Capstone Project - The Battle of Neighbourhoods

Norbert Bogya – bnorbert88@gmail.com

June 12, 2021

Business Problem

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. In general, the decision can be made by several parameters, for example renting price, size and scalability of the office, but in large office block, these parameters can be varied quite easily. In this analysis, we focus on the human part of the problem.

Shortly, an ideal spot to rent an office should be a youthful neighbourhood near university buildings that provide a motivating working environment. In details, to choose the best neighbourhood, we consider the following three aspects.

- 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.

By the above concept, 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 provide a suggestion to the spot that is suitable for a new office.

Data

We use two different kind of data. We need static data about the neighbourhoods of Budapest and venue data about what can we find in the appropriate areas.

(1) `https://hu.wikipedia.org/wiki/Budapest_v%C3%A1rosr%C3%A9szeinek_list%C3%A1ja`
This website contains table about the neighbourhoods of the Hungarian capital, Budapest, see Figure 1. Budapest is divided onto 23 districts, and each district may contain several

A városrészek listája [szerkesztés]

–	Név	Kerület	Lakosság (2001)	Megjegyzés	Területi lehatárolás[2]
1	Adyliget	Budapest II. kerülete	856		Rézsű utca 82. számú telek (50860 helyrajzi szám) északnyugati oldalától-Rézsű utca-Nagykovácsi út-Feketefej utca-a főváros határa a Rézsű utca 82. számú telek (50860 helyrajzi szám) északnyugati oldaláig. 1950 előtt Nagykovácsi része volt.
2	Akadémiaújtelep	Budapest XVII. kerülete	2895		Pesti út a X. és XVII. kerület közigazgatási határától-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-Alacskai 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 -Rákos-patak-Füredi utca-Nagy Lajos király útja-Teleki Blanka utca-Szegedi út-Tatai utca-Kámfor 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-Kámfor 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-Huszti 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 közig-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-Budapesti út-Szlovák 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.
11	Békásmegyer	Budapest III. kerülete	38 169	2012. december 12-én hegyvidéki kiterjedése csökkent Csillaghegy javára	Budapest közigazgatási határa-Duna folyam a Pünkösdfürdő utcáig-Pünkösdfürdő utca-Árpád utca-Ipartelep utca-Szent István utca-Madzsar József utca-Hollós Korvin Lajos utca-Dózsa György utca-65536 hrsz.-ú közterület-Rókavár utca-Márton köz-Hegyláb utca-64869/4, 64871/4, 64876/1, 64879 hrsz.-ú ingatlanok déli telekhatára-Tamás utca-65164 hrsz.-ú ingatlan délnyugati, majd nyugati telekhatára-65163 hrsz.-ú ingatlan nyugati telekhatára és északkeleti telekhatárának meghosszabbítása Budapest közigazgatási határáig.
12	Bélatelep	Budapest XVIII. kerülete	1285		A XVII. és XVIII. kerület határa a Tünde utcától-Csévéző utca-a MÁV szolnoki vonala-Tünde utca a XVII. és XVIII. kerület határáig.
13	Belső-Ferencváros	Budapest IX. kerülete		2012. december 12-én alakult	Kálvin tér-Üllői út-Ferenc körút-Boráros tér-Duna folyam-Fővám tér-Vámház körút.
14	Belsőmajor	Budapest XVIII. kerülete	1687		A MÁV lajosmizsei vonala a Dózsa György utcától-a főváros határa-a XVIII. és XXIII. kerület közigazgatási határa-Dózsa György utca a MÁV lajosmizsei vonaláig.
15	Belváros	Budapest V. kerülete	12 244		Vigadó tér déli oldala-Deák Ferenc utca-Deák Ferenc tér déli oldala-Károly körút-Múzeum körút-Vámház körút és meghosszabbított tengelye-Duna folyam a Vigadó tér déli oldaláig.
16	Bókaytelep	Budapest XVIII. kerülete	5093		Üllői út a Bartók Lajos utcától-Dalmady Győző utca-Wlassics Gyula utca-Madarász utca-Bókay Árpád utca-Garay utca-Varjú utca-Fürst Sándor utca-Sallai Imre utca-Margó Tivadar utca-Bartók Lajos utca az Üllői útig.

Figure 1: Image about the table from Wikipedia

neighbourhoods. We use only the “Név” and “Kerület” columns of the table, which correspond to the names of the neighbourhoods and districts, respectively.

(2) `https://api.foursquare.com`
We use the *Foursquare* API the explore the areas and retrieve the necessary information that correspond to the above problem description. During the API calls, we filters for the

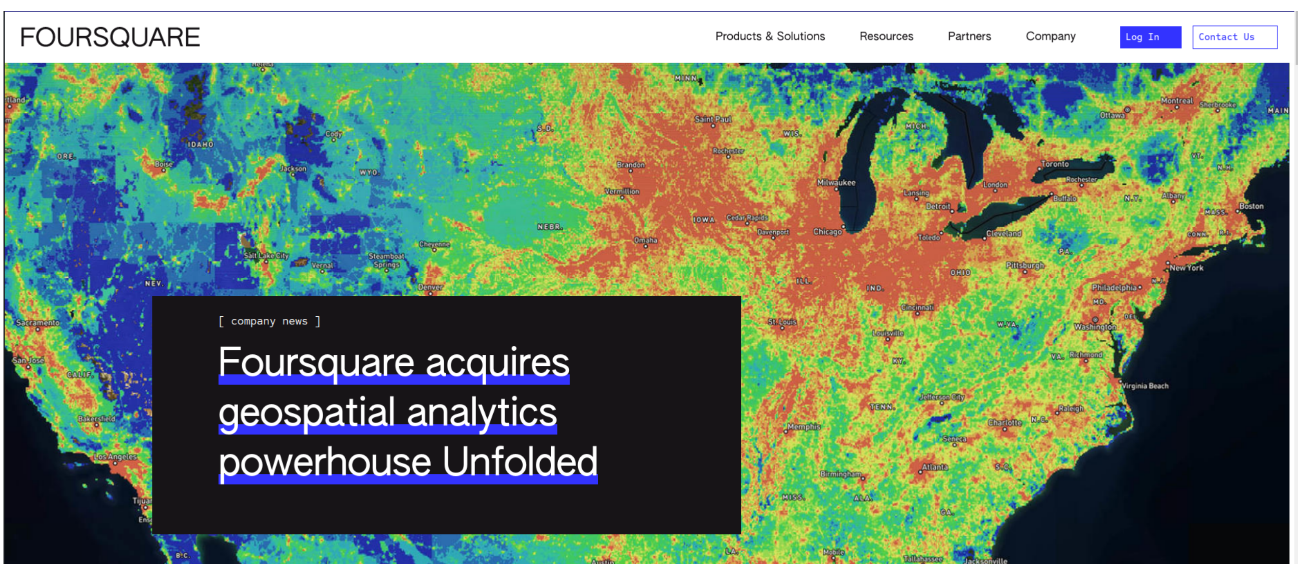

Figure 2: Homepage of the Foursquare application

mentioned categories of the venues: cafés, universities and offices. Foursquare API has an option to restrict our exploration to several type of venues, see Figure 3.

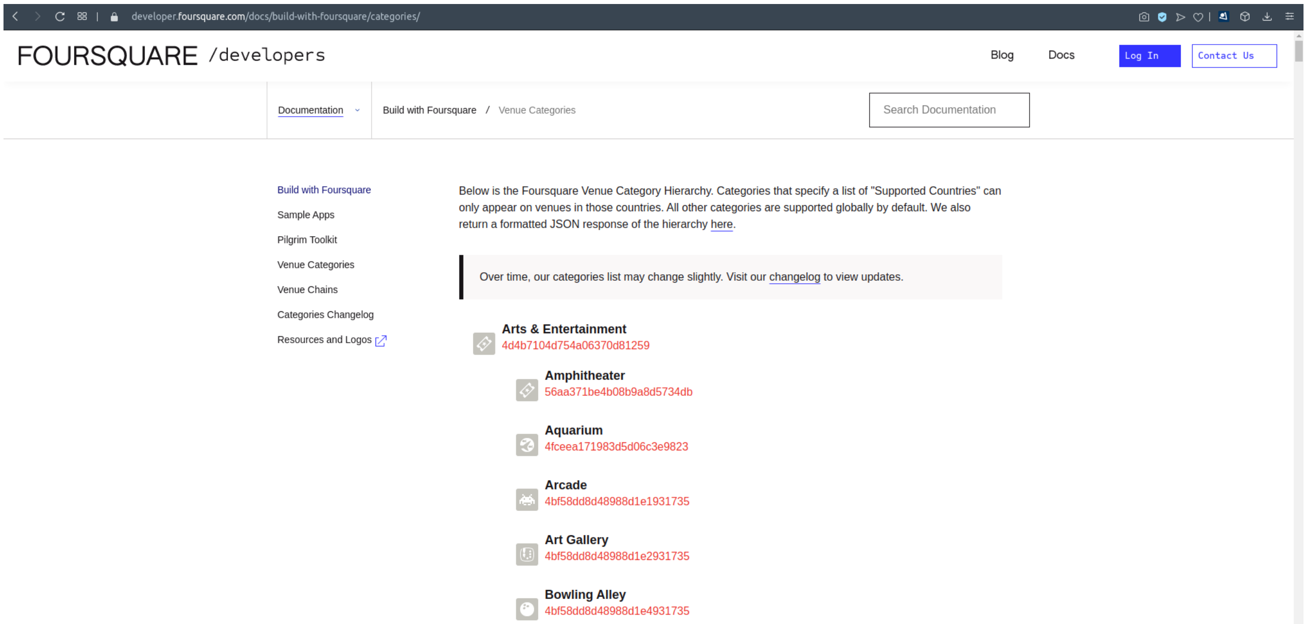

Figure 3: List of available categories in Foursquare API

To demonstrate this, I illustrate the data on the neighbourhood Gellérthegy of the district I od Budapest. If we want to focus for the nearby office buildings, we make the API call

`https://api.foursquare.com/v2/venues/explore?&client_id=CLIENT_ID&client_secret=SECRET_CLIENT_ID&v=20180605&ll=47.492064,19.037200&radius=300&limit=100&categoryId=4bf58dd8d48988d124941735`

that retrieves with the following JSON data with 4 different venues from the office category.


```
{'groups': [{'items': [{'reasons': {'count': 0,
      'items': [{'reasonName': 'globalInteractionReason',
        'summary': 'This spot is popular',
        'type': 'general'}]},
     'referralId': 'e-0-590c76cb18d43b3d8ca62088-0',
     'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/shops/internetcafe_',
         'suffix': '.png'},
        'id': '4bf58dd8d48988d1f0941735',
        'name': 'Internet Cafe',
        'pluralName': 'Internet Cafes',
        'primary': True,
        'shortName': 'Internet Cafe'}],
      'id': '590c76cb18d43b3d8ca62088',
      'location': {'address': 'Attila út 27',
       'cc': 'HU',
       'city': 'Budapest',
       'country': 'Magyarország',
       'distance': 255,
```

```
      'formattedAddress': ['Budapest',
       'Attila út 27',
       '1013',
       'Magyarország'],
      'labeledLatLngs': [{'label': 'display',
        'lat': 47.49351,
        'lng': 19.03983}],
      'lat': 47.49351,
      'lng': 19.03983,
      'postalCode': '1013',
      'state': 'Budapest'},
     'name': 'UrbanFood Cafe & Coworking',
     'photos': {'count': 0, 'groups': []},
     'venuePage': {'id': '429929520'}}},
   {'reasons': {'count': 0,
     'items': [{'reasonName': 'globalInteractionReason',
       'summary': 'This spot is popular',
       'type': 'general'}]},
    'referralId': 'e-0-4d90c2222d86d7ce637b93cb-1',
    'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/building/default_',
        'suffix': '.png'},
       'id': '4bf58dd8d48988d124941735',
       'name': 'Office',
       'pluralName': 'Offices',
       'primary': True,
       'shortName': 'Office'}],
     'id': '4d90c2222d86d7ce637b93cb',
     'location': {'address': 'Lisznyai utca 38.',
      'cc': 'HU',
      'city': 'Budapest',
      'country': 'Magyarország',
      'distance': 182,
      'formattedAddress': ['Budapest',
       'Lisznyai utca 38.',
       '1016',
       'Magyarország'],
      'labeledLatLngs': [{'label': 'display',
        'lat': 47.49314253872398,
        'lng': 19.035370820541853}],
      'lat': 47.49314253872398,
      'lng': 19.035370820541853,
      'postalCode': '1016',
      'state': 'Budapest'},
     'name': 'TradeTracker Hungary',
     'photos': {'count': 0, 'groups': []},
```

```
      'venuePage': {'id': '33332663'}}},
    {'reasons': {'count': 0,
      'items': [{'reasonName': 'globalInteractionReason',
        'summary': 'This spot is popular',
        'type': 'general'}]},
     'referralId': 'e-0-4d035e238620224bfe77a240-2',
     'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4
        sqi.net/img/categories_v2/building/default_',
         'suffix': '.png'},
        'id': '4bf58dd8d48988d124941735',
        'name': 'Office',
        'pluralName': 'Offices',
        'primary': True,
        'shortName': 'Office'}],
      'id': '4d035e238620224bfe77a240',
      'location': {'address': 'Lisznyai utca 38.',
       'cc': 'HU',
       'city': 'Budapest',
       'country': 'Magyarország',
       'distance': 187,
       'formattedAddress': ['Budapest',
        'Lisznyai utca 38.',
        '1016',
        'Magyarország'],
       'labeledLatLngs': [{'label': 'display',
         'lat': 47.49314030042303,
         'lng': 19.03527699700466}],
       'lat': 47.49314030042303,
       'lng': 19.03527699700466,
       'postalCode': '1016',
       'state': 'Budapest'},
      'name': 'Hamu és Gyémánt',
      'photos': {'count': 0, 'groups': []}}},
    {'reasons': {'count': 0,
      'items': [{'reasonName': 'globalInteractionReason',
        'summary': 'This spot is popular',
        'type': 'general'}]},
     'referralId': 'e-0-561f5c92498eef5e3a4bd4ca-3',
     'venue': {'categories': [{'icon': {'prefix': 'https://ss3.4
        sqi.net/img/categories_v2/building/default_',
         'suffix': '.png'},
        'id': '52e81612bcbc57f1066b7a3d',
        'name': 'Advertising Agency',
        'pluralName': 'Advertising Agencies',
        'primary': True,
        'shortName': 'Advertising Agency'}],
```

```
      'id': '561f5c92498eef5e3a4bd4ca',
      'location': {'address': 'Lisznyai u. 38.',
       'cc': 'HU',
       'city': 'Budapest',
       'country': 'Magyarország',
       'crossStreet': 'Fém u.',
       'distance': 203,
       'formattedAddress': ['Budapest',
        'Lisznyai u. 38. (Fém u.)',
        '1016',
        'Magyarország'],
       'labeledLatLngs': [{'label': 'display',
         'lat': 47.4932540068575,
         'lng': 19.035140993789717}],
       'lat': 47.4932540068575,
       'lng': 19.035140993789717,
       'postalCode': '1016',
       'state': 'Budapest'},
      'name': 'HG Media',
      'photos': {'count': 0, 'groups': []}}}],
   'name': 'recommended',
   'type': 'Recommended Places'}],
 'headerFullLocation': 'Krisztinaváros, Budapest',
 'headerLocation': 'Krisztinaváros',
 'headerLocationGranularity': 'neighborhood',
 'query': 'office',
 'suggestedBounds': {'ne': {'lat': 47.4947640027, 'lng': 19.041188
    444744087},
  'sw': {'lat': 47.489363997299996, 'lng': 19.03321155525591}},
 'totalResults': 4}
```

Methodology

Loading Data

Data of the neighbourhoods and district of Budapest is from a Wikipedia webpage, so we need to pull these data out of HTML tables. We do this by a *Python* library, that is *Beautiful Soup*. Name of the districts, and neighbourhoods of Budapest are saved into a *pandas dataframe*. The following packages are used during the data preparation:

`pandas, numpy, requests, bs4`.

Some data cleaning is necessary, because in the original table there are some neighbourhoods that correspond to more than one districts, hence we separate them by an easy string manipulation technique to get unique district-neighbourhood pairs. This part of the data preparation contains basic dataframe operations, and its result is a clean dataframe.

Then, with another *Python* client, *GeoPy*, we locate the centres of neighbourhoods using *Nominatim* and *Photon* geocoding services. Geopy makes it easy for Python developers to locate the coordinates of addresses, cities, countries, and landmarks across the globe using third-party geocoders and other data sources. We need to find the location of neighbourhoods to illustrate them on the map, but the exact coordinates are not considered in the further calculations. At this point, we have a full list of neighbourhoods of Budapest with geometric location data. With the Python package *folium*, we can illustrate them on a map.

It is followed by a lot of Foursquare API calls to get information from the venues in the neighbourhoods. Foursquare is a webpage to explore venue information near a specific location, that can be exact address or geographic coordinates. We focus on the number of cafés, university buildings and office buildings near the centres of the neighbourhoods. The `https://api.foursquare.com/v2/venues/explore` API has a flag called "categoryId", that can filter the nearby venues onto a certain category. We use the following ID-s:

- office: 4bf58dd8d48988d124941735,
- university: 4d4b7105d754a06372d8125,
- cafe: 4bf58dd8d48988d16d94173

Clustering

We cluster these information as a three dimensional space to locate the similar neighbourhoods using k-means clustering algorithm. The neighbourhoods are mapped into the space $\mathbb{R}^3$ by the corresponding (office, university, cafe)-triplet.

To perform a k-means clustering, the first thing is to estimate the best choice of the value k. With the scikit-learn machine learning library, we can easily build k-means models with different values of k with a simple for loop, saving the distortions of the models. Using the elbow rule, the distortion plot will suggest the ideal value of k.

Results

Loading Data

The webpage about neighbourhoods of Budapest contains a very good table about the list of neighbourhoods. It is a well structured html table, so it is easy to read and load it with Beautiful Soup library into a dataframe, like Table 1.

Budapest is divided into 23 different districts and the 1-level smaller administrative units are the neighbourhoods. After a small data cleaning, we split some row into unique (district, neighbourhood) pairs, see Table 2.

After that, latitude and longitude coordinates are appended to the previous dataframe (Table 3), using *GeoPy* client.

With the *folium* package, we can illustrate these neighbourhoods on a map (Figure 4).

Appending the number of cafés, university buildings and office buildings near the given coordinates, we get a dataframe like Table 4.

	district	neighborhood
0	Budapest II. kerülete	Adyliget
1	Budapest XVII. kerülete	Akadémiaújtelep
2	Budapest XI. kerülete	Albertfalva
3	Budapest XVIII. kerülete	Almáskert
4	Budapest XIV. kerülete	Alsórákos
5	Budapest XIII. kerülete	Angyalföld
6	Budapest III. kerülete	Aquincum
7	Budapest III. kerülete	Aranyhegy-Ürömhegy-Péterhegy
8	Budapest XVI. kerülete	Árpádföld
9	Budapest XXII. kerülete	Baross Gábor-telep

Table 1: Dataframe from the Wikipedia webpage

	district	neighborhood
0	I	Gellérthegy
1	I	Krisztinaváros
2	I	Tabán
3	I	Vár
4	I	Víziváros
5	II	Adyliget
6	II	Budakeszierdő
7	II	Budaliget
8	II	Csatárka
9	II	Erzsébetliget

Table 2: Dataframe unique (district, neighbourhood) pairs

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

Table 3: Dataframe with geographic coordinates

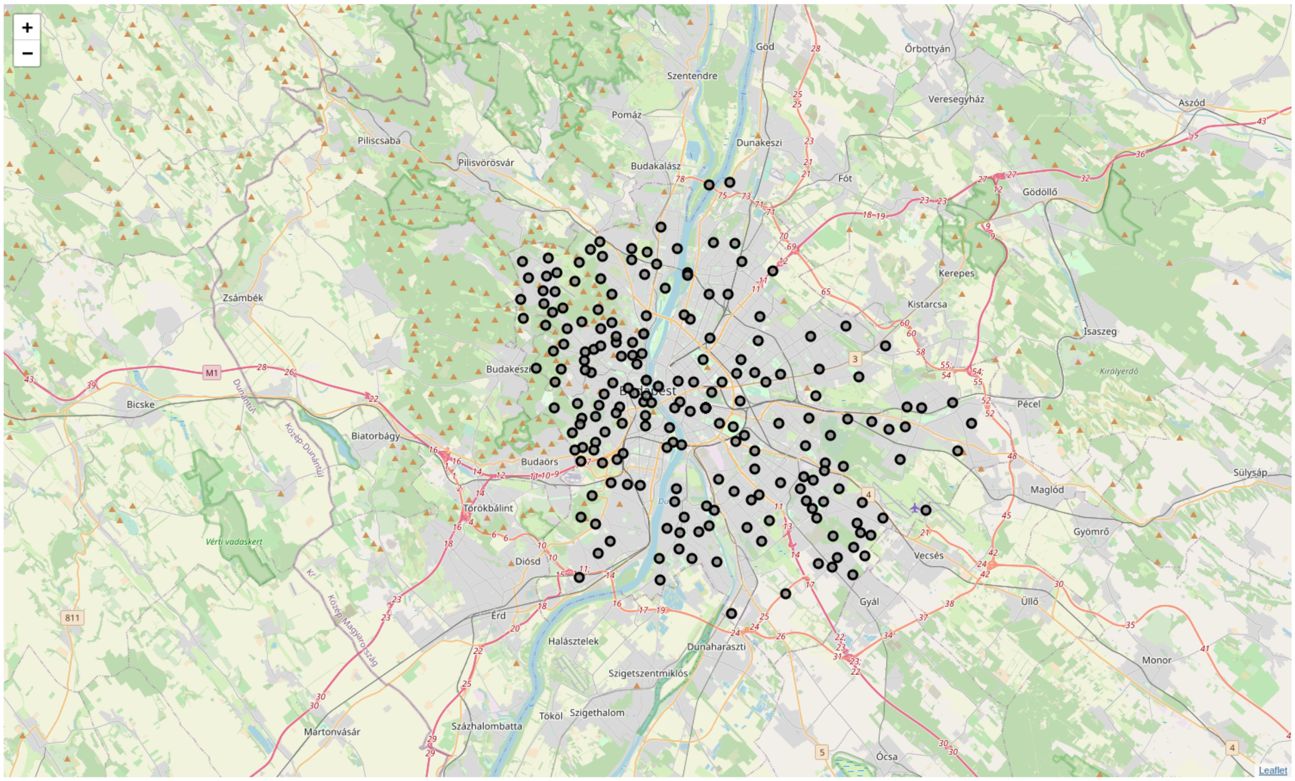

Figure 4: Neigbourhoods of Budapest

Clustering

Creating models with $k = 2, 3, \ldots 20$, we get the distortion plot on Figure 5.

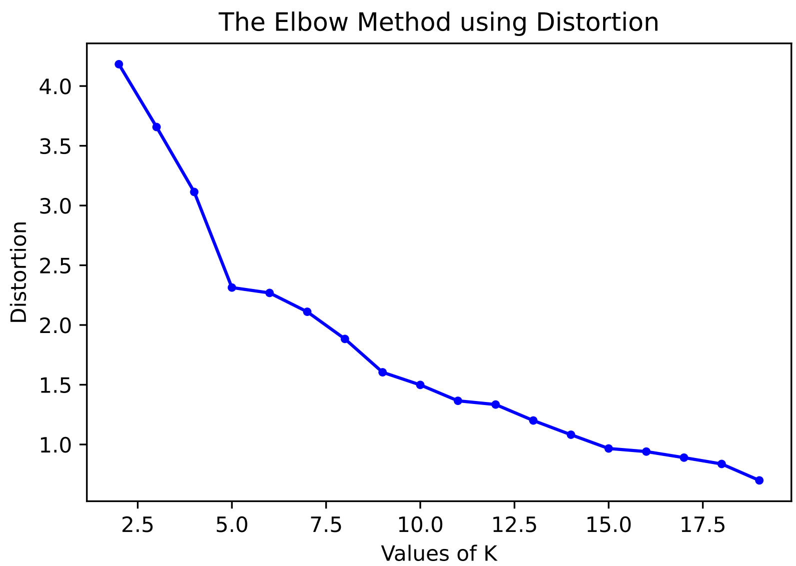

Figure 5: Distortion diagram

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

	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

Table 4: Dataframe with number of venues

will focus on 5 different clusters. Clustering the triplets, we get the clusters that are illustrated in the Figure 6.

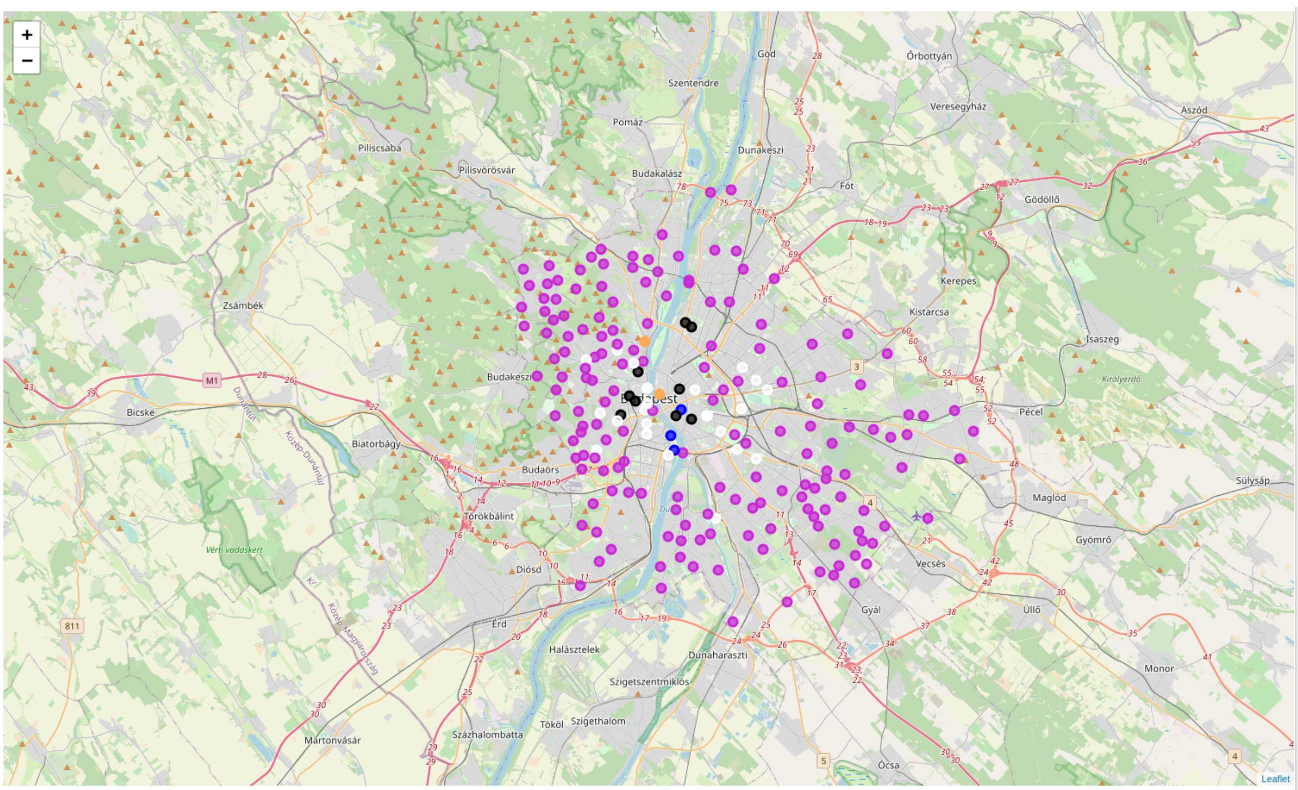

Figure 6: Clustering of neighbourhoods by the number of universities, cafés and offices.

We can summarise the size of clusters in the Table 5.

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

Table 5: Size of clusters

Discussion

We investigate the above clusters one-by-one. Let us start with the **cluster with label 3**. Most of the neighbourhoods lie in this cluster, however this cluster is the least suitable for us. This one contains neighbourhoods whose (universities,cafés,offices) triplet has low numbers, the frequencies of the specific venues considering the 164 neighbourhoods can be observed in the Figure 7. None of these neighbourhoods can be suggested to open a new office, so most of the neighbourhoods do not require further investigations. These neighbourhoods are coloured as **purplish magenta**.

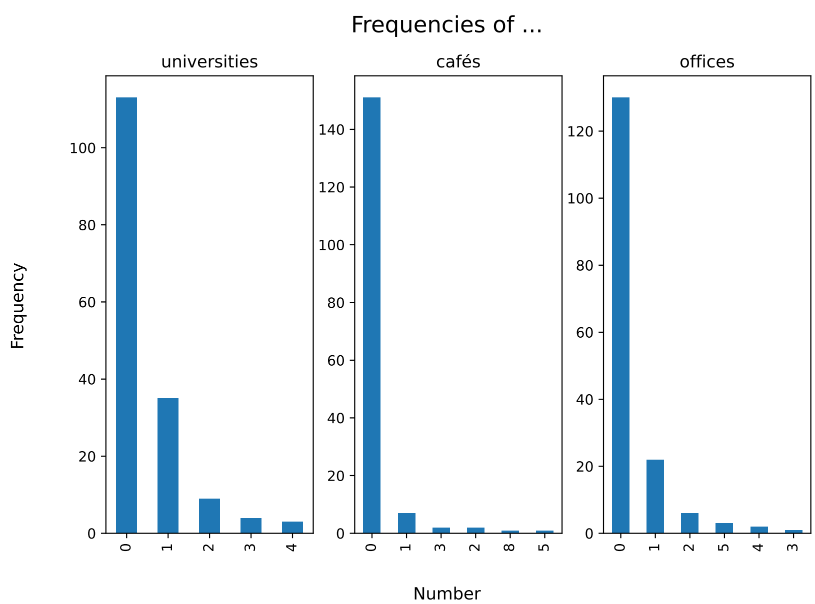

Figure 7: Frequencies of the numbers of venues in cluster 3.

Let us continue with cluster 0, that is represented on the map with white dots. It contains spots where there are a few university buildings and quite small number of cafés and universities, see in Figure 8. This cluster is neither recommended, because of the small number of important factors.

Cluster 2 is a very small cluster with only 3 elements, see Table 6. These spots has a large number of university buildings, however the number of offices and cafés are quite small. This is a very unique situation in Budapest, because these areas contain the main buildings of the

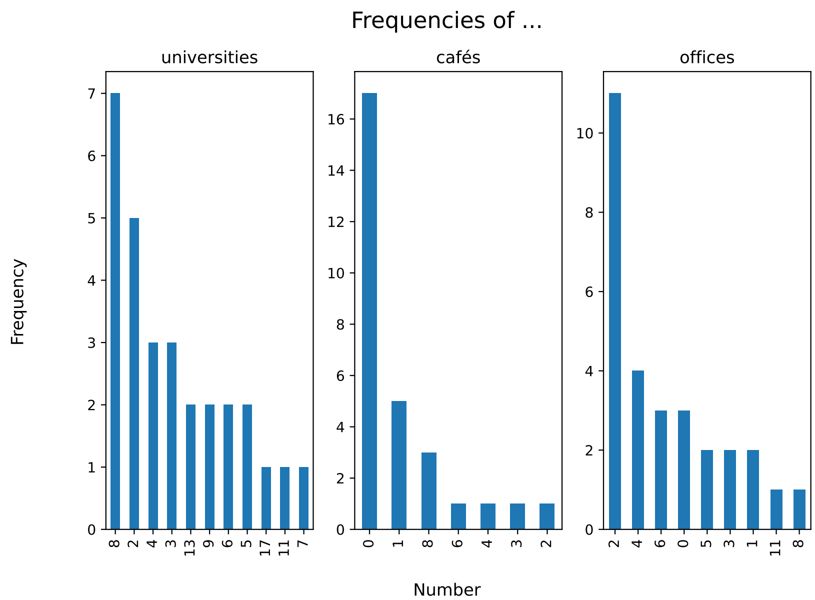

Figure 8: Frequencies of the numbers of venues in cluster 0.

	Cluster Labels	district	neighborhood	Latitude	Longitude	office	university	cafe
71	2	VIII	Palotanegyed	47.492185	19.066103	6	49	13
94	2	XI	Infopark	47.470484	19.060338	33	49	2
99	2	XI	Lágymányos	47.478441	19.057715	4	61	1

Table 6: Elements of cluster with label 2

two biggest universities of Budapest (ELTE and BME) and besides this campus region there is a huge office area, called Infopark. In this are there are few number of reported cafés, because the cafés and restaurants in these neighbourhoods are inside the universities and office building that focus students and employees, respectively. These spots are highly recommended to open new office, but it is an overwhelmed area, and to open offices is not too cheap.

Cluster 4 and **Cluster 1** are similar clusters. Their common property is that, they contain large number of office building 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. On the other hand, Cluster 1 is more balanced with a lot of office building and moderate number of cafés and university buildings.

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

Conclusion

In this report we used a clustering algorithm to find a recommendation to the business problem described in the first section, namely what is an ideal spot to open a new office for a new start-up IT company. We have considered the number of cafes, universities and other office buildings in the neighbourhoods, but of course with more data we can consider other important parameters, too such as renting price, noise pollution, distance from the main public transport hubs, and so on... Using a classical k-means algorithm, we could suggest an area (and prioritize others) that is ideal for a new office.