Exploring restaurant locations in Villeneuve d'Ascq, France

Artem Kondratskyi

April, 2020

Contents

Introduction

Villeneuve d'Ascq, where I'm currently living, is one of the main cities of the Hauts-de-France region in France. This town is located near Lille and is famous for its scientific, business and sports facilities. Villeneuve d'Ascq is the main academic pole of the Lille metropolitan area and hosts two main campuses of the University of Lille along with five graduate schools. Villeneuve d'Ascq has more than 63000 inhabitants with a population density of 2271 people per square kilometer. Villeneuve-d'Ascq is also one of the main economic spots of the Hauts-de-France region and many multinational corporations, including Bonduelle, Cofidis and Decathlon, have their head offices there.

Villeneuve d'Ascq is a great place to live and is attractive for businesses, including a restaurant business. Therefore, restaurant owners thinking about developing their business as well as entrepreneurs planning to start a restaurant business in Villeneuve d'Ascq, or investors could benefit from the analysis of the restaurant business in this town.

So, this project is devoted to exploring restaurants in Villeneuve d'Ascq, clustering the neighborhoods by most popular restaurants, and finding an optimal place for a future restaurant. By optimal location I mean the place with the least number of restaurants per habitant in a specific neighborhood of this town. Therefore, we'll analyze and compare different neighborhoods/clusters in Villeneuve d'Ascq, in terms of population density and restaurant type/density, and try to come up with the best places for a new restaurant.

Data

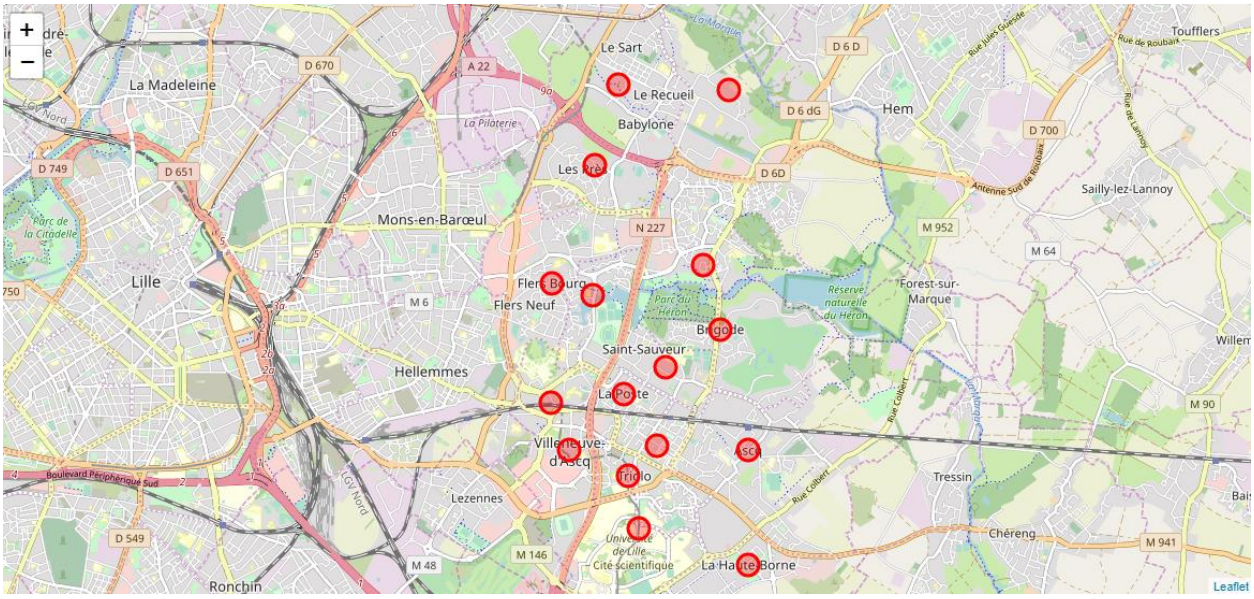
In order to understand restaurant business in Villeneuve d'Ascq, I used the following datasets:

- Villeneuve d'Ascq's Wikipedia page - https://fr.wikipedia.org/wiki/Villeneuve-d%27Ascq
- Various websites, including https://sig.ville.gouv.fr/, to get the number of of inhabitants in each neighborhood.
- Foursquare - to find food venues in each neighborhood.
- Tripadvisor – to complete data from Foursquare.

Methodology

First I scraped Villeneuve d'Ascq's Wikipedia page in order to get the names of neighborhoods. For this I used Beautiful Soup package. I had to select the right table, as there are several tables with common classes on that page. I extracted the names of neighborhoods into a list and then, using GeoPy client I got the geographical coordinates for each neiborhood in this list. The results were transformed to a pandas dataframe containing neighborhood names, as well as latitude and longitude for each neighborhood. So, I got 17 neighborhoods.

These neigborhoods where visualized on the map using Folium library.



I also searched for population data in the neighborhoods of Villeneuve d'Ascq. These data are not readily available for downloading, and the most helpful ressourse was:
https://sig.ville.gouv.fr/tableaux/loadTableauIris/codezone/59009/tableCode/ind111

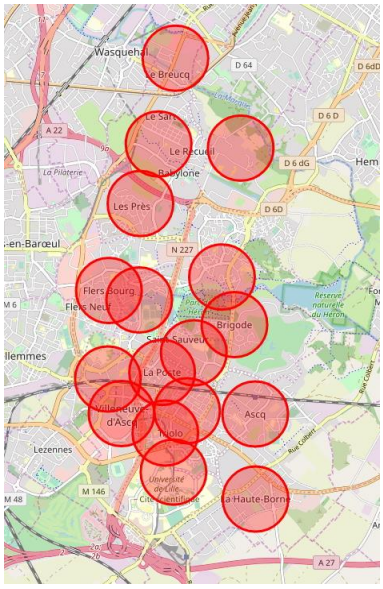
However, I needed to work on it, in order to fit the data in 17 neighborhoods. Given the small number of neighborhoods, I did it manually, saved the data in populationva.csv file and then extracted data from this file into pandas dataframe.

Then I used Foursquare API to explore the neigborhoods and segment them. Given that I decided to concentrate on restaurants, I searched only for FOOD VENUES. This could be done by properly creating the API request:

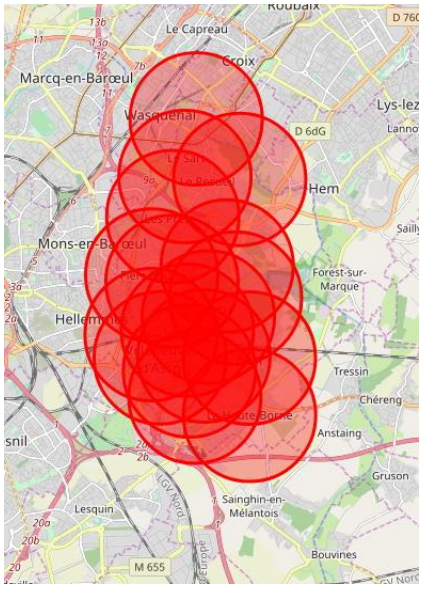
```
# create the API request URL
url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&categoryId={}&limit={}'.format(
    CLIENT_ID,
    CLIENT_SECRET,
    VERSION,
    lat,
    lng,
    radius,
"4d4b7105d754a06374d81259", #food category ID        https://developer.foursquare.com/docs/build-with-foursquare/categories/
    LIMIT)
```

Importantly, I decided to use a radius of 1500 meters in order to cover the total area of Villeneuve d'Ascq. The smaller radius values will lead to missing venues.

Radius = 500 meters



Radius = 1500 meters



Of course, this approach leads to many duplicate venues, because the circles intersect, so any venue situated at the intersection will be attributed to both neighborhoods. We'll need to work on it later.

So, our Foursquare request gave us 474 food venues. I then removed duplicates according to venue name, category and geographical coordinates.

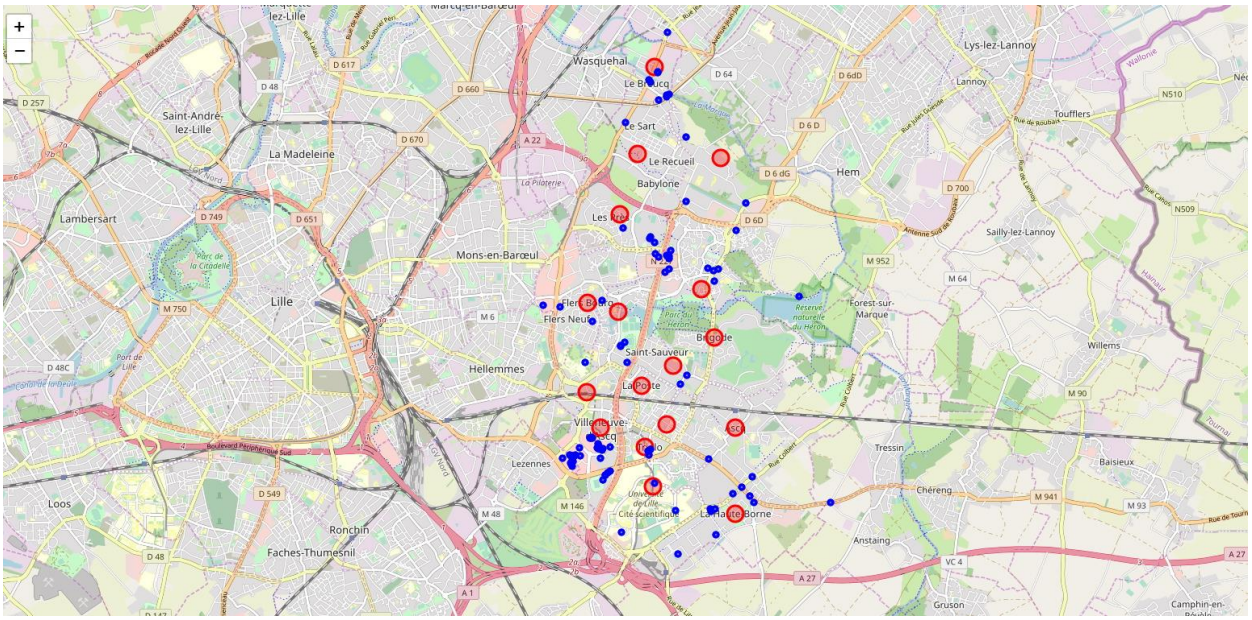
```
drop_duplicates(subset = ['Venue', 'Venue Latitude', 'Venue Longitude', 'Venue Category'], keep='first').reset_index(drop=True)
```

This reduced the list to 136 food venues. But, due to the radius of 1500 meters, we've also covered a part of neighbor towns. So, it's quite possible, that in this list we have some venues which are not situated in Villeneuve d'Ascq, but near Villeneuve d'Ascq. We needed to remove these venues from the list. And I did it by finding the address from latitude/longitude pair for each venue, and then removing all the venues, which are not situated in Villeneuve d'Ascq.

The venue addresses where obtained by using GeoPy's reverse function.

	Venue	Venue Latitude	Venue Longitude	Venue Category	Address
0	La Maison Commune	50.626619	3.149605	French Restaurant	Monument aux Morts, Place de la République, Annappes, Villeneuve-d'Ascq, Lille, Nord, Hauts-de-France, France métropolitaine, 59491, France
1	McDonald's	50.640515	3.145051	Fast Food Restaurant	McDonald's, 211, Rue de la Convention, La Cousinerie, Cousinerie, Villeneuve-d'Ascq, Lille, Nord, Hauts-de-France, France métropolitaine, 59650, France
2	Pic Nic	50.615279	3.154220	Burger Joint	Friterie Picnic, Rue Gaston Baratte, Ascq, Villeneuve-d'Ascq, Lille, Nord, Hauts-de-France, France métropolitaine, 59491, France
3	Zio	50.625351	3.148280	Italian Restaurant	Pizza Zio, 4, Rue de la Station, Annappes, Villeneuve-d'Ascq, Lille, Nord, Hauts-de-France, France métropolitaine, 59650, France
4	Class' Croute	50.640988	3.145950	Sandwich Place	10, Rue de la Cousinerie, La Cousinerie, Cousinerie, Villeneuve-d'Ascq, Lille, Nord, Hauts-de-France, France métropolitaine, 59650, France

Then I dropped all the venues, that are not situated in Villeneuve d'Ascq and I got 97 venues (see below in blue).



Great, but because of big radius, we got a lot of venues attributed to several neighborhoods, not just one. So, after dropping duplicates, we don't know what the right neighborhood for each venue is.

To find the right neighborhood for each venue, I calculated the distance between venue coordinates and neighborhood coordinates and have chosen the neighborhood with the smallest distance to venue (using GeoPy's distance function).

In the end, I had a table of 97 food venues attributed to right neighborhoods.

But, as I live here, I found this number (97) rather small and suspected that, in reality, we have a lot more, so I decided to complete this table with the data from Tripadvisor.

When looking for the restaurant in Villeneuve d'Ascq in Tripadvisor you'll get 50 pages of results. These restaurants are situated not only in Villeneuve

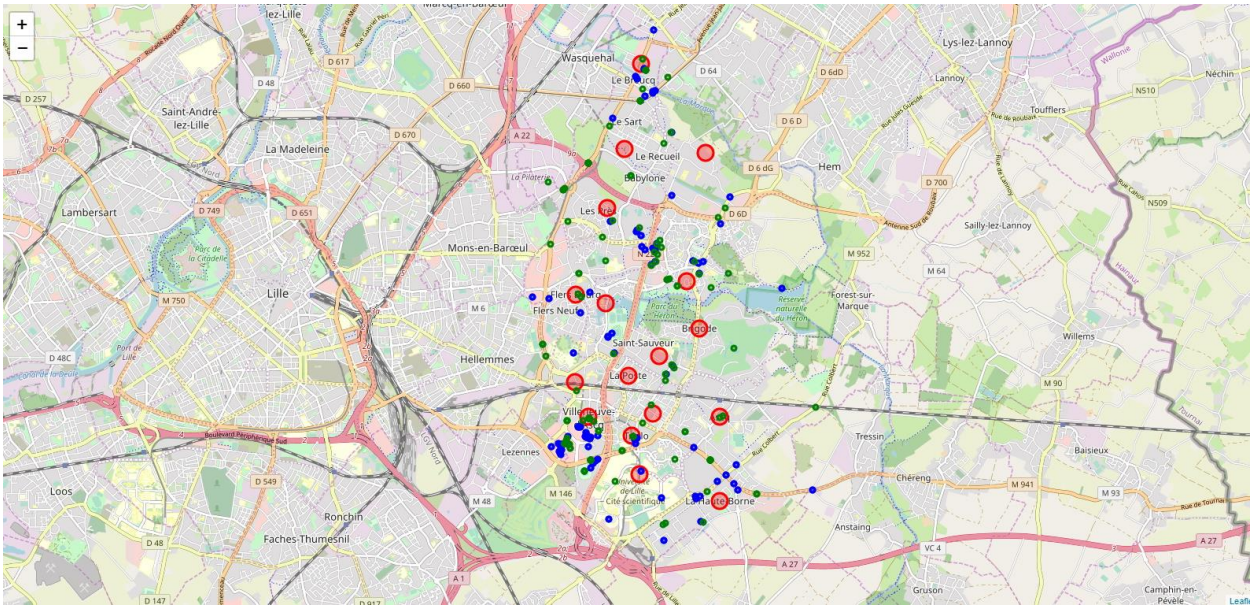
d'Ascq but in its surroundings as well. So, I decided to download all the restaurants, and then drop those not located in Villeneuve d'Ascq.

Scraping Tripadvisor gave me 1730 restaurants in Villeneuve d'Ascq and its surroundings. I needed to clean these data before analyzing them.

After dropping duplicates (because of sponsored content, we can have one restaurant displayed several times) and restaurants outside Villeneuve d'Ascq, I got 153 restaurants, but some of these restaurants contained wrong addresses and GeoPy couldn't come up with geographic coordinates for these restaurants.

After cleaning addresses of the restaurants, I got latitude and longitude for each restaurant using GeoPy. Finally, one restaurant had a completely wrong address and was not situated in Villeneuve d'Ascq, so I dropped it.

As it the case of Foursquare data I found the right neighborhood for each restaurant by calculating the distance between restaurant coordinates and neighborhood coordinates and chosing the neighborhood with the smallest distance to venue. Here is what I got: the neighborhoods are in red, Foursquare data in blue and Tripadvisor data in green.



The next step was to build the final table containing data from Foursquare and Tripadvisor. Obviously, after concatenation of two tables I had a lot of duplicates – the venues that were in both Foursquare and Tripadvisor lists.

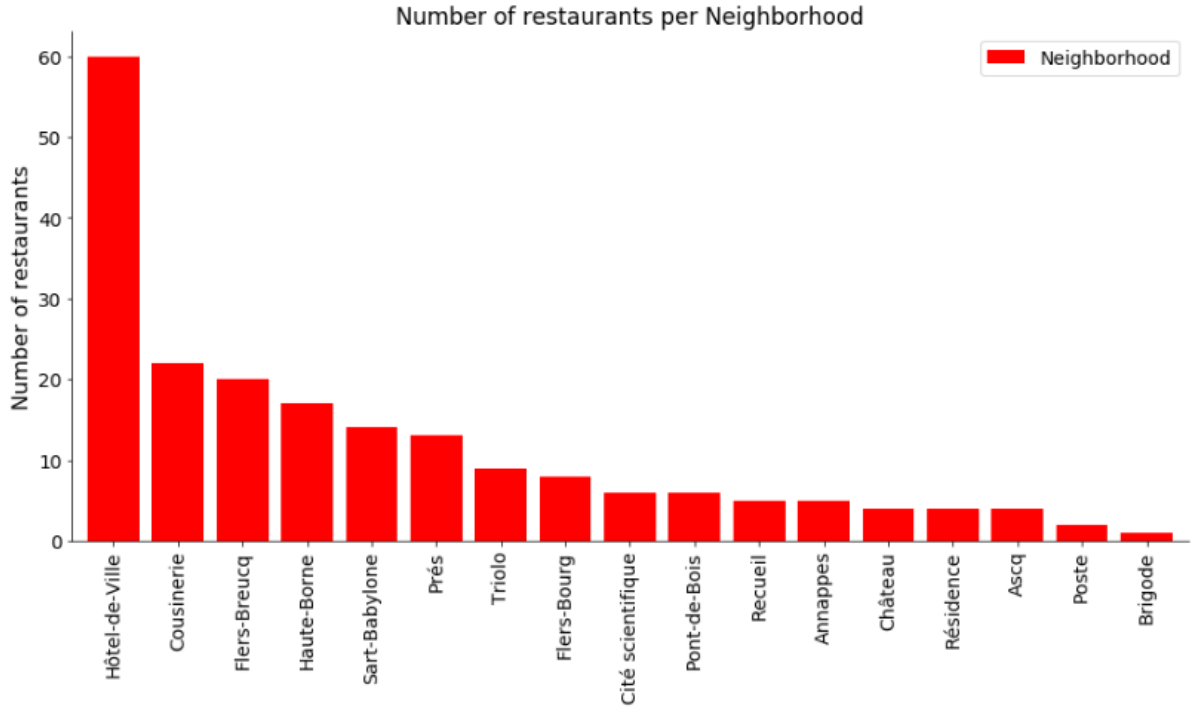
Dropping duplicates was rather tricky here because the same restaurant could have a slightly different spelling in Foursquare and Tripadvisor, so I needed to make some data wrangling before dropping duplicates. Finally, I ended up with 200 unique restaurants in Villeneuve d'Ascq.

The next problem that I needed to solve is the difference between Foursquare categories and Tripadvisor categories. I needed to unify the categories. I invite you to check my Jupyter notebook in order to see how I did it.

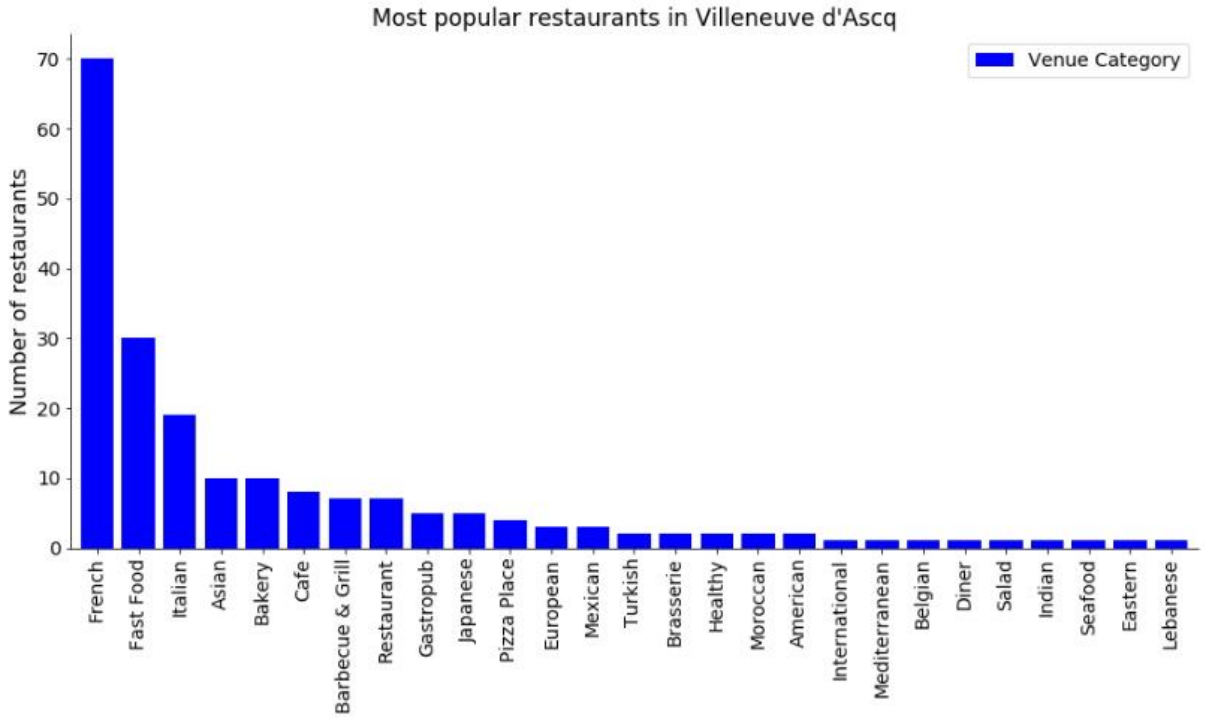
Finally, I got the final table with 200 restaurants, right categories and right neighborhoods – all the required information to do my analysis.

Exploratory data analysis

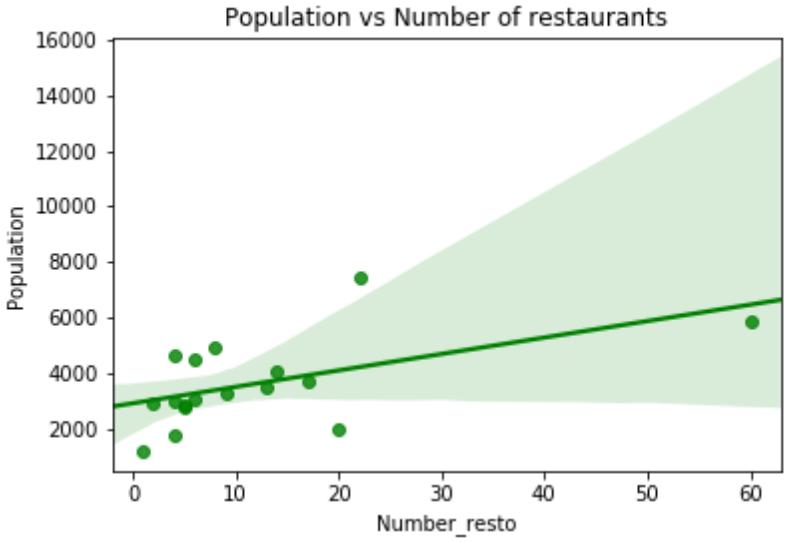
First, I analyzed how many food venues are in each neighborhood.



Then I checked what are the most popular restaurant types in Villeneuve d'Ascq.



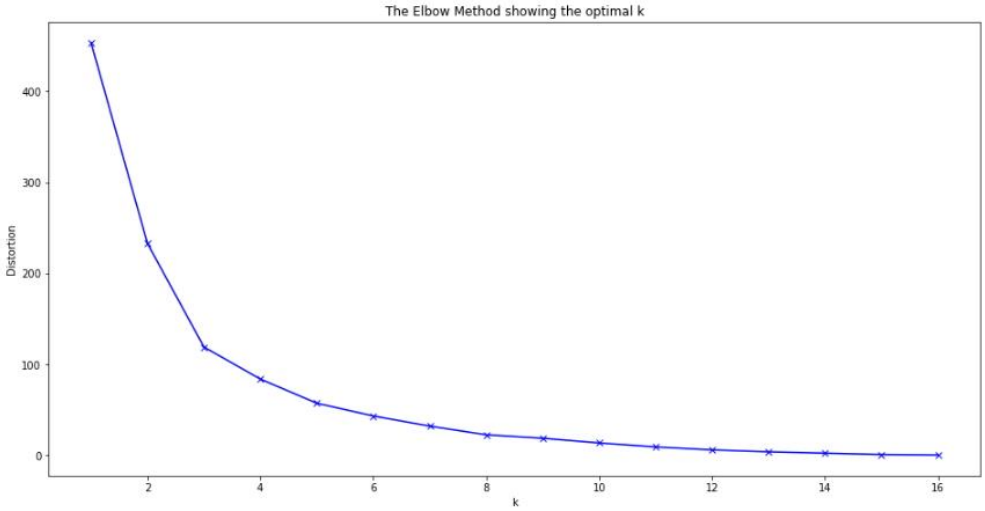
Then I checked if there is any linear correlation between the number of restaurants and the number of people in neighborhoods.



Apparently, there is no correlation between these values.

Then, I analyzed each neighborhood by the restaurant category. I used *one hot encoding* for this. It means that I converted categorical variables into dummy/indicator variables for subsequent analysis. I found 5 most common restaurants in each neighborhood and based on this info performed *k-means clustering* of neighborhoods.

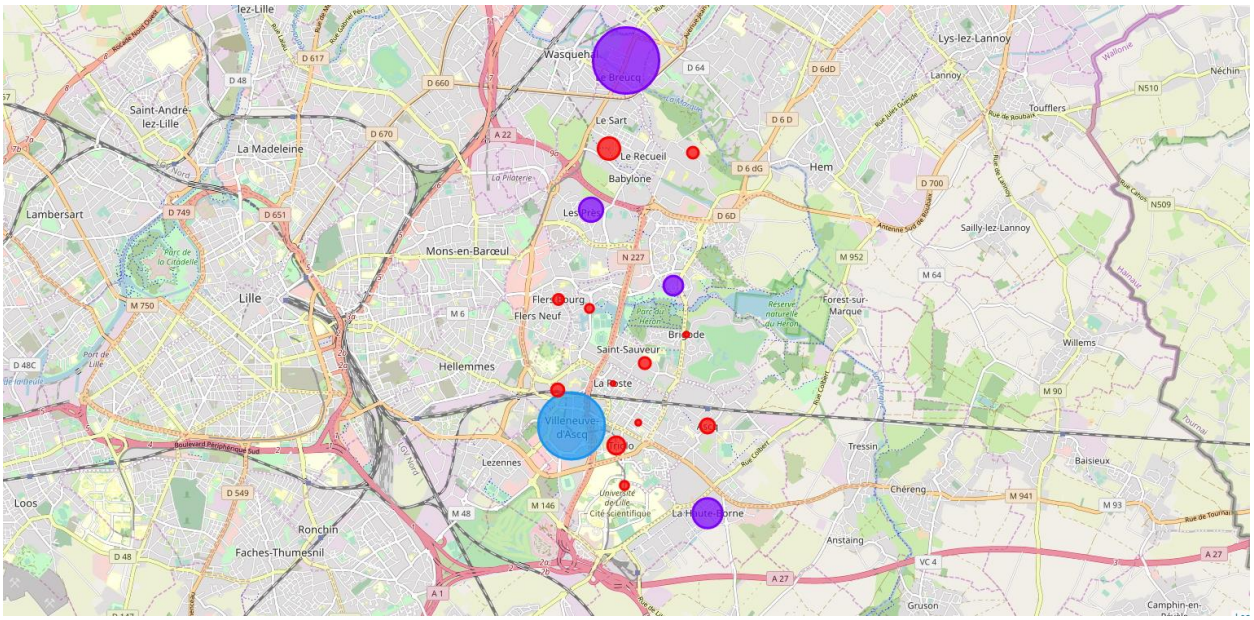
In order to get the optimal number of clusters I used *Elbow method*.



Out of this, I found the optimal number of clusters = 3. Then I performed k-means clustering and come up with this table:

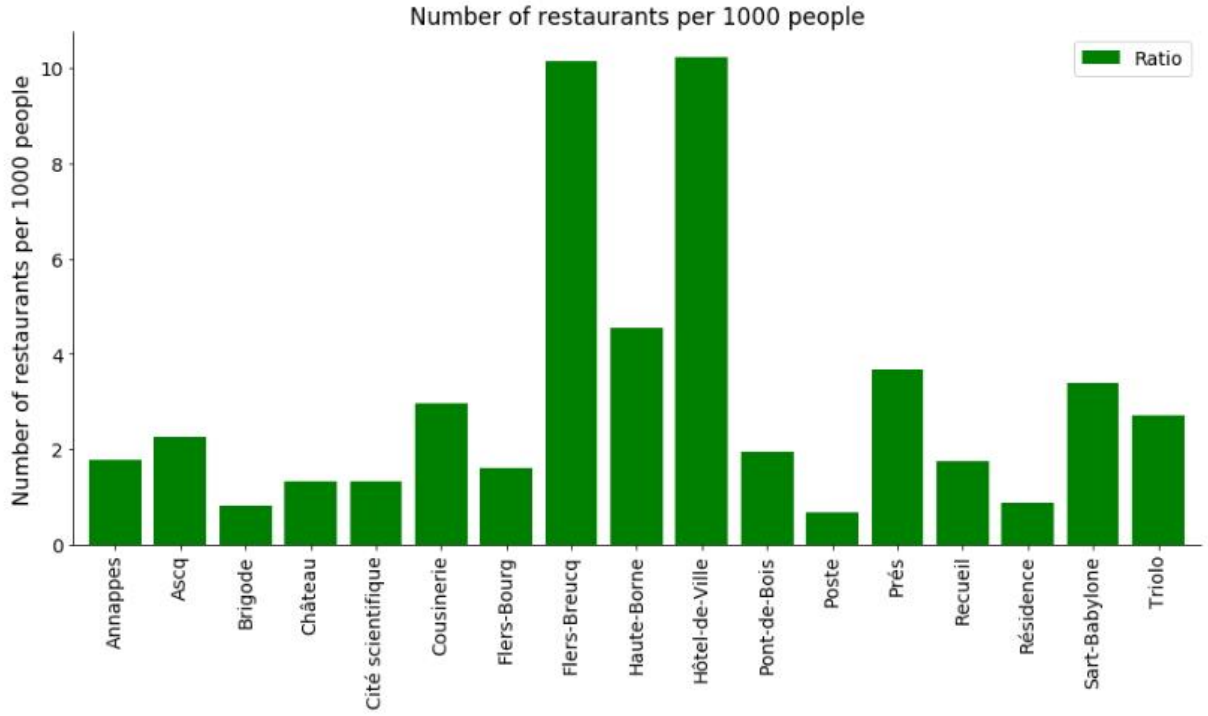
	Neighborhood	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Annappes	0	French	Italian			
1	Ascq	0	French	Italian			
2	Brigode	0	French				
3	Château	0	French	Restaurant	Japanese	Gastropub	
4	Cité scientifique	0	Fast Food	French			
5	Cousinerie	2	French	Fast Food	Bakery	Barbecue & Grill	Brasserie
6	Flers-Bourg	0	French	Restaurant	Bakery	Italian	
7	Flers-Breucq	2	French	Italian	Fast Food	Restaurant	Pizza Place
8	Haute-Borne	2	French	Fast Food	Asian	Restaurant	Pizza Place
9	Hôtel-de-Ville	1	French	Asian	Italian	Fast Food	Cafe
10	Pont-de-Bois	0	Cafe	French	Gastropub	Fast Food	
11	Poste	0	French				
12	Prés	2	French	Fast Food	Bakery	Barbecue & Grill	Italian
13	Recueil	0	French	Brasserie	Eastern		
14	Résidence	0	French	Restaurant	Italian		
15	Sart-Babylone	0	Italian	French	Turkish	Asian	Bakery
16	Triolo	0	Fast Food	French	Restaurant	Lebanese	

On the map my clusters look like this:

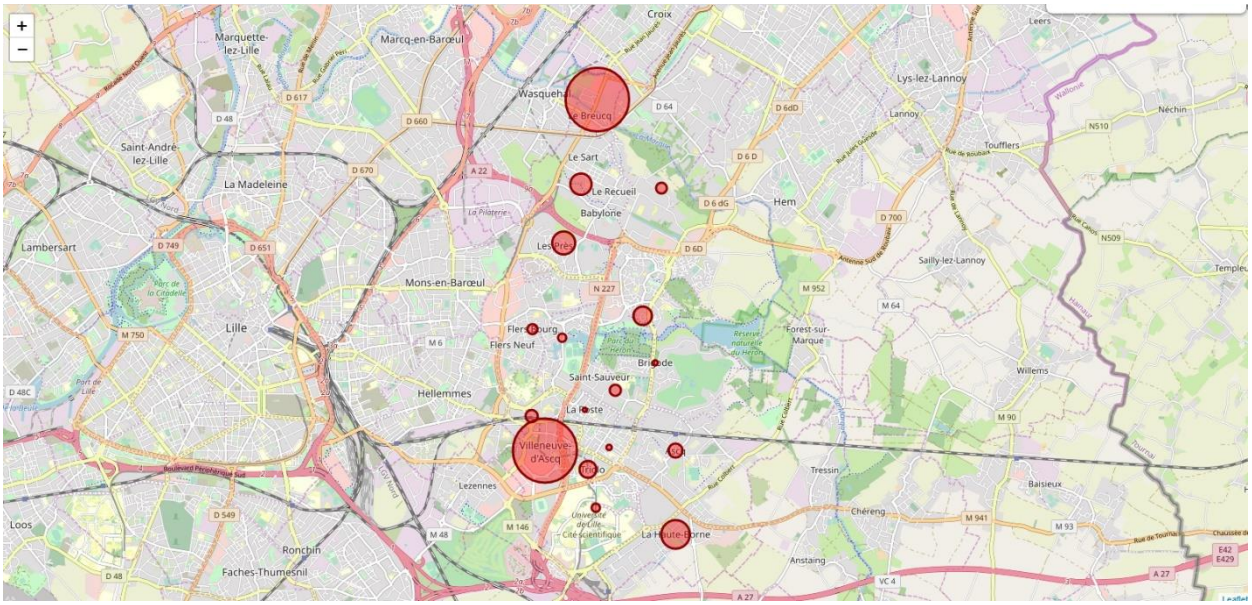


Results

First, I saw how the neighborhoods are distributed by the number of restaurants per 1000 people.



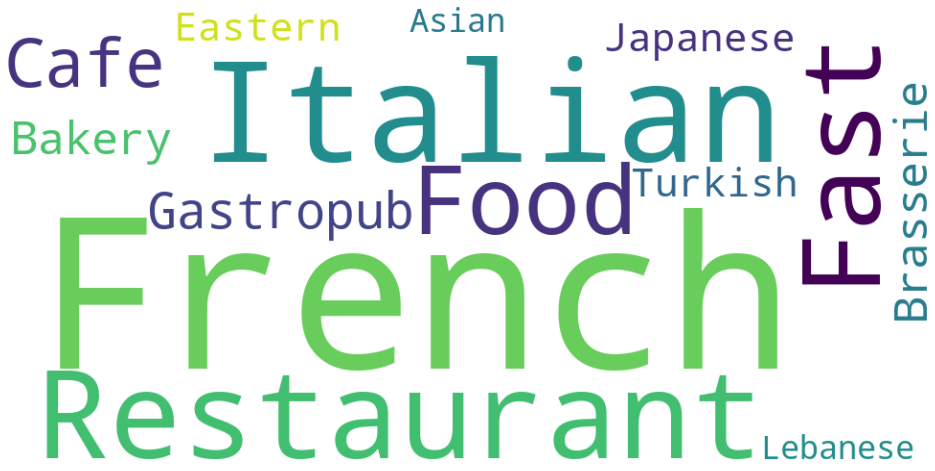
So, here are the most restaurant-reach places per 1000 people in Villeneuve d'Ascq: Hotel-de-Ville and Flers-Breucq. On the other hand, Brigode, Poste and Residence are underserved by restaurants.



As for clustering, I got 3 clusters of neighborhoods.

Cluster 0. French-Italian.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Annappes	French	Italian			
1	Ascq	French	Italian			
2	Brigode	French				
3	Château	French	Restaurant	Japanese	Gastropub	
4	Cité scientifique	Fast Food	French			
5	Fiers-Bourg	French	Restaurant	Bakery	Italian	
6	Pont-de-Bois	Cafe	French	Gastropub	Fast Food	
7	Poste	French				
8	Recueil	French	Brasserie	Eastern		
9	Résidence	French	Restaurant	Italian		
10	Sart-Babylone	Italian	French	Turkish	Asian	Bakery
11	Triolo	Fast Food	French	Restaurant	Lebanese	



Cluster 1. French – Asian.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Hôtel-de-Ville	French	Asian	Italian	Fast Food	Cafe

Cluster 2. French – Fast Food.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Cousinerie	French	Fast Food	Bakery	Barbecue & Grill	Brasserie
1	Flers-Breucq	French	Italian	Fast Food	Restaurant	Pizza Place
2	Haute-Borne	French	Fast Food	Asian	Restaurant	Pizza Place
3	Prés	French	Fast Food	Bakery	Barbecue & Grill	Italian

Discussion

I performed an analysis of a restaurant business in Villeneuve d'Ascq. Well, obviously, restaurants are not distributed evenly over the town. There are some hotspots as well as restaurant-free areas. Hotspots are the commercial

centers with many shops and restaurants accumulated in one place, whereas residential areas with individual houses are mostly restaurant-free.

According to my analysis, the most restaurant-rich places in Villeneuve d'Ascq are Hotel-de-Ville, Cousinerie and Flers-Breucq. At the same time, the neighborhoods having the largest number of restaurants per 1000 inhabitants are: Hotel-de-Ville, Flers-Breucq and Haute Borne. On the other hand Brigode and Poste lack restaurants.

The most popular by far is French cuisine, followed by Fast food, Italian and Asian cuisine.

When deciding where is the best place for a new restaurant, you need to take into account different factors, including restaurant type, neighborhood, population, the wealth of the population in the particular neighborhood, the taste of people and so on. It's very important to perform a complete analysis of neighborhoods before deciding where to open a restaurant.

It this project I clustered neighborhoods in Villeneuve d'Ascq based on the most popular restaurants. This can give an idea of where to open a new restaurant. We see that although French restaurants are the most popular choises in all 3 clusters, these clusters differ greatly in their second most popular food venue. Cluster 0 is more Italian, Cluster 1 is more Asian, whereas cluster 2 is more Fast Food.

If I were planning a restaurant opening, I would most probably focus on claster 2 and, more specifically, on Haute Borne district. Haute Borne is a new district of Villeneuve d'Ascq, which hosts the European scientific park, which is a home for many high-tech and R&D companies. Haute Borne is situated very close to the University of Lille. It is a rapidly developing district and it's a very attractive place for a new restaurant. From our analysis, we also see that Italian restaurants are underrepresented in Haute Borne, so, that could be an idea to develop.

Conclusion

Neighborhoods are like small worlds in a big world – city. The population could differ strikingly between different neighborhoods. While this is mostly true for big cities, it could be much less pronounced in small towns. Nevertheless, "rich" and "poor" districts exist in almost every town,

regardless of its size. So, the strategy of a restaurant entrepreneur should be built according to neighborhood realities.

Therefore, one should critically assess the results and approaches of this particular project. In real-life situations, we need to take into account more factors in order to decide where we can start the restaurant in a particular town.