1. **DATA**

**2.1. Data Sources**

The following data were collected for this project:

* The latest available census of Metro Manila which can be collected from the PSA website or from Wikipedia is of 2015 and was scraped using the python *BeautifulSoup* library. Collected with the census were the list of different barangays and cities in Metro Manila.
* The geocoder library was used to acquire the coordinates of the barangays and cities.
* The manila.json file was taken from github account of altcoder.
* The list of the Yummy Tokyo branches was prepared in an excel sheet. The locations of each branch were collected from google maps.
* The names and locations of other ramen restaurants were collected using FourSquare API, though the list that was generated were of the popular restaurants only.
* The venues near the five Yummy Tokyo branches were collected using the FourSquare API.
* The venues near the other stores were collected using the FourSquare API.
* The venues for each barangay were collected using the FourSquare API.
  1. **Data Cleaning**
* The census collected had missing population data, so these barangays were removed from the data set. There were originally 519 barangays from the original list but were cut down to 424 after removing the barangays with missing population data.
* There were originally 17 cities/municipalities listed, but due to unavailable population data, it was reduced to 14. Included in the cities/municipalities that were dropped are Caloocan, Pasay, and Pateros.
* Zip codes from the scraped data were removed from the data set since the coordinates of each place can be collected even with using only the barangay and city names.
* The population data were of type string and there were commas as thousand separators. The commas were removed, and the values were converted into type int.
* Some of the barangays of different cities have the same names, so using the *‘groupby’* function by ‘Barangay’ names merges these redundant names. In order to address this problem, after the coordinates of each locations were acquired, the barangay and city names were joined into one column using the *‘aggregate’* and ‘join’ functions.