Analysis of Budapest Neighborhoods

DATE : 8th Aug 2020

By
Sethumadhavan Aravindakshan

INTRODUCTION

Budapest is ranked as the top Eastern European ecosystem for scaleups due to the city's availability of capital. The tech infrastructure is exceptional and continues to develop further creating many opportunities for entrepreneurs and working professionals.

Aim of this project is to cluster the Budapest neighborhood according to their similarities and different categories into groups as :

- Fully developed neighborhood
- Moderately developed neighborhood
- Developing neighborhood
- Under developed neighborhood

This will aid anyone who visits Budapest city to decide the place of stay based on what each neighborhood has to offer and their personal preferences.

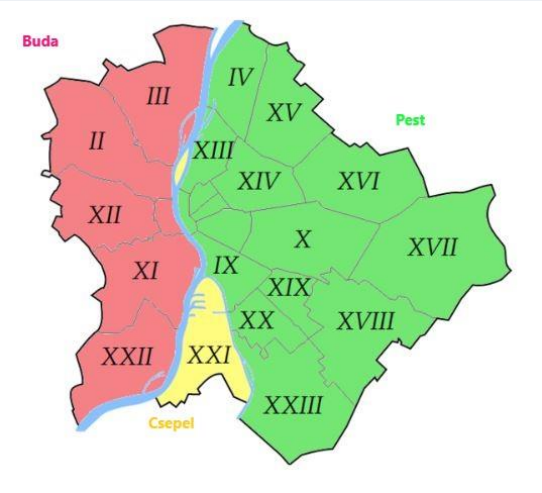
Data

In order to perform a meaningful analysis, we need data from reliable sources. To understand our problem and quantify results, the following data is used:

1. List of all districts of the city and postal codes are extracted from the below url: https://en.wikipedia.org/wiki/List_of_postal_codes_in_Hungary

	Postal Code	District names	Neighborhood
0	1010	Várkerület	I.
1	1020	2nd district	II.
2	1030	Óbuda-Békásmegyer	III.
3	1040	Újpest	IV.
4	1050	Belváros-Lipótváros	V.

2. Then the Borough names are added as per the below map:



	Postal Code	District names	Neighborhood	Borough
0	1010	Várkerület	I.	Buda
1	1020	2nd district	II.	Buda
2	1030	Óbuda-Békásmegyer	III.	Buda
3	1040	Újpest	IV.	Pest
4	1050	Belváros-Lipótváros	V.	Pest

3. Geographical co-ordinates of each neighborhood is extracted using the Gecoder from Geopy python library.

	Postal Code	District names	Neighborhood	Borough	Latitude	Longitude
0	1010	Várkerület	I.	Buda	47.499163	19.035143
1	1020	2nd district	II.	Buda	47.538887	18.982636
2	1030	Óbuda-Békásmegyer	III.	Buda	47.567611	19.036780
3	1040	Újpest	IV.	Pest	47.558687	19.079662
4	1050	Belváros-Lipótváros	V.	Pest	47.499945	19.050549

4. Foursquare developer access to venue data : https://foursquare.com/developers//
   Foursquare API is used to explore and extract the list of all venues and their frequencies of the neighborhood of the Budapest city. This data is used to further cluster the neighborhood based on their similarities.

Later a list of all venues for the top 10 categories are extracted using the Foursquare API and then frequency for all categories for each neighborhood is determined. Finally, Unsupervised Machine Learning Clustering techniques are applied to determine a similar pattern between the neighborhoods to group them to similar clusters for further analysis.