Data Report

Neo Reinmann

June 2024

1 Research Question

This project aims to address the following research question:

To what extent does weather influence energy consumption, and does an increase in extreme weather events affect energy consumption patterns?

To investigate this, the electric energy production data from Germany will be compared with weather data from selected regions within Germany. The analysis covers the period from January 1, 2015, at 00:00 to December 31, 2023, at 23:00.

2 Data Sources

2.1 Electric Energy Production

The data on German electric energy production were sourced from **smard.de**, a platform managed by the **Bundesnetzagentur**. The dataset includes hourly electric energy production figures categorized by production means, such as solar power and coal power. Table 1 details the data categories.

The **Bundesnetzagentur** is a government-run organization, ensuring that the data is reliable and of high quality. The data is provided under the CC BY 4.0 license, permitting its use, sharing, and modification. The licensing information is available on the data source website, promoting open data usage.

2.2 Weather Data

Weather data was obtained from the **Deutscher Wetterdienst (DWD)**, which operates under the **Federal Ministry for Digital and Transport** of the German government. This data is highly trustworthy. The data is openly accessible, with individual tables available for each weather station. From a

Date From	Date To	Biomass [MWh]	Hydropower [MWh]	Offshore Wind [MWh]	Onshore Wind [MWh]	Photovoltaics [MWh]	Other Renewables [MWh]
Nuclear [MWh]	Lignite [MWh]	Hard Coal [MWh]	Natural Gas [MWh]	Pumped Storage [MWh]	Other Conventional [MWh]		

Table 1: Structure of the electric energy production data.

Station ID	Start Date	End Date	Altitude [m]	Latitude	Longitude	Station Name	State
04931	19880101	20240521	371	48.6883	9.2235	Stuttgart-Echterdingen	Baden-Württemberg
03668	19510101	20240521	314	49.5030	11.0549	Nürnberg	Bayern
03379	19970701	20240521	515	48.1632	11.5429	München-Stadt	Bayern
01420	19810101	20240521	100	50.0259	8.5213	Frankfurt/Main	Hessen
01078	19760301	20240521	37	51.2960	6.7686	Düsseldorf	Nordrhein-Westfalen
02014	19490101	20240521	55	52.4644	9.6779	Hannover	Niedersachsen
01228	20020101	20240521	0	54.1651	6.3460	UFS TW Ems	Hamburg
00433	19510101	20240521	48	52.4676	13.4020	Berlin-Tempelhof	Berlin

Table 2: Selected weather stations with their coordinates.

Station ID	Date	QN_9	Temperature [°C]	Relative Humidity	eor

Table 3: Column structure of each weather station's data.

meta table containing information on all weather station codes, eight stations were selected. These stations are listed in Table 2.

Each station's data table includes several columns, as shown in Table 3. For this study, the columns *MESS_DATUM* and *TT_TU*, which contain the date and temperature, are of primary interest.

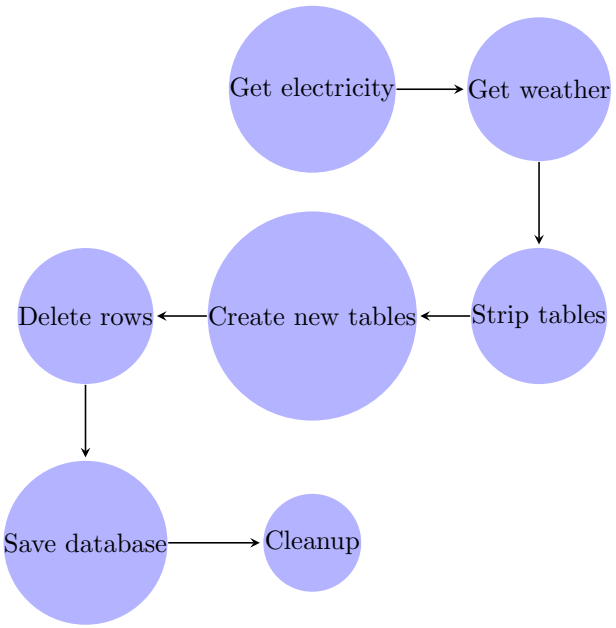
Like the energy production data, the weather data is also under the CC BY 4.0 license, allowing adaptation and sharing. This information is detailed in the *README* file on the **DWD** website.

3 Data Pipeline

The data pipeline involves several steps, illustrated in the following flowchart. Python was used to fetch the data and perform initial adaptations before saving it in *.csv* format. The initial step involves obtaining the data. For **SMARD**, a POST request is required to retrieve the *.csv* file. For the weather data, the station ID must be obtained, followed by downloading a ZIP file containing the data. Despite being in *.txt* format, the file is compatible with *.csv* format.

The transformation of energy production data involves categorizing it into sustainable and conventional energy production, with an additional column for total energy production (the sum of the previous two categories).

For each weather station, only the time and temperature columns are extracted and saved into a *.csv* file.



In the subsequent step, Jayvee is employed as a dedicated data pipeline language. Rows not within the specified period (01.01.2015, 00:00 to 31.12.2023, 23:00) are removed, and each column is validated for correct data types. Finally, all weather station tables and the electric energy production table are stored in an SQLite 3 database.

Any changes to URLs or table formats will require manual updates to the code.

4 Results and Limitations

The output of the pipeline is an SQL database containing nine tables: eight for each weather station and one for energy production. While the energy production data is complete, some weather stations lack hourly temperature measurements. However, the data is highly accurate, consistently formatted, and relatively recent.

With an SQL database as the output, the data is well-organized and ready for further analysis.

Potential limitations include the generality of the data. Since the dataset encompasses the entire energy production of Germany and broadly selected weather stations, localized weather phenomena and their effects on energy production might be obscured by aggregate data.