Introduction

In this project, I analyzed a dataset containing hourly information about energy consumption and renewable energy sources (wind, solar, hydro). Each record represents one hour over an extended period and includes variables such as the hour of the day, day of the week, total consumption, and the contribution of each renewable source.

The main objective was to understand the distribution and efficiency of green energy, identify patterns by hour and day, estimate monthly and yearly trends, and develop predictive models using statistical and machine learning techniques.

Initial Data Processing

The dataset contains hourly records of:

- **Total energy consumption** (Consumption)
- **Hour of the day** (Hour) and **day of the week** (Weekday)
- Renewable energy production:
    - **Wind energy** (Wind)
    - **Hydropower** (Hydroelectric)
    - **Solar energy** (Solar)

Based on these variables, I calculated the following additional indicators:

- **total_green**: the total amount of energy produced from renewable sources
- **green_percentage**: the share of green energy in total consumption
- **efficiency_level**: classification of energy efficiency, defined as:
    - **Low**: below 20%
    - **Medium**: between 20% and 50%
    - **High**: above 50%

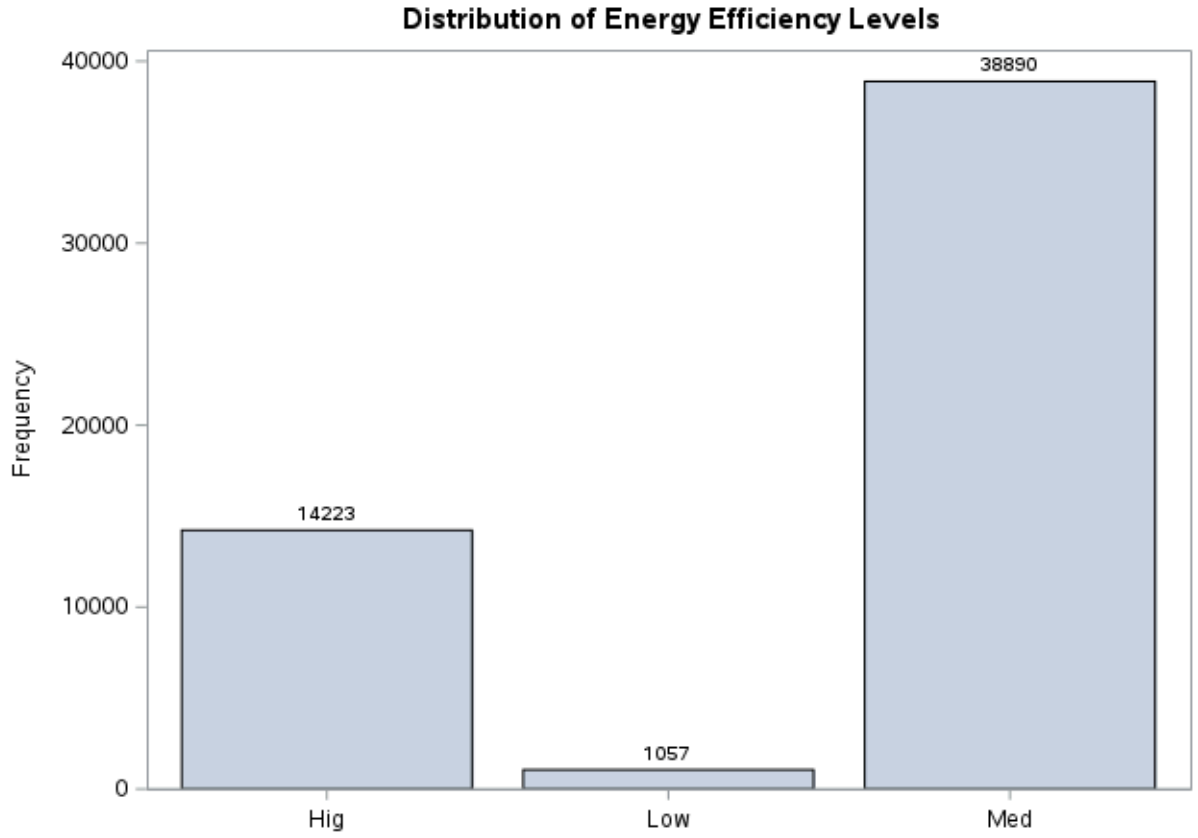
	Consumption	Wind	Hydroelectric	Solar	Hour	Weekday	total_verde	procent_verde	nivel_eficienta
1	6352	79	1383	0	0	Tuesday	1462	23.016372796	Mediu
2	6116	96	1112	0	1	Tuesday	1208	19.75147155	Scazut
3	5873	142	1030	0	2	Tuesday	1172	19.95572961	Scazut
4	5682	191	972	0	3	Tuesday	1163	20.468145019	Mediu
5	5557	159	960	0	4	Tuesday	1119	20.136764441	Mediu
6	5525	91	958	0	5	Tuesday	1049	18.986425339	Scazut
7	5513	98	938	0	6	Tuesday	1036	18.791946309	Scazut
8	5524	93	1187	0	7	Tuesday	1280	23.171614772	Mediu
9	5510	51	1325	8	8	Tuesday	1384	25.117967332	Mediu
10	5617	15	1398	61	9	Tuesday	1474	26.241766067	Mediu

Key Observations

- Solar energy is absent during nighttime hours but starts contributing around 8–9 AM.
- Hydropower is consistently present and plays a major role in total renewable production.

1. Frequency of Energy Efficiency Levels

A vertical bar chart shows the frequency of records for each of the three energy efficiency levels (efficiency_level), based on the percentage of green energy.



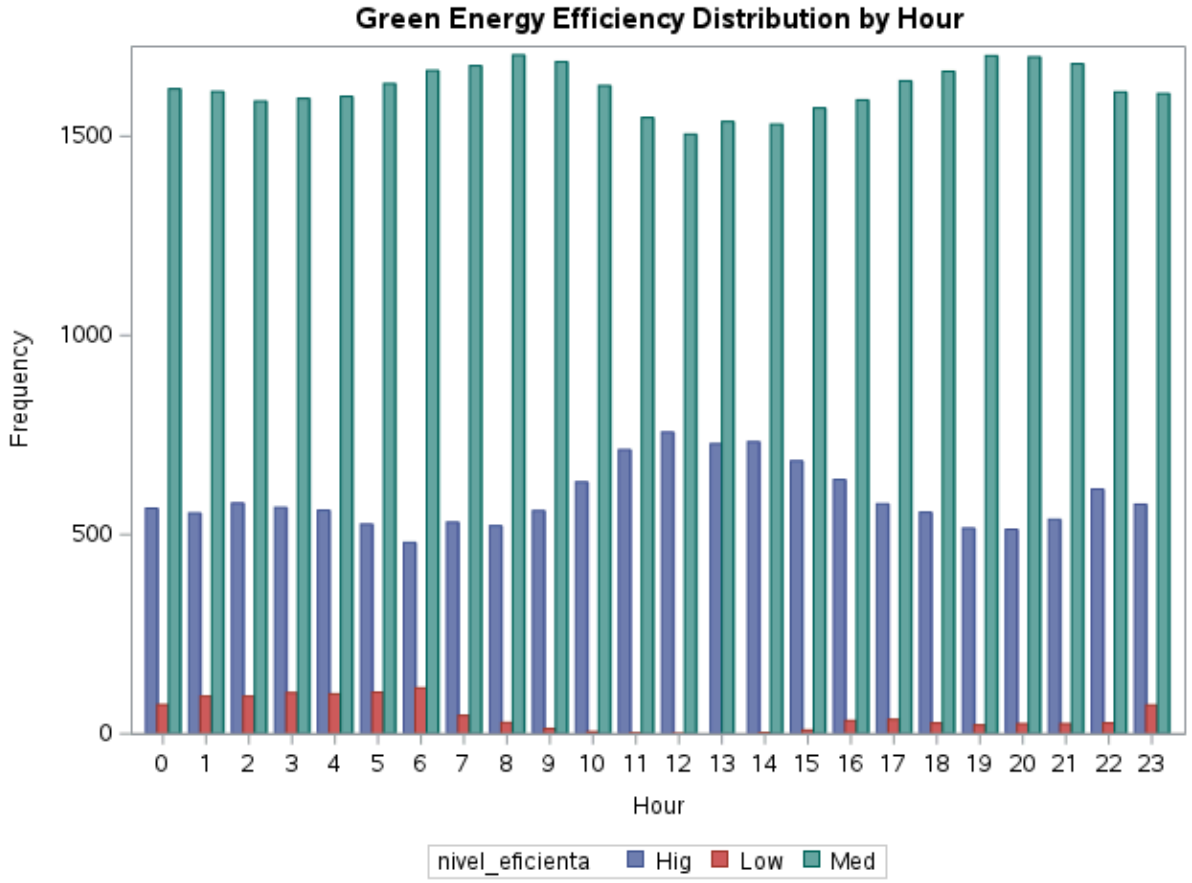
2. Green Energy Efficiency Distribution by Hour

This chart illustrates the frequency of the three efficiency levels (Low, Medium, High) across the 24 hours of the day. Each hour is represented by a group of color-coded bars:

- **Blue** – High efficiency
  Red – Low efficiency
- **Green** – Medium efficiency

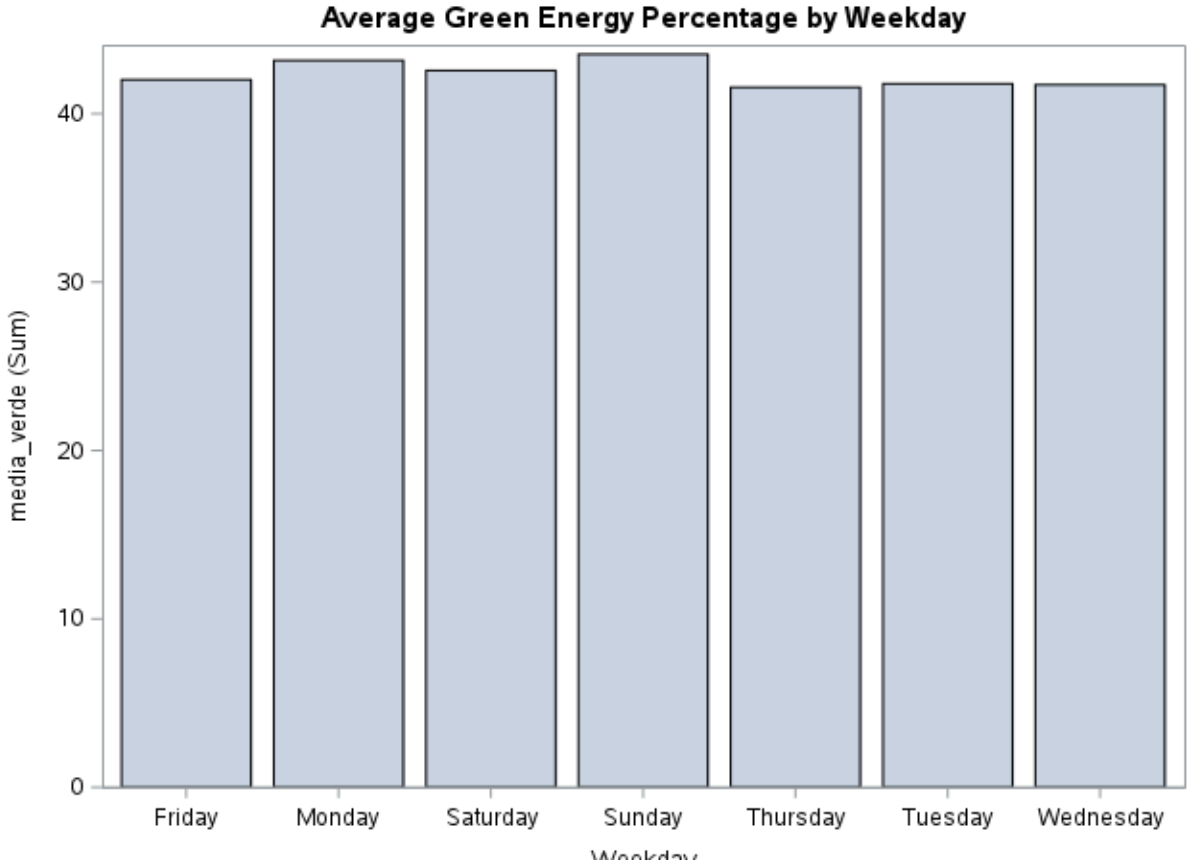
Interpretation

- The **Medium** level is predominant across all hours, indicating that renewable sources typically cover 20–50% of total consumption.
- **Low** efficiency is more common between 0:00 and 7:00, when solar energy is unavailable and wind contribution is variable.
- **High** efficiency is rare but shows a slight increase between 10:00 and 16:00, when solar energy is also available.

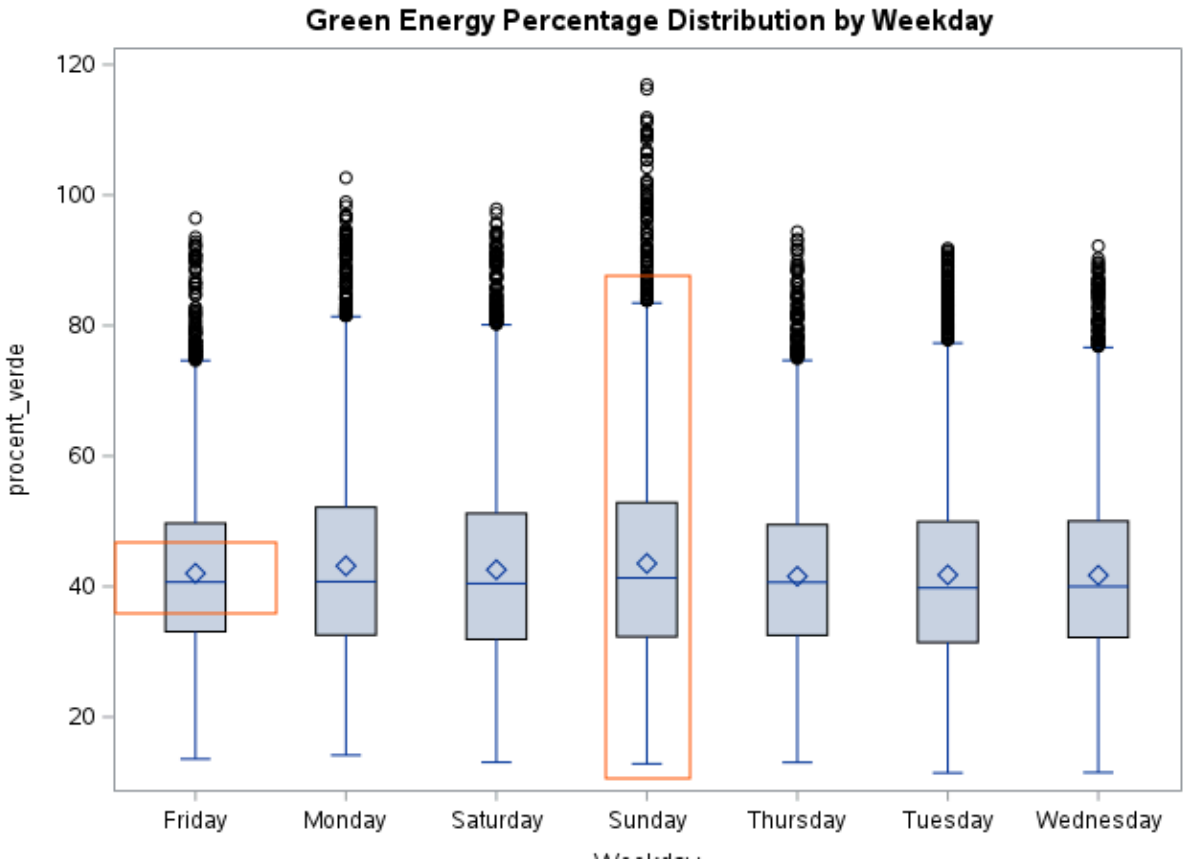


3. Average Green Energy Percentage by Day of the Week

This chart displays the average green energy percentage (green_percentage) for each day of the week. Each vertical bar represents the average of all hourly values for that day.



The daily average of green energy is relatively stable throughout the week, with a slight increase during weekends.



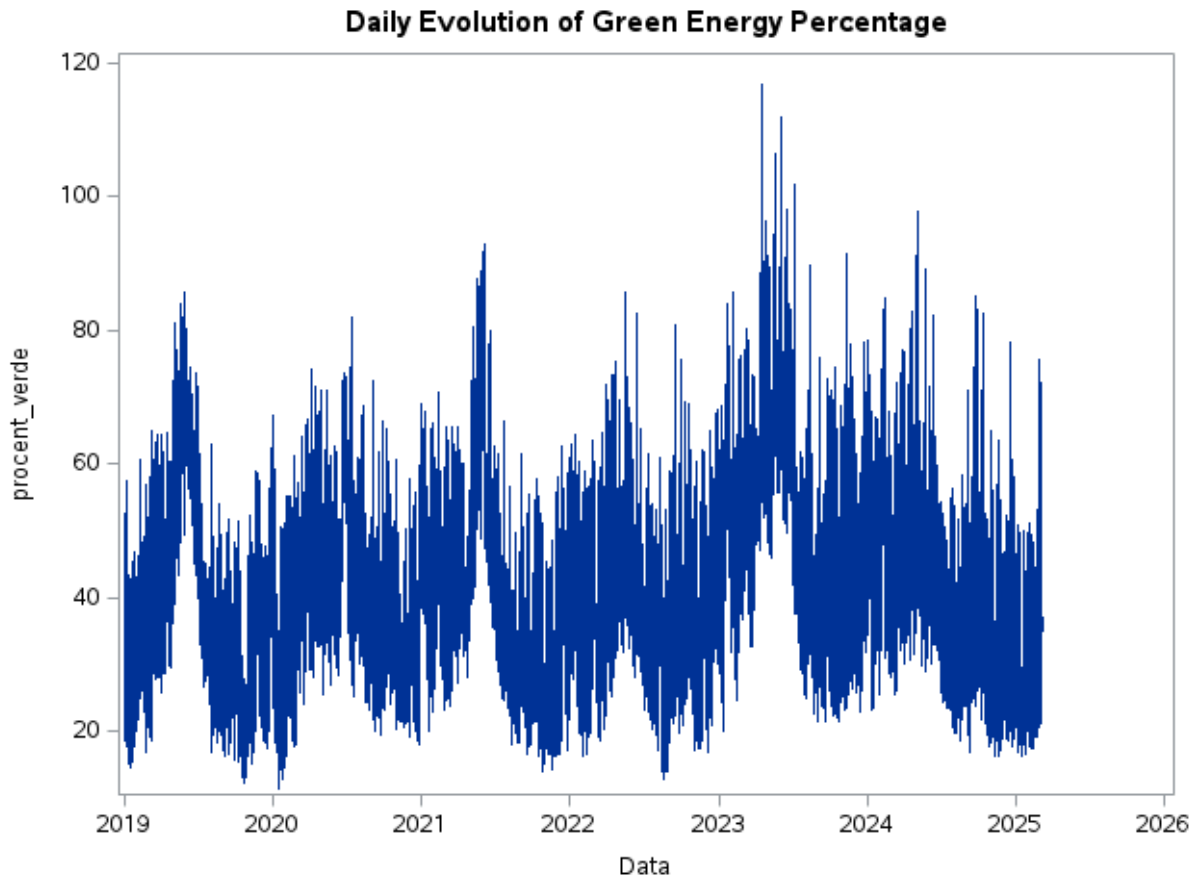
The percentage of green energy shows significant variability within each day. While the median remains around 40%, there are many cases where renewable sources exceed 80–90%, indicating potential for energy autonomy during certain hours.

4. Daily Evolution of Green Energy Percentage

To analyze the data over time, I created a simulated calendar:

- Each record represents one hour.
- 24 records were grouped into calendar days.
- A date variable was assigned to each observation.

```
zi = int((_N_-1) / 24);
Data = '01JAN2019'd + zi;
```



The daily evolution of green energy shows a fluctuating pattern, likely influenced by natural variations in production (e.g., lack of wind or cloudy weather). While some days demonstrate remarkable efficiency, the absence of a clear upward trend suggests the need for a combination of different renewable sources to ensure consistency.

5. Simulated Evolution of Renewable Energy (2019–2021)

To test a predictive model for energy efficiency, I created a hypothetical simulation:

- First 9,000 rows → 2019
- Next 9,000 rows → 2020
- Remaining rows → 2021

Additionally, I applied an artificial 3% annual increase in the green energy percentage.

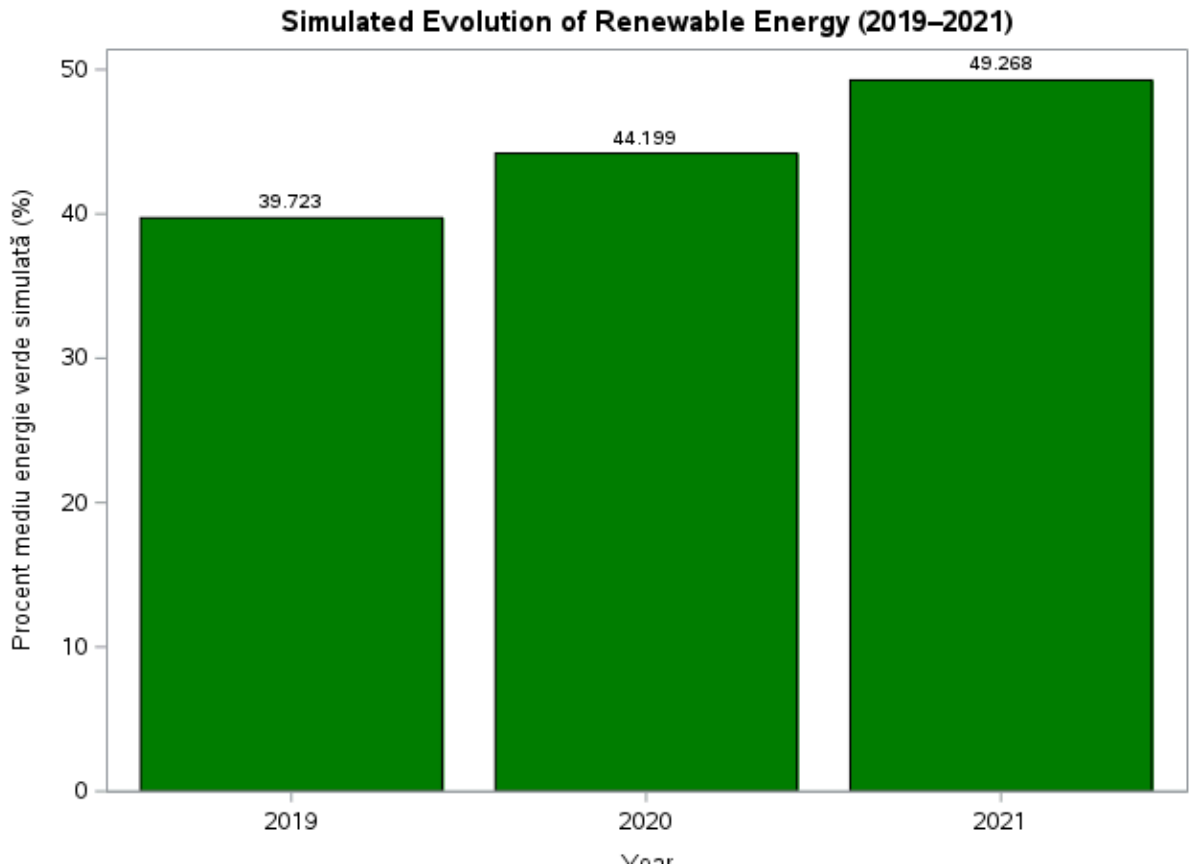
```
data curs1.simulat;
    set curs1.lunar;
    if _N_ <= 9000 then Year = 2019;
    else if _N_ <= 18000 then Year = 2020;
    else Year = 2021;

    procent_verde_sim = procent_verde + (Year - 2019) * 3;
run;
```

A vertical bar chart illustrates the simulated average annual green energy percentage:

- **2019:** 39.72%
- **2020:** 44.20%
- **2021:** 49.27%

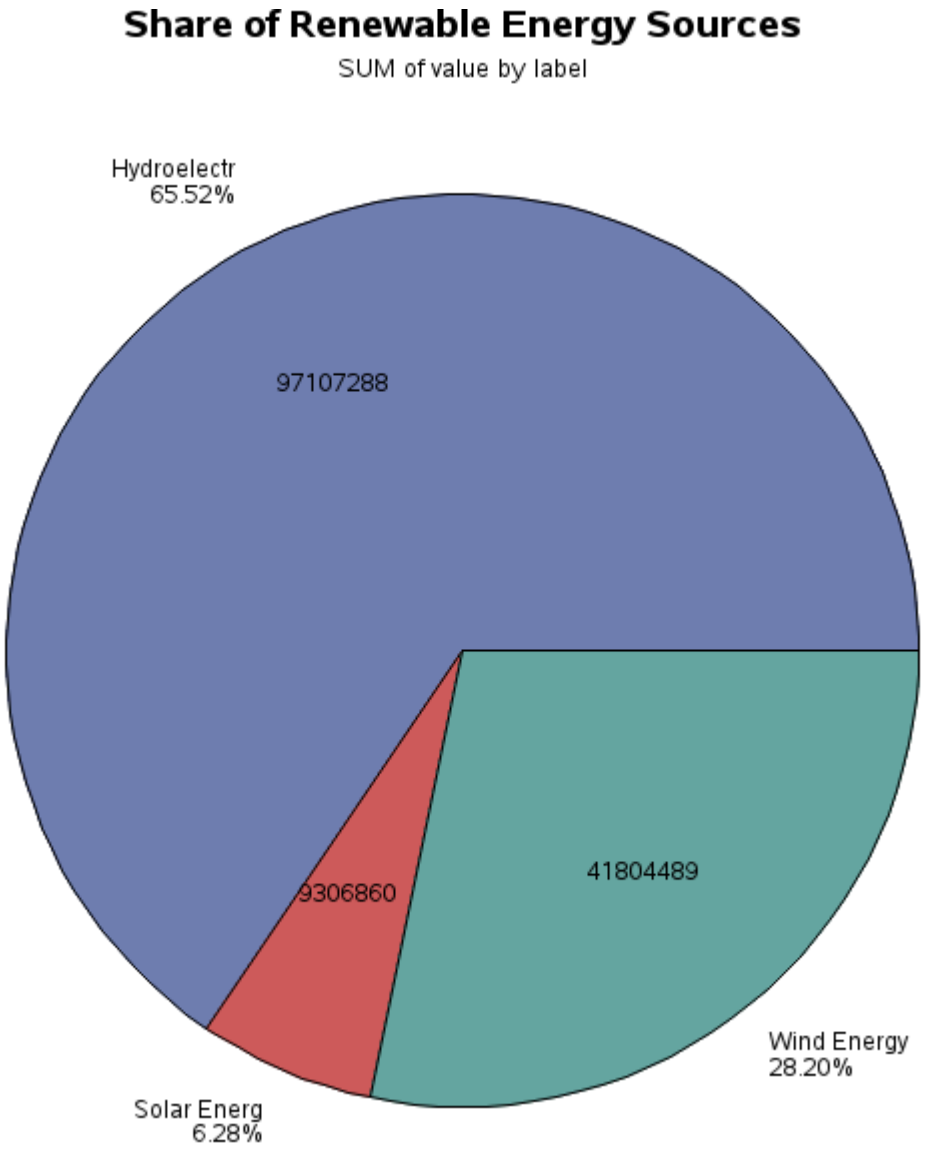
This reflects a steady annual increase of approximately 4.5%, based on the simulation assumption of a +3% yearly adjustment relative to the initial real values.



6. Share of Renewable Energy Sources

The pie chart presents the overall contribution of the three renewable sources:

- Wind energy: **28.20%**
- Hydropower: **65.52%**
- Solar energy: **6.28%**



Interpretation

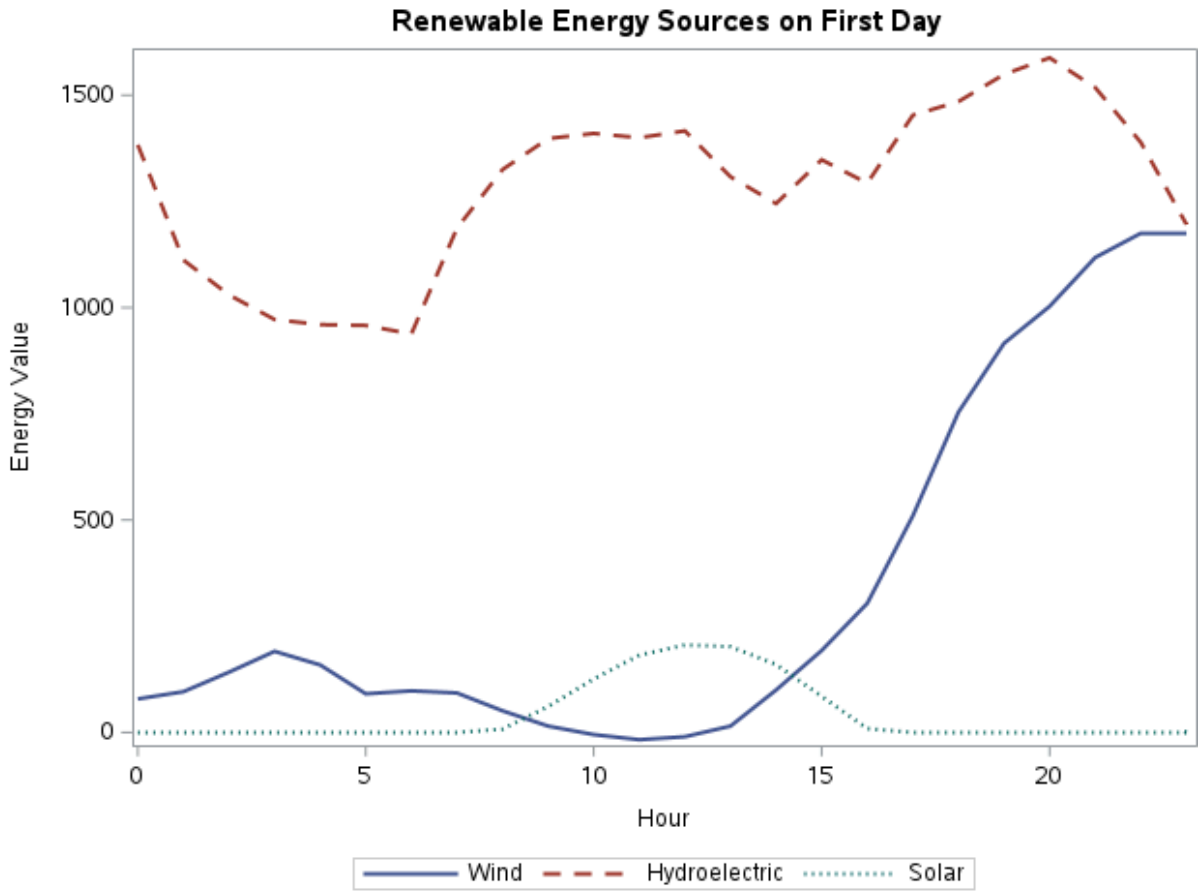
- Hydropower is the dominant source, contributing over 65% of total green energy.
- Wind energy has a significant share, accounting for over one-quarter of renewable production.
- Solar energy has the smallest share (only 6.28%), mainly due to its absence at night.

This distribution highlights the strong dependence on hydropower. Although wind energy contributes meaningfully, the limited role of solar energy indicates potential for future expansion, particularly through photovoltaic development.

7. Evolution of Renewable Sources During the First Day

A line chart shows the hourly values of the three renewable sources on the first day:

- **Blue line:** Wind energy
- **Red line:** Hydropower
- **Green line:** Solar energy

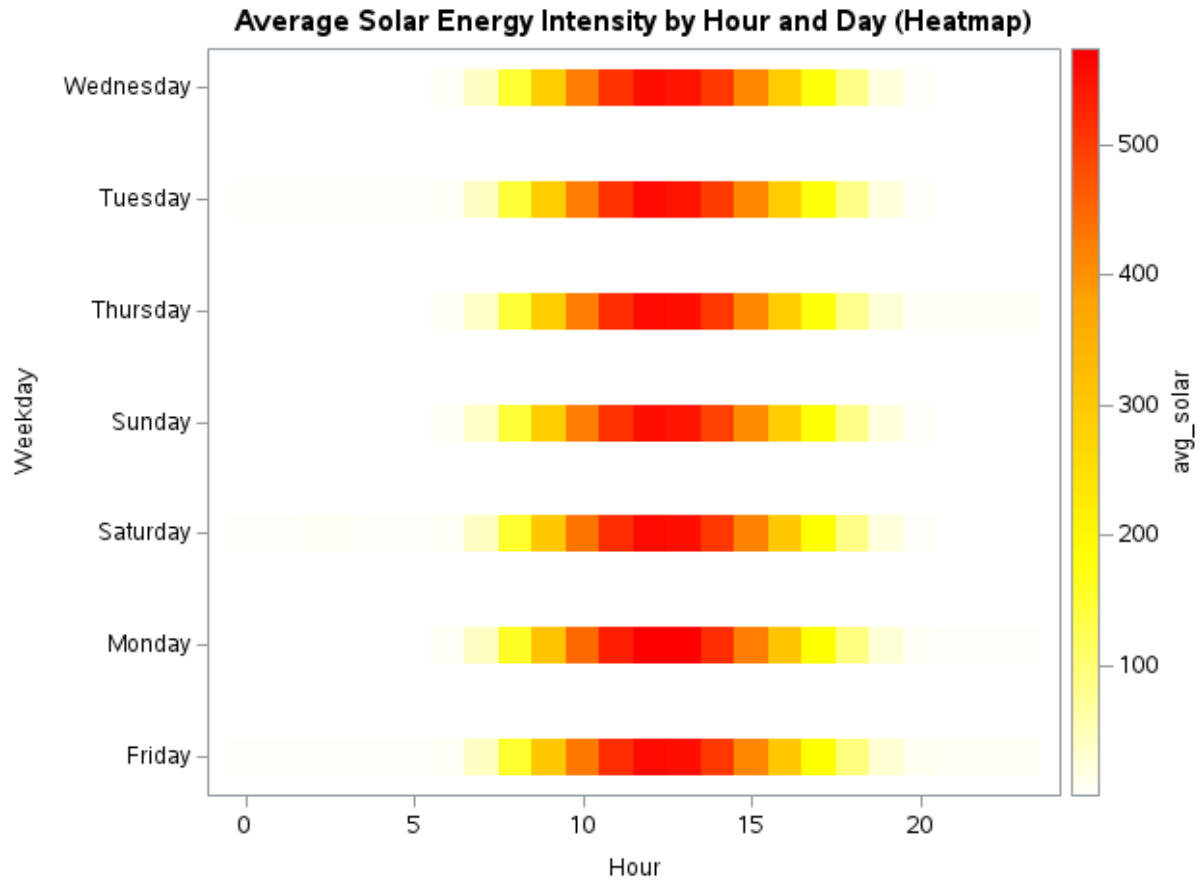


Renewable energy sources exhibit complementary behavior throughout the day. Hydropower remains the stable base source, wind energy adds variability, and solar energy provides a limited contribution during daylight hours. This pattern demonstrates the importance of a balanced energy mix to meet daily demand.

8. Average Solar Energy Intensity by Hour and Day

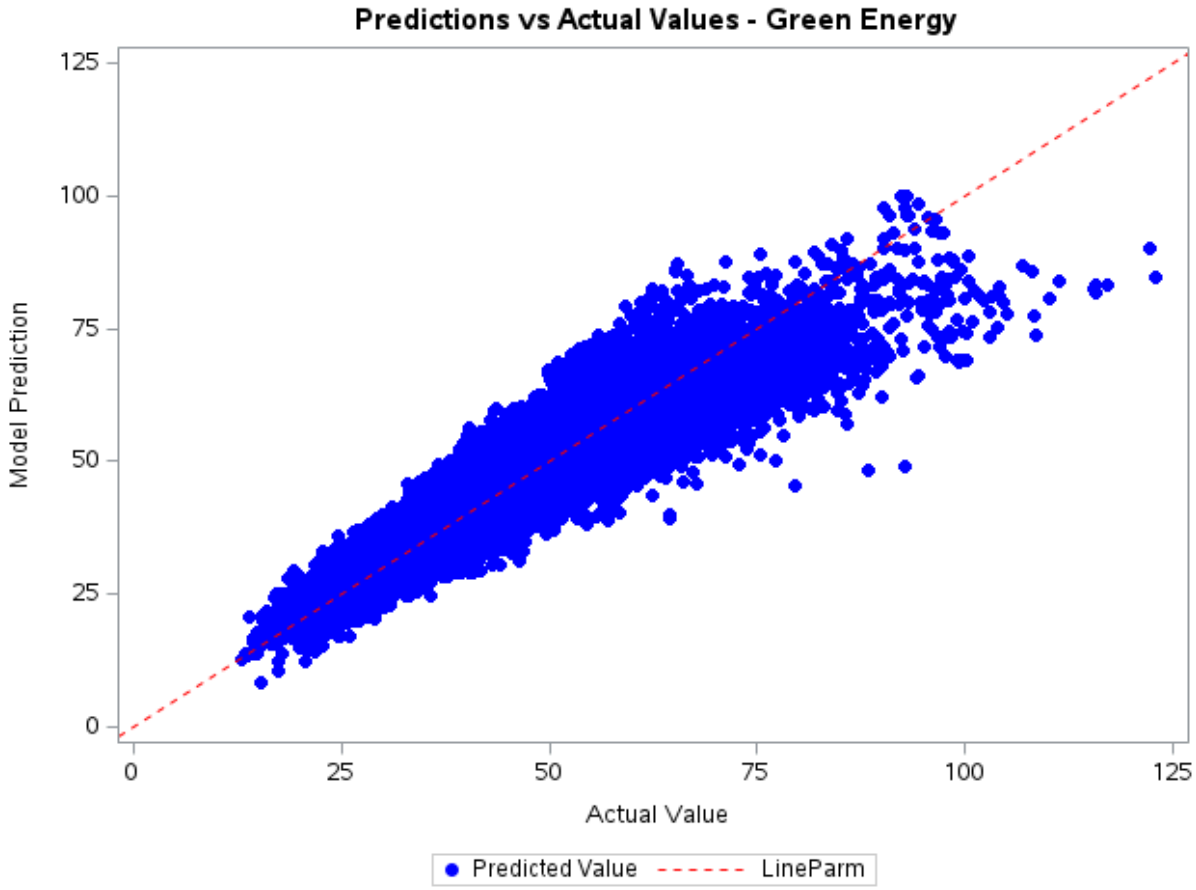
This visualization simulates a heatmap using colored markers:

- X-axis: Hour of the day (0–23)
- Y-axis: Day of the week
- Color: Average solar energy intensity for each hour/day combination



Interpretation

- Solar energy is absent at night (0:00–6:00 and 20:00–23:00).
- The 9:00–16:00 interval is the most active for solar production across all days.
- Solar energy peaks around midday, with a consistent pattern across the week. Early morning and evening hours require complementary energy sources.


Predictions vs Actual Values - Green Energy
Model Prediction
Actual Value
0
25
50
75
100
125
Predicted Value
LineParm