
Project 1 : Exploring Renewable Energy Evolution Between 1965 and 2019 Using SQL

Introduction

This project aims to gather various data related to the development of renewable energy worldwide from 1965 to 2019. It seeks to address the following questions:

- Is the progression of renewable energy development similar across different regions of the world in terms of development speed and types of renewable energy used?
- Which entities are the largest users of renewable energy in 2019?
- How can we explain the differences in renewable energy use between countries? Is there a link between a country's economic development stage and its use of renewable energy? Is there a link between the number of renewable energy patents in a country and its use of renewable energy?

Data Sources

The data sources for this project include multiple tables containing information on renewable energy use, economic indicators, and patent filings related to renewable energies across different countries and years. These tables provide detailed metrics on the share of various types of renewable energy (solar, wind, hydro), economic growth (GDP for specific years), and innovation indicators (renewable energy patents).

Subject	Web site	Link
Renewable energies between 1965 and 2021	Kaggle	https://www.kaggle.com/code/mehmetisik/01-the-global-portrait-of-renewable-energy/input
GDP between 1960 and 2020	Kaggle	https://www.kaggle.com/datasets/zackerym/gdp-annual-growth-for-each-country-1960-2020?resource=download&select=GDP_annual_growth.csv
Patents for renewable energies	Our world in data	https://ourworldindata.org/grapher/patents-for-renewables-by-country

Database Structure

The following initial tables, contained in csv files, were imported on SQL :

- **share_renewables**: This table consists of the columns *Entity*, *code*, *year*, and *share_of_renewables*. The *share_of_renewables* column represents the percentage of equivalent primary energy from renewable sources for each country or region in a given year. This measure aims to indicate the amount of fossil fuels needed to produce the same amount of energy generated by renewable sources, accounting for heat conversion losses during fossil fuel transformation.
- **solar_energy**: This table consists of the columns *entity*, *code*, *year*, and *share_of_solar* (as a percentage of equivalent primary energy).
- **wind_energy**: This table consists of the columns *entity*, *code*, *year*, and *share_of_wind* (as a percentage of equivalent primary energy).
- **hydro_energy**: This table consists of the columns *entity*, *code*, *year*, and *share_of_hydro* (as a percentage of equivalent primary energy).
- **gdp_1965**: This table consists of the columns *entity*, *code*, *Indicator_Name*, *Indicator_Code*, and *gdp_in_1965* (in USD).
- **gdp_2019**: This table consists of the columns *entity*, *code*, *Indicator_Name*, *Indicator_Code*, and *gdp_in_2019* (in USD).
- **patents_renewable_energies**: This table consists of the columns *entity*, *code*, *year*, and *total_patents*. The *total_patents* column refers to the total number of patents filed in the renewable energy sector in 2019 for each country.

Data Preparation

Data Format Conversion: Some data were initially in text format and were converted to numeric format to enable calculations in SQL. For instance, GDP values were in string format, making it impossible to calculate GDP growth from 1965 to 2019.

Creation of IDs: To join different tables, an ID column was created in each table. This ID is composed of two elements: the country and the observation date.

SQL Script presentation

- **Importing Initial Tables**: The initial tables include *share_renewables*, *solar_energy*, *wind_energy*, *hydro_energy*, *gdp_1965*, *gdp_2019*, and *patents_renewable_energies*.
- **Creation of the *all_renewables_table***: This table aggregates all variables related to the share of renewable energy consumed by each country from 1965 to 2021. It includes columns for

share_of_renewables, *share_of_solar*, *share_of_wind*, and *share_of_hydro*. A new column, *most_used_renewable*, was created to show which renewable energy type is most used in each country.

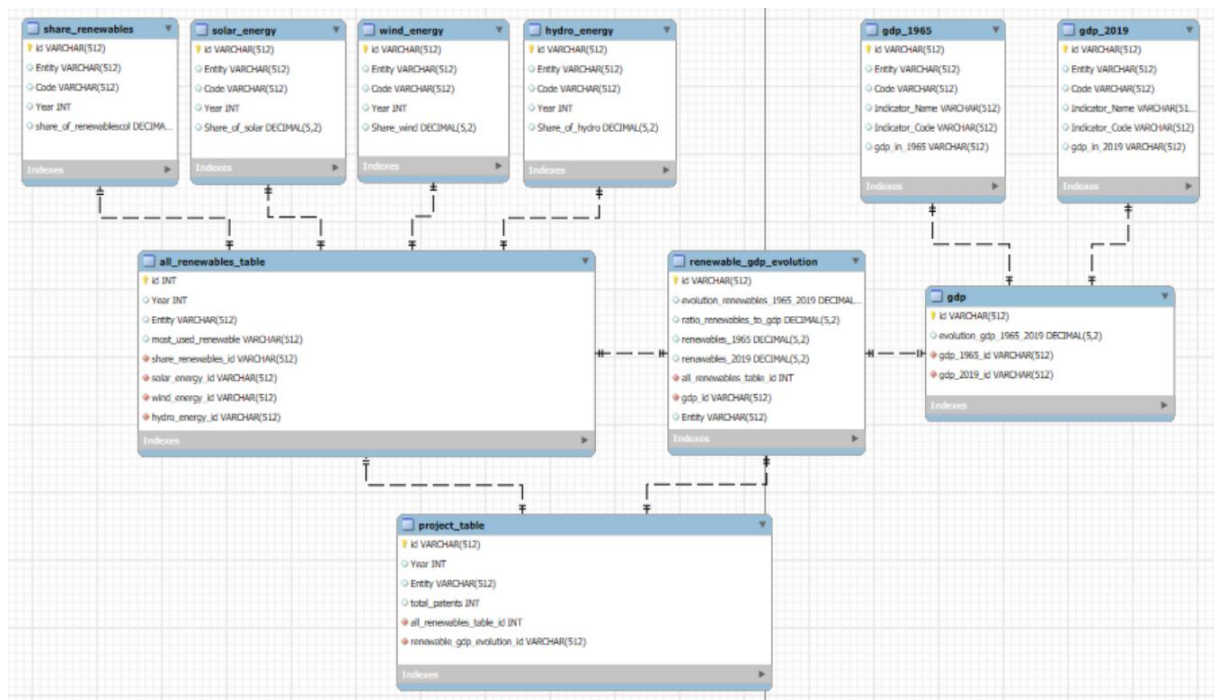
- **Creation of the *gdp* table:** This table calculates GDP growth for each country from 1965 to 2019. It contains columns for *gdp_in_1965*, *gdp_in_2019*, and *evolution_gdp_1965_2019*. The last column shows the percentage change in GDP from 1965 to 2019 for each country.
- **Creation of the *renewables_gdp_evolution* table:** This table includes columns for *Entity*, *renewable_1965*, *renewable_2019*, *evolution_renewables_1965_2019*, *evolution_gdp_1965_2019* and *ratio_renewables_to_gdp*.

Two new columns were added: *evolution_renewables_1965_2019* to calculate the growth in renewable energy from 1965 to 2019 for each country, and *ratio_renewables_to_gdp* to assess each country's commitment to transitioning to a greener economy. If the ratio is less than 1, renewable energy is growing more slowly than GDP, indicating continued economic reliance on fossil fuels.

- **Creation of the *project* table:** This table consolidates all columns needed to answer the project's research questions. It includes columns for *share_of_renewables*, *share_of_solar*, *share_of_wind*, *share_of_hydro*, *most_used_renewable*, *evolution_renewables_1965_2019*, *evolution_gdp_1965_2019*, *ratio_renewables_to_gdp*, and *total_patents*.
- **Row Ordering:**
 - The row for "world" appears in the first position.
 - Other rows are sorted in descending order based on the share of equivalent primary energy from renewable sources.

Relationship Schema

The following diagram was created in SQL to highlight the relationships between the different imported and created tables.



Script Usage

This project consists of several SQL files. The main file, named *Assignment_1*, contains code to create the various tables. Files such as *renewable_energy*, *solar_energy*, *wind_energy*, *hydro_energy*, *gdp_1965*, *gdp_2019*, and *patents_renewable_energy* populate the initial tables.

It is essential to execute these SQL files after creating the initial tables in *Assignment_1*. Detailed instructions are included in the code comments.

Precautions and Limitations

- **Missing or Incomplete Data:** Some countries may have missing data. Rows containing missing values ('Null'), such as for GDP, have been kept because the primary focus of the analysis is on the percentage of primary energy equivalent coming from renewable sources.
- **Analysis Period:** Historical data, especially for years before 2000, may be less reliable or less frequently updated for some countries, particularly developing nations.
- **Causation vs. Correlation:** A strong correlation between renewable energy growth and economic growth does not imply direct causation. Both variables may be influenced by other factors like public policies or external shocks (e.g., energy crises).
- **Renewable Energy Share:** Even if renewable energy use increases, its share in the energy mix may remain low if overall energy consumption also rises, limiting the real impact on energy sustainability.