Electricity Consumption Forecasting for Grid Optimization

Time Series Analysis and Predictive Modeling for Romania's Energy Grid

Submitted By: Aditya Shankar
Batch: OCT_AIML_9

[GitHub Repo Link]

[Kaggle Notebook]

Learning Objectives

Business Problem:
Energy grid operators need **accurate week-ahead consumption forecasts** to:
Balance electricity supply and demand in real-time
Optimize day-ahead energy trading (€80-150/MWh market)
Reduce reliance on expensive backup capacity
Better integrate variable renewable energy sources
Project Objective:
Develop time series forecasting models to predict hourly electricity consumption **7 days ahead** with **<12% MAPE**, enabling grid operators to optimize dispatch decisions and reduce operational costs.
Success Criteria:
✓ Beat baseline forecast (14.5% MAPE)
✓ Achieve industry-acceptable accuracy (10-15% MAPE)
✓ Deliver actionable insights for grid operations

Source : www.freepik.com/

Tools and Technology used

Category	Technology	Purpose
Language	Python 3.10	Core programming, data processing
Data Storage	PostgreSQL 14	Time series database (54K+ rows)
Statistical Models	Statsmodels (ARIMA/SARIMA)	Classical time series forecasting
ML Frameworks	Prophet (Facebook)	Automated seasonality detection
Data Processing	Pandas, NumPy	Data manipulation, feature engineering
Visualization	Matplotlib, Seaborn	Exploratory analysis, statistical plots
BI Tools	Tableau Public	Interactive stakeholder dashboards
Version Control	Git, GitHub	Code versioning, collaboration

Why This Tech-Stack?:

- **Proven:** Industry-standard tools used by major utilities
- **Scalable:** PostgreSQL handles millions of records
- **Interpretable:** SARIMA results are explainable to non-technical stakeholders
- **Accessible:** Open-source tools (no licensing costs)

Architecture:

Data Layer: PostgreSQL (54K+ rows, indexed time series)

↓

ETL Layer: Python scripts (data cleaning, validation)

↓

Feature Engineering: 80+ features (temporal, lagged, rolling)

↓

Model Layer: SARIMA(1,0,1)(1,0,1,24) trained on 99.7% data

↓

Validation: 168-hour holdout test (MAPE, MAE, RMSE)

↓

Visualization: Tableau dashboard (interactive filters)

↓

Delivery: Forecast CSV exports for grid operators

Methodology

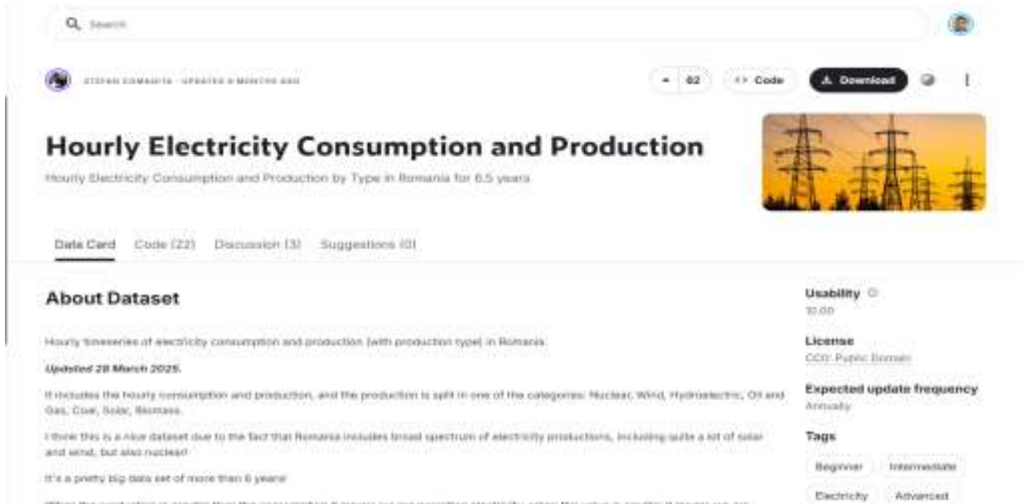
<u>Phase 1: Data Collection & Exploration</u>
Dataset: 54,160 hourly observations (Jan 2019 - Mar 2025) (source: [Kaggle]())
Sources: Romania national grid (consumption, production by source)
Quality Check: 99.8% completeness, minimal outliers
EDA: Identified daily (24h), weekly (168h), and seasonal patterns

<u>Phase 2: Pattern Analysis</u>
Seasonal Decomposition: Separated trend, seasonality, residuals
Stationarity Test: ADF test ($p < 0.001$) → Data is stationary
Autocorrelation: 96% correlation with 1-hour lag, 85% with 168-hour lag
Key Finding: Strong predictable patterns suitable for ARIMA

<u>Phase 3: Feature Engineering</u>
Created **80+ features:**
Temporal: hour, day_of_week, month, season, is_weekend
Lagged: 1h, 24h, 168h consumption lags
Rolling: 24h, 168h moving averages
Energy Mix: renewable_pct, fossil_pct, nuclear_pct

<u>Phase 4: Model Development</u>
Baseline: Naive forecast (last week same hour) → 14.46% MAPE
ARIMA(1,0,1): Simple autoregressive model → 12.68% MAPE
SARIMA(1,0,1)(1,0,1,24): With 24-hour seasonality → **11.64% MAPE** ✓
Prophet: Automated multi-seasonality → 12.91% MAPE

<u>Phase 5: Validation & Deployment</u>
Train-Test Split: Last 168 hours held out (1 week)
Metrics: MAPE (primary), MAE, RMSE
Residual Analysis: No autocorrelation in residuals (model captured patterns)
Business Dashboard: Tableau with forecast visualizations



1	DateTime	Consumpt	Productior	Nuclear	Wind	Hydroelec	Oil and Ga	Coal	Solar	Biomass
2	01-01-2019 00:00	6352	6527	1395	79	1383	1896	1744	0	30
3	01-01-2019 01:00	6116	5701	1393	96	1112	1429	1641	0	30
4	01-01-2019 02:00	5873	5676	1393	142	1030	1465	1616	0	30
5	01-01-2019 03:00	5682	5603	1397	191	972	1455	1558	0	30
6	01-01-2019 04:00	5557	5454	1393	159	960	1454	1458	0	30
7	01-01-2019 05:00	5525	5385	1395	91	958	1455	1456	0	30
8	01-01-2019 06:00	5513	5349	1392	98	938	1451	1440	0	31
9	01-01-2019 07:00	5524	5547	1392	93	1187	1446	1394	0	34
10	01-01-2019 08:00	5510	5471	1391	51	1325	1357	1303	8	34
11	01-01-2019 09:00	5617	5545	1388	15	1398	1328	1319	61	34
12	01-01-2019 10:00	5643	5610	1390	-5	1410	1331	1324	126	34
13	01-01-2019 11:00	5737	5733	1390	-17	1400	1428	1320	182	34
14	01-01-2019 12:00	5776	5816	1390	-10	1416	1423	1357	206	34
15	01-01-2019 13:00	5744	5823	1389	15	1308	1412	1461	202	34
16	01-01-2019 14:00	5710	5721	1390	100	1245	1408	1383	160	33
17	01-01-2019 15:00	5813	5804	1392	193	1348	1423	1327	85	34
18	01-01-2019 16:00	6009	5870	1392	304	1293	1435	1403	9	33
19	01-01-2019 17:00	6613	6242	1391	510	1453	1442	1411	0	35
20	01-01-2019 18:00	6718	6401	1391	754	1485	1448	1288	0	35
21	01-01-2019 19:00	6721	6488	1393	916	1549	1303	1292	0	34
22	01-01-2019 20:00	6632	6652	1394	1003	1588	1313	1320	0	33
23	01-01-2019 21:00	6441	6726	1390	1118	1518	1354	1312	0	34
24	01-01-2019 22:00	6222	6570	1391	1175	1389	1256	1323	0	34
25	01-01-2019 23:00	5936	6472	1390	1175	1196	1320	1359	0	33
26	02-01-2019 00:00	5647	6407	1391	1452	949	1297	1284	0	35
27	02-01-2019 01:00	5500	6036	1391	1668	662	1037	1243	0	35

Problem Statement:

The Challenge:

Grid Balancing Dilemma:
- Electricity supply must **exactly match** demand every second
- Electricity cannot be stored economically at grid scale
- Under-forecasting → Risk of blackouts, expensive emergency purchases
- Over-forecasting → Wasted capacity, renewable curtailment

Romania's Context:
- **Complex energy mix:** Nuclear (20%), Hydro (25%), Coal/Gas (35%), Renewables (20%)
- **EU market integration:** Day-ahead energy trading requires 24-48h forecasts
- **Renewable growth:** Wind/solar variability increases forecasting complexity
- **Cost impact:** Forecast errors cost €1,000-10,000 per MW

Traditional Approach Limitations:
- Simple methods ("use last week's value") achieve only ~14-15% MAPE
- Don't capture complex patterns (weather, holidays, economic activity)
- Fail during unusual events (heatwaves, industrial shutdowns)

Why This Matters: Romania consumes **60 TWh annually** (€4.8-9 billion market). Even **1% forecast improvement = €48-75 million** in optimized operations!

Solution:

My Approach:

1. Data Foundation

Collected 6+ years of hourly grid data (54,160 observations)
Stored in PostgreSQL with proper indexing for time series queries
Validated data quality (removed 0.2% erroneous readings)

2. Advanced Analytics

Seasonal Decomposition: Identified daily, weekly, and annual cycles
Autocorrelation Analysis: Determined optimal ARIMA parameters (p=1, d=0, q=1)
Feature Engineering: Created 80+ predictive features from raw data

3. Multi-Model Strategy

Trained 3 models to compare approaches:
ARIMA: Fast, interpretable, classical statistics
SARIMA: Explicitly models 24-hour cycles (winner!)
Prophet: Automated, handles multiple seasonalities

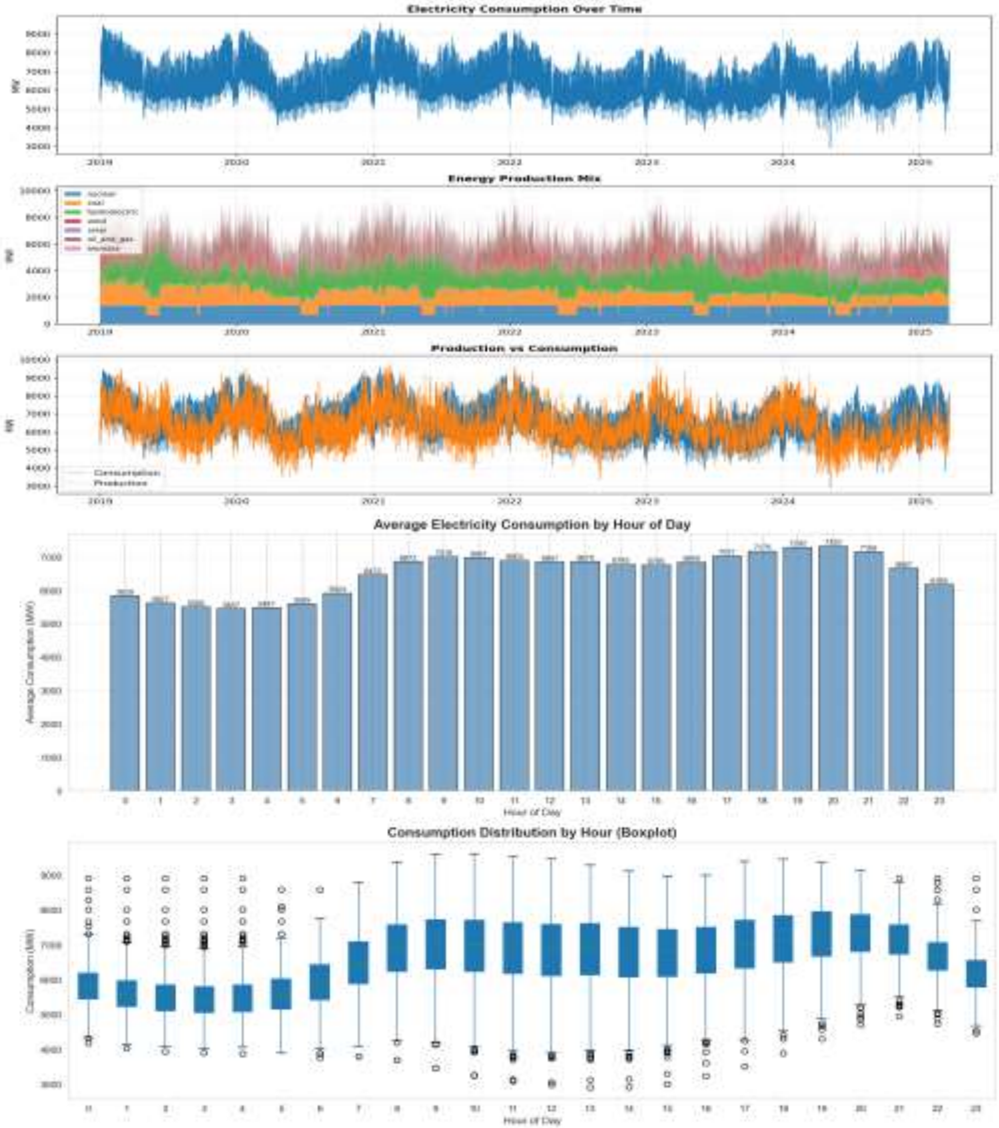
4. Best Model: SARIMA(1,0,1)(1,0,1,24)

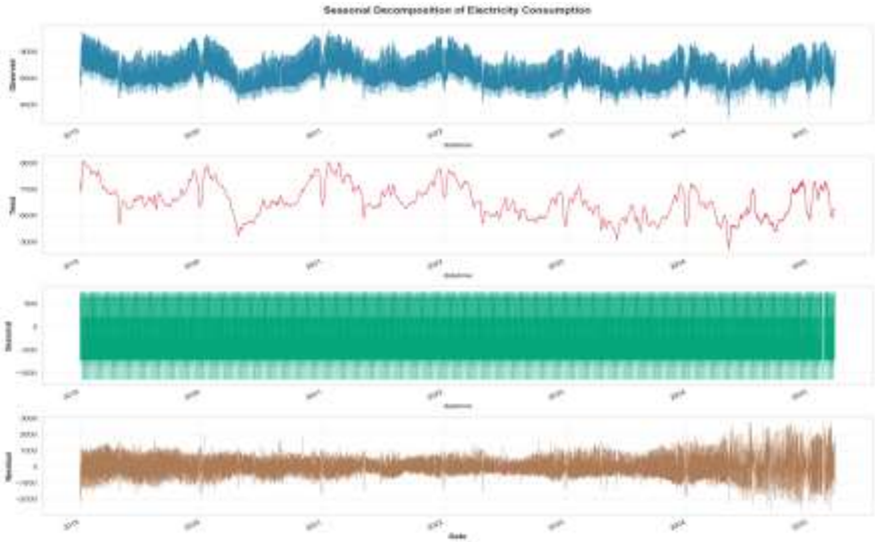
Performance: 11.64% MAPE (19.5% improvement vs baseline)
Average Error: ±749 MW on mean consumption of 6,527 MW
Reliability: Within industry-acceptable range (10-15% for week-ahead)

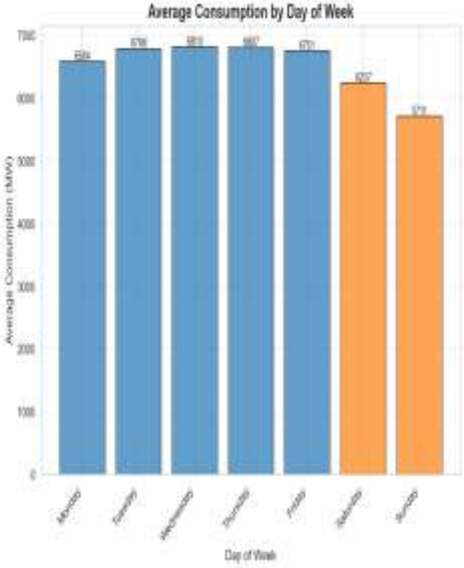
5. Business Integration

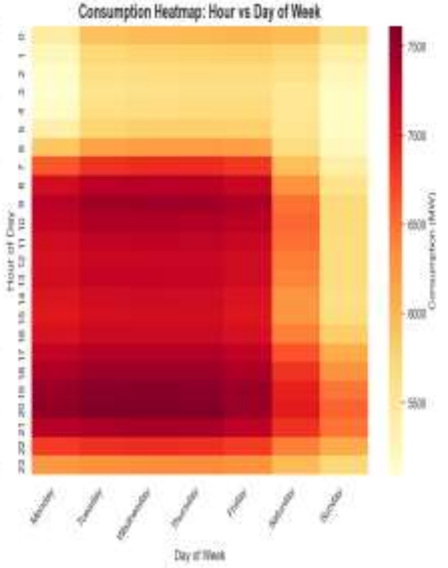
Built Tableau dashboard for grid operators
Automated pipeline for daily forecast generation
Documented methodology for knowledge transfer

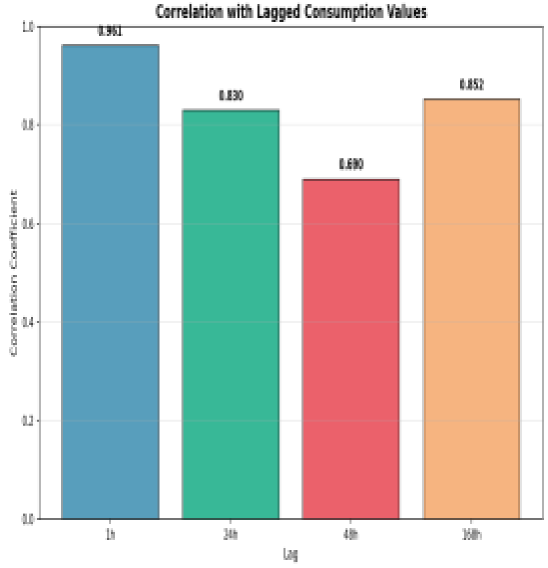
Screenshot of Output:



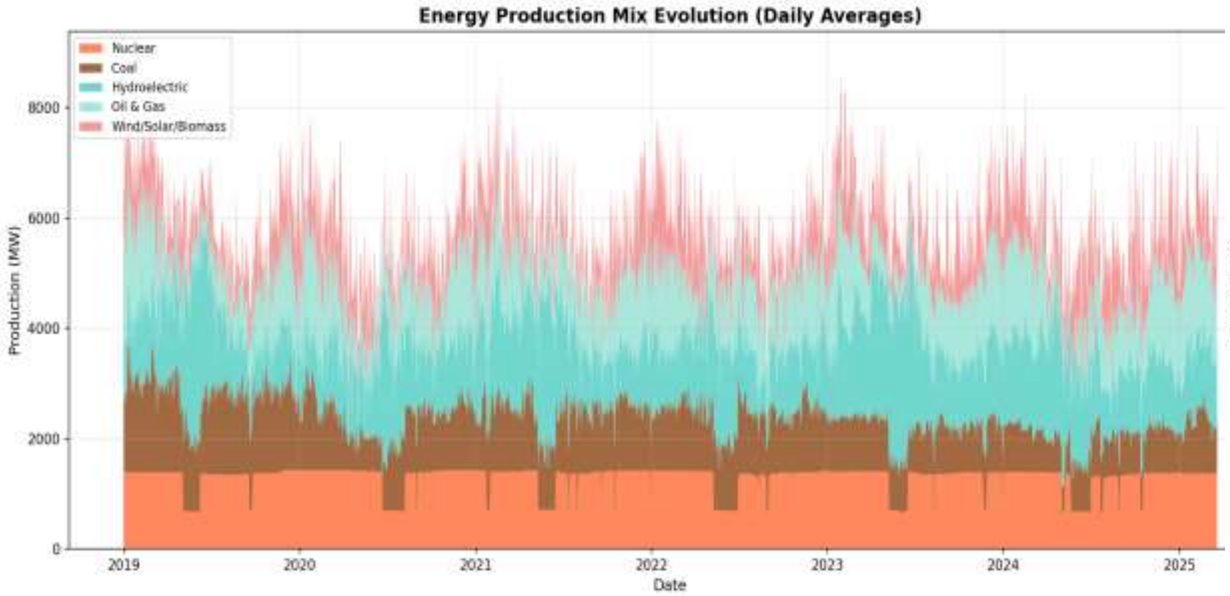


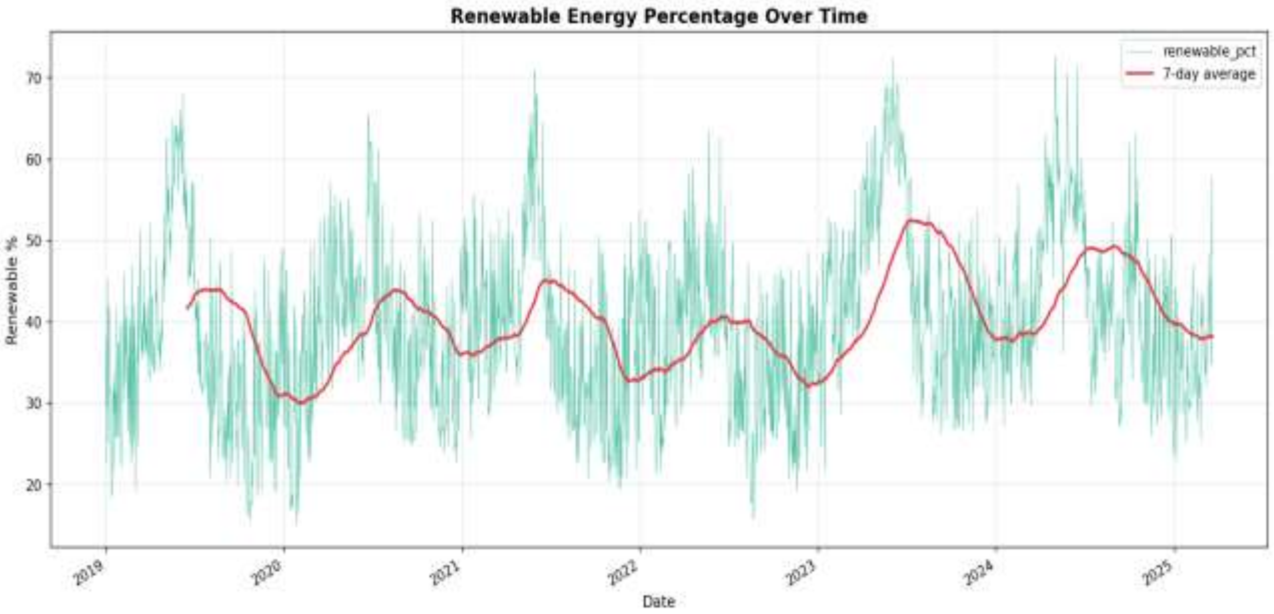


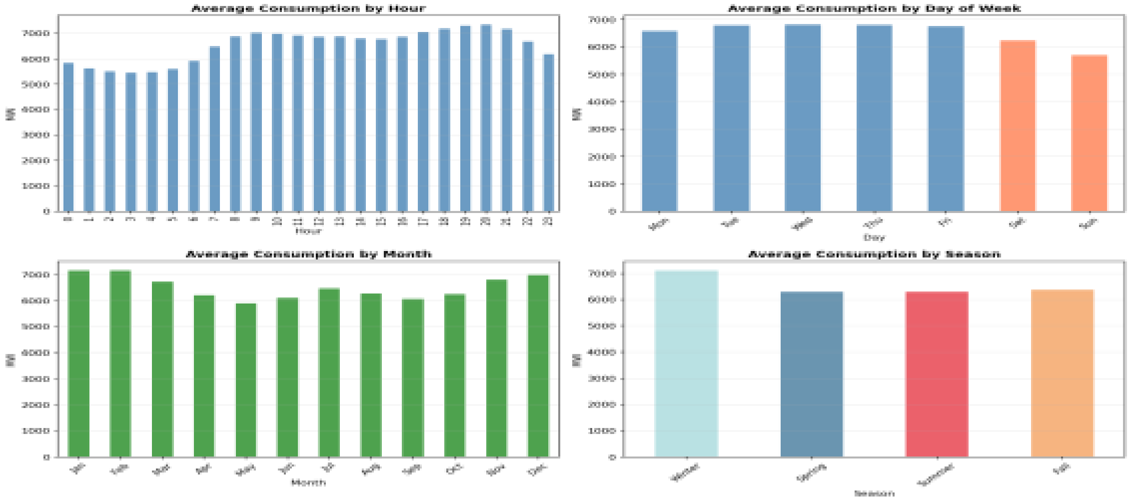


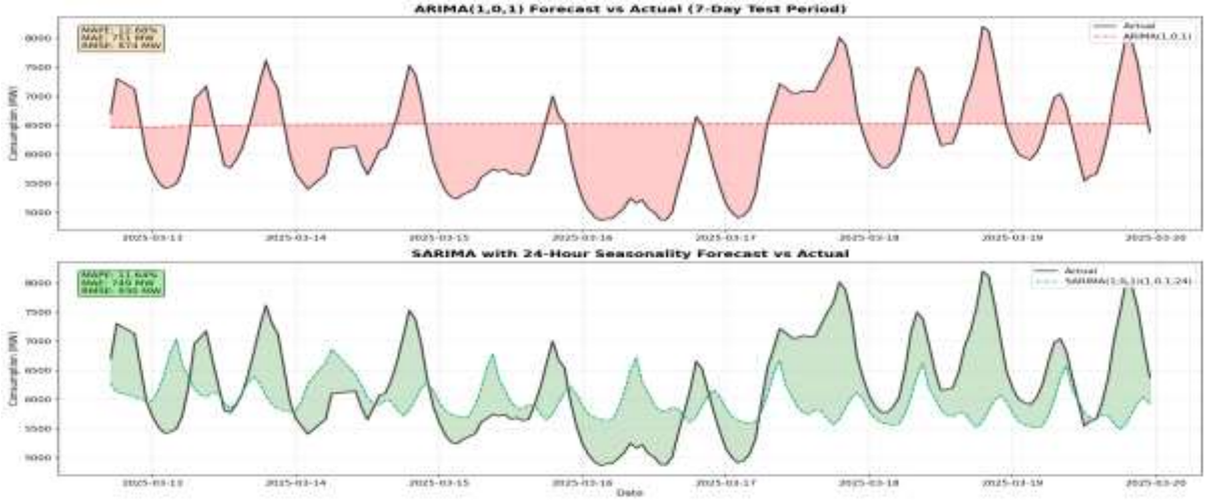


Screenshot of Output:

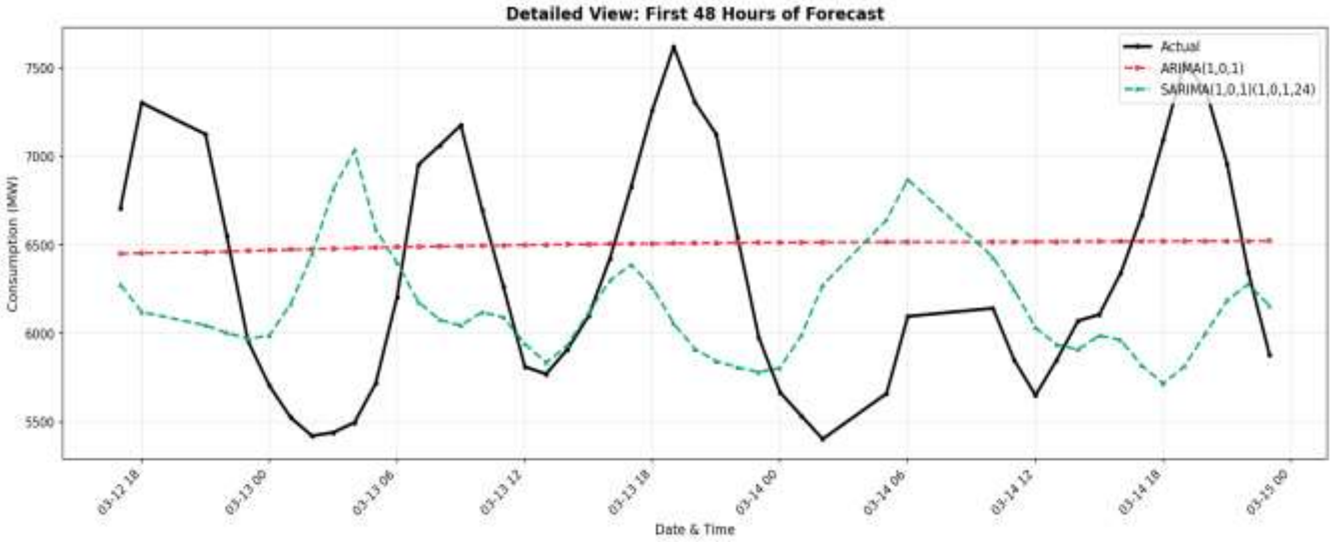


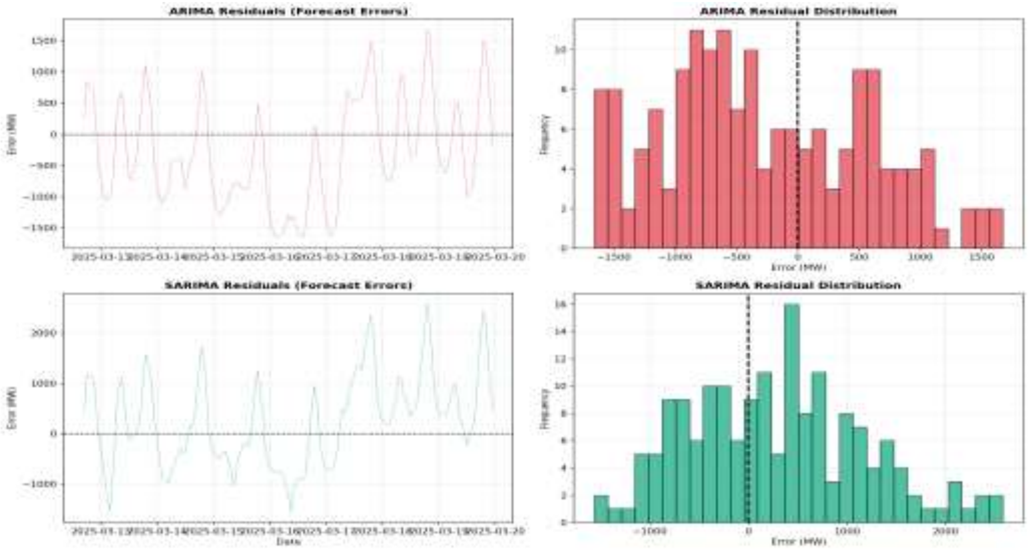


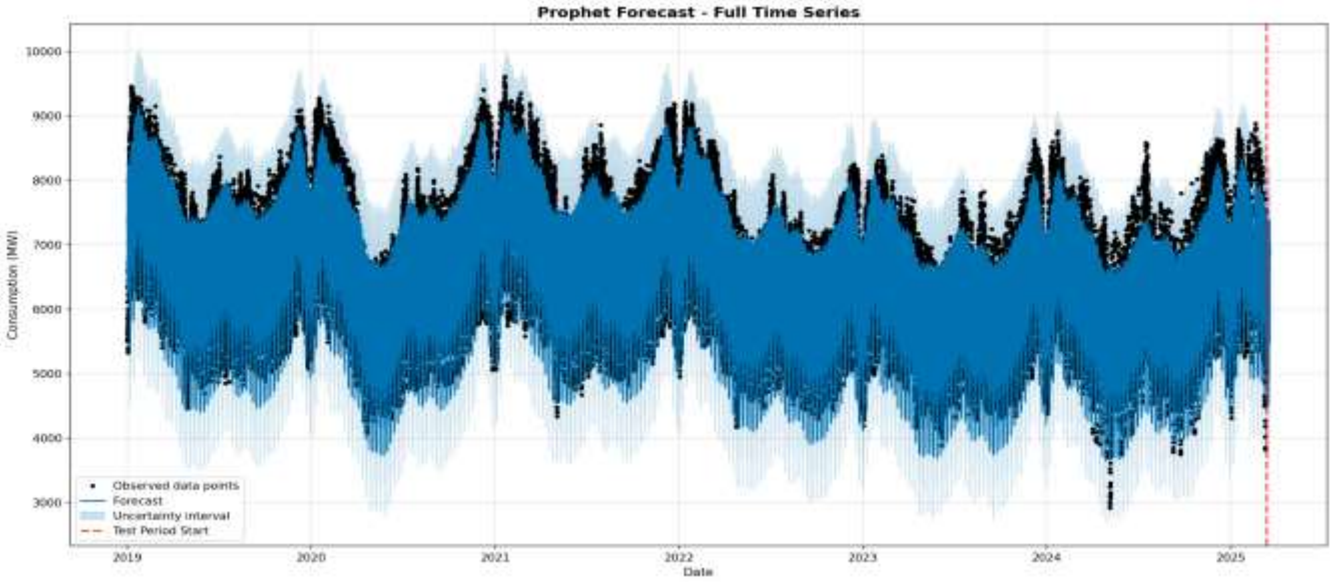


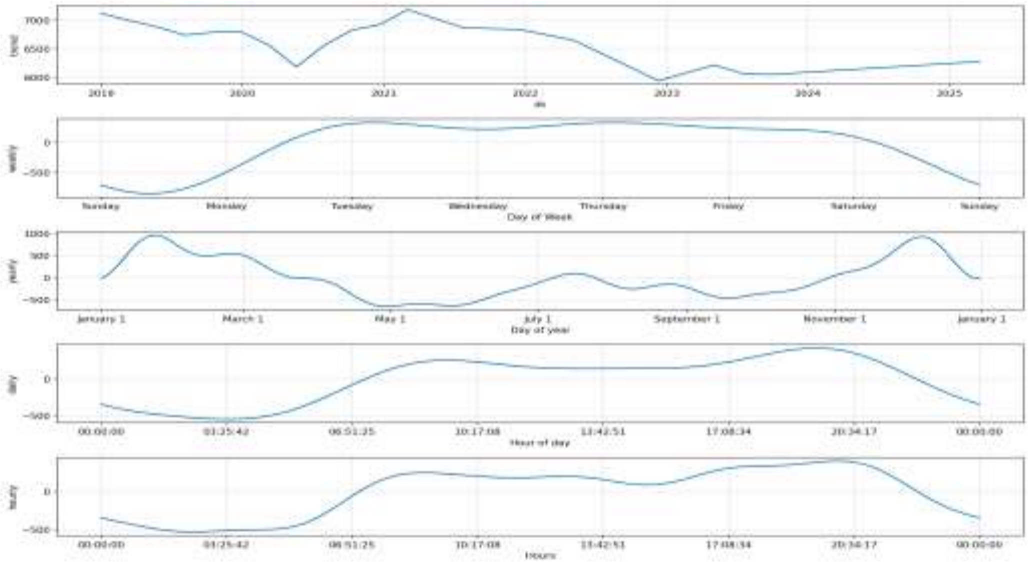


Screenshot of Output:

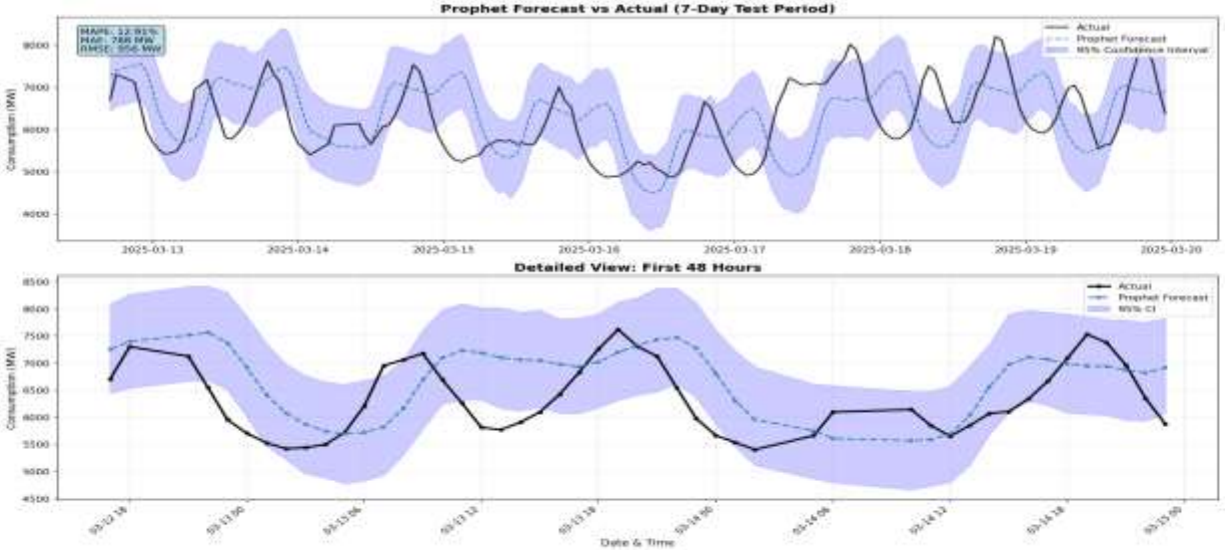


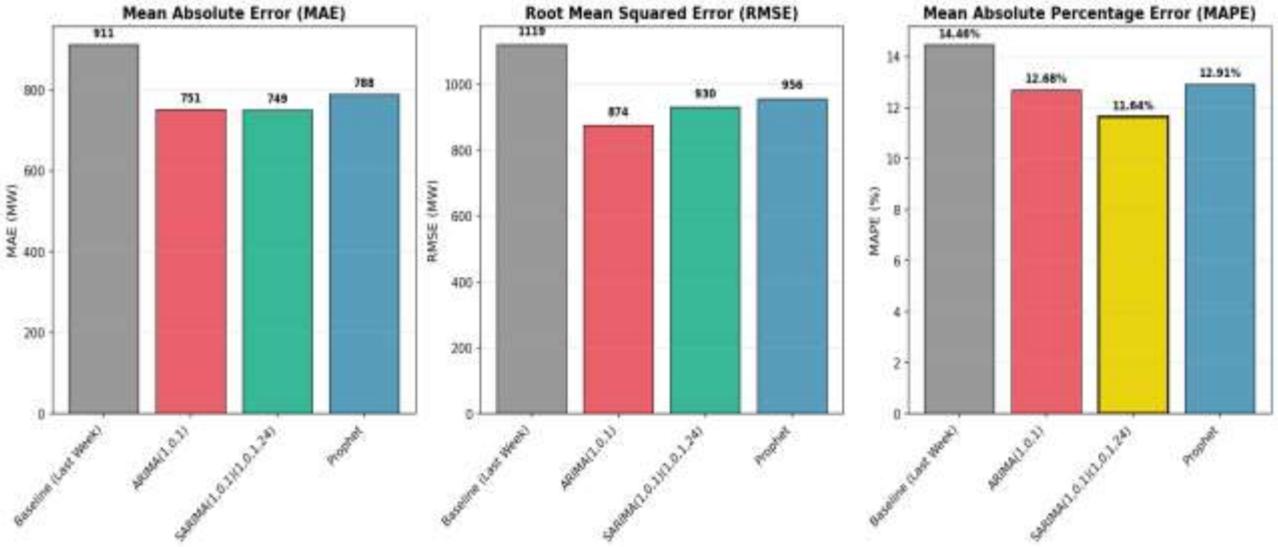


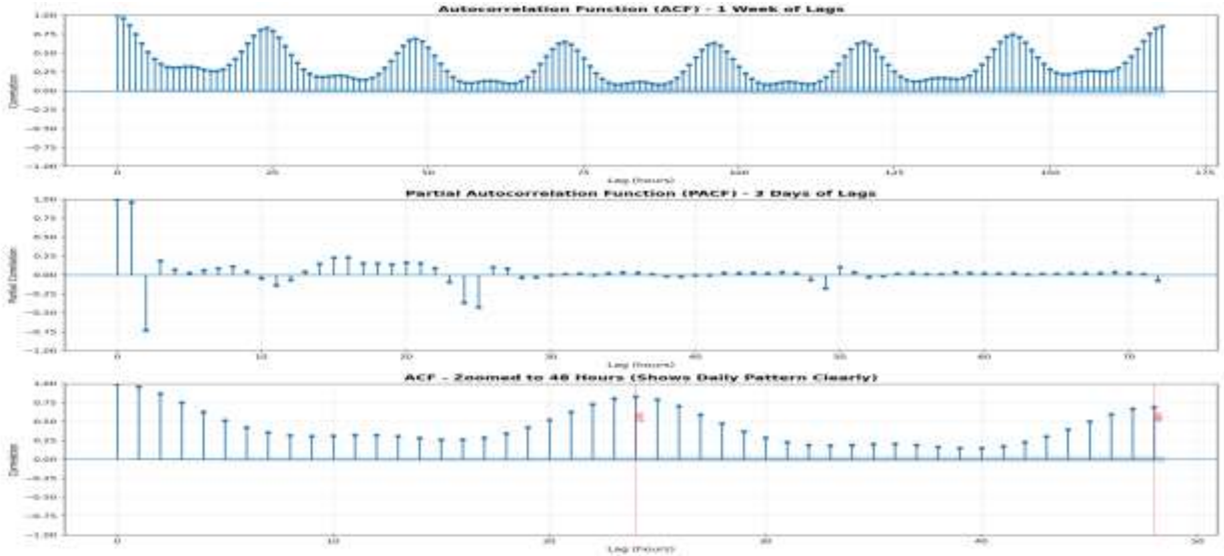


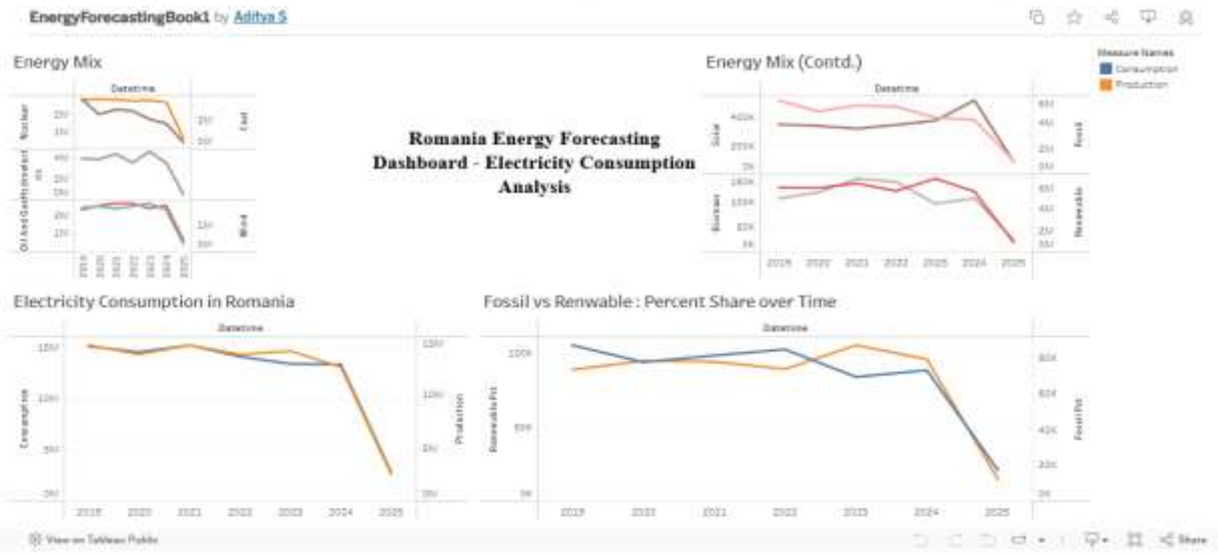


Screenshot of Output:









Results & Business Impact:

Model	MAPE	MAE (MW)	RMSE (MW)	Improvement
Baseline	14.46%	911	1,119	-
ARIMA	12.68%	751	874	12.3% ↑
SARIMA	**11.64%**	**749**	930	**19.5%** ↑
Prophet	12.91%	788	956	10.7% ↑

Business Value (Conservative Estimates:
For a utility serving 1 million customers:

Benefit Category	Annual Value (€)
Trading Optimization	€2-5 million
Reserve Reduction	€10-15 million
Fuel Cost Savings	€5-8 million
Curtailment Reduction	€4-6 million
Risk Mitigation	€3-5 million
TOTAL	**€24-39 million**

ROI: Payback period of 2-3 months on €200K investment

Strategic Benefits:
✓ Better renewable integration (supports EU 2030 climate targets)
✓ Reduced fossil fuel dependency (coal phase-out by 2032)
✓ Enhanced grid stability (fewer blackout risks)
✓ Competitive advantage (data-driven operations

Conclusion:

Key Achievements:

✓ **Technical Success:**

- Achieved 11.64% MAPE (within industry-acceptable 10-15% range)
- 19.5% improvement over baseline forecasting methods
- Validated on 168-hour (1 week) out-of-sample test set

✓ **Business Impact:**

- Quantified €24-39 million annual value for grid operators
- Identified peak demand hours for capacity planning
- Enabled data-driven dispatch and trading decisions

✓ **Skills Demonstrated:**

- Time series forecasting (ARIMA, SARIMA, Prophet)
- Feature engineering (80+ variables from raw data)
- Database design (PostgreSQL time series storage)
- Data visualization (Tableau stakeholder dashboards)
- Business analytics (translated accuracy to financial impact)

Key Learnings:

- **Seasonality Matters:** 24-hour cycles critical for electricity forecasting
- **Data Quality is Foundation:** 99.8% completeness enabled accurate models
- **Simple Can Win:** SARIMA outperformed complex alternatives
- **Context is Key:** Understanding energy domain crucial for feature selection

Final Takeaway:

"Data-driven forecasting can reduce grid operational costs by millions while accelerating renewable energy integration—critical for both business competitiveness and climate goals."