# Battle of the neighbourhoods (Week 1)

*Using Machine Learning to find locations to open up a Sports Nutrition shop*

*Robert Ashton-Jones (https://github.com/ RobEAJ123)*

**1.Introduction**

**1.1 Background**

For this Capstone project, I am creating a hypothetical scenario, where a Canadian entrepreneur is looking to open a new sports nutrition store. This project aims to clarify where the ideal location to open a store in Toronto would be. This will help the store to gain reputation quickly with the idea of expanding quickly. The starting location for a store like this is important, as it will define how profitable the store can be (and will ultimately affect how quickly the next store will be opened as a result). Finding the location to open such a restaurant is one of the most important decisions for this entrepreneur and I am designing this project to help him find the most suitable location.

**1.2 Business Problem**

The objective of this capstone project is to find the most suitable location for the entrepreneur to open a new sports nutrition store in Toronto. Using data science, including machine learning methods (e.g. clustering), this project aims to answer the question: “Where to start?”

**1.3 Target Audience**

The entrepreneur looking to begin his sports nutrition start-up.

**2.Data**

To examine this problem, I will use data from the sources below:

● List of neighbourhoods in Toronto.

● Latitude and Longitude of said neighbourhoods.

● Venue data related to the neighbourhoods in Toronto. This will help us to maximise the footfall for people who may visit the store.

I will use the data above to determine the frequency of different types of venue. This will help us to infer whether a neighbourhood is more interested in restaurants (in which case, a nutrition store will see less footfall) or other venues including: Gyms, parks, or health food stores.

**3. Extracting Data**

● Scraping of Toronto neighbourhoods (from Wikipedia)

● Latitude and Longitude data of the above neighbourhoods (from the Geocoder package).

● Venue data related to these neighbourhoods (from Foursquare API)

**Methodology & Approach:**

First, I obtained a list of neighbourhoods in Toronto, which was done by extracting them from the list found on this Wikipedia page:

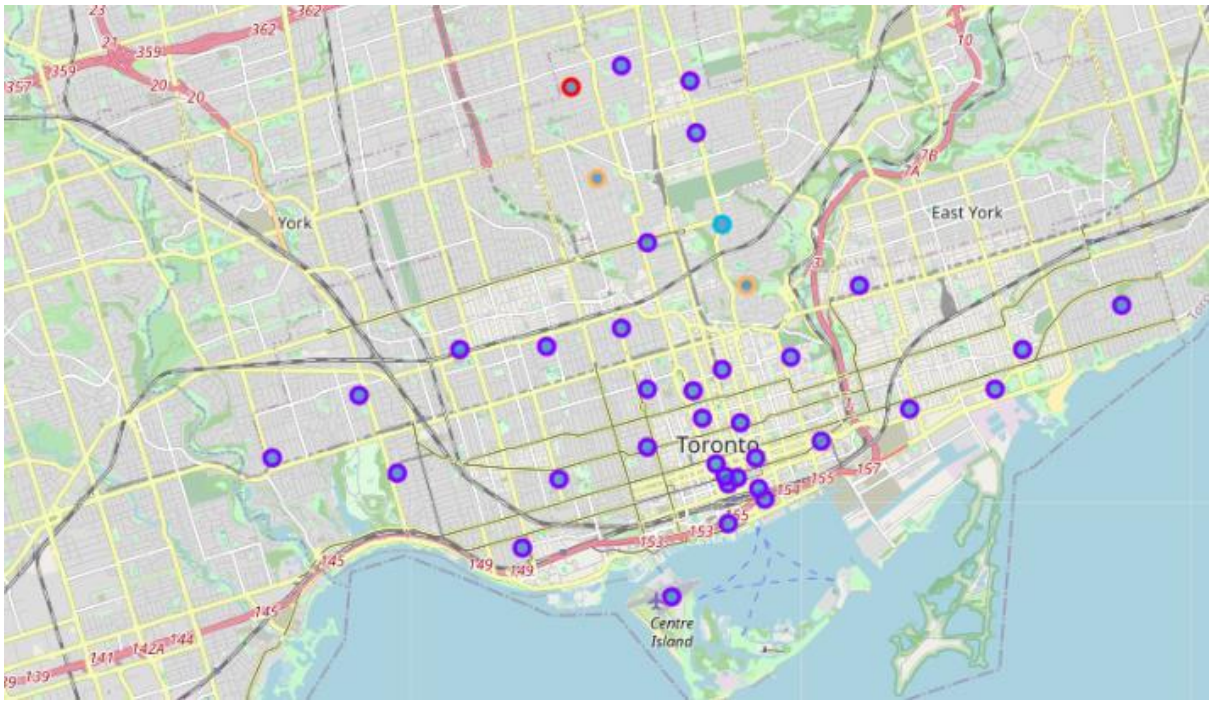
("https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M ")

I used pandas to web scrape the tabular data directly from a web page into a data frame. I know had a list of neighbourhoods' names and postal codes. I then needed to associate coordinates with each of the neighbourhoods and then use Foursquare to pull the list of venues that were nearby to each neighbourhood. I attempted to use the Geocoder package but ran into a couple of issues, so I used the CSV file provided by IBM. After gathering these coordinates from the CSV, I visualized the map of Toronto using Folium to verify whether the coordinates were correct.

I then used Foursquare's API to pull the top 100 venues within 500 meters radius of the neighbourhoods. From Foursquare, I was able to pull the names, categories and the latitude & longitudes of the venues. This allowed me to see the unique categories of the venues. I then analysed each neighbourhood by grouping the rows by neighbourhood before taking the mean of the frequency of occurrence of each venue category. This was done to prepare for clustering. Here, I looked for areas that were high in occurrence of gyms, sporting goods shops and health food stores. Finally, I performed the clustering method by using k-means clustering. K-means clustering algorithm to identify k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible. It is a simple, popular, unsupervised machine learning algorithm which is well suited for this project. I have clustered the neighbourhoods in Toronto into clusters based on their frequency of occurrence for health based shops. Based on the results (the concentration of clusters), I will be able to recommend the ideal location to open the restaurant.

**Results**:

*Clusters*



The results from k-means clustering show that we can categorize Toronto neighbourhoods into 3 clusters of interest, based on the presence of health and fitness-based shops in each neighbourhood:

● Cluster 1: Neighbourhoods with few health-based stores and a high concentration of restaurants/cafes (coloured purple)

● Cluster 5: Neighbourhoods with some health-based stores (coloured orange)

● Cluster 4: Neighbourhoods with a lot of health-based stores (coloured red)

**Recommendations:**

Most of the gyms were based around Cluster 4 which is around Adelaide, King, Richmond areas and lowest (close to zero) in Cluster 1 areas which are North Toronto West and Parkdale areas.

Looking at nearby venues, it seems Cluster 1 is the least favourable area. As there are a large number of restaurants and comparatively few health food stores/gyms in the area. Whereas Cluster 4 has a higher amount of the aforementioned. Therefore, this project recommends the entrepreneur to open a Sports Nutrition shop in the locations seen in Cluster 4, where people who are interested in going to the gym are more likely to see the store.

**Limitations and Suggestions for Future Research:**

Future research can take into consideration additional factors, outside the scope of this project due to the short time frame of this project. A few examples of additional datasets that could help further illustrate the most suitable areas could include: this project could include the frequency of obesity/heart problems in each neighbourhood; population density; competitor store locations or even the average income for each neighbourhood.

**Conclusion:**

In this project, we have gone through the process of identifying the business problem, specifying the data required, extracting and preparing the data, performing the machine learning by utilizing k-means clustering and providing recommendation to the stakeholder.

**References:**


- List of neighbourhoods in Toronto: https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M
- Foursquare Developer Documentation: https://developer.foursquare.com/docs