

Data Management Project Proposal

Group Members:

- Samuele Cervo 1883147
- Agnese M. Capparelli 1794326

Type of project: Data Warehouse

Description of the dataset:

- <https://github.com/pcm-dpc/COVID-19>
- <https://www.kaggle.com/datasets/arthurio/italian-vaccination>
- <https://github.com/italia/covid19-opendata-vaccini>

Brief description:

This project studies the connection between the COVID-19 pandemic and the vaccination campaign in different regions of Italy. We will use three datasets: daily COVID-19 data from the Italian Civil Protection Department, vaccination data from the Italian government's open data, and population data to adjust values based on how many people live in each region. We will ETL (Extract, Transform, Load) steps to clean and combine the data, using the date and region as common links. Then, we will create a data model and built a star schema in PostgreSQL to help with analysis. The final data warehouse lets us explore how vaccines affected the spread of COVID-19 in each region over time.