Capstone Project - The Battle of Neighbourhoods (Part-1)

Introduction:

New York City was home to nearly 8.4 million people in 2018, accounting for over 40% of the population of New York State. Due to New York growing in population each year with a finite amount of space, the state of New York is offering incentives to small fitness owners to open gyms/fitness areas in the New York area to get the population to remain fit. The challenge is to ensure the need for gyms and thus the reason all gym franchises are looking to do an in-depth study of the New York area and determine the best possible solution/area to open a gym.

Problem:

A new gym franchise is looking to open a gym in one of New York's neighbourhoods. The franchise does not understand the area and the availability of gyms in each neighbourhood and requires an investigation to take place in order to determine the best place for the franchise to open a new gym based on the decision metrics below.

Decision Metrics:

The following decision metrics are requested in order to make an informed decision:

1. Density of people for each Borough

2. Number of Neighbourhoods in each Borough

3. Number of gyms in each Borough

4. Gyms in the Neighbourhood with the best density metric per gym

5. Cluster Gyms in Neighbourhood with the best density metric per gym

Data Requirements and sources:

For the investigation the following data sources will be used:

1. Wikipedia to obtain density of each Borough in New York city. a. Source: https://en.wikipedia.org/wiki/New_York_City b. Description: New York Boroughs and the density of each Borough in the New York area.

2. New York City data that contains list Boroughs, Neighbourhoods along with their latitude and longitude. a. Source: https://cocl.us/new_york_dataset b. This contains the data as mentioned above and will be used for investigating the Borough and Neighbourhoods using Foursquare API.

3. Gyms in each neighbourhood of New York city. a. Source: Foursquare API b. The API will return all known gyms in each Borough and Neighbourhood.