Project proposal

A. TITLE

- "Coping with Covid: Signs for Success"

B. ELEVATOR PITCH

- We propose to analyse covid data for different countries as well as demographic and social data. What are the "signs of trouble" and "signs for success" in a country? We believe that we may find patterns in the geodemographic data that can serve as indicators/markers of how well or how poorly a country is capable of handling a crisis such as Covid-19. This would enable us to, in the case of future virus outbreaks, possibly predict which countries are going to get most affected by the virus.
- There are primary and secondary target groups for our mini-project:
    - first and foremost it would benefit global decision-makers who need to plan national strategies or can influence the national strategies of different countries.
    - Secondly, it benefits the general public since it behoves us all to be well informed on matters that directly influence us all.
- There is sadly a need for this kind of study as pandemics tend to be a cyclic type of event (the last big one in the Western World having been the Spanish flu). We need all the take-away lessons that we can get for better and faster future decisions and improving the "signs for success" in the different countries.

C. DATA

- **Data Sources:**
    - Collected COVID19 data: https://coronavirus.jhu.edu/about/how-to-use-our-data?fbclid=IwAR34LrKBj1-Ic4kT9F89Gu6r1oOQdBEriqa7PYlSPGVCcWKizH_Ezk90OMo
    - Collected Geodemographic and Social data: https://unstats.un.org/unsd/demographic-social/products/dyb/index.cshtml/?fbclid=IwAR277ZK_PLqKARFt8y9o-mDeoknwnq44lFJcbiPPVzQhUgau7R8Uqt7uYBo#censusdatasets
- Which is the data management plan?

    Data Collection

    - Collecting pre-existing open source and publicly available data (mainly csv format) from the Internet from the sources mentioned above.

    Documentation and Metadata

- Our presentation will contain links to the data sources which contain the descriptions of the data and pertaining metadata.

  Data Sharing

- We will be making a blog post with our findings and data visualizations.

- Data cleaning/wrangling involved? – Yes. For example, there are several spaces with NA and similar which will have to be replaced with values. We will also not use all of the data but only selected portions of it.

D. DATA ANALYSIS

- What will we try to learn?
  - What are the "signs of trouble" and "signs for success" in a country? We believe that we may find patterns in the geodemographic data that can serve as indicators/markers of how well or how poorly a country is capable of handling a crisis such as Covid-19.
- Which variables are we more likely to use?
  - Country, gender, infection numbers, vaccination numbers…
- Which (machine/statistical) learning methods might be more beneficial and why?
  - Exploratory data analysis – to have a good look at the data before making any assumptions.
  - ML: supervised learning as it will enable us to make predictions (on vaccination rates, etc).

E. COMMUNICATION OF RESULTS

- Online page with graphics and explanations of our findings and a nice visual presentation using the software from gapminder.org .

F. OPERATIONALIZATION

- There is sadly a need for this kind of study as pandemics tend to be a cyclic type of event (the last big one in the Western World having been the Spanish flu). We need all the take-away lessons that we can get for better and faster future decisions and improving the "signs for success" in the different countries.