

## E-SouP

Github Repository: [GitHub - bengiada/CS306-GroupProject-E-SoUp](https://github.com/bengiada/CS306-GroupProject-E-SoUp)

Group Members: Bengi Ada Tüfekci 30720

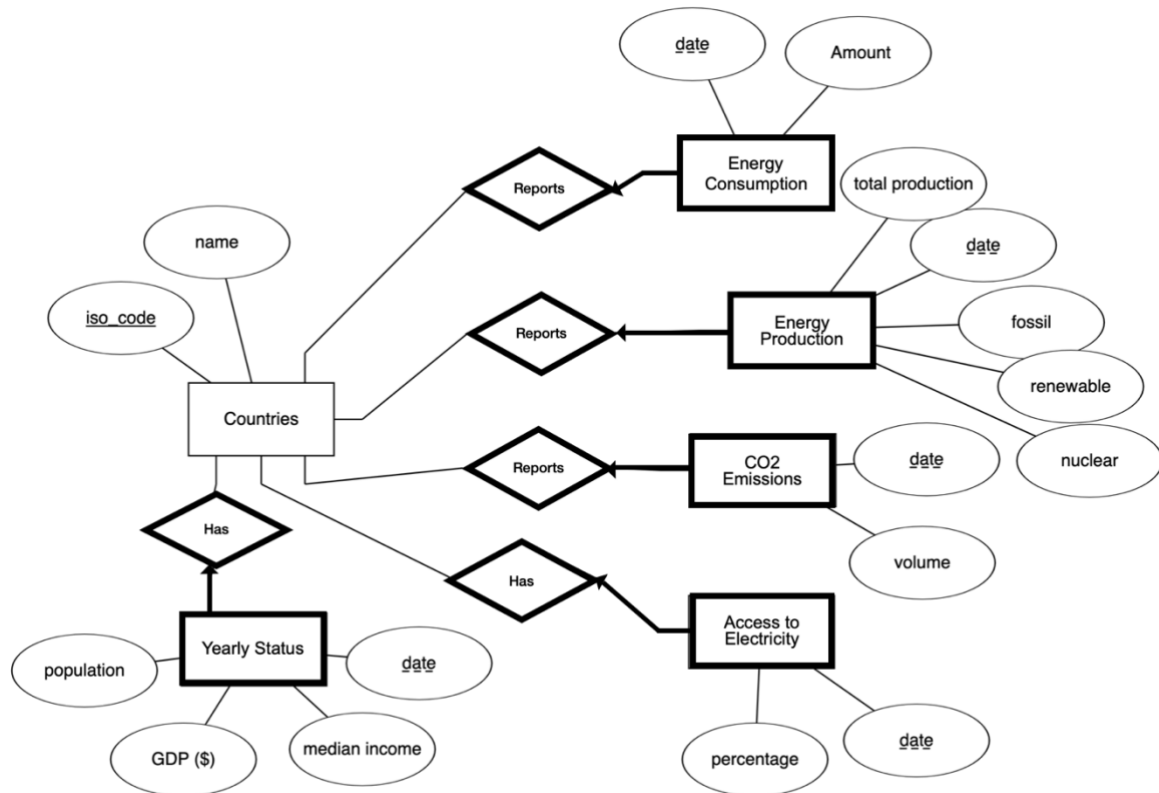
Can İnanır 31159

Doğa Turak 27975

Eren Bilen 25485

Zeynep Merve Yaman 28876.

We are going to make a database containing countries and their yearly energy trends. Our dataset will include 6 entities; Countries, Yearly Status, Energy Consumption, Energy Production, CO2 Emissions, Access to Electricity. Our database will store these entities and their attributes as data and will have relations for the sake of seeing the correlation. To describe the characteristics of these entities we will have basic attributes and a key attribute unique to each data. For instance, iso\_code is the key attribute for Countries since every country has a distinct iso\_code. Basically these entities are the most important components that a country should keep track of for identifying how the change in yearly status relates to the energy production sources, consumption and accessibility. With this database we are aiming to form several relations between these entities that a country can benefit from. Some of these relations are: countries' yearly status based on its population, GDP, median income and countries reported energy consumption, countries' reported energy production, countries' CO2 emissions, countries' access to electricity. We plan to name our Project E-SoUp, as in energy sources Project.



In our CSV files, we utilized seven different datasets. We merged data from GDP per capita, GDP total, and population data files to create the "yearly status" table. Each of the four datasets for access to electricity, annual CO2 emissions, consumption of energy per capita, and sources of electricity production had their own separate table.

Most of the datasets we used had additional parameters that they tracked. For instance, all four electricity datasets included data for GDP and population for their given year, which was redundant since we already had separate datasets for GDP and population information. Therefore, most CSV files were stripped down to only three crucial variables: country code, year, and one to three rows of relevant data. We additionally filtered all datasets to exclude any data that:

- did not have a country code or date (as these are our key constraints);
- expanded past 1985 (as most of our data only extends as far as approximately 1990);
- was continent or world data as these are not compatible with the system we have in mind.

Apart from these filters, any extra data from the base tables that was not relevant to our research was removed.