

IBM Applied Data Science Capstone Project

Neighborhoods to invest in for office properties in Budapest, Hungary

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. Due to its central role in the services sector, Budapest attracts the most office real estate investments in the country (Hegedűs, 2019). Gaál (2019) reports a record low office vacancy rate of 6.3% in Budapest, which signifies the need for these buildings in the capital. The substantial need for office buildings creates an opportunity for real estate developers due to the consistent rental income gained which is also supported by the average gross prime office yield of 6.2% (Hegedűs, 2019). Nevertheless, the choice of real estate development depends on a multitude of factors, one of them being the exact location.

Business Problem

The purpose of this project is to identify suitable districts / neighborhoods in Budapest for office building development. 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 real estate developers and investors that are looking to open or invest new office buildings in the capital of Hungary. As was described in the *Introduction* section, this project can bring significant value to the target audience due to the immense need for office buildings in Budapest.

Data

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

part of the project, especially the mapping of neighborhoods and venues. Lastly, office building venue 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 scarping 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

- Hegedűs O. (2019). Hungary Office Market Snapshot Third Quarter 2019. *Cushman & Wakefield*. Retrieved from <http://www.cushmanwakefield.com/en/research-and-insight/hungary/hungary-office-snapshot>
- Gaál B. (2019). Budapest office market vacancies hit record low. *Budapest Business Journal*. Retrieved from https://bbj.hu/real-estate/budapest-office-market-vacancies-hit-record-low_168953
- OECD (2019). Regional Economy. *OECD*. Retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=REGION_ECONOM