

# Battle of Neighbourhoods: Data

---

## Data source

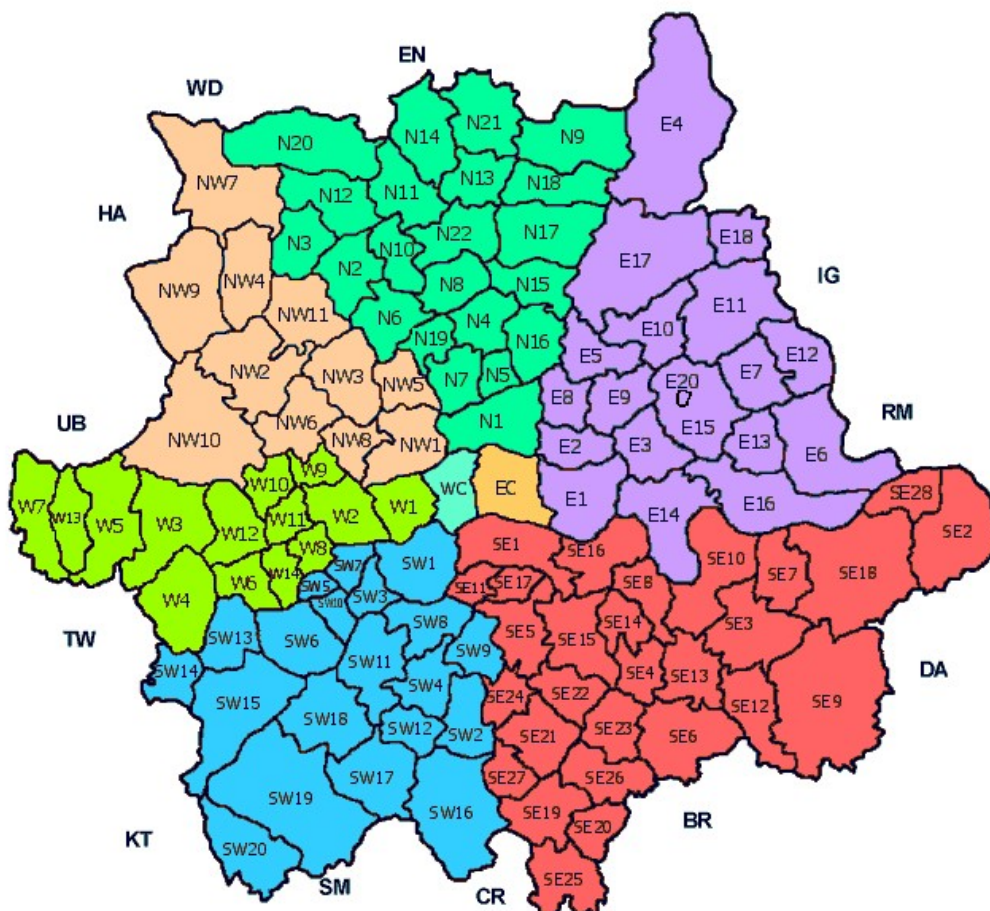
For this study we will use the “outcode area” (OA) postcodes as the area unit (also known as postal districts). Information regarding London postcodes, including its coordinates, can be extracted from multiple sources. We have selected a combination of two, for simplicity. First one will be obtained scrapping the following webpage, that offers a list of the London postcodes:

<https://www.milesfaster.co.uk/london-postcodes-list.htm>

This dataset contains a simple list, that we will merge with another dataset that provides the coordinates of all the OA postcodes:

<https://www.freemaptools.com/download/outcode-postcodes/postcode-outcodes.csv>

A distribution of those OA postcodes can be seen on the following map:



(source: [https://www.doogal.co.uk/london\\_postcodes.php](https://www.doogal.co.uk/london_postcodes.php))

Greater granularity levels can be obtained using smaller geographical divisions, as the “full” postcodes.

The information about the shopping establishments will be extracted from the Foursquare database.

## Data description

The starting point will be the geographical information of the areas that we will use as the study unit. The minimum information needed will be an identifier (OA postcode, each of them associated to a Post district), the name of the area and its geographical coordinates.

The rest of the information will be extracted from the Foursquare database. This database will provide different “venues” (places) in an area centred in specific coordinates and inside an area defined by a specific radius (distance to the centre). Each of these “venues” will have a category, that will be the main parameter used for clustering.

An example of the data records of the geographical areas can be the following one (the result from merging the information from both already mentioned data sources:

postcode	location	latitude	longitude
E1	Whitechapel, Stepney, Mile End	51.51766	-0.05841
E1W	Wapping	51.50775	-0.05739

Foursquare will provide different information about the venues, in our case we will take only four fields, here we can find a generic example based in Toronto:

venue	venue latitude	venue longitude	venue category
Roselle Desserts	43.653447	-79.362017	Bakery
Tandem Coffee	43.653559	-79.361809	Coffee Shop