NOVA IMS
Information Management School

MGI

Mestrado em Gestão de Informação
Master Program in Information Management

WASTE PRODUCTION PROFILING USING SOCIOECONOMIC FACTORS

A CASE STUDY OF LISBON

Marta Marinho Amarante Malta Dias

Project Work presented as partial requirement for obtaining the Master’s degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

WASTE PRODUCTION PROFILING USING SOCIOECONOMIC FACTORS: A CASE STUDY OF LISBON

by

Marta Marinho Amarante Malta Dias

Project Work presented as partial requirement for obtaining the Master's degree in Information Management with a specialisation in Knowledge Management and Business Intelligence.

Advisor: Professor Doutor Miguel de Castro Simões Ferreira Neto

Co Advisor*:* Doutor Pedro Alexandre Reis Sarmento

November 2022

DEDICATION

This work closes a chapter in my life and the end of my academic journey at NOVA IMS, where I was able growth my knowledge as well as a person, and I believe it will give me opportunities to work in this area, which fascinates me more and more by the day.

I want to thank everyone that directly and indirectly contributed to my master's dissertation, particularly the People that work in Câmara Municipal de Lisboa and Instituto Nacional de Estatística for making available the data I worked with.

In particular, I would like to thank Professor Doutor Miguel de Castro Simões Ferreira Neto and Doutor Pedro Alexandre Reis Sarmento for their dedication and availability throughout this project.

To my family, I want to thank them for the support they have always given, and to my partner for his patience, affection, motivation and support.

ABSTRACT

Urban Waste Management is a public service provided by municipalities, managed by both private and public companies. Ensuring good waste management in municipalities provides the quality of waste collection service, meeting environmental targets and offering clean, healthy, and safe cities.

Municipal waste management is already a topic of concern in recent decades in Portugal. From the literature analysed, there is ample evidence about the legislation and policies implemented to achieve better waste management and increase recycling to promote the country's sustainable development.

This study aims to classify the socioeconomic profile of producers of various types of waste in Lisbon, the capital of Portugal. That is, to find the characteristics of the different groups of people who produce waste, and the variables that can better describe these groups, to understand whether their social and economic characteristics impact the amount of waste produced.

The methodology to be implemented to develop the study is the application of clustering techniques to identify the characteristics of the groups of people who produce waste in the city of Lisbon. It is expected to use clustering techniques, such as k-means and Dynamic Time Warping.

Was concluded that in mixed waste, people over 65 years of age, with a low level of education and whose income is also low, are the ones who produce the most waste, followed by the youngest between 0 and 24 years of age, with a high level of education and with better incomes. As for recycled waste, was concluded, in general, that residents with a higher level of education and better incomes produce more recycled waste, unlike those with less education level and lower incomes that produce less.

KEYWORDS

Municipal Waste Management; Socioeconomic; Waste Generation; Waste Collection; Clustering; Dynamic Time Warping.

INDEX

1. Introduction 1
   1.1. Study Objective 3
   1.2. Problem Framing 4
2. Literature review 6
   2.1. Variables Study 6
   2.2. Waste in Portugal 7
   2.3. Waste Collection Circuits 8
      2.3.1. Waste Management Integrated System in Lisbon 8
      2.3.2. Waste Process 9
      2.3.3. Waste disposal and collection system 10
3. Methodology 12
4. Conceptual Model 13
   4.1. Model for Waste Production Profiling 13
      4.1.1. Data Understanding 13
      4.1.2. Data Preparation 15
      4.1.3. Modelling 17
      4.1.4. Evaluation 19
   4.2. Socioeconomic and Context Variable Association 20
5. Results and discussion 22
   5.1. Data Preparation 22
      5.1.1. Mixed Waste 22
      5.1.2. Recycled Waste 24
   5.2. Modelling 29
      5.2.1. Mixed Waste 29
      5.2.2. Recycled Waste 33
   5.3. Evaluation 43
      5.3.1. Mixed Waste 43
      5.3.2. Recycled Waste 47
6. Conclusions 60
7. Bibliography 62

8. Annexes 67
8.1. Annex I 67
8.2. Annex II – Data Preparation 72
8.3. Annex III – Data Modelling 78
8.4. Annex IV – Evaluation 80

LIST OF FIGURES

Figure 1 – Organization of service by supporting zones (Município de Lisboa, 2020) 9
Figure 2 – Eco-islands in the Municipality of Lisbon (Município de Lisboa, 2020) 10
Figure 3 – Waste Collection Points and each location (Source Lisboa Aberta (*Circuitos Contentores*, 2021)) 15
Figure 4 - Mixed Waste Time-Series result after data cleaning 23
Figure 5 - Plastic Waste Time-Series result after data cleaning. 26
Figure 6 - Paper Waste Time-Series result after data cleaning 28
Figure 7 – Cluster Constellation Plot for Mixed Waste 29
Figure 8 – Cluster Dendrogram for Mixed Waste 30
Figure 9 – Elbow Graph for Mixed Waste 30
Figure 10 – Segment Size, *k*=3 and *k*=4 (Mixed Waste) 31
Figure 11 – Linear Plot by Segment *k*=3 (Mixed Waste) 32
Figure 12 – Linear Plot by Segment *k*=3 after normalization (Mixed Waste) 33
Figure 13 – Cluster Constellation Plot for Plastic Waste 33
Figure 14 – Cluster Dendrogram for Plastic Waste 34
Figure 15 – Elbow Graph for Plastic Waste 34
Figure 16 – Segment Size, *k*=4 and *k*=5 (Plastic Waste) 35
Figure 17 – Linear Plot by Segment *k*=5 (Plastic Waste) 36
Figure 18 – Linear Plot by Segment *k*=5 after normalization (Plastic Waste) 36
Figure 19 – Cluster Constellation Plot for Paper Waste 37
Figure 20 – Cluster Dendogram for Paper Waste 37
Figure 21 – Elbow Graph for Paper Waste 37
Figure 22 – Segment Size, *k*=3 and *k*=4 (Plastic Waste) 38
Figure 23 – Linear Plot by Segment *k*=3 (Paper Waste) 39
Figure 24 – Linear Plot by Segment *k*=3 after normalization (Paper Waste) 39
Figure 25 – Cluster Constellation Plot for Glass Waste 40
Figure 26 – Cluster Dendogram for Glass Waste 40
Figure 27 - Elbow Graph for Glass Waste 40
Figure 28 – Segment Size, *k*=3 and *k*=4 (Glass Waste) 41
Figure 29 – Linear Plot by Segment *k*=4 (Glass Waste) 42
Figure 30 – Linear Plot by Segment *k*=4 after normalization (Glass Waste) 42

Figure 31 – PowerBI Dashboard 3 Clusters, Mixed Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow). 43
Figure 32 – PowerBI Dashboard 5 Clusters and an Outlier, Plastic Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow, Cluster 4 – orange, Cluster 5 – Red, Circuit EESA01MF – light blue). 47
Figure 33 – PowerBI Dashboard 3 Clusters and an Outlier, Paper Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow, Circuit P3101 – Red). 52
Figure 34 – PowerBI Dashboard 4 Clusters and an Outlier, Glass Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow, Cluster 4 – red, Circuit V1305 – light blue). 56
Figure 35 – Undifferenciated Waste Time-Series result before data cleaning 72
Figure 36 - Plastic Waste Time-Series result before data cleaning. 73
Figure 37 - Paper Waste Time-Series result before data cleaning 74
Figure 38 - Glass Waste Time-Series result before data cleaning 75
Figure 39 - Glass Waste Time-Series result after data cleaning 76
Figure 40 –Segment Size Mixed Waste, K=2, K=3, K=4, K=5, K=6, respectively 78
Figure 41 –Segment Size Plastic Waste, K=2, K=3, K=4, K=5, K=6, respectively 78
Figure 42 –Segment Size Paper Waste, K=2, K=3, K=4, K=5, K=6, respectively 79
Figure 43 –Segment Size Glass Waste, K=2, K=3, K=4, K=5, K=6, respectively 79
Figure 44 – PowerBI Mixed Waste, Cluster 1, 2 and 3 83
Figure 45 - Linear Plot by Segment *k*=5 and EESA01MF Plastic Waste. 87
Figure 46 - PowerBI Plastic Waste, Cluster 1, 2, 3, 4, 5 and EESA01MF 89
Figure 47 - Linear Plot by Segment *k*=3 and P3101 Paper Waste. 92
Figure 48 - PowerBI Paper Waste, Cluster 1, 2, 3, and P3101 93
Figure 49 - Linear Plot by Segment *k*=4 and V1305 Glass Waste. 96
Figure 50 - PowerBI Glass Waste, Cluster 1, 2, 3, 4 and V1305 97

LIST OF TABLES

Table 1 - Municipal waste collected per inhabitant (kg/inhabitant) by Geographical location; Annual (Adapted INE - National Institute of Statistics (*INE*, 2020) 4
Table 2 – Summary table of Databases and Data Sources... 14
Table 3 - Example circuit I0102 from mixed waste agregated by week for 22 weeks (see Table 24 for full table).. 16
Table 4 - Variables that better influence waste generation 21
Table 5 - Comparison between the average of mixed waste circuits. 23
Table 6 - Variable description and classification 24
Table 7 - Missing values description for Mixed Waste 24
Table 8 - Comparison between the average of plastic waste circuits. 25
Table 9 - Missing values description for Plastic Waste 27
Table 10 - Comparison between the average of paper waste circuits. 27
Table 11 - Comparison between the average of all glass waste circuits. 29
Table 12 - Number of circuits in each segment, *k*=3 and *k*=4 (Mixed Waste) 31
Table 13 - Number of circuits in each segment, *k*=4 and *k*=5 (Plastic Waste)........................ 35
Table 14 - Number of circuits in each segment, *k*=3 and *k*=4 (Paper Waste)......................... 38
Table 15 - Number of circuits in each segment, *k*=3 and *k*=4 (Glass Waste).......................... 41
Table 16 - Summary table Mixed Waste 43
Table 17 - Amount per each Variable Mixed Waste, Z-Score 44
Table 18 - Summary table Plastic Waste.. 47
Table 19 - Amount per each Variable Plastic Waste, Z-Score... 48
Table 20 - Summary table Paper Waste... 52
Table 21 - Amount per each Variable Paper Waste, Z-Score.. 52
Table 22 - Summary table Glass Waste.. 56
Table 23 - Amount per each Variable Glass Waste, Z-Score... 56
Table 24 - Example circuit I0102 from mixed waste agregated by week for 52 weeks........... 67
Table 25 – Socioeconomic and Context Variables 68
Table 26 – Plastic circuit considered as an outlier 77
Table 27 – Paper circuit considered as an outlier 77
Table 28 – Glass circuit considered as an outlier 77
Table 29 – Amount per each Variable Mixed Waste, Z-Score, full table. 80

Table 30 - Amount per each Variable Plastic Waste, Z-Score, full table................................. 84
Table 31 - Amount per each Variable Paper Waste, Z-Score, full table.................................. 90
Table 32 - Amount per each Variable Glass Waste, Z-Score, full table................................... 94

LIST OF ABBREVIATIONS AND ACRONYMS

CO2	Carbon Dioxide
INE	National Institute of Statistics
EU	European Union
GIS	Geographic Information Systems
WEEE	Wasted Electrical and Electronic Equipment
CDW	Construction and Demolition Waste
ELVs	End of Life Vehicles
SCE	Sorting Centre and Ecocenter
MSWTC	Municipal Solid Waste Treatment Centre
STRF	Slag Treatment and Recovery Facility
SSE	Sum of Squared Errors
SPV	Sociedade Ponto Verde
GDP	Gross Domestic Product
HDI	Human Development Index
DTW	Dynamic Time Warping
TS	Time Series

1. INTRODUCTION

Urban Waste Management is a public service provided by municipalities, managed by both private and public companies. According to Beigl (2008), the amount and composition of waste generated prove to be essential information on the planning, operation, and optimisation of the waste management process.

Ensuring good waste management in municipalities provides the quality of waste collection service, meeting environmental targets and offering clean, healthy, and safe cities.

Maity (2018) indicates that it is increasingly important to carry out good waste management, an integral element of modern society. It emphasises, among others, several consequences of poor or inefficient waste management, such as: representing a threat to the environment and a market value decreased in the affected areas, in addition to the fact that it represents a significant risk to public health. Having said that, and since waste is present in the day-to-day life of modern society, it is inevitable to ensure its proper functioning for a sustainable future.

Different authors agreed that urban waste management is not only related to money and equipment problems but also with societies' cultural and social characteristics, beliefs and relationships. It is directly connected to daily human activities, and it tends to be very local (Namlis & Komilis, 2019; Oribe-Garcia et al., 2015; Vieira & Matheus, 2018)

With cities development and modernisation, different technologies provided mechanisms to face this challenge and brought new resources that produce more waste. As more developed a society is, with more income and a higher standard of living, more resources are used, and more waste tends to be produced. In that sense, an efficient waste management system must be developed by those societies to not compromise their safety and health. This kind of service is provided by both developed and developing countries' municipal authorities (Khan et al., 2016; Maity, 2018)

Several entities provide the exact expectations concerning the future of waste. According to The World Bank and Statista (*Trends in Solid Waste Management*, 2021; Tiseo, 2020), in 2016 were produced 2.02 billion tonnes of municipal waste worldwide, where 33% were not managed safely. On average, 0.74 kilograms of waste are made every day per person and range from 0.11 to 4.54 kilograms.

In 2030 it is expected that the total waste will reach 2.59 billion tonnes and in 2050, 3.4 billion tonnes. It is also expected that in low and middle-income countries, waste will increase by 40% since it is likely that the population of those countries will double or triple. This is due to the increase of the population, urban and economic development, and the rise in living standards of developing countries (Abdel-Shafy & Mansour, 2018). The research also says that high-income countries collect at least 90% of waste in the opposite of low-income countries, that only collect about 30%. Moreover, a critical step to managing waste is waste collection.

According to The World Bank Group (*Trends in Solid Waste Management*, 2021), global waste comprises food and green waste, representing 44% of the waste produced, followed by paper and cardboard with 17% and plastic with 12%, representing 73% of the global waste composition. Based on waste generation, design and management, it is estimated to generate 1.6 billion tonnes of carbon dioxide (CO_2), or 5% global emissions in 2016, where food waste represents nearly 50% of those emissions. Emissions which are expected to increase to 2.32 billion tonnes of CO_2 if waste-related issues are not anticipated per year by 2050.

There are several studies related to waste management; it is not a new topic. Nonetheless, policy authorities need help to deal with globalisation, population increase, and resource consumption.

Municipal waste management is already a topic of concern in recent decades in Portugal. From the literature analysed, there is ample evidence about the legislation and policies implemented to achieve better waste management and increase recycling to promote the country's sustainable development. According to Magrinho (2006), European Directives influenced the development of municipal waste management, with incentives and guidelines applied in the portuguese legislation.

Other literature shows that socioeconomic and demographic factors such as education, employability, income, gender, population density, buildings, age, and others are essential when analysing waste production and are the variables used to present different models.(Kannangara et al., 2018; Niska & Serkkola, 2018; Oribe-Garcia et al., 2015; Soukiazis & Proença, 2020; Talalaj & Walery, 2015).

Waste Management, as mentioned before, is not only related to equipment or money. It is also associated with cultural and social characteristics in a specific region. To collect and analyse such data, we can resort to open-source data, that sometimes is sparse or difficult to find and with low accuracy (Niska & Serkkola, 2018) Fortunately, Industry 4.0 provided the development of digital technologies and information that contribute to economic changes and new trends in cities. Open-source data contains increasingly more information that can be analysed that contributes to decision-making. In this case, open-source data is fundamental to provide future analysis related to waste management because society's registries are necessary (Neves et al., 2020; Niska & Serkkola, 2018).

In this way, open-source data can be used to analyse and forecast waste generation in time in a specific region accurately.

1.1. Study Objective

This study aims to classify the socioeconomic profile of producers of various types of waste in Lisbon, the capital of Portugal. That is, to find the characteristics of the different groups of people who produce waste, and the variables that can better describe these groups, to understand whether their social and economic characteristics impact the amount of waste produced. Besides that, this study pretends to identify how the production of waste of those groups evolve during the year, to identify waste production patterns along time.

Thus, it allows the Municipality of Lisbon to reassess its municipal waste management strategies, implementing more objective measures in the different waste collection circuits, depending on the characteristics identified in the different groups and their location.

This study allows for a long-term approach to planning by the Lisbon City Council, as the characteristics of the population vary little over time.

To this end, the following objectives were identified:

- Define the type of waste to be studied;
- Choose the most relevant variables to define the waste production profile;
- Identify the best methodology to be implemented in the case study;
- Understand the relationship between the socioeconomic profile and the amount of waste produced;
- Describe possible solutions and/or strategies that improve the waste management strategy implemented, depending on the identified groups and their location.

1.2. PROBLEM FRAMING

According to Statistics Portugal (*INE*, 2020), Portugal registered 474 kg/inhabitant waste production in 2016, and the trend recorded in the following years was always higher, reaching 514 kg/inhabitants in 2019. Which comparing with The European Union, in the same period, registered a total of 489 kg/inhabitants in 2018 (*Gestão de resíduos na UE*, 2020), much lower than that recorded in Portugal in the same period, 507 kg/inhabitants.

In the Metropolitan Area of Lisbon were produced on average, 536 kg/inhabitant in 2019. The highest values of waste production were recorded for the municipality of Palmela with 715 kg/inhabitant and Cascais with 670 kg/inhabitant. Furthermore, Lisbon recorded 666 kg/inhabitants, also for the year 2019, as shown in the table below:

Table 1 - Municipal waste collected per inhabitant (kg/inhabitant) by Geographical location; Annual (Adapted INE - National Institute of Statistics (*INE*, 2020)

Geographic Location	Data reference period			
	2019	2018	2017	2016
	kg/ inhab.	kg/ inhab.	kg/ inhab.	kg/ inhab.
Portugal	514	507	486	474
Continent	512	505	484	472
Lisbon Metropolitan Area	536	535	516	499
Alcochete	584	555	501	521
Almada	601	606	594	576
Amadora	413	409	402	371
Barreiro	469	454	424	429
Cascais	670	675	642	636
Lisboa	666	664	635	621
Loures	421	434	450	421
Mafra	561	556	518	488
Moita	501	522	513	516
Montijo	499	450	432	460
Odivelas	x	x	x	x
Oeiras	458	466	429	438
Palmela	715	670	629	631
Seixal	501	477	467	422
Sesimbra	669	673	636	651
Setúbal	623	631	644	565
Sintra	469	471	449	429
Vila Franca de Xira	405	404	362	361

A waste management plan requires data related to its production, the factors that influence it, and ways to predict the amount of waste produced to ensure the proper functioning of services and avoid consequences for the future of municipalities, such as the degradation of public health (Lebersorger & Beigl, 2011).

It is possible to identify the main population characteristics related to the amount of waste produced to help create or remodel municipal waste management strategies using clustering techniques to combine data from waste collection circuits with different socioeconomic and demographic data sources such as Census (*INE – Censos 2021*, 2021), infrastructure context (counting of school establishments, restaurants, tourist establishments, among others) (*Lisboa Aberta*, 2021) and price of sale of housing (*Confidencial Imobiliário*, 2022).

Using information collected through the Census (variables), infrastructure context and price of sale of housing will allow building a model adapted to any municipality in the country, given the consistency in managing and treating socioeconomic data.

2. LITERATURE REVIEW

2.1. VARIABLES STUDY

The literature found that there were made several studies related to waste management because this is a living problem, where policy authorities' have more challenges trying to decide the best practices and strategies to overtake this problem or to turn it sustainable for future generations.

Most of the studies use data analytics approaches to use models or create models to predict future waste generation to be previously controlled (Beigl et al., 2008; Kannangara et al., 2018; Oribe-Garcia et al., 2015).

Beigl (2008) proposed an implementation guideline where 45 models were studied. The published methodologies revealed a high heterogeneity of the models, although the issues to be solved were very similar. Beigl also indicates that regression and correlation analysis are helpful in most cases. Still, group comparisons and cluster criteria were beneficial because they allow testing the relationship between waste generation and its effects on recycling quotas or the level of affluence.

Oribe-Garcia (2015) developed two models, one for the whole province of the city of Biscay and the second dividing Biscay in groups with similar characteristics (clustering). He concluded that there were very heterogeneous observations without the grouping, and the second model was tested to prove that clustering analysis allows suggesting models with higher forecasting abilities. So previous clustering analysis for very heterogenous data provides better outcomes for posterior prediction analysis with more accurate results.

In that sense, clustering analysis was the approach for this study. Since that, the objective is to classify the socioeconomic profile of producers of mixed, glass, plastic and paper and cardboard waste, to find the different groups of waste producers and find possible solutions for waste mitigation according to their characteristics. Nevertheless, it will also provide a solid base for posterior prediction analysis of waste generation.

The literature shows that one of the most challenging parts of the research was identifying the variables. It also shows that the country and its development have a more significant impact on the chosen variables (Khan et al., 2016; Namlis & Komilis, 2019; Vieira & Matheus, 2018).

GDP (Gross Domestic Product) is an instrument commonly used in several economic and public policies studies, mainly because it is easy to calculate and provides a clear and broad perception of a country's economic and development level (Van den Bergh, 2009). This indicator is also used in studies to forecast waste generation because it reflects changes in the population's well-being (Giannakitsidou et al., 2016).

Olga Giannakitsidou (2016) studied the possibility to substitute this indicator with another multidimensional index such as HDI (Human Development Index)1. She concluded that although GDP

[1] Human Development Index is calculated based on life expectancy index, education index and income per capita.

provides an incomplete picture of a nation's state, it remains the most appropriate index to predict future municipal waste management. However, it suggests also that HDI is fundamental to express cultural and social aspects, which, as mentioned before, are crucial for waste management politics.

Studies considering the European countries, considered as developed countries, determined that GDP, a proxy of income, explains countries' waste generation and that GDP and HDI are correlated considering waste streams. Most of the studies mentioned earlier have shown that income positively affects solid waste generation, especially in the secondary and tertiary sectors, where more resources are allocated and more revenue is generated (Soukiazis & Proença, 2020; Vieira & Matheus, 2018). However, Namlis and Komilis (2019) studied the influence of socioeconomic indices in an economic crisis on solid waste generation. He concluded that Portugal, highly affected by the crisis, produced more waste in specific waste streams, although the income was diminished.

Regarding other socioeconomic factors that could influence a waste generation, the literature indicated: citizen's age, household size (m2), household surface area (m2), household living space (m2), and employment rate (%) (Niska & Serkkola, 2018); family size, education, health, and inequality (Khan et al., 2016; Vieira & Matheus, 2018). Some of them have more influence regarding waste generation than others, and it is mainly related to the level of social development and economic status. For instance, family member numbers significantly impact the waste generation rate in developing countries.

Niska and Serkkola (2018) collected information on container level monitoring in Helsinki, Finland, collecting data on the weight and quantity of waste from containers from public data, combining this information with socioeconomic data. It was also used time-series values that reflect the necessary time-series properties of the study.

Data from waste collection circuits can be combined with the socioeconomic and demographic information of the population. This is because each circuit has a different destination and covers specific city areas and different parishes, which can demonstrate social and economic differences.

However, according to the same study, few demographic and socioeconomic variables were used to define the main population characteristics that influence the production of larger and smaller amounts of waste.

2.2. WASTE IN PORTUGAL

Municipal waste management is already a topic of concern in recent decades in Portugal. From the literature analysed, there is ample evidence about the legislation and policies implemented to achieve better waste management and increase recycling to promote the country's sustainable development. According to Magrinho (Magrinho et al., 2006), European Directives influenced the development of municipal waste management, with incentives and guidelines applied in Portuguese legislation.

According to the United Nations (UNFCCC, 2019), 8% of mitigation actions in developed countries corresponds to waste management. One of the action support trends is to enhance waste management practices at the local level, which leads directly to co-benefits for local communities.

In 2016, waste represented 5% of global greenhouse gas emissions managed efficiently by the local authorities.

The European Directive 2008/98/EC defines Waste Management as the "collection, transport, recovery and disposal of waste, including the supervision of such operations and the after-care of disposal sites, including actions taken as a dealer or broker" (2008, p. 7).

The Directive also defines that the Member States must establish general environmental objectives and frameworks to prevent and reduce the sources of pollution for the management of waste within the Community. So, every Member State must align its strategies to accomplish the Communities objectives. However, they are free to decide the plan that best fits their society.

Reducing the amount of waste generated and, when unavoidable, promoting it as a resource and achieving higher levels of recycling and safe disposal of waste is the aim of EU policies (Waste Statistics, 2018).

The Eurostat shows that in 2018 Portugal is the country that generated more waste in the category of household waste compared to the European Union countries, with 32,8% and 9,5% of total waste, respectively (Waste Statistics, 2018). Household waste is defined by the wastes strictly produced at home and is collected in a mixed way (Oribe-Garcia et al., 2015)

2.3. Waste Collection Circuits

The higher the economic growth and the more consuming society is, the higher the solid waste generation is expected to be due to their capacity to consume more and greater availability of products.(Fasihi & Parizadi, 2021; Giannakitsidou et al., 2016; Khan et al., 2016; Vieira & Matheus, 2018).

As mentioned before, data from waste collection circuits are essential for urban waste management. Moreover, to improve analytics, Geographic Information Systems (GIS) were implemented. In past years, they have been a vital resource to represent information of the geographical space and a base to perform analysis. GIS optimises routes for collection and management of circuits, integration between alphanumeric and geographical information, automatic calculation of routes, and information availability via intranet/internet (Liu et al., 2017; Santos, 2011).

The Municipality of Lisbon has a program that provides different datasets about the city. The datasets are open-sourced to allow data analytics research and services creation; data for waste collection circuits is one of the datasets available, making possible this analysis.

2.3.1. Waste Management Integrated System in Lisbon

Lisbon's Council municipal waste system integrates several phases of waste management, from production, disposal, transport, and temporary storage to its referral for recycling or other forms of recovery, treatment, and destination final (Município de Lisboa, 2020).

According to the literature, the Municipal Solid Waste Management system of Lisbon has innovated through time to be able to respond to the cities' necessities. Different waste deposits and collection types depend on the waste types.

This study used mixed, paper, glass, and plastic waste because they are the most common and used for the community in general. Streams of waste like wasted electrical and electronic equipment (WEEE), construction and demolition waste (CDW), used oils, batteries (portable-batteries, Pb-acid batteries, Ni-Cd batteries), End of Life Vehicles (ELVs), used car tires, waste metals (ferrous and non-ferrous) and wood waste were not considered in this study. They are separately collected and collected by private operators (Município de Lisboa, 2020; Namlis & Komilis, 2019).

The Municipal Waste Management Plan for the Lisbon Council (Município de Lisboa, 2020) defines that the paper, glass, and plastic waste are collected for further recycling through the Sorting Centre and Ecocenter (SCE). The mixed waste is collected for incineration with energy recovery through Municipal Solid Waste Treatment Centre (MSWTC).

Concerning the waste collection circuits, the city of Lisbon is divided and organised into four areas encompassing a total of 8 support posts of waste removal.

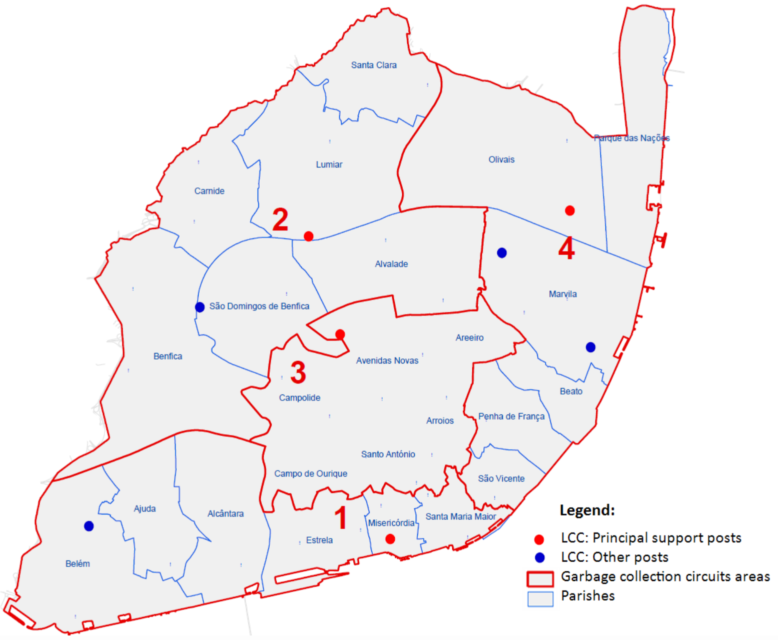


Figure 1 – Organization of service by supporting zones (Município de Lisboa, 2020)

2.3.2. Waste Process

Mixed waste is subjected to an energy recovery process at the Municipal Solid Waste Treatment Center (MSWTC), giving rise to electricity, slag, and residual solids production. During the incineration process, combustion gases are released, that are treated before being emitted into the atmosphere. The residual solids produced are inertized at the Inertization Station and subsequently deposited in the Landfill. The slags are treated at the Slag Treatment and Recovery Facility (STRF), where inert waste, ferrous and non-ferrous metals are separated (Município de Lisboa, 2020).

Selectively collected glass packaging is deposited at a specific location in the Sorting Center, where a manual selection of the most problematic contaminants for the glass industry (e.g. ceramics) is made (Município de Lisboa, 2020)

Paper or cardboard packaging of plastic, metal and cardboard for selectively collected food liquids is also delivered to the Sorting Centre. In the sorting line of plastic and metal packaging, the various materials are used using mechanical and manual processes. The packaging is then sent, in bales, to the recycling industry, through the Sociedade Ponto Verde (SPV) (Município de Lisboa, 2020).

2.3.3. Waste disposal and collection system

The municipality has different waste disposal and collection solutions depending on each area's producers, urban morphology, and socioeconomic characteristics. The existing main collection systems can be classified into (i) collective collection; (ii) door-to-door collection; (iii) pneumatic collection and (iv) reception sites waste (Município de Lisboa, 2020; Santos, 2011)

2.3.3.1. Collective Collection

They consist of large-capacity container assemblies located for public and collective use. They are commonly known as eco points and are intended for the selective deposition of three flows of recycled materials: (i) plastic packaging, (ii) paper and (iii) glass. Commonly is also composed of containers for mixed waste, forming a battery of four waste streams. This battery of four waste streams is called an eco-island in the municipality of Lisbon (Município de Lisboa, 2020; Santos, 2011)

This kind of collection is done at different times, depending on the waste stream. Eco-islands waste is collected from two times a week to one time a month for recycled materials like paper, glass and plastic packages, and four times a week for mixed waste. The singular eco points (no mixed waste containers) are collected from one time a week to one time a month (Município de Lisboa, 2020; Santos, 2011)

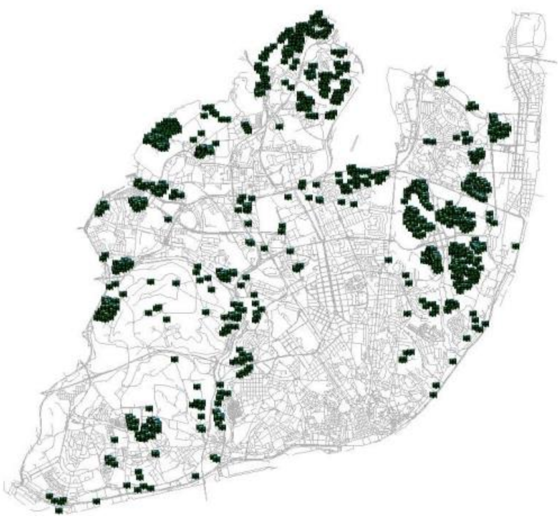


Figure 2 – Eco-islands in the Municipality of Lisbon (Município de Lisboa, 2020)

2.3.3.2. Selective Collection – Door to Door

Consist in individual containers assigned to houses, buildings or economic activities located in residential areas and aside entities like restaurants, bars, and hotels (Município de Lisboa, 2020; Santos, 2011)

This kind of collection is also done at different times and depending on the waste stream. For houses and medium/large buildings, the mixed waste is collected three to six times a week, and the recycled materials are collected one to two times a week. For the entities with economic activity, waste collection frequency is higher because more waste is produced, where either mixed and recycled materials are collected one to six times a week (Município de Lisboa, 2020; Santos, 2011)

3. METHODOLOGY

The methodology implemented to develop the study was based on the application of clustering techniques to identify groups of waste collection circuits that have a similar waste collection profile along the year and identify the socioeconomic characteristics of the population that is covered and makes use of those waste collection circuits, in the city of Lisbon. This methodology will take a Design Science approach.

A Data Mining project was implemented following the subsequent phases: identification of the problem, collection, and analysis of the data, performing a cleaning and the necessary transformations to the data, applying the algorithms chosen, and evaluation of the results obtained.

To this end, we had access to data about the waste loads for 557 circuits provided by the Urban Hygiene Department of Lisbon (*Lisboa Aberta*, 2021) regarding the period from 01/01/2017 and 01/11/2020, with 267,185 observations. This data set has 46 variables, with all the information about each circuit performed, such as the day, the hour, the total weight of waste transported by the truck (in kg), the type of waste transported, the origin of the transport (location), among others.

Initially, to evaluate the relationship between the waste loads in the different waste collection circuits, and identify the different groups, it is expected to use clustering techniques, such as k-means and Dynamic Time Warping.

K-means is the most commonly used model for clustering, that is, for partitioning a dataset into k-groups, depending on the homogeneity of the data characteristics (Likas et al., 2003).

Dynamic Time Warping is an algorithm that allows the user to compare time series, which do not contain a linear alignment, allowing to find patterns between measurements of events with different rhythms (Berndt & Clifford, 1994).

Additionally, since the objective of the study focuses on the contextual and socioeconomic characteristics of waste producers, other sources of information were used that are essential for the study, such as cultural events and school establishments data from Lisboa Aberta (*Lisboa Aberta*, 2021) and the preliminary results of the 2021 Census (counting) from the National Institute of Statistics (*INE – Censos 2021*, 2021), occupation and land use (% occupancy) from Territory General Directorate (*Lisboa Aberta*, 2021), infrastructure (counting of markets, restaurants, tourist establishments, among others) from Open Street Maps (*OpenStreetMap*, 2021) and Portugal Tourism (*Turismo de Portugal*, 2020) and the average price of sale of housing (m2, between 2017 and 2020) from (*Confidencial Imobiliário*, 2022) (Table 2).

The 2021 Census are used as a data source because it is the only source containing all the socioeconomic information of the population (level of education, employability, gender, number of dependents). However, the available data is only provisory since this study's definitive data is unavailable. This will be considered a limitation of the study. As for the average price of sale of housing, it will be used as a proxy level for the value of the population's income.

4. CONCEPTUAL MODEL

In this chapter is presented the conceptual model developed in this study. After defining the business objectives and requisites and problem identification, the data is identified, collected, and analyzed to describe, explore, and verify the quality of the data. This phase defines how reliable and viable the results can be.

On a second phase, the data was selected to be analyzed and perform a "cleanup" of the data, eliminating duplicates, redundant data, errors, outliers, and blank values. The Dynamic Time Warping technique was used since the data is time-dependent; it calculates the distance between different sequences coping with time deformations and different speeds (Müller, 2007)

The segments (Clusters) are created through the K-means algorithm in the modelling phase. K-means is the most commonly used model for clustering, that is, for partitioning a dataset into k groups. It starts by choosing the number of initial clusters (k) and randomly placing the so-called centroids, originating an iterative process (the attributes have to be numeric)

Following the evaluation, the model's results are compared against the objectives desired. Generally, a good result is obtained when the data points belong to the same cluster and are close to each other. An elbow graph and averages between clusters were used to find the number of clusters.

The clustering results were combined with the socioeconomic, demographic, and context variables, proceeding with the profiling phase. Each cluster was analyzed which are the most relevant socioeconomic and contextual variables, that can describe the clusters and understand the factors that could influence waste generation.

After the previous phases had been carried out, from analysis of the results was determined which policies were most appropriate to implement, given the study objectives, through the conversion of the identified clusters.

According to Thuraisingham (2000), the data mining process is iterative, meaning there is a need to repeat the process to improve data analysis and results. In this sense, the mixed type of waste was selected to be used as a test to ensure data quality, and a more detailed explanation will follow. And then, the model defined was applied to the other types of waste, always considering the specificities of each type of waste information.

4.1. MODEL FOR WASTE PRODUCTION PROFILING

4.1.1. Data Understanding

Lisbon is the main engine for the country's overall performance, given the centralization of resources, production and consumption, and also due to the concentration of the population itself (Ribeiro, 2017), as it is the city with the largest number of inhabitants in Portugal, about 545 thousand inhabitants (*INE*, 2020). Given its size and centralization, Lisbon is a tourist attraction with great relevance in Portugal, registering 6 million foreign and 2.2 million national guests in 2019 (LCG, 2019). Furthermore, given the size of inhabitants and tourists during the year, the last have great influence in waste production in the city.

The data used in this study was collected through an open data portal site Lisboa Aberta, the Department of Urban Hygiene of Lisbon, where all the pick-up events were collected, gathering 267,185 observations from January 2017 to November 2020, from the city of Lisbon. The foremost information was the type of waste (mixed, glass, paper and plastic), timestamp (year, month, week, day and hour), measure waste weight (kg) and circuit identification (circuit id).In the same database, the waste collection point's locations were identified in each circuit for each waste type through coordinates data.

Additionally, a different database was gathered, containing information about the socioeconomic and demographic data of the citizens of Lisbon (Census) and other context information such as buildings (schools, restaurants, hospitals and others) and price per square meter (m2). The Census data were selected considering the factors that most influence waste generation in the literature review.

The Table 2 indicates which databases are collected and their respective sources summarized.

Table 2 – Summary table of Databases and Data Sources

Data	Data Sources	Category
Collection Circuits	Lisboa Aberta (*Lisboa Aberta*, 2021)	Waste
Collection Loads	Department of Urban Hygiene of Lisbon (*Resíduos*, 2022)	Waste
Census 2021	National Institute of Statistics (*INE – Censos 2021*, 2021)	Socioeconomic and demographic
School establishments, hospitals, health centres, public transportation	Lisboa Aberta (*Lisboa Aberta*, 2021)	Context
Restaurants, Markets, Coffee Shops	Open Street Maps (*OpenStreetMap*, 2021)	Context
Culture Spots, Cultural and Sporting Events	Open Street Maps (*OpenStreetMap*, 2021)	Context
Tourism Information (hotels, local accommodation)	Tourism (*Turismo de Portugal*, 2020)	Context
Average price of sale of housing (m2, between 2017 and 2020)	Confidencial Real State (*Confidencial Imobiliário*, 2022)	Context

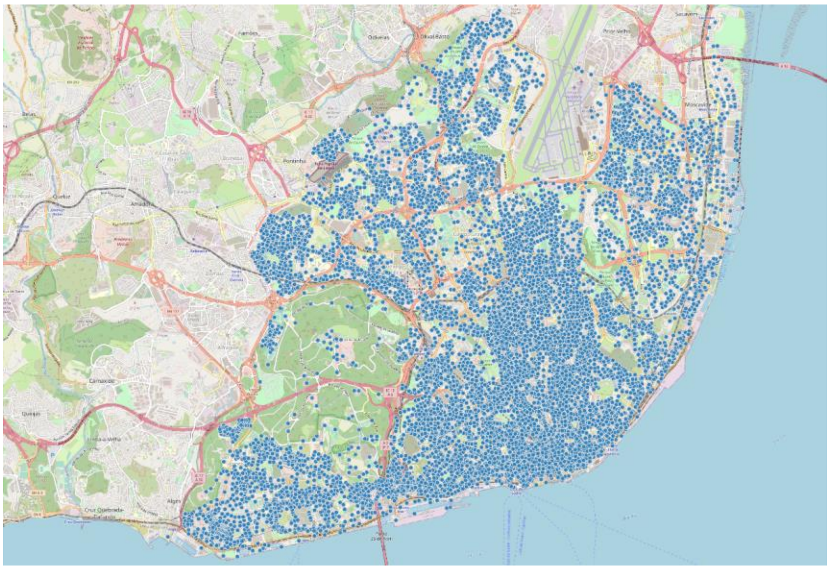


Figure 3 – Waste Collection Points and each location (Source Lisboa Aberta (*Circuitos Contentores*, 2021))

4.1.2. Data Preparation

4.1.2.1. Waste Collection Loads

The waste loads data was divided into four types of waste: mixed, glass, paper and plastic, simplifying the analysis by establishing a lower number of circuits and having different clusters for the different types, creating a separate database for each type.

Each circuit on the database has a weight collection associated with each freight collected and its geographical information. The data provided in the freight database indicates the quantity of waste collected in each waste collection load. Moreover, waste collection circuits have different waste collection frequencies (daily, weekly and monthly). Consequently, to guarantee that the waste collection in each circuit is comparable, was aggregated weekly for each waste collection circuit, using sum as the aggregator function. To have a better understanding of the average waste production in each week of the year, waste collection data was averaged in each week considering the years of 2017, 2018, 2019 and 2020.

Table 3 - Example circuit I0102 from mixed waste agregated by week for 22 weeks (see Table 24 for full table)

Sum of Weight ton		**Years**						
Circuit	**Week**	**2017**	**2018**	**2019**	**2020**	**Grand Total**	**Average**	**Average 2017-2019**
I0102	1	28,14	26,18	14,04	23,84	92,2	23,05	22,79
I0102	2	28,58	27,88	25,14	24,76	106,36	26,59	27,20
I0102	3	24,78	27,84	25,78	26	104,4	26,1	26,13
I0102	4	26,62	27,76	24,7	34,04	113,12	28,28	26,36
I0102	5	27,04	28,26	22,28	20,8	98,38	24,595	25,86
I0102	6	28,6	31,98	27,76	38,86	127,2	31,8	29,45
I0102	7	27,46	25,88	18,86	28,58	100,78	25,195	24,07
I0102	8	28,14	24,72	30,58	30,22	113,66	28,415	27,81
I0102	9	25,84	25,9	25,1	26,36	103,2	25,8	25,61
I0102	10	27,82	28,38	21,6	26,3	104,1	26,025	25,93
I0102	11	28,68	28,46	25,34	21,76	104,24	26,06	27,49
I0102	12	27,56	29,22	21,26	42,62	120,66	30,165	26,01
I0102	13	28,1	26,54	24,72	19,56	98,92	24,73	26,45
I0102	14	26,5	26,64	24,76	27,82	105,72	26,43	25,97
I0102	15	27,52	21,56	23,8	23,16	96,04	24,01	24,29
I0102	16	26,98	25,9	25,76	34,44	113,08	28,27	26,21
I0102	17	29,5	26,3	21,96	29,44	107,2	26,8	25,92
I0102	18	28,04	22,94	24,68	36,72	112,38	28,095	25,22
I0102	19	27,84	23,62	22,96	40,3	114,72	28,68	24,81
I0102	20	29,78	30,06	27,36	40,2	127,4	31,85	29,07
I0102	21	28,04	23,88	25,76	36,46	114,14	28,535	25,89
I0102	22	29,46	32,44	25,78	37,34	125,02	31,255	29,23

The data was aggregated weekly, producing a weekly time-series over the geographical area and generating a regular geospatial waste generation time series. In that sense, all the noisy temporal trends would become smooth. Noise or outliers is a characteristic of data resulting from external actions and elements of the waste producers (Korhonen & Kaila, 2015).

Although the data became smoother, it also needs to be clean, eliminating the circuits that did not occur during the range of time being analysed or were discontinued, and the circuits that do not include any information regarding the quantity of waste collected. They are also considered as outliers and missing data, deteriorating the data.

Although 2020 figures are in the table, the year 2020 was excluded because it was a year where due to the COVID-19 pandemic, waste production suffered a massive decline after the imposition of the lockdown in Portugal (Sarmento et al., 2022). Thus, the data only aggregate information from the years 2017 to 2019.

Afterwards, the waste generation scale was normalised to the annual waste amount reflecting the relative waste generation values.

This process has been applied for all types of waste, guaranteeing that data assumes the same range of years to be analysed, mitigating the differences between different kinds of waste, and ensuring data quality.

4.1.2.2. Sociodemographic and Contextual Data

The socioeconomic and context database and the waste load database were spatially combined using the spatial join technique to get specific information about each geo-point, associating each socioeconomic detail.

Spatial Join is a Geographic Information System operation that compares spatially one target feature with other feature layers through proximity, using the points inside each land parcel and moving the point table columns into the land parcel layer (Jacox & Samet, 2007).

Since the waste loads database contain more circuits than the socioeconomic collection points database, there was the need to match both databases capturing the circuits that appear in both databases. This process will filter the number of circuits for each waste type to be analysed.

4.1.2.3. Data Transformation

In the attempt to improve the clustering results, it was decided that the weight data would be replaced by the amount of waste collected by week and by the cubic meter (m3) during the three years (on average), considering a data mining project iterative.

For that, it was possible to calculate the total cubic meter for each circuit using the data contained in the freight database. In the freight database, waste collection points data such as the maximum container capacity variable was available in litres (L). Using the formula below, it was possible to calculate the maximum capacity or volume by the cubic meter (m3) referred to by Niska and Serkola (2018).

$$m^3 = \frac{L}{1000}$$

Equation 1 - Conversion litre to the cubic meter

However, the results showed that there was no improvement, and so, the initial approach was implemented.

4.1.3. Modelling

This phase involves selecting the modelling technique, which is usually related to the business problem or data.

According to the methodology selected, the software SAS Enterprise Miner allows the use of the DTW (Dynamic Time Warping) method to compare time series using the node TS Similarity (Schubert & Lee, 2011), and the K-means algorithm to determine the clusters of the circuits, that is, the groups of circuits that possess similar characteristics. To find the ideal number of clusters, the Elbow Graph was used.

Dynamic Time Warping - DTW

Instead of the Euclidean distance that measures the distance between the data points, the DTW allows comparing time series. This technique measures the similarity between temporal sequences, not ignoring the time dimensions and varies according to the time shifts, contrary to Euclidean distance (Amidon, 2021). This technique was applied to the time series of all waste types to understand the time profile of waste production in Lisbon.

K-Means Algorithm

K-means is the most commonly used model for clustering, that is, for partitioning a dataset into *k* groups. It starts by choosing the number of initial clusters (*k*) and randomly placing the so-called *centroids*, originating an iterative process (Syakur et al., 2018):

- For each point, looks for the nearest centroid by calculating the distance between the two and assign each point to the closest *centroid*, forming a cluster;
- The average of points assigned to each identified cluster is calculated, and the *centroid* is changed to the calculated mean;
- This process is repeated until it is not possible to change the clusters, having reached the final clusters.

This method sorts the data into *k* groups, meeting the following conditions:

- Each group contains at least one object;
- Each object fits only one cluster. The separation is improved by moving objects from one group to another using an iterative restructuring technique.

Elbow Graph

In order to detect the ideal number of clusters, the Elbow Graph, or Elbow Method, was used. This method expresses the percentage of variance explained according to the number of clusters. The optimal number of clusters should be one that, when adding another cluster, will not improve data modelling (Syakur et al., 2018).

The Elbow Graph method clusters K-means in the existing dataset, that is, groups the data into groups where that data has very similar characteristics, for a range of example, 10 clusters, and for each value, calculates the sum of squared errors (SSE).

The idea is to minimize the SSE to obtain a graph as an arm, and the "elbow" of the arm indicates the best number of clusters. In addition, the SSE tends to decrease to 0 as the number of clusters increases, so we can say that the SSE is 0 when the number of clusters is equal to the number of *datapoints* within the dataset so that each *datapoint* will represent a particular cluster and there will be no error between them and the centre of their cluster. Thus, the Elbow Method aims to discover a small number of clusters, which still represent a low SSE, and where the "elbow" usually indicates where it begins to have decreasing returns with an increasing number of clusters.

4.1.4. Evaluation

This chapter contains the evaluation of the four types of waste: mixed, plastic, paper, and glass.

After the previous chapter, the data were conciliated, prepared, and cleaned, and the DTW and K-means, along with the Elbow graph, were applied to create the clusters for each waste type. In this chapter, the socioeconomic and context variables were combined to analyse each cluster's social and demographic characteristics. Four different approaches were tested to find the best results, with the help of another normalization method, Z-Score.

Z-Score is a normalization method to standardize data. The z-score indicates how far from the mean a data point is. It takes the difference between the field value and the mean value and scales it by the standard deviation of the field's value. In this case, the z-score allows the user to compare the differences between each cluster for each variable. Usually, the considered range is [-3,3]. The closer to the range limit, the more significant the variable is to the description of the cluster.

Furthermore, to help the analysis, there was the need to visualize the location of each waste collection point associated with each cluster and understand to which city zones each cluster was inserted. Each point's longitude and latitude coordinates were available in the waste collection point data. The best way to provide this visualization was through the PowerBI tool.

Below the different approaches are explained:

Absolute Amount

For each socioeconomic variable, it was calculated the average amount for each cluster, which allowed us to identify which cluster had the biggest and the lowest value, considering the absolute amount. To understand better the differences between clusters, the z-score was calculated, verifying which variables had more significance. The results were the ones expected. However, there was a need to comprehend if another approach could better explain the differences.

Total Amount per each Variable

Another approach was to calculate the percentage for each variable. Although there are too many features, there was the possibility of calculating the total value of a feature group. For example, four features represent different age groups for female and male people. However, with the sum of all four features amount we have the total amount of people for each cluster, allowing us to calculate the percentage, for instance, of the total people inside cluster 1 with ages between 1 and 14 years. The same procedure was used for the remaining variables. Then, the z-score was calculated, and the results were better than the absolute amount.

Total Amount per each Collection Point

This approach came with the fact that different clusters had a different number of circuits and, by itself, a different number of waste collection points. One circuit could be longer than another one. To normalize that situation, it was decided to have a proportion of each variable with the total number of collection points for each cluster. The result was not the best because it was concluded that although there are differences between the number of waste collection points, it does not mean that a circuit with more collection points collects more waste. After all, the waste containers could have different sizes. In this sense, this approach was disregarded.

Total Amount per Maximum Capacity (in m^3)

Considering the result above, waste containers could have different sizes, and in some ways, they could not be comparable. To minimize this issue, a proportion of each variable with the maximum capacity of each cluster was calculated. As mentioned before, there was available information to calculate the container size by the cubic meter, so the total capacity of each cluster was known by summing the maximum capacity of all circuits that belong to the same cluster. However, the results could not have shown what was expected.

It was concluded that the most reasonable approach is the total amount per each variable.

4.2. Socioeconomic and Context Variable Association

After detecting the clusters for each type of waste, the socioeconomic data was combined to classify each cluster's socioeconomic, demographic, and context characteristics, using descriptive statistics. As previously mentioned, the spatial join technique combined socioeconomic and context information with the loads database (Table 2) to get specific information about each geo-point, associating each socioeconomic detail.

Census 2021 provisory data was used. In this sense, the data is only available at a parish level, being only possible to calculate the mean value for each socioeconomic variable for each circuit. For instance, if one waste collection circuit comprises two different parishes, the result of the number of people with ages between 0 and 14 years would be the mean value of both parishes rather than the actual value. Table 25 in Annex I shows all the socioeconomic and context variables and their brief description.

Incorporating the literature review and the socioeconomic and context variables retrieved, the variables that better influence waste generation are:

Table 4 - Variables that better influence waste generation

Variables	
HM_0_14_years	Citizen's Age - Male and Female
HM_15_24_years	
HM_25_64_years	
HM_65_and_more_years	
No Schooling level	Level of Education
Basic Education	
1st Cycle	
2nd Cycle	
3rd Cycle	
Secondary and post-secondary education	
University Education	
Size Less than 30 m	Household Living Space
Size 30 m to 39 m	
Size 40 m to 49 m	
Size 50 m to 59 m	
Size 60 m to 79 m	
Size 80 m to 99 m	
Size 100 m to 119 m	
Size 120 m to 149 m	
Size 150 m to 199 m	
Size 200 m or more	
Rent Less than 20 euros	House rent - proximity of family income
Rent 20 to 49,99 euros	
Rent 50 to 99,99 euros	
Rent 100 to 199,99 euros	
Rent 200 to 399,99 euros	
Rent 400 to 649,99 euros	
Rent 650 to 999,99 euros	
Rent 1000 or more euros	
Family size with 1 person	Family Size
Family size with 2 people	
Family size with 3 people	
Family size with 4 people	
Family size with 5 or more people	
Rest Coffe Shop Bar Market	Cultural and Social aspects
Culture Spot	
Beds Tourist Establishments	
Beds Local accommodation	
Cultural and Sporting events	
Price per sqm	Rent - proximity of family income

5. RESULTS AND DISCUSSION

5.1. Data Preparation

5.1.1. Mixed Waste

As mentioned before, mixed waste was selected as the first data to be analysed since more circuits, frequency, and waste collection points allow testing and adjusting the process. Furthermore, considering that the different types of waste came from the same database, they will follow the same line of thinking.

Preparing the data follows the conciliation, aggregation, and normalization noted above, matching the different databases to ensure only the circuits in all databases were analysed and the remaining ones were disregarded. It resulted in a total of tons of waste collected by week on average for the three years being analysed (2017, 2018 and 2019). For instance, in table 3 for the circuit I0102, the amount of waste collected in the first week for the four years is, on average, 23,05 tons. And that is the base considered for the clustering model.

This result allows the user to compare the quantity of waste collected during the year, considering some points when the amount of waste produced can vary according to the events during the year. These events include Easter, Popular Saints, summer holidays, Halloween, Christmas, and New Year's Eve.

The graphic in Annex II Figure 35 shows the first result of this process. There was the need to start the year in February ending in January to permit considering the time-series behaviour in the Christmas and New Year's festivities.

5.1.1.1. Data Cleaning

After the first result, the mixed waste was composed of 103 circuits. It was necessary to understand that not all of these circuits contain relevant information for the analysis because some were interrupted or only started at the end of the analysed range, corrupting the data.

The available data is from 01/01/2017 to 01/11/2020, and it is evident that the range from November to December has a significant decrease since there is only data in that range period in the years between 2017 and 2019. Also, 2020, as cited before, was an unusual year, with the Covid-19 pandemic situation, where the data can be influenced. In that sense, 2020 was excluded from the data considering only the periods between 2017 and 2019.

From the 103 circuits, six showed that occurred a few times during the analysed range, having the average of the waste collected very low compared with the average of waste collected from all circuits, assuming some flaws.

Table 5 - Comparison between the average of mixed waste circuits.

Circuit	All	I0101MK	I0204B	I0217MK	I0308micK	IESA01Cx	IMF01PqN
Waste collected (avg) ton	1571,55	67,405	6,44	40,1438	2,045	61,635	54,53

After all the cleaning steps, the final number of mixed waste circuits is 97, with a range of periods between 2017 and 2019. The final graphic can be found in Figure 4.

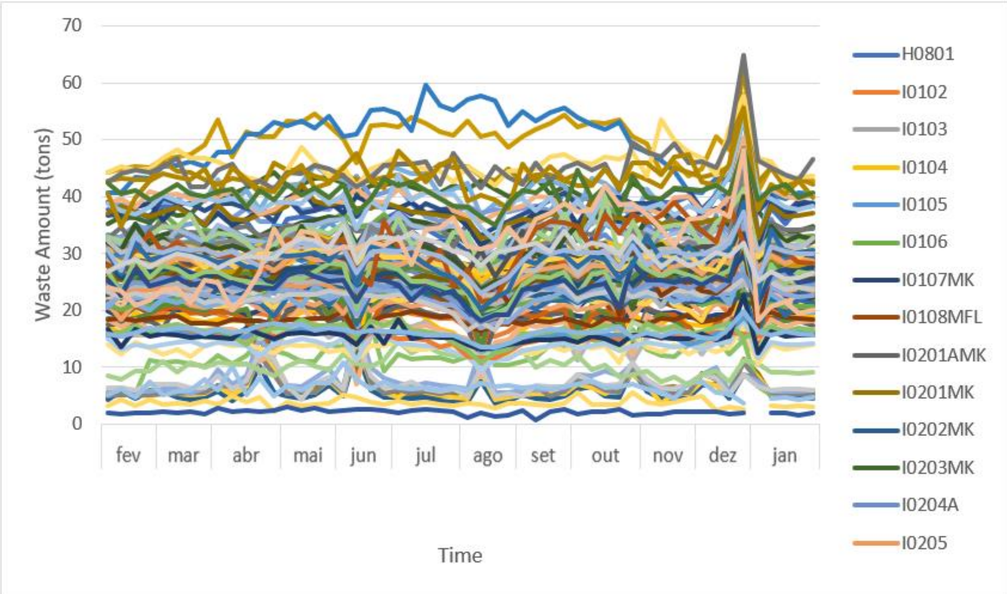


Figure 4 - Mixed Waste Time-Series result after data cleaning

To continue the analysis, SAS Enterprise Miner was used as a tool and software to prepare the time-series and clustering analysis. This data mining software allows the possibility to analyse complex data, discover patterns, and build models (*SAS Enterprise Miner*, 2022). The goal is to understand the similarity between time series of waste collection, dividing them into clusters or groups that possess the same time behaviour.

The choice of variables for the clustering analysis was based on building clusters established on time series. In this sense, only three variables were chosen from the aggregated database, Circuit ID, Weight, and Date, considering only the 97 circuits.

The table below shows the description of the variables and the classification of the specific roles associated with SAS for time series analysis (*SAS Center: SAS Enterprise Miner Data Sources*, 2022):

Table 6 - Variable description and classification

Variable	Description	Role
Circuit	Contains the circuit number identification	Cross ID
Weight (ton)	Average amount of waste collected in three years by week in tones	Input
Week	Provides timestamp or sequential information	Time ID

Outliers

As previously indicated, during the aggregation process, the data become smoother, eliminating the possibility of having outliers or noisy data. In this sense, no outliers were detected.

Missing Values

Besides the outliers, it is also necessary to analyse the missing values because data can be deteriorated, providing incorrect results.

Although the data was assembled and normalized, and circuits were eliminated, there were some failures where no data was available. For example, in the last week of December, there was no waste collection during the three analysed years for circuit I0215MK, meaning there was a null value.

The table below demonstrates the number of missing values or null observations found for the mixed waste before and after the aggregation process:

Table 7 - Missing values description for Mixed Waste

Missing Values (Observations)	
Original data	**After data aggregation**
207	3

For the mixed waste, it was only detected three missing values. Since it represents less than 1% of the observations (from 5044), it was considered that a constant value, 0, replaced the missing data, not affecting the dataset.

5.1.2. Recycled Waste

The same procedure was performed for the types of recycled waste. Furthermore, considering that different types of waste came from the same database, they will follow the same line of thinking.

As in mixed waste, the recycled waste data were conciliated, aggregated, and normalized, matching the different databases to ensure only the circuits in all databases were analysed and the remaining ones were disregarded. It resulted in a total of tons of waste collected by week on average for the three years being analysed (2017, 2018 and 2019).

However, there was a need to aggregate the data differently for glass waste. After the first result, it was found that of the 59 circuits, only 9 occurred weekly, while the remaining 50 occurred monthly. This is because there is less need to collect glass waste. In this sense, the glass waste was aggregated monthly, following the same reasoning, resulting in tons of waste collected by month on average for the three years being analysed (2017, 2018 and 2019).

The charts in Annex II shows the first result of this process[2].

5.1.2.1. Data Cleaning

Plastic Waste

After the first result, the plastic waste was composed of 60 circuits. It is necessary to understand that not all these circuits contain relevant information for the analysis because some were interrupted or only started at the end of the analysed range, corrupting the data.

As in mixed waste, 2020 was excluded from the data considering only periods between 2017 and 2019, due to the unusual year, with the Covid-19 pandemic situation, where data can be influenced.

Seven of the 60 circuits showed that the average of the waste collected was lower compared with the average of waste collected from all circuits, assuming some flaws.

Table 8 - Comparison between the average of plastic waste circuits.

Circuit	All	E0812	EE0001	E0816	E0204MK	E0105	E0805	E0101MK
Waste collected (avg) ton	155,92	26,785	29,29	29,235	62,68	15,095	29,56	7,875

After all the cleaning steps, the final number of plastic waste circuits considered were 53, with a range of periods between 2017 and 2019. The final chart can be found in Figure 5.

[2] There was also the need to start the year in February ending in January to permit considering the time-series behaviour in the Christmas and New Year's festivities

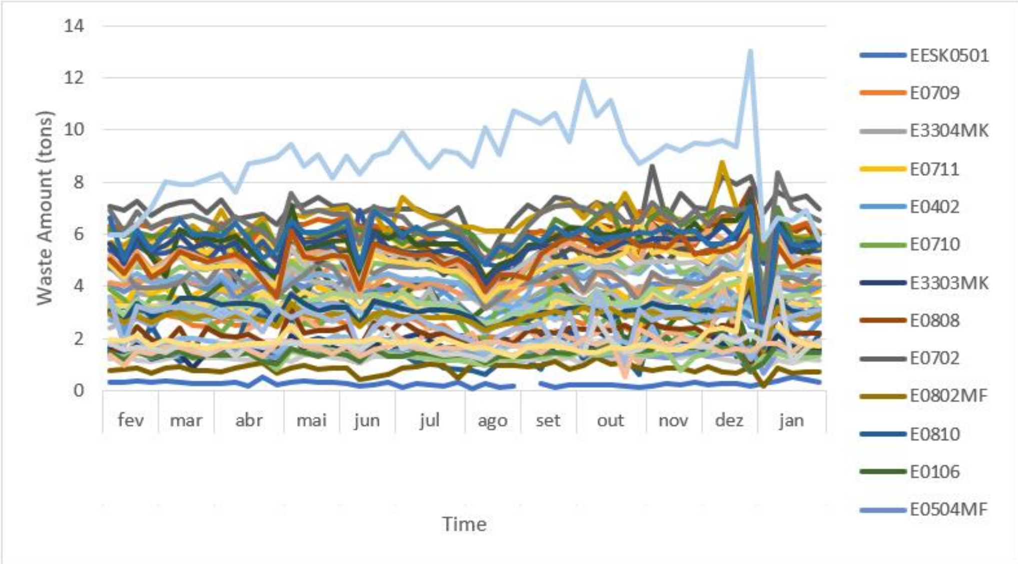


Figure 5 - Plastic Waste Time-Series result after data cleaning.

To continue the analysis, like in mixed waste, SAS Enterprise Miner was used as a tool and software to prepare the time-series and clustering analysis. An identical number of variables were chosen from the aggregated database (See Table 6).

Outliers

It was also verified that one of the plastic waste circuits contained a much higher amount of waste than the average. It was decided that this circuit would be excluded from clustering analysis since it could affect the data considering an outlier. As such, the number of final circuits went from 53 to 52. This circuit was analysed in the evaluation phase along with the clusters found to understand what impact a circuit has that has collected much waste.

In Annex II, Table 26 compares the average waste collected from all circuits and the circuit considered an outlier.

Missing Values

Besides the outliers, it is also necessary to analyse the missing values because data can be deteriorated, not providing good results.

Although the data was assembled and normalized, and circuits were eliminated, there were some failures where no data was available.

The table below demonstrates the number of missing values or null observations found for the plastic waste before and after the aggregation process:

Table 9 - Missing values description for Plastic Waste

Missing Values (Observations)	
Original data	**After data aggregation**
63	2

For the plastic waste, it was only detected three missing values. Since it represents less than 1% of the observations (from 2756), it was considered that a constant value, 0, replaced the missing data, not affecting the dataset.

Paper Waste

Following the first result, the paper waste was composed of 77 circuits. It is necessary to understand that not all these circuits contain relevant information for the analysis because some were interrupted or only started at the end of the analysed range, corrupting the data.

As in mixed waste, 2020 was excluded from the data considering only periods between 2017 and 2019, due to the unusual year, with the Covid-19 pandemic situation, where data can be influenced.

Seven of the 77 circuits showed that the average of the waste collected was inferior compared with the average of waste collected from all circuits, assuming some flaws.

Table 10 - Comparison between the average of paper waste circuits.

Circuit	All	P0812	P0710	P0810	P0204MK	P0106	PMF01PqN	P0101MK
Waste collected (avg) ton	160,11	76,06	51,03	70,72	95,25	18,86	36,52	6,37

After all the cleaning steps, the final number of paper waste circuits is 70, with a range of periods between 2017 and 2019. The final graphic can be found in Figure 6.

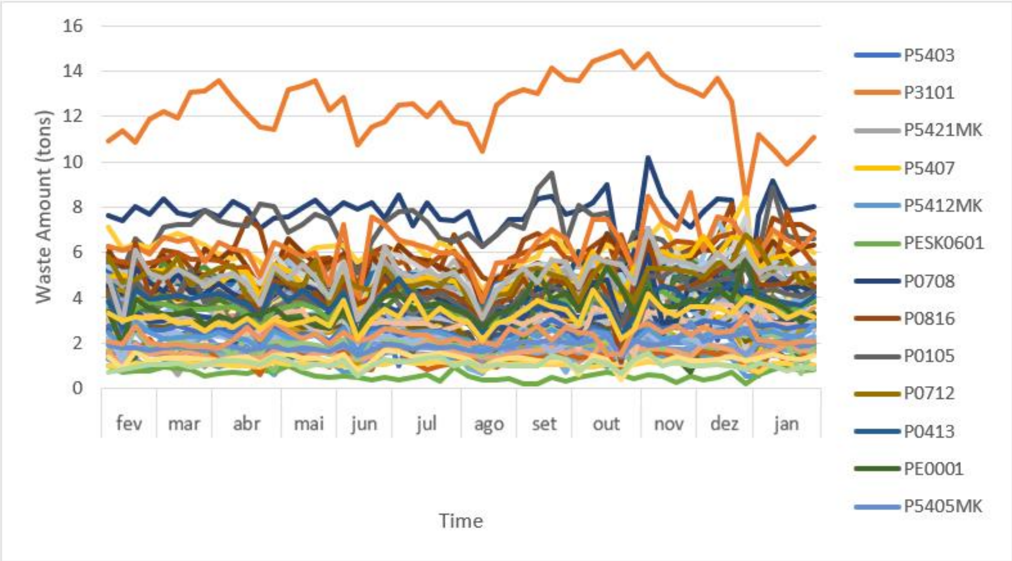


Figure 6 - Paper Waste Time-Series result after data cleaning

To continue the analysis, like in mixed waste, SAS Enterprise Miner was used as a tool and software to prepare the time-series and clustering analysis. An identical number of variables were chosen from the aggregated database (See Table 6).

Outliers

As per plastic waste, it was also verified that one of the paper waste circuits contained a much higher amount of waste than the average. It was decided that this circuit would be excluded from clustering analysis since it could affect the data considering an outlier. As such, the number of final circuits went from 70 to 69. This circuit was analysed in the evaluation phase along with the clusters found to understand what impact a circuit has that has collected much waste.

In Annex II, Table 27 compares the average waste collected from all circuits and the circuit considered an outlier.

Missing Values

After aggregation and elimination of circuits, it was found that no missing values or null values were identified, not requiring any treatment.

Glass Waste

Considering the first result, after aggregating the data monthly and not weekly as in the other types of waste, the glass waste was composed of 59 circuits. Although, there was a need to analyse the data to understand if some circuits would corrupt the data.

As in mixed waste, 2020 was excluded from the data considering only periods between 2017 and 2019, due to the unusual year, with the Covid-19 pandemic situation, where data can be influenced.

Of the 59 circuits, one showed that it occurred only during the year 2017; in this way, it was disregarded, counting 58 circuits.

Outliers

Another six circuits showed that the average of the waste collected was higher compared with the average of waste collected from all circuits, having the possibility to be considered outliers.

Table 11 - Comparison between the average of all glass waste circuits.

Circuit	All	V1303	V1402	V1305	V1309	VE0001	V1301
Waste collected (avg) ton	100,07	245,57	159,76	653,04	400,66	321,19	241,36

Despite having an average of collected waste much higher than the total average, these six circuits represent about 10.34% of the data, so they cannot be eliminated, and the final number of glass waste circuits is 58. However, the V1305 circuit has a much higher amount of glass waste collected than average, significantly affecting the results. Therefore, this is considered an outlier being analysed in the evaluation phase along with the clusters found to understand the circuit's impact that collected much waste.

Missing Values

After aggregation and elimination of circuits, it was found that no missing values or null values were identified, not requiring any treatment.

5.2. MODELLING

5.2.1. Mixed Waste

After running the node TS Similarity, it is possible to check already the potential number of clusters with the Cluster Constellation Plot and the Cluster Dendrogram. It is to be noted that these plots do not demonstrate the final number of clusters but an overview of the DTW.

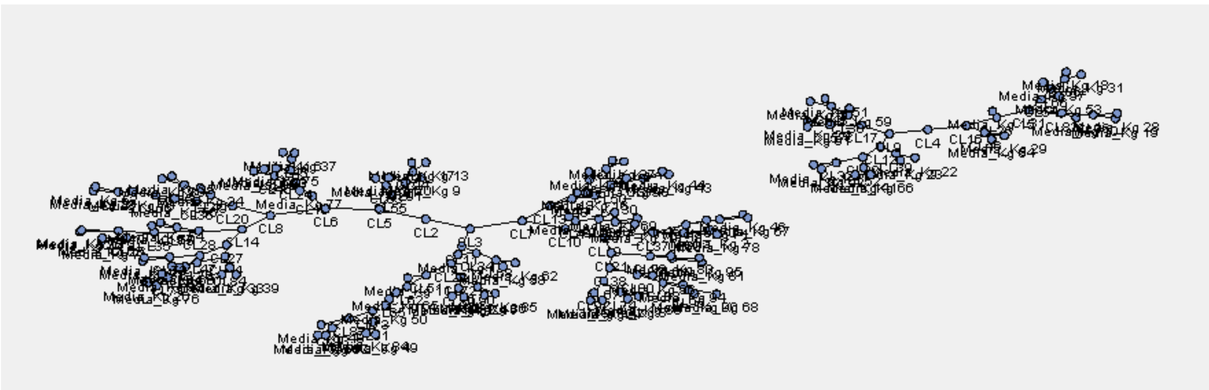


Figure 7 – Cluster Constellation Plot for Mixed Waste

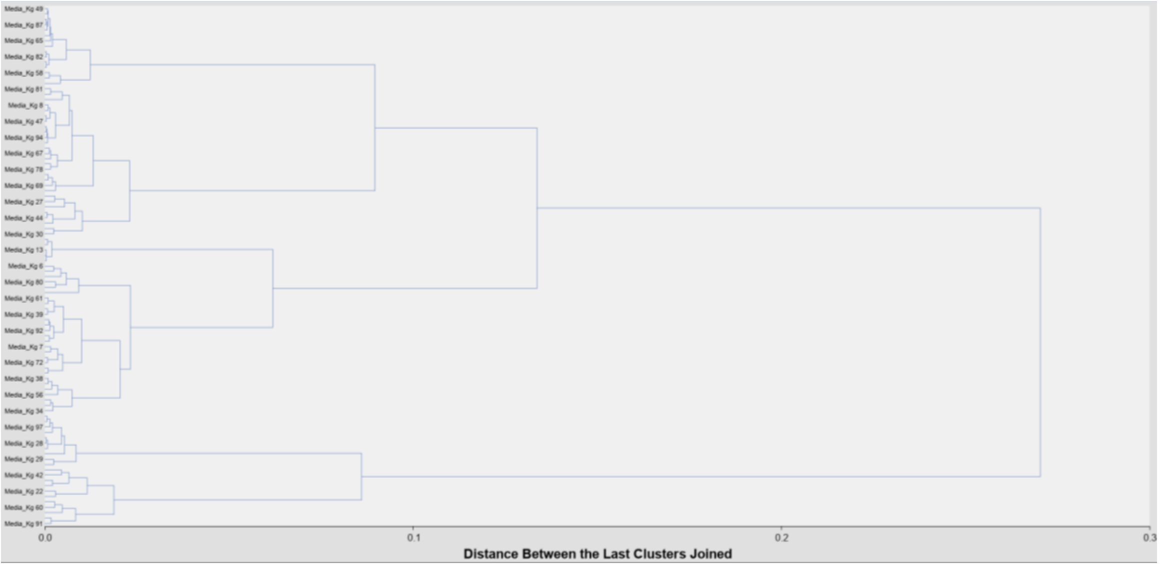


Figure 8 – Cluster Dendrogram for Mixed Waste

According to Schubert and Lee (2011), all other specific plots, like Line Plots, related to DTW were suppressed because it is costly to develop all combinations.

By analysing the different graphics above, it is possible to understand that there are at least two clusters. However, the Constellation Plot shows that more clusters can be formed. To understand the best number of clusters to use for the analysis is by use the Elbow Graph. The TS Similarity node combined all the time series through the Dynamic Time Warping method.

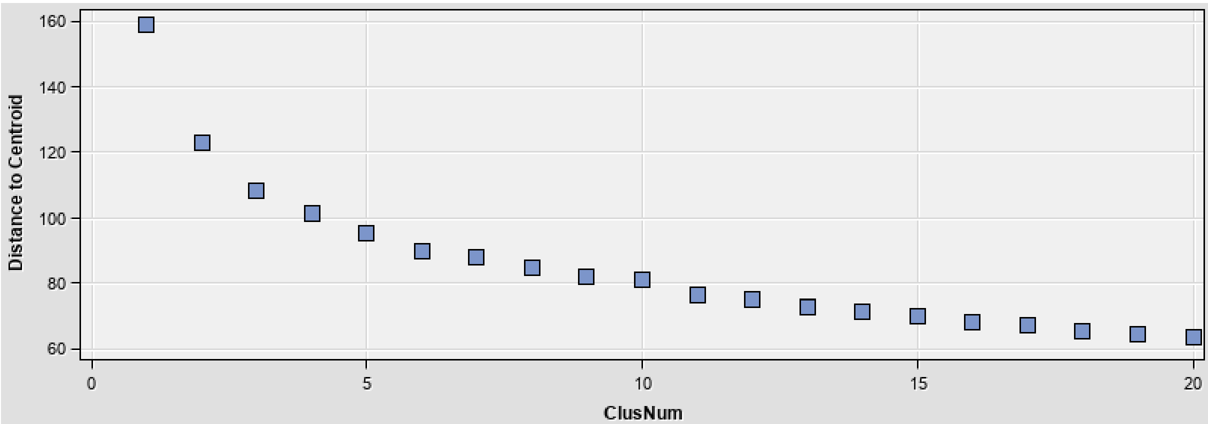


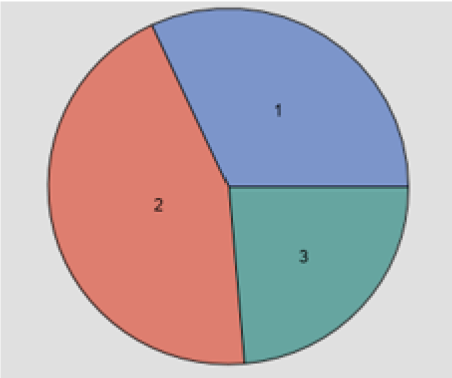
Figure 9 – Elbow Graph for Mixed Waste

According to the methodology, the Elbow Graph indicates that the best number of clusters is possibly three. However, further analysis and graphs were made to make the most appropriate choice regarding the ideal number of clusters.

Considering three clusters, the number of observations is well distributed, as shown in the table and Segment Size charts below, while comparing with four clusters:

Table 12 - Number of circuits in each segment, *k*=3 and *k*=4 (Mixed Waste)

Segment ID	Number of circuits (k=3)	Segment ID	Number of circuits (k=4)
1	31	1	29
2	43	2	40
3	23	3	10
-	-	4	18



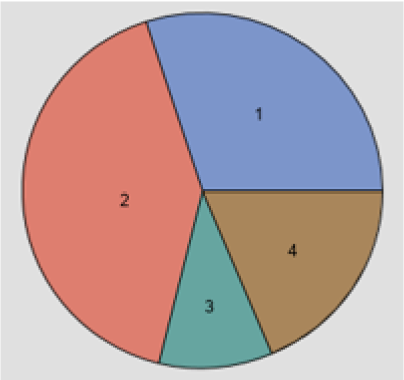


Figure 10 – Segment Size, *k*=3 and *k*=4 (Mixed Waste)

Comparing the two options, it is possible to see that segment ID 3 (in k=3) was divided into two clusters if it is considered k=4. In this way, having k=3 is considered the best solution. Annex III can be found different results considering different cluster numbers, from k=2 to k=6.

Extracting the results, where each circuit is associated with the segment ID or cluster number, a Linear Plot by Segment was performed to examine the differences between each cluster visually.

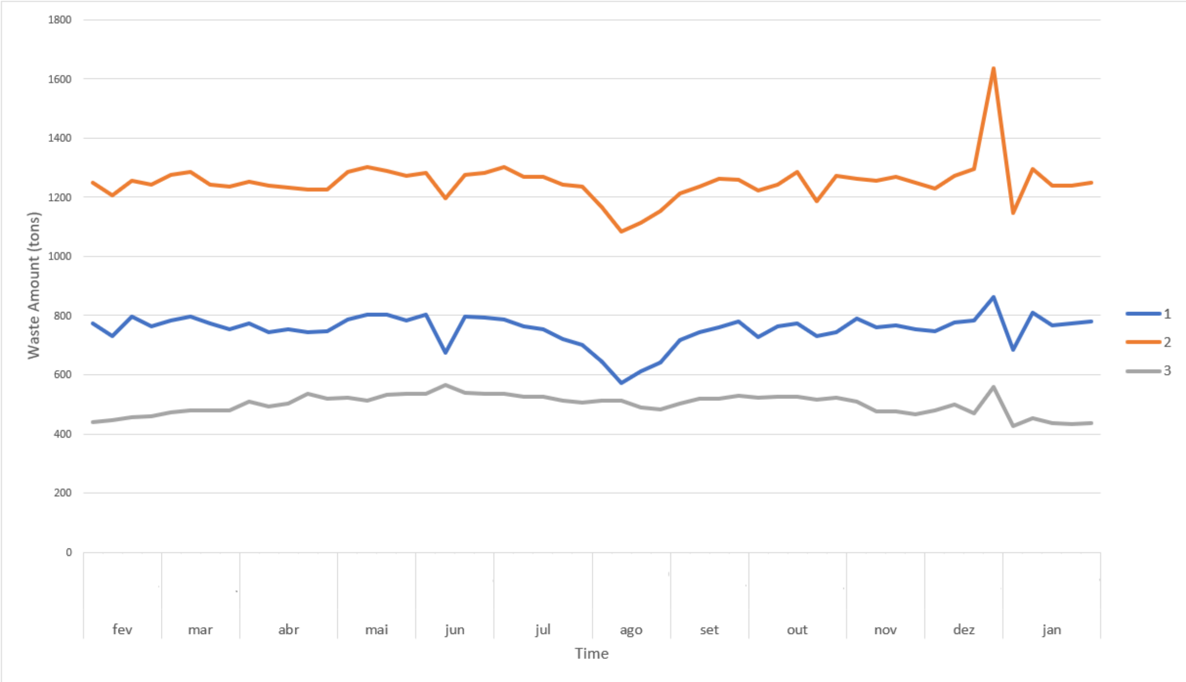


Figure 11 – Linear Plot by Segment *k*=3 (Mixed Waste)

This visualization makes it possible to recognize the discrepancies between each cluster, considering the absolute weight value. However, to understand the actual time series behaviour during the year, the Min-Max method was required to normalize the data (Patro & Sahu, 2015).

Normalization can make a continuous variable fall within a specific range while maintaining the relative differences between the values for the variables. In this case, for the Min-Max normalization method, the considered range is [0,1].

Figure 12 shows the resulting type profiles of mixed waste generation. This is the result of DTW and K-means. In line with a discovery in Finland by Korhonen and Kaila (2015), some seasonal and festivities influences in the waste profiles are identified. The three clusters identified a significant decrease in summer holidays in July and August and a peak in Christmas December-January. However, cluster 3 identified an increase in waste production during April-Jun, possibly related to Easter and Popular Saints in Portugal, contrary to clusters 1 and 2.

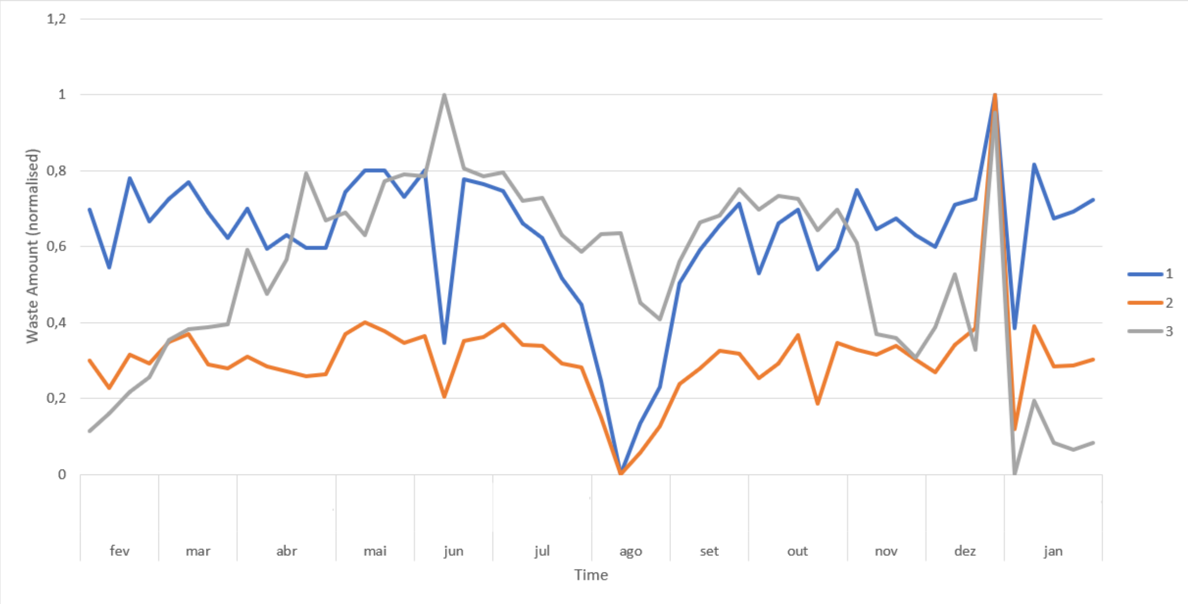


Figure 12 – Linear Plot by Segment *k*=3 after normalization (Mixed Waste)

5.2.2. Recycled Waste

After analysing the number of clusters for mixed waste, recycled waste was analysed: plastic, paper, and glass. To this end, the same procedure was followed as the mixed waste, the same methodology: using the DTW method to compare the similarity of the time series and the k-means and the elbow graph to reach the ideal number of clusters.

To make it easier to understand, each type of recycled waste is analysed individually, following the same process.

5.2.2.1. Plastic Waste

In the first approach, through the TS Similarity node, it was possible to verify the potential number of clusters with a Cluster Constellation Plot and the Cluster Dendrogram charts.

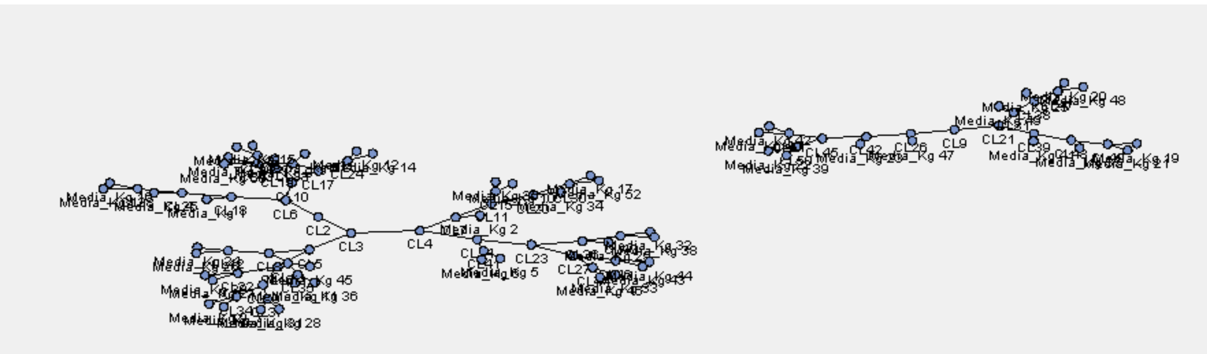


Figure 13 – Cluster Constellation Plot for Plastic Waste

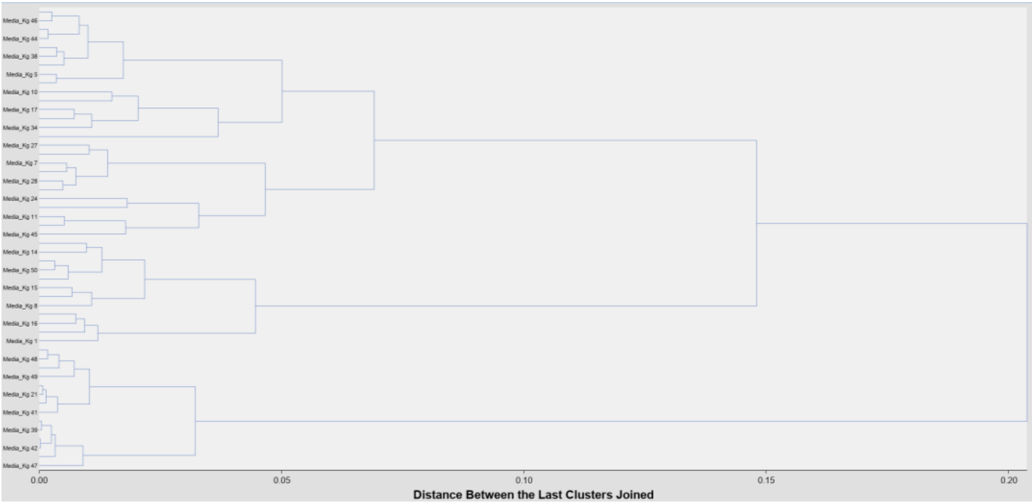


Figure 14 – Cluster Dendrogram for Plastic Waste

By analysing the different graphics above, it is possible to understand that there are at least two clusters. However, the Constellation Plot shows that more clusters can be formed. To understand the best number of clusters to use for the analysis is by use the Elbow Graph. The TS Similarity node combined all the time series through the Dynamic Time Warping method.

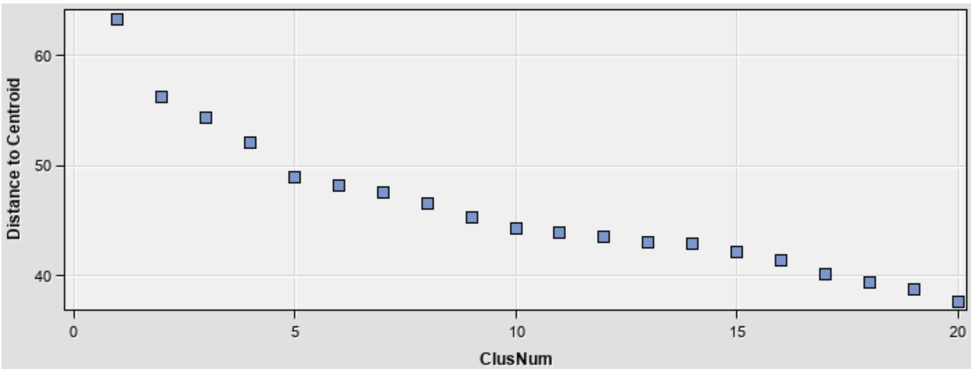


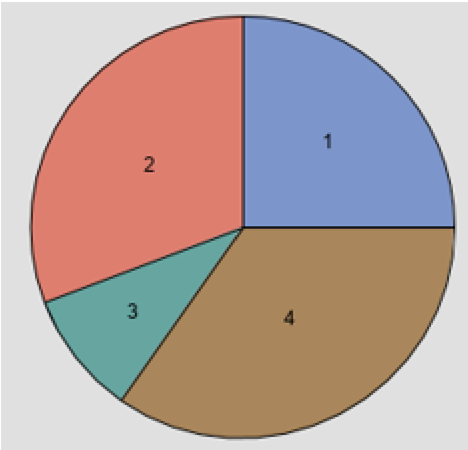
Figure 15 – Elbow Graph for Plastic Waste

According to the methodology, the Elbow Graph indicates that the best number of clusters is possibly four. However, further analysis and graphs were made to make the most appropriate choice regarding the ideal number of clusters.

The difference between four and five clusters was analysed because, according to the Elbow Graph, there is also a possibility of considering five clusters. Considering the tables and Segment Size charts below, it was found that there is a better distribution considering five clusters:

Table 13 - Number of circuits in each segment, *k*=4 and *k*=5 (Plastic Waste)

Segment ID	Number of circuits (k=4)	Segment ID	Number of circuits (k=5)
1	13	1	12
2	16	2	16
3	5	3	5
4	18	4	7
-	-	5	12



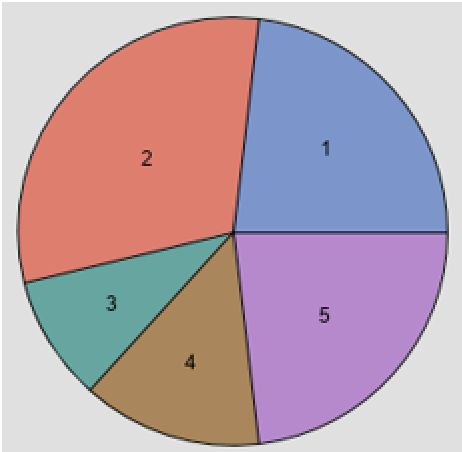


Figure 16 – Segment Size, *k*=4 and *k*=5 (Plastic Waste)

As analysed in mixed waste, a Linear Plot by Segment was constructed to examine the differences between the different clusters. It was observed that k=5 shows better results when compared with k=4.

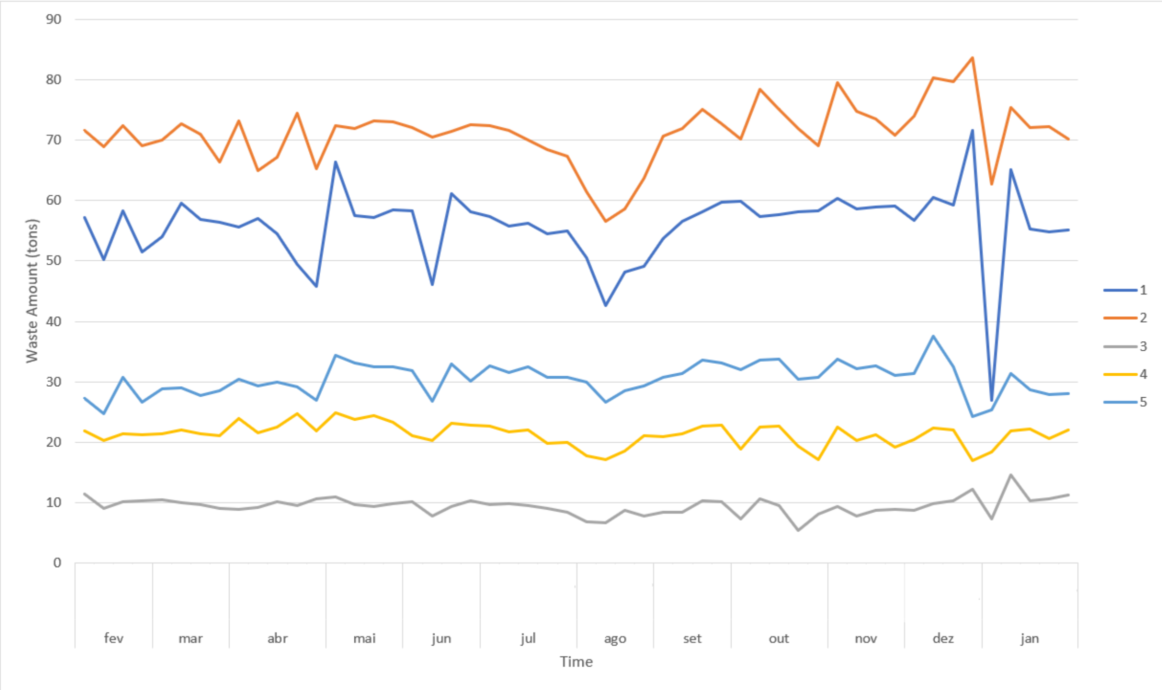


Figure 17 – Linear Plot by Segment *k*=5 (Plastic Waste)

Normalization was also used using the Min-Max method to verify the discrepancies between the different clusters, resulting from DTW and K-means, considering the same range [0.1].

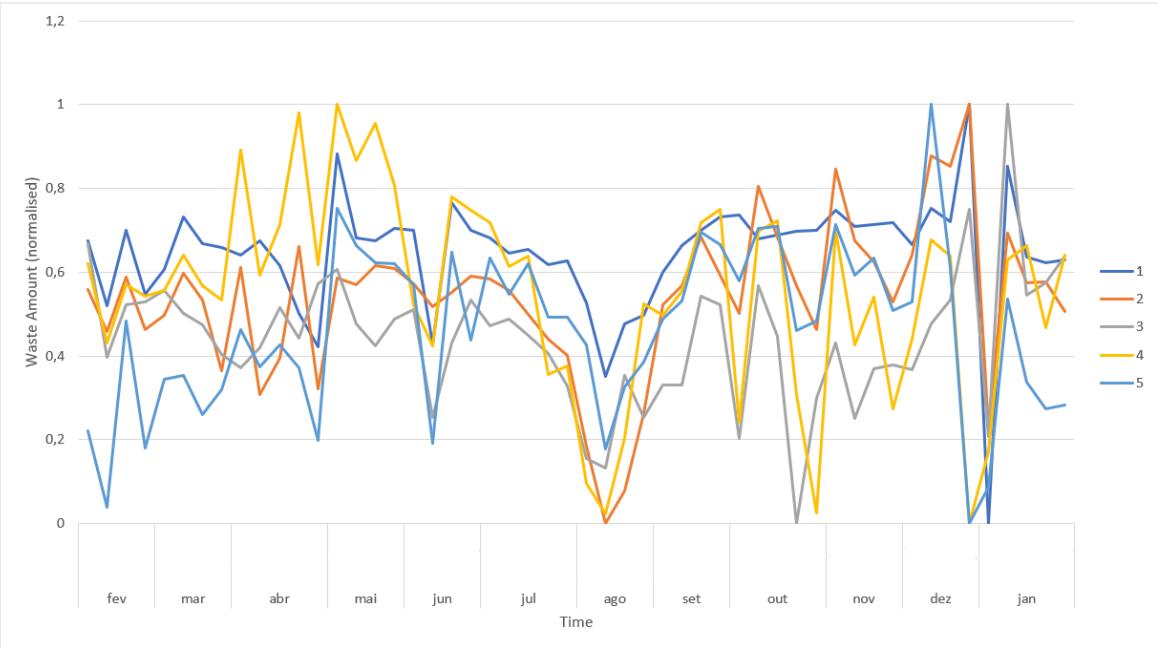


Figure 18 – Linear Plot by Segment *k*=5 after normalization (Plastic Waste)

All clusters showed significant variability throughout the year, indicating that plastic production in Lisbon is not constant. There is also some influence of seasons as it is verified in all clusters, like the considerable drop in the summer holidays, July-August. There are also substantial peaks at Christmas and New Year's Eve, December-January, and some decreases in the time of the Popular Saints in June.

5.2.2.2. Paper Waste

In the first approach, through the TS Similarity node, it was possible to verify the potential number of clusters with a Cluster Constellation Plot and the Cluster Dendrogram charts.

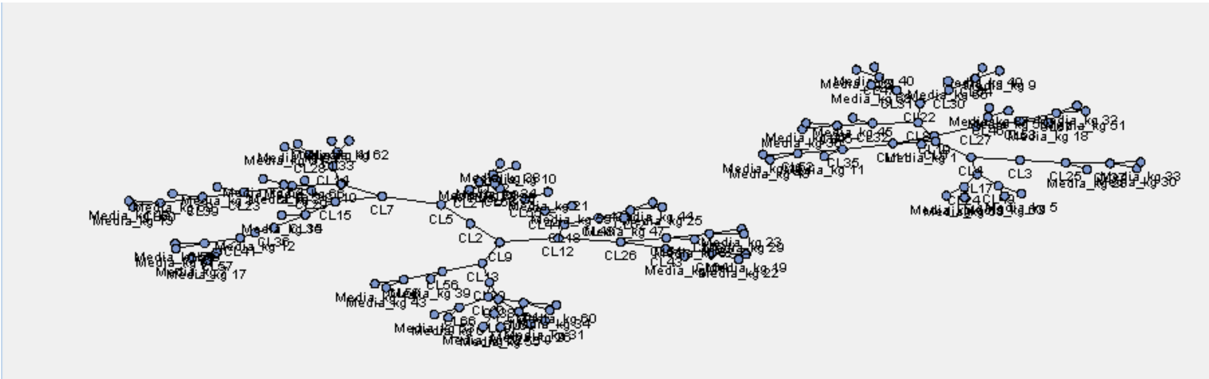


Figure 19 – Cluster Constellation Plot for Paper Waste

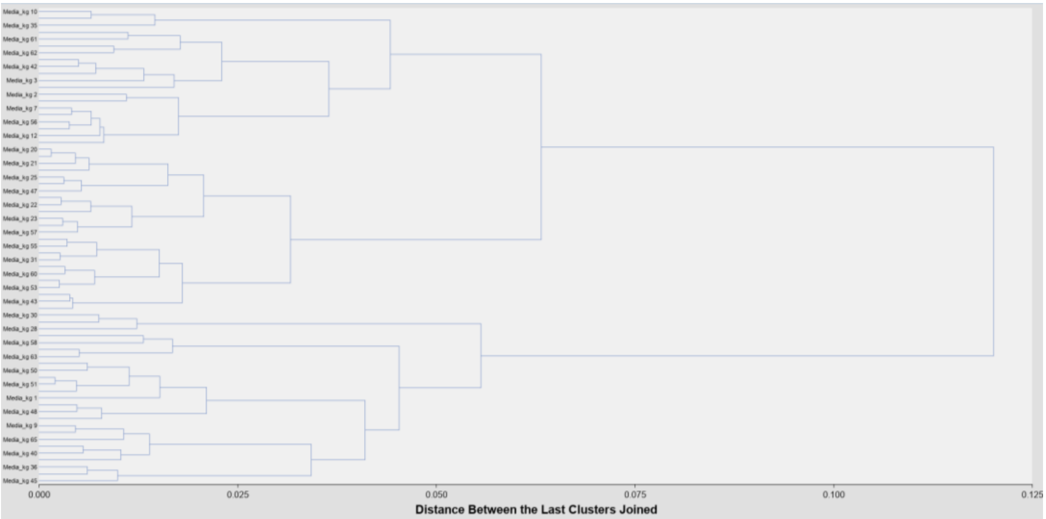


Figure 20 – Cluster Dendogram for Paper Waste

By analysing the different graphics above, it is possible to understand that there are at least two clusters. However, the Constellation Plot shows that more clusters can be formed. To understand the best number of clusters to use for the analysis is by use the Elbow Graph. The TS Similarity node combined all the time series through the Dynamic Time Warping method.

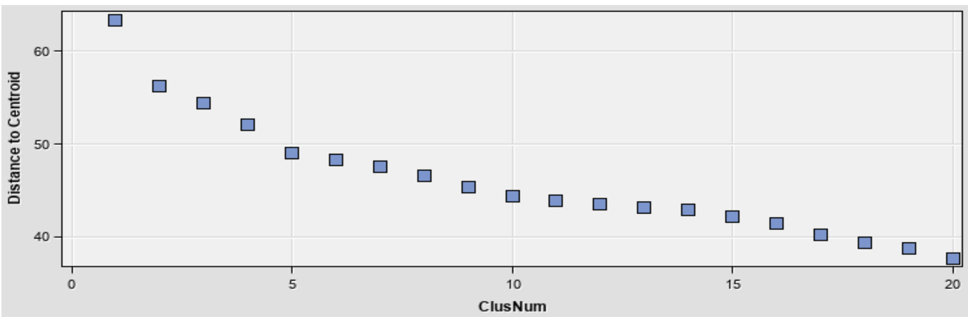


Figure 21 – Elbow Graph for Paper Waste

Following the methodology, it is difficult to understand the ideal number of clusters. The Elbow Graph indicates that the best number of clusters is possibly three. However, further analysis and graphs were made to make the most appropriate choice regarding the ideal number of clusters.

The difference between the three and four clusters is minimal, both of which have good distribution, as can be seen in the following tables and Segment Size plots:

Table 14 - Number of circuits in each segment, *k*=3 and *k*=4 (Paper Waste)

Segment ID	Number of circuits (k=3)	Segment ID	Number of circuits (k=4)
1	32	1	17
2	21	2	22
3	16	3	18
-	-	4	12

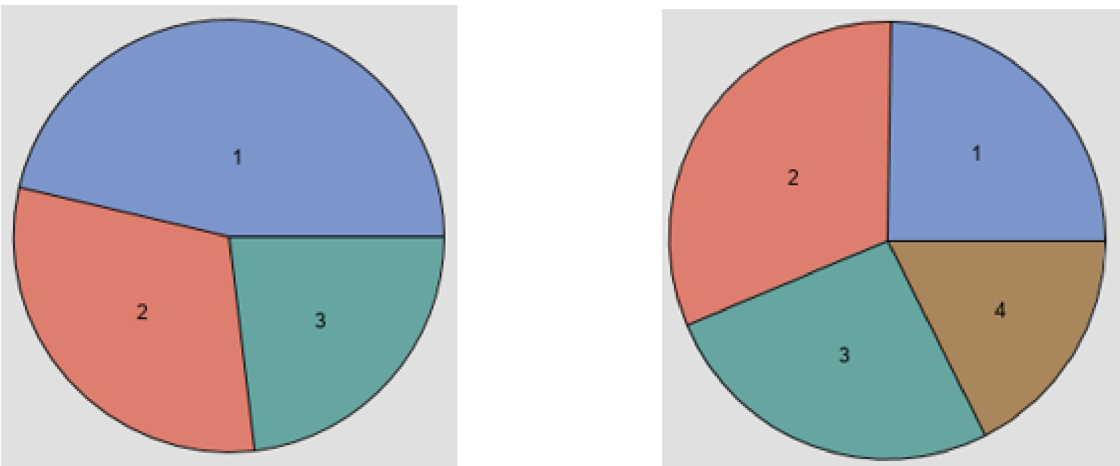


Figure 22 – Segment Size, *k*=3 and *k*=4 (Plastic Waste)

Analysing the tables and charts above was found that considering k=4, better results were obtained. However, a Linear Plot by Segment was constructed to examine the differences between the different clusters. It was found that although k=4 showed a better distribution, k=3 presented a greater difference between clusters.

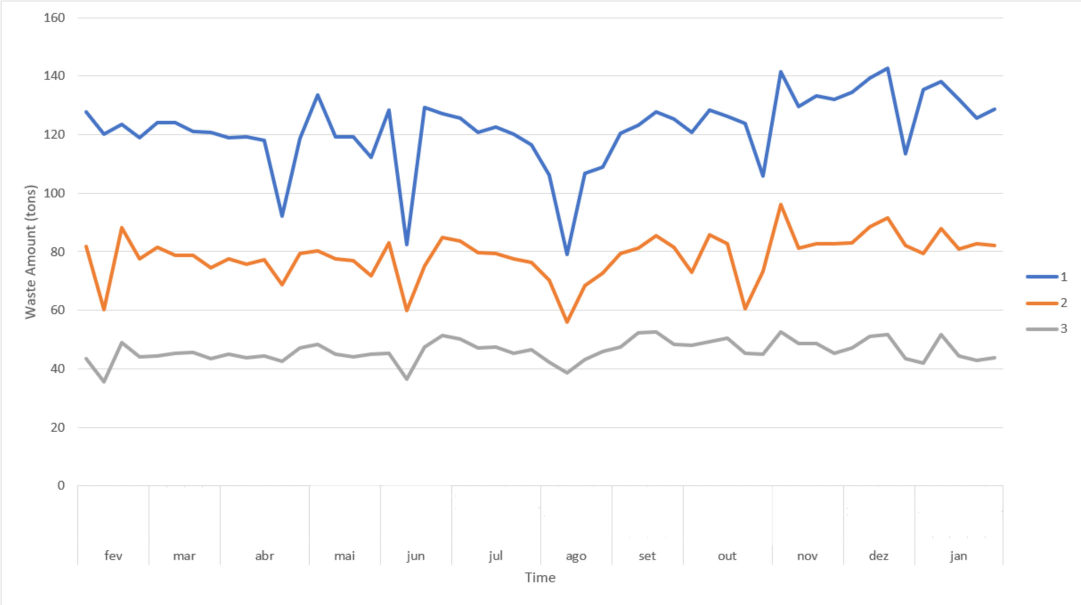


Figure 23 – Linear Plot by Segment *k*=3 (Paper Waste)

Normalization was also used using the Min-Max method to verify the discrepancies between the different clusters, resulting from DTW and K-means, considering the same range [0.1].

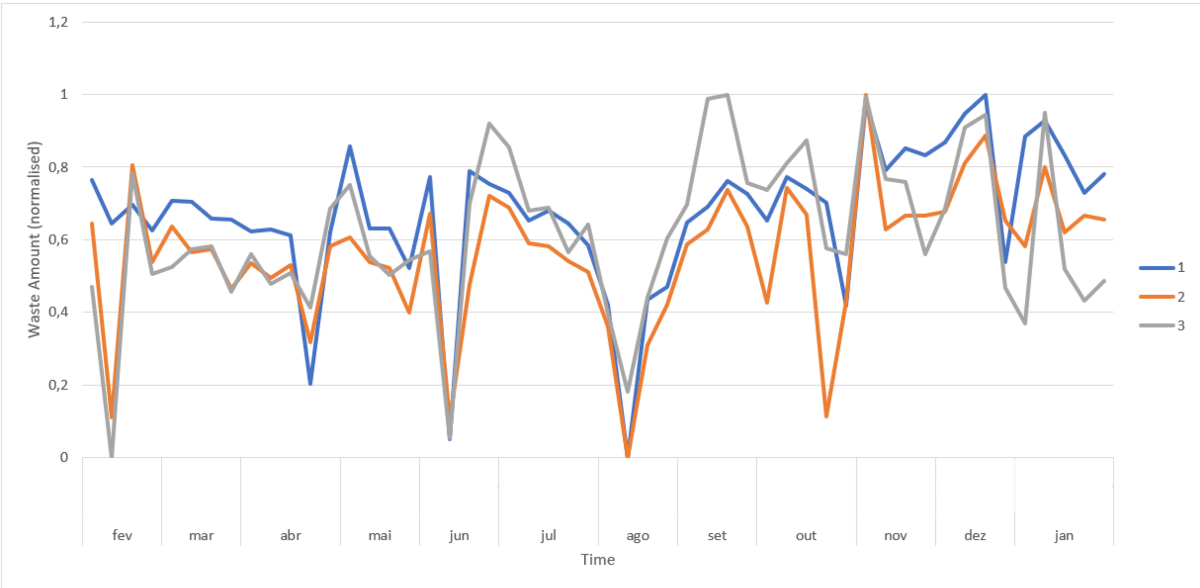


Figure 24 – Linear Plot by Segment *k*=3 after normalization (Paper Waste)

As seen in plastic waste, all clusters exhibit notable variability during the year. Especially at the time of the summer holidays, July-August, and during the festivities of Easter, Popular Saints and Halloween, April, June and October, respectively, with a substantial decrease.

5.2.2.3. Glass Waste

As mentioned earlier, glass waste was initially aggregated monthly rather than weekly, as in other types of waste. However, the same procedure of the previous waste types was applied through the

TS Similarity node with a Cluster Constellation Plot and the Cluster Dendrogram charts to verify the number potential of clusters.

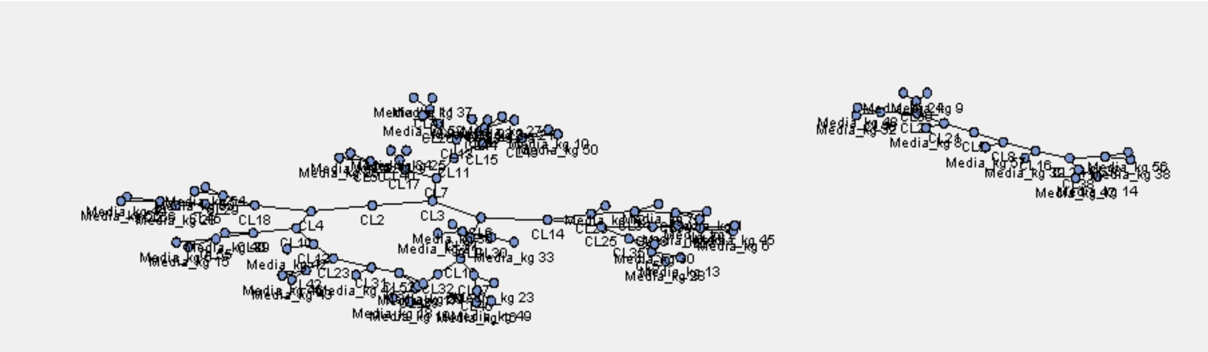


Figure 25 – Cluster Constellation Plot for Glass Waste

Figure 26 – Cluster Dendogram for Glass Waste

By analysing the different graphics above, it is possible to understand that there are at least two clusters. However, the Constellation Plot shows that more clusters can be formed. To understand the best number of clusters to use for the analysis is by use the Elbow Graph. The TS Similarity node combined all the time series through the Dynamic Time Warping method.

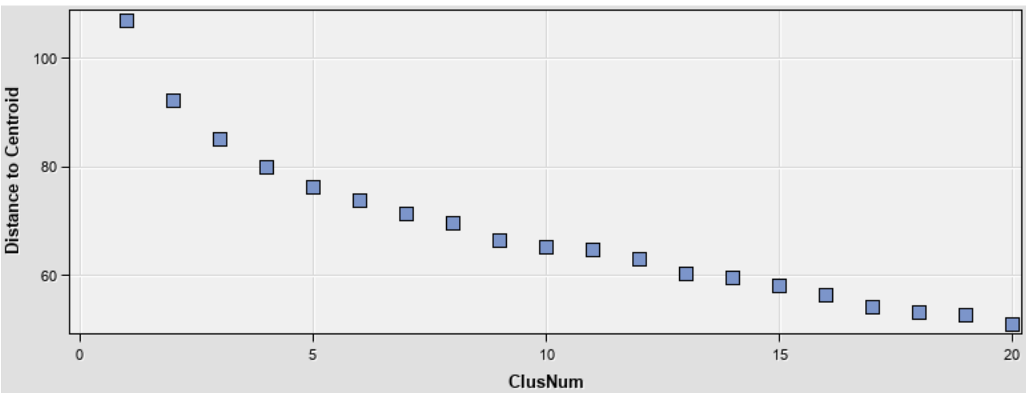


Figure 27 - Elbow Graph for Glass Waste

According to the methodology, the Elbow Graph indicates that the best number of clusters is possibly four. However, further analysis and graphs were made to make the most appropriate choice regarding the ideal number of clusters.

The difference between three and four clusters was analysed because, according to the Elbow Graph, there is also a possibility of considering three clusters. Considering the tables and Segment Size charts below, it was found that there is a better distribution considering four clusters:

Table 15 - Number of circuits in each segment, *k*=3 and *k*=4 (Glass Waste)

Segment ID	**Number of circuits (k=3)**	**Segment ID**	**Number of circuits (k=4)**
1	28	1	18
2	22	2	19
3	7	3	13
-	-	4	7

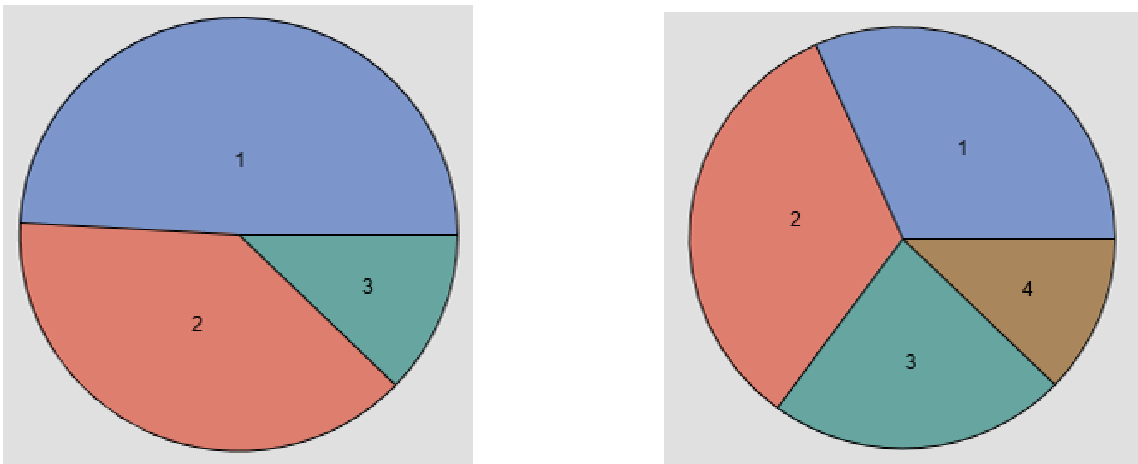


Figure 28 – Segment Size, *k*=3 and *k*=4 (Glass Waste)

As analysed in mixed waste, a Linear Plot by Segment was constructed to examine the differences between the different clusters. It was observed that k=4 shows better results when compared with k=3.

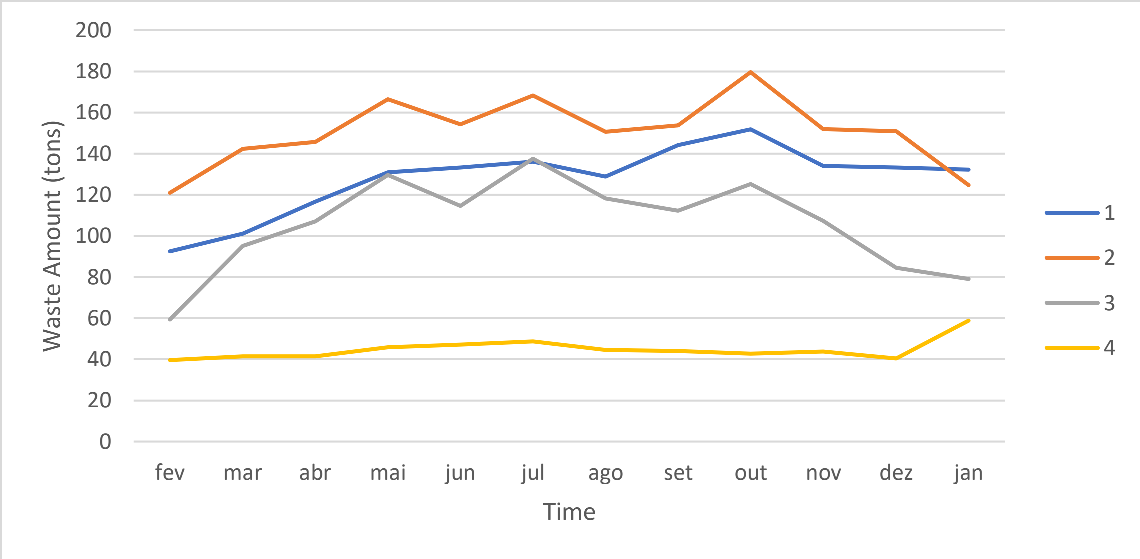


Figure 29 – Linear Plot by Segment *k*=4 (Glass Waste)

Normalization was also used using the Min-Max method to verify the discrepancies between the different clusters, resulting from DTW and K-means, considering the same range [0.1].

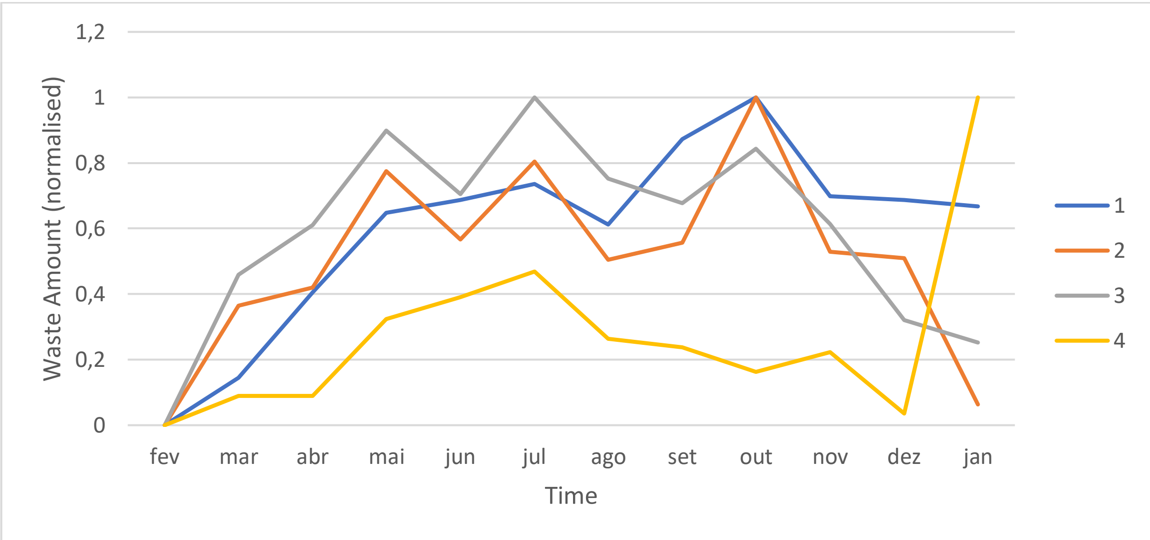


Figure 30 – Linear Plot by Segment *k*=4 after normalization (Glass Waste)

Although the data were aggregated monthly, it is possible to verify that there is some variability in the production of glass waste during the year. Like other types of waste, the seasonal season indicates a decrease, especially in the July-September summer holidays. And then, it was found that although in most clusters there is a decrease in the time of Christmas and New Year, December-January, in cluster 4, there is an increase in the same period.

5.3. Evaluation

5.3.1. Mixed Waste

In this section is explained the results obtained for the mixed waste. Each cluster was analysed according to the PowerBI tool, which provides an overview of the location of the clusters and the total amount per each variable approach.

The PowerBI dashboard is designed to provide easy access to the choice of cluster to analyse. In this way, a slicer was added that allows the choice of the cluster to view on the map.

For mixed waste, the number of clusters chosen was 3. Moreover, below the dashboard for 3 clusters can be found:

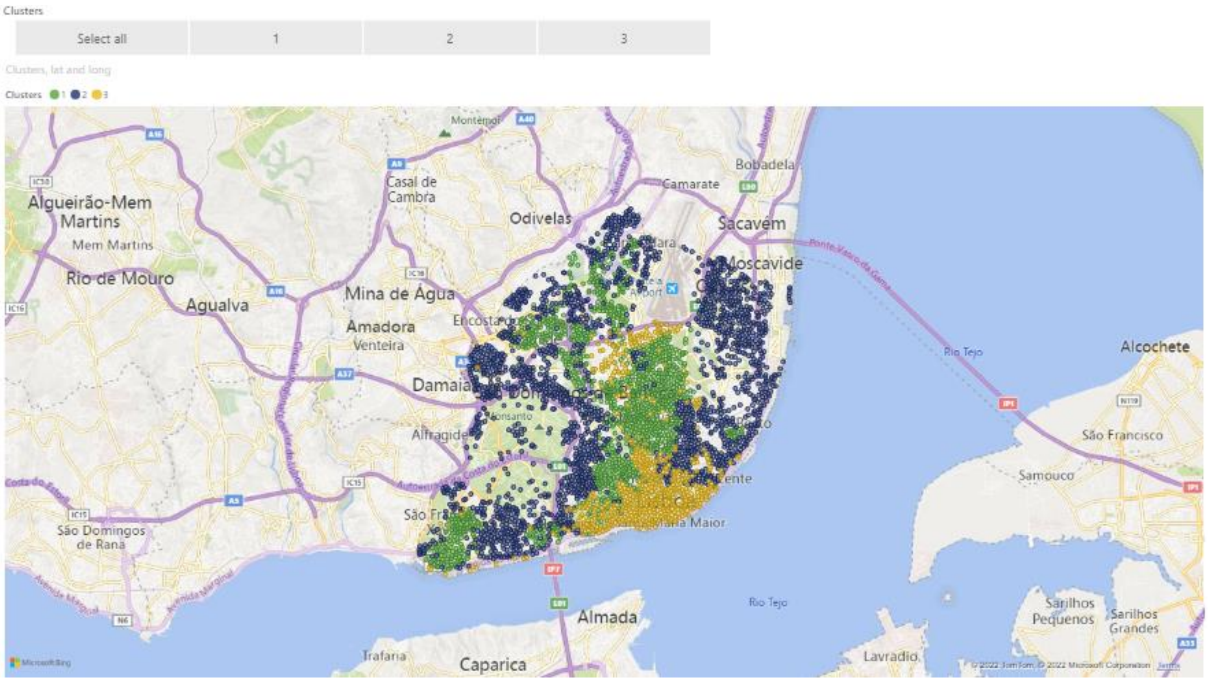


Figure 31 – PowerBI Dashboard 3 Clusters, Mixed Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow).

The following tables expresses a summary of the mixed waste characteristics that support the analysis:

Table 16 - Summary table Mixed Waste

Cluster	1	2	3
Total Waste Collected (ton)	39 159	65 007	25 951
Waste Collection Points	2 707	3 813	3 219
Number of Circuits	31	43	23
Maximum Clusters' Capacity (m3)	2 125	2 722	3 323

Table 17 - Amount per each Variable Mixed Waste, Z-Score

		Clusters		
	Variables	**1 *(n=31)***	**2 *(n=43)***	**3 *(n=23)***
Citizen's Age - Male and Female	HM_0_14_years	0,843	0,262	-1,105
	HM_15_24_years	0,930	0,127	-1,058
	HM_25_64_years	-0,379	-0,755	1,134
	HM_65_and_more_years	-0,243	1,099	-0,856
Citizen's Age - Male	H_0_14_years	0,755	0,379	-1,134
	H_15_24_years	0,782	0,345	-1,127
	H_25_64_years	-0,446	-0,700	1,145
	H_65_and_more_years	0,007	0,997	-1,003
Citizen's Age - Female	M_0_14_years	0,957	0,081	-1,038
	M_15_24_years	1,113	-0,289	-0,824
	M_25_64_years	-0,169	-0,905	1,074
	M_65_and_more_years	-0,777	1,128	-0,351
Level of Education	No Schooling level	-0,112	-0,939	1,051
	Basic Education	-0,915	1,067	-0,152
	1st Cycle	-1,021	0,977	0,044
	2nd Cycle	-0,794	1,123	-0,329
	3rd Cycle	-0,735	1,139	-0,403
	Secondary and post-secondary education	-0,783	-0,343	1,126
	University Education	1,033	-0,963	-0,070
Household Living Space	Size Less than 30 m	-0,710	-0,434	1,144
	Size 30 m to 39 m	-0,758	-0,375	1,133
	Size 40 m to 49 m	-0,853	-0,247	1,100
	Size 50 m to 59 m	-1,119	0,312	0,807
	Size 80 m to 99 m	0,007	0,996	-1,004
	Size 100 m to 119 m	1,084	-0,196	-0,887
	Size 60 m to 79 m	-0,861	1,097	-0,236
	Size 120 m to 149 m	1,122	-0,324	-0,798
	Size 150 m to 199 m	1,139	-0,732	-0,408
	Size 200 m or more	0,872	-1,091	0,219
House rent - proximity of family income	Rent Less than 20 euros	-0,385	1,135	-0,750
	Rent 20 to 49,99 euros	-0,637	1,153	-0,516
	Rent 50 to 99,99 euros	-0,484	1,150	-0,665
	Rent 100 to 199,99 euros	-1,020	0,979	0,042
	Rent 200 to 399,99 euros	-1,154	0,604	0,550
	Rent 400 to 649,99 euros	-0,547	-0,607	1,154
	Rent 650 to 999,99 euros	1,051	-0,939	-0,112

	Rent 1000 or more euros	0,969	-1,029	0,060
Family Size	Family size with 1 person	-0,378	-0,756	1,134
	Family size with 2 people	-0,334	1,124	-0,790
	Family size with 3 people	0,223	0,869	-1,093
	Family size with 4 people	0,758	0,375	-1,133
	Family size with 5 or more people	0,533	-1,154	0,620
Cultural and Social aspects	Rest Coffee Shop Bar Market	0,804	-1,120	0,316
	Culture Spot	1,083	-0,889	-0,194
	Beds Tourist Establishments	-0,405	-0,734	1,139
	Beds Local accommodation	0,002	-1,001	0,999
	Cultural and Sporting events	0,124	-1,056	0,932
Rent - proximity of family income	Price per sqm	0,033	-1,016	0,983

Each cluster was analysed to understand the results concerning socioeconomic, demographic, and context variables that influence waste production, taking into consideration the literature review.

- **Cluster 1:**

Of all clusters, it shows the most significant variability of waste collection throughout the year, demonstrating peaks in waste production at Christmas time and a significant decrease in the time of the summer holidays and Popular Saints, having in consideration the average of the three years under analysis, 2017 to 2019 (Figure 12).

Although it demonstrates that it has the smallest capacity in cubic meters and has the smallest number of collection points, other clusters collected the least amount of waste (Table 16).

In Lisbon, the cluster is located mainly in the centre of Belém, Alcântara, Amoreiras, Campo Pequeno, Alvalade, Carnide, Lumiar, and Madre de Deus, generally covering the most central area of Lisbon (Figure 44).

This cluster is characterized mainly by the most significant number of residents with higher education and the lowest number of residents with a level of education from elementary to secondary education. The cluster has the most extensive number of younger citizens, between 0 and 24 years, predominantly women from 15 to 24 years (Table 17).

As for buildings, the cluster contains the largest household area, from 100 m2, whose rents prove to be the most expensive compared to the other clusters, starting on 650 euros per month. Because this cluster is located essentially in the residential areas of central Lisbon, it has a price per square meter within the average (Table 17).

- **Cluster 2:**

Of all clusters, cluster 2 shows less variability in waste production over time, only demonstrating a peak in Christmas and a decrease in the summer holidays. The rest of the year remains constant, considering the average of the three years under analysis, 2017 to 2019 (Figure 12).

Cluster 2 is the cluster that has the most waste collection quantity since it also has the most extensive number of circuits and waste collection points (Table 16).

The cluster shows to be located mainly in the peripheral area of Lisbon, with greater coverage by Industry, around Olivais, Santa Clara, Benfica, São Domingos de Benfica, Ajuda, Campo de Ourique, and Penha de França (Figure 44).

The cluster is characterized by containing the most significant number of residents over 65 years, that is, the oldest population of Lisbon, predominantly men. With the highest number of residents with a basic education level and the lowest with a higher education level. The household comprises 2 or 3 people (Table 17).

As for buildings, the household area is only between 60 and 99 square meters. The value of rents within this cluster is shown to be the lowest in Lisbon, from less than 20 euros to 400 euros, which crosses the lowest price per square meter (Table 17). This cluster is also characterized for showing the lowest number of restaurants, coffee shops and bars, less available local accommodation, and touristic beds along with less cultural and sport events.

- **Clusters 3:**

Cluster 3 demonstrates a different behaviour from the other clusters, demonstrating several small peaks throughout the year, at Easter, Popular Saints, and Christmas, showing to be influenced by the festivities, considering the average of the three years under analysis, 2017 to 2019. At the time of the summer holidays has a slight decrease, but nothing compared to the other two clusters (Figure 12).

Given the variation in waste production demonstrating increased waste production during the festivities, it is the cluster with the smallest amount of waste collected and the smallest number of circuits. Otherwise, it holds the highest capacity in terms of cubic meters (Table 16).

Cluster 3 is located mainly in Lapa, Baixa, Bairro Alto, and São Jorge's Castle, a typical tourist area. Given that, it shows that it has a greater number of local accommodation beds, cultural points, and cultural and sporting events (Figure 44).

This cluster is characterized by having more residents between 25 and 64 years old, predominantly men. The level of education verified is secondary and post-secondary education or with no education. The household is mainly composed of one person (Table 17).

For buildings, the cluster consists of the most significant number of buildings with the smallest living area, only up to 59 square meters. The cluster also encompasses the buildings with the highest price per square meter compared to the other clusters; however, it is not reflected in the value of rents, having only greater relevance in rents between 400 and 649 euros. This cluster also has the largest number of available tourist establishments and local accommodation beds, restaurants, bars, cafes, and markets. The waste production dynamic is strongly associated with tourist, holidays, and cultural/sports events activity (Table 17).

5.3.2. Recycled Waste

5.3.2.1. Plastic Waste

In this section is explained the results obtained for the plastic waste.

For plastic waste, the number of clusters chosen was 5. As mentioned above, an outlier circuit was identified that would be analysed along with the clusters found to understand the circuit's impact that has collected much waste. Moreover, below the dashboard for 5 clusters and the circuit EESA01MF can be found:

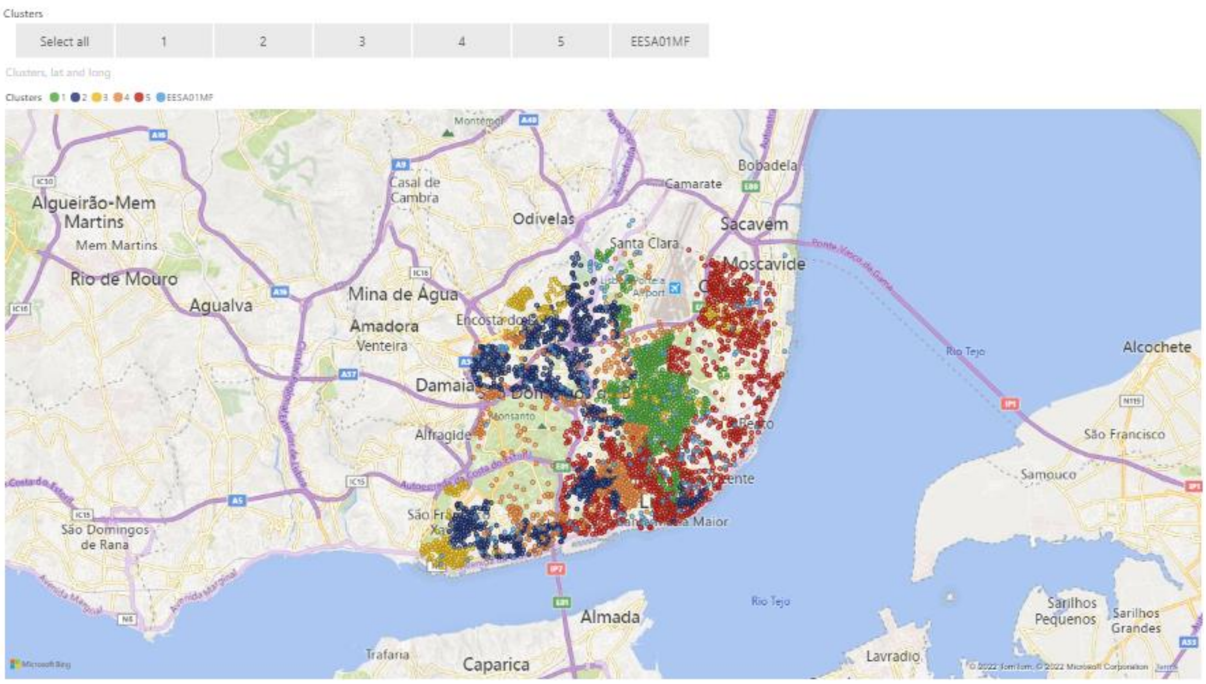


Figure 32 – PowerBI Dashboard 5 Clusters and an Outlier, Plastic Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow, Cluster 4 – orange, Cluster 5 – Red, Circuit EESA01MF – light blue).

The following tables expresses a summary of the plastic waste characteristics that support the analysis:

Table 18 - Summary table Plastic Waste

Cluster	1	2	3	4	5	EESA01MF
Total Waste Collected (ton)	2 900	3 697	492	1 109	1 583	457
Waste Collection Points	1394	1571	588	762	1857	262
Number of Circuits	12	16	5	7	12	1
Maximum Clusters' Capacity (m3)	2 113	2 665	590	895	1 575	1 223

Table 19 - Amount per each Variable Plastic Waste, Z-Score

Variables		Cluster 1 (n=12)	Cluster 2 (n=16)	Cluster 3 (n=5)	Cluster 4 (n=7)	Cluster 5 (n=12)	EESA01MF
Citizen's Age - Male and Female	HM_0_14_years	0,073	0,584	1,332	0,285	-1,462	-0,812
	HM_15_24_years	0,413	0,209	1,635	-1,245	-0,350	-0,662
	HM_25_64_years	1,266	-0,701	-1,475	-0,190	0,337	0,762
	HM_65_and_more_years	-1,841	0,589	0,851	0,466	0,366	-0,431
Citizen's Age - Male	H_0_14_years	-0,262	0,856	1,068	0,604	-1,385	-0,880
	H_15_24_years	0,006	0,391	1,773	-1,021	-0,547	-0,602
	H_25_64_years	1,192	-0,921	-1,397	-0,107	0,552	0,681
	H_65_and_more_years	-1,737	0,864	1,013	0,216	0,043	-0,400
Citizen's Age - Female	M_0_14_years	0,425	0,249	1,490	-0,057	-1,450	-0,658
	M_15_24_years	1,067	-0,068	1,166	-1,402	-0,037	-0,727
	M_25_64_years	1,319	-0,255	-1,537	-0,275	-0,114	0,862
	M_65_and_more_years	-1,824	0,156	0,601	0,622	0,855	-0,411
Level of Education	No Schooling level	1,640	-1,318	-0,524	-0,366	0,311	0,258
	Basic Education	-1,287	-0,387	0,177	-0,240	1,762	-0,025
	1st Cycle	-1,193	-0,503	0,129	-0,245	1,802	0,009
	2nd Cycle	-1,193	-0,377	0,247	-0,355	1,798	-0,120
	3rd Cycle	-1,535	-0,127	0,228	-0,135	1,599	-0,031
	Secondary and post-secondary education	0,609	-1,099	-1,040	-0,510	1,041	0,999
	University Education	1,106	0,558	-0,064	0,299	-1,835	-0,063
Household Living Space	Size Less than 30 m	-0,367	-0,410	-1,625	1,052	0,531	0,819
	Size 30 m to 39 m	-0,522	-0,471	-1,517	1,013	0,576	0,921
	Size 40 m to 49 m	-0,611	-0,659	-1,306	1,240	0,494	0,840
	Size 50 m to 59 m	-0,869	-0,920	-0,712	0,483	1,566	0,452
	Size 80 m to 99 m	-0,533	0,892	-0,666	-1,334	1,239	0,402
	Size 100 m to 119 m	1,396	0,565	0,347	-0,536	-1,500	-0,272
	Size 60 m to 79 m	-1,226	-0,294	-0,478	-0,031	1,764	0,265
	Size 120 m to 149 m	1,062	0,834	0,572	-0,372	-1,555	-0,541
	Size 150 m to 199 m	0,958	0,295	1,074	-0,128	-1,551	-0,647
	Size 200 m or more	0,368	-0,407	1,739	0,076	-1,103	-0,672
House rent - proximity of family income	Rent Less than 20 euros	-1,065	-0,221	0,810	-0,721	1,580	-0,383
	Rent 20 to 49,99 euros	-1,588	0,323	0,777	-0,466	1,218	-0,264
	Rent 50 to 99,99 euros	-1,470	0,168	0,730	-0,527	1,385	-0,285
	Rent 100 to 199,99 euros	-1,671	0,078	0,565	-0,320	1,334	0,015
	Rent 200 to 399,99 euros	-0,757	0,611	-1,662	0,370	0,471	0,967
	Rent 400 to 649,99 euros	0,497	0,204	-1,707	0,631	-0,628	1,003
	Rent 650 to 999,99 euros	1,660	-0,024	-0,248	0,142	-1,469	-0,061

	Rent 1000 or more euros	1,343	-0,632	0,825	0,359	-1,255	-0,641
Family Size	Family size with 1 person	0,490	-0,197	-1,319	1,414	-0,867	0,479
	Family size with 2 people	-1,272	1,369	-0,810	-0,159	-0,038	0,909
	Family size with 3 people	-0,587	-0,022	0,790	-1,221	1,520	-0,479
	Family size with 4 people	0,037	0,486	1,624	-1,192	-0,125	-0,831
	Family size with 5 or more people	0,336	-1,104	1,427	-0,902	0,750	-0,507
Cultural and Social aspects	Rest Coffee Shop Bar Market	1,047	1,058	-1,092	0,221	0,009	-1,242
	Culture Spot	0,712	0,296	-0,564	-0,436	1,392	-1,399
	Beds Tourist Establishments	1,156	0,230	-1,005	-0,085	0,981	-1,277
	Beds Local accommodation	0,373	1,185	0,006	0,526	-0,339	-1,751
	Cultural and Sporting events	0,271	0,564	-0,840	-0,051	1,424	-1,368
Rent - proximity of family income	Price per sqm	0,291	-0,762	0,346	1,659	-1,146	-0,388

Each cluster and the outlier were analysed to understand the results concerning socioeconomic, demographic, and context variables that influence waste production, taking into consideration the literature review.

- **Cluster 1:**

Of all clusters, cluster 1 is the most stable during the period, demonstrating peaks in waste production at Christmas time and a decrease in the time of the summer holidays and Popular Saints, considering the average of the three years under analysis, 2017 to 2019 (Figure 45).

According to the summary table, it is a cluster whose collected amount goes according to the number of waste collection points, circuits, and cluster capacity. Which means, there are no significant differences that make attention (Table 18).

It is located mainly in a more central city area, namely in Areeiro, Campo Pequeno and Alvalade (Figure 46).

The cluster is characterized by having more residents between 25 and 64 years old, predominantly men and fewer residents over 65 years. It also has the highest number of residents without any level of education and, at the same time, has the highest number of residents with higher education (Table 30

As for buildings, the household area is only between 100 and 149 square meters. The value of rents within this cluster is shown to be the highest in Lisbon, from 650 euros, being one of the clusters whose value per square meter is one of the highest (Table 19).

This cluster also covers the most significant number of local accommodation beds and restaurants, coffee shops, bars and markets, as it is located in a more central and bustling area of Lisbon (Table 19).

- **Cluster 2:**

Cluster 2 has erratic behaviour throughout the year, with a considerable decrease in the time of the summer holidays and a peak at Christmas time. It should also be noted that this cluster shows a minor variation in waste production during Popular Saints festivities, considering the average of the three years under analysis, 2017 to 2019 (Figure 45).

This cluster shows the largest amount of waste collected; however, it also has the largest number of circuits, collection points and capacity in cubic meters (Table 18).

The areas to which this cluster is inserted are Restelo, Santo Amaro, Campo de Ourique, Benfica, Carnide and Telheiras, located more in the west of Lisbon (Figure 46).

The cluster is characterized by having a good population distribution concerning age and level of education, with only the smallest number of residents with no level of education. In terms of households, it is the cluster with the biggest number of families composed of 2 people (Table 19).

It is also characterized by covering a significant number of schools, from pre-school to secondary school, and also covers the largest number of cultural points (Table 19).

- **Cluster 3:**

Cluster 3 has irregular behaviour over the period, as other clusters. Nevertheless, of all clusters is the one that produces the least amount of waste, with a total of 492 tons, not getting far behind the EESA01MF outlier circuit with 457 tons. This is also because it has the smallest number of collection points and the smallest capacity in cubic metres (Figure 45 and Table 18).

This cluster is relatively distributed in different city areas, such as Belém and Olivais and a small area of Carnide (Figure 46).

It is characterized by the significant number of residents, from 0 to 24 years and over 65 years, where from 15 to 24 are notably men and from 0 to 14 are women. Regarding the level of education, the cluster presents fewer people with secondary and post-secondary education (Table 19).

The cluster also holds many buildings with a larger living area, from 150 square meters. Furthermore, whose house rent is from 20 euros to 199 euros, as well as 1000 euros or more. The household consists mainly of 4 or more people (Table 19).

- **Cluster 4:**

Cluster 4 has irregular behaviour over the period, as other clusters, being one of the little clusters within the plastic waste (Figure 45 and Table 18).

It is distributed by Rato, Picoas, Alcântara and Ajuda, Bairro da Boavista and some points by Campo Grande, namely the Cidade Universitária (Figure 46).

It is characterized by a larger number of small buildings with a household area from 30 to 49 square meters, with the highest price per square meter. The household of this cluster is composed of 1 person (Table 19).

- **Cluster 5:**

Cluster 5 has erratic behaviour over the period, as other clusters, demonstrating higher peaks at Christmas and during May, such as a descent into the summer holidays (Figure 45).

Compared to cluster 1, which has the same number of circuits, cluster 5 collected much less waste than cluster 1, although it also has a larger number of collection points. This is due to the capacity of the containers of the clusters themselves, to which the capacity of cluster 5 is smaller (Table 18).

The cluster is located essentially in Olivais, Avenida da Liberdade and Baixa, representing the east and south of Lisbon (Figure 46).

The cluster is characterized by having the smallest number of young people, from 0 to 14 years old, and a significant number of people with a basic education level up to secondary and post-secondary school. At the same time aggregates the smallest number of people with higher education. The cluster household is 3 people (Table 19).

As for buildings, the cluster admits household areas between 50 and 79 square meters. The value of rents within this cluster is shown to be the lowest, until 199 euros, the clusters whose value per square meter is one of the lowest (Table 19).

The cluster contains many tourist establishments and cultural and sporting events (Table 19).

- **Circuit EESA01MF:**

This circuit proves to be constant, with some waste collection peaks in October and Christmas (Figure 45).

Being only one circuit, the amount of waste collection is greatly reduced compared to the other clusters but has a maximum capacity higher than clusters 3 and 4. This circuit takes place daily and only collects plastic waste from eco-islands (mentioned in the literature review). In this way, it stands out (Table 18).

This circuit is located throughout Lisbon, especially in the central and northern areas of the city (Figure 46).

Being widely distributed, it is characterized by adding residents aged 25 to 64. As for the level of education, it includes a greater number of people with secondary and post-secondary education (Table 19).

It is also characterized by encompassing housing whose income varies between 200 and 650 euros (Table 19).

5.3.2.2. Paper Waste

In this section is explained the results obtained for the paper waste.

For paper waste, the number of clusters chosen was 3. As mentioned above, an outlier circuit was identified that would be analysed along with the clusters found to understand the circuit's impact that has collected much waste. Moreover, below the dashboard for 3 clusters and the circuit P3101 can be found:

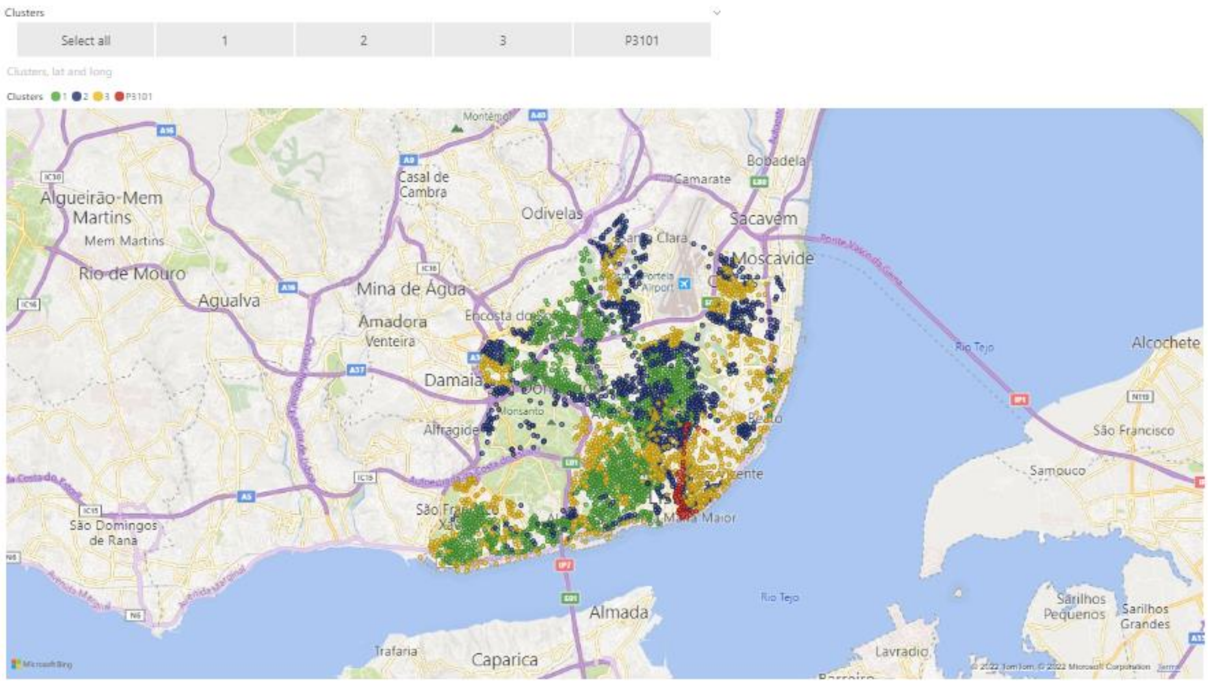


Figure 33 – PowerBI Dashboard 3 Clusters and an Outlier, Paper Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow, Circuit P3101 – Red).

The following table expresses a summary of the plastic waste characteristics that support the analysis:

Table 20 - Summary table Paper Waste

Cluster	1	2	3	P3101
Total Waste Collected (ton)	6 333	4 077	2 398	646
Waste Collection Points	2 678	1 755	1 542	97
Number of Circuits	32	21	16	1
Maximum Clusters' Capacity (m3)	4 915	3 312	1 819	73

Table 21 - Amount per each Variable Paper Waste, Z-Score

Variables		Cluster			
		1 (n=32)	2 (n=21)	3 (n=16)	P3101
Citizen's Age - Male and Female	HM_0_14_years	0,728	0,489	0,254	-1,472
	HM_15_24_years	0,742	0,704	-0,049	-1,397
	HM_25_64_years	-0,699	-0,402	-0,383	1,484
	HM_65_and_more_years	0,643	0,240	0,593	-1,476

Citizen's Age - Male	H_0_14_years	0,743	0,464	0,264	-1,471
	H_15_24_years	0,762	0,576	0,104	-1,441
	H_25_64_years	-0,726	-0,410	-0,342	1,479
	H_65_and_more_years	0,691	0,301	0,489	-1,481
Citizen's Age - Female	M_0_14_years	0,722	0,528	0,217	-1,467
	M_15_24_years	0,691	0,923	-0,379	-1,235
	M_25_64_years	-0,662	-0,367	-0,459	1,489
	M_65_and_more_years	0,516	0,006	0,879	-1,401
Level of Education	No Schooling level	-0,776	-0,293	-0,399	1,467
	Basic Education	-0,998	-0,146	1,387	-0,242
	1st Cycle	-1,187	-0,291	1,194	0,284
	2nd Cycle	-0,728	0,107	1,375	-0,754
	3rd Cycle	-0,422	0,040	1,363	-0,981
	Secondary and post-secondary education	-0,764	-0,384	-0,324	1,471
	University Education	1,261	0,307	-0,990	-0,578
Household Living Space	Size Less than 30 m	-0,657	-0,679	-0,112	1,447
	Size 30 m to 39 m	-0,678	-0,690	-0,067	1,435
	Size 40 m to 49 m	-0,766	-0,716	0,108	1,374
	Size 50 m to 59 m	-1,005	-0,608	0,408	1,205
	Size 80 m to 99 m	-0,046	0,811	0,629	-1,394
	Size 100 m to 119 m	0,845	0,848	-0,590	-1,103
	Size 60 m to 79 m	-1,131	-0,093	1,306	-0,081
	Size 120 m to 149 m	0,991	0,660	-0,494	-1,157
	Size 150 m to 199 m	1,098	0,519	-0,478	-1,139
	Size 200 m or more	1,143	0,150	0,000	-1,293
House rent - proximity of family income	Rent Less than 20 euros	-0,264	0,415	1,096	-1,248
	Rent 20 to 49,99 euros	-0,308	0,101	1,302	-1,095
	Rent 50 to 99,99 euros	-0,279	0,237	1,218	-1,176
	Rent 100 to 199,99 euros	-0,897	-0,538	1,379	0,056
	Rent 200 to 399,99 euros	-0,961	-0,587	0,257	1,291
	Rent 400 to 649,99 euros	-0,531	-0,557	-0,409	1,497
	Rent 650 to 999,99 euros	0,960	0,612	-1,270	-0,302
	Rent 1000 or more euros	1,214	0,409	-0,669	-0,955
Family Size	Family size with 1 person	-0,332	-0,763	-0,377	1,472
	Family size with 2 people	0,312	0,154	0,943	-1,410
	Family size with 3 people	0,171	0,768	0,515	-1,454
	Family size with 4 people	0,701	0,660	0,077	-1,438
	Family size with 5 or more people	-1,046	-0,192	-0,123	1,361
Cultural and Social aspects	Rest Coffee Shop Bar Market	1,230	0,146	-0,178	-1,198
	Culture Spot	1,264	0,183	-0,323	-1,124

	Beds Tourist Establishments	1,072	0,141	0,134	-1,346
	Beds Local accommodation	1,329	0,078	-0,359	-1,048
	Cultural and Sporting events	1,116	-0,256	0,383	-1,242
Rent - proximity of family income	Price per sqm	-0,028	-0,880	-0,498	1,406

Each cluster and the outlier were analysed to understand the results concerning socioeconomic, demographic, and context variables that influence waste production, taking into consideration the literature review.

- **Cluster 1:**

Cluster 1, like the remaining clusters, varies significantly over the period, with substantial decreases followed by large increases in festive seasons such as Easter, Popular Saints, Halloween and Christmas, and the summer holidays. These variations may be related to the festivities as well as national holidays (Figure 47).

This cluster is the one that has the most considerable amount of waste collected, as well as the one that presents the highest number of collection points and circuits (Table 20).

The cluster is located mainly in Belém, Alcântara, Campo de Ourique, Rato, Picoas, Areeiro, Alvalade, São Domingos de Benfica, Carnide and Telheiras, covering the west and north of Lisbon (Figure 48).

It is characterized by a large number of residents from 0 to 24, as well as by over 65 years. The population that is part of this cluster has mostly higher education. Moreover, the household is mainly composed of 4 people (Table 21).

As for buildings, this cluster features a large number of buildings with the largest household living space, from 150 square meters or more (Table 21).

It is also characterized by having the biggest number of tourist establishments, local accommodation beds, cultural points, cultural and sporting events, restaurants, coffee shop, bars and markets. It also includes the largest number of schools, from pre-school to secondary schools (Table 21).

- **Cluster 2:**

Cluster 2, like the remaining clusters, varies significantly over the period, with substantial decreases followed by large increases in festive seasons such as Easter, Popular Saints, Halloween and Christmas, and the summer holidays. These variations may be related to the festivities as well as national holidays (Figure 47).

The cluster is located mainly in Arroios, Campo Pequeno, Entrecampos, Alameda, Madre de Deus, Lumiar, Benfica, and Olivais do Sul, focusing mainly on the city centre area (Figure 48).

The characteristics that best describe the cluster are household size, with areas between 80 and 119 square meters, the lowest price per square meter and the high number of women between 15 and 24 years (Table 21).

- **Cluster 3:**

Cluster 3, like the remaining clusters, varies significantly over the period, with substantial decreases followed by large increases in festive seasons such as Easter, Popular Saints, Halloween and Christmas, and the summer holidays. These variations may be related to the festivities as well as national holidays (Figure 47).

Of all clusters, the one that collects the least amount of waste, not considering the outlier circuit, and the one that represents the smallest number of collection points, circuits and capacity (Table 20).

The cluster is in Caselas, Campolide, São Jorge's Castle, Avenida da Liberdade, Olivais, Benfica and Ameixoeira, distributed throughout Lisbon, covering industrial areas (Figure 48).

It is characterized essentially by having residents with a basic education level and a household composed of 2 people (Table 21).

It contains the largest number of houses with a household size of 60 to 79 square meters and the lowest rents, up to 200 euros (Table 21).

- **Circuit P3101:**

The circuit also proves variable over the period. However, it remains stable between September and November and has a significant drop at Christmas. Being only one circuit, the absolute amount of waste collection is greatly reduced compared to the other clusters (Table 20 and Figure 47).

The circuit only encompasses a small area from Terreiro do Paço to Penha de França (Figure 48).

However, it is characterized by mainly having residents between 25 and 64 years. And residents with no education level or higher education. Furthermore, the household consists only one person and 5 or more (Table 21).

As for buildings, they include areas up to 59 square meters, with rents between 200 and 649 euros. The circuit also consists of buildings with the highest price per square meter (Table 21).

5.3.2.3. Glass Waste

In this section is explained the results obtained for the glass waste.

For glass waste, the number of clusters chosen was 4. As mentioned above, an outlier circuit was identified that would be analysed along with the clusters found to understand the circuit's impact that has collected much waste. Moreover, below the dashboard for 4 clusters and the circuit V1305 can be found:

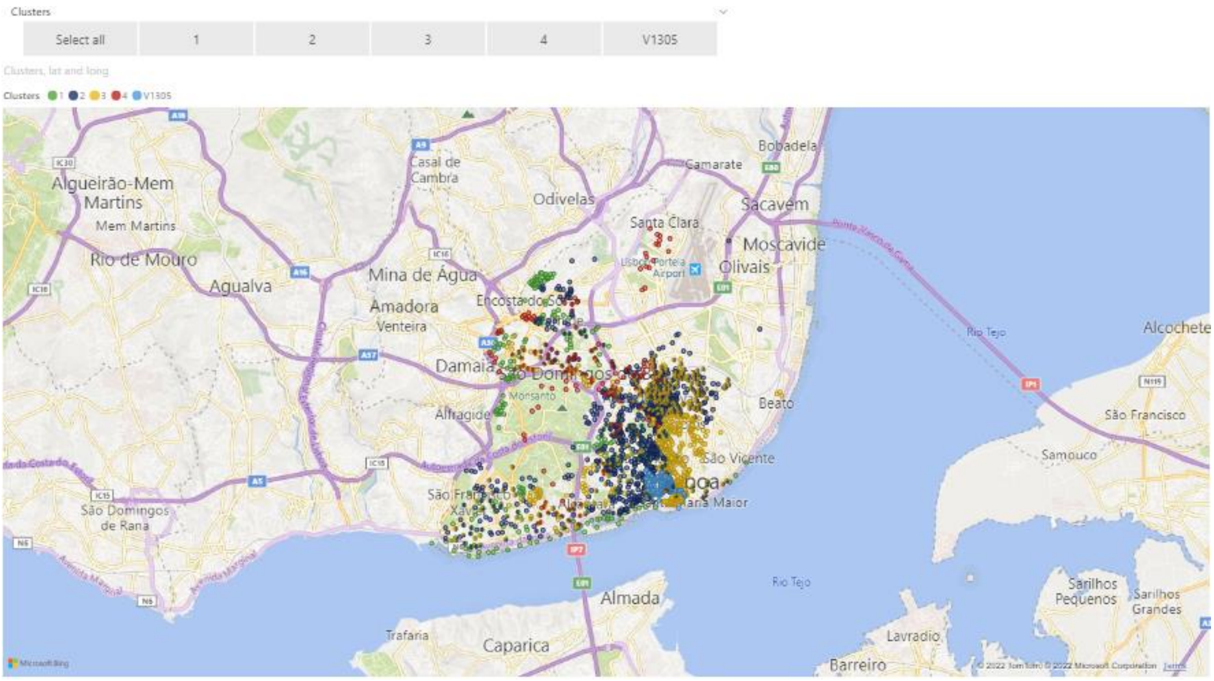


Figure 34 – PowerBI Dashboard 4 Clusters and an Outlier, Glass Waste (Cluster 1 – green, Cluster 2 – blue, Cluster 3 – yellow, Cluster 4 – red, Circuit V1305 – light blue).

The following table expresses a summary of the plastic waste characteristics that support the analysis:

Table 22 - Summary table Glass Waste

Cluster	**1**	**2**	**3**	**4**	**V1305**
Total Waste Collected (ton)	1 534	1 809	1 270	538	653
Waste Collection Points	523	1297	857	185	211
Number of Circuits	18	19	13	7	1
Maximum Clusters' Capacity (m3)	1 105	1 105	697	402	42

Table 23 - Amount per each Variable Glass Waste, Z-Score

Variables		**Cluster 1 *(n=18)***	**Cluster 2 *(n=19)***	**Cluster 3 *(n=13)***	**Cluster 4 *(n=7)***	**Cluster V1305**
Citizen's Age - Male and Female	HM_0_14_years	0,425	0,680	0,165	0,488	-1,758
	HM_15_24_years	0,280	0,634	0,004	0,786	-1,703
	HM_25_64_years	-0,625	-0,342	-0,107	-0,668	1,742
	HM_65_and_more_years	1,015	-0,337	0,072	0,758	-1,509
Citizen's Age - Male	H_0_14_years	0,458	0,616	0,132	0,551	-1,758
	H_15_24_years	0,373	0,570	0,108	0,694	-1,744
	H_25_64_years	-0,567	-0,365	-0,077	-0,726	1,735

	H_65_and_more_years	0,799	-0,110	-0,017	0,914	-1,586
Citizen's Age - Female	M_0_14_years	0,388	0,739	0,205	0,423	-1,756
	M_15_24_years	-0,019	0,772	-0,364	1,070	-1,459
	M_25_64_years	-0,772	-0,273	-0,122	-0,564	1,731
	M_65_and_more_years	1,467	-1,234	0,095	0,200	-0,528
Level of Education	No Schooling level	-0,776	0,287	0,276	-1,162	1,375
	Basic Education	0,024	-0,950	-0,570	-0,161	1,657
	1st Cycle	-0,108	-0,745	-0,449	-0,441	1,743
	2nd Cycle	-0,147	-0,737	-0,861	0,107	1,638
	3rd Cycle	0,886	-1,358	-0,293	1,087	-0,322
	Secondary and post-secondary education	-0,477	-0,649	0,122	-0,689	1,692
	University Education	0,182	0,795	0,355	0,411	-1,743
Household Living Space	Size Less than 30 m	-0,477	-0,276	-0,205	-0,785	1,743
	Size 30 m to 39 m	-0,444	-0,298	-0,185	-0,811	1,738
	Size 40 m to 49 m	-0,442	-0,280	-0,111	-0,881	1,714
	Size 50 m to 59 m	-0,454	-0,373	-0,079	-0,819	1,726
	Size 80 m to 99 m	0,649	-0,300	0,056	1,102	-1,507
	Size 100 m to 119 m	0,415	0,345	0,141	0,831	-1,732
	Size 60 m to 79 m	0,539	-1,708	0,378	0,805	-0,014
	Size 120 m to 149 m	0,238	0,448	0,119	0,903	-1,707
	Size 150 m to 199 m	0,155	0,957	0,151	0,427	-1,690
	Size 200 m or more	0,263	1,532	-0,117	-1,125	-0,553
House rent - proximity of family income	Rent Less than 20 euros	0,530	-0,020	-0,591	1,338	-1,257
	Rent 20 to 49,99 euros	0,346	-0,738	-1,341	0,996	0,738
	Rent 50 to 99,99 euros	0,507	-0,153	-0,479	1,382	-1,257
	Rent 100 to 199,99 euros	-0,185	-0,585	-0,615	-0,376	1,762
	Rent 200 to 399,99 euros	-0,489	-0,994	-0,102	-0,076	1,661
	Rent 400 to 649,99 euros	-0,486	-0,643	0,713	-0,966	1,382
	Rent 650 to 999,99 euros	0,055	0,759	0,659	0,238	-1,712
	Rent 1000 or more euros	0,269	1,322	0,309	-0,577	-1,323
Family Size	Family size with 1 person	-0,420	-0,349	-0,203	-0,777	1,748
	Family size with 2 people	0,511	-0,207	0,340	0,977	-1,621
	Family size with 3 people	0,344	0,248	0,160	0,949	-1,701
	Family size with 4 people	0,403	0,493	0,099	0,745	-1,740
	Family size with 5 or more people	0,317	1,258	0,314	-0,479	-1,411
Cultural and Social aspects	Rest Coffe Shop Bar Market	0,587	0,974	0,550	-0,759	-1,352
	Culture Spot	0,661	1,128	0,002	-0,310	-1,481
	Beds Tourist Establishments	0,385	0,928	0,813	-0,879	-1,248
	Beds Local accommodation	0,698	1,155	0,148	-0,724	-1,277
	Cultural and Sporting events	0,988	0,925	0,127	-0,829	-1,210

Rent - proximity of family income	Price per sqm	-0,248	-0,115	-0,240	-1,057	1,659

Each cluster and the outlier were analysed to understand the results concerning socioeconomic, demographic, and context variables that influence waste production, taking into consideration the literature review.

- **Cluster 1:**

In cluster 1, waste collection is relatively constant throughout the year, and the month of February is when there is less production of glass waste, and October is when there is greater production. There is also a slight decline in the summer holidays, but waste production remains stable at Christmas (Figure 49).

It is one of the clusters with greater waste production. However, compared to cluster 3, there is more waste collection, even though cluster 3 has a more significant number of collection points due to cluster 1 having containers with higher capacity (Table 22)

It is located mainly in Avenidas Novas, Benfica, Carnide, Bairro da Boavista, São Francisco Xavier and Ajuda, taking the west side of Lisbon, around the Monsanto Natural Park (Figure 50).

The cluster is characterized by a population over 65 years, mainly women, and having a greater number of cultural and sporting events (Table 23).

- **Cluster 2:**

Cluster 2 shows the highest variation over time, presenting peak waste production in May, July and October, descending at Christmas (Figure 49).

It is the cluster with the highest production of glass waste since it has a substantial number of circuits and waste collection points; however, it has the same capacity as cluster 1 (Table 22).

It is mainly located in the Saldanha area, Campo Pequeno, Rato, Campo de Ourique and Carnide. It is characterized by having more residents between 0 and 14 years and a higher number of residents with higher education. The household of this cluster consists of 5 or more people (Figure 50).

The buildings that make up this cluster have a household living space size of 150 square meters or more, and rents are the most expensive, starting at 650 euros (Table 23).

This cluster is also characterized by having many restaurants, coffee shop, bars and markets, local accommodation beds and tourist establishments, cultural spots, and cultural and sporting events (Table 23).

- **Cluster 3:**

The temporal behaviour of cluster 3 is very similar to cluster 2, although it demonstrates a smaller amount of glass waste collected because it has fewer circuits and collection points (Table 22 and Figure 49).

It is located in the Baixa Pombalina, Arroios and Avenidas Novas, taking the east-central area of Lisbon (Figure 50).

The cluster has few characteristics compared to the other clusters, indicating only that it has the smallest number of buildings whose rent is from 50 to 99 euros and has the least number of residents with the second cycle of education (Table 23).

- **Cluster 4:**

This cluster has a different temporal behaviour than all other clusters. It is constant throughout the year and only shows a peak in Christmas and New Year, December-January (Figure 49).

It is the cluster with the lowest production of glass waste, with a total of only 538 tons (Table 22)

It has few collection points but is largely located between Benfica and São Domingos de Benfica, a more residential area (Figure 50).

It is characterized by low number of residents aged between 25 and 64 years. As for the level of education, the residents present that they have mostly the third cycle of elementary school. It also consists of households with 2 to 4 members (Table 23).

It has a more significant number of buildings with household sizes between 60 and 150 square meters, with greater emphasis on rents up to 50 euros (Table 23).

- **Circuit V1305:**

The circuit has a higher waste production between April and November, but then there is a decrease between December and February, showing a different behaviour of the remaining clusters (Figure 49).

Being only one circuit, it has more collection points than cluster 4, and the total amount of waste collected is higher, even having a much lower capacity than cluster 4. This circuit occurs daily with door-to-door removal (mentioned in the literature review), which can be an influence (Table 22)

The circuit is in the Bairro Alto area, known for being a nightlife area (Figure 50).

The circuit is characterized by having a large number of residents between 25 and 64 years, with a level of education mainly up to secondary and post-secondary. The household of this circuit has mostly only one person (Table 23).

As for buildings, the circuit is characterized by buildings with a household size of up to 59 square meters, with rents between 100 and 649 euros, and concentrated in a small area with the highest price per square meter (Table 23).

6. CONCLUSIONS

This study has the objective to find the characteristics of the different groups of people who produce waste, and the variables that can better describe these groups, to understand whether their social and economic characteristics impact the amount of waste produced, in the city of Lisbon. Moreover, this document used the design science research methodology, creating and following a conceptual model to identify and understand the different groups, and the socioeconomic and demographic characteristics.

A waste management plan requires data related to its production, the factors that influence it, and ways to predict the amount of waste produced to ensure the proper functioning of services and avoid consequences for the future of municipalities, such as the degradation of public health. The city of Lisbon shows records of increased waste produced in recent years, with a tendency to continue. Thus, it allows the Municipality of Lisbon to reassess its municipal waste management strategies, implementing more objective measures in the different waste collection circuits, depending on the characteristics identified in the different groups and their location.

Through the groups found and their characteristics, it is possible to enrich the existing literature and provide value to the Municipality of Lisbon.

Was concluded that in mixed waste, people over 65 years of age, with a low level of education and whose income is also low, are the ones who produce the most waste in absolute terms, followed by the youngest between 0 and 24 years of age, with a high level of education and with better incomes. On the contrary, it was found that residents between 25 and 64 years old are the ones who produce the least amount of mixed waste. The tourist sites and as many restaurants, coffee shops, bars and markets also influence the production of mixed waste.

As for recycled waste, was concluded, in general, that residents with a higher level of education and better incomes produce more recycled waste, unlike those with less education level and lower incomes that produce less. Residents over 65 years of age produce more paper and glass waste, and those between 25 and 64 years old are the ones who produce the most plastic waste. It is also interesting to note that the areas of Lisbon where there is a greater concentration of tourists and restaurant establishments, coffee shops, bars and markets also contribute to waste recycling. It was also found that there were no significant differences between clusters.

To improve the waste management strategy implemented, awareness measures for poorer residents with lower levels of education should be implemented to resort to recycling waste by inserting the 3'Rs policy: recycle, reduce, and reuse. Measures such as using digital equipment reducing the use of paper, using products more recyclable and reusable to decrease plastic consumption, and preferring products labelled *Eco Label* reducing the amount of mixed waste.

During the data analysis and treatment process, some limitations affected the results.

As mentioned above, data from the provisional 2021 Census were used. This is because, at the time of paper implementation and development, the final data was not available at census block level. For this reason, it was used the provisional data from Census 2021 that was available at parish level. This aspect led to the development of an aggregation approach of socioeconomic data based on the intersection of waste collection points with the census data only available at parish level using the mean as aggregator function for each waste collection circuit.

In this sense, it is recommended for future work the use of Census 2021, with the census information at census block level to perform a more accurate analysis. Also, it can be considered that more updated information on waste collection, namely from 2022, possibly already incorporates possible changes in waste production after the COVID-19 pandemic.

7. BIBLIOGRAPHY

Abdel-Shafy, H. I., & Mansour, M. S. (2018). Solid waste issue: Sources, composition, disposal, recycling, and valorization. *Egyptian journal of petroleum*, *27*(4), 1275–1290.

Amidon, A. (2021, agosto 27). *How to Apply K-means Clustering to Time Series Data*. Medium. https://towardsdatascience.com/how-to-apply-k-means-clustering-to-time-series-data-28d04a8f7da3

Beigl, P., Lebersorger, S., & Salhofer, S. (2008). Modelling municipal solid waste generation: A review. *Waste Management*, *28*(1), 200–214. https://doi.org/10.1016/j.wasman.2006.12.011

Berndt, D. J., & Clifford, J. (1994). *Using dynamic time warping to find patterns in time series. 10*(16), 359–370.

Circuitos Contentores. (2021). https://geodados-cml.hub.arcgis.com/datasets/circuitoscontentores?geometry=-9.324,38.720,-8.997,38.767

Confidencial Imobiliário (Ci): Estatísticas sobre preços reais de venda. (2022). https://www.confidencialimobiliario.com/

Directive 2008/98/EC of the European Parliament and of the Council of 19 November 2008 on waste and repealing certain Directives (Text with EEA relevance). (2008). [Text]. Https://Webarchive.Nationalarchives.Gov.Uk/Eu-Exit/Https://Eur-Lex.Europa.Eu/Legal-Content/EN/TXT/?Uri=CELEX:02008L0098-20180705; Queen's Printer of Acts of Parliament. https://www.legislation.gov.uk/eudr/2008/98/contents

Fasihi, H., & Parizadi, T. (2021). Analyzing household's environmental behavior on solid waste management and its relations with population and housing characteristics (The case: Amlash city, Iran). *Journal of Environmental Management*, *292*, 112686.

Gestão de resíduos na UE: Factos e números | Atualidade | Parlamento Europeu. (2020, dezembro 3). https://www.europarl.europa.eu/news/pt/headlines/society/20180328STO00751/gestao-de-residuos-na-ue-factos-e-numeros

Giannakitsidou, O., Tsagkanos, A., & Giannikos, I. (2016). Correlation of municipal solid waste production and treatment with socioeconomic indexes. *International Journal of Environment and Waste Management*, *18*(4), 303–316.

INE. (2020, novembro 11). https://www.ine.pt/xportal/xmain?xpid=INE&xpgid=ine_indicadores&indOcorrCod=0008657&contexto=bd&selTab=tab2

INE - Plataforma de divulgação dos Censos 2021 – Resultados Provisórios. (2021). https://www.ine.pt/scripts/db_censos_2021.html

Jacox, E. H., & Samet, H. (2007). Spatial join techniques. *ACM Transactions on Database Systems (TODS)*, *32*(1), 7-es.

Kannangara, M., Dua, R., Ahmadi, L., & Bensebaa, F. (2018). Modeling and prediction of regional municipal solid waste generation and diversion in Canada using machine learning approaches. *Waste Management*, *74*, 3–15.

Khan, D., Kumar, A., & Samadder, S. (2016). Impact of socioeconomic status on municipal solid waste generation rate. *Waste Management*, *49*, 15–25.

Korhonen, P., & Kaila, J. (2015). Waste container weighing data processing to create reliable information of household waste generation. *Waste Management*, *39*, 15–25.

LCG. (2019). *TravelBI by Turismo de Portugal—Turismo em Portugal | 2019*. https://travelbi.turismodeportugal.pt/turismo-em-portugal/turismo-em-portugal-2019/

Lebersorger, S., & Beigl, P. (2011). Municipal solid waste generation in municipalities: Quantifying impacts of household structure, commercial waste and domestic fuel. *Waste management*, *31*(9–10), 1907–1915.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern recognition*, *36*(2), 451–461.

Lisboa Aberta. (2021). LISBOA ABERTA. https://lisboaaberta.cm-lisboa.pt/index.php/pt/

Liu, X., Wang, X., Wright, G., Cheng, J. C., Li, X., & Liu, R. (2017). A state-of-the-art review on the integration of Building Information Modeling (BIM) and Geographic Information System (GIS). *ISPRS International Journal of Geo-Information*, *6*(2), 53.

Magrinho, A., Didelet, F., & Semiao, V. (2006). Municipal solid waste disposal in Portugal. *Waste management*, *26*(12), 1477–1489.

Maity, S. (2018). Importance of Municipal Solid Waste Management. *International Journal of Advanced Engineering Research and Science*, *5*(5), 237504.

Müller, M. (2007). Dynamic time warping. *Information retrieval for music and motion*, 69–84.

Município de Lisboa. (2020). *Plano Municipal de Gestão de Resíduos do Município de Lisboa | 2015-2020*. 102.

Namlis, K.-G., & Komilis, D. (2019). Influence of four socioeconomic indices and the impact of economic crisis on solid waste generation in Europe. *Waste management*, *89*, 190–200.

Neves, F. T., de Castro Neto, M., & Aparicio, M. (2020). The impacts of open data initiatives on smart cities: A framework for evaluation and monitoring. *Cities*, *106*, 102860.

Niska, H., & Serkkola, A. (2018). Data analytics approach to create waste generation profiles for waste management and collection. *Waste Management*, *77*, 477–485.

OpenStreetMap. (2021). OpenStreetMap. https://www.openstreetmap.org/

Oribe-Garcia, I., Kamara-Esteban, O., Martin, C., Macarulla-Arenaza, A. M., & Alonso-Vicario, A. (2015). Identification of influencing municipal characteristics regarding household waste generation and their forecasting ability in Biscay. *Waste management*, *39*, 26–34.

Patro, S., & Sahu, K. K. (2015). Normalization: A preprocessing stage. *arXiv preprint arXiv:1503.06462*.

Resíduos. (2022). MUNICÍPIO de LISBOA. https://www.lisboa.pt/cidade/ambiente/residuos-e-reciclagem/residuos

Ribeiro, M. B. (2017). *O impacto do turismo no centro histórico de Lisboa*.

Santos, P. H. M. do R. (2011). *Avaliação de circuitos de recolha de resíduos urbanos: Indicadores operacionais*.

Sarmento, P., Motta, M., Scott, I. J., Pinheiro, F. L., & de Castro Neto, M. (2022). Impact of COVID-19 lockdown measures on waste production behavior in Lisbon. *Waste Management*, *138*, 189–198. https://doi.org/10.1016/j.wasman.2021.12.002

SAS Center: SAS Enterprise Miner Data Sources. (2022). https://documentation.sas.com/doc/en/emref/15.2/p0ldp4l9cnob3gn1dytmimkunaa6.htm

SAS Enterprise Miner. (2022). https://support.sas.com/en/software/enterprise-miner-support.html

Schubert, S., & Lee, T. (2011). *160-2011: Time Series Data Mining with SAS® Enterprise Miner™*. 12.

Soukiazis, E., & Proença, S. (2020). The determinants of waste generation and recycling performance across the Portuguese municipalities–A simultaneous equation approach. *Waste Management*, *114*, 321–330.

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, *336*, 012017. https://doi.org/10.1088/1757-899X/336/1/012017

Talalaj, I. A., & Walery, M. (2015). The effect of gender and age structure on municipal waste generation in Poland. *Waste management*, *40*, 3–8.

Thuraisingham, B. (2000). A primer for understanding and applying data mining. *IT Professional*, *2*(1), 28–31. https://doi.org/10.1109/6294.819936

Tiseo, I. (2020, novembro 26). *Topic: Waste generation worldwide*. Statista. https://www.statista.com/topics/4983/waste-generation-worldwide/

Trends in Solid Waste Management. (2021). https://datatopics.worldbank.org/what-a-waste/trends_in_solid_waste_management.html

Turismo de Portugal—Open Data. (2020). https://dadosabertos.turismodeportugal.pt/

UNFCCC. (2019). https://unfccc.int/

Van den Bergh, J. C. (2009). The GDP paradox. *Journal of Economic Psychology*, *30*(2), 117–135.

Vieira, V. H. A. M., & Matheus, D. R. (2018). The impact of socioeconomic factors on municipal solid waste generation in São Paulo, Brazil. *Waste Management and Research*, *36*(1), 79–85. Scopus. https://doi.org/10.1177/0734242X17744039

Waste statistics. (2018). Eurostat. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Waste_statistics

8. ANNEXES

8.1. ANNEX I

Table 24 - Example circuit I0102 from mixed waste agregated by week for 52 weeks

Sum of Weight_ton		Years						
Circuit	Week	2017	2018	2019	2020	Grand Total	Average	Average 2017-2019
I0102	1	28,14	26,18	14,04	23,84	92,2	23,05	22,79
I0102	2	28,58	27,88	25,14	24,76	106,36	26,59	27,20
I0102	3	24,78	27,84	25,78	26	104,4	26,1	26,13
I0102	4	26,62	27,76	24,7	34,04	113,12	28,28	26,36
I0102	5	27,04	28,26	22,28	20,8	98,38	24,595	25,86
I0102	6	28,6	31,98	27,76	38,86	127,2	31,8	29,45
I0102	7	27,46	25,88	18,86	28,58	100,78	25,195	24,07
I0102	8	28,14	24,72	30,58	30,22	113,66	28,415	27,81
I0102	9	25,84	25,9	25,1	26,36	103,2	25,8	25,61
I0102	10	27,82	28,38	21,6	26,3	104,1	26,025	25,93
I0102	11	28,68	28,46	25,34	21,76	104,24	26,06	27,49
I0102	12	27,56	29,22	21,26	42,62	120,66	30,165	26,01
I0102	13	28,1	26,54	24,72	19,56	98,92	24,73	26,45
I0102	14	26,5	26,64	24,76	27,82	105,72	26,43	25,97
I0102	15	27,52	21,56	23,8	23,16	96,04	24,01	24,29
I0102	16	26,98	25,9	25,76	34,44	113,08	28,27	26,21
I0102	17	29,5	26,3	21,96	29,44	107,2	26,8	25,92
I0102	18	28,04	22,94	24,68	36,72	112,38	28,095	25,22
I0102	19	27,84	23,62	22,96	40,3	114,72	28,68	24,81
I0102	20	29,78	30,06	27,36	40,2	127,4	31,85	29,07
I0102	21	28,04	23,88	25,76	36,46	114,14	28,535	25,89
I0102	22	29,46	32,44	25,78	37,34	125,02	31,255	29,23
I0102	23	29,2	27,78	23,76	31,78	112,52	28,13	26,91
I0102	24	21,62	25,72	22,56	33,6	103,5	25,875	23,30
I0102	25	33,66	23,06	25,32	41,36	123,4	30,85	27,35
I0102	26	27,82	21,58	26,74	33,24	109,38	27,345	25,38
I0102	27	28,46	26,18	25,44	22,74	102,82	25,705	26,69
I0102	28	27,4	25,38	25,18	22,54	100,5	25,125	25,99
I0102	29	23,5	25,04	27,32	22,16	98,02	24,505	25,29
I0102	30	25,18	24,36	24,76	21,88	96,18	24,045	24,77
I0102	31	24,9	26,22	23,44	20,1	94,66	23,665	24,85

I0102	32	23,6	22,66	21,16	17,98	85,4	21,35	22,47
I0102	33	21,22	21,56	18,12	18,18	79,08	19,77	20,30
I0102	34	22,1	17,64	21,52	18,28	79,54	19,885	20,42
I0102	35	23,42	23	18,4	17,54	82,36	20,59	21,61
I0102	36	26,4	25,7	20,34	12,42	84,86	21,215	24,15
I0102	37	26,08	27,38	23,98	26	103,44	25,86	25,81
I0102	38	27,46	25,56	26,18	20,72	99,92	24,98	26,40
I0102	39	27,24	26,46	24,96	23,94	102,6	25,65	26,22
I0102	40	23,32	22,52	24,42	22,9	93,16	23,29	23,42
I0102	41	29,98	27,12	22,84	21,38	101,32	25,33	26,65
I0102	42	29,44	25,78	24,38	21,98	101,58	25,395	26,53
I0102	43	20,46	18,8	21,76	22,62	83,64	20,91	20,34
I0102	44	33,76	25,88	25,32	22,12	107,08	26,77	28,32
I0102	45	26,38	26,3	29,62	6,66	88,96	22,24	27,43
I0102	46	26,26	26,04	23,04		75,34	18,835	25,11
I0102	47	25,24	24,18	23,34		72,76	18,19	24,25
I0102	48	25,14	29,38	23		77,52	19,38	25,84
I0102	49	13,92	33	26,3		73,22	18,305	24,41
I0102	50	41,78	24,6	20,5		86,88	21,72	28,96
I0102	51	27,28	26,34	24,34		77,96	19,49	25,99
I0102	52	32,2	23,56	32,48		88,24	22,06	29,41
I0102	53	8,1	7,06	6,46		21,62	5,405	7,21

Table 25 – Socioeconomic and Context Variables

Socioeconomic, Demographic and Context Variables	Description
HM_0_14_years	Resident population, male (H) and female (M), aged from 0 to 14
HM_15_24_years	Resident population, male (H) and female (M), aged from 15 to 24
HM_25_64_years	Resident population, male (H) and female (M), aged from 15 to 64
HM_65_and_more_years	Resident population, male (H) and female (M), aged 65 or more
H_0_14_years	Resident population, male (H), aged from 0 to 14
H_15_24_years	Resident population, male (H), aged from 15 to 24

Socioeconomic, Demographic and Context Variables	Description
H_25_64_years	Resident population, male (H), aged from 15 to 64
H_65_and_more_years	Resident population, male (H), aged 65 or more
M_0_14_years	Resident population, female (M), aged from 0 to 14
M_15_24_years	Resident population, female (M), aged from 15 to 24
M_25_64_years	Resident population, female (M), aged from 15 to 64
M_65_and_more_years	Resident population, female (M), aged 65 or more
No Schooling level	Resident population with no schooling education
Basic Education	Resident population with basic education level
1st Cycle	Resident population with 1st cycle education level
2nd Cycle	Resident population with 2nd cycle education level
3rd Cycle	Resident population with 3rd cycle education level
Secondary and post-secondary education	Resident population with secondary and post-secundary education level
University Education	Resident population with university education level
No charges	Monthly family charges: No charges
With charges	Monthly family charges: With charges
Charges Less than 100 euros	Monthly family charges: less than 100 euros charges
Charges 100 to 199,99 euros	Monthly family charges: from 100 to 199,99 euros charges
Charges 200 to 299,99 euros	Monthly family charges: from 200 to 299,99 euros charges
Charges 300 to 399,99 euros	Monthly family charges: from 300 to 399,99 euros charges
Charges 400 to 649,99 euros	Monthly family charges: from 400 to 649,99 euros charges
Charges 650 to 999,99 euros	Monthly family charges: from 650 to 999,99 euros charges
Charges 1000 or more euros	Monthly family charges: from 1000 euros or more charges
Size Less than 30 m	Household space living: Less than 30 m
Size 30 m to 39 m	Household space living: from 30 m to 39 m
Size 40 m to 49 m	Household space living: from 40 m to 49 m
Size 50 m to 59 m	Household space living: from 50 m to 59 m
Size 60 m to 79 m	Household space living: from 60 m to 79 m
Size 80 m to 99 m	Household space living: from 80 m to 99 m
Size 100 m to 119 m	Household space living: from 100 m to 119 m
Size 120 m to 149 m	Household space living: from 120 m to 149 m
Size 150 m to 199 m	Household space living: from 150 m to 199 m
Size 200 m or more	Household space living: from 200 m or more
Rent Less than 20 euros	House income: Less than 20 euros

Socioeconomic, Demographic and Context Variables	Description
Rent 20 to 49,99 euros	House income from: 20 to 49,99 euros
Rent 50 to 99,99 euros	House income from: 50 to 99,99 euros
Rent 100 to 199,99 euros	House income from: 100 to 199,99 euros
Rent 200 to 399,99 euros	House income from: 200 to 399,99 euros
Rent 400 to 649,99 euros	House income from: 400 to 649,99 euros
Rent 650 to 999,99 euros	House income from: 650 to 999,99 euros
Rent 1000 or more euros	House income from: 1000 or more euros
Family size with 1 person	Household size with 1 person
Family size with 2 people	Household size with 2 people
Family size with 3 people	Household size with 3 people
Family size with 4 people	Household size with 4 people
Family size with 5 or more people	Household size with 5 or more people
Before 1919	Building construction season: Before 1919
1919 - 1945	Building construction season: 1919 - 1945
1946 - 1960	Building construction season: 1946 - 1960
1961 - 1980	Building construction season: 1961 - 1980
1981 - 2000	Building construction season: 1981 - 2000
2001 - 2010	Building construction season: 2001 - 2010
2011 - 2021	Building construction season: 2011 - 2021
Hospital	Number of Hospitals
Health Center	Nmber of Health Centre
Pre Scholar	Number of Pre Scholar
1st cycle	Number of Schools with 1st cycle
2nd 3rd cycle	Number of Schools with 2nd 3rd cycle
Secondary School	Number of Secondary School
University	Number of University
Train Station	Number of Train Station
Metro Station	Number of Metro Station
Bus Stop	Number of Bus Stop
Rest Coffe Shop Bar Market	Number of Restaurants, Coffe Shop, Bars and Markets
Culture Spot	Number of Culture Spots
Beds Tourist Establishments	Number of Beds in Tourist Establishments
Beds Local accommodation	Number of Beds in Local Accommodation (AL)
Cultural and Sporting events	Number of Cultural and Sporting events
Coverage percentage by building built continuous predominantly vertical	Areas of continuous built-up buildings in which buildings with a height greater than or equal to 3 floors occupy a surface greater than or equal to 50% of the plot.

Socioeconomic, Demographic and Context Variables	Description
Coverage percentage by building built continuous predominantly horizontal	Areas of continuous built-up buildings in which buildings with a height of less than 3 floors occupy an area greater than or equal to 50% of the plot
Percentage of coverage by building built discontinuous	Building areas for the most part occupied by residential type constructions. The waterproofed surface occupies a superior area or equal to 50% and less than 80% of the total surface.
Coverage percentage by building built discontinuous sparse	Building areas for the most part occupied by residential type constructions. The waterproofed surface occupies a superior area or equal to 30% and less than 50% of the total surface.
Industry	Areas occupied by industrial production.
Trade	Large commercial surfaces, warehouses and other miscellaneous equipment.
Sports facilities	Areas occupied by sports facilities. Includes football stadiums and infrastructure adjacent areas, hockey stadiums, swimming pools and tennis courts, cycling tracks, racetracks and athletics tracks, whether or not included in built-up fabric.
Leisure Equipment	Recreation spaces and structures, including zoos and botanical gardens not included.
Cultural Equipment	Open-air archaeological complexes, religious temples and associated spaces, and cultural facilities such as theaters, planetariums and concert halls.
Price per sqm	Price per Square Meter in Lisbon

8.2. Annex II – Data Preparation

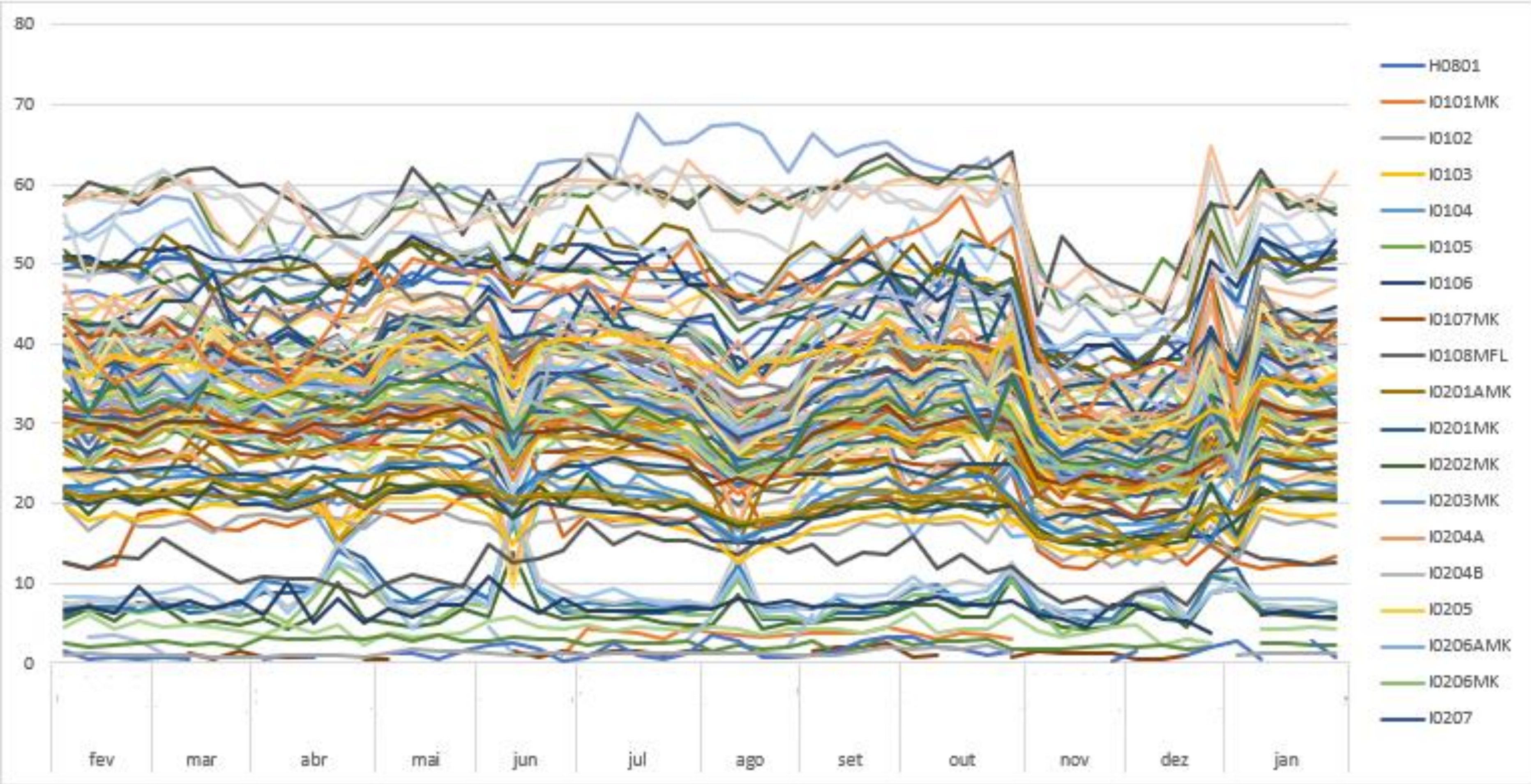


Figure 35 – Undifferenciated Waste Time-Series result before data cleaning

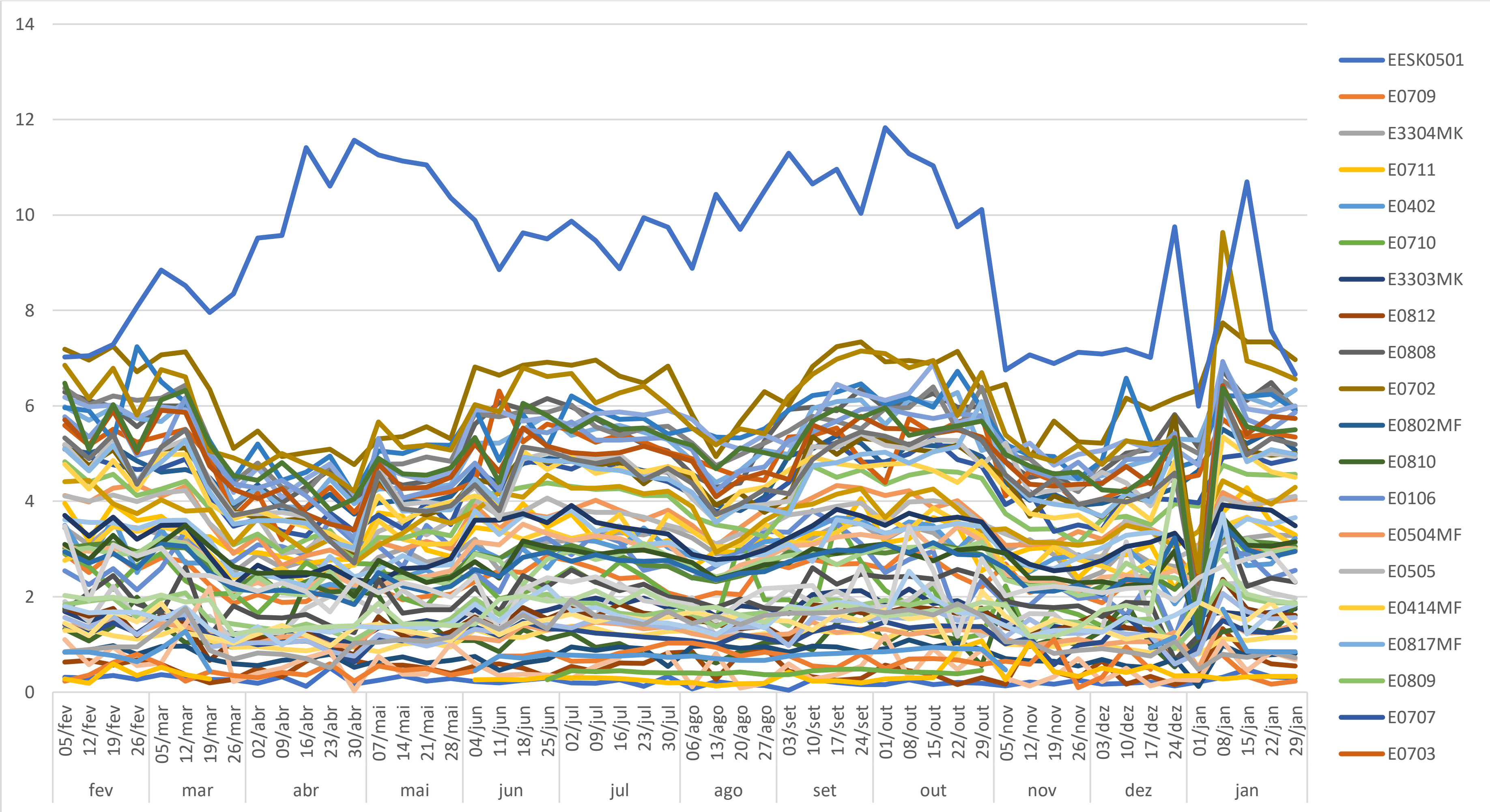


Figure 36 - Plastic Waste Time-Series result before data cleaning.

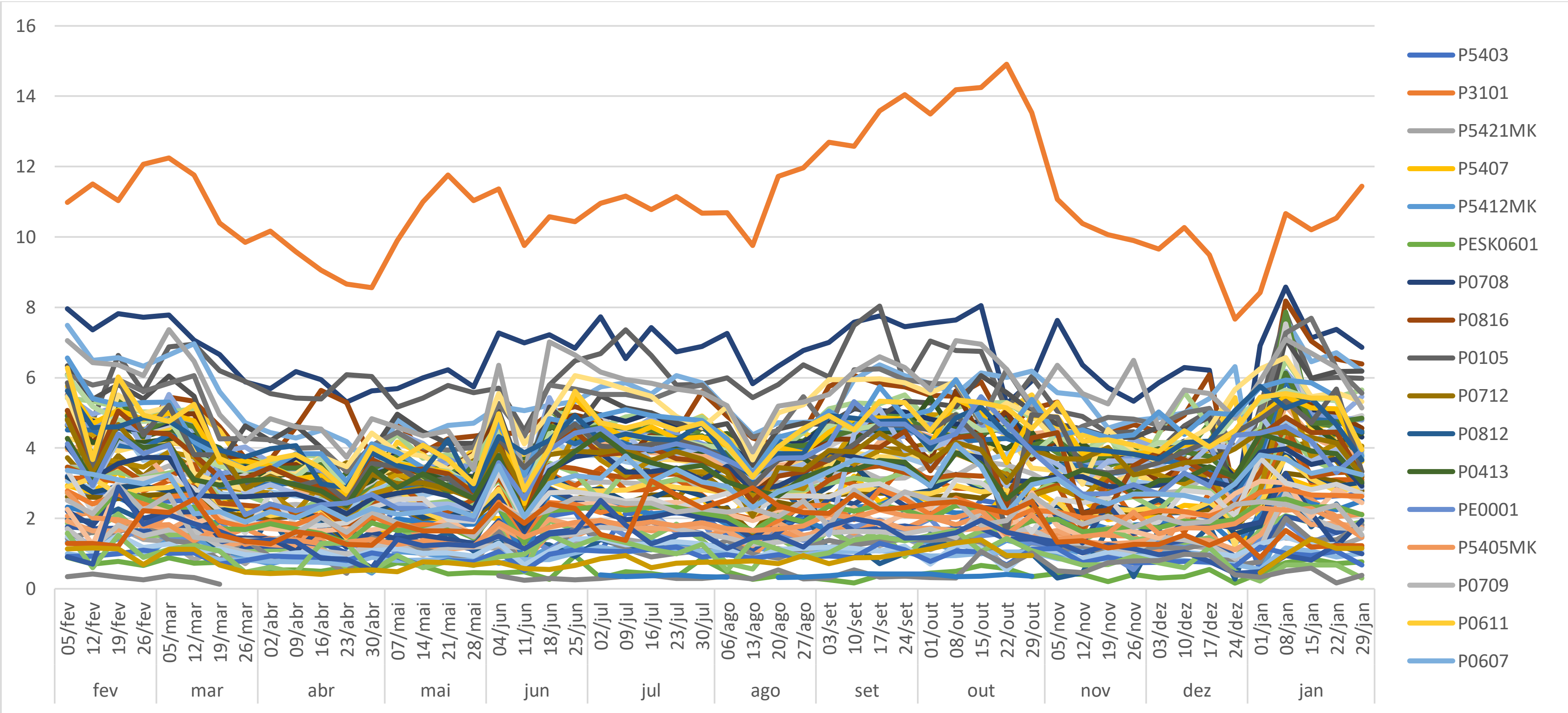


Figure 37 - Paper Waste Time-Series result before data cleaning

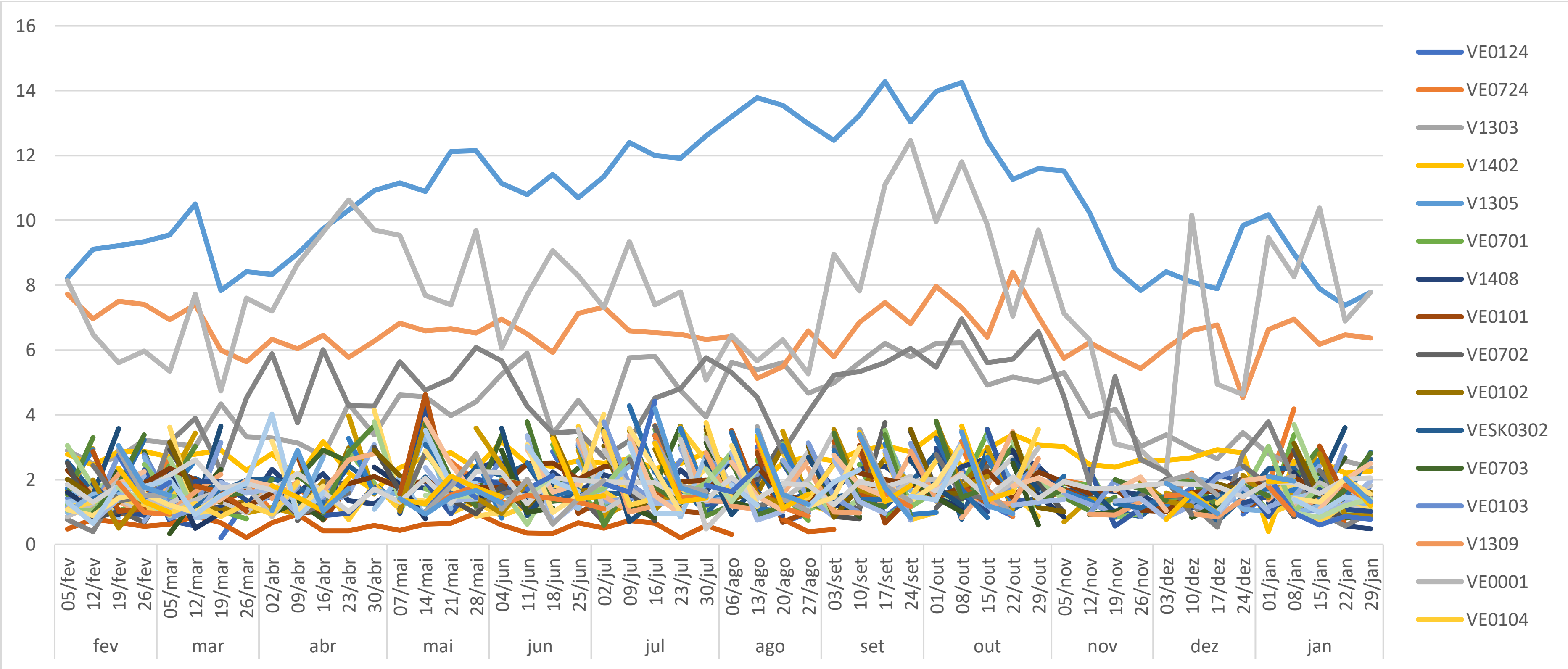


Figure 38 - Glass Waste Time-Series result before data cleaning

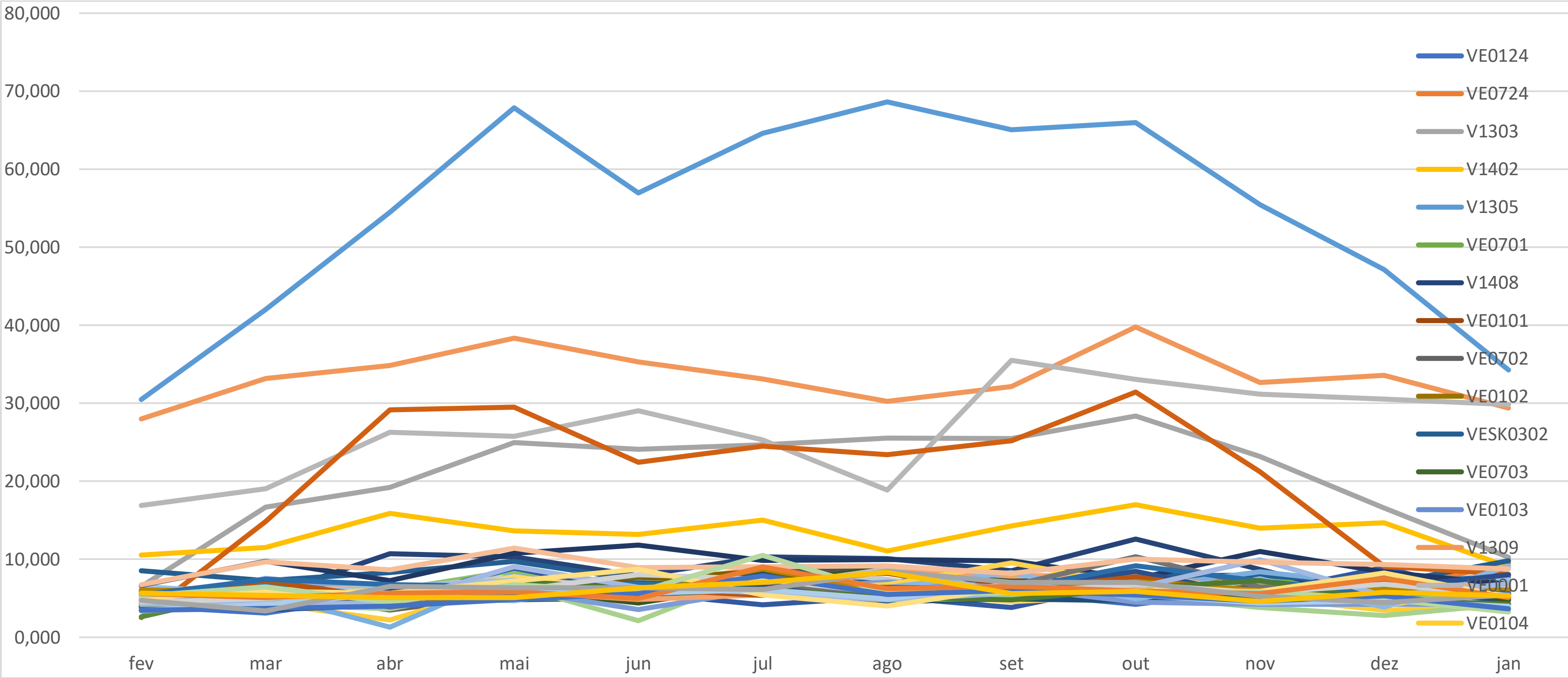


Figure 39 - Glass Waste Time-Series result after data cleaning

Circuit	All	EESA01MF
Waste collected (avg) ton	155,92	479,855

Table 26 – Plastic circuit considered as an outlier

Circuit	All	P3101
Waste collected (avg) ton	155,92	573,6

Table 27 – Paper circuit considered as an outlier

Circuit	All	V1305
Waste collected (avg) ton	100,07	653,04

Table 28 – Glass circuit considered as an outlier

8.3. Annex III – Data Modelling

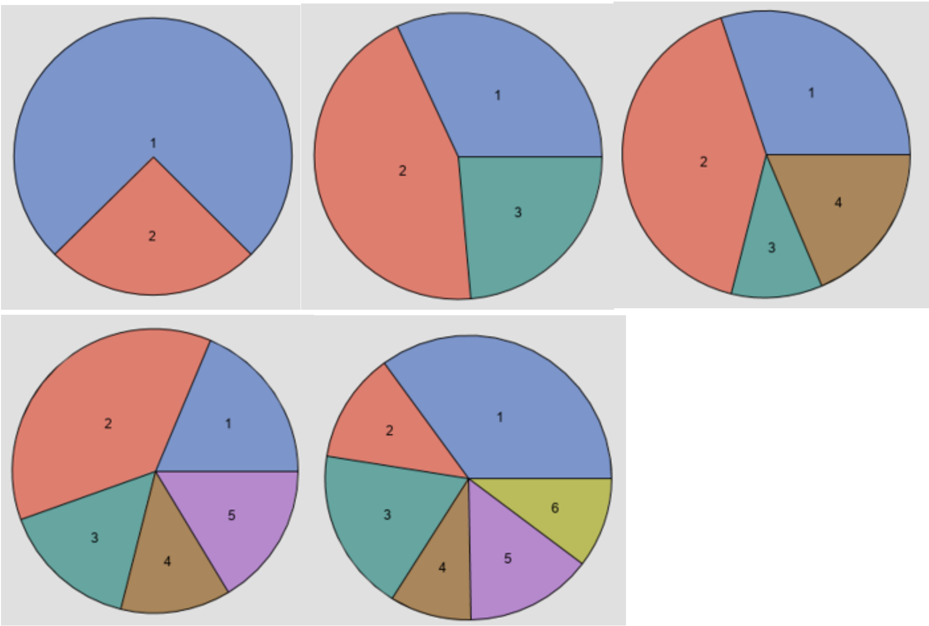


Figure 40 –Segment Size Mixed Waste, K=2, K=3, K=4, K=5, K=6, respectively

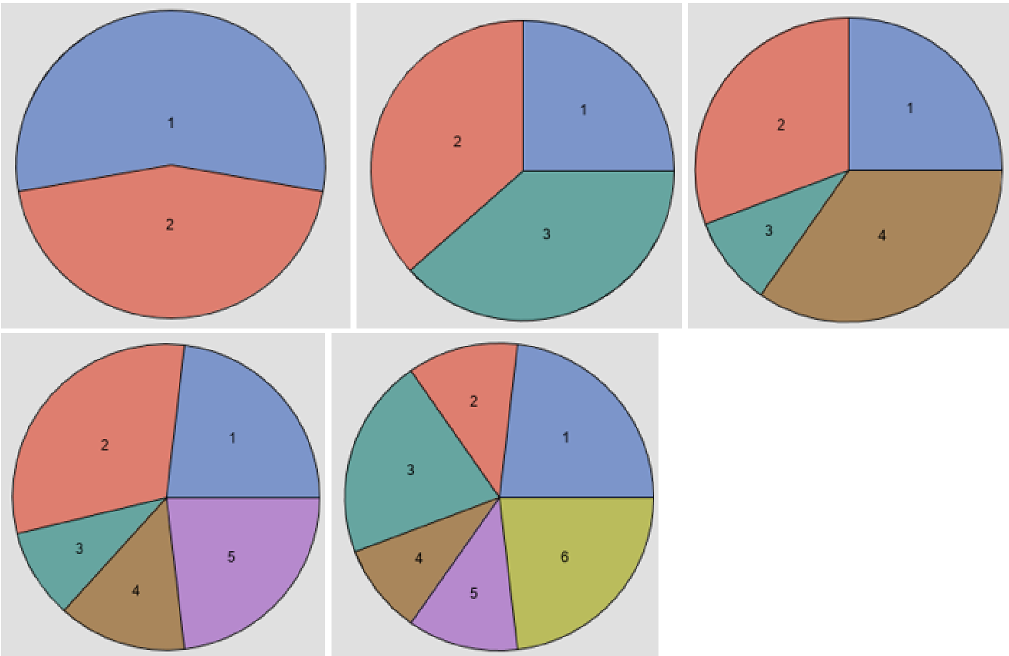


Figure 41 –Segment Size Plastic Waste, K=2, K=3, K=4, K=5, K=6, respectively

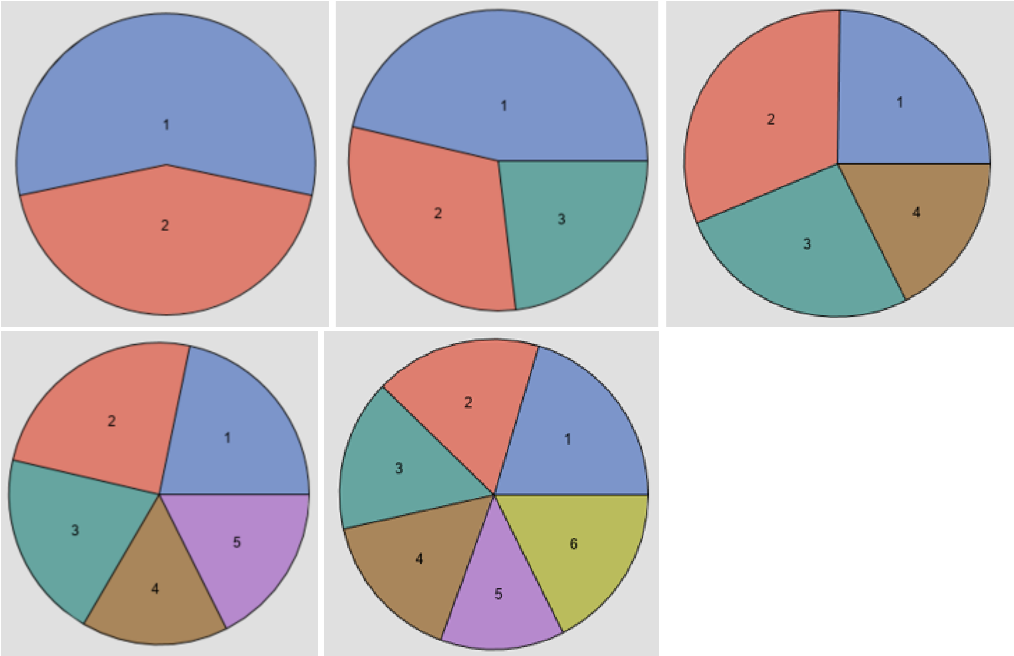


Figure 42 –Segment Size Paper Waste, K=2, K=3, K=4, K=5, K=6, respectively

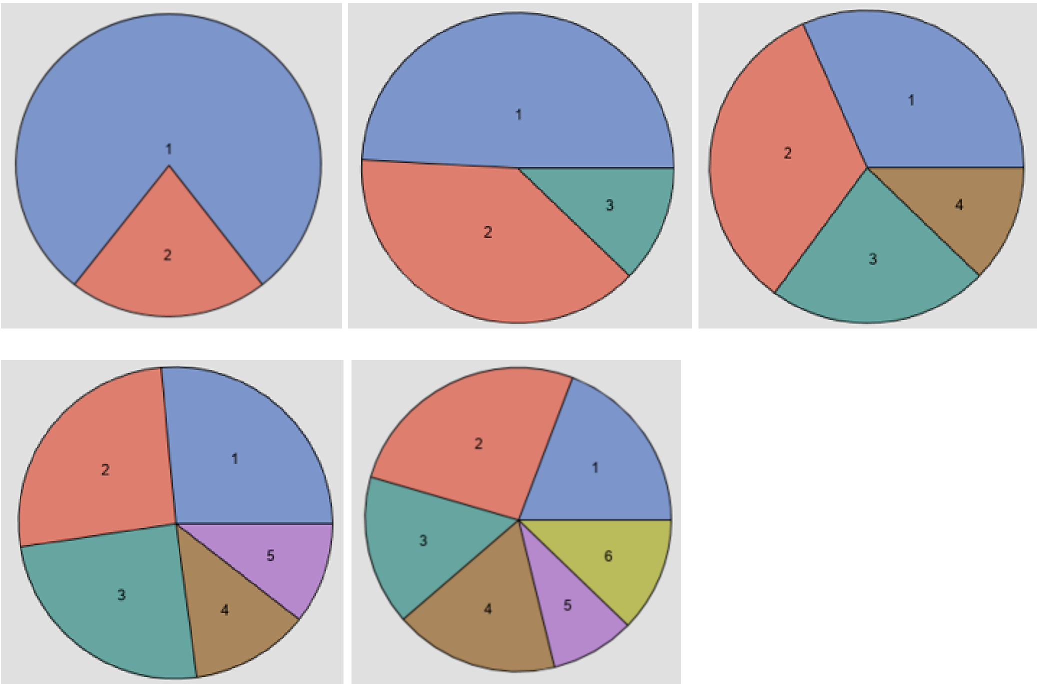


Figure 43 –Segment Size Glass Waste, K=2, K=3, K=4, K=5, K=6, respectively

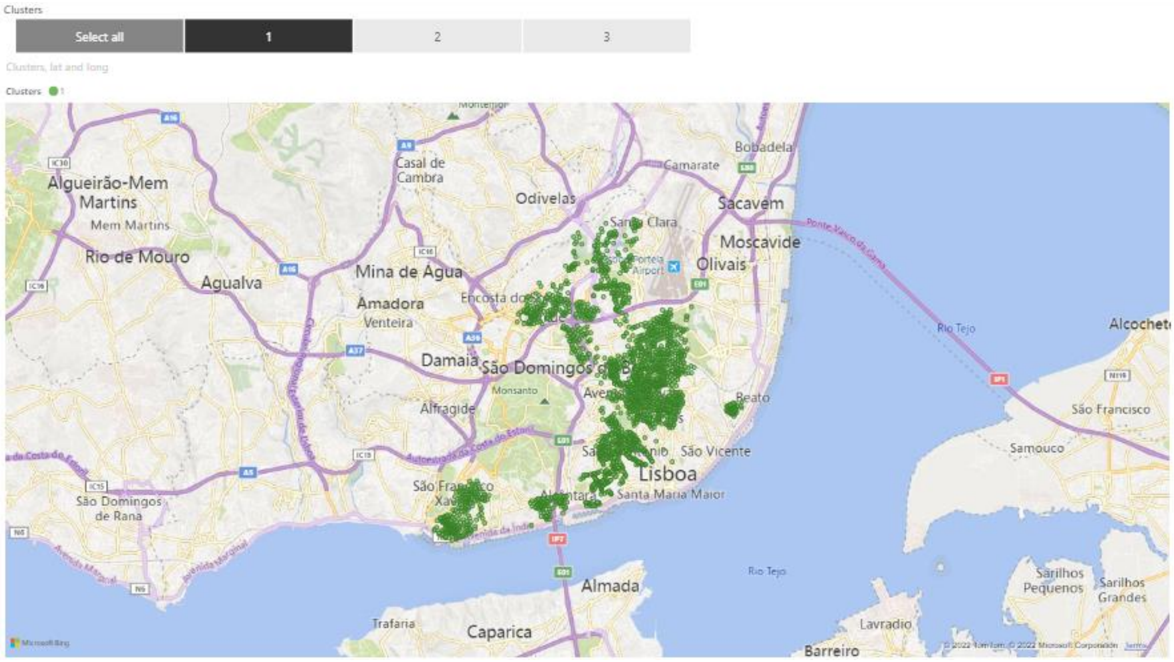
8.4. Annex IV – Evaluation

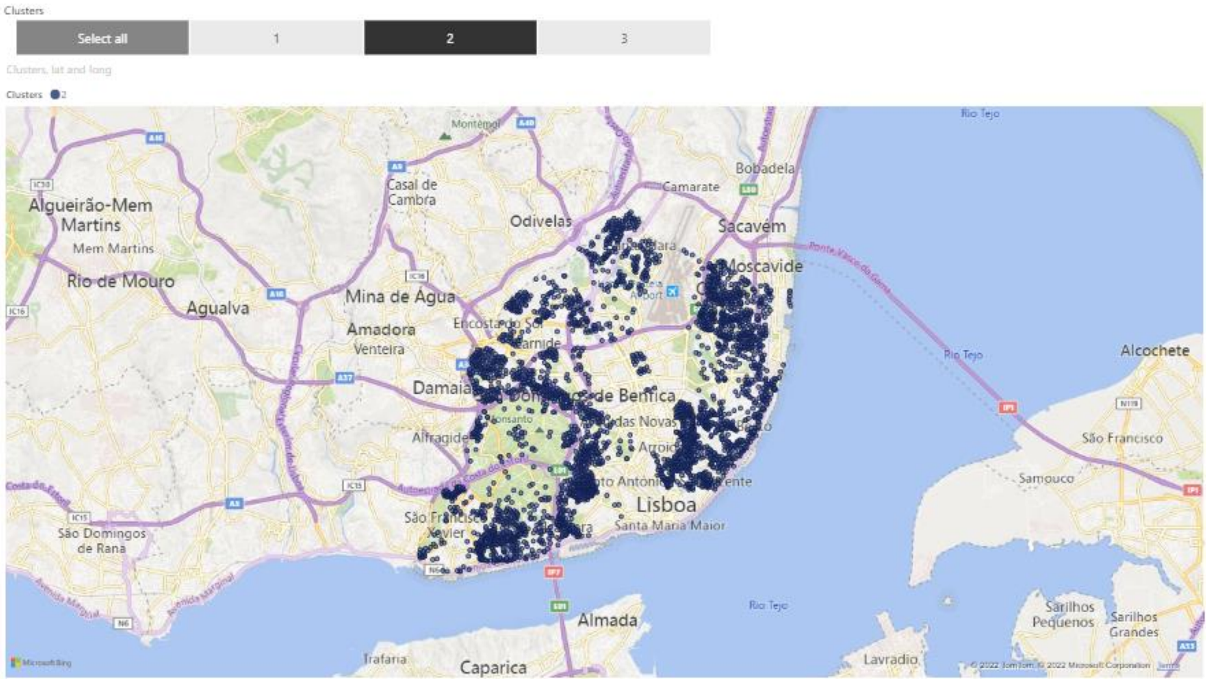
Table 29 – Amount per each Variable Mixed Waste, Z-Score, full table.

Variables	Clusters		
	1	2	3
HM_0_14_years	0,842996025	0,261889342	-1,104885367
HM_15_24_years	0,930325421	0,127180904	-1,057506325
HM_25_64_years	-0,378782043	-0,755274593	1,134056636
HM_65_and_more_years	-0,243120848	1,099143766	-0,856022918
H_0_14_years	0,755427369	0,378591226	-1,134018595
H_15_24_years	0,781659374	0,34521139	-1,126870764
H_25_64_years	-0,445790682	-0,699575797	1,145366479
H_65_and_more_years	0,006951951	0,996505901	-1,003457852
M_0_14_years	0,957155225	0,080788463	-1,037943688
M_15_24_years	1,112592959	-0,288709708	-0,823883251
M_25_64_years	-0,168735993	-0,90489745	1,073633443
M_65_and_more_years	-0,777292586	1,128147596	-0,35085501
No Schooling level	-0,112048174	-0,93925673	1,051304904
Basic Education	-0,915308832	1,067291444	-0,151982612
1st Cycle	-1,021417342	0,977109743	0,044307599
2nd Cycle	-0,794035747	1,123055355	-0,329019608
3rd Cycle	-0,735317258	1,138684827	-0,403367569
Secondary and post-secondary education	-0,78323151	-0,343170814	1,126402324
University Education	1,033352384	-0,962924035	-0,070428349
No charges	-1,153487325	0,53091525	0,622572075
With charges	1,153487325	-0,53091525	-0,622572075
Charges Less than 100 euros	-0,596316907	1,154490521	-0,558173614
Charges 100 to 199,99 euros	-0,891687983	1,081193821	-0,189505838
Charges 200 to 299,99 euros	-0,991576953	1,008215408	-0,016638455
Charges 300 to 399,99 euros	-0,764752626	1,131619308	-0,366866682
Charges 400to 649,99 euros	1,105920186	-0,265374615	-0,840545571
Charges 650 to 999,99 euros	0,95614942	-1,03872943	0,082580009
Charges 1000 or more euros	0,470798944	-1,148504759	0,677705815
Size Less than 30 m	-0,709745768	-0,433920947	1,143666715
Size 30 m to 39 m	-0,758317588	-0,374973706	1,133291294
Size 40 m to 49 m	-0,853332529	-0,247030771	1,1003633
Size 50 m to 59 m	-1,118811894	0,312029796	0,806782098
Size 80 m to 99 m	0,007163247	0,996399134	-1,003562381
Size 100 m to 119 m	1,083538931	-0,196143166	-0,887395765
Size 60 m to 79 m	-0,860739133	1,096962477	-0,236223345
Size 120 m to 149 m	1,121775212	-0,323789986	-0,797985225
Size 150 m to 199 m	1,139488083	-0,7315315	-0,407956583

Size 200 m or more	0,872251011	-1,091399212	0,219148201
Rent Less than 20 euros	-0,385274374	1,135331593	-0,750057219
Rent 20 to 49,99 euros	-0,636563976	1,152601908	-0,516037931
Rent 50 to 99,99 euros	-0,484475267	1,149961593	-0,665486326
Rent 100 to 199,99 euros	-1,020147608	0,978554479	0,041593129
Rent 200 to 399,99 euros	-1,15427863	0,604169539	0,550109091
Rent 400 to 649,99 euros	-0,547460858	-0,606732493	1,154193351
Rent 650 to 999,99 euros	1,05146117	-0,939036835	-0,112424335
Rent 1000 or more euros	0,968716202	-1,02859306	0,059876858
Family size with 1 person	-0,377621311	-0,756203424	1,133824734
Family size with 2 people	-0,334259088	1,124314556	-0,790055468
Family size with 3 people	0,223310296	0,869466345	-1,092776641
Family size with 4 people	0,75849026	0,374757121	-1,133247381
Family size with 5 or more people	0,533307977	-1,153607737	0,62029976
Before 1919	-0,539222275	-0,61465664	1,153878916
1919 - 1945	-0,357070381	-0,772451432	1,129521814
1946 - 1960	0,795402261	0,327213645	-1,122615906
1961 - 1980	0,127714345	0,930007398	-1,057721743
1981 - 2000	0,4651705	0,682680998	-1,147851498
2001 - 2010	0,755648384	0,378315107	-1,133963491
2011 - 2021	0,868127501	-1,093431615	0,225304114
Hospital	0,689333656	-1,146922606	0,45758895
Health Center	0,056262463	-1,026943477	0,970681014
Pre Scholar	0,449321542	0,696524668	-1,145846211
1st cycle	0,420113861	0,72140872	-1,141522581
2nd 3rd cycle	0,177752987	0,899203922	-1,076956908
Secondary School	0,701640248	0,443393778	-1,145034026
University	-0,740528745	-0,397011903	1,137540648
Train Station	-0,993673758	1,006208401	-0,012534643
Metro Station	0,89699799	0,181219954	-1,078217944
Bus Stop	-0,141563412	1,063238179	-0,921674766
Rest Coffe Shop Bar Market	0,80389608	-1,119801311	0,315905231
Culture Spot	1,082799416	-0,88875985	-0,194039566
Beds Tourist Establishments	-0,405206679	-0,733802374	1,139009053
Beds Local accommodation	0,002222423	-1,001109359	0,998886936
Cultural and Sporting events	0,124085243	-1,056251925	0,932166682
Coverage percentage by building built continuous predominantly vertical	-0,098280595	-0,947230965	1,045511561
Coverage percentage by building built continuous predominantly horizontal	-0,269452994	1,107118554	-0,83766556
Percentage of coverage by building built discontinuous	0,617346448	0,536408146	-1,153754594
Coverage percentage by building built discontinuous sparse	0,610580026	0,543470274	-1,154050299
Industry	-0,647236391	1,151756946	-0,504520556

Trade	0,441512585	0,703256682	-1,144769268
Sports facilities	-0,57236785	1,154686247	-0,582318397
Leisure Equipment	0,739588426	0,398161983	-1,13775041
Cultural Equipment	0,038064033	-1,018488542	0,980424509
Price per sqm	0,032710545	-1,01595395	0,983243404





3

[3] Variables overlaid in gray are those selected by Literature Review.

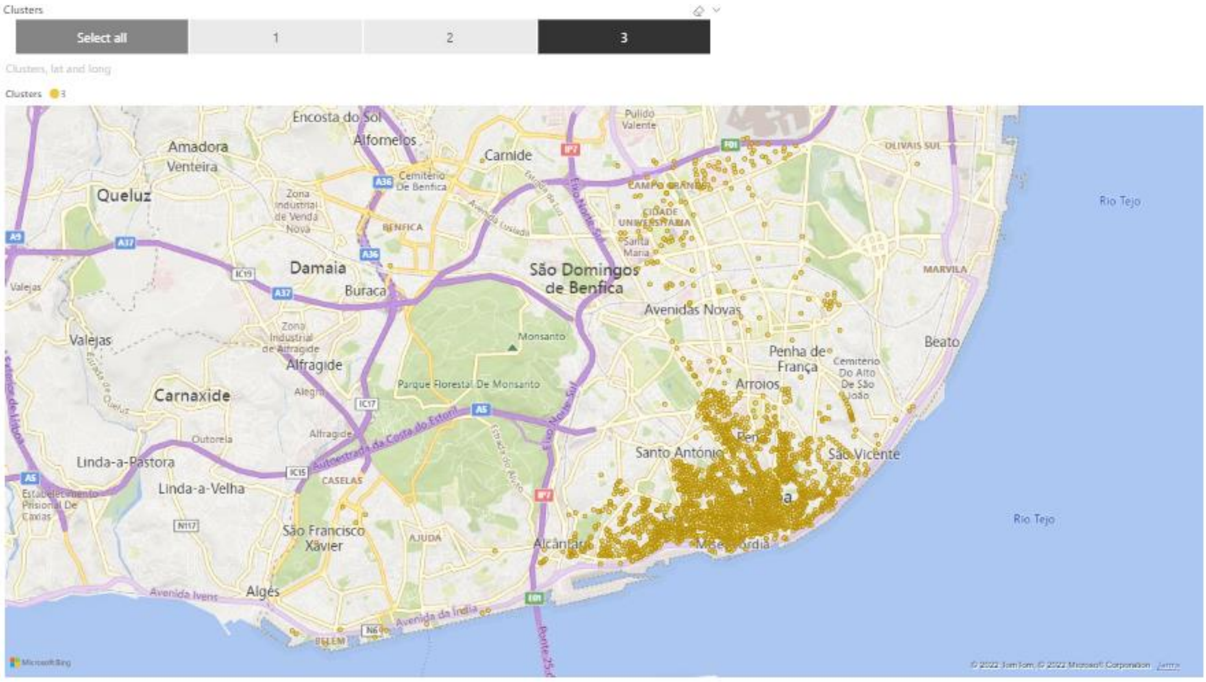


Figure 44 – PowerBI Mixed Waste, Cluster 1, 2 and 3

Table 30 - Amount per each Variable Plastic Waste, Z-Score, full table

	Cluster					
Variables	1	2	3	4	5	EESA01MF
HM_0_14_years	0,072734802	0,584011089	1,332229775	0,28527621	-1,46217302	-0,812078856
HM_15_24_years	0,412871331	0,20935905	1,635127106	-1,24545823	-0,349538084	-0,662361174
HM_25_64_years	1,265938173	-0,700847277	-1,474868516	-0,189811594	0,337191881	0,762397332
HM_65_and_more_years	-1,841357895	0,58932197	0,851447993	0,465834754	0,366128668	-0,43137549
H_0_14_years	-0,262259979	0,855574683	1,067845561	0,604347675	-1,385160681	-0,880347259
H_15_24_years	0,006223343	0,390745194	1,773006666	-1,02062759	-0,546869307	-0,602478306
H_25_64_years	1,191776065	-0,920619335	-1,397249147	-0,10692534	0,552379127	0,68063863
H_65_and_more_years	-1,737143762	0,864193982	1,013411801	0,216376039	0,04295344	-0,3997915
M_0_14_years	0,425485332	0,249024968	1,489733443	-0,056840433	-1,449831606	-0,657571705
M_15_24_years	1,067065274	-0,067539426	1,166303512	-1,402483128	-0,036536487	-0,726809745
M_25_64_years	1,319174282	-0,254693851	-1,537212439	-0,275096995	-0,114237988	0,862066991
M_65_and_more_years	-1,823918627	0,155561209	0,601368051	0,622373801	0,855271989	-0,410656423
No Schooling level	1,640038121	-1,318437299	-0,524234131	-0,366104085	0,310951698	0,257785696
Basic Education	-1,286812886	-0,387494044	0,177334684	-0,239637952	1,761943598	-0,0253334
1st Cycle	-1,192830304	-0,502622018	0,12923455	-0,244959685	1,802148338	0,009029119
2nd Cycle	-1,193471496	-0,376987661	0,24709884	-0,35466357	1,797860191	-0,119836305
3rd Cycle	-1,534733874	-0,127160447	0,228055726	-0,134801274	1,599159133	-0,030519265
Secondary and post-secondary education	0,608605402	-1,098855136	-1,039714661	-0,509796763	1,041149681	0,998611477
University Education	1,105739392	0,557582648	-0,064035778	0,299359215	-1,835409118	-0,063236359
No charges	-1,053770286	-0,634831299	1,153825543	0,158389621	1,20291568	-0,826529259
With charges	1,053770286	0,634831299	-1,153825543	-0,158389621	-1,20291568	0,826529259
Charges Less than 100 euros	-0,73934272	-0,278094173	-0,648981562	-0,09616318	1,974969138	-0,212387503
Charges 100 to 199,99 euros	-1,082881496	-0,269542125	-0,385237038	-0,229459455	1,883247425	0,083872689
Charges 200 to 299,99 euros	-1,033938624	-0,370814887	-0,179014492	-0,497441664	1,861657818	0,219551849
Charges 300 to 399,99 euros	-1,007752442	-0,036512029	-0,566884869	-0,62746711	1,736692487	0,501923963

Charges 400to 649,99 euros	0,951381214	1,004428157	-0,442917727	-0,160106665	-1,663604221	0,310819243
Charges 650 to 999,99 euros	1,157466849	0,254317075	0,23186897	0,425389704	-1,815675914	-0,253366682
Charges 1000 or more euros	0,672532668	-0,343142651	1,02804905	0,74977008	-1,597087652	-0,510121495
Size Less than 30 m	-0,367312651	-0,409640193	-1,624522776	1,051628688	0,531334102	0,81851283
Size 30 m to 39 m	-0,521549258	-0,471135834	-1,516819769	1,012567108	0,575883072	0,921054681
Size 40 m to 49 m	-0,610729193	-0,658638907	-1,305550854	1,240188464	0,494273587	0,840456903
Size 50 m to 59 m	-0,868962843	-0,920203641	-0,71216443	0,483381982	1,566304519	0,451644413
Size 80 m to 99 m	-0,533187901	0,891592042	-0,666162139	-1,333901636	1,239270701	0,402388933
Size 100 m to 119 m	1,396272736	0,564719581	0,347168661	-0,536333845	-1,499773678	-0,272053455
Size 60 m to 79 m	-1,225942317	-0,294158958	-0,478254088	-0,030686162	1,763641105	0,26540042
Size 120 m to 149 m	1,062186587	0,833766591	0,57192501	-0,371981745	-1,55526141	-0,540635032
Size 150 m to 199 m	0,957909044	0,295471876	1,073749171	-0,128367253	-1,551273106	-0,647489731
Size 200 m or more	0,367791136	-0,407498608	1,738832507	0,076417614	-1,103433765	-0,672108883
Rent Less than 20 euros	-1,064682696	-0,220726242	0,809921611	-0,720855387	1,579626204	-0,38328349
Rent 20 to 49,99 euros	-1,587904527	0,322636031	0,777395167	-0,466234938	1,217859658	-0,263751391
Rent 50 to 99,99 euros	-1,470456847	0,168151997	0,730214582	-0,527148405	1,384550459	-0,285311786
Rent 100 to 199,99 euros	-1,671255865	0,078358116	0,565069966	-0,320437798	1,333628564	0,014637017
Rent 200 to 399,99 euros	-0,756785201	0,610782404	-1,661557409	0,369991217	0,470538457	0,967030533
Rent 400 to 649,99 euros	0,496568072	0,203847957	-1,7069875	0,631410477	-0,627590577	1,00275157
Rent 650 to 999,99 euros	1,659832352	-0,023817584	-0,248212218	0,142361909	-1,469287964	-0,060876495
Rent 1000 or more euros	1,343433596	-0,632204942	0,825053452	0,359441569	-1,255032409	-0,640691266
Family size with 1 person	0,490241019	-0,197465298	-1,318763046	1,414218795	-0,867227387	0,478995916
Family size with 2 people	-1,271538923	1,369200563	-0,809808556	-0,15855848	-0,038189249	0,908894645
Family size with 3 people	-0,587415011	-0,021820138	0,78979951	-1,221440058	1,519701902	-0,478826205
Family size with 4 people	0,037161101	0,486223929	1,623682748	-1,191712284	-0,124598826	-0,830756667
Family size with 5 or more people	0,335657848	-1,103834038	1,427004142	-0,902084522	0,749800252	-0,506543681
Before 1919	-0,986121933	-0,128360096	-1,132076098	1,401600306	-0,029201646	0,874159467
1919 - 1945	1,331635584	-0,917581724	-1,330530171	0,117579346	0,024257309	0,774639657
1946 - 1960	1,483055262	-0,904820351	0,698233219	0,346111803	-0,981791278	-0,640788654

1961 - 1980	-1,239546881	0,502361297	0,57250327	-0,741559084	1,433643532	-0,527402135
1981 - 2000	-0,747450232	1,295986485	1,179465525	-1,072165067	-0,273148707	-0,382688003
2001 - 2010	-0,311528794	1,562849607	0,824117642	-0,72558077	-1,089510374	-0,260347311
2011 - 2021	-0,939050827	0,657601115	-0,643870002	1,689948679	-0,630226234	-0,134402732
Hospital	-0,081315913	1,577120238	0,249935796	0,300132381	-0,661871866	-1,384000636
Health Center	1,089753283	0,280092859	-1,264989526	0,4466097	0,66826618	-1,219732496
Pre Scholar	0,288059812	1,642039023	-0,75073285	-0,197282482	0,262604451	-1,244687954
1st cycle	0,102898905	1,502124162	-0,710491431	-0,504205487	0,778961229	-1,169287379
2nd 3rd cycle	0,031104186	1,324993897	-0,802392444	-0,323065499	0,995675275	-1,226315415
Secondary School	0,928688357	1,427869495	-0,853185709	-0,132457605	-0,230730739	-1,140183799
University	-0,227580613	-0,079811605	-0,425536537	1,991294746	-0,585832242	-0,67253375
Train Station	0,986247761	-0,599306318	-0,74117625	-0,62121597	1,550068005	-0,574617227
Metro Station	0,644307365	0,995011407	-1,180132848	-0,16429797	0,884940037	-1,179827992
Bus Stop	-0,094740746	1,208299404	-0,719374561	0,282330919	0,824710666	-1,501225683
Rest Coffe Shop Bar Market	1,046565839	1,058305312	-1,091919899	0,220801071	0,008653706	-1,242406031
Culture Spot	0,712060845	0,295648625	-0,564219138	-0,436331187	1,392272762	-1,399431906
Beds Tourist Establishments	1,155648353	0,230278068	-1,005406879	-0,084636104	0,981144001	-1,27702744
Beds Local accommodation	0,373412845	1,184681124	0,005539856	0,526235574	-0,339115367	-1,750754032
Cultural and Sporting events	0,271294486	0,563638495	-0,8395605	-0,051404851	1,424375173	-1,368342802
Coverage percentage by building built continuous predominantly vertical	0,924106938	0,691061787	-1,563620917	-0,096936424	-0,756126232	0,801514848
Coverage percentage by building built continuous predominantly horizontal	-1,188549479	-0,524586579	0,621488649	0,17132516	1,572155298	-0,651833049
Percentage of coverage by building built discontinuous	-0,364510995	-0,482175145	1,654085764	-1,0550216	0,738184755	-0,490562778
Coverage percentage by building built discontinuous sparse	-0,762400971	0,213955483	-0,762400971	0,308173347	1,765074083	-0,762400971
Industry	-0,709435282	-0,265029058	-0,717420482	0,848713472	1,612380685	-0,769209334
Trade	-0,465973552	1,757846793	0,612536324	-0,674080295	-0,38450986	-0,84581941
Sports facilities	-0,780443165	-0,022464231	-0,59275413	1,612644217	0,732604442	-0,949587134

Leisure Equipment	-0,40824829	2,041241452	-0,40824829	-0,40824829	-0,40824829	-0,40824829
Cultural Equipment	-0,13325463	0,586104202	-1,436128955	-0,230472998	-0,338254348	1,552006729
Price per sqm	0,291088611	-0,761765071	0,345612831	1,658902941	-1,145893721	-0,387945591

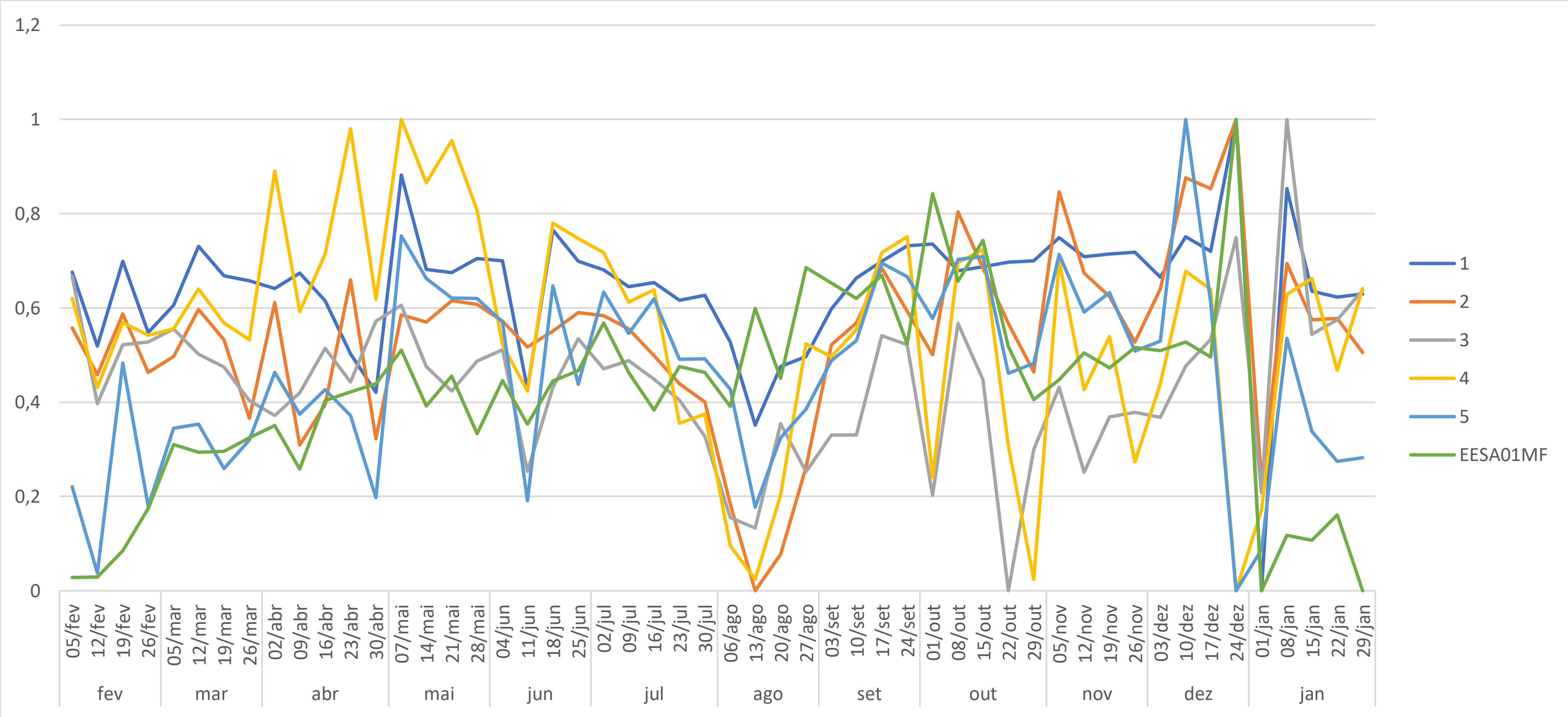
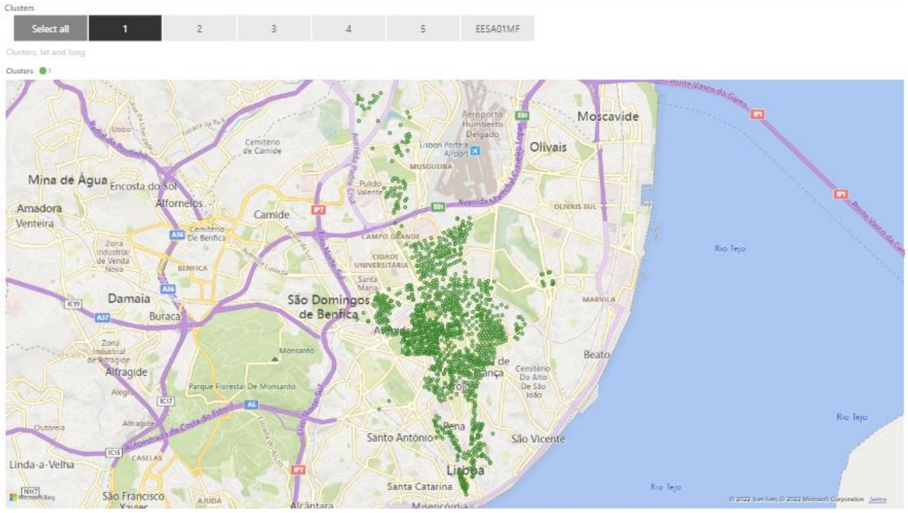
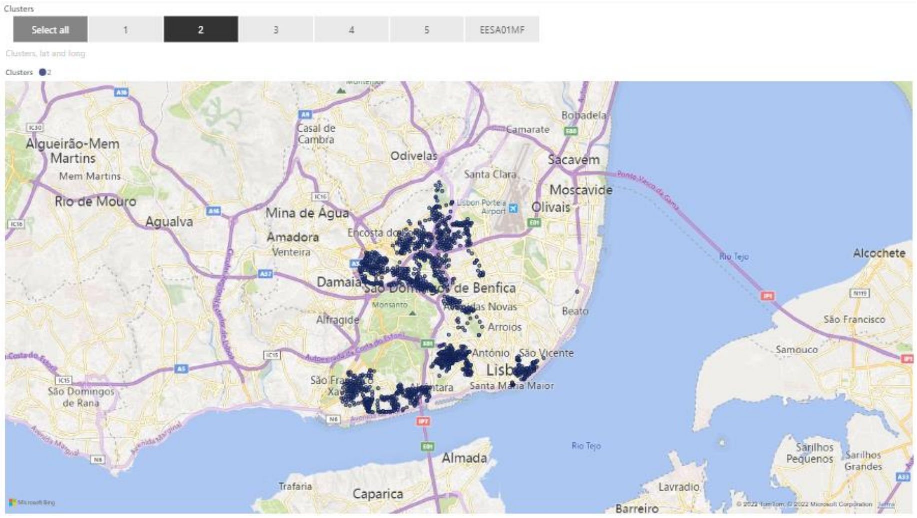
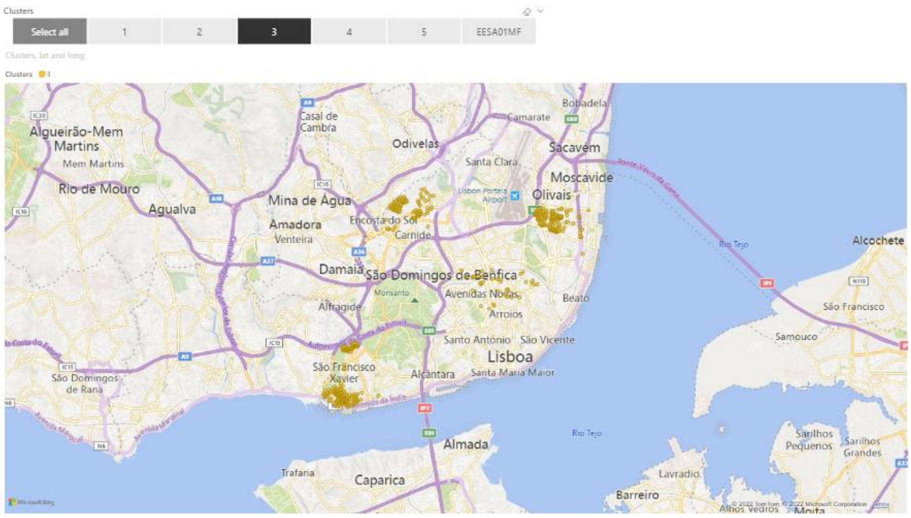


Figure 45 - Linear Plot by Segment *k*=5 and EESA01MF Plastic Waste.


Clusters
Select all
1
2
3
4
5
EESA01MF
Clusters, lat and long
Clusters 1
Moscavide
Olivais
Mina de Água
Encosta do Sol
Amadora
Venteira
Alfornelos
Carnide
Damaia
Buraca
São Domingos de Benfica
Alfragide
Monsanto
Parque Florestal De Monsanto
Beato
Rio Tejo
Santo António
São Vicente
Lisboa
Santa Catarina
Linda-a-Velha
São Francisco Xavier
Alcântara
Misericórdia


Clusters
Select all
1
2
3
4
5
EESA01MF
Clusters, lat and long
Clusters 2
Algueirão-Mem Martins
Mem Martins
Rio de Mouro
Agualva
Casal de Cambra
Odivelas
Camarate
Bobadela
Sacavém
Santa Clara
Moscavide
Olivais
Mina de Água
Amadora
Venteira
Damaia
São Domingos de Benfica
Avenidas Novas
Arroios
Alfragide
Beato
Alcochete
São Francisco
Samouco
São Vicente
Lisboa
Santa Maria Maior
São Domingos de Rana
Almada
Rio Tejo
Sarilhos Pequenos
Sarilhos Grandes
Trafaria
Caparica
Lavradio
Barreiro


Clusters
Select all
1
2
3
4
5
EESA01MF
Clusters, lat and long
Clusters 3
Algueirão-Mem Martins
Mem Martins
Rio de Mouro
Agualva
Casal de Cambra
Odivelas
Camarate
Bobadela
Sacavém
Santa Clara
Moscavide
Olivais
Mina de Água
Amadora
Venteira
Encosta do Sol
Carnide
Damaia
São Domingos de Benfica
Avenidas Novas
Arroios
Alfragide
Monsanto
Beato
Alcochete
São Francisco
Samouco
Santo António
São Vicente
Lisboa
São Francisco Xavier
Alcântara
Santa Maria Maior
São Domingos de Rana
Almada
Rio Tejo
Sarilhos Pequenos
Sarilhos Grandes
Trafaria
Caparica
Lavradio
Barreiro
Moita

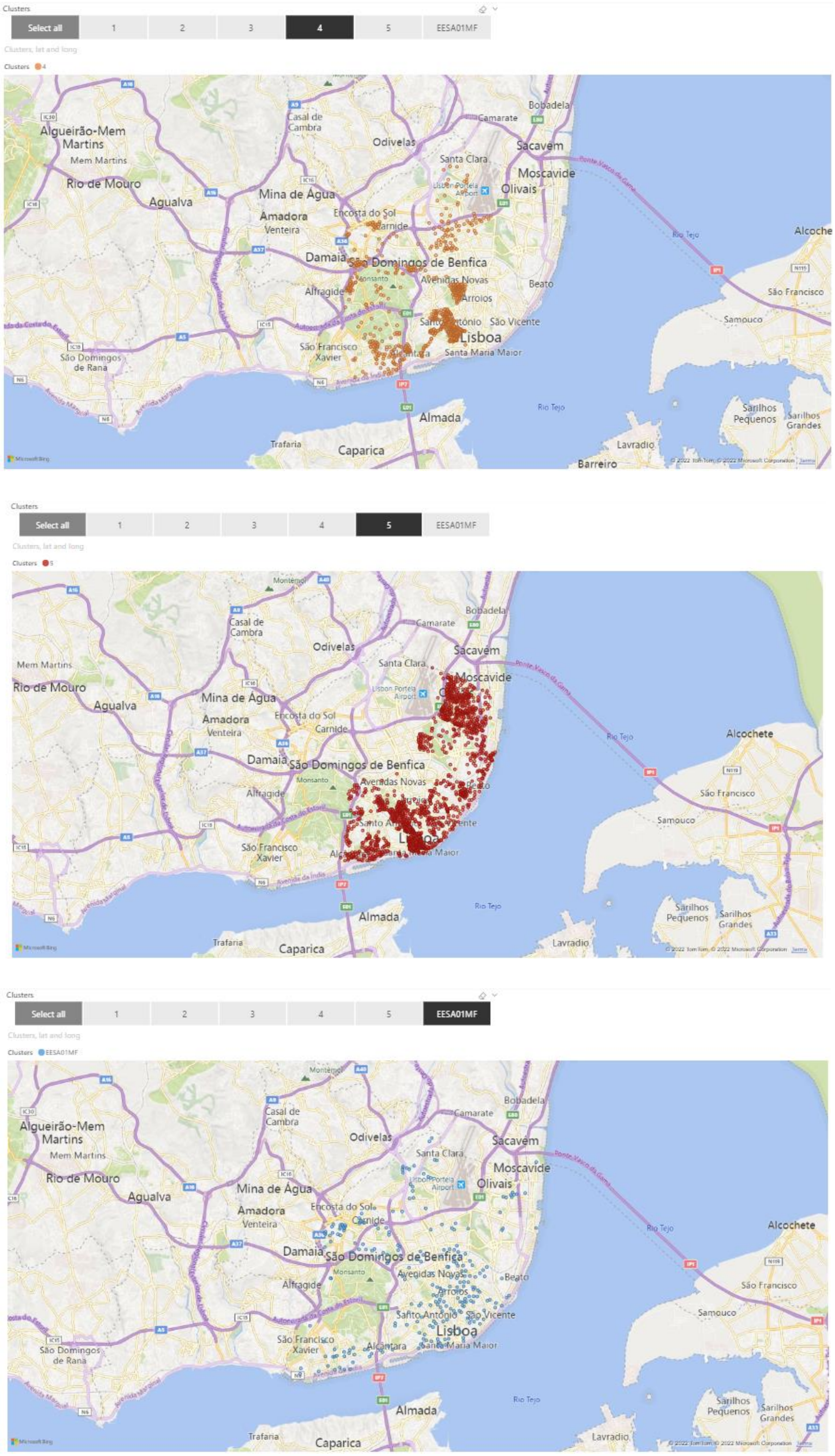


Figure 46 - PowerBI Plastic Waste, Cluster 1, 2, 3, 4, 5 and EESA01MF

Table 31 - Amount per each Variable Paper Waste, Z-Score, full table

	Cluster			
Variables	**1**	**2**	**3**	**P3101**
HM_0_14_years	0,728313664	0,488979851	0,254358503	-1,471652018
HM_15_24_years	0,741822237	0,704069845	-0,04889728	-1,396994802
HM_25_64_years	-0,698505776	-0,402302676	-0,383499574	1,484308026
HM_65_and_more_years	0,642765961	0,239957829	0,592966755	-1,475690546
H_0_14_years	0,742519431	0,464053017	0,26429984	-1,470872288
H_15_24_years	0,76152679	0,576107547	0,103693101	-1,441327438
H_25_64_years	-0,726322662	-0,4103096	-0,342215299	1,478847562
H_65_and_more_years	0,690910391	0,300751734	0,489180191	-1,480842317
M_0_14_years	0,722113469	0,527902588	0,217191748	-1,467207804
M_15_24_years	0,69120689	0,923378369	-0,379412596	-1,235172662
M_25_64_years	-0,662333981	-0,366860299	-0,459339918	1,488534198
M_65_and_more_years	0,515548857	0,005885831	0,879069405	-1,400504093
No Schooling level	-0,77558626	-0,292903912	-0,398985225	1,467475397
Basic Education	-0,998466415	-0,146449908	1,386767854	-0,241851531
1st Cycle	-1,18690203	-0,291020129	1,194177457	0,283744702
2nd Cycle	-0,728346196	0,107106773	1,374826164	-0,753586741
3rd Cycle	-0,421573998	0,039928351	1,362969229	-0,981323582
Secondary and post-secondary education	-0,763884618	-0,383729327	-0,323612442	1,471226387
University Education	1,261406824	0,306829485	-0,990304881	-0,577931428
No charges	0,685194532	0,17601484	0,601065426	-1,462274798
With charges	-0,685194532	-0,17601484	-0,601065426	1,462274798
Charges Less than 100 euros	0,173952874	0,309053357	0,932715542	-1,415721773
Charges 100 to 199,99 euros	-1,069101009	-0,085502023	1,346443361	-0,191840329
Charges 200 to 299,99 euros	-1,289067505	-0,263026913	0,596292832	0,955801585
Charges 300 to 399,99 euros	-1,307564927	-0,139869058	0,389667568	1,057766417
Charges 400to 649,99 euros	0,1408222	0,399733713	-1,426626027	0,886070114
Charges 650 to 999,99 euros	1,306033157	0,230749009	-0,941859484	-0,594922683
Charges 1000 or more euros	1,330006974	-0,145636664	-0,087950639	-1,096419671
Size Less than 30 m	-0,656826453	-0,678937218	-0,111698492	1,447462163
Size 30 m to 39 m	-0,677909222	-0,690122793	-0,067121071	1,435153086
Size 40 m to 49 m	-0,766120467	-0,716052985	0,107877249	1,374296203
Size 50 m to 59 m	-1,005287585	-0,608166946	0,408034351	1,205420179
Size 80 m to 99 m	-0,045578564	0,810850782	0,629199555	-1,394471773
Size 100 m to 119 m	0,845491145	0,848179624	-0,590441439	-1,103229331
Size 60 m to 79 m	-1,131315645	-0,092892048	1,305700898	-0,081493205
Size 120 m to 149 m	0,990853623	0,660207432	-0,494370488	-1,156690566
Size 150 m to 199 m	1,097931531	0,518970285	-0,478276367	-1,138625449
Size 200 m or more	1,142910089	0,15026021	-0,000429716	-1,292740583
Rent Less than 20 euros	-0,263811392	0,415269063	1,096125559	-1,24758323
Rent 20 to 49,99 euros	-0,307786609	0,100958531	1,302097869	-1,095269792

Rent 50 to 99,99 euros	-0,279244323	0,237079606	1,217938018	-1,1757733
Rent 100 to 199,99 euros	-0,897248876	-0,538112525	1,379212328	0,056149073
Rent 200 to 399,99 euros	-0,960693222	-0,58714034	0,25698333	1,290850232
Rent 400 to 649,99 euros	-0,530820234	-0,557247762	-0,408793067	1,496861062
Rent 650 to 999,99 euros	0,959969172	0,611964554	-1,269945558	-0,301988167
Rent 1000 or more euros	1,214021846	0,409349978	-0,668655295	-0,954716529
Family size with 1 person	-0,332376892	-0,762658071	-0,376692312	1,471727275
Family size with 2 people	0,312463307	0,154158005	0,943493122	-1,410114435
Family size with 3 people	0,171274214	0,768143513	0,515016442	-1,454434169
Family size with 4 people	0,701357456	0,659604469	0,076787479	-1,437749403
Family size with 5 or more people	-1,046306219	-0,192135263	-0,122890668	1,361332149
Before 1919	-0,41639968	-0,865762843	-0,151097896	1,43326042
1919 - 1945	-0,678137005	-0,538519032	-0,260579772	1,477235809
1946 - 1960	0,573442164	0,644495626	0,261118238	-1,479056029
1961 - 1980	0,38761723	0,768743066	0,31350433	-1,469864625
1981 - 2000	0,67943311	0,742452212	-0,01228093	-1,409604392
2001 - 2010	0,7205794	0,766226795	-0,11556335	-1,371242845
2011 - 2021	1,130494098	-0,15931645	0,297037636	-1,268215284
Hospital	0,656542112	0,920076379	-0,298641513	-1,277976979
Health Center	1,077026356	0,28075951	-0,03105148	-1,326734386
Pre Scholar	1,225843393	0,13916924	-0,159863918	-1,205148715
1st cycle	1,17531422	0,017566894	0,076926064	-1,269807178
2nd 3rd cycle	1,251812889	-0,115810174	0,054212998	-1,190215714
Secondary School	1,177018022	0,434298366	-0,552830928	-1,05848546
University	0,448046926	0,959856312	-0,038055962	-1,369847275
Train Station	-0,278261168	-0,413589074	1,466680764	-0,774830522
Metro Station	1,386525557	-0,125920889	-0,264837737	-0,995766931
Bus Stop	1,159410534	0,033496373	0,090253094	-1,28316
Rest Coffe Shop Bar Market	1,229576071	0,146283577	-0,177905847	-1,197953802
Culture Spot	1,2643478	0,182897346	-0,323176607	-1,124068539
Beds Tourist Establishments	1,072109795	0,140511425	0,133828572	-1,346449791
Beds Local accommodation	1,329359804	0,077702295	-0,359399543	-1,047662557
Cultural and Sporting events	1,115555702	-0,255749936	0,382653105	-1,242458871
Coverage percentage by building built continuous predominantly vertical	0,230202357	-0,40946811	-1,085797818	1,265063571
Coverage percentage by building built continuous predominantly horizontal	-0,545657945	0,500736991	1,129373059	-1,084452105
Percentage of coverage by building built discontinuous	0,387849465	0,982550833	0,002247365	-1,372647662
Coverage percentage by building built discontinuous sparse	-0,401929497	1,399441665	-0,061434438	-0,93607773
Industry	0,069469336	0,195687029	1,076083608	-1,341239973
Trade	-0,31087754	0,620425896	0,956647607	-1,266195964
Sports facilities	0,364885989	0,601517342	0,526270326	-1,492673657

Leisure Equipment	1,439190741	-0,081143845	-0,679023448	-0,679023448
Cultural Equipment	-1,174383004	-0,360214191	0,371333937	1,163263257
Price per sqm	-0,027732324	-0,880432997	-0,497675327	1,405840647

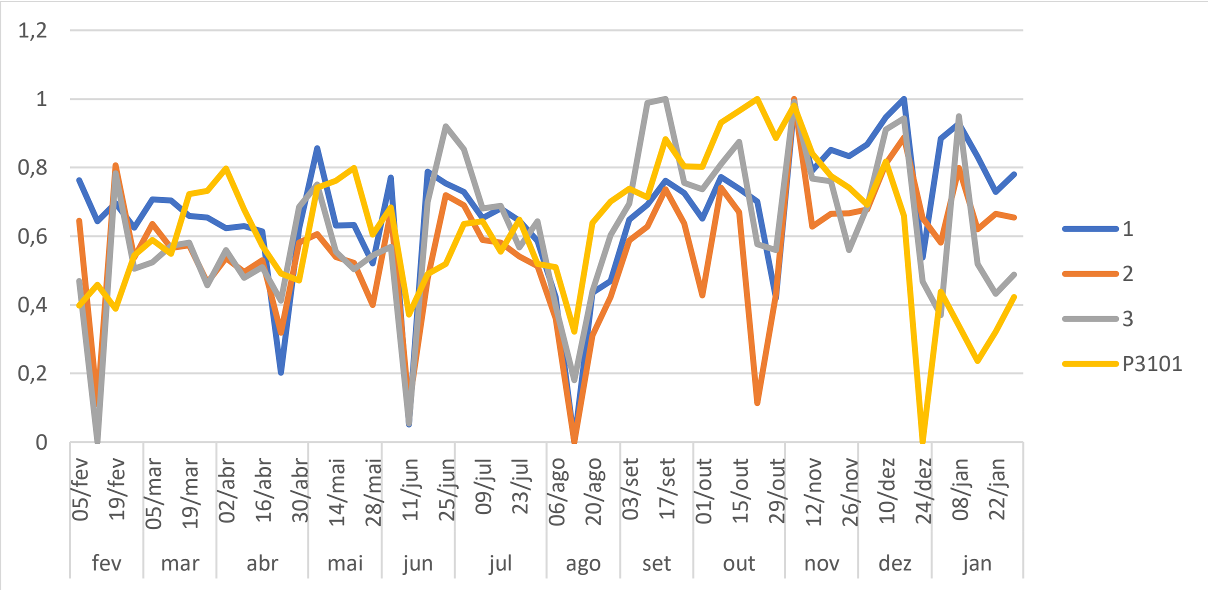
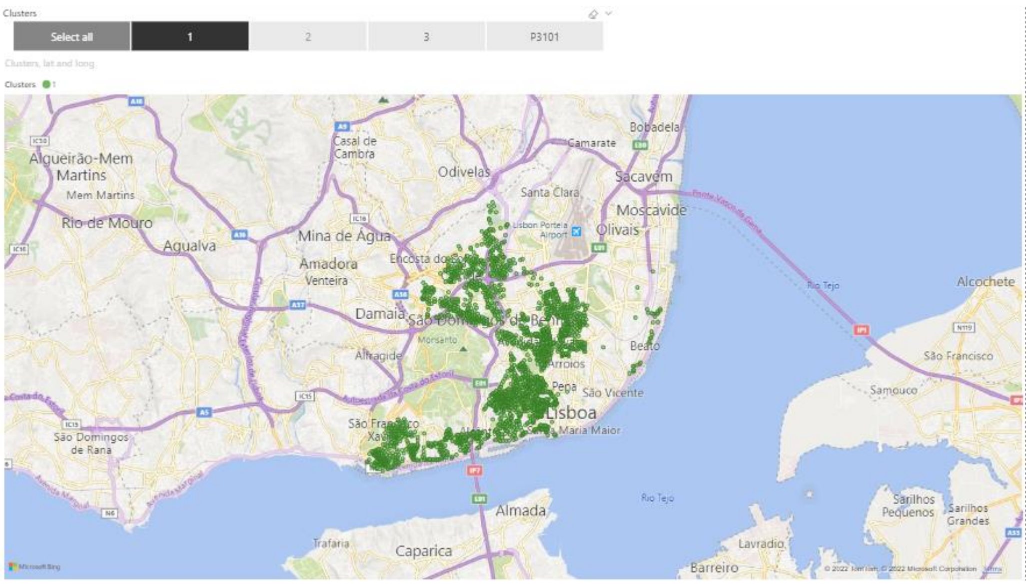


Figure 47 - Linear Plot by Segment *k*=3 and P3101 Paper Waste.



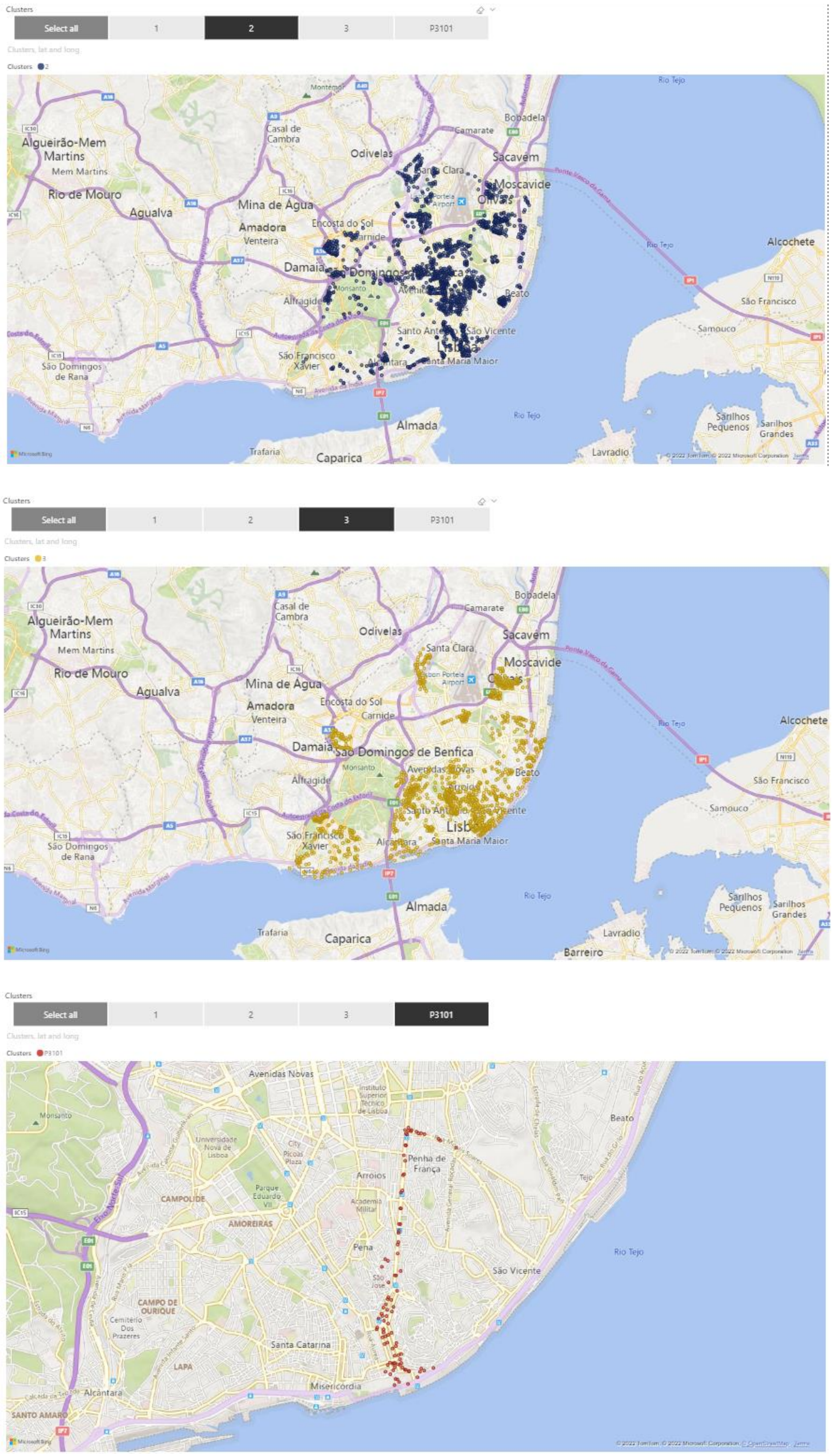


Figure 48 - PowerBI Paper Waste, Cluster 1, 2, 3, and P3101

Table 32 - Amount per each Variable Glass Waste, Z-Score, full table

	Cluster				
Variables	**1**	**2**	**3**	**4**	**V1305**
HM_0_14_years	0,424815	0,680285	0,165202	0,487929	-1,758231
HM_15_24_years	0,279873	0,634451	0,003512	0,785654	-1,703490
HM_25_64_years	-0,624699	-0,342006	-0,106797	-0,668482	1,741984
HM_65_and_more_years	1,014920	-0,337079	0,072326	0,758431	-1,508598
H_0_14_years	0,458330	0,615818	0,131961	0,551489	-1,757597
H_15_24_years	0,372701	0,570346	0,107750	0,693638	-1,744436
H_25_64_years	-0,567351	-0,364917	-0,077022	-0,726144	1,735434
H_65_and_more_years	0,798873	-0,109974	-0,016787	0,913656	-1,585768
M_0_14_years	0,387597	0,739337	0,205151	0,423416	-1,755500
M_15_24_years	-0,018525	0,771521	-0,364244	1,069779	-1,458531
M_25_64_years	-0,771918	-0,273098	-0,122465	-0,563660	1,731141
M_65_and_more_years	1,466510	-1,233531	0,095005	0,199971	-0,527954
No Schooling level	-0,775676	0,287246	0,275770	-1,161967	1,374627
Basic Education	0,024418	-0,950060	-0,569960	-0,161463	1,657065
1st Cycle	-0,108025	-0,745198	-0,448793	-0,440774	1,742789
2nd Cycle	-0,146883	-0,736793	-0,861022	0,106773	1,637926
3rd Cycle	0,886252	-1,358063	-0,292957	1,086643	-0,321876
Secondary and post-secondary education	-0,476536	-0,648601	0,122449	-0,689149	1,691837
University Education	0,181676	0,795421	0,354882	0,411339	-1,743318
No charges	0,289144	-0,444259	-0,219399	-1,154526	1,529040
With charges	-0,289144	0,444259	0,219399	1,154526	-1,529040
Charges Less than 100 euros	0,295268	-1,108803	-0,075862	1,512899	-0,623503
Charges 100 to 199,99 euros	-0,309068	-1,047315	-0,424565	0,168855	1,612092
Charges 200 to 299,99 euros	-0,423422	-1,028046	-0,277428	0,092331	1,636565
Charges 300 to 399,99 euros	0,089430	-1,570420	-0,234187	0,861104	0,854073
Charges 400to 649,99 euros	0,463467	0,255521	0,326820	0,715288	-1,761096
Charges 650 to 999,99 euros	0,060191	1,258703	0,410231	-0,252673	-1,476452
Charges 1000 or more euros	-0,286766	0,822934	-0,191500	-1,426223	1,081555
Size Less than 30 m	-0,477256	-0,275906	-0,204606	-0,785153	1,742921
Size 30 m to 39 m	-0,444013	-0,297835	-0,185227	-0,811237	1,738313
Size 40 m to 49 m	-0,442199	-0,279847	-0,110829	-0,881065	1,713939
Size 50 m to 59 m	-0,453541	-0,373432	-0,079336	-0,819217	1,725526
Size 80 m to 99 m	0,649010	-0,299965	0,056453	1,101744	-1,507242
Size 100 m to 119 m	0,414805	0,345223	0,140936	0,830734	-1,731698
Size 60 m to 79 m	0,539035	-1,708132	0,377924	0,805418	-0,014245
Size 120 m to 149 m	0,238126	0,447779	0,118627	0,902548	-1,707081
Size 150 m to 199 m	0,155383	0,956896	0,150531	0,426944	-1,689753
Size 200 m or more	0,262752	1,531731	-0,116794	-1,125026	-0,552663
Rent Less than 20 euros	0,530177	-0,020206	-0,590911	1,337785	-1,256845
Rent 20 to 49,99 euros	0,346175	-0,738402	-1,341309	0,995689	0,737848
Rent 50 to 99,99 euros	0,506847	-0,152686	-0,479159	1,382042	-1,257044

Rent 100 to 199,99 euros	-0,185323	-0,585448	-0,615261	-0,375586	1,761617
Rent 200 to 399,99 euros	-0,489218	-0,993570	-0,101548	-0,076197	1,660532
Rent 400 to 649,99 euros	-0,486137	-0,642684	0,712631	-0,965826	1,382015
Rent 650 to 999,99 euros	0,055254	0,758859	0,659295	0,238218	-1,711626
Rent 1000 or more euros	0,269008	1,322269	0,308931	-0,577005	-1,323203
Family size with 1 person	-0,420172	-0,348605	-0,202551	-0,777104	1,748432
Family size with 2 people	0,510748	-0,207018	0,339590	0,977191	-1,620511
Family size with 3 people	0,344323	0,247860	0,160144	0,948857	-1,701185
Family size with 4 people	0,402929	0,493429	0,099096	0,745031	-1,740484
Family size with 5 or more people	0,317440	1,257919	0,313652	-0,478507	-1,410505
Before 1919	-0,393577	-0,302567	-0,218558	-0,824726	1,739428
1919 - 1945	-0,528191	0,270061	0,653632	-1,451273	1,055771
1946 - 1960	0,583194	0,448883	0,270775	0,474708	-1,777561
1961 - 1980	0,352631	0,023899	0,085366	1,140387	-1,602283
1981 - 2000	0,386561	0,121674	-0,259810	1,242792	-1,491218
2001 - 2010	0,069660	0,101518	-0,215648	1,425329	-1,380860
2011 - 2021	0,534415	0,878192	0,679887	-0,664365	-1,428129
Hospital	-0,460964	1,346559	0,772944	-0,802123	-0,856415
Health Center	1,126732	0,465311	0,279092	-0,350122	-1,521014
Pre Scholar	0,594721	1,095990	0,128745	-0,287905	-1,531551
1st cycle	0,549849	0,807654	0,533671	-0,248348	-1,642827
2nd 3rd cycle	0,243893	0,912790	0,432239	0,116078	-1,705000
Secondary School	0,245032	1,182791	0,501287	-0,501443	-1,427667
University	0,801920	0,740888	-0,054511	0,177134	-1,665431
Train Station	1,734389	-0,142623	-0,288886	-0,511171	-0,791709
Metro Station	0,834322	1,034927	0,204494	-0,894023	-1,179720
Bus Stop	0,630327	1,149657	0,093563	-0,428279	-1,445268
Rest Coffe Shop Bar Market	0,587078	0,973735	0,550290	-0,758921	-1,352183
Culture Spot	0,661085	1,128178	0,002344	-0,310461	-1,481145
Beds Tourist Establishments	0,384595	0,928391	0,813307	-0,878554	-1,247738
Beds Local accommodation	0,698261	1,155394	0,147604	-0,723872	-1,277388
Cultural and Sporting events	0,987895	0,924726	0,127096	-0,829374	-1,210343
Coverage percentage by building built continuous predominantly vertical	-0,705064	-0,627675	-0,057201	-0,339358	1,729298
Coverage percentage by building built continuous predominantly horizontal	1,107619	0,965004	-0,248165	-0,671356	-1,153101
Percentage of coverage by building built discontinuous	1,047536	0,943986	-0,150930	-0,536313	-1,304280
Coverage percentage by building built discontinuous sparse	0,623717	-0,697049	1,467429	-0,697049	-0,697049
Industry	0,198572	0,081698	1,135509	0,203784	-1,619563
Trade	-0,602058	0,603206	-0,829160	1,473100	-0,645088
Sports facilities	0,067894	1,442420	0,012851	-0,147322	-1,375843
Leisure Equipment	1,788854	-0,447214	-0,447214	-0,447214	-0,447214
Cultural Equipment	0,804663	0,146550	1,036076	-0,617528	-1,369761

Price per sqm	-0,247713	-0,114583	-0,239991	-1,056528	1,658814

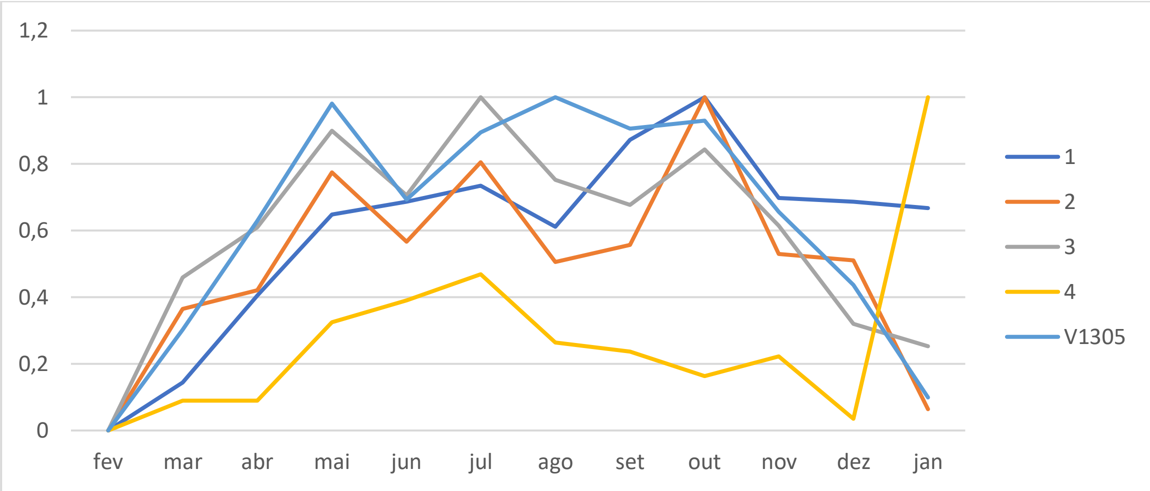
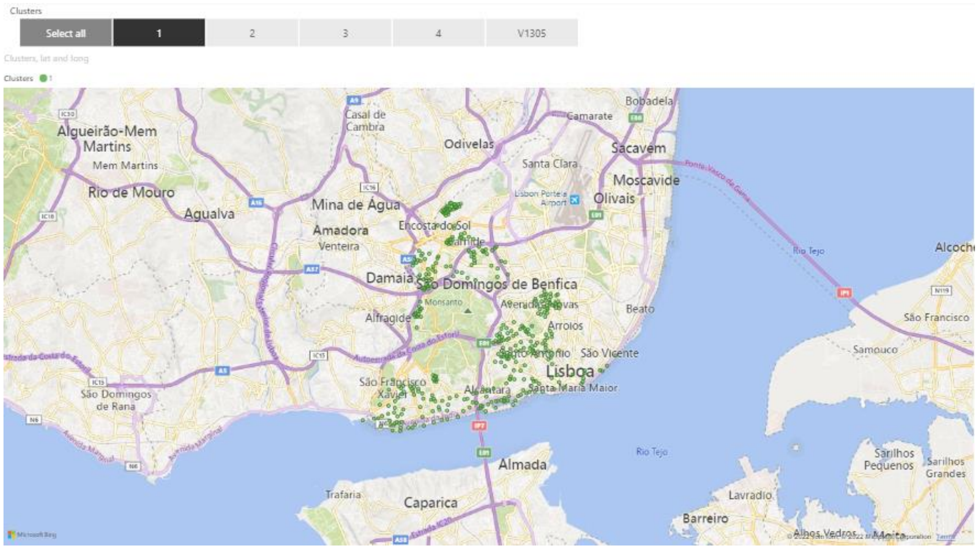
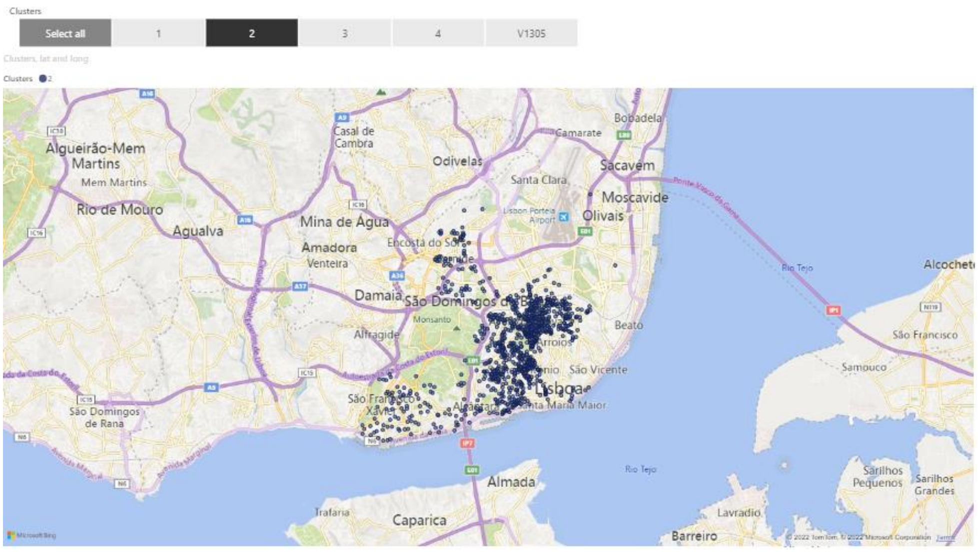


Figure 49 - Linear Plot by Segment *k*=4 and V1305 Glass Waste.





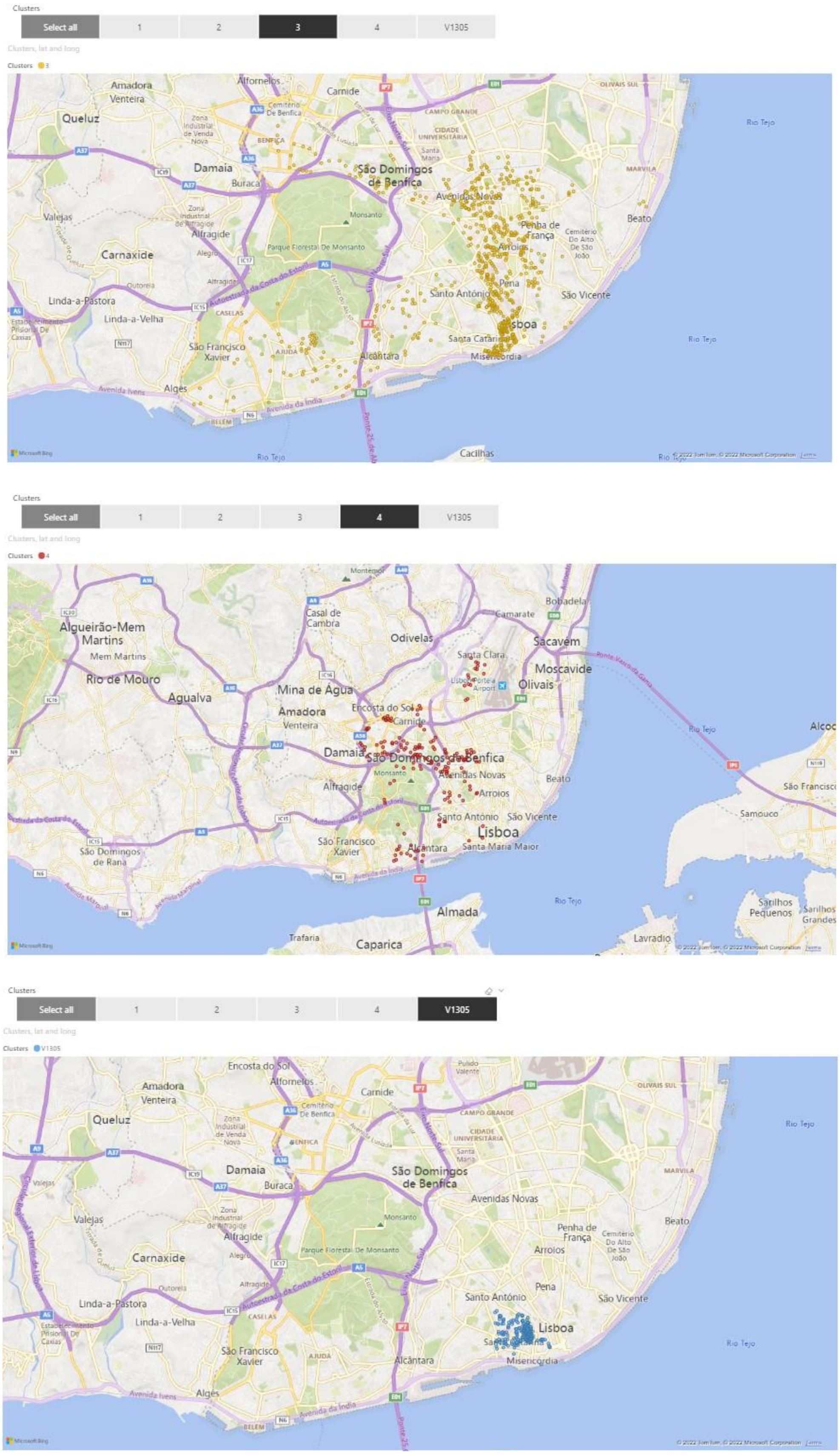


Figure 50 - PowerBI Glass Waste, Cluster 1, 2, 3, 4 and V1305