

# **Power Patterns: Harnessing Electricity for Innovation**

Welcome to our submission for the **Power Patterns: Harnessing Electricity for Innovation Challenge**! In this collaborative notebook, we dive into the electrifying world of load profile clustering to reveal hidden rhythms in building energy usage.

Just like fingerprints, every building has its own unique energy signature. By analyzing these consumption patterns, we’ve grouped buildings with similar behaviors—laying the foundation for smarter, more responsive energy systems. Through this process, we unlock powerful insights that can fuel targeted demand-response initiatives and strengthen grid resilience.

Our mission in this challenge is to design a robust clustering model that uncovers meaningful groupings of buildings—each cluster telling a different story about how electricity is consumed. From deep-dive exploratory analysis to crafting thoughtful features, applying unsupervised learning techniques, and interpreting real-world implications—we guide you through each step of our data-driven discovery.

# Notebook Overview

### **🔍 Data Loading**

We kick off our analysis by importing essential libraries and loading the building-level energy consumption dataset. This section sets the stage by organizing and previewing the data, offering an overview of both building metadata and their corresponding electricity usage profiles.

### **🛠️ Feature Engineering & Extraction**

To uncover meaningful patterns, we dive into feature engineering—transforming raw data into insightful attributes that reflect variations in energy consumption. We explore temporal features, consumption summaries, and time-series behaviors to enrich our dataset and prepare it for clustering.

### **📊 Exploratory Data Analysis (EDA)**

Before building models, we get to know the data inside and out. Through visualizations and descriptive statistics, we analyze the shape, seasonality, trends, and anomalies of load profiles. This step is crucial to guide our modeling decisions and understand the story the data wants to tell.

### **🤖 Unsupervised Learning: Clustering**

Here, we apply unsupervised learning techniques to group buildings with similar electricity usage patterns. We test multiple clustering algorithms, tune parameters, and assess cluster quality—ultimately choosing the approach that best captures the diversity of consumption behaviors.

### 🔍 **Data Profiling**

Before diving into analysis, we perform comprehensive **data profiling** to understand the structure and quality of the dataset. This involves examining data types, detecting missing values, identifying outliers, and analyzing distributions. Profiling provides a foundational understanding of the dataset and helps uncover hidden issues, ensuring cleaner and more reliable feature engineering and modeling downstream.

### **💡 Interpretation & Demand Response (DR) Recommendations**

With our clusters defined, we interpret their unique characteristics and explore how they can support targeted demand-response strategies. This section translates data science into actionable insights—offering cluster-specific recommendations for smarter energy management.

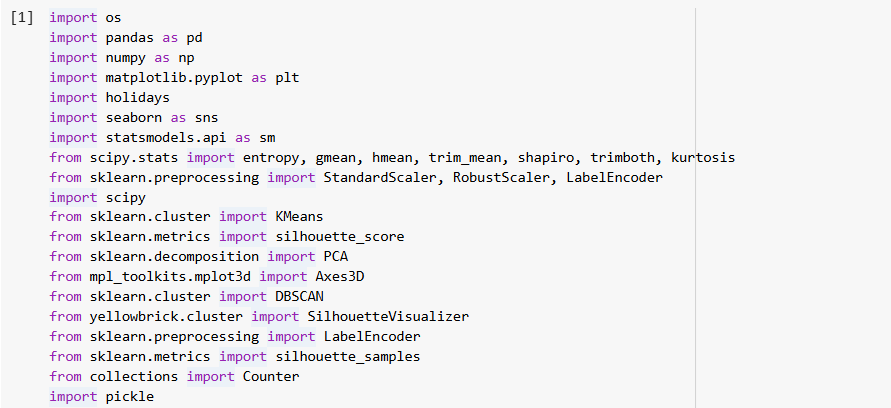
### **🧾 Conclusion**

### We wrap up our journey by summarizing the key discoveries and evaluating the effectiveness of our clustering approach. Finally, we outline future directions, including ways to refine the model, scale deployment, and further align with real-world energy efficiency goals.

# Data Loading

**Library Imports**

We begin our analysis by importing a comprehensive suite of libraries that support various stages of the workflow. For data manipulation and numerical computations, we utilize **Pandas** and **NumPy**, which provide powerful tools for handling structured data and performing array-based operations efficiently. To visualize energy consumption patterns and exploratory insights, we employ **Matplotlib** and **Seaborn**, which offer flexible plotting capabilities and aesthetically pleasing graphics for better data interpretation. For statistical analysis and model diagnostics, we incorporate **Statsmodels** and **SciPy**, which enable robust statistical testing and modeling. The **Holidays** library is used to account for the influence of public holidays on energy consumption patterns, allowing us to incorporate calendar-based features into our analysis. From **Scikit-learn**, we leverage tools for preprocessing and scaling data, implementing clustering algorithms such as KMeans, performing dimensionality reduction through Principal Component Analysis (PCA), and evaluating clustering quality using metrics like the silhouette score. For advanced 3D visualizations, we utilize **mpl\_toolkits**, which allow us to represent multidimensional clustering results in an intuitive format. Finally, the **OS** module is used to interface with the operating system, enabling seamless navigation of directories and management of file paths throughout the project.



**Mount Google Drive**



### **Overview of Building Data and Energy Consumption**

**Building Data:**

The core dataset consists of timestamped electricity load profiles for **1,277 buildings**. Each building's data is stored in a separate Parquet file, capturing its electricity consumption over the course of a year—from **January 1, 2018, to December 31, 2018**. The data is recorded at **15-minute intervals**, providing high-resolution insights into daily and seasonal usage patterns.

**Data Format:**

All files are stored in **Parquet format**, a columnar storage file type optimized for efficient data retrieval. Each file contains a time series of energy consumption values for one building, with consistent timestamp intervals of 15 minutes.

**Energy Consumption:**

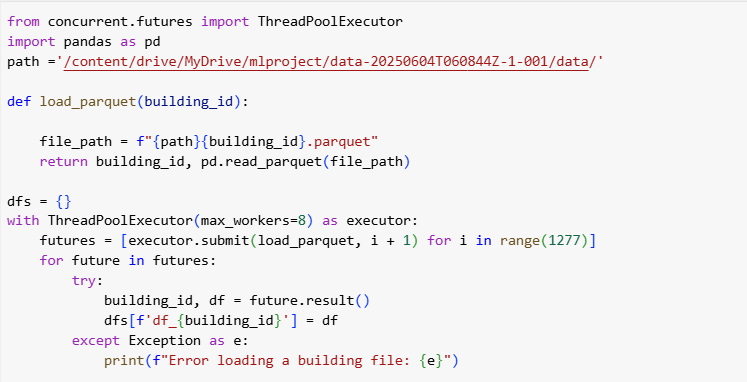
The energy usage values are recorded in **kilowatt-hours (kWh)**. Each value reflects the total energy consumed during the **15-minute interval ending at the corresponding timestamp**, enabling precise tracking of consumption patterns throughout the year.

**Building Data Organization:**

To streamline access and manipulation, the data is loaded into a **Python dictionary** where:

* **Keys** range from 'df\_1' to 'df\_1277', serving as unique identifiers for each building.
* **Values** are **Pandas DataFrames**, each containing the complete time-series energy data for one building.

This structure allows for efficient iteration, analysis, and feature extraction across the full set of buildings.



## Feature Extraction

We are implementing a set of functions to enrich our dataset with additional features derived from the timestamped electricity load profiles and electricity consumption of 1277 buildings.

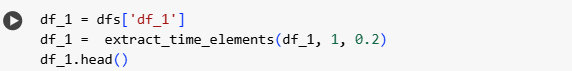
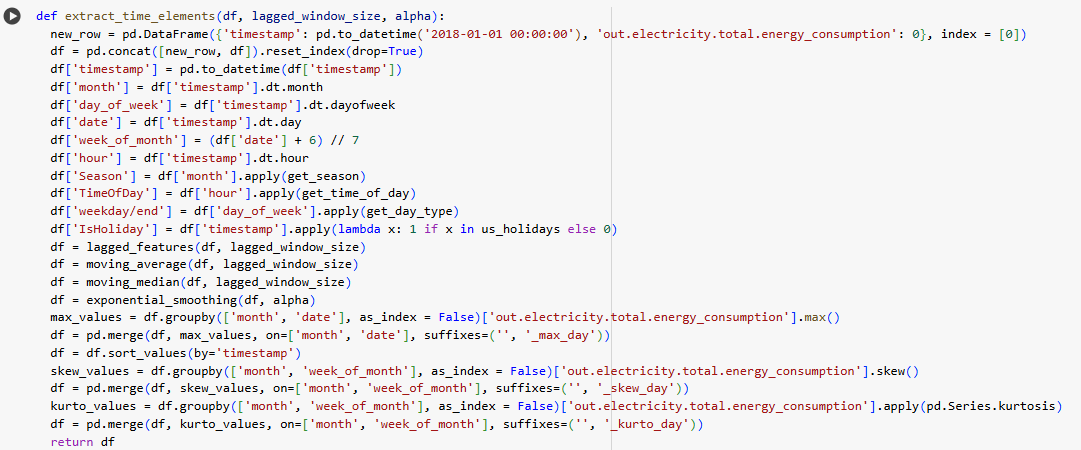
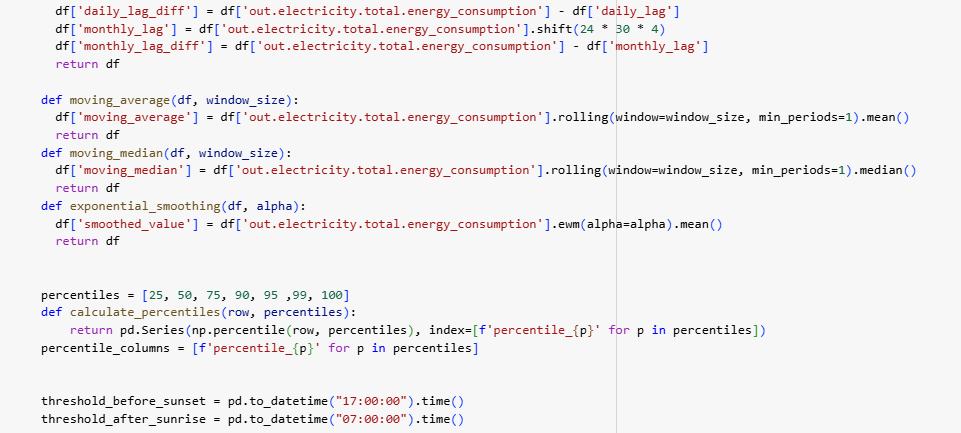
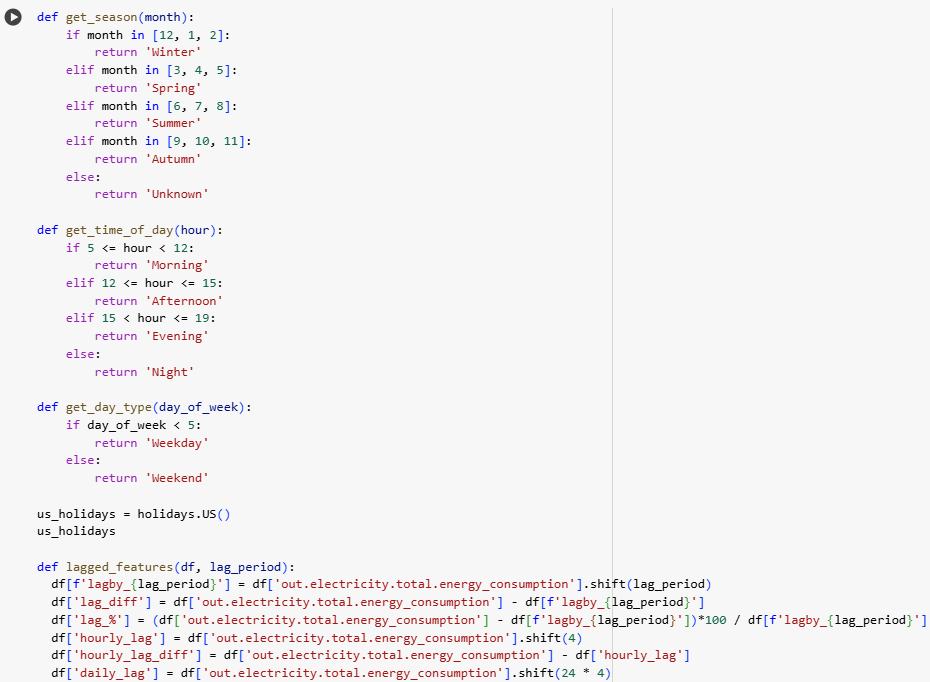
We created functions based on the following data considerations:

Seasons:

|  |  |
| --- | --- |
| Season | Months |
| Winter | December, January, February |
| Spring | March, April, May |
| Summer | June, July, August |
| Autumn | September, October, November |

Time of the Day:

|  |  |
| --- | --- |
| Time of Day | Time Range |
| Morning | 5am-11am |
| Afternoon | 12pm-3pm |
| Evening | 4pm-7pm |
| Night | 8pm-4am |

**#output**To enable clustering, we first need to consolidate the data into a unified format. We apply a characteristic-based clustering approach, which involves extracting a comprehensive set of features that summarize each building’s time series energy data. For every building, we generate 191 features (including the building ID), resulting in a Data Frame of shape (1277, 191). These features fall into two categories:

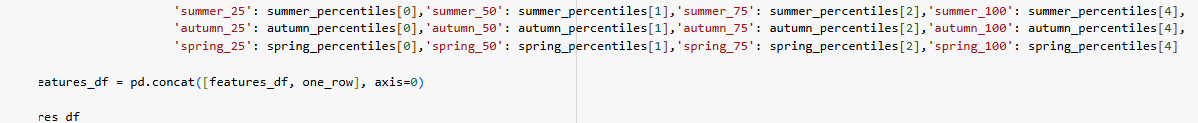
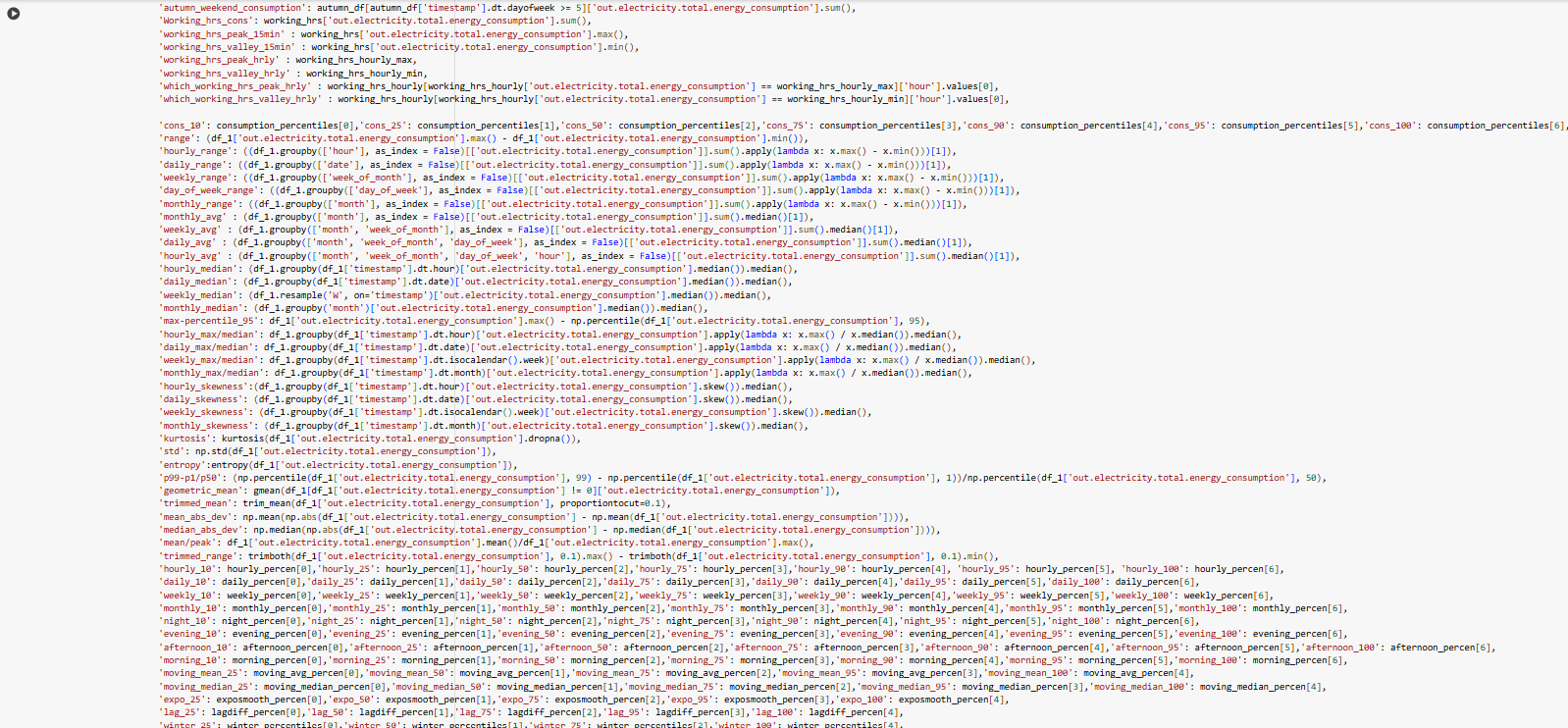
**distribution features**, used for clustering, and **descriptive features**, which support post-clustering interpretation and profiling of building energy behavior.

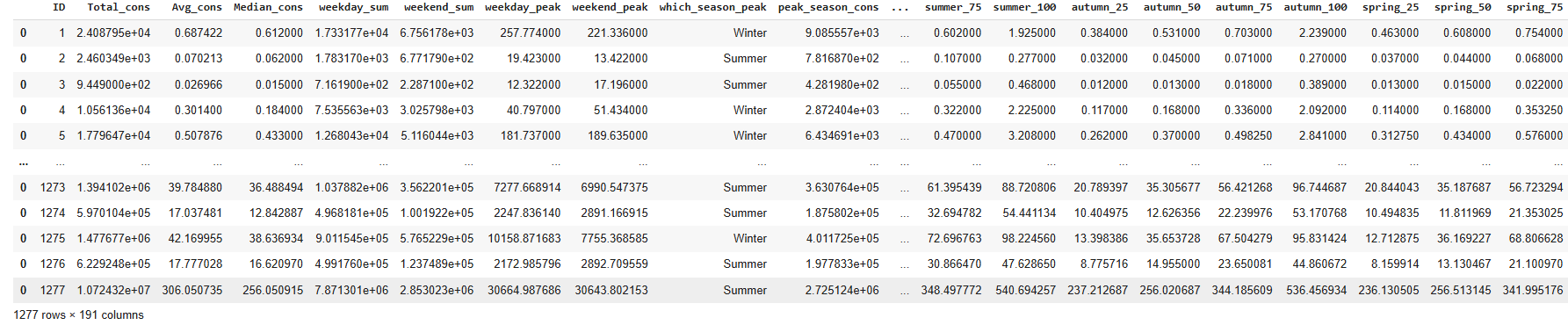
Distribution features:

|  |  |
| --- | --- |
| Feature name | Description of Feature |
| Percentiles (10,25,50,75,90,100) of the following:  Consumption hourly, daily, weekly, monthly night, evening, afternoon, morning, moving\_median, moving\_mean, exponential smoothing, winter, summer, autumn, spring, lag | For each of the variable, we calculated 6 percentiles to get a grasp of consumption distribution. |
| Hourly/daily/weekly/day\_of\_week/monthly\_range  Hourly/daily/weekly/day\_of\_week/monthly\_avg  Hourly/daily/weekly/day\_of\_week/monthly\_median  Hourly/daily/weekly/day\_of\_week/monthly\_max/median  Hourly/daily/weekly/day\_of\_week/monthly\_skewness | Range of the consumption calculated hourly,daily,weekly,day of week and monthly  average of the consumption calculated hourly,daily,weekly,day of week and monthly  median of the consumption calculated hourly,daily,weekly,day of week and monthly  max/median of the consumption calculated hourly,daily,weekly,day of week and monthly  skewness of the consumption calculated hourly,daily,weekly,day of week and monthly |
| Kurtosis,std,entropy, p99-p1/p50  , geometric mean, trimmed mean, trimmed range, mean/median\_abs\_dev, mean/peak | These are some more statistical features that we will create to extract time series data. |

Descriptive features:

|  |  |
| --- | --- |
| Feature name | Description of features |
| Total\_cons,avg\_cons,median\_cons | Total, average,median consumption of a building |
| Weekdays/weekend\_sum | Sum of the consumption for all weekends and weekdays |
| Weekdays/weekend\_peak | Peak consumption on the day with the highest consumption level |
| Which\_season/week/day\_of\_week/hour\_peak | In which season/week/day\_off\_week/hour\_peak,peak consumption for that building happened |
| Peak\_season/week/day\_of\_week/hour\_cons | Peak consumption for the above feature timestamps |
| Weekday/weekend\_timeofday\_max | In weekends/weekdays how much consumption at any time of the day when it is highest |
| Weekday/weekend\_timeofday | In weekends/weekdays at which time of the day, consumption is highest |
| Weekday/weekend\_hour\_max | In weekends/weekdays how much consumption at any hour when it is highest |
| Weekday/weekend\_hour | In weekends/weekdays ,at which time of the day, consumption is highest |
| Night/morning/evening/afternoon\_cons | Total night,morning,autumn and spring consumption |
| Peak\_hour/TimeOfday/day/month | Maximum consumption taken after grouping by hour/timeofday/day/month |
| Summer/winter/spring/autumn\_peak\_hour | At which hour in different season does peak happened |
| Summer/winter/spring/autumn\_max\_consumption\_TimeOfday | At which time of the day in different seasons does peak happened |
| Summer/winter/spring/autumn\_max\_consumption\_value | Peak consumption values for the above feature timestamp |
| Summer/winter/spring/autumn\_weekday\_consumption | Peak consumption values when it is highest at weekdays in different seasons |
| Summer/winter/spring/autumn \_weekend\_consumption | Peak consumption values when it is highest at weekends in different seasons |
| Working\_hrs\_cons | Total working hours consumption where working hours are taken as 9 am - 5pm |
| Working\_hrs\_peak/valley\_15min | Working hours peak/valley taken for 15 minutes intervals |
| Working\_hrs\_peak /valley\_hrly | Working hours peak/valley taken by aggregating consumption hourly |
| Which\_Working\_hrs\_peak /valley\_hrly | At which hours peak/valley is taken above |

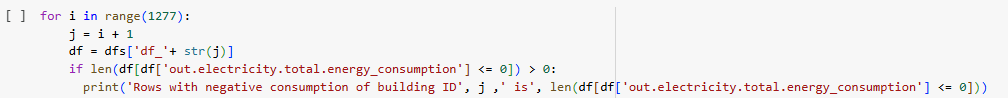


#output

# Exploratory Data Analysis

**Building 963 with Negative Consumption of Electricity**

In the dataset, we observed that a single building with Building ID 963 has exhibited non-positive electricity consumption.



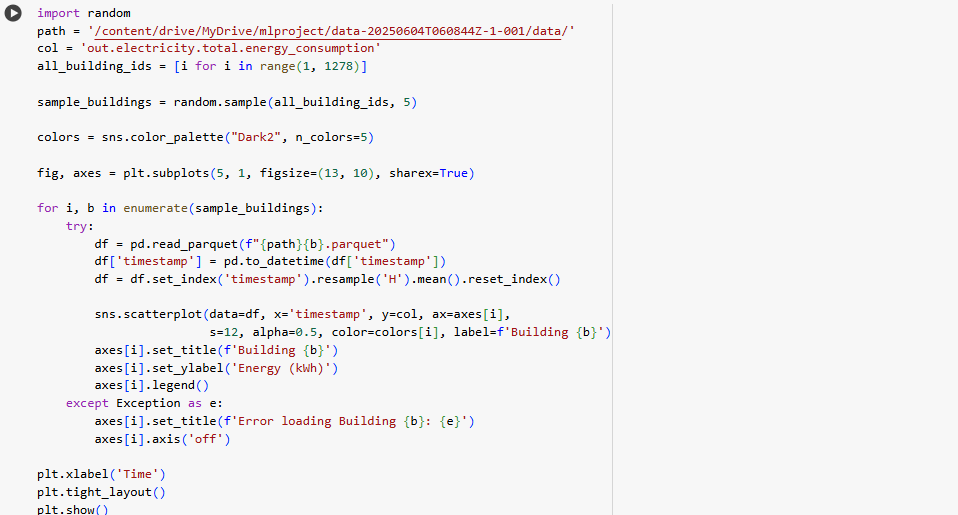
**Preview of Building 963 with non-positive Electricity Consumption**

Even with the presence of non-positive electricity consumption, we are not dropping the building 963 from our dataset because the number of consumption rows associated with it is significantly low. Therefore, removing this building ID would have minimal impact on our analysis as it represents a small portion of the dataset and does not significantly affect the overall trends or patterns observed.



#total yearly consumption graph

The dataset reveals a distinct difference in **total yearly energy consumption** between buildings—those in the first half of the dataset consumes significantly less energy than those in the second half. This separation likely stems from **data leakage** caused by not shuffling the building IDs. However, a more reliable distinction can be made using a **cutoff point of around 35,000 kWh**, based on the consumption distribution. Buildings below this threshold are likely **residential**, while those above are assumed to be **commercial**. To enable more meaningful analysis, the data was **standardized** to minimize the effect of building size and focus on consumption patterns.

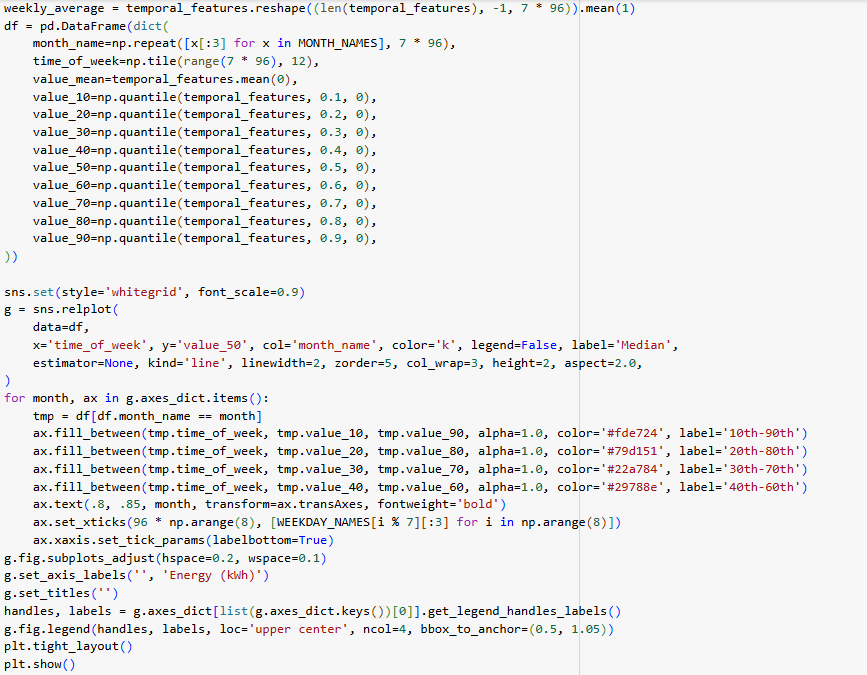
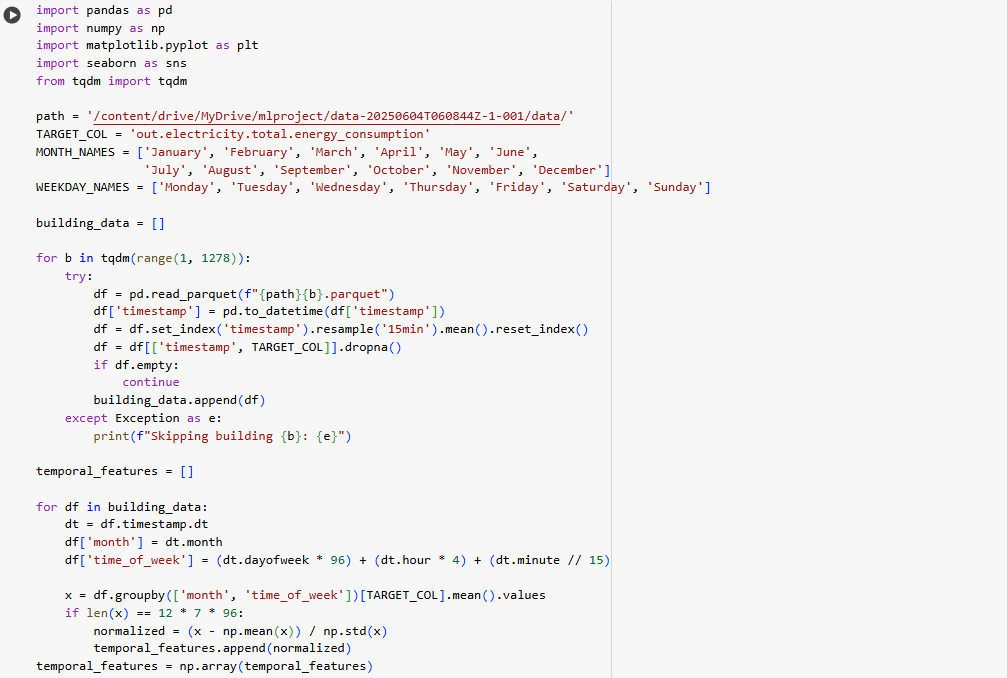


#seasonal consumption graph

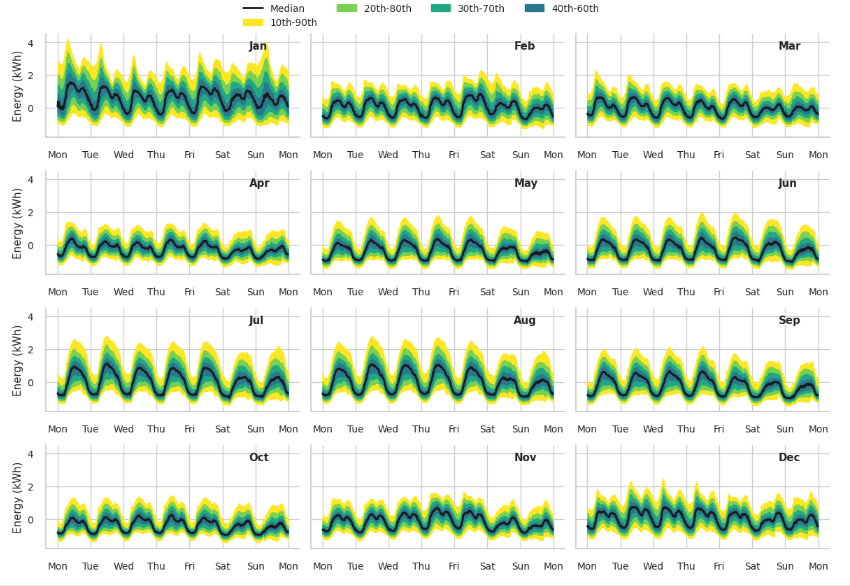
The figure displays the empirical distribution of seasonality vectors across all buildings. The black line indicates the average seasonality pattern, while the shaded areas represent quantiles, offering insights into the variability in energy usage throughout the dataset.

From the figure, we observe several key trends:

* **Winter months** show higher overall energy consumption and greater variability.
* A **dual-peak pattern** appears during winter—one in the morning and one in the evening.
* In contrast, **summer months** typically exhibit a **single daily peak**.



**#OUTPUT**

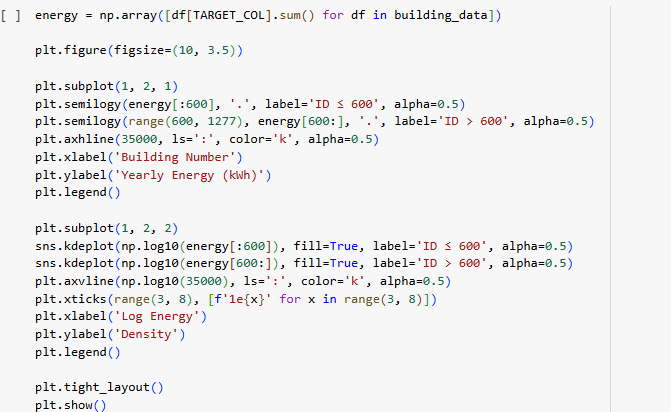


This visualization provides a broad overview of seasonal consumption trends. However, individual buildings may show significant deviations from these patterns. Through clustering, we aim to uncover more distinct and nuanced patterns that may not be evident from the overall distribution alone.

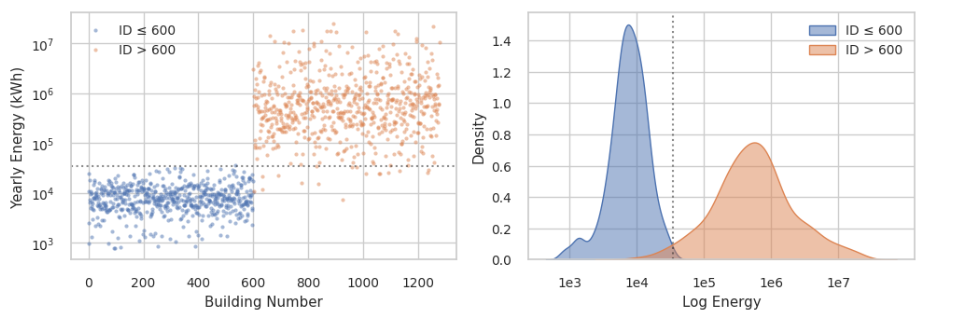
#weekday and weekend consumption graph

To distinguish buildings based on their weekday and weekend energy usage, we compute the **weekday\_weekend\_consumption\_ratio**—the average weekday consumption divided by the average weekend consumption. A ratio above 1 signifies greater consumption on weekdays; below 1 indicates higher weekend usage. For instance, a ratio of 3 implies that weekday consumption is three times that of the weekend.

These insights are valuable for **demand response programs**. Buildings with higher weekday usage could benefit from **time-of-use pricing**, encouraging load shifting to weekends. Conversely, those with greater weekend demand may be ideal candidates for **load-shedding initiatives**, helping reduce stress on the grid during peak times.



**#OUTPUT**



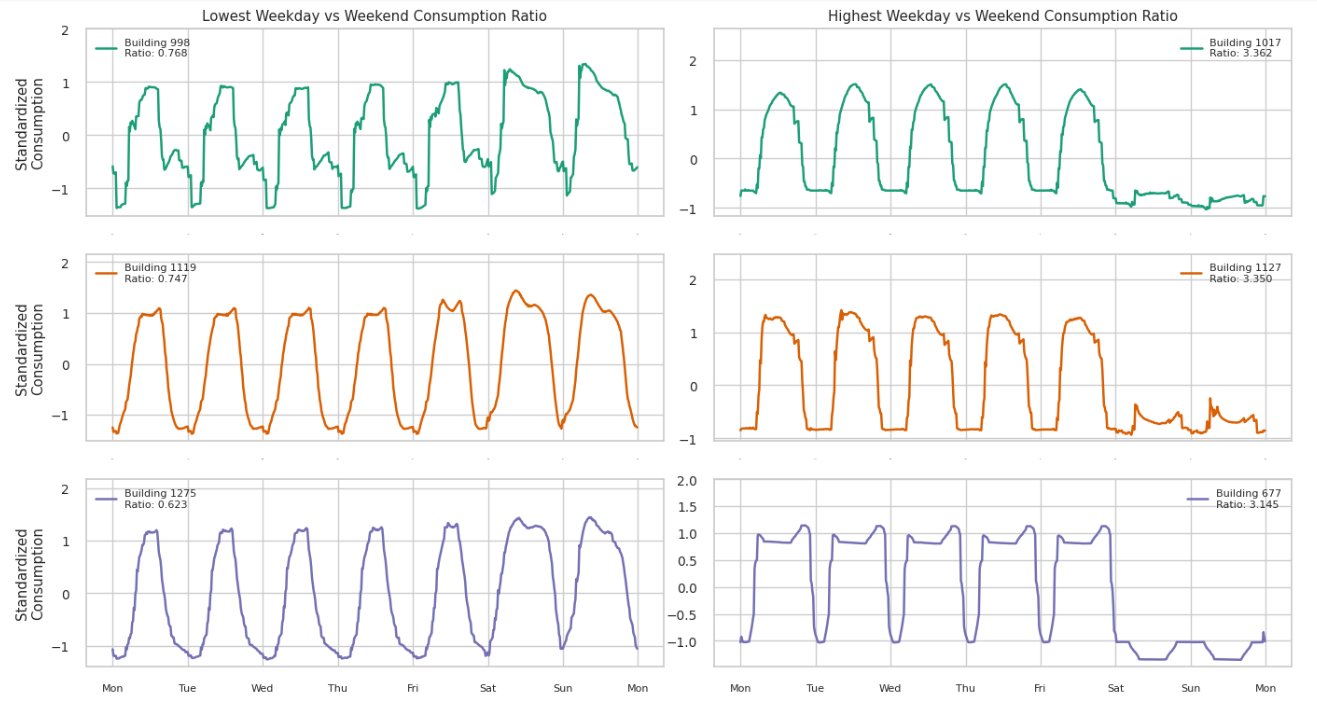
The figure illustrates the **weekly average energy consumption** for the top three buildings with the **lowest** (left) and **highest** (right) values of the **weekday\_weekend\_consumption\_ratio**. Each subplot displays the building ID and its corresponding ratio. Buildings on the left consume more energy on **weekdays**, whereas those on the right consume more on **weekends**. Interestingly, all these buildings are **commercial** (with bldg\_id > 600). This highlights the importance of analyzing weekday vs. weekend consumption separately for residential and commercial buildings.

#predictiability

The figure presents the actual energy consumption patterns for the top three buildings with the **lowest** (left) and **highest** (right) levels of **predictability** during the month of March. Each subplot includes the building ID and its corresponding predictability score. Buildings on the left display **irregular and sporadic spikes** in energy usage, indicating low predictability, whereas buildings on the right exhibit **consistent and stable** consumption, reflecting high predictability. The next step is to compare these patterns between **residential and commercial buildings**.



**#OUTPUT**



### **Unsupervised Learning**

**Unsupervised learning** is a type of machine learning where the algorithm learns patterns and structures from **unlabeled data**—meaning there are **no predefined categories or outcomes**. Instead of predicting a specific result, unsupervised learning focuses on **discovering hidden structures**, such as clusters, associations, or anomalies, within the data.

The two most common types of unsupervised learning are:

* **Clustering** (e.g., K-Means, DBSCAN): Grouping similar data points together.
* **Dimensionality Reduction** (e.g., PCA, t-SNE): Simplifying data while retaining key patterns.

### **What Insights Can We Gain from Building Energy Data Using Unsupervised Learning?**

By applying unsupervised learning (especially clustering) to building energy data, we can uncover **meaningful patterns in electricity consumption** without prior labeling. Some valuable insights include:

#### **1. Building Typology Identification**

* Grouping buildings with similar usage patterns can help distinguish between **residential**, **commercial**, **institutional**, etc., even if that metadata isn't explicitly provided.

#### **2. Load Profile Clustering**

* Buildings can be grouped based on their **daily or seasonal consumption behaviors**, such as:
  + Morning vs. evening peak usage
  + Weekday vs. weekend consumption differences
  + High vs. low variability across seasons

#### **3. Demand Response Targeting**

* Clusters with high weekday usage might be suitable for **time-of-use pricing**, while those with high weekend consumption could benefit from **load-shifting incentives**.
* Highly predictable buildings are good candidates for **automated energy management systems**.

#### **4. Anomaly Detection**

* Outliers or buildings with unusual consumption patterns (e.g., excessive spikes, very low usage) can be identified for further investigation—indicating **inefficiencies**, **malfunctions**, or **waste**.

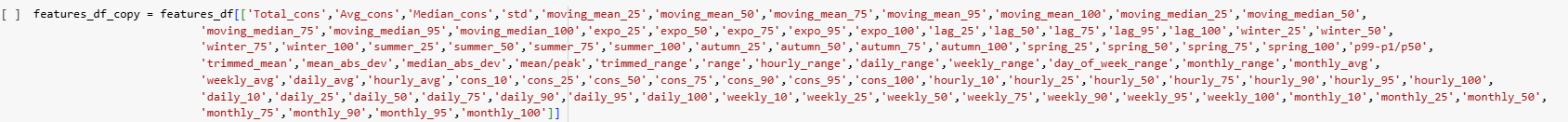
#### **5. Infrastructure Planning & Optimization**

* Utilities can design **localized energy strategies**, optimize **grid resources**, or prioritize infrastructure upgrades based on cluster behavior.

**Clustering**

# **KMeans**

We will implement the KMeans algorithm with a range of cluster numbers (2,14) to identify the optimal number of clusters that best represent the underlying structure of the data. For each value of cluster within the range, we calculate the inertia and silhouette scores, which serve as metrics for evaluating the clustering performance.

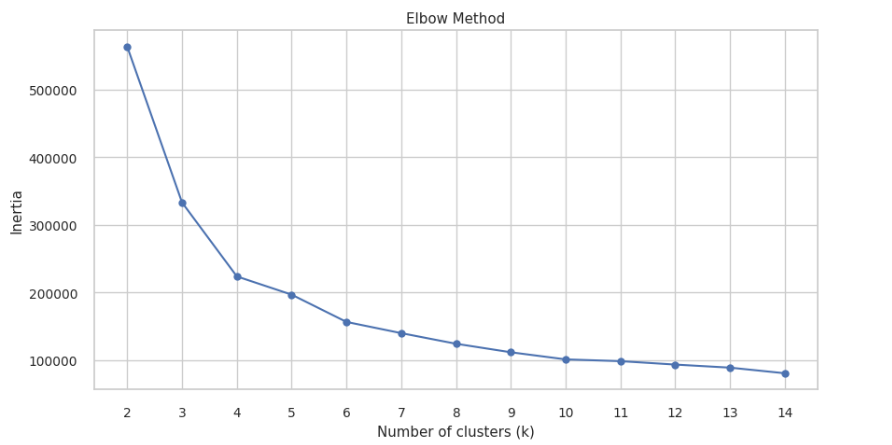


**Elbow Method**:

To determine the optimal number of clusters, we utilize the elbow method by plotting the inertia values against the number of clusters. The following code scales a dataset using RobustScaler, then iterates through cluster numbers (2-14) to perform KMeans clustering, storing inertia and silhouette scores, and plots the elbow curve.



**#OUTPUT**



**Silhouette Visualizer**

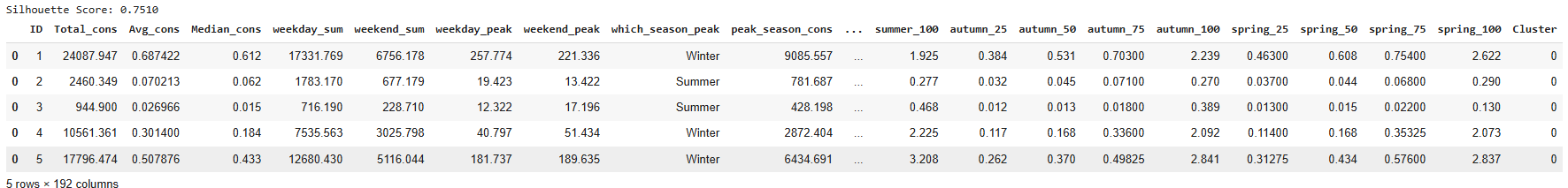
The Silhouette Visualizer displays the silhouette coefficient for each sample on a per-cluster basis, visualizing which clusters are dense and which are not. It also displays average silhouette score.



**#OUTPUT**



**#OUTPUT**



Upon analysis, the elbow method suggests that the optimal number of clusters for our dataset is 4. From Silhouette Visualizer we can infer that cluster 0 will contain maximum data points while cluster 3 will contain least. This clustering can make sense if we have maximum buildings sharing same energy consumption pattern. The Silhouette score obtained 0.75 at 4 clusters indicates better-defined clusters, well-separated clusters. Also, 4 clusters are better in profiling point of view too.

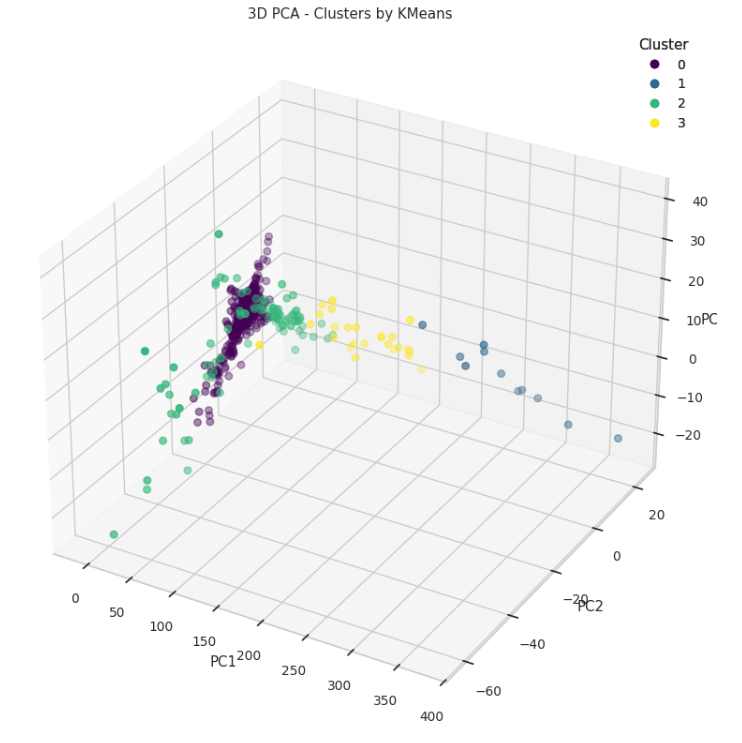
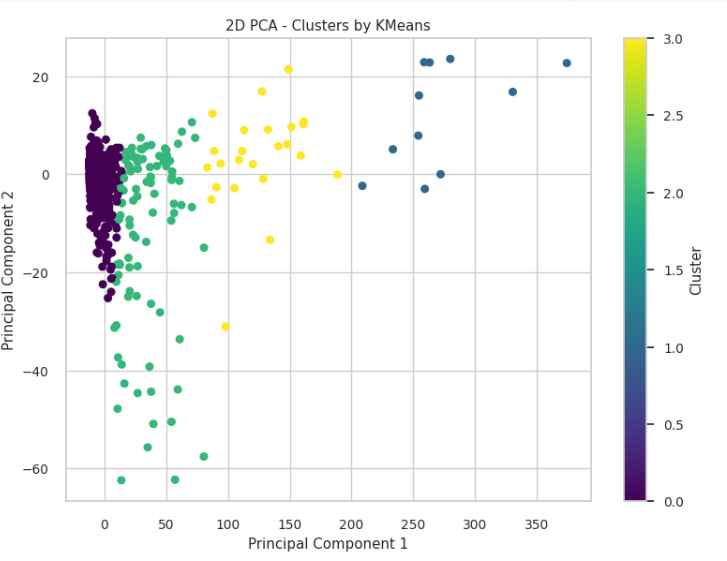
Since, we have obtained optimal cluster count as 4, we fit the cluster labels to features\_df as 'Cluster' column.

**PCA(2D)/PCA(3D):**

PCA is applied with both two and three components to visualize and interpret the underlying structure of the data across four clusters, capturing more variance and revealing intricate patterns in a lower-dimensional space.

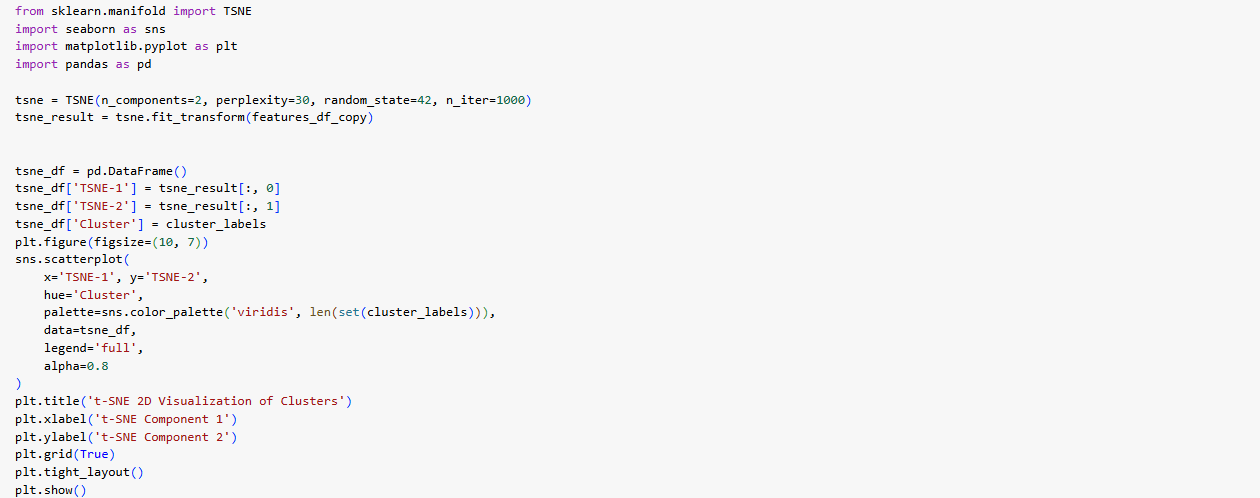


**#OUTPUT**

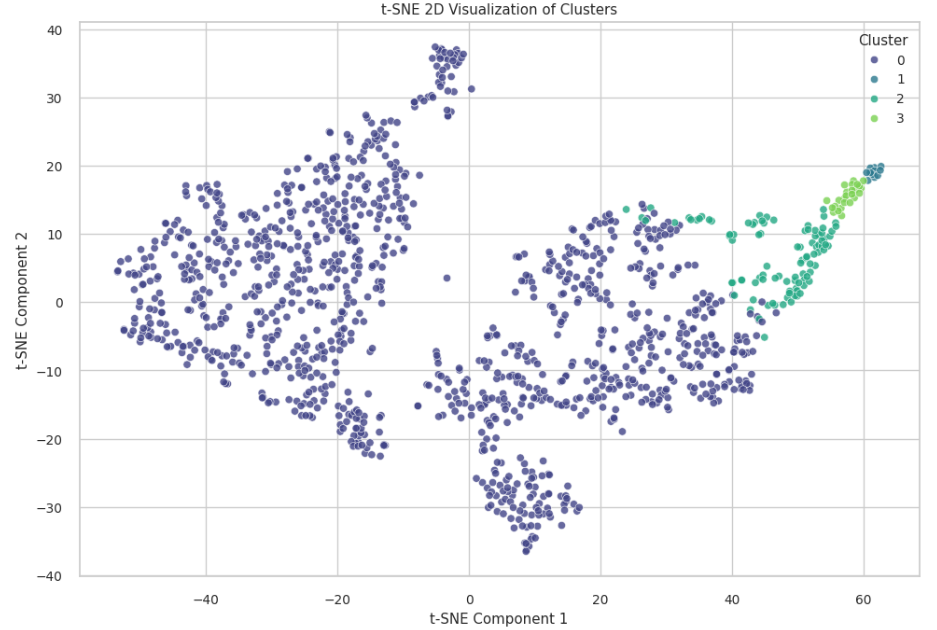


***Given that Cluster 0 comprises 90% of all buildings, we will further divide it into subclusters to investigate whether these subgroups exhibit any distinct patterns or differences in energy consumption behavior.***

**TSNE**



**#output**



***Clustering of cluster 0***

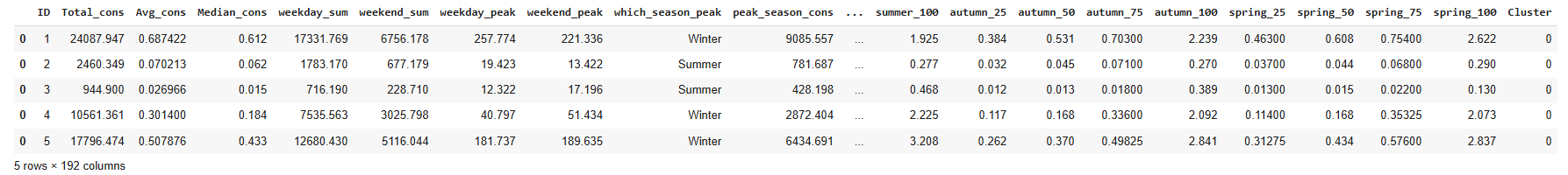
**Defining a dataframe for cluster 0**



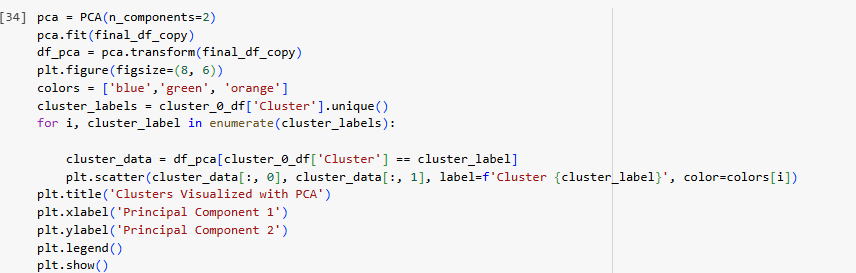
**Elbow Method for Cluster 0****Silhouette Visualizer**

**Performing KMeans for Cluster 0**

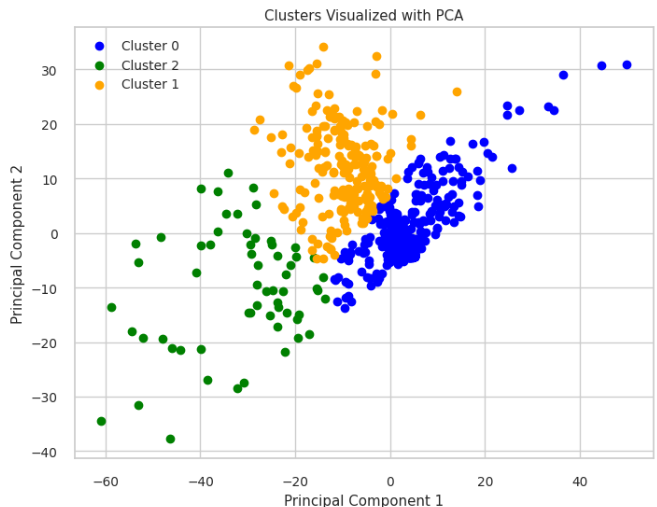
**#output**



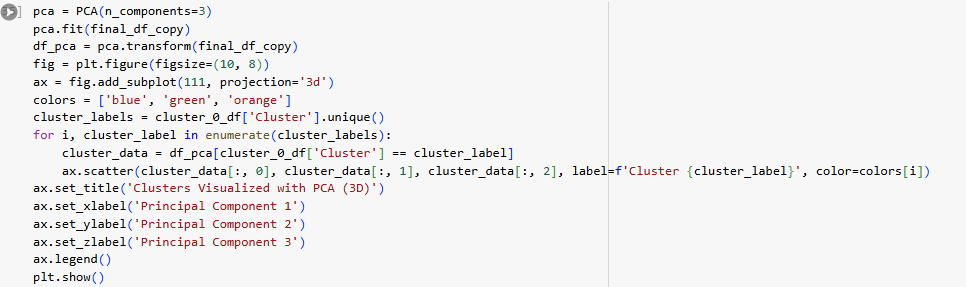
**PCA(2D) of Cluster 0**



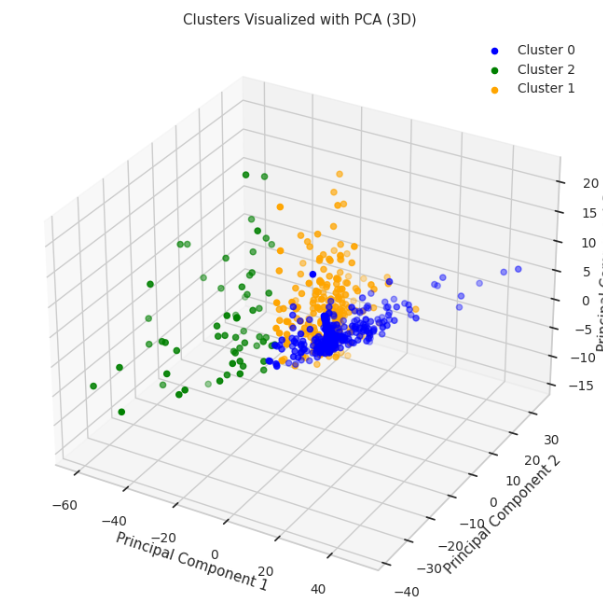
**#output**



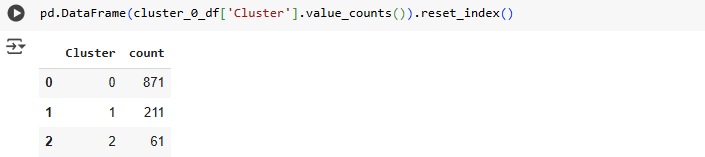
**PCA(3D) of Cluster 0**

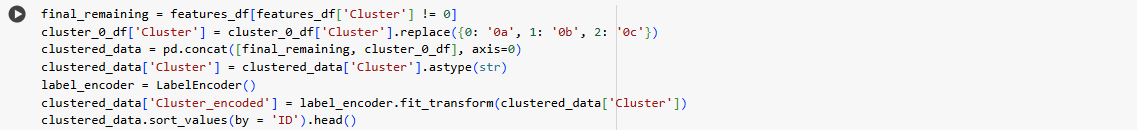


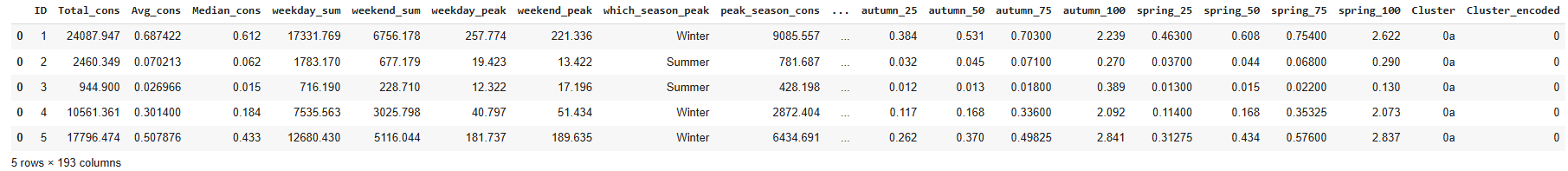
**#OUTPUT**



**Building Count within Sub-Clusters of Cluster 0**

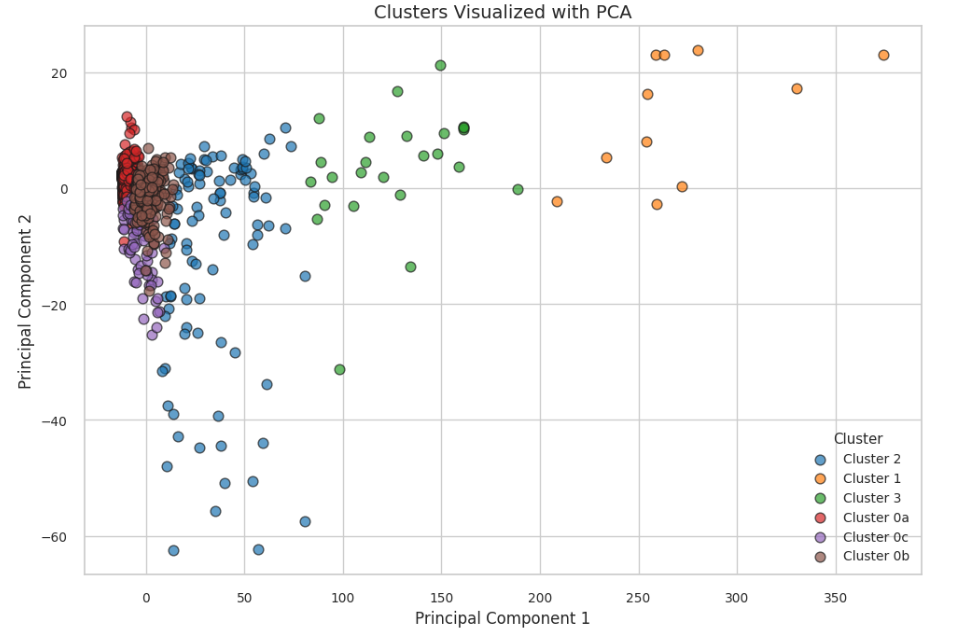


**#output**

2D Visualization of all 6 clusters



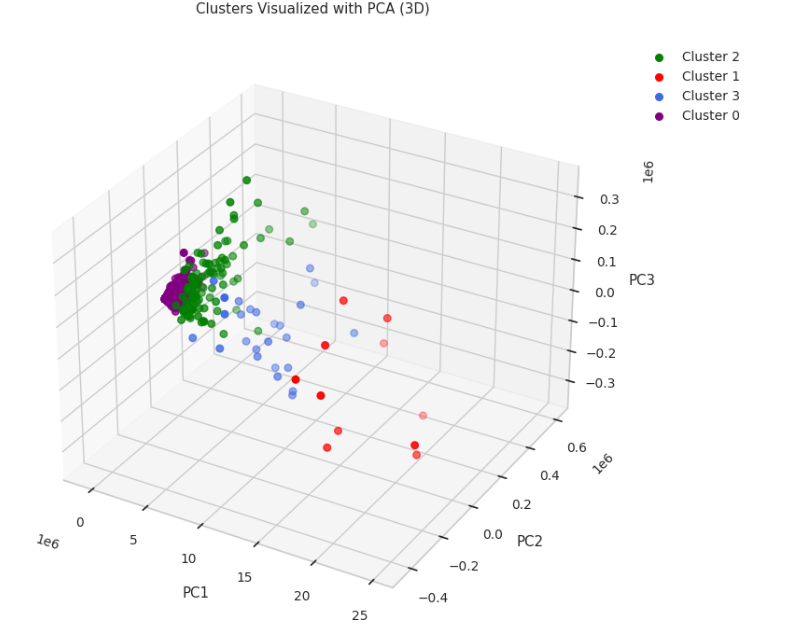
**#output**



3D Visualization of all 6 clusters



**#OUTPUT**



# **DBSCAN**

We utilize DBSCAN to group closely packed points into clusters and identify points in low-density areas as outliers. In DBSCAN, we have two parameters: ε (maximum distance for points to be neighbors) and MinPts (minimum points to form a cluster)

**Knee Plot**

Knee plot visualizes distances to nearest neighbors, helping identify a significant change, known as the "knee" point. This point indicates the optimal parameter value for epsilon in DBSCAN, balancing cluster density and noise.

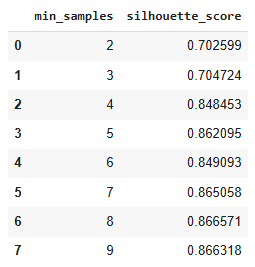
**Installing 'kneed' package**

From the Knee Plot, we can identify the "knee point" which is marked on the distance curve. The Value of knee point is 22.685694902332738. Hence, we infer that the best epsilon value is 22.685694902332738.





**#output**



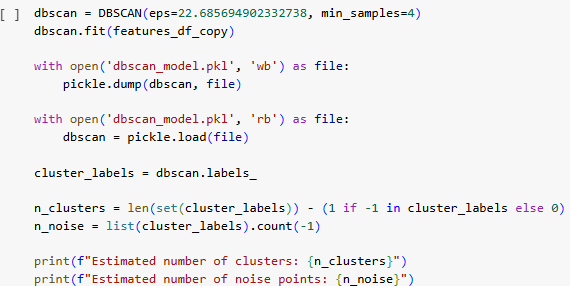
**Utilization of Silhouette Score to determine min\_samples value**

By computing the silhouette scores over a range of values (2, 10), we determine the optimal min\_samples value.

From the silhouette scores, to obtain lesser noise points we choose min\_samples value as 4.

**Performing DBSCAN**

With epsilon value as 22.685694902332738 and min\_samples as 4, we perform DBSCAN to obtain clusters and noise points.



**#output**



**Silhouette Plot**

Silhouette plot evaluates the quality of clusters by measuring how similar points are to their own cluster versus other clusters. It helps determine the appropriateness of clustering parameters.

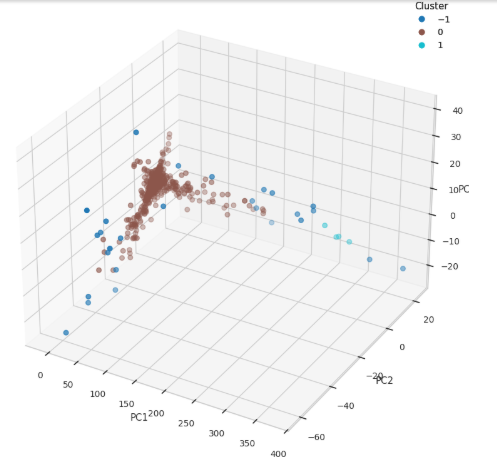
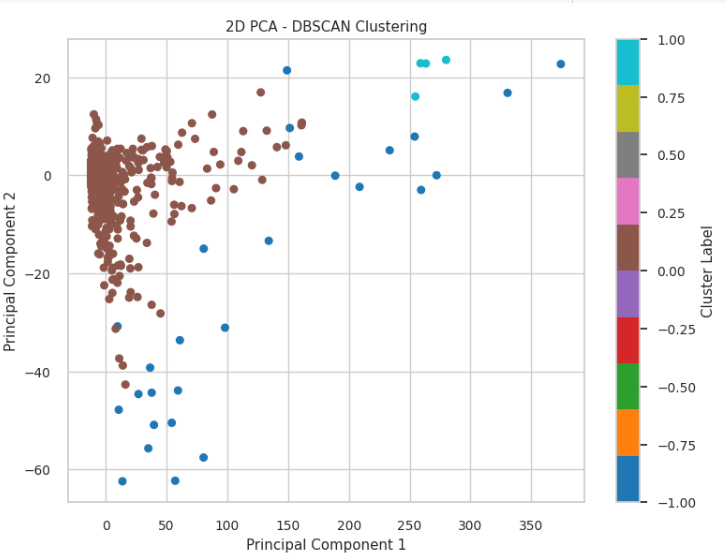
The silhouette plot indicates that clusters are well-defined, but Cluster 0 is disproportionately large, diminishing the overall effectiveness of the clustering.

**PCA (2D)/(3D):**

PCA in 2D reduces high-dimensional data into two principal components, simplifying visualization and effectively revealing the structure of clusters and identifying noise points, which aids in interpreting DBSCAN results. Extending this to a 3D PCA plot offers a more comprehensive view of cluster formations and outliers, further enhancing our understanding of the underlying patterns and the performance of the DBSCAN algorithm.



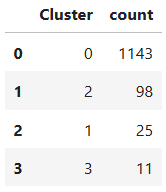
**#output**



**Building Count** **Of Each Cluster**



**#output**



***After evaluating the clustering results, it was clear that DBSCAN was not well-suited for this dataset, as it produced an imbalanced cluster distribution and fewer clusters than expected. Due to these shortcomings and its suboptimal performance, we opted to switch to the KMeans algorithm. K-Means is anticipated to deliver more balanced and meaningful clusters, enhancing the accuracy and effectiveness of the overall clustering process.***

# **Hierarchical Clustering**

We utilize **Hierarchical Clustering** to build a nested grouping of data points based on their similarity, without requiring a pre-defined number of clusters. The algorithm progressively merges (or splits) clusters based on a **distance metric** and a **linkage criterion**.

In hierarchical clustering, two key components are:

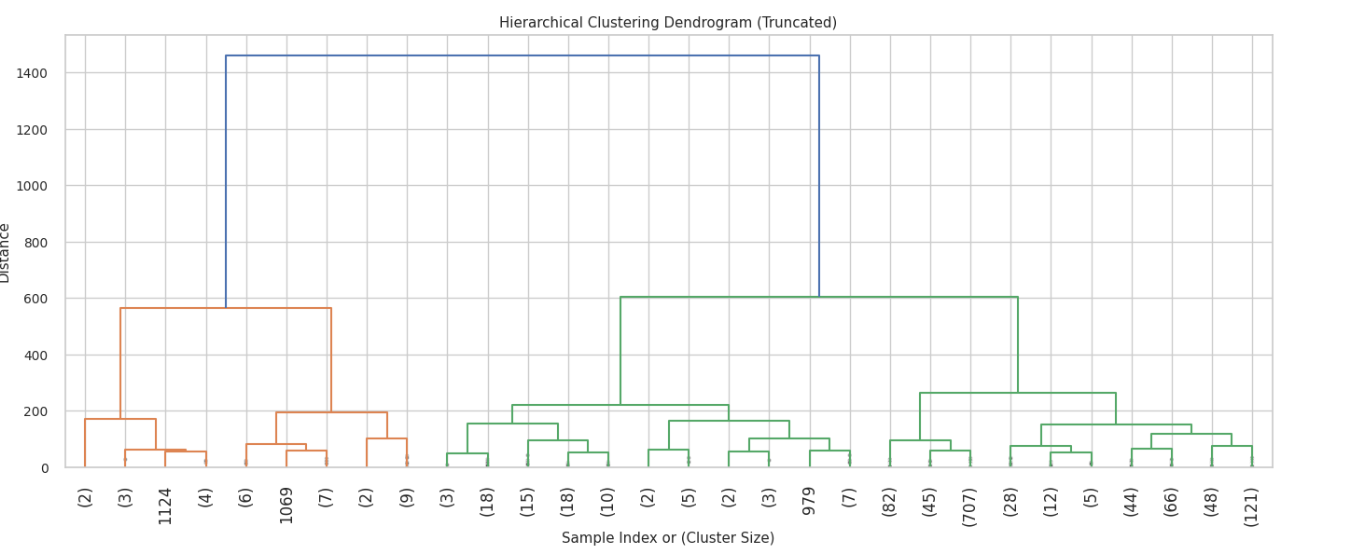
* **Distance Metric (e.g., Euclidean, Manhattan):** Defines how similarity between points (or clusters) is measured.
* **Linkage Criteria:** Determines how distances between clusters are calculated, including:
  + **Single linkage:** Minimum distance between points across clusters.
  + **Complete linkage:** Maximum distance between points across clusters.
  + **Average linkage:** Mean distance between all points across clusters.

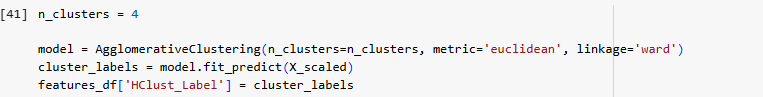
The result is visualized through a **dendrogram**, which shows the hierarchy of merges. By cutting the dendrogram at an appropriate height, we can determine the final number of clusters.

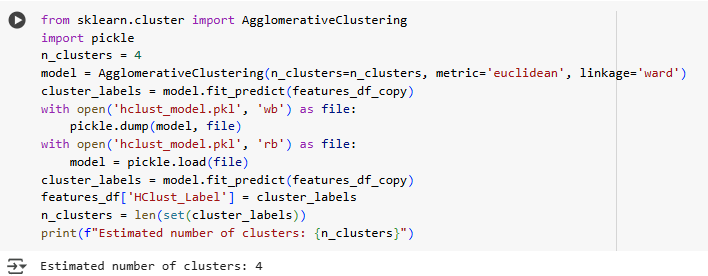
**Dendrogram:**

In a hierarchical clustering dendrogram, the y-axis shows the distance at which clusters merge, indicating their dissimilarity. The x-axis represents data points or clusters in arbitrary order. As height increases, more dissimilar clusters merge. Cutting the dendrogram at a specific y-value determines the final number of clusters.

**#output**



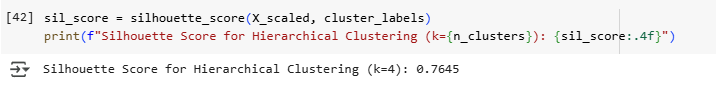




**Silhouette Plot:**

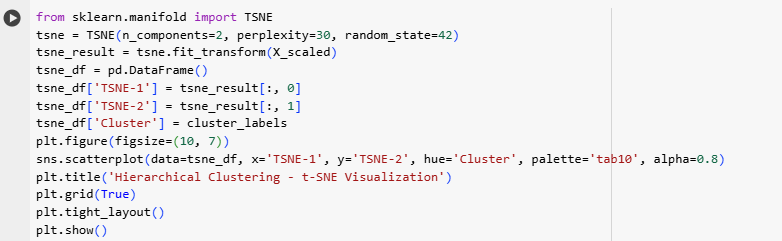
Silhouette plot evaluates the quality of clusters by measuring how similar points are to their own cluster versus other clusters. It helps determine the appropriateness of clustering parameters.

The silhouette plot indicates that clusters are well-defined, but Cluster 0 is disproportionately large, diminishing the overall effectiveness of the clustering.

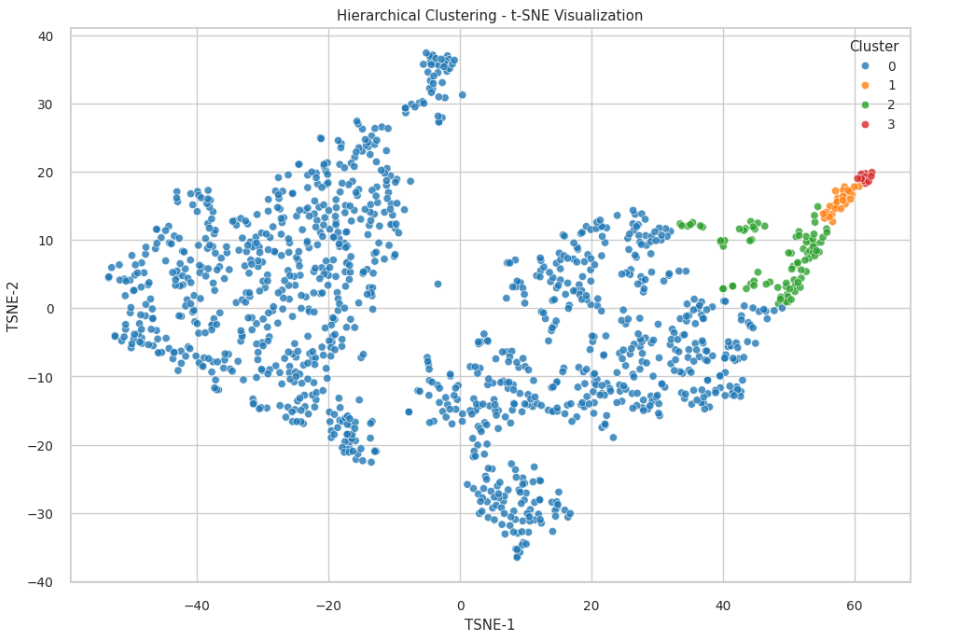


**t-SNE plot:**

t-SNE (t-distributed Stochastic Neighbor Embedding) is a powerful dimensionality reduction technique that maps high-dimensional data into 2D, or 3D space while preserving local structure. When applied before or after **hierarchical clustering**, it enables clearer visualization of how data points group together based on similarity. Unlike PCA, which preserves global variance, t-SNE focuses on local neighborhoods, making it particularly effective at revealing **tightly knit clusters** and **subtle separations**. This can help validate the **merging structure seen in a dendrogram** and provide deeper insights into the cluster hierarchy.

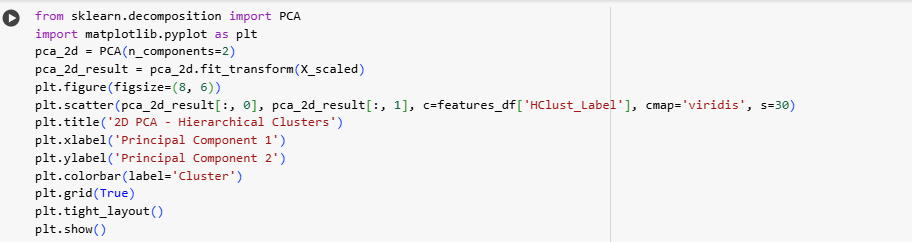


**#output**

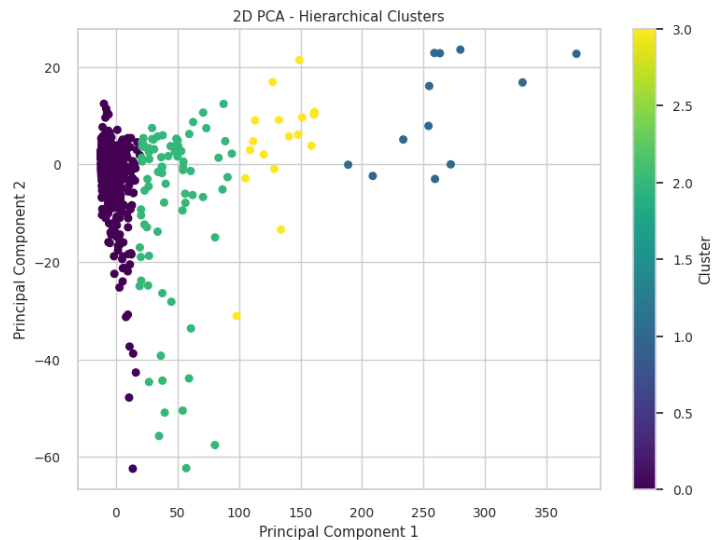


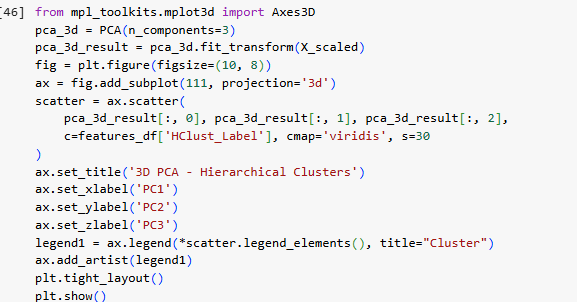
**PCA(2D)/(3D):**

PCA in 2D reduces high-dimensional data into two principal components, making it easier to visualize the cluster structure and interpret the results of hierarchical clustering. It helps reveal how data points group together and at what distances they merge. Expanding to a 3D PCA plot provides a deeper view of the spatial relationships between clusters, highlighting the hierarchy and separation between them. This enhanced visualization supports better understanding of the clustering process and aids in selecting an appropriate level to cut the dendrogram.

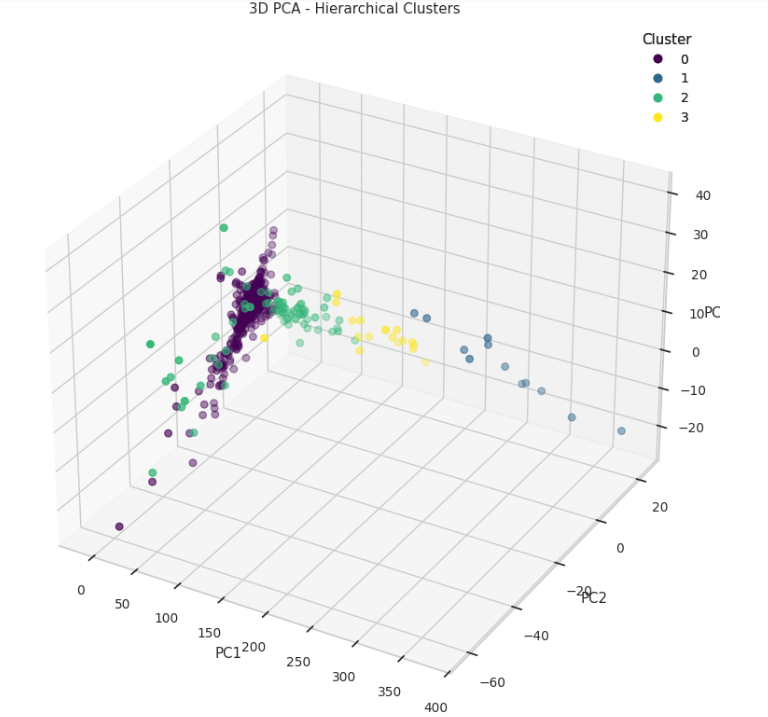


**#output**



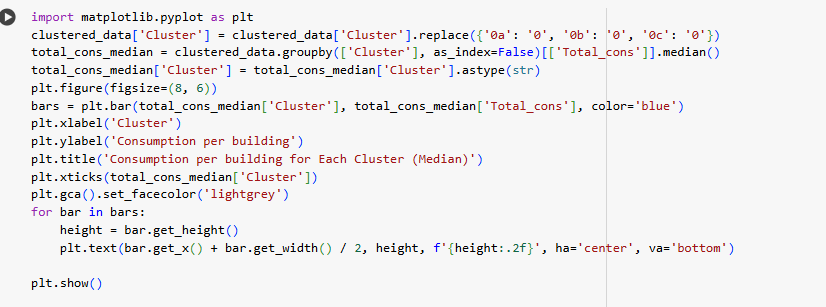


**#output**

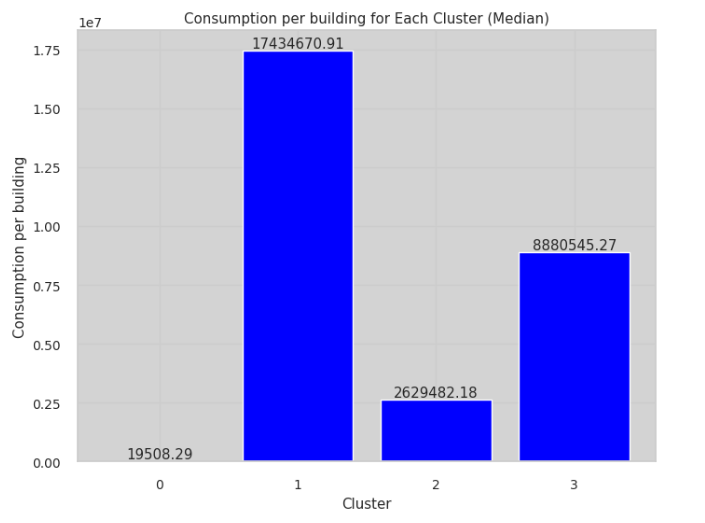


# **Profiling**

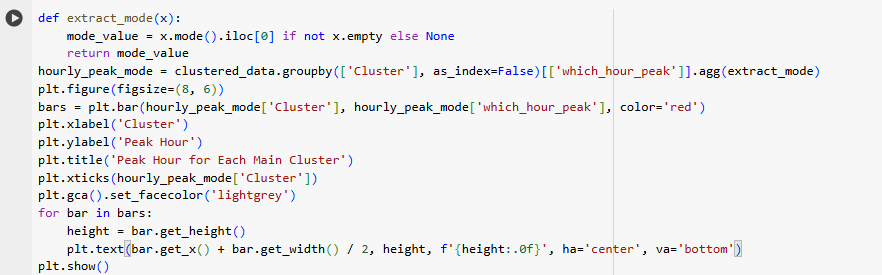
**Consumption per building for Each Cluster (Median)**



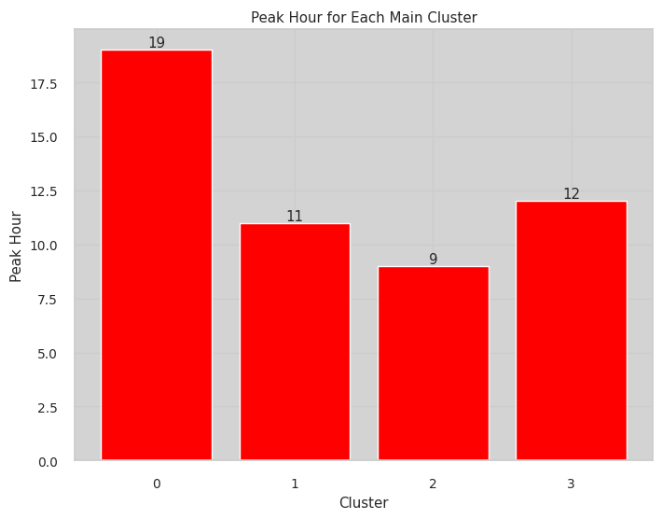
**#output**



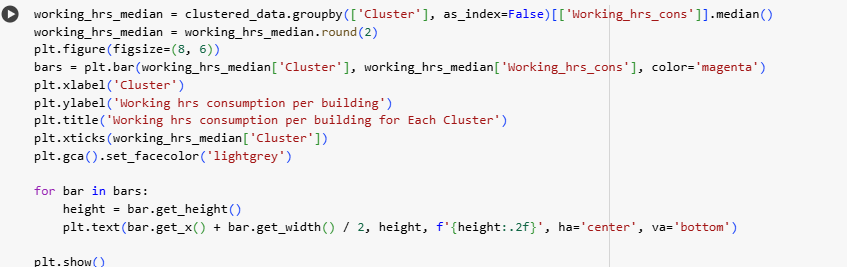
**Peak Hour for Each Cluster**



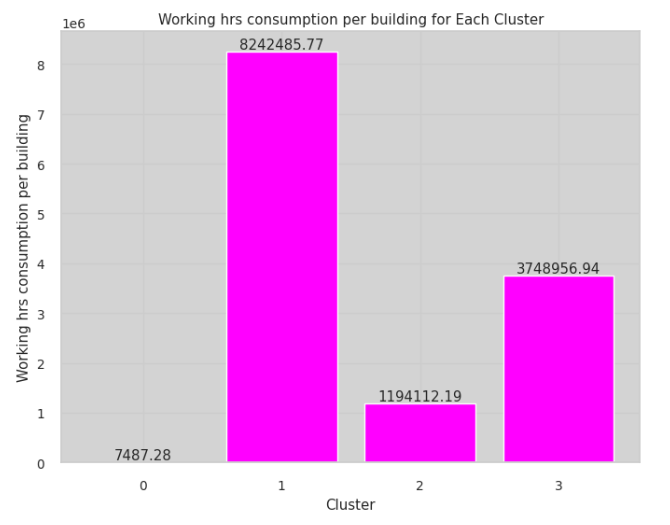
**#output**



**Working hours consumption per building for Each Cluster**



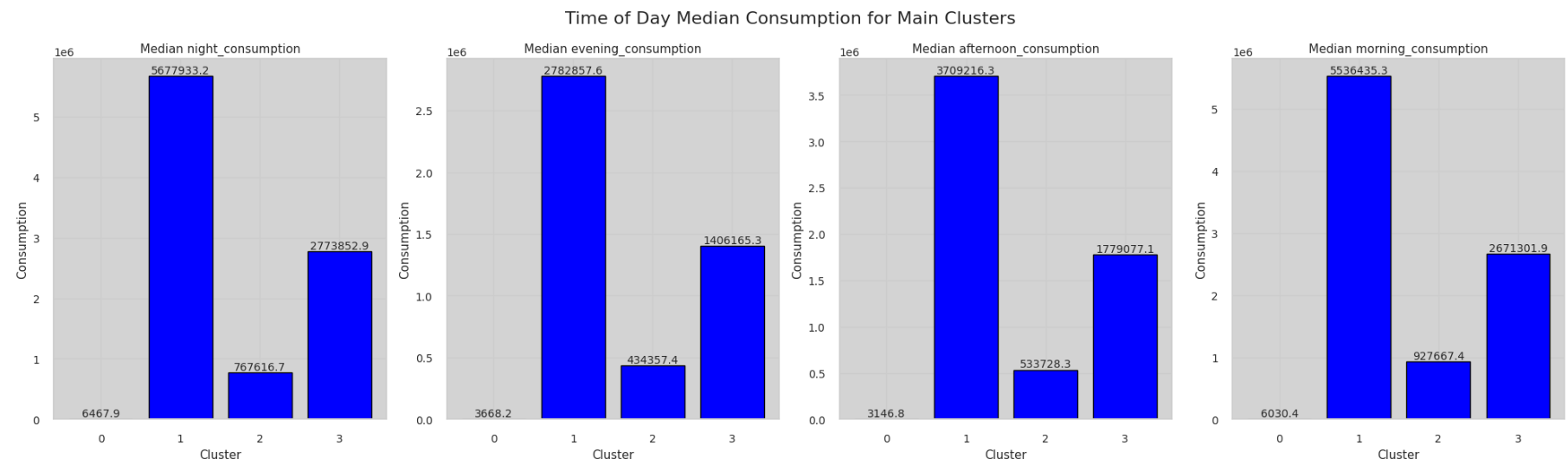
**#output**



**Median Time of Day Consumption for Each Clusters**



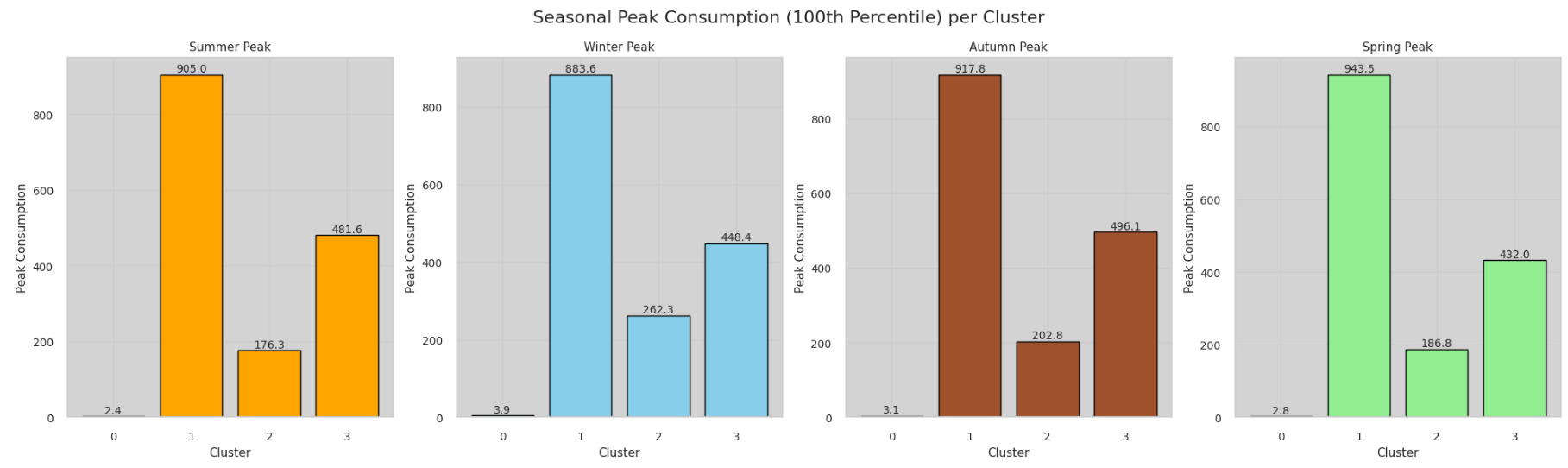
**#output**



**Seasonal Peak Consumption for Each Clusters**



**#output**



Here are the features used for profiling, along with their variations across each cluster.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Building profile | Residential | Restaurant/Hospital/Hotel | Schools | Offices/  Shop |
| Number of buildings | 1143 | 25 | 98 | 11 |
| Cluster | 0 | 1 | 2 | 3 |
| which\_hour\_peak | 19.0 | 11.0 | 9.0 | 12.0 |
| % in total median\_consumption | 3.5% | 30.7% | 9.1% | 60.2% |
| weekday\_sum\_per\_day | ~765950 | 12673968 | 1989343 | 6450162 |
| weekend\_sum\_per\_day | 220000 | 4743701 | 648041 | 229966 |
| %diff | 71.3 | 59.3 | 51.5 | 64.3 |
| Weekday\_timeofday | Evening | Afternoon | Morning | Afternoon |
| working\_hours\_  consumption | 586743 | 8242486 | 1194112 | 3748957 |
| working\_hrs\_peak\_hrly | ~970 | ~2000 | ~885 | ~3850 |
| working\_hrs\_valley\_hrly | ~210 | ~604 | ~127 | ~1180 |
| Seasonal\_peak\_  consumption | 3.9 | 943.5 | 262.3 | 491.6 |
| Season peak | Winter | Spring | Winter | Autumn |

|  |  |  |  |
| --- | --- | --- | --- |
| Building profile | Small residential | Duplex/  Bungalow | Multistory |
| Number of Buildings | 871 | 211 | 61 |
| cluster | 0 | 1 | 2 |
| which\_hour\_peak | 19.0 | 10.0 | 7.0 |
| % in total median\_consumption | ~0.06% | ~4.5% | ~2.6% |
| weekday\_sum\_per\_day | 14228 | 765949 | 441468 |
| weekend\_sum\_per\_day | 5116 | 361521 | 220287 |
| %diff | 64 | 52.8 | 50.1 |
| Weekday\_timeofday | Evening | Morning | Morning |
| working\_hours\_consumption | 4420.02 | 361521.85 | 220287.45 |
| working\_hrs\_peak\_hrly | 13.7 | ~1400 | ~900 |
| working\_hrs\_valley\_hrly | ~0.7 | ~700 | ~300 |
| Seasonal\_peak\_  consumption | 2.4 | 79.7 | 81.3 |
| Season peak | Winter | Winter | Winter |

## ***Cluster Analysis based on Energy Consumption Patterns***

### **Main Cluster 0 (Residential):**

#### Sub Cluster 0a: Small Residential Buildings

* **Number of Buildings:** 871
* **Peak Hour:** 7 PM
* **Median Total Consumption:** 11,036 units
* **Median Total Weekday Consumption:** 14228 units
* **Median Total Weekend Consumption:** 5116 units
* **Percentage Difference between Per Day Consumption between Weekday and Weekend:** 64%
* **Time of Day Peak in Weekdays:** Evening
* **Working Hours Median Consumption:** 4,420 units
* **Hourly Peak during Working Hours:** 7.8 units
* **Hourly Valley during Working Hours:** 0.4 units
* **Seasonal Peak:** Winter, 2.4units

**Interpretation**:

* Evening peak (7 PM) suggests occupancy-driven usage after work hours.
* Lowest total consumption among all clusters, matching small household sizes (2–4 people).
* Drastic drop in usage during weekends implies minimal home activity.
* Extremely low seasonal and weekly peaks further confirm small energy needs.

**Explanation:**

* **Evening peak (7 PM)**: This is consistent with residents returning home after work, which results in increased electricity use from lights, cooking, TV, etc.
* **Lowest total consumption**: These buildings have small family sizes (2–4 people) and limited appliances, explaining the very low energy use.
* **Drastic weekend drop**: Suggests that weekends are quieter, possibly due to residents going out or lesser appliance use.
* **Low seasonal/weekend peaks**: Minimal fluctuation shows there's no significant change in appliance usage across seasons, and occupancy is steady.

#### Sub Cluster 0b: Duplex/Bungalow

* **Number of Buildings:** 211
* **Peak Hour:** 10 am
* **Median Total Consumption:** 7,65,950 units
* **Median Total Weekday Consumption:** ~3,61,522 units
* **Median Total Weekend Consumption:** ~2,78,470 units
* **Percentage Difference between Per Day Consumption between Weekday and Weekend:** 52.8%
* **Time of Day Peak in Weekdays:** Morning
* **Working Hours Median Consumption:** 3,61,522 units
* **Hourly Peak during Working Hours:** ~1400 units
* **Hourly Valley during Working Hours:** ~700 units
* **Seasonal Peak:** Winter, 79.7 units

**Interpretation**:

* Moderate to high consumption shows these homes are larger with 5–10 occupants.
* Activity throughout the day is evident from the strong working-hours usage.
* High winter peak suggests dependence on heating systems or general indoor activity.
* Slightly lower weekend activity but still significant, possibly due to occupancy throughout the week.

**Explanation:**

* **Moderate to high consumption**: Duplexes and bungalows are larger, typically with 5–10 occupants and more appliances (e.g., multiple ACs, washing machines), explaining the overall higher usage.
* **Activity during the day**: High working-hour usage suggests these homes are occupied during the day—maybe by elderly, homemakers, or work-from-home individuals.
* **High winter peak**: Indicates heating appliances (heaters, water heaters) are commonly used in winter.
* **Weekend activity**: While there's a dip, it’s not drastic suggesting that occupancy patterns remain consistent through the week.

#### Sub Cluster 0c: Multistory Buildings

* **Number of Buildings:** 61
* **Peak hour (Maximum for all building):** 7 AM
* **Median Total Consumption:** 4,41,468 units
* **Median Total Weekday Consumption:** 4,41,468 units
* **Median Total Weekend Consumption:** 2,20,287 units
* **Percentage Difference between Per Day Consumption between Weekday and Weekend:** 50.1%
* **Time of Day Peak in Weekdays:** Morning
* **Working Hours Median Consumption:** 2,20,287 units
* **Hourly Peak during Working Hours:** ~900 units
* **Hourly Valley during Working Hours:** ~300 units
* **Seasonal peak:** Winter, 81.3units

**Interpretation**:

* Morning peak implies shared services or early activity in large complexes (>100 occupants).
* Large differences in weekday vs. weekend activity suggest weekday-only services (e.g., maids, maintenance).
* Moderate-high winter and weekly peak usage reflect dense population and appliance use.

**Explanation:**

* **Morning peak (7 AM)**: Residents start their day with heavy appliance use (geysers, cooking, lights), and peak is earlier due to combined load across units.
* **Shared services**: Higher morning consumption could also include elevators, corridor lighting, and water pumps.
* **Weekday/weekend difference**: Suggests weekday-only services like housekeeping, society office operations, etc.
* **Winter/weekly peaks**: Larger population density means more residents using devices simultaneously during cold weather or weekdays.

### **Main Cluster 1 (Hotel/Restaurant/Hospital):**

* **Number of Buildings:** 25
* **Peak hour (Maximum for all building):** 11:00 am
* **Median Total Consumption:** 88,80,545.27 units
* **Median Total Weekday Consumption:** 1,26,73,968 units
* **Median Total Weekend Consumption:** 47,43,701units
* **Percentage Difference between Per Day Consumption between Weekday and Weekend:** 59.3%
* **Time of Day Peak in Weekdays:** Afternoon
* **Working Hours Median Consumption:** 82,42,486 units
* **Hourly Peak during Working Hours:** ~2000units
* **Hourly Valley during Working Hours:** ~604 units
* **Seasonal Peak:** Spring, 943.5units

**Interpretation**:

* Highest spring peak shows seasonal demand variability (e.g., air conditioning).
* Strong weekday/weekend difference may reflect higher commercial activity during weekdays.
* Peak during late morning to noon aligns with check-in/out or service operations.
* Very high overall and working hour consumption shows 24/7 operations typical of hotels and hospitals.

**Explanation:**

* **Spring peak**: Likely due to cooling demand (air conditioning), especially in guest rooms, kitchens, and lobbies.
* **Weekday/weekend difference**: Hotels may have more guests and events on weekdays (corporate bookings), while hospitals run consistently, yet outpatient services may dip on weekends.
* **Late morning peak**: Matches meal prep times, patient care, check-ins/check-outs, and peak customer service windows.
* **High working hour usage**: Confirms continuous operation, with HVAC systems, commercial kitchens, medical equipment, and lighting running most of the day.

### **Main Cluster 2 (Schools):**

* **Number of Buildings:** 98
* **Peak hour (Maximum for all building):** 9:00 (Morning)
* **Median Total Consumption:** 26,29,482.17 units
* **Median Total Weekday Consumption:** 19,89,343 units
* **Median Total Weekend Consumption:** 6,48,041 units
* **Percentage Difference between Per Day Consumption between Weekday and Weekend:** 51.5%
* **Time of Day Peak in Weekdays:** Morning
* **Working Hours Median Consumption:** 11,94,112units
* **Hourly Peak during Working Hours:** 885units
* **Hourly Valley during Working Hours:** 127 units
* **Seasonal Peak:** Winter, 262.3units

**Interpretation**:

* Morning peak reflects school start times.
* Weekend consumption drops significantly, confirming no weekend operation.
* High winter peak aligns with heating usage.
* Moderate total consumption and strong peak-time patterns indicate consistent, scheduled activity.

**Explanation:**

* **Morning peak**: Reflects classroom operations starting around 8–9 AM—when lights, fans, AV equipment, and admin offices are active.
* **Low weekend consumption**: Most schools are closed on weekends, hence the sharp drop.
* **Winter peak**: Heaters or increased lighting due to shorter daylight may drive up consumption.
* **Structured pattern**: Predictable usage mirrors school timetables—almost no energy used outside working hours, with sharp peaks and valleys.

### **Main Cluster 3 (Office, Shops):**

* **Number of Buildings:** 11
* **Peak hour (Maximum for all building):** 12:00 (Afternoon)
* **Median Total Consumption:** 1,74,34,670.91 units
* **Median Total Weekday Consumption:** 64,50,162units
* **Median Total Weekend Consumption:** 22,99,660units
* **Percentage Difference between Per Day Consumption between Weekday and Weekend:** 64.3%
* **Time of Day Peak in Weekdays:** Afternoon
* **Working Hours Median Consumption:** 37,48,957units
* **Hourly Peak during Working Hours:** ~3850 units
* **Hourly Valley during Working Hours:** ~1180 units
* **Seasonal Peak:** Autumn, 491.6 units

**Interpretation**:

* Highest peak in autumn suggests heavy usage from HVAC or lighting systems.
* Strong peak during noon shows alignment with business operations.
* Second-highest total and weekday consumption highlights full operational load.
* Large difference in weekday vs. weekend use confirms typical Mon–Fri business schedules.

**Explanation:**

* **Autumn peak**: May relate to increased HVAC needs during seasonal transitions (humid weather needing dehumidifiers or cooling).
* **Noon peak**: Business operations reach maximum activity during midday—full staffing, lights, computers, and equipment are in use.
* **Second-highest weekday consumption**: Reflects the intense load during weekdays from commercial operations.
* **Large weekday/weekend difference**: Most offices follow a 5-day work week. Shops may remain open but still contribute less to weekend load compared to weekdays.

**Conclusion:**

* **Cluster 1 (Hotels/Restaurants)** leads in **total and working-hour consumption**, particularly in **spring**.
* **Cluster 3 (Offices)** has the **highest hourly peaks**, especially during **autumn**.
* **Residential clusters (0a, 0b, 0c)** show distinct peak times and consumption patterns aligned with home occupancy.
* **Cluster 2 (Schools)** shows structured and time-bound usage patterns.

## **DR Program for each profile**

**Small Residential (Cluster 0)**

1. **Smart Plug-Based Load Control**

Smart plug-based load control lets utilities or homeowners remotely manage small appliances like fans, lights, TVs, and chargers using internet-enabled plugs. These devices can automatically shut off or delay non-essential loads during peak hours, reducing grid strain. Users retain manual override options for convenience. Utilities may offer incentives for participation, and real-time monitoring helps identify energy waste. This cost-effective and scalable solution is ideal for small to mid-sized homes that do not have advanced home automation systems.

1. **Prepaid Energy Plans**

Prepaid energy plans let consumers pay for electricity in advance, offering greater financial control and awareness of usage. Users receive frequent balance updates, encouraging mindful consumption, particularly during peak pricing hours. These plans are well-suited for small households with limited budgets, helping avoid unexpected bills and supporting energy-saving habits. Some utilities also provide usage alerts and consumption history, enabling smarter decisions and load shifting to off-peak times. Combined with time-of-use pricing, prepaid plans effectively promote demand response and more efficient energy behavior.

1. **SMS and App based DR Notification**

SMS and app-based demand response (DR) notifications offer a simple, low-cost way for utilities to communicate directly with residential users. During peak demand or grid stress, users receive alerts suggesting voluntary load reductions, like delaying appliance use or adjusting thermostats. These messages may include tips and link to incentive programs. Easy to implement and widely accessible, especially for homes without smart systems, this approach leverages growing smartphone use to quickly encourage energy-saving actions across a broad range of residential consumers.

**Multistory Buildings (Cluster 0)**

1. **In-unit Smart Thermostat Programs**

In-unit smart thermostat programs encourage residents to install intelligent thermostats that automatically adjust set-point temperatures during peak demand hours, reducing energy use while preserving comfort. These systems optimize air conditioning without manual input, making them especially effective in multistory buildings with high AC consumption. Utilities often support participation through rebates or bill credits. By shifting cooling loads away from peak periods, these programs help stabilize the grid, lower energy costs, and promote long-term efficiency through automated, demand-responsive climate control.

1. **Common Area Battery/Storage Use**

Installing batteries in the common areas of multistory residential buildings allows for energy storage during off-peak hours when electricity is more affordable and environmentally friendly. This stored energy can be used during peak demand periods, lowering the building’s reliance on the grid. It also offers backup power for critical services such as elevators and lighting. Strategically using battery systems helps reduce electricity costs, enhances grid stability, and enables participation in utility demand response or ancillary service programs, benefiting both residents and the energy system.

1. **Real-Time Feedback via Resident Portals**

Resident energy portals offer real-time insights into electricity usage and demand response participation at the individual flat level. These platforms enable residents to track their consumption, compare it with neighbors, and see savings from load reduction efforts. Features like gamification and personalized tips promote energy-conscious behavior. By making energy data accessible and interactive, these portals empower residents to manage usage more effectively, foster community-wide engagement, and enhance the overall success of demand response programs through increased awareness and participation.

**Duplex/Bungalow (Cluster 0)**

1. **Dynamic Pricing + Smart Home Integration**

Dynamic pricing, combined with smart home systems, encourages users to delay high-energy activities—like using ACs, water heaters, or EV chargers—until off-peak times. By adjusting usage based on electricity rates, households can reduce costs while easing grid demand during peak hours through automated appliance scheduling.

1. **Thermostat-Controlled HVAC Load Shedding**

Smart thermostats automatically increase the AC setpoint by 2–3°C during peak hours or DR events, reducing HVAC load without compromising comfort. This automated load shedding lowers electricity demand during critical periods and helps maintain grid stability while offering users incentives for participating in such programs.

1. **EV Charging Control**

EV charging can be scheduled during low-demand, off-peak hours—typically between midnight and early morning. This reduces pressure on the grid during peak times, lowers electricity costs for users, and aligns EV energy use with overall demand response goals, especially in high-consumption homes like bungalows or duplexes.

**Hotel / Restaurant / Hospital (Cluster 1)**

1. **HVAC Load Shifting (with Priority Zoning)**

Adjust HVAC settings by raising AC setpoints in non-essential areas like lobbies or hallways during peak hours, while maintaining full comfort in critical zones such as guest rooms or patient wards. This reduces load without compromising essential comfort or operations in hotels and hospitals.

1. **Thermal Energy Storage (TES)**

Utilize thermal storage systems to pre-cool or pre-heat water or air during off-peak hours, then use the stored thermal energy during peak periods. This strategy helps large facilities like hospitals and hotels reduce grid demand during critical times without affecting comfort or operations.

1. **Kitchen Equipment Scheduling**

Reschedule the operation of high-load kitchen appliances such as ovens, dishwashers, and fryers to off-peak hours. By staggering their use in restaurants or hospital kitchens, facilities can significantly lower energy demand during peak times without disrupting food service schedules.

1. **Laundry and Sterilization Timing (Hospitals)**

Shift energy-intensive laundry and sterilization tasks in hospitals to early morning or late-night hours. These flexible processes can be scheduled outside peak periods to reduce electricity demand while maintaining hygiene and operational standards, making it an effective non-intrusive DR strategy.

**Schools (Cluster 2)**

1. **Daytime HVAC Pre-Cooling**

Pre-cool classrooms early in the day before students arrive, then reduce HVAC output during mid-day peak hours. This maintains comfort while cutting energy use during high-demand periods. It’s an effective and non-disruptive DR method for schools operating on predictable daytime schedules.

1. **Lighting Control in Unused Areas**

Use occupancy sensors or automated schedules to turn off lights in unoccupied spaces like hallways, labs, or offices. Reducing unnecessary lighting during peak hours conserves energy, lowers electricity bills, and supports DR goals without affecting core academic activities.

1. **Time-of-Use Education & Awareness**

Launch awareness campaigns, challenges, or gamified programs to educate students and staff about energy use. Encourage actions like unplugging devices or shifting loads. These initiatives build energy-conscious habits and enhance DR participation across campus through behavioral engagement.

1. **Computer Lab Load Shifting**

Reschedule software updates, backups, or other high-energy computing tasks for after school hours. Shifting these operations avoids peak-hour demand, reducing pressure on the grid while maintaining lab availability for student use during regular hours

**Offices / Shops (Cluster 3)**

1. **Smart HVAC Scheduling & Optimization**

Use occupancy sensors and weather forecasts to adjust HVAC operation, ensuring efficient cooling or heating only when and where needed. Reducing HVAC usage during peak hours lowers energy demand without compromising occupant comfort, making it a smart, automated solution for offices and retail spaces.

1. **Lighting Dimming / Daylight Harvesting**

Install daylight sensors to automatically dim artificial lighting in areas receiving sufficient natural light, especially during peak hours. This strategy significantly reduces electricity usage while maintaining adequate illumination, making it ideal for office floors, showrooms, or retail spaces with large windows or skylights.

1. **Demand Response Alerts for Tenants**

Send real-time DR alerts to office or shop tenants via mobile apps or display screens, requesting voluntary load reduction during peak periods. Suggested actions may include adjusting thermostats or turning off non-essential lights. This approach encourages cooperation while offering potential incentives and fostering energy-aware behavior.

1. **Elevator Load Management**

During periods of low traffic or peak electricity demand, reduce the number of active elevators or delay non-essential trips. This lowers the overall building load temporarily without impacting accessibility, especially in commercial complexes with multiple elevator banks or alternative stair access.

1. **Office Equipment Control**

Automatically power down idle office equipment—such as printers, monitors, or conference room displays—during lunch breaks or non-working hours. Scheduling or sensor-based control of plug loads reduces unnecessary consumption and supports demand response efforts without interfering with productivity or daily operations.

# **CONCLUSION**

In this unsupervised clustering project, we extracted around 190 features from various sources—including U.S. holidays, seasonal patterns, time series metrics, and percentiles—for 1,277 buildings. By selecting the most informative subset of these features, we applied K-means clustering and identified four primary clusters.

Given the large number of buildings in Cluster 0, we conducted a deeper analysis and found it to be a mix of residential building types. We further subdivided this cluster into three meaningful sub-clusters. This refinement led to six well-defined building profiles: small residential homes, multistory residential buildings, bungalows/duplexes, large offices/shops, schools, and hospitals/restaurants/hotels.

Based on this in-depth profiling, we proposed targeted Demand Response (DR) programs and customized energy-saving strategies for each cluster. This tailored approach enhances the effectiveness of DR participation across diverse building types, promoting smarter energy consumption and improved grid reliability.