# Coursera Capstone Project:

## The Battle of the Neighbourhoods

**Author:**

Michiel Schinkel

**Introduction**

In today’s society, the ability to use data to make decisions is becoming increasingly important. Data driven approaches to healthcare, marketing and business have shown great results. Data science provides us with the opportunity to learn from past events with an unprecedented accuracy and detect patterns that could otherwise go unnoticed.

Suppose we would like to open up a new sandwich place in New York City (NYC). We could start looking for a place to open our shop at random and hope for the best. We could also make a more informed decision about there to open up our new place by looking at data of other venues in NYC and see where other sandwich places are located.

This project aims to find clusters of neighbourhoods based on geolocation and compare their characteristics to find the place within NYC that may offer the greatest chances of success for opening up a new sandwich place.

**Data**

Several data sources will be used for this project. First of all, a list of neighbourhoods in NYC is downloaded from the following location:

* <https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json>

Secondly, the location of data from NYC is obtained through the geolocator function from the geopy library.

Finally, we sent a GET request through the Foursquare API to obtain venues surrounding all the different NYC neighbourhoods within a radius of 500 meters. These venues will be used as characteristics of the clusters and to decide upon the best location to start a new sandwich place.