Outdoor Air Pollution

Members: Asmaa Bashir (30401) - Kemal Ayhan (22511) - Mehmet Emin Er (27748) - Şükrü Baktır (23730) - Atahan Bozkuş (28471)

Repository link: https://github.com/baktirsukru/CS306Project

Pollution is a serious problem that has significant health and environmental impacts. Our database focuses on data relevant to outdoor air pollution. We used the datasets available on https://ourworldindata.org/outdoor-air-pollution to analyze and quantify the severity of the issue. Each dataset provides data on a country level. We have selected the following tables to approach the issue:

1. CAUSES_OF_DEATH: This dataset compares between various risk factors and the number of deaths they caused in 2019. Outdoor air pollution is one of the major causes of death globally. While cleaning the CSV file, columns for other minor risk factors were deleted, in order to focus on 5 major ones. The data was filtered so it only displayed values from 2019. Country codes were deleted to avoid duplication.

2. CONCENTRATIONS_OF_AIR_POLLUTION: This dataset has the information of air pollution concentration by years. Comparison of air pollution by year is useful for death rates of people and other datasets. While cleaning the CSV file, duplicates and empty blocks have been removed from the dataset.

3. PERCENTAGE_OF_DEATHS_BY_AIR_POLLUTION: This dataset contains the information about the percentage of deaths caused by outdoor air pollution. It can help us to compare outdoor air pollution with other causes of death. Country codes are deleted from this dataset to remove the duplicates since they are stored in another table.

4. OUTDOOR_POLLUTION_RATES_BY_AGE: This dataset compares the death rates of different age groups attributed with outdoor air pollution per 100.000 individuals. The country codes are deleted to get a cleaner dataset.

5. OZONE/PARTICULATE_MATTER_POPULATİON: This dataset has information on mortality rates from ozone and particulate matter pollution by year. Comparing ozone and particulate matter pollution by year is useful for human mortality and other datasets. The dataset was cleaned by removing duplicates and empty blocks from the CSV file.

6. COUNTRIES: This dataset lists countries and country codes. For regions without a country code, empty cells were filled with NULL.

First of all we created our entity sets for the ER model with the taken datasets from the website Our World in Data. We determined the attributes of our entity sets. Each group member has an entity set and we have determined the relations between different sets. Lastly we determined participation constraints, weak entities and finalized the ER model.