DOCTORAL (PhD) DISSERTATION

UWE RADTKE

HUNGARIAN UNIVERSITY OF
AGRICULTURE AND LIFE SCIENCES, KAPOSVÁR
CAMPUS

2023

HUNGARIAN UNIVERSITY OF
AGRICULTURE AND LIFE SCIENCES

Doctoral School of Economic and Regional Sciences

Head of the Doctoral (PhD) School
Prof. Dr. ZOLTÁN BUJDOSÓ, PhD

Supervisor:
Dr. GYÖNGYI BÁNKUTI, PhD
Associate Professor

URBAN ENERGY SYSTEMS - ANALYSIS OF SMART METERING DISTRICT HEATING SYSTEM OF KAPOSVÁR

DOI: 10.54598/002700

Created
UWE RADTKE

Kaposvár, Hungary

2023

Contents

LIST OF FIGURES

LIST OF TABLES

1 Introduction

Challenged by climate change and coupled with the need to secure sustainable economic growth and social cohesion, Europe must achieve a genuine energy revolution to reverse present-day unsustainable trends and live up to the ambitious policy expectations. Toward this direction, district heating (and cooling) systems need to be more efficient, intelligent and cheaper. "Contrary to electricity smart meter data analysis, little research regarding district heating (smart) meter data has been published" (Tureczek et al. 2019).

The work was started during the Covid crises in 2020/2021. With the beginning of 2022 the main perception was the crisis is over and the European world would prosper again. But that in Feb. 2022 the Ukrainian war started and the European Union decided on drastic economic measures and sanctions packages – which at least during the ongoing study drove the European market into a recession. Many countries suffered a dramatic increase in energy prices and regulated prices became standard for most European countries.

The research will consist of several parts contributing to several questions regarding the reason and the use of smart meters in a district heating system. The aim is to focus on local needs and take a scientific, methodological approach to local problem-solving. The district heating system of Kaposvár will be explored and used as example to perform an economic and (customer) service research. With the analyses and the results, the heating plant has access to more information and could easier explore strategic suggestions. This can indirectly contribute to the efficiency of district heating, the reduction of environmental load and higher consumer satisfaction.

District heating (also known as heat networks or teleheating) (Goia, May, and Fusai 2010) is a system for distributing heat generated in a centralized location through a system of insulated pipes. This system is primarily designed to meet residential and commercial heating requirements, including space heating and water heating. The heat is often obtained from a cogeneration plant burning fossil fuels or biomass, but other sources such as heat-only boiler stations, geothermal heating, heat pumps, and central solar heating can also be utilized. Heat waste from factories and nuclear power electricity generation is also used and common. District heating plants can provide higher efficiencies and better pollution control than localized boilers. According to some research, district heating with combined heat and power (CHP) is the cheapest method of cutting carbon emissions, and has one of the lowest carbon footprints of all fossil generation plants. (Andrews 2009)

Consumer behavior is the study of how people make buying decisions. It attempts to understand how buyers choose and use products and services. By understanding how buyers think, feel and decide, businesses can determine how best to market their products and services (Southeastern Oklahoma State University 2022). Understanding buyers can help marketers connect with them and influence their behavior. In regard to sustainability – personal behavior is one of the key success factors. Understanding energy users' consumption patterns benefits both utility companies and consumers, as it can support improving energy management and usage strategies. The heat usage of customers is crucial for effective district heating (DH) operations and management.

1.1 District Heating (DH)

District heating refers to the provision of heat to buildings through a dedicated heating network that distributes thermal energy. This system involves the utilization of various heat sources such as power plants, solar thermal or geothermal installations, and large heat pumps to heat water. The heated water is then transported through a network of insulated pipes, typically buried underground, directly to the connected buildings. Within each building, the water passes through a handover station and enters the building's own heat distribution system, supplying heating energy and hot water. Once the water has cooled, it returns to the original heat source, forming a continuous cycle. Consequently, buildings that receive district heating are not reliant on individual heating systems and chimneys, as the centralized system efficiently provides the necessary heat and hot water. (Federal Ministry for Economic Affairs and Climate Germany 2023) District heating systems encompass a comprehensive supply infrastructure comprising district heating plants, facilities for pressure and volume maintenance, water treatment, district heating transport and distribution networks, and customer transfer stations. These systems operate in a state of balance, ensuring a constant equilibrium between the output and generation of district heating. To achieve efficient control, an additional heating center oversees the management of the system.

District heating represents a highly valuable product primarily generated through combined heat and power generation, wherein electricity and heat are produced concurrently. The process of district heating supply is straightforward: The heat is conveyed to customers through a pipeline system employing a transport medium, typically hot water. At the building level, heat transfer occurs in the house transfer station, enabling the efficient delivery of district heating to individual premises. (Rezaie and Rosen 2012)

District heating serves a dual purpose by not only decreasing the demand for resources but also safeguarding the climate. "…In the Net Zero Emissions by 2050 Scenario, the combined share of renewable sources and electricity in global district heat supplies together rises from 8% today to about 35% in the current decade, helping to slash heat generation carbon emissions by more than one-third..." (IEA 2022b). The Net Zero Emissions by 2050 Scenario (NZE) is a scenario proposed by the International Energy Agency (IEA) that outlines a prescribed pathway for the global energy sector to attain net zero CO2 emissions by the year 2050. This scenario emphasizes the advanced economies' capability to achieve net zero emissions earlier than other regions. Furthermore, the NZE scenario aligns with significant energy-related United Nations Sustainable Development Goals (SDGs), particularly in terms of attaining universal energy access by 2030 and substantial enhancements in air quality (IEA 2023).

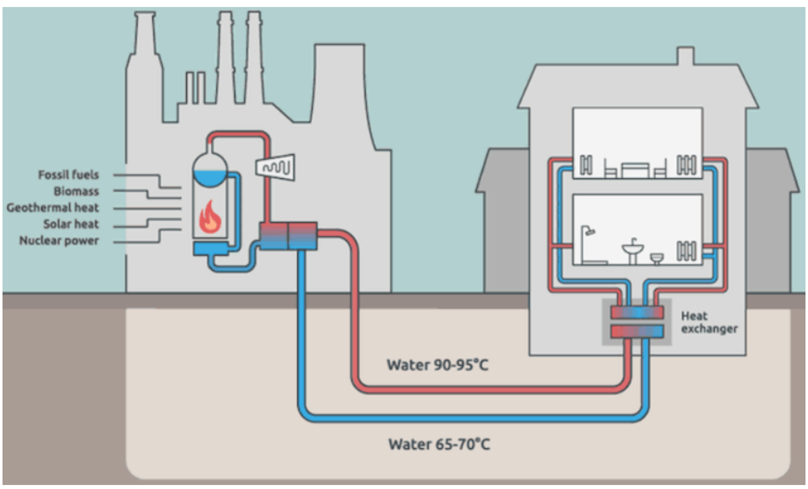


Figure 1: Image showing how district heating works (Toffetti 2015)

1.2 The role of district heating and its importance in Hungary

Power export relies on sufficient capacity at local interconnectors and demand for electricity in the importing country, taking into account the economic benefits of such export arrangements. The ideal energy storage solution would possess large capacity, fast charging capabilities, high recovery efficiency, and cost-effectiveness. However, achieving high scores across all these criteria can be challenging when focusing solely on a single energy sector. By adopting a broader perspective and leveraging the synergies between different energy sectors, it is possible to attain favorable outcomes in terms of all three criteria (Boldrini et al. 2022).

With the increasing presence of non-dispatchable power capacities, the need for flexibility is growing. District heating systems (DHs) have the potential to contribute to frequency containment reserves (FCR), automatic frequency restoration reserves (aFRR), and manual frequency restoration reserves (mFRR) markets by leveraging technologies such as CHP or power-to-heat and incorporating thermal storage (Gudmundsson, Thorsen, and Brand 2018).

District heating represents approximately 15% of the Hungarian residential heating market, serving around 650,000 heated dwellings. However, there has been a stagnation in the number of connected dwellings since 1990 due to the cessation of large prefab building projects, which make up more than 75% of the dwelling heating market (Sigmond 2009). Over the same period, residential heat demand has decreased by 33% due to changes in consumer habits resulting from the metering of hot water consumption and the gradual refurbishment of buildings and heating systems (Mezősi et al. 2017).

In Hungary, district heating is present in all major cities and most medium-sized cities, totaling 92 cities. There is even a single village with a local district heating system that serves over 60% of its cottages. The capacity of district heating systems varies widely, with approximately 36-36% of all heated dwellings located in Budapest and 10 other large cities, while 148 out of the total 202 systems have capacities of less than 10 MW (Mezősi et al. 2017).

1.3 Regulated and/or Deregulated

Regulated markets are described as follows: Vertical integration in the power sector can be defined as a situation where a sole utility company holds complete ownership and control over every stage of the power supply chain, from power generation to grid transmission and end-user distribution. In this arrangement, the utility company assumes the responsibility of generating electricity, delivering it to the grid, and ultimately supplying it to residential and commercial consumers. As a result, customers are left with limited or no options to select alternative power providers, and they have little influence over the selection of power generation sources or the determination of pricing structures (Tarassova 2020). In deregulated markets, the distribution and maintenance of electricity infrastructure, including wires and poles, are under the control of the utility. However, the delivery or supply of electricity to customers, both residential and commercial, is managed by specialized entities known as Retail Electricity Providers (REPs) or suppliers. These REPs take on the responsibility of delivering electricity to end-users, ensuring a reliable and continuous power supply. This separation of roles allows customers to choose their preferred REP, giving them the flexibility to select an electricity provider that best meets their needs in terms of pricing, renewable energy options, customer service, and other factors. (Tarassova 2020).

Starting from July 1, 2007, in accordance with the latest directives from the European Union on Gas and Electricity Market, all customers of electricity and gas have the freedom to choose their supplier without any restrictions. The Hungarian Parliament passed Act No. 86 of 2007, known as the Electricity Act (Körmendi 2018), with the aim of promoting the complete liberalization of the electricity market. This legislation was implemented to enhance economic competitiveness and ensure a sustainable and secure energy supply, aligning with the requirements of the European Union. The majority of the provisions within the act came into effect in 2007, and by the beginning of 2008, the electricity market in Hungary was fully liberalized. However, 2008 was considered a transition period during which market participants had to adjust to the new regulations. It is important to note that non-residential electricity consumers in Hungary bear significant subsidies for the renewable energy sector, which are incorporated into their tariff payments (Körmendi 2018). In 2016, the Hungarian Parliament approved a new premium-based renewable support scheme called METÁR, which was subsequently implemented in 2017. This scheme aims to provide financial support and incentives for the development and integration of renewable energy sources in the country. In the last couple of years, a change has been observed in the activity of the Hungarian State in the Hungarian energy market. The regulated energy and public utility prices for household residential customers have been gradually decreased, and price cuts are planned to be extended to industrial customers. Hungarian residential customers have been enjoying decreased energy and public utility prices since 2013 (Soyamedia.com 2022), as the end prices of electricity and natural gas universal service, district heating, water utility, chimney sweeping, and waste management services have been gradually decreased by law in Hungary. To ensure that customers are actually able to reap the benefit of these price cuts, new consumer protection rules have been implemented. There are plans to extend the price cuts to a certain extent to industrial customers as well.

While the markets might be deregulated – the prices for electricity as well as for district heating are regulated within Hungary. This has to be taken into account upon the further content of this thesis.

1.4 Demand, Consumption and Capacity

In the context of district heating, "demand”, “capacity " and "consumption" refer to different aspects of energy usage. The difference can be explained as follows:

Consumption: Consumption refers to the actual energy used by a heating system over a specific period. It represents the real energy expenditure required to meet the heating needs. Consumption is measured in units like kilowatt-hours (kWh).

Capacity: Capacity refers to the maximum capability of a heating system to generate heat or provide a certain level of heating output. It represents the system's ability to handle a specific heating load. Capacity is measured in units such as kilowatts (kW).

Demand: Demand refers to the amount of heat required or desired by a building or system at a particular time. It represents the heat load that needs to be met. Demand can vary based on factors such as weather conditions, insulation, and user preferences. It is typically expressed in units like kilowatts (kW).

Therefore, demand is indeed a meaningful concept in the context of heating as it represents the specific amount of heat required or desired, while capacity refers to the maximum heat output capability of the heating system, and consumption reflects the actual energy used to meet the heating needs. The thesis refers only on the demand measured within the heat transfer stations. Heat transfer stations in district heating systems play a key role in meeting the heat demand of connected buildings or zones. They regulate the flow and temperature of the heat-carrying medium to ensure that the required amount of heat is delivered efficiently. Heat transfer stations monitor and manage the energy need from buildings or zones, adjusting the heat flow and control parameters to optimize the distribution process. While consumption is measured at individual households, heat transfer stations focus on meeting the heat demand and ensuring effective heat delivery within the district heating system.

1.5 Research aim and objectives

The deployment of smart meters offers a unique opportunity for researchers and district heating utilities to analyze large-scale data and discover both typical, as well as atypical, patterns in the network. Within the research a data-driven approach shall be used to partition district heating users into separate clusters such that users in the same cluster possess similar consumption and behavior pattern. Because of the unavailability of high-resolution, hourly or sub-hourly meter data before the installation of smart meters, the literature on analytics in district heating is still in its infancy. There are not many studies focusing on the analysis of heat load patterns in district heating systems. So far only a few reference papers could be identified.

The proposed methods for this research will include the use of the K-means algorithm to segment the different groups based on demand intensity and representative patterns according to measured values. Understanding of consumers' opinion about district heating with Q-method-based opinion categorization method is planned to be a major part of the research. By quantifying subjective data, the research will get information about consumers opinions and mentalities – e.g. regarding the tariff system and expected (sustainable) behavior.

The objective of the research project is to address the questions: RQ1: Does fully anonymous (but identified with a unique ID) data measured by smart meters provide new insight? Along with that it shall also be explored if personalized data will be needed for direct influence and strategies. RQ2: Can sampled data effectively capture the cluster information of large datasets in clustering analysis, potentially reducing the hardware requirements for processing and analysis? RQ3: Which strategies could the district heating company apply to leverage the available measured data the most in regard to the general sustainability goals? RQ4: Can the strategies be related to the personal behavioral style? RQ5: Would consumers follow the strategy and support it?

1.6 Structure of the dissertation

The dissertation will include a literature review that focuses on the organization of district heating measurements. It will also examine the existing literature related to pricing aspects of district heating systems. The final part of the literature review will explore customer opinions and perceptions of district heating. Through this review, it was discovered that there is a limited number of research papers available on consumers' opinions regarding district heating in general, as well as the specific billing methods employed.

The first part of the literature review provided a lack of analysis of structured district heating data within the Eastern European countries – but revealed the main method of choice is K-means. Therefore, the provided data was clustered based in the K-means methodology in the next part of this work. Due to data privacy and protection requirements the set of provided data includes a huge volume of data but only very few additional parts. Only the meter number itself, the measurement, the time of the measurement and the operating hours of the meters were provided. The size or location of the dwellings measured by the meter were not available. It could be part of an additional analysis/research to cluster the available data based on house size, house age and number of inhabitants. Nevertheless, the clustering provided a first impression on the structure of the metered data.

The third main part of this work will contain the opinion research of a user group of district heating users. The group is based and supplied in Kaposvár – a city with 64,872 inhabitants (Város 2017). It shall be checked which general opinion the users have towards district heating, smart meters and corresponding smartphone apps and sustainability. The user opinions are retrieved and analyzed using Q-methods. Q methodology is a research method used in psychology and in social sciences to study people's "subjectivity"—that is, their viewpoint.

The last part of the dissertation will be the discussion and summary of the chapters and found users opinion before.

2 Literature Review

The following chapter analyses literature on selected aspects of district heating. It's split into three major parts – clustering, pricing procedures and consumer opinion towards different aspects of district heating. The goals of this study are as follows: (1) to identify and discuss fundamental research issues related to the topic; (2) to review and analyze previous investigations in order to establish the connection between this research and the existing body of knowledge; (3) to identify gaps in the current understanding of the subject matter. Although this dissertation focuses on Hungarian district heating, mainly international literature was reviewed. Even until recently, Hungarian aspects of many fields of research are performed in Hungarian language. But publications in English are increasingly common for the last decade (Camerlink and Pongrácz 2022), which could be an additional reason why district heating research long suffered from inadequate exchange with the international community. This holds true for many other strong native languages such as German as well. Furthermore, the international scope of this literature review is motivated by the aim to explore the field of urbans energy systems, especially district heating in a European context from a broader perspective. Within the image (Figure 2) below a short comparison between the term 'Hungary' and 'Magyarország' demonstrates the distribution between English speaking open access documents and Hungarian speaking documents. The search criteria is included within the figure – it was a search for only journal articles, starting from 2015 and only scientific documents. Based on (Holl 2022) the system Hungarian Scientific Bibliography Database (MTMT) can be used for monitoring OA mandate compliance.

Figure 2: Research in English and Hungarian language in Hungary

The search was limited by the following conditions: publication year greater or equal to 2015, scientific journal articles and the search term ‘Hungary’ or‘ Magyarország’. For the English term 6642 results were identified and for Magyarország’ 4952. The language of the publication was not available as limiting criteria; therefore, the given terms were used. Nevertheless, it seems as there are still a huge amount papers published in Hungarian language and therefore it can be assumed there are several papers available in Hungarian language on district heating.

The next subchapter was already published within a journal as (Radtke 2022a).

2.1 Structuring district heating data based on measured values

A literature review of the existing literature regarding district heating (DH) data and its clustering will follow within the next parts. Different approaches are used to structure the measured data – some based on consumption data and others based on heat load/demand data. Common methodology shall be researched and checked. New methods shall be double checked if reused in related work or developed single purposed only. Real-world databases are highly susceptible to being inconsistent, incomplete (lacking attribute values) and/or noisy (containing errors or outlier values). The major obstacle to obtaining knowledge is indeed poor data. There is the need to ensure that the knowledge discovery from the databases is in fact reliable. A research gap regarding the coverage of district heating in Eastern Europe has been identified.

Unfortunately, existing knowledge about customers and their heat load behaviors is quite scarce (Calikus et al. 2019). Previous studies have primarily focused on small-scale analyses, which may not provide sufficient representation to comprehend the behavior of the entire network. Contrary to electricity smart meter data analysis, little research regarding district heat smart meter data has been published. The deployment of smart meters offers a unique opportunity for researchers and district heating utilities to analyze large-scale data and discover both typical and atypical patterns in the network.

Sustainable development (SD) is a multi-aspect and complicated concept, and measuring it requires many data from a wide range of indicators (Mirghaderi and Ghiri 2019). In order to assist and facilitate cities in their pursuit of sustainability, various companies and organizations have established diverse networks and benchmark systems. These platforms are designed to gather, assess, and rank sustainability-related data, enabling cities to formulate and implement practical strategies, utilize effective tools and methodologies, and foster the exchange of best practices and lessons learned in the realm of sustainability. Analyzing the data enables integration of renewable energy resources (Radtke and Kaempf 2021). In addition, however, local conditions and the requirements and support from the European Union play a role that should not be underestimated.

Clustering district heating data is mainly based on two different dimensions: demand and/or consumption. Smart meter data are mainly available for electricity, and here, consumption-based data are available and measured (Völker et al. 2021). For district heating, very few installations use consumption-based data, while demand of mainly heat substations or large buildings (consisting of several flats) is available and used. A systematic literature review shall shed light on the methodology used to cluster data, the dimension measured, and the localized distribution shall be captured. Based on the literature review an own clustering approach could be derived – fitting best to the data available and the geographical location.

2.1.1 Methods and Materials

According to (Radtke 2022b) it’s important to use structured methods for unbiasing science. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) scheme was applied for the systematic review (Fink 2020; Guba 2008). PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. PRISMA primarily focuses on the reporting of reviews evaluating the effects of interventions, but can also be used as a basis for reporting systematic reviews with objectives other than evaluating interventions (Moher et al. 2009). Table 1 presents an overview of the research framework.

Table 1: Review protocol according to (Wohllebe 2020)

Review question	Which methods are used to cluster measured district heating data – differentiated by dimension?
Literature search	*Sources*: Microsoft Academic, Science Direct, google scholar and BASE
	Search Term: (“clustering method” AND (“district heating” OR “district heat”) AND (“data”) AND (“consumption-based” OR “based on consumption” OR “demand based” OR “heating capacity” OR “heat load pattern” OR “peak load” OR “load pattern”)
Filter criteria	*Type of work*: All type of publications *Years*: 2010 - 2022
Exclusions	*By title*: Examination of topics in a broader sense, exclusion of publications related to electricity (only)
	By abstract: Exclusion of articles not related to the combination *clustering* and *district heating*
Evaluation	*Full-text assessment*: Inclusion of those articles which are engaged with clustering accompanying district heat

The search for relevant records was conducted in the mentioned databases from October 10th until March 19th, 2022. The keywords were intended to cover the combination of district heating data and clustering methods. Because today's smart meter data are almost exclusively electricity-based, the term was excluded. The inclusion criteria were as follows: clustering heating data, which were used in combination with the corresponding dimensions (consumer, consumption vs. peak load, load profile, pattern).

The exclusion criteria were as follows:

- electricity
- estimated values, replacement values, substitute values

- Results about stand-alone installations
- Geographically originated data except Europe were excluded.

To provide three good reasons to exclude papers on:

1. Latency and time-dependent aspects: District heating systems can be influenced by latency and time-dependent factors, such as heat transfer rates, thermal inertia, and response times. Including papers on district heating allows for a thorough examination of these dynamics, which are not applicable to electricity. By excluding papers on electricity, the focus can remain on understanding and addressing the challenges specific to district heating systems.
2. Temperature dependency: District heating heavily relies on outside temperature conditions, as they directly impact the heat demand and energy consumption. By including papers on district heating, the study can delve into the influence of external temperature variations on the performance, efficiency, and optimization of district heating systems. Electricity, on the other hand, is not affected by external temperature changes unless it is specifically used for heating or cooling purposes.
3. Specificity of heating-related issues: District heating is a specialized area that involves unique considerations, such as heat distribution, pipe networks, thermal losses, and demand-side management. By focusing on district heating papers, it’ possible to address these specific issues and explore solutions tailored to this domain. Excluding papers on electricity, which may not be directly relevant to the heating aspects, allows for a more targeted analysis and a deeper understanding of district heating systems.

2.1.2 Results of the searches

The number of retrieved materials according to the search databases during the defined period is presented in Table 2.

Table 2: Initial results of the literature search

Search Term	Science Direct	Google Scholar	BASE	Microsoft Academic	Results after removing duplicates
(“clustering method” AND (“district heating” OR “district heat”) AND (“data”) AND (“consumption-based” OR “based on consumption” OR “demand based” OR “heating capacity” OR “heat load pattern” OR “peak load” OR “load pattern”)	833	153	549	80	821

In this systematic review, no type of publication was excluded due to its type. According to the mentioned Prisma Flow Chart, the following search steps were conducted:

- Identification. In that step, duplicates were removed, resulting in 821 of 1615 records remaining in the review process.
- Screening. The titles of these records were screened, and 567 were removed due to a missing relation to the research topic. Most of these removed records were in the field of traditional consumption clustering or in energy-related sectors. Additionally, a search

yielded various records from the field of computer science, including those related to smartphone technologies. However, these records were also rejected due to their lack of relevance to the research question. Following this step in the review process, the initial database was narrowed down to 253 publications that advanced to the next stage. The abstracts of these 253 records were carefully examined, resulting in the exclusion of 187 records that were deemed irrelevant.

- Eligibility. was then determined for the remaining 67 records, which were read in full and evaluated for their potential use in the systematic review. Out of these, 56 records were excluded for various reasons. As in the screening phase, some had a wrong setting, i.e., The combination of apps and smart meter data was not given, or the methodology used was based on electricity devices. Afterwards, the references of the records were screened completely to identify additional records.
- Included. Initially only eleven papers were identified. Based on the references of the analyzed papers one more article was reviewed and included into the work. The further discovered paper was accordingly noted in the including step of the Prism Flow Chart (Figure 3). Twelve of the papers were then assigned to the qualitative synthesis and consequently listed with their scientific results (Table 3). They were grouped according to their characteristics into the following sections: i) consumption-based cluster analysis and ii) demand-based cluster analysis. If several records appeared in the same year, they were arranged according to the first letter of the main author.

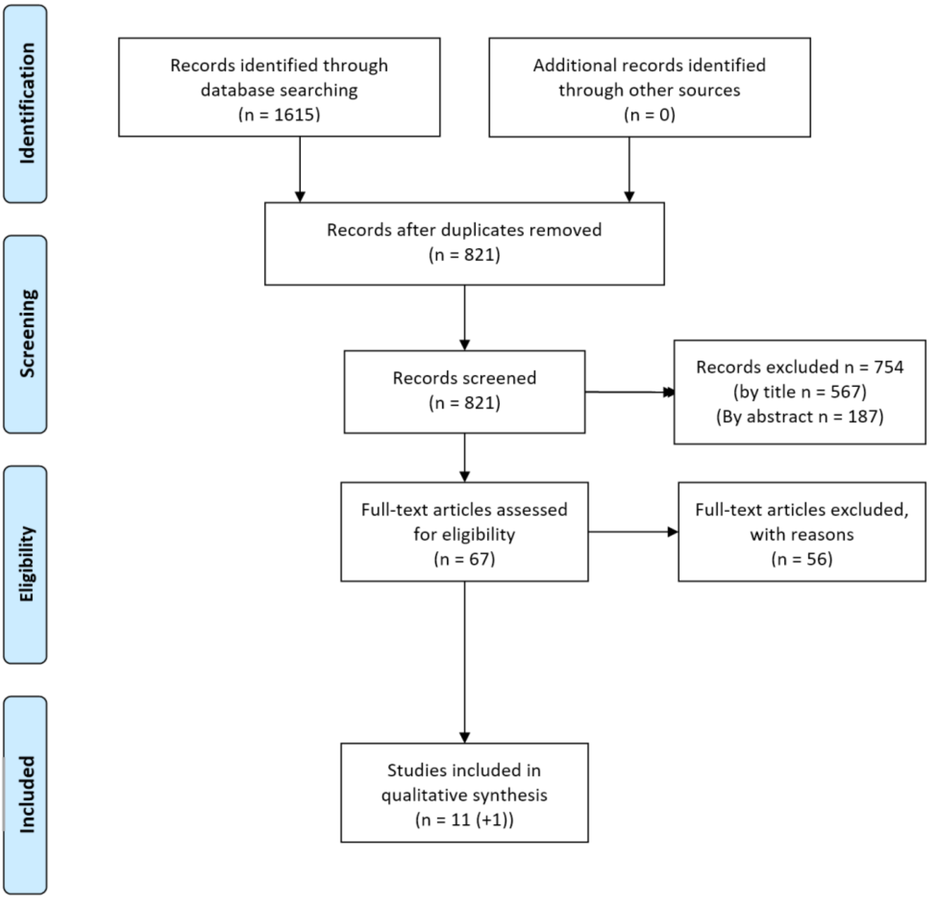


Figure 3: Prisma flow chart

2.1.3 Summary on used clustering methods

The focus is on investigating the clustering methodologies of the publications.

Table 3: Summary of the review's results

Author/Title	Key findings and methodology used	Geographical origin of the data
Demand based clustering (measurement in Watt, Kilowatt, Megawatt, Terawatt)		
(Calikus et al. 2019) A data-driven approach for discovering heat load patterns in district heating	• data-driven approach that enables large-scale automatic analysis of heat load patterns in district heating networks • three step patterns: data preprocessing, clustering, and visualization • K-means algorithm for clustering and removal of abnormal heat load profiles, before re-clustering again • clustering not by household, but by building • hourly measured values (taken by Smart Meters)	• Sweden
(Le Ray and Pinson 2019) Online adaptive clustering algorithm for load profiling	• clustering that consists into an iterative process based on the K-means algorithm that connects time steps • tried to generate four slowly changing typical load profiles by using generated profiles taken from ENTSO-E (European Network of Transmission System Operators for Electricity) • the clustering algorithm has been tested on two real-world datasets, (1) central district heating loads from 97 buildings in Copenhagen at hourly resolution for a month, (2) 13 241 electrical loads from industries, businesses and households with PV	• Denmark

	• for the district heating the consensus clustering is using a modified version of the K-means algorithm • for district heating data Root Mean Square Error (RMSE) does not depend on the number of clusters and globally decreases over the period • the methodology could be extended to multienergy profiling	
(Sala, R. Li, and Christensen 2019) Clustering and classification of energy meter data: A comparison analysis of data from individual homes and the aggregated data from multiple homes	• data used: heating consumption data of two different apartments and a district heating substation, as well as the outdoor temperature and solar irradiance • heating substation consists of hourly heating consumption of the apartment building with 72 different households • methodology used 2 approaches: (1) heating data is framed with each day as one observation and with hourly data as variables to identify the different daily heating patterns and (2) three different datasets were used, each of them consists of the daily mean value of outdoor temperature, solar irradiance and heating consumption of the apartments and the substation • clustering was done with several algorithms (1) nbclust() function in R which is equal to k-means (2) artificial neural network (ANN) (3) decision tree (DT) and (4) Random Forest (RF) • results showed that the models performed differently to the data of two individual homes and the data of the substation • for individual households, heating consumption is not necessarily dependent on weather	• Denmark

	• using data with regular patterns, i.e., heat substation data in this study, the prediction of future trend is reliable and accurate	
(Guelpa, Deputato, and Verda 2018) Thermal request optimization in district heating networks using a clustering approach	• clustered into 7 groups, clustering is performed through a K-means approach • used the model to simulate the heat flux daily evolution • measures and clustering based on buildings as they tried to respect the thermal conductivity and the inverse of the thermal capacity of the buildings • outcome: optimization is performed by modifying the thermal requests of buildings, anticipating the time the heating systems are switched on, i.e., by virtual storage	• Italy
(Gadd and Werner 2013) Heat load patterns in district heating substations	• clustering depending on the building properties but also of the type of activity that takes place in the buildings • two descriptive parameters: (1) annual relative daily variation and (2) annual relative seasonal variation • different types of reading included: Continuous operation control, Night setback control, Time clock operation control 5 days a week, Time clock operation control 7 days a week • 3 conclusions: (1) normal heat load patterns vary with applied control strategy, season, and customer category, (2) it is possible to identify obvious outliers compared to normal heat loads with the two descriptive parameters (3) the developed method can probably be enhanced by redefining the customer categories by their indoor activities	• Sweden

(Goia, May, and Fusai 2010) Functional clustering and linear regression for peak load forecasting	• focused on peak load, not on average consumption • used hourly observations data • summarized four series by plotting the daily mean data, observed a seasonal trend of each period • did not consider weather variables such as temperature but did use the seasonality trend and did not consider differences between weekdays and weekends or holiday as the data were taken for civil residences, which does not change considerably depending on the days of the week • forecast model based on a functional linear regression model which was good for December, January and February • another model was based on curve classification • evaluated the out-of-sample performances of the functional models	• Italy
Consumption base clustering		
(Du et al. 2019) Clustering Heat Users Based on Consumption Data	• two clustering methods: (1) clustering via daily consumption profile, (2) clustering via duration curve • K-means clustering scheme was applied to perform clustering • clustering performed by buildings	• Sweden
(Iglesias and Kastner 2013)	• similarity measures & Euclidean distance • Dynamic Time Warping (DTW) distance • clustered-vector balance is the self-developed and mainly used methodology	• Spain

	• FCM clustering □ k-means (soft k-means)	
(Tureczek et al. 2019) Clustering district heat exchange stations using smart meter consumption data	• applied learning from smart meter electricity consumption clustering to district heat exchange station clustering • clustering technique K-Means on different preparation of data: normalized data, standardized data, mean-centered data and mean-divided data • additional used: autocorrelation feature extraction and wavelet feature extraction for the cluster performance • data used was gathered by smart meters installed at Heat Ex-change stations	• Denmark
(Wang et al. 2019) New methods for clustering district heating users based on consumption patterns	• data basis: hourly heat consumption readings (in MW) of 561 users (Multifamily houses, offices and school, Hospital and social service) • clustering by GMM (Gaussian Mixture Models) mechanism which is based on a probabilistic model which assumes data points to be generated from a mixture of k (possibly multidimensional) Gaussian distribution • several different ways of clustering were used: (1) modified daily load profile (MDLP), (2) discretized duration curve (DDC) and (3) consumption-production consistency (CPC) • results: almost all users ambient temperature has strong impacts on the heat demand of all users, discretized duration curve can be used to group DH users, consumption-production consistency can be used to reflect similarity level between DH user’s	• Sweden

(Gianniou et al. 2018) Clustering-based analysis for residential district heating data	• three step patterns: data preprocessing, clustering and analysis • K-means algorithm (with (for hourly measures) and without normalization (for daily measures)) • segmented the customers into five consumption groups • observed seasonal variation • general clustering by households (not by buildings in the first place) • clustering by age of building, area of building, size of households	• Denmark
Neither or combined clustering approach		
(Marquant et al. 2018) A new combined clustering method to Analyse the potential of district heating networks at large-scale	• multiple energy systems in a MILP (mixed integer linear programming problem) problem becomes computationally demanding in terms of solving time when increasing the problem space by augmenting the number of integer variables (exponential increases of the solving time) • multiscale hierarchical approach for DES (distributed energy systems) optimization • a density-based and hierarchical algorithm is employed for clustering • combined electricity and heating • based on 32 buildings, hourly measured heat and electricity data	• Switzerland

2.1.4 Details of clustering district heating data

Demand-based clustering

In (Calikus et al. 2019) work, clustering was used to discover heat load patterns in DH networks automatically. A data-driven approach is used. They use a three-step pattern: data preprocessing, clustering and visualization followed by the K-means algorithm for clustering and removal of abnormal heat load profiles. The clustering is performed per building – which several other researchers have used as well (e.g. (Du et al. 2019)). Hourly measured values (taken by smart meters) provide a data-based fundament. The work concludes three main findings derived from the analysis of heat load behaviors among district heating (DH) customers. Firstly, a novel technique is introduced for grouping buildings based on similarities in their heat load profiles, preserving the shapes of these profiles and extracting representative patterns that capture the typical behavior within each cluster. Secondly, the study presents a method for identifying buildings with abnormal heat load profiles, indicating significant deviations from the expected patterns. Finally, the research highlights the identification of buildings with control strategies that are inappropriate for their customer category through visual examination and validation by domain experts, utilizing the discovered heat load patterns.

For (Le Ray and Pinson 2019) it was questionable to be included or excluded as the work partially contributes to data measured for electrical consumption. However, the main part of the work included hourly measured district heating meter data, and the work was considered relevant to be included in the review. The clustering methodology, which was always compared to K-Means, was enhanced to fit multi-energy clustering. (Sala, R. Li, and Christensen 2019) included the measured data for (1) two individual homes in an apartment building and (2) the district heating substation of the apartment building, which includes 72 homes. The K-means algorithm was applied to cluster the days with similar patterns based on the heating consumption, outdoor temperature and solar irradiance. Four different classification models were proposed to predict the heating consumption using the clustering results and weather conditions. For individual households, heating consumption is not necessarily dependent on weather conditions due to the high uncertainty and variability in occupants' daily activities and energy use behavior.

(Guelpa, Deputato, and Verda 2018) clustered their data into 7 groups, and clustering was performed through a K-means approach. They used the model to simulate the heat flux daily evolution. They used measures and clustering based on buildings as they tried to respect the thermal conductivity and the inverse of the thermal capacity of the buildings. This also underlined the importance of the usability of historical energy meter data.

(Gadd and Werner 2013) reached three key conclusions based on their K-Means-based approach: (1) normal heat load patterns vary with applied control strategy, season, and customer category, (2) it is possible to identify obvious outliers compared to normal heat loads with the two descriptive parameters, and (3) the developed method can probably be enhanced by redefining the customer categories by their indoor activities.

The study conducted by (Goia, May, and Fusai 2010) primarily examined peak load rather than average consumption, utilizing hourly observation data. They summarized four series by representing the daily mean data and identified a distinct seasonal pattern in each period. Notably, the study did not take into account weather variables like temperature but did consider the presence of seasonality. Additionally, differences between weekdays, weekends, or holidays were not considered as the data focused on civil residences, which exhibit minimal variation based on different days of the week

Consumption-based clustering

(Du et al. 2019) used two clustering methods: (1) clustering via a daily consumption profile and (2) clustering via a duration curve. They based the analysis on the K-means clustering scheme. The clusters themselves were based on buildings and not on single households. (Tureczek et al. 2019) was also almost out sorted as they applied learning from smart meter electricity consumption clustering to district heat exchange station clustering. However, as the study was rather derived from electricity, the work was considered relevant. Their clustering technique was K-Means on different preparations of data: normalized data, standardized data, mean-centered data and mean divided data. Based on the findings earlier from electrical clustering, autocorrelation feature extraction and wavelet feature extraction were used for the cluster performance. Similar to some others, the data used were gathered by smart meters installed at heat exchange stations (and not on single households).

(Iglesias and Kastner 2013) do not use the K-means algorithm directly but mainly base their work on similarity measures using the Euclidean distance, Mahalanobis distance, distance based on Pearson's correlation and Dynamic Time Warping (DTW) distance. Later they base the analysis on fuzzy clustering module that uses the FCM algorithm to compute clusters. The fuzzy c-means algorithm is similar to the k-means algorithm and is rated as an extension to it.

(Wang et al. 2019) also used hourly heat consumption readings (in MW) of 561 users (Multifamily houses, offices and school, Hospital and social service). As already seen within the demand-based clustering, authors prefer large buildings or rather combined usage. Single household measurements on an hourly basis are mainly not available for the district heating market. This could be the reason for using combined consumer data. The authors clustered the data by the GMM (Gaussian Mixture Models) mechanism, which is based on a probabilistic model that assumes data points to be generated from a mixture of k (possibly multidimensional) Gaussian distributions. Several different ways of clustering the data were used: (1) modified daily load profile (MDLP), (2) discretized duration curve (DDC) and (3) consumption-production consistency (CPC). They found that almost all users' ambient temperatures have strong impacts on the heat demand of all users, discretized duration curves can be used to group DH users, and consumption-production consistency can be used to reflect the similarity level between DH users.

(Gianniou et al. 2018) also used the K-means algorithm for hourly measured data. For single consumer households, the method was applied without normalization. As a result, they segmented the customers into five consumption groups. Similar to several others, the authors observed seasonal variations. Uniquely, the general clustering was done by households (not by buildings in the first place). Most interesting part was the clustering by age of building, area of building, size of households to discover efficiency reserves.

Neither or combined clustering approach

(Marquant et al. 2018) combined the clustering. Multiple energy systems in a MILP (mixed integer linear programming problem) problem become computationally demanding in terms of solving time when increasing the problem space by augmenting the number of integer variables (exponential increases of the solving time). They used a multiscale hierarchical approach for DES (distributed energy systems) optimization with a density-based and hierarchical algorithm employed for clustering. Similar to (Tureczek et al. 2019; Le Ray and Pinson 2019), they used a combined approach to cluster electricity and heating data. However, they used demand-driven and consumption-based data from 32 buildings and hourly measured heat and electricity data.

Conclusion

For the reviewed literature on clustering, mainly the methodology of clustering by K-means is described. The K-means algorithm has been considered to be the best known and most frequently used for clustering, which divides the data set into k clusters by minimizing the sum of all distances to the respective cluster centers (Ramos et al. 2015). Using K-means as a clustering algorithm is well covered by the literature and can serve as a basis for further tests on models and other clustering methods. Several alternative methods have been described and tested, but within the reviewed literature, no common basis for further cluster methods was found. Each researcher who used an alternative approach only did compare the results using K-means. The main difference between the articles is the data basis and the data preparation. Each article used its own set of data, some very small amount from only two apartments until up to over 500 measuring places within district heating (networks) while for comparable electricity measures close to 15.000 measuring points were analyzed.

However, the literature reviewed observed additional differences for those who performed a comparative analysis between electricity and heating measures: the influencing factors indicate that the outside temperature has a significant effect, including the area of living. The data gathered in Genova showed fewer peaks and less significant differences than the evaluation of data taken in northern Europe (GB, Sweden). The conclusion was rather identical. The most influential factor for volatile consumption was temperature and the type of building for which the measures were taken. The results did show an effect of modern isolation vs. no isolation at all. However, all researchers adjusted the data, which indicated that consumption remained stable during the period analyzed.

The usage of K-means as a clustering algorithm is mandatory and should not be skipped. Any additional clustering approach must be compared to the results of the K-means method. According to the reviewed literature, the K-means algorithm (which was included in several others such as nbclust(), GMM,…) did show similar results to all other used methodologies. Outliers will have to be respected and treated – but that's the general rule when using K-Means algorithm. The best documented approach according to the reviewed literature is K-Means; all other algorithms were only used by single papers, while two-thirds of the papers included statements and results regarding K-means.

When checking the geographical location of the data origin, the coverage of Eastern European countries is nonexistent. None of the observed materials were created using data from Eastern Europe. Reasons for this could be lack of interest in English speaking research or lack of research interest. But also, a lack of available data could be a reason for the research gap. The fourth explanation could be the geographical fixation of publishing journals. Here seems to be a research gap.

2.2 Review on prices and pricing mechanisms in district heating and the customer perspective to it

Initially only the customer perspective or the consumers opinion on district heating (all aspects) was supposed to be included, but there were not enough published results. Almost all documents about consumer opinion were in relation to the pricing or sustainability. To broaden the basis, the different pricing models were included into the review scope. For pricing mechanisms in district heating and the customer opinion towards DH, similar methods were applied as already described in chapter 2.2.1 and 2.2.2. The summary of those methods can be described as follows:

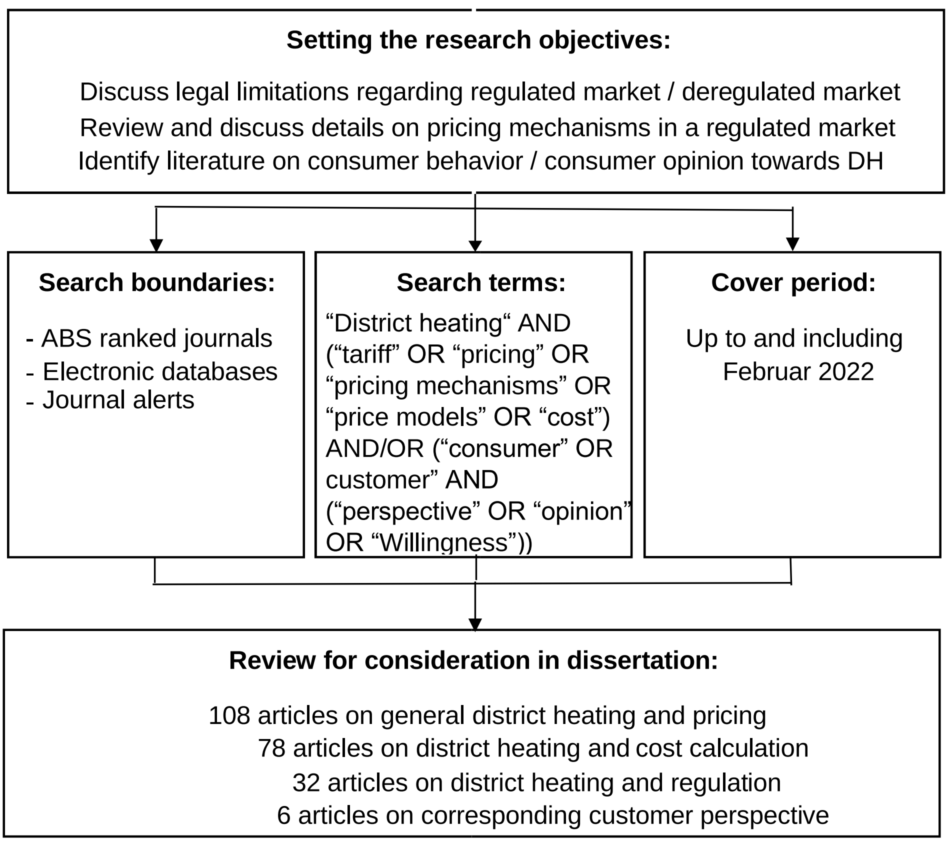


Figure 4: Summary on literature review

Given the fact that especially Hungary works with regulated prices throughout the energy sector in total, the publications having no relation to regulated price procedures were mainly excluded (some exceptions were needed).

The most prominent work of (Kerekes 2022) had to be excluded as it's available only in Hungarian language. It provides an overview of the available biomass and geothermal resources, the technical characteristics and cost elements of the different technologies and estimates the costs of increasing the share of heat generation from biomass and/or geothermal and CHP in the district heating systems studied.

The full Prisma flow charts for both areas (prices and pricing mechanisms and consumer behavior) can be found in the Appendix A.4 and A.5.

2.2.1 Details on prices and pricing mechanisms

The most common is the model of the heat energy market operating under two pricing conditions: a) free (liberalized) pricing and b) tariff regulation for consumers (Stennikov and Penkovskii 2020). The costs of DH depend on three main factors: (1) the connexion costs for customers, (2) the costs of a distribution network, which depend on the size of the DH network and its thermal loads, and (3) the production costs of thermal energy (H. Li et al. 2015). Mostly Swedish traditions were reported by the authors. Correspondingly, the price of heating mainly comprises a connexion fee, a standing cost and a unit cost. (Jingjing Song et al. 2016) show the same components but refer to a different naming: Fixed component (FxC) is the fixed price a user need to pay for being connected to the network. Load Demand Component (LDC) is essentially a variable component charged basing on user's consumption pattern (load demand), it usually covers DH company's nonproduction costs caused by investment on fixed assets, depreciation, salary, etc. All pricing schemes include energy demand component (EDC), which is based on the user's energy consumption. This component is supposed to cover DH companies' production costs.

(Stadtwerke Bonn 2021) also uses three price components to bill the end user with Annual base price + Commodity price + Emission price. Heating via district heating is very environmentally friendly. Nevertheless, CO2 emissions are also produced when district heating is generated, although these are much lower than with many other forms of heat generation. (Stadtwerke Bonn 2021)is obliged to present emission certificates for this purpose. A part of the certificates is provided to (Stadtwerke Bonn 2021) free of charge, the rest must be purchased in addition (Wehrle and Schmidt 2018). This part is distributed equally to all customers over their respective consumption. The consumption measured at the customer's district heating meter is multiplied by the emission price. Similar description can be found at (Stadtwerke Kaiserslautern 2021).

The synergies with power generation in CHP plants in district heating prices have been analyzed by (Åberg, Widén, and Henning 2012) and (Linden and Peltola-Ojala 2010). Without the synergy in price fixation, the papers would have been excluded as the analyzed markets are de-regulated.

In regulated markets, the price of DH is regulated by government and the regulated price dictates the profit made by DH companies. The price for district heating is equal to the sum of costs to be recovered and reasonable profits for DH companies – at least according to (H. Li et al. 2015). The equation for that can be described as:

(1) $$Price_{DH} = OA + AD + PP$$

where OA is operating cost, AD is annual depreciation, and PP is permitted profit. This method is called the cost-plus pricing method. Permitted profits (PP) can be calculated as:

(2) $$PP = WACC * RAB$$

where WACC is the weighted average cost of capital, and RAB = Depreciated fixed cost + new investment + labour cost (H. Li et al. 2015) and (Gudmundsson, Thorsen, and L. Zhang 2013). Under a cost-plus pricing mechanism, DH companies have incentives to increase profits by inflating costs, since permitted profits are usually related to costs. If they are operating on a lower cost than the reported level, the DH companies would be punished through the imposition of a lower level of permitted profits (Poputoaia and Bouzarovski 2010). Consequently, the cost-plus pricing method undermines suppliers' incentives to reduce costs and to upgrade their technologies. In addition to this, changes in real fuel costs cannot be transferred to consumers due to the use of historic measured metered data, and this prevents DH producers from generating enough profit to budget for necessary maintenance and improvements (Oh and Kim 2022). Within state-regulated pricing models it is likely that the tariff paid by consumers cannot cover the costs, and municipalities or regional governments fill the gap when this occurs.

The marginal-cost method is widely used in the deregulated market (Li et al. 2019). In a marginal cost-based pricing model, the total price normally involves two parts: fixed cost and variable cost, as formula (3) shows below:

(3) $$\mathrm{MC} = \frac{d(TC)}{\mathrm{d(Q)}} = \frac{d(FC+VC)}{\mathrm{d(Q)}} = \frac{d(VC)}{\mathrm{d(Q)}})$$

where TC is total cost, FC is fixed cost, VC is variable cost and Q represents the volume of heat production (Zhang, Ge, and Xu 2013). VC mainly consists of energy cost, labor cost and other variable operation cost, such as the cost for marketing. Energy cost or fuel cost can be calculated as (4):

$$(4) \qquad Fuel\ cost = Fuel\ price + Sulphur\ tax + NOx\ tax + Carbon\ tax + Energy\ tax$$

Another pricing method – so far only used within deregulated markets – is the method of levelized cost of heat (LCOH). It is the cost of generating heat for a particular system at a particular temperature of the working fluid. It is an economic assessment of the cost of the heat-generating system including all the cost over its lifetime: initial investment, operations and maintenance, cost of fuel, cost of capital (Gabbrielli et al. 2014). LCOH is the minimum price at which heat must be sold for a heat generating system, at a defined maximum temperature of the working fluid, to break even. Typically, LCOH is calculated over 20 year lifetime, and it is given in units of currency per kilowatt-hour, for example $/kWh or €/kWh or per megawatt-hour (Stanytsina et al. 2021). As this rather new method only applies to deregulated markets (Simón-Martín 2022a) with also deregulated prices, the methodology behind was not further researched. LCOH has its origin in the mainly deregulated electricity market which is indicated by several papers regarding the levelized cost of energy (Simón-Martín et al. 2022b) and (Li et al. 2019).

Regarding the prices itself, the work of (Werner 2016) is the most extensive and summarized paper for European market. The outputs from this price collection project consist of long time series of national average district heating prices until 2013 and the corresponding annual revenues and heat sales. In all, 560 annual average national district heating prices have been estimated. (Werner 2016) gathered an overview for a relatively long time series of district heating prices. The study encompassed 23 European countries, including the current membership of 20 European Union nations. The remaining countries consisted of Iceland, Norway, and Switzerland. Notably, the eight EU member countries excluded from the analysis were omitted due to the absence of significant district heating activity. Luxembourg, Belgium, Ireland, Portugal, Spain, and Greece have only a limited number of district heating systems, while Cyprus and Malta do not have any district heating infrastructure. During the analysis (Werner 2016) was able to find a huge variation in district heating prices. The study revealed that the highest prices are found in Denmark, Slovak Republic, Germany, Norway, and Sweden, while the lowest prices are obtained in Iceland, Bulgaria, Switzerland, Hungary, and Poland. Croatia and the United Kingdom show slightly higher prices – the rest shows significantly higher prices.

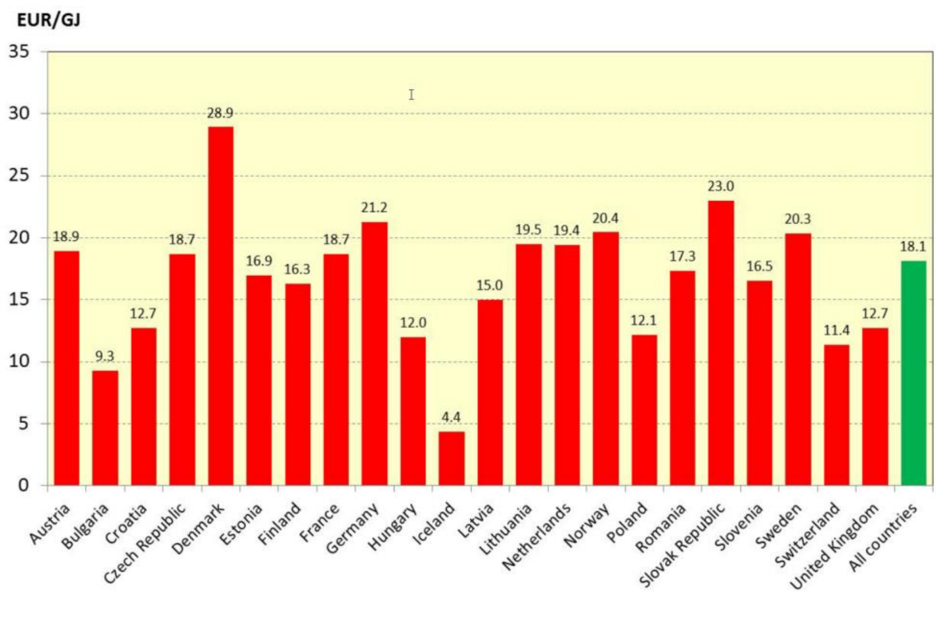


Figure 5: National average district heating prices for 22 European countries in 2013 (Werner 2016).

While the research of (Werner 2016) is based on reliable past figures, the political environment has change. The implementation of a utility cost reduction program in Hungary, initiated in multiple stages since 2013, has led to a significant decrease in regulated energy prices. This reduction has played a crucial role in providing affordable electricity, district heating, and natural gas to households. According to (Weiner and Szép 2022) the utility cost reduction program discourages energy conservation and energy efficiency; erodes the competitiveness of renewables; reduces gross capital formation in the energy sector; deteriorates security of supply; and increases energy prices for non-household customers. The effectiveness of the implemented measures remains uncertain, and various negative impacts have also been observed. Despite these drawbacks, the utility cost reduction program is expected to continue with some adjustments at most.

2.2.2 Customer perspectives on different aspects of district heating

While for pricing and pricing methods the number of identified articles was still quite high – for the consumers perspective or consumers view the result was very manageable. All found papers were included in the review. Regarding district heating the published knowledge on customer perspective is scarce. Only few relevant papers were identified which dealt with consumers opinion or customer behavior in district heating. For this review no geographical restriction was used.

(Krog et al. 2020) performed a literature review on transition towards 4th generation district heating (4GDH) with special interest on how consumers can be meaningfully and strategically included in the transition towards 4GDH. Their goal was to evaluate the consumer level’s role during 4GDH in the transition towards 100% renewable energy systems. They came to the conclusion that direct involvement of consumers either is not yet researched further, (too) difficult or even counterproductive. The authors found details on a free mobile phone app, called WATTS (initially provided by the Danish electricity supplier SEAS-NVE), to enable consumers to monitor their electricity consumption based on hourly smart meter data. Several other energy supply companies have joined a collaboration with SEAS-NVE with the objective to develop the app such that it can several types of consumption (electricity, heat, gas, water). The enhanced app allows consumers to follow their DH consumption hourly, daily, weekly and quarterly as well as their expected heat consumption use based on their consumption in previous years. Users can see their DH costs (refers to the financial expenditure associated with the consumption of district heating services). and are presented with a budgeted DH consumption (refers to the estimated or planned amount of heat energy that is expected to be consumed by users within a specific period, typically a fiscal year or budgeting period) based on previous consumption patterns. Colors indicate – green, yellow or red – when users are below, within or above their budgeted consumption. The authors concluded that more information technology (IT) is needed for a better link between demand and supply side. It’s needed for improving availability and exchange of information on a specific building’s performance. Furthermore, it’s needed to link building energy efficiency research with a consumer focus to 4GDH research.

(Sernhed, Gåverud, and Sandgren 2017) observed within their literature review: the Swedish market on district heating is better researched than any other country. They concluded especially for newer and more complex pricing models that most recent studies about price models for DH that were found in scientific journals are Swedish studies. Their work is to research how customers responded to more complex price models (introduced in Sweden). The methodology used to investigate was via focus group interviews and through interviews with companies that have changed their price models. Their results show that several important customer requirements are suffering with the new price models. The most important finding was the dislike of the fact in regard to the new pricing models, that energy savings do not provide financial savings, when costs are hard to predict and are perceived to be out of control. According to (Sernhed, Gåverud, and Sandgren 2017) factors like weather dependency, sunk costs from fixed assets and new competition on the heat market constitute challenges and business risks for the DH industry that must be considered. Most important finding in regard to any new pricing system in a deregulated (price) market: dissatisfied customers voting with their feet constitutes another financial risk for the DH business.

(B. Xu, Fu, and Di 2009) researched on dynamic consumer behavior, hydraulic performance and energy consumption of a DH system in China. While the other parts of the literature review exclude areas other than Europe, the scarcity of available papers is the reason to include this work within the consumer behavior part of the review as well. Traditionally consumers did not have a possibility to adjust their heat consumption, and the billing system was only based on the floor space of an apartment. The systems were operated with constant water flow rate and variable water temperature. Consumers did not have enough interest in saving energy, and always opened their windows to dissipate heat when a room was too hot. The China Ministry of Construction proposed national heat reforms with the primary objective of implementing a heat metering and billing mechanism in central heating systems to promote energy efficiency prior to the study already. As methodology questionnaires were used to explore how consumers behave differently because of their various temperature preferences, lifestyles and indoor thermal environment. Based on the answers tests and analysis of hydraulic performance within the district heating system were performed. The authors found that the influence of consumers' stochastic regulation behavior (by using thermostats instead of simply opening the window) on the hydraulic behavior within the district heating system is very slight. Additionally, they found heat metering billing systems can lead to about 10% energy savings compared with traditional billing systems. But main finding is the missing understanding of consumers on the self-adjustment of a thermostat. The authors suggest to advise and educate consumers on the use of the controls and the effect the controls have, e. g. set the radiator thermostats to reasonable levels, switch off the heating before opening the window, close the valves in unused rooms, and close all the valves only when going out for a long time.

(Ueno et al. 2006) performed a quantitative analysis on an on-line interactive energy consumption information system (ECOIS). This system was constructed to evaluate the motivating energy-saving activities in nine residential houses. Electricity consumption and space heating were measured in a combined approach as energy savings in total. During the research the energy awareness and energy-saving activities induced by this system were described and measured based on questionnaires and comparisons of power consumption before and after installation of an ECOIS. The result revealed that the power consumption of many appliances had been reduced by 9% after the usage of ECOIS. Furthermore, the analysis of daily-load curves and load-duration curves for individual appliances, both before and after installation, unveiled diverse energy-saving practices adopted by household members, including minimizing standby power and enhancing appliance operation control. The installation of an ECOIS influenced the energy-saving awareness of the customers.

What will consumers pay for convenience - was the question raised by (Yoon, Y. Ma, and Rhodes 2015). As method a double-bounded dichotomous choice method estimates consumer value for convenience, in a hypothetical market was used. The found that families with higher household income, higher heating expenditures during winter, higher educational achievement, and residence at relatively expensive apartments – in other words, those with higher living standards – assign higher value to the user convenience of DH. The Willingness to pay (WTP) for DH over individual heating (IH) as a percentage of current monthly heating costs was 4.03% in the used 800-household sample, and 7.92% for DH users only. IH users unfamiliar with DH expect little greater convenience (0.1% WTP), whereas the WTP for DH users runs to 7.9%, demonstrating consumer loyalty. As result it's recommended, according to their study, in order to foster DH, the many external benefits of DH systems should be stressed more and not their lower cost, but convenience, comfort, and safety.

(Krikser et al. 2020) performed similar research – the WTP for District Heating from Renewables of Private Households in Germany. The evaluated the willingness-to-pay for district heating and district heating from renewables compared to gas condensing boilers and heat pumps (individual heating). As method a discrete-choice experiment and collected data on attitudes towards sustainability, economic aspects and demands was used for providers of heat supply as dimensions for a factor and cluster analysis in order to apply a market segmentation.

Their results show a preference towards 'district heating from renewable energies'. Other alternatives like 'district heating from fossil fuels', 'heat pump' and 'gas' rank lower in the analyzed household's opinion. The participants revealed a significant additional WTP for ‘district heating’ just for the fact that it is from renewable energies.

2.3 Summary of the literature review

K-Means is a well-known approach for clustering data. The best documented approach according to the reviewed literature is K-Means; all other algorithms were only used by single papers, while two-thirds of the papers included statements and results regarding K-means. Although the K-Means algorithm has some possible disadvantages, it’s the bases for almost all reviewed work. Other methods, for example the Pearson Correlation Coefficient (PCC) could be used for specific interests e.g., to determine the dissimilarity measure to group the daily load profiles on the basis of the variation similarity instead of the magnitude similarity (Z. Ma, Yan, and Nord 2017). When checking the geographical location of the data origin, the coverage of Eastern European countries is nonexistent. Even before the functional limitation for heat load and patterns or even district heating, none of the observed materials were created using data from Eastern Europe. The research gap towards Eastern Europe as part of Europe might be related to language dependency, but more likely a research gap is identified.

For prices and pricing procedures – as long as a regulated price is guaranteed by government and local law, the tendency to invest into more sustainable technology is very low. Within the reviewed literature that was already concluded. New pricing mechanisms and pricing models are more likely to be accepted if the information is transparent and the end users can significantly influence their price by own behavior. A large common part for infrastructural means with a lower part for direct consumption provide less potential for an actual behavior change. Information is the key to success to any new model, but even to achieve behavior change, information is a key factor. While in earlier years email-based information system were tested and provided to end users, more and more apps running on a mobile taking over the information part. It remains to be researched to which intent end users will trust the new technology like smart meters with smart regulation and automated supervision.

In 2006 the availability of smart phones and tablet PCs was not given for average consumers, but already started – nevertheless almost all papers concluded that information is one of the key success factors towards any behavioral change of customers. While ECOIS was still email driven, the WATTS system run on mobile phones. But both systems included several divisions, not only district heating. The more the end user gets educated and information provided, the more the WTP increases as well for sustainable behavior. Without proven or trusted information, the users will stay to well-known habits such as opening the window instead of smart usage of thermostats will be kept. New prices are also more likely to be accepted in case the behavioral change has a direct influence and measurable payment reduction although the overall bill might be higher than before. The reviewed papers did show a high tensity towards more sustainable behavior but also a very high-cost sensitivity. The usage of gamification like presentation of sustainable behavior and the consumption profile in comparison to other similar households or end users was also a very common part in the information presentation. So even while not explicitly mentioned – signal colors like Red, Yellow and Green are commonly used and understood by the consumers. Visualisation in a graphical manner can be assumed to be highly contributing towards higher acceptance of any new information providing technology. Another common finding between the investigated papers was the used methodology – all used at least interviews and questionnaires in the research. The methodology of Q-methods was not used.

Within the literature review several research gaps could be identified. Some are obvious like the missing research data and research results from Eastern European countries. Others like the missing of research towards regulated pricing methods were not identified so easily. But especially for the last the research interest is very low as several aspects of the consumer behavior indicated a high tendency towards sustainable behavior in favor of low prices. When the heating source of a district heating network is clearly documented, and end user behavior demonstrates an influence on consumption – that is more likely to be accepted by consumers than simple price adoption. According to the literature it always has to be combined – sustainability, sustainable behavior and the information about it.

3 Material and methods

Within the research several methodological measures have to be combined. For the first part - the data-driven approach – to cluster the data. For that a behavioral clustering method will be used. This involves three major steps: (a) data preprocessing, (b) clustering and pattern discovery, and (c) visualization. In the first step, the data is cleaned, transformed and normalized. In the second step, k-shape clustering is performed to group demand having similar heat load behaviors. Abnormal heat load profiles, which do not conform to behavior in any group, are detected (and removed).

For the second part - a survey for statements has to be conducted and evaluated via quantifying subjective data using Q-methodology . Q-method is used to investigate the perspectives of consumers who represent different stances on district heating, sustainability and general attitude, by having participants rank and sort a series of statements.

3.1 Technical Overview

This chapter provides a comprehensive exploration of the technical background of the Kaposvár district heating. Additionally, it investigates the innovative methods employed for the effective utilization of waste heat within the system. By understanding the technical foundations and harnessing waste heat, Kaposvár aims to optimize energy usage, reduce environmental impact, and enhance the overall efficiency of its district heating infrastructure.

3.1.1 Technical Background on Kaposvár District Heating

The number of apartments/homes using district heating in Kaposvár in 2021 was 6 900, which represents 30% of the dwellings in the city, plus

305 other heated buildings are supplied. This is an exceptional, outstanding proportion in Hungary (Bánkuti and Zanatyné Uitz 2021).

In 2019 the natural gas-based heating plant operated with an installed thermal capacity of 51.7 MWt, the combined electricity generation part was 1.9 MWt. The installed gas engine electrical capacity is 1.35 MWe.

The length of the pipeline system is 38.6 km. The number of heat exchange stations is 385. The number of substations is quite high which is required within the regulation laws. The annual revenue of the district heating is HUF 2.4 billion (about 6.7 million EUR). The Kaposvári Vagyonkezelő és Szolgáltató Zrt.” (Asset Management and Services joint-stock Incorporation) was established in 1992 with several independent units, besides the district heating, like, building management, maintenance, and condominium management. Within the last couple of years – due to cost reduction and sustainability reasons a new 15 MW wood chip fired biomass heating plant installation project was implemented under the KEHOP 5.3.2 project. Figure 6 shows the close location of the biomass plant in the north of the city. The main supplier of it will be SEFAG. (SEFAG Zrt. Forest Management and Wood Industry Share Co.) (Kaposvári Municipal Property Management and Service Co., Ltd. 2022) It will be used for the production of locally produced wood materials that cannot be used for other purposes (cuttings, roots, other chips), without having to be transported far away.

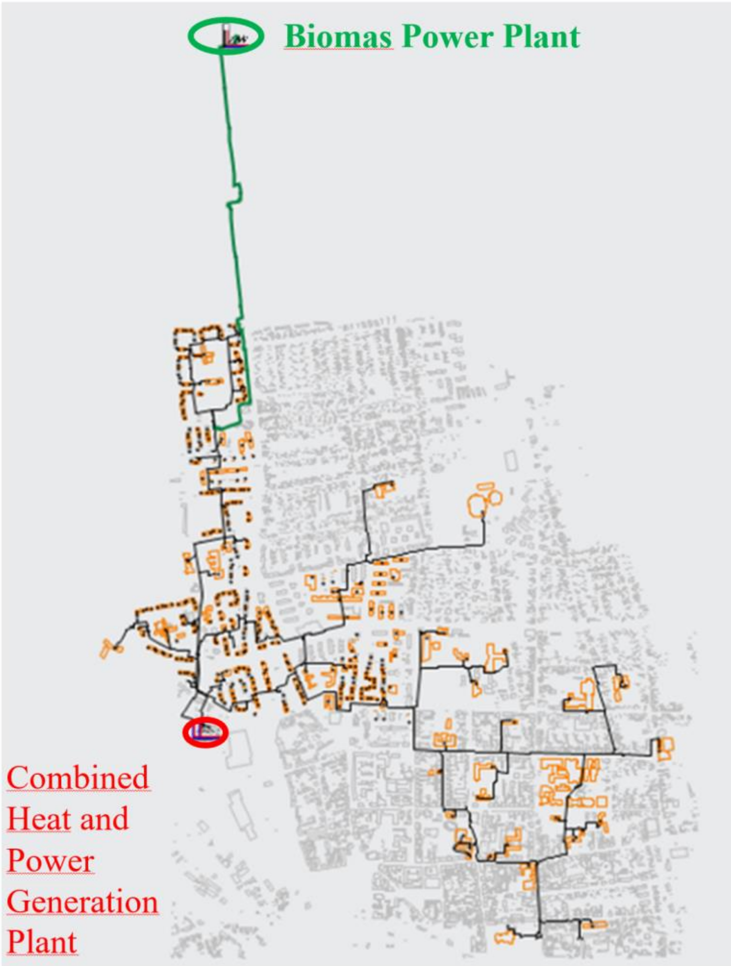


Figure 6: District heating grid and biomass plant of Kaposvár

In 2015, the city of Kaposvár completely replaced the outdated diesel-powered city buses with 40 compressed natural gas (CNG) vehicles, which significantly improved the city's air quality. The CNG filling station was installed on the site of the district heating company, where the necessary high-capacity electricity and high-pressure gas infrastructure was available. Namely, 2 x 132 kW of electrical power to create the necessary 200 bar – for the filling - from the available 6 bar pressure. For this high pressure, there must be high-quality, high-safety devices. In the district heating company, there is 1350 kW of low-cost electrical power available, from its generation. The accessibility of the heating plant is also ideal, as it is close to the route, terminal of the buses.

The installation of a biogas plant in 2007 contributed to the survival of the AGRANA subsidiary Magyar Cukor Zrt. (Hungarian Sugar Private Limited Company, hereinafter referred to as “the sugar factory”), in the worsening condition in producing beet sugar. The capital invested was HUF 1.7 billion, (about EUR 6.8 million). The power plant - with its two extremely large fermenters of 12,000 m^3 useful volume - was that to date is unique in the European sugar industry. Local professionals found the production reached the value of 140,000 m^3 with an average methane content of 53% (Csima and & Szendefy 2009). More than 50% of the factory's energy demand was fulfilled by the methane equivalent of 76-77,000 m3. Later further digesters and fermenters were added which increased the biogas production capacity to 2,500,000 m^3, covering about 80% of the needs during the production phase of sugar. However, the demand among the phases is minimal. Therefore, in the years after the start-up, the fermenters were tried to stabilize at very low level, almost shut down the gas production in the inter-production periods. This is a technically difficult and risky task and was a pioneering solution at the time. (Bánkuti and Zanatyné Uitz 2021). To balance the production volume, contracts were signed with the nearby swimming pool & spa and the heating plant. The gas from the fermenter is used directly in the sugar factory, in steam boilers, and steam turbines. It is used there also for cogeneration, to produce the energy (heat and electricity) needed for their beet sugar production process. The biogas is piped directly to the city spa (where there is a swimming pool, thermal baths, sauna, adventure pool, and beach in summer). It is located very close to the sugar factory and has a relatively high energy demand.

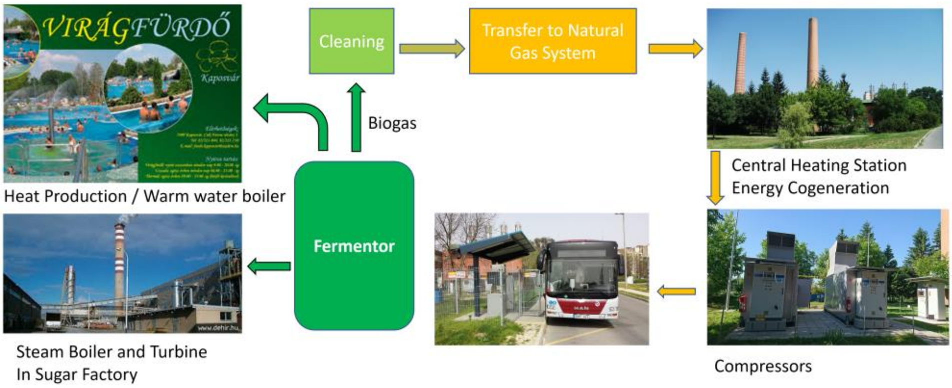


Figure 7: Urban Energy System of Kaposvár - The way of the biogas, from Agrana Sugar Factory

The biomethane after purification - according to the regulations - can be supplied to the gas network. The biogas produced in the Sugar Factory is filled into the gas utility system of EON and used where it is needed. The district heating division meets its energy needs by sourcing gas from the grid. The district heating consumption benefits from a price that is lower than the market rate. It generates synergistic benefits for all involved parties. This solution is unique in Hungary (Bánkuti and Zanatyné Uitz 2021).

3.1.2 Waste heat utilization

In the realm of waste heat utilization, cross-company integration presents a promising solution wherein surplus waste heat, which cannot be internally utilized, is directed to third-party entities such as commercial or residential buildings. This approach offers significant opportunities for energy optimization and sustainability. However, certain key challenges must be addressed, primarily revolving around the availability of accurate data to effectively match waste heat potential with corresponding demands.

One obstacle stem from the mismatch between waste heat availability and demand, making it necessary to carefully navigate this discrepancy. Currently, the most economically viable utilization of waste heat relies on the spatial proximity between waste heat sources and the locations where demand exists. This chapter explores the technical intricacies, challenges, and potential strategies related to cross-company waste heat utilization, shedding light on its feasibility, benefits, and areas for improvement (Fang et al. 2013). The utilization of waste heat through heat recovery or heat displacement represents the most efficient and straightforward technological approaches to enhance overall energy and cost efficiency. Heat exchangers are commonly employed in this process, facilitating the transfer of waste heat to a transport medium, which subsequently redistributes the heat to other units. However, it is important to note that some losses may occur during this heat transfer process. Furthermore, transferring waste heat to third parties requires additional transport infrastructure such as local and district heating pipes, buffer storage etc.. Local and district heating networks offer the distinct advantage of leveraging diverse heat sources in a flexible manner, encompassing both centralized and decentralized options. These networks operate seamlessly by integrating various energy sources at different levels and locations, irrespective of seasonal variations. This means that any economically viable waste heat can be extracted and effectively channeled into the heating network, leading to several benefits. Firstly, the company can reduce expenses associated with cooling water, while simultaneously generating revenue through the sale of heat energy. Additionally, this approach plays a crucial role in minimizing CO_2 emissions since the heat utilized in the network would otherwise need to be generated elsewhere. Thus, local and district heating networks provide a comprehensive solution that not only enables cost savings but also contributes significantly to environmental sustainability (Angelidis et al. 2023).

Glass stands out as a remarkably sustainable packaging material due to its composition from natural elements, extensive reusability, and complete recyclability. However, the manufacturing process of glass is characterized by high energy requirements, with furnaces operating continuously at temperatures exceeding 1500°C, around the clock. Consequently, glass production generates a substantial amount of waste heat, posing a significant opportunity for its utilization. Typically, the placement of an additional heat exchanger occurs prior to the flue gas treatment in a plant. This configuration eliminates the need to reduce the temperature, as long as it remains within the limit supported by the filter, avoiding dilution of the flue gas with external air or the use of water spraying methods (such as a quenching tower). The recoverable heat quantity and temperature from a single production line are often modest, which limits the feasibility of utilizing recovered heat for power generation using steam turbines, particularly when additional fuels are required to prevent steam overheating. In such cases, the Organic Rankine Cycle (ORC) emerges as an appealing solution for electricity generation from waste heat, even in situations involving low power and intermittent flows of hot gases with temperatures around 300ºC or even lower. The ORC demonstrates reduced sensitivity to changes in hot gas temperature and flow rate, offering greater operational simplicity and eliminating the need for specialized personnel. It also boasts lower operating costs and does not necessitate water treatment or consumption (Jouhara et al. 2018). Initially, the waste heat is harnessed for the generation of high-pressure steam or supplied to consumers with high-temperature requirements. However, this utilization results in the creation of waste heat at a lower temperature level. This surplus waste heat presents an additional potential resource, which can be effectively employed for various purposes such as heating products, serving as feed water, or acting as boiler water. Subsequently, what remains is low-temperature waste heat below 100 degrees Celsius, typically without any internal consumers. Instead of simply discarding this valuable energy, the optimal approach involves transferring it to a district or local heating network. These networks are typically designed to operate within specific temperature ranges, typically between 70 to 100 degrees Celsius. By integrating the low-temperature waste heat into such networks, its potential can be fully harnessed, further enhancing the overall efficiency and sustainability of the heating system (Jouhara et al. 2018). The significance of this development becomes evident as Şişecam, an esteemed industrial enterprise with a rich corporate legacy spanning over 85 years, expands its operations with a substantial investment exceeding 200 million EUR. This investment is dedicated to the construction of a state-of-the-art glass packaging plant in Kaposvár, located in the south-western region of Hungary. Originally established to address Turkey's fundamental glass product requirements, Şişecam has evolved into one of the most influential industrial conglomerates in the country, extending its influence globally across multiple sectors within the glass industry, as well as soda and chromium compounds business lines. This strategic investment in Kaposvár solidifies Şişecam's commitment to furthering its presence and impact in the glass packaging sector (Glass online 2022). The plant, which will be Şişecam’s first glass packaging factory in Europe, will have the capacity to produce 330,000 tons of glass packaging material a year (Glass online 2022). More details or figures on the amount of usable waste heat can’t be provided.

In general, highly energy-efficient glass production processes typically generate lower amounts of waste heat compared to less efficient processes. However, the exact amount of waste heat generated will depend on many factors such as the type of furnace used, the production process, the size of the plant, and the specific energy efficiency measures in place.

With the identified 4 clusters of heat demand, additional waste heat sources might be explored and included to the sustainable heat generation for the Kaposvár district heating. Above one potential additional source was mentioned. But again, further research and exploration would be needed as well as the technically feasibility needs to be researched.

3.1.3 Used Software for Comprehensive Analysis

Smart meters in district heating systems often require software for efficient monitoring, data collection, and analysis. The software helps manage and process the data obtained from the smart meters, allowing for real-time monitoring of heat consumption, remote meter reading, and advanced analytics. It enables utilities and operators to optimize energy distribution, detect anomalies or inefficiencies, and make informed decisions regarding system operation and maintenance. Additionally, the software may provide features such as billing calculations, customer management, and integration with other energy management systems. Overall, the software plays a crucial role in maximizing the benefits of smart metering in district heating.

There are mainly two important parts used in Kaposvár to be mentioned: TERMIS (Thermal Energy Analysis and Optimization of District Heating Systems) software. TERMIS is a software tool specifically designed for the analysis and optimization of district heating systems (Schneider Electric SE 2022). It provides a range of features to support the modeling, simulation, and analysis of thermal energy systems, including district heating networks. TERMIS software enables users to model the components and parameters of district heating systems, such as heat sources, heat exchangers, pipes, and consumers. It allows for dynamic simulation of system behavior, including heat flows, temperatures, and pressures, to assess system performance under different scenarios. The software also supports optimization functions to identify the most efficient operating conditions and configurations for district heating systems (Schneider Electric SE 2022).

READy is a software solution developed by Kamstrup, a leading provider of energy metering solutions. READy is designed to support efficient data management and analysis for utilities and energy service companies. The READy software offers functionalities for remote reading and management of energy meters, including smart meters used in district heating systems (Kamstrup 2023). It allows for automated data collection from the meters, eliminating the need for manual reading and enabling real-time access to consumption data.

With READy, utilities can efficiently monitor energy consumption, detect anomalies, and analyze usage patterns. The software provides tools for data visualization, reporting, and advanced analytics, allowing for informed decision-making and optimization of energy distribution. In addition to meter data management, READy software often integrates with other systems, such as billing systems and customer management platforms, to streamline processes and enhance overall operational efficiency (Kamstrup 2023).

Within this work the measured data provided by the READy system is used for the K-means cluster analysis. The following variables are available (inclusive their dimension). The used variables are marked in bold:

- **Cons_ID - Consumption place identifier [unit: --]**
- **Dat_Ro - Date of Readout [unit: --] -> transformed to YYYY-MM DD HH:MM:SS**
- E1 - Energy (meter reading) [unit: GJ]
- V1 - Volume (meter reading) [unit: m3]
- **Op_hrs - Operation hours counter [unit: hour]**
- **P1 – Power/Demand (instantaneous value) [unit: kW]**
- Constant for an hour [unit: kWh = Consumption]
- F1 - Flow (instantaneous value) [unit: m3/h]
- T1 - Flow-temperature (instantaneous value) [unit: °C]
- T2 - Return-temperature (instantaneous value) [unit: °C]

3.2 K-means Cluster analysis of hourly measured power demand.

As found and described within the literature review, the clustering approach shall be done via K-Means. K-Means clustering is a widely used method in vector quantization, originally developed for signal processing (Oti et al. 2021). Its objective is to divide a set of n observations into k clusters, where each observation is assigned to the cluster with the closest mean (also known as cluster centers or cluster centroid), serving as a representative of that cluster. This process results in a partitioning of the data space into Voronoi cells. The main goal of k-means clustering is to minimize the within-cluster variances, specifically the squared Euclidean distances. However, it does not optimize regular Euclidean distances, which is a more challenging problem known as the Weber problem (Bose, Maheshwari, and Morin 2003). While the mean value optimizes squared errors, only the geometric median can minimize Euclidean distances. To address the limitations of k-means clustering in terms of Euclidean solutions, alternative approaches like k-medians and k-medoids can be employed. These methods offer improved solutions in terms of Euclidean distances. By considering these variations, it is possible to explore different clustering techniques that suit specific data characteristics and optimization objectives (MacQueen 1967). It is a well-known approach for clustering data (Jain 2010).

The most common algorithm uses an iterative refinement technique. Due to its ubiquity, it is often called "the k-means algorithm"; it is referred to as native k-means (Giordani 2020):

Given an initial set of k means $m_1^{(1)}, ..., m_k^{(1)}$ (see below), the algorithm proceeds by alternating between two steps:

Assignment step: Assign each observation to the cluster with the nearest mean: that with the least squared Euclidean distance (Mathematically, this means partitioning the observations according to the Voronoi diagram generated by the means (Reddy and Jana 2012)).

$$S_i^{(t)} = \left\{x_p : \left\|x_p - m_i^{(t)}\right\|^2 \leq \left\|x_p - m_j^{(t)}\right\|^2 \ \forall j, 1 \leq j \leq k\right\}$$

,

where each is assigned to exactly one $S^{(t)}$, even if it could be assigned to two or more of them. $S^{(t)}$ are sets $S = \{S_1, S_2, ..., S_k\}$.

Update step: Recalculate means (centroids) for observations assigned to each cluster:

$$m_i^{t+1} = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_j \in S_i^{(t)}} x_j$$

The k-means algorithm is considered to have converged when the assignments of objects to clusters no longer change. However, it's important to note that the algorithm is not guaranteed to find the globally optimal solution. Typically, the k-means algorithm is presented as assigning objects to the nearest cluster based on their distance. It commonly utilizes (squared) Euclidean distance for this purpose. It's worth mentioning that using a different distance function, other than (squared) Euclidean distance, can potentially prevent the algorithm from converging (Li and Wu 2012). To address this limitation and enable the use of alternative distance measures, various modifications of k-means have been proposed. Examples include spherical k-means, which incorporates spherical distance metrics, and k-medoids, which employs medoids (representative objects) instead of cluster means. These adaptations allow for the utilization of different distance measures, enhancing the flexibility and applicability of the clustering algorithm. (MacKay 2003; Pelleg and Moore 1999)

The data was provided by the district heating company of Kaposvár. The used data consists of rd. 300 devices, measured hourly for 365 days. With a set of 3.3 million measures the analysis can't be performed on a simple PC anymore. But the research did prove - similar results can be achieved using randomized samples of 10.000 sets of data. As only two dimensions were available for public use and calculation - the result still will reveal 4 clusters in a 2-dimensional space. Large demand with least operating hours is something probably no district heating company wants. Of course - the method used to analyze the data contained standards like removing unmeasured sets, check if the data is skewed and transforming data.

3.2.1 Data description and preliminary steps

After the initial analysis, the data was converted from the given SQL Server format into a readable STATA format. Using STATA for transforming the date and time as well as the measured demand (P1) into computerized forms was the first step. But the later steps were performed using R respectively R Studio as the library and documentation was easier to access and the R Studio did allow a computerized output. Additionally, it was possible to convert the STATA computed file with the name "measures_corr.dta" into the R readable format. The complete script computed with R can be found at the appendix (A.1) of this work.

Data

The first step to get a first impression was to show the basic statistics of the whole dataset:

Data description:

```
obs:     3,332,901
vars:            6                          24 Nov 2021 21:31

              storage   display    value
variable name   type    format     label      variable label
cons_id         long    %12.0g                Cons_id
Dat_Ro          str19   %19s                  Dat_Ro
Op_hrs          int     %8.0g                 Op_hrs
p1              str15   %15s                  P1
P1_numeric      float   %9.0g                 P1
datetime        double  %tc
```

Figure 8: Description of the unprocessed dataset

The variable cons_id describes the unique device ID, P1 and P1_numeric represent the measured demand in the unit kilowatt (kW). Op_hrs is the parameter for the operating hours and datetime is the already converted value for date and time store in dat_ro. P1 and dat_ro were provided as string values which can't be used for further analysis and had to be converted already in the very initial step.

Cleaning of data and outlier handling

Data Cleaning, Outlier Handling (Winsorizing):

Data cleaning is a crucial step in the data preprocessing phase, which involves identifying and addressing issues such as missing values, inconsistencies, and outliers. Outliers are data points that deviate significantly from the majority of the data and can potentially distort analysis and modeling results. One approach to handle outliers is through a technique called Winsorizing.

Winsorizing is a statistical method used to mitigate the impact of outliers by replacing extreme values with less extreme ones. Instead of removing outliers entirely, Winsorizing modifies their values to be closer to the rest of the data distribution. This helps maintain the integrity of the dataset while minimizing the influence of outliers on subsequent analyses.

In Winsorizing, the extreme values are replaced with a predetermined percentile value, often the highest and lowest values within a certain range. For example, in a 5% Winsorization, the top 5% of the values would be replaced with the value at the 95th percentile, and the bottom 5% would be replaced with the value at the 5th percentile.

By Winsorizing outliers, the dataset's statistical properties and relationships among variables can be preserved, allowing for more accurate analysis and modeling. However, it's important to carefully select the appropriate percentile threshold for Winsorization, as it can impact the results and should be chosen based on the specific context and requirements of the analysis. Overall, data cleaning and outlier handling techniques like Winsorizing are essential in ensuring the quality and reliability of data for further analysis and modeling tasks.

Data completeness assessment

Within the next step, the number of unique device IDs were detected , the datetime was rounded to full hours and unique moments were detected. Additionally, the determination minimal date und maximum date was performed. The dataset initially contains more than 365 unique dates with measurements, resulting in more than 365 days * 24 hours worth of unique data. To ensure consistency and have exactly 8760 measurements per device, duplicate rows were removed. In total, 288310 duplicate rows were dropped from the dataset.

The next step involves forming a cross-join from the unique time points (dates) and unique IDs to complete the framework of the data frame. This is achieved by creating a new data frame called "dt" using the CJ (cross-join) function, combining the unique dates and IDs.

By comparing the number of rows in the cross-joined data frame (a) with the original data frame (b), the difference (a - b) indicates the number of missing Date/ID combinations. In this case, the result is 60745, implying that there are 60745 Date/ID combinations that are missing from the data. These missing combinations are filled with NA values in the cross-joined data table.

Furthermore, it is mentioned that negative values and zero values in the data need to be set to NA. These values could arise due to various reasons such as late start of measurements, calibration issues, device defects, or environmental influences impacting the accuracy of measurements. Setting these problematic values to NA helps ensure the reliability and validity of the data for subsequent analyses.

3.2.2 Data processing and empirical results

Report metric statistics for non-factor features:

Table 4: Statistics after data pre-processing (computed by R, own representation)

Variable/measure	vars	n	mean	sd	median	trimmed
P1	1	3044437	16807.51	6057.13	17423.0	16948.06
Operating hours	2	1795745	19.24	44531	16.2	17.51
Variable/measure	**min**	**max**	**range**	**skew**	**kurtosis**	**se**
P1	5583.0	26840.0	21257.0	-0.20	-0.88	3.47
Operating hours	4.4	51.3	46.9	1.16	0.73	0.01

At this stage the mean and standard deviation (sd) of both variables for later re-transformation of cluster centers have to be stored/kept. Further, the distribution of these features with histograms are shown:

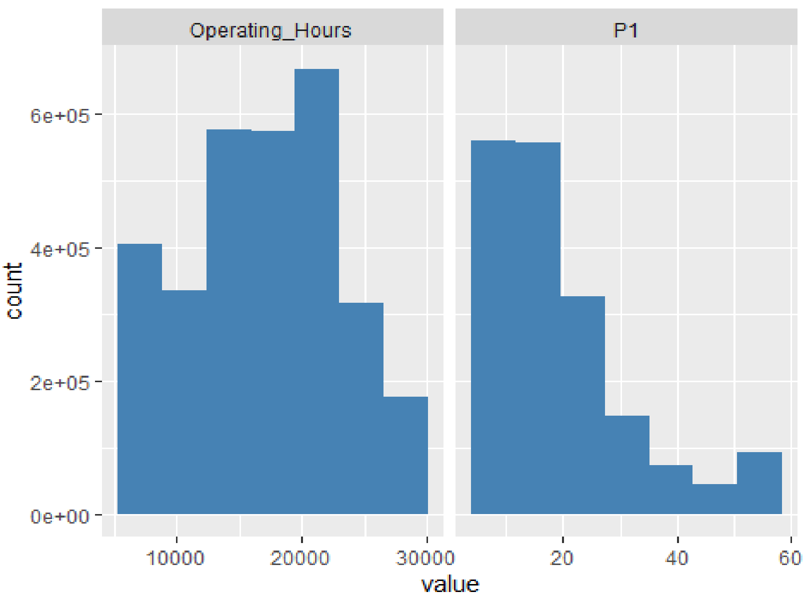


Figure 9: Distribution of P1 and operating hours untransformed

P1 (unit is kW) heavily skewed and is therefore log-transformed in the next step. Visualize the histograms again to see the effect of the transformation:

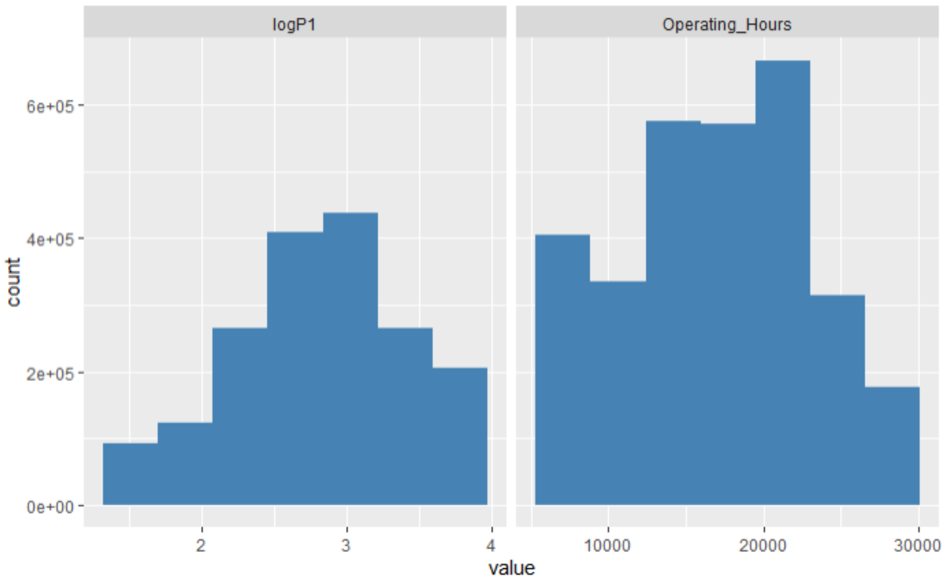


Figure 10: Distribution of logged P1 and operating hours after transformation

P1 normal now (see Figure 10) and can be used for modelling. Additionally, the column is renamed to logP1 for easier access and reading. The boxplot (Figure 11) also reveals the normal distribution of the now used values. To visualize the boxplot, it was needed to transform data from a wide format to a long format.

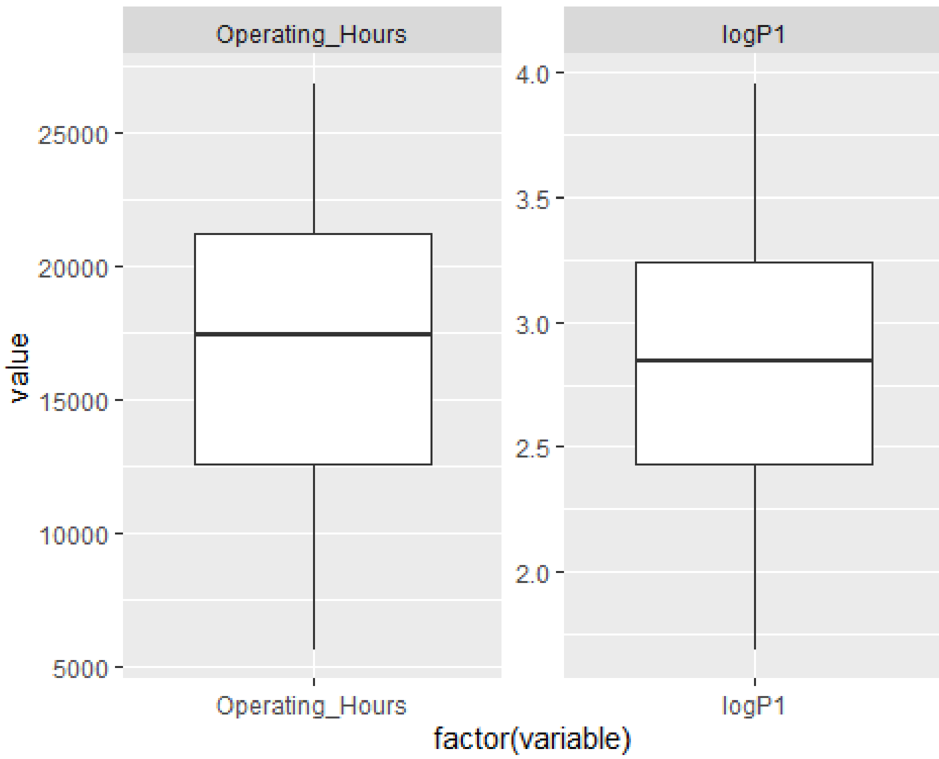


Figure 11: Boxplot of operation hours and transformed P1

Finally, the data preparation is almost finished. Last steps include the removal of lines with missing values (already marked via NA). It is also needed to standardize the further used variables by subtracting the mean and dividing by the standard deviation. Standardization, also known as feature scaling, is often applied to data for various reasons. Within the context of this work, it's used to achieve a comparison of variables: Standardizing variables ensures that different variables are on a similar scale, which allows for meaningful comparisons between them. When variables have different scales or units, it becomes difficult to assess their relative importance or make accurate comparisons. Standardizing variables eliminates this bias and ensures fair treatment of all variables.

The correlation (see next table) shows the correlation coefficients between Operating_Hours and logP1.

Table 5: Correlation between Operating_Hours and logP1

	Operating_Hours	logP1
Operating_Hours	1	-0.009
logP1	-0.009	1

The correlation coefficient between Operating_Hours and logP1 is approximately -0.009. This indicates a very weak, almost negligible, negative correlation between the two variables. The negative sign suggests that as Operating_Hours increases, there is a slight tendency for logP1 to decrease, and vice versa. However, the correlation coefficient being close to zero indicates that there is no substantial linear relationship between the two variables.

3.3 Consumer's opinion on district heating using Q method

Q Method was developed by William Stephenson to examine individuals' psychological attitudes. Q method is a factor analysis, which analyses the people themselves not their characteristics (William Stephenson 1955). The method employed in this study centers on individual differences and shares a similar mathematical foundation with factor analysis. By utilizing the Q method, a considerable number of statements can be evaluated with a relatively small sample size of participants. Correlation coefficients that are calculated by the method show correlation between people. The Q method can be considered as an inverse factor analysis. The typical benefits of Q methodology include insight into the perceptions of individuals at a level where broad social forces are enacted within individual awareness. According to (Ramlo 2016): "…methodological aspects of Q offer the ability to scientifically study subjectivity." "The qualitative methods of Q-methodology allow participants to express their (subjective) opinions and the quantitative methods of Q-methodology use factor analytic data-reduction and induction to provide insights into opinion formation as well as to generate testable hypotheses" (Valenta and Wigger 1997). Q-methodology research focuses on understanding the qualitative aspects of how and why individuals think in particular ways. This methodology places importance on exploring the subjective experiences and perspectives of individuals rather than quantifying the frequency or prevalence of specific thoughts or opinions among a group of people (Valenta and Wigger 1997). For that purpose, Q provides a more systematic approach and higher methodological transparency than purely qualitative methods (Brown 1996). Distinct viewpoints on any topic are limited and therefore, any set of statements clearly reflecting a broad heterogeneous range of opinions. Manifested by diverse participants it will reveal the existence of groups with similar viewpoints. Q-methodology integrates qualitative and quantitative approaches to explore the subjective perspectives of individuals who have direct involvement in a specific topic. The primary objective of Q-methodology is to reveal distinct patterns of thought rather than focusing on their numerical distribution within a larger population. Studies employing Q-methodology typically involve small sample sizes, and the findings are less susceptible to the impact of low response rates when compared to survey-based studies (Valenta and Wigger 1997).

The Q methodology is quite well established within research in the field of medicine and health care as well as psychology. Of course – it's originated there. The steps and the usage differ between publications. Some refer to 5, others to 6 and some to 7 phases as well.

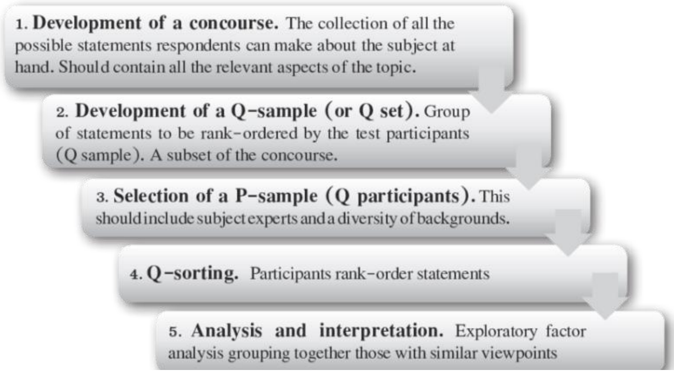


Figure 12: 5 phases of Q methodology according to (Stone and Turale 2015)

For phase (1), development of the concourse, a set of statements that reflect the range of perceptions about the research topic has to be developed, either by application of primary or secondary data. Next, for phase (2), developing a set of statements about the issue (Q-sample or Q-set) has to be performed. Phase (3) consists of selecting participants representing different perspectives (P-sample or P-set), phase (4) requires the participants sorting (Q-sort). Phase (5) is about analyzing and interpreting the data from the Q-sort. These five phases can be found for example with (Balch and Brown 1982), (Stone and Turale 2015) and (Yeun 2021).

(Ladan, Wharrad, and Windle 2018) and (Ordóñez et al. 2020) describe the 6 phase approach as follows:

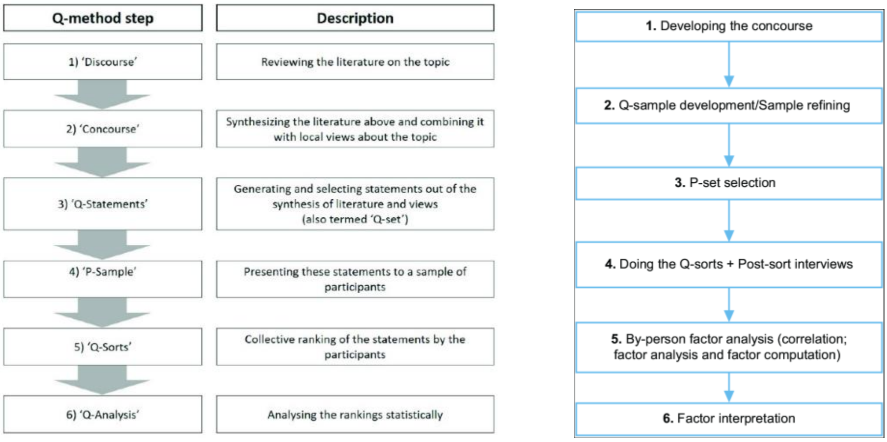


Figure 13: 6 phases of Q methodology according to (Ordóñez et al. 2020) vs. (Ladan, Wharrad, and Windle 2018)

Within the work of (Ordóñez et al. 2020) the phases are very close to the already described 5-phase approach. Only the first step – Discourse – differs. This first step is used by the authors to develop the concourse while other authors perform similar steps. (Ladan, Wharrad, and Windle 2018) created an own adoption of the Q methodology but kept the key phases as all the other authors very close to the original approach.

(Damio 2016) and (Churruca et al. 2021) refer to a 7 phasis approach. But here the steps were defined and used identical by both authors. Step (1) includes defining and building the concourse, step (2) developing the Q Set, step (3) selection of P Set. Conducting the Q Sorting is defined es step (4) and step (5) is defined as post a Q Interview. Steps (6) refers to analysis and step (7) as interpretation of the result.

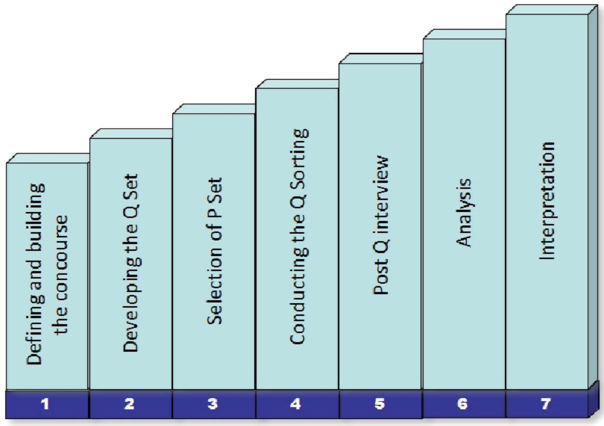


Figure 14: 7 steps of Q methodology according to (Damio 2016)

All three methodology descriptions are very close to each other within the steps. Within this work the 5-step approach will be used with some specialties in regards to develop and selecting the Q Set. Within this dissertation, the Q method is just a methodology used like a tool. There is no desire to develop further the Q methodology itself. The Q methodology shall be used for retrieving consumers opinion on district heating and possibly help policy makers with the result in guiding society to a more sustainable behavior and consumption.

3.3.1 Defining the concourse, Developing Q-Set and Selection of P-set

Q-methodologists have established diverse sources and methodologies to construct a concourse, which includes scientific literature, expert interviews, focus groups, social media, and websites. These sources gather existing opinions and arguments, encompassing the viewpoints of representative organizations, professionals, and other experts pertaining to the subject matter. Prior to the development of the concourse, a research question must be identified to guide the entire research process. Therefore, the following questions were applied to define the concourse:

- What influences the opinion on district heating?
- What can be identified as everyday behavior of consumers?
- Which digital experience is expected and can contribute to shaping opinions?
- Is sustainable behavior known to consumer?

To determine the statements, several sources were used, including the already reviewed literature, lectures and discussions with professors, expert interviews and also statements from websites. At the end of the concourse development, 61 statement items were identified with four themes: General district heating, prices, requirements/expectations for software/apps and personal behavior.

The Q method often raises the question of the reliability of the method. In order to objectivize the Q-set definition, a Delphi-like technique was applied. Delphi technique is a structured method to facilitate consensus of expert opinion. Although initially developed for military forecasting (Dalkey and Helmer 1963), it has since been applied to many research areas, for example healthcare (Keeney, Hasson, and McKenna 2011). The technique involves a panel of experts who undertake a series of questionnaire rounds.

For this study, a two-round Delphi-like technique was used to further reduce the number of statements and to achieve expert consensus on the statements to include in the Q sample. The Delphi component was conducted from December 2021 to January 2022. An international multidisciplinary team of experts were selected. Experts were defined as researchers or professionals in the area of energy billing or even specialized with heating. Finding experts only in the area of district heating was lacking a direct connection and the possibility to interact via online conferencing.

Using the broader definition enabled the Delphi-like technique to be applied via MS Teams and Skype meetings:

- Round one

A defined number of experts was invited to participate in the first round. They were invited to rate their agreement on statements. In rating each statement, the experts were requested to consider the relevance of each statement based on their experience and their knowledge. The overall intention is to reduce the number of statements. Experts were advised that participation involved a commitment to several rounds.

- Round Two

At the commencement of the round, panel members were informed that, based on feedback, statements would be reworded where necessary. Statements that achieved the median requirement were kept. Any new statements suggested by the panel in round one were included and presented for rating.

The panel members were asked to rate the presented statements according to the following scale: 1) Fully relevant regarding the intended research questions 2) Matches the basic idea 3) Not very relevant 4) Shall be replaced

The feedback of round one was evaluated and quantitively calculated. Each statement has been given a score. The score determines if the statement is kept within the next round or is replaced by another statement. Statements which have been majorly receiving a score of (3) were replaced by a statement provided within the free text. Round two was sent out after round one was evaluated, and the statements were rewritten to the same number of experts. Round two did provide the (semi-)final set of statements to be used within the Q Methods approach. An additional round was done – as the panel members were not familiar with the Q-methodology and the statements had to be shortened more. The general text was kept but the explanatory additional hints were totally removed so the participants can create a strong opinion towards each statement. Q methodology has been criticized for lack of transparency and detail around Q sample construction. For example, it has been stated that "the QM (Q methodology) literature remains uncomfortably silent with respect to how to assemble and verify completeness of a concourse, and how to verify or falsify the representativeness of a sample drawn therefrom" (Kampen and Tamás 2014). Therefore (Kirschbaum, Barnett, and Cross 2019) and (Wallis, Burns, and Capdevila 2009) proposed a Delphi based approach which was adopted to a Delphi-like approach within this dissertation. At the end of the process 39 statements were selected and presented as P-set to the rating participants (see Appendix A.3 for all statements).

3.3.2 Design of the study procedure and selection of participants

In designing the study procedure, where participants perform a Q sort, a sorting range with nine categories was utilized following the recommendation for Q samples smaller than n = 40. Although (Webler, T., Danielson, S., & Tuler, S. 2009) suggested to add labels as "least like how I think" to "most like how I think", the labels were set a bit broader. From "Disagree, Least Preferred, Least important" to "Agree, Most preferred, Most important" nine categories were created. The sorting arrangement is supposed to represent a quasi-normal distribution that is symmetrical over the middle and represents a normal distribution. For 39 statements, the best way to force this distribution was through the following arrangement of positions into which participants would sort statements (see Figure 15).

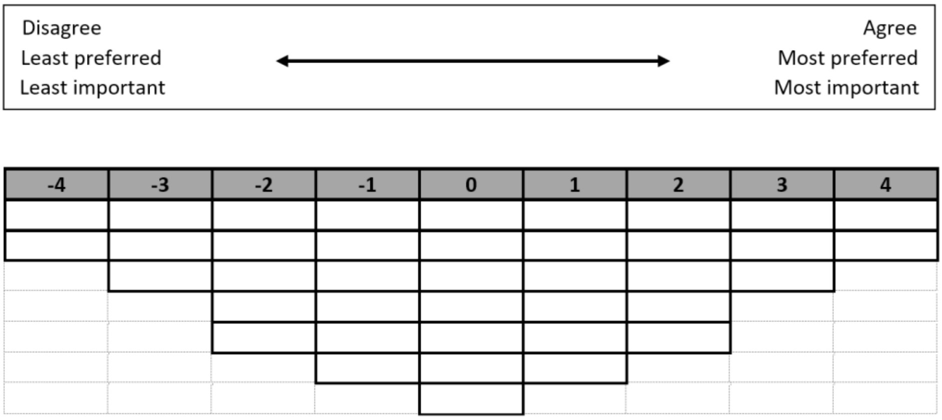


Figure 15: Q-sorting grid for 39 statements

The Q sort, the study procedure of Q methodology, can be conducted classically by printing out the questionnaire and distributing the given statements or in an electronical way using already available software. This depends on the personal preference of each participant. Some might use paper and scissors while others make use of modern technical devices and use Google sheets or MS Excel.

Q methodology combines qualitative and quantitative methods to investigate the subjective views of those directly involved in a particular topic. According to (Webler, T., Danielson, S., & Tuler, S. 2009) the number of participants should match the number of statements. Minimum ratio of Q-Participants and Q-Statements is 3: 1. The ratio that should be used is 2: 1. In this research with 39 statements, 13 (39/3 = 13) to 20 (39/2 = 20) participants will have to be expected.

In order to determine some additional criteria, the questionnaire was sent out with the socio-demographic questions (see Appendix A.3).

The P-set consists of district heating users of Kaposvár. The opinion on the Kaposvár district heating users shall be explored and analyzed. In order to avoid language barriers, the Q-sets and the surrounding questionnaire was provided in English and Hungarian language. The distribution of the questionnaire itself was done by the district heating technical manager of the Kaposvár Municipality Asset Management and Service Co, Zsuzsanna Zanatyné Uitz. The results were gathered, scanned and anonymously sent to me. There is no trace to the participants or any personal data. The survey was not just confidential but truly anonymous. Responses to analyze the data by socio-demographic factors such as education or age-range. Participants could give an incorrect answer, but high level of inaccuracy is not expected.

4 Results

The Results chapter presents the findings of the study, highlighting the outcomes obtained through data analysis and Q-methodology. The chapter includes an analysis of the data using K-means clustering, providing insights into the patterns and structures discovered. Additionally, the results of the Q-sorts are presented, offering valuable information on subjective views and opinions.

4.1 K-means results

4.1.1 Data analysis using K-means

Sampling needed using seed (by setting the seed also at subsampling you force the same output at multiple runs) – a sample of 10 000 values was used. Sampling was needed in order to enable reproducible results. The initial dataset could not be handled on a PowerBook workstation with Xeon CPU with 6 cores and 12 Threads at 2.7 GHz and 64 GB RAM. The clustering was performed once using a virtual server cluster with 8 CPUs and 1024 GB RAM – but that system was not available after an initial run. Therefore, the sub-sampling approach was used. The results between the full dataset and sampled dataset will be shown and compared.

Initially for testing and setting up the algorithm a random number of k = 5 were used. The detailed data distribution is removed of the below script, only the sum of square was kept for later comparison.

```
# testing with k = 5
kmeans(df, 5)

## K-means clustering with 5 clusters of sizes 2170, 2226, 1949, 1838, 1817
##
## Cluster means:
##   Operating_Hours         logP1
## 1      -0.1773503  0.0006644251
## 2       1.0755369  0.5087342724
## 3      -1.2743402 -0.7885253689
## 4      -0.6833280  1.2495590751
## 5       0.8686841 -1.0934166095
```

```
##
## Clustering vector:
##
## Within cluster sum of squares by cluster:
## [1]  546.4458 1041.9130 1331.6470 1206.7509  999.0389
##  (between_SS / total_SS =  74.4 %)
##
```

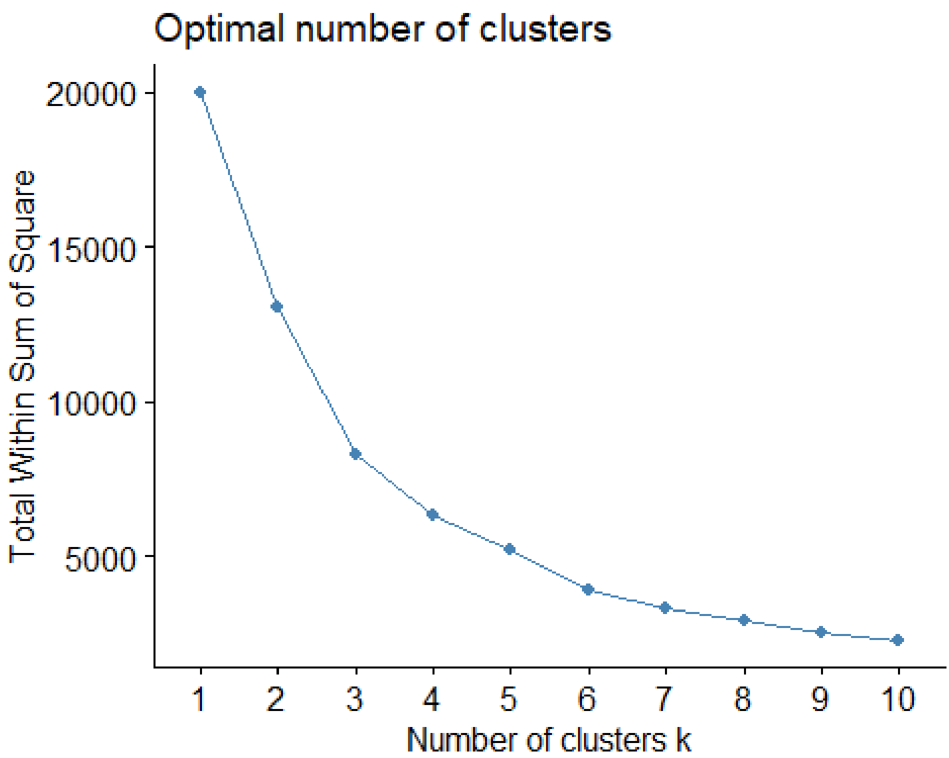


Figure 16: Elbow method for finding optimal k

The number of k was not easily identified so far, additionally to the elbow method the gap statistics has to be used:

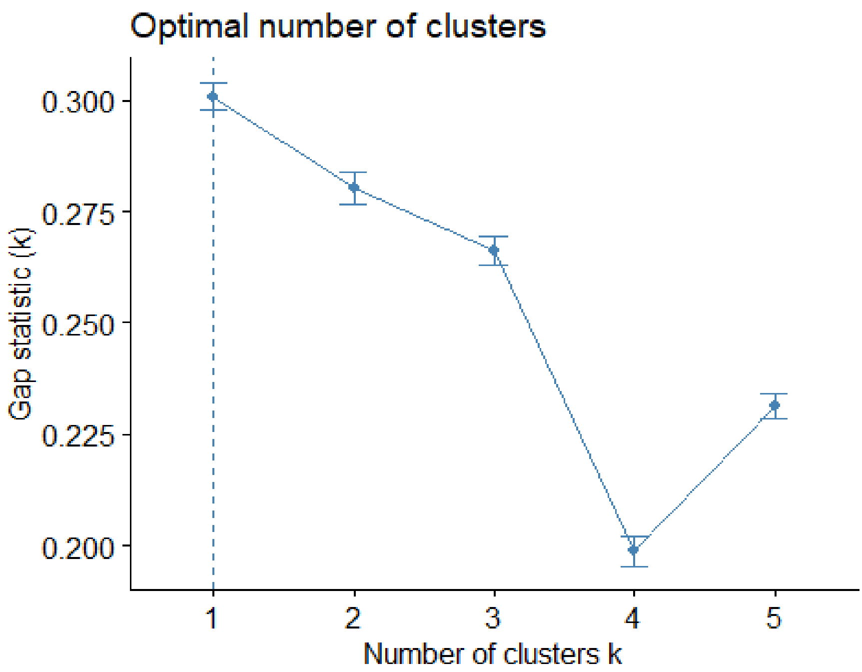


Figure 17: Optimal number of k using the gap statistics

The optimal k is calculated and visible with 4. So within the next steps the clustering will be performed using k = 4. This indicates there are 4 clusters – the same result was achieved with the full set of data. For the number of clusters there is no difference between the sampled data and the full set of data. One cluster represents rd. 25% of each used (sampled-)value.

```
# Perform k-Means Clustering with optimal k = 4
km <- kmeans(df, centers = 4, nstart = 25)

# show results
km## K-means clustering with 4 clusters of sizes 2930, 2319, 2194, 2557
## ## Cluster means:
##   Operating_Hours      logP1
## 1      -0.5165050  0.9126801
## 2       1.0258957  0.4999305
## 3       0.7613733 -0.9752235
## 4      -1.0512699 -0.6988077
## Within cluster sum of squares by cluster:
## [1] 2049.577 1129.078 1323.251 1816.090
##  (between_SS / total_SS =  68.4 %)
```

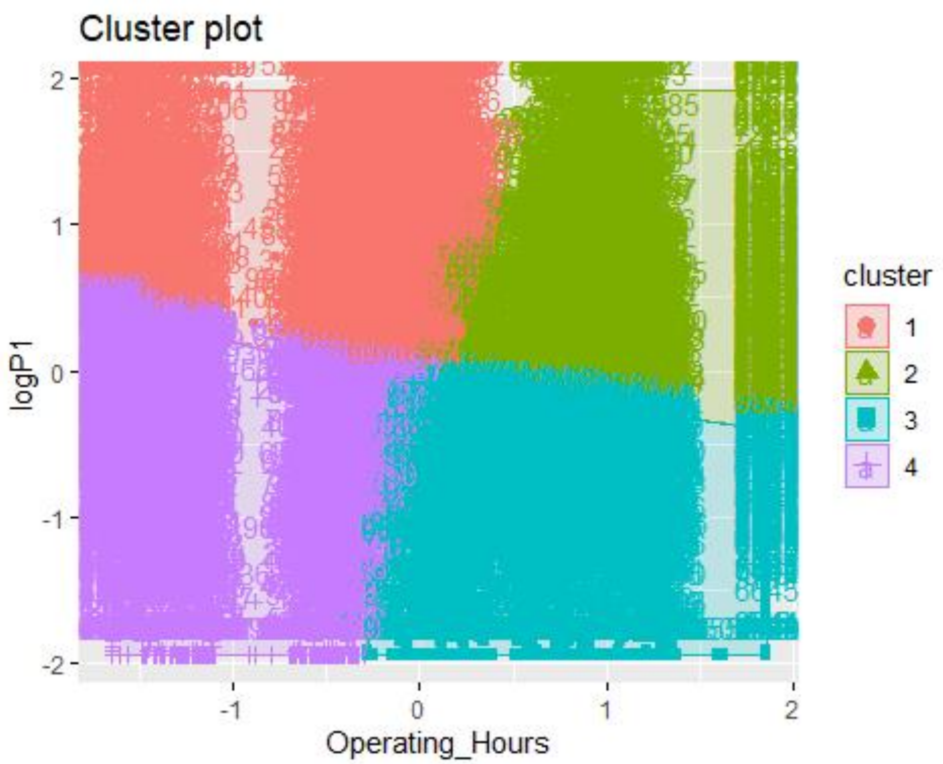


Figure 18: Cluster plot sampled data

Retransforming of the cluster centers is needed in order to evaluate the result. Just the log-transformed demand value will not enable to evaluate the found clusters. The cluster centers can be easily determined using the corresponding command. But for the retransformed centers the earlier calculated values for standard deviation and the mean are needed.

$$RC_k = CV * \sigma + \mu$$

For RC_k is the retransformed center value of each cluster k, CV is the value of the center as determined by the k-means algorithm, σ is the standard deviation and μ is the mean. The values for standard deviation and mean were already calculated within the metric statistics and if the formula is now applied to. For the P1 values which was transformed to a logarithmic value, the calculation has to be reversed using the exponential function.

μ_{OH} = 16807.51 μ_{P1} = 19.24387 σ_{OH} = 6057.133 σ_{P1} = 12.21368

Table 6: Retransformed cluster values

cluster	Operating_Hours	P1
1	13 678.97	37.45
2	23 021.50	27.17
3	21 419.25	11.64
4	10 439.83	13.10

4.1.2 Comparison between sampled value and complete dataset

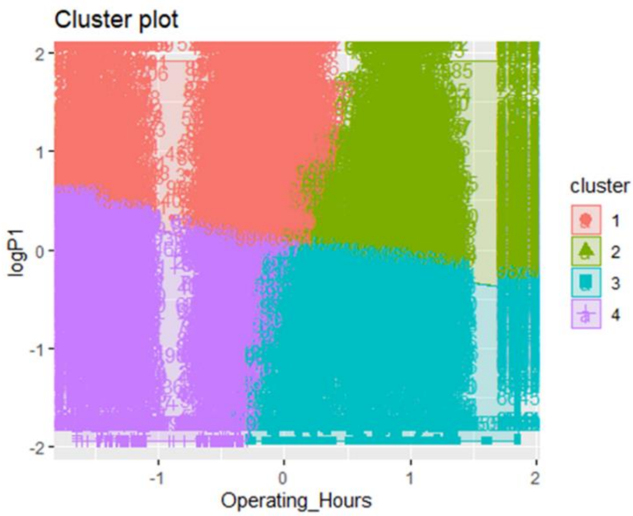


Figure 19: Sampled data set result of k-means analysis

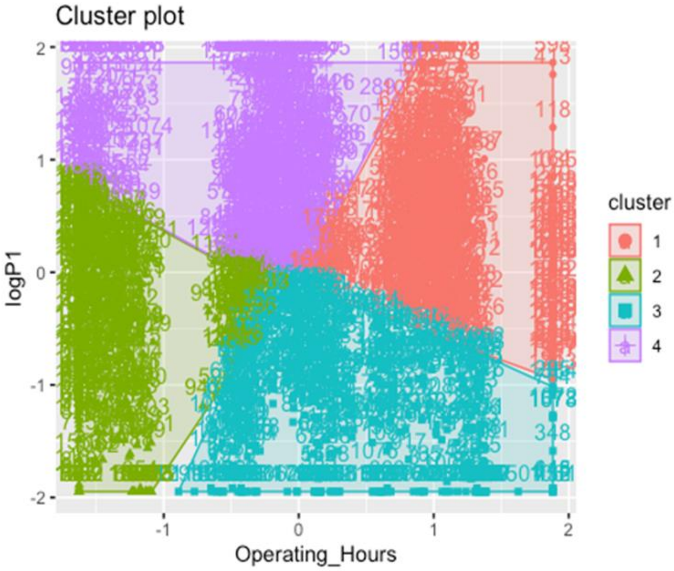


Figure 20: Result of k-means full data set analysis

Although the sampled data (Figure 19) shows different colors for the clusters then the full set of data (Figure 20), both reveal 4 clusters, and the clusters are visually within the same area and close by each other. Within the below table (Table 7), the clusters were adopted for easier comparison. Cluster 4 in the full set is set in comparison to cluster 1 of the sampled set, cluster 1 of the full set and cluster 2 of the sampled set represent the same cluster, cluster 3 is kept for full set and sampled set and cluster 4 of the sampled set is cluster 2 of the full set. Figure 21 shows a graphical view on the cluster centers for sampled and full dataset.

Table 7: Result of transferred cluster values

Sampled data set			Full data set		
Cluster	Operating hours	P1 (in kW)	Cluster	Operating hours	P1 (in kW)
1	13678.97	37.45434	4	14264.07	40.21686
2	23021.50	27.16575	1	22963.40	24.17497
3	21419.25	11.63607	3	18798.65	10.98849
4	10439.83	13.10257	2	8609.02	15.10806

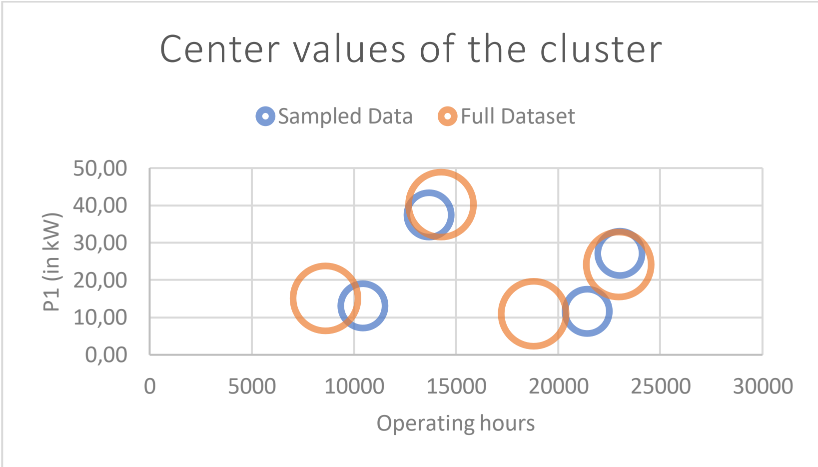


Figure 21: Center values of the clusters (own representation)

An additional test was conducted to assess the significance of the difference between the cluster centers of the full dataset and a sampled dataset using a one-sample t-test. The one-sample t-test determines whether the mean of a single sample differs significantly from a known or hypothesized population mean, that was known in the case used as the full dataset was investigated. The test compares the sample mean (x̄) to a specified value (μ) and assesses the statistical significance of the difference. The formula for the one-sample t-test is:

$$t = \frac{(\bar{x} - \mu)}{\frac{\sigma}{\sqrt{n}}}$$

Where: t is the t-value; x̄ is the sample mean; μ is the population mean; σ is the sample standard deviation; n is the number of observations in the sample. The calculated t-value is then compared to the critical t-value corresponding to the desired significance level and degrees of freedom (df = 3) to determine whether the difference between the sample mean and the hypothesized population mean is statistically significant.

H_0: the cluster center parameters of the full and sampled datasets are the same.

H_1: the cluster center parameters of the full and sampled datasets are not same.

The t-test was concluded at a significance level of 0.05. The standard deviation for $P1_{sample}$ is 12.27 and for the $Operating\_hours_{sample}$ it is 6049.11. The obtained p-values for the cluster center parameters of the P1 variables are 0.06179378, 0.78434675, 0.16137930, and 0.32346609. For the cluster center parameters of the operating hours variable: 0.41184022, 0.14983914, 0.62157135, and 0.06670868. The p-values suggest that there is no strong evidence to reject the null hypothesis, indicating that the cluster center parameters of the full and sampled datasets are the same.

Consequently, based on the results of the one-sample t-tests, it can be concluded that the sampled data does not exhibit significant differences from the cluster center parameters for both the P1 variable and the operating hours variable.

As the hardware demands were huge for the full set of data, the usage of a subset is recommended. Additional tests with sampled data, but more and less than the finally used 10.000 values, were performed but did not show any new finding. The number of clusters stayed equal and the cluster areas as well. For the data given, all together four clusters were identified. One cluster represents rd. 25% of each measured value.

4.1.3 K-means result summary and discussion

Of course, some further details would have to be analyzed before any action could be taken to use the above discovered results. Such actions could be the adjustment of tariffs or even the recommendation of isolation for certain buildings. Mainly the anonymized data would have to be reviewed in the quartiles. As only the heat transfer stations were given and measured, the houses behind those transfer stations have to be identified. For the hospital – for which a higher demand through the whole year could be expected – different measured would have to be taken into account as for any single-family home or a house with several rental flats.

As summary the following results can be taken:

- K-means clustering on large data requires high performing hardware
- Data sampling reveals the same number of clusters
- Even data without private data like building type or addresses clustering can be performed
- Four clusters were identified in a two-dimensional space of heat demand and operating hours
- LOW operating hours with a HIGH demand is the least preferrable option under normal circumstances (see chapter technical details and the inclusion of the Kaposvár sugar factory). Due to the nature of anonymous data it can't be determined how many users/consumers would be counted to the group as the data was received from the heat transfer stations and not end users.
- Longer and stable operating hours are most preferred – with low demand. Also, low demand with shorter operating hours is welcome

Any additional steps like identifying meters and the behind consumers for cluster #1 from sampled data or performing the test for maybe an additional year should be part of a follow-up research. These have to be kept as gaps as the primary goal was to evaluate the possibility of clustering and the identification of the clusters. Follow-up research could also perform structuring / data analysis containing private data (house size, type of use (private, office, ...)).

The result did reveal a large heat demand at low operating hours – given the starting date of the measurements, this is in the early phase of the heating period. The use of waste heat like the heat and methane gas produced by the sugar factory could be enhanced by additional sources as well. That would be a measure which could be performed without any further research on the details of each cluster.

4.2 Results on Q-methods

4.2.1 Unprocessed results of the Q-sorts

The results were transferred into an MS excel file format for further processing (see Table 8 and Table 9). In total twenty-two questionnaires were answered by the participants from which two had to be excluded due to wrong values or a not unique identifiable answer(see Figure 22):

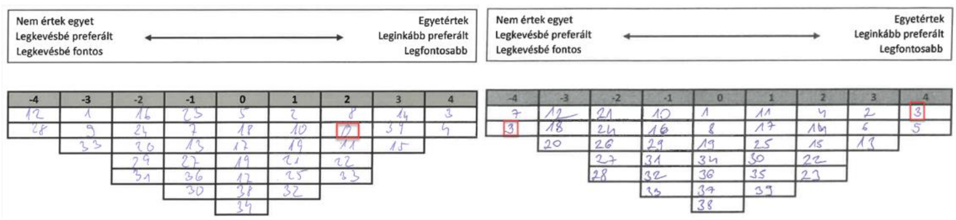


Figure 22: Results with wrong/none-unique answer, self-marked

The answers of the left-hand side picture could probably be calculated (missing value) but might lead to uncertainty. The statement 3 on the right-hand side for the rating of 4 can be uniquely identified – but there is also (see marked value) a value for -4 which could be interpreted as a 3 or a 9. To achieve stable results only the other 21 responses will be used.

The socio-demographic results show a large tendency to employed working owners of houses or flats. More participants consider themselves as having a low demand for heat but during a longer period, than having low demand for short period of time. All participants own an electronical devise and except for one also perform already payments via an electronical device.

Table 8: Summary of the socio-demographic answers

Questions/ Participant	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
P01	Owner	LDSP	employed	R4	single	HM1	Yes	Yes	D1	E6
P02	Owner	LDSP	employed	R3	partnership	HM3	Yes	Yes	D1	E2
P03	Owner	LDLP	employed	R2	partnership	HM3	Yes	Yes	D1	E5
P04	Owner	LDLP	employed	R3	single	HM1	Yes	Yes	D1	E3
P05	Owner	LDLP	employed	R2	partnership	HM2	Yes	Yes	D2	E2
P06	Owner	LDLP	employed	R1	single	HM2	Yes	Yes	D2	E4
P07	Owner	LDLP	employed	R2	single	HM2	Yes	Yes	D2	E3
P08	Owner	None	employed	R2	partnership	HM2	Yes	Yes	D1	E5
P09	Owner	LDLP	employed	R2	partnership	HM3	Yes	Yes	D1	E4
P10	Tenant	LDSP	employed	R2	partnership	HM4	Yes	Yes	D2	E5
P11	Owner	LDLP	employed	R3	partnership	HM3	Yes	Yes	D2	E4
P12	Owner	LDLP	employed	R2	partnership	HM3	Yes	Yes	D2	E2
P13	Owner	LDLP	employed	R2	partnership	HM2	Yes	Yes	D2	E4
P14	Owner	LDLP	employed	R4	partnership	HM2	Yes	Yes	D2	E4
P15	Tenant	LDSP	employed	R4	partnership	HM3	Yes	No	D2	E3
P16	Tenant	LDSP	employed	R1	single	HM3	Yes	Yes	D2	E4
P17	Tenant	LDSP	employed	R3	partnership	HM3	Yes	Yes	D2	E4
P18	Owner	LDSP	Employed	R4	Single	HM1	Yes	Yes	D1	E6
P19	Tenant	LDLP	Employed	R3	partnership	HM2	Yes	Yes	D2	E3
P20	Owner	LDSP	Employed	R2	partnership	HM4	Yes	Yes	D1	E6
P21	Owner	LDLP	Employed	R4	Partnership	HM2	Yes	Yes	D1	E4

Explanation to the table: All questions are marked as Q1 to Q10 in the headline of the Table 8. Question 1 (Q1) simply distinguished between owner and tenant. For question 2 the quartiles as detected in chapter 4.1 of this thesis were taken: usage quartile: HDSP: High demand & for a short period of time (e.g., some hours a day for only a few month) HDLP = High demand & long period LDSP = Low demand & for a short period LDLP = Low demand & long period. Question 3 included only three options – but only one user group was detected by the answers. For question 4 a range of age was supplied: R1: <30 years old R2: 31 to 45 years old; R3: 46 to 60 years old; R4: 61 years to 75 years old; R5: > 75 years old. Q5 was also only a two-dimensional question: single or living in a partnership. Q6 was more open: number of household members: HM1 = 1, HM2 = 2 or 3; HM3 = 4 or 5; HM4 = 6 or 7; HM5 = 8 or more. Here it could happen that people living as a single still had 4 or 5 persons living in the household. Question 7 and 8 were straight forward and simple yes/no questions. Question 9 asked for the preference of usage: D1 = PC, Notebook, Tablet; D2 = Smart Phone; D3 = none of the before. Question 10 regarding the highest degree earned in education (E) was translated to E1 = undergraduate; E2 = skilled/specialist worker; E3 = High school; E4 = Vocational school, with baccalaureate; E5 = Higher education (College,- BSc degree); E6 = University, Master's degree, Bachelor's degree; E7 = PhD and above.

Most of the respondents own the property and live in a partnership. In regard to reached education a broad variety can be observed. All participants are employed. A majority has more than one person living in the household which a broad range of age.

The results of the Q-sort were also transferred into an MS excel file format for further processing (see Table 9):

Table 9: Rated statements by the participants

Statement/ Participant	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21
1	0	3	-4	0	-1	2	0	-1	1	-2	-4	-4	-4	-4	-4	-3	-2	-4	-4	0	-2
2	4	3	4	0	2	4	2	1	4	2	1	-2	-4	3	3	3	3	2	2	0	3
3	2	4	3	0	4	4	2	1	3	3	1	2	2	2	4	4	4	3	3	1	2
4	3	3	3	-2	3	2	-1	2	4	-1	4	3	3	2	2	3	3	3	2	1	1
5	0	1	-1	-2	0	0	3	0	0	1	3	2	4	1	2	2	2	0	1	-1	-1
6	-3	-1	0	-2	1	-1	-2	0	-1	3	0	-1	4	1	1	2	4	2	2	4	0
7	1	0	2	1	1	0	-3	2	0	2	2	1	1	-1	1	2	2	0	1	3	-3
8	0	2	1	-2	-1	0	4	-3	-2	1	1	3	3	-3	0	1	1	1	1	3	-4
9	-4	-2	-3	-4	-2	-4	-4	-3	-4	-4	-3	-3	-3	0	0	-4	-4	-4	-2	-4	2
10	0	-1	1	4	2	1	-4	-1	1	2	-2	0	0	2	-3	-1	-1	1	-1	1	-1
11	1	0	2	1	1	2	1	1	2	2	2	4	2	1	-1	-2	0	1	-2	2	0
12	-1	-1	2	2	-3	-4	-3	-1	-2	0	2	0	1	-4	2	0	2	-3	0	2	2
13	1	4	3	-3	4	3	3	2	0	3	3	4	3	-2	4	4	3	0	2	3	4
14	1	0	1	-3	-1	2	0	2	1	0	2	3	2	-3	-1	3	0	4	1	4	1
15	2	1	1	4	1	1	1	1	1	4	1	2	2	4	2	2	1	2	0	2	3
16	3	2	-2	0	-3	0	-3	-1	0	-3	-3	-2	-3	0	-2	-3	-1	-1	-3	0	-2
17	1	-1	-1	0	-2	1	-1	1	0	0	2	2	-2	-1	-1	0	0	1	0	0	-1
18	-2	1	0	1	2	0	-2	1	-1	0	-2	-1	-2	-3	-1	-3	-2	-1	0	0	-3
19	0	2	2	1	-2	0	2	-1	-1	1	1	-4	1	-1	0	0	0	0	-1	-1	-4
20	2	1	0	1	-1	-1	3	0	-1	1	3	2	1	2	3	1	0	-1	0	2	0
21	2	-2	0	1	-2	-1	0	3	-1	-3	1	1	2	-2	1	-4	-4	0	-3	2	1
22	1	1	1	2	1	0	0	4	3	0	4	1	1	-1	-1	0	0	1	1	-1	1
23	-2	2	-1	-2	1	1	1	-2	-4	0	0	1	1	-2	-4	-2	-3	-2	0	-2	3
24	-3	2	-1	-1	0	-1	2	-1	1	-1	0	1	-1	0	-3	-2	-3	-3	-2	-1	2
25	-1	0	-2	2	0	-2	0	-2	1	-1	0	0	0	-1	-2	-1	-3	2	3	-1	1
26	2	-4	0	-1	0	2	2	3	3	0	0	1	0	0	-3	-2	-1	-1	1	1	-2
27	-4	-4	-4	-1	3	1	-2	4	0	-4	0	-3	-3	-2	-2	-2	-2	-3	-4	-4	-1
28	-2	-3	-3	-1	-4	-2	0	3	-3	-1	0	-3	-1	-2	0	0	1	-2	-1	-3	-2
29	-3	0	-2	-1	-1	-3	-2	0	-3	-2	-3	-2	0	-1	-2	-1	-2	-2	-3	0	-3
30	-2	-2	-2	-1	2	3	1	0	-1	-3	-1	0	0	0	0	1	1	0	3	0	-2
31	-1	-2	-1	2	0	-2	1	-2	2	2	-1	-1	0	3	-2	1	-1	-1	2	1	0
32	0	-3	-1	3	0	-1	-1	0	2	-2	-1	-1	-1	2	1	2	2	1	-1	-2	2
33	3	0	0	3	2	-1	-1	0	-2	-2	-1	0	-2	0	0	1	1	-2	-1	1	0
34	-1	0	-2	0	-2	-2	0	-2	2	4	-1	0	0	1	1	0	2	0	4	-2	1
35	-1	-2	4	-3	3	1	-1	-2	-2	-2	-2	-2	-2	4	3	-1	-2	2	-2	-2	4
36	-2	-1	0	3	-3	-2	-2	-4	-2	-1	-4	0	-1	3	0	-1	1	-2	4	-2	0
37	-1	-1	2	0	-1	-3	1	-3	0	-1	-1	-1	-2	1	1	1	-1	3	0	-3	-1
38	0	-3	-3	-4	-4	-3	-1	-4	-3	1	-2	-2	-1	0	2	0	0	-1	-2	-3	-1
39	4	1	1	2	0	3	4	2	2	1	-2	-1	-1	1	-1	-1	-1	4	-1	-1	0

For the following analysis the software RStudio with the package “qmethod” (Zabala A 2014) was used.

4.2.2 Qualitative data analysis

The analysis begins by employing a multivariate data reduction technique to condense the correlation matrix of Q-sorts into components. Subsequently, a subset of the initial components is chosen, and a mathematically optimized rotation is applied to achieve a more distinct and simplified structure of the data. The size of the eigenvalue is used to determine the number of factors. The factors with the largest eigenvalues have to be retained. But, for example, using the Kaiser Gutman criterion, all the factors with eigenvalues that are greater than 1 should be used according to (Braeken and van Assen 2017). After some experimentation with different sets of factors, the decision for a four-factor model has been taken, although also a 5-factor model would have been applicable (DiStefano, M. Zhu, and Mîndrilã 2009) as visible in Figure 23 for the unrotated factors.

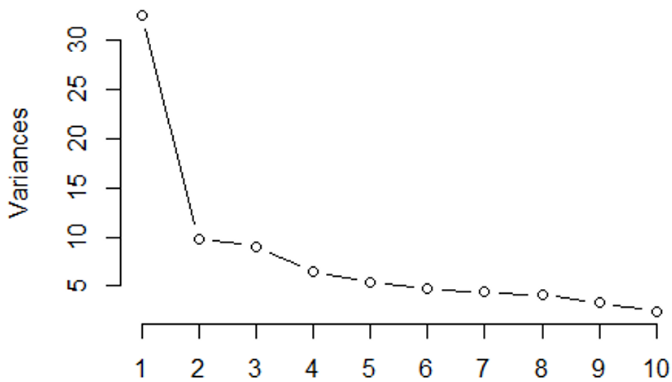


Figure 23: Variances/Eigenvalue of the unrotated factors

Investigation of factor loadings, eigenvalues, explained variance, factor correlations and composite reliability scores suggests the four-factor solution, accounting for 57.4 % of the variance (see Table 10).

Table 10: characteristics of factor loading

	F1	F2	F3	F4
Average reliability coefficient	0.80	0.80	0.80	0.80
Number of loading Q-sorts	9	4	7	1
Eigenvalues	4.1	3.3	3.2	1.5
Percentage of explained variance	19.5	15.6	15.3	7.1
Composite reliability	0.97	0.94	0.97	0.80
Standard error of factor scores	0.16	0.24	0.19	0.45

As the loading of four factors did provide already more than 55% of explanation, this number was taken for further evaluation. The composite reliability as above the recommended value of 0.7 (Hair, Black, William & Anderson 2010) for all four extracted factors. Unrotated factor loadings are often difficult to interpret (Yamamoto and Jennrich 2013). Factor rotation simplifies the loading structure, allowing to interpret the factor loadings more easily. However, one method of rotation may not work best in all cases. Different rotations were tried and “cluster” rotation was used as it created the most interpretable results. There are the two standard criteria for automatic flagging used in Q method analysis (Brown 1980):

1) Q-sorts which factor loading is higher than the threshold for p-value < 0.05, and
2) Q-sorts which square loading is higher than the sum of square loadings of the same Q-sort in all other factors.

PQMethod uses two or more Q sorts with significant loading to 'flag' Q sorts in its automatic mode Threshold for significant loadings at

$$0.01 = 2.58 * (1 + \mathrm{sqrt}(y))$$

y being the number of statements (Zabala A 2014).

The used software package “qmethod” based on ‘R’ provide the following automated flagging for the factor loadings (Table 11):

Table 11: Factor matrix provided by qmethod()

Participants	F1	F2	F3	F4
P01	0.11	0.21	0.52*	0.36
P02	0.03	0.26	0.43*	-0.13
P03	0.48*	0.17	0.33	0.04
P04	0.05	-0.07	-0.01	0.86*
P05	0.23	-0.16	0.68*	-0.25
P06	-0.06	0.04	0.91*	-0.18
P07	0.13	0.22	0.37*	-0.19
P08	-0.31	0.36	0.61*	0.02
P09	0.20	-0.10	0.64*	0.35
P10	0.49*	0.36	-0.01	0.20
P11	0.11	0.72*	0.22	-0.15
P12	0.10	0.66*	0.24	-0.06
P13	0.19	0.78*	-0.06	-0.11
P14	0.82*	-0.58	0.03	0.26
P15	0.81*	0.06	-0.23	-0.21
P16	0.75*	0.28	-0.06	-0.07
P17	0.70*	0.32	-0.17	0.06
P18	0.54*	-0.01	0.37	0.10
P19	0.63*	0.23	-0.13	0.03
P20	-0.05	0.75*	0.16	0.21
P21	0.54*	-0.30	0.21	-0.29

Values calculated after factor rotation; marked values (*) indicating a defining sort (a significant loading) automatically by “qmethod”.

Table 12 contains the four factors or main perspectives with the factor scores which represent the strength of agreement or disagreement with all statements.

Table 12: Factor scores for each of the four extracted factors

Statements	F1	F2	F3	F4
1	-4	-4	1	0
2	4	-2	4	0
3	4	1	4	0
4	3	3	3	-2
5	2	2	0	-2
6	2	2	-1	-2
7	1	2	1	1
8	-1	3	-1	-2
9	-2	-4	-4	-4
10	-1	0	1	4
11	0	3	2	1
12	0	1	-4	2
13	3	4	3	-3
14	0	4	2	-3
15	3	2	2	4
16	-2	-3	0	0
17	-1	1	1	0
18	-3	-1	0	1
19	-1	0	0	1
20	2	2	-1	1
21	-2	1	-1	1
22	0	1	1	2
23	-3	0	0	-2
24	-2	0	-1	-1
25	-1	0	-2	2
26	-2	1	2	-1
27	-4	-3	1	-1
28	-1	-2	-2	-1
29	-3	-1	-3	-1
30	1	0	2	-1
31	1	0	-2	2
32	1	-1	-1	3
33	0	-1	0	3
34	2	-1	-2	0
35	2	-3	0	-3
36	1	-2	-3	3
37	1	-2	-2	0
38	0	-2	-3	-4
39	0	-1	3	2

For example, perspective (factor) F4 is in strong disagreement with statement 8 (scoring -3), whereas F1 rather takes the opposite opinion (scoring 3) and perspectives F2 and F3 show an ambivalent opinion (scoring 1 and -1, respectively). The qualitative analysis and interpretation follow a narrative approach, aiming to establish connections between themes and statements to form a comprehensive understanding of participants' perspectives. The findings are supplemented with descriptive data on participants' factor loadings, which include information on age, education, and the number of household members.

The analysis assumes that the relationship (correlation) between variables is due to a set of underlying factors (latent variables) that are being measured by the variables. The rotation method used in Q-methodology is typically by-person factor analysis or cluster rotation. This approach groups participants based on their factor scores and performs separate factor analyses for each cluster or group, allowing for the interpretation of factors based on individual perspectives. (DTREG 2022). Table 13 displays a table showing the correlations the rotated factor z-scores:

Table 13: Correlations between rotated factor Z-Scores

	F1	F2	F3	F4
F1	1.0000	0.4537	0.3586	0.0661
F2	0.4537	1.0000	0.3796	-0.0098
F3	0.3586	0.3796	1.0000	0.0432
F4	0.0661	-0.0098	0.0432	1.0000

Based on the matrix, following is observed: There is a positive correlation between F1 and F2, with a correlation coefficient of 0.4537. F1 and F3 also have a positive correlation, with a coefficient of 0.3586. F2 and F3 have a moderate positive correlation, with a coefficient of 0.3796. The correlation between F1 and F4 is weakly positive, with a coefficient of 0.0661. F2 and F4 have a weak negative correlation, with a coefficient of -0.0098. F3 and F4 also have a weak positive correlation, with a coefficient of 0.0432. The correlation reflects the relationship between each participant's rankings or Q-sorts and the overall pattern or consensus among the participants. It helps identify the degree of agreement or disagreement among participants' subjective viewpoints. The correlation between participants and their sorting helps in interpreting the factor analysis results, understanding the variation in viewpoints, and identifying the underlying dimensions or factors that capture the shared perspectives.

Within the Table 14 below, the correlation between the sorts is analyzed. The correlation coefficient ranges from -1 to 1, where values closer to 1 indicate a strong positive correlation, values closer to -1 indicate a strong negative correlation, and values close to 0 indicate a weak or no correlation. The correlation matrix revealed several positive relationships among participants. Pairs such as P01-P03, P01-P18, P03-P18, P08-P11, P09-P14, and P10-P14 displayed correlations of 0.5 or higher, indicating similarities or agreements between their variables (highlighted with gray in Table 14). However, no correlations below or equal to -0.5 were found. These findings suggest that certain participants share patterns or similarities, warranting further investigation and contextual analysis.

Table 14: Correlation matrix between Sorts

	P01	P02	P03	P04	P05	P06	P07
P01	1.0000	0.4000	0.5000	0.2824	0.1471	0.5176	0.3588
P02	0.4000	1.0000	0.4118	-0.0059	0.3118	0.4412	0.3941
P03	0.5000	0.4118	1.0000	0.1235	0.4588	0.4059	0.2529
P04	0.2824	-0.0059	0.1235	1.0000	0.0118	-0.0471	-0.0941
P05	0.1471	0.3118	0.4588	0.0118	1.0000	0.6471	0.1529
P06	0.5176	0.4412	0.4059	-0.0471	0.6471	1.0000	0.4294
P07	0.3588	0.3941	0.2529	-0.0941	0.1529	0.4294	1.0000
P08	0.3412	0.0059	0.1529	0.0882	0.3706	0.4941	0.1176
P09	0.5412	0.2765	0.3529	0.3059	0.4059	0.5471	0.2941
P10	0.3000	0.3706	0.4529	0.1647	0.1824	0.2412	0.3647
P11	0.3588	0.2824	0.4706	-0.0588	0.2706	0.2765	0.3765
P12	0.3647	0.3176	0.4412	0.0059	0.3118	0.3294	0.3647
P13	0.1118	0.2294	0.3588	-0.0882	0.1765	0.1059	0.3235
P14	0.2294	-0.0529	0.3235	0.2412	0.2529	0.1118	0.0824
P15	0.2941	0.1647	0.5353	-0.1412	0.1765	0.0588	0.1824
P16	0.3353	0.3118	0.5059	-0.0353	0.3647	0.3059	0.3294
P17	0.3294	0.2706	0.4353	0.0412	0.2294	0.2765	0.1588
P18	0.5000	0.2118	0.6118	0.0765	0.3353	0.4471	0.3176
P19	0.1000	0.2000	0.3353	0.0824	0.2000	0.1471	0.2471
P20	0.4176	0.3412	0.4588	0.1235	0.2412	0.3235	0.1529
P21	0.1118	0.1529	0.3353	-0.0471	0.3353	0.1353	0.0706

	P08	P09	P10	P11	P12	P13	P14
P01	0.3412	0.5412	0.3000	0.3588	0.3647	0.1118	0.2294
P02	0.0059	0.2765	0.3706	0.2824	0.3176	0.2294	-0.0529
P03	0.1529	0.3529	0.4529	0.4706	0.4412	0.3588	0.3235
P04	0.0882	0.3059	0.1647	-0.0588	0.0059	-0.0882	0.2412
P05	0.3706	0.4059	0.1824	0.2706	0.3118	0.1765	0.2529
P06	0.4941	0.5471	0.2412	0.2765	0.3294	0.1059	0.1118
P07	0.1176	0.2941	0.3647	0.3765	0.3647	0.3235	0.0824
P08	1.0000	0.4235	0.0118	0.5235	0.2353	0.1824	-0.1941
P09	0.4235	1.0000	0.3529	0.3471	0.2941	0.0824	0.3353
P10	0.0118	0.3529	1.0000	0.3765	0.4235	0.5118	0.2824
P11	0.5235	0.3471	0.3765	1.0000	0.6706	0.6412	-0.0824
P12	0.2353	0.2941	0.4235	0.6706	1.0000	0.6882	0.0529
P13	0.1824	0.0824	0.5118	0.6412	0.6882	1.0000	0.0000
P14	-0.1941	0.3353	0.2824	-0.0824	0.0529	0.0000	1.0000
P15	0.0294	0.0412	0.3412	0.4000	0.2412	0.3176	0.3588
P16	0.1471	0.3412	0.5471	0.5353	0.4176	0.4706	0.2941
P17	0.1588	0.3235	0.5706	0.4294	0.3235	0.4294	0.2529
P18	0.1941	0.5176	0.4235	0.3588	0.4000	0.3765	0.3824
P19	-0.0941	0.3118	0.5176	0.3059	0.4294	0.4118	0.2765
P20	0.2824	0.2294	0.4588	0.4471	0.6000	0.6000	-0.1118
P21	0.0059	0.1824	0.1412	0.2059	0.2647	0.0706	0.3588

	P15	P16	P17	P18	P19	P20	P21
P01	0.2941	0.3353	0.3294	0.5000	0.1000	0.4176	0.1118
P02	0.1647	0.3118	0.2706	0.2118	0.2000	0.3412	0.1529
P03	0.5353	0.5059	0.4353	0.6118	0.3353	0.4588	0.3353
P04	-0.1412	-0.0353	0.0412	0.0765	0.0824	0.1235	-0.0471
P05	0.1765	0.3647	0.2294	0.3353	0.2000	0.2412	0.3353
P06	0.0588	0.3059	0.2765	0.4471	0.1471	0.3235	0.1353
P07	0.1824	0.3294	0.1588	0.3176	0.2471	0.1529	0.0706
P08	0.0294	0.1471	0.1588	0.1941	-0.0941	0.2824	0.0059
P09	0.0412	0.3412	0.3235	0.5176	0.3118	0.2294	0.1824
P10	0.3412	0.5471	0.5706	0.4235	0.5176	0.4588	0.1412
P11	0.4000	0.5353	0.4294	0.3588	0.3059	0.4471	0.2059
P12	0.2412	0.4176	0.3235	0.4000	0.4294	0.6000	0.2647
P13	0.3176	0.4706	0.4294	0.3765	0.4118	0.6000	0.0706
P14	0.3588	0.2941	0.2529	0.3824	0.2765	-0.1118	0.3588
P15	1.0000	0.6353	0.6353	0.3647	0.3412	0.1706	0.3765
P16	0.6353	1.0000	0.8353	0.5529	0.5941	0.4000	0.2235
P17	0.6353	0.8353	1.0000	0.3765	0.5824	0.3882	0.0882
P18	0.3647	0.5529	0.3765	1.0000	0.4059	0.3176	0.1882
P19	0.3412	0.5941	0.5824	0.4059	1.0000	0.2471	0.1765
P20	0.1706	0.4000	0.3882	0.3176	0.2471	1.0000	-0.0235
P21	0.3765	0.2235	0.0882	0.1882	0.1765	-0.0235	1.0000

4.2.3 Quantitative data analysis

For statement 14 “District heating supports (large) investments in renewable energies.” all factors distinguish. But for statement 19: “An app should always have several divisions integrated (district heating, electricity, water, gas, ...)” consensus is achieved. The implications will be drawn later. To investigate how the perspectives (factors) differ within this thesis the analysis will focus on the distinguishing statements.

Factor 1: Space Savings and Convenience: The Appeal of District Heating

Factor 1 represents a multifaceted perspective on district heating, highlighting various aspects that influence customer preferences. Space savings emerge as the least important driver for customers. The simplicity of operation and the utilization of combined heat and power (CHP) technology are valued. Moreover, customers emphasize neutral stance towards government intervention, regulations, and support for renewable energy investments in the district heating sector. Based on the ratings provided, it can be concluded that customers generally perceive district heating as offering price stability (rated with 2) and express a desire for the option to choose between different district heating providers (also rated with 2). These ratings suggest that customers see value in the stability of prices provided by district heating systems and value the ability to have choices and options when it comes to selecting their district heating service providers.

Characterization: Factor 1 encompasses customers' perspectives on district heating, emphasizing the appeal of space savings, convenience, environmental considerations, and the role of government intervention and support.

Table 15: Table distinguishing statements for factor 1 ($p < 0.05$)

Statement	Q-SV	Z-Score
1. It is mainly the space savings that encourage many customers to buy district heating.	-4	-2.26
12. The state should intervene even more in the energy market and make and enforce regulations regarding the technologies used.	0	-0.11
14. District heating offers price stability.	2	0.04
21. District heating should be supported and promoted by the government.	1	-1.08
22. Customers should have the option to choose between different district heating providers.	2	-0.10
35. District heating is a sustainable and environmentally friendly heating solution.	1	0.93
38. District heating should be the primary heating solution for new residential buildings.	0	0.00

Factor 2: Advantages of District Heating

For factor 2, the highest-rated statements are "Security of supply is always the top priority in a district heating system" and "District heating supports (large) investments in renewable energies." These ratings indicate that customers highly value the security of energy supply provided by district heating systems and recognize the positive impact of district heating on promoting investments in renewable energy sources. The high ratings suggest that customers perceive these aspects as important and desirable qualities of district heating systems. The least important part – similar to factor 1 – is the space saving aspect.

Characterization: Factor 2 emphasizes the advantages of district heating, highlighting its accessibility and modern features. The demand for district heating in sparsely populated areas underscores the importance of accessibility and local resources. Price stability is another attractive feature, providing customers with confidence in their heating costs. Additionally, the integration of graphical app is not seen as significantly in terms of convenience and energy usage management.

Table 16: Table distinguishing statements for factor 2 ($p < 0.05$)

Statement	Q-SV	Z-Score
1. It is mainly the space savings that encourage many customers to buy district heating.	-4	-1.66
8. District heating should also be available in sparsely populated areas. (Preferably from a local resource.)	3	1.43
14. District heating offers price stability.	4	1.52
39. Consumption and billing via app (if available) must be displayed graphically.	-1	-0.69

Factor 3: Energy Market Dilemma: Convenience in District Heating

Factor 3 emphasizes the importance of efficiency, sustainability, and market dynamics in the context of district heating. The integration of smartphone apps that offer cost-effective rate suggestions based on consumption provides convenience for customers. Furthermore, the willingness to embrace changes in habits and adopt energy-conscious behaviors contributes to the sustainability of district heating. However, there may be differing opinions regarding the extent of state intervention in the energy market and the enforcement of regulations. Balancing market dynamics with sustainable practices remains a key consideration for the successful implementation of district heating systems.

Characterization: Factor 3 The integration of smartphone apps that offer personalized rate suggestions based on consumption caters to the convenience and preferences of customers. Additionally, the willingness to embrace changes in habits and adopt energy-conscious behaviors reflects a positive attitude towards sustainability. However, differing perspectives may arise concerning the optimal level of state intervention and the enforcement of regulations in the energy market. Striking a balance between market dynamics and sustainable practices is vital for the successful implementation of district heating systems.

Table 17: Table distinguishing statements for factor 3 ($p < 0.05$)

Statement	Q-SV	Z-Score
5. Combined heat and power (CHP) make optimum use of fuels.	0	0.11
12. The state should intervene even more in the energy market and make and enforce regulations regarding the technologies used.	-4	-1.76
14. District heating supports (large) investments in renewable energies.	2	0.75
26. A smartphone app must offer/suggest the cheapest rate for me depending on consumption.	2	-0.97
30. I would be willing to change my habits (turn off the heating on vacation, ventilate now and then instead of constantly ventilate, …).	2	0.87
35. The state should let the market regulate prices (market economy).	0	0.15

Factor 4: Debunking Myths: District Heating's Value and Sustainability

Factor 4 examines the aspects of simplicity, security, and sustainability in district heating systems based on varying viewpoints. While opinions differ on the simplicity of district heating for housing temperature control, there is a strong rating on governmental influence in investments and the need for insulation. District heating is recognized as a catalyst for investments in renewable energies.

Characterization: Factor 4 explores the dimensions of simplicity, security, and sustainability in district heating systems. While opinions vary on the simplicity of housing temperature control, the importance of security of supply is widely acknowledged. District heating is not seen as supportive of investments in renewable energies. Individuals express readiness to adopt energy-saving measures suggested by apps. Additionally, aligning energy taxation with climate protection goals is considered significant. These perspectives highlight the multifaceted nature of district heating systems.

Table 18: Table distinguishing statements for factor 4 ($p < 0.05$)

Statement	Q-SV	Z-Score
4. Simple operation: In terms of comfort, district heating is probably the simplest way of housing temperature control.	-2	-0.95
5. Combined heat and power (CHP) make optimum use of fuels.	0	-0.95
10. No more research and investment should be made in district heating.	1	1.89
13. Security of supply is always the top priority in a district heating system.	-3	-1.42
14. District heating supports (large) investments in renewable energies.	-3	-1.42
25. If an app suggests savings measures, e.g., home insulation, different thermostat, ... I would take the appropriate steps to implement the recommendations.	2	0.95
33. The taxation of energy products in Hungary / in the EU must be more closely aligned with climate protection aspects.	3	1.42

Statement 29: “I would rather heat with another energy source than with district heating.” has not reached consensus with all factors but is rated negatively by all 4 factors. is consistently rated negatively across all four factors, it indicates a general disagreement or preference against using district heating as an energy source for heating. The negative ratings across the factors suggest that individuals tend to agree with or have a higher preference for district heating compared to other energy sources. It could mean district heating is well supported.

Finding common socio-demographic factors for each factor group was not possible. For example, factor group 3 included participant 5, 6, 7, 8 and 9 – all having different educational degree, different preferences on used devices and partnership. The age range also differs. Only the persons living in the household (HM2 or HM3) were close by. But with P01 also being part of F3 that did not match any more. Therefore, the factor groups represent different attitudes towards sustainability and district heating – but there is no unified socio-demographic characteristic identified.

4.2.4 Summary and discussion in consumer opinion

The four identified factors were given a comprehensible name to identify the attitude towards district heating in general and sustainability in detail together with price sensitiveness. The study indicates that government intervention, price stability, sustainability, provider choice, and integration in building standards play significant roles in shaping attitudes towards district heating. Consumers are attracted to district heating due to its various advantages, and there is a strong consensus in regards to state regulation and support. Price stability and sustainability are highly valued, and consumers appreciate the option to choose from multiple providers.

Furthermore, integrating district heating in new building standards is seen as essential for maximizing its benefits and widespread adoption.

Overall, these findings shed light on the factors influencing the perception and acceptance of district heating. Policymakers and industry stakeholders can leverage these insights to develop effective strategies, regulations, and marketing approaches to promote district heating as an attractive and sustainable energy solution.

The necessity for a paper bill – is desired by only one factor. As this attitude was a differentiator to factor group three – the option to access a paper bill will not be the main driving and innovation feature for any future solution. With a rather strong attitude towards what is needed for a smartphone app, it should be possible to convince this group also to overcome the need for printed bills. The trust in smartphone apps seems high already as all (except one) participants indicated to use electronical devices for banking already. But the consumers would not trust any app to regulate the complete household as the reaction to statement 23 (I would trust an app enough to let it regulate my heating on its own) revealed. Here the factors provided a very diverse response. While smartphones still need to get more sustainable (Tu, X.-Y. Zhang, and Huang 2018) themselves, they are one key to more digital information and behavioral change. Smartphones offer the potential for enhanced access to advice and sustainable behavior by enabling virtual appointments, particularly in regions where individuals would typically face challenges in accessing expert guidance.

The answer to statement 7 (More digitalization leads to more environmentally friendly generation processes (e.g., through education, visualization)) was throughout slightly positive. The consumers agree digitalization helps to grow environment friendly generation of heat.

Digitalization is also able to provide the base figures and transfer the data when needed. Here the socio-demographic part indicated the end users are technically able to work with electronical devices and some even prefer the smartphone already over a standard PC or laptop. With that the finding (Krog et al. 2020) could be confirmed and validated also for East European consumers: the integrated usage of several divisions is a must for any smart meter app.

Generally, the attitude towards district heating is positive. It’s seen as key to more sustainability. Although there was no full consensus on statement 9 (No more research and investment should be made in district heating) this statement was rated strong negative. Conversely, this means the consumers would support and approve further research in the area of district heating.

5 Conclusion, recommendations, and limitations

The objective is to elucidate the findings of the dissertation, establish connections with the existing literature, examine limitations, and propose avenues for future research.

The challenge of heat decarbonization requires an immediate policy and research response. The best documented approach to cluster a large set of data, especially measured heat demand or heat consumption according to the reviewed literature is K-Means; all other algorithms were only used by single papers. When checking the geographical location of the data origin of the reviewed literature, the coverage of Eastern European countries is nonexistent.

The K-Means algorithm is the most widely used approach for clustering data in the reviewed literature and is the basis for almost all of the reviewed work. However, other methods, such as the Pearson Correlation Coefficient, may be more appropriate for specific interests. There is a lack of research on data from Eastern European countries, which may be due to language dependencies or other factors. Information is key to the success of new pricing models and to achieving behavior change and can be provided through email-based information systems or mobile apps. Visualization in a graphical manner, such as through the use of signal colors, can also contribute to higher acceptance of new technology. Common methodologies used in the research include interviews and questionnaires, but the Q-method is not used. Research gaps identified in the literature review include a lack of research on data from Eastern European countries and on regulated pricing methods, as well as a lack of research on the impact of personalized data on behavior change.

The Q-method revealed the consumers do not require to be informed about the cheapest rate (statement 26 - A smartphone app must offer/suggest the cheapest rate for me depending on consumption) but have a rather diverse attitude. Statement 35 (The state should let the market regulate prices (market economy)) was positively rated by factor group 1, while the other three factors rated that as neutral or negatively. Given the fact that even Germany implemented a state regulated price lately (Website of the Federal Government | Bundesregierung 2022) (or will implement the same in early 2023), this finding was rather unexpected.

Within the reviewed literature that was already concluded: New pricing mechanisms and pricing models are more likely to be accepted if the information is transparent and the end users can significantly influence their price by own behavior.

The results of the data analysis suggest that K-means clustering can be performed on large data sets, and that data sampling can reveal the same number of clusters as the full data set. Four clusters were identified in a two-dimensional space, and the results indicate that longer and stable operating hours with low demand are the most preferred, followed by low demand with shorter operating hours. The least preferred option is low operating hours with high demand. Further research could include identifying meters and the consumers behind them for cluster #1 from sampled data or performing the analysis for an additional year. Follow-up research could also include private data such as house size and type of use. The results also suggest that the use of waste heat, such as that produced by the sugar factory, could be enhanced by additional sources.

Four factors were identified in the research to understand consumer attitudes towards district heating and sustainability. The majority of consumers were convinced that district heating is the best choice for urban energy systems and sustainable consumption and were willing to use intelligent artificial agents and smart meter apps to manage their heating. However, there was little interest in accessing paper bills and some hesitation about trusting an app to regulate the household's heating on its own. Digitalization was seen as a way to promote environmentally friendly heat generation through education and visualization, and most consumers were technically proficient in using electronic devices. Overall, the attitude towards district heating was positive, with a majority supporting further research and investment in the area.

5.1 Contribution and practical implications

5.1.1 Developing a dynamic pricing method

Actually, Hungary has state regulated prices – and many other European countries (e.g. Germany) are introducing state regulated prices as well. Nevertheless, the results of the previous chapters of this dissertation can be used to describe a pricing model. The best pricing method for district heating varies depending on several factors (IEA 2022a), including the characteristics of the system, the customer base, and the regulatory framework. Ultimately, the best pricing method for district heating will depend on the specific needs and goals of the system and its customers. Conservation pricing might work for electricity, but heating periods depend on outside temperature – when it’s warm outside it rare to find consumers needing a lot of heat. Dynamic pricing is something which would have to be explored deeper based on the findings of this dissertation, but the idea development shall be started.

Dynamic pricing is a pricing method that adjusts the price of a product or service in real-time based on market conditions and supply and demand. In the context of district heating, dynamic pricing uses advanced metering and billing systems to charge customers for the energy they consume based on the prevailing market price for energy at the time of consumption.

Dynamic pricing can provide several benefits for district heating systems, including (Li et al. 2019):

1. Improved efficiency: By charging customers based on real-time market conditions, dynamic pricing can encourage customers to reduce their energy consumption during peak demand periods, when energy prices are highest.
2. Increased revenue: By charging higher prices during peak demand periods, district heating systems can increase their revenue and better recover the cost of providing energy.
3. Better alignment of supply and demand: By adjusting prices based on market conditions, dynamic pricing can help to better match the supply of energy with customer demand, reducing the risk of over- or under-supply.
4. Dynamic pricing can also provide benefits for customers, such as increased transparency and control over their energy costs. However, it may also be seen as unfair or confusing by some customers, particularly if they are not familiar with how dynamic pricing works.

Overall, dynamic pricing can be an effective pricing method for district heating, but it requires careful implementation and communication to ensure that it is well understood and accepted by customers. Dynamic pricing involves adjusting the price of energy based on real-time market conditions, such as the supply and demand for energy, the availability of alternative energy sources, and the cost of producing energy.

Based on the results of the Q-method analysis, it seems that consumers have a positive attitude towards district heating and see it as a key to more sustainability. They are prepared for the use of smart meters and intelligent artificial agents, as long as these devices provide an integrated experience for all utilities and allow for comparison to average consumption. Although the cheapest price is not a primary concern, consumers want to be able to compare their own consumption to others. Digitalization is seen as a way to increase environmentally friendly generation processes and access to expert advice.

The pricing method would also need to be accessible through smartphones and provide an integrated experience for all utilities, as consumers have a high level of trust in these devices and prefer them over traditional paper bills. Finally, the pricing method would need to take into account the need for further research and investment in district heating, as consumers generally support this idea.

The q-method findings indicate that consumers:

- Mostly have a positive attitude towards district heating as a key to sustainability
- Prefer smart meter devices over traditional paper bills and some are willing to pay more for environmentally friendly options

With this in mind, the dynamic pricing method could be designed as follows (also based on the result of the K-Means clusters analysis):

A. High demand for a long period of time (cluster 1)

During peak periods of high demand, prices could be adjusted to reflect the increased usage

For example, if the average price per unit of energy during off-peak periods is €0.10, during peak periods of high demand the price could increase to €0.15

Additionally, the price could be made transparent by showing users how it compares to the average consumption of a selected comparison group

This would allow users to adopt more sustainable habits and make more informed decisions about their energy usage

B. High demand for a short period of time (cluster 2)

During short periods of high demand, prices could be adjusted to reflect the increased usage

For example, if the average price per unit of energy during off-peak periods is €0.10, during a short period of high demand the price could increase to €0.12

This would encourage users to shift their energy usage to off-peak periods, which would help to balance the demand on the system

C. Low demand for a long period of time (cluster 3)

During low demand periods, prices could be adjusted to reflect the reduced usage

For example, if the average price per unit of energy during off-peak periods is €0.10, during low demand periods the price could decrease to €0.08

This would encourage users to continue their energy usage during these periods, which would help to balance the demand on the system

D. Low demand for a short period of time (cluster 4)

During short periods of low demand, prices could be adjusted to reflect the reduced usage

For example, if the average price per unit of energy during off-peak periods is €0.10, during a short period of low demand the price could decrease to €0.09

Short periods of time mean in this context only a few hours per day resp. week while a long period would reflect in several consecutive hours(or even days). This would encourage users to shift their energy usage to the corresponding periods, which would help to balance the demand on the system.

It's important to note that this is just one example, and the actual prices and pricing structures used in dynamic pricing for district heating may differ in practice, depending on the specific needs and goals of the system and its customers, as well as the regulatory framework in place. Dynamic pricing models can vary widely depending on the specific needs and goals of the district heating system and its customers, as well as the regulatory framework in place. The example above is just one possible scenario, and the actual prices and pricing structures used in dynamic pricing may differ in practice.

5.1.2 Technical consumer information enablement

A large common part for infrastructural means with a lower part for direct consumption provide less potential for an actual behavior change. Information is the key to success to any new model, but even to achieve behavior change, information is a key factor. While in earlier years email-based information system were tested and provided to end users, more and more apps running on a mobile taking over the information part. Within several publications it was proven (Atilla Wohllebe et al. 2021b) that even early in grade school the usage and consumption of smart phones and Apps is already very common. The growing generation might not even remember what a paper bill is or how that was sent to the end user. In retail business (Atilla Wohllebe et al. 2021a) the usage of push notifications a common and broadly researched. The Technology Acceptance Model (TAM) was adopted for the retail use case. But for consumer apps in regards to sustainable behavior so far nothing comparable was designed. Especially with a combination of weather data and smart home applications more sustainable awareness and consumption can be reached and would be welcomed and used by the consumers.

5.2 Limitations and deviations

Hungary faced the possibility of entering a technical recession by the end of 2022 due to high inflation and elevated interest rates, reaching levels not seen in nearly three decades. Historical data indicates that previous economic crises in Hungary coincided with recessions following the global financial crisis in 2008 and the Covid-19 pandemic in 2020 (Lepcha 2022).

Comparing the severity of the crises, the economic downturn in Hungary was more pronounced in 2008-2009 than in 2020. During that period, Hungary experienced a 6.6% decline in annual GDP growth, whereas in 2020, the decline was 4.5%. However, the subsequent economic recovery from the 2020 lows was stronger, with Hungary achieving a GDP growth of 7.1% in 2021. In contrast, the post-2008-2009 crisis recovery in Hungary was slow, with GDP growth rates of 1.1% in 2010 and 1.9% in 2011. Economists expressed concerns that the surge in food and fuel costs could lead to demand destruction and prolong the economic crisis in Hungary (Lepcha 2022).

This is important as a deeper recession means sustainability policies will receive less support (Anbumozhi and Bauer 2010), so what options do governments, business leaders and citizens have? In 2011 a similar question was raised leaving the 2009 recession due to the Lehman bankruptcy. With the Rio+20 sustainable development conference 2012 a global economic downturn will gut prospects of agreement? According to the statement, green office buildings were primarily appealing to a restricted set of tenants, such as government agencies and prestigious companies aiming to establish their prominent headquarters. Unless there were additional regulations in place, including stricter building standards, increased energy costs, or other market forces, this market would continue to be limited in scope. Factors related to sustainability, such as water conservation, energy efficiency, and labor-friendly design, were not considered when making decisions about office spaces.– all already happened 2011 and 2012.

Once again, in the year 2023, governments, business leaders, NGOs, and citizens are presented with three overarching strategic choices. The first option is to acknowledge these presumed realities and adopt a wait-and-see approach. This entails pursuing strategies that are competitive and cost-effective in the present market conditions while accepting the potential sustainability risks and higher costs in the future. This also entails downplaying the significance of the upcoming United Nations Conference on Sustainable Development and reserving resources for future opportunities, whether it be a breakthrough in renewable energy costs or a series of disruptive climate events.

The second choice is to concentrate efforts on supporting one or two initiatives that have the potential to catalyze widespread positive transformation. Examples include the UN Environment Programme's Green Economy initiative or the Corporate Sustainability Reporting Coalition's advocacy. While financial commitments were made during the 2022 conference to aid countries affected by climate change, the allocated funds appear relatively modest compared to military investments. Globally there are more investments in military technique and research than in sustainable consumption and production.

The third option involves revitalizing and intensifying efforts towards sustainability on a global scale, even if it requires increased monetary supply. This approach may resemble the economic stimulus incentives implemented during the 2008/9 financial crisis in terms of objectives, but with clearer definitions of drivers and targets. Rather than focusing solely on recession prevention, the emphasis would be on preventing a potentially enduring global economic downturn by fostering innovation and enhancing competitiveness. Additionally, this approach would prioritize the security of food, water, and energy in a world where competition for limited resources and heightened climate variability are the new "normal" business environment. This is a concrete agenda where many businesses have already invested heavily and shown promising results (Nick Robins 2009).

As over half of the global population resides in urban areas and construction contributes approximately 40% of greenhouse gas emissions, the real estate sector emerges as a pivotal battleground in the pursuit of sustainability for humanity. The challenges for (smart) urban energy systems remain.

5.3 Critical thinking aspects

The complete roadmap towards achieving a decarbonized global heating system is still uncertain. Moreover, even if a definitive path were identified, ongoing technological advancements suggest that the perceived optimal pathway at present is bound to evolve and change over time. Some recent publications in the longer term, the impact of heat supply electrification on energy systems can be limited if electrification occurs in a coordinated manner that takes into account our principles and overall technological developments. Energy efficiency and electrification are crucial in countries with high space heating demand. The combination of these approaches, along with ongoing innovation in electrification, supports and strengthens this pathway (Itten et al. 2021; Lowes et al. 2020). Additionally, the electrification of heating demand is expected to play a crucial role in increasing the utilization of renewable energy within energy systems. Generally, the idea sounds reasonable and challenging. Only problem – usually the photovoltaic energy generation works less efficient during the main heating periods (U.S. Energy Information Administration 2022). The efficiency of photovoltaic (PV) systems, or the amount of energy they can convert from sunlight, is typically around 20%. However, the cost-effectiveness of PV systems is affected by other factors, such as electricity production costs, space requirements, resource usage, and carbon emissions savings (Harry Wirth 2022). The maximum amount of electricity a generator can support at the point of connection to the transmission and distribution system during summer and winter is known as net summer and net winter electricity generation capacity respectively. These are usually determined by performance tests. Two main factors that influence the capacity difference between summer and winter are the temperature of cooling water for thermal power plants and the temperature of ambient air for combustion turbines and water flow and reservoir storage characteristics for hydropower plants (Laughlin 2017).

“Heating electrification is one of the biggest mistakes of the energy transition” (Simon 2019) - Heating electrification is one of the biggest mistakes of the energy transition says Christian Holter, who calls for allocating scarce renewable energy resources to economic sectors where they can bring the most in terms of carbon reduction. “It “makes no sense” to bring electric power into heating, because winter demand for heat is 5 to 10 times bigger than the entire electricity system, which won’t be able to cope:” – is another citation found by Holter (Simon 2019).

The EU has set a goal to increase the use of renewable energy in heating and cooling by 1.3% annually until 2030, but some experts believe this goal is too ambitious while others think it is not enough to fully decarbonize the energy system by 2050. The challenges in achieving this goal include lack of education and understanding about renewable energy, difficulties in obtaining financing for long-term investments, and preference for short-term returns on investments over long-term financial exposure. Renewable energy infrastructure projects require long-term perspectives of 25-30 years and returns on investments in 15-20 years to be sustainable. This is where the finance industry and the needs for sustainable development are in opposite directions. The city of Kaposvár is already on the right track. Science could and will support the chosen path. The current work did show the need for additional sustainable heating sources as there are high demand clusters and also consumers who support further investments in district heating and urban energy systems. Just the combined approach to use waste heat together with classical heat generation plants truly accounts for the term urban energy systems. Within the literature there is usually a singular approach, an approach to combine different sources and supplies is rarely found. The focus is on electricity as it's easier to transmit and to consume (Sarbu and Sebarchievici 2018).

Acknowledging the anticipated demand for substantial heat electrification in numerous countries, alongside the simultaneous decarbonization of electricity. Current research focuses mainly on electricity and tries to electrify several aspects (e.g., cars). It is still not proven that electrification is the right approach – but it reduces uncertainties. Thermal energy storage(TES) systems are also an answer to sustainable and green house gas reducing approaches – which affect heating (Laughlin 2017).

5.4 Further research aspects

Further research on the topic of district heating and its impact on energy systems could include:

- Examining the potential for energy efficiency measures to be combined with electrification in order to meet heating demand where it is significant.
- Examining local challenges facing implementation of renewable energy infrastructure projects and investigating solutions to overcome them.
- Investigating innovative solutions for the urban energy systems, like waste heat recovery and combined heat and power.
- Examining the potential for demand-side management strategies, such as load shifting and peak shaving, to reduce the strain on urban energy systems during times of high demand.

- Investigating consumer preferences and behavior in relation to energy consumption and willingness to support further investments in district heating and urban energy systems.

Especially the last part would be crucial to expanding involvement of the European union. Country-specific distinctions would result in more specific programs dependent on local needs and possibilities. Countries such as Norway with their huge availability of waterpower generations are in need for totally different funding programs than especially East European countries. Here is still gap on research towards consumer behavior and expectation (in Eastern Europe).

The work has shown it’s possible to get some results on anonymous measured values – but better and further recommendation could only be given on more detailed data. Building size, number of inhabitants or detailed definition of business hours could further contribute to recommendation such as insolation of the buildings itself or even indicate missing pipelines for easier consumption and low-loss transmission. The transmission itself could also be part of further research. The infrastructure is already provided and available in Kaposvár – but could there be secondary usages such as de-icing: Heat transmission pipelines can be used to melt snow and ice on roads and sidewalks to improve safety and mobility during winter conditions. Another option could be research in the area of soil remediation: heat transmission pipelines can be used to heat contaminated soil to temperatures high enough to destroy pollutants and make the soil safe for use again or enable and help farmers grow crops year-round, regardless of the outside temperature.

6 New scientific findings

The new scientific results are based on answering the research questions of the dissertation. The dissertation consists of mainly three parts: literature review, K-means analysis of the measured values and the consumer opinion.

The research gaps identified in the literature review could be closed to some extent. Mainly the lack of research in Eastern Europe based data and the lack of research in consumer opinion in district heating. According to the literature review, the most used and efficient method for clustering district heating data is K-means. Other methods have been discussed and evaluated, but no consensus on an alternative method was found among the reviewed literature. Each researcher who tried a different approach only compared their results to K-means. Main findings are:

1. As the coverage of Eastern European countries in this field is nonexistent, this research can be considered to be a pilot one at this field.

While performing the actual K-means analysis, the smart metering based (containing hourly measured data) district heating dataset is extremely large and requires high-performance hardware and careful data analysis. Cloud-based techniques are not sufficient for this type of investigation. It is possible to use K-means clustering on them, although it requires high-performance hardware. Main finding here is:

2. The dataset was sampled to reduce hardware demands, and the same number of clusters and values were identified.
3. Four clusters were identified in the two-dimensional space (of operating hours and heat demand variables) by using K-means

algorithm, The cluster centers from the full and from the sample database have proved to be the same – was proven by t-test at significancy level of 0.05. The result also revealed a large heat demand at low operating hours as well as on a larger number of operating hours.

4. The dataset was sampled to reduce hardware demands, and the

While analyzing the consumer opinion closes another gap within the research in general, the performed method reveals all consumers generally accept the necessity and effectiveness of district heating. To conclude the finding:

5. Q-method analysis reveals four distinct perceptual factors influencing attitudes towards district heating: environmental sustainability, convenience and reliability, cost-effectiveness, and concerns about individual autonomy.

Based on the identified K-means clusters and groups of the Q-method findings, a new idea came up:

6. In case of smart metering-based dynamic regulation available for the customers, a dynamic pricing model should be created as it encourages customers to reduce energy consumption during peak demands, that makes a step into sustainable direction.

At the peak demand times the heat comes mostly from less-sustainable resources thus, this financial technique makes a step into sustainable direction. Obviously, this will only be implementable if the official prices for district heating get liberalized. This method can encourage customers to reduce energy consumption during peak demand, increase revenue, and align supply and demand.

Based on the technical details(and/or additional parameters) it is desirable to use the waste heat for district heating from the glass factory that is under construction in Kaposvár. This solution is not only CO_2 emission reducing but also environmentally friendly by reducing the environmental heat load that the cooling technology would imply. It is also economical friendly, as it eliminates the need to install cooling technology, which would require a significant investment.

The utilization of cross-company waste heat, where waste heat that cannot be used internally is used by third parties in commercial or residential buildings, is a promising solution to increase energy efficiency and reduce CO_2 emissions. The most economically feasible utilization of waste heat requires spatial proximity of the waste heat source and demand, and heat recovery or heat displacement through the use of heat exchangers is the most efficient and simplest technological approach. Thus, not only the technology of energy creation as such should be considered, but also the context from which the technology originates. The last finding would be:

7. The Utilization of Waste Heat for District Heating from a Glass Factory in Kaposvár Promotes Energy Efficiency and Sustainability.

7 Summary

The research aims to analyze the use of smart meters in a district heating system, focusing on local needs and taking a scientific approach to problem-solving. The objectives of the research include:

1. Perform economic and service research on the district heating system of Kaposvár, Hungary.
2. Explore the benefits and potential of smart meters in providing new insights and improving district heating operations.
3. Use data-driven approaches, specifically the K-means algorithm, to cluster district heating users based on consumption patterns and behavior.
4. Analyze consumers' opinions and mentalities regarding district heating through Q-method-based opinion categorization.
5. Investigate strategies that the district heating company could apply to leverage the available measured data in line with sustainability goals.
6. Examine the relationship between strategies and personal behavioral styles.
7. Assess the willingness of consumers to follow and support the proposed strategies.

The research methodology involves a literature review using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) scheme. The review focuses on customer perspectives, pricing models, and sustainability aspects of district heating. The K-means clustering algorithm is applied to analyze hourly measured power demand and segment users into different clusters. Data transformation and analysis are performed using STATA and R Script.

For the Q-methodology research, a concourse of statements is developed to capture consumers' opinions on district heating. The Q-Set is created based on various sources, including literature, interviews, and websites. Q-method analysis is conducted to identify factors that represent different perspectives on district heating. Factor loadings, eigenvalues, factor correlations, and composite reliability scores are analyzed to determine the optimal number of factors.

The results of the research include empirical findings on demand-based measurements, cluster analysis results, and the outcomes of the Q-method analysis. Empirical results show the distribution of operating hours and power demand after transformation. The Q-method analysis identifies four factors representing different perspectives on district heating, with factor loadings and distinguishing statements for each factor.

Overall, the research aims to contribute to the efficiency of district heating, reduction of environmental impact, and higher consumer satisfaction by leveraging smart meter data, understanding consumer opinions, and developing effective strategies.

8 Acknowledgements

Since the beginning of my personal PhD journey in 2019, I have been supported by my family, especially by my wife. So, the first thanks has to go to her and my family.

I would like to express my sincere gratitude to Dr. Gyöngyi Bánkuti, my supervisor, for her exceptional guidance throughout the entire process. Her invaluable insights, constructive feedback, and meticulous attention to detail constantly motivated me to surpass the ordinary standards and exceed the anticipated outcomes. I would like to thank Zsuzsanna Zanatyné Uitz. Without her this research would not have been possible – she can be seen as a second and external supervisor.

Furthermore, I extend my heartfelt appreciation to the esteemed professors and lecturers at MATE Hungarian University of Agriculture and Life Sciences - Kaposvár Campus, Kaposvár, Hungary, who have been instrumental in my academic journey. Their extensive expertise and valuable guidance not only enriched my knowledge but also contributed to insightful discussions on research methods and publication strategies.

In particular, I would like to thank Dr. Imre Fertő as Head of Doctoral School of Management and Organizational Sciences and Dr. Kinga Szabó and Zsófia Móga for the great support in all organizational concerns. Special thanks go to Dr. Sándor Kerekes who introduced us to Q-Methods methodology and also provided advise in the actual usage.

Furthermore, I would like to thank all colleagues, professors and academics who supported me in various ways at my start in the world of science, including Prof. Dr. Uwe Wloka who unfortunately passed away before I could finish, Dr. Volker Persch and Dr. Daniel Spielmann from SAP who helped with the Q-Method, as well as Dr. Reiner Meyer and Felix Moehler who completed the expert round. Jiantao He and Andre Clauss always provided open feedback on some details within the energy market and the billing solution of SAP.

Throughout my journey, I have had the privilege of encountering numerous individuals who have provided me with invaluable support in various ways. I would like to express my sincere gratitude to all of these remarkable individuals for their unwavering support and assistance.

Lastly, I would like to extend my heartfelt appreciation to my fellow students who shared this challenging journey with me. It was a privilege to have their companionship and support throughout this transformative period. I would like to mention Atilla Wohllebe, Doreen Kaempf, Thomas Stoyke and Dirk Siegfried Huebner who authored or co-authored several publications and always motivated me to achieve the best possible results.

The research was supported by the District Heating company of Kaposvár by providing the needed data and the colleagues from the Novustat GmbH who provided insight and expertise that assisted the research on the details of the R programming for the K-means analysis.

Dresden, Germany, January 2023

9 Publications of the author

9.1 Topic related publications

Radtke, U. (2022). K-MEANS CLUSTER ANALYSIS OF HOURLY MEASURED POWER DEMAND IN THE DISTRICT HEATING SYSTEM IN KAPOSVÁR. Regional and Business Studies (2022) Vol 14 No 1, 57–72. https://doi.org/10.33568/rbs.3343

Radtke, U. (2022). A Brief Literature Review of Structuring District Heating Data based on Measured Values. The Ohio Journal of Science, 122(2), 75–84. https://doi.org/10.18061/ojs.v122i2.8845

Radtke, U. & Kaempf, D. (2021). RELATION BETWEEN DISTRICT HEATING SYSTEM OF KAPOSVÁR AND SUSTAINABILITY: in Challenges in the Carpathian Basin - Global Challenges - Local Answers. Interdependencies or Slobalisation?, 1197–1215. http://csik.sapientia.ro/en/news/global-challenges-local-answers-interdependencies-or-slobalisation

Uwe Radtke & Doreen Kaempf (2021). Relation between Urban energy systems - analysis of smart metering district heating system of Kaposvár and sustainability. IOSR Journal of Environmental Science, Toxicology and Food Technology, 15(3 Ser. II), 59–65. https://www.iosrjournals.org/iosr-jestft/papers/Vol15-Issue3/Series-2/G1503025965.pdf

9.2 Additional publications

Radtke, U. (2022). Unbiasing Science by Using Search Engines and Structured Methods – Analyzing or Measuring Scientific Output. International Journal of Applied Research in Business and Management, *3*(1), 19–28. https://doi.org/10.51137/ijarbm.2022.3.1.3

Uwe Radtke (2022). Employee satisfaction equally important to customer satisfaction – motivation model according to Maslow's pyramid of needs: In: Szilárd, BERKE; Katalin, SZABÓ; Beáta SZÜCS, Pató Gáborné (eds.) Organizational Behaviour and Leadership Theory in Practice, 47–55.

Radtke, U., Kaempf, D., Stoyke, T., & Huebner, D. S. (2022). Case study of small and very small businesses in Germany during COVID-19 in 2021. International Journal of Management and Enterprise Development, 21(2), Article 122106, 164. https://doi.org/10.1504/IJMED.2022.122106

Wohllebe, A., Hübner, D.-S., Radtke, U., & Podruzsik, S. (2022). Recommending a Retailer's Mobile App – Influence of the Retailer and the Mediating Role of Push Notifications. In M. E. Auer & T. Tsiatsos (Eds.), Lecture Notes in Networks and Systems. New Realities, Mobile Systems and Applications (Vol. 411, pp. 361–371). Springer International Publishing. https://doi.org/10.1007/978-3-030-96296-8_32

Radtke, U., & Kaempf, D. (2021). Lifelong learning-studying in the European higher education area with Covid-19. International Journal of Learning and Change, 1(1), Article 10039866, https://doi.org/10.1504/IJLC.2021.10039866

Stoyke, T., & Radtke, U. (2021). Economic Implications for Stationary Trade Under the Influence of SARS-CoV-2. International Journal of Applied Research in Business and Management, 2(2), 1–13. https://doi.org/10.51137/ijarbm.2021.2.2.1

Wohllebe, A., Hübner, D.-S., Radtke, U., & Podruzsik, S. (2021). Mobile apps in retail: Effect of push notification frequency on app user behavior.

Innovative Marketing, 17(2), 102–111. https://doi.org/10.21511/im.17(2).2021.10

Wohllebe, A., Hübner, D.-S., Radtke, U., & Wohllebe, A. (2021). Smartphones and Mobile Apps: Case Study on Usage Behavior of Elementary School Students – Insights from a Rural Area in Northern Germany. International Journal of Interactive Mobile Technologies (IJIM), 15(12), 184. https://doi.org/10.3991/ijim.v15i12.22565

Uwe Radtke, & Doreen Kaempf (2020). Postgrowth Economists - Christian Felber: "Change everything: Creating an economy for the common good": In: Diána, Koponicsné Györke; Róbert, Barna (eds.) Proceedings of the International Conference on Sustainable Economy and Agriculture, 333–339. https://m2.mtmt.hu/gui2/?mode=browse&params=publication;31384060

9.3 Conferences

- 61st ERSA Congress with “K-means Cluster analysis of hourly measured power demand”; Pécs Hungary (2022)
- 15th International Conference on Economics and Business with “RELATION BETWEEN DISTRICT HEATING SYSTEM OF KAPOSVÁR AND SUSTAINABILITY”, Cluj, Kolozsvár, Romania (2021)
- International Conference on Sustainable Economy and Agriculture with “New economic visions - Postgrowth Economists: Christian Felber. Establishes and advocates the economy for the common good movement, transmitting post-growth economics to businesses and banks”; Kaposvár, Hungary : Kaposvár University (2019)

10 Short professional CV

Time	Company	Position
since 06/2003	SAP SE	Chief Development Architect
01/2002 – 05/2003	SDV AG	Consultant & IT Lead for complete IT landscape
07/1999 – 10/2001	Verdi Information Consult AG	Development Consultant and Team Lead for document management software
11/1998 – 06/1999	GfD – Gesellschaft für Datenverarbeitung mbG	Consultant
10/1995 – 10/1998	GfD – Gesellschaft für Datenverarbeitung mbG	Vocational Student

References

Åberg, M., J. Widén, and D. Henning. 2012. "Sensitivity of District Heating System Operation to Heat Demand Reductions and Electricity Price Variations: A Swedish Example." *Energy* 41 (1): 525–40. https://doi.org/10.1016/j.energy.2012.02.034.

Anbumozhi, Venkatachalam, and Armin Bauer. 2010. "Impact of Global Recession on Sustainable Development and Poverty Linkages." *SSRN Journal*. https://doi.org/10.2139/ssrn.1653700.

Andrews, Dave. 2009. ""Carbon Footprints of Various Sources of Heat - Biomass Combustion and CHPDH Comes Out Lowest " - William Orchard." *Claverton Group*, February 7, 2009. https://claverton-energy.com/carbon-footprints-of-various-sources-of-heat-chpdh-comes-out-lowest.html.

Angelidis, O., A. Ioannou, D. Friedrich, A. Thomson, and G. Falcone. 2023. "District Heating and Cooling Networks with Decentralised Energy Substations: Opportunities and Barriers for Holistic Energy System Decarbonisation." *Energy* 269:126740. https://doi.org/10.1016/j.energy.2023.126740.

Balch, George, and Steven R. Brown. 1982. "Political Subjectivity: Applications of Q Methodology in Political Science." *Journal of Marketing Research* 19 (1): 162. https://doi.org/10.2307/3151542.

Bánkuti, G., and S. Zanatyné Uitz. 2021. "A Good Practice in Urban Energetics in a Hungarian Small Town, Kaposvar." In *Handbook of Smart Energy Systems*, edited by Mahdi Fathi, Enrico Zio, and Panos M. Pardalos, forthcomming. Cham: Springer International Publishing; Imprint: Springer.

Boldrini, A., J. P. Jiménez Navarro, W.H.J. Crijns-Graus, and M. A. van den Broek. 2022. “The Role of District Heating Systems to Provide Balancing Services in the European Union.” *Renewable and Sustainable Energy Reviews* 154:111853. https://doi.org/10.1016/j.rser.2021.111853.

Bose, Prosenjit, Anil Maheshwari, and Pat Morin. 2003. “Fast Approximations for Sums of Distances, Clustering and the Fermat–Weber Problem.” *Computational Geometry* 24 (3): 135–46. https://doi.org/10.1016/S0925-7721(02)00102-5.

Braeken, Johan, and Marcel A. L. M. van Assen. 2017. “An Empirical Kaiser Criterion.” *Psychological methods* 22 (3): 450–66. https://doi.org/10.1037/met0000074.

Brown, Steven R. 1980. *Political Subjectivity: Applications of Q Methodology in Political Science.* New Haven. https://qmethod.org/wp-content/uploads/2016/01/brown-1980-politicalsubjectivity.pdf.

Brown, Steven R. 1996. “Q Methodology and Qualitative Research.” *Qual Health Res* 6 (4): 561–67. https://doi.org/10.1177/104973239600600408.

Calikus, Ece, Sławomir Nowaczyk, Anita Sant'Anna, Henrik Gadd, and Sven Werner. 2019. “A Data-Driven Approach for Discovering Heat Load Patterns in District Heating.” *Applied Energy* 252:113409. https://doi.org/10.1016/j.apenergy.2019.113409.

Camerlink, Irene, and Péter Pongrácz. 2022. “A Note on the Use of English Language Editing Services.” *Applied Animal Behaviour Science*, 105675. https://doi.org/10.1016/j.applanim.2022.105675.

Churruca, Kate, Kristiana Ludlow, Wendy Wu, Kate Gibbons, Hoa Mi Nguyen, Louise A. Ellis, and Jeffrey Braithwaite. 2021. “A Scoping Review of Q-Methodology in Healthcare Research.” *BMC medical research methodology* 21 (1): 125. https://doi.org/10.1186/s12874-021-01309-7.

Csima, F., and J. & Szendefy. 2009. “Synergic Consideration of Competitiveness and Ecology in Biogas Production and Use.” *Regional and Business Studies* 1 (1): 45–48.

Dalkey, Norman, and Olaf Helmer. 1963. “An Experimental Application of the DELPHI Method to the Use of Experts.” *Management Science* 9 (3): 458–67. https://doi.org/10.1287/mnsc.9.3.458.

Damio, Siti Maftuhah. 2016. “Q Methodology: An Overview and Steps to Implementation.” *Asian Journal of University Education* 12 (1): 105–17. https://files.eric.ed.gov/fulltext/EJ1207820.pdf.

DiStefano, Christine, Min Zhu, and Diana Mîndrilã. 2009. “Understanding and Using Factor Scores: Considerations for the Applied Researcher.”.

DTREG. 2022. “Correlation & Factor Analysis | Solution | DTREG.” Accessed December 11, 2022. https://www.dtreg.com/solution/correlation.

Du, Yuyan, Chao Wang, Hailong Li, Jingjing Song, and Bo Li. 2019. “Clustering Heat Users Based on Consumption Data.” *Energy Procedia* 158:3196–3201. https://doi.org/10.1016/j.egypro.2019.01.1010.

Fang, Hao, Jianjun Xia, Kan Zhu, Yingbo Su, and Yi Jiang. 2013. “Industrial Waste Heat Utilization for Low Temperature District

Heating.” *Energy Policy* 62:236–46. https://doi.org/10.1016/j.enpol.2013.06.104.

Federal Ministry for Economic Affairs and Climate Germany. 2023. “What Exactly Is 'District Heating'?” Accessed June 16, 2023. https://www.bmwi-energiewende.de/EWD/Redaktion/EN/Newsletter/2021/03/Meldung/direkt-account.html.

Fink, Arlene. 2020. *Conducting Research Literature Reviews: From the Internet to Paper*. Fifth edition. Los Angeles: SAGE.

Gabbrielli, R., P. Castrataro, F. Del Medico, M. Di Palo, and B. Lenzo. 2014. “Levelized Cost of Heat for Linear Fresnel Concentrated Solar Systems.” *Energy Procedia* 49:1340–49. https://doi.org/10.1016/j.egypro.2014.03.143.

Gadd, Henrik, and Sven Werner. 2013. “Heat Load Patterns in District Heating Substations.” *Applied Energy* 108:176–83. https://doi.org/10.1016/j.apenergy.2013.02.062.

Gianniou, Panagiota, Xiufeng Liu, Alfred Heller, Per Sieverts Nielsen, and Carsten Rode. 2018. “Clustering-Based Analysis for Residential District Heating Data.” *Energy Conversion and Management* 165:840–50. https://doi.org/10.1016/j.enconman.2018.03.015.

Giordani, Paolo. 2020. *An Introduction to Clustering with R*. With the assistance of M. B. Ferraro and F. Martella. Behaviormetrics v.1. Singapore: Springer Singapore Pte. Limited. https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6320880.

Glass online. 2022. “Sisecam to Build Glass Packaging Plant in Kaposvár, Hungary | GlassOnline.Com - the World's Leading Glass Industry Website.” Accessed October 03, 2022. https://www.glassonline.com/sisecam-to-build-glass-packaging-plant-in-kaposvar-hungary/.

Goia, Aldo, Caterina May, and Gianluca Fusai. 2010. “Functional Clustering and Linear Regression for Peak Load Forecasting.” *International Journal of Forecasting* 26 (4): 700–711. https://doi.org/10.1016/j.ijforecast.2009.05.015.

Guba, Beate. 2008. “Systematische Literatursuche.” [Systematic literature search]. *Wiener medizinische Wochenschrift (1946)* 158 (1-2): 62–69. https://doi.org/10.1007/s10354-007-0500-0.

Gudmundsson, O., J. E. Thorsen, and M. Brand. 2018. “The Role of District Heating in Coupling of the Future Renewable Energy Sectors.” *Energy Procedia* 149:445–54. https://doi.org/10.1016/j.egypro.2018.08.209.

Gudmundsson, O., J. E. Thorsen, and L. Zhang. 2013. “Cost Analysis of District Heating Compared to Its Competing Technologies.” In *Energy and Sustainability IV*, edited by C. A. Brebbia, A. M. Marinov, A. A. Mammoli, and C. A. Safta, 3–13. WIT Transactions on Ecology and The Environment: WIT PressSouthampton, UK.

Guelpa, Elisa, Stefania Deputato, and Vittorio Verda. 2018. “Thermal Request Optimization in District Heating Networks Using a Clustering Approach.” *Applied Energy* 228:608–17. https://doi.org/10.1016/j.apenergy.2018.06.041.

Hair, Black, William & Anderson. 2010. *Multivariate Data Analysis.* 7th ed. Upper Saddle River NJ: Prentice Hall.

Harry Wirth. 2022. “Recent Facts About Photovoltaics in Germany.” https://www.ise.fraunhofer.de/en/publications/studies/recent-facts-about-pv-in-germany.html.

Holl, András. 2022. “Hungary Open Access Case Study.” http://real.mtak.hu/30113/1/Hungary%20Case%20Study.pdf. Accessed June 26, 2022.

IEA. 2022a. “District Heating.” *IEA*. https://www.iea.org/reports/district-heating. Accessed December 20, 2022.

IEA. 2022b. “District Heat Production by Region, 2020, and World Average Carbon Intensities of District Heat Supply in the Net Zero Scenario, 2020-2030.” Accessed April 10, 2022. https://www.iea.org/reports/district-heating.

IEA. 2023. “Net Zero Emissions by 2050 Scenario (NZE) – Global Energy and Climate Model – Analysis - IEA.” Accessed June 16, 2023. https://www.iea.org/reports/global-energy-and-climate-model/net-zero-emissions-by-2050-scenario-nze.

Iglesias, Félix, and Wolfgang Kastner. 2013. “Analysis of Similarity Measures in Times Series Clustering for the Discovery of Building Energy Patterns.” *Energies* 6 (2): 579–97. https://doi.org/10.3390/en6020579.

Itten, Anatol, Fionnguala Sherry-Brennan, Thomas Hoppe, Aarthi Sundaram, and Patrick Devine-Wright. 2021. “Co-Creation as a Social Process for Unlocking Sustainable Heating Transitions in Europe.”

Energy research & social science 74:101956. https://doi.org/10.1016/j.erss.2021.101956.

Jain, Anil K. 2010. “Data Clustering: 50 Years Beyond K-Means.” *Pattern Recognition Letters* 31 (8): 651–66. https://doi.org/10.1016/j.patrec.2009.09.011.

Jouhara, Hussam, Navid Khordehgah, Sulaiman Almahmoud, Bertrand Delpech, Amisha Chauhan, and Savvas A. Tassou. 2018. “Waste Heat Recovery Technologies and Applications.” *Thermal Science and Engineering Progress* 6:268–89. https://doi.org/10.1016/j.tsep.2018.04.017.

Kampen, Jarl K., and Peter Tamás. 2014. “Overly Ambitious: Contributions and Current Status of Q Methodology.” *Qual Quant* 48 (6): 3109–26. https://doi.org/10.1007/s11135-013-9944-z.

Kamstrup. 2023. “Kamstrup - Smart Metering Solutions.” Accessed June 22, 2023. https://www.kamstrup.com/en-en.

Kaposvári Municipal Property Management and Service Co., Ltd. 2022. “Green Heating Plant - the Freedom of District Heating Reaches Apartments.” Accessed September 18, 2022. https://www.kaposho.hu/hirek/zoldfutomu-a-lakasokig-er-a-tavhoszabadsag.html.

Keeney, Sinead, Felicity Hasson, and Hugh P. McKenna. 2011. *The Delphi Technique in Nursing and Health Research.* Oxford: Wiley-Blackwell.

Kerekes, Lajos. 2022. “Assessment of Price Regulation in Hungarian District Heating Systems.” Accessed June 20, 2023.

https://rekk.hu/analysis-details/285/assessment-of-price-regulation-in-hungarian-district-heating-systems.

Kirschbaum, Melissa, Tony Barnett, and Merylin Cross. 2019. “Q Sample Construction: A Novel Approach Incorporating a Delphi Technique to Explore Opinions About Codeine Dependence.” *BMC medical research methodology* 19 (1): 101. https://doi.org/10.1186/s12874-019-0741-9.

Körmendi, G. 2018. “Hungary 2018.” Accessed June 19, 2023. https://www-pub.iaea.org/MTCD/Publications/PDF/cnpp2018/countryprofiles/Hungary/Hungary.htm.

Krikser, Thomas, Adriano Profeta, Sebastian Grimm, and Heiko Huther. 2020. “Willingness-to-Pay for District Heating from Renewables of Private Households in Germany.” *Sustainability* 12 (10): 4129. https://doi.org/10.3390/su12104129.

Krog, Louise, Karl Sperling, Michael Kvist Svangren, and Frede Hvelplund. 2020. “Consumer Involvement in the Transition to 4th Generation District Heating.” 141-152 Pages / International Journal of Sustainable Energy Planning and Management, Vol. 29 (2020). https://doi.org/10.5278/IJSEPM.4627.

Ladan, Muhammad Awwal, Heather Wharrad, and Richard Windle. 2018. “Towards Understanding Healthcare Professionals' Adoption and Use of Technologies in Clinical Practice: Using Qmethodology and Models of Technology Acceptance.” *Journal of innovation in health informatics* 25 (1): 965. https://doi.org/10.14236/jhi.v25i1.965.

Laughlin, Robert B. 2017. “Pumped Thermal Grid Storage with Heat Exchange.” *Journal of Renewable and Sustainable Energy* 9 (4): 44103. https://doi.org/10.1063/1.4994054.

Le Ray, G., and P. Pinson. 2019. “Online Adaptive Clustering Algorithm for Load Profiling.” *Sustainable Energy, Grids and Networks* 17:100181. https://doi.org/10.1016/j.segan.2018.100181.

Lepcha, Mensholong. 2022. “Hungary Recession: Surging Inflation Pushes Hungary to the Brink of a Technical Recession.” *Capital Com SV Investments Limited*, December 28, 2022. https://capital.com/hungary-recession.

Li, J. Song, Q. Sun, F. Wallin, and Q. Zhang. 2019. “A Dynamic Price Model Based on Levelized Cost for District Heating.” *Energ. Ecol. Environ.* 4 (1): 15–25. https://doi.org/10.1007/s40974-019-00109-6.

Li, and Wu. 2012. “A Clustering Method Based on K-Means Algorithm.” *Physics Procedia* 25:1104–9. https://doi.org/10.1016/j.phpro.2012.03.206.

Li, Hailong, Qie Sun, Qi Zhang, and Fredrik Wallin. 2015. “A Review of the Pricing Mechanisms for District Heating Systems.” *Renewable and Sustainable Energy Reviews* 42: 56–65. https://www.sciencedirect.com/science/article/pii/S136403211400820X.

Linden, Mikael, and Päivi Peltola-Ojala. 2010. “The Deregulation Effects of Finnish Electricity Markets on District Heating Prices.” *Energy Economics* 32 (5): 1191–98. https://doi.org/10.1016/j.eneco.2010.03.002.

Lowes, Richard, Jan Rosenow, Meysam Qadrdan, and Jianzhong Wu. 2020. “Hot Stuff: Research and Policy Principles for Heat Decarbonisation Through Smart Electrification.” *Energy research & social science* 70:101735. https://doi.org/10.1016/j.erss.2020.101735.

Ma, Zhenjun, Rui Yan, and Natasa Nord. 2017. “A Variation Focused Cluster Analysis Strategy to Identify Typical Daily Heating Load Profiles of Higher Education Buildings.” *Energy* 134:90–102. https://doi.org/10.1016/j.energy.2017.05.191.

MacKay, David J. C. 2003. *Information Theory, Inference and Learning Algorithms.* Cambridge: Cambridge University Press. http://www.loc.gov/catdir/description/cam032/2003055133.html Publisher description.

MacQueen, J. B. 1967. “Some Methods for Classification and Analysis of Multivariate Observations.” *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability* 1: 281–97. https://projecteuclid.org/Proceedings/berkeley-symposium-on-mathematical-statistics-and-probability/proceedings-of-the-fifth-berkeley-symposium-on-mathematical-statistics-and/toc/bsmsp/1200512974.

Marquant, Julien F., L. Andrew Bollinger, Ralph Evins, and Jan Carmeliet. 2018. “A New Combined Clustering Method to Analyse the Potential of District Heating Networks at Large-Scale.” *Energy* 156:73–83. https://doi.org/10.1016/j.energy.2018.05.027.

Mezősi, András, Enikő Kácsor, Ákos Beöthy, Ágnes Törőcsik, and László Szabó. 2017. “Modelling Support Policies and Renewable Energy Sources Deployment in the Hungarian District Heating Sector.”

Energy & Environment 28 (1-2): 70–87. https://doi.org/10.1177/0958305X16685473.

Mirghaderi, Seyed Hadi, and Zahra Mohit Ghiri. 2019. “Measuring Sustainable Development: Linear Regression Approach.” *IJSD* 22 (1/2): 110. https://doi.org/10.1504/IJSD.2019.104736.

Moher, David, Alessandro Liberati, Jennifer Tetzlaff, and Douglas G. Altman. 2009. “Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.” *PLoS medicine* 6 (7). https://doi.org/10.1371/journal.pmed.1000097.

Nick Robins. 2009. “Will Sustainable Investing Survive the Recession?” Accessed December 28, 2022. https://www.gresham.ac.uk/watch-now/will-sustainable-investing-survive-recession.

Oh, Saesin, and Sang-Kee Kim. 2022. “Impact of Heat Price Regulation on the Optimal District Heating Production Mix and Its Policy Implications.” *Energy* 239:122305. https://doi.org/10.1016/j.energy.2021.122305.

Ordóñez, Camilo, Dave Kendal, Caragh G. Threlfall, Dieter F. Hochuli, Melanie Davern, Richard A. Fuller, Rodney van der Ree, and Stephen J. Livesley. 2020. “How Urban Forest Managers Evaluate Management and Governance Challenges in Their Decision-Making.” *Forests* 11 (9): 963. https://doi.org/10.3390/f11090963.

Oti, Eric U., Michael O. Olusola, Francis C. Eze, and Samuel U. Enogwe. 2021. “Comprehensive Review of K-Means Clustering Algorithms.” *IJASRE* 07 (08): 64–69. https://doi.org/10.31695/IJASRE.2021.34050.

Pelleg, Dan, and Andrew Moore. 1999. “Accelerating Exact K -Means Algorithms with Geometric Reasoning.” In *KDD-99: The 5th ACM*

SIGKDD International Conference on Knowledge Discovery and Data Mining, edited by Surajit Chaudhuri and David. E. Madigan, 277–81. New York, New York, USA: ACM Press.

Poputoaia, Diana, and Stefan Bouzarovski. 2010. "Regulating District Heating in Romania: Legislative Challenges and Energy Efficiency Barriers." *Energy Policy* 38 (7): 3820–29. https://doi.org/10.1016/j.enpol.2010.03.002.

Radtke, Uwe. 2022a. "A Brief Literature Review of Structuring District Heating Data Based on Measured Values." *OJS* 122 (2): 75–84. https://doi.org/10.18061/ojs.v122i2.8845.

Radtke, Uwe. 2022b. "Unbiasing Science by Using Search Engines and Structured Methods – Analyzing or Measuring Scientific Output." *IJARBM* 3 (1): 19–28. https://doi.org/10.51137/ijarbm.2022.3.1.3.

Radtke, Uwe, and Doreen Kaempf. 2021. "RELATION BETWEEN DISTRICT HEATING SYSTEM of KAPOSVÁR and SUSTAINABILITY: In Challenges in the Carpathian Basin - Global Challenges - Local Answers. Interdependencies or Slobalisation?", 1197–1215. http://csik.sapientia.ro/en/news/global-challenges-local-answers-interdependencies-or-slobalisation.

Ramlo, Sue. 2016. "Centroid and Theoretical Rotation: Justification for Their Use in Q Methodology Research." *Mid-Western Educational Researcher* 28 (1). https://www.mwera.org/MWER/volumes/v28/issue1/v28n1-Ramlo-SPECIALIZED-RESEARCH-STATISTICAL-METHODS.pdf. Accessed December 17, 2022.

Ramos, Sérgio, João M. Duarte, F. Jorge Duarte, and Zita Vale. 2015. “A Data-Mining-Based Methodology to Support MV Electricity Customers’ Characterization.” *Energy and Buildings* 91:16–25. https://doi.org/10.1016/j.enbuild.2015.01.035.

Reddy, Damodar, and Prasanta K. Jana. 2012. “Initialization for K-Means Clustering Using Voronoi Diagram.” *Procedia Technology* 4:395–400. https://doi.org/10.1016/j.protcy.2012.05.061.

Rezaie, Behnaz, and Marc A. Rosen. 2012. “District Heating and Cooling: Review of Technology and Potential Enhancements.” *Applied Energy* 93:2–10. https://doi.org/10.1016/j.apenergy.2011.04.020.

Sala, Juan, Rongling Li, and Morten H. Christensen. 2019. “Clustering and Classification of Energy Meter Data: A Comparison Analysis of Data from Individual Homes and the Aggregated Data from Multiple Homes.” *Build. Simul.* https://doi.org/10.1007/s12273-019-0587-4.

Sarbu, Ioan, and Calin Sebarchievici. 2018. “A Comprehensive Review of Thermal Energy Storage.” *Sustainability* 10 (1): 191. https://doi.org/10.3390/su10010191.

Schneider Electric SE. 2022. “Termis Engineering | Schneider Electric.” Accessed June 22, 2023. https://www.se.com/il/en/product-range/61613-termis-engineering/#overview.

Sernhed, Kerstin, Henrik Gåverud, and Annamaria Sandgren. 2017. “Customer Perspectives on District Heating Price Models.” 47-60 Pages / International Journal of Sustainable Energy Planning and Management, Vol 13 (2017). https://doi.org/10.5278/IJSEPM.2017.13.4.

Sigmond, G. 2009. “The Situation of District Heating, District Cooling and Energy Supply in Hungary.” Scientific Society of Measurement, Automation and Informatics, Branch of Thermal Engineering and Thermogrammetry.

Simon, Frédéric. 2019. “Heating Electrification Is One of the Biggest Mistakes of the Energy Transition.” *EURACTIV*, April 16, 2019. https://www.ise.fraunhofer.de/en/publications/studies/recent-facts-about-pv-in-germany.html.

Simón-Martín, Miguel de, ed. 2022a. *Levelized Cost of Energy in Sustainable Energy Communities: A Systematic Approach for Multi-Vector Energy Systems.* SpringerBriefs in Applied Sciences and Technology. Cham: Springer.

Simón-Martín, Miguel de, Stefano Bracco, Giorgio Piazza, Luisa Carlotta Pagnini, Alberto González-Martínez, and Federico Delfino. 2022b. “The Levelized Cost of Energy Indicator.” In *Levelized Cost of Energy in Sustainable Energy Communities: A Systematic Approach for Multi-Vector Energy Systems*, edited by Miguel de Simón-Martín, 31–76. SpringerBriefs in Applied Sciences and Technology. Cham: Springer.

Song, Jingjing, Fredrik Wallin, Hailong Li, and Björn Karlsson. 2016. “Price Models of District Heating in Sweden.” *Energy Procedia* 88:100–105. https://doi.org/10.1016/j.egypro.2016.06.031.

Southeastern Oklahoma State University. 2022. “Gain a Better Understanding of Consumer Behavior.” Accessed March 20, 2022. https://online.se.edu/articles/mba/gain-a-better-understanding-of-consumer-behavior.aspx.

Soyamedia.com. 2022. "Change in State Activity in the Hungarian Energy Market (Detail) - Kinstellar." Accessed April 10, 2022. https://www.kinstellar.com/insights/detail/232/change-in-state-activity-in-the-hungarian-energy-market.

Stadtwerke Bonn. 2021. "Erläuterung Der Fernwärmepreise Von SWB Energie Und Wasser - SWB Energie Und Wasser - Stadtwerke Bonn." Accessed November 21, 2021. https://www.stadtwerke-bonn.de/fuer-zuhause/produkte/fernwaerme/erlaeuterung-der-fernwaermepreise-von-swb-energie-und-wasser/.

Stadtwerke Kaiserslautern. 2021. "District Heating: Our Offer for You | SWK Kaiserslautern." Accessed November 21, 2021. https://www.swk-kl.de/en/products-services/energy/district-heating.

Stanytsina, Valentyna, Volodymyr Artemchuk, Olga Bogoslavska, Iryna Zinovieva, and Nataliia Ridei. 2021. "The Influence of Environmental Tax Rates on the Levelized Cost of Heat on the Example of Organic and Biofuels Boilers in Ukraine." *E3S Web Conf.* 280:9012. https://doi.org/10.1051/e3sconf/202128009012.

Stennikov, Valery, and Andrey Penkovskii. 2020. "The Pricing Methods on the Monopoly District Heating Market." *Energy Reports* 6:187–93. https://doi.org/10.1016/j.egyr.2019.11.061.

Stone, and Turale. 2015. "Q Methodology: An Introduction." *Pacific Rim International Journal of Nursing Research* 19 (3): 183–86. https://he02.tci-thaijo.org/index.php/PRIJNR/article/view/38444.

Tarassova, D. 2020. "Energy Markets Explained: Deregulated Vs. Regulated Markets." *Dash Energy*, September 29, 2020.

https://dash.energy/2020/09/29/energy-markets-explained-deregulated-vs-regulated-markets/.

Toffetti, L. 2015. “File:District Heating.Gif - Wikimedia Commons.” Accessed June 05, 2022. https://commons.wikimedia.org/wiki/File:District_heating.gif.

Tu, Jui-Che, Xiu-Yue Zhang, and Sin-Yi Huang. 2018. “Key Factors of Sustainability for Smartphones Based on Taiwanese Consumers’ Perceived Values.” *Sustainability* 10 (12): 4446. https://doi.org/10.3390/su10124446.

Tureczek, Alexander Martin, Per Sieverts Nielsen, Henrik Madsen, and Adam Brun. 2019. “Clustering District Heat Exchange Stations Using Smart Meter Consumption Data.” *Energy and Buildings* 182:144–58. https://doi.org/10.1016/j.enbuild.2018.10.009.

U.S. Energy Information Administration. 2022. “Frequently Asked Questions (FAQs) - U.S. Energy Information Administration (EIA).” Accessed January 15, 2023. https://www.eia.gov/tools/faqs/faq.php?id=101&t=3.

Ueno, Tsuyoshi, Fuminori Sano, Osamu Saeki, and Kiichiro Tsuji. 2006. “Effectiveness of an Energy-Consumption Information System on Energy Savings in Residential Houses Based on Monitored Data.” *Applied Energy* 83 (2): 166–83. https://doi.org/10.1016/j.apenergy.2005.02.002.

Valenta, A. L., and U. Wigger. 1997. “Q-Methodology: Definition and Application in Health Care Informatics.” *Journal of the American Medical Informatics Association : JAMIA* 4 (6): 501–10. https://doi.org/10.1136/jamia.1997.0040501.

Város, Kaposvár Megyei Jogú. 2017. “Infrastructure - Kaposvár Megyei Jogú Város Hivatalos Weboldala.” Accessed June 19, 2023. https://otthon.kaposvar.hu/en/infrastructure/.

Völker, Benjamin, Andreas Reinhardt, Anthony Faustine, and Lucas Pereira. 2021. “Watt’s up at Home? Smart Meter Data Analytics from a Consumer-Centric Perspective.” *Energies* 14 (3): 719. https://doi.org/10.3390/en14030719.

Wallis, Jennifer, Jan Burns, and Rose Capdevila. 2009. “Q Methodology and a Delphi Poll: A Useful Approach to Researching a Narrative Approach to Therapy.” *Qualitative Research in Psychology* 6 (3): 173–90. https://doi.org/10.1080/14780880701734545.

Wang, Chao, Yuyan Du, Hailong Li, Fredrik Wallin, and Geyong Min. 2019. “New Methods for Clustering District Heating Users Based on Consumption Patterns.” *Applied Energy* 251:113373. https://doi.org/10.1016/j.apenergy.2019.113373.

Webler, T., Danielson, S., & Tuler, S. 2009. “Using Q Method to Reveal Social Perspectives in Environmental Research.” https://www.researchgate.net/profile/Stentor-Danielson/publication/273697977_Using_Q_Method_to_Reveal_Social_Perspectives_in_Environmental_Research/links/582a4e1608aef19cb805583d/Using-Q-Method-to-Reveal-Social-Perspectives-in-Environmental-Research.pdf. Accessed June 03, 2021.

Website of the Federal Government | Bundesregierung. 2022. “EU Energy Council: Reducing Prices, Securing Supplies | Federal Government.” Accessed December 30, 2022.

https://www.bundesregierung.de/breg-en/search/eu-gas-price-cap-2155126.

Wehrle, Sebastian, and Johannes Schmidt. 2018. "District Heating Systems Under High CO2 Emission Prices: The Role of the Pass-Through from Emission Cost to Electricity Prices." https://arxiv.org/pdf/1810.02109.

Weiner, Csaba, and Tekla Szép. 2022. "The Hungarian Utility Cost Reduction Programme: An Impact Assessment." *Energy Strategy Reviews* 40:100817. https://doi.org/10.1016/j.esr.2022.100817.

Werner, Sven. 2016. *European District Heating Price Series.* Print. Accessed November 21, 2021. https://energiforskmedia.blob.core.windows.net/media/21926/european-district-heating-price-series-energiforskrapport-2016-316.pdf.

William Stephenson. 1955. "The Study of Behavior: Q-Technique and Its Methodology." *American Journal of Sociology* 61 (2): 167–69. https://doi.org/10.1086/221704.

Wohllebe. 2020. "Consumer Acceptance of App Push Notifications: Systematic Review on the Influence of Frequency." *Int. J. Interact. Mob. Technol.* 14 (13): 36. https://doi.org/10.3991/ijim.v14i13.14563.

Wohllebe, Atilla, Dirk-Siegfried Hübner, Uwe Radtke, and Szilárd Podruzsik. 2021a. "Mobile Apps in Retail: Effect of Push Notification Frequency on App User Behavior." *Innovative Marketing* 17 (2): 102–11. https://doi.org/10.21511/im.17(2).2021.10.

Wohllebe, Atilla, Dirk-Siegfried Hübner, Uwe Radtke, and Arabel Wohllebe. 2021b. "Smartphones and Mobile Apps: Case Study on Usage Behavior of Elementary School Students – Insights from a Rural

Area in Northern Germany." *Int. J. Interact. Mob. Technol.* 15 (12): 184. https://doi.org/10.3991/ijim.v15i12.22565.

Xu, Baoping, Lin Fu, and Hongfa Di. 2009. "Field Investigation on Consumer Behavior and Hydraulic Performance of a District Heating System in Tianjin, China." *Building and Environment* 44 (2): 249–59. https://doi.org/10.1016/j.buildenv.2008.03.002.

Yamamoto, Michio, and Robert I. Jennrich. 2013. "A Cluster-Based Factor Rotation." *The British journal of mathematical and statistical psychology* 66 (3): 488–502. https://doi.org/10.1111/bmsp.12007.

Yeun, Eun Ja. 2021. "Q-Methodology: Understanding and Applying It to Nursing and Health Research." *Nursing & health sciences* 23 (4): 777–81. https://doi.org/10.1111/nhs.12744.

Yoon, Taeyeon, Yongsun Ma, and Charles Rhodes. 2015. "Individual Heating Systems Vs. District Heating Systems: What Will Consumers Pay for Convenience?" *Energy Policy* 86:73–81. https://doi.org/10.1016/j.enpol.2015.06.024.

Zabala A. 2014. "Qmethod: A Package to Explore Human Perspectives Using Q Methodology." *The R Journal* 6 (2): 163–73. https://journal.r-project.org/archive/2014-2/zabala.pdf. Accessed March 10, 2022.

Zhang, Ge, and Xu. 2013. "An Equivalent Marginal Cost-Pricing Model for the District Heating Market." *Energy Policy* 63:1224–32. https://doi.org/10.1016/j.enpol.2013.09.017.

Addendum

A.1 R Script for K-Means cluster analysis

```
---
title: "Clusteranalysis 3"
author: "Uwe Radtke"
date: "01/07/2021"
output: word_document
---
```{r include=FALSE}
# load functions and packages
library(data.table)
library(psych)
library(class)
library(caTools)
library(e1071)
library(ggplot2)
library(tidyverse)
library(haven)
library(dplyr)
library(sjlabelled)
library(labelled)
library(factoextra)
library(cluster)
```
Data

```{r}
df <- as.data.table(read_dta(file = "measures_corr.dta"))
```

Cleaning

Data Cleaning, Outlier Handling (Winsorizing)
```{r}
df <- df[, c("datetime", "cons_id", "op_hrs", "P1_numeric")]
colnames(df)[1] <- "Date"
colnames(df)[2] <- "ID"
colnames(df)[3] <- "Operating_Hours"
colnames(df)[4] <- "P1"
```
Einzigartige Geräte
```{r}
length(unique(df$ID))
```

Runde Datetime auf volle Stunde

```{r}
df$Date <- format(round(df$Date, units="hours"), format="%Y-%m-%d
%H:%M")
```

Einzigartige Zeitpunkte

```{r}
length(unique(df$Date))
```
```

````
MinDate und MaxDate

```{r}
min(df$Date)
max(df$Date)
```

Im Datensatz sind etwas mehr als 365 Tage zu den Messdaten, daher
existieren mehr als 365*24 einzigartige Daten.

Verwerfe Duplicates

```{r}
a <- nrow(df)
df <- df[!duplicated(df), ]
b <- nrow(df)
a - b
```

288310 doppelte Reihen wurden bereinigt.

Bilde aus einzigartigen Zeitpunkten und IDs einen Cross-Join, um den
Rahmen des DataFrames zu vervollständigen

```{r}
Date <- unique(df$Date)
ID <- unique(df$ID)
dt <- CJ(Date, ID)
a <- nrow(dt)
b <- nrow(df)
a - b
```

60745 Date/ID Kombinationen fehlen und werden im cross-joined data
table mit NA aufgefüllt.

```{r}
df <- remove_all_labels(df)
dt <- merge(dt, df, by = c("Date", "ID"), all = TRUE)
```

Setze negative Werte und 0 zu NA

```{r}
dt[Operating_Hours <= 0] <- dt[Operating_Hours <= 0][,
Operating_Hours := NA]
dt[P1 <= 0] <- dt[P1 <= 0][, P1 := NA]
df <- dt
df <- remove_attributes(df, "format.spss")
df <- remove_attributes(df, "format.stata")
```

Bereinige Outlier zum 95%-Percentil

```{r}
df$Operating_Hours <- winsor(df$Operating_Hours, trim = 0.05)
df$P1 <- winsor(df$P1, trim = 0.05)
```
````

````
We visualize the histograms again to see the effect of the
transformation:

```{r}
df_metrical <- df[, c(3,4)]
ggplot(gather(df_metrical), aes(value)) +
    geom_histogram(bins = 7, fill = "steelblue") +
    facet_wrap(~key, scales = 'free_x')
```
P1 looks log-normal now and this help with the modeling.

```{r}
colnames(df)[4] <- "logP1"
```



```{r}
meltData <- melt(df[, c(3,4)])
par(cex.axis=0.5)
#boxplot(data=meltData, value~variable, las = 2, xlab = "")

p <- ggplot(meltData, aes(factor(variable), value))
p + geom_boxplot() + facet_wrap(~variable, scale="free")
```

#### Additional stats by groups

```{r}
#describeBy(df$Operating_Hours, df$ID)
```

```{r}
#describeBy(df$logP1, df$ID)
```

#### Modeling

- Clusteranalyse K-Means

```{r}
# Zeilen mit fehlenden Werten entfernen
df <- na.omit(df)
```

```{r}
# Normalisierung des Datensatzes
df[,c(3,4)] <- lapply(df[,c(3,4)], scale)
#df <- df[!is.finite(rowSums(df)),]
df <- df[, c(3,4)]
```

```{r}
rm(meltData)
rm(df_metrical)
rm(p)
```
````

````
Subsample df to n rows
Mit setzen des seeds auch beim subsampling forciert man den gleichen
output bei Mehrmaligem durchlaufen.

```{r}
set.seed(1337)
df <- sample_n(df, 10000)
```
```{r}
set.seed(1337)
# Beispieldurchlauf mit k = 5
kmeans(df, 5)
```
```{r}
# Ellbogenmethode zur Findung eines optimalen k
fviz_nbclust(df, kmeans, method = "wss")
```
```{r}
# Lückenstatistik zur Findung eines optimalen k
gap_stat <- clusGap(df,
                    FUN = kmeans,
                    nstart = 25,
                    K.max = 5,
                    B = 50)
# Plotten der Anzahl der Cluster vs. Lückenstatistik
fviz_gap_stat(gap_stat)
```

```{r}
# Durchführen von k-Means Clustering mit optimalen k = 4
km <- kmeans(df, centers = 4, nstart = 25)

# Ergebnisse anzeigen
km
```



```{r}
# Plotten der Ergebnisse des endgültigen k-means-Modells
fviz_cluster(km, data = df)
```

Retransformation der Cluster-Zentrum

```{r}
centers <- as.data.table(km$centers)
centers
```
```{r}
centers_retransformed <- centers
centers_retransformed$logP1 <- exp(centers_retransformed$logP1) - 1
colnames(centers_retransformed)[2] <- "P1"
centers_retransformed$Operating_Hours <-
centers_retransformed$Operating_Hours * sd_OH + mean_OH
centers_retransformed$P1 <- centers_retransformed$P1 * sd_P1 + mean_P1
centers_retransformed
```
````

A.2 R-Script for Q-Method analysis

````
---
title: "Q-Method Analysis"
author: "Uwe Radtke"
date: "12/20/2022"
output: word_document
###source and idea: https://aiorazabala.github.io/qmethod/Cookbook
---
### load functions and packages
```{r include=FALSE}
library(qmethod)
```
### Data
```{r}
setwd("C:/Users/D041144/OneDrive - SAP
SE/Daten/privat/Studium/Thesis/actual/Q-method/data/K2")
mydata <- read.csv("Q_REPLY_KAPOSVAR.csv")
```
###Data Analysis
Data Analysis - Dimensions
```{r}
dim(mydata)
```
Correlation Matrix
```{r}
cor(mydata)
```
### Set distribution
```{r}
distri<-c(rep(-4, 2), rep(-3, 3), rep(-2, 5), rep(-1, 6), rep( 0, 7),
rep( 1, 6), rep( 2, 5), rep( 3, 3), rep( 4, 2))
```
### Run the analysis
### Create an object called 'results', and put the output
### of the function 'qmethod()' into this object
### (replace the number of factor 'nfactors' as necessary)
Calculate results
test with rotation
```{r}
results <- qmethod(mydata, nfactors = 4, extraction="PCA", rotation =
"cluster", forced = FALSE, distribution = distri, cor.method
="spearman", silent = FALSE)
```
###Eigenvalues, total explained variability, and number of Q-sorts
significantly loading
```{r}
results$f_char$characteristics
```
### See the factor loadings
```{r}
round(results$loa, digits = 2)
```
### See the flagged Q-sorts: those indicating 'TRUE'
Results
```{r}
results$flag
```
````

```
### # Print the table of factor loadings with an indication of
flags,
### as extracted from the object results$flag:
```{r}
loa.and.flags(results)
summary(results)
plot(results)
```
###Decide upon the number of factors to extract
```{r}
results$f_char$characteristics
screeplot(prcomp(mydata), main = "Screeplot of unrotated factors",
          type = "l")
```
Summary: general characteristics and factor scores
Summary: general characteristics and factor scores
```{r}
summary(results)
results
plot(results)
```
Reorder the statements from highest to lowest scores for each
factor
Put z-scores and factor scores together
Reorder the statements from highest to lowest scores for each
factor
Put z-scores and factor scores together
```{r}
scores <- cbind(round(results$zsc, digits=2), results$zsc_n)
nfactors <- ncol(results$zsc)
col.order <- as.vector(rbind(1:nfactors, (1:nfactors)+nfactors))
scores <- scores[col.order]
scores
```
invlid sorting - try with another later
scores[order(scores$zsc_f1, decreasing = T),]

Explore the table of distinguishing and consensus statements
Explore the table of distinguishing and consensus statements
```{r}
results$qdc
results$qdc[which(results$qdc$dist.and.cons == "Consensus"), ]
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes
all"), ]
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes f1
only"), ]
```

export results
```{r}
write.csv(results$zsc,   file = "zscores.csv")
write.csv(results$zsc_n, file = "factorscores.csv")
write.csv(results$loa,   file = "loadings.csv")
export.qm(results, file = "myreport.txt", style = "R")
export.qm(results, file = "myreport-pqm.txt", style = "PQMethod")
```
```

A.3 Q-Questionaires

A3.1 In English language

Dear participants,

as part of my dissertation at MATE Kaposvár Campus, Hungary, I am conducting a survey on the topic of Urban Energy Systems - Analysis based on the smart meters of the district heating system in Kaposvár. To better understand and interpret the findings from the literature, I would like to ask you to answer the statistical questions as far as possible and fill in the questionnaire or sort the corresponding statements according to personal preferences.

District heating is the supply of heat to buildings from a power or heating plant. The heat generated there reaches you through a pipe system. District heating customers therefore do not need their own heating system at home. Different materials and processes are used as fuels, e.g., natural gas and wood chips in the case of Kaposvár.

Objective: to determine the preferences of residents or users of district heating in terms of sustainability, digitalization of district heating and opinion on district heating in general.

Filling out the sheet with the 39 statements should take about 30 minutes, and your answers will be evaluated completely anonymously.
If you have any questions about the survey, please email me: radtke.uwe@phd.uni-mate.hu.

Thank you very much.
With kind regards
Uwe Radtke

Statistical data query – please mark only one answer:

1) You are ☐ tenant ☐ owner.
2) In which quartile do you think you are fitting:
   ☐ High demand & for a short period of time (e.g., some hours a day for only a few month)
   ☐ High demand & long period
   ☐ Low demand & for a short period
   ☐ Low demand & long period
3) You are ☐ employed ☐ self-employed ☐ without paid employment.
4) You are between ☐ <30 years old ☐31 to 45 years old ☐46 to 60 years old ☐ 61 years to 75 years old ☐ > 75 years old
5) You are ☐alone ☐married/living in a partnership.
6) How many persons live in your household? ☐ 1 ☐ 2 or 3 ☐ 4 or 5 ☐ 6 or 7 ☐ 8 or more persons living in the household.
7) You own and use mobile devices (smartphone, tablet, …) ☐ Yes ☐ No
8) You use electronic media for bank transfers ☐ Yes ☐ No
9) You prefer to use ☐ PC, Notebook, Tablet ☐ Smart Phone ☐ none of the before
10) What is your highest educational degree? ☐undergraduate ☐ skilled/specialist worker ☐High school ☐Vocational school, with baccalaureate ☐Higher education (College,- BSc degree) ☐University, Master's degree, Bachelor's degree ☐PhD and above

Disagree		Agree
Least preferred	⟷	Most preferred
Least important		Most important

-4	-3	-2	-1	0	1	2	3	4

To fill in, place each number of the statements in a separate box, depending on the agreement, disagreement of your preference. If you agree with a statement, the higher the level of agreement, the statement should be placed further to the right. Statement to which you are neutral - have to be placed into the middle columns. The numbers of the statements you disagree, has to be placed left, below the negative column headings. With one statement there is some additional explanation, which is to explain the statement possibly somewhat more nearby. This serves only the understanding and are not a component of the statement.

1) It is mainly the space savings that encourage many customers to buy district heating.	2) No maintenance costs, no responsibility: there is no heat production, no firing inside the properties of the district heating users, therefore no chimney is required.	3) The district heating customers does not have to worry about the procurement, pre-financing and storage capacities for the fuels - they get the heat "ready" delivered to their house.
4) Simple operation: In terms of comfort, district heating is probably the simplest way of housing temperature control.	5) Combined heat and power (CHP) make optimum use of fuels.	6) Wherever possible, a biomass-based CHP plant should be used.
7) More digitalization leads to more environmentally friendly generation processes (e.g., through education, visualization).	8) District heating should also be available in sparsely populated areas. (Preferable from local resource.)	9) No more research and investment should be made in district heating.
10) State-regulated prices prevent expensive investments in sustainable technologies for heat generation.	11) District heating offers price stability.	12) The state should intervene even more in the energy market and make and enforce regulations regarding the technologies used.
13) Security of supply is always the top priority in a district heating system.	14) District heating supports (large) investments in renewable energies.	15) The investments in thermal insulation of buildings should also be supported like the investments in sustainable CO_2 free production of thermal energy.
16) I would be very happy to hear comparative statements, such as: "With this energy, you can drive your 5-year-old car to Brussels and back."	17) The comparable average consumption should always be indicated with the own consumption.	18) I would be willing to spend more money on district heating in the case of proven sustainable production.
19) An app should always have several divisions integrated (district heating, electricity, water, gas, ...).	20) I would be willing to change my habits (turn off the heating on vacation, ventilate now and then instead of constantly ventilate, ...).	21) Grid- and system-serving consumer behavior should be rewarded.
22) Weather data must be integrated into the app in order to control the heating specifically.	23) I would trust an app enough to let it regulate my heating on its own.	24) Even when I am not at home, I need information about heat consumption. I would like to have the consumption transmitted via the Internet.
25) If an app suggests savings measures, e.g., home insulation, different thermostat, ... I would take the appropriate steps to implement the recommendations.	26) A smartphone app must offer/suggest the cheapest rate for me depending on consumption.	27) I always need a paper invoice/bill.
28) It is good for consumers if the government regulates energy prices, even if it means that companies may invest less in renewable energy production.	29) I would rather heat with another energy source than with district heating.	30) Environmentally friendly heating and ventilation does not require financial incentives.
31) Users of district heating are becoming more environmentally aware	32) Laws and the technical systems must be adapted so that data	33) The taxation of energy products in Hungary / in the EU must be more

and just need the right information to act in an environmentally conscious way.	protection and data security are ensured, and consumption data cannot be passed on.	closely aligned with climate protection aspects.
34) No basic charge - district heating prices should be charged exclusively according to consumption.	35) The state should let the market regulate prices (market economy).	36) The European Union should become even more involved in the energy industry.
37) Only sufficient financial incentives would induce me to change my behavior with regard to heating and ventilation.	38) District heating is the most expensive form of heating.	39) Consumption and billing via app (if available) must be displayed graphically. Explanation: see below Figure 1

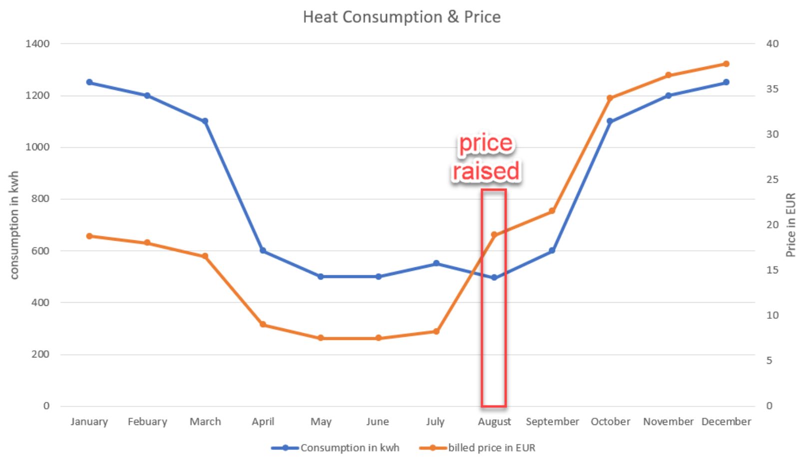


Figure 24: Example for graphical display, own creation, no real values, just an example
Blue – consumption in [kW]   Red – Billed prices in [Euro]
Price raised in August for the rest of the year – highlighted with red rectangle

A3.2 In Hungarian language

Kedves Kitöltő!

A MATE Kaposvári Campusán készülő disszertációm keretében felmérést készítek a “URBAN ENERGY SYSTEMS - ANALYSIS OF SMART METERING DISTRICT HEATING SYSTEM OF KAPOSVÁR” témakörben. A szakirodalomból származó megállapítások jobb megértése és értelmezése érdekében szeretném megkérni Önt, hogy töltse ki a kérdőívemet. A statisztikai kérdések megválaszolása után személyes preferenciái alapján rendezze sorrendbe az állításokat.
A távfűtés az épületek hőellátása egy erőműből vagy hőerőműből. Az ott megtermelt hő egy csőrendszeren keresztül jut el Önhöz, tehát nincs szüksége saját otthoni fűtési rendszerre. A távfűtéshez tüzelőanyagként különböző anyagokat és eljárásokat használnak, Kaposvár esetében földgázt, faaprítékot.
A kérdőív célja a lakosok, illetve a távfűtést használók preferenciáinak meghatározása a fenntarthatóság, a távfűtés digitalizálása és általában, a távfűtésről alkotott vélemény szempontjából.
A 39 állítást tartalmazó felmérés kitöltése körülbelül 30 percet vesz igénybe, és a válaszokat teljesen anonim módon értékelem.
Ha bármilyen kérdése van a kérdőívvel kapcsolatban, kérem, írjon e-mailt a következő címre: radtke.uwe@phd.uni-mate.hu.
Köszönettel és tisztelettel,
Uwe Radtke

Statisztikai adatok - kérem, csak egy választ jelöljön meg:

1) Ön ☐ bérlő ☐ tulajdonos.
2) Ön szerint melyik csoportba tartozik:
   ☐ Magas fűtés & rövid ideig (rövid üzemidő) –&
   ☐ Magas fűtés & hosszú ideig –
   ☐ Alacsony fűtés & rövid ideig
   ☐ Alacsony fűtés és hosszú ideig
3) Ön ☐ alkalmazott ☐ egyéni vállalkozó ☐ állás nélküli.
4) Az Ön kora
   ☐ <30 éves
   ☐31-45 éves
   ☐46-60 éves
   ☐ 61-75 éves
   ☐ > 75 éves
5) Ön ☐egyedülálló ☐házas/élettársi kapcsolatban élő.
6) Hány személy él az Ön háztartásában? ☐ 1 ☐ 2 vagy 3 ☐ 4 vagy 5 ☐ 6 vagy 7 ☐ 8 vagy többszemély
7) Ön rendelkezik és használ mobil eszközöket (okostelefon, táblagép, ...) ☐ Igen ☐Nem
8) Ön elektronikus médiát használ banki átutalásokhoz ☐ Igen ☐Nem

9) Ön melyik elektronikus médiát használja inkább? ☐ számítógépet, notebookot, táblagépet ☐ okostelefont ☐ egyiket sem

1) Mi a legmagasabb iskolai végzettsége? ☐Szakképzetlen ☐Szakmunkás ☐Középiskola ☐Szakközépiskola, érettségivel ☐Felsőfokú végzettség (Főiskolai,- BSc diploma) ☐Egyetem, mesterképzés, osztatlan egyetemi képzés ☐Phd illetve a fölött

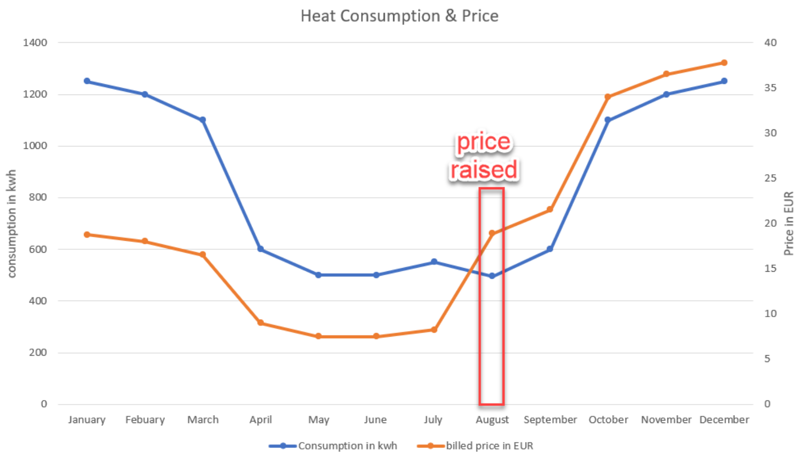
Nem értek egyet	Egyetértek
Legkevésbé preferált	Leginkább preferált
Legkevésbé fontos	Legfontosabb

-4	-3	-2	-1	0	1	2	3	4

A Q módszernél a kitöltés technikája: a fenti 11 oszlopos rácshálóban a bal oldali, (negatív számozású) oszlopokba kell tenni az alábbi 39 állítás közül azoknak a számait, amelyekkel legkevésbé ért egyet a kitöltő, jobb oldali, pozitív feliratú oszlopokba pedig amellyel a leginkább egyetért, tartva a fenti piramis struktúrát, azaz az oszlopokbeli cellák előre megadott számát. Így középre kerülnek a kitöltő számára leginkább semleges állítások (számai). Az oszlopon belüli a soroknak nincs jelentősége, (azonos oszlopba kerülő állításokkal azonos mértékben ért egyet a válaszadó). Az utolsó állításnál szereplő magyarázat csak a megértést szolgálja, nem alkotóeleme az állításnak, nem kell a preferencia elrendezésben figyelembe venni.

1) Elsősorban a helytakarékosság az, ami sok ügyfelet távfűtés vásárlására ösztönöz.	2) Nincs fenntartási költség, nincs felelősség: a távhőfelhasználók ingatlanjain belül nincs hőtermelés, nincs tüzelés, ezért nincs szükség kéményre.	3) A távhő fogyasztónak nem kell aggódnia a tüzelőanyagok beszerzése, előfinanszírozása és tárolása miatt - a hőt "készen" kapja házhoz szállítva.
4) Egyszerű működés: A kényelem szempontjából a távfűtés valószínűleg a legegyszerűbb módja a lakáshőmérséklet szabályozásának.	5) A kapcsolt hő- és villamosenergia-termelés optimálisan hasznosítja a tüzelőanyagokat.	6) Ahol csak lehetséges, biomassza alapú, hőt és villamos energiát kapcsoltan termelő erőművet kell használni.
7) A fokozottabb digitalizáció környezetbarátabb termelési folyamatokhoz vezet (pl. oktatás, vizualizáció révén).	8) A távfűtésnek a ritkán lakott területeken is elérhetőnek kell lennie (preferáltan helyben elérhető forrásból).	9) Nem szabad több kutatást és beruházást végezni a távfűtés területén.
10) Az államilag szabályozott árak megakadályozzák a hőtermelés fenntartható technológiáiba történő drága beruházásokat.	11) A távfűtés árstabilitást kínál.	12) Az államnak még jobban be kellene avatkoznia az energiapiacba, és az alkalmazott technológiákra vonatkozó előírásokat kellene megalkotnia és betartatnia.
13) A távfűtési rendszerben mindig az ellátás biztonsága a legfontosabb prioritás.	14) A távfűtés támogatja a megújuló energiákba történő (nagy) beruházásokat.	15) Az épületek hőtechnikai korszerűsítésére irányuló beruházásokat ugyanúgy támogatni kell, mint a fenntartható CO_2-mentes hőenergia-termelésre irányuló beruházásokat.
16) Nagyon szívesen hallanék összehasonlító nyilatkozatokat, például: " “Ennyi energiával elmehetsz az 5 éves autóddal Brüsszelbe és vissza.”	17) Az összehasonlítható átlagfogyasztást mindig a saját fogyasztással együtt kell feltüntetni.	18) Fenntartható termelés esetén hajlandó lennék több pénzt költeni a távfűtésre.
19) Egy az energia felhasználóknak szánt alkalmazásnak mindig több területet kell integrálnia (távfűtés, villamos energia, víz, gáz, ...).	20) Hajlandó lennék változtatni a szokásaimon (kikapcsolni a fűtést nyaraláskor, időközönként szellőztetni az állandó szellőztetés helyett, ...).	21) A hálózatot és a rendszert támogató fogyasztói magatartást jutalmazni kell.
22) A fűtés vezérléséhez használt applikációba az időjárási adatokat is integrálni kell.	23) Megbíznék egy alkalmazásban annyira, hogy hagynám, hogy egyedül szabályozza a fűtésemet.	24) Még akkor is szükségem van a hőfogyasztással kapcsolatos információkra, amikor nem vagyok otthon. Szeretném, ha a fogyasztási adatokat az interneten keresztül is elérhetném.
25) Ha egy alkalmazás takarékossági intézkedéseket javasol (például a lakás szigetelése, más termosztát, stb.) megtenném a megfelelő lépéseket az ajánlások megvalósítására.	26) Az okostelefonos alkalmazásnak a fogyasztástól függően a legolcsóbb díjat kell felajánlania/javasolnia számomra.	27) Mindig papíralapú számlára van szükségem a fűtési fogyasztásomról.
28) A fogyasztóknak jó, ha a kormány szabályozza az energiaárakat, még akkor is, ha ez azt jelenti, hogy a vállalatok esetleg kevesebbet fektetnek be a	29) Inkább más energiaforrással fűtenék, mint távfűtéssel.	30) A környezetbarát fűtés és szellőztetés nem igényel pénzügyi ösztönzőket.

megújuló energiatermelésbe.		
31) A távfűtés felhasználói egyre környezettudatosabbak, és csak a megfelelő információkra van szükségük ahhoz, hogy még környezettudatosabban cselekedjenek.	32) A jogszabályokat és a technikai rendszereket úgy kell átalakítani, hogy az adatvédelem és az adatbiztonság garantált legyen, és a fogyasztási adatok ne legyenek továbbadhatók.	33) Az energiatermékek adóztatását Magyarországon/az EU-ban jobban össze kell hangolni az éghajlatvédelmi szempontokkal.
34) Ne legyen alapdíj - a távfűtés árát kizárólag a fogyasztás alapján kellene számítani.	35) Az államnak hagynia kellene, hogy a piac szabályozza az árakat (piacgazdaság).	36) Az Európai Uniónak még jobban be kellene kapcsolódnia az energiaiparba.
37) Csak megfelelő pénzügyi ösztönzők késztetnének arra, hogy megváltoztassam a fűtéssel és szellőztetéssel kapcsolatos viselkedésemet.	38) A távfűtés a fűtés legdrágább formája.	39) A fogyasztást és az alkalmazáson keresztüli számlázást (ha van) grafikusan kell megjeleníteni. Magyarázat: lásd az alábbi 1. ábrát



1. ábra: Példa grafikus megjelenítésre. Saját készítésű ábra példaként kitalált értékekkel,

***kék** – a fogyasztás [kW]-ban* ***piros** – a kiszámlázott összeg [Euro]-ban*

Mj:- augusztusban az év hátralévő részére is érvényes áremelés történt. (Kiemelve a piros téglalappal.)

A.4 Prisma flow chart on pricing mechanisms

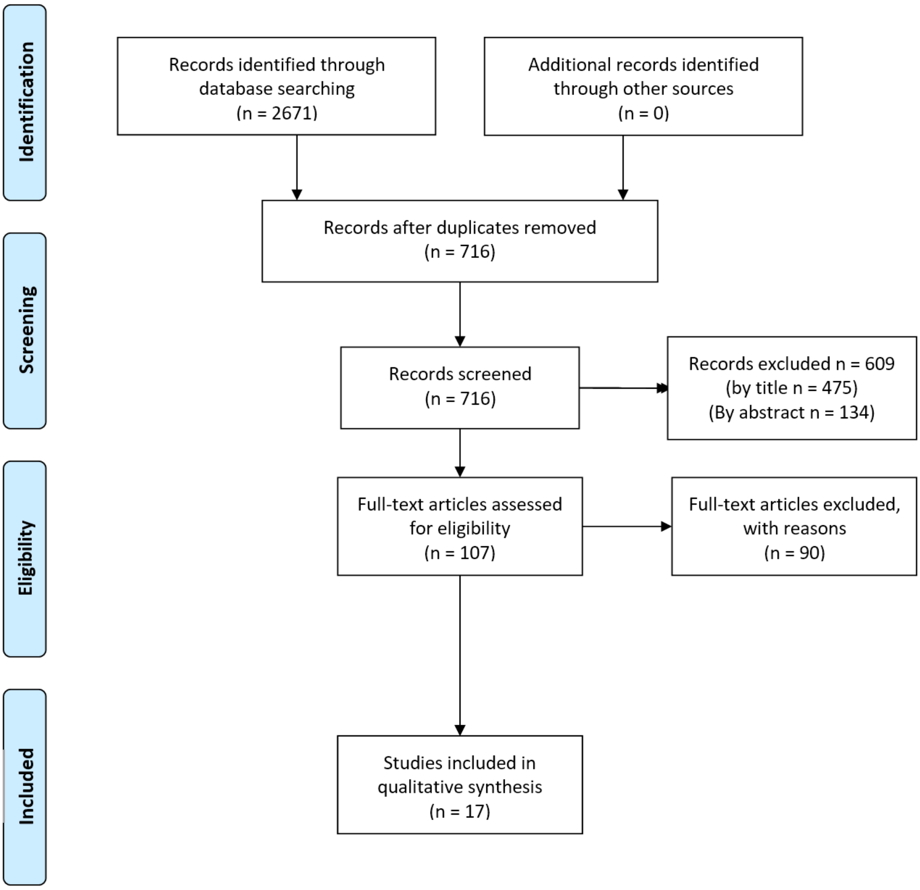


Figure 25: Prisma flow chart on pricing mechanisms

A.5 Prisma flow chart in consumer behavior

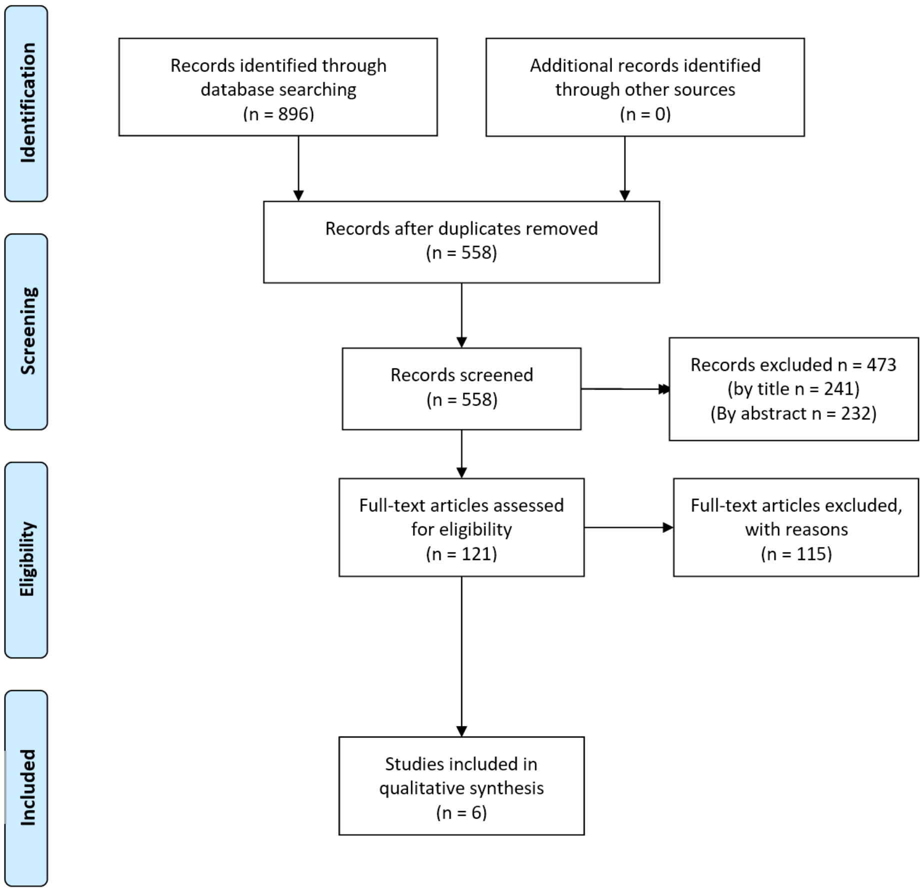


Figure 26: Prisma flow chart in consumer behavior

A.6 PQMethod-like output

```
Q-method analysis.
Finished on:               Mon Jun 26 14:06:10 2023
'qmethod' package version: 1.8
Original data:             39 statements, 21 Q-sorts
Forced distribution:       FALSE
Number of factors:         4
Extraction:                PCA
Rotation:                  cluster
Flagging:                  automatic
Correlation coefficient:   spearman

Correlation Matrix Between Sorts
         P01     P02    P03     P04    P05     P06     P07     P08    P09
P01    1.0000  0.4000 0.5000  0.2824 0.1471  0.5176  0.3588  0.3412 0.5412
P02    0.4000  1.0000 0.4118 -0.0059 0.3118  0.4412  0.3941  0.0059 0.2765
P03    0.5000  0.4118 1.0000  0.1235 0.4588  0.4059  0.2529  0.1529 0.3529
P04    0.2824 -0.0059 0.1235  1.0000 0.0118 -0.0471 -0.0941  0.0882 0.3059
P05    0.1471  0.3118 0.4588  0.0118 1.0000  0.6471  0.1529  0.3706 0.4059
P06    0.5176  0.4412 0.4059 -0.0471 0.6471  1.0000  0.4294  0.4941 0.5471
P07    0.3588  0.3941 0.2529 -0.0941 0.1529  0.4294  1.0000  0.1176 0.2941
P08    0.3412  0.0059 0.1529  0.0882 0.3706  0.4941  0.1176  1.0000 0.4235
P09    0.5412  0.2765 0.3529  0.3059 0.4059  0.5471  0.2941  0.4235 1.0000
P10    0.3000  0.3706 0.4529  0.1647 0.1824  0.2412  0.3647  0.0118 0.3529
P11    0.3588  0.2824 0.4706 -0.0588 0.2706  0.2765  0.3765  0.5235 0.3471
P12    0.3647  0.3176 0.4412  0.0059 0.3118  0.3294  0.3647  0.2353 0.2941
P13    0.1118  0.2294 0.3588 -0.0882 0.1765  0.1059  0.3235  0.1824 0.0824
P14    0.2294 -0.0529 0.3235  0.2412 0.2529  0.1118  0.0824 -0.1941 0.3353
P15    0.2941  0.1647 0.5353 -0.1412 0.1765  0.0588  0.1824  0.0294 0.0412
P16    0.3353  0.3118 0.5059 -0.0353 0.3647  0.3059  0.3294  0.1471 0.3412
P17    0.3294  0.2706 0.4353  0.0412 0.2294  0.2765  0.1588  0.1588 0.3235
P18    0.5000  0.2118 0.6118  0.0765 0.3353  0.4471  0.3176  0.1941 0.5176
P19    0.1000  0.2000 0.3353  0.0824 0.2000  0.1471  0.2471 -0.0941 0.3118
P20    0.4176  0.3412 0.4588  0.1235 0.2412  0.3235  0.1529  0.2824 0.2294
P21    0.1118  0.1529 0.3353 -0.0471 0.3353  0.1353  0.0706  0.0059 0.1824
         P10     P11    P12     P13     P14     P15     P16    P17    P18
P01    0.3000  0.3588 0.3647  0.1118  0.2294  0.2941  0.3353 0.3294 0.5000
P02    0.3706  0.2824 0.3176  0.2294 -0.0529  0.1647  0.3118 0.2706 0.2118
P03    0.4529  0.4706 0.4412  0.3588  0.3235  0.5353  0.5059 0.4353 0.6118
P04    0.1647 -0.0588 0.0059 -0.0882  0.2412 -0.1412 -0.0353 0.0412 0.0765
P05    0.1824  0.2706 0.3118  0.1765  0.2529  0.1765  0.3647 0.2294 0.3353
P06    0.2412  0.2765 0.3294  0.1059  0.1118  0.0588  0.3059 0.2765 0.4471
P07    0.3647  0.3765 0.3647  0.3235  0.0824  0.1824  0.3294 0.1588 0.3176
P08    0.0118  0.5235 0.2353  0.1824 -0.1941  0.0294  0.1471 0.1588 0.1941
P09    0.3529  0.3471 0.2941  0.0824  0.3353  0.0412  0.3412 0.3235 0.5176
P10    1.0000  0.3765 0.4235  0.5118  0.2824  0.3412  0.5471 0.5706 0.4235
P11    0.3765  1.0000 0.6706  0.6412 -0.0824  0.4000  0.5353 0.4294 0.3588
P12    0.4235  0.6706 1.0000  0.6882  0.0529  0.2412  0.4176 0.3235 0.4000
P13    0.5118  0.6412 0.6882  1.0000  0.0000  0.3176  0.4706 0.4294 0.3765
P14    0.2824 -0.0824 0.0529  0.0000  1.0000  0.3588  0.2941 0.2529 0.3824
P15    0.3412  0.4000 0.2412  0.3176  0.3588  1.0000  0.6353 0.6353 0.3647
P16    0.5471  0.5353 0.4176  0.4706  0.2941  0.6353  1.0000 0.8353 0.5529
P17    0.5706  0.4294 0.3235  0.4294  0.2529  0.6353  0.8353 1.0000 0.3765
P18    0.4235  0.3588 0.4000  0.3765  0.3824  0.3647  0.5529 0.3765 1.0000
P19    0.5176  0.3059 0.4294  0.4118  0.2765  0.3412  0.5941 0.5824 0.4059
P20    0.4588  0.4471 0.6000  0.6000 -0.1118  0.1706  0.4000 0.3882 0.3176
P21    0.1412  0.2059 0.2647  0.0706  0.3588  0.3765  0.2235 0.0882 0.1882
          P19     P20     P21
P01     0.1000  0.4176  0.1118
P02     0.2000  0.3412  0.1529
P03     0.3353  0.4588  0.3353
P04     0.0824  0.1235 -0.0471
P05     0.2000  0.2412  0.3353
P06     0.1471  0.3235  0.1353
P07     0.2471  0.1529  0.0706
P08    -0.0941  0.2824  0.0059
P09     0.3118  0.2294  0.1824
P10     0.5176  0.4588  0.1412
P11     0.3059  0.4471  0.2059
P12     0.4294  0.6000  0.2647
P13     0.4118  0.6000  0.0706
P14     0.2765 -0.1118  0.3588
P15     0.3412  0.1706  0.3765
P16     0.5941  0.4000  0.2235
P17     0.5824  0.3882  0.0882
P18     0.4059  0.3176  0.1882
P19     1.0000  0.2471  0.1765
P20     0.2471  1.0000 -0.0235
P21     0.1765 -0.0235  1.0000

Unrotated Factor Matrix
Loadings:
       PC1    PC2    PC3    PC4    PC5    PC6    PC7    PC8
```

```
P01     0.611  0.366  0.207  0.242 -0.075 -0.112  0.488  0.120
P02     0.509  0.180 -0.071 -0.065 -0.600  0.018  0.167 -0.360
P03     0.750 -0.029  0.173 -0.080  0.015  0.160  0.301 -0.186
P04     0.094  0.170  0.359  0.713  0.199  0.264  0.055 -0.085
P05     0.528  0.313  0.237 -0.413  0.109  0.124 -0.350 -0.356
P06     0.590  0.586  0.159 -0.221 -0.173 -0.181 -0.168 -0.117
P07     0.496  0.143 -0.106 -0.087 -0.550 -0.106 -0.033  0.492
P08     0.355  0.635 -0.181 -0.098  0.519 -0.199 -0.032  0.110
P09     0.593  0.416  0.375  0.200  0.056 -0.057 -0.220  0.171
P10     0.684 -0.259 -0.031  0.309 -0.205  0.043 -0.124 -0.047
P11     0.705  0.081 -0.425 -0.126  0.249  0.042  0.090  0.241
P12     0.702  0.033 -0.370 -0.005  0.013  0.415 -0.056  0.133
P13     0.624 -0.228 -0.559  0.055  0.069  0.230 -0.111  0.096
P14     0.321 -0.301  0.735  0.038  0.046  0.151 -0.057  0.162
P15     0.564 -0.509  0.111 -0.282  0.176 -0.210  0.390  0.008
P16     0.792 -0.341  0.010 -0.048  0.087 -0.312 -0.075 -0.063
P17     0.708 -0.367 -0.003  0.114  0.159 -0.433 -0.047 -0.199
P18     0.707  0.019  0.241  0.054  0.019 -0.023 -0.001  0.205
P19     0.586 -0.435  0.037  0.165 -0.075 -0.008 -0.443  0.021
P20     0.614  0.130 -0.424  0.279  0.066  0.172  0.117 -0.327
P21     0.328 -0.156  0.348 -0.512  0.087  0.485  0.110  0.053

                 PC1   PC2   PC3   PC4   PC5   PC6   PC7   PC8
SS loadings    7.277 2.191 2.017 1.435 1.216 1.057 0.985 0.919
Proportion Var 0.347 0.104 0.096 0.068 0.058 0.050 0.047 0.044
Cumulative Var 0.347 0.451 0.547 0.615 0.673 0.723 0.770 0.814


Cumulative Communalities Matrix
       X1     X2     X3     X4     X5     X6     X7     X8
1  0.3733 0.5072 0.5501 0.6086 0.6142 0.6269 0.8652 0.8797
2  0.2596 0.2920 0.2970 0.3011 0.6615 0.6619 0.6899 0.8197
3  0.5624 0.5632 0.5931 0.5994 0.5996 0.6253 0.7160 0.7505
4  0.0088 0.0378 0.1669 0.6757 0.7152 0.7851 0.7882 0.7954
5  0.2784 0.3766 0.4325 0.6029 0.6148 0.6303 0.7525 0.8795
6  0.3482 0.6916 0.7167 0.7655 0.7955 0.8283 0.8565 0.8703
7  0.2459 0.2662 0.2775 0.2851 0.5877 0.5989 0.5999 0.8416
8  0.1259 0.5292 0.5618 0.5715 0.8413 0.8809 0.8819 0.8939
9  0.3514 0.5244 0.6653 0.7051 0.7083 0.7115 0.7600 0.7891
10 0.4677 0.5348 0.5357 0.6309 0.6729 0.6747 0.6901 0.6923
11 0.4968 0.5034 0.6838 0.6997 0.7615 0.7633 0.7713 0.8294
12 0.4930 0.4941 0.6312 0.6313 0.6314 0.8037 0.8068 0.8244
13 0.3890 0.4408 0.7532 0.7562 0.7610 0.8138 0.8262 0.8354
14 0.1032 0.1936 0.7337 0.7351 0.7372 0.7602 0.7635 0.7897
15 0.3183 0.5770 0.5894 0.6690 0.7001 0.7442 0.8960 0.8960
16 0.6270 0.7430 0.7431 0.7454 0.7530 0.8506 0.8562 0.8602
17 0.5012 0.6357 0.6357 0.6486 0.6740 0.8615 0.8637 0.9034
18 0.5000 0.5004 0.5587 0.5616 0.5619 0.5625 0.5625 0.6044
19 0.3431 0.5320 0.5334 0.5606 0.5663 0.5663 0.7626 0.7631
20 0.3767 0.3937 0.5733 0.6512 0.6556 0.6851 0.6988 0.8058
21 0.1073 0.1317 0.2531 0.5152 0.5227 0.7582 0.7705 0.7732


Factor Matrix and Defining Sorts
$`Q-sort factor loadings`
           f1      f2      f3      f4
P01     0.1105  0.2061  0.5207  0.3561
P02     0.0319  0.2569  0.4348 -0.1269
P03     0.4773  0.1737  0.3342  0.0447
P04     0.0509 -0.0733 -0.0085  0.8642
P05     0.2288 -0.1636  0.6794 -0.2501
P06    -0.0642  0.0415  0.9084 -0.1773
P07     0.1289  0.2243  0.3748 -0.1916
P08    -0.3138  0.3600  0.6121  0.0207
P09     0.2017 -0.1009  0.6449  0.3500
P10     0.4942  0.3642 -0.0116  0.2001
P11     0.1075  0.7157  0.2220 -0.1483
P12     0.1029  0.6620  0.2413 -0.0572
P13     0.1894  0.7844 -0.0593 -0.1117
P14     0.8187 -0.5753  0.0322  0.2571
P15     0.8131  0.0602 -0.2264 -0.2076
P16     0.7463  0.2776 -0.0562 -0.0666
P17     0.6950  0.3214 -0.1711  0.0577
P18     0.5399 -0.0082  0.3652  0.1023
P19     0.6273  0.2329 -0.1263  0.0277
P20    -0.0505  0.7453  0.1578  0.2058
P21     0.5359 -0.2956  0.2081 -0.2934

$`Flagged Q-sorts`
       flag_f1 flag_f2 flag_f3 flag_f4
ï..P01   FALSE   FALSE    TRUE   FALSE
P02      FALSE   FALSE    TRUE   FALSE
P03       TRUE   FALSE   FALSE   FALSE
P04      FALSE   FALSE   FALSE    TRUE
P05      FALSE   FALSE    TRUE   FALSE
```

```
P06      FALSE   FALSE    TRUE   FALSE
P07      FALSE   FALSE    TRUE   FALSE
P08      FALSE   FALSE    TRUE   FALSE
P09      FALSE   FALSE    TRUE   FALSE
P10       TRUE   FALSE   FALSE   FALSE
P11      FALSE    TRUE   FALSE   FALSE
P12      FALSE    TRUE   FALSE   FALSE
P13      FALSE    TRUE   FALSE   FALSE
P14       TRUE   FALSE   FALSE   FALSE
P15       TRUE   FALSE   FALSE   FALSE
P16       TRUE   FALSE   FALSE   FALSE
P17       TRUE   FALSE   FALSE   FALSE
P18       TRUE   FALSE   FALSE   FALSE
P19       TRUE   FALSE   FALSE   FALSE
P20      FALSE    TRUE   FALSE   FALSE
P21       TRUE   FALSE   FALSE   FALSE

Free Distribution Data Results -- not calculated

Factor Scores (z-scores)
   zsc_f1 zsc_f2 zsc_f3 zsc_f4
1   -2.26  -1.66   0.63   0.00
2    1.90  -0.80   1.98   0.00
3    2.13   0.85   2.01   0.00
4    1.43   1.52   1.38  -0.95
5    0.78   1.17   0.11  -0.95
6    1.09   1.21  -0.50  -0.95
7    0.42   0.99   0.15   0.47
8   -0.25   1.43  -0.21  -0.95
9   -1.15  -1.83  -2.11  -1.89
10  -0.25  -0.11   0.32   1.89
11  -0.12   1.33   0.93   0.47
12  -0.11   0.74  -1.76   0.95
13   1.38   1.79   1.55  -1.42
14   0.04   1.52   0.75  -1.42
15   1.51   0.99   0.63   1.89
16  -1.13  -1.14  -0.17   0.00
17  -0.31   0.11   0.20   0.00
18  -1.17  -0.72   0.04   0.47
19  -0.26  -0.25  -0.15   0.47
20   0.78   1.07  -0.25   0.47
21  -1.08   0.89  -0.32   0.47
22  -0.10   0.65   0.57   0.95
23  -1.24  -0.01   0.01  -0.95
24  -1.07  -0.22  -0.31  -0.47
25  -0.54  -0.15  -0.69   0.95
26  -0.75   0.25   0.97  -0.47
27  -1.60  -1.44   0.41  -0.47
28  -0.58  -0.94  -1.09  -0.47
29  -1.20  -0.60  -1.35  -0.47
30   0.08  -0.13   0.87  -0.47
31   0.28  -0.09  -0.67   0.95
32   0.73  -0.71  -0.29   1.42
33  -0.05  -0.34  -0.18   1.42
34   0.79  -0.43  -0.78   0.00
35   0.93  -1.12   0.15  -1.42
36   0.52  -0.99  -1.34   1.42
37   0.42  -1.04  -1.20   0.00
38   0.00  -1.09  -1.74  -1.89
39   0.02  -0.69   1.44   0.95

Correlations Between Factor Scores
       zsc_f1  zsc_f2 zsc_f3  zsc_f4
zsc_f1 1.0000  0.4537 0.3586  0.0661
zsc_f2 0.4537  1.0000 0.3796 -0.0098
zsc_f3 0.3586  0.3796 1.0000  0.0432
zsc_f4 0.0661 -0.0098 0.0432  1.0000

Factor Scores
$`-- For Factor 1`
   zsc_f1
3   2.130
2   1.901
15  1.507
4   1.433
13  1.384
6   1.093
35  0.933
34  0.788
20  0.783
5   0.778
32  0.725
36  0.517
37  0.418
7   0.416
```

```
31  0.284
30  0.081
14  0.037
39  0.019
38  0.001
33 -0.048
22 -0.098
12 -0.113
11 -0.117
10 -0.248
8  -0.248
19 -0.264
17 -0.313
25 -0.542
28 -0.581
26 -0.749
24 -1.067
21 -1.082
16 -1.126
9  -1.151
18 -1.174
29 -1.201
23 -1.245
27 -1.600
1  -2.264

$`-- For Factor 2`
   zsc_f2
13  1.790
14  1.524
4   1.519
8   1.427
11  1.332
6   1.209
5   1.170
20  1.073
15  0.994
7   0.988
21  0.890
3   0.846
12  0.736
22  0.655
26  0.252
17  0.107
23 -0.012
31 -0.086
10 -0.111
30 -0.130
25 -0.148
24 -0.224
19 -0.255
33 -0.342
34 -0.426
29 -0.597
39 -0.692
32 -0.710
18 -0.724
2  -0.799
28 -0.937
36 -0.995
37 -1.038
38 -1.092
35 -1.124
16 -1.137
27 -1.445
1  -1.655
9  -1.834

$`-- For Factor 3`
   zsc_f3
3   2.006
2   1.976
13  1.553
39  1.445
4   1.385
26  0.975
11  0.928
30  0.867
14  0.749
15  0.635
1   0.631
22  0.565
27  0.408
10  0.323
17  0.197
```

```
7   0.152
35  0.151
5   0.107
18  0.037
23  0.010
19 -0.154
16 -0.166
33 -0.184
8  -0.208
20 -0.249
32 -0.292
24 -0.306
21 -0.321
6  -0.498
31 -0.667
25 -0.694
34 -0.777
28 -1.094
37 -1.196
36 -1.343
29 -1.345
38 -1.736
12 -1.759
9  -2.109
```

```
$`-- For Factor 4`
   zsc_f4
10  1.891
15  1.891
32  1.418
33  1.418
36  1.418
12  0.946
22  0.946
25  0.946
31  0.946
39  0.946
7   0.473
11  0.473
18  0.473
19  0.473
20  0.473
21  0.473
1   0.000
2   0.000
3   0.000
16  0.000
17  0.000
34  0.000
37  0.000
24 -0.473
26 -0.473
27 -0.473
28 -0.473
29 -0.473
30 -0.473
4  -0.946
5  -0.946
6  -0.946
8  -0.946
23 -0.946
13 -1.418
14 -1.418
35 -1.418
9  -1.891
38 -1.891
```

```
Descending Array of Differences Between Factors
$`1 and 2`
     zsc_f1 zsc_f2   f1_f2 sig_f1_f2                      dist.and.cons
2   1.90055 -0.799  2.6996        6*
35  0.93345 -1.124  2.0573        6* Distinguishes f1 Distinguishes f3
36  0.51661 -0.995  1.5114        6*
37  0.41810 -1.038  1.4563        6*
32  0.72516 -0.710  1.4352        6*
3   2.12985  0.846  1.2837       ***
34  0.78839 -0.426  1.2142       ***
38  0.00095 -1.092  1.0928       ***                  Distinguishes f1
39  0.01944 -0.692  0.7110         *                  Distinguishes f2
9  -1.15061 -1.834  0.6833         *
15  1.50726  0.994  0.5130
31  0.28409 -0.086  0.3698
28 -0.58061 -0.937  0.3561
```

```
33 -0.04770 -0.342  0.2941                       Distinguishes f4 only
30  0.08146 -0.130  0.2111                       Distinguishes f3 only
16 -1.12578 -1.137  0.0117
19 -0.26402 -0.255 -0.0092                                   Consensus
4   1.43320  1.519 -0.0861                       Distinguishes f4 only
6   1.09281  1.209 -0.1158
10 -0.24829 -0.111 -0.1371                           Distinguishes f4
27 -1.59993 -1.445 -0.1550
20  0.78279  1.073 -0.2906
5   0.77806  1.170 -0.3914          Distinguishes f3 Distinguishes f4
25 -0.54184 -0.148 -0.3938                       Distinguishes f4 only
13  1.38389  1.790 -0.4060                       Distinguishes f4 only
17 -0.31278  0.107 -0.4200
18 -1.17402 -0.724 -0.4504
7   0.41565  0.988 -0.5721
29 -1.20093 -0.597 -0.6038         *
1  -2.26394 -1.655 -0.6085         * Distinguishes f1 Distinguishes f2
22 -0.09780  0.655 -0.7525         *             Distinguishes f1 only
24 -1.06660 -0.224 -0.8425        **
12 -0.11350  0.736 -0.8490        ** Distinguishes f1 Distinguishes f3
26 -0.74872  0.252 -1.0009       ***                 Distinguishes f3
23 -1.24492 -0.012 -1.2330       ***
11 -0.11662  1.332 -1.4487        6*
14  0.03736  1.524 -1.4868        6*                Distinguishes all
8  -0.24839  1.427 -1.6749        6*             Distinguishes f2 only
21 -1.08207  0.890 -1.9722        6*                 Distinguishes f1

$`1 and 3`
     zsc_f1  zsc_f3  f1_f3 sig_f1_f3                     dist.and.cons
36  0.51661 -1.3427  1.859        6*
38  0.00095 -1.7356  1.737        6*                 Distinguishes f1
12 -0.11350 -1.7594  1.646        6* Distinguishes f1 Distinguishes f3
37  0.41810 -1.1957  1.614        6*
6   1.09281 -0.4985  1.591        6*
34  0.78839 -0.7766  1.565        6*
20  0.78279 -0.2490  1.032       ***
32  0.72516 -0.2921  1.017       ***
9  -1.15061 -2.1086  0.958       ***
31  0.28409 -0.6671  0.951       ***
15  1.50726  0.6349  0.872       ***
35  0.93345  0.1506  0.783        ** Distinguishes f1 Distinguishes f3
5   0.77806  0.1070  0.671        ** Distinguishes f3 Distinguishes f4
28 -0.58061 -1.0940  0.513         *
7   0.41565  0.1523  0.263
25 -0.54184 -0.6942  0.152                       Distinguishes f4 only
29 -1.20093 -1.3451  0.144
33 -0.04770 -0.1843  0.137                       Distinguishes f4 only
3   2.12985  2.0061  0.124
4   1.43320  1.3849  0.048                       Distinguishes f4 only
8  -0.24839 -0.2083 -0.040                       Distinguishes f2 only
2   1.90055  1.9755 -0.075
19 -0.26402 -0.1544 -0.110                                   Consensus
13  1.38389  1.5530 -0.169                       Distinguishes f4 only
17 -0.31278  0.1970 -0.510         *
10 -0.24829  0.3230 -0.571         *                 Distinguishes f4
22 -0.09780  0.5650 -0.663        **             Distinguishes f1 only
14  0.03736  0.7491 -0.712        **                Distinguishes all
24 -1.06660 -0.3059 -0.761        **
21 -1.08207 -0.3210 -0.761        **                 Distinguishes f1
30  0.08146  0.8674 -0.786        **             Distinguishes f3 only
16 -1.12578 -0.1657 -0.960       ***
11 -0.11662  0.9280 -1.045       ***
18 -1.17402  0.0368 -1.211       ***
23 -1.24492  0.0097 -1.255        6*
39  0.01944  1.4446 -1.425        6*                 Distinguishes f2
26 -0.74872  0.9748 -1.723        6*                 Distinguishes f3
27 -1.59993  0.4076 -2.008        6*
1  -2.26394  0.6306 -2.895        6* Distinguishes f1 Distinguishes f2

$`1 and 4`
     zsc_f1 zsc_f4  f1_f4 sig_f1_f4                     dist.and.cons
13  1.38389  -1.42  2.802        6*             Distinguishes f4 only
4   1.43320  -0.95  2.379        6*             Distinguishes f4 only
35  0.93345  -1.42  2.352        6* Distinguishes f1 Distinguishes f3
3   2.12985   0.00  2.130       ***
6   1.09281  -0.95  2.038       ***
2   1.90055   0.00  1.901       ***
38  0.00095  -1.89  1.892       ***                 Distinguishes f1
5   0.77806  -0.95  1.724       *** Distinguishes f3 Distinguishes f4
14  0.03736  -1.42  1.456        **                Distinguishes all
34  0.78839   0.00  0.788
9  -1.15061  -1.89  0.741
8  -0.24839  -0.95  0.697                       Distinguishes f2 only
30  0.08146  -0.47  0.554                       Distinguishes f3 only
37  0.41810   0.00  0.418
```

```
20  0.78279   0.47  0.310
7   0.41565   0.47 -0.057
28 -0.58061  -0.47 -0.108
26 -0.74872  -0.47 -0.276                            Distinguishes f3
23 -1.24492  -0.95 -0.299
17 -0.31278   0.00 -0.313
15  1.50726   1.89 -0.384
11 -0.11662   0.47 -0.589
24 -1.06660  -0.47 -0.594
31  0.28409   0.95 -0.661
32  0.72516   1.42 -0.693
29 -1.20093  -0.47 -0.728
19 -0.26402   0.47 -0.737                                    Consensus
36  0.51661   1.42 -0.902
39  0.01944   0.95 -0.926                            Distinguishes f2
22 -0.09780   0.95 -1.043         *              Distinguishes f1 only
12 -0.11350   0.95 -1.059         * Distinguishes f1 Distinguishes f3
16 -1.12578   0.00 -1.126         *
27 -1.59993  -0.47 -1.127         *
33 -0.04770   1.42 -1.466        **              Distinguishes f4 only
25 -0.54184   0.95 -1.487        **              Distinguishes f4 only
21 -1.08207   0.47 -1.555        **                  Distinguishes f1
18 -1.17402   0.47 -1.647       ***
10 -0.24829   1.89 -2.139       ***                  Distinguishes f4
1  -2.26394   0.00 -2.264       *** Distinguishes f1 Distinguishes f2

$`2 and 3`
   zsc_f2  zsc_f3  f2_f3 sig_f2_f3                        dist.and.cons
12  0.736 -1.7594  2.495        6* Distinguishes f1 Distinguishes f3
6   1.209 -0.4985  1.707        6*
8   1.427 -0.2083  1.635        6*             Distinguishes f2 only
20  1.073 -0.2490  1.322       ***
21  0.890 -0.3210  1.211       ***                 Distinguishes f1
5   1.170  0.1070  1.063       *** Distinguishes f3 Distinguishes f4
7   0.988  0.1523  0.835        **
14  1.524  0.7491  0.775         *                Distinguishes all
29 -0.597 -1.3451  0.748         *
38 -1.092 -1.7356  0.644         *                 Distinguishes f1
31 -0.086 -0.6671  0.581
25 -0.148 -0.6942  0.546                       Distinguishes f4 only
11  1.332  0.9280  0.404
15  0.994  0.6349  0.359
34 -0.426 -0.7766  0.351
36 -0.995 -1.3427  0.348
9  -1.834 -2.1086  0.275
13  1.790  1.5530  0.237                       Distinguishes f4 only
37 -1.038 -1.1957  0.157
28 -0.937 -1.0940  0.157
4   1.519  1.3849  0.134                       Distinguishes f4 only
22  0.655  0.5650  0.090                       Distinguishes f1 only
24 -0.224 -0.3059  0.082
23 -0.012  0.0097 -0.022
17  0.107  0.1970 -0.090
19 -0.255 -0.1544 -0.100                                  Consensus
33 -0.342 -0.1843 -0.158                       Distinguishes f4 only
32 -0.710 -0.2921 -0.418
10 -0.111  0.3230 -0.434                           Distinguishes f4
26  0.252  0.9748 -0.723         *                 Distinguishes f3
18 -0.724  0.0368 -0.760         *
16 -1.137 -0.1657 -0.972        **
30 -0.130  0.8674 -0.997        **             Distinguishes f3 only
3   0.846  2.0061 -1.160       ***
35 -1.124  0.1506 -1.274       *** Distinguishes f1 Distinguishes f3
27 -1.445  0.4076 -1.853        6*
39 -0.692  1.4446 -2.136        6*                 Distinguishes f2
1  -1.655  0.6306 -2.286        6* Distinguishes f1 Distinguishes f2
2  -0.799  1.9755 -2.775        6*

$`2 and 4`
   zsc_f2 zsc_f4  f2_f4 sig_f2_f4                     dist.and.cons
13  1.790  -1.42  3.208        6*          Distinguishes f4 only
14  1.524  -1.42  2.942        6*              Distinguishes all
4   1.519  -0.95  2.465       ***          Distinguishes f4 only
8   1.427  -0.95  2.372       ***          Distinguishes f2 only
6   1.209  -0.95  2.154       ***
5   1.170  -0.95  2.115       *** Distinguishes f3 Distinguishes f4
23 -0.012  -0.95  0.934
11  1.332   0.47  0.859
3   0.846   0.00  0.846
38 -1.092  -1.89  0.799                        Distinguishes f1
26  0.252  -0.47  0.725                        Distinguishes f3
20  1.073   0.47  0.601
7   0.988   0.47  0.515
21  0.890   0.47  0.417                        Distinguishes f1
30 -0.130  -0.47  0.343                    Distinguishes f3 only
```

```
35 -1.124  -1.42  0.295           Distinguishes f1 Distinguishes f3
24 -0.224  -0.47  0.249
17  0.107   0.00  0.107
9  -1.834  -1.89  0.057
29 -0.597  -0.47 -0.124
12  0.736   0.95 -0.210           Distinguishes f1 Distinguishes f3
22  0.655   0.95 -0.291                      Distinguishes f1 only
34 -0.426   0.00 -0.426
28 -0.937  -0.47 -0.464
19 -0.255   0.47 -0.728                                  Consensus
2  -0.799   0.00 -0.799
15  0.994   1.89 -0.897
27 -1.445  -0.47 -0.972
31 -0.086   0.95 -1.031         *
37 -1.038   0.00 -1.038         *
25 -0.148   0.95 -1.094         *            Distinguishes f4 only
16 -1.137   0.00 -1.137         *
18 -0.724   0.47 -1.196         *
39 -0.692   0.95 -1.637        **                Distinguishes f2
1  -1.655   0.00 -1.655        ** Distinguishes f1 Distinguishes f2
33 -0.342   1.42 -1.760       ***            Distinguishes f4 only
10 -0.111   1.89 -2.002       ***                Distinguishes f4
32 -0.710   1.42 -2.128       ***
36 -0.995   1.42 -2.413       ***

$`3 and 4`
    zsc_f3 zsc_f4 f3_f4 sig_f3_f4                   dist.and.cons
13  1.5530  -1.42  2.97        6*            Distinguishes f4 only
4   1.3849  -0.95  2.33       ***            Distinguishes f4 only
14  0.7491  -1.42  2.17       ***                Distinguishes all
3   2.0061   0.00  2.01       ***
2   1.9755   0.00  1.98       ***
35  0.1506  -1.42  1.57        ** Distinguishes f1 Distinguishes f3
26  0.9748  -0.47  1.45        **                Distinguishes f3
30  0.8674  -0.47  1.34        **            Distinguishes f3 only
5   0.1070  -0.95  1.05         * Distinguishes f3 Distinguishes f4
23  0.0097  -0.95  0.96         *
27  0.4076  -0.47  0.88
8  -0.2083  -0.95  0.74                      Distinguishes f2 only
1   0.6306   0.00  0.63           Distinguishes f1 Distinguishes f2
39  1.4446   0.95  0.50                          Distinguishes f2
11  0.9280   0.47  0.46
6  -0.4985  -0.95  0.45
17  0.1970   0.00  0.20
24 -0.3059  -0.47  0.17
38 -1.7356  -1.89  0.16                          Distinguishes f1
16 -0.1657   0.00 -0.17
9  -2.1086  -1.89 -0.22
7   0.1523   0.47 -0.32
22  0.5650   0.95 -0.38                      Distinguishes f1 only
18  0.0368   0.47 -0.44
28 -1.0940  -0.47 -0.62
19 -0.1544   0.47 -0.63                                  Consensus
20 -0.2490   0.47 -0.72
34 -0.7766   0.00 -0.78
21 -0.3210   0.47 -0.79                          Distinguishes f1
29 -1.3451  -0.47 -0.87
37 -1.1957   0.00 -1.20         *
15  0.6349   1.89 -1.26        **
10  0.3230   1.89 -1.57        **                Distinguishes f4
33 -0.1843   1.42 -1.60       ***            Distinguishes f4 only
31 -0.6671   0.95 -1.61       ***
25 -0.6942   0.95 -1.64       ***            Distinguishes f4 only
32 -0.2921   1.42 -1.71       ***
12 -1.7594   0.95 -2.70        6* Distinguishes f1 Distinguishes f3
36 -1.3427   1.42 -2.76        6*

Factor Q-Sort Values for Each Statement
   fsc_f1 fsc_f2 fsc_f3 fsc_f4
1      -4     -4      1      0
2       4     -2      4      0
3       4      1      4      0
4       3      3      3     -2
5       2      2      0     -2
6       2      2     -1     -2
7       1      2      1      1
8      -1      3     -1     -2
9      -2     -4     -4     -4
10     -1      0      1      4
11      0      3      2      1
12      0      1     -4      2
13      3      4      3     -3
14      0      4      2     -3
15      3      2      2      4
16     -2     -3      0      0
```

17	-1	1	1	0
18	-3	-1	0	1
19	-1	0	0	1
20	2	2	-1	1
21	-2	1	-1	1
22	0	1	1	2
23	-3	0	0	-2
24	-2	0	-1	-1
25	-1	0	-2	2
26	-2	1	2	-1
27	-4	-3	1	-1
28	-1	-2	-2	-1
29	-3	-1	-3	-1
30	1	0	2	-1
31	1	0	-2	2
32	1	-1	-1	3
33	0	-1	0	3
34	2	-1	-2	0
35	2	-3	0	-3
36	1	-2	-3	3
37	1	-2	-2	0
38	0	-2	-3	-4
39	0	-1	3	2

Factor Q-Sort Values for Statements sorted by Consensus vs. Disagreement (Variance across Factor Z-Scores)

	fsc_f1	fsc_f2	fsc_f3	fsc_f4
17	-1	1	1	0
28	-1	-2	-2	-1
7	1	2	1	1
19	-1	0	0	1
24	-2	0	-1	-1
9	-2	-4	-4	-4
29	-3	-1	-3	-1
22	0	1	1	2
15	3	2	2	4
20	2	2	-1	1
30	1	0	2	-1
16	-2	-3	0	0
11	0	3	2	1
23	-3	0	0	-2
34	2	-1	-2	0
31	1	0	-2	2
25	-1	0	-2	2
18	-3	-1	0	1
26	-2	1	2	-1
37	1	-2	-2	0
33	0	-1	0	3
38	0	-2	-3	-4
21	-2	1	-1	1
5	2	2	0	-2
27	-4	-3	1	-1
39	0	-1	3	2
32	1	-1	-1	3
10	-1	0	1	4
8	-1	3	-1	-2
3	4	1	4	0
6	2	2	-1	-2
35	2	-3	0	-3
4	3	3	3	-2
12	0	1	-4	2
14	0	4	2	-3
36	1	-2	-3	3
1	-4	-4	1	0
2	4	-2	4	0
13	3	4	3	-3

Factor Characteristics

	f1	f2	f3	f4
Average reliability coefficient	0.80	0.80	0.80	0.80
Number of loading Q-sorts	9.00	4.00	7.00	1.00
Eigenvalues	4.10	3.27	3.20	1.48
Percentage of explained variance	19.54	15.59	15.25	7.05
Composite reliability	0.97	0.94	0.97	0.80
Standard error of factor scores	0.16	0.24	0.19	0.45

Standard Errors for Differences in Factor Z-Scores

	f1	f2	f3	f4
f1	0.2324953	0.2930027	0.2480117	0.4764735
f2	0.2930027	0.3429972	0.3054608	0.5087470
f3	0.2480117	0.3054608	0.2626129	0.4842342
f4	0.4764735	0.5087470	0.4842342	0.6324555

```
Distinguishing Statements
$`for Factor 1`
   zsc_f1 zsc_f2 zsc_f3 zsc_f4                      dist.and.cons sig_f1_f2
1   -2.26  -1.66   0.63   0.00 Distinguishes f1 Distinguishes f2         *
12  -0.11   0.74  -1.76   0.95 Distinguishes f1 Distinguishes f3        **
14   0.04   1.52   0.75  -1.42                  Distinguishes all       6*
21  -1.08   0.89  -0.32   0.47                  Distinguishes f1        6*
22  -0.10   0.65   0.57   0.95              Distinguishes f1 only        *
35   0.93  -1.12   0.15  -1.42 Distinguishes f1 Distinguishes f3        6*
38   0.00  -1.09  -1.74  -1.89                  Distinguishes f1       ***
   sig_f1_f3 sig_f1_f4 sig_f2_f3 sig_f2_f4 sig_f3_f4
1         6*       ***        6*        **
12        6*         *        6*                  6*
14        **        **         *        6*       ***
21        **        **       ***
22        **         *
35        **        6*       ***                  **
38        6*       ***         *

$`for Factor 2`
   zsc_f1 zsc_f2 zsc_f3 zsc_f4                      dist.and.cons sig_f1_f2
1   -2.26  -1.66   0.63   0.00 Distinguishes f1 Distinguishes f2         *
8   -0.25   1.43  -0.21  -0.95              Distinguishes f2 only       6*
14   0.04   1.52   0.75  -1.42                  Distinguishes all       6*
39   0.02  -0.69   1.44   0.95                  Distinguishes f2         *
   sig_f1_f3 sig_f1_f4 sig_f2_f3 sig_f2_f4 sig_f3_f4
1         6*       ***        6*        **
8                             6*       ***
14        **        **         *        6*       ***
39        6*                  6*        **

$`for Factor 3`
   zsc_f1 zsc_f2 zsc_f3 zsc_f4                      dist.and.cons sig_f1_f2
5    0.78   1.17   0.11  -0.95 Distinguishes f3 Distinguishes f4
12  -0.11   0.74  -1.76   0.95 Distinguishes f1 Distinguishes f3        **
14   0.04   1.52   0.75  -1.42                  Distinguishes all       6*
26  -0.75   0.25   0.97  -0.47                  Distinguishes f3       ***
30   0.08  -0.13   0.87  -0.47              Distinguishes f3 only
35   0.93  -1.12   0.15  -1.42 Distinguishes f1 Distinguishes f3        6*
   sig_f1_f3 sig_f1_f4 sig_f2_f3 sig_f2_f4 sig_f3_f4
5         **       ***       ***       ***         *
12        6*         *        6*                  6*
14        **        **         *        6*       ***
26        6*                   *                  **
30        **                  **                  **
35        **        6*       ***                  **

$`for Factor 4`
   zsc_f1 zsc_f2 zsc_f3 zsc_f4                      dist.and.cons sig_f1_f2
4    1.43   1.52   1.38  -0.95              Distinguishes f4 only
5    0.78   1.17   0.11  -0.95 Distinguishes f3 Distinguishes f4
10  -0.25  -0.11   0.32   1.89                  Distinguishes f4
13   1.38   1.79   1.55  -1.42              Distinguishes f4 only
14   0.04   1.52   0.75  -1.42                  Distinguishes all       6*
25  -0.54  -0.15  -0.69   0.95              Distinguishes f4 only
33  -0.05  -0.34  -0.18   1.42              Distinguishes f4 only
   sig_f1_f3 sig_f1_f4 sig_f2_f3 sig_f2_f4 sig_f3_f4
4                   6*                 ***       ***
5         **       ***       ***       ***         *
10         *       ***                 ***        **
13                  6*                  6*        6*
14        **        **         *        6*       ***
25                  **                   *       ***
33                  **                 ***       ***

results$qdc[which(results$qdc$dist.and.cons == "Consensus"), ]
  dist.and.cons       f1_f2 sig_f1_f2     f1_f3 sig_f1_f3     f1_f4 sig_f1_f4
19     Consensus -0.00923515           -0.109597           -0.736811
       f2_f3 sig_f2_f3      f2_f4 sig_f2_f4     f3_f4 sig_f3_f4
19 -0.1003618           -0.7275758           -0.627214
results$qdc[which(results$qdc$dist.and.cons == "Distinguishes all"), ]
       dist.and.cons     f1_f2 sig_f1_f2      f1_f3 sig_f1_f3    f1_f4
14 Distinguishes all -1.486759        6* -0.7117396        ** 1.455727
   sig_f1_f4    f2_f3 sig_f2_f3    f2_f4 sig_f2_f4    f3_f4 sig_f3_f4
14        ** 0.775019         * 2.942485        6* 2.167466       ***
```