The Battle of the Neighborhoods - Report

Helping people relocate in Brussels

Charles-Emmanuel de Lannoy

04/2020

1. Introduction

1.1. Background

Brussels in Belgium is the capital of European Union. Therefore all European Countries members have delegates in this city to represent them.

Brussels is made of several municipalities, each of them having its own characteristics. Some are more adapted to Families, other more to single young people.

There are 19 municipalities in Brussels. Some have a really active life (lot of restaurants, shops, bars,…) and some are more quiet and more adapted to families with childrens. The centre of the city is more for Business and is principally made of Offices while the periphery is more used for living and therefore is made from home and appartments.

In order to install the European delegates and their families (if any) according to their needs, European commission ask me to find a way to determine the best place to live in Brussels for each member delegate.

1.2. Problem

Data can be used to categorise each Municipalities and their characteristics and a classification approach can be used to determine which municipality would best fit the wishes of European delegates.

Municipalities should be studied according to their facilities. To allow identification of municipality that would allow classification that would suit both single and families, people who look for calm or intense city life, we will compare municipalities in terms of proximity of restaurants, public transport stations, parks, shops, population density, average age of population and schools.

The objective is to find the best municipality in the city in order to optimize the satisfaction in term of area.

2. Data

The 19 municipalities of Brussels will be studied to identify those who can suit a particular European delegate.

A dataset containing all the data from Brussels Municipalities will be built.

PostalCode	Municipality	Latitude	Longitude	size(km^2)	population	Population_density	population_average_age
1000	1000 Bruxelles	50.8465573	4.351697	32.61	179797	5514	41.86
1030	Schaerbeek	50.8676041	4.3737121	8.14	131547	16161	32.99
1040	Etterbeek	50.8361447	4.3861737	3.15	48008	15241	35.75

Source are

For the list of municipalities:

https://fr.wikipedia.org/wiki/Liste_des_communes_de_la_r%C3%A9gion_de_Bruxelles-Capitale

For statistics about population:

https://bestat.statbel.fgov.be/bestat/crosstable.xhtml?view=47d4f3be-0523-4a03-b062-819257075fff

Latitude and longitude are obtained from Geopy & Nominatim.

For each municipalities we will use foursquare to find the number of restaurants, public transport stations, parks, shops and schools at proximity.

A radius of 1000m will be applied for restaurants, public transport stations and shops as it would be important to have them close to the location. For parks and schools, a radius of 2000m will be applied as it would be less important to have them quite close to the location.

Resulting data set will be in the following form

	PostalCode	Municipality	Latitude	Longitude	size(km^2)	population	Population_density	population_average_age	restaurants	schools	public_transport_station	shops	parks
0	1000	1000 Bruxelles	50.84656	4.351697	32.61	179797	5514	41.86	30	10	18	30	2
1	1030	Schaerbeek	50.8676	4.373712	8.14	131547	16161	32.99	17	6	7	10	3
2	1040	Etterbeek	50.83615	4.386174	3.15	48008	15241	35.75	25	6	8	30	2

3. Methodology

3.1. Business understanding

Our goal is to find a way to determine the best place to live in Brussels for each member delegate according to its preferences.

The following criterions will be used to represent the preferences: the number of restaurants, public transport stations, parks, shops and schools at proximity in the municipality.

3.2. Analytic approach

Brussels has a total of 19 municipalities, each with its own characteristics mostly obtained from Foursquare.

We first start with a visual analysis of the data by creating bins for each of the continuous variables and look at their Map to draft a description of each municipalities.

Then we use a decision tree to set a way to decide which municipality would best fit the expectation of EU delegates.

We conclude by comparing decision tree results and first description we visually made.

3.3. Exploratory data analysis

Brussels geographical data were obtained combining Wikipedia data with Belgian statistics and use Geopy to obtain the geographical coordinates. We then use Folium to generate a Map of Brussels and its municipalities. Wikipedia and Belgian statistic data give the following

	PostalCode	Municipality	Latitude	Longitude	size(km^2)	population	Population_density	population_average_age
0	1000	1000 Bruxelles	50.846557	4.351697	32.61	179797	5514	41.86
1	1030	Schaerbeek	50.867604	4.373712	8.14	131547	16161	32.99
2	1040	Etterbeek	50.836145	4.386174	3.15	48008	15241	35.75
3	1050	Ixelles	50.833114	4.366828	6.34	86675	13671	40.95
4	1060	Saint-Gilles	50.826741	4.345668	2.52	49715	19728	34.82
5	1070	Anderlecht	50.839098	4.329653	17.74	118920	6938	37.05
6	1080	Molenbeek-Saint-Jean	50.854596	4.338636	5.89	96501	16384	36.15
7	1081	Koekelberg	50.860604	4.331550	1.17	21961	18770	38.80
8	1082	Berchem-Sainte-Agathe	50.864923	4.294673	2.95	25195	8541	39.05
9	1083	Ganshoren	50.870327	4.307798	2.64	24817	9400	37.89
10	1090	Jette	50.875959	4.324570	5.04	52417	10400	37.79
11	1140	Evere	50.872010	4.403418	5.02	41588	8284	37.42
12	1150	Woluwe-Saint-Pierre	50.837025	4.427464	8.85	41789	4722	41.92
13	1160	Auderghem	50.817236	4.426898	9.03	33970	3762	40.46
14	1170	Watermael-Boitsfort	50.798106	4.417644	12.93	25172	1947	42.52
15	1180	Uccle	50.803544	4.333844	22.91	82742	3612	34.64
16	1190	Forest	50.811795	4.318119	6.25	55925	8948	39.13
17	1200	Woluwe-Saint-Lambert	50.843045	4.425673	7.22	56496	7825	41.10
18	1210	Saint-Josse-ten-Noode	50.850820	4.369163	1.14	27087	23761	35.61

From the preceding we can see that population density and average age can also be used as discriminant variables

Indeed, performing some basic statistics gives

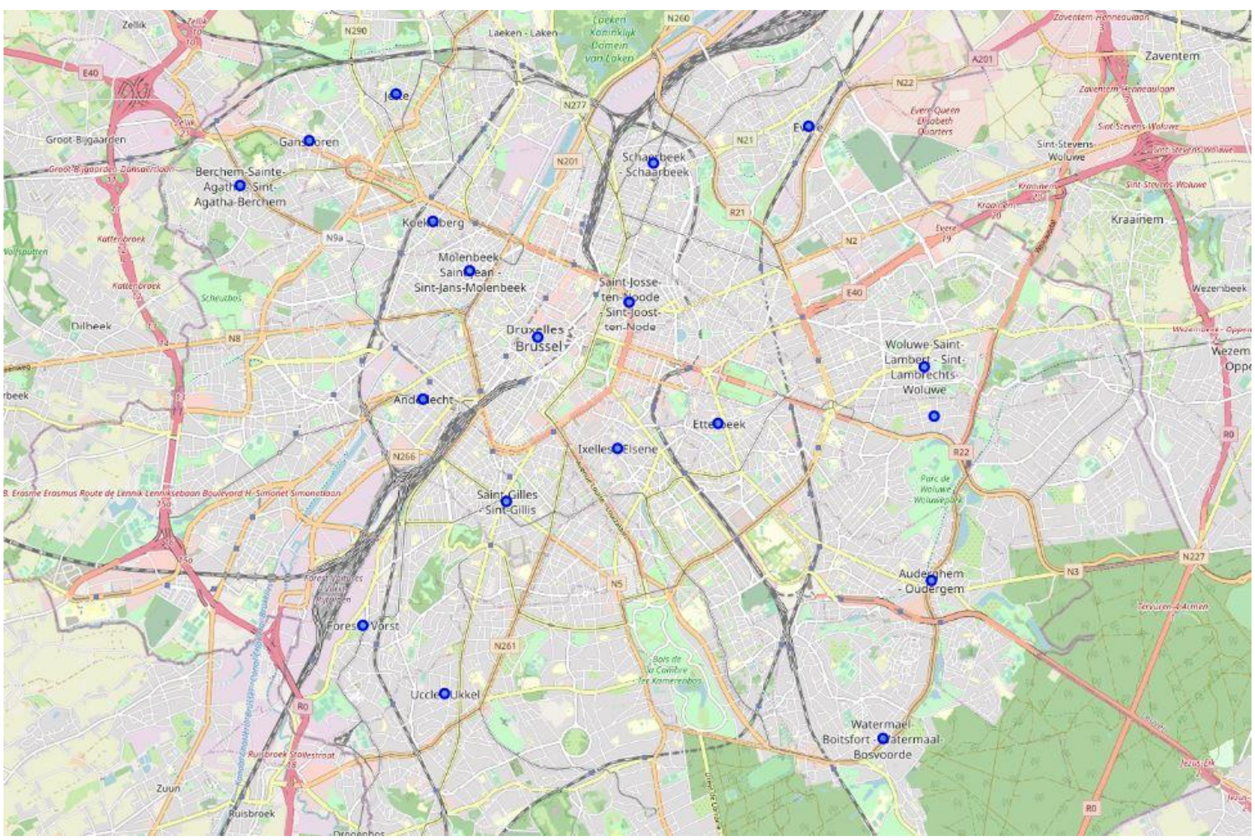
	Population_density	**population_average_age**
min	1,947	33
median	8,948	38
max	23,761	43

We see that 50% of the municipalities have a population that have an average age of population less than 38, with a minimum of 33 and a max of 43. Preference of living around young or old people can have some importance.

Population density show strong differences among municipalities min 2.000 people /km² vs max 23.761. Living around few or a lot of people also can have quite importance.

Therefore, we will use those two variable in the determination of the best place to live for people.

Visually, we could map the Brussels' municipalities



We can see that some municipalities are greener than other and that some include forest. These are mainly the municipalities in the periphery. We could notice that the municipality with the lowest population density include a large part of forest and parks. While the one with the highest population density is in the center of the city what allow to think that most of the people live in apartment buildings VS home with garden for the other.

Adding foursquare data give the following results

	PostalCode	Municipality	Latitude	Longitude	size(km^2)	population	Population_density	population_average_age	restaurants	schools	public_transport_station	shops	parks
0	1000	1000 Bruxelles	50.846557	4.351697	32.61	179797	5514	41.86	30	10	18	30	2
1	1030	Schaerbeek	50.867604	4.373712	8.14	131547	16161	32.99	17	6	7	10	3
2	1040	Etterbeek	50.836145	4.386174	3.15	48008	15241	35.75	25	6	8	30	2
3	1050	Ixelles	50.833114	4.366828	6.34	86675	13671	40.95	30	6	8	30	1
4	1060	Saint-Gilles	50.826741	4.345668	2.52	49715	19728	34.82	26	10	2	28	0
5	1070	Anderlecht	50.839098	4.329653	17.74	118920	6938	37.05	25	7	9	15	0
6	1080	Molenbeek-Saint-Jean	50.854596	4.338636	5.89	96501	16384	36.15	30	12	7	30	2
7	1081	Koekelberg	50.860604	4.331550	1.17	21961	18770	38.80	9	6	6	7	2
8	1082	Berchem-Sainte-Agathe	50.864923	4.294673	2.95	25195	8541	39.05	2	2	1	8	0
9	1083	Ganshoren	50.870327	4.307798	2.64	24817	9400	37.89	5	3	0	16	0
10	1090	Jette	50.875959	4.324570	5.04	52417	10400	37.79	11	2	2	14	2
11	1140	Evere	50.872010	4.403418	5.02	41588	8284	37.42	7	2	3	5	3
12	1150	Woluwe-Saint-Pierre	50.837025	4.427464	8.85	41789	4722	41.92	8	3	4	5	4
13	1160	Auderghem	50.817236	4.426898	9.03	33970	3762	40.46	8	6	3	10	8
14	1170	Watermael-Boitsfort	50.798106	4.417644	12.93	25172	1947	42.52	2	7	2	1	6
15	1180	Uccle	50.803544	4.333844	22.91	82742	3612	34.64	4	6	1	19	0
16	1190	Forest	50.811795	4.318119	6.25	55925	8948	39.13	3	10	2	7	0
17	1200	Woluwe-Saint-Lambert	50.843045	4.425673	7.22	56496	7825	41.10	9	4	4	17	3
18	1210	Saint-Josse-ten-Noode	50.850820	4.369163	1.14	27087	23761	35.61	30	8	20	30	2

To have a more understanding of those new data we will create bins to categorise them between 3 levels (low, Medium and High). This gives the following

	PostalCode	Municipality	restaurant-binned	schools-binned	pts-binned	shops-binned	parks-binned	density-binned
0	1000	1000 Bruxelles	High	High	High	High	Low	Low
1	1030	Schaerbeek	Medium	Medium	Medium	Low	Medium	Medium
2	1040	Etterbeek	High	Medium	Medium	High	Low	Medium
3	1050	Ixelles	High	Medium	Medium	High	Low	Medium
4	1060	Saint-Gilles	High	High	Low	High	Low	High
5	1070	Anderlecht	High	Medium	Medium	Medium	Low	Low
6	1080	Molenbeek-Saint-Jean	High	High	Medium	High	Low	Medium
7	1081	Koekelberg	Low	Medium	Low	Low	Low	High
8	1082	Berchem-Sainte-Agathe	Low	Low	Low	Low	Low	Low
9	1083	Ganshoren	Low	Low	Low	Medium	Low	Medium
10	1090	Jette	Low	Low	Low	Medium	Low	Medium
11	1140	Evere	Low	Low	Low	Low	Medium	Low
12	1150	Woluwe-Saint-Pierre	Low	Low	Low	Low	Medium	Low
13	1160	Auderghem	Low	Medium	Low	Low	High	Low
14	1170	Watermael-Boitsfort	Low	Medium	Low	Low	High	Low
15	1180	Uccle	Low	Medium	Low	Medium	Low	Low
16	1190	Forest	Low	High	Low	Low	Low	Low
17	1200	Woluwe-Saint-Lambert	Low	Low	Low	Medium	Medium	Low
18	1210	Saint-Josse-ten-Noode	High	Medium	High	High	Low	High

We can see that there is a huge difference among municipalities. This reinforce the initial choice of them as discriminant variables for our analysis.

Now lets study the correlations of those variables to check if there could be redundancies among them.

Correlation matrix is the following

	size(km^2)	population	Population_density	population_average_age	restaurants	schools	public_transport_station	shops	parks
size(km^2)	1.000000	0.730042	-0.587526	0.229117	0.113168	0.267473	0.286639	0.105935	0.033845
population	0.730042	1.000000	-0.090796	-0.116484	0.535886	0.435574	0.473034	0.392177	-0.185860
Population_density	-0.587526	-0.090796	1.000000	-0.573735	0.573261	0.311885	0.401839	0.505397	-0.323114
population_average_age	0.229117	-0.116484	-0.573735	1.000000	-0.268987	-0.161889	-0.038644	-0.303096	0.422763
restaurants	0.113168	0.535886	0.573261	-0.268987	1.000000	0.570380	0.750968	0.849810	-0.198026
schools	0.267473	0.435574	0.311885	-0.161889	0.570380	1.000000	0.431083	0.504741	-0.105527
public_transport_station	0.286639	0.473034	0.401839	-0.038644	0.750968	0.431083	1.000000	0.576517	-0.005980
shops	0.105935	0.392177	0.505397	-0.303096	0.849810	0.504741	0.576517	1.000000	-0.357942
parks	0.033845	-0.185860	-0.323114	0.422763	-0.198026	-0.105527	-0.005980	-0.357942	1.000000

Correlations are generally not significant except for restaurants and shops whose correlation is close to 0.85. We could conclude that if a municipality has a lot of restaurants it will also have a lot of shops and conversely.

Therefore we will remove "restaurant" from our study and include restaurants in shops keeping only shops as discriminant variable.

Now that we have determined all the variable that can be used to model the best municipality to live for people, we can switch to modeling.

3.4. Modeling

To obtain the best discriminant way to fit the best municipality to live in using the variable chosen previously, we will use a classification approach. Decision tree has been chosen to give us expected results.

Variable of concern are "Population_density", "population_average_age", "shops" , "schools", "public_transport_station" and "parks".

As we use already processed data as source (i.e. already aggregated by municipality), we will not be able to train and test our model, we will use decision tree directly on them and will check if we can consider results as sufficiently discriminant to fit our goal.

4. Results

The rsulting of our work is a decision tree.

Obtained tree is presented in appendix:

The tree can be summed up as (where PTS = Public Transport Stations, pop dens = population density & average = average age)

level 1	level 2	level 3	level 4	level 5	Municipality
shops <=14.5	schools <=4.5	PTS <=2.5	pop dens <=9470		Berchem-Sainte-Agathe
shops <=14.5	schools <=4.5	PTS <=2.5	pop dens > 9470		Jette
shops <=14.5	schools <=4.5	PTS > 2.5	PTS <=3.5		Evere
shops <=14.5	schools <=4.5	PTS > 2.5	PTS > 3.5		Woluwe-Saint-Pierre
shops <=14.5	schools > 4.5	PTS <=2.5	parks <=3		Forest
shops <=14.5	schools > 4.5	PTS <=2.5	parks >3		Watermael-Boitsfort
shops <=14.5	schools > 4.5	PTS > 2.5	pop dens <=9961		Auderghem
shops <=14.5	schools > 4.5	PTS > 2.5	pop dens > 9961	parks <=2.5	Koekelberg
shops <=14.5	schools > 4.5	PTS > 2.5	pop dens > 9961	parks >2.5	Schaerbeeck
shops > 14.5	parks <=1.5	schools <=6.5	shops <=17.5		Ganshoren
shops > 14.5	parks <=1.5	schools <=6.5	shops > 17.5	pop dens <=8641	Uccle
shops > 14.5	parks <=1.5	schools <=6.5	shops > 17.5	pop dens > 8641	Ixelles
shops > 14.5	parks <=1.5	schools > 6.5	shops <=21.5		Anderlecht
shops > 14.5	parks <=1.5	schools > 6.5	shops > 21.5		Saint Gilles
shops > 14.5	parks > 1.5	pop dens <=11553	parks <=2.5		1000 Bruxelles
shops > 14.5	parks > 1.5	pop dens <=11553	parks > 2.5		Woluwe-Saint-Lambert
shops > 14.5	parks > 1.5	pop dens > 11553	aver age <= 35.68		Saint-Josse-ten-Noode
shops > 14.5	parks > 1.5	pop dens > 11553	aver age > 35.68	pop dens <=15812	Etterbeek
shops > 14.5	parks > 1.5	pop dens > 11553	aver age > 35.68	pop dens > 15812	Molenbeek-Saint-Jean

Comparing with the datasets and binned ones stated previously, we can see that this tree is in line with datas.

We can see that according to the tree, the first discriminant factor id the number of shops in close area. Threshold is put around 15 shops in close area. Then we have schools or Park that comes as second criterion. Population density, public transport and average age come next.

So, if people want to make shopping and go to the restaurant close to their living place, they would prefer to live in one of the last 10 municipalities of the table hereover. Otherwise, they would choose among the first 9. Note that they would have the opportunity to go to other municipalities with public transport or with car even as for any European capital cities the use of the car is not the best way to move in the city.

If they have children that need to go to school, they would choose a municipality with schools and parks at proximity

5. Discussion

The obtained tree gives good results comparing to people already living in Brussels. We should not forget that before selecting a living place is always better to go there and see if we like it.

This model gives only a general way to find a living place. Some people may want to choose according to the dataset got from Foursquare.

There could be other discriminant aspects like the cost of living. Note that for European Union members, money is in general not an issue, so for them (our target) the model should suit.

To be fully complete we should add types of shops and restaurants to fit with nationality of European members. Indeed, an Italian would prefer a place close several Italian restaurant. This should be an improvement possibility for the model but with the risk it becomes too complicated to be used.

We could think to make a bespoke model to fit with specific individuals.

6. Conclusion

In this project we use Foursquare data combined to statistical data of Brussels to find a way to help European members people identifying their place to live while relocating.

We used a decision tree to find a model after selecting discriminant variables through data exploration.

The obtained model is good to give a first insight but could be improved with bespoke criterions. This should be balanced between added value and complication of the model.

Appendix: decision tree