IBM Applied Data Science Capstone Project
Districts in Budapest by restaurant categories

Introduction

Budapest is a large Central-European city which accounts for more than 35% of the total GDP of Hungary, therefore, it has a central role in the Hungarian economic outlook (OECD, 2019). Hence, Budapest is not only the capital of Hungary, but the most important economic region of the country from the perspective of the services sector and tourism. In 2018, Budapest registered 12.5 million guests, which is a year-to-year increase of 5%. Moreover, it is expected that tourism will continue to flourish in the capital (Berende, 2019). On top of the growth in the number of visitors, Budapest also received the European Best Destination 2019 award, ahead of such cities as Athens and Florence ("Best places to travel in 2019", 2019). Tourism also has a great impact on the Hungarian economy as it contributes with 6.3% to GDP, which is composed by commercial accommodation & catering (amounts to 2% contribution to GDP) and investments in tourism ("Broader tourism sector generates more than 6% of GDP", 2019). The commercial catering sector is greatly sustained by the thriving restaurant market in Budapest, where one can find a variety of different cuisines. Nevertheless, there are defining restaurant categories one might find in the capital.

Business Problem

The purpose of this project is to segment districts in Budapest based on restaurant categories that can be found in the given district. A combination of descriptive statistics and machine learning techniques, such as clustering, will be used to support the aforementioned goal. The analysis will support the decision-making process of tourists and anyone that plan to dine in Budapest to look for a restaurant in the district that most suits their needs.

Data

To support the analysis, data will be collected from various sources. Firstly, a list of the districts of Budapest will be needed. Next, geographical data is needed for the respective districts. The geographical data will support the data visualization part of the project, especially the mapping

of neighborhoods and venues. Lastly, restaurant categories data will be needed to perform the clustering on the neighborhoods.

To supply the data about the neighborhoods, Wikipedia will be used, which provides a list of districts with the corresponding neighborhoods in Budapest. Moreover, Wikipedia also provides a list of postal codes of the districts, which will be used in combination with geographical data from geodatos. Geodatos maintains the latitude and longitude coordinates of each district. Moreover, Foursquare API will be used to supply venues data from each neighborhood in Budapest.

The data from Wikipedia will be extracted using web scraping techniques with the pandas package. The geographical data from Geodatos will be extracted into a .csv file and joined to the main neighborhoods dataframe. Lastly, GET calls will be utilized from the Foursquare API based on the geographical data to have the venues data needed to perform the analysis.

Bibliography

Berende, P. (2019). Hungary attracts record number of visitors. *welovebudapest.com*. Retrieved from https://welovebudapest.com/en/2019/02/13/hungary-attracts-record-numbers-of-visitors/

Best places to travel in 2019. (2019). Retrieved from https://www.europeanbestdestinations.com/european-best-destinations-2019/

Broader tourism sector generates more than 6% of GDP. (2019). Retrieved from https://bbj.hu/economy/broader-tourism-sector-generates-more-than-6-of-gdp_159634

OECD (2019). Regional Economy. *OECD*. Retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM