

BakerZ Expansion Strategy

IBM Data Science Capstone Project

Shail Rajput, Nov 23, 2020

Introduction/Business Problem

BakerZ, Inc. Expansion Strategy for New York and Toronto

BakerZ, Inc. is a multi-national corporation that specializes in wholesale manufacturing and distribution of fine baked goods to coffee shops. Quality, freshness, and punctual delivery of the product have been their key success factors in sustaining their customers. They are expanding their business to New York and Toronto.

We were tasked us to determine the best neighborhood locations in Toronto and New York where they can set up distribution centers.

Their requirements for each distribution center are as follows:

- It should serve multiple neighborhoods with the maximum number of coffee shops that are closest to each other;
- It should only attempt to target neighborhoods with at least 10 or more coffee shops
- It should itself be located in the neighborhood with the maximum number of coffee shops

Deliverable: A priority list in establishing the distribution centers so they have the best return on investment.

Data Sources

Data sources for Toronto, ON

Neighborhood data shall be scraped from Wikipedia located at the following URL:

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

The second source is located in a CSV file called toronto_postal_code_geoloc.csv at the following URL: https://cocl.us/Geospatial_data

Data sources for New York, NY

JSON data is available in a file located at the following URL:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDriverSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json

Sourcing Coffee Shop data using Foursquare API

Foursquare Venues Search feature (<https://api.foursquare.com/v2/venues/search/<parameters>>)

- Using neighborhood geolocation to determine all the coffee shops in a neighborhood, count them and create NumCoffeeShops feature

Primary Features Used for New York and Toronto

Neighborhood Name, Latitude, Longitude, and NumCoffeeShops

This data shall be compiled into a data frames that map New York and Toronto neighborhoods to their geolocations.

Methodology

Data Preparation, Filtering, Machine Learning, Visualization

Prepared Clean Data Frames for New York and Toronto with following features:

Postal Code, Borough, Neighborhood, Latitude, Longitude, and NumCoffeeShops

Removed Neighborhoods for New York and Toronto with less than 10 Coffee Shops

Used Unsupervised Machine Learning (DBSCAN) to cluster neighborhoods in New York and Toronto

Determined Best DBSCAN Epsilon values for New York and Toronto using NearestNeighbor Approach

Combined Resulting Data into single data frame with the following features:

City, ClusterID, Neighborhood (with most coffee shops), NumCoffeeShops, and TotalCoffeeShopsInCluster

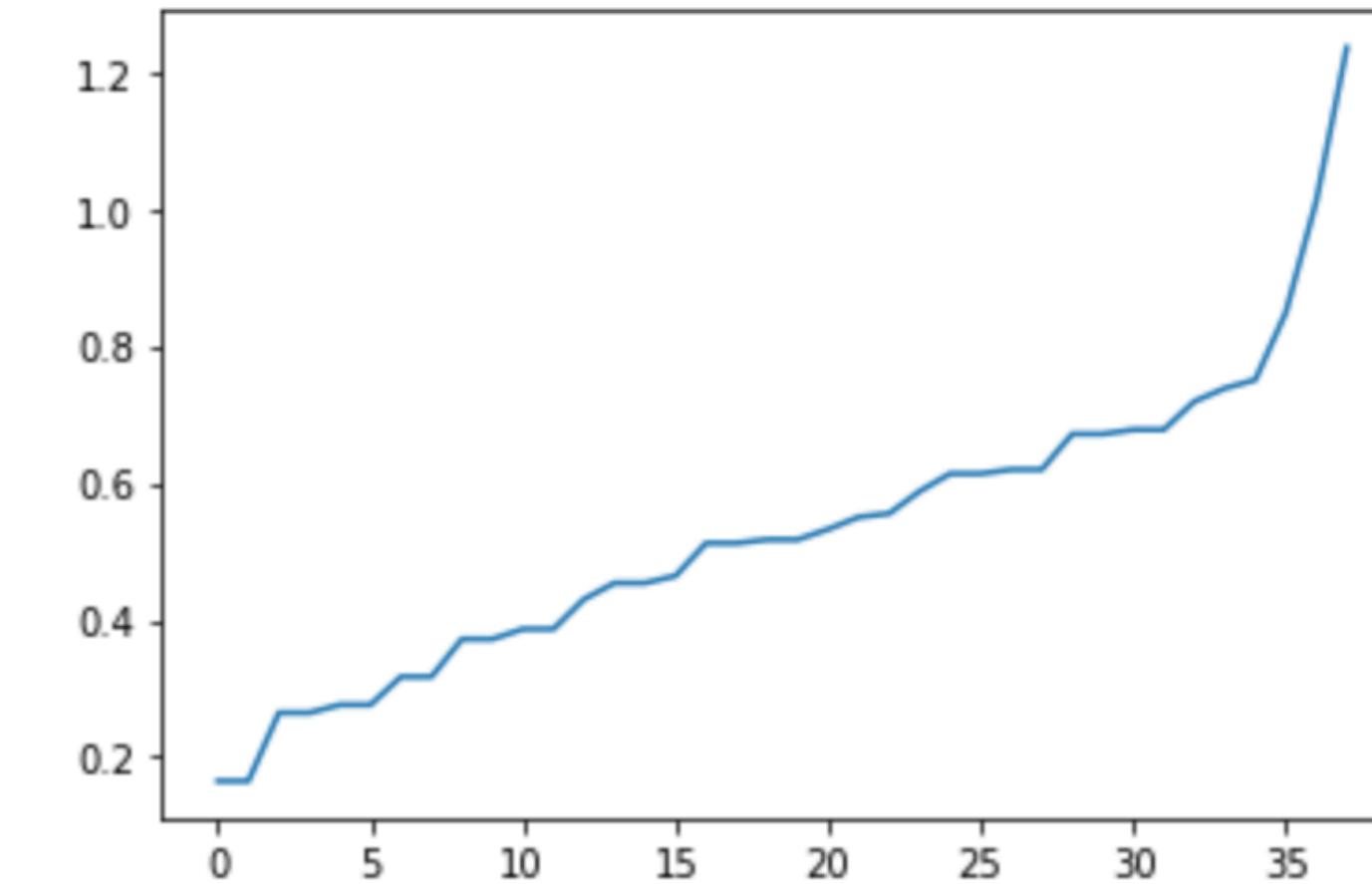
Created the Final Prioritized List for BakerZ by sorting on TotalCoffeeShopsInCluster feature

Map Visualizations were created for each stage

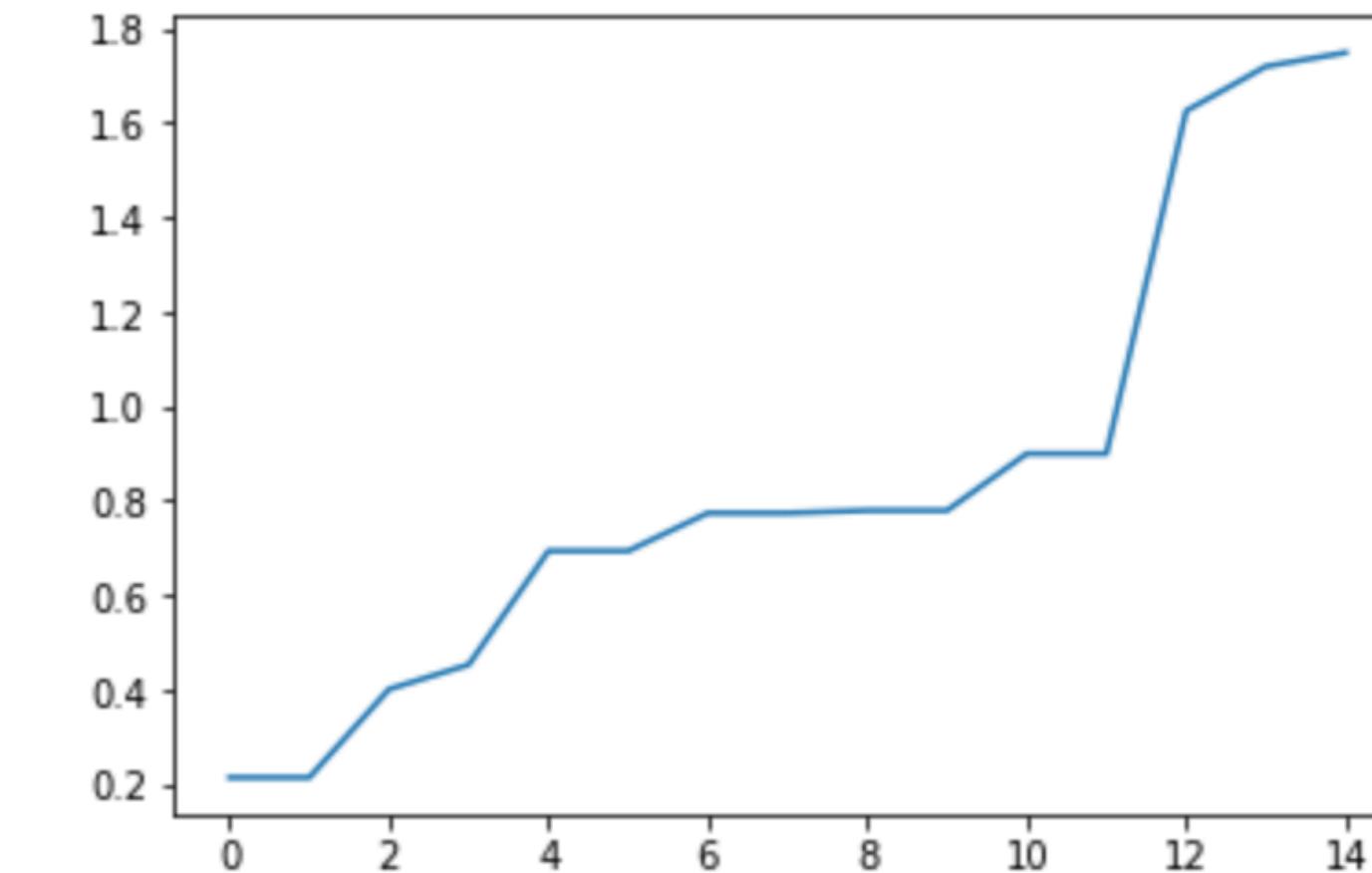
DBSCAN

Determining Epsilon parameter values for New York and Toronto

Using NearestNeighbor, the best Epsilon value for New York was determined to be 0.75



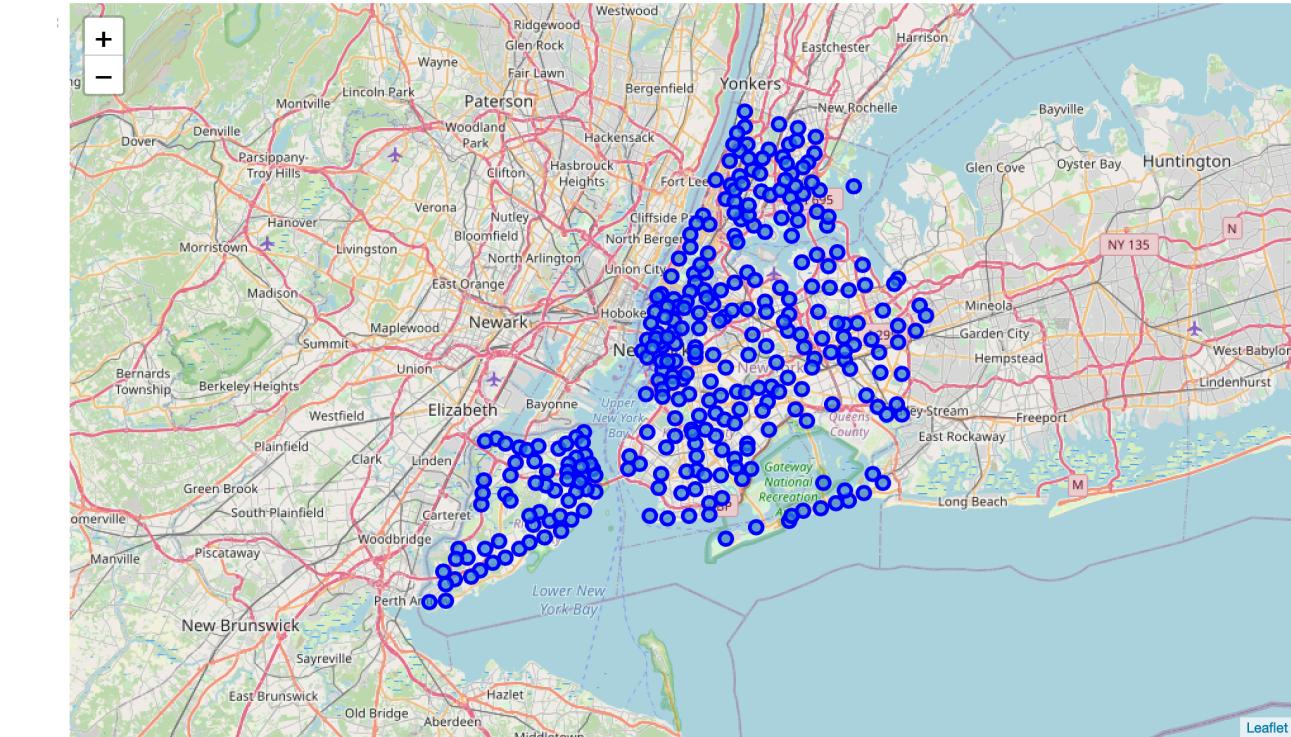
Using NearestNeighbor, the best Epsilon value for Toronto was determined to be 0.90



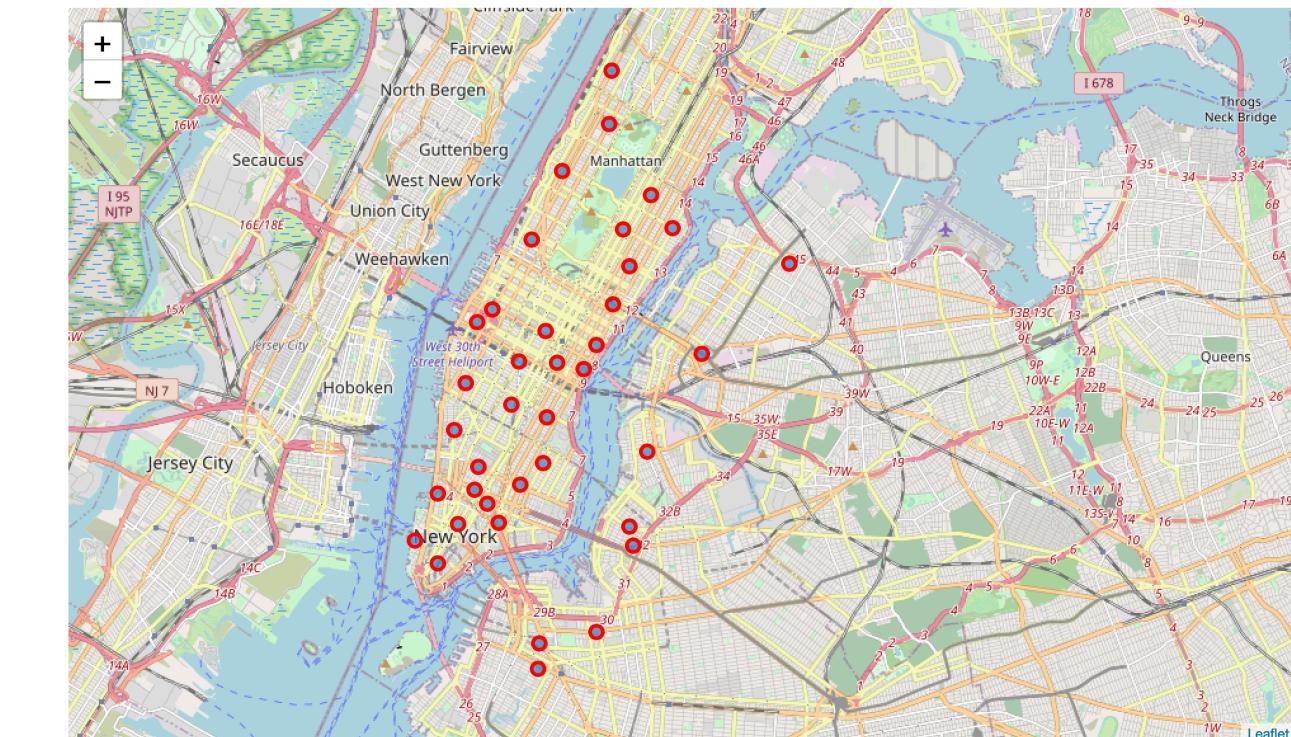
Visualizations

New York

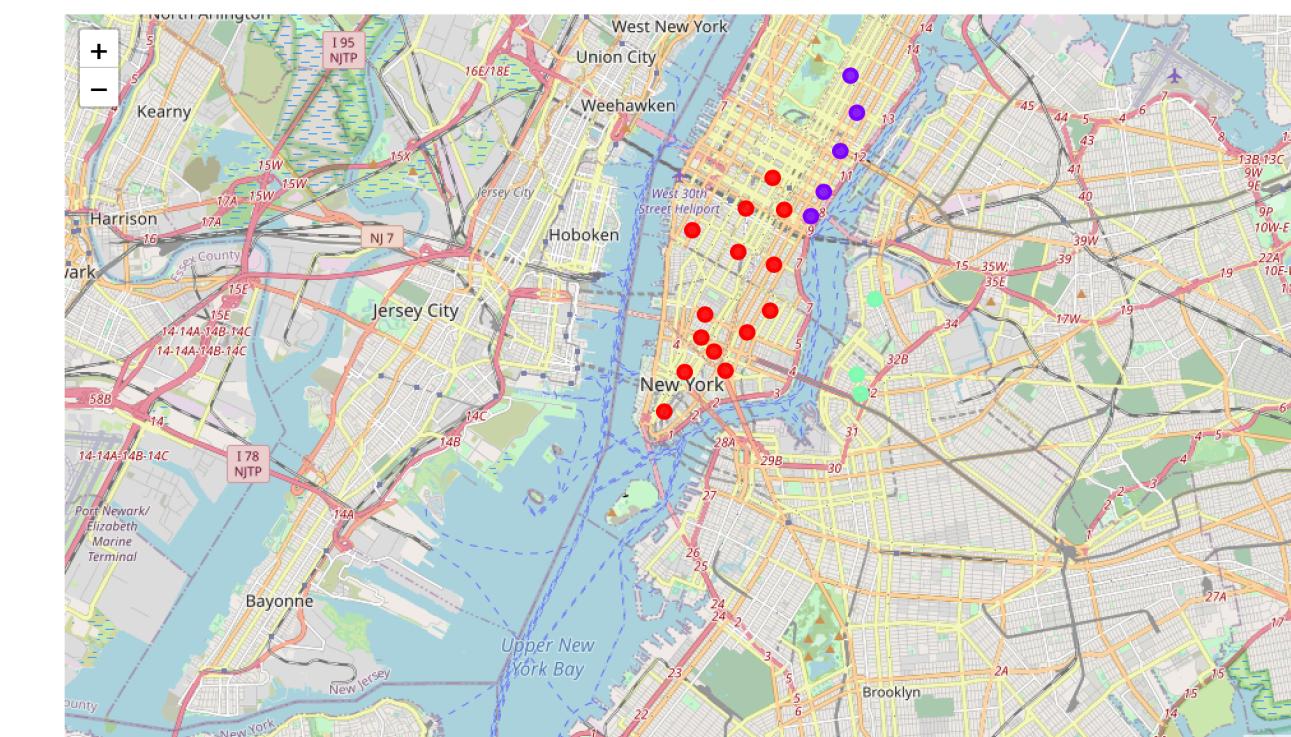
All Neighborhoods in New York



Neighborhoods in New York with 10 or more Coffee Shops



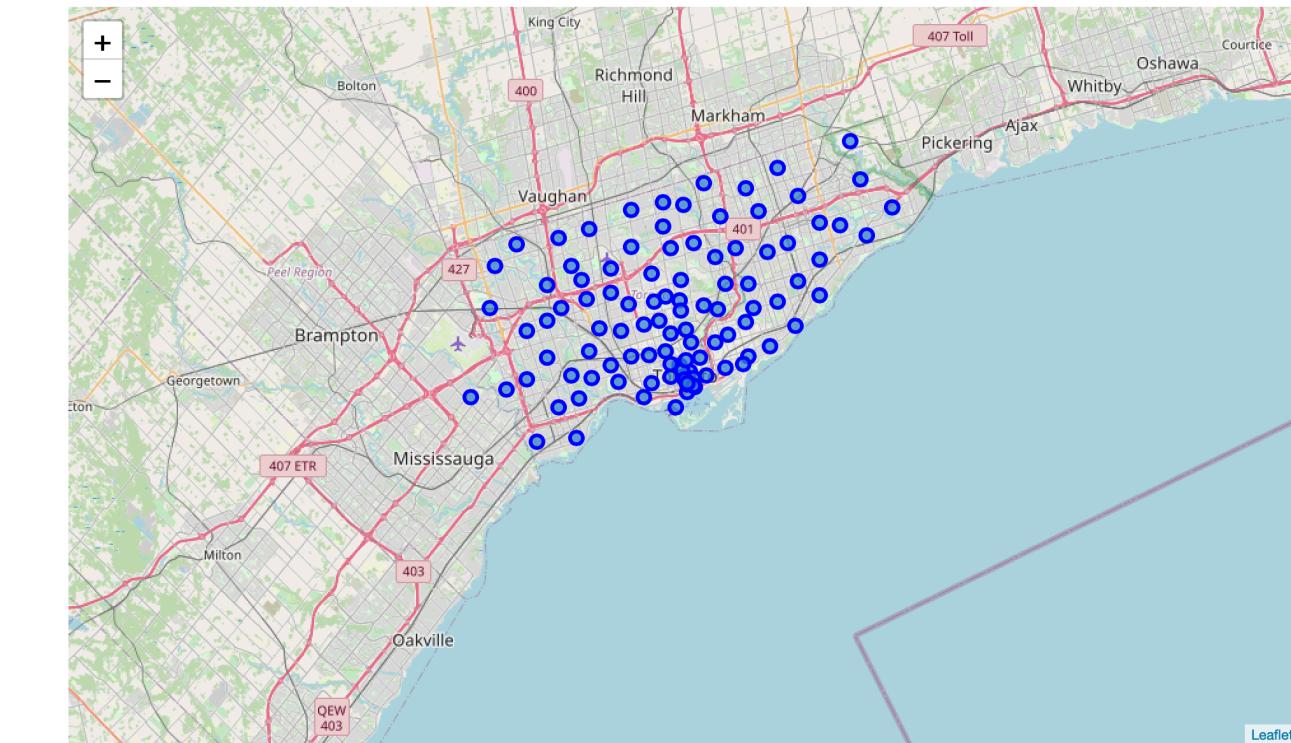
Clustered Neighborhoods in New York using DBSCAN



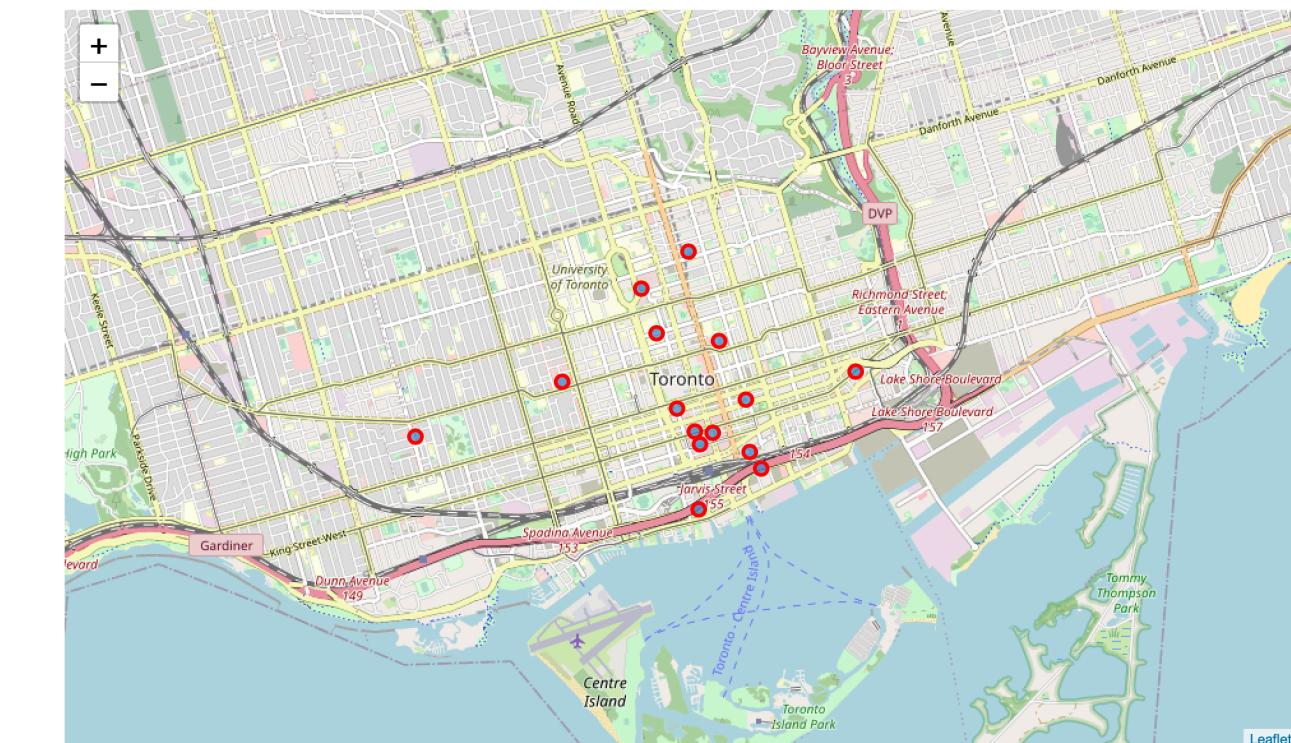
Visualizations

Toronto

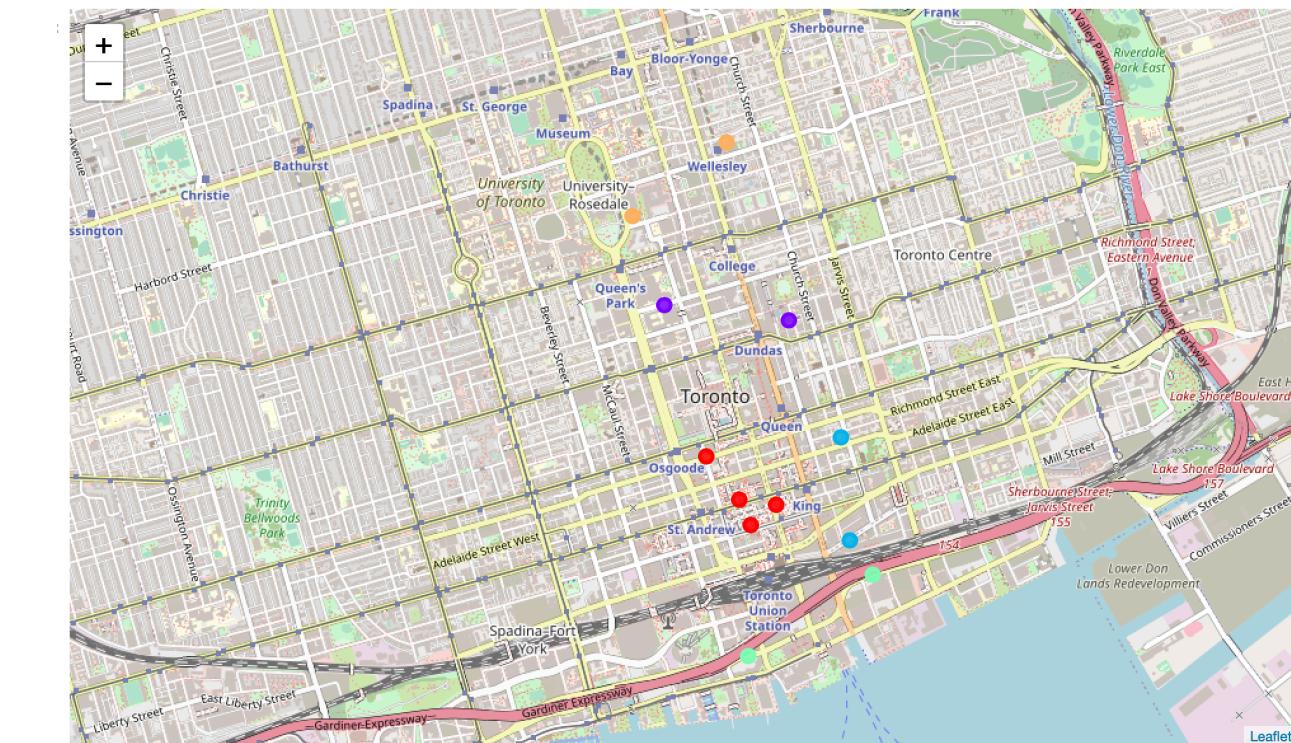
All Neighborhoods in Toronto



Neighborhoods in Toronto with 10 or more Coffee Shops



Clustered Neighborhoods in Toronto using DBSCAN



DBSCAN Results

New York

**22 Neighborhoods
Grouped into 3 Clusters
representing highest density
of Coffee Shops
closest to each other.**

Number of Neighborhood Clusters in New York is 3

ClusterID	Borough	Neighborhood	Latitude	Longitude	NumCoffeeShops
0	0	Manhattan	Murray Hill	40.748303	-73.978332
1	0	Manhattan	Greenwich Village	40.726933	-73.999914
2	0	Manhattan	Flatiron	40.739673	-73.990947
3	0	Manhattan	Midtown	40.754691	-73.981669
4	0	Manhattan	Soho	40.722184	-74.000657
5	0	Manhattan	Midtown South	40.748510	-73.988713
6	0	Manhattan	Financial District	40.707107	-74.010665
7	0	Manhattan	Chelsea	40.744035	-74.003116
8	0	Manhattan	Civic Center	40.715229	-74.005415
9	0	Manhattan	Little Italy	40.719324	-73.997305
10	0	Manhattan	Noho	40.723259	-73.988434
11	1	Manhattan	Turtle Bay	40.752042	-73.967708
12	1	Manhattan	Upper East Side	40.775639	-73.960508
13	1	Manhattan	Lenox Hill	40.768113	-73.958860
14	1	Manhattan	Tudor City	40.746917	-73.971219
15	0	Manhattan	Chinatown	40.715618	-73.994279
16	1	Manhattan	Sutton Place	40.760280	-73.963556
17	0	Manhattan	East Village	40.727847	-73.982226
18	0	Manhattan	Gramercy	40.737210	-73.981376
19	2	Brooklyn	North Side	40.714823	-73.958809
20	2	Brooklyn	South Side	40.710861	-73.958001
21	2	Brooklyn	Greenpoint	40.730201	-73.954241

DBSCAN Results

Toronto

**12 Neighborhoods
Grouped into 5 Clusters
representing highest density
of Coffee Shops
closest to each other.**

Number of Neighborhood Clusters in Toronto is 5

ClusterID	Postal Code	Borough	Neighborhood	Latitude	Longitude	NumCoffeeShops
0	0	M5H Downtown Toronto	Richmond, Adelaide, King	43.650571	-79.384568	43.0
1	0	M5X Downtown Toronto	First Canadian Place, Underground city	43.648429	-79.382280	40.0
2	0	M5L Downtown Toronto	Commerce Court, Victoria Hotel	43.648198	-79.379817	39.0
3	0	M5K Downtown Toronto	Toronto Dominion Centre, Design Exchange	43.647177	-79.381576	35.0
4	1	M5G Downtown Toronto	Central Bay Street	43.657952	-79.387383	32.0
5	1	M5B Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937	30.0
6	2	M5C Downtown Toronto	St. James Town	43.651494	-79.375418	26.0
7	2	M5W Downtown Toronto	Stn A PO Boxes	43.646435	-79.374846	26.0
8	3	M5E Downtown Toronto	Berczy Park	43.644771	-79.373306	14.0
9	4	M7A Downtown Toronto	Queen's Park, Ontario Provincial Government	43.662302	-79.389494	14.0
10	3	M5J Downtown Toronto	Harbourfront East, Union Station, Toronto Islands	43.640816	-79.381752	13.0
11	4	M4Y Downtown Toronto	Church and Wellesley	43.665860	-79.383160	11.0

Combined Results

New York and Toronto

15 New York and Toronto Neighborhoods Grouped into Clusters representing highest density of Coffee Shops closest to each other.

Key is identifying the Neighborhood with the most Coffee Shops in the Cluster along with the Total Coffee Shops in the Cluster

Note: Multiple Neighborhoods show up in same city cluster as they may have same high number of Coffee Shops (capped by Foursquare)

	City	ClusterID	Neighborhood	NumCoffeeShops	TotalCoffeeShopsInCluster
0	Toronto	0	Richmond, Adelaide, King	43.0	157.0
1	Toronto	1	Central Bay Street	32.0	62.0
2	Toronto	2	St. James Town	26.0	52.0
3	Toronto	2	Stn A PO Boxes	26.0	52.0
4	Toronto	3	Berczy Park	14.0	27.0
5	Toronto	4	Queen's Park, Ontario Provincial Government	14.0	25.0
6	New York	0	Murray Hill	50.0	593.0
7	New York	0	Greenwich Village	50.0	593.0
8	New York	0	Flatiron	50.0	593.0
9	New York	0	Midtown	50.0	593.0
10	New York	0	Soho	50.0	593.0
11	New York	0	Midtown South	50.0	593.0
12	New York	0	Financial District	50.0	593.0
13	New York	1	Turtle Bay	35.0	154.0
14	New York	2	North Side	20.0	44.0

Sorted Results

New York and Toronto

15 New York and Toronto Neighborhoods Grouped into Clusters representing highest density of Coffee Shops closest to each other.

Key is identifying the Neighborhood with the most Coffee Shops in the Cluster along with the Total Coffee Shops in the Cluster

Note: Multiple Neighborhoods show up in same city cluster as they may have same high number of Coffee Shops (capped by Foursquare)

	City	ClusterID	Neighborhood	NumCoffeeShops	TotalCoffeeShopsInCluster
0	New York	0	Murray Hill	50.0	593.0
1	New York	0	Greenwich Village	50.0	593.0
2	New York	0	Flatiron	50.0	593.0
3	New York	0	Midtown	50.0	593.0
4	New York	0	Soho	50.0	593.0
5	New York	0	Midtown South	50.0	593.0
6	New York	0	Financial District	50.0	593.0
7	Toronto	0	Richmond, Adelaide, King	43.0	157.0
8	New York	1	Turtle Bay	35.0	154.0
9	Toronto	1	Central Bay Street	32.0	62.0
10	Toronto	2	St. James Town	26.0	52.0
11	Toronto	2	Stn A PO Boxes	26.0	52.0
12	New York	2	North Side	20.0	44.0
13	Toronto	3	Berczy Park	14.0	27.0
14	Toronto	4	Queen's Park, Ontario Provincial Government	14.0	25.0

Final Prioritized Neighborhood Selection

Remove Redundant Neighborhoods

8 New York and Toronto
Neighborhoods
each in a separate Cluster
representing highest density
of Coffee Shops
closest to each other.

	City	ClusterID	Neighborhood	NumCoffeeShops	TotalCoffeeShopsInCluster
0	New York	0	Murray Hill	50.0	593.0
7	Toronto	0	Richmond, Adelaide, King	43.0	157.0
8	New York	1	Turtle Bay	35.0	154.0
9	Toronto	1	Central Bay Street	32.0	62.0
10	Toronto	2	St. James Town	26.0	52.0
12	New York	2	North Side	20.0	44.0
13	Toronto	3	Berczy Park	14.0	27.0
14	Toronto	4	Queen's Park, Ontario Provincial Government	14.0	25.0

Final Strategic Expansion Plan for BakerZ

Maximizes Return On Investment

BakerZ should establish 8 Distribution Centers in the following order:

- 1. Murray Hill (New York)**
- 2. Richmond, Adelaide, King (Toronto)**
- 3. Turtle Bay (New York)**
- 4. Central Bay Street (Toronto)**
- 5. St. James Town (Toronto)**
- 6. North Side (New York)**
- 7. Berczy Park (Toronto)**
- 8. Queen's Park, Ontario Provincial Government (Toronto)**