Methodology

To complete the exercise requested, a search was conducted to collect the required data for the exercise. Some location data would be needed for both Toronto and New York City to use in Foursquare to investigate other restaurants in the cities to investigate potential for market saturation with the same type of restaurants. Furthermore, some demographics information, particularly population and income per neighborhood, would also be needed to get a measure of the market size and potential clientele of the new restaurant. The data collected is described in the following section.

Data Collection

The following key data were collected to facilitate the analysis required for the building of the model.

1. **New York Neighborhood Location Data:** The neighborhood location data for New York City came from the site https://cocl.us/new_york_dataset. The link to the site was provided as part of the Capstone Project class. The data from the site is a json format that had to be scraped with the JSON library in Python. After scraping with the library, a data frame of New Yorks neighborhood location data was created from the data.
2. **Toronto Neighborhood Location Data:** The neighborhood location data was found on the Wikipedia site https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M. The site was provided as part of the Capstone Project course. The data table from the site was scraped with BeautifulSoup library in Python and loaded into a data frame. The data from this site only had postal code information for each Toronto neighborhood and the latitude and longitude data had to sourced differently. The latitude and longitude information was sourced from the site https://cocl.us/Geospatial_data, and read directly as csv format. The two files were combined to create a data frame of neighborhoods location data.
3. **New York Neighborhood Population Data:** Population data for New York neighborhoods was sourced from the site https://data.cityofnewyork.us/api/views/swpk-hqdp/rows.csv?accessType=DOWNLOAD, owned by the City of New York. The link provided neighborhood population data as downloadable csv format. The data was directly downloaded into a data frame.
4. **New York Neighborhood Income Data:** Income data for New York neighborhoods was sourced from the site https://ny.curbed.com/2017/8/4/16099252/new-york-neighborhood-affordability, a NYC real estate report provided by Curbed New York. Scraping the site with BeautifulSoup was impossible. Instead, the table was downloaded into a csv and loaded into a data frame. The income and population data were combined into a single demographic data frame.
5. **Toronto Neighborhood Population and Income Data:** Neighborhood income and population data for Toronto were sourced from the Wikipedia site https://en.wikipedia.org/wiki/Demographics_of_Toronto_neighbourhoods. Information on the site included multiple tables, which were scraped with BeautifulSoup in Python and put in a data frame.