**Measure Energy Consumption**

**Problem Statement:**

The measurement of energy consumption is critical in understanding and optimizing energy usage in various sectors, including manufacturing sites, homes, commercial buildings, and transportation. However, the manual collection and analysis of energy consumption data can be time-consuming and error-prone. Therefore, there is a need for an automated approach to collect, analyze and visualize energy consumption data for better decision-making.

**Phase 1: Problem Definition and Design Thinking**

**Problem Definition:**

The problem at hand is to create an automated system that measures energy consumption, analyzes the data, and provides visualizations for informed decision-making. This solution aims to enhance efficiency, accuracy, and ease of understanding in managing energy consumption across various sectors.

**Design Thinking:**

**1. Data Source:**

We are choosing a dataset from kaggle this dataset contains information related to electricity load and power consumption in various regions, possibly within the PJM (PJM Interconnection) electric grid system in the United States. Here's a brief interpretation of some of the column names:

**COMED\_MW:** Likely electricity load data for the Commonwealth Edison (COMED) service area.

**DAYTON\_MW:** Likely electricity load data for the Dayton Power and Light (DAYTON) service area.

**DEOK\_MW:** Likely electricity load data for the Duke Energy Ohio and Kentucky (DEOK) service area.

**DUQ\_MW:** Likely electricity load data for the Duquesne Light (DUQ) service area.

**EKPC\_MW:** Likely electricity load data for the East Kentucky Power Cooperative (EKPC) service area.

**FE\_MW:** Likely electricity load data for the FirstEnergy (FE) service area.

**NI\_MW:** Likely electricity load data for the Northern Indiana (NI) service area.

**PJM\_Load\_MW:** Likely the aggregated electricity load data for the entire PJM Interconnection region.

**AEP, PJME, PJMW:** These abbreviations likely refer to other specific regions or utilities within the PJM system.

**Dataset Link:**[**https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption**](https://www.kaggle.com/datasets/robikscube/hourly-energy-consumption)

**2.Data Preprocessing:**

Clean, transform, and prepare the dataset for analysis.

**Handling Missing Values:** Address missing data by either removing or filling missing values.

**Data Transformation:** Convert date-time columns to a consistent format and scale numerical features.

**Outlier Detection:** Identify and handle outliers using statistical methods.

**Data Splitting:** Split data into training, validation, and test sets if building models.

**Quality Assurance:** Ensure data cleanliness and accuracy throughout the process.

**3. Feature Extraction:**

Feature extraction involves selecting and transforming relevant information from the dataset to create new features that are more informative for analysis or modeling. For the energy consumption dataset you mentioned, some feature extraction ideas include:

**Time-Based Features:**  Extract day of the week, month, or year from the datetime column.

**Statistical Aggregates:** Compute statistical aggregates for each region, such as mean, median, standard deviation, or percentiles, to summarize energy consumption patterns.

**Seasonal Trends:** Create features to represent seasonal variations, such as temperature data if available, to account for weather-related energy consumption patterns.

**Correlations:** Compute correlations between energy consumption in different regions to identify regions with similar or opposite trends.

**Frequency Domain Features:** Apply Fourier or wavelet transforms to reveal periodic patterns in the data.

**4. Model Development:**

Certainly, here's a concise step-by-step guide for developing a Linear Regression model for energy consumption prediction:

**Data Preparation:** Clean and preprocess the dataset.

**Data Splitting:** Split data into training and testing sets.

**Feature Selection:** Choose relevant features.

**Model Selection:** Choose Linear Regression for prediction.

**Model Training:** Train the Linear Regression model.

**Model Evaluation:** Assess model performance using metrics like MAE, MSE, RMSE, and R-squared.

**Visualization:** Visualize predictions vs. actual values.

**Interpretation:** Analyze coefficients for feature impact.

**Fine-Tuning:** Optimize the model as needed.

**Deployment and Monitoring:** Deploy for predictions and monitor over time.

**Communication:** Share results and recommendations with stakeholders.

**5.Visualization:**

Certainly, here are five types of concise visualizations for presenting energy consumption trends and insights:

**Time Series Plot:** Show long-term consumption trends and seasonality.

**Residual Plot:** Display model errors to identify patterns.

**Feature Importance Plot:** Highlight the impact of each feature on consumption.

**Histogram of Residuals:** Check the distribution of model errors.

**Regression Line Plot:** Visualize the linear relationship between a key feature and consumption.

**6. Automation:**

Certainly, here's a concise overview of automating data collection, analysis, and visualization:

**Data Collection:** Automate data retrieval from online sources.

**Data Analysis:** Develop scripts for preprocessing, modeling, and analysis.

**Visualization:** Create automated chart and graph generation.

**Integration:** Combine data collection, analysis, and visualization into a single workflow.

**Scheduled Execution:** Set up automatic execution at defined intervals.

**Error Handling:** Implement error checks and logging for reliability.

**Continuous Improvement:** Regularly update and optimize the automation process.