Capstone Project – Identifying Restaurant Location in London City

1. Introduction: Description of the problem

In this project, I will try to identify specific area and location, where a restaurant might be of interest for stakeholders in **London city**. This is partly based on the lessons that were part of Data Science course and the Berlin City example provided in the course.

The output of the project is to provide probable addresses where a stakeholder can begin their search.

2. Data Sources:

The neighbourhood and borough locations have been scrapped from below Wikipedia link using Beautiful Soup. https://en.wikipedia.org/wiki/List_of_areas_of_London

The borough boundaries have been taken from below link. Though borough boundaries are not utilised for analysis, but have been used for general visualisation purpose. https://github.com/martinjc/UK-GeoJSON/blob/master/json/administrative/eng/lad.json

London has about 33 boroughs, including City of London. There are about 533 neighbourhoods. All the data has been cleaned and filtered for missing & abnormal values, in order to consider for the analysis. **Figure 1** shows the neighbourhoods (in red), borough boundaries (in green) and the 6000m search radius (in blue) marking.

Ideally, I would have wanted to survey all the neighbourhoods, but considering limitations on the sandbox account of Foursquare, I have limited the study to about 6000 metres radius of London (51.5073219, -0.1276474).

Below are the Python packages used in this project to extract, visualise and analyse the data;

- BeautifulSoup – used for web scrapping the London neighbourhood and borough data.
- Pandas – used for analysing the data.
- Folium – used for map building.
- Geopy – used for picking up the coordinates, calculating distances and extracting addresses.
- Utm – used for converting latitude and longitude to UTM system, in order to serve as an input to machine learning algorithms.
- Sklearn – It is a machine learning library. In this project it is used to implement clustering algorithm.
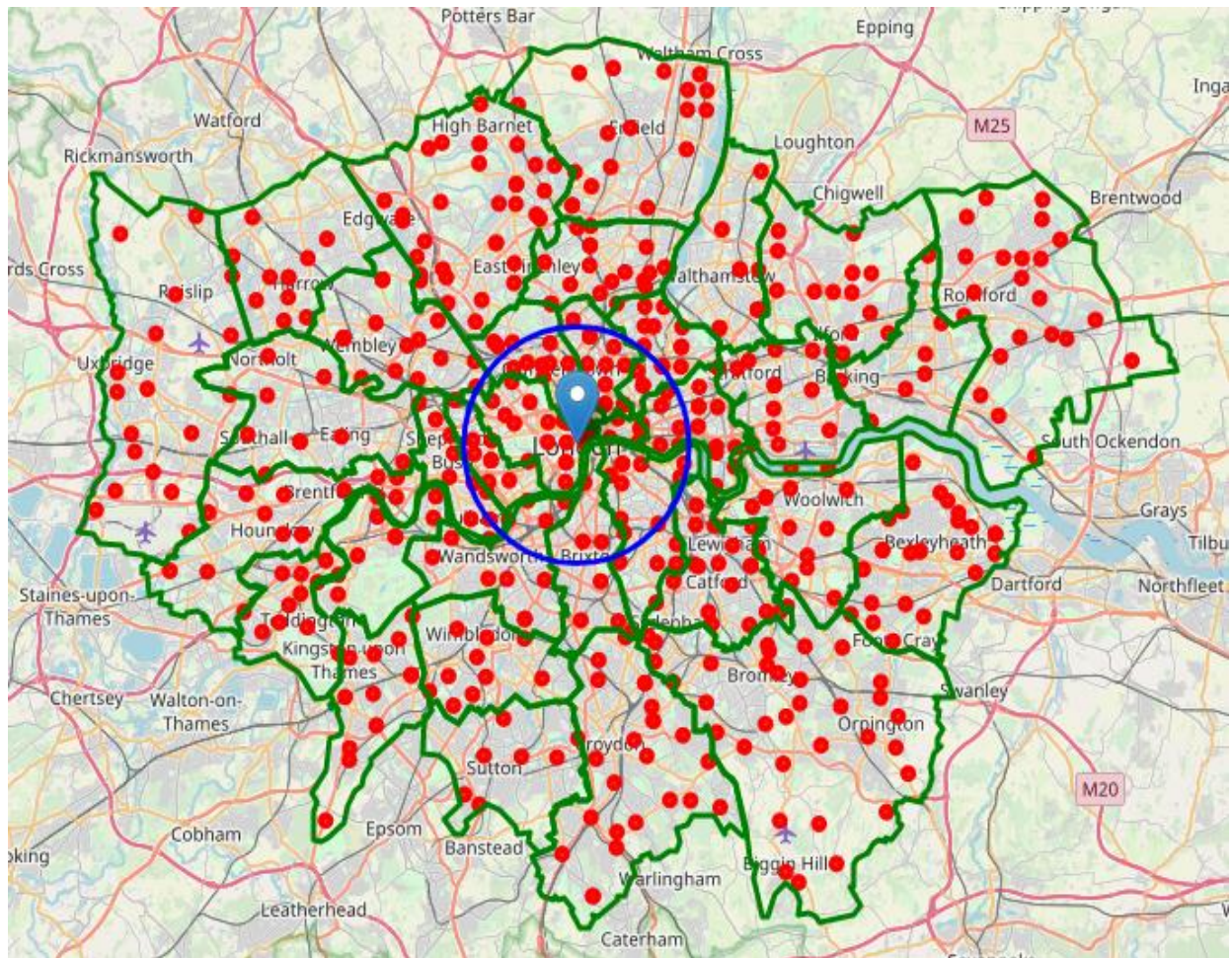
Figure 1: London Neighbourhoods and Borough Boundaries

3. Methodology:

A simple methodology is identified and listed followed;

- Divide the region into several areas.
- Search for all restaurants in the areas.
- Identify the density of restaurants and pick a region with least restaurant density.
- Run the Kmeans clustering algorithm to identify cluster centres, which become the candidates of search.

The neighbourhoods of the London will not be specifically considered. The region under 6000 metres radius of London is divided into several areas to begin the survey of restaurants. The restaurant data is extracted from Foursquare using sandbox account.

The search radius of each area is set to 300m. This gives us about 349 areas to study. While utilising Foursquare, the search radius is increased to 430m in order to cover the missing regions between the area circles.

Figure 2 shows all the areas identified with their numbers and centres. **Figure 3** shows all the restaurants (in blue) in the 349 areas. Out of interest I have plotted the Italian restaurants (in red) in order to see how it fares in terms of location.

There are about 5400 restaurants out of which about 430 are Italian restaurants. It is to be mentioned that food places like Café, Bakery etc have not been considered.

Upon visual inspection and based on the restaurant density, the south east of London seems to have lesser restaurant density than the rest of London. This region falls under Areas 120 to 194. Clustering algorithm is applied on these areas.

The clustering algorithm is applied to shortlisted region with 15 clusters. **Figure 4** shows the shortlisted areas 120 to 194 and the 15 cluster locations.
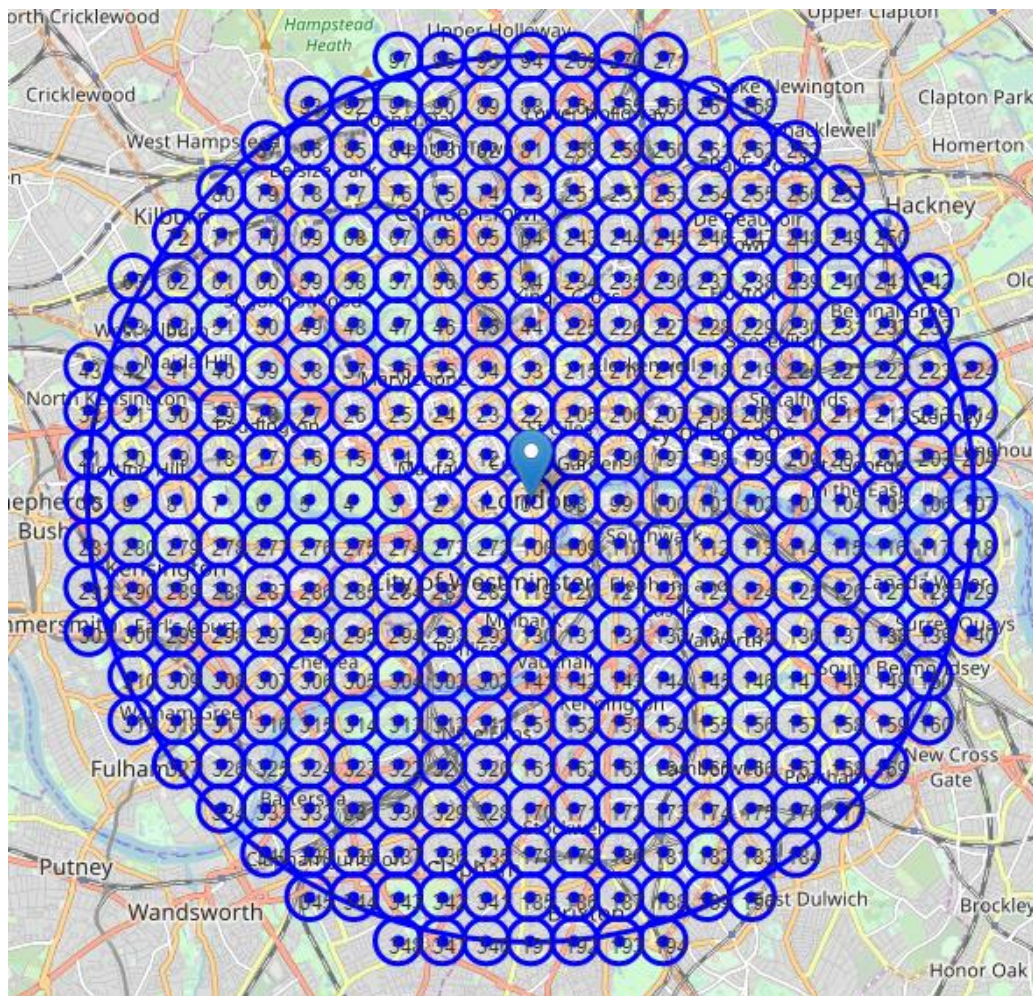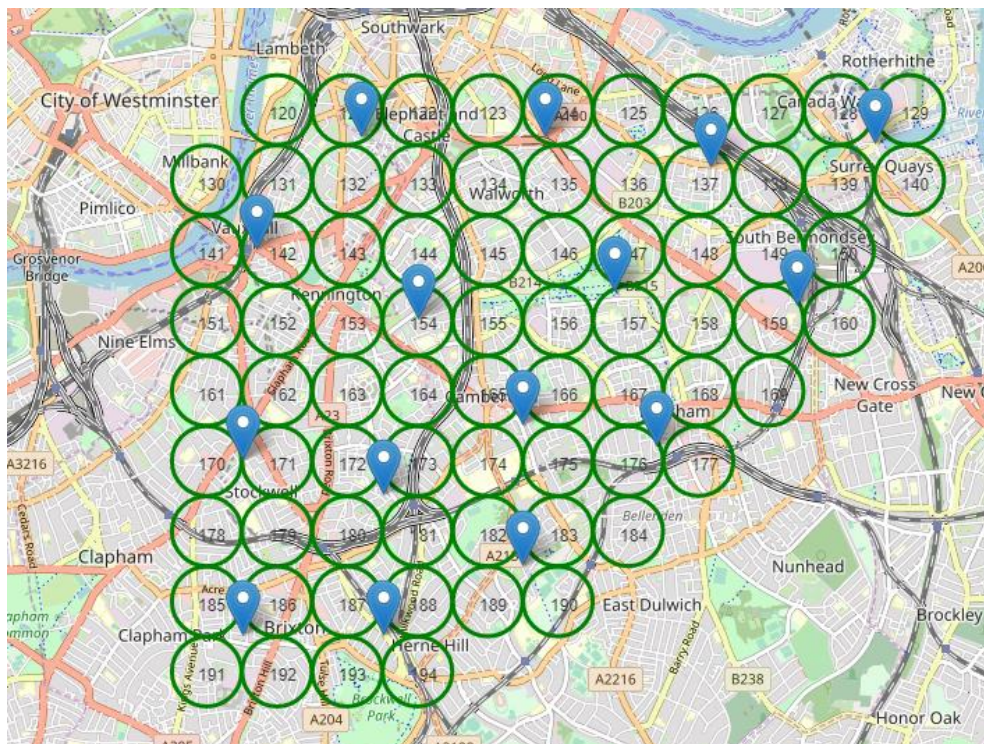

Figure 2: Areas of search
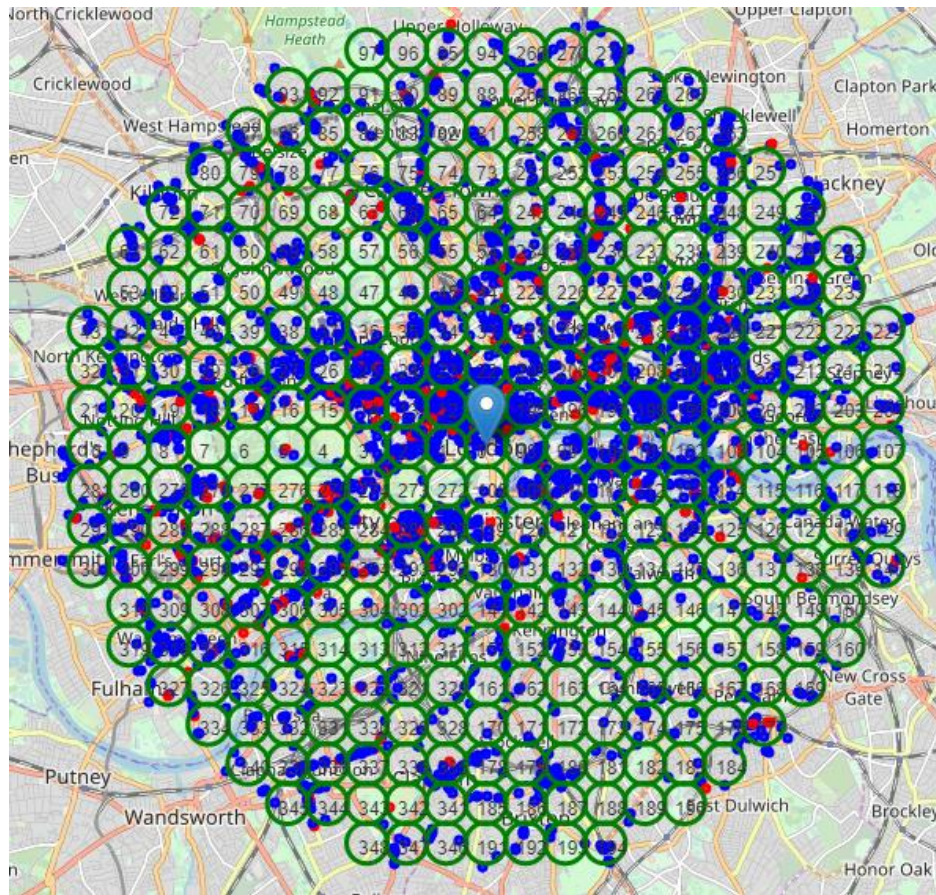
Figure 3: All the Restaurants including Italian Restaurants

Figure 4: Areas 120 to 194 with 15 cluster locations

4. Conclusion & Discussion:

The 15 clusters centres are the probable locations where a stakeholder can begin their search. The addresses are listed below. The result is taken straight from the program.

```
----------------Search Locations to set up a Restaurant-----------

1. Royal Borough of Greenwich, London, Greater London
2. Children's play area, Park Vista, East Greenwich, Royal Borough of Greenwich, London, Greater London, SE10 9LY
3. 18, Baizdon Road, Blackheath Vale, Blackheath, London Borough of Lewisham, London, Greater London, SE3 0US
4. GR40, Tunnel Avenue, Greenwich Peninsula, Royal Borough of Greenwich, London, Greater London, SE10 0PT
5. GR40, Tunnel Avenue, Greenwich Peninsula, Royal Borough of Greenwich, London, Greater London, SE10 0PT
6. Lassell Street, East Greenwich, Royal Borough of Greenwich, London, Greater London, SE10 9GL
7. 4-20, Northbrook Road, Blackheath Vale, Hither Green, London Borough of Lewisham, London, Greater London, SE13 5QT
8. 2-14, Blessington Road, Blackheath Vale, Blackheath, London Borough of Lewisham, London, Greater London, SE13 5ED
9. Whitfield's Mound, Goffers Road, Blackheath Vale, Blackheath, London Borough of Lewisham, London, Greater London, SE3 7BN
10. GR40, Tunnel Avenue, Greenwich Peninsula, Royal Borough of Greenwich, London, Greater London, SE10 0PT
11. GR38, Morden Wharf Road, Greenwich Peninsula, Royal Borough of Greenwich, London, Greater London, SE10 0NU
12. Cross Avenue, East Greenwich, Royal Borough of Greenwich, London, Greater London, SE10 8XG
13. Arches Leisure Centre, Greenwich Park Street, East Greenwich, Royal Borough of Greenwich, London, Greater London, SE10 9LT
14. Blackheath Vale, Blackheath, London Borough of Lewisham, London, Greater London
15. 4-20, Northbrook Road, Blackheath Vale, Hither Green, London Borough of Lewisham, London, Greater London, SE13 5QT
```

Figure 5: Probable Restaurant search locations

It is observed that broadly, **Royal Borough of Greenwich** and **London Borough of Lewisham** are good boroughs to search for a restaurant location. There are several factors that go into decision making. The above analysis is pretty basic and is based on simple restaurant density in the area. Future work can include population, income etc layers on to the existing one. And can also include a specific cuisine restaurant.