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

During this process I eliminated rehabilitation clinics that appeared twice, bearing different names but the same street address. This sometimes happens when there are multiple divisions in the clinic to keep patients and visitors from getting lost. It's a lot easier to look for and differentiate between the "Klinik an der Weser" and "Paracelsus Rheuma-Klinik" for example than it is using the same name and only adding the division like "Orthopädische Klinik an der Weser" and "Klinik für Rheumatische Beschwerden an der Weser" - which is pretty much the same for the layman.

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

Clinic Locations

The names and street addresses of the clinics were compiled by hand, mainly using a single website. During the process of building up the list I already eliminated any clinics that were listed twice because the website had several lists of rehabilitation clinics, depending on the complaint. However, many clinics are not specialized and so were listed on several lists.

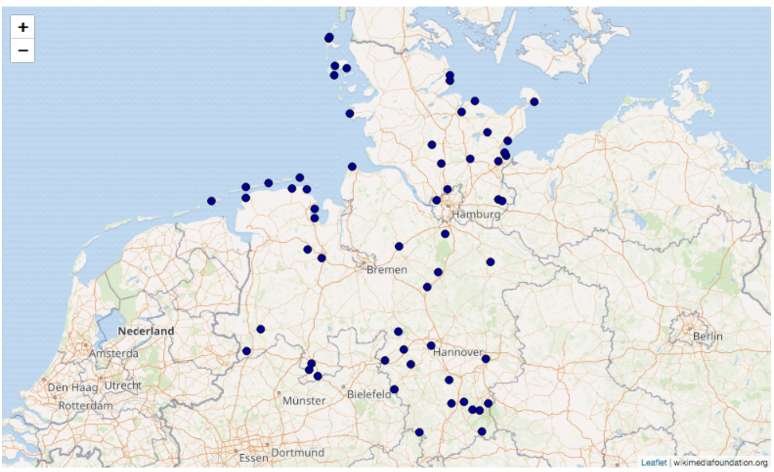
Being a native of Germany I was also able to spot those addresses that did not quite fit the official postal codes and representation of the address. These I checked against each clinics website and if that didn't help asked the official postal service website to point me to the correct representation. This was important because the GeoPy package needs the full, correct address to find the geolocation.

Once I was done I concatenated all parts of the address (street and house number, postal code, village/town) into one column and ran the results through the GeoPy package which returned each clinic's coordinates as in longitude and latitude and appended them to the dataframe. Checking for none-values there were none but the code is there for future use to both list and drop the rows. Of course, any future data scientist should then go and double check any addresses not found.

Since these are simple addresses and geolocations there was no need to actually analyse the data but I did add a sanity check by printing out the lowest and highest values of both latitude and longitude in order to eyeball them.

minimum values	maximum values
`Latitude     51.615529`	`Latitude     54.941989`
`Longitude     6.664696`	`Longitude    11.205156`
`dtype: float64`	`dtype: float64`

As they seemed reasonable enough I then mapped them using folium to get an idea how all the rehabilitation clinics were actually distributed in the area.



This map could also be used as a sanity check: The distribution is relatively even throughout the area, with some clustering near the coast and mountainous areas.

Clinic Venues

Using the geolocation data now contained in the dataframe of rehabilitation clinics I decided to only call to the Foursquare API once instead of three or more times separately for every complaint since the basic parameters would have been the same for all of them.

Having lived here all my life I know that rehabilitation clinics are usually located outside villages or suburbs in their own, often extensive grounds. After some deliberation I chose a radius of 5,000 m or 5 km around the geolocations of the clinics. This is a distance that is, barely, within walking distance, and certainly easily within driving distance without the need to drive longer routes and take a lot of time out of the day. I also set a limit of 100 venues both to keep the time needed to retrieve the venues down to a reasonable value and also to keep from exceeding the daily limit as "venues nearby" is a premium service.

Overall 1767 were retrieved and matched with the relevant clinic. These were divided in 221 unique venue categories which are listed in the appendix. At this point I did not employ any other methods like statistics or visualizations but decided to look at each of the problems and the venues retrieved separately.

Places to Stay

Since the list of venues retrieved is in American English I searched through the list by hand to get all venues intended for anyone to stay the night. This got me the list of Hotel, Hostel, Bed & Breakfast and Vacation Rental. I deliberately left out camping grounds since it seems unlikely that a significant number of people visiting a patient actually own a camper van or trailer.

Overall 115 places to stay were filtered out from all venues. While it might be expected that these would be distributed evenly between all clinics this was not actually the case. 16 or around 1/3 did not have an places to stay listed. However, as mentioned above clinics are often located away from villages/suburbs and it may well be that there are a number of hotels just outside this radius. I could have made another call to Foursquare with a larger radius of, say, 7500 m or 7.5 km but decided against it. The complaint we are looking at here was that there are no places to stay nearby and 7.5 km can hardly be defined as nearby.

Places to Eat At

Here I first filtered out all venue categories that contain the word "Restaurant" since this is the obvious place for someone to eat out at. Moreover, there were many kinds of Restaurants and one of them might easily have been missed when scanning the list manually. On top of that I did scan the list manually for other venues people might decide to eat at - a burger joint might not appeal to the older generation but these days it's a lot more likely that there are younger patients and there visitors will be younger as well. Again I had to scan the list manually because nobody in Germany would be talking about a burger Joint or a BBQ Joint. The result was a list of 40 venues:

```
['American Restaurant', 'Italian Restaurant', 'Middle Eastern Restaurant', 'Modern European Restaurant', 'German Restaurant', 'Vietnamese Restaurant', 'Spanish Restaurant', 'Sushi Restaurant', 'Restaurant', 'Doner Restaurant', 'Fast Food Restaurant', 'Greek Restaurant', 'Asian Restaurant', 'Seafood Restaurant', 'Grilled Meat Restaurant', 'Falafel Restaurant', 'Molecular Gastronomy Restaurant', 'Syrian Restaurant', 'Indian Restaurant', 'Turkish Restaurant', 'Mediterranean Restaurant', 'Russian Restaurant', 'French Restaurant', 'Mexican Restaurant', 'Halal Restaurant', 'Kebab Restaurant', 'Vegetarian / Vegan Restaurant', 'Eastern European Restaurant', 'Chinese Restaurant', 'Thai Restaurant', 'Japanese Restaurant', 'Brazilian Restaurant', 'North Indian Restaurant', 'Swiss Restaurant', 'Steakhouse', 'Bistro', 'Trattoria/Osteria', 'Pizza Place', 'Burger Joint', 'BBQ Joint']
```

Filtering the dataframe then found 466 restaurants for all clinics. As was already seen for places to stay these were not evenly distributed and 10% of the clinics did not have any places to eat at nearby. Please see the Results section for a further discussion of this point as it is more relevant to the results than to the methodology.

Interesting Venues Nearby

Rather than building a list of interesting venues I decided to drop a number of venues from the dataframe:

- all places to stay at since these were part of their own problem.
- all places to eat at since again these were part of their own problem
- places nobody (or very, very few people) would think of to either visit themselves or take visitors to:
  ```
  ['Rental Car Location', 'Train Station', 'Light Rail Station', 'Airport', 'Airport Terminal', 'Gas Station', 'Intersection', 'Insurance Office', 'Advertising Agency', 'Business Service', 'Big Box Store', 'Train', 'City Hall', 'Military Base', 'Construction & Landscaping', 'Truck Stop', 'Nightclub', 'ATM']
  ```
- shops or stores to buy food or essentials from. While this might be of interest to the patient in case they were unsatisfied with what the clinic provided or had forgotten something at home it's not relevant for this problem.

This left me with 140 unique categories.

140 categories are a quite a lot for a machine learning algorithm to deal with and search for underlying patterns in. However, dimensionality reduction by grouping single categories into a broader one did not seem to make much sense here. There *is* a difference between a history museum and science museum after all as well as between a bar (generic term) and a beach bar. So I decided to go ahead without making any changes.

Filtering the dataframe for these 140 categories again showed a diverse result with the number of venues per clinic reaching from just 1 to 56:

Klinik Gesundheitszentrum Hannover	56	Reha-Klinik Damp	2
Städtisches Klinikum Braunschweig gGmbH	53	Klinikum Bad Salzdetfurth	1
MediClin Klinikum Soltau	46	REHA-Klinik Lehmrade	1
Klinik Schloß Warnsdorf	44	Dörenberg-Klinik GmbH	1
Rehabilitations-Zentrum Oldenburg GmbH	39	Rehabilitationsklinik Fallingbostel	1

There was also one clinic that did not have any venues nearby which will be discussed further in the results section.

Machine Learning

This a problem for unsupervised machine learning as there are no labels provided for known data. The goal here is to group all clinics together in group that have the same or similar interesting venues nearby.

The algorithm used to build these groups, or clusters, is k-means which is fast and reliable. Its goal is to cluster data points by determining similarities and differences between them. Data points that are similar to each other - or lie near other - are clustered together. K-means clustering tries to minimize distances within a cluster and maximize the distance between different clusters. In this case we are taking about the distribution of the different venues, not the actual distances in longitude and latitude.

Since kmeans cannot work with venue categories the data has be transformed first. I used one-hot encoding so that each category was placed into one column. This produced one row per clinic and venue which is not useful for machine learning so this dataframed was grouped to represent each clinic in just one row. Rather than using mean values I opted for using the sum of each venue since I felt that this represented this particular situation better. Patients are interested in the actual number of venues not any mean values and this should also be represented in the clusters.

kmeans is not able to determine the number of clusters by itself. This value had to be provided by me. The very unbalanced dataset as well as the problem itself represented a challenge here. There is general rule of thumb that suggest taking the square root of half the number of samples, here clinics. In this case that would be the square root of 30, approximately 5. Using this would probably only take the top 5 venues (see Results section) into account and group the clinic accordingly. This does not really fit the problem's goal, which is to give patients a wider choice of clinics. I decided to go with a value a quarter of the number of clinics: 15. This seems unreasonably high but did a great job at clustering for three main groups and showing up outliers which would otherwise have found themselves as part of one of the left-over two groups had I used just 5.

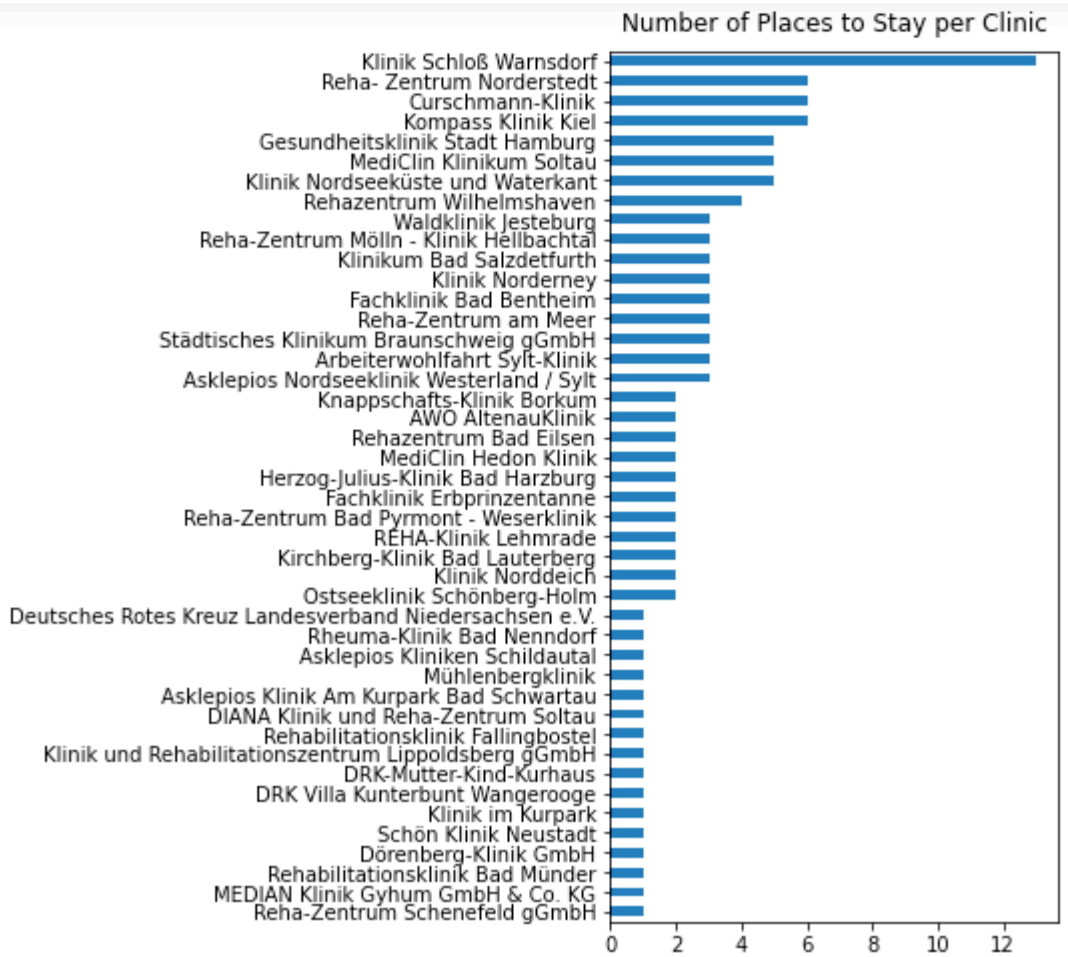
A definite disadvantage of using unsupervised learning is that the algorithm used is quite opaque and it's very difficult even with visualizing and printing out the venues associated with each cluster to make any sense of the underlying mechanism of the clusters and why some that seemed similar were not put into one cluster. For a further discussion of the clusters build by the algorithm see Results section.

Results

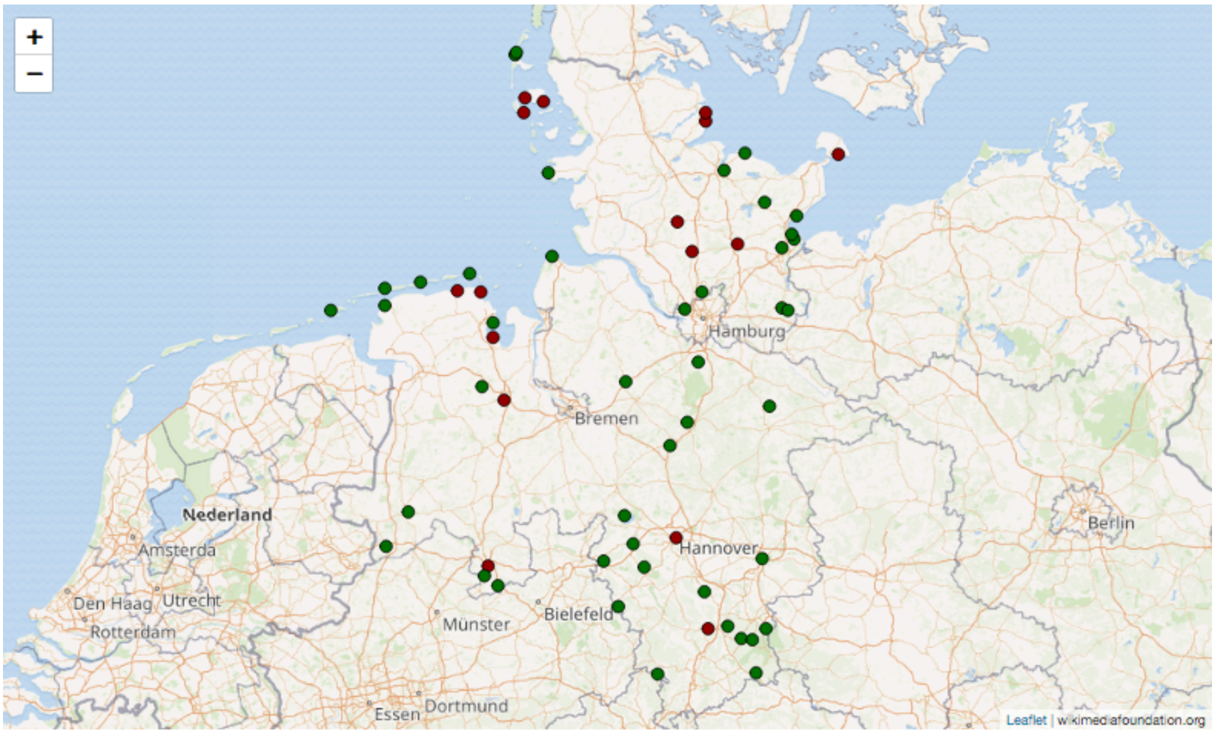
Places to Stay at for Visitors

To determine places to stay the location data obtained for each rehabilitation clinic was filtered to contain only hotels, hostels, holiday rentals and bed & breakfast places. Holiday rentals may be less interesting to visitors but with the recent covid-19 crises significantly more people prefer to cater for themselves rather than share space with other travellers. I didn't include camp sites as these would only be relevant for people who own a camp trailer or camper and are willing to use it for just two or three nights. Usually neither is available for rent short-term.

Over all clinics 115 places to stay at were found but only 44 out of all 60 clinics have at least on hotel, hostel, bed & breakfast or holiday rental nearby. The number of places within this group also varies from just one up to thirteen.



Rather than using the count of places of stay as shown in the plot above I opted to enrich the dataframe containing the clinics, their addresses and geographic location by a column indicating whether there is at least one place to stay at. This column were then used to show the clinics on the map, using green circles for clinics with places to stay at and red ones for those that didn't have even one place nearby.

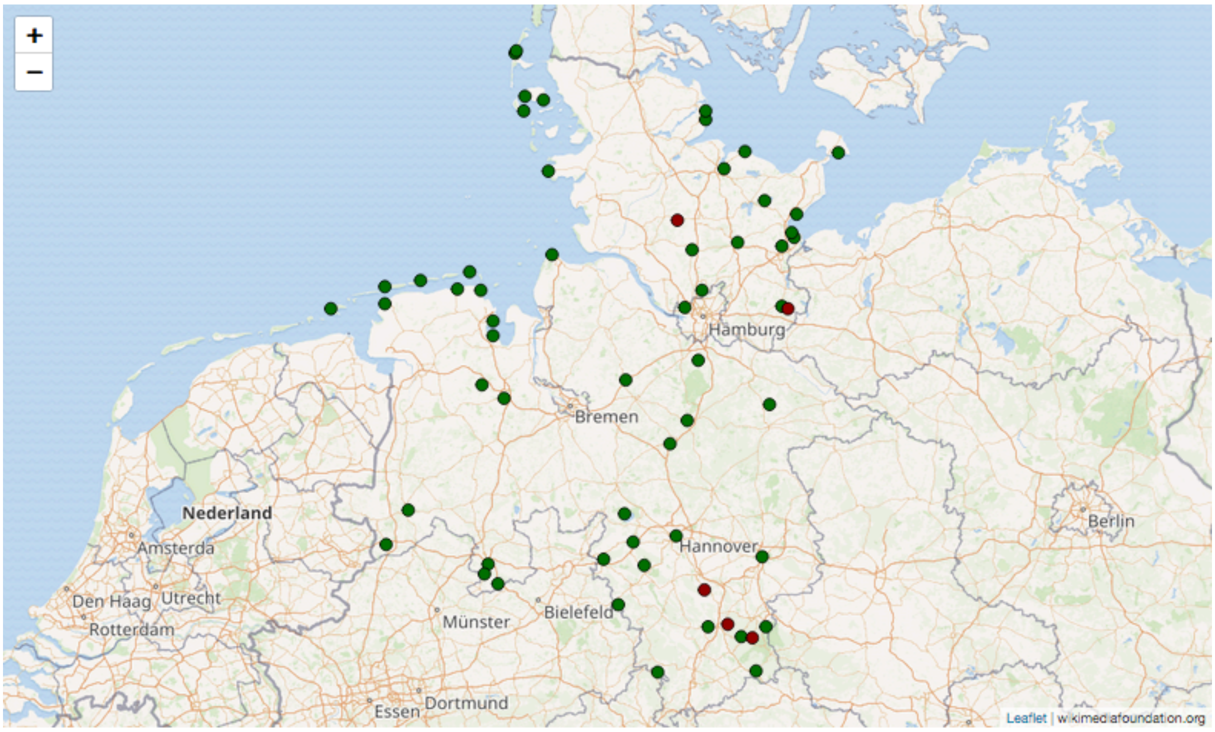


It is interesting to see that about half of the clinics near the seaside don't have any places to stay at nearby while all of them are favourite holiday destinations. There is, however, no need to over-interpret this as it might simply be the case that the clinic is shoved outside the community to make more place for paying holiday-makers and the hotels are actually near the seaside and more than 5 km from the clinic. Except for the one in Hannover all the others are in rural areas and there probably really are no places to stay at for casual visitors as hotels in these regions tend to cater to firms that host retreats for their employees as well as seminars.

Places to Eat at

Using the same location data as obtained earlier I filtered vanue categories for anything containing the word "restaurant" augmented by the list Steakhouse, Bistro, Trattoria/Osteria, Pizza Place, Burger Joint and BBQ Joint. There may have been a couple more venues to have a full meal at that I didn't recognize as such because of the American notation. The whole list gave me 40 different kinds of restaurants over all clinics. 466 restaurants were found for all clinics together, getting a statistical mean of 7.7 restaurants per clinic. The reality looks different and is as varied as with places to stay. 6 clinics don't have any restaurants within a radius of 5 km, many have only one or two and a handful has more than thirty to choose from.

For the map I took the same approach as I have done for places to stay and simply calculated a column showing whether there were any restaurants nearby or not. Again green circles stand for clinics with nearby restaurants while any without are depicted by a red circle.

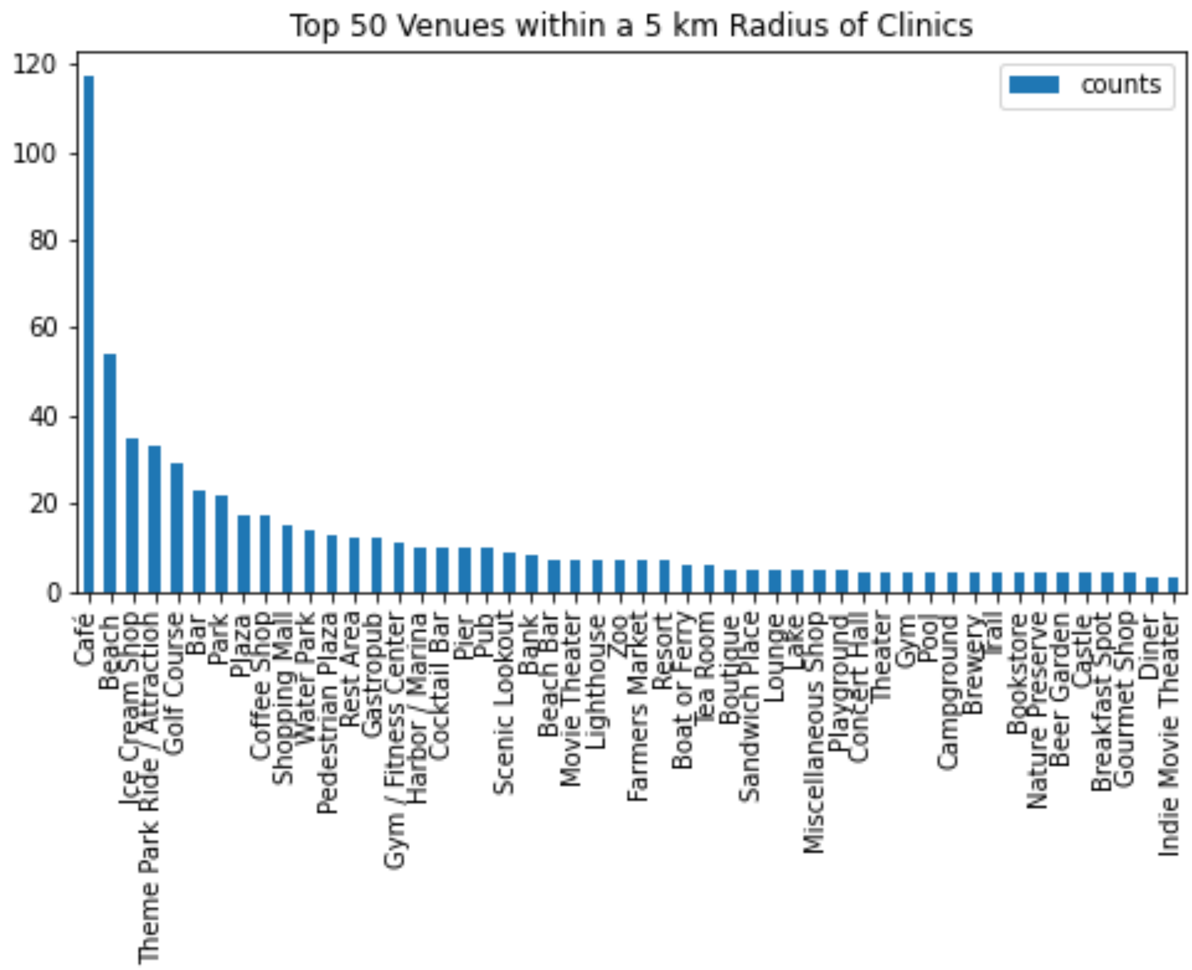


Again the fact that there were no restaurants found nearby may be accounted for by the fact that clinics tend to be some distance away from village commons and town centres. It must also be noted that the covid-19 crises has let to a lot of restaurants being closed for either the duration or even for good as they haven't been allowed to be open to patrons for over four months by the time of writing this report. This may have let to them being dropped from Foursquare data. Personally, I have been to the place for the red circle just below Hannover and I know for a fact that there used to be several restaurants in the area before the crises. In fact, Google maps shows me a couple even now. This section will have to be re-evaluated once the crises is over, restaurants are allowed to open without restrictions and our world settles into a new normal.

Nearby Venues

Using the location data obtained from Foursquare some more extensive cleaning had to be applied which is detailed in the methodology section. Suffice it to say that all places to stay and places to eat as well as any places not of interest (such as "Advertising Agency") were filtered out. This left me with 140 unique venue categories that might be of interest to go to with a group of other patients or with visitors. I decided not to reduce this further, partly in order not to introduce a bias into the data, partly to have some data for the machine learning algorithm to work on. With reducing dimensionality to a readily comprehensible amount of broader categories machine learning would probably not have been necessary and no new insight into the data would have been generated.

While there is again a high variety of venues there is also a high varied in how often these occur. Even among the top 50 there are only a couple that are really common overall:

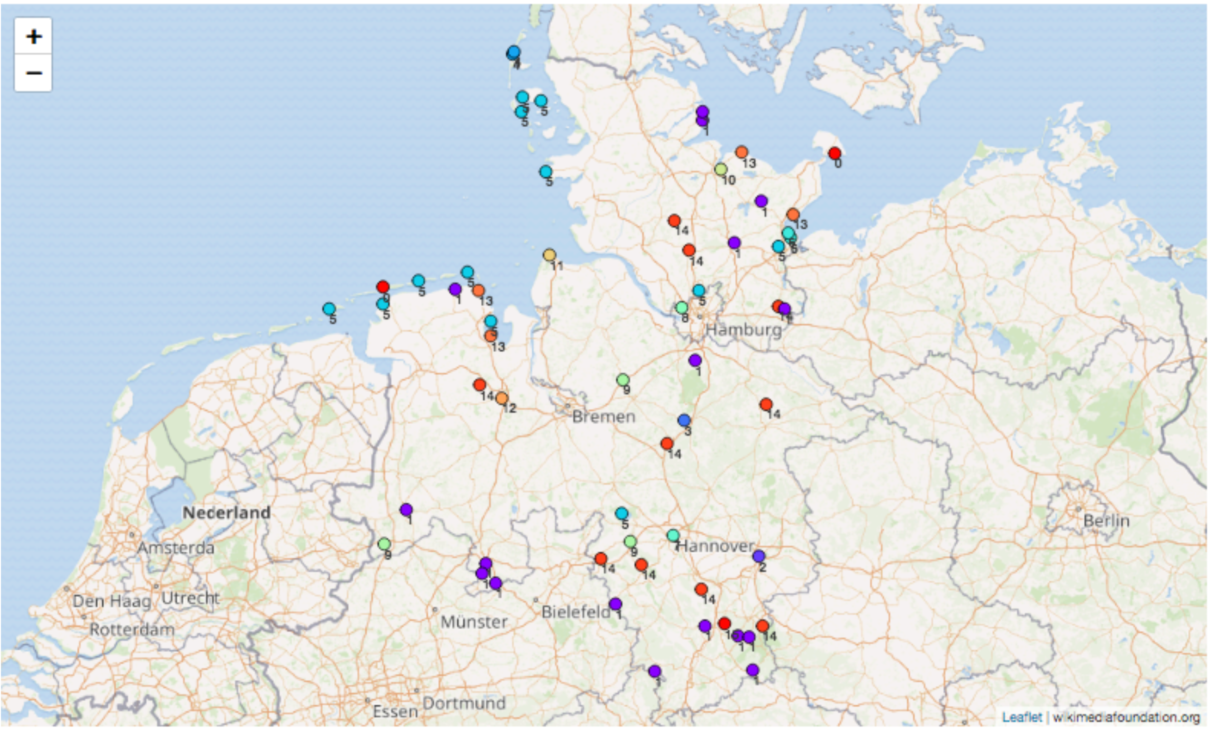


It seems that Cafés and Ice Cream Shops are really popular in North-West Germany! I was surprised to find that a golf course seems to be present at at least a third of the clinics. This might be unintentional bias though because many of the clinics tend to be situated on mostly flat land to cater to patients who would otherwise be unable to enjoy the grounds. This sort of ground is just as well suited to golf courses.

Once the dataframe was filtered two of the clinics did not have a interesting venues nearby. This was later reduced to just one clinic as the other two bear the same name, are near to each other and could be clustered together as far as this report is concerned. The one clinic left over with no venues was placed into a cluster all by itself. All the others were places into 15 clusters (0 to 14) by the machine learning algorithm. While logic tells me that this would mean that there are around four clinics per cluster, giving a patient a good way to choose the sort of location that appeals to them, the reality was somewhat surprising. There were three dominant clusters and quite a few that contained only one clinic:

cluster	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
counts	2	17	1	1	2	12	2	1	1	3	1	1	1	4	10	1

This result can be shown on a map showing every cluster in a different colour. In order to make it easier to distinguish between similarly coloured clusters and to help colourblind people to make sense of the map I also added the number of the cluster beneath the circle.



One thing that is really obvious even to the casual glance is that almost all clinics from cluster 5 are located near the seaside. This made me surmise that "Beach" which is the second most common venue overall would be the most common venue for this cluster. However, the most common venue is a café. It's top 1 for every of the 12 clinics in this cluster. The top 2 venue includes the beach only twice, the most common one is an ice cream shop.

Cluster 0 could probably be added to cluster 5 as both are located on islands off the coast and for both the most common venue is also a café. The second most common venues are Harbor/Marina and Golf Course respectively which might be why the algorithm did not place them in the same cluster.

Cluster 13 is also located near the seaside but on the shore side rather than an island. For all of the them the top 1 venue is the beach followed by Harbor/Marina which is clearly different from cluster 5 and most likely the reason to place them into a cluster of their own.

Most clinics in cluster 14 which is also the third largest cluster are located in rural area. The most common venue for them is the golf course and in three cases it is also the only venue within a radius of 5 km. So this cluster would be aimed at the sports-minded individual who isn't adverse to learning something new beyond health topics.

Clinics in cluster 1 are also rather farther away from large towns and also tend to be in mountainous or hilly terrain. There is no clear top 1 venue but points of interest like lake, reservoir, nature preserver and hot spring are found in the list of venues overall. It should also be noted that none of the clinics has more than six venues nearby and eight have no more than two. While this is not part of the clustering mechanism it can also be noted that in this cluster 7 out of 17 clinics do not have a any places to stay at nearby.

Cluster	Most Common Top 1 Venue	Most Common Top 2 Venue	Remarks if any
0	Café	Boat or Ferry and Ice Cream Shop	located on islands off the coast
1	-	-	venue for every neighbourhood different; many don't have more than two venues altogether
2	Café	Park	
3	Theme Park Ride / Attraction	Boutique	
4	Café, Beach	Beach, Café	both are located on the northernmost island
5	Café	Ice Cream Shop (only 3 out of 12)	
6	Café	Beach	might be grouped with 4
7	Café	Coffee Shop	only clinic that is located in a big city
8	Shopping Mall	Ice Cream Shop	
9	Rest Area	-	this is also the Top 2 for the only outlier
10	Café	Gym / Fitness Center	
11	Beach	Ice Cream Shop	
12	Café	Ice Cream Shop	
13	Beach	Harbor / Marina	
14	Golf Course	-	
15	-		this was the only clinic that did not have any interesting venues nearby

Looking at the table makes it different to decide why exactly the algorithm did not put a couple of them into the same cluster. However, since geographic location did not play a role and the whole dataset was seriously unbalanced it seems sensible that the overall top venues are also the top venues in the different clusters. This table should not be used for patients and doctors as it will very likely add to confusion rather than help defuse it.

Seeing the different clusters on the map will help them decide as most people have at least some idea about the greater area and state they live it. It's not very common to send a patient to a clinic located in another state or jurisdiction so specialized knowledge of the topography of all of Germany is not necessary for them.

Observations and Recommendations

Observations

After analysing the location data provided by Foursquare it is easy to see that there is a great diversity between the clinics:

- 30% of them don't have a place for visitors to stay within a radius of 5 km
- The same goes for places to have a meal at which is absent at 10% of the clinics.
- A number of clinics also don't have any interesting venues within the same radius at all and what is on offer at those that do is rather limited as well.

This objective data makes it more likely that complaints and dissatisfaction are real rather than the result of general disgruntlement on the patient's side. It was probably quite normal and expected not to have much contact outside the clinic in former times but these times have changed and people have different expectations now. As complaints have been rising these changes were likely not taken into account in recent times. Even if places to stay and/or eat at don't come into being overnight it might have been possible by the clinics to set aside a couple of rooms for visitors and provide a decent if simple meal when asked for well in advance. As far as places to visit this is a more difficult problem to solve for the clinics themselves but using a location service to give patients a choice beforehand might head off a lot of these problems as patients would be much more likely to be placed in an environment that appeals to them.

It must be made clear here that the above points are also not necessarily the fault of the clinics and their administrations. In the case of places to stay it might simply never have occurred to someone that this can be a reliable way of earning money. The same may also apply to restaurants. That is also dependent on the community where the clinic is located. If there is little likelihood that a restaurant is also going to be used by the local population and general holiday makers there might not be the incentive to run one as it might well operate at a loss.

The consortium will have to think about this hard and perhaps work with the clinics to rethink some of their policies.

Recommendations

Based on the results and observations three main recommendations can be made:

Web Application

A web application has the advantage that it is easily accessible on nearly every device and does not need the maintenance efforts an app for both iOS and Android would need.

This application would then consists of three different maps showing

1. clusters of clinics to chose from based on interesting venues then can be reached easily on foot or by bike/car.
2. symbols that mark for every clinic whether there are any hotels or bed & breakfast relatively places nearby for visitors to stay over the weekend.

3. symbols that mark for every clinic whether there are any restaurants relatively nearby to have a meal at with visitors or while out with a couple of other patients.

In all three cases a click on a specific clinic could zoom the map to show just the vicinity of each clinic, clearly marking where places to stay and places to eat at are located as well interesting venues. The latter could be divided into sub-categories rather than showing each individually and relying on mouse-over functions to provide more information.

The application should also contain information about each clinic individually, both listing any information pertaining to the clinic itself and the results returned by the clustering, e.g. citing the top 5 interesting venues for each clinic without the need to refer to the map.

Application Form

It would take little effort to add three questions to the application form to find out whether patients are interested in receiving visitors and/or going out for a meal. To keep the resulting process as simple as possible for the health insurers this should simply be yes/no-questions. Based on the answers a database based on the results from the location service could quickly provide a list of clinics that are suitable for this particular patient.

On top of that there should be question which cluster of clinics the patient wishes for and has been able to choose using the above mentioned web application. Allocating a specific clinic could then be made easy by enriching the database with the clusters resulting from the machine learning process.

However, this process needs to be thought through carefully and priorities defined as there will be instances when the wishes and cluster chosen collide, especially when there are no places to stay and eat at nearby. In this case the yes/no-questions should have priority over the cluster and a similar cluster chosen.

Future Plans for New Clinics

As the results have shown there are a lot of differences between the locations of the clinics and the venues associated with them.

An intensive study of the complaints and negative ratings of the surveys should be made and a database build from them which is constantly updated by and compared to survey made after above measures have been implemented.

The results from both have to be taken into consideration when plans for building or upgrading a clinic are drawn up. There will often be the case that no land is available near venues but if many guests have wished for a place to stay nearby then a hostel or bed & breakfast could be build on the grounds and an external operator for a restaurant found. Alternatively a bus shuttle service could be planned together with the town's council to connect the clinic to the community while still serving the regular population as well.

Conclusion

When this project has been completed the consortium has a couple of objective and reliable tools on hand to

- judge and classify outright complaints and general dissatisfaction
- give patients and their doctors an equally reliable way for taking a active role in choosing the location of a clinic

It is essential that both points are taken seriously over many years and not just being used as a tool for the present. People's wishes and priorities change over time and it is important to constantly monitor which direction they are taking. It is very likely that in the upcoming months following the pandemic people want to have visitors for every single weekend - and once memory of the restrictions during the pandemic fade it is well within scope that patients don't want to see anyone from home while away for a health cure. Both details concerning dissatisfaction as the statistics of patients' wishes provide and excellent way by being able to rely on readily available and objective data.

In general however, while some conflicts are certain to arise, using up to date location data and making it public, too, should help increase patient satisfaction as far as the location of the clinic is concerned.

Appendix

List of all Venues Retrieved, Sorted Alphabetically

```
['ATM', 'Accessories Store', 'Adult Boutique', 'Advertising Agency',
'Airport', 'Airport Terminal', 'American Restaurant', 'Aquarium', 'Asian
Restaurant', 'Athletics & Sports', 'Automotive Shop', 'BBQ Joint', 'Baby
Store', 'Bakery', 'Bank', 'Bar', 'Basketball Stadium', 'Beach', 'Beach
Bar', 'Bed & Breakfast', 'Beer Bar', 'Beer Garden', 'Beer Store', 'Big Box
Store', 'Bistro', 'Board Shop', 'Boarding House', 'Boat or Ferry',
'Bookstore', 'Botanical Garden', 'Boutique', 'Bowling Alley', 'Brazilian
Restaurant', 'Breakfast Spot', 'Brewery', 'Building', 'Burger Joint', 'Bus
Stop', 'Business Service', 'Cafeteria', 'Café', 'Campground', 'Canal Lock',
'Castle', 'Cave', 'Chinese Restaurant', 'Chocolate Shop', 'City Hall',
'Climbing Gym', 'Clothing Store', 'Cocktail Bar', 'Coffee Roaster', 'Coffee
Shop', 'Comic Shop', 'Concert Hall', 'Construction & Landscaping',
'Creperie', 'Cupcake Shop', 'Currywurst Joint', 'Deli / Bodega',
'Department Store', 'Dessert Shop', 'Diner', 'Doner Restaurant',
'Drugstore', 'Eastern European Restaurant', 'Electronics Store', 'Event
Space', 'Falafel Restaurant', 'Farm', 'Farmers Market', 'Fast Food
Restaurant', 'Fish & Chips Shop', 'Fish Market', 'Flower Shop', 'Food',
'Food & Drink Shop', 'Food Court', 'Forest', 'French Restaurant', 'Fried
Chicken Joint', 'Frozen Yogurt Shop', 'Furniture / Home Store', 'Garden',
'Garden Center', 'Gas Station', 'Gastropub', 'General Entertainment',
'German Restaurant', 'Gift Shop', 'Golf Course', 'Gourmet Shop', 'Greek
Restaurant', 'Grilled Meat Restaurant', 'Grocery Store', 'Gym', 'Gym /
Fitness Center', 'Gym Pool', 'Halal Restaurant', 'Harbor / Marina',
'Hardware Store', 'History Museum', 'Hockey Rink', 'Hostel', 'Hot Spring',
'Hotel', 'Ice Cream Shop', 'Indian Restaurant', 'Indie Movie Theater',
'Intersection', 'Irish Pub', 'Island', 'Italian Restaurant', 'Japanese
Restaurant', 'Juice Bar', 'Kebab Restaurant', 'Kitchen Supply Store',
```

```
'Lake', 'Light Rail Station', 'Lighthouse', 'Liquor Store', 'Lounge',
'Mediterranean Restaurant', "Men's Store", 'Mexican Restaurant', 'Middle
Eastern Restaurant', 'Military Base', 'Miscellaneous Shop', 'Mobile Phone
Shop', 'Modern European Restaurant', 'Molecular Gastronomy Restaurant',
'Motel', 'Mountain', 'Movie Theater', 'Multiplex', 'Museum', 'Music Store',
'Music Venue', 'National Park', 'Nature Preserve', 'Nightclub', 'North
Indian Restaurant', 'Opera House', 'Optical Shop', 'Organic Grocery',
'Outdoor Sculpture', 'Outdoor Supply Store', 'Outdoors & Recreation',
'Outlet Mall', 'Outlet Store', 'Palace', 'Park', 'Pastry Shop', 'Pedestrian
Plaza', 'Performing Arts Venue', 'Perfume Shop', 'Pet Service', 'Pharmacy',
'Photography Studio', 'Pier', 'Pizza Place', 'Playground', 'Plaza', 'Pool',
'Pub', 'Public Art', 'RV Park', 'Rental Car Location', 'Reservoir',
'Resort', 'Rest Area', 'Restaurant', 'Rock Climbing Spot', 'Rooftop Bar',
'Russian Restaurant', 'Salad Place', 'Sandwich Place', 'Scenic Lookout',
'Science Museum', 'Seafood Restaurant', 'Shoe Store', 'Shopping Mall',
'Shopping Plaza', 'Skate Park', 'Skating Rink', 'Ski Area', 'Snack Place',
'Soccer Field', 'Soccer Stadium', 'Spanish Restaurant', 'Sports Bar',
'Stadium', 'Steakhouse', 'Supermarket', 'Surf Spot', 'Sushi Restaurant',
'Swiss Restaurant', 'Syrian Restaurant', 'Taverna', 'Tea Room', 'Thai
Restaurant', 'Theater', 'Theme Park', 'Theme Park Ride / Attraction',
'Tourist Information Center', 'Track', 'Trail', 'Train Station',
'Trattoria/Osteria', 'Tree', 'Truck Stop', 'Turkish Restaurant', 'Vacation
Rental', 'Vegetarian / Vegan Restaurant', 'Vietnamese Restaurant', 'Water
Park', 'Waterfront', 'Wine Bar', 'Wine Shop', 'Zoo', 'Zoo Exhibit']
```