Francesco Saverio Sconocchia Pisoni, 1889241

Project:

[DW] Select a set of data sources, integrate them by means of ETL operations (you can either use an integration tool or write

simple scripts to solve common ETL problems), define the DFM model for your data, define and populate the corresponding

star schema or snowflake schema, by using a relational DB (e.g., Postgres)

Datasource on pollution: https://ourworldindata.org/grapher/average-exposure-pm25-pollution

Datasource on deaths, cause Cancer https://vizhub.healthdata.org/gbd-results/

Datasource on health expenditures / GDP per country https://databank.worldbank.org/source/world-development-indicators

Datasource on location global positions https://datahub.io/core/country-codes

Datasource on iso3 codes https://www.kaggle.com/datasets/wbdill/country-aliaseslist-of-alternative-country-names

The project consists on integrating the three datasets through ETL operations, integrating them per

country and year, selecting the minimum common range of years between the datasets and selecting

the minimum common subset of countries, recognizing when the same country is named in different

ways through an ETL tool.

The DFM schema focuses on the deaths (cause Cancer) with respect to the GDP and the PM2.5

pollution level per year and country