Project Report For CS661A: BIG DATA VISUAL ANALYTICS
2022-2023 Semester II

PowerViz – Energy Consumption Visualization Framework for Smart Home

Group - **11**
Team members: Antreev Brar *190163*, Suvam Basak *22111277*, Anshumann *190162*, Akshay Gupta *190093*, Akshan Agrawal *180061*

Member emails: antreev@iitk.ac.in, suvambasak22@iitk.ac.in, anshmn@iitk.ac.in, akshay@iitk.ac.in, akshan@iitk.ac.in
IIT Kanpur

1. Introduction

Nowadays, with cheaper sensing and actuation devices, the number of connected devices has exponentially increased in day-to-day life. In the context of smart homes, various home appliances are available. These appliances connect to the home network and provide a mobile application as a remote control. Also, these appliances have some smart sensors to work automatically without interaction from the user. In recent years, plenty of literature has been published about the security threat these connected devices bring [1, 2]. And a bunch of visualization software or web-based dashboard has been developed to show sensor data [3, 4]. Therefore, we found some gaps in visualization and analytical study in the domain of power consumption of such smart homes full of intelligent appliances. In this work, we proposed a new interactive data exploration and visualization system for a smart home. This will allow users to explore the power consumption and weather information for unexpected insight. The interactivity of the application provides data inspection on the web with a high degree of user experience.

2. Dataset

We used the Smart Home Dataset with weather Information [5] data set from *Kaggle*[1]. The dataset consists of 32 attributes of weather information, generated power, and consumed power by individual home appliances. The data is recorded with *1 minute* of granularity over *350 days*. The main **challenge** with this dataset is it doesn't contain the absolute timestamp in it [6]. The sequence number only gives the order of the data point being recorded.

3. Tasks

For exploration and visualization of the dataset [5] following tasks are carried out to build the proposed system:

1. **Timestamp correction**: Although we do not know where the data has been recorded, the weather condition also varies widely across the globe. However, it is a general intuition that the temperature rises in the daytime and starts falling after sunset. Therefore, the timestamp is adjusted with the fact temperature peaks at mid-day.

2. **Exploration of dataset**: The dataset [5] includes record of 32 attributes over *350 days* with *1 minute* of granularity. Which is high dimensional data. Just blindly plotting the dataset will get cluttered and overlapped, which will be the

[1] https://www.kaggle.com/

poorest visualization experience. Therefore, the *Exploration* technique must query the dataset on demand based on the user input and gives an interactive way to explore a certain part of the fetched data.

3. **Correlation of data attributes**: As the dataset contains 32 attributes, the visualization system should show a *Correlation* between the selected attributes by the user.

4. **Analysis of power generation and consumption**: The dataset has overall total power generation and consumption by individual appliances. Users should be able to explore day-wise generation and usage statistics with a comparative study of power usage of home appliances.

5. **Prediction from past trends**: From the previous 350 days of records, the framework should predict the future values from the past trends.

6. **Dimensionality reduction**: This dataset with 32 attributes is essentially high dimensional data. Our visual system is somewhat limited to 3D. Humans don't understand well more than three dimensions. Therefore, it is challenging to identify some meaningful patterns while analyzing a high-dimensional dataset. Therefore, many methods have been introduced for reducing dimensionality and embedding the data point from high to low dimensional space [7]. The visualization system should employ the following dimensionality reduction techniques:

   (a) *Principal component analysis (PCA)*: will show the linear relationship between the data points from the high dimensional space to the low dimensional space [7].

   (b) *t-Distributed Stochastic Neighbor Embedding (t-SNE)*: a nonlinear dimensionality reduction technique will show nonlinear relationships between data points [7].

   (c) *Uniform Manifold Approximation and Projection (UMAP)*: t-SNE also embeds the nonlinear relationship between points from higher to lower dimensions [8].

4. Proposed Solution

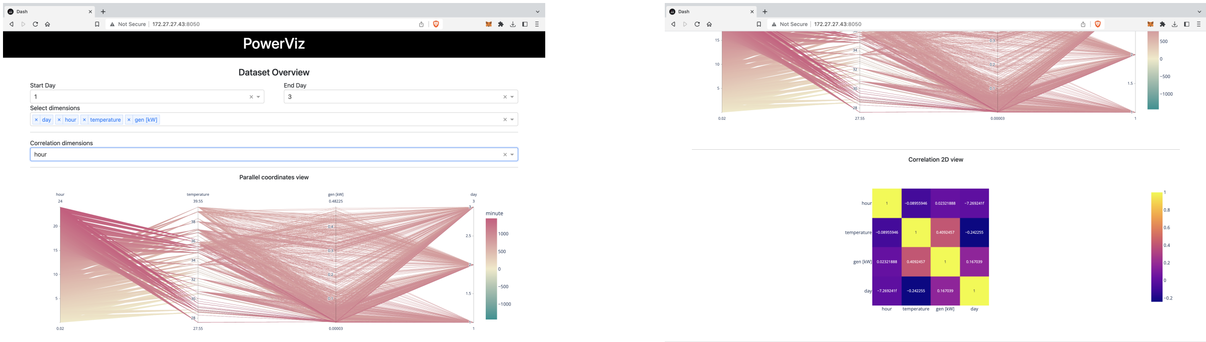
The proposed *Energy Consumption Visualization Framework for Smart Home - PowerViz* is a web-based visualization dashboard built with *Plotly Dash*[2] (uses Flask[3] for web server, React.js[4] for rendering UI, Plotly.js[5] for generating charts at the core).

The PowerViz dashboard is divided into four major sections:

1. **Dataset Overview** in Figure 1a, 1b allows users to explore the dataset on demand and shows correlations between attributes in the charts. It takes *Start Day*, *End Day*, *Dimensions* (attributes), and *Correlation dimensions* as input and queries the dataset and builds a Parallel Coordinates Plot[6] (PCP) dimensions order by the

[2] https://dash.plotly.com/
[3] https://flask.palletsprojects.com/en/2.2.x/
[4] https://react.dev/
[5] https://plotly.com/javascript/
[6] https://plotly.com/python/parallel-coordinates-plot/

correlation with Correlation dimensions input followed by one 2D Grid of correlation metrics[7] with all selected dimensions.



(a) Parallel Coordinates Plot   **(b)** Correlation Heatmaps

Figure 1: Dataset Overview

2. **Power and usage** in Figure 2 allows the user to inspect the power generation and usage daily through a Bar chart[8]. It also includes an interactive appliances input-driven Pie chart[9] for a comparative power usage analysis.

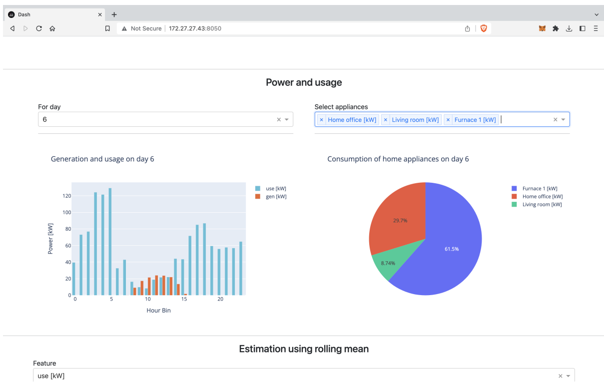


Figure 2: Power and usage - Bar and Pie chart

3. **Estimation using rolling mean** in Figure 3 shows a Line chart[10] of predicted value from past trends of the input feature by the users. The naive prediction hypothesis of any variable at time t can be modelled as: $\hat{y}_t = y_{t-1}$, i.e., value of variable in future just depends on its value at the past timestamp. However, this model can show great fluctuations at some timestamps. Therefore, we assume that the future value of our variable is dependent on its average of previous m values. Thus, this rolling moving average can be modelled as:

$$\hat{y}_t = \frac{1}{m} \sum_{n=1}^{m} y_{t-n} \tag{1}$$

The rolling moving average value for different attributes associated with power consumption, power generation and weather can be plotted.

[7] https://plotly.com/python/heatmaps/
[8] https://plotly.com/python/bar-charts/
[9] https://plotly.com/python/pie-charts/
[10] https://plotly.com/python/line-charts/

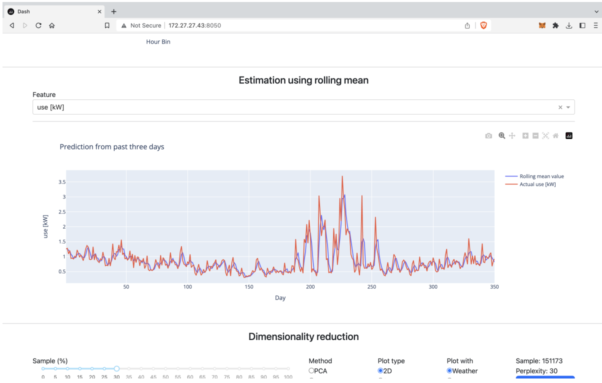


Figure 3: Estimation - rolling mean vs. actual use

4. **Dimensionality reduction** in Figure 4 allows three methods (i) PCA[11], (ii) t-SNE[12], and (iii) UMAP[13], and shows results in 2D[14] and 3D Scatter[15] plots. As the dataset is huge, it provides a slider for users to sample the part of the dataset with Simple Random Sampling (SRS)[9]. Moreover, another slider gets enabled to set the hyperparameters `n_neighbors`, `perplexity` when the reduction method UMAP, t-SNE is selected, respectively. The Dimensionality reduction can be done with only (i) Weather attributes, (ii) Power usage, and also with aggregated (iii) all data attributes.

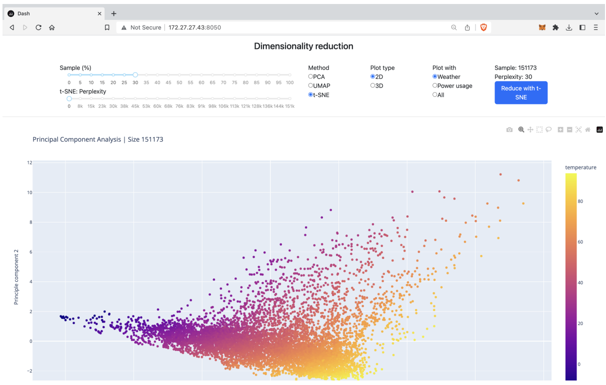


Figure 4: Dimensionality reduction

5. Results

The results we found from the *PowerViz* framework are shown in the following sections:

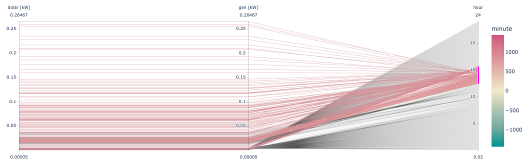
Dataset Overview

The exploration of the dataset gives the following insights.

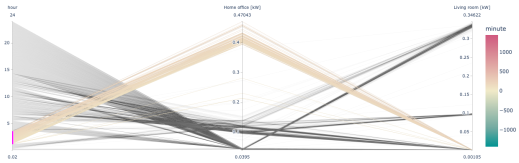
- In Figure 5a, the straight line between two axes, Solar [KW] and gen [KW], tells that the solar panel generates most of the power in the daytime. The rate of generation fluctuates because of clouds. At night no power is generated because of no sunlight.

[11] https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.PCA.html
[12] https://scikit-learn.org/stable/modules/generated/sklearn.manifold.TSNE.html
[13] https://github.com/lmcinnes/umap.git
[14] https://plotly.com/python/line-and-scatter/
[15] https://plotly.com/python/3d-scatter-plots/

- Figure 5b give some insight into the activity of the people. Sliding the selected region on the hour axis gives the idea of activity like working at night, staying in the living room, or going outside at night.



(a) Exploring power generation data



(b) Most of the power consumed by furnaces

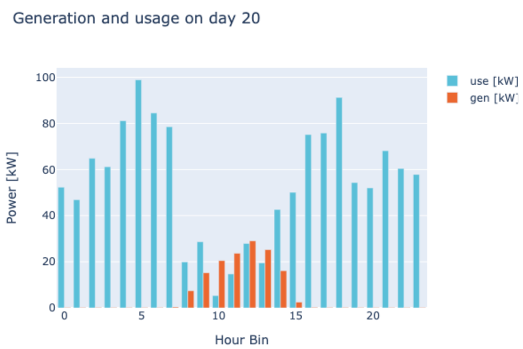
Figure 5: Presence of people from power consumption

Power and usage

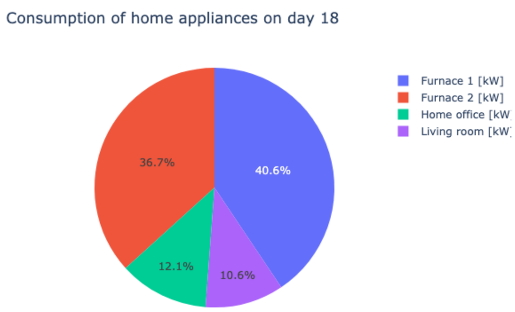
This section gives insight like

- The power generated from the solar panel from 7-8 AM to 4-5 PM, as shown in Figure 6a. At the same time, people probably are not at home (at work); hence the power consumption is high at night and morning.

- Figure 6b shows that Furnaces are the most power-hungry appliances. That also tells us the time is winter, as we don't have the time and date given implicitly in the dataset.



(a) Power generation and usage



(b) Most of the power consumed by furnaces

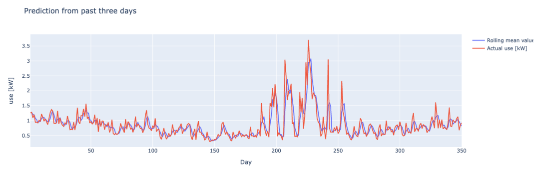
Figure 6: Insight from power and usage section

Estimation using rolling mean

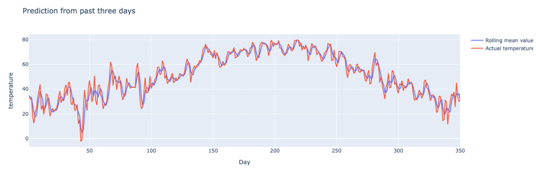
The estimation of future value from the past three days of data for power generation and temperature is shown in Figure 7a and 7b, respectively. The red line of actual values almost overlaps with the estimated values, showing the high accuracy of the estimation.

Dimensionality reduction

The dimension reduction with three different methods gives three types of results:



(a) Actual and predicted power generation



(b) Actual and predicted temperature generation

Figure 7: Prediction in estimation using mean rolling section

- The result of PCA gives some strips of the data points in 2D plots. The data point in these strips has some similar weather conditions.

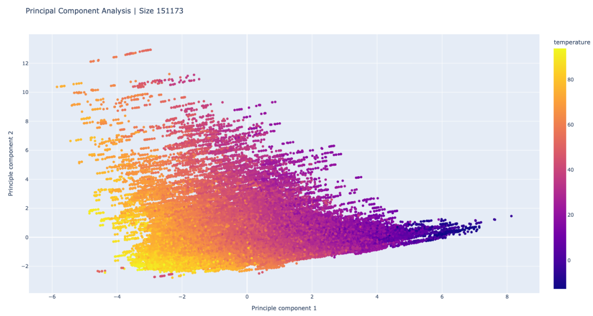
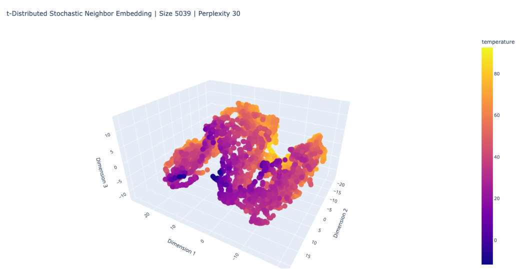
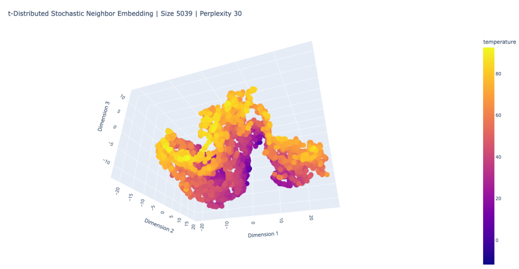


Figure 8: PCA

- The t-SNE method with 1% sample of the dataset with `perplexity` 30 produces a wave-like structure in in Figure 9. The structure essentially shows the weather transaction of a year from summer to winter. The same patterns probably repeat themselves yearly and could be verified with a few years of data.



(a) View from top with summer side points

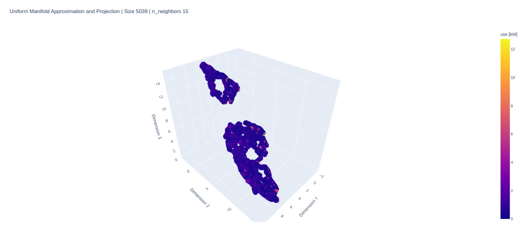


(b) View from bottom with winter side points

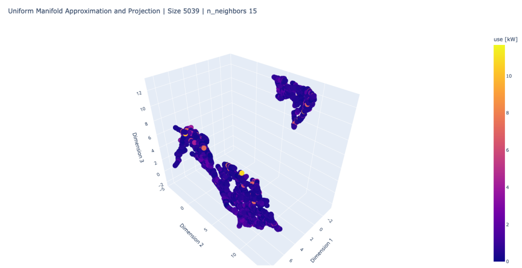
Figure 9: t-SNE with weather data attributes

- The UMAP method with 1% sample of the dataset with `n_neighbors` 15 produces a wave-like structure in in Figure 9. The structure essentially shows two clusters of points.

6. Conclusion



(a) View from one side



(b) View from another side

Figure 10: UMAP with all data attributes

PowerViz offers a powerful suite of visualization capabilities for analyzing weather, energy production and energy consumption data, allowing users to gain valuable insights into energy demand and production trends. This visualization tool, to some extent, can also help to provide a holistic overview of the day to day activities a person was indulged in. Moreover, PowerViz also offers forecasting capabilities to predict weather conditions, future energy demand and production based on historical power usage data and weather patterns. Moreover, it supports the dimensional reduction techniques that play a vital role in identifying and understanding important characteristics associated with the data. With its user-friendly interactive interface and advanced features, PowerViz makes it easy for analysts, business owners and scientists to create dynamic visualizations that can help to identify trends, patterns, and outliers in data. Thus, it is an invaluable tool for making data-driven decisions and optimizing energy production and consumption. Overall, PowerViz provides users with a comprehensive range of weather and power-related data that can facilitate a deeper understanding of intricate data sets and assist in making informed decisions.

7. Link to source code: https://github.com/suvambasak/PowerViz.git

Note: All the steps to run the project are given in `README.md`[16] file.

[16] https://github.com/suvambasak/PowerViz/blob/main/README.md

References

[1] Badis Hammi, Sherali Zeadally, Rida Khatoun, and Jamel Nebhen. Survey on smart homes: Vulnerabilities, risks, and countermeasures. *Computers & Security*, 117:102677, 2022.

[2] Haseeb Touqeer, Shakir Zaman, Rashid Amin, Mudassar Hussain, Fadi Al-Turjman, and Muhammad Bilal. Smart home security: challenges, issues and solutions at different iot layers. *The Journal of Supercomputing*, 77(12):14053–14089, 2021.

[3] Microsoft. Tutorial: Visualize real-time sensor data from Azure IoT Hub using Power BI. https://learn.microsoft.com/en-us/azure/iot-hub/iot-hub-live-data-visualization-in-power-bi, 2023. [Online; accessed 24-Apr-2023].

[4] 7 patterns for IoT data ingestion and visualization- How to decide what works best for your use case . https://aws.amazon.com/blogs/iot/, 2022. [Online; accessed 24-Apr-2023].

[5] TARANVEER SINGH ANTTAL. Smart Home Dataset with weather Information. https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information?select=HomeC.csv, 2019. [Online; accessed 24-Apr-2023].

[6] Yaswanth kumar Reddy. Smart Home Dataset with weather Information, Discussion (11). https://www.kaggle.com/datasets/taranvee/smart-home-dataset-with-weather-information/discussion/122943?select=HomeC.csv, 2020. [Online; accessed 24-Apr-2023].

[7] Soumya Dutta. Big Data Visual Analytics (CS 661A) - Lecture 11: High Dimensional Data Analysis and Visualization 1. https://hello.iitk.ac.in/sites/default/files/cs661asem223/6536f3634ac196a483a9ae0b7b331bc8ffc88b3727802921c7f118e3b6aab43a_Lecture11_130223.pdf, 2023. [Online; accessed 24-Apr-2023].

[8] Soumya Dutta. Big Data Visual Analytics (CS 661A) - Lecture 11: High Dimensional Data Analysis and Visualization 2. https://hello.iitk.ac.in/sites/default/files/cs661asem223/f0514fcd871a67262cf5d7596a8424e4f67d05a11d93b1aed47fff4eab3b8751_Lecture12_150223.pdf, 2023. [Online; accessed 24-Apr-2023].

[9] Soumya Dutta. Big Data Visual Analytics (CS 661A) - Lecture 17: Sampling for Big Data Visual Analysis 1. https://hello.iitk.ac.in/sites/default/files/cs661asem223/ed79142e44c1fc5cd7d38c0794870ca9c494e2c27238bd2f1d219e1358cec67f_Lecture17_200323.pdf, 2023. [Online; accessed 24-Apr-2023].