# **Capstone Project – The Battle of Neighbourhoods in London**

## **1. Introduction**

**1.1 Problem Statement**  
This research aims to analysis the popularity and geolocation information of restaurant in London, United Kingdom to facilitate the decision of new store location and category.

### **1.2 Background**

In 2020, tremendous restaurants in London closed due to COVID-19 since it was discovered in China early January. It is believe the economy will be recovered in 2021 alongside with mass production of vaccine. Thus, it is a perfect opportunity to conduct a market research on existing restaurant in London although it is not the best time to start a new one.

So, how could we leverage Foursquare location data and machine learning to help us make decision and find appropriate neighbourhoods? This is the problem I would like to address in this capstone project taking Tokyo as an example. In this project, I am going to use Foursquare location data and clustering methods to group the districts to different group by their restaurant venues information.

## **2. Data Requirement**

For this project we need following data:

* Tokyo data that contains list districts (Wards) along with their latitude and longitude.

Datasource: <https://en.wikipedia.org/wiki/London_boroughs#Former_authoritie>

Description: We will Scrap London Borough Table from Wikipedia and get the coordinates of these Borough using geocoder class of Geopy client.

* Restaurants in each neighbourhood of London:

Data source: Foursquare APIs

Description: By using this API we will get all the venues in each neighbourhood. We can filter these venues to get only restaurants.

## **3. Methodology**

### **3.1 Data Preparation**

3.1.1 Scraping London Borough information from Wikipedia

3.1.2 Retrieving Coordinates of London Borough

3.2 Exploratory Data Analysis

3.2.1 Applying Foursquare Location Data

3.3 Apply k-Mean clustering to Data

4. Discussion

5. Conclusion