

Forecasting Energy price advanced time series prediction

Presented By:
Aadip Thapaliya and Helge Röpke

CONTENT

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

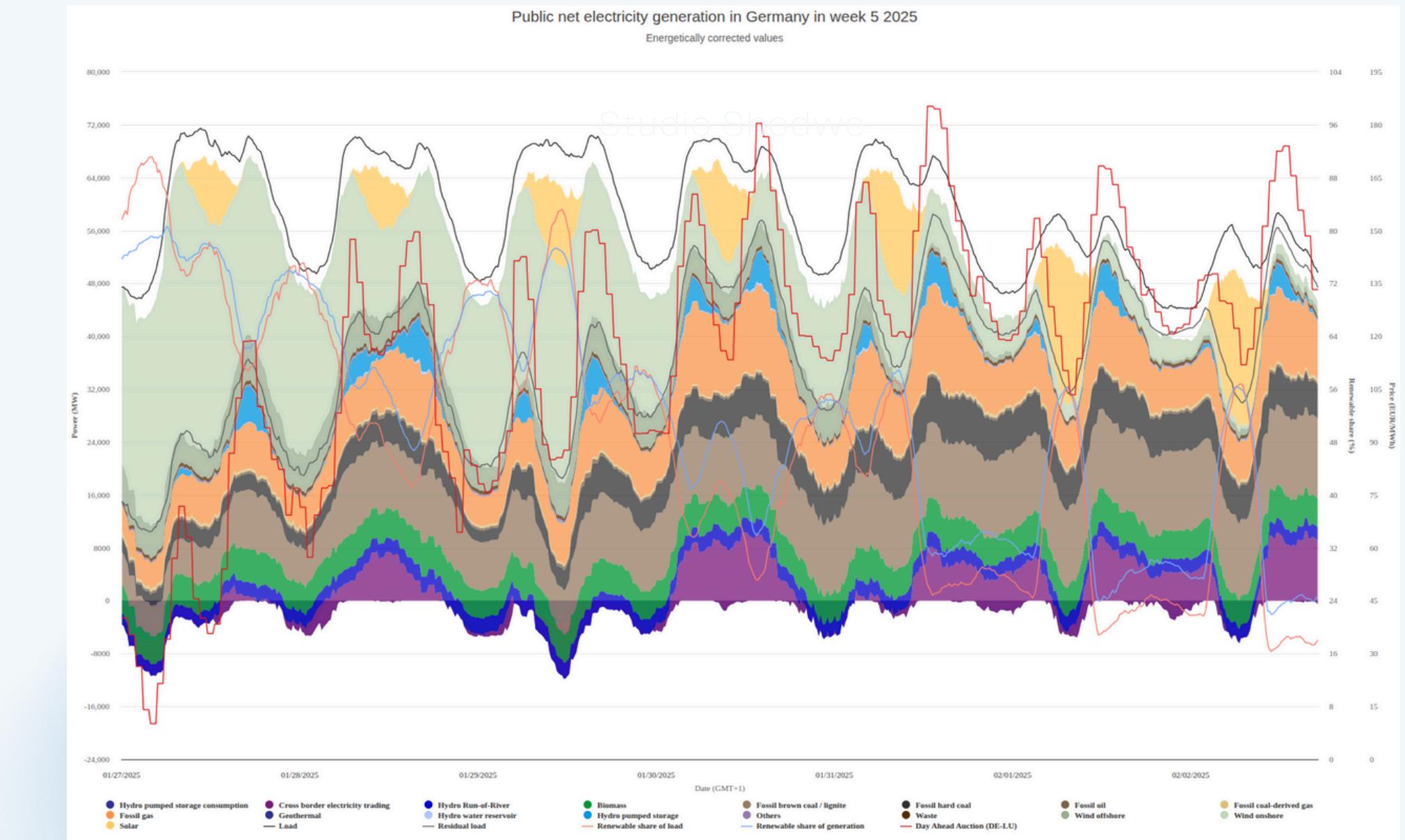
MOTIVATION

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**.

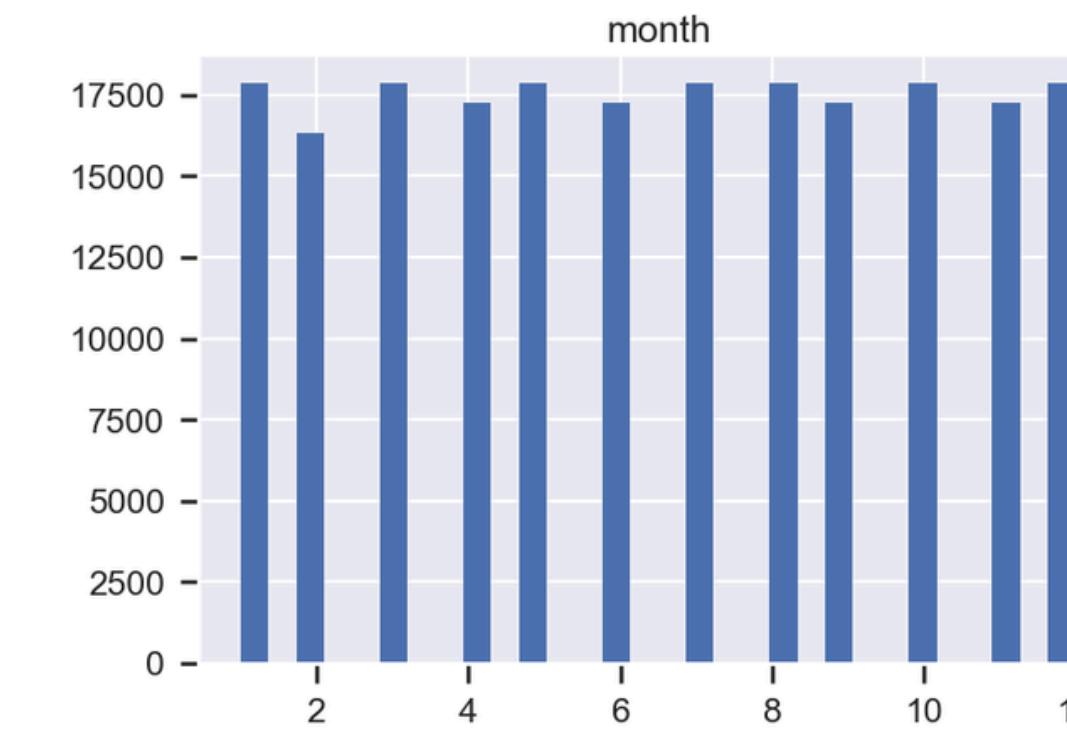
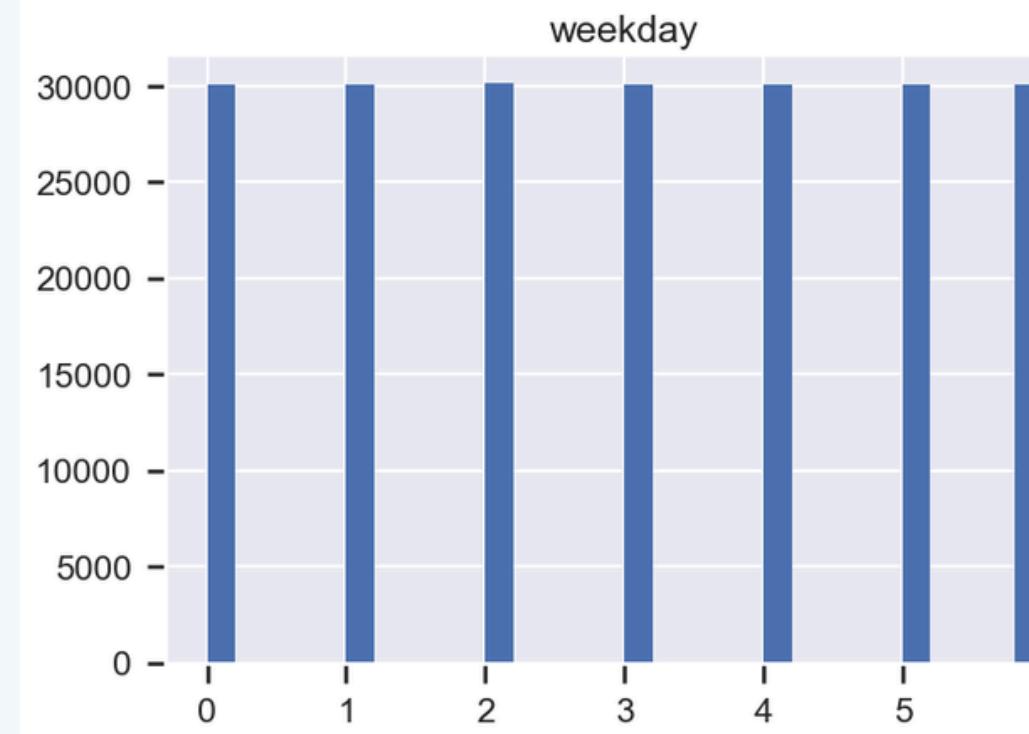
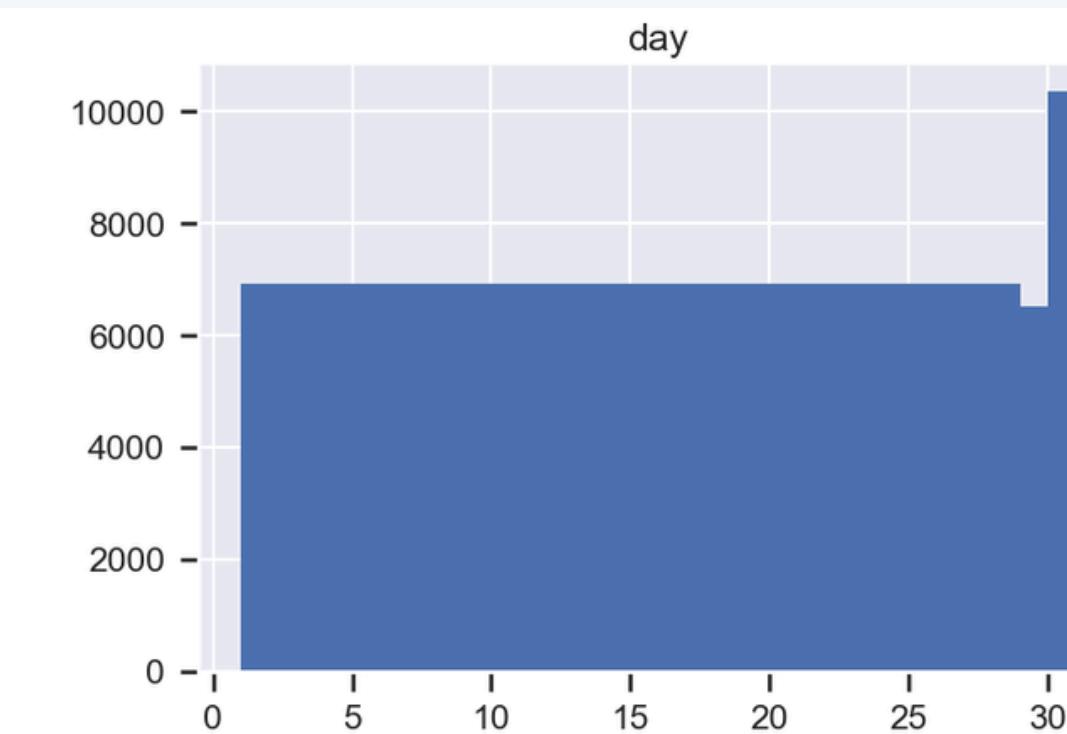
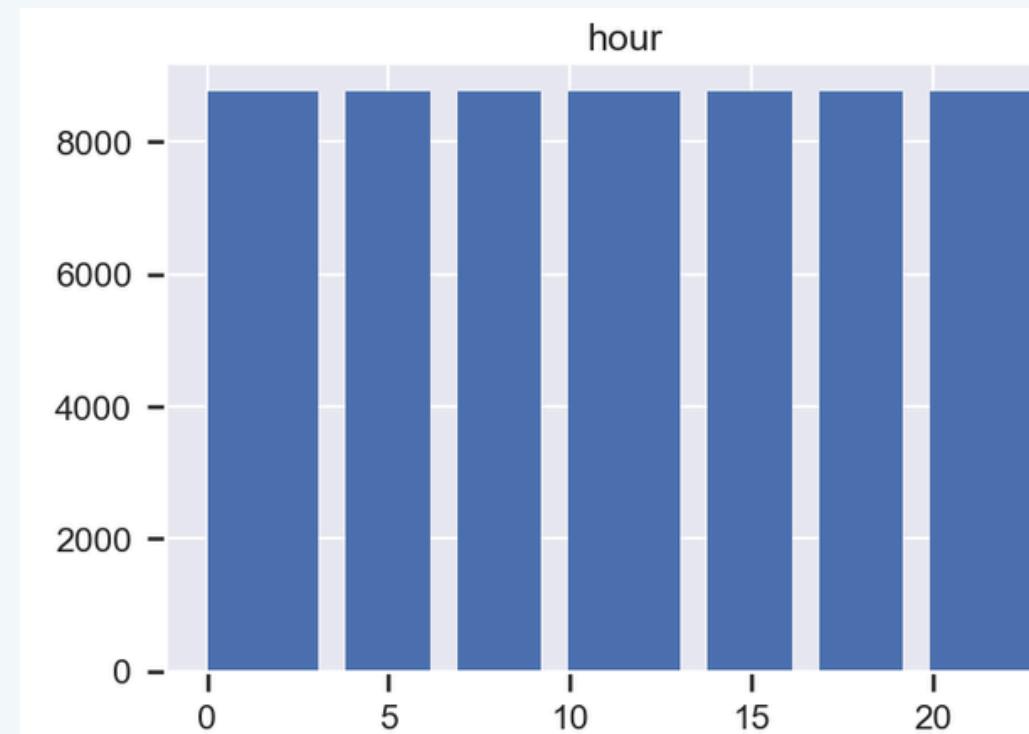
- Get practical experiencing for prediction time series analysis
- Compare classic and contemporary models across sectors
- uncover patterns driven by renewables, demand, and market behavior
- build a reliable forecasting system that supports smarter decisions

DATA SET

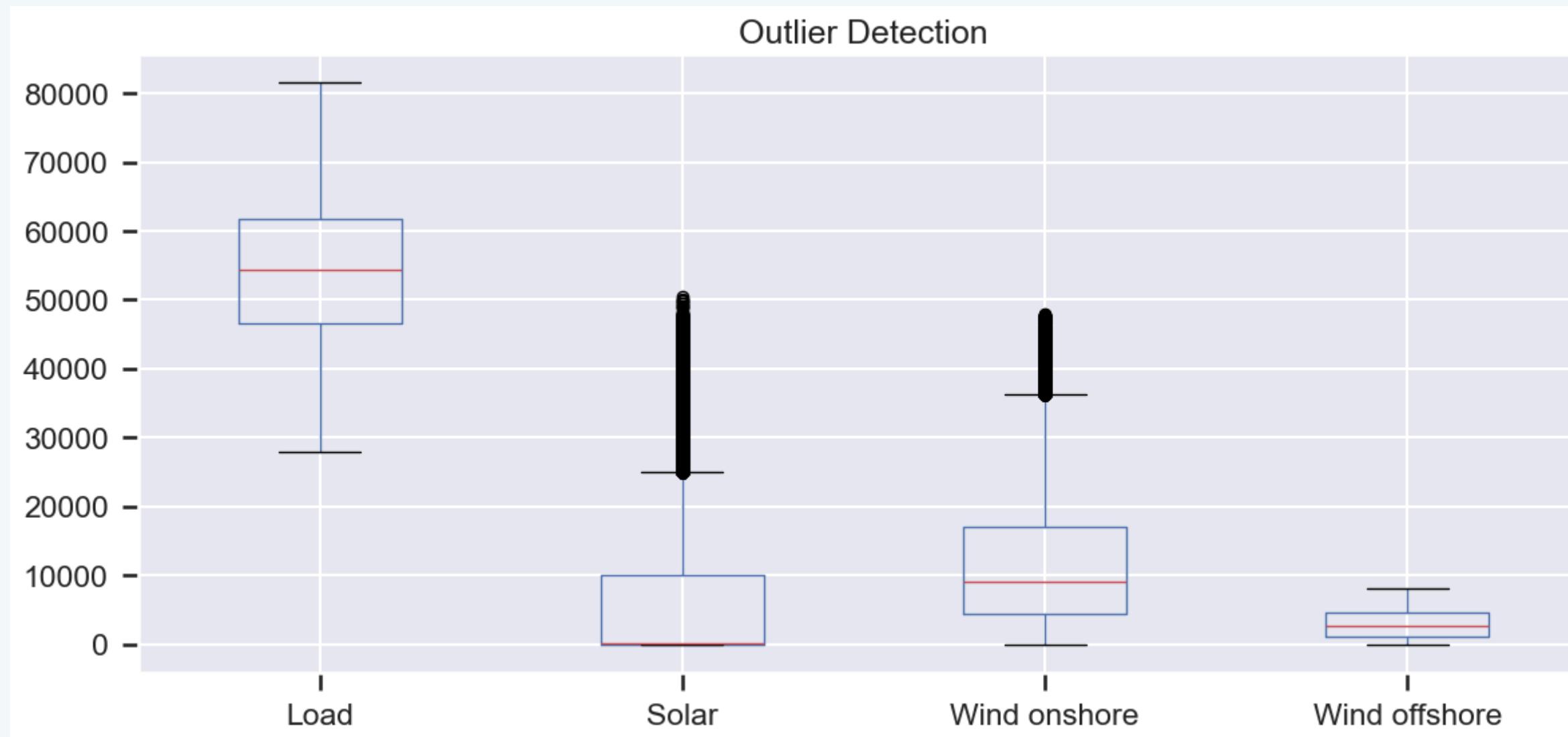
- energy data from 2024 and 2025
- Data was sourced from Fraunhofer-Institut für Solare Energiesysteme ISE
- High seasonal component
- Day Ahead Auction (DE-LU), bids until 12 noon, published at around 12:40 p.m.
- Hydro pumped storage consumption
- Cross border electricity trading
- Hydro Run-of-River
- 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**



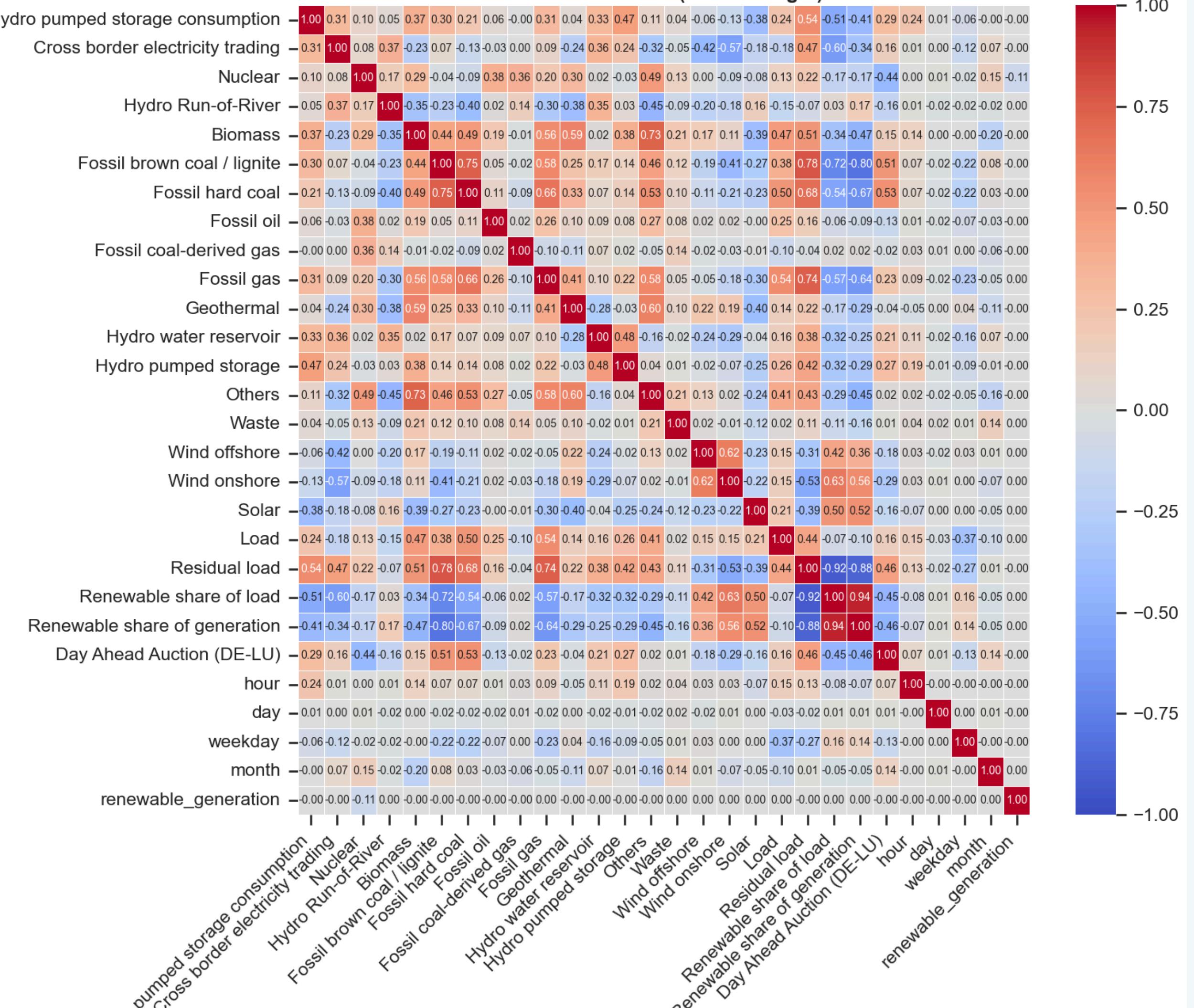
HISTOGRAMS



BOX PLOT (OUTLINERS)

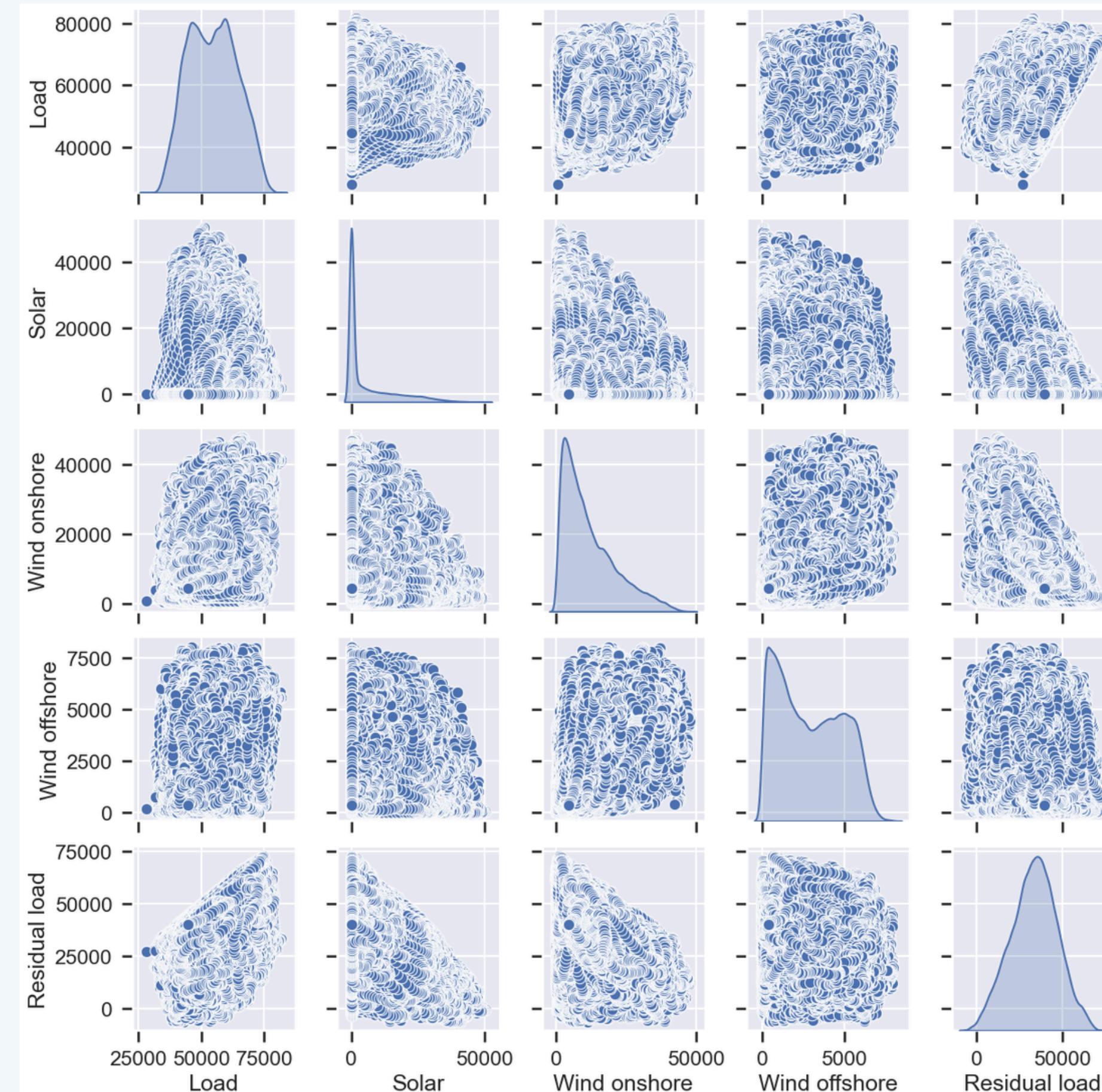


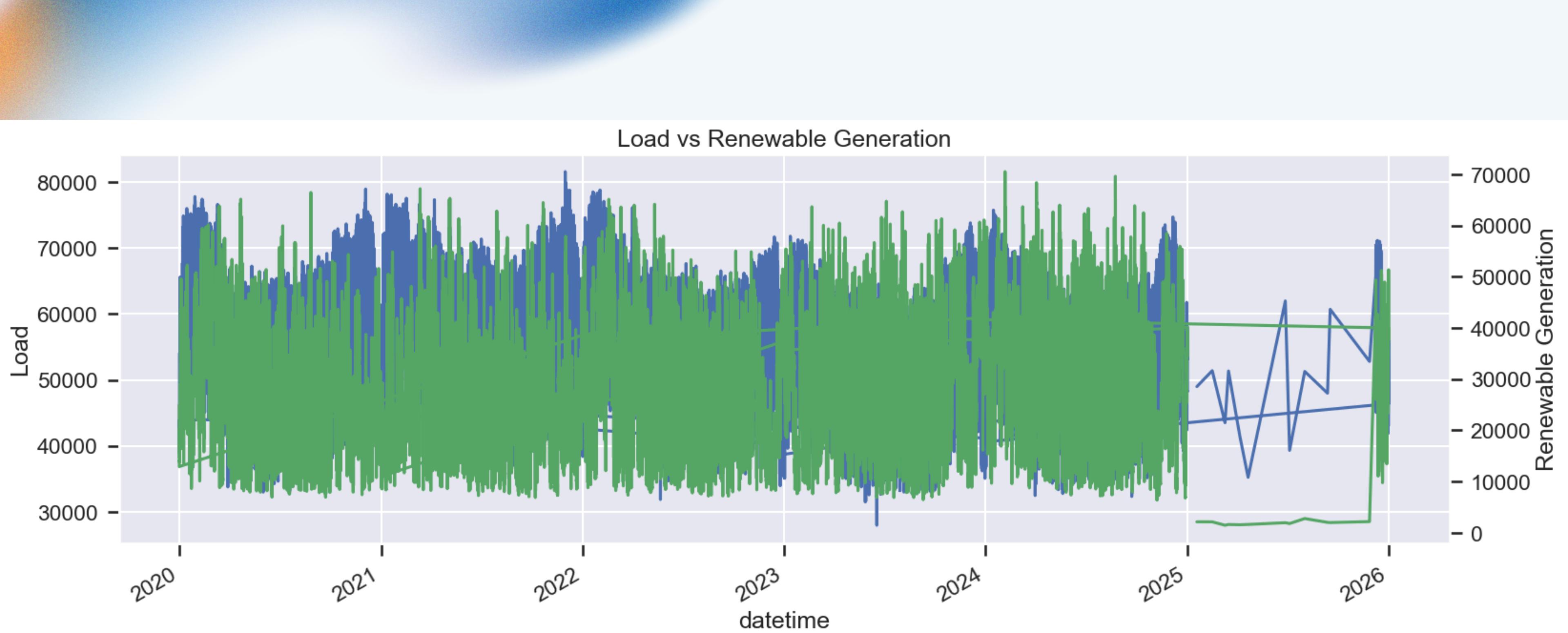
Correlation Matrix (with Strength)

**TARGET VALUE**

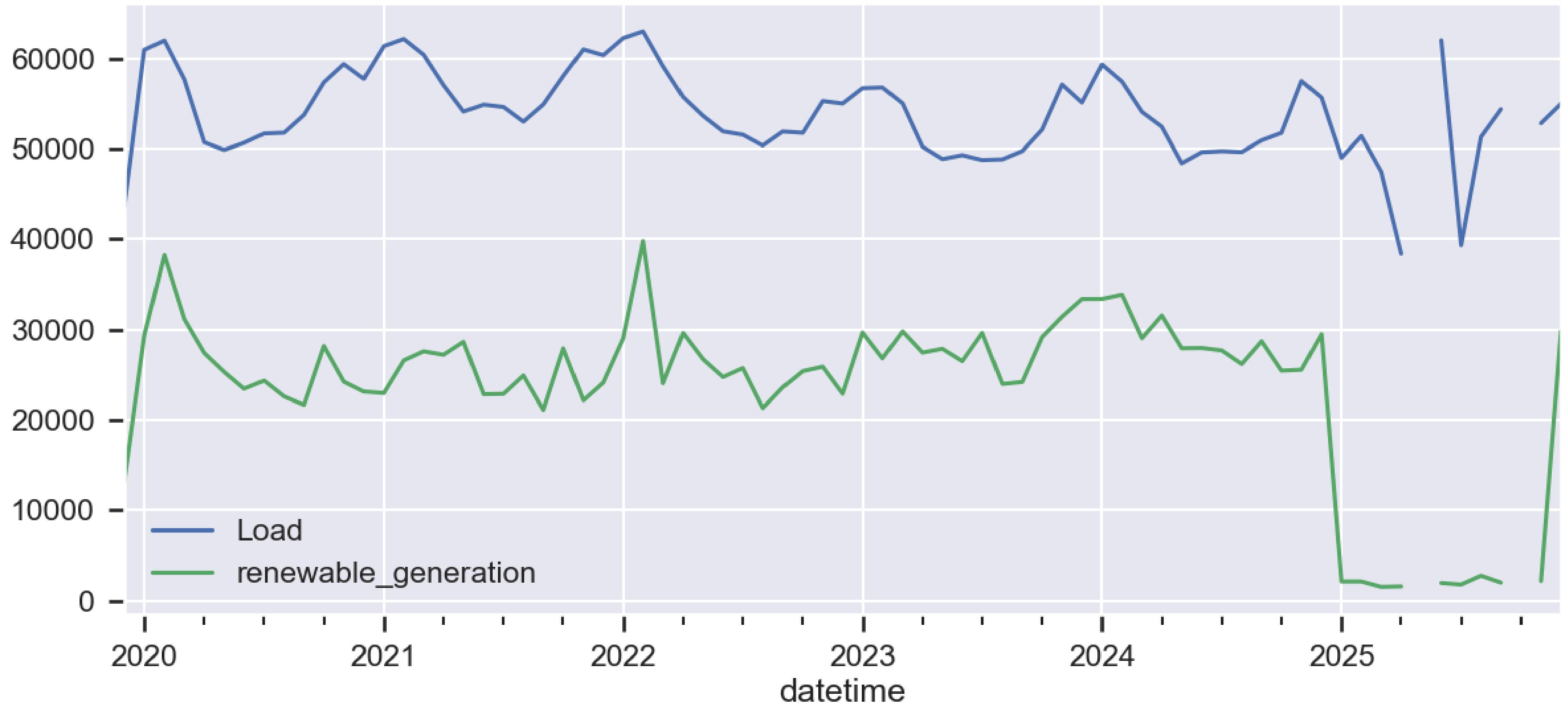
Day Ahead Auction (DE-LU)	1.000000
Residual load	0.742104
Fossil brown coal / lignite	0.596684
Fossil gas	0.545497
Hydro pumped storage consumption	0.544739
Fossil hard coal	0.516677
Cross border electricity trading	0.515919
Hydro pumped storage	0.484154
Biomass	0.478509
Hydro water reservoir	0.381108
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

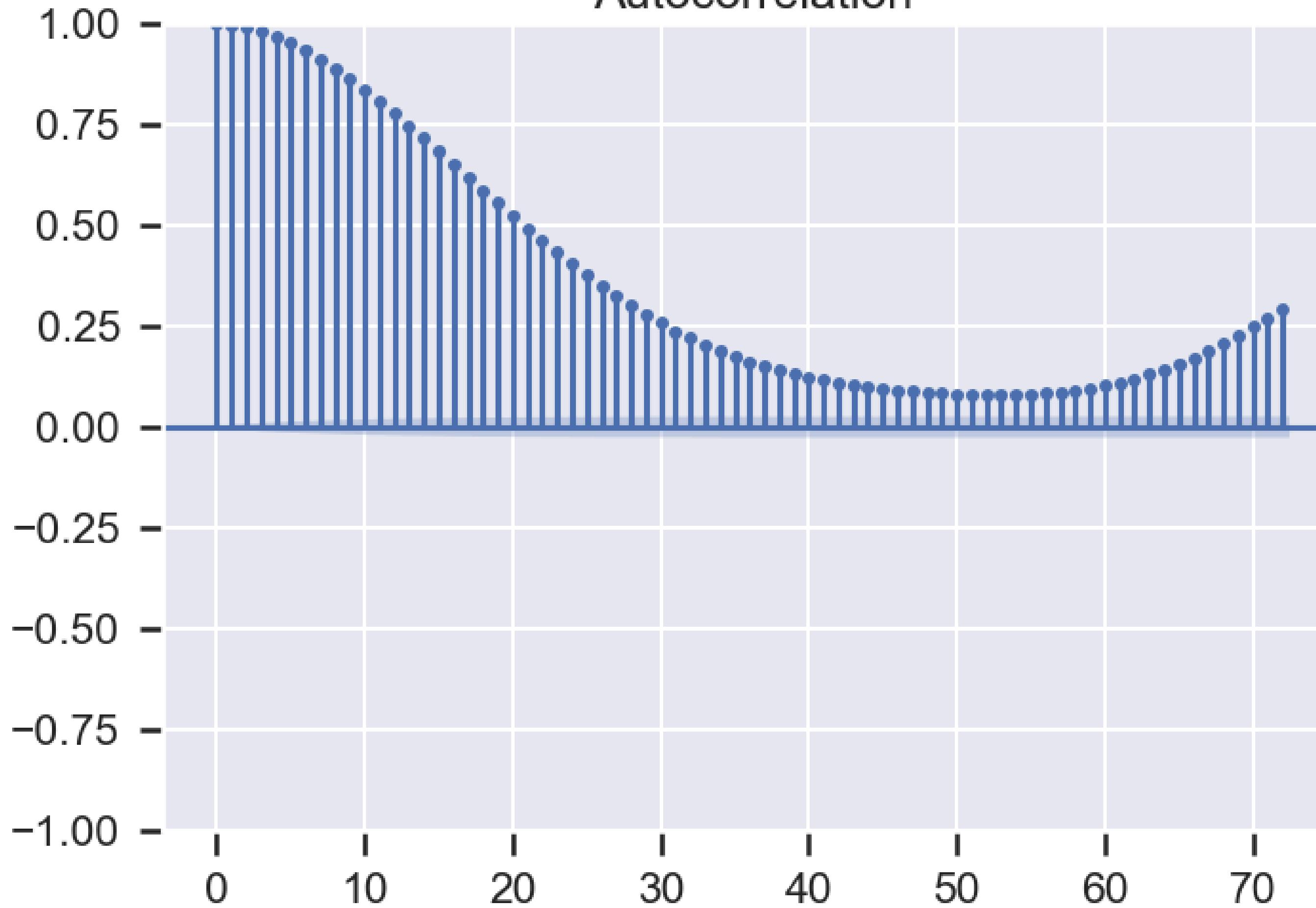




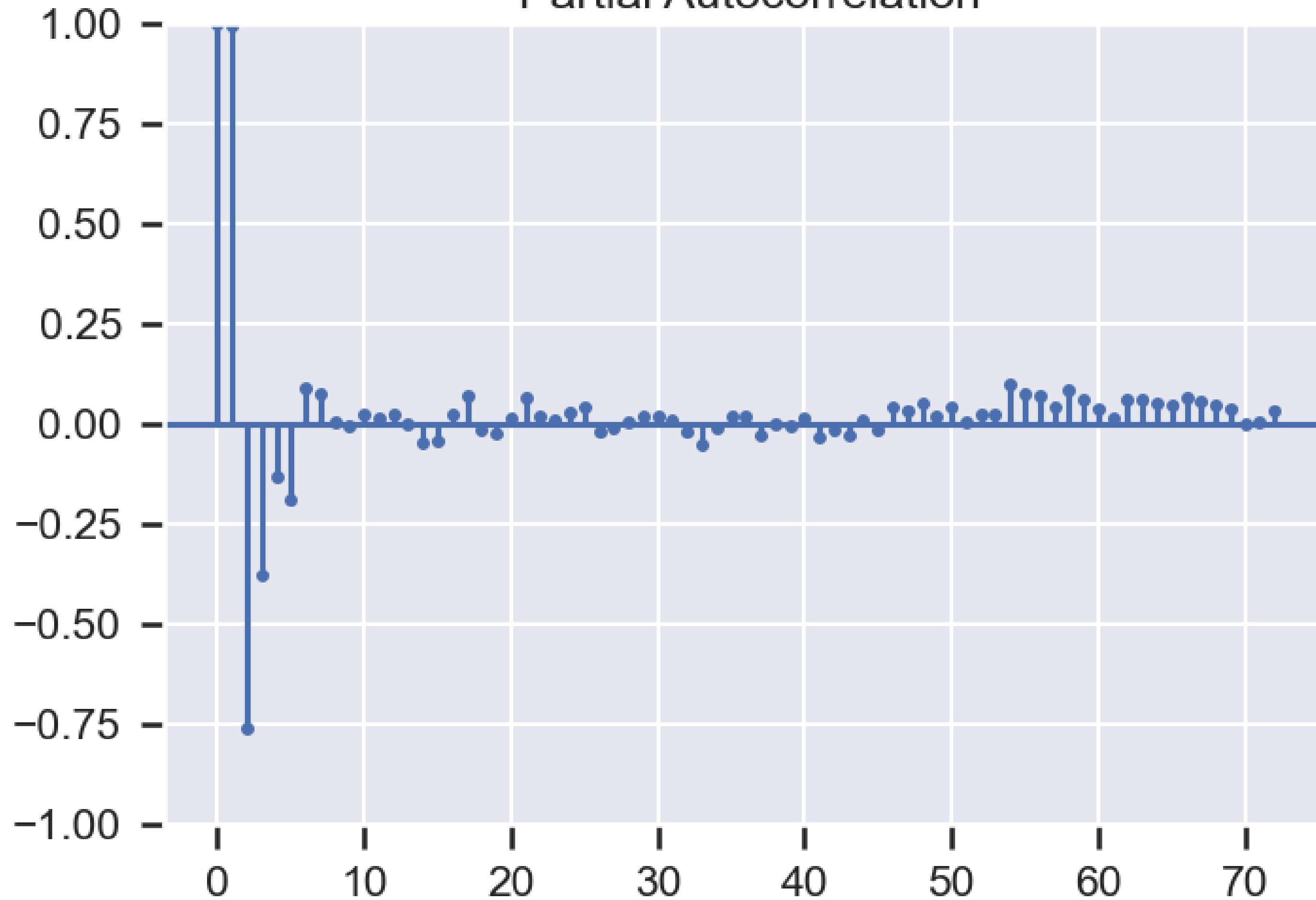
Monthly Trends



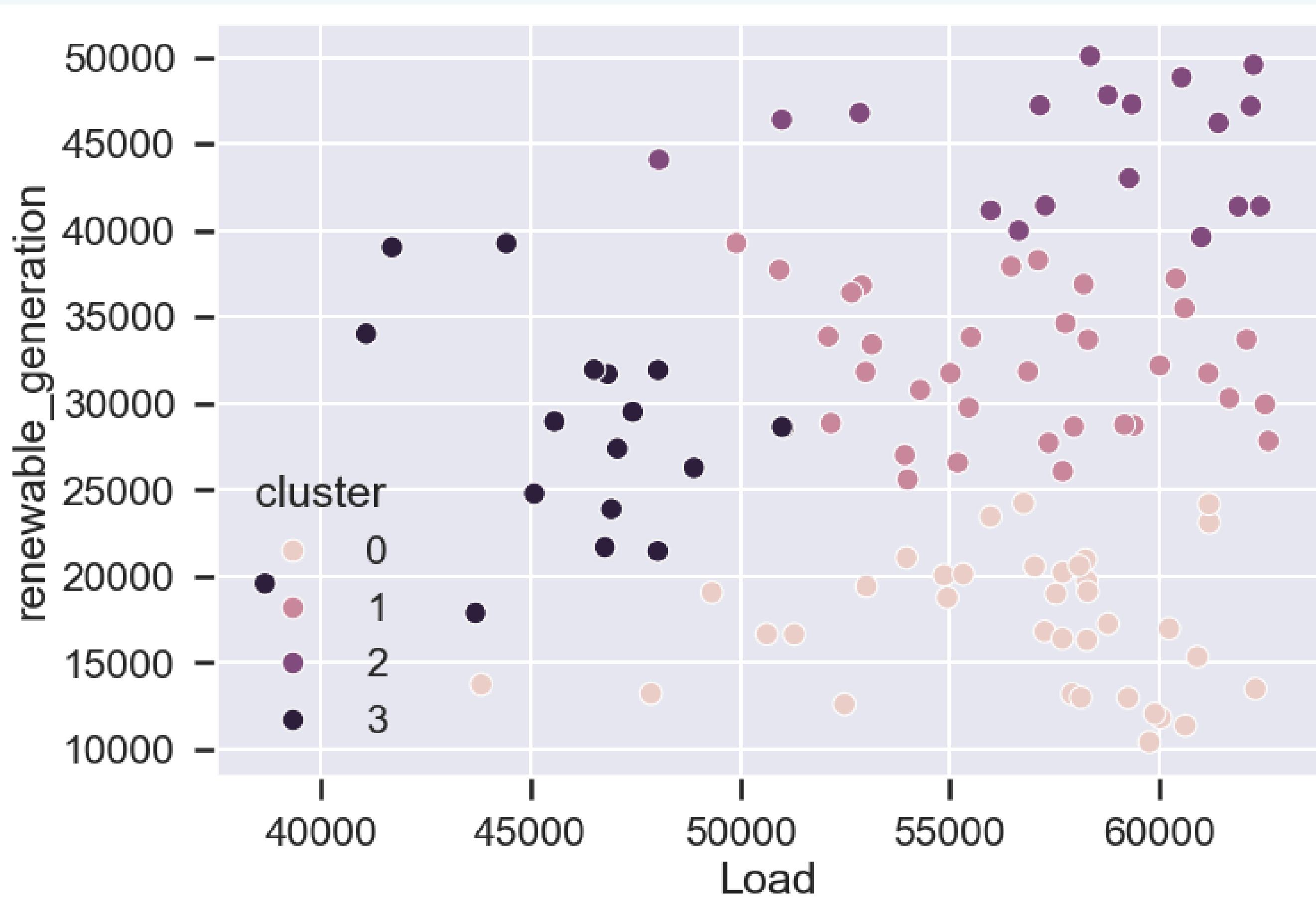
Autocorrelation



Partial Autocorrelation



CLUSTERING



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 (exogenous features). 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