Toronto Neighborhoods Clustering

Introduction

Toronto is the capital city of the Canadian province of Ontario and the most populous city in Canada. Downtown Toronto is the main central business district of Toronto.

When someone want to move from one neighborhood to another neighborhood in Downtown Toronto, the person would like to move to another neighborhood with similar amenities.

In this project we will use k-means clustering algorithm to find the neighborhoods with similar amenities. The area around the Downtown Toronto will be clustered using their venue data.

Data

Postal code, neighborhood and borough information are available from the following link:

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

FourSquare API will also be used to get venues data in Downtown Toronto based on the latitude and longitude of the neighborhoods.

Methodology

Using the url and use BeautifulSoup package for webscrapping, information regarding postal code, neighborhood and borough can be gathered.

	PostalCode	Borough	Neighborhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Regent Park, Harbourfront
3	M6A	North York	Lawrence Manor, Lawrence Heights
4	M7A	Queen's Park	Ontario Provincial Government

Methodology

Combining the Postal Code data and Latitude and Longitude information for Downtown Toronto, the following data can be gathered.

	PostalCode	Borough	Neighborhood	Latitude	Longitude
0	M5A	Downtown Toronto	Regent Park, Harbourfront	43.654260	-79.360636
1	M5B	Downtown Toronto	Garden District, Ryerson	43.657162	-79.378937
2	M5C	Downtown Toronto	St. James Town	43.651494	-79.375418
3	M5E	Downtown Toronto	Berczy Park	43.644771	-79.373306
4	M5G	Downtown Toronto	Central Bay Street	43.657952	-79.387383

Downtown Toronto Map

Methodology

To get the venues information, an API call to the FourSquare API is performed. The information received in JSON format. The venues for all the neighborhoods in Downtown Toronto are stored in new dataframe.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Regent Park, Harbourfront	43.65426	-79.360636	Roselle Desserts	43.653447	-79.362017	Bakery
1	Regent Park, Harbourfront	43.65426	-79.360636	Tandem Coffee	43.653559	-79.361809	Coffee Shop
2	Regent Park, Harbourfront	43.65426	-79.360636	Cooper Koo Family YMCA	43.653249	-79.358008	Distribution Center
3	Regent Park, Harbourfront	43.65426	-79.360636	Body Blitz Spa East	43.654735	-79.359874	Spa
4	Regent Park, Harbourfront	43.65426	-79.360636	Impact Kitchen	43.656369	-79.356980	Restaurant

Methodology

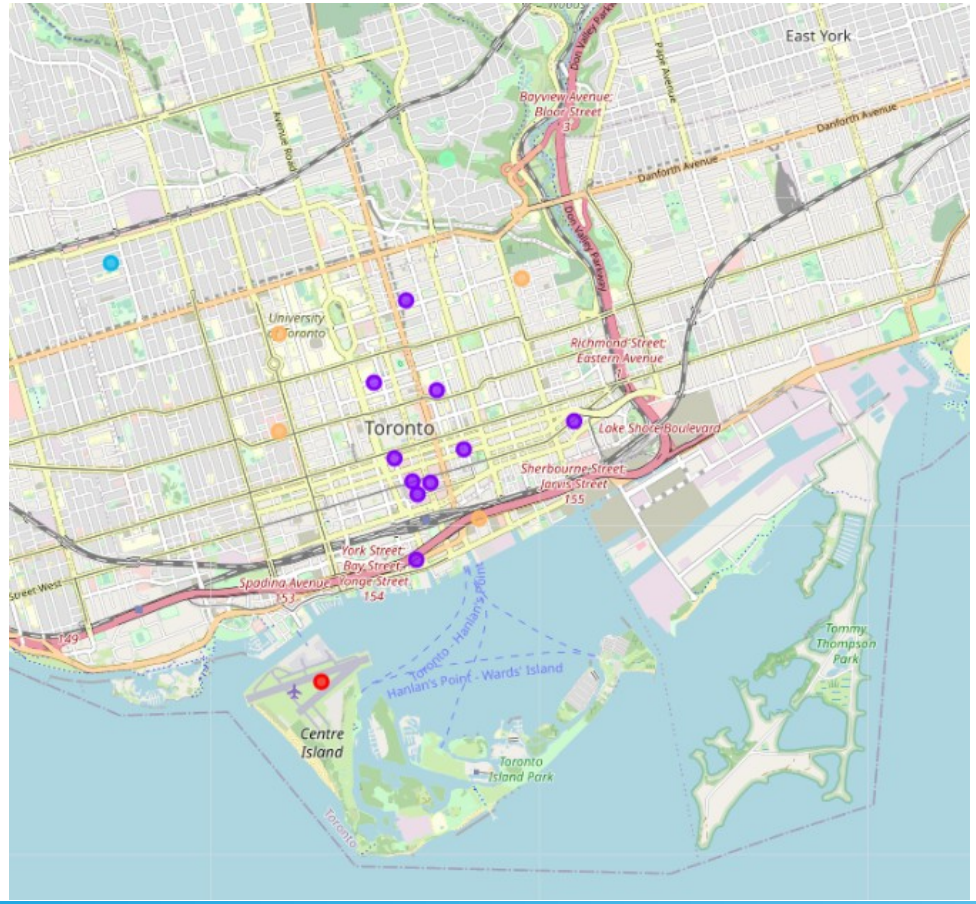
The next step is to analyze each neighborhood and grouping the venues by neighborhood and take the mean of the frequency of occurrence of each venue category. A new dataframe is created and top 10 venues for each neighborhood are stored.

	Neighborhoods	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Berczy Park	Cocktail Bar	Bakery	Coffee Shop	Restaurant	Seafood Restaurant
1	CN Tower, King and Spadina, Railway Lands, Har...	Airport Service	Airport Lounge	Airport Terminal	Coffee Shop	Harbor / Marina
2	Central Bay Street	Coffee Shop	Italian Restaurant	Sandwich Place	Café	Salad Place
3	Christie	Grocery Store	Café	Park	Athletics & Sports	Restaurant
4	Church and Wellesley	Coffee Shop	Japanese Restaurant	Sushi Restaurant	Restaurant	Gay Bar

Methodology

After that, k-means clustering algorithm is run to cluster the neighborhood into 5 clusters. A new dataframe is created to store the cluster as well as the top 10 venues for each neighborhood.

Clustering


East York
Bayview Avenue
Bloor Street
Danforth Avenue
Don Valley Parkway
University of Toronto
Richmond Street
Eastern Avenue
Toronto
Lake Shore Boulevard
Sherbourne Street
Jarvis Street
155
York Street
Bay Street
Yonge Street
154
Spadina Avenue
153
Toronto - Hanlan's Point - Wards' Island
Centre Island
Toronto Island Park
Tommy Thompson Park

Results and Discussion

From the figure, it can be seen that most neighborhood in Downtown Toronto are grouped in 1 cluster.

By examining the biggest cluster, with top 1 common venue as Coffee Shop, this cluster can be named as Coffee Shop cluster. If a person wants to move between the neighborhood, the person can select the following neighborhood with similar venu

Neighborhoods
Central Bay Street
Church and Wellesley
Commerce Court, Victoria Hotel
First Canadian Place, Underground city
Garden District, Ryerson
Harbourfront East, Union Station, Toronto Islands
Regent Park, Harbourfront
Richmond, Adelaide, King
St. James Town
Toronto Dominion Centre, Design Exchange

Conclusion

The neighborhoods in Downtown Toronto are now clustered and a map containing the results is available. Most neighborhoods in Downtown Toronto are not unique and located in the same cluster. To get more insights, more clusters can be added and more features can be included in the observation such as demographics and median age of each neighborhood.