

A black and white photograph of the Toronto skyline, featuring several prominent skyscrapers and a dense urban landscape. The image is split horizontally, with the top half showing a wider view of the city and the bottom half showing a more detailed, closer-up view of the buildings.

Wellness in Toronto

IBM Applied Data Science Capstone
By Joshua Kehler

20
Million

**Cases of
COVID19**

744
Thousand

**Deaths
from
COVID19**

7X
*Relative Risk of
Death from
COVID19*

**Pre existing
Heart
Disease**

Global Crisis of Wellness

As doctors and scientist work towards treatment and vaccines, the general public can strive towards solvency as well.

Health & Wellness

In observance of social distancing and CDC best practice, the local gym may be the best place to program the initiative.



Cardiovascular Training

Males who run for more than an hour per week at moderate intensity may realize a 42% reduction in risk of heart disease than those who do not



Strength Training

Resistance training has shown to decrease glycosylated hemoglobin (HbA1c) levels, a condition associated with diabetes and cardiovascular disease



Diet

Diets such as the Mediterranean diet, high in mixed nuts, fish, vegetables and low in red meats, sugars may reduce risk of heart disease by 31%

1. Tanasescu M, Leitzmann MF, Rimm EB, Willett WC, Stampfer MJ, Hu FB. Exercise type and intensity in relation to coronary heart disease in men. *Journal of the American Medical Association* 2002; 288:1994–2000.
2. Dunstan DW, Daly RM, Owen N, Jolley D, De Courten M, Shaw J, Zimmet P. High-intensity resistance training improves glycemic control in older patients with type 2 diabetes. *Diabetes Care*. 2002; 25: 1729–1736
3. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Arós F, Gómez-Gracia E, Ruiz-Gutiérrez V, Fiol M, Lapetra J, Lamuela-Raventós RM, Serra-Majem L, Pinto X, Basora J, Muñoz MA, Sorlí JV, Martínez JA, Fito M, Gea A, Hernán MA, Martínez-González MA; PREDIMED Study Investigators. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extravirgin olive oil or nuts. *N Engl J Med* 2018;378:e34.

BUSINESS PROBLEM

To identify optimal locations for future gym sites based upon current density of related venues in Toronto, Canada.

- Leverage an array of data science techniques; such as web scraping, Foursquare API, machine learning algorithms and visualization methods to develop a tool useful to real estate developers, sole proprietors, and fitness enthusiasts looking to add gyms or wellness centers to their community.



DATA

Scrape, Process & Explore

1

Wikipedia

List of postal codes from Canada to establish the scope of data to be refined to the city, neighborhood, and venue level.

.

2

Geocoder

Geographical coordinates for neighborhoods to plot the map and return venue data.

3

Foursquare API

Gym Venue data used to explore and cluster neighborhoods

(https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)



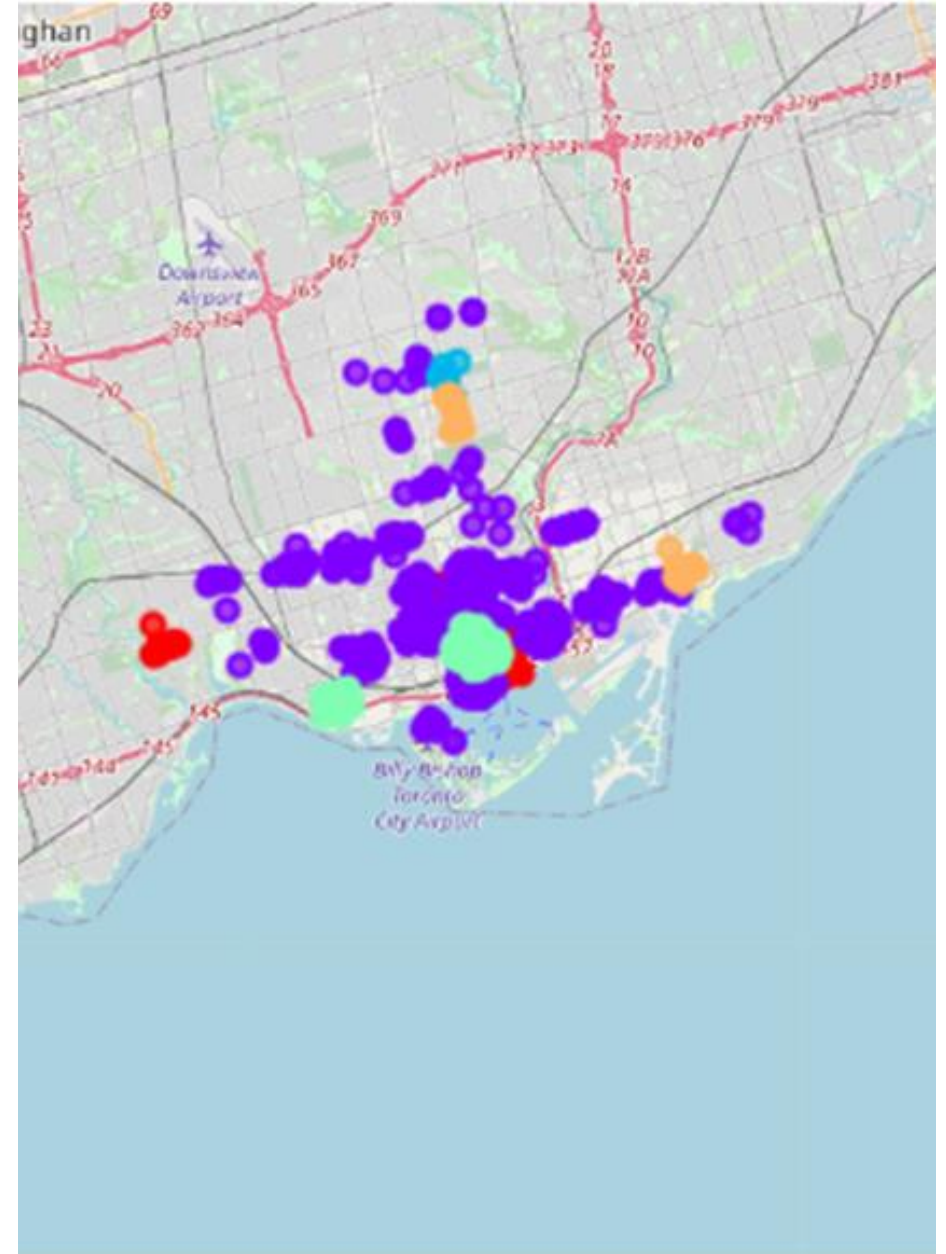
Methodology

- Webscrape a list of postal codes of Toronto Canada, within a table format on the Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M).
- Acquire geographical coordinates with Geocoder package
- Foursquare API to retrieve the top 100 venues within 500 meters.
- Group rows by neighborhood and by mean frequency of occurrence per venue, limit to “Gym”
- K-means clustering
- Generate a map of each cluster by folium

Results

Five Clusters

- Cluster 0: Low Gym Density
- Cluster 1: No Gym Density
- Cluster 2: High Gym Density
- Cluster 3: Low- Moderate Gym Density
- Cluster 4: Moderate Gym Density



Discussion, Conclusions & Recommendations

In summation, we present a methodology to determine an optimal location for future gym or wellness center development by identifying data sources, extracting and wrangling the required data and applying machine learning techniques.

Underserved Areas

- Cluster 0 and cluster 1 have the lowest density of gyms, thus these clusters may represent a population of peoples currently underserved in the domain of fitness & health.
- In contrast, cluster 2 has the highest density of gyms and may be saturated by over-supply.

Recommendation to our Stakeholders

- Cluster 0 and cluster 1 represent two areas of underserved populations which would benefit from the development of gyms and wellness centers.
- Thus, improving the community as a whole by providing greater access to health and wellness resources, a vital component in the initiative against chronic disease.



**THANK
YOU!**

Joshua Kehler