

Integrating Health Data for Sustainable Development

Data Integration Project
Step 2: Integrating



Huthayfa Mutan

SUSTAINABLE DEVELOPMENT GOALS





Experiments. Chosen Datasets:

Out of hundreds of overwhelming datasets about multiple of large sets of indicators, the choosing data sets are the following:

Global Health Spending 1995-2019

<https://ghdx.healthdata.org/record/ihme-data/global-health-spending-1995-2019>

Global Burden of Disease Study 2019 (GBD 2019) Healthcare Access and Quality Index 1990-2019

<https://ghdx.healthdata.org/record/ihme-data/gbd-2019-healthcare-access-and-quality-1990-2019>

Reasons of choosing:

- 1- Contains multiple valuable indicators over a wide range of countries.
- 2- Well documented datasets.
- 3- Valuable data for the goal of the project.



Pipelining:

Since the Healthcare Access and Quality Index dataset have similar values for multiple rows, such as: country name, indicator name, the first step is to convert these into a format easier to integrate which will be in the shape of:

(Country,Year) -> Set of indicators



Algorithms:

String matching to match country names and codes and using the documentation of each dataset to match correctly.

Date Matching to match the dates.

ER diagram

