# **Green Coffee - Comparing Neighborhoods of Budapest**

## Introduction

**Background**

In Budapest a well-known coffeshop owner would like to open a Green Café, which sells only fair-trade coffes from certified sources.

He is targeting well-educated, wealthy citizens as he is offering fair-trade arabica, which costs a bit more than the regular coffee. He would like to see which district of the city would be the most suitable for his future business. He firmly believes that the presence of other expensive services in the district (eg. spa, yoga studios, gastropubs, etc) could be a good indicator to solvent source of business.

**Problem**

The owner do not know the city and its various districts, in this point he would like to gather some insights. He thinks that the most frequent places are the good ones in the inner city, because there are a lot of tourists.

My hypothesis is that it could be a bad choice as they won’t be loyal customers on the long run.

**Interest**

My aim is to find the best district for the business in Budapest, based ont he already operating venues in the neighbourhood, using very similar method what we used in the example in the course but designed to Budapest. The capital of Hungary has 24 districts, some of them sharing similar characteristics. The difficulty is to find the exact centerpoints of the districts, as it is not accesssible on a wikipedia page. I will do this manually, as it is only 24 rows, and there is a running site where could we find them.

## Data Acquisition

**Data Sources**

As there is no open source data regarding the centrum of the districts, we will use the data of a running site, which could provide the centrums of different districts: https://www.futas.net/hungary/Budapest/gps-koordinatak.php

Then I will use the Foursquare API, in order to gather the venues nearby the given centrums.

I will use Python (numpy, folium, kmeans) packages in order to calculate the means and other necessary calculations.

**Data Cleaning**

At the end of the process I examined that there are some districts, which containing only a limited number of venues, therefore I was excluding those from our analysis, using the following rule: below 10 results, we won’t use the data on the district.

**Feature selection**

I will make a top 10 category list in each district, based on the most frequent categories in the examined radiuses. For this we will need the results and count them, and making a descending order

## Methodology

Firstly, I will use the data of a running site, which could provide the centrums of different districts. I will make a dataframe out of this data manually, so simply make an excel with 72 cells.

Then make requests from the Foursquare API, in 2000m radius of the centrums (mind that those districts are bigger than the ones used in Toronto) in order to find the venues nearby.

After that I will check the most frequent categories (top10), in order to see the most typical places in each district.

Then I make clusters (k-means), based on the dataframe. I will examine the districts which are similar and exclude those ones which already containing competitors, meaning one of the top3 venues are wheather Restaurant, Caffee or Coffe Shop.

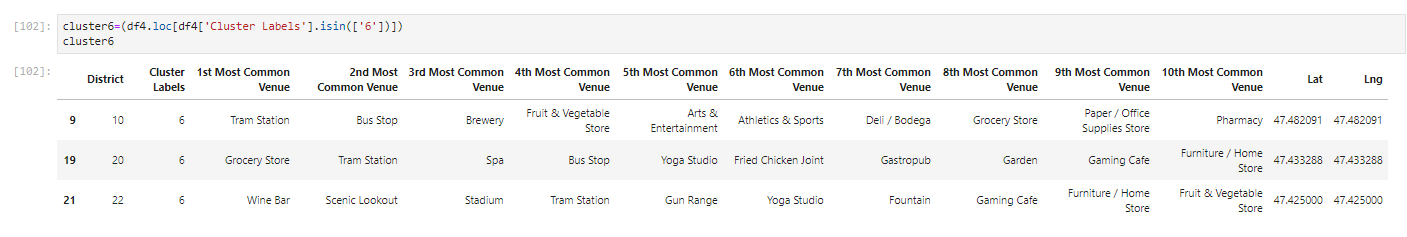
Then I will check the results and try to find a cluster, then a picking the most suitable.

## Results

Based on the accessible data the most suitable district would be one of the 6th cluster. I would suggest to open it in the 10th district, as both of the other two options has only a few result on Foursquare (I tried to widen the circle limit, it came back with the same results), thus the 10th is more reliable.

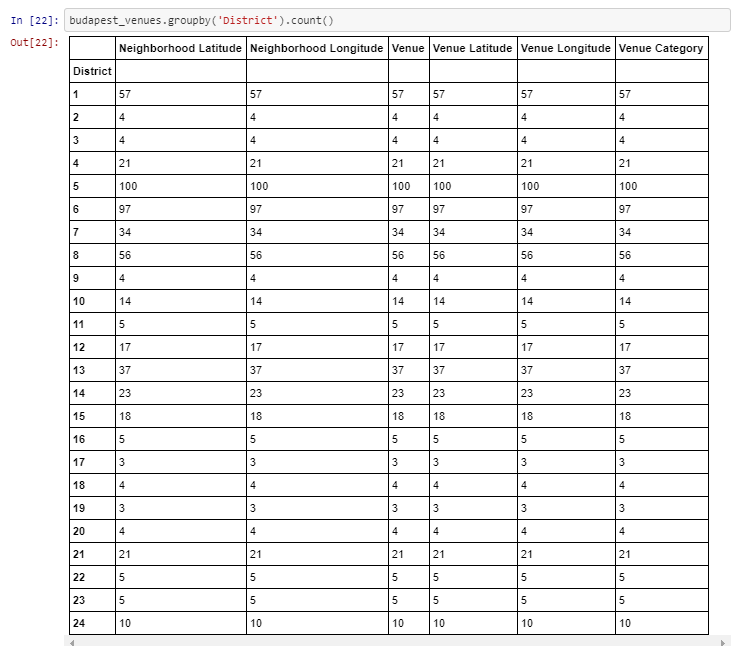
The 10th district is quite central, still it is not flooded by the tourists, as the inner city part is considered as the closest districts to the Danube, thus those districts are out of league.

As the first two common venues are public transport stops, it is easily accessible. If we see the data closer, we can see that two of the most beloved craft beer breweries are also located here, therefore resourceful consumers are presumably located here.



## Future Directions

It is important to highlight that the process needs further examinations, as the Foursquare API could provide only a few venues in Budapest in some cases, despite that I tried to use higher limits on API requests.



Still, it is a good insight for the beginning of this validation process. I would suggest to compare the housing prices with the examined data and try to gather more information on the district itself.

For further info, please visit this site.

<https://en.wikipedia.org/wiki/K%C5%91b%C3%A1nya>