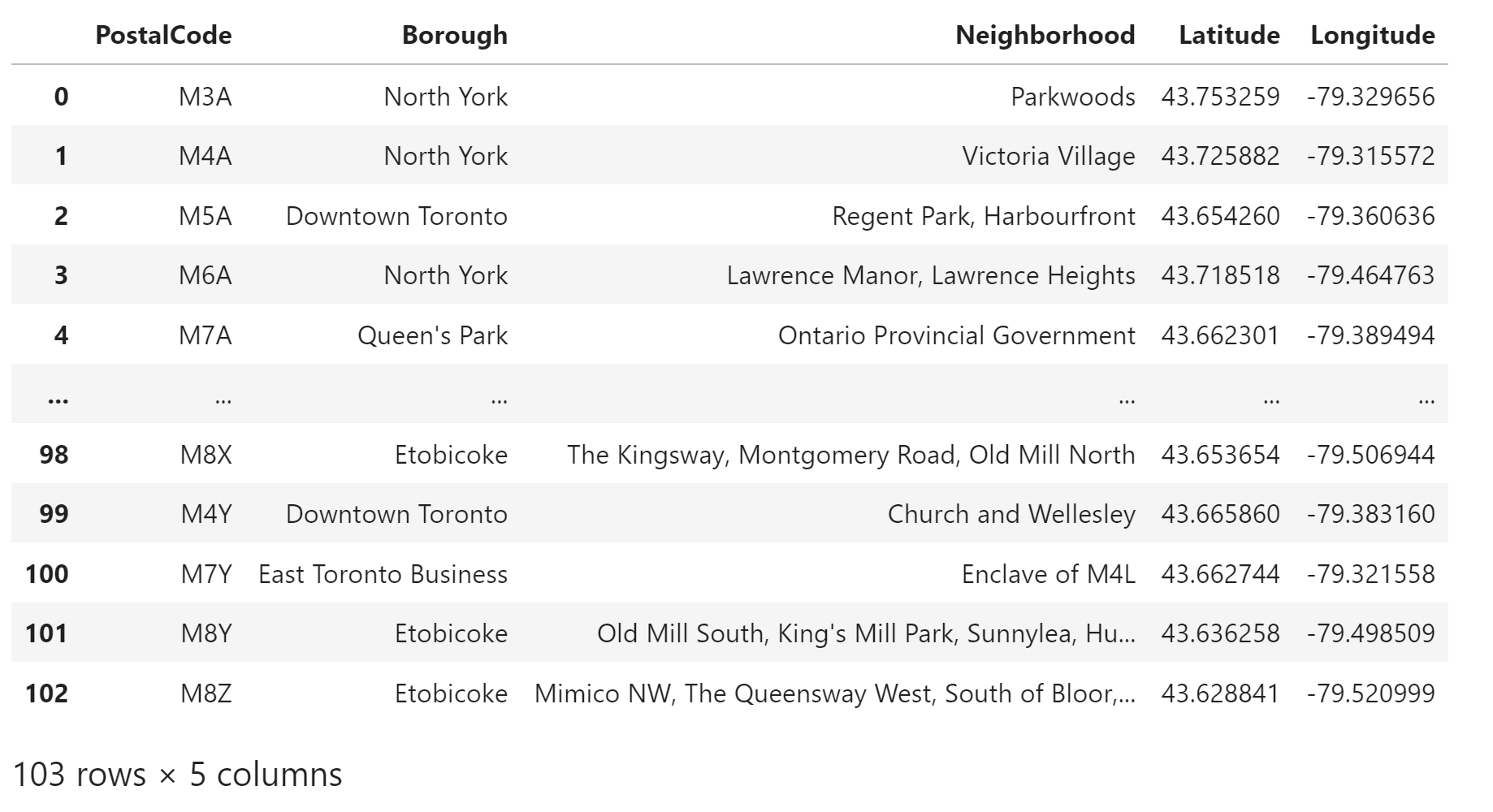
Data

To attaint the objective for this businessman, we will to have collect the below data:

1. PostalCode, Borough, Neighborhood and their respective latitude and longitude.
   * We find get the data from scrapping website: <https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M>
   * We get the latitude and longitude values of the Neighborhood by geopy library which are used as input into Foursquare
   * This is useful to categorize and breakdown the city of Toronto to smaller areas for the ease of our clustering and analysis
2. We deploy Foursquare to generate a list of restaurants close to the location of every neighborhood.
   * We selected the top 200 restaurants within 500m of the neighborhood with the using of API from Foursquare.
   * This would be helpful for us to find the most popular type of restaurant in the city of Toronto and in each neighborhood.
   * With the popular type of restaurants, we can cluster the neighborhood into various groups.
3. We gathered population data of Toronto, latest as of 2016, by Forward Sortation Area (FSA) from the Statistics Canada to show if there is a relationship between population density and the type of popular food in particular area
   * Link: <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/hlt-fst/pd-pl/Tables/CompFile.cfm?Lang=Eng&T=1201&OFT=FULLCSV>
4. To create a choropleth map of Toronto, we also need a geojson file
   * This helps to create a handy tool to visualize the population density in each area
   * Link: <https://github.com/BlizzWiz/IBM_Capstone_Project/raw/master/toronto_m.geojson>

Data processing and cleansing

1. We used Beautifulsoap 4 as a web scraping package to get the neighborhood data and combined it with the latitude and longitude values from geopy package.
2. We removed values that contain null or not assigned.



1. Then we downloaded the all restaurants near each neighborhood via Foursquare API.



1. We downloaded the population data from Statistics Canada and extracted the useful columns such as geographic code, province and population data. Then we extracted the data that are related to Ontario only.