Wellness in Toronto: IBM Applied Data Science Capstone

By Joshua Kehler

Introduction:

On March 11th, 2020, the World Health Organization determined that COVID-19 had reached the level of pandemic at 118,000 cases and 4,291 deaths across 114 countries (1). As of August 12th, 2020 the number of Global cases surges to 20,439,000 with 744,941 deaths to date (2). With a reproduction number estimated at 3.28, the number of infected will continue to increase (3) with proportional loss of life. To anticipate need and safeguard those most vulnerable, risk factors and the elucidation of exacerbating comorbidities are of great interest. Research has shown that those with heart disease may be up to 7X more likely to die from COVID-19 (4). As doctors and scientist work towards treatment and vaccines, the general public can strive towards solvency as well.

Studies indicate that 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 (5). Resistance training has shown to decrease glycosylated hemoglobin (HbA1c) levels, a condition associated with diabetes and cardiovascular disease (6). 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% (7). Taken together, cardiovascular training, strength training, and nutrition guidance coalesce into a treatment plan readily deployed and exercised by the general public. In observance of social distancing and CDC best practice, the local gym may be the best place to program the initiative.

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.

Target Audience:

The tools and data herein will prove useful for real estate developers, sole proprietors, and fitness enthusiasts looking to add gyms and wellness centers to their community. Moreover, it is valuable to those looking to encourage greater health and wellness in the face of crisis by establishing more venues designed to facilitate healthier living. As gyms tighten hygiene and social distancing standards, these venues may very well establish a first line of defense against future pandemics.

Data:

- List of postal codes from Canada to establish the scope of data to be refined to the city, neighborhood, and venue level.
- Geographical coordinates for neighborhoods to plot the map and return venue data.
- Gym Venue data used to explore and cluster neighborhoods.

Sources of data:

Wikipedia page (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M) will be scraped to retrieve the data contained in the table of postal codes and transformed into a pandas dataframe for further analysis. Python requests and beautfulsoup packages will be employed towards this end. Retrieve geographical coordinates such as latitude and longitude for neighborhoods via Python geocoder package. Use the Foursquare API to extract venue data for all neighborhoods and focus our attention on the gym category.

References:

1. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. https://web.archive.org/web/20200502133342/https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefingon-covid-19–11-march-2020 (2020).
2. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Accessed 8/12/20.
3. Liu Y, Gayle AA, Wilder-Smith A, Rocklov J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med. 2020.
4. Deng, G., Yin, M., Chen, X. & Zeng, F. Clinical determinants for fatality of 44,672 patients with COVID-19. Crit. Care 24, (2020).
5. 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.
6. 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
7. Estruch R, Ros E, Salas-Salvado J, Covas MI, Corella D, Ar os F, G omez-Gracia E, Ruiz-Gutie´rrez V, Fiol M, Lapetra J, Lamuela-Raventos RM, Serra-Majem L, Pinto X, Basora J, Mu ~noz MA, Sorlı JV, Martınez JA, Fito M, Gea A, Hern an MA, Martınez-Gonzalez 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.