PERFORMANCE INDICATORS FOR EMISSIONS REPORTING BASED ON ARTIFICIAL INTELLIGENCE

Victor de Almeida Xavier

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientadores: Priscila Machado Vieira Lima
Felipe Maia Galvão França

Rio de Janeiro
Janeiro de 2022

PERFORMANCE INDICATORS FOR EMISSIONS REPORTING BASED ON ARTIFICIAL INTELLIGENCE

Victor de Almeida Xavier

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientadores: Priscila Machado Vieira Lima
Felipe Maia Galvão França

Aprovada por: Prof. Nelson Maculan Filho
Prof. Josefino Cabral Melo Lima
Profª. Priscila Machado Vieira Lima
Prof. Felipe Maia Galvão França

RIO DE JANEIRO, RJ – BRASIL
JANEIRO DE 2022

Xavier, Victor de Almeida

Performance Indicators for Emissions Reporting based on Artificial Intelligence/Victor de Almeida Xavier. – Rio de Janeiro: UFRJ/COPPE, 2022.

XVII, 159 p.: il.; 29, 7cm.

Orientadores: Priscila Machado Vieira Lima

Felipe Maia Galvão França

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2022.

Referências Bibliográficas: p. 87 – 90.

1. KPI. 2. Performance. 3. Indicators. 4. Maturity Model. 5. CDP. 6. Carbon Disclosure Project. 7. Emissions Reporting. 8. Clustering. 9. WiSARD. 10. ClusWiSARD. 11. Hierarchical. 12. k-means. 13. DBSCAN. I. Lima, Priscila Machado Vieira *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

...to every person that works
hard every day to make the world
a better place to live...

Acknowledgements

Gostaria de agradecer, primeiramente, a Deus, que me permitiu chegar até aqui. À minha esposa, Lúcia, por todo o suporte e carinho nos momentos mais decisivos. Aos meus filhos, Pedro, João Victor e Ângelo, amigos e familiares, por todo o incentivo que tive ao longo do processo e pela paciência pela privação da minha presença. Agradeço aos meus pais, José Carlos e Márcia, pela minha vida e por acreditarem no meu potencial. Agradeço à minha dindinha Aninha, por todo o apoio que permitiu, entre outras conquistas, eu poder escrever essa dissertação em inglês. E à minha querida Vovó Anna (in memoriam), pelo seu amor, pelo seu carinho, pelo seu silêncio, pelo seu apoio e por sua máquina de costura Singer class 127...

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

INDICADORES DE PERFORMANCE PARA REGISTRO DE EMISSÕES BASEADO EM INTELIGÊNCIA ARTIFICIAL

Victor de Almeida Xavier

Janeiro/2022

Orientadores: Priscila Machado Vieira Lima
Felipe Maia Galvão França

Programa: Engenharia de Sistemas e Computação

As mudanças climáticas e o aquecimento global têm sido um tema discutido em todo o mundo desde a conferência Eco-92. Entretanto, poucos avanços na redução das emissões de gases de efeito estufa (GEE) foram verificados até agora. Os problemas e desafios relacionados às emissões são complexos e exigem um esforço comum e amplo para serem enfrentados. O Relatório de Emissões é um dos aspectos centrais nas políticas de redução de emissões de GEE e, por isso, é o foco do presente trabalho.

Este trabalho apresenta um método para explorar, agrupar e analisar dados de iniciativas de registro de emissões. Utilizando tecnologias de inteligência artificial, conceitos de indicadores de desempenho e abordagens de análise qualitativa, o método proposto é implementado através de um processo de desenvolvimento de indicadores de performance (PIDP), cujo objetivo é o de procurar por indicadores de performance entre os dados provenientes de bases de dados de emissões.

Durante a execução do PIDP, os resultados indicaram que um novo modelo para tratar os registros de emissões era necessário. Assim, esse trabalho propõe um novo modelo de avaliação de processos relacionados ao registro de emissões implementados pelas cidades, e que é baseado em conceitos herdados do modelo de maturidade de capacidade (CMM). O objetivo principal deste modelo é prover orientação a essas cidades ao tratarem com os desafios de redução de emissões através do melhoramento dos processos e áreas relacionados ao registro de emissões.

Ao longo deste estudo, esse modelo e como ele pode ser utilizado no contexto das tarefas de registro de emissões será descrito em detalhes, assim como os experimentos e outros resultados obtidos durante o seu desenvolvimento.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

PERFORMANCE INDICATORS FOR EMISSIONS REPORTING BASED ON ARTIFICIAL INTELLIGENCE

Victor de Almeida Xavier

January/2022

Advisors: Priscila Machado Vieira Lima
Felipe Maia Galvão França

Department: Systems Engineering and Computer Science

Climate change and global warming have been a trending topic worldwide since the Eco-92 conference. However, little progress has been made in reducing greenhouse gases (GHGs). The problems and challenges related to emissions are complex and require a concerted and comprehensive effort to address them. Emissions reporting is a key component of GHG reduction policy and is therefore the focus of this work.

This work presents a method for examining, clustering, and analysing data from emissions reporting initiatives. Using artificial intelligence clustering technologies, performance indicator concepts and qualitative analysis approaches, the proposed method is implemented through a performance indicator development process (PIDP), which aims to search for performance indicators (PIs) among data selected from emissions databases.

During the implementation of the PIDP, the results showed that a new model is essential to deal with emission reporting information. Therefore, this study proposes a new model to evaluate emissions reporting processes implemented by cities, which is based on concepts inherited from the capability maturity model (CMM). The main objective of this model is to help cities address the challenges of emission reduction by leveraging the areas and processes associated with emission reporting.

This model and how it can be used in the context of emissions reporting is described in detail in the methodology, as are the experiments and other results obtained during the development of this study.

Contents

List of Figures

List of Tables

List of Abbreviations

AI	artificial intelligence, p. 4
ANN	artificial neural networks, p. 13
CCI	climate change indicators, p. 3
CDP	Carbon Disclosure Project, p. 3
CEM	capability evaluation matrix, p. 44
CMMII	Capability Maturity Model Integration Institute, p. 13
CMM	capability maturity model, p. vii
CO2MM	carbon management models, p. 6
CSV	comma-separated values, p. 7
DBSCAN	Density based spatial clustering of applications with noise, p. 19
DMC	data management context, p. 44
DMMM	data management maturity model, p. 13
ECI	emissions correlations indicators, p. 46
EDI	electronic data interchange, p. 85
EPA	U.S. Environmental Protection Agency, p. 1
ERM-L	emissions reporting maturity level, p. 8
ESG	environmental, social and governance, p. 6
GCoM	Global Covenant of Mayors for Climate and Energy, p. 4
GDP	gross domestic product, p. 7
GHGs	greenhouse effect gases, p. vii

HDI	human development index, p. 7
IBGE	Instituto Brasileiro de Geografia e Estatística, p. 25
IND	quality indicator, p. 75
IPCC	Intergovernmental Panel on Climate Change, p. 1
ISC	International Science Council, p. 1
IoT	Internet of things, p. 86
KPIs	key performance indicators, p. 3
KRI	key result indicator, p. 11
OECD	organisation for Economic Co-operation and Development, p. 65
OPEC	organisation of the Petroleum Exporting Countries, p. 76
OWID	Our World In Data, p. 24
PIDP	performance indicators development process, p. vii
PIER	performance indicators for emissions reporting, p. 46
PIs	performance indicators, p. vii
RAM	random access memory, p. 14
ROI	Return Over Investment, p. 86
SCI	smart cities index, p. 7
SHDI	sub-national human development index, p. 7
UML	unified modelling language, p. 25
UNFCC	United Nations Framework Convention on Climate Change, p. 1
WCRP	World Climate Research Programme, p. 1
WNN	weightless neural networks, p. 14
WiSARD	Wilkie, Stonham & Aleksander's Recognition Device, p. 15

Chapter 1

Introduction

1.1 Climate Change and Global Warming

The Paris Agreement [1], implemented by the United Nations Framework Convention on Climate Change (UNFCC) [2] conducted in 2015, states that it is imperative to keep the increase in global average temperature to well below two degrees Celsius above pre-industrial levels. Ideally, nations should take all necessary measures to reach the 1.5 degrees Celsius target, as this can reduce the risks and impacts of climate change.

The Intergovernmental Panel on Climate Change (IPCC)[3], the United Nations body responsible for the scientific assessment of climate change, has produced a special report on the impacts of global warming of 1.5 degrees Celsius above pre-industrial levels and associated global greenhouse gas emission pathways[4]. The report was prepared in response to the Paris Agreement proposals and highlights the implications by comparing the two scenarios of 1.5 and 2 degrees Celsius, as well as the mitigation alternatives that can be applied as part of a global effort, to strengthen the global response to the threat of climate change and sustainable development that can contribute in the poverty eradication efforts.

Other leading organisations involved in climate change research, policymaking and education such as the International Science Council (ISC)[5], U.S. Environmental Protection Agency (EPA)[6], World Climate Research Programme (WCRP)[7], all point in the same direction: the urgency of effective policies to reduce greenhouse gas emissions, negotiated globally and implemented locally.

---

[1] https://unfccc.int/sites/default/files/english_paris_agreement.pdf
[2] https://unfccc.int
[3] https://www.ipcc.ch
[4] https://www.ipcc.ch/sr15/
[5] https://council.science/
[6] https://www.epa.gov/
[7] https://www.wcrp-climate.org/

1.2 Greenhouse Effect Gases (GHG) Emissions

Greenhouse gas emissions contribute significantly to the rise in global temperature. For this reason, reducing emissions of these gases should be a central component of strategies to mitigate global warming and the effects of climate change. The figure 1.1 illustrates how greenhouse gas emissions are distributed globally by looking at the emission totals of the main gas (CO_2).

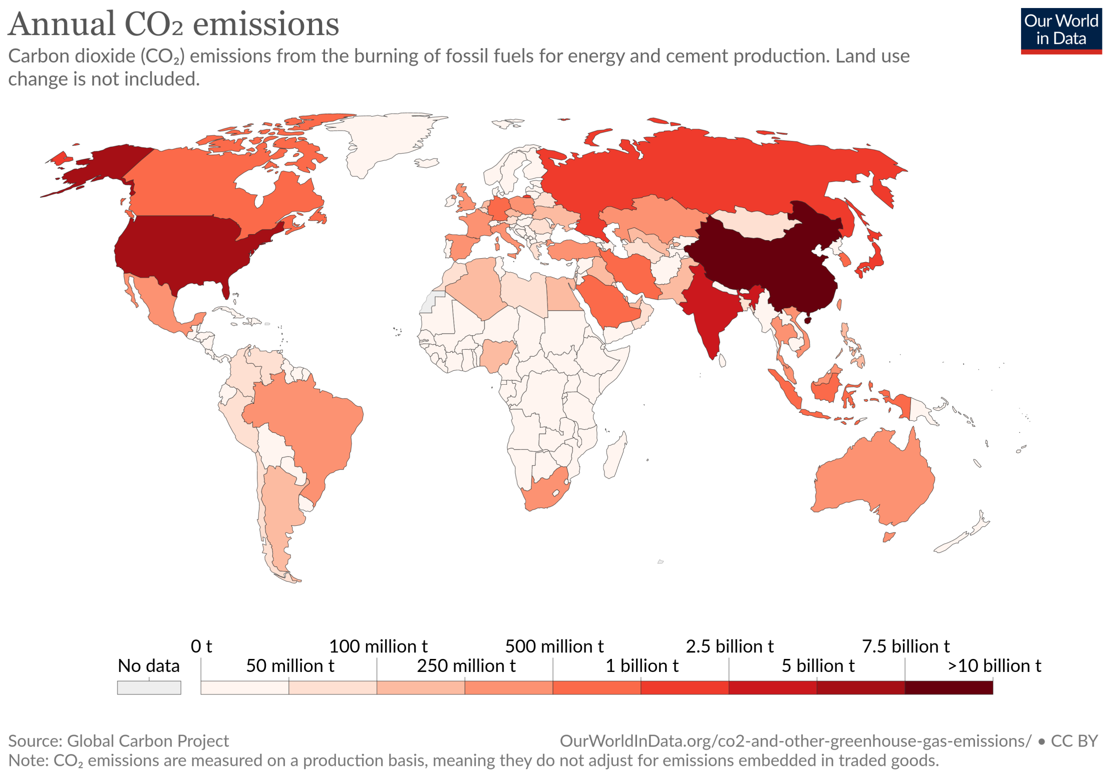

Figure 1.1: CO_2 emissions world map 2019. Source: https://ourworldindata.org/co2-emissions

Greenhouse gas emissions include various gases and are normalised using carbon equivalents [8]. The table 1.1 shows the carbon equivalent relationships between CO_2 and other gases that make up greenhouse gas emissions [9].

Although the link between the rise in global temperature and the increase in extreme weather events has been scientifically proven, governments still have to contend with disbelief and lobbies that mislead measures to reduce local GHG emissions. According to BROADSTOCK *et al.* [1], the link between GHG emissions and economic activity is well established, as is the disconnect between environmental and social responsibility in measuring corporate performance. One of the reasons

[8] https://unfccc.int/process-and-meetings/transparency-and-reporting/greenhouse-gas-data/frequently-asked-questions

[9] https://unfccc.int/process/transparency-and-reporting/greenhouse-gas-data/greenhouse-gas-data-unfccc/global-warming-potentials

Table 1.1: Global Warming Potentials (IPCC Second Assessment Report).

Species	**Chemical formula**	**Lifetime (years)**	**Global Warming Potential (Time Horizon)**		
			20 years	**100 years**	**500 years**
Carbon dioxide	CO_2	variable	1	1	1
Methane	CH_4	12±3	56	21	6.5
Nitrous oxide	N_2O	120	280	310	170
HFC-23	CHF_3	264	9100	11700	9800
HFC-32	CH_2F_2	5.6	2100	650	200
HFC-41	CH_3F	3.7	490	150	45
HFC-43-10mee	$C_5H_2F_{10}$	17.1	3000	1300	400
HFC-125	C_2HF_5	32.6	4600	2800	920
HFC-134	$C_2H_2F_4$	10.6	2900	1000	310
HFC-134a	CH_2FCF_3	14.6	3400	1300	420
HFC-152a	$C_2H_4F_2$	1.5	460	140	42
HFC-143	$C_2H_3F_3$	3.8	1000	300	94
HFC-143a	$C_2H_3F_3$	48.3	5000	3800	1400
HFC-227ea	C_3HF_7	36.5	4300	2900	950
HFC-236fa	$C_3H_2F_6$	209	5100	6300	4700
HFC-245ca	$C_3H_3F_5$	6.6	1800	560	170
Sulphur hexafluoride	SF_6	3200	16300	23900	34900
Perfluoromethane	CF_4	50000	4400	6500	10000
Perfluoroethane	C_2F_6	10000	6200	9200	14000
Perfluoropropane	C_3F_8	2600	4800	7000	10100
Perfluorobutane	C_4F_{10}	2600	4800	7000	10100
Perfluorocyclobutane	c-C_4F_8	3200	6000	8700	12700
Perfluoropentane	C_5F_{12}	4100	5100	7500	11000
Perfluorohexane	C_6F_{14}	3200	5000	7400	10700

highlighted by the authors is the non-reporting of emissions, a recurring problem also seen in emissions reporting by local governments. Emissions data are widely available from a variety of sources. EPA maintains a catalogue of four climate change indicators (CCI) related to GHG emissions. The Figure 1.2 shows the increase in GHG emissions from 1990 to 2015, but examples of emissions reporting that efficiently and effectively contribute to emissions reduction through mitigation actions are still hard to find.

1.3 GHG Impacts Mitigation Initiatives

Carbon Disclosure Project (CDP)[10] is an initiative that promotes collaboration on emissions reduction and focuses on obtaining reliable data from cities and businesses worldwide to help them manage their environmental impacts. To drive the exploration and analysis of the data, CDP enlisted the infrastructure and expertise of Kaggle[11] to promote a competition whose main objective was to discover key performance indicators (KPIs) among the responses provided. The database provided

[10] https://www.cdp.net/en
[11] https://www.kaggle.com/c/cdp-unlocking-climate-solutions

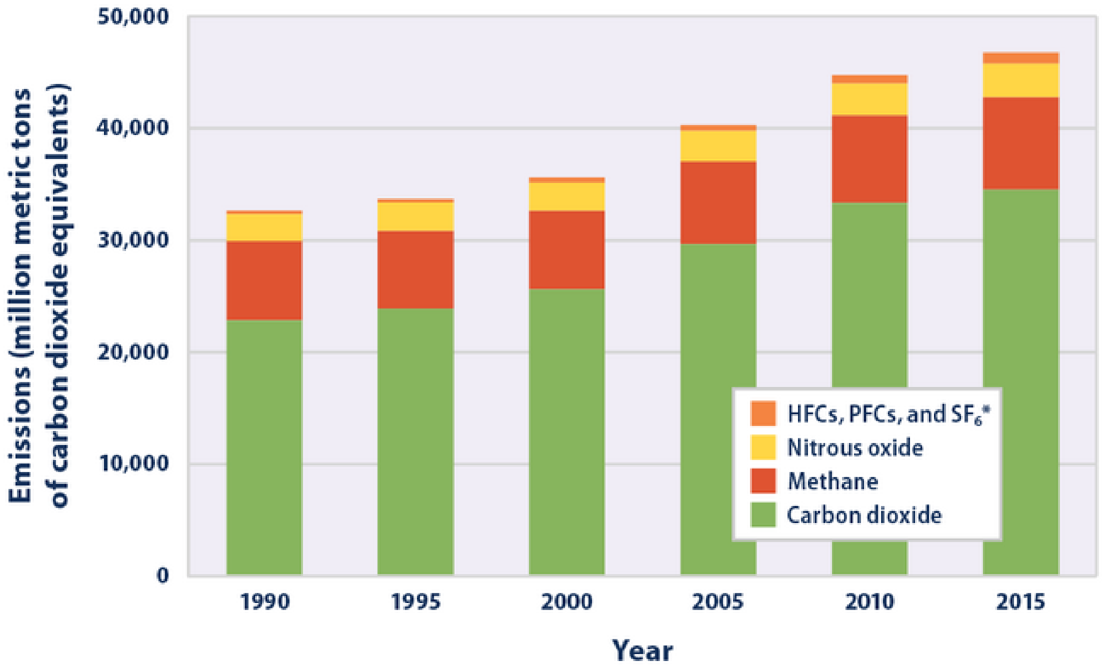

Figure 1.2: Global Greenhouse Gas Emissions by Gas, 1900-2015. Source: EPA.

is based on questionnaires that CDP deployed in 2018, 2019 and 2020 to some cities and companies around the world. Thus, this work attempts to show the relationship between the information provided and the policies already in place that lead to emissions reduction and the associated benefits, both locally and globally.

Other initiatives, such as the Global Covenant of Mayors for Climate and Energy (GCoM)[12] and C40 cities[13], are also the subject of this work, as they provide complementary and useful information on emissions at the city level. Despite the efforts of the selected cities, there are still some problems to be solved in emissions reporting in order for these cities to effectively contribute to the reduction of GHG emissions.

1.4 Goals and Limits

The main objective of this work is to use artificial intelligence (AI) to support initiatives to improve emissions reporting. It is crucial to improve the process efficiency of emissions reporting in order to achieve better emissions reduction results, as there is a direct link between effective emissions policies implemented by cities and emissions reduction (or increase) due to the effectiveness of these policies.

To achieve this goal, this work proposes a series of steps to investigate, search and develop performance indicators (PIs) for emissions reporting. These perfor-

[12]https://www.globalcovenantofmayors.org/
[13]https://www.c40.org/

mance indicators are based on the data provided by cities on the processes they go through to address emission problems. PIs can be used to guide and optimize the policies responsible for implementing emission reduction measures at the city level. Therefore, as a by-product of this work, a process for developing these indicators is proposed to organize the steps necessary to find candidate performance indicators among the data provided by cities.

1.4.1 Emissions Reporting Analysis using AI

Emissions reporting analysis can be made using statistical tools and techniques, also known as Analytics. This approach has already been used to produce relevant information in the field of emissions impact analysis [2][3], such as indicators and correlations with external indices[4], but it lacks a qualitative view of the data, which AI can also help with, and this is one of the analysis mechanisms used in this work.

More than ever, algorithms and artificial intelligence techniques play a key role in every field of knowledge, especially when it comes to solving problems through optimization. Also, in the challenges and problems related to emissions reporting, these algorithms and techniques can be used to address and even solve some of them, such as data processing, integrity and usefulness.

1.4.2 Performance Indicators for Emissions

Performance indicators (PIs) are one of the most commonly used tools for evaluating processes in terms of their effectiveness. Therefore, processes related to emissions reporting can also benefit from the concepts and formalization of the performance indicator development process. To achieve this goal, a performance indicator development process (PIDP) should be applied to emissions reporting processes to proceed with assessments based on available data.

The emissions reporting processes are subject to the PIDP, in which the analysis of candidate PI plays an important role, as this PI will be used to improve them. Thus, the candidates for PI can be used both to evaluate the effectiveness of the overall emissions reporting process and to search for other PIs candidates among the relationships with external indices and indicators.

1.4.3 Scope of this work

Although the techniques and procedures developed in this work are not exclusively applied in emissions-related fields, this work focuses on one particular field: emissions reporting. For this reason, this work will not address issues that are not clearly related to emissions reporting. Another boundary established is the focus on the city as the fundamental unit of the information provider, represented by the data it processes. The same occurs with the business implications of emissions are applied to city outcomes. Even if they are related to emissions reporting, they are outside the scope of this work because the relationships between businesses and cities, like regulations, are complex enough to require specialized work in this area. Reporting on environmental, social and governance (ESG) aspects, for example, is outside the scope of this work.

Emissions reporting processes can be considered as part of emissions reduction initiatives that can be organised into programs such as carbon management powered by carbon management models (CO2MM). Figure 1.3 shows these relationships between the different parts. As noted, the emissions reporting process, when active, can be part of the implementation of a carbon management model, but is not dependent on such a model being in place to deliver results, as explained in the Contribution section of this work.

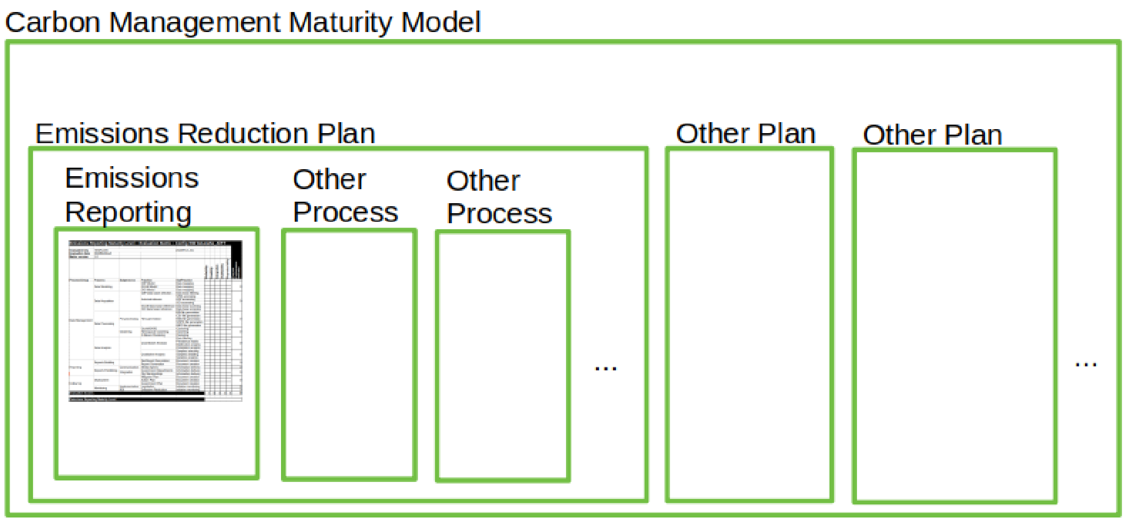

Figure 1.3: Example of emissions reporting as part of carbon management model. Other processes and plans are supposed to be part of CO2MM, but they are not shown due to not being part of the scope of this work.

The methodology of this work uses qualitative analysis to achieve the objectives of identifying candidate performance indicators. The methodology is strongly linked to the available data and to the proposed frame work. Thus, this work is limited to the application of the proposed methods, but is not intended to evaluate or review the application of the methods presented in the qualitative analysis outside the framework used in this work. This is mainly because the qualitative methods had to be adapted to the data already selected and processed, thus bypassing the interviews proposed with these methods to achieve the objectives set.

1.5 Contributions

This study contributes to the performance of emission reporting processes by analysing indicators and proposing a maturity model. After processing and analysing the emissions reporting data, three major contributions were identified, which are further explained in the next sections: the process of developing performance indicators, performance indicators based on emissions indicators and other indices, and a major performance indicator based on a maturity model developed to increase the efficiency of emissions reporting processes.

1.5.1 Performance Indicators Development Process

The performance indicators development process (PIDP) proposed in this work is an attempt to find candidate performance indicators based on emissions reporting data. These data are provided through a series of operations performed with CDP and other related databases. PIDP begins by selecting public data sources that can provide emissions-related data. The databases downloaded from these selected data sources are analysed and processed to produce data input files in comma-separated values (CSV) format. Thus, the steps that compose PIDP aim at preparing, processing and analysing the data present in the input file and generating by-products during the execution that are stored in output files.

The results obtained in each step of PIDP guide the next step to be taken in the process. Both quantitative and qualitative analyses are performed to select the most promising data to be examined and validated for use as a performance indicator. In the quantitative part of the process, the data is subjected to AI clustering methods to select the most promising configurations (experiment context). If the selected data are still promising after analysing the results, validation with other AI techniques will be applied to the selected samples, which will be part of the next phase - qualitative analysis. The constraints and rules applied during the PIDP are presented in more detail in the methodology proposed by this work in chapter 3.

1.5.2 Correlations between Emissions and Performance Indicators

One contribution of this work is to examine the relationships between established indices and indicators, such as the smart sities index (SCI), the human development index (HDI), the subnational human development index (SHDI), the gross domestic product (GDP), and the CDP database data.

The resulting correlations may serve as an indication for further data collection and future studies, depending on the complexity and magnitude of the results. Some of these correlations have already been presented in studies on emission reduction or mitigation and social and economic indicators. This work goes further by analysing the relationships between these indices and indicators with how cities answered emissions reporting questions. Other categorizations such as geographic region, country, population, urban area, or year of affiliation are also used to refine the analysis and final results of the correlations. Although the correlations identified cannot be used by decision makers on their own, they can still serve as a basis for future initiatives that will explore these correlations in more detail.

1.5.3 Emissions Reporting Maturity Model

The emission reporting maturity model (ERMM) is the main contribution of this work. Based on the results obtained in the implementation of the PIDP, a model has been developed to assess the maturity of a set of processes identified as part of emissions reporting initiatives. In this way, a maturity level for emissions reporting undertaken by a city is calculated based on an assessment of the capabilities of these processes. The emissions reporting maturity level (ERM-L) can be used to create a checklist for cities in areas relevant to emissions reporting.

When the ERM-L is used to compare the cities in a region or using another categorization, this maturity level can be seen as a key performance indicator of the city. Thus, the main goal of the ERMM is to make it possible to evaluate the city in terms of emissions reporting and to point to improvements that should be made so a city can perform better in emissions reduction challenges and issues.

1.6 Structure of this work

This work is organised into 5 chapters. Chapter 1 (Introduction), which this section is part of, describes a general view of the motivation and the proposed solution. In chapter 2 (Basic Concepts) are described some concepts that compose the necessary background to better understand some technical aspects of this work. In chapter 3 (Methodology) it is described how this work aims to achieve the proposed goals. In this chapter, it is depicted the steps necessary to produce the results through experiments and analysis, which are detailed in chapter 4 (Results) along with the configurations and hyperparameters used. Chapter 5 (Conclusion) consolidates the observations made along with the work and proposes future initiatives. In the appendixes are presented a *wordcloud* built upon the words written in this study and their frequencies (A), the set of questions selected from CDP database (B); a more detailed result of an evaluation obtained from the emissions reporting maturity level (C); the OECD countries list (D); an example log to illustrate the self-test execution

(E); and an example log to illustrate the preprocessing execution step (F).

Chapter 2

Basic Concepts

The main concepts related to performance indicators, maturity models, neural networks, grounded theory and case studies are presented in this chapter. These concepts implement the techniques and methodologies used to achieve the proposed goals and act as the building blocks of the methodology proposed in this work which is described in more detail in the next chapter chapter 3.

2.1 Performance Indicators

> "When you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfying kind" William Thomson – Lord Kelvin (1824 -1907).

In the scope of this work, emissions reporting is a crucial component in the decision-making process of emissions reduction policies. One way to help the cities to implement better policies is to use well-defined performance indicators that can measure the progress of the implementation and the results of the process that support these policies. The set of performance indicators used in this matter can compose a key performance indicator (KPI) to consolidate the view of the overall progress of adopting these policies. In this case, a KPI can be used alone or with other KPIs to implement a health check of one or more policies.

Therefore, a KPI development focus on identifying the best representatives among the performance indicators found by analysing the CDP forms database and additional data and how to use them to identify performance levels for the cities regarding emissions. In addition, developing a KPI requires a complete mapping of the data flow and business processes to make the KPI a reliable source of information.

As defined by BADAWY *et al.* [5], KPI allows gathering knowledge and exploring the best way to achieve organisational goals. Many researchers have provided different ideas for determining KPI's either manually, or semi-automatic, or automatic, depending on the application field. Based on the works of PARMENTER [6] and ECKERSON [7], the table 2.1 shows a way to identify KPI through some characteristics, summarized by BADAWY *et al.* [5].

Table 2.1: KPI characteristics proposed by PARMENTER [6] and ECKERSON [7].

Characteristic	**Description**
Sparse	The fewer KPIs, the better. A KPI is based on established PIs.
Drillable	Users can drill into the details of a KPI and then to its PIs.
Simple	Users understand the KPIs. They can indicate what action is required by staff.
Actionable	Users know how to affect outcomes. The KPIs expected results should be publishable.
Owned	KPIs have an owner. Are acted on by the CEO and senior management team.
Referenced	Users can view origins and context of KPIs.
Correlated	KPIs drive desired outcomes. They encourage appropriate action from other KPIs.
Balanced	KPIs consist of both financial and non-financial metrics.
Aligned	KPIs don't undermine each other.
Validated	Workers can't circumvent the KPIs or temper with them.
Regulated	Are measured frequently (e.g., 24/7, daily, or weekly).
Distributed	Are measures that tie responsibility down to a team.

Figure 2.1 shows a general schema for KPI development based on evaluating the available data in an evolving approach. In this view, a considered result indicator will provide the core value used by a performance indicator related to it, as this performance indicator adds a comparison dimension to the value provided. A key result indicator (KRI) can also be built upon available result indicators, as a KPI is built on one or more PIs. Both KPI and KRI can be correlated at executive level. For example, a KRI that summarizes all emissions reported by a city can provide the core (absolute) values to its correspondent KPI which indicates the percentage of rising or reduction of total emission for a certain period.

2.2 Capability Maturity Models

The capability maturity model (CMM) presents sets of recommended practices in some vital process areas to enhance software development and maintenance capabilities, as defined by PAULK *et al.* [8]. Thus, the CMM is based on knowledge acquired from software-process assessments and extensive feedback from both industry and government.

As pointed by MONTEIRO e MACIEL [9] and METTLER *et al.* [10], maturity can be considered a measure of a process related to its state or condition: defined, managed, measured, and controlled. CMM should be viewed as a set of "best

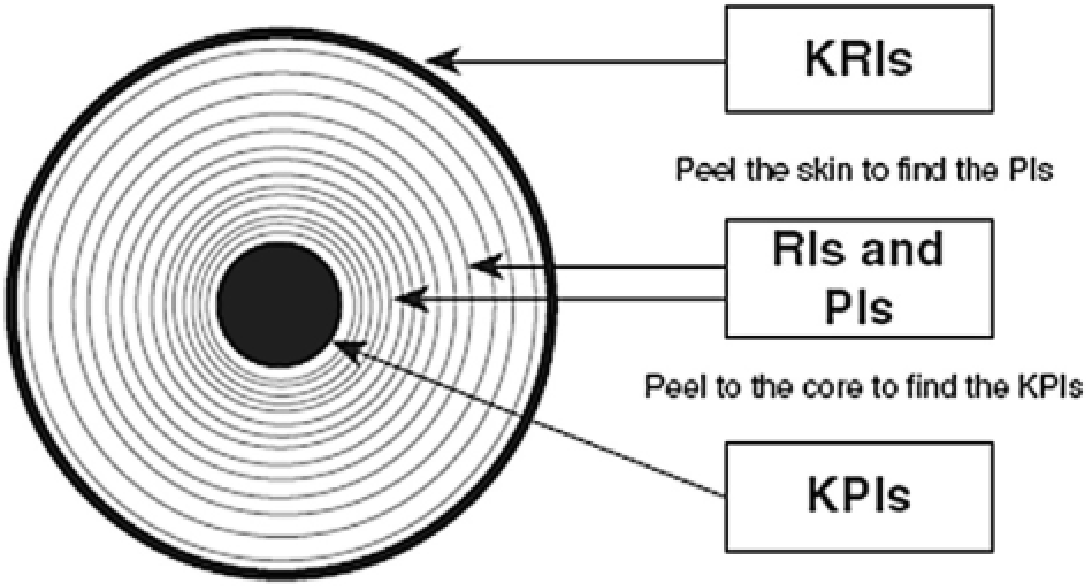

Figure 2.1: General schema of result and performance indicators. Source: PARMENTER [6].

practices" more than a straight list of steps to be implemented. Surveys, third party verification, and certification[11] can evaluate the level of adoption of these best practices.

According to PAULK *et al.* [12], the CMM is composed of five levels of maturity: initial, repeatable, defined, managed and optimized. These levels reflect processes, goals, and practices to be developed during the model implementation. However, the number of levels and what they represent can vary depending on the model to be implemented [9]. Figure 2.2 shows a general schema of CMM as defined by MONTEIRO e MACIEL [9].

Each level of CMM indicates a general process capability, and it has processes and goals to be achieved with the execution of these processes. Some standard features and, by them, some practices are identified by analysing these processes. The junction of implementation capability by a process relies on the same capability regarding the underlying practices. The evolution of processes and goals in complexity and completeness generates detailed sub-processes and sub-practices. These detailed components are checked to evaluate the adherence of the processes to the CMM level. Thus, to reach the next CMM level, these processes should pass the checklist based on defined capabilities.

The ISO/IEC [13] defines a capability maturity model for the software development process. Although this model has been updated in the years, the core components described above remain the same. The variations (new models derived from it) and improvements in the process are a direct result of the success of this

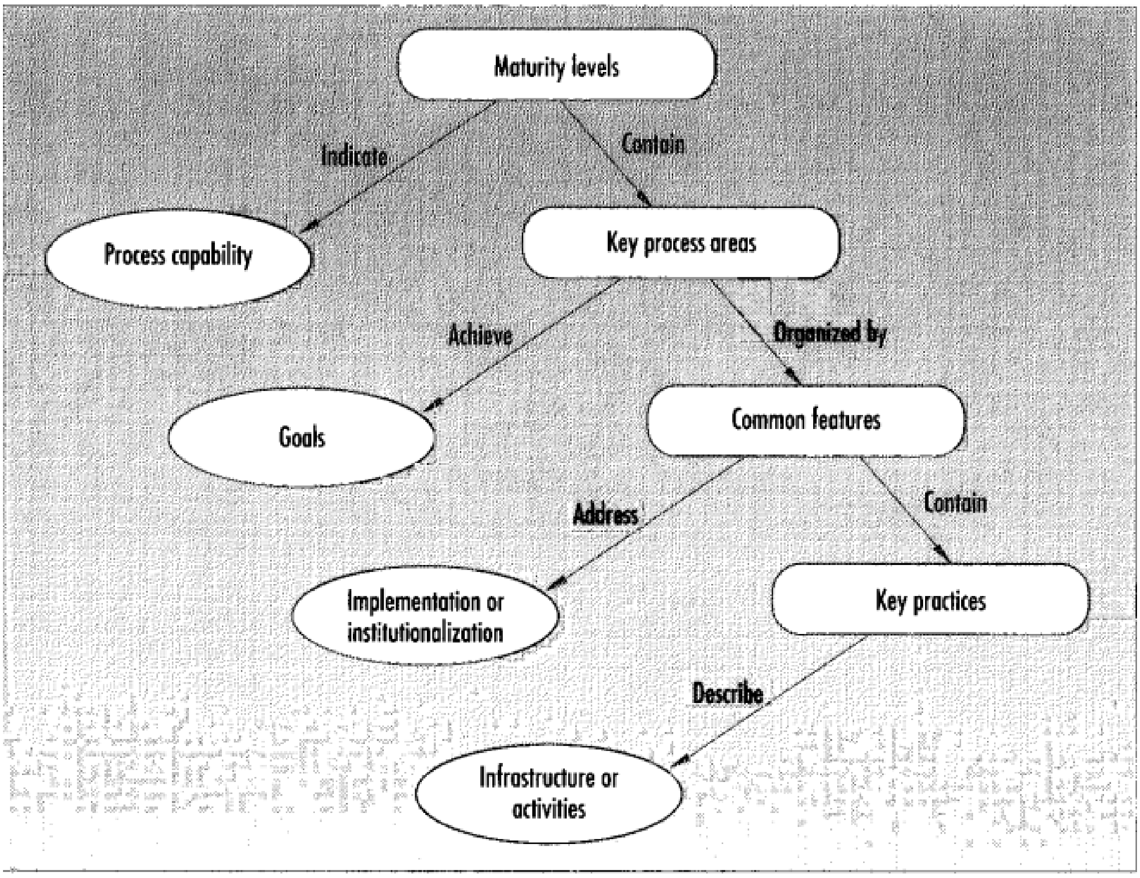

Figure 2.2: Capability maturity model general process view. Source: [8].

model in help measuring the efficiency of normalizing and standardizing development processes, independently or application area or previous expertise in dealing with maturity models.

The data management maturity model (DMMM)[1], built by the Capability Maturity Model Integration Institute (CMMII)[2], is a derivation of CMM that was considered in this work. It deals with data management challenges in any sector and organisation, which has been more necessary than ever when organisations have to process high volumes of unstructured data daily. To implement DMMM, the organisation has to asses the processes and interactions to/from data sources within the organisation and third parts. Figure 2.3 shows an example of a general view of the assessment and the impacted areas.

2.3 Weightless Neural Networks

Despite the fact that mainstream artificial neural networks (ANN) are heavily used in all almost all areas of knowledge, they are still implemented using models based on

[1]https://cmmiinstitute.com/data-managementmaturity
[2]https://cmmiinstitute.com/

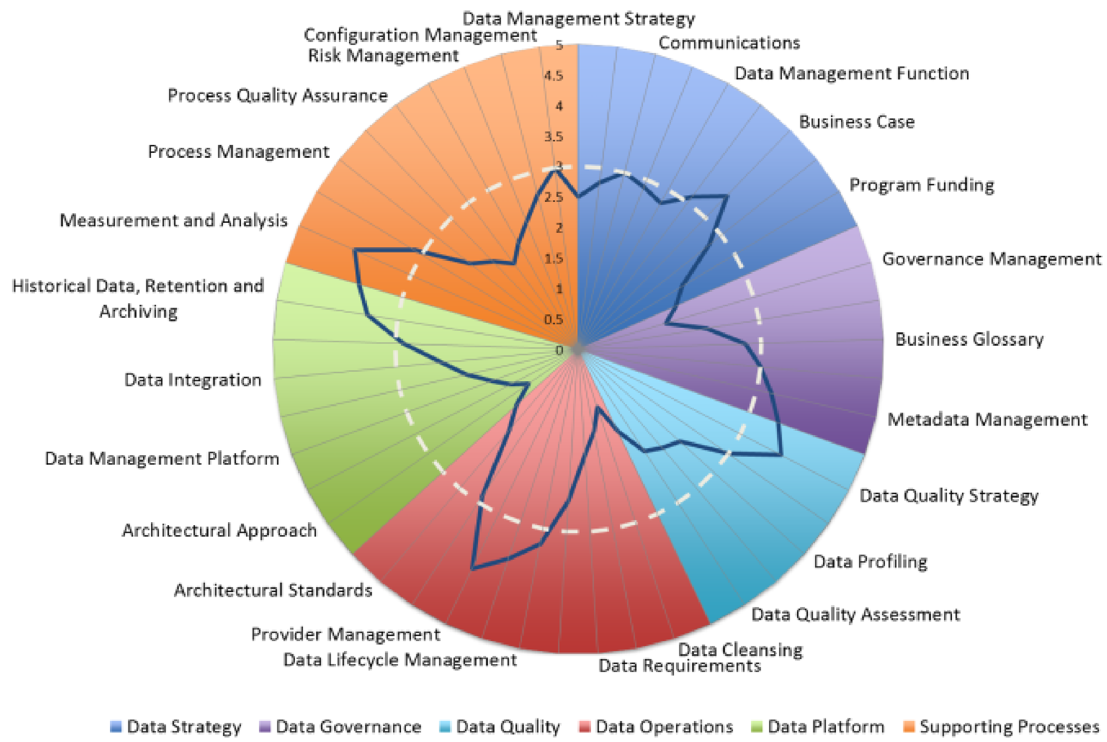

Figure 2.3: Data management maturity model assessment example, showing numerous areas in which DMM can be applied. Source: CMMII.

weighted-sum-and-threshold artificial neurons, as the pioneering **Threshold Logic Unit**[14], as as pointed by ALEKSANDER *et al.* [15].

The artificial neurons can be defined as structures that map the synaptic strength between an output transmitted by the neuron's axon and a post-synaptic neuron. This impulse creates a chain of signals based on pseudo-continuous numerical weights, terminating at neuron's *soma* - the central part of a neuron [15].

Conversely, weightless neural networks (WNN) models use another approach. The weight, which is responsible for representing the strength of the input signals, is replaced with the position (height) of the origin of the signals along an emulation of a neuron's dendritic tree [15]. This model is closer to the implementation of a "random access memory" (RAM) addresses decoding, which is a core concept in any implementation of the WNN model.

According to GREGORIO e GIORDANO [16], the WNN model was developed by BLEDSOE e BROWNING [17] as an n-tuple recognition method also known in the literature as "RAMnet" and based on an emulation of a neuron. Thus, a neuron system corresponds to a RAM with 2^n memory cells or address lines. These n-tuples use n bits samples from input data and are used to access memory cells to write or read neuron contents, emulating the learn/test phases of the system.

This work is based on implementations of WNNs as they are more likely to discover relations among the answers from the cities. These implementations look at how "strong" (clear) is a set of answers, considering the context in which the answering process is happening compared to another set of answers provided by other cities. It is possible to retrieve quality information in both scenarios: when the set of answers is firmly connected and, in contrast, when the answers seem to have no fit in the same set.

2.3.1 WiSARD

As summarized by ALEKSANDER *et al.* [15], the **Wi**lkie, **S**tonham & **A**leksander's **R**ecognition **D**evice (WiSARD)[18] is a system formed by various RAM-discriminators that was created as a pattern recognition device used in image processing, as pointed by GREGORIO e GIORDANO [16]. A RAM-discriminator consists of a set of X one-bit word RAMs with n inputs and a summing device (Σ). The system works by storing the result of processing an input of n bits in the training phase. Thus, the RAM input lines are connected to the input pattern by using a one-to-one pseudo-random mapping. The summing device enables this network of RAMs to exhibit, like other ANN models based on synaptic weights, generalisation and noise tolerance. As asserted by LUSQUINO FILHO *et al.* [19], during the classification, all discriminators are accessed, and they are assigned a score formed by the number of non-null positions accessed. The discriminator with the highest score will determine the class of the entry. Bleaching is a technique used when more than one discriminator is available for the input. The validation of another sample with n bits is based on checking the RAM-discriminator informed against the one generated. The Figure 2.4 shows an example of WiSARD.

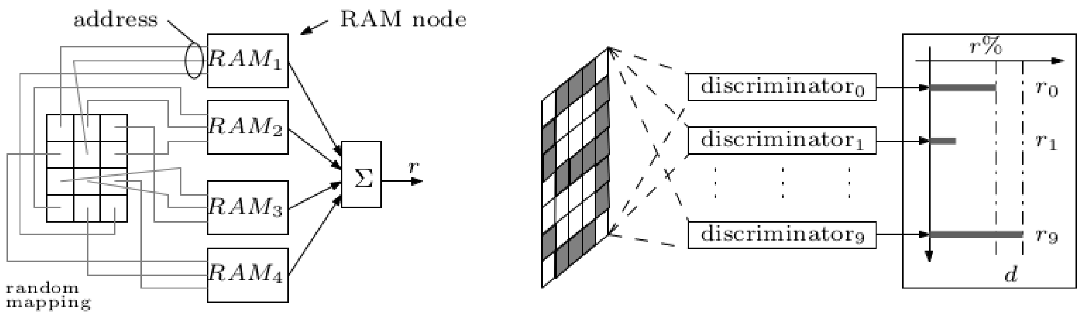

Figure 2.4: Example of a RAM-discriminator and of a WiSARD. Source: ALEKSANDER *et al.* [15].

As pointed by LUSQUINO FILHO *et al.* [19] and GREGORIO e GIORDANO [16], the main advantages of WiSARD, as any WNN model, is that it takes less time to be trained due to the direct relation between the RAM-based address mapping and the entry. Furthermore, the high level of generalisation is another characteristic of the WNN that can be seen in WiSARD implementation. Thus, the pseudo-random mapping, generalisation and noise tolerance make WiSARD an excellent alternative to pure analytical methods for processing the answers.

For the present work to achieve the proposed goals, the relevance of how the cities provide the answers to the questions related to emissions and the answers themselves should be the same.

2.3.2 ClusWiSARD

ClusWiSARD is a derivation of the original WiSARD model that allows the same class to have more than one discriminator, as pointed by LUSQUINO FILHO *et al.* [19] and CARDOSO *et al.* [20]. It is necessary to have more than one discriminator to implement the clustering capability of the method. The sub-profiles (clusters) of the same class that are not similar are learned in different places. It can inhibit the saturation of a discriminator's learning with the superposition of extremely heterogeneous patterns but that still belong to the same class. For this reason, ClusWiSARD could simultaneously handle supervised, semi-supervised and unsupervised learning. In this case, unlabeled data will be trained on the discriminator with the highest score. Figure 2.5 shows a ClusWiSARD multi-discriminator schema.

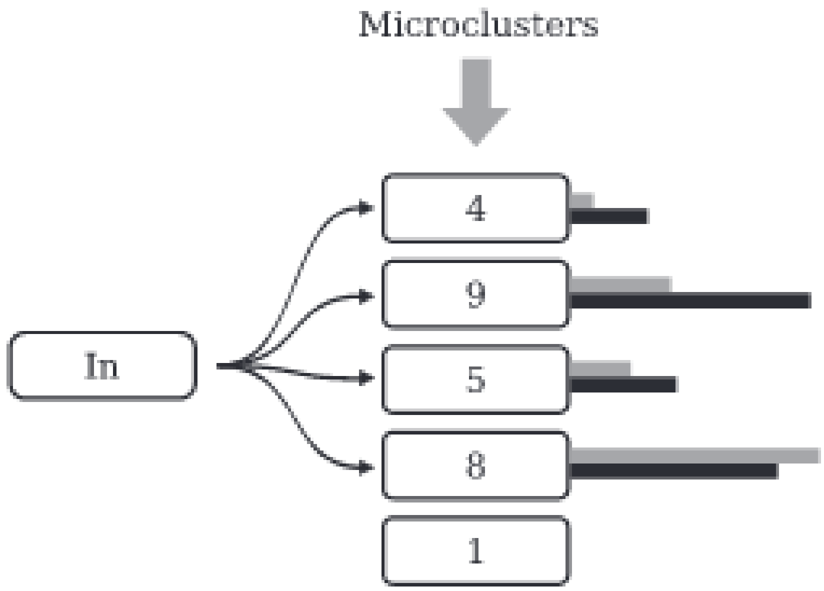

Figure 2.5: ClusWiSARD multidiscriminator schema. An example of a pattern In being presented to the microclusters. Each microcluster contains a number corresponding to the quantity of samples stored in it. The darker bar on the right represents the cluster's threshold, while the gray bar corresponds to the pattern activation. Observe that the discriminator containing 8 samples is the one that will learn the new observation. Source: CARDOSO *et al.* [20].

As pointed by LUSQUINO FILHO *et al.* [19], ClusWiSARD was created based on the plasticity-stability dilemma: a cluster can receive new information but still maintains the homogeneity of the data already associated with it. The idea that supports the model is the ability of a discriminator to understand that an example (sample) is not similar enough to the ones already associated with it. In this case, a new discriminator should be created to store the new sample. Still, in this model, the same example can be learned in more than one discriminator.

To start the clustering process, ClusWiSARD is initialised with only one discriminator of each class. While learning new examples, the method verifies if it is necessary to create a new discriminator based on the threshold value informed. Other hyperparameters, such as the discrimintator_limit, also interfere with how the method can discover clusters among the given samples. In the scope of this work, ClusWiSARD will be used to find clusters based on discriminators found and tuned by the threshold hyperparameter. Furthermore, as the discriminators can naturally vary between the execution of ClusWiSARD through the experiments, the computation of this variation is also registered to be used in the clustering process.

2.4 Other clustering methods

Other clustering methods are used in this work to complement the results of ClusWiSARD to help guide the selection of the most suitable experiments' configurations for established goals. Each additional clustering method deals with a different evaluation dimension.

2.4.1 Hierarchical Clustering

The essence of a cluster analysis process is to partition a set of N objects into C clusters such that objects within a cluster should be similar to each other and objects in different clusters should be dissimilar with each other, as summarized by K.SASIREKHA e P.BABY [21]. Clustering techniques can be used to associate a quantity to the available data, to extract a set of cluster prototypes for the compact representation of the data, targeting the generation of homogeneous subsets. As pointed by K.SASIREKHA e P.BABY [21], clustering can be seen as a mathematical tool that attempts to discover structures or specific patterns in a data set, where the objects inside each cluster show a specific degree of similarity. It can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to find them efficiently.

Cluster analysis is not an automatic task but an iterative process of knowledge discovery or interactive multi-objective optimization. In clustering, one of the most widely used algorithms is agglomerative algorithms, such as hierarchical clustering. The results are usually presented in a dendrogram - a graph representation of clusters construction. An example of this is shown in Figure 2.6.

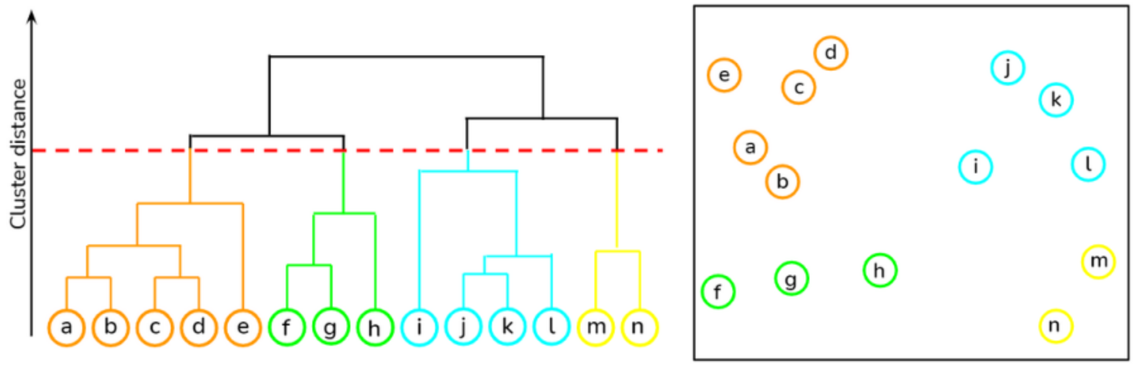

Figure 2.6: Example of dendogram produced by a hierarchical clustering method execution. Source: https://towardsdatascience.com/hierarchical-clustering-explained-e59b13846da8.

2.4.2 K-means

As pointed by LIKAS *et al.* [22], the simplest form of clustering is partitional clustering, in which specific clustering criteria are optimized to produce clusters (sub-sets) of objects based on the samples database given to the method. To achieve this result, the most widely used criterion to segregate data in clusters is the clustering error based on the evaluation of the squared distance from the corresponding cluster centre to the sample point being clustered.

According to LIKAS *et al.* [22], a popular clustering method that minimizes the clustering error is the k-means algorithm. Even though searching for an optimum value can lead to important drawbacks related to performance, the k-means algorithm is still a good choice in finding locally optimal solutions concerning clustering error. Still, it is a fast iterative algorithm that has been used in many clustering applications. K-means clustering method starts with the cluster centres initially placed at arbitrary positions.

K-means method moves the cluster centres at each step to minimize the clustering error. The main disadvantage of the method lies in its sensitivity to the initial positions of the cluster centres. One alternative to leverage the performance of k-means is to have another way than arbitrary to choose the initial position. The idea is to run the method several times with different initial positions, check for the near-optimal solutions obtained by the method, and compare them.

The variations in the centroids (k-means centre points) are an advantage in the present work. K-means is used to validate the clustering behaviour of hierarchical clustering and complement ClusWiSARD results in the overall clustering effort. The different approaches of the clustering methods used in this work aim to leverage the

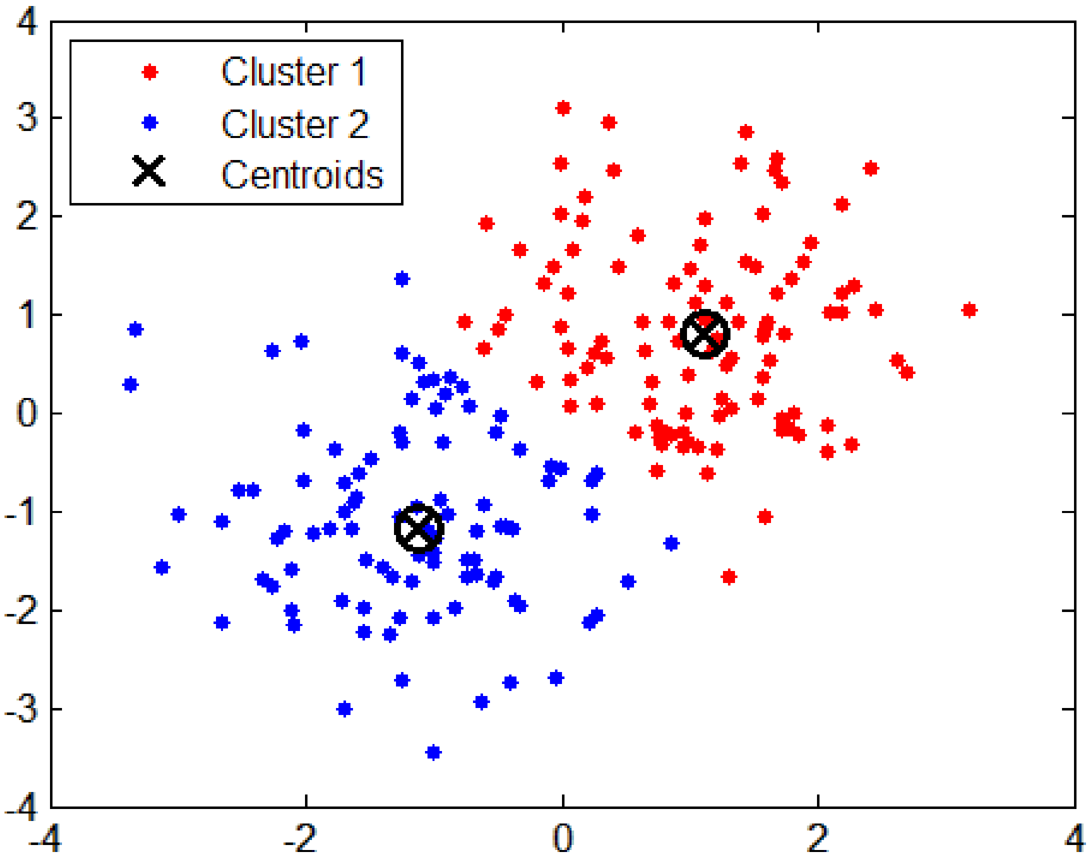

Figure 2.7: Simple example of k-means clustering of two clusters. Source: http://mines.humanoriented.com/classes/2010/fall/csci568/portfolio_exports/mvoget/cluster/cluster.html .

selection of samples. A simple example of k-means clustering based on centroids is shown in Figure 2.7.

2.4.3 DBSCAN

Density-based clustering methods are proposed to cluster spatial databases with noise [23]. So, Density-based spatial clustering of applications with noise (DBSCAN) is a clustering method that can discover clusters of arbitrary shape and also handles outliers effectively, which is its main advantage among other clustering methods. DBSCAN can achieve it by computing the distances from a given point to all other points in the database and then obtaining the clusters by finding the number of points within the specified distance from this given point.

The main advantage of DBSCAN over conventional index-based methods that construct a hierarchical structure over the data set to speed up the neighbour search operations is that DBSCAN scales better in terms of performance than those methods, mainly when applied to data sets of dimensionality above 20 features. Furthermore, although the performance of DBSCAN degrades due to unnecessary distance computations introduced by noise points, it is still robust to noise by pruning out noise points early and eliminating the unnecessary distance computations.

As also described by KUMAR e REDDY [23], the DBSCAN algorithm defines a cluster as a region of densely connected points separated by regions of non-dense points. This distinction of types of points is crucial to the method. DBSCAN algorithm takes two input hyperparameters called *eps* and *minpts* to support these characteristics. The Euclidean distance[24] is used to compute the distance from a given point in the region of a hypersphere of radius *eps* having at point p as its centre. Some essential definitions arise from that:

- *eps*: specifies the maximum distance neighbourhood for a given point.
- *eps*-neighborhood: for a point X , the *eps*-neighbourhood denotes the set of points whose distance from X is less than or equal to *eps*.
- the cardinality of *eps*-neighborhood defines the threshold density of X
- *eps*-connected: for a pair of points X and Y in the database, if the distance of X and Y is less than *eps*, then X and Y are *eps*-connected points.
- *minpts* is the minimum number of points required in the *eps*-neighbourhood of a point to form a cluster.

Hereafter, from the view of a DBSCAN method, every point in the database will fall into either core point or border point, which can be either a noise point or density connected point:

- core point: a point with threshold density greater than or equal to *minpts*.
- border point: a point with threshold density less than *minpts*.
- noise point: a point p is a noise point if the threshold density of p is less than *minpts* and all points in the *eps*-neighbourhood of p are border points.
- density-connected point: a border point with at least one core point in its *eps*-neighbourhood.

DBSCAN starts selecting an unvisited point from the unvisited points data set initially built with all points in the data set. If the number of points in its *eps*-neighborhood is less than *minpts*, it is marked as noise or outlier. Otherwise, it is considered as a dense point, and a new cluster is created. The next point is taken and added to the cluster by finding dense points for each point in the *eps*-neighborhood of the cluster. If there is no unvisited point to be added to a cluster, the new cluster is complete, and no points will be added to the cluster in subsequent

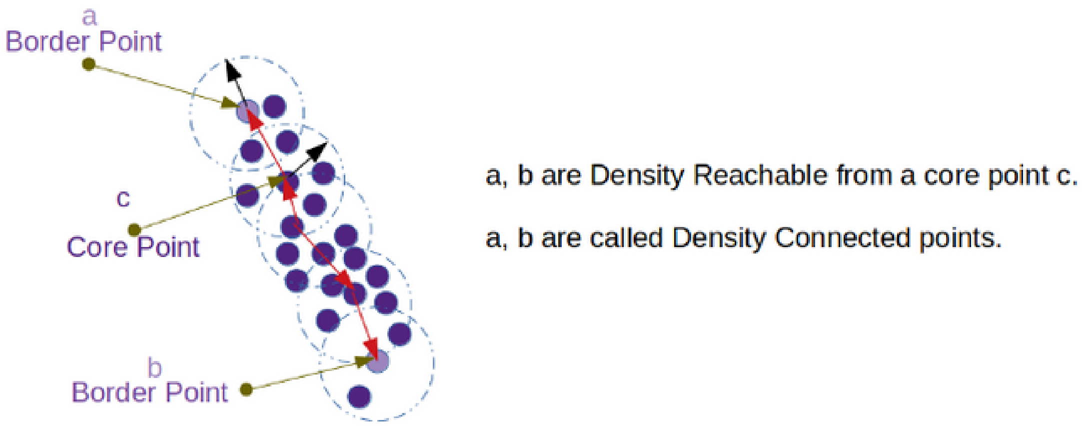

Figure 2.8: Simple example of DBSCAN clustering with core point and two borders points and five clusters. Source: https://towardsdatascience.com/dbscan-algorithm-complete-guide-and-application-with-python-scikit-learn-d690cbae4c5d .

iterations. Figure 2.8 shows an example with a core point, two borders points and five clusters in it.

Thus, the process is finished when all the points in the database are either assigned to some cluster or marked noise. The next cluster is found repeating the process to seek for *eps*-connected points. Every point in a cluster is *eps*-connected with at least one point in the same cluster to which it belongs and is not *eps*-connected with any other points in the remaining clusters. The number of *eps*-neighborhood operations performed is equal to the size of the data set.

2.5 Grounded Theory method

The main goal of a grounded theory study is to produce or discover a theory based on the (grounded) data provided using a process, action or interaction, as pointed by CRESWELL [25], CORBIN e STRAUSS [26] and DENZIN e LINCOLN [27]. The theory was first proposed by GLASER e STRAUSS [28] and was followed by other books and studies and spread to other areas of knowledge than Sociology.

Participants in the study would all have experienced the process, and the development of the theory helps explain the practices or provide a framework for further research. The key idea is that this theory-development does not come *off the shelf*, but instead is generated or *grounded* in data from participants who have experienced the process. Thus, grounded theory is a qualitative research design in which the inquirer generates a general explanation (a theory) of a process, action, or interaction shaped by the views of numerous participants [26].

The grounded theory can be when there is a process or a similar schema of execution in steps available that people can explain how it works and being part of its execution. There are two approaches established:

- systematic procedures: when an investigation search for some categorisation or similar organisation of the processes, actions or interactions [26];
- constructivist grounded theory: when data selection and process selection happen alternatively [29].

This last approach occurs more in social science fields. Regarding the codification of the theory being developed, it can be based on open coding, as the mapping of casual conditions lead to strategies for selecting data and then to the mapping of interview/analysis conditions that converge in a conditional matrix, or it can be based on selective coding when the proposition (hypothesis) is searching using a "storyline" based on the selected data.

2.6 Case Study method

As pointed by CRESWELL [25], case study research involves the study of an issue explored in one or more cases within a bounded system (i.e., a setting, a context). Although STAKE [30] states that case study research is not a methodology but a choice of what is to be studied (i.e., a case within a bounded system), others present it as a strategy of inquiry, a methodology, or a comprehensive research strategy [27][31][32]. This work follows the last approach.

A case study can have multiple sources of information with different levels of detailed answers, and it can be applied to many disciplines/fields. The case study research method can be used when there are clearly identifiable cases with boundaries. Nevertheless, the goal is to identify the meaning of a case: learned from the issue or from an unusual situation found.

During the investigation, the samples are selected and entered in a loop, considering how useful the information is, where the most promising case in terms of data is selected to be analysed. Regarding the data analysis, it can happen both overall issues presented by the data (holistic approach) or over a specific aspect of an issue (embedded approach).

Chapter 3

Methodology

The performance indicators development process (PIDP) described in the following sections is the core of the methodology of this work. The PIDP begins by exploring and processing the available data regarding emissions to look for candidates for performance indicators. The main products of PIDP are the performance indicators based on emissions and their relations to external indicators and rankings. The clustering techniques were used during the process to group cities with similar answers to CDP forms questions.

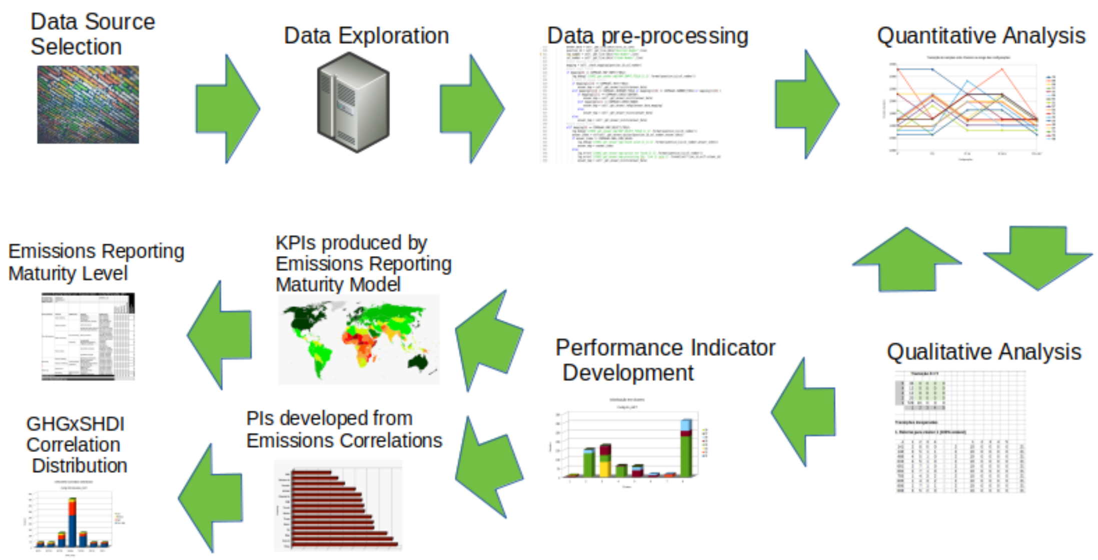

Figure 3.1: Performance indicators development process (PIDP) general view

The results of the clustering methods are used as the input to analyses to be done in the quantitative analysis phase. First, the samples with similar answers are analyzed to search for performance indicators among the features representing these answers. The qualitative phase uses the samples and features to confirm (or discard) performance indicators candidates.

The quantitative analysis identifies potential performance indicator candidates among the questions used to segregate the samples into groups. The confirmation of a candidate can be validated by looking into selected samples' data to check how stable is the candidate in separating the groups in the presence of other data. This step, executed using two experimental techniques, results in the qualitative analysis of the candidates to performance indicator. If the performance indicator found during the process can be used as the main source in a decision-making process, it is promoted to a key performance indicator. One example of this is shown in the emissions reporting maturity model (ERMM), a product of this process. The PIDP workflow overview is synthesized in Figure 3.1.

3.1 Data sources selection and processing

The first step in the performance indicators development process (PIDP) is to obtain reliable data about emissions among a set of cities representing as best as it can be the diversity found in the development level of cities along with the world. The intent to choose cities as a minimal comparison unit was based on the available literature and works regarding the crucial role of cities in the emissions reduction effort. The general view of this step is described in the Figure 3.2

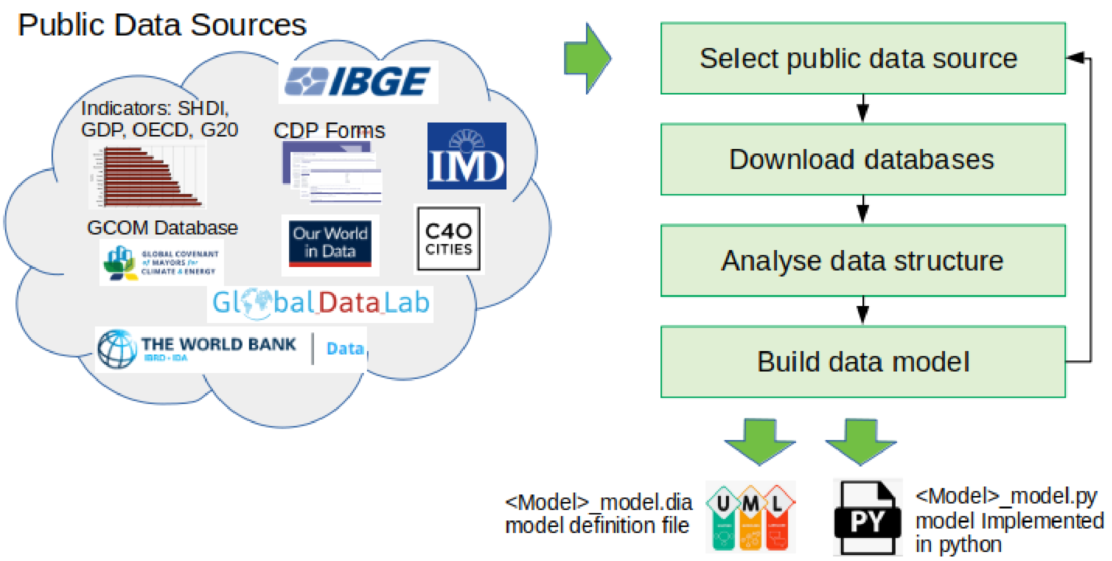

Figure 3.2: Data source selection general view.

According to the evaluation of the literature review focused on emissions reporting related topics [1][4][33][34] and data sources with emissions information from cities (GCoM[1], C40 cities[2], Our World In Data (OWID)[3], Global Data Lab[4], World

[1]https://www.globalcovenantofmayors.org/our-cities/
[2]https://www.c40.org/cities
[3]https://github.com/owid/owid-datasets/tree/master/datasets
[4]https://globaldatalab.org

Bank[5], Instituto Brasileiro de Geografia e Estatística (IBGE)[6]), the CDP disclosure database, obtained through a Kaggle contest in this topic, demonstrated to be the the most promising emissions reporting database.

CDP is an initiative that gets together 814 cities worldwide and significant private companies operating inside them to leverage policies and actions regarding reducing GHG emissions and their effects. CDP database is used as the primary source of information, even as being retrieved from a secondary data set. However, as pointed by LIEBCHEN e SHEPPERD [35], the pattern of using secondary data, typically data sets that have been made publicly available through various repositories, remains the norm.

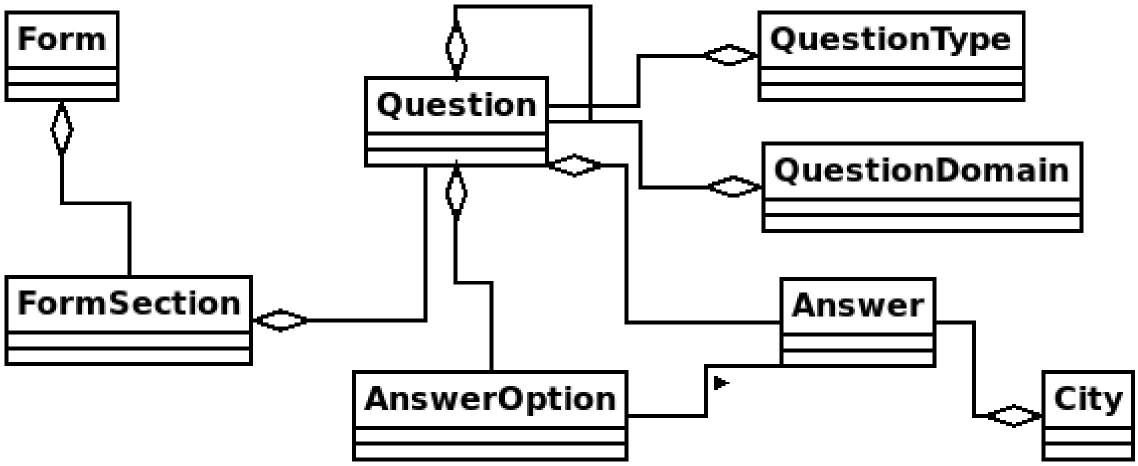

Figure 3.3: CDP UML model schema.

The process starts with accessing one data source from the public data source's list built manually through analysing previous works in the area. Then, using the means provided by the data source host, the databases available from the data source are downloaded and checked for consistency. After downloading and checking the databases, the data structure is mapped to build a model to process the data.

The model is built using unified modelling language (UML), proposed by BOOCH *et al.* [36], for simplification and standardisation. In Figure 3.3 is shown an example of a UML schema of the CDP model developed in this work. Implementing the model using a python file is also a product of this phase. It will be used in the data preprocessing phase and the clustering step of the quantitative analysis phase.

[5]https://data.worldbank.org
[6]https://cidades.ibge.gov.br/

3.2 Data exploration

After the databases of emissions-related data have been downloaded, mapped and initially analysed, the next step in the PIDP is to find the data units that can provide insights on candidates to performance indicators. This schema view of data exploration step is shown in Figure 3.4.

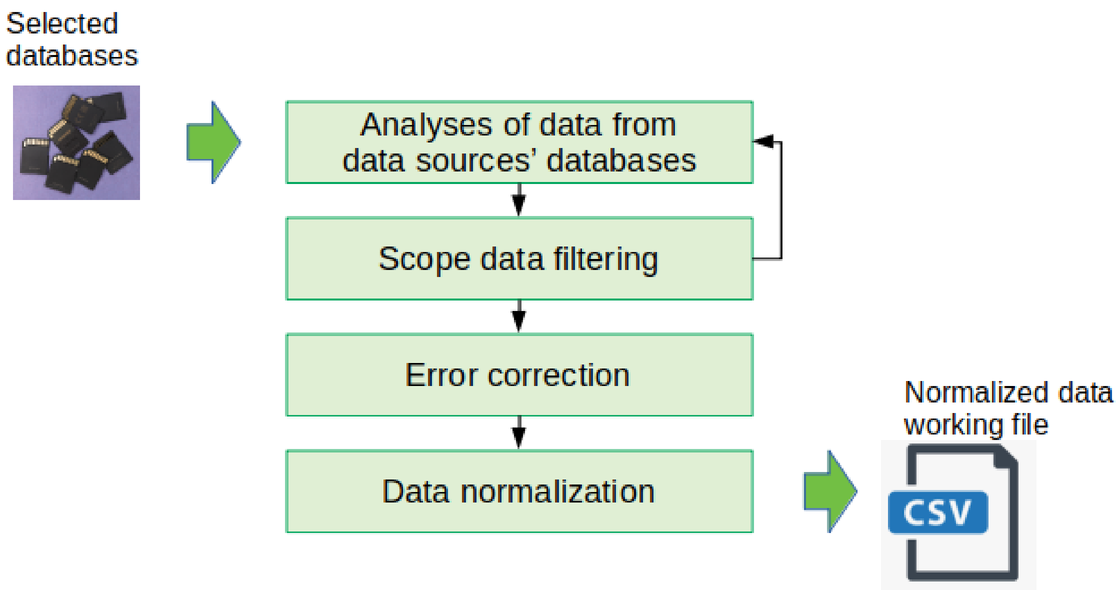

Figure 3.4: Data exploration general view.

The data to be explored should be analysed with the help of scope filtering. The objective of scope filtering is to filter the valuable data among the data available from the databases. As an example, even though the CDP disclosure database has much information about related areas like transportation, energy and employment level, the focus of this work is on emissions direct information, as presented in the sections described in table 3.1. In its third column (points) is represented the number of questions and sub-questions (tables) potentially used as the source of information. It indicates the potential of providing useful information on each forms section.

After obtaining a stable set of data from scope filtering, the process initiates attempts of error correction. In the context of data exploration, it is considered an error any inconsistency found in the analysed data: e.g. wrong data type, empty value in "selection" or "multi-selection" answer type, empty value in "not null" answer. If the error cannot be recovered using other data from the same record, the record is discarded.

Other sources of emissions information were used to complete the information extracted from the CDP database. The GCoM has a database with more than ten thousand cities in it. This database was used to provide additional data about total emissions per year (2019), the presence of preparation (planning) to face emissions hazards and mitigation targets. Other indicators like gross domestic product (GDP), sub-national human development index (SHDI) and smart cities index (SCI) were also used to support the indicators discovery and validation along the performance indicators development process.

The data normalisation occurs when these additional data are joined to CDP data to produce useful information. Finally, the external indicators as GDP and SHDI are examples of this. The result of the processing is saved in a working file to be used as input in the data preprocessing phase.

Table 3.1: CDP disclosure Sections. The column "Points" represents how many questions and sub-questions could be used to retrieve useful information.

Section	**Description**	**Points**
0:Introduction	General information	6
1:Governance and Data Management	Data management related information	32
4:City-wide Emissions	Emissions produced by the city, its companies and citizens	86
5:Emissions Reduction	Emissions reduction inventory reporting	110
7:Emissions Reduction by local government	Emissions reduction inventory of government scope	40

3.3 Data preprocessing

The data preprocessing step is responsible for preparing the available data to be correctly used by the clustering algorithms. The primary source of input data is the CDP forms database. However, after an initial inspection of the data in the forms, inconsistencies and errors were found that could jeopardise the clustering process.

According to SHEPPERD *et al.* [37], the better preprocessing strategy is that first, the problem data should be treated. Some cases of either conflicting feature values or implausible values should be discarded before data can be used. Therefore, it was necessary to build a support system to deal with these issues and leverage the quantitative and qualitative analysis steps. Figure 3.5 shows the data preprocessing schema with the generated output files. The generation of the files, their usages and which goals they address is detailed in the following sections.

3.3.1 Input data

A working file provides the input normalised data and additional data files in CSV format. Each one of the composed databases has a related model to support the processing of the underlying information. The models define the fields, the fields' types, and the operations realised over the data. For example, for fields of type

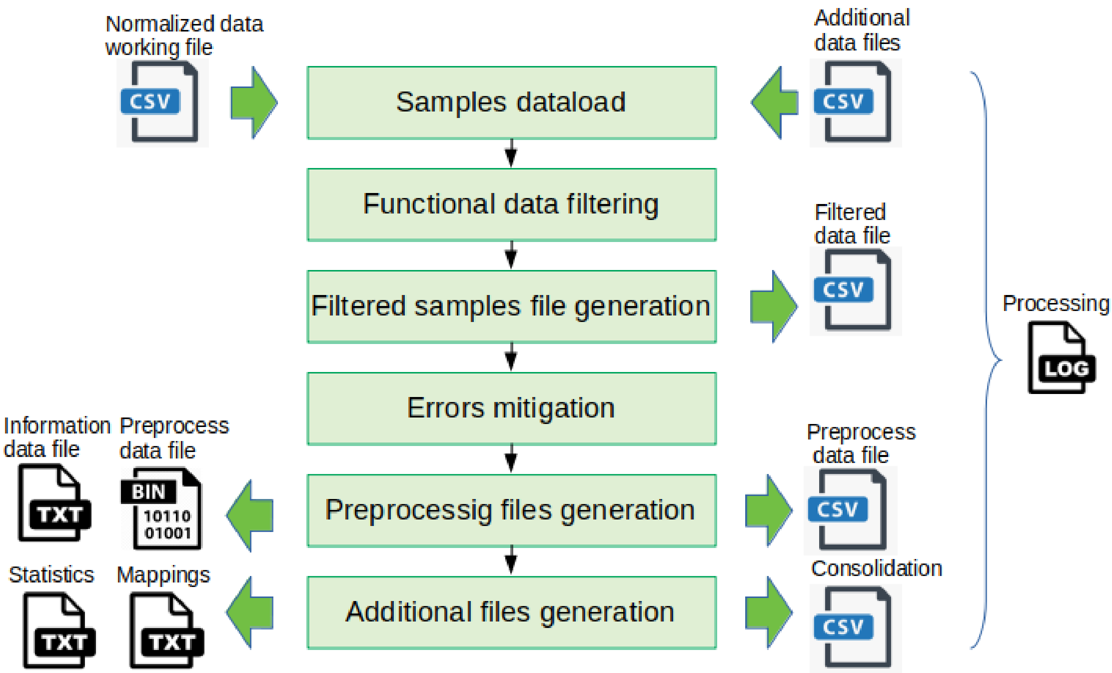

Figure 3.5: Data preprocessing view.

"single selection" and "multiselection", the models also checks the values provided. The same occurs with fields of "date" and "year" data domains: the ranges are defined to help validate the values informed.

For the CDP database, the classes that implement the concepts of fields, types, domains and operations are listed in table 3.2.

Table 3.2: CDP database classes mapping. The mappings show the relation between the CDP data model and data processing concepts.

Class	**Mapped concept**
Form	Data set
FormSection	Sub data set
Question	Field
QuestionType	Field type: null, not null, single-select, multi-select
QuestionDomain	Field type: DATE, YEAR, NUMBER, INTEGER, TEXT
Answer	Field values
AnswerOption	Field values options; case of single-select of multi-select types
City	Record

Samples dataload

The samples data-load is the first step inside data preprocessing, in which the normalised data working file is loaded along with the additional information present in additional data files. The CDP disclosures are organised in forms in which fields, represented by the questions, are subject to the model used to map the questions and constrain the answers. Each question and sub-question is represented by a line in the input data file. The table 3.3 details the cells presented in the line to process.

Table 3.3: Structure of a line in the input data file from CDP database in CSV format.

Cell	Description	Remarks
Questionnaire	Form identification	Filtered: Cities 2019
Year Reported to CDP	Base year for answers in CDP database	Filtered: 2019
Account Number	Unique identification for city in CDP database (Sample Id)	Unique Id
Organisation	City Name	Normalized to include State Name for clarification
Country CDP Region	CDP regions defined in table 4.6	
Parent Section	Group of sections	
Section	Group of answers	
Question Number	Question unique identification	
Question Name	Question unique name	
Column Number	Column unique identification inside question	Column 0 indicates direct answer
Column Name	Column name to identify tabled answer	
Row Number	Row unique identification inside question	Row 0 indicates direct answer
Row Name	Row name to identify tabled answer	
Response Answer	Answer value	
Comments		Used to clarify the answer
File Name		Complementary information about external file
Last update	Data time of last update of the record	

Functional data filtering

Functional data filtering occurs when filtering parameters are passed to preprocessing execution module to segregate only the information needed in the context of a preprocessing configuration and optimise the drill down during quantitative and qualitative analyses. The filtering engine can be used to select a set of questions and sub-questions, a set of samples (listed using a samples file) or all samples in which a field type is present. The filtering engine permits include (I:) or exclude (E:) operators, acting to compose the filtering rules to be applied over the data. Some filtering examples are shown in table 3.4.

The byproduct of this step is to generate a "filtered data file", which is an exact copy of the filtered samples. This file can be used to accelerate the drill-down process of the investigation of quantitative and qualitative analyses.

Table 3.4: Filtering examples used during the experimental phase of this work.

Filtering scope	**Filtering options**
Question "0" and its sub-questions	I:Question Number=0*
All questions "0,1,4,5,7" and their sub-questions	I:Question Number=0*,1*,4*,5*,7*
All questions "0,1" and their sub-questions, excluding fields with type YN	E:#FieldType=YN;I:Question Number=0*,1*
All questions reported by cities in samples.txt file	"I:#SampleId=@samples.txt

3.3.2 Errors mitigation

The errors mitigation step in the data preprocessing phase of PIDP is responsible for leveraging defected data in the subsequent phases. The techniques applied in errors correction depend on the nature of the error: e.g. domain-value matching, invalid value type, and values out-of-ranges.

One problem identified in the CDP forms data entry is the text representation for questions with single and multi-selection options. To solve this issue, the CDP model implements unique codes and associate them with the available options. However, in some samples, the text informed does not match the text of any option available to that question. In this situation, the use of techniques to correct the string representation based on the number of changes, like Damerau-Levenshtein distance, as presented by ZHAO e SAHNI [38]. The table 3.5 shows some examples found in CDP database preprocessing.

Table 3.5: Examples of application of Damerau-Levenshtein distance to answers correction. The text differences are presented in bold.

CDP Id	**City Name**	**Question**	**Original Answer**	**Correct Answer**
1093	Atlanta	1.1a:Please select any commitments to climate adaptation and/or mitigation your city has signed and attach evidence	Individual city **c**ommitment	Individual city **C**ommitment
1184	Austin	1.13:What tools does your city/department use to analyse its environmental related data? Select all that apply.	Visualization / Analysis Software - Tableau **;** Qlik etc	Visualization / Analysis Software - Tableau **,** Qlik**,** etc
1184	Austin	5.0a:Please provide details of your total city-wide base year emissions reduction (absolute) target.	Larger **–** covers the whole city and adjoining areas	Larger **-** covers the whole city and adjoining areas

The invalid value type occurs when a numeric value is expected, and a "null" or other value type is provided to an answer instead. The mitigation, in this case, is to convert the text representation to the best number representation, when it is possible, and to set the answer to "zero value" and "not answered" when it is not. The values out-of-range issue is mitigated using statistics tools (e.g.variance) to check and correct scaling errors. To achieve this, the model used to support the processing of the database holds the expected min and max (range) values that are supposed to happen and a "mark" in the question in the model indicating that it has to be range-checked.

3.3.3 Preprocessing files generation

The main products of the preprocessing phase are the CSV format output files with the processed data. The internal representation of the files differs based on the target clustering engine that will be used. The textual representation will be used as the input file for Hierarchical, K-means and DBSCAN clustering methods. The file with binary representation, on the other hand, will be used as input for ClusWiSARD.

Table 3.6: Conversion mechanisms used to transform the input CSV file into processed textual representation file, also in CSV format. The result of the data processing is saved in the correspondent textual processed data file.

Field type	Conversion mechanism
TABLE	Conversion of each field's value inside table (multi-select) using the correspondent field type rule described here. Each value is separated by ":" in a list representing each row of the table of multi-select fields.
SELECT	Conversion to numeric value represented the text informed in the field value. If the field value is not found among the predefined answering options, error mitigation techniques try to choose the best available option. If it is not possible, the conversion uses "0" to represent the "not found" answering option.
TEXT	Conversion to "0" if field is empty or "1" on the contrary.
NUMBER	Conversion to the log of the field's value to try to narrow to a common scale to be used with the other questions. The log value is then converted to text representation.
INTEGER	Conversion straight to text representation as-is.
YEAR	Conversion of difference from base year value to text.
DATE	Conversion to ISO data format (ISO 8061) without hyphenation.

During processing, each field generates an output in text format, based on the rules defined by the model. The field type and specification define the field's value conversion mechanism. The table 3.6 details the conversion mechanisms used. Due to optimisation, during the preprocessing of the numeric fields (NUMBER, INTEGER and YEAR types), some statistics are collected to be used in the next binarisation step. Another important measure to be taken is the number of bits to represent an answer. To obtain the best minimum value for the number of bits, the data preprocessing uses the number of options for an answer, the single and multi-selection fields, and the number of digits in the answer for the numeric fields. The number obtained is registered as the **binary slot size** in the information data file generated by the process. It is used to define the same number of bits applied to all answers.

Table 3.7: Conversion mechanisms used to transform the processed textual representation file in CSV format into processed binary representation file, also in CSV format. The result of the data processing is saved in the correspondent binary representation processed data file.

Field type	**Conversion mechanism**
TABLE	Conversion is applied to each value in the list of preprocessed text values according to the rules described here. The final binary value is a superposition ("OR" operation) of each bit of each binary value of each field in the table.
SELECT	Conversion to bits-value representation using two technique: bit-mapping and thermometers. The bit-mapping is used for multi-select fields and maps the numeric value of the option chosen as an index to the position in the bit string, which is filled by s-bits "1". The number of s-bits is a result from the *slot size* divided by the total number of options for the answer. The thermometer technique adds "1"s bits as to fill the string (from left to right) until reached the position of the option. This technique is used when processing single-select fields.
TEXT	Conversion to full "0"s or "1"s depending on the processed value.
NUMBER	Conversion to the thermometer representation of the processed value. The mechanism is the same as the one applied to SELECT field, but using min and max values computed along the answers to establish the *scale* of the thermometer. Thus, the number of bits used is the result of slot size times the field value minus min value divided by the max value minus min value.
INTEGER	Conversion to thermometer representation as described in NUMBER field.
YEAR	Conversion to thermometer representation as described in NUMBER field.
DATE	Conversion to thermometer representation as described in NUMBER field.

Hereafter, the binarisation step occurs when the binary representation file is generated based on another conversion mechanism applied over the processed textual representation file. It is necessary to guarantee that different clustering methods use the same clustering information in different formats. During the generation of the binary file, the text values are converted into binary (0's and 1's) representation, based on the field type and specification. The table 3.7 details the conversion mechanisms used to generate the binary representation of the processed textual data. To avoid misinterpretation of which file should be used as input to ClusWiSARD, the file with binary representation content receives a *.bin* extension.

An example depicting the data processing of the city of Rio de Janeiro's data extracted from CDP forms, present in the CDP forms database file, is shown in Figure 3.6.

3.3.4 Additional files generation

During preprocessing, some additional files are generated as important byproducts. The consolidation data file holds information about the processed numeric values:

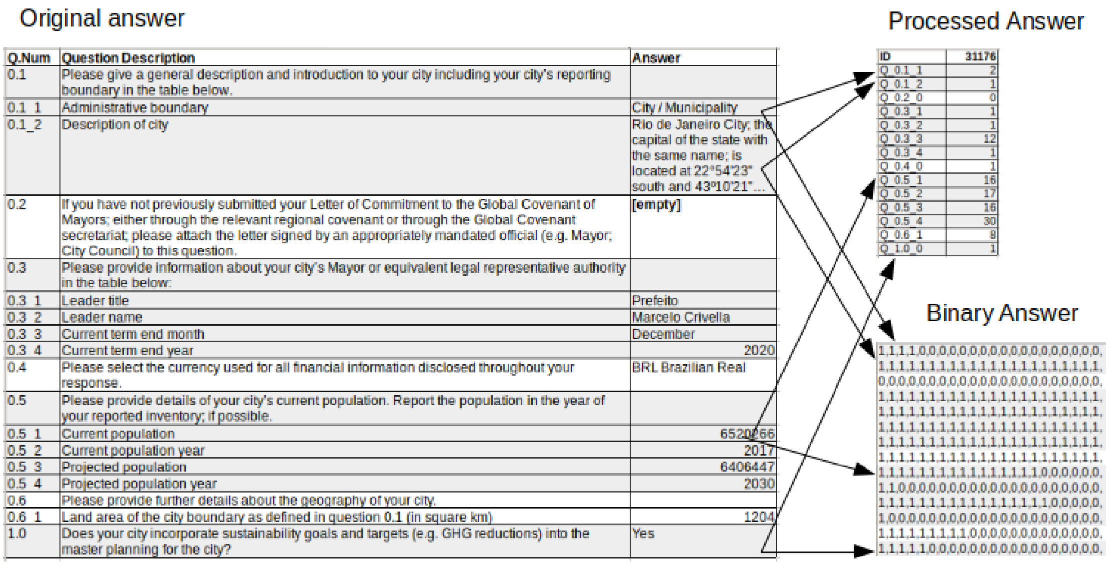

Figure 3.6: Preprocessing generation using Rio de Janeiro (31176) data sample example.

min, max, mean, and frequency of not empty answers. These values make it possible to check the distribution behaviour observed using the thermometer technique to process binary data output. An example of summary data file is shown in Figure 3.7.

Qnum	Frequency	Max Value	Min Value	Mean Value
0.1_1	814	2030	0	12
0.1_2	814	792	0	1
0.2_0	595	0	0	0
0.3_1	814	803	0	1
0.3_2	814	805	0	1
0.3_3	814	5663	0	12
0.3_4	814	763	0	1
0.4_0	814	775	0	1
0.5_1	814	10084	0	22
0.5_2	814	16170	0	219
0.5_3	814	8156	0	21
0.5_4	814	18173	0	700
0.6_1	814	5370	0	21
1.0_0	814	1238	0	5

Figure 3.7: Preprocessing consolidation file output example.

The questions filtered in the preprocessing are put in a list with question_id and question_name. At the end of the preprocessing step, a text file is saved with the number and description of the question. It is used to facilitate the qualitative analysis based on the applicability of the questions. For questions which underlying field of type multi-select, the options are also listed to help calibrate the quantitative analysis as needed. The questions of configuration 0a1a4a5a are listed in the appendix of this work.

The processing statistics output file holds quantitative and qualitative information about the processing of questions for each city. The table 3.8 shows the details of the obtained statistics.

Table 3.8: Preprocessing statistics details collected during the execution of the experiments.

Statistic	Detail
SampleId	City unique identification
TABLE	Count of fields of type "table" with answer
SELECT	Count of multi-select field with answer
TEXT	Count of fields of type "text" with answer
NUMBER	Count of fields of type "number" with answer
INTEGER	Count of fields of type "integer" with answer
YEAR	Count of fields of type "year" with answer
DATE	Count of fields of type "date" with answer
CC_R	Count of characters in the answer
CC_C	Count of characters in the comments
WC_R	Count of words in the answer
WC_C	Count of words in the comments
WU_R	Count of unique words in the answer
WU_C	Count of unique words in the comments
WD_R	Count of dictionary words in the answer
WD_C	Count of dictionary words in the comments

One extraction to exemplify the statistics obtained during preprocessing is shown in Figure 3.8. The differences between the cities are established, even being part of the same south-east region. For example, despite having the best GDP, São Paulo is far from being the best information provider.

SampleId	CityName	StateCode	SELECT	TEXT	NUMBER	YEAR	DATE	CC_R	CC_C	WC_R	WC_R	WU_R	WU_C
31176	Rio de Janeiro	RJ	240	554	177	3	12	25417	93	4700	18	3600	16
31184	São Paulo	SP	59	41	18	3	4	7242	0	1411	0	856	0
35848	Belo Horizonte	MG	116	115	41	3	16	23138	0	4380	0	2926	0
35897	Campinas/SP	SP	184	205	173	3	4	23362	0	4429	0	3164	0
45219	Aparecida/SP	SP	16	20	9	3	2	302	61	58	12	48	12
50383	Sorocaba	SP	48	249	22	3	4	5259	0	951	0	757	0
50387	Guarulhos	SP	14	18	9	3	2	1032	0	213	0	141	0
50392	Vitória	ES	88	210	71	3	4	8493	0	1376	0	1152	0
50396	Santos	SP	54	50	9	3	2	22007	0	4038	0	2510	0
54623	Betim	MG	87	95	45	3	4	7639	0	1461	0	1104	0

Figure 3.8: Preprocessing statistics extraction example listing ten cities in Brazil.

3.3.5 Processing Logs

The logging information generated during the preprocessing step is used to check the overall process and validate the information's reliability. The indication of errors in the logs interrupts the (next) output generation step, forcing checking what is causing it. For example, the CDP database has some errors in field mapping, domain values, and rules applied to form filling. These errors were marked or fixed to continue the form processing. Another use for general logging is to set up the proper provisioning for machine power and memory needed in preprocessing and the following steps. A logging extraction of the preprocessing phase is listed in the appendix of this work.

3.4 Quantitative Analysis

Quantitative analysis is based on the results from the clustering methods applied to the CDP disclosures database. The clustering results are treated and viewed as an alternative to purely statistical ones. However, the main goal is to search for similarities and answers that indicate different approaches implemented by the cities that are grouped in the same cluster. The nuances of the clustering process, the comparative data generated, and validation techniques are shown in the following sections.

Figures 3.9 and 3.10 show a general view of this step of the process.

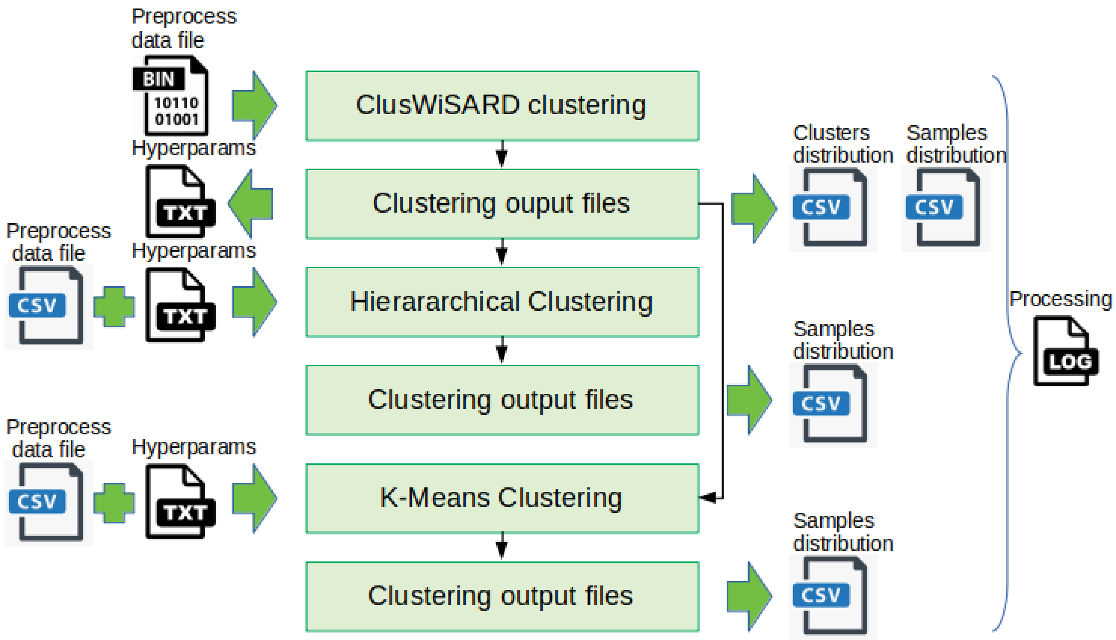

Figure 3.9: Quantitative analysis schema view

3.4.1 Using ClusWiSARD

ClusWiSARD is the primary clustering mechanism used to group the samples (cities) with similar or related answers. The other clustering mechanisms were used to validate and narrow the quantitative analysis process in pursuing performance indicators based on the answers. The ClusWiSARD results can be seen as "pictures" taken from the binary correspondence of the CDP forms' responses and additional data. The similarities in the answers are registered and used to group the samples into clusters.

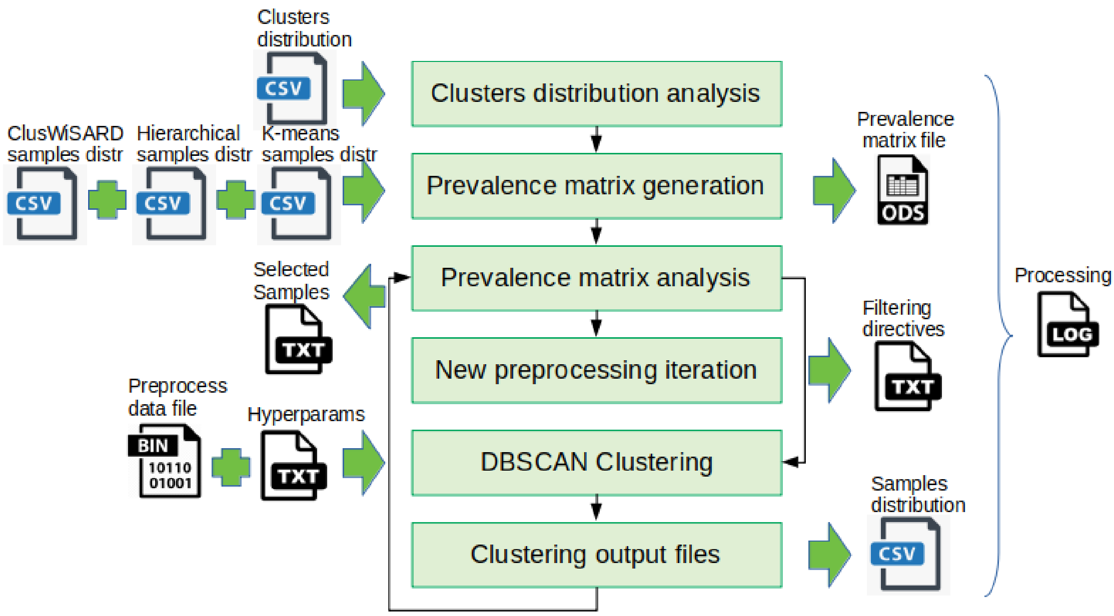

Figure 3.10: Quantitative analysis schema view (continuation)

Even though cities provide different answers to the same question as expected, the differences in the pictures are more subtle, and it is not easy to extract a pattern among them. The main advantage of ClusWiSARD application to this work is the generalisation capacity of the method. Even though different, the answers set tends to generate patterns in the responses used to identify candidates to performance indicators. The Figure 3.11 shows an example of "pictures" processed by ClusWiSARD.

This step is the generation of two CSV files: a clusters distribution and the distribution of a sample. The clusters distribution file holds information about how the clusters were consolidated. The number of clusters in which a sample can be grouped is registered along with the cluster chosen as the best choice (group) for this sample. This measures how stable is the clustering process given the hyperparameters informed to the ClusWiSARD algorithm. The Figure 3.12 shows a clusters and a samples distributions examples.

The ClusWiSARD is executed in "discover" mode when the hyperparameters threshold and discriminator_limit are set to "auto" value. In this case, a text file with the best values for these two hyperparameters and the other hyperparameters used to execute the method is saved from being used in another process of hierarchical and k-means clustering methods.

3.4.2 Using other clustering methods

Some other clustering methods were used in this work to validate the results of ClusWiSARD regarding the processing of the available data, as these other methods

Figure 3.11: Preprocessing binary representation extraction example as "pictures".

use different approaches to identify the groups of data (clusters). Among differences in implementation, ClusWiSARD uses a non-deterministic approach to group similar "pictures" from the data, as the other methods use a deterministic one. The Hierarchical Clustering method uses the aggregation (agglomerative or bottom-up approach) of similar features of the samples to compose the groups. The maximum number of groups (clusters) is pre-defined, and it is set as the same as the one used in ClusWiSARD. In addition to it, K-means uses another approach that uses the Euclidean distance between the field values to k centroids (or geometric centres) to group the samples. Both methods have the results compared to decide the use (or not) of the DBSCAN method to complete the analysis.

The execution of hierarchical and k-mean clustering methods uses the processed CSV format data file as input and the hyperparameters used in the ClusWiSARD method. The byproducts of this step are the files with the distributions of the samples that will be used to compose the prevalence matrix in the further step of the process.

The result of this step is the generation of the prevalence matrix file, as shown in Figure 3.13. The prevalence matrix analysis leads to four possible paths:

- a new preprocessing iteration with a new configuration: when the analysis of the prevalence matrix indicates a "dead-end", a new filtering configuration is established and the preprocessing phase is executed again.

Clusters Distribution

	Number of clusters								ClusterSum
ClusterId	1	2	3	4	5	6	7	8	ClusterSum
1	0	0	327	44	9	5	0	1	386
2	0	2	19	8	17	10	45	0	101
3	1	2	59	15	13	3	0	0	93
4	0	0	9	12	12	12	1	1	47
5	1	0	2	15	6	7	2	2	35
6	0	2	2	16	10	3	2	0	35
7	0	0	13	4	4	0	1	0	22
8	0	3	3	2	1	3	2	0	14
9	0	2	7	3	0	0	0	0	12
10	0	0	0	5	2	3	1	1	12
11	0	4	4	0	2	0	0	0	10
12	0	0	5	4	0	0	0	0	9
13	1	2	2	2	1	0	0	0	8
14	0	0	2	2	1	3	0	0	8
15	0	1	1	1	3	1	0	0	7
16	0	1	0	0	2	2	0	0	5
17	0	2	0	2	0	0	0	0	4
18	0	1	1	0	1	0	0	0	3
19	0	0	0	0	0	2	0	0	2
20	0	0	0	0	0	0	0	1	1
									814

Samples Distribution

SampleId	SortedClusterId
1093	2
1184	8
1499	9
2028	5
2185	13
2430	3
3203	9
3417	8
3422	17
3429	17
8242	10
10495	4
10894	16
11315	15
13067	2
14088	18
14344	11
14874	8
16581	1
19233	1
20113	18
...	...

Figure 3.12: Clusters and samples distributions examples.

- a new preprocessing iteration with new filters and the selected samples generated by the prevalence matrix analysis: this path is based on the "drill-down" of the analysis of the set of samples that can hold information to lead to identify performance indicators candidates, but still have to be verified through another iteration of the quantitative analysis so far.

- a selected samples set that will be analyzed in the qualitative analysis step: this path occurs when the analysis of the prevalence matrix indicates that the configuration being evaluated has a good chance to produce a performance indicator candidate. In this case, a selected samples file is generated to be used in the qualitative analysis phase.

- a DBSCAN clustering execution with the same hyperparameters as the previous methods: this happens when the analysis of the relations between ClusWiSARD, hierarchical clustering and k-Means clustering did not point to a clear result. In this case, a DBSCAN method is executed to help identifying a more clear path reducing the plausible "noise" in the samples analyzed so far in this step.

Thus, the validation step in the process is based on comparing the behaviour of ClusWiSARD with the other clustering methods. The samples in each cluster should be compared to their corresponding in the other clustering methods, generating a prevalence matrix P.

This matrix is built using the formula 3.1.

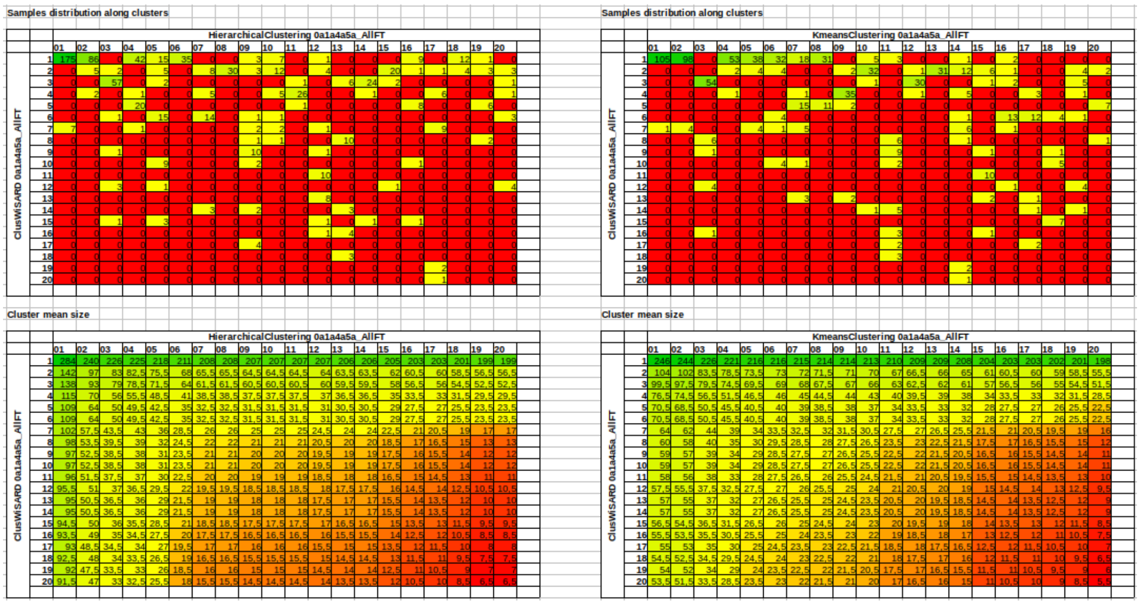

Figure 3.13: Prevalence matrix example.

$$P_{i,j} = (2 * b_{i,j})/(c_i + v_j) \tag{3.1}$$

where $P_{i,j}$ is the prevalence index in i, j; $i \in C$; $j \in V$; C is the ClusWiSARD clusters set; V is the validation clustering mechanism (hierarchical clustering, k-means or DBSCAN) clusters set; $b_{i,j}$ is the number of samples present both in C_i and V_j; C_i is a subset of C with samples in cluster i; V_j is a subset of V with samples in cluster j; c_i is the number of samples in C_i; v_j is the number of samples in V_j.

The samples present in the clusters with a higher prevalence index in P are then selected, and another experiment is executed using the same hyperparameters as the original experiment. This analysis and verification processes repeat as long as the mean global prevalence index (mpi) is greater or equal to the prevalence index of the last experiment.

The mpi is built using the formula 3.2.

$$mpi = \sum_{\substack{1<i<m \\ 1<j<n}} P_{i,j}/i * j \tag{3.2}$$

where mpi is the mean prevalence index; $P_{i,j}$ is the prevalence index in i, j ; m is the number of clusters identified by ClusWiSARD; n is the number of clusters identified by the validation clustering mechanism (hierarchical clustering, k-means or DBSCAN).

3.5 Qualitative Analysis

Qualitative analysis is based on the visualisation of specific values from selected questions. The main objective is to compare the responses of different cities present in the same cluster, indicating a convergent approach over the data. Another strategy is to compare different responses from cities in different clusters, indicating a divergent approach in this case. The following sections detail the techniques involved in the step of the process.

3.5.1 Using Grounded Theory

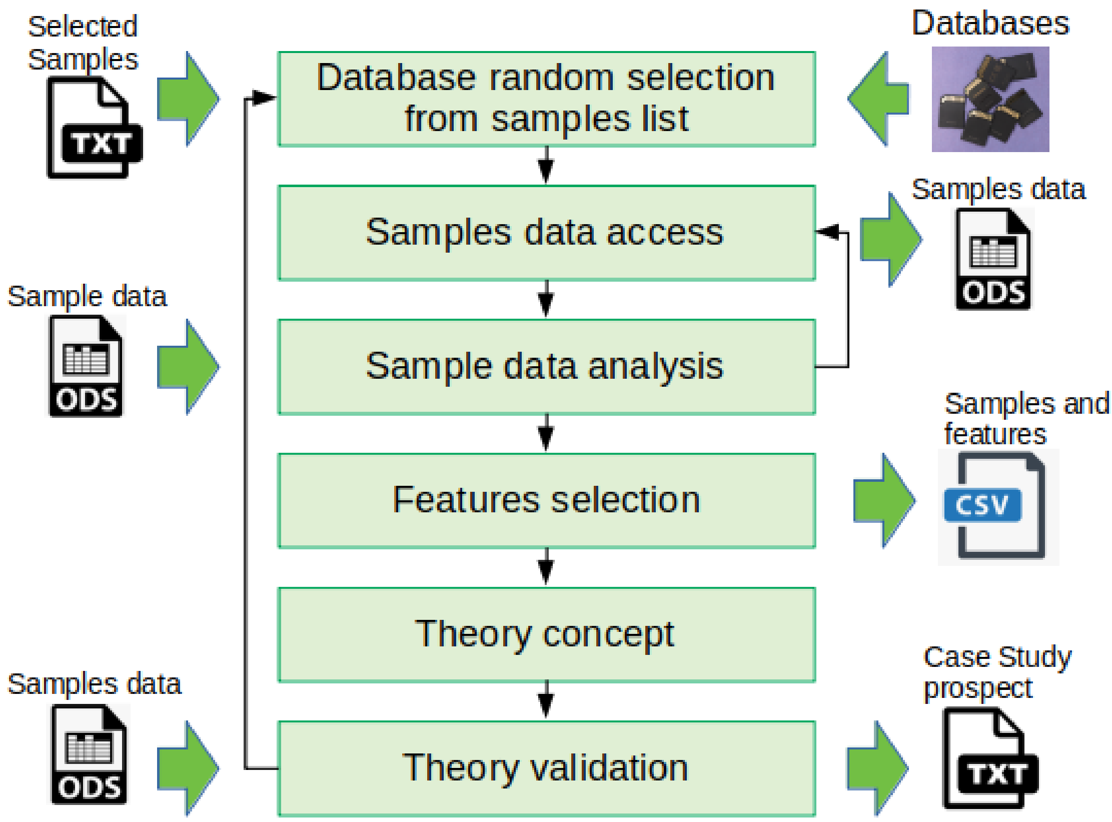

Figure 3.14: Qualitative Analysis: Grounded Theory application general view.

This work uses an adaptation over what is proposed by grounded theory to facilitate analysing responses to the same question from different cities. Here, a random set of cities is gathered from the cities in a cluster, which was selected as the most promising from the quantitative analysis step. A set of questions of interest is chosen, and a matrix is built to allow the visual analysis. According to the results, another round is performed to select other cities for comparison. This procedure is performed when the results are inconclusive or show a possible tendency in the answering process. This tendency composes a theory of answering that should be confirmed or denied in the further steps. The next set of cities can be used to confirm the tendency, reinforce the theory, or deny it, resetting the process to look for another theory based on other tendencies. The analysis continues until more than 50% of the cities are selected. Hereafter, if the tendency pattern remains, the process involves finding samples that represent exceptions to the theory (or candidate rule), using the subsequent (case study) approach.

3.5.2 Using Case Study

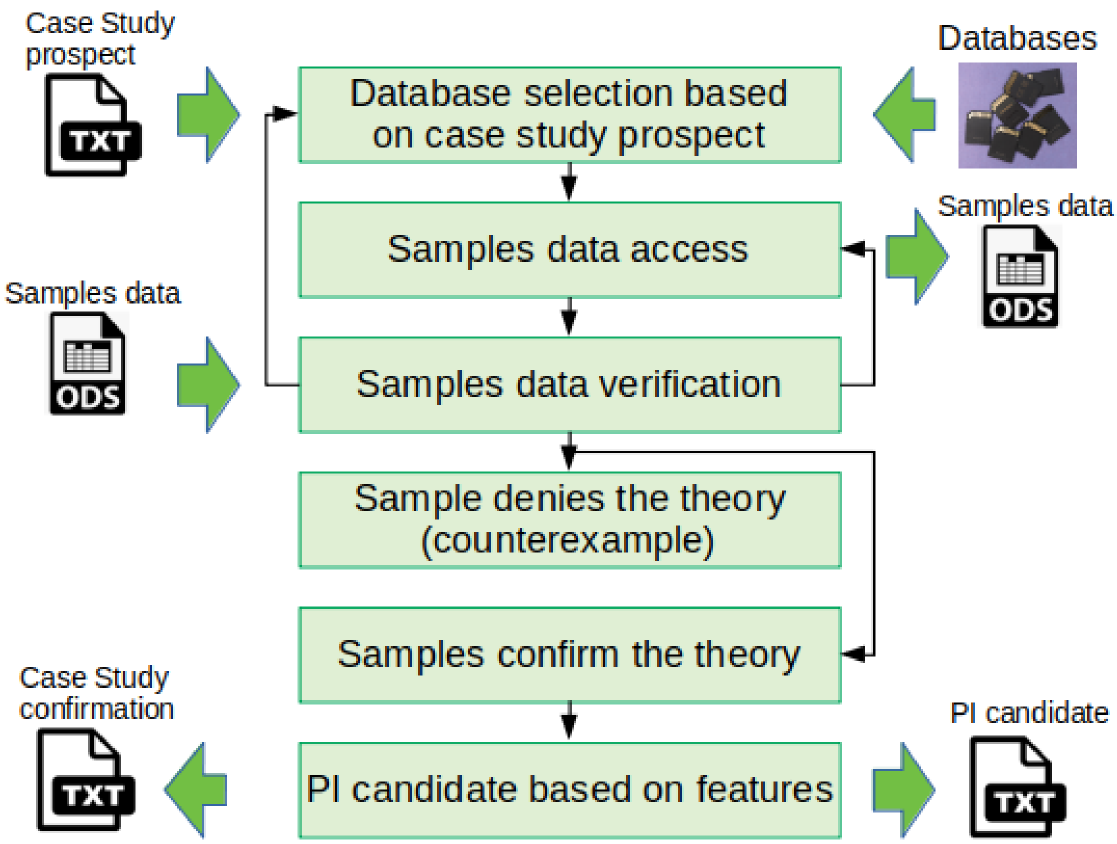

Figure 3.15: Qualitative Analysis: Case Study application general view.

The case study approach uses all available questions from a single city selected from any other than the selected cluster being analysed to check for inconsistencies that confirm or discredit a tendency found through the grounded theory approach. Suppose it is impossible to proceed with the confirmation or denial of the theory. In that case, another city is selected from another cluster, and the analysis continues until all clusters have been visited at least once.

3.6 Emissions Reporting Maturity Model

The emissions reporting maturity model (ERMM) stands for a methodology to select, process, classify and deliver evaluations of emissions-related processes based on the information presented in emissions reports.

The main goal is to help the cities better structure the information related to emissions so it can effectively and efficiently be used in the policy-making processes regarding emission reduction. The ERMM describes a six-level evolutionary path that aims to leverage the quality of the information provided to stakeholders in city administration along the time, as described in table 3.9.

Table 3.9: Emissions Reporting Maturity Levels summary. The processing contexts from which the ERM-L can be obtained is described for each level.

ERM Level	Contexts from which ERM level is extracted
0:Unavailable	Emissions information is not available to be used whatsoever.
1:Initial	Emissions information is available, but it is not part of any government plan or it cannot be validated or trusted.
2:Managed	Emissions information has been used to help plan the emissions policies, but cannot be independently validated.
3:Established	Emissions information is part of the government's general plan for the city and it can be validated using in-house (local) methods.
4:Predictable	Emissions information is part of general and departments plans for the city and it can validated both internally and externally, by an independent auditing contractor.
5:Optimized	Emissions information selecting, processing and using processes are integrated in cities both short-term general and departments plans and long-term policies (laws) and the actions resulted from these can be verified independently and has their effectiveness measured. The policies derived from emissions information can also be replicated to other cities

The model is loosely based on ISO/IEC TS 33061:2021 (Process assessment model for software life cycle processes) and uses some techniques proposed by the data management maturity model (DMMM), built by Capability Maturity Model Integration Institute (CMMII). The general view of the model is shown in Figure 3.16.

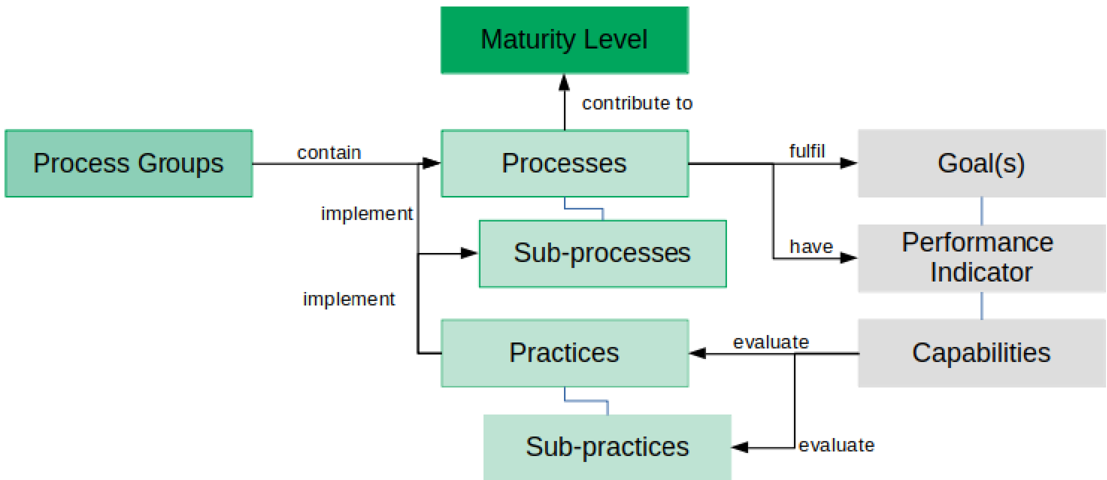

Figure 3.16: Emissions reporting maturity model general view.

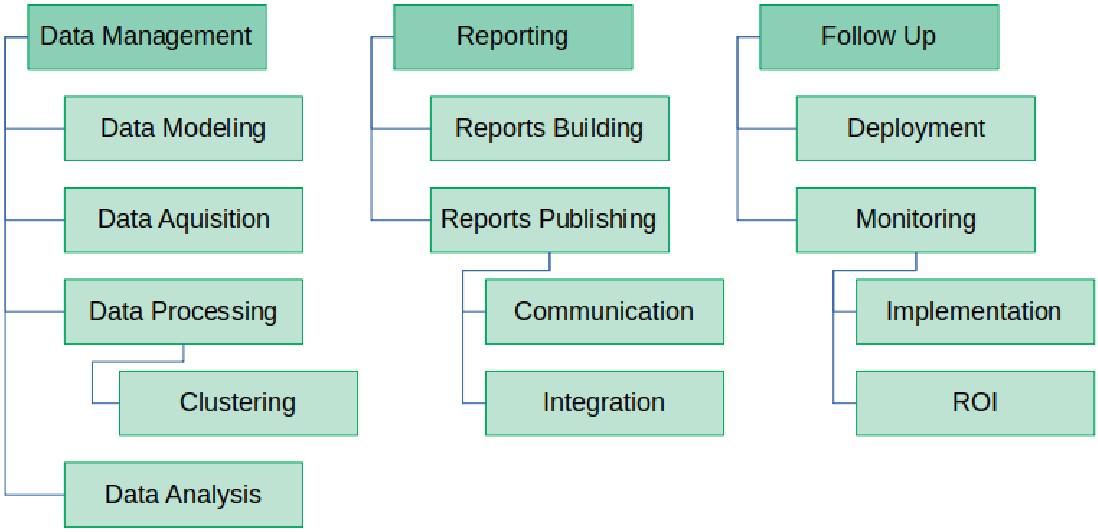

Figure 3.17: Emissions reporting maturity model processes view.

Table 3.10: Capabilities analyzed in the context of execution of ERMM over the cities in CDP database. Each capability is used to evaluate tune the evaluation of the processes present in ERMM.

Capability	**Application Example**
Reliability	how reliable is the information being processed. Automated practices of data acquisition is an example of reliability level 5.
Usability	how useful is the information to the processes. Information acquired from the available data that may compose a performance indicator is considered most useful, receiving value 5.
Integration	how integrated to other sources and targets is the information. If the information is provided or validated with the help of an external source, this capability is at level 4, at least. If the channel is automated, for example, this raises to 5.
Auditability	how auditable is the process and the information is treated by it. The auditability will be as good as the auditing process and resources. For example, if the information is audited by a known auditing provider with good auditing results, the level would be set to 5.
Reproducibility	how much a process can be reproduced in other scenarios and contexts. As an example, if the process cannot be reproduced by another city because of lack of documentation or resources, the level of capability would be set to 0. On the contrary, if conditions of reproducibility are fulfilled like human and economic resources available associated with full knowledge of the process and its pitfalls, the level would be set to 5, in this case.

Each level of ERMM defines some goals and processes to address these goals. The processes and sub-processes are composed of practices and sub-practices to describe the mechanisms better to achieve established goals. The practices and sub-practices are evaluated by five capabilities listed in table 3.10. The evaluation scale from 0 (incapable) to 5 (most capable). In the way of structuring the emissions reporting information, these capabilities are verified through the levels. The processes are also organised in areas to indicate different contexts or areas of application. The proposed ERMM process schema is shown in Figure 3.17 and the processes, sub-processes and capabilities exercised are shown in 3.19.

The capability evaluation matrix (CEM) is used to associate the capabilities to the practices and sub-practices, defining the weights and levels of applying. The Figure 3.18 shows a template of it.

Emissions Reporting Maturity Level – Evaluation Matrix – Config WW 0a1a4a5a_AllFT										
Evaluated City	TEMPLATE			[SAMPLE_ID]						
Evaluation Date	DD/MM/AAAA									
Matrix version	1.0									
ProcessGroup	**Process**	**Subprocess**	**Practice**	**SubPractice**	**Reliability**	**Usability**	**Integration**	**Auditability**	**Reproducibility**	**Process Performance Indicator**
Data Management	Data Modeling		CDP Model	Data mapping						
			GCoM Model	Data mapping						0
			C40 Model	Data mapping						
	Data Acquisition		CDP data base selection	Data base filtering						
			External indexes	SHDI accessing						
				GDP accessing						
				SCI accessing						0
			GCoM data base reference	Data base accesing						
			C40 data base reference	Data base accesing						
	Data Processing	Pre-processing	Files generation	BIN file generation						
				CSV file generation						
				RAW file generation						0
				STATS file generation						
				INFO file generation						
		Clustering	ClusWiSARD	Clustering						
			Hierarquical Clustering	Clustering						0
			K-Means Clustering	Clustering						
	Data Analysis		Quantitative Analysis	Data filtering						
				Prevalence matrix						
				Distribution analysis						0
				Correlation analysis						
			Qualitative Analysis	Samples selecting						
				Samples detailing						0
				Samples analysis						
Reporting	Reports Building		Dashboard Generation	Document creation						0
			Report Generation	Document creation						
	Reports Publishing	Communication	Media Agents	Information delivery						0
		Integration	Government Departments	Information delivery						0
			City Memberships	Information delivery						
Follow Up	Deployment		Mitigaton Plan	Document creation						
			Action Plan	Document creation						0
			Government Plan	Document creation						
	Monitoring	Implementation	Legislation	Initiative monitoring						0
		ROI	Emissions Reduction	Initiative monitoring						0
Evaluation Scores					0	0	0	0	0	**0**
Emissions Reporting Maturity Level										

Figure 3.18: Emissions reporting maturity model evaluation matrix.

The data management context (DMC) of ERMM can be defined through a usage example using the available data in this work. CDP, GCoM and C40 databases can be used to compose a virtual data model used in ERMM. The practices of modelling this virtual database are subjects to the capabilities to define a performance indicator related to the owner process. The same occurs with other practices of other processes, summing the values of the performance indicators up in the execution chain. On the other side, the level of the fulfilment of the goal(s) associated with a process is also added to the performance indicator of the process.

The reporting group of processes is composed of construction and publication-related steps. Reports building concentrates on the generation of the document at

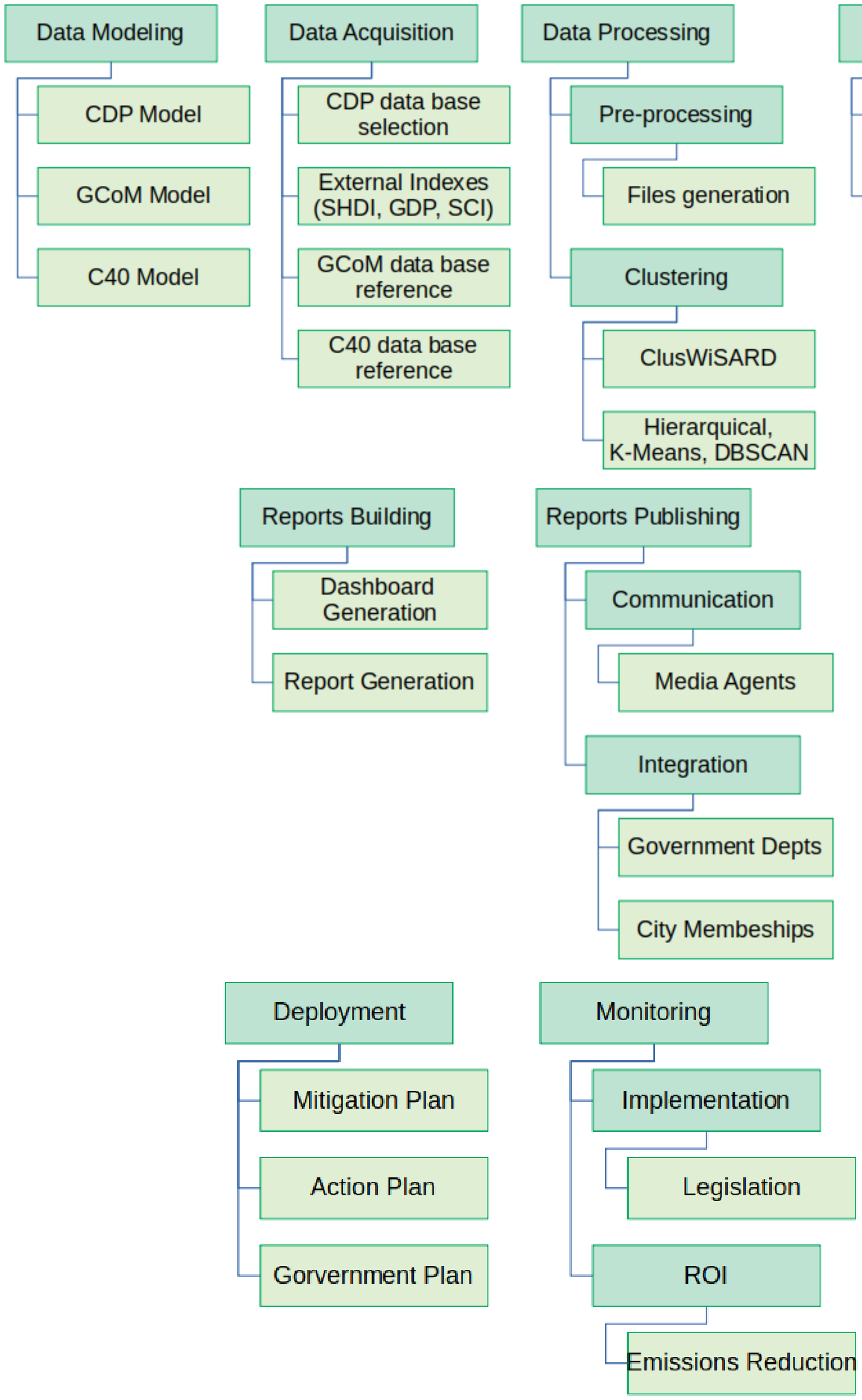

Figure 3.19: Emissions reporting maturity model processes and capabilities detail view

the high administration level, using the publication of dashboards with the summary of emissions reports, and at the administrative/technical level, in which projects for future laws or mayor's decrees are built.

Another area of interest present in the emissions reduction initiatives viewed so far is the follow-up emissions policies. This group of processes deals with the ability to receive and process feedback information regarding the emissions reports applied data, which are part of the city's plans to mitigate emissions impacts, to action over presented challenges and to be a permanent part of the city's general management plans.

3.7 Performance Indicators for Emissions Reporting

Performance indicators for emissions reporting (PIER) can be obtained from both the correlations of CDP data and external (additional) indicators, and the maturity model processes developed to address emissions reporting challenges. The lack and the low-quality information provided in CDP, GCoM, C40 and additional databases is the seed of the initiative of proposing a set of methods, encapsulated in a maturity model, to address these issues. The following sections detail both sets of performance indicators.

3.7.1 Performance Indicators based on Emissions Correlations

Performance indicators (PI) based on emissions correlations indicators (ECI) can be developed analysing the correlations between available emissions information in CDP forms database and other databases, indicators and indexes normally used to measure emissions and their effects.

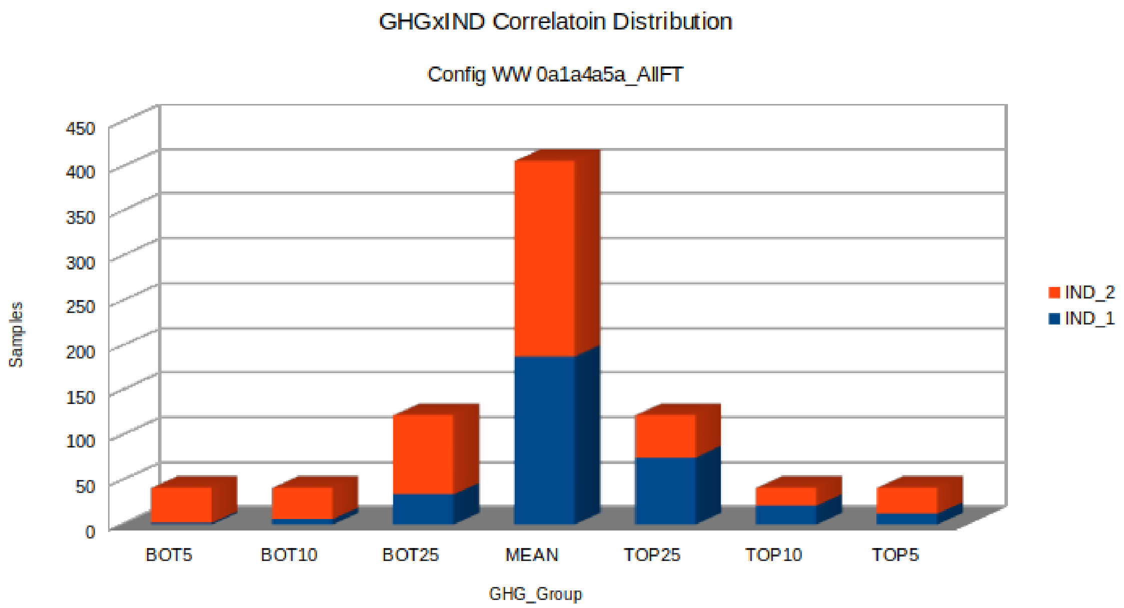

Figure 3.20: Example of correlation that can be used as a performance indicator.

Although, some correlations can also be established looking at a country being part of which multilateral organisation. The main difference between this work and others is that the focal point is the cities listed in the CDP database, forming a minimal baseline to promote the necessary analysis to support those correlations. An example of a correlation distribution based on CDP data is shown in Figure 3.20.

3.7.2 Performance Indicators based on Emissions Reporting Maturity Model

Performance indicators (PI) based on the emissions reporting maturity model (ERMM) can be discovered using qualitative analysis over provided data. The set of performance indicators associated with the processes implemented by ERMM can be integrated into key performance indicators, which can better segregate cities into groups of interest.

Table 3.11: Decision table for ERMM KPI candidates. The aspects evaluated were used to select the most feasible KPI based on the available data.

KPI Candidate	**Self-contained ?**	**Self-explainable ?**	**Can be used alone ?**	**CEO friendly ?**	**Short-term implemented ?**	**Implementation stable ?**
ERM Level	Yes	Yes	Yes	Yes	Yes	Yes
ERM Rank	Yes	No	Yes	Yes	Yes	No
Data Management	No	Yes	Yes	No	Yes	Yes
Reporting Issues	No	No	Yes	No	Yes	Yes
Follow Up Issues	No	Yes	Yes	No	Yes	Yes

This work will present a key performance indicator based on the acquired maturity level, as it is congregated to help in the decision-making process. In addition, some other KPIs were also considered, and the decision table listed in 3.11 helped to select the most promising KPI.

Although the ERM-Level evaluation is a continuous process, it is stable enough to indicate the success of the overall initiatives for emissions reporting. Figure 3.21 shows an example of the evaluation spreadsheet for the city of Rio de Janeiro.

Emissions Reporting Maturity Level – Evaluation Matrix – Config WW 0a1a4a5a_AllFT										
Evaluated City	RIO DE JANEIRO			31176						
Evaluation Date	09/10/21									
Matrix version	1.0									
ProcessGroup	**Process**	**Subprocess**	**Practice**	**SubPractice**	**Reliability**	**Usability**	**Integration**	**Auditability**	**Reproducibility**	**Process Performance Indicator**
			CDP Model	Data mapping						
	Data Modeling		GCoM Model	Data mapping						1
			C40 Model	Data mapping						
			CDP data base selection	Data base filtering						
				SHDI accessing						
	Data Acquisition		External indexes	GDP accessing						111
				SCI accessing						
			GCoM data base reference	Data base accesing						
			C40 data base reference	Data base accesing						
				BIN file generation						
				CSV file generation						
Data Management		Pre-processing	Files generation	RAW file generation						
	Data Processing			STATS file generation						1
				INFO file generation						
			ClusWiSARD	Clustering						
		Clustering	Hierarquical Clustering	Clustering						
			K-Means Clustering	Clustering						
				Data filtering						
			Quantitative Analysis	Prevalence matrix						
				Distribution analysis						
	Data Analysis			Correlation analysis						4
				Samples selecting						
			Qualitative Analysis	Samples detailing						
				Samples analysis						
	Reports Building		Dashboard Generation	Document creation						1
			Report Generation	Document creation						
Reporting		Communication	Media Agents	Information delivery						
	Reports Publishing	Integration	Government Departments	Information delivery						0
			City Memberships	Information delivery						
			Mitigaton Plan	Document creation						
	Deployment		Action Plan	Document creation						1
Follow Up			Government Plan	Document creation						
	Monitoring	Implementation	Legislation	Initiative monitoring						1
		ROI	Emissions Reduction	Initiative monitoring						
Evaluation Scores										7
Emissions Reporting Maturity Level							3			

Figure 3.21: Example of emissions reporting maturity level applied to the city of Rio de Janeiro.

Chapter 4

Results

This section presents the results of the experiments and how they can corroborate the performance indicators developed based on the proposed methodology. The experiments set up and execution is described to illustrate the achievement of the results.

4.1 Experimental setup

The experiments that support the performance indicators development (PID) were executed in a controlled environment using a python-based system developed to help obtain and process the experimental results, statistics and logs. Python was chosen because of its well-known use in artificial intelligence solutions.

All the AI algorithms implementations needed in data exploration and data processing steps of the performance indicators development process can be found in python libraries such as wisardpkg (cluswisard) and sklearn (hierarchical cluster, k-means and dbscan). A general view of the execution schema is shown in Figure 4.1.

The support system operates in three modes:

- preprocessing: used in data preprocessing step and is responsible for transforming the input data from CDP's forms and additional data sets into outputs files used in the clustering phase inside the quantitative analysis step. The running mode executes the clustering algorithms using the CSV and BIN data files produced in the data preprocessing step. The data file type used depends on the clustering algorithm and its executing parameters.

- running: can be executed multiple times, varying the hyperparameters specific to each algorithm. The information about one execution is registered in both the general log and auxiliary log files. For example, for the ClusWiSARD algorithm, one execution can have multiple epochs to help find the best values for threshold and discriminator-limit hyperparameters. Each experiment's execution runs inside a renewed and isolated environment, avoiding any interference between executions.

- and post-processing: used to process additional statistics obtained from a set of executions, and it can be used to compare the results from different algorithms.

The hardware and software used to run the experiments are described in the table 4.1.

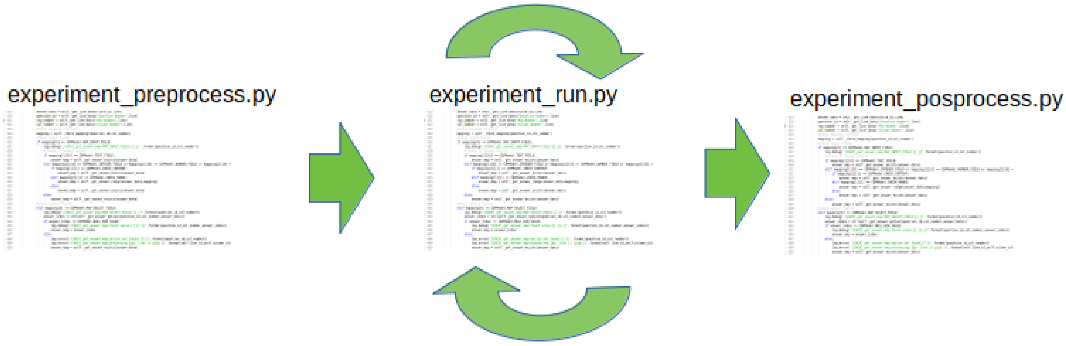

Figure 4.1: Experiment execution schema.

Table 4.1: Environment hardware and software details.

Parameter	Description and use
Machine	Dell T610 Server
CPU Cores	24
RAM	128 GB
Storage	120 GB
OS	Ubuntu 18.04.3 LTS
Python Version	3.6.7

4.1.1 Execution support programs

Experiment preprocess

The preprocessing step is executed using the python program `experiment_preprocess.py` with the following parameters: `cluster: <ClusWiSARD|HierarchicalCluster|KMeans|DBSCAN> CSV -d <dimension> -N <num_samples> -in <input_filename> [-out <output_filename>] [-bin] [-csv] [-f <filter>] [-D <datasetId>|<dataset_oper>:<dataset_filepath>] [-J <joinfieldId>@<dataset_id>:<dataset_fieldId>;<dataset_fields>] [-i <Info>] [-v None|Plot|Save|Review|Debug] [-o <options>]` . The parameter details are shown in table 4.2. An execution example of preprocessing for clustering using ClusWiSARD and CDP debugging database follows:

```
python3 experiment_preprocess.py cluster: ClusWiSARD
CSV -N 1000000 -d 2 -v Save -i CDP_Preprocess_AllCities_-
Config0a_AllFT -in ./input/2019_Emissions_Cities_Dataset_-
DEBUG.csv -bin -csv -f I:Question Number=0* -o copy_-
dat=./input/cdp/cluster_allcities_0a_AllFT.dat,copy_-
out=./input/cdp/cluster_allcities_0a_AllFT_out.csv,copy_-
stats=./input/cdp/cluster_allcities_0a_AllFT_stats.csv
```

Table 4.2: Experiment preprocess parameters description.

Parameter	**Description and use**
<ClusWiSARD | HierarchicalCluster | KMeans | DBSCAN>	main target engine of preprocessing
-d <dimension>	number of dimensions of input data set
-N <numSamples>	number of samples to process
-in <inputFilename>	input data set file
-out <outputFilename>	optional output filename
-bin	command to generate binary output
-csv	command to generate text (csv) output
-f <filter>	filters applied to select samples from input data set
-D	data set to be loaded:
<datasetId>	data set identification to be referenced internally
<datasetOper>	operation to be applied over data set (ex: cluster)
<datasetFilepath>	data set file path
-J	join field values from different data sets:
<joinfieldId>	field id in data set source of information
<datasetId>	source data set
<datasetFieldId>	match field
<datasetFields>	data fields to be retrieved
-i <Info>	mnemonic used to identify the execution in the future
-v	verbose level:
None	nothing is printed in auxiliary files nor saved in logs
Plot	auxiliary files are generated but nothing in logs
Save	both auxiliary files and logs are generated
Review	logs receive additional information
Debug	logs receive debug information
-o <options>	options specific to each target engine. Examples:
copy_dat=<filepath>	also copy generated output binary (dat) file to <filepath>
copy_out=<filepath>	also copy generated output csv file to <filepath>
copy_stats=<filepath>	also copy generated statistics file to <filepath>
Filter construction:	<I|E>:<#SampleId|#FieldType|<FieldName»=<[@]value>;...
<I|E>	include/exclude rule
<#SampleId|#FieldType|<	filter type
#SampleId	filter samples which ids are in the list <value> or in filename <@value>
#FieldType	filter samples which field type is defined by <value> (YN, NUMBER, DATE or YEAR)
FieldName	filter samples in which field content is equal to <value>

Experiment run

The python program experiment_run.py is used to run the experiments, using the following command-line syntax from a terminal: cluster: <ClusWiSARD|HierarchicalCluster|KMeans|DBSCAN> Grp -d <dimension> -N <num_samples> [-e <executions>] [-D <datasetId>|<datasetOper>:<datasetFilepath>[,<datasetId>:<datasetFilepath>]] [-i <Info>] [-v None|Plot|Save|Review|Debug] [-o <options>] . The parameter details are shown in table 4.3. An execution example using ClusWiSARD to cluster a data set with midle-east (ME) cities and questions "0" and "1" and its subquestions:

```
python3 experiment_run.py cluster: ClusWiSARD Grp -N
1000000 -d 2 -e 10 -v Save -i ClusWiSARD_N100000_ME_0a1a_-
AllFT_e10 -D cluster:./input/cdp/cluster_mecities_0a1a_-
AllFT.dat -o config=me0a1a_AllFT,update_clusters=true,save_-
analytics=true,threshold=auto,discriminatorLimit=auto,sufix=exec
params,dump_data=true,configs_log=true
```

Experiment post-process

The post-processing step is executed using the python program experiment_-postprocess.py with the following parameters: -N <num_samples> -in <input_filename> [-out <output_filename>] [-i <Info>] [-v None|Plot|Save|Review|Debug] [-o <options>] . The parameter details are shown in table 4.4. Execution example:

```
python3 experiment_postprocess.py -N 1000000 -v Save -i CDP_-
Postprocess_AllCities_Config0a_AllFT -in ./input/stats_-
config0a_AllFT.csv -o copy_stats=./input/cdp/cluster_-
allcities_0a_AllFT_stats_postprocess.csv
```

Execution jobs

The experiments' executions can be run in parallel with the help of experiment_-job.sh utility. The experiments' configurations can be organised in text (jobs) files and be executed in parallel. The job execution general view is shown in Figure 4.2. The experiment_job.sh utility cab be controlled using the parameters:

experiment_job.sh <module> <command>

experiment_job.sh <module> <job_id> [<max_num_procs>]

Table 4.3: Experiment run parameters details.

Parameter	Description and use
<ClusWiSARD | HierarchicalCluster | KMeans | DBSCAN>	main target engine of preprocessing
-d <dimension>	number of dimensions of input data set
-N <numSamples>	number of samples to process
-e <executions>	number of executions (epochs)
-D	data set to be loaded:
<datasetId>	data set identification to be referenced internally
<datasetOper>	operation to be applied over data set (ex: cluster)
<datasetFilepath>	data set file path
-i <Info>	mnemonic used to identify the execution in the future
-v	verbose level:
None	nothing is printed in auxiliary files nor saved in logs
Plot	auxiliary files are generated but nothing in logs
Save	both auxiliary files and logs are generated
Review	logs receive additional information
Debug	logs receive debug information
-o <options>	options specific to each target engine. Examples:
config=<config_name>	config identification
update_clusters=<true|false>	update (or not) clusters aggregation logs
save_analytics=<true|false>	save analytics information
threshold=<auto|value>	threshold value for ClusWiSARD hyperparameter or discovery best value with auto option
discriminatorLimit=<auto|	threshold value for ClusWiSARD
	hyperparameter or discovery best value with auto option
sufix=exec_params	add execution params to gererated files
dump_data=<true|false>	dump (or not) data used in clustering
configs_log=<true|false>	save (or not) the configuration logs

module: run|preprocess|postprocess

command: pause|resume|finish|kill|update

job_id: job filename present in ./jobs folder, without _module.job sufix

* update command requires <max_num_procs> to be informed . The parameter details are shown in table 4.4. Execution example:

```
./experiment_job.sh preprocess cluswisard_allcities 8
```

4.1.2 Experiments configurations

The experiments' configurations are a central part of the performance indicator development process. The variations in the configurations are mapped and used to calibrate the progress in finding potential candidates for performance indicators. The questions in CDP form are identified by numbers and the sub-questions by

Table 4.4: Experiment postprocess parameters details.

Parameter	Description and use
-N <numSamples>	number of samples to process
-in <inputFilename>	input data set file
-out <outputFilename>	optional output filename
-i <Info>	mnemonic used to identify the execution in the future
-v	verbose level:
None	nothing is printed in auxiliary files nor saved in logs
Plot	auxiliary files are generated but nothing in logs
Save	both auxiliary files and logs are generated
Review	logs receive additional information
Debug	logs receive debug information
-o <options>	options specific to each target engine. Examples:
copy_stats=<filepath>	also copy generated statistics file to <filepath>

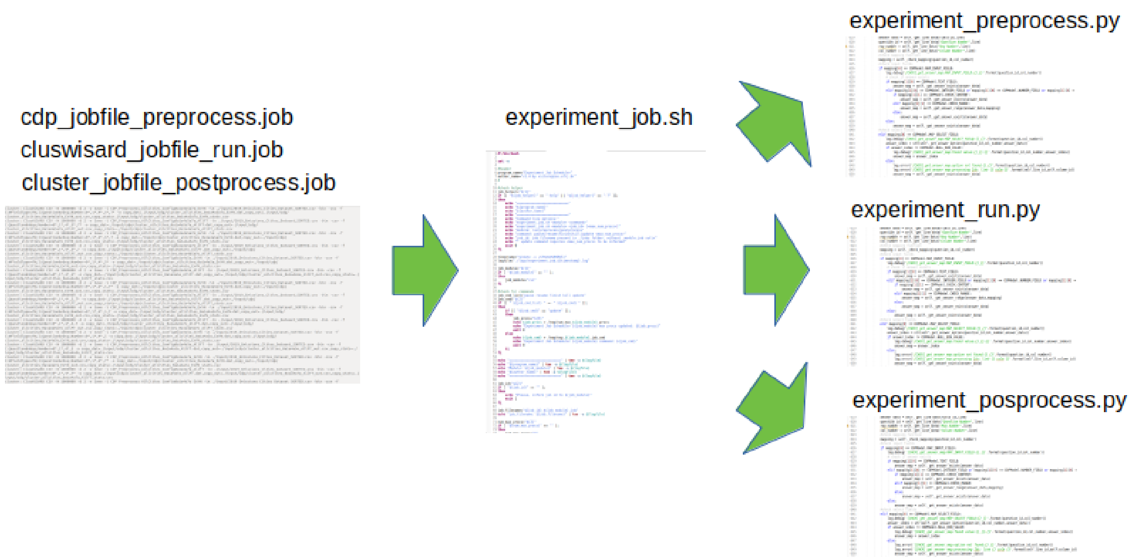

Figure 4.2: Experiment jobs schema general view.

numbers and letters. The configuration settings can use wildcards to indicate sub-questions inclusion (or exclusion) in the filters. Along with this, the use (or not) responses computed as just answered/not answered, indicated by field type YN, summed up more possible combinations. Other filters can narrow results like the geographic region or extend the analysis with external data.

Configuration pattern applied in preprocessing step:

```
<Geo><Questions>_<FieldType>[_<ExtData>]
Geo:  ww|br|latam|...
Question:  {[01457][ao ],}
Sub-questions:  a(all),o(one-level)
ExtData:  SHDI|GDP|SHDI&GDP
```

In the running phase, the configuration can receive hyper parameters information as they can vary among different experiments:

Table 4.5: Experiment job parameters details.

Parameter	Description and use
<module>	correspondent step in experiment execution step
run	experiment running step
preprocess	experiment preprocessing step
postprocess	experiment post-processing step
<command>	command to be executed
pause	pause jobs execution
resume	resume jobs execution after been paused
finish	stop jobs execution but wait for experiments to finish
kill	stop jobs execution and force experiments to quit
update	update the number of experiments executed in parallel
<job_id>	job file identification
<max_num_procs>	max number of experiments in parallel

```
<Geo><Questions>_<FieldType>[_<ExtData>][_<Algorithm_Specific>]
```

ClusWiSARD example:

```
ww0a1a_AllFT_tdauto_dl20
```

4.1.3 Self-test experiments

The self-test experiments are used to provide execution information about the internal mechanisms of the support system to help plan the execution of the interest experiments. The configuration and data sets are simple to be manually verified but complete enough to exercise all support system functions that the interest experiments will use. The logs and statistics are checked in debug mode to confirm the expected execution. The configuration used to accomplish it was the 0a_AllFT. The sequence of execution and checking is shown below:

```
• python3 experiment_preprocess.py cluster:  ClusWiSARD CSV
  -N 1000000 -d 2 -v Debug -i CDP_Preprocess_AllCities_-
  Config0a_AllFT -in ./input/2019_Emissions_Cities_-
  Dataset_DEBUG.csv -bin -csv -f I:Question Number=0*
  -o copy_dat=./input/cdp/cluster_allcities_0a_-
  AllFT.dat,copy_out=./input/cdp/cluster_allcities_0a_-
  AllFT_out.csv,copy_stats=./input/cdp/cluster_allcities_-
  0a_AllFT_stats.csv

• python3 experiment_checklogs.py -v Save -i CDP_-
  Preprocess_AllCities_Config0a_AllFT

• python3 experiment_run.py cluster:
  ClusWiSARD Grp -N 1000000 -d 2 -e 10 -v
  Debug -i ClusWiSARD_N100000_0a_AllFT_e10 -D
```

```
cluster:./input/cdp/cluster_allcities_0a_AllFT.dat
-o config=ww0a_AllFT,update_clusters=true,save_-
analytics=true,threshold=auto,discriminatorLimit=auto,sufix=
params,dump_data=true,configs_log=true
```

- ```
  python3 experiment_checklogs.py -v Save -i CDP_-
  Preprocess_AllCities_Config0a_AllFT
  ```

- ```
  python3 experiment_postprocess.py -N 1000000 -v
  Save -i CDP_Postprocess_AllCities_Config0a_AllFT
  -in ./input/stats_config0a_AllFT.csv -o copy_-
  stats=./input/cdp/cluster_allcities_0a_AllFT_stats_-
  postprocess.csv
  ```

- ```
  python3 experiment_checklogs.py -v Save -i CDP_-
  Postprocess_AllCities_Config0a_AllFT
  ```

4.1.4 Clustering experiments

The clustering experiments are based on ClusWiSARD to group samples (cities) with similar or related answers and other clustering mechanisms to validate and narrow the quantitative analysis process. The experimental results using ClusWiSARD can be seen as "pictures" taken from the binary correspondence of the CDP forms' responses. The similarities in the answers are registered and used to group the samples into clusters.

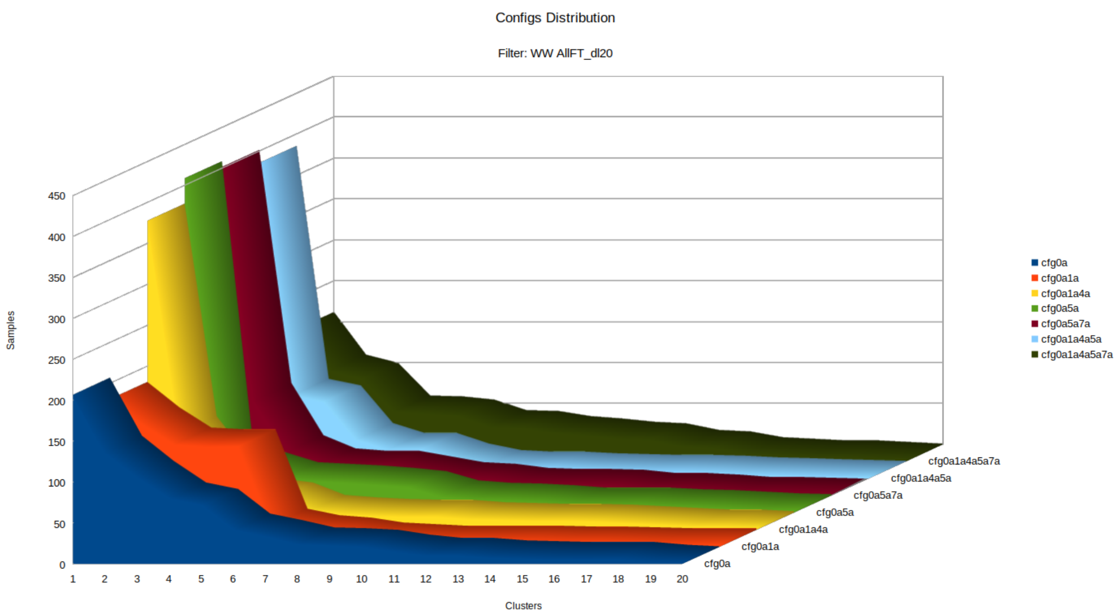

Figure 4.3: Configuration distributions.
  ```

The best configurations' distributions are those with a higher level of concentration of samples in the first clusters, as it is shown in Figure 4.3.

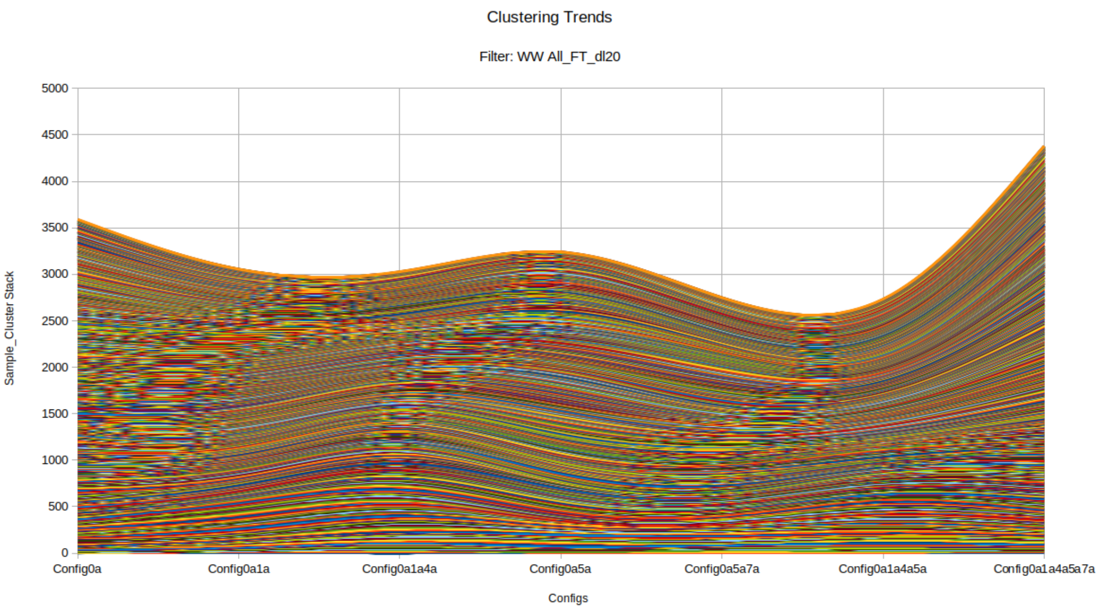

Figure 4.4: Cluster cumulative distribution.

Trends

Trends are another way to analyse which configurations have more chances to produce performance indicators based on the distributions of the samples and clusters. The computation of the trends is based on the formula 4.1.

$$t(c) = \sum_{k=1}^{N} \|S_{k,i} - S_{k,j}\| \tag{4.1}$$

where c is the configuration being analysed; $i <> j$; $0 < i,j < \max(\text{clusters})$; $S_{k,i}$ is the $k-ist$ sample in sample list that has cluster i as its main cluster choice; $S_{k,j}$ is the $k-ist$ sample in sample list that has cluster j as other clusters choices.

Trend Points

Trend points are the configurations that can indicate higher or lower variability among the analysed configurations. The trends and trend points can be seen in the Figure 4.4.

1. a higher variability among cluster distribution indicates some hindrance in choosing possible clusters to group a sample. Thus, the configuration being

evaluated is not a good candidate to offer clear answers for performance indicator candidates.

2. a lower variability among cluster distribution indicates some uniformity among the answers. This uniformity can indicate a better place to focus on the analysis to find performance indicators candidates among CDP questions.

Clustering using all cities

The first attempt to discover candidates to performance indicators used all available emissions data in CDP forms related to all partnered cities. The idea was to select and process as much available data as possible, varying the configurations' parameters, filtering different questions levels and types of fields. The best configuration using all cities' data was WW_0a1a4a5a_AllFT_dl20, as shown in Figure 4.3.

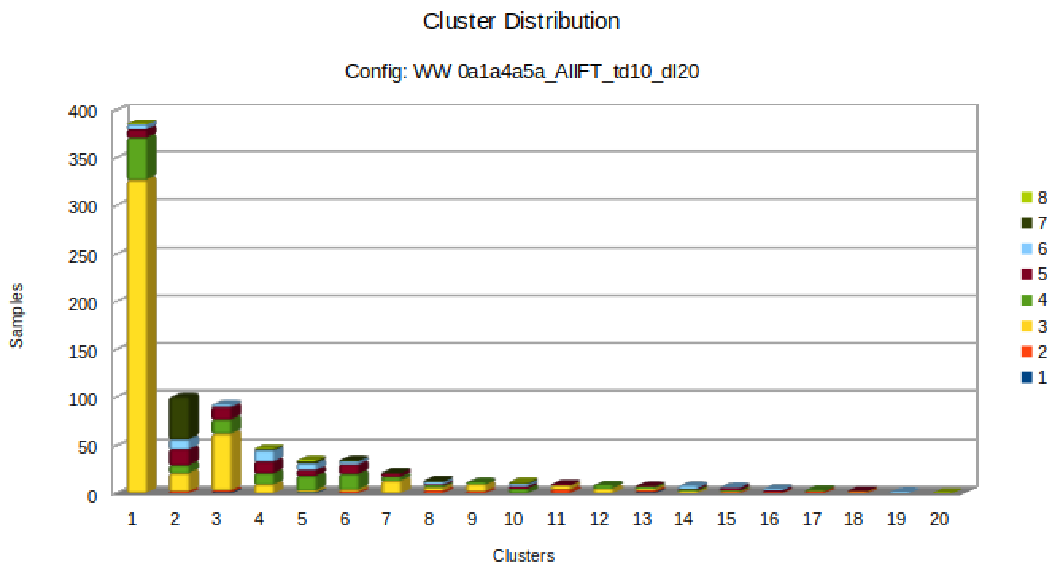

Figure 4.5: Clusters distribution using configuration ww0a1a4a5a_AllFT.

The cluster distribution of ClusWiSARD method can be seen in Figure 4.5. The instability level of the distribution is measured by indexes 1 through 8, which represent the number of clusters chosen along with the ten executions of each ClusWiSARD experiment. Here, 87% of the samples clustered in cluster 1 was also clustered in 2 other clusters along the clustering process.

The self-test experiments indicate that ten executions are enough to generate stable sets of clusters for ClusWiSARD using the available data.

After running the ClusWiSARD experiment using the configuration WW_-0a1a4a5a_AllFT_dl20, the results were narrowed using the other two methods (hierarchical clustering and k-means). The prevalence rates are shown in Figure 4.6.

Prevalence rate: #samples / cluster mean size

		HierarchicalClustering 0a1a4a5a_AllFT																			
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
ClusWiSARD 0a1a4a5a_AllFT	1	0,62	0,36	0	0,19	0,07	0,17	0	0	0,01	0,03	0	0	0	0	0	0,04	0	0,06	0,01	0
	2	0	0,05	0,02	0	0,07	0	0,12	0,46	0,05	0,19	0	0,06	0	0	0,32	0,02	0,02	0,07	0,05	0,05
	3	0	0	0,72	0	0,03	0	0	0	0	0	0,02	0	0,1	0,4	0,03	0	0	0	0	0,02
	4	0	0,03	0	0,02	0	0	0,13	0	0	0,13	0,69	0	0	0,03	0	0	0,18	0	0	0,03
	5	0	0	0	0,4	0	0	0	0	0	0	0,03	0	0	0	0	0,29	0	0	0,26	0
	6	0	0	0,02	0	0,35	0	0,43	0	0,03	0,03	0	0	0	0	0	0	0	0	0	0,13
	7	0,07	0	0	0,02	0	0	0	0	0,08	0,08	0	0,04	0	0	0	0	0,44	0	0	0
	8	0	0	0	0	0	0	0	0	0,05	0,05	0	0	0,5	0	0	0	0	0	0,15	0
	9	0	0	0,03	0	0	0	0	0	0,5	0	0	0,05	0	0	0	0	0	0	0	0
	10	0	0	0	0	0,29	0	0	0	0,1	0	0	0	0	0	0	0,06	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	0	0,54	0	0	0	0	0	0	0	0
	12	0	0	0,08	0	0,03	0	0	0	0	0	0	0	0	0	0,06	0	0	0	0	0,38
	13	0	0	0	0	0	0	0	0	0	0	0	0,46	0	0	0	0	0	0	0	0
	14	0	0	0	0	0	0	0,16	0	0,11	0	0	0	0,18	0	0	0	0	0	0	0
	15	0	0	0,03	0	0,11	0	0	0	0	0	0	0,06	0	0,06	0	0,07	0	0	0	0
	16	0	0	0	0	0	0	0	0	0	0	0	0,06	0,26	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0,25	0	0	0	0	0	0	0	0	0	0	0
	18	0	0	0	0	0	0	0	0	0	0	0	0	0,21	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,19	0	0	0
	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,1	0	0	0

Figure 4.6: Prevalence rate for configuration WW_0a1a4a5a_AllFT.

The next step was to check the best prevalence rate found to get the samples to be processed: 0,72 for cluster 03 in the hierarchical clustering method and cluster 3 in the ClusWiSARD's experiment. The result of the reprocessing is shown in Figure 4.7.

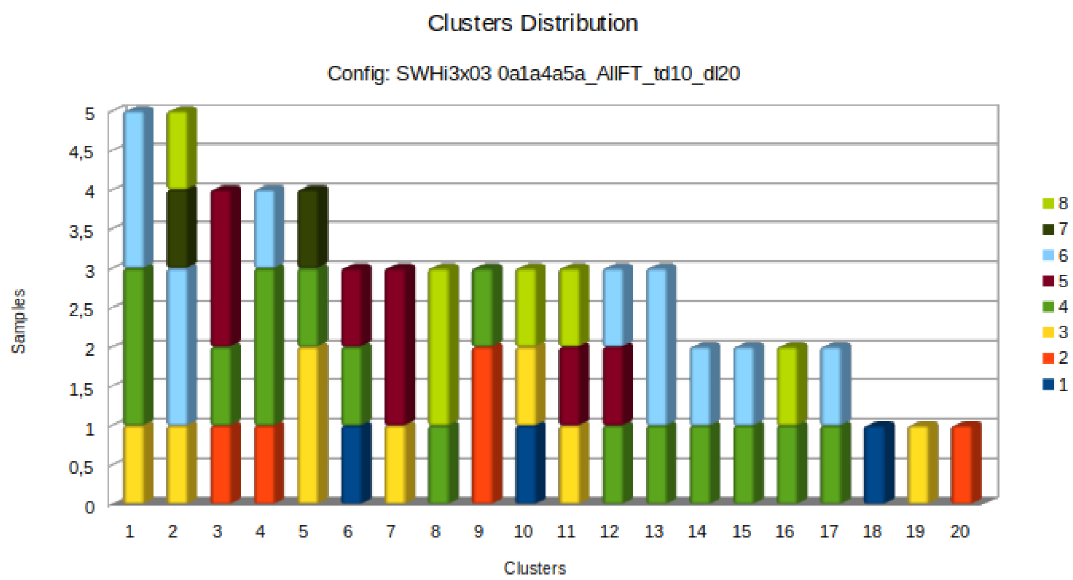

Figure 4.7: Clusters distribution using configuration swhi3x03 0a1a4a5a_AllFT.

As expected, the distribution is much more equal since the selected samples share features obtained using the first step in the process: clustering and validating the intersection between ClusWiSARD and other clustering methods. Again, the prevalence rate was checked to narrow the results, as shown in Figure 4.8

Prevalence rate: #samples / cluster mean size

		HierarchicalClustering 0a1a4a5a_AllFT																			
		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
ClusWiSARD 0a1a4a5a_AllFT	1	0,23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0,19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,33
	3	0,19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	4	0,19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0,14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,4	0	0	0
	6	0,1	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	0,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0,5	0	0	0	0	0
	8	0,15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0,05	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0,5	0	0	0	0
	10	0,05	0	0	0,5	0	0,5	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	11	0,1	0	0	0	0	0	0	0	0	0	0	0	0,5	0	0	0	0	0	0	0
	12	0,05	0	0	0	0	0	0	0	0	0,5	0	0	0	0,5	0	0	0	0	0	0
	13	0,05	0	0	0	0	0	0	0	0	0	0,5	0	0	0	0	0	0	0,5	0	0
	14	0,1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	15	0,05	0	0	0	0	0	0	0,67	0	0	0	0	0	0	0	0	0	0	0	0
	16	0,05	0	0	0	0	0	0	0	0,67	0	0	0	0	0	0	0	0	0	0	0
	17	0	0	0	0	0	0	0	0	0	0	0	0,67	0	0	0	0	0	0	0,67	0
	18	0,05	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	20	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 4.8: Prevalence rate for configuration SWHi3x03_0a1a4a5a_AllFT.

The next step was to apply a qualitative analysis over the results to look for candidates to performance indicators based on the answers provided by the cities in the selected samples. The qualitative analysis considers both the content of the answers and how good they represent reality.

A subset of the selected samples (20%) was picked at random to be checked. Then, the questions were analysed one by one, comparing the answers to each other and to external data when it is available. The comparison of the answers generates a quality indicator as described in table 4.9, in which are computed if the answers exist or not, the variation in their values (if measurable) and how different they are from other sources. In addition, the availability, difficulty of access and reliability of other sources are also registered and used to bind the answers.

Clustering using CDP regions

The CDP geographic regions are a categorisation used in CDP forms that can reflect the expansion of CDP partnerships with the cities. The regions are listed in 4.6 and the diversity each region represents affected the way the answers were informed. This diversity can be seen in the results of experiments that use CDP regions as filtering in the preprocessing phase, as it is shown in Figure 4.9

The clustering distribution using CDP regions is uneven, indicating the independence of the answers from which region a city is classified in the CDP database. Therefore, CDP regions cannot be used as a reliable bias to help find performance

Table 4.6: CDP codes and regions list.

Region Label	Region Name
NORAM	North America
EURO	Europe
SEASOC	Southeast Asia and Oceania
EAS	East Asia
AF	Africa
LATAM	Latin America
SWAS	South and West Asia
ME	Middle East

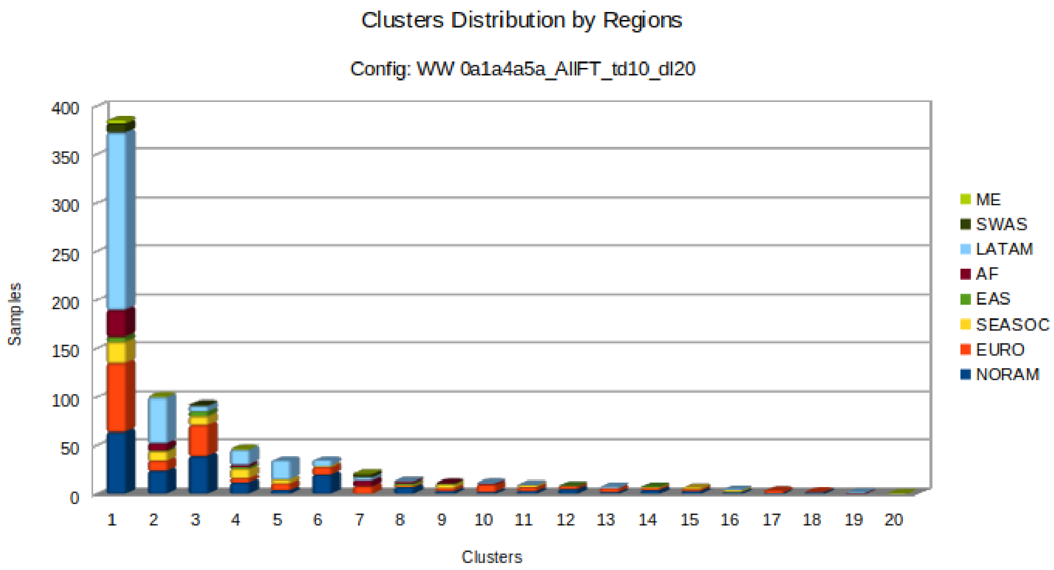

Figure 4.9: Clusters distribution by regions using configuration WW_0a1a4a5a_-AllFT.

indicators among the answers. However, they can still help measure the level of quality of the answering process.

Clustering using cities of Brazil

Another alternative approach to finding candidates for performance indicators was to use only the cities of Brazil present in the CDP database and compare the patterns of cluster's distributions. There are 111 cities from different regions in Brazil that joined the CDP initiative in different periods. Figure 4.10 shows the distribution of the cities along the clusters is also uneven. It indicates a different pattern than that observed in the whole database cluster distribution.

According to Instituto Brasileiro de Geografia e Estatística (IBGE), Brazil is one of the few countries with a human development index (HDI) gathered at the city level. The sub-national human development index (SHDI) database was used to compare the clustering results of Brazil's cities. The buildup of SHDI follows the same rule of HDI, but with data provided and restricted do the sub-national level,

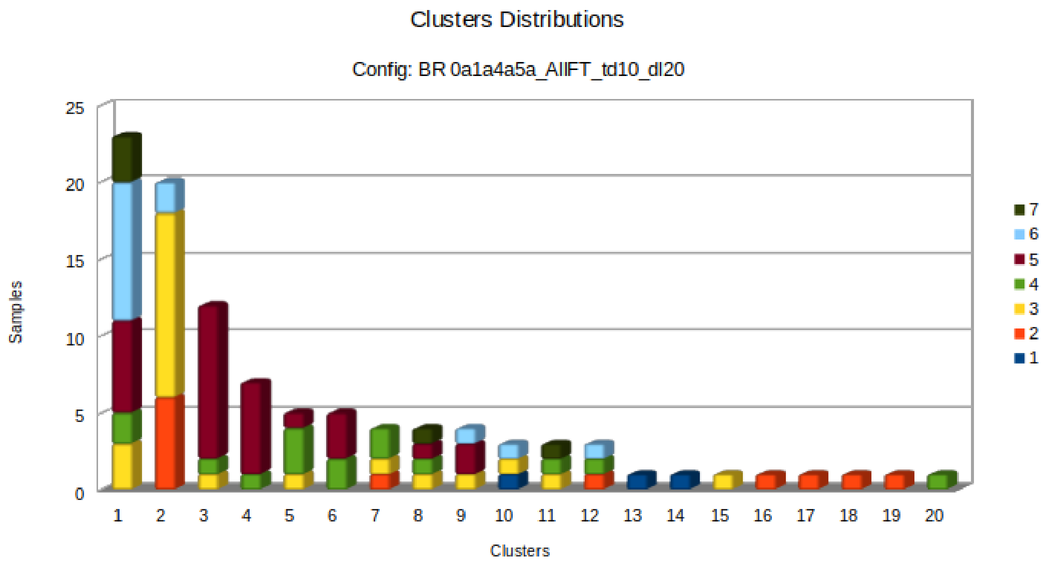

Figure 4.10: Clusters distribution using configuration BR_0a1a4a5a_AllFT.

like states in a federation or a city, town or local government alike. The result is shown in Figure 4.11.

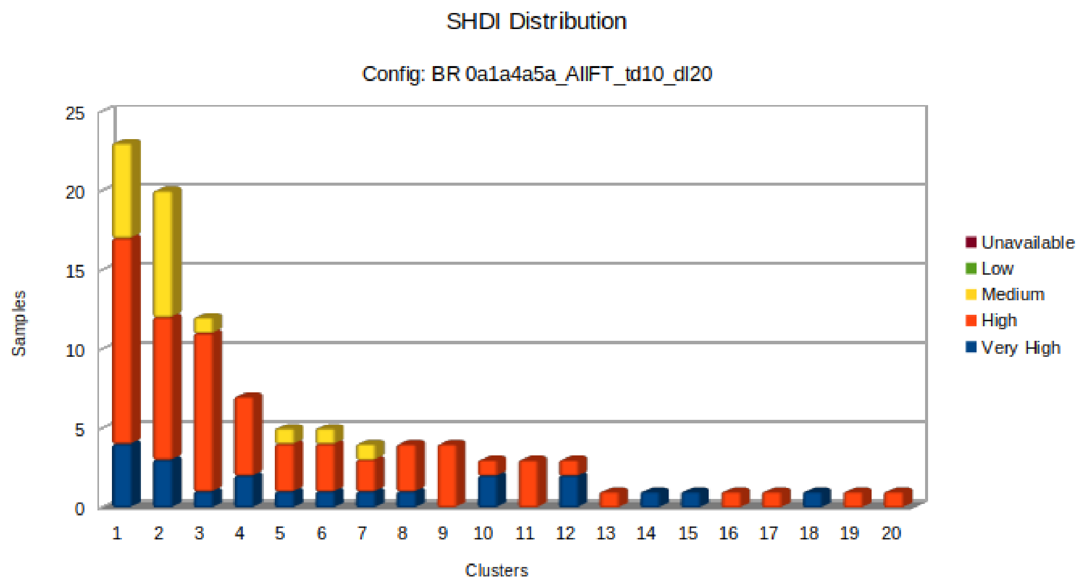

Figure 4.11: Clusters distribution using configuration BR_0a1a4a5a_AllFT.

Clusters 9 and 11 have only high SHDI cities on them, so they were used in the drill down to select city samples for qualitative analysis in this case. An example of the analysed samples are in table 4.7. The first attempt was to take the first question's values with two parts (columns) (Q_0.1_1:Administrative boundary and Q_0.1_2:Description of the city) for all samples and compare them. Here, all values for Q_0.1_1 were chosen "2: City / Municipality" and "1" to indicate that the "description of the city" was informed.

Table 4.7: Selected samples from clusters 9 and 11 and which have "high SHDI". The answers for the Question 0.1 and its sub-questions are shown.

CDP Id	City Name	Cluster	Q_01_1	Q_01_2
50383	Sorocaba	11	2	1
54681	Araçatuba	9	2	1
55380	Cubatão	9	2	1
60267	Guarujá	9	2	1
60292	Jaú	11	2	1
60318	Porto Velho	11	2	1
60349	São Leopoldo	9	2	1

The next attempt was to get another question to check. The question "Q_-1.0:Does your city incorporate sustainability goals and targets (e.g. GHG reductions) into the master plan for the city?" was chosen, and the values for the selected samples were "1: Yes" and "5: Don't Know".

Table 4.8: Selected samples from clusters 9 and 11 adn which have "high SHDI". The answers for the Question 1.0 and its sub-questions are shown.

CDP Id	City Name	Cluster	Q_10
50383	Sorocaba	11	5
54681	Araçatuba	9	1
55380	Cubatão	9	1
60267	Guarujá	9	1
60292	Jaú	11	5
60318	Porto Velho	11	5
60349	São Leopoldo	9	1

These results show a direct relationship between "a plan for the city that incorporates sustainability goals and targets" and cluster 9. However, cluster 11, in this extraction, remained with samples that "don't know" the answer to the question.

4.2 PIs developed from Emissions Reporting Correlations

Analysing the emissions reporting data correlations are another way to look for performance indicators. After some experiments with broadly used indicators like GDP, HDI and their variants, and others like OECD, C40, GCoM and SCI memberships, some results indicate correlations between then and emissions reported by cities. The cities total emissions clusters distribution is shown in Figure 4.12.

Some of these correlations are already a theme of published studies. Despite the correlation level that can be obtained using statistical methods like covariance, it is not in the scope of this work to compare the results obtained with other published

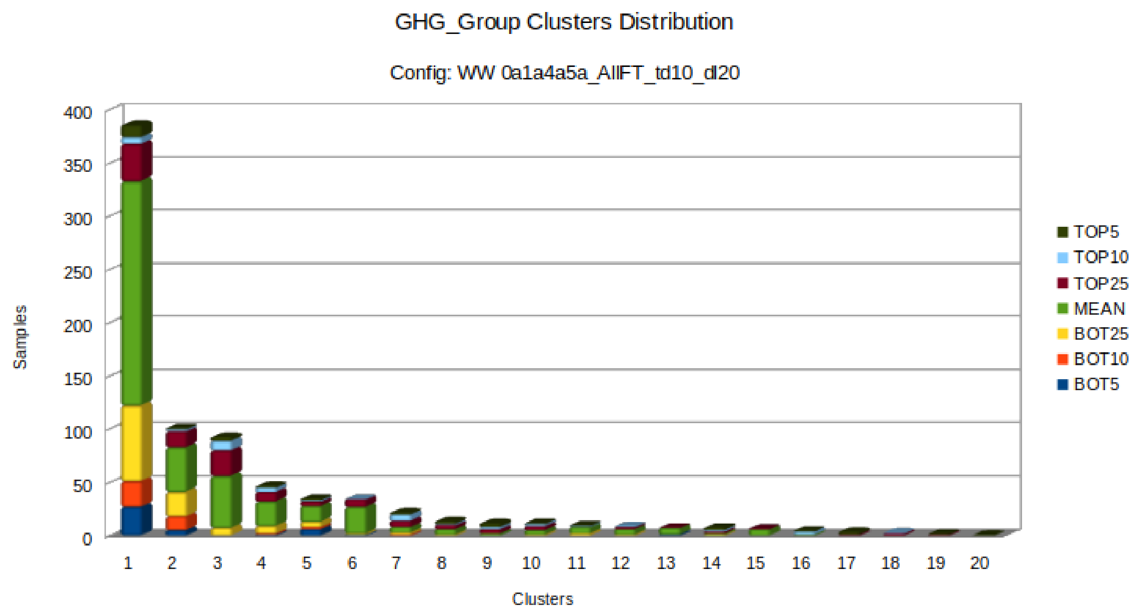

Figure 4.12: Cities total emissions in 2019 clusters distribution using configuration WW_0a1a4a5a_AllFT.

results using only statistics. Instead, this work focuses on the correlations resulting from using ClusWiSARD method, validated with other clustering methods hierarchical clustering and K-means, and analysed with the help of a qualitative view over the data.

The GHG_Group represents a normal distribution of total emissions registered and provided by the cities in the database of CDP. The values used as percentiles of the distribution are shown in table 4.9. The GHG_Group percentiles are uneven along with the clusters, indicating that different clusters that represent groups of cities with similar answers, at some level, are subject to different levels of emissions.

Table 4.9: Cities total emissions in 2019 distributions values and percentiles.

Label	**Percentile**	**Min Value**	**Max Value**
TOP5	Up 5%	28,005915	4011,500633
TOP10	Up 10%	13,957336	28,005915
TOP25	Up 25%	3,319459	13,957335
MEAN	Mean	0,219558	3,319458
BOT25	Least 25%	0,046210	0,219557
BOT10	Least 10%	0,014927	0,046209
BOT5	Least 5%	0	0,014926

Based on this fact, two hypotheses can be presented: the total emissions of a city are not related to its level of development and the cities, in general, are incapable of correctly registering emissions information. The results and analysis that help answer these hypotheses are detailed in the following sections.

4.2.1 PI: SHDI x Cities Emissions

The correlations between emissions and human development index (HDI) can be seen in studies at the country level. Nevertheless, this work aims to use available subnational human development index (SHDI) information to investigate the correlation between this index and cities total emissions. Figure 4.13 shows the SHDI along with the cluster distribution.

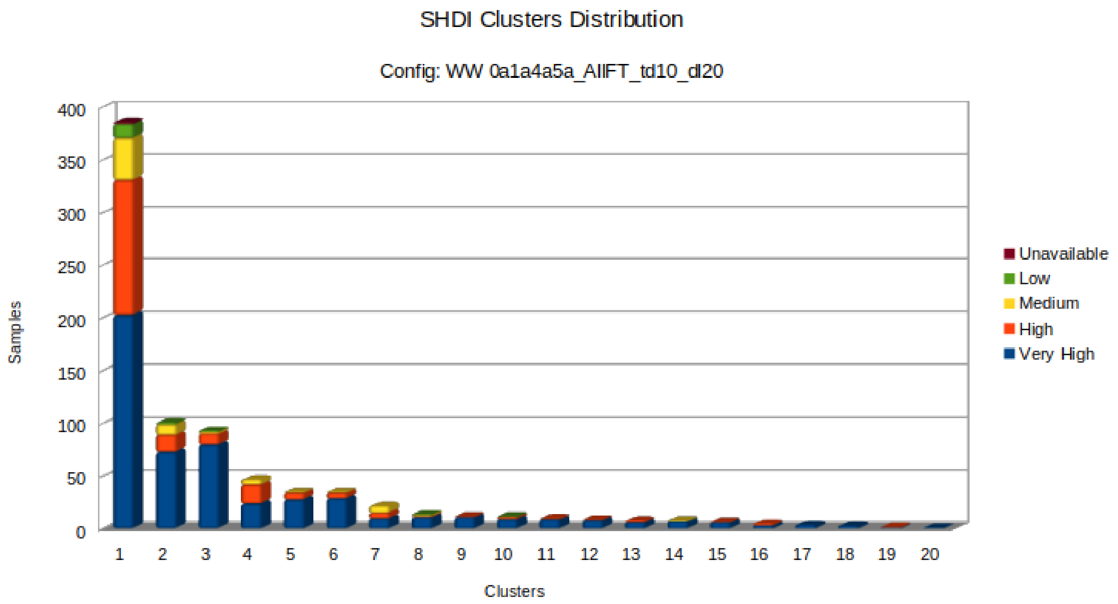

Figure 4.13: SHDI clusters distribution using configuration WW_0a1a4a5a_AllFT.

The SHDI is distributed along with all clusters, even with different frequencies, so it cannot be used alone as a performance indicator. Although, when SHDI is mixed with total emissions, the correlation appears, as shown in Figure 4.14.

Different classes combining both variables can be used to indicate some level of development of a city in terms of emissions mitigation and reduction. These trade-offs can be seen in Figure 4.15, as at least two correlations can be extracted. The "high" and "very high" SHDI indicators are in the top 5% of cities with more emissions, indicating that high development is related to more emissions. However, 25% of cities with fewer emissions tend to have SHDI distribution more even along with the clusters, indicating a more direct relationship between development and emissions reductions policies.

4.2.2 PI: OECD x Cities Emissions

The organisation for Economic Co-operation and Development (OECD) is a membership of developed countries that cooperate in economic matters. The cities in these countries follow the policies and directives negotiated in forums promoted by

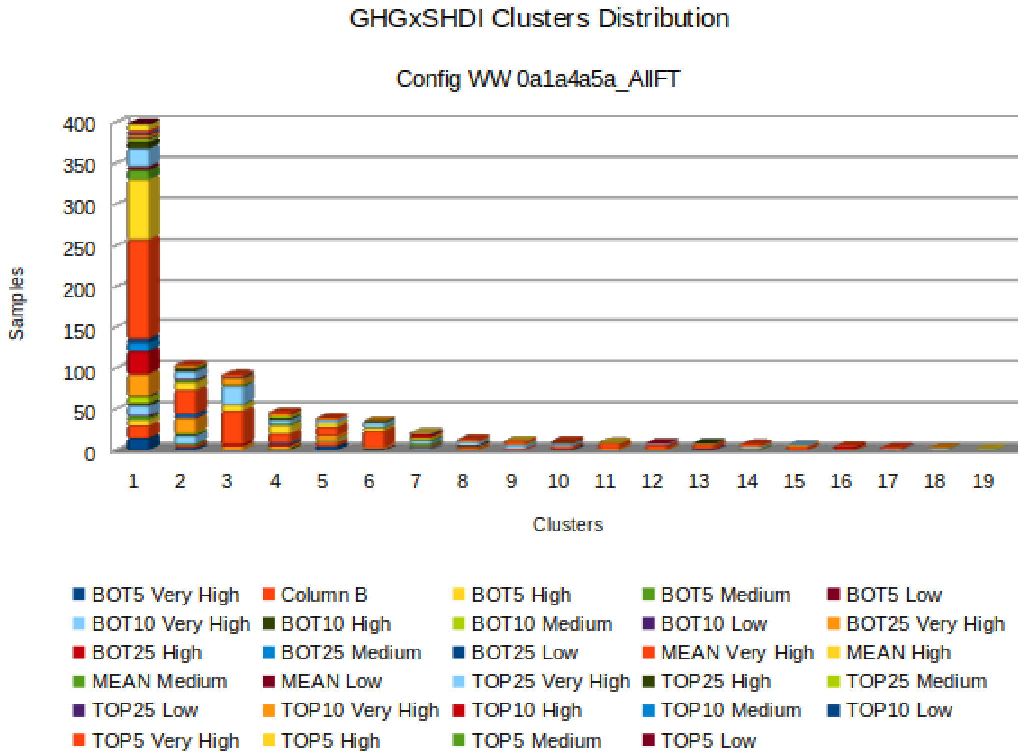

Figure 4.14: Cities total emisisons x SHDI clusters distribution using configuration WW_0a1a4a5a_AllFT.

OECD. To some extent, part of these policies is applied as is by the cities. The list of countries members of OECD is available in the appendix.

The correlation between cities total emissions in 2019 and belonging to a country being part of OECD is shown in Figure 4.16.

Based on the results shown in Figure 4.17 there is a direct relationship between being part of OECD countries and high emission rates. As mentioned before in this work, studies confirm this correlation at the country level, taking all emissions produced by the country members of the OECD and comparing them to those in emerging and developing countries.

Even though the distribution of the cities in the CDP database does not reflect what is observed in the real world, the method and results still point to the same result. The level of economic development of a country can be used as a performance indicator about how much emissions that country proportionally produces.

4.2.3 PI: GCoM Membership

The Global Covenant of Mayors for Climate & Energy (GCoM) has more than ten thousand participant cities. It aims to monitor hazardous events related to climate and energy that occur inside cities' boundaries through mitigation plans and other additional information. Some of this information is related to emissions (total and

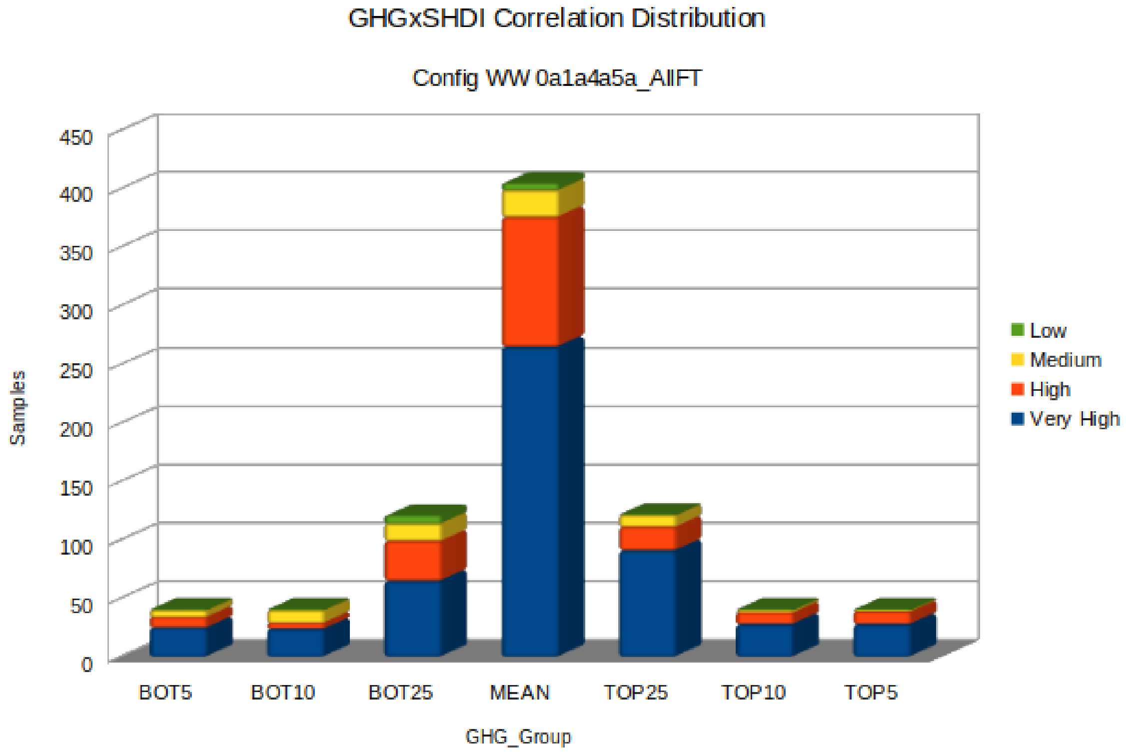

Figure 4.15: Cities total emisisons x SHDI correlation distribution using configuration WW_0a1a4a5a_AllFT.

by sectors) and energy assessment and plans, as explained in table 4.10.

Table 4.10: GCoM database information: fields and descriptions.

Field	**Description**
GCoM Id	Identification of city in GCoM database
City Name	City identification
Compliance Year	Year that city joined the initiative
Bagdes	Indication of the presence of assessments, targets and plans
Emissions	Emissions information (total, buildings, transportation
	industry, waste and other
Hazards	Indication of hazardous events (flood, fire, chemical, etc)
Action Plan	Hyperlink to the mitigation plan
Joint Plan	Hyperlink to the joint plan
Adaptation Plan	Hyperlink to the an adaptation plan

The GCoM membership cluster distribution is shown in Figure 4.18 and those GCoM members that have planning actions registered in GCoM database is shown in 4.19.

During the analysis, the prevalence between GCoM and CDP database was 73%, indicating that 219 cities reached by CDP partnership are still not part of GCoM by 2019. This analysis aims to verify the viability of the information provided using CDP forms to help join other city partnerships and GCoM among them. Considering the clusters with a prevalence ratio over 90%, clusters 5, 12, 16, 17, 18, 19, and 20 had 59 samples reprocessed. The new clusters distributions can be seen in Figures

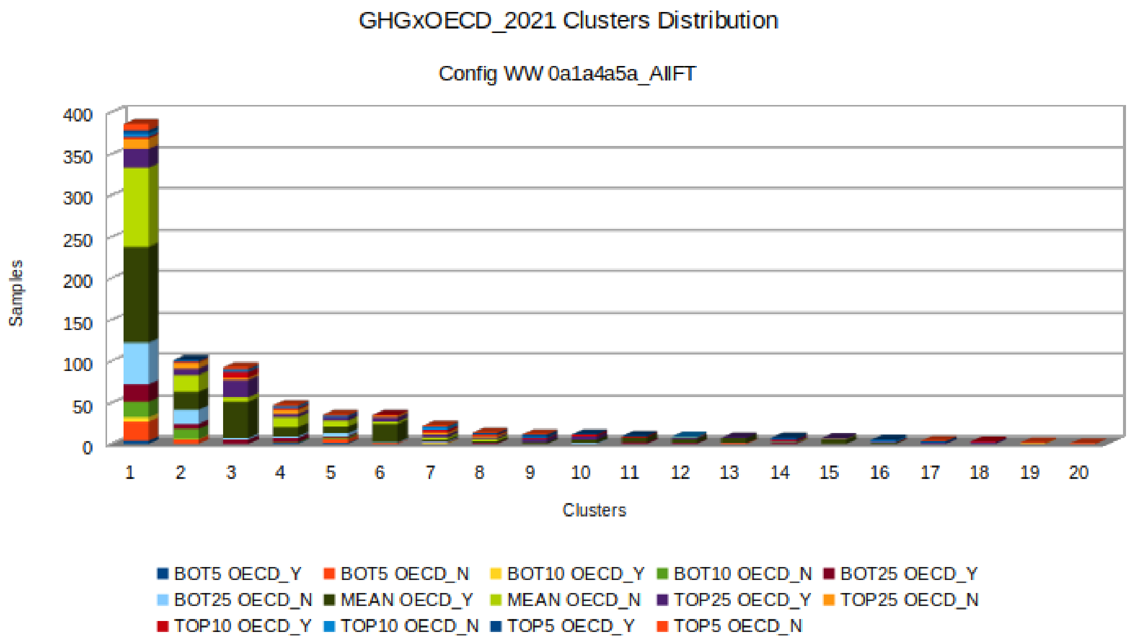

Figure 4.16: Cities total emisisons x OECD clusters distribution using configuration WW_0a1a4a5a_AllFT.

4.20.

Analysing the new cluster distribution, the stable clusters (samples distributed to only one cluster along with ten executions), which was clusters 8, 9, 13, 14, 15, 16 and 17, selected ten samples that were analysed in detail to look for similarities in the responses to the questions. Starting the qualitative analysis with question 1.0 ("Q_1.0:Does your city incorporate sustainability goals and targets (e.g. GHG reductions) into the master plan for the city?"), all but one city answered "1: Yes" to it, as shown in table 4.11. Only Hollywood/FL indicates that the city has produced the plan ("2: In Progress").

Reinforcing the tendency to the application of sustainability-oriented plans, all cities responded "1: Yes" to questions "4.0:Does your city have a city-wide emissions inventory to report?" and "5.5:Does your city have a climate change mitigation or energy access plan for reducing city-wide GHG emissions?". This indicates that planning for emissions reductions using energy access plan can be used as a performance indicator when comparing different cities which are partners in emissions reductions initiatives like GCoM. The results are listed in table 4.11.

However, the data obtained during the implementation of the plans are not uniform. A reason is that even though these cities were selected by representing overall performance, their methodology varies based on many aspects not covered by the questions set. An example of it is the question "4.3:Please give the name of the primary protocol; standard; or methodology you have used to calculate your city's city-wide GHG emissions".

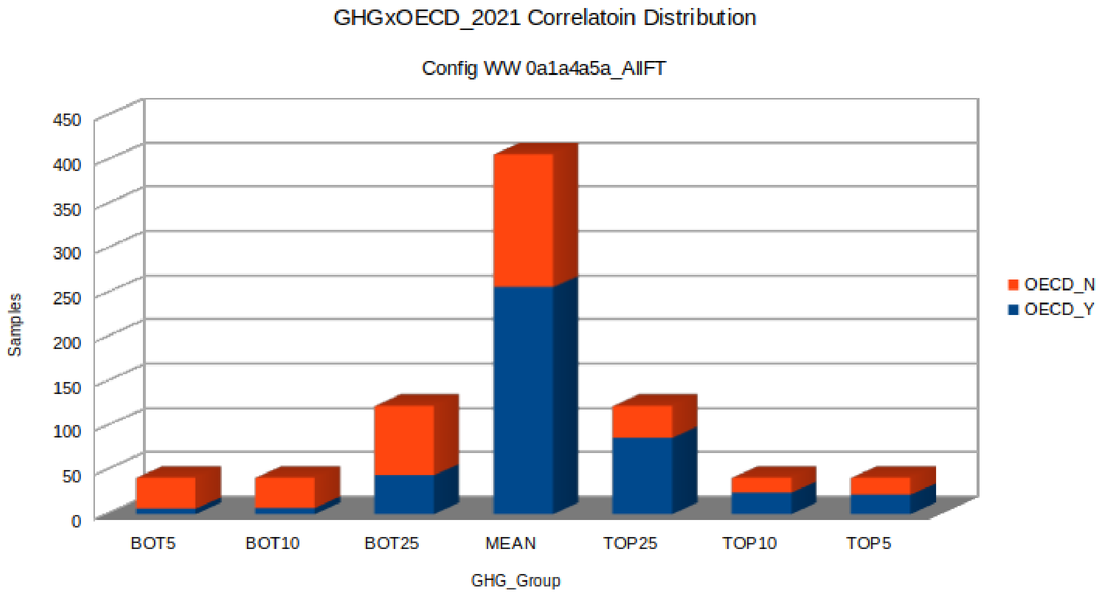

Figure 4.17: Cities total emisisons x OECD correlation distribution using configuration WW_0a1a4a5a_AllFT

Table 4.11: Selected samples clusters 8, 9, 13, 14, 15, 16 and 17 which are also present in GCoM database. The answers to the questions 1.0, 4.0 and 5.0 are shown.

CDP Id	**City Name**	**Cluster**	**Q_1.0**	**Q_4.0**	**Q_5.0**
31051	Coventry	9	1	1	1
31177	Salt Lake City	9	1	1	1
32480	Adelaide	13	1	1	1
35993	Singapore	14	1	1	1
43938	Dubai	15	1	1	1
49334	Richmond/VA	8	1	1	1
53959	Fayetteville/AR	8	1	1	1
54082	Hollywood/FL	16	2	1	1
54517	Örebro	17	1	1	1
57616	Lake Forest/IL	8	1	1	1

Most of the cities use "4: U.S. Community Protocol for Accounting and Reporting of Greenhouse Gas Emissions (ICLEI)", which indicates the adherence to United States regulations by cities in the U.S. Options "5:Regional or country-specific methodology" and "6:City specific methodology" are also implemented. However, only two cities used global initiatives protocol like "1:Global Protocol for Community Greenhouse Gas Emissions Inventories (GPC)". The results are listed in table 4.12.

Although the planning over energy aspects provided at the city level can be used as a performance indicator for emissions reduction, the distribution of cities through CDP regions is uneven as it is the distribution of emissions production. These can be viewed in Figures 4.21 and 4.22. The region distribution indicates a higher number of Latin American cities in cluster 1, indicating high participation

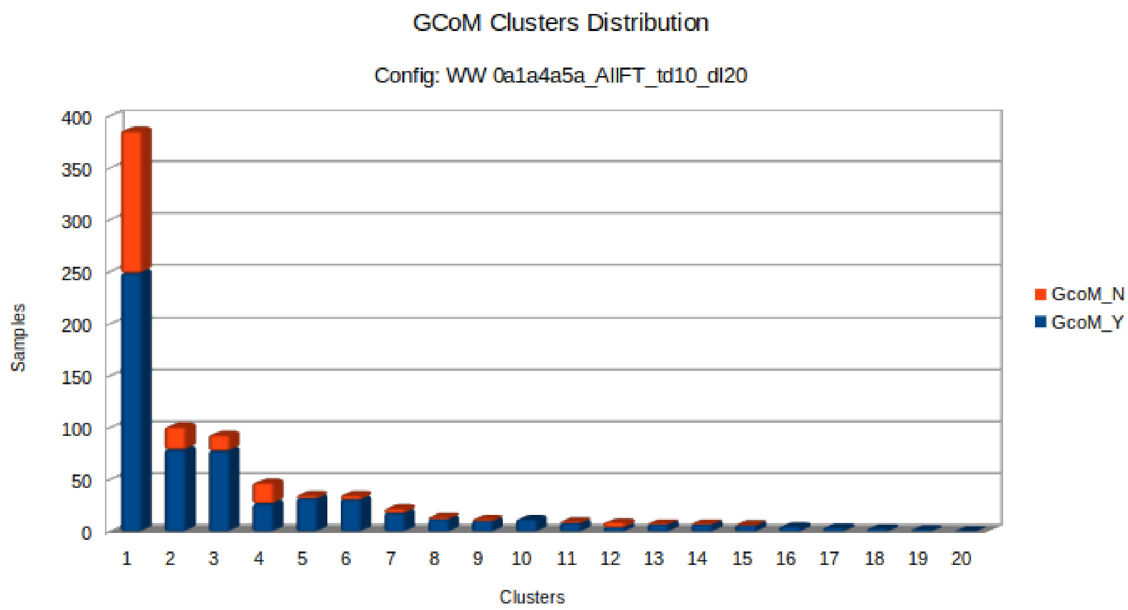

Figure 4.18: GCoM clusters distribution using configuration WW_0a1a4a5a_-AllFT.

Table 4.12: Selected samples clusters 8, 9, 13, 14, 15, 16 and 17 which are also present in GCoM database. The answers to the Question 4.3 are shown.

CDP Id	City Name	Cluster	Q_43
31051	Coventry	9	5
31177	Salt Lake City	9	4
32480	Adelaide	13	1
35993	Singapore	14	3
43938	Dubai	15	6
49334	Richmond/VA	8	4
53959	Fayetteville/AR	8	4
54082	Hollywood/FL	16	4
54517	Örebro	17	5
57616	Lake Forest/IL	8	1

of Brazil's cities in the GCoM database and with similar responses, reinforcing the hypothesis of different levels of development of tools to face emissions reduction challenges. However, the GCoM database confirms the concentration of cities with high levels of emissions generation. The presence of TOP5, TOP10 and TOP25 groups in cluster 1 is a clear indication of that.

4.2.4 PI: C40 Membership

The C40 is a coalition of cities to share experiences dealing with climate challenges like emissions produced by the cities. The list of participant cities in the C40 organisation in 2019 is available in the appendix. The cluster distribution of C40 members is shown in Figure 4.23.

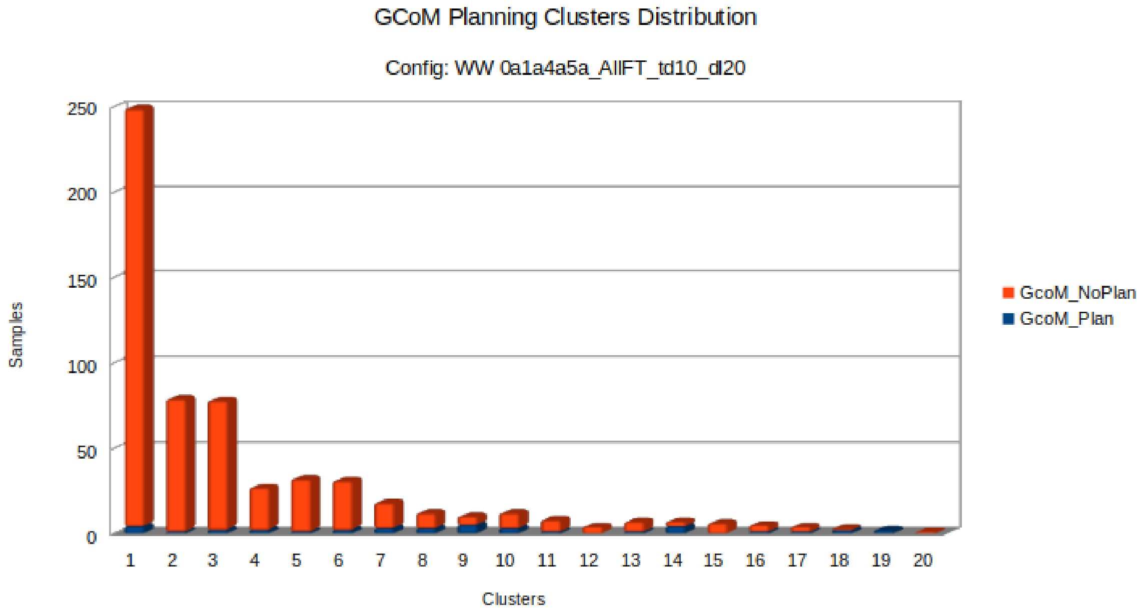

Figure 4.19: GCoM planning clusters distribution using configuration WW_-0a1a4a5a_AllFT.

Of 98 cities that participate in the C40 consortium, 76 is also in the CDP initiative. Looking at cluster 8, which 81% of samples are part of C40, and clusters 17, 18, 19 and 20, with 100% C40 members, they were selected and the distribution of the questions 1.0, 4.0 and 5.0 was verified. The distribution is shown in table 4.13 for comparison.

Table 4.13: Selected samples clusters 8, 17, 18, 19 and 20 which are also in C40 cities database. The answers frequencies to the questions 1.0, 4.0 and 5.0. are shown.

Question	**Positive Answers**	**Negative Answers**
1.0: Does your city incorporate sustainability goals and targets (e.g. GHG reductions) into the master planning for the city?	97.37%	2.63%
4.0: Does your city have a city-wide emissions inventory to report	94.74%	5.26%
5.0: Do you have a GHG emissions reduction target in place at the city-wide level?	80.26%	19.74%

Although quite all C40 cities (97.37%) have sustainability goals in place, almost 20% of them does not have a target of GHG reduction being executed. Considering that C40 cities have similar SHDI on average (821) as cities in CDP (826), the SHDI level cannot explain the lack of emissions reduction by part of the C40 cities. The distribution of clusters along 19 slots indicates instability in deciding which cluster is the best representative of the data set provided by the city. This behaviour is shown in Figure 4.24.

On the other side, the cluster distribution by regions has a better segregation as it is shown in Figure 4.24. The most populated clusters in this case (1, 2 and 3)

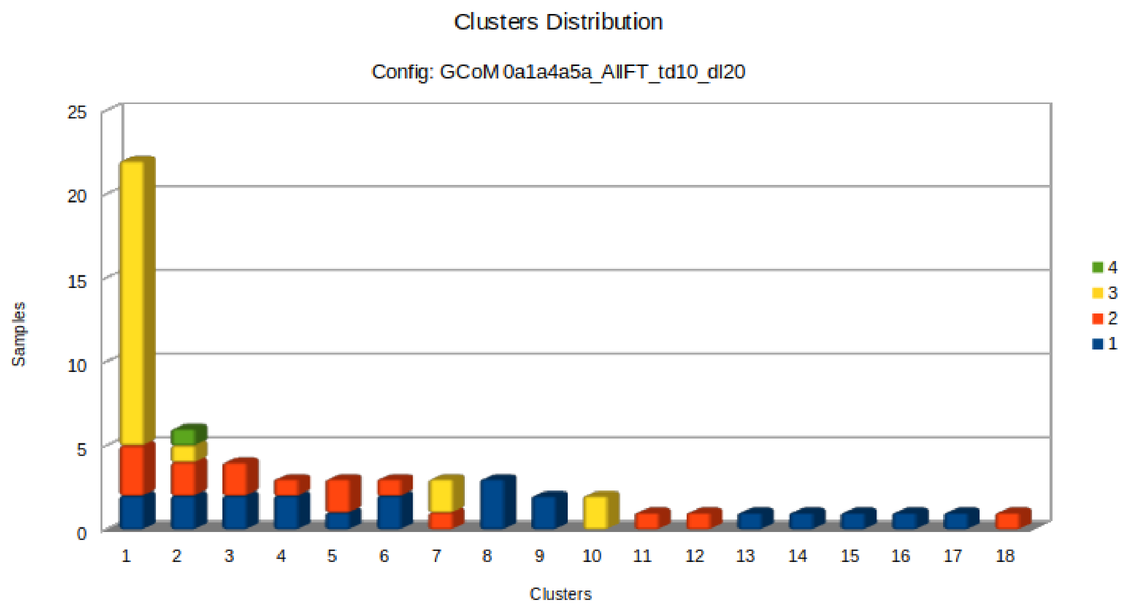

Figure 4.20: GCoM selected samples clusters distribution using configuration WW_0a1a4a5a_AllFT.

has only the EURO region in all of them. The other regions are distributed spatially along with the clusters. It reinforces even more differences in cities' responses from different regions, although the efforts in considering C40 a global initiative to promote emissions reduction.

The concentration of higher emissions rates is another characteristic of C40 cities. In both correlation and samples distribution view, as shown in Figures 4.26 and 4.27, the indication of TOP25, TOP10 and TOP5 emissions is evident and point to the right choice for picking high emissions cities but have conditions to be organised around emission reduction goals.

4.2.5 PI: Smart Cities Index Membership

A consortium of international organisations maintains the Smart Cities Index (SCI) to rank cities in terms of developing connectivity to services provided by the cities. The list of cities in SCI in 2019 is available in the appendix. The cluster distribution of SCI members is shown in Figure 4.28.

The SCI 2019 had 111 cities, and 63 of them are part of the CDP initiative. Even though it is not expected that all SCI cities have any direct goal related to emissions reduction, selecting the ones that are part of the CDP initiative introduces a bias that needs to be considered during the analysis. The clusters with SCI minimal prevalence were 1, 7, 9, 16, 18, 19, and 20, and the frequency of positive answers is similar to those of C40, although the execution of the actions is better. The distribution is shown in table 4.14 for comparison.

The distribution of the answers is a little more stable in this case. Still, 10%

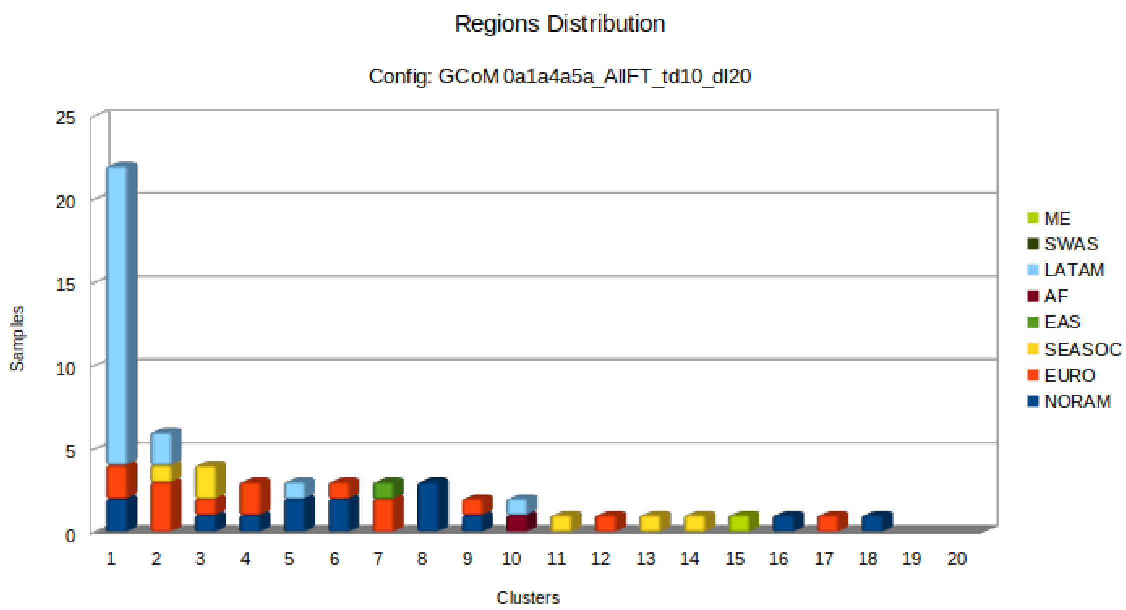

Figure 4.21: GCoM selected samples regions distribution using configuration WW_-0a1a4a5a_AllFT.

Table 4.14: Selected samples clusters 1, 7, 9, 16, 18, 19 and 20 which are also in SCI database. The answers frequencies to the questions 1.0, 4.0 and 5.0 are shown.

Question	**Positive Answers**	**Negative Answers**
1.0: Does your city incorporate sustainability goals and targets (e.g. GHG reductions) into the master planning for the city?	93.65%	6.35%
4.0: Does your city have a city-wide emissions inventory to report	96.83%	3.17%
5.0: Do you have a GHG emissions reduction target in place at the city-wide level?	90.84%	9.52%

of the selected cities does not have a GHG emissions reduction target in place. In table 4.15 is a list of cities that have at least one negative answer. In clusters 1 and 2, we have situations that represent errors during data processing or cities entered; the value "0" indicates that no information was provided. However, in clusters 4 and 7, there are some indications of "planning and not executing", confirming some cities have problems running the last mile and, for this, not getting the full benefits from emissions reduction initiatives.

The average of SHDI for cities in SCI is very high (879) and it is inclined to some uniformity in answers for at least 90% of them. It can be seen in distributions by clusters and by regions in Figures 4.29 and 4.30. The GHG emissions distribution tends to be more elevated in SCI also then average CDP cities, as shown in Figure 4.31.

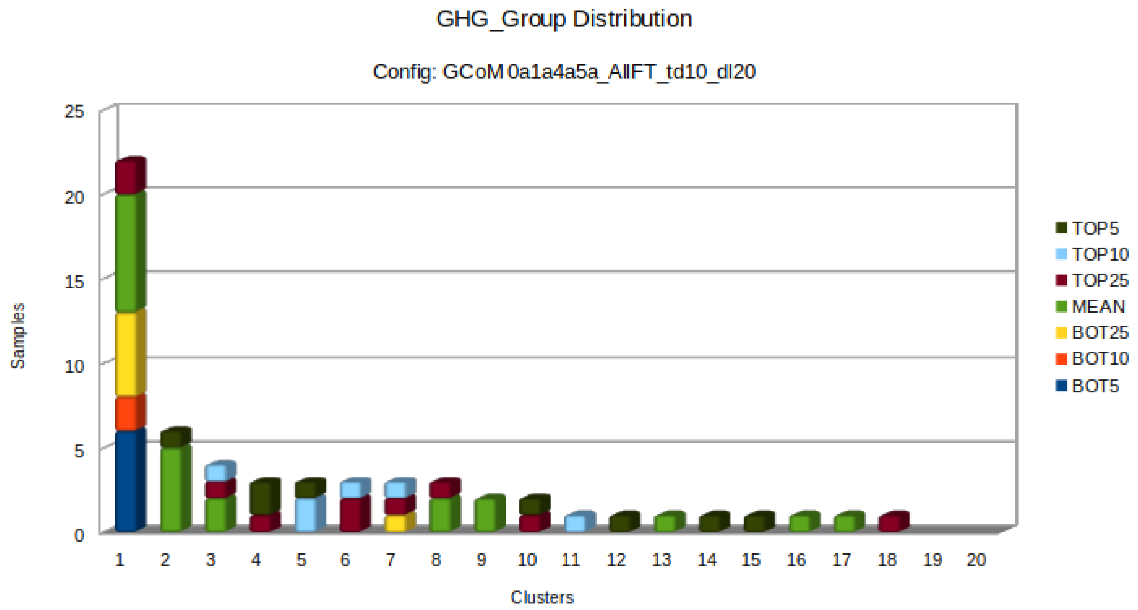

Figure 4.22: GCoM selected samples GHG_Group distribution using configuration WW_0a1a4a5a_AllFT.

Table 4.15: Selected samples clusters 1, 2, 4, 7 from SCI samples featuring questions 1.0, 4.0 and 5.0 and SHDI.

CDP Id	Cluster	City Name	Country	Q_1.0	Q_4.0	Q_5.0	SHDI
31167	7	Lagos	Nigeria	1	1	5	673
31171	7	Madrid	Spain	4	1	1	928
31180	4	Santiago	Chile	1	1	5	845
35885	7	Tel Aviv-Yafo	Israel	3	1	5	906
35913	7	Nairobi	Kenya	1	3	5	644
51075	1	Shenzhen	China	1	4	3	791
54291	1	Chengdu	China	1	0	0	716
54306	1	Medan	Indonesia	0	3	1	718
54457	1	Hamburg	Germany	0	1	4	975
59595	2	Brisbane/CA	United States	0	1	1	930
826237	4	Madrid	Colombia	1	1	5	767

4.3 KPIs developed from the Emissions Reporting Maturity Model

4.3.1 KPI: Emissions Reporting Maturity Level

The emissions reporting maturity level (ERM-L) can be used to measure the overall capability of a city to select, process and deliver information about emissions in both city-wide and city-administration scopes. The ERM-L can vary from 0 to 5, as established in the emissions reporting maturity model (ERMM). The processes defined in ERMM were evaluated based on the data provided by the cities to obtain the ERM-L. The processing results for the 814 cities in the CDP database are

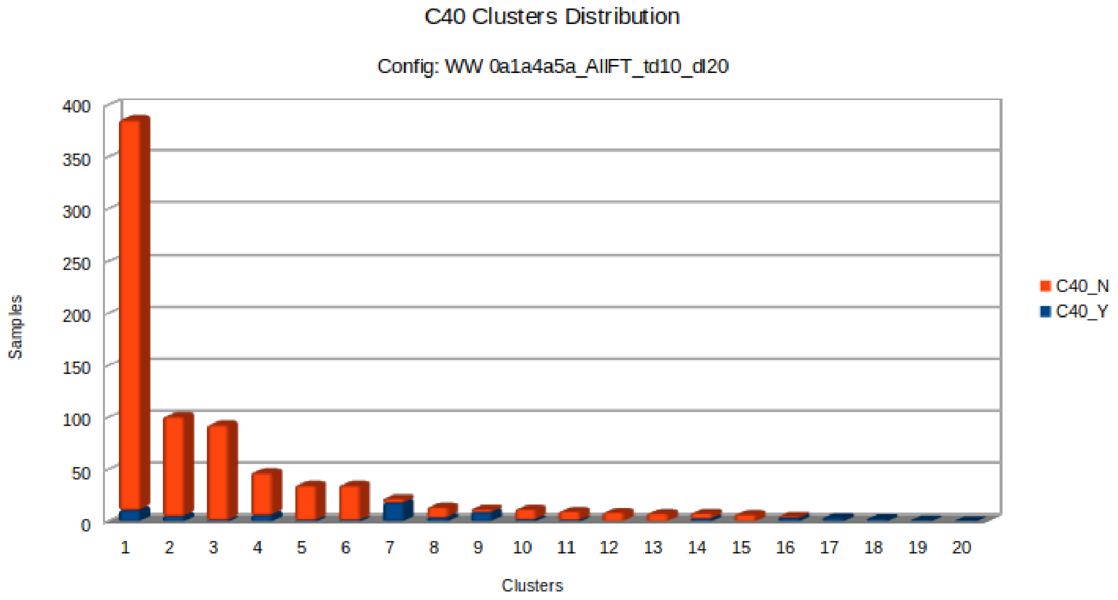

Figure 4.23: C40 clusters distribution using configuration WW_0a1a4a5a_AllFT.

presented in the appendix of this work. The table 4.16 shows the ERM-L for some cities in Brazil. The PI values for the processes are also shown: data modelling, data acquisition, data processing, data analysis, build, publish, deployment, monitoring.

One of the processes evaluated to obtain the ERM-L is data acquisition. One of the practices evaluated is the quality of answering from those cities. The distribution of the quality indicator (IND) is shown in Figure 4.32

Analysing the clusters distribution and quality indicator labels, clusters 5 and 7 do not have any samples in the best 10% in terms of answering quality. It indicates the uneven balance between the answers provided by the cities and the quality of the answering process.

4.3.2 KPI: Emissions Reporting Maturity Level by Regions

The findings obtained from the execution of the EMM-L process over the cities in the CDP database indicate differences when using the CDP region information as a filter. Further experiments executed with other region-based distributions (Country, e.g.) show similar behaviour in clusters distribution. The region type attributes can interfere in the level of achievement of the processes and the evaluation of some capabilities. Thus, to achieve better results with ERMM, it is essential to consider region alike attributes, even to use them to the obtained results from the method.

The distributions of quality indicator (IND) are shown in Figures 4.33 and 4.34, clearly indicating the differences between the quality of the answers and the CDP regions, taking into consideration the clusters distribution of the answers from the cities.

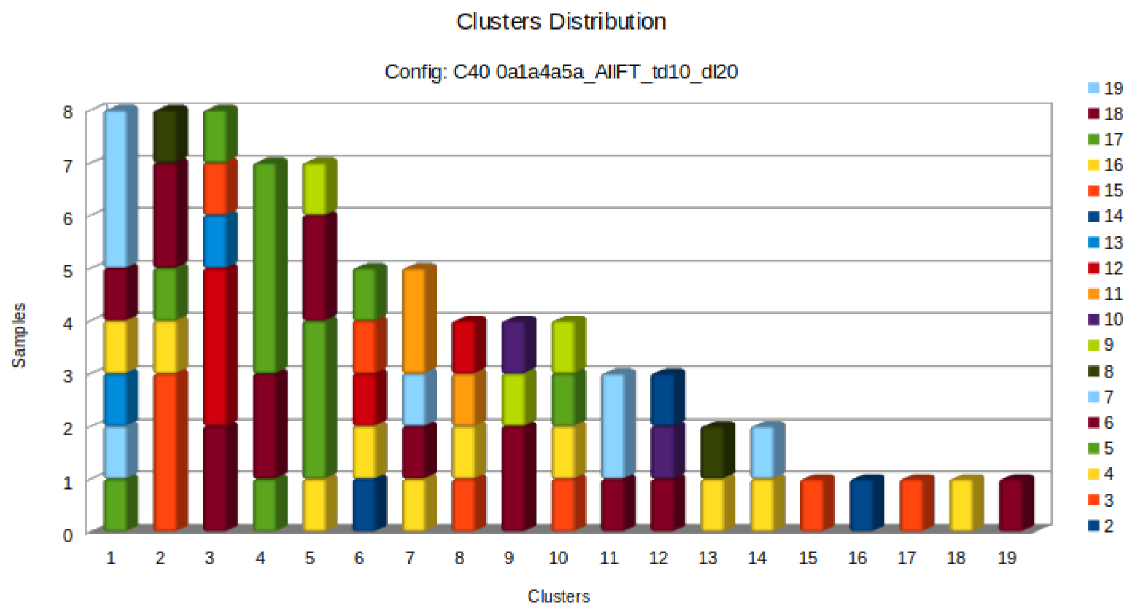

Figure 4.24: C40 samples clusters distribution using configuration WW_0a1a4a5a_-AllFT.

4.4 Other Observations

During the experiments, some analyses demonstrated cities' behaviours that should be objects in future studies. Among some examples of this are the few cities from OPEC countries, the differences in the distribution of clusters for G7 and G20 cities, and the economic bias encountered in the metrics used in some cities rank systems. The results of these analyses are detailed in the following sections.

4.4.1 OPEC cities participation

The organisation of the Petroleum Exporting Countries (OPEC) was founded to coordinate and unify the petroleum production policies of member countries. Some experiments presented in this work used global memberships as OPEC to guide some analyses and validate the results. Only 4 in the 814 cities belong to countries members of OPEC. The distribution of clusters is shown in Figure 4.35.

The cities are distributed along with the clusters 1, 7, and 20, indicating that even participating in an organisation like OPEC, these cities answered CDP questions differently.

4.4.2 G20 and G7 comparison

The G7 is the group of the most developed countries (Canada, France, Germany, Italy, Japan, United Kingdom, United States), representing 30.84% of the CDP

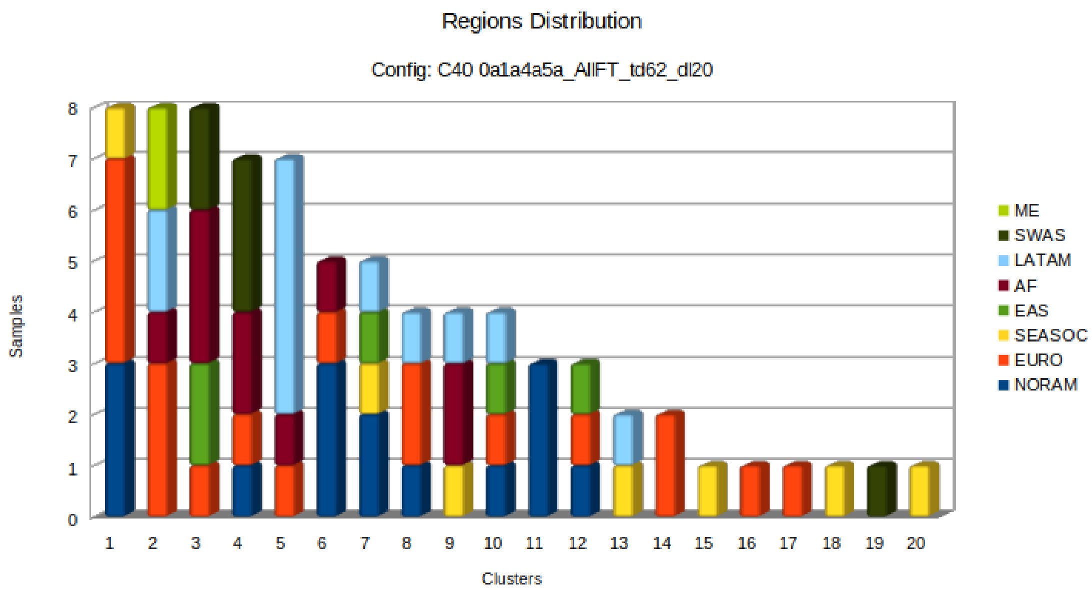

Figure 4.25: C40 samples regions distribution using configuration WW_0a1a4a5a_-AllFT.

database. Adding the group of developed and emerging countries (Argentina, Australia, Brazil, China, India, Indonesia, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey), this rate rises to 65.11%. The cluster distributions are shown in Figures 4.36 and 4.37. The distributions in some clusters represent the variation of answers provided by the cities and, for instance, the variation in the levels of development of the cities.

4.4.3 Ranks defined by short-term economic goals

Analysing the additional (external) data used to validate the results from the experiments, some aspects in the questions related to clarity, quality and application, show some characteristics that indicate short-term economic challenges only. For example, even in the CDP database, some questions have the option "2: Yes, in 2 years", but we cannot find any other question pointing to mid or long-term goals.

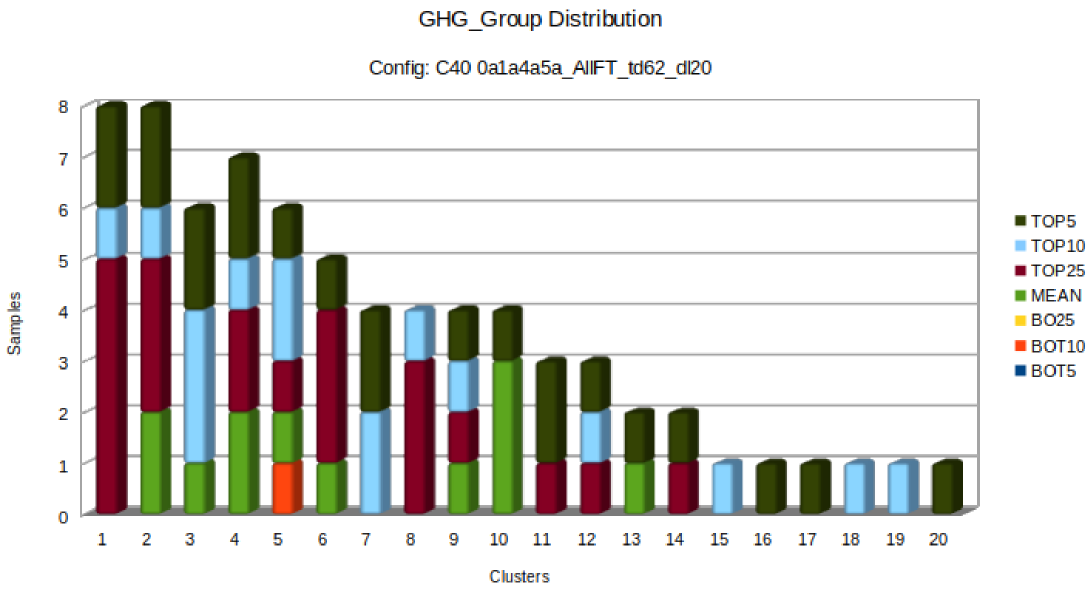

Figure 4.26: C40 samples GHG_Group distribution using configuration WW_-0a1a4a5a_AllFT.

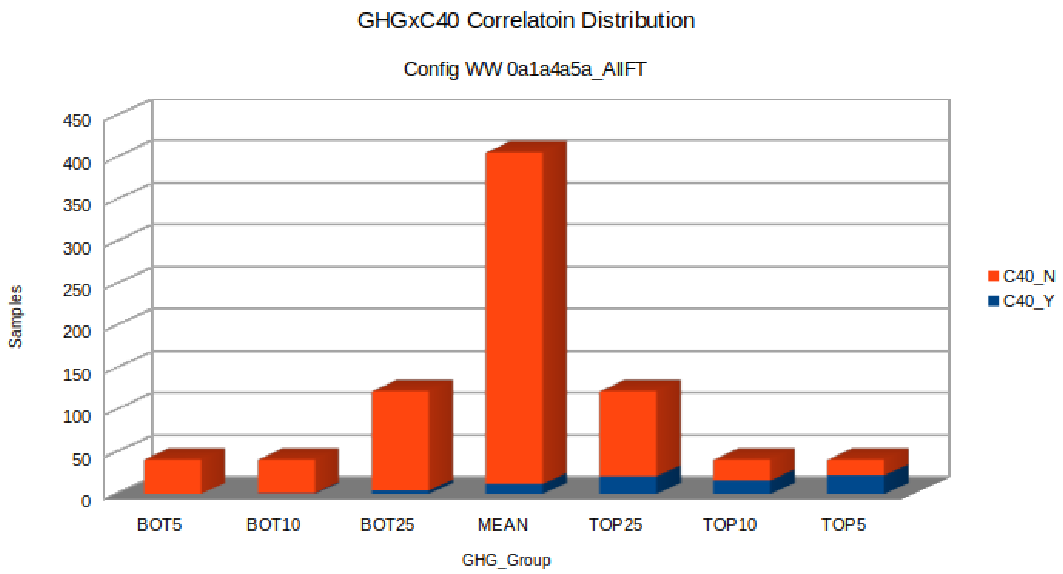

Figure 4.27: GHGxC40 correlation using configuration WW_0a1a4a5a_AllFT.

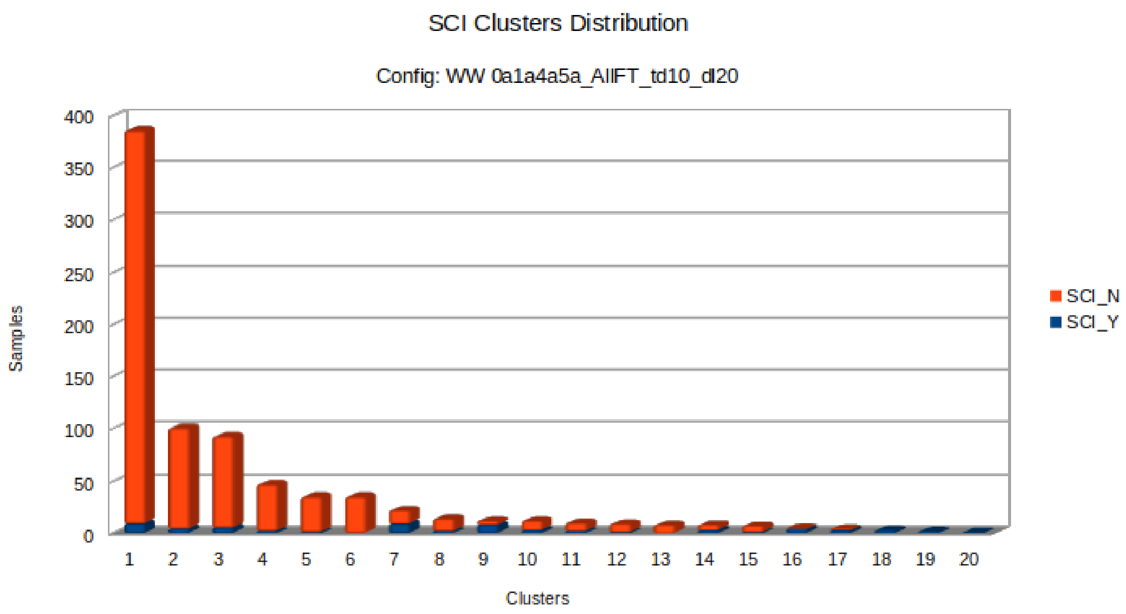

Figure 4.28: SCI clusters distribution using configuration WW_0a1a4a5a_AllFT.

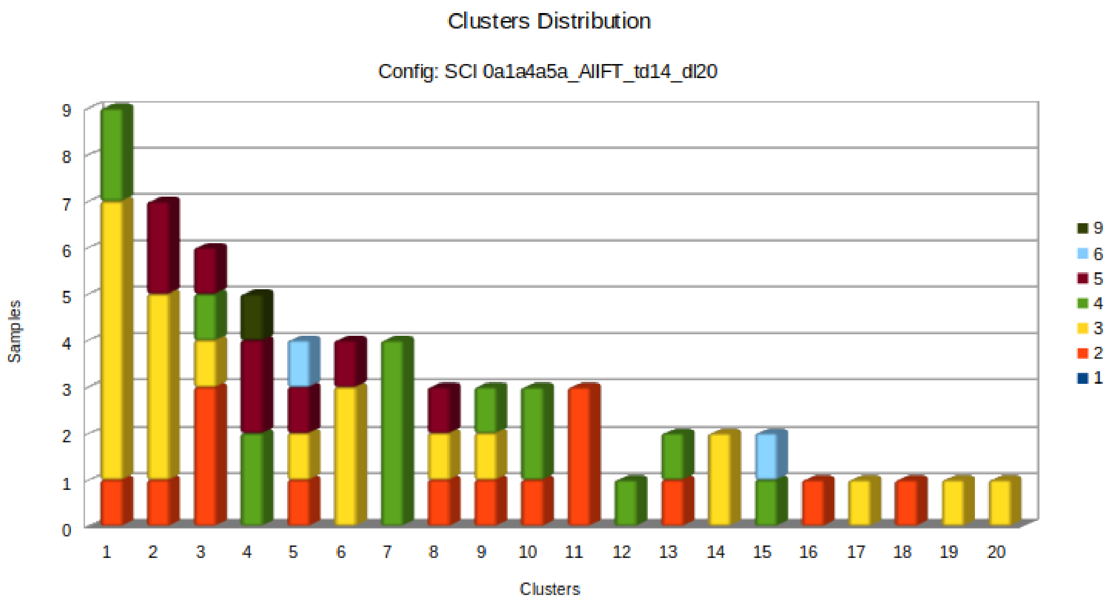

Figure 4.29: SCI samples clusters distribution using configuration WW_0a1a4a5a_-AllFT.

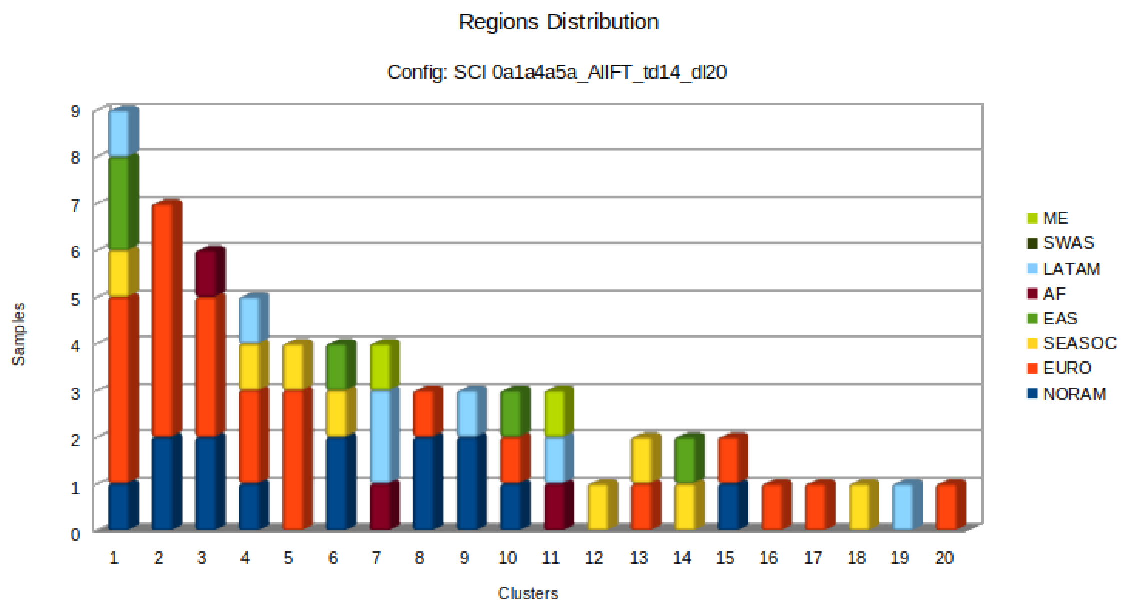

Figure 4.30: SCI samples regions distribution using configuration WW_0a1a4a5a_-AllFT.

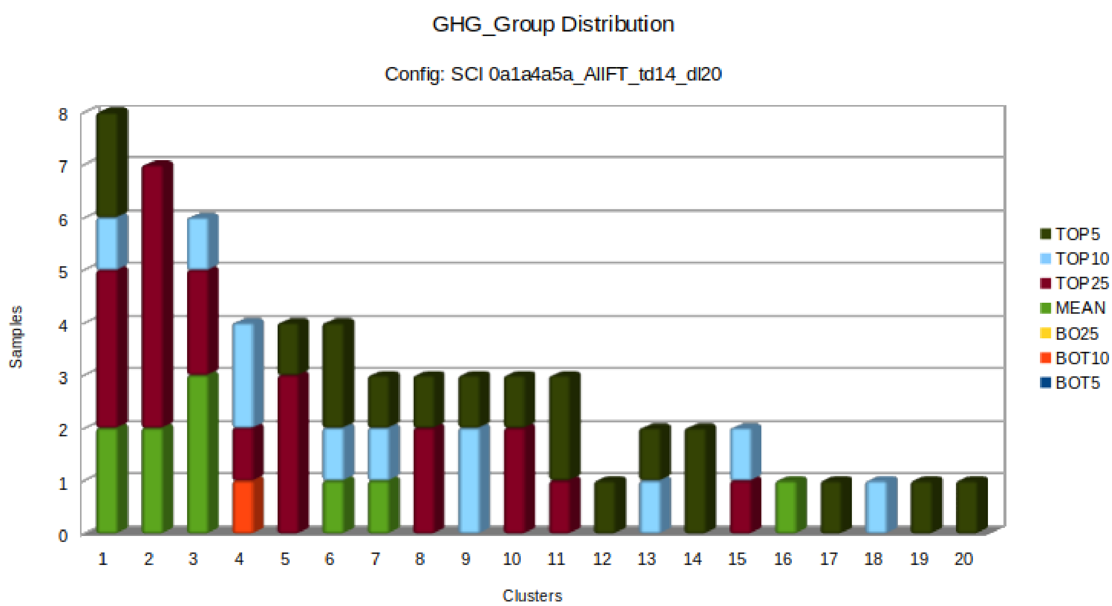

Figure 4.31: SCI samples GHG_Group distribution using configuration WW_-0a1a4a5a_AllFT.

Table 4.16: ERM-L method execution for Brazil cities.

CDP Id	City Name	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
31156	Curitiba	1	1	110	0	1	0	0	1	1
31176	Rio de Janeiro	3	1	111	1	4	1	0	1	1
31184	São Paulo	2	1	101	1	1	1	0	1	1
35848	Belo Horizonte	1	1	100	0	2	1	0	1	1
35865	Fortaleza	1	1	100	0	1	1	0	1	1
35872	Recife	0	0	100	0	2	0	0	1	1
35880	Porto Alegre	2	1	100	1	1	0	0	1	0
35897	Campinas	3	1	100	1	2	0	0	1	1
36041	Belém	0	1	000	1	1	0	0	1	0
42120	Salvador	1	1	110	0	1	1	0	1	1
42123	Goiânia	2	1	100	1	1	1	0	1	0

Note: ERM-L values can vary from 0 to 5. The range values for the performance indicators are: Data Modeling (0-1); Data Acquisition (0-1) in each sub-item; Data Processing (0-1); Data Analysis (0-5); Report Building (0-1); Report Publishing (0-1); Deployment (0-1); Monitoring (0-1)

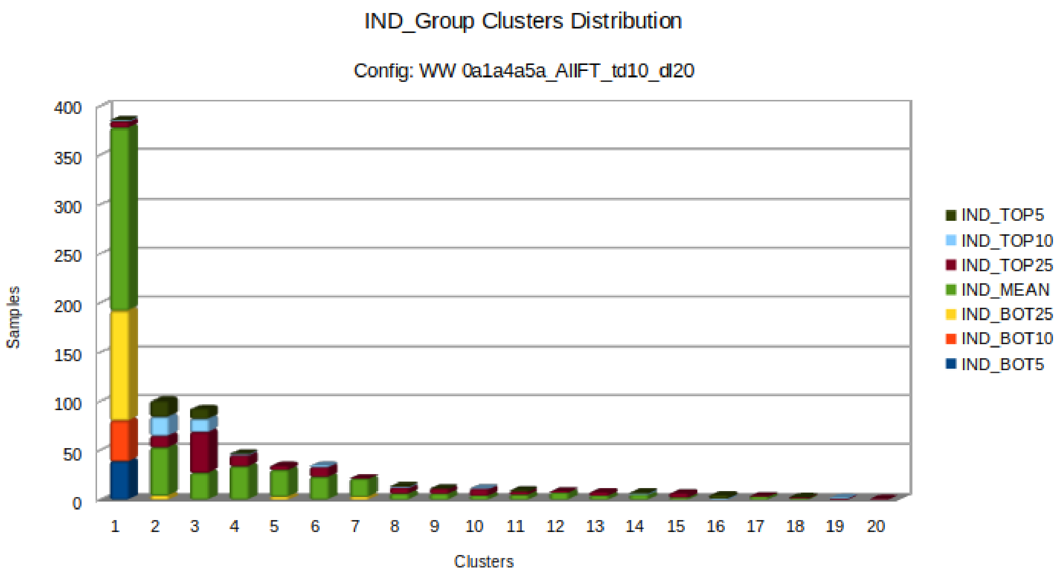

Figure 4.32: Quality indicator distribution using configuration WW_0a1a4a5a_-AllFT.

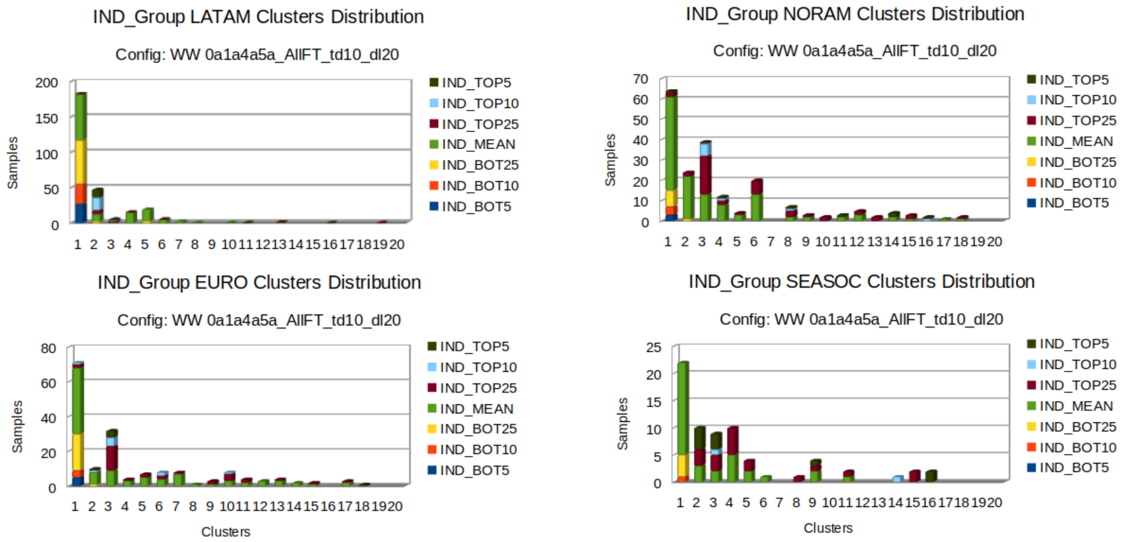

Figure 4.33: Quality indicator distribution using configuration WW_0a1a4a5a_-AllFT for regions NORAM, LATAM, EURO, SEASOC.

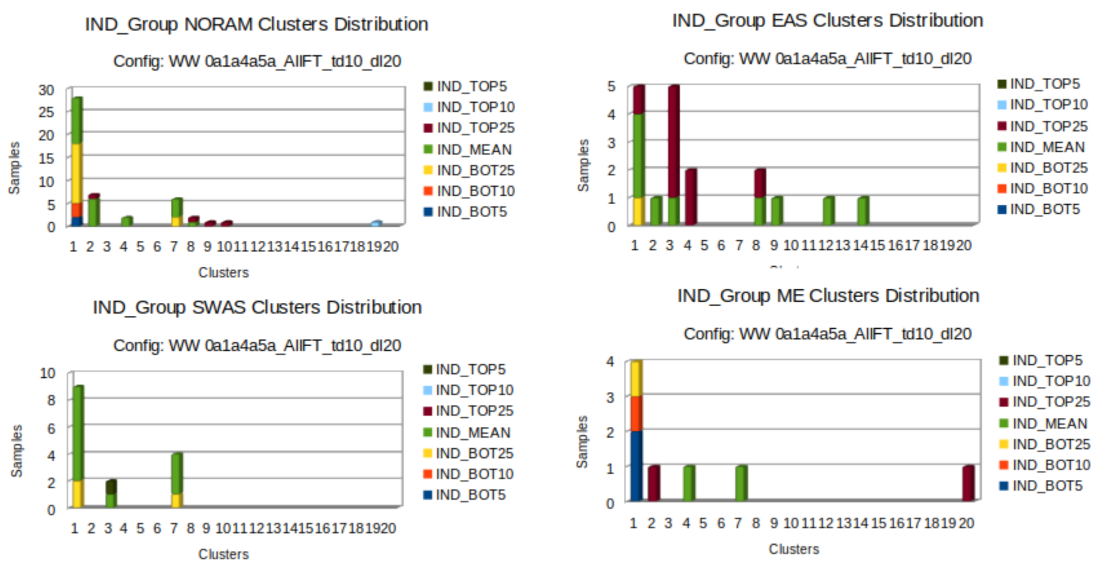

Figure 4.34: Quality indicator distribution using configuration WW_0a1a4a5a_-AllFT for regions AF, SWAS, EAS, ME.

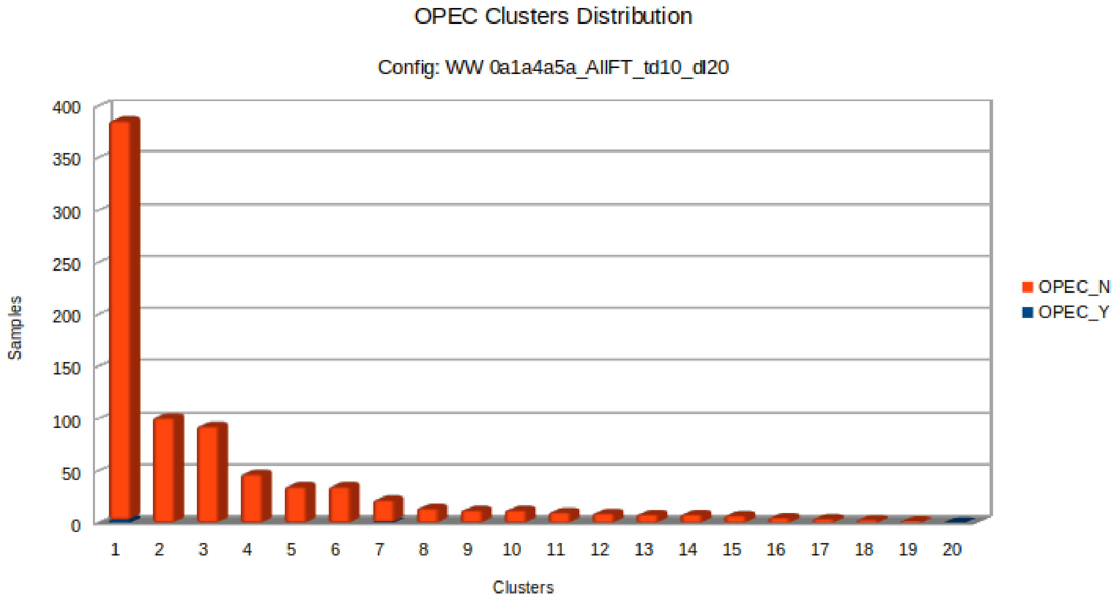

Figure 4.35: OPEC clusters distribution using configuration WW_0a1a4a5a_-AllFT.

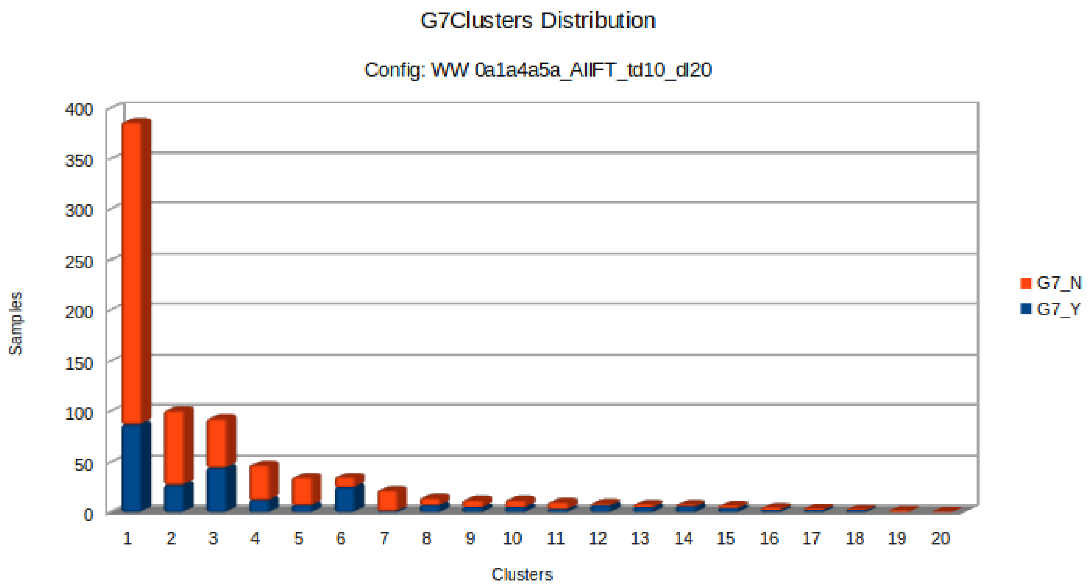

Figure 4.36: G7 clusters distribution using configuration WW_0a1a4a5a_AllFT.

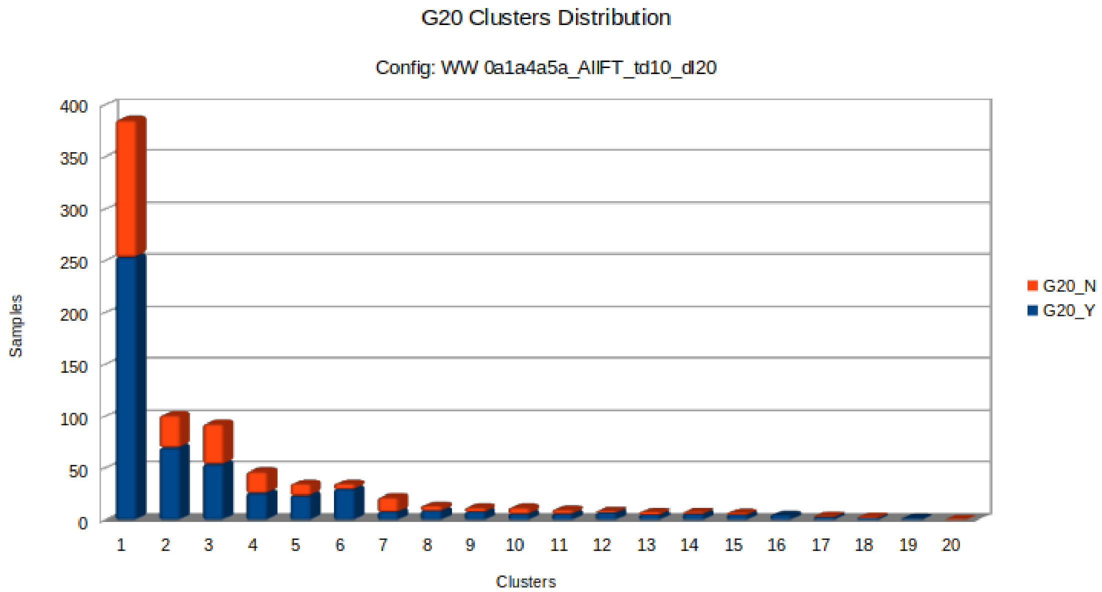

Figure 4.37: G20 clusters distribution using configuration WW_0a1a4a5a_AllFT.

Chapter 5

Conclusion and future works

This session shows the main findings of this study and their impacts. Emissions reporting empowerment is one of the highlights of this work to help leverage the overall capacity of the cities to deal with emissions reduction issues and challenges. For example, the analysed cities struggle to convert emissions reporting information into actionable processes to enforce emissions reduction policies. This work points to the lack of reliable information or efficient means to correctly inform emissions facts along the decision chain as the leading cause. It also occurs when comparing the data from databases provided by cities consortia and memberships like GCoM and C40 with the disclosure data provided by the cities in the CDP database. Another issue is the absence of patterns for exchanging information about emissions-related data among the major databases, such as electronic data interchange (EDI).

The performance indicators development process (PIDP), which is an important contribution of this work, searches for PIs among the analysed data. For example, some correlations of the emissions reporting data and external indicators and indexes can establish the basis for PIs. However, the search for PIs looking into patterns for the answers provided by the cities failed. The main reason was the lack of quality in the data made available. Therefore, a qualitative analysis was made based on the data produced by the clustering iterations. These analyses indicated a gap between the responses provided by the cities and the related indicators used to show emissions levels, impacts, and mitigation policies. It happened due to the low-reliability level of the information found within the sample data analysed.

However, the analyses promoted in the scope of PIDP over the data could expose the inefficiencies found in the emissions reporting processes. For example, consistency errors in the forms and between the information reported and external sources were constantly found in the majority of the cities. The motives for this are not established, but based on the diversity of the cities analysed that showed these difficulties, the lack of standardisation and effectiveness of the emissions reporting processes can explain that.

Thus, this work proposes the emissions reporting maturity model (ERRM) as the main contribution. This model can be used to leverage the emissions reporting processes efficiency and, by doing this, to achieve better results in the emissions reduction policies implementation. In this case, the PIDP results can be used to guide the survey of the processes subject to the ERMM: a city that aims to build an ERMM should apply a survey over the processes owned by the areas that deal with emissions. In this processes survey, the main goal is to identify processes impacted by or executed by emissions reduction initiatives. The survey maps processes, related goals of each process and the practices exercised by them. Thus, performance indicators are defined to gauge the impact of the implementation of these processes. Nevertheless, it is expected of a maturity model to have improvements over time, mainly because its effectiveness is tightly related to its application.

The findings of this work also suggest the need to investigate if the reporting issues associated with the emissions policies in the cities apply to other areas of interest: energy, transportation, employment are some areas that can benefit from a reporting maturity model. The ERMM is flexible enough to embrace these other areas and their challenges. The mapped processes, goals, practices and capabilities can transcend the challenges specific to each area of interest.

Another possible future contribution is to use the ERMM to help design an AI-based helper system toward e-government full implementation. The ERMM can map the processes that use "Internet of things" (IoT) to provide reliable information about emissions. Furthermore, the ERMM can use AI to search for patterns, best performance cases, successfully applied policies and social and economic return over investment (ROI). Finally, the evolutive aspect of ERMM is an advantage to the cities to adopt and share expertise in emissions reduction policies.

References

[1] BROADSTOCK, D. C., COLLINS, A., HUNT, L. C., et al. "Voluntary disclosure, greenhouse gas emissions and business performance: Assessing the first decade of reporting", *The British Accounting Review*, v. 50, pp. 48–59, fev. 2018.

[2] GIEST, S. "Big data analytics for mitigating carbon emissions in smart cities: opportunities and challenges", *European Planning Studies*, v. 25, n. 6, pp. 941–957, fev. 2017.

[3] TIWARI, K., KHAN, M. S. "Sustainability accounting and reporting in the industry 4.0", *Journal of Cleaner Production*, v. 258, n. 120783, jun. 2020.

[4] ZHANG, L., ZHOU, P. "A non-compensatory composite indicator approach to assessing low-carbon performance", *European Journal of Operational Research*, 2018.

[5] BADAWY, M., EL-AZIZ, A. A., IDRESS, A. M., et al. "A survey on exploring key performance indicators", *Future Computing and Informatics Journal*, v. 1, pp. 47–52, abr. 2016.

[6] PARMENTER, D. *Key performance indicators: developing, implementing, and using winning KPIs.* 1 ed. New York, John Wiley & Sons, 2010.

[7] ECKERSON, W. W. "Performance management strategies", *Business Intelligence Journal*, v. 14, n. 1, pp. 24–37, 2009.

[8] PAULK, M. C., CURTIS, B., CHRISSIS, M. B., et al. "Capability maturity model, version 1.1", *IEEE Software*, v. 10, n. 4, pp. 18–27, jul. 1993.

[9] MONTEIRO, E. L., MACIEL, R. S. P. "Maturity Models Architecture: A large systematic mapping", *Revista Brasileira de Sistemas de Informação (Brazilian Journal of Information Systems)*, v. 13, n. 2, pp. 110–140, 2020.

[10] METTLER, T., ROHNER, P., WINTER, R. "Towards a Classification of Maturity Models in Information Systems", *Management of the Interconnected World*, pp. 333–340, 2010.

[11] CMMI. *CMMI for Development, Version 1.3*. CMU SEI-2010-TR-033, Capability Maturity Model Integration Institute, 2010.

[12] PAULK, M. C., CURTIS, B., CHRISSIS, M. B., et al. "Capability maturity model, version 1.2", *IEEE Software*, v. 10, n. 4, pp. 18–27, 2006.

[13] ISO/IEC. *Process assessment model for software life cycle processes*. ISO/IEC TS 33061, International Organization for Standardization, 2021.

[14] MCCULLOCH, W., PITTS, W. "A logical calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical Biophysics*, v. 1, n. 7, pp. 115–133, 1943.

[15] ALEKSANDER, I., GREGORIO, M. D., FRANÇA, F., et al. "A brief introduction to Weightless Neural Systems", *ESANN'2009 proceedings, European Symposium on Artificial Neural Networks - Advances in Computational Intelligence and Learning*, pp. 22–24, abr. 2009.

[16] GREGORIO, M. D., GIORDANO, M. "An experimental evaluation of weightless neural networks for multi-class classification", *Applied Soft Computing Journal*, 2018.

[17] BLEDSOE, W. W., BROWNING, I. *Pattern recognition and reading by machine*. In: Papers presented at the eastern joint ire-aiee-acm computer conference, IRE-AIEE-ACM '59 (Eastern), New York, NY, USA, 1959.

[18] ALEKSANDER, I., THOMAS, W., BOWDEN, P. "WISARD, a radical new step forward in image recognition", *Sensor Rev.*, v. 4, n. 3, pp. 120–124, 1984.

[19] LUSQUINO FILHO, L. A. D., OLIVEIRA, L. F. R., CARNEIRO, H. C. C., et al. "A weightless regression system for predicting multi-modal empathy", *IEEE International Conference on Automatic and Gesture Face Recognition*, v. 15, pp. 657–661, 2020.

[20] CARDOSO, D. O., CARVALHO, D. S., ALVES, D. S., et al. "Financial credit analysis via a clustering weightless neural classifier", *Douglas O. Cardoso and Danilo S. Carvalho and Daniel S.F. Alves and Diego F.P. Souza and Hugo C.C. Carneiro and Carlos E. Pedreira and Priscila M.V. Lima and Felipe M.G. França*, v. Neurocomputing, n. 183, pp. 70–78, 2016.

[21] K.SASIREKHA, P.BABY. "Agglomerative Hierarchical Clustering Algorithm - A Review", *International Journal of Scientific and Research Publications*, v. 3, n. 3, pp. 1–3, mar. 2013.

[22] LIKAS, A., VLASSIS, N., VERBEEK, J. J. “The global k-means clustering algorithm”, *The Journal of Pattern Recognition Society*, v. 36, n. 1, pp. 451–461, mar. 2003.

[23] KUMAR, K. M., REDDY, A. R. M. “A fast DBSCAN clustering algorithm by accelerating neighbor searching using Groups method”, *The Journal of Pattern Recognition Society*, v. 58, n. 1, pp. 39–48, mar. 2016.

[24] LIBERTI, L., LAVOR, C., MACULAN, N., et al. “Euclidean Distance Geometry and Applications”, *SIAM Review*, v. 56, n. 1, pp. 1–64, maio 2012.

[25] CRESWELL, J. W. *QUALITATIVE INQUIRY & RESEARCH DESIGN Choosing Among Five Approaches*. 2 ed. Lincoln, USA, Sage Publications, Inc., 2007.

[26] CORBIN, J., STRAUSS, A. “Grounded Theory Research: Procedures, Canons, and Evaluative Criteria”, *Qualitative Sociology*, v. 13, n. 1, pp. 3–21, 1990.

[27] DENZIN, N. K., LINCOLN, Y. S. “The Discipline and Practice of Qualitative Research”. In: *Sociology*, 2005.

[28] GLASER, B., STRAUSS, A. *The Discovery of Grounded Theory Strategies for Qualitative Research*. 1 ed. New Brunswick (U.S.A.) and London (U.K.), Aldine Transaction, 1967.

[29] CHARMAZ, K. *Constructing Grounded Theory - A Practical Guide Through Qualitative Analysis*. 1 ed. London (U.K.), SAGE Publications, 2006.

[30] STAKE, R. E. “Qualitative case studies”. In: *The Sage handbook of qualitative research*, 3 ed., pp. 443–466, Thousand Oaks, CA, USA, Sage Publications, Inc., 2005.

[31] MERRIAM, S. B. *Qualitative Research and Case Study Applications in Education*. 1 ed. San Francisco, USA, Jossey-Bass Publishers, 1998.

[32] YIN, R. K. *Case Study Research: Design And Methods*. 1 ed. San Francisco, USA, SAGE Publications, 2003.

[33] PILLAIN, B., GEMECHU, E., SONNEMANN, G. “Identification of Key Sustainability Performance Indicators and related assessment methods for the carbon fiber recycling sector”, *Ecological Indicators*, v. 72, pp. 833–847, 2017.

[34] LOU, Y., JAYANTHA, W. M., SHEN, L., et al. "The application of low-carbon city (LCC) indicators - A comparison between academia and practice", *Sustainable Cities and Society*, v. 51, n. 101677, 2019.

[35] LIEBCHEN, G., SHEPPERD, M. "Data Sets and Data Quality in Software Engineering: Eight Years On". In: *Proceedings of the The 12th International Conference on Predictive Models and Data Analytics in Software Engineering*, v. 122, *PROMISE 2016*, Association for Computing Machinery, pp. 7:1–7:4, New York, NY, USA, 2016.

[36] BOOCH, G., RUMBAUGH, J., JACOBSON, I. *Unified Modeling Language User Guide, The, 2nd Edition*. 2 ed. Hoboken, NJ, USA, Addison-Wesley Professional, 2005.

[37] SHEPPERD, M., SONG, Q., SUN, Z., et al. "Data Quality: Some Comments on the NASA Software Defect Datasets", *IEEE TRANSACTIONS ON SOFTWARE ENGINEERING*, v. 39, n. 9, pp. 1208–1215, set. 2013.

[38] ZHAO, C., SAHNI, S. "String correction using the Damerau-Levenshtein distance". In: *7th IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS 2017)*, pp. 20–46, Orlando, FL, USA, out. 2017.

Appendix A

Wordcloud

Figure A.1: Wordcloud image built with the words presented in this work.

Appendix B

Fields List in Configuration 0a1a4a5a

0.1:Please give a general description and introduction to your city including your city's reporting boundary in the table below.

0.1_1:Administrative boundary

0.1_1[01]:Local government area within a city / metropolitan area

0.1_1[02]:City / Municipality

0.1_1[03]:Independent city

0.1_1[04]:Special city

0.1_1[05]:Federal district

0.1_1[06]:Sovereign city-state

0.1_1[07]:Metropolitan area

0.1_1[08]:Province / County

0.1_1[09]:Independent province

0.1_1[10]:Intercommunality

0.1_1[11]:Sub-municipal district

0.1_1[12]:Other, please specif

0.1_2:Description of city

0.2:If you have not previously submitted your Letter of Commitment to the Global Covenant of Mayors; either through the relevant regional covenant or through the Global Covenant secretariat; please attach the letter signed by an appropriately mandated official (e.g. Mayor; City Council) to this question. 0.3:Please provide information about your city's Mayor or equivalent legal representative authority in the table below:

0.3_1:Leader title

0.3_2:Leader name

0.3_3:Current term end month

0.3_3[01]:January

0.3_3[02]:February

0.3_3[03]:March

0.3_3[04]:April

0.3_3[05]:May

0.3_3[06]:June

0.3_3[07]:July

0.3_3[08]:August

0.3_3[09]:September

0.3_3[10]:October

0.3_3[11]:November

0.3_3[12]:December

0.3_4:Current term end year

0.4:Please select the currency used for all financial information disclosed throughout your response.

0.5:Please provide details of your city's current population. Report the population in the year of your reported inventory; if possible.

0.5_1:Current population

0.5_2:Current population year

0.5_3:Projected population

0.5_4:Projected population year

0.6:Please provide further details about the geography of your city.

0.6_1:Land area of the city boundary as defined in question 0.1 (in square km)

1.0:Does your city incorporate sustainability goals and targets (e.g. GHG reductions) into the master planning for the city?

1.0a:Please detail which goals and targets are incorporated in your city's master plan and describe how these goals are addressed in the table below.

1.0a_1:Goal type

1.0a_1[01]:Emissions reduction targets

1.0a_1[02]:Adaptation targets

1.0a_1[03]:Renewable energy targets

1.0a_1[04]:Energy efficiency targets

1.0a_1[05]:Water security targets

1.0a_1[06]:Waste management targets

1.0a_1[07]:Other, please specify

1.0a_2:How are these goals/targets addressed in the city master plan?

1.1:Has the Mayor or city council committed to climate adaptation and/or mitigation across the geographical area of the city?

1.11:How would you characterize the data management of your city and department?

1.11_1:City

1.11_1[01]:Initial. Our city does not have a stable, consistent environment for information management

1.11_1[02]:Recognised. Our city has recognised that we are not managing our sustainability data and are in the process of planning and establishing a system planning and establishing a system
1.11_1[03]:Repeatable. Our city has undocumented policies and procedures in place to repeat some information processes
1.11_1[04]:Defined. Our city has documented policies and procedures for the management of information across the organisation
1.11_1[05]:Managed. Our city has established organisational wide metrics for each department and results are measured
1.11_1[06]:Optimised. Our city is focussed on continuous process improvement through the use of data
1.11_2:Department
1.11_2[01]:Initial. Our department does not have a stable, consistent environment for information management
1.11_2[02]:Recognised. Our department has recognised that we are not managing our sustainability data and are in the process of planning and establishing a system planning and establishing a system
1.11_2[03]:Repeatable. Our department has undocumented policies and procedures in place to repeat some information processes
1.11_2[04]:Defined. Our department has documented policies and procedures for the management of information across the organisation
1.11_2[05]:Managed. Our department has established organisational wide metrics for each department and results are measured
1.11_2[06]:Optimised. Our department is focussed on continuous process improvement through the use of data
1.12:What tools does your city / department use to manage its environmental related data? Select all that apply.
1.13:What tools does your city / department use to analyse its environmental related data? Select all that apply.
1.14:Does your city have a team dedicated to data analysis (e.g.; data analytics staff; performance management staff; evaluation staff; chief data officer; etc.)?
1.15:Has your city's Mayor or equivalent legal authority communicated their commitment to governing with data publicly to city residents (e.g. through public remarks; press releases; etc.)?
1.1a:Please select any commitments to climate adaptation and/or mitigation your city has signed and attach evidence.
1.1a_1:Name of commitment and attach document
1.1a_1[01]:Global Covenant of Mayors for Climate & Energy
1.1a_1[02]:Durban Adaptation Charter

1.1a_1[03]:Mexico City Pact

1.1a_1[04]:UNISDR, Making Cities Resilient Campaign

1.1a_1[05]:100 Resilient Cities

1.1a_1[06]:Resilient Communities for America

1.1a_1[07]:STAR Communities

1.1a_1[08]:LEED for Cities

1.1a_1[09]:Mayors National Climate Action Agenda

1.1a_1[10]:Chicago Climate Charter

1.1a_1[11]:Klimakommune (Climate Municipality)

1.1a_1[12]:1001.1a_1[13]:Building Efficiency Accelerator

1.1a_1[14]:District Energy in Cities Initiative

1.1a_1[15]:One Planet City Challenge

1.1a_1[16]:EcoMobility Alliance

1.1a_1[17]:ICLEI's Green Climate Cities Program

1.1a_1[18]:Deadline 2020 - Delivering the 1.5 degree ambition of the Paris Agreement in a resilient, inclusive way

1.1a_1[19]:Individual city Commitment

1.1a_1[20]:Other: please specify

1.1a_2:Type of commitment

1.1a_2[01]:Adaptation

1.1a_2[02]:Mitigation

1.1a_2[03]:Both

1.1a_2[04]:Other

1.1a_3:Comments

1.1b:Please explain why your Mayor or city council does not have any commitments to climate adaptation and/or mitigation.

1.1b_1:Reason

1.1b_1[01]:Lack of political will

1.1b_1[02]:Commitments are under consideration

1.1b_1[03]:Lack of funding / resources

1.1b_1[04]:Lack of expertise / knowledge

1.1b_1[05]:Actions prioritised over commitment

1.1b_1[06]:Other: please specify

1.1b_2:Explanation

1.2:Please attach the letter from your city's Mayor requesting the relevant local government department to participate in the Green Climate Cities (GCC) program.

1.3:Please list the local government departments involved in the GCC program and its role. It is important to specify the program coordinator; action plan developer; GHG inventory accountant; verifier and action plan implementer.

1.3_1:Name of the department

1.3_2:Number of employees in the department

1.3_3:Role in the GCC program

1.3_4:Attach awareness raising and capacity building plan for the municipal staff

1.3_5:Attach organigram or other relevant reference document

1.4:Please list the key development challenges; barriers and opportunities within the GCC Program.

1.4_1:Type

1.4_2:Please describe the selected development; challenge; barrier or opportunity

1.4_3:Attach SWOT analysis or SOAR analysis result

1.5:Please list the stakeholder engagement activities for each relevant stakeholder group

1.5_1:Name of the stakeholder group

1.5_2:Role in the GCC program

1.5_3:Name of the engagement activities

1.5_4:Aim of the engagement activities

1.5_5:Please attach stakeholder engagement and communication plan

1.5_6:Attach reference document such as meeting minutes; pictures or webpage

1.6:Does the Mayor have a statutory duty (legal responsibility) to reduce greenhouse gases?

1.7:How many staff (FTE) work on topics related to climate change mitigation and adaptation?

1.7_1:Mitigation

1.7_2:Adaptation

1.8:Please describe your city's climate data management plan including data collection; storing; quality assurance/checking (QA/QC) and updating of the plan; and attach reference document.

1.9:How many staff (FTE) does your city have for environmental related data management (including collecting; storing; analyzing and communicating)?

4.0:Does your city have a city-wide emissions inventory to report?

4.1:Please state the dates of the accounting year or 12-month period for which you are reporting your latest city-wide GHG emissions inventory.

4.11:Has the city-wide GHG emissions data you are currently reporting been externally verified or audited in part or in whole?

4.11a:Please provide the following information about the city-wide emissions verification.

4.11a_1:Name of verifier and attach verification certificate

4.11a_2:Year of verification

4.11a_3:Please explain which parts of your inventory are verified

4.11b:Please explain why your city-wide emissions inventory is not verified and describe any plans to verify your city-wide emissions in the future.

4.11b_1:Reason

4.11b_1[01]:Verification under consideration

4.11b_1[02]:Lack of funding / resources

4.11b_1[03]:Lack of expertise / knowledge

4.11b_1[04]:Verification is not prioritised

4.11b_1[05]:Data is internally verified

4.11b_1[06]:Other

4.11b_2:Comments

4.12:Please provide details on any historical and base year city-wide emissions inventories your city has; in order to allow assessment of targets in the table below.

4.12_1:Inventory date from

4.12_2:Inventory date to

4.12_3:Scopes / boundary covered

4.12_3[01]:Total emissions

4.12_3[02]:Scope 1 (direct)

4.12_3[03]:Scope 2 (indirect)

4.12_3[04]:Scope 3 (other indirect)

4.12_3[05]:Other

4.12_4:Previous emissions (metric tonnes CO2e)

4.12_5:Is this inventory used as the base year inventory?

4.12_5[01]:Yes

4.12_5[02]:No

4.12_6:Methodology

4.12_6[01]:Global Protocol for Community Greenhouse Gas Emissions Inventories (GPC)

4.12_6[02]:International Standard for Determining Greenhouse Gas Emissions for Cities (UNEP and World Bank)

4.12_6[03]:2006 IPCC Guidelines for National Greenhouse Gas Inventories

4.12_6[04]:U.S. Community Protocol for Accounting and Reporting of Greenhouse Gas Emissions (ICLEI)

4.12_6[05]:Regional or country specific methodology

4.12_6[06]:City specific methodology

4.12_6[07]:Other

4.12_7:File name and attach your inventory

4.12_8:Comments

4.13:Since your last submission; have you needed to recalculate any past city-wide GHG emission inventories previously reported to CDP?

4.13a:Please provide your city's recalculated total city-wide emissions figures for any previous inventories along with Scope 1; 2 and 3 breakdowns where applicable. 4.13a_1:Inventory date from

4.13a_2:Inventory date to

4.13a_3:Scope

4.13a_3[01]:Total emissions

4.13a_3[02]:Scope 1 (direct)

4.13a_3[03]:Scope 2 (indirect)

4.13a_3[04]:Scope 3 (other indirect)

4.13a_3[05]:Other

4.13a_4:Previous emissions (metric tonnes CO2e)

4.13a_5:Updated emissions (metric tonnes CO2e)

4.13a_6:Updated methodology

4.13a_6[01]:Global Protocol for Community Greenhouse Gas Emissions Inventories (GPC)

4.13a_6[02]:International Standard for Determining Greenhouse Gas Emissions for Cities (UNEP and World Bank)

4.13a_6[03]:2006 IPCC Guidelines for National Greenhouse Gas Inventories

4.13a_6[04]:U.S. Community Protocol for Accounting and Reporting of Greenhouse Gas Emissions (ICLEI)

4.13a_6[05]:Regional or country specific methodology

4.13a_6[06]:City specific methodology

4.13a_6[07]:Other

4.13a_7:File name and attach your new inventory

4.13a_8:Reasoning for recalculation

4.1_1:From

4.1_2:To

4.2:Please indicate the category that best describes the boundary of your city-wide GHG emissions inventory.

4.2_1:Boundary of inventory relative to city boundary (reported in 0.1)

4.2_1[01]:Same - covers entire city and nothing else

4.2_1[02]:Smaller - covers only part of the city

4.2_1[03]:Larger - covers the whole city and adjoining areas

4.2_1[04]:Partial - Covers part of the city and adjoining areas

4.2_2:Excluded sources / areas

4.2_3:Explanation of boundary choice where the inventory boundary differs from the city boundary (include inventory boundary; GDP and population)

4.3:Please give the name of the primary protocol; standard; or methodology you have used to calculate your city's city-wide GHG emissions.

4.3_1:Primary protocol

4.3_1[01]:Global Protocol for Community Greenhouse Gas Emissions Inventories (GPC)

4.3_1[02]:International Standard for Determining Greenhouse Gas Emissions for Cities (UNEP and World Bank)

4.3_1[03]:2006 IPCC Guidelines for National Greenhouse Gas Inventories

4.3_1[04]:U.S. Community Protocol for Accounting and Reporting of Greenhouse Gas Emissions (ICLEI)

4.3_1[05]:Regional or country specific methodology

4.3_1[06]:City specific methodology

4.3_1[07]:Other

4.3_2:Comment

4.3a:The Global Covenant of Mayors requires committed cities to report their inventories in the format of the new Common Reporting Framework; to encourage standard reporting of emissions data. If your city is reporting an updated inventory; we encourage reporting this in the CRF format; for which guidance can be found in the link below. Would you like to report your inventory in the CRF format or continue to report in the GPC format? Please ensure you respond to this question in order for the correct emissions breakdown questions to be displayed.

4.4:Which gases are included in your city-wide emissions inventory? Select all that apply.

4.5:Please attach your city-wide inventory in Excel or other spreadsheet format and provide additional details on the inventory calculation methods in the table below.

4.5_1:Emissions inventory format

4.5_1[01]:GPC format: City Inventory Reporting and Information System (CIRIS) GPC Reporting tool

4.5_1[02]:GPC format: ClearPath (ICLEI)

4.5_1[03]:Custom or older GPC format

4.5_1[04]:SCATTER

4.5_1[05]:This inventory is in a format other than the GPC

4.5_2:Document title and attachment

4.5_3:Emissions factors used

4.5_3[01]:IPCC

4.5_3[02]:LCA

4.5_3[03]:Other: EPA

4.5_4:Global Warming Potential (select relevant IPCC Assessment Report)

4.5_4[01]:IPCC 2nd AR (1995)

4.5_4[02]:IPCC 3rd AR (2001)

4.5_4[03]:IPCC 4th AR (2007)

4.5_4[04]:IPCC 5th AR (2013)

4.5_5:Please select which additional sectors are included in the inventory

4.5_5[01]:Industrial process and/or product use

4.5_5[02]:Agriculture, forestry or other land use sectors

4.5_6:Population in inventory year

4.5_7:Overall Level of confidence

4.5_7[01]:High

4.5_7[02]:Medium

4.5_7[03]:Low

4.5_8:Comment on level of confidence

4.6a:The Global Covenant of Mayors requires committed cities to report their inventories in the format of the new Common Reporting Framework; to encourage standard reporting of emissions data. Please provide a breakdown of your city-wide emissions by sector and sub-sector in the table below. Where emissions data is not available; please use the relevant notation keys to explain the reason why.

4.6a_1:Direct emissions / Scope 1 (metric tonnes CO2e)

4.6a_2:If you have no direct emissions to report; please select a notation key to explain why

4.6a_3:Indirect emissions from the use of grid-supplied electricity; heat; steam and/or cooling / Scope 2 (metric tonnes CO2e)

4.6a_4:If you have no indirect emissions to report; please select a notation key to explain why

4.6a_5:Emissions occurring outside the city boundary as a result of in-city activities / Scope 3 (metric tonnes CO2e)

4.6a_6:If you have no emissions occurring outside the city boundary to report as a result of in-city activities; please select a notation key to explain why

4.6a_7:Please explain any excluded sources; identify any emissions covered under an ETS and provide any other comments

4.6c:Please provide a breakdown of your GHG emissions by scope. Where values are not available; please use the comment field to indicate the reason why.

4.6c_1:Scope 1 emissions excluding emissions from grid-supplied energy generation

4.6c_10:Calculated total Scope 1 + Scope 2 emissions

4.6c_11:Total (Scope 1 + Scope 2) emissions - please ensure this matches the total calculated field above

4.6c_12:Level of confidence

4.6c_12[01]:High

4.6c_12[02]:Medium

4.6c_12[03]:Low

4.6c_13:Total Scope 3 emissions

4.6c_14:Level of confidence

4.6c_14[01]:High

4.6c_14[02]:Medium

4.6c_14[03]:Low

4.6c_2:Level of confidence

4.6c_2[01]:High

4.6c_2[02]:Medium

4.6c_2[03]:Low

4.6c_3:Scope 1 emissions from grid-supplied energy generation within the city boundary

4.6c_4:Level of confidence

4.6c_4[01]:High

4.6c_4[02]:Medium

4.6c_4[03]:Low

4.6c_5:Calculated Total Scope 1 emissions

4.6c_6:Total Scope 1 emissions - please ensure this matches the calculated total above

4.6c_7:Level of confidence

4.6c_7[01]:High

4.6c_7[02]:Medium

4.6c_7[03]:Low

4.6c_8:Total Scope 2 emissions

4.6c_9:Level of confidence

4.6c_9[01]:High

4.6c_9[02]:Medium

4.6c_9[03]:Low

4.6d:Where it will facilitate a greater understanding of your city-wide emissions; please provide a breakdown of these emissions by IPCC sector in the table below.

4.6d_1:IPCC sector

4.6d_1[01]:Energy

4.6d_1[02]:Industrial processes and product use (IPPU)

4.6d_1[03]:Agriculture, Forestry and other land use (AFOLU)

4.6d_1[04]:Waste

4.6d_1[05]:Other

4.6d_2:Sector

4.6d_2[01]:Stationary energy (buildings)

4.6d_2[02]:Residential buildings

4.6d_2[03]:Public buildings

4.6d_2[04]:Commercial buildings

4.6d_2[05]:Industrial buildings

4.6d_2[06]:Transportation

4.6d_2[07]:Road

4.6d_2[08]:Rail

4.6d_2[09]:Waste

4.6d_2[10]:Wastewater

4.6d_2[11]:Other

4.6d_3:Scope

4.6d_3[01]:Scope 1

4.6d_3[02]:Scope 2

4.6d_3[03]:Total figure

4.6d_4:Emissions (metric tonnes CO2e)

4.6e:Where it will facilitate a greater understanding of your city-wide emissions; please provide a breakdown of these emissions by the US Community Protocol sources.

4.6e_1:US Community Protocol Sources

4.6e_1[01]:Built environment

4.6e_1[02]:Transportation and other mobile sources

4.6e_1[03]:Solid waste

4.6e_1[04]:Wastewater and water

4.6e_1[05]:Agricultural livestock

4.6e_1[06]:Upstream impacts of community-wide activities

4.6e_2:Sector

4.6e_2[01]:Stationary energy (buildings)

4.6e_2[02]:Residential buildings

4.6e_2[03]:Public buildings

4.6e_2[04]:Commercial buildings

4.6e_2[05]:Industrial buildings

4.6e_2[06]:Transportation

4.6e_2[07]:Road

4.6e_2[08]:Rail

4.6e_2[09]:Waste

4.6e_2[10]:Wastewater

4.6e_2[11]:Other

4.6e_3:Scope

4.6e_3[01]:Scope 1

4.6e_3[02]:Scope 2

4.6e_3[03]:Total figure

4.6e_4:Emissions (metric tonnes CO2e)

4.6f:Where it will facilitate a greater understanding of your citywide emissions; please provide a breakdown of these emissions by end user (buildings; water; waste; transport); economic sector (residential; commercial; industrial; institutional); or any other classification system used in your city.

4.6f_1:Source

4.6f_2:Sector

4.6f_2[01]:Stationary energy (buildings)

4.6f_2[02]:Residential buildings

4.6f_2[03]:Public buildings

4.6f_2[04]:Commercial buildings

4.6f_2[05]:Industrial buildings

4.6f_2[06]:Transportation

4.6f_2[07]:Road

4.6f_2[08]:Rail

4.6f_2[09]:Waste

4.6f_2[10]:Wastewater

4.6f_2[11]:Other

4.6f_3:Scope

4.6f_3[01]:Scope 1

4.6f_3[02]:Scope 2

4.6f_3[03]:Total figure

4.6f_4:Emissions (metric tonnes CO2e)

4.7:If the submitted GHG inventory is baseline inventory for target setting; please provide the Baseline Synthesis Report and stakeholder consultation process and results to this inventory.

4.7_1:Year of inventory as baseline of the target

4.7_2:Baseline synthesis report

4.7_3:Data gap analysis report

4.7_4:Stakeholder consultation reference document for this inventory; including consultation process and results

4.8:Please indicate if your city-wide emissions have increased; decreased; or stayed the same since your last emissions inventory; and describe why.

4.8_1:Change in emissions

4.8_1[01]:Increased

4.8_1[02]:Decreased

4.8_1[03]:Stayed the same

4.8_1[04]:This is our first year of calculation

4.8_1[05]:Do not know

4.8_2:Primary reason for change

4.8_2[01]:Increased energy/electricity consumption

4.8_2[02]:Population increase Text field

4.8_2[03]:Improved data accuracy

4.8_2[04]:Emissions reduction actions not implemented

4.8_2[05]:Change in weather conditions

4.8_2[06]:Change in accounting methodology

4.8_2[07]:Change in calculation following verification

4.8_2[08]:Behavioural change

4.8_2[09]:Technological change

4.8_2[10]:Legislative change

4.8_2[11]:Change in available data

4.8_2[12]:Change in data collection methods

4.8_2[13]:Policy change

4.8_2[14]:Financial conditions

4.8_2[15]:Other

4.8_2[21]:Lack of resource / funding overcome

4.8_2[22]:Lack of knowledge overcome

4.8_2[31]:No new inventory to report

4.8_2[32]:Emissions have not changed

4.8_2[41]:Change in staff

4.8_2[42]:Lack of documentation

4.8_2[43]:Change in methodology

4.8_3:Please explain and quantify changes in emissions

4.9:Does your city have a consumption-based inventory to measure emissions from consumption of goods and services by your residents?

4.9_1:Response

4.9_1[01]:Yes

4.9_1[02]:In progress

4.9_1[03]:Intending to incorporate in the next 2 years

4.9_1[04]:Not intending to incorporate

4.9_1[05]:Do not know

4.9_2:Provide an overview and attach your consumption-based inventory if relevant

5.0:Do you have a GHG emissions reduction target in place at the city-wide level? Select all that apply.

5.0a:Please provide details of your total city-wide base year emissions reduction (absolute) target. In addition; you may add rows to provide details of your sector-specific targets; by providing the base year emissions specific to that target.

5.0a_1:Sector

5.0a_10:Percentage of target achieved so far

5.0a_11:Does this target align with the global 1.5 - 2 °C pathway set out in the Paris Agreement?

5.0a_11[01]:Yes - 1.5 c

5.0a_11[02]:Yes - 2°c

5.0a_11[03]:No

5.0a_11[04]:Do not know

5.0a_12:Please indicate to which sector(s) the target applies

5.0a_12[01]:Energy industry

5.0a_12[02]:Heating and cooling supply

5.0a_12[03]:Commercial buildings

5.0a_12[04]:Residential buildings

5.0a_12[05]:Public facility

5.0a_12[06]:Industrial facilities

5.0a_12[07]:Transport

5.0a_12[08]:Water

5.0a_12[09]:Other

5.0a_13:Does this target align to a requirement from a higher level of sub-national government

5.0a_13[01]:Yes

5.0a_13[02]:Yes, but it exceeds its scale or requirement

5.0a_13[03]:No

5.0a_13[04]:Do not know

5.0a_14:Please describe your target. If your country has an NDC and your city's target is less ambitious than the NDC; please explain why.

5.0a_1[01]:All emissions sources included in city inventory

5.0a_1[02]:Energy

5.0a_1[03]:Transport

5.0a_1[04]:Waste

5.0a_1[05]:Other

5.0a_2:Where sources differ from the inventory; identify and explain these additions / exclusions

5.0a_3:Boundary of target relative to city boundary (reported in 0.1)

5.0a_3[01]:Same - covers entire city and nothing else

5.0a_3[02]:Smaller - covers only part of the city

5.0a_3[03]:Larger - covers the whole city and adjoining areas

5.0a_3[04]:Partial - Covers part of the city and adjoining areas

5.0a_4:Base year

5.0a_5:Year of target implementation

5.0a_6:Base year emissions (metric tonnes CO2e)

5.0a_7:Percentage reduction target

5.0a_8:Target year

5.0a_9:Target year absolute emissions (metric tonnes CO2e)

5.0b:Please provide details of your total fixed level target.

5.0b_1:Sector

5.0b_10:Please indicate to which sector(s) the target applies

5.0b_10[01]:Energy industry

5.0b_10[02]:Heating and cooling supply

5.0b_10[03]:Commercial buildings

5.0b_10[04]:Residential buildings

5.0b_10[05]:Public facility

5.0b_10[06]:Industrial facilities

5.0b_10[07]:Transport

5.0b_10[08]:Water

5.0b_10[09]:Other

5.0b_11:Does this target align to a requirement from a higher level of government?

5.0b_11[01]:Yes

5.0b_11[02]:Yes, but it exceeds its scale or requirement

5.0b_11[03]:No

5.0b_11[04]:Do not know

5.0b_12:Please describe your target. If your country has an NDC and your city's target is less ambitious than the NDC; please explain why.

5.0b_1[01]:All emissions sources included in city inventory

5.0b_1[02]:Energy

5.0b_1[03]:Transport

5.0b_1[04]:Waste

5.0b_1[05]:Other

5.0b_2:Where sources differ from the inventory; identify and explain these additions / exclusions

5.0b_3:Boundary of target relative to city boundary (reported in 0.1)

5.0b_3[01]:Same - covers entire city and nothing else

5.0b_3[02]:Smaller - covers only part of the city

5.0b_3[03]:Larger - covers the whole city and adjoining areas

5.0b_3[04]:Partial - Covers part of the city and adjoining areas

5.0b_4:Year of target implementation

5.0b_5:Target year

5.0b_6:Projected population in target year

5.0b_7:Target year absolute emissions goal (metric tonnes CO2e)

5.0b_8:Percentage of target achieved

5.0b_9:Does this target align with the global 1.5 -2 °C pathway set out in the Paris agreement?

5.0b_9[01]:Yes - 1.5 c

5.0b_9[02]:Yes - 2°c

5.0b_9[03]:No

5.0b_9[04]:Do not know

5.0c:Please provide details of your total city-wide base year intensity target. An intensity target is usually measured per capita or per unit GDP. If you have an absolute emissions reduction target; please select "Base year emissions (absolute) target" in question 5.0.

5.0c_1:Sector

5.0c_10:Target year

5.0c_11:Target year absolute emissions (metric tonnes CO2e)

5.0c_12:Percentage of target achieved

5.0c_13:Does this target align with the global 1.5 - 2 °C pathway set out in the Paris agreement?

5.0c_13[01]:Yes - 1.5 c

5.0c_13[02]:Yes - 2°c

5.0c_13[03]:No

5.0c_13[04]:Do not know

5.0c_14:Please indicate to which sector(s) the target applies

5.0c_14[01]:Energy industry

5.0c_14[02]:Heating and cooling supply

5.0c_14[03]:Commercial buildings

5.0c_14[04]:Residential buildings

5.0c_14[05]:Public facility

5.0c_14[06]:Industrial facilities

5.0c_14[07]:Transport

5.0c_14[08]:Water

5.0c_14[09]:Other

5.0c_15:Does this target correspond to a requirement from a higher level of government?

5.0c_15[01]:Yes

5.0c_15[02]:Yes, but it exceeds its scale or requirement

5.0c_15[03]:No

5.0c_15[04]:Do not know

5.0c_16:Please describe your target. If your country has an NDC and your city's target is less ambitious than the NDC; please explain why.

5.0c_1[01]:All emissions sources included in city inventory

5.0c_1[02]:Energy

5.0c_1[03]:Transport

5.0c_1[04]:Waste

5.0c_1[05]:Other

5.0c_2:Where sources differ from the inventory; identify and explain these additions / exclusion

5.0c_3:Boundary of target relative to city boundary (reported in 0.1)

5.0c_3[01]:Same - covers entire city and nothing else

5.0c_3[02]:Smaller - covers only part of the city

5.0c_3[03]:Larger - covers the whole city and adjoining areas

5.0c_3[04]:Partial - Covers part of the city and adjoining areas

5.0c_4:Base year

5.0c_5:Year of target implementation

5.0c_6:Intensity unit (Emissions per)

5.0c_6[01]:Metric tonnes of CO2e per capita

5.0c_6[02]:Metric tonnes of CO2e per unit GDP

5.0c_6[03]:Other

5.0c_7:Base year emissions per intensity unit (metric tonnes CO2e per denominator)

5.0c_8:Base year absolute emissions (metric tonnes CO2e)

5.0c_9:Percentage reduction target in emissions intensity

5.0d:Please provide details of your total city-wide baseline scenario target; including projected business as usual emissions.

5.0d_1:Sector

5.0d_10:Percentage of target achieved

5.0d_11:Does this target align with the global 1.5 - 2 °C pathway set out in the Paris agreement?

5.0d_11[01]:Yes - 1.5 c

5.0d_11[02]:Yes - 2°c

5.0d_11[03]:No

5.0d_11[04]:Do not know

5.0d_12:Please describe the target and the modelling methodology(ies) and parameters used to define it

5.0d_13:Please indicate to which sector(s) the target applies

5.0d_13[01]:Energy industry

5.0d_13[02]:Heating and cooling supply

5.0d_13[03]:Commercial buildings

5.0d_13[04]:Residential buildings

5.0d_13[05]:Public facility

5.0d_13[06]:Industrial facilities

5.0d_13[07]:Transport

5.0d_13[08]:Water

5.0d_13[09]:Other

5.0d_14:Does this target correspond to a requirement from a higher level of government?

5.0d_14[01]:Yes

5.0d_14[02]:Yes, but it exceeds its scale or requirement

5.0d_14[03]:No

5.0d_14[04]:Do not know

5.0d_15:Please describe your target. If your country has an NDC and your city's target is less ambitious than the NDC; please explain why.

5.0d_1[01]:All emissions sources included in city inventory

5.0d_1[02]:Energy

5.0d_1[03]:Transport

5.0d_1[04]:Waste

5.0d_1[05]:Other

5.0d_2:Where sources differ from the inventory; identify and explain these additions / exclusions

5.0d_3:Boundary of target relative to city boundary (reported in 0.1)

5.0d_3[01]:Same - covers entire city and nothing else

5.0d_3[02]:Smaller - covers only part of the city

5.0d_3[03]:Larger - covers the whole city and adjoining areas

5.0d_3[04]:Partial - Covers part of the city and adjoining areas

5.0d_4:Base year

5.0d_5:Year of target implementation

5.0d_6:Base year emissions (metric tonnes CO2e)

5.0d_7:Target year

5.0d_8:Estimated business as usual absolute emissions in target year (metric tonnes CO2e)

5.0d_9:Percentage reduction target from business as usual

5.0e:Please explain why you do not have a city-wide emissions reduction target

and any plans to set one in the future.
5.0e_1:Reason
5.0e_1[01]:Emissions not calculated
5.0e_1[02]:Not intending to set a target
5.0e_1[03]:Lack of resources
5.0e_1[04]:Lack of available data
5.0e_1[05]:Policies/projects prioritized over target setting
5.0e_1[06]:Target is set at regional level
5.0e_1[07]:Target is set at national level
5.0e_1[08]:Target is in development
5.0e_1[09]:Target already achieved
5.0e_1[10]:Other
5.0e_2:Comment
5.1:Please describe how the target(s) reported above align with the global 1.5 - 2 °C pathway set out in the Paris agreement.
5.2:Is your city-wide emissions reduction target(s) conditional on the success of an externality or component of policy outside of your control?
5.2a:Please identify and describe the conditional components of your city-wide emissions reduction target(s).
5.3:Does your city-wide emissions reduction target(s) account for the use of transferable emissions units?
5.3a:Please provide details on the use of transferable emissions.
5.3a_1:Type of transferable emissions
5.3a_1[01]:Renewable energy generation produced within the geographic boundary, or reflecting an investment by the city
5.3a_1[02]:Renewable energy credits
5.3a_1[03]:Offset credit transactions generated within the boundary and sold
5.3a_1[04]:Offset credit transactions purchased from outside of the boundary
5.3a_1[05]:Other
5.3a_2:Emissions saved (metric tonnes CO2e)
5.3a_3:What percentage of the target does this unit represent?
5.3a_4:Please identify which target this refers to and describe the transferable emissions unit in particular the source of the transferable units
5.4:Describe the anticipated outcomes of the most impactful mitigation actions your city is currently undertaking; the total cost of the action and how much is being funded by the local government.
5.4_1:Mitigation action
5.4_10:Action description
5.4_11:Finance status

5.4_11[01]:Pre-feasibility study status

5.4_11[02]:Feasibility undertaken

5.4_11[03]:Feasibility finalized, and finance partially secured

5.4_11[04]:Finance secured

5.4_12:Total cost of the project

5.4_13:Total cost provided by the local government

5.4_14:Primary fund source

5.4_14[01]:Local

5.4_14[02]:(Sub) national

5.4_14[03]:International (ODA)

5.4_14[04]:Climate finance (carbon credits)

5.4_14[05]:Public-private partnership

5.4_14[06]:Other

5.4_15:Web link to action website

5.4_16:Name of the stakeholder group

5.4_17:Role in the GCC program

5.4_18:Name of the engagement activities

5.4_19:Aim of the engagement activities

5.4_2:Action title

5.4_20:Attach reference document

5.4_3:Means of implementation

5.4_4:Implementation status

5.4_4[01]:Pre-feasibility study

5.4_4[02]:Pre-implementation

5.4_4[03]:Implementation

5.4_4[04]:Implementation complete

5.4_4[05]:Operation

5.4_4[06]:Monitoring and reporting

5.4_5:Estimated emissions reduction (metric tonnes CO2e)

5.4_6:Energy savings (MWh)

5.4_7:Renewable energy production (MWh)

5.4_8:Timescale of reduction / savings / energy production

5.4_8[01]:Per year

5.4_8[02]:Projected lifetime

5.4_8[03]:Other

5.4_9:Co-benefit area

5.5:Does your city have a climate change mitigation or energy access plan for reducing city-wide GHG emissions?

5.5a:Please attach your city's climate change mitigation plan below. If your city has both action and energy access plans; please make sure to attach all relevant documents below.

5.5a_1:Publication title and attach document

5.5a_10:Has there been a stakeholder engagement plan to develop the plan?

5.5a_11:Primary author of plan

5.5a_11[01]:Dedicated city team

5.5a_11[02]:Relevant city department

5.5a_11[03]:Consultant

5.5a_11[04]:International organisation

5.5a_11[05]:Community group

5.5a_11[06]:Regional / state / provincial government

5.5a_11[07]:National / central government

5.5a_11[08]:Other

5.5a_2:Year of adoption from local government

5.5a_3:Web link

5.5a_4:Areas covered by action plan

5.5a_5:Boundary of plan relative to city boundary (reported in 0.1)

5.5a_5[01]:Same - covers entire city and nothing else

5.5a_5[02]:Smaller - covers only part of the city

5.5a_5[03]:Larger - covers the whole city and adjoining areas

5.5a_5[04]:Partial - Covers part of the city and adjoining areas

5.5a_6:If the city boundary is different from the plan boundary; please explain why and any areas/other cities excluded or included

5.5a_7:Stage of implementation

5.5a_7[01]:Plan in development

5.5a_7[02]:Plan developed but not implemented

5.5a_7[03]:Plan in implementation

5.5a_7[04]:Implementation complete

5.5a_7[05]:Measurement in progress

5.5a_7[06]:Plan update in progress

5.5a_7[07]:Other

5.5a_8:Has your local government assessed the synergies; trade-offs; and co-benefits; if any; of the main mitigation and adaptation actions you identified?

5.5a_8[01]:Yes

5.5a_8[02]:In progress

5.5a_8[03]:Intending to undertake in the next 2 years

5.5a_8[04]:Not intending to undertake

5.5a_8[05]:Don’t know

5.5a_9:Comment or describe the synergies; trade-offs; and co-benefits of this interaction

5.5b:Please explain why you do not have a city climate change mitigation plan and any future plans to create one.

5.5b_1:Reason

5.5b_1[01]:No plans yet to create an action plan

5.5b_1[02]:Resources lacking to create an action plan

5.5b_1[03]:Action plan in early stages of project planning

5.5b_1[04]:Action planning in progress

5.5b_1[05]:Lack of budget/resources

5.5b_1[06]:Other

5.5b_2:Comment

7.0:Do you have an emissions inventory for your local government operations to report? Reporting a Local Government Operations emissions inventory is optional.

7.1:Please state the dates of the accounting year or 12-month period for which you are reporting an emissions inventory for your local government operations.

7.1_1:From

7.1_2:To

7.2:Please indicate the category that best describes the boundary of your local government operations emissions inventory.

7.3:Please give the name of the primary protocol; standard; or methodology used to calculate your local government operations emissions inventory and attach your inventory using the attachment function.

7.3_1:Primary protocol and attach inventory

7.3_1[01]:Greenhouse Gas Protocol: Public Sector Standard

7.3_1[02]:International Emissions Analysis Protocol (ICLEI)

7.3_1[03]:ISO 14064

7.3_1[04]:Local Government Operations Protocol (ICLEI/The Climate Registry/California Climate Action Registry/ California Air Resources Board)

7.3_1[05]:Australian National Greenhouse and Energy Reporting (Measurement) Determination

7.3_1[06]:Global Protocol for Community-Scale Greenhouse Gas Emissions Inventories (GPC), (WRI, C40 and ICLEI)

7.3_1[07]:2006 IPCC Guidelines for National Greenhouse Gas Inventories

7.3_1[08]:Other

7.3_2:Comment

7.4:Which gases are included in your emissions inventory? Select all that apply.

7.5:Please give the total amount of fuel (refers to Scope 1 emissions) that your local government has consumed this year.

7.5_1:Source

7.5_1[01]:Airport (s)

7.5_1[02]:Buildings

7.5_1[03]:Buses

7.5_1[04]:Electricity generation

7.5_1[05]:Electricity transmission and distribution

7.5_1[06]:Employee commuting

7.5_1[07]:Incineration of waste

7.5_1[08]:Landfills

7.5_1[09]:Local trains

7.5_1[10]:Maritime port

7.5_1[11]:Municipal vehicle fleet

7.5_1[12]:Regional trains

7.5_1[13]:Roads / highways

7.5_1[14]:Street lighting and traffic signals

7.5_1[15]:Subway / underground

7.5_1[16]:Thermal energy

7.5_1[17]:Waste collection

7.5_1[18]:Wastewater treatment

7.5_1[19]:Water supply

7.5_1[20]:Unknown source

7.5_1[21]:Total

7.5_1[22]:Other

7.5_2:Fuel

7.5_2[01]:Natural gas

7.5_2[02]:Compressed Natural Gas (CNG)

7.5_2[03]:Liquefied Petroleum Gas (LPG)

7.5_2[04]:Methane

7.5_2[05]:Butane

7.5_2[06]:Propane

7.5_2[07]:Town gas or city gas

7.5_2[08]:Coal (Bituminous or Black coal)

7.5_2[09]:Coking coal

7.5_2[10]:Crude oil

7.5_2[11]:Diesel/Gas oil

7.5_2[12]:Motor gasoline (petrol)

7.5_2[13]:Aviation gasoline

7.5_2[14]:Jet gasoline

7.5_2[15]:Jet kerosene

7.5_2[16]:Kerosene

7.5_2[17]:Residual fuel oil

7.5_2[18]:Distillate fuel oil No 1

7.5_2[19]:Distillate fuel oil No 2

7.5_2[20]:Distillate fuel oil No 3

7.5_2[21]:Distillate fuel oil No 4

7.5_2[22]:Distillate fuel oil No 5

7.5_2[23]:Distillate fuel oil No 6

7.5_2[24]:Liquified petroleum gas (LPG)

7.5_2[25]:Bitumen

7.5_2[26]:Petroleum coke

7.5_2[27]:Wood or wood waste

7.5_2[28]:Biodiesel

7.5_2[29]:Biogasoline

7.5_2[30]:Ethanol

7.5_2[31]:E85

7.5_2[32]:Other liquid biofuel

7.5_2[33]:Landfill gas

7.5_2[34]:Other biogas

7.5_2[35]:Waste (municipal)

7.5_2[36]:Other

7.5_3:Amount

7.5_4:Units

7.5_4[01]:GWh

7.5_4[02]:MWh

7.5_4[03]:kWh

7.5_4[04]:TJ

7.5_4[05]:GJ

7.5_4[06]:MJ

7.5_4[07]:Therms

7.5_4[08]:Btu m3

7.5_4[09]:L

7.5_4[10]:Metric tonnes

7.5_4[11]:Short tons

7.5_5:Emissions (tonnes CO2e)

7.6:Please provide total (Scope 1 + Scope 2) GHG emissions for your local government operations; in metric tonnes CO2e. Scopes are a common categorization method. 7.6_1:Total Scope 1 + Scope 2 emissions (metric tonnes CO2e)

7.6_2:Total Scope 1 emissions (metric tonnes CO2e)

7.6_3:Total Scope 2 emissions (metric tonnes CO2e)

7.6_4:Comment

7.7:Do you measure local government Scope 3 emissions?

7.7a:Please complete the table.

7.7a_1:Source of Scope 3 emissions

7.7a_1[01]:Employee commuting

7.7a_1[02]:Employee business travel

7.7a_1[03]:Emissions from contracted services

7.7a_1[04]:Upstream production of materials and fuels

7.7a_1[05]:Upstream and downstream transportation of materials and fuels

7.7a_1[06]:Waste related Scope 3 emission sources

7.7a_1[07]:Other

7.7a_2:Emissions (metric tonnes CO2e)

7.7a_3:Comment

7.7b:Please explain why not and detail your plans to do so in the future; if any.

7.7b_1:Reasoning

7.7b_1[01]:Lack of data

7.7b_1[02]:Low data quality

7.7b_1[03]:Lack of knowledge/capacity

7.7b_1[04]:Lack of funding/resources

7.7b_1[05]:Scope categorization not used

7.7b_1[06]:Not required by national authorities

7.7b_1[07]:Not required by international agreements

7.7b_1[08]:Local government Scope 3 emissions have been assessed as insignificant

7.7b_1[09]:Other

7.7b_2:Explanation

7.8:Please indicate if your local government operations emissions have increased; decreased; or stayed the same since your last emissions inventory; and please describe why.

7.8_1:Change in emissions

7.8_1[01]:Increased

7.8_1[02]:Decreased

7.8_1[03]:Stayed the same

7.8_1[04]:This is our first year of calculation

7.8_1[05]:Do not know

7.8_2:Primary reason for change

7.8_2[01]:Increased energy/electricity consumption

7.8_2[02]:Population increase

7.8_2[03]:Improved data accuracy

7.8_2[04]:Emissions reduction actions not implemented

7.8_2[05]:Change in weather conditions

7.8_2[06]:Change in accounting methodology

7.8_2[07]:Change in calculation following verification

7.8_2[08]:Behavioural change

7.8_2[09]:Technological change

7.8_2[10]:Legislative change

7.8_2[11]:Change in available data

7.8_2[12]:Change in data collection methods

7.8_2[13]:Policy change

7.8_2[14]:Financial conditions

7.8_2[15]:Lack of resource / funding overcome

7.8_2[16]:Lack of knowledge overcome

7.8_2[17]:No new inventory to report

7.8_2[18]:Emissions have not changed

7.8_2[19]:Change in staff

7.8_2[20]:Lack of documentation

7.8_2[21]:Change in methodology

7.8_2[22]:Other

7.8_3:Please explain

7.9:Has the GHG emissions data you are currently reporting been externally verified or audited in part or in whole?

7.9a:Please provide the following information about the emissions verification process.

7.9a_1:Name of verifier and attach verification certificate

7.9a_2:Year of verification

7.9a_3:Please explain which parts of your inventory are verified

7.9b:Please explain why your local government operations inventory is not verified and describe any future plans for verification.

7.9b_1:Reason

7.9b_1[01]:Verification under consideration

7.9b_1[02]:Lack of funding / resources

7.9b_1[03]:Lack of expertise / knowledge

7.9b_1[04]:Verification is not prioritised

7.9b_1[05]:Data is internally verified

7.9b_1[06]:Other

7.9b_2:Explanation

Appendix C

Detailing of experimental results

Table C.1: ERM-L execution for all CDP cities.

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
1093	Atlanta/GA	USA	1	1	100	0	1	1	0	1	0
1184	Austin/TX	USA	2	1	110	1	1	1	0	1	0
1499	Barcelona	Spain	3	1	111	1	2	1	0	1	0
2028	Bonn	Germany	1	1	100	0	2	1	0	1	0
2185	Bristol	UK	1	1	100	0	2	1	0	1	1
2430	Burlington/VT	USA	3	1	100	1	2	0	0	1	0
3203	Chicago	USA	1	1	111	0	1	1	0	1	0
3417	New York City	USA	3	1	100	1	3	1	0	1	0
3422	London	UK	3	1	111	1	2	1	0	1	0
3429	Stockholm	Sweden	2	1	111	1	1	0	0	1	1
8242	Helsinki	Finland	1	1	101	0	2	1	0	1	1
10495	Las Vegas/NV	USA	1	1	100	0	3	0	0	1	0
10894	Los Angeles/CA	USA	3	1	111	1	2	0	0	1	0
11315	Manchester	UK	1	1	101	0	2	1	0	1	1
13067	New Orleans/LA	USA	0	0	110	1	1	0	0	1	0
14088	Oslo	Norway	3	1	111	1	3	1	0	1	1
14344	Park/UT	USA	1	1	100	0	3	1	0	1	0
14874	Portland/OR	USA	3	1	110	1	2	1	0	1	1
16581	Seattle/WA	USA	3	1	111	1	3	1	0	1	1

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
19233	Torres Vedras/11	Portugal	0	0	100	1	1	0	0	1	1
20113	Vancouver	Canada	3	1	111	1	2	1	0	1	1
31009	Copenhagen	Denmark	2	1	111	1	1	0	0	1	0
31051	Coventry	UK	0	1	000	1	1	1	0	1	0
31052	Cardiff	UK	1	1	100	0	2	0	0	1	1
31055	Glasgow	UK	0	0	100	1	1	1	0	1	1
31090	Washington/DC	USA	3	1	100	1	4	1	0	1	0
31108	Houston/TX	USA	1	1	110	0	1	0	0	1	1
31109	Melbourne	Australia	1	1	111	0	2	1	0	1	0
31110	Rome	Italy	0	0	111	0	1	0	0	1	1
31111	Tokyo	Japan	3	1	111	1	2	1	0	1	0
31112	Kaohsiung	Taiwan	0	1	000	1	2	1	0	1	1
31113	Yokohama	Japan	2	1	100	1	1	0	0	1	0
31114	Sydney	Australia	3	1	111	1	3	1	0	1	1
31115	Johannesburg	South Africa	2	1	110	1	1	1	0	1	0
31117	Toronto	Canada	3	1	111	1	5	1	0	1	0
31146	Addis Ababa	Ethiopia	2	1	110	1	1	1	0	1	0
31148	Amsterdam	Netherlands	1	1	111	0	1	1	0	1	1
31149	Athens	Greece	2	1	111	1	1	1	0	1	0
31150	Bangkok	Thailand	3	1	011	1	2	1	0	1	0
31151	Basel	Switzerland	1	1	110	0	3	1	0	1	1
31153	Berlin	Germany	3	1	111	1	2	1	0	1	1
31154	Bogotá	Colombia	0	0	100	0	2	0	0	1	0
31155	Buenos Aires	Argentina	1	1	111	0	2	0	0	1	0
31156	Curitiba	Brazil	1	1	110	0	1	0	0	1	1
31157	Delhi	India	2	1	010	1	1	0	0	1	0
31163	Istanbul	Turkey	1	1	110	0	2	0	0	1	1
31165	Heidelberg	Germany	2	1	110	1	1	1	0	1	0
31166	Jakarta	Indonesia	1	1	111	0	3	1	1	1	1
31167	Lagos	Nigeria	1	1	111	0	1	1	0	1	1
31168	Karachi	Pakistan	2	1	110	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
31169	Hong Kong	Hong Kong	3	1	101	1	2	0	0	1	0
31170	Lima	Peru	1	1	110	0	1	0	0	1	1
31171	Madrid	Spain	1	1	111	0	2	0	0	1	0
31172	Mexico City	Mexico	1	1	111	0	3	1	0	1	0
31173	Milano	Italy	2	1	100	1	1	1	0	1	0
31174	Moscow	Russia	1	1	111	0	2	1	0	1	1
31175	Paris	France	3	1	111	1	2	1	0	1	1
31176	Rio de Janeiro	Brazil	3	1	111	1	4	1	0	1	1
31177	Salt Lake City	USA	1	1	100	0	2	1	0	1	0
31179	Rotterdam	Netherlands	3	1	111	1	2	1	0	1	0
31180	Santiago	Chile	1	1	111	0	1	1	0	1	1
31181	Philadelphia/PA	USA	2	1	111	1	1	1	0	1	1
31182	San Francisco	USA	3	1	111	1	3	1	0	1	0
31184	São Paulo	Brazil	2	1	101	1	1	1	0	1	1
31185	Warsaw	Poland	1	1	111	0	2	1	0	1	0
31187	Seoul	Republic of Korea	3	1	011	1	2	1	0	1	1
31446	Taipei	Taiwan	0	1	000	1	2	1	0	1	1
32480	Adelaide	Australia	1	1	100	0	2	1	0	1	1
32550	Denver/CO	USA	3	1	101	1	2	1	0	1	1
35268	Boston/MA	USA	2	1	111	1	1	1	0	1	1
35274	Portland/ME	USA	0	0	110	1	1	1	0	0	0
35393	Saint Louis	USA	0	0	000	1	2	1	0	1	1
35449	Zürich	Switzerland	1	1	100	0	2	1	0	1	1
35475	Calgary	Canada	0	1	000	1	2	1	0	1	1
35755	Kadiovacik	Turkey	0	1	000	1	1	1	0	1	0
35848	Belo Horizonte	Brazil	1	1	100	0	2	1	0	1	1
35853	Baltimore/MD	USA	3	1	100	1	2	1	0	1	1
35854	Brussels	Belgium	3	1	101	1	2	0	0	1	1
35857	Cincinnati	USA	2	1	100	1	1	1	0	1	1
35858	Cape Town	South Africa	1	1	111	0	3	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
35859	Cleveland/OH	USA	0	0	100	1	3	1	0	1	1
35860	Dallas/TX	USA	0	1	000	0	1	1	0	1	1
35862	Detroit	USA	0	1	000	1	1	0	0	1	0
35863	eThekwini	South Africa	3	1	100	1	2	1	0	1	0
35864	Ekurhuleni	South Africa	1	1	100	0	1	1	0	1	0
35865	Fortaleza	Brazil	1	1	100	0	1	1	0	1	1
35867	Guadalajara	Mexico	1	1	110	0	1	0	0	1	1
35870	Miami/FL	USA	0	0	100	1	2	1	0	1	0
35872	Recife	Brazil	0	0	100	0	2	0	0	1	1
35873	Medellín	Colombia	0	0	100	0	1	1	1	1	1
35874	Phoenix/AZ	USA	1	1	111	0	1	1	0	1	0
35877	Pittsburgh	USA	1	1	100	0	2	1	0	1	1
35878	Sacramento	USA	0	1	000	1	1	0	0	1	0
35879	Minneapolis/MN	USA	1	1	100	0	1	0	0	1	0
35880	Porto Alegre	Brazil	2	1	100	1	1	0	0	1	0
35883	San José	USA	3	1	100	1	2	1	0	1	0
35884	San Diego/CA	USA	3	1	100	1	2	1	0	1	0
35885	Tel Aviv-Yafo	Israel	1	1	001	0	1	0	0	1	1
35886	Torino	Italy	0	0	100	1	2	1	0	1	0
35887	Valencia	Spain	0	0	100	1	2	1	0	1	0
35893	Dar es Salaam	Tanzania	1	1	100	0	1	0	0	1	0
35894	Montreal	Canada	2	1	100	1	1	1	0	1	0
35897	Campinas/SP	Brazil	3	1	100	1	2	0	0	1	1
35898	Greater Manchester	UK	3	1	100	1	2	1	0	1	1
35903	Casablanca	Morocco	0	1	000	1	1	0	0	1	0
35904	Kolkata	India	2	1	010	1	1	1	0	1	0
35905	Chennai	India	1	1	010	0	1	1	0	1	1
35907	Bangalore/KA	India	0	1	000	1	1	0	0	1	0
35910	Pune	India	0	1	000	1	2	1	0	1	0
35913	Nairobi	Kenya	1	1	111	0	1	0	0	1	0
35915	Jaipur	India	2	1	010	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
35993	Singapore	Singapore	3	1	111	1	2	1	0	1	0
36002	Kinshasa	Democratic Republic Of Congo	2	1	100	1	1	1	0	0	0
36004	Abidjan	Côte d'Ivoire	1	1	110	0	1	1	0	0	0
36032	Dakar	Senegal	1	1	110	0	1	1	0	1	0
36036	Ibadan	Nigeria	0	1	000	0	1	0	0	1	0
36037	Cali	Colombia	1	1	100	0	2	1	0	1	0
36039	Accra	Ghana	1	1	110	0	2	1	0	1	1
36041	Belém/PA	Brazil	0	1	000	1	1	0	0	1	0
36043	Abuja Federal Capital Territory	Nigeria	0	1	000	1	1	0	0	1	0
36045	Guayaquil	Ecuador	2	1	100	1	1	0	0	1	0
36158	Napoli	Italy	0	0	100	0	2	0	0	1	0
36159	Lisbon/11	Portugal	1	1	111	0	2	1	0	1	1
36223	Antananarivo	Madagascar	0	1	000	1	1	1	0	0	0
36254	Provincia di Venezia	Italy	3	1	100	1	2	0	0	1	1
36261	Bolzano	Italy	0	0	100	0	2	1	0	1	1
36262	Genova	Italy	3	1	100	1	2	1	0	1	1
36263	Ravenna	Italy	3	1	100	1	2	0	0	1	0
36274	Bologna	Italy	3	1	101	1	2	1	0	1	1
36282	Chieti	Italy	2	1	100	1	1	0	0	1	0
36285	Florence/FI	Italy	2	1	100	1	1	0	0	1	1
36286	Ferrara	Italy	0	0	100	0	3	0	0	0	0
36410	Memphis/TN	USA	3	1	100	1	3	0	0	1	1
36426	Riga	Latvia	1	1	100	0	2	1	0	1	0
36469	L'Aquila	Italy	0	0	100	0	1	1	0	1	0
36470	La Spezia	Italy	1	1	100	0	1	1	0	1	0
36477	Lucca	Italy	0	0	100	0	1	1	0	1	0
36491	Pesaro	Italy	0	0	100	1	2	1	0	1	0
36492	Parma	Italy	3	1	100	1	2	0	0	1	0

Continued on next page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
36493	Pescara	Italy	0	0	100	1	1	0	0	1	0
36494	Padova	Italy	2	1	100	1	1	0	0	1	0
36495	Piacenza	Italy	0	0	100	0	1	0	0	1	0
36501	Prato/45	Italy	0	0	100	1	1	1	0	1	0
36504	Rimini	Italy	2	1	100	1	1	1	0	1	0
36512	Teramo	Italy	2	1	100	1	1	1	0	1	0
36522	Verbania	Italy	1	1	100	0	1	1	0	1	0
37038	Cologne	Germany	0	0	100	1	1	0	0	0	0
37241	Berkeley/CA	USA	1	1	100	0	1	1	0	1	0
37261	Pietermaritzburg	South Africa	0	1	000	0	1	0	0	1	0
42120	Salvador	Brazil	1	1	110	0	1	1	0	1	1
42123	Goiânia	Brazil	2	1	100	1	1	1	0	1	0
42178	Quito	Ecuador	1	1	110	0	1	1	0	1	1
42388	Montevideo	Uruguay	2	1	100	1	1	0	0	1	0
43905	San Antonio/TX	USA	2	1	100	1	1	0	0	1	0
43907	Indianapolis	USA	2	1	100	1	1	1	0	1	1
43909	Orlando	USA	0	0	100	1	1	1	0	1	0
43910	Columbus/OH	USA	3	1	100	1	2	1	0	1	1
43911	Ottawa	Canada	0	0	100	0	2	0	0	1	1
43912	Edmonton	Canada	1	1	100	0	2	1	0	1	1
43914	Charlotte/NC	USA	1	1	100	0	1	1	0	1	0
43917	Sofia	Bulgaria	1	1	101	0	1	0	0	1	0
43920	Ljubljana	Slovenia	3	1	100	1	2	1	0	1	1
43921	Zagreb	Croatia	0	0	100	0	2	1	0	1	0
43923	Hannover	Germany	0	0	100	1	2	1	0	1	0
43928	Canberra	Australia	3	1	100	1	2	1	0	1	1
43930	The Hague	Netherlands	2	1	101	1	1	1	0	1	1
43932	Auckland	New Zealand	2	1	111	1	1	1	0	1	0
43934	Perth/WA	Australia	1	1	100	0	1	1	0	1	1
43937	Wellington	New Zealand	2	1	100	1	1	0	0	1	0
43938	Dubai	United Arab	1	1	111	0	2	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
		Emirates									
43940	Malmö	Sweden	3	1	100	1	2	1	0	1	0
43969	Asuncion	Paraguay	0	0	100	1	1	1	0	1	0
43975	Lima	Peru	1	1	110	0	1	0	0	1	0
44076	Bursa	Turkey	1	1	100	0	2	0	0	1	0
44077	Kampala	Uganda	1	1	100	0	2	0	0	1	1
44185	Suwon	Republic of Korea	0	1	000	0	1	1	0	1	0
45219	Aparecida/SP	Brazil	0	0	100	1	1	0	0	0	0
46470	Vitoria-Gasteiz	Spain	0	0	100	1	2	1	0	1	1
46473	Zaragoza	Spain	0	0	101	0	2	1	0	1	0
46514	Porto	Portugal	0	0	100	1	2	1	0	1	0
49172	St. Petersburg	USA	1	1	101	0	1	1	0	1	0
49327	Providence/RI	USA	1	1	100	0	2	1	0	1	0
49330	Kansas City/KS	USA	0	1	000	0	2	1	0	1	1
49333	Louisville/KY	USA	1	1	100	0	2	1	0	1	0
49334	Richmond/VA	USA	1	1	100	0	1	1	0	1	0
49335	Nashville/TN	USA	2	1	100	1	1	1	0	1	1
49339	Honolulu	USA	0	1	000	1	2	1	0	1	1
49342	Rochester/NY	USA	3	1	100	1	2	1	0	1	0
49347	Omaha	USA	0	1	000	1	1	1	0	0	0
49359	Harare	Zimbabwe	1	1	100	0	1	0	0	1	0
49360	Tshwane	South Africa	1	1	100	0	1	0	0	1	1
49367	Freetown	Sierra Leone	2	1	010	1	1	1	0	1	0
49787	Benicia	USA	2	1	100	1	1	0	0	1	0
50154	Turku	Finland	3	1	100	1	2	1	0	1	1
50203	Gaziantep	Turkey	3	1	100	1	2	1	0	1	1
50211	Tbilisi	Georgia	0	0	100	0	1	0	0	1	0
50220	Nice	France	1	1	100	0	2	1	0	1	0
50354	Tegucigalpa	Honduras	1	1	100	0	1	0	0	1	0
50356	Morelia	Mexico	0	1	000	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
50357	Mendoza	Argentina	1	1	100	0	1	1	0	1	1
50358	Toluca de Guadalupe	Mexico	0	0	100	1	1	1	0	0	0
50359	León de los Aldamas	Mexico	0	1	000	1	3	1	0	1	1
50361	Hermosillo/SON	Mexico	0	1	000	0	2	1	0	1	1
50362	Concepción	Chile	0	0	000	1	1	0	0	1	0
50364	La Paz	Bolivia	2	1	100	1	1	0	0	0	0
50368	Provincia de Arequipa	Peru	0	1	000	1	1	0	0	1	0
50370	Tampico	Mexico	2	1	100	1	1	0	0	1	0
50371	Córdoba	Argentina	2	1	100	1	1	1	0	1	1
50373	Rosario	Argentina	1	1	100	0	3	0	0	1	0
50375	Chihuahua/CHH	Mexico	0	1	000	1	2	0	0	1	0
50377	Querétaro/QUE	Mexico	0	1	000	0	1	0	0	1	0
50378	San José/SJ	Costa Rica	3	1	100	1	2	1	0	1	0
50380	Bucaramanga	Colombia	1	1	100	0	1	1	0	1	0
50381	Torreón	Mexico	0	1	000	0	1	1	0	1	0
50382	Mérida/YUC	Mexico	0	1	000	1	2	1	0	1	1
50383	Sorocaba	Brazil	1	1	100	0	2	0	0	1	1
50384	Florianópolis	Brazil	2	1	100	1	1	0	0	1	0
50385	Campo Grande/MS	Brazil	2	1	100	1	1	0	0	0	0
50386	Cuiabá	Brazil	1	1	100	0	1	0	0	1	0
50387	Guarulhos	Brazil	0	0	100	1	1	0	0	0	0
50388	Natal	Brazil	2	1	100	1	1	0	0	0	0
50389	Maceió	Brazil	2	1	100	1	1	0	0	1	0
50390	Teresina	Brazil	2	1	100	1	1	0	0	1	0
50391	Manaus	Brazil	2	1	100	1	1	0	0	0	0
50392	Vitória	Brazil	0	0	100	1	2	0	0	1	0
50394	João Pessoa	Brazil	1	1	100	0	1	0	0	1	1
50395	São Luís/MA	Brazil	0	1	000	1	1	0	0	1	0
50396	Santos	Brazil	0	1	000	1	1	1	0	1	0
50398	Juárez/CHH	Mexico	0	1	000	1	1	1	0	1	0
50401	Madison/WI	USA	0	1	000	0	1	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
50541	Greensboro/NC	USA	0	1	000	1	1	0	0	1	0
50543	Halifax Regional	Canada	1	1	100	0	1	1	0	1	0
50544	Aurora/IL	USA	1	1	100	0	1	1	0	1	0
50549	Fort Worth	USA	0	1	000	1	1	0	0	0	0
50550	Buffalo/NY	USA	0	1	000	0	2	1	0	0	0
50551	Long Beach/CA	USA	1	1	100	0	1	0	0	1	0
50555	Hamilton	Canada	1	1	100	0	2	1	0	1	0
50557	Norfolk/VA	USA	2	1	100	1	1	0	0	1	1
50558	London/ON	Canada	3	1	111	1	2	1	0	1	0
50559	St Catharines/ON	Canada	0	0	100	0	1	0	0	1	0
50560	Oakland/CA	USA	3	1	100	1	3	0	0	1	0
50562	Chula Vista	USA	1	1	100	0	1	1	0	1	0
50565	Toledo/OH	USA	0	1	000	0	1	0	0	1	0
50566	Anchorage/AK	USA	1	1	100	0	1	0	0	1	0
50568	Saskatoon	Canada	2	1	100	1	1	1	0	1	0
50571	Victoria	Canada	3	1	100	1	2	1	0	1	1
50572	Saint Paul/MN	USA	1	1	100	0	1	0	0	1	0
50578	Windsor/ON	Canada	3	1	100	1	2	1	0	1	0
50579	Winnipeg	Canada	0	1	000	0	1	1	0	1	1
50650	Gibraltar	Gibraltar	1	1	100	0	2	1	0	1	0
50665	Ovar	Portugal	0	0	100	1	1	0	0	1	0
50671	Fafe	Portugal	1	1	100	0	2	1	0	1	1
50672	Santarém	Portugal	2	1	100	1	1	0	0	1	0
50673	Faro	Portugal	1	1	100	0	1	0	0	0	0
50674	Viseu	Portugal	1	1	100	0	1	0	0	1	0
50679	Barreiro	Portugal	2	1	100	1	1	0	0	1	0
50680	Cascais/11	Portugal	1	1	100	0	2	0	0	1	0
50681	Funchal	Portugal	2	1	100	1	1	0	0	1	0
50782	Dhaka	Bangladesh	2	1	110	1	1	0	0	1	0
50792	Ville de Monaco	Monaco	1	1	100	0	1	1	0	1	1
51075	Shenzhen	China	0	0	011	0	1	1	0	0	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
51079	Guatemala City	Guatemala	2	1	100	1	1	0	0	1	0
51374	Rio Branco/AC	Brazil	0	0	100	1	1	0	0	1	0
52638	Aracaju	Brazil	2	1	100	1	1	0	0	1	1
52894	Winston-Salem	USA	0	1	000	1	1	0	0	0	0
52897	Aspen	USA	1	1	100	0	2	1	0	1	0
53254	Hobart	Australia	1	1	100	0	2	1	0	1	1
53829	Kingston/ON	Canada	0	0	100	0	1	1	0	1	0
53860	Wilmington/NC	USA	0	1	000	0	1	1	0	1	0
53921	Tempe/AZ	USA	2	1	100	1	1	1	0	1	1
53959	Fayetteville/AR	USA	0	1	000	0	2	1	0	1	1
54026	Tacoma	USA	2	1	100	1	1	1	0	1	0
54029	Spokane	USA	0	0	100	1	1	0	0	1	1
54030	Little Rock/AR	USA	0	0	000	1	1	0	0	1	0
54037	Des Moines/IA	USA	0	0	100	0	1	1	0	1	0
54048	Knoxville/TN	USA	1	1	100	0	2	0	0	1	0
54057	Lancaster/CA	USA	1	1	100	0	2	0	0	1	0
54060	Greater Sudbury	Canada	0	1	000	0	1	0	0	1	0
54066	Fort Collins	USA	1	1	100	0	2	1	0	1	0
54070	Eugene	USA	3	1	100	1	2	0	1	1	1
54075	Lakewood/CO	USA	3	1	100	1	2	1	0	1	0
54078	Hayward/CA	USA	2	1	100	1	1	1	0	1	0
54082	Hollywood/FL	USA	0	0	000	1	2	0	0	1	0
54084	Guelph	Canada	1	1	100	0	2	1	0	1	1
54085	Savannah/GA	USA	1	1	100	0	1	1	0	1	0
54088	Peterborough	Canada	0	1	000	0	2	1	0	1	0
54092	Ann Arbor	USA	0	0	100	0	2	1	0	1	0
54098	Thunder Bay	Canada	0	0	100	0	1	1	0	1	0
54100	Columbia/MO	USA	1	1	100	0	2	0	0	1	1
54102	Albany/NY	USA	3	1	100	1	2	1	0	1	0
54104	Boulder/CO	USA	3	1	100	1	2	1	0	1	0
54108	Durham/NC	USA	0	1	000	0	1	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
54109	Bloomington/IN	USA	3	1	100	1	3	1	0	1	1
54110	Santa Monica/CA	USA	1	1	100	0	3	1	0	1	1
54111	Iowa	USA	1	1	100	0	1	1	0	1	0
54113	Flagstaff	USA	4	1	100	1	3	1	1	1	0
54114	Asheville	USA	0	1	000	0	1	1	0	0	0
54116	Dubuque	USA	0	0	100	0	2	1	0	1	1
54119	Palo Alto/CA	USA	3	1	100	1	2	0	0	1	0
54124	Fremont/CA	USA	3	1	100	1	2	1	0	1	0
54253	Wollongong	Australia	0	0	100	1	1	0	0	1	1
54270	Palmerston North	New Zealand	0	1	000	0	1	1	0	1	0
54274	Rotorua	New Zealand	1	1	100	0	1	0	0	1	1
54277	New Plymouth District	New Zealand	2	1	100	1	1	0	0	1	0
54291	Chengdu	China	0	0	011	1	1	1	0	0	0
54305	Rajkot	India	1	1	100	0	3	1	0	1	1
54306	Medan	Indonesia	1	1	001	0	1	0	0	1	0
54318	Tangerang	Indonesia	0	0	100	0	1	0	0	1	0
54327	Semarang	Indonesia	0	1	000	0	1	0	0	1	0
54329	Bogor	Indonesia	3	1	100	1	2	0	0	1	1
54335	Yogyakarta	Indonesia	0	0	000	1	1	0	0	1	1
54337	Amman	Jordan	1	1	110	0	1	1	0	1	1
54342	Jbail	Lebanon	0	0	000	1	1	0	0	0	0
54345	Davao City/DVO	Philippines	0	1	000	1	2	0	0	1	0
54347	Pasig	Philippines	0	0	100	0	4	0	0	1	0
54348	Quezon/03	Philippines	0	1	000	1	2	0	0	1	1
54349	Balikpapan	Indonesia	1	1	100	0	2	1	0	1	1
54352	Muntinlupa	Philippines	1	1	100	0	3	1	0	1	1
54356	Parañaque	Philippines	0	1	000	0	1	1	0	1	0
54360	Shah Alam	Malaysia	0	1	000	0	3	1	0	1	0
54361	Petaling Jaya	Malaysia	0	1	000	1	3	1	0	1	0
54364	Kuala Lumpur	Malaysia	2	1	111	1	1	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
54367	Melaka	Malaysia	3	1	100	1	3	1	0	1	1
54370	George Town	Malaysia	0	1	000	0	2	0	0	0	0
54386	Tainan	Taiwan	0	0	100	1	2	1	0	1	1
54388	Iskandar Puteri	Malaysia	2	1	100	1	1	1	0	1	0
54389	Taichung	Taiwan	0	1	000	1	1	1	0	1	1
54391	Nonthaburi	Thailand	0	1	000	1	2	0	0	1	1
54395	Taoyuan	Taiwan	3	1	100	1	2	1	0	1	1
54402	Lahti	Finland	3	1	100	1	2	1	0	1	1
54409	Espoo	Finland	1	1	100	0	2	1	0	1	0
54430	Le Havre	France	0	0	100	0	2	1	0	1	1
54457	Hamburg	Germany	1	1	101	0	1	0	0	1	0
54459	Reykjavík	Iceland	1	1	100	0	2	1	0	1	1
54478	Gemeente Nijmegen	Netherlands	0	1	000	0	2	1	0	1	0
54488	Trondheim	Norway	3	1	100	1	2	0	0	1	1
54491	Málaga	Spain	2	1	100	1	1	0	0	1	1
54493	Kristiansand	Norway	1	1	100	0	1	0	0	1	0
54497	Wroclaw/SL	Poland	0	1	000	0	1	1	0	1	1
54498	Murcia	Spain	0	0	100	1	1	1	0	1	0
54510	Umeå	Sweden	3	1	100	1	3	1	0	1	1
54513	Uppsala	Sweden	3	1	100	1	3	1	0	1	1
54517	Örebro	Sweden	3	1	100	1	2	1	0	1	1
54518	Helsingborg	Sweden	3	1	100	1	3	0	0	1	0
54521	Bournemouth	UK	3	1	100	1	2	1	0	1	1
54529	Leicester	UK	2	1	100	1	1	0	0	1	0
54538	Bath and North East Somerset	UK	0	1	000	0	2	1	0	1	0
54579	Sekhukhune District Municipality	South Africa	0	1	000	1	1	0	0	1	0
54588	West Coast District	South Africa	0	0	000	1	1	0	0	1	0
54603	Pasto	Colombia	1	1	100	0	1	0	0	1	1
54605	Cusco	Peru	2	1	100	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
54608	Montería	Colombia	2	1	100	1	1	1	0	1	1
54609	Trujillo	Peru	1	1	100	0	1	0	0	1	0
54611	Manizales	Colombia	2	1	100	1	1	0	0	1	0
54612	Valledupar	Colombia	1	1	100	0	1	1	0	1	0
54613	Valdivia	Chile	0	1	000	0	1	1	0	1	0
54617	Pereira	Colombia	2	1	100	1	1	0	0	1	0
54619	Piura	Peru	0	1	000	1	1	1	0	1	0
54620	la Paraná	Argentina	1	1	100	0	3	0	0	1	0
54623	Betim	Brazil	2	1	100	1	1	1	0	1	1
54625	Londrina	Brazil	2	1	100	1	1	1	0	1	0
54627	Joinville	Brazil	2	1	100	1	1	0	0	1	0
54633	Lorena	Brazil	0	1	000	1	1	1	0	1	0
54637	Cuenca	Ecuador	1	1	100	0	2	1	0	1	1
54641	Limeira	Brazil	2	1	100	1	1	0	0	0	0
54650	Palmas/TO	Brazil	3	1	100	1	2	1	0	1	0
54651	Santo André/SP	Brazil	2	1	100	1	1	0	0	1	0
54652	Osasco	Brazil	0	1	000	1	1	0	0	0	0
54654	São João da Boa Vista	Brazil	0	1	000	1	1	0	0	0	0
54656	Vinhedo	Brazil	0	1	000	1	1	0	0	1	0
54662	Maringá	Brazil	0	0	100	1	1	0	0	1	0
54667	Contagem	Brazil	2	1	100	1	1	0	0	1	0
54670	Capivari	Brazil	2	1	100	1	1	0	0	1	0
54678	Porto Feliz	Brazil	1	1	100	0	1	1	0	0	0
54681	Araçatuba	Brazil	1	1	100	0	1	1	0	1	0
54683	Franco da Rocha	Brazil	2	1	100	1	1	0	0	0	0
54687	São José dos Campos	Brazil	2	1	100	1	1	0	0	1	0
54692	Sertãozinho/SP	Brazil	0	1	000	1	1	1	0	0	0
54696	XIV La Paz	Mexico	0	1	000	0	2	1	0	1	1
54697	Cerquilho	Brazil	0	1	000	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
54699	Barueri	Brazil	0	0	000	1	1	0	0	0	0
54700	Sumaré	Brazil	0	1	000	0	1	0	0	1	0
54703	Mairiporã	Brazil	0	1	000	1	1	0	0	0	0
54706	Boa Vista/RR	Brazil	2	1	100	1	1	0	0	1	0
54709	Blumenau	Brazil	0	1	000	1	1	0	0	1	0
55324	Guimarães	Portugal	3	1	100	1	2	1	0	1	1
55325	Águeda	Portugal	0	0	100	0	2	0	1	1	0
55331	Ílhavo	Portugal	0	0	100	1	1	0	0	1	0
55334	Braga	Portugal	1	1	100	0	1	0	0	1	1
55371	Vicente López	Argentina	0	1	000	0	1	1	0	1	0
55372	Canoas	Brazil	2	1	100	1	1	1	0	1	0
55373	Cabreúva	Brazil	0	0	000	1	1	1	0	0	0
55379	Santa Fe	Argentina	1	1	100	0	3	0	0	1	1
55380	Cubatão	Brazil	0	1	000	0	1	1	0	1	0
55419	Miramar	USA	2	1	100	1	1	0	0	1	0
55799	Arlington/VA	USA	1	1	100	0	1	0	0	1	1
55800	Cambridge/MA	USA	1	1	100	0	2	1	0	1	0
55801	West Palm Beach	USA	2	1	100	1	1	1	0	1	0
56276	Taipei	Taiwan	0	1	000	1	2	1	0	1	1
57347	Pingtung County	Taiwan	0	0	000	0	1	1	0	1	1
57509	Niterói	Brazil	2	1	100	1	1	0	0	1	1
57616	Lake Forest/IL	USA	0	1	000	1	1	0	0	1	0
58310	Roanoke/VA	USA	2	1	100	1	1	1	0	1	0
58346	Plymouth	UK	0	1	000	0	1	0	0	1	0
58357	West Hollywood/CA	USA	1	1	100	0	1	1	1	1	0
58395	Bærum	Norway	3	1	100	1	3	1	0	1	1
58413	Carmel/IN	USA	2	1	100	1	1	1	0	1	0
58424	Gdańsk	Poland	1	1	100	0	1	0	0	1	0
58482	Laval	Canada	3	1	100	1	2	0	0	1	0
58485	Abington/IL	USA	0	1	000	1	2	0	0	1	0
58489	Hoeje-Taastrup	Denmark	1	1	100	0	2	1	0	1	1

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
	Kommune										
58511	Richmond/CA	USA	3	1	100	1	2	0	0	1	1
58513	Medford/MA	USA	3	1	100	1	2	0	0	1	0
58530	Northampton/MA	USA	1	1	100	0	1	1	0	1	0
58531	Somerville/MA	USA	3	1	100	1	2	1	0	1	0
58543	Byron Shire	Australia	1	1	100	0	2	1	1	1	1
58569	Podgorica	Montenegro	1	1	100	0	1	1	0	1	0
58590	Easton/PA	USA	2	1	100	1	1	0	0	1	0
58591	Greenbelt/MD	USA	0	0	100	1	1	0	0	1	0
58595	Belén/G	Costa Rica	0	1	000	0	2	1	0	1	1
58597	La Unión	Costa Rica	3	1	100	1	2	0	0	1	1
58609	Ærøskøbing	Denmark	0	1	000	1	1	1	0	1	0
58621	Blacksburg/VA	USA	1	1	100	0	2	0	0	1	0
58626	Racine/WI	USA	1	1	100	0	1	0	0	1	0
58627	Alton/IL	USA	1	1	100	0	1	0	0	1	0
58668	New Bedford/MA	USA	1	1	100	0	2	0	0	1	0
58670	Monrovia	Liberia	1	1	100	0	1	0	1	1	0
58671	Helsingør Kommune / Elsinore	Denmark	0	1	000	1	2	1	0	1	1
58783	Province du Ziro	Burkina Faso	0	1	000	1	1	0	0	0	0
58795	Blantyre	Malawi	1	1	100	0	2	1	0	1	1
58796	Odder Kommune	Denmark	0	1	000	1	1	0	0	0	0
58797	Hørsholm Kommune	Denmark	2	1	100	1	1	0	0	1	1
58865	Jammerbugt Kommune	Denmark	0	0	100	0	1	0	0	1	1
58871	Salem/MA	USA	2	1	100	1	1	0	0	1	0
59124	Natchez/MS	USA	2	1	100	1	1	0	0	0	0
59151	Akureyri	Iceland	1	1	100	0	1	0	0	1	0
59158	Moroni	Comoros	0	0	100	0	1	0	0	1	1
59160	Nyon	Switzerland	1	1	100	0	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
59163	Tirana	Albania	0	0	100	0	1	0	0	1	0
59165	Gladsaxe Kommune	Denmark	3	1	100	1	2	1	0	1	0
59166	Independencia	Chile	2	1	100	1	1	1	0	1	0
59167	Providencia	Chile	1	1	100	0	1	1	0	1	0
59168	Dioudoubou	Senegal	2	1	100	1	1	0	0	0	0
59180	Middelfart Kommune	Denmark	2	1	100	1	1	0	0	1	0
59298	Yaoundé 6	Cameroon	1	1	100	0	1	0	0	0	0
59531	Santa Barbara/CA	USA	2	1	100	1	1	0	0	1	1
59535	Town of Vail/CO	USA	1	1	100	0	1	1	0	1	0
59536	Kitchener	Canada	0	0	100	0	2	1	0	1	0
59537	Denton/TX	USA	0	1	000	0	1	0	0	1	1
59538	Mississauga	Canada	1	1	100	0	1	0	0	1	1
59545	Charlottesville/VA	USA	1	1	100	0	1	0	0	1	0
59552	Davis/CA	USA	0	0	100	1	1	0	0	1	0
59562	Urbana/IL	USA	0	0	100	1	2	0	0	1	0
59563	Takoma Park/MD	USA	1	1	100	0	2	1	0	1	0
59580	Town of Dedham/MA	USA	0	1	000	1	1	0	0	1	1
59595	Brisbane/CA	USA	0	0	101	0	1	0	0	1	0
59631	San Leandro/CA	USA	1	1	100	0	1	1	0	1	1
59633	Santa Cruz/CA	USA	2	1	100	1	1	0	0	1	0
59642	Dublin/CA	USA	1	1	001	0	1	0	0	1	1
59644	Culver/CA	USA	0	1	000	0	2	0	0	1	0
59653	Manhattan Beach/CA	USA	3	1	100	1	2	1	0	1	0
59657	Beaverton/OR	USA	3	1	100	1	2	0	0	1	0
59669	North Vancouver	Canada	1	1	100	0	2	1	0	1	0
59678	Evanston/IL	USA	3	1	100	1	2	1	0	1	1
59681	Town of East Hampton/NY	USA	0	0	100	0	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
59697	Lake Worth/FL	USA	0	1	000	1	1	0	0	1	0
59707	Town of Princeton/NJ	USA	0	0	000	0	2	0	0	1	0
59956	Banda Aceh	Indonesia	1	1	100	0	2	1	0	1	1
59969	Mandurah	Australia	2	1	100	1	1	0	0	1	0
59971	Melton	Australia	0	0	100	1	1	0	0	1	0
59996	Batangas	Philippines	1	1	100	0	2	1	0	1	1
60003	Baguio/15	Philippines	1	1	100	0	1	1	0	1	0
60007	Santa Rosa/03	Philippines	1	1	100	0	2	1	0	1	0
60029	Cagayan de Oro	Philippines	3	1	100	1	2	1	0	1	1
60050	Guwahati	India	0	1	000	0	1	1	0	1	0
60053	Indore	India	0	1	000	1	2	0	0	1	0
60073	Wolverhampton	UK	0	1	000	1	1	0	0	1	0
60104	Cambridge	UK	1	1	100	0	1	1	0	1	0
60114	Gdynia	Poland	0	0	100	1	1	1	0	1	0
60125	Klaipeda	Lithuania	0	1	000	1	1	1	0	1	0
60126	Tartu	Estonia	1	1	100	0	1	1	0	1	0
60127	Thessaloniki	Greece	0	0	100	1	1	1	0	1	0
60140	Nakuru	Kenya	2	1	100	1	1	0	0	1	0
60142	Kisumu	Kenya	2	1	100	1	1	1	0	1	0
60216	Växjö	Sweden	3	1	100	1	2	1	0	1	0
60218	Karlskrona	Sweden	1	1	100	0	2	1	0	1	0
60223	Panevėžys	Lithuania	2	1	100	1	1	0	0	1	0
60229	Arendal	Norway	1	1	100	0	2	1	0	1	1
60233	Pärnu	Estonia	0	0	000	1	1	0	0	0	0
60236	Trelleborg	Sweden	1	1	100	0	1	0	0	0	0
60258	Brusque	Brazil	0	1	000	0	1	0	0	1	0
60264	Botucatu	Brazil	0	1	000	0	1	0	0	1	0
60267	Guarujá	Brazil	0	1	000	1	1	1	0	1	0
60268	Brumadinho	Brazil	0	1	000	1	1	0	0	1	0
60271	Bertioga	Brazil	0	1	000	1	1	0	0	0	0
60272	Campina Grande	Brazil	0	0	100	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
60273	Extrema/MG	Brazil	0	1	000	1	1	0	0	1	1
60274	Cruzeiro do Sul/AC	Brazil	2	1	100	1	1	0	0	0	0
60276	Assis	Brazil	2	1	100	1	1	0	0	1	0
60278	Fernandópolis	Brazil	0	1	000	1	1	0	0	0	0
60279	Campos dos Goytacazes	Brazil	2	1	100	1	1	0	0	0	0
60284	Angra dos Reis	Brazil	0	1	000	1	1	0	0	0	0
60292	Jaú	Brazil	0	0	100	1	1	0	0	1	0
60307	Nova Friburgo	Brazil	0	1	000	1	1	0	0	1	0
60318	Porto Velho	Brazil	1	1	100	0	1	0	0	1	0
60320	Presidente Prudente	Brazil	2	1	100	1	1	0	0	1	0
60328	Pirenópolis	Brazil	0	1	000	1	1	0	0	1	0
60332	São Carlos/SP	Brazil	0	0	000	1	1	0	0	0	0
60340	Rio Verde	Brazil	0	1	000	1	1	0	0	0	0
60349	São Leopoldo	Brazil	2	1	100	1	1	1	0	1	0
60361	Tangará da Serra	Brazil	2	1	100	1	1	0	0	1	0
60369	Armenia/ANT	Colombia	0	1	000	0	2	0	0	1	0
60371	Leticia	Colombia	0	0	000	1	1	0	0	0	0
60374	Ibagué	Colombia	0	1	000	1	1	1	0	1	0
60375	Archipiélago de San Andrés	Colombia	0	1	000	1	1	0	0	1	0
60381	Alcaldía Distrital de Santa Marta	Colombia	0	1	000	0	1	0	0	1	0
60384	Yopal	Colombia	0	1	000	1	1	0	0	1	0
60385	Villavicencio	Colombia	1	1	100	0	1	1	0	0	0
60387	Alcaldíade Sincelejo	Colombia	0	0	000	1	1	0	0	0	0
60388	Chiclayo	Peru	2	1	100	1	1	0	0	1	0
60391	San Borja	Peru	1	1	100	0	2	1	0	1	1
60392	San Isidro	Peru	1	1	100	0	1	0	0	1	0
60393	Santiago	Chile	1	1	111	0	1	1	0	1	0
60394	Tarija	Bolivia	1	1	100	0	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
60399	Miraflores	Peru	2	1	100	1	1	1	0	1	0
60400	Temuco	Chile	0	0	100	1	1	0	0	1	0
60408	Talca	Chile	0	1	000	1	1	0	0	1	0
60409	Callao	Peru	2	1	100	1	1	1	0	1	0
60410	Peñalolén	Chile	0	1	000	1	3	0	0	1	0
60414	Venado Tuerto	Argentina	1	1	100	0	3	0	0	1	0
60416	San Isidro	Argentina	2	1	100	1	1	0	0	1	0
60417	Bariloche	Argentina	0	0	100	0	1	0	0	1	0
60419	Rio Grande	Argentina	2	1	100	1	1	0	0	1	1
60433	Hvidovre	Denmark	1	1	100	0	1	1	0	1	0
60577	Frederikshavn Kommune	Denmark	2	1	100	1	1	0	0	0	0
60588	Alba-Iulia	Romania	2	1	100	1	1	1	0	1	0
60599	Town of Bridgewater/NS	Canada	0	1	000	0	1	0	0	1	0
60603	Prince George/BC	Canada	0	1	000	1	1	1	0	1	1
60621	Lilongwe	Malawi	0	1	000	1	1	0	0	0	0
60633	Bujumbura	Burundi	2	1	100	1	1	0	0	0	0
60638	Walvis Bay	Namibia	2	1	100	1	1	1	0	1	0
60656	Piedmont/CA	USA	2	1	100	1	1	0	0	1	0
60898	Naucalpan de Juárez	Mexico	0	1	000	1	1	0	0	1	0
60906	Vitacura	Chile	0	0	100	0	1	0	0	1	1
61427	Nacala	Mozambique	0	0	100	0	2	0	0	1	1
61467	Dipolog	Philippines	2	1	100	1	1	1	0	1	0
61753	Yilan	Taiwan	0	1	000	0	1	0	0	1	0
61790	Emeryville/CA	USA	3	1	100	1	2	1	0	1	1
61876	Mazabuka	Zambia	2	1	100	1	1	0	0	1	0
62791	Botoșani	Romania	0	1	000	1	1	0	0	0	0
62817	Ithaca/NY	USA	0	0	000	1	2	0	0	1	0
62855	Egedal	Denmark	2	1	100	1	1	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
62864	Lancaster/PA	USA	0	0	100	1	1	0	0	1	0
62868	Eskişehir	Turkey	0	0	100	1	1	0	0	1	0
63543	Fredensborg Kommune	Denmark	1	1	100	0	2	1	0	1	0
63562	South Bend/IN	USA	3	1	100	1	2	0	0	1	1
63615	Hillerød Kommune	Denmark	1	1	100	0	1	0	0	1	0
63616	Abasan Al-Kabira	Palestine	1	1	100	0	2	0	0	1	1
63862	Ashland/OR	USA	0	0	100	0	2	0	0	1	0
63919	Saratoga Springs/NY	USA	0	0	000	1	1	1	0	1	0
63941	Broward County/FL	USA	1	1	100	0	1	1	0	1	0
63999	Miami Beach/FL	USA	0	0	100	0	1	0	0	1	0
64014	Cupertino	USA	3	1	100	1	2	1	0	1	0
68290	Wyndham	Australia	1	1	100	0	1	1	0	1	1
68296	Hobsons Bay	Australia	1	1	100	0	1	0	0	1	0
68337	Bekasi	Indonesia	0	0	000	0	1	0	0	1	1
68373	Pedreira	Brazil	0	0	000	1	1	0	0	1	0
68378	Santiago de Surco	Peru	2	1	100	1	1	0	0	1	0
68383	Itatiba	Brazil	0	1	000	1	1	0	0	1	0
68385	Chorrera	Panama	2	1	100	1	1	1	0	1	0
69822	Kristianstad	Sweden	1	1	100	0	2	0	0	1	0
69823	Visby	Sweden	0	1	000	0	1	1	0	1	0
69824	Västervik	Sweden	1	1	100	0	1	1	0	1	1
69834	General Alvear	Argentina	1	1	100	0	3	0	0	1	0
69840	Itacoatiara	Brazil	2	1	100	1	1	1	0	1	0
69848	Loja	Ecuador	1	1	100	0	2	0	0	1	1
69850	Comas	Peru	2	1	100	1	1	0	0	1	0
69968	Rionegro/ANT	Colombia	1	1	100	0	1	1	0	0	0
69985	Sillamäe	Estonia	0	0	000	1	1	1	0	1	0
69995	Kemi	Finland	0	1	000	1	1	1	0	1	0
69999	Greifswald	Germany	1	1	100	0	1	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
70005	Tauragė	Lithuania	2	1	100	1	1	0	0	0	0
70017	Palmira	Colombia	0	1	000	1	1	0	0	1	0
73240	Tuzla	Turkey	0	0	000	1	1	0	0	0	0
73252	Pemba	Mozambique	2	1	100	1	1	0	0	1	0
73293	LaGrange/MO	USA	0	1	000	1	1	0	0	0	0
73295	La Crosse/WI	USA	3	1	100	1	2	0	0	1	1
73301	Gretna/LA	USA	0	1	000	1	1	0	0	1	0
73302	Port Allen/LA	USA	2	1	100	1	1	0	0	0	0
73365	Ithaca/NY	USA	0	0	000	0	2	1	0	1	0
73413	Abidjan	Côte d'Ivoire	1	1	110	0	2	1	0	1	1
73530	Town of Lexington/MA	USA	3	1	100	1	2	0	0	1	1
73637	Nkangala	South Africa	0	1	000	0	1	0	1	1	1
73645	KwaDukuza	South Africa	0	0	100	0	1	1	1	1	1
73648	Arias	Argentina	1	1	100	0	2	0	0	1	0
73650	Armstrong	Argentina	0	0	100	0	1	0	0	1	0
73652	Caseros	Argentina	1	1	100	0	2	0	0	1	0
73663	Iriondo	Argentina	0	1	000	0	3	0	0	1	0
73665	Villa Pehuenia	Argentina	1	1	100	0	1	0	0	1	0
73666	Cuyahoga County	USA	0	0	100	1	2	1	0	1	0
73668	Malabrigo	Argentina	1	1	100	0	2	0	0	1	1
73671	Godoy Cruz	Argentina	1	1	100	0	2	0	0	1	0
73676	Umhlathuze	South Africa	0	0	100	1	1	0	0	1	0
73678	Chañar Ladeado	Argentina	0	0	100	0	1	0	0	1	0
73679	Cruz Alta	Argentina	0	0	100	0	1	0	0	1	0
73680	Carcarañá/S	Argentina	0	0	100	0	1	0	0	1	0
73684	Carlos Tejedor	Argentina	0	0	100	0	1	0	0	1	0
73686	Mendoza	Argentina	0	0	100	0	1	0	0	1	0
73690	Villa General Belgrano	Argentina	1	1	100	0	3	0	0	1	0
73693	La Rioja	Argentina	0	0	100	0	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
73694	Chacabuco	Argentina	1	1	100	0	2	0	0	1	0
73695	Pueblo Uranga	Argentina	0	0	000	0	1	0	0	1	1
73701	San Carlos Sur	Argentina	0	0	100	0	1	0	0	1	0
73706	Alameda	USA	1	1	100	0	1	0	0	1	0
73707	Totoras	Argentina	1	1	100	0	3	0	0	1	0
73709	Los Surgentes	Argentina	0	0	100	0	1	0	0	1	0
73712	Camilo Aldao	Argentina	1	1	100	0	2	0	0	1	0
73713	San Miguel	Argentina	0	0	100	0	1	0	0	1	0
73715	Cosquín	Argentina	0	0	100	0	1	0	0	1	0
73722	Mornington/VIC	Australia	0	0	100	0	1	0	0	1	0
73724	Caseros	Argentina	1	1	100	0	2	0	0	1	0
73725	Guaminí	Argentina	0	0	100	0	1	0	0	1	0
73732	Monte Buey	Argentina	1	1	100	0	2	0	0	1	0
73736	Tafí Viejo	Argentina	0	0	100	0	1	0	0	1	0
73738	Khon Kaen	Thailand	3	1	100	1	2	1	0	1	1
73746	Probolinggo	Indonesia	1	1	100	0	2	0	0	1	0
73749	Nagpur	India	1	1	100	0	1	1	0	1	0
73750	Tarakan	Indonesia	2	1	100	1	1	1	0	1	0
73752	Bontang	Indonesia	3	1	100	1	2	1	0	1	0
73754	Panaji	India	0	0	100	0	1	0	0	1	0
73759	Jambi	Indonesia	2	1	100	1	1	1	0	1	0
73762	Malang	Indonesia	3	1	100	1	2	1	0	1	0
73763	San Carlos/MDC	Philippines	1	1	100	0	2	0	0	1	1
73787	Tapalqué	Argentina	0	0	100	0	1	0	0	1	0
73788	Salliqueló	Argentina	1	1	100	0	2	0	0	1	0
73789	Rafaela	Argentina	0	0	100	0	1	0	0	1	0
73801	Juana Koslay	Argentina	0	0	100	0	1	0	0	1	1
73802	Crespo	Argentina	0	0	100	0	1	0	0	1	0
73803	Herrera	Argentina	0	0	100	0	1	0	0	1	0
73806	Llambi Campbell	Argentina	0	0	100	0	1	0	0	1	0
73879	Roskilde	Denmark	2	1	100	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
74309	Nakhon Sawan	Thailand	0	1	000	1	2	1	0	1	1
74386	Gangtok	India	2	1	100	1	1	0	0	0	0
74401	Encinitas/CA	USA	0	1	000	1	2	1	0	1	0
74414	Boulder County	USA	1	1	100	0	1	1	0	1	0
74418	Town of Breckenridge/CO	USA	0	1	000	1	1	0	0	1	0
74423	Key West/FL	USA	0	0	000	0	1	1	0	1	0
74427	Sarasota	USA	0	0	100	0	1	1	0	1	0
74428	South Miami/FL	USA	0	0	100	1	1	0	0	1	0
74453	Highland Park/IL	USA	1	1	100	0	2	0	0	1	0
74466	Village of South Barrington/IL	USA	0	1	000	1	1	0	0	0	0
74488	Beverly/MA	USA	0	0	000	1	1	0	0	1	0
74508	Winona/MN	USA	0	0	000	1	1	1	0	1	0
74531	Santa Fe County	USA	2	1	100	1	1	0	0	0	0
74534	Erie County/NY	USA	0	1	000	0	2	1	0	1	0
74558	Summit County/UT	USA	1	1	100	0	1	1	0	1	0
74560	Moab/UT	USA	1	1	100	0	2	1	0	1	1
74563	Town of Guilford/VT	USA	0	0	100	1	1	0	0	0	0
74573	Snoqualmie/WA	USA	0	1	000	0	1	1	0	0	0
74575	Dane County	USA	0	1	000	1	1	0	0	0	0
74594	Boynton Beach	USA	1	1	100	0	2	0	0	1	0
74631	Lubumbashi	Democratic Republic Of Congo	2	1	100	1	1	0	0	1	0
74673	İzmir	Turkey	1	1	100	0	3	1	0	1	0
74677	Cluj-Napoca	Romania	1	1	100	0	1	0	0	1	0
74678	Galaţi	Romania	0	0	000	1	1	0	0	1	0
74680	Iaşi	Romania	0	0	100	0	1	0	0	1	0
826167	Tapalpa	Mexico	0	0	000	1	1	1	0	1	0
826182	Tonalá/JAL	Mexico	0	1	000	1	1	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
826207	Junta Intermunicipal de Medio Ambiente Region Valles (JIMAV)	Mexico	3	1	100	1	2	1	0	1	1
826208	Junta intermunicipal de Medio Ambiente de Sierra Occidental Y Costa (JISOC)	Mexico	2	1	100	1	1	0	0	0	0
826209	Aipromades Lago de Chapala	Mexico	2	1	100	1	1	1	0	1	0
826210	Junta Intermunicipal de Medio Ambiente de la Costa Sur (JICOSUR)	Mexico	1	1	100	0	1	0	0	1	0
826212	Junta intermunicipal para la gestión integral de la cuenca del Río Coahuayana (JIRCO)	Mexico	0	0	100	1	1	0	0	0	0
826236	Tremembé	Brazil	2	1	100	1	1	0	0	0	0
826237	Madrid	Colombia	2	1	011	1	1	0	0	1	1
826239	Sopó	Colombia	0	1	000	1	1	1	0	1	0
826380	Junta Intermunicipal de la Cuenca Baja del Rio Ayuquila (JIRA)	Mexico	0	0	100	1	1	0	0	0	0
826396	Sintra/11	Portugal	3	1	100	1	2	1	0	1	1
826407	Mirandela	Portugal	0	1	000	1	1	0	0	0	0
826429	Figueira da Foz	Portugal	0	0	000	1	1	0	0	0	0
827048	Zhenjiang	China	0	0	010	0	2	0	0	1	0
831152	San Pedro de Urabá	Colombia	0	1	000	1	1	0	0	1	0
831230	Al Marsá	Tunisia	0	1	000	1	1	0	0	1	0
831433	Ataliva	Argentina	0	1	000	1	1	0	0	0	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
831616	Tsévié	Togo	2	1	100	1	1	0	0	1	1
831617	Bouaké	Côte d'Ivoire	1	1	100	0	1	0	0	1	1
831618	Yaoundé 4	Cameroon	1	1	100	0	1	1	0	1	0
831620	Yaoundé 3	Cameroon	1	1	100	0	1	1	0	1	0
831674	Amarante	Portugal	0	0	000	1	1	0	0	0	0
831823	Massa Marittima	Italy	0	0	100	0	1	0	0	0	0
831926	Ramallah	Palestine	0	0	100	1	1	0	0	0	0
831999	Monte Verde	Costa Rica	0	0	100	0	2	0	0	1	1
832000	Desamparados/SJ	Costa Rica	2	1	100	1	1	1	0	1	0
832078	Mafra/11	Portugal	0	0	000	1	1	0	0	1	0
832097	Lagos	Portugal	2	1	011	1	1	0	0	1	0
832274	Odemira	Portugal	0	0	000	1	1	0	0	0	0
832610	Orange County/NC	USA	1	1	100	0	1	0	0	1	0
832838	Town of Wellfleet/MA	USA	1	1	100	0	1	0	0	1	0
832909	Coruche/14	Portugal	0	0	100	0	2	0	0	1	0
833379	Bani-Suhaila	Palestine	1	1	100	0	1	0	0	1	0
834058	Bogor Regency	Indonesia	0	1	000	1	2	1	0	1	1
834083	Eau Claire/WI	USA	1	1	100	0	2	1	0	1	1
834120	Tanjung Pinang	Indonesia	0	0	100	0	1	0	0	1	0
834153	Melaka	Malaysia	1	1	100	0	1	0	0	1	0
834161	Kinmen County	Taiwan	0	1	000	1	1	0	0	1	0
834163	Si Satchanalai	Thailand	0	0	000	0	2	1	0	1	1
834167	Cochin	India	0	1	000	0	1	0	0	1	0
834202	Mogale City	South Africa	0	1	000	0	1	1	1	0	0
834219	Corrientes	Argentina	1	1	100	0	2	0	0	1	0
834226	Bell Ville	Argentina	1	1	100	0	2	0	0	1	0
834229	Bragado	Argentina	0	0	100	0	1	0	0	1	0
834238	Centeno	Argentina	0	0	100	0	1	0	0	1	0
834246	Gemona	Italy	1	1	100	0	2	1	0	1	0
834251	Coronel Domínguez	Argentina	0	0	100	1	1	0	0	1	0
834255	Guaymallén	Argentina	0	0	100	0	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
834258	Inriville	Argentina	1	1	100	0	1	0	0	1	0
834259	Lobos	Argentina	0	0	100	0	1	0	0	1	0
834260	Olavarría	Argentina	1	1	100	0	3	0	0	1	0
834261	Irapuato	Mexico	0	1	000	1	1	1	0	1	0
834277	Patagones	Argentina	1	1	100	0	3	0	0	1	0
834278	Resistencia	Argentina	1	1	100	0	3	0	0	1	0
834280	Pérez	Argentina	1	1	100	0	2	0	0	1	0
834287	Terra Nova do Norte	Brazil	1	1	100	0	2	1	0	1	1
834289	Rauch	Argentina	1	1	100	0	2	0	0	1	0
834300	Villanueva	Guatemala	0	0	000	1	1	0	0	1	0
834301	San Antonio de Areco	Argentina	1	1	100	0	2	0	0	1	0
834313	Tópaga	Colombia	1	1	100	0	1	1	0	1	0
834323	Patong	Thailand	0	1	000	1	2	1	0	1	1
834347	Seberang Perai	Malaysia	1	1	100	0	4	1	0	1	1
834362	Sigtuna	Sweden	0	1	000	1	2	1	0	1	1
834370	Town of Secaucus/NJ	USA	0	0	100	1	1	0	0	1	0
834374	Tagum/DVO	Philippines	0	1	000	1	2	1	0	1	1
834403	San Martín de los Andes	Argentina	1	1	100	0	3	0	0	1	0
834405	Soldini	Argentina	1	1	100	0	3	0	0	1	0
834406	Tlaquepaque	Mexico	0	0	100	0	1	1	0	1	0
834413	Provincia de Tahuamanú	Peru	0	1	000	0	1	1	0	1	0
838939	İzmit	Turkey	0	1	000	0	1	0	0	1	1
839648	Victoria	Mexico	2	1	100	1	1	0	0	0	0
839650	Uriangato	Mexico	0	1	000	1	1	1	0	1	0
839665	Celaya	Mexico	0	1	000	1	3	1	0	1	1
839666	Escuintla	Guatemala	0	1	000	0	1	1	0	1	1
839667	Guanagazapa	Guatemala	0	1	000	0	1	1	0	1	0
839668	Iztapa	Guatemala	0	1	000	1	1	1	0	1	1

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
839669	San José	Guatemala	0	1	000	1	1	1	0	1	1
839670	Santa Catarina Pinula	Guatemala	0	1	000	0	1	1	0	1	0
839673	Jesús María	Peru	1	1	100	0	1	1	0	1	0
839931	Dong Hoi	Vietnam	0	1	000	0	2	0	0	1	1
839954	Vinh	Vietnam	0	1	000	0	2	0	0	1	1
839963	Alpa Corral	Argentina	0	0	100	0	1	0	0	1	0
839964	Florentino Ameghino	Argentina	0	0	100	0	1	0	0	1	0
839965	Dolores	Argentina	0	0	100	0	1	0	0	1	0
839966	Loncopue	Argentina	0	0	100	0	1	0	0	1	0
839967	Malargüe	Argentina	1	1	100	0	3	0	0	1	0
839970	San Justo/S	Argentina	1	1	100	0	4	0	0	1	0
839971	San Nicolás de los Arroyos	Argentina	0	0	100	0	1	0	0	1	0
839972	Villa Elisa	Argentina	0	0	100	0	1	0	0	1	0
839980	Avellaneda	Argentina	1	1	100	0	2	1	0	1	0
839982	Sepang	Malaysia	0	1	000	0	3	1	0	1	0
840018	Ate	Peru	1	1	100	0	1	0	0	1	0
840024	Perth and Kinross	UK	0	1	000	1	1	1	0	1	0
840030	Reconquista	Argentina	1	1	100	0	3	0	0	1	0
840033	Laborde	Argentina	0	0	100	0	1	0	0	1	0
840034	Morón	Argentina	0	0	100	0	1	0	0	1	0
840036	La Paz	Argentina	0	0	100	0	1	0	0	1	0
840037	Tilisarao	Argentina	0	0	100	0	1	0	0	1	0
840039	Arequito	Argentina	0	0	100	0	1	0	0	1	0
840042	Gislaveds Kommun	Sweden	3	1	100	1	2	1	0	1	1
840070	Somerset West and Taunton	UK	0	1	000	0	1	0	0	1	0
840161	Del Carmen/03	Philippines	0	1	000	1	1	0	0	1	0
840244	Águas da Prata	Brazil	0	1	000	1	1	0	0	1	0
840253	Pedra Bela	Brazil	0	0	000	1	1	0	0	0	0
840269	Town of Whitby/ON	Canada	0	1	000	0	2	1	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
840309	Markaryds Kommun	Sweden	0	1	000	1	1	1	0	0	0
840313	Cerro Navia	Chile	0	0	100	1	1	1	0	0	0
840328	San Pedro Carchá	Guatemala	1	1	100	0	1	1	0	0	0
840349	St Davids	UK	0	0	100	1	1	0	0	1	0
840370	Upplands-Bro	Sweden	0	1	000	1	2	1	0	1	1
840371	Falkoping Kommun	Sweden	1	1	100	0	2	1	0	1	0
840419	Mahasarakham	Thailand	0	1	000	1	2	1	0	1	1
840425	Skövde kommun	Sweden	1	1	100	0	2	1	0	1	0
840490	La Carlota	Philippines	0	1	000	1	1	1	0	1	0
840492	Malolos/03	Philippines	0	1	000	1	1	0	0	1	0
840507	Dura	Palestine	0	0	100	1	1	0	0	0	0
840514	Blitar	Indonesia	0	1	000	0	2	1	1	1	1
840521	Denizli	Turkey	1	1	100	0	1	0	0	1	0
840529	Victoria/TAM	Mexico	0	1	000	1	1	0	0	1	0
840601	San Miguel de Allende	Mexico	1	1	100	0	1	0	0	1	0
840693	Maneiro	Venezuela	0	0	000	1	1	0	0	1	0
840914	Cáceres	Brazil	1	1	100	0	1	0	0	1	0
840916	Igarassu	Brazil	2	1	100	1	1	0	0	1	0
840917	Pau Brasil	Brazil	2	1	100	1	1	0	0	0	0
840918	Pilões/PB	Brazil	0	1	000	1	1	0	0	1	0
840919	Fraiburgo	Brazil	2	1	100	1	1	0	0	0	0
840924	Alexânia	Brazil	0	0	100	1	1	0	0	1	0
840925	Indiaroba	Brazil	2	1	100	1	1	0	0	1	0
840926	Serra Talhada	Brazil	2	1	100	1	1	0	0	0	0
840927	São Cristóvão	Brazil	2	1	100	1	1	0	0	1	0
840930	Venâncio Aires	Brazil	0	0	100	1	1	0	0	1	0
840931	Cordeirópolis	Brazil	1	1	100	0	1	1	0	1	0
840935	Brasiléia	Brazil	0	0	100	1	1	0	0	0	0
840936	Guanhães	Brazil	2	1	100	1	1	0	0	1	0
840937	Epitaciolândia	Brazil	2	1	100	1	1	0	0	1	0

Continued on next page

Table C.1 – continued from previous page

Id	City Name	Country	ERM-L	Data Modeling	Data Acquisition	Data Processing	Data Analysis	Report Building	Report Publishing	Deployment	Monitoring
840938	São Luís de Montes Belos	Brazil	0	0	100	1	1	1	0	1	0
840941	Vila Nova dos Martírios	Brazil	1	1	100	0	1	0	0	1	0
840943	Boa Ventura	Brazil	2	1	100	1	1	0	0	1	0
840944	Carnaúba dos Dantas	Brazil	2	1	100	1	1	0	0	1	0
841003	Ciudad Apodaca	Mexico	0	1	000	1	1	0	0	0	0
841098	Chimbote	Peru	2	1	100	1	1	0	0	1	0
841153	Bellavista/SAM	Peru	2	1	100	1	1	0	0	1	0
841154	Independencia/LIM	Peru	0	0	100	1	1	0	0	1	0
841155	Tarapoto	Peru	0	0	100	1	1	0	0	1	0
841269	Montecarlo	Argentina	0	0	100	0	1	0	0	1	0
841326	Oro Verde	Argentina	1	1	100	0	3	0	0	1	0
841416	Puerto Esperanza	Argentina	1	1	100	0	1	0	0	1	0

Appendix D

OECD Countries List

Table D.1: OECD Countries List

Country	Joined OECD
Australia	1971
Austria	1961
Belgium	1961
Canada	1961
Chile	2010
Colombia	2020
Costa Rica	2021
Czech Republic	1995
Denmark	1961
Estonia	2010
Finland	1969
France	1961
Germany	1961
Greece	1961
Hungary	1966
Iceland	1961
Ireland	1961
Israel	2010
Italy	1962
Japan	1964
South Korea	1966
Latvia	2016
Lithuania	2018
Luxembourg	1961
Mexico	1994
Netherlands	1961
New Zealand	1973
Norway	1961
Poland	1996
Portugal	1961
Slovakia	2000
Slovenia	2010
Spain	1961
Sweden	1961
Switzerland	1961
Turkey	1961
United Kingdom	1961
United States	1961

Appendix E

Example of self-test experiments logs

```
2021-06-13 17:39:41,681 - INFO - ***********************************************************
2021-06-13 17:39:41,682 - INFO - *** MScHelper:ExplorationData ***
2021-06-13 17:39:41,683 - INFO - *** Projeto Kaggle/CDP ***
2021-06-13 17:39:41,683 - INFO - *** Aluno:  Victor de Almeida Xavier ***
2021-06-13 17:39:41,683 - INFO - ***********************************************************
2021-06-13 17:39:41,683 - INFO - [INIT]Module:Experiment
2021-06-13 17:39:41,686 - INFO - [INIT]Processing experiment
parameters:['cluster:', 'ClusWiSARD', 'Grp', '-N', '1000000',
'-d', '2', '-e', '1', '-v', 'Save', '-i', 'ClusWiSARD_N100000_WW_-
0a1a_ExYN_e1', '-D', 'cluster:./input/cdp/cluster_allcities_0a1a_-
ExYN.dat', '-o','config=ww0a1a_ExYN,update_clusters=true,save_-
analytics=true,threshold=auto,discriminatorLimit=auto,sufix=exec_params,dump_-
data=true,configs_log=true']
2021-06-13 17:39:41,688 - DEBUG - [INIT]Loading data sets:cluster:./input/cdp/cluster_-
allcities_0a1a_ExYN.dat
2021-06-13 17:39:41,688 - INFO - [INIT]Loading data set:./input/cdp/cluster_-
allcities_0a1a_ExYN.dat
2021-06-13 17:39:41,689 - DEBUG - [»»]Data set cleared before load!
2021-06-13 17:39:47,769 - DEBUG - [»»]DataSet category distribution:
2021-06-13 17:39:47,769 - DEBUG - [»»] -1:0
2021-06-13 17:39:47,770 - DEBUG - [»»]DataSet category distribution:
2021-06-13 17:39:47,770 - DEBUG - [»»] 1:1
2021-06-13 17:39:47,770 - INFO - Loaded 814 sample points
2021-06-13 17:39:47,771 - INFO - [DONE]Loading data set:./input/cdp/cluster_-
allcities_0a1a_ExYN.dat
2021-06-13 17:39:47,771 - DEBUG - [»»]Loaded?True
2021-06-13 17:39:47,771 - INFO - Loaded 1 data sets
2021-06-13 17:39:47,771 - INFO - Load errors:  0
```

2021-06-13 17:39:47,772 - DEBUG - [DONE]Loading data sets:cluster:./input/cdp/cluster_-allcities_0a1a_ExYN.dat
2021-06-13 17:39:47,772 - DEBUG - [»»]Loaded?True
2021-06-13 17:39:47,772 - INFO - [DONE]Processing experiment parameters:['cluster:', 'ClusWiSARD', 'Grp', '-N', '1000000', '-d', '2', '-e', '1', '-v', 'Save', '-i', 'ClusWiSARD_N100000_WW_-0a1a_ExYN_e1', '-D', 'cluster:./input/cdp/cluster_allcities_0a1a_-ExYN.dat', '-o', 'config=ww0a1a_ExYN,update_clusters=true,save_-analytics=true,threshold=auto,discriminatorLimit=auto,sufix=exec_params,dump_-data=true,configs_log=true']
2021-06-13 17:39:47,773 - DEBUG - [»»]Before running experiment:ClusWiSARD
2021-06-13 17:39:47,774 - DEBUG - [»»]Before running experiment:ClusWiSARD
2021-06-13 17:39:50,817 - INFO - ***
2021-06-13 17:39:50,817 - INFO - Experiment Log
2021-06-13 17:39:50,817 - INFO - ***
2021-06-13 17:39:50,818 - INFO - Experiment Id:20210613173939176087
2021-06-13 17:39:50,818 - INFO - Experiment Info:ClusWiSARD_N100000_WW_0a1a_-ExYN_e1
2021-06-13 17:39:50,818 - INFO - Experiment parameters:
2021-06-13 17:39:50,819 - INFO - Machinery cluster:ClusWiSARD
2021-06-13 17:39:50,819 - INFO - Operation:Grp
2021-06-13 17:39:50,819 - INFO - Number of Sample Points:1000000
2021-06-13 17:39:50,819 - INFO - Dimension:2
2021-06-13 17:39:50,819 - INFO - Number of Executions:1
2021-06-13 17:39:50,820 - INFO - K_Fold:1
2021-06-13 17:39:50,820 - INFO - W_Vector Initial:
2021-06-13 17:39:50,820 - INFO - Verbose:3
2021-06-13 17:39:50,820 - INFO - Options:config=ww0a1a_ExYN,update_clusters=true,save_-analytics=true,threshold=auto,discriminatorLimit=auto,sufix=exec_-params,dump_data=true,configs_log=true 2021-06-13 17:39:50,820 - DEBUG - ***
2021-06-13 17:39:50,821 - INFO - ***
2021-06-13 17:39:50,821 - INFO - *** ClusWiSARD ***
2021-06-13 17:39:50,821 - INFO - ***
2021-06-13 17:39:50,821 - INFO - [INIT]Experiment 20210613173939176087 Started at 2021-06-13 17:39:50.821607.
2021-06-13 17:39:50,822 - DEBUG - [INIT]reset view:cluster
2021-06-13 17:39:50,902 - DEBUG - [INIT]Ini view space:cluster

```
2021-06-13 17:39:50,903 - DEBUG - [INIT]Init view space data:cluster
2021-06-13 17:39:51,191 - DEBUG - [DONE]Init view space data:cluster
2021-06-13 17:39:51,192 - DEBUG - [DONE]Init view space:cluster
...
2021-06-13 17:39:51,794 - INFO - [INIT]Execution 1_1 Started at 2021-06-13
17:39:51.794282.
2021-06-13 17:39:51,794 - DEBUG - [»»]Set Target Function:ClusWiSARD
2021-06-13 17:39:51,795 - INFO - [»»]Set Target Function:ClusWiSARD
2021-06-13 17:39:51,795 - DEBUG - [»»]Set Cluster DataSet:ClusWiSARD
2021-06-13 17:39:51,796 - INFO - [»»]Set Cluster DataSet:ClusWiSARD
2021-06-13 17:39:51,796 - INFO - [INIT]Cloning Clustering_Dataset
2021-06-13 17:39:51,890 - INFO - [INIT]Saving data set:20210613173939176087
2021-06-13 17:39:51,891 - DEBUG - [»»]Data set replaced!
2021-06-13 17:39:52,215 - INFO - Saved 814 sample points
2021-06-13 17:39:52,215 - INFO - [DONE]Saving data set:20210613173939176087
2021-06-13 17:39:52,216 - DEBUG - [»»]Saved?True
2021-06-13 17:39:52,216 - INFO - [DONE]Cloning Clustering_Dataset
2021-06-13 17:39:52,216 - DEBUG - [»»]Clonned?True
2021-06-13 17:39:52,217 - DEBUG - [IMPL]Cluster Dataset Visualziation...
2021-06-13 17:39:52,217 - INFO - [INIT]Clustering using ClusWiSARD...
2021-06-13 17:39:52,217 - DEBUG - [»»]Clustering:ClusWiSARD
2021-06-13 17:39:52,217 - INFO - [»»]Clustering:ClusWiSARD
2021-06-13 17:39:52,218 - DEBUG - [INIT]ClusWiSARD.clustering
2021-06-13 17:39:52,351 - INFO - [»»]ClusWiSARD Parameters:
2021-06-13 17:39:52,351 - INFO - [»»]addressSize:9
2021-06-13 17:39:52,352 - INFO - [»»]minScore:0.1
2021-06-13 17:39:52,352 - INFO - [»»]threshold:auto
2021-06-13 17:39:52,352 - INFO - [»»]discriminatorLimit:auto
2021-06-13 17:39:52,353 - INFO - [»»]bleachingActivated:True
2021-06-13 17:39:52,353 - INFO - [»»]ignoreZero:False
2021-06-13 17:39:52,353 - INFO - [»»]completeAddressing:True
2021-06-13 17:39:52,353 - INFO - [»»]verbose:False
2021-06-13 17:39:52,354 - INFO - [»»]indexes:[]
2021-06-13 17:39:52,354 - INFO - [»»]base:2
2021-06-13 17:39:52,354 - INFO - [»»]returnActivationDegree:False
2021-06-13 17:39:52,354 - INFO - [»»]returnConfidence:False
2021-06-13 17:39:52,355 - INFO - [»»]returnClassesDegrees:False
2021-06-13 17:39:52,355 - INFO - Clustering...
2021-06-13 17:39:52,355 - INFO - [INIT]Tunning exec params
```

```
2021-06-13 17:39:52,356 - DEBUG - _get_threshold_value:  exec_attempt=0, num_-
samples=814, threshold_value=444
2021-06-13 17:39:52,356 - DEBUG - _get_discriminatorLimit_value:  exec_attempt=0,
num_samples=814, discriminatorLimit_value=100
2021-06-13 17:39:52,356 - INFO - Exec tuning with params:  exec_attempt=1,
threshold_value=444, discriminatorLimit_value=100
2021-06-13 17:39:52,357 - DEBUG - Executing Tuning...1
2021-06-13 17:39:52,477 - DEBUG - Clusters:
2021-06-13 17:39:52,477 - DEBUG - Number of clusters found:25
2021-06-13 17:39:52,477 - DEBUG - Cluster:1 Count:184
2021-06-13 17:39:52,477 - DEBUG - Cluster:2 Count:65
2021-06-13 17:39:52,478 - DEBUG - Cluster:3 Count:52
2021-06-13 17:39:52,478 - DEBUG - Cluster:4 Count:47
2021-06-13 17:39:52,478 - DEBUG - Cluster:5 Count:50
2021-06-13 17:39:52,479 - DEBUG - Cluster:6 Count:179
2021-06-13 17:39:52,479 - DEBUG - Cluster:7 Count:159
2021-06-13 17:39:52,479 - DEBUG - Cluster:8 Count:194
2021-06-13 17:39:52,479 - DEBUG - Cluster:9 Count:173
2021-06-13 17:39:52,480 - DEBUG - Cluster:10 Count:177
2021-06-13 17:39:52,480 - DEBUG - Cluster:11 Count:195
2021-06-13 17:39:52,480 - DEBUG - Cluster:12 Count:149
2021-06-13 17:39:52,480 - DEBUG - Cluster:13 Count:154
2021-06-13 17:39:52,481 - DEBUG - Cluster:14 Count:152
2021-06-13 17:39:52,481 - DEBUG - Cluster:15 Count:147
2021-06-13 17:39:52,481 - DEBUG - Cluster:16 Count:125
2021-06-13 17:39:52,481 - DEBUG - Cluster:17 Count:147
2021-06-13 17:39:52,482 - DEBUG - Cluster:18 Count:131
2021-06-13 17:39:52,482 - DEBUG - Cluster:19 Count:125
2021-06-13 17:39:52,482 - DEBUG - Cluster:20 Count:128
2021-06-13 17:39:52,482 - DEBUG - Cluster:21 Count:93
2021-06-13 17:39:52,483 - DEBUG - Cluster:22 Count:45
2021-06-13 17:39:52,483 - DEBUG - Cluster:23 Count:61
2021-06-13 17:39:52,483 - DEBUG - Cluster:24 Count:41
2021-06-13 17:39:52,483 - DEBUG - Cluster:25 Count:8
2021-06-13 17:39:52,484 - WARNING - Discriminators activation count (2981)
greater than number of samples (814)!
2021-06-13 17:39:52,484 - DEBUG - Executing Tuning...1 DONE
2021-06-13 17:39:52,484 - INFO - Exec tuning results:  discr_count=25, discr_-
sum=2981
```

```
...
2021-06-13 17:39:52,485 - DEBUG - Executing Tuning...2
2021-06-13 17:39:52,601 - DEBUG - Clusters:
2021-06-13 17:39:52,601 - DEBUG - Number of clusters found:25
2021-06-13 17:39:52,601 - DEBUG - Cluster:1 Count:172
...
2021-06-13 17:39:52,610 - DEBUG - Executing Tuning...3
2021-06-13 17:39:52,765 - DEBUG - Clusters:
2021-06-13 17:39:52,765 - DEBUG - Number of clusters found:25
...
2021-06-13 17:39:53,221 - DEBUG - Clustering dumping...
2021-06-13 17:39:53,223 - DEBUG - [»»]setting path:  ./work/20210613173939176087/dumps
2021-06-13 17:39:53,223 - DEBUG - [»»]filenamepath:  ./work/20210613173939176087/dumps/cluswisard.
2021-06-13 17:39:53,233 - DEBUG - Clustering dumping...OK
2021-06-13 17:39:53,234 - DEBUG - Clustering details...
2021-06-13 17:39:53,235 - INFO - [INIT]Finding Clusters
2021-06-13 17:42:58,128 - DEBUG - [»»]setting path:  ./work/20210613173939176087/dumps
2021-06-13 17:42:58,129 - DEBUG - [»»]filenamepath:  ./work/20210613173939176087/dumps/1093.data
...
```

Appendix F

Example of experimental logs

```
2021-10-10 19:58:44,517 - INFO - [INIT]Module:Preprocess
2021-10-10 19:58:44,518 - INFO - [INIT]Processing experiment
parameters:['cluster:', 'ClusWiSARD', 'CSV', '-N', '1000000', '-d', '2', '-v',
'Save', '-i',
'CDP_Preprocess_WW_Config0a1a4a5a_AllFT', '-in', './input/2019_Emissions_Cities_-
Dataset_SORTED.csv', '-bin', '-csv', '-f', 'I:Question Number=0*,1*,4*,5*',
'-o',
'copy_dat=./input/cdp/cluster_ww_0a1a4a5a_AllFT.dat,copy_-
out=./input/cdp/cluster_ww_0a1a4a5a_AllFT_out.csv,copy_-
stats=./input/cdp/cluster_ww_0a1a4a5a_AllFT_stats.csv']
2021-10-10 19:58:44,518 - INFO - [DONE]Processing experiment
parameters:['cluster:', 'ClusWiSARD', 'CSV', '-N', '1000000', '-d', '2', '-v',
'Save', '-i',
'CDP_Preprocess_WW_Config0a1a4a5a_AllFT', '-in', './input/2019_Emissions_Cities_-
Dataset_SORTED.csv', '-bin', '-csv', '-f', 'I:Question Number=0*,1*,4*,5*',
'-o',
'copy_dat=./input/cdp/cluster_ww_0a1a4a5a_AllFT.dat,copy_-
out=./input/cdp/cluster_ww_0a1a4a5a_AllFT_out.csv,copy_-
stats=./input/cdp/cluster_ww_0a1a4a5a_AllFT_stats.csv']
2021-10-10 19:58:44,519 - INFO - ***********************************************************
2021-10-10 19:58:44,519 - INFO - Experiment Log
2021-10-10 19:58:44,519 - INFO - ***********************************************************
2021-10-10 19:58:44,519 - INFO - Experiment Id:20211010195842884929
2021-10-10 19:58:44,520 - INFO - Experiment Info:CDP_Preprocess_WW_-
Config0a1a4a5a_AllFT
2021-10-10 19:58:44,520 - INFO - Experiment parameters:
2021-10-10 19:58:44,520 - INFO - Machinery cluster:ClusWiSARD
2021-10-10 19:58:44,520 - INFO - Operation:PreProcess
```

```
2021-10-10 19:58:44,520 - INFO - Number of Sample Points:1000000
2021-10-10 19:58:44,520 - INFO - Dimension:2
2021-10-10 19:58:44,520 - INFO - Number of Executions:1
2021-10-10 19:58:44,520 - INFO - K_Fold:1
2021-10-10 19:58:44,520 - INFO - W_Vector Initial:
2021-10-10 19:58:44,521 - INFO - Verbose:3
2021-10-10 19:58:44,521 - INFO - Options:copy_dat=./input/cdp/cluster_ww_-
0a1a4a5a_AllFT.dat,copy_out=./input/cdp/cluster_ww_0a1a4a5a_AllFT_out.csv,copy_-
stats=./input/cdp/cluster_ww_0a1a4a5a_AllFT_stats.csv
2021-10-10 19:58:44,521 - INFO - ***********************************************************
2021-10-10 19:58:44,521 - INFO - *** ClusWiSARD ***
2021-10-10 19:58:44,521 - INFO - ***********************************************************
2021-10-10 19:58:44,521 - INFO - [INIT]Experiment 20211010195842884929 Started at
2021-10-10 19:58:44.521542.
2021-10-10 19:58:45,307 - INFO - [INIT]Execution 1 Started at 2021-10-10
19:58:45.307803.
2021-10-10 19:58:45,308 - INFO - [»»]Set Target Function:ClusWiSARD
2021-10-10 19:58:45,308 - INFO - [»»]Set Input DataSetClusWiSARD
2021-10-10 19:58:45,308 - INFO - [»»]Set Input DataSetClusWiSARD
2021-10-10 19:58:45,308 - INFO - [»»]Preprocessing:ClusWiSARD
2021-10-10 19:58:45,308 - INFO - [»»]Preprocessing:ClusWiSARD
2021-10-10 19:58:45,341 - INFO - [INIT]Hyperparams
2021-10-10 19:58:45,341 - INFO - [PARM]lbda=0.01
2021-10-10 19:58:45,341 - INFO - [PARM]kfold=10
2021-10-10 19:58:45,341 - INFO - [PARM]maxsamples=10000000
2021-10-10 19:58:45,341 - INFO - [PARM]threshold_cost=1
2021-10-10 19:58:45,341 - INFO - [PARM]validation_limit=10
2021-10-10 19:58:45,342 - INFO - [INIT]Preprocessing...
2021-10-10 19:58:45,342 - INFO - [FILE]Processing file ./input/2019_Emissions_-
Cities_Dataset_SORTED.csv
2021-10-10 19:58:45,344 - INFO - [INIT]Processing...
2021-10-10 19:58:47,065 - INFO - Number of lines found:  236936
2021-10-10 19:58:47,065 - INFO - Number of lines to process:  236936
2021-10-10 19:58:47,067 - INFO - Header:  ['Questionnaire', 'Year Reported
to CDP', 'Account Number', 'organisation', 'Country', 'CDP Region', 'Parent
Section', 'Section', 'Question Number', 'Question Name', 'Column Number', 'Column
Name', 'Row Number', 'Row Name', 'Response Answer', 'Comments', 'File Name',
'Last update']
2021-10-10 19:58:47,067 - INFO - [»»] 1 of 236936 lines processed.
```

```
2021-10-10 19:58:47,068 - INFO - [»»] 2 of 236936 lines processed.
2021-10-10 19:58:47,068 - INFO - [»»] 3 of 236936 lines processed.
2021-10-10 19:58:47,068 - INFO - [»»] 4 of 236936 lines processed.
2021-10-10 19:58:47,069 - INFO - [»»] 5 of 236936 lines processed.
2021-10-10 19:58:47,069 - INFO - [»»] 6 of 236936 lines processed.
2021-10-10 19:58:47,069 - INFO - [»»] 7 of 236936 lines processed.
2021-10-10 19:58:47,070 - INFO - [»»] 8 of 236936 lines processed.
2021-10-10 19:58:47,070 - INFO - [»»] 9 of 236936 lines processed.
2021-10-10 19:58:47,070 - INFO - [»»] 10 of 236936 lines processed.
...
2021-10-10 19:58:47,076 - INFO - [»»] 22 of 236936 lines processed.
2021-10-10 19:58:47,076 - INFO - [»»] 23 of 236936 lines processed.
2021-10-10 19:58:47,077 - WARNING - _get_answer_option 1.1a:1 not found; using
aproximation for answer [Individual city commitment]
2021-10-10 19:58:47,094 - WARNING - _get_answer_option 1.1a:1 nearest key choosed
[Individual city Commitment]
2021-10-10 19:58:47,095 - INFO - [»»] 24 of 236936 lines processed.
2021-10-10 19:58:47,095 - INFO - [»»] 25 of 236936 lines processed.
...  2021-10-10 19:58:47,109 - INFO - [»»] 55 of 236936 lines processed.
2021-10-10 19:58:47,109 - INFO - [»»] 56 of 236936 lines processed.
2021-10-10 19:58:47,109 - WARNING - _get_answer_option 5.0a:3 not found; using
aproximation for answer [Smaller - covers only part of the city]
2021-10-10 19:58:47,116 - WARNING - _get_answer_option 5.0a:3 nearest key choosed
[Smaller - covers only part of the city]
2021-10-10 19:58:47,116 - INFO - [»»] 57 of 236936 lines processed.
2021-10-10 19:58:47,117 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.0a_3_-
2):empty answer for select field
2021-10-10 19:58:47,117 - INFO - [»»] 58 of 236936 lines processed.
2021-10-10 19:58:47,118 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.0a_3_-
3):empty answer for select field
2021-10-10 19:58:47,118 - INFO - [»»] 59 of 236936 lines processed.
2021-10-10 19:58:47,119 - INFO - [»»] 60 of 236936 lines processed.
...
2021-10-10 19:58:47,126 - INFO - [»»] 74 of 236936 lines processed.
2021-10-10 19:58:47,127 - WARNING - _get_answer_range 5.0a:9 empty answer when
integer value should be informed!
2021-10-10 19:58:47,127 - INFO - [»»] 75 of 236936 lines processed.
2021-10-10 19:58:47,127 - WARNING - _get_answer_range 5.0a:9 empty answer when
integer value should be informed!
```

```
2021-10-10 19:58:47,128 - INFO - [»»] 76 of 236936 lines processed.
2021-10-10 19:58:47,128 - WARNING - _get_answer_range 5.0a:9 empty answer when
integer value should be informed!
2021-10-10 19:58:47,129 - INFO - [»»] 77 of 236936 lines processed.
2021-10-10 19:58:47,129 - WARNING - _get_answer_range 5.0a:10 empty answer when
integer value should be informed!
2021-10-10 19:58:47,129 - INFO - [»»] 78 of 236936 lines processed.
2021-10-10 19:58:47,130 - WARNING - _get_answer_range 5.0a:10 empty answer when
integer value should be informed!
2021-10-10 19:58:47,130 - INFO - [»»] 79 of 236936 lines processed.
2021-10-10 19:58:47,130 - WARNING - _get_answer_range 5.0a:10 empty answer when
integer value should be informed!
2021-10-10 19:58:47,130 - INFO - [»»] 80 of 236936 lines processed.
2021-10-10 19:58:47,131 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.0a_11_-
1):empty answer for select field
2021-10-10 19:58:47,131 - INFO - [»»] 81 of 236936 lines processed.
2021-10-10 19:58:47,131 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.0a_11_-
2):empty answer for select field
2021-10-10 19:58:47,132 - INFO - [»»] 82 of 236936 lines processed.
2021-10-10 19:58:47,132 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.0a_11_-
3):empty answer for select field
2021-10-10 19:58:47,132 - INFO - [»»] 83 of 236936 lines processed.
2021-10-10 19:58:47,133 - INFO - [»»] 84 of 236936 lines processed.
...
2021-10-10 20:01:25,472 - INFO - [»»] 236537 of 236936 lines processed.
2021-10-10 20:01:25,472 - INFO - [»»] 236538 of 236936 lines processed.
2021-10-10 20:01:25,473 - INFO - [»»] 236539 of 236936 lines processed.
2021-10-10 20:01:25,473 - WARNING - _get_answer_range 5.4:6 empty answer when
integer value should be informed!
2021-10-10 20:01:25,473 - INFO - [»»] 236540 of 236936 lines processed.
2021-10-10 20:01:25,474 - WARNING - _get_answer_range 5.4:6 empty answer when
integer value should be informed!
2021-10-10 20:01:25,474 - INFO - [»»] 236541 of 236936 lines processed.
2021-10-10 20:01:25,474 - WARNING - _get_answer_range 5.4:6 empty answer when
integer value should be informed!
...
2021-10-10 20:01:25,700 - INFO - [»»] 236865 of 236936 lines processed.
2021-10-10 20:01:25,701 - INFO - [»»] 236866 of 236936 lines processed.
2021-10-10 20:01:25,701 - WARNING - [CHCK]_get_answer_map(MAP_SKIPPED_-
```

```
FIELD:4.6b_1_1)
2021-10-10 20:01:25,701 - INFO - [»»] 236867 of 236936 lines processed.
2021-10-10 20:01:25,701 - WARNING - [CHCK]_get_answer_map(MAP_SKIPPED_-
FIELD:4.6b_1_2)
2021-10-10 20:01:25,701 - INFO - [»»] 236868 of 236936 lines processed.
...
2021-10-10 20:01:25,727 - INFO - [»»] 236918 of 236936 lines processed.
2021-10-10 20:01:25,727 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.4_11_-
0):empty answer for select field
2021-10-10 20:01:25,727 - INFO - [»»] 236919 of 236936 lines processed.
2021-10-10 20:01:25,728 - WARNING - _get_answer_range 5.4:12 empty answer when
integer value should be informed!
2021-10-10 20:01:25,728 - INFO - [»»] 236920 of 236936 lines processed.
2021-10-10 20:01:25,728 - WARNING - _get_answer_range 5.4:13 empty answer when
integer value should be informed!
2021-10-10 20:01:25,728 - INFO - [»»] 236921 of 236936 lines processed.
2021-10-10 20:01:25,729 - WARNING - [CHCK]_get_answer_map(SELECT_FIELD:5.4_14_-
0):empty answer for select field
2021-10-10 20:01:25,729 - INFO - [»»] 236922 of 236936 lines processed.
2021-10-10 20:01:25,730 - INFO - [»»] 236923 of 236936 lines processed.
2021-10-10 20:01:25,730 - INFO - [»»] 236924 of 236936 lines processed.
2021-10-10 20:01:25,730 - INFO - [»»] 236925 of 236936 lines processed.
2021-10-10 20:01:25,731 - INFO - [»»] 236926 of 236936 lines processed.
2021-10-10 20:01:25,731 - INFO - [»»] 236927 of 236936 lines processed.
2021-10-10 20:01:25,732 - INFO - [»»] 236928 of 236936 lines processed.
2021-10-10 20:01:25,732 - INFO - [»»] 236929 of 236936 lines processed.
2021-10-10 20:01:25,733 - INFO - [»»] 236930 of 236936 lines processed.
2021-10-10 20:01:25,733 - INFO - [»»] 236931 of 236936 lines processed.
2021-10-10 20:01:25,734 - INFO - [DONE]Processing...OK
2021-10-10 20:01:25,734 - INFO - [»»]218421 lines processed.
2021-10-10 20:01:25,734 - INFO - [»»]18512 lines skipped by filter.
2021-10-10 20:01:25,734 - INFO - [INIT]Classification...
2021-10-10 20:01:29,745 - INFO - [DONE]Classification...OK
2021-10-10 20:01:29,745 - INFO - [INIT]Binarization...
2021-10-10 20:01:29,747 - INFO - [»»]Slot Size:  22
2021-10-10 20:02:03,534 - INFO - [DONE]Binarization...OK
2021-10-10 20:02:03,534 - INFO - [INIT]Saving results...
2021-10-10 20:02:03,534 - INFO - [INIT]Saving Sample Ids output
20211010195842884929.ids...
```

```
2021-10-10 20:02:03,543 - INFO - [DONE]Saving Sample Ids output...OK
2021-10-10 20:02:03,544 - INFO - [INIT]Saving Statistics output
20211010195842884929.stats...
2021-10-10 20:02:03,586 - INFO - [DONE]Saving Statistics output...OK
2021-10-10 20:02:03,586 - INFO - [INIT]Coppying stats
file:./work/20211010195842884929/20211010195842884929.stats ./input/cdp/cluster_-
ww_0a1a4a5a_AllFT_stats.csv ...
2021-10-10 20:02:03,586 - INFO - [DONE]Coppying stats
file:./work/20211010195842884929/20211010195842884929.stats ./input/cdp/cluster_-
ww_0a1a4a5a_AllFT_stats.csv ...OK
2021-10-10 20:02:03,586 - INFO - [INIT]Saving Information output
20211010195842884929.info...
2021-10-10 20:02:03,593 - INFO - [DONE]Saving Information output...OK
2021-10-10 20:02:03,593 - INFO - [INIT]Saving RawData output
20211010195842884929_raw.csv...
2021-10-10 20:02:04,473 - INFO - [DONE]Saving RawData output...OK
2021-10-10 20:02:04,474 - INFO - [INIT]Saving PreClassData output...
2021-10-10 20:02:05,183 - INFO - [DONE]Saving PreClassData output...OK
2021-10-10 20:02:05,183 - INFO - [INIT]Saving CSV output...
2021-10-10 20:02:05,377 - INFO - [DONE]Saving CSV output...OK
2021-10-10 20:02:05,378 - INFO - [INIT]Coppying out file:./work/20211010195842884929/2021101019584
out.csv ./input/cdp/cluster_ww_0a1a4a5a_AllFT_out.csv ...
2021-10-10 20:02:05,378 - INFO - [DONE]Coppying out file:./work/20211010195842884929/2021101019584
out.csv ./input/cdp/cluster_ww_0a1a4a5a_AllFT_out.csv ...OK
2021-10-10 20:02:05,378 - INFO - [INIT]Saving Summary output...
2021-10-10 20:02:05,381 - INFO - [DONE]Saving Summary output...OK
2021-10-10 20:02:05,381 - INFO - [INIT]Saving Field Map output...
2021-10-10 20:02:05,388 - INFO - [DONE]Saving Field Map output...OK
2021-10-10 20:02:05,388 - INFO - [INIT]Saving DAT output...
2021-10-10 20:02:13,285 - INFO - [DONE]Saving DAT output...OK
2021-10-10 20:02:13,285 - INFO - [INIT]Coppying dat file:./work/20211010195842884929/2021101019584
out.dat ./input/cdp/cluster_ww_0a1a4a5a_AllFT.dat ...
2021-10-10 20:02:13,295 - INFO - [DONE]Coppying dat file:./work/20211010195842884929/2021101019584
out.dat ./input/cdp/cluster_ww_0a1a4a5a_AllFT.dat ...OK
2021-10-10 20:02:13,296 - INFO - [INIT]Saving Filtered Lines output...
2021-10-10 20:02:16,193 - INFO - [DONE]Saving Filtered Lines output...OK
2021-10-10 20:02:16,193 - INFO - [DONE]Saving results
2021-10-10 20:02:16,193 - INFO - [»»]Processed 236932 lines
2021-10-10 20:02:16,193 - INFO - [DONE]Processing with ./input/2019_Emissions_-
```

```
Cities_Dataset_SORTED.csv
2021-10-10 20:02:16,193 - INFO - [FILE]236932 records processed.
2021-10-10 20:02:16,193 - INFO - [DONE]Preprocessing
2021-10-10 20:02:16,360 - INFO - [INIT]Saving results.
2021-10-10 20:02:16,360 - INFO - [DONE]Saving results.
2021-10-10 20:02:16,360 - INFO - [DONE]Execution 1 Finished at 2021-10-10
20:02:16.360929.
2021-10-10 20:02:16,361 - INFO - [»»]Execition 1 Elapsed time:  0:03:31.053126
2021-10-10 20:02:16,361 - INFO - [DONE]Experiment 20211010195842884929 Finished
at 2021-10-10 20:02:16.361093.
2021-10-10 20:02:16,361 - INFO - [»»]Experiment 20211010195842884929 Elapsed
time:  0:03:31.839551
2021-10-10 20:02:16,361 - INFO - [INIT]Saving statistics.
2021-10-10 20:02:16,361 - ERROR - 'PreProcess'
2021-10-10 20:02:16,361 - ERROR - [EXPT]get_exp_stats
2021-10-10 20:02:16,361 - ERROR - 'PreProcess'
2021-10-10 20:02:16,364 - INFO - [DONE]Saving statistics.
2021-10-10 20:02:16,364 - INFO - [INIT]Saving Experiment Logs.
2021-10-10 20:02:16,365 - INFO - [DONE]Saving Experiment Logs.
2021-10-10 20:02:16,365 - INFO - [DONE]Module:Preprocess
```