Coursera - Applied Data Science Course – Final Assignment

CAPSTONE - THE BATTLE OF NEIGHBORHOODS PROJECT

Peter Hrabovszki

2019 JULY

Table of Contents

Introduction

The real estate prices have significantly increased in Budapest, Hungary in the last couple of years. It is very difficult to find flats for reasonable prices nowadays. As the prices are really high one needs to do a very thorough discovery before buying or renting a flat. When choosing the neighborhood and the flat different age groups chose flats based on different factors and it is difficult to find information in one place. For example for youngers it might be more important to have many places near by where they can live social life ie. bars, restaurants etc. However, for younger couples for example it might be more important if day care for children, schools are available but such data or view is not available combined with other type of venues and prices. Using this project my goal is to help those who consider to buy or rent a flat in Budapest. Specially helping younger couples with analyzing the different neighborhoods in Budapest so they can hopefully easier find the right flat.

Data

I have used the following data sources in the analysis.

- **Budapest district listing** data has been pulled from Wikipedia: https://en.wikipedia.org/wiki/List_of_districts_in_Budapest The latitude and longitudes will be pulled using the geocoder service.
- Venue data has been pulled via **Foursquare API** that contains several categories of venues. The type and location of the venue has been used in the analysis.
- *Data about **children daycare, schools** has been pulled from http://budapest.imami.hu. This data contains the address and name of the venue. The longitude and latitude has been pulled based on the address.

Methodology

Data collection and cleansing

The wikipedia data with the Budapest district data contained the listing of the districts but no latitude, longitude data. It has been pulled using the geopy library. Budapest has 23 districts

and one area that is not part of any district (Margaret-island) and it has no inhabitants therefore it was excluded.

District number	District name	Neighborhoods	Sights
I.	Várkerület(Castle District)	Buda Castle, Tabán, Gellérthegy, Krisztinaváro...	Buda Castle, Matthias Church, Hungarian Nation...
II.	none	Adyliget, Budakeszierdő, Budaliget, Csatárka, ...	Tomb of Gül Baba, Mechwart Park, Cave of Szeml...
III.	Óbuda-Békásmegyer	Óbuda, Aquincum, Aranyhegy, Békásmegyer, Csill...	Ruins of Aquincum, Aquincum Military Amphithea...
IV.	Újpest(New Pest)	Újpest, Megyer, Káposztásmegyer, Székesdűlő, I...	Queen of Heavens Church, Synagogue of Újpest, ...
V.	Belváros-Lipótváros(Inner City-Leopold Town)	Inner City, Lipótváros	Parliament, Hungarian Academy of Sciences, Gre...
VI.	Terézváros(Theresa Town)	Terézváros	Andrássy Avenue, Hungarian State Opera House, ...
VII.	Erzsébetváros(Elizabeth Town)	Erzsébetváros	Dohány Street Synagogue, Rumbach Street Synago...
VIII.	Józsefváros(Joseph Town)	Józsefváros, Kerepesdűlő, Tisztviselőtelep	Hungarian National Museum, Erkel Theatre, Orcz...
IX.	Ferencváros(Francis Town)	Ferencváros, Gubacsidűlő, József Attila-lakótelep	National Theatre, Palace of Arts, Kálvin Squar...
X.	Kőbánya(Quarry)	Felsőrákos, Gyárdűlő, Keresztúridűlő, Kőbánya-...	Népliget (People's Park), Planetarium, St. Lás...
XI.	Újbuda(New Buda)	Albertfalva, Dobogó, Gazdagrét, Gellérthegy, H...	Gellért Hill, Citadella, Liberty Statue, Budap...
XII.	Hegyvidék(Highlands)	Budakeszierdő, Csillebérc, Farkasrét, Farkasvö...	Elizabeth Lookout Tower, Normafa
XIII.	none	Újlipótváros, Angyalföld, Vizafogó, Népsziget,...	Comedy Theatre, St. Margaret of Árpád House Ch...
XIV.	Zugló	Alsórákos, Herminamező, Istvánmező, Kiszugló, ...	City Park, Heroes' Square, Zoo, Széchenyi Medi...
XV.	none	Rákospalota, Pestújhely, Újpalota	Water Tower
XVI.	none	Mátyásföld, Sashalom, Cinkota, Rákosszentmihál...	Mátyásföld Airport,
XVII.	Rákosmente	Rákoskeresztúr, Rákoscsaba, Rákoscsaba-Újtelep	Statue of Heroes, Statue of Pope John Paul II,...

Figure 1 - Budapest Districts

Figure 2 - Budapest Districts on the map

The data about the daycares in Budapest has been pulled using the request library and the geospatial data has been added using the geopy library. For 60 daycares no geospatial data was returned so such entries have been removed. The final dataset contained 506 entries.

Name	Address	District	Latitude	Longitude
Budavári Kolping Katolikus Óvoda	1014 Budapest, Dísz tér 3.	1	47.499511	19.035977
Naphegyi Napköziotthonos Óvoda	1016 Budapest, Nyárs u. 2-4.	1	47.493485	19.031064
Vízivárosi Napköziotthonos Óvoda	1011 Budapest, Iskola u. 44.	1	47.504468	19.037362
Akadémiai Óvoda és Bölcsőde	1022 Budapest, Bimbó u. 33.	2	47.429733	19.016248
Bolyai Utcai Napközi Otthonos Óvoda	1023 Budapest, Bólyai u. 15.	2	47.518883	19.028548

Figure 3 - Daycares in Budapest Sample

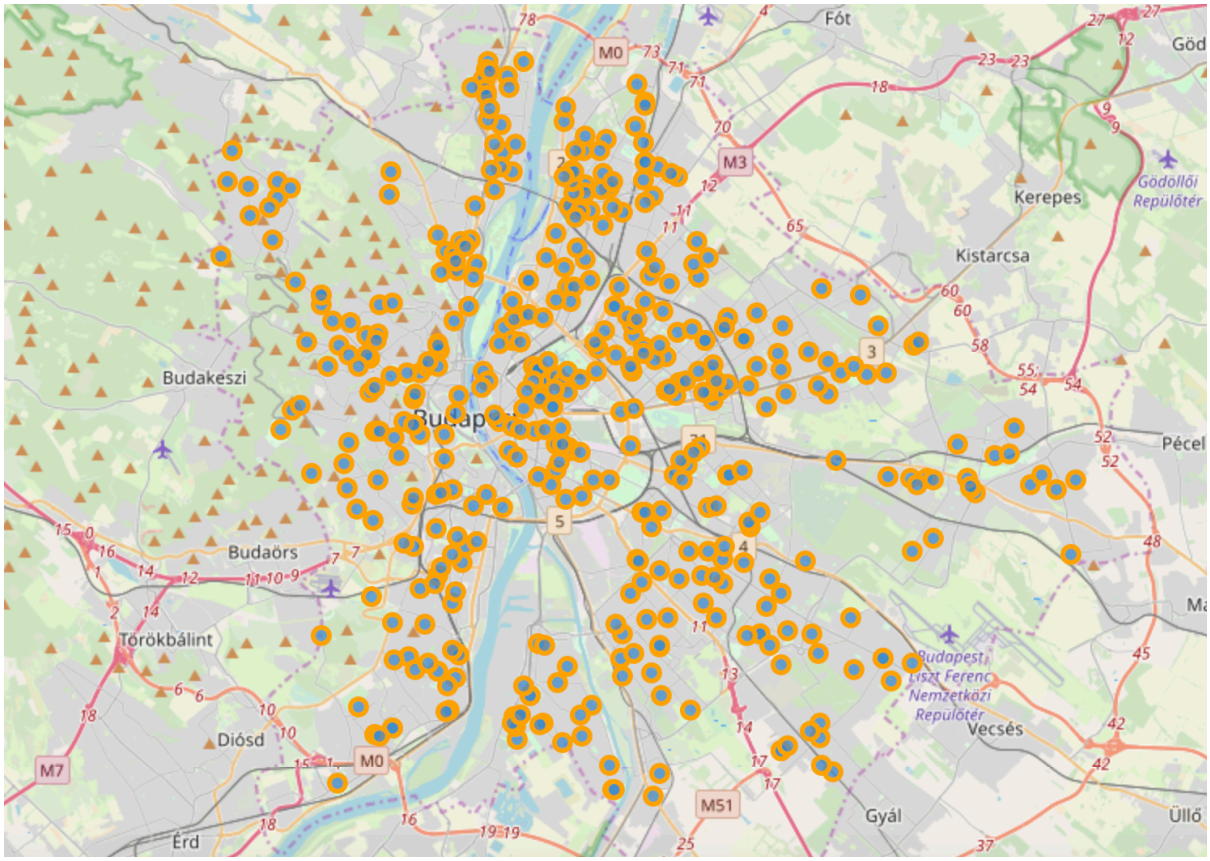

Figure 4 - All daycares on the map

The datasets have been saved to csv so they could be imported again as needed.

The TOP100 Venue data has been pulled using the Foursquare API for all districts in Budapest. The API returned 513 venues. By looking at the Foursqaure data retrieved it turned out that no 'daycare' venue type was included. Therefore the daycare data crawled from the internet has been combined so daycares became a new 'venue type'

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
1	47.4968	19.0375	Budavári Palota	47.496198	19.039543	Castle
1	47.4968	19.0375	Magyar Nemzeti Galéria | Hungarian National Ga...	47.496082	19.039468	Art Museum
1	47.4968	19.0375	Várhegy	47.497570	19.038747	Scenic Lookout
1	47.4968	19.0375	Alagút-tető	47.498166	19.039510	Scenic Lookout
1	47.4968	19.0375	Savoyai Terasz	47.496455	19.040133	Scenic Lookout
1	47.4968	19.0375	Zhao Zhou Teashop & Lab	47.497354	19.041026	Tea Room

Figure 5 - Sample output from Foursquare API

10. District	47.4821	19.1575	Komplex ovoda, altalanos Iskola, Keszsegfejles...	47.471630	19.122916	Daycare
10. District	47.4821	19.1575	Kobanyai onkormanyzat Napkozi Otthonos ovoda	47.409647	19.008769	Daycare
10. District	47.4821	19.1575	Mackokucko Angol-Magyar ketnyelvu ovoda	47.501529	19.167858	Daycare
10. District	47.4821	19.1575	ovoda	47.472947	19.146371	Daycare
10. District	47.4821	19.1575	Jazmin ovoda	47.480339	19.132222	Daycare
11. District	47.4593	19.0187	Megafitness	47.457480	19.019410	Gym
11. District	47.4593	19.0187	Don Pepe	47.461886	19.015849	Pizza Place
11. District	47.4593	19.0187	Neszmelyi ut (150,250)	47.457404	19.020078	Bus Stop
11. District	47.4593	19.0187	ormezo Gyogyszertar	47.461726	19.015735	Pharmacy
11. District	47.4593	19.0187	Pikolo etterem	47.459563	19.013307	Food
11. District	47.4593	19.0187	Manna Ice	47.461834	19.014718	Ice Cream Shop
11. District	47.4593	19.0187	Bukkony ovoda	47.445056	19.034785	Daycare
11. District	47.4593	19.0187	Szent Gellert ovoda	47.476179	19.027711	Daycare

Figure 6 - Example for Daycare data combined with Foursquare API data

Data analysis

During the data analysis the frequency of the daycares and the distribution have been looked at by district and also represented visually.

For the venues the k-means clustering modeling was used to cluster the districts in Budapest.

Important to mention that the number of daycare datapoints were 506 while the Foursquare venues were 513, meaning that the daycare data would have distort the analysis so when evaluating the venue types daycares were not considered.

Results

The statistical analysis revealed that districts 1, 5, 6 and 23 have less than 10 daycares.

On the other hand districts 2, 3 and 14 have more than 30 daycares.

District	Name	Ad
1	3	
2	32	
3	38	
4	27	
5	8	
6	8	
7	11	
8	15	
9	8	
10	23	
11	28	
12	26	
13	21	
14	41	
15	13	
16	24	
17	20	
18	27	
19	15	
20	12	
21	23	
22	19	
23	4	

Figure 7 - Number of daycares by district

The k-mean clustering modeling created 5 clusters in Budapest.

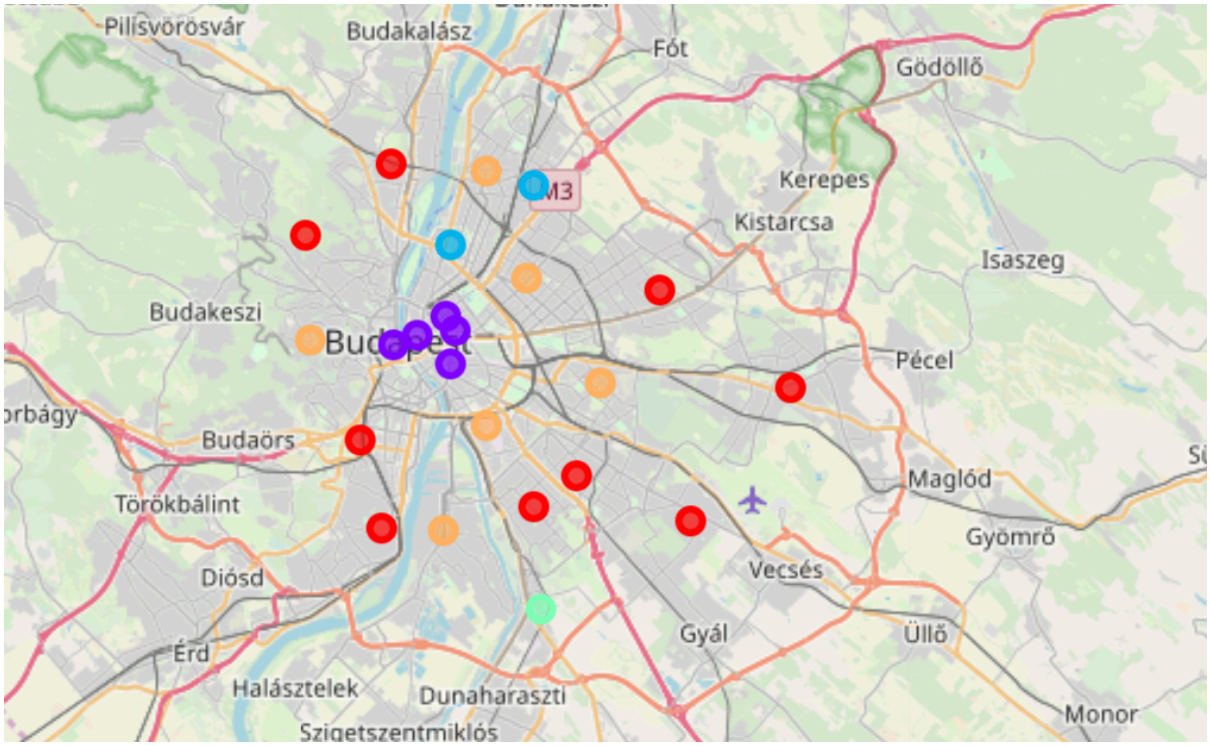

Figure 8 - K-means clustering output on map

Cluster 1

9 districts are in this cluster, the main venues in this cluster are parks, dog run, playgrounds.

Cluster 2

5 districts are in this cluster, the main venues are hotels, plazas, coffee shops.

Cluster 3

2 districts in this cluster, the main venues are pizza places, restaurants, grocery stores.

Cluster 4

1 district in this cluster with bakery and grocery stores.

Cluster 5

6 districts in this cluster, interestingly main venues are related to public transportation: bus stops, train stations.

Discussion

As presented in the results section districts 1, 5, 6 and 23 have less than 10 daycares which indicates that such districts might not be good choice for couples with young children as it might be difficult to find a place or would require lots of travel to get to the daycare.

On the other hand districts 2, 3 and 14 have more than 30 daycares therefore one of these district might be a good choice for couples with young children, likely it is easy to get a place in one of the daycares and also there is a daycare in close proximity therefore travel between home and the daycare does not take much time.

The clustering revealed that 9 districts (cluster 1) might be ideal for those with children as several playgrounds, parks are available. 2 districts have plenty of options for restaurants so can be a good choice for those who frequently visit such places.

Conclusion

Although this analysis revealed several interesting observations like which districts are good choice for couples with young children there are many other factors to be considered that might influence the choice. For example the public transportation coverage, housing prices, access to medical venues (ie. hospitals) and so on. Using this analysis one can at least narrowing down where to start looking for buying a house or flat.

References

Websites used for data collection:

1) Daycare data: https://holmivan.valami.info
2) Budapest districts: https://en.wikipedia.org/wiki/List_of_districts_in_Budapest

List of figures