

***Categorization of Neighborhoods as per availability of
healthcare facilities within 2-kilometer radius
– A data analysis approach***

Manash Sarma

Table of Contents

Table of Contents	2
Introduction:	3
Business problem:	3
Place selected for grouping neighbourhood:	4
Data description:	4
Methodology:	5
Exploration:	5
Pre-processing of data:	6
Clusters determination:	8
Determining number of clusters Technically:	8
Results:	9
Discussion and Inferences:	13
Recommendation:	13
Conclusion	13

Introduction:

Health care is one of the basic needs of all living beings. To run life smoothly, we need healthy life and for that health care facilities are essential.

It is often observed:

- Some places do not have a hospital / clinic.
- Some places may have a hospital but there may not be medical center with better facilities.
- Again, some places do not have Urgent care center.
(This bulleted list can increase)

Information about non availability of a health care facilities is raised in public sites/newspapers often mainly from a humanitarian aspect. But information is available in bits and pieces for some specific places without a reasonable accuracy. Sometimes, it may even be biased. Ultimately, it does not help address the problem either way.

This happens because often survey is manual and data collection method and their processing is error prone.

Business problem:

We need a system that can scientifically take a survey of different neighborhoods inside a region / state for finding different types of medical facilities available within 2 kilometer radius of the neighborhoods.

So, our business goal is to:

- Do a statistical survey / data analysis of existing medical facilities in different neighborhoods inside a region / state within a radius of 2000 meters (2 Kilometers).
 - Data should be very relevant – like number medical center, hospital, lab etc.
- Make a grouping or categorization of the neighborhoods based on health care facilities found in the analysis that should reflect business expansion scope (or raise as a concern if medical facility is insufficient in a locality)

- Visually mark the neighborhoods as per group/category they belong inside a map.

Place selected for grouping neighbourhood:

I was initially exploring some places in developing / under-developed countries. However, because of lack of expected data, I am sticking to Toronto. Existing healthcare facilities are explored for each neighborhood in Toronto.

Data description:

Since, the place is Toronto itself, I consider the almost the same approach for collecting the data set as it will serve the purpose of solving the business problem:

1. Build code to scrape the following Wikipedia page, https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M, and takeout relevant information of the table of postal codes and then to transform the data into a pandas dataframe.
2. The dataframe will consist of three columns: PostalCode, Borough, and Neighborhood as shown:

	PostalCode	Borough	Neighborhood
0	M5G	Downtown Toronto	Central Bay Street
1	M2H	North York	Hillcrest Village
2	M4B	East York	Parkview Hill, Woodbine Gardens
3	M1J	Scarborough	Scarborough Village
4	M4G	East York	Leaside
5	M4M	East Toronto	Studio District
6	M1R	Scarborough	Wexford, Maryvale
7	M9V	Etobicoke	South Steeles, Silverstone, Humbergate, Jamest...
8	M9L	North York	Humber Summit
9	M5V	Downtown Toronto	CN Tower, King and Spadina, Railway Lands, Har...
10	M1B	Scarborough	Malvern, Rouge
11	M5A	Downtown Toronto	Regent Park, Harbourfront

3. Geographical coordinates of each postal code can be found from http://cocl.us/Geospatial_data. From this dataset, a data frame of geospatial data can be constructed that includes longitude and latitude information:

So, finally a merged dataframe would be constructed from the two datasets that will have two additional columns latitude and longitude . This latitude and longitude column information would be essential to for using Foursquare API to achieve the next steps.

Methodology:

Exploration:

For any data analysis and inference work, exploration is the first step after selection of data as mentioned in data section:

- Type of data is verified
- Different tables, fields are verified.
- Explored neighbourhood data with Foursquare API for determining medical venues within a radius of 2 kilometres.
- Since data is mainly geospatial, data is to draw a map and verified.

	Postalcode	Borough	Neighborhood	Latitude	Longitude
0	M1B	Scarborough	Rouge,Malvern	43.806686	-79.194353
1	M1C	Scarborough	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497
2	M1E	Scarborough	Guildwood,Morningside,West Hill	43.763573	-79.188711
3	M1G	Scarborough	Woburn	43.770992	-79.216917
4	M1H	Scarborough	Cedarbrae	43.773136	-79.239476

Figure 1 - Snapshot of the table with geospatial data

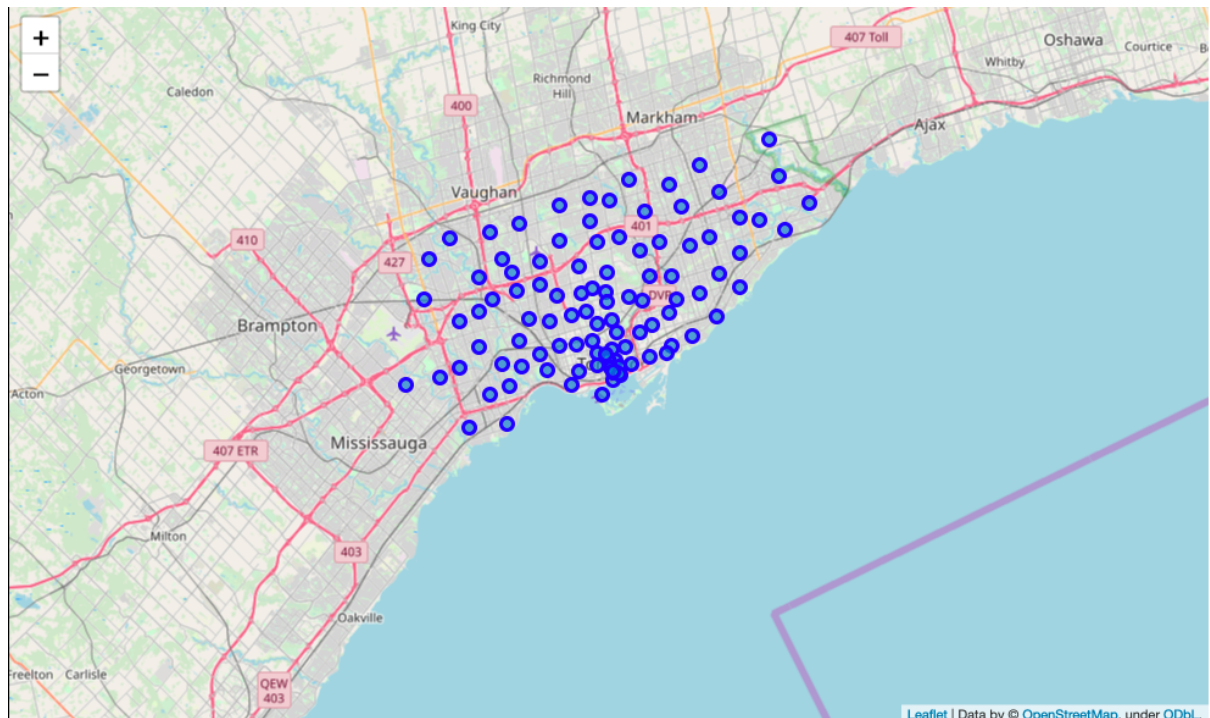


Figure 2 - Toronto Neighbourhoods marked with blue

Pre-processing of data:

We need to convert the data shown in figure 1 to another tabular form to fulfil actual goal of survey / analysis of medical facilities.

I have considered five major facilities in the table –

- Medical center,
- Hospital,
- Medical emergency,
- Urgent health care,
- Medical lab

For this to achieve, 'Four Square API' and venue categories support in <https://developer.foursquare.com/docs/resources/categories> is followed.

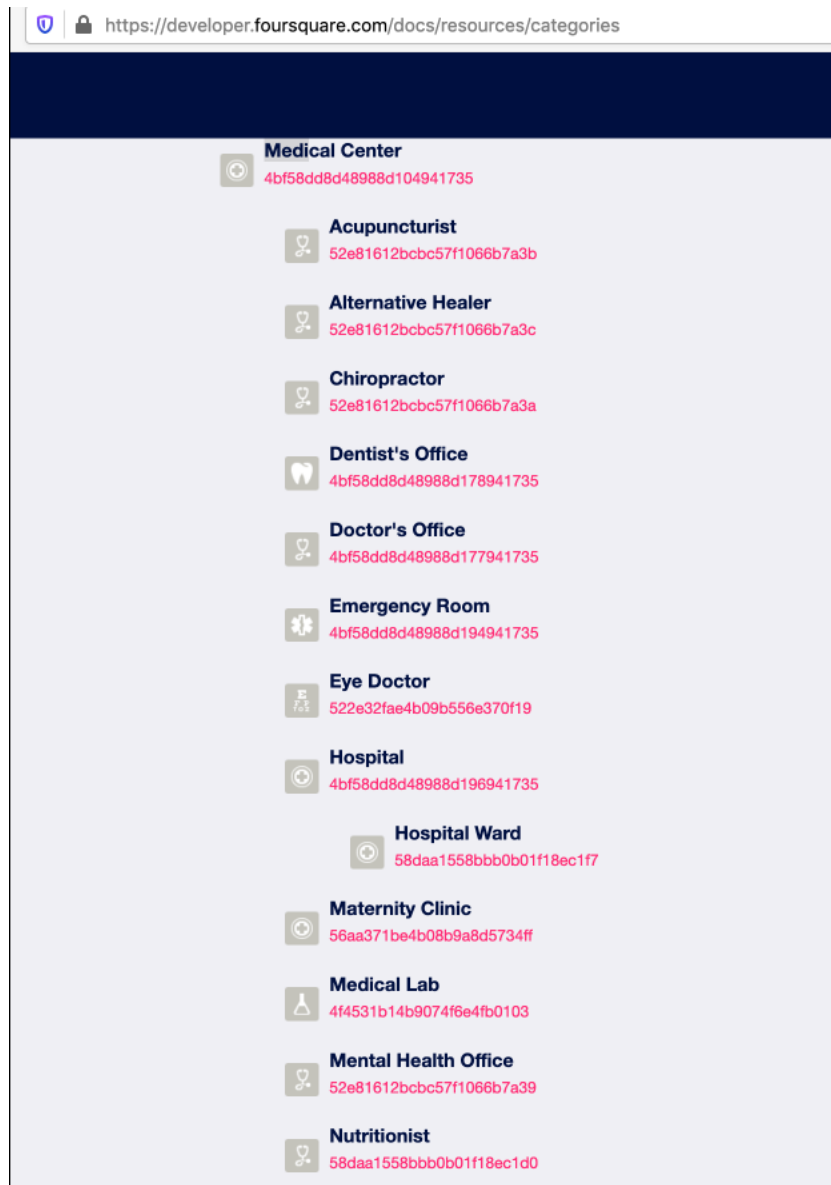


Figure 3 - Categories for medical establishments

As per the study and other websites, it is assumed that:

- A medical center has all the facilities like emergency room, medical lab, urgent care. So, if a neighborhood has a medical center, it will also have these facilities.
- It is assumed that a hospital will have lab facilities.
- While doing cluster analysis, a medical center is assigned three times weightage mentioned in 'cluster determination' section below.

With this assumption and information from table in figure1, following table/data-frame is constructed:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Medical Center	Urgent care	Medical Lab	Medical Emergency	Hospital
0	Rouge,Malvern	43.806686	-79.194353	1	1	1	1	0
1	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0	0	0	0	0
2	Guildwood,Morningside,West Hill	43.763573	-79.188711	3	3	4	3	1
3	Woburn	43.770992	-79.216917	6	6	8	7	2
4	Cedarbrae	43.773136	-79.239476	6	6	6	7	0
5	Scarborough Village	43.744734	-79.239476	4	4	6	5	2
6	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029	0	0	0	0	0
7	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577	1	1	1	1	0
8	Cliffcrest,Cliffside,Scarborough Village West	43.716316	-79.239476	0	0	0	0	0

Figure 4 - Data frame after preprocessing

Clusters determination:

Scoring method: I have not directly used attributes of data frame at Figure 4 for grouping. For each neighbourhood , a score evaluated based on number of medical facilities in the locality:

Score of a neighbourhood = 3 * No of Medical centers + No of Hospitals + No of urgent cares + No of Medical labs + No of emergency facilities.

Please note that a medical center's weightage is considered as 3 times more.

Determining number of clusters Technically:

Elbow method is followed that suggests 3 clusters from the dataset mentioned figure 4.

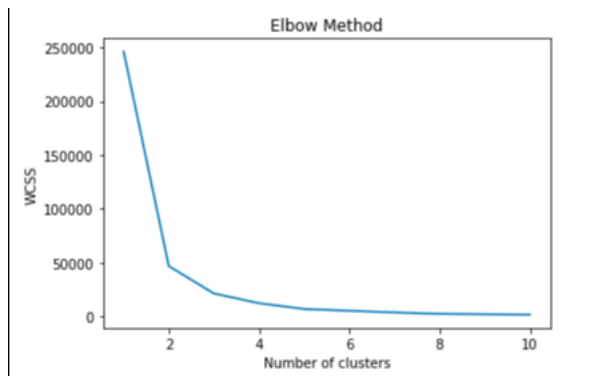


Figure 5 - Elbow method to find number of clusters

Now with score data for each neighborhood and Kmeans python library method with Kmean algorithm approach, three groups constructed. Results are mentioned result section.

Results:

As a result, there are three tables and three maps categorized with health care organization availability:

- Figure 6 is a table with comprehensive information of different types of health organization and result category of the neighborhood at the last column.
- Figure 7 is a bar graph showing the count of each type health organizations in the neighborhoods in Toronto.
- Figure 8, 10, 12 are bar chart that reflect the count of different types of medical organizations in each neighborhood for category1, category2 and category3 respectively.
- Figure9, 11, 13 are the marked neighborhood inside map for category1, 2 and 3 respectively.
- Number of neighborhoods placed in category1, 2 and 3 are found to be 71, 17 and 15 respectively ; related visualization Figure 14.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Medical Center	Urgent care	Medical Lab	Medical Emergency	Hospital	Medical support category
0	Rouge,Malvern	43.806686	-79.194353	1	1	1	1	0	0
1	Highland Creek,Rouge Hill,Port Union	43.784535	-79.160497	0	0	0	0	0	0
2	Guildwood,Morningside,West Hill	43.763573	-79.188711	3	3	4	3	1	0
3	Woburn	43.770992	-79.216917	6	6	8	7	2	0
4	Cedarbrae	43.773136	-79.239476	6	6	6	7	0	0
5	Scarborough Village	43.744734	-79.239476	4	4	6	5	2	0
6	East Birchmount Park,Ionview,Kennedy Park	43.727929	-79.262029	0	0	0	0	0	0
7	Clairlea,Golden Mile,Oakridge	43.711112	-79.284577	1	1	1	1	0	0
8	Cliffcrest,Cliffside,Scarborough Village	43.716316	-79.239476	0	0	0	0	0	0

Figure 6 : Table with group formation

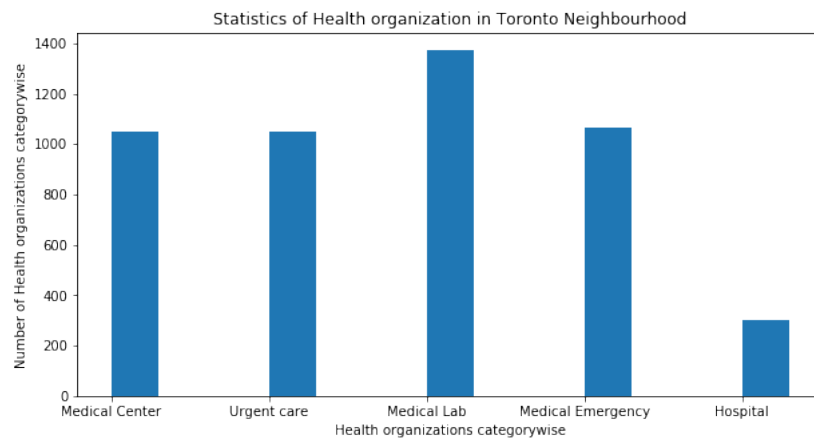


Figure 7 - Bar graph of health organizations

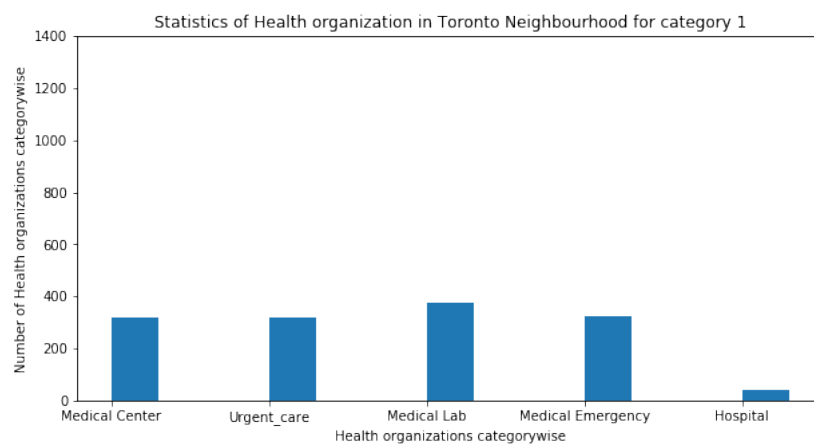


Figure 8 - Medical organizations count in category 1

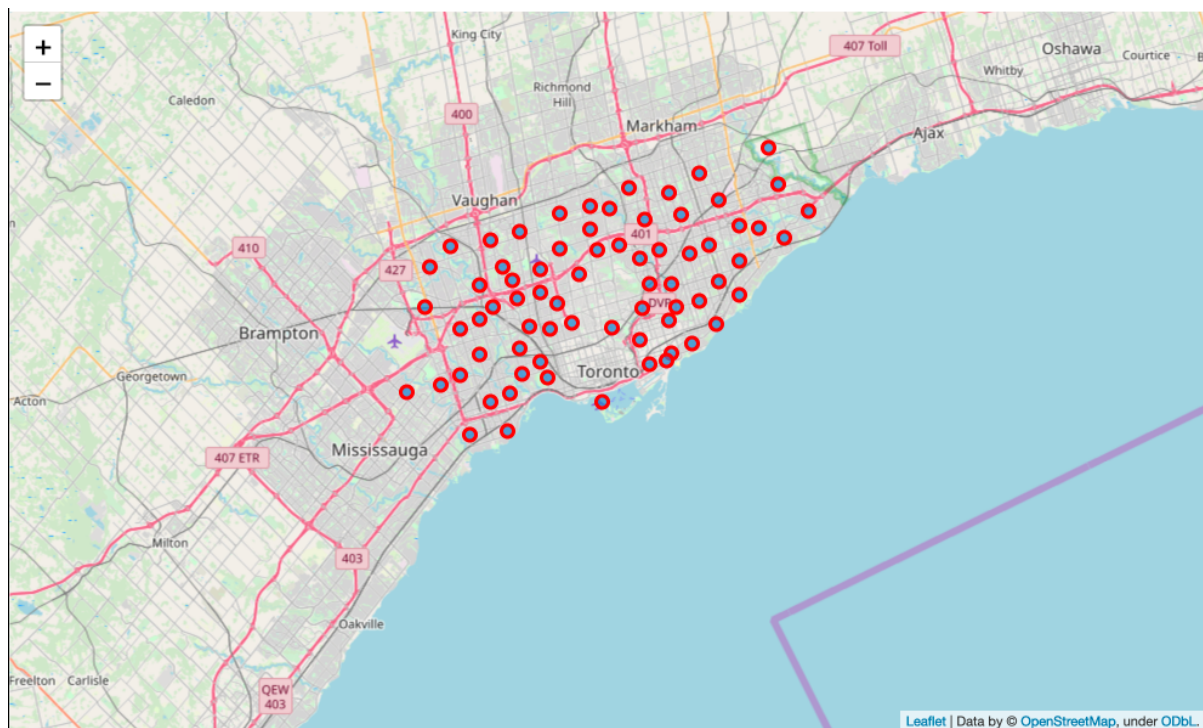


Figure 9 – Neighbourhoods with Medical venues of Category1 highlighted with red color

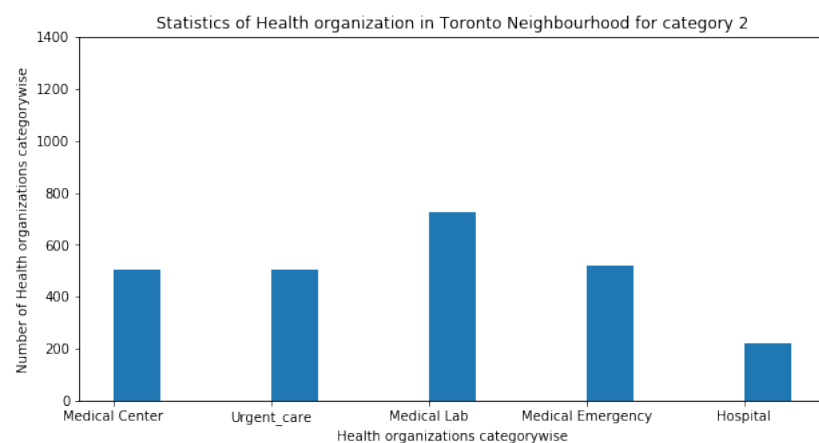


Figure 10 - Medical organizations count in category 2 Neighbourhood

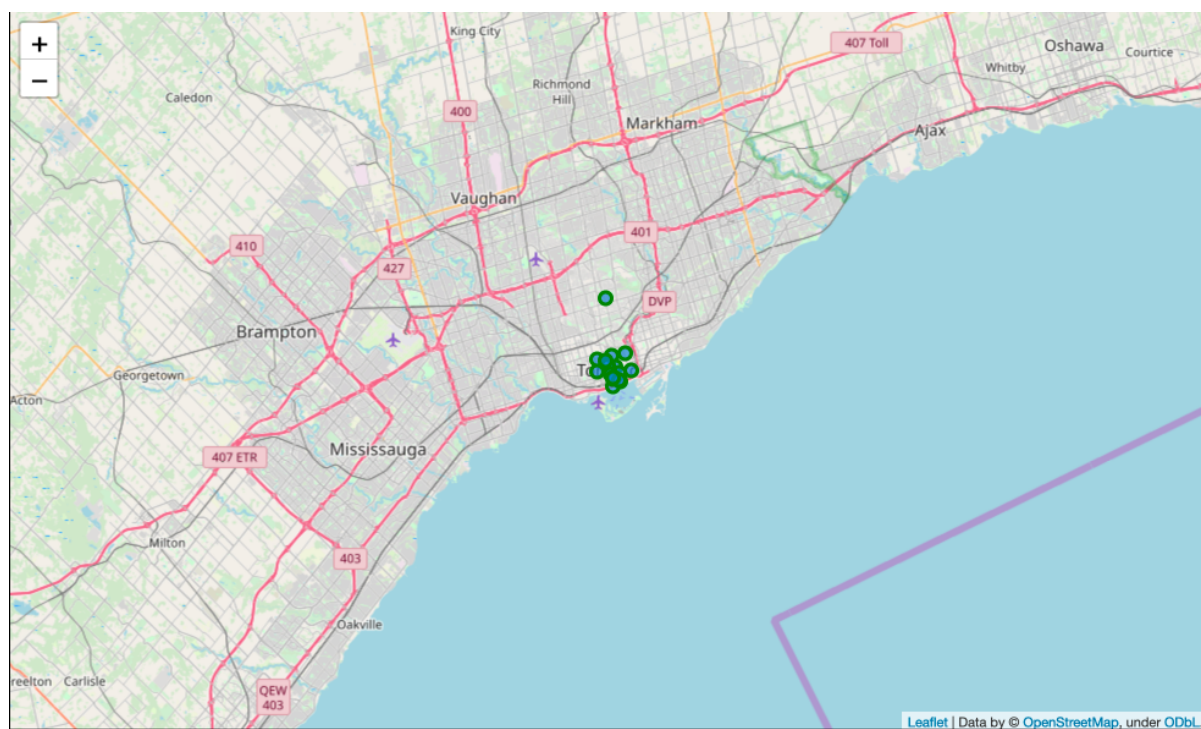


Figure 11 – Map with neighbourhoods with Medical venues of Category2 highlighted with green colour

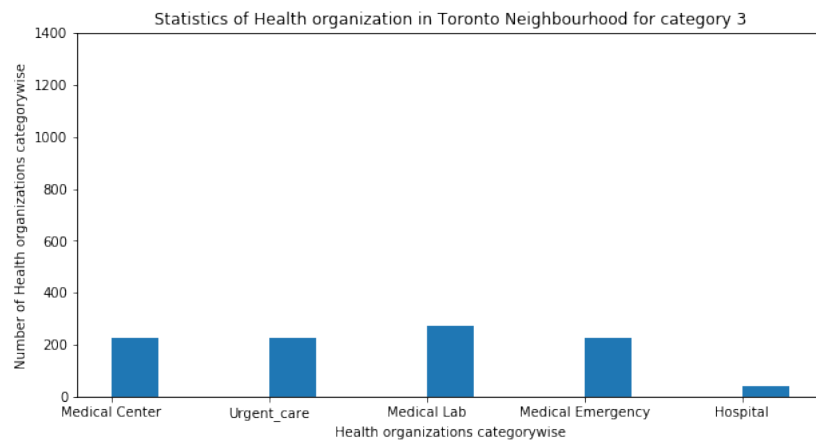


Figure 12 - Medical organizations count in category 3 Neighbourhood

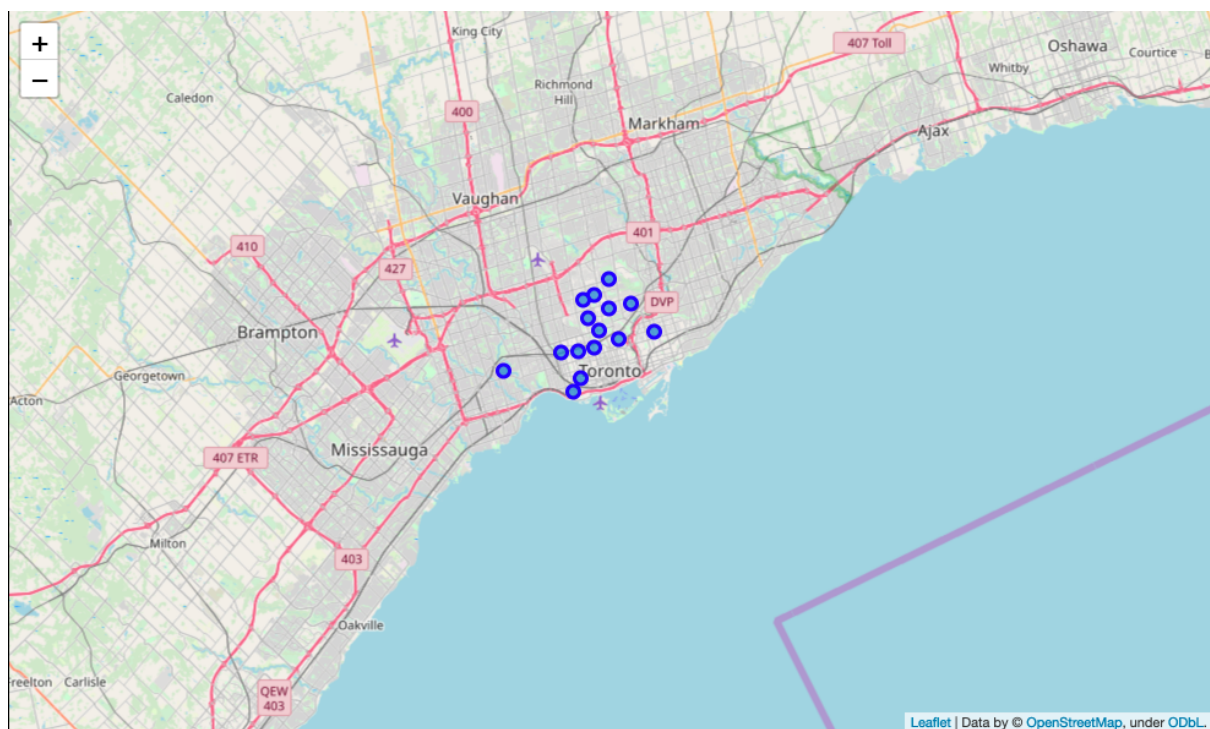


Figure 13 – Map with neighbourhoods with Medical venues of Category3 highlighted with blue colour

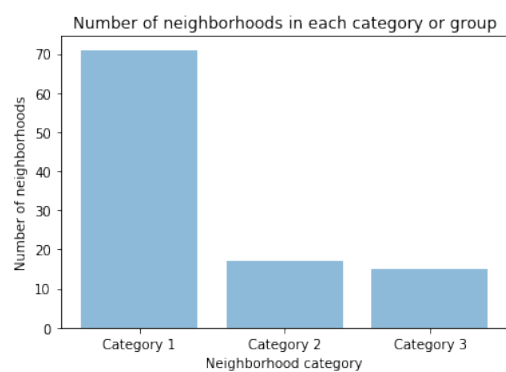


Figure 14 - Number of neighbourhoods in each category

Discussion and Inferences:

Following points observed/inferred from the bar-charts and maps:

- As per figure 14 and 8 and bulleted comments in result section, there are highest number of neighbourhoods in category 1 multiple times the other two categories. But number of health care facilities very few.
- As per Figure 14 and 10 and bulleted comments in result section, number of health care facilities in category 2 are much more than the sum of other two while the number of neighbourhoods is multiple times lower than category 1.

Recommendation:

Based on discussion points, I would recommend:

- Interested investors / business persons can look for business opportunities in neighbourhood areas grouped in category 1. This is also marked in map of figure 9.
- Best avoid exploring in category2 neighbourhood, marked in map of figure 11.

Conclusion

The business goal of survey of health care facilities in neighborhoods of Toronto and categorization of neighborhoods is successfully completed. Consequently, a recommendation is provided.

However, this work needs to be further explored from a humanitarian angle:

- In many of 2nd and 3rd world countries, health care facilities are insufficient or not available.
- Relevant dataset should be explored or made available for effective analysis.
- The findings can be shared with medical support organizations, NGOs etc.