UNIVERSITY OF LJUBLJANA

MASTER'S THESIS

Development and Analysis of new Activation Based Load Profiles

Author:
Jakob JENKO, BSC

Supervisor:
Dr. Carolina FORTUNA and
Dr. Marko MEŽA and

A thesis submitted in fulfillment of the requirements for the degree of Masters of Electrical Engineering

in the

-
ICT

March 8, 2023

Declaration of Authorship

I, Jakob JENKO, BSC, declare that this thesis titled, "Development and Analysis of new Activation Based Load Profiles" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

"In science, great oaks grow from little acorns."

D. Everett

UNIVERSITY OF LJUBLJANA

Abstract

Faculty of Electical Engineering
ICT

Masters of Electrical Engineering

Development and Analysis of new Activation Based Load Profiles

by Jakob JENKO, BSC

An increasing amount of energy data holds enormous amounts of untapped potential. The efficient presentation of energy data to humans and machines through load profiles is the constant narrative throughout the thesis.

Besides the amount of data, the consumption of electricity itself is increasing. A third of electrical energy in the EU is consumed in residential areas. Therefore, optimizing consumption would leave a significant impact on reducing the human footprint.

Another problem that the EU is facing is an aging population. Using the energy data, it would be possible to design an anomaly detection system that would detect accidents such as falls, strokes or dementia-induced altered behavior.

The idea of solving big issues using big data led us to perform a comprehensive review of existing publications and use cases. It was found that some publications have already tried to solve the mentioned issues. Using the related work publications, we have made a table of possible ways to present the data, which left us with a lot of gaps. These gaps present load profiles that were not yet researched or used. The question we tried to answer next was, are these profiles really useful? It could be, that they are impractical, which could be the reason they were not used before.

During the thesis, we utilized these previously unused load profiles. This was done first by looking at how data and load profiles are related in high dimensional space using t-SNE. Newly obtained knowledge aided us when designing and constructing the elderly care system. The successful implementation of newly proposed load profiles proved that the new profiles can be efficiently utilized. While we filled in a few gaps, many are left to be researched.

Acknowledgements

I would like to thank the following people without whom this thesis would not be possible. Dr. Carolina Fortuna and Dr. Marko Meža helped me with the research, by providing continuous feedback and support. The scientific and open-source community provided the data and the tools used throughout the thesis. Friends and colleagues, who through discussion helped induce the ideas that were put forward through the thesis. My family, who inspired, supported and encouraged me to pursue my studies in the first place. And my partner Nika, who supported me through all my years of studies in all aspects possible.

Contents

Declaration of Authorship iii

Abstract vii

Acknowledgements ix

1 Exploratory data analysis of LPs using t-SNE 1
- 1.1 Introduction . . . 1
- 1.2 Methodology . . . 2
  - 1.2.1 LPs . . . 2
    - Weekly-Daily LP . . . 2
    - Bag of Appliances LP . . . 2
  - 1.2.2 Data . . . 3
  - 1.2.3 T-SNE Algorithm . . . 3
- 1.3 Results . . . 4
  - 1.3.1 Results for Per-Building LPs . . . 4
    - Normalized LPs . . . 6
  - 1.3.2 Per-Appliance . . . 8
    - Single Appliance Over Many Buildings . . . 8
    - Per-Appliance LPs - Comparing Appliances . . . 14
  - 1.3.3 Per-Appliance Per-Building . . . 18
    - Bag of Appliances . . . 18
- 1.4 Discussion . . . 19
- 1.5 Conclusion . . . 20

A The source code, high-resolution figures and datasets 21
- A.1 The source code . . . 21
- A.2 High resolution figures . . . 21
- A.3 Data and datasets . . . 22

B Expanded General Table 23

Bibliography 25

List of Figures

1.1 Weekly per-appliance LP . . . 2
1.2 Universal presentation of per-building per-appliance LP . . . 3
1.3 Projection of per-building LPs . . . 5
1.4 Projection of per-building LPs with actual samples . . . 6
1.5 Projection of normalised per-building LPs . . . 7
1.6 Projection of normalised per-building LPs with actual samples . . . 8
1.7 Projection of fridge LPs for various buildings . . . 9
1.8 Projection of fridge LPs for various buildings with actual samples . . . 10
1.9 Projection of kettle LPs for various buildings . . . 11
1.10 Projection of kettle LPs for various buildings with actual samples . . . 12
1.11 Projection of TV LPs for various buildings . . . 13
1.12 Projection of TV LPs for various buildings with actual samples. . . . 14
1.13 Projection of filtered per-appliance LPs . . . 15
1.14 Projection of grouped per-appliance LPs . . . 16
1.15 Projection of grouped per-appliance LPs with actual samples . . . 17
1.16 Projection of grouped per-appliance LPs with actual samples . . . 17
1.17 Projection of a bag of appliances LPs for various buildings . . . 18
1.18 Projection of a bag of appliances LPs for various buildings with actual samples . . . 19

List of Tables

B.1 Expanded general table of load profiles 23

List of Abbreviations

LP	**L**oad **P**rofile
TP	**T**able of load **P**rofiles
P	**P**ower (profile)
A	**A**ctivation (profile)
ZEB	**Z**ero **E**nergy **B**uilding
DR	**D**emand **R**esponse
AD	**A**nomaly **D**etection
EC	**E**lderly **C**are
DER	**D**istributed **E**nergy **R**esources
HVAC	**H**eating, **V**entilation and **A**ir **C**onditioning
EV	**E**lectric **V**ehiecle
PV	**P**hoto **V**oltaics
EU	**E**uropean **U**nion
NILM	**N**on **I**ntrusive **L**oad **M**onitoring
t-SNE	**t**-distributed **s**tochastic **n**eighbor **e**mbedding
PCA	**P**rincipal **C**omponent **A**nalysis
EDA	**E**xploratory **D**ata **A**nalysis

Chapter 1

Exploratory data analysis of LPs using t-SNE

1.1 Introduction

In this chapter, we will explore the use of t-SNE for Exploratory Data Analysis (EDA) on LPs. LPs are a valuable tool for understanding and analyzing the consumption patterns of appliances or buildings. However, they do not inherently allow for the comparison of activation patterns between different households or buildings. By using t-SNE, we can visualize the similarity of activation patterns and better understand the differences and similarities in consumption behavior. We will delve into the details of how t-SNE can be applied to LPs and the insights it can provide.

To achieve this goal, we will first provide a brief overview of t-SNE and its application to LPs. We will then describe our methodology for using t-SNE to analyze LPs and compare activation patterns. Finally, we will present the results of our analysis and discuss their implications for understanding energy consumption patterns.

The clustering of similar LPs was researched many times before, as it was described in related work Chapter **??**. We will be working with dimensionality reduction, where clusters are usually formed as a side product. The following clustering publications are worth mentioning. We have seen that authors [12], [15] and [1] have clustered regular one-dimensional LPs, as well as with 2D image-based load profiling in publications published by authors [22].

The publication by authors [2] compared various dimensionality reduction techniques for clustering and visualization of LPs. Their goal was to compare Principal Component Analysis, Isometric Feature Mapping, Sammon Mapping, Locally Linear Embedding and Stochastic Neighbor Embedding. They used daily power LPs from residential and industrial areas. This publication was of the closest resemblance to our goals, that we were able to find.

In all cases, work has been done with the power LP, whereas in this case, we will try to find similarities between activation profiles using a t-SNE algorithm. Most of the publications used single-time dimensions, whereas we will use two-time dimensions.

Although the use-cases were presented in-depth in Chapter **??**, it is worth mentioning one specific use case. The increasing price of energy resources, could lead to over-saving and living in cool homes. By using similarity metrics between profiles across different buildings, it would be possible to detect outliers when it comes to heating. With this approach, it would be possible to detect users, that are living in below-average cool homes and offer them cheaper plans.

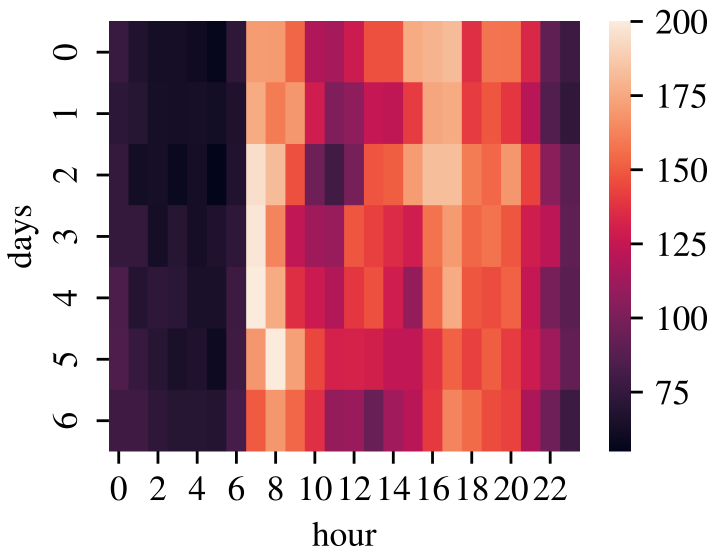
1.2 Methodology

1.2.1 LPs

Weekly-Daily LP

During testing, a weekly-daily LP constructed from a month of data will be used. Y-axis will present the days in a week and X-axis presents the hours in a day. Weekdays are labeled from 0 to 6, and hours from 0 to 23. Since we are working with images, the origin is placed in the upper-left corner. This means that a pixel in the upper-left corner presents the first hour of a week, this would be a Monday from midnight to one o'clock. The lower-right corner presents the last hour of the week. Since there are roughly 4 weeks in each month, each pixel will present 4 samples. One such example of profiles that we will use, was already presented in Chapter **??** with Figure **??**. For practical reasons, we are presenting it again here with Figure 1.1.

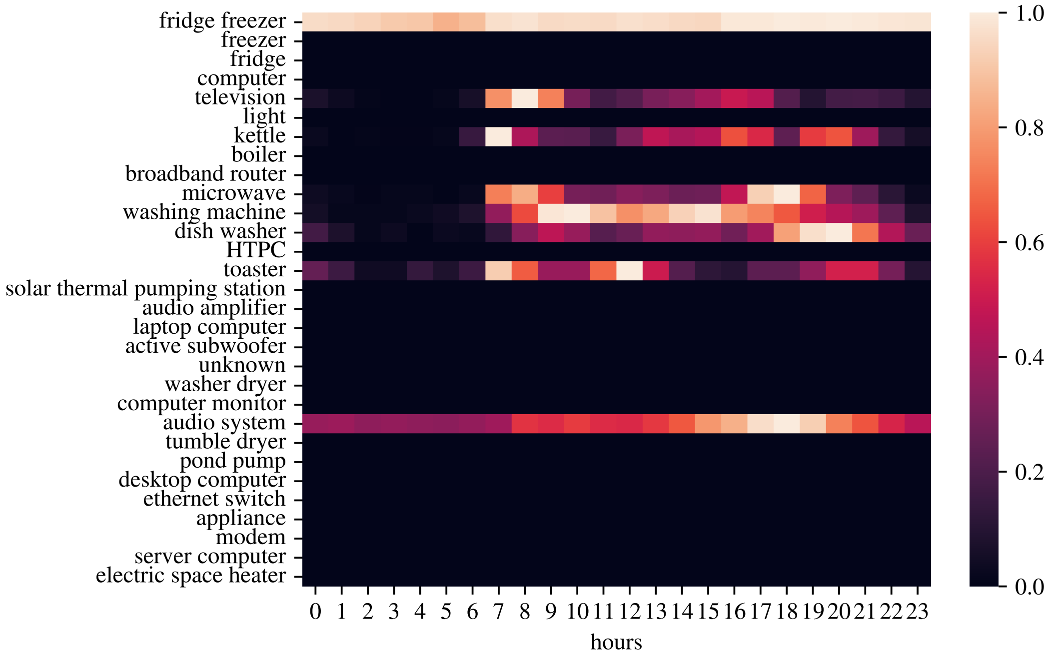
FIGURE 1.1: Weekly per-appliance LP



Bag of Appliances LP

Another LP that will be used at the end of this Chapter will be the bag-of-appliances LP. The profile was presented and analyzed in depth in Chapter **??** and was presented in Figure **??**. But again, for ease-of-use purposes, we will summarize the profile here.

FIGURE 1.2: Universal presentation of per-building per-appliance LP



To build the profile seen in Figure 1.2, we used the data from all 5 datasets and made a list of the most commonly used appliances. Only the top 30 appliances were selected. This enables us to have the same LP for all buildings, and thus enables us to see how the usage differs across them. One problem that arises here is the missing appliances. These appliances present themselves as a black line. A lot of missing appliances may cause the image to be primarily black, which could cause trouble for the algorithm processing this as an image.

1.2.2 Data

We have on average roughly one year of data per building. In some cases few weeks and in others up to 5 years for some appliances. By slicing this data into 1-month-long intervals and converting them to LPs we were able to obtain 5218 samples.

More detailed methodological approaches were discussed in Chapter **??** Subsection **??**.

1.2.3 T-SNE Algorithm

The t-SNE [21] or t-distribution stochastic neighboring embedding is a method for portraying high dimensional data in low dimensional space. This process is also known as dimensionality reduction.

One of the well-known dimensionality reduction algorithms is PCA. The key difference between the two is that one is linear, and the other is non-linear. PCA, linear, projects data in new space and finds the one with the least variance between data points. SNE [13], non-linear, is composed of two main parts. The first one is converting the high-dimensional Euclidean distances between data points into conditional probabilities that represent similarities [13]. The pairs with high similarity have a high probability, and pairs with lower a low probability. Second, it uses Kullback-Leibler divergence to minimize it with respect to a location on a map. To achieve this it uses gradient descent to minimize the cost function. Over many iterations, similar data points should be close together and far away from dissimilar objects. Similar data points usually form clusters.

t-SNE uses SNE as a basis, except that it uses t-student distribution instead of normal to calculate the similarity.

In our case, two dimensions will be used. Since this is a non-linear dimensionality reduction, the axis usually presents dimensions that are hard to comprehend by the brain. It is important to keep in mind that the resulting low-dimensional representation is not necessarily interpretable in the same way as the original high-dimensional data. This also means that the axes on the graphical presentations are meaningless. In our case, we labeled the two axes as comp-1 and comp-2.

Since the algorithm uses similarity at the base of the algorithm, it is possible to see which samples are more similar to each other. This is also the main idea behind the algorithm.

The t-SNE algorithm is useful for visualizing patterns and relationships in the data, such as clusters of similar data points and relationships between different data points. By observing these patterns, it is possible to gain insights into the data for further analysis.

1.3 Results

The results will be presented in three subsections

- Per-building LP
- Per-appliance LP
- Per-building per-appliance LP

Most of the focus will be done on the per-appliance LP since it is the most universal.

1.3.1 Results for Per-Building LPs

This LP is useful when it comes to comparing how activation patterns change over buildings and datasets. Per-building data uses combined activations of all appliances to present the aggregated usage pattern.

Figure 1.3 is using non-normalized data, meaning the number of appliances in a building will affect the end LP. The algorithm could pick up on how many appliances are being used. In some cases, such as energy poverty detection, this information is useful, again in others we would like to find a more complex usage pattern.

FIGURE 1.3: Projection of per-building LPs

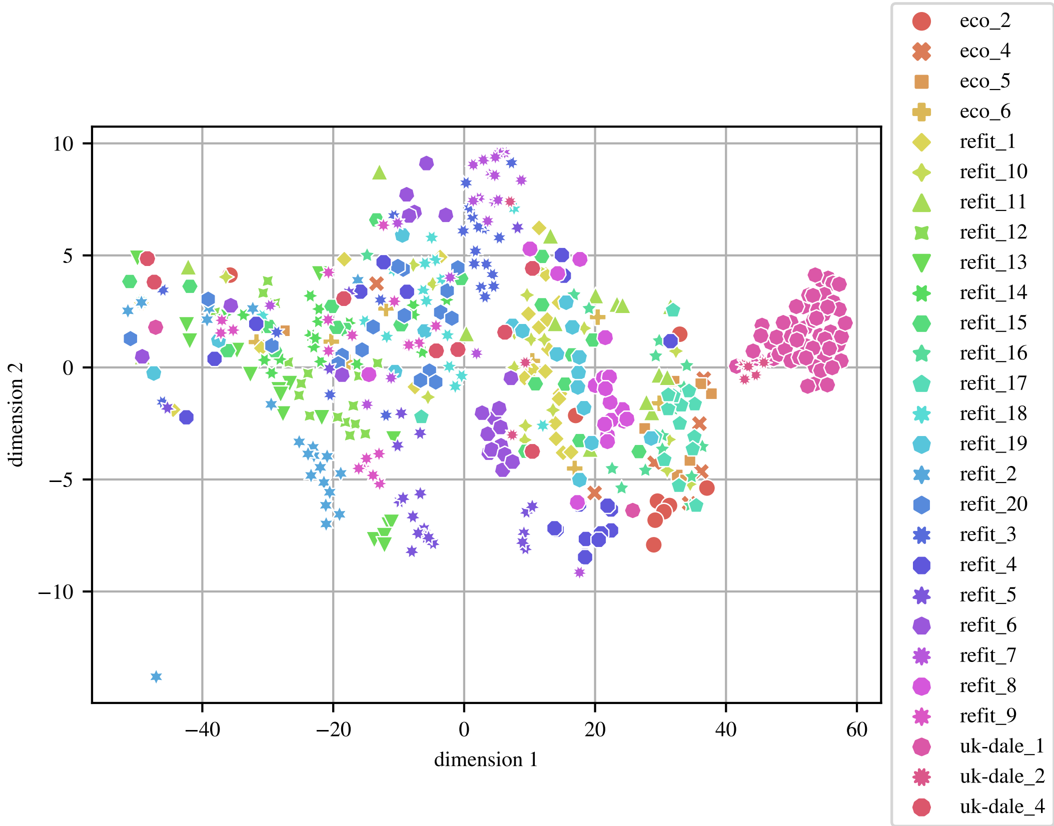


Figure 1.4 below presents the actual LP for each sample. It is possible to see that on the left there are mostly samples with very little activity, and on the right, we see samples with more activity. Since the two plotted components are of a higher dimension, it is hard to determine what they present. As said t-SNE gives us the intuition of how LPs are connected in higher-dimensional space.

The following figures are best viewed in color and a digital format. Readers reading the digital version should have the ability to zoom into each cluster, and see the actual samples. Readers reading a paper version can still explore the high-resolution figures online via the provided link at the end of the document.

FIGURE 1.4: Projection of per-building LPs with actual samples

Normalized LPs

To solve the issue mentioned in Subsection 1.3.1 have to normalize the data between 0 and 1. Figure 1.5 shows how normalizing samples affect the algorithm.

When comparing figures 1.3 and 1.5, it is possible to see that the samples on the latter are much closer to each other, while it is still possible to see the individual clusters. This could imply that the normalized usage pattern of users is more similar to the activation pattern of users. A normalized activation pattern tells us at what part of the day the appliances will most likely be used, and the activation pattern tells us how much will the appliance be used in each part of the day. Based on that, we can conclude the time when the appliance is used is more consistent than how much it will be used.

FIGURE 1.5: Projection of normalised per-building LPs

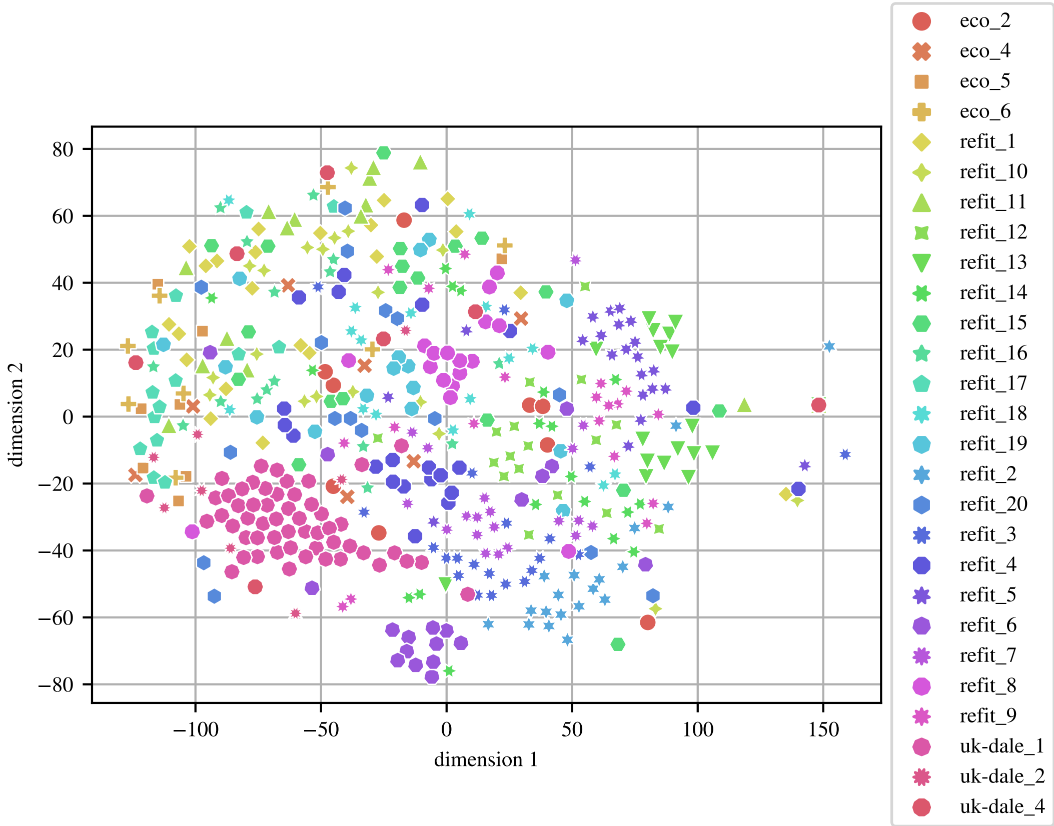


Figure 1.6 presents only the main cluster of samples. Since the smaller cluster presents mostly low entropy data, it was cut out. If the reader wants to see the samples in the cluster, the very same cluster can be found on the far left in Figure 1.4.

FIGURE 1.6: Projection of normalised per-building LPs with actual samples

In Figure 1.6 it is possible to find various usage patterns. But the general pattern is that there is less activity during the night with one peak in the morning and evening hours. Some buildings are more active during the week and again some more during the weekend. A lot of the data is from UK-DALE building 1 (pink box). It is possible to see that the building has one big cluster where activations are generally similar, with few outliers, where the pattern completely changed. Albeit less obvious, this pattern is the same for all buildings. This happens due to events such as vacations, holidays or weather-induced behavioral changes.

1.3.2 Per-Appliance

We can use per-appliance LPs to examine how different appliances are used in a single building, how a single appliance is being used across other buildings or how many appliances are being used in many buildings.

Per appliance LPs are built using sub-meter data, meaning each LP should present each appliance.

Single Appliance Over Many Buildings

Using one appliance and the building as a label, allows us to examine how the same type of appliance is being used across different buildings.

Fridges are generally a bad indicator when it comes to user behavior since the user does not affect its operation. The only case when the user interacts with it is when opening the door and turning on the light inside. Usually, this event is dwarfed by the activations of a compressor. This also means that the usage pattern should be the same across all buildings. This can be seen in Figure 1.7, where apart from REFIT buildings 1 and 11, there are no clusters.

FIGURE 1.7: Projection of fridge LPs for various buildings

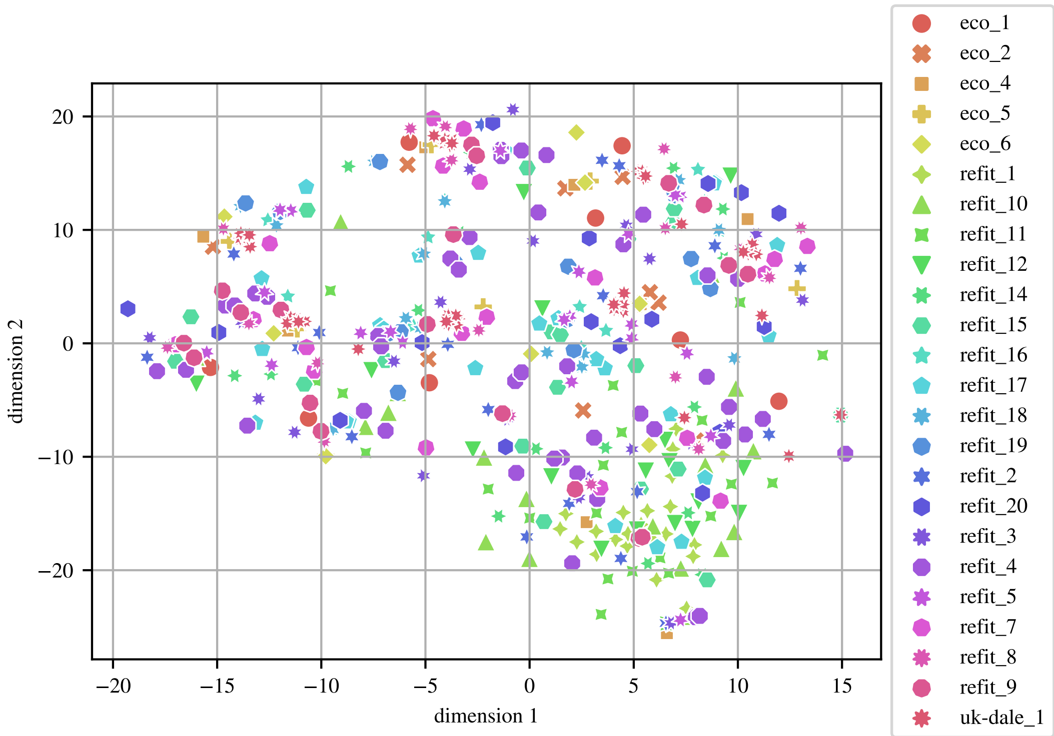


Figure 1.8 Shows mostly bright images, apart from a few outliers. LPs scattered in a circle are generally less dynamic than the ones at the bottom. Figure 1.8 is a good example of how LPs with little to no human interaction, can look a lot different. This could be due to different makes of the appliances, malfunctions of the appliance or the meter measuring it.

FIGURE 1.8: Projection of fridge LPs for various buildings with actual samples

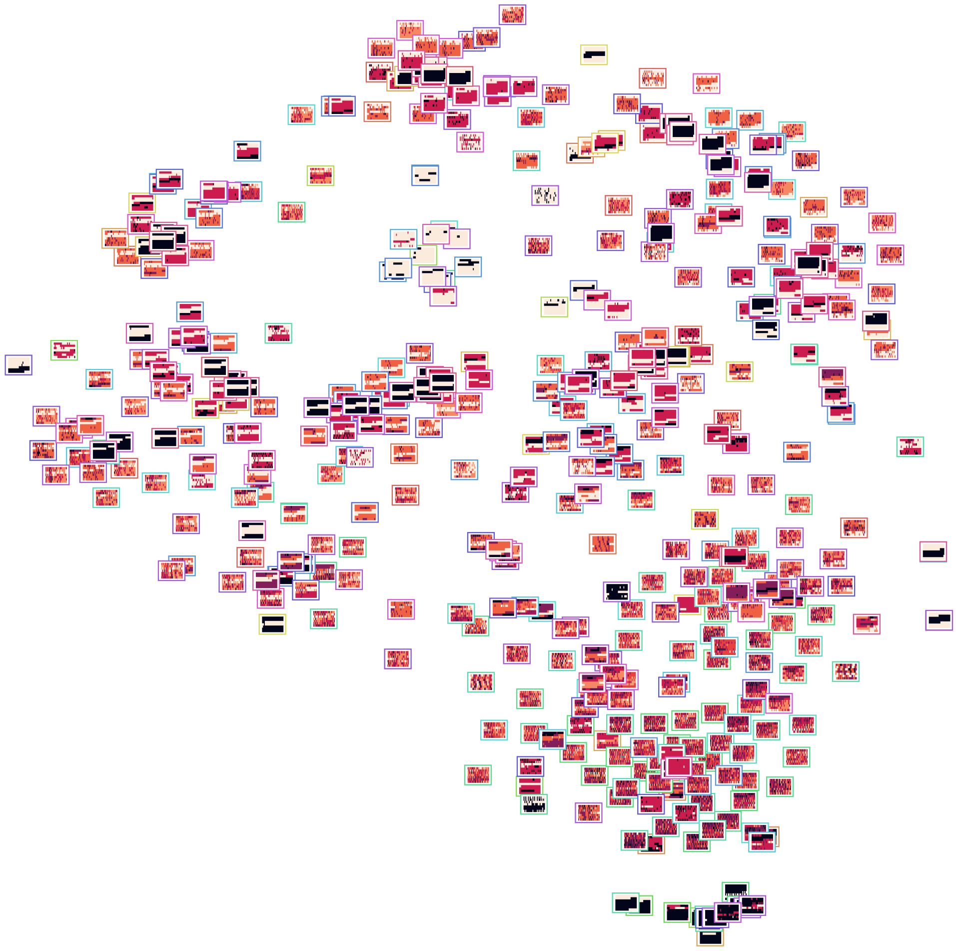


Figure 1.9 shows how, compared to fridges, kettles have many clear clusters that are spaced out between each other. This could mean that every household uses a kettle a bit differently. This cluster is a good example where we can see how strong is a routine of a user. The closer together the clusters, the higher the routine since samples are more similar to each other.

FIGURE 1.9: Projection of kettle LPs for various buildings

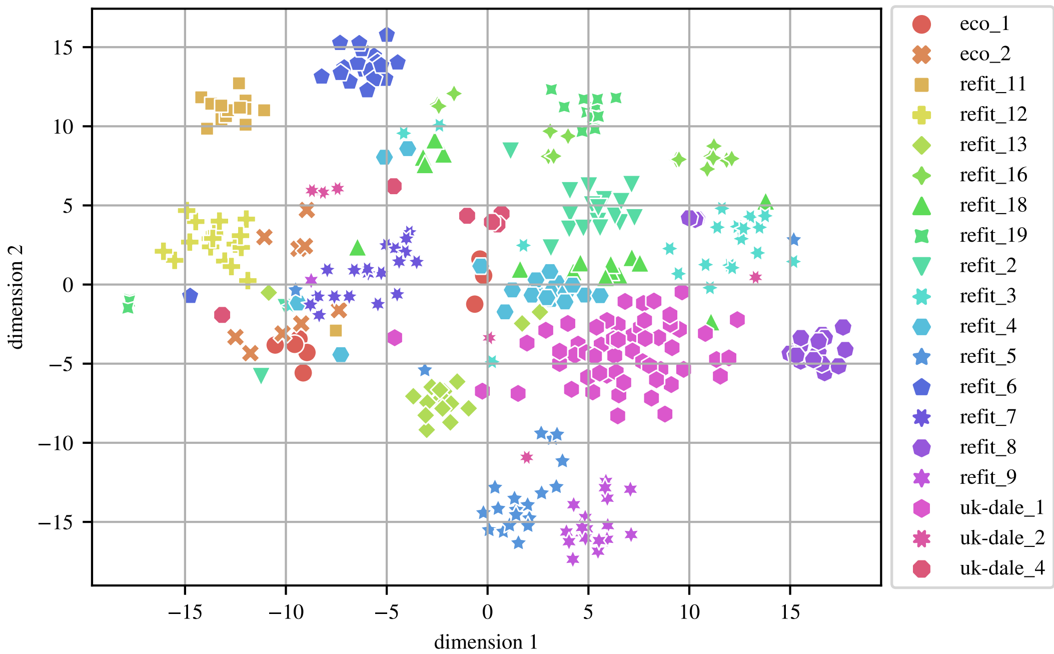
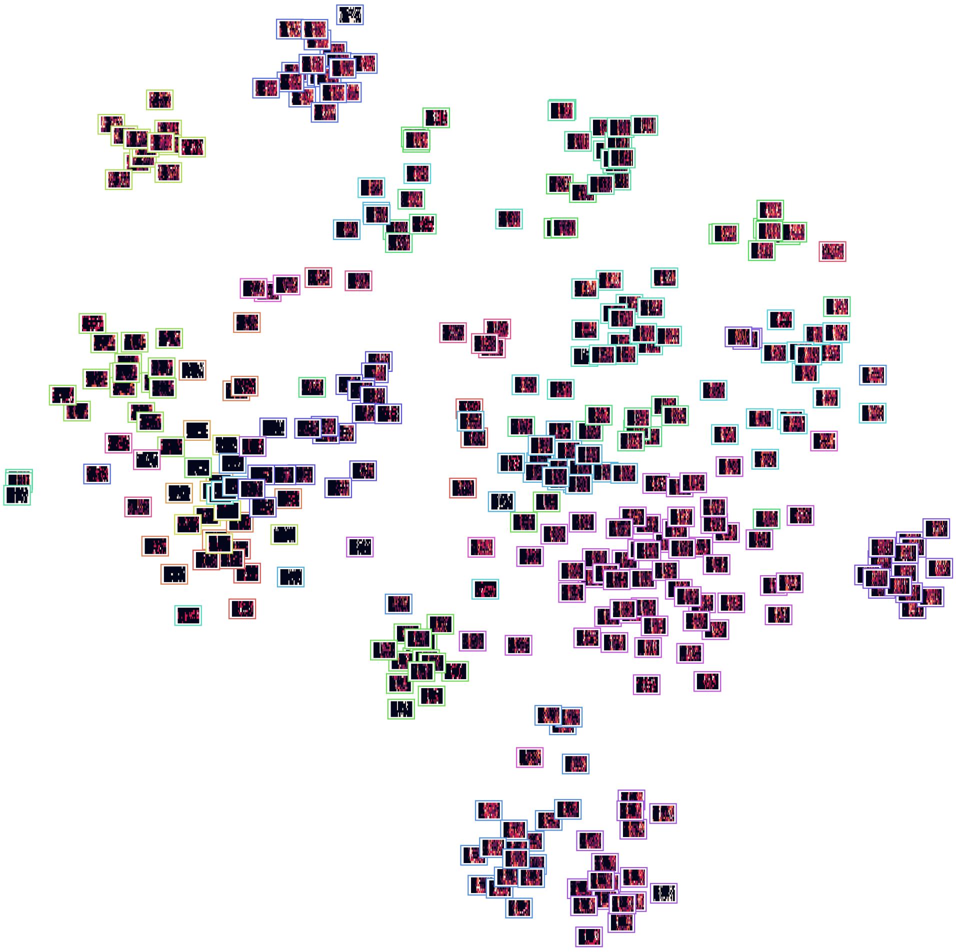


Figure 1.10 shows us that images on the lower part of the plot contain less activity than the others. LPs that are closer together have more similar activation patterns. Similar activation patterns are caused by similar behavior, which is essentially a routine. This means that this projection could be used to calculate how much a behavior variates in time for each building. This could be calculated by measuring the scattering of samples (variance) for each building.

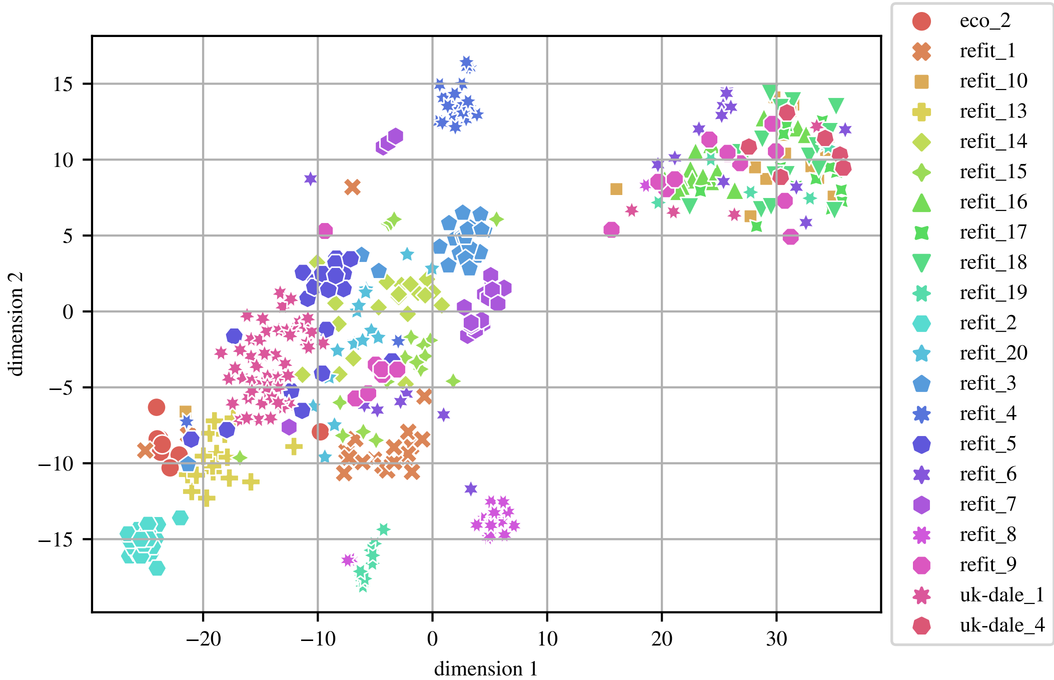
If we find samples that always activate in the same morning buckets, we would see that they form a straight line on the y-axis. This is the daily routine. One such example can be seen in Figure 1.9 in cluster refit 5 and refit 9, where we can see the lines and the pattern throughout the day. Since the routine is present, the samples look more similar and are therefore closer together. This does not necessarily mean that the closer the samples higher the routine. They could also be together in case of "ordered chaos" such as can be seen in Figure 1.9 for building refit 16 and refit 8 where there is no pattern through the day. So the scattering is not a precise metric when it comes to the routine, but it gives us a rough idea of its presence. The strength of a routine is an important feature that will be used in Chapter **??** to build an elderly care anomaly system.

FIGURE 1.10: Projection of kettle LPs for various buildings with actual samples



The last per-appliance example is television presented in Figure 1.11. Television was chosen since it is the most commonly occurring appliance. Interestingly enough, televisions form nice clusters with a few outliers. Clusters are separated but close together, this could mean that usage patterns across buildings are unique but not that different from one another. The LPs in some clusters are also close to each other, which could also indicate a higher routine.

FIGURE 1.11: Projection of TV LPs for various buildings



The images in Figure 1.12 prove the fact that outliers' consumption is a lot different. Again the bright images could be the results of faulty appliances, faulty meters or simply odd behavior. Figure 1.12 also enables us to see that TVs are primarily used in the evening hours. Outliers from the main cluster show slightly different behavior. One such example is the blue cluster (building REFIT 4), where appliances are mostly used in the morning hours. One other interesting observation can be made when looking at the purple cluster. This is the far low cluster for building REFIT 8. Here, the TV is being consistently used every day in the early morning hours. This is portrayed as a straight line. There could be two possible explanations for this. First is simply a high routine of a user, who turns on the TV every morning to listen to the news. The other is that the TV updates itself every morning. This is probably not the case since updates do not occur on regular basis. What is also interesting, is that the very same pattern can be observed in a few other buildings, one example being building REFIT 19.

FIGURE 1.12: Projection of TV LPs for various buildings with actual samples.

Per-Appliance LPs - Comparing Appliances

To get a general idea of where each appliance group lies, let's filter out all appliances that have less than 150 samples. Applying this filter yields Figure 1.13.

FIGURE 1.13: Projection of filtered per-appliance LPs

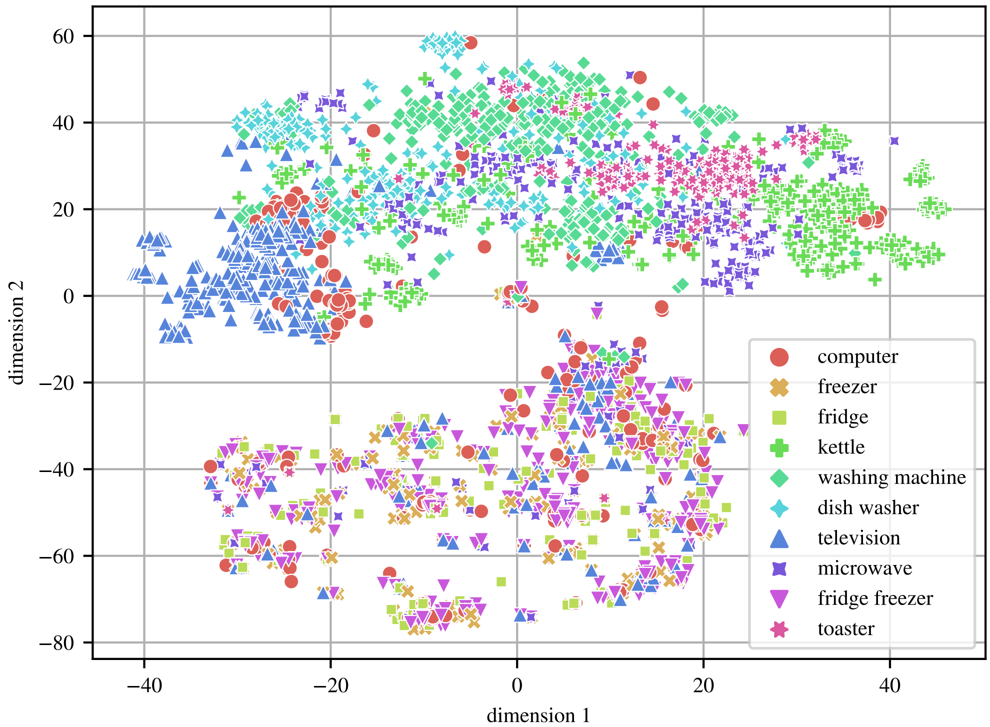


Figure 1.13 shows how these 10 appliances are connected in high dimensional space. Kettles, microwaves and toasters are quite similar when it comes to usage patterns. They are operated for a short amount of time and are usually used in users' routines in the morning or evening. These appliances are located in the upper left part of the plot.

The second group of appliances that are quite near each other is white goods (without fridges) such as washing machines, dishwashers, dryers etc. Let's say that they are white goods with a program. This group of appliances is located in the upper right part of the plot.

The third group of appliances is white goods with a compressor. They are usually not affected by human interaction and are therefore harder to cluster. They are located in the lower part of the plot.

The final group of appliances is televisions and computers. They lie on a bridge between the fridges and other groups.

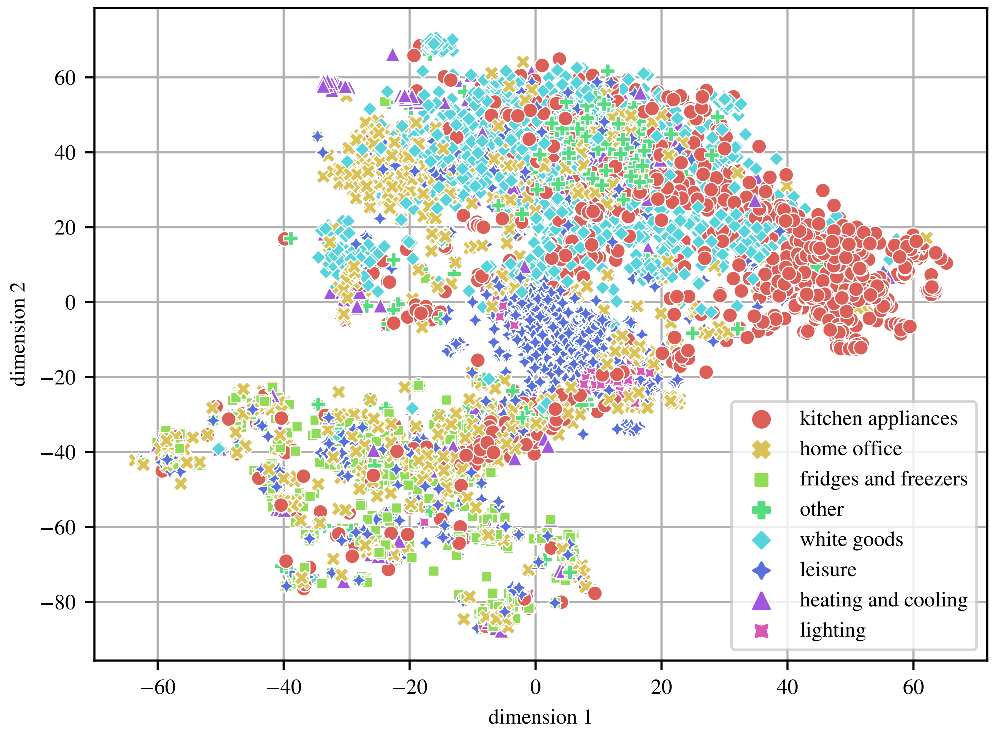
Knowing that a pattern exists, we can use the newly found group to define new appliance groups. The following 8 groups will be defined

- Kitchen appliances - toasters, ovens, microwaves, etc.
- Fridges and freezers - contains fridges, freezers and fridge freezers or white goods with a compressor
- White goods - washers, dryers, dishwashers i.e. white goods with a program
- heating and cooling - Electric radiators, dehumidifiers and HVACs
- leisure - Living room appliances such as TVs, games consoles, audio amps, HTPCs, etc.
- home office - Computer, laptops, printers, network equipment, chargers, etc.

- lightning - lights and lamps
- Others - unknown and unlabeled appliances

Applying these groups yields Figure 1.14. The new plot shows how, although appliances could be used by a different user, maybe even by users in a different part of the EU or world, they can be grouped in a high-dimensional space.

FIGURE 1.14: Projection of grouped per-appliance LPs



The Figure 1.15 below is the same as the first Figure **??** in the subsection, except it is easier to use color to see the appliance they present

FIGURE 1.15: Projection of grouped per-appliance LPs with actual samples

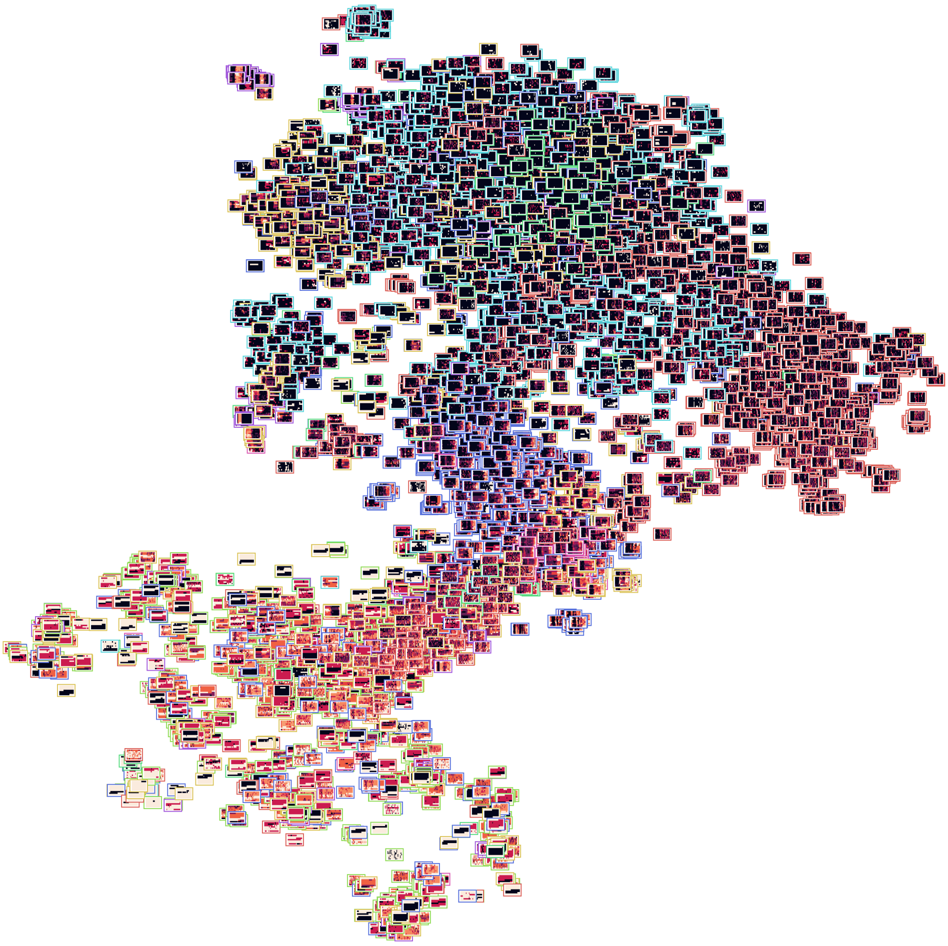
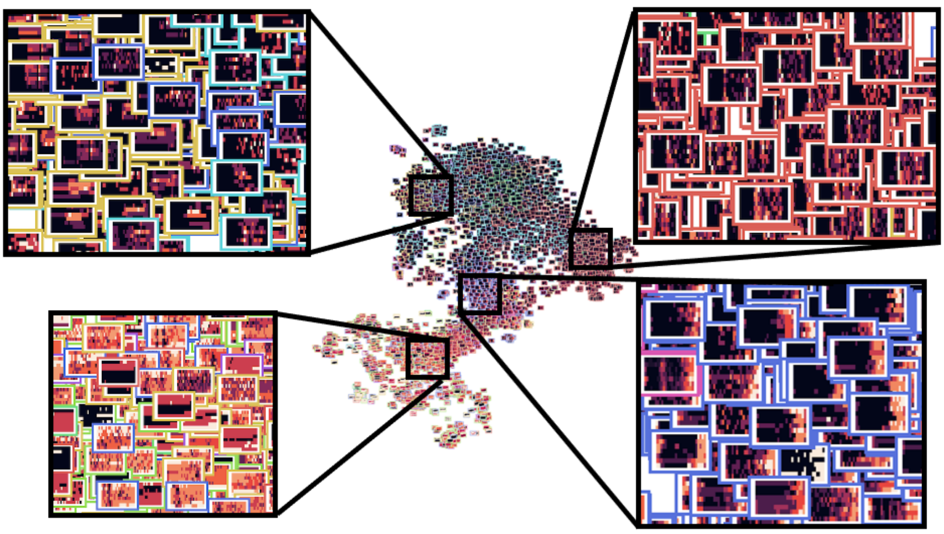


Figure 1.16 shows the four main types of profiles for readers that cannot zoom in.

FIGURE 1.16: Projection of grouped per-appliance LPs with actual samples



1.3.3 Per-Appliance Per-Building

To study the usage by comparing all appliances between buildings, we have to use one of the proposed LPs and in this case, this is a Bag of appliances.

Bag of Appliances

This LP is a combination of the LPs above, except it offers a larger detail when observing groups of appliances. Since we are using one dimension for appliances, we will use only the daily dimension.

To construct such a profile we need a universal way of constructing it. This is done by measuring how many times each appliance occurs in the datasets, then this list is sorted from most common to least common, and finally, the top 30 are selected.

The problem with such a comparison is, that it is best if all buildings would use the same appliances. Since that is not the case, missing appliances are portrayed as always off.

This is the main reason why we can see in Figure 1.17 the clusters are separated quite a bit. We can still see that some clusters are closer than others, meaning they are more similar.

FIGURE 1.17: Projection of a bag of appliances LPs for various buildings

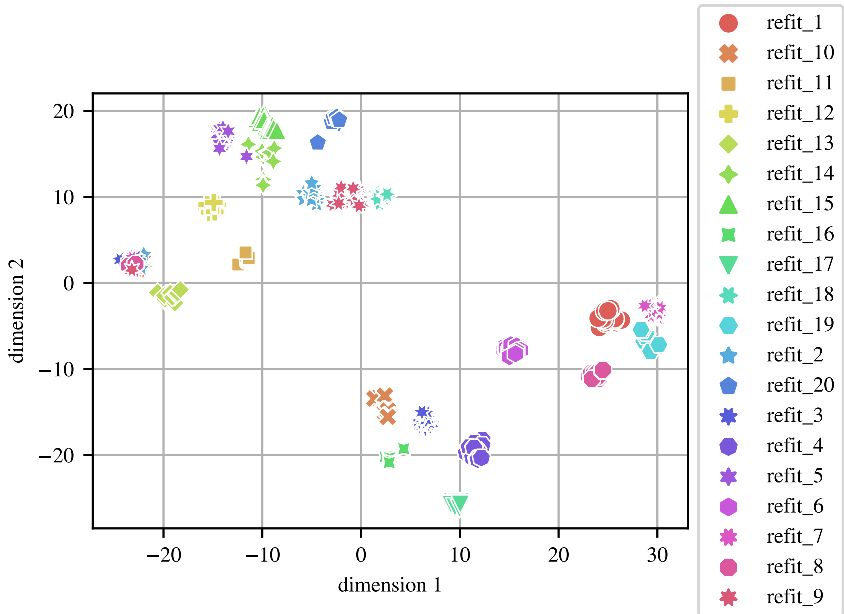
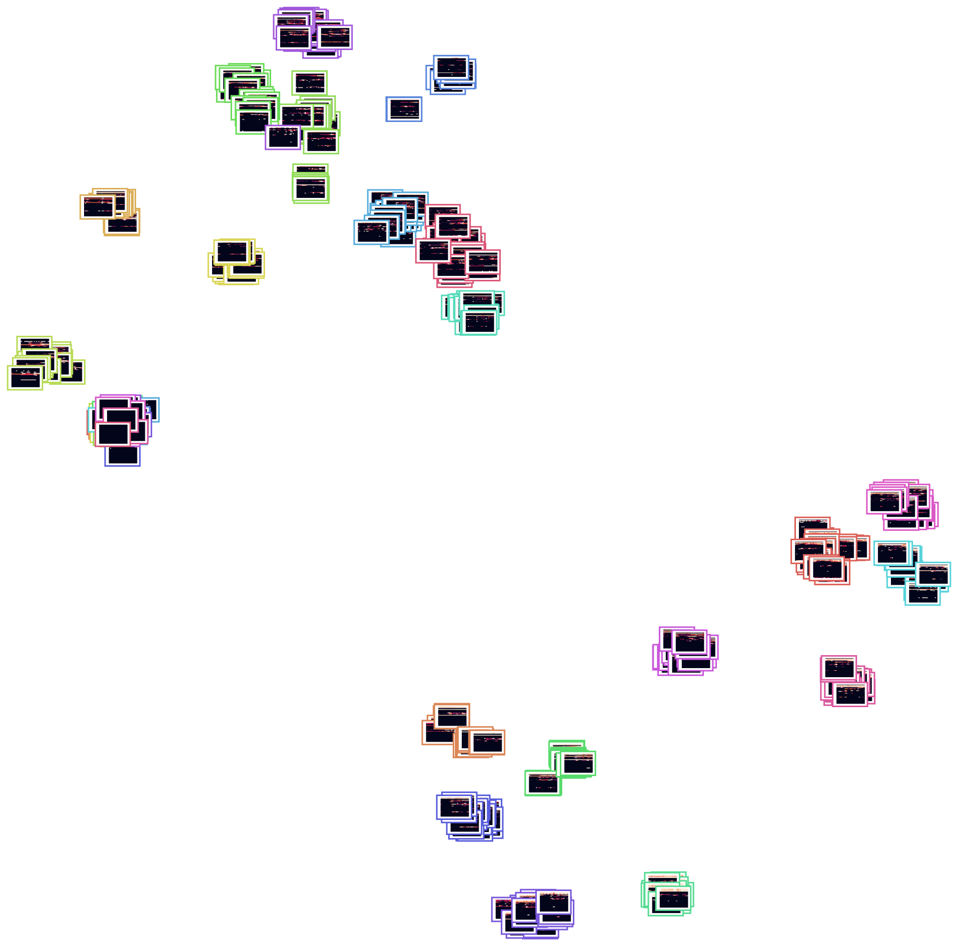


Figure 1.18 shows that LPs are split between two poles. By observing the Figure it is possible to see that all the bottom clusters have more than one active white good with a compressor (fridges and freezers), while the top ones have only one. In general, the bottom buildings have more appliances, with more activity than the top ones.

FIGURE 1.18: Projection of a bag of appliances LPs for various buildings with actual samples



1.4 Discussion

In this chapter, we used t-SNE dimensionality reduction to compare and analyze load profiles (LPs) in order to understand how data is related in high-dimensional space. We used three different types of LPs: per-building, per-building per-appliance, a bag of appliances, and per-appliance. Through our analysis, we were able to identify patterns in energy consumption across different buildings and datasets, as well as compare the usage patterns of different appliances within and across buildings.

One of the key findings of this chapter was the formation of appliance groups. These groups allowed us to more clearly understand the connections between samples and revealed the existence of routine and persistent usage patterns. This information can be particularly valuable for identifying patterns of energy usage in assisted living settings that may be related to health or well-being. This will be further explored in the next chapter.

In addition, our analysis showed the versatility of per-appliance LPs in understanding energy consumption patterns. These LPs allowed us to compare appliance usage across buildings, compare different appliances with each other, and understand the connections between appliance usage patterns within a single building. This highlights the importance of considering the usage patterns of individual appliances in efforts to optimize energy consumption.

Overall, our analysis using t-SNE has provided valuable insights into the relationships between LPs and the patterns of energy consumption. We believe that this work has the potential to contribute new tools and techniques for studying and optimizing energy consumption patterns, and we hope that it will inspire further research in this field.

1.5 Conclusion

In conclusion, this chapter demonstrated the usefulness of t-SNE in visualizing and comparing load profiling data in high-dimensional space. We were able to identify patterns and connections between different types of LPs, and we were able to group appliances into categories based on their usage patterns. These findings will be valuable in the next chapter as we continue to explore the potential applications of load profiling in assisted living settings. By analyzing and understanding the data in this way, we hope to contribute new insights and techniques for studying energy consumption patterns and identifying opportunities for optimization.

Appendix A

The source code, high-resolution figures and datasets

The following appendix contains mostly links that point to GitHub. GitHub should be a valid and safe repository for such projects, where links should persist indefinitely. In case the links do eventually break, you can find the repository of thesis and demos under the user name "jenkoj", under "msc" and "appliance-profiling". In case the Google Drive link breaks, send an email to jakob.jenko@gmail.com and I will try to forward the documents.

A.1 The source code

The source code used in this chapter can be found in a GitHub repository:
https://github.com/jenkoj/appliance-profiling

Individual scripts can be found in the following Jupyter Notebooks:

The source code for generating the figures can be found at:
https://github.com/jenkoj/appliance-profiling/blob/main/profilng_slices.ipynb

The source code for t-SNE can be found at:
https://github.com/jenkoj/appliance-profiling/blob/main/profilng_slices.ipynb

the source code for elderly care can be found at:
https://github.com/jenkoj/appliance-profiling/blob/main/elderly_care_demo.ipynb

A.2 High resolution figures

High-resolution figures can be found in the thesis repository https://github.com/jenkoj/msc.
More precisely in the "figures" subfolder https://github.com/jenkoj/msc/tree/main/Figures.

A.3 Data and datasets

We cannot share the data since it is not ours to share, but we can share the spreadsheet that points to each dataset. The spreadsheet includes other datasets that could be used for the very same purpose. The spreadsheet can be found here.

https://github.com/jenkoj/msc/blob/main/Appendices/datasets_and_sources.pdf

An overview of the data in the datasets was made, and it can be seen in the following spreadsheet.
https://github.com/jenkoj/msc/blob/main/Appendices/dataset_overview.pdf

The sliced hourly datasets can be found here.
https://drive.google.com/drive/folders/1tIsG-bqxoJdbU1p8xa_LCTaKNSk_LylZ?usp=sharing

Appendix B

Expanded General Table

TABLE B.1: Expanded general table of load profiles

	frequency	appliances	number of activations	power (avg)	operating time
appliances		X	X	X	X
number of activations	X	[6] [18]	X	X	X
power (avg)	X	[25]		X	X
power (array)	[18]	X	X	X	X
power (histogram)			X	X	X
operating time	X	[16]	[24] [23] [3]	[3]	X
time array	X	X	[6] [18]	[8] [9] [4] [17] [26] [12] [11] [15] [1] [19] [24] [23] [14] [3] [7] [20] [8] [5] [22] [18] [11]	[10]

Bibliography

1. Joana M. Abreu, Francisco Câmara Pereira, and Paulo Ferrão. "Using pattern recognition to identify habitual behavior in residential electricity consumption". In: *Energy and Buildings* 49 (2012), pp. 479–487. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2012.02.044. URL: https://www.sciencedirect.com/science/article/pii/S0378778812001363.
2. A. Aréchiga et al. "Comparison of Dimensionality Reduction Techniques for Clustering and Visualization of Load Profiles". In: *2016 IEEE PES Transmission and Distribution Conference and Exposition Latin America*. 2016, 1 to 6. DOI: 10.1109/TDC-LA.2016.7805661.
3. Elnaz Azizi, Mohammad T. H. Beheshti, and Sadegh Bolouki. "Appliance-Level Anomaly Detection in Nonintrusive Load Monitoring via Power Consumption-Based Feature Analysis". In: *IEEE Transactions on Consumer Electronics* 67.4 (Nov. 2021), pp. 363–371. ISSN: 1558-4127. DOI: 10.1109/TCE.2021.3129356.
4. R. Bitterer and Prof. Dr. habil. B. Schieferdecker. *Repräsentative VDEW-Lastprofile Aktionsplan Wettbewerb, M-32 99*. VDEW. 1999.
5. A. Capasso et al. "A bottom-up approach to residential load modeling". In: *IEEE Transactions on Power Systems* 9.2 (May 1994), pp. 957–964. ISSN: 1558-0679. DOI: 10.1109/59.317650.
6. Cristina Carpino et al. "Energy performance gap of a nearly Zero Energy Building in Denmark the influence of occupancy modelling". In: *Building Research Information* 48 (Jan. 2020).
7. Marco Castangia et al. "Detection of Anomalies in Household Appliances from Disaggregated Load Consumption". In: *2021 International Conference on Smart Energy Systems and Technologies (SEST)*. 2021, pp. 1–6. DOI: 10.1109/SEST50973.2021.9543232.
8. Luo Chuan, D. M. K. K. Venkateswara Rao, and Abhisek Ukil. "Load profiling of Singapore buildings for peak shaving". In: *2014 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)*. 2014, pp. 1–6. DOI: 10.1109/APPEEC.2014.7065998.
9. Tamás Csoknyai et al. "Analysis of energy consumption profiles in residential buildings and impact assessment of a serious game on occupants' behavior". In: *Energy and Buildings* 196 (2019), pp. 1–20. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2019.05.009. URL: https://www.sciencedirect.com/science/article/pii/S0378778818334790.
10. David Fischer, Andreas Härtl, and Bernhard Wille-Haussmann. "Model for electric load profiles with high time resolution for German households". In: *Energy and Buildings* 92 (2015), pp. 170–179. ISSN: 0378-7788. DOI: https://doi.org/10.1016/j.enbuild.2015.01.058. URL: https://www.sciencedirect.com/science/article/pii/S0378778815000845.

11. Bingtuan Gao, Xiaofeng Liu, and Zhenyu Zhu. "A Bottom-Up Model for Household Load Profile Based on the Consumption Behavior of Residents". In: *Energies* 11.8 (2018). ISSN: 1996-1073. DOI: 10.3390/en11082112. URL: https://www.mdpi.com/1996-1073/11/8/2112.
12. D. Gerbec et al. "Allocation of the load profiles to consumers using probabilistic neural networks". In: *IEEE Transactions on Power Systems* 20.2 (May 2005), pp. 548–555. ISSN: 1558-0679. DOI: 10.1109/TPWRS.2005.846236.
13. Geoffrey Hinton and Sam Roweis. "Stochastic Neighbor Embedding". In: *Advances in neural information processing systems* 15 (2003). Ed. by S Thrun S Becker and KEditors Obermayer, 833 to 840.
14. Fatih Issi and Orhan Kaplan. "The Determination of Load Profiles and Power Consumptions of Home Appliances". In: *Energies* 11.3 (2018). ISSN: 1996-1073. DOI: 10.3390/en11030607. URL: https://www.mdpi.com/1996-1073/11/3/607.
15. Hyun Cheol Jeong et al. "Clustering of Load Profiles of Residential Customers Using Extreme Points and Demographic Characteristics". In: *Electronics* 10.3 (2021). ISSN: 2079-9292. DOI: 10.3390/electronics10030290. URL: https://www.mdpi.com/2079-9292/10/3/290.
16. Emily Kattan, Suleiman Halasah, and Tareq a hamed. "Practical Challenges of Photovoltaic Systems in the Rural Bedouin Villages in the Negev". In: *Journal of Fundamentals of Renewable Energy and Applications* 8 (May 2018). DOI: 10.4172/20904541.1000258.
17. Amir Kavousian, Ram Rajagopal, and Martin Fischer. "Determinants of residential electricity consumption: Using smart meter data to examine the effect of climate, building characteristics, appliance stock, and occupants' behavior". In: *Energy* 55 (2013), pp. 184–194. ISSN: 0360-5442. DOI: https://doi.org/10.1016/j.energy.2013.03.086. URL: https://www.sciencedirect.com/science/article/pii/S0360544213002831.
18. Jack Kelly and William Knottenbelt. "The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes". In: *Scientific Data* 2.1 (Mar. 2015), p. 150007. ISSN: 2052-4463. DOI: 10.1038/sdata.2015.7. URL: https://doi.org/10.1038/sdata.2015.7.
19. Elke Klaassen, Jasper Frunt, and Han Slootweg. "Assessing the impact of distributed energy resources on LV grids using practical measurements". In: June 2015.
20. Wilhelm Kleiminger et al. "Occupancy Detection from Electricity Consumption Data". In: Nov. 2013, pp. 1–8. DOI: 10.1145/2528282.2528295.
21. L.J.P. van der Maaten and G.E. Hinton. "Visualizing High-Dimensional Data Using tSNE". In: (2008).
22. Keon-Jun Park and Sung-Yong Son. "A Novel Load Image Profile-Based Electricity Load Clustering Methodology". In: *IEEE Access* 7 (2019), pp. 59048–59058. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2914216.
23. Haroon Rashid et al. "Can non-intrusive load monitoring be used for identifying an appliances anomalous behaviour?" In: *Applied Energy* 238 (2019), pp. 796–805. ISSN: 0306-2619. DOI: https://doi.org/10.1016/j.apenergy.2019.01.061. URL: https://www.sciencedirect.com/science/article/pii/S0306261919300613.

24. Haroon Rashid et al. "Evaluation of Non-intrusive Load Monitoring Algorithms for Appliance-level Anomaly Detection". In: *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. May 2019, pp. 8325–8329. DOI: 10.1109/ICASSP.2019.8683792.

25. Mercy shuma-iwisi. "Domestic appliances standby power losses: The case of eleven suburbs in the greater Johannesburg," in: June 2009.

26. C.F. Walker and J.L. Pokoski. "Residential Load Shape Modelling Based on Customer Behavior". In: *IEEE Transactions on Power Apparatus and Systems* PAS-104.7 (1985), pp. 1703–1711. DOI: 10.1109/TPAS.1985.319202.