

# Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 1):

## Introduction/Business Problem-

Given the COVID-19 outbreak and the subsequent quarantines and shutdowns, individuals and businesses have begun to have to abide with social distancing and quarantining rules in order to flatten the curve and reduce the number of infections at any given time. In order to stay informed, individuals, businesses, or even the government may wish to use data science in order to determine the most densely populated zones in order to better allocate resources towards the area for COVID treatment or support for at risk individuals who would be more likely to contract COVID if they visited or lived in those zones.

Data science may be used to identify and visualize the major areas of a city, in this project I will be using New York, and determining the density of residential and commercial venues that have the highest probability of either inhabitants or visitors. Determining where these venues are most likely to be clumped together may give insight into which zones require more support or should be avoided by at risk individuals.

## Data-

There are three sets of data that will be used to analyze the problem: the previously given New York City data that contains lists of the Boroughs and Neighborhoods with their respective latitudes and longitudes, the GeoSpace data to draw the borders on the maps, and the Foursquare venues to determine the density of venues in an area.

The Borough and Neighborhood data will come from a source we previously used in this course: [https://cocl.us/new\\_york\\_dataset](https://cocl.us/new_york_dataset). It will consist of a list of Boroughs, Neighborhoods, and their coordinates in order to generate a list of zones.

The Foursquare data will come from accessing the Foursquare API to generate the number of venues in each zone.

The GeoSpace data will come from <https://data.cityofnewyork.us/City-Government/Borough-Boundaries/tqmj-j8zm> and it will provide the geo spatial data to generate a map visualization of the data analysis.