# 📊 Insight Summary Report – Energy Demand Forecasting Pipeline

## 📁 Project Context:

This pipeline analyzes electricity demand patterns using classical ML models, deep learning (LSTM), and time-series forecasting (SARIMAX) across distinct calendar periods (e.g., COVID, CNY, weekends). The models are benchmarked using RMSE, MAE, and R² to assess prediction quality.

## 🔍 Key Insights by Scenario

### 🧭 Scenario-Based Model Performance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Scenario | Best Model | RMSE | MAE | R² | Notable Insight |
| Weekdays | XGBoost | ✓ Low | ✓ Low | ✓ High | Tree-based models capture weekday trends |
| Saturday | LightGBM | ✓ Low | ✓ Low | ✓ High | Lags + forecast features boost accuracy |
| Sunday | LSTM | ✓ Low | ✓ Low | ✓ High | Captures longer-term dependencies |
| CNY | CatBoost | ✓ Low | ✓ Low | ✓ High | Handles holiday volatility effectively |
| COVID Period | SARIMAX | ✓ Low | ✓ Low | ✓ High | Temporal autoregression performs well |

## 📈 Visual Insights

- Model Comparison Bar Charts: Clear differentiation of RMSE/MAE across models.  
- Residual Plots: All models show low residual bias; LSTM has slight heteroscedasticity on weekends.  
- Box Plots of Errors: XGBoost and LightGBM have the most consistent error distribution.  
- Feature Importances: Forecasted demand and lagged actuals are top contributors.  
- Forecast Deviation Plots:  
 - COVID: SARIMAX and LSTM models handle lockdown-driven anomalies well.  
 - CNY: CatBoost and LightGBM maintain consistent performance.

## 🧠 LSTM + SHAP Interpretability

- SHAP summary plots confirm the importance of lagged actual values.  
- Force plots reveal individual predictions influenced heavily by prior 3-period demand.  
- SHAP results align with temporal seasonality.

## 📌 Final Recommendations

1. ✅ XGBoost or LightGBM for general forecasting across weekdays and weekends.  
2. ✅ LSTM recommended for use cases with high sequential dependencies.  
3. ✅ SARIMAX excels in periods with strong autocorrelation (COVID).  
4. ✅ CatBoost can be used in volatile/holiday periods due to categorical robustness.

## 📦 Next Steps

- Incorporate weather and temperature data for seasonal adjustments.  
- Introduce ensemble blending for final demand prediction.  
- Evaluate models on a rolling time window to capture temporal drift.

\*Prepared for DSS5104 — Semester 2 AY2024/25\*