

Data:

As discussed in the previous section on what the problem is, we will need to analyze what makes a neighborhood a livable neighborhood by categorizing them into different clusters. The initial data we will need will be a table of neighborhoods and their postal codes along with the borough that they belong to within Toronto. This data can be easily found on Wikipedia¹.

This data only contains the postal codes, neighborhoods and boroughs though, so to do a proper analysis we will need to somehow grab a list of venues that are nearby to the postal code's locations. Luckily, Foursquare has this data readily available, but before we can grab this data, we need the latitude and longitude coordinates, which we can grab from the python geocoder² module. Alternatively, if the geocoder module does not work, a csv with the latitude and longitude locations of all the postal codes in Toronto was provided to us by the IBM team at coursera³, so we elected to use that data since, in our case, the geocoder module did not work properly.

	Postal Code	Borough	Neighbourhood	Latitude	Longitude
0	M3A	North York	Parkwoods	43.753259	-79.329656
1	M4A	North York	Victoria Village	43.725882	-79.315572
2	M5A	Downtown Toronto	Harbourfront	43.654260	-79.360636
3	M6A	North York	['Lawrence Heights', 'Lawrence Manor']	43.718518	-79.464763
4	M7A	Queen's Park	Queen's Park	43.662301	-79.389494

An example of the database after latitude/longitude values have been added

Once all the latitude and longitude coordinates were in the database, we could do a call to the Foursquare API to collect a list of attractions that were within a certain set radius (in our case, 250 meters). Once this data was collected, we stuck it all into a pandas data frame and started our analysis.

Neighbourhood	Neighbourhood Latitude	Neighbourhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Parkwoods	43.753259	-79.329656	Brookbanks Park	43.751976	-79.332140	Park
Parkwoods	43.753259	-79.329656	Variety Store	43.751974	-79.333114	Food & Drink Shop
Victoria Village	43.725882	-79.315572	Victoria Village Arena	43.723481	-79.315635	Hockey Arena
Victoria Village	43.725882	-79.315572	Tim Hortons	43.725517	-79.313103	Coffee Shop
Victoria Village	43.725882	-79.315572	Portugril	43.725819	-79.312785	Portuguese Restaurant
Victoria Village	43.725882	-79.315572	Eglinton Ave E & Sloane Ave/Bermondsey Rd	43.726086	-79.313620	Intersection

An example of the database after venue information was added

¹ https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

² <https://geocoder.readthedocs.io/>

³ <https://www.coursera.org/learn/applied-data-science-capstone/home/week/3>

