



KTH Royal Institute of Technology and Faculty of Engineering LTH,  
Lund University

# Turning Smart Water Meter Data Into Useful Information

A case study on rental apartments in Södertälje

Philip Dahlström

Anna Söderberg

September 2017

## *Abstract*

Managing water in urban areas is an ever increasingly complex challenge. Technology enables sustainable urban water management and with integrated smart metering solutions, massive amounts of water consumption data from the end users can be collected. However, the possibility of generating data from the end user holds no value in itself. It is with the use of data analysis the vast amount of the collected data can provide more insightful information creating potential benefits. It is recognized that a deeper understanding of the end user could potentially provide benefits for operational managers as well as for the end users. A single case study of a data set containing high frequency end user water consumption data from rental apartments has been conducted, where the data set was analyzed in order to see what possible information that could be extracted and interpreted based on an exploratory data analysis (EDA). Furthermore, an interview with the operational manager of the buildings under study as well as a literature review have been carried out in order to understand how the gathered data is used today and to which contexts it could be extrapolated to provide potential benefits at a building level.

The results suggests that the EDA is a powerful method approach when starting out without strong preconception of the data under study and have successfully revealed patterns and a fundamental understanding of the data and its structure. Through analysis, variations over time, water consumption patterns and excessive water users have been identified as well as a leak identification process. Even more challenging than to make meaning of the data is to trigger actions, decisions and measures based on the data analysis. The unveiled information could be applied for an improved operational building management, to empower the customers, for business and campaign opportunities as well as for an integrated decision support system. To summarize, it is concluded that the usage of smart water metering data holds an untapped opportunity to save water, energy as well as money. In the drive towards a more sustainable and smarter city, smart water meter data from end users have the potential to enable smarter building management as well as smarter water services.

***Keywords:*** *End User Water Consumption, Smart Water Meter Data, Exploratory Data Analysis, Water Management, Building Operational Management*

# *Acknowledgements*

This master thesis is written for the fulfillment of a Master's degree in Master of Science and Engineering in a collaboration between KTH Royal Institute of Technology and Faculty of Engineering LTH, Lund University. The thesis have been written during five months in Stockholm at the headquarter of the consultant company Tyréns AB.

I, Anna, would like to start giving my warmest and deepest thanks to Luigia Brandimarte, assistant professor at the Department of Hydraulic Engineering and my supervisor at KTH Royal Institute of Technology. Your valuable suggestions and support have been of great impact for the outcome of this thesis. Furthermore, I would like to thank the program director of my Master's program in Environmental Engineering and Sustainable Infrastructure as well as the Department of Hydraulic Engineering for supporting my ideas of an interdisciplinary approach and the choice of a rather unconventional topic out of personal interest.

I, Philip, would like to extend my thanks and gratitude to my supervisors at Faculty of Engineering LTH - Lars Bengtsson, professor at Innovation Engineering, and Emil Åkesson, doctoral student at Innovation Engineering. Your support and valuable feedback have been imperative in the making of this study. I would also like express my sincere thanks to the Department of Design Science for making this collaboration possible.

We would like to express our gratitude towards Per Bjälnes - BIM and IoT strategist and likewise our supervisor at Tyréns AB, for sharing your knowledge, enthusiasm and for giving us the opportunity of involvement in the project.

A special thanks goes to Daniel Bäcklin at Telge Bostäder for your participation in the case study and your time and efforts to answer all our questions. Furthermore, we would like to thank Johan Genzell as well as Robert Westerberg for the study visit in Södertälje as well as providing access to the database used in the case study.

Finally, our warmest thanks to Tyréns AB and the people in the BIM department for welcoming us and making this study possible.

Stockholm, August 2017

*Philip Dahlström and Anna Söderberg*

# Contents

<b>Abstract</b>	<b>i</b>
<b>Acknowledgements</b>	<b>ii</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Tables</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Purpose . . . . .	4
1.3 Research Questions . . . . .	5
1.4 Delimitation . . . . .	5
1.5 Definitions . . . . .	7
1.6 Interdisciplinarity . . . . .	8
1.7 Disposition of the Study . . . . .	8
1.8 Work Distribution . . . . .	9
<b>2 Fundamental Concepts and Enabling Technologies</b>	<b>12</b>
2.1 Smart Cities . . . . .	12
2.2 Data Gathering of High Frequency Data . . . . .	13
2.2.1 The Internet of Things . . . . .	13
2.2.2 Smart Sensors and Water Meters . . . . .	14
2.3 Exploratory Data Analysis . . . . .	16
2.4 Data Driven Decision-Making . . . . .	17
<b>3 Method</b>	<b>19</b>
3.1 Research Design . . . . .	19
3.2 Research Methods . . . . .	21
3.2.1 Literature Review . . . . .	21
3.2.2 The Case Study . . . . .	21
3.3 Detailed Description of the Case Study . . . . .	22
3.3.1 The Case Study Area . . . . .	22
3.3.2 Description of the Data Set . . . . .	23
3.4 Methods and Techniques for Data Analysis . . . . .	26
3.4.1 Data Pre-processing . . . . .	26
3.4.2 Data Analysis Method . . . . .	27

---

3.4.2.1	Resistance Procedures . . . . .	28
3.4.2.2	Residual Analysis . . . . .	28
3.4.2.3	Revelations . . . . .	28
3.5	Research Ethics . . . . .	29
<b>4</b>	<b>Literature Review and Interview Results</b>	<b>30</b>
4.1	Previous Studies . . . . .	30
4.2	Interview with Telge Bostäder . . . . .	35
4.2.1	Data Gathering . . . . .	36
4.2.2	Data Usage . . . . .	37
4.2.3	Assessing Potential in Data Driven Operational Management . . . . .	38
<b>5</b>	<b>Analysis</b>	<b>39</b>
5.1	Average Consumption and Baselines . . . . .	39
5.2	Seasonality . . . . .	41
5.3	Regression Analysis . . . . .	44
5.4	Clustering . . . . .	45
5.5	Leak Detection . . . . .	47
<b>6</b>	<b>Discussion</b>	<b>48</b>
6.1	Discussion of the Analysis . . . . .	48
6.2	Discussion of Potential Benefits and Applications . . . . .	52
6.3	Discussion of Limitations . . . . .	55
<b>7</b>	<b>Conclusion</b>	<b>57</b>
7.1	Conclusions . . . . .	57
7.2	Contributions . . . . .	58
7.3	Future Research . . . . .	59
	<b>References</b> . . . . .	<b>60</b>
<b>A</b>	<b>Appendix A</b>	<b>67</b>
A.1	Obtaining the Data Set Under Study . . . . .	67
A.2	Obtaining the Baseline Values . . . . .	67
A.3	Residual Analysis . . . . .	68
A.4	Seasonality . . . . .	68
A.5	Leak Detection . . . . .	69
<b>B</b>	<b>Appendix B</b>	<b>71</b>
B.1	The Interview Guide . . . . .	71

# List of Figures

1.1	The number of published papers on Science Direct when the search terms "IoT" and "Water consumption" were used [1]. . . . .	2
1.2	Groundwater levels in Sweden during April 2017 [2]. . . . .	3
1.3	Scope of the study. . . . .	5
2.1	Elements of a smart city [3]. . . . .	13
2.2	Process flowchart of the IoT ecosystem [4]. . . . .	14
2.3	Methods for recording water volumes measured [5]. . . . .	15
2.4	The DIKW pyramid presented by Ackoff. . . . .	17
3.1	Research design and methods. . . . .	19
3.2	The reference method practice for conducting a case study [6]. . . . .	20
3.3	Södertälje city [7]. . . . .	23
3.4	A 2-room apartment of 63 $m^2$ with sensor locations marked in red. . . . .	25
3.5	Installed water meters at the incoming pipe to the kitchen. . . . .	25
5.1	A. Histogram of average daily consumption per apartment. B. Density curve with summary statistics over average daily consumption. . . . .	39
5.2	The ratio between total consumed hot and cold water per apartment. . . . .	40
5.3	A. Bar chart visualizing monthly seasonality on average consumption. B. Box plot of monthly seasonality on average consumption. . . . .	41
5.4	A. Bar chart visualizing monthly seasonality segmented into average hot and cold water consumption. B. Box plot of monthly seasonality segmented into average hot and cold water consumption. . . . .	41
5.5	A. Bar chart visualizing the daily variations over the weekdays of average consumption. B. Box plot of daily variations distributed over the weekdays of average consumption. . . . .	42
5.6	Hourly variations of average consumption visualized as a bar chart. . . . .	43
5.7	Hourly variations of average consumption for every apartments visualized as a box plot. . . . .	43
5.8	Graph visualizing correlation between apartment size and total water consumption with a linear regression line. . . . .	44
5.9	A. Clustering of minimum/maximum month for each apartment. B. Histogram visualizing the distribution of each apartments peak hour during the day. . . . .	45
5.10	Total consumption per apartment segmented per meter. . . . .	46
5.11	Violin plot of average daily consumption and number of rooms within each apartment. . . . .	46
5.12	Visualization of the leak detection procedure, obtained at 2016-10-03. . . . .	47

# List of Tables

1.1	Distribution of responsibility . . . . .	10
3.1	Properties of the installed water meters . . . . .	24
3.2	Type and number of errors found in the data set. . . . .	27
5.1	Summary Table with statistics for the whole population. . . . .	40
B.1	General interview information . . . . .	71
B.2	Introduction and important background information questions . . . . .	72
B.3	Questions regarding gathering and usage of data . . . . .	73
B.4	Questions regarding future prospects of data driven operational management	74

# Chapter 1

## Introduction

*This introductory chapter gives a brief background of the research topic and the identified knowledge gap. Purpose, research questions and delimitation of the study are presented as well as definitions of terms and concepts used throughout the thesis. Lastly, an overview of the disposition of the study can be found.*

### 1.1 Background

More people now live in cities than in rural areas around the world [8]. Due to the rapid urbanization, cities are growing and the density of cities are increasing, which creates new demands on services and infrastructure. At the same time, with the rising awareness of the importance of sustainability, there is an overarching goal to enable a transition towards a more sustainable city [3]. The digital revolution with its new technologies such as the Internet of Things (IoT) and how these technologies could be incorporated in services and infrastructure have emerged into the term of smart cities. Smart cities have enormous potential and it is recognized that smart cities could meet these new challenges posed by an increasing complexity [9]. Already a couple of years ago, smart cities were pointed out as a future emerging market which is expected to drive the digital economy forward [10].

IoT technology enables collection of massive amounts of high frequency data from smart sensors and could be used to monitor and measure usage and performance of different technical systems. The recent explosion of IoT enables new technical capabilities such as finer granular real time monitoring [4]. A relevant application of this technology would be in the water sector, where the technology could be used throughout the water supply infrastructure [11]. The interest of IoT in the water sector is growing rapidly, which



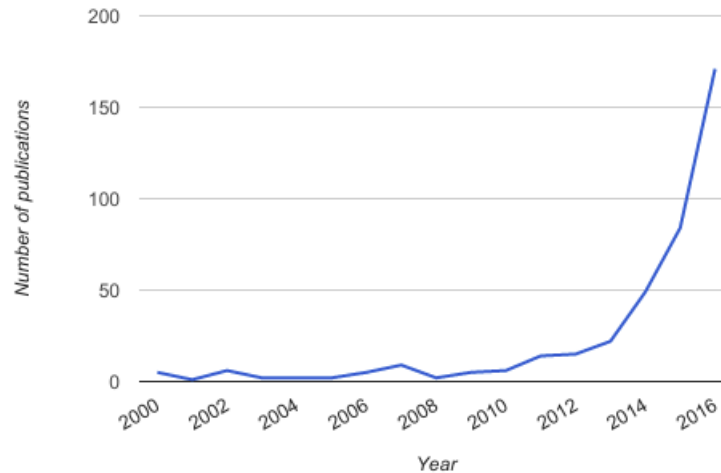


FIGURE 1.1: The number of published papers on Science Direct when the search terms "IoT" and "Water consumption" were used [1].

can be seen by the number of scientific papers that have been written and published on the topic during the last two years as seen in Figure 1.1. The theoretical objectives of introducing IoT technology in the water supply system is to enable for both water utilities, operational managers and consumers to proactively manage their water usage [11] and hence, achieve higher levels of sustainability. An implementation of smart sensors would feed the dedicated information platform with huge amounts of data. However, the technology development in the water sector are generally set behind compared to other sectors, as for example the energy sector [5], and full-scale implementations within the research field of urban water are still very rare [12]. Further on, incitements of investigating the possibilities of utilizing smart sensors have not been homogeneously distributed and have primarily been seen where the impacts of growing population demands, water scarcity and challenges posed by climate change have been more noticeable [5, 13–16].

In Sweden, improving existing technologies in order to conserve water or utilize water more efficiently have not been of priority, since water often is both oversupplied and under priced. However, groundwater levels have decreased in several parts of the country and major parts of Götaland, east of Svealand and the coastal areas of south Norrland have reached critical groundwater levels [17]. As Figure 1.2 shows, groundwater levels in April 2017 are below or even well below normal levels in major parts of Sweden. Impacts of posing water crises have recently been noticeable, as in Örebro during the winter of 2016/2017 where water restrictions and savings were necessary in order to be able to meet the demand. Most recently, a posing water crisis in the municipalities of Södertälje and Nykvarn have led to an irrigation prohibition from 2017-05-02 and onwards [18]. The prohibition is for everyone connected to the water distribution network and there is a recognized challenge to meet the demand during the summer. One approach to reduce

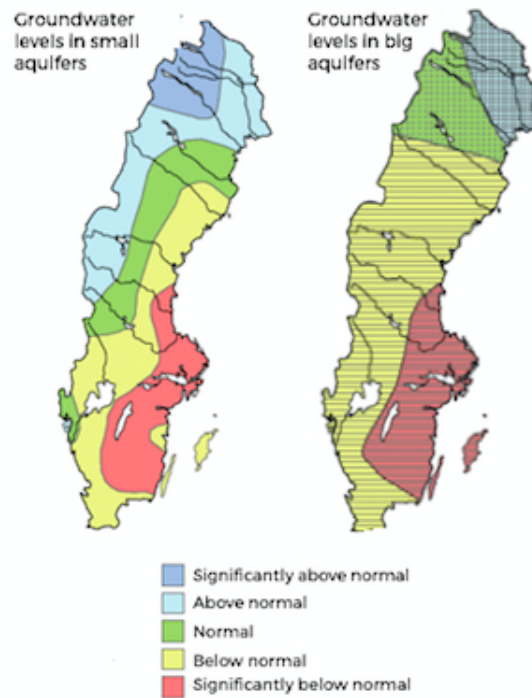


FIGURE 1.2: Groundwater levels in Sweden during April 2017 [2].

demand could be to use smart water sensors in order to reveal untapped water savings potential by the end user by measuring and monitoring at an apartment level.

Moreover, end user water consumption is linked to energy consumption by the heating and usage of hot tap water. The 2012 Energy Efficiency Directive of the European Union establish measures towards the goal of reducing the prime energy consumption by 20% at 2020 [19]. Hence, all EU countries are required to use energy more efficiently. This applies to all stages of the energy chain, i.e from production to final consumption, and further on for all sectors. Measures to reduce hot water consumption have been pointed out as a significant part in reducing total energy consumption for residential buildings. According to the 2012 Energy Efficiency Directive, the consumers should be empowered to better manage their own consumption. One step is through individual metering which provides access to data on individual consumption per household.

Telge Bostäder, the public utility real estate company of the municipality Södertälje, decided to install sensors for individual metering as a strategy to control and debit the hot water consumption at an apartment level a couple of years ago [20] with the goal to minimize energy consumption. Around 2000 of their rental apartments were equipped with sensors measuring end user consumption. After installing the sensors, no agreement on how to debit the tenants for their hot water consumption have been established and the project to install sensors in the remaining properties have been put on hold, since no other benefits of collecting the data of consumption at an apartment level can be seen.

In general, high frequency end user consumption data offers new possibilities. More accessible, more detailed, and more frequent data enables interpretation of consumption that was not available before and an increasing number of end-use studies highlights and emphasizes the importance of detailed knowledge about the end users [21]. At the same time, the water sector is confronted with new and complex challenges. In particular, data management, interpretation and analysis requirements of the data are pointed out as major challenges. To utilize the processed high frequency data to create information on issues such as end use consumption is still at a developmental stage [5] and requires further investigation.

With a growing amount of data available regarding water consumption, it is important to be able to transform this data into useful information [22] since the possibility of generating data from the end user holds no value in itself. Through data analysis, the data can be transformed into information which could foster an in-depth knowledge and insights to manage water more efficiently. End user consumption data could potentially provide both the consumers as well as the operational managers with tools to control and monitor the water consumption within the building. Knowledge of by whom, when and how water is being consumed poses a potential for improved operational building management. However, more investigation is needed to fully understand the role of smart metering data and its applications.

## 1.2 Purpose

The purpose of the thesis is twofold. First, the objective is to explore what possible information about the end user water consumption that can be extracted and interpreted from high frequency data by the use of an exploratory data analysis. Secondly, the usefulness of this information will be evaluated by investigating how the revealed information potentially could provide benefits at a building level. This is done through a single case study, where a data set from Södertälje is analyzed. The rental apartments under study are equipped with sensors collecting and storing data of accumulated water consumption at an hourly basis. By extracting information about the end user water consumption, a potential greater understanding of by whom, when and how water is being consumed can be provided. However, the value of the gathered data can be questioned if not applied in order to provide benefits, either for managers or consumers.

### 1.3 Research Questions

A clear research issue is a must for all good research [23]. In order to establish clear objective, research questions have been formulated which the thesis seeks answer to.

- *RQ1*: What possible information can be extracted and interpreted from a data set of end user water consumption based on an exploratory data analysis?
- *RQ2*: How can the revealed information potentially provide benefits at a building level?

### 1.4 Delimitation

The scope of this thesis is limited to the building level and the individual apartment level, illustrated by Figure 1.3. The individual tenants have not been included due to the aggregation level of the data set under study as well as integrity purpose. Therefore, the data set was already anonymised beforehand and the authors do not know which existing buildings in Södertälje the data set under analysis corresponds to. Hence, no other information about the tenants or the building standards and performance that were not included in the data set could be obtained. Furthermore, according to RQ2 an investigation if the revealed information could provide benefits and insights at the building level limit the scope to include the operational building manager and the end

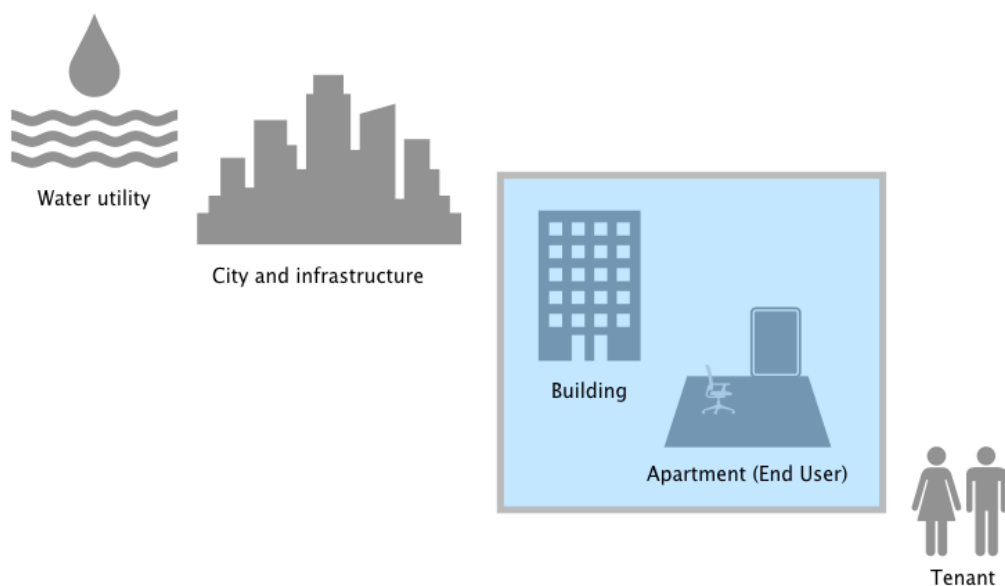


FIGURE 1.3: Scope of the study.

user. Even though other benefits might have been possible to obtain when analyzing and interpreting data, they are not covered by the scope of this thesis. As example, the literature suggest several benefits associated with end user consumption data for e.g planning purposes in the perspective of the water utility but these are however not further discussed in this study.

‘End use’ refers to where water is used. Here in this thesis the end use site is in the individual apartment, i.e tap water consumption such as water use for flushing toilets, shower and washing machines. The definition does only include water usage from indoors activities occurring in the individual apartment. This means that water consumed in outdoors activities such as irrigation are not considered. Furthermore, water used in shared spaces are not included since the water volume used can not be separated and therefore not assigned to an individual end user (an individual apartment). The thesis only addresses residential apartment buildings and does not include other end users and their water consumption, as for example water consumption in public buildings, industry buildings or single family houses.

The case study is limited to one data set from two properties of an ongoing project in Södertälje, which in this thesis is seen as one system or one building. Due to the lack of control of background factors such as socio-demographic and socio-economic aspects, the two properties could be assumed to be similar in these aspects and therefore seen as one building in order to provide more statistical significance in the analysis.

The data set used was collected from smart water meters and consist of readings of water volume used as a meter reading per hour. The water meters were already installed and the data collected before the study started and hence, there was a lack of influence over the collected data and the authors had no impact on which data to record or on which format. No adjustment in the locations of the sensors, fabricate of sensor, type of sensor, resolution or other changes in the enabling technology were possible. With that said, the focus of the study is on how to analyze and make meaning of the existing data set, not on the data gathering process. However, it is important to understand the basics around these concepts and enabling technologies and hence, they are presented in Chapter 2.

Data could be analyzed in various ways, obtaining different results. Here, the analyses made were limited to the beforehand stated types of analyses in an exploratory data analysis and the corresponding analyses in Chapter 3.4.2. The limited time of the study also restricted the number of possible analyses. Apartments with one or more error-prone sensors have been neglected from the analysis.

The aim was to identify promising areas of application in which the information extracted from the end user water consumption data could be put into use and potentially provide

benefits. The effects, impacts or an evaluation have not been part of this thesis. As example, the effect of the long term process of influencing tenants in changing behavior in regard to water consumption have not been covered in and the economical effects have not been quantified.

Finally, since the study is based on a single case study, it is hard to draw general conclusions and the thesis does not provide general knowledge. Instead, it should be seen as a possible approach or method to explore and unveil structures and patterns of a data set of end user water consumption data, that potentially could provide insights and benefits.

## 1.5 Definitions

In order to avoid ambiguity throughout the thesis, some recurring important terms and concepts are stated stipulative below.

- *End user*: Refers to where water is used, here end user is referred to as an individual apartment (single household level).
- *Operational management*: Operational management is the area of management responsible of ensuring that operations are efficient in terms of using as few resources as needed and effective in terms of meeting customer requirements, here limited to the building operational management.
- *End use consumption*: Tap water consumption at an apartment level, including both cold water and hot water consumption [Liter].
- *Accumulated water consumption*: The total water volume passing through the water meter during the chosen resolution interval [Liter/time interval].
- *Smart water meter*: A smart sensor that capture water use information in combination with a communication system that transmit information of water use in real time or near real time.
- *Individual metering*: Monitoring and measuring consumption at an apartment level, i.e the end user consumption.
- *Water demand*: The measure of the total amount of water used by the customers within a water system.
- *High frequency data*: Data with a temporal resolution of at least one reading per hour.

- *Data mining*: The gathering and analysis of large quantities of data.
- *Data driven decision making (DDD)*: An organization structure where decision making is heavily influenced by data analysis instead of intuition or experience.
- *Internet of Things (IoT)*: A world-wide network of interconnected objects uniquely addressable, based on standard communication protocols.

## 1.6 Interdisciplinarity

Interdisciplinarity, the combining of two or more different academic disciplines into one common research project, have been recognized to promote creativity, new thinking and innovation by thinking across boundaries [24] and may be regarded as a response to challenges of an increasingly complex world [25]. Moreover, according to Nissani [24], interdisciplinarians enjoy greater flexibility in their research, which governs the outcome of a good scientific research. This thesis is an interdisciplinary effort from researchers in the fields of urban water systems and ICT technologies. The approach has been collaborative with an emphasize on knowledge sharing. Anna has been studying civil engineering with a master within environmental engineering and sustainable infrastructure. Her knowledge lays among other within urban water engineering. Philip has been studying industrial engineering and management with a specialization within software intense systems and innovation. His expertise is within information technology, data analysis and business strategy.

Traditionally, the disciplinary institutions “regulate which questions to ask and which truth claims to make” [25]. However, interdisciplinary research may help to push boundaries of each discipline. Here, methods and insights of two traditional fields of study are used and combined to achieve a greater understanding of the problem. Historically there have been a major research gap between ICT technologies and water management. Lately, a growing interest have emerged into several research projects within the research area, but it is still in its infancy.

## 1.7 Disposition of the Study

This thesis is divided into seven chapters followed by a list of references and Appendixes. To get a quick overview of the disposition of the study, the seven chapters and their main content are briefly introduced below.

- *Chapter 1:* Gives a brief background to the study and problematizes the identified knowledge gap, arriving at the reason to conduct the study. Purpose and research questions the thesis seeks answers to are stated as well as the delimitation in order to clarify and specify the scope of this thesis. Further on, important terms and concepts are stipulatively defined, an explanation of the interdisciplinary concept as well as individual contributions made from each researcher is presented.
- *Chapter 2:* Gives a short theoretical introduction to concepts and enabling technologies important to understand in order to grasp the content of this thesis.
- *Chapter 3:* Justifies and describes the choice of research design and methods as well as presents a comprehensive review of the case study and its set-up. Techniques used for data analysis are emphasized and research ethics are presented.
- *Chapter 4:* Presents the literature review made and the current "state of the art" within the research field. Moreover, the interview result is presented.
- *Chapter 5:* Presents the results of the exploratory data analysis made and visualize its results.
- *Chapter 6:* Discusses how to interpret the analysis as well as how to apply the results combined with the interview results and the literature review in order to understand how potential benefits could be obtained. Moreover, important limitations and their impact on the study are discussed.
- *Chapter 7:* Concludes the most important findings and contributions of this study and suggest issues for further research.

## 1.8 Work Distribution

Since this thesis is made collaborative between KTH Royal Institute of Technology and Faculty of Engineering LTH, Lund University as the fulfillment for the degree of Master of Science and Engineering for the two authors, it is important for the academia to be able to examine the individual researcher and his/hers work load and contribution to the final thesis. The work distribution and individual contributions made are presented chapter by chapter in the overview Table 1.1.

In order to give the reader a proper introduction to the topic and its underlying drivers, Anna summarized previous research and current trends in a background chapter. This was done after thorough research in the areas of end user water consumption, water demand management, smart sensors and the use of IoT-technology within the water sector.



TABLE 1.1: Distribution of responsibility

Chapter	Part	Responsible author
Chapter 1	Background	Anna
	Scope of the study	Shared
Chapter 2	Smart cities and sensors	Anna
	Data analysis and decision making	Philip
Chapter 3	Research design	Shared
	Description of the case study	Anna
	Methods for data analysis	Philip
Chapter 4	Literature review of previous studies	Anna
	Interview with Telge Bostäder	Shared
Chapter 5	Extraction and validation of data	Philip
	Exploratory data analysis	Philip
	Visualization	Anna
Chapter 6	Interpretation of the analysis	Philip
	Application of data to provide benefits	Anna
	Limitations	Shared
Chapter 7	Conclusions and future research	Shared

The iterative process to generate a clear and well defined scope of the study was considered as one of the major obstacles throughout the study. By returning collaborative brainstorming sessions, the scope of the study was narrowed down to its final version.

Choice of research design and methods were made commonly and based upon appropriateness of the problems nature. The concepts of the methods chosen were well-known for both researchers beforehand and regarded as appropriate and justifiable methods within both disciplines. The research design was one of the main key issues and a lot of emphasis was put on designing the case study in order to produce reproducible and reliable results. For the data analysis methods and techniques, Philip used his knowledge in data analysis to construct the analysis method and Anna collected and organized all the relevant information for the case study that was not present in the used data set.

For the literature review, both authors were responsible for finding new and relevant studies on our research topic. Philip searched LTHs database LUBsearch and Anna searched KTHs database Primo. Anna conducted the literature review results.

The interview with Telge Bostäder was conducted at the 12th of June in Södertälje and both authors were present and engaging with the interview questions. An interview guide was prepared beforehand with interview questions and topics to cover in order to gather valid and relevant data.

Philip was responsible for the analysis of the data set. Firstly, an understanding of how this data had been acquired and structured into the database was essential for further analysis of the data set. After initial inspection the data was validated and pre-processed

which enabled Philip to conduct the exploratory data analysis with the help of Microsoft SQL Server and Microsoft Excel. The visualizations were made with the statistical programming language R by Anna.

In the discussion chapter, both authors have contributed with their acquired knowledge and expertise in the different subsections. Philip was responsible to look deeper into the numerical results from the analysis while Anna discussed the results with respect to previous research, the interview conducted and from a water management perspective.

However, in the pursuit of a common task the capacity to work collaboratively is an essential characteristics of interdisciplinary research [25]. To achieve the highest possible quality of the research, emphasis have been on collaboration and knowledge sharing. Perspectives and insights from both disciplines have been treated equally and with respect.

## Chapter 2

# Fundamental Concepts and Enabling Technologies

*The emerging possibilities to generate and analyze large data sets consisting of high frequency data are the result of new technological breakthroughs. In this Chapter a few key technological elements and concepts are presented relevant for the reader to grasp the content of this thesis.*

### 2.1 Smart Cities

Urban areas are responsible for major consumption of resources which puts an urge on the creation of smarter cities. The term "smart city" was introduced in the 1990s [26], focusing on the technological perspective of a city with major incorporation of new ICT technology within infrastructure and services. Many definitions of smart cities exists and there is no agreed upon definition. According to Hancke et al.[3], a smart city is defined by integrating its infrastructure and services in an intelligent way into a coherent unit. By the use of IoT for monitoring and control, higher levels of sustainability and efficiency can be ensured.

Through the use of sensors, real world data is captured and integrated into a computing platform. The collected data from the sensors become "smart" when complex analytics, modelling, optimization and/or visualization are included and applied in order to assist improved operational decisions [26]. The concept of a smart city entails a strategic decision basis, targeting sustainable development, economic growth and an increased quality of life for citizens [27].

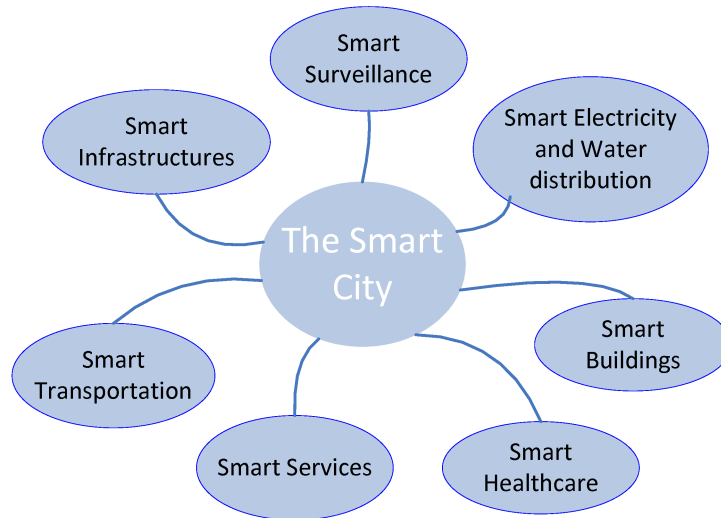


FIGURE 2.1: Elements of a smart city [3].

The water sector will play a significant role in achieving smarter cities. As understood by Figure 2.1, control and monitoring at an end user level by smart sensors is an essential part in achieving sustainability and resource efficiency. Smart buildings and smart services as well as smart water distribution are all interconnected to the gathering of high frequency end user water consumption by smart sensors.

## 2.2 Data Gathering of High Frequency Data

The velocity of information and data gathering have changed immensely in the last couple of decades. Technology is moving from storing information in batches to continuous data streams of near real time resolution. This increased spatial and temporal resolution is often called high frequency data and is defined in this study as data with a temporal resolution of at least one reading per hour.

IoT is one of the main drivers in this increased velocity of information gathering but one has to remember that not all high frequency data is generated by smart solutions or IoT devices. A sensor in its traditional meaning is fully capable of registering and communicating high frequency data.

### 2.2.1 The Internet of Things

There is no unanimously definition of what “The Internet of Things (IoT)” is and descriptions vary depending on the focal point of the research in question. One overarching definition is “a world-wide network of interconnected objects uniquely addressable, based

on standard communication protocols” [28] and the term was first introduced by Ashton in 1999 [29]. In practice this refers to a scenario where network connectivity and computing capabilities are extended to an object not normally associated with computing. Objects are enabled to generate and communicate data with minimal human interaction [30]. However, it is first in recent years that the term and usage of IoT have seen real popularity, mostly thanks to a number of enabling technologies within the fields of identification, sensing and communication [30]. Advancements within manufacturing enables small scale computation and communication units to be incorporated onto small objects and increased computer economics keeps the cost down. Access to more computational power and storage, either from in-house resources or a cloud computing service allows for aggregation, correlation and analytics of large dynamic data sets to access previously "hard to find"-information and knowledge[30].

The possibility of generating data from more and more sources holds no value in itself and is just one part of the concept called the IoT. It is with the use of algorithms and automation the vast amount of collected data can provide more insightful information and in the end create value. Figure 2.2 shows an overview of the different steps and processes in the chain that leads from data gathering to actionable information.

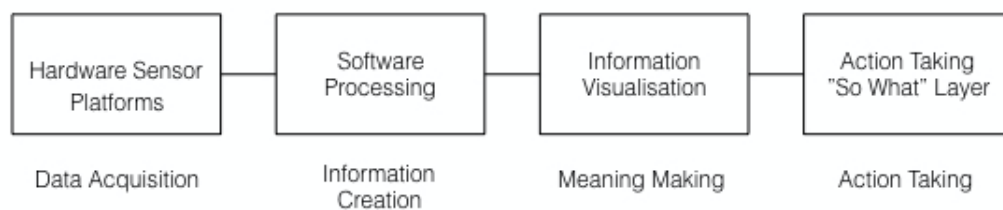


FIGURE 2.2: Process flowchart of the IoT ecosystem [4].

### 2.2.2 Smart Sensors and Water Meters

The term smart sensors was first used in the mid 1980s and has since been used to describe a number of different types of sensors [31]. There is no single consensus on what it means for a sensor to be classified as ‘smart’ but Darby [32] concludes that the key features of a smart sensor are the storing and transmission of measurements at a set frequent interval. Smart sensors used at a household level are in essence connected devices communicating over a large network connected to the internet utilizing the advancements in ICT and manufacturing technology. These smart sensors often communicate with one or several end nodes, so called sinks, that forwards the information to a dedicated server or cloud service. These kind of networks are generally referred to as wireless sensor networks (WSN). The increased quality of sensor networks regarding energy efficiency, scalability,

reliability and robustness have played a significant role for the rapid expansion of services and products using IoT [33]. The sensors associated with household monitoring can measure and transmit a number of different data points depending on sensor type, ranging from water flow and temperature to humidity and  $CO_2$  levels.

Data of water flow is acquired by a smart water meter, a smart sensor that capture water use information combined with a communication system that transmit information of water use in real time or near real time (e.g every hour or 15 min) [5]. Conventional water meters transmits low-resolution data, with maybe just a single data point per year when the meter is manually read. Smart meters generates high-resolution data, both in temporal and spatial scale.

Several types of water meters have been developed utilizing different technologies and physical properties of the water flow, such as displacement meters, velocity meters and electromagnetic meters. Obtaining the data for this study, displacement meters were used. Displacement meters requires the movement of water to mechanically record water flow. To record the volumes measured, three different methods to record consumption can be utilized as shown in Figure 2.3. An accumulation meter is the simplest form of a water meter. At a certain resolution, the total accumulated consumption during the resolution interval is sent and there is no information stored in between the sent data points. Normally, pulse or interval meters are used to measure end user water consumption which enables more easy end use analysis. A pulse is generated when a quantum of water passes through the pulse water meter and both the pulses recorded and a time stamp of the pulse are stored in the data [5]. An interval meter (also called time-of-use meters) constantly monitor the water flow through the meter and after a set time interval the volume water that passed through the meter within the interval is recorded.

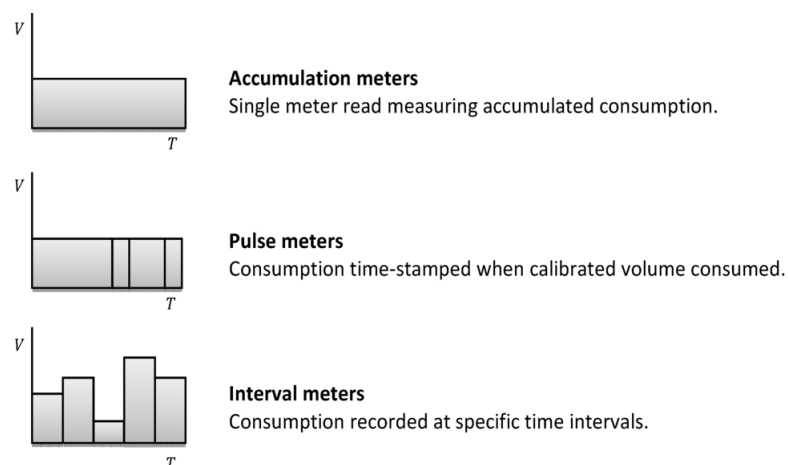


FIGURE 2.3: Methods for recording water volumes measured [5].

## 2.3 Exploratory Data Analysis

Exploratory data analysis (EDA) is an approach to data analysis that in many ways differ from the more traditional confirmatory data analysis (CDA) and was introduced by Tukey et al. in 1962 [34]. Unlike CDA which aims to confirm a given hypothesis with the support of data, EDA primary aim is to identify main characteristics of a data set as well as unveil new insights about the observed topic. This is not to be confused with browsing the data aimlessly and without goal, modeling and preconception are still required but the EDA approach simply urges the researcher not to start with a strong preconception of the data [35].

EDA as a philosophy or attitude have seen increased popularity due to the evolution of data mining within increasingly large data sets [36]. Like EDA, data mining does not stem from a strong preconception or specific hypothesis but rather looks for patterns, relationships and useful information already present in the data set [35]. This approach corresponds well with the intended goals and the research topic for this thesis.

The definition and methods associated with EDA have been a discussed topic by researchers for a long time. Behrens and Yu [37] provides one of the more recent and define the four fundamental tools for EDA as:

- *Residual analysis*. Residuals are the measured difference between a predicted outcome and the measured outcome from a validation data set, thus representing data not present in the model.
- *Re-expression (Data transformation)*. Used to improve interpretability of data by for example replacing a variable with a function of said variable.
- *Resistance procedures*. Parametric tests are sensitive to outliers and skewed distributions. Resistance procedures are used to account for this.
- *Revelation (Data visualization)*. Different forms of graphing and visualizations can be used to reveal hidden patterns and relationships.

These are all tools that can be used to uncover underlying structures, detect outliers and anomalies, maximize insight and provide a basis for further data gathering through surveys or experiments.

## 2.4 Data Driven Decision-Making

To understand the potential value of increased data gathering one has to differentiate between data, information and knowledge. The hierarchy and transformation of data to insights are often discussed within information management, information system management and knowledge management [38]. A often cited model is the Data–Information–Knowledge–Wisdom hierarchy (DIKW), as can be seen in Figure 2.4, which was first presented by Ackoff in 1989 [39]. Chaffey and Wood [40] give a good explanation of the three first steps:

- *Data*. Data are discrete, objective facts or observations, which are unorganized and unprocessed, and do not convey any specific meaning.
- *Information*. Information is data which adds value to the understanding of a subject.
- *Knowledge*. Knowledge is the combination of data and information, to which is added expert opinion, skills and experience, to result in a valuable asset which can be used to aid decision making.

The transformation processes between different steps in the hierarchy are still discussed within academia. One definition of the transformation between data and information is that information is organized and structured data, giving it relevance to a specific context and thus making it meaningful. Knowledge on the other hand can be described as "actionable information", where information is combined with understanding and capability [38].

It is this actionable information fueled by collected data that is the corner stone in making data driven decisions (DDD). DDD in this context implies that a decision is made from evidence in data rather than from personal experience and intuition. Brynjolfsson

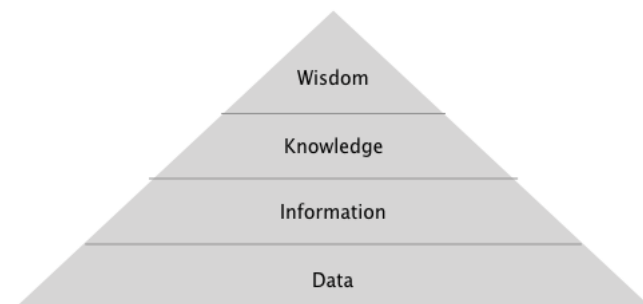


FIGURE 2.4: The DIKW pyramid presented by Ackoff.



and McElheran [41] concludes in their studies that more data open up opportunities to make better decisions and that the improvements to digital technologies vastly increases the availability of data to managers and other decision makers. However, Brynjolfsson and McElheran [41] also states that although data driven management can provide great benefits to an organization, adaptation can be slow and costly and that new techniques takes time to spread. Organization learning, scale and employee education and IT proficiency are all factors that influences a organizations ability to successfully implement DDD.

The transformation between knowledge and wisdom and its definition is a topic of great discussion without consensus between different fields of academia. However, in DDD wisdom is generally not included [38].

# Chapter 3

## Method

*The following Chapter describes and justifies the research design and research methods. Further on, the case study is described in detail and thereafter the methods and techniques for data analysis are described. Finally, the important aspects of research ethics are brought up.*

### 3.1 Research Design

Science is a mean to obtain and increase knowledge and is expected to provide answers to global challenges and to guide decision making processes that shape our societies. To be able to produce reproducible and reliable results, a transparent strategy for carrying out scientific research is needed. In Figure 3.1 the used strategy based on common methodology concepts [42] is visualized. In the design of the study, appropriate and justifiable decisions along the whole process have to be made in order to produce good scientific research.

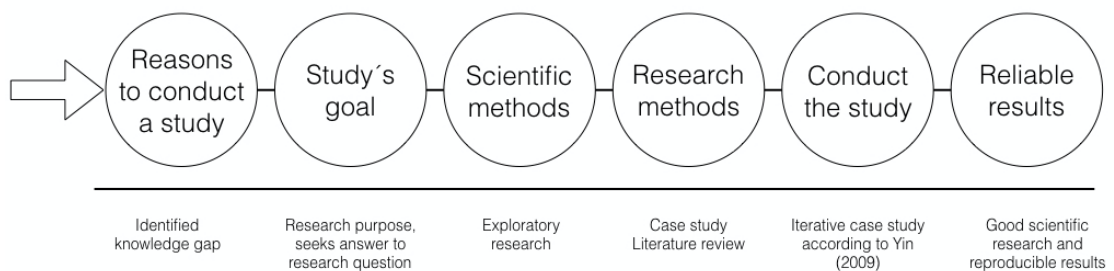


FIGURE 3.1: Research design and methods.

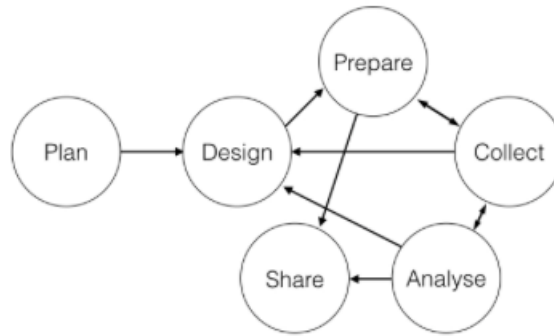


FIGURE 3.2: The reference method practice for conducting a case study [6].

The choice of scientific and research method is a crucial aspect [42]. In this thesis, an explorative research approach is used. To justify the choice of research method, one should consider the nature of the problem under study. As stated by Yin [6] “case studies are the preferred method when the investigator has little control over events and the focus is on a contemporary phenomenon”, which corresponds well to our topic of research. A case study is exploratory, descriptive and explanatory [42]. The aim is not only to increase knowledge, but also to create change in the phenomenon being studied. The approach of a case study does not require a strict boundary between the object of the study and its environment and hence, the case study is particularly appropriate for complex problems within the context they occur [43]. Even though many different research methods are available for exploratory research, the method choice of a case study seems justifiable.

Since applying proper research method practices are fundamental for the research outcome, the method described by Yin [6] have been chosen as the reference method practice due to the simple but however distinctive fact that it is the most cited book of case study research. According to Yin [6], doing case study research is a linear but an iterative process and the case study have been conducted according to the steps in Figure 3.2.

How to design the research is the most difficult step and equally the most important one. The case study design is a holistic single-case study. According to Yin [6] the single case study is fully justifiable under certain circumstances. Examples of such circumstances are when the case represents a critical test of existing theory, a representative or typical case, or where the case serves a revelatory purpose.

Case study as a method have been criticized to be less valuable, since it differs from analytical and controlled empirical studies and hence, generalization is not always appropriate. However, by applying proper research method practices, the critique can be met. It is fundamental that the research method procedures are made explicit, which governs the credibility of the study [42]. It is also important to understand the strengths

and limitations of the chosen method. Research by case studies is said to "allow for in-depth review of new or unclear phenomena whilst retaining the holistic and meaningful characteristics of real-life events" [6]. One has to remember that knowledge can be more than statistical significance.

## 3.2 Research Methods

### 3.2.1 Literature Review

A literature review has been conducted [44] in order to penetrate the research field and find the best available knowledge or "state of the art". According to Machi and McEvoy [44], six steps are to be followed in order to successfully review literature and could be regarded as a flowchart of the literature review process; 1) Select a topic, 2) Develop the tools of argumentation, 3) Search the literature, 4) Survey the literature, 5) Critique the literature, 6) Write the review.

Starting from a broad personal interest, a more specific research problem have been narrowed down during the literature search process. Firstly, available research within the research field have been used to define the conceptual structure, delineate research questions and setting limits for the investigation. Secondly, a literature review have been written.

For critically reviewing the research, searches have been made within the databases Google Scholar, KTH's database Primo and LTH's database LUBsearch. Primary sources are published peer-reviewed scientific journal papers. Examples of keywords used in the search are: IoT, end user water consumption, end use analysis, high frequency data, smart sensors, smart water meter system, water management, data driven decision making and exploratory data analysis. Primarily papers published from 2010 and onwards have been reviewed, since this specific research field is still in its infancy. However, even if the primary source is scientific peer-review papers, reports are by no means avoided. Reports are a natural way of presenting research results within the field.

### 3.2.2 The Case Study

A case study has been conducted on a data set where high frequency data of end user water consumption from a rental apartment complex in Södertälje have been analyzed. The study is limited to one data set of two apartment buildings (Property 1 and Property 2). Property 1 and Property 2 are seen as one system under study since there are major

uncertainties regarding unknown background factors affecting the end user consumption and hence, they can be assumed to be similar in those aspects and therefore regarded as one building.

The study is a holistic single-case study and the unit of analysis is the accumulated end user water consumption at an apartment level. The data set consists of water readings from 79 apartments. Four sensors are installed in every apartment and the data set consist of data between 2015-12-31 and 2017-03-31.

In addition to the data set used, data have been gathered through an interview with the public utility real estate company Telge Bostäder. Telge Bostäder own and manage the properties under study and the corresponding data set. The interview has been conducted in order to build understanding how the smart metering data is used today and how it can be applied in the future. A semi structured interview method was chosen and the interview guide used can be found in Appendix B.1. The interview guide was followed, but flexibility to ask follow up question from brought up answers was allowed. The answers were of open type.

### 3.3 Detailed Description of the Case Study

#### 3.3.1 The Case Study Area

The case study area is a rental apartment complex located in Södertälje, south west of Stockholm. A satellite map of Södertälje can be seen in Figure 3.3. According to the latest statistics, the municipality of Södertälje have a population of 94 631 inhabitants [45] and the population is expected to increase with approximately 11% within 10 years. The drinking water source in Södertälje is groundwater where water from lake Mälaren is infiltrated through the Malmsjö esker. The average water consumption in Södertälje is 160 liter per person and day [46], compared to a Swedish average of 140 liter per person and day [47]. The water cost is approximately 0,02 SEK per liter excluding fixed fees.

Telge Bostäder owns and manage around 9100 rental apartments in Södertälje. It is a public utility real estate company under the Telge concern of Södertälje municipality. The first building under study consists of 48 rental apartments. The size of the apartments vary from  $37 m^2$  to  $81 m^2$ . Every apartment has one bathroom and a kitchen. Washing machines are not generally located in the apartment due to a common laundry area in the basement. The property has 57 tenants registered. The second property consists of 36 rental apartments. The size of the apartments vary from  $37,5 m^2$  to  $71 m^2$ . Every



FIGURE 3.3: Södertälje city [7].

apartment has one bathroom and a kitchen. Similar to property 1, washing machines are not generally located in the apartment. The property has 42 tenants registered.

### 3.3.2 Description of the Data Set

The original database contains data from 47 rental apartment buildings. A majority of the buildings only have sensors installed to measure hot water consumption. Six buildings have sensors installed for both hot water and cold water consumption, which corresponds to the total end user water consumption. Out of these six residential buildings, one building was eliminated due to inconsequential installation of the sensors. Of the five remaining buildings only the two largest once were chosen due to time and software limitations of the study. These buildings, referred to as Property 1 and Property 2 are together making up the data set under study. In Appendix A.1, a description of how the data set under study can be obtained from the original database can be found.

The first building under study, Property 1, consists of 48 rental apartments. Out of the 48 apartments, five apartments have one or more sensors that are inactive or otherwise error-prone. These apartments were excluded from the study, since the total end user water consumption can not be calculated for. This means that there are 43 apartments under study in Property 1. Property 2 consists of 36 rental apartments and all sensors were correctly installed. To achieve a higher degree of statistic significance when analyzing the data, the two properties are assumed to be seen as one system or one building making up the single case-study. Technical specifications of building performance, existing fixtures and white goods etc. affect the efficiency and hence, the volume needed for a specific task such as flushing the toilet. Furthermore, socio-economic factors are significant for water consumption behaviour. Lacking information regarding these aspects, both the technical and the behavioural, they are neglected and assumed to be similar for the two apartment buildings under study since they are located in the same area. Hence, the total data set used consists of water readings from 79 apartments seen as one building.

Property 1 and Property 2 were equipped with smart meters for water measurements in 2015. Four sensors, with the properties seen in Table 3.1, were installed in every apartment. Two sensor were installed at the incoming pipe to the kitchen (hot and cold water) and the other two at the incoming pipe to the bathroom (hot and cold water). An example of a typical apartment plan with the location of the sensors can be seen in Figure 3.4. The installed water meters can be seen in a close up in Figure 3.5.

A data point of the passing volume per time unit is sent to the sink which in turn propagates the information to a dedicated server. The data set consists of over one year of data collected between 2015-12-31 and 2017-03-31, totaling approximately 3 500 000 collected data points. Each data point holds data regarding current meter readings of accumulated consumption, medium (hot or cold water), time stamp, sensor ID, sensor settings and spatial information.

TABLE 3.1: Properties of the installed water meters

<b>Meter information and properties</b>	
Fabricate	Bmeter
Model	GSD5
Recording method	Accumulated consumption
Unit type	Volume
Metric	Liter
Time resolution	One reading per hour

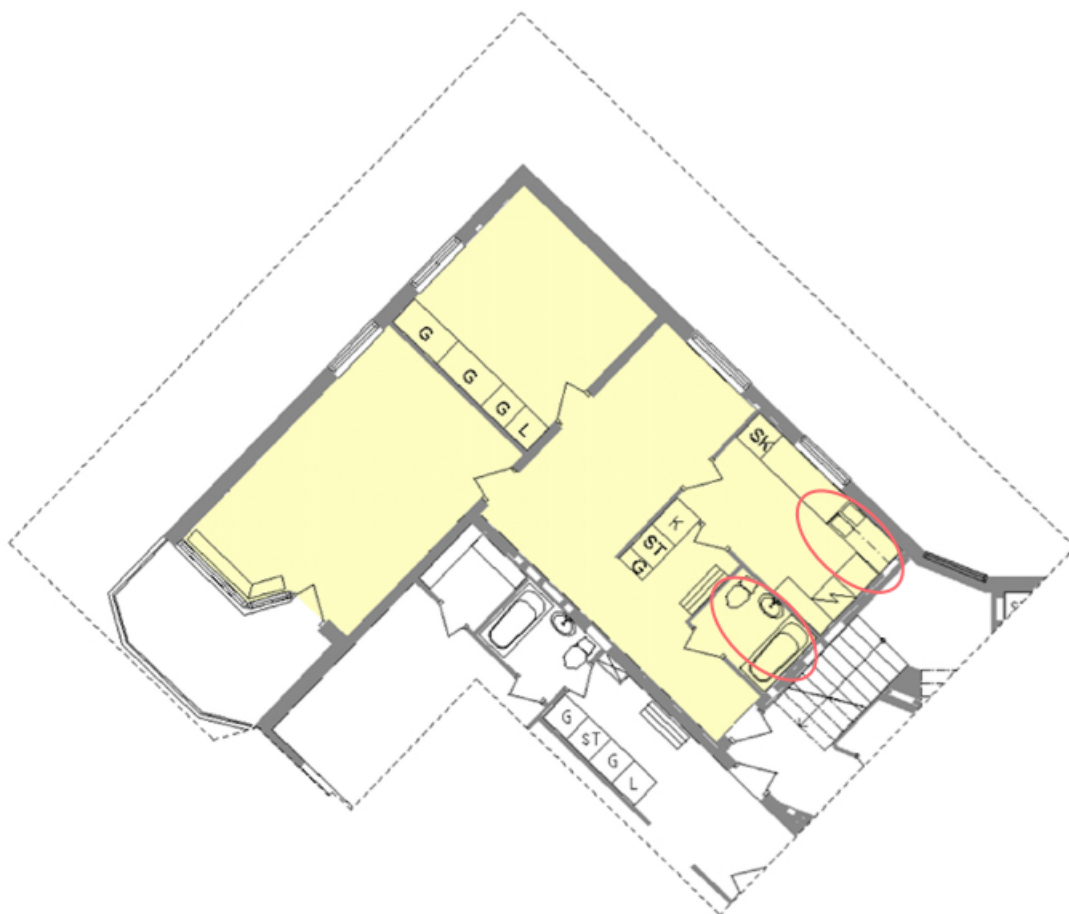


FIGURE 3.4: A 2-room apartment of  $63 \text{ m}^2$  with sensor locations marked in red.



FIGURE 3.5: Installed water meters at the incoming pipe to the kitchen.



## 3.4 Methods and Techniques for Data Analysis

The best research collect and uses data in an original way or offers interpretation of existing data in an new and innovative way [48]. The data is of high temporal and spatial resolution and although from a relatively small number of apartments consists of a large number of data points. Generation of high resolution data enables applications of data analytics tools [16].

### 3.4.1 Data Pre-processing

In order to produce reliable results and enable accurate interpretations of the collected data, a deeper understanding of the data structures and potential flaws had to be achieved. The data set analyzed in this study had several properties that made it imperative to pre-process and validate before proper analysis could begin.

- The sensors had been installed over a long time period and had known inconsistencies within the installation and gathering process.
- The data set had been stored in different parts of the database with some overlapping and redundancies.
- The data had been transferred between different systems.

In addition to this, the chance of faulty sensors and corrupt data previously being detected where minimal, since the data is not used regularly. In summary, precaution had to be made in order to produce reliable results.

Detecting outliers by pre-processing the data is an important precaution. An outlier can be either a natural variation in the data or the consequence of wrongly recorded measurement, the latter which should be excluded in order to achieve correct results [49]. Determining if an outlier exist due to extreme variation or incorrect measurement is a challenge and up to each researcher to determine.

The data was first scanned for faulty or incorrectly installed sensors which includes sensors that had been sending negative values, decreasing values for aggregated consumption or registered unreasonably high values regularly. Due to the large amount of data, these checks were initially conducted on an apartment aggregate level to reveal obvious flaws.

As mentioned, the installation process of the sensors had been inconsistent and some sensors were installed backwards, making sensors record decreasing values of accumulated consumption. To identify those sensors, minimum and maximum monthly meter reading

values for each sensor were compared to assure that meter reading values were increasing. All negative values found in the data set and their corresponding sensors were excluded in the analysis.

Once obvious faulty sensors were identified and excluded from further analysis, a more thorough investigation was done to verify that all sensors had continuously been transmitting data throughout the time period under study. This was done by counting the number of data points associated with each sensor, ensuring a correct number of data points transmitted over the given time period.

Each data point consists of a meter reading value. Identifying improbable spikes in consumption was conducted in a top-bottom approach. Monthly consumption per apartment were calculated and apartments with major outliers were identified. Next, the consumption for those apartments were calculated on a daily basis. This was done in order to investigate if the high monthly values were due to a continuous high consumption or due to an unreasonable increase in reader values. This process of investing abnormal consumption on an increasingly more granular level continued until several hourly readings with unrealistic high readings were identified and excluded.

Table 3.2 shows a summary of the described and identified errors in the data set.

TABLE 3.2: Type and number of errors found in the data set.

Type of error	Errors found	Description
Faulty sensors	20	Sensors continuously sending negative values or counting backwards.
Sensors with downtime	144	Sensors which record and send correct data but with instances of downtime.
Unreasonable consumption	3	Sensors with sudden unexplained high values but overall correct data.

### 3.4.2 Data Analysis Method

Due to the nature of the research problem, an EDA was conducted where the goal was to identify characteristics and unveil new insights. Analysis was conducted with the help of a wide array of analysis tools, such as SQL-editors, Microsoft Excel and the statistical programming language R.

Finding relevant information in the data set requires using several analysis methods in conjunction as well as cross referencing findings from multiple methods. Below follows

a method description of utilized methods to deduct information from the data set and how they correspond to the fundamental concepts of an EDA.

### 3.4.2.1 Resistance Procedures

Baseline values have been extrapolated in order to put other de-aggregated values into perspective. This was done by extracting consumption data for each apartment using SQL. The SQL-code can be found in Appendix A.2. After extraction, the data was imported to Microsoft Excel and modified using excel functionality.

Box plots are an important tool to understand the distribution, spread and assert robustness for different metrics. The following properties were calculated for each metric under study: minimum, maximum, median, 1st quartile, 3rd quartile, interquartile range (IQR) and whiskers. Whiskers are calculated as  $\pm 1.5 * IQR$  from the third and first quartile respectively. The calculations were made using R studio.

### 3.4.2.2 Residual Analysis

An analysis of the relation between apartment size and water consumption was conducted with the help of a linear regression analysis. The total water consumption per apartment and apartment size was extracted using the SQL-code in Appendix A.3. The data was imported to R Studio and a regression analysis was made using the R language regression tool.

### 3.4.2.3 Revelations

To investigate seasonality in water consumption on different time scales, data was extracted and grouped by time stamp. The different grouping criteria were by hour, week-day and month. The relevant SQL-code can be found in Appendix A.4. The average total consumption was also segmented into hot and cold water consumption for the monthly seasonality analysis.

An attempt to detect potential leaks have been done in order to prove possibility of such actions. Hourly data from all 316 sensors where analyzed over a time period of one month between 2016-10-01 and 2016-10-31. In order to see if there were sensors registering water consumption continuously over a 24 hour period, thus indicating leakage, each data point was categorized into two categories; "Consumption" or "No consumption". A detailed description can be found in Appendix A.5.

Visualizations of the data analysis have been done in R studio, a free and open-source integrated development environment for statistical computing and graphics. The free extension package "ggplot2" have been used. Visualizations have been designed in an exploratory manner, where a good exploratory visualization explain for the viewer what is going on and allows to intuitive develop an understanding of the data.

### 3.5 Research Ethics

Research ethics process moral questions that arises within science and research [50]. The scientific community is built upon trust; trust that the reported research results are based on an honest and accurate reflection of the scientific work and that the researcher have been using appropriate methods and techniques for analysis of data [51]. Decisions about research design and how to present results have been justified in order to achieve trustworthy research. One must also remember the limitations of the research and in particular the chosen research method's limitations.

Ethical principles following the CODEX guidelines of the Swedish research council have been applied in this thesis [52]. One research ethics aspect that needs extra consideration in this thesis is the ethical principles during gathering and treatment of data. Personal integrity of tenants have been protected by obtaining anonymized data. The spatial location of the rental apartment buildings under study, as exact address or property name, are not known and have not been published.

Misleading or inaccurate data as well as inappropriate interpretation of data have to be considered a major violation against the code of conduct of scientific research [51]. Hence, it can not be repeated to many times that the research design is of critical importance for maintaining the integrity of the research.

## Chapter 4

# Literature Review and Interview Results

*This Chapter presents the results from the literature review, which was conducted in order to find the best available theoretical knowledge or “state of the art” within the research field, delineate research questions and identify knowledge gaps in present literature. Prior research on end user water consumption by smart metering and its applications and corresponding benefits have been reviewed. Further on, this Chapter presents the results from the interview with Telge Bostäder, the owner and operational manager of the buildings under study.*

### 4.1 Previous Studies

The water sector is increasingly focused on the installation and usage of smart water meters, since they have several recognized benefits compared to conventional meters [5, 53] and the capacity to deliver increasing amounts of data to both planners, water utilities, managers, government organizations and customers [54, 55]. However, with growing amounts of data available, new questions are raised. The technical capabilities, such as data gathering and information systems that store and manage data, do not pose as much of a problem as the question of what to do with all this collected data [5]. This study is addressing this problem by not only gathering and analyzing data, but applying the information found for a specific purpose.

High frequency data in real-time or near real-time is communicated and enables instant information on consumption. As up to today, the average daily water consumption in Sweden is estimated to 140 liters per person [47] and approximately distributed between

the end use categories as; Personal hygiene 60 [L], Toilet flushing 30 [L], Dishes 15 [L], Laundry 15 [L], Drinking and preparing food 10 [L], Other 10 [L]. In a report by Sweden's Energy Council [56] an average end user water consumption of 184 liters per person and day was found when investigating data of hot water consumption from nine apartments over one month. The average use of hot water per person and day was 58 liters, which accounts for 32% of the total water volume used. Another cited consumption estimation in Sweden is the SABO-estimation which estimates the yearly average end user hot water consumption by living area as  $0.40 [m^3/m^2]$  [57]. However, it is recognized that the average consumption distribution is neither linear nor normally distributed. Results in the report by Sweden's Energy Council [56] showed considerably different consumption patterns between different households and concluded that more knowledge is needed to understand how water is used at a household level. In a study by Chen et al. [58], it is concluded that water consumption is affected by a number of different variables such as number of household members and socio-demographic factors.

It seems basic, but there is a defined lack of important knowledge on how we use water in our homes [56, 59]. An increasing number of studies focus on the end user and emphasizes the importance of detailed knowledge [21]. Knowledge of by whom, when and how water is being consumed is becoming more and more accessible. However, interpretation and analysis is needed in order to transform the data into information [22]. There is no standard method how to use the processed high frequency data to create information and the research field is still at a developmental stage [5]. This study proposes a new and an unexplored method approach to reveal information by the use of an EDA and could be seen as a further advance within the research field.

Overall, analysis of data enables a new and, at this time, underexplored opportunity to manage water more efficiently [60]. Of all studies reviewed, a great majority are small-scale investigations or implementations. In general, benefits found in the studies reviewed could be divided into a few categories depending on applications as well as purpose.

- Improved operational management
- Cost-effective measures
- Empowering end users
- Business and campaign opportunities
- Integrated Decision Support System

Worth mentioning, even if not included in this study, are the applications of high frequency data within infrastructure planning and management of the distributions system.

Several applications and their associated benefits could be found in the literature. As example, mining of end user consumption data could forecast future water consumption trends, provide a basis for end user modelling for a better understanding of the systems hydraulics and help to develop effective and powerful water demand strategies in order to reduce overall consumption [5, 11–14, 16, 61].

End user consumption data have the ability to assist in the building operational management [62]. End users with excessive consumption could be highlighted and indications of inefficient use or leakage could be given [55]. It is common knowledge that leakage within the distribution system is excessive and much has been written about leakage identification and control of the system network. However, post-meter household leakage, i.e. water losses located within the residential property boundaries, have not received as much attention by the utilities or the scientific community [54]. Post-meter household leakage is often harder to detect since current conventional metering systems are not able to provide detailed water use information. However, post meter leakages is estimated to account for up to 10% of the total water consumption and is particularly noticeable in the residential sector. Britton et al. [54] concludes that smart metering data is a powerful tool for managers to rapidly identify leakage. The ability of customer leakage identification and corresponding action measures is according to Stewart et al. [55] one of the key benefits of smart metering [54].

Furthermore, by highlighting indicative consumption patterns such as excessive consumption, the operational management can be improved. By identifying water consumption patterns of different types of consumers, a deeper understanding of different categories of end users are enabled. In a study from 2013, water consumption data were disaggregated into a registry of end use events in order to pin-point end use activities [63]. In a study from 2016, a new algorithm to find routine behaviours was suggested and further on it was demonstrated that knowledge extracted by finding routines can be used to manage water more efficient [64]. An interesting application can be found in a study from Ferreira et al. [62] where unusual patterns are pin-pointed in order to find patterns caused by wastage. A few softwares have been developed to categorize end user consumption, such as Identiflow and Trace Wizard, which are based on a decision tree algorithm, or HydroSense, which is a probabilistic-based classification approach [16]. Chen et al. [58] developed a benchmarking model and the results indicate a successful prediction of water usage at the household level. By using historic end user data, the model learned and became better and better in predicting consumption, even for a specific purpose as kitchen water usage. However, in general there are still needs for development to decrease the human interaction and increasing accuracy.

Data can also be used in a cost-benefit perspective, to target resources and options that provide the greatest savings at lowest cost [55]. No-cost or low-cost measures, frequently called “low hanging fruits”, are measures yielding the greatest water savings at lowest cost, often by corrective measures. In a study from Ferreira et al. [62] it concluded that most of these potential savings were not physically visible and hence only detected through analysis of data. Savings as well as an overall improved operational management were achieved. One of the most common of these “low hanging fruits” is the leak identification. By correction measures, such as repairing water leaks, major water savings as well as energy savings could easily and cost-effectively be achieved [62].

One should remember that a key challenge is to ensure that there are benefits for both the manager and the customer [5]. One way to use data from smart meters is to empower the end users, i.e. the customers and provide them with more frequent and detailed consumption information and feedback in order to encourage water savings [5, 15, 60]. To use consumer feedback to encourage conservation have been widely applied and evaluated in other sectors, such as the energy sector, with strong evidence for consumption reductions ranging from 5 % to 20 % [53]. Fielding et al. [65] were the first to use smart water metering data as a tool for behavioural change as late as in 2013. To achieve behavioural change, the importance of high-frequency, comprehensive and individual customized feedback including easily interpretable visualizations is emphasized. In a review article, the effectiveness of the use of different feedback technologies and methods designed to promote water conservation by the end user were highlighted [53]. However, no conclusive evidence could be drawn such as for the energy sector. Some studies indicate that the feedback is effective in reducing consumption in a short term perspective, but long term effects of feedback are not sustained and water consumption often returns to baseline levels after some time [65]. Nevertheless, several recent studies indicate effectiveness in managing water by the use of feedback, suggesting reductions ranging between 2.5% up to 28.6% [53]. Liu et al. [60] evaluated both the effect of providing end users with feedback as well as the important aspect of the feedback design. According to research, feedback should preferably include consumption patterns, changes over time as well as social comparisons [60]. However, more research is needed to fully understand which kind of feedback that is preferable when aiming for behavioural change.

If seeing things differently, high frequency data has the potential to provide an improved customer experience and customer service [5]. By providing the end user with easily interpreted information regarding their consumption at near real-time, they are empowered to actively manage and control their consumption. For example, consumers could track their consumption by their personal water consumption web page [55]. Information on daily, weekly and monthly consumption could be easily presented to the



customer as well as comparisons, categories of water end-use, alerts for leaks, high consumption alerts etc. Water bills based on actual consumption rather than estimated consumption could be used [53], i.e individual metering, and billing can be updated at a daily or even hourly basis [55]. In addition to access to information and services, a greater transparency for the end user is achieved, which promotes action taking and proactive management of consumption.

Another application is the potential for data mining business opportunities [5, 15]. One simple example is marketing and advertising, where products or services can be promoted to different end users depending on consumption patterns, such as beauty products for users using the shower more frequently. On a neighborhood level, plumbers could be alerted if the area has a high occurrence of leaks and hence, the plumbers may choose to extend their advertising in this area. Cardell-Oliver et al. [15] analyzed consumption data to find different groups of end users who utilize water in a similar way. Prevalence and significance by each group as well as their peak hour, peak month, frequency and intensity were found and significant differences between the groups were discovered. With the use of these consumption groups, situations were identified where small-scale interventions could be targeted with a more effective result in reducing consumption instead of broad scale campaigns. Worth mentioning, even if not further discussed in this thesis, is the data privacy issues when analyzing end user consumption data. As brought up in several studies, there is an obvious need for regulations that govern the privacy of customer data and information, however it is not clear how to handle those questions up to today [5].

As seen by the different applications described, a wide range of reports could possibly be generated manually or automated by processing end user consumption data and made accessible to different users. Current approaches to water end use analysis are time consuming and requires manual processing [5]. Automated reporting tools utilizing the processed data are still at a developmental stage, but are needed in order to reach scale. To transform high frequency data into useful information, Stewart et al. [55] proposed a Web-Based Knowledge Management System, which integrates end-use consumption data, wireless communication networks and information management systems. Information on how, when and where water is being consumed could then be provided in real-time or near real-time for consumers as well as managers. Such a system have the possibility to enable data to be used by both customers and managers and hence transfer water consumption data and information into water consumption knowledge.

The European Union funded project Integrated Support System for Efficient Water Usage and Resources Management (ISS-EWATUS) was initiated in 2014 and is an innovative integrated decision support system under development which aims efficient management of water resources by recognizing and exploiting untapped potentials to save water [59,

66–68]. An information system for gathering, interpreting and sharing data about water consumption at the household level is planned, i.e a household decision support system, in order to increase awareness and further save water [66]. The interpreted data will be used for behavioural change and presented to end users using mobile devices. A social media platform is also planned in order to "reinforce the water-saving behavior of consumers by means of social interactions among people and also to link consumers and experts on water-saving techniques" [66]. At a higher urban level, the main goal is to reduce water leaks.

Despite an increasingly large number of papers published over the last years, there is a clear identified need to shift research efforts to a more integrated approach. As Cominola et al. [16] concludes, the majority of the studies focus on a specific and specialized method for analyzing data and its corresponding benefits. This study makes an effort to broaden the perspective and give a more conclusive overview of the potential benefits associated with smart meter water data. Addressing the question about the usefulness of the information rather than analyzing the data for a specific purpose have not been done in the reviewed studies.

## 4.2 Interview with Telge Bostäder

The interview with Telge Bostäder (TB) was conducted at the 12th of June 2017. Daniel Bäcklin (DB), engineer and responsible for energy questions at TB, was interviewed. Due to a reorganization and personnel shortages within the organization, DB is responsible operation manager at TB and hence, suitable to answer questions regarding end user water consumption. General information about the interview, contacts and the interview guide used can be found in Appendix B.1.

Hitherto, there is no outspoken interest or strategy in place for questions regarding water consumption and demand management in general. However, DB is quite sure that the issue will be higher regarded in a more long term strategic planning. In operational management, water is a large financial cost. The total cold water consumption of TB and their rental apartments population is approximated to 40 millions (SEK) per year, a major part of the total operational costs.

Sustainability is important to TB and the primary target is to reduce energy consumption according to the 2012 Energy Efficiency Directive of the European Union. Indirectly, the hot water consumption is a target of interest since there could be major savings in energy associated with a decrease in end user hot water consumption. More resource efficient

fixtures have already been installed with the primary goal to increase the efficiency of hot water consumption at the end user level.

#### **4.2.1 Data Gathering**

Water consumption within TB's properties are monitored and measured in two different ways. Two thirds of the population have conventional measuring and is monitored by one single water meter for the whole property. The meter is located in the mechanical room. The water meter is manually read every six months by the building manager where the meter reading is written down on paper and handed in to DB. DB then transfer the handwritten data point to an excel-sheet, which is saved and sent to the distribution system operator, Telge Nät, who uses the data point to bill TB for the total water volume consumed. The measured unit is accumulated consumption as a meter reading, which has to be compared to the previous meter reading to know the actual accumulated consumption within the past six months.

Approximately one third of the population of apartments have installed sensors at an end user level measuring either hot water consumption and cold water consumption, or as in most cases, only hot water consumption. Irrespective, the total water volume for the building is monitored by a single conventional water meter as described above in addition to the sensors. The project of installing sensors for individual metering was initiated in 2013 due to the 2012 Energy Efficiency Directive. According to the 2012 Energy Efficiency Directive, energy consumers should be empowered to better manage consumption. This includes easy and free access to data on consumption through individual metering. At the moment, around 2000 rental apartments within the population of TB are generating data of high or medium resolution.

TB is the owner of the data gathered, both the data collected with sensors and the conventional manually way. The data gathered from the sensors is stored in a database. A limited number of people have access to the database; DB, property managers at TB and the database administrators. However, within TB its only DB who uses the database now and then. It is from this database that the data set used in the case study originate.

When asking about how the rental apartment buildings for the sensor installation were chosen, as well as technical parameters such as resolution, type of sensor etc, no answer can be given. DB have tried to find information specified above, without results. If a proper decision basis existed, it can not be found.

### **4.2.2 Data Usage**

One direct usage of the data collected by the sensors is the possibility to individually debit the tenants for their actual consumption, which was the primary goal for TB when initiating the project of installing individual metering at an apartment level. Economical as well as environmental benefits were expected, since a more effective monitoring would provide an increased control. Today 24 end users are being debited for their hot water consumption. Except from the individual debiting from these 24 apartments, the data gathered is fairly ever used. The project of installing sensors have now been put on hold. Financially, the installation costs per apartment have almost doubled compared to the calculated cost and furthermore, the organization representing the tenants turned their proposal of individual metering down. No agreement of a normal consumption could be established and hence, neither individual debiting. Except the possibility to debit the tenants by individual metering, there is not any outspoken interest in how the end user consumes water.

In small scale at a building level, the data generated by conventional meters is manually used as identification of errors such as unusual meter readings or identification of possible leakage. The two data points per property and year are manually transferred to an excel-sheet and then sent to the distribution system operator, who sends a bill for the consumed water within a property. When DB gets the manually read data from the water meters, he quickly checks if the data seems plausible. Plausibility is judged upon number of apartments the water meter supply and a comparison to previous meter readings from the same property. If the data point is way above the previous one, one could easily suspect a leakage. When a suspected leakage is identified, DB often initiates an investigation. He often make the investigation by himself by inspecting the property and the mechanical room. Sometimes a leak can be identified and found by having a look around. If nothing unusual can be seen by inspection, its harder to identify the location of the leak. The apartments are not inspected. Another investigation method is to put up a camera in front of the single water meter in the mechanical room to see if it is still running at high speed during the nights, when the consumption is expected to be close to zero. If yes, one may suspect a leakage.

Sometimes it is when the invoice reaches the building manager at TB it can trigger actions. If the invoice has increased since the last one, DB often receives questions regarding the water consumption. The manager wants to know if there has been a specific event causing the increase or if the increase is plausible, to eliminate the chance that the operator have made an incorrect invoice.

### **4.2.3 Assessing Potential in Data Driven Operational Management**

There is a strong ongoing digital trend, which is also affecting TB. As example the tenants nowadays have digital access to a personal web page, there are digital meeting places and digital portal screens. To digitize management operations, DB mention a more excessive database, where all information for every specific property will be gathered and stored.

Lack of knowledge how to use the data, except than for debiting the tenants, is one of the main reasons the data is not used within operational management according to DB. However there is a will and an openness to use the data if it can generate valuable information. What valuable information for TB could be does not DB know. In general, sensors are promoted to be able to generate valuable information. However, any specific cases and applications are hard to find and DB would like to know what specific value a data driven management would generate to their operational management. Further on, there is not any information about the end user that DB thinks would have been useful that they don't have access to today. The first step is to use the data TB already are gathering and have stored.

When asking DB how he thinks the data could be applied, DB have had some ideas of his own, but up to today the ideas have not left his office. One option seen is to use the data to influence the tenants and curbing their consumption. By influencing behaviour, as example through visualization of the consumption via the end users individual digital page ("My page") or by a water saving competition between different residential buildings. However, since the apartments are rental apartments, DB have some concerns and it could be questionable if the tenants would engage with the question of water consumption if not being debited for their individual consumption.

Another possibility seen is to be able to regulate heating at an apartment level in order to save energy. When hot water is being consumed, as when showering, the input temperature for the radiators could be decreased since the hot water is carrying energy and contributing in heating the apartment. Another application area of the data could be for diagnostic purposes. On a building level, a general increase of cold water consumption could reveal an overall low pressure in the building. Further on, leakage identification might be relevant. Automated alarms for different diagnostic purposes would be a scenario that could come handy.

# Chapter 5

## Analysis

*This Chapter presents the results from the EDA, where the data set under study have been analyzed without beforehand strong preconception of the data in order to see what information about the end user that possible could be extracted. The Chapter is divided into different sections according to the type of analysis made.*

### 5.1 Average Consumption and Baselines

The average daily consumption for the population is 174 [L/day]. Figure 5.1 shows a histogram of the consumption distribution over the total population of apartments as well as a density plot combined with a few selected summary statistics for the average

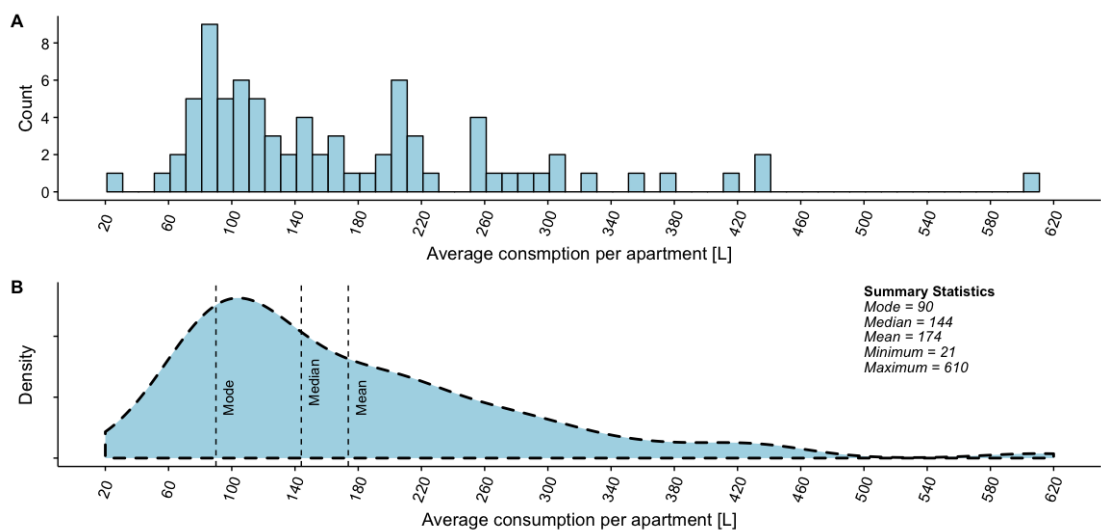


FIGURE 5.1: A. Histogram of average daily consumption per apartment. B. Density curve with summary statistics over average daily consumption.

daily consumption. The consumption distribution ranges from a minimum of 21 [L/day] to maximum 610 [L/day], with a majority of the end-users utilizing between 60 and 220 [L/day]. As can be seen in Figure 5.1, the distribution is right skewed, with a mode, median and mean displayed as dotted lines within the density graph. Looking closer at the spread of consumption, the top 20% (16 in total) apartments with the highest average consumption consume a bit over 40% of the total water volume.

Baseline values are provided in order to get an overview of the consumption. Table 5.1 presents a summary for the whole population between 2016-04-15 and 2017-03-19. As seen in Table 5.1, the ratio between cold and hot water consumption is on average around 50%. However, this ratio differs considerable throughout the population, which can be seen by the spread in Figure 5.2. In general, more water is consumed in the bathroom compared to the kitchen, as seen by the ratios in Table 5.1.

TABLE 5.1: Summary Table with statistics for the whole population.

Metric	Value	Ratio of total
Number of apartments	79 [-]	
Number of tenants	99 [-]	
Total consumption	4 710 700 [L]	
Average consumption per apartment	59 629 [L/A]	
Average consumption per tenant	47 583 [L/T]	
Total consumption hot water	2 246 980 [L]	
Total consumption cold water	2 463 720 [L]	
Average consumption hot water	28 443 [L/A]	47,7%
Average consumption cold water	31 186 [L/A]	52,3%
Average consumption kitchen	16 305 [L/A]	27,3%
Average consumption bathroom	43 324 [L/A]	72,7%

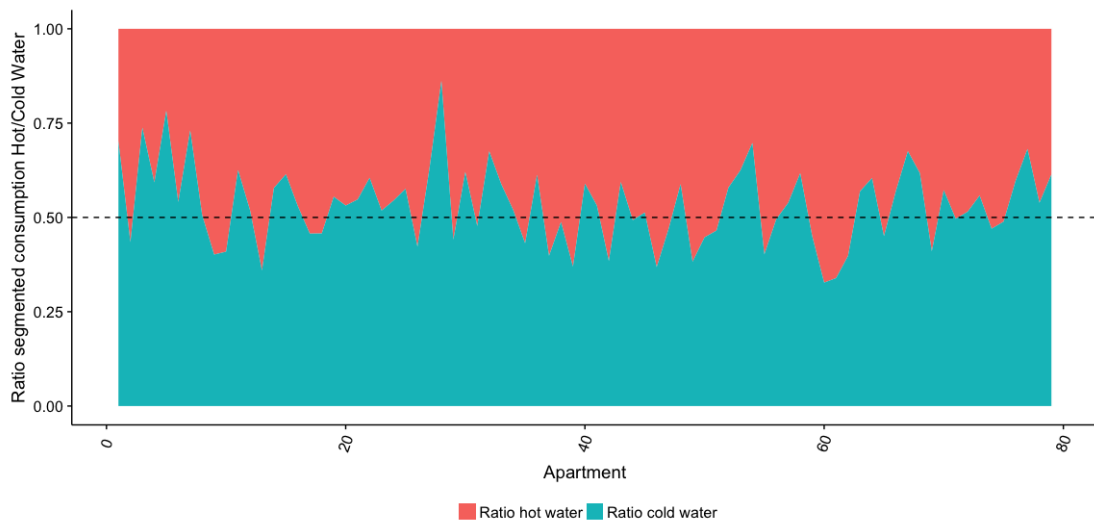


FIGURE 5.2: The ratio between total consumed hot and cold water per apartment.

## 5.2 Seasonality

Seasonality analyzes on different time scales were made and are presented below. Figure 5.3 shows the difference in total average consumption per month and Figure 5.4 shows the average monthly consumption segmented by hot and cold water. The box plots to the right in both Figure 5.3 and Figure 5.4 show the distribution between apartments within each month. The data used was collected between 2016-01-01 and 2017-02-01.

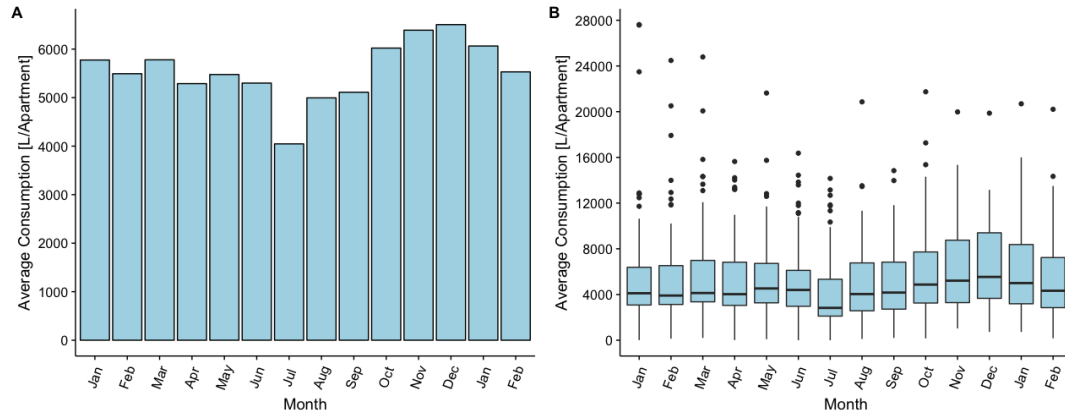


FIGURE 5.3: A. Bar chart visualizing monthly seasonality on average consumption. B. Box plot of monthly seasonality on average consumption.

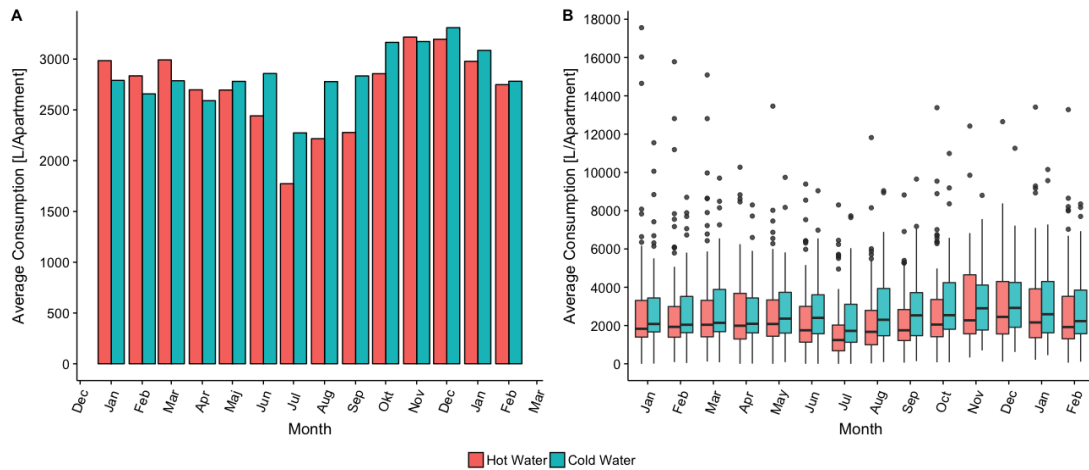


FIGURE 5.4: A. Bar chart visualizing monthly seasonality segmented into average hot and cold water consumption. B. Box plot of monthly seasonality segmented into average hot and cold water consumption.

As can be seen by Figure 5.3, there are some variations during the year, with a minimum in July and a maximum in December. For January and February, the average values are almost equal for 2016 and 2017. Looking at the segmented water consumption, Figure



5.4 shows an overall more or less equal distribution of hot and cold water consumption during the studied time scale with an exception of an increased cold water consumption during the summer months.

Figure 5.5 shows the average daily water consumption per apartment for different weekdays. The box plot to the right shows the distribution between the apartments for each weekday. The data used was collected between 2016-04-15 and 2017-03-19. As can be seen by Figure 5.5, there are no significant differences in water consumed during the week.

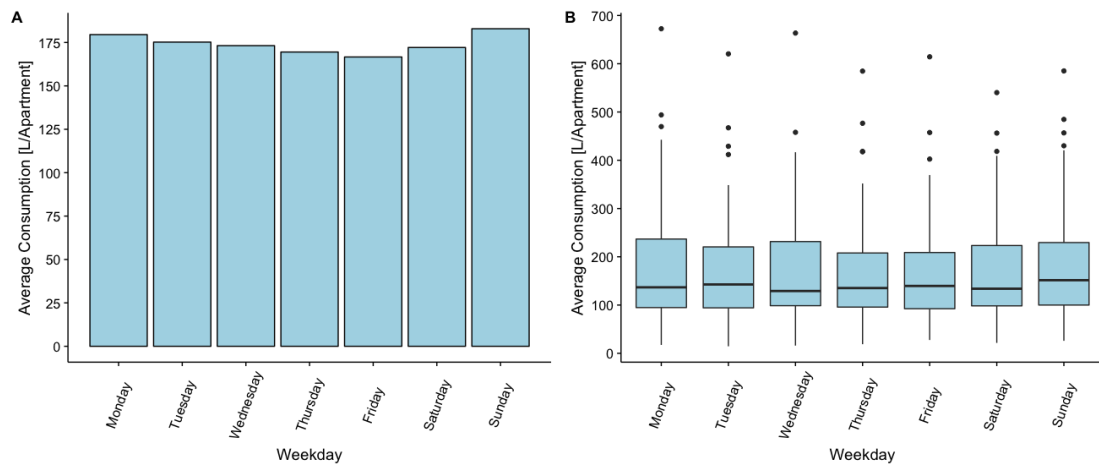


FIGURE 5.5: A. Bar chart visualizing the daily variations over the weekdays of average consumption. B. Box plot of daily variations distributed over the weekdays of average consumption.

Figure 5.6 shows the hourly average consumption per apartment. A higher consumption are shown at noon and evening time and significantly lower values between midnight and 6 am. Figure 5.7 shows a box plot over the distribution between apartments for each hour. The data used was collected between 2016-04-15 and 2017-03-19 for both figures.

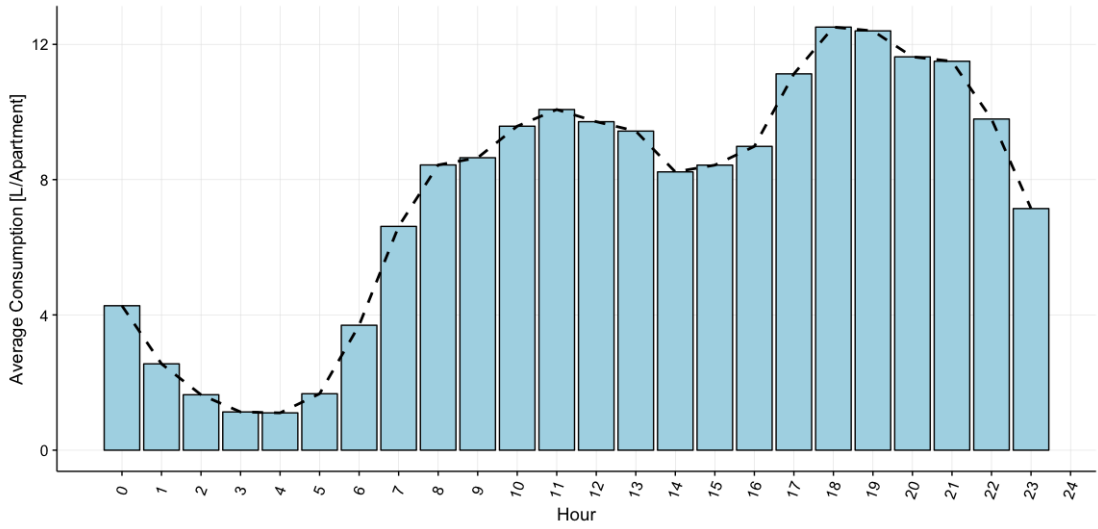


FIGURE 5.6: Hourly variations of average consumption visualized as a bar chart.

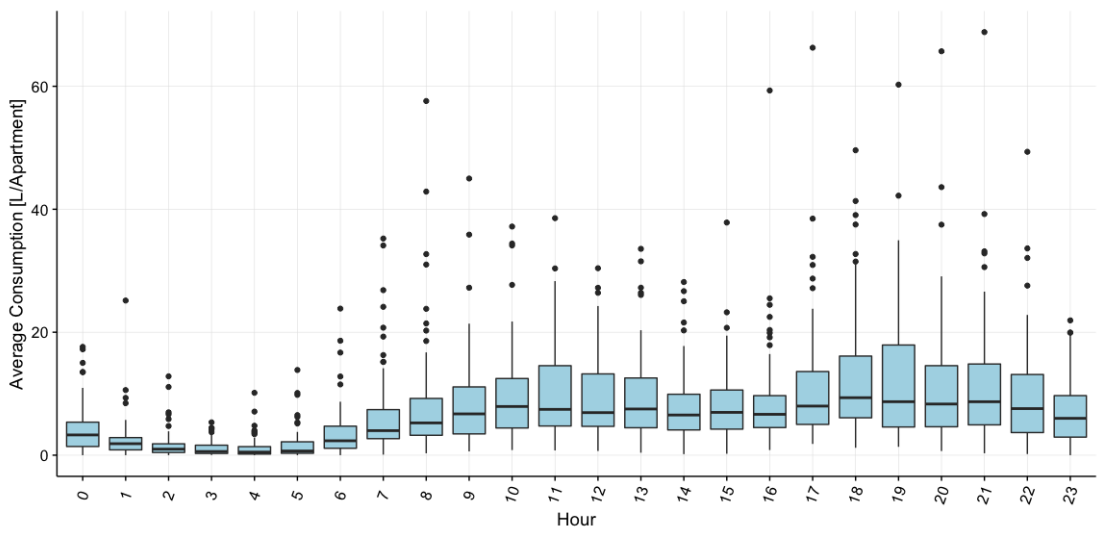


FIGURE 5.7: Hourly variations of average consumption for every apartments visualized as a box plot.

### 5.3 Regression Analysis

Regression analysis is used to identify the relationship and dependencies between two or more variables, where the end user water consumption is a variable in the analysis. The data used was collected between 2016-04-15 and 2017-03-19.

Guidelines for water consumption is sometimes expressed as liters per square meter [57]. Figure 5.8 shows the relation between apartment size and consumed water per apartment. The dotted line shows a linear fit trend line. The dots are representing the total consumption per apartment and are slightly displaced to the left or right for visualization purpose due to several identical apartment sizes. A summary of the regression statistics can be found in the upper left corner of Figure 5.8

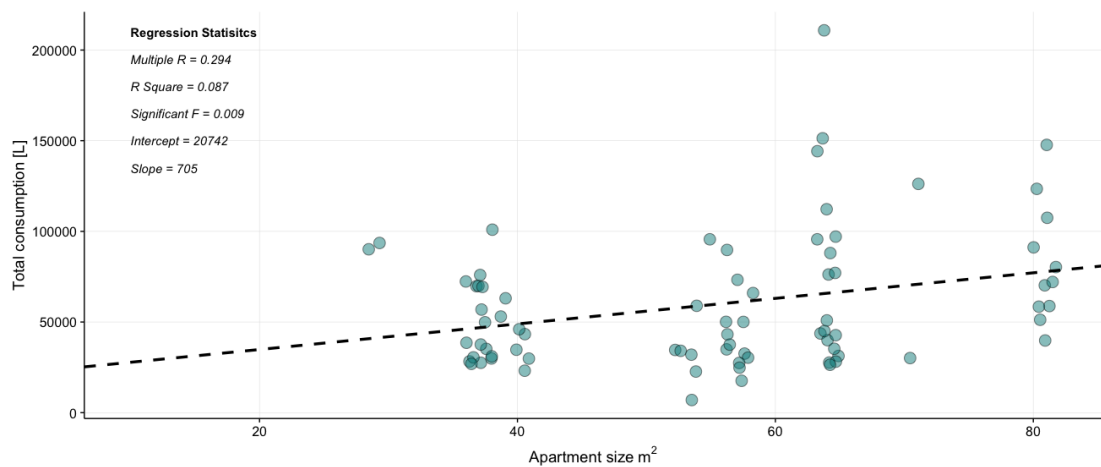


FIGURE 5.8: Graph visualizing correlation between apartment size and total water consumption with a linear regression line.

## 5.4 Clustering

Clustering is used to identify groups of users with common behavior by finding signature patterns. Figure 5.9 shows each apartments peak hour in a histogram (to the right in Figure 5.9). Peak hour consumption is calculated as the hour with the highest average total consumption over the time period 2016-04-15 to 2017-03-19. Figure 5.9 (to the left) shows a scatter plot over each apartments minimum and maximum water consumption month. The blue dots are slightly displaced to the left or right for visualization purpose. The red transparent dots are helping to visualize an area where several end users have their maximum and minimum. The data used was collected between 2016-01-01 and 2017-01-01.

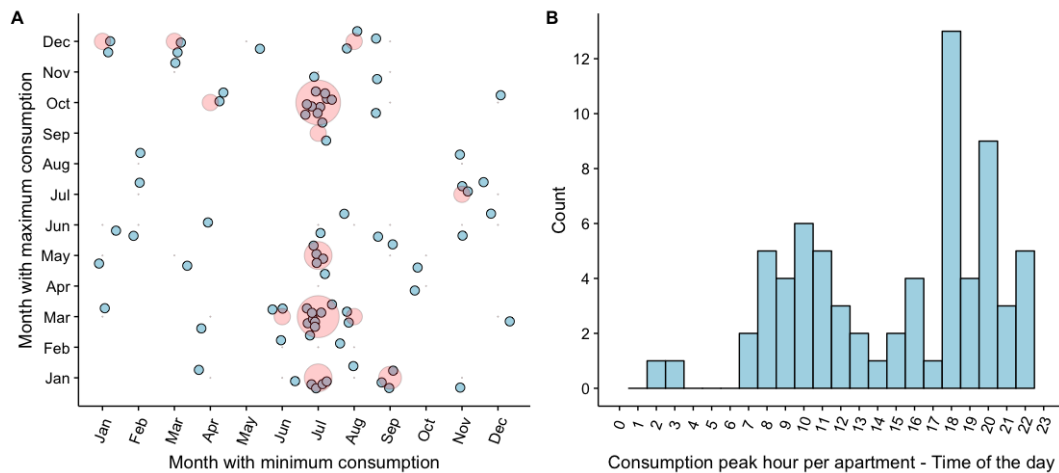


FIGURE 5.9: A. Clustering of minimum/maximum month for each apartment. B. Histogram visualizing the distribution of each apartments peak hour during the day.

Figure 5.10 shows a violin plot over the total consumption distribution per apartment and specific meter, hence the consumption distribution is segmented into both hot and cold water consumption as well as the location in either bathroom or kitchen. A violin plot is a compact display of a continuous distribution and the mirrored distribution is displayed in the same way as a box plot. The data used was collected between 2016-04-15 and 2017-03-19.

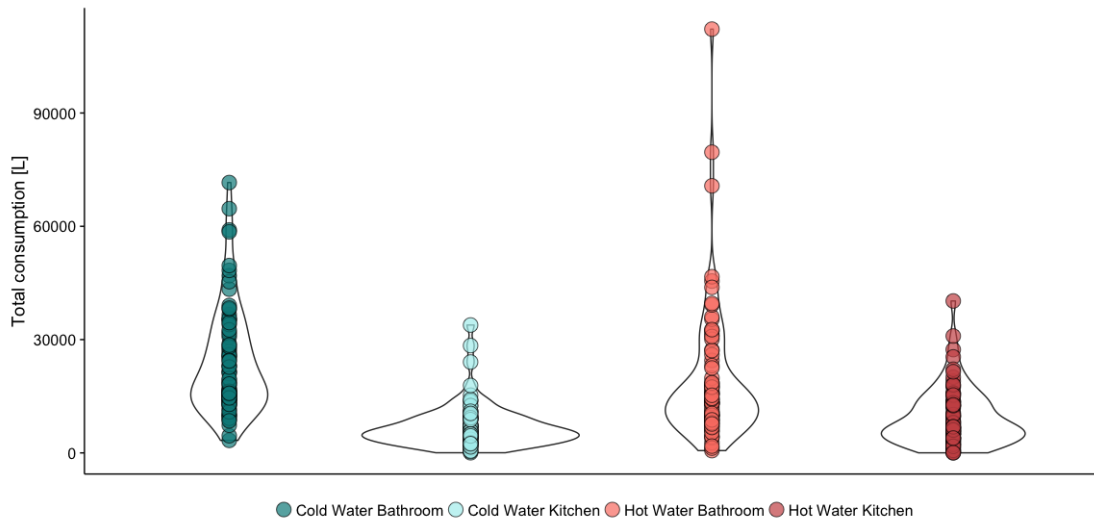


FIGURE 5.10: Total consumption per apartment segmented per meter.

Figure 5.11 shows a violin plot of the average daily consumption per apartment distributed over number of rooms within each apartment. One and two room apartments have their highest density at roughly the same level while three room apartments have a more uniform distribution. As can be seen, category two has the largest spread. The data used was collected between 2016-04-15 and 2017-03-19.

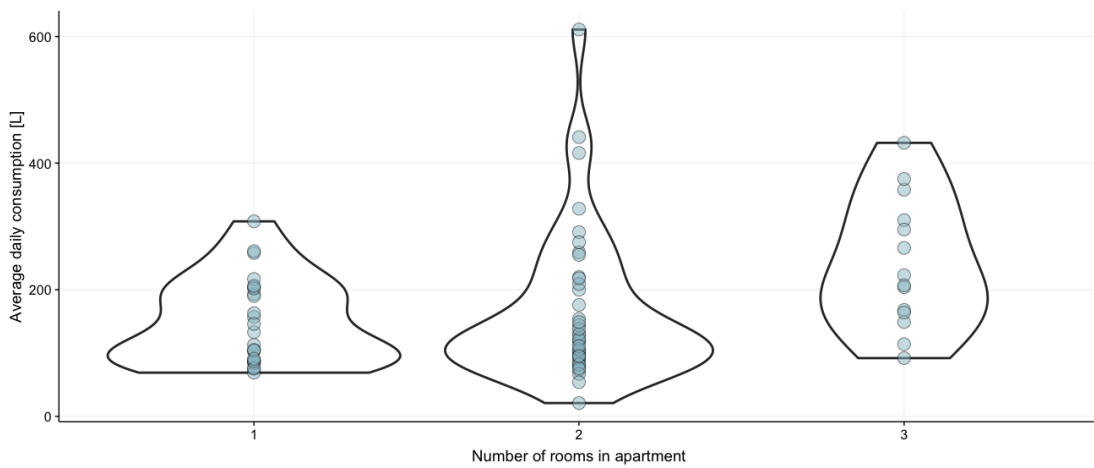


FIGURE 5.11: Violin plot of average daily consumption and number of rooms within each apartment.

## 5.5 Leak Detection

During the investigated time period (2016-10-01 to 2016-10-31) no meter registered consumption continuously for 24 hours, thus no signs of leaking fixtures in the apartments were found. As example, Fig 5.12 visualize the leak identification procedure for one apartment during 24 hours. Each square represents one hour within the specific day and for each hour the data point sent was grouped by "Consumption" or "No consumption". As seen, there is no sign of leakage since consumption hours are ranging between 1 to 4 hours for the different sensors during the day. Worth noting is that the squares representing each hour are grouped to the respectively category and hence, not visualized in a sequential hourly order.

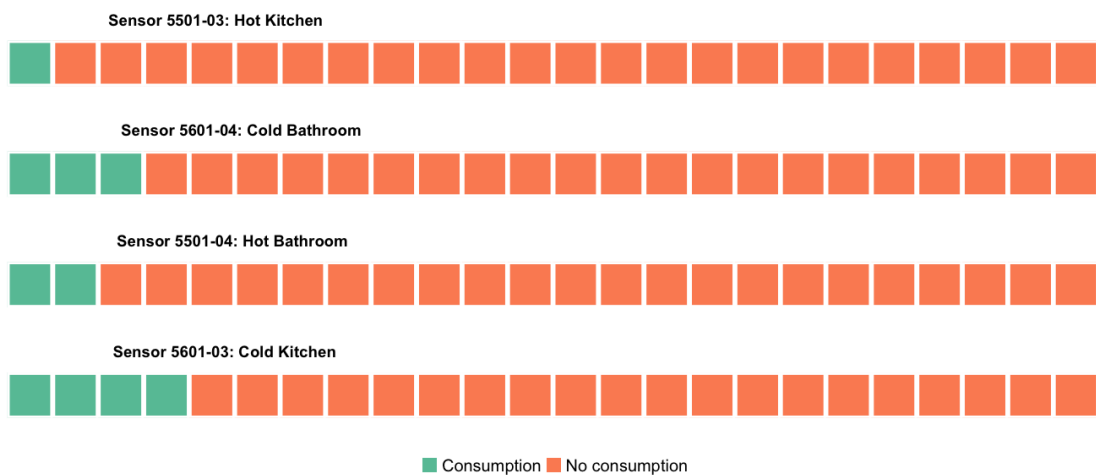


FIGURE 5.12: Visualization of the leak detection procedure, obtained at 2016-10-03.

# Chapter 6

## Discussion

*In this Chapter the results from the EDA combined with the interview result and the literature review are discussed. The Chapter is divided in three sections. Firstly, the analysis and its interpretation is discussed. Secondly, the applications of end user consumption data and the potential abilities to provide benefits are discussed. Finally, important limitations of this study and their affect on the outcome are discussed.*

### 6.1 Discussion of the Analysis

Smart water metering in conjunction with analysis of end user consumption enables access to more detailed information on household water consumption and how it is being consumed than possible before [60]. The information needs to be made accessible and understandable in order to make meaning. How to interpret the analysis is not always straight forward and this section discusses the result of the analysis and its interpretation combined with important features and findings.

The baseline results are meant as a first orientation of the data and its content gives a frame of reference to further analysis and other studies. The average daily water consumption is 174 [L/Apartment]. Noteworthy is that this number is per apartment and not per person. This value seems reasonable, since it is within the similar range of other reported average consumption values such as 140-184 [L/person and day] [47, 56]. However, no information regarding the exact number of tenants per apartment are present in this study making the numbers not directly comparable. 99 registered tenants at 79 apartments gives an average of 1.25 persons per apartment and hence, the average consumption per person would be 143 [L]. This value is lower but still in line with previous studies and estimations.

An uncertainty in this result is whether the water volume consumed for laundry is accounted for or not, since no information regarding installments of washing machines in the individual apartments are to be found. If laundry is being done in the common laundry area, a significant amount of water is not registered at an end user level and not accounted for in the data. Water for laundry is estimated to 15 [L] per person and day [47] which if added to the average consumption of 143 [L/person and day] would add up to 158 [L/person and day] of average daily consumption for the studied population. This stands in line with the average consumption of 160 [L/person and day] in Södertälje as a whole.

The analysis presents a ratio between cold and hot water consumption of around fifty-fifty, which differs from previous results by Sweden's Energy Council concluding a hot water consumption ratio of 32% [56]. Based on the observed seasonality in this study's analysis for hot and cold water, minor differences are shown in the ratio between different months. The exception is for the summer months between June and September where a higher cold water ratio is seen. If this is due to warmer weather or other parameters is hard to say and will not be discussed further. There are also major differences in the average ratio between different end users. Given the sample size in the report by Sweden's Energy Council [56] of only nine apartments and the time frame studied isolated to one month, no general conclusions can be drawn regarding the differences in hot and cold water ratios between the studies.

The average consumption distribution shows a heavily right-skewedness which indicates that the end user water consumption is not normally distributed and hence statistics assuming normal distribution should be used with care. Typical for right-skewed data is that the average value (mean) is pushed closer to the third quartile than the first quartile. Since average consumption is an often used statistic in both this study and other reports [47, 56, 57], this can cause misinterpretation of the underlying consumption behavior. To only report an average consumption even though it's recognized as well as shown in this study that the distribution is heavily right-skewed, can give the impression of overconsumption when in reality the mean is pushed up due to a few large consumers in the population. This is something that has been accounted for in this study and should be kept in mind for further studies.

Water consumption is known to vary over time. By visualizing consumption differences over time, consumption patterns and behaviors can be identified. The monthly seasonality analysis shows an even consumption over the year with a slight decrease during July, August and September where July stands out with the most noticeable decrease. However, great variations within the population for each month with several outliers above the upper fence are seen. This goes in line with previous results showing a few end



users having considerable higher water consumption than the average. In general, data from several consecutive years are needed to verify if this is a reoccurring phenomena or not in order to provide any strong evidence or indications. If speculating, the decrease during the summer could be due to summer vacations.

In the weekly variations analysis, a steady consumption without significant differences between the weekdays can be seen. The box plot shows a similar story with a large spread in distribution throughout the week. The hourly variations over a day shows a curve that has a heavy decrease in consumption during the night hours and peak consumption around noon and evening time. The spread is the largest during the morning and evening as seen in the box plots. The increase between 6:00 and 8:00 seems natural, however one could expect an even higher increase due to morning routines. The continued high consumption during the day together with low variations of consumption volume during the week might indicate that parts of the end users stay at home during the day, also on weekdays.

In the clustering analysis consumption or derivatives of it have been plotted against different variables in order to see if there are non obvious patterns in the data set. Each apartments maximum and minimum water consumption month have been plotted against each other and four clusters can be seen in an otherwise quite uniform distribution. All of the clusters have July as minimum consumption month and January, March, May or October as maximum consumption month. This confirms July as a low consumption month but the intercepts between minimum and maximum months are too spread out to say anything conclusive about the relation between them. Furthermore, a peak hour distribution similar to the hourly seasonality can be seen if looking at the spread of the individual peak hour per apartment. Differences exist tho and a high average consumption ranges between 10:00 am and 9:00 pm, even though most apartments have their peak consumption during either the morning or evening. Interestingly two apartments have their peak hour average consumption between 1:00 am and 3:00 am.

The SABO-estimation of average water consumption per year is expressed in relationship to the living area [ $m^3/m^2$ ] [57]. The linear regression analysis clearly shows that apartment size is a bad estimator for water consumption in the analyzed data set. A correlation coefficient of 0.294 (Multiple R) is very low and only 8.7% of the changes in water consumption can be explained by apartment size (R Square). This conclusion stands in line with for example a study by Chen et al. [58] concluding that water consumption is effected by a number of different variables and that household living area not always have a significant effect on water consumption in comparison to other variables such as number of household members or socio-demographic parameters. Furthermore,

the clustering of water consumption by number of rooms in the apartment confirms the regression analysis that living area have no significant correlation to water consumption.

When comparing total water consumption for the four different meters installed in each apartment, a higher consumption by the meters located in the bathroom can be seen. As analysis shows, both hot and cold water is consumed to a higher degree in the bathroom compared to the kitchen. Water consumption in the bathroom also has a larger spread than the more dense consumption spread in the kitchen, indicating that a larger behavior variation can be seen in the bathroom area. This can also indicate that the number of people living in the apartment affect bathroom consumption more than it affects water usage in the kitchen.

The leak detection method in this study was conducted through the identification of continuous flow for a meter over a longer period of time. The definition of a potential leak was set at a continuous flow for 24 hours, but if this is the most effective time interval required for detecting a leak is not accounted for and needs further investigation. A shorter time frame for an alert results in a faster potential identification but with a higher chance of a false alarm compared to a longer time period. It is also important to take into consideration how the normal consumption patterns looks like. A continuous flow during the night hours can indicate leakage but only if this differs from the normal consumption pattern.

To summarize, the EDA method has been a powerful approach generating a number of different results, each one in itself a step to a deeper understanding of the underlying data set. Each result in the process is not necessarily deemed valuable in itself but as the understanding of the structures and patterns of the underlying data increases, the ability to ask and answer more context relevant questions improve. Improved understanding of the underlying data set also reduces the risk of drawing false assumptions about the present situation and misinterpret the data. The analyses made do not provide a complete understanding but gives a rather conclusive picture of consumption and its patterns for the studied data set.

However, manually analysis is time consuming and this thesis time frame have limited the amount of analyses. Starting out without a clear hypothesis, a lot of effort and energy have been put into the analysis to unveil what possible information that could be extracted and interpreted by an EDA. As mentioned above, once knowing the objectives of an analysis or what to look for, further and more specific analyses could be made.

## 6.2 Discussion of Potential Benefits and Applications

Even more challenging than to interpret data by visualizations to make meaning is the "So What"-layer that urges action taking [4]. As Boyle et al. [5] concludes, the most relevant question is what to do with all this collected data. There is no surprise that the reaction to some of the new data flows might be "So What?". It is recognized that completely new data flows, especially on a consumer basis, is associated with confusion regarding the necessity and relevance of the data [4]. This correlates well with the interview answers by Daniel Bäcklin [20] where the major identified obstacle towards a data driven operational management is the confusion and lack of knowledge how to utilize the data.

Recent literature suggest several application areas of end user water consumption aiming to improve the buildings operational management [55, 62]. Leak identification is according to Stewart et al. [55] one of the key benefits of utilizing smart metering data and Britton et al. [54] concludes that it is a powerful tool for rapid identification, saving both water, money and time. Further on, as Ferreira et al. [62] brought up, the possibility to identify leakage by utilizing data analysis was significant compared to physically inspection. For Telge Bostäder, where physical inspection is the major part of the leak detection process, an adjustment and establishment of a leak identification process utilizing high frequency data could result in an improvement in comparison to the existing procedure. This should not be too unfamiliar since the operational manager already is using data (however at a very low resolution as a data point twice a year) at a small-scale. In the case study, a leak identification analysis was proposed and even if no leak was found during the studied time-frame, it shows how the data stored could be easily used for a simple leak detection. An alarm could automatically be sent if a sensor register unusual consumption. Further on, a work order could be generated if leakage is indicated and used as a basis for the decision to investigate the cause.

Quick correction measures when a leak occurs could pose one of the described "low hanging fruits", where a low-cost or a no-cost measure yielding great water savings and hence, also energy savings. In addition to this, if identifying a leak in real-time or near real-time instead of, in best cases, after the manually meter reading every six months, repairing costs for water damages within the apartments or building could decrease as well. However, once these "low hanging fruits" have been picked, further savings might be harder to achieve [62]. Nevertheless, continuously monitoring and analysis have the potential to ensure an efficient building operation at all times, which should be desirable by the operational manager. One could argue that having control and knowing that the building is operating efficiently by continuous monitoring contributes to an improved operational management in itself.

It is recognized that smart water meters have the potential to provide valuable information to both operational managers as well as to the consumer [5] and that it is important to ensure that there are benefits for both associated with the deployment of sensors at an apartment level. By empowering the consumers to proactively manage their consumption, potential behavioural change as well as transparency could be achieved [5, 15, 53, 60]. To use consumer feedback to encourage consumption conservation have been widely applied and evaluated in other sectors but have not been documented to the same extent in managing water consumption [53]. By providing more frequent and detailed consumption information and feedback, several studies reveal water savings, whereas others are critical arguing that there is little evidence at the time whether high frequency feedback is effective in reducing consumption in the water sector or not. Nevertheless, smart metering offers an improved possibility to supply end users with feedback that potentially could promote water savings. In the digital age, information and feedback in real-time or near-real time could easily be provided, as example through the existing individual digital page already available to Telge Bostäder's tenants. Daily, weekly and monthly consumption could be presented to the customer as well as comparisons and alerts for leaks and high consumption. To include consumption patterns, changes over time as well as social comparisons are suggested in the literature [60] and this information is already existing in the studied data set and could be obtained through data analysis.

Building on the argument of the importance of individual and personalized feedback [60], end users with different consumption patterns may benefit from different information and feedback. By identifying consumer groups or end users with different kind of consumption patterns, as done in several studies with varying methods and techniques according to the literature review, targeted educational campaigns or conservation advises could be distributed to end users with an continuous excessive water consumption thus, yielding the potential greatest water savings. By this study's analysis some outliers could easily be detected who had an continuous excessive water usage and thus, corresponding for approximately 40 % of the buildings total water consumption. Curbing their water consumption holds a potential low-cost water and energy saving.

However, since the apartments under study are rental apartments, it could be questionable if the tenants would engage with the question of water consumption if not being debited for their individual consumption as brought up by Daniel Bäcklin [20]. No previous studies on rental apartments were found by the authors in the literature reviewed and the concern seems legitimate. Even though transparency is a goal in itself and according to the 2012 Energy Efficiency Directive, consumers should be empowered to better manage consumption which includes easy and free access to data on consumption through individual metering. Even if the directive was not made mandatory in Sweden

for hot water consumption, positive effects can be associated with empowering the consumer even if it is questionable which impact it has. Furthermore, as brought up by Boyle et al. [5], high frequency data has the potential to provide an improved customer experience and customer service which is becoming more and more important.

Whether targeting marketing campaigns, finding cost-effective measures to reduce consumption or improving operational management, data is used to produce actionable insights and measures. As example, in the study by Cardell-Oliver et al. [15], smart meter time series are used as inputs and the outputs are actionable insights to support decisions. One could easily argue that the applications of data mentioned above are generating decision support driven by data analysis, i.e data driven decision making. There is a need to bridge the gap from information being available to enabling decisions and action taking. Data driven decisions could be applied for a more evident based work flow in most aspects. As example, the leak identification approach described above based on data analysis rather than physically inspection probably generates a measure whenever a leak is detected. The obvious need of action to repair the leak can be seen as a data driven decision, since the measure is initiated due to the input data. In the same way, data could be used as a decision basis for well grounded decisions such as targeted actions towards high consuming tenants. Decisions based on the real world rather than opinions would benefit management at all levels. Further on, as brought up during the interview, a better organized database, tools and measure reports are requested by the operational manager at Telge Bostäder. Automated reports such as work-orders or alarms are requested not only by Telge Bostäder but are still at a developmental stage according to literature.

Managing end user data should not be underestimated [5] and a significant investment of establishing individual metering in Södertälje have been the cost to develop the data management system. The fact that the system and underlying data is rarely used confirms the difficulties with new information flows brought up by Brynjolfsson et al. [41], stating that there is a transition time with significant challenges during adaption to a more data driven organization and it require not only monetary resources but also knowledge and motivation from the people in the organization. Generally, there is a lag time between availability of new technology and the maturation process around them. The same goes for new data flows, which unwrap conceptual and behavioural changes in how we interact with data. As suggested by Swan [4], perhaps a change in mindset is required to understand as well as applying appropriate usage of these new data flows. In that sense, the scientific community with the recent explode of written paper within IoT technology and water consumption could be seen as sign for maturation. The former confusion of what to do with the collected data is replaced with a number of application areas and revealed benefits. As seen by the interview with Telge Bostäder, there is a

change and an open mindset to understand what can be improved and accomplished with the use of data.

Overall, analysis of end consumer data enables a new and, at this time, underexplored and untapped opportunity to manage water more efficiently at a building level. In the common and overarching drive towards a smarter city, the water sector will have a significant role. Smart water meter data applications have the possibility to enable smarter building management as well as smarter services and hence a higher level of sustainability could be achieved.

### 6.3 Discussion of Limitations

When analyzing water consumption data, one must remember that water consumption is determined and influenced by a number of factors. Climate and seasonality as well as socio-economic and socio-demographic factors and characteristics are significant [58]. As presented within the delimitation of the study, both the socio-economic and socio-demographic characteristics have been neglected even though number of household members, age, income level and educational level are highly relevant factors and recognized to affect consumption volume and consumption patterns significantly. The absence of information about the tenants is the main limitation of the study and hence, further research could benefit significantly from such data in the pursuit to build a better understanding of the topic. By disclosing information about the tenants and include it in the analyses, new patterns and underlying explanations on water consumption could possibly be revealed. However, the data privacy issues when including information about the tenants as shortly mentioned previously have to be kept in mind. Nevertheless, it would be interesting to complement this study with detailed data about the tenants.

The limited time-frame of this study as well as the lack of knowledge about the tenants when accessing the anonymized data set are explanatory reasons to not include the socio-economic and socio-demographic factors. However, one could argue that these characteristics are most certainly similar and could be seen as representative for data obtained from the same geographic area. Hence, if not applying the results obtained to another geographical context, the results are relevant and reliable. The same goes for the climate factors affecting consumption.

Further on, this study does not seek to provide answers to the behavioural question of why we consume water in a certain way. Rather has this study answered to the question on how we consume water. Socio-economic and socio-demographic characteristics would provide important additional information that would increase our understanding and

provide further insights, but are however not essential when assessing how and when consumption occurs and the application areas of such information.

The sensors measuring water consumption are utilizing accumulated consumption, which in itself has several limitations compared to other types of sensors as for example pulse-sensors [5]. By data containing accumulated consumption, it is practicable impossible to pin-point the water usage to a certain end user consumption activity, as flushing the toilet or taking a shower due to aggregation [15]. For example, concurrent activities such as taking a shower while the washing machine is running, and sequential activities such as showering and breakfasting in the morning, are all aggregated into single, hourly volume readings. Therefore, several analyses and methods developed [58, 64] can not be used. A new method is needed to identify usage patterns in accumulated consumption meter data. Even if the resolution would be higher there would be no possibility to pin-point the human activities due to the unit of analysis utilizing accumulated consumption.

Lastly, as with most single case studies, results from this study are confined to this particular area and setting. No general conclusions can reliably be drawn. Results and insights can be discussed and applied in a broader context but one have to keep in mind that the presented analysis and data might not be representative for water consumption patterns in general. This is also important to have in mind when making inferences and drawing conclusions based on the interview with Telge Bostäder. The information and opinions of the operational manager is important in order to generate a more complete result and as a foundation for discussion but a more thorough investigation with more operational managers have to be done in order to draw any general conclusions about the possibilities and obstacles regarding data driven management. This study present a method how data of accumulated water consumption can be turned into information and applied for stakeholders within the studied building. It should be seen as one out of many possibilities to turn data into information and knowledge.

# Chapter 7

## Conclusion

*In this last Chapter, the main findings of the study and its contributions to the scientific community are concretized in a conclusion. Lastly, recommendations of further research are suggested.*

### 7.1 Conclusions

*RQ1: What possible information can be extracted and interpreted from a data set of end user water consumption based on an exploratory data analysis?*

EDA is a powerful analysis method when seeking a deeper understanding of the data under study and main characteristics of the data set could be unveiled and patterns of consumption found. By focusing on by whom (at an apartment level), when and how water is being consumed, this study proposes a novel method analyzing an existing data set from 79 rental apartments. No up to today known studies have utilized an EDA as a method approach. Further on, the number of apartments under study and the time frame studied are both larger and longer than most of previous studies and it is one out of few studies made in Sweden.

Information regarding the large and skewed spread for water consumption within the population as well as consumption seasonality over different time frames have been unveiled. Segmentation for hot and cold water as well as for location of use have revealed the distributions and ratios between the segmented categories of consumption. Living area have been proven to have no significant correlation to consumption and a method for leak detection have been conducted as a proof of concept.

An EDA is shown to establish a fundamental understanding of the content, structure and weaknesses of the data when the objectives of analyzing data is not fully understood.



However, once data have been transformed into information and further on to knowledge, one has the ability to ask and answer more context relevant questions with a lower risk of misinterpretation. More in-depth analyzes can be preformed when a clear objective is present.

*RQ2: How can the revealed information potentially provide benefits at a building level?* Data have the ability to assist in the building operational management as identifying leaks and excessive water consumption. Some of these diagnostic analyses are "low hanging fruits", where a low-cost or no-cost corrective measure yields great water savings and at the same time achieves an efficient control of the building.

Data could also be applied to empower the end users. By information and feedback on consumption, improved customer service as well as transparency is achieved which possibly could lead to positive behavioural change. Daily, weekly and monthly consumption could be presented to the customer as well as social comparisons and alerts for leaks or high consumption.

A goal of processing end user consumption data is to create a better decision basis. By utilizing smart meter data as input and transforming this data through different and purpose dependent analyses, generated outputs are potentially actionable insights to support decisions. Data driven decisions have recognized benefits and pose an efficient tool for improved water management.

However, there is generally a lag time between the availability of new data flows and an understanding in how we should interact with the data. The recent explosion of published papers within the research field as well as newly initiated and founded projects within the European Union poses several signs for an increased understanding how to utilize the data in order to provide benefits. The maturation enables a shift towards a more data driven water management at all levels, which is necessary in the drive towards a smarter and more sustainable city.

## 7.2 Contributions

This study has particularly contributed to the scientific community within the below stated areas;

- Knowledge contribution in how we consume water in our homes.
- A novel study utilizing real life data from rental apartments.

- Suggest that the EDA method is a powerful approach when no beforehand strong preconceptions are held and well suited for this type of exploratory studies.
- How smart meter data can potentially benefit stakeholders at a building level and improve water management.
- A good example of an integrated and interdisciplinary approach to tackle complex challenges.

### 7.3 Future Research

Building on the findings in this thesis, continued research would benefit if socio-economic and socio-demographic data is included into the analysis. This would add another dimension to the analysis in order to build an understanding on how we consume water in our homes.

Based on this thesis findings and discussion, there are several identified challenges that would need further investigation. One aspect that needs further research is the question about scalability. To reach scale, i.e to collect and analyze data from much larger number of apartments, is an important feature in order to achieve the benefits associated with smarter buildings and services as well as the transition towards smarter cities. In this study and in the previous studies reviewed, small-scale investigations and implementations are one of the common denominators. More data transfers to deeper knowledge and more research is needed in order to understand how to enable and achieve scale within smart water meter data analytics and its applications.

Another future research area is the implementation of data driven decision making and management based on smart meter data. This study shows the potential in analyzing data and the theoretical objectives of such actions. However, in order to utilize the data, an implementation is needed at an organizational level. Further research is needed to understand how the found benefits can be translated into implemented processes.

This thesis has brought up the potential benefits of smart water meters at a building level. However how to quantify the economic effects of said benefits have not been touched upon and needs further investigation. The economic aspects of such a cost-benefit analysis for potential savings and improved management are important factors to consider and the topic is up to today scarcely researched.

# References

- [1] Science Direct, 2017. URL <http://www.sciencedirect.com>. Retrieved 2 May 2017.
- [2] Sveriges geologiska undersökning. Månadens grundvattennivå, 2017. URL <http://www.sgu.se/grundvatten/grundvattennivaer/>. Retrieved 11 May 2017.
- [3] G.P. Hancke, B. de Carvalho e Silva, and G.P. Hancke Jr. The role of advanced sensing in smart cities. *Sensors*, 13(1):393–425, 2013. URL [doi:10.3390/s130100393](https://doi.org/10.3390/s130100393).
- [4] M. Swan. Sensor mania! the internet of things, wearable computing, objective metrics, and the quantified self 2.0. *Journal of Sensor and Actuator Networks*, 1(3): 217–253, November 2012. URL [doi:10.3390/jsan1030217](https://doi.org/10.3390/jsan1030217).
- [5] T. Boyle, D. Giurco, P. Mukheibir, A. Liu, C. Moy, S. White, and R. Stewart. Intelligent metering for urban water: A review. *Water*, 5(3):1052–1108, July 2013. URL [doi:10.3390/w5031052](https://doi.org/10.3390/w5031052).
- [6] R.K Yin. *Case study research, design and method*. Sage Publications Ltd, London, 2009.
- [7] Google Kartdata. Södertälje, 2017. URL <https://www.google.se/maps>. Retrieved 1 August 2017.
- [8] United Nations. World’s population increasingly urban with more than half living in urban areas, 2014. URL <http://www.un.org/en/development/desa/news/population/world-urbanization-prospects-2014.html>. Retrieved 11 May 2017.
- [9] C. Laspidou, E. Papageorgiou, K. Kokkinos, S. Sahu, A. Gupta, and L. Tassioulas. Exploring patterns in water consumption by clustering. *Procedia Engineering*, 119: 1439–1446, 2015. URL <https://doi.org/10.1016/j.proeng.2015.08.1004>.
- [10] J.M. Hernández-Muñoz, J. Bernat Verche, L. Muñoz, J.A. Galache, M. Presser, L.A. Hernández Gómez, and J. Pettersson. Smart cities at the forefront of the future internet. (6656), 2011. URL [https://link.springer.com/chapter/10.1007/978-3-642-20898-0\\_32](https://link.springer.com/chapter/10.1007/978-3-642-20898-0_32).

- [11] D. Koo, K. Piratla, and J. C. Matthews. Towards sustainable water supply: Schematic development of big data collection using internet of things (iot). *Procedia Engineering*, 118:489–497, September 2015. URL <https://doi.org/10.1016/j.proeng.2015.08.465>.
- [12] M. Mutchek and E. Williams. Moving towards sustainable and resilient smart water grids. *Challenges*, 5(1):123–137, March 2014. URL [doi:10.3390/challe5010123](https://doi.org/10.3390/challe5010123).
- [13] T. Ram Gurung, R.A. Stewart, C.D. Beal, and A.K. Sharma. Smart meter enabled informatics for economically efficient diversified water supply infrastructure planning. *Journal of Cleaner Production*, 135:1023–1033, November 2016. URL <https://doi-org.focus.lib.kth.se/10.1016/j.jclepro.2016.07.017>.
- [14] C.D. Beal and R.A. Stewart. Identifying residential water end uses underpinning peak day and peak hour demand. *Journal of Water Resources Planning and Management*, 140(7), July 2014. URL [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000357](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000357).
- [15] R. Cardell-Oliver, J. Wang, and H. Gigney. Smart meter analytics to pinpoint opportunities for reducing household water use. *Journal of Water Resources Planning and Management*, 142(6), January 2016. URL [DOI:10.1061/\(ASCE\)WR.1943-5452.0000634](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000634).
- [16] A. Cominola, M. Giuliani, D. Piga, A. Castelletti, and A.E. Rizzoli. Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling Software*, 72:198–214, October 2015. URL <https://doi.org/10.1016/j.envsoft.2015.07.012>.
- [17] J. Kihlberg. Vattenbrist hotar flera delar av landet, 2017. URL <http://www.dn.se/nyheter/sverige/vattenbrist-hotar-fler-delar-av-landet/>. Published in Dagens Nyheter 2017-03-02.
- [18] Telge Nät. Bevattningsförbud i södertälje och nykvarns kommuner, 2017. URL <http://www.telge.se/bevattningsforbud>. Retrieved 2 May 2017.
- [19] The European Parliament and the Council of European Union. Directive 2012/27/eu of the european parliament and of the council of 25 october 2012 on energy efficiency, amending directives 2009/125/ec and 2010/30/eu and repealing directives 2004/8/ec and 2006/32/ec. *Official Journal of the European Union*, (L 315/1):1–56, October 2012. URL <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012L0027>.
- [20] Daniel Bäcklin. Operational Manager at Telge Bostäder. Interview, June 12th 2017.

- [21] K. Rathnayaka, H. Malano, S. Maheepala, B. George, B. Nawarathna, M. Arora, and P. Roberts. Seasonal demand dynamics of residential water end-uses. *Water*, 7(1):202–216, January 2015. URL [doi:10.3390/w7010202](https://doi.org/10.3390/w7010202).
- [22] L.M. Irons, J. Boxall, V. Speight, B. Holden, and B. Tam. Data driven analysis of customer flow meter data. *Procedia Engineering*, 119:834–843, 2015. URL <https://doi.org/10.1016/j.proeng.2015.08.947>.
- [23] S.O Hansson. *The art of doing science*. 2007.
- [24] M. Nissani. Ten cheers for interdisciplinarity: The case for interdisciplinary knowledge and research. *The Social Science Journal*, 34(2):201–216, 1997. URL [https://doi.org/10.1016/S0362-3319\(97\)90051-3](https://doi.org/10.1016/S0362-3319(97)90051-3).
- [25] V. Castán Broto, M. Gislason, and M-H. Ehlers. Practising interdisciplinarity in the interplay between disciplines: experiences of established researchers. *Environmental Science Policy*, 12(7):922–933, November 2009. URL <https://doi.org/10.1016/j.envsci.2009.04.005>.
- [26] U. Berardi, V. Albino, and R.M. Dangelico. Smart cities: definitions, dimensions, and performance. *Journal Of Urban Technology*, 22(1):3–21, February 2015. URL <http://dx.doi.org/10.1080/10630732.2014.942092>.
- [27] A.P Lara, E.M Da Costa, T. Zilinski Furlani, and T. Yigitcanlar. Smartness that matters: towards a comprehensive and human-centred characterisation of smart cities. *Journal of Open Innovation: Technology, Market, and Complexity*, 2(8), December 2016. URL <http://dx.doi.org/10.1186/s40852-016-0035-y>.
- [28] RFID INFISO G.2 Micro & Nanosystems INFISO D.4 Networked Enterprise and RFID Working group of the European Technology Platform on smart systems integration (EPOSS). Internet of things in 2020, roadmap for the future. 2008.
- [29] K. Ashton. That 'internet of things' thing. 2009. URL <http://www.rfidjournal.com/articles/view?4986>.
- [30] K. Rose, S. Eldridge, and L. Chapin. The internet of things (iot): An overview. 2015. URL <https://www.internetsociety.org/doc/iot-overview>.
- [31] R. Frank. *Understanding Smart Sensors, Second Edition*. Artech House, Boston, 2013.
- [32] S Darby. Smart metering: what potential for householder engagement? *Building Research Information*, 38(5):442–457, 2010. URL <http://dx.doi.org.focus.lib.kth.se/10.1080/09613218.2010.492660>.

- [33] L. Atzori, A. Iera, and G. Morabito. The internet of things: A survey. *Computer Networks*, 54:2787–2805, May 2010. URL [doi:10.1016/j.comnet.2010.05.010](https://doi.org/10.1016/j.comnet.2010.05.010).
- [34] J.W. Tukey. The future of data analysis. *Annals of Mathematical Statistics*, 33(1): 1–67, March 1962. URL [doi:10.1214/aoms/1177704711](https://doi.org/10.1214/aoms/1177704711).
- [35] C.H. Yu. Exploratory data analysis in the context of data mining and resampling. *International Journal of Psychological Research*, 3(1):–, 2010. URL [dx.doi.org/10.21500/20112084.819](https://dx.doi.org/10.21500/20112084.819).
- [36] J. Han and M. Kamber. *Data mining: Concepts and Techniques*. Elsevier Inc, 2006. URL <http://www.sciencedirect.com/science/book/9780123814791>.
- [37] J.T Behrens and C.H Yu. *Handbook of psychology: Exploratory data analysis*. John Wiley Sons, Inc, 2003. URL [doi:10.1002/0471264385.wei0202](https://doi.org/10.1002/0471264385.wei0202).
- [38] J. Rowley. The wisdom hierarchy: representations of the dikw hierarchy. *Journal of Information Science*, 33(163), February 2007. URL [doi:10.1177/0165551506070706](https://doi.org/10.1177/0165551506070706).
- [39] R.L. Ackoff. From data to wisdom. *Journal of Applied Systems Analysis*, (16):3–9, 1989. URL <http://faculty.ung.edu/kmelton/Documents/DataWisdom.pdf>.
- [40] D. Chaffey and S. Wood. *Business Information Management: Improving Performance using Information Systems*. Financial Times/Prentice Hall, 2005.
- [41] E. Brynjolfsson and K. McElheran. The rapid adoption of data-driven decision-making. *American Economic Review*, 106(3):133–139, May 2016. URL [http://dx.doi.org/10.1257/aer.p20161016](https://dx.doi.org/10.1257/aer.p20161016).
- [42] A. Dresch, D. Pacheco Lacerda, and J.A.V Antunes Jr. *Design Science Research A Method for Science and Technology Advancement*. Springer International Publishing, 2015. URL [doi:10.1007/978-3-319-07374-3](https://doi.org/10.1007/978-3-319-07374-3).
- [43] L. Dubé and G. Paré. Rigor in information systems positivist case research: current practice. *MIS Quarterly*, 27(4):597–635, 2003. URL <http://www.jstor.org/stable/30036550>.
- [44] L.A Machi and B.T McEvoy. *The literature review: six steps to success (2nd ed.)*. Corwin Press, Thousand Oaks, California, 2012.
- [45] Södertälje kommun. Snabbstatistik om södertälje kommun, 2016. URL <https://www.sodertalje.se/kommun-och-politik/statistik-och-nyckeltal/snabbstatistik-sodertalje/>. Retrieved 11 May 2017.

- [46] Telge Nät. Vårt dricksvatten, 2017. URL <http://www.telge.se/vatten-avlopp/vart-dricksvatten/>. Retrieved 11 May 2017.
- [47] Svenskt Vatten. Dricksvattenfakta, 2017. URL <http://www.svensktvatten.se/fakta-om-vatten/dricksvattenfakta/>. Retrieved 11 May 2017.
- [48] W. Olsen. *Data Collection: Key Debates and Methods in Social Research*. Sage Publications Ltd, London, 2012. URL [doi:10.4135/9781473914230](https://doi.org/10.4135/9781473914230).
- [49] F.E. Grubbs. Procedures for detecting outlying observations in samples. (11):1–21, 1969. URL <http://dx.doi.org/10.1080/00401706.1969.10490657>.
- [50] J. Ahlin. Ethical thinking. 2016.
- [51] National Academy of Sciences. On being a scientist: A guide to responsible conduct in research: Third edition. 2009. URL [www.nap.edu](http://www.nap.edu).
- [52] S. Eriksson. Naturvetenskaplig och teknisk forskning, 2017. URL <http://www.codex.vr.se/forskningteknat.shtml>. Retrieved 25 April 2017.
- [53] A. L. Sønderlund, J. R. Smith, C. J. Hutton, K. Zoran, and D. Savic. Effectiveness of smart meter-based consumption feedback in curbing household water use: Knowns and unknowns. *Journal of Water Resources Planning and Management*, 142(12), December 2016. URL [https://doi-org.focus.lib.kth.se/10.1061/\(ASCE\)WR.1943-5452.0000703](https://doi-org.focus.lib.kth.se/10.1061/(ASCE)WR.1943-5452.0000703).
- [54] T. Britton, R. Stewart, and K.R. O’Halloran. Smart metering: enabler for rapid and effective post meter leakage identification and water loss management. *Journal of Cleaner Production*, 54:166–176, September 2013. URL <https://doi.org/10.1016/j.jclepro.2013.05.018>.
- [55] R.A. Stewart, R. Willis, D. Giurco, K. Panuwatwanich, and G. Capati. Web-based knowledge management system: linking smart metering to the future of urban water planning. *Australian Planner*, 47(2):66–74, 2010. URL <http://dx.doi.org.focus.lib.kth.se/10.1080/07293681003767769>.
- [56] Statens Energimyndighet. Mätning av kall- och varmvattenanvändning i 44 hushåll, delrapport i energimyndighetens projekt förbättrad energistatistik i bebyggelsen och industrin. (ER 2009:26), 2009.
- [57] Sabo Sveriges Allmännyttiga Bostadsföretag. Snabbanalys: Normalanvändning – varmvatten. 2014. URL <http://www.sabo.se/aktuellt/sabotycker/snabbanalyser/Documents/Snabbanalys/>.

- [58] X. Chena, S-H. Yang, L. Yang, and X. Chen. A benchmarking model for household water consumption based on adaptive logic networks. *Procedia Engineering*, 119: 1391–1398, 2015. URL <https://doi.org/10.1016/j.proeng.2015.08.998>.
- [59] Y. Shan, L. Yang, K. Perren, and Y. Zhang. Household water consumption: Insight from a survey in greece and poland. *Procedia Engineering*, 119:1409–1418, 2015. URL <https://doi.org/10.1016/j.proeng.2015.08.1001>.
- [60] D. Giurco A. Liu and P. Mukheibir. Urban water conservation through customised water and end-use information. *Journal of Cleaner Production*, 112(4):3164–3175, January 2016. URL <https://doi.org/10.1016/j.jclepro.2015.10.002>.
- [61] M.K Tiwari and J. Adamowski. Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resources Research*, 49:6486–6507, October 2013. URL [doi:10.1002/wrcr.20517](https://doi.org/10.1002/wrcr.20517).
- [62] V. Ferreira, P. Fleming, G. Stuart, P. Patel, P. Webber, and S. Conway. Low hanging fruits or cost effective energy and water savings using intelligent metering and monitoring systems? 2007. URL [http://plone.fred.nu/library/conference\\_proceedings/eceee\\_Summer\\_Studies/2007/Panel\\_4/4.282/paper](http://plone.fred.nu/library/conference_proceedings/eceee_Summer_Studies/2007/Panel_4/4.282/paper).
- [63] K. Anh Nguyena, H. Zhanga, and R.A Stewart. Development of an intelligent model to categorise residential water end use events. *Journal of Hydro-environment Research*, 7(3):182–201, September 2013. URL <https://doi.org/10.1016/j.jher.2013.02.004>.
- [64] J. Wang, R. Cardell-Oliver, and W. Liu. An incremental algorithm for discovering routine behaviours from smart meter data. *Knowledge-Based Systems*, 113(1):61–74, December 2016. URL <https://doi.org/10.1016/j.knosys.2016.09.016>.
- [65] K.S. Fielding, A. Spinks, S. Russell, R. McCrea, R. Stewart, and J. Gardner. An experimental test of voluntary strategies to promote urban water demand management. *Journal of Environmental Management*, 114:343–351, January 2013. URL <https://doi.org/10.1016/j.jenvman.2012.10.027>.
- [66] E. Magiera and W. Froelich. Integrated support system for efficient water usage and resources management (iss-ewatus). *Procedia Engineering*, 89:1066–1072, 2014. URL <https://doi.org/10.1016/j.proeng.2014.11.226>.
- [67] E. Magiera, W. Froelich, T. Jach, K. Kurcius Berbeka, S. Bhulai, K. Kokkinos, E. Papageorgiou, C. Laspidou, L. Yang, K. Perren, S.H. Yang, A. Capiluppi, S. El-Jamal, and Z. Wang. Iss-ewatus an example of integrated system for efficient water management. 2016. URL [http://issewatus.eu/pluginfile.php/1038/mod\\_resource/content/1/Magiera\\_Sanjai\\_Amsterdam.pdf](http://issewatus.eu/pluginfile.php/1038/mod_resource/content/1/Magiera_Sanjai_Amsterdam.pdf).



- 
- [68] W. Froelich and E. Magiera E. Iss-ewatus decision support system - overview of achievements. *Intelligent Decision Technologies 2017: Smart Innovation, Systems and Technologies*, 72:1–10, 2017. URL [https://link-springer-com.focus.lib.kth.se/chapter/10.1007/978-3-319-59421-7\\_18](https://link-springer-com.focus.lib.kth.se/chapter/10.1007/978-3-319-59421-7_18).

## Appendix A

# Appendix A

### A.1 Obtaining the Data Set Under Study

LISTING A.1: SQL-code for obtaining the complete data set

---

```
with lebanon1(complexID,apartment,meter,house) as (  
    select complexID,apartment,meter,house from complexObfuscated where house  
    IN ('Lebanon','Romania') and information LIKE ('%vatten%')  
    )  
    select  
    *  
    from lebanon1  
join meter ON lebanon1.complexID = meter.complexID  
join template ON meter.meterID = template.meterID  
join templatesetting ON templatesetting.templateId = template.templateID  
join value ON templatesetting.templatesettingId = value.templateSettingId  
join value2 ON templatesetting.templatesettingId = value2.templateSettingId  
join valuetype ON valuetype.valueTypeId = value.valueTypeId  
where apartment NOT IN (10066, 10979, 11001, 11274, 12005) and value.date <  
    2017-03-20'
```

---

### A.2 Obtaining the Baseline Values

LISTING A.2: SQL-code for obtaining the data used for baseline values

---

```
with lebanon1(complexID,apartment,house,meter,size) as (  
    select complexID,apartment,house,meter,size from complexObfuscated where  
    house IN ('Lebanon','Romania') and information LIKE ('%vatten%')  
    )  
    select  
    apartment,  
    lebanon1.meter,  
    min(value) as 'min',  
    max(value) as 'max'
```

---

```

        from lebanon1
    join meter ON lebanon1.complexID = meter.complexID
    join template ON meter.meterID = template.meterID
    join templatesetting ON templatesetting.templateId = template.templateId
    join value ON templatesetting.templatesettingId = value.templateSettingId
    join valuetype ON valuetype.valueTypeId = value.valueTypeId
    where apartment NOT IN (10066, 10979, 11001, 11274, 12005) and date <'2017-03-20'
    group by apartment,lebanon1.meter order by apartment ASC

```

---

### A.3 Residual Analysis

---

LISTING A.3: SQL-code for obtaining data for residual analysis

---

```

with lebanon1(complexID,apartment,house,meter,size) as (
    select complexID,apartment,house,meter,size from complexObfuscated where
    house IN ('Lebanon','Romania') and information LIKE ('%vatten%')
)
select
    apartment,
    lebanon1.complexID,
    lebanon1.house,
    lebanon1.size,
    lebanon1.meter,
    min(value),
    max(value)
from lebanon1
join meter ON lebanon1.complexID = meter.complexID
join template ON meter.meterID = template.meterID
join templatesetting ON templatesetting.templateId = template.templateId
join value ON templatesetting.templatesettingId = value.templateSettingId
join valuetype ON valuetype.valueTypeId = value.valueTypeId
where apartment NOT IN (10066, 10979, 11001, 11274, 12005) and date <'2017-03-20'
group by lebanon1.complexID,apartment,lebanon1.house,lebanon1.size,lebanon1.meter
order by apartment ASC

```

---

### A.4 Seasonality

---

LISTING A.4: SQL-code for obtaining the data for seasonality analysis per month

---

```

with lebanon1(complexID,apartment,meter) as (
    select complexID,apartment,meter from complexObfuscated where house IN ('
    Lebanon','Romania') and information LIKE ('%vatten%')
)
select
    -- lebanon1.complexID,
    lebanon1.apartment,
    CAST(date as date),
    month,
    day,

```

---

```

        hour,
        SUM(CASE WHEN lebanon1.meter IN('5601-4','5601-3') THEN value ELSE 0 END)
as coldWater,
        SUM(CASE WHEN lebanon1.meter IN('5501-3','5501-4') THEN value ELSE 0 END)
as warmWater,
        SUM(value) as total
        from lebanon1
join meter ON lebanon1.complexID = meter.complexID
join template ON meter.meterID = template.meterID
join templatesetting ON templatesetting.templateId = template.templateID
join value ON templatesetting.templatesettingId = value.templateSettingId
join valuetype ON valuetype.valueTypeId = value.valueTypeId
where apartment NOT IN (10066, 10979, 11001, 11274, 12005) and hour = '0' and day
='1'
group by lebanon1.apartment,CAST(date as date), month,day,hour
order by CAST(date as date), lebanon1.apartment

```

---

LISTING A.5: SQL-code for obtaining the data used in seasonality analysis per hour  
and day

---

```

with lebanon1(complexID,apartment,house,meter,size) as (
        select complexID,apartment,house,meter,size from complexObfuscated where
        house IN ('Lebanon','Romania') and information LIKE ('%vatten%')
        )
        select
        apartment,
        date,
        hour,
        SUM(value)- LAG (SUM(value),1) over (partition by apartment order by date
        )
        from lebanon1
join meter ON lebanon1.complexID = meter.complexID
join template ON meter.meterID = template.meterID
join templatesetting ON templatesetting.templateId = template.templateID
join value ON templatesetting.templatesettingId = value.templateSettingId
join valuetype ON valuetype.valueTypeId = value.valueTypeId
where apartment NOT IN (10066, 10979, 11001, 11274, 12005) and date < '2017-03-20
,
group by apartment,date,hour order by apartment,date ASC

```

---

## A.5 Leak Detection

Values, meter ID and date for each sensor and hour were extracted from the database. The data was imported to Microsoft Excel and the hourly consumption was calculated by taking the difference in accumulated values per hour for each sensor. The hours for each sensor was given a binary status, 1 or 0, 1 for hours with registered consumption and 0 otherwise. A function then summarized the binary values over a 24 hour rolling window to identify if any sensor had been transmitting "Consumption" continuously for 24 hours, thus indicating a possible leak.

## LISTING A.6: SQL-code for obtaining the data used for leak detection

---

```
with lebanon1(complexID,apartment,house,meter,size) as (  
    select complexID,apartment,house,meter,size from complexObfuscated where  
    house IN ('Lebanon','Romania') and information LIKE ('%vatten%')  
    )  
    select  
    apartment,  
    lebanon1.complexID,  
    lebanon1.meter,  
    value,  
    date  
    from lebanon1  
join meter ON lebanon1.complexID = meter.complexID  
join template ON meter.meterID = template.meterID  
join templatesetting ON templatesetting.templateId = template.templateID  
join value ON templatesetting.templatesettingId = value.templateSettingId  
join valuetype ON valuetype.valueTypeId = value.valueTypeId  
where apartment NOT IN (10066, 10979, 11001, 11274, 12005) and date between '  
    2016-10-01' and '2016-10-31'  
order by lebanon1.complexID,date ASC
```

---

## Appendix B

# Appendix B

### B.1 The Interview Guide

This semi-structured interview with Telge Bostäder (TB) was a part of the data acquisition for the case study. General information about the interview and its set up can be seen in Table B.1.

TABLE B.1: General interview information

Interview Date	2017-06-12
Interview Location	Storgatan 42, Södertälje
Participant	Daniel Bäcklin (Operational manager at Telge Bostäder)
Contact Address	Daniel.Backlin@telge.se
Time used	2 hours
Original language	Swedish

The participant had been informed about the context of the interview beforehand, but had not received the questions. The interview was voluntary and could have been interrupted at any time during the interview. During the interview, we were engaging with two main topics:

- The current situation; Gathering and usage of data
- Future prospects of data driven operational water management

The questions were of open type and the beforehand written questions were followed, but flexibility to ask follow up question from brought up answers or thoughts in the conversation was allowed.

Firstly, introductory questions and some general background questions according to Table B.2 were asked for the general context.

TABLE B.2: Introduction and important background information questions

Have you received enough information to be able to participate and give your informed consent?
Are we allowed to publish your name and use you as a reference in our thesis?
What is your position at TB?
What are your responsibilities at TB?
For how long have you been working at TB?
Can you tell us a little bit about TB?
Do TB actively work with sustainability and environmental issues? Which and how?
Are there any identified and common challenges regarding water consumption?
What are your thoughts on the irrigation prohibition in Södertälje?
Do TB have any strategies for managing end user water consumption?

Secondly, questions in Table B.3 were asked to build an understanding of the data gathering process, usage of data and the general interest in the end user water consumption.

TABLE B.3: Questions regarding gathering and usage of data

What interest does TB have in questions regarding water consumption?
Do you handle questions regarding water consumption within building operational management? How?
When planning and developing construction, are aspects regarding water consumption brought up?
Are there any strategies for managing water consumption at a building level?
What are used as a decision basis for questions and issues regarding water consumption?
What interest do you have of how the end user and how they consume water?
How do you monitor and measure water consumption? At which level?
Why did you start to install sensors?
What drivers existed?
What was the goal?
How did you chose properties?
How did you chose technical parameters such as resolution?
What is the status of the project?
Do you have any rough numbers of the costs?
Is it correctly understood that TB owns the data?
Where is the data saved?
Who has access to it?
Do you use the data gathered?
For which purposes? Any examples? If no, why?
What are useful and relevant information to you?



Thirdly, questions regarding the future prospects and potential in using data were asked according to Table B.4.

TABLE B.4: Questions regarding future prospects of data driven operational management

How does the digital trend affect TB?
Have you seen a shift within the sector due to digitization?
Are your long term strategic plans including digitization?
What resources would you need to digitize the management?
Do you think TB will take that path?
Do you have any suggestions how the data gathered could be applied?
Do you think that your daily work could be improved by data?
Do you think someone else could benefit from this data?
Is there any information about the end user that you do not have today that would have been valuable to you in your work?
What are your thoughts on data driven operational management?
What are the main obstacles, as you see it, towards a data driven management?
Are there any decisions that could be improved by data?