APPLIED DATA SCIENCE WITH IBM SPECIALIZATION

BY MANASH SARMA

PRESENTATION OUTLINE

- Problem statement
- Dataset
- Use cases
- Data exploration
- Data preprocessing
- Results
- Discussion and Recommendation
- Conclusion

PROBLEM STATEMENT

Categorization of neighborhoods as per availability of healthcare facilities within 2-kilometer radius

– A data analysis approach .

DATASET AND SOURCE

Wikipedia site for Canada with table of postal codes https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M

Geographical coordinates of each postal code can be found from http://cocl.us/Geospatial_data

The geospatial data set provides latitude and longitude information for each postal code mentioned in the Wikipedia site.

USE CASES

Identify	Do a statistical survey / data analysis of existing medical facilities in different neighborhoods in Toronto within a radius of 2000 meters (2 Kilometers)
Identify	Make a group or clustering of the neighborhoods based on existing health care facilities available.
Achieve	Help explore opportunities in health care industry in the region with clustering analysis (thereby help humanity as all).

METHODOLOGY

Data Exploration

Data type verification of wikipedia site data of Toronto

```
df_merged.dtypes.sort_values()

Latitude        float64
Longitude       float64
Postalcode       object
Borough          object
Neighborhood     object
dtype: object
```

Explore the data in tabular form

	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

Explored neighbourhood data with four square API to get different medical venue detail.

METHODOLOGY

Preprocessing of data set

Convert the data to another tabular form to fulfil actual goal of survey / analysis of medical facilities

Considered five major facilities in the table –

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

	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

METHODOLODY

- **Assumptions based on foursquare API and other websites:**

- 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.

METHODOLOGY

- **Scoring method for clustering:**

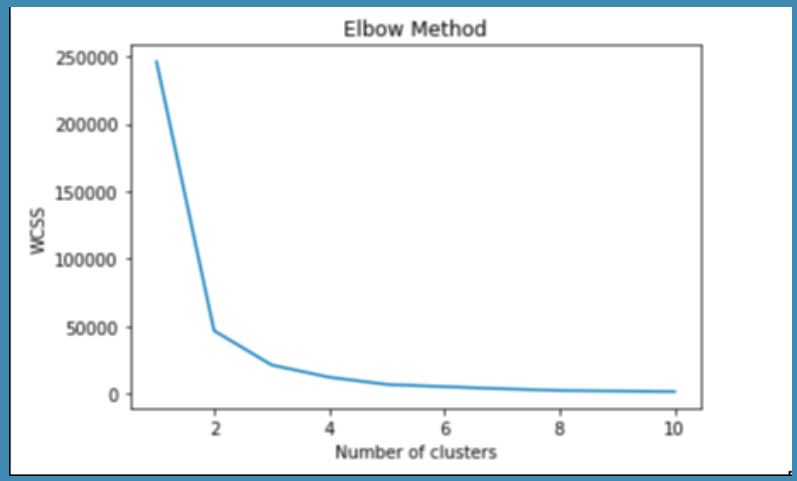
- For each neighbourhood , a score evaluated based on number of medical facilities in the locality:
- Score of a neighbourhood = 3 * No of Medical centers + No of Hospitals + No of urgent cares + No of Medical labs + No of emergency facilities.
- (Please note that a medical center's weightage is considered as 3 times more.)
- Score provided to each neighbourhood is used for clustering all the neighbourhoods.

METHODOLOGY

Determining number of clusters Technically:

With score assigned, number of clusters for all neighborhoods is determined by using elbow method.

Elbow method suggests three clusters in the dataset:



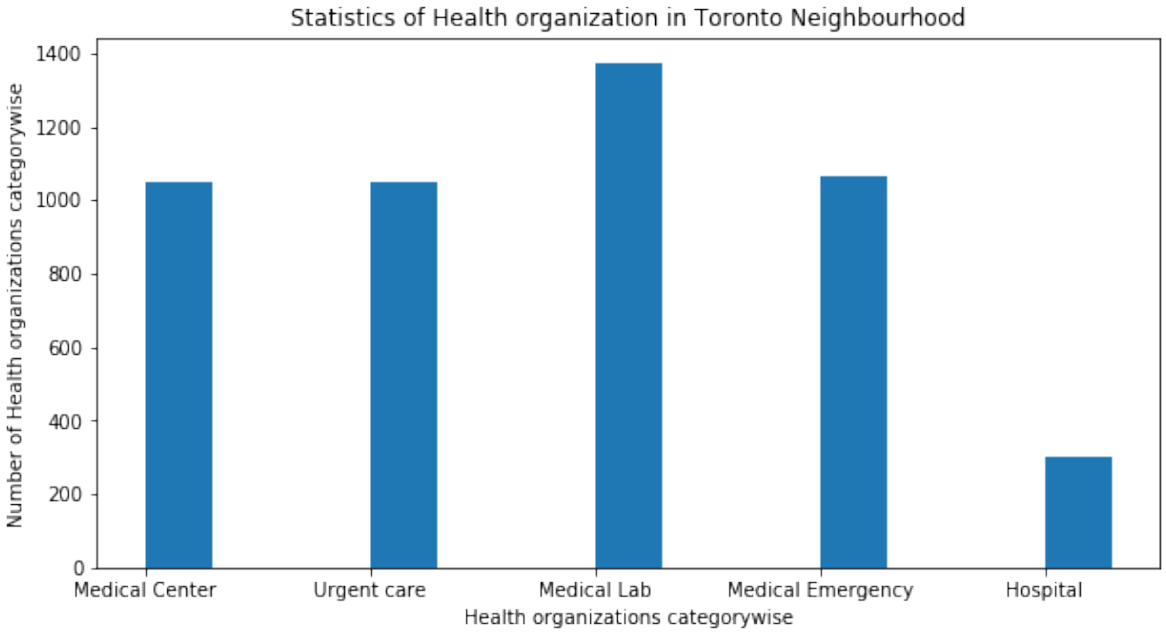
RESULTS

Table with clustered information at the last column with label ‘Medical Support Category’

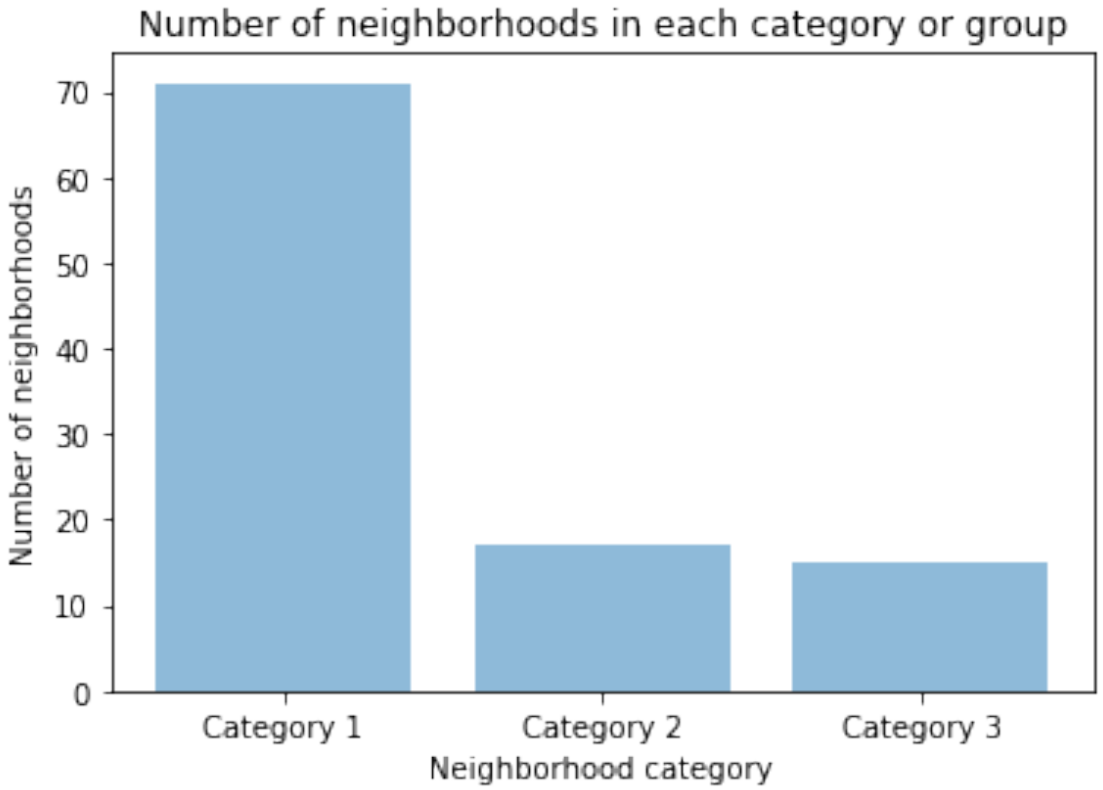
Three clustered identified with 0, 1 and 2.

	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

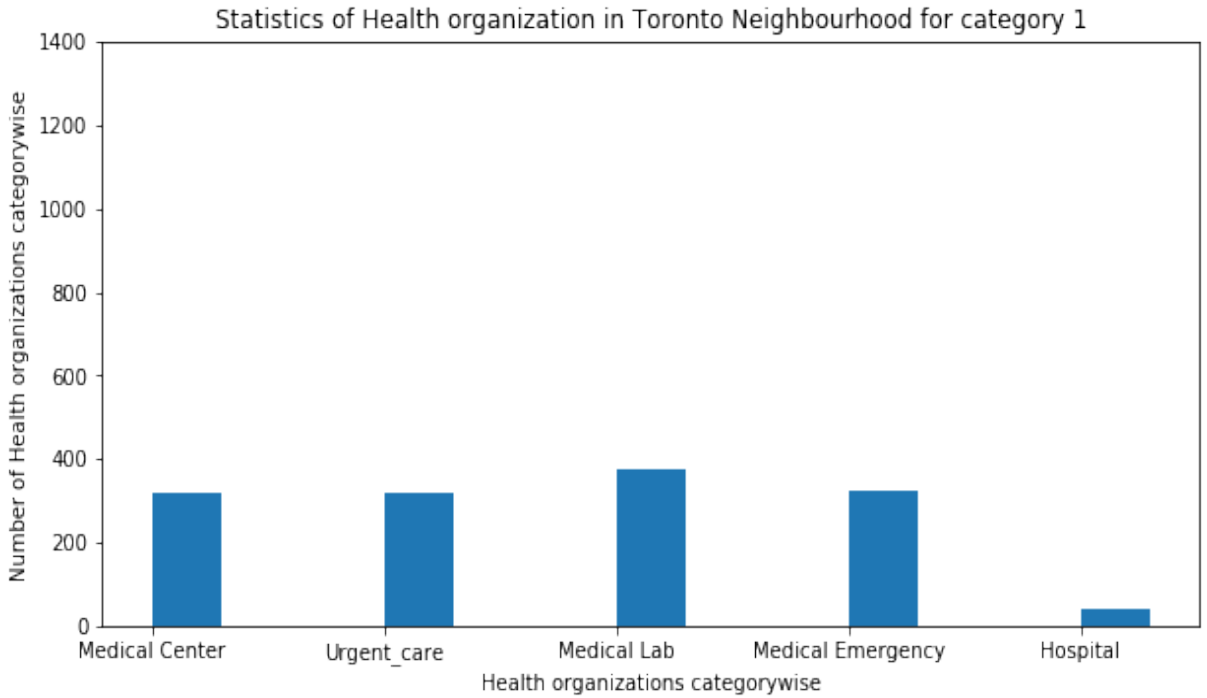
RESULTS


Statistics of Health organization in Toronto Neighbourhood
Number of Health organizations categorywise
0
200
400
600
800
1000
1200
1400
Medical Center
Urgent care
Medical Lab
Medical Emergency
Hospital
Health organizations categorywise

RESULTS

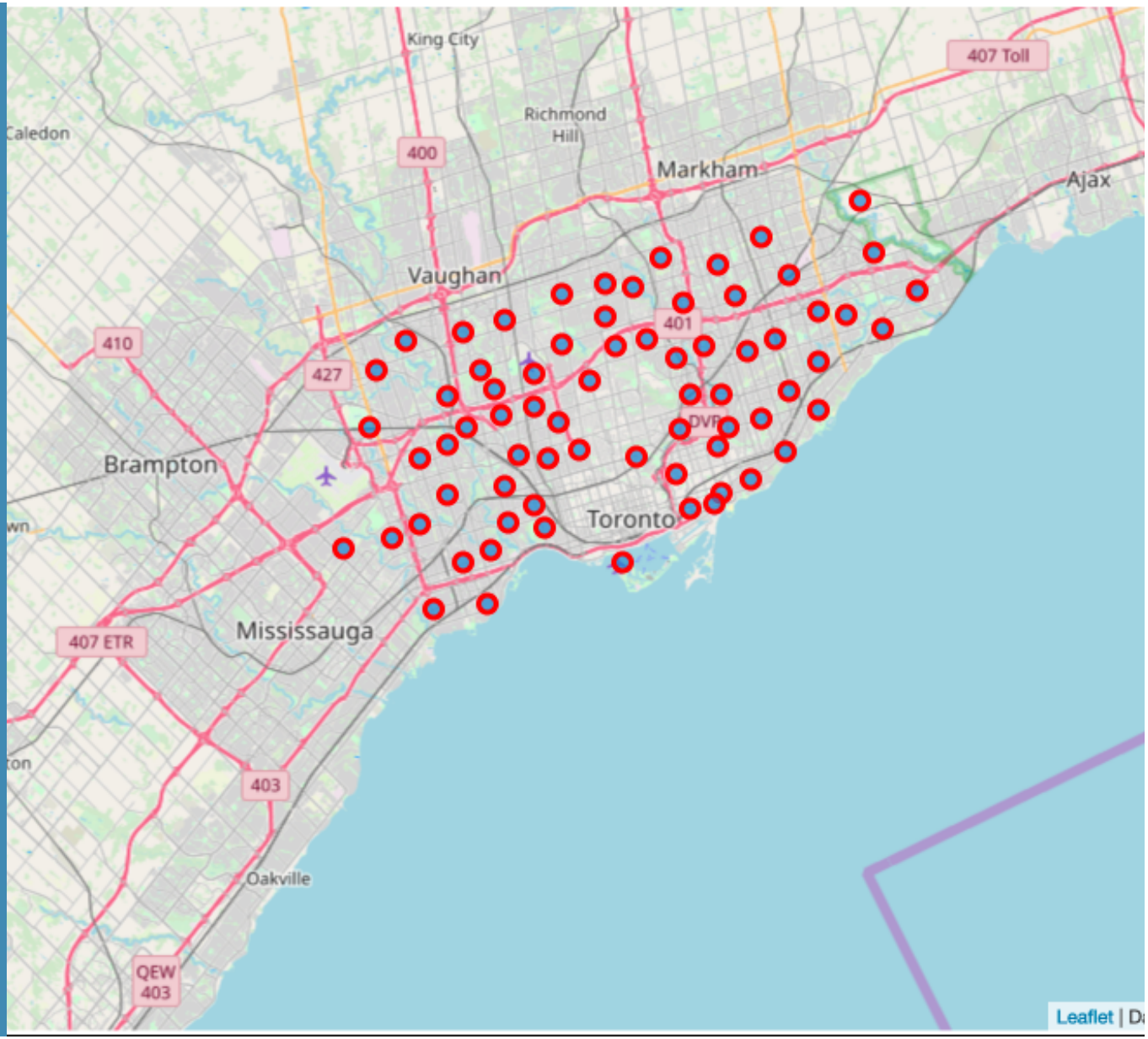

Number of neighborhoods in each category or group
Number of neighborhoods
0
10
20
30
40
50
60
70
Category 1
Category 2
Category 3
Neighborhood category

RESULTS

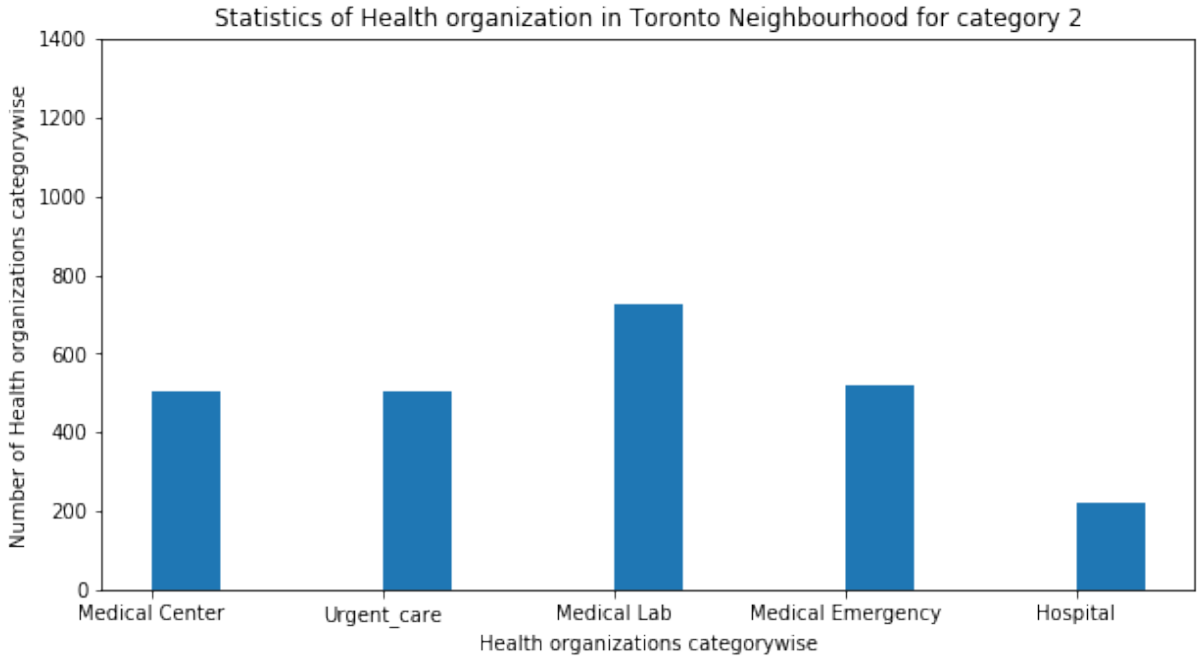

Statistics of Health organization in Toronto Neighbourhood for category 1
Number of Health organizations categorywise
1400
1200
1000
800
600
400
200
0
Medical Center
Urgent_care
Medical Lab
Medical Emergency
Hospital
Health organizations categorywise

RESULTS

(RED MARKED NEIGHBORHOOD IN MAP – CATEGORY1)

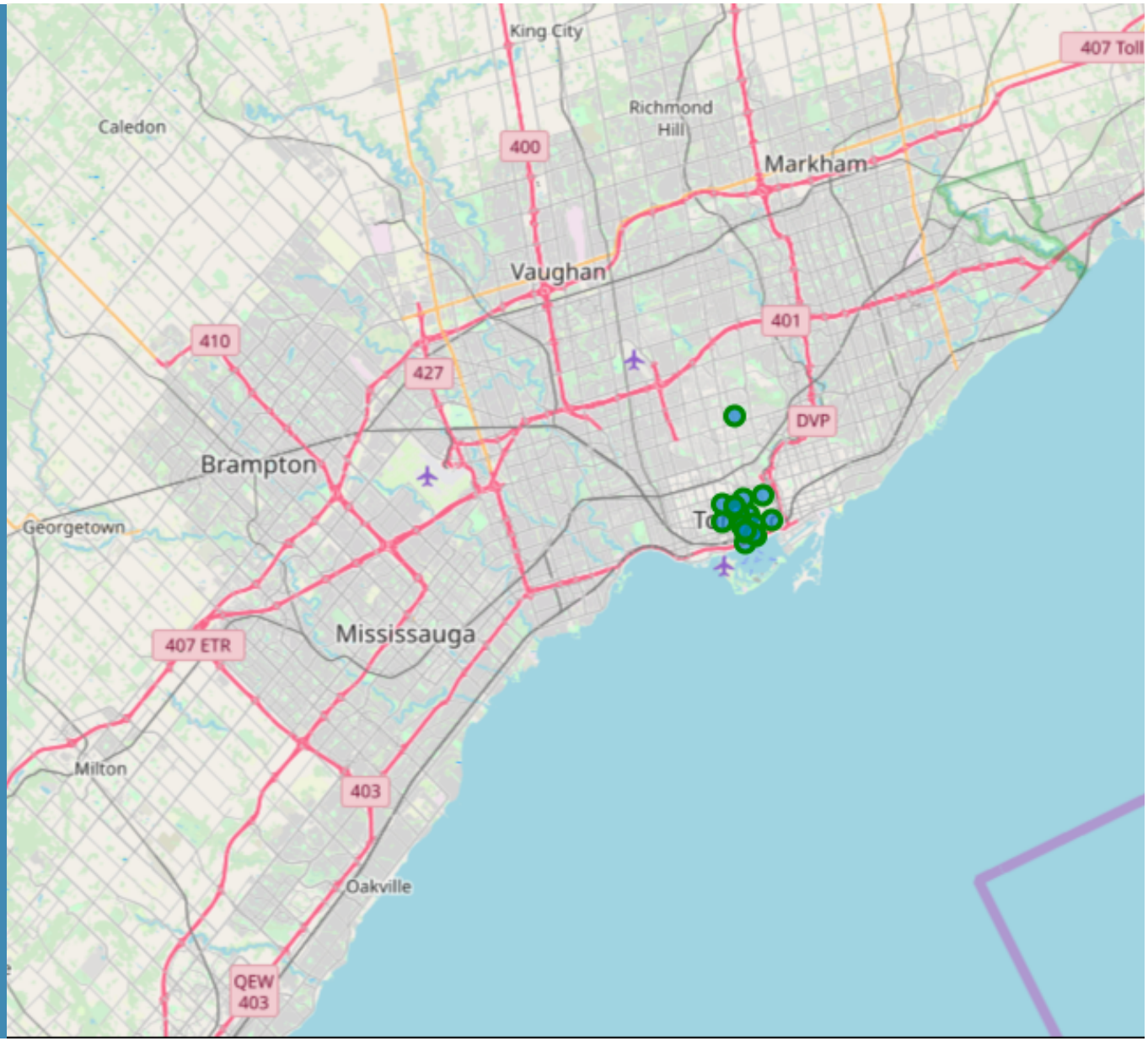


RESULTS

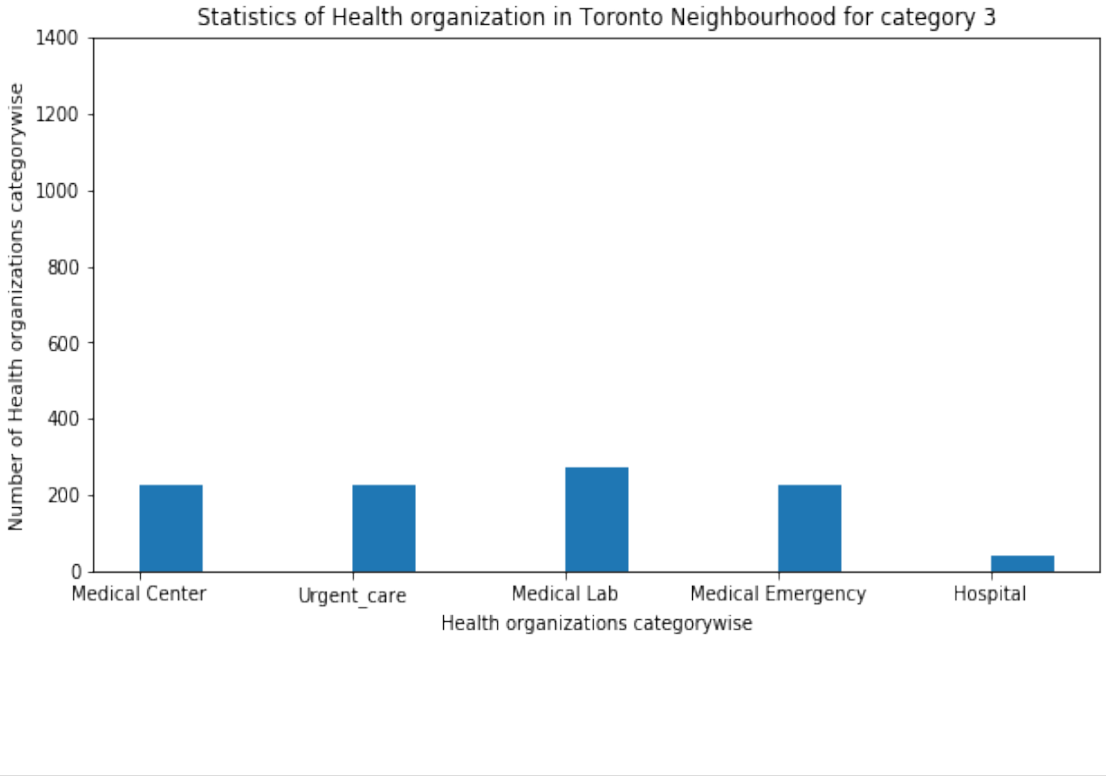

Statistics of Health organization in Toronto Neighbourhood for category 2
Number of Health organizations categorywise
1400
1200
1000
800
600
400
200
0
Medical Center
Urgent_care
Medical Lab
Medical Emergency
Hospital
Health organizations categorywise

RESULTS

(GREEN MARKED NEIGHBORHOOD IN MAP - CATEGORY2)

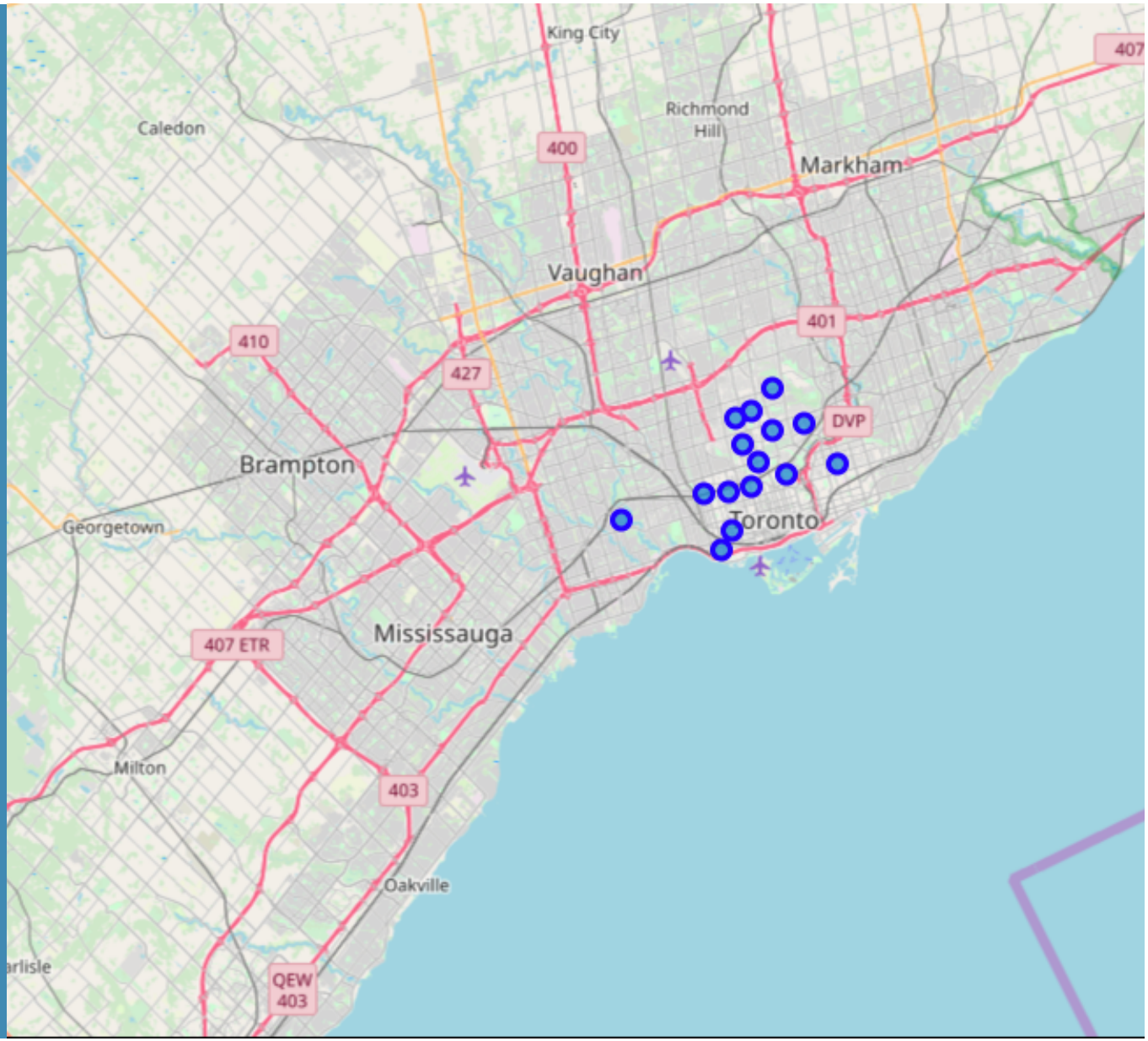


RESULTS


Statistics of Health organization in Toronto Neighbourhood for category 3
Number of Health organizations categorywise
1400
1200
1000
800
600
400
200
0
Medical Center
Urgent_care
Medical Lab
Medical Emergency
Hospital
Health organizations categorywise

RESULTS

(BLUE MARKED NEIGHBORHOOD IN MAP – CATEGORY3)



DISCUSSION AND INFERENCES

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

- There are highest number of neighbourhoods in category 1 multiple times the other two categories. But number of health care facilities very few.

- 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, marked in map of category 1.
- Best avoid exploring in category2 neighbourhood, marked in map of slide.

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 places of 2nd and 3rd word 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.