MODS203: Data analysis in economics I :

Collection and visualization

Baala Mathis - Bernhard Quentin - Lkhaoua Alexandre

1 - Give an introduction about your topic.

Nowadays, an important issue has appeared in many public health care systems due to too long waiting times for medical services. Especially when people live in a medical desert where the number of doctors is low. It is all the more so when it deals with specialists. The minister is often criticized for not dividing the new doctors equally, it means that a lot of them go to big cities or places where it feels good to live. For some people, that boils down to inequalities in waiting times between town and countryside. Even within cities, excess fees played a major role because it is often discussed that doctors give preference to patients who pay more.

The main driver of our project will be the following question :“Is it true that the richer we are, the better access to medical services we have ?” and it will lead to the question “To what extent can we establish a link between waiting times for medical services and location across the country as well as willingness to pay?”. Indeed, as it has been said before, people who live in high population density areas can probably get access to medical services much more quickly than people living in medical deserts. Moreover, one can wonder if patients who are willing to pay more through copayments can get rid of a long waiting time for their desired medical service.

To answer this question, we will collect data from a French medical service provider named Doctolib. As Doctolib is almost a monopolist, we think that a part of its database will be enough to be a representative sample. The collected data will be ordered in different categories : the waiting times to have an appointment, the specialization of the doctor, his location and his copayment fees. Then we will analyze the data collection in order to verify whether there exists a link between waiting time, location, medical specialization and copayments.

This type of research has been already made by journalists and is widely discussed as it was an all-important issue during the Covid-19 crisis, which pointed out a lot of dysfunctions in health systems. Indeed, hardly a month goes by without a new article on this topic. For instance Christelle Millien, Hélène Chaput and Marie Cavillon highlights the difference between having an appointment with a generalist and with other specializations (https: //drees.solidarites-sante.gouv.fr/IMG/pdf/er1085-2.pdf). And the French Newspaper *Le Monde* points out the issue of medical deserts in France, in the lower population density areas. Le monde, Déserts médicaux : l’accès aux spécialistes est de plus en plus difficile, Camille Stromboni.

https://www.lemonde.fr/societe/article/2022/09/27/deserts-medicaux-l-acces-aux-specialistes-l-autre-fracture-qui-s-aggrave_6143321_3224.html
Youtube - Le monde - Pourquoi y a-t-il une pénurie de médecins en France ?

2 - Describe your data collection strategy.

We plan to use France's most important platform to take appointments with medical services named Doctolib. It is very simple to understand the principle of this device : write what you are looking for (specialization) and in what place (city), then a list of calendars appears for each doctor available in the area. The purpose of this data is to notice the waiting time to have an appointment in function of the specialization and the city chosen. Collect these data in different cities and organize them by specializations to respond to the problematic.

Our research is based on an observation of appointments. The focus is the possibility to have an appointment with the right specialist in the city asked. The data has 4 dimensions. Indeed, with an appointment we extract the following things : the doctor (in a way, the specialization), the waiting time to have the appointment, the price of the consultation and the place (the city) where the appointment is supposed to take place. After making a few tests, we noticed that all these characteristics around the appointment are available on Doctolib and can be extracted to be used in our work.

As we said before, we really want to extract the specialization, the waiting time, the price and the city in order to look for a correlation between all these variables. The goal is to have a tab of the waiting time for an appointment in function of the specialization, the city and the fact that the doctor resorts to copayments. For instance, the tab could look like the following one.

City	Paris		Strasbourg	
Specialization	Copayments	No Copayments	Copayments	No Copayments
Cardiology	15 days	110 days	20 days	33 days
Ophthalmology	…	…	…	…
Gynecology	…	…	…	…
Dermatology	…	…	…	…

We will get access to the data through the Doctolib site directly, our idea is to scrap the data looking for the main features we are interested in. We plan on using the

selenium library (only works with google chrome) because we already are a bit familiar with it.

At first sight there are no blocking pop ups or captchas that we will have to deal with. Furthermore, we do not need an account, we can scrape the site as a guest. There are other websites like KelDoc or Ameli that let one find doctors around a certain area but they are approximately the same as Doctolib so we believe that we are going to stick with this site.

Our strategy is to iterate between every city and every type of specialization to find all the consultations possible. Thus, we will apply a filter on the cities, the type of specialization and if there are any available consultations (if it is not the case we will not take into account the said doctor).
We have no idea of how many data points we will get as for each city and specialization the number of doctors is highly variable. If we had to give a range we would say between 10.000 and 100.000 data points.

We believe that the coverage of the data will not be complete as Doctolib does not identify all of the existing doctors. However, by looking at Doctolib's statistics we found out that this lack of doctors is not statistically significant compared to the ones present on the website as almost every doctor is present in Doctolib.

3 - Describe your first ideas for a possible outline for the analysis

Lucky we are, we don't expect to have to pre-process the data because we will take only the variables that we want. We will sort the data upstream to keep only what we need about the appointment. In case of missing or null values we will eliminate this data so as not to distort our results.

Once we collect data, we would like to establish a potential link between waiting time, location, medical specialization and copayments. We expect to observe differences of waiting time and price depending on the location, and differences of price depending on the waiting time.

First of all, making simple graphs with the data collected to easily compare the waiting time between different cities or different specializations. But the most important task will be to create a map of France, a dynamic one. Use color to highlight the fact that it is easy to have an appointment or not, a color gradient that covers the map in function of time that you need to wait to have a consultation. The filter will be the specialization and/or the fact that it is practiced copayments. It could be also possible to zoom on the map and to have more precise details about the different cities.

We could use regressions to see if our models are right, for instance if we suppose that the richer one is, the faster he will find a doctor, then with a regression we could prove if it is either right or wrong. We could also build a model to predict the waiting time for a certain type of specialization, etc …

We believe that one of the conclusions we will find is that there is a negative correlation between a doctor's consultation price and its waiting time. Indeed, if a doctor is able to have high fees it could be because he knows that lots of people want his expertise and thus it will be harder to have an appointment with him. And, there could be huge differences between the waiting time if you live in a town or in the countryside because of the number of doctors available. But, even in town, we could put forward differences between the doctors that resort to copayments and those who do not.