Moving to Toronto: A Study in the Livability Factor of the City

Problem:

When moving to a new city, it is imperative that one understands what kind of neighborhoods have what amenities nearby. The nearby amenities will be shaped by several factors, namely size of city, various city laws, and geographic and national location of the city. As such, I have chosen the city of Toronto with an intent to analyze where the best place to live in the city would be based on a certain set of criteria. I believe this report will help those who are interested in moving to a Canadian city make a better choice as to where to live.

Hypothesis:

When presented with the possibility of moving to Toronto, one must think about some of the underlying motivations that may be present for moving to the city. My assumption that I will use to test livability will be that people moving to a city in Canada are looking for a closer connection to nature, due to Canada's reputation for vast untamed wildernesses. As such, I will be looking for neighborhoods in Toronto that have a close proximity to a natural area, such as a park or river.

Data:

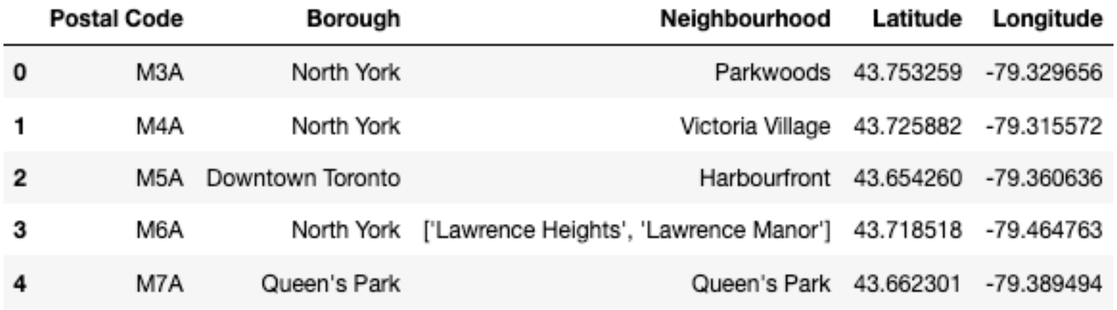
As discussed in the previous section on what the problem is, we will need to analyze what makes a neighborhood a livable neighborhood by categorizing them into different clusters. The initial data we will need will be a table of neighborhoods and their postal codes along with the borough that they belong to within Toronto. This data can be easily found on Wikipedia[1].

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 co-ordinates, which we can grab from the python geocoder[2] 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 coursera[3], so we elected to use that data since, in our case, the geocoder module did not work properly.

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

[2] https://geocoder.readthedocs.io/

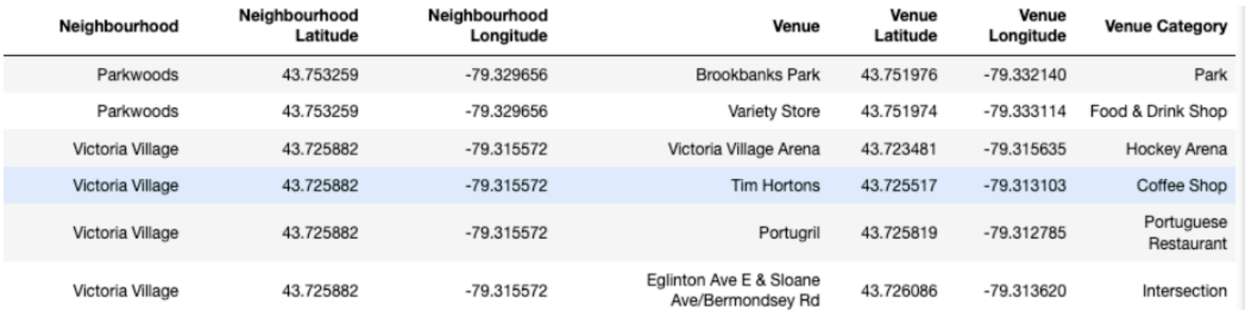
[3] https://www.coursera.org/learn/applied-data-science-capstone/home/week/3



	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

Fig1, 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

Fig2, An example of the database after venue information was added

Methodology:

Much of the exploratory data analysis that we did was performed in the data collection stage. While collecting data, we would continuously look to see what the shapes of the data were like, what data types that we had, and how we could use this.

When first obtaining the data from the wikipedia page, the first thing we noticed was that all postal codes shared the same borough, even if they had differing neighbourhoods. As an example, there were two entries for postal code M6A, both of which shared the same borough of “North York”, though they each had separate neighbourhoods, namely “Lawrence Heights” and “Lawrence Manor”.

After we had done some data grabs from a csv of coordinate data shared by the IBM team at coursera, we put all the coordinate data together with the neighbourhoods, and then used folium to visualize the location of each borough, giving us a nice map (fig3).

This map gave a nice graphical representation of where our data lies. Once we had all this data, we needed to move onto the next part, namely grabbing locational

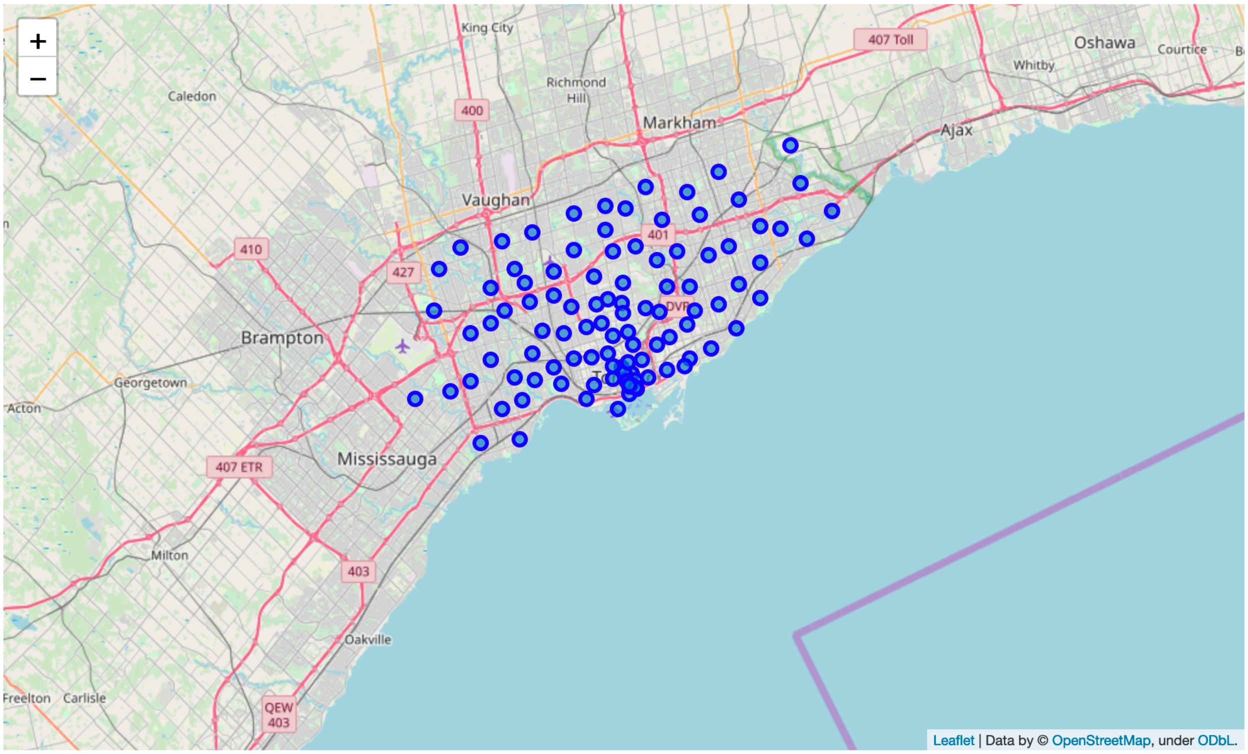


Fig3, a map with blue dots representing the lat/lon coordinates of every borough

data of different venues around each borough. We did a call to foursquare for each location to get the various venues located around each location. We limited the search radius to 250 meters so as to prevent as much overlap as possible while still getting a fairly robust search area.

Next up was to create a data frame that contained all the locations with a list of the top ten venues sorted by distance, along with what type of venue it was. This data was important for the KNN model we would use to sort the locations. We sorted this data frame and encoded all the venues using one-hot encoding such that our KNN model could intake the data.

From here, we put the data into a KNN algorithm, and it sorted the venues into five different categories.

Results:

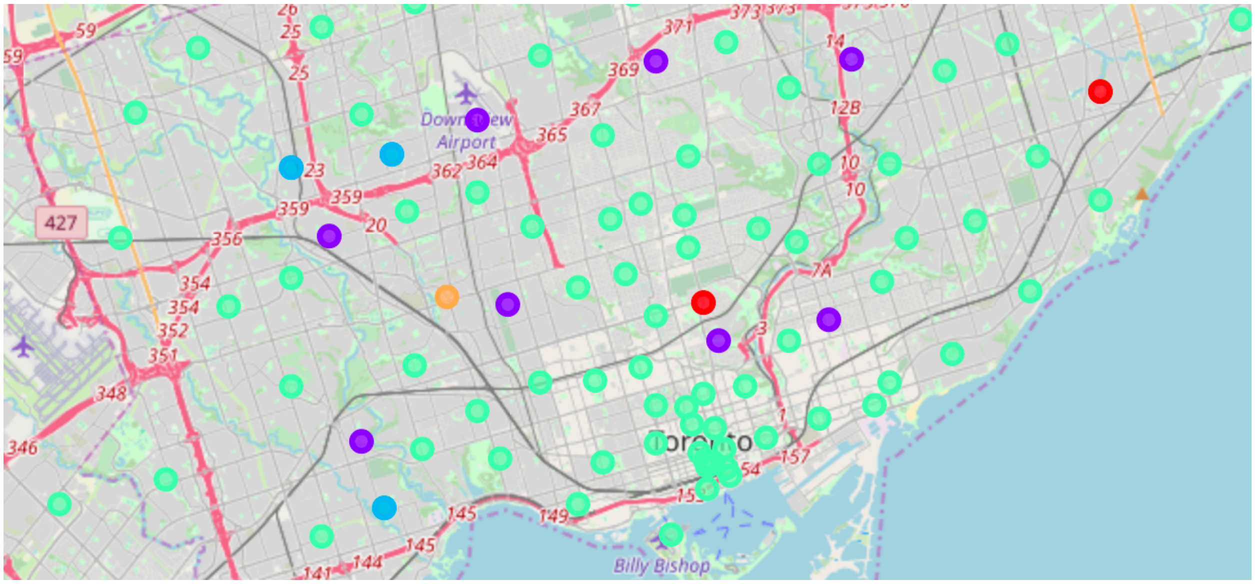
In the five categories that the KNN algorithm gave us, we were highly interested in two of those five categories. We were interested in category 1 and category 3. These two categories had high groupings of locations that were very similar, and as such were ripe for analyzing what made them unique.

Cluster label 3 gave us many results where the most common venue would fall into a category that one might call “mall locations”. This includes coffee shops, fast food chains, and stores. While this could be a livable location, it is more reminiscent of an American city, and as such does not fall under our definitions of reasons to move as stated in the hypothesis.

Cluster label 1 is are the results we actually want to be looking at. In cluster label one, the majority of most common venue results fall under the categories of either parks, rivers, or some other form of public area. These are the areas that we are most interested in, and would be the areas that we would recommend moving to.

Discussion:

There is one thing about cluster label 1 that does warrant some discussion. Pictured below is a snapshot of the map of Toronto, with the dots representing the "center" of the different neighbourhoods/boroughs that we are looking at.



Purple is cluster 1, Green is cluster 3, and the other 3 clusters are denoted by the other 3 color. As can be seen on the map, most of the purple dots are near large public parks and/or rivers. However, there is one glaring exception, and that is the purple dot on Downsview Airport. While there is the large Downsview park nearby, the closest location to this is the Downsview airport. I believe that what cluster 1 was truly measuring was whether something was a "public area" or not, and as such an airport is a public area. While most public areas it would seem are parks/trails/river areas, exceptions can occur, and this is one of those notable exceptions.

As far as recommendations for living spaces go, we would recommend neighbourhoods such as Parkwoods, The Kingsway, Montgomery North, Old Mill North, and Rosedale.

Conclusion:

In conclusion, we have found through our KNN Algorithm that the neighborhoods in "cluster 1" as labeled by the algorithm present the best qualities for neighborhoods to move to, provided that your reason for moving to a Canadian city is that you want to be closer to nature, with a good selection of public parks around you.