

## **2. Data**

Datasets used in this analysis are from three sources: Foursquare location data, Wikipedia and The London Datastore. The data extracted from these sources will become the inputs of the machine learning model to classify the ratings of grocery stores.

### **2.1 Foursquare API**

Foursquare API is the main source of data in this report. It is a popular location data vendor that provides venues locations, user ratings and tips, etc. It is utilised to find the coordinates of London's Tesco stores, explore the neighbourhood of these stores and obtain the ratings of these stores.

### **2.2 Wikipedia**

The "List of areas of London" ([https://en.wikipedia.org/wiki/List\\_of\\_areas\\_of\\_London](https://en.wikipedia.org/wiki/List_of_areas_of_London)) is used. It provides the areas and boroughs of London and the corresponding post codes, which will be used for the locations of the grocery stores in the model.

### **2.3 The London Datastore**

The London Datastore is a free and open data-sharing portal where anyone can access data relating to the capital. The median house prices of each areas in London is downloaded from this website (<https://data.london.gov.uk/dataset/average-house-prices>).