Unsupervised ML

Rana Hossny Hassan
GIZA SYSTEMS | RANA.HOSNY@GIZASYSTEMS.ONMICROSOFT.COM

Contents

1. Daily Energy Profile Preparation:

Develop a daily energy profile in which each row represents a day, and each column corresponds to hourly energy consumption over a 24-hour period. The preparation process involves the following preprocessing steps:

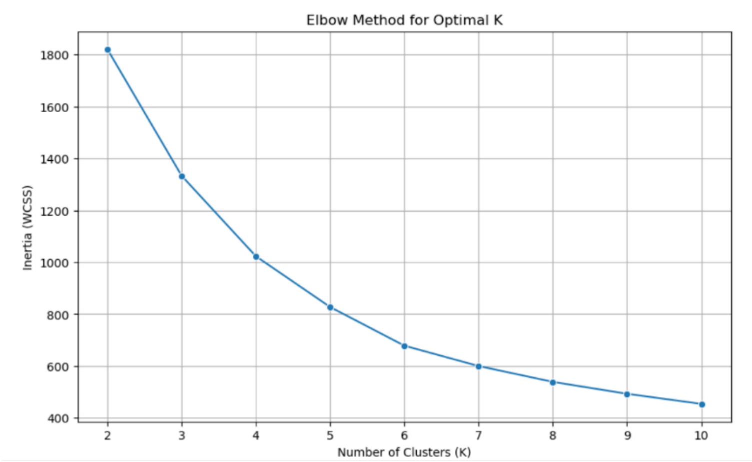
1. **Remove Duplicates**: Eliminate any duplicate entries to ensure the integrity of the data.
2. **Handle Missing Values**: Address gaps in the dataset to maintain completeness and accuracy.
3. **Manage Outliers**: Identify and mitigate the influence of outliers on the analysis.
4. **Feature Scaling**: Normalize the data to enhance consistency and improve performance in subsequent analyses.

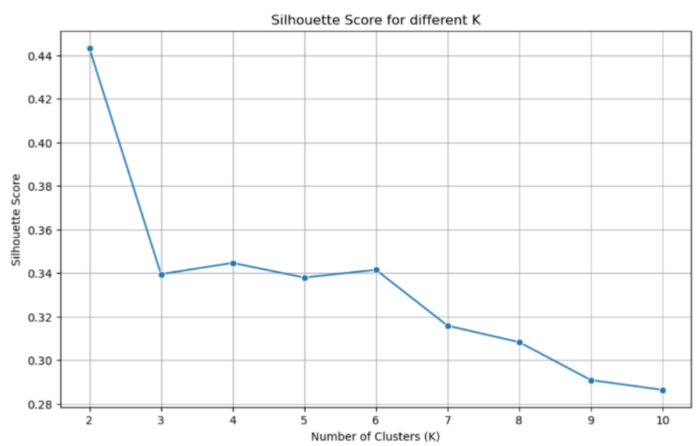
2. Clustering with K-Means:

To determine the optimal number of clusters (k) for K-Means clustering, I employed two methods:

1. **Elbow Method**: Using the elbow method, I calculated the Within-Cluster Sum of Squares (WCSS) for different values of k. The analysis indicated that the best k is 2, with a WCSS of 1800.
2. **Silhouette Method**: The silhouette method evaluates the cohesion and separation of clusters by computing the silhouette score for each k. A higher score indicates better-defined clusters. In this case, the analysis also identified k=2 as optimal, with a silhouette score of approximately 0.443.

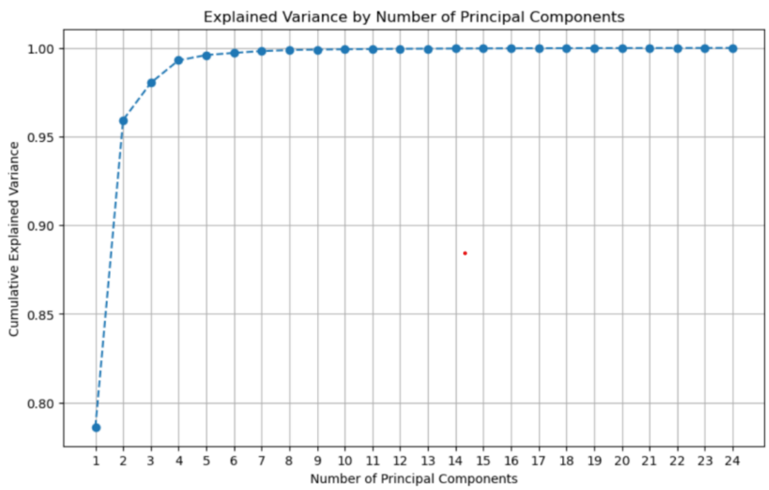
Both methods consistently suggested that k=2 is the most suitable choice for clustering the data.



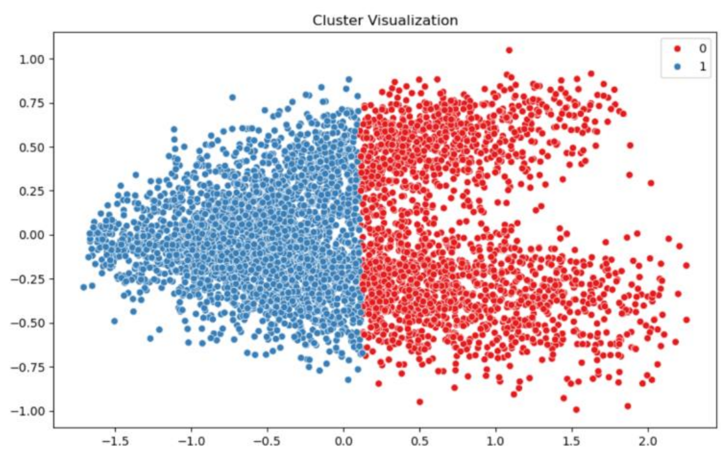


3. Visualization:

To assess the effectiveness of the clustering, Principal Component Analysis (PCA) was utilized to visualize the data. The optimal number of components was determined to be 2, which captures 95% of the explained variance.



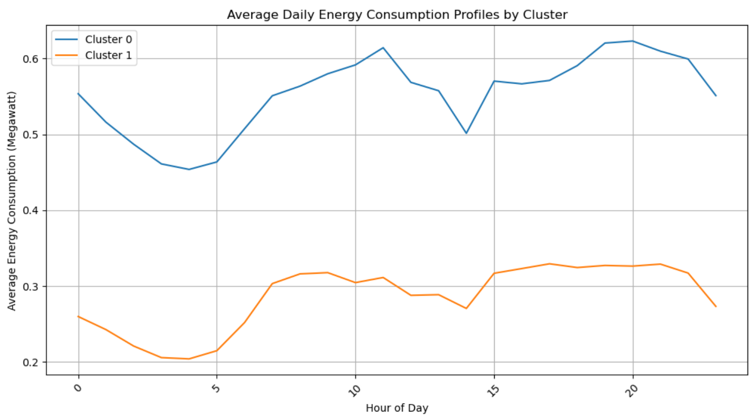
The cluster visualization for k=2:

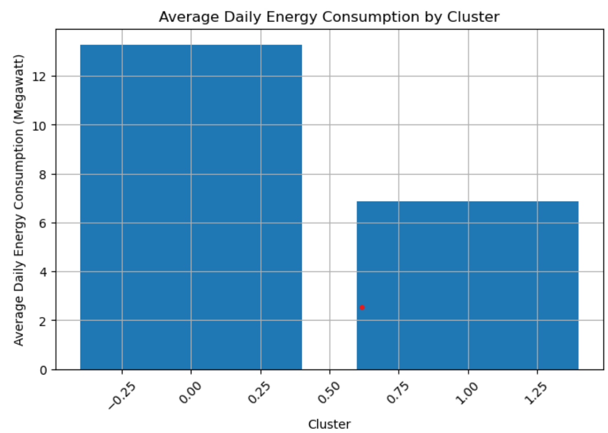


4. Cluster Result Analysis:

For the optimal number of clusters (k=2), the resulting clusters were interpreted as follows:

- **Cluster 0**: This cluster represents periods of high energy usage during peak hours. It indicates times when demand is significantly elevated, likely due to increased activity in residential or commercial settings.
- **Cluster 1**: This cluster signifies low energy usage throughout the day. The data reflects a more stable consumption pattern, suggesting minimal fluctuations in energy demand, typical of off-peak hours or periods of reduced activity.

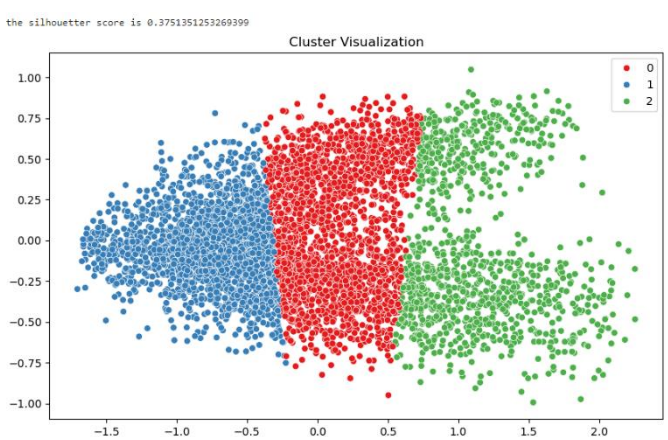


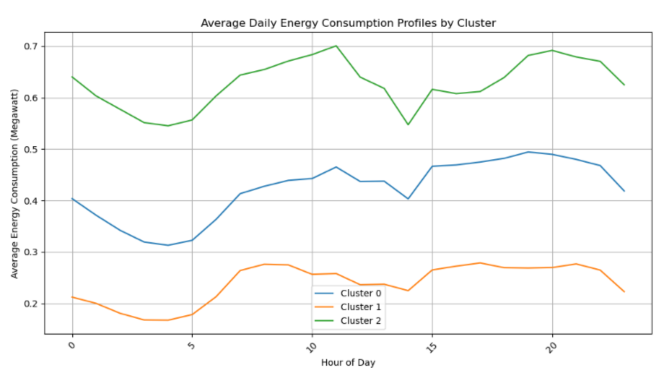


5. Tring different k:

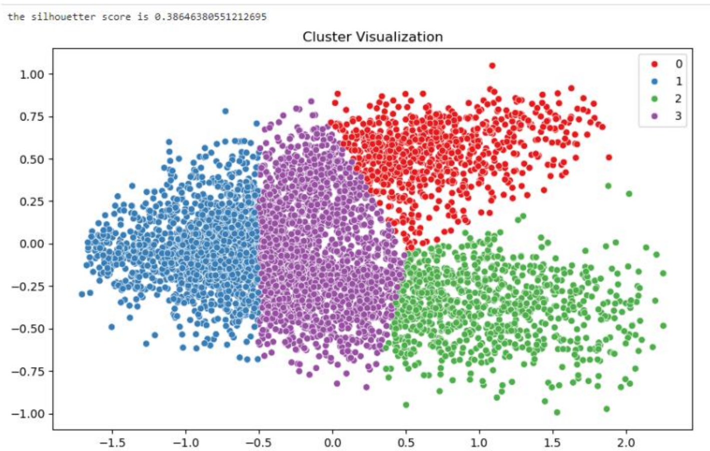
When exploring different values for k, we tested both k=3 and k=4 to gain deeper insights into the energy consumption patterns.

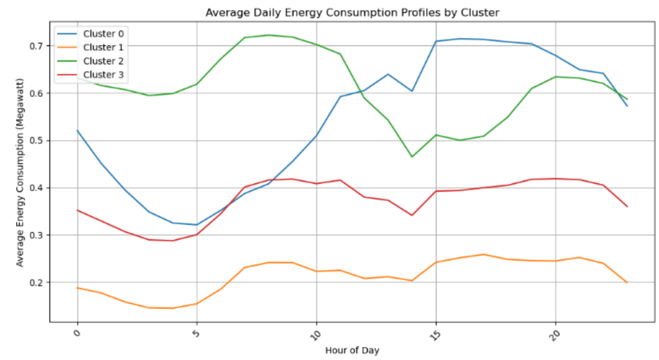
- **k=3**: This configuration allows us to categorize the data into three distinct clusters:
    - **Cluster 0**: Represents high energy usage during peak hours, characterized by significant demand spikes.
    - **Cluster 1**: Represents moderate energy usage throughout the day, indicating a balanced consumption pattern.
    - **Cluster 2**: Represents low energy usage, showing minimal fluctuations and typically associated with off-peak times.





- **k=4**: Attempting to define four clusters resulted in difficulties in interpreting the outcomes. The additional cluster did not provide clear





6. Optional Extensions: DBSCAN Clustering Analysis

1. The initial application of the DBSCAN algorithm yielded a low silhouette score of approximately **0.238**, indicating suboptimal clustering performance, likely due to inadequate hyperparameter settings.
2. After fine-tuning the hyperparameters, the optimal values were determined to be **eps = 0.700** and **min_samples = 2**, resulting in an improved silhouette score of **0.366**.
3. Despite this enhancement, the silhouette score for the tuned DBSCAN model remained lower than that of K-Means, suggesting that K-Means provided better-defined clusters in this analysis.
4. This comparison highlights the effectiveness of K-Means as a more suitable choice for clustering this specific dataset.

7. Data Insights:

1. **Weekday vs. Holiday Energy Usage**:

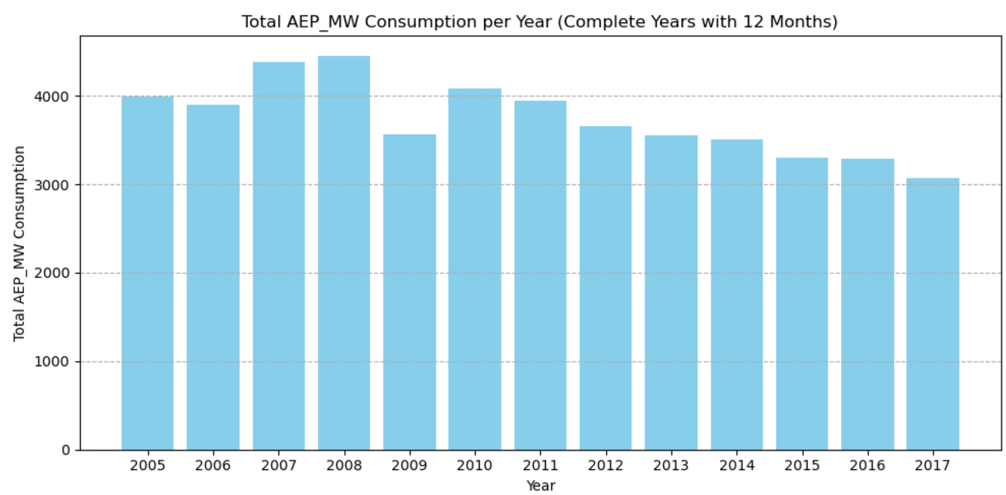
    - The average energy usage on weekdays (0.431719 AEP_MW) is higher than on holidays (0.396524 AEP_MW). This increase during weekdays can be attributed to heightened industrial, commercial, and residential activities.

2. **Monthly Average Energy Usage**:

    - Analysis of monthly energy consumption shows that January, December, July, and August consistently exhibit the highest power consumption rates across most years. This trend highlights a strong correlation between weather conditions and energy usage, emphasizing the impact of seasonal variations on power consumption.

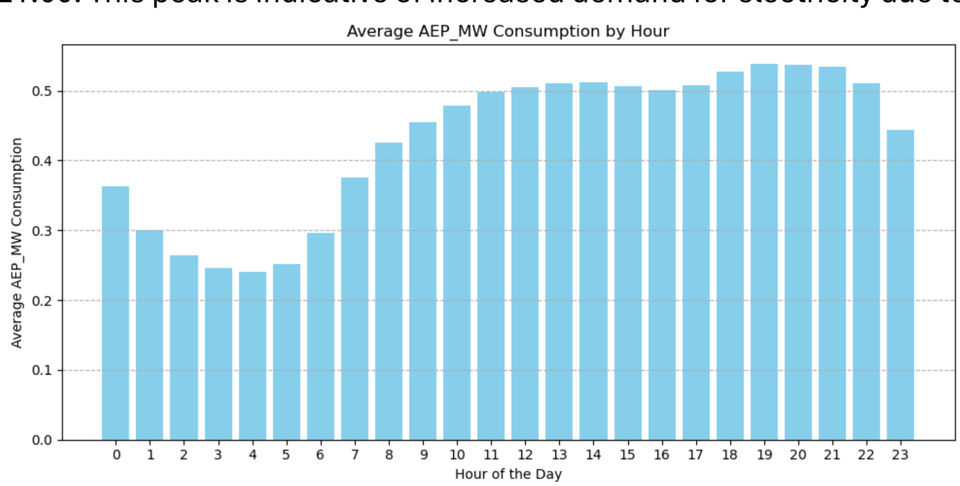
3. **Average Energy Consumption per Year**:

    - The year 2009 marked the lowest power consumption following a peak in 2008. Although there was a slight increase in 2010, energy consumption continued to decline thereafter. The 2009 recession significantly impacted all economic sectors in the EU, leading to reduced energy demand and demonstrating the relationship between economic activity and energy consumption. Post-2010, there has been increased awareness of energy consumption, which drives efforts toward energy efficiency and conservation.



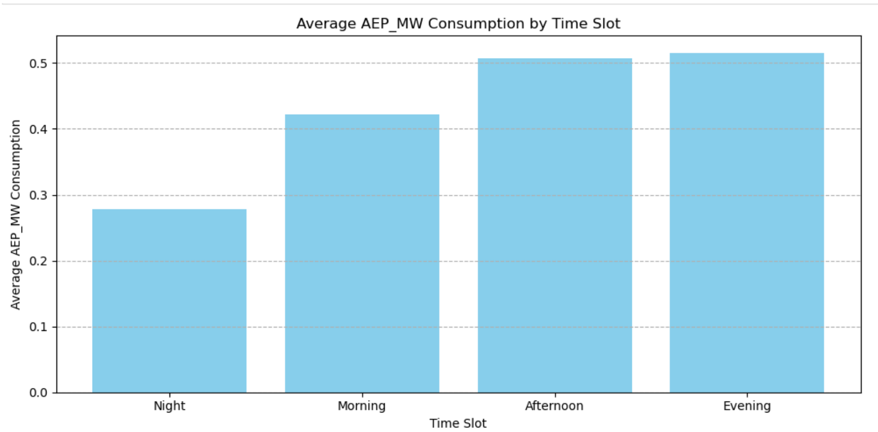
4. **Hourly Energy Consumption Patterns**:

    - The analysis reveals that the highest energy usage occurs between 19:00 and 21:00. This peak is indicative of increased demand for electricity due to



activities such as cooking, heating, and lighting as individuals return home from work.

5. **Energy Usage in Time Intervals**:
   - Energy consumption analysis across different time intervals shows maximized usage during the afternoon and evening. This trend suggests that demand peaks when people are most active, particularly in the hours following work and school.




```
Average AEP_MW Consumption during Night: 0.28 MW
Average AEP_MW Consumption during Morning: 0.42 MW
Average AEP_MW Consumption during Afternoon: 0.51 MW
Average AEP_MW Consumption during Evening: 0.52 MW
```

Solutions:

- **Peak Demand Management**: Utilities can implement time-based pricing strategies to encourage users to shift their consumption to off-peak hours, thereby managing peak energy demand effectively.
- **Energy Efficiency Initiatives**: Promoting smart home technologies and enhancing energy efficiency programs can significantly lower overall consumption during peak hours, contributing to more sustainable energy use.