University Institute of Lisbon

Department of Information Science and Technology

Waste Collection in Smart Cities: The Frequency Capacity Problem

André Oliveira

A Dissertation presented in partial fulfillment of the Requirements
for the Degree of
Master in Computer Science

Supervisor
João Carlos Amaro Ferreira, Assistant Professor, Ph.D.
ISCTE-IUL

October 2019

Resumo

Esta dissertação tem como objetivo o desenho e implementação de algoritmos para a resolução de um problema de otimização de capacidade de contentores versus frequência de recolha de lixo, tirando partido da informação do volume de resíduos nos contentores obtida através de sensores.

Inicialmente será feito o estado da arte relativo a cidades inteligentes através da utilização de sensores e as áreas onde a *internet das coisas* é melhor aplicada. Serão revistos problemas de otimização de recolha de lixo estudados na literatura.

Posteriormente, através de dados fornecidos por sensores que medem o volume ocupado em cada contentor, é feita uma análise estatística sobre os mesmos de forma a perceber quais os contentores críticos (cuja capacidade se torna nula com alguma frequência) e quais os contentores secundários (cuja frequência de recolha poderia ser diminuída).

Em seguida são criados algoritmos para a resolução do problema de otimização de capacidade de contentores versus frequência de recolha de lixo, gerados novos modelos de capacidade frequência e são apresentados resultados sobre os mesmos.

Por último, é feita a validação dos modelos gerados através de modelos de previsão e retiradas conclusões sob a performance de cada um dos algoritmos apresentados.

Palavras-chave: Análise de Dados, Otimização, Internet das Coisas.

Abstract

This dissertation aims the design and implementation of algorithms for solving an optimization problem of container capacity versus collection frequency, taking advantage of the information about the waste volume in containers obtained through sensors.

Initially, state of the art will be made relative to intelligent cities through the use of sensors and the areas where *internet of things* is best applied. Waste collection optimization problems studied in the literature will be reviewed.

Subsequently, through data provided by sensors that measure the volume occupied in each container, a statistical analysis is done on the same in order to understand which critical containers (whose capacity becomes null with some frequency) and which secondary containers (whose frequency of collection could be reduced).

Next, algorithms are created to solve the problem of container capacity versus collection frequency, new models of frequency capacity are generated and results are present.

Finally, the models generated are validated through prediction models and conclusions are drawn about the performance of each of the presented algorithms.

Keywords: Data Analytics, Optimization, Internet of Things.

Acknowledgements

I would like to thank my family for all their support. Even with all my unavailability in these passed two years, they were always there for me.

I would like to thank my supervisor, for always reaching out to me even when I disappeared for days, for its quick reviews and for all the help delivering the article associated with this dissertation.

I would like to thank my colleagues Pedro Camacho, Ruben Ribeiro and Nuno Mendonça. Pedro for always pushing me to do better and go further. Ruben for understanding the pains of working and studying at the same time. Nuno for never letting me give up and bringing joy until the end.

I would like to thank my friends that are still there even after two years of oblivion.

And I would specially like to thank my girlfriend Inês Loureiro for all the patience and support during this troubled year. She never stopped believing me for a second, something I'll always remember.

Contents

List of Figures

List of Algorithms

List of Tables

Chapter 1

Introduction

In smart cities, the use of technology is common to optimize several services provided by the city council [23]. One of the areas where technology can be used is in waste collection. With the addition of sensors in containers, with the ability to measure the volume of waste in it each time the container is opened, it's possible to know the volume of waste in every container of the city. In Portugal this is already used in cities like Castelo Branco, but all the data generated by the sensors are typically used for routing optimization only. Problems like frequency-capacity optimization with a fixed frequency of waste collection or the correlation of waste data with other datasets are not typically addressed.

The frequency-capacity optimization problem consists in, given a set of containers in a city and its historical volume data, find the best frequency of waste collection and container volume adjustments so that no container is ever full and the overall cost is minimum. Or, in a more generic form, find a set of fixed frequencies and capacity adjustment so that the overall cost is minimum.

This dissertation aims to explore the data generated by the sensors and the correlation of that data with other data sets like weekday, events or atmospheric conditions. It also aims the design and implementation of an algorithm-based analysis to solve the problem of container frequency-capacity optimization. In order to achieve this, real data was analyzed on the volume of containers over time in Castelo Branco between 2017 and 2018. In the next sections, motivation and context about this subject and research questions are addressed. Next, the objectives of the dissertation are defined, as well as the research methodology to achieve them.

1.1 Motivation

The study on how technology can help cities to provide better services to their citizens is a great motivation itself and one of the reasons why computer engineering is so interesting and at the same time important.

The problem of finding the best frequency of waste collection and container volume adjustments so that no container is ever full and the overall cost is minimum, has not yet been addressed in the literature, which increases the interest of the topic addressed in this dissertation. The results obtained can be used to save resources and reduce costs to the city councils that decide to implement the algorithms under study. The correlation between waste volume data sets can also provide interesting information about the citizens habits. Moreover, the application of the knowledge obtained during this master directly to a real problem, makes this dissertation of interest to those involved.

1.2 Context

The technological advances that have taken place since the beginning of the 2000s allowed more and more the use of technology for the optimization of resources [21]. The necessity of resource management is one of the reasons for the creation of the smart city and sustainable cities concept.

The smart city concept was first defined in 1994 and its definition has evolved over time [12]. It can be viewed as a city that makes investments in human and social capital and modern communication infrastructure, fuel sustainable economic growth and high quality of life, with a wise management of natural resources, through participatory governance [10]. A sustainable city is a city in which its conditions of production do not destroy the conditions of its reproduction, over time [11].

Smart cities use the Internet of Things technology, like sensors, to collect data for servers in the Cloud and then treat that data using Data Mining, Big Data and Machine Learning algorithms. The treatment of data has numerous applications [16] such as optimization of the water distribution system, optimization of the electrical distribution system, intelligent construction, bridge monitoring, seismic monitoring, etc. One of the areas where technology can be of great use is in the optimization of the waste collection system.

1.3 Research Questions

The study of the problem of the waste collection can go beyond the creation of algorithms for the optimization of routes. With the use of sensor data in containers to measure the volume of the contents in them, it is intend in this dissertation to answer the following questions:

1. Is it possible, with the volume data of waste in the containers, to figure out which containers require more or less frequent collection (or the addition or removal of a container)?

2. Can this information be used to optimize waste collection routes? If so, can it be done in real-time?

3. Is it possible, given the history of a container, to predict the volume of waste deposited by the year, events or atmospheric conditions?

4. Can we relate this data with information on population density or families income?

5. By setting a collection frequency, is it possible to estimate the required capacity (in containers) by geographical area?

6. Given the past volume data of a container, is it possible to define the optimal collection frequency and required capacity by geographical area?

7. If we consider clusters of containers next to each other, can we improve the solution of collection frequency and required capacity by geographical area?

To answer question 1, exploitation of the data sets and initial analysis has to be made. Questions 2 and 3 are answered by literature review and with the application of new algorithms in the data sets. Question 4 are also reviewed in the literature and a statistical study is made, to find correlation between different data sets. Focusing more on the frequency-capacity optimization problem, answering questions 5, 6 and 7 are the main objective of this dissertation.

1.4 Goals

The main goal of this dissertation is to answer the research questions and provide new algorithms to solve optimization problems regarding waste collection in smart cities. It is expected to obtain information about the capacity of containers given a fixed frequency and depending on some factors like population density, time of the year or events to happen. Furthermore, it is intended for the exploration of data and the validation of all algorithms, by a comparison between them and algorithms presented in the literature.

1.5 Research Methodology

In the previous sections, a first definition of the problem was addressed and main goals were defined. In the next chapter, a literature review is done relative to smart cities using sensors and the areas where internet of things is best applied. Related work about waste collection optimization, including routing problems studied in the literature are to be addressed. This allows the identification of the problem and a justification of the value of the solution.

Next, a definition of the objectives for a solution will be presented. The solution shall consist of a set of algorithms that can provide valuable information from the data sets and solve the problems addressed.

Subsequently, through data provided by sensors that measure the volume occupied in each container, a statistical analysis is done and algorithms are applied to realize which critical containers, whose capacity becomes full with some frequency, and which secondary containers, whose frequency of collection could be reduced. This will lead to the design and development of models for solving the frequency-capacity optimization problem and a demonstration of the results on them.

To measure the quality of the algorithms created, it will be used only 80% of the data to estimate the capacity needed in the future given a frequency and then using the result to see how it would perform on the other 20% of the data. This can also be done to measure predictions about waste collection needs by population density or time of the year. The analysis of the results and the solutions presented, with potential cost reduction, will allow us to take advantage of the study.

Finally, a communication of the problem and its importance will be made with the writing of this dissertation. It is expected that the results compose an interesting new approach to the waste collection problem in smart cities.

Chapter 2

Related Work

The role of technology in smart cities contributes to the contextualization of the theme addressed in this dissertation, so we start by reviewing the concept of smart city and its main objectives. Then we move to the main papers about waste collection optimization and its different approaches, including routing problems and frameworks proposed in literature.

2.1 Smart Cities

Smart cities are a concept that covers several purposes. One complete definition of smart city is given in [15], that says that a smart city is "*a city that monitors and integrates conditions of all of its critical infrastructures, including roads, bridges, tunnels, rails, subways, airports, seaports, communications, water, power, even major buildings, can better optimize its resources, plan its preventive maintenance activities, and monitor security aspects while maximizing services to its citizens*". Although this definition is from 1999, it's still accurate in it's objectives, the only difference is that more technology is used to help to achieve them. The evolution of its definition can be viewed in [12]. Initially, in 1993, the concept of Digital City was created and until 2010, that was the most common term for Smart City. After 2011, the concept of Smart City has grown exponentially and it's now the most common term in literature. To study the several advances in smart cities, [16] looks at the top areas where technology is used to make a city a smart city, like Smart Infrastructures, Smart Surveillance, Smart Electricity and Water Distribution, Smart Buildings, Smart Healthcare, Smart Services and Smart Transportation. To distinguish these Smart areas of the classic ones, they point out the advantages and disadvantages of using advanced sensing on them, which can be viewed in Table 2.1.

	Without Advanced Sensing	**With Advanced Sensing**
Structural Health Monitoring	High costs due to the number of personnel required for scheduled inspections	Autonomous monitoring system reduces costs of scheduled inspections and provides continuous monitoring
	Visual Inspection is not always effective	Enables a more accurate analysis of the structure's state than visual inspection
Water Distribution	High costs in disasters caused by missed or late leak detections	Enables mitigation of costs caused by possible accidents due to late leak detections
		Monitoring the quality of water ensures that the water is safe for human consumption
Electricity Distribution	Inaccurate metering and demand prediction	Advanced sensing enables more accurate metering and demand prediction
Smart Buildings	High electricity and water consumption	Reduction in water and electricity consumption due to HVAC and light control
Intelligent Transportation	Inefficient traffic control schemes causing traffic jams	Improved traffic control schemes adaptive to traffic conditions
Surveillance	Need for a human operator who is prone to distraction	Intelligent detection of abnormal situations without the need of an operator
Environmental Monitoring	Hazardous conditions, like the presence of dangerous gases, maybe detected too late	Continuous environmental gas sensing ensures that hazardous conditions can be detected timely

Table 2.1: Impact of advanced sensing in smart cities [16]

It's common to see the concepts of smart cities and sustainable cities mixed in literature, an accurate study on their differences can be viewed in [4], where is explored to what extent smart cities address the same issues that sustainable cities. It is concluded that some of the smart cities frameworks studied lack of environmental indicators, which leads to less sustainability than expected. Smart cities tend to focus more on Economy, Education Culture Science and Innovation, Information and Communication Technology and Governance and Citizen Engagement. Sustainable cities focus more on Natural and Built Environment, Water Management, Transport, Energy and Waste Management.

Going deeply through the role of Internet of Things (IoT) and the current solutions available, we can see a study of different technologies and protocols in [26] that are close to being standardized. They start by referring the huge role that IoT can provide in smart cities.

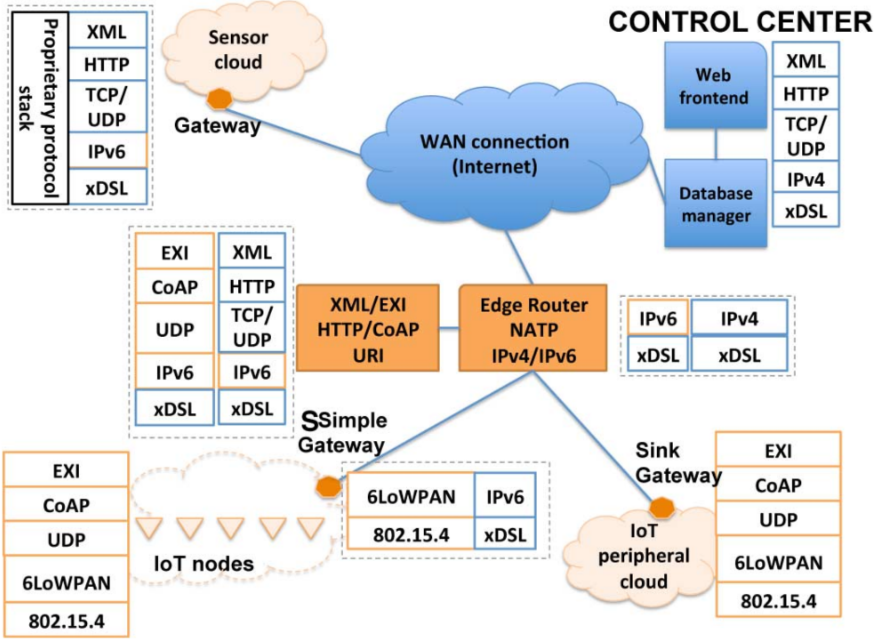


Figure 2.1: Conceptual Representation of an urban IoT network [26]

With IoT it's possible to enable easy access and interaction with home appliances, monitoring sensors, vehicles, surveillance cameras, etc. A conceptual representation of an urban IoT network framework is presented and an experimental study for a Smart City is made.

Smart cities are a huge topic and according to Pike's Research[1], it is expected that the smart cities technology market to top $20 billion in annual value by 2020. The use of technology in waste collection is just one example of how smart cities and IoT are continually changing our life for the better.

2.2 Waste Collection Problem

Waste collection management is a service that every city provides to its residents [17]. As cities grow in size and number of citizens, optimizing this service and reducing its costs is imperative. Many articles can be found on this subject, either on routing optimization or new architectures to include sensors and help the waste collection process. Some of the most important are reviewed in the following sections.

2.2.1 Optimizing Routes

Most of what has been studied about the waste collection are focused on routing problems. It's possible to associate the waste collection routing problem with the generic Traveling Salesman Problem (TSP) or Vehicle Routing Problem (VRP)[22]. The TSP consists in, given a set of n cities and the distance between them, find the best path for a Salesman to visit all the cities once and only once and return to the initial city. In the VRP, instead of one salesman or vehicle, we have m vehicles to visit n cities. In waste collection optimization, the containers represent the cities and the garbage trucks represent the vehicles. To limit the waste collection schedule, it can be added time windows restrictions to this problem [22]. Despite their simple statement, both these problems are too complex to solve obtaining the optimal solution when the number of containers is large, so it's typical to see heuristic approaches to obtain good solutions in less time [7].

Several articles study this problem, proposing algorithms for the calculation of good routes using optimization and/or machine learning. In [13], a mathematical formulation of the problem is presented and several papers in the literature are classified by the type of algorithms proposed. In [18], a genetic algorithm (GA) is presented for the identification of optimal routes for Municipal Solid Waste collection, supported by a geographic information system. Good solutions were achieved but for a small and simplified waste collection routing problem. In practice, the authors reduced the problem to the TSP and applied GA to it's resolution.

In [5], the proposed algorithms differ from the previous ones in the literature because they are dynamic algorithms and at the same time robust, being prepared for the recalculation of the routes in the event of any failure or of a collection truck reaching the limit of capacity. In their system, they take into account the trips to dumps when the storage capacity of the trucks is full and consider two types of trucks: low capacity trucks and high capacity trucks. The low capacity trucks transport waste from bins to depots. The high capacity trucks transport waste from depots to dumps outside the cities. This way, the amount of trips from bins to dumps is reduced.

FIGURE 2.2: IoT-enabled System Model Overview [5]

Some papers focus on optimizing time and costs of waste collection in particular cities, like Xangai (Pudong area) [20] or Allahabad [24], proposing municipal solid waste management systems suitable for those particular places. [9] summarizes similar papers for the United Kingdom.

Focused on the logistics involved in waste collection and recycling in several European cities, [8] carries out a detailed study on how to manage waste collection and what standards are imposed by the European Union.

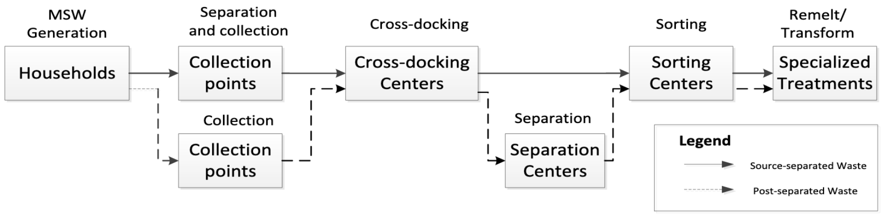


FIGURE 2.3: The recycling network flow of municipal solid waste [8]

One interesting topic about this study is the recycling network process of municipal solid waste which includes the process of collection, separation, sorting and re-processing. It provides a set of current information about the waste management problem as well as what is expected in its resolution.

2.2.2 System Architectures

More focused on cloud technologies, the article [19] presents a whole system for the collection of waste in smart cities, proposing different solutions for different stakeholders in the city. To collect data, the authors use not only the sensors but also the surveillance system of a city and it addresses several possible problems in the collection of waste in the containers, like inaccessible waste bins. The result

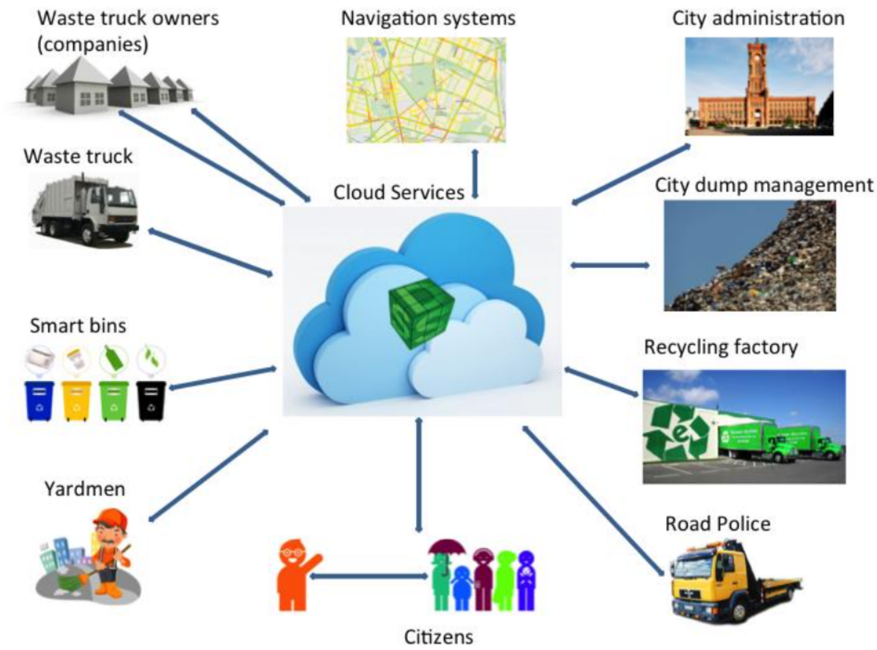


FIGURE 2.4: The big picture of a waste collection management system [19]

was a cloud-based system for waste collection in smart cities with a set of web applications for the different stakeholders.

Similarly, [14] used sensors that can read, collect and transmit trash volume and used this data to calculate new routes in real-time, guaranteeing that when trashcans become full, they are collected on the same day. However, by doing that, they increased the waste collection frequency too much, incrementing the daily collection cost between $13 - 25\%$.

In [25], the authors focused on forecasting quantity and variance of solid waste and its correlation with other sets of data, like residential population, consumer index and season, in Shanghai. The work [6] proposes a new architecture for the dynamic scheduling of waste collection considering the capacity of the same using sensors for their measurement. This is one of the complete articles in the use of measurements of capacities of the containers for the calculation of the frequency of garbage collection and the calculation of routes in real-time taking these data into account.

2.2.3 Frequency Capacity Problem

Even though there's a large number of articles dedicated to routing optimization, it can't be found in literature a study about the frequency-capacity optimization with a fixed frequency of waste collection. This can be modeled by a generic optimization problem where we want to find the minimum number of containers needed by geographic area that guarantees sufficient capacity (or maximize that capacity) with the constrain of the collection frequency. It can also be viewed as a multi-objective optimization problem where we want to minimize the total number of containers while maximize the capacity by geographic area.

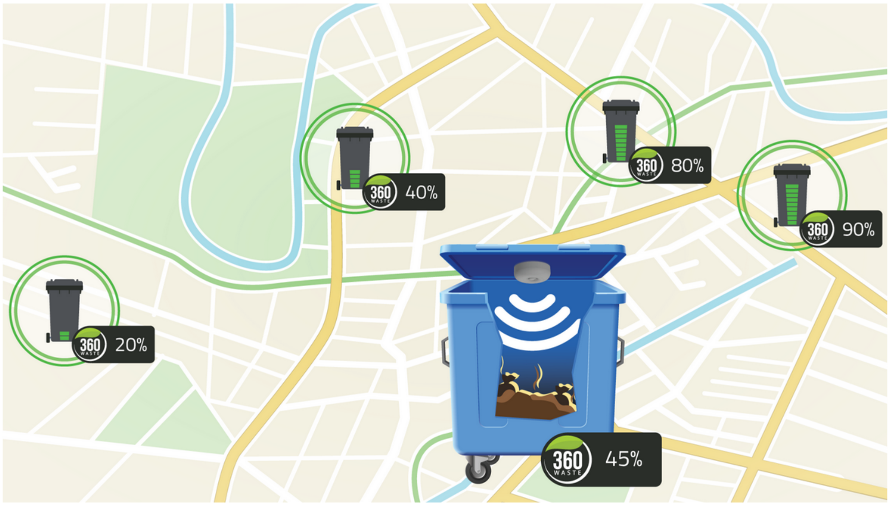


FIGURE 2.5: Waste Capacity Measured in Real-time [3]

In this dissertation we define the frequency-capacity optimization problem and propose new algorithms to its resolution. We use other sets of data to take advantage of the information provided by the sensors, like in [25] and use the information on the waste volumes over time like in [6] and [14]. The improvement we expect to obtain over [14] is that we use the volume data to change not only the frequency of collection but the capacity of the containers too. The goal is to propose models that not only prevent containers from becoming full but provide interesting solutions that may reduce waste collection costs.

Chapter 3

Data Exploration

In this chapter, we present the data sets used for this problem and data exploration using those data sets. First, an analysis of the data sets for the study and tools is made. Next, data exploration and organization are done to define what is the potential of these data sets and what problems we expect to solve with them. Lastly, data visualization and some conclusions about its study are presented.

3.1 Data Sets and Tools

To study the problem of capacity optimization given a fixed frequency, we start by analyzing sets of data of containers volume in time. Real data from the Portuguese company Evox relative to containers of the city Castelo Branco are to be used. The containers have each a sensor that measures the volume of waste in it, each time the container is opened. The company already uses this data for container volume control, waste collection, container washing and mostly for routing optimization. Based on a pre-defined filled volume a collection route is defined, like the example provided in Figure 3.1.

Our approach is the data analysis to identify deposition and collection patterns, correlate with special events and weather conditions to identify what container

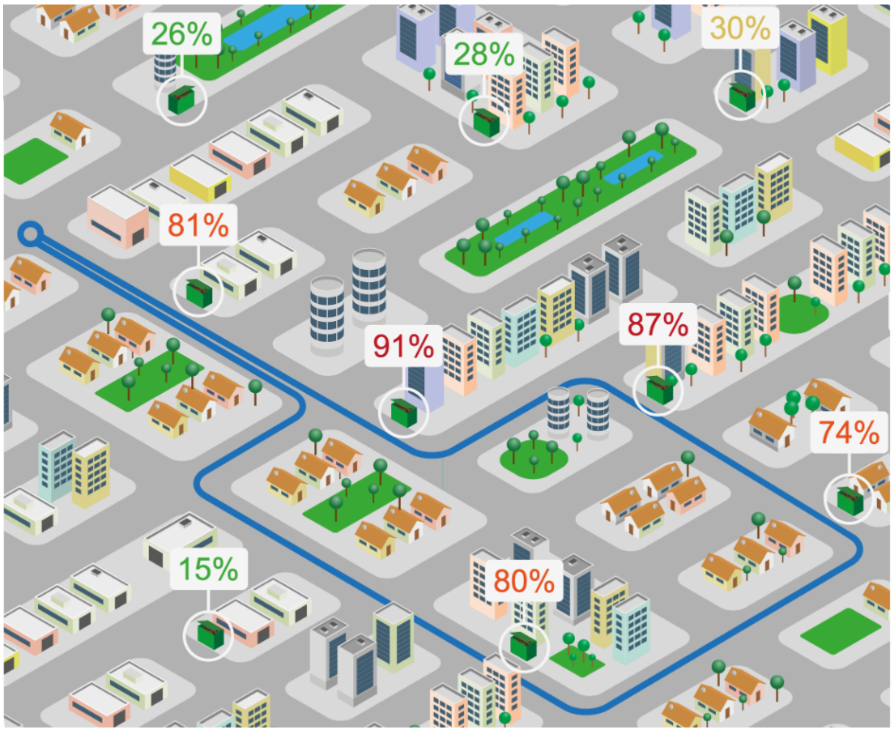


FIGURE 3.1: Route optimization based on container capacity [2]

capacity should be installed, for a weekly uniform waste collection. To study this problem, we start by analyzing sets of data of containers volume in time.

The data from each container consists on the following elements: **container id**, description, container type, waste type, **geographic localization**, address, localization zone and **sets of reading date and time and respective volume**. Table 3.1 shows an example of those elements. This represents the core data of the container and also data about the volume reading.

Field	**Example**
Container Id	15415
Description	Container 611
Container Type	Four weel with 1000 litres
Waste Type	Solid Urban Waste
Geo. Localization	39.826069 / -7.493849
Address	R. do Arco do Bispo 21
Localization Zone	Castle zone
Reading Date and time	08/06/2018 12:04; 08/06/2018 17:21;...
Volume	59; 83; ...

TABLE 3.1: Data set an example

This data must be cleaned and organized in appropriate structures to begin their mining. To do so, we started by defining the tool/framework that is to be used. There are several tools that can be used when it comes to doing data science work, like:

- **Apache Hadoop:** Apache Hadoop is an open-source software for distributed and scalable computing. It is one of the most used tools to solve big data problems and contains Hadoop Distributed File System (HDFS), Hadoop MapReduce modules and Hadoop YARN;
- **Python:** It is one of the most popular programming languages for data science. Provides a simple but complete set of tools to explore data and it emphasizes productivity and code readability;
- **R:** The R project is a programming language initially used by statisticians. R focuses on user-friendly data analysis, statistics and graphical models.

To explore the data available, we decided to work with the R project because of its simplicity when working with data sets. Besides, its a powerful tool for statistical studies which can be useful for the study of the correlation between different data sets. Having the framework decided, we then proceed to the exploitation of data.

We start by reading the data from the files into sets of data frames, containing all the information about the containers and their waste volume readings through time. A ***container data frame*** is created containing all generic data about the containers, excluding the reading date and time and respective volume. This data frame provides, for each container, information about its type and location, to be of use later.

Then considering the set C of the containers in the study, the volumes readings v_t^i, for $i \in C$, can be worked as they can provide more information that the volume itself. For that, let us consider the following definition:

Definition 3.1. Given a volume reading of a container i, v_t^i, measuring the waste volume within the container at t, when compared with the container previous reading v_{t-1}^i, is considered a:

1. **Deposit** if $v_t^i > v_{t-1}^i$;
2. **Compression** if $0 < v_{t-1}^i - v_t^i \leq \epsilon$ and $v_t^i > 0$;
3. **Collection** if $v_{t-1}^i - v_t^i > \epsilon$ or $v_t^i = 0 \wedge v_{t-1}^i \neq 0$,

where ϵ is the collection threshold.

This means that whenever the volume of waste increases on a container, a waste deposit was made. When the volume of waste decreases to zero, its for sure a collection, but if it decreases to another value and the difference between the volumes is less than the collection threshold (which in our case is set to 20% waste volume), we consider that a waste compression.

With this, it's possible to create a ***reading data frame***, where each line contains information about the container id, date, hour, volume and reading type. This way, we have a data frame that joins the volume reading of all containers. To this new data frame, is possible to add more context about the time of each reading. The following information was added:

- **season** of the year;
- **typeDay**, if it's a holiday or there's a special event in town;
- **weekDay**, day of the week;
- **dayFase**, by the hour, if it's morning (5h-12h), afternoon (12h-19h) or night (19h-5h);

An example of some entries of this data frame is shown in Table 3.2.

Id	**Season**	**Day**	**WeekDay**	**D Fase**	**Date**	**Hour**	**Vol**	**Read**
...	...	...	...	...	...	...	...	...
48843	Spring	N	Thursday	Afternoon	07/06/2018	16:36	51	Deposit
48843	Spring	N	Friday	Morning	08/06/2018	08:03	66	Deposit
...	...	...	...	...	...	...	...	...

Table 3.2: Reading Data Frame example

An analysis of the volumes and reading types is made using this complete data frame. This analysis allows the measurement of the current waste collection frequency.

The data sets containing the information mentioned provide a big portion of the information we intend to use in the study of the capacity-frequency problem. However, because the volume is measured each time a container is opened, these discrete data doesn't have a fixed time period between readings. One container can be opened ten times in a day, while others might not be opened in that same day.

To deal with this, we created a function that generates another dataset in which the ***time period reading data frame*** is defined with a fixed time period of every x hours (2 hours, 8 hours, or even 1 day). Each line of the data frame has, for each container, information about the last measured volume and the mean and median measure of volume in that time period. If there is no volume information in that time period, the volume is considered to be the same as the previous volume measured. It brings much more information to the data set and makes any period of time comparable between different containers. This can also be viewed as a continuous dataset in which the volume of a container on a datetime is the last measure or the average volume in the time period containing that datetime. The algorithm used is the following:

input : $byHours$, $currentDf$
output: $newDf$

$Dates \leftarrow$ sequence from $min(currentDf['Dates'])$ and $max(currentDf['Dates'])$ every $24/byHours$;
$MinHour \leftarrow$ sequence every $byHours$ starting at 00:00;
$MaxHour \leftarrow MinHour + byHours$;
$Vol \leftarrow \{\}$;
for i in $length(Dates)$ **do**
  $samples \leftarrow$ samples in $currentDf$ where $currentDf['Dates'] == Date[i]$ and $currentDf['Hour'] > MinHour[i]$ and $currentDf['Hour'] < MaxHour[i]$;
  if $length(samples) > 0$ **then**
    $Vol \leftarrow Vol \cup samples[1]$;
  else
    $Vol \leftarrow Vol \cup Vol[i-1]$;
  end
end
$newDf \leftarrow dataframe(Dates, MinHour, MaxHour, Vol$;

Algorithm 1: Time period dataframe generation

With this, the *time period reading data frame* is similar to the *reading data frame*, in terms of columns, but instead of *Hour*, we have a *MinHour* and *MaxHour* column, as shown in Table 3.3.

Id	Season	Day	WeekDay	D Fase	Date	MinHour	MaxHour	Vol	Read
...	...	...	...	...	...	...	...	...	
48843	Spring	N	Thursday	Afternoon	07/06/2018	12:00	20:00	51	Deposit
48843	Spring	N	Thursday	Night	07/06/2018	20:00	04:00	51	None
48843	Spring	N	Friday	Morning	08/06/2018	04:00	12:00	66	Deposit
...	...	...	...	...	...	...	...	...	

TABLE 3.3: Time Period Reading Data Frame example

For visualization purposes, another data frame was created with the volumes of the containers, in each time period, side by side. The first three columns of the data frame are the date, initial time and final time (of the period) and the remaining columns are volume measures, each column for each container in study. A visualization of these volumes can be viewed in Table 3.4.

	Date	**IT**	**FT**	**V1**	**V2**	**V3**	**...**
1	21/12/2017	00:00	06:00	0	0	0	...
2	21/12/2017	06:00	12:00	100	54	78	...
3	21/12/2017	12:00	18:00	100	24	10	...
4	21/12/2017	18:00	23:59	100	26	10	...
5	22/12/2017	00:00	06:00	100	26	10	...
6	22/12/2017	06:00	12:00	45	12	69	...
7	22/12/2017	12:00	18:00	45	61	100	...
8	22/12/2017	18:00	23:59	66	61	69	...
9	23/12/2017	00:00	06:00	66	100	69	...
...	...	...	...	...	...	...	...

Table 3.4: Volume Data Frame every 6 hours

This allows a direct comparison between the volumes of the different containers. On top of that, if we consider a similar data frame with the deposits and collections information instead of the volumes, it's easy to see patterns on the current waste collection frequency.

As an addition to this information, we also considered weather information from the National Centers for Environmental Information (NCEI) using the information on air temperature and rain. For the following sections, all these dataframes are used and the results provided with a reading data frame (RDF) and a time period reading data frame (TPRDF) are to be compared.

3.2 Data Analysis

In this section we analyse the data and try to extract information taking advantage of the different dataframes. We start by considering the RDF to get information about the deposits and collections. The variation of volume by day is too big, as it's shown in Figure 3.2.

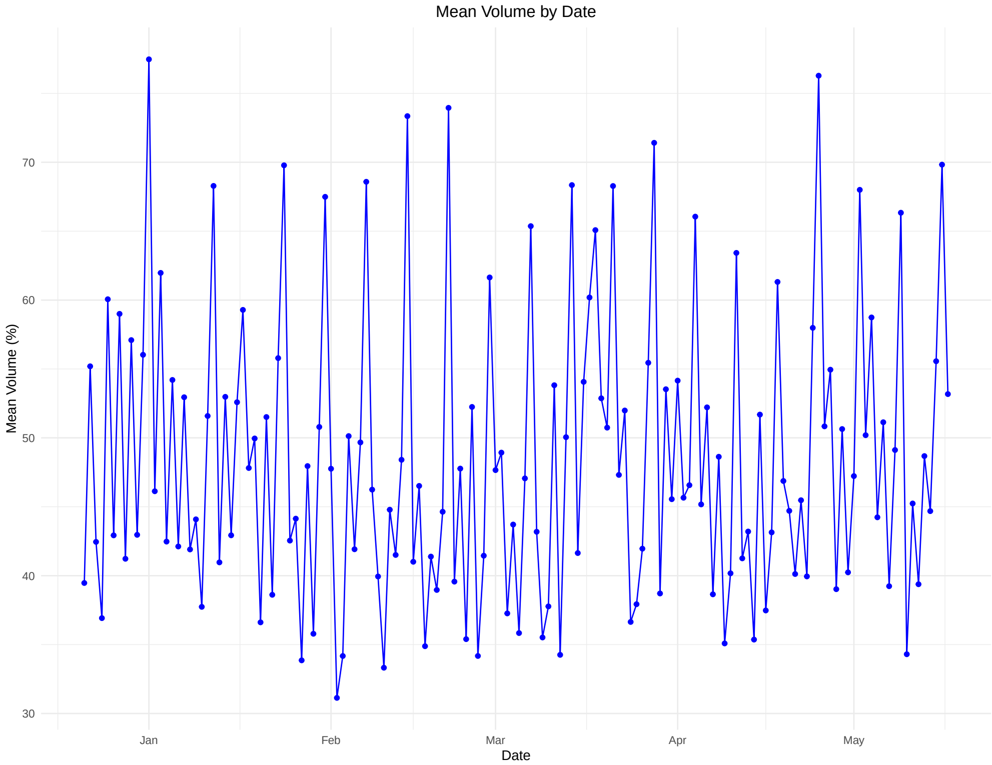


FIGURE 3.2: Mean of Waste Volume of all Containers

To get more precise information, in the following sections we start by visualizing the containers in the study across the city and their average volume by street. Then, a study on the deposits and collections in those containers is made, using the TPRDF, with time periods shorter than a day. To classify the collection efficiency we introduce the concepts of *critical points* and *needless collections*. Finally, we correlate the waste volumes with other data like weekday, season, air temperature and precipitations and summarize the major findings of the data analysis.

3.2.1 Visualization by zone

In this problem, we have access to data from 18 different containers that can be of one of three types: the standard ones, with 800 litres (31450, 48843, 49619, 50443, 50708, 50856, 51698, 52910 and 54452) and 1000 litres (15415, 41483, 44289, 44776, 53181 and 54494) capacity and the underground containers (44263, 44966 and 50419) which can also store 1000 litres. The containers are split across the district of Castelo Branco, on eight different streets, as shown in Figure 3.3, where we can visualize the number of containers that are for disposal for each street, following by their id number.

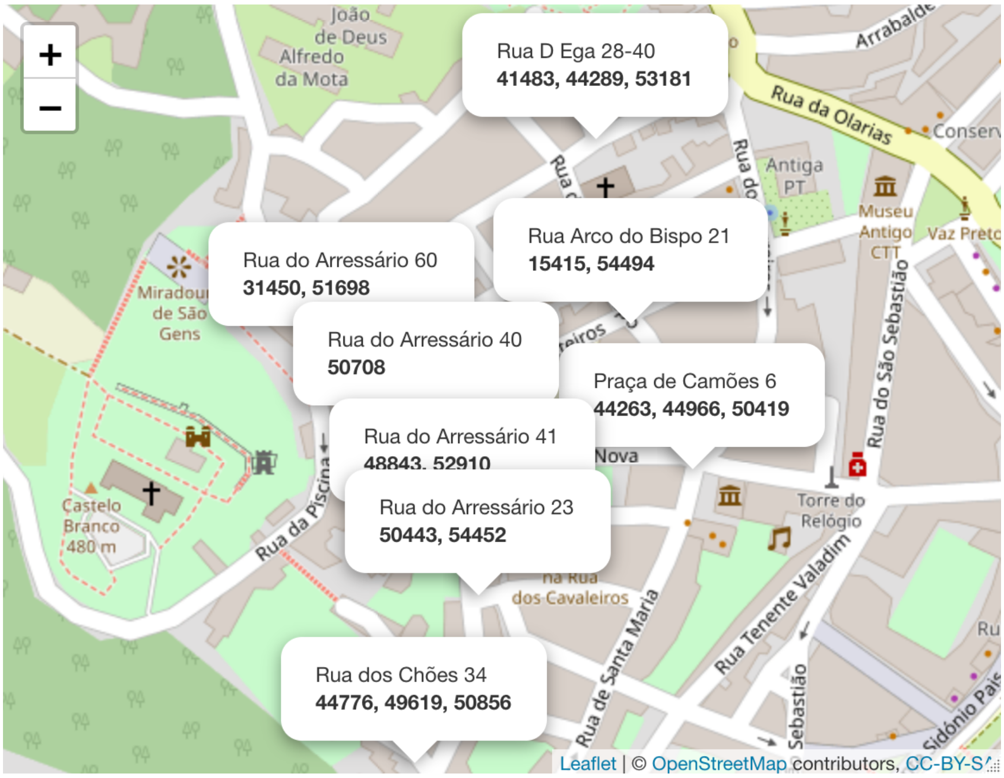


FIGURE 3.3: Container Locations

Considering the waste volume data between 2017-11-01 and 2018-05-31 (period with the most amount of data), Table 3.5 shows the average volume of waste of each cluster (street) by each month. We can see that even for an average calculation, the values seem to appear quite aleatory, however, seems to be increasing over time. Despite the fluctuation, we can notice that most of the volumes are between the range of 30% to 60%.

Dates	2017-11	2017-12	2018-01	2018-02	2018-03	2018-04	2018-05
Arco do Bispo 21	NA	55	62	61	53	58	64
Arressario 60	43	46	47	51	47	54	58
Arressario 41	40	48	47	26	41	63	49
Arressario 23	42	44	56	52	47	52	55
D Ega	51	55	62	47	44	46	48
Praca de Camoes	22	32	33	30	42	34	31
Choes 34	43	46	54	47	57	51	52
Arressario 40	44	43	49	51	47	56	60

TABLE 3.5: Mean of Volume by Street and Month

3.2.2 Deposits and Collections

The most important parameters for the frequency capacity problem, are the deposits and collections that currently happen. Considering Definition 3.1, we start by analyzing the number of deposits and collections that happen on these containers. To this, we consider a period of time between 2017-12-21 and 2018-05-17, which is the period of time with readings from all the containers. The total amount of deposits and collections are shown in Figure 3.4.

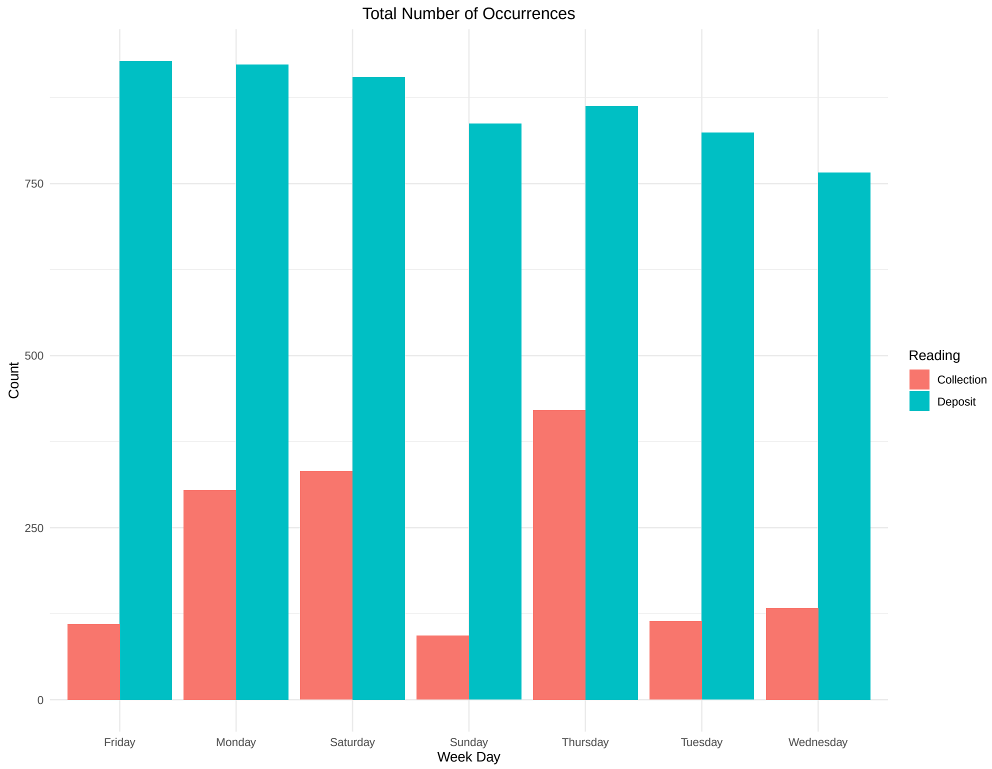


FIGURE 3.4: Total number of Deposits and Collections

To have a notion of the frequency of deposits and collections, we consider the number of weeks between 2017-12-21 and 2018-05-17 (approximately 52 weeks) and the number of containers considered (18). The average of deposits and collections by week and container is shown in Figure 3.5.

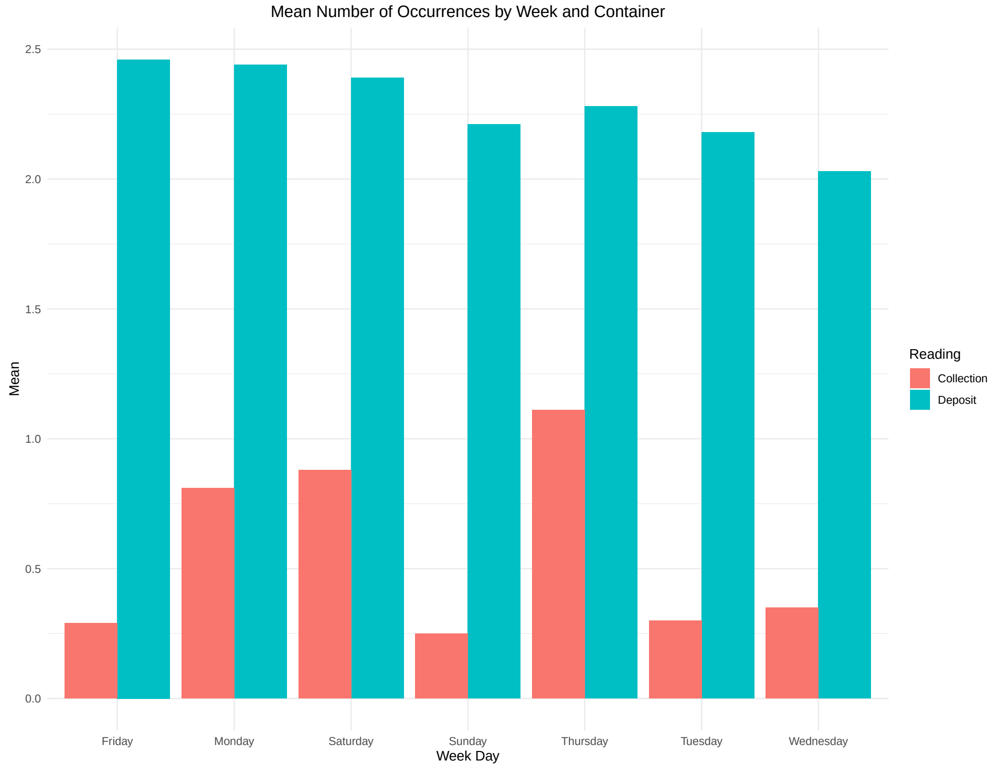


FIGURE 3.5: Mean of Deposits and Collections By Week and Container

We can see that we usually have waste collections on Monday, Saturday and Thursday and sometimes on the other weekdays. This indicates that the waste collection frequency for these data is between 3 to 4 times a week. On the other hand, the number of deposits in each day of the week doesn't fluctuate much in mean, by day of the week.

Regarding the deposits, an interesting statistic to check is the average of the volume of waste that is deposit every day. Figure 3.6 shows this by weekday. We can see that the amount of waste deposit is very similar in each weekday, meaning that there's no pattern on waste deposit by weekday. We check in the following section this statistic against other variables.

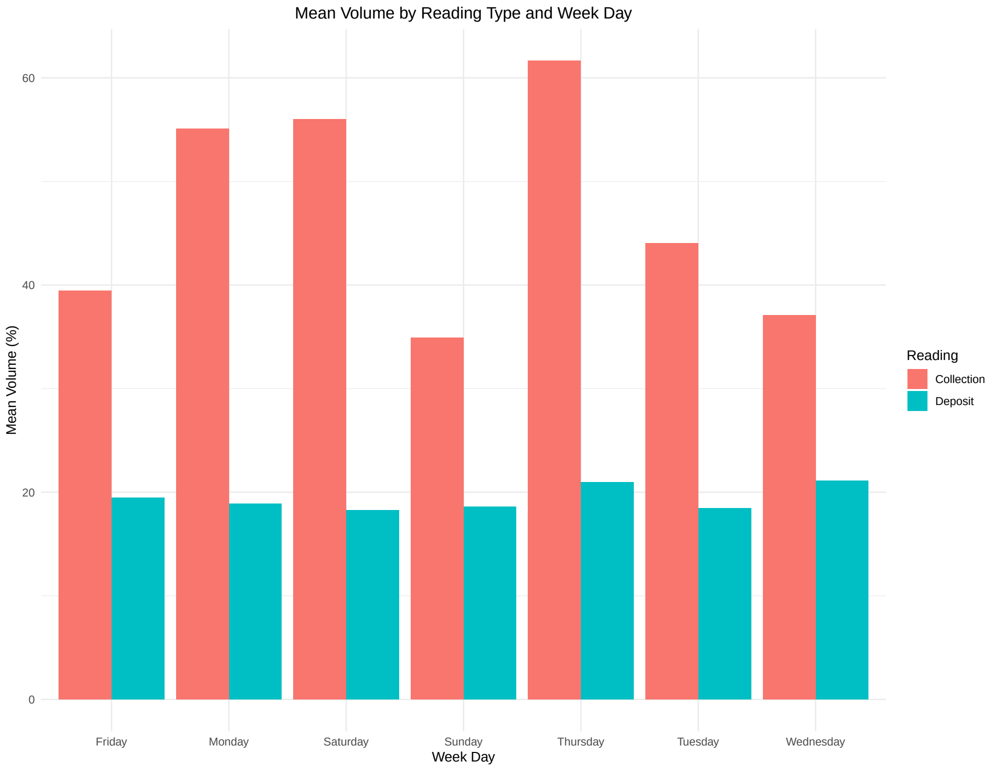


FIGURE 3.6: Mean Volume by Reading Type and Week Day

The high values of the volume collected on Monday, Saturday and Thursday are consistent with the number of times the waste is collected on those days.

3.2.3 Collection analysis

Considering the waste volume in the containers when waste collections are made, it is possible to evaluate, for each container, how well the current waste collection frequency performs. To do so, let us consider the following definitions:

Definition 3.2. Given a reading of a container i corresponding to a waste collection, with volume v_t^i, we consider it a ***needless collection*** if the waste in the container was less than 35%, that is, $v_{t-1}^i < 35$.

Definition 3.3. Given a TPRDF with time period of one hour and a reading of a container i, with volume v_t^i, we consider it a ***critical point*** if the volume of the container is 100% for more than one day, that is $v_k^i = 100, k \in \{t, t-1, ..., t-24\}$.

According to the collections of each container, the percentage of needless collections are presented in Figure 3.7.

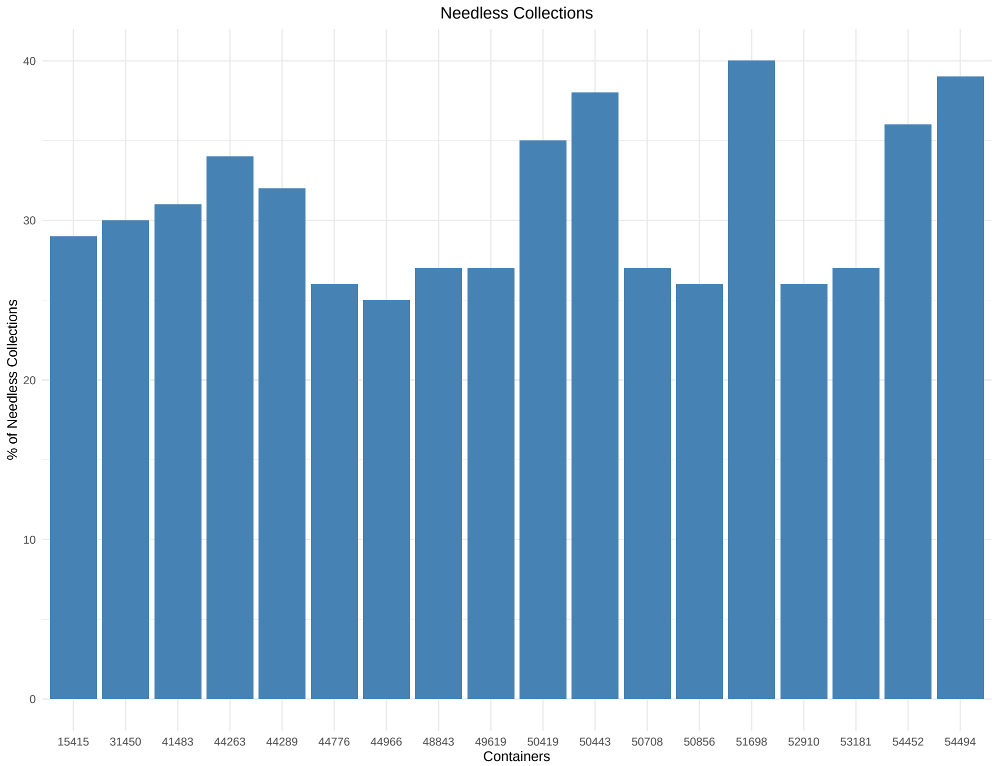


FIGURE 3.7: Percentage of *Needless Collectionns* by Container

We can see by this chart that at least 1/4 of the collections made, in each container, are needless and could have been done later. For seven of the eighteen containers, the amount of needless collections is more than 1/3. This indicates that there is progress margin to decrease the waste collections frequency of these containers.

On the other hand, it's important to check for critical points. The best way to see critical points is to use a time period data frame, that can easily provide information like how much time a container was full. We can see the total amount of critical points for each container in Figure 3.8.

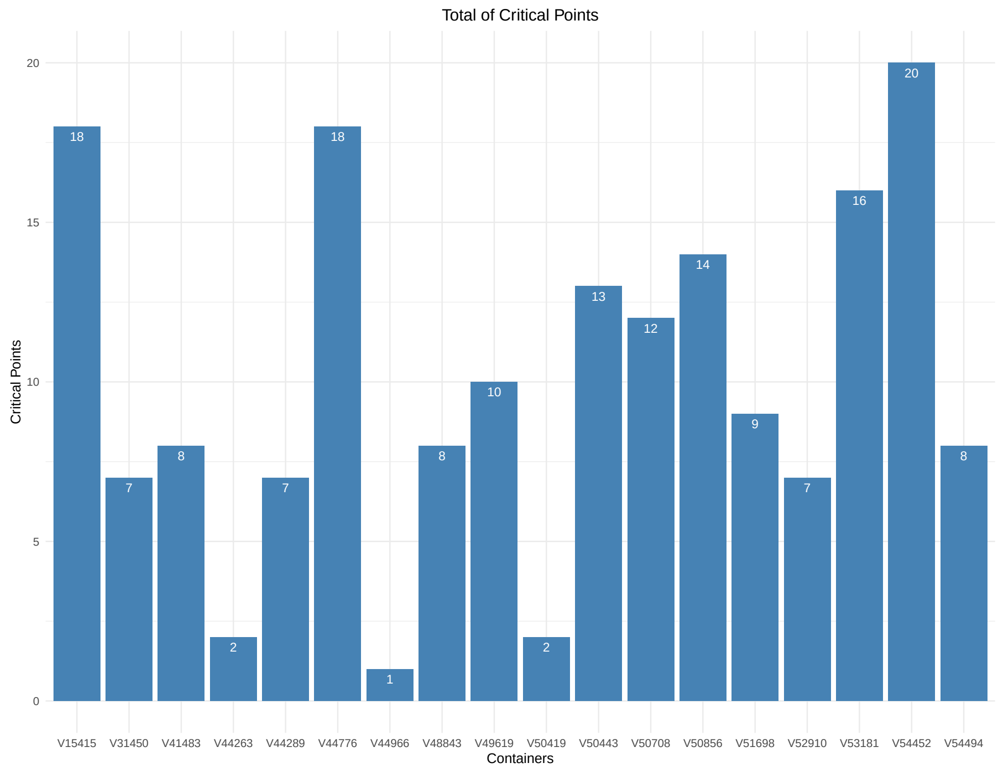


FIGURE 3.8: Number of *Critical Points* for each Container

We can see that the containers 5415, 44776, 53181 and 54452 have a high number of critical points, which suggests that they should have a more frequent waste collection frequency. Even the remaining containers seem to have been full for more than one day too many times, with the exception of 44263, 44966 and 50419. On the other hand, they all presented a high percentage of needless collections.

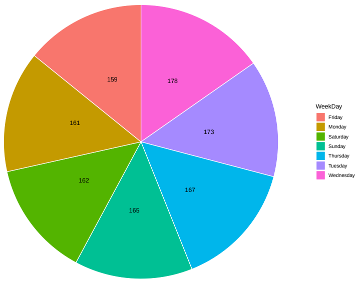
This data shows that the waste collections frequency and/or the capacity of the containers can be changed and improved for each container. It suggests that many times the collections are needless and other times the waste should be collected and it isn't. This is where the information of the volume of waste in each container, at real-time, can prove to be of great usage when optimizing waste collection.

Ideally, there would be no critical points or needless collections, but our focus is not to minimize these points individually for each container but to consider all the containers grouped by their location and address the frequency capacity problem.

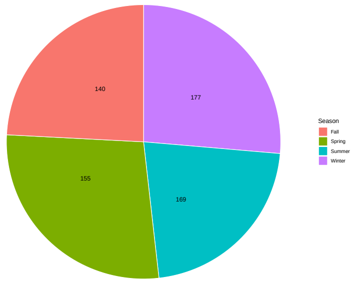
3.2.4 Data Correlation

In this subsection, we try to find patterns that may influence the number and volume amount of waste deposits. To do this, we use TPRDF to fill any gaps in the volume data and have a more significant data set.

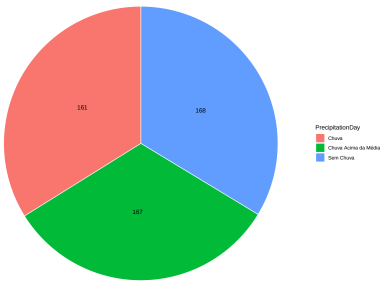
This case studies four scenarios, according to the class weekday, with every day of the week, season, which represents the partition of the database into the different seasons, precipitation [mm], which can represent days without rain, rainy days or very rainy days and air-temperature, that vary from a mild day, cold day, very cold day, hot day or very hot day. The mean volumes of the waste deposits can be seen in Figure 3.9.



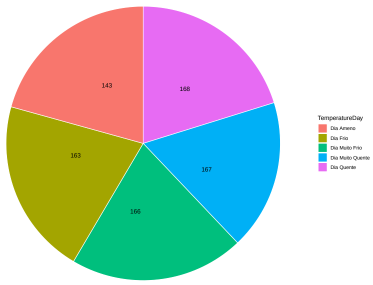
(A) WeekDay



(B) Season



(C) Precipitation



(D) Temperature

FIGURE 3.9: Mean Volumes of Waste Deposits

Relatively to the levels of precipitation, we can notice the average waste deposits are very close to each other, showing our lower value of 161 on a normal rainy day, and the higher value of 168, in days without rain. Concerning to the air temperature, the verified mean values of waste deposits differ from 143 to 168 litres. The amounts are also very similar, with the exception of the variable very hot day, which is a much lower value. This may be due to the seasonal time corresponding to the summer or simply because the tendency to deposit waste is less on very hot days. Comparatively to the season, one can observe the volume of deposits in the summer is significantly lower than in the rest of the seasons. There seem to appear some correlation between the variables summer and very hot day and so, the values can be interpreted as the seasonal time of the year, where a set of families go out to another city which decreases the demographic population of Castelo Branco.

Another variable seen was the type of day and, in average, the amount of waste deposits is similar between the type of days, which are normal days, holidays and event days. The values obtained were close to each other but they do not provide much value because the sample of other types of day is of very small size.

In short, we can observe that the most promising class is the season, probably due to the reduced number of people in summer. That can be used as a natural division to solve the capacity frequency problem in multiple periods of time. We expect that summer may require a lower waste collection frequency that the other seasons.

3.3 Major Findings

In this section it was shown a lot of information about the dataset and a good data visualization and analysis, which is to be used to leverage information for the algorithms coming in the following sections.

Regarding the class day-of-week we saw that the frequency of collection is dynamic, as it may vary according to the time of the year. Also, the days of week for collection are fixed on Monday, Thursday and Saturday. Tuesday is also added when the frequency is increased to four.

The daily average volume of the containers is mostly between 30% and 60%, and the few days where the volume is higher than 60%, most of them correspondent to Wednesdays. As the volume of deposits is, on average, about 180 litres per day and the containers have capacity between 800 to 1000 litters, it is expected to be easily possible to decrease the frequency to three times a week and in the summer to two times a week.

A quick analysis on the current collection showed that a large percentage of the collections are needless collections and some of the containers have a considerable amount of critical points, which leads to the idea that waste collection can be generally improved, not only by its frequency but also by the days and hours that the waste is collected.

On data correlation to waste deposits, it was found that the classes weekday, precipitation and air temperature, haven't shown concrete results, as the variation was very low. Relatively to the class season, this indicated us that in the summer season, the volume of waste deposits decayed, which may be due to less population density and we must take that into account.

Chapter 4

Frequency-Capacity Problem

The Frequency-Capacity Problem is the major focus of this dissertation and it is important to define it correctly before moving on to its resolution. In this chapter, we start by defining the problem and some variants. Then we move to its resolution: we propose three different algorithms to help solve the Frequency-Capacity Problem.

4.1 Definition

The Frequency-Capacity Problem (FCP) is the problem of, given a set of containers in a city and its historical volume data, find the best frequency of waste collection and container volume adjustments so that no container is ever full and the overall cost is minimum. Volume adjustments are the replacement of containers for bigger containers or the addition of containers in specific zones to increase container capacity on those zones.

A container in the problem can be viewed as a single container or a set of containers close together (container cluster). The overall cost is defined by the waste collection cost (in a year) plus the cost of the containers addition. Each container or container cluster has a maximum capacity that cannot be exceeded. Otherwise, we could arrive at solutions where the streets were full of containers.

4.1.1 Problem Formulation

Considering the set of n containers with $\{c_1, ..., c_n\}$ and its historical volume data $\{\{v_{11}, ..., v_{1m}\}, ..., \{\{v_{n1}, ..., v_{nm}\}\}$, we want to find the waste collection frequency f and the new containers capacity $\{c_1^{'}, ..., c_n^{'}\}$ so that the the new volume data $\{\{v_{11}^{'}, ..., v_{1m}^{'}\}, ..., \{\{v_{n1}^{'}, ..., v_{nm}^{'}\}\}$ is always less than 100% and we minimize the waste collection cost w_c and container addition cost a_i for each container i. This is an optimization problem and can be formulated as it follows:

$$\text{minimize } w_c f + \sum_{i=1}^{n} a_i$$

$$\text{s.a. } v_{ij}^{'} < 100, i = 1, ..., n, j = 1, ..., m \tag{4.1}$$

$$c_i^{'} < c_{max}, i = 1, ..., n, \tag{4.2}$$

Restrictions (4.1) guarantee that the new volume of waste is never above 100% for every container at every moment. Restrictions (4.2) guarantee that no container of a cluster of containers has more than a maximum capacity that as to be defined to prevent too many containers on the streets. This problem has a total of $m \times n + n = (m + 1) \times n$ restrictions, where n is the number of containers and m is the number of readings. As we consider for this problem a fixed time period and the time period data frame with readings every x hours, we have a fixed m equal for each container.

One of the challenges of this problem is to, given the new frequency and volume capacity, calculate the new volumes of waste for every container at every moment. This dependency suggests that numerical algorithms that tend to the solution over iterations are a natural approach to solve this problem.

4.1.2 Variants

With the historical volume data on a time period it's possible, as shown in Chapter 3, to deduce when the waste collections were done on that time period. With that, it's possible to take into account needless waste collections and critical points. A variant of the FCP is to find the best frequency of waste collection and container volume adjustments so that no needless or critical points occur and the overall cost is minimum.

A more generic approach would be to not fix a frequency for the whole year but to find a set of fixed frequencies for example by year station or even month. This problem could take advantage of some historical variables like season, month or even event days.

In the following sections we propose algorithms to solve the FCP using containers individually and clusters of containers by zone. We then proceed to the FCP in its generic variant.

4.2 Algorithms

To solve the FCP problem making the most of the data available, we divide the problem in three steps that can help to find a good solution after few iterations. The first step is to, after defining a fixed waste collection frequency (for example one/two/three times a week), to see what happens to the waste volume in the containers every hour. This is to be done with the *Volume Simulation Algorithm*. A second step is to check if the current capacity of the containers is enough to satisfy the problem restrictions or if it needs to be readjusted. This is made with the *Capacity Readjustment Algorithm*. After that, a new validation is made. This is made until a solution is found for the problem.

4.2.1 Volume Simulation

With the historical data from each container, it's possible to simulate what happens to the volume waste after fixing a new waste collection frequency, for every container. By fixing a set of dates D' and times T' for the new collection and, from the historical volume data by hour $V = \{\{v_{11}, ..., v_{1m}\}, ..., \{v_{n1}, ..., v_{nm}\}\}$, being n the number of containers and n the number of records over time, we generate an entire new set of volume data $V' = \{\{v'_{11}, ..., v'_{1m}\}, ..., \{v'_{n1}, ..., v'_{nm}\}\}$. The algorithm used is the following:

```
input  : V, D', T'
output: V'
numberContainers ← length(V) ;
for i in numberContainers do
    gap ← 0;
    currentV ← V_i;
    V'_i ← {};
    for v in currentV do
        d ← date of v;
        t ← time of v;
        if d ∈ D' and t ∈ T' then
            gap ← −v;
            status ← Model Collection ;
        else
            if prev_v − v > δ then
                gap+ = prev_v − v;
                status ← Original Collection;
            else
                status ← Normal Register;
            end
        end
        V'_i ← V'_i ∪ max(v + gap, 0)
    end
    V' ← V' ∪ V'_i
end
```

Algorithm 2: Volume Simulation Algorithm

This means that for every entry of the historical volume data set we check date d, hour h and volume v. If $d \in D'$ and $t \in T'$ it's time for a new collection, so we set $gap = -v$. This is done because the new volume is always the sum of the old volume with the gap, and this way, the new volume is 0. Otherwise, we check if it was an old collection and we set $gap+ = prev_v - v$ otherwise gap stays the same. Then we set the new volume for this date and time $v^{'} = v + gap$.

4.2.2 Capacity Readjustment

After obtaining the new set of waste volumes, it's easy to see if the capacity of the container is enough for every container cluster, since its waste volume has to be under 100%. If constrains (4.1) are not satisfied, the capacity of each container cluster is increased by the maximum waste volume registered for each cluster. If that increasing violates constrains (4.2), the waste collection frequency has to be changed. Otherwise, the new volume set V' is recalculated accordingly with the capacity of the new clusters.

input : $V^{'}$, C
output: $V^{''}$, $C^{'}$

$numberContainers \leftarrow length(V)$;
$Ctp \leftarrow \{\}$; % Containers to Update
$C^{'} \leftarrow C$;
$V^{''} \leftarrow V^{'}$;
for i *in* $numberContainers$ **do**
  $maxVol_i \leftarrow max(v^{'} \in V_i^{'})$;
  if $maxVol_i > 100$ **then**
    $c_i^{'} = c_i^{'} * maxVol_i$;
    $Ctp \leftarrow Ctp \cup i$;
end
for i *in* Ctp **do**
  for $v^{''}$ *in* $V^{''}$ **do**
    $v^{''} \leftarrow (v^{'} * c_i)/c_i^{'}$
  end
end

Algorithm 3: Capacity Readjustment Algorithm

Joining this algorithm with the *Volume Simulation Algorithm*, we can simulate the volume of waste over time of the new waste collection module with the new waste collection frequency and containers capacity.

Note that the use of containers clusters is highly important for these algorithms because it's easy that a single container reach a volume waste above 100%. When considering clusters of containers next to each other, it's important just to ensure that the sum of their waste volumes is less than $n \times 100\%$, where n is the number of containers in the cluster.

4.2.3 Multiple Period Algorithm

The idea of the previous algorithms is to take advantage of the locations of the containers to form clusters of containers and the historical volume data to define a good frequency-capacity for a time period. But, as refered in Chapter 3, some periods of time may require more waste collections that others.

The purpose of the multiple period algorithm is to define a good set of time periods so that the previous algorithms can be applied to each period. Considering g_t the daily volume growth rate of a time period, we define that:

- If $g_t \leq 12\%$, that period may require only a waste collection per week;
- If $12\% < g_t \leq 25\%$, that period may require only two collections per week;
- If $g_t > 25\%$, that period may require three or more waste collections per week.

In order to define specific rules for each period (to make the solution doesn't fit only to the current volume data) we can take into account the correlation between events, year seasons, month and weather conditions with the waste volume growth rate for each container or cluster of containers.

4.3 Finding Solutions

In order to find good solutions (and the best) for the FCP, a combination of the algorithms above is to be used. Given a set of volumes, the first thing to do is to calculate the current waste collection frequency and its total cost. Next, is to chose a new collection frequency, apply the *Volume Simulation Algorithm* and the

Capacity Readjustment Algorithm and finally calculate the total cost of the new solution.

The costs associated with waste collection are typically far bigger than the increase of container capacity, so finding the better solution is on a standard case, trying to get possible solutions for a waste collection frequency of once, twice or three times a week.

1. Calculate current frequency and total cost;
2. Define new frequency;
   (a) *Volume Simulation Algorithm* -> new volume data;
   (b) *Capacity Readjustment Algorithm*-> validate if this is a possible solution and readjust container capacity;
3. Calculate current cost and check improvement;

Algorithm 4: Standard Algorithm

When multiple period are involved, the *Multiple Period Algorithm* must be applied first and then the *Volume Simulation Algorithm* and the *Capacity Readjustment Algorithm* must be applied to each time period, in order to find the overall solution with the minimum cost.

1. Calculate current frequency and total cost;
2. Find patterns to define time periods
3. Apply **Standard Algorithm** to each period;
4. Calculate current cost and check improvement;

Algorithm 5: Dynamic Algorithm

It's possible that the priority of a city council is to avoid critical points and needless collections. The previews algorithm may be used for that, calculating the critical points and needless collections for each period.

1. Calculate current frequency, critical points and needless collections;
2. Find patterns to define time periods
3. Apply **Standard Algorithm** to each period;
4. Calculate current critical points and needless collections;

Algorithm 6: Dynamic Algorithm on Critical Points and Needless Collections

These algorithms provide a complete information on the cost reduction produced by the solutions. As we do not know the current waste collection costs or the costs of adding new containers, in the following Chapter we just focus on finding the best solutions in terms of reducing the waste collection frequency and find the necessary capacity adjustments for those solutions.

Over the next chapter, we solve the FCP using the standard algorithm, considering each container separately and then moving on to clusters of containers to check if, for these set of containers, by using clusters we can reduce the cost of the capacity readjustment. We then try to solve the FCP using the dynamic algorithm, by defining first the time periods considering other data like season. For simplification purposes, we just try to get a better waste collection frequency for summer season, which has the least waste deposits.

Chapter 5

Problem Resolution

The study presented in Chapter 3 suggested that the current waste collection frequency in Castelo Branco is between three to four times a week, most of the time on Monday, Thursday and Saturday. On the other hand, if we consider that we just need to collect a container waste if the container has more than 35% of waste volume, it was shown that more than 40% of the past collections were needless collections, meaning the collection frequency should be easily decreased.

In this chapter we apply the algorithms mentioned in Chapter 4 to analyze what would happen if the waste collection frequency were reduced to one, two or three times a week. To do this, we start by applying the standard algorithm which consists of applying the volume simulation algorithm and the capacity readjustment algorithm in order to find solutions to the FCP. Next we see what changes in the solutions if we consider clusters of containers and how that influence the capacity readjustment. Lastly, we see how to apply the dynamic algorithm by defining multiple time periods and solving the problem for each one of them. Conclusions on the results are made.

5.1 Solving with Standard Algorithm

In this section, we want to analyze what is the capacity needed if we reduce the waste frequency to one, two or three times a week and present good models to find the best day or days for waste collection. To validate these models, an analysis of the containers overload (new volumes provided by the model higher than 100%) is made.

5.1.1 One Time a Week

To start, we simulate a situation where the waste collection frequency would be once a week considering each container as a container cluster in separate.

Considering a time period data set with the time period of 2 hours we start by applying the Volume Simulation Algorithm to generate the new volumes of waste if the waste collection frequency was once a week (Thursday at 10P.M.). The figure 5.1 shows an example of the new model volume, compared to the real volume with the current collection frequency:

As we can see, a waste collection frequency of once a week is not enough for container 49619 between 06/08/2017 and 15/10/2017, with too many occurrences of waste overload. Table 5.1 shows the mean of the new volume by container and the number of waste overloads for each container.

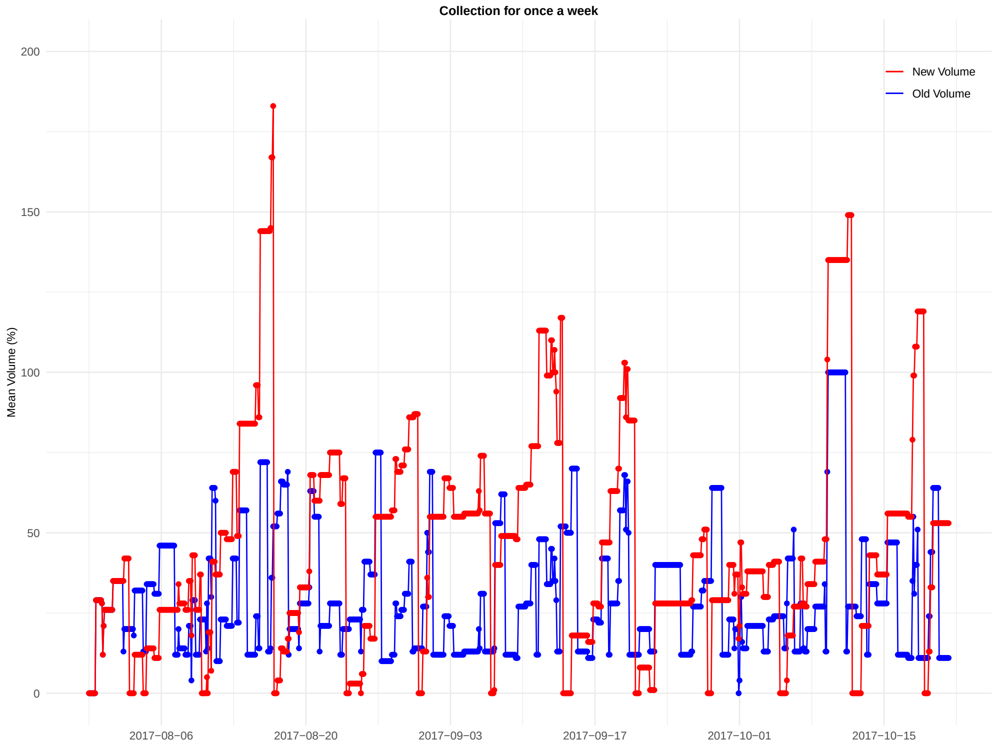


FIGURE 5.1: First Records of Once Week Frequency for container 49619

Container	Average Volume	Nº of Overloads
15415	114	81
31450	72	85
41483	103	100
44263	32	26
44289	108	95
44776	101	104
44966	125	114
48843	103	103
49619	51	44
50419	66	56
50443	104	96
50708	96	102
50856	105	126
51698	112	104
52910	82	85
53181	116	112
54452	127	114
54494	102	84
Mean	**95%**	**90**

TABLE 5.1: Once a Week Frequency Results

We can see by the results that a collection frequency of once a week is clearly not enough for these containers. The average volume in the containers is almost their full capacity and the number of overloads are incredibly high which asks for improvement of the container capacity or the collection frequency.

Applying the capacity readjustment algorithm to this data we see the improvement needed on the capacity of the container in order to never have an overload of waste, in each container. The problem with this approach is that the new capacity is as big as the maximum waste volume ever seen on the new model.

Container	Old Capacity	New Capacity
15415	1000	3650
31450	800	2560
41483	1000	3910
44263	1000	1430
44289	1000	2970
44776	1000	2470
44966	1000	3800
48843	800	2688
49619	800	1640
50419	1000	2090
50443	800	2256
50708	800	2344
50856	800	2272
51698	800	2920
52910	800	2360
53181	1000	3160
54452	800	3208
54494	1000	3150
Total	**16200**	**48878**

TABLE 5.2: Capacity Readjustment Needed for Waste Collection on Thursday

The results are shown in Table 5.2, using the volume data generated by the volume simulation algorithm. Some of the containers capacity improvements need a capacity increase of around three times the current capacity. That certainly violates restrictions (4.2) of the FCP, when considering any reasonable value for c_{max}.

This results are when we consider a waste collection of one time a week on Thursday at 10P.M., but that may not be the best day and hour to collect the waste. To take better conclusions on the sufficiency of a collection of once a week, we applied the volume simulation algorithm and the capacity readjustment algorithm for every day of the week (we considered that the waste can only be collected at night around 10P.M.) and considered the best day. The best day is the day in which the sum of the new capacities is the least.

In this case, the best day of the week to collect waste, when considering a waste collection frequency of once a week is Sunday, with the overall capacity increase of 55044 litres. Table 5.3 shows the capacity readjustment needed for each container.

Container	**Old Capacity**	**New Capacity**
15415	1000	3360
31450	800	2528
41483	1000	3080
44263	1000	1670
44289	1000	3190
44776	1000	2420
44966	1000	3150
48843	800	2864
49619	800	1896
50419	1000	2000
50443	800	2088
50708	800	1984
50856	800	2240
51698	800	3592
52910	800	2256
53181	1000	2890
54452	800	2592
54494	1000	2970
Total	**16200**	**46770**

Table 5.3: Capacity Readjustment Needed for Waste Collection on Sunday

Even on the best day, the overall capacity would have to increase around three times the current capacity, which is not practicable. With this result we immediately conclude that a waste collection of one time a week is not suitable for our problem data.

5.1.2 Two times a Week

After trying a waste collection of once a week, we simulate a situation where the waste collection frequency would be twice a week considering each container as a container cluster in separate. We start by showing an example similar to the previous shown and then apply the volume simulation algorithm and the capacity readjustment algorithm for every pair of days of the week.

Considering the same container and time period, Figure 5.2 shows the simulation for a twice a week frequency (Wednesday and Sunday at 10P.M.):

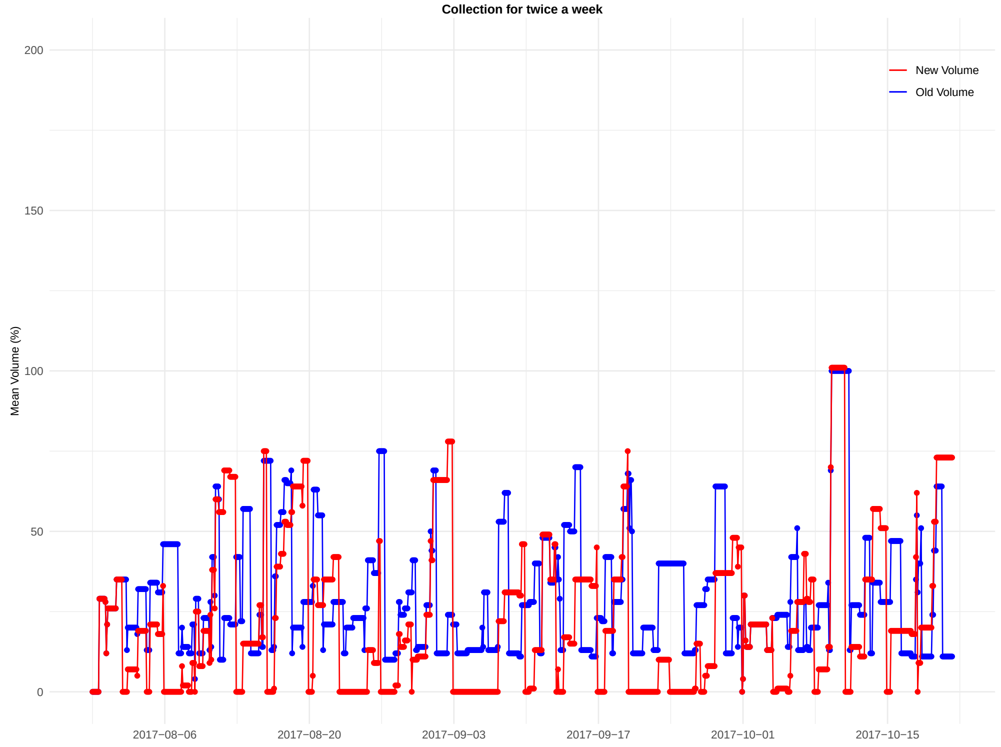


Figure 5.2: First Records of Twice Week Frequency for container 49619

We can see in this example that a waste frequency of twice a week is perfectly enough for container 49619 between 06/08/2017 and 15/10/2017, with only two occurrences of waste overload.

Table 5.4 shows the mean of the new volume by container and the number of waste overloads for each container.

Container	**Average Volume**	**N° of Overloads**
15415	62	35
31450	36	22
41483	53	34
44263	15	2
44289	56	29
44776	49	22
44966	53	35
48843	49	36
49619	27	8
50419	27	4
50443	46	21
50708	46	16
50856	56	37
51698	58	37
52910	37	19
53181	57	37
54452	63	41
54494	52	25
Mean	**46%**	**25**

TABLE 5.4: Twice a Week Frequency Results

The results are much more reasonable, with a total average of 46% of volume. Besides, the number of waste overloads are much less than when considering once a week, as expected.

Applying the Capacity Readjustment Algorithm using the volumes generated by considering a waste collection frequency of two times a week on Wednesday and Sunday at 10P.M., we get the results shown in Table 5.5.

Container	**Old Capacity**	**New Capacity**
15415	1000	2480
31450	800	1296
41483	1000	1910
44263	1000	1190
44289	1000	2040
44776	1000	1470
44966	1000	2290
48843	800	1600
49619	800	1200
50419	1000	1350
50443	800	1720
50708	800	1152
50856	800	1384
51698	800	2328
52910	800	1520
53181	1000	1930
54452	800	1712
54494	1000	1790
Total	**16200**	**30362**

Table 5.5: Capacity Readjustment Needed for Waste Collection on Wednesday and Sunday

As we can see from these results, the capacity increase needed is still a little too high for us to consider this a possible solution, but again, these days may be a bad pair of days to collect waste. Running both algorithms for every pair of days of the week, Monday and Friday are the best days for waste collection.

Container	**Old Capacity**	**New Capacity**
15415	1000	1830
31450	800	1344
41483	1000	1790
44263	1000	1050
44289	1000	2250
44776	1000	1280
44966	1000	1960
48843	800	1456
49619	800	1392
50419	1000	1290
50443	800	1328
50708	800	1432
50856	800	1432
51698	800	1472
52910	800	1592
53181	1000	1630
54452	800	1600
54494	1000	1930
Total	**16200**	**28058**

Table 5.6: Capacity Readjustment Needed for Waste Collection on Monday and Friday

Table 5.6 shows the results for the best pair of days in twice a week waste collection frequency. These results show that, for a waste collection frequency of twice a week, we had to increase the capacity of the overall container by almost the double capacity. This is because we are trying to prevent the worst-case scenario of having any waste overload. The average volume of 46% seem to suggest that if we consider clusters of containers for the capacity readjustment, instead of individual containers, it's possible that that readjustment would be much less than doubling the capacity. For now, however, this is not a solution to consider.

5.1.3 Three Times a Week

As seen in Chapter 3, the current waste collection frequency in Castelo Branco is between three to four times a week. This means that the three times a week model is the less attractive because it provides little improvement on the current waste collection cost, but it can be the only solution for when considering individual containers and not considering multiple time periods.

Running the volume simulation algorithm and the capacity readjustment algorithm for every three days of the week, we see that the best days for waste collection are Monday, Thursday and Saturday. For these days, Table 5.7 shows the mean of the new volume by containers and the number of waste overloads for each container.

We can see that there are containers that don't even have any overloads, which means they won't need a capacity readjustment. Other containers like 41483, 44289 and 48843 still have some overloads, which is probably inevitable.

Container	Average Volume	Nº of Overloads
15415	33	6
31450	21	5
41483	34	16
44263	12	0
44289	36	10
44776	32	1
44966	37	8
48843	33	11
49619	19	5
50419	23	0
50443	33	4
50708	32	8
50856	31	6
51698	34	8
52910	29	6
53181	36	6
54452	38	9
54494	32	6
Mean	**30%**	**6**

Table 5.7: Twice a Week Frequency Results

We can see in Table 5.8 the necessary capacity improvements that need to be made to have a waste collection frequency of three times a week for the whole year.

Container	Old Capacity	New Capacity
15415	1000	1330
31450	800	1048
41483	1000	1440
44263	1000	1000
44289	1000	1870
44776	1000	1010
44966	1000	1710
48843	800	1056
49619	800	1008
50419	1000	1000
50443	800	976
50708	800	1160
50856	800	1056
51698	800	1112
52910	800	1352
53181	1000	1350
54452	800	1192
54494	1000	1340
Total	**16200**	**22010**

Table 5.8: Capacity Readjustment Needed for Waste Collection on Monday and Friday

We can confirm that containers 44263 and 50419 don't need a capacity improvement. Container 44776 improvement is residual and containers 31450, 48843, 49619, 50443, 50856 and 51698 need to be upgraded from a 800 litres to a 1000 litres container. Containers 44289 and 44966 need to practically double their capacity. The remaining ones just need little improvement.

This proves to be a good solution for the FCP, for this set of data.

5.1.4 Results Conclusion

In this first attempt to solve the FCP, we saw that a waste collection frequency of once and twice a week is not enough when considering individual containers and an overall period of a year. However, the results when considering a waste collection frequency of two times a week are quite interesting as they provide an overall average volume of 46% on containers, which leads to believe that if we consider clusters of containers for the capacity readjustment, this may be a possible solution for the problem. This case is study on the next section.

For the three times a week waste collection frequency case, we saw that with little improvements on the capacity of some containers, this frequency is enough for this set of historical volume data. This result alone is already an improvement of the current waste collection frequency in Castelo Branco, that is between three to four times a week. As the containers capacity improvement have a one time cost, this solution provides cost reduction to the city council.

In the following section, we consider clusters with several containers instead of individual containers, to see how that affects the container capacity improvement and we check if the solution of two times a week is enough for those clusters.

5.2 Models Considering Clusters

As shown in Chapter 3, this is possible to high proximity of some containers (for example containers 41483, 44289 and 53181), so we can assume that if a container is full, one can simply use another container on the cluster.

To do this, we agregate the containers by their location, which in this case means considering the following clusters:

- Cluster of *Rua D Ega 28-40*: 41483, 44289, 53181;
- Cluster *Rua Arco do Bispo 21*: 15415, 54494;
- Cluster *Praça te Camões 6*: 44263, 44966, 50419;
- Cluster *Rua do Arressário 60*: 31450, 51698;
- Cluster *Rua do Arressário 40*: 50708;
- Cluster *Rua do Arressário 41*: 48843, 52910;
- Cluster *Rua do Arressário 23*: 50443, 54452;
- Cluster *Rua dos Chões 34*: 44776, 49619, 50856.

The data frames associated with each cluster have the capacity and volume (by each date and time) equal to the sum of corresponding capacities and volumes of the containers on each cluster.

5.2.1 Two times a Week

Considering, for example, the cluster *Rua D Ega* and like previously a period dataset with the time period of 2 hours, Figure 5.3 shows the simulation for a twice a week frequency on Monday and Friday, the best days for collection as seen in the previous section:

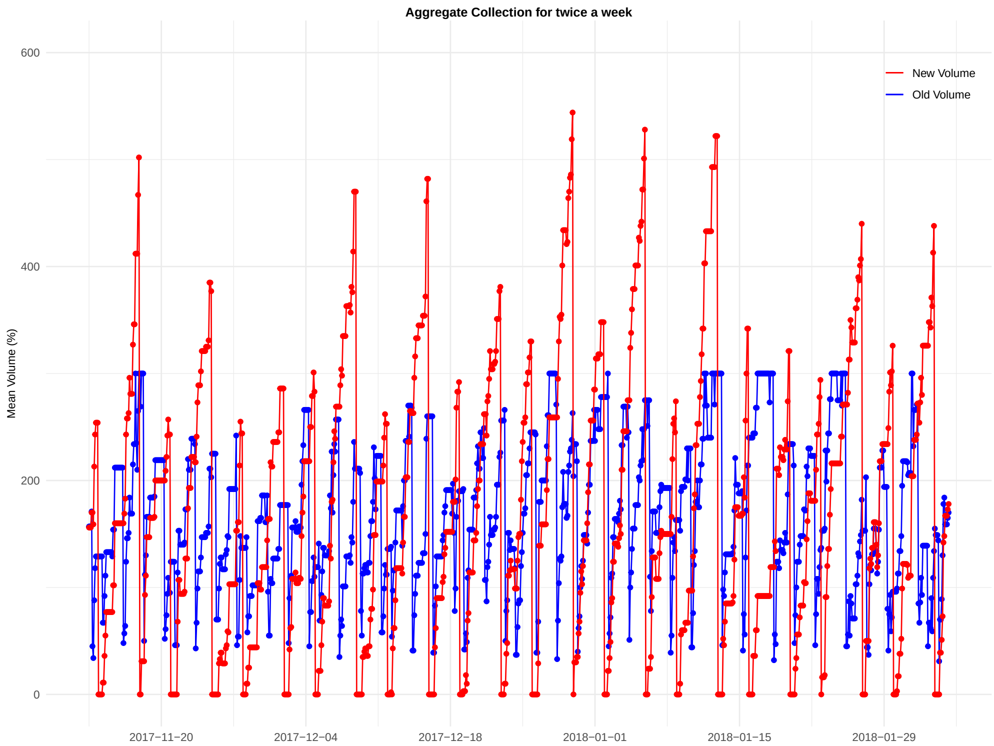


FIGURE 5.3: First Records of Twice Week Frequency for cluster *Rua D Ega*

In this graphic, we see values between 0% and 300% volume due to having three containers in this cluster. This is just an example of the aggregated volumes for this container.

Cluster	Mean V	N Containers	Mean V/Container	Overloads
Arco do Bispo 21	97	2	49	13
Arressario 60	92	2	46	12
Arressario 41	92	2	46	19
Arressario 23	107	2	54	19
D Ega	165	3	55	24
Praca de Camoes	103	3	34	0
Choes 34	124	3	41	10
Arressario 40	50	1	50	24
Mean	**103**	**2,25**	**46**	**15**

TABLE 5.9: Aggregate Twice a Week Frequency Results

Mean V	Mean V/Container
71	36
53	27
52	26
60	30
105	35
91	30
100	33
46	46
72	**33**

TABLE 5.10: After Capacity Readjustment

We can see in Table 5.10 the average volume of the clusters and of each container in the clusters. The number of waste overloads is far less than when considering individual containers for the same waste collection frequency.

Using the capacity readjustment algorithm, the new capacities for each clusters can be seen in Table 5.11.

Container	Old Capacity	New Capacity
Arco do Bispo 21	2000	3450
Arressario 60	1600	2288
Arressario 41	1600	2192
Arressario 23	1600	2416
D Ega	3000	4430
Praca de Camoes	3000	3000
Choes 34	2600	3284
Arressario 40	800	1432
Total	**16200**	**22492**

TABLE 5.11: Capacity Readjustment Needed for Twice Week Frequency With Clusters

In this case, the new capacity of *Arressario 40* is the same as the new capacity of container 50708, because that's the only container in this cluster. Cluster *Praca de Camoes* doesn't need changing and cluster *Arco do Bispo* and cluster *D Ega* need two new containers of 800 litters. The remaining clusters just need an extra container of 800 litres.

We can see by these results that the capacity improvement is almost as good as the capacity improvement that has to be made when considering single containers and collecting three times a week. The new overall capacity of 22492 is a major improvement when compared with the 28058 obtained in the previous section. This means that considering clusters is definitely worth it.

5.2.2 Three Times a Week

Considering the cluster *Rua D Ega* and like previously a period dataset with the time period of 2 hours, Figure 5.4 shows the simulation for three times a week frequency on Monday, Thursday and Saturday, the best days for collection as seen in the previous section:

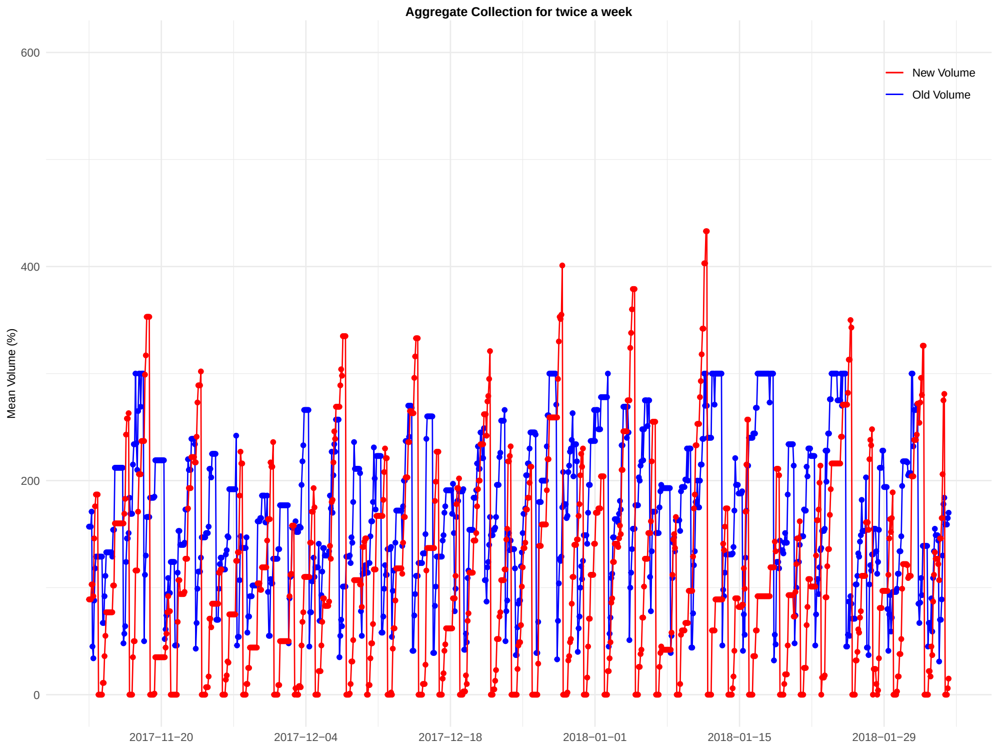


FIGURE 5.4: First Records of Three Times a Week Frequency for cluster *Rua D Ega*

The number of overloads in this example is immediately way less than when considering a twice a week frequency, as expected. The mean of the volume by cluster and container and the total number of waste overloads can be seen in Table 5.12.

The only cluster with a non residual number of overloads is *Arressario 40*, that is the cluster with only one container, the other ones seem to do well with a three times a week waste collection frequency. Running the capacity readjustment algorithm for this case, we can see in Table 5.13 that we almost don't need capacity improvement for clusters with more than one container.

Cluster	Mean V	N Containers	Mean V/Container	Overloads
Arco do Bispo 21	65	2	33	2
Arressario 60	57	2	29	0
Arressario 41	61	2	31	3
Arressario 23	72	2	36	2
D Ega	108	3	36	2
Praca de Camoes	76	3	25	0
Choes 34	82	3	27	2
Arressario 40	32	1	32	8
Mean	**69**	**2,25**	**31**	**2,4**

Table 5.12: Aggregate Three Times a Week Frequency Results

Container	Old Capacity	New Capacity
Arco do Bispo 21	2000	2270
Arressario 60	1600	1600
Arressario 41	1600	1816
Arressario 23	1600	1760
D Ega	3000	3150
Praca de Camoes	3000	3000
Choes 34	2600	2799
Arressario 40	800	1160
Total	**16200**	**17555**

Table 5.13: Capacity Readjustment Needed for Three Times a Week Frequency With Clusters

With these results we confirm the results of the previous section, that the waste collection frequency of three times a week is perfectly enough for our historical volumes data set.

5.2.3 Results Conclusion

In this section, we concluded that, when considering clusters of containers close together, the waste collection frequency of twice a week is a possible solution for the FCP with this set of volume historical data. Although this solution requires the addition of 10 containers to improve the clusters capacity, it's a solution that may require a much lower overall cost.

The waste collection frequency of three times a week proves to be the safest solution for the future, with more cost but almost no need for capacity improvement.

5.3 Multiple Time Periods

In this section, we analyse what happens in the summer regarding the waste collection frequency. As referred on Chapter 3 this is the period with less deposits of waste volume, so it may be useful to check the required waste collection frequency for this period alone.

5.3.1 Summer Collection Frequency

In this period, only containers 31450, 44263, 49619 and 50856 have a significant amount of data to analyse their waste collection frequency. We applied to these containers the volume simulation algorithm and the capacity readjustment algorithm for every day of the week, for summer data. The best day for waste collection is again Wednesday, similar to the results for the whole year. The figure 5.5 shows an example of the new volume model, side by side with the real volume:

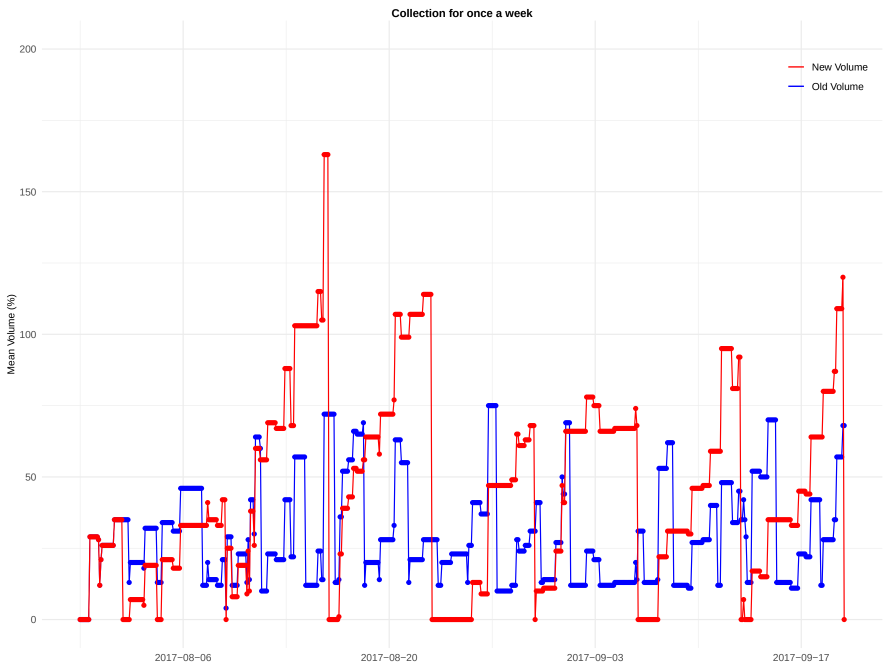


FIGURE 5.5: Summer Records of Once Week Frequency for container 49619

Table 5.14 shows the average of waste volume and waste overloads during summer, if the collection frequency was set to once a week.

Container	Average Volume	Nº of Overloads
31450	90	14
44263	37	9
49619	45	4
50856	92	1
Mean	**66%**	**7**

Table 5.14: Once a Week Frequency in Summer Results

As we can see these are much more acceptable results than the 95% average of waste volume seen for the whole year. Besides, the mean of overloads is much less, although we have to consider that there we are analyzing 1/4 of the period previously analyzed. The results from the capacity readjustment algorithm are shown in Table 5.15.

Container	Old Capacity	New Capacity
31450	800	1528
44263	1000	1520
49619	800	944
50856	800	936
Total	**3400**	**4928**

Table 5.15: Capacity Readjustment Needed for Waste Collection on Wednesday in Summer

We can see that containers 31450 and 44263 need an extra container to prevent waste overloads. As for containers 49619 and 50856, an upgrade from 800 litres to 1000 litres would be enough.

5.3.2 Results Analysis

Considering the model of waste collection frequency of twice a week using clusters, and with the addition of the 10 containers, we see that in summer, it's perfectly possible to consider a waste collection frequency of once a week. This is another improvement on the current waste collection frequency.

Chapter 6

Model Validation

In this chapter, we validate the solutions obtained in Chapter 5 by predicting how they behave in the future in terms of waste collections needs. We start by making predictions over the original data, in order to validate the prediction models themselves. Then, we apply the same predictions to the models created in order to have an idea of how they behave in the future.

6.1 Predictions

Using information of the season, weekday, precipitation and air temperature can provide good predictions on whether a container waste must be collected or not. To do so, we used data from the main data set and several datasets with fixed time periods. In both cases we considered that a container waste must be collected if the amount of waste in the container is higher than 60%. We pretend to compare the results between the original datasets.

6.1.1 Data Preparation

Data with categorical values, in this case season, day-of-week, precipitation and air temperature, are pre-processed where the number of columns is equal to the number of categories. The target (volume) is what we want to predict. More specifically we want to predict if a container has to be collected. To improve the performance and match the points of interest of the article, we transferred the values, which vary from 0% to 100%, to binary data. When the volume filled is inferior to 60%, is considered *Full* (in the sense that needs collection), otherwise is considered *Not Full*.

This data is prepared for each container. The summary of the data used for container 31450 can be seen in Table 6.1

Season		**WeekDay**		**Precipitation**		**Temperature**		**Volume**	
Fall	1092	Friday	504	Rainy Day	228	Mild Day	1212	Full	1269
Spring	960	Monday	504	Very Rainy Day	468	Cold Day	1008	Not Full	2235
Summer	372	Saturday	492	Day without Rain	2808	Very Cold Day	120		
Winter	1080	Sunday	492			Hot Day	1020		
		Thursday	504			Very Hot Day	144		
		Tuesday	504						
		Wednesday	504						

TABLE 6.1: Data for Machine Learing

The data is divided in two sets: the training set, with 80% of the data, and the validation set, with 20% of the data.

6.1.2 Prediction by container

Considering a dataset with volume values every 8 hours, for every container and using information about season, weekday, precipitation and temperature, we predicted if a container waste should be collected using, in this case, five algorithms: k-nearest neighbours (KNN), Latent Dirichlet Allocation (LDA), decision tree (cart) and random forest (RF). An example of the training results is presented in Figure 6.1

In this example, we can see that the random forest algorithms present the better results, with up to 80% of accuracy. With this, the random forest was the elected algorithm for the remaining tests. Although this is not a bad accuracy result, when making predictions to compare with the validation set, the accuracy of the predictions doesn't go further than 75%, for most of the containers data used.

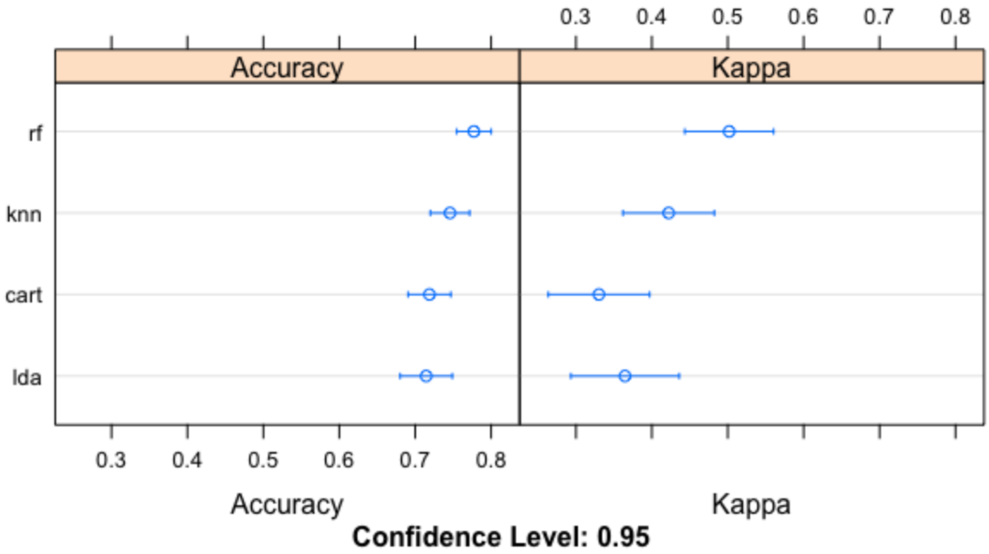


FIGURE 6.1: Training Results

Container/TP	6H	8H	12H	24H
15415	0,65	0,66	0,65	0,71
31450	0,76	0,71	0,8	0,74
41483	0,69	0,68	0,78	0,67
44263	0,91	0,91	0,92	0,9
44289	0,79	0,67	0,78	0,72
44776	0,69	0,64	0,76	0,77
44966	0,75	0,71	0,84	0,72
48843	0,67	0,71	0,73	0,61
49619	0,86	0,82	0,84	0,81
50419	0,92	0,92	0,92	0,93
50443	0,73	0,64	0,74	0,7
50708	0,76	0,66	0,75	0,6
50856	0,63	0,58	0,69	0,79
51698	0,64	0,58	0,67	0,68
52910	0,8	0,77	0,81	0,78
53181	0,66	0,62	0,75	0,65
54452	0,66	0,63	0,74	0,7
54494	0,7	0,66	0,71	0,62
Mean	**0.73**	**0.70**	**0.77**	**0.72**

TABLE 6.2: Prediction results by time period

All results are shown in Table 6.2. The mean of the accuracy obtained was around 70% for the datasets with volume values every 8 hours. The same algorithm was applied for datasets with time periods of 6 hours, 12 hours and a day. The mean accuracy obtained was 73%, 77% and 72% respectively. The results presented show that information like season, weekday, temperature and precipitation provide good predictions on whether a container waste must be collected or not.

6.1.3 Prediction by cluster

In this section we see how well predictions perform on clusters instead of individual containers. We check if the sum of the volume of the containers in the cluster divided by the number of containers in the cluster are above or below 60%. The algorithm used was again the random forest algorithm. The results can be seen in Table 6.4.

Cluster/TP	**6H**	**8H**	**12H**	**24H**
Arco do Bispo 21	0,65	0,55	0,62	0,83
Arressario 60	0,71	0,7	0,78	0,73
Arressario 41	0,82	0,76	0,82	0,71
Arressario 23	0,75	0,7	0,83	0,8
D Ega	0,74	0,7	0,67	0,81
Praca de Camoes	0,92	0,9	0,92	0,9
Choes 34	0,77	0,76	0,81	0,84
Arressario 40	0,76	0,66	0,75	0,6
Mean	**0.77**	**0.72**	**0.78**	**0.78**

TABLE 6.3: Prediction results by time period

We can see that there's no particular improvement when considering the clusters of containers instead of the containers individually. Still, these predictions have a pretty good accuracy to predict if a container or cluster should be collected or not. This is useful, in the following section, to validate the solutions obtained in Chapter 5, providing a way to study how they change in terms of the amount of waste volume, not only for the dates on the datasets but also to predict how they behave in the future.

6.2 Model Comparison

In this section, we compare the number of times we have to collect waste given the current frequency versus the new frequency of models obtained on chapter 5. If it's they are similar, it means that the new models are probably good solutions, not only for the current volume data but for the future too. This is always considering the 77% of average accuracy on the previous section.

For this study, we only consider the best solution obtained for this data, which is a waste collection frequency of twice a week and considering clusters of containers. We start by applying the same predictions to the new sets of data, obtained after running the volume simulation algorithm to time periods of 6h, 8h and 12h. The results on the predictions accuracy are shown in Table

Cluster/TP	**6H**	**8H**	**12H**
Arco do Bispo 21	0,81	0,85	0,78
Arressario 60	0,78	0,89	0,82
Arressario 41	0,84	0,84	0,8
Arressario 23	0,82	0,81	0,8
D Ega	0,89	0,88	0,86
Praca de Camoes	0,89	0,92	0,92
Choes 34	0,89	0,9	0,87
Arressario 40	0,84	0,88	0,82
Mean	**0,85**	**0,87**	**0,83**

Table 6.4: Prediction results by time period

The accuracy obtained for the new data is better than in the original data. This happens because in practice, the amount of times a cluster is consider full is even less than in the original data, due to the capacity improvement in this model. Table 6.5 shows the difference between the original data and the data from the model with two times a week collection frequency.

	Original Data		New Data	
Clusters	**Full**	**Not Full**	**Full**	**Not Full**
Arco do Bispo 21	275	317	173	419
Arressario 60	250	586	202	634
Arressario 41	217	695	221	691
Arressario 23	271	545	292	524
D Ega	742	2	265	479
Praca de Camoes	69	731	83	717
Choes 34	254	650	170	734
Arressario 40	317	603	233	687
Total	**2395**	**4129**	**1639**	**4885**

Table 6.5: Amount of Full Containers in Original and New Models

These values present another good argument to have a waste collection frequency of two times a week, as they indicate that this model is not only better for the current data, but it's likely to be better for future waste collections.

Chapter 7

Conclusions

In this final chapter, we provide a summary of what was studied in this dissertation. We start by revisiting the research questions made in Chapter 1. Next we talk about the major findings in this study. Lastly, we draw some conclusions and talk about future work.

7.1 Questions Answered

Going back to the research questions made in Chapter 1, we can see in this dissertation the answer to almost all of them.

Regarding question 1, we saw in Chapter 3 and 5 that this question could be answered considering critical points in each container and the frequency needed for each container individually to avoid waste overload. For this set of waste volume historical data, the containers 15415, 44776, 53181 and 54452 have the most critical points. On the other hand, containers 41483, 44289 and 48832 need more than twice a week collection for the best overall pair of days. So yes, with the volume data of waste in the containers, it's possible to know which containers require more or less frequent collection.

As mention in Chapter 3, Evox, the company that provided the data sets for this study, already use the volume data for collection routes optimization in real-time, answering to question 2.

Although this is not something yet addressed in the literature, on Chapter 6 it was shown that we could predict with an accuracy of around 78% if a container waste volume is above 60%, using information like season, weekday, air temperature and precipitation.

Question 4 was not answered in this dissertation. We saw on Chapter 6 that the waste deposit volume is less in the summer, where probably the population density is less, but to truly answer this question, we would need a much larger set of containers and information on families income.

The capacity readjustment algorithm presented in Chapter 4 and applied in Chapter 5 solve Question 5.

Question 6 is solved using the volume simulation algorithm combined with the capacity readjustment algorithm combined. This was done in Chapter 5 and for this set of data and considering individual containers, the best collection frequency is three times a week with the addition of two containers and the upgrade from 800 litres to 1000 litres of six containers.

Question 7 is also solved in Chapter 5. When considering clusters of containers next to each other, we saw that a waste collection frequency of twice a week with the capacity improvement of adding ten containers is an optimal solution for these set of volume data. This is a major improvement over the current three to four times week collection in Castelo Branco.

7.2 Major Findings

In this study, it was shown that grouping the containers by streets, the monthly average volume was always between 30% to 60% and the average volume of deposits were never above 20% of waste a day. On the other hand, the waste collection was, in most cases, done wrong, with a high number of needless collections and critical points. It was also found that the classes weekday, precipitation and air temperature had a week correlation with the volume data while season had a strong correlation.

Using machine learning algorithms, we predicted if a container waste has to be collected or not with a 78% accuracy, just using the information on season, weekday, air temperature and precipitation. These predictions can be used to propose more complex models where the waste collection frequency varies by season.

We propose three different solutions of frequency-capacity. The first proposal was a waste collection frequency of one time a week. For this model, we saw that almost every container had an average waste volume over 95% which shows that a frequency of once a week is not enough for this case. The second proposal was a waste collection frequency of twice a week considering individual containers. This model shows a great improvement over the previous one with an average waste volume over 46% volume by container, but it needed an increase of the capacity of the overall container by almost the double of the current capacity. The third proposal was a waste collection of three times a week. We saw that the average volume was around 30% and this frequency requires addition of two containers and the upgrade of six containers from 800 litres to 1000 litres, but it provides a better solution than the current solution of three to four times a week. The last proposal was a frequency of two times a week considering clusters of containers next to each other. We saw that if when a container is full, a person uses another container of the cluster, to prevent waste overloads it would only need to be added ten containers to the current ones. On top of that, we saw that the collection frequency may be reduced to once a week in summer. With predictions, it was shown that this solutions is likely to do well in the future, due to having less probability of having containers with waste volume above 60%. Because the cost of collecting is much greater over time than the cost of adding containers, this is a major improvement over the current three to four times week collection in Castelo Branco, making this a successful capacity-frequency solution for our containers.

7.3 Conclusion

In this dissertation we addressed the waste collection process with a different approach by studying the capacity-frequency problem.

It was possible analyze waste deposition volume and to identify patterns for a determinist and uniform waste collection. For this case, we concluded that a uniform collection of twice a week, with small improvements in containers capacity, proved to be enough for these containers, which is a major improvement to the current collection frequency of three to four times a week.

This process, composed mostly by the volume simulation algorithm and the capacity readjustment algorithm is easy to implement for other sets of data because the processes to generate the model's new volume data and capacities required are scalable, so it's easy to apply this study for other use cases.

For future work, we pretend to use more information to propose mixed different waste collected frequencies by season, month or other periods of time, providing an automated dynamic calculation of frequency-capacity solutions.

Bibliography

[1] Pike research. https://smartcitiescouncil.com/tags/pike-research, 2011. Accessed: 2019-04-22.

[2] Evox 360 waste. https://www.360waste.pt, 2018. Accessed: 2019-06-06.

[3] Evox 360 waste dashboard. https://www.360waste.pt/dashboardpt, 2018. Accessed: 2019-06-03.

[4] Hannele Ahvenniemi, Aapo Huovila, Isabel Pinto-Seppä, and Miimu Airaksinen. What are the differences between sustainable and smart cities? *Cities*, 60:234 – 245, 2017.

[5] T. Anagnostopoulos, A. Zaslavsky, and A. Medvedev. Robust waste collection exploiting cost efficiency of iot potentiality in smart cities. In *2015 International Conference on Recent Advances in Internet of Things (RIoT)*, pages 1–6, April 2015.

[6] T. Anagnostopoulos, A. Zaslavsy, A. Medvedev, and S. Khoruzhnicov. Top – k query based dynamic scheduling for iot-enabled smart city waste collection. In *2015 16th IEEE International Conference on Mobile Data Management*, volume 2, pages 50–55, June 2015.

[7] Joaquín Bautista, Elena Fernández, and Jordi Pereira. Solving an urban waste collection problem using ants heuristics. *Computers And Operations Research*, 35(9):3020 – 3033, 2008. Part Special Issue: Bio-inspired Methods in Combinatorial Optimization.

[8] Xiaoyun Bing, Jacqueline M. Bloemhof, Tania Rodrigues Pereira Ramos, Ana Paula Barbosa-Povoa, Chee Yew Wong, and Jack G.A.J. van der Vorst. Research challenges in municipal solid waste logistics management. *Waste Management*, 48:584 – 592, 2016.

[9] Stephen J. Burnley. A review of municipal solid waste composition in the united kingdom. *Waste Management*, 27(10):1274 – 1285, 2007. Wascon 2006 6th International Conference: Developments in the re-use of mineral waste.

[10] Andrea Caragliu, Chiara Del Bo, and Peter Nijkamp. Smart cities in europe. *Journal of Urban Technology*, 18(2):65–82, 2011.

[11] Manuel Castells. Urban sustainability in the information age. *City*, 4(1):118–122, April 2000.

[12] Renata Dameri and A Cocchia. Smart city and digital city: Twenty years of terminology evolution. *X Conference of the Italian Chapter of AIS, ITAIS2013*, pages 1–8, 01 2013.

[13] G. Ghiani, D. Laganà, E. Manni, R. Musmanno, and D. Vigo. Operations research in solid waste management: A survey of strategic and tactical issues. *Computers and Operations Research*, 44:22 – 32, 2014.

[14] Jose M. Gutierrez, Michael Jensen, Morten Henius, and Tahir Riaz. Smart waste collection system based on location intelligence. *Procedia Computer Science*, 61:120 – 127, 2015. Complex Adaptive Systems San Jose, CA November 2-4, 2015.

[15] Peter Hall. Creative cities and economic development. *Urban Studies*, 37(4):639–649, 2000.

[16] Gerhard P. Hancke, Bruno de Carvalho e Silva, and Gerhard P. Hancke, Jr. The role of advanced sensing in smart cities. *Sensors*, 13(1):393–425, 2013.

[17] Perinaz Hoornweg, Daniel; Bhada Tata. *What a Waste : A Global Review of Solid Waste Management*. 2012.

[18] Nikolaos Karadimas, Katerina Papatzelou, and Vassili Loumos. *Genetic Algorithms for Municipal Solid Waste Collection and Routing Optimization*, volume 247, pages 223–232. 09 2007.

[19] Alexey Medvedev, Petr Fedchenkov, Arkady Zaslavsky, Theodoros Anagnostopoulos, and Sergey Khoruzhnikov. Waste management as an iot-enabled service in smart cities. In Sergey Balandin, Sergey Andreev, and Yevgeni Koucheryavy, editors, *Internet of Things, Smart Spaces, and Next Generation Networks and Systems*, pages 104–115, Cham, 2015. Springer International Publishing.

[20] Zhu Minghua, Fan Xiumin, Alberto Rovetta, He Qichang, Federico Vicentini, Liu Bingkai, Alessandro Giusti, and Liu Yi. Municipal solid waste management in pudong new area, china. *Waste Management*, 29(3):1227 – 1233, 2009.

[21] M. Naphade, G. Banavar, C. Harrison, J. Paraszczak, and R. Morris. Smarter cities and their innovation challenges. *Computer*, 44(6):32–39, June 2011.

[22] Teemu Nuortio, Jari Kytöjoki, Harri Niska, and Olli Bräysy. Improved route planning and scheduling of waste collection and transport. *Expert Systems with Applications*, 30(2):223 – 232, 2006.

[23] S. Pellicer, G. Santa, A. L. Bleda, R. Maestre, A. J. Jara, and A. G. Skarmeta. A global perspective of smart cities: A survey. In *2013 Seventh International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, pages 439–444, July 2013.

[24] Mufeed Sharholy, Kafeel Ahmad, Rakesh Chandra Vaishya, and Rana Datta Gupta. Municipal solid waste characteristics and management in allahabad, india. *Waste management*, 27 4:490–6, 2007.

[25] F. Vicentini, A. Giusti, A. Rovetta, X. Fan, Q. He, M. Zhu, and B. Liu. Sensorized waste collection container for content estimation and collection

optimization. *Waste Management*, 29(5):1467 – 1472, 2009. First international conference on environmental management, engineering, planning and economics.

[26] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi. Internet of things for smart cities. *IEEE Internet of Things Journal*, 1(1):22–32, Feb 2014.