Classifying Urban Sectors of the Greater Toronto Area of Canada

By Het Shah, 11th July 2020

Introduction:

The Greater Toronto Area is one of the most vibrant, multicultural and ever-developing regions of North America. The population of GTA is growing every day and it gets diverse with each passing year. This is due to the continuing migration from other parts of Canada and the world.

As the city grows, it is empirical for the City Planners, Developers, Investors, and Entrepreneurs to understand and identify the opportunities that may arise with the growing urban population. It is also important for local authorities and officials to quantitatively examine the underdeveloped neighborhoods and take action to establish equity within the Greater Toronto Area.

We will use Foursquare API and population data for the following analysis:

1) Classifying neighborhoods as highly developed, downtown and less/under developed

2) Understanding Toronto's make up by learning the demography.

3) Exploring underdeveloped neighborhoods.

4) Identifying and examining business opportunities in various neighborhoods.

Data:

To fulfill our research goal we use reliable and publicly available datasets. The following sources of data have been used

1) Population and Demographic data from : https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/335e3407-d3ca-422a-bfac-e3807cfbe0d5

This dataset has population and demographic categories for 140 neighborhoods that make up the city of Toronto. It also has data on Age and Sex breakdown of the population. It also identifies families by their first language which helps in understanding the demographics of the city.

2) Latitude and Longitude using the GeoJSON and csv file of the borders of the neighborhood made available by the city of Toronto.

    GeoJSON file: https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=geojson&projection=4326

    CSV file: https://ckan0.cf.opendata.inter.prod-toronto.ca/download_resource/a083c865-6d60-4d1d-b6c6-b0c8a85f9c15?format=csv&projection=4326

3) Foursquare Developers Access

Methodology:

The methodology used explores themes learned throughout the course. It uses publicly available data and works on that data in the Jupyter labs. The methodology includes:

1) K- means clustering and segmenting neighborhoods.

2) Retrieving data and wrangling it using python.

3) Visualizing clusters and population data using geospatial analysis.

Data Retrieving, Curating and Wrangling:

We had access to the Population data set made available by StatsCan and the City of Toronto open access data. This data summarizes the distribution of population among different neighborhoods of Toronto. We used the following columns of the data that was available.

1) **Area-** This column contains the names of the neighbourhoods of the Toronto Metropolitan Area. There are some neighbourhoods that stretch beyond the Toronto CMA but that only helps the study to understand how the city is progressing.

2) **Total Area-** This column describes the size of the neighbourhood in square kilometers. We later find out that the size of the neighbourhood increases as we move outwards of the city.

3) **Total Population-** This column describes the total population in the neighborhood. The density of population decreases as we move outwards of the city.

Data on the population and the linguistic characteristics of the city was transformed into a Pandas dataframe. We removed the unnecessary columns using the drop condition.

We then used the Geocoder function from Geopy Library of python to retrieve the Latitudes and Longitudes of each neighbourhood. Following is a snapshot of the data.

[29]:		Area	Latitude	Longitude
	0	Clairville, ON, Canada	43.773492	-79.651051
	1	Jamestown, ON, Canada	43.810844	-81.197743
	2	Kipling, ON, Canada	43.637593	-79.535494
	3	Old Rexdale, ON, Canada	43.721362	-79.565513
	4	The Westway, ON, Canada	43.688707	-79.542703

We then used Foursquare API Developer's Access to explore different type of venues and categories. The API found these 10 categories:

Arts and Entertainment	Outdoors and Recreation
College and University	Professional and Other Places
Events	Residence
Food	Shop and Service
Nightlife Spot	Travel and Transport

The following is how the resulting dataframe looks like:

|	Area	Latitude	Longitude	Arts & Entertainment	College & University	Event	Food	Nightlife Spot	Outdoors & Recreation	Professional & Other Places	Residence	Shop & Service	Travel & Transport
0	Clairville, ON, Canada	43.773492	-79.651051	1	0	0	5	2	2	5	0	6	3
1	Jamestown, ON, Canada	43.810844	-81.197743	0	0	0	0	0	0	0	0	0	0
2	Kipling, ON, Canada	43.637593	-79.535494	8	3	0	38	8	4	22	6	28	6
3	Old Rexdale, ON, Canada	43.721362	-79.565513	4	1	1	5	0	3	3	0	12	1
4	The Westway, ON, Canada	43.688707	-79.542703	1	2	0	4	2	4	6	1	8	5

We then create a boxplot of the frequencies of the type of the venue
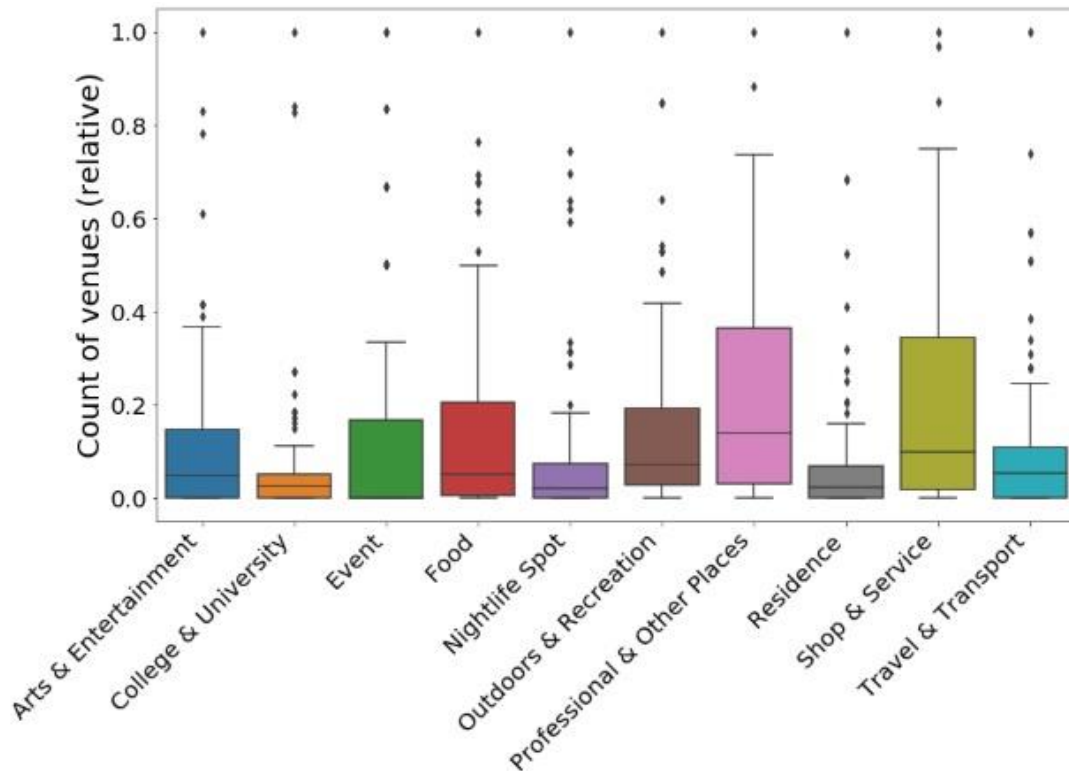
## Using K-Means clustering and segmenting neighborhoods.

We use 4 clusters in this study as it is extremely important to understand the levels of division among the neighborhoods. This will help us in understanding how different amenities are concentrated in particular neighbourhoods. This would be able to help us understand the Uptown and Downtown divide as well as the clusters within the Uptown and Downtown neighbourhoods.

The following is a visual representation of the 4 clusters.
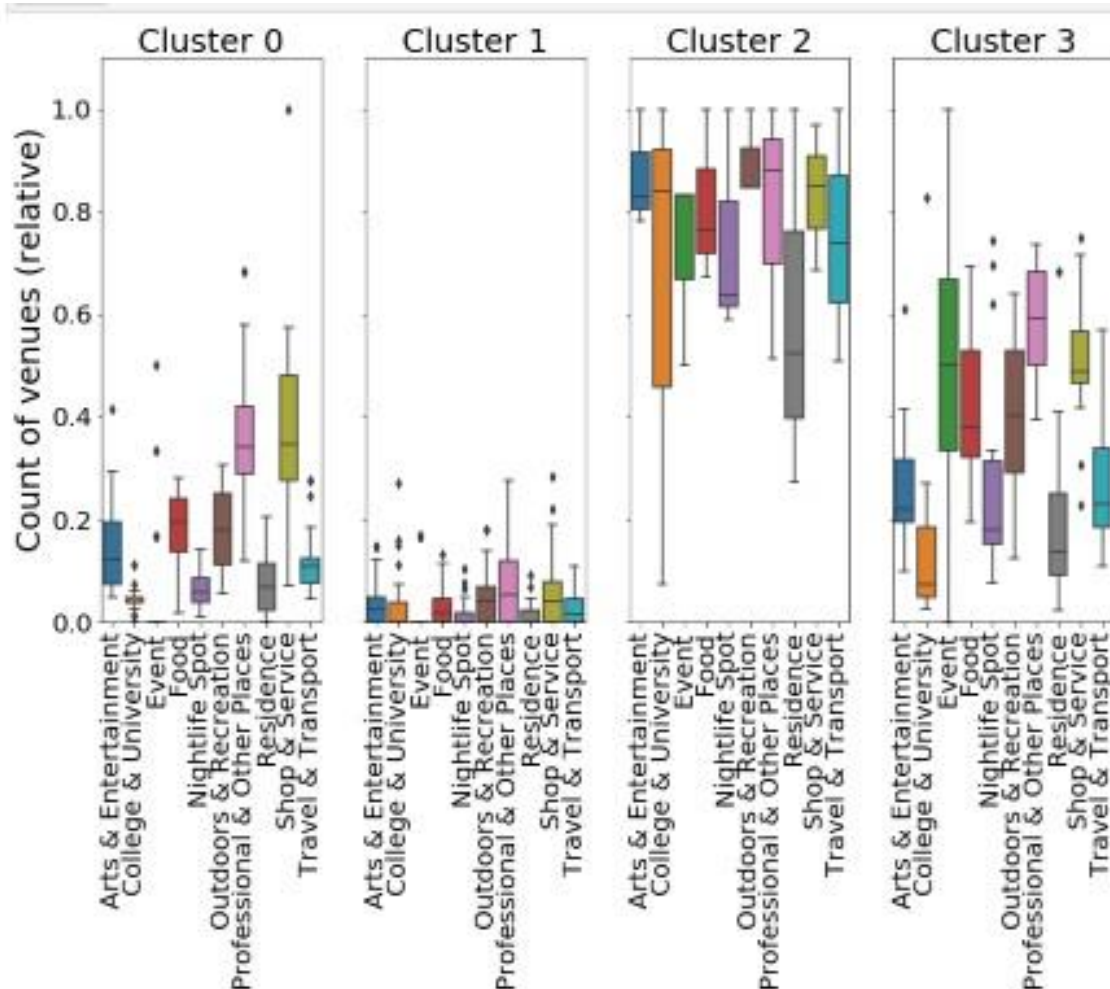
The interpretation of these clusters can be made as follows:

1) Cluster 0 (blue) was found to have medium to low frequencies for all categories. These neighbourhoods were found to be between the downtown areas and the suburbs.

2) Cluster 1 (green) was found to have the lowest frequencies for all categories among all the clusters. These neighbourhoods are far outside the downtown areas and are sparsely populated compared to the downtown areas.

3) Cluster 2 (Red) was found to have the highest frequencies for all categories and have the highest density of population. These neighbourhoods are kind of the *creame de la creame* of the downtown Toronto with areas such as Yonge-Dundas.

4) Cluster 3 (Pink) was found to have medium to high frequencies of most categories. Most of these neighborhoods were found to be in the downtown to midtown of Toronto.
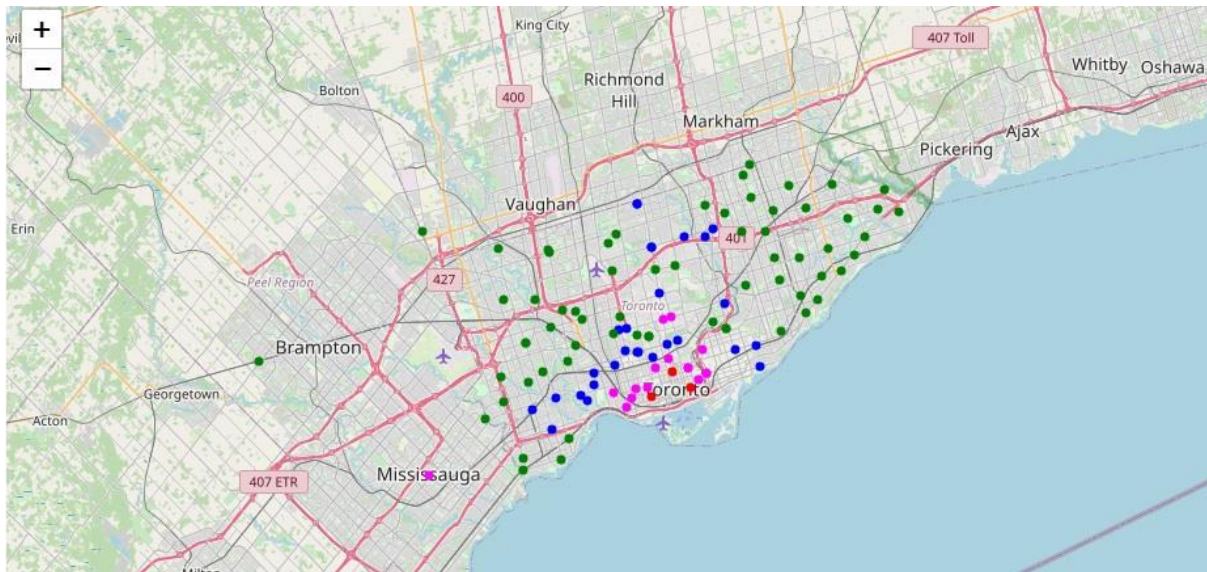
Results

After our analysis we were able to synthesize the following findings.

1) Only a few neighbourhoods of Toronto Metropolitan area can be identified as highly developed. All of these areas are located at the heart of downtown Toronto and mostly commercial and non-residential areas. There is a huge disparity between the most developed cluster and the next most developed neighbourhood.

2) The next most developed neighbourhoods were all also located in the downtown and some in midtown Toronto. These areas are densly populated and more residential than commercial.

3) The next most developed areas were just outside downtown Toronto and areas that are mostly residential. Most of the amenities and categories are present but are sparsely located.

4) The least developed areas were found to be the suburbs of Toronto. These areas are located between the more developed suburbs such as Mississauga and Oakville and the downtown Toronto areas.

Discussion and Recommendations

My aim before starting this study was to identify areas of Toronto that have the most number of amenities and community centers.

Due to the limited scope of this study we were not able to explore deeper demographics of the city and relate the community centers with it. It would be interesting to see how income structures and ethnic composition of an area shapes up the community centers set up in various neighbourhoods. The city needs to pay more attention to the regions left out of these amenities and develop these areas more.