Smart Home Energy Consumption Analysis

By
Lakshmi Venkataramu (202304395)
Hariprasad Sivapatham Anand (202205341)

Introduction

- Smart homes use interconnected devices and sensors to automate and control various home functions, like lighting, temperature, and security, often leading to more efficient energy use.
- Technologies involved include the Internet of Things (IoT), smart thermostats, smart plugs, and intelligent lighting systems, all of which help manage energy use more effectively.

Why Analyze Energy Consumption?

- As smart homes become more common, understanding and optimizing energy consumption is crucial for both environmental and economic reasons.
- Insightful energy analysis can lead to cost savings for homeowners and reduce the overall environmental impact.

Objectives

Seasonal Trend Analysis

- Identify seasonal variations in energy consumption, such as differences across seasons, weekdays vs. weekends, and specific times of day.
- Helps in adapting energy management strategies to seasonal needs, promoting efficiency year-round.

Occupancy-Based Energy Usage Segmentation?

- Segment energy usage based on home occupancy status, distinguishing between times when the home is occupied and unoccupied.
- Provides insights into energy-saving opportunities during unoccupied periods, encouraging optimal usage.

Correlation Analysis

- Analyze correlations between energy usage and variables such as time of day, temperature, and occupancy.
- Understanding these relationships enables smarter, data-driven adjustments to energy settings and consumption.

Outlier Detection and Forecasting

- Detect unusual consumption events (outliers) and forecast future energy demand.
- Outliers can signal inefficiencies or potential issues, while accurate forecasts aid in planning and resource allocation for future needs.

Dataset Overview

- **Source:** Dataset hosted on Kaggle, originally used for energy analysis and forecasting in smart home environments.
- **Volume:** Large dataset with over 5 lakh records and 32 attributes enabling robust analysis, reflecting typical big data characteristics.
- **Data Structure:** Time-series data capturing energy consumption patterns, with features including:
- **Timestamp:** Records date and time of each observation.
- **Energy Consumption:** The primary feature of interest, measuring energy usage in real-time.
- **Occupancy Status:** Indicates whether the home is occupied, a key factor in segmenting energy usage.
- **Environmental Variables:** Includes factors such as temperature or humidity, affecting energy demand.

Dataset

HomeC.csv (130.95 MB)

Detail Compact Column

10 of 32 columns

# time	# use [kW]	# gen [kW]	# House overall [k...	# Dishwasher [kW]	# Furnace 1 [kW]	# Furnace 2 [kW]
time	Total energy consumption	Total energy generated by means of solar or other power generation resources	overall house energy consumption	energy consumed by specific appliance	energy consumed by specific appliance	energy consumed by specific appliance
1.45b – 1.45b	0 – 14.7	0 – 0.61	0 – 14.7	0 – 1.4	0 – 1.93	0 – 0
1451624407	1.4319	0.003416667	1.4319	0.00025	0.477866667	0.178633333
1451624408	1.6273	0.003416667	1.6273	0.000183333	0.44765	0.3657
1451624409	1.735383333	0.003416667	1.735383333	1.67E-05	0.17155	0.6825
1451624410	1.585083333	0.003416667	1.585083333	5.00E-05	0.0221	0.678733333
1451624411	1.510316667	0.003433333	1.510316667	3.33E-05	0.021966667	0.620666667

Data Preprocessing

	time
0	1451624400
1	1451624401

	time
0	2016-01-01 05:00:00
1	2016-01-01 05:01:00

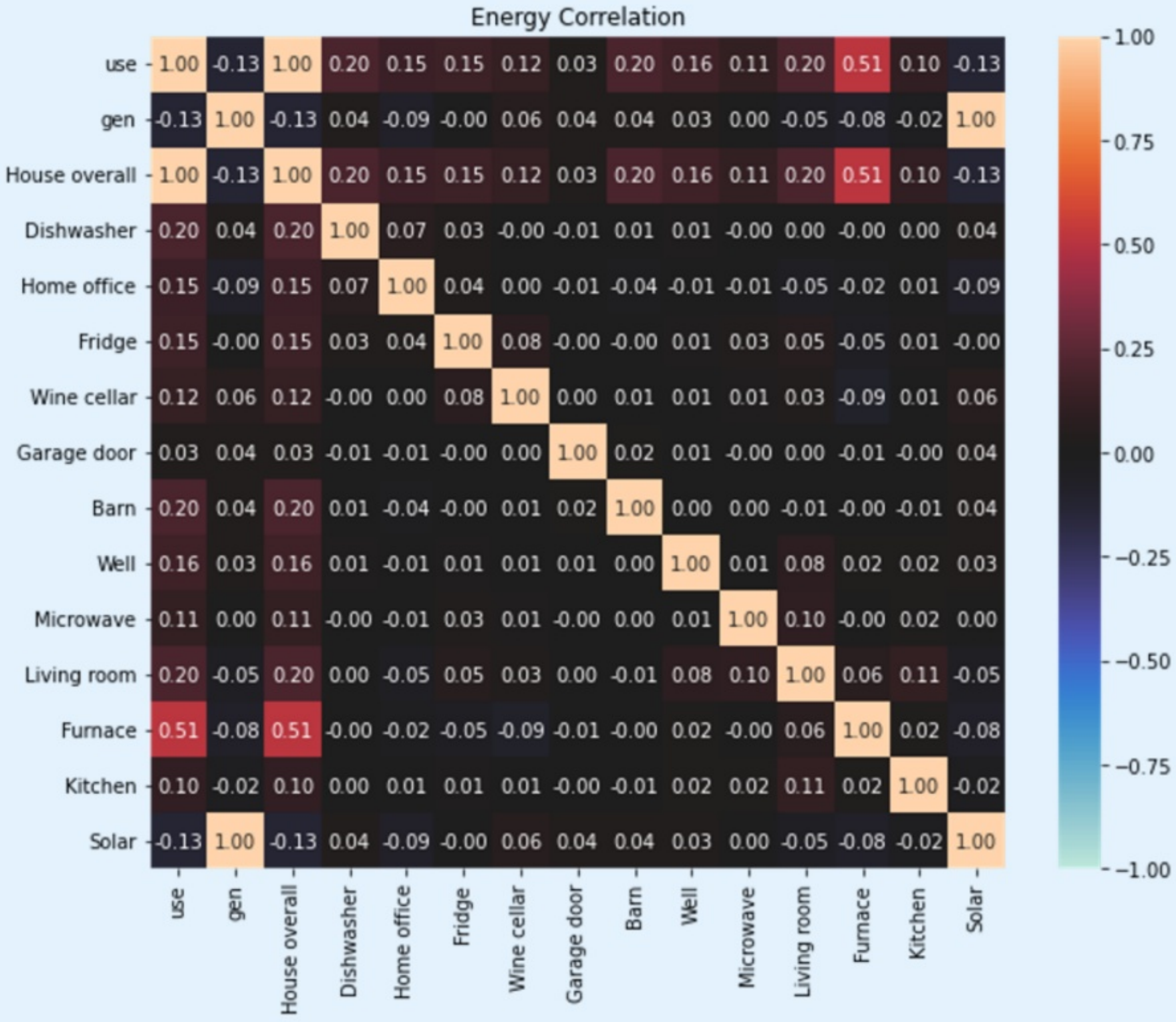
Kitchen 12 [kW]	Kitchen 14 [kW]	Kitchen 38 [kW]
0.000417	0.00015	0.0

Kitchen
0.000567

Datetime Conversion: Converted Unix timestamps to a human-readable datetime format for easier interpretation and time-series analysis.
For example, 1700224800 corresponds to 2016-01-01 05:00:00.
Resampling Data: Resampled the dataset to a consistent frequency, such as hourly or daily, to handle irregular intervals and standardize the time-series data.
Feature Engineering: Extracted new features from the datetime column, such as hour, day, or month, to capture temporal patterns in energy usage.
Combining Feature: Aggregated similar features, such as summing energy usage across different kitchens.
Encoding Categorical Variables: Transformed categorical features, such as occupancy or weather conditions, into numerical formats using one-hot encoding.

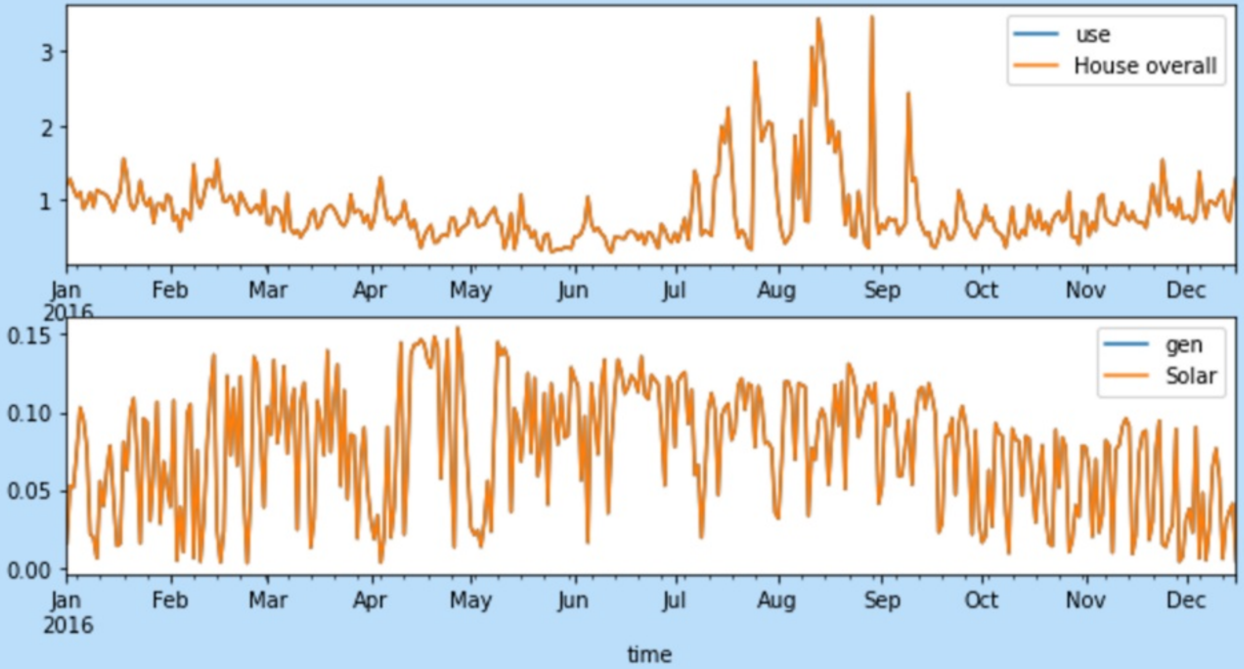
Energy Correlation with house appliances

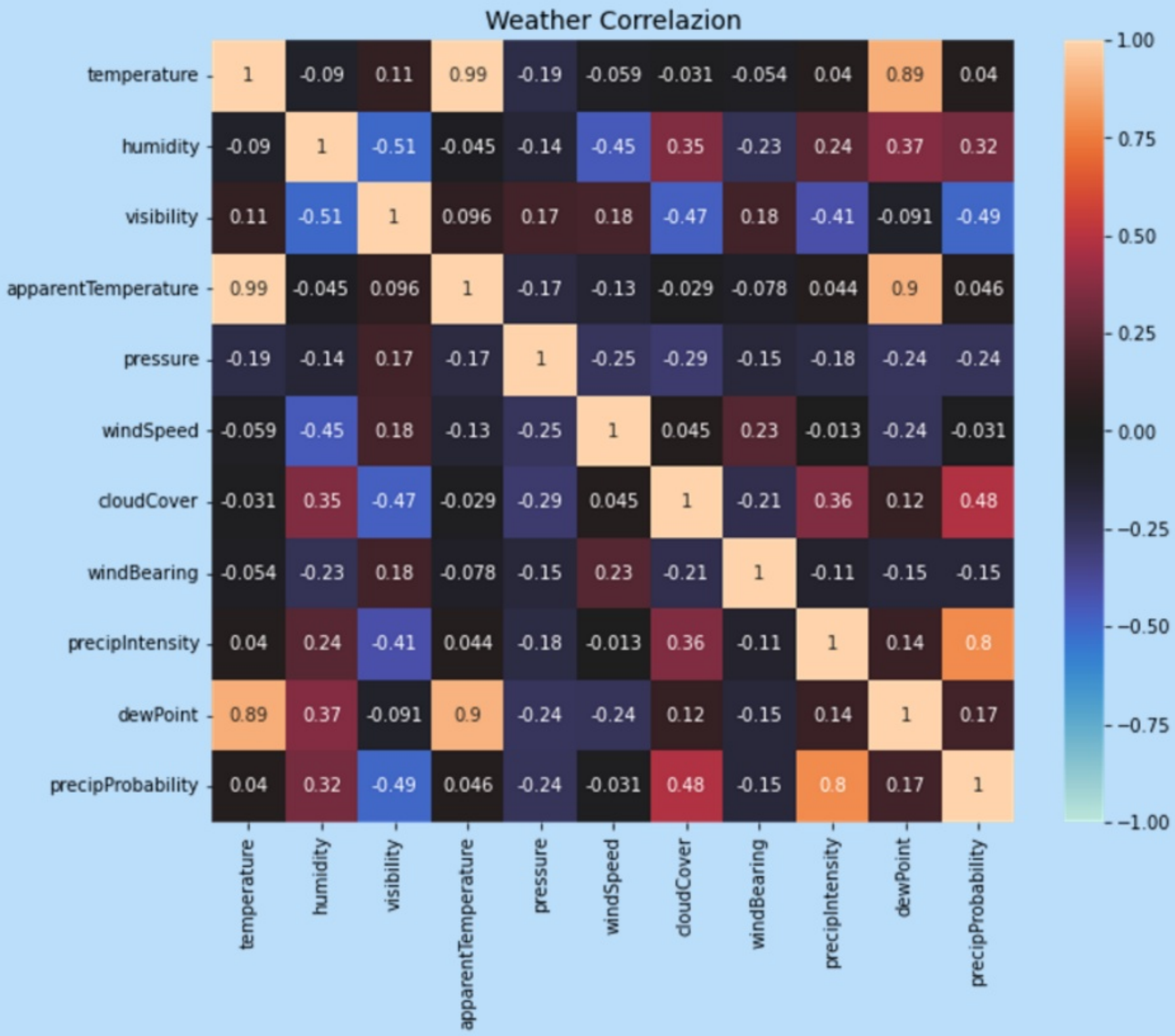
Represents the correlation matrix of different house appliances variables in the dataset, where values range from -1 (strong negative correlation) to 1 (strong positive correlation). Notable correlations include the strong positive relationship (0.51) between two variables, indicating potential dependencies, while most variables show weak or no significant correlations.



Overlapping data removal:

The analysis confirmed that the columns “use” and House overall”, as well as “gen” and “solar”, are identical. These duplicate columns were removed to reduce redundancy.





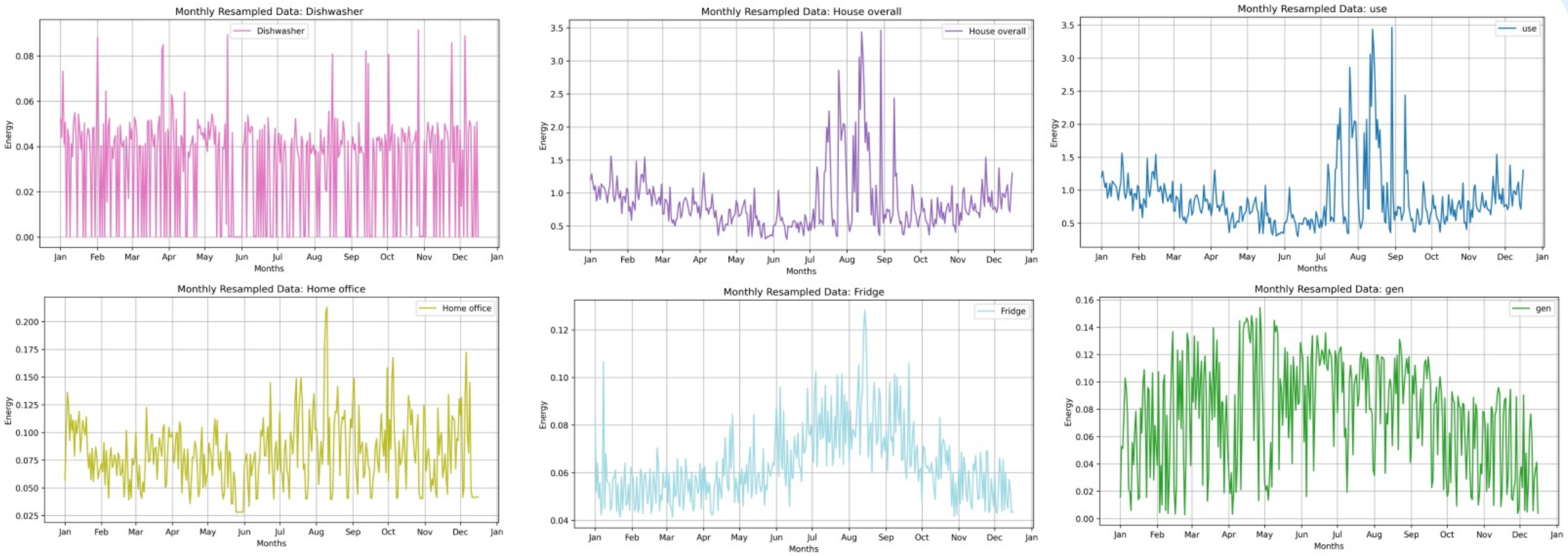
Weather Correlations

This Correlation matrix shows relationships between numerical variables, with a high positive correlation (0.99) between two variables and a significant negative correlation (-0.51) indicating inverse relationships, highlighting dependencies useful for analysis.

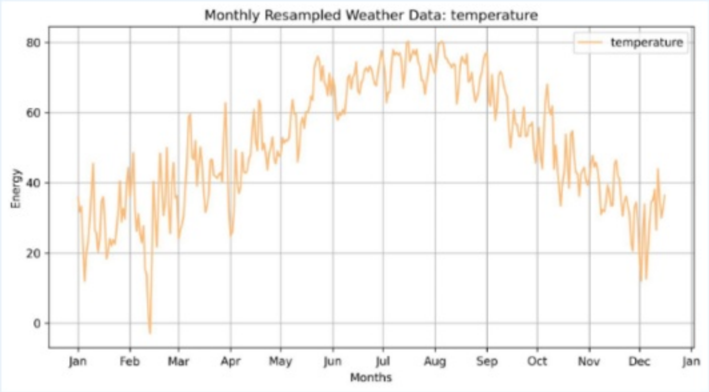
Time series Analysis

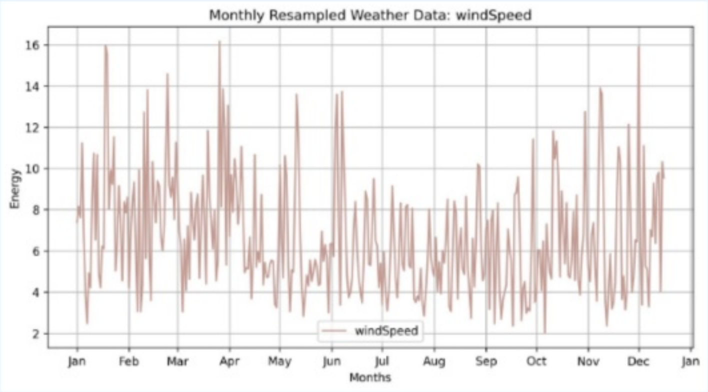
Understanding effective analysis techniques is crucial for leveraging smart home energy consumption data to uncover patterns and anomalies.

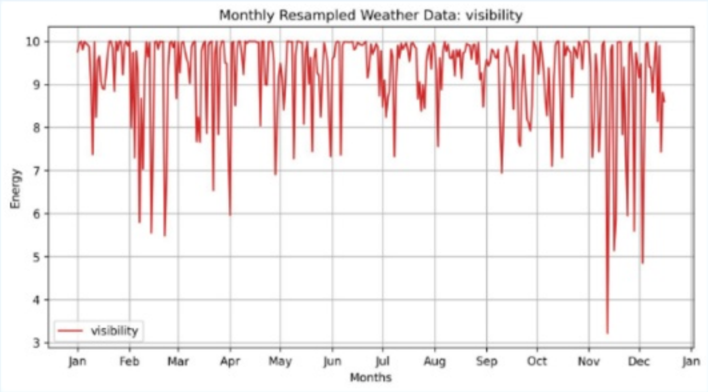
Energy Consumption analysis of house appliances

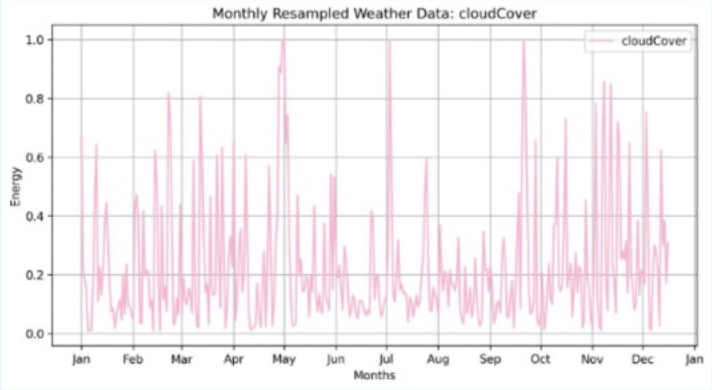


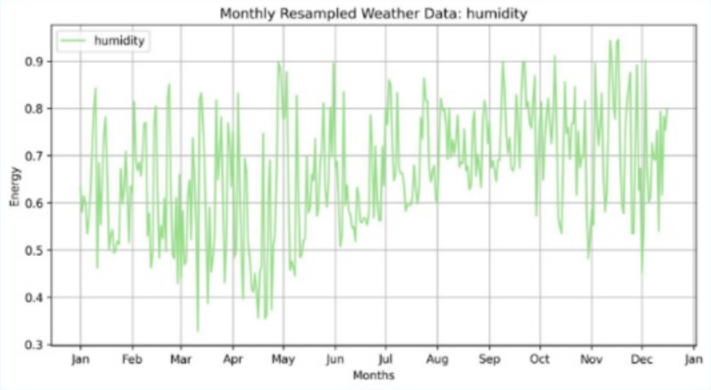
Weather Information Analysis

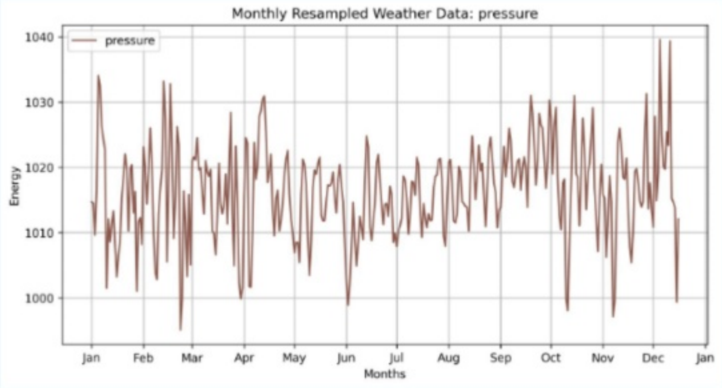












Anomaly Detection

Why anomaly detection?

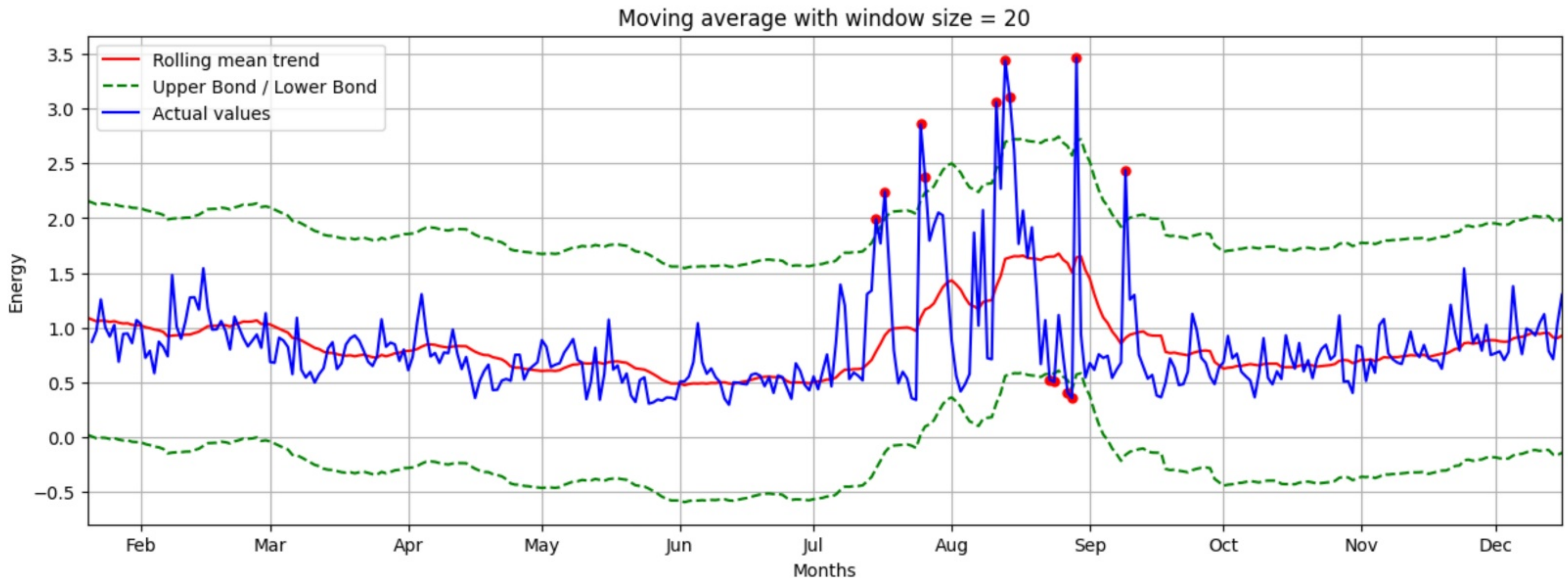
- Identify Unexecped Events
- Optimize Resource Usage
- Preventative Maintenance
- Improve Energy Management
- Correlate with External Factors
- Enhanced Home Automation

Moving Average Analysis

Definition: Moving Average Analysis is a statistical technique used to smooth out short-term fluctuations and highlight long-term trends or cycles in time-series data. It helps identify patterns by averaging subsets of the dataset over a specified period.

Types: There are several types of moving averages, including simple moving average (SMA), where each data point is equally weighted, and exponential moving average (EMA), which assigns more weight to recent data points for quicker trend identification.

Applications: It is widely used in fields like finance (to analyze stock price trends), weather forecasting, inventory management, and signal processing, providing insights into trends without the noise of short-term variations.



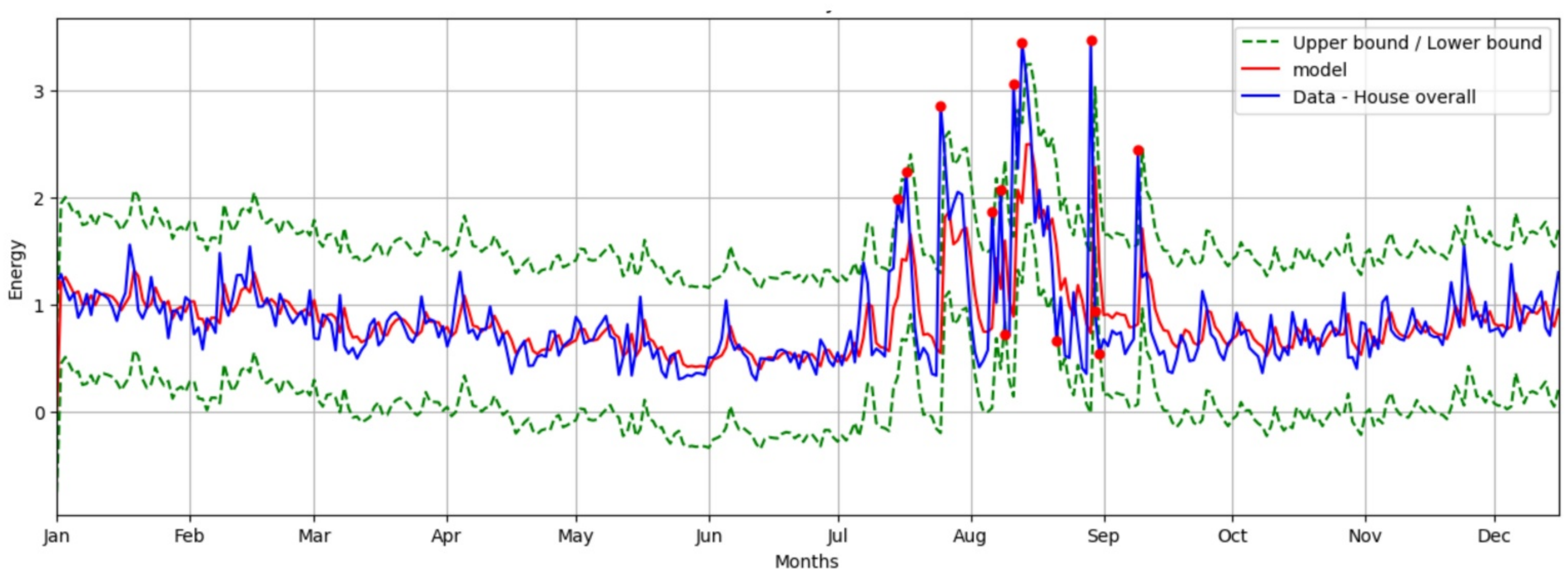
- Calculates a rolling mean over a defined window 20 days smooths fluctuations to reveal long-term trends.
- Add upper and lower bounds to highlight normal ranges calculated using mean absolute error and standard deviation of residuals.
- Flags values that deviates significantly from normal ranges.

ARIMA (Autoregressive Integrated Moving Average)

Definition and Components: AutoRegressive Integrated Moving Average (ARIMA) is a statistical modeling technique used for analyzing and forecasting time-series data. It combines three components: AutoRegressive (AR), which models the relationship between current and past values; Integrated (I), which involves differencing to achieve stationarity; and Moving Average (MA), which models the relationship between current values and past forecast errors.

Key Features: ARIMA is versatile and can handle non-stationary data by differencing. The model is parameterized by three integers (p, d, q), representing the order of the AR, I, and MA components respectively. It requires careful selection of these parameters using methods like ACF, PACF, or grid search.

Applications: ARIMA is widely used for forecasting in various fields, including finance (stock prices), economics (GDP growth), inventory management, and demand prediction. It is effective for univariate time series with patterns or trends.



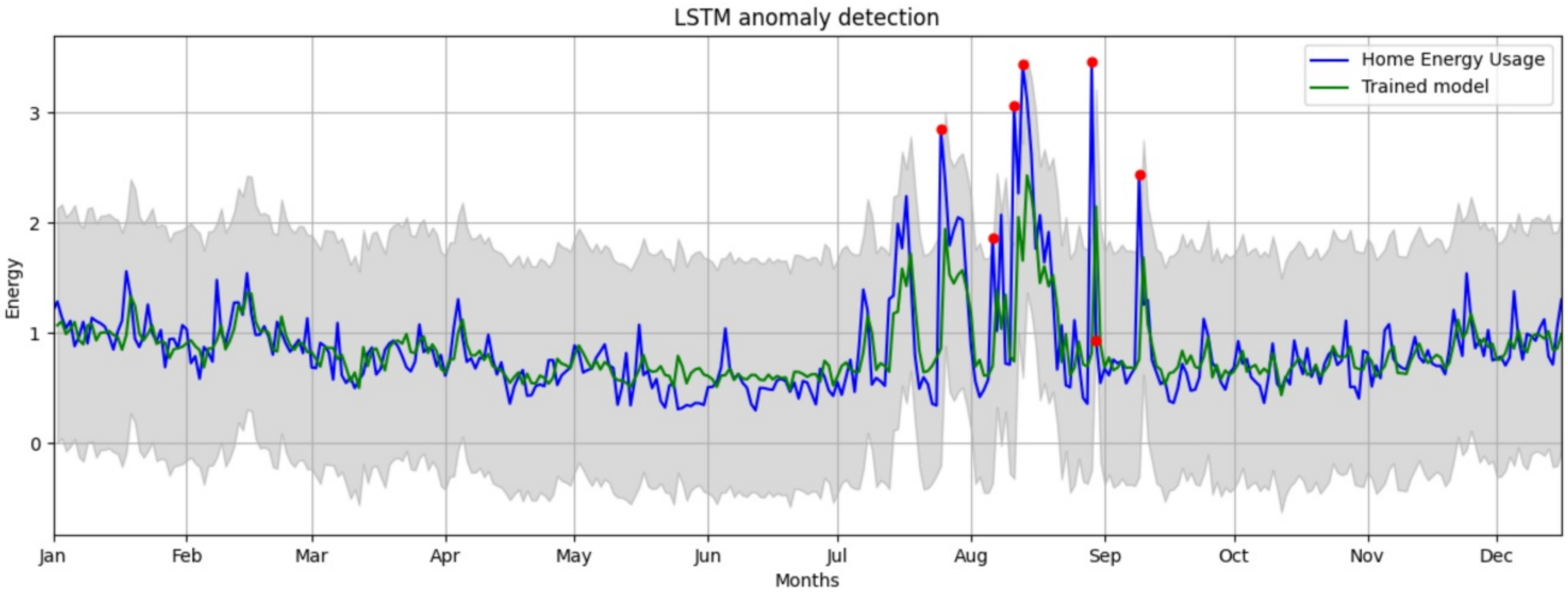
- The Integrated (I) component in ARIMA makes it capable of handling non-stationary data by applying differencing. This allows ARIMA to model trends and seasonality effectively, while MA assumes the data is stationary and cannot handle these patterns.
- ARIMA is specifically designed for forecasting future values, making it useful for time-series predictions. Additionally, the inclusion of prediction intervals enables ARIMA to identify anomalies (values outside the bounds), which is not possible with simple MA.

LSTM(Long Short-Term Memory networks)

Sequential Data Modeling: LSTMs (Long Short-Term Memory networks) are designed to model sequential dependencies, making them ideal for detecting anomalies in time-series data. They can capture long-term patterns and relationships, identifying deviations that do not align with learned normal behavior.

Reconstruction and Prediction Errors: LSTMs are commonly used in autoencoder architectures or forecasting frameworks. The network learns to reconstruct or predict sequences, and anomalies are flagged when reconstruction or prediction errors exceed a defined threshold.

Dynamic Adaptability: LSTMs adapt to complex temporal patterns, including non-linear trends, seasonality, and noise, enabling accurate anomaly detection in diverse domains such as financial transactions, network security, and industrial equipment monitoring.



- Unlike MA and ARIMA, which primarily model linear relationships, LSTMs can learn **nonlinear and complex temporal dependencies**. This makes them effective for datasets with intricate patterns or sudden changes, such as energy usage spikes.
- LSTMs excel at capturing long-term dependencies in the data due to their architecture, which includes memory cells. This allows them to retain information from past time steps, even if the relationship spans a long time, whereas ARIMA and MA models focus on short-term dependencies.

Future Work

Clustering: Group similar energy usage patterns or weather conditions to identify distinct user behavior profiles or seasonal trends.
For example, classify high-consumption days based on weather parameters (e.g., temperature, humidity).

Regression: Use regression models to predict energy consumption based on weather conditions and other factors, improving forecasting accuracy and actionable insights.

Conclusion

- **Seasonal analysis** reveals that energy consumption spikes in both summer (cooling) and winter (heating).
- **Insights and Optimization:** Identified energy spikes for cost-saving and operational improvements like the furnace and microwave highlight areas for potential optimization.
- **Explored Smart Home Energy Usage:** Applied models to detect trends and anomalies effectively.
- **Model Comparison:** LSTM excelled with multivariate inputs and non-linear dependencies.

References

1. M. Kuzlu, M. Pipattanasomporn, and S. Rahman, "Review of Communication Technologies for Smart Homes/ Building Applications," IEEE Innovations in Smart Grid Technologies (ISGT), Washington, DC, USA, 2010.

2. M. D. Tascikaraoglu and S. Boynuegri, "Smart Home Energy Management System Based on an Internet of Things Approach," IEEE Access, 2019.

3. R. Herero, C. Andujar, and M. Molina, "Energy Data Analysis for Smart Homes: Techniques for Data Preprocessing and Clustering," IEEE Transactions on Industrial Informatics, vol. 16, no. 1, pp. 110-120, Jan. 2020.

4. S. S. Ozturk, B. T. Yildirim, and S. Ekinci, "Exploratory Data Analysis of IoT-Based Smart Home Data," IEEE 3rd International Conference on Data Science and Machine Learning Applications (CDMA), Riyadh, Saudi Arabia, 2020.

5. G. Marques, R. Agarwal, and R. P. de Jesus, "Exploratory Data Analysis and Visualization of a Smart Home Monitoring System," IEEE Access, 2020.