

IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

The Battle of Neighborhoods – A new gym in Toronto

Prepared by Maciej Skoczny

Background

Our client, Jane Doe, is looking for a suitable location in Toronto to open her new gym. She has experience in the fitness industry and runs a handful of fitness studios in Canada's other regions. She would like to expand her Canadian business to Toronto; however, she doesn't know the city well.

Business Problem

What Toronto neighborhood/borough should Jane open her new gym in? It is crucial to solving this problem by providing appropriate analysis and conclusions so that Jane can make profitable business decisions.

Target Audience

Jane Doe is our target audience, as she's our client. The report could be useful for all the entrepreneurs who want to open a new fitness venue in Toronto.

Data and Data Sources

- List of Toronto neighborhoods and boroughs (including postal codes): https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Geographical coordinates of Toronto, including neighborhoods and boroughs (latitude, longitude): GeoCoder pack
- Gym venue data, including addresses and types - <https://foursquare.com/>

Proposed Solution

To solve the business problem and help Jane find a suitable sport for her new gym in Toronto, we'll use the data science methods to determine which neighborhood of Toronto has the smallest number of gyms. We'll combine the data extracted by Foursquare API with the geographical coordinates from Wikipedia and find an area with the smallest number of gyms. We'll use the Jupyter Notebook on the Anaconda platform and Python programming language. We'll use the following data science techniques:

- Data extraction
- Data scrapping
- Data segmentation
- Data clustering
- Data visualization

Methodology - Phase 1 - Toronto Data

- Libraries import
- Extraction of Toronto neighborhoods and associated postal codes from Wikipedia
- Scrapping web data and creating and cleaning data frame.
- Extracting the geographical coordinates and merging both data frames

	index	Postalcode	Borough	Neighbourhood	Latitude	Longitude
	0	2	M3A	North York	Parkwoods	43.75245 -79.32991
	1	3	M4A	North York	Victoria Village	43.73057 -79.31306
	2	4	M5A	Downtown Toronto	Regent Park, Harbourfront	43.65512 -79.36264
	3	5	M6A	North York	Lawrence Manor, Lawrence Heights	43.72327 -79.45042
	4	6	M7A	Downtown Toronto	Queen's Park, Ontario Provincial Government	43.66253 -79.39188

	98	160	M8X	Etobicoke	The Kingsway, Montgomery Road, Old Mill North	43.65319 -79.51113
	99	165	M4Y	Downtown Toronto	Church and Wellesley	43.66659 -79.38133
	100	168	M7Y	East Toronto	Business reply mail Processing Centre, South C...	43.64869 -79.38544
	101	169	M8Y	Etobicoke	Old Mill South, King's Mill Park, Sunnylea, Hu...	43.63278 -79.48945
	102	178	M8Z	Etobicoke	Mimico NW, The Queensway West, South of Bloor,...	43.62513 -79.52681

103 rows x 6 columns

Methodology - Phase 2 - Venues Data

- Extracting venue data from Foursquare API and creating data frame
- Filtering and sorting gym/fitness venue data
- Clustering neighborhoods into 3 clusters
- Merging venues data with Toronto geographical data

	Neighborhood	Gym / Fitness Center	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Berczy Park	0.0	2	43.64536	-79.37306	The Keg Steakhouse + Bar - Esplanade	43.646712	-79.374768	Restaurant
0	Berczy Park	0.0	2	43.64536	-79.37306	Fresh On Front	43.647815	-79.374453	Vegetarian / Vegan Restaurant
0	Berczy Park	0.0	2	43.64536	-79.37306	LCBO	43.642944	-79.372440	Liquor Store
0	Berczy Park	0.0	2	43.64536	-79.37306	Meridian Hall	43.646292	-79.376022	Concert Hall
0	Berczy Park	0.0	2	43.64536	-79.37306	St. Lawrence Market (South Building)	43.648743	-79.371597	Farmers Market
...
38	University of Toronto, Harbord	0.0	2	43.66311	-79.40180	Second Cup (Miles Nadal JCC Fitness)	43.666527	-79.403872	Coffee Shop
38	University of Toronto, Harbord	0.0	2	43.66311	-79.40180	Shoppers Drug Mart	43.666562	-79.405007	Pharmacy
38	University of Toronto, Harbord	0.0	2	43.66311	-79.40180	Al Green Theatre	43.666547	-79.404053	Concert Hall
38	University of Toronto, Harbord	0.0	2	43.66311	-79.40180	Kenzo Japanese Noodle House	43.666295	-79.406002	Ramen Restaurant
38	University of Toronto, Harbord	0.0	2	43.66311	-79.40180	Swiss Chalet	43.666577	-79.404338	Restaurant

Results - Clusters

Cluster 0

	Neighborhood	Gym / Fitness Center	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
3	CN Tower, King and Spadina, Railway Lands, Har...	0.025641	0	43.64082	-79.39818	FlirtyGirl Fitness	43.644005	-79.397590	Gym / Fitness Center
3	CN Tower, King and Spadina, Railway Lands, Har...	0.025641	0	43.64082	-79.39818	Cykl	43.642778	-79.402361	Gym / Fitness Center
4	Central Bay Street	0.016129	0	43.65609	-79.38493	Hard Candy Fitness	43.659556	-79.382440	Gym / Fitness Center
24	Regent Park, Harbourfront	0.041667	0	43.65512	-79.36264	The Extension Room	43.653313	-79.359725	Gym / Fitness Center

Cluster 0

	Neighborhood	Gym / Fitness Center	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
9	Davisville North	0.125	1	43.71276	-79.38851	900 Mount Pleasant - Residents Gym	43.711671	-79.391767	Gym / Fitness Center

Cluster 0

	Neighborhood	Gym / Fitness Center	Cluster Labels	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	Business reply mail Processing Centre, South C...	0.010000	2	43.64869	-79.38544	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center
6	Church and Wellesley	0.012658	2	43.66659	-79.38133	Verve Gym	43.666702	-79.376539	Gym / Fitness Center
7	Commerce Court, Victoria Hotel	0.010000	2	43.64840	-79.37914	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center
11	First Canadian Place, Underground city	0.010000	2	43.64828	-79.38146	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center
13	Garden District, Ryerson	0.010000	2	43.65739	-79.37804	Hard Candy Fitness	43.659556	-79.382440	Gym / Fitness Center
25	Richmond, Adelaide, King	0.010000	2	43.64970	-79.38258	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center
31	Stn A PO Boxes	0.010000	2	43.64869	-79.38544	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center
37	Toronto Dominion Centre, Design Exchange	0.010000	2	43.64710	-79.38153	Adelaide Club Toronto	43.649279	-79.381921	Gym / Fitness Center

Results - Map & Recommendation

Based on the criteria of gym/fitness venues clustering and competitors numbers, our recommendation for Jane Doe is to find a place for her new business venue within cluster 1, Dashville North, consisting of only one gym/fitness venue. Hence, Jane Doe would face a smaller number of competitors than in the other two clusters.

