Battle of the Neighborhoods

Rehabilitations Clinics in North-West Germany

A Data Science Project Using Foursquare Location Data

Author: Brita von Bartenwerffer
Date: xx.xx.2021

Battle of the Neighborhoods: Rehabilitation Clinics in North-West Germany

Table of Contents

Business Problem

General Introduction

In Germany every person has the right to apply for a health cure either as an outpatient while living at home or away from home living and having treatment at a rehabilitation clinic. The patients own doctor is always involved in the application process which is then submitted to the patient's health insurer or the Germany Federal Pension Fund.

There is no time pressure for these cures and usually the patient has been suffering symptoms for some time but has not been signed off work because of them. This is the

basis for this analysis and report. It does not deal with health cures that are part of the treatment of serious illnesses where it's often a matter of urgency, time and available places where a patient is send to.

The clinics used for this project offer treatment for many different complaints, most notably orthopaedic and cardiovascular problems many people suffer from through today's sedentary lifestyle. They also treat patients for stress-related problems as long as those don't call for specialized treatment at a neurological or psychiatric facility. This gives a wide choice of clinics to send a patient to. In many cases the turnover is on same Monday across many clinics so there is less of a question of available places.

The Problem

Patients may include some wishes as to the location of the clinic but unless they actually know about a specific one through word-of-mouth recommendations this is difficult for them to do. There are a couple of internet portals that provide some information about clinics but none of them take the amenities in the vicinity of the clinic into account. This usually leads to people saying that they don't want to be too far/near their place of residence but other than that there is little for them to go on.

Over several decades this has not been a problem. In general patients were expected to stay on or near the clinic grounds, even on weekends. People were much less mobile than they are today. Many were brought by car by relatives or came by train and taxi. During weekdays their schedules were often packed with treatments and specific rest periods so there was also little time to spend time elsewhere. Communications with their loved ones was by letter and telephone; visits from relatives or friends were frowned upon and not encouraged. While this was and still is a health cure it was widely accept that it is a medical measure.

Nowadays patients view a health cure more along the lines of a relaxed but organized holiday, in the same way people go on tours packed with cultural events. Patients usually driven themselves up and expect to be mobile during their stay. Due to staff bottlenecks their schedules are also less packed and they don't wish to spend all their free time in or around the clinic. Many like to have visitors and clinics no longer discourage visits as it has been accepted that social interaction is very important for the patient's wellbeing. However, many patients find that there is nowhere to stay for their visitors and only very little choice of restaurants for a relaxed midday meal. Most clinics don't cater to visitors other than offering coffee & cake in the afternoons.

Over the past decade this has let to a significant rise in complaints and long drawn-out correspondences between the insurance carriers, patients and doctors. This has let to rise of cost for the insurers and a serious trend of negative reviews of the clinics on the internet, taking things as far as patients refusing to go to specific ones regardless of their medical value.

The Goal

Health cures are brokered through different insurance carriers. All insurance companies whose area of operation lies in the north-western part of Germany have joined forced to

tackle this problem, decrease the number of complaints and increase patient's satisfaction. The consortium want

- to get an overview of what amenities are offered in the vicinity of the clinics in order to handle complaints more efficiently
- provide doctors and patients with databases and maps to help them formulate their wishes for the location of a clinic.

They are looking for a way to cluster clinics within similar neighborhoods into one cluster, helping them to decide on an appropriate clinic when the patient has stated a wish for a clinic within a certain cluster.

Rather than assessing the complaints which are always subjective and only represent the patients that actually take the time and effort to complain the companies want to rely on data science practices using objective location data to build the clusters. There are a couple of location services available over the internet, providing Application Programming Interfaces (APIs) to their databases. Anyone can contribute to their data and give ratings and review of the venues as well. This frees the insurers from having to compile and keep their own location data up to date.

Once this service has been built up it can also be used for similar issues, even ones outside of the scope of the insurers' business.

Data

The Clinics

In the scenario above the data would be compiled from the databases of the consortium. Since I don't have access to their data I used mainly one website to make a list of the clinics in north-western Germany: https://www.medfuehrer.de/Reha-Kliniksuche. There was no way to scrape the data from the website so I opted for copying the addresses of the clinics into a csv-file containing the following columns:

- Klinik: The name of the clinic
- Str_Hnr: Street address
- PLZ: Postcode
- Ort: village or town

Running into a few problems with addresses on the website obviously not conforming to the official postcodes of Germany I also needed to check a couple of the clinics' own websites and the postal service to get the official addresses.

Overall 60 clinics were compiled in this list. The list was then used to obtain the latitudes and longitudes of each clinic using the Geopy package in the Python environment. This is the prerequisite to make calls to the location data service.

Location Venues Data

To get a list of all venues in the vicinity of each clinic the Foursquare City Guide is used through their API. This has the advantage that the service is free to use for a limited

amount of calls per day. The consortium can gain experience without incurring even higher costs than those caused by the complaints. There's also the fact that there is no need to make calls to the service every day of the week or even several times during one day since the venues don't change that fast. It is quite sufficient to update the list once a month or even once every quarter. Once they have gained experience and want to roll this out to tackle other issues in the same way it's easy to upgrade to another membership tier in order to be able to place more calls to the API.

Foursquare is a location data service based in the United States in America, providing data for locations world-wide. In their own words their " mission is to build the most trusted, independent platform for understanding how people move through the real world." They have data for more than 100 million points of interest globally. They partner with other companies well-known all over the world, e.g. Twitter, Microsoft, Samsung, Apple. Location data used in their map services are likely to come from Foursquare.

Their API is very flexible, allowing the user to define the kind of venues as well as the radius around the point of interest in which they should be located. The API not only return the venues and their categories but also latitudes and longitudes, making it easy to generate maps that give a quick visual overview. There is also the ability to get venues' overall ratings plus a photo and a tip for each but these won't be used for this project because they are not objective and, at least in Germany, there may not be enough ratings to provide an average rating that is fairly reliable. In fact in preliminary analysis using just their website I found that many venues don't have ratings at all.

Methodology

Data Analysis and Cleaning

blablabla

Machine Learning

blablabla

Results

blablabla

Observations and Recommendations

blablabla

Conclusion

blablabla