**Comprehensive Report on Household Electricity Consumption**

****

**Introduction:**

This report summarizes the findings, challenges encountered, and recommendations for optimizing household electricity consumption based on the analysis done on household consumption data.

**Objectives:**

The main objectives of the study were:

* To build time series models and evaluate their metrics.
* To predict future household electricity consumption.

**Data Description:**

The dataset used for the analysis included the following features:

* Date: Date of the electricity consumption recording.
* Time: Time of the electricity consumption recording.
* Global\_active\_power: Total active power consumed by the household.
* Global\_reactive\_power: Total reactive power consumed by the household.
* Voltage: Voltage level during the electricity consumption period.
* Global\_intensity: Total current intensity consumed by the household.
* Sub\_metering\_1: Electricity consumption in sub-metering 1 (e.g., kitchen).
* Sub\_metering\_2: Electricity consumption in sub-metering 2 (e.g., laundry).
* Sub\_metering\_3: Electricity consumption in sub-metering 3 (e.g., water heater).

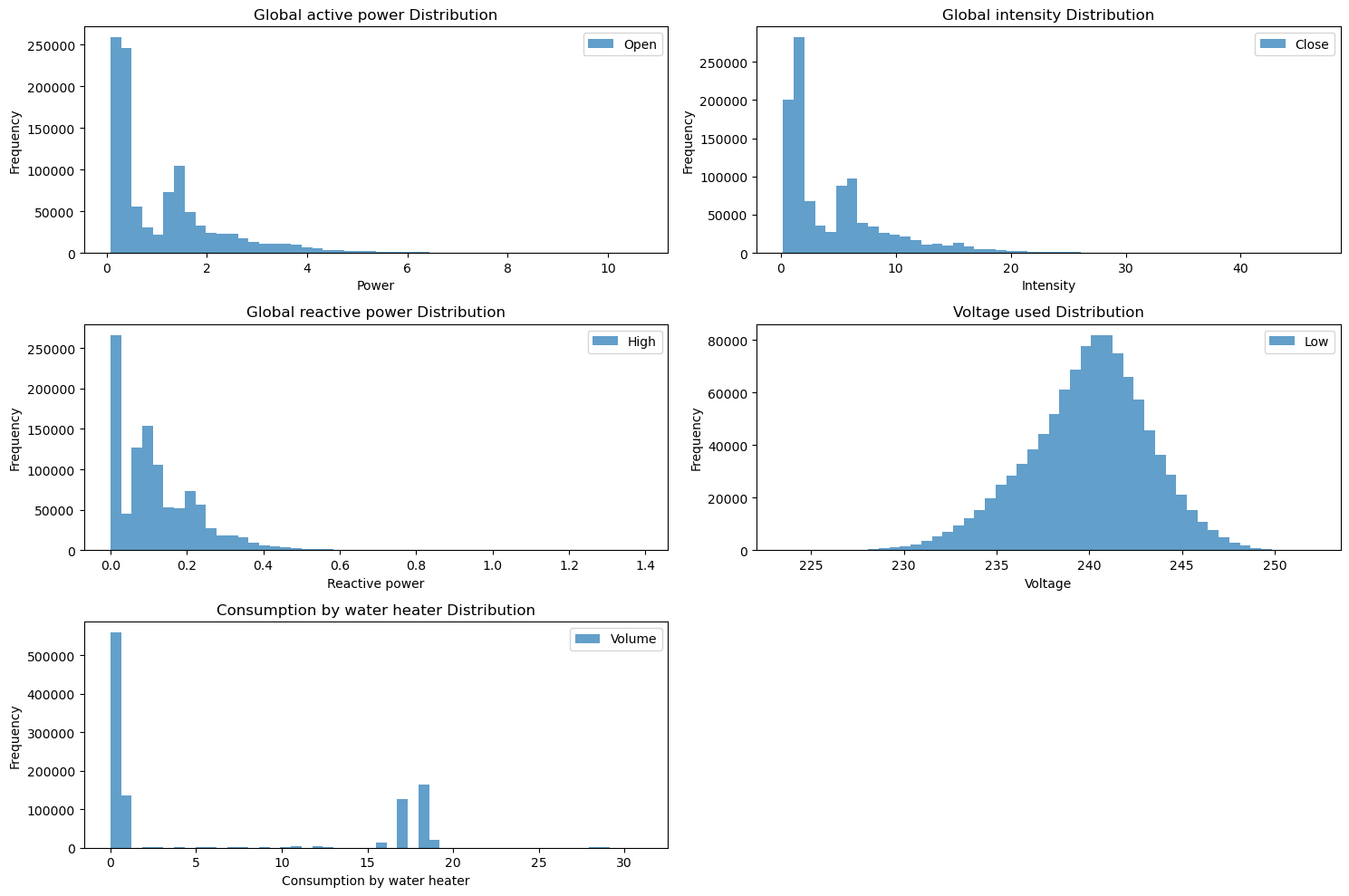
**Data Preparation:**

The data preparation steps included:

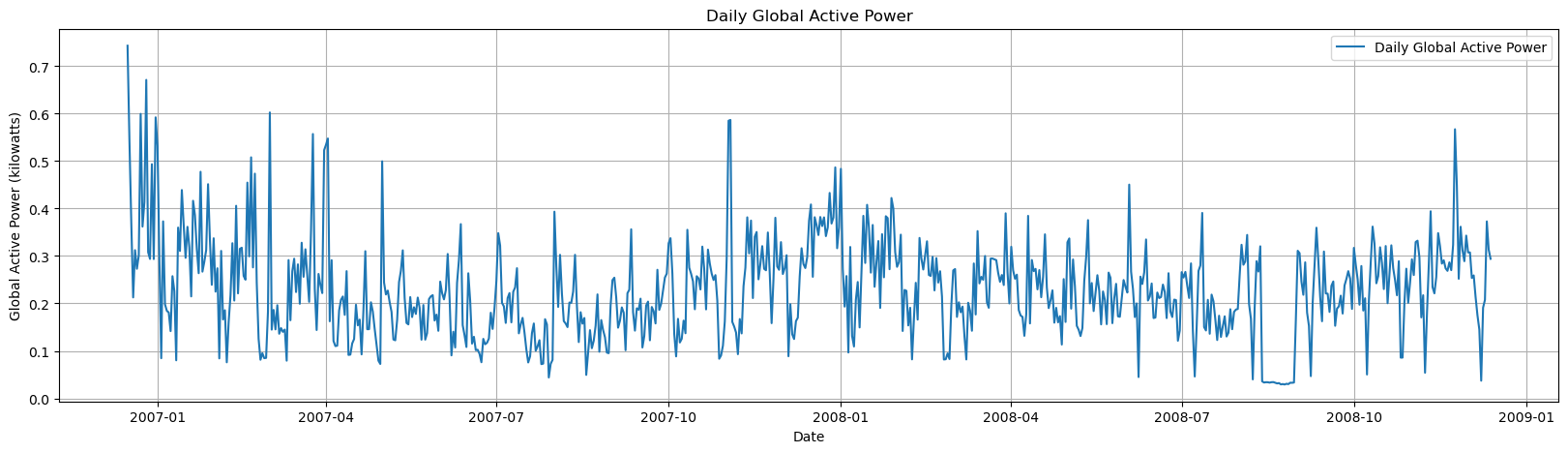
* Checking for null values: No null values were found.
* Converting data to appropriate types.
* Filtering outliers.
* Normalizing the data to bring it to the same scale (0-1).

**Exploratory Data Analysis (EDA):**

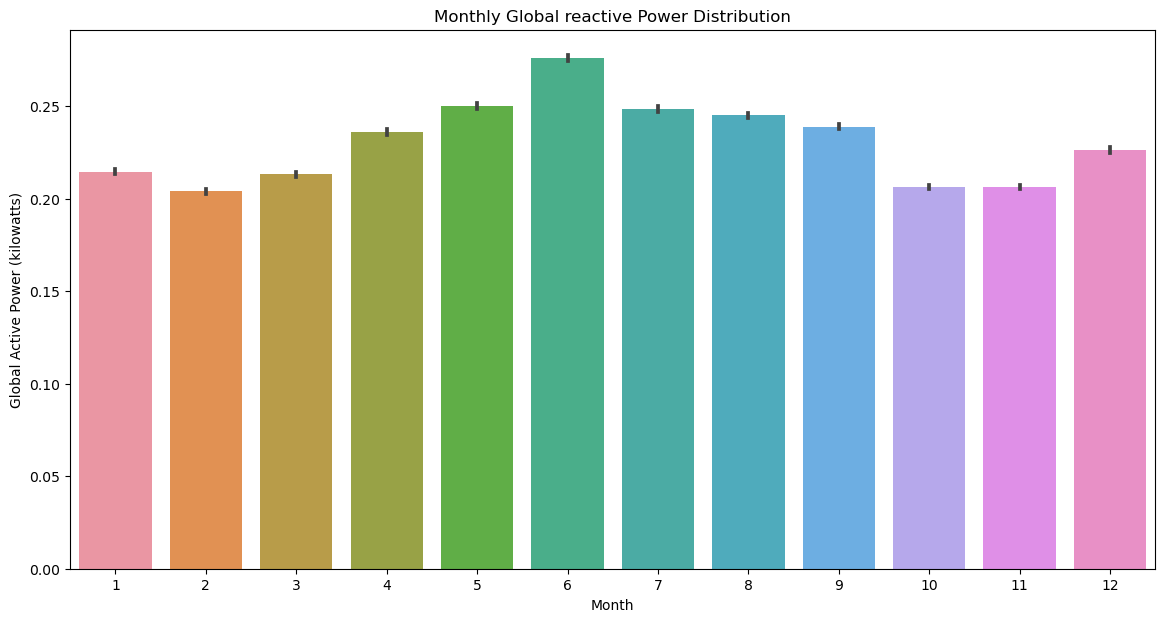
The EDA included distribution visualizations that revealed key insights:



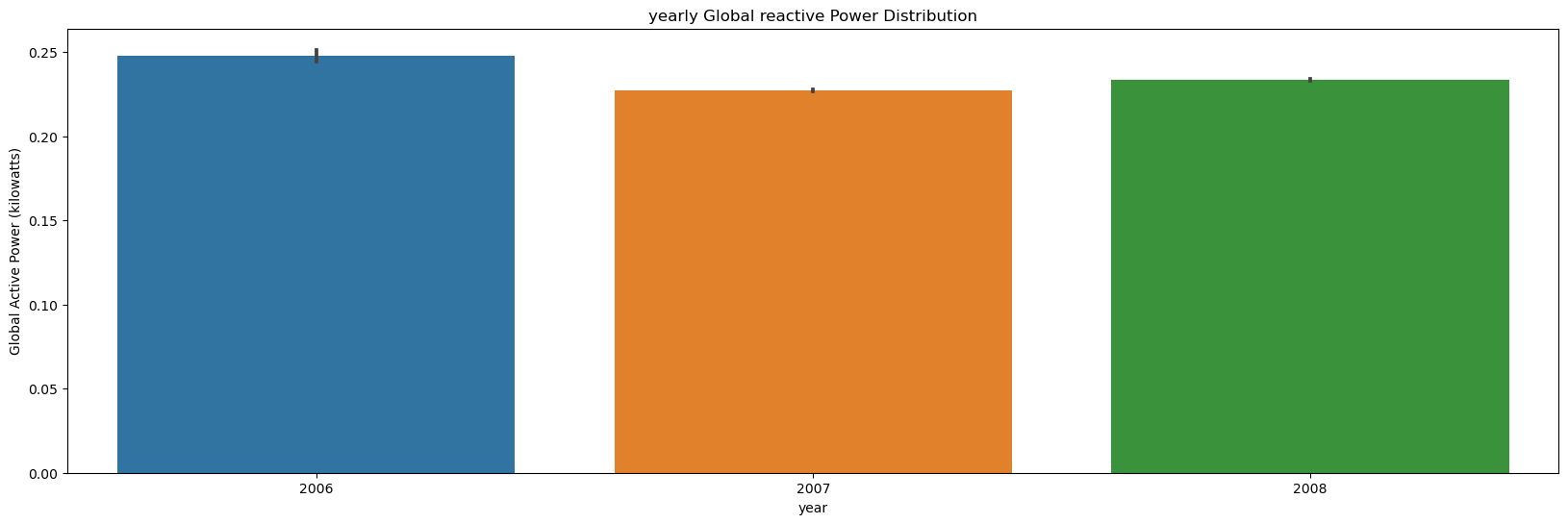
* Global active power and global intensity: Right-skewed, indicating most consumption values were low with occasional high spikes.
* Reactive power: Similar pattern with generally lower values.
* Voltage: Approximately normal distribution, showing stable voltage levels around 240 volts.
* Sub\_metering\_3 (water heater consumption): Frequently registered zero, indicating many periods of no usage with occasional high consumption peaks.



* Daily consumption visualizations: Showed an increasing and decreasing trend with a noticeable pattern during the middle and end of the year. There was a spike in consumption at the beginning of each year with reduced consumption.

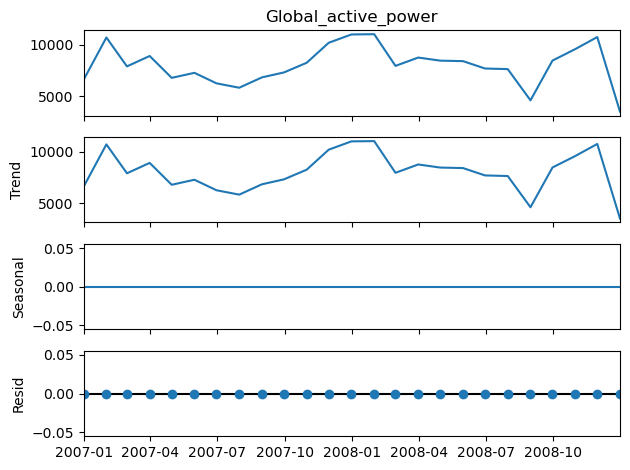


* Monthly consumption indicates the electricity consumption hiked during thr 6th month and the least used in the 2nd month



* yearly consumption visualizations: Indicated the highest consumption in 2006, a reduction in 2007, and a slight increase in 2008.

**Time Series Decomposition:**

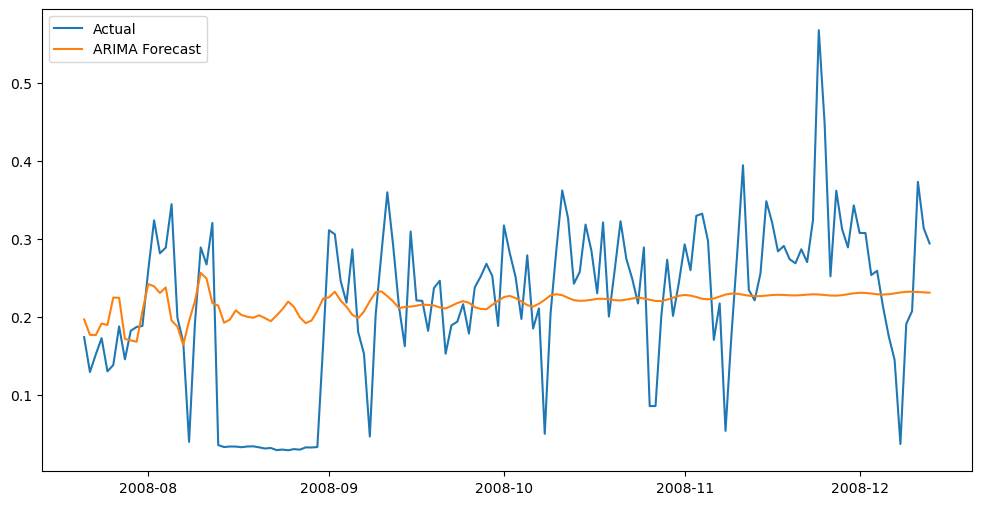
****

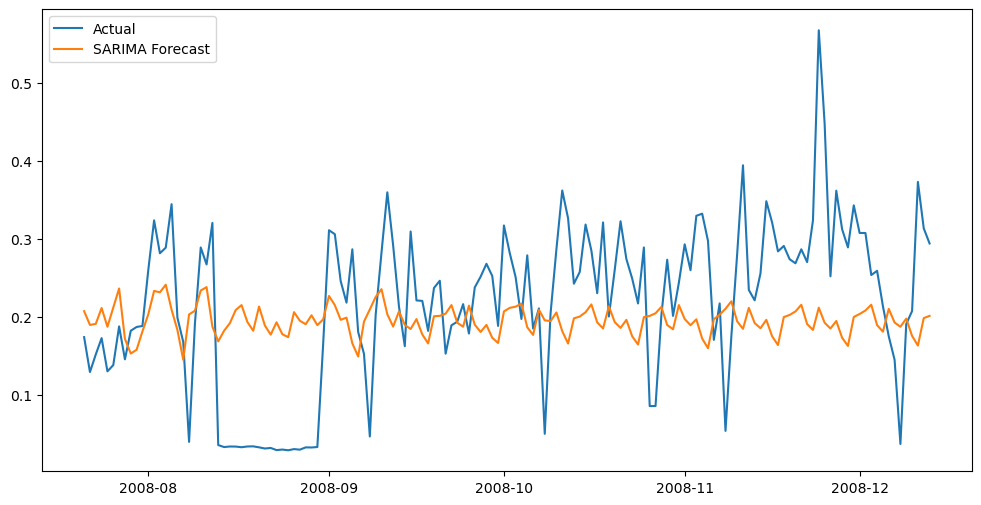
The time series decomposition resampled to monthly data revealed no significant trend or seasonality in the electricity consumption.

**Predictive Modelling:**

The predictive models used included ARIMA and SARIMA models:

* ARIMA Model: RMSE score of 0.0957, indicating a relatively low error in the model's predictions.
* SARIMA Model: RMSE score of 0.1053, indicating a slightly higher prediction error compared to the ARIMA model.
* Both models performed well, but the ARIMA model had a slight edge in prediction accuracy.



****

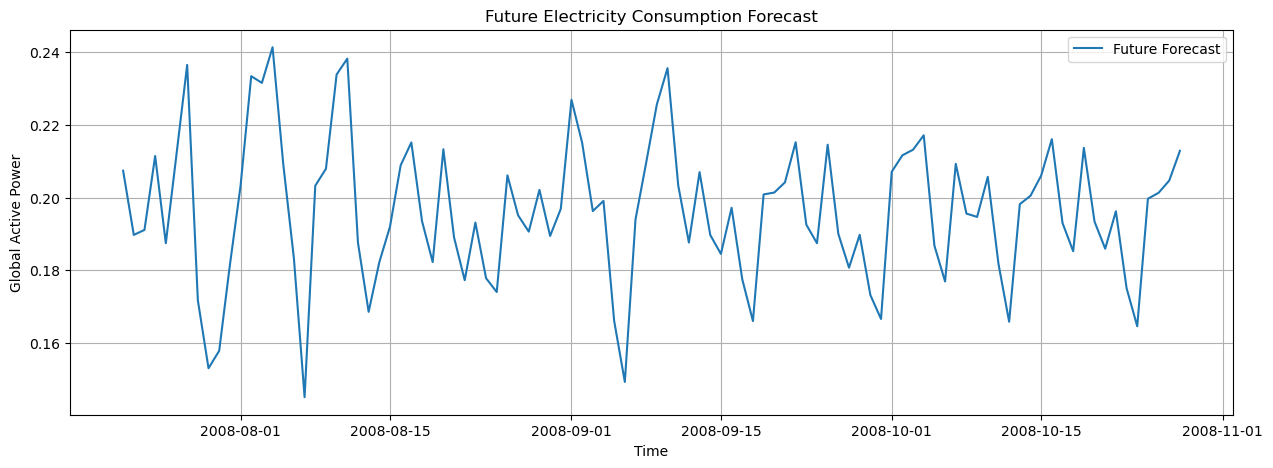
**Feature Engineering:**

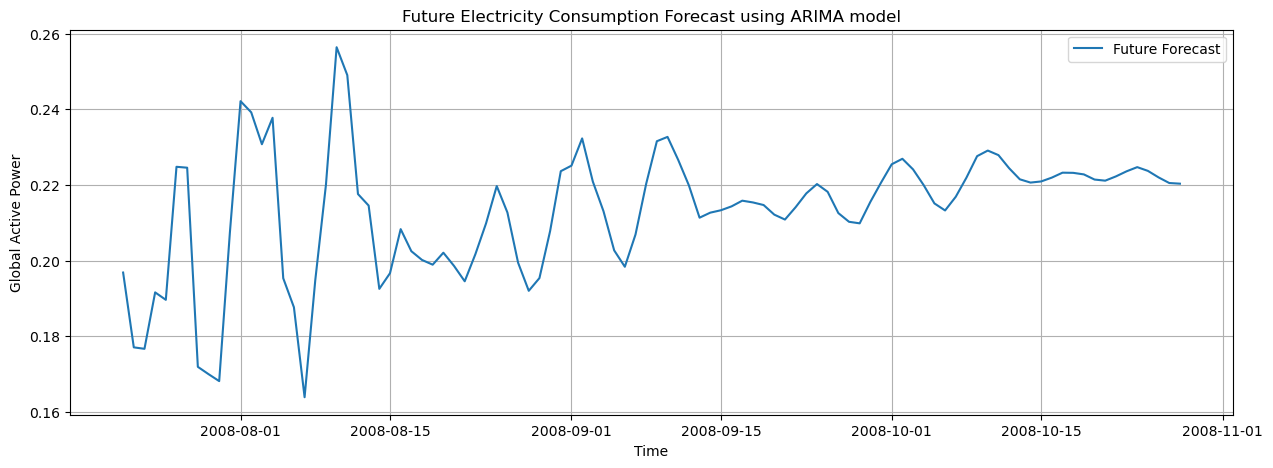
Rolling statistics features and critical features only were created as part of the feature engineering process.

* The moving average and data variability were captured via the creation of rolling mean and standard deviation features.
* Use a correlation matrix to identify key characteristics, such as "global intensity" and "sub metering 3," which have a strong correlation with global active power.

**Future Consumption Prediction:**

* Both the ARIMA and SARIMA models were used to anticipate future consumption and the results showed how well the models could predict future power consumption numbers.





**Challenges Encountered:**

The main challenges encountered during the analysis included:

* Managing the consumption data's right-skewedness.
* Ensuring scaling and standardization of data in order to train models accurately.
* Huge amount of data with lot of cleaning procedures especially in date and time column
* Memory issues so couldn’t perform advanced models like LSTM.

**Recommendations:**

Based on the analysis, the following recommendations are made to optimize household electricity consumption:

* Regularly track and evaluate consumption trends: Determine peak use times and areas where you could save money.
* Use energy-saving equipment and procedures: Cut back on consumption all around.
* Apply forecasting models: Project future consumption and make plans in line with it.
* Promote energy-saving practices: Particularly during times of high use.