II. Data description:

Toronto is the heart of the Canada's economy. Being a major economy center of America, Toronto has been explored by many investors to find the best opportunities in this area. Therefore, it is not surprising to see that there's very good data sets of Toronto neighborhoods specifications on the web.

FourSquare API is used to extract the list of venues and their types. The first step in data extraction is to extract the Toronto neighborhoods list and population from the following Wikipedia page: https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods

We can extract the name of the neighborhoods from the given page and select the required data.

For each neighborhood, the name, population, population density and average income is extracted and ordered in a tabular format. At the next step the coordinates (latitude and longitude) are found by using geocoders library. Then the geospatial location is sent to FourSquare API. Using the "explore" endpoint, a list of surrounding venues in a pre-defined radius is returned by FourSqure. The occurrence of each venue type in neighborhoods would then be counted and one hot encoding is applied to turn each venue type into a column with their occurrence as the value.

The neighborhoods population, population density, average income and geospatial coordinates are placed in a table format dataframe as shown by figure 1.

[65]:

	Name	Population	Density (people/km2)	Average Income	Latitude	Longitude
0	Agincourt	44,577	3580	25,750	43.7854	-79.2785
1	Alderwood	11,656	2360	35,239	43.6017	-79.5452
2	Alexandra Park	4,355	13,609	19,687	43.6508	-79.4043
3	Allenby	2,513	4333	245,592	43.7114	-79.5534
4	Amesbury	17,318	4,934	27,546	43.7062	-79.4835
5	Armour Heights	4,384	1914	116,651	43.7439	-79.4309
6	Banbury	6,641	2442	92,319	43.7428	-79.37
7	Bathurst Manor	14,945	3187	34,169	43.7639	-79.4564
8	Bay Street Corridor	4,787	43,518	40,598	43.6628	-79.3863
9	Bayview Village	12,280	2,966	46,752	43.7692	-79.3767

Figure 1 - Neighborboods dataset

Each row represents a neighborhood dataset and each column is the properties of that neighborhood. The dataset has 5 features and 174 samples.