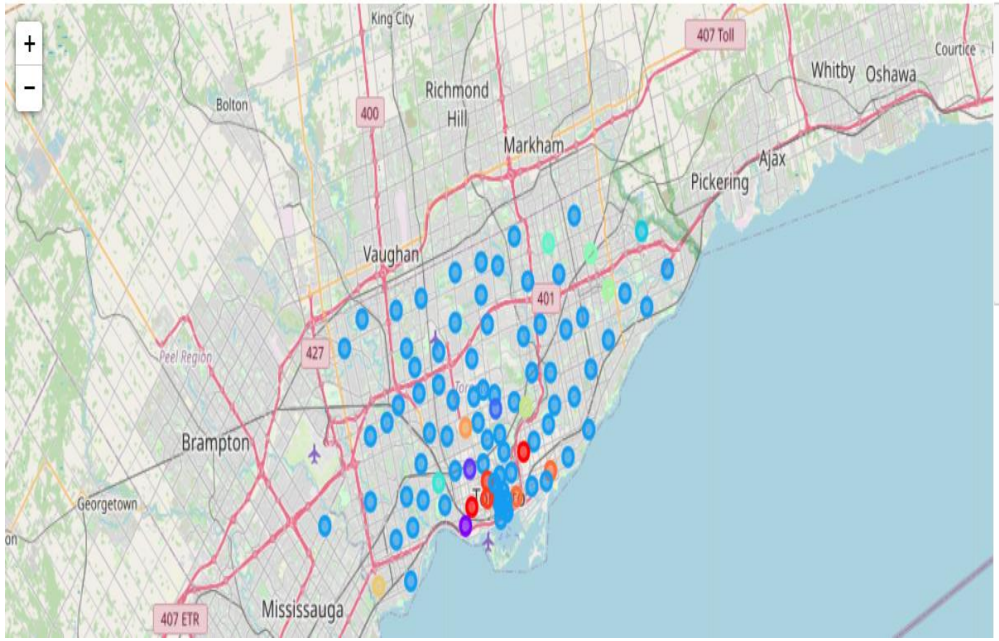




Top 5 Neighborhood Restaurant Venue Categories

	Neighborhood	1st Top Venue Category	2nd Top Venue Category	3rd Top Venue Category	4th Top Venue Category	5th Top Venue Category
0	Agincourt	Sri Lankan Restaurant	Malay Restaurant	Filipino Restaurant	Cantonese Restaurant	Chinese Restaurant
1	Alderwood, Long Branch	Moroccan Restaurant	Ramen Restaurant	Vegetarian / Vegan Restaurant	Taiwanese Restaurant	Filipino Restaurant
2	Bathurst Manor, Wilson Heights, Downsview North	Mediterranean Restaurant	Middle Eastern Restaurant	Sushi Restaurant	Restaurant	Ramen Restaurant
3	Bayview Village	Japanese Restaurant	Chinese Restaurant	Restaurant	Korean Restaurant	Vegetarian / Vegan Restaurant
4	Bedford Park, Lawrence Manor East	Comfort Food Restaurant	Greek Restaurant	Italian Restaurant	American Restaurant	Indian Restaurant

Kmeans Restaurants Cluster Map of Toronto




```
kclusters = 15 # num of clusters

venue_grouping = venue_mean.drop('Neighborhood', 1)
kmeans = KMeans(n_clusters = kclusters, random_state=0).fit(venue_grouping)
 #display clusters in array
kmeans.labels_
```

```
array([ 8, 11,  4,  4,  4,  4,  4,  1,  4,  4,  4,  9,  4,  2,  4,  4,  4,
        3,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,  6,
        4,  4, 12, 13,  4, 14,  4,  4,  4,  0,  5,  4,  4,  4,  4,  4,  4,
        4,  4,  4,  4,  4,  4, 13,  4,  4,  4,  4,  4,  4,  4,  4,  4,  4,
        7,  4,  4,  4,  4,  4,  0,  4, 10,  4, 14,  4,  4,  4,  4,  4,  4,
        4,  4,  4,  4,  4])
```

Conclusion and Future Directions

- Very similar cluster patterns for food venue categories throughout the city of Toronto (not a large preferences variance)
- Foursquare's 'Place API' helps leverage important multi-dimensional datasets that can help us better determine patterns and similarities within our dataframes.
- Foursquare's API could be further integrated and updated to help better determine proper 'restaurant-related' venues.
- Would be interesting to test K-Means clustering algorithm to DBSCAN to see which gives a better performance and categorization of our restaurant data.