**Introduction**

We have seen previously that different neighbourhoods in New York and Toronto can be clustered together into similar groups. However geographical information were not included, and we have seen cases of clusters which were formed of disjoint neighbourhoods, and separated by other clusters geographically.

In this project, I will look to group neighbourhoods in Sutton, UK into clusters based on both venues in each area as well as their geographical locations. This will be useful to both local business and councils, to determine local town centres, as well as possible catchment areas that will be of interest for each area. For example, when trying to open a new store, business may consider an area which have wide geographical reach, yet currently lacks competitors in the industry it is in.

**Data Used**

1. Ordnance Survey data, which is public information, that contains the latitude and longitude for postal areas. In our project, we will be looking at all the postal sectors in Sutton (all except for last 2 letters of postcode), and the latitude and longitude for each sector. This will be useful to our project from the geographic location point of view.

Sample dataset

|  |  |  |
| --- | --- | --- |
| Postal Sector | Latitude | Longitude |
| SM1 1 | 51.3646 | -0.19466 |

1. Foursquare data, by making calls to its API, we can obtain the venues of interest in each postal sector, such as restaurants, supermarket etc. This will be useful to our project from the business point of view, to help cluster neighbourhoods together based on similar businesses

Sample dataset

|  |  |  |  |
| --- | --- | --- | --- |
| Venue | Venue Latitude | Venue Longitude | Venue Category |
| Holiday Inn - Sutton | 51.361713 | -0.196645 | Hotel |