

Data Visualization Report

Shadi Farzankia, Mina Aminpour, Zoha Karimzadeh

July 2025

1 Introduction

This project investigates electricity consumption and renewable energy generation patterns in Germany using hourly time-series data. We explore how solar and wind energy production, electricity demand, and temperature vary across time and geography.

By integrating energy data from **ENTSO-E** and weather data from the German Meteorological Service (**DWD**), we conduct visual and statistical analyses to understand:

- Temporal and seasonal patterns in renewable production and demand
- Forecast accuracy of electricity loads
- Relationship between temperature and consumption behavior

The complete code is available in the R Markdown file within the project directory.

2 Packages Used

We used the following R packages throughout the project for data wrangling, visualization, and geospatial mapping:

- `library(tidyverse)`: Core wrangling and plotting
- `library(lubridate)`: Parsing and extracting components of timestamps (ymd_hms, year, month, hour)
- `library(slider)`: Rolling window functions
- `library(rnaturalearth)`: Natural Earth data for country borders
- `library(sf)`: Handling and plotting of spatial data
- `library(plotly)`: Interactive visualizations
- `library(viridis)`: Used in Choropleth map

These packages enabled robust preprocessing, time-based feature extraction, rolling averages, geospatial visualizations, and interactive graphics, all of which were integral to our analysis.

3 Data Sources

We used two primary data sources for our analysis, covering the years 2014 to 2020. The first is the ENTSO-E dataset, which provides time series data on electricity consumption across various European countries. The second is the DWD weather dataset, offering hourly temperature records for several German cities. In the sections below, we provide a detailed description of each dataset.

3.1 European Network of Transmission System Operators for Electricity (ENTSO-E)

The ENTSO-E Time Series Dataset is commonly used in electricity market and power systems analysis. It typically contains hourly time series data from the European Network of Transmission System Operators for Electricity (ENTSO-E). This dataset contains the following key data fields:

- **Timestamp (in UTC):** Hourly time resolution for each observation
- **Electricity consumption (load):**
 - Actual load
 - Forecasted load
- **Renewable energy generation:**
 - Solar Generation
 - Wind Generation
- **Total generation:** Sum of all power generation types
- **Country-level coverage:** Data available for multiple European countries (e.g., DE = Germany, FR = France)

Each field is typically labeled using a standard naming convention, such as 'DE_load_actual_entsoe_transparency' or 'FR_solar_generation_actual', where the prefix represents the country.

3.2 Deutscher Wetterdienst (DWD)

The DWD dataset refers to meteorological data provided by the Deutscher Wetterdienst (DWD) — the German National Meteorological Service. It offers a wide range of weather and climate data for Germany, and it's commonly used in energy modeling, climate research, agriculture, hydrology, and environmental monitoring. For this analysis, we used hourly meteorological data from five major German cities — **Berlin, Frankfurt, Munich, Cologne, and Hamburg**—and computed the average of their temperatures to represent the overall temperature for Germany.

This dataset contains the informations such as:

- **timestamp_iso:** ISO 8601 timestamp in local time with hourly resolution. Format: YYYY-MM-DD HH:MM:SS.
- **STATIONS_ID:** Numeric station ID assigned by DWD. 1975 corresponds to a weather station in Hamburg.
- **QN_9:** Quality level of the measurement (DWD code, typically 3 = validated data).
- **TT_TU:** Air temperature at 2 meters above ground in °C.
- **RF_TU:** Relative humidity at 2 meters in %.

The table below shows sample data from the DWD dataset for Berlin.

Table 1: Excerpt from Hourly Weather Data (Station ID: 433)

timestamp_iso	STATIONS_ID	QN_9	TT_TU	RF_TU	timestamp
2014-12-31 23:00:00	433.0	3.0	5.1	98.0	2014-12-31T23:00:00Z
2015-01-01 00:00:00	433.0	3.0	5.2	95.0	2015-01-01T00:00:00Z
2015-01-01 01:00:00	433.0	3.0	5.1	95.0	2015-01-01T01:00:00Z
2015-01-01 02:00:00	433.0	3.0	4.8	94.0	2015-01-01T02:00:00Z
2015-01-01 03:00:00	433.0	3.0	4.8	93.0	2015-01-01T03:00:00Z
2015-01-01 04:00:00	433.0	3.0	4.4	89.0	2015-01-01T04:00:00Z
2015-01-01 05:00:00	433.0	3.0	3.7	91.0	2015-01-01T05:00:00Z
2015-01-01 06:00:00	433.0	3.0	3.5	89.0	2015-01-01T06:00:00Z

4 Data Preprocessing

To ensure the quality, consistency, and readiness of the data for analysis, we applied a comprehensive preprocessing pipeline to both the electricity time series dataset and the weather dataset. This step was crucial for enabling accurate temporal comparisons and robust visualizations. The following preprocessing actions were carried out:

1. Temperature Data Cleaning (Weather Dataset)

- In the raw weather data, missing temperature values were encoded as -999.
- These placeholder values were replaced with the mean temperature for that specific city to preserve daily averages without introducing artificial bias.

2. Electricity Data Cleaning (Load & Generation)

- Several electricity variables (e.g., actual load, solar and wind generation) had missing values.
- We imputed these missing values using the mean of each respective variable, ensuring continuity in time series computations such as rolling averages and totals.

5 Analytical Questions

In the sections below, we address the key analytical questions through a series of data visualizations. Each plot is designed to highlight specific patterns, trends, or relationships within the electricity and weather datasets.

1. How Does Average Electricity Load Vary Across European Countries?

The choropleth map in Figure 1 visualizes the average electricity load (in megawatts) across European countries. The values were aggregated over the full time period and mapped using color intensity:

- Brighter colors (yellow) indicate higher average electricity demand.
- Darker shades (blue/purple) indicate lower average demand.

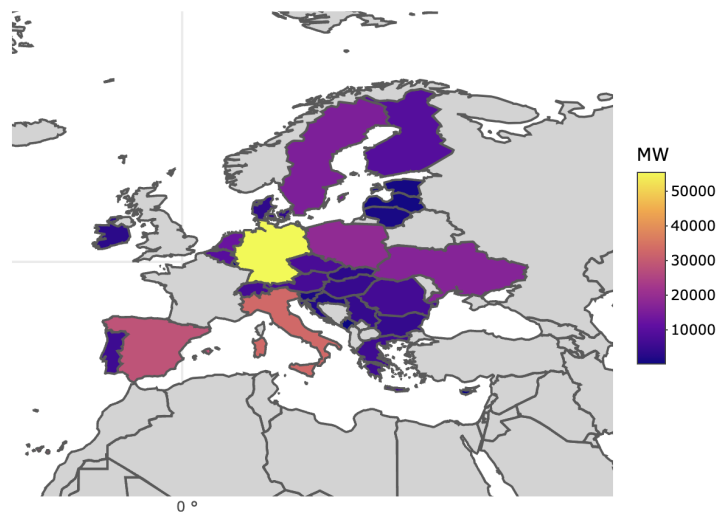


Figure 1: Average Electricity Load by Country

Interpretation

- Germany stands out with the highest average load, exceeding 50,000 MW, shown in bright yellow. This reflects its large population, industrial base, and role as a central energy consumer in Europe.

- France, Italy, and Spain also show relatively high demand (orange/pink tones), indicating their significant energy usage.
- Eastern European and Balkan countries generally show moderate to low average loads (purple/blue), likely due to smaller populations, lower industrial demand, or different energy consumption patterns.
- Some Nordic and Baltic states show lower load levels despite cold climates—possibly due to energy efficiency measures, lower population, or alternative heating methods.
- Overall, western European countries (e.g., Germany, France) have higher average loads, aligning with their economic size and energy-intensive industries.
- In contrast, Central and Eastern Europe shows lower average demand, possibly reflecting different energy infrastructures or economic scales.

Figure 1 provides a continental overview, highlighting Germany’s leading role in electricity consumption — a key motivation for our deeper analysis of German energy flow and renewable integration in the sections that follow.

2. How have solar and wind energy generation evolved annually in Germany and France?

We now shift our focus to Germany and France, comparing their yearly renewable energy production. The chart in Figure 2 illustrates the total annual electricity generation (in megawatts, MW) from solar and onshore wind sources for Germany and France between 2014 and 2020.

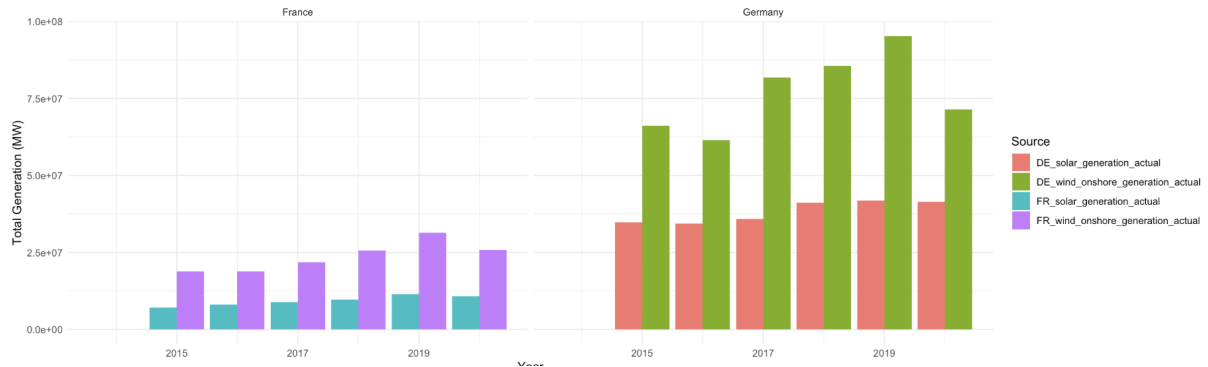


Figure 2: Yearly Renewable Energy Production Trends (Germany & France). Energy generation is significantly higher than France

Interpretation

- This plot demonstrates how Germany leads in renewable energy production in the region.
- Wind energy is the most important contributor in both countries, but its scale in Germany is substantially higher.
- Solar energy contributes less overall but shows consistent year-over-year growth.
- These trends are vital for informing energy policy, infrastructure planning, and climate action strategies.

Germany

- Wind energy is the dominant renewable source, showing a strong upward trend from 2014 to 2020 and peaking in 2019.
- Solar energy displays consistent and moderate growth across all years.
- Both sources slightly declined in 2020, which may reflect changes in weather, capacity, or operational conditions.

France

- Wind energy has steadily increased, though at a lower absolute level than in Germany.
- Solar energy remains modest and shows only minor growth from 2015 to 2020.

3. What Are the Temporal Patterns in Germany's Weekly Solar and Wind Output?

This line chart in Figure 3 displays the 7-day rolling average of solar and onshore wind electricity generation in Germany from 2014 to 2020. By smoothing daily fluctuations, the chart highlights long-term seasonal and interannual patterns in renewable output.

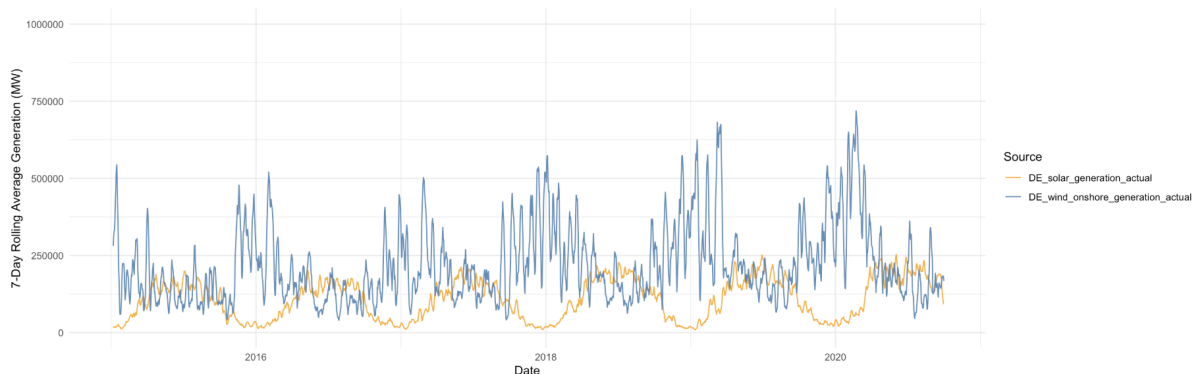


Figure 3: Weekly Solar and Wind Generation in Germany

Interpretation

- The combination of both helps balance renewable supply across seasons, mitigating the intermittency of each individual source.
- Solar is cyclical and stable → reliable for forecasting.
- Wind is erratic but powerful → useful for high-capacity contributions.

Solar Generation

Displays a clear and consistent seasonal cycle:

- Peaks around May to August each year (summer).
- Drops to near-zero values in winter (November to January).
- The pattern is highly predictable due to the natural solar cycle.
- Shows a gradual upward trend over the years, likely reflecting growing installed capacity and improved efficiency.

Wind Generation

- Exhibits high short-term variability with frequent spikes and dips.
- Slight tendency for higher output during late autumn and winter months.
- More irregular compared to solar, likely due to fluctuating wind conditions.
- Overall, wind generation shows growth in variability and output over time, especially from 2018 onwards.
- The increased variability and rising peaks in wind output suggest expanded wind capacity or better harnessing of wind resources in recent years.

4. How does Germany's total renewable energy production compare to its electricity consumption over time?

The plot in Figure 4 shows a stacked area chart of Germany's annual electricity generation from solar and onshore wind, compared against the total electricity consumption, represented by the black dashed line, from 2014 to 2020.

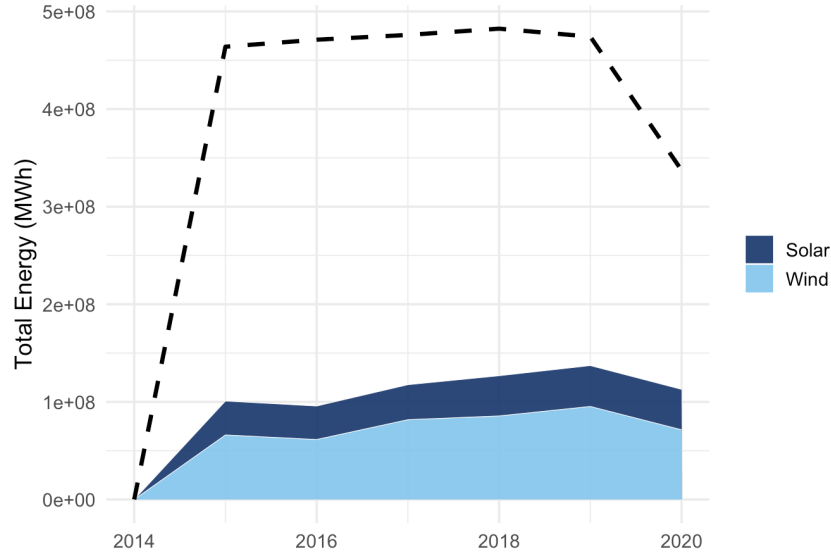


Figure 4: Germany: Renewable Energy Generation vs. Total Consumption

Interpretation

- Wind energy (light blue area) contributes the majority of renewable generation, showing a gradual upward trend from 2015 to 2019.
- Solar energy (dark blue area) also increases steadily, but at a lower absolute scale than wind.
- Total consumption (dashed line) remains relatively flat across most years, with a noticeable decline in 2020, likely related to the COVID-19 pandemic's economic effects.
- Despite growth in renewables, total generation from solar and wind remains significantly below total consumption, leaving a visible gap.

5. How Does Electricity Demand in Germany Vary by Hour of the Day and Month of the Year?

5.1 Hour of the Day

The boxplot in Figure 5 illustrates the distribution of electricity demand (in megawatts) for each hour of the day in Germany, aggregated across all days in the dataset. Each box represents the range and central tendency of load values for a specific hour.

Interpretation

- Demand rises sharply starting around 5:00 AM, reflecting the beginning of daily activities.
- Load reaches its peak between 8:00 AM and 2:00 PM, consistent with typical business and industrial operations.
- After 2:00 PM, demand gradually declines, with a steady drop from 6:00 PM onward.
- The lowest electricity demand consistently occurs between midnight and 5:00 AM, corresponding to nighttime inactivity.

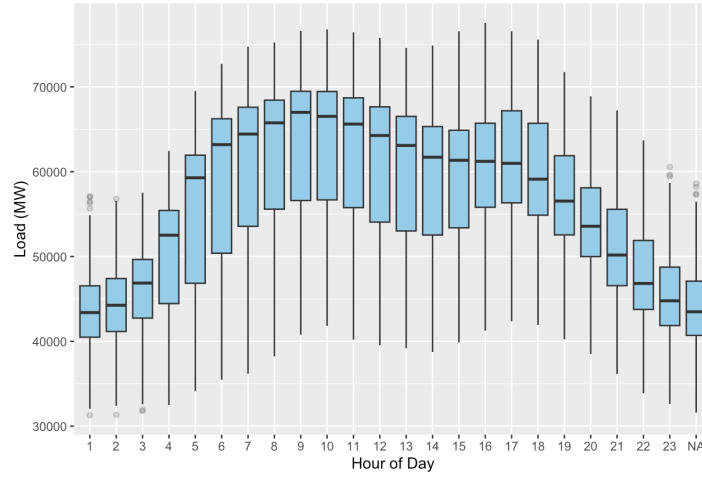


Figure 5: Germany: Hourly Distribution by Hour of the Day

- The interquartile range (box height) is widest during the midday hours, indicating greater variability in daytime electricity use.
- Outliers are more common in early morning and late evening, possibly due to unusual load spikes or drops.

5.2 Month of the Year

This heatmap in Figure 6 displays the average hourly electricity load (in megawatts) across each month in Germany. The color gradient represents the intensity of demand, with darker purples indicating lower demand and bright yellows indicating higher demand. The x-axis corresponds to the hour of the day, and the y-axis represents the month of the year.

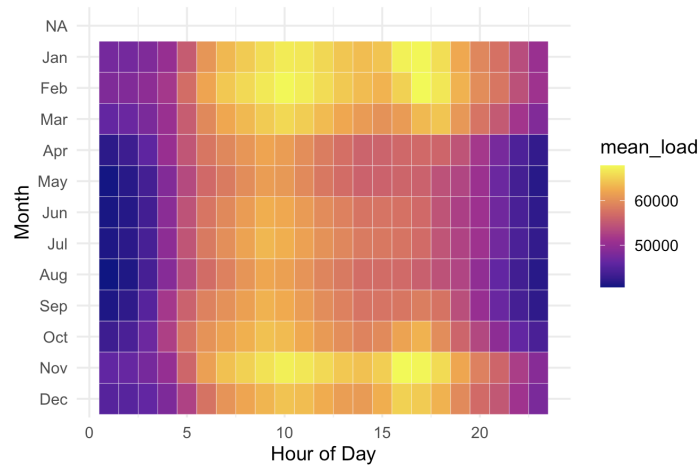


Figure 6: Germany: Average Load by Hour and Month

Interpretation

- Winter months (January, November, December) exhibit the highest overall demand, especially between 8:00 AM and 6:00 PM.
- Summer months (June, July, August) show lower average loads, particularly during night hours.
- Higher loads in winter suggest increased electricity use for heating, lighting, and possibly industrial activities, while summer demand is lower, possibly due to longer daylight hours and reduced heating needs.

6. How accurate are load forecasts compared to actual demand?

This scatter plot in Figure 7 compares the forecasted electricity load with the actual measured load in Germany. Each point represents one hourly observation. The black dashed line is the ideal 1:1 line, where forecast and actual values would match exactly.

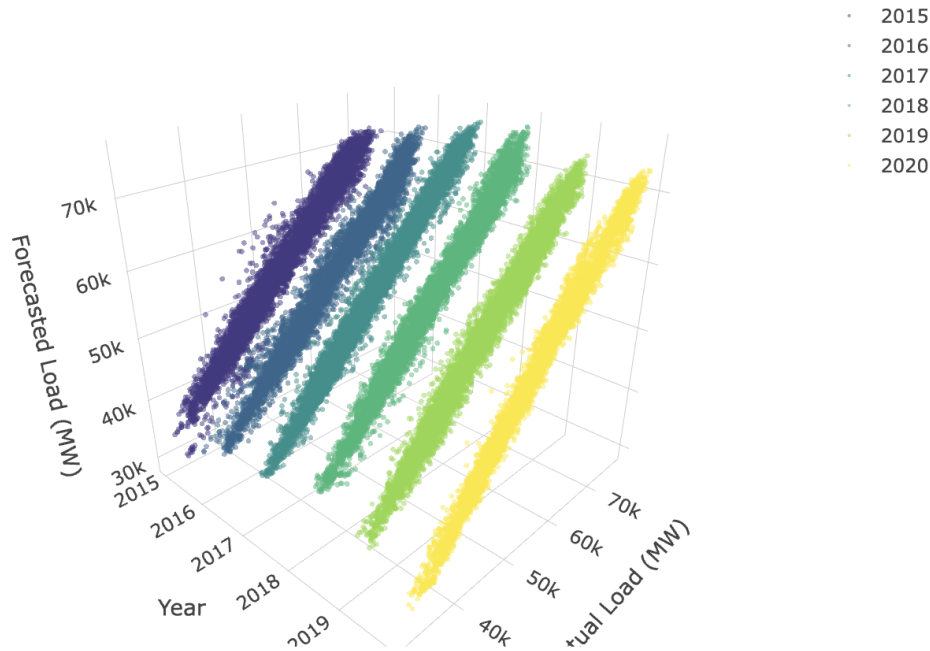


Figure 7: Germany: Forecasted vs Actual Load

Interpretation

- Most points lie closely along the diagonal line, indicating a strong agreement between forecasted and actual load.
- The tight, dense cluster of points around the 1:1 line suggests that the forecasting model used is highly accurate across a broad range of load values.
- Deviations become slightly larger at very high and very low load values, suggesting some forecast error during extreme demand conditions.
- No major signs of systematic over- or underestimation—the deviations appear random and balanced, which is a good indication of unbiased forecasting.

7. What Is the Relationship Between Temperature and Electricity Demand in Germany, and How Is It Affected by Extreme Values?

This scatter plot in Figure 8 visualizes the relationship between daily average temperature (°C) and total electricity demand (GWh/day) in Germany. Each point represents a daily observation. The blue line indicates a linear regression trend fitted across the full data range.

Interpretation

There is a clear negative correlation between temperature and electricity demand:

- Colder days (left side of the plot, below 10°C) are generally associated with higher electricity usage.
- Warmer days (right side of the plot) correspond to lower demand, particularly above 15°C.
- The trend line confirms this pattern, sloping downward, indicating that as temperature increases, average demand decreases.

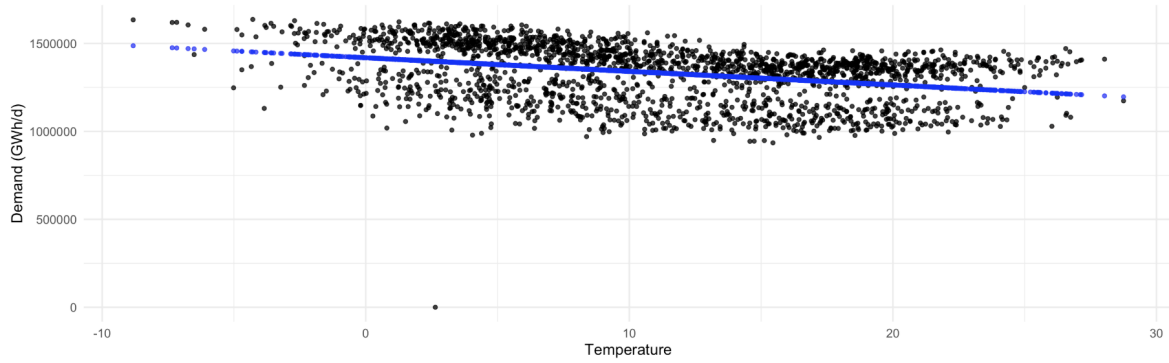


Figure 8: Demand vs Temperature (Daily)

- Some high-demand outliers occur even at moderate temperatures, possibly reflecting exceptional days (e.g., holidays, industrial activity, or anomalies in data reporting)

6 Conclusion

This project provided a comprehensive visualization-based analysis of electricity demand and renewable energy generation in Germany between 2014 and 2020, integrating both meteorological and energy datasets. Key findings include:

- **Temporal Patterns:** Solar energy showed strong seasonal cycles, peaking in summer, while wind energy exhibited irregular fluctuations with increased output during colder months.
- **Electricity Demand Trends:** Demand followed daily and seasonal rhythms—rising during working hours and winter months, and dropping at night and in summer.
- **Forecast Accuracy:** Forecasted electricity loads closely matched actual loads, demonstrating strong model reliability with only minor deviations during extreme load conditions.
- **Temperature-Demand Relationship:** A clear negative correlation exists between temperature and electricity demand—colder days correspond to higher usage, confirming heating demand impacts.
- **Renewables vs Consumption:** While renewable generation (especially wind) increased significantly, it still fell short of covering total national electricity consumption, underlining the need for continued expansion and integration.

7 Limitations

Despite the thorough visual and statistical analysis, this study has several limitations:

- **Missing and Imputed Data:** Both temperature and electricity datasets contained missing values. Imputation using mean values may smooth out true variability, especially during extreme weather or demand events.
- **Geographic Aggregation:** Renewable generation and demand were analyzed at the national level. Local-level disparities and transmission bottlenecks were not accounted for.
- **Renewable Scope:** The analysis only included solar and onshore wind generation. Other important sources such as hydro, biomass, and offshore wind were excluded.
- **Temporal Range:** The dataset ends in 2020, missing recent developments in the energy sector, especially those related to post-COVID demand recovery and geopolitical energy shifts.
- **Forecast Model Transparency:** While forecast accuracy was assessed, the underlying forecasting methodology (used by ENTSO-E) is not examined in detail.

8 References

Links for the Dataset:

- ENTSO-E Electricity Time
- Deutscher Wetterdienst (DWD)