Forecasting energy price advanced time series prediction

Presented By:
Aadip Thapaliya and Helge Röpke

CONTENT

- Motivation
- Dataset
- Baseline Models
- Models
- Model comparison
- Conclusion

MOTIVATION

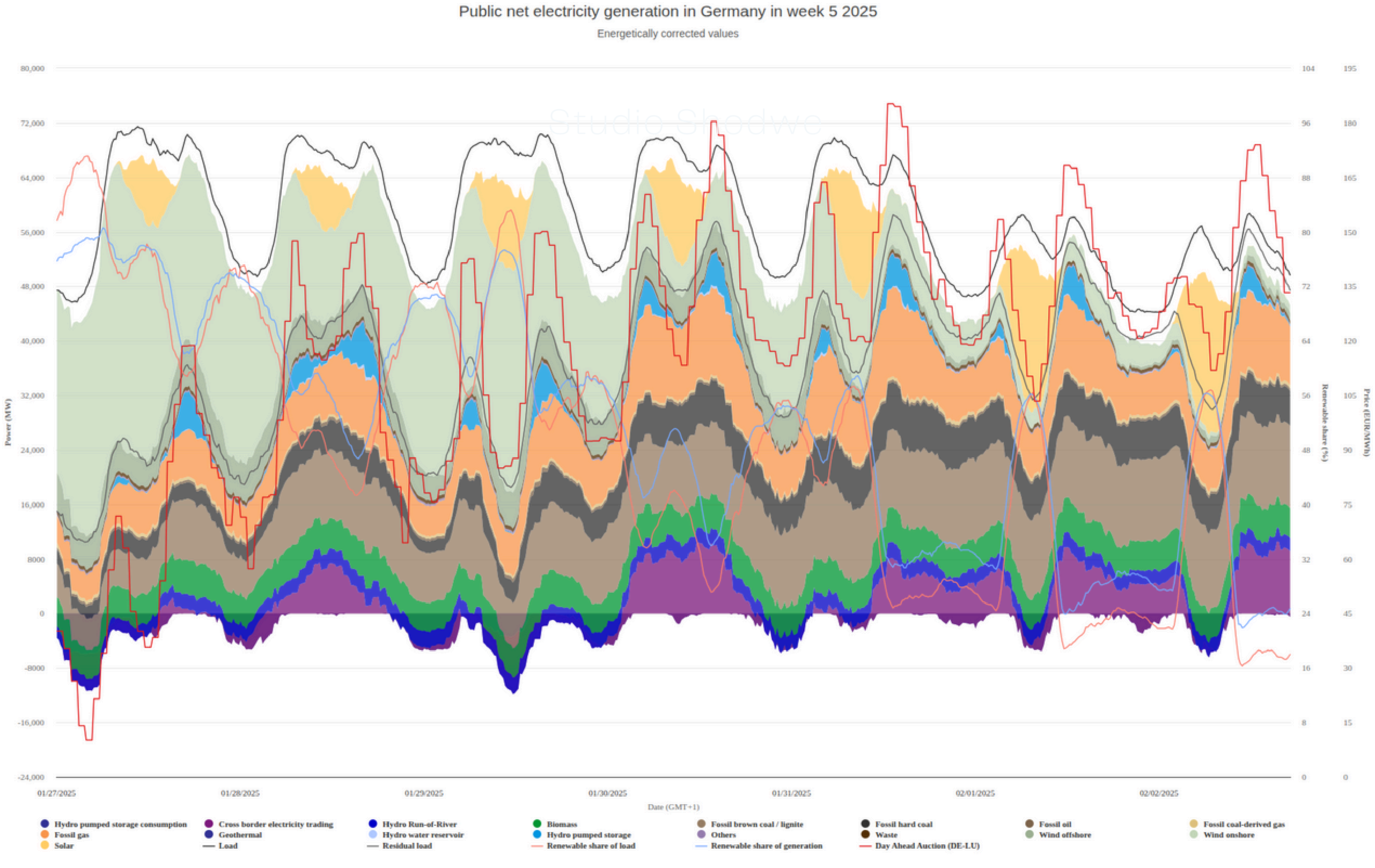
Our motivation is to explore how modern time-series forecasting techniques can be applied to real-world energy markets. We want to deepen our understanding of data analysis and machine learning by working with high-resolution **energy price data from 2024 and 2025.**

- Get practical experiencing for prediction time series analysis
- Compare classic and contemporary models across sectors
- uncover patterns driven by renewables, demand, and market behavior
- build a reliable forecasting system that supports smarter decisions

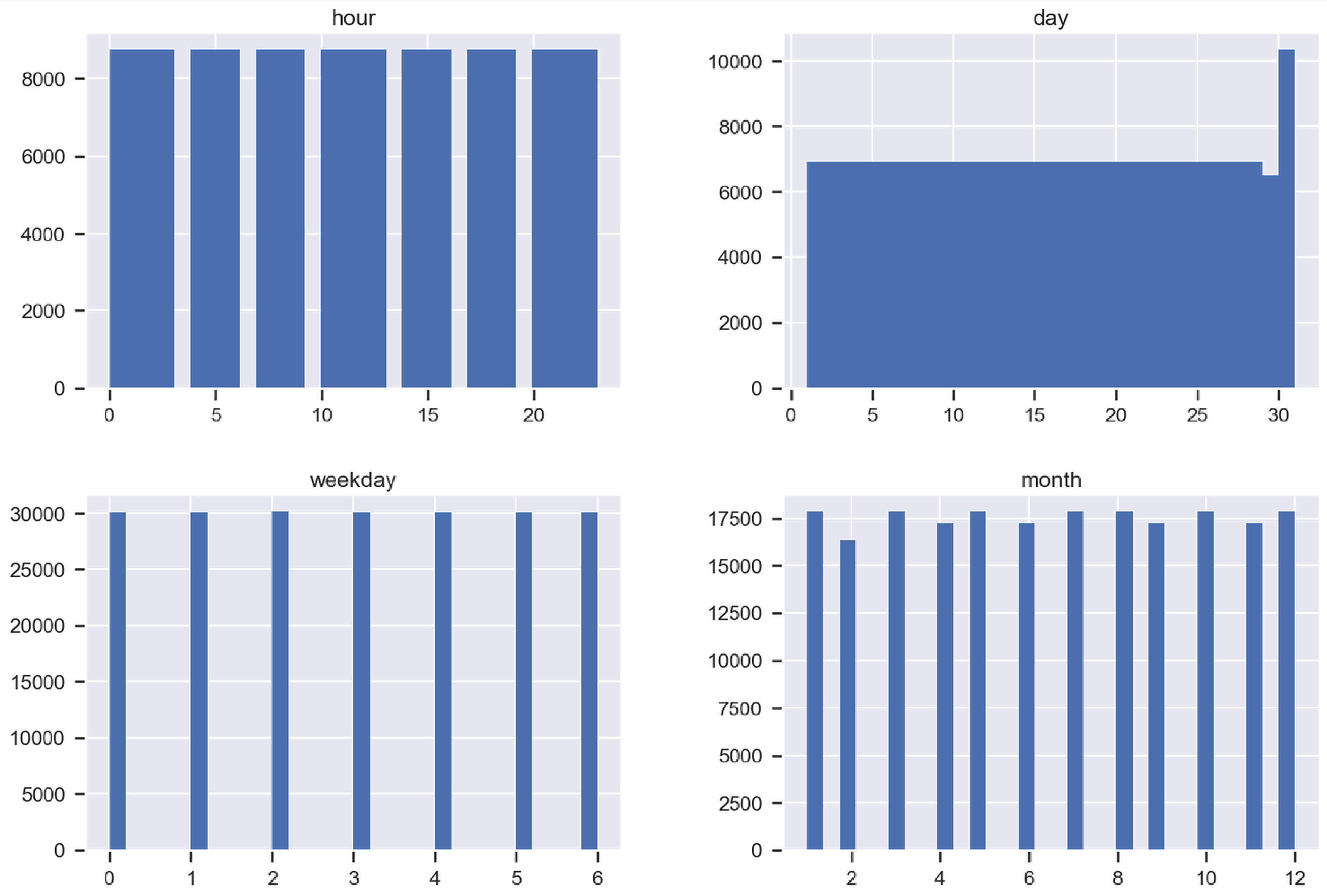
DATA SET

- energy data from 2024 and 2025
- Data was sourced from Fraunhofer-Institut für Solare Energiesysteme ISE
- High seasonal component
- Day Ahead Auction (DE-LU), bids until 12 noon, published at around 12:40 p.m.

- Hydro pumped storage consumption
- Cross border electricity trading
- Hydro Run-of-River
- Biomass
- Fossil brown coal / lignite
- Fossil hard coal
- Fossil oil
- Fossil coal-derived gas
- Fossil gas
- Geothermal
- Hydro water reservoir
- Hydro pumped storage
- Others
- Waste
- Wind offshore
- Wind onshore
- Solar
- Load
- Residual load
- Renewable share of load
- Renewable share of generation
- **Day Ahead Auction (DE-LU) - target value**

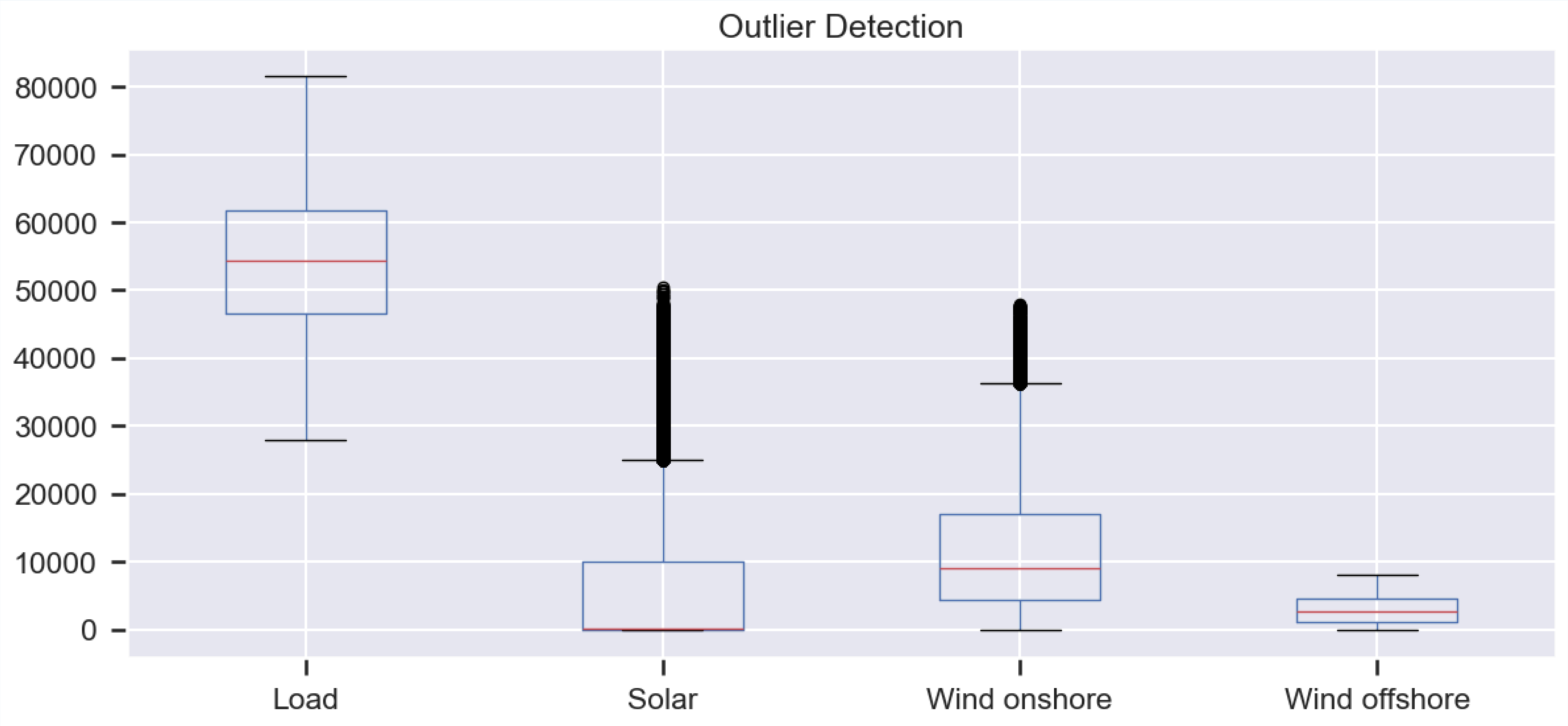


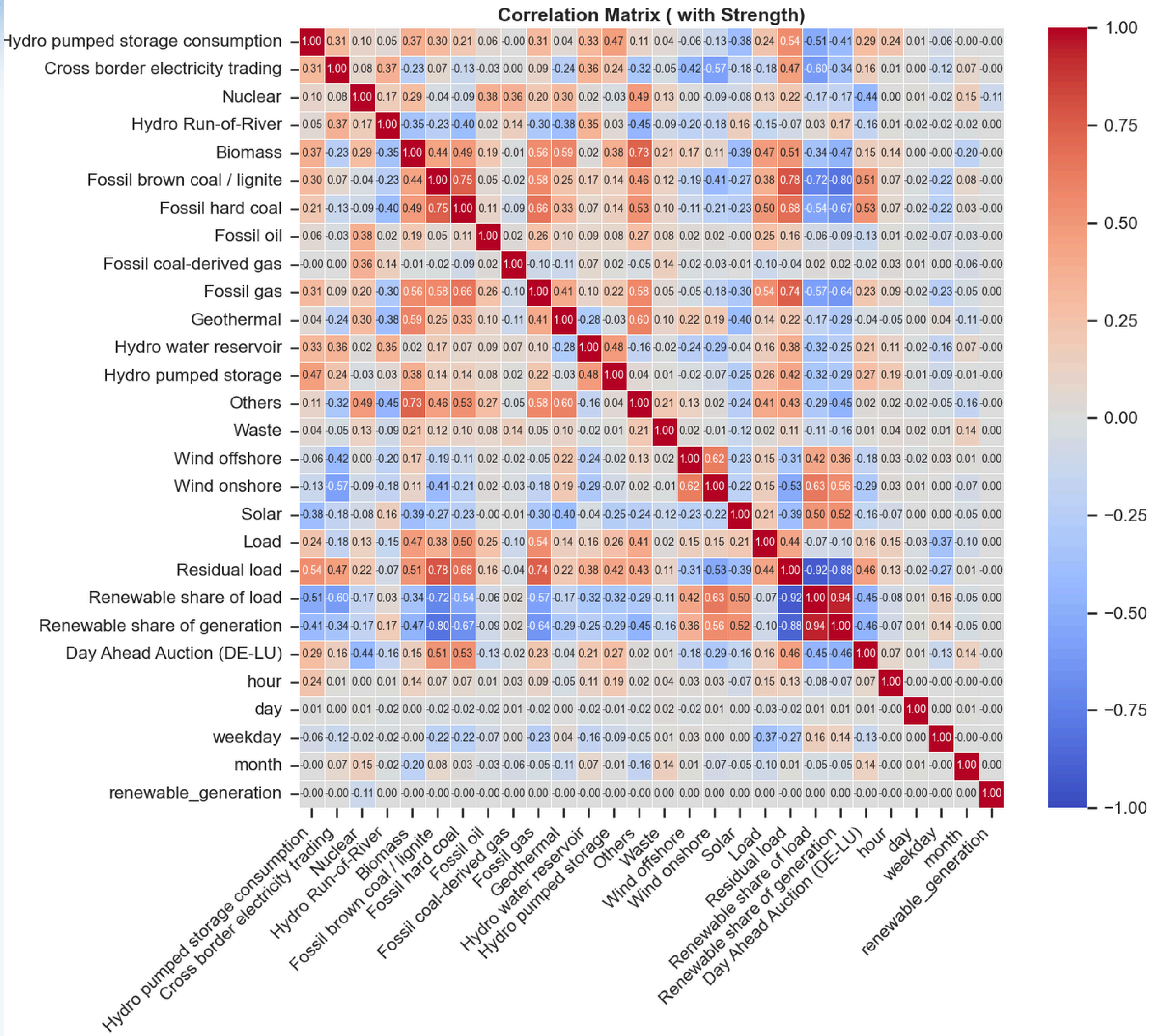
HISTOGRAMS



hour
8000
6000
4000
2000
0
0
5
10
15
20
day
10000
8000
6000
4000
2000
0
0
5
10
15
20
25
30
weekday
30000
25000
20000
15000
10000
5000
0
0
1
2
3
4
5
6
month
17500
15000
12500
10000
7500
5000
2500
0
2
4
6
8
10
12

BOX PLOT (OUTLINERS)

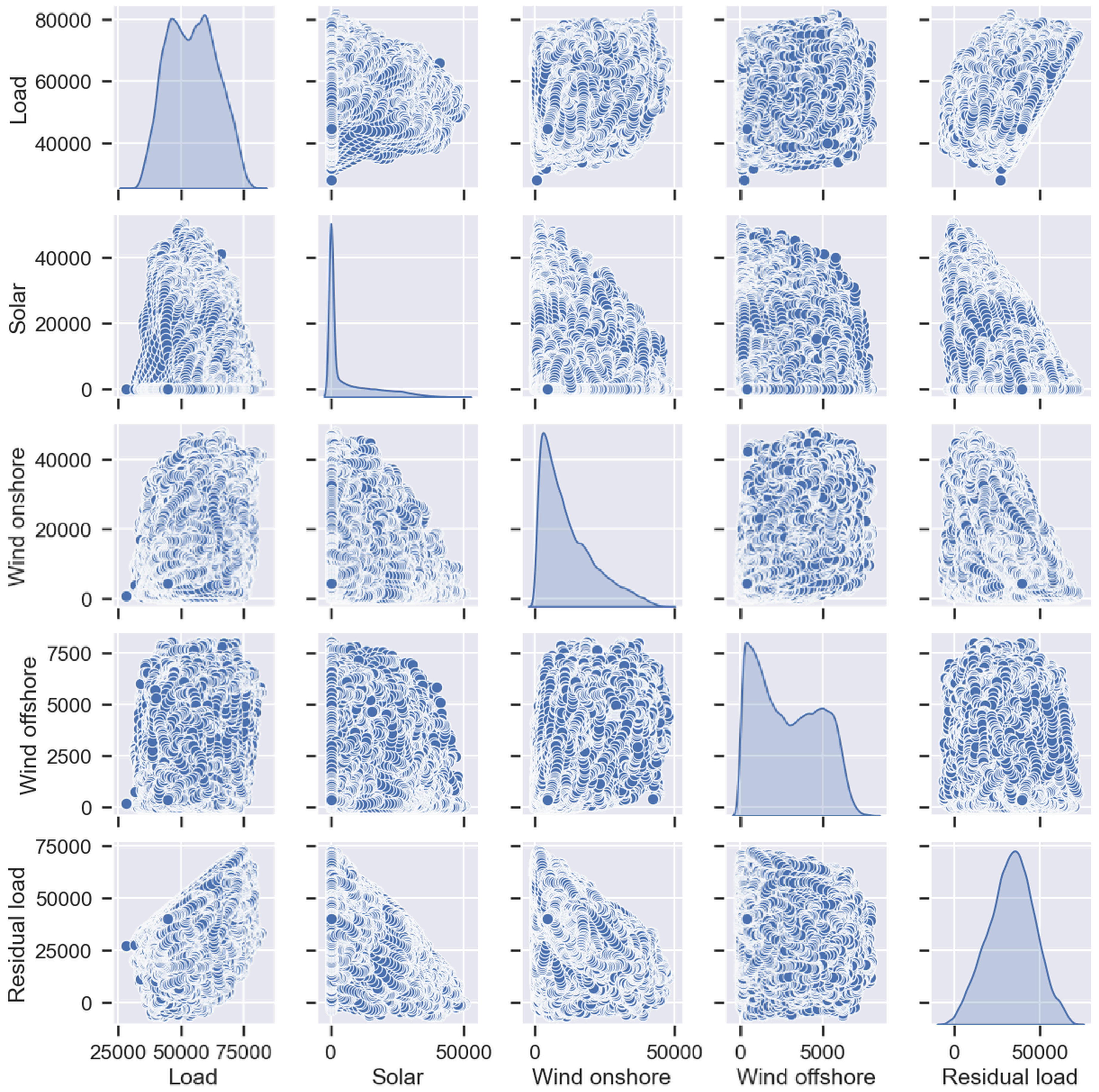


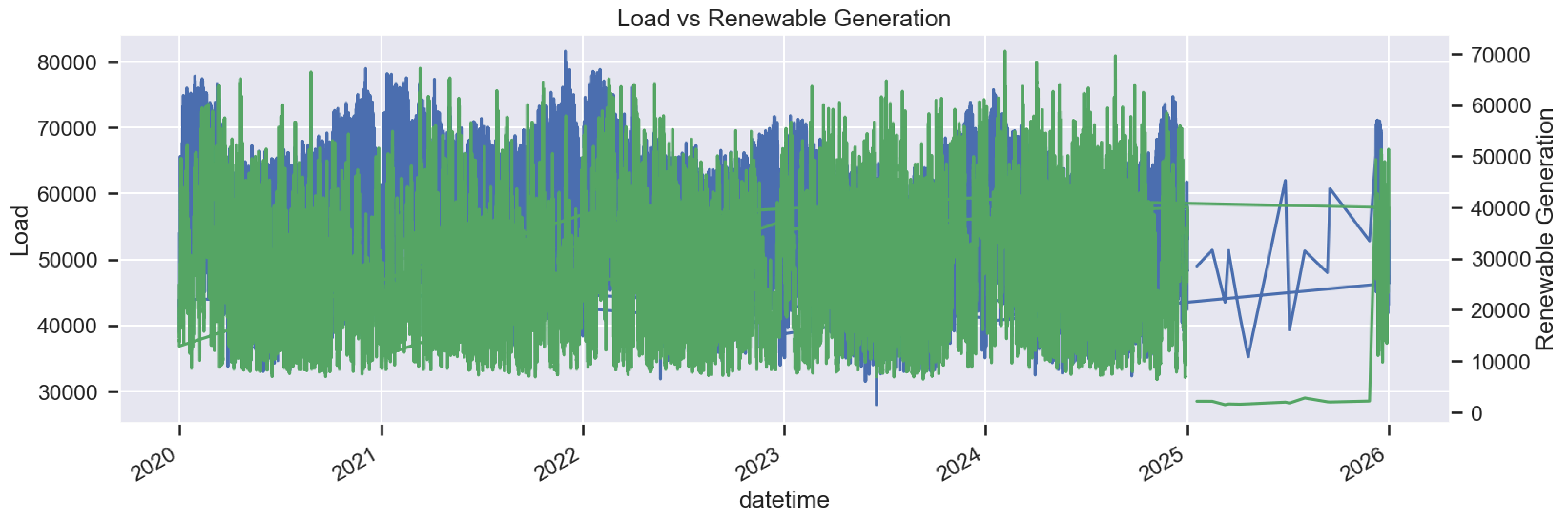


TARGET VALUE

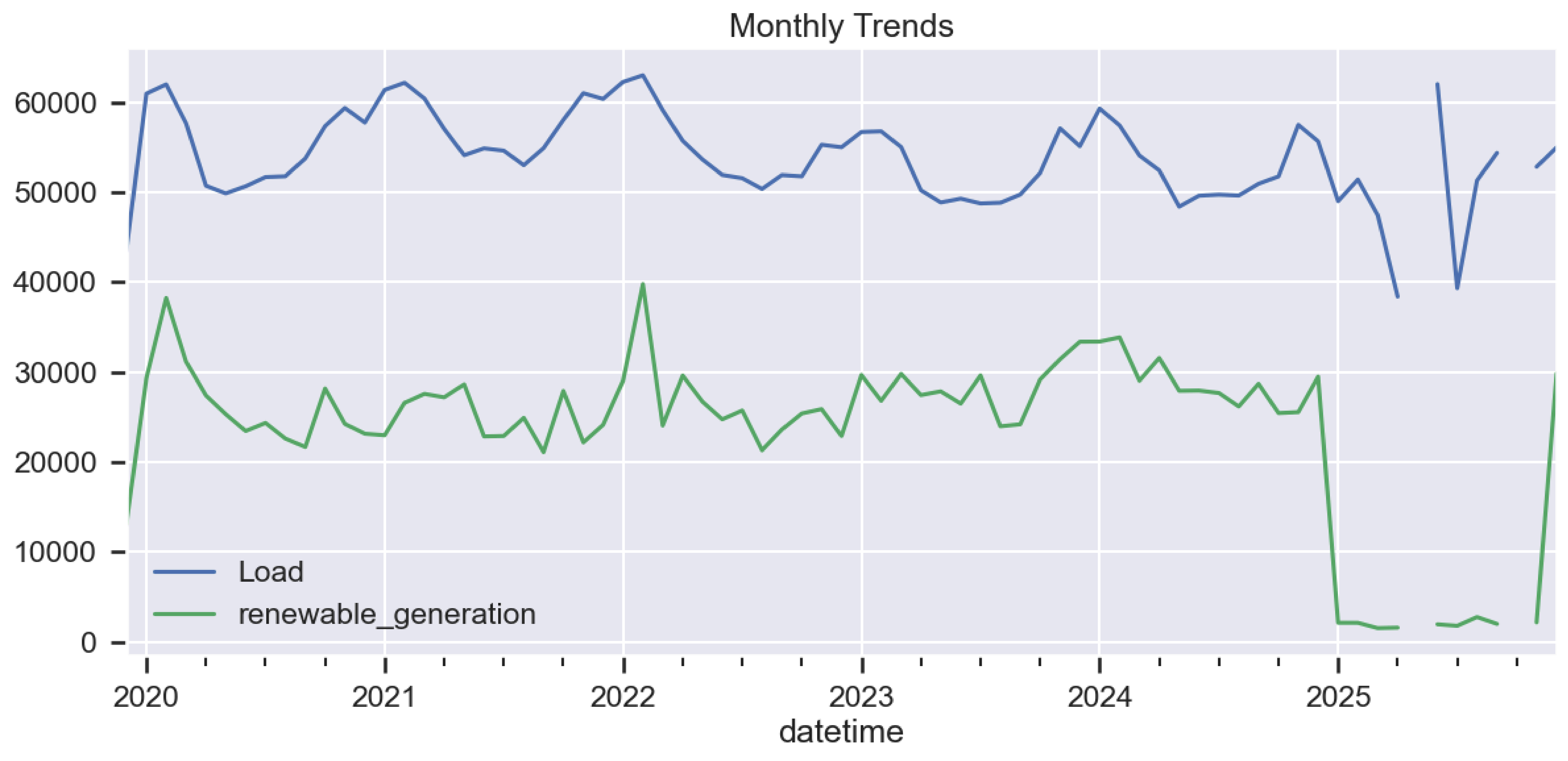
```
Day Ahead Auction (DE-LU)            1.000000
Residual load                        0.742104
Fossil brown coal / lignite          0.596684
Fossil gas                           0.545497
Hydro pumped storage consumption     0.544739
Fossil hard coal                     0.516677
Cross border electricity trading     0.515919
Hydro pumped storage                 0.484154
Biomass                              0.478509
Hydro water reservoir                0.381108
Others                               0.309137
Load                                 0.274796
Geothermal                           0.174299
Fossil oil                           0.118479
Waste                                0.020347
Hydro Run-of-River                  -0.021132
Fossil coal-derived gas             -0.036501
Wind offshore                       -0.144816
Wind onshore                        -0.296296
Solar                               -0.425950
Renewable share of generation       -0.683314
Renewable share of load             -0.695510
```

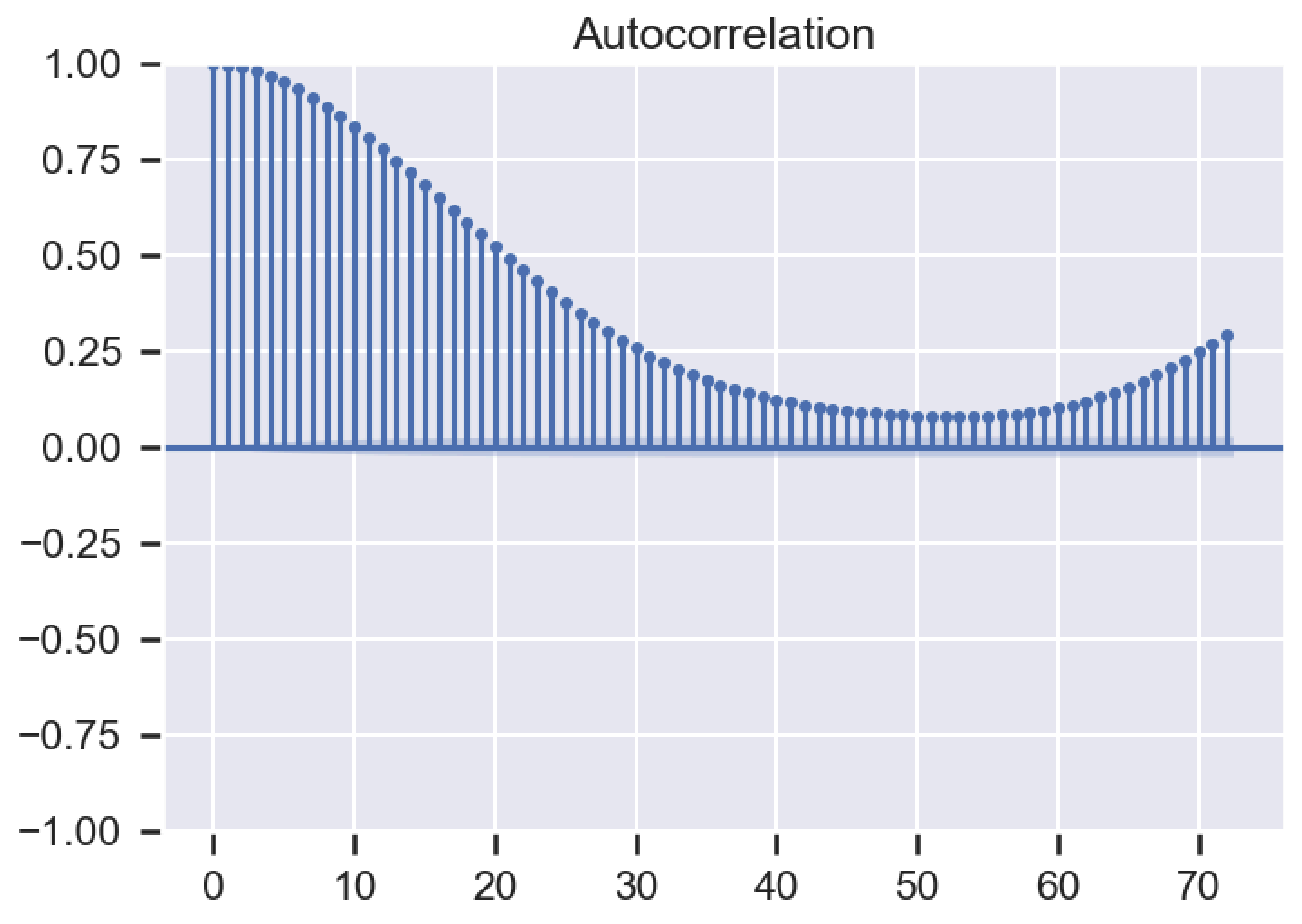
KEY PAIR RELATIONSHIP

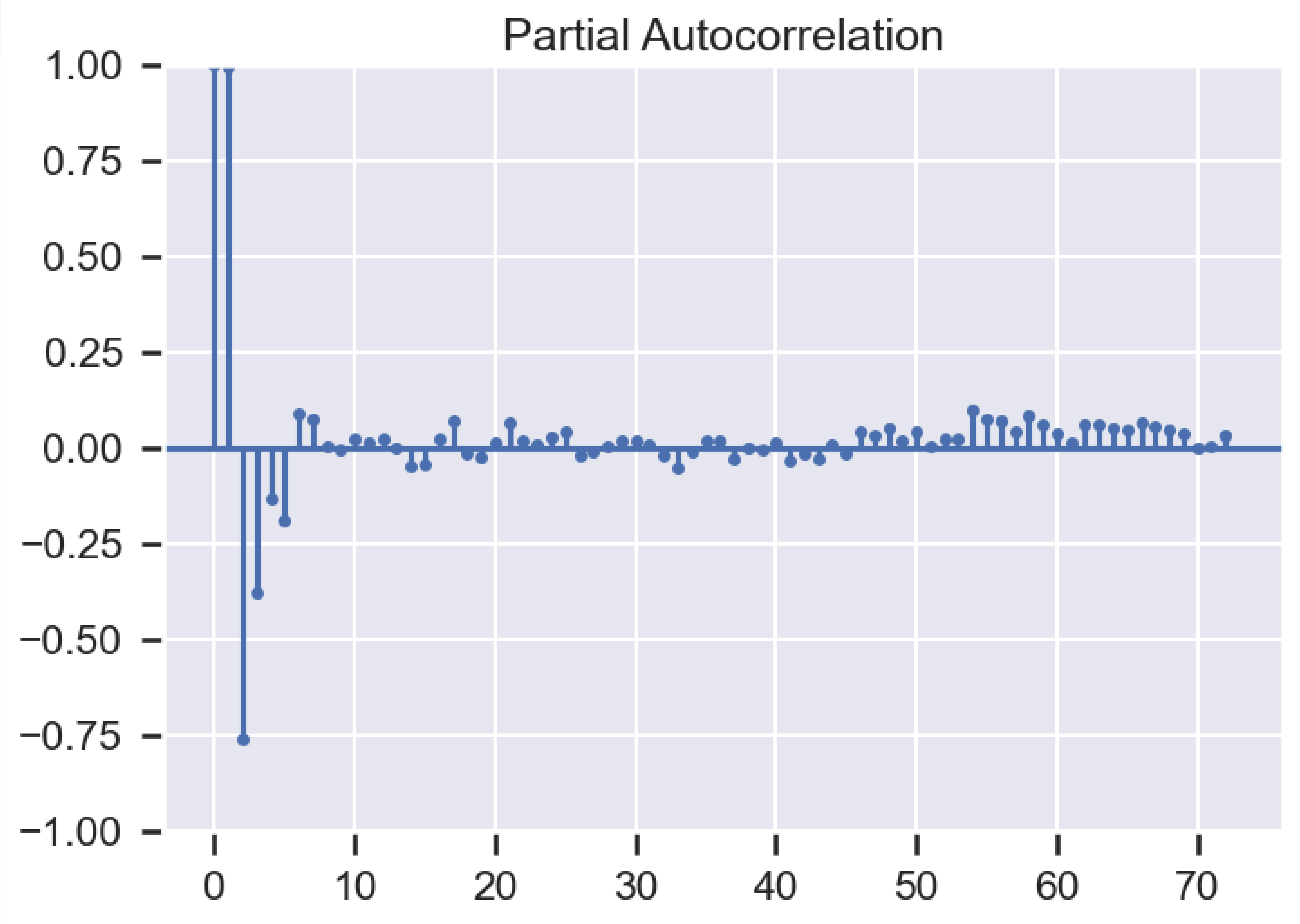




Load vs Renewable Generation
Load
80000
70000
60000
50000
40000
30000
Renewable Generation
70000
60000
50000
40000
30000
20000
10000
0
2020
2021
2022
2023
2024
2025
2026
datetime


Monthly Trends
60000
50000
40000
30000
20000
10000
0
Load
renewable_generation
2020
2021
2022
2023
2024
2025
datetime


Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50
60
70


Partial Autocorrelation
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0
10
20
30
40
50
60
70

CLUSTERING

renewable_generation
50000
45000
40000
35000
30000
25000
20000
15000
10000
40000
45000
50000
55000
60000
Load
cluster
0
1
2
3

BASELINE MODELS

We have used SARIMAX as our baseline model

- **SARIMAX:** is a statistical time-series model that extends ARIMA by adding seasonality and external explanatory variables (exogenousfeatures). It captures both past patterns in the target series and the influence of related factors such as renewable share or load.

The SARIMAX model will serve as our foundational benchmark for the day ahead price prediction.

SARIMAX MODEL

Data Preparation

- All values were converted to the correct format
- Gaps were filled
- And the index was set

Parameter Selection

- The non-seasonal parameters (p,d,q) were chosen using ACF and PACF plots together with stationarity checks (e.g., Augmented Dickey–Fuller test)
- The seasonal parameters (P,D,Q,s) were selected based on the known data frequency (15-minute intervals → seasonal period such as 96 for one day).

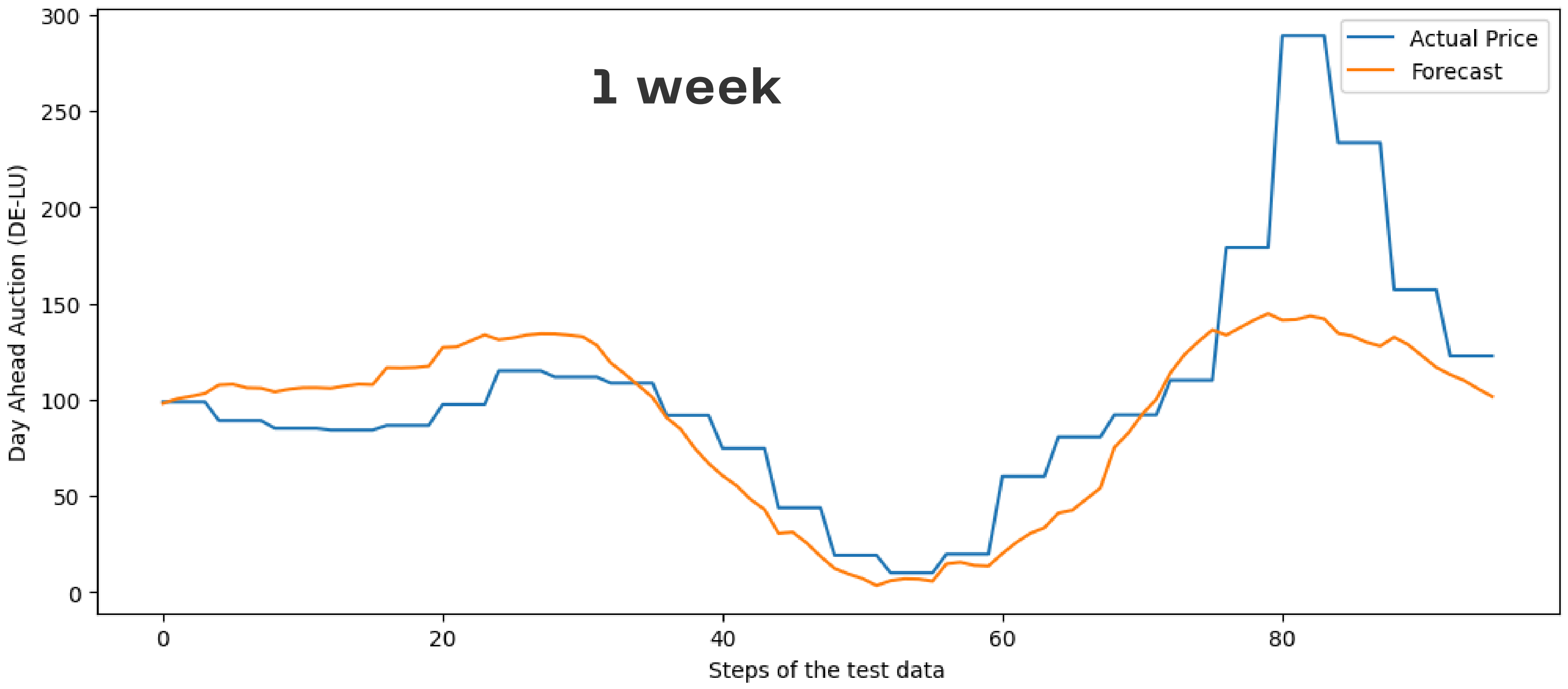
Model Fitting

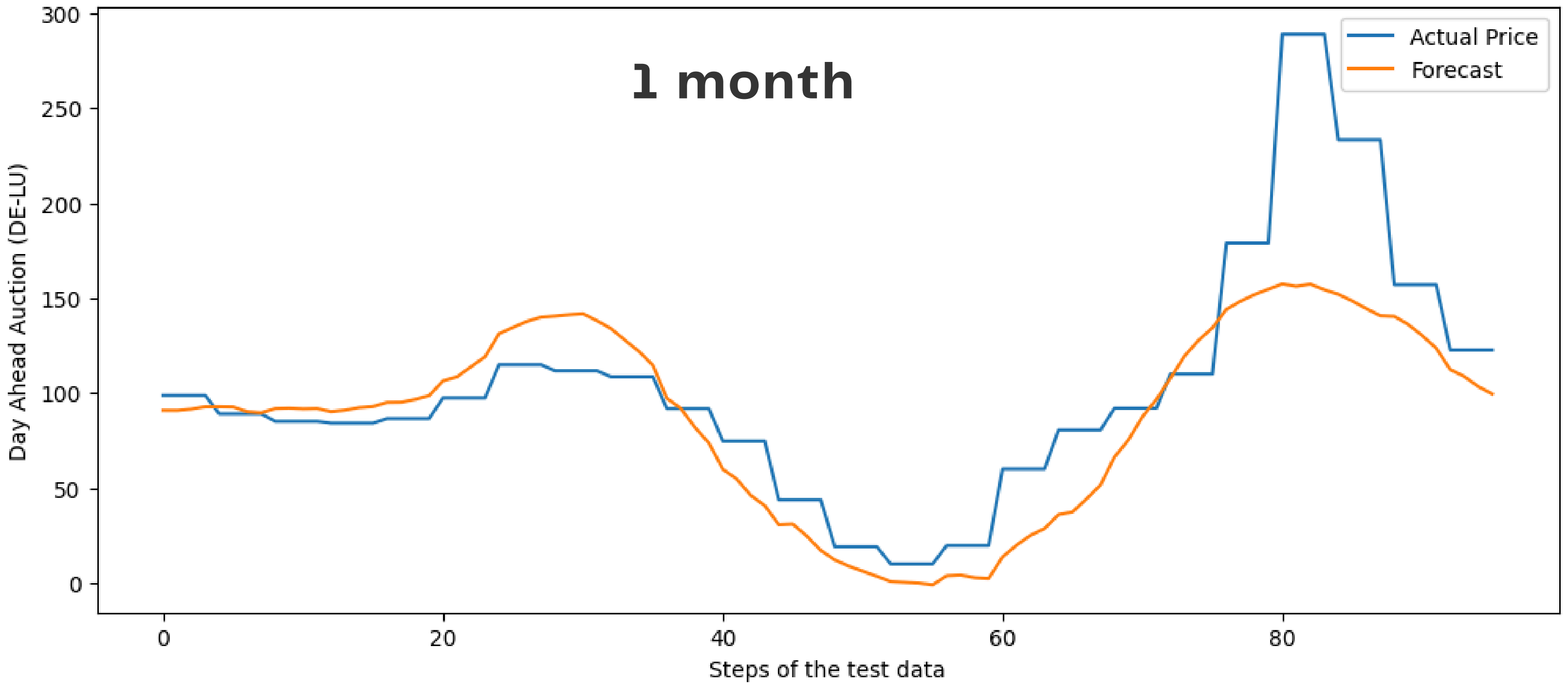
- We took three periods, one week, two weeks, and one month, as training data due to the long training duration.
- As exogenous features, we took the 4 features with the highest correlation.

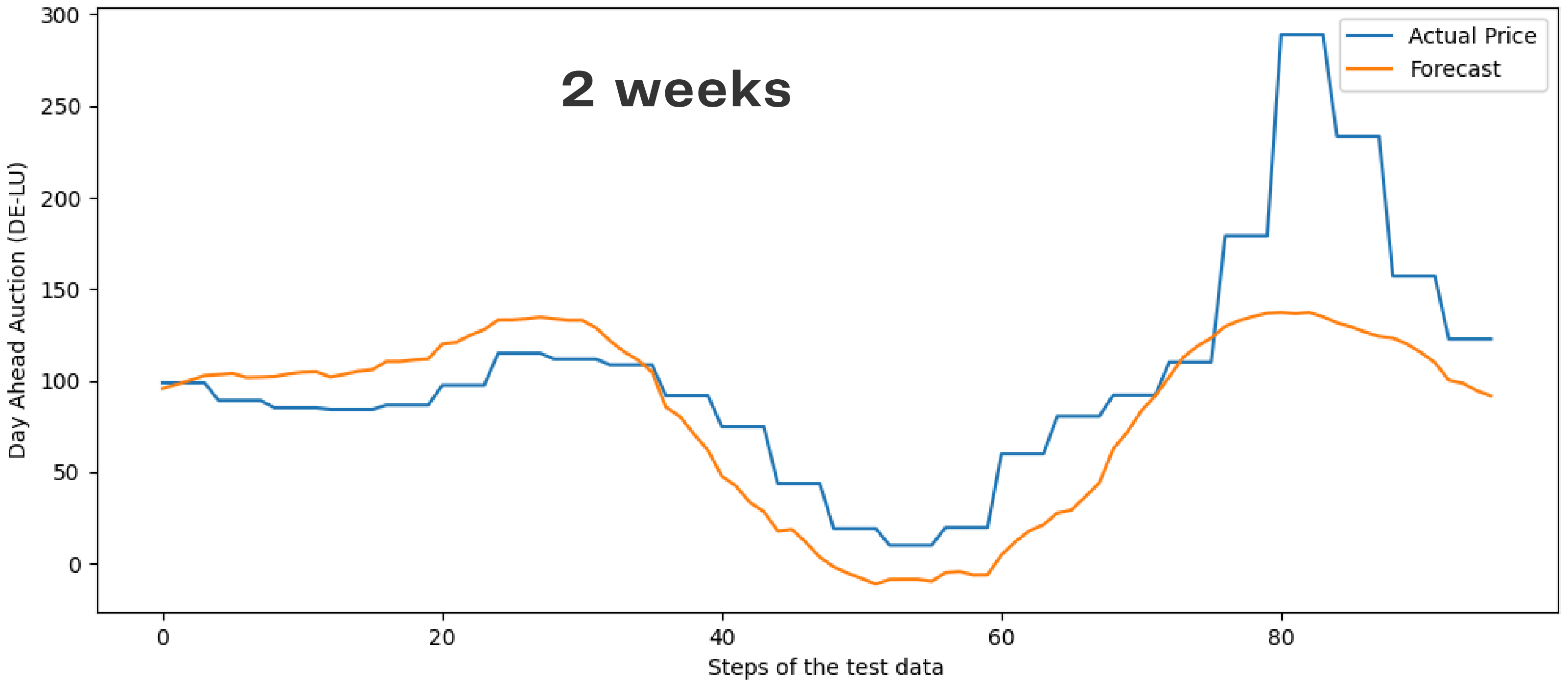
Evaluation

- measured the prediction error using MSE and RMSE
- RMSE is particularly intuitive here, because it represents the average deviation of the forecast from the actual energy price in EUR/MWh

SARIMAX MODEL







SARIMAX MODEL

A RMSE of 80.5 means 80.5 EUR/MWh wrong in avarage
Range: 25-300 EUR/MWh

Test data 1 week
MSE: 1790.91
RMSE: 42.31

Test data 2 weeks
MSE: 2126.99
RMSE: 46.11

Test data 4 weeks
MSE: 1424.25
RMSE: 37.73

An MSE: 1425.25 and a RMSW: 37.73 will be the baseline

TREE-BASED MODELS

- XGBoost
- LightGBM

XGBOOST

Data Preprocessing

- Feature engineering: ---->
- Target defined: 1 day ahead
- Normalization: Applied RobustScaler to reduce the impact of outliers

XGBoost Model Architecture

- Input: feature correlation >= 0.3
- Target: 1-day forward return (96x 15min intervalls)
- TimeSeriesSplit with 3 folds (preserves temporal order)
- RobustScaler for outlier-resistant normalization
- Hyperparameters: 500 trees, LR = 0.05, max depth = 6

XGBoost Model Architecture no tuning

- Input: feature correlation >= 0.3
- Target: 1-day forward return (96x 15min intervalls)
- same train and test data

Features

Residual load,
Fossil brown coal / lignite,
Renewable share of load,
Renewable share of generation,
....

LIGHTGBM

Data Preprocessing

- Applied same preprocessing pipeline as XGBoost

LightGBM Model Architecture

- default parameters are used

TREE-BASED MODELS

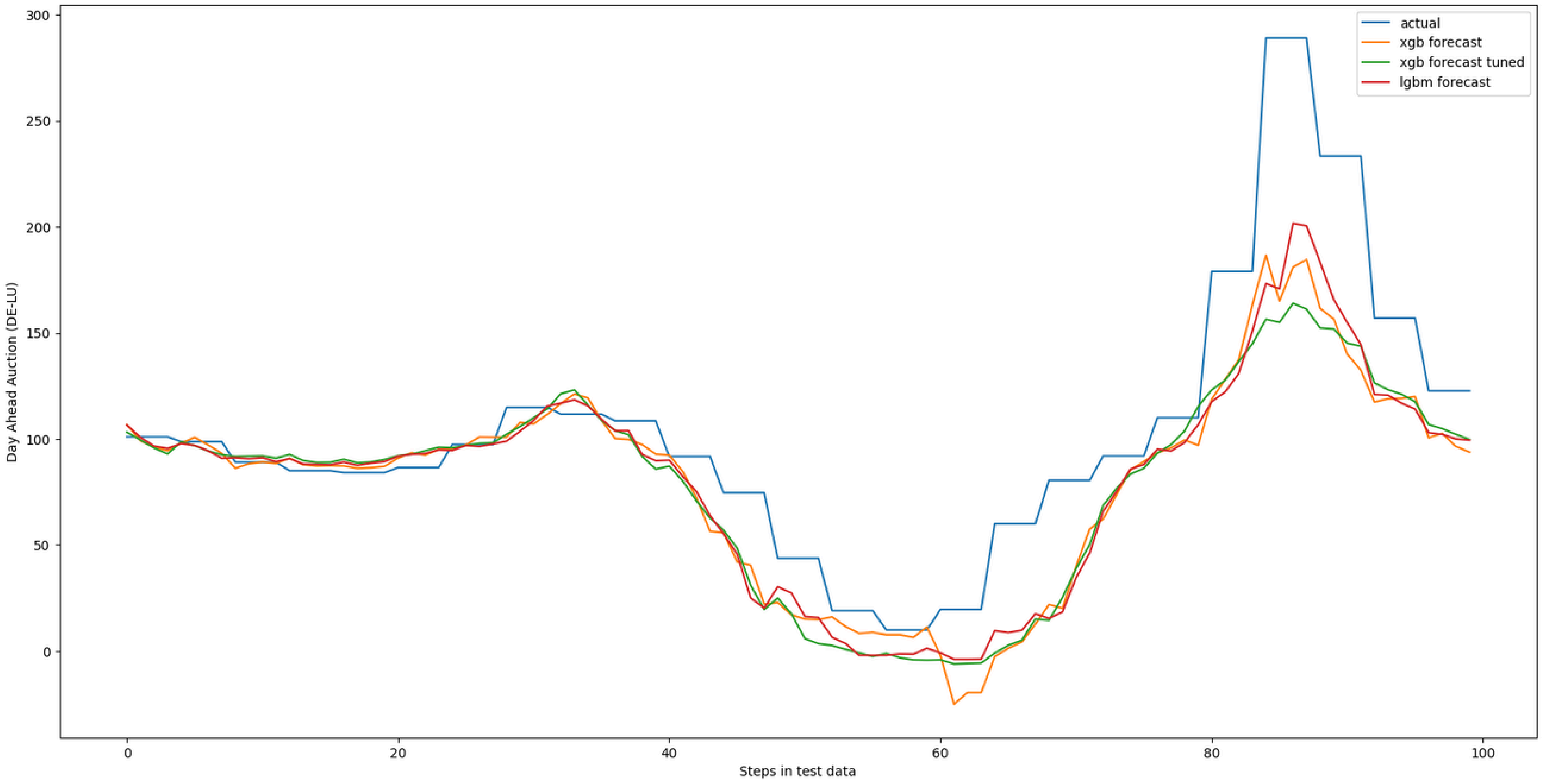
XGBoost
MSE: 1358.92
RMSE: 36.86

XGBoost tuned
MSE: 1532.83
RMSE: 39.151

LightGBM
MSE: 1645.50
RMSE: 40.56

Baseline Model
MSE: 1424.25
RMSE: 37.73

XGBoost and LightGBM run much faster
Best one so far XGBoost



RECURRENT NEURAL NETWORKS

- LSTM (Long Short-Term Memory)

LSTM/GRU

- Input: sliding windows shaped as (look_back=5, n_features=23)
- Recurrent layer: LSTM with 128 units and swish activation
- Output layer: Dense(1) for single-step forecasting
- Loss function: Mean Squared Error (MSE)
- Optimizer: Adam
- Metric: Mean Absolute Percentage Error (MAPE)
- Training: 50 epochs, batch size 100

LSTM/GRU

Comparative Analysis

Baseline Model (SARIMAX)
MSE: 1424.25
RMSE: 37.73

XGBoost (Best so far)
MSE: 1358.92
RMSE: 36.86

LSTM_GRU
MSE: 331.04
RMSE: 18.19

The LSTM achieves the best performance overall, with lower errors than both SARIMAX and XGBoost. Its MSE (331) and RMSE (18.19) show that it captures the temporal patterns in the data far more effectively, making it the strongest forecasting model in this comparison.

SPECIALIZED TIME SERIES MODELS

- N-Beats Model
- TFT (Temporal Fusion Transformers)

N-BEATS

Data Preprocessing

- Used all numeric columns as multivariate input features
- Applied a sliding window: 5-step look-back → 1-step ahead forecast
- Created supervised samples using a custom window_multivariate function
- Flattened each window into a single vector for the fully-connected N-BEATS model
- Performed an 80/20 time-ordered train–test split

Model Architecture

- Input: flattened window of shape (look_back=5 × n_features=23)
- Stacks: 3 sequential N-BEATS blocks
- Each block:
- 4 fully-connected layers (256 units, ReLU)
- Linear theta layer → produces block parameters
- Backcast head: reconstructs input residual
- Forecast head: predicts 1-step ahead value
- Residuals updated after each block; forecasts summed across stacks
- Loss: MSE
- Optimizer: Adam (learning rate 1e-3)
- Training: 50 epochs, batch size 128, 10% validation split

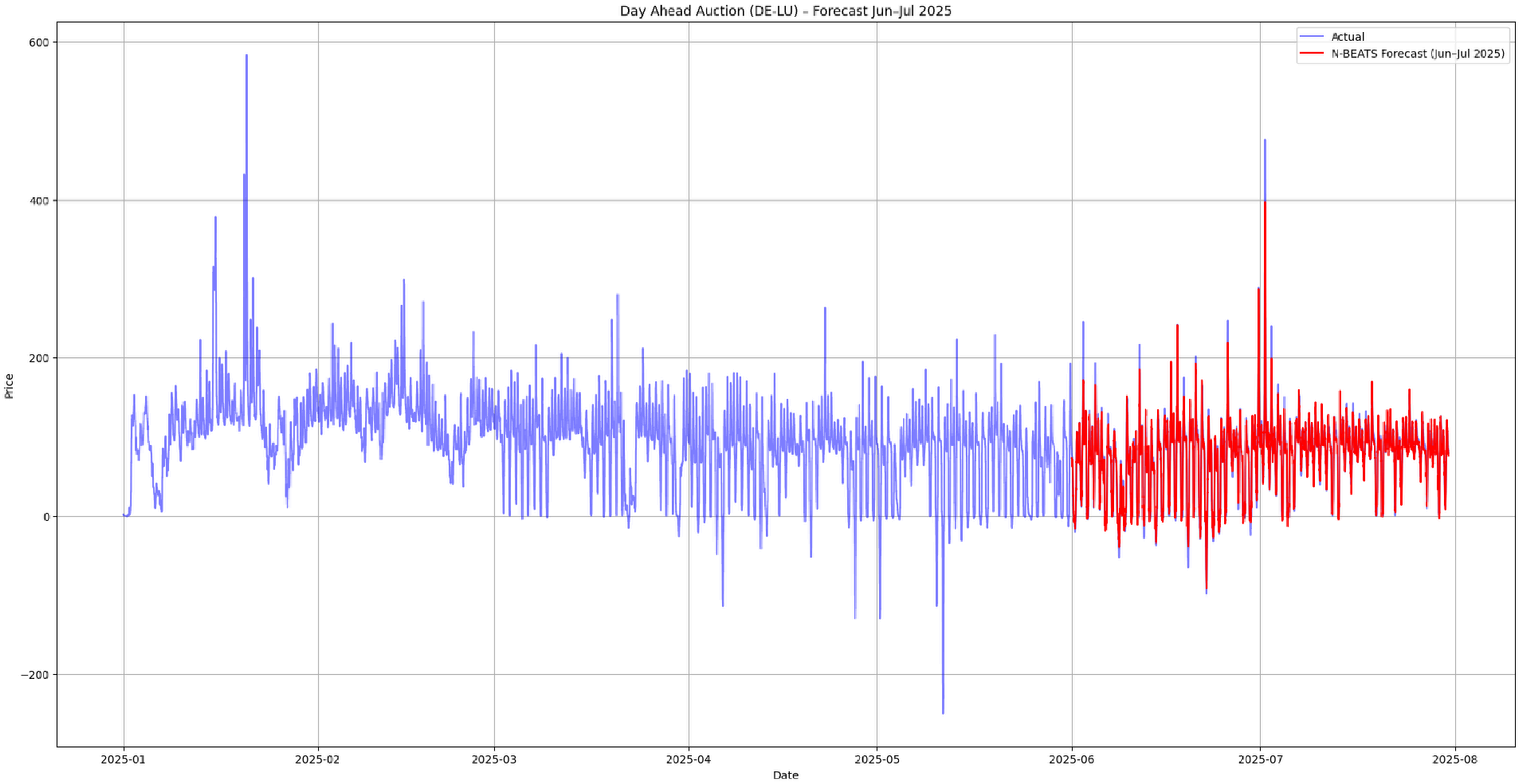
N-BEATS

Comparative Analysis

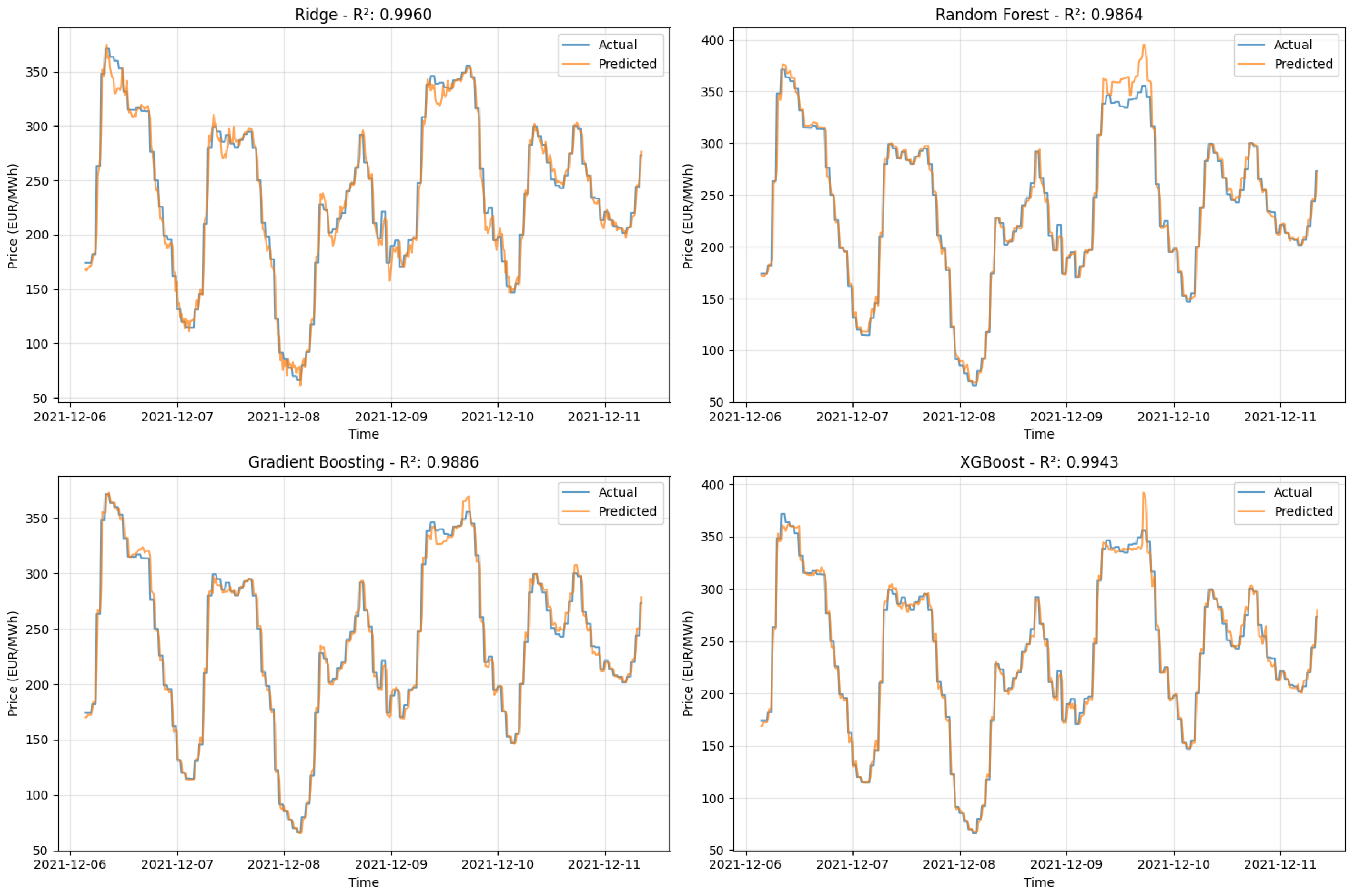
Baseline Model (SARIMAX)
MSE: 1424.25
RMSE: 37.73

LSTM_GRU (Best so far)
MSE: 331.04
RMSE: 18.19

N-Beats
MSE: 3900.33
RMSE: 62.45



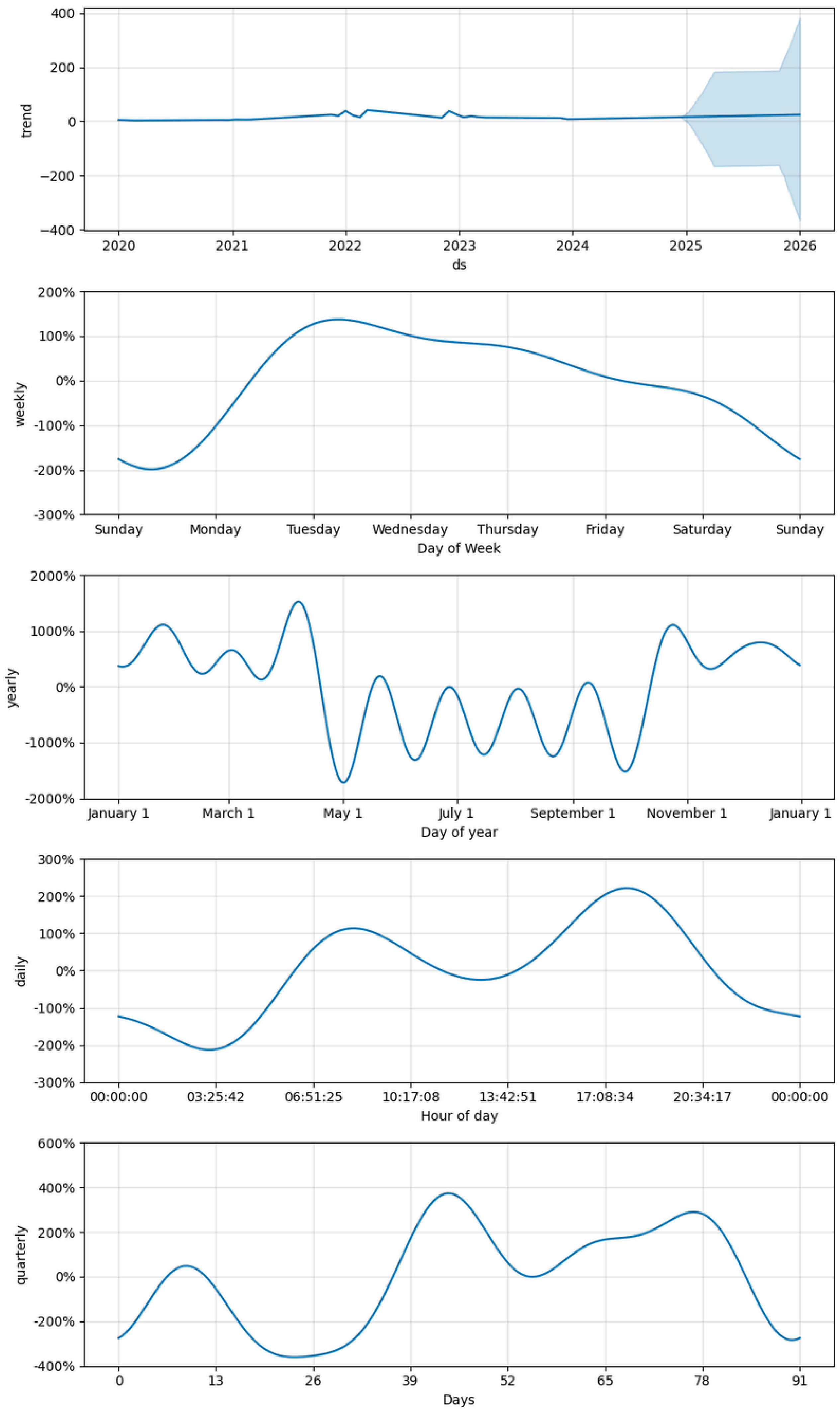
The LSTM achieves still the best performance.



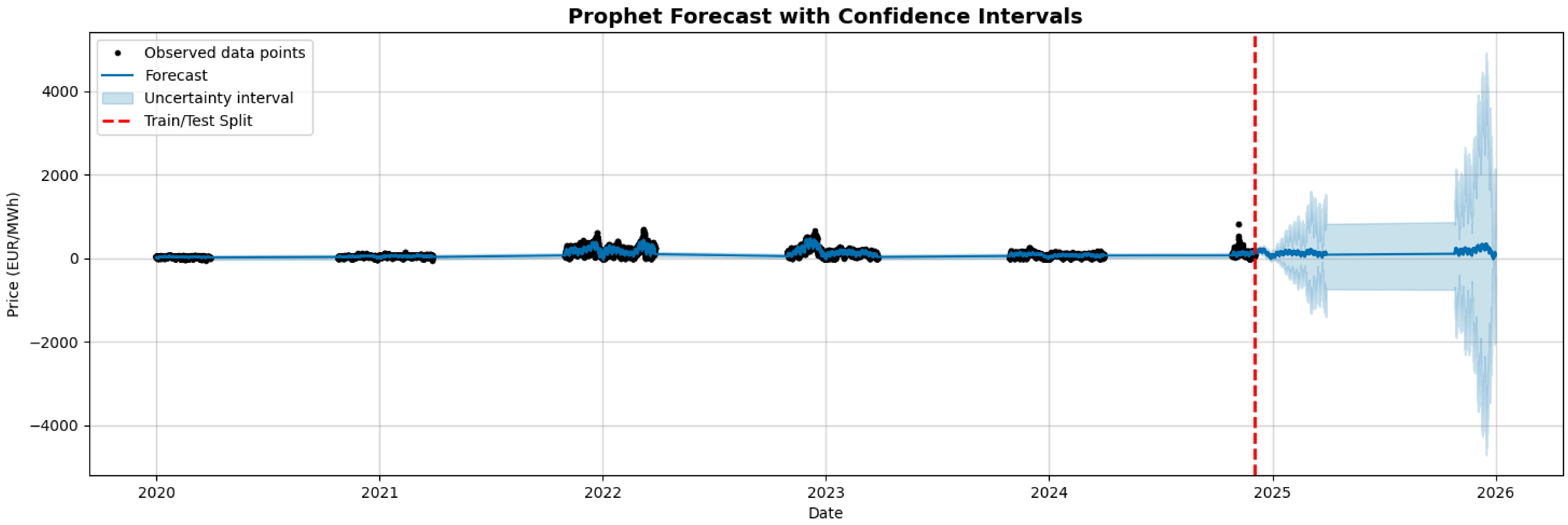

```
==================================================
MODEL COMPARISON
==================================================
```

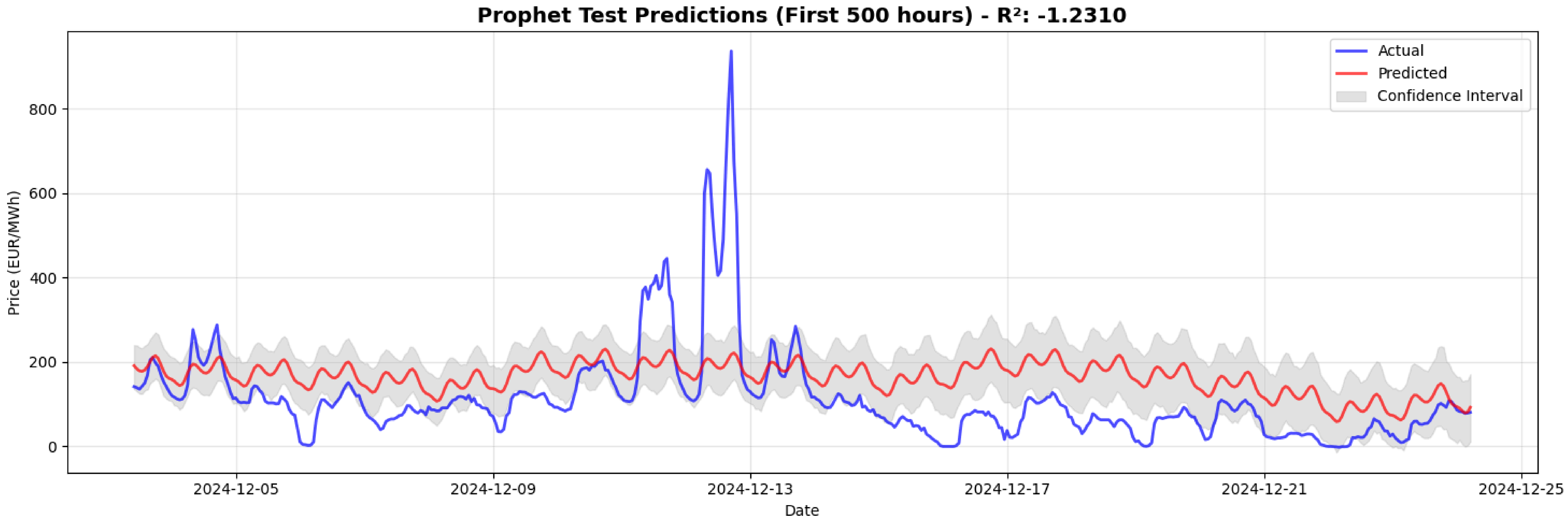
	MAE	RMSE	R2
Ridge	2.866442	5.376392	0.995991
XGBoost	3.246775	6.403631	0.994313
Gradient Boosting	4.955492	9.079562	0.988568
Random Forest	5.669502	9.897852	0.986414

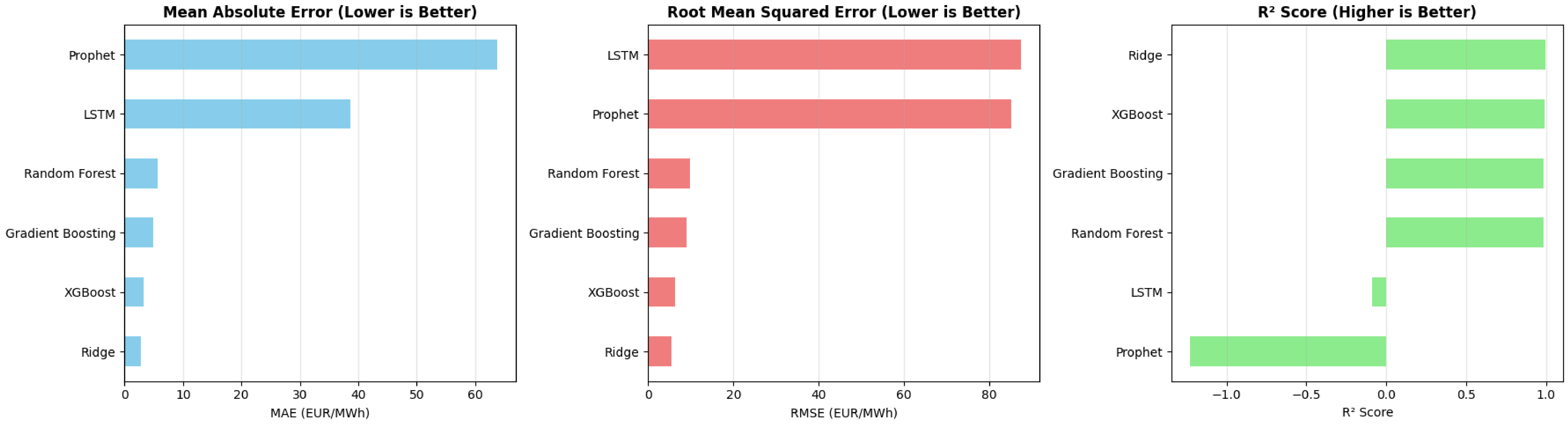
PROPHET



trend
400
200
0
−200
−400
2020
2021
2022
2023
2024
2025
2026
ds
weekly
200%
100%
0%
-100%
-200%
-300%
Sunday
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of Week
yearly
2000%
1000%
0%
-1000%
-2000%
January 1
March 1
May 1
July 1
September 1
November 1
January 1
Day of year
daily
300%
200%
100%
0%
-100%
-200%
-300%
00:00:00
03:25:42
06:51:25
10:17:08
13:42:51
17:08:34
20:34:17
00:00:00
Hour of day
quarterly
600%
400%
200%
0%
-200%
-400%
0
13
26
39
52
65
78
91
Days


Prophet Forecast with Confidence Intervals
Observed data points
Forecast
Uncertainty interval
Train/Test Split
Price (EUR/MWh)
4000
2000
0
−2000
−4000
2020
2021
2022
2023
2024
2025
2026
Date


Prophet Test Predictions (First 500 hours) - R²: -1.2310
Actual
Predicted
Confidence Interval
Price (EUR/MWh)
800
600
400
200
0
2024-12-05
2024-12-09
2024-12-13
2024-12-17
2024-12-21
2024-12-25
Date


Mean Absolute Error (Lower is Better)
Prophet
LSTM
Random Forest
Gradient Boosting
XGBoost
Ridge
0 10 20 30 40 50 60
MAE (EUR/MWh)
Root Mean Squared Error (Lower is Better)
LSTM
Prophet
Random Forest
Gradient Boosting
XGBoost
Ridge
0 20 40 60 80
RMSE (EUR/MWh)
R² Score (Higher is Better)
Ridge
XGBoost
Gradient Boosting
Random Forest
LSTM
Prophet
−1.0 −0.5 0.0 0.5 1.0
R² Score

CONCLUSION

The LSTM/GRU model clearly outperforms all other approaches. While XGBoost and SARIMAX deliver similar mid-range errors, the LSTM/GRU cuts the RMSE roughly in half and achieves a strong R^2 of 0.88, showing it captures temporal patterns far more effectively. N-Beats performs worst in this setting, indicating a poor fit for the structure of this dataset. Overall, the LSTM/GRU stands out as the most accurate and reliable forecasting model in the comparison.

Model	**Strenght**	**Weakness**	**Total avg MSE**	**Total avg RMSE**
SARIMAX	Interpretable, seasonal,	manual tuning, poor nonlinear fit	1424.25	37.73
XGBoost	Robust for tabular data	Requires feature engineering for	1358.92	36.86
LightGBM	Fast than XGBoost, scalable	Similar limitations as XGBoost	1645.5	40.56
LSTM_GRU	Handles long sequences	Requires significant data & tuning	331.04	18.19
N-Beats	Learns patterns end-to-end	Computationally heavy	3900.33	62.45

THANK YOU