Foursquare data project

Open
a new bar
in Brussels center

specific market approach
missing gap:
students

Table of Content

Inhoudsopgave

Abstract

If a person wants to open a new bar in the center of a big city, one can expect heavy competition. After COVID-19, it could be that a lot of bars needed to close down business, so many places could come vacant to start your own business. Important in the entrepreneurship is to investigate where opportunities lay, what specific market approach needs to be chosen.

Brussels has the **most students** of Belgium (+100,000) and this number is still increasing. Universities and non-Universities ('Hogescholen' in Dutch) are scattered around in Brussels and are also very present in City center (see figure 1). Nevertheless there is not much visibility of students in the city center. Students blend in and there are not many bars focused on them. **The nightlife in the center is more focused on locals or tourists.**

There are some student bars but they are located **out of the center**. But this has downsides for the student:
  – Brussels is not a very safe or friendly environment for bikes and steps: altitude differences, bumpy road, traffic, …
  – significant amount of students feel unsafe to walk large distances at night or get in a cab with someone they don't know.
  – Using a private car is not a good alternative for most students and it is difficult to find parking spots.
  – public transportation stops around 01h00.

Of course going out near the schools and the places **where the students live and sleep** is much more convenient. Especially for new students who don't have a large peer group yet. Freshmen and international students can have a hard time to **make new friends.** There is a lot of diversity and already the choice of language of the school (Dutch, French or English) can make it more difficult to meet new people.

New students (Freshmen and international students) don't have habits yet and there is a lot of opportunity. In **Dorms** (in the notebook referred to as a dataframe with the name "Sleeplaces"), a lot of students are gathered and it is a perfect spot for getting to know fellow students. For finding those places I use 3 different categories in Foursquare: "Sorority House", "Fraternity", "College Residence Hall". New students who don't know the city or people

from Brussels yet can find dorms useful to pick .

Summary:

In this paper I will investigate how much nightlife there is around the dorms. The place with less bars has less competition for a new bar which makes it easier to attract new customers.
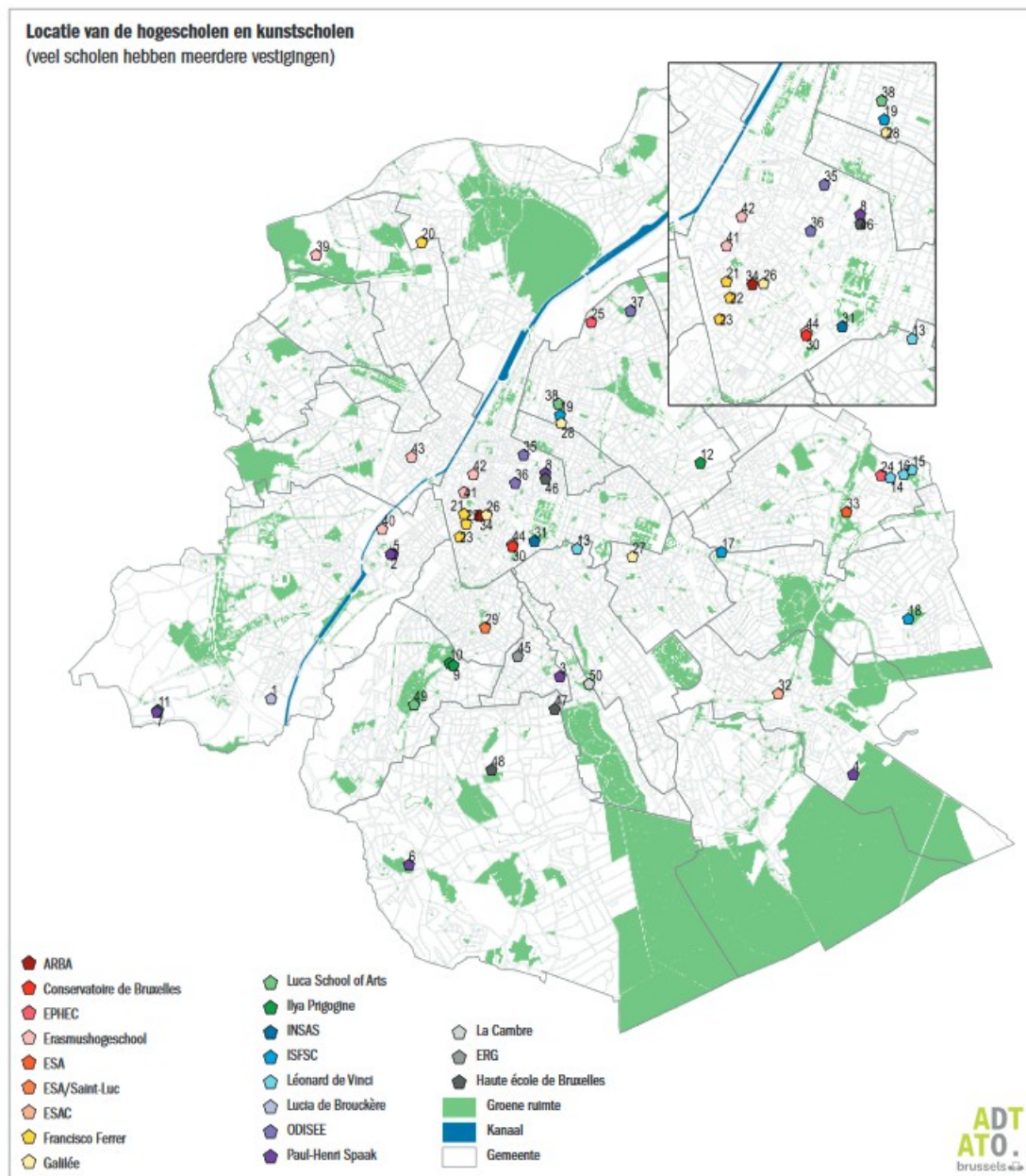

Figure 1: Placement of schools in 2014
https://perspective.brussels/sites/default/files/documents/NL_ADT_Synt_HE_v04.pdf

Methodology

A) Tools

The data is from **Foursquare** and is reached by API calls. Foursquare data is crowd sourced. Crowd sourcing can yield large numbers of data points, but the measures are inherently biased - you don't know what part of any given population you've missed. Also, crowd sourced data skews to people who know how to use smart-phones and are adept with social media.

But because the project is focused on digital natives, the data is self selected and representative. Foursquare data is ideal for precise locations and close distances (*reference 1*) and in contrast to what most people think, the data is up to date even without people actively checking in on Foursquare *(reference 2)* or have the Foursquare-app installed *(reference 3)*!

The Foursquare data is reached by sending **API-calls** selected by CategoryId
"https://api.foursquare.com/v2/venues/explore?&categoryId={}&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}"
(reference 4)

Juypter Notebook with Python 3 is used and the data is loaded into Python Pandas dataframes for processing. It is visualised by Folium because we work with exact geographical locations. Also Matplotlib is used to present and visualize the outcomes.

B) Usage

With the 3 categories ("Sorority House" + "Fraternity" + "College Residence Hall") I make a total dataframe "Sleepplaces". In my example I use a radius of 2000m around the center. I focus on geographical city center (latitude = 50.85, longitude = 4.35) which is not the same as the touristic center but very close to a spot with a vivid nightlife (Beurs, Hallen van Saint-Géry, Ancienne Belgique, Sainte-Catherine, Dansaertstraat, ...)

For each of those dorms (sleepplaces) I look how much Nightlife (categoryId ='4d4b7105d754a06376d81259') is close by. I classify the sleepplaces by amount of bars nearby.

Discussion

With a radius of 9000 m around the center (latitude = 50.85, longitude = 4.35) I cover all of Brussels to get an idea of where dorms are situated.
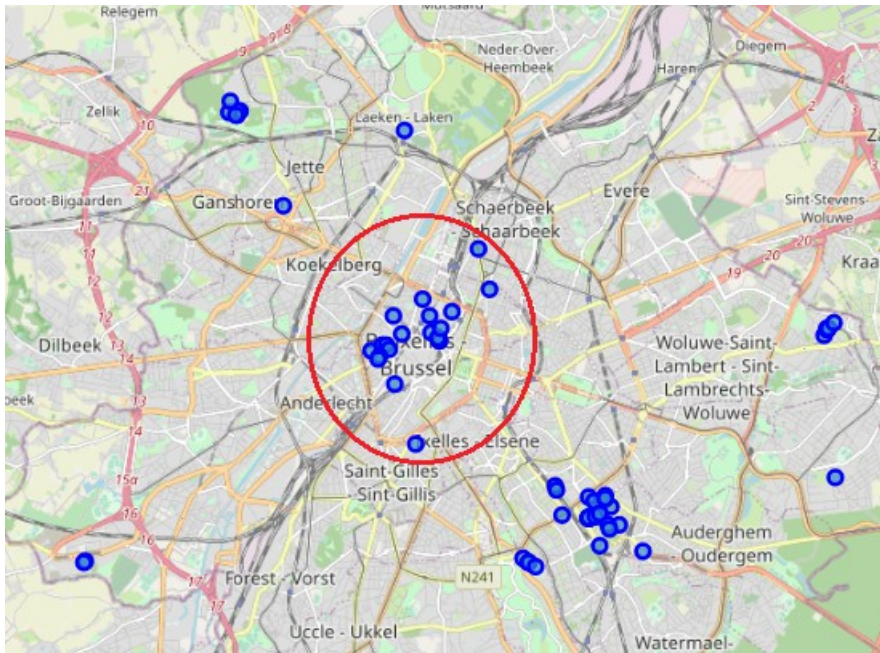

Figure 2: See notebook

Dorms are situated in two clusters: one around Oudergem – Watermaal-Bosvoorde and one I will get my attention to: around the city center.

Because I only want to work with the cluster in the center I use radius 2000m with still the same latitude and longitude to address the sleepplaces in the red circle on Figure 2.

With category "College Residence Hall" I encounter a problem. Apparently *"Br(ik Koten"* is an organization with different locations. In Foursquare all sites have the exact same name. Of course each different place has to be uniquely identified so I need to add a suffix.

For all different sites of "Br(ik Koten" I add a number and process it again in the same format of the other 2 dataframes.

Foursquare data project

Now that I have cleaned and pre-processed the data of all dorms, I need to append the 3 dataframes in one dataframe with the latitude and longitude.

My solutions has 20 locations which I store in the dataframe "Sleepplaces" and can visualize with folium. I now can use the latitudes and longitudes to again search Foursquare for all Nightlife around each sleepplace.
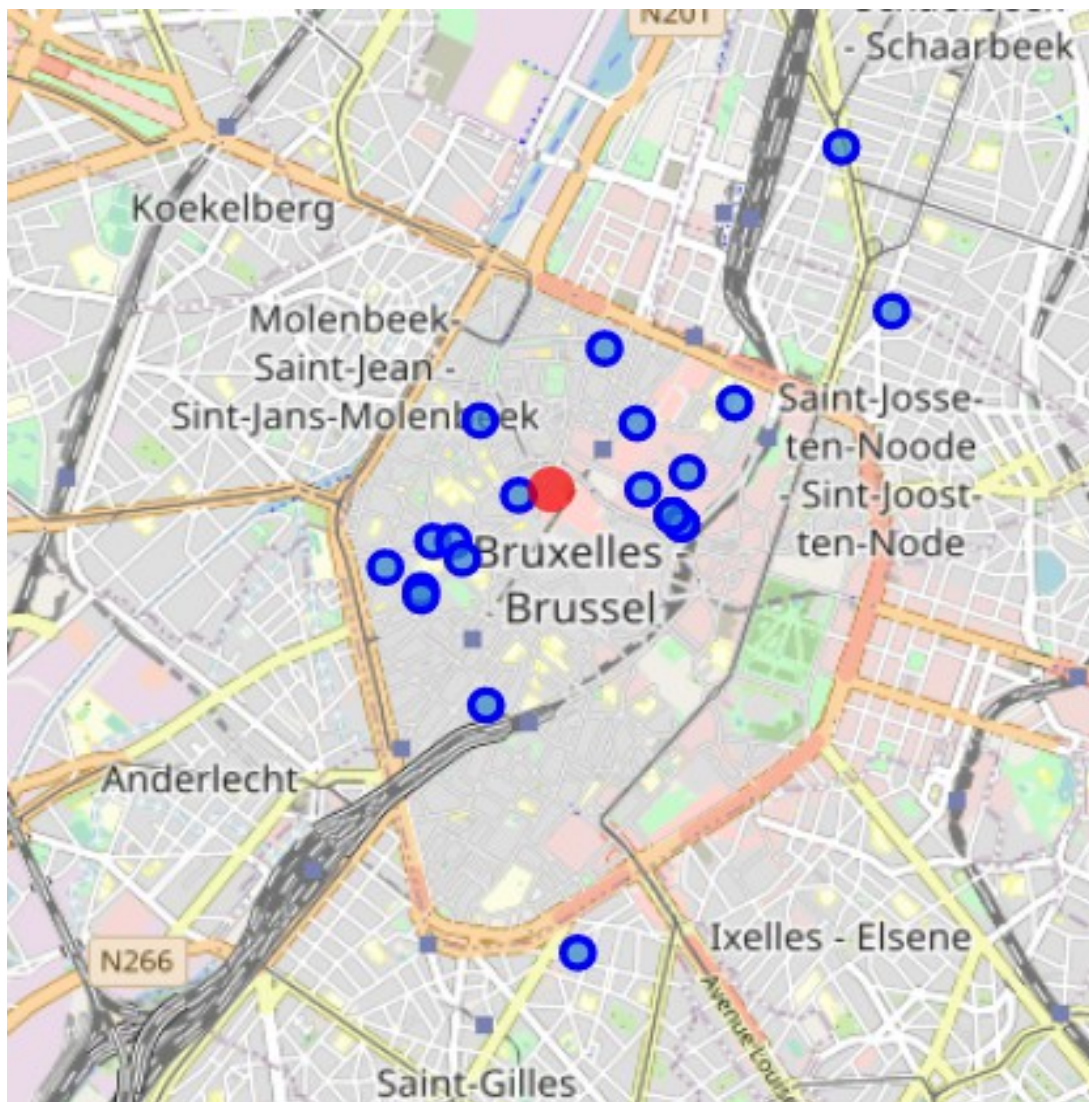

Figure 3: See notebook. The red spot is the geographical center and all sleepplaces in blue around it with a radius of 2000m are shown. The main roads around the city center mark an area with a pentagon-shape. 3 dorms are located nearby but just outside, the other 17 are located even closer to

the center.

With the dataframe I now can also use those latitudes and longitudes to find the Nightlife around each spot. I loop in the dataframe and send out an API call for each location and store all found Nightlife for each dorm in a new dataframe. I can do this for any given radius.

Since My focus is how much nightlife there is in close proximity of the dorms/sleepplaces I choose a small radius.

It is very interesting to know how many bars are within 50m of the sleepplace but of course if by coincidence a hotspot for bars is situated some 150m of the sleepplace this has a considerable effect. Therefore I choose radius 50m, 100m, 150m en 200m. Bigger radius than 200m is not relevant anymore since all dorms are located in a radius of 2000m and sometimes 2 different dorms are closer than 200m. Using a radius that is too large will only present a bar nearby a lot of dorms.

I sort by r200m and rename columns for clarity. I group the nightlife for each dorm and add up the number of bars nearby for each place.
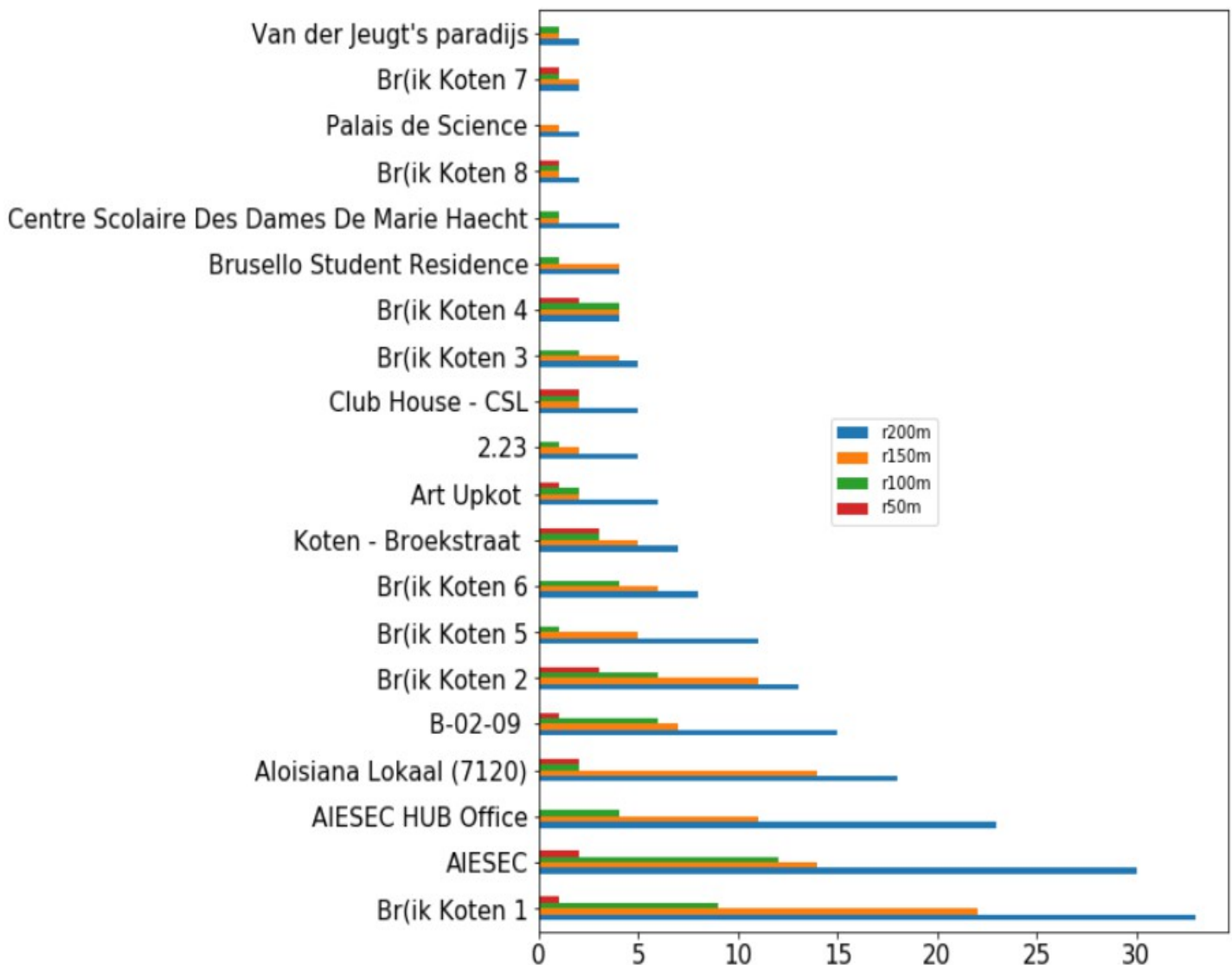
Foursquare data project

|	r200m	r150m	r100m	r50m
Br(ik Koten 1	33	22	9.0	1.0
AIESEC	30	14	12.0	2.0
AIESEC HUB Office	23	11	4.0	NaN
Aloisiana Lokaal (7120)	18	14	2.0	2.0
B-02-09	15	7	6.0	1.0
Br(ik Koten 2	13	11	6.0	3.0
Br(ik Koten 5	11	5	1.0	NaN
Br(ik Koten 6	8	6	4.0	NaN
Koten - Broekstraat	7	5	3.0	3.0
Art Upkot	6	2	2.0	1.0
2.23	5	2	1.0	NaN
Club House - CSL	5	2	2.0	2.0
Br(ik Koten 3	5	4	2.0	NaN
Br(ik Koten 4	4	4	4.0	2.0
Brusello Student Residence	4	4	1.0	NaN
Centre Scolaire Des Dames De Marie Haecht	4	1	1.0	NaN
Br(ik Koten 8	2	1	1.0	1.0
Palais de Science	2	1	NaN	NaN
Br(ik Koten 7	2	2	1.0	1.0
Van der Jeugt's paradijs	2	1	1.0	NaN

Figure4: see notebook.

Conclusion

This table can be visualized in different ways, depending on what the bar opener is focused on.
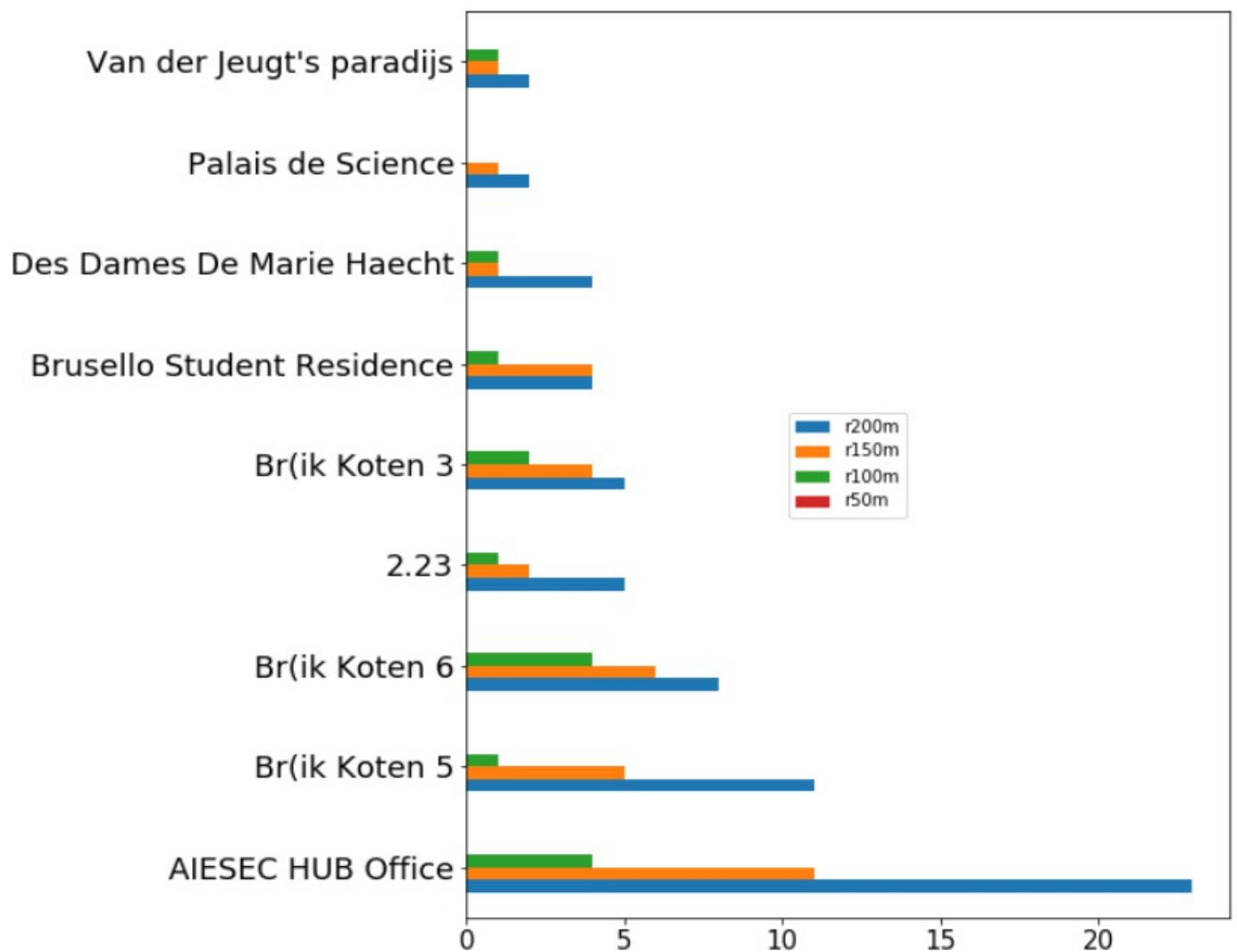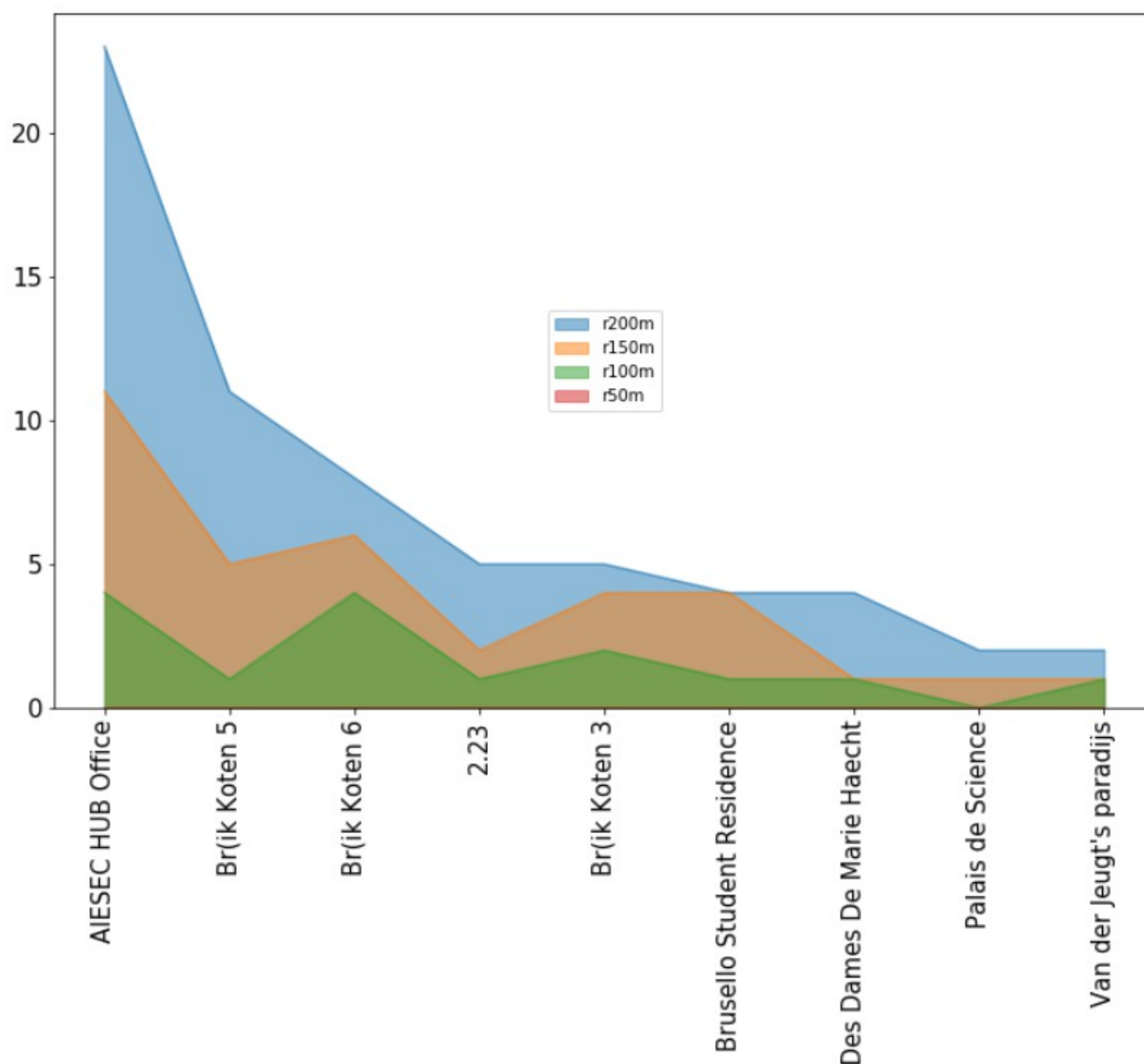
Not much nightlife around in total

Foursquare data project

Not much nightlife closeby (sorted by "r50m")
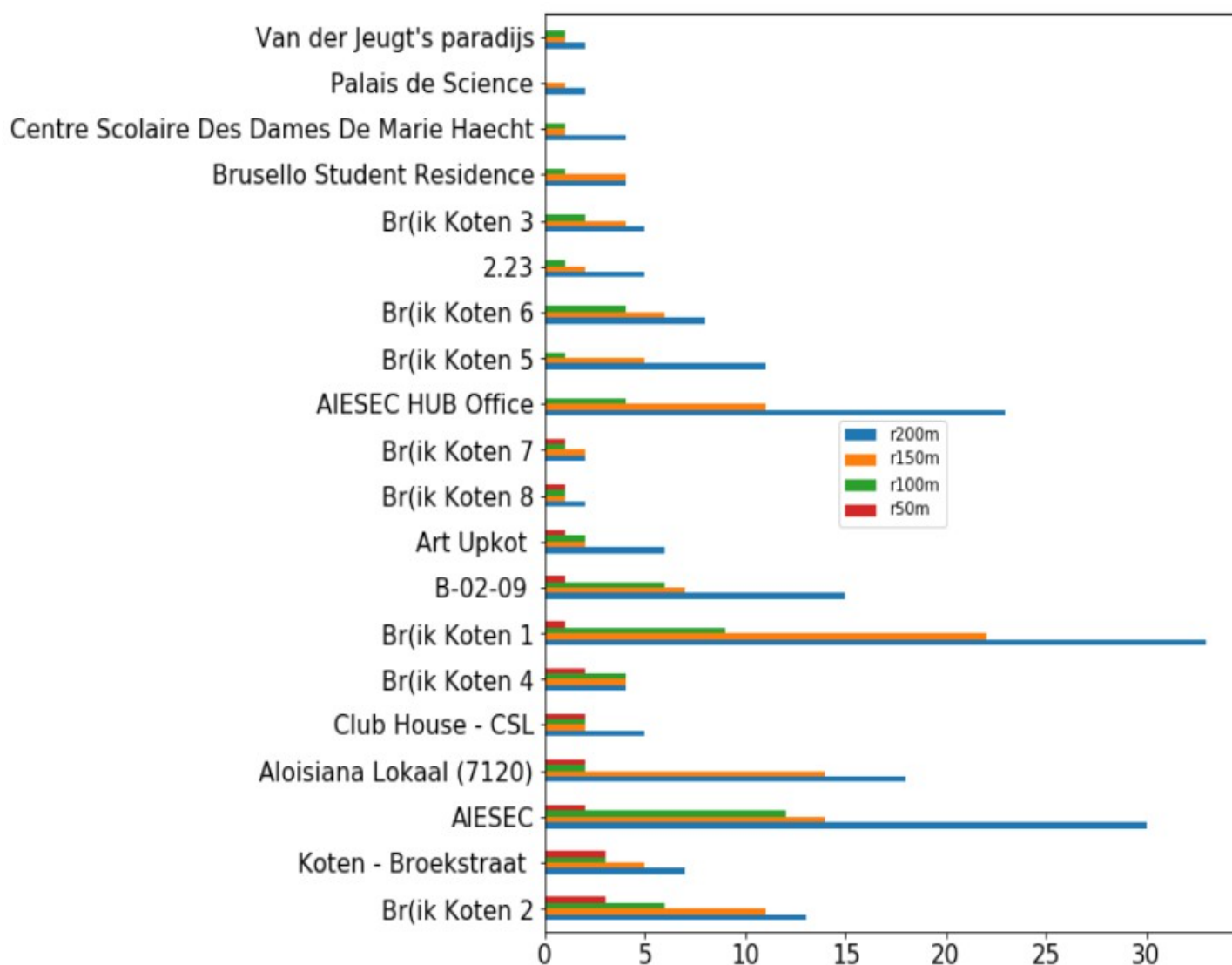
Focused on places with no nightlife in radius of 50m

This script can be used several times to see of something would have changed within a couple of weeks / months.

Six of them are very good candidates with a total of <=5 bars in radius 200m. In those neighborhoods can be searched and they can be compared against other parameters.

Foursquare data project

References

Figure 1: https://perspective.brussels/sites/default/files/documents/NL_ADT_Synt_HE_v04.pdf

(1) "Because of the "corpus" of data generated by people like Uncle Tony, Foursquare knows when the dimensions of storefronts change and can tell the difference between an office on the eighth floor and one of the ninth floor." https://nymag.com/intelligencer/2019/08/ten-years-on-foursquare-is-now-checking-in-to-you.html

(2) "In addition to all of those active check-ins, at some point Foursquare began collecting passive data using a "check-in button you never had to press." https://nymag.com/intelligencer/2019/08/ten-years-on-foursquare-is-now-checking-in-to-you.html

(3) "You might think you don't use Foursquare, but chances are you do. Foursquare's technology powers the geofilters in Snapchat, tagged tweets on Twitter; it's in Uber, Apple Maps, Airbnb, WeChat, and Samsung phones, to name a few." https://www.wired.com/story/you-may-have-forgotten-foursquare-it-didnt-forget-you/

(4) https://developer.foursquare.com/

Figures on notebook: https://github.com/Nick0165/coursera-capstone-project/tree/Nick0165-Data

https://labs.cognitiveclass.ai

www.coursera.com

https://cloud.ibm.com