Introduction on Battle of Neighborhoods - New York City

Problem Statement

This project is a requirement in completing Coursera's IBM Applied Data Science Capstone. It is the final course in a series for IBM Data Science Professional Certificate program.

Location data analysis will be conducted on New York City neighborhood in the attempt to figure out which is the best location between 5 of its boroughs and its surrounding neighborhood is best for a Food Truck business start-up.

Specific attributes of its neighborhood such as population, housing and crime trends will be looked into including performing segmentation and clustering of the neighborhood's social networking data made available by Foursquare.

Main goal is to find the best neighborhood area that will be covered by the new Food Truck business.

Data Utilized on the Research

The data utilized in the research will be sourced from these locations:

- NYC Open Data
- Foursquare

Note however that due to the size of some of the historical datasets used for the research and the space limitation of the Cognitive Class lab environment, there was the need to extract the data directly into a personal computer and load into a relational database and conduct the data wrangling and data transformation from it.

Datasets sourced from NYC Open Data

- NYPD Complaint Data Historic
  Source : https://data.cityofnewyork.us/Public-Safety/NYPD-Complaint-Data-Historic/qgea-i56i
- NYC Borough Boundaries
  Source: https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm
- NYC Borough Total Population 1900 to 2010
  Source: https://www1.nyc.gov/site/planning/data-maps/nyc-population/historical-population.page
- NYC Borough Total Housing Units 1940 to 2010
  Source: https://www1.nyc.gov/site/planning/data-maps/nyc-population/historical-population.page