Expatriated executive to Toronto City looking for a new family home

This is the last part of the IBM Capstone Project.

Objectives of the final assignments were to define a business problem, look for data in websites and use Foursquare location data.

1. Discussion and Background of the Business Problem:

Problem Statement: **An executive manager has been expatriated to Toronto City and has to look for a new family home in one of its neighbourhoods.**

The objective is to evaluate the selection of a suitable neighbourhood in a city based on key criteria predefined.

In this case the target city is Toronto, the largest in Canada, and one of the largest in North America (behind only Mexico City, New York and Los Angeles). With a population just short of 3 million people, but The Greater Toronto Area (GTA) includes around 6.5 million people, stretching along the shore of Lake Ontario and including suburban communities further inland.

Toronto is also one of the most multicultural cities in the world with more than 140 languages and dialects are spoken in the city, and almost half the population Toronto were born outside Canada.

Although not the capital city of the country – that particular honour rests with Ottawa – Toronto is nonetheless the centre of many of Canada's industries, and therefore it offers many economic opportunities to new arrivals.

Consistently ranked as one of the most liveable cities in the world, Toronto enjoys a reputation as an exciting, diverse, clean, and safe city to set up home. It has 50 kilometres of waterfront with beaches, parks, marinas and waterfront trail.

Selection criteria:

The key criteria have to take into consideration his family needs and personal needs.

In order to simplify our quest:

we are going to consider as family needs the proximity of **good rated elementary schools** for his young children and the presence of plenty of **malls** for his wife.

With the intention of meeting his personal needs it would be the existence of **gyms** in the chosen neighbourhood

We are not going to consider the criminality factor due to Toronto is well known to have low crime rates for such a big city nor cost the cost of rental housing as it is included in the expatriation package benefits of the executive

Target Audience

What type of stakeholders would be interested in this project?

1. Investors who could benefit from the model to assess real estate investments in high qualified potential neighbourhoods

2. Commercial Real Estate Brokers (CBRE, Cushman & Wakefield, etc..) encouraged to offer commercial and brokerage services related to the new locations.

3. Big Corporations, with no presence in the city, but willing to expand their business and operate in the city. They would need to know the impact of key parameters to be taken into consideration in a relocation process for their expatriated candidates.

4. Public Administration who can grant immigration permits, get taxes from large groups and would like to consider the factors of attractiveness to right size their infrastructures

5. Toronto residents who could benefit from the assessment model to take data driven decisions

6. Individuals in expatriation situation who may have to face a similar situation

2. Data Preparation:

We'll install the necessary packages, if missing, as

beautifulsoup4 to scrape websites

geopy to geocode web services and to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources

folium to visualize data that's been manipulated in Python on an interactive leaflet map. It enables both the binding of data to a map for choropleth visualizations as well as passing rich vector/raster/HTML visualizations as markers on the map

and libraries as:

numpy # to handle data in a vectorized manner
pandas # for data analysis
json # to handle JSON files
Nominatim # to convert an address into latitude and longitude values
requests # to handle requests
json_normalize # to tranform JSON file into a pandas dataframe
matplotlib and associated # to plotting graphsmodules
sklearn # to use machine learning k-means at clustering stage
folium # to map rendering
geocoder # to get coordinates

Expatriated executive to Toronto City looking for a new family home

For this project we need the following data:

- Toronto data that contains the list of Boroughs and Neighborhoods
    - Data source :

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

    - Description : This data set contains the required information. And we will use this data set to explore various neighbourhoods of Toronto city.

- Venues in each neighbourhood of Toronto city.
    - Data source : Fourquare API

      https://developer.foursquare.com/docs/resources/categories

    - Description : By using this API we will get all the venues in each neighborhood. We can filter these venues to get only those that meet the predefined criteria.

- GeoSpace data
    - Data source : https://github.com/jasonicarter/toronto-geojson/
    - Description : By using this geo space data we will get the Toronto Borough boundaries.

To simplify this project, we will only use Toronto neighbourhoods where Borough contains Toronto.

In order to obtain ratings data for elementary schools we will use ontario.compareschoolrankings.org

2.1. Scrapping Toronto Neighbourhoods from Wikipedia

I first make use of Wikipedia on its page
https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M page

to scrap the table to create a data-frame.

For this, I've used requests and Beautifulsoup4library to create a data-frame containing the PostalCode, the Borough and the Neigbourhood

[29]:	PostalCode	Borough	Neighbourhood
0	M3A	North York	Parkwoods
1	M4A	North York	Victoria Village
2	M5A	Downtown Toronto	Kingsway Park South West,Mimico NW,The Queensw...
3	M6A	North York	Lawrence Heights,Lawrence Manor
4	M7A	Queen's Park	Queen's Park
5	M9A	Etobicoke	Islington Avenue
6	M1B	Scarborough	Rouge,Malvern
7	M3B	North York	Don Mills North
8	M4B	East York	Woodbine Gardens,Parkview Hill
9	M5B	Downtown Toronto	Ryerson,Garden District
10	M6B	North York	Glencairn

2.2. Getting coordinates of Boroughs: Geopy Client

The following step is to get the coordinates of the Boroughs, and the 103 Neighbourhoods using geocoder class of Geopy client along with their latitude and longitude.

2.3. Using Foursquare Location Data:

Foursquare data is very comprehensive and it powers location data for Apple, Uber etc.

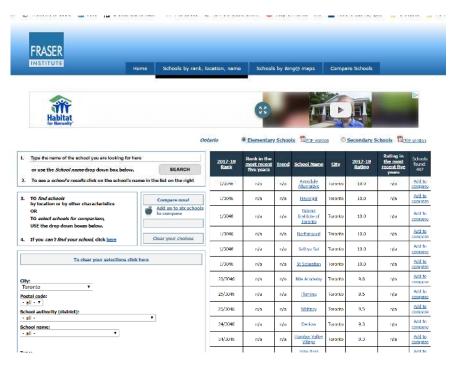# Expatriated executive to Toronto City looking for a new family home

For this business problem I have used, as a part of the assignment, the Foursquare API to retrieve information about the pre-defined criteria to find the suitable Neighborhood in Toronto

The call returns a JSON file that we need to turn that into a data-frame.

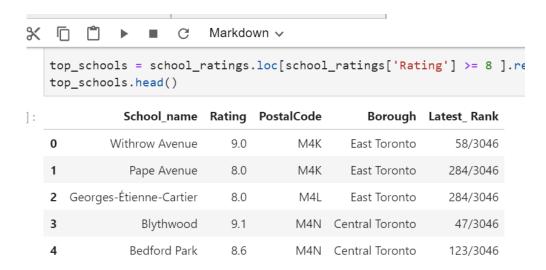We repeated the process twice, once for the Gym criterium and another for the Mall.

To get the ranking for the elementary schools we build a dataframe based on the scrapped information from

Expatriated executive to Toronto City looking for a new family home

Until we get a dataframe with the name of the school, the postal code, the Borough and the Latest Rank position

```
top_schools = school_ratings.loc[school_ratings['Rating'] >= 8 ].re
top_schools.head()
```

	School_name	Rating	PostalCode	Borough	Latest_ Rank
0	Withrow Avenue	9.0	M4K	East Toronto	58/3046
1	Pape Avenue	8.0	M4K	East Toronto	284/3046
2	Georges-Étienne-Cartier	8.0	M4L	East Toronto	284/3046
3	Blythwood	9.1	M4N	Central Toronto	47/3046
4	Bedford Park	8.6	M4N	Central Toronto	123/3046

The above school information will be merged with the Foursquare information into a new dataframe to start the analysis and the K-means clustering.