Data

To identify various post codes and cluster them we will be using the following data:

- List of Post codes and their coverage - This will define the area/city in scope for this project which is Bristol, UK. [Wikipedia](#) will be used to source this data. This page has a table with Post code districts, Post town, their coverage and the Local authority area. Python's Beautifulsoup package will be used to scrape data from this page.
- Latitude and Longitude co-ordinates of these post codes which will be used to plot the map and also to get the popular venues list. This can either be found using Python's Geocoder package or pulled from a compiled list readily available on various sites. 2-Outcode file from the [link](#) was used to get the co-ordinate details for Bristol.
- Venue data will be used to perform clustering of the post code areas - Foursquare is a location technology platform which has one of the largest data of over 105 million+ locations and is used by over 150k developers. Explore function of this API will be used to get the most common venue categories in each post code, and then use this feature to group the various post codes of Bristol into clusters. K-means clustering will be used to cluster the various post codes based on similarity to gain insights.