**Capstone Project - The Battle of Neighborhoods**

**Recommendation location for Yoga Studio in Toronto**

* 1. **Background**

Physical exercise is not only good for body but also it is necessary for mind and spirit. Regular physical activity supply many advantages from having a control on maintaing the weight to improving self-esteem and confidence builder. People are generally interested with gym or fitness saloon for meeting their needs.

Nowadays, people are interested in their minds an spirits. For this reasons, taking a yoga class once or twice a week might make a great addition to pysical activities. Yoga class makes helps to improve both physcial and spirit improvements.

**1.2 Problem**

Data that might contribute to determining finding the best locations in Toronto by examining the all of the stores and the categorization of the type of the stores that describe which place is suitable for Yoga Studio. This project aims that guessing the best locations were to open a Yoga Studio in Toronto based on the data.

**1.3 Interest**

Obviously, Foursquare API helps to explore neighborhoods in Toronto. Once the data is in a structured format then making exploration, segmentation, and clustering the neighborhoods in the city of Toronto.

**2. Data Requirements**

The postal code of Toronto is obtained from wikipedia which is offered by Coursera. The data is include postal code, borough and neigbourhood information of Canada. It is found but is not in a useable form, data wrangling and cleaning will have to be performed.

The cleaned data will then be used alongside Foursquare data, which is readily available. Foursquare location data will be available to explore or compare districts around Toronto, identifying areas where people go for gym, restaurants informations such as salad restaurants, vegan or vegetarion foods and some kind of healthy restaurants and cafes which are the areas where the opening a Yoga Studio.

For this project we need following data:

* The data is include postal code, borough and neigbourhood information of Canada from <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
* Read downloaded csv from <http://cocl.us/Geospatial_data> for getting geographical coordinates of each postal code.
* FourSquare API is used in order to explore neighborhoods in Toronto. FourSquare API to collect data about venue, venue latititude and venue category of the neighborhood. In this project top 100 venues that are in Toronto within a radius of 500 meters are getting.

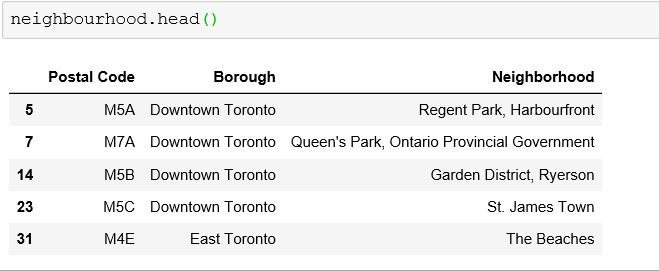
# 3. Methodology

# 3.1 Data Preparation

Firstly, make use of List\_of\_postal\_codes\_of\_Canada:\_M page from Wiki to scrap the table to create a data-frame. For this, pandas have used to transform the data in the table on the Wikipedia page into a dataframe. After a little data cleaning operation such as

* Remove rows where borough equals "Not assigned"
* If a cell has a borough but a Not assigned neighborhood, then the neighborhood will be the same as the borough.

The dataframe is containing name of the 39 Borough of Toronto as below:



# 3.1.2 Getting Coordinates of Major Districts : [Geopy Client](https://geopy.readthedocs.io/en/stable/)

# Next objective is to get the coordinates of Toronto using geocoder class of Geopy client.

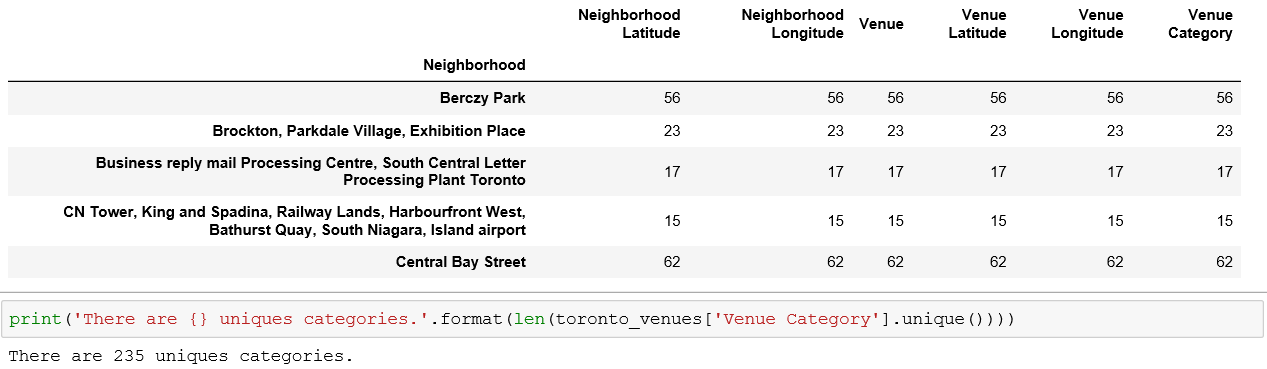
The geograpical coordinate of Toronto, Ontario are 43.6534817, -79.3839347.

#### Read downloaded csv from (<http://cocl.us/Geospatial_data>) for getting geographical coordinates of each postal code and mapping lattidues and longitudes to postal codes in Toronto data as follow:



# 3.2 Explore Neighborhoods in Toronto

#### FourSquare API is used in order to explore neighborhoods in Toronto. let's get the top 100 venues that are in Toronto within a radius of 500 meters. According to the data there are 235 uniques categories.. . .

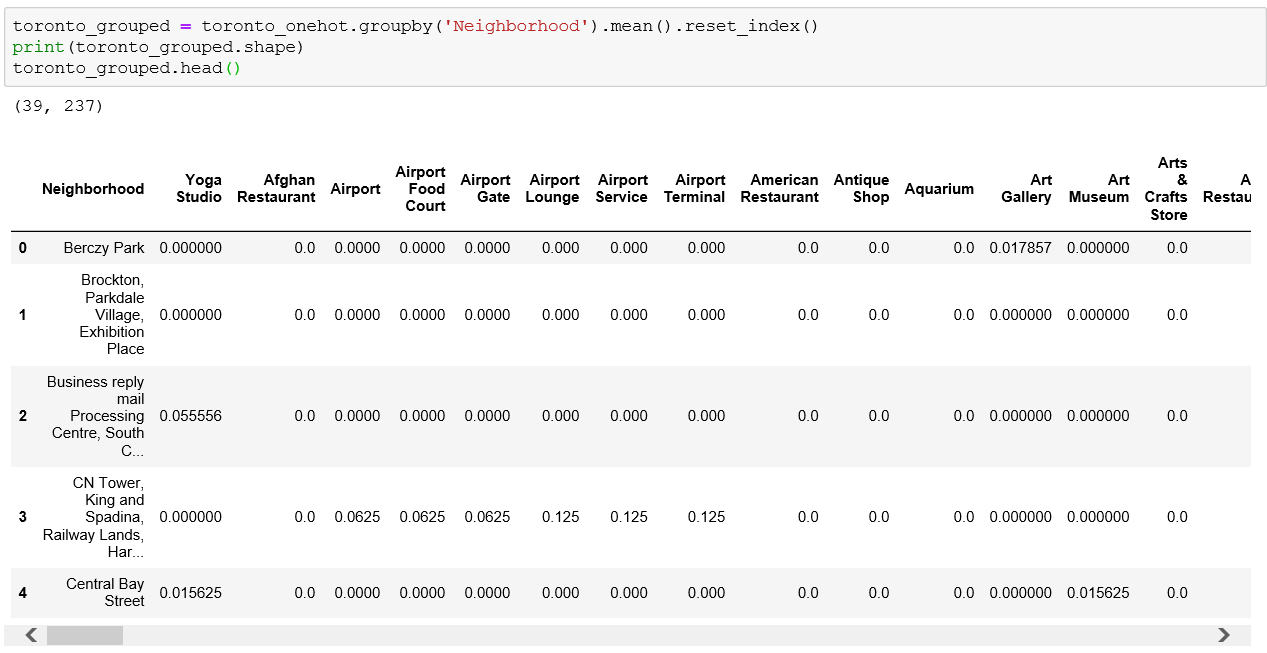


# 3.3 Analyze Each Neighborhood

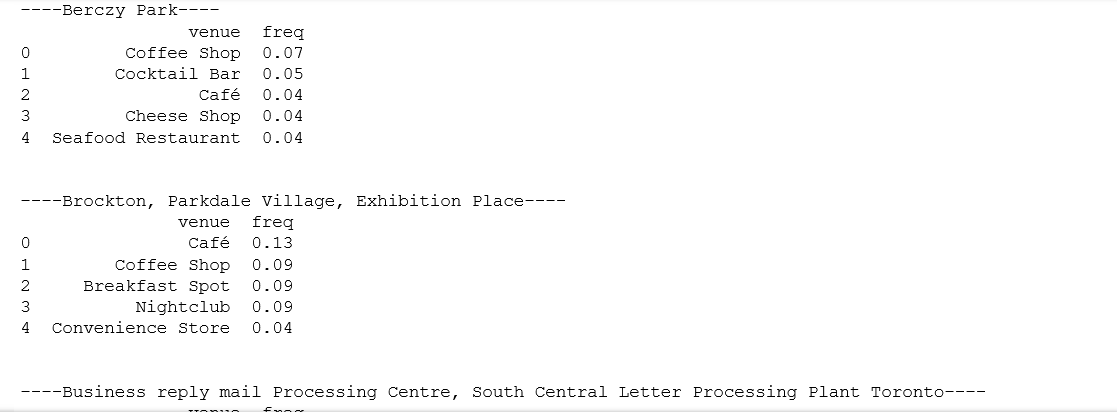
# Create a data-frame with [pandas one hot encoding](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.get_dummies.html) for the venue categories.

# 

* Use pandas groupby on neighborhood column and calculate the mean of the frequency of occurrence of each venue category.



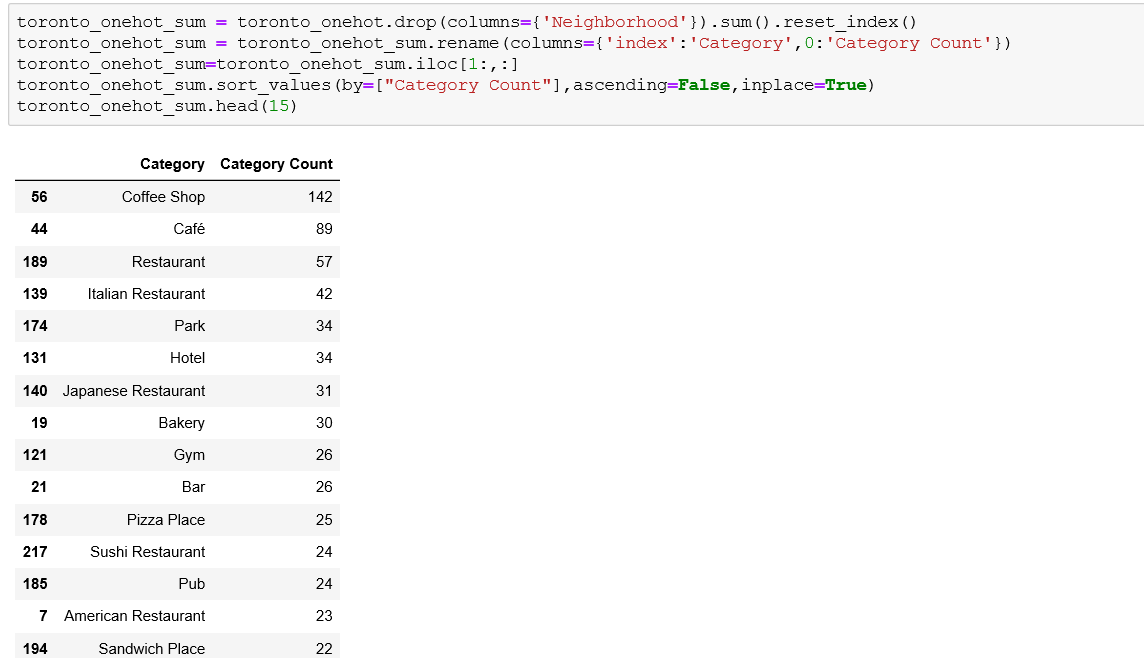
* Output each neighborhood along with the top 5 most common venues:



* Display the top 10 venues for each neighborhood, sort the venues in descending order. The example of the result is:

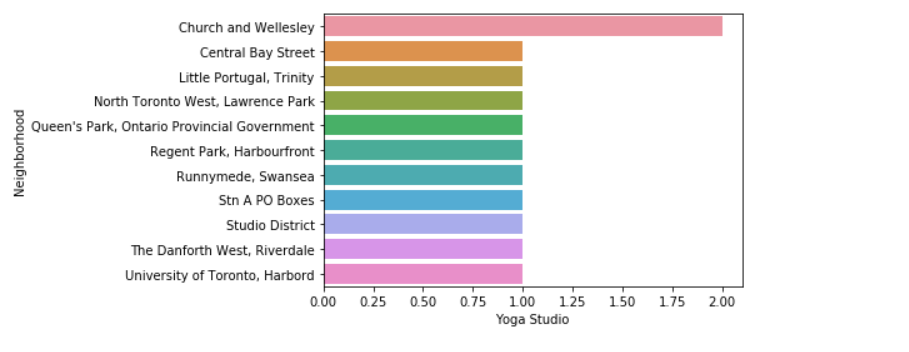


This is a very useful results table that can provide at a glance information for all of the neighborhood. That table can supply meaningful insights about the top categories of businesses in all the neighborhoods.

* The total distribution according to the categories are:

# 3.3.1 Analyze Yoga Studios

Let's look at their frequency of occurance for all the Yoga Studios in Toronto, isolating the categorical venues. The visualise frequency distribution datasets is like:

According to the results, the Yoga Studios are in which neighborhood, the studios will not open studios in cities like Church and Wellesley or Central Bay Street, etc. as a business necessary.

## **3.3.2. Definition of related categories**

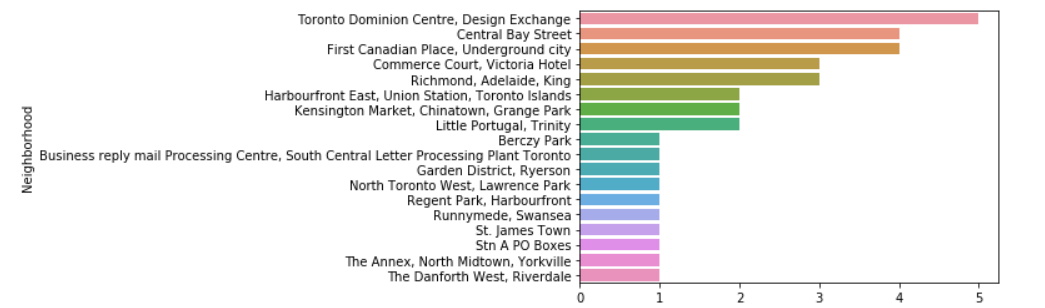
The top 2 venue types as specified by the client for each neighborhood are used for the plotting.

These venue types are determined by choosing due to the location for healthy places and gym places.

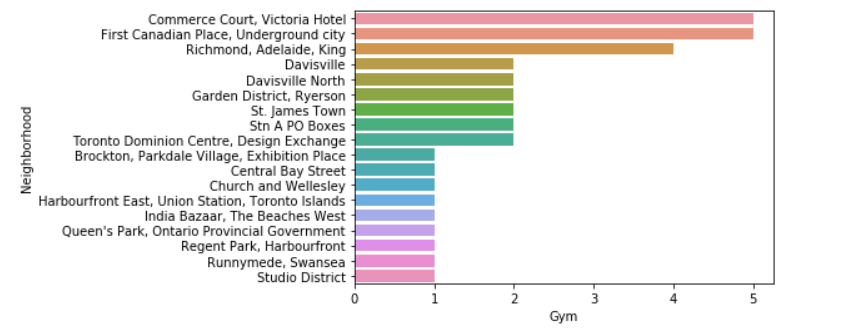
Whereas the healthy places are grouped as Salad Place, Vegetarian / Vegan Restaurant and Spa,

the sports places are grouped as Gyms and Gym / Fitness Center.

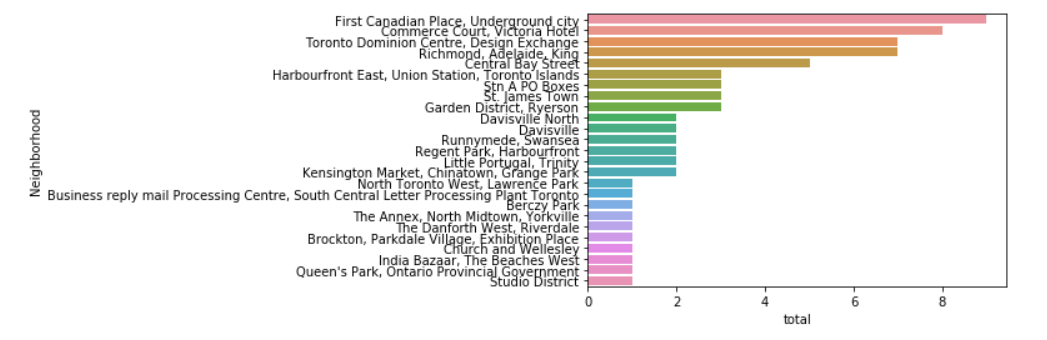
Firstly, healthy places are found that there is no yoga studios inside it. The neighbourhood of these healthy places are ranked by from highest to lowest according to the count of healthy places in each neighbourhood. The plot is shown below:



Secondly, the sports places are found that there is no yoga studios inside it. The neighbourhood of these sports places are ranked by from highest to lowest according to the count of sports places in each neighbourhood, again. The plot is shown below:



As a conclusion these two results of healthy and sports are combined together to find a best location for yoga studios in Toronto. The plot is shown below:



## **4. Inferences and Discussion**

Inferential analysis using the data, business allows the list to be focussed to just 3 neighbourhoods.

Identifying neighbourhoods that are lively with healthy places (Salad Place, Vegetarian / Vegan Restaurant and Spa) and the gym places (Gyms and Gym / Fitness Center) adding YogaStudios into the that places is a good option.

So the final 3 neighborhoods for new Yoga Studio locations are where these two criterias are mentioned above are meet:

* First Canadian Place, Underground City
* Commerce Court, Victoria Hotel
* Toronto Dominion Centre, Design Exchange