Best locations for establishing new hotels in Budapest

Ferenc Farkas, PhD (2019-02-24)

Applied Data Science Capstone by IBM/Coursera (Week 2)

Abstract

There is a steady state growing in the tourism of Budapest and similar growth is expected in the coming years. Thus, there is a business case to establish new hotel(s) in the city. And several stakeholders are eager to do so. This analysis try to explore the current market and propose possible locations for establishing new hotels in Budapest in aiming stakeholders to choose the optimal location for their investment.

Table of contents

1. Introduction

Every year, more and more people visit Budapest, the capital of Hungary and, even better, those visitors spend increasingly more time in the city (Budapest tourism). The passenger traffic of Budapest International Airport (BUD) increased heavily in the last 5 years (annual growth rate well above 10%) and by the end of this year is expected to almost double compared to 2013 (BUD traffic). Several developments have been carried out in the airport and more are planned for the near future. As a result, Budapest Airport has been awarded the Skytrax title for "Best Airport in the region" for the fifth time in a row. In the history of the most prestigious award in the industry based on passengers' votes, winning the title in five consecutive years by the same airport in the region has been unprecedented. (Skytrax award).

And this impressive increase should continue, as Budapest took first place in European best destination' voting for "BEST EUROPEAN TRAVEL DESTINATION". The notification of the winning the prize states that "no other winning European travel destination has received such international support, i.e. votes from outside the country concerned. 77% of the votes in support of Budapest came from outside Hungary, in particular the UK, USA, Germany, France, Austria and Italy (BEST EUROPEAN TRAVEL DESTINATION). The EU remains an attractive destination for Chinese tourists, and while the U.K. is still the most popular looking at sheer numbers, Hungary's 25.1% growth in arrivals in 2018 puts the country in third place in terms of relative growth (Chinese tourist arrivals).

Fig. 1. *Budapest was the winner for "BEST EUROPEAN TRAVEL DESTINATION"*

2. Business understanding

The increasing number of tourists visiting Budapest need to be accommodated somewhere. Thus, there is a great potential in establishing new hotels in Budapest in the coming years. But before starting to build a new hotel (either from the ground or by renovating an existing old building) requires a good understanding of the best locations which guarantee a good percentage of occupancy of the hotel over the whole year. For this reason, one should avoid locations where there are already plenty of hotels, and choose locations where tourists are still frequent, but hotels are rear. Hotel location shall also count the proximity of the metro station and popular sites, like landmarks, monuments, historic sites, museums, and even spas. This data analysis tries to help stakeholders in selecting the best locations in Budapest for establishing new hotels.

3. Data collection and analysis

Based on definition of our problem, factors that will influence our decision are:

- number of existing hotels in the neighborhood
- distance to nearest metro station
- distance from the city center
- number of nearby popular sites (landmarks, monuments, historic sites, museums, spas)

For gathering data we use the Foursquare API. Subcategory IDs are taken from Foursquare web site https://developer.foursquare.com/docs/resources/categories. Google map for obtaining geographical location was also used.

Finding the geographical location of Budapest

Budapest geographical location can be obtained from coordinates of Budapest which represents the so called 0km statue from where the counting is started for all main motorways and roads of Hungary going out of Budapest. This 0km statue is on the Clark Adam square near the Chain Bridge (see Fig. 2.) and Budavár Castle (from where the above picture was taken). On the picture below you can see the 0km statue with the funicular in the back which takes you to the Budavár castle.

Fig. 2: The 0 km statue

Collecting geographical locations for metro stations in Budapest

There is a direct relation between flat prices and proximity of metro station (Subway proximity effect), thus during analysis the distance of the metro stations are also counted. For this reason geographical location of metro stations are also queried using Foursquare API with category 'metro station', explore endpoint, and 10km radius. The query result is converted to a pandas dataframe. The data provided by Foursquare API need be cleaned, thus unnecessary rows from the dataframe is removed. The total number of metro stations in Budapest is 52 (see Fig. 3.), but the list is reduced to 48 because locations where you can switch to another metro is counted only once.

Fig. 3: Budapest metro map

Collecting geographical locations of popular sites in Budapest

From my personal experience, tourist are in favor of hotels which are in close proximity to most of the popular sites, like monuments, landmarks, historic sites, museums and spas in the city. Spas is included here because Budapest is famous of the dozens of spas, one of the famous being the Széchenyi thermal bath in the City Park.

The following categories are used to query the popular sites using Foursquare API with explore endpoint, and 15km radius:

- Monument / Landmark

- Historic Site
- Museum
- Spa

The obtained results are converted to pandas dataframe and merged together removing duplicated sites (which have identical geographical locations).

As it can be seen on the heat map shown in Fig. 4. the popular sites are crowded in the city center, but several popular sites can be found outside of it.

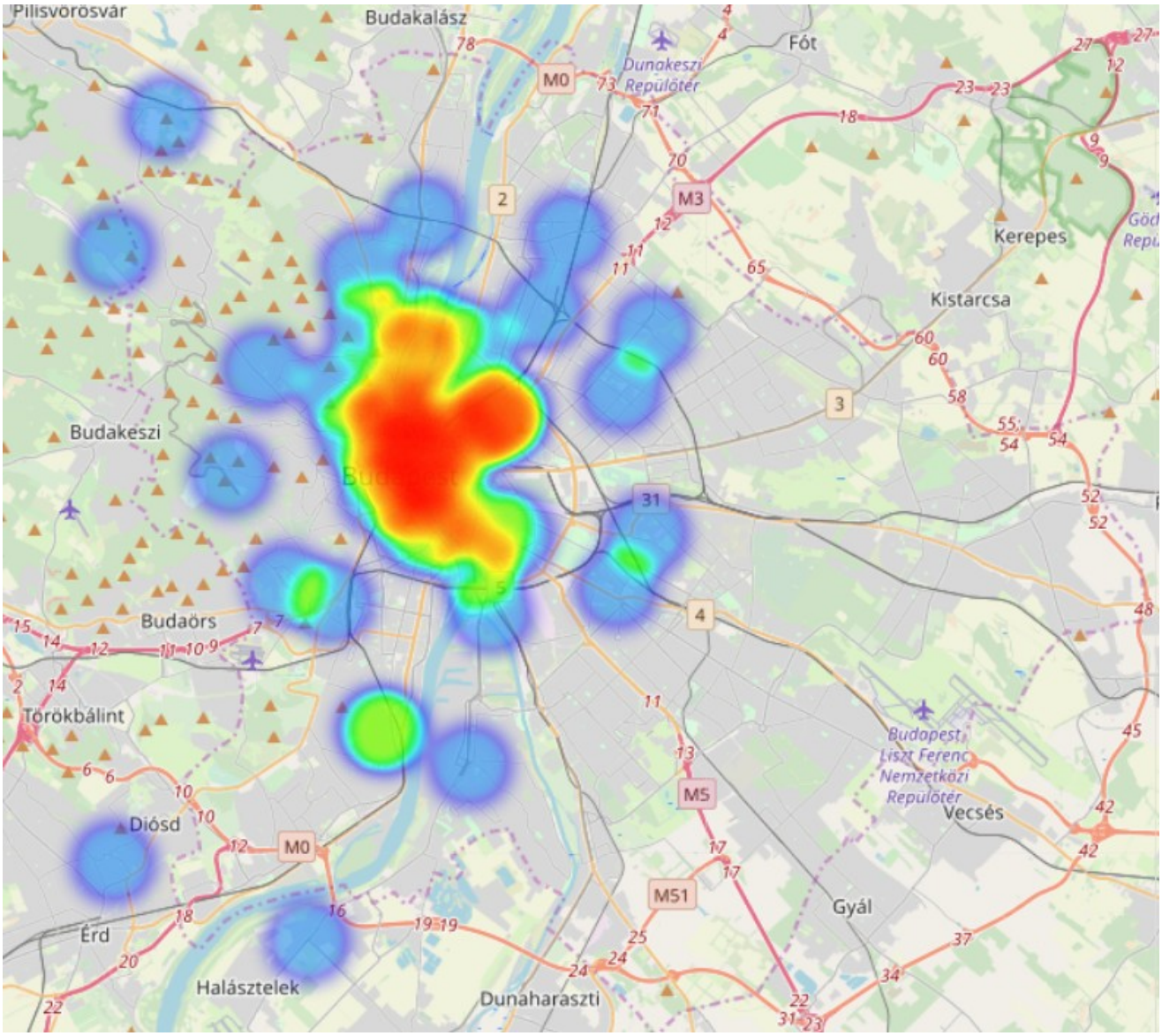

Fig. 4: Heat map of popular sites (landmarks, monuments, historical sites, museums, spas) in Budapest

Collecting geographical locations for hotels

At first, the geographical location of each district for querying the hotel locations was used. Unfortunately, due to limitation of max. 100 returned findings of Foursquare API that was not a possible choice. After

several hours of struggling the final idea was to create a fine grid around the downtown (9 geographical locations with 708m radius, 500m distance between the grid points) and create additional 5 geographical locations outside of this fine grid, one to north, one to south, one to west, and one to east, each at 5km from the edge of the fine grid shown on the map of Fig. 5. Because Budapest has an elongated shape toward its airport, then we should add another location to south-east. Some small outside portion of outskirt districts are not covered on the east side of Budapest, but these locations are far away from the city center and no nearby metro stations, so we are not interested in these locations anyway. To have a better understanding of the above description see the map created with folium in Fig. 5.

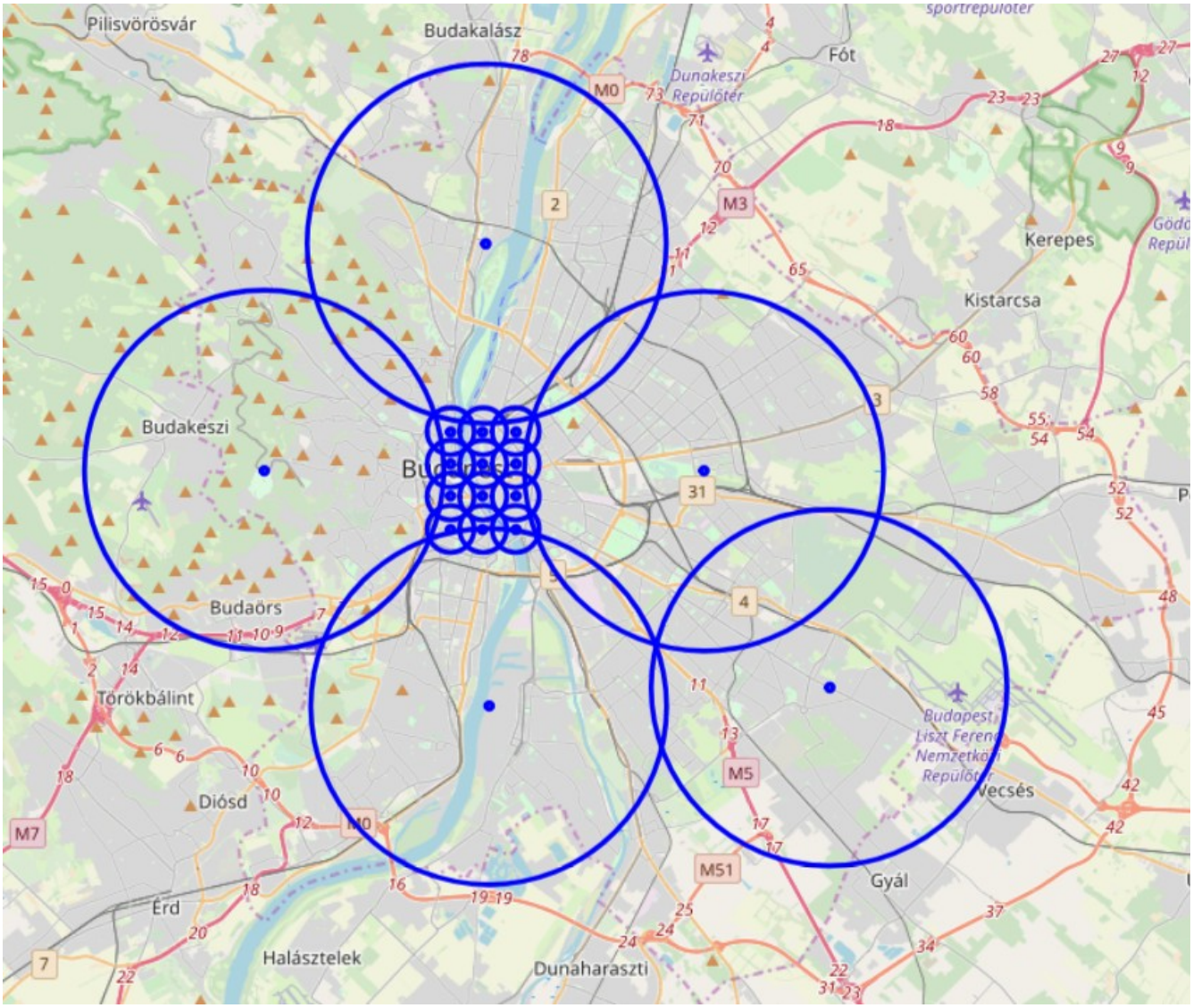

Fig. 5. Geographical locations with radius for querying hotel locations in Budapest

The query results are cleaned (only 'hotel' subcategories are kept using 'shortname' of venues) converted to pandas dataframe and merged together. In total 314 hotels have been found with Foursquare API

From the map above we can see that there are overlapped regions, thus duplicated hotel locations (which have identical geographical data) are removed.
As one would expect the dense regions of existing hotels are in the proximity of the city center as one can see on the heat map created with folium in Fig. 6.

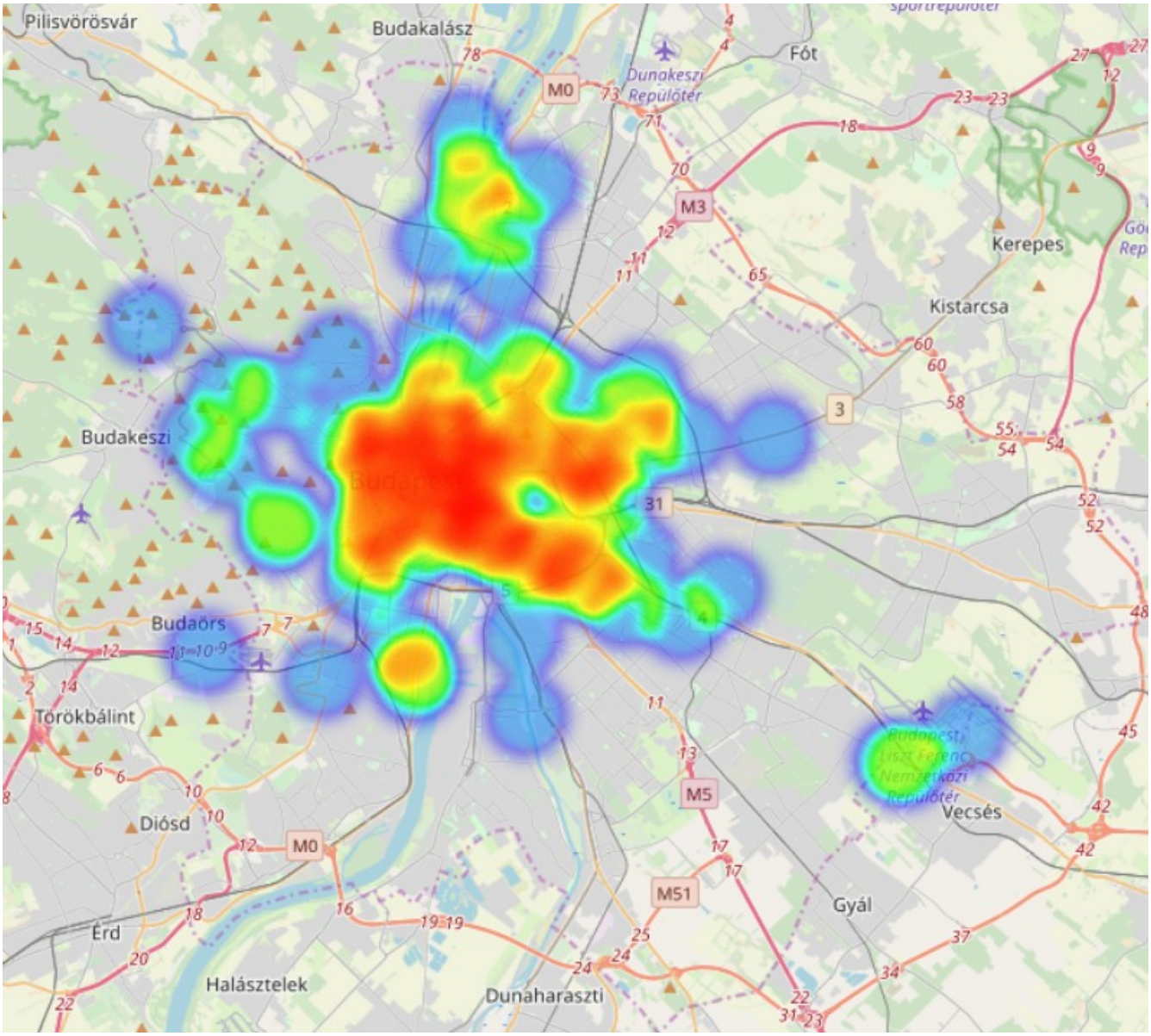

Fig. 6. Heat map of hotels in Budapest

4. Methodology

In this project we will direct our efforts on detecting areas of Budapest that have low hotel density, particularly those which have high number of popular sites (landmarks, monuments, historic sites, museums, and spa) in their proximity, are close to a metro station, and are not far away from the city center.

In the first step we have collected the required data:

- location of metro stations,
- the location of popular sites, and
- location of every hotels in Budapest.

In the second step in our analysis we will be identifying ***high hotel density*** and ***low hotel density*** areas. We already could identify that visually using the heat map created with folium as presented earlier in Fig. 6. However, this is very high level, so we should provide more exact locations for the stakeholders.

In third step we will use a machine learning algorithm to find locations which are not so dense regarding existing hotels.

For this purpose we choose the **Density-based spatial clustering of applications with noise** (**DBSCAN**) algorithm which ia a density-based clustering algorithm. In the model creation the article and acompanying code of Geoff Boing was used [1].

Because Budapest has big areas where hotels cannot be established, like the river Danube, big islands, hills, public parks, forests, or fields without infrastructure, identifying regions without any hotel might not be the best option. Thus, we will look for low density locations which are represented by outliers (noise) of the DBSCAN algorithm. We will also check locations which are the borders of the high density regions (clusters defined by DBSCAN). Based on the market analysis of stakeholders, a **cluster of hotel** is represented by 3 hotels being not more than 250 m away from each other.

When performing DBSCAN clustering methodology we are not interested in the number of clusters obtained, only the outliers and borders of the clusters. Based on the requirements stated by the stakeholders, the existing hotels can be categorized as:

1. **'*core*' hotel**, a hotel which has at least another 2 neighboring hotels within 250 m distance
2. **'*border*' hotel**, a hotel which has another hotel within 250 m being a 'core' hotel
3. **'*outlier*' hotel**, a hotel which has no no member of cluster ('core' or 'border' hotel) within 250 m

Note: Because we set the min. number of hotels to form a cluster to 3, two hotels being 250 m away from each other but not being neighbors of 'core' or 'border' hotels still represents 'outlier' hotels.

For model creation the DBSCAN of scikit-learn library is used.

6. Results

The DBSCAN algorithm – with hyperparameter of epsilon equal to 250m and number of members in a cluster of 3 – found 10 clusters, 107 'outlier' hotels, and 10 'border' hotels. As one would expect clusters are formed in the city center as shown on the map with gray in Fig. 7. 'Border' hotels are shown with blue, while 'outlier' hotels are shown with red. Thus, our analysis shows that there are several locations close to city center with low density hotel regions (defined by a circle with 250 m radius having only one hotel). These

locations are close to metro stations (within 10 minute walking distance) which has several nearby popular sites (within 15 minutes walking distance).

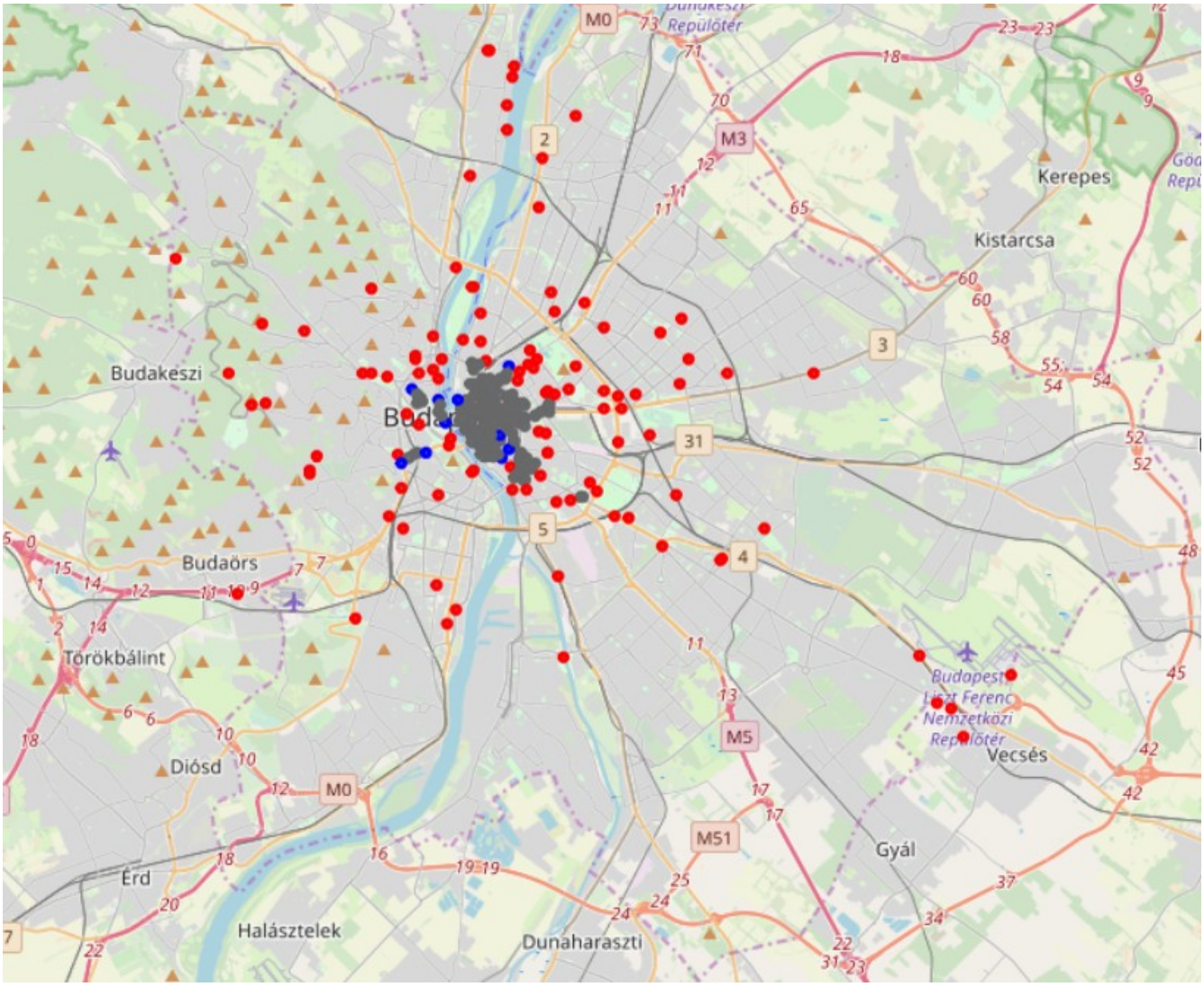

Fig. 7. DBSCAN result: 'Core' hotels are with gray, 'border' hotels with blue, and 'outlier' hotels with red

According to stakeholders they are interested in hotel locations not further away than 3 km from the city center. Thus, we impose the following limitations to the potential new hotel locations:

- closest metro station max. 800m (10 minute walking distance from the hotel)
- distance from city center is max. 3 km (15-20 minute to reach the hotel from the city center)
- counting the number of popular sites being at max. 1200 m from the hotel location (15 minute walking distance from the hotel)

With that limitation the number of potential new hotel locations is 36, ordered based on the distance from the city center and the number of reachable popular sites. In Table 1. the first 20 potential locations are

shown out of the 36 (these represent potential new hotel locations at max. ~2 km from the city center). The last column, called 'Singleton', states whether at that location there is only one hotel within 250 m (is an 'outlier' hotel), or there are two ('border' hotel) and other hotels might exist at further distance.

Table 1. List of locations for establishing new hotels

	Latitude	Longitude	Dist. from metro	Dist. from center	Nr. of nearby popular sites	Singleton
0	47.502768	19.046838	0.298830	0.762249	38	Yes
1	47.493722	19.063393	0.256393	0.845260	10	Yes
2	47.496589	19.041748	0.649746	0.933512	42	Yes
3	47.492477	19.044258	0.657939	0.950875	32	Yes
4	47.502888	19.039457	0.413285	1.225130	39	Yes
5	47.489924	19.066590	0.356582	1.297544	7	Yes
6	47.487460	19.064133	0.280573	1.389737	10	Yes
7	47.484063	19.053157	0.169639	1.542466	15	Yes
8	47.483852	19.052578	0.174630	1.568391	15	No
9	47.508530	19.038986	0.220343	1.632684	31	Yes
10	47.496548	19.031788	0.701577	1.676900	25	Yes
11	47.508134	19.070548	0.152365	1.682670	9	Yes
12	47.513068	19.057917	0.315421	1.709346	9	Yes
13	47.485042	19.067252	0.230270	1.743709	10	Yes
14	47.511679	19.066470	0.346790	1.792968	10	Yes
15	47.510270	19.070980	0.158904	1.873156	12	Yes
16	47.494425	19.078543	0.139480	1.883273	1	Yes
17	47.510831	19.037091	0.488550	1.917431	18	Yes
18	47.499143	19.027166	0.250171	2.021608	19	Yes
19	47.505612	19.028881	0.644720	2.072689	17	Yes
20	47.493806	19.080957	0.326012	2.075178	1	Yes

In Fig. 8. the 36. potential new hotel locations are shown with circles, the middle of a blue circle representing the 'outlier' hotel location with a radius of 250 m, while the yellow circles are representing the 'border' hotel location with a radius of 250 m. Red points represents the location of popular sites.

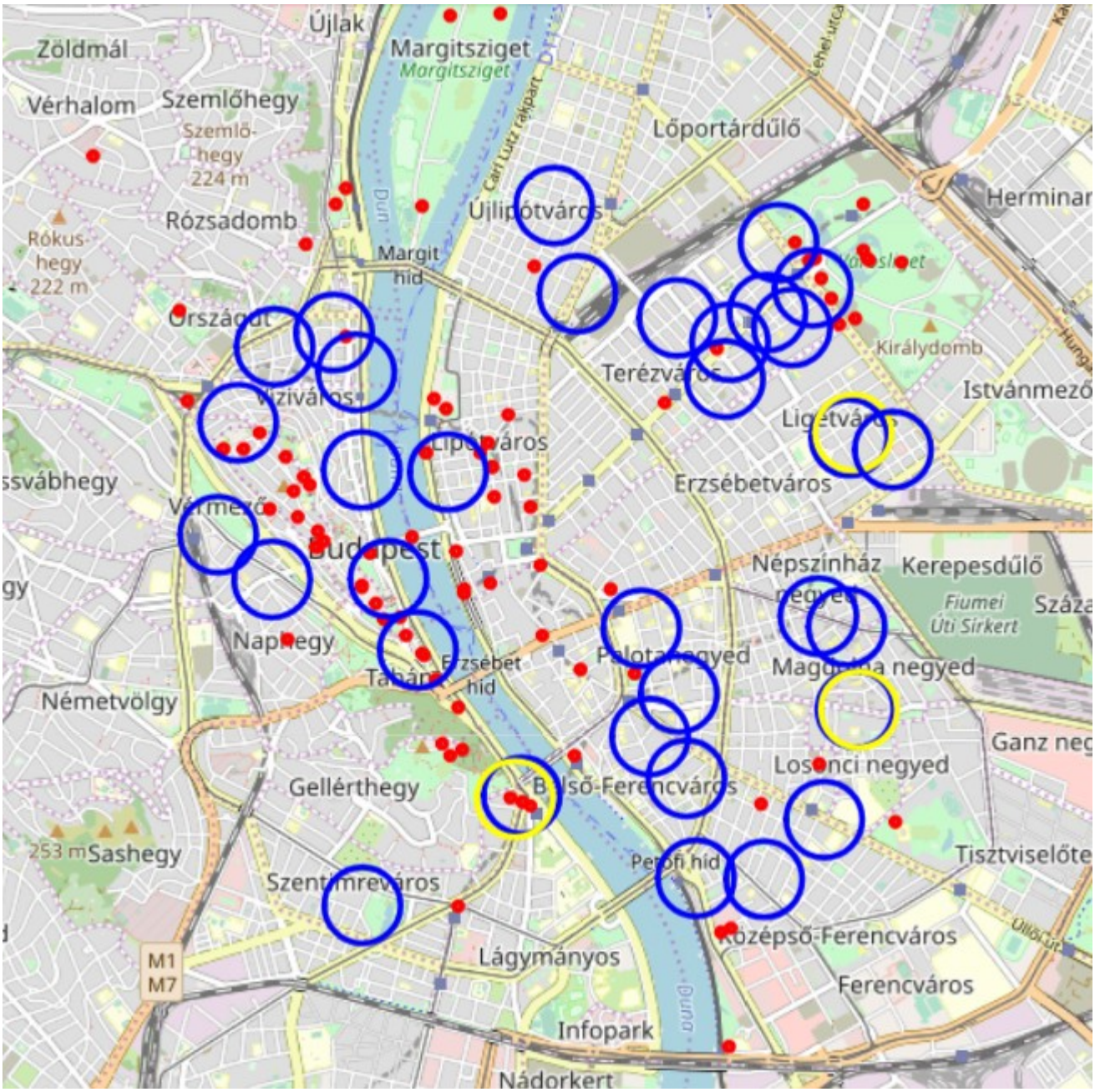

Fig. 8. Potential new hotel locations (blue shows the 'outlier', and yellow the 'border' hotel location)

6. Discussion

With a simple approach like the DBSCAN algorithm we could localize area of Budapest which has low density of hotel. However, more locations could be found if not only metro stations, but tram stations are also counted. Moreover, there are locations where there are not hotels at all, which are not

identified by DBSCAN algorithm. However, identifying those locations are not simple, because Budapest has big areas, even close to the city center, where hotels cannot be established, simply because there is the river Danube, there are big islands, hills, public parks, forests, or fields without any infrastructure. That is why the simplistic approach of looking for outliers of DBSCAN was chosen.

7. Conclusion

Purpose of this project was to identify areas in Budapest close to city center with low number of hotels in order to aid stakeholders in narrowing down the search for optimal location for a new hotel. When looking for the optimal location the distance to the metro station and the number of nearby popular sites were also considered.

Because Budapest has big areas, even close to the center, which are not suitable for establishing hotels, like the river Danube, big islands (the most famous being the Margaret island near the city center or the Obuda island, the latter being known from the Sziget Festival)), hills (like Gellért with Citadella), public parks (like City Park), forests, or fields without infrastructure, the simplest approach was chosen, namely we looked for locations where at least one hotel exist, but no other hotel within 250 m (‘outlier’ hotel) or max. one hotel (‘border’ hotel).

Our analysis found 36 such locations not further than 3km from the city center (out of which 33 are ‘outlier’ hotels, and 3 are ‘border’ hotels. Final decision on optimal hotel location will be made by stakeholders based on specific characteristics of neighborhoods and locations in every recommended location.

References

[1] Boeing, Geoff. “Clustering to Reduce Spatial Data Set Size.” SocArXiv, 22 Mar. 2018
https://osf.io/preprints/socarxiv/nzhdc/
(*Note: The rest of the references are embedded in the text as hyperlinks.*)