COURSERA IBM DATA SCIENCE CAPSTONE ASSIGNMENT

WHICH NEIGHBORHOODS OF TORONTO LACK SHOPPING MALLS

By Viatcheslav Liachenko

December 01, 2020

1. Introduction

1.1 Background of the problem

A shopping mall (aka mall, shopping center or supermarket) is a North American term for a large indoor shopping center, usually anchored by department stores (Wikipedia, 2020).

In recent years, the number of “dead” malls increased significantly in the early 21st century because the economic health of malls has been in decline, as identified by high vacancy rates.

Unfortunately, this is not only e-commerce and online shopping that caused such a problem. Some real estate experts say the "fundamental problem" is a glut of malls in some parts of the country creating a market that is "extremely over-retailed" (Schwartz, Nelson D. 2015. "The Economics (and Nostalgia) of Dead Malls". *The New York Times*). On the other hand, there are many neighborhoods, even in large cities, which do not have a shopping mall within a distance of a few kilometers.

In this capstone project, I used available data to study the shopping malls distribution in the city of Toronto to see which part of the city have too much and which parts are in lack of shopping malls.

1.2 Research relevance

This research will be of value to the municipalities of the neighborhoods of Toronto as well as to real estate developers. It will help both mentioned stakeholders to better plan the development of the city of Toronto in a manner that would make new shopping malls more profitable and being placed within the reach of the people residing in Toronto neighborhoods.

2. Data Acquisition & Processing

2.1 Data sources

For the final capstone project assignment, we continue to build upon the data obtained during the third week of the studies.

The necessity of data for the Toronto's neighborhoods was solved by utilization of Wikipedia's source of Toronto, CA neighborhoods, which was scrapped into the notebook's Pandas dataframe using BeautifulSoup library. The geographical coordinates were imported from the available geospatial_data file.

2.2 Data processing

Data preprocessing, first of all, included removing of "\n" tags and "not assigned" boroughs in some postal codes from the original scrapped dataframe. Also, due to the nature of the task, I removed "Borough" column as it was unnecessary for the research.

Then, the neighborhoods and coordinates were merged into a single dataframe presenting 103 neighborhoods in Toronto, with their postal codes and coordinates, shown in the table below.

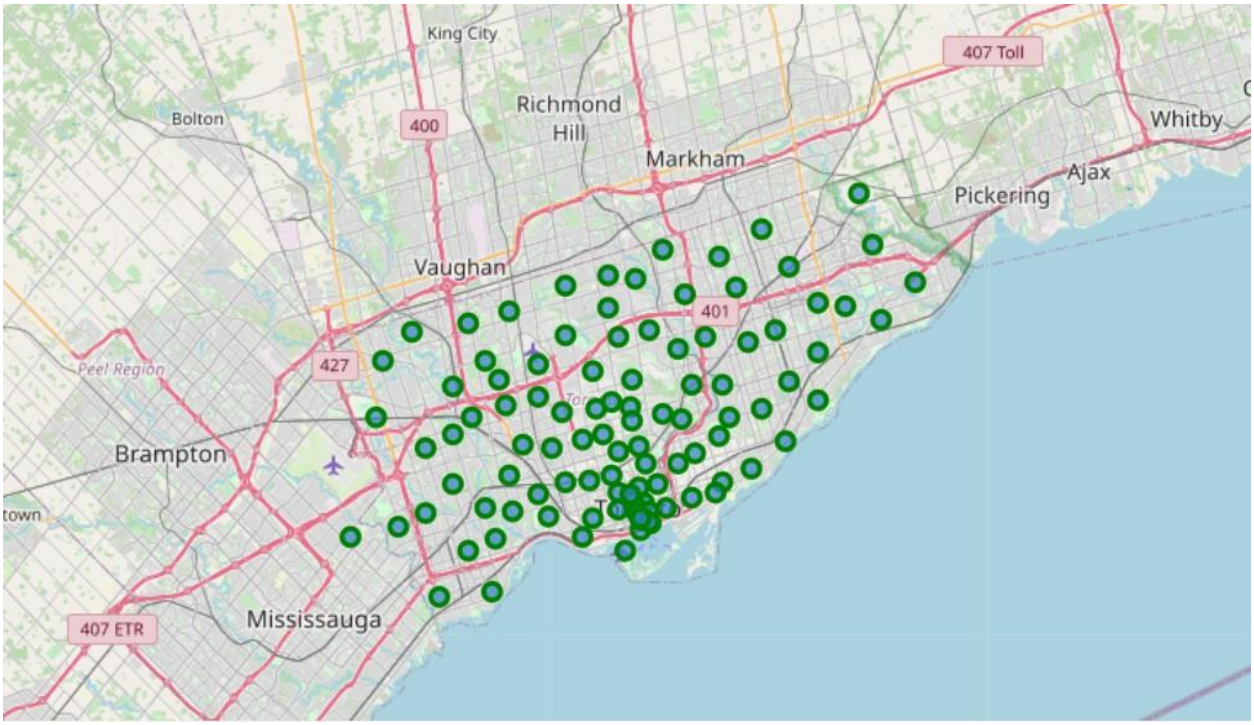
Exhibit 1: Toronto Neighborhoods Count

	PostalCode	Neighborhood	Latitude	Longitude
1	M1B	Malvern, Rouge	43.806686	-79.194353
2	M1C	Rouge Hill, Port Union, Highland Creek	43.784535	-79.160497
3	M1E	Guildwood, Morningside, West Hill	43.763573	-79.188711
4	M1G	Woburn	43.770992	-79.216917
5	M1H	Cedarbrae	43.773136	-79.239476
...	...	...	...	...
170	M9N	Weston	43.706876	-79.518188
171	M9P	Westmount	43.696319	-79.532242
172	M9R	Kingsview Village, St. Phillips, Martin Grove ...	43.688905	-79.554724
175	M9V	South Steeles, Silverstone, Humbergate, Jamest...	43.739416	-79.588437
176	M9W	Northwest, West Humber - Clairville	43.706748	-79.594054

103 rows × 4 columns

After the installation of the necessary Python libraries including Nominatim, Jason_Normalize, Folium and K-Means, and adding markers, I was able to produce the following Toronto map with neighborhoods.

Exhibit 2: Toronto Map with Neighborhood Markers



The following step was to use Foursquare to find and filter the venues, grouping them by the neighborhood and using OneHot encoding to convert categorical data into numerical values.

The obvious following action was to filter the dataframe leaving only the neighborhoods with the “Shopping Mall” venues grouped by the neighborhoods. The resulting dataframe showed 99, out of 103, neighborhoods that had shopping malls.

Exhibit 3: Toronto Neighborhoods with Shopping Malls

	Neighborhoods	Shopping Mall
0	Agincourt	0.010000
1	Alderwood, Long Branch	0.010000
2	Bathurst Manor, Wilson Heights, Downsview North	0.010000
3	Bayview Village	0.020000
4	Bedford Park, Lawrence Manor East	0.010000
...	...	...
94	Willowdale, Willowdale West	0.000000
95	Woburn	0.000000
96	Woodbine Heights	0.000000
97	York Mills West	0.011765
98	York Mills, Silver Hills	0.010000

99 rows × 2 columns

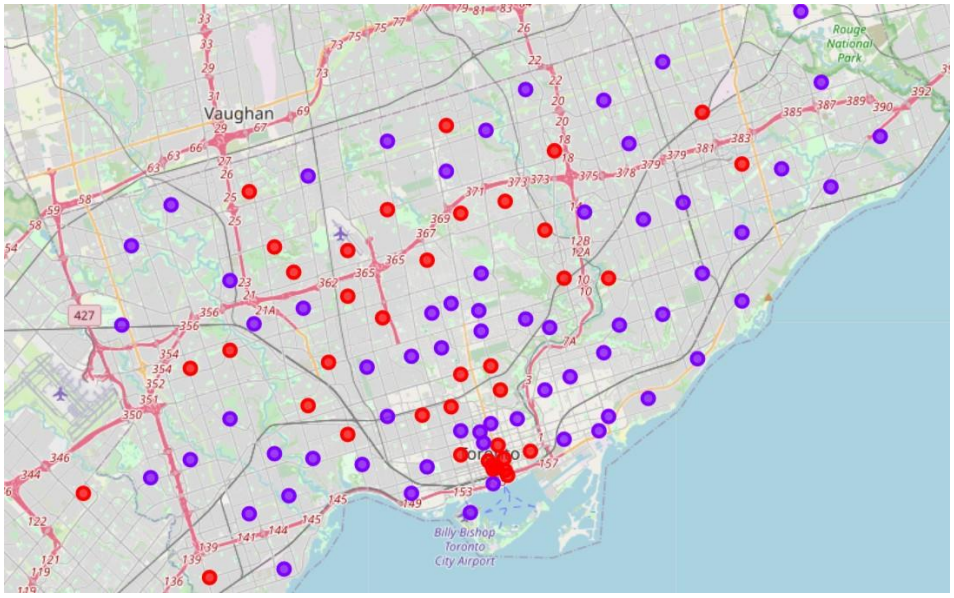
The consequent usage of K-Means library allowed me to cluster the original dataframe with 103 neighborhoods with postal codes and coordinates (Exhibit 4 below) into three clusters, which, after having them sorted by the neighborhood, produced the following Toronto clusters map that is shown in Exhibit 5.

Exhibit 4: Toronto Clusters Dataframe

	Neighborhood	Shopping Mall	Cluster Labels	PostalCode	Latitude	Longitude
0	Agincourt	0.010000	1	M1S	43.794200	-79.262029
1	Alderwood, Long Branch	0.010000	1	M8W	43.602414	-79.543484
2	Bathurst Manor, Wilson Heights, Downsview North	0.010000	1	M3H	43.754328	-79.442259
3	Bayview Village	0.020000	2	M2K	43.786947	-79.385975
4	Bedford Park, Lawrence Manor East	0.010000	1	M5M	43.733283	-79.419750
...	...	...	...	...	...	...
94	Willowdale, Willowdale West	0.000000	0	M2R	43.782736	-79.442259
95	Woburn	0.000000	0	M1G	43.770992	-79.216917
96	Woodbine Heights	0.000000	0	M4C	43.695344	-79.318389
97	York Mills West	0.011765	1	M2P	43.752758	-79.400049
98	York Mills, Silver Hills	0.010000	1	M2L	43.757490	-79.374714

103 rows × 6 columns

Exhibit 5: Toronto Clusters Map



3. Exploratory Data Analysis

3.1 Examination of clusters 0,1 and 2

There are 55 neighborhoods in Toronto that are marked with cluster 0 showing a lack of shopping malls in those neighborhoods, which is reflected in the Exhibit 6 partially shown the corresponding dataframe (the full list shown in the following Exhibit 6a).

Exhibit 6: Cluster 0 – Lack of Shopping Malls

	Neighborhood	Shopping Mall	Cluster Labels	PostalCode	Latitude	Longitude
47	Lawrence Park	0.0	0	M4N	43.728020	-79.388790
30	Forest Hill North & West, Forest Hill Road Park	0.0	0	M5P	43.696948	-79.411307
72	South Steeles, Silverstone, Humbergate, Jamest...	0.0	0	M9V	43.739416	-79.588437
33	Golden Mile, Clairlea, Oakridge	0.0	0	M1L	43.711112	-79.284577
34	Guildwood, Morningside, West Hill	0.0	0	M1E	43.763573	-79.188711
35	Harbourfront East, Union Station, Toronto Islands	0.0	0	M5J	43.640816	-79.381752
71	Scarborough Village	0.0	0	M1J	43.744734	-79.239476
37	Hillcrest Village	0.0	0	M2H	43.803762	-79.363452
38	Humber Summit	0.0	0	M9L	43.756303	-79.565963
39	Humberlea, Emery	0.0	0	M9M	43.724766	-79.532242
40	Humewood-Cedarvale	0.0	0	M6C	43.693781	-79.428191
41	India Bazaar, The Beaches West	0.0	0	M4L	43.668999	-79.315572
43	Kennedy Park, Ionview, East Birchmount Park	0.0	0	M1K	43.727929	-79.262029
69	Runnymede, Swansea	0.0	0	M6S	43.651571	-79.484450
68	Rouge Hill, Port Union, Highland Creek	0.0	0	M1C	43.784535	-79.160497
67	Roselawn	0.0	0	M5N	43.711695	-79.416936

Exhibit 6a: List of Neighborhoods Lacking Shopping Malls

['Willowdale, Willowdale East', 'St. James Town, Cabbagetown', 'Bayview Village', 'Islington Avenue, Humber Valley Village', 'University of Toronto, Harbord', "Queen's Park, Ontario Provincial Government", 'Parkwoods', 'Church and Wellesley', 'Central Bay Street', 'Don Mills', 'Canada Post Gateway Processing Centre', 'York Mills West', 'Westmount', 'Moore Park, Summerhill East', 'Lawrence Manor, Lawrence Heights', 'Alderwood, Long Branch', 'Richmond, Adelaide, King', 'Kingsview Village, St. Phillips, Martin Grove Gardens, Richview Gardens', 'Kensington Market, Chinatown, Grange Park', 'Regent Park, Harbourfront', 'Agincourt', 'Rosedale', 'Runnymede, The Junction North', 'St. James Town', 'Glencairn', 'Stn A PO Boxes', 'Summerhill West, Rathnelly, South Hill, Forest Hill SE, Deer Park', 'The Annex, North Midtown, Yorkville', 'Toronto Dominion Centre, Design Exchange', 'Victoria Village', 'Willowdale, Newtonbrook', 'High Park, The Junction South', 'York Mills, Silver Hills', 'Commerce Court, Victoria Hotel', 'First Canadian Place, Underground city', 'Bathurst Manor, Wilson Heights, Downsview North', 'Garden District, Ryerson', 'Cedarbrae', 'Christie', 'Del Ray, Mount Dennis, Keelsdale and Silverthorn', 'Bedford Park, Lawrence Manor East', 'Berczy Park', 'Fairview, Henry Farm, Oriole', 'Downsview', 'The Danforth West, Riverdale', 'Cliffside, Cliffcrest, Scarborough Village West', "Clarks Corners, Tam O'Shanter, Sullivan", "Steeles West, L'Amoreaux West", 'South Steeles, Silverstone, Humbergate, Jamestown, Mount Olive, Beaumond Heights, Thistletown, Albion Gardens', 'Studio District', 'Scarborough Village', 'Caledonia-Fairbanks', 'The Beaches', 'Golden Mile, Clairlea, Oakridge', 'The Kingsway, Montgomery Road, Old Mill North', 'Thorncliffe Park', 'CN Tower, King and Spadina, Railway Lands, Harbourfront West, Bathurst Quay, South Niagara, Island airport', 'Runnymede, Swansea', 'Upper Rouge', 'Brockton, Parkdale Village, Exhibition Place', 'West Deane Park, Princess Gardens, Martin Grove, Islington, Cloverdale', 'Birch Cliff, Cliffside West', 'Weston', 'Wexford, Maryvale', 'Willowdale, Willowdale West', 'Woburn', 'Woodbine Heights', 'Business reply mail Processing Centre, South Central Letter Processing Plant Toronto', 'Davisville', 'Rouge Hill, Port Union, Highland Creek', 'Roselawn', 'Harbourfront East, Union Station, Toronto Islands', 'Hillcrest Village', 'Humber Summit', 'Humberlea, Emery', 'Humewood-Cedarvale', 'India Bazaar, The Beaches West', 'Forest Hill North & West, Forest Hill Road Park', 'Kennedy Park, Ionview, East Birchmount Park', 'Eringate, Bloordale Gardens, Old Burnhamthorpe, Markland Wood', 'Lawrence Park', 'Leaside', 'East Toronto, Broadview North (Old East York)', 'Malvern, Rouge', "Milliken, Agincourt North, Steeles East, L'Amoreaux East", 'Mimico NW, The Queensway West, South of Bloor, Kingsway Park South West, Royal York South West', 'Dufferin, Dovercourt Village', 'New Toronto, Mimico South, Humber Bay Shores', 'North Park, Maple Leaf Park, Upwood Park', 'North Toronto West, Lawrence Park', 'Northwest, West Humber - Clairville', 'Northwood Park, York University', "Old Mill South, King's Mill Park, Sunnylea, Humber Bay, Mimico NE, The Queensway East, Royal York South East, Kingsway Park South East", 'Parkdale, Roncesvalles', 'Parkview Hill, Woodbine Gardens', 'Dorset Park, Wexford Heights, Scarborough Town Centre', 'Davisville North', 'Guildwood, Morningside, West Hill', 'Little Portugal, Trinity']

The moderate number of shopping malls is shown in cluster 1, which counts for 39 neighborhoods, which is reflected in the following dataframe in Exhibit7 below.

Exhibit 7: Cluster 1 – Moderate Number of Shopping Malls

	Neighborhood	Shopping Mall	Cluster Labels	PostalCode	Latitude	Longitude
76	Stn A PO Boxes	0.010000	1	M5W	43.646435	-79.374846
78	Summerhill West, Rathnelly, South Hill, Forest...	0.010000	1	M4V	43.686412	-79.400049
92	Willowdale, Newtonbrook	0.010000	1	M2M	43.789053	-79.408493
73	St. James Town	0.010000	1	M5C	43.651494	-79.375418
70	Runnymede, The Junction North	0.010000	1	M6N	43.673185	-79.487262
65	Richmond, Adelaide, King	0.010000	1	M5H	43.650571	-79.384568
89	Westmount	0.012048	1	M9P	43.696319	-79.532242
66	Rosedale	0.010000	1	M4W	43.679563	-79.377529
87	Victoria Village	0.010000	1	M4A	43.725882	-79.315572
84	Toronto Dominion Centre, Design Exchange	0.010000	1	M5K	43.647177	-79.381576
79	The Annex, North Midtown, Yorkville	0.010000	1	M5R	43.672710	-79.405678
64	Regent Park, Harbourfront	0.010000	1	M5A	43.654260	-79.360636
0	Agincourt	0.010000	1	M1S	43.794200	-79.262029
22	Don Mills	0.015000	1	M3C	43.725900	-79.340923
1	Alderwood, Long Branch	0.010000	1	M8W	43.602414	-79.543484

The following 9 neighborhoods, belonging to cluster 2, show the highest number of shopping malls (Exhibit 8 below).

Exhibit 8: Cluster 2 – Highest Number of Shopping Malls

	Neighborhood	Shopping Mall	Cluster Labels	PostalCode	Latitude	Longitude
3	Bayview Village	0.020000	2	M2K	43.786947	-79.385975
93	Willowdale, Willowdale East	0.022222	2	M2N	43.770120	-79.408493
13	Central Bay Street	0.020000	2	M5G	43.657952	-79.387383
15	Church and Wellesley	0.020000	2	M4Y	43.665860	-79.383160
85	University of Toronto, Harbord	0.020000	2	M5S	43.662696	-79.400049
74	St. James Town, Cabbagetown	0.020000	2	M4X	43.667967	-79.367675
63	Queen's Park, Ontario Provincial Government	0.020000	2	M7A	43.662301	-79.389494
42	Islington Avenue, Humber Valley Village	0.020000	2	M9A	43.667856	-79.532242
62	Parkwoods	0.020000	2	M3A	43.753259	-79.329656

4. Conclusion & Recommendation

According to the results obtained from my Jupyter notebook, there are 44 shopping malls in Toronto city. The good news is that most of the shopping malls (cluster 2) in the city of Toronto are more or less evenly spread throughout the city, with a slight overconcentration in the downtown Toronto as it is shown on the Toronto Cluster Map in red color (Exhibit 5). Moderate number of shopping malls, which are also spread rather evenly, are in cluster 1, which is reflected in blue color on the Toronto Cluster Map.

On the other hand, cluster 0 reflects lacking shopping malls in the neighborhoods. This cluster contains 55 neighborhoods shown in the partial dataframe in Exhibit 6, with a full list of neighborhoods shown in Exhibit 6a. This represents a great opportunity for the municipalities and real estate developers as it has a high potential for opening new shopping malls in a manner that would be more profitable, less competitive and providing greater access to the malls by the people who live in those neighborhoods.

Overconcentration of shopping malls, as it has been already mentioned in the introductory part of this report, may lead to so-called “glut of malls” in some part of Toronto.

Therefore, the city planners of Toronto have done relatively good job by evenly spreading the malls over the territory of the city. This, however, does not mean that there are no overconcentration or lacking. It is recommended to further study the necessity of shopping malls construction in the areas that lack shopping malls.