Introduction

With over 6 million people living in 33 different boroughs, London is the capital of the United Kingdom and one of the most important financial cities in the world. As you can imagine, whenever any family needs to move into London, or just one to relocate within the city, the decision is not easy: Not all of the neighbourhoods have the same access to services or places. Some areas could have really good transportation links however they may not be close enough to supermarkets or grocery shops. Some others could have extraordinary access to hospitals and or doctors but may not have any park, or school for the kids in the family near to them.

The target audience of this analysis is a young family with children that has to make the decision of which area of London is going to be next place to his new home and it is aimed to provide enough good information about each of the 33 boroughs of the London metropolitan Area to help in the decision making process.

We would like to classify each boroughs of the London metropolitan area based on the population and the number of services that each borough has into main categories such as: Transportation, Health Services, Education, Well-Being and Daily needs that are considered the most relevant for the target audience.

Data

To do the analysis a number of online sources in conjunction with Forsquare location data can be used. The details of them are as follows:

* **Population Data of the UK by Authority Area**

This can be found online in Web from the Office for National Statistics of the British Government (<https://www.ons.gov.uk/>). The link is: <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland> In the file we can find the spreadsheet MYE2 - Persons with the estimates of the British population by Authority area of 2019.

Each row of the data represents one Authority Area of UK (in the case of the London Metropolitan Area, the file will have a row for each of the neighbourhoods or boroughs of London) The columns represents the number of people living in 2019 for each age. There is also a column with the Total that is the one that we are interested in.

In the file we can see that each Authority Area has a code. The codes for the 33 London boroughs are the ones Starting in E09000001 and ending in E09000033

* **GeoJson file with the boundaries of all district areas for the UK**

This can be found online in the British Government data portal <https://data.gov.uk/>

The link to the particular GeoJson File is: [http://geoportal1-ons.opendata.arcgis.com/datasets/fab4feab211c4899b602ecfbfbc420a3\_3.geojson?outSR={%22latestWkid%22:4326,%22wkid%22:4326}](http://geoportal1-ons.opendata.arcgis.com/datasets/fab4feab211c4899b602ecfbfbc420a3_3.geojson?outSR=%7B%22latestWkid%22:4326,%22wkid%22:4326%7D)

This file is providing the boundaries of each Authority Area (borough or neighbourhood) for the whole United Kingdom. For each area we can see the name, the code of the neighbourhood (the same as in the population data file) and the geometry of the area by points (Longitude, Latitude) of the boundary

* **Foursquare location data** for all the venues within the London Metropolitan Area of the next types: Hospital, Doctor, Pharmacy, Train Platforms, Underground, Light Rail, Park, Pool, Gym, Fruit and Vegetable shop, Supermarket, Shopping mall, Elementary School, Middle School and Preschool*.*

Methodology

# Data preparation

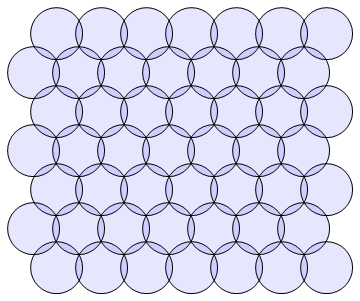
## GeoJson data:

The data obtained from the British government provides all the boundaries of all British Local Authority Areas. Since the analysis is focused in London a modification in the file is needed dropping all the areas that are not within London metropolitan Area. In future, this new file will be used to exclude any venues that are outside London and to display the chlorophet map with only the boroughs in the analysis

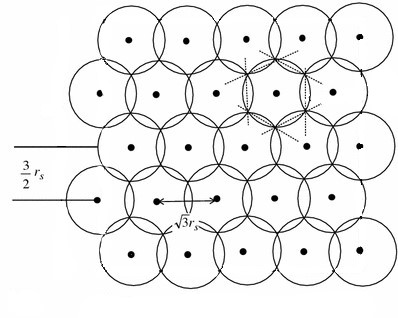
## Foursquare data:

The objective is to obtain all the venues of the selected categories that are inside the London metropolitan Area. The challenge in this is the limits in number of requests and number of venues in the results that Foursquare set for free accounts. A Foursquare request, centred in the centre of London with a radius large enough to cover all the metropolitan area and with all the categories we are interested could be used, however only 50 venues would be retrieved due to Foursquare limits. To ensure that all the venues of each category are retrieved from the Foursquare database a sweep of the London metropolitan area with circles of a radius small enough to not saturate the request with over 50 results is used. Going further a sweep of the area looking for venues of only one particular category is done. In that way the number of sweeps to the London metropolitan area will match exactly the number of distinct types of venues that the target audience of the analysis are interested in.

### Creation of the mesh of points

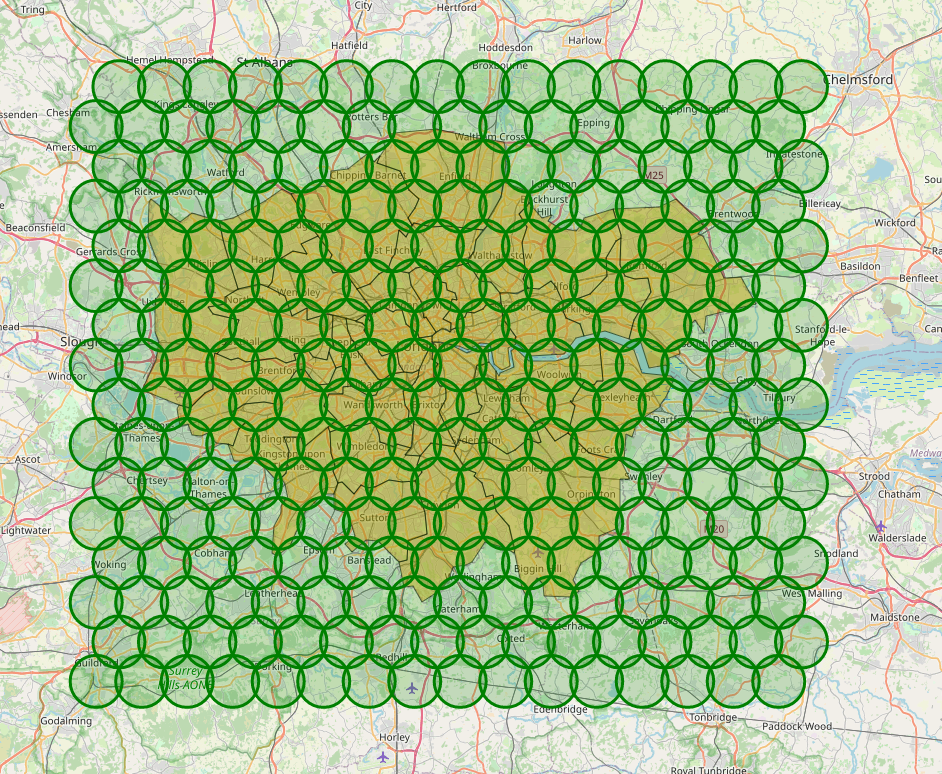
Foursquare only allows to get venues by giving it a centre and a radius. In this way we need to create a mesh of points (Longitude, Latitude) that will be used as centres in the Foursquare request and that will cover the entire area of all London neighbourhoods. The layout of circles that will cover the area with less overlapping zones is similar to the one in the next picture.

The distance between the centre of each circle in function of the radius are defined by the next expressions:

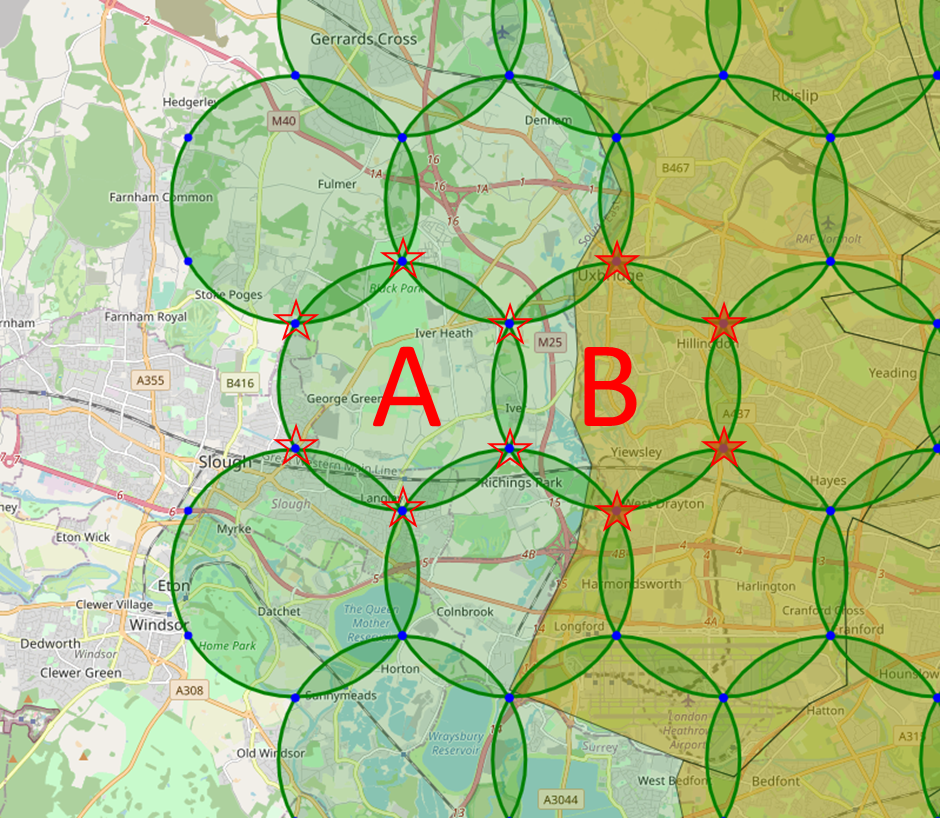


Once the distance between points is known we only need to define where to place the mesh. The answer to that lies with the GeoJson data from the previous step: by checking what are the minimum and maximum values of the coordinates longitude and latitude of the geographical points defining the boundary of all the neighbourhoods of London we obtain the 4 vertex of the rectangle that set the boundaries of the mesh.

The next picture shows the mesh obtained:

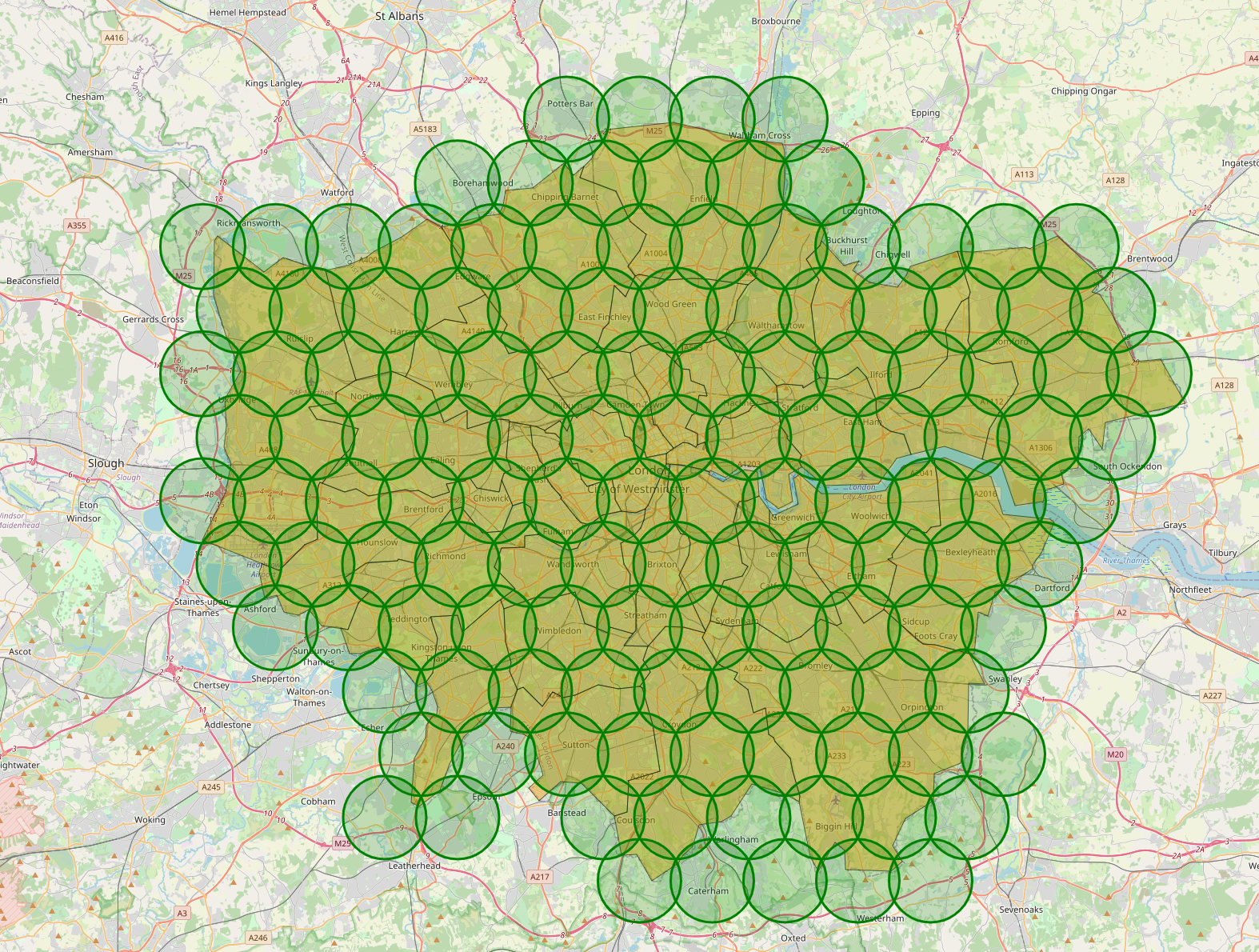


As we can see we cover the whole area of London, however there are a number of circles that are entirely out of any London area. The next step is to remove them from the mesh as this will lead to requests to Foursquare API that will not bring useful data for our analysis. The resulting mesh of points is displayed in the next figure:

To get rid off the circles a check in 6 points of the its circumference is used to discard any circle with all the 6 points out of London.

Circle A is removed from the mesh as none the 6 points belongs to London area. Circle B is kept as 4 points fall under London area.

The next image shows the final mesh obtained:



### Foursquare request

In total 15 different venues categories have been selected for the analysis. They all fall into the main groups that are considered the lead for the project that can be see next including the Foursquare venue category code. This piece of information has been obtained from the Foursquare documentation (<https://developer.foursquare.com/docs/build-with-foursquare/categories/>)

Health:

Hospital: 4bf58dd8d48988d196941735

Doctor: 4bf58dd8d48988d177941735

Pharmacy: 4bf58dd8d48988d10f951735

Transport:

Trains\_Platforms: 4f4531504b9074f6e4fb0102

Underground: 4bf58dd8d48988d1fd931735

Light\_rail: 4bf58dd8d48988d1fc931735

Well-Being:

Park: 4bf58dd8d48988d163941735

Pool: 4bf58dd8d48988d15e941735

Gym: 4bf58dd8d48988d175941735

Dailyneeds:

Fruit\_And\_Veg\_Shop: 52f2ab2ebcbc57f1066b8b1c

Supermarket: 52f2ab2ebcbc57f1066b8b46

Shopping\_Mall: 4bf58dd8d48988d1fd941735

Education:

Elementary\_School: 4f4533804b9074f6e4fb0105

Middle\_School: 4f4533814b9074f6e4fb0106

Preschool: 52e81612bcbc57f1066b7a45

exploratory data analysis

inferential statistical testing

if any, and

what machine learnings were used and why.

Results

section where you discuss the results.

Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.

Conclusion

section where you conclude the report.