

Analysis of Brussels Neighborhoods for future residents

1. Introduction

The present project is an analysis of the neighborhoods of the city of Brussels and its surrounding municipalities (Brussels Capital Region), intended to cluster neighborhoods according to the availability of services and comparing them to see which neighborhood fits my personal interests best.

The Brussels Capital region is located at the center of Belgium, and it consists of 19 different municipalities, including the City of Brussels, which is the capital of Belgium. Because of its central location, the region forms part of both the French Community of Belgium, and the Flemish Community.

Brussels is the most populated and richest region of Belgium, covering 162km² with 1.2 million inhabitants.

Since the end of the Second World War, Brussels has played an important role in international politics, becoming the capital of the European Union, and the headquarters of NATO. On a cultural note, Brussels is known for its gastronomy and its architectural landmarks, some of which are categorized as UNESCO World Heritage sites (e.g. La Grand Place, Manneken Pis, Atomium, etc.)[1].

As a resident of Brussels, I decided to use the Foursquare API to cluster the neighborhoods of Brussels, with the goal to create a tool to help aspiring residents when they look for accommodation in the city. I will first create a basic recommendation system, based on my top 10 preferred venue categories and their mean occurrence per neighborhood; the neighborhood with the highest score wins!

In order to provide a second layer to the analysis, I decided to cluster all the 19 municipalities into different groups, depending on how similar are they in terms of the type of venues they have. This provides more alternatives in case living in your top-ranked neighborhood is not really a viable option.

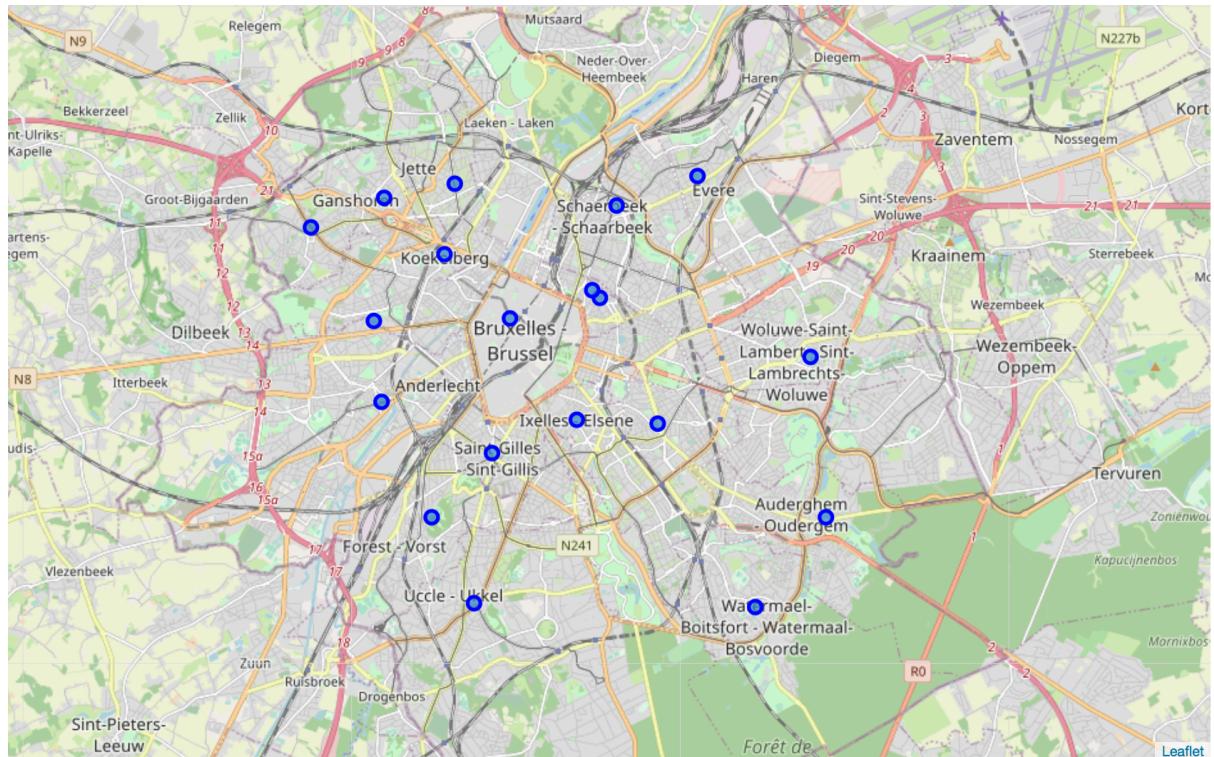


Figure 1 - Brussels and its municipalities

2. Methods and data

In order to pull information from the Foursquare API, it is first necessary to extract the geographical coordinates of the neighborhoods of Brussels. For this, we will have to use the Geocoder Python Package, which extracts the coordinates of a neighborhood/municipality based on the respective postal code.

Now, to obtain the lists of postal codes and neighborhood names, I will scrape this table from the Wikipedia List of Brussels' Municipalities [2], using the BeautifulSoup web-scraping package.

	French name	Dutch name	Flag	CoA	post code	Population (1/1/2017)	Area	Population density (km ²)	Ref.
1	Anderlecht	Anderlecht			1070	118,241	17.7 km ² (6.8 sq mi)	6,680	[7]
2	Auderghem	Oudergem			1160	33,313	9.0 km ² (3.5 sq mi)	3,701	[8]
3	Berchem-Sainte-Agathe	Sint-Agatha-Berchem			1082	24,701	2.9 km ² (1.1 sq mi)	8,518	[9]
4	Bruxelles-Ville*	Stad Brussel*			1000 1020 1120 1130	176,545	32.6 km ² (12.6 sq mi)	5,415	[10]
5	Etterbeek	Etterbeek			1040	47,414	3.1 km ² (1.2 sq mi)	15,295	[11]
6	Evere	Evere			1140	40,394	5.0 km ² (1.9 sq mi)	8,079	[12]
7	Forest	Vorst			1190	55,746	6.2 km ² (2.4 sq mi)	8,991	[13]
8	Ganshoren	Ganshoren			1083	24,596	2.5 km ² (1.0 sq mi)	9,838	[14]
9	Ixelles	Elsene			1050	86,244	6.3 km ² (2.4 sq mi)	13,690	[15]
10	Jette	Jette			1090	51,933	5.0 km ² (1.9 sq mi)	10,387	[16]
11	Koekelberg	Koekelberg			1081	21,609	1.2 km ² (0.5 sq mi)	18,008	[17]
12	Molenbeek-Saint-Jean	Sint-Jans-Molenbeek			1080	96,629	5.9 km ² (2.3 sq mi)	16,378	[18]
13	Saint-Gilles	Sint-Gillis			1060	50,471	2.5 km ² (1.0 sq mi)	20,188	[19]
14	Saint-Josse-ten-Noode	Sint-Joost-ten-Noode			1210	27,115	1.1 km ² (0.4 sq mi)	24,650	[20]
15	Schaerbeek	Schaarbeek			1030	133,042	8.1 km ² (3.1 sq mi)	16,425	[21]
16	Uccle	Ukkel			1180	82,307	22.9 km ² (8.8 sq mi)	3,594	[22]
17	Watermael-Boitsfort	Watermaal-Bosvoorde			1170	24,871	12.9 km ² (5.0 sq mi)	1,928	[23]
18	Woluwe-Saint-Lambert	Sint-Lambrechts-Woluwe			1200	55,216	7.2 km ² (2.8 sq mi)	7,669	[24]
19	Woluwe-Saint-Pierre	Sint-Pieters-Woluwe			1150	41,217	8.9 km ² (3.4 sq mi)	4,631	[25]

Figure 2 - List of Brussels' Municipalities - Wikipedia [2]

	Municipalities	Postal Codes	Latitude	Longitude
0	Anderlecht	1070	50.8362	4.3145
1	Auderghem	1160	50.8167	4.4333
2	Berchem-Sainte-Agathe	1082	50.8657	4.2956
3	Bruxelles-Ville*	1000	50.8504	4.3488
4	Etterbeek	1040	50.8327	4.3884
5	Evere	1140	50.8744	4.3990
6	Forest	1190	50.8168	4.3278
7	Ganshoren	1083	50.8707	4.3153
8	Ixelles	1050	50.8333	4.3667
9	Jette	1090	50.8731	4.3342
10	Koekelberg	1081	50.8612	4.3314
11	Molenbeek-Saint-Jean	1080	50.8499	4.3125
12	Saint-Gilles	1060	50.8276	4.3439
13	Saint-Josse-ten-Noode	1210	50.8538	4.3728
14	Schaerbeek	1030	50.8694	4.3774
15	Uccle	1180	50.8022	4.3394
16	Watermael-Boitsfort	1170	50.8015	4.4144
17	Woluwe-Saint-Lambert	1200	50.8439	4.4291

Figure 2 - Municipality dataframe

3. Exploratory Analysis

After formatting the municipality dataframe, I created a function (*getNearbyVenues*) to query the venue data using the Foursquare API; I decided to limit the amount of venues to 50, but this can be changed as desired. The function extracted a total of 567 venues in all the 19 neighborhoods.

(567, 7)							
	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Anderlecht	50.8362	4.3145	Friture René	50.835846	4.311632	Belgian Restaurant
1	Anderlecht	50.8362	4.3145	Le Chapeau Blanc	50.835034	4.307790	Restaurant
2	Anderlecht	50.8362	4.3145	Snack Mirvan	50.835176	4.308543	Snack Place
3	Anderlecht	50.8362	4.3145	Erasmushuis / Maison d'Erasme (Erasmushuis)	50.836507	4.307860	History Museum
4	Anderlecht	50.8362	4.3145	Le Nil	50.836718	4.313628	Middle Eastern Restaurant
5	Anderlecht	50.8362	4.3145	Iberico	50.832146	4.311816	Gourmet Shop
6	Anderlecht	50.8362	4.3145	Proxy Delhaize	50.835256	4.310246	Grocery Store
7	Anderlecht	50.8362	4.3145	Carrefour Express	50.835387	4.310213	Convenience Store
8	Anderlecht	50.8362	4.3145	Panos	50.835544	4.310776	Sandwich Place

Figure 3 - Venue dataframe

Below you can see a summary of the total amount of venues pulled per neighborhood.

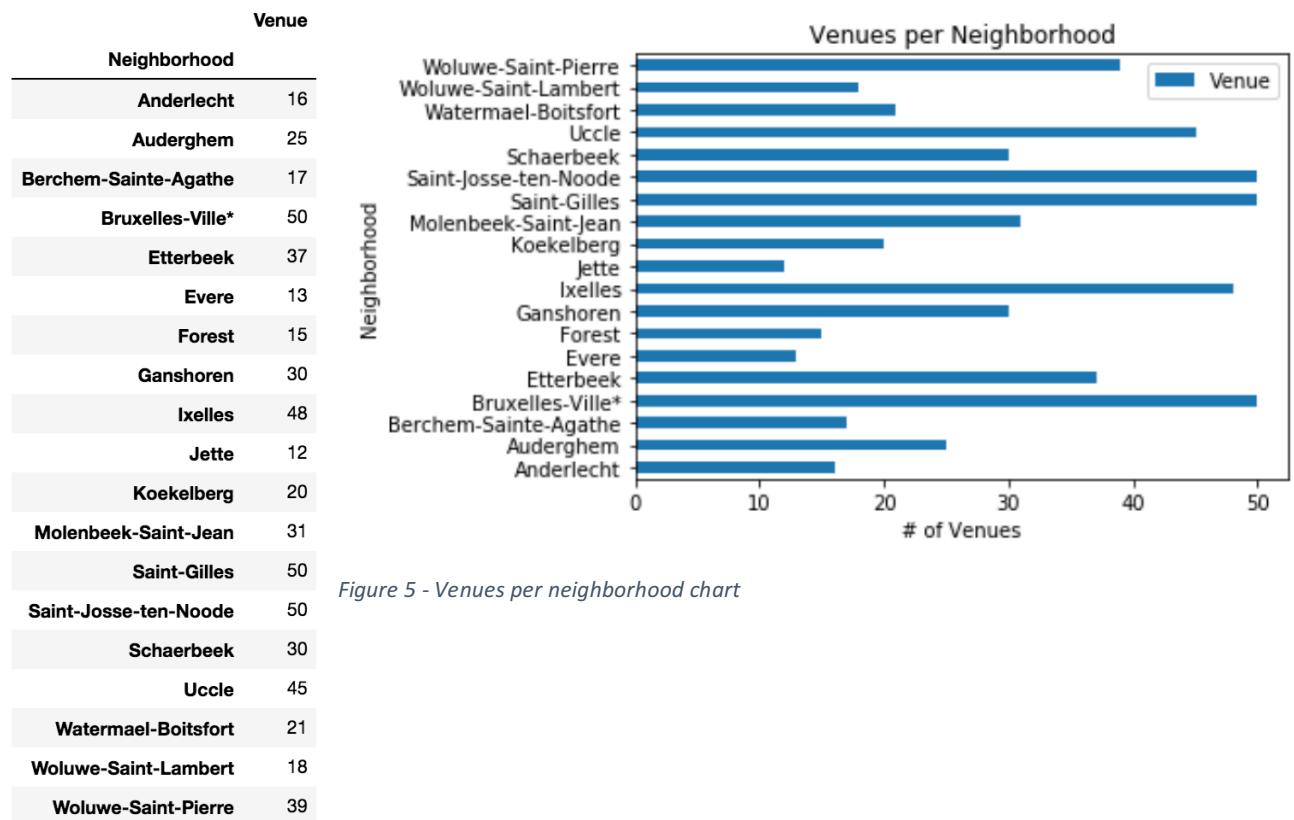


Figure 4 - Venues per neighborhood

Figure 4 - Venues per neighborhood

From the total of 567 venues, there are 156 unique venue categories, with Italian Restaurants and Supermarkets being the most common overall.

Venue Count	
Italian Restaurant	28
Supermarket	27
Plaza	26
Bar	26
Pizza Place	21
Snack Place	19
Sandwich Place	14
Bakery	14
Restaurant	14
French Restaurant	12

Figure 6 - Top 10 most common venue categories

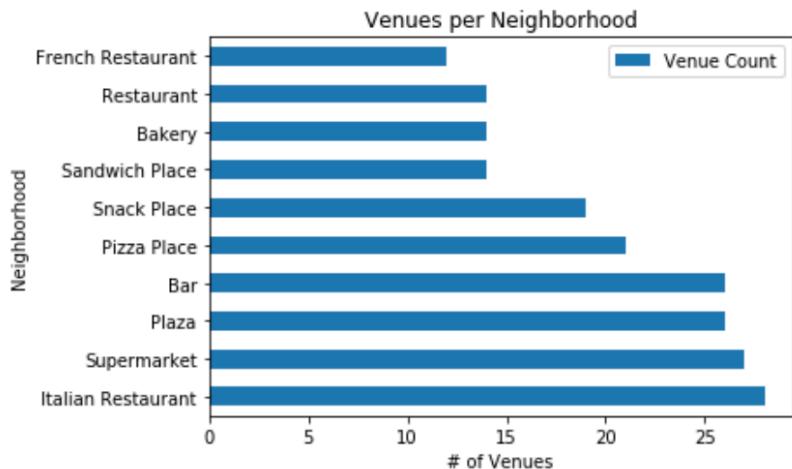


Figure 7 - Top 10 most common venue categories chart

After having pulled the venue data, I am interested in calculating the mean occurrence of each venue category in each neighborhood. For this, I dummy coded each venue into a venue category per neighborhood. The result is a dataframe with 567 rows and 156 columns. After dummy coding each venue, I grouped all the venues per neighborhood and calculated the mean frequency per venue category. The resulting dataframe has 19 rows and 156 columns, and displays the relative frequency of each venue category per neighborhood.

	Neighborhood	African Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Automotive Shop	BBQ Joint	Bagel Shop	Bakery	Bar	Beer Bar	Beer Store
0	Anderlecht	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	Auderghem	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.040000	0.040000	0.000000	0.000000
2	Berchem-Sainte-Agathe	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.058824	0.000000	0.000000	0.000000
3	Bruxelles-Ville*	0.000000	0.000000	0.000000	0.02	0.02	0.000000	0.000000	0.020000	0.000000	0.020000	0.160000	0.000000	0.000000
4	Etterbeek	0.000000	0.000000	0.000000	0.00	0.00	0.027027	0.000000	0.000000	0.000000	0.000000	0.081081	0.000000	0.000000
5	Evere	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.076923	0.000000	0.000000	0.000000
6	Forest	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.066667	0.000000	0.000000	0.000000
7	Ganshoren	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.066667	0.100000	0.000000	0.000000
8	Ixelles	0.020833	0.020833	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.041667	0.062500	0.020833	0.020833
9	Jette	0.000000	0.000000	0.083333	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
10	Koekelberg	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	Molenbeek-Saint-Jean	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.032258	0.000000	0.032258	0.000000	0.000000	0.000000
12	Saint-Gilles	0.020000	0.000000	0.000000	0.00	0.00	0.020000	0.000000	0.020000	0.020000	0.020000	0.060000	0.020000	0.000000
13	Saint-Josse-ten-Noode	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.020000	0.000000	0.000000
14	Schaerbeek	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.033333	0.000000	0.000000	0.033333	0.000000	0.000000	0.000000
15	Uccle	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.022222	0.044444	0.044444	0.000000	0.000000	0.000000
16	Watermael-Boitsfort	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.000000	0.047619	0.000000	0.000000	0.000000
17	Woluwe-Saint-Lambert	0.000000	0.000000	0.000000	0.00	0.00	0.055556	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
18	Woluwe-Saint-Pierre	0.000000	0.000000	0.000000	0.00	0.00	0.000000	0.000000	0.000000	0.025641	0.000000	0.000000	0.000000	0.000000

Figure 8 - Relative frequency of venue categories per neighborhood

With this dataframe, we can see the most common venue categories per neighborhood, therefore I proceeded to create another dataframe and sorted the top 10 most common venue categories per neighborhood.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Anderlecht	History Museum	Discount Store	Fish Market	Snack Place	Supermarket	Business Center	Sandwich Place	Belgian Restaurant	Theater	Middle Eastern Restaurant
1	Auderghem	Italian Restaurant	Sushi Restaurant	French Restaurant	Convenience Store	Middle Eastern Restaurant	Snack Place	Fast Food Restaurant	Salad Place	Restaurant	Cultural Center
2	Berchem-Sainte-Agathe	Greek Restaurant	Tennis Court	Sports Club	Bar	Burger Joint	Supermarket	Restaurant	Tailor Shop	Café	French Restaurant
3	Bruxelles-Ville*	Bar	Plaza	Seafood Restaurant	Fish & Chips Shop	Belgian Restaurant	Yoga Studio	Comfort Food Restaurant	Sandwich Place	Restaurant	Pub
4	Etterbeek	Snack Place	Plaza	Bar	Pizza Place	Supermarket	Sandwich Place	Cosmetics Shop	Kebab Restaurant	Diner	Department Store

Figure 9 - Venue categories sorted by occurrence

What we have just done is just some exploratory analysis of the data that we pulled using the Foursquare API, now let's actually start ranking the neighborhoods according to my own personal preferences. First, I will create a scorecard, where I will specify my top 10 preferred venue categories and rank them from 1 to 10.

	venue	rating	After creating this dataframe, we have to create another dataframe with the relative frequency of these venues per neighborhood. It is important to note that if you are interested in recreating this analysis, but tailored to your own personal preferences, you need to modify this dataframe and also the one with the relative frequencies of the venue categories.								
0	Supermarket	10.0	Neighborhood	Anderlecht	Auderghem	Berchem-Sainte-Agathe	Bruxelles-Ville*	Etterbeek	Evere	Forest	Ganshoren
1	Gym	10.0	Supermarket	0.0625	0.04	0.0588235	0	0.0540541	0	0.0666667	0.0666667
2	Metro Station	8.0	Gym	0	0	0.0588235	0	0.027027	0	0	0
3	Bar	6.5	Metro Station	0.0625	0	0	0	0	0	0	0
4	Breakfast Spot	6.5	Bar	0	0.04	0.0588235	0.16	0.0810811	0	0.0666667	0.1
5	Tram Station	8.0	Breakfast Spot	0	0	0	0.02	0	0	0	0
6	Sushi Restaurant	8.0	Tram Station	0	0.04	0	0	0	0	0.133333	0
7	Snack Place	7.0	Sushi Restaurant	0	0.08	0	0.02	0	0	0	0
8	Pharmacy	8.0	Snack Place	0.0625	0.04	0.0588235	0	0.0810811	0.0769231	0	0
9	Grocery Store	10.0	Pharmacy	0	0	0	0	0	0	0	0.0333333
			Grocery Store	0.0625	0.04	0	0	0	0	0	0

Figure 10 - User Input

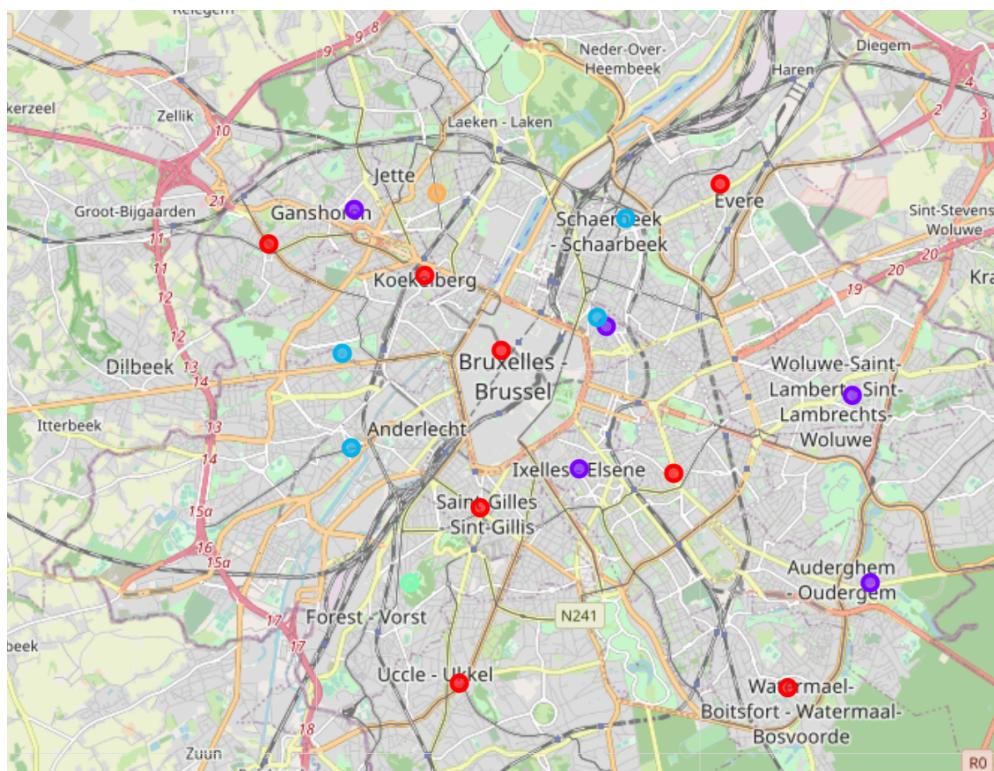
Figure 11 - Relative frequencies of preferred venue categories

After formatting these two dataframes, I calculate so to say the weighted average for each neighborhood by multiplying the score I assigned to the venue times the frequency of that venue in each neighborhood. The resulting scores are displayed in another dataframe, with a grand totals column for the total score per neighborhood.

venue	Supermarket	Gym	Metro Station	Bar	Breakfast Spot	Tram Station	Sushi Restaurant	Snack Place	Pharmacy	Grocery Store	Total score
Neighborhood											
Molenbeek-Saint-Jean	0.967742	0.645161	0.00	0.000000	0.000000	0.258065	0.000000	0.677419	0.258065	0.322581	3.129032
Schaerbeek	1.666667	0.000000	0.00	0.000000	0.000000	0.533333	0.000000	0.233333	0.266667	0.000000	2.700000
Jette	2.500000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	2.500000
Auderghem	0.400000	0.000000	0.00	0.260000	0.000000	0.320000	0.640000	0.280000	0.000000	0.400000	2.300000
Anderlecht	0.625000	0.000000	0.50	0.000000	0.000000	0.000000	0.000000	0.437500	0.000000	0.625000	2.187500
Forest	0.666667	0.000000	0.00	0.433333	0.000000	1.066667	0.000000	0.000000	0.000000	0.000000	2.166667
Woluwe-Saint-Lambert	1.111111	0.000000	0.00	0.000000	0.000000	0.000000	0.444444	0.000000	0.444444	0.000000	2.000000
Berchem-Sainte-Agathe	0.588235	0.588235	0.00	0.382353	0.000000	0.000000	0.000000	0.411765	0.000000	0.000000	1.970588
Etterbeek	0.540541	0.270270	0.00	0.527027	0.000000	0.000000	0.000000	0.567568	0.000000	0.000000	1.905405
Ganshoren	0.666667	0.000000	0.00	0.650000	0.000000	0.000000	0.000000	0.000000	0.266667	0.000000	1.583333
Koekelberg	0.000000	1.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.350000	0.000000	0.000000	1.350000
Bruxelles-Ville*	0.000000	0.000000	0.00	1.040000	0.130000	0.000000	0.160000	0.000000	0.000000	0.000000	1.330000
Uccle	0.666667	0.000000	0.00	0.288889	0.144444	0.177778	0.000000	0.000000	0.000000	0.000000	1.277778
Saint-Josse-ten-Noode	0.400000	0.000000	0.16	0.130000	0.000000	0.000000	0.000000	0.560000	0.000000	0.000000	1.250000
Ixelles	0.000000	0.000000	0.00	0.406250	0.135417	0.000000	0.166667	0.000000	0.000000	0.208333	0.916667
Woluwe-Saint-Pierre	0.256410	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.538462	0.000000	0.000000	0.794872
Evere	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.538462	0.000000	0.000000	0.538462
Saint-Gilles	0.000000	0.000000	0.00	0.390000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.390000
Watermael-Boitsfort	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Figure 12 - Neighborhood ranking

From the dataframe above we see that Molenbeek-Saint-Jean is the neighborhood that best fits my personal interests, but it can also be the case that for work purposes, living in Molenbeek is not actually a convenient idea. Therefore, the next part of the analysis is to cluster all of the 19 neighborhoods in order to see which neighborhoods are similar to each other, and thereby providing some alternatives to your top ranked neighborhoods. I used a KMeans algorithm in order to cluster each neighborhood into 5 clusters, determined by how similar there top 10 venue categories are.



From the clustered map, we see that there are many alternatives to living in Molenbeek, in case there is any reason not to live there.

4. Conclusions

During this analysis, it was possible to rank the neighborhoods in Brussels according to how well the fit my personal preferences when it comes to the availability of services. In addition, during the analysis, the top 10 most common venue categories were calculated for a given neighborhood, as well as their relative frequency. In order to make the results of this analysis more flexible, it was also decided to cluster all the 19 municipalities of Brussels, since the results from the neighborhood ranking may be too strict and could often be inconvenient. In my own case, Molenbeek-Saint-Jean was the top-ranking neighborhood, however, assuming that I am working in Evere, Schaerbeek seems to be a similar option and more convenient because of its location.

References:

- [1] Brussels – Wikipedia
- [2] List of municipalities in the Brussels Capital Region – Wikipedia