

# The Battle of Neighbourhoods

## Applied Data Science Capstone Project

Katja Szalay - May 2021

---

### 1. Introduction

This project primarily helps people who have to move to Brussels, Belgium for professional reasons to find a suitable neighbourhood to live in. The analysis is intended to help future expats to find a suitable area more easily and to find a similar neighbourhood as they were used to before. For this purpose, a comparison is made between the different neighbourhoods in Brussels.

#### Background

Brussels grew from a small rural settlement on the river Senne to become an important city-region in Europe. Since the end of the Second World War, it has been a major centre for international politics and home to numerous international organisations, politicians, diplomats and civil servants. Brussels is the de facto capital of the European Union, as it hosts a number of principal EU institution. Its name is sometimes used metonymically to describe the EU and its institutions. The secretariat of the Benelux and headquarters of NATO are also located in Brussels. As the economic capital of Belgium and one of the top financial centres of Western Europe with Euronext Brussels, it is classified as an Alpha global city.

#### Problem

This circumstance includes the fact that a large number of employees, also called expats, who come from all countries in Europe and worldwide, have to find accommodation in Brussels on a temporary, but also on a longer-term basis and look for suitable residential areas to feel at home.

The analysis of neighbourhoods in Brussels is intended to help future expats find a suitable area to live in more easily and to help people moving from Brussels to Berlin and vice versa to find a similar neighbourhood as they were used to before.

Berlin is one of the capitals with the highest number of established diplomatic representations. Relocations between these cities take place frequently.

### 2. Data

In this project, three different datasets will be used to solve the problem – Monitoring of the Neighbourhoods in Brussels, Brussels Recorded Crime and Foursquare API. After scraping them from original and reliable sources, they will be wrangled and cleansed for further analysis into more useful forms.

## Neighbourhood Monitoring Brussels ([monitoringdesquartiers.brussels](http://monitoringdesquartiers.brussels))

Code	Territoire	Densité de po -2020-	Densité de bu -2018-	Part des bât -1997-	Part des surt -2006-	Taux de mobi -2001-2006-	Part des isolé -2019-	Part des isolé -2019-	Part des isolé -2019-	Part des coup -2019-	Part des coup -2019-	Part de l'Euro -2019-	Part des nou -2015-	Revenu impos -2012-	Part de la pop -2001-	Part des mén -2001-	Part des m -2001-
1	Grand Place	8896,96	535254,58	15,43	94,05	119,11	20,25	50,18	6,79	7,46	12,66	32,51	4,34	4	60,89	89,18	9
2	Dansert	17105,34	227800,38	17,84	89,49	101,46	12,77	45,17	7,98	14,44	13,50	19,67	4,56	2	60,99	88,62	10
3	Béguinage - C	117973,30	379895,97	17,40	86,41	103,49	15,76	42,16	6,98	15,45	13,50	19,47	3,87	2	100,00	90,77	8
4	Martyrs	7075,01	908364,20	28,47	95,35	108,55	19,73	47,02	7,20	9,12	12,90	20,48	5,15	3	77,93	95,24	4
5	Notre-Dame-è	8855,44	1711022,91	22,53	90,68	122,91	20,51	46,86	5,59	10,29	11,49	25,04	6,08	3	100,00	92,24	6
6	Quartier Roye	458,18	851468,97	33,93	67,00	108,89	12,78	36,67	4,44	17,78	17,78	33,94	3,94	5	99,99	86,36	11
7	Cette	8894,45	345664,07	15,29	97,00	109,49	19,45	54,20	10,27	10,45	12,45	34,70	4,70	4	87,94	86,45	10

Shown above is a section of the table of the Neighbourhood Monitoring in Brussels. The dataset includes a list of all neighbourhoods in the Brussels-Capital Region and the related and required data:

- Municipality and Postal Code
- Population data like population density
- Building structure like share of high-rise buildings, office density
- Income structure
- Resident structure like family share, single person share
- Environment like access to green spaces
- Mobility like access to public transport

The dataset shown above will be simplified and prepared for further analysis.

## Brussels Recorded Crime ([stat.policefederale.be](http://stat.policefederale.be))

The second data set contains the crime statistics of the individual municipalities of Brussels. The crime statistics will be transferred to the individual neighbourhoods.

Nombre total des délits	2000	2012	2013	2014	2015	2016	2017	2018	2019	2020 (*)
▼ National	1.002.087	1.040.232	997.802	977.506	921.805	893.873	883.834	891.441	898.057	479.775
▼ Région de Bruxelles-Capitale	158.064	178.347	167.119	159.975	146.937	142.463	145.917	149.903	158.452	78.276
▼ Bruxelles-Capitale	158.064	178.347	167.119	159.975	146.937	142.463	145.917	149.903	158.452	78.276
Bruxelles (Bruxelles-Capitale)	158.064	178.347	167.119	159.975	146.937	142.463	145.917	149.903	158.452	78.276
► BRUXELLES CAPITALE IXLLES	60.567	71.975	67.054	64.372	55.702	52.609	55.535	59.248	64.964	29.932
► BRUXELLES-OUEST	19.339	20.378	19.584	19.364	19.504	20.051	18.727	18.858	19.043	12.130
Pas de ZP attribuée - Eurostar	1.976	821	265	0	0	0	1	1	0	0
► ZP POLBRUNO	21.821	21.496	21.576	20.454	19.659	19.179	19.423	20.342	20.483	9.578
► UCLE/W-B/AUDERGHEM	13.481	14.644	13.470	13.225	11.402	11.067	11.101	11.114	11.724	5.669
► ZONE MONTGOMERY	12.372	12.836	12.498	12.387	11.452	11.422	11.666	10.930	11.261	6.122
► ZONE MIDI	28.508	36.197	32.672	30.173	29.218	28.135	29.464	29.410	30.977	14.845

The crime statistics are transferred to the individual neighbourhoods. For further analysis, crimes are not differentiated by category and are calculated by proportion of the population in the municipality. An average of the last years will be calculated.

## Foursquare API

Finally, Foursquare API will be used to call the top 10 popular venues in each neighbourhood. This will be done by the “explore” function of requesting URL. The data will be looking like the following section of the dataset.

	BoroughName	Borough Latti	Borough Latti	Venue	Venue Lattitu	Venue Longitu	venue Category		
0	Grand Place	51.4253	0.4587	Grand Place	51.4253	0.4587	Place		
1	Grand Place	51.4254	0.4588	Restaurants	51.4254	0.4588	Restaurants		
2	Grand Place	51.4255	0.4589	Cafe	51.4255	0.4589	Cafe		
3	Grand Place	51.4256	0.4590	Bar	51.4256	0.4590	Bar		
4	Grand Place	51.4257	0.4591	Hotel	51.4257	0.4591	Hotel		

### 3. Methodology

Three data sets were used in this project, with monitoring\_data and foursquare\_data used for cluster analysis.

#### Exploratory Analysis

To better understand and interpret the data, some visualizations are shown here. All datasets have been simplified and prepared for further analysis.

#### Neighbourhood Monitoring Brussels (monitoringdesquartiers.brussels)

The figure below shows the adjusted table with the monitoring data.

	Code	Commune	Latitude	Longitude	Neighbourhood	Postal Code	Population	Rent	Mobility	Tax	Singles	Families	European	Skyscrapers	Apartments	Houses	GreenSpace	Metro
0	1	Bruxelles	50.846714	4.352514	Grand Place	1000	8887	947	119	18275	90	7	37	15	89	9	61	98
1	2	Bruxelles	50.850453	4.346755	Dansaert	1000	17105	711	101	18277	79	14	24	18	89	11	61	94
2	3	Bruxelles	50.855866	4.350933	Béguinage - Dixmude	1000	17973	826	103	18856	78	15	23	17	91	9	100	100
3	4	Bruxelles	50.851834	4.356594	Martyrs	1000	7075	663	109	19626	87	9	26	28	95	4	78	100
4	5	Bruxelles	50.850006	4.363218	Notre-Dame-aux-Neiges	1000	8855	663	123	20667	84	10	31	23	92	7	100	100

The dataframe includes the following variables:

- Population = Number of inhabitants per neighborhood,
- Rent = Average value of the rents in €,
- Mobility = Frequency of residential moves (abs.),
- Tax = average taxable salary in €,
- Singles = Households without children,
- Families = Households with children,
- European = Proportion of other European residents (except Belgium),
- Skyscrapers = Proportion of buildings with 5 or more floors,
- Apartments = Percentage of apartments (not houses),
- Houses = Percentage of houses (not apartments),
- GreenSpace = Percentage of the population near a publicly accessible green space,
- Metro = Percentage of the population near a metro or a Tram stop.

	Population	Rent	Mobility	Tax	Singles	Families	European	Skyscrapers	Apartments	Houses	GreenSpace	Metro
count	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000	118.000000
mean	12429.644068	769.322034	78.898305	20881.940678	72.194915	24.364407	23.474576	5.983051	68.601695	30.635593	81.983051	45.415254
std	7013.398339	118.591656	15.431896	3588.173914	9.924724	7.355056	10.426554	6.359572	17.137523	17.158221	18.095182	37.513605
min	458.000000	593.000000	49.000000	15104.000000	47.000000	7.000000	8.000000	0.000000	19.000000	4.000000	19.000000	0.000000
25%	6814.500000	673.500000	70.000000	18369.750000	65.000000	19.250000	15.000000	2.000000	62.000000	18.250000	72.000000	2.250000
50%	11759.500000	744.500000	76.500000	20434.500000	73.000000	26.000000	21.500000	4.000000	72.000000	27.000000	87.000000	46.500000
75%	17420.250000	840.750000	86.750000	23169.500000	79.750000	29.750000	29.750000	8.000000	80.750000	37.000000	96.750000	82.750000
max	37531.000000	1113.000000	124.000000	30879.000000	93.000000	45.000000	58.000000	35.000000	95.000000	80.000000	100.000000	100.000000

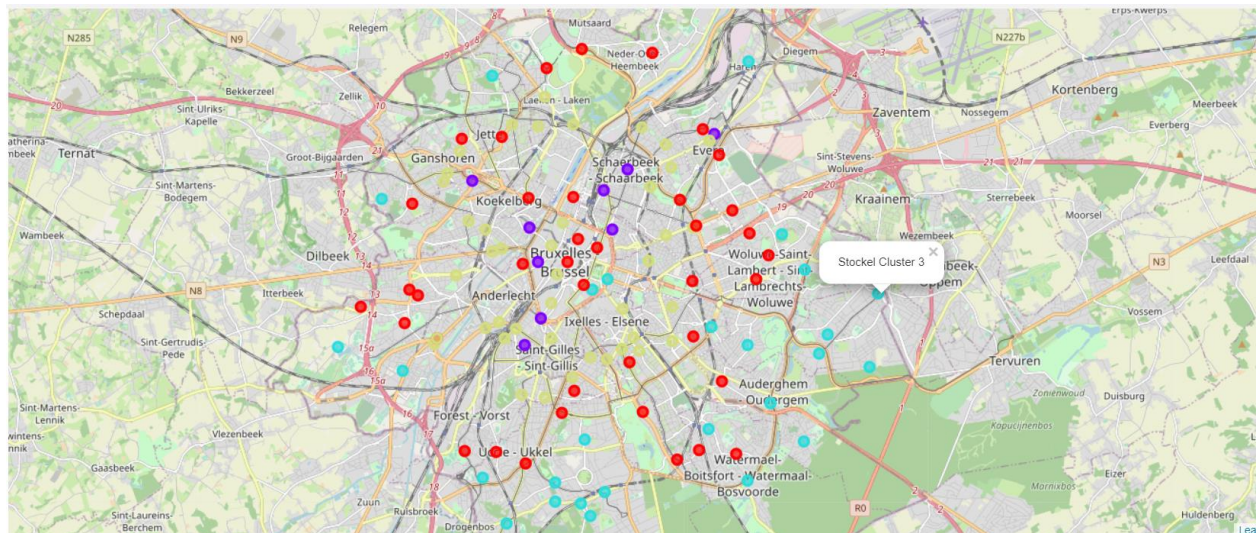
The figure above shows some basic statistical details of the monitoring data.

After some descriptive statistics k-means clustering was conducted in order to group the neighbourhoods according to the variables shown above. The analysis revealed four clusters. The figure below shows the map of Brussels and the neighbourhoods. The colour dots represent the neighbourhoods and every cluster is marked with another colour.

Each cluster is given a name that refers to the main character of the environment:

- Cluster 0 – yellow = "Below average price and high population area"
- Cluster 1 – red = "Higher-priced and well-situated area"
- Cluster 2 –violet = "Affordable and very densely populated area"
- Cluster 3 – green = "High-priced and upscale area"

The main characters of the neighbourhoods is the population density, the average rent paid, the purchase power, the structure of the buildings and the housing-properties.



For further analysis, only neighborhoods from **Cluster 1 and Cluster 2** are considered. This is based on a subjective decision, as it can be assumed that future expats would like to live in a more upscale neighborhood.

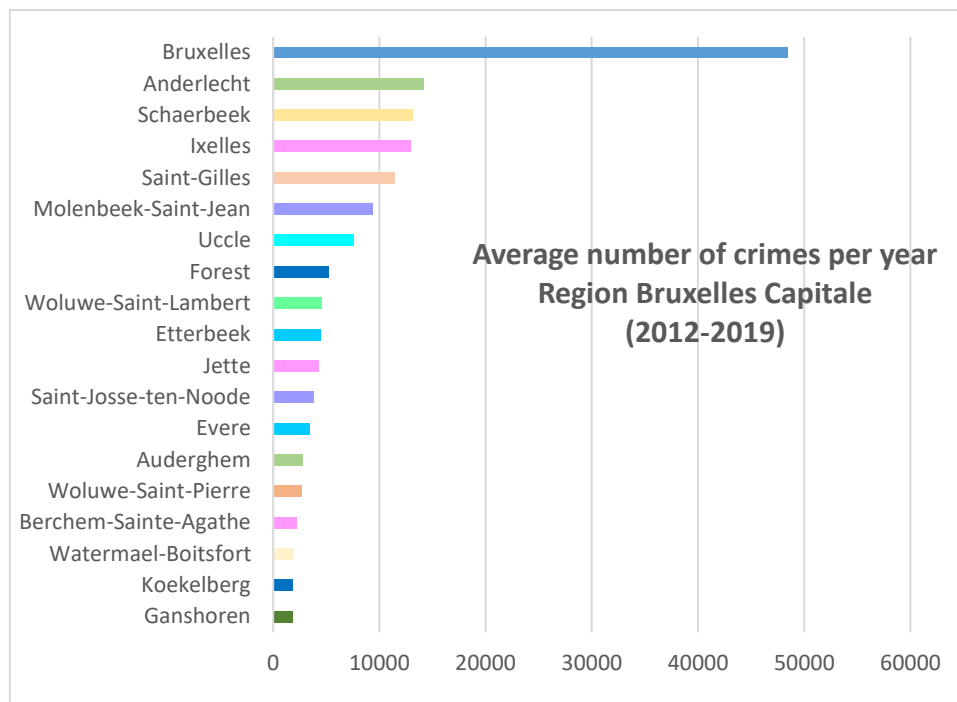
### Brussels Recorded Crime (stat.policefederale.be)

The figure below shows Statistics of Total Number of Crimes by Year and Location in Brussels. The data has been processed: for simplicity, the crimes of the last eight years (2012-2019) have been averaged and divided by the number of inhabitants of each municipality, giving a value that indicates the average crimes per 1000 per month.

Code	Quartier	Commune	Postal Code	Total population (Number of inhabitants)	Crimes per year average	Crimes per month (average)	Crimes per 1000 inhabitants per month on average
1	Grand Place	Bruxelles	1000	185103	48456.375	4038.03125	21.81505
2	Dansaert	Bruxelles	1000	185103	48456.375	4038.03125	21.81505
3	Béguinage - Dixmude	Bruxelles	1000	185103	48456.375	4038.03125	21.81505
4	Martyrs	Bruxelles	1000	185103	48456.375	4038.03125	21.81505
5	Notre-Dame-aux-Neiges	Bruxelles	1000	185103	48456.375	4038.03125	21.81505

The number of crimes was then sorted according to their level - in five levels from very high to very low. The number of crimes relating to the communities have been transferred to the neighbourhoods. For the further analysis, only neighbourhoods with at least the crime level "ok" are considered. This level corresponds to a number of offences that is below the average. Finally, the two relevant tables - monitoring\_data and crime\_data - have been merged (the figure below shows the table).

Code	Commune	Neighbourhood	Postal Code	monitoring level	Crime level	
0	1	Bruxelles	Grand Place	1000	2 - Higher-priced and well-situated area	very high
1	2	Bruxelles	Dansaert	1000	3 - Below average price and high population area	very high
2	3	Bruxelles	Béguinage - Dixmude	1000	3 - Below average price and high population area	very high



The chart above shows the average number of crimes per year in the Brussels Capital Region.

## Foursquare API

Finally, Foursquare was used to call the top 50 popular venues in each neighbourhood. This was done using the 'explore' function of requesting URL. As the first step, one hot encoding was conducted to give binary variables to each venue category. The figure below shows the table.

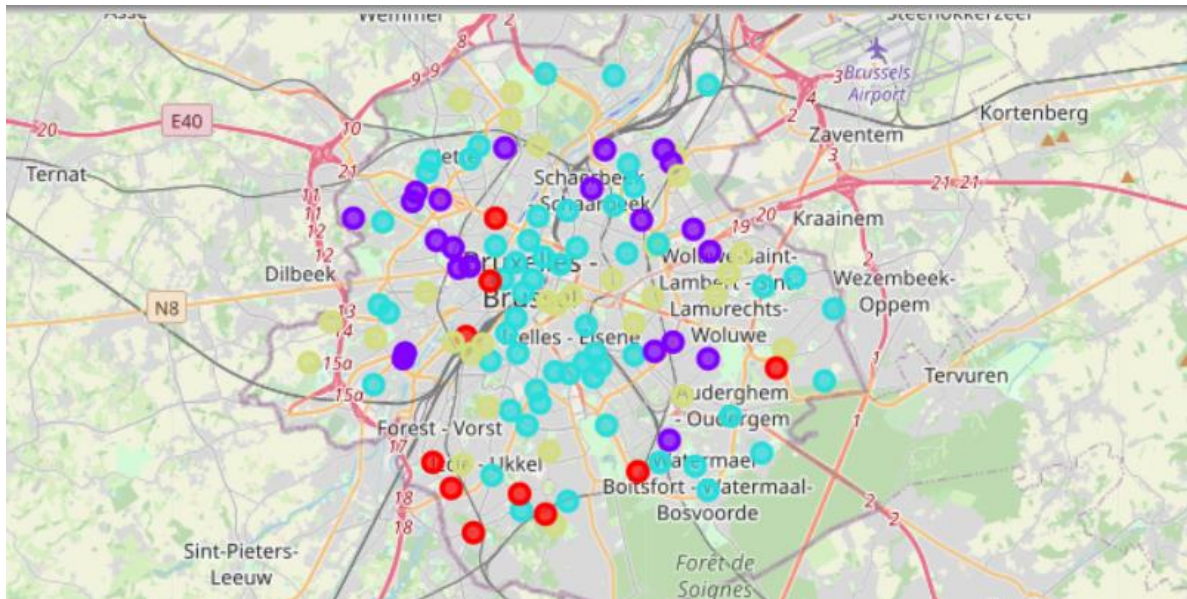
	Neighborhood	African Restaurant	Amphitheater	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Auto Dealership	BBQ Joint	...	Train Station	Tram Station	Trattoria/Osteria	Turkish Restaurant	Veget / v Resta
0	Altitude 100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
1	Anderlecht - Centre - Wayez	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.1	0.0	0.0	0.0
2	Anneessens	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
3	Auderghem centre	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0
4	Avenue	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0

The. The data was grouped by borough names to find out how many venues of each category exist in the boroughs within the top 50. Based on the frequency, a list has been created, as shown in the figure below. The first till 10<sup>th</sup> common venue of each neighbourhood is presented.

Code	Commune	Neighbourhood	Latitude	Longitude	Cluster Labels	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
1	Bruxelles	Grand Place	50.846714	4.352514	3	Chocolate Shop	Concert Hall	Toy / Game Store	Dessert Shop	Plaza	Cheese Shop	Shopping Mall	Hotel	Auto Dealership
2	Bruxelles	Dansart	50.850453	4.346755	3	Plaza	Bakery	French Restaurant	Sushi Restaurant	Fish & Chips Shop	Bookstore	Bar	Seafood Restaurant	Moroccan Restaurant
3	Bruxelles	Béguinage - Dixmude	50.855666	4.350933	3	Yoga Studio	Restaurant	Bookstore	Butcher	Cultural Center	Ethiopian Restaurant	Grocery Store	Organic Grocery	Plaza
4	Bruxelles	Martyrs	50.851834	4.356594	3	Department Store	Clothing Store	Bookstore	Cosmetics Shop	Coffee Shop	Sporting Goods Shop	Seafood Restaurant	Event Service	Fish Market
5	Bruxelles	Notre-Dame-aux-Neiges	50.850006	4.363218	3	Ice Cream Shop	Bar	Gastropub	Deli / Bodega	Concert Hall	Coffee Shop	Sandwich Place	Smoke Shop	Turkish Restaurant



Then, the neighbourhoods were clustered in four groups, based on the frequency of similarity of the venues, which will be presented in the next figure below.



The coloured dots also represent each neighbourhood of Brussels, whereby the colours represent the different characteristics.

Each cluster is given a name according to the characteristics that can be derived from the popular venues:

- Cluster 0 = yellow = **"Busy and touristic"** - everything is nearby (Restaurants, Pubs, Supermarkets, Sports, Culture) - these are the most hectic areas,
- Cluster 1 = red = **"Green and Sports"** - the main focus is on a green environment as well as the possibility to do sports,
- Cluster 2 = violet = **"Foods and Quiet"** - Restaurants, Take aways, Convenient Shops, but not too busy,
- Cluster 3 = green = **"Lively"** - everything is nearby, but it is not too busy and less touristic.

For further analysis, Cluster 0 will be excluded, based on subjective opinion. It is assumed that future expats will not want to live in an overly busy area.

In a final step, the Foursquare data was merged with the monitoring-and-crime data to create a concise table containing all neighbourhoods.

Code	Latitude	Longitude	Commune	Neighbourhood	Postal Code	monitoring level	Crime level	foursquare level
1	50.846714	4.352514	Bruxelles	Grand Place	1000	2 - Higher-priced and well-situated area	very high	Busy and touristic
2	50.850453	4.346755	Bruxelles	Dansaert	1000	3 - Below average price and high population area	very high	Busy and touristic
3	50.855666	4.350933	Bruxelles	Béguinage - Dixmude	1000	3 - Below average price and high population area	very high	Busy and touristic
4	50.851834	4.356594	Bruxelles	Martyrs	1000	2 - Higher-priced and well-situated area	very high	Busy and touristic
5	50.850006	4.363218	Bruxelles	Notre-Dame-aux-Neiges	1000	2 - Higher-priced and well-situated area	very high	Busy and touristic

From this table, all unnecessary neighbourhoods can be removed, so that a table is created that contains only the relevant, filtered neighbourhoods (final results).

## 4. Results and Recommendation

The following selection of suitable neighborhoods is simplified and shaped according to subjective preferences.

**a) Monitoring Level** The monitoring level describes neighborhoods based on facts such as population density, rent levels, housing characteristics, metro and green space access.

- 1 - High-priced and upscale area
- 2 - Higher-priced and well-situated area
- 3 - Below average price and high population area
- 4 - Affordable and very densely populated area

For further analysis, only **Levels 1 and 2** are relevant, as these areas have higher rents but lower population density, are family-oriented, well connected, and offer access to green space.

**b) Crime Level** The Crime Level describes the crimes of the last eight years (2012-2019), averaged and divided by the number of inhabitants of each municipality, giving a value that indicates **the average crimes per 1000 habitants per month**. The crime level ranges from "Very high" to "best", where "best" represent the least crimes:

- very high
- high
- relatively high
- ok
- good
- best

For further analysis, only the **crime levels ok, good and best** are relevant. These are areas where less than 10 crimes per 1000 habitants per month occurs.

**c) Foursquare Level** The Foursquare Level based on the top 10 venues nearby every neighbourhood. Every cluster represents another main focus and characteristic of the neighbourhood:

- **"Busy and touristic"** - everything is nearby (Restaurants, Pubs, Supermarkets, Sports, Culture - these are the most hectic areas
- **"Green and Sports"** - the main focus is on a green environment as well as the possibility to do sports
- **"Foods and Quiet"** - Restaurants, Take aways, Convenient Shops, but not too busy
- **"Lively"** - everything is nearby, but it is not too busy and less touristic

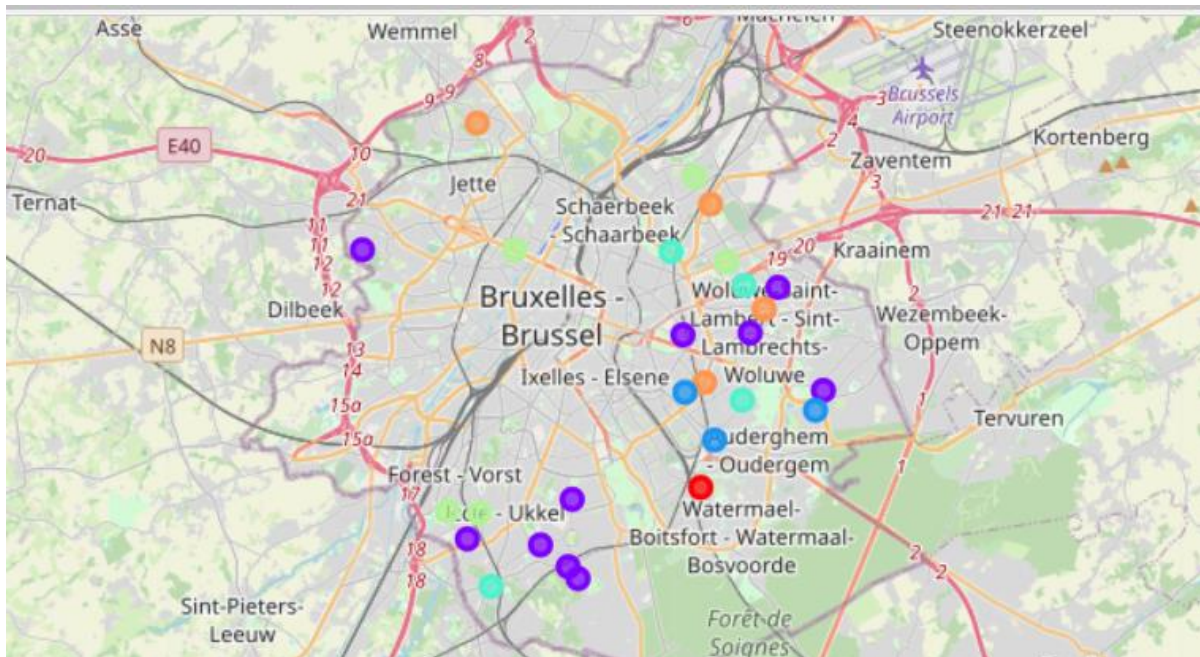
For further analysis, only three levels are relevant - **"Green and Sports"**, **"Foods and Quiet"** and **"Lively"**. The **"Busy and touristic"** area will be excluded for subjective reason.

Based on this selection a new data frame is created (figure below).

Code	Latitude	Longitude	Commune	Neighbourhood	Postal Code	monitoring level	Crime level	foursquare level
18	50.861169	4.338667	Molenbeek-Saint-Jean	Quartier Maritime	1080	2 - Higher-priced and well-situated area	ok	Green and Sports
30	50.842452	4.397510	Etterbeek	Porte Tervueren	1040	2 - Higher-priced and well-situated area	ok	Lively
31	50.831890	4.404529	Etterbeek	Saint-Michel	1040	1 - High-priced and upscale area	ok	Foods and Quiet
65	50.829735	4.398131	Molenbeek-Saint-Jean	Korenbeek	1080	2 - Higher-priced and well-situated area	ok	Foods and Quiet
66	50.861080	4.285760	Berchem-Sainte-Agathe	Potaarde	1082	1 - High-priced and upscale area	good	Foods and Quiet
73	50.888936	4.325685	Jette	Heymbosch - AZ-Jette	1090	1 - High-priced and upscale area	good	Lively
79	50.876910	4.401511	Evere	Paix	1140	2 - Higher-priced and well-situated area	good	Foods and Quiet

## 5. Conclusion

**A total of 27 suitable neighborhoods have been identified.**



There are six different clusters regarding a suitable neighbourhood in Brussels:

- Red: High-priced and upscale area - Foods and Quiet,
- Violet: High-priced and upscale area - Green and Sports,
- Blue: High-priced and upscale area - Lively,
- Green: Higher-priced and well-situated area - Foods and Quiet,
- Light Green: Higher-priced and well-situated area - Green and Sports,
- Orange: Higher-priced and well-situated area - Lively.

Through the analysis, 27 neighborhoods have been identified that are suitable for future expats to live in. These 27 neighborhoods can be divided into six different groups, each with a different focus. For example, the first group contains neighborhoods that are very expensive, upscale, safe, and rather quiet, while still having plenty of dining options. So everyone can subjectively decide what they prefer.

## 6. Discussion – Recommendation for future study

- More explicitly delineate and demarcate neighbourhoods. They are very close to each other. There can be overlaps and distortions.
- Decisions have been made subjectively. This should be taken into account.
- Simplified presentation due to the scope.
- Discriminatory power of the clusters not always given.
- It was a lot of fun to delve into this project!