Rimsha Aziz and Shreya Sachdev

05/04/2020

**ETL Project Report**

***Sources***

[**FAOSTAT**](http://www.fao.org/faostat/en/#data/GN)

Energy emissions - CO2 and CO2 equivalents

By country

1970-2012

CSV file

Cleaning:

* Drop code columns ( Year Code, Element Code, Domain Code) - Drop
* Aggregate regions (China, Ethiopia) - used .replace function
* Drop currently non-existent countries (USSR,Yugoslav)
* Change Column headers to omit special characters and replace with underscores

[**Energy**](https://ourworldindata.org/energy)

Energy consumption by source and region

By country

1965-2018

CSV file

Cleaning:

* Drop currently non existent countries (USSR)
* Drop total world values
* Renaming “Entity” column to country so we can join
* Drop country code column

***Transformations***

Joining two tables on countries in SQL using pgAdmin to examine possible correlations between CO2 emissions and sources of energy consumption between countries from 1965 - 2018 which can be filtered through to select a specific year.

***Load***

Final database is relational and has two tables. One table has data from CO2 emissions and CO2 equivalents for countries and the other table has sources of energy consumption by countries. These two tables were chosen because when joining them on countries, the database can show both the emissions created by that country as well as which energy source is being consumed the most. This could provide insight into a cause for high CO2 emissions levels and a lack of use in renewable sources. Having two separate tables is beneficial because it still allows us to see a larger scope. For example, if there were any countries that existed in one table but not the other.

***Challenges***

We were able to load the data into SQL, however we had challenges performing queries on the tables such as joining. The data would be loaded into the tables however, the schema would change from the originally outlined schema. This happened because all the table headers, when imported, were strings with quotation marks around them. To accommodate for this issue and others when importing, including special characters. We had to rename all the column headers to lowercase and remove all special characters.