Capstone Project - The Battle of Neighborhoods

Mateus Teruyuki Nakata

04/11/2020

Introduction

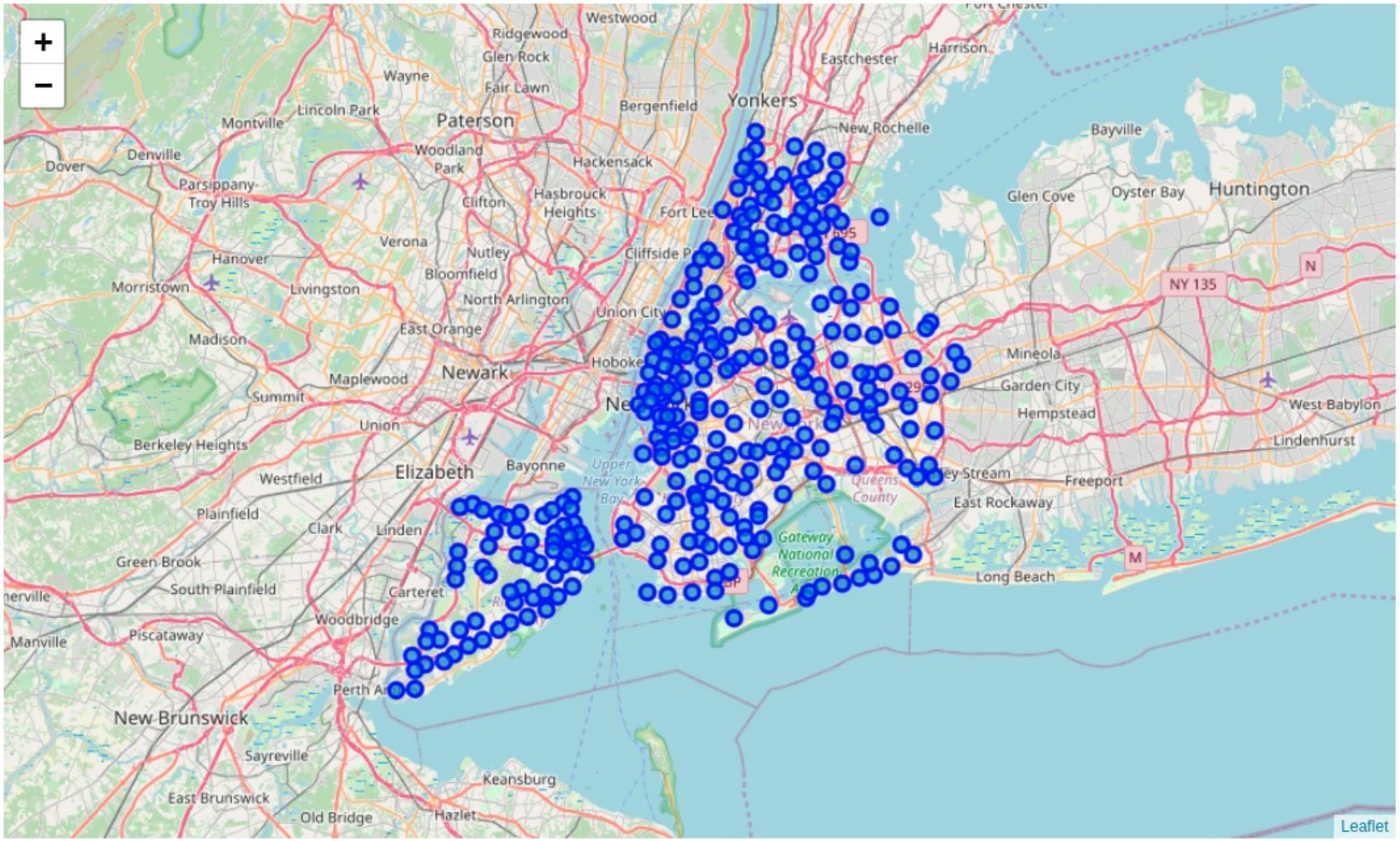
In a globalized world, It is very common for a company to transfer an employee to another city or even country, offering better career prospects or salary increase. Or someone could find a better job in other company in other city. No matter the reasons, moving to other city is always very stressful, and people are afraid about not get used with the new city . The problem is worse when the person is married with children. To minimize the problems, it is better to move a place similar to his last location, with compatible venues, like schools, restaurants, swimming pools, gyms, coffee-shops, supermarkets, etc. To help people in this situation, the goal of this project is to develop a system to find out which neighborhoods are similar to the current location. We will simulate a situation where a person is moving from New York to Toronto and vice versa.

Data Preparation

- New York
  - 5 boroughs and 306 neighborhoods
  - dataset:https://geo.nyu.edu/catalog/nyu_2451_34572
  - Use of Pgeocode to get the latitude and longitude

	Postcode	Borough	Neighborhood	Latitude	Longitude
83	M6R	West Toronto	Parkdale / Roncesvalles	43.6469	-79.4521
84	M6S	West Toronto	Runnymede / Swansea	43.6512	-79.4828
85	M7A	Downtown Toronto	Queen's Park / Ontario Provincial Government	43.6641	-79.3889
86	M7R	Mississauga	Canada Post Gateway Processing Centre	43.6370	-79.6158
87	M7Y	East Toronto	Business reply mail Processing CentrE	43.7804	-79.2505
88	M8V	Etobicoke	New Toronto / Mimico South / Humber Bay Shores	43.6075	-79.5013
89	M8W	Etobicoke	Alderwood / Long Branch	43.6021	-79.5402
90	M8X	Etobicoke	The Kingsway / Montgomery Road / Old Mill North	43.6518	-79.5076
91	M8Y	Etobicoke	Old Mill South / King's Mill Park / Sunnylea /...	43.6325	-79.4939
92	M8Z	Etobicoke	Mimico NW / The Queensway West / South of Bloo...	43.6256	-79.5231

Visualization

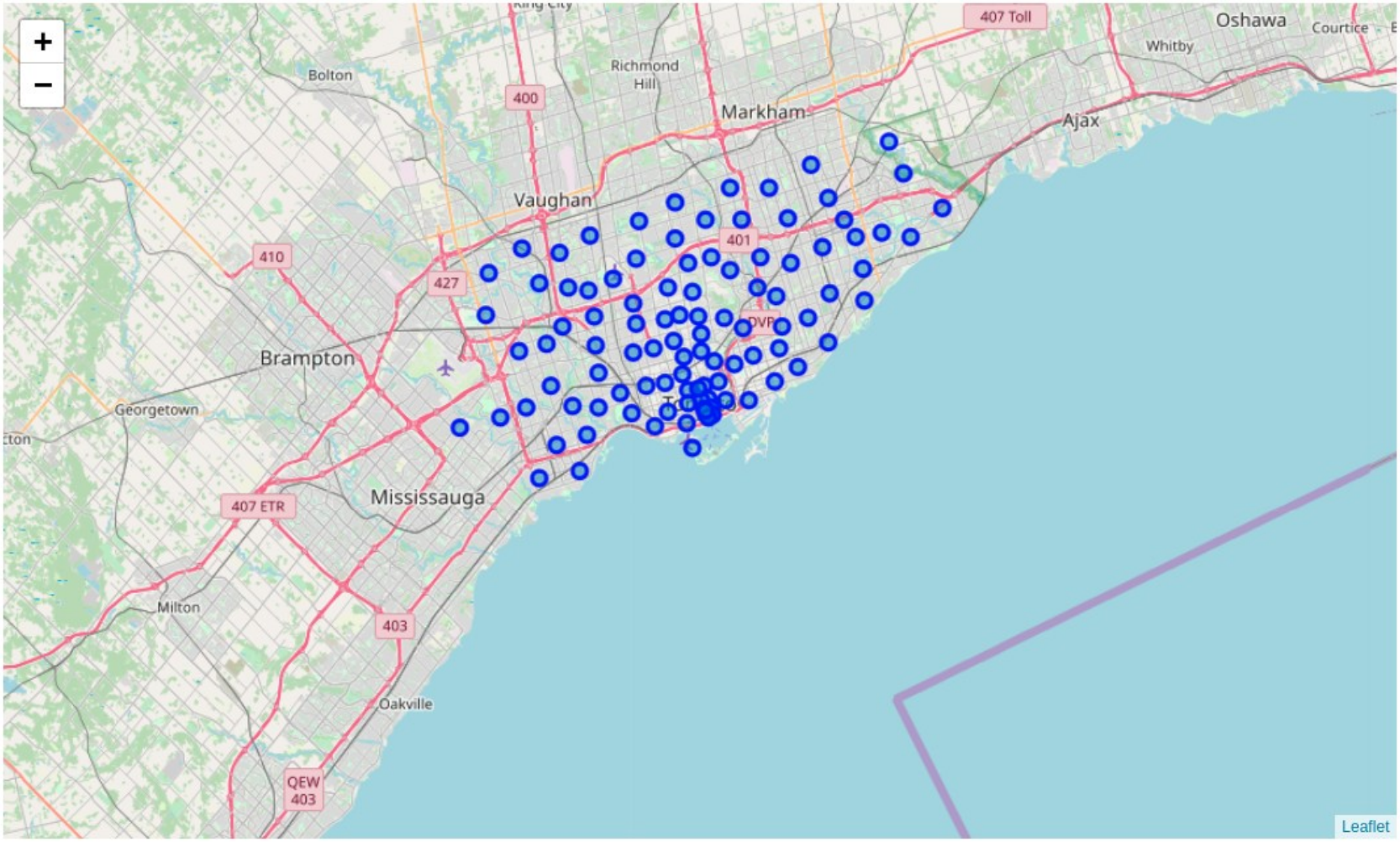


Data Preparation

- Toronto
  - 10 boroughs and 103 neighborhoods
  - Wikipedia page, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - json format and it can be transformed into pandas dataframe

	Borough	Neighborhood	Latitude	Longitude
0	Bronx	Wakefield	40.894705	-73.847201
1	Bronx	Co-op City	40.874294	-73.829939
2	Bronx	Eastchester	40.887556	-73.827806
3	Bronx	Fieldston	40.895437	-73.905643
4	Bronx	Riverdale	40.890834	-73.912585

Localization



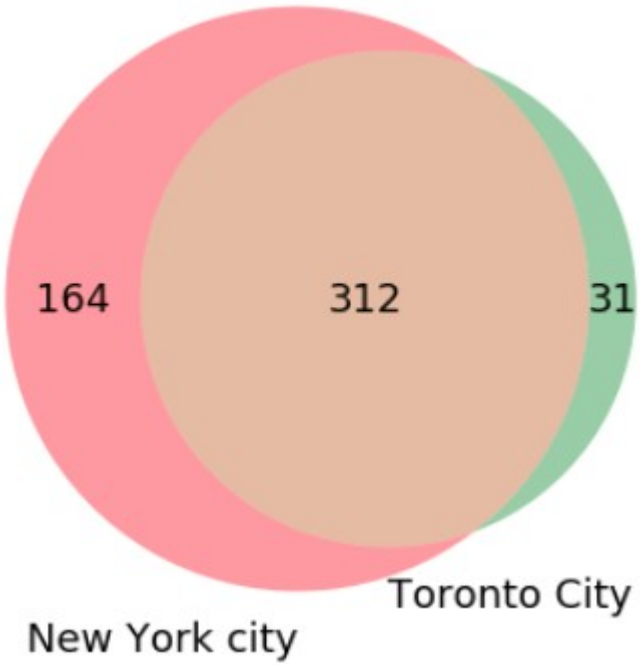
Foursquare to Get Venue Data

- Free online API access
- Information about the location venues

	Borough	Neighborhood	Latitude	Longitude	ATM	Accessories Store	Adult Boutique	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Amphithea
0	Staten Island	St. George	40.644982	-74.079353	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.025000	
1	Staten Island	New Brighton	40.640615	-74.087017	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.020000	
2	Staten Island	Stapleton	40.626928	-74.077902	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	
3	Staten Island	Rosebank	40.615305	-74.069805	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.000000	
4	Staten Island	West Brighton	40.631879	-74.107182	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.016667	

Methodology

- New York and Toronto dataset are very similar:
- Number of common venue categories in both data are :312
- Number of different venue categories in New York city are : 164
- Number of different venue categories in Toronto city are : 31
- As we had 312 common categories, we used only these categories to compare the venues of both cities.
- The measure the similarity of the locations, we used the Cosine similarity.

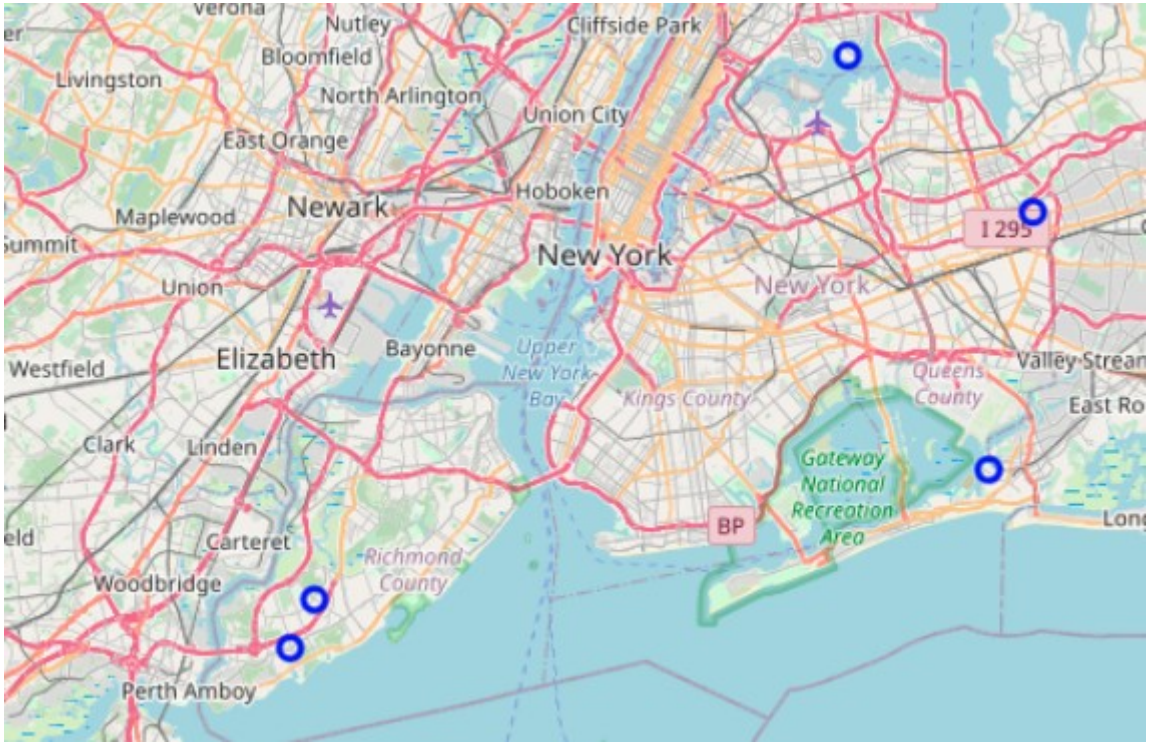


Results

- Moving from Wingate, Brooklyn, NY, to Toronto

Results

- Moving froMoving from Woburn, Scarborough, to New Yorkm Woburn, Scarborough, to New York



Conclusion

- Help someone to choose a new place to live
- Use location data
- New cities need new data
- Foursquare API is very useful
- Folium aid to visualize the results
- Cosine similarity to measure the similarity
- Neighborhood localization can be found in the internet