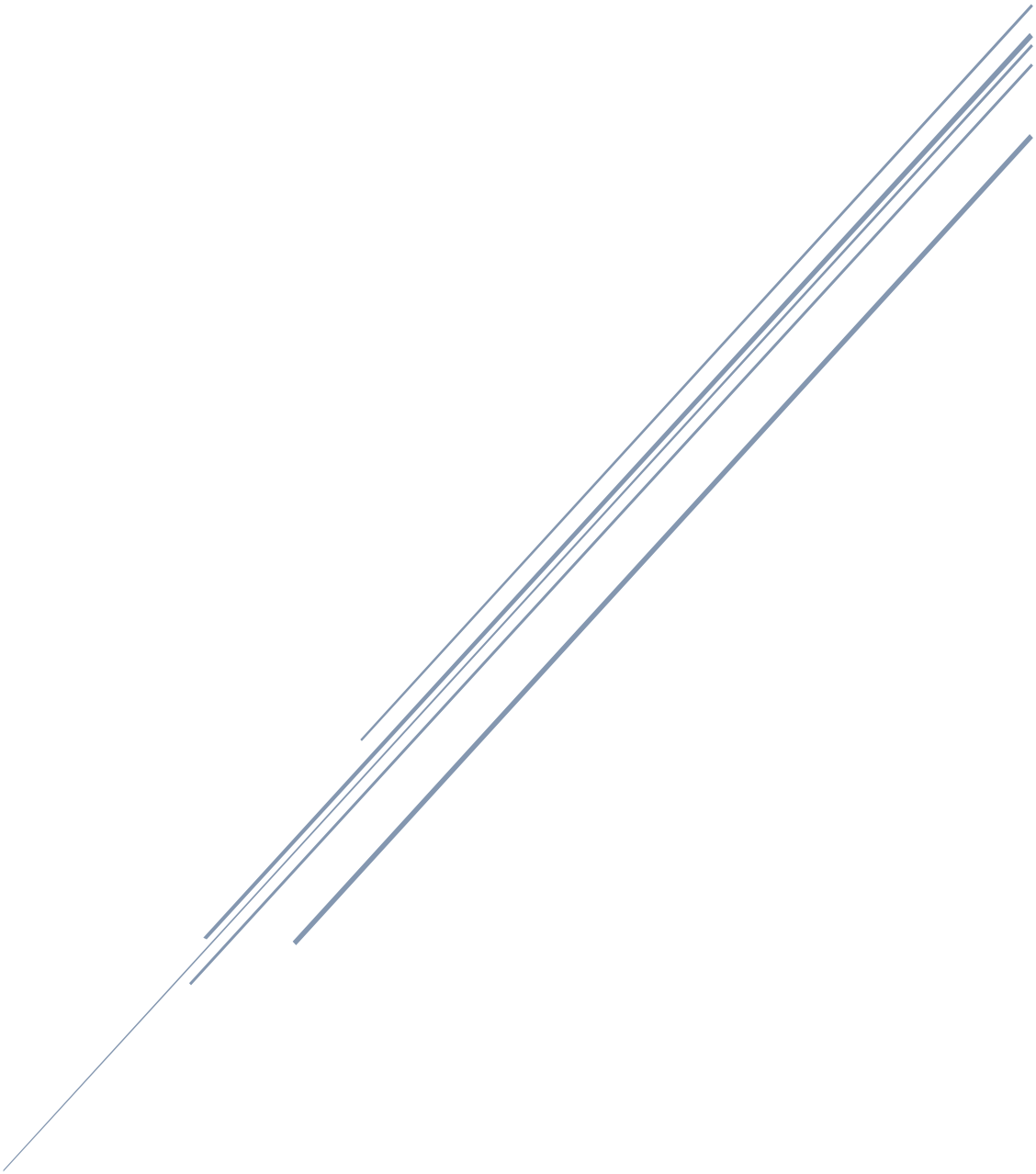
BATTLE OF THE NEIGHBORHOODS

A neighborhood classifier system



by Gergely Karácsonyi

Capstone of IBM Data Science Professional course

Table of Contents

Introduction....2
Background....2
Data Description....2
Municipal area data (districts and neighborhoods)....2
Nominatim....2
Foursquare....3
Personal preferences data....3
Data usage....3
Methodology....3
Results....3
Discussion....3
Conclusion....3
References used....3

Introduction

Background

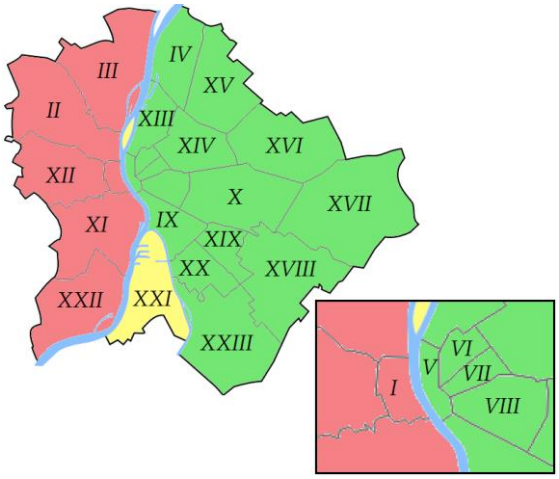
Moving to a big city is not simple. Without any local knowledge, it can be very challenging to pick the right area to buy or rent our new home. In this study I will set up a recommender system which uses different personae and various location based data as input to help us select a proper place to live.

Data Description

For the study I selected Budapest, the capital of Hungary. It's the most populous city in the country with a population of approx. 1.7M. I chose Budapest because it's a popular place to live, with a thousands of venues and sights. I utilized the following data for the study:

Municipal area data (districts and neighborhoods)

Budapest has 23 districts, pictured on the right. Each district can be associated with one or more neighbourhoods so we will dig into neighborhood level later if possible, because district-level resolution is quite coarse. I've extracted the data (both the district and the neighborhood names) from Wikipedia.[1] and after that I've done a little cleanup and left only the names in the dataframe.



	District
0	I.
1	II.
2	III.
3	IV.
4	V.

Nominatim

I used Openstreetmap's Nominatim engine [2] to assign geocoordinates to each district and neighborhood.

	District	Latitude	Longitude
0	I.	47.499163	19.035143
1	II.	47.538887	18.982636
2	III.	47.568691	19.027668
3	IV.	47.577779	19.093164
4	V.	47.499945	19.050549

Foursquare

I used data from Foursquare API to get the most common venue categories for each neighborhood. [3]

	District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	I.	47.499163	19.035143	Dísz tér	47.499100	19.036163	Plaza
1	I.	47.499163	19.035143	Halászbástya | Fisherman's Bastion (Halászbástya)	47.502029	19.035058	Historic Site
2	I.	47.499163	19.035143	Stand25 Bisztró	47.497673	19.032679	Bistro
3	I.	47.499163	19.035143	Várnegyed	47.501195	19.032261	Scenic Lookout
4	I.	47.499163	19.035143	Honda Dream	47.498561	19.031825	Motorcycle Shop

Personal preferences data

This custom dataset contains personal venue preferences on a scale of 1 to 10 (least preferred to most preferred).

Example:

```
userInput = [   {'venue':'Beach', 'rating':10},
                {'venue':'Garden', 'rating':8}  ]
```

Data usage

We'll create clusters of the districts based on their most common venue categories. The user can get information about these clusters and by using the recommender system she can get recommendations matching her persona. This persona will be created by classification based on her venue preferences (e.g.: sports fan, family person, coffee lover, party animal, etc.). After she got a district recommendation, she can dig deeper if possible, to neighborhood granularity (It's not always possible, given there are districts which have only one neighborhood - practically the whole district is the neighborhood itself). Then we could repeat the process to get an even more specific recommendation. In the end, we'll visualize the results on a map.

Methodology

Results

Discussion

Conclusion

References used

[1] https://en.wikipedia.org/wiki/List_of_districts_in_Budapest

[2] https://nominatim.openstreetmap.org/

[3] https://developer.foursquare.com/docs/places-api/