**Household Energy Usage Analysis**

**Overview**

This project analyzes household energy usage data to uncover meaningful patterns and insights through clustering techniques. By exploring the relationships between key metrics such as **Global Active Power (kW)** and **Global Intensity**, we aim to group energy usage behaviours into clusters. These clusters help identify trends, anomalies, and potential areas for optimization in household energy consumption.

The project leverages clustering algorithms (e.g., K-means, hierarchical clustering) to segment the data and provide actionable insights into energy usage. The results are visualized through scatter plots and other graphical representations, highlighting the distribution and correlation of key variables.

**Objective**

The primary goals of this project are:

1. To identify patterns and trends in household energy consumption.
2. To cluster data points into distinct groups based on their energy usage characteristics.
3. To visualize the relationships between key energy metrics.
4. To provide actionable insights for energy optimization and anomaly detection.

**Dataset**

The dataset used for this analysis contains real-world energy consumption data, including:

* **Global Active Power (kW):** The household's total active power consumption.
* **Global Intensity:** The intensity of electricity usage.
* **Other Metrics:** Sub-metering data, voltage levels, and reactive power.

**Data Preprocessing**

1. **Handling Missing Values:** Missing or inconsistent data points were handled through imputation or removal to ensure a clean dataset.
2. **Feature Scaling:** Numerical features were normalized to ensure clustering accuracy.
3. **Feature Selection:** Key variables like **Global Active Power** and **Global Intensity** were chosen for clustering.

**Methodology**

1. **Exploratory Data Analysis (EDA):**
   * Visualized relationships between energy metrics to understand patterns and correlations.
   * Checked for outliers and data inconsistencies.
2. **Clustering:**
   * **Algorithm Used:** K-means clustering was implemented to group data points into clusters.
   * **Optimal Cluster Determination:** The elbow method and silhouette scores were used to identify the optimal number of clusters.
3. **Visualization:**
   * Scatter plots were created to represent clusters in two dimensions, with **Global Active Power** on the x-axis and **Global Intensity** on the y-axis.
   * A color-coded legend was added to differentiate between clusters.
4. **Tools and Libraries:**
   * **Programming Language:** Python.
   * **Libraries:** Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn.
   * **Environment:** Jupyter Notebook.

**Results**

* **Clustering Output:**
  + The data was segmented into **three distinct clusters**:
    - **Cluster 0:** Represents low energy usage.
    - **Cluster 1:** Represents moderate energy usage.
    - **Cluster 2:** Represents high energy usage.
  + The clustering results provide a clear differentiation between varying energy usage patterns.
* **Correlation:**
  + A strong positive correlation between **Global Active Power** and **Global Intensity** was observed. This linear relationship indicates that higher power consumption corresponds to higher intensity usage.
* **Visualization:**
  + The scatter plot clearly illustrates the distribution of data points across the clusters, with distinct color codes for each group.

**Conclusion**

This project successfully demonstrated the application of clustering techniques to household energy consumption data and also shows that the energy consumption on weekends is more than the weekdays for climate control systems which is sub-metering 3 which is more at the nighttime as compared to the midday time. The key findings are:

1. Energy consumption patterns can be effectively grouped into clusters, providing insights into typical household behaviors.
2. The identified clusters can help energy providers design tailored strategies for energy management, such as:
   * Targeting **low-usage clusters** for promoting energy-saving products.
   * Advising **high-usage clusters** on strategies for reducing consumption during peak hours.
3. The strong correlation between **Global Active Power** and **Global Intensity** reaffirms the importance of monitoring these metrics for energy optimization.