Geographical Information Management and Applications

Driving Sustainable Development in Spanish Municipalities: The Circular City Index as an Open Data Tool for Policymakers in Spain

MSc. GIMA Master's Thesis
Final Thesis Report

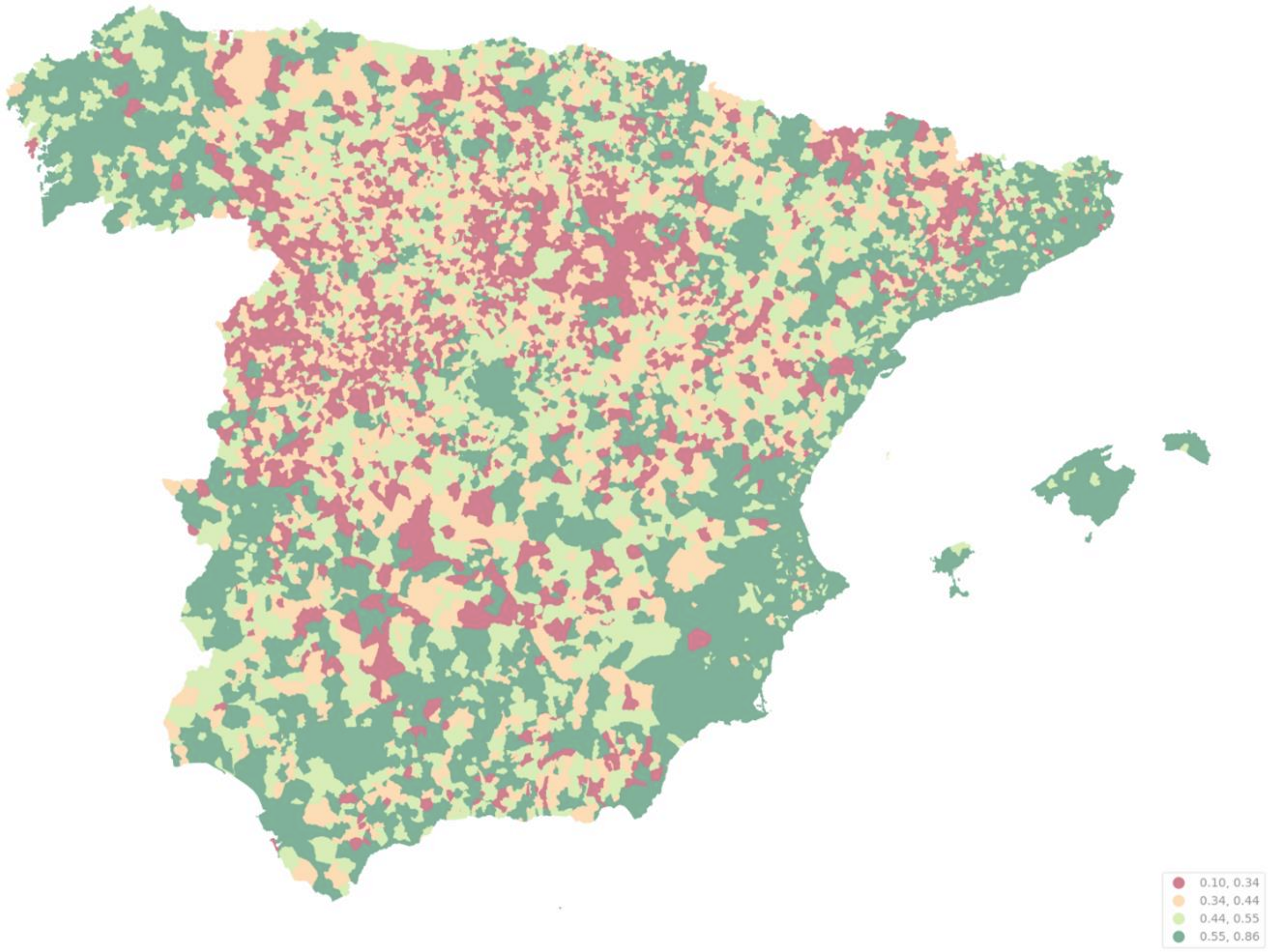
Rens Wiebe van Wijk (0653349)
r.w.vanwijk@students.uu.nl
MSc. Geographical Information Management and Applications (GIMA)
Utrecht University

03-03-2023
Barcelona Supercomputing Center (BSC)

Day-to-day Supervisor: dr. Patricio Reyes (BSC)
University Supervisor: prof. dr. Ioannis N. Athanasiadis (WUR)
Responsible Professor: dr. ir. Ron van Lammeren (WUR)

DRIVING SUSTAINABLE DEVELOPMENT IN SPANISH MUNICIPALITIES: THE CIRCULAR CITY INDEX AS AN OPEN DATA TOOL FOR POLICYMAKERS IN SPAIN

MSc. GIMA Master’s Thesis



Rens Wiebe van Wijk (0653349)
MSc. Geographical Information Management and Applications (GIMA)
Utrecht University

03-03-2023
Barcelona Supercomputing Center (BSC)

Day-to-day Supervisor: dr. Patricio Reyes (BSC)
University Supervisor: prof. dr. Ioannis N. Athanasiadis (WUR)
Responsible Professor: dr. ir. Ron van Lammeren (WUR)

Abstract

In recent years, Spain has made commendable policy commitments to tackle climate change. However, the country ranks lower than other European nations regarding environmental policy and global environmental protection. Urban areas are particularly vulnerable to ecosystem disruptions, and without significant improvements, the situation will only exacerbate. Unfortunately, there is a lack of comprehensive and universal approaches to evaluating sustainable development challenges at city and municipal level. To address this gap, Italian researchers developed the Circular City Index (CCI), which measures urban circularity and green transition based on open data principles. However, the CCI has only been assessed on an Italian case study, and there is a need for universal methods in Spain. This thesis aims to fulfill this need.

This thesis reformulates and redefines the CCI for Spain and evaluates the level of preparedness of Spanish municipalities towards urban circularity and green transition by applying the index. The study uses a quantitative research design with a case study of 8,217 Spanish municipalities, identifying key socio-demographic and economic characteristics that impact municipal performance. The research demonstrates that the CCI is a valuable tool for policymakers to identify areas for improvement and develop policies to address sustainable development challenges in alignment with European targets.

The research findings indicate that population size is the primary socio-demographic characteristic impacting urban circularity, with larger municipalities generally being more prepared for green transition than smaller ones. This implies that policies encouraging population growth and urbanization could lead to better sustainable outcomes. This research aims to fill the current research gap regarding the lack of comprehensive and universal strategies to quantify and evaluate sustainable development challenges at the municipal level and offers insights to assist policymakers in defining policies to address these challenges. The thesis is fully reproducible, scalable, and built on open-source data, methodologies, and software.

Preface

I present to you my thesis entitled *"Driving Sustainable Development in Spanish Municipalities: The Circular City Index as an Open Data Tool for Policymakers in Spain."* This research was conducted between September 2022 and March 2023 as part of the master's program in Geographical Information Management and Applications (GIMA). This research was performed in collaboration with the Barcelona Supercomputing Center-Centro Nacional de Supercomputación (BSC-CNS), a Spanish multidisciplinary research center specializing in high-performance computing, and supported by SoBigData++ (Gr. No. 871042).

I am grateful to the individuals who provided guidance and support throughout this research process. Patricio Reyes, my day-to-day supervisor from BSC-CNS, guided me through the challenging research process and provided me with support when the research became more complex. Ioannis N. Athanasiadis, my supervisor from Wageningen University and Research (WUR), provided me with useful advice and feedback. Angelo Facchini, researcher at the IMT School for Advanced Studies Lucca and one of the founders of the Circular City Index, provided me with the necessary in-depth information to carry out the research.

For policymakers, I hope that this thesis will assist you in developing policies that promote sustainable development in Spanish municipalities. For researchers, I hope that this thesis inspires you to continue future work towards meeting the growing necessity for a comprehensive and universal approach to evaluate sustainable development challenges at the municipal level.

Finally, do not forget to check out the data pipeline, it is open source![1]

Sincerely,

Rens Wiebe van Wijk

[1] https://doi.org/10.5281/zenodo.7682093

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criterion
CCI	Circular City Index
EEA	European Environment Agency
GSA	Global Spatial Autocorrelation
GWR	Geographically Weighted Regression
KNN	K-Nearest Neighbor
KPI	Key Performance Indicator
LSA	Local Spatial Autocorrelation
MAE	Mean Absolute Error
MAUP	Modifiable Areal Unit Problem
OLS	Ordinary Least Squares
RMSE	Root Mean Squared Error
SDG	Sustainable Development Goal
UTM	Universal Transverse Mercator
VIF	Variance Inflation Factor

1. Introduction

1.1 Context

From 1960 to 2015, the global economy grew more than sixfold, while the world's ecosystem decreased by 60% (Jackson, 2011; Our World in Data, 2017). A similar trend was concluded by the IPBES Global Assessment Report on Biodiversity and Ecosystem Services, which analyzed 15,000 sources, and found that approximately 75% of the land-based environment and 66% of the marine environment have been drastically disturbed by the effects of economic growth and the corresponding human actions (IPBES, 2019). Most of these disruptions take place in urban areas, which are home to 50% of the global population and are globally accountable for 80% of the GDP, 75% of the energy-related CO_2 pollution, and 50% of the waste (IPBES, 2019; Lucertini & Musco, 2020; Merino-Saum et al., 2020; Muscillo et al., 2021).

These cautionary numbers emphasize the need to address sustainable development challenges in urban areas. Over the last few years, Spain has made major policy commitments to mitigate and adapt to climate change, including becoming carbon neutral by 2050, reducing GHG emissions by 90%, and utilizing only renewable sources for energy production (Banco de España, 2021; Kölling et al., 2022). The foundations of these commitments are the Climate Change and Energy Transition Law, the National Energy and Climate Plan 2021-2030, and the National Plan for Adapting to Climate Change (Banco de España, 2021; BOE, 2021, 2021; Ministerio para la Transición Ecológica y el Reto Demográfico, 2020). Despite Spain its significant policy commitments to address climate change, the country ranks fourteenth on environmental policy, and tenth on global environmental protections in comparison to other European countries (Kölling et al., 2022).

1.2 Problem Statement

The rankings of Spain's environmental policy and global environmental protections highlight the need for improvements on governing sustainable development challenges at all administrative levels. One concept that strives to tackle these sustainable development challenges is the circular economy concept, which enables the quantification of environmental effects, and introduces innovative strategies to deal with emerging climate challenges (Petit-Boix & Leipold, 2018; United Nations, 2019). In the last decade, the circular economy concept gained interest in the field of sustainable development, and the European Commission has made it a major component of the Urban Agenda, the Partnership on Circular Economy, and the Circular Economy Action Plan (European Commission, 2020; Urban Agenda Partnership on Circular Economy, 2019). The United Nations has also incorporated the concept in the Sustainable Development Goals (SDGs), specifically targeting the targets of SDG 6 (clean water and sanitation), SDG 7 (affordable and clean energy), SDG 8 (decent work and economic growth), SDG 12 (responsible consumption and production), and SDG 15 (life on land) (Schroeder et al., 2019; United Nations, 2015).

Despite the growing interest, the circular economy concept is still a relatively new area of research. Researchers have incorporated the concepts such as circularity, sustainability, technology, and urban metabolism in the circular economy concept to quantify sustainable development challenges in urban areas (Kębłowski et al., 2020; Muscillo et al., 2021; Remøy et al., 2019; Sánchez Levoso et al., 2020). However, there is ongoing debate among researchers and policymakers regarding the interpretation and implementation of the circular economy concept, as the overlap with other related concepts is vague and ambiguous (Geissdoerfer et al., 2017; Mhatre et al., 2021). As result, current literature is lacking a distinctive and unequivocal set of strategies to quantify and evaluate sustainable development challenges at the scale of cities and municipalities (Muscillo et al., 2021; Petit-Boix & Leipold, 2018).

To address this gap, researchers from the universities of Siena, Bari and Lucca developed the Circular City Index (CCI) (Enel, 2021; Muscillo et al., 2021). The index is based on open data principles and defines four key areas that can influence circularity and green transition at municipal level: digitalization, environment and energy, mobility, and waste. The CCI is designed to enable data assessment of various elements related to the performance of municipalities in these four areas, thereby supporting the formulation of green policies at both local and national level (Muscillo et al., 2021).

1.3 Research Objectives

The main objective of this thesis is to reformulate and redefine the CCI for application in the Spanish context, with a focus on evaluating the level of preparedness of Spanish municipalities towards urban circularity and green transition. To achieve this, it is necessary to improve the interpretability of the index results by identifying key socio-demographic and economic characteristics that impact the performance of municipalities. The research aims to fill the current research gap regarding the lack of comprehensive and universal strategies to quantify and evaluate sustainable development challenges at municipal level and offer insights to assist policymakers in defining policies to address these challenges. Figure 1 visualizes the research objectives.

The reformulation and redefinition of the original CCI methodology is essential to effectively apply the index to the Spanish context since local contexts may differ from the original situation. The development of non-spatial and spatial models could support the identification of significant socio-demographic and economic characteristics that impact the level of preparedness of Spanish municipalities in relation to urban circularity and green transition. Understanding these characteristics might improve the interpretability of the performance of municipalities, thereby supporting policy-makers in defining effective green policies to address sustainable development challenges.

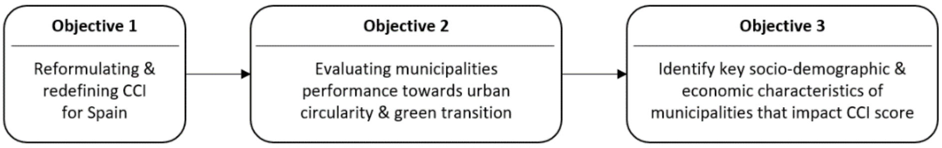


Figure 1 - Research Objectives of the Thesis.

1.4 Research Questions

In line with these research objectives, the following main research question is formulated:

What is the level of preparedness of Spanish municipalities towards urban circularity and green transition, according to the Circular City Index, and how do socio-demographic and economic characteristics impact this level of preparedness?

To realize the general research objective of this thesis, and to answer the main research question, three research sub-questions are formulated:

Sub-question 1: "What extent of reformulation and redefinition is required for the original Circular City Index to be effectively applied to the Spanish context?"

Sub-question 2: "How does the Circular City Index evaluate the preparedness of Spanish municipalities towards urban circularity and green transition?"

Sub-question 3: "What are the key socio-demographic and economic characteristics that impact the performance of Spanish municipalities in the Circular City Index, as revealed by non-spatial and spatial models, and how could these improve the interpretability of the results of the Circular City Index?"

1.5 Research Design

This study employs a quantitative research design with a spatial case study. This approach is selected to improve the reproducibility and scalability of the findings. Each sub-question is designed to answer one part of the main research question, as visualized in Figure 2.

The first sub-question focuses on reformulating and redefining the original CCI methodology to suit the Spanish context. Data is identified and collected to test the suitability of indicators, and where necessary, modifications are made to guarantee the accurate measurement of the index. The second sub-question concentrates on the application of the modified index to a Spanish case study and aims to evaluate the performance of each municipality by (spatial) analyses. The third sub-question centers on the development and employment of non-spatial and spatial models to identify key socio-demographic and economic characteristics correlating with the degree of urban circularity.

In accordance, this thesis study adopts a multi-level approach to study the urban circularity in Spain, with the unit of analysis being the circularity of economies at all administrative levels. To ensure this, the unit of observation is the Key Performance Indicators (KPIs) at the municipal level. The gathering and monitoring data at the municipal level enables aggregation to higher administrative levels, while avoiding ecological fallacy. This approach ensures that the findings of this study accurately reflect the urban circularity across administrative levels in Spain.

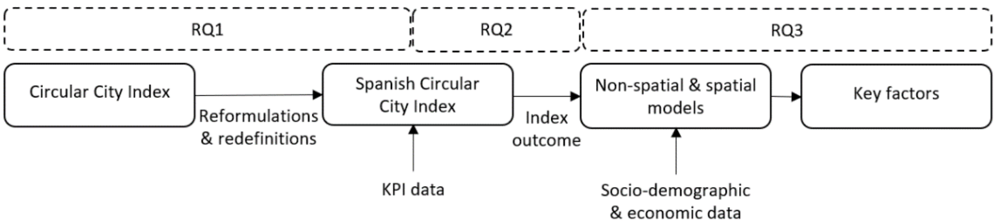


Figure 2 – Simplified Research Design.

1.6 Significance of Thesis

This study represents an important contribution to the emerging field of circular economies by modifying and testing a universal set of strategies for quantifying and evaluating sustainable development challenges at the municipal level. This development is important given the growing necessity for a more comprehensive and universal approach for quantifying and evaluating sustainable development challenges at the municipal level. The modified CCI methodology in combination with the non-spatial and spatial models developed in this study will provide a tool for policymakers to in-depth understand local contexts, identify areas of improvement and align their policies with European targets, as the index incorporates aspects of the European Union's Green Deal. The transparency, reproducibility and scalability of the results will ensure that all stakeholders can benefit from this open data approach, leading to more effective and sustainable solutions to local challenges[2].

1.7 Structure

The rest of this thesis is organized as follows. Chapter two provides an extensive literature review of the fundamental concepts relevant to this thesis. Chapter three offers a comprehensive overview of the methodological framework, and chapter four provides a detailed description of the data. Chapter five presents the results of this thesis. Chapter six critically discusses the results and implications of this study. Lastly, chapter seven draws the conclusions and provides recommendations for future work.

[2] https://doi.org/10.5281/zenodo.7682093

2. Literature Review

As introduced in the introductory chapter, the main research objective of this study is to reformulate and redefine the original CCI to assess the preparedness level of Spanish municipalities to urban circularity and green transition. In addition, the research aims to improve the interpretability of the index' results by identifying important socio-demographic and economic characteristics that impact the CCI performance of municipalities. To achieve this objective, a literature review is necessary.

In this chapter, the main concepts of the thesis are presented through a review of the current literature. First, the central circular economy concept is discussed, including its theoretical foundations and practical implications. Secondly, the current developed frameworks to assess circularity of economies are reviewed, including the CCI, considering their strengths and limitations. By conducting a comprehensive review of the literature, this chapter aims to provide a clear and insightful understanding of the theoretical and practical aspects of the central concept, as well as the different approaches used to measure circularity in the economy.

2.1 Circular Economy

The circular economy concept is one of the main aspects in this thesis. Over the past years, the circular economy concept has gained attention in the knowledge domain of sustainable development (European Commission, 2020; Schroeder et al., 2019; Tura et al., 2019; United Nations, 2015; Urban Agenda Partnership on Circular Economy, 2019). This attention resulted out of the idea that the spillover effects of instant economic expansion and urbanization are approaching physical thresholds, which make the shift towards a more sustainable future crucial (Fan et al., 2019; Tura et al., 2019). An effective resource and utility management approach is needed for this shift, focusing on the minimization of waste, pollution, and losses of resources (Fan et al., 2019). The development of a circular economy could fulfill this need. As an instrument at different implementation scales, the circular economy could enable the quantification of environmental effects and introduces innovative strategies to deal with emerging sustainable development challenges (Petit-Boix & Leipold, 2018; United Nations, 2019).

In general, the circular economy concept refers to the transition from a traditional linear economy towards a circular economy. According to this concept, the linear economy only follows a loop of taking natural resources, producing materials, and disposing waste. In contrast, the circular economy intends to stop the loop by ending, decelerating, and reducing the resource flows and ideally return back to the production stage when a product's life is ended (Nikolaou & Tsagarakis, 2021). To add on this, Ghisellini et al. (2015) recognize three general levels of intervention of the circular economy concept. The micro-level refers to single companies or customers, the meso-level implies ecological industrial parks, and the macro-level represents administrative areas, such as nations or cities (Ghisellini et al., 2015). In this paper, the focus will be on the macro-level of intervention of the circular economy concept.

Nevertheless, researchers and policymakers debate about the exact interpretation and application of the concept on macro-level (Geissdoerfer et al., 2017; Mhatre et al., 2021). In some studies, the concepts of circular economy and sustainability are considered to be identical, while other studies distinguish differences between these concepts (Nikolaou et al., 2021). In line with this, the overlap between the circular economy concept and other related concepts is often vague and ambiguous. Various studies integrated the concepts of circularity, environmental sustainability, technology, and urban metabolism in the circular economy concept (Fan et al., 2019; Kębłowski et al., 2020; Muscillo et al., 2021; Remøy et al., 2019; Sánchez Levoso et al., 2020). Kirchherr et al. (2017) theorized the circular economy concept by performing a systematic analysis on 114 concept definitions. The researchers concluded that the concept is most often defined as reduce, reuse, and recycle activities. In addition, economic prosperity and environmental equality are mentioned most frequently as objective of the concept (Kirchherr et al., 2017). Recent studies also strive for the implementation of a more sustainable format, the sustainable circular economy concept. This concept strongly focuses on transforming consumption patterns and producer responsibilities to develop a circular economy without negative environmental impacts of waste (Maitre-Ekern, 2021). In reaction to these varying interpretations of the concept, Korhonen et al. (2018) state that the concept has become an "essentially contested concept", making it impossible to outline it by one single definition. In line with this, Kirchherr et al. (2017) think that the substantially differing explanations of the circular economy concept could sooner or later cause the end of the concept.

In addition, researchers disagree about the focus of indicators to assess the circular economy concept. Fan et al. (2019) emphasize that the key indicators should focus on the aspects of energy and water management, waste management, and green policy and pollution minimization management. In contrast, Tura et al. (2019) argue that an framework should contain indicators that drive or limit circular economies, concentrating on the environment, economy, society, institutions, technology and information, supply chain, and organizations. Paoli et al. (2022) claim that indicators should focus on waste, energy, water, green, food, buildings, and mobility. In addition, other cross-sectional indicators with an environmental, economic and financial, or social and cultural dimension should be taken into account (Paoli et al., 2022). Based on these studies, corresponding indicators could be identified. These corresponding indicators could be categorized into the seven topics: waste, energy, water, environment, mobility, information and technology, and economy. These categories could be used to evaluate indices and frameworks relating to circular economies. Therefore, this thesis uses these topics to test indices by creating an indicator criteria framework, which is outlined in Table 1. In short, to enable the assessment of circular economies at macro-level, an index is needed that implements a set of indicators in line with this indicator criteria framework.

Table 1 - Indicator Criteria Framework for Testing Circular Economy Indices.

Categorized topic	**Sub-topics examples**
Waste	e-waste; food waste; industrial waste; municipal waste; (solid) waste generation; waste recycle;
Energy	energy consumption; energy dependence; energy productivity; industrial energy consumption; renewable energy
Water	water consumption; water irrigation; water leakage; wastewater recycling; water withdrawal
Environment	CO_2 emissions; green policy; natural resource renewables; NO_x emissions; NO_2 emissions; emission reduction targets; SO_2 emissions
Mobility	15-minute city; cycleways; e-charging stations; footways; pedestrian areas; public transport
Information and Technology	data accessibility; environmental technology; green institutions; green organizations; internet broadband connection; public digital identity system; technology patents;
Economy	circular jobs; green employment; green finance; green economic growth; green economic investments; supply chain

2.2 Circular Economy Indices and Frameworks

By incorporating different sets of indicators, various indices and frameworks have been developed to assess the circularity of economies on different macro-levels. At national level, fourth indices have been discussed in literature. The European Commission (2018) implemented a Circular Economy Monitoring Framework to monitor the degree of circularity of national economies within the European Union. The framework consists of four thematic areas: production and consumption, waste management, secondary raw materials, and competitiveness and innovation (European Commission, 2018). Additionally, the Global Green Growth Institute (GGGI) established the Green Growth Index (GGGI, 2020). Divided into four main categories, the index intends to evaluate the performance of countries in realizing the targets of the SDGs, which are partly in line with the idea of the circular economy concept. The main categories consist of efficient and sustainable resource use, natural capital protection, green economic opportunities, and social inclusion. In addition, Geng et al. (2012) discussed the national circular economy indicator system of the Chinese National Development and Reform Commission. At macro-level, the indicator system contains of four categories, focused on resource output rate, resource consumption rate, integrated resource utilization rate, and waste disposal and pollutant emission (Geng et al., 2012). Finally, Su et al. (2013) built upon this Chinese indicator system, and created a standardized index that added four extra indicators to the system. To better measure the circularity of economies at city level, they proposed to add the indicators on resource efficiency and waste discharge, waste treatment, and waste reclamation.

At regional level, the following two indices have been proposed by researchers. First, Guo-gang (2011) developed the evaluation system of regional circular economy development level, which aims to evaluate the circularity of economies of single regions in China. The evaluation index focusses on the topics of resources consumption, environmental disturbance, recycling, and social development. Based on analytic hierarchy process (AHP), the indicators related to water use, energy consumption and waste utilization are given the highest weight-factor (Guo-gang, 2011). In line with the first index, Chun-rong & Jun (2011) developed the comprehensive index of circular economy to evaluate the circular economy development in China. The index focuses on economic, environmental, and social aspects of circular economies, and these topics are assessed by the categorization into reduce, recycle, and reuse indicators. By using five grade standards, the continuous datasets are categorized to represent the degree of circularity in a region.

At city level, three indices must be highlighted. To begin with, Holcim & Bloomberg Media (2022) developed the Circular Cities Barometer to better understand how urban areas are managing the transition to a circular economy. The barometer groups indicators into four categories: buildings, systems, living, and leadership. Holcim & Bloomberg Media (2022) weighted their data based on the three criteria: the impact of the category or indicator on circular economies, the quantity of cities reporting the specific indicator data, and the quantity of cities participating in the given indicator. Additionally, Nurdiana et al. (2021) created a framework to determine the circularity of cities in Indonesia, based on a systematized literature review. The designed framework contains twenty-three equally weighted indicators, separated over the following six categories: economic, social, environment and energy, environment and land, environment and water, and environment and pollutant. Lastly, Zaman & Lehmann (2013) suggested a tool - the Zero Waste Index - to measure the potency of raw materials to be compensated by zero waste management systems in cities. By measuring this potency, it could be clarified how circular economies are at city level.

Table 2 - Overview of the Circular Economy Indices at Macro-Level.

Research	Index	Total categories	Total indicators (sub-indicators)	Macro-level of measurement
European Commission (2018)	Circular Economy Monitoring Framework	4	10 (19)	National
GGGI (2020)	Green Growth Index	4	16 (36)	National
Geng et al. (2012)	National Circular Economy Indicator System	4	22	National
Su et al. (2013)	Circular Economy approach for China	4	9	National
Guo-gang (2011)	Evaluation System of Regional Circular Economy Development Level	4	16	Regional
Chun-rong & Jun (2011)	Comprehensive Index of Circular Economy	3	10	Regional
Holcim & Bloomberg Media (2022)	Circular Cities Barometer	4	12	City level
Nurdiana et al. (2021)	Circular Cities in Indonesia	6	23	City level
Zaman & Lehmann (2013)	Zero Waste Index	2	6	City level

2.2.1 Circular City Index

In line with the shortcomings of the existing indices, the urgency for the development of a unique and unequivocal set of strategies to assess the circularity of economies at city level has grown. To tackle this urgent challenge, Muscillo et al. (2021) developed the CCI and tested the index in a case study of all municipalities in Italy. The index aims to 'provide data and a succinct measurement of the attributes related to municipalities performances that can support the definition of green policies at national and local level' (Muscillo et al., 2021, p. 1). In line with open data principles, the researchers identified four sets of KPIs that may impact circularity and green transition in cities and municipalities. As showed in Figure 3 and Table 3, these sets of KPIs concentrate on digitalization; energy, climate, and resources; mobility; and waste.

After validating the index, the findings proved discrepancies under the territorial and urban-size points of view. Therefore, results of the CCI could provide useful knowledge to researchers and policy-makers when defining effective investment tools and developing policies (Muscillo et al., 2021). In addition, the researchers emphasize the index' usefulness to 'enhance enabling factors of the green transition that may differ across territories, helping policymakers to promote a smooth and fair transition by fostering the preparedness of municipalities in addressing the challenge' (Muscillo et al., 2021, p. 2). By integrating existing offered open data and citizen science, the CCI could fulfill this urgency to develop a unique and unequivocal strategy set to successfully executes, manages, and evaluates indicators that characterize the circularity of economies at city level.

Figure 3 - Schematic diagram of areas and KPIs composing the CCI. Source: Muscillo et al. (2021, p. 9).

2.2.2 Benefits and Drawbacks of Indices

Analyzing the benefits and drawbacks of the indices is important to evaluate their ability to perform a complete evaluation. Except for the CCI, each index lacks a focus on one or more essential elements to enable the measurement of circular economies at different macro-levels. At national level, the aforementioned indices are unable to fully implement circular economy measurement. The Circular Economy Monitoring Framework of European Commission (2018) complicates comparisons between nations, since a weighted index total per entity is missing. Also, the index focuses mainly on resources and waste management, overlooking topics of energy, water, environment, and mobility. The Green Growth Index largely focusses on non-urban areas on national level, while taking no notices of topics related to mobility. In addition, even if the additional proposed indicators by Su et al. (2013) are included, the Chinese national circular economy indicator system is lacking attention on sustainable development challenges regarding to the topics of mobility, information and technology, and economy.

At regional level, the indices of Guo-gang (2011) and Chun-rong & Jun (2011) merely focus on topics related to waste, energy, water, and the environment. These indices ignore therefore at least the themes linked to mobility, information, and technology. In addition, the indices lack the ability to analyze cities or local administrative areas due to their regional focus.

At city level, the framework for Circular Cities in Indonesia and the Zero Waste Index are missing vital topics to measure the circularity of economies. Both are overlooking the topics of mobility and information and technology. The Circular Cities Barometer of Zaman & Lehmann (2013) is the only index capable to incorporate six out of seven categorized topics at city level. However, the index does not consider topics related to information and technology. In conclusion, these indices and frameworks are currently unsuitable to evaluate the circularity of economies.

In contrast to the previously discussed indices and frameworks, the CCI considers a broader range of circular economy related topics: the CCI incorporates six out of seven categorized topics of the indicator criteria framework to measure the economic circularity. The index' Digitalization set of KPIs covers the topic of information and technology. The Energy, Climate and Resources set relates to the topics of energy, water, and environment. The Mobility and Waste sets of KPIs deal with the corresponding topics of mobility and waste. The only indices that almost incorporate all topics six out of seven subjects are the Green Growth Index and the Circular Cities Barometer. However, the Green Growth Index measures its indicators at a national level. At city level, the only comparable index to the CCI is the Circular Cities Barometer. Though, this barometer has not yet been applied at the municipal level, which gives the CCI an advantage for multi-level implementation.

Table 3 - Categorized Topics of the Circular Economy Incorporated by the Indices.

Research	**Incorporated Circular Economy Topic**						
	Waste	**Energy**	**Water**	**Environment**	**Mobility**	**Information & Technology**	**Economy**
European Commission (2018)	✓					✓	✓
GGGI (2020)	✓	✓	✓	✓		✓	✓
Geng et al. (2012)	✓	✓	✓	✓			
Su et al. (2013)	✓	✓	✓	✓			
Guo-gang (2011)	✓	✓	✓	✓			✓
Chun-rong & Jun (2011)	✓	✓	✓				
Holcim & Bloomberg Media (2022)	✓	✓	✓	✓	✓		✓
Nurdiana et al. (2021)	✓	✓	✓	✓			✓
Zaman & Lehmann (2013)	✓	✓	✓	✓			
Muscillo et al. (2021)	✓	✓	✓	✓	✓	✓	

3. Methodology

In the introductory chapter the general research objective of this thesis is presented. In the second chapter, the circular economy concept and the circular economy index are theorized, including the CCI. In this third chapter, the methodological framework of this research is operationalized.

This chapter is structured in line with the sub-questions of this thesis. In the first section, the methodological framework of the original CCI was described that assessed the extent of reformulations and redefinitions needed to apply the index in the Spanish context, in line with the first sub-question. Afterwards, the methods used for the spatial analysis of the index outcomes are described, which enable the evaluation of the preparedness of municipalities, in line with the second sub-question. Thirdly, the methodology of the data models is outlined, which supports the identification and explanation of the socio-demographic and economic characteristics that impact the index outcome, in line with the third sub-question.

It is important to notice that this chapter focuses on discussing the methodology rather than delving into the data itself. Since data processing constitutes a significant part of this thesis, it has been decided to dedicate a separate chapter to this in order to provide a more comprehensive understanding of the steps taken.

3.1 Circular City Index

The first section of this methodology support sub-question one, aiming to assess the extent of reformulation and redefinition needed to effectively apply the CCI to the Spanish context. As a tool, the index quantifies sustainable development challenges by indicators. By dividing the index into measurable sub-levels, the index provide insights into the degree of urban circularity readiness of municipalities and could encourage policymakers to improve municipality's urban circularity readiness by implementing green transition policies. Muscillo et al. (2021) identified four sets that could indicate the level of preparedness towards urban circularity and green transition in cities and municipalities:

- Digitalization (D)
- Energy, climate, and resources (ECR)
- Mobility (M)
- Waste (W)

These sets could be categorized into two types: (i) factors that have a direct impact; and (ii) factors that have an indirect impact. The first group consists of tools that can directly influence sustainable and circular outcomes when implemented in policies. The ECR and W sets belong to this group. The second group comprises factors that indirectly impact urban circularity and green transition by supporting the observation and implementation of policies. The D and M sets belong to this second group (Muscillo et al., 2021).

3.1.1 Key Performance Indicators (KPIs)

Muscillo et al. (2021) subdivided the 4 sets into 17 KPIs to measure the related factors. Table 4 presents this subdivision. By utilizing KPIs, the index assessed the factors that influence the level of urban circularity and green transition at the municipal level, either directly or indirectly. In this thesis, the data had to be collected in line with these indicators.

Furthermore, the researchers applied different weights to each set and KPI in the index, dependent on their specific importance. Specifically, the weight of each set and KPI increased when they held greater relative significance for policy-makers, political agenda, or environmental standards (Muscillo et al., 2021). The collected KPI data was weighted based on this defined weighting system.

Table 4 - Overview of (Sets of) KPIs in original CCI. Source: (Muscillo et al., 2021)

Level	Sub-level	KPI	Definition
Digitalization (D)	Public digital identity system	D1	Presence in ANPR (public digital service platform)
		D2	Adoption of SPID in PA digital properties
	Broadband internet connection	D3	Percentage of people with broadband connection (>30Mb/s)
	Data accessibility	D4	Accessibility of local government digital properties
Energy, Climate and Resources (ECR)	Emission reduction targets	ECR 1	Covenant of Mayors - Subscription
		ECR 2	Covenant of Mayors - Level of commitment
	Energy consumption	ECR 3	Percentage self-consumption (household only)
	Air quality	ECR 4	Annual average concentration of PM10
		ECR 5	Annual average concentration of NOx
	Water efficiency	ECR 6	Percentage of water leaks
Mobility (M)	Pedestrian areas	M1	Pedestrian areas (m2/100 inhab.)
	Charging stations for electric vehicles	M2	Charging stations (charging station/1,000 inhab.)
	Cycle lanes	M3	Cycleways (km/100 km2)
	Public transportation availability	M4	Bus stops (bus stops/100 inhab.)
Waste (W)	Production of solid waste	W1	Per capita production of solid waste (t/inhab.)
	Recycling of waste	W2	Percentage of solid waste recycling
	Collection of e-waste	W3	Collection of e-waste

3.1.2 Index Formula

To assess the level of preparedness of municipalities towards urban circularity and green transition, Muscillo et al. (2021, p. 9) introduced an index formula. This was measured by quantifying the direct or indirect factors that influences these processes at municipality level, and weighting these factors, based on its relative significance to policymakers. The following formula was implemented by Muscillo et al. (2021, p. 9):

$$CCI_c = \Sigma_{A\epsilon\ Areas}(W_A \Sigma_{k\ \epsilon\ KPI(A)} W_k \times S_{kc})$$

The CCI value for each city or municipality, denoted as CCI_c, was quantified by Muscillo et al. (2021) based on a specific area by A, a set of KPIs by $KPI(A)$, and a particular k in $KPI(A)$. In addition, weightings were calculated for area A by $W_A\epsilon$, and for k in by $W_k\epsilon$. Accordingly, for each area A, the sum of weights was equal to 1 ($\Sigma_{A\epsilon\ Areas} W_A\ = 1$). The same applied for each particular k in area A, as the sum of weights was equal to 1 ($\Sigma_{k\ \epsilon\ KPI(A)} W_k = 1$). In short, a higher weight suggested greater importance, and a higher outcome for CCI_c indicated a better score on the CCI.

Finally, to determine the score for each municipality, the computed area value S_{kc} of each municipality had to be linked to the aforementioned equations (Muscillo et al., 2021). Before the area value S_{kc} could be calculated, the relative value of each index area had to be determined by using a benchmark. These benchmarks were built upon current and future policies in Italy as well as in Europe. The weights and benchmarks used by the original CCI are presented in Table 5.

Table 5 - Weights and Benchmarks per KPI in Original CCI. Source: (Muscillo et al., 2021)

Level	**Level Weight**	**KPI**	**KPI weight**	**Benchmark**	**Ideal Outcome**
Digitalization (D)	0.2	D1	0.3	Yes	Yes
		D2	0.3	Yes	Yes
		D3	0.3	High	100%
		D4	0.1	High	High
Energy, Climate and Resources (ECR)	0.3	ECR 1	0.2	Yes	Yes
		ECR 2	0.2	2020-30	2020-30
		ECR 3	0.3	55%	100%
		ECR 4	0.1	40μg/m3	0μg/m3
		ECR 5	0.1	40μg/m3	0μg/m3
		ECR 6	0.1	Low	0
Mobility (M)	0.2	M1	0.2	900 m^2 / 100	900 m^2 / 100
		M2	0.3	1 / 1,000	1 / 1,000
		M3	0.2	100 km / 100 km^2	100 km / 100 km^2
		M4	0.3	1 / 100 inhabitants	1 / 100 inhabitants
Waste (W)	0.3	W1	0.4	Low	0 t
		W2	0.4	65%	100%
		W3	0.2	Yes	Yes

By implementing these benchmarks in an area value computation formula, the area value S_{kc} could be computed. Muscillo et al. (2021) differentiated five formulas to compare the KPI values with the defined benchmarks. The outcomes were normalized between 0 (no match with the benchmark) and 1 (fully match with the benchmark). The following groups of KPI types were identified: binary, percentage, threshold down, threshold up, and quartile down. Each KPI type had a particular area value computation formula:

- *Binary formula*: Formula applied to binary KPIs. This formula was used for the KPIs D1, D2, D4, ECR1, ECR2, and W3.
- *Percentage formula:* Formula utilized for KPIs that assign scores through a percentage. This formula was applied for the KPIs D3, ECR3, and W2.
- *Threshold down formula*: Formula used when the KPI value should not exceed a maximum threshold (the defined benchmark). This formula was used for the KPIs ECR4 and ECR5.
- *Threshold up formula*: Formula implemented when the KPI value should not lower than a minimum threshold (the defined benchmark). This formula was applied for the KPIs M1, M2, M3, and M4.
- *Quartile down formula*: Formula computed when no threshold (benchmark) was defined, and the KPI value should preferably be in the lower bracket. This formula was utilized for the KPIs ECR6 and W1.

Finally, the computed area values S_{kc} were implemented in the index formula. An overview of the formulas could be found in Appendix A.

3.2 Spatial Analysis
In this section, the methods are discussed that are used to evaluate the preparedness of municipalities, as indicated by the index outcomes. These methods support answering sub-question two and were devoted to summarizing and recognize patterns of spatial clustering by two methods: (i) global spatial autocorrelation (GSA); and (ii) local spatial autocorrelation (LSA).

3.2.1 Spatial Weights Matrix
These methods of spatial autocorrelation rely on the definition of spatial systems. These systems are summarize by spatial weights matrices, which represent the spatial relationships between individual units of analysis. Depending on the type of data, there are several main types: (i) contiguity-based neighborhoods (Rooks Neighborhood or Queens Neighborhood), (ii) k-nearest neighborhoods (KNN), (iii) threshold distance-based neighborhoods, and (iv) interaction-based neighborhoods (inverse distance weighting or squared inverse distance weighting) (Chi & Zhu, 2008). The choice for a specific neighborhood matrix influences the outcomes of the model, also referred to as the Modifiable Areal Unit Problem (MAUP).

In this study, a distance based KNN spatial weight matrix was used to summarize the spatial system. This matrix identifies the k-nearest neighbors of each observation based on the Euclidean distance, and assigns a weighted value to each neighbor to determine the spatial relationships (Chi & Zhu, 2008). This type of matrix suited the context of the thesis, because it avoided the "island effects" and was appropriate for analyzing units of varying sizes. Moreover, this type of matrix addressed the issue of normal distance-based spatial weight matrices, which could overestimate the total neighbors in urban areas and underestimate them in rural areas (Chi & Zhu, 2008). In specific, a value of five distance-based k-nearest neighbors was selected.

3.2.2 Global Spatial Autocorrelation
After the spatial weights matrix was defined, the GSA could be quantified to assess spatial clustering in the index. GSA is statistical measure to globally analyze spatial clustering by evaluating the extent of spatial similarity or dissimilarity of geographically defined observations (Anselin, 2005; Chi & Zhu, 2008). A spatial lag – a spatially weighted average of the geographically defined observation - is necessary to calculate the GSA (Anselin, 1995). This spatial lag serves as a dependent variable, and enables the evaluation of its relationship with the independent variable.

In this thesis, the Global Moran's I statistic was utilized to evaluate the GSA of the index. This statistic provided the first spatial statistical insights into the index by describing the spatial characteristics of globally summarized patterns. The Global Moran's I statistic was represented by one value, ranging between -1 and 1. A value greater than zero represented positive spatial autocorrelation, indicating that similar observed values are spatially close-by. A value less than zero exemplified a negative spatial autocorrelation, suggesting that dissimilar observed values are spatially close-by. A value equal or close to zero indicates the absence of spatial autocorrelation, implying the spatially random distribution of observed values (Anselin, 2005; Chi & Zhu, 2008).

3.2.3 Local Spatial Autocorrelation
In addition to GSA, the patterns of spatial clustering were evaluated by LSA, which is statistical measure that reveals the degree of significant spatial clustering of (dis)similar values around individual geographically defined observations (Anselin, 1995). A technique to measure LSA is the Local Moran's I statistic, which is a local disaggregation of the global coefficient of the Moran's I statistic. In contrast to the global coefficient, it has the ability to detect attribute similarity for each unit in comparison to its neighbors, and this enables the identification of spatial variations within a study area (Anselin, 2005; Chi & Zhu, 2008). Furthermore, the Local Moran's I statistic could recognize hot spots (high-

high), cold spots (low-low) and spatial outliers (high-low or low-high). Hot spots indicate concentrations of areas with a relatively observed value compared to the surrounding areas. Cold spots indicate areas where there were concentrations of observed values is relatively low compared to the surrounding areas. In addition, spatial outliers exemplify regions that performed significant better or worse than their neighboring areas (Anselin, 1995). Finally, the Moran Scatterplot was used to display the relationship between municipality's index values and its neighbors.

3.3 Data Modelling

Data modelling was the last stage of the case study and aimed to identify the main socio-demographic and economic characteristics that influenced municipalities' preparedness towards urban circularity and green transition, in line with sub-question three. The analysis involved the construction of both non-spatial and spatial models to explore and describe spatial patterns and relationships between the index outcome and the socio-demographic and economic characteristics. In addition, the spatial model could be used to explain and predict these patterns and relations. To realize this, a three-phase model construction process was employed.

In the first phase, a first version of the non-spatial model was developed to lay the foundation for the second model. In the second phase, the modelling process of the first model was modified to generate an improve second model. In the third phase, a final spatial model was constructed, which was built upon the first and second model. This model was a GWR and incorporated the spatial component. All models in this thesis try to explain the relationship between the CCI index scores (dependent variable) and the socio-demographic and economic characteristics (independent variables), as visualized in Figure 4.

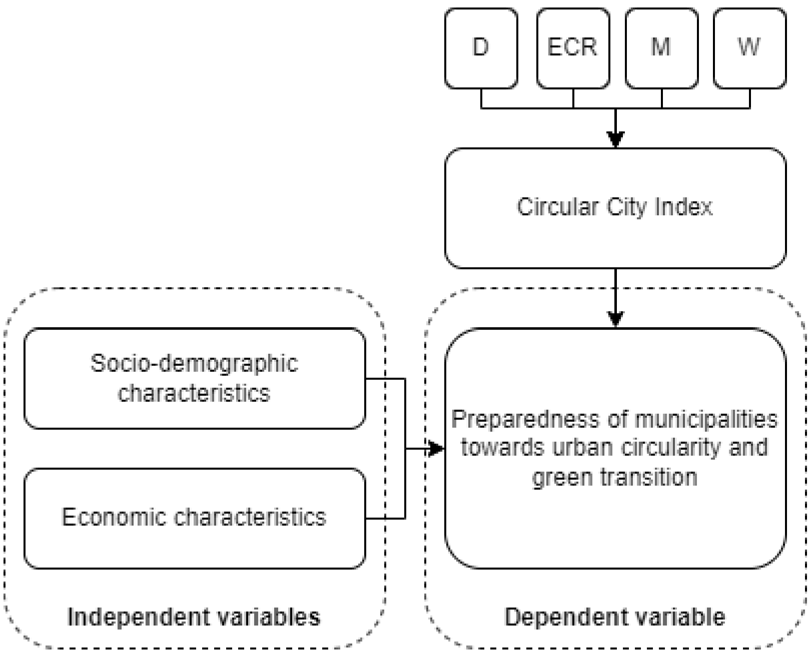


Figure 4 - Dependent and Independent Variables in Data Models.

3.3.1 Multiple Linear Regression Model

In the first and second model, the multiple linear regression was utilized to identify relationships between key socio-demographic and economic characteristics and the index performance of municipalities, without taking into account the spatiality of variables. This type of multiple linear regression model is a global regression technique, and utilizes an Ordinary Least Squares (OLS) method to statistically determine the relation between a dependent variable and multiple continuous independent variables over a whole research area (Dobson, 2002; Kutner et al., 2005).

The output of multiple linear regression models provides information on two main statistics: (i) coefficients, and (ii) p-value. The coefficient (R) represents the modelled slope between the dependent and independent variables in an equation and describes the relation between these variables. In addition, the p-value (or null-hypothesis significance test) assess the likelihood of finding a test statistic at least as extreme as the observed statistic (Randolph & Myers, 2013). In this thesis, if the p-value was below the threshold value of 0.05, then the null hypothesis was false, and the observed correlation was significant. In the case of a significant test value, other non-spatial statistical analyses could be performed (Anselin, 2005; Beale et al., 2010; Randolph & Myers, 2013).

3.3.1.1 First Model

The first model laid the foundation for the data modelling. As a first step, the data had to be preprocessed. The data was imputed using k-nearest neighbors, and normalized using the standard score (Z-score) normalization method. This method scaled the numerical data by subtracting the data mean from the data score, and divided the data score by the standard deviation (Kreyszig, 2011). The values of the normalized data corresponds to the number of standard deviations it differs from the mean (which is represented by zero) (Kreyszig, 2011). The aim of normalizing the socio-demographic and economic dataset was to reduce the impact of different units of measurement and enable the implementation of the data in spatial analysis and models.

3.3.1.2 Second Model

After the first model was validated, adjustment was necessary to improve the initial model results. These required modifications were implemented in the modelling process that generated a second multiple linear regression model.

First, the normal distribution of model data was tested using the D'Agostino's K^2 test, which is a suitable normality test for very large data observations (D'Agostino & Pearson, 1973). In contrast, other normality tests, such as the Shapiro-Wilk test, are not reliable for large datasets (Royston, 1995; Shapiro & Wilk, 1965). Assessing the normality of data was a necessary step for the later data analysis and modelling.

Second, the missing values in the datasets were imputed by using the k-nearest neighbors' method. Afterwards, the data was normalized by the Yeo-Johnson power transformation function, which enables the normalization of different types of input distributions, and allows the input of both positive and negative values (Raju et al., 2020).

Third, a feature selection was performed to simplify the model and avoid multicollinearity, and this was done by grouping features based on their similar sub-characteristics. The best feature of each class was selected using the F-statistic and p-values of a univariate linear regression. This selection reduced the number of features to a total of 11. In addition, the root mean squared error (RMSE) was calculated per feature to determine the best combination of selected features. By recursively removing features with the lowest RMSE, the best combination of features was determined to improve the model. Hereby was a minimum of 8 features specified.

3.3.2 Geographically Weighted Regression (GWR) Model

The first and second model formed the foundation of the final model: the Geographically Weighted Regression (GWR). In contrast to the standard OLS regression models, the GWR is a spatially extension of locally weighted regressions and accounts for spatial heterogeneity and spatial weighting (Beale et al., 2010; Chi & Zhu, 2008; Helbich & Griffith, 2016). This enhances the predictability power of the model (Bidanset & Lombard, 2014). As with the OLS-based multiple linear regression, the GWR its output provides information of two main measures: (i) coefficients (per municipality), and (ii) p-value.

In alignment with the second model, the selected features were implemented and preprocessed using the k-nearest neighbors imputation method and the Yeo-Johnson power transformation function.

In addition, the Variance Inflation Factor (VIF) was used to test multicollinearity and avoid highly uncertain or numerical instable outcomes in models due to correlation among the estimated regression coefficients (Wheeler & Tiefelsdorf, 2005). A VIF higher than 5 was indicative for multicollinearity between features, indicating problems in the model (Tay, 2017).

Subsequently, the spatial weightings of the model were specified, which is important in a GWR model, as these influence the accuracy and uniformity of model outcomes (Bidanset & Lombard, 2014). In this thesis, the bi-square kernel function was specified to allocate weights. This functions allocated weights based on Tobler's first law of geography and assigned zero-values to observations outside the given bandwidth (Bidanset & Lombard, 2014; Tobler, 1970). Additionally, the bandwidth was defined using adaptive kernels. This method defines bandwidths by the optimal proportion of neighborhoods, based the lowest root mean square prediction error in a model calibration (Bidanset & Lombard, 2014). Since the municipalities were unevenly scattering over space, and varying in size, the use of adaptive kernels was an optimal solution. To apply this method in the case study, the bandwidth selector (Sel_BW) from the Python library MGWR was used (PySAL, 2018).

Finally, the local multicollinearity was evaluated. In a GWR, the presence of multicollinearities among local coefficients is of greater significance compared to a global regression, because the complex estimation of coefficients in a GWR can be significantly affected by the correlations between local features at each single GWR (Wheeler & Tiefelsdorf, 2005). Therefore, the Variance Inflation Factor (VIF) was quantified and mapped to measure the multicollinearity for each municipality. As discussed earlier on, a VIF higher than 5 indicates difficulties with multicollinearity. The assessment of the local multicollinearities could indicate potential model validity problems at municipal level.

3.3.3 Model Validation

For each of the three models, a validation was performed. This step included the following evaluations: (i) model fit; (ii) model residuals; and (iii) sensitivity analysis. First, the model fit was assessed by the R-squared, mean absolute error (MAE), and root-mean-squared error (RMSE) measures. The R-squared described the proportion of variance of the dependent variable explained by the independent variable, and examined how well the model fits the data, ranging between 0 and 1. A value closer to 1 represented a higher variation that could be explained, and thus a better model fit (Anselin, 2005; Randolph & Myers, 2013). In particular for the GWR, the local R-squared was measured to indicate the model's goodness-of-fit was per municipality. This measure was used to identify municipalities where the model had no appropriate goodness-of-fit. Additionally, for all models the MAE assessed the mean variation between the absolute predicted values and the absolute observed values. In contrast, the RMSE calculated the mean variation between the predicted values and the observed values through the root-mean-squared error between those values (James et al., 2013).

Second, the model residuals were examined. The residuals provided insights into the differences between observed and predicted values in the model. This model residuals were analyzed through the relative errors of the observed values of each municipality. Ideally, the residuals do not show patterns of spatial clustering, and thus display homoscedasticity (Chi & Zhu, 2008).

Third, a sensitivity analysis was performed on each model. This analysis determines the effect of input features on the model output, and measures how sensitive the features are to changes in the model's parameters. For this all models in this study, the sensitivity analysis was performed through a perturbative study of five k-fold cross-validation. By repeating this cross-validation five times, the root mean squared error (RMSE) could be calculated to determine the coefficient robustness.

3.3.4 Model Comparison

The model comparison was performed to validate the quality of the final model. For this, the R-squared and the Akaike Information Criterion (AIC) were used. The AIC combines the model fit and model complexity to compare the goodness-of-fit of between different modes. A smaller AIC-value related to a better model fit (Anselin, 2005; Beale et al., 2010; Chi & Zhu, 2008).

4. Data

This chapter discusses the data of this thesis, building upon the previous methodology chapter. As a sizable portion of the thesis was dedicated to data processes, an additional chapter was deemed necessary to provide a clear understanding of the steps taken.

The section is structured as follows. First, this chapter examines open-source data in the Spanish context, which aligns with sub-question one. Subsequently, the criteria for data quality are discussed, followed by an explanation of the data collection process per data group, aligning with sub-questions two and three. The subsequent sections address the steps taken for data cleaning, homogenization, computation, merge, and geocoding. Finally, the data engineering pipeline is discussed, emphasizing this as an essential component of this thesis in line with the principles of open data.

4.1 Open-Source Data in Spain

In this section, the open-source data context of Spain is described in more detail. Spain - the case study area of this thesis is governed through three administrative divisions: autonomous communities (17) and autonomous cities (2), provinces (52), and municipalities (8,131) (INE, 2022). In certain autonomous communities, a supplementary administrative division - called "*núcleo de población*" or population nucleus - exists within the municipalities. In addition to this, several minor overseas territories - known as "minor plazas de soberanía" – are part of Spanish territory, and classified as unincorporated areas under protection of the Ministry of Defense (del Valle Gálvez, 2011).

A similar administrative organization could be recognized in the open governmental data structure. On national level, open data is available and accessible through data portals of Datos[3], INE[4], CNIG[5], IGN[6], and several open data portal of ministries. On autonomous community level, each autonomous community publish their data in an separate open data portal, which is often not centralized, and forms an obstacle for data collection on autonomous community level data. On province level, data availability and accessibility differ per region. While several provinces provide open data at an individual portal, other provinces incorporate this data into the portal of the autonomous community. On municipal level, the majority of the municipalities does not provide open data through an individual owned data portal. A few exceptions to this municipalities of main urban areas, such as Madrid[7] and Barcelona[8].

This decentralized way of publishing open data forms a barrier to data collection in Spain. The lack of municipal data portals to share open data forms an additional obstacle to the collection of municipal level data. As a consequence, this case study relied on the data portals of national governments and autonomous communities. On these data portals, the published municipal data includes data from the above mentioned population nucleus and minor overseas territories. As a result, it was considered necessary to integrate these local entities into the index, resulting in a total of 8,217 case study areas.

---

[3] www.datos.gob.es
[4] www.ine.es
[5] www.centrodedescargas.cnig.es
[6] www.ign.es
[7] www.datos.madrid.es
[8] www.opendata-ajuntament.barcelona.cat

4.2 Data Quality

Prior to the data collection, the data quality criteria were established to ensure the reliability and comparability of the data used in this thesis. The following criteria were identified, ranked by their level of importance:

(i) Datasets had to be collected from open data sources, in line with the open data principle.
(ii) Datasets had to be at municipal level.
(iii) Datasets had to be collected from official governmental sources.
(iv) Datasets had to be comparable to those used in the original CCI.
(v) Datasets had to be up to date to reflect the current state of municipalities.

Criteria 1 emphasizes the high priority and non-negotiability of open sources, as it aligned with the core values of transparency and reproducibility of the study. This criterion must always be fulfilled. Criteria two was important to avoid 'ecological fallacy' or 'fallacy of division'. This criterion is negotiable if the indicator is part of cross-municipal processes. Criteria three was important to safeguard the data reliability, as official governmental sources were considered to be highly trustworthy. This criterion could be ignored if data was only available at external data sources. In addition, the level of data availability or accessibility of official sources was considered to be a reflection of the priorities of (local) governments, and thus the importance of these topics towards politicians. Criteria 4 was important to ensure the development of an index comparable to the original CCI. Not fulfilling this criteria might indicate the need for modifications to the index. Criteria 5 was important to safeguard a reflection of the current state of municipalities. Data from 2021 or later was considered to be up to date.

4.3 Data Collection

This thesis classified the data into three main categories: g*eneral municipal data*, KPI data, and socio-demographic and economic data. General municipal data included information such as the official names (in Castilian), area code (CMUN), geometry, and population number of each municipality. KPI data was used to compute indicators that were necessary to implement the index, in line with sub-question two. Socio-demographic and economic data was used to gain insights into the relation between key socio-demographic and economic characteristics and the CCI results, in line with sub-question three.

4.3.1 General Municipal Data

The first data group - general municipal data - was collected for the implementation of both the KPI and socio-demographic and economic data. The data was derived from two datasets: the municipal geospatial dataset, and the municipal population register dataset. The first dataset was collected from IGN (2022) and contained information of four regions: (i) Iberian Peninsula and Balearic Islands; (ii) Canary Islands; (iii) Ceuta; and (iv) Melilla. These datasets contained the official name, area code, and geometry of each municipality, which was essential for the geocoding of other datasets. In addition, the municipal population register dataset was acquired from INE (2021), and facilitated the computation of KPI values relative to their number of inhabitants. Both datasets meet the five data quality criteria previously established. The information is summarized in Table 6.

Table 6 – Collected General Municipal Data.

Dataset	Definition	Type	Target	Result	Data Source
Geospatial municipalities dataset	Geospatial dataset containing the geometry of the official administrative borders of municipalities inside the Spanish territory.	ESRI Shapefile	Official dataset of 2021	Official dataset of 2022	(IGN, 2022)
Municipal population register dataset	Dataset containing information about the official population number per municipality inside the Spanish territory.	Non-spatial tabular data	Official dataset of 2021	Official dataset of 2021	(INE, 2021)

4.3.2 KPI Data

The second data group - KPI data - was acquired to compute KPIs, which were require for the index implementation. The original CCI methodology categorized seventeen quantifiable indicators into four levels: digitalization (D); energy, climate, and resources (ECR); mobility (M); and waste (W).

4.3.2.1 Digitalization (D)

The first level – digitalization (D) – contained four indicators in the original CCI (Muscillo et al., 2021):

- D1 – Presence in ANPR (public digital service platform).
- D2 – Adoption of SPID in PA digital properties.
- D3 – Percentage of people with broadband connection (>30Mb/s).
- D4 – Accessibility of local government digital properties.

For D1, D2, and D4, a similar dataset was collected via OBSAE (2023). The dataset provided information of services implemented by each municipality in the public digital service platform in 2022. The dataset was a Microsoft Excel-file with the data given by municipality name and type of public digital service implemented by the municipality. For D1, information was obtained regarding the presence of each municipality in the PAe, and this was the case if they had implemented at least one service. For D2, data was gathered about municipalities that incorporated the Cl@ve-service. Cl@ve is a system for the electronic access to digital administration and public services of citizens in Spain (Cl@ve, 2014). For D4, data was collected of each individual implemented services in local government digital properties. This provided information about the accessibility of local government digital properties for citizens.

In contrast, the D3 dataset was collected via MINECO. This dataset provides a percentage of the population that had a broadband internet connection of 30 MB/s or higher per municipality in June 2021. The dataset was a Microsoft Excel-file with the data given by municipality code (CMUN) and percentage of population with broadband connection.

4.3.2.2 Energy, Climate, and Resources (ECR)

The second level – energy, climate, and resources (ECR) – was measured by six indicators in the original CCI (Muscillo et al., 2021):

- ECR1 – Subscription of the Covenant of Mayors.
- ECR2 – Level of commitment in the Covenant of Mayors.
- ECR3 – Percentage local energetic self-sufficiency (from renewables).
- ECR4 – Annual average concentration of PM10.
- ECR5 – Annual average concentration of NOx.
- ECR6 – Percentage of water leaks.

The ECR1 and ECR2 data were manually collected via the website of the Covenant of Mayors (CoM)[9] in October 2022. This dataset contained information on signatures and level of commitment per municipality in the CoM. The level of commitment was classified into four groups: CoM by 2020, CoM by 2030, CoM by 2050, and adaption to CoM.

The ECR4 and ECR5 data was obtained from the European Environment Agency's data portal[10]. The raster datasets contained average interpolated values of PM10 (for ECR4) and NOx (for ECR5), calculated by using a regression-interpolation-merging mapping methodology (European Environment Agency, 2022), and had a spatial resolution of 1 kilometer grid, which is suitable for computing the average PM10 and NOx values per municipality.

The ECR3 and ECR6 data were not collected in this study. The ECR3 data required information on photovoltaic and wind generators, which was only available on province level, making the data implementation unreliable at municipal level. The ECR6 data required information on the water distribution system leakage, but this was not openly available in Spain.

<u>Mobility (M)</u>

The third level – mobility (M) – was evaluated by four indicators in the original CCI (Muscillo et al., 2021):

- M1 – Total surface of pedestrian areas (in square meters per 100 inhabitants).
- M2 – Total charging stations (per 1,000 inhabitants).
- M3 – Total length of cycleways (in kilometers per 100 square kilometers).
- M4 – Total bus stops (per 100 inhabitants).

The mobility level datasets were acquired from OpenStreetMap (2022) by using the OSMnx Python package (Boeing, 2022), and were categorized based on tags defined in Table 7, following the original CCI methodology. To ensure accurate calculation of length and surfaces, the spatial projections were transformed. The data was then aggregated to the geometry of each municipality.

Table 7 – Data Collection by OpenStreetMap (2022) Tags of KPI.

M1	**M2**	**M3**	**M4**
place=square	amenity=charging_station	cycleway=*	highway=bus_stop
highway=path		highway=cycleway	
highway=pedestrian			
leisure=park			
highway=footway			
foot=designated			

<u>Waste (W)</u>

The fourth level – waste (W) – was assessed by three indicators in the original CCI (Muscillo et al., 2021):

- W1 – Production of solid waste (in tons per inhabitant).
- W2 – Percentage of solid waste recycling.
- W3 – Collection of e-waste.

[9] https://eu-mayors.ec.europa.eu/en/home
[10] https://www.eea.europa.eu/data-and-maps/

The data of W1, which required information of the total produced solid waste, was not available in all autonomous communities in Spain, making it impossible to obtain a complete and reliable dataset. As a consequence, the data is not incorporated in the index. The W2 data had to be modified, as it initially incorporated the W1 data for computing the percentage of solid waste recycling. As alternative, the total recycled glass (packaging) waste was used, which was manually collected from the website of EcoVidrio[11]. The dataset provided information of total recycled glass packages (in kilograms) per municipality in Spain. The W3 dataset was manually collected from the website of Punto Limpio[12], a project of the non-profit organization ECOTIC.

Table 8 - Collected KPI Datasets in Thesis.

Level	KPI	Data Type	Data Source
Digitalization	D1	Binary	OBSAE
	D2	Binary	OBSAE
	D3	Percentage	MINECO
	D4	Levels	OBSAE
Energy, climate, and resources	ECR 1	Binary	Covenant of Mayors
	ECR 2	Categorical	Covenant of Mayors
	ECR 4	Scalar (non-negative value)	European Environment Agency (EEA)
	ECR 5	Scalar (non-negative value)	European Environment Agency (EEA)
Mobility	M1	Scalar (non-negative value)	OpenStreetMap
	M2	Scalar (non-negative value)	OpenStreetMap
	M3	Scalar (non-negative value)	OpenStreetMap
	M4	Scalar (non-negative value)	OpenStreetMap
Waste	W2	Scalar (non-negative value)	EcoVidrio
	W3	Binary	Punto-Limpio / ECOTIC

4.3.3 Socio-Demographic and Economic Data

The third data group – socio-demographic and economic data - was collected to conduct both non-spatial and spatial models. For socio-demographic characteristics, data was collected in two subclasses: demographic, and socio-economic variables. Demographic variables refer to population characteristics, and information was collected for population, gender, age, and nationality. Socio-economic variables reflect the social and economic status of a population, and data was collected on the topics of housing, income, and employment status. In contrast, economic characteristics pertain to the economic situation of a population, and information was obtained on wealth distribution, public finance, entrepreneurship, agriculture, and tourism in local economies.

The data was collected via the data portal of INE[13], as shown in Table 9. Unlike data portals of other national institutions and autonomous communities, NIE did not publish data on population nucleus and minor overseas territories. Therefore, the data could only be acquired for 8,087 case study areas.

[11] https://www.ecovidrio.es/en/recycling/recycling-data
[12] https://punto-limpio.info/
[13] https://www.ine.es/

Table 9 - Overview of Socio-Demographic and Economic Datasets.

Type of Characteristic	Type of Sub-Characteristic	Variable	Year	Data Source
Socio-Demographic	Population	Population	2020	INE
		Population density (per km^2)	2020	INE
		Natural population growth (in %)	2020	INE
	Gender	Population female (in %)	2020	INE
	Age	Average age of population	2020	INE
		Population below 18 years (in %)	2020	INE
		Population above 65 years (in %)	2020	INE
	Nationality	Population non-Spanish citizen (in %)	2020	INE
	Housing	Average household size	2020	INE
		Single persons households (in %)	2020	INE
		Total residential buildings (per capita)	2011	INE
	Income	Income (per capita)	2020	INE
		Income (per household)	2020	INE
		Unemployment benefits (in % of average salary)	2020	INE
Economic	Wealth distribution	GINI index	2020	INE
	Public finance	Municipal debt (per capita)	2021	INE
	Economy	Total companies (per capita)	2020	INE
	Agriculture	Total agricultural livestock units (per km^2)	2020	INE
		Total agricultural cattle farms (per km^2).	2020	INE
	Tourism	Total tourist houses (per capita)	2022	INE

4.4 Data Cleaning and Homogenization

The initial step after the data collection included the cleaning and homogenization of raw data. The data cleaning was performed to exclude irrelevant information that not contained information on the municipal name or area code, the (raw) indicator data, and the municipal geometry.

The data homogenization was performed to ensure a coherence, efficiency, and interoperability of the data. Columns were renamed to either 'Municipality' (if referring to the municipality name), 'CMUN' (if referring to the 5-digit municipality area code), or 'CTOT' (if referring to the 7-digit area code). Next, the column of indicator data was labeled in line with the KPI. Finally, the municipality names were homogenized to the official government names to account for naming differences due to languages, punctuation marks, or administrative names. The FuzzyWuzzy Python library was used to match unofficial names with official names using the Levenshtein Distance and a minimum similarity threshold of 90% (PyPi, 2020). The scores lower than the threshold was assigned a missing value.

Additionally, the coordinate reference systems and map projections were transformed, which was essential for the integration of local spatial datasets with global spatial datasets, maps or remote sensing data (Knippers & Tempfli, 2013). These transformations were necessary since spatial datasets used distinct projection coordinates and geographic coordinates systems. By transforming them, the data could accurately represent the geographical region, and geometric measurements could be calculated with high accuracy. For mapping all Spanish territorial administrative regions together, the EPSG:4258 was used, as proposed by the Spanish government (IGN, z.d.). For each specific region, Table 10 provides an overview of the transformation procedures.

Table 10 - Coordinate Reference Systems and Map Projection per Geographical Region in Spain.

Geographical Region	Geodetic Coordinate Reference System	Map Projection	EPSG-code
Canary Islands	REGCAN95	Universal Transverse Mercator (UTM)	4083
Balearic Islands; Iberian Pensinula; Plazas de Soberanía (Ceuta, Melilla, and Minor Plazas de Soberanía)	ETRS89	Universal Transverse Mercator (UTM)	4258

4.5 Data Computation

Next, the data was computed. This was required when the desired dataset could not be directly gathered from the data source, needed to be scaled as a relative metric, or required spatial aggregation to the municipal level. In such cases, new datasets were created by computing one or more (alternative) datasets. An overview of the data computation could be found in Appendix C and Appendix D.

4.6 Data Merge & Geocoding

Following on this data computation, the data was merged and geocoded. The datasets were merged using the official municipal names, CMUNs, or CTOTs. The datasets were geocoded during this merge since the geospatial municipalities' dataset contained information on geometry. Geocoding was a vital procedure in this thesis, as it was necessary for the spatial analysis and spatial modelling. The geocoded datasets were stored in both ESRI Shapefile and OGC GeoPackage file formats, which enabled the data to be applicated in open source software.

4.7 Data Engineering Pipeline

The final step was to incorporate the data workflow into a data pipeline for each of the above-discussed steps. This data pipeline incorporated open data principles, and ensured the full reproducibility and scalability of the thesis. The pipeline consisted of four main steps: data collection, data preprocessing, index computation, data analytics and data modelling. For each step in the pipeline, a Jupyter Notebook was developed and integrated. An overview of the data pipeline could be found in Figure 5.

For data collection, a script was developed that acquired the required datasets by downloading files from specified websites. The script was designed to ensure that all collected data was publicly available and met the open data principles. For data preprocessing, two scripts were developed that preprocess each separate dataset as described in the methodology by first cleaning and homogenizing the datasets, afterwards merging and geocoding them into one dataset for each data group. The preprocessing scripts ensured that the data was consistent, properly formatted, and ready for index computation. For the index computation, a Jupyter Notebook was developed to calculate the index formula, incorporating the redefined benchmarks, area values computations, and weightings per (set of) indicator. The script was designed to be scalable and easily adaptable to different datasets and parameters.

For the data analytics and data modelling, separate scripts were created that were integrated into the pipeline. For data analytics, one script was created for initial descriptive statistics of the index dataset, while two scripts were made to analyze the spatial autocorrelation of the index data, incorporating the global and local spatial autocorrelation. For data modelling, two scripts were redeveloped, based on a data pipeline developed for the Urbana project by researchers of BSC-CNS (Gregori & Reyes Valenzuela, 2021). This data pipeline incorporated scripts for both the multiple linear regression model as for the geographically weighted regression model. However, the scripts had to be redefined to suit the current thesis research. Both scripts incorporated the procedures for training the model by analyzing and modifying the model's data normality, fit, residuals, sensitivity analysis, and multicollinearity. In addition, these scripts allowed for data visualization and modeling, which enabled a more comprehensive understanding of the preparedness of Spanish municipalities towards urban circularity and green transition.

All collected data is public, and the methods and software used were open source, making the pipeline a fully reproducible, scalable method. In line with the open data principles, the data script is openly available on Zenodo[14], a sharing platform maintained by the European Union. This approach enhanced the transparency and reproducibility of the research process and findings, as these are available to other researchers and policymakers. The pipeline is easy to modify because the input data and model parameters are clearly stated at the beginning of each script, and the same structure is applied throughout the pipeline. The flexibility of the data pipeline allows for easy adaptation to various projects and contexts, and can be run directly for specific autonomous communities by adjusting the parameters to the specific community, providing a useful and easy to use tool for policymakers.

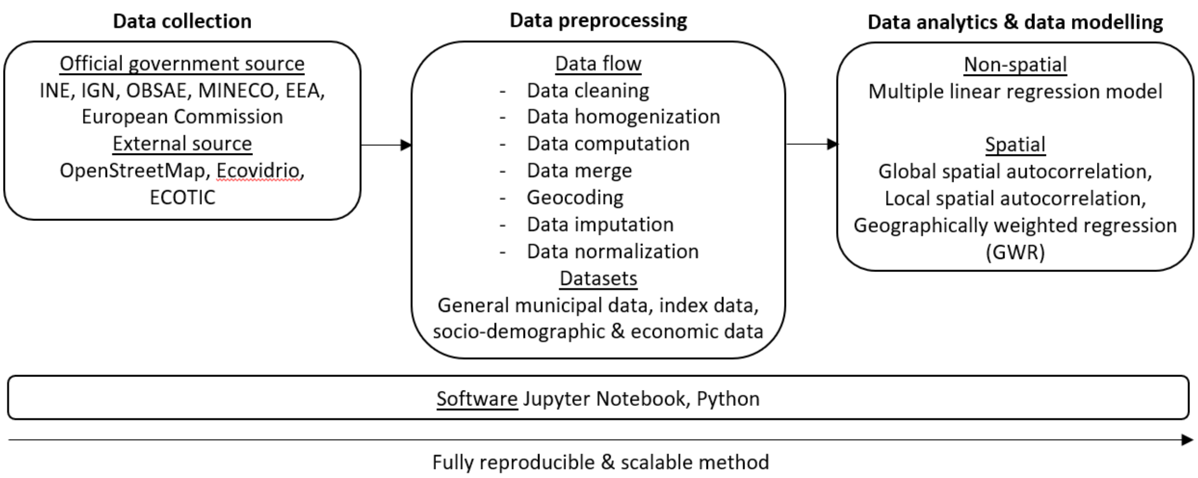


Figure 5 - Overview of Data Pipeline of Thesis.

[14] https://doi.org/10.5281/zenodo.7682093

5. Results

In the introductory chapter the general research objective of this thesis was presented. In the second chapter, a comprehensive literature review was provided. The third chapter operationalized the methodological framework of this study, and the fourth chapter discussed the data. Finally, in this fifth chapter, the results of the thesis were presented.

This chapter is structured corresponding to the three sub-questions. In the first section, the suitability assessment of the CCI for the Spanish context is presented. In the second section, the preparedness of Spanish municipalities towards urban circularity and green transition was quantified by the CCI outcomes. In the third section, the results of the spatial and non-spatial models are presented, in order to analyze and predict the relation between significant socio-demographic and economic factors and the level of preparedness of Spanish municipalities.

5.1 Reformulation and Redefinition of the Circular City Index

In this first section, the results are presented to answer the first sub-question: *"What extent of reformulation and redefinition is required for the original Circular City Index to be effectively applied to the Spanish context?".* A detailed explanation is provided of the modifications made to the CCI for the Spanish context. These modifications included adjustments to the KPIs, indicator benchmarks, area values computations, and weighting system.

5.1.1 Index Formula

With regards to the index formula, the Spanish CCI implemented the original index formula to evaluate the preparedness of municipalities, as developed by Muscillo et al. (2021, p. 9):

$$CCI_c = \Sigma_{A\epsilon\ Areas}(W_A \Sigma_{k\ \epsilon\ KPI(A)} W_k \times S_{kc})$$

No reformulations or redefinitions were implemented to this formula because the formula is not bounded to the Italian context. In addition, the implementation of a similar formula preserved the characteristics of the original index, and enabled a comparison of the original index with the Spanish index.

5.1.2 Key Performance Indicators (KPIs)

For the first level - digitalization – two indicators had to be reformulated to align with the related concepts in Spain. These consisted of the indicators D1 and D2, and they were reformulated as follows:

- *Original D1: “presence in ANPR (public digital service platform)”*
- *Reformulated D1: “presence in PAe (public digital service platform)”*

- *Original D2: “adoption of SPID in PA digital properties”*
- *Reformulated D2: “adoption of Cl@ve in PA digital properties”*

The second level – energy, climate, and resources – contained two indicators that were not suitable for the Spanish context. The ECR3 indicator could not be implemented in the Spanish CCI, since data was not available at municipal level. The ECR6 indicator could not be incorporated, because data of the water distribution system was not openly available or accessible in Spain.

The third level – mobility – contained of four indicators that did not have to be adapted.

The fourth level – waste – contained one excluded indicator. For the W1 indicator, the data of the produced solid waste was not available at municipal level in all autonomous communities. The incompleteness of the nationwide data coverage at this level of aggregation caused problems with data collection. In addition, due to the unavailability of the aforementioned produced solid waste data, the W2 indicator had to be reformulated to incorporate absolute numbers instead of percentage numbers. The indicator W2 was reformulated as follows:

- *Original W2: "percentage of solid waste recycling"*
- *Reformulated W2: "total recycled waste (in kilograms per inhabitant)"*.

5.1.3 Indicator Benchmark

The Spanish version of the CCI implemented a similar approach to the original index for determining the indicator benchmarks. Originally, the benchmarks were partly defined for the Italian context. Consequently, the benchmarks of five indicators had to be reconsidered.

First, the benchmarks for ECR4, ECR5, M1, and M2 were based on Italian policies and legislations. These policies and legislations were not in implemented in Spain, making the benchmarks unsuitable in this context. Second, the unavailability of the W1 dataset in Spain meant that W2 could not be computed as in the original CCI, resulting in the benchmark set to zero. While the other eight benchmarks were suitable for the Spanish context, the redefinition of new benchmarks was beyond the research scope. Therefore, all original benchmarks were preserved in the index, including the benchmarks of ECR4, ECR5, M1, and M2, with only the benchmark W2 being nullified. An overview could be found Appendix F.

5.1.4 Area Values Computation

With regards to the area values (S_{kc}) computation, no major redefinitions were implemented. The five equations to compute the area values – as defined by Muscillo et al. (2021) - were implemented to the Spanish index. Nonetheless, a minor change was applied to computation of area values for the W2 indicator because it was reformulated as an absolute measure and no longer contained relative data. Therefore, the new defined *quartile up*-equation was used, instead of the initial *percentage*-equation. The *quartile up*-equation is an inverted version of the *quartile down*-equation, as implemented in the original CCI. The *quartile up*-equation is used for indicators that are missing a defined threshold (benchmark), and that should preferably be in the higher bracket. The computed area values S_{kc} were then integrated in the index formula. An overview of the formulas is shown in Appendix B.

5.1.5 Index Weighting

As the KPIs ECR3, ECR6, and W1 were excluded from the index, the single KPI weights ($W_k\epsilon$) in the energy, climate, and resources set and waste set had to be reweighted. The following formula had been used for the calculated weighting $W_k\epsilon$ for single KPI k:

$$W_k\epsilon = \frac{original\ W_k\epsilon}{\Sigma\ (original\ W_k\epsilon\ in\ KPI(A))}$$

In the abovementioned formula, $W_k\epsilon$ represents the new calculated weight for single KPI k. Further, $original\ W_k\epsilon$ embodies the original weight for single KPI k in the original CCI, and $\Sigma\ (original\ W_k\epsilon\ in\ KPI(A))$ symbolizes the sum of these original weights per set of KPIs. The new formula for $W_k\epsilon$ was used in the index formula. With regards to the weighting of KPI sets ($W_A\epsilon$), no redefinitions were needed as none of the original sets was dropped from the index.

Table 11 – Conceptualization of Weights per (Sets of) KPIs.

Level	Level weight	KPI	KPI weight
Digitalization	0.2	D1	0.3
		D2	0.3
		D3	0.3
		D4	0.1
Energy, Climate and Resources	0.3	ECR 1	0.33
		ECR 2	0.33
		ECR 4	0.16
		ECR 5	0.16
Mobility	0.2	M1	0.2
		M2	0.3
		M3	0.2
		M4	0.3
Waste	0.3	W2	0.66
		W3	0.33

5.2 Preparedness of Spanish Municipalities Towards Urban Circularity and Green Transition

In this second section, the findings are presented to support the second sub-question: *"How does the Circular City Index evaluate the preparedness of Spanish municipalities towards urban circularity and green transition?".* To evaluate the level of preparedness of Spanish municipalities towards urban circularity and green transition, the index was implemented and quantified. In this section, the findings of the index quantification are outlined.

First, the initial results are presented by descriptive statistics. Second, the geographic distribution of the index outcomes is presented through maps. Third, the index outcomes are tested on spatial autocorrelation to enable a more in-depth understanding of the geographic distribution.

5.2.1 Descriptive Statistics

The reformulated and redefined index – as described in the section 3.1 – was implemented and assessed for all 8,217 case study areas in Spain. In Table 12, the descriptive statistics of the input data are illustrated for the non-binary and non-leveled KPIs. The binary and categorical KPIs have not been considered, as descriptive statistics do not provide appropriate insights for these types of indicators.

Table 12 - Descriptive Statistics of Non-Binary and Non-Leveled KPIs used in the Spanish CCI.

	D3	ECR4	ECR5	M1	M2	M3	M4	W2
count	8131	8217	8217	5757	667	1699	8131	8131
mean	0,84	12.07	6.88	11068.94	0.97	29.72	0.16	507.89
std	0,30	4.13	5.07	647340.69	2.85	52.41	0.64	4504.12
min	0.00	0.00	0.00	3.36	0.00	0.00	0.00	0.00
25%	0,89	9.50	4.00	277.11	0.07	3.23	0.00	2.58
50%	0,98	11.0	5.95	609.08	0.20	11.29	0.00	16.37
75%	1.00	13.5	8.00	1329.19	0.62	31.72	0.12	106.71
max	1.00	51.00	86.00	48959398.18	37.04	615.83	40.00	289605.71

Table 13 presents this index output data. The most noteworthy observations relate to the mean, standard deviation, minimum, and maximum scores. On average, Spanish municipalities achieved a mean score of 0.454 and the standard deviation of 0.145. The maximum score of 0.857 was accomplished by the municipality of León, located in the autonomous community of Castile and Leon. In contrast, some municipalities achieved the lowest score of 0.100. With regards to the four index levels, the results show that municipalities scored on average 0.806 on digitalization, and 0.145 on mobility. In addition, the standard deviation was highest for waste (0.311), and lowest for mobility (0.135). Furthermore, none of the 8,217 municipalities reached a perfect score of 1.0 on digitalization or mobility, or an imperfect score of 0.0 on energy, climate, and resources.

Table 13 - Descriptive Statistics of the CCI Score.

	CCI	Digitalization (D)	Energy, Climate, and Resources (ECR)	Mobility (M)	Waste (W)
count	8217	8217	8217	8217	8217
mean	0.454	0.806	0.496	0.145	0.385
std	0.145	0.190	0.232	0.135	0.311
min	0.100	0.000	0.167	0.000	0.000
25%	0.339	0.638	0.333	0.011	0.167
50%	0.445	0.929	0.333	0.129	0.500
75%	0.551	0.954	0.767	0.207	0.667
max	0.857	0.985	1.000	0.868	1.000

Figure 6 visualize the descriptive statistics of Table 13 in more detail. Figure 6 provides a five-number summary of the CCI scores by a boxplot. For the overall CCI, the scores were almost evenly distributed between the minimum (0) and maximum (1), while the median score was approximately 0.5. For digitalization, more than 75% of the municipalities achieved a score higher than 0.6. The opposite outcome was noticeable for mobility, where about 75% of the municipalities scored less than 0.2. For both levels were outliers detectable. For both energy, climate, and resources, and waste, a median of 0.5 could be recognized. Nevertheless, these indicators levels had a different distribution of observations. Figure 7 presents the distribution of the overall and separate levels CCI scores. Energy, climate, and resources had the majority of the observations between 0.2 and 0.4, while waste was more evenly distributed. In addition, bimodal distributions can be recognized in all distributions.

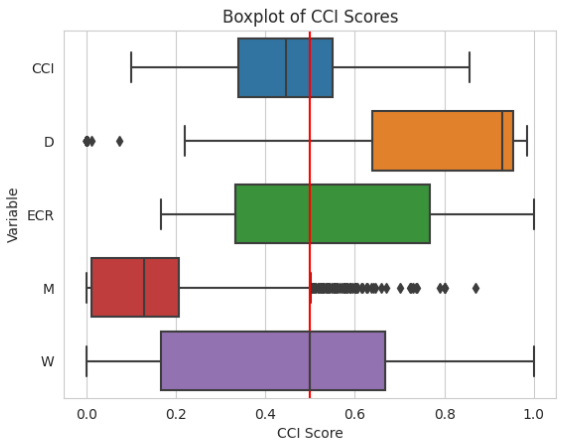


Figure 6 - Boxplot of CCI Scores in Spain.

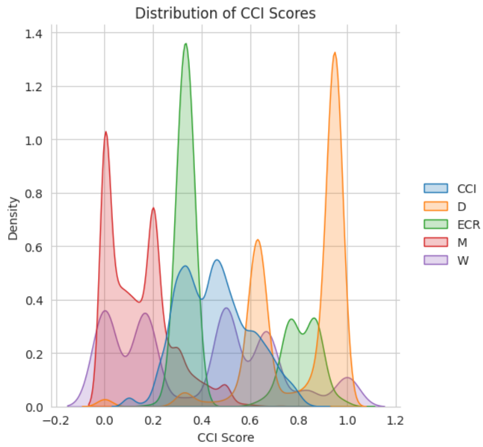


Figure 7 - Distribution of CCI Scores in Spain.

Finally, the results were analyzed by the population size of municipalities. The CCI scores were categorized by the municipality size classification of Muscillo et al. (2021, p. 10). The distribution of population numbers per municipality is shown in Table 14. The table reveals that 75% of the population lived in medium-sized (15 – 100 thousand inhabitants) or large (more than 100 thousand inhabitants) municipalities. In contrast, out of the 8,217 municipalities included in this thesis, 6,818 had a population of less than 5 thousand. Together, these municipalities were home to 12% of the Spanish population (INE, 2021).

Table 14 - Distribution of Population per Municipality Size in Spain (2021). Source: (INE, 2021)

Population Size of Municipality (in thousands)	Number of Municipalities	Population (in millions)	Population (in %)
0 – 5	6818	5.69	12.00
5 – 15	766	6.48	13.68
15 – 100	484	16.33	34.46
> 100	63	18.89	39.86
Total	8217	47.39	100

Figure 8 provides more a context to these numbers, by normalizing the distribution of the CCI scores by municipality size. The figure reveals that municipalities with a population size greater than 15 thousand had a right-skewed distribution, and a larger percentage of municipalities achieved a score closer to 1. The municipalities with a population size between 5 and 15 thousand followed a more normal distribution, and the scores of municipalities were more or less evenly distributed between 0.4 and 0.8. The municipalities with a population less than 5 thousand inhabitants followed a left-skewed distribution, and a larger share of municipalities achieved a score closer to zero.

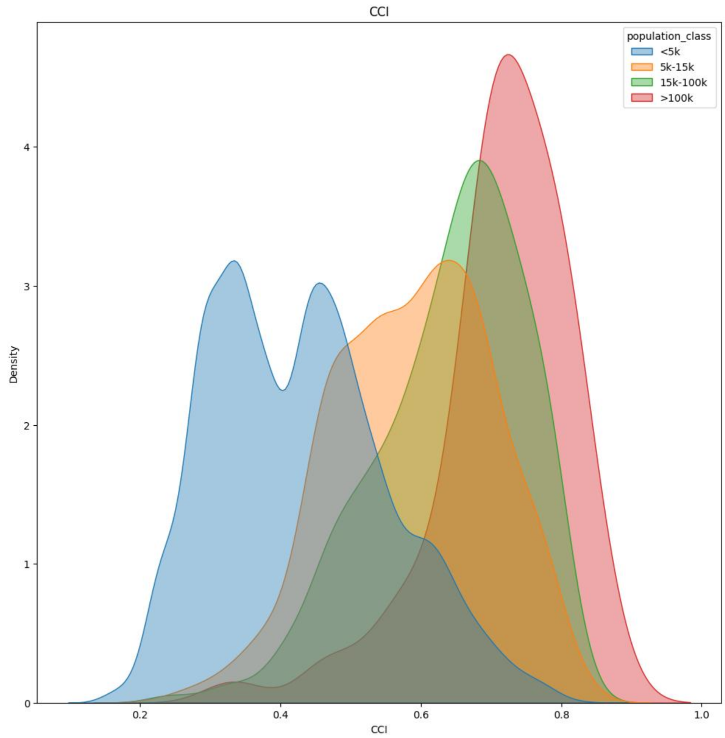


Figure 8 - Distribution of CCI Scores, Classified by Municipality Size.

5.2.2 Geographic Distribution of Index

To further understand the first results, it is necessary to study the geographical distribution of the CCI scores. Figure 9 maps this geographic distribution, classified by quantiles to ease the identification of spatial patterns. Upon first glance, the map presents different spatial patterns. In particular, the lowest 25% of CCI scores were more or less located in the central and northwestern regions of Spain, with the exception of the Madrid and Galicia regions. Contrariwise, the highest 25% of CCI scores concentrated in the Balearic Islands, Galicia, Madrid area, and eastern coastal regions.

Figure 10 provide a similar representation. The figure illustrates the distribution of the CCI scores, aggregated to province and autonomous community level. Corresponding to the figure, the bottommost 25% of scores at the province level were predominantly originated in central-northwest Spain, whereas the top 25% scores concentrated in the Balearic, Galicia, the regions near the Strait of Gibraltar, and the coastal areas of Catalonia and Valencian Community. Similarly, at the autonomous community level, the lowest 25% of scores were clustered in Asturias, Castile and Leon, and Castilla-La Mancha. Meanwhile, the higher 25% scores were located in the autonomous communities of the Balearic Islands, Galicia, Murcia, and Valencian Community.

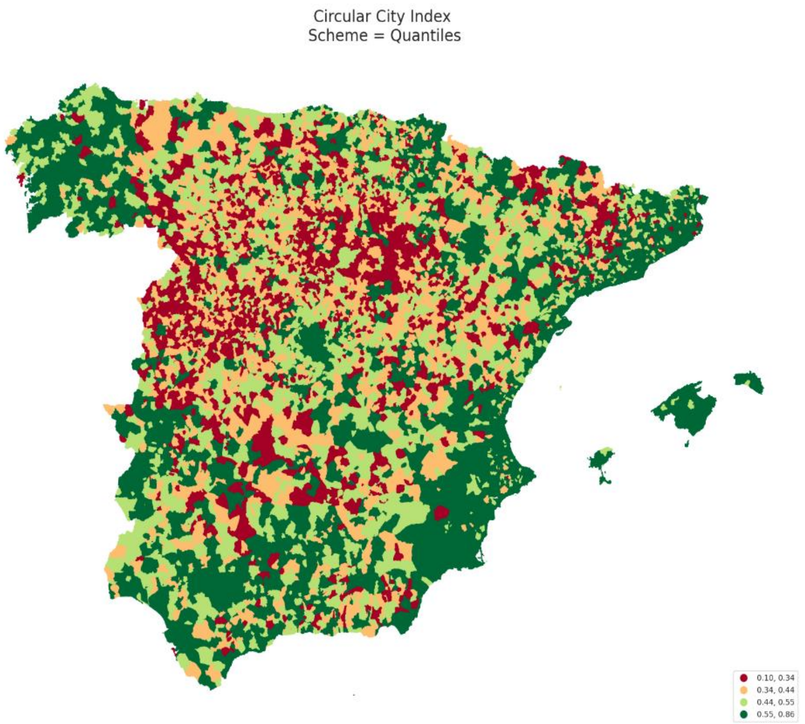


Figure 9 - Spatial Distribution of the CCI scores at municipal level in Spain.

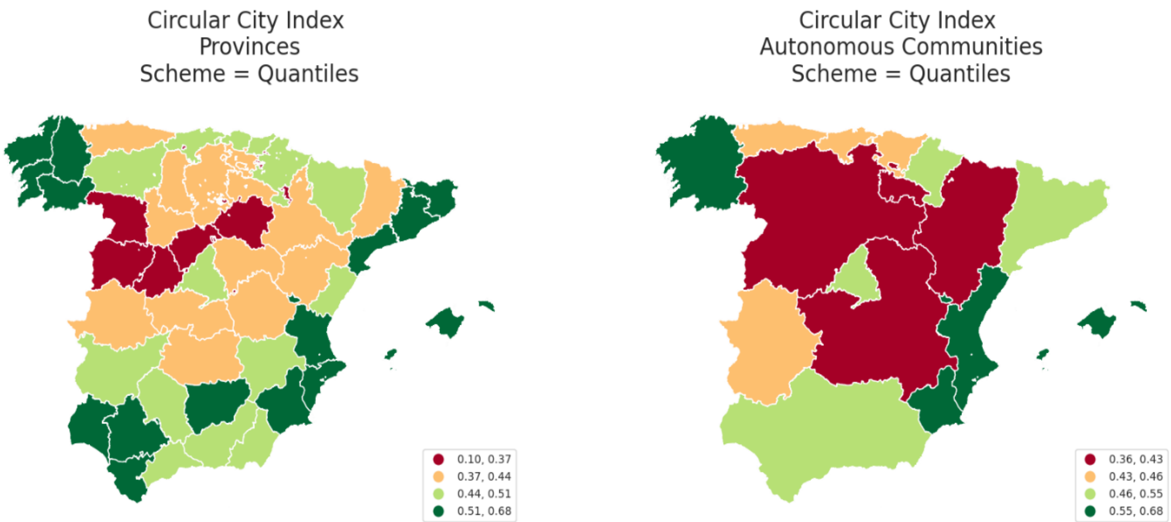


Figure 10 - Spatial Distribution of the CCI scores at province and autonomous community level in Spain.

Figure 11 reveals the outcomes per CCI level for each municipality. For digitalization, the scores related to the lowest quartile were found in the southern, northern, and northeastern parts of Spain. In specific, these scores were observed in almost the entire autonomous communities of Asturias, Basque Country, Catalonia, and Navarra. In contrast, the higher scores were more evenly distributed. For energy, climate, and resources, scores between 0.17 and 0.33 were noted in the central regions, while scores between 0.77 and 1.0 were observed around Galicia, Valencian Community, and the Strait of Gibraltar. For mobility, the highest 25% of the scores were located mainly in the central-northern parts of Spain, while the other 75% of the score were more or less randomly distributed across the country. For waste, scores below 0.17 were mainly located in central-northwestern and central-southern regions, while scores above 0.5 were spread across the country.

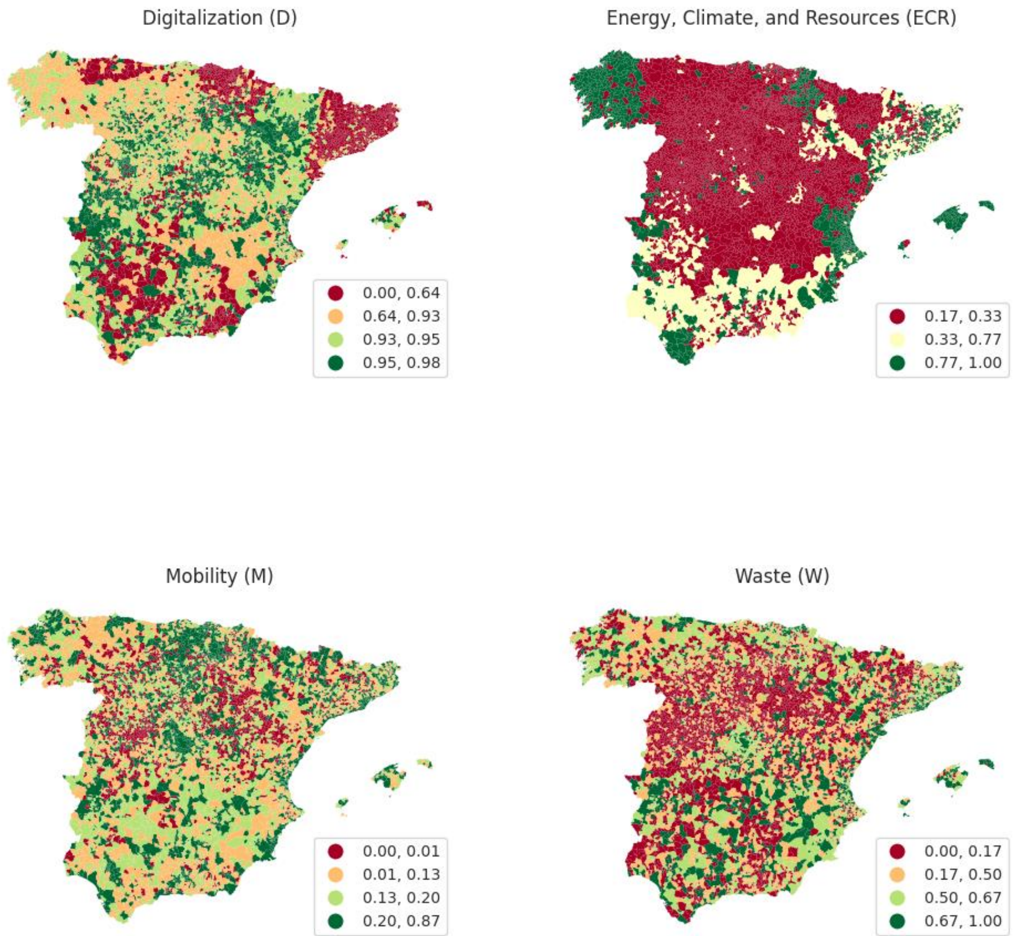


Figure 11 - Spatial Distribution of the CCI scores per index level at municipal level.

5.2.3 Spatial Clustering

The previous paragraphs discussed the initial results of the descriptive statistics, and delved into the identification of (spatial) patterns in the CCI scores. To obtain further insights into these patterns, the spatial clustering was investigated by global and local spatial autocorrelation.

Figure 12 visualizes the reference distribution and Moran Scatterplot. The calculated Global Moran's I value for the CCI index is 0.46, with a p-value of 0.001. Additionally, Figure 13 presents the outcomes of the Local Moran's I on a cluster map. Each color represents a particular type of spatial clustering. The red areas denote hotspots, representing spatial clusters that have a relatively high degree of urban circularity readiness. The blue areas indicate cold spots, which are clustered areas with a relatively low urban circularity readiness compared to surrounding areas. The orange and light-blue areas correspond to regions with high-low or low-high spatial outliers. The gray areas suggest that no significant spatial autocorrelation was detected, as the Moran's I p-value was higher than 0.05. Figure 13 illustrates that hotspots were primarily located in the Balearic Islands, Galicia, Madrid area, in the eastern coastal regions of Catalonia, Murcia, and Valencian Community, and around the Strait of Gibraltar. In contrast, the cold spots were located in the central-northwestern parts of Spain, mainly around the autonomous community of Castile and Leon. Moreover, some minor high-low spatial outliers areas could be recognized in the central-northern part of Spain, around La Rioja.

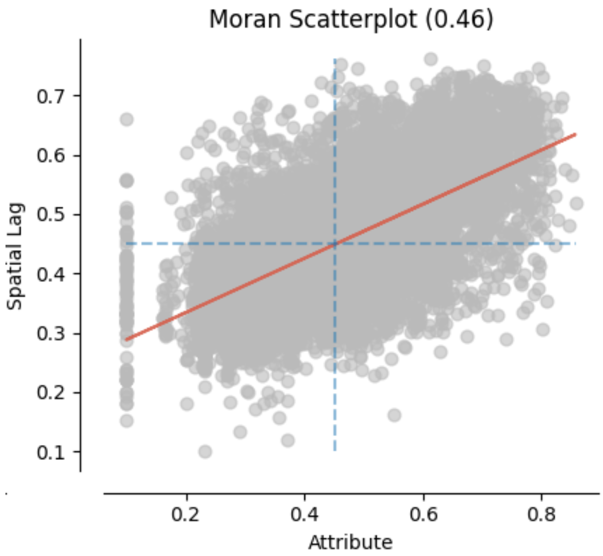


Figure 12 - Moran Scatterplot of the Global Moran's I of CCI Scores in Spain.

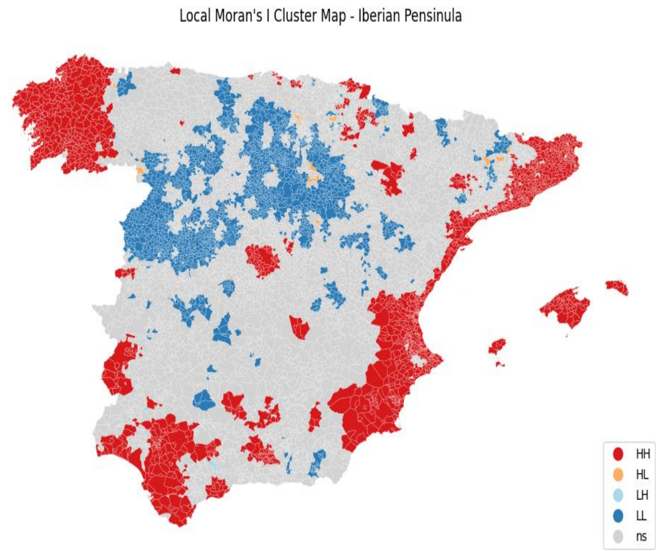


Figure 13 - Cluster Map of the Local Moran's I of CCI Scores in Spain.

5.3 Key Socio-Demographic and Economic Characteristics That Impact the Preparedness of Spanish Municipalities

In this third section, the outcomes are presented to address the third sub-question: *"What are the key socio-demographic and economic characteristics that impact the performance of Spanish municipalities in the Circular City Index, as revealed by non-spatial and spatial models, and how could these improve the interpretability of the results of the Circular City Index?".* Building on these previous sections, this section tries to identify the key socio-demographic and economic characteristics that impact the performance of municipalities by presenting the results of non-spatial and spatial models. The modelling of data contribute to the improvement of the interpretability of the index results.

5.3.1 Data Description

As discussed in the case study chapter, data was collected for twenty features that represented socio-demographic and economic characteristics. Important to mention here is that the socio-demographic and economic data were collected via the online data portal of INE[15], thus incorporated the datasets information of 8,087 instead of 8,217 case study areas. As the features related to income per capita (6,519) and unemployment benefits (6,289) were missing observation of approximately 20% of the municipalities, these features were excluded from the data model. The income per household feature had 8 missing values, and this effect was expected to be negligible.

5.3.2 Descriptive Statistics

Out of the 18 features analyzed, 12 showed a significant positive correlation with the CCI score, while 4 demonstrated a significant negative correlation, and 2 demonstrated no significant correlation. Upon closer inspection, the features related to natural population growth, gender, age, housing, and wealth distribution exhibit a Pearson's R greater than (-)0.30. In contrast, features associated to the population size, population density, income per household, and municipal debt showed lower correlation coefficients, with a Pearson's R between 0.11 and 0.21. The remaining features demonstrate a Pearson's R of (-)0.05 or smaller. Lastly, the two features related to agriculture did not reveal a significant relationship with the CCI scores, as their p-values were higher than 0.05. An overview of the correlation matrix is presented in Appendix G.

5.3.3 Data Modelling

The presented correlations present the first initial insights into the relation between the CCI scores and the socio-demographic and economic characteristics. To form a foundation for an accurate analyzation and prediction of these relations, two non-spatial were constructed ahead of the final spatial model.

5.3.3.1 First Model

The sensitivity analysis of the first model revealed that in the perturbative approach six features had coefficients that were both positive and negative. The range of the boxplots for most features was narrow, except for the population size feature.

[15] https://www.ine.es/

5.3.3.2 Second Model

The second model tried to stabilize the model coefficients by implementing modifications to the first model. One of the initial steps was to test for normality using the D'Agostino's K-squared test, which showed that the data of 17 features did not follow a normal distribution and had a p-value smaller than the alpha of 0.05. To address this, the data was normalized using by the Yeo-Johnson power transform function. Afterwards, the feature selection process resulted in the best model, consisting of 11 features with a MAE of 0.73. The following features were selected as best model:

1. Population
2. Population female (in %)
3. Average age of population
4. Population non-Spanish citizen (in %)
5. Total residential buildings (per capita)
6. Income (per household)
7. GINI index
8. Municipal debt (per capita)
9. Total companies (per capita)
10. Total agricultural cattle farms (per km2)
11. Total tourist houses (per capita)

Model Output

By integrating these selected features in the model, the model was performed. In the estimates of the regression coefficients were given. The estimates revealed that each feature had a different effect on the dependent variable differed. The population feature had the most substantial estimated effect on the CCI outcome, with a positive coefficient of 0.681. In addition, income per household had the second largest effect with an estimated coefficient of 0.102. Among the features analyzed, only the GINI index (β of -0.087), total agricultural cattle farms per square kilometer (β of -0.071), average age of population (β of -0.036), and total companies per capita (β of -0.024) showed significant negative estimated coefficients. All other features were positive correlated to the CCI score. Finally, all model outputs were significant as all t-values were greater than 2, or less than -2, and all p-values were less than the alpha of 0.05.

Table 15 - Model Outputs of Transformed Second Model.

Transformed Second Model Outputs				
Goodness-of-fit measure				**Value**
AIC				17763.197
R-squared				0.475
Feature	**Estimate (β)**	**Standard Error (SE)**	**t-value**	**Significance (p-value)**
Population	0.681	0.018	38.229	0.000
Population female (in %)	0.071	0.010	7.098	0.000
Average age of population	- 0.036	0.015	- 2.343	0.019
Population non-Spanish citizen (in %)	0.054	0.010	5.596	0.000
Total residential buildings (per capita)	0.081	0.017	4.865	0.000
Income (per household)	0.102	0.010	10.636	0.000
GINI index	- 0.087	0.011	- 8.201	0.000
Municipal debt (per capita)	0.036	0.010	5.596	0.000
Total companies (per capita)	- 0.024	0.010	- 2.345	0.019
Total agricultural cattle farms (per km^2)	- 0.071	0.009	4.865	0.000
Total tourist houses (per capita)	0.071	0.009	8.158	0.000

<u>Model Validation</u>

Third, the model was validated. This was done by analyzing the model fit, model residuals, and sensitivity analysis before (second model) and after data normalization (transformed second model). In terms of the model fit, the second model produced an R-squared of -8.02, which was outside the acceptable range of 0 to 1. In contrast, the transformed second model produced a R-squared of 0.47. Figure 14 shows the model fits by a scatterplot. The left scatter plot clearly shows that the second model was experiencing issues. The right scatter plot shows improvements in the transformed second model, as the scattered values are plotted more linearly. In addition, this transformed second model has a MAE of 0.41 and an RMSE of 0.42. In terms of the model residuals, differences between the models were observed.

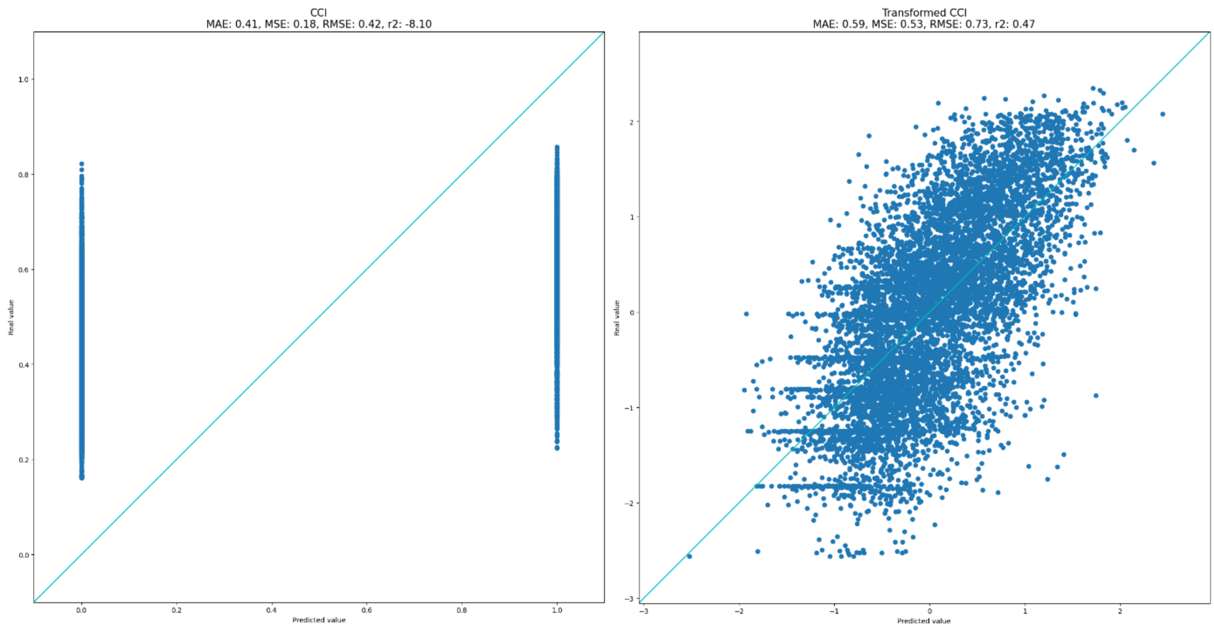


Figure 14 - Model Fit - Second Model (L) vs Transformed Second Model (R).

Figure 15 displays the relative errors of both models on a map. The left map displays relative errors of the second model and shows a clear patterns of spatial clustering in the relative errors. In contrast, the right map refers to the relative errors of the transformed second model, and displays less patterns of spatial clustering.

Finally, a sensitivity analysis was performed on the transformed second model. As for the sensitivity analysis of the first model, the sensitivity was analyzed through a perturbative study. In contrast to the first model, the sensitivity analysis of the second model reflected that no feature coefficients were both negative and positive. An overview of the sensitivity analysis of the second model could be found in the Appendix H.

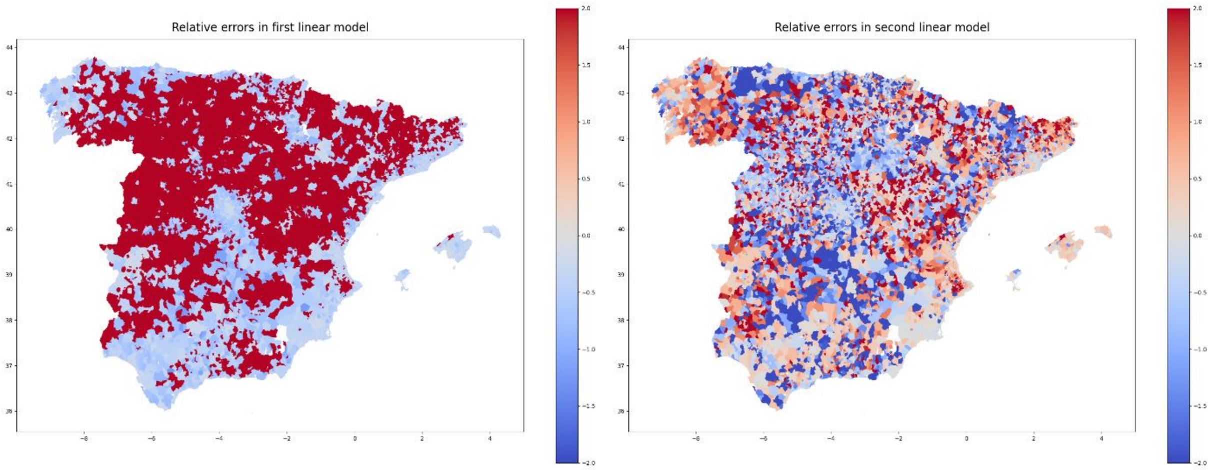


Figure 15 - Relative Errors per Municipality - Second Model (L) vs Transformed Second Model (R).

5.3.3.3 Geographically Weighted Regression (GWR) Model

The final model built on the aforementioned models. Integrating the selected features of the second model, the data was imputed and normalized. Afterwards, the features multicollinearity was tested by the VIF. Three features scored a relatively high value compared to the other features: population (VIF of 4.88), total residential buildings per capita (VIF of 4.26), and average age of population (VIF of 3.61). However, none of the features indicated a VIF-value higher than 5 - which was set as threshold – and thus none are highly correlated with each other. Subsequently, the model was trained. First, the spatial weightings were specified by an adaptive bi-square kernel, and the optimal bandwidth of 451 k-nearest neighbors was selected.

Model Outputs

Afterwards, the model was performed. First, the GWR model reflected the varied effect of each feature on the dependent variable, as well as the varying effect per case study area, as given by the estimates of the regression coefficients in Table 16. The population feature had the highest global correlation, with a mean estimated coefficient of 0.635. The feature with the second largest estimated effect was the feature of total residential buildings per capita, with a mean estimate coefficient of 0.086.

In contrast, the GINI index (mean β of -0.076), total agricultural cattle farms per square kilometer (mean β of -0.051), average age of population (mean β of -0.078), and total companies per capita (mean β of -0.001) revealed significant negative mean estimated coefficients. For all these outcomes, the adjusted alpha had a value of 0.001 and the adjusted critical value was 3.265. The standard deviation of the estimated coefficient, as given in Table 16, reflects that these numbers could differ between municipalities. In specific, the features related to agricultural cattle farms (0.198), population size (0.157), residential buildings (0.129), and GINI index (0.126) show relative high standard deviations.

Table 16 - Model Outputs of GWR Model.

Transformed Second Model Outputs		
Goodness-of-fit measure		**Value**
AIC		16587.712
R-squared		0.598
Adjusted alpha (95%)		0.001
Adjusted critical t-value		3.265
Feature	**Mean Estimate (β)**	**Standard Deviation of Estimate (β)**
Population	0.635	0.157
Population female (in %)	0.054	0.089
Average age of population	- 0.078	0.109
Population non-Spanish citizen (in %)	0.026	0.090
Total residential buildings (per capita)	0.086	0.129
Income (per household)	0.039	0.118
GINI index	- 0.076	0.126
Municipal debt (per capita)	0.040	0.063
Total companies (per capita)	-0.001	0.075
Total agricultural cattle farms (per km^2)	- 0.051	0.198
Total tourist houses (per capita)	0.032	0.061

To put these model outputs in a better context, the results were showed on a map. Figure 16 displays that the population feature was significant for almost all municipalities in Spain, with higher estimated coefficients up to 1.2 in north-western Catalonia. On the contrary, population was not a significant feature in the Mallorca and central-southern regions of Spain. Figure 17 shows that the residential buildings per capita had the largest significant positive coefficient in small regions in central and northeastern Spain, with estimates up to 0.4. In northeastern Spain, two small clusters of municipalities had a significant correlation for the residential building features. In contrast, this feature had a negative coefficient of approximately -0.3 in the metropolitan area of Barcelona.

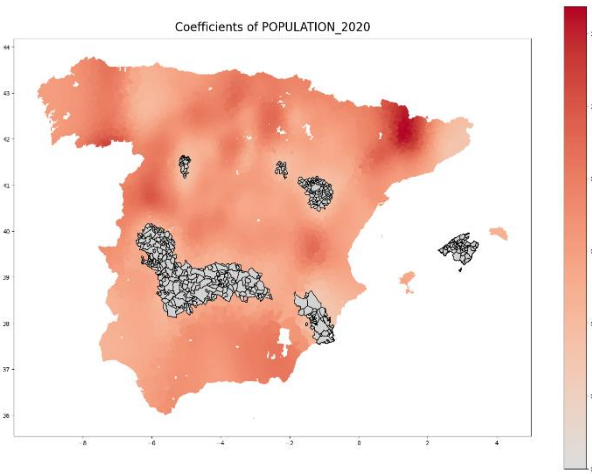


Figure 16 - GWR Model Output: Significant Estimated Coefficients per Municipality. Feature: Population.

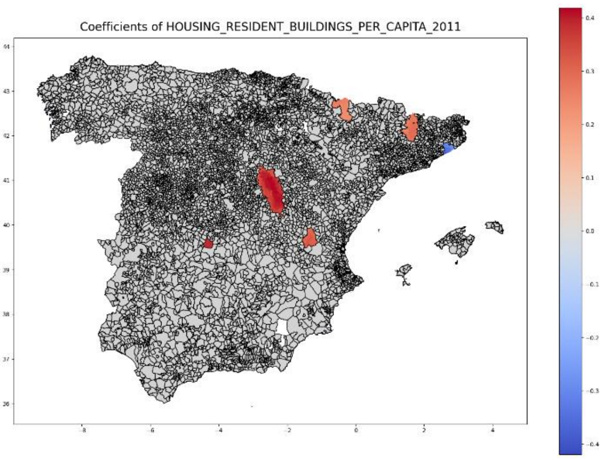
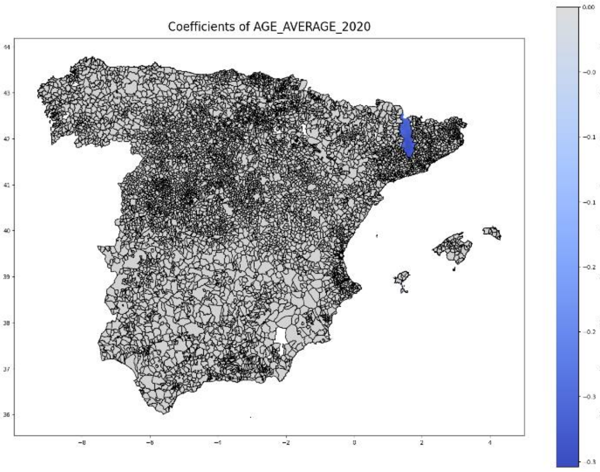


Figure 17 - GWR Model Output: Significant Estimated Coefficients per Municipality. Feature: Residential Buildings per Capita.



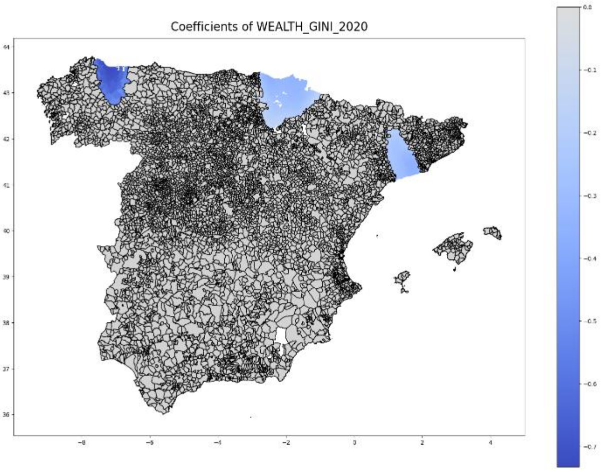
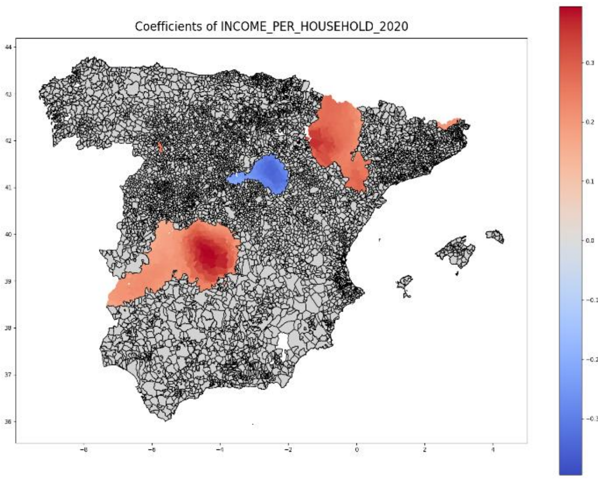


Figure 18 - GWR Model Output: Significant Estimated Coefficients per Municipality. Feature: Average Age of Population.

Figure 19 - GWR Model Output: Significant Estimated Coefficients Municipality. Feature: GINI Index

Figure 18 reveals that the average age of the population was only significant correlated in Catalonia, with a negative coefficient of -0.35. For the rest of the municipalities in Spain, this coefficient was not significant. Figure 19 indicates that the GINI index was significantly correlated in parts of Asturias, Basque Country, Catalonia, and Navarre. In particular, the feature had an estimated coefficient of -0.7 in parts of Galicia.



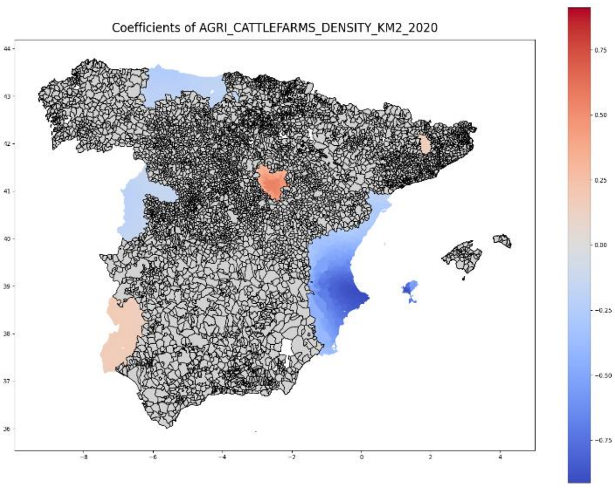


Figure 20 - GWR Model Output: Significant Estimated Coefficients per Municipality. Feature: Income per Household

Figure 21 - GWR Model Output: Significant Estimated Coefficients per Municipality. Feature: Agricultural Cattle Farms (per square kilometer).

Figure 20 illustrates that the income per household was estimated to be both significantly positive and negative correlated with the index scores, depending on the municipality. The significant positive coefficients were estimated in parts of Aragon, Castile-La Mancha, Catalonia, Extremadura, and Navarre. In particular, the feature was estimated to have a positive correlation with the index score of almost 0.4 in westerns parts of Castile-La Mancha. In contrast, the income per household was estimated to have a significant negative correlation in parts of north Castile-La Mancha and west Castile and Leon. Finally, Figure 21 presents the estimated coefficients for feature of agriculture cattle farms. Interesting about these results is that this feature had a both highly significant negative and positive estimate coefficients. Positive coefficients were estimated in western Andalusia, southwestern Extremadura, and northern Castile-La Mancha. The last-mentioned region was estimated to have a coefficients up to 0.5. In opposition, negative coefficients were estimated in western parts of Castile and Leon, and in almost the entire autonomous communities of Asturias, Cantabria, and Valencia. The index outcomes in regions around the city of Valencia were estimated to be highly negative correlated to this feature.

In addition to these detailed results per feature, a more global number of significant features per municipality could indicate how well the model features in general were performing. The population feature was significantly correlated to the index score in more than 90% of the municipalities, while the residential buildings was significant in almost 20%, and agriculture cattle farms in approximately 10%. On the contrary, the other features were significant in 5% or less of the municipalities.

Figure 22 shows the number of significant features per municipality. The majority of municipalities had at least 1 or 2 significant features. Municipalities with more than 2 significant features were mostly clustered in the autonomous communities Castile and Leon, Castile-La Mancha, Catalonia, La Rioja, Navarre, and Valencia. Nonetheless, there were no features estimated to be significant for municipalities in the Mallorca and several central-southern regions.

Finally, Figure 23 presents the most important features per municipality. For most municipalities in Spain, population was the most important feature to predict the index outcome. In addition, the feature agriculture cattle farms was the largest coefficient around Valencia, and the features GINI index and total companies were most critical in southern parts of Castile-La Mancha. On a more detailed level, municipal debt was the most significant feature in municipalities around the city of Valladolid, and the total companies was the most important feature in the Catalan capital Barcelona.

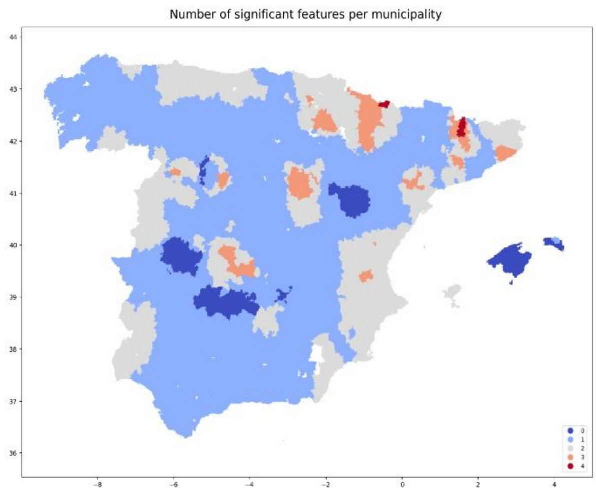


Figure 22 - Number of Significant Features per Municipality.

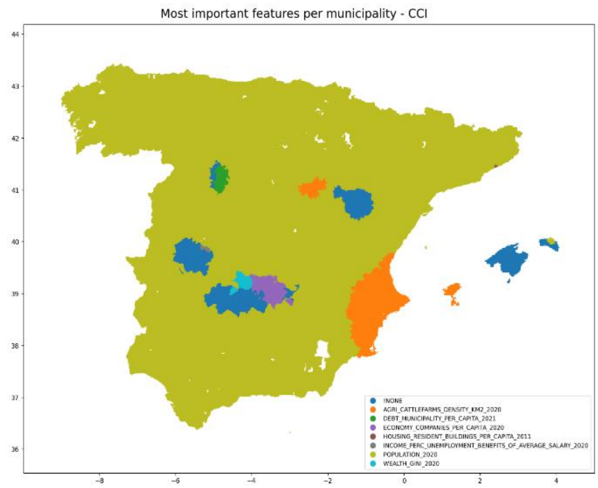


Figure 23 - Feature with Highest Estimate Coefficient per Municipality.

Model Validation

To validate these model outputs, a validation was performed. First, the model had a R-squared of 0.598. In addition, the model produced a MAE of 0.51 and a RMSE of 0.63. Additionally, the Local R-squared per municipality provided clearer results of the model fit, which is presented in Figure 24. For the majority of municipalities, the model had a R-squared of around 0.50. For areas in the Valencian Community and the Madrid area, the model had a R-squared of around 0.60, resulting in a higher goodness-of-fit. Remarkably, the western part of Catalonia had a significant higher model fit, as the R-squared in these areas peaked up to almost 0.70. With regards of the model residuals, Figure 25 shows no clear patterns of spatial clustering for the relative errors in the model. Thus, the variance of model residuals was more or less constant across the model features.

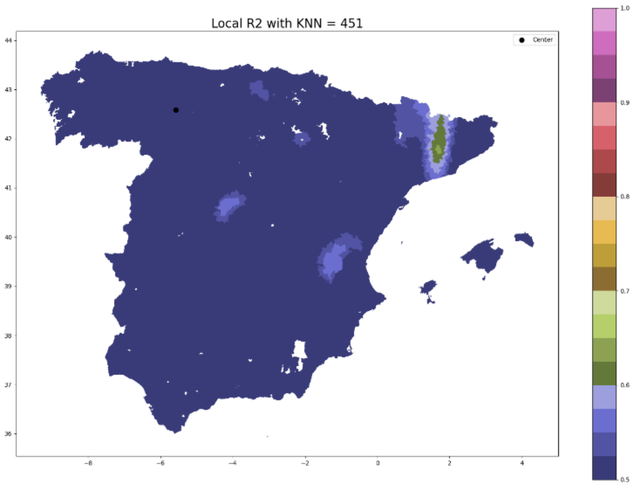


Figure 24 - Local R-Squared per Municipality - GWR Model.

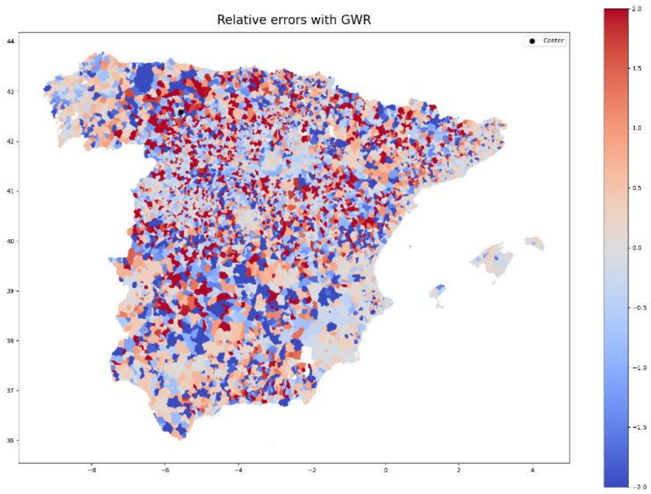


Figure 25 - Relative Errors per Municipality - GWR Model.

As a third part of the model validation, the sensitivity analysis was performed. The perturbative study shows that several model features reflect both negative and positive coefficients. In particular, this was the case for the features of agricultural cattle farms, companies, and non-Spanish population. In addition, the features related to residential buildings, and income per household had outliers that showed negative estimated coefficients, instead of positive. Nevertheless, the majority of the model features seemed to act robust. An overview of the sensitivity analysis of the GWR model could be found in the Appendix H.

Finally, the local multicollinearity of the features was tested, in addition to the earlier multicollinearity test. This local variant was estimated by computing the VIF for each single GWR. Figure 26 shows that the highest measured VIF was 3.44, which is below the maximum threshold of 5. In short, none of the single GWRs faced multicollinearity issues.

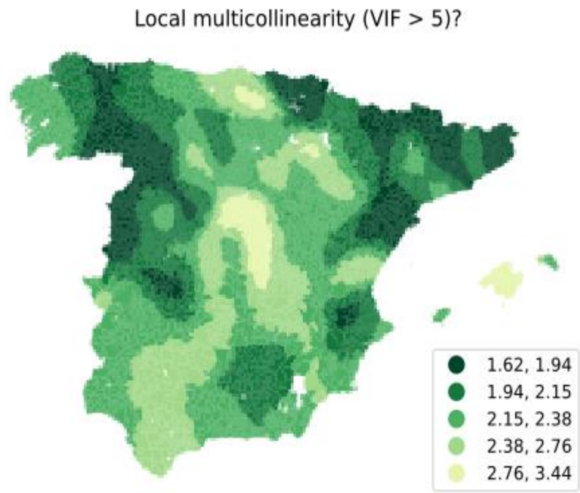


Figure 26 - Local Multicollinearity in the GWR Model.

6. Discussion, Conclusion and Recommendation

In this thesis, the main objective was to reformulate and redefine the CCI for Spain, and to evaluate the level of preparedness of Spanish municipalities towards urban circularity and green transition. In addition, the thesis aimed to improve the interpretability of the index results by identifying key socio-demographic and economic characteristics that impact the performance of municipalities.

This concluding chapter is structured as follows. In the first sections, the main findings of the thesis are discussed in relation to the research questions. Afterwards, the general conclusion of this thesis is drawn, based on the discussed results. Finally, recommendations for future work are given.

6.1 Discussion

In this section, the three research sub-questions are answered, based on the methodology and results presented in this index. First, the research sub-question related to the extent of reformulation and redefinition of the original CCI is answered. Second, the research sub-question related to implementation of the index and evaluation of the preparedness of municipalities is discussed. Third, the research sub-question related to the identification of the key socio-demographic and economic characteristics that impact the index outcomes is debated.

6.1.1 Extent of Reformulation and Redefinition for the Circular City Index in Spain

By discussing the methodology of section 3.1, and the results of section 5.1, this section answers the first research sub-question: "What extent of reformulation and redefinition is required for the original Circular City Index to be effectively applied to the Spanish context?"

In Italy, the CCI demonstrated to be an effective tool for assessing the readiness of municipalities towards urban circularity and green transition at municipal level. However, the original index was established for one specific context and was not directly appropriate for Spain. The methodology had to be adapted to suit local contexts and guarantee the effective capturing of urban circularity. Therefore, reformulations and redefinitions of the CCI methodology were necessary to make it more relevant and effective for measuring the preparedness of Spanish municipalities.

The first key adjustment this thesis makes to the CCI methodology is the exclusion of indicators. Although the original CCI methodology comprises a total of 17 indicators, the reformulated index includes 14. The exclusion of three indicators (ECR3, ECR6, and W1) is a direct result of the lack of open data availability, a harmonized approach to share data formats, data types, and metadata in Spain. In addition, three indicators (D1, D2, and W2) are reformulated to suit the local contexts. As a consequence, the application of an identical copy of the original index is no longer possible, which complicates a possible comparative study between the two indices in the future. As a matter of course, researchers need to exercise caution when making comparisons between the original index and the Spanish version. Any comparisons made between Spain and Italy must also be considered in light of the excluded indicators.

Additionally, the thesis studies the index benchmarks, which are partly derived from policies and legislation implemented in Italy. Several benchmarks are integrated into the index, despite the policies they are based on not being implemented in Spain. As the redefinition of the benchmarks is out of research scope, no adjustment are made to the implement benchmarks. Consequently, the lack of adjustments leads to peculiar results for certain indicators. For instance, the benchmarks of indicators related to pollution (ECR4 and ECR5) are met by almost all municipalities, resulting in the lack of meaningful insights into their actual performance in these fields. To address this issue, this thesis suggest reevaluating the benchmarks and the associated area values computations used in the CCI. This would include the reviewing of the appropriateness of benchmarks in the context of local policies and legislations. A more accurate capturing of the actual situation could be ensured by this, and more meaningful understandings into the performance of municipalities could be obtained.

Finally, this thesis identifies the need to adjust the single KPI weights ($W_k \epsilon$) in the Spanish context, because of the exclusion of three indicators in the energy, climate, and resources set and waste set. The reweighting preserves the balance of the index, and avoids a highly influenced index due to the overvaluation of specific KPIs. In order to adjust the weighting of the indicators, three potential approaches are considered. The first approach concerns the modification of the original KPIs ECR3 and ECR6 to make them more suitable for data accessibility and availability in the Spanish context. Nonetheless, this would create new challenges when comparing the original CCI with the Spanish version. The second approach is to redefine the entire weight system of single KPIs to preserve their original relative weighted value in relation to the total index. This would mitigate the impact of missing indicators, but would complicate the comparison between the two indices. The third approach involves redefining the weights of simply the KPIs in the second and fourth sets. This approach would require the calculation of new weights based on the original relative value of each KPI in comparison to the other KPIs in the same set, and thus preserving the comparability of the indices while accounting for missing indicators. In this research, the third approach is implemented, and only the weights of the second and fourth set are reweighted. This approach is considered to be most similar the original KPI weight, and avoids major problems when making comparisons between the original and Spanish index. While these adjustments to the weighting system are necessary to accurately measure the preparedness of municipalities, they also do imply that the modified CCI no longer fully aligns with the core principles of the original index. Therefore, policymakers must carefully consider the measured outcomes of indicators where the weighting has been adjusted, as this may give a misleading understanding of reality.

6.1.2 Evaluation of the Preparedness of Municipalities

By discussing the methodology of section 3.2, and the results of section 5.2, this section answers the second research sub-question: "How does the Circular City Index evaluate the preparedness of Spanish municipalities towards urban circularity and green transition?"

After implementing the reformulated and redefined CCI, the preparedness of Spanish municipalities towards urban circularity and green transition was assessed. The results provide useful insights into the degree of preparedness of municipalities. This section delves deeper into the index results, exploring trends in the data based on descriptive statistics, municipality size and geographic distribution. The discussion highlights the importance of the index as a valuable instrument for policymakers to realize a more sustainable future in municipalities.

Descriptive statistics are first used to present the input data. Remarkably is that there are numerous missing values in the data of indicators M1, M2, and M3. The missing values could represent the absence of the measured elements, or a lack of OpenStreetMap data contributors in these municipalities. Additionally, the table indicate that D3, M4, and W2 have 8,131 observations instead of 8,217, and this implies that these indicators only provided information at municipal level, and did not incorporate data on population nucleus or minor overseas territories.

In addition, the descriptive statistics show that the Spanish municipalities achieved an average score of 0.454, with a standard deviation of 0.145. This indicates the degree of preparedness varies significantly between municipalities. Further, municipalities scored on average high on digitalization. This could imply that local governments prioritize efforts to digitalization and make considerable progress on this topic. In contrast, the politicians might lack attention - or motivation - to put emphasis on transportation and mobility, as municipalities scored in general low on mobility. Likewise, the index reveals a high standard deviation for waste, implying that particular local governments concentrate on waste management, while other municipalities may not consider this a priority.

Moreover, when comparing municipality population size with the index scores, the results reveal that larger and medium-sized municipalities had a right-skewed distribution. This might imply that municipalities with larger population sizes are more prepared towards urban circularity and sustainability compared to smaller-sized municipalities. By benefitting from the lessons learned from larger municipalities, policymakers should prioritize the development of policies and interventions that address sustainability and circularity challenges faced by small municipalities. These findings also highlight the importance of incorporating the context of municipality-size when designing policies and interventions towards sustainability and circularity.

Finally, the spatial analysis provides a more in-depth understanding of the observed geographic distributions. A Global Moran's I of 0.46 was measured for the CCI scores, which indicates a significant moderate to strong positive global spatial autocorrelation in Spain. In other words, the preparedness of municipalities is influenced by the geographic proximity to other municipalities, as the results tend to spatially cluster. Furthermore, the analysis indicates that spatial clusters with a relatively high degree of preparedness are mainly located in coastal areas, such as the Balearic Islands, Catalonia, Galicia, Murcia, Valencia, and around the Strait of Gibraltar. These results imply that the proximity to the sea might be a contributing factor to a higher degree of preparedness, possibly due to the increased exposure to impacts of climate change in these areas. Conversely, the central-northwestern regions of Spain located spatial clusters with a comparatively lower degree of preparedness. These regions achieved lower scores on climate and waste-related indicators, providing evidence to strengthen the assumption that the preparedness in areas might be linked to its vulnerability to effects of climate change. Nevertheless, the Madrid area is an exception to this implication, as it does not have a close proximity to sea, but revealed to have a relatively high degree of preparedness. This might imply that other local characteristics play a significant role in the degree of preparedness.

6.1.3 Identification of the Key Socio-Demographic and Economic Characteristics

By discussing the methodology of section 3.3, and the results of section 5.3, this section answers the third research sub-question: "What are the key socio-demographic and economic characteristics that impact the performance of Spanish municipalities in the Circular City Index, as revealed by non-spatial and spatial models, and how could these improve the interpretability of the results of the Circular City Index?"

The study implemented a three-phase modelling process to develop a dependable model for interpreting index outcomes. The first model initially lacked robustness. By selecting optimal model features and normalizing data, improvements were made to the second model in terms of normality, model fit, residuals, and robustness. The final model incorporated the spatial component, resulting in a better goodness-of-fit compared to the second non-spatial model. In particular, the spatial model had a higher R-squared of 0.598 and a lower AIC of 16,587, indicating that the model explains 59.8% of the variance in the dependent variable, compared to the non-spatial model's R-squared of 0.475 and AIC of 17,763. Furthermore, the residual analysis revealed no clear patterns of spatial clustering, suggesting that residuals behave constant and that no additional features were required. Although the non-spatial model exhibited better coefficient robustness, the spatial model captures more accurately key socio-demographic and economic characteristics, explaining nearly 60% of the index results. This proves that the model could provide valuable insights for policymakers seeking to interpret the index. It should be noted that the lack of significant correlation among model features in municipalities in the Canary Islands and Mallorca suggests that the model may be less suitable for Spanish regions beyond the Iberian Peninsula.

The results of the model indicate that socio-demographic and economic characteristics play a significant role in the degree of preparedness towards urban circularity and green transition. In particular, the model reveals that the population variable has the largest global effect on the index outcome, and is significant in almost all municipalities in Spain. In other words, municipalities with larger populations tend to be more prepared towards urban circularity and the green transition than those with smaller populations. A possible explanation could be that a large population often comes with greater economic and social resources, for instance higher levels of economic activity, or education. This might generate more tax revenues, which could give municipalities more opportunities to implement circular or sustainable policies. On top of this, municipalities with a large population may generate a greater demand for sustainable and circular services. This could incentivize policy-makers to prioritize sustainable development, and invest in more circular solutions.

Furthermore, the total residential buildings per capita had a significant positive global effect on the index, indicating that municipalities with a higher number of residential buildings per capita are more ready towards urban circularity and green transition. Normally, this relation could be clarified by the fact that higher numbers of residential buildings per capita suggest greater urban density and urbanization, which often relate to circular characteristics of more walkable environments, better-connected public transportation networks, or more efficient waste management. However, the spatial model output shows that the opposite is the case. In rural areas, the residential buildings per capita are significantly positively correlated, while in Barcelona the relationship is significantly negatively correlated. Therefore, external local contexts may explain the relation in these areas.

On the other hand, both the average age of the population and the wealth inequality showed to be negatively correlated to the preparedness of municipalities. For the average age of the population, an explanation could be that older population groups may stick to traditional practices and habits, and may be less willing or able to adapt to new practices. In contrast, a younger population may be more willing to adapt to circular practices in their lives, such as public transportation or electronic public services. For wealth inequality, an explanation could be that municipalities that have a greater unequal distribution of wealth, may be more likely to prioritize initiatives that benefit certain subclasses, instead of the entire community. Specifically, this significant correlation was observed in municipalities of autonomous communities that strive for more autonomy. Therefore, in the communities of Basque Country, Catalonia, and Navarre, a supplementary explanation could be that a significant part of the population emphases local identities. This could contribute to a higher degree of wealth inequality due to the presence of more segregated subgroups within the society. These results may imply that policies encouraging population growth and urbanization could lead to better sustainable outcomes. The significant negative impact of variables such as wealth inequality and average age of the population indicates that policies targeted at reducing income inequality and involving elderly in the participation processes could lead to improved urban circularity and green transition readiness.

6.2 Conclusion

In this thesis, a reformulated and redefined version of the Circularity City Index (CCI) is presented that contributes to the growing necessity for a more comprehensive and universal approach to quantify and assess circular economies. The modified index was tested in a case study of 8,217 municipalities in Spain to evaluate their preparedness towards urban circularity and green transition. As the index centers on open data principles, the thesis is fully reproducible, scalable, and built on open source data, methodologies, and software.

This study highlights the importance of adapting the CCI methodology to fit local contexts in order to ensure its effectiveness. The study developed a reliable spatial model that explained the relationships between the preparedness of municipalities and local socio-demographic and economic characteristics. The findings reveal that population size is the primary socio-demographic characteristic that impacts urban circularity, with larger populated municipalities generally being more prepared for green transition than smaller ones.

The study findings provide a foundation for future research on this topic. When taking into consideration socio-demographic and economic characteristics, the index is a valuable tool for policymakers to identify challenges, and to develop sustainable policies at the municipal level. The results of this study can therefore be applied to wider contexts, and support the transition towards more sustainable and circular urban systems.

6.3 Recommendations for Future Research

Future research is recommended to reconsider the measurement level of indicators, as some indicators measure processes that might take place on a higher aggregation than the municipal level, which might cause issues related to ecological fallacy. In addition, there should be a focus on redefining the implemented benchmarks, based on relevance of local policies and legislation. The original benchmarks do not match the context of Spain, and can therefore give a distorted depiction of the outcomes. In addition, a more nuanced index weighting system should be developed to incorporate the essence of the original index, while accommodating the effect of excluded indicators. The current modified weighting system fails to fully capture the fundamental aspects of the original CCI. Finally, to avoid the MAUP, future research should incorporate tests of model fit with different specified spatial weightings.

By integrating the aforementioned limitations and recommendations in future research, the validity and reliability of the index can be improved, making it a valuable tool for evaluating circular economies at the municipal level. Therefore, this study provides a strong basis for advancing the Circular City Index as a comprehensive and universal approach to sustainable development assessment.

References

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L. (2005). *Exploring Spatial Data with GeoDaTM : A Workbook: Vol. Revised Version*. Center for Spatially Integrated Social Science.

Banco de España. (2021). *Annual Report 2021. Chapter 4: The Spanish Economy and the Climate Challenge* [Annual Report]. Banco de España. https://www.bde.es/f/webbde/SES/Secciones/Publicaciones/PublicacionesAnuales/InformesAnuales/21/Files/InfAnual_2021-Cap4_En.pdf

Beale, C. M., Lennon, J. J., Yearsley, J. M., Brewer, M. J., & Elston, D. A. (2010). Regression analysis of spatial data. *Regression Analysis of Spatial Data*, *13*(2010), 246–264.

Bidanset, P. E., & Lombard, J. R. (2014). The Effect of Kernel and Bandwidth Specification in Geographically Weighted Regression Models on the Accuracy and Uniformity of Mass Real Estate Appraisal. *Journal of Property Tax Assessment & Administration*, *10*(3), 5–14.

BOE. (2021). *Disposición 5106 del BOE núm. 77 de 2021* (BOE-A-2021-5106; BOLETÍN OFICIAL DEL ESTADO, pp. 36796–37220). BOE. https://www.boe.es/boe/dias/2021/03/31/pdfs/BOE-A-2021-5106.pdf

Boeing, G. (2022). *OSMnx 1.2.2*. https://osmnx.readthedocs.io/en/stable/

Chi, G., & Zhu, J. (2008). Spatial Regression Models for Demographic Analysis. *Population Research and Policy Review*, *27*(1), 17–42. https://doi.org/10.1007/s11113-007-9051-8

Chun-rong, J., & Jun, Z. (2011). Evaluation of Regional Circular Economy Based on Matter Element Analysis. *Procedia Environmental Sciences*, *11*, 637–642. https://doi.org/10.1016/j.proenv.2011.12.099

Cl@ve. (2014). *¿Qué es? - Cl@ve*. https://clave.gob.es/clave_Home/clave/queEs.html

D'Agostino, R., & Pearson, E. S. (1973). Tests for Departure from Normality. Empirical Results for the Distributions of b2 and √ b1. *Biometrika*, *60*(3), 613–622. https://doi.org/10.2307/2335012 del Valle Gálvez, A. (2011). Ceuta, Melilla, Chafarinas, Vélez y Alhucemas: Tomar la iniciativa. *Real Instituto Elcano*, *163/2011*, 1–10.

Dobson, A. J. (2002). *An introduction to generalized linear models* (2nd ed). Chapman & Hall/CRC. https://reneues.files.wordpress.com/2010/01/an-introduction-to-generalized-linear-models-second-edition-dobson.pdf

Enel. (2021, september 29). *Enel X launches the first "Circular City Index" for Italian municipalities*. Enel X Launches the First "Circular City Index" for Italian Municipalities. https://www.enel.com/media/explore/search-press-releases/press/2021/09/enel-x-launches-the-first-circular-city-index-for-italian-municipalities

European Commission. (2018). *Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: On a Monitoring Framework for the Circular Economy* (COM(2018) 29 final). https://ec.europa.eu/environment/circular-economy/pdf/monitoring-framework.pdf

European Commission. (2020). *Communication From the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. A new Circular Economy Action Plan: For a Cleaner and More Competitive Europe* (COM(2020) 98 final; A New Circular Economy Action Plan). https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1583933814386&uri=COM:2020:98:FIN

European Environment Agency. (2022). *EEA geospatial data catalogue*. https://sdi.eea.europa.eu/catalogue/srv/eng/catalog.search#/metadata/3c72b3a2-af82-4595-82b2-b205e74a90cf

Fan, Y. V., Lee, C. T., Lim, J. S., Klemeš, J. J., & Le, P. T. K. (2019). Cross-disciplinary approaches towards smart, resilient and sustainable circular economy. *Journal of Cleaner Production*, *232*, 1482–1491. https://doi.org/10.1016/j.jclepro.2019.05.266

Geissdoerfer, M., Savaget, P., Bocken, N. M. P., & Hultink, E. J. (2017). The Circular Economy – A new sustainability paradigm? *Journal of Cleaner Production*, *143*, 757–768. https://doi.org/10.1016/j.jclepro.2016.12.048

Geng, Y., Fu, J., Sarkis, J., & Xue, B. (2012). Towards a national circular economy indicator system in China: An evaluation and critical analysis. *Journal of Cleaner Production*, *23*(1), 216–224. https://doi.org/10.1016/j.jclepro.2011.07.005

GGGI. (2020). *Green Growth Index 2020: Measuring performance in achieving SDG targets* (Nr. 16; GGGI Technical Report). Global Green Growth Institute. https://greengrowthindex.gggi.org/wp-content/uploads/2021/01/2020-Green-Growth-Index.pdf

Ghisellini, P., Cialani, C., & Ulgiati, S. (2015). *A review on circular economy: The expected transition to a balanced interplay of environmental and economic systems*. https://doi.org/DOI: 10.1016/j.jclepro.2015.09.007

Gregori, E., & Reyes Valenzuela, P. (2021). *Urbana* [Python / Jupyter Notebook; GitHub Repository]. Barcelona Supercomputing Center. https://github.com/egregorimar/urbana

Guo-gang, J. (2011). Empirical Analysis of Regional Circular Economy Development—Study Based on Jiangsu, Heilongjiang, Qinghai Province. *Energy Procedia*, *5*, 125–129. https://doi.org/10.1016/j.egypro.2011.03.023

Helbich, M., & Griffith, D. A. (2016). Spatially varying coefficient models in real estate: Eigenvector spatial filtering and alternative approaches. *Computers, Environment and Urban Systems*, *57*(2016), 1–11. https://doi.org/10.1016/j.compenvurbsys.2015.12.002

Holcim, & Bloomberg Media. (2022). *The Circular Cities Barometer*. Bloomberg. https://sponsored.bloomberg.com/immersive/holcim/the-circular-cities-barometer#overview

IGN. (2022). *Información Geográfica de Referencia: Límites municipales, provinciales y autonómicos*. Centro de Descargas del CNIG. http://centrodedescargas.cnig.es

IGN, I. G. (z.d.). *Cartografía y Datos geográficos—Instituto Geográfico Nacional*. Geoportal oficial del Instituto Geográfico Nacional de España. Geraadpleegd 15 februari 2023, van http://www.ign.es

INE. (2021). *Población residente por fecha, sexo y edad (2021)*. INE. https://www.ine.es/jaxiT3/Datos.htm?t=31304

INE. (2022). *Number of municipalities by Province, Autonomous Communities and island*. https://www.ine.es/en/daco/daco42/codmun/cod_num_muni_provincia_ccaa_en.htm

IPBES. (2019). *Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*. IPBES secretariat, Bonn, Germany. https://ipbes.net/sites/default/files/inline/files/ipbes_global_assessment_report_summary_for_policymakers.pdf

Jackson, T. (2011). *Prosperity without growth: Economics for a finite planet* (Reprinted). Earthscan.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications in R*. Springer. http://www.pdfdrive.com/an-introduction-to-statistical-learning-with-applications-in-r-d181506025.html

Kębłowski, W., Lambert, D., & Bassens, D. (2020). Circular economy and the city: An urban political economy agenda. *Culture and Organization*, *26*(2), 142–158. https://doi.org/10.1080/14759551.2020.1718148

Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. *Resources, Conservation and Recycling*, *127*, 221–232. https://doi.org/10.1016/j.resconrec.2017.09.005

Knippers, R., & Tempfli, K. (2013). Spatial referencing and satellite-based positioning. In V. A. Tolpekin & A. Stein, *The core of GIScience: A systems-based approach* (p. 524). University of Twente Faculty of Geo-Information and Earth Observation (ITC). https://issuu.com/itc-utwente/docs/corebook2013_03_spatialreferencing

Kölling, M., Rodriquez Teruel, J., & Colino, C. (2022). *Spain Report: Sustainable Governance Indicators 2022*. BertelsmannStiftung. https://www.sgi-network.org/docs/2022/country/SGI2022_Spain.pdf

Korhonen, J., Nuur, C., Feldmann, A., & Birkie, S. E. (2018). Circular economy as an essentially contested concept. *Journal of Cleaner Production*, *175*, 544–552. https://doi.org/10.1016/j.jclepro.2017.12.111

Kreyszig, E. (2011). *Advanced Engineering Mathematics* (10de dr.). John Wiley & Sons, Inc. https://www.bau.edu.jo/UserPortal/UserProfile/PostsAttach/59003_3812_1.pdf

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (Red.). (2005). *Applied linear statistical models* (5th ed). McGraw-Hill Irwin. https://users.stat.ufl.edu/~winner/sta4211/ALSM_5Ed_Kutner.pdf

Lucertini, G., & Musco, F. (2020). Circular Urban Metabolism Framework. *One Earth*, *2*(2), 138–142. https://doi.org/10.1016/j.oneear.2020.02.004

Maitre-Ekern, E. (2021). Re-thinking producer responsibility for a sustainable circular economy from extended producer responsibility to pre-market producer responsibility. *Journal of Cleaner Production*, *286*, 125454. https://doi.org/10.1016/j.jclepro.2020.125454

Merino-Saum, A., Halla, P., Superti, V., Boesch, A., & Binder, C. R. (2020). Indicators for urban sustainability: Key lessons from a systematic analysis of 67 measurement initiatives. *Ecological Indicators*, *119*, 106879. https://doi.org/10.1016/j.ecolind.2020.106879

Mhatre, P., Panchal, R., Singh, A., & Bibyan, S. (2021). A systematic literature review on the circular economy initiatives in the European Union. *Sustainable Production and Consumption*, *26*, 187–202. https://doi.org/10.1016/j.spc.2020.09.008

Ministerio para la Transición Ecológica y el Reto Demográfico. (2020). *PLAN NACIONAL DE ADAPTACIÓN AL CAMBIO CLIMÁTICO*. 246.

Muscillo, A., Re, S., Gambacorta, S., Ferrara, G., Tagliafierro, N., Borello, E., Rubino, A., & Facchini, A. (2021). *Circular City Index: An Open Data analysis to assess the urban circularity preparedness* *of cities to address the green transition -- A study on the Italian municipalities* (arXiv:2109.10832). arXiv. https://doi.org/10.48550/arXiv.2109.10832

Nikolaou, I. E., Jones, N., & Stefanakis, A. (2021). Circular Economy and Sustainability: The Past, the Present and the Future Directions. *Circular Economy and Sustainability*, *1*(1), 1–20. https://doi.org/10.1007/s43615-021-00030-3

Nikolaou, I. E., & Tsagarakis, K. P. (2021). An introduction to circular economy and sustainability: Some existing lessons and future directions. *Sustainable Production and Consumption*, *28*, 600–609. https://doi.org/10.1016/j.spc.2021.06.017

Nurdiana, J., Franco-Garcia, M. L., & Heldeweg, M. A. (2021). How Shall We Start? The Importance of General Indices for Circular Cities in Indonesia. *Sustainability*, *13*(20), Art. 20. https://doi.org/10.3390/su132011168

OBSAE. (2023). *DATAOBSAE - Área Atencion Ciudadano y Empresa*. https://dataobsae.administracionelectronica.gob.es/cmobsae3/dashboard/Dashboard.action

OpenStreetMap. (2022). *OpenStreetMap—Website*. https://www.openstreetmap.org/

Our World in Data. (2017). *World GDP over the last two millennia*. Our World in Data. https://ourworldindata.org/grapher/world-gdp-over-the-last-two-millennia

Paoli, F., Pirlone, F., & Spadaro, I. (2022). Indicators for the Circular City: A Review and a Proposal. *Sustainability*, *14*(19), Art. 19. https://doi.org/10.3390/su141911848

Petit-Boix, A., & Leipold, S. (2018). *Circular economy in cities: Reviewing how environmental research aligns with local practices*. https://doi.org/10.1016/j.jclepro.2018.05.281

PyPi. (2020). *fuzzywuzzy: Fuzzy string matching in python* (0.18.0) [Python]. https://github.com/seatgeek/fuzzywuzzy

PySAL. (2018). *Mgwr.sel_bw.Sel_BW — mgwr v2.1.1 Manual*. https://mgwr.readthedocs.io/en/latest/generated/mgwr.sel_bw.Sel_BW.html

Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *2020*

Third International Conference on Smart Systems and Inventive Technology (ICSSIT), 729–735. https://doi.org/10.1109/ICSSIT48917.2020.9214160

Randolph, K. A., & Myers, L. L. (2013). *Basic Statistics in Multivariate Analysis*. OUP USA.

Remøy, H., Wandl, A., Ceric, D., & Van Timmeren, A. (2019). Facilitating Circular Economy in Urban Planning. *Urban Planning*, *4*(3), 1–4. https://doi.org/10.17645/up.v4i3.2484

Royston, P. (1995). Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *44*(4), 547–551. https://doi.org/10.2307/2986146

Sánchez Levoso, A., Gasol, C. M., Martínez-Blanco, J., Durany, X. G., Lehmann, M., & Gaya, R. F. (2020). Methodological framework for the implementation of circular economy in urban systems. *Journal of Cleaner Production*, *248*, 119227. https://doi.org/10.1016/j.jclepro.2019.119227

Schroeder, P., Anggraeni, K., & Weber, U. (2019). The Relevance of Circular Economy Practices to the Sustainable Development Goals. *Journal of Industrial Ecology*, *23*(1), 77–95. https://doi.org/10.1111/jiec.12732

Shapiro, S. S., & Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, *52*(3/4), 591–611. https://doi.org/10.2307/2333709

Su, B., Heshmati, A., Geng, Y., & Yu, X. (2013). A review of the circular economy in China: Moving from rhetoric to implementation. *Journal of Cleaner Production*, *42*, 215–227. https://doi.org/10.1016/j.jclepro.2012.11.020

Tay, R. (2017). *Journal of the Eastern Asia Society for Transportation Studies*. *12*, 2006–2015.

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234–240. https://doi.org/10.2307/143141

Tura, N., Hanski, J., Ahola, T., Ståhle, M., Piiparinen, S., & Valkokari, P. (2019). Unlocking circular business: A framework of barriers and drivers. *Journal of Cleaner Production*, *212*, 90–98. https://doi.org/10.1016/j.jclepro.2018.11.202

United Nations. (2015). *Transforming our world: The 2030 Agenda for Sustainable Development.* (A/RES/70/1; General Assembly Resolution 70/1). United Nations. https://documents-dds-ny.un.org/doc/UNDOC/GEN/N15/291/89/PDF/N1529189.pdf?OpenElement

United Nations. (2019). *Circular Economy for the SDGs: From Concept to Practice General Assembly and ECOSOC Joint Meeting*. United Nations. https://www.un.org/en/ga/second/73/jm_conceptnote.pdf

Urban Agenda Partnership on Circular Economy. (2019). *Indicators for circular economy (CE) transition in cities—Issues and mapping paper* (Issues and mapping paper Version 4; Urban Agenda for the EU). Ecorys. https://ec.europa.eu/futurium/en/system/files/ged/urban_agenda_partnership_on_circular_economy_-_indicators_for_ce_transition_-_issupaper_0.pdf

Wheeler, D., & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, *7*(2), 161–187. https://doi.org/10.1007/s10109-005-0155-6

Zaman, A. U., & Lehmann, S. (2013). The zero waste index: A performance measurement tool for waste management systems in a 'zero waste city'. *Journal of Cleaner Production*, *50*, 123–132. https://doi.org/10.1016/j.jclepro.2012.11.041

Appendices

Appendix A: Index Area Values Computation in the Original CCI.

KPI Type	Area Value Computation Formula	KPI
Binary	$S_{kc} = kpi * 1$	D1, D2, D4, ECR1, ECR2, W3
Percentage	$If\ benchmark > 0 \rightarrow\ S_{kc} = kpi * benchmark$ $If\ benchmark = 0 \rightarrow\ S_{kc} = \frac{kpi - \min(K)}{\max(K) - \min(K)}$	D3, ECR3, W2
Threshold down	$If\ kpi\ value\ in\ I1 = [-1, 0.5 * benchmark] \rightarrow kpi = 4$ $If\ kpi\ value\ in\ I2 = [0.5 * benchmark, benchmark] \rightarrow kpi = 3$ $If\ kpi\ value\ in\ I3 = [benchmark, 1.5 * benchmark] \rightarrow kpi = 2$ $If\ kpivalue\ in\ I4 = [1.5 * benchmark, 2 * benchmark] \rightarrow kpi = 1$ $If\ KPI\ value\ in\ I5 = [2 * benchmark, \rightarrow] \rightarrow kpi = 0$ $S_{kc} = \frac{kpi - \min(K)}{\max(K) - \min(K)}$	ECR4, ECR5
Threshold up	$S_{kc} = \frac{kpi}{benchmark}$	M1, M2, M3, M4
Quartile down	$Q1 = \frac{1}{4}(N+1) \mid Q2 = \frac{2}{4}(N+1) \mid Q3 = \frac{3}{4}(N+1)$ $If\ kpi\ value < Q1 \rightarrow kpi = 4$ $If\ Q1 \leq kpi\ value < Q2 \rightarrow kpi = 3$ $If\ Q2 \leq kpi\ value < Q3 \rightarrow kpi = 1$ $If\ Q3 \leq kpi\ value \rightarrow kpi = 0$ $S_{kc} = \frac{kpi - \min(K)}{\max(K) - \min(K)}$	ECR6, W1

Appendix B: Index Area Values Computation in the Spanish CCI.

KPI Type	Area Value Computation Formula	KPI
Binary	$S_{kc} = kpi * 1$	D1, D2, D4, ECR1, ECR2, W3
Percentage	$If\ benchmark > 0 \rightarrow\ S_{kc} = kpi * benchmark$ $If\ benchmark = 0 \rightarrow\ S_{kc} = \frac{kpi - \min(K)}{\max(K) - \min(K)}$	D3
Threshold down	$If\ kpi\ value\ in\ I1 =\ [-1, 0.5 * benchmark] \rightarrow kpi = 4$ $If\ kpi\ value\ in\ \ I2 = [0.5 * benchmark, benchmark] \rightarrow kpi = 3$ $If\ kpi\ value\ in\ I3 = [benchmark, 1.5 * benchmark] \rightarrow kpi = 2$ $If\ kpivalue\ in\ I4 = [1.5 * benchmark, 2 * benchmark] \rightarrow kpi = 1$ $If\ KPI\ value\ in\ I5 = [2 * benchmark, \rightarrow]\ \rightarrow kpi = 0$ $S_{kc} = \frac{kpi - \min(K)}{\max(K) - \min(K)}$	ECR4, ECR5
Threshold up	$S_{kc} = \frac{kpi}{benchmark}$	M1, M2, M3, M4
Quartile up	$Q1 = \frac{1}{4}(N+1) \mid Q2 = \frac{2}{4}(N+1) \mid Q3 = \frac{3}{4}(N+1)$ $If\ kpi\ value < Q1 \rightarrow kpi = 0$ $If\ Q1 \leq kpi\ value < Q2 \rightarrow kpi = 1$ $If\ Q2 \leq kpi\ value < Q3 \rightarrow kpi = 3$ $If\ Q3 \leq kpi\ value \rightarrow kpi = 4$ $S_{kc} = \frac{kpi - \min(K)}{\max(K) - \min(K)}$	W2

Appendix C: Data Computation of KPI Data per Indicator.

Level	KPI	Method	Explanation
Digitalization (D)	D1	Standardization (binary)	Add 1 for all municipalities in PAe, and add 0 for all missing municipalities.
	D2	Standardization (binary)	Add 1 for all municipalities with Cl@ve service, and add 0 for all municipalities without Cl@ve service.
	D4	Standardization (categorization)	Add D4-value to 1/14 to each row. Drop row if CMUN and type of service are duplicated. Group values by CMUN and sum D4-value. Add 0 for all municipalities without Cl@ve service.
Energy, Climate and Resources (ECR)	ECR 1	Standardization (binary)	Add 1 for all municipalities that signed CoM, and add 0 for all missing municipalities.
	ECR 2	Standardization (categorization)	Add values to level of commitment: 2020 = 0.3; 2030 = 0.6; and 2050 = 0.9. Assign value based on highest level of commitment per municipality. Add value 0.1 if also adaption is committed. Add 0 for all missing municipalities.
	ECR 4	Aggregation (Map Algebra)	Overlay GeoTIFF raster file by administrative geospatial dataset with geometry of municipality. Aggregate data by calculation of average value of GeoTIFF raster value per geometry.
	ECR 5	Aggregation (Map Algebra)	Overlay GeoTIFF raster file by administrative geospatial dataset with geometry of municipality. Aggregate data by calculation of average value of GeoTIFF raster value per geometry.
Mobility (M)	M1	Aggregation	Spatial intersect OSMnx network data by administrative geospatial dataset with geometry of municipality. Calculate surface of separate features in square meters. Aggregate data by CMUN and sum feature surfaces per geometry.
		Scaling	Divide aggregated sum of surface by population of municipality, and multiple with 100.
	M2	Aggregation	Spatial intersect OSMnx network data by administrative geospatial dataset with geometry of municipality. Aggregate data by CMUN and sum features per geometry.
		Scaling	Divide aggregated sum of features by population of municipality, and multiple with 1,000.
	M3	Aggregation	Spatial intersect OSMnx network data by administrative geospatial dataset with geometry of municipality. Calculate length of individual features in kilometers. Aggregate data by CMUN and sum feature lengths per geometry.
		Scaling	Divide aggregated sum of lengths by total surface in square kilometers of municipality, and multiple with 100.
	M4	Aggregation	Spatial intersect OSMnx network data by administrative geospatial dataset with geometry of municipality. Aggregate data by CMUN and sum features per geometry.
		Scaling	Divide aggregated sum of features by population of municipality, and multiple with 100.
Waste (W)	W2	Scaling	Divide sum of produced recycled glass waste by population of municipality.
	W3	Standardization (binary)	Add 1 for all municipalities with Punto Limpio, and add 0 for all missing municipalities.

Appendix D: Data Computation of Socio-Demographic and Economic Features.

Variable	Method	Explanation
Population density (per km^2)	Density	Divide population of municipality by surface in square kilometers of municipality.
Natural population growth (in %)	Percentage	Divide natural population growth of municipality by total population of municipality. Multiply by 100.
Population female (in %)	Percentage	Divide female population of municipality by total population of municipality. Multiply by 100.
Population non-Spanish citizen (in %)	Percentage	Subtract percentage population of Spanish citizens from 100.
Total residential buildings (per capita)	Scaling	Sum total of two types of residential buildings per municipality. Divide summed total residential buildings per municipality by total population of municipality.
Unemployment benefits (in % of average salary)	Percentage	Divide average unemployment benefits per capita of municipality by average salary per capita of municipality. Multiply by 100.
Municipal debt (per capita)	Scaling	Divide total outstanding debt of municipality by total population of municipality.
Total companies (per capita)	Scaling	Divide total companies in municipality by total population of municipality.
Total agricultural livestock units (per km^2)	Density	Divide total livestock units in municipality by total surface in square kilometers of municipality.
Total agricultural cattle farms (per km^2).	Density	Divide total cattle farms in municipality by total surface in square kilometers of municipality.
Total tourist houses (per capita)	Scaling	Divide total tourist houses per municipality by total population of municipality.

Appendix E: Overview of (Sets of) KPIs in Spanish CCI.

Level	Sub-level	KPI	Definition
Digitalization (D)	Public digital identity system	D1	Presence in PAe (public digital service platform)
		D2	Adoption of Cl@ve in PA digital properties
	Broadband internet connection	D3	Percentage of people with broadband connection (>30Mb/s)
	Data accessibility	D4	Accessibility of local government digital properties
Energy, Climate and Resources (ECR)	Emission reduction targets	ECR 1	Covenant of Mayors - Subscription
		ECR 2	Covenant of Mayors - Level of commitment
	Air quality	ECR 4	Annual average concentration of PM10
		ECR 5	Annual average concentration of NOx
Mobility (M)	Pedestrian areas	M1	Pedestrian areas (m2/100 inhab.)
	Charging stations for electric vehicles	M2	Charging stations (charging station/1,000 inhab.)
	Cycle lanes	M3	Cycleways (km/100 km2)
	Public transportation availability	M4	Bus stops (bus stops/100 inhab.)
Waste (W)	Recycling of waste	W2	Recycled waste (kg/inhab.)
	Collection of e-waste	W3	Collection of e-waste

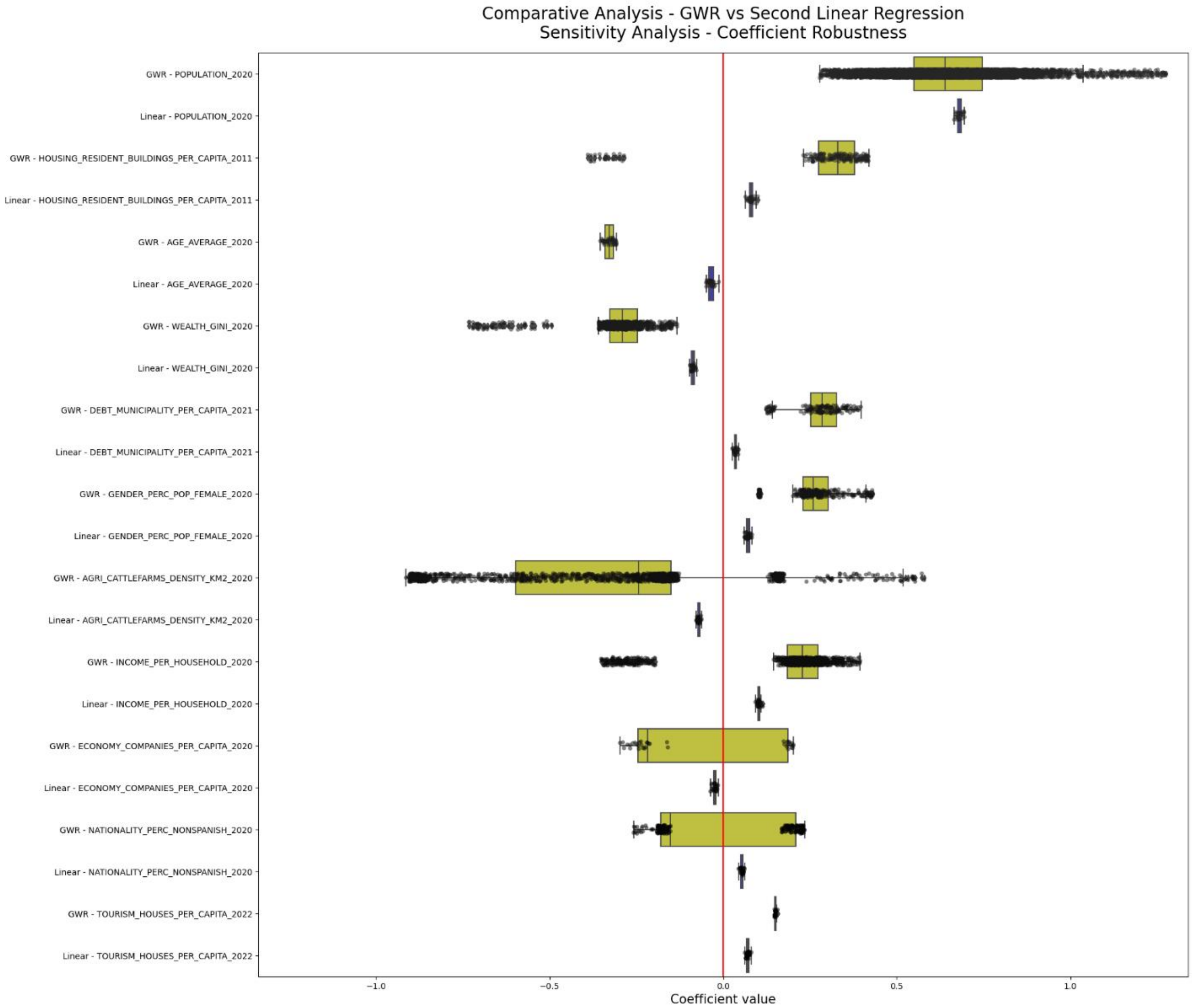
Appendix F: Benchmarks of the Original CCI and Spanish CCI.

Level	KPI	Original CCI Benchmark	Spanish CCI Benchmark	Suitable in Spanish Context
Digitalization (D)	D1	Yes	Yes	Yes
	D2	Yes	Yes	Yes
	D3	High	High	Yes
	D4	High	High	Yes
Energy, Climate and Resources (ECR)	ECR 1	Yes	Yes	Yes
	ECR 2	2020-30	2020-30	Yes
	ECR 4	40µg/m3	40µg/m3	No
	ECR 5	40µg/m3	40µg/m3	No
Mobility (M)	M1	900 m^2 / 100 inhabitants	900 m^2 / 100 inhabitants	No
	M2	1 / 1,000 inhabitants	1 / 1,000 inhabitants	No
	M3	100 km / 100 km^2	100 km / 100 km^2	Yes
	M4	1 / 100 inhabitants	1 / 100 inhabitants	Yes
Waste (W)	W2	65%	-	No
	W3	Yes	Yes	Yes

Appendix G: Correlation Matrix of Socio-Demographic and Economic Features.

Type of Sub-Characteristic	**Feature**	**Correlation with CCI Score (Pearson's R)**	***Significance (p-value)***	***Sample size (N)***
Population	Population	0.1756	0.000	8,087
	Population density (per km^2)	0.1570	0.000	8,087
	Natural population growth (in %)	0.3032	0.000	8,087
Gender	Population female (in %)	0.4147	0.000	8,087
Age	Average age of population	- 0.5239	0.000	8,087
	Population below 18 years (in %)	0.5128	0.000	8,087
	Population above 65 years (in %)	- 0.4724	0.000	8,087
Nationality	Population non-Spanish citizen (in %)	0.2556	0.000	8,087
Housing	Average household size	0.4263	0.000	8,087
	Single persons households (in %)	- 0.4590	0.000	8,087
	Total residential buildings (per capita)	- 0.3814	0.000	8,087
Income	Income (per household)	0.2053	0.000	8,079
Wealth distribution	GINI index	0.3427	0.000	8,087
Public finance	Municipal debt (per capita)	0.1127	0.000	8,087
Economy	Total companies (per capita)	0.0440	0.000	8,087
Agriculture	Total agricultural livestock units (per km^2)	- 0.0140	0.211	8,087
	Total agricultural cattle farms (per km^2)	- 0.0137	0.219	8,087
Tourism	Total tourist houses (per capita)	0.0490	0.000	8,087

Appendix H: Sensitivity Analysis of GWR Model and Second Multiple Linear Regression Model



Appendix I: GWR Model Output: Significant Estimated Coefficients per Municipality. Other Features.