IBM Data Science Capstone Project- The Battle of the Neighborhoods

2.0 Data Required

In order to address our previously stated business problem, we would require various kinds of data from numerous sources which have been listed and briefly described below:

- Crime Data from the Chicago City Data Portal which maintains an extensive record of all crimes that have been reported in Chicago over the last two decades including- date, type of crime, exact location, district, police ward etc.
    - Data Source: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2
    - Description: As the original dataset is extremely large and cannot be handled properly for analysis, only the data from the year 2016 has been used as representative. This data has been used to build a Choropleth map of the city of Chicago showing the intensity of crime in various neighborhoods.

- Chicago city data which includes the names and boundaries of all neighborhoods was obtained from the Chicago Data Portal in 'csv' format as well as Geospatial data was obtained in 'json' format.
    - Data Source: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Neighborhoods/bbvz-uum9
    - Description: Using the above data in conjunction with the *GeoPy* geocoding libraries for Python, latitudes and longitudes for each neighborhood was obtained and organized into a *Pandas* dataframe.

- Data about the venues and amenities within a 1km radius of each neighborhood and the locations of established Japanese restaurants in Chicago was obtained using the *Foursquare API*.
    - Data Source: Foursquare credentials can be obtained by creating a developer account on https://developer.foursquare.com/
    - Description: Foursquare's API was used to obtain the names of venues as well as related information such as location, customer ratings, tips etc.