FORECASTING RESIDENTIAL ELECTRICITY DEMAND USING R: A TIME SERIES APPROACH FOR A U.S. UTILITY FIRM

1. Background

A regional energy utility serving 3 million households in the U.S. Midwest aimed to improve its short-term and long-term electricity demand forecasting. Rising usage during extreme weather and increasing adoption of smart devices made past static models unreliable. The utility sought a transparent R-based forecasting model, capable of incorporating both seasonal patterns and weather variables for grid planning and rate setting.

2. Objective

- To forecast monthly residential electricity demand over the next 18 months
- To isolate and model trend, seasonality, and temperature-driven effects in demand
- To build a forecasting pipeline that aids infrastructure and pricing strategy

3. Data Used

Sources:

- Smart meter data from residential users
- NOAA historical temperature data by ZIP code
- Historical electricity rates

Structure:

- Monthly aggregated usage from Jan 2015 to Dec 2023
- Data columns: Month, Region, kWh_Consumed, Avg_Temperature, Peak_Demand_Hours, Holiday_Indicator

4. Methodology

4.1 Data Preparation

- Cleaned for missing temperature and meter reading anomalies

- Aggregated data to monthly frequency by region
- Merged external NOAA datasets using left_join() on region and month

4.2 Decomposition and Pattern Recognition

- Used stl() decomposition for trend, seasonality, and residual extraction
- Detected peak in demand during July (cooling) and January (heating)
- Calculated Cooling_Degree_Days and Heating_Degree_Days as regressors

4.3 Forecasting Models in R

- Built baseline ARIMA and ETS models on deseasonalized data
- Developed **dynamic regression models** (tslm()) with external regressors:
    - kWh ~ Temperature + Holiday_Indicator + Trend
- Applied auto.arima() with external regressors for flexible modeling

```
model_dynamic <- tslm(kwh_ts ~ Temperature + Holiday_Indicator + trend)
forecast_dynamic <- forecast(model_dynamic, newdata = future_df)
```

4.4 Validation

- Used rolling-origin cross-validation for 2019–2023 window
- Accuracy assessed using RMSE, MAPE, and MAE
- Best models selected by region using lowest error score and residual independence

5. Results

Region	**Best Model**	**RMSE**	**MAPE (%)**
Urban Core	ARIMA + regressors	2,200	4.8%
Suburban	ETS(A,A,N)	2,950	6.1%
Rural	ARIMA(1,0,1)	1,980	5.4%
Coastal Edge	Dynamic tslm()	2,160	4.3%

6. Interpretation and Recommendations

- **Temperature was the most significant predictor**, especially during summer months

- Dynamic regression models provided **better accuracy** than pure ARIMA/ETS in high-variance regions
- Recommend switching to **temperature-linked pricing alerts** for peak season load
- Urged deployment of **smart thermostat partnerships** to flatten peak-hour curves
- Created a monthly **“strain index”** that combines kWh + peak-hour flags for internal load tracking

7. Reporting Output

- **R Markdown Report (PDF, 38 pages)**:
    - Forecasts with 80% and 95% confidence intervals
    - STL plots, residual diagnostics, and region-wise comparisons
    - Scenario simulation under different temperature bands
- **Excel Planner File**:
    - Monthly demand forecasts for each service region
    - Weather-adjusted projections
    - Peak-hour prediction overlay for maintenance scheduling
- **Reusable R Components**:
    - Forecasting function by ZIP: forecast_demand_by_zip(zip, horizon = 12)
    - Pre-built visualization templates using ggplot2
    - Parameter-tuning notebook for retraining with new temperature data

8. Business Impact

- Helped **reduce power purchase cost variance by 11%** during 2024 Q1
- Informed **rate adjustments** tied to projected demand spikes
- Supported **grid capacity investment planning** for 2025 infrastructure expansion
- Adopted by the firm’s data science team for integration with **smart meter APIs**