ML Project Proposal

Alvaro Guijarro May
226883@students.hertie-school.org

Niklas Pawelzik
221791@students.hertie-school.org

Justus v. Samson-Himmelstjerna
222918@students.hertie-school.org

1. Introduction[1]

The "Europe Abortion Access Project", a coordinated study between the University of Barcelona and the European Research council focused on cross country abortion travel examination, concludes that the purpose of seeking abortion care abroad is a "phenomenon [that] remains poorly understood".[1] The researchers argue that the lack of scientific insight in this aspect of European abortion regimes is linked to a general lack of sufficient "quantitative and qualitative data" on the matter. Our research aims to contribute to a closure of this gap by investigating data on abortion numbers on the state-level within Germany. We believe that based on this data, we can contribute to finding answers regarding cross country abortion travel. Our goal is to develop a classification model that allows predictions on abortions performed on foreigners in comparison to abortions performed on German citizens. We expect a number of factors to potentially play a role when it comes to increase the share of abortions on foreigners:

1. An important factor is the share of foreigners in the overall population in a given state. There is a straight-forward reasoning behind this expectation: Foreign residents are expected to seek abortions, a higher share of foreigners can be expected to result in a higher share of abortions performed on foreigners.

2. However, we also expect a less direct positive effect of higher foreign populations to play a role. Women from abroad that consider travelling to Germany for an abortion might be inclined to do so in an area where they have a social network such as family members or friends (for emotional support, for logistical/language support, for shelter). A bigger foreign community in a state could therefore be a pull factor for women to travel from abroad to this particular state in order to seek abortion care. This effect can be expected to be stronger for countries of origin that have a more restrictive abortion policy in place than Germany.

3. In addition to a social network in a state, proximity to the country of origin might be a pull factor for women to travel to one state instead of another. Shorter (and possibly cheaper) traveling might be a reason to seek a therapy in a closer state. Again, this effect can be expected to be stronger for countries of origin that have a more restrictive abortion policy in place than Germany. In other words: states within Germany that are closer to borders with neighboring countries that have stricter abortion laws can be expected to have higher shares of foreigners seeking abortion treatment.

We intend to cover these effects on our dependent variable - the share of abortions performed on foreigners on overall population per state in proportion to the share of foreign population on overall population per state - in our model by combining a number of different data sets, as explained in greater detail further below. We hope to train our model based on a sample of roughly 100.000 abortions performed in Germany in the year 2021. Given the size of our sample and the nature of our research interest, we are confident that we can build a SGD classifier model that distinguishes between "foreign" and "domestic" abortions.

Dependent Variable:

- Share of Abortions performed on foreigners on overall population per state in proportion to: Share of foreign population on overall population per state.

Independent Variables:

- Proximity of neighboring countries (geographical distance).

  - Restrictiveness of respective countries (using abortion atlas scores).

---

[1]Here's a link to Alvaro's GitHub account https://github.com/Alvaroguijarro97/ML_Group_Project, Justus's https://github.com/jvsamson and Niklas's https://github.com/nikpaw

- Proximity of neighboring countries (geographical distance).
  - Restrictiveness of respective countries (using abortion atlas scores).
- Foreign population per state (share of overall population).
  - Interesting for network effects (foreigners going for their abortion to states where they have relatives/social network)
  - If feasible/possible: Control for share of female between 12-50 years
- Population with a migrant background per state.

Research question: Is there an over-proportionally big share of abortions performed on foreigners in some states?

Classification approach: Is a given abortion (based on the input data/independent variables) predicted to be classified by our model to be performed on a foreigner or a German?

Binary classification: foreigner/German abortion.

Model used: SGD Classifier (TBA)

2. Motivation

2.1. Societal Relevance:

Europe's health systems are embedded in their respective social welfare landscapes. However, this also means that health systems in the EU are diverse and reflect different social choices. On the other hand, health systems in the EU are increasingly interacting with each other. Especially since the Directive on Patients' Rights in Cross-Border Healthcare.[2] For patients in Europe, this means that the costs for treatment abroad are reimbursed by the respective health insurance fund up to the amount that would have been incurred for the corresponding treatment at home.

In keeping with this vision, we believe that our project could contribute to improving access to safe and legal abortion care throughout Europe. Cross-border patient movements are a fact of our time. With our model, countries could better prepare themselves for this fact. Social actors could anticipat and estimat effects of stricter policies in neighboring countries on their own healthcare system. Governments have a responsibility to safeguard the right of women and girls to lead free and safe reproductive lives without discrimination and coercion.

2.2. Research relevance:

As already mentioned in the Europe Abortion Access Project, it writes on its research website that the phenomenon of transnational abortion travel remains poorly understood. They write that it is known that many European women travel for abortions outside and/or within the borders of their country of residence, even from countries with relatively liberal abortion laws.[1]

2.3. Learning perspective:

In terms of our own academic progress, our motivation is that we hope to get a sense of how well data sources from the European multi-level governance system (in this case: state, country, EU) can be combined in order to gain insights. Furthermore, our team shares an interest in health policy. This project thus allows us to take the first step towards applying our newly learned machine learning skills to policy analysis in this area.

We are looking forward to working with large data sets (about 100,000 abortions in Germany per year). Here we are particularly interested in combining different data sets, as well as integrating different types of data sources, in order to derive new information in our combined field of interest.

3. Evaluation

Our Research model aims to shine a light on the question if the proximity of German states to bordering nations with a more restrictive stance on abortion leads to a higher percentage of abortions performed on foreigners in these states. Therefore, our classification approach would be successful if we can predict if a given abortion, based on the input data/independent variables, was performed on a foreigner or a German.

For classification predictions, there are four possible types of outcomes: true positives, true negatives, false positives, false negatives. With this outcomes we can evaluate our classification model by its accuracy, precision, and recall [3]. The most important evaluation metrics are precision (true positives / all positives) and recall (true positives / (false negatives + true positives)). For our initial research question, it will be more important to obtain a higher precision score, since we will be looking to develop a ML model that can accurately predict if an abortion procedure realized on German territory was applied to a German citizen or a foreigner.

4. Resources

4.1. Data Sets

1. **The European Abortion Policies Atlas**
   An in-depth analysis of abortion policies across Europe, which scores 53 European countries and territories in accordance with their legal frameworks in reference to their access to safe abortion care. The Atlas is an initiative which is powered by the European Parliamentary Forum for Sexual and Reproductive Rights

(EPF) and International Planned Parenthood Federation European Network (IPPF EN).

The questions and structures for the Atlas were designed by a group of experts in sexual and reproductive health and rights. They came up with an overall score on a scale from 0-100 composed of four sections with several sub-categories: "access", "legal status of abortion care", "clinical care and service delivery", "information and on-line information".[2]

2. **Abortion Statistics for Germany ID: 23311-0006**
The data set concerning abortions of the Federal Statistical Office contains information at the level of the Federal Länder on abortions performed annually, subdivided into the individual quarters. Furthermore, information on the origin of the patient is listed for the individual procedures, whereby a distinction is made between the various federal states and the category "foreigners". The dataset contains information for all 16 federal states with an annual number of about 100,000 abortions performed in Germany from 1996 up to 2021.[3]

3. **Foreigner statistics ID: 12521-0022**
The statistics on foreigners of the Federal Statistical Office contain different representations of the regional distribution of the foreign population. Among other things, they list the federal states, sex, years of age and country groupings/citizenship. With the help of this database we hope to identify driving factors for higher shares of abortions performed on foreigners, checking for variables such as composition of foreign population or proximity to neighboring countries with stricter abortion policies.[4]

4. **World Cities Database**
With the help of this database, which contains information like City Name, Longitude, Latitude, Country, State, Region, City Status, Population Density, Population, and Id of more than 4 million unique cities around de world, we will create our own database with the information needed for the European Continent. With this database we expect to calculate the shortest distance from all of the German State's Capitals (or their biggest city) with the closest city from their neighboring European countries, in order to determine the distance between said states and countries.[5]

---

[2] https://www.epfweb.org/node/857
[3] https://www-genesis.destatis.de/genesis/
/online?operation=table&code=23311-0006&
bypass=true&levelindex=1&levelid=1664961110565#
abreadcrumb
[4] https://www-genesis.destatis.de/genesis/
online#astructure
[5] https://simplemaps.com/data/world-cities

4.2. Computational tools

GitHub will be used for version control and collaboration, and coding will be performed in Jupyter Notebooks and Visual Studio Code.

5. Contributions

Since all group members have little to no experience when it comes to bigger coding projects, we have a big interest to involve each team member in the foundational steps of the data operation. Given that we combine different data sets as outlined above, we will all be able to for instance clean a data set individually. Given Alvaro´s experience with Machine Learning, he will thereby focus on the World Cities Database in order to wrangle it according to our needs. The decision for a SGD Classifier model resulted from a group discussion. Since we consider the implementation of this model to be one of the core pieces of our machine learning project, we intent to work on as many steps of its developement as possible collaboratively. As Niklas has previously worked on a research paper that touched on a related topic, his input was particualry valuable during the preparatory work phase when it comes to identifying potential research interests and data sources that allow to tackle the questions we developed (such as the European abortion atlas, or the respective data sets from the Federal Statistical Bureau). Justus will contribute especially when it comes to visualization and communication of our research results, for instance via maps showing the different effects across the different German states. Given our little experience with this kind of project, this is only a preliminary frame we will adapt accordingly, once we can better estimate the workload of each step in the work flow.

References

[1] Europe abortion access project, 2020.
[2] R. d. E. U. Europäisches Parlament. Richtlinie 2011/24/eu des europäischen parlaments und des rates vom 9. märz 2011 über die ausübung der patientenrechte in der grenzüberschreitenden gesundheitsversorgung. 2011.
[3] J. Jordan. Evaluating a machine learning model., July 2017.