



COURSERA CAPSTONE PROJECT

OPENING A WELLNESS & SPA CHAIN IN KANTON OF
ZUG, SWITZERLAND

SUMMARY

Evaluation of the 5 most expensive regions of kanton of Zug, Switzerland. Utilisation of various data science techniques in Python. FourSquare leveraging and Folium maps generation.

Mikołaj Wlazło

IBM APPLIED DATA SCIENCE PROFESSIONAL COURSE

JUNE, 2020

1. Introduction

Wellness & spa is an active process of making choices toward a healthy and fulfilling life. Wellness/spa tourism is generally understood to be traveling for the purpose of enhancing health and wellbeing through the use of spas, preventative treatments and therapies, and it becomes more and more popular.

For many people, visiting wellness & spa centre is a great way to relax and enjoy themselves during weekends and holidays. Locations like this are like a one-stop destination for both locals and tourists. Of course, as with any business decision, opening a chain of wellness & spa centres requires serious consideration, and is a lot more complicated than it seems. Particularly, the location of the centre is one of the most important decisions that will determine whether it will be a success or a failure.

2. Business Problem

A private investor is seeking a perfect location to open a chain of wellness & spa centres around Switzerland. The most expensive (in terms of gross annual rent per m²) kantons are being targeted, and within them the locations with the same feature, and the lowest possible competition. This is motivated by the fact that this type of luxury centres mainly attract busy, mid-/high-class people.

Thus, the main objective of the project is to select the best locations where such centres can be put up, aiming at the above demographic, thereby helping the owners to achieve maximum profits.

Using data science methodology and machine learning techniques, this project aims to provide solutions to answer the business question: In Switzerland, if an investor is looking to open a wellness & spa chain, where would you recommend to open it?

3. Data

To solve the problem, we will need the following data:

- List of kantons in Switzerland;
- List of gemeindens in kanton of Zug;
- Respective values for each region in the area of housing prices. This defines the scope of this project which is confined to the kanton of Zug, one of the main regions in the country of Switzerland, in Central Europe;
- Latitude and longitude coordinates of those gemeindens. This is required in order to plot the map and also to get the venue data;
- Venue data, particularly data related to services and leisure. We will use this data to perform clustering on the gemeindens and evaluate potential for the given problem.

The data for this project has been retrieved and processed through multiple sources, giving careful considerations to the accuracy of the methods used.

3.1 Regions & Gemeindens

The data for specific kanton in Switzerland and gemeinden in the kanton of Zug can be extracted from the webpage: <https://realadvisor.ch/en/property-prices>. The generated excel file was uploaded over to the notebook and read using pandas library for Python.

3.2 Geocoding

The file contents from *3 Zug Data.xlsx* are retrieved into a Pandas Data Frame. The latitude and longitude of the gemeindens were gathered from: <https://www.mapplus.ch>. The geometric location values are then stored into the initial dataframe.

3.3 Venue Data

From the location data obtained, after Data Frames creation and Geocoding, the venue data is found out by passing in the required parameters to the FourSquare API, and creating another Data Frame to contain all the venue details along with the respective gemeindens.

FourSquare has one of the largest databases of 105+ million places, and is used by over 125,000 developers. FourSquare API will provide categories of the venue data for clustering and decision making.

This is a project that will make use of many data science skills, from web scraping, working with API (FourSquare), data cleaning, data wrangling, data visualization (matplotlib), to machine learning (K-means clustering) and map visualization (Folium).

Next, the Methodology section is presented where discussion of the steps taken in this project takes place, along with the data analysis, and the used machine learning techniques.