**Automated Energy Consumption Analysis System**

**Problem Definition and Design Thinking Document**

**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. The data source for this project is hourly power consumption data from PJM Interconnection LLC (PJM), which operates an electric transmission system serving multiple regions in the United States.

**Design Thinking**

To solve the problem of automated energy consumption analysis, we will follow a structured approach:

**1. Data Source Identification**

* The primary data source for this project is PJM's website, which provides hourly power consumption data in megawatts (MW).

**2. Data Preprocessing**

* Clean, transform, and prepare the dataset for analysis. This step involves handling missing values, data format consistency, and removing outliers.

**3. Feature Extraction**

* Extract relevant features and metrics from the energy consumption data. These features may include time of day, day of the week, and seasonality factors.

**4. Model Development**

* Utilize statistical analysis to uncover trends, patterns, and anomalies in the data. This can involve time series analysis, clustering, and regression modeling to understand the factors affecting energy consumption.

**5. Visualization**

* Develop visualizations (graphs, charts) to present the energy consumption trends and insights. Visualization is crucial for stakeholders to easily understand and interpret the data. Possible visualizations include line charts, heatmaps, and geographical maps.

**6. Automation**

* Build a script that automates data collection, analysis, and visualization processes. This script should be able to fetch the latest data from PJM's website, preprocess it, apply the developed models, and generate updated visualizations.

**Data Source Details**

The data source for this project is PJM Interconnection LLC (PJM), a regional transmission organization (RTO) in the United States. PJM operates an electric transmission system serving the following regions:

* + Delaware
  + Illinois
  + Indiana
  + Kentucky
  + Maryland
  + Michigan
  + New Jersey
  + North Carolina
  + Ohio
  + Pennsylvania
  + Tennessee
  + Virginia
  + West Virginia
  + District of Columbia

Please note that the regions served by PJM have changed over the years, so data may only appear for certain dates per region. This historical data is provided in megawatts (MW) and is available in hourly intervals.

**Next Steps**

**The next steps in this project involve the implementation of the design thinking process:**

**1. Data Collection and Preprocessing**

* Develop a data collection script to regularly fetch the latest energy consumption data from PJM's website.
* Preprocess the data to handle missing values, format consistency, and outliers.

**2. Feature Extraction**

* Identify relevant features that can provide insights into energy consumption trends .
* Extract and engineer these features from the dataset.

**3. Model Development**

* Implement statistical models to analyze the data, including time series analysis and predictive modeling.
* Detect anomalies and unusual consumption patterns.

**4. Visualization**

* Create interactive visualizations and dashboards to present the analysis results.
* Ensure that stakeholders can easily access and interpret the visualizations.

**5. Automation**

* Build an end-to-end automation script that integrates data collection, preprocessing, modeling, and visualization.
* Schedule the script to run at regular intervals for real-time insights.

By following this structured approach, we aim to create an automated energy consumption analysis system that empowers decision-makers with valuable insights for efficient energy management across various sectors served by PJM.