

Applied Data Science Capstone

Where to open a new office in Budapest

Norbert Bogya
bnorbert88@gmail.com

June 12, 2021

When a **new start-up IT company** levels up and can afford a real **office** instead of working home, it is quite important to **open** it in as suitable neighbourhood as possible.

Parameters, should be considered:

- renting price
- size and scalability of the office
- human factors

Ideal spot: youthful neighbourhood near university buildings that provide a motivating working environment.

Considered human factors

- To create a **motivating** working environment, a **dynamic office-block area** is needed to choose. If workers are surrounded by similar workers, who enjoys to work there, it will help their productivity.
- As the potential **new workers** will come from the **universities** of Budapest, it is convenient to find a place that is near the main university buildings and colleges. In this case it will be much more attractive for young, agile students who are taking classes at a university in parallel their job.
- To increase the **youthfulness** of the area, it is recommended to choose a neighbourhood with places that are preferred in the circle of young people. For example, **cafés** in the area will empower the attractiveness among them.

We try to cluster the neighbourhoods in Budapest, considering the number of existing office buildings, university buildings and cafés, and with the help of this clustering to provide a suggestion to the spot that is suitable for a new office.

Data: Wikipedia page about the neighbourhoods of Budapest

https://hu.wikipedia.org/wiki/Budapest_v%C3%A1rosr%C3%A9szainek_list%C3%A1ja

Budapest is divided onto 23 districts, and each district may contain several neighbourhoods. We use only the “Név” and “Kerület” columns of the table, which correspond to the names of the **neighbourhoods** and **districts**, respectively.

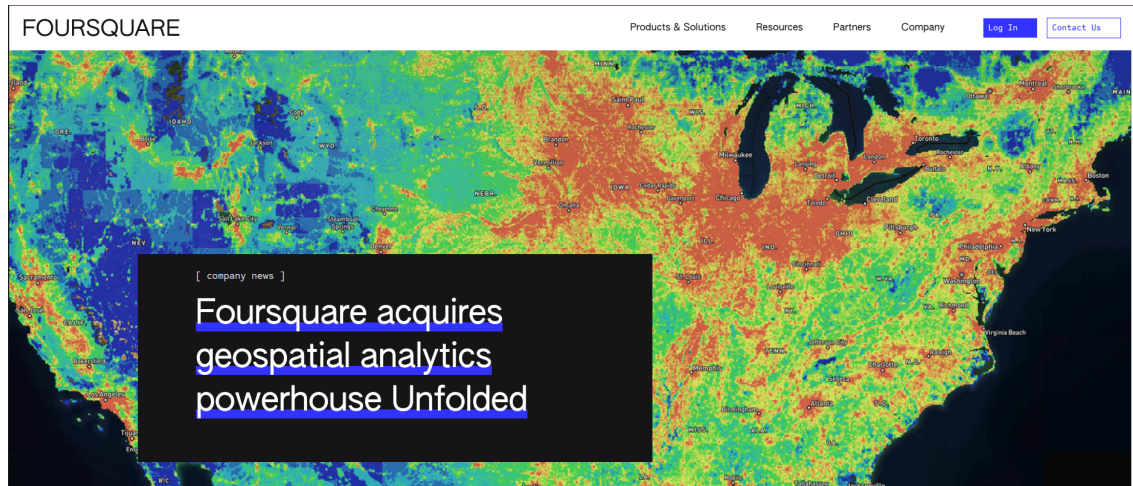
A városrészek listája [\[szerkesztés \]](#)

–	Név	Kerület	Lakosság (2001)	Megjegyzés	Területi lehatárolás ^[2]	↕
1	Adyliget	Budapest II. kerülete	856		Rézsú utca 82. számú telek (50860 helyrajzi szám) északnyugati oldalától-Rézsú utca-Nagykovácsi út-Feketefej utca-a főváros határa a Rézsú utca 82. számú telek (50860 helyrajzi szám) északnyugati oldaláig. 1950 előtt Nagykovácsi része volt.	
2	Akadémiaútelep	Budapest XVII. kerülete	2895		Pesti út a X. és XVII. kerület közigazgatási határáról-513. utca-a MÁV ecseri vonala-a X. és XVII. kerület közigazgatási határa a Pesti útig.	
3	Albertfalva	Budapest XI. kerülete	11 845		Kondorosi út a Solt utcától-Duna folyam-Hosszúréti-patak-MÁV pécsi vonala-Solt utca a Kondorosi útig.	
4	Almáskert	Budapest XVIII. kerülete	—		145111/2091 és 145111/386 hrsz.-ú névtelen közterület-Alacsony út-Ganz Ábrahám utca-Kerékkötő út-a főváros közigazgatási határa-Határ utca 145111/2091 és 145111/386 hrsz.-ú névtelen közterületig.	
5	Alsórákos	Budapest XIV. kerülete	29 023		Madridi utca a Szent László úttól-MÁV Körvasút-Vezseny utca-Vazul utca-Körvasút sor-Szolnoki út-Kerepesi út -Rákospatak-Füredi utca-Nagy Lajos király útja-Teleki Blanka utca-Szegedi út-Tatai utca-Kármor utca-Szent László út a Madridi utcáig.	
6	Angyalföld	Budapest XIII. kerülete	62 006		A MÁV esztergomi vonala a Duna folyamtól-Újpalotai út-Dugonics utca-Madridi utca-Szent László út-Kármor utca-Tatai utca-Szegedi út-Dévényi utca-Vágány utca-a MÁV váci vonala a Bulcsú utcai vasúti aluljáróig-Bulcsú utca-Lehel utca-Lehel tér keleti és nyugati oldala-Váci út-Meder utca-Duna folyam a MÁV esztergomi vonaláig.	
7	Aquincum	Budapest III. kerülete	761	2012. december 12-én területe megnőtt	Duna folyam-Bogdáni út-Szentendrei út-Kazal utca-Husztói út folytatása északi irányba a Zsófia utcáig-Zsófia utca-23215 hrsz.-ú ingatlan déli telekhatára-23152/45 hrsz.-ú ingatlan északkeleti telekhatára a Péter utcáig-Péter utca meghosszabbítása déli irányban a vasútvonalig-vasútvonal.	
8	Aranyhegy- Ürömhegy- Péterhegy	Budapest III. kerülete		Aranyhegy és Ürömhegy egyesítéséből és bővítéséből alakult 2012. december 12-én	Budapest közigazgatási határa-Bécsi út-Aranyvölgy utca-Aranyhegyi út-Pusztakúti út a Pusztakúti köz-Pusztakúti köz meghosszabbítása a Héthalom utcáig-névtelen közterület (22890/2 hrsz.) - 22374/1 hrsz.-ú ingatlan keleti telekhatára Budapest közigazgatási határáig	
9	Árpádföld	Budapest XVI. kerülete	6186		Budapest közigazgatási határa a Csömöri úttól-Budaörsi út-Szolovák út-Csömöri út a Budapest közigazgatási határáig.	
10	Baross Gábor- telep	Budapest XXII. kerülete	3016		Szabadkai utca a Kamaraerdei úttól-Csiperke utca-Klauzál Gábor utca-XI. utca és meghosszabbított vonala-Rózsakert utca-Minta utca-Dózsa György út-a főváros határa-Kamaraerdei út és meghosszabbított vonala a Szabadkai útig.	
				2012. december 12-én	Budapest közigazgatási határa-Duna folyam a Pünkösdfürdő utcáig-Pünkösdfürdő utca-Árpád utca-Ipartelep utca-Szent István utca-Madzsar József utca-Hollós	

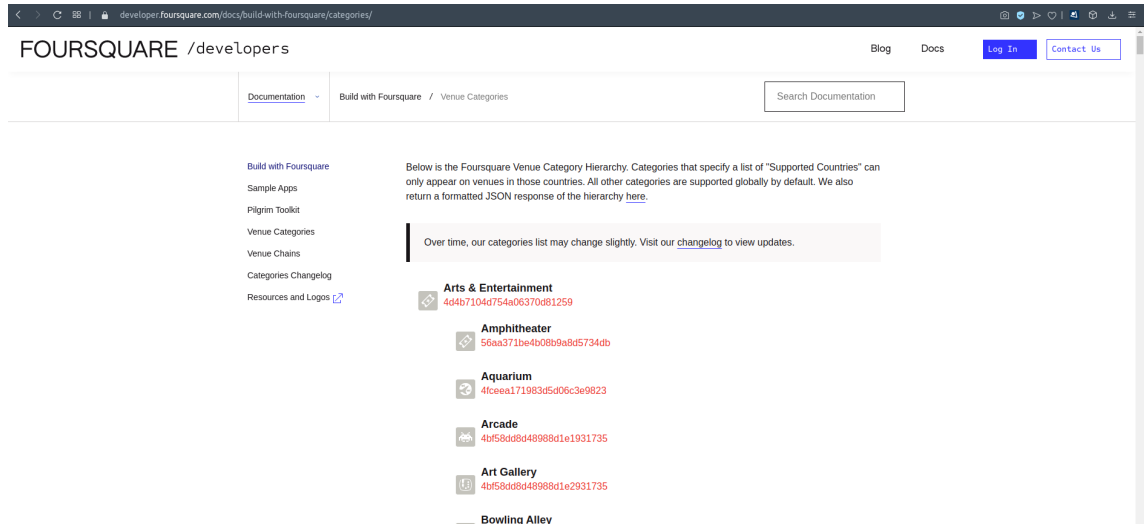
Data: *Foursquare* API

<https://api.foursquare.com>

We use the *Foursquare* API to explore the venues of areas.



During the API calls, we filters for the mentioned categories of the venues: **cafés, universities and offices**. Foursquare API has an option to restrict our exploration to several type of venues.



The screenshot shows the Foursquare developer documentation page for Venue Categories. The page has a dark header with the Foursquare logo and navigation links for Blog, Docs, Log In, and Contact Us. Below the header, there's a breadcrumb trail: Documentation > Build with Foursquare > Venue Categories. A search bar is also present. The main content area is divided into a left sidebar with links to various resources and a main content area. The main content area starts with a paragraph explaining the Venue Category Hierarchy and a link to a changelog. Below this, there's a list of venue categories, each with an icon, a name, and a unique identifier (Foursquare ID).

FOURSQUARE /developers

Documentation > Build with Foursquare > Venue Categories

Search Documentation

Build with Foursquare

Sample Apps

Pilgrim Toolkit

Venue Categories

Venue Chains

Categories Changelog

Resources and Logos

Below is the Foursquare Venue Category Hierarchy. Categories that specify a list of "Supported Countries" can only appear on venues in those countries. All other categories are supported globally by default. We also return a formatted JSON response of the hierarchy [here](#).

Over time, our categories list may change slightly. Visit our [changelog](#) to view updates.

Arts & Entertainment
4d4b7104d754a06370d81259

Amphitheater
56aa371be4b08b9a8d5734db

Aquarium
4fcea171983d5d06c3e9823

Arcade
4bf58dd8d48988d1e1931735

Art Gallery
4bf58dd8d48988d1e2931735

Bowling Alley

Data manipulation:

- Loading data from Wikipedia: *Beautiful Soup* Python library
- Store data: *pandas* dataframes
- Geographic locations of neighbourhoods: *GeoPy* Python client with *Nominatim* and *Photon* geocoding services
- Venue information: Foursquare API calls

Clustering:

- *k*-means algorithm by the *scikit-learn* Python package
- Elbow rule to choose the optimal value of *k*
- Plot data points on map: *Folium* package

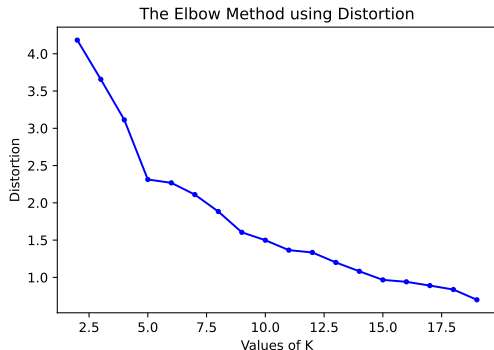
Dataframe after data manipulation

Beautiful Soup \Rightarrow pandas \Rightarrow GeoPy \Rightarrow Foursquare

	district	neighborhood	Latitude	Longitude	office	university	cafe
0	I	Gellérthegy	47.492064	19.037200	4	3	3
1	I	Krisztinaváros	47.496866	19.029776	17	5	7
2	I	Tabán	47.491613	19.043169	2	1	3
3	I	Vár	47.495334	19.039546	3	4	8
4	I	Víziváros	47.503719	19.039128	4	3	8
5	II	Adyliget	47.547550	18.938984	0	0	0
6	II	Budakeszierdő	47.542471	18.972903	0	0	0
7	II	Budaliget	47.567579	18.940664	0	1	0
8	II	Csatárka	47.531525	19.002578	0	1	0
9	II	Erzsébetliget	47.561714	18.967558	0	0	0

Optimal value of k for clustering

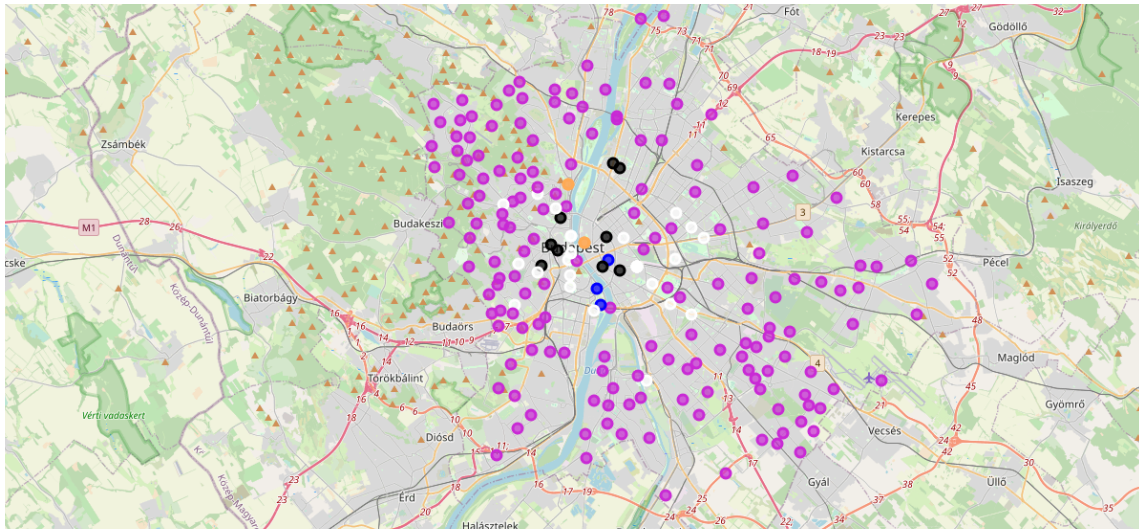
Creating models with $k = 2, 3, \dots, 20$, we get the following distortion plot.



Using the elbow method, we can see that the optimal choice for k is 5.

Clusters

Considering the number of cafés, university and office buildings

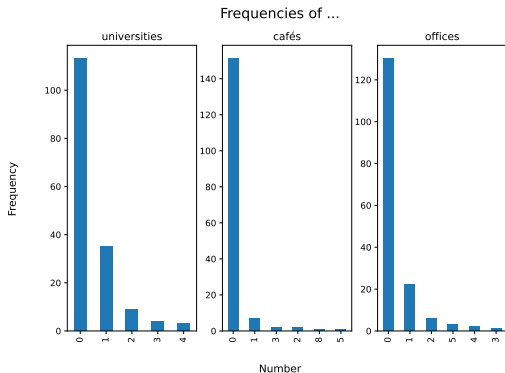


Size of clusters

Cluster Labels	Size
0	29
1	9
2	3
3	164
4	4

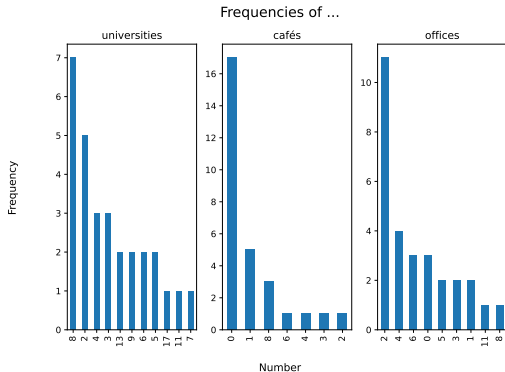
Cluster 3

- Most of the neighbourhoods lie in this cluster.
- This **cluster** is the least suitable for us.
- Triplets (universities,cafés,offices) have low numbers.
- **None** of these neighbourhoods can be **recommended** to open a new office.



Cluster 0

- White dots.
- Few university buildings and quite small number of cafés and universities.
- This cluster is **neither recommended**, because of the small number of important factors.



- Blue dots.
 - Only 3 elements.
 - Large number of university buildings, however the number of offices and cafés are quite small.
-
- Main buildings of the two biggest universities of Budapest.
 - Huge office area, called Infopark.
 - Few number of reported cafés, they are inside the universities and office buildings
 - These spots are highly recommended to open new office, but it is an overwhelmed area, and to open offices is not too cheap.

Cluster 1 and 4

- **Cluster 4** and **Cluster 1** are similar clusters.
- Large number of office buildings but small number of cafés and university buildings.
- Cluster 4 has only few element, and these neighbourhoods are the most crowded spots of Budapest, including the inner city. There are a lot of office buildings here, but they are far from the university campuses.
- Cluster 1 is more balanced with a lot of office building and moderate number of cafés and university buildings.
- They can be recommended as a spot for a new office.

Taking everything into account, by this clustering method we can recommend the clusters in the following priority order.

- 1 Cluster 2
- 2 Cluster 1
- 3 Cluster 4
- 4 Cluster 0
- 5 Cluster 3

Thank you for your attention!