FIELD NOTE

PROJECT WEEKLY REPORT - W1 - 0218-0220

FIELD NOTE

PROJECT WEEKLY REPORT: SWEDISH ELECTRICITY PRICE FORECASTING

Period: Week 1 (Feb 18 – Feb 22, 2026)
Scope: Infrastructure, Data Acquisition, and Advanced EDA

---

1. EXECUTIVE SUMMARY

This week focused on laying a robust foundation for the machine learning pipeline. Key milestones include project architecture initialization, multi-source data integration from European energy APIs, and a comprehensive Exploratory Data Analysis (EDA) uncovering regional market dynamics across Sweden's four bidding zones (SE1–SE4).

---

2. COMPLETED TASKS

A. Project Infrastructure & Version Control

1. **Repository Setup**: Initialized Git repository on GitHub.
2. **Standardized Directory Structure**:

```
swedish-electricity-forecast/
├── data/ (raw, processed, external)
├── src/ (utils.py, features.py, models.py)
├── notebooks/ (EDA and validation)
├── scripts/ (automation scripts)
├── reports/figures/ (analysis exports)
└── requirements.txt
```

3. **Core Utility Library (`src/utils.py`)**:
   - Implemented a base `APIClient` class with retry logic and rate limiting.
   - Built a custom logger with rotating file handlers.
   - Developed `convert_to_swedish_time` to solve Daylight Saving Time (DST) complexities.

B. Data Acquisition & Multi-Source Strategy

Data Type	Source	Primary Tool	Frequency
Price Data	ENTSO-E / Energi Data Service	REST API	Hourly
Production/Consumption	ENTSO-E	API / CSV	Hourly
Weather Data	SMHI	METOBS API	Hourly

Key Achievement: Successfully merged fragmented datasets from ENTSO-E and Energi Data Service, filling coverage gaps for Northern Sweden (SE1/SE2).

C. Advanced Exploratory Data Analysis (EDA)

Produced 4 high-fidelity diagnostic figures in `reports/figures/`:

- **Fig 1: Price Timeline (01_price_timeline_hq.png)**: Visualized 2-year trends with 30-day moving averages.
- **Fig 2: Distribution & Volatility (02_distribution_volatility.png)**: Identified "Negative Prices" in the North and extreme volatility in the South (SE4).
- **Fig 3: Regional Correlation (03_regional_correlation.png)**: Discovered a strong North-South market split (SE1-SE2 corr: 0.97 vs. SE1-SE4 corr: 0.45).
- **Fig 4: Seasonal Heatmap (04_hourly_seasonality_se3.png)**: Confirmed dual daily peaks (08h & 18h) and summer solar-driven dips.

---

3. TECHNICAL CHALLENGES & SOLUTIONS

Challenge	Impact	Engineering Solution
DST Transition	Duplicate or missing hours in time series.	Handled via `tz_localize` with `ambiguous='infer'`.
Duplicate Entries	Pivot Errors during data merging.	Implemented `groupby().mean()` during the consolidation phase.
Market Fragmentation	Missing or inconsistent data across zones.	Developed a dual-source fallback strategy (Energi Data + Manual ENTSO-E).

---

4. DEVELOPMENT PROGRESS & CODE STATUS

- **Scripts**: Automated data downloader (`scripts/download_data.py`) is fully functional.
- **Features**: `DataProcessor` class skeleton initialized in `src/features.py`.

---

5. PLANS FOR NEXT WEEK (WEEK 2)

- [ ] **Feature Engineering**:
  - Handling missing values (interpolation/deletion strategies)
  - Outlier detection and handling (extreme electricity prices)
  - Time zone unification (UTC+1/CET)
- [ ] **SMHI Integration**: Integrate temperature data into the main training pipeline.
- [ ] **EDA Feature Analysis**
  - Feature Distribution Visualization
  - Feature-Target Relationship Graph (Scatter Plot, Heatmap)