

Master's Degree in Environmental Pathways for Sustainable  
Energy Systems

---

## PROJECT REPORT District Heating Substation Efficiency:

---

AUTHORS Ankit Bid  
Amrou Ghattas  
Damien Pantalos  
Julian Martinez  
Santhosh Subramani

COURSE  
*Data driven challenges for energy engineers*

2025-01-21

# Contents

|          |                                                                                    |           |
|----------|------------------------------------------------------------------------------------|-----------|
| <b>1</b> | <b>Introduction</b>                                                                | <b>3</b>  |
| 1.1      | Objectives . . . . .                                                               | 4         |
| <b>2</b> | <b>Opportunity Analysis</b>                                                        | <b>5</b>  |
| 2.1      | Opportunity Analysis . . . . .                                                     | 5         |
| 2.2      | Market Potential . . . . .                                                         | 5         |
| 2.3      | Target Audience . . . . .                                                          | 6         |
| 2.4      | Supporting UK Policies . . . . .                                                   | 6         |
| <b>3</b> | <b>Business idea and strategy</b>                                                  | <b>7</b>  |
| 3.1      | Target Market . . . . .                                                            | 8         |
| 3.2      | Value Proposition . . . . .                                                        | 8         |
| 3.3      | Channels . . . . .                                                                 | 8         |
| 3.4      | Revenue Streams . . . . .                                                          | 8         |
| 3.5      | Key Partners . . . . .                                                             | 9         |
| 3.6      | Cost Structures . . . . .                                                          | 9         |
| 3.7      | Final remarks . . . . .                                                            | 9         |
| <b>4</b> | <b>Methodology</b>                                                                 | <b>10</b> |
| 4.1      | Main Dataset . . . . .                                                             | 10        |
| 4.1.1    | Data Acquisition and Preprocessing: Electricity demand data . . . . .              | 10        |
| 4.1.2    | Data Acquisition and Preprocessing: Weather Dataset Cleaning . . . . .             | 12        |
| 4.2      | AI as a load disaggregation method . . . . .                                       | 13        |
| 4.2.1    | NILMTK: Non-Intrusive Load Monitoring Toolkit [9] . . . . .                        | 13        |
| 4.2.2    | Challenges Faced During Implementation . . . . .                                   | 15        |
| 4.2.3    | Energy Disaggregation Using Deep Learning (Energy-Disaggregation-DL) [8] . . . . . | 16        |
| 4.2.4    | Comparison of NILMTK and Energy-Disaggregation-DL Methodologies . . . . .          | 18        |
| 4.3      | Alternative methodology: Manual disaggregation . . . . .                           | 19        |
| <b>5</b> | <b>Results</b>                                                                     | <b>23</b> |
| <b>6</b> | <b>Challenges and limitations</b>                                                  | <b>24</b> |
| 6.1      | Technical Challenges . . . . .                                                     | 24        |
| 6.2      | Methodological Challenges . . . . .                                                | 24        |
| 6.3      | Business Limitations . . . . .                                                     | 25        |
| 6.4      | Resource Constraints . . . . .                                                     | 25        |
| <b>7</b> | <b>Implementation and Usage</b>                                                    | <b>26</b> |
| 7.1      | Implementation Plan . . . . .                                                      | 26        |
| 7.1.1    | Integration with existing tools (e.g., Power BI dashboards) . . . . .              | 27        |

|                                      |           |
|--------------------------------------|-----------|
| 7.2 Stakeholder Engagement . . . . . | 28        |
| 7.3 Scalability . . . . .            | 29        |
| <b>8 Conclusion</b>                  | <b>30</b> |
| <b>Bibliography</b>                  | <b>30</b> |

# Chapter 1

## Introduction

Heating and cooling account for over 50% of global final energy consumption and more than 40% of worldwide CO<sub>2</sub> emissions. As electricity grids become increasingly decarbonized, electrifying heating presents a promising alternative to the current fossil fuel-dominated systems. However, challenges arise due to the high penetration of renewable energy, the pronounced seasonality and hourly fluctuations in heat demand, and the growing domestic need for energy services like cooling. These factors complicate both real-time system operations and long-term investment planning for electricity infrastructure. To achieve low-carbon heating and cooling while integrating a larger share of renewable energy and managing peak loads, a combination of demand-response strategies and flexible systems will be essential [1].

Heat pumps, which use electricity to provide both heating and cooling, are a critical technology in the shift towards sustainable energy. By electrifying climate control, households can significantly cut their carbon emissions, particularly when operating within grids powered by renewable sources like wind and solar. In addition to their environmental advantages, switching to electric systems offers cost savings and improved energy efficiency, benefiting both consumers and the planet.

Furthermore, as climate change intensifies weather variability, properly sizing heating and cooling systems becomes increasingly critical. Using historical energy consumption data, the most suitable heat pump models can be recommended, ensuring they are tailored to meet each household's unique heating and cooling demands. This data-driven strategy not only improves system efficiency but also accelerates the adoption of these eco-friendly solutions in homes yet to transition to electrified climatization.

### Electric Heat Pumps: A Promising Alternative

Electric heat pumps are becoming a prominent technology in the shift toward electrified heating and cooling systems. Compared to traditional fossil fuel-based systems, heat pumps offer several notable advantages:

- **Energy Efficiency:** Heat pumps are highly efficient, measured by their coefficient of performance (COP), which represents the ratio of heat output to electricity consumed [2].
- **Reduced Greenhouse Gas Emissions:** By utilizing electricity rather than fossil fuels, heat pumps can significantly reduce carbon emissions. Their environmental benefits are even greater when paired with renewable energy sources like solar or wind power [2].
- **Improved Air Quality:** Unlike combustion-based systems, heat pumps do not release harmful by-products

such as carbon monoxide, nitrogen oxides, or particulate matter, contributing to healthier indoor and outdoor air quality [2].

## Problem Statement

Current heating and cooling systems face several challenges. Fossil fuels like natural gas, oil, and coal remain the main energy sources for heating and cooling in many countries. These systems contribute significantly to greenhouse gas emissions and negatively affect air quality and public health [2]. Also, many traditional systems are highly energy-intensive, leading to increased operational costs and reduced performance.

While electric heat pumps offer many advantages, several challenges and obstacles continue to limit their widespread adoption:

- **High Initial Costs:** Installing heat pumps can be more expensive upfront compared to traditional systems. However, long-term energy savings and reduced maintenance expenses often compensate for these costs [2].
- **Low Awareness:** Many consumers and professionals lack knowledge about heat pump technology and its advantages, resulting in limited demand and slow uptake [2].
- **Challenges in Extreme Climates:** Heat pumps generally work well in most regions, but their efficiency can drop in freezing temperatures. Advancements, like cold-climate heat pumps, are improving their performance in such conditions [2].
- **Grid and Infrastructure Limitations:** Increased adoption may strain electrical grids, requiring significant upgrades to support higher demand, especially during peak usage periods [2].

## 1.1 Objectives

The primary objective of this project is to optimize heat pump recommendations by utilizing load disaggregation techniques on historical energy consumption data, coupled with regional weather information. This analysis aims to accurately identify heating and cooling loads, ensuring precise sizing of heat pumps for both electrified and non-electrified homes. The project enhances energy efficiency and system performance by doing recommendations to specific household needs.

Another key goal is to promote the electrification of heating and cooling systems by identifying homes that have not yet transitioned from fossil-fuel-based systems. For these households, the project will recommend suitable heat pump models, encouraging the adoption of sustainable, low-carbon energy solutions.

## Chapter 2

# Opportunity Analysis

### 2.1 Opportunity Analysis

The market for heat pumps, particularly correctly sized ones, is growing rapidly as part of the UK's efforts to decarbonize heating systems. The UK government aims to install 600,000 heat pumps annually by 2028, driven by substantial financial incentives and supportive policies [3]. The Boiler Upgrade Scheme (BUS) has increased heat pump subsidies, now offering up to £7,500 per unit in England and Wales [2]. Scotland provides additional incentives for rural households and interest-free loans [1]. Cumulatively, these efforts align with the government's roadmap, which anticipates £28 billion in investment across the heat pump supply chain by 2028. This includes manufacturing incentives and targeted support for installation infrastructure.

### 2.2 Market Potential

The United Kingdom Heat Pumps Market is anticipated to hold a significant share by 2033, growing at a CAGR of 11.05% from 2023 to 2033. The growing recovering office space demand & HVAC retrofits and government regulations & initiatives are driving the growth of the heat pumps market in the UK. The market potential is significant due to:

- **Electrified Homes:** Rising electrification in urban settings demands efficient heat pump solutions to offset traditional heating.
- **Non-Electrified Homes:** Remote and rural homes reliant on oil or LPG boilers are transitioning to low-carbon alternatives as regulations phase out these systems by 2035.

In 2023, the total addressable market (TAM) for heat pump installations in the UK is estimated at £280 billion, given the 28 million households, each potentially needing a £10,000 unit including installation. However, the serviceable available market (SAM), which represents the 3 million urban households with an identified electrification need, is valued at £3 billion. This reflects a more targeted segment of the population likely to adopt heat pumps for energy efficiency and decarbonization efforts. Within this, the serviceable obtainable market (SOM) is projected at £60 million, assuming the installation of 6,000 heat pumps annually by 2028, capturing 1% of the SAM. This represents a modest yet growing market share within the electrification transition, highlighting opportunities for businesses targeting urban households focused on sustainable energy solutions. Although the potential for growth

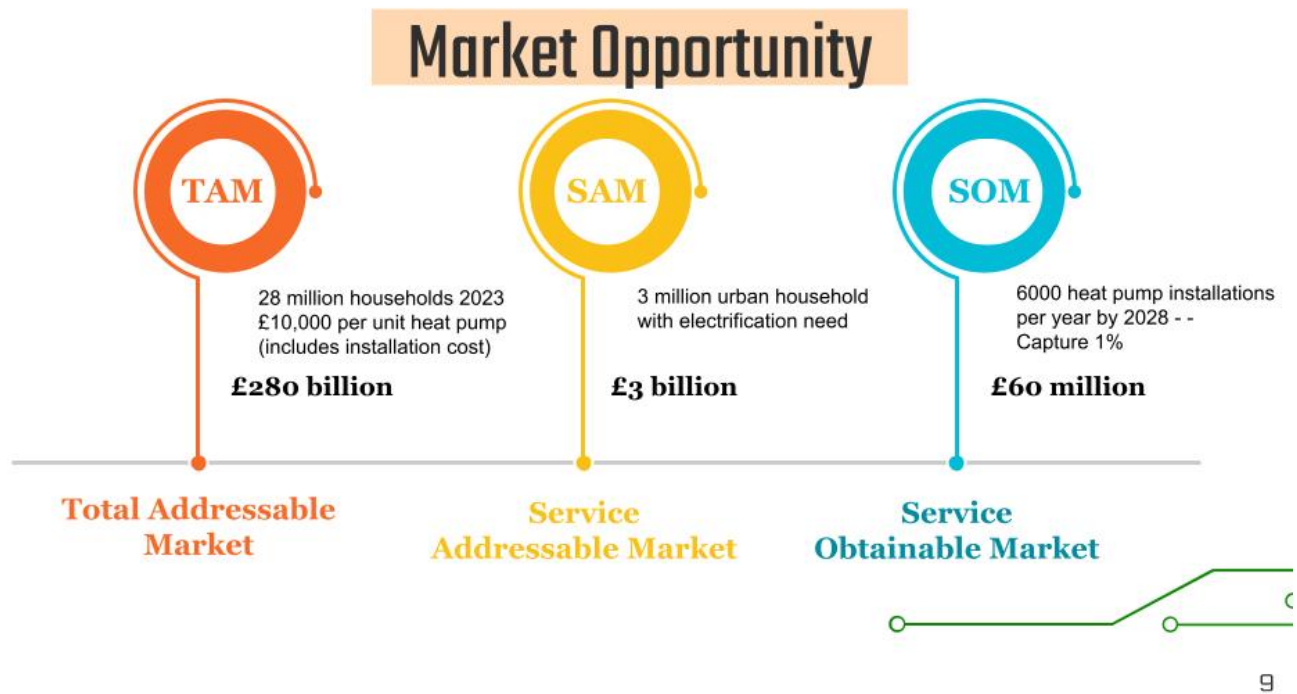


Figure 2.1: Market Anlysis for heat pumps in UK

is significant, achieving these targets will depend on factors such as government incentives, consumer awareness, and the affordability of heat pump systems relative to conventional heating methods.

## 2.3 Target Audience

Our solution is designed specifically for **high demand homes with non-electrified heating or cooling systems**, such as those relying on oil, propane, or natural gas. By offering a transition to modern, efficient heat pumps, we empower homeowners to achieve significant energy savings while contributing meaningfully to sustainability goals.

## 2.4 Supporting UK Policies

- **Boiler Upgrade Scheme (BUS):** Increased subsidies to promote adoption, ensuring up to £7,500 per unit for installations[2].
- **Future Homes and Buildings Standard:** Mandates that new homes must adopt low-carbon heating solutions by 2025.[4]

# Chapter 3

## Business idea and strategy

This business aims to provide heat pump installation and consulting services for high-demand non-electrified households, focusing on urban homes in the UK. The service is designed to meet the growing need for electrification in households and align with the UK’s energy efficiency and sustainability goals. By offering cost-effective, energy-saving solutions, the business seeks to create value for customers, reduce their utility bills, and drive long-term savings in both energy and time.

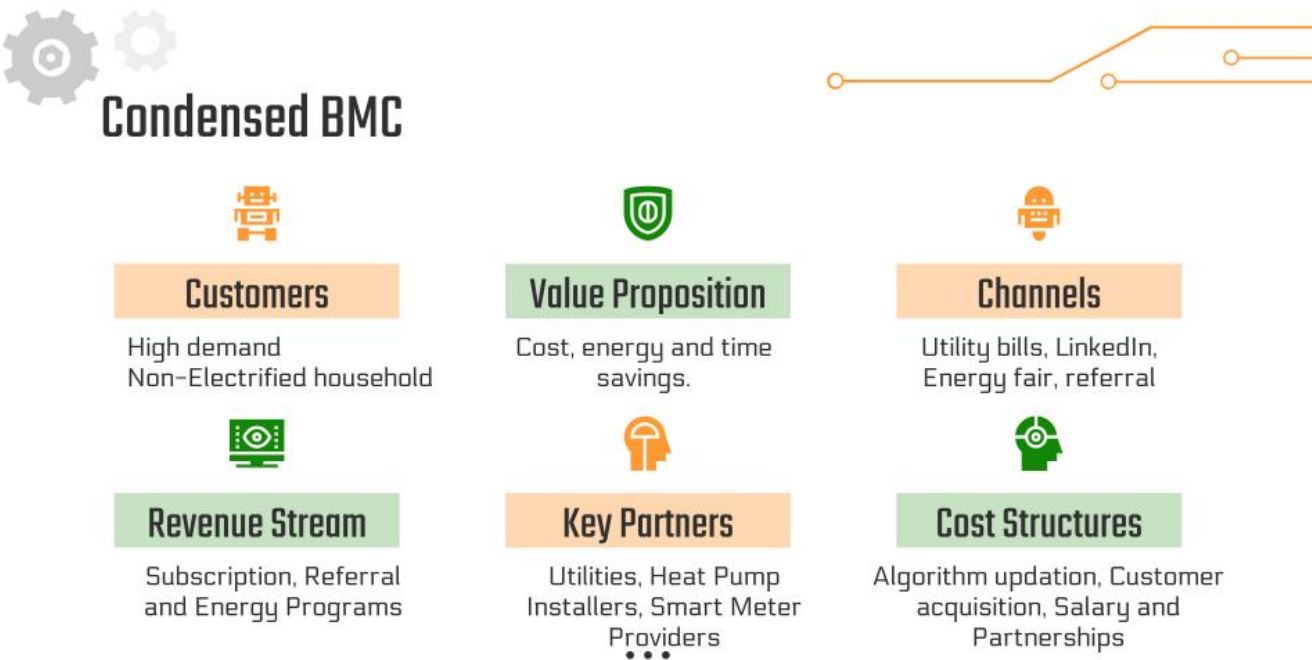


Figure 3.1: Condensed Business Model Canvas

### 3.1 Target Market

The business will focus on the 3 million urban households in the UK that need electrification. These homes are prime candidates for heat pump installations, as the push to decarbonize heating in the UK intensifies. With a serviceable available market (SAM) valued at £3 billion and a serviceable obtainable market (SOM) projected at £60 million by 2028 (based on capturing 1% of the SAM), this presents a significant opportunity.

### 3.2 Value Proposition

This business offers a simple yet powerful value proposition: it helps homeowners save money, time, and energy by installing heat pumps. Heat pumps can lower monthly utility bills through better energy efficiency, and they also provide a cleaner, more sustainable alternative to traditional heating systems. Beyond savings, we handle the whole process—from consultation to installation and ongoing maintenance—giving customers peace of mind and saving them time.

### 3.3 Channels

To connect with potential customers, the business will use several channels:

- **Utility Bills:** Collaborating with utility companies to reach customers via their energy bills, showing them how heat pumps can reduce costs.
- **LinkedIn:** Using LinkedIn to connect with property developers, utilities, and homeowners interested in energy-efficient solutions.
- **Energy Fairs:** Attending local and national energy events to showcase heat pumps, meet potential clients, and build awareness.
- **Referral Programs:** Building a referral program where existing customers can recommend the service to others, earning incentives while helping to grow the customer base.

### 3.4 Revenue Streams

To keep things financially sustainable, the business will generate revenue in multiple ways:

- **Subscription Model:** By offering ongoing maintenance, monitoring, and support for heat pump systems, we can secure regular payments from customers who need our help to keep everything running smoothly.
- **Referral Fees:** When we partner with heat pump installers and other energy services, we'll earn a commission for each successful referral.
- **Energy Programs:** Through government-backed programs that incentivize energy efficiency, we can generate revenue from heat pump installations funded by these programs.

### 3.5 Key Partners

The success of this business depends on strong partnerships with other key players in the energy industry:

- **Utilities:** Working with utility companies will allow us to tap into their customer base and support ongoing energy-efficiency programs.
- **Heat Pump Installers:** Collaborating with experienced installers ensures high-quality installations and expands our reach.
- **Smart Meter Providers:** Partnering with smart meter companies will let customers track their energy savings, making heat pump adoption more attractive.

### 3.6 Cost Structures

Running the business comes with several important costs, including:

- **Algorithm Updates:** Continuously improving algorithms to optimize system design and ensure efficiency.
- **Customer Acquisition:** Investing in marketing strategies such as LinkedIn ads, energy fairs, and utilities promotions to get new customers on board.
- **Salaries:** Paying employees who handle consultations, installations, and customer support.
- **Partnerships:** The costs involved in maintaining relationships with utilities, installers, and smart meter companies to ensure we deliver on our value proposition.

### 3.7 Final remarks

This business taps into a rapidly growing need for energy-efficient solutions in the UK's housing market, especially among high-demand, non-electrified homes. The market is large, with a potential £3 billion SAM, and with strategic partnerships and the right outreach channels, this business can carve out a significant share of that market. By providing value through cost savings, sustainability, and a streamlined customer experience, we can help homeowners make the switch to cleaner energy while benefiting from recurring revenue through maintenance subscriptions and energy programs. This plan not only addresses a real market gap but also positions the business as an important player in the UK's green energy transition.

# Chapter 4

## Methodology

### 4.1 Main Dataset

The primary dataset utilized in this work is the \*Electricity Demand\* dataset, provided by the EDS Lab and hosted on the Hugging Face platform. This dataset comprises three distinct files:

- **Electricity Demand Data:** (`data/demand.parquet`) This file contains electricity demand records for various users, sampled at 30-minute intervals over a three-year period (2012–2014).
- **Weather Data:** (`data/weather.parquet`) This file provides meteorological measurements for multiple cities during the same time period, recorded hourly. The weather attributes include wind speed, temperature, apparent temperature, rainfall, snowfall, and other relevant variables.
- **Metadata:** (`data/metadata.csv`) This file connects the demand and weather datasets through unique user identifiers (`unique_id`) and location (`city`). Additionally, it provides contextual information, such as whether the users are residential or commercial and their geographic distribution across different areas of the United States, Canada and the United Kingdom.

Together, these datasets allow us to do a comprehensive analysis of electricity consumption patterns and their correlation with weather conditions, providing the bases for our methodology.

#### 4.1.1 Data Acquisition and Preprocessing: Electricity demand data

The first step in our analysis involved connecting to the dataset’s API and exporting the data into a `.parquet` file. This was necessary due to the large size of the dataset, particularly the electricity consumption file, which contains approximately 237,944,171 rows.

#### Data Matching and Cleaning

To ensure data integrity, it was verified the consistency between the `unique_id` fields in the electricity demand data and the metadata file. We identified a mismatch, with 7,514 unique IDs in the demand dataset and 7,572 in the metadata. To address this, it was removed entries that did not match, ensuring alignment between the two datasets.

## Data Visualization and Selection

Next, it was visualized the geographic distribution of the data, summarized in Table 4.1. This analysis revealed that the majority of the data originates from London, UK. London not only provides the largest subset of data but also offers diverse seasonal conditions, including both heating and cooling requirements. Based on this observation, it was decided to filter the dataset to include only entries from London.

**Table 4.1:** Geographic Distribution of Data

| City, Country               | Number of Records |
|-----------------------------|-------------------|
| London, UK                  | 5,634             |
| PT                          | 370               |
| Washington DC, USA          | 305               |
| Minneapolis, Minnesota, USA | 163               |
| Cardiff, UK                 | 147               |
| Phoenix, Arizona, USA       | 137               |
| Orlando, Florida, USA       | 136               |
| Austin, Texas, USA          | 124               |
| Ithaca, New York, USA       | 124               |
| Oakland, California, USA    | 92                |
| Princeton, New Jersey, USA  | 47                |
| Groningen, NL               | 36                |
| Ottawa, CA                  | 20                |

To further tune the dataset for the analysis, it was applied an additional filter to focus exclusively on `building_class = residential`. By combining these filters, it ensured that the analysis targets high-quality data specifically for residential properties in London.

## Data cleaning

This stage of the methodology, combining data visualization and cleaning, aims to identify and select the best-quality data subset to perform an in-depth analysis of residential electricity consumption patterns in London. By narrowing our focus, it ensured that the dataset is both representative of seasonal variations and suitable for our objective of heat pump sizing.

The first step in our preprocessing involved analyzing the extent and distribution of missing values in the dataset. In total, it was identified approximately 614,425 missing entries, which represent only 2–3% of the total data. While this may seem significant in absolute terms, the dataset’s overall quality remains high.

To further understand the impact of missing values, it was analyzed the percentage of missing entries per `unique_id`. For visualization purposes, it was focused on the worst 1% of `unique_ids`, revealing that 23 out of 5,634 `unique_ids` had more than 25% missing data. This confirmed that the majority of the dataset maintains a high level of completeness, even though a small fraction of `unique_ids` exhibit significant data gaps.

After this it was conducted an iterative filtering process to identify an appropriate threshold for retaining high-quality data. By removing `unique_ids` with more than 20% missing data, it retained 3,653 entries. Tightening this threshold to 5% reduced the dataset to 2,534 `unique_ids`, and applying a 1% threshold yielded the same result. Consequently, it was decided to retain only the `unique_ids` with at least 99% of their data intact, ensuring the dataset’s overall quality and reliability.

The next step was identifying potential outliers in the dataset. To accomplish this, it was employed box plots and histograms, tools widely used in data analysis for understanding the distribution and detecting anomalies. This step aimed to verify whether any extreme values come from errors in the meters (e.g., faulty readings) or were valid data points caused by cumulative energy reporting.

For instance, it is possible that a meter might fail to register data for three consecutive hours and subsequently record a disproportionately high consumption in a single half-hour interval to account for the missing period. By visualizing the data, it was carefully assessed whether such anomalies were plausible. After the analysis, it was found that no extreme outliers has appeared unreasonably large or inconsistent with typical energy consumption patterns. Therefore, no additional cleaning of the data was required based on outlier detection.

Following the cleaning process, the next step involved verifying whether all households in the dataset started and ended their data recording on the same dates. It was observed that the recording periods varied. To address this, the two years with the most complete data were identified: 2012-02-01 to 2013-01-31 and 2013-02-01 to 2014-01-31. Consequently, the dataset was split into two separate dataframes corresponding to these periods.

For each dataframe, the focus shifted to analyzing zero values in the data. Random sampling revealed that households generally do not report zero electricity consumption, as a house typically demands at least some electricity. Possible reasons for zero values could include households turning off power, disconnection due to unpaid bills, or meter errors. To ensure high-quality data, households with more than 8% zero values (equivalent to approximately 2 hours of missing data per day in 30-minute resolution) were excluded. For the 2012-2013 dataframe, this resulted in the removal of 15 households, reducing the total from 448 to 433.

## **Data filling**

Before filling missing values, the longest consecutive periods of missing entries were analyzed to ensure the chosen imputation method would remain accurate. The analysis revealed a maximum gap of six consecutive missing entries (equivalent to three hours), making linear interpolation a suitable method for filling missing values. Linear interpolation was selected for its ability to approximate missing values effectively without introducing significant bias.

## **Data transformation**

To reduce dataset complexity, processing time, and address file size limitations on GitHub, the electricity consumption data for both periods (2012-2013 and 2013-2014) was aggregated into hourly resolution. The average of the first and second 30-minute intervals was computed for each hour. The resulting datasets were exported as `load_2012_2013_hourly.parquet` and `load_2013_2014_hourly.parquet` for further analysis.

### **4.1.2 Data Acquisition and Preprocessing: Weather Dataset Cleaning**

For the weather dataset, only data corresponding to London was considered. An analysis for outliers and missing entries was conducted, and it was found that all values were consistent and complete. Therefore, no further cleaning or transformation was required.

For the purpose of load disaggregation and to remove unnecessary information, a subset of features from the weather dataset was selected. The retained columns included `timestamp`, `temperature-2m`, `relative-humidity-2m`,

apparent-temperature, location-id, precipitation, wind-speed-10m, and is-day. This reduced complexity and ensured that only the most relevant features were used in subsequent analyses.

## 4.2 AI as a load disaggregation method

Energy disaggregation, is a process of analyzing aggregate energy consumption data from a single meter to estimate the power usage of individual appliances within a building. This approach provides valuable insights into appliance-level energy consumption without the need for individual meters on each device, making it a cost-effective solution for monitoring and optimizing energy use. [5]

The importance of energy disaggregation lies in its ability to promote energy efficiency. By identifying the consumption patterns of individual appliances, homeowners and businesses can make informed decisions to reduce energy waste, improve operational efficiency, and transition toward sustainable energy practices. Additionally, it supports the development of smart grids and energy management systems by offering granular data for demand-side management. Relevant to this project, it is helpful in order to identify the demand on the HVAC systems of the houses [6].

Traditional methods for energy disaggregation, such as rule-based approaches and statistical models, rely on pre-defined heuristics or probabilistic frameworks like Hidden Markov Models (HMMs). While these methods are computationally lightweight and interpretable, they are often limited in accuracy, especially when appliances have overlapping power signatures or operate in multiple states [7].

Artificial Intelligence (AI), particularly Deep Learning (DL), has emerged as a superior alternative for energy disaggregation due to its ability to learn complex, non-linear relationships between aggregate power signals and individual appliance loads. Deep learning models, such as Sequence-to-Point (Seq2Point) and Recurrent Neural Networks (RNNs), perform good at capturing temporal dependencies in energy consumption data, enabling higher disaggregation accuracy even in challenging scenarios. In the Table 4.2, it is presented a brief comparison between the two methodologies presented.

Lastly, AI-driven methods also offer scalability, making them ideal for processing large datasets with diverse appliance types. Furthermore, these models can generalize across different households and conditions when trained on representative data, a significant advantage over traditional methods.

Now, for the project, it was decided to pursue AI methodologies. Therefore, next will be presented two different models that during the course of the project it was intended to use.

### 4.2.1 NILMTK: Non-Intrusive Load Monitoring Toolkit [9]

#### What is NILMTK?

Non-Intrusive Load Monitoring Toolkit (NILMTK) is an open-source Python library designed for the analysis, and development of energy disaggregation algorithms. It simplifies the process of working with NILM by providing tools for data preprocessing, algorithm implementation, and evaluation using standardized datasets. The toolkit supports a wide range of datasets and disaggregation algorithms, making it a resource for researchers and practitioners in the energy domain.

**Table 4.2:** Comparison of Traditional and AI-Based Methods for Energy Disaggregation

| Aspect                   | Traditional Methods                                                                                                                     | AI-Based Methods (Deep Learning)                                                                                                         |
|--------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Brief Explanation</b> | Predefined rules (rule-based) or statistical models (e.g., Hidden Markov Models) identify appliance usage patterns.                     | Neural networks learn complex, non-linear relationships between aggregate power data and appliance consumption.                          |
| <b>Advantages</b>        | Simple and interpretable, low computational cost, effective for appliances with distinct signatures.                                    | High accuracy for noisy and overlapping signals, handles multi-state appliances, scalable with large datasets, learns subtle patterns.   |
| <b>Disadvantages</b>     | Limited accuracy for overlapping signals, struggles with variable or multi-state appliances, requires manual tuning, lacks flexibility. | Computationally intensive (requires GPUs), needs large labeled datasets, models are less interpretable (black-box), long training times. |
| <b>Data Requirement</b>  | Minimal labeled data (rule-based) or moderate data (statistical models).                                                                | Requires large amounts of high-resolution labeled data.                                                                                  |
| <b>Model Examples</b>    | Rule-based heuristics, Hidden Markov Models (HMM), Factorial HMM, Gaussian Mixture Models (GMM).                                        | Sequence-to-Point, Sequence-to-Sequence (Seq2Seq), CNNs, RNNs, Attention-based Models.                                                   |
| <b>Best Use Case</b>     | Simple scenarios with fewer appliances or distinct power signatures.                                                                    | Complex scenarios with high-resolution data, overlapping appliances, and multi-state behavior.                                           |

## How NILMTK Works

NILMTK provides a structured framework to process aggregate energy data and perform load disaggregation. The workflow involves the following steps:

1. **Data Loading and Preprocessing:** Aggregated energy data from NILM datasets, such as UK-DALE or REDD, are loaded into NILMTK. This data often includes power measurements over time (e.g., active power, apparent power) and metadata about appliances and buildings.
2. **Training Disaggregation Models:** NILMTK enables the use of predefined disaggregation algorithms such as Combinatorial Optimization (CO), Factorial Hidden Markov Models (FHMM), and deep learning models like Sequence-to-Point (Seq2Point).
3. **Evaluation:** After training, the models are evaluated on test data using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and F1-score. NILMTK provides built-in functions to compute these metrics and generate visualizations of the results.
4. **Visualization and Analysis:** Results, including appliance-level energy consumption, are visualized through NILMTK's tools, providing insights into energy usage patterns.

## Data Requirements

NILMTK requires datasets that provide the following:

- **Aggregate Power Data:** Time-series data of total power consumption (e.g., active power in watts, apparent power in VA).
- **Appliance-Level Data:** Individual power consumption readings for appliances (used for training and evaluation).
- **Metadata:** Information about appliances (e.g., type, location, and state transitions) and the building (e.g., number of rooms, total area).

## Dataset Format

Datasets need to be converted into a format compatible with NILMTK. The typical file format is **HDF5**, a hierarchical data format optimized for large datasets. NILMTK provides tools to convert popular datasets like UK-DALE and REDD into the required format. Each dataset contains:

- **Aggregate Energy Data:** Stored in a hierarchical structure, indexed by building, meter, and time.
- **Metadata Files:** YAML files containing descriptions of appliances, buildings, and power meters.

## How Algorithms Work in NILMTK

NILMTK supports various algorithms for load disaggregation:

- **Combinatorial Optimization (CO):** This algorithm uses predefined appliance signatures to find the best combination of appliances that matches the aggregate power signal.
- **Factorial Hidden Markov Model (FHMM):** FHMM models the states of multiple appliances and their power usage as hidden variables. It predicts appliance states from the aggregate signal.
- **Deep Learning Models:** Algorithms like Seq2Point and Sequence-to-Sequence use neural networks to predict individual appliance loads from a time window of aggregate data. These models learn complex patterns in the data, improving accuracy over traditional methods.

In summary, NILMTK is a powerful toolkit for advancing energy disaggregation research. By providing standardized tools for data handling, model training, and evaluation, it enables researchers to develop and compare algorithms efficiently. However, its effectiveness depends on the availability of high-quality datasets, appropriate metadata, and careful handling of data preprocessing and conversion steps.

### 4.2.2 Challenges Faced During Implementation

While NILMTK offers a robust framework for energy disaggregation, we encountered several challenges during its implementation:

1. **Learning Curve:** As newcomers to NILMTK, there was a significant learning curve in understanding its functionality, installing the toolkit, ensuring compatibility with dependencies, and grasping the required data structure (e.g., what an HDF5 file is). Running examples and configuring the system required continuous learning and troubleshooting.

2. **Metadata Requirements:** From our understanding, NILMTK mandates the use of detailed metadata in a specific YAML format. For our main dataset, this meant converting the data manually into a hierarchical structure. Given that we were dealing with thousands of homes, the task of organizing data and creating metadata for appliances without knowing the exact list of existing appliances in each house proved challenging. Although assumptions could be made, the absence of pre-existing YAML metadata increased complexity.
3. **Training Models:** We faced difficulties in training the models, as specific inputs required by NILMTK were not fully clear in practice. Despite attempting examples with three different datasets to understand the workflow, errors persisted during the training process. This limited our ability to explore NILMTK’s full potential for disaggregation.
4. **Metadata Creation:** Accurate and detailed metadata is essential for proper dataset conversion and model training.
5. **Dataset Conversion:** Converting raw datasets into NILMTK-compatible formats can be time-intensive and requires careful validation.
6. **High-Resolution Data:** Many NILM algorithms perform best with high-resolution data, which may not always be available or feasible to collect.
7. **Generalization:** Models trained on one dataset may not generalize well to others due to differences in appliance usage patterns and sampling rates.

Given these challenges, we recognized that the implementation pathway for NILMTK was complex and resource-intensive. As a result, we decided to explore an alternative AI-based methodology for energy disaggregation, which is discussed in the subsequent section of this report.

#### 4.2.3 Energy Disaggregation Using Deep Learning (Energy-Disaggregation-DL) [8]

Here, it is presented the *Energy Disaggregation DL* methodology, which uses deep learning for Non-Intrusive Load Monitoring (NILM) tasks. Specifically, this approach utilizes an attention-based encoder-decoder model to disaggregate aggregate power consumption data into individual appliance-level usage patterns. The methodology is implemented in the open-source GitHub repository <https://github.com/inesylla/energy-disaggregation-DL>.

##### Overview of the Methodology

The core of the methodology is a neural network architecture that incorporates an encoder-decoder structure enhanced with an attention mechanism. The model predicts appliance-level power consumption from a sequence of aggregated power readings.

##### Key Features:

- **Encoder-Decoder Architecture:** The encoder compresses the input sequence of aggregate power readings into a latent representation, while the decoder reconstructs appliance-specific power consumption from this representation.
- **Attention Mechanism:** The attention layer assigns varying levels of importance to different positions in the aggregated signal, improving the model’s ability to focus on relevant patterns associated with each appliance.

- **Dual Output:** The model supports both regression (predicting appliance power usage) and classification (determining appliance state: ON/OFF).
- **Data-Driven Approach:** The methodology relies heavily on labeled, high-resolution datasets for training and evaluation.

## Data Requirements and Format

The methodology requires the following:

- **Aggregate Power Data:** Time-series data of total power consumption.
- **Appliance-Specific Data:** Ground truth power consumption data for individual appliances, used for training and evaluation.
- **High Resolution:** Datasets with sampling rates of 1 second to 1 minute are preferred to capture detailed appliance behavior.
- **File Format:** The data should be structured and stored in formats like HDF5 or CSV for compatibility with deep learning frameworks.

## Implementation Process

1. **Data Preprocessing:** Normalization and synchronization of aggregate and appliance-level data.
2. **Model Training:** Train the encoder-decoder model using sequences of aggregate power data as input and corresponding appliance-level data as output.
3. **Evaluation:** Evaluate the model's performance using metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE).
4. **Prediction:** Apply the trained model to unseen aggregate power data to predict appliance-specific power usage.

## Advantages and Limitations

### Advantages:

- Superior accuracy compared to traditional methods due to the attention mechanism.
- Ability to handle overlapping appliance signals effectively.
- Supports both regression and classification tasks, making it versatile.

### Limitations:

- Requires extensive labeled data for training.
- Computationally intensive, necessitating GPU resources for training.
- Limited interpretability due to the black-box nature of deep learning models.

## Challenges faced during the implementation

Although the *Energy Disaggregation DL* methodology offers significant advantages in terms of accuracy and advanced capabilities, we encountered several challenges that led us to decide against its implementation in our project:

1. **Data Requirements and Granularity:** The methodology relies heavily on high-resolution data, with sampling intervals typically in the range of seconds to 2 minutes. However, our dataset only provides power consumption readings every 15 minutes. This lower granularity significantly reduces the accuracy of deep learning-based disaggregation models, as fine-grained patterns and transitions in appliance usage are not captured.
2. **Data Requirements:** The methodology requires labeled, high-resolution datasets that include both aggregate and appliance-level power consumption data. However, our primary dataset lacks labeled appliance-level data, which is critical for training the model effectively. Generating such labels for thousands of homes would be a resource-intensive and time-consuming process.
3. **Computational Complexity:** The deep learning models used in this methodology, such as attention-based encoder-decoders, are computationally intensive. Training these models requires substantial GPU resources, which were beyond the computational capacity available for this project.
4. **Time Constraints:** Implementing this methodology involves multiple steps, including data preprocessing, model training, and hyperparameter tuning. Given the limited time frame of the project, these steps posed a significant challenge to achieving meaningful results within the deadline.
5. **Black-Box Nature:** The deep learning approach offers limited interpretability, making it harder to explain the disaggregation process to stakeholders. This contrasts with more traditional methods, which are relatively easier to interpret and justify.

Given these challenges, we concluded that this methodology, while powerful, was not feasible for our specific project constraints. Instead, we chose to explore a more manual, structured approach, which is described in the following section of the report.

### 4.2.4 Comparison of NILMTK and Energy-Disaggregation-DL Methodologies

Both methodologies serve distinct purposes and have unique advantages. NILMTK provides a comprehensive framework for analyzing and benchmarking NILM algorithms, making it suitable for academic research and initial implementations. However, its dependency on metadata and traditional algorithms limits its performance for overlapping or noisy signals.

In contrast, the Energy-Disaggregation-DL methodology uses state-of-the-art deep learning techniques, providing superior accuracy and versatility. It excels in handling complex appliance interactions but requires substantial computational resources and labeled data. Given these factors, the choice of methodology depends on the project's goals, available resources, and the complexity of the data.

**Table 4.3:** Comparison of NILMTK and Energy-Disaggregation-DL Methodologies

| Aspect                        | NILMTK                                                               | Energy Disaggregation DL                                        |
|-------------------------------|----------------------------------------------------------------------|-----------------------------------------------------------------|
| <b>Model Type</b>             | Traditional methods (e.g., CO, FHMM) and basic deep learning models. | Attention-based encoder-decoder with advanced deep learning.    |
| <b>Ease of Implementation</b> | Requires metadata, dataset conversion, and manual adjustments.       | Requires labeled, high-resolution datasets and GPU resources.   |
| <b>Accuracy</b>               | Moderate for complex, overlapping appliance signals.                 | High accuracy due to attention mechanisms and advanced DL.      |
| <b>Data Requirements</b>      | Moderate (structured metadata, aggregate data).                      | High (labeled appliance data, aggregate data, high resolution). |
| <b>Output</b>                 | Appliance-level power estimates.                                     | Appliance power estimates and state classification (ON/OFF).    |
| <b>Scalability</b>            | Limited by metadata complexity and dataset compatibility.            | Scalable but resource-intensive.                                |
| <b>Interpretability</b>       | Higher due to traditional methods like CO and FHMM.                  | Lower due to the black-box nature of deep learning.             |

### 4.3 Alternative methodology: Manual disaggregation

While artificial intelligence approaches for energy disaggregation, such as NILMTK and deep learning-based methods, offer significant potential, they proved impractical for this project due to several limitations. Beyond the challenges discussed previously—such as the lack of appliance-level labeled data, insufficient data granularity, and computational complexity—these AI methodologies also fail to consider exogenous variables like weather. Incorporating external factors, such as temperature variations, can provide valuable insights into energy consumption patterns, particularly for heating and cooling loads.

Given these limitations, we developed an alternative manual methodology that integrates weather data with household energy consumption. This approach focuses on identifying correlations between temperature and load demand, allowing us to establish a baseline energy consumption for days with neutral or "comfort" temperatures. By analyzing deviations from this baseline during cold and hot days, we can estimate the contribution of heating and cooling to the overall energy demand. This methodology provides a transparent and data-driven way to disaggregate energy consumption without relying on labeled appliance-level data or complex AI models.

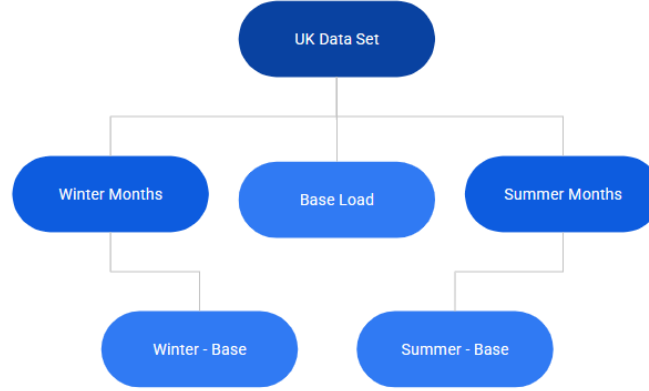
The manual disaggregation methodology was selected because it:

- Uses the available weather data we have, which covers the entire dataset consistently.
- Bypasses the need for high-resolution energy data or appliance-level labels, making it well-suited for datasets with 15-minute granularity.
- Offers interpretability, allowing for clear insights into how temperature variations impact energy consumption.

This methodology represents a practical and scalable alternative to AI-based disaggregation, particularly for datasets with limited resolution and missing metadata. In the following sections, we detail the steps taken to implement this approach and the results obtained.

The process began with the original dataset, which underwent cleaning and standardization as outlined in Section

4.1. This step eliminated errors and ensured consistency for further analysis. Following this, the dataset was segregated into three primary categories based on outside temperature: **cooling months**, **heating months**, and **base load months**. Base load months were defined as periods without significant heating or cooling needs. Figure 4.1 illustrates the data segregation process in a structured flowchart.



**Figure 4.1:** Data Process Flow Chart

The segregation of the dataset was guided by thresholds derived from external temperature data. Specifically:

- **Cooling months:** Periods when the outside temperature exceeded **27°C**, indicating a need for cooling systems.
- **Heating months:** Periods when the outside temperature fell below **18°C**, signaling heating requirements.
- **Base load months:** Periods when the outside temperature ranged between **18°C and 27°C**, implying negligible heating or cooling needs.

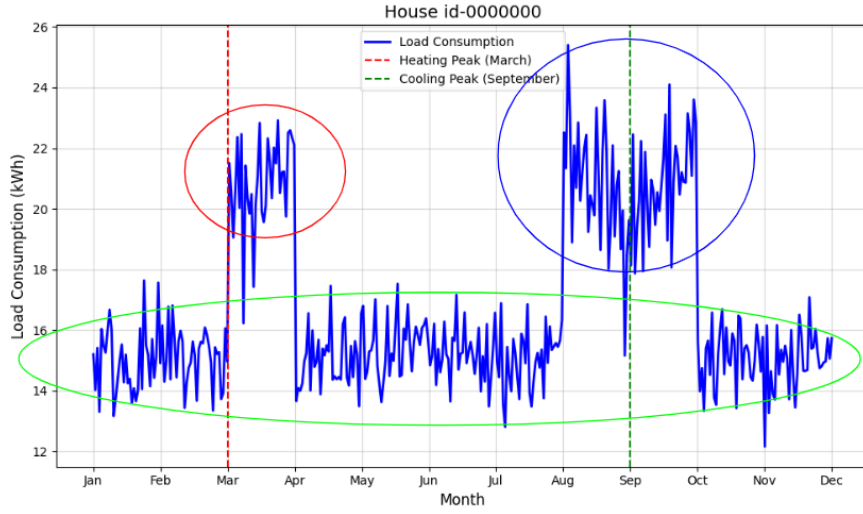
This classification aligns with findings from Source [10], which suggests a comfort indoor temperature in the UK of approximately **22°C**. Figure 4.2 demonstrates the load profile of a sample household across the year, highlighting the periods classified as cooling, heating, and base load months.

The dataset was further refined into two additional subsets:

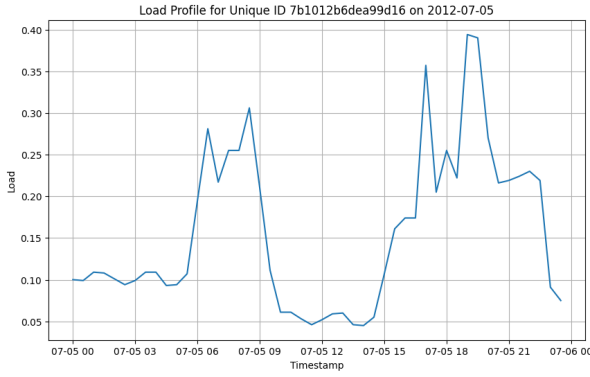
1. **Winter-Base Dataset:** Combines heating months with base load months.
2. **Summer-Base Dataset:** Merges cooling months with base load months.

Table 3.4 presents the threshold-based classification methodology used to distinguish between houses with electrified heating or cooling systems and those without. A **20% threshold** of the difference between the minimum and maximum load was applied to identify electrified systems. If a house's load exceeded this threshold, it was assumed to use electrified heating or cooling.

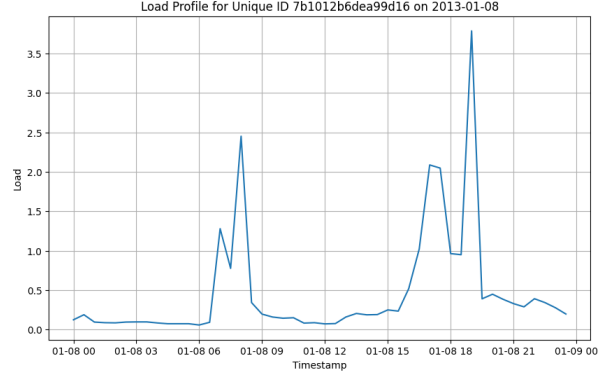
During the analysis, it was observed that using a representative load profile for the coldest day (for heating analysis) and the hottest day (for cooling analysis) provided comparable results to year-round data. This realization significantly reduced computational demands while maintaining accuracy. Figure 4.3 compares load profiles for a cold day (January 8, 2013) and a hot day (July 5, 2012), clearly illustrating the differences in energy consumption during these periods.



**Figure 4.2:** Demo Load Profile



**(a)** Profile for - 2012-07-05



**(b)** Profile for - 2013-01-08

**Figure 4.3:** Comparison of profiles for different dates.

Once this methodology was validated, it was applied to all unique house IDs in the dataset. Houses identified as having non-electrified heating and cooling systems were isolated into a new dataset. This refined dataset was merged with location data, enabling the mapping of potential customers for electrified heating and cooling systems. The results are visualized in Figures 4.1 and 4.2, while Table 4.4 provides a detailed breakdown of the classification outcomes.

**Table 4.4:** Hourly Load Data Comparison for House-Id "7b1012b6dea99d16"

| <b>Time</b>  | <b>05-07-2012</b> | <b>08-01-2013</b> | <b>Difference</b> |
|--------------|-------------------|-------------------|-------------------|
| 0:00         | 0.1               | 0.124             | 0.024             |
| 0:30         | 0.099             | 0.188             | 0.089             |
| 1:00         | 0.109             | 0.095             | -0.014            |
| 1:30         | 0.108             | 0.087             | -0.021            |
| 2:00         | 0.101             | 0.085             | -0.016            |
| 2:30         | 0.094             | 0.095             | 0.001             |
| 3:00         | 0.099             | 0.097             | -0.002            |
| 3:30         | 0.109             | 0.097             | -0.012            |
| 4:00         | 0.109             | 0.084             | -0.025            |
| 4:30         | 0.093             | 0.074             | -0.019            |
| 5:00         | 0.094             | 0.074             | -0.020            |
| 5:30         | 0.107             | 0.074             | -0.033            |
| 6:00         | 0.195             | 0.058             | -0.137            |
| 6:30         | 0.281             | 0.092             | -0.189            |
| 7:00         | 0.217             | 1.279             | 1.062             |
| 7:30         | 0.255             | 0.776             | 0.521             |
| 8:00         | 0.255             | 2.451             | 2.196             |
| 8:30         | 0.306             | 0.342             | 0.036             |
| 9:00         | 0.210             | 0.196             | -0.014            |
| 9:30         | 0.111             | 0.159             | 0.048             |
| 10:00        | 0.061             | 0.144             | 0.083             |
| 10:30        | 0.061             | 0.150             | 0.089             |
| 11:00        | 0.053             | 0.082             | 0.029             |
| 11:30        | 0.046             | 0.087             | 0.041             |
| 12:00        | 0.052             | 0.072             | 0.020             |
| 12:30        | 0.059             | 0.076             | 0.017             |
| 13:00        | 0.060             | 0.159             | 0.099             |
| 13:30        | 0.046             | 0.204             | 0.158             |
| 14:00        | 0.045             | 0.187             | 0.142             |
| 14:30        | 0.055             | 0.190             | 0.135             |
| 15:00        | 0.107             | 0.249             | 0.142             |
| 15:30        | 0.161             | 0.234             | 0.073             |
| 16:00        | 0.174             | 0.519             | 0.345             |
| 16:30        | 0.174             | 1.020             | 0.846             |
| 17:00        | 0.357             | 2.087             | 1.730             |
| 17:30        | 0.205             | 2.047             | 1.842             |
| 18:00        | 0.255             | 0.963             | 0.708             |
| 18:30        | 0.222             | 0.949             | 0.727             |
| 19:00        | 0.394             | 3.786             | 3.392             |
| 19:30        | 0.390             | 0.390             | 0.000             |
| 20:00        | 0.270             | 0.448             | 0.178             |
| 20:30        | 0.216             | 0.384             | 0.168             |
| 21:00        | 0.219             | 0.329             | 0.110             |
| 21:30        | 0.224             | 0.288             | 0.064             |
| 22:00        | 0.230             | 0.391             | 0.161             |
| 22:30        | 0.219             | 0.342             | 0.123             |
| 23:00        | 0.091             | 0.277             | 0.186             |
| 23:30        | 0.075             | 0.197             | 0.122             |
| <b>Total</b> | <b>7.573</b>      | <b>22.778</b>     | <b>66.753%</b>    |

## Chapter 5

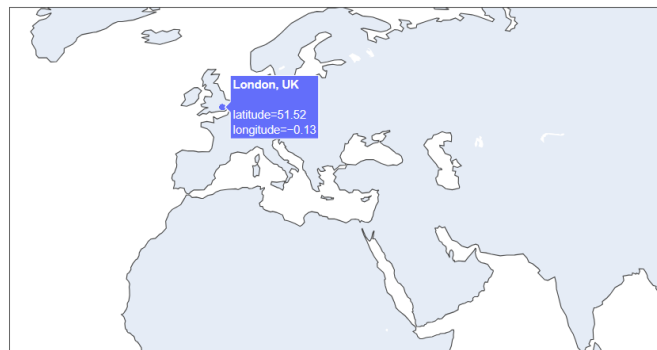
### Results

Following the methodology applied to all the unique house IDs, we generated a final merged dataset. This dataset includes only the unique house IDs, their corresponding latitude and longitude, and an indicator of whether the house uses electrified heating or cooling (see Figure 5.1).

|       | unique_id        | y     | location_id  | latitude | longitude | Electrified |
|-------|------------------|-------|--------------|----------|-----------|-------------|
| 0     | 012b7d955200d435 | 0.031 | gcpvj4cmfb0f | 51.52    | -0.13     | Yes         |
| 17568 | 018e5014e97c6d05 | 0.116 | gcpvj4cmfb0f | 51.52    | -0.13     | No          |
| 35136 | 024a2fda927fc576 | 0.123 | gcpvj4cmfb0f | 51.52    | -0.13     | Yes         |
| 52704 | 026dbf16600025c6 | 0.549 | gcpvj4cmfb0f | 51.52    | -0.13     | No          |
| 70272 | 028597f745cbea07 | 0.030 | gcpvj4cmfb0f | 51.52    | -0.13     | No          |

**Figure 5.1:** Final Dataset for Mapping

The original dataset contained coordinates for the houses; however, all coordinates pointed to the same location in central London. This limitation in the data underscores the importance of accurate geospatial information for meaningful mapping and analysis. Despite this, the methodology demonstrates that if precise location data were available, it would be possible to map the zones with non-electrified heating and cooling users effectively 5.2.



**Figure 5.2:** Illustrative Map for Non-Electrified Users

## Chapter 6

# Challenges and limitations

This project encountered several hurdles that shaped its journey and final outcomes.

### 6.1 Technical Challenges

1. **Data Collection and Quality:** Managing data inconsistencies was one of the biggest technical hurdles. Approximately 2-3% of the dataset was riddled with missing entries, requiring exhaustive preprocessing efforts. Cleaning and filling gaps while ensuring accuracy demanded both time and creativity.
2. **Granularity of Data:** The dataset provided readings at 15-minute intervals, which is sufficient for broad analyses but limited the potential of advanced energy disaggregation techniques. These methods typically require high-resolution data—such as second-by-second readings—to detect nuanced appliance behaviors.
3. **Model Implementation:** Implementing AI tools like NILMTK and deep learning frameworks proved to be a complex task. The lack of detailed metadata for appliance-level data and compatibility issues with existing computational infrastructure created significant bottlenecks. Additionally, debugging these tools required expertise and time that stretched the team’s capabilities.

### 6.2 Methodological Challenges

1. **Metadata Creation:** The absence of predefined metadata—describing appliances and their usage—posed a unique challenge. Generating this information manually across thousands of households was not only labor-intensive but also prone to inaccuracies.
2. **Balancing Manual and AI Methods:** While the manual disaggregation methodology developed during this project was practical and flexible, it fell short in precision when compared to AI-based models. This was especially evident in scenarios involving overlapping appliance signals or multi-state devices that required more granular insights.

## 6.3 Business Limitations

- **Customer Awareness and Adoption:** Many homeowners lack an understanding of heat pump technology and its advantages. Additionally, high upfront costs, despite long-term savings, acted as a deterrent for adoption.
- **Market Targeting:** Effective market segmentation and outreach were hampered by the lack of precise geospatial data in the provided dataset. Without specific location information, identifying and targeting households with non-electrified heating systems became a significant challenge.
- **Policy Influence:** The project's success hinges on alignment with government policies and incentives, which vary regionally and are subject to frequent changes. This variability complicates the creation of universally applicable recommendations.

## 6.4 Resource Constraints

- **Time and Budget:** Limited project timelines restricted in-depth exploration of advanced methodologies and their validation. Budget constraints further impacted the ability to secure additional resources for experimentation.
- **Computational Power:** Advanced AI methodologies require substantial computational resources, particularly GPUs, to train and implement effectively. The unavailability of such infrastructure constrained the scope of the analysis.

## Recommendations for Improvement

To address these challenges and enhance future projects, it is recommended to acquire high-resolution, labeled datasets to improve modeling accuracy and robustness, and leverage cloud-based computational resources to enable advanced methodologies without significant infrastructure investments. Strengthening stakeholder engagement will ensure alignment with policy frameworks and market trends, while public awareness campaigns should highlight the environmental and financial benefits of electrifying heating systems. Additionally, investigating hybrid approaches that combine manual and AI methods can maximize both accuracy and feasibility within project constraints.

## Chapter 7

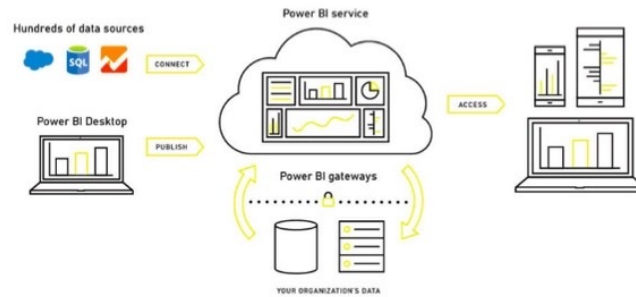
# Implementation and Usage

### 7.1 Implementation Plan

Many modern or newly constructed commercial buildings are likely equipped with some form of heat pump technology, especially those designed for habitation or frequent human use. In contrast, older commercial properties, such as warehouses and workshops, often lack this technology [3]. The UK government is making a strong effort to promote heat pump technology, with plans to phase out traditional boilers by 2035 [4]. In addition, it expects to increase its installation of heat pumps from approximately 35,000 annually to 600,000 per year by 2028. This significant growth is driven by the fact that heat pumps are a reliable and scalable solution to reduce carbon emissions in heating, as long as they are properly designed and installed [5].

Selecting the appropriate heat pump system for commercial spaces is essential for ensuring a comfortable and efficient environment. Heat pumps help regulate temperature, while also improving energy efficiency and reducing costs. Since commercial spaces have distinct needs compared to residential settings, it is important to thoroughly evaluate all factors before choosing a heat pump system. So, we aim to outline the steps for implementing heat pump sizing recommendations in commercial buildings:

- **Reduce heat demand:** It is crucial to reduce a building's energy consumption, especially in existing structures, to ensure efficient heat pump performance. Higher peak loads necessitate a larger heat pump and could result in the need for a more costly electrical grid connection.
- **Characterise the heating and cooling demand:** Evaluate heating and cooling load profiles to decide if a thermal store is necessary to improve heat pump capacity and ensure resilience. Identify the required outlet temperatures at the heat pump, as these will impact its efficiency. This is especially critical for retrofit applications, where existing systems often set temperature limitations.
- **Select an appropriate technology and size for the heat pump:** The designer must assess which technologies are suitable for the specific location and whether they can meet the required heating and cooling needs. Once the system configuration is decided, it is essential to size the heat pump correctly. An oversized unit will incur higher upfront and operating costs, occupy more space, produce more noise, and have a larger embodied carbon footprint.
- **Allocate space:** The placement of the heat pump system may be limited by factors such as noise, vibration,



**Figure 7.1:** Power BI workflow

airflow, and refrigerant safety, depending on the type of system chosen. Additionally, the visibility of the plant might be a concern due to planning regulations or architectural goals.

- **Optimise system performance:** Designers should create a controls narrative that defines the system’s operation and specifies the parameters to be monitored for triggering actions. Moreover, a thorough metering strategy is essential for optimizing system performance.
- **Installation and commissioning:** Installers should be adequately qualified, it is crucial that the scope of work is clearly outlined between the ground source contractor and the primary mechanical system contractor.
- **Operation:** User guides should be provided that cover: how the system is meant to operate and how to maintain optimal performance; key factors to consider when making modifications; safety precautions; troubleshooting steps; and routine maintenance procedures.

### 7.1.1 Integration with existing tools (e.g., Power BI dashboards)

One of the steps that can be done after the methodology of disaggregating the load has been optimized. The purpose is to use the results in a way they are compatible and synchronized to a tool that is able to show the data in an easy manner, for sales and marketing purposes. An example of this could be Power BI. One of its benefits is its ability to process and present large volumes of data related to energy consumption, performance, and system efficiency. Through interactive dashboards, we can easily track the performance of the heat pump system, monitor energy usage patterns, and identify areas where energy efficiency can be improved. This enables informed decision-making. Additionally, Power BI allows for the integration of data from various sources (see Figure 7.1, such as sensors within the building or external weather data, providing a comprehensive view of factors influencing energy consumption and system performance.

Another significant advantage of Power BI is its ability to improve collaboration, as we can access shared insights and data, ensuring alignment and consistency throughout the process. Furthermore, Power BI’s user-friendly interface allows non-technical users to interpret complex data, making it easier to understand the system’s performance and make data-driven adjustments.

Finally, Power BI supports the creation of custom reports and dashboards that can be tailored to the needs of different stakeholders, providing them with the most relevant data in a format that’s easy to comprehend. This flexibility helps in tracking key performance indicators (KPIs), such as energy savings and cost reductions, and aligns with the goal of ensuring the heat pump system is delivering optimal performance throughout its lifecycle.

## 7.2 Stakeholder Engagement

This project leverages data-driven methodologies, including load disaggregation and heat pump sizing algorithms, to provide tailored recommendations for optimizing heat pump performance. By understanding household energy consumption patterns and heating/cooling demands, this study offers valuable insights for utilities, energy companies, and governments to design policies and strategies that promote energy efficiency, reduce carbon emissions, and support the widespread adoption of electrified heating systems.

### Policy Development:

- **Energy Efficiency Standards:** Inform regulations or building codes that require the installation of energy efficient heating systems. By using accurate data on optimal heat pump sizing, governments can set guidelines that improve energy efficiency in new builds. Policy: "Energy Performance of Buildings Directive: Establishes a framework for the gradual transition away from fossil fuels in buildings, promotes the adoption of renewable solutions like heat pumps, enhances building energy efficiency, and tackles issues such as energy poverty." [7]
- **Carbon Reduction Targets:** Governments aiming to meet climate goals can incorporate these findings into their strategies to reduce carbon emissions by encouraging the adoption of energy-efficient heat pumps, which reduce fossil fuel dependence. Policy: "Fit-for-55 policy package: Speeds up the decarbonisation efforts across all sectors through a comprehensive legislative and policy framework. This initiative will boost renewable energy adoption, enhance energy efficiency, and introduce innovative financial mechanisms. Additionally, it incorporates building emissions into the new Emissions Trading System starting in 2027." [7]
- **Incentivizing Electrification:** Governments can use the findings to create policies that incentivize the electrification of heating and cooling systems. For instance, offering rebates or tax credits for households installing correctly sized heat pumps could accelerate the transition to clean energy solutions.

### Commercial Strategies:

- **Personalized Recommendations:** Based on the study's load disaggregation algorithm, utilities could create personalized recommendations for customers on the ideal heat pump size. This kind of targeted advice could help consumers understand the most cost-effective and efficient option for their home's specific needs.
- **Supply Chain Optimization:** Based on the demand forecasts from the study, energy companies could collaborate with heat pump manufacturers to ensure the availability of appropriately sized units for different households. They could work with manufacturers to streamline the installation process, provide bulk discounts, or organize group-buying campaigns for communities interested in heat pump adoption.

By analyzing and utilizing the results from this project, utilities, energy companies, and governments can design more effective policies and commercial strategies that promote the adoption of energy-efficient heat pumps. These strategies would not only accelerate the shift toward electrification but also optimize grid management, increase consumer savings, and contribute to overall environmental goals such as carbon emission reductions.

## 7.3 Scalability

Our project has significant potential to be scaled to other regions and types of homes, leveraging its data-driven methodology and universal applicability of heat pump technology. Below are two aspects that enable scalability:

- **Energy Policy Alignment:** It is a crucial aspect of scaling our project, as it allows the findings to directly contribute to national or regional energy goals, especially those focused on decarbonization and sustainability. Many countries, states, and municipalities have committed to ambitious energy transition policies, such as achieving net-zero carbon emissions by a certain year, increasing renewable energy generation, and reducing reliance on fossil fuels. The insights and recommendations from this project can help align with these energy policies by offering practical solutions to support these goals.
- **Regional Certifications and Compliance:** Regional Certifications are essential for ensuring that this project's recommendations are not only effective but also legally recognized and aligned with widely accepted energy efficiency standards. These standards, set by organizations like Energy Star (USA), the EU Ecodesign Directive, and the International Energy Agency (IEA), play a crucial role in driving the adoption of energy efficient technologies and ensuring that products meet the required environmental and performance criteria.

## Chapter 8

# Conclusion

This project has highlighted the significant potential of data-driven approaches in transforming heating systems and accelerating the shift toward electrification. By carefully analyzing energy consumption patterns and correlating them with weather data, the project delivered actionable recommendations for heat pump sizing tailored to specific household needs.

Despite the challenges—ranging from data limitations to resource constraints and market outreach difficulties—the project has laid a strong foundation for future initiatives. Some key insights include:

1. **Scalability:** The manual methodology, though less precise, has proven to be a practical approach that can be adapted to different regions and datasets. Its simplicity makes it a viable option for large-scale implementations with limited data.
2. **Policy Alignment:** The findings provide a valuable resource for government and regulatory bodies aiming to meet decarbonization and energy efficiency targets. They can inform policies and incentives designed to promote sustainable heating solutions.
3. **Consumer Benefits:** For homeowners, the project’s insights emphasize the tangible advantages of switching to heat pumps—including lower energy bills, reduced carbon footprints, and enhanced comfort.

Refining methods, adopting advanced technologies, and collaborating with stakeholders are essential to advancing cleaner energy systems, using data-driven insights for environmental and economic progress.

# Bibliography

- [1] Scottish Government. *Home Energy Scotland: Support for Heat Pumps*. Accessed: 2025-01-19. 2023. URL: <https://www.homeenergyscotland.org/find-funding-grants/renewables/heat-pumps/>.
- [2] UK Government. *Boiler Upgrade Scheme: What You Can Get*. Accessed: 2025-01-19. 2022. URL: <https://www.gov.uk/apply-boiler-upgrade-scheme/what-you-can-get>.
- [3] UK Government. *Energy Security Bill Factsheet: Low Carbon Heat Scheme*. Accessed: 2025-01-19. 2022. URL: <https://www.gov.uk/government/publications/energy-security-bill-factsheets/energy-security-bill-factsheet-low-carbon-heat-scheme>.
- [4] UK Government. *Future Homes Standard: Building a Greener Future*. Accessed: 2025-01-19. 2022. URL: <https://www.gov.uk/government/consultations/the-future-homes-standard-changes-to-part-l-and-part-f-of-the-building-regulations-for-new-dwellings>.
- [5] J. Kelly and W. Knottenbelt. “Neural NILM: Deep neural networks applied to energy disaggregation”. In: *Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments*. Accessed: 2024-08-23. 2015. URL: <https://dl.acm.org/doi/10.1145/2821650.2821672>.
- [6] J. Z. Kolter and M. J. Johnson. “REDD: A public data set for energy disaggregation research”. In: *Workshop on Data Mining Applications in Sustainability (SIGKDD)*. Accessed: 2024-08-23. 2011. URL: <https://zicokolter.com/publications/kolter2011redd.pdf>.
- [7] D. Rolnick, P. L. Donti, L. H. Kaack, et al. “Tackling Climate Change with Machine Learning”. In: *arXiv preprint arXiv:1906.05433* (2019). Accessed: 2024-08-23. URL: <https://arxiv.org/abs/1906.05433>.
- [8] Ines Syllame. *Energy Disaggregation using Deep Learning*. <https://github.com/inesylla/energy-disaggregation-DL>. Accessed: 2024-08-23. 2023.
- [9] NILMTK Development Team. *NILMTK: Non-Intrusive Load Monitoring Toolkit*. <https://github.com/nilmTk>. Accessed: 2024-08-23. 2024.
- [10] HomeBuilding UK. *This is the minimum house temperature to avoid damp, mould and condensation*. Accessed: 2024-08-23. URL: <https://www.homebuilding.co.uk/advice/minimum-house-temperature>.