

DATA FOR THE BATTLE OF NEIGHBORHOODS

To solve the problem, we need the following data

- New York city data containing the neighborhoods and boroughs.
- Latitude and longitude coordinates of those neighborhoods. This is required to plot the map and get the venue data.
- Venue data, particularly data related to the restaurants. We are going to use this data to perform further analysis of the neighborhoods.

Data Source and methods to extract them

New York City data containing the neighborhoods and boroughs will be obtained from the open data source: https://cocl.us/new_york_dataset. After it, we will get the geographical coordinates of the neighborhoods (latitude and longitude) using Python Geocoder package. Finally, we will use Foursquare API to get the venue data for the neighborhoods defined at the previous step. Foursquare has one of the largest databases of 105+ million places and over 125,000 developers use this application. Foursquare API provides many categories of the venue data; we are particularly interested in the restaurant data to solve the business problem defined above. This project will require using of many data science skills, from web scrapping (open-source dataset), working with API (Foursquare), data cleaning, data wrangling, to map visualization (Folium). In the next Methodology section, we will discuss and describe any exploratory data analysis that we did, any inferential statistical testing that we performed, and what machine learning techniques were used.