IBM Data Science Capstone Project
Problem Definition and Data

Problem Definition

For this project, I chose a theoretical business problem. The question that we are trying to answer is the following.

A successful owner of multiple mid to high-end restaurants decided to open a new restaurant in Budapest, Hungary. Having visited the city many times in recent years, he couldn't disregard the big boom in gastronomy. He is keen on opening a new unit, which will focus on the European and Asian fusion kitchen.

Taking into account the price level at which the restaurant will operate, the intent is to find an optimal location in an area, where gastronomy is booming and which is easily accessible for tourists and for wealthier local citizens as well.

Assumptions, business logic

The assumption behind the analysis is that we can use unsupervised machine learning to create clusters of districts that will provide us with a list of areas for consideration for the restaurant. The intent is that the restaurant to be situated close to one of the gastronomical centres.

Audience

While here we are assuming a concrete business owner to whom we are addressing this report, but actually this restaurant owner can be treated as a persona and thus this analysis could be useful for a group of market players (restaurant owners).

Data

To perform this analysis, we will need the following data:

1. List of the districts of Budapest
2. Geo-coordinates of the districts in Budapest
3. Top venues of districts

List of districts will be obtained from wikipedia (https://en.wikipedia.org/wiki/List_of_districts_in_Budapest)

Geo-coordinates of districts will be obtained with the help of the geocoder tool in the notebook.

Top venues data will be obtained from Foursquare through an API.

Use of Data and Methodology

After tidying up and exploring the data, we will apply the K-means machine learning technique for creating clusters of districts. We will use the silhouette score for choosing the optimal number of clusters.