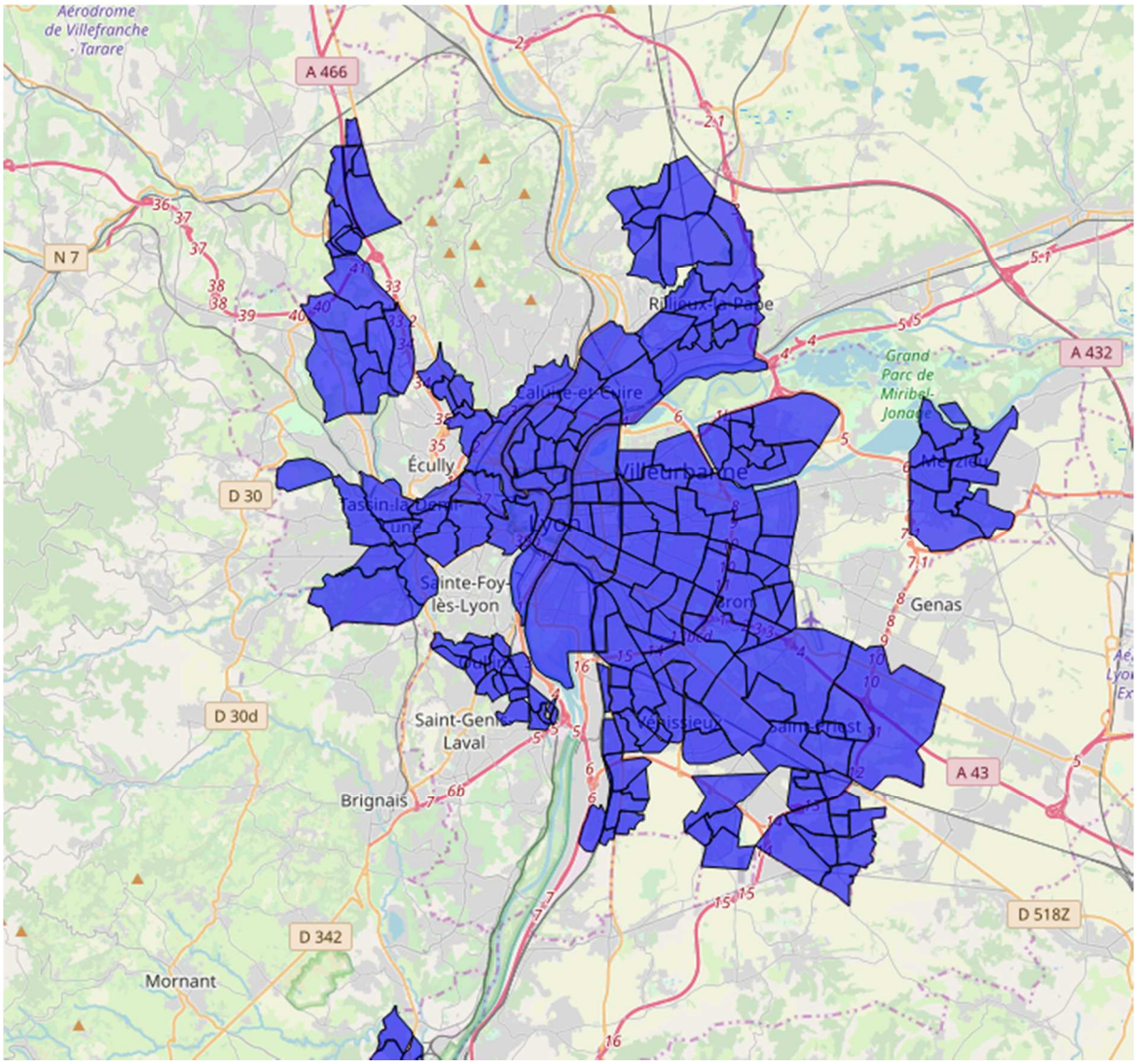
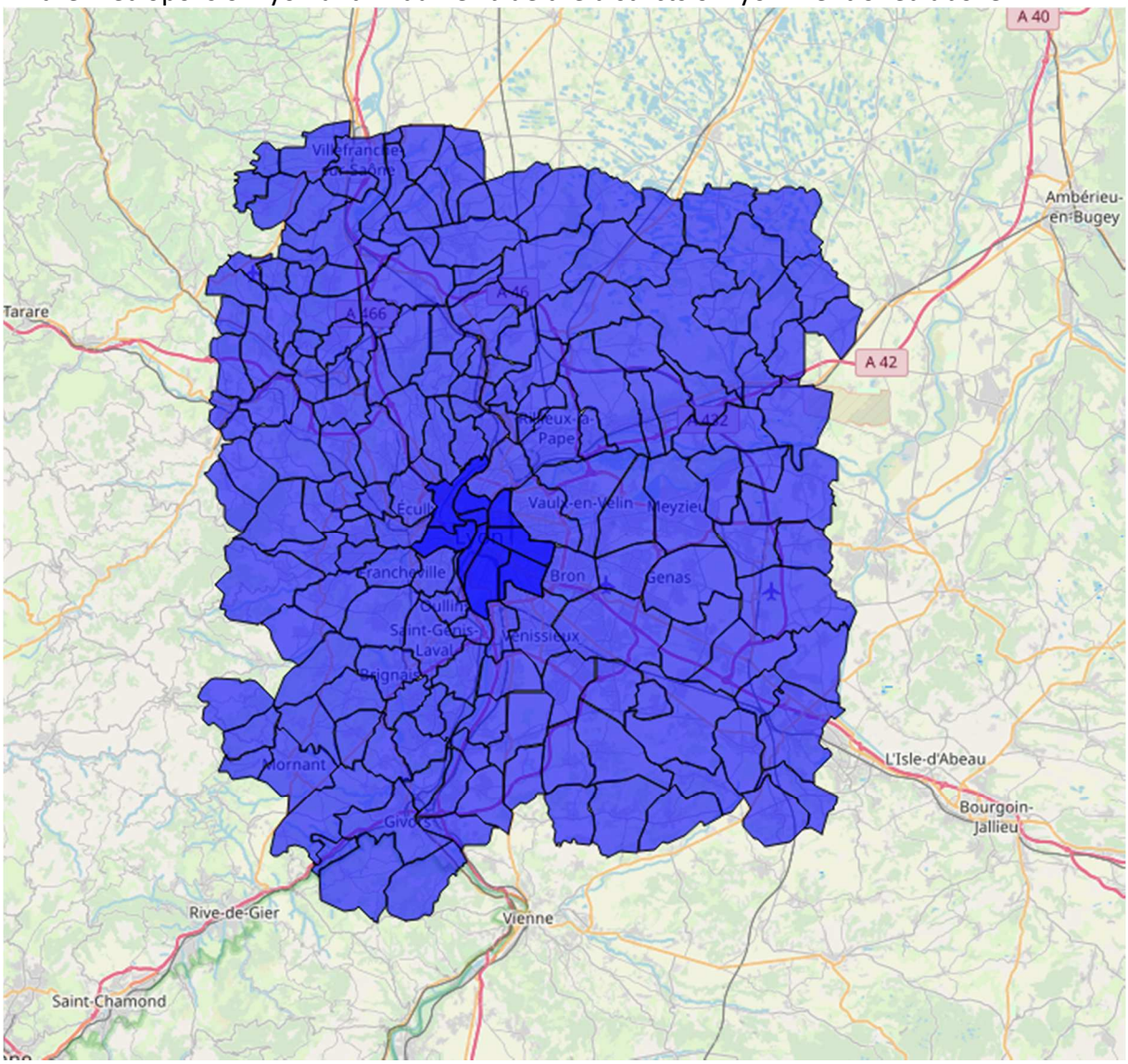
IBM Data Science Capstone
The Battle of Neighborhoods – Week 1

2. **Description of the data and how it will be used to solve the problem.**
   First we have to identify the neighborhoods of the metropolis of Lyon. On the open platform for French public data: data.gouv.fr we can search for appropriate data including geo-data.
   We investigate and find three kind of sources which can be our neighborhoods on a subdivision of the platform about Lyon: data.grandlyon.com. The first are the administrative districts of the town of Lyon. It is not good because it doesn't cover the metropolis and also, the administrative districts (called "arrondissement" in French) are much larger than the idea of a neighborhood. The second are the neighborhoods of the metropolis "Quartiers des communes de la Métropole de Lyon". You can see below the map.



We note that not all neighborhoods are represented, and some are not contiuous. So we look for the third source which are the cities of the metropolis, in the sense of the French administration called a “commune”. You can see below the map of the cities of the metropolis of Lyon and in darker blue the districts of Lyon mentioned above.



Conversely, this perimeter is much too wide for our purpose. So we decide to keep the second source “Quartiers des communes de la Métropole de Lyon” which is the closest to our needs.
With more time on it, it will be appropriate to complete those data with the missing neighborhoods, but they are not the main places of the metropolis of Lyon therefore it will not be so limiting for us.

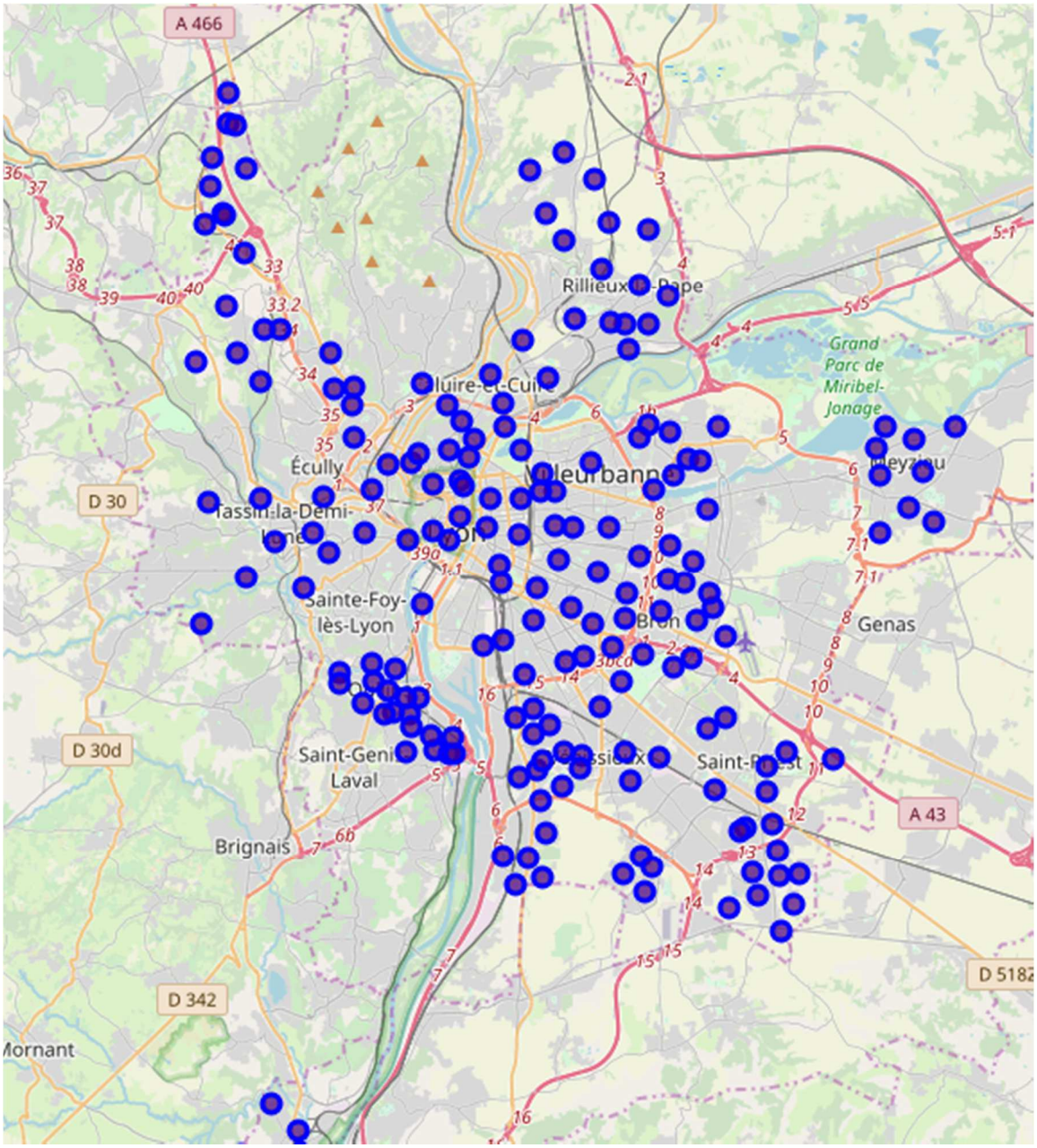
FOURSQUARE

We also use the data of foursquare to get the different venues of the neighborhoods we identify based on what we study previously in this course.

We also want to add some data that are not in the foursquare venues. This are the location of schools (primary, secondary and high schools) and hospitals. We find those

geo-data on the same platform data.gouv.fr, three sources for the schools (one for each kind of school) and one for hospital.

The data for the neighborhoods give us the coordinates of the limit and not a latitude and longitude, so we calculate the centroid to get the lat and long of each neighborhoods as you can see on the map below.



Based on those data, we can run Foursquare API to get the venues of each neighborhoods in a radius of 1000 meters. After that we concatenate to it the schools and hospital in the same radius to complete our data.

We can then identify the five most common venues (including schools and hospitals) of each neighborhood as for example below.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Bussière	Restaurant	PrimarySchool	Pharmacy	FoodStore	Flea Market
1	Cadière	Harbor / Marina	Supermarket	Shopping	Services	Restaurant
2	Centre-ville	PrimarySchool	Theater	TakeAway	SecondarySchool	FoodStore
3	Chassagnes	HighSchool	Park	Sport	SecondarySchool	Pool
4	Clavelière	HighSchool	Pharmacy	Shopping	Services	SecondarySchool

This is what we will use to answer to the initial question. The next section will explain the methodology.