Capstone Project: Battle of Neighborhoods

Suitable locations for a techno club in Brussels

Tom Nel

Introduction

Although the techno scene in Brussels is not on the same level as Berlin's, the Belgian capital has an exciting nightlife and many techno clubs offer a great night out till early morning. There is reason enough to believe however that another techno temple could add to the various options Brussels' night owls already have: many artists in the techno scene reside in Brussels and sometimes it is difficult for them to show off their talent as the established techno clubs already have limited available slots to perform. Moreover, a frustrating aspect of the clubs are the waiting lines, which indicates that there are many enthusiasts of techno, though mostly only a limited amount of people are allowed inside at the same time. That's why I've decided to use location data and other considerations to offer the best feasible location in Brussels if a new techno club were to be built.

Approach and Criteria

1st Analysis

In Berlin, most of the techno clubs are located in the district "Friedrichshain". In general, there is a thriving nightlife scene and therefore we will set this neighborhood against all municipalities in Brussels and analyze which municipality resembles this Berlin neighborhood the most.

2nd Analysis

If multiple municipalities are similar to Friedrichshain after performing a general comparison, a deeper analysis will be made to further distinct the most favorable neighborhood in Brussels. When searching the perfect location for a new techno club, many other aspects could be considered such as safety, accessibility, type of neighborhood, etc. In my project I've decided to further distinguish a favorable location based on following criteria.

First of all, a lot of nightclubs are coping with noise disturbance and this is something that can be reduced but not disappear. As such, a distinction will be made between neighborhoods and residential areas will be left out of the options. The new location should be in a neighborhood where already **a lot of bars are present**, as many people tend to roll over from bars to nightclubs once they have a few beers in. Finally, we will compare **the amount of nightclubs already present** in the neighborhoods and make appropriate recommendations based on the findings.

Outcomes

After these two analyses, we should be able to narrow it down to one single most favorable neighborhood. It has to be emphasized that this is merely a hypothetical question as the data foursquare provides us gives recommended venues, not all the venues in a neighborhood and as such is not complete, which might skew a neighborhood's properties.

To summarize, the goal is to identify the best districts in Brussels to open a new techno club. A district which is similar to Friedrichshain *and* meets the most of the criteria laid out in the above section. The results will be translated in a simple form that will convey the data-driven analysis for the best locations.

Target Audience

This analysis could be helpful for future business owners in the nightlife industry or hospitality in general, to identify criteria that are important to consider when opening a new venue.

Data

For the project objectives to be achieved, python geolocation libraries will used, along with the Foursquare API. Also, in order to create clusters of similar candidate locations, the K- Means machine learning clustering algorithm will be used.

The following data has been used to make several analyses:
- Data information on the specific Berlin "Friedrichshain" district
- List of the different municipalities in Brussels (https://en.wikipedia.org/wiki/List_of_municipalities_of_the_Brussels-Capital_Region)
- Foursquare venue data for the respective municipalities
- Foursquare venue data for the Berlin district

Using geolocation libraries, lattitude and longitude values of the respective districts are added to allow comparisons using foursquare location data. This data is used to explore the Brussels' neighborhoods on the above criteria and to compare the districts with Friedrichshain to find similar neighborhoods in Brussels.

How the data is used to solve the problem
Analysis 1

A comparison is made based on the type of venues that are most common in the Friedrichshain district and municipalities of Brussels. A K-means clustering algorithm determines which of the municipalities are most similar to the Friedrichshain district.

The goal of K-means clustering is to find groups in the data. K stands for the number of groups. Data points are clustered based on the similarity of the features. So the more similar the data features, the more likely they will be grouped. The municipalities that we ultimately look for are the ones in the same cluster with Friedrichshain.

Analysis 2

Once the clusters are made, the foursquare and geopy data was used to find the best one out of those neighborhoods. The amount of nightclubs and amount of bars present and whether the neigbhorhood is residential or not was used to differentiate the best neighborhood from the good ones.

The processing of this data allowes us to answer the key question:

Which neighborhood in Brussels, similar to Friedrichshain, has the best properties to open a new techno club, given our limited data availability and assumptions?

After extracting the data from the source and performing multiple data wrangling operations, the following dataframe consisting of the neighborhoods and their coordinates was obtained:

	Neighborhood	Population	Latitude	Longitude
0	Anderlecht	118241	50.839098	4.329653
1	Auderghem	33313	50.817236	4.426898
2	Berchem-Sainte-Agathe	24701	50.864923	4.294673
3	Bruxelles-Ville	176545	50.846557	4.351697
4	Etterbeek	47414	50.836145	4.386174
5	Evere	40394	50.872010	4.403418
6	Vorst	55746	50.811795	4.318119
7	Ganshoren	24596	50.870327	4.307798
8	Ixelles	86244	50.833114	4.366828
9	Jette	51933	50.875959	4.324570
10	Koekelberg	21609	50.860604	4.331550
11	Molenbeek-Saint-Jean	96629	50.854596	4.338636
12	Saint-Gilles	50471	50.826741	4.345668
13	Saint-Josse-ten-Noode	27115	50.850820	4.369163
14	Schaerbeek	133042	50.867604	4.373712
15	Uccle	82307	50.803544	4.333844
16	Watermael-Boitsfort	24871	50.798106	4.417644
17	Woluwe-Saint-Lambert	55216	50.843045	4.425673
18	Woluwe-Saint-Pierre	41217	50.837025	4.427464
19	Friedrichshain	133754	52.512215	13.450290

The following images show the different districts on the map, in Brussels and Berlin respectively.
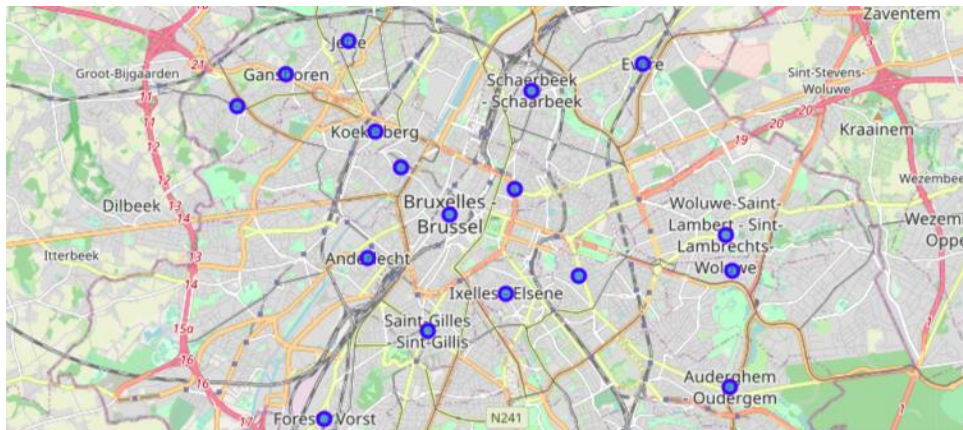
As an exploratory analysis, a comparison of the populations of each neighborhood was made. We can conclude that the population of Friedrichshain is within range of the respective other population numbers of the Brussels' districts.
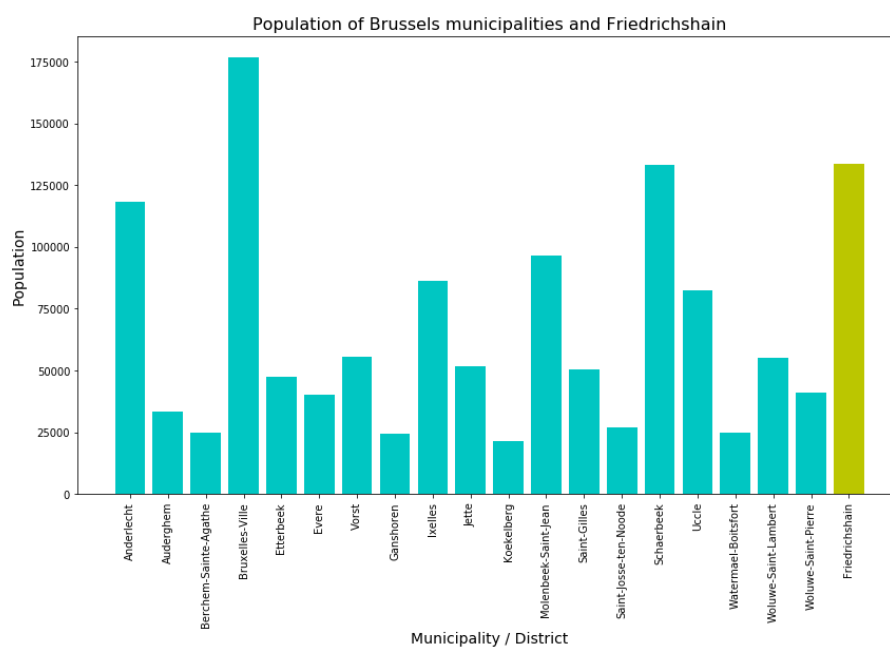
Methodology

The goal is to identify the best district in Brussels to open a new techno club. A district which is similar to Friedrichshain and meets the criteria laid out in the data section.

Now that we have the districts and their respective coordinates in a dataframe, we continued as follows:

- Foursquare location data extraction for each district: getting a list of nearby venues in each district (non-exhaustive!) with general explore function
- One-hot encoding to make the data ready for a K-means clustering algorithm
- K-means clustering algorithm
- Cluster Analysis
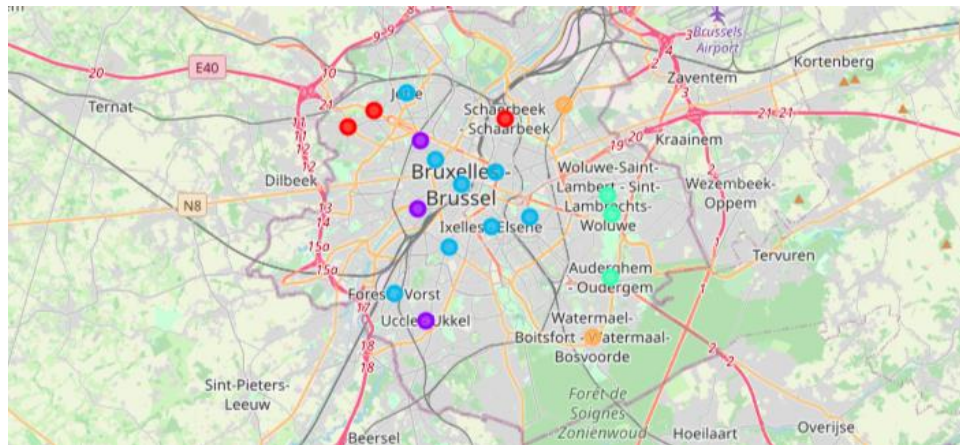
The two figures above show the respective clusters. Once we had the different clusters, a further analysis was made for the cluster with Friedrichshain in it. A new foursquare query with the explore function, but this time the extracted venues restricted to the "Nightlife" category. A specific category ID was used in the explore function. After doing a few operations on the obtained dataset, we visualized the frequency of bars and nightclubs for each neighborhood in the relevant cluster (see image below).

Frequency of bars and nightclubs for each neighborhood (click to enlarge)

Results & Discussion

1st Analysis

When performing the clustering algorithm on the dataset, Friedrichshain was put in a cluster with Auderghem, Woluwe-Saint-Lambert and Woluwe-Saint-Pierre. Friedrichshain is an outer skirt district in Berlin, as are these Brussels' districts. Its plausible that being an outskirt district, the neighborhood consists of different venues compared to city-center districts.

2nd Analysis

When analyzing the third cluster (with Friedrichshain), the explore function was specifically set to search for 'nightlife' venues only. After plotting the frequency of bars and nightclubs we can conclude the following:

- In terms of bars present, Auderghem outperforms Woluwe-Saint-Lambert and Woluwe-Saint-Pierre with more than twice the amount.
- When we compare the nightclubs in each district, the data didn't show nightclubs for Woluwe-Saint-Lambert and Woluwe-Saint-Pierre, however, the foursquare explore function found even more nightclubs in Auderghem than in Friedrichshain.

This tends to favor the neighborhood **Auderghem** over the others, however, there are a few sidemarks that are worth mentioning:

- The initial clustering algorithm clustered the districts based on a general explore function, which gives *recommended* venues in the area. If all relevant venues would be considered and the dataset would be bigger (now 1700-ish rows), a more accurate dataset of venues would probably give us a better clustering of districts.
- Even if we restricted the explore function to nightlife venues only, I'm confident not all clubs are being considered as Friedrichshain should have more clubs than all districts in Brussels combined. No clubs present in Woluwe is also doubtful.

- Auderghem does have a lot of nightclubs and bars in the area and assuming people go out in places where there is already a lot of nightlife, favors Auderghem above the others.
- An extra advantage for Auderghem is the presence of more nature and being a less residential area than Woluwe-Saint-Lambert and Woluwe-Saint-Pierre, which helps the case of noise disturbance reduction.

Conclusion

Based on our analysis and the clustering algorithm, we can conclude that **Auderghem** best meets the proposed criteria. The area has a lot of bars, nightclubs and has an environmental advantage. Further analyses could be made to further investigate to where exactly open a new techno club.

There are many ways this analysis could've been performed based on different data sources or different machine learning algorithms. I opted for the k-clustering algorithm as it is straightforward and simple in many ways. Assumptions have been made and the data retrieved from foursquare is non-exhaustive, so Auderghem being the most favorable district is merely an outcome of this analysis in a limited environment and may or may not be the best district in real life. Domain knowledge of the industry, government regulations and restrictions regarding nightlife business and a more detailed demographic analysis for each district could benefit this research, and will probably change the outcome. We also didn't differentiate between types of nightclubs, as techno clubs attract a certain type of people.

However, the result is here and based on what's been proposed and assumed, this is the best outcome. The data approach helps us to think out of the box, providing a new way of thinking and data-driven analyses will lead to critical thinking and a better solution in the end.

Thanks for taking part in my Data Science journey!