Version: October 10, 2025

Data Science I

Project Report: Predicting Energy Prices

Submission Deadline: January 25 2026, 21:00 UTC

University of Oldenburg

Winter 2025/2026

Overall Points: / 100

Instructors: Jannik Schröder, Wolfram "Wolle" Wingerath

Submitted by: <Writwik Ray>

Introduction

Day-ahead prices are the prices set for electricity delivery on the next day, determined through a daily auction process based on anticipated demand and supply. These prices provide a forecast that helps producers and consumers plan their generation and consumption, reflecting expected market conditions. Predicting these prices is crucial for energy companies and traders to optimize their operations, manage risks, and improve market efficiency.

Over the course of this semester project, you will build a model to predict hourly day-ahead energy prices for Germany on **January 27, 2026** as reported by the SMARD.de information platform (see Figure 1 below for an example).

alternative text
Figure 1: Day-Ahead Prices for October 10, 2025.

Report Structure & Semester Overview

This module is designed in a way that you can reuse and repurpose work from the exercise sheets when working on the project report. By building this semester project report, you can thus apply the knowledge you pick up in the lectures in a practical context. There are 4 phases overall that are loosely aligned with the topics of the lectures and exercise sessions.

Note: We may update the instructions with clarifications or additional resources during the semester, so **please check Stud.IP for new versions** before starting to work on a new phase. You can find the publication date of the project report instructions on the top-right corner.

Phase 1: Gathering domain knowledge and Data Sources

Phase 1: Gathering Domain Knowledge & Data Sources

/ 25

For building a model that predicts energy prices, you are required to gather not only historical data on energy prices, but also gain a strong understanding of the energy sector's dynamics. This entails collecting data for recent years on energy prices as well as demand and production capacities, weather conditions, supply metrics, and any significant events impacting energy prices. Additionally, acquiring domain knowledge is crucial for understanding the factors that influence energy markets.

You can think of your report as a story about data discovery, where the background information sets the stage for understanding the significance and narrative of your analytical findings. It should therefore include important facts on the domain that you are exploring.

- Research the energy market's structure, including how prices are set and the role of different energy sources.
- Understand regulatory policies, market reforms, and technological advancements that could impact energy prices.
- Review energy sector publications, expert analyses, and videos or other resources.

Possible data sources include energy market databases, weather archives, government publications on energy, and news articles reporting on the energy sector. Combining comprehensive data collection with in-depth domain knowledge will lay a solid foundation for developing accurate predictive models and meaningful analysis.

Market Structure, Pricing system and role of energy sources, and data source

Datasets and Sources:

This dataset contains average hourly, daily and monthly wholesale day-ahead electricity prices for European countries. The prices shown are what generators get for selling electricity on the spot market. They're different from what consumers pay, which can include extra costs like taxes, network fees, subsidies, and supplier markups.

The DWD (Deutscher Wetterdienst) dataset provides historical and current weather data for Germany, including temperature, wind speed, solar radiation, and precipitation. Using DWD data alongside an energy price dataset can help analyze how weather patterns impact energy demand and prices. For example, high wind speeds might boost wind power generation, potentially lowering prices, while cold snaps could increase heating demand and raise prices. Combining these datasets can improve forecasting models.

For data on several sectors that produce energy we collected them from the SMARD website.

Market Structure and Pricing System

Germany's electricity market uses marginal pricing with a merit order system.

Merit Order: Power plants are ranked and switched on according to their marginal operating costs, basically the cost of operation, specifically excluding construction costs energy transition etc. This creates an ascending cost curve .

Marginal Pricing Mechanism: Energy suppliers bid their marginal costs in day-ahead auctions. The most expensive bid needed to meet demand sets the price everyone gets – that's uniform pricing. All accepted plants get this price, even if they bid lower. For example, if the price is set at 60 EUR/MWh, all dispatched plants get 60 EUR/MWh, which means cheaper plants like renewables make a nice profit.

The Merit Order Effect: Renewable energy substantially reduces the amount of highly priced peak electricity that transmission companies need to buy when solar and wind power are available. A Fraunhofer Institute study found solar power reduced German electricity prices by 10% on average and as much as 40% in early afternoon.
The system faced intense scrutiny during the 2022 energy crisis when skyrocketing natural gas prices drove electricity costs to over €1,000/MWh, since gas plants often set the marginal price.

In a nutshell:

- Power plants are ranked by marginal operating costs (cheapest first, merit order).
  - Waste & renewables (cheapest)
  - Nuclear & coal (mid-range)
  - Gas & oil (most expensive, used for peaks)
- Bids are submitted; the most expensive bid sets the price for all.
- All accepted plants get the same price (this is known as uniform pricing).

A note on regulatory policies, market reforms and technological advancements that affect energy prices

Regulatory Policies and their effects on energy prices:

- Capacity Market (2027-2028): Backup power plants get support to stay viable. Plan to build 20 GW of new gas-fired capacity by 2030. This may result in slight increment in prices through socialized costs but prevents extreme price spikes by ensuring backup power availability.
- Renewables Support Reform: Adjusting subsidies to EU rules. Future projects must respond to price signals (e.g., stop production during oversupply). Reduces price volatility by curtailing production during oversupply, eliminating negative price hours while preventing artificial price suppression.
- No more solar subsidies: Rooftop solar and battery combinations can now go solo. Minimally impacts wholesale prices but enables distributed storage to reduce peak demand and lower evening prices.
- Climate Protection Contracts: Government compensates for CO_2 price gaps (like a price guarantee). Neutral effect on electricity prices while de-risking industrial clean energy investments.
- Electricity Tax Cut: Tax reduced to EU minimum (0.05 cents/kWh) for companies (2024-2025, maybe permanent). Modestly increases wholesale prices by stimulating industrial demand and accelerating electrification.

Market Reforms and their effects on energy prices:

- EU Market Design Reform (June 2024): Boosting PPAs and CfDs to stabilize energy prices. Long-term PPAs and CfDs stabilize prices by reducing spot market exposure, though may slightly raise average prices as generators secure minimum price guarantees.
- Germany's Market Modernization: Reforms focus on: renewables support, flexible power use, local pricing signals and generation capacity. Local pricing signals and flexibility incentives reduce grid congestion costs and improve price efficiency across regions.
- Storage gets a boost: Can now participate in frequency response markets, making investments more viable. New revenue from frequency markets accelerates storage deployment, which flattens price volatility by arbitraging between cheap and expensive periods.

Technological Advancements:

- Renewables hit big numbers:62.7% of electricity from renewables in 2024 and solar capacity hit 100 GW (added 15.5 GW in 2024). Massively increases zero-marginal-cost generation, pushing down average wholesale prices through the merit order effect while increasing frequency of negative prices during oversupply periods.
- Battery storage is exploding: 600,000 new battery storage systems in 2024 (+50% capacity), 18.2 GWh storage in Jan 2025; ~24 GWh by end-2025. Reduces price volatility by storing cheap renewable energy and releasing it during expensive peak hours, flattening price spikes and making the market more predictable.
- Large-scale storage surges: over 81% growth and 9,710 battery storage grid-connection requests (400 GW power, 661 GWh storage). Future deployment will further compress peak-to-trough price spreads as storage arbitrage becomes more competitive, potentially reducing profitability of peaker plants and lowering maximum prices.

```
Requirement already satisfied: requests in /root/venv/lib/python3.11/site-packages (2.32.5)
Requirement already satisfied: pandas in /root/venv/lib/python3.11/site-packages (2.1.4)
Requirement already satisfied: openpyxl in /root/venv/lib/python3.11/site-packages (3.1.5)
Requirement already satisfied: charset_normalizer<4,>=2 in /root/venv/lib/python3.11/site-packages (from requests) (3.4.4)
Requirement already satisfied: idna<4,>=2.5 in /root/venv/lib/python3.11/site-packages (from requests) (3.11)
Requirement already satisfied: urllib3<3,>=1.21.1 in /root/venv/lib/python3.11/site-packages (from requests) (2.6.3)
Requirement already satisfied: certifi>=2017.4.17 in /root/venv/lib/python3.11/site-packages (from requests) (2026.1.4)
Requirement already satisfied: numpy<2,>=1.23.2 in /root/venv/lib/python3.11/site-packages (from pandas) (1.26.4)
Requirement already satisfied: python-dateutil>=2.8.2 in /root/venv/lib/python3.11/site-packages (from pandas) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /root/venv/lib/python3.11/site-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.1 in /root/venv/lib/python3.11/site-packages (from pandas) (2025.3)
Requirement already satisfied: et-xmlfile in /root/venv/lib/python3.11/site-packages (from openpyxl) (2.0.0)
Requirement already satisfied: six>=1.5 in /root/venv/lib/python3.11/site-packages (from python-dateutil>=2.8.2->pandas) (1.17.0)

[notice] A new release of pip is available: 24.0 -> 25.3
[notice] To update, run: pip install --upgrade pip
Note: you may need to restart the kernel to use updated packages.
```

	Country object	ISO3 Code object	Datetime (UTC) o...	Datetime (Local) o...	Price (EUR/MWhe)
96...	Germany	DEU	2026-01-21 02:0...	2026-01-21 03:0...	100.88
96...	Germany	DEU	2026-01-21 03:0...	2026-01-21 04:0...	99.09
96...	Germany	DEU	2026-01-21 04:0...	2026-01-21 05:0...	98.85
96...	Germany	DEU	2026-01-21 05:0...	2026-01-21 06:0...	118.44
96...	Germany	DEU	2026-01-21 06:0...	2026-01-21 07:0...	118.44

5 rows, 5 cols 10 / page Page 1 of 1

```
Data downloaded: stundenwerte_FF_01766_akt.zip
```

```
   STATION_ID        date  quality  wind_speed  wind_direction  eor
0        1766  2024072000       10         0.5              40  eor
1        1766  2024072001       10         0.6              40  eor
2        1766  2024072002       10         0.9              40  eor
3        1766  2024072003       10         0.8              40  eor
4        1766  2024072004       10         0.6              50  eor
```

Phase 2: Data Cleaning and Exploratory Data Analysis

/ 25

Phase 2: Data Cleaning & Exploratory Data Analysis (EDA)

This phase is critical for uncovering insights from your collected data and setting the stage for predictive modeling.

- Prepare your data for analysis by handling missing values, outliers, and ensuring data quality.
- Use statistical summaries and visualizations to understand the basic characteristics of the data. Explore the distribution of key variables and identify patterns or anomalies.
- Investigate the relationships between different variables, especially how various factors such as weather conditions, demand, and significant events correlate with energy prices. Utilize statistical methods like Pearson's or Spearman's rank correlation coefficient to quantify the strength and direction of these relationships. This step is crucial for identifying potential predictors for your models.

This initial analysis will not only help you understand the data better but also guide your choice of features for the predictive modeling phase. It's essential to document any insights or interesting findings during this stage, as they could be valuable for the final report and presentation.

	price float64	date int64
96...	100.88	2026012102
96...	99.09	2026012103
96...	98.85	2026012104
96...	118.44	2026012105
96...	118.44	2026012106

5 rows, 2 cols 10 / page Page 1 of 1

	date int64	quality int64	wind_speed float...	wind_direction in...
0	2024072000	10	0.5	40
1	2024072001	10	0.6	40
2	2024072002	10	0.9	40
3	2024072003	10	0.8	40
4	2024072004	10	0.6	50

5 rows, 4 cols 10 / page Page 1 of 1

	date int64	quality int64	wind_speed float...	wind_direction in...	price float64
0	2024072000	10	0.5	40	86.42
1	2024072001	10	0.6	40	88.02
2	2024072002	10	0.9	40	89.47
3	2024072003	10	0.8	40	89.7
4	2024072004	10	0.6	50	84.9

5 rows, 5 cols 10 / page Page 1 of 1

```
-0.021953212573105306
```

The negative correlation tells us that when wind generation is high, it pushes the expensive coal/gas plants out of the "merit order." The price is then set by the next most expensive plant still running, which is now a cheaper one. This effectively suppresses the wholesale price. However it is very close to 0 this suggests other factors (like gas prices, nuclear outages, or demand spikes) are driving the price more than wind speed. Which is why we will focus on other production methods rather than just wind or renewable sources.

```
/tmp/ipykernel_604/2333080659.py:2: UserWarning: Could not infer format, so each element will be parsed individually, falling bac
  actual_gen['Start date'] = pd.to_datetime(actual_gen['Start date'])
```

	date int64	Biomass [MWh] ...	Hydropower [M...	Wind offshore [M...	Wind onshore [M...	Photovoltaics [M...	Other renewable ...	N
0	2024072000	3,883.25	2,336.25	2,303.50	5,769.50	3.25	116.5	-
1	2024072001	3,808.00	2,500.75	2,630.25	5,763.50	3.25	117	-
2	2024072002	3,775.50	2,362.00	2,559.25	5,658.00	4.00	117.5	-
3	2024072003	3,787.50	2,413.25	2,125.25	5,487.00	3.50	118.25	-
4	2024072004	3,852.50	2,458.00	1,900.50	5,267.75	4.00	118.5	-

5 rows, 13 cols 10 / page Page 1 of 1

	date int64	quality int64	wind_speed float...	wind_direction in...	price float64	Biomass [MWh] ...	Hydropower [M...	V
0	2024072000	10	0.5	40	86.42	3,883.25	2,336.25	2
1	2024072001	10	0.6	40	88.02	3,808.00	2,500.75	2
2	2024072002	10	0.9	40	89.47	3,775.50	2,362.00	2
3	2024072003	10	0.8	40	89.7	3,787.50	2,413.25	2
4	2024072004	10	0.6	50	84.9	3,852.50	2,458.00	1

5 rows, 17 cols 10 / page Page 1 of 1

/ 25

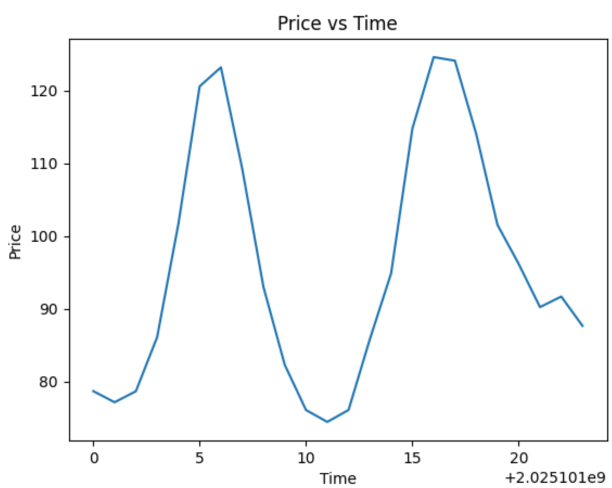
Phase 3: Visualization & Storytelling

Effective visualization is key to interpreting the data and predictive modeling results comprehensively.

- Create visualizations that illustrate the historical trends in energy prices and highlight the key factors influencing price fluctuations.
- Use graphs to compare actual vs. predicted energy prices to evaluate the model's performance visually.
- Explain your findings by connecting the dots between visuals, forecasts, and insights, making it easier to understand how energy prices change and what that means for people and businesses.

Visualizations should not only aim to support the narrative of your findings but also uncover insights that might not be immediately apparent from the raw data or statistical metrics alone. A well-crafted visualization can make complex data more accessible and insights more compelling.

```
       price        date
94440  78.67  2025101000
       price        date
94441  77.14  2025101001
       price        date
94442  78.65  2025101002
       price        date
94443  86.09  2025101003
        price        date
94444  101.66  2025101004
        price        date
94445  120.56  2025101005
        price        date
94446  123.18  2025101006
        price        date
94447  109.25  2025101007
       price        date
94448  92.97  2025101008
       price        date
94449  82.29  2025101009
       price        date
94450  76.07  2025101010
       price        date
94451  74.44  2025101011
       price        date
94452  76.08  2025101012
       price        date
94453   85.8  2025101013
       price        date
94454  94.85  2025101014
```



We can see the plot above shows us a "M" like figure which indicates what we would generally expect from a energy price graph. There are two peaks where price is high and one valley in the mornings. This is because we use the most amount of energy during early morning and afternoons.