**Business Problem**

The objective of this analysis is to provide hotels, and possibly other organizations including local lifestyle magazines or newspapers in the city of Toronto with a listing of the top restaurants in specific areas with breakdowns by type of food, so they can provide this data to their guests to serve as recommendations. We can charge a subscription fee to provide this data on a quarterly basis in order to include the newer establishments and the latest trends. In order to provide a list relevant to their market, we will cluster Toronto based on Postal Codes, then based on the groupings, we will identify the top restaurants within the clusters.

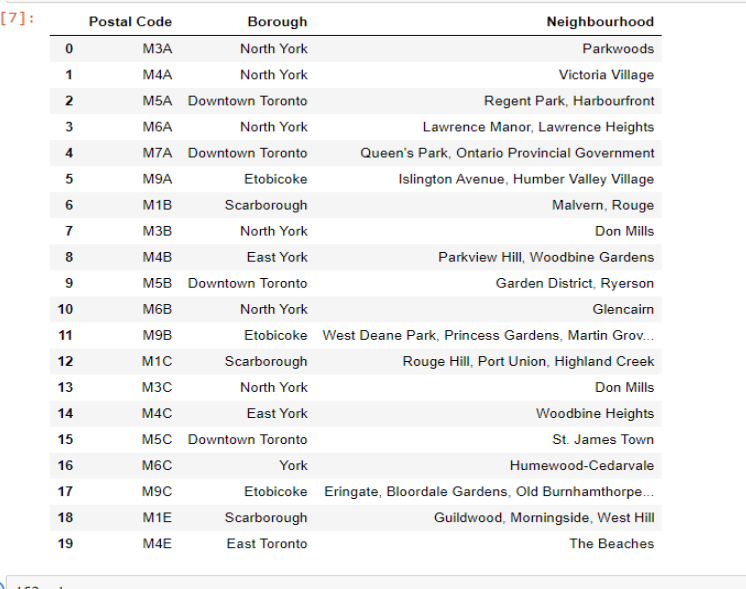
**Data**

In order to support this analysis, we need geographical data for Toronto including neighborhoods, post codes, boroughs and corresponding latitudes and longitudes. This data will be used to define geographical clusters. Additionally, we can tap into data from Four Square to analyze the top eateries that correspond to these clusters.

Wikipedia, generally speaking, is a great source for a wide variety of information and data. We can tap into Wikipedia to capture a listing of Toronto’s boroughs, post codes and neighborhoods. This will serve as the foundation for the geographical clusters. The following URL provides us with a listing of post codes, boroughs and neighborhoods in Toronto: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>

According to Wikipedia, this list includes postal codes that begin with ‘M’ which indicates that these post codes are within the city of Toronto.

In order for this data to be useful, we need to do a small amount of data cleansing. First, we must exclude any boroughs with a value of ‘Not Assigned’. Second, we need to collapse the rows resulting from multiple neighborhoods per post code and borough by creating a comma delineated list of neighborhoods aggregated by post code and borough. This step is necessary as we will capture latitude and longitudes per post code, not neighborhood. See the sample data below.



Next, we need to extend the data set with latitude and latitudes per Post Code. The following URL provides this mapping in order to attach the original data set with the corresponding latitude and longitudes: <https://cocl.us/Geospatial_data>.

Foursquare is a location technology platform that offers an API where we can extract location-venue data, with only a few constraints for the “community” (free) version. We can use an API to extract current venue data for Toronto. We can associate the venues with the clusters we create and provide summaries regarding the most visited venues by type of venue.