# A Central Europe investment opportunity analysis between Warsaw and Budapest

## Federico Noverano

### September 7, 2020

## Business problem description

A very successful Italian food chain, have recently opened 5 restaurants in Warsaw, which share similar atmosphere, menu, furniture-style, processes, tools and raw material providers. The restaurants have been very successful due to online marketing messages, link to blogs/internet influencers, fashion newspaper as well as due to efficient cost structure and with an high pricing point. The company which owns the company had come to know that there is the possibility to get both EU and local funding for investing in Hungary and is considering the opportunity to launch five restaurants in Budapest, so it turned to our company to make a similarity analysis between Warsaw and Budapest in terms of existing venues and distribution, to validate the idea. The target audience are the decision maker for the investment: the CEO, COO and CMO, however an extract of the report shall be made available to the shareholders to document the outcome of the decision

### The locations

The analysis will be performed comparing Warsaw and Budapest districts, the existing restaurants, pubs, and other attractions

## Methodology and tools

### Foursquare API

This project would use Four-square API as its prime data gathering source as it has a database of millions of places, especially their places API which provides the ability to perform location search, location sharing and details about a business.

### Work Flow

Using credentials of Foursquare API features of near-by places of the neighborhoods would be mined. Due to http request limitations the number of places per neighborhood parameter would reasonably be set to 100 and the radius parameter would be set to 500.

### Clustering Approach

To compare the similarities of the two cities, we decided to explore neighborhoods, segment them, and group them into clusters to find similar neighborhoods between them. To be able to do that, we need to cluster data which is a form of unsupervised machine learning: k-means clustering algorithm

### Libraries Which are Used to Develop the Project

Pandas: For creating and manipulating dataframes.

Folium: Python visualization library would be used to visualize the neighborhoods cluster distribution of using interactive leaflet map.

Scikit Learn: For importing k-means clustering.

JSON: Library to handle JSON files.

XML: To separate data from presentation and XML stores data in plain text format.

Geocoder: To retrieve Location Data.

Beautiful Soup and Requests: To scrap and library to handle http requests.

Matplotlib: Python Plotting Module.

## Data Description

The main data sources will be

1) the list of districts of Warsaw and Budapest with their population which are available at

<https://pl.wikipedia.org/wiki/Podzia%C5%82_administracyjny_Warszawy>

<https://en.wikipedia.org/wiki/List_of_districts_in_Budapest>

2) The longitude and latitude of each district which will be retrieved from the Geocoder API

3) The foursquare API data from venues

### Foursquare API

We will need data about different venues in different neighborhoods of that specific borough. In order to gain that information we will use "Foursquare" locational information.

Foursquare is a location data provider with information about all manner of venues and events within an area of interest. Such information includes venue names, locations, menus and even photos. As such, the foursquare location platform will be used as the sole data source since all the stated required information can be obtained through the API.

After finding the list of neighborhoods, we then connect to the Foursquare API to gather information about venues inside each and every neighborhood. For each neighborhood, we have chosen the radius to be calculated as the radius of the area of the district (approximating each district shape as circle).

The data retrieved from Foursquare contained information of venues within a specified distance of the longitude and latitude of the postcodes. The information obtained per venue as follows:

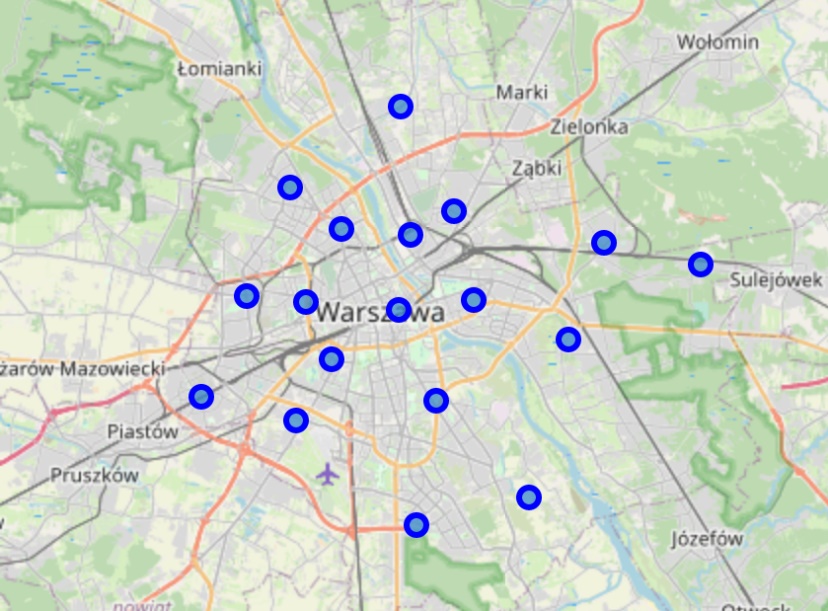
* Neighborhood
* Neighborhood Latitude
* Neighborhood Longitude
* Venue
* Name of the venue e.g. the name of a store or restaurant
* Venue Latitude
* Venue Longitude
* Venue Category

## Warsaw Districts Data

In Warsaw there are 18 districts, according to below table showing their population and area:



The geographic distribution of Warsaw district is quite homogeneous as presented in below picture, the majority of the districts are on the west side of the Vistula river.



## Budapest Districts Data

In Budapest there are 23 districts, according to below table showing their population and area:

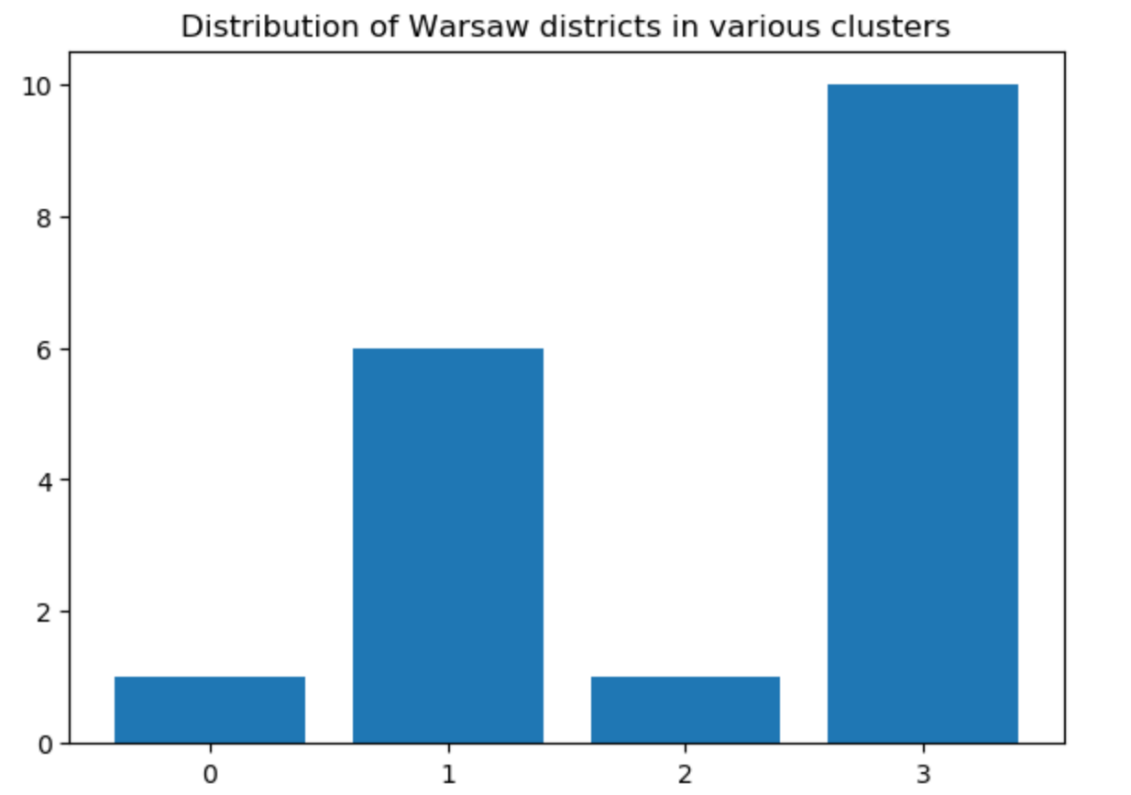


The geographic distribution of Budapest district is presented below, more districts concentrated in the city center whilst other are distant. The majority of the districts are on the west side of the Danube river.

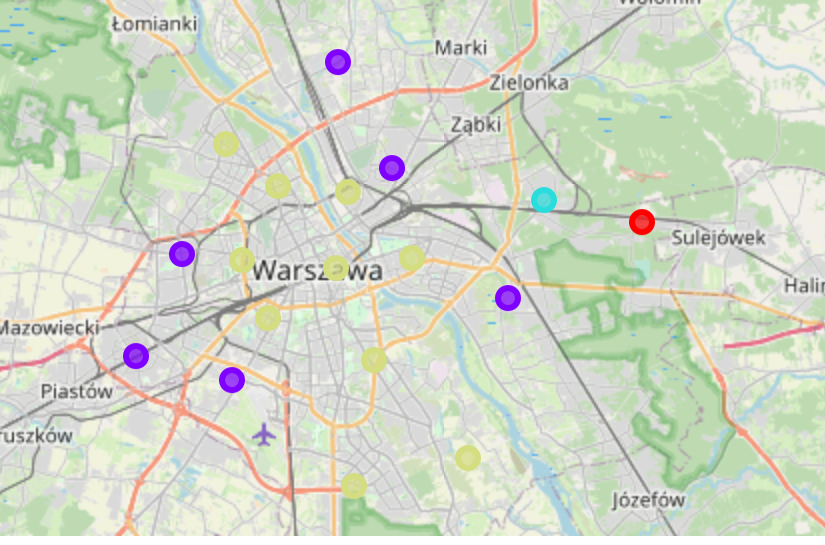


## Warsaw districts clustering, by district venues

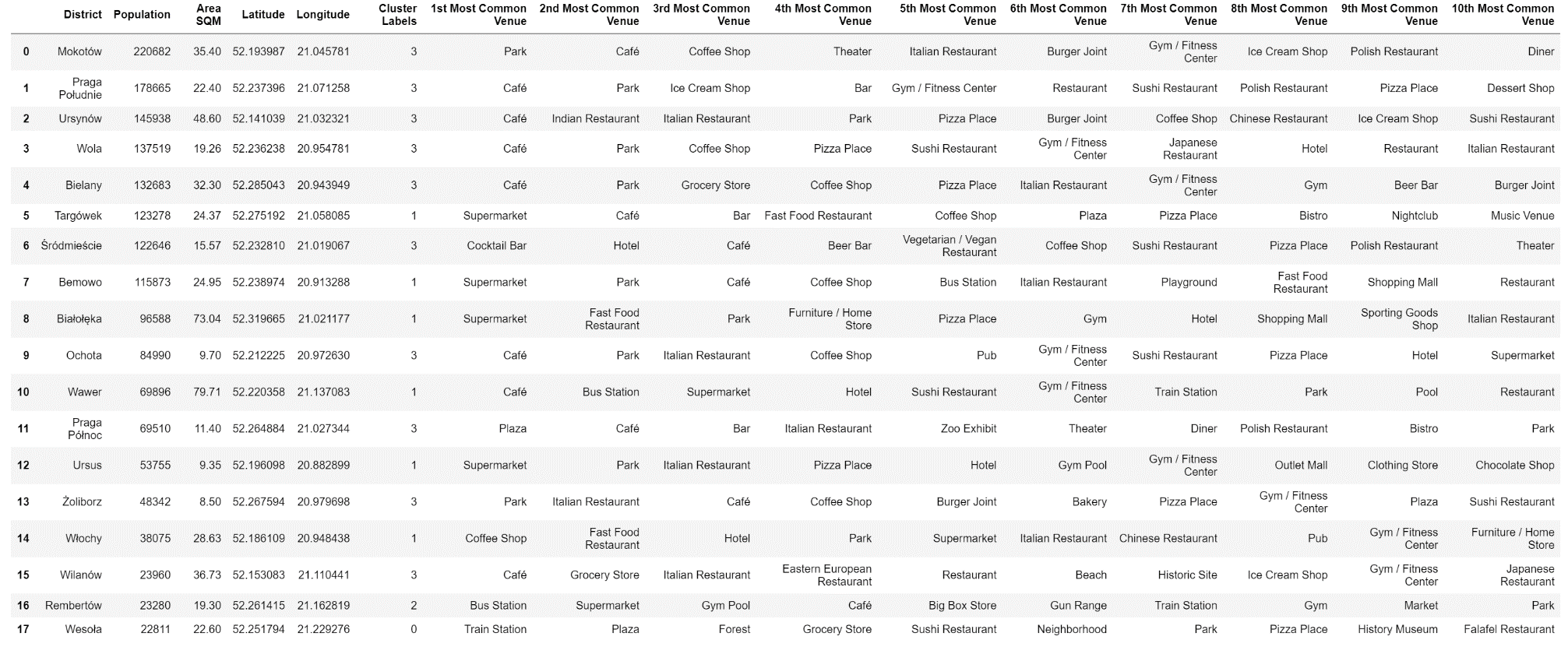
Upon downloading the list of venues (up to 100 per district) of Warsaw districts we have proceed with clustering them in 4 group (the 3 and 5 groups distribution were tested and rejected) we can observe the Warsaw districts are split in two major groups and two residual, less relevant, groups.



As we can see from the below map, the most wealthy districts are the majority, and they are concentrated in the center (yellow dots, corresponding to cluster 3 above), than there is a second tier of very similar good-level neighborhoods (purple dots, cluster 1 above), than just two districts are not recommended for living and investment.

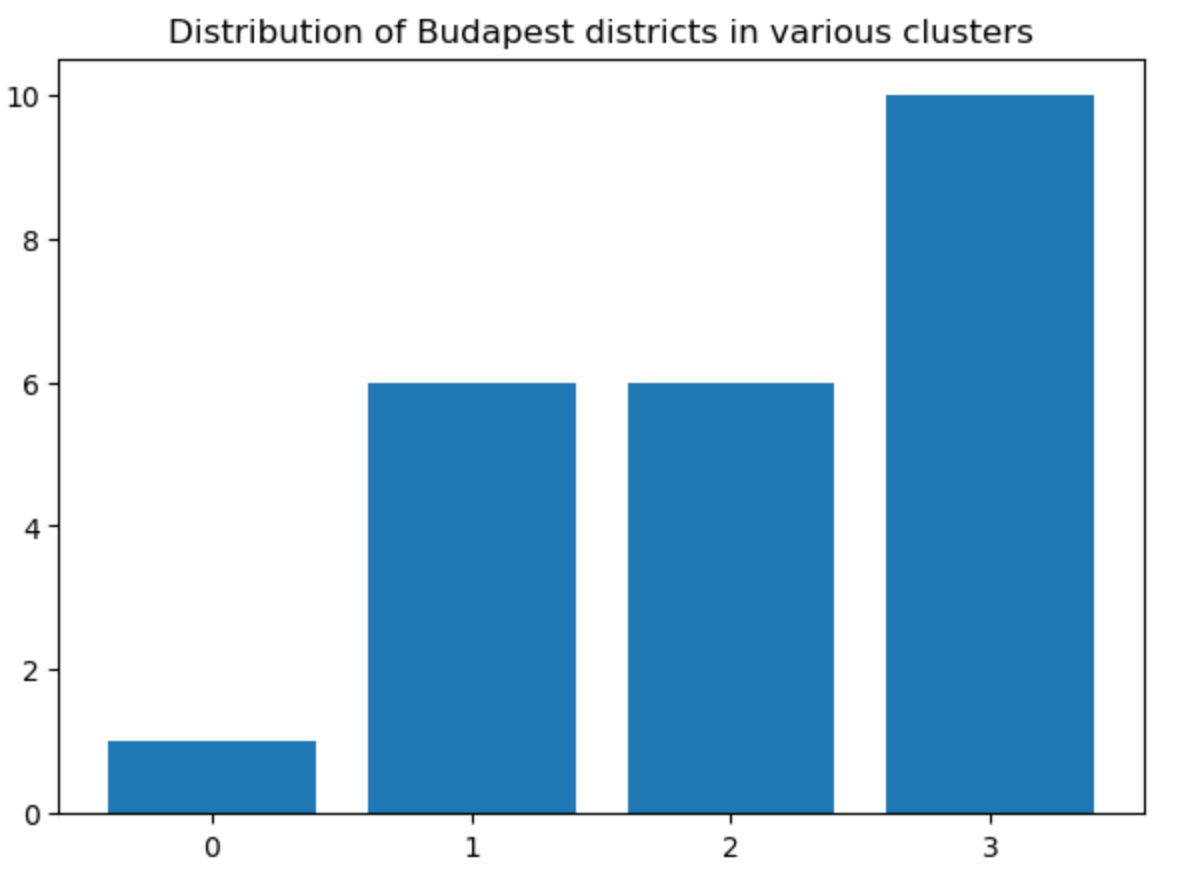


If we look at the venue distribution per cluster we can see a fair amount of parks, gyms and restaurants of various kind in both cluster 3, as well as a certain amount of supermarkets in cluster 1, which is a understandable being those areas outside the city center.

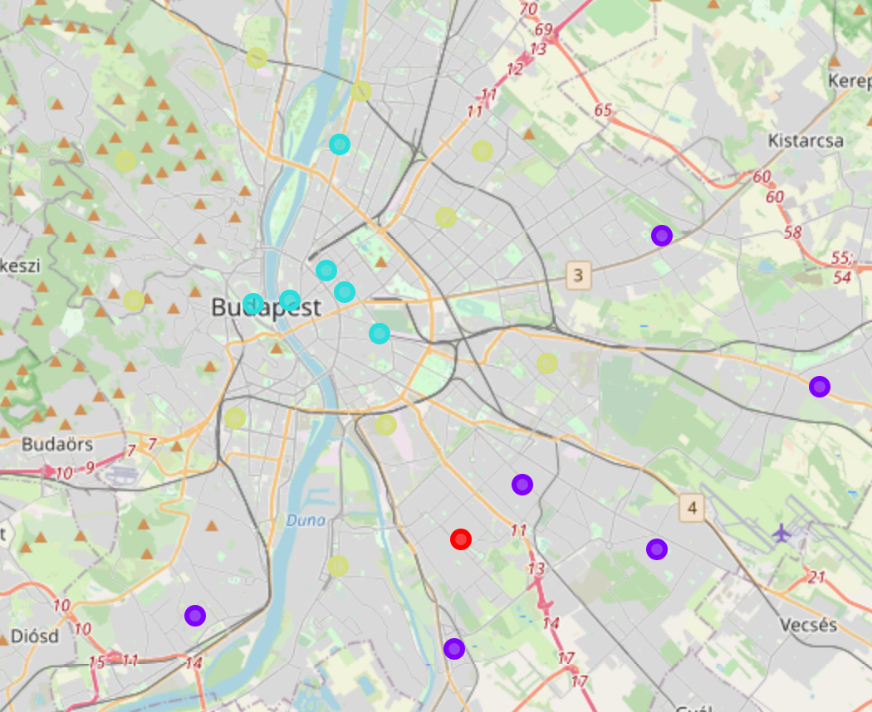


## Budapest districts clustering, by district venues

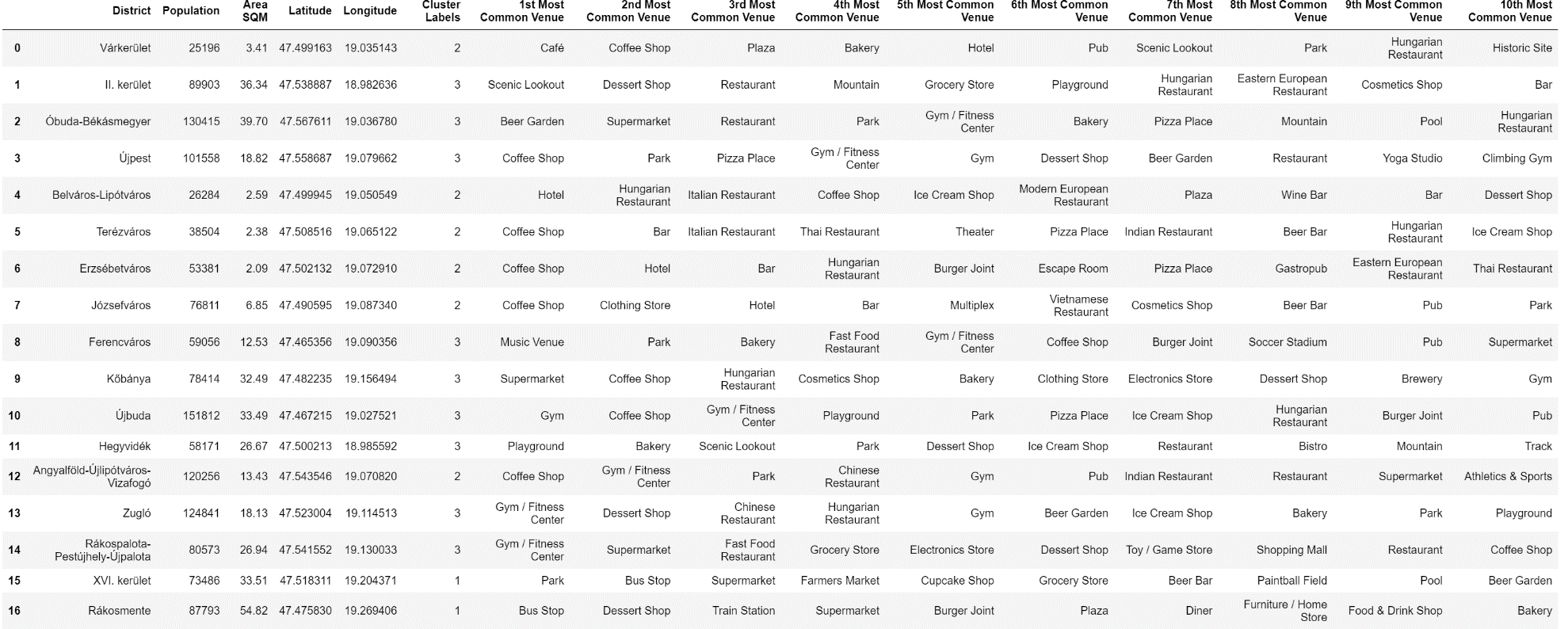
Upon downloading the list of venues (up to 100 per district) of Budapest districts we have proceed with clustering them in 4 group (as for Warsaw case) we can observe that the districts are distributed along three main clusters, and only one is belonging to a residual, less relevant group.



Differently from Warsaw, the most wealthy districts are starting from the center (blue dots, cluster 2 above) and then decreasing their attractiveness with distance (yellow dots, cluster 3 above, purple dots, cluster 1 above).



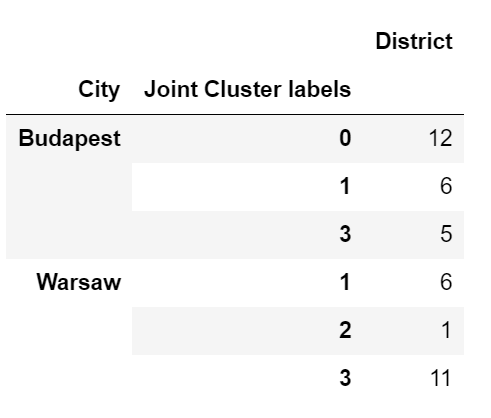
If we look at the venue distribution per cluster we can see the most attracting venues, are concentrated in clusters 2 and 3, whilst in cluster 1 there is a far share of public transport and some lower-end venues like groceries, so those are typically areas for lower income people.





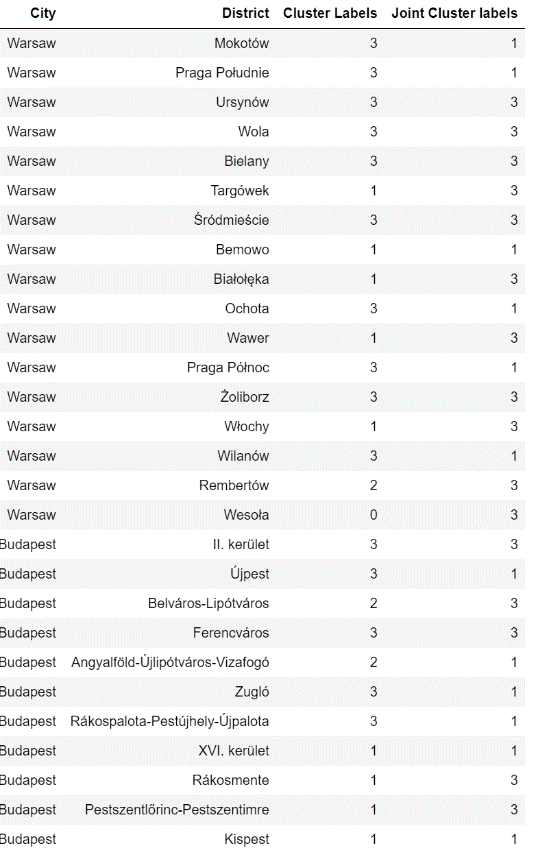
## Results of the comparison analysis

As a final step from our analysis we have grouped all districts and venues form Warsaw and Budapest and create a maxi cluster to check how well characteristics of districts from two city matches and finally recommend our client if to proceed with the investment or not.

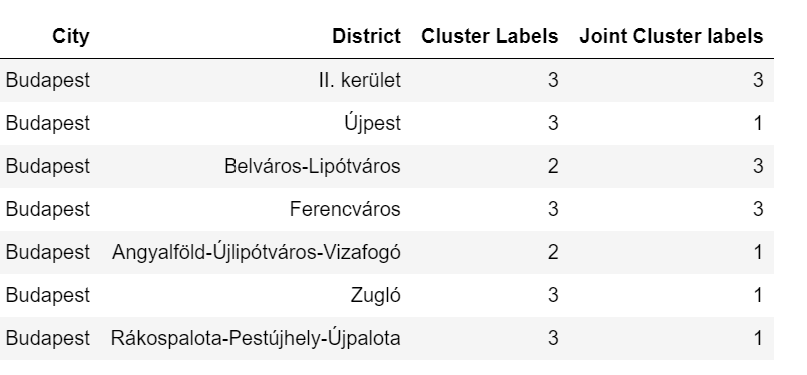


We can observe when joining the two cities there are only three clusters applicable to each of them, cluster 0 is applicable only to Budapest, cluster 2 only to Warsaw.

If we deep dive in cluster 1 and 3 we can also observe how the previous clustering was reshuffled in the new joint clustering.



So if we want to make sure we catch those areas from Budapest which nicely match Warsaw characteristics we should focus on the ones that were previously classified as 2 and3 in Budapest, i.e. the following ones



## Discussion of the results

Based on the initial data collected, the geographical and venue analysis of Warsaw and Budapest confirm some nice similarity among the two cities:

* a large set of venues of different kind
* a concentration of wealth in the central area and similar type of attractions
* a pretty clear distinction between the city center/wealthier areas and areas which are more a kind of dormitory for low income people

Given the identified similarity we are well encouraged to suggest our client to proceed with the investment in Budapest, furthermore the joint clustering exercise has identified which are the more promising area of Budapest to consider the opening of new restaurants given both their localization in the highest end of Budapest living as well as their similarity with Warsaw districts from cluster where the previous investment was successful.

## Conclusion

In conclusion of this report, based on the performed analysis we are going to recommend to proceed with the investment in opening a new set of restaurant in Budapest, however to fine tune the exact localizations starting from the shortlist of proposed districts we would suggest to perform additional statistical analysis, leveraging appropriate machine learning algorithms of the following kind of data:

* presence of people in the various areas at different time of the day
* more detailed analysis of target group being Hungarian residents, as well as foreigners (tourism in Budapest is definitely much more relevant than in Warsaw)
* analysis of economic growth in Hungary perspective in the longer term and comparison with similar data from Poland

Is also worth suggesting some analysis of the political situation and monetary policy in Hungary in order to mitigate risks deriving from import/export regulation, currency fluctuation, legislation benefits or barriers from foreign investments.