Analysis of healthcare providers' distribution in Vienna
IEOR 256: Project report

Student Name: Guillaume Domenge
Student ID: 3039675732

February 9, 2025

Introduction

Distance and availability are two very important factors that patients take into consideration when looking for healthcare practitioners, this statement follows the notion that individuals will try to minimize the disruption that a healthcare visit will have on their day to day life, and therefore, will prefer to spend less time traveling to and waiting for a certain doctor.

In this project we intend to use both information regarding the location of healthcare practitioners in the city of Vienna and demographic data for the same city, in order to see whether we can draw some conclusions on the distribution of this healthcare practitioners and possibly devise a plan on how to better the current distribution. For this we used two datasets.

The first dataset consists of a list of healthcare providers registered on the https://www.praxisplan.at/ website downloaded from https://www.kaggle.com/datasets/kamyababedi/doctors-locations-vienna?resource=download where it was uploaded by the Kaggle user kamyababedi, who appears to have since been deleted.

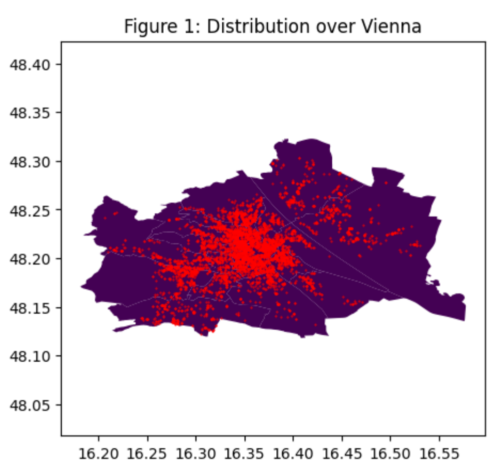
The second dataset consists of a simple data table of the census data from the city of Vienna, which we obtained by using the Austrian Statistical Database at https://statcube.at/.

Data Analysis

In this section we'll walk through different steps of the analysis process, all the while talking about the tools that are being used and interpreting the results.

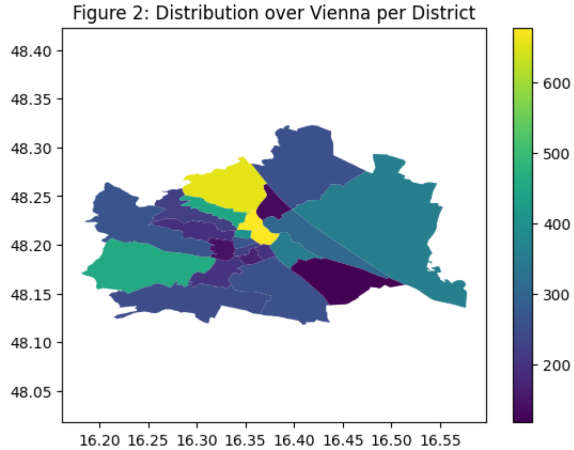
GPS Data Only

We firstly start by simply plotting all of the doctors' GPS coordinates over Vienna, this can be seen in the following graph.



This first graph, although quite pretty, doesn't provide much information other than showing that generally healthcare practitioners tend to work near the center district of the city, Innere Stadt.
If we wish to show more meaningful information, one thing we can do is group the healthcare practitioners by district. This can be achieved with the following Python packages, overpy for the Overpass API allows us to extract and work with the district borders, and shapely allows us to turn these borders into polygons in order to quickly filter every doctor into the district they belong in. This gives us the following graph.

Figure 2: Distribution over Vienna per District

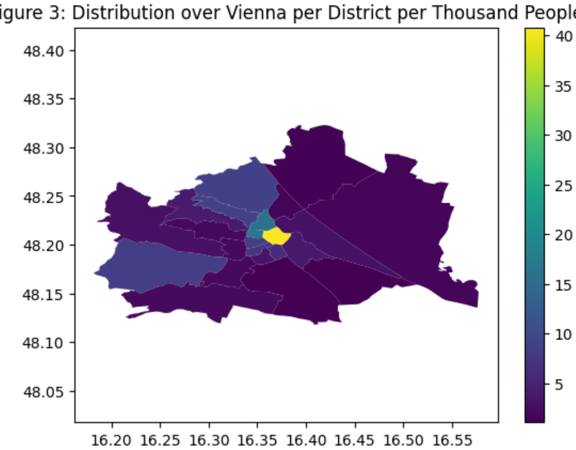


With this graph we can now see that rather than the doctors focusing around the center of the city, they gather primarily in Innere Stadt, Alsergrund and Döbling (the three district in yellow), as we can see other districts that are also close to the city center do not have such high doctor density and Hietzing (center-west) has a particularly high density despite being in the western end of the city. Ultimately this graph does show us more precisely the distribution of doctors, but doesn't quite help us understand why they are distributed this way.

Doctors per Thousand People per District

A plausible explanation for what we see in Figure 2 is that the distribution of doctors follows the demography of the area, to test this hypothesis we decided to graph the distribution per thousand people and this gives us Figure 3.

Figure 3: Distribution over Vienna per District per Thousand People

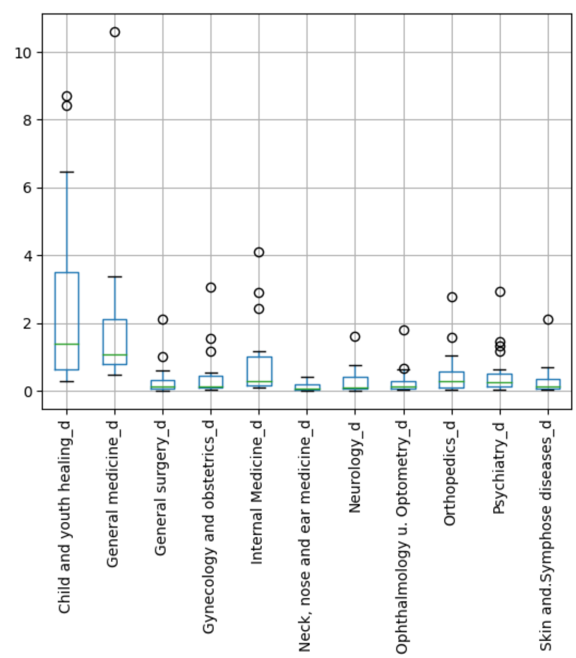


This result shows us that most districts in Vienna will have the same or very similar density of doctors per thousand people, around 5 to 10 doctors. However, some districts will have much higher doctors per thousand people, this is specially true for Innere Stadt with 40 doctors per thousand. This leads us to believe that demography is generally a good tell for how many doctors are in an area, however other factors might also influence this. We suspect it could be explained partly by the fact that most office buildings will generally be found near the city center and this causes the city center to be close to where most people spend time during the day, and, that a city center is naturally a central point for most people. We unfortunately did not find a database of working locations to test this.

Distribution per Specialty

If we want to concentrate what specialty might be lacking in some district, we have to group our healthcare providers by specialty and by district and then plot the box plots in Figure 4. It's important to take in consideration that for Pediatrics (in the graph as "Child and youth healing") is divided by the number of individuals under 14 years old instead of the whole population, as well as General medicine is divided by the number of people who are not under 14. For the sake of simplicity we have also decided to focus on the specialties more than 150 doctors city-wide, reduces the number of specialties to deal with and overlooks specialties such as Nuclear Medicine, that would have no doctors in almost all districts.

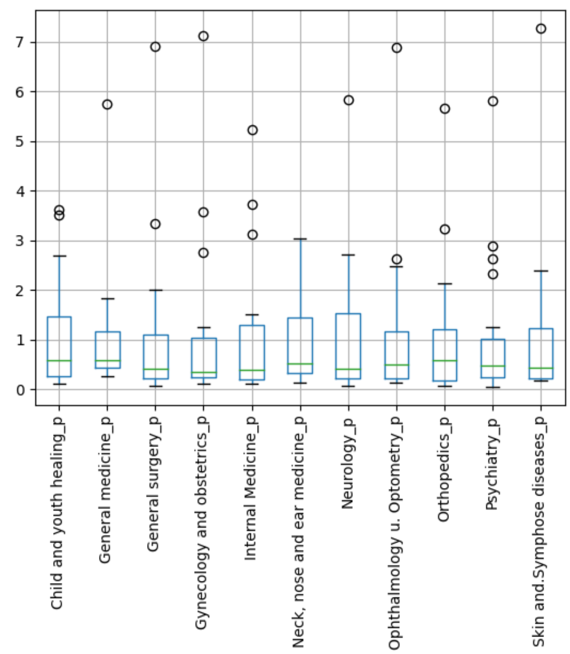
Figure 4: Box plots of specialty per thousand people



Now these graphs might hold a lot of information, however they don't answer whether specialty A is lacking in some place or not. At least not on their own, we'd need to know what is the target concentration of doctors of some specialty per thousand people. Unfortunately, we do not know this and will therefore naively assume the target for each specialty to be their city average.

If we now divide each result by their specialty target we should get comparable results since the target is the same, for the sake of using a short notation, we'll call this normalization (we use no other kind of normalization in this project).

Figure 5: Box plots of specialty per thousand people normalized



In Figure 5 we have plotted the box plots of the normalized results, as explained earlier. Here we can see for example that general medicine relative to its target has a tighter inter-quartile range than the other specialties. This graph also allows us to define a metric of lack for a certain specialty that is comparable between specialties, the distance from its minimum to its target (one). Here we can make out that the specialty lacking the most healthcare providers is Psychiatry since its minimum is the lowest out of all specialties.

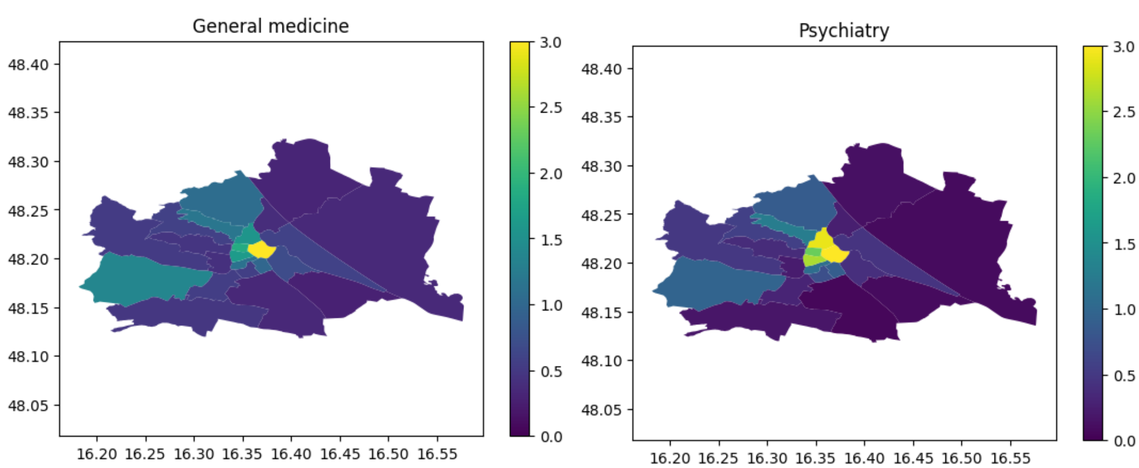


Figure 6: Comparison of normalized density between General Medicine and Psychiatry

In figure 6 we can see how General Medicine is better distributed and most districts are closer to the target of one than in Psychiatry, where a lot of districts are far below the target.

Decision Making

In this section, we want to focus on answering one question, if Praxisplan had enough budget to incorporate X new doctors, from what specialty and what district would they be in order to maximize the impact they'd have?

In order to answer this question we rely on the results from the previous section where we had observed that the minimum of the normalized results corresponds to the specialty and district which would benefit the most in terms of being closer to the specialty target the target per thousand people. This means that for X new doctors we can sequentially take the minimum out of all the normalized scores simulate adding a new doctor of that specialty and district by adding one to its original doctor count and dividing by its specialty target then repeating until the X doctors are added.

The following table is the result of applying the steps had talked previously this section for 100 doctors.

District	Child	General M	General S	Gyne.	Internal	Neck	Neuro.	Ophtha.	Ortho.	Psych.	Skin
Innere Stadt	0	0	0	0	0	0	0	0	0	0	0
Leopoldstadt	1	0	0	1	0	0	1	0	1	0	1
Landstraße	0	0	0	0	0	0	0	0	0	0	0
Wieden	0	0	1	0	0	0	0	0	0	0	0
Margareten	0	0	0	1	0	0	0	0	0	0	0
Mariahilf	0	0	0	0	0	0	0	0	0	0	0
Neubau	0	0	0	0	0	0	0	0	0	0	0
Josefstadt	0	0	0	0	0	0	0	0	0	0	0
Alsergrund	0	0	0	0	0	0	0	0	0	0	0
Favoriten	1	1	1	1	1	1	1	1	1	1	1
Simmering	1	1	1	1	1	1	1	1	1	1	1
Meidling	1	0	1	1	1	1	1	1	1	1	1
Hietzing	0	0	0	0	0	0	0	0	0	0	0
Penzing	0	0	1	1	1	0	0	1	0	0	1
Rudolfsheim	1	0	1	1	1	1	1	0	1	1	1
Ottakring	1	0	1	1	1	1	1	1	1	0	1
Hernals	0	0	0	0	0	0	0	0	0	0	0
Währing	0	0	0	0	0	0	0	0	0	0	0
Döbling	0	0	0	0	0	0	0	0	0	0	0
Brigittenau	1	0	1	1	1	1	1	1	1	1	1
Floridsdorf	1	1	1	1	1	1	1	1	1	1	1
Donaustadt	1	0	1	1	1	1	1	0	1	1	1
Liesing	1	0	1	1	1	0	1	1	0	1	1

We can see one of the benefits of this approach is that it makes the decision making process extremely simple.

Physical Location of New Doctors Within a District

The previous algorithm helped us determine in which district we'd profit from x amount of doctors, but we'd still need to know exactly where physically. For example, if Praxisplan wanted to incorporate three new healthcare practitioners of the General Medicine specialty in Briggetenau, where would be the best place to put them? Generally you want to put new practitioners in a position where they reduce the distance from any point within the district to the closest doctor as much as possible.

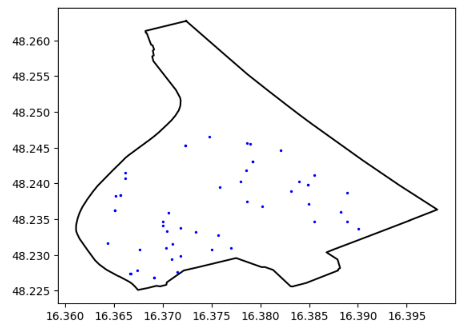


Figure 7: Current General Medicine Practitioners in Briggetenau

To be able to solve this problem we need a way to evaluate the sum of distances to the nearest doctor in the district. And in order to do that we start by generating m, $y_1, ..., y_m$ equally spaced points in the district by generating them in a square containing the district the applying the shapely polygon function to only keep those that are inside the district. Then we define the following minimization problem:

$$\min_{x_1,x_2,x_3} \sum_{i=1}^{m} \min\left(\{d_i, d(x_1, y_i), d(x_2, y_i), d(x_3, y_i)\}\right)$$

where d_i is the distance from point i to its nearest original doctor (blue point in Figure 7), and $d()$ is the euclidean distance.

If we had a set of candidates we could simply iterate over them and simply choose the best subset, but since we don't, we had to treat it as a continuous problem. This causes problems since the min inside the sum makes the problem non differentiable. Due to this we've decided to solve the problem using Genetic Algorithms, more precisely by using the Pymoo package, to which we had to define the coordinates of our three new doctors as variables and the sum in the minimization problem above as the objective function. This resulted in the solution shown in Figure 8, where the red dots are the new doctors.

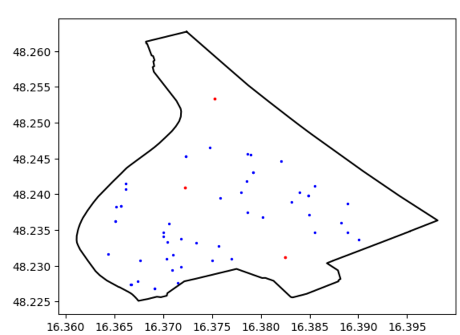


Figure 8: Three new General Medicine Practitioners in Briggetenau

Conclusions and future works

Throughout this project, we have collected, merged, and analyzed real-world data to uncover potential gaps in the Austrian healthcare system. Additionally, we have identified tools that could address these gaps effectively. However, there are limitations to our analysis and the tools we have developed. Firstly, the Praxisplan database may not fully represent the entire healthcare landscape in Vienna. We recognize this as a limitation of the dataset, and we believe that similar work on a larger dataset would provide more comprehensive insights. Secondly, gathering data on people's daytime locations could offer valuable insights into the distribution of healthcare practitioners across districts. Thirdly, establishing explicit targets for the number of doctors per specialty per thousand people is crucial for informed decision-making. Lastly, our analysis of the physical distribution of practitioners assumes uniform population distribution within each district, which does not accurately reflect reality. Incorporating a population density function to generate the relevant data points could significantly enhance the accuracy of this aspect of the analysis.