Finding the best value real estate for the foreigner moving to Brussels city

Introduction to the problem

Brussels is a dynamic and diversified city. It's been historically formed as 19 different communities as 'Boroughs' in NYC or Toronto). Due to the different historical and economical development of those communities, there is a big disparity between them on different aspects, such as availability of restaurants, shops and other venues. This is also reflected in the average price of the apartments by neighboorhoods. However, recently thanks to the support of the Brussels government some distressed areas of the city started to develop social offerings, support infrastructure, improve the crime situation. A newcomer who is coming over to Brussels has often a difficulty to find the best place to stay given the access to the venues.

As such, the objective of my work would be to find the best price/quality neighboorhood in Brussels and to rank those venues according to certain criteria. For this purpose I will be using different statistical techniques, namely K-means clustering to cluster specific areas of the city and match it with the normalised prices of the apartments. This info will give a substantial information to the person on the best possible choice of residence given his or her interests and considering the budget limits.

As a result of my project, several main deliverables will be given to the target audience:

1. Information on the main types of venues by residence clusters in the city which are better target a certain person. For instance, if a person is in his 30-40s, he will feel more comfortable living in a neighbourhood with a good choice of restaurants vs someone who is in their 20s who would prefer night clubs or bars. Also
2. Within certain cluster find better deals on the apartment price by selecting the best price within a given cluster.

Data sources

For this project I will be using public sources of information- prices of the real estate in Brussels (for simplicity I have chosen the average price reported for the apartment by district in Brussels).

The website for the data of real estate prices is located here: https://statbel.fgov.be/fr/open-data?category=66.

For the venues I will be using the data of Foursquare by major types. There are certainly limitations of the use of Foursquare, which I will describe more in details in the discussion section.

Methodology

I will be using statistical techniques and methodologies to solve my problem.

First, I will describe my data using statistical techniques, then after merging the data by district on the apartment prices and merge with Foursquare data, I will be using K-means algorithm to determine clusters for the analysis.

Finally, detailed analysis of the clusters will give us some overview on the main venue types, distribution of the venues and then give some recommendation in terms of residences.

Approach to the analysis

The data on the real estate prices (EUR per sqm) show the great disparity between the districts as shown below, despite relatively low number of people living in Brussels city (1.21 mln according to the public sources) and small area of 162 km^2. The price varies from as low as 1940 EUR/m2 in Anderlecht to 3750 EUR/m2 in Woluwe-St-Pierre.

Price EUR/m2 of the apartment in Brussels, 2018 by district
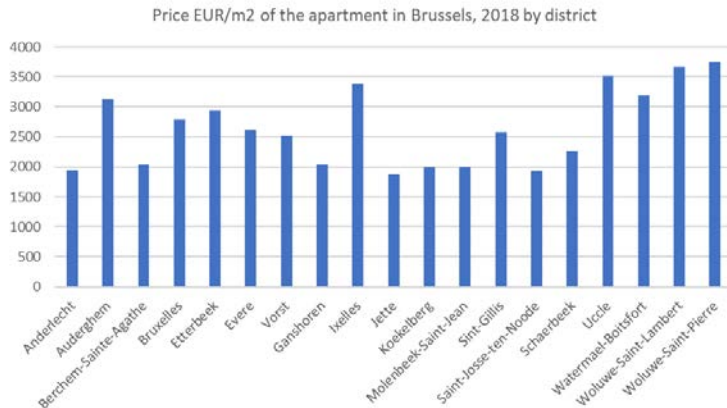
Many of those discrepancies are linked with historical aspects, urban development, public transport offering etc, which I will not discuss in this paper. In this analysis, we will focus exclusively on the availability and types of venues in each district. Also, due to the limitation of non-commercial license of Foursquare, we will be focusing on the venues located within the radius of 2000 meters from the center of each district.

 For the clustering of the data we'll be using the methodology called K-means which allows to effectively segment the data in K different clusters without pre-defined conditions. This approach is called unsupervised machine learning.

Results of the study

As a result of the study, we have identified 5 clusters with similar venue structure, as shown below:
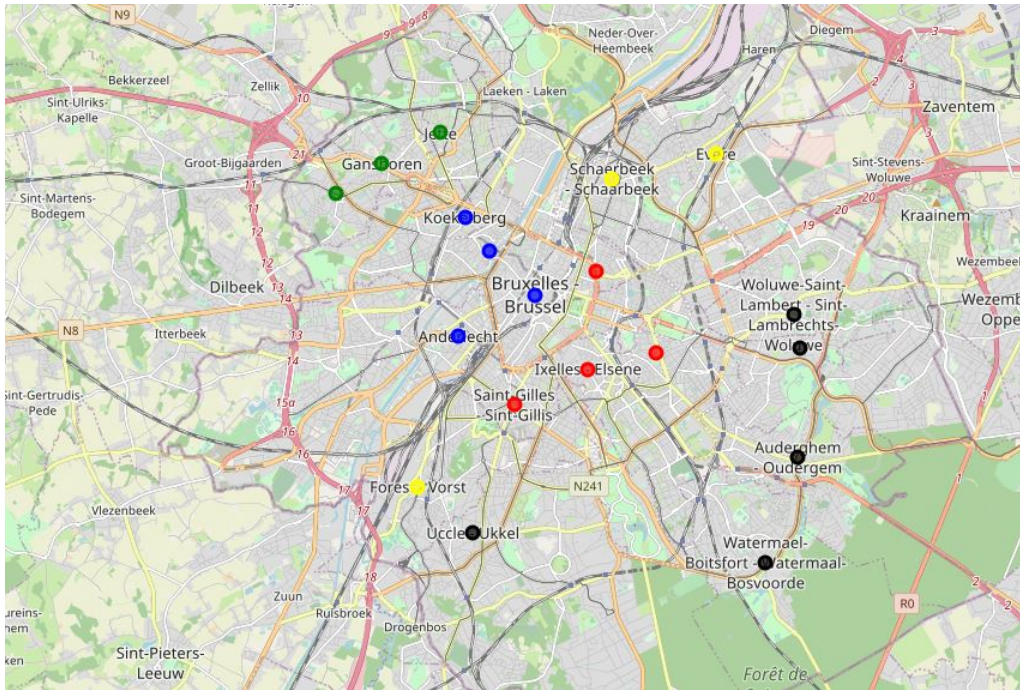
As one can see, districts in those clusters are located next to each other. As I mentioned earlier, this is mainly due to the historical reasons and the way the districts were developed. Unsurprisingly, those districts show similar price ranges. However, there are some price differences within each segment, which can be used to select the best value real estate.

Let's look at the clusters more in details. When grouping by clusters, similar venues appear as described in the table below:

Cluster	District	Price per sqm	Main venues
0	Woluwe St Pierre	3750	Italian and French Restaurants, Parks, Bakery, Gastropub
	Auderghem	3130	
	Watermael Boitsfort	3190	
	Uccle	3520	
	Woluwe St Lambert	3660	
1	Molenbeek	2000	Bar, Hotel, Plaza, Brewery
	Koekelberg	2000	
	Anderlecht	1940	
	Bruxelles	2790	
2	Ixelles	3380	Plaza, Coffee shop, bar
	Etterbeek	2940	
	Sint Gillis	2570	
	Saint-Josse-ten-Noode	1930	
3	Vorst	2520	

	Evere	2610	Bar, Snack place, supermarket
	Schaerbeek	2260	
	Berchem St Agathe	2040	Bar, Pizza place, Chinese restaurant
	Ganshoren	2040	
4	Jette	1880	

As you can see, cluster 0 is a posh area where there are mainly high-level restaurants with French and Italian cuisine, parks and gastropubs which make them a perfect place to live for middle class 40+ population. In that cluster, best value commune is Auderghem which has the lowest price.

If you are a young professional looking to spend some time to go out in the weekends, then the cluster 2 is the choice, which is also close to the city center. There are number of pubs and coffee shops in those areas whereas the price for the real estate is rather comfortable. The district of St Josse ten Noode is a good value and not far from the city center.

Finally, if you have a family and you might want the proximity of large supermarkets, then segment 3 is your choice with the district of Schaerbeek offering a good value.

Next steps and other considerations

This study of course is just a first look at the problem of selection of the ideal place to stay in Brussels. Due to the limitation of the data, the following could be done to improve the analysis:

1. Incorporate the venues rating in the analysis to have a selection of top choice
2.   Extend the selection of venues not only to the radius of 2000 meters from the center of the district but to include all the venues with a specific project code
3. Group venues into larger segments and perform the analysis on the broader groups.

Having mentioned all above, current data analysis with visualization already gives a good overview of the Brussels districts for the newcomer.

WELCOME TO BRUSSELS!