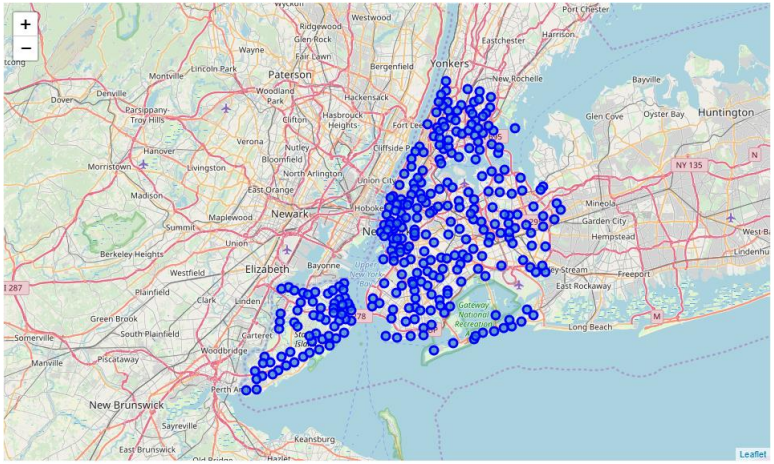
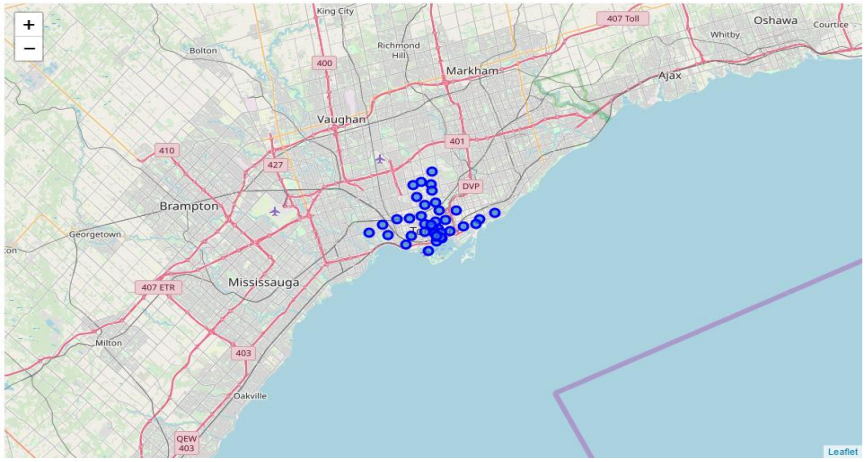
The Battle of Neighborhoods

Applied Data Science Capstone
Project by IBM

Introduction

- It is interesting to compare the neighborhoods of two cities.
- New York vs Toronto as an example
- The clustering technique is useful for this comparison by segmenting the neighborhoods into different groups. Objects within a cluster are similar and objects across different clusters are different.





Data

- **Data sources**
  - The Borough, Neighborhood, Latitude and Longitude information for New York and Toronto are required.
  - New York: https://cocl.us/new_york_dataset
  - Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
  - Venue information is obtained via Foursquare API for each neighborhood in New York and Toronto.
- **Data preparation**

	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

	Postcode	Borough	Neighbourhood	Latitude	Longitude
37	M4E	East Toronto	The Beaches	43.676357	-79.293031
41	M4K	East Toronto	The Danforth West, Riverdale	43.679557	-79.352188
42	M4L	East Toronto	The Beaches West, India Bazaar	43.668999	-79.315572
43	M4M	East Toronto	Studio District	43.659526	-79.340923
44	M4N	Central Toronto	Lawrence Park	43.728020	-79.388790

Methodology

- To simplify the analysis, Manhattan from New York and Downtown Toronto from Toronto are selected. Note that the method can be generalized to any number of cities.
- The final dataset contains 59 neighborhoods.
- K-means is one of the most popular clustering algorithms.
- K-means clustering divide the neighborhoods into different clusters by their venue information. Each cluster represents neighborhoods with similar venue distribution.
- Different K values (5 and 10) are used.

Results

- K = 5
  - Harbord, University of Toronto, China Town, Grange Park, and Kensington Market in Downtown Toronto are similar to Manhattan.

- K = 10
  - Queen's Park in Downtown Toronto is also found to be similar to Manhattan Ville in Manhattan.

- Please refer to the Jupyter notebook for the entire analysis.

Discussion and Conclusion

- Large K value should be used for K-mean clustering for this cities comparison.

- The neighborhoods in Manhattan in New York and Downtown Toronto are compared for similarity. Harbord, University of Toronto, China Town, Grange Park, Kensington Market and Queen's Park are found similar to Manhattan.

- The work provides a prototype and can be easily extended any number of cities.