

# Capstone Project: Battle of the Neighborhoods

IBM Data Science

June 9, 2021

## 1 Data

### 1.1 Data Sources

The rating, likes, tips, venue name, and venue type data for individual restaurants will be taken from **Foursquare**. This will be combined with location data for NYC [1] containing latitude and longitude values for each venue.

### 1.2 Data Cleaning

The **NYC** location data will be compiled into a data frame containing the borough, neighborhood, latitude, and longitude. This data will be merged with borough, neighborhood, rating, likes, tips, and restaurant name data from **Foursquare**.

The **NYC** location data [1] will only be necessary to generate a map populated with points that display the neighborhood, borough, and average rating per neighborhood when clicked.

From the **Foursquare** data a restaurant count per neighborhood will be determined. Additionally, an average rating per neighborhood will be calculated. All this will be used to determine the restaurant density per neighborhood in order to apply a statistical significance to the average ratings per neighborhood.

## References

[1] [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset)