Decision Making Tool for Smart Cities

Orestis Trasanidis

Main academic Supervisor:	Associate Professor, Dr. Christos, TJORTJIS, IHU.
Partner supervisor:	Dr, Johanna, TZANIDAKI, ERTICO ITS Europe

A Master Thesis submitted for the Erasmus Mundus Joint Master Degree on Smart Cities and Communities (SMACCs)

June 2021

University of Mons, Heriot Watt University, International Hellenic University, University of the Basque Country

Acknowledgements

The researcher wishes to acknowledge the support given to him by expressing his warm appreciation and sincere gratitude to the persons who were involved in helping him overcome the challenges of this study.

My Family John, Antigone, Helen and Lazarus for being extremely supportive throughout my SMACCs studies by always believing in me.

Dr. Johanna Tzanidaki, ERTICO ITS Europe, Partner Supervisor -who has been a tremendous mentor in every sense of the word.

Prof. Christos Tjortjis, IHU, Academic Supervisor- for sharing his recommendations and knowledge and for helping me to make this work better.

Dr. Ioannis Magnisalis European Commission DIGIT, Advisor- for supporting my ideas and their implementation allowing me to grow and develop as a researcher.

My colleagues at ERTICO ITS Europe and especially Zeljko, Vlad, Lydia and Cassandre with whom I spent a lot of hours working and discussing on the City Moonshot survey and on Urban Mobility plans.

Dr. Sesil Koutra for her advice and suggestions regarding the theoretical foundations of smart cities.

Coordinators, Professors, Lecturers, Administration Staff and students of SMACCs and especially Associate Professot Jon Teréz, Dr. Benoit Couraud, Mrs Ruken Karakus and, Mrs Rallou Taratori for offering me the opportunity to have fruitful and inspiring discussions with them during this programme. A special thanks to Prof. Iñaki Gomez for helping me during my COVID-19 quarantine in Bilbao.

Last but not least my best friends Fotis, George, Nikos, Thodoris, Dimitra, Alexandros, Stella Stefanos, Zef, and others who were a constant source of happiness and courage during times of research and solitude.

Abstract

Smart cities and their link to sustainability especially in the mobility sector has become a major concern on a global scale. Nonetheless, the challenges of this link must be first dealt at the local scale. Urbanization is rapidly transforming the cities and creates new urban behaviors pushing for a holistic assessment of cities. Furthermore, the complexity of the resulting urban problems necessitates an evidence based decision and policy making approach. This study aspires to propose a methodological framework for enabling an exploratory analysis of the smart mobility characteristics of European cities by utilizing the capabilities of ICT tools in tackling the obstacles of large scale assessment. This work involves the combination of a qualitative survey (questionnaire) and the creation of new quantitative datasets in order to cluster cities according to a proposed system of indicators. This will serve as a basis for developing an interactive tool that will dynamically visualize the correlations between sustainability assessment indicators, cities and clusters allowing the identification of key influencers, city profiles and characteristics. The methodology is tested on 57 European cities and incorporates the use of composite indicators, data mining techniques and state-of-the-art analysis of the theoretical background and a detailed review of relevant past studies.

Keywords: Decision-making, Mobility, Indicators, Smart cities, dashboard, Machine Learning, city intelligence, sustainability, clustering.

Student Name: Orestis Trasanidis

Date: 21/06/2021

Table of Contents

List of Tables

List of Figures

Nomenclature

AI	Artificial intelligence
BEVs	Battery Evs
BI	Business Intelligence
C-ITS	Cooperative Intelligent Transport Systems
CAV	Connected Automated Vehicle
CCAM	Cooperative, connected and automated mobility
CCAV	Cooperative, Connected and Automated Vehicle
CO2	Carbon Dioxide
CO2e	Carbon Dioxide equivalent
CoM	Covenant of Mayors
DDT	Dynamic Driving Task
EAFO	European Alternative Fuels Observatory
EC	European Commission
EEA	European Environment Agency
EFTA	European Free Trade Association
ELTIS	European Local Transport Information Service
EU	European Union
EVs	Electric Vehicles
FP7	Seventh Framework Programme
GDP	Gross Domestic Product
GHG	Greenhouse Gases
H2020	Horizon 2020
HEVs	Hybrid EVs
ICT	Information and Communication Technology
IT	Information Technology
ITS	Intelligent Transport Systems
JRC	Joint Research Centre
MaaS	Mobility as a Service
ML	Machine Learning
MoD	Mobility on Demand
NO2	Nitrogen dioxide
NUTS	Nomenclature of territorial units for statistics
OECD	Organisation for Economic Cooperation and Development
PHEV	Plug-in HVEs

PM10	Particulate Matter 10
R&D	Research and Development
RES	Renewable Energy Sources
RIS	Regional Innovation Scoreboard
RT	Research Tasks
SMEs	Small Medium Enterprises
SUMP	Sustainable Urban Mobility Plan
TRIMIS	Transport Research and Innovation Monitoring and Information System
UK	United Kingdom
UVARs	Urban Vehicle Access Regulation Schemes
WHO	World Health Organization

1 Introduction

Plato in 360 BC gave us the first insight: for the development of human communities (cities) in Laws V, the land, must be sufficient to support no more than a number of people living in moderation. Nowadays this concept of a city that should be able to sustain itself and its population has been redefined and is more relevant than ever. According to the OECD (OECD, 2020), by 2050 the world population living in cities will increase by 2 billion. As evidence of the rapid urbanization, half of the world's cities did not exist four decades ago. However, the projected overpopulation, is not the only source of concern, as it is combined with overconsumption. The date of 29 July 2019 was set as the Earth's Overshoot Day (Global Footprint Network, 2020), which indicated that the world had spent the natural resources that the earth could renew within a year. The day raised global awareness that at the rate the world currently consumes, we will soon need more than three planets to meet our needs while during 6 days 1 million new city dwellers (as much as Thessaloniki's population for example) will be born.

As a result, of the anticipated adverse effects of excessive consumption of resources and unsustainable population growth the concept of smart cities emerged in the last two decades proposing that new technologies will improve urban efficiency and hence enhance overall urban sustainability. In that context, the United Nations (UNECE, 2011) identified mobility and access to it, as high priorities for improving the environment and quality of life at the urban level. Even though, there is currently among the population a general awareness of the positive impact of smartness on cities, as well as a wide understanding of the problems caused by the absence of smartness, there is lack of awareness with regards to the nature of this smartness, "we will know it when we see it?" (Beatley & Manning, 1997). This notion is based on the premise "you can't manage something that you can't understand, evaluate and see". Hence, decision makers in urban policy departments around the world encounter the difficulty of addressing urban challenges while they do not possess the appropriate knowledge and understanding of what is needed, how to measure smartness and what it entails. For that purpose, the use of assessment indicators on city smartness has rapidly spread in academia aiming to prevent the negative consequences of this phenomenon by providing decision-makers and analysists the ability to track and compare concrete policy objectives.

The aim of this study is to utilize existing fragmented data by deploying an aggregation system that uses indicators in order to create an interactive tool for exploring the data that can support decision making. It is within this framework that this research objective attempts to identify the nature of smartness. By proposing a new system of indicators for accessing and characterizing smart mobility in European Cities, the thesis aspires to introduce a new, updated approach towards solving the problem of data availability and data visualization by the use of Information and Communications Technology (ICT) tools. Ultimately, this study attempts to support the creation of a European framework for analyzing cities' urban mobility behavior so as to enable citizen, and data-driven decision making in cities with regards to smartness. This study allows cities to extract valuable knowledge on some of their important urban trends in order to take effective decisions, and remedial actions. Finally, the results of the research presented here are expected to contribute towards raising awareness on city intelligence in terms of challenges and opportunities for an evidence based decision making.

The contributions of this thesis are:

- A conceptual theoretical framework of smartness, and sustainability along with a summarization of future transport technologies.
- A deeper understating of urban data availability, openness and indicator analysis for city characterization.
- A methodological framework for assessing the smartness of mobility of European cities.
- An investigation of a new technique for estimating missing data based on data mining in the smart cities context
- A process for enabling interactive exploratory analysis of urban data towards contributing to city intelligence.

The subsequent two chapters, are dedicated to literature review of similar theoretical and empirical works. In these chapters, the concepts of *smartness, sustainability, intelligent transport* and *data mining* will be thoroughly discussed in order to set the foundational framework for the chapters that follow. The fourth chapter explains the case study, and the methodology of fieldwork together with a review of the state of the art related studies. In the fifth chapter, the main results and insights stemming from this research are summarized. Finally, the conclusion and discussion are presented in the sixth chapter together with recommendations, and reference is also made on the practical limitations of this research. Recommendations for next steps close the work for this Thesis.

2 Smart Cities

This chapter thoroughly investigates the definition, and the conceptual interconnections of smartness with sustainability and governance for cities. The theoretical foundations of this review are driving essential research decisions in this study. Moreover, this chapter discuss the most impactful concepts and factors for the future of intelligent transport systems. This analysis is the theoretical background of the case study that it is explained in the next chapters. Overall this chapter is the intellectual cornerstone upon which the assessment of cities was built.

2.1 Theoretical foundations

2.1.1 Smartness & Sustainability

The broad concept of *sustainable development* gives rise to multiple interpretations and definitions (Tanguay, et al., 2010). Nevertheless, the most popular definition for sustainable development was provided by the Brundtland Report: "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" (WCED, 1987). Sustainability, was implied as a notion of a none declining utility and suggests a new organization of societies based on the relation between human and nature, and by emphasizing the intergenerational responsibility for equity (Baumgärtner & Quaas , 2010). Therefore, it is shaped as equilibrium of environmental quality, economic prosperity, and social justice or 'the triple bottom line' (Elkington, 1997). In that context, urban sustainability is extremely challenging for cities because the economic and social benefits resulting from the agglomeration effect are resulting in the degradation of the natural recourses almost equally. Consequently, environmental urban challenges are giving a push for smart solutions which are rapidly gaining a rising attention as they promise an urban utopia (Datta, 2015).This alleged convergence of urban sustainability with 'smartness' provides a starting point for further investigating the concept of the latter.

Recent advances in technology have enabled innovative digital scenarios that provide citizens, and communities with cohesive and tailored solutions for their urban life. At the same time, technological achievements have improved the methods and tools towards city management from the standpoint of different urban stakeholders. In this direction, an innovative vision of a 'clever and integrated' city has emerged under the name of "smart city". Zuccalà and Verga (Zuccalà & SergioVerga, 2017) introduce the "smart city" as a sustainable area, where every aspect is supported by ICT efficiently taking Plato's insight into the digital era of the second millennium.

Figure 1:Triple bottom line

The word "smart" is treated in the literature as an adjective or a normative concept (Höjer & Wangel, 2014). Its definition as an "adjective" has several meanings in dictionary (Oxford University, 1998), such as "mentally alert", 'very good at learning or thinking, "intelligent", "knowledgeable", etc. applied for "persons, objects, places, etc"., while as an instrumental concept it implies the creation of "products, services, infrastructure, etc.", in which ICT play an important role (Höjer & Wangel, 2014). On the other hand, a "smart place", a city/district/building, is usually described as being able to manage its resources 'smartly' basically connected on the technologies (Bessis & Dobre, 2014). Among the various definitions given for "smart city" concept, the most relevant reported in the literature review are provided at Table 1, (Koutra, et al., 2019).

Table 1: 'Smart City' Definitions Most Cited in the Literature (Koutra, et al., 2019)

Authors	Definition	Citations
Caragliu, Del Bo & Nijkamp- 2011	"We believe a city to be smart when investments in human and social capital and traditional (transport) and modern communication (ICT) infrastructure fuel sustainable economic growth and a high quality of life with a wise management of natural resources through participatory governance".	358
Komninos- 2011	"The Smart Cities concept (…) is connected to notions of global competitiveness, sustainability, empowerment and quality of life, enabled by broadband networks and modern ICTs".	291
Giffinger, Fertner, Milanovic & Meijers,-2007	"A Smart City is a city well-performing in a forward looking way in six characteristics built on 'smart' combination of endowments and activities of self-decisive, independent and aware citizens".	207
Nam & Pardo-2011	"Smart City integrates technologies, systems, infrastructures, services and capabilities into an organic network that is sufficiently complex for unexpected emergent properties to develop".	103
Batty - 2013	"A Smart City is a city in which ICT is merged with traditional infrastructure, coordinated and integrated using new digital technologies. Smart cities are also instrumenting for improving competitiveness in such a way that community and the QoL are enhanced".	87

The origin of “smart cities” is founded in the concept of “smart growth” (Bollier, 1998) during the ‘90s, which advocated new policies for urban planning and was driven, until recently, by a number of big corporates in an attempt to apply complex information systems to the integration of infrastructure and services; since then it has been evolved to mean any form of technologically based innovation in the planning and city development. Nonetheless, the idea of “smart in cities” is not completely new. Already in the mid-1800s, researchers (Eger, 2009) focused their works in efficient and self-governed cities of the American West. Gabrys (Gabrys, 2014) stated the idea in ‘60s under what is called the “cybernetically planned cities”. Historically, the “smart cities” are a gradual evolution of digital, ubiquitous, and intelligent cities towards the manipulation of information in order to create wisdom. The underlying concept is to use knowledge to improve our cities, something that is made possible by ICT as mentioned above. Today it is agreed that the wisdom that is created should aim is to make cities liveable, enable innovation, and promote entrepreneurship and finally improve the quality of life. To achieve this target, we must be able to think through systems that manage, and operate the infrastructure and the information that is key to connect these systems to citizens and other elements of urban life. This information is acquired by developing some urban capabilities so as to gain valuable knowledge, and then transfer it back to processing (Khan, et al., 2015).

Clearly there is an interconnectivity between “smartness” and “sustainability”, especially when it becomes apparent that cities and communities are widely understood as socio-technical systems and therefore treated as such. In one of the most cited references in the literature of “smart cities” (Komninos, 2011), the concept of smart cities is connected with the concept of sustainability, stating that this is made possible through broadband networks and modern ICTs. This is also mentioned in Toppeta (Toppeta, 2010) who claims that a city that incorporates ICT tools in its strategy, can ameliorate sustainability, and liveability. Barrionuevo et al., 2012 delve a bit more into this, highlighting that in order for this to be achieved, the incorporation must be done “in an intelligent and coordinated manner”. In accordance with the above, in the more recent literature, smart cities are most commonly defined as cities that exploit information technologies in order to improve their sustainability, and well-being (Boob, 2015).

2.1.2 Smartness & Governance

Innovative governments and public organizations are implementing “smart " solutions in order to tackle the demands of an increasingly urban population, utilizing information, and communications technology tools (ICT) (Belissent, 2010). Smart cities are rising as the only actors, which can effectively manage the natural resources, and support citizens to change their behaviour, enabling in that sense sustainable development. Horbaty (Horbaty , 2014) defines “smart cities” from their ability to give citizens a good quality of life through successful resource management and the appropriate use of ICT tools. The idea behind this definition is to create sustainable cities that address natural resource problems, have an efficient infrastructure, are interconnected and are comfortable, attractive and secure to live in (Lazaroiu & Roscia, 2012). Connectivity of the infrastructure is an important factor to the smart cities solution. IoT systems are believed to be fundamental for the future development of large-scale heterogeneous infrastructures, connecting various ICT tools, development platforms and apps that help citizens' wellbeing and enable the use of maintenance, sustainability, technical, social and economic key performance indicators (KPIs) by the decision makers (Mehmood, et al., 2017). ICT can provide integrated information and intelligence for the purpose of better urban management and governance, sustainable socio-economic growth and policy development using participatory processes (Schaffers, 2012).

In the literature that was analysed for the purposes of this Thesis on smart cities, three different types of ideal-typical definitions were found: smart cities as cities using smart technologies (technological focus), smart cities as cities with smart people (human resource focus) and smart cities as cities with smart collaboration (governance focus) (Meijer, 2016). A prominent and sophisticated definition for "smart city" has been developed by Caragliu et al. (Caragliu, 2011): *"We believe a city to be smart when investments in human and social capital and traditional (transport) and modern (ICT) communication infrastructure fuel sustainable economic growth and a high quality of life, with a wise management of natural resources, through participatory governance."*

The interest in "smart cities" is growing fast (Komninos & Mora, 2018) as a way to facilitate and improve citizens' life by integrating information and communication technologies. In a comprehensive study, Lee et al. (Lee, 2014) propose a reflection of six dimensions to analyse the "smart city" concept: (1) urban openness; (2) service innovation; (3) partnership and collaboration; (4) proactiveness; (5) infrastructure integration and (6) governance, which is considered as a key driving force to enable the smart city development. Thus, Lee et al. and Caragliou offered an interesting perspective to the matter, adding participatory governance to the equation for smart and sustainable development. More specifically, they stretch the importance of educating citizens in the long term in order to achieve public participation on governance. Marijn Janssen in 2015 (Janssen , et al., 2015) states that "a smart city only becomes smart when there are smart citizens, businesses, civil servants and other stakeholders". In that sense, smart cities and smart citizens are interdependent concepts while informed and engaged citizens, who exploit their knowledge in order to reduce their consumption, are the path to fully implementing the smart cities idea. The expectation of the citizens for an improved form of governance based on the rapid development of new smart technologies has pushed city authorities to redefine governmental structures (Al-Khouri, 2015) (Sankowska, 2018).

According to Belissent (Belissent, 2010), governance is the core of smart city initiatives. The challenge is great because the adaptation of digitalization is not only a technological and economical incorporation but rather a new process to communicate which involves new knowledge holders and stakeholders (German Federal Institute for Research on Building, Urban Affairs and Spatial Development,, 2017). Barrionuevo et al. (Barrionuevo, et al., 2012) delve a bit more into this, highlighting that in order for this to be achieved, the incorporation must be done "in an intelligent and coordinated manner". In other words, the interaction between stakeholders who actively share data through a cooperation process is enabled by ICT leading to a new urban paradigm (Sankowska, 2018). Therefore, data-driven concepts like e-Governance and open- Government become broadly adopted, as they introduce a more transparent and citizen-friendly way. Bernardo identified that two of the fundamental factors of data-driven Governance are e- participation and e-consultation which are bringing smart initiatives to citizens (Bernardo, 2017). Participatory governance and citizen involvement are key concepts in many smart city action plans, strategies and frameworks (Castelnovo, et al., 2016). In that sense, "participatory governance" is defined as the engagement of all stakeholders in decision-making by utilizing cutting-edge ICT to provide unique opportunities for citizen participation in city management actions, offering ultimately effective governance, combined with enhanced legitimacy and justice. According to Arnstein (1969), the understanding of the public participation as power is the key to improved future cities. As he thoroughly explained, citizen participation has multiple levels and the evolution of participation from informing-citizens' awareness into consultation and later partnership is a critical pathway towards democratization and sustainability. Furthermore, in order to take advantage of its full potential the focus of governance must be beyond Arnstein's Ladder on enabling social learning (Collins & Ison, 2006). This notion highlights the complementary nature of data-driven governance, co creation and participatory governance. Together they secure a transparent process of decision-making and enable better citizen participation in implementing, monitoring, and evaluating smart initiatives (Albino, et al., 2015).

Smartness leads the development of governance but only that of processes, their activator is the collective decision- making that includes both public and private actors. In that context, the necessity for a thorough sustainability & smartness assessment of cities has risen, in order to provide the stakeholder with valuable data and information. This information should be open and connected to the people so as to engage the public (Janssen , et al., 2015). Therefore, the sustainability & smartness assessment for cities has emerged to fill the urban information gap since an "issue that is not clearly measured is also difficult to improve" (Bohringer & Jochem, 2017) by helping decision makers to have a holistic view of their cities towards more effective and smart solutions. This idea inspired many researchers to develop several characterization systems that would be further described in section 4.2.2. Apparently, the absence of clear understanding of smartness and sustainability in urban context is causing major difficulties of addressing problems such as greenhouse gases and congestion and hinders the integration of planning tools (Garau, et al., 2016) which otherwise are fragmented and therefore not smart.

2.2 Intelligent Transport Systems

2.2.1 Future mobility concepts

In 2007 Giffinger et al. (Giffinger, et al., 2007) described "smart cities" using six dimensions as the following figure illustrates. Clearly mobility is one of the fundamental concepts towards future smart cities and the activator which integrates these dimensions is the Information and Communication Technologies. Mobility is at the heart of human existence and considerable resources have been put into shaping the transport of the future. Mobility can be simply described as "the ability to move or be moved freely and easily" (Ertico ITS Europe, 2018) but this conceals some much wider implications. When mobility services are linked to IT systems, telecommunication networks and a wide range of sensors, it is transformed into a smart mobility. In that context, the rapid concentration of people in dense areas has resulted in some impactful urban problems, such as public transport inefficiencies, inadequate mobility services, and more air and noise pollution. The transport systems in cities the last decades were highly depended on cars, causing productivity losses from congestion equal to 1-2 % of the EU's GDP, approximately 850 million CO2 emissions and higher housing expenses as a result of the increasing commuting time.

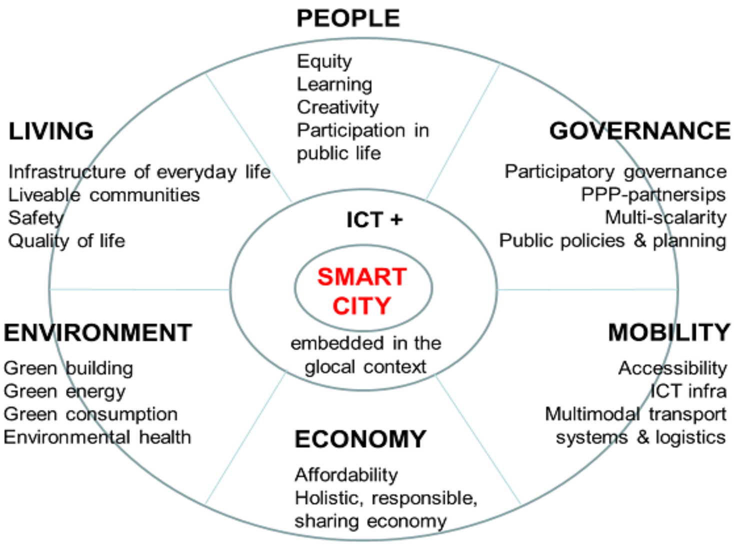


Figure 2: Six dimensions for the "smart cities" proposed by (Giffinger, et al., 2007)

Therefore, the implementation of the latest technologies and ICT can promote environmentally friendly and sustainable new transport schemes improving the overall efficiency of the systems and so the citizens' quality of life (Caragliu, 2011). For that reason, the European Commission

(EC) in 2019 published an insightful report "The future of road transport" to explore the "intelligent" revolution in mobility and identify the technological and business drivers that rapidly transform the cities. Artificial Intelligence (AI), decarbonisation, automation and digitalization are major disruptive forces promising an inclusive, safe, efficient and viable future for cities and citizens alike. Factors such as city governance processes, infrastructure requirements, data governance, ICT and legislation/policies are either contributing or hindering the mobility evolution of the urban societies. Hence, societal implications like economic growth, unemployment, skills, citizen's behaviour/decision making and urban planning have a crucial role for society's readiness and capacity to incubate and deploy new mobility paradigms. However, the development of the cities is not perceived as the deployment of new ICT concepts alone but rather the improvement of the existing transport systems and policies based on ICT. The intelligent transport systems that are currently deployed in urban mobility networks are representing just a fraction of what is proposed by the scientific community in terms of city development. Three rapidly developing mobility concepts have the potential to drastically reshape urban transport. Moreover, the possible combination of these three mobility concepts and their integration strengthens their individual impact and paves the way for the new future mobility schemes.

Connectivity: Connectivity is usually referred to as Cooperative Intelligent Transport Systems (C-ITS) and conceptualizes the communication between vehicles and roadside infrastructure. Thus, it is strictly correlated with the automation resulting the Cooperative Connected and Automated Vehicles (CCAV), "Connectivity, Cooperation and Automation are complementary technologies that reinforce each other and will over time merge completely". Automation, meaning the systems that are capable to "perform part or all of the Dynamic Driving Task (DDT) excluding the strategic functions" (SAE International, 2016) is actually an equally important concept with connectivity, but the implications with the abandonment of car ownership, the driving as experience, the user acceptance and technology challenges are resulting in lower maturity levels of adaptation. Specifically, according with the Gartner Hype Cycle the Autonomous Driving has just left the peak of inflated expectations (Gartner, 2019) and has more than a decade to reach the plateau of productivity.

Decarbonisation: In order to reduce the greenhouse gas emissions several experts propose the use of alternatives fuels aiming to reduce the dependence on fossil fuels. Among biofuels (bioethanol-biodiesel)], natural gas, hydrogen and electrification the latter is widely accepted as the currently most viable strategy to reduce the environmental impact of urban mobility. The market penetration of Electric Vehicles (EVs), Hybrid EVs (HEVs), EVs which include batteries (BEVs) and plug-in HEVs (PHEVs) is increasing mainly due to reduction of their cost. Generally, three elements are essential for the deployment of electrification: the increase of the usage, higher durability of the technology and higher accessibility to charging infrastructure.

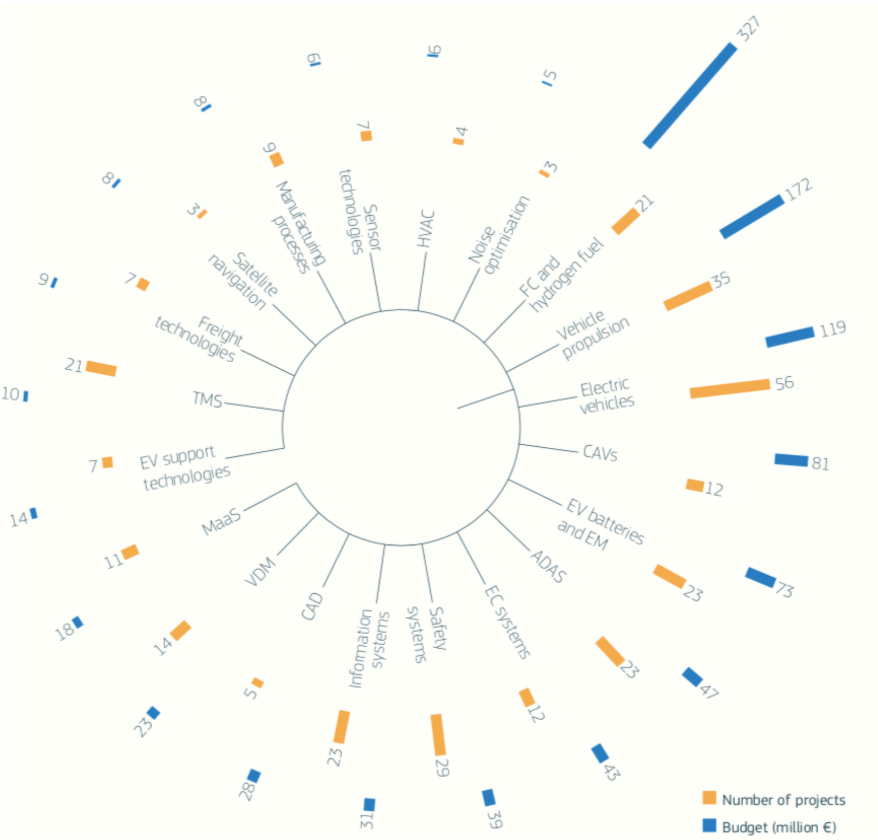
Sharing: In one of the most cited papers on the subject, shared mobility was defined as “the shared use of a vehicle, bicycle, or other low-speed mode that enables users to have short-term access to mobility modes on an “as-needed” basis” (Shaheen & Chan, 2016). The individual mobility modes will need to be effectively interconnected physically and digitally at the appropriate nodes to integrate them into a single mobility service. The concept of Mobility as a Service (also known as MaaS) involves several stakeholders and is realized through new business models based on digital economy such as car-sharing, bike-scouter sharing, fractional ownership and other flexible transit services.

2.2.2 Key factors towards ITS

Coordination and management in transport are critical factors for maximising the efficiency of transport systems. CAVs for example are focusing on user comfort, safety and reducing the individual travel cost rather than smoothing congestion and increasing the network capacity, unless the public authorities or regulators request otherwise (Makridis, et al., 2018). Urban Vehicle Access Regulation Schemes (UVARs) together with vehicle cooperation call for both centralized and decentralized strategies to keep the balance between freedom and traffic improvement. Hence, new mobility governance is needed to secure that the increasing capabilities of the new technologies won’t lead to a future mobility based solely on vehicles because that will contribute to inefficiency and inequality (Alonso Raposo, et al., 2019). Cycling and walking (micro mobility) are vital mobility alternatives as they increase accessibility, promote better health conditions and create an attractive urban environment (Stevenson, et al., 2016). Hence, the goal of modern transport policies is to align shared mobility operators, accessible infrastructure, public transport with new emerging mobility concepts in order to reduce the “price of anarchy”, the losses from mobility inefficiencies that may be caused by lack of coordination.

The need for coordination and integration was the cause of development of digital mobility service platforms during the last decade. Users nowadays are able to easily pick among various options the service that matches their needs and create flexible mobility combinations revolutionizing the typical transport process. Thus, these platforms can activate a dynamic management of supply and demand and boost the efficiency of the mobility networks. Mobility service platforms are already generating data for personal destinations, utilising personal information, parking capacity, delays, mechanical car etc. These datasets are crucial for the integration of the mobility systems also for a wide variety of other services provided within the urban settings such as accommodation and urban planning, business and services and others. Therefore, the access to this data is valuable and it also involves concerns about the ownership of the data and the privileged position of some stakeholders over others in terms of having access to it. In that sense, the competition between mobility service platforms and the extended vehicle concept (the car manufactures’ privileged position on data-accessibility) might create “walled gardens” (Alonso Raposo, et al., 2019) able to hinder the exploitation of connectivity. Thus, a new mobility governance involving the public sector should promote the coordination of the mobility services to use the market power in the most beneficiary way for the citizens to use mobility options to their advantage. Finally, data sharing becomes more and more important so that standardization and the interoperability standards like DATEX II for road transport are the key for ensuring the quality of these operations.

In this evolving mobility dialogue, the European Commission (EC) is investing in the Research and Innovation (R&I) projects that involve transport technologies in order to trigger the private sector, more than any other major global economy does (Pasimeni, et al., 2018). More specifically, the EC funded over 300 Horizon 2020 projects involving more than 1000 unique organizations covering various mobility fields, as it is shown in the following figure.



Note: FC = fuel cells; EV = electric vehicles; ADAS = advanced driver assistance systems; VDM = vehicle design and manufacturing; TMS = traffic management systems; EM = energy management; EC = emission control; MaaS = mobility as a service; HVAC = heating, ventilation and air-conditioning; CAD = computer aided design

Figure 3: Mobility technologies funded under Horizon 2020

The results of these projects are actually reshaping urban mobility by introducing new innovative ideas, which will have varying impacts on economy and the job market demanding new skills. It is important to note that the necessity for ICT reskilling and upskilling of the automotive specialists has already been identified as a major priority for the mobility market (European Commission, 2016).

Furthermore, the EC has been supporting the adaptation of two fundamental action plans for urban mobility to prepare the take-up of future mobility, the Sustainable Urban Mobility Plan (SUMP) and the Covenant of Mayors (CoM), which have the central goal of introducing a bottom up planning of a clear long-term implementation plan and of promoting sustainable solutions in an ever-changing urban climate (ELTIS, 2020). Also, both are providing city authorities and decision makers a high-quality framework for understanding the current performance of the urban systems in terms of energy, climate and mobility. However, a holistic assessment of the mobility smartness is an essential element of the urban metabolism (Lopez-CarreiroI. & Monzon, 2010).

Overall, urban plans including governance together with the promotion of innovation, active mobility infrastructure and accessible public transport are the key factors that bring cities close to future mobility (Vandecasteele, et al., 2019).

3 Information and Communication Technologies

This chapter is highlighting the core technologies of smart cities focusing on the management and dissemination of data, information and knowledge which are seen as the basis for better governance. This theoretical framework is the structure that support the methodology of this study.

3.1 ICT & Governance

In the previous Chapters, it was highlighted that the main enabler for smart cities and for the related sustainability in modern societies is the management of data, information and knowledge which is seen in the literature as serving to be the basis for better governance. Nowadays people, private and public sector are endlessly producing and systemizing data as prosumers (consumers involved in designing or customising products for their own needs) mainly due to new innovation-driven applications and e –government services (Bauer, et al., 2006). The explosive growth of data generation has caused an information glut, as the stored data alone does not create knowledge that can improve decision making services or help the development of sustainable projects (Sumathi & Sivanandam, 2006). As a result, the growing need for, collection and utilization of urban data has introduced complexity in information management and knowledge discovery. Transparency, accountability, seamless services and integration are issues that must be taken into account when transforming conventional governance to address those trends (Brown, 2007). Indeed, these issues generate fundamental challenges to the conventional structure of processes and governmental internal and external collaborations. Strategy design, funding, implementation and performance management necessitate evidence-based decision making and policy by adapting a data mining approach in order to improve efficiency (Heinrich, 2007).

Data mining is considered to be the process of automatically extracting relevant or evident information through manipulating data from large databases (Tan, et al., 2014). Of course, there are many Data mining definitions but the most widely accepted is the one introduced by J.Leskovec, A.Rajaman and D. Ulman in 2012 (Leskovec, et al., 2012)“ Data Mining: This term refers to the process of extracting useful models of data.”. Ultimately, as Han Kamber (Han & Kamber, 2000) highlighted, data mining combines the computers’ high performance computing with human’s ability to detect partners. Therefore, data mining has the capacity to enable prospective analysis beyond the assessment of the past. It can thus be deduced that data mining allows the prediction of future trends and behaviours offering evidence steered leadership for public administration (Stenvall, et al., 2007) by extracting hidden or unknown valid information from large information and data sets. Undoubtedly, data mining as a sophisticated and analytical process is a powerful tool which follows a multidisciplinary approach drawing works from database technology, high-performance computing, data visualization information theory relying primarily on statistics, Artificial Intelligence and Machine Learning (ML) (Zhou, 2003).

However, even though data mining is saving time and adds new organizational capabilities it is as beneficial as the performance of the evidence-based information management (Heinrich, 2007). Veenstra 2017 (Veenstra & Kotterink, 2017) highlights that policy-makers rarely have the luxury to use homogeneous data sets in order to measure the performance of their decisions. In the public administration domain, services are driven by information which is originated from various sources. Under high-stakes pressure by decision-makers to show improvements and in the absence of high quality information management systems local governments sometimes are misusing and manipulating data in order to have readily available information on possible outcomes. In addition, ICT and e-government management is continuously challenged by a rapidly changing environment with quite high expectations for high quality services (Bannister, 2005), especially from citizens. Moreover, the evolution of ICT itself demands the quick transformation of conventional governance and administration structure, towards a unique direction for each city, pushing for an innovative redesign of their procedures. As a result, it has been stated that this disruptive nature of ICT affects the levels of peoples' work satisfaction, and that must be taken under consideration by high management (Golden & Veiga, 2005). For that reason and for other ones more operational oriented, cities are increasingly investing in ICT in order to address complexities caused by ambiguous interorganizational networks. Overall, ICT is the tool for managing information and through that to explicit mine data in a specific context. On the other hand, information management systems influence the public administration itself and actually they are e-government systems because they are responsible for the production and delivery of services.

In conclusion, ICT management and data mining are key elements for modern public administration. Both are offering support to decision makers and tools for information coordination. In extent, ICT-driven governance allows a bottom-up community planning towards a more efficient policy implementation even in the small scale (Misuraca, et al., 2010). In addition, Data mining was characterized by Piatetsky-Shapiro during a workshop in 1989 as "knowledge discovery in databases" so the linkage to traditional decision making, and managerial or informational tasks is clear (Fayyad, et al., 1996). The socio-technical results from applying the above are redefining smart city's governance and planning and are thus transforming their fundamental principles.

3.2 Data mining

Data mining is a relatively new technology, initially, it emerged in the late 80' and it rapidly evolved during the next decades with great success in building business intelligence. In this competitive world of accelerating change, the potential to help companies focus on the most important costumer behaviour as knowledgeable observers caused an explosive growth in the development of apps. In the short term, it offers a leverage point to initiate efficient and impactful decision making one step ahead of competition for many industries. To elaborate, in the Mobility sector data mining is usually used for predicting traffic congestion (Mystakidis & Tjortjis, 2020), also it has many applications in the Energy (Christantonis, et al., 2020) and Climate (Avramidou & Tjortjis, 2021) sectors for estimating future needs and performance but it is equally vital in many other sectors like Bioscience, and e-Commerce.

Data mining as was explained in the previous part of this chapter, is the computational process of discovering novel and potentially useful patterns from large amounts of data stored in various data repositories and the use algorithms to harvest hidden meaningful information. In order to achieve those data mining many up-to-date technologies such as statistics, big data, neural networks, machine learning and evolutionary algorithm by exploiting techniques and functions are involved, such as:

Classification: This method distributes objects of a dataset to specific predetermined categories -tuple. It has a phase of supervised learning where it assigns a category label to a set of unclassified instances which are the given training data. The output of classification is a discrete class for the given objects and their attributes. Based on the relationship between the input and output the deduced model can predict the class of new objects. Therefore, the model can identify the patterns that define each category and so the description of the classes. Several classification techniques exist such as: naïve Bayes classifier (as seen in Figure 4, below), decision trees, neural networks, k-nearest neighbour.

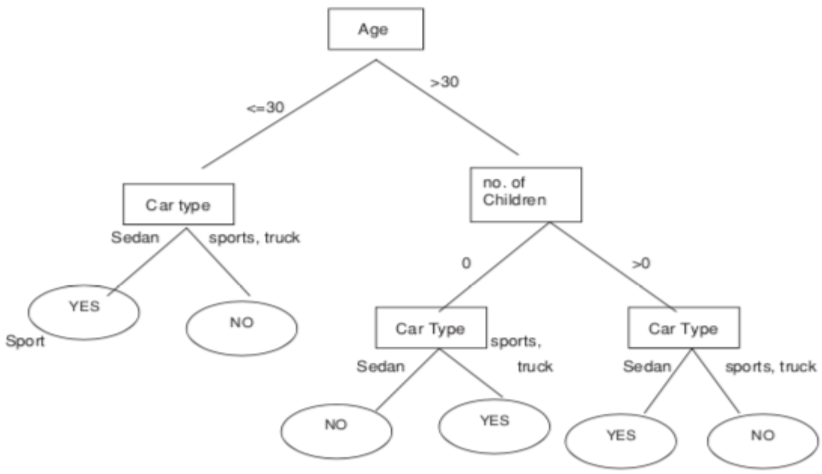


Figure 4: Classification Tree, Magazine subscription example (Sumathi & Sivanandam, 2006)

Association analysis: Association rules discover the set of interesting correlations present in a dataset between items. This mining technique has two phases. First, the search of the most frequent items and next the rules that can be created based on the frequent item sets. It is worth mentioning that in order to improve the performance of the mining process some scalability techniques must be applied in the first stage to mitigate the impact of time consumption, since usually datasets have million potentially frequent items. In this context, the association rules method explores the dataset for frequent trends and then utilizes them to reveal the most significant relations. The association rules are usually categorized as follows: constraint-based association, one-dimensional and multi- dimensional, multilevel association.

Characterization –Discrimination- Time series analysis: Data Characterization is the process of finding a short description (summarization) of user-defined data's general characteristics. Moreover, data discrimination consists of the different treatment of predefined data types of certain general characteristics. Finally, the time series or trend analysis is the statistical analysis of data that is in a series of particular time intervals.

Clustering: The purpose of Clustering is to categorize data so that "similar" data points are grouped together as a result of a similarity function. This function summarizes dense regions which correspond to clusters as depicted in the figure below. Clustering is an unsupervised methodology meaning that the model itself will classify the data and determine the groups according to the similarity and proximity of the data points based on a certain attribute. Moreover, the partitioning of the data is selected by another attribute in order to be mutually exclusive. Apart of the partitioning or density- based clustering there is the hierarchical one that generates nested clusters in a form of a hierarchical tree. Some of the most frequently used clustering methods are hierarchical clustering, density-based clustering and k-means. A clustering example can be seen in Figure 5 below.

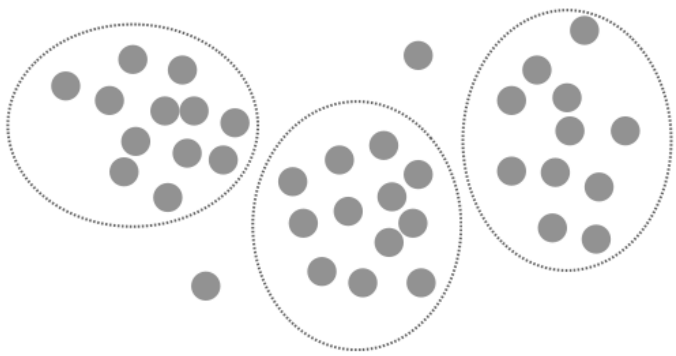


Figure 5: Clustering Example

Outlier analysis/Deviation analysis: Outliers are data points that do not fit within a general model and hence are often the outcome of interference, noise and exceptions. Outlier detection and removal are fundamental steps for smoothing the data and improve the quality of the final output. Furthermore, outliers are expressing deviation and are thus usually found through deviation detection. This operation is focusing in exploring the most important changes (temporal or group deviations) in the data between the initial content and the normative values-expected content. The following figure graphically represents an example of outlier data (O). An example of this is shown in Figure 6 below.

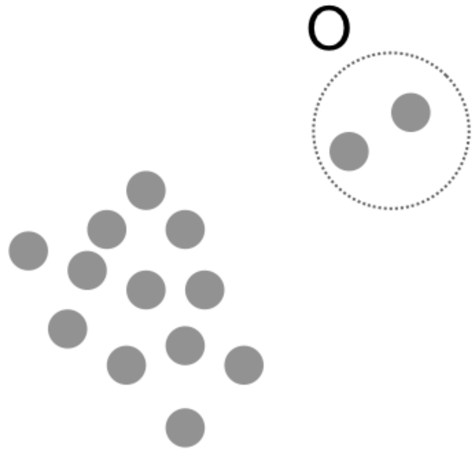


Figure 6: Outlier Analysis

Generally, the data mining models/techniques can be conventionally divided to two types. First, the predictive models such as classification are developing a process (learning) by utilizing given data and then predict the outcome of new data. On the other hand, descriptive models are discovering previously unknown informative trends, correlations, data categories and patterns by identifying data relationships. One of the most representative examples of descriptive models is the clustering model. The data mining objective is to involve both predictive and descriptive models in order to generate new data and knowledge (Mukhopadhyay, et al., 2014). In order to understand how these various models may work together it is important to look at a typical data mining procedure. According to literature (Chen, et al., 2015), (Sumathi & Sivanandam, 2006) a typical data mining procedure consists of the following steps, as seen in Figure 7:

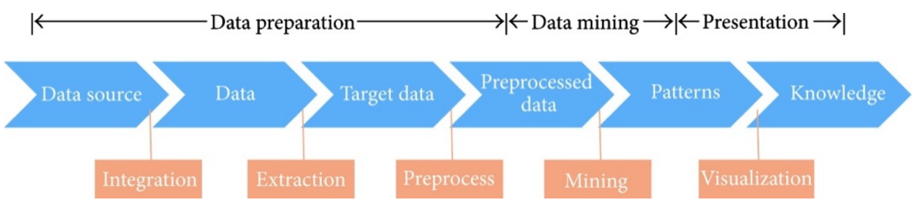


Figure 7: The data mining overview (Chen, et al., 2015)

Data preparation: is a step before mining which consists of 3 sub-steps: Data integration or enrichment from the various data sources, Data cleansing for removing 'pollution'/ outlier removal or fixing the coarse data by modifying corrupted parts of data into data mining system and

Data pre-processing to transform data into understandable format, apply dimensionality reduction and overall facilitate the mining.

Data mining: The core process is using learning algorithms to identify patterns in the data and applies an assessment for knowledge discovery. Usually, a hybrid learning is indicated to mitigate the disadvantages of each technique case-based learning, online analytical processing, genetic algorithm etc. It is worth mentioning that the preliminary analysis is an important step as it holds a large amount of interesting information by using a simple method like query tools (naïve prediction).

Data presentation: Involves the visualisation of the data through different graphical techniques and the illustration of the relationship between datasets indicating the mined knowledge to the user. Moreover, Visualization techniques might also be used towards discovering patterns methods that the projection pursues. Thus, techniques like scatter diagram or three-dimensional environment are useful for both the dissemination and the analysis of the data.

3.3 Dashboards & Open data

One of the most important areas of data science is data visualization using dashboards (Matheus, et al., 2018). Data visualization tools are widespread and used for quick mine data correlations, spotting anomalies, trends and patterns (Vartak, et al., 2016). A visual interpretation of the combination of different data analytics is enabling multidisciplinary investigation of hidden insights. The easy-to-use interactive features of dashboards allow for in depth user oriented analysis of the data and can increase the impact of conventional business intelligence applications. More importantly, dashboards support the facilitation of transparency, trust and governance by enabling citizen-led decision making (Allio, 2012) As a result, they are reducing the information asymmetry in the society (Janssen , et al., 2015). Dashboards can be utilized for various applications such as planning, performance monitoring and decision making. Dashboards are effective and efficient tools for public operations and policy as they gather knowledge, mobilize external capacity and activate open government by combining various tools and datasets (Matheus, et al., 2018). Typically, dashboards are using open data and activate a strategic reading of city's performance on sustainable mobility, energy efficiency, air quality etc (Lluïsa & Llacuna, 2020)

In general, the use, the sharing and the display of data are essential elements in bridging the gap between the public authorities and citizens. Therefore, dashboards are an essential mean for creating value by sharing open data. In 2015, the EC published a report "Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resources" (European Commission, 2015) which was dedicated to exploring the potential and impact of Open Data on Europe. Two of the key indicators measured for the assessment was cost saving and efficiency gains. The indirect benefits of the generation and re-use of Open Data in the public sector was estimated to have the potential to reach 1.7 billion EUR. However, the efficiency gains were presented as a combination of insights, namely 7000 lives saved due to quicker emergency response and 629 million hours saved in the mobility sector. In other words, data-driven decision making also known as open data sharing increased efficiency of public services and supported citizens' personal decision-making capabilities resulting in higher levels of transparency, inclusiveness and accountability.

In 2015, the World Bank issued the report "Transport and ICT| Open data for sustainable development" (WorldBankGroup, 2015) highlighting that open data are the key for smart cities. More specifically, open data is fundamentally transforming the cities and urban transportation in particular via hundreds of applications utilizing in the most suitable way open data. Subsequently, open data and dashboards are two sides of the same coin. Combined they can facilitate information sharing improving governance and environment by avoiding and exposing corruption and mismanagement.

4 Case Study & Methodology

This chapter presents the case study of the City Moonshot initiative of ERTICO ITS Europe for assessing city needs together with the problem statement and the methodological design deployed on them.

4.1 Case Study

ERTICO ITS Europe is the leading European organization for Intelligent Transport Systems (ITS). It was founded 30 years ago from the EC and 15 industry leaders to form a private- public partnership in order to fill the gap between deployment and research in the sector of ITS. Today ERTICO ITS Europe numbers 120 Partners across the world and has organized approximately 40 ITS European and World Congresses. Moreover, it has developed several networks, alliances and platforms while it has participated in numerus Horizon 2020 and FP7 projects as result of its key role in mobility thought leadership in Europe. In April 2020 ERTICO ITS Europe launched an ambitious initiative aiming to map cities' behaviour and needs. The City Moonshot initiative is a global series of interviews with 300 cities worldwide based a set of question, which was the result of an intensive stakeholder consultation performed by Ertico with its 120 members, with a goal of understanding their needs and requirements when it comes to transport and mobility. In Figure 8 the first cities interviewed in Europe are illustrated.

Figure 8: City Moonshot interviewed cities in Europe.

The cities in a series of individual interviews are sharing their insights, strategies and policies on addressing climate change, Mobility on Demand (MoD), data sharing, citizen engagement and policy priorities on sustainability and the role of transport. The cities' representatives are experts, which are either public officers, or transport/mobility researchers working for each particular city. The City Moonshot initiative aims to map the needs and unique characteristics of the interviewed cities and present the results at the ITS World Congress in Hamburg in October 2021. This is the first time that such an initiative is taking place in the field of mobility, as it is usually the mobility industry that claims to know the city needs it aims to address via its products and services.

In the post COVID world the mobility in cities is going to be transformed again giving the opportunity to public authorities to reshape urban mobility through more sustainable, responsive and smart decision making. The assessment of practices and needs of cities based on those principles is an essential element if European stakeholders (public and private) wish to understand city behaviour and if the cities themselves with to better understand their own performance, needs, priorities and as a consequence rethink their strategies and action plans based on a data-driven decision making. Also, the mapping and the dynamic visualization of cities' urban behaviour is a vital factor for sharing the knowledge and enabling the empowerment of citizens and stakeholders towards smarter decisions. Furthermore, the integration of the above via the creation of an interactive tool would be an extremely beneficiary tool for supporting the overall governance of cities by providing city-centric knowledge. So far, the lack of sufficient data and the absence of reliable data "streams" have limited any relevant endeavor in literature as they have been traditionally focusing on a small number of cities, which are usually the country capitals. As a result, there has not yet been developed any assessment framework or a system of indicators that is widely agreed and officially adapted for analyzing cities on specific sectors, such as mobility is.

The main problem statement of this thesis is expressed by the following research questions: "How can data mining address the issues of data availability and integrity to enable a holistic assessment of cities in Europe; and how can city profiles be formed?". This thesis utilizes ICT tools in order to deal with the lack of data availability and by adapting a data-driven approach using the EU Databases it succeeds in creating an exploratory assessment tool for the European Cities. The tool, combined with cluster analysis consist the basis on which the identification of city mobility profiles in Europe can be built. It is worth mentioning that an assessment at a global scale is extremely complicated because the cities do not share similar policies and they do not adhere to the same legislation framework. Specifically, the group of cities presented in Figure 8 constitute the focus group of this Thesis because only they have been interviewed in Europe so far and they thus adhere to the same general EU umbrella legislation. The main strategy utilized in this Thesis was to develop a system of indicators and leverage widely used open software to increase the quality of data and the impact of the results. The Thesis aimed at transforming business intelligence into city intelligence. Under these conditions, the problem statement can be analyzed by the following research tasks (RT). They are also presented in Figure 9 as parts of Thesis's flow chart.

- RT1: Data extraction and preprocess from EU databases in order to enrich the City Moonshot database with quantitative data regarding cities' present state. The purpose is to characterize the cities according to sustainability and smartness (present state) and compare the results with their priorities and plans.
- RT2: Creation of a mobility indicators system by identifying appropriate indicators and their corresponding categories in order to assess the cities. Also, the creation of indexes to support the overall analysis.
- RT3: Use of Machine Learning as method to accurately estimate missing data avoiding generalizations.
- RT4: Clustering of the cities based on the final main database in order to identify mobility profiles.
- RT5: Creation of an interactive tool for dynamically visualizing and explore data in order to enable personalized mobility insights.

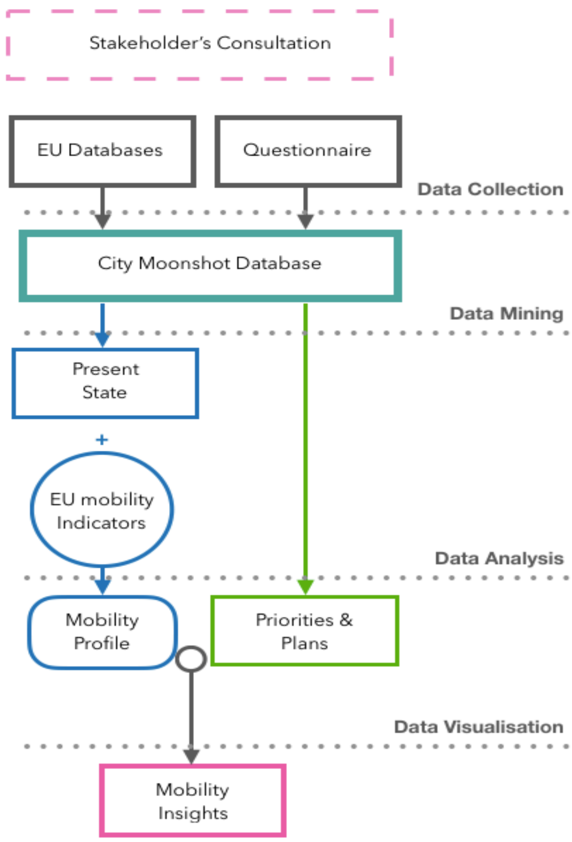


Figure 9: Thesis's flow chart.

4.2 Methodology

The methodology of the thesis follows the recommendations of EC and OECD on creating composite indicators. The reason is that even though the ultimate goal of thesis is not to create a ranking or to assign a score to each city but rather to cluster the cities, the use of composite indicators is necessary to form the pillars –categories, sectors upon which the clustering is performed. To elaborate on this, clustering allows for the assessment of the citie's behaviour without losing the dimension of the individual sub-indicators. In that context, the set of technical guidelines to improve the quality of composite indicators published by the EC and OECD (European Commission & OECD, 2008) is the basis of the methodology adopted in this Thesis together with other scientific studies on similar subjects as they offered an empirical insight on how indexes are constructed (Haghshenas & Vaziri, 2012), (Alonso, et al., 2015), (Lopez-CarreiroI. & Monzon, 2010), (Battara, et al., 2018), (Lopez-CarreiroI. & Monzon, 2010). Finally, the proposed framework based on the methodological assessment of several sustainable city rankings, benchmarking and indexes issued by Sáez et al. in 2019 was taken under consideration.

4.2.1 Selection of the Cities

The size of the cities matters when it comes to evaluate and their sustainability (Alonso, et al., 2015), for that reason cities are carefully grouped by the European Commission depending on the selected scope but the following clusters are identified by the European Commission in the Statistics for European cities-Population (Eurostat, 2020), as shown in Table 2. European cities are considered the cities located in the 27 countries of the European Union together with cities from EFTA, Switzerland and the former EU member: the UK. According to the degree of urbanisation (Eurostat, UN, OECD, The World Bank, 2021) issued by Eurostat and OECD, 72 % of Europe's population lives in urban areas. This percentage refers to, the harmonized definition of urban centres-cities which defines them as the areas with more than 50,000 inhabitants and at least 1,500 inhabitants per km2.

Table 2: Classification of European Cities and

Group	Population	City Moonshot Cities
1	< 100.000	4
2	100.000-< 250.000	11
3	250.000-< 500.000	13
4	500.000 -< 1.000.000	15
6	1.000.000-< 3.000.000	12
7	3.000.000-<	2

Almost, 6% of EU27-EFTA-UK population lives in metro-cities of 3 million and another 6% lives in small or medium-sized cities of populations between 50,000 and 100,000. Moreover, there are 31 European cities of 1 million or more and 214 cities of 250,000 or more which account for 29 % of the total population. In addition, some 240 million people live in European cities of 50,000 or more inhabitants. In that context, the selection of cities is based on their territorial typology considering the distribution of European cities as analyzed above. Also, where the cities are located was another criterion used in selecting the city as the aim of the Thesis is to include as much European cities as possible. This happened so that the complexity of comparing cities with completely diverse trajectory, governance and history such as capitals is mitigated by including cities from same countries. Finally, it is important to clarify that the number of City Moonshot cities which are listed in Table 2 in correlation to population values are defined according to the population of cities (urban areas-agglomerations) and sometimes to the population of municipalities.

4.2.2 System of Indicators

The information deduced by the City Moonshot is descriptive and prescriptive. Thus, in order to utilize it for the purpose of the assessment information needed to be transformed into objectively measurable, specific and interpretable indicators. Therefore, research continued with an initial desk research phase in order to collect quantitative indicators and complement it with an additional secondary desk research by applying a data-driven approach.

In the relevant literature, a plethora of indicators and indexes related to urban mobility smartness are available for a sustainability characterization. Indicators are variables applied as decision making tools that evaluate and measure the progress of a specific phenomenon –policy impact (Joumard & Gudmundsson, 2010), (Gudmundsson, 2003) towards an objective. It is well established in the scientific world that indicators are useful to assess the multidisciplinary nature of urban smartness (Escolar, et al., 2018), (Giffinger, et al., 2007). Nevertheless, still a well-defined system of indicators to measure urban smartness doesn't exist due to diverse parameters, controversial methodologies and subjectivity (Sáez, et al., 2019), (Lopez-CarreiroI. & Monzon, 2010). Hence, as Alonso et al. the selection indicator should be in accordance with the established definition of urban mobility (Alonso, et al., p. 2015). However, an essential factor for measuring through quantitative indicators and thus identifying the most appropriate indicators is the availability of necessary valid data (Garau, et al., 2016), (Tafidis, et al., 2017).

The selection of the indicators followed three different stages. In the first stage, a literature review of recently published papers and reports with similar subject was conducted. The outcome of this process was the identification of numerous indicators, but it gave the framework for the subsequent indicator analysis working as a foundational step. In that context, in order to have a solid basis and to guarantee the consistency and the ability to make insightful comparisons, with the results of other studies preference was given to similar works as it is shown in Table 3.

Table 3: Information related to past studies

No	References	Authors (Year)	Indicators
1	Cagliari and smart urban mobility: Analysis and comparison.	(Garau, et al., 2016)	18
2	Evaluating sustainability and innovation of mobility patterns in Spanish cities. Analysis by size and urban typology	(Lopez-CarreiroI. & Monzon, 2010)	15
3	Smart mobility in Italian metropolitan cities: A comparative analysis through indicators and actions	(Battara, et al., 2018)	28
4	Catapult Urban Mobility Innovation Index	(Taborda, et al., 2017)	22
5	2020 Deloitte City Mobility Index	(Deloitte Insights , 2020)	82
6	Cebr Urban Mobility Index	(Cebr, 2017)	20

In this review, scientific papers published with almost the same scope were included, which aimed to assess the smartness of urban mobility by adapting similar methodologies, and well-known mobility indexes created by international organizations. The second stage was the identification of the indicators which were suitable to express the multi- dimensionality of urban mobility and strengthen better decision making based on the widely used criteria shown at Table 4 (Haghshenas & Vaziri, 2012), (Lopez-CarreiroI. & Monzon, 2010) (Alonso, et al., 2015). It is worth mentioning that the theoretical analysis in Chapter 2 was of great importance to analyse the relevance and sensitivity of the indicators in this stage.

Table 4: Criteria for indicator assessment

No	Criteria	Description
1	Target relevance	Indicators must illustrate one aspect of urban smart mobility
2	Standardised	Indicator should be standardized to enable cities comparison
3	Sensitivity	Indicators must be able to reveal cities' mobility changes
4	Independent	Indicators should be independent of each other
5	Transparency	Indicators should be easy to understand and possible to reproduce
6	Validity	Indicators must measure the aspect it is supposed to measure
7	Data availability	Indicators must be linked with regularly updated databases seeking measurable and reliable data

Finally, the last stage was the assessment of the frequency of the indicators' usage in past studies. On a general note, the indication of the most frequently used indicators in the pursuit of smartness is an extremely impactful factor as it secures compatibility, transparency, reusability and inter-connection with other systems and studies. The indicators of the past studies were collected and categorized, in order to better access the hidden questions each indicator is addressing (Haghshenas & Vaziri, 2012).Table 5 shows the results of this summarization which were compared and enriched with the results of, similar reviews but with an overall sustainability focus, in order to be aligned with the theoretical foundations of smartness, which is crucial for this Thesis. The Appendix A, is shown the result of 17 studies review that was published by Hashenas & Vaziri in 2012 which also was the basis for Alonso et al. in 2015 (Alonso, et al., 2015) proposing a holistic system for mobility characterization of cities in the global and European scales respectively.

Table 5: Categories of Smart Mobility

No	**Categories**	**Number of Indicators**	**Frequency of Use**
1	Climate Change (CO2e-Energy)	3	4
2	Bicycle-Walkability	5	8
3	Mobility support systems	5	15
4	PT Accessibility - Infrastructure	5	12
5	Safety	2	2
6	Urban Structure/Land Consumption	3	5
7	Air Quality	4	3
8	Congestion	3	3
9	Investment	3	8
10	Alternative fuels –Renewable Energy	4	8
11	Private transportation-Car	2	4
12	Autonomous Vehicles	2	2
13	Shared Mobility - MaaS -Integrated Transport-seamless	4	11
14	Modal Diversity	2	4
15	Vision and strategy	3	9
16	Openness	2	6
17	Regulatory environment-incentives	4	12
18	Wellbeing	2	5
19	Engagement	1	3
20	Satisfaction	1	2

In the light of the above, the selection of the indicators for this Thesis was conducted with the aim to maximize the number of cities that can be assessed based on the final system of indicators. As it was explained above, the comparison between cities is rising as an extremely useful additional decision-making tool for smart cities, so the indicators must provide the ability for such an analysis.

Furthermore, a critical criterion for narrowing down the indicators was their reference area (e.g. Country, Region, cities etc). In literature, so far, most of the European Cities indexes were developed for the comparison of Capitals allowing the use of indicators, which are available only at country level. Moreover, other studies are focusing on specific country's cities using, not city databases but national ones extrapolating values from nationally meaningful indicators. In that context, the use of indicators which were referring on regional and prefecture level NUTS2-NUTS3 respectively, were highly preferred for the research of this Thesis. The reason is that the scale used in these is sufficiently representative of the differences between cities but more importantly because they are contributing to holistically characterize the cities sustainability by including the indirect effects of cities on areas and ecosystems beyond their boundaries, as (Mori & Christodoulou, 2012) thoroughly discussed. Cities are pursuing agglomeration benefits on other areas to support their economic and social development.

The research for this Thesis was carried out between January 2021 and March 2021. The information found was mostly (but not exclusively) in English, while the literature used was exclusively in English due to the vast majority of related journals, papers, articles etc. and their credibility on the topic. For the internet search, Google search was used (keywords for the query: innovation index city, mobility index, transport, indicator, sustainability, smart, sustainability and similar terms), while the literature research was conducted using Google Scholar. The keywords used for the research were "city mobility", "transport", "indicator", "innovation", "smart" and similar terms were included, together with keywords such as "smart sustainability" and "index".

4.2.3 Data collection

Data collection for characterizing cities is a fundamental step for establishing a reliable and international comparable set of data (Sáez, et al., 2019). The assessment of the European cities was undertaken through a combination of qualitative and quantitative data collection methods. The main qualitative phase of the research consisted of semi-structured in-person interviews conducted under the City Moonshot Initiative (the survey) with local stakeholders via GoMeeting. The interviewees were either city officers or transport experts working in local public authorities. During the interviews, a set of 68 pre-determined, closed-ended and open-ended questions were asked, which covered the spectrum of urban mobility by exploring both the ongoing activities and the vision and needs of cities for the future (survey available upon request to Ertico ITS Europe).

Also, they allowed in the most cases a fruitful unscripted discussion between the interviewee and the interviewer (the role of latter usually taken over by a Senior manager of ERTICO –ITS Europe and accompanied by the author of this Thesis) digging into more depth on particular cities' areas of interest. The coordination of each interview was performed by an Ertico Mobility expert together with another expert who supported the interview by keeping notes and clarifying any misunderstandings. The result until today was a database of 110 cities worldwide and one of 57 European cities.

On the other hand, the quantitative data was obtained using mainly official European/national datasets, and cities' documentation More specifically, in order to have a solid basis and to guarantee the consistency and availability of adequately reliable data for the indicators, preference was given to official European Commission datasets and affiliated studies:

- Eurostat's general Urban Audit European Environmental Agency repositories for transport, environment and Energy. In addition, datasets offered by EC Joint Research Centre open data platform and Europe Union open data portal as well. Also, other databased and studies affiliated with the EC.
- EC's Interactive tools which are providing unstructured data published by European Agency, TRIMIS Smart mobility and Services dashboard, EU's Regional Innovation Dashboard.
- Internationally accepted mobility indexes which offer open data such as TomTom Traffic index.

Moreover, aiming to increase the integrity and transparency of the information collection process, the documentation that was officially provided from the interviewees was considered as cities' documentation. The ultimate goal was to create a unified homogeneous database dedicated to mobility for a sufficient number of European cities. In order to avoid overly biased analysis and enable a wider European scope of comparison between cities the following criteria for selecting the databases were formed:

- The databases should have data for a variety of European Cities in terms of size referring to most European countries, including always cities from southern, norther and central Europe.
- The majority of the cities' data should be available for the period 2016-2019, as the 2019 was set as the reference year for this Thesis. If for a minority of the cities' the data was less timely the integrity of the data was secured by applying imputation techniques (Hollanders, et al., 2019a). Also for some indicators like population, density etc. a larger gap was considered acceptable because during this period any changes occurring are insignificant.

The aggregation of the different databases necessitated the manual connection of various NUTS levels and cities. Eurostat developed a hierarchical system for distinguishing the economic territory of the EU called Nomenclature of territorial units for statistics-NUTS. Thus, depending on the activities the Eurostat monitors the progress of specific indicators in the appropriate NUTS level. According to the NUTS classification Europe is divided in three levels: NUTS 1 large regions which are capturing major socio-economic activities, NUTS 2 basic regions which are formed based on the application of regional policies, and NUTS 3 smaller regions or prefecture for closer diagnoses. Therefore, in order to aggregate the various databases, the manual assignment of the appropriate NUTS2, NUTS2 and NUTS1 level for each city was a prerequisite for structuring the final database. This was manually performed by the author for the purposes of this Thesis.

Moreover, another important step was the transformation of the data that originated from databases which was referring to urban areas and other NUTS levels into a per capita version in order to be assigned to each city accordingly and allow meaningful comparisons. Finally, it is worth mentioning that some data are not directly available in any editable form. To collect this data a manual extraction was performed from the original source.

Under these circumstances, Microsoft Excel was used as data management tool. Spreadsheets allow for storing both quantitative and qualitative data easily. The functions such as VLOOKUP(), HLOOKUP(), SEARCH(), SUMIFS() etc are provide flexible customization features for reorganizing and analysing data. Moreover, data stored in spreadsheets can work seamlessly with other applications maximizing the impact of the analysis. For that reason, most of the online databases found are available in xlsx format, making any compatibility operations extremely smooth. These capabilities of Microsoft Excel spreadsheets were a perfect fit for handling the problems of the available medium sized dataset, some of these were identified as being the following:

- Inconsistent variable names and subject identifiers: The databases are using different variations of the city names while city and NUTS identifiers are not available for most of them.
- Inconsistent layout in multiple files: Data was fragmented in a lot of different databases which were originated with various layouts.
- Inconsistent categorical variables: The context of each qualitative question is not a single common value.

4.2.4 Missing Data Imputation

The extended City Moonshot database contains more than 4.500 cells (56 cities, 21 indicators, 68+ questions). The selected group of cities is not systematically monitored according to the final proposed system of indicators and even though European Commission Databases and other original sources cover a wide range of cities some data is still missing. However, the availability of data was a prerequisite for selecting the indicators, so imputation techniques were used to estimate the missing data in order to fulfill the criteria. The following process is the general framework which this Thesis followed:

- If city data is not available from the main data source, different official sources will be used.
- If city data is available for the past year, then this data will be utilized together with the corresponding higher aggregate NUTS level (NUTS3, NUTS2, NUTS1 or country level) which has available data too, in order for the missing data to be imputed by the formula:

$$D_C^t = \frac{D_C^{t-1}}{D_R^{t-1}} * D_R^t, C \ denotes \ city's \ level \ and \ R \ higher \ NUTS \ level, t \ year \quad (1)$$

- If city data is available for the past and next year and there is no available data for a corresponding higher NUTS level, then the missing data will be estimated using linear extrapolation.

$$D_C^t = \frac{D_C^{t+1} - D_C^{t-1}}{Y_C^{t+1} - Y_C^{t-1}} * (Y_C^t - Y_C^{t-1}) + D_C^{t-1},$$
$$C \ denotes \ city's \ level \ and \ Y \ denotes \ year \ (2)$$

- If there is no city data then a higher aggregate NUTS level (Urban Area, NUTS3, NUTS2, NUTS1 or country level) will be used for the most recent available year – this technique is used depending on the context of the indicator in aiming to avoid generalizations.

$$D_C^t = D_R^t, \qquad C \ denotes \ city's \ level \ and \ R \ higher \ NUTS \ level, t \ year \quad (3)$$

- If there is no available data for the city or any other NUTS level then data is predicted by using ML as explained in the next section.

4.2.5 Machine Learning (ML)

Machine learning was used in this Thesis in order to predict missing values for indicators by creating new test datasets. These datasets were developed by using different indicators to avoid creating any unintentional correlation between the composite indicators. The selection of those supportive indicators was based on the literature review. According to the H2020 projects' report "Urban Mobility: Preparing for the Future, learning from the Past" based on the assessment of several cities worldwide both GDP and density are factors contributing to road congestion. The rapid urbanization of the last decades stressed the cities' transport and land use planning. Density of a city especially in the wider urban areas contributes to a low urban sprawl increasing the degree of car dependency (Peters, 2018) and is therefore resulting in higher population and density increased traffic rates (Sang, et al., 2018). Moreover, the combination of GDP with the availability of cheaper cars, lower fuel prices and better public transport infrastructure makes this a very useful indicator, which has a positive correlation to urban transportation (Haghshenas & Vaziri, 2012). In 2019, Albalate and Fagate (Albalate & Fageda, 2019) discussed the importance of GDP and statistically found relevant influence to mobility. Also, GDP per capita expresses the economic growth which together with its energy intensity drives the CO2e emissions mainly due to up to this day high carbon intensity of energy (Edenhofer, et al., 2014). Also, the influence of weather and climate on transport demand (Rudloff, et al., 2015), (Böckera, et al., 2019) and effect of the unique landscape of each on the citizens' mobility behaviours are equally important factors in measuring smartness. Hence, the following set of indicators to assess city smartness was formed with the addition of geographical indicators and by applying some of the aforementioned imputation techniques too:

- Population: The main source was Eurostat Regions and Cities Repository (Eurostat, 2021a). In case of missing values and in order to cross check the values, official sources found in Wikipedia (Wikipedia , 2021) were used. The reason, is that despite the general sense there is inconsistency in defining the various urban entities in literature and datasets– municipalities, cities, urban area, metro area and agglomeration.
- Density weighted: The typical density is not always representative of the concentration of citizens in a given territory due to cities urban planning, historic spatial structure etc for that reason the European Commission developed a methodology to calculate the weighted density (Poelman, et al., 2020). In case of missing values the data were imputed using the density of cities and the corresponding average weighted density factor of same sized cities of the same country.
- GDP per Capita (€/capita): The main source for EU cities and Norway (Eurostat, 2021a) for United Kingdom (UK National Statistics, 2021), and for Switzerland (Switzerland's Federal Statistical Office, 2021).
- Energy Intensity: The source was the Energy Intensity Report of EEA (EEA, 2019a). The missing data was found using linear extrapolation.
- Car ownership: The main source was Eurostat Regions and Cities Repository (Eurostat, 2021a) at the cities level in case of missing values the values from NUTS2 level were used.

- Coastal or no Costal city: This information is available in Eurostat Regions and Cities Repository (Eurostat, 2021a) at NUTS3 level.
- Climate type Classification: A simplified version of Köppen-Geiger classification (PVsites, 2016) was manually assigned to the corresponding NUTS2.
- Corresponding NUTS2 and Country: The source was Eurostat Regions and Cities Repository (Eurostat, 2021a).
- Date: The reference year of target data.

Based on those indicators and the indicators with the missing values, different databases were created after the necessary cleansing of the data due to the differences in cities' data availability. These manually made datasets were the input for the respective prediction processes. The calculations were performed using one of the leading data science platforms (KNIME) according to Gartner (Gartner, 2017). KNIME is an open-source platform with a rapidly growing community that came into being two decades ago and it is supported by the European Commission. The main characteristics are summarized below:

- Graphical user Interface: A Node repository contains all kind of operations which can be used by drag and drop to the main analysis flow. Then these operations are linked by drawing lines between operations according to the process scope.
- Methods: It supports more than 1000 standard operations both analysis techniques such as neural networks, decision trees and data preparations techniques. Also, it is connected with Weka offering the opportunity to apply flexible custom techniques.
- Cooperation with other Platforms: KNIME is able to cooperate with Weka, SQL, Python, R and utilizes files in various forms xlsx, csv, xml and others.

Since the datasets are created based on theoretical principles and also as they are subject to data availability, the designing of an integrated deployment mining strategy that is automatically adapted based on each content is vital. KNIME (KNIME, 2021) provides a special single component for such problems, the Automated Machine Learning (AutoML). This flexible solution automates the training and validation of up to nine classification algorithms. After carefully configuring the previous steps and executing AutoML, the performance (metadata) of each algorithm is shown in "interactive view" option. For the selection of the training model the highest accuracy (average class accuracy- arithmetic mean (Kelleher, 2020)) is preferred as the most suitable metric among the other available options in KNIME. The following figure illustrates the design of the developed process.

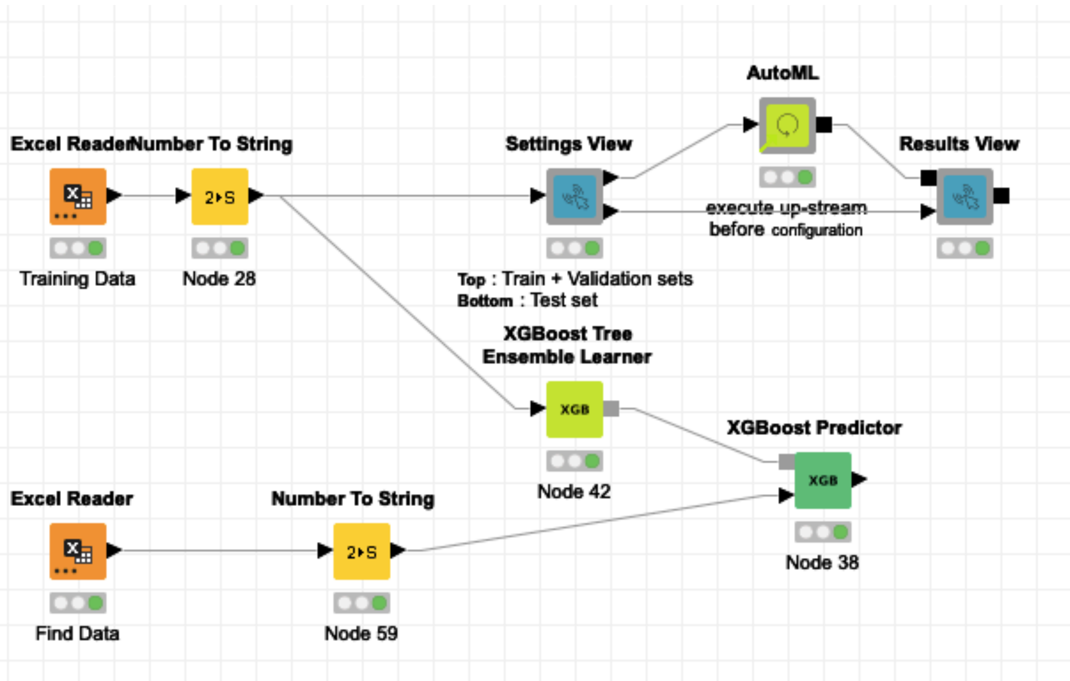


Figure 10: Missing Data Workflow

In more detail, the first step of the process created for addressing the challenge of missing data is the data preparation stage, the dataset after cleansing is split into train and missing sets. The first is connected with the Number to String component in order to transform the type of the target column into String in order to be complied with the AutoML operations. The data then is passed to the Settings View component, which allows the manual configuration (Open Views –interactive view) of the upcoming AutoML process. The AutoML as explained above, performs an automated unsupervised ML training of both binary and multiclass classification techniques (Naive Bayes, Logistic Regression, Gradient Boosted Trees, Decision Tree, Random Forest, XGBoost Trees, Generalized Linear Model (H2O), Deep Learning (Keras).

The result view component shows (Interactive View) the final ranking of the techniques based on the previous selected criteria together with stats regarding the selected ML technique. Depending on the results the Learner component of the corresponding technique is used. In some cases, the use of the Domain Calculator component is necessary before the Learner in order to retain possible value domain and restrict the number of possible values. The output of the Learner component is passed to the first port of the appropriate Predictor; its second receiver is connected with the data based on which the prediction will be calculated. This process is shown in Figure 11 here below.

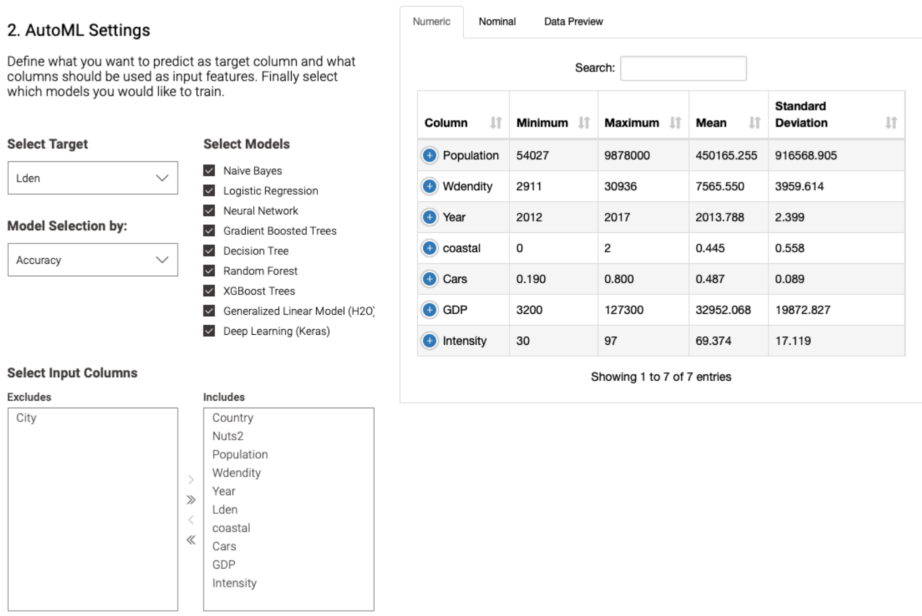


Figure 11: AutoML Configuration

4.2.6 Composite Indicators

"… it is hard to imagine that debate on the use of composite indicators will ever be settled…" (Saisana, Tarantola, & Saltelli, 2005). In theory, a composite indicator is the compilation of a group of individual indicators with no standard units of measurement into a synthetic index. The use of these is criticised in the literature mainly because sometimes they lack the relevant focus (Escolar, et al., 2018) and their arbitrary weighing process hinders their advantages. To elaborate, synthetic indexes are enabling a more effective evaluation and interpretation of complex realities rather than searching for a common trend in a multi-dimensional concept. In contrast, they are usually misused to serve political and corporate interests and also, they introduce biases by adopting a generalist approach for performance conclusions. However, if the methodological framework is based on mathematical and conceptual principles they allow the creation of a powerful communication and analysis tool to support cities identify strategic weaknesses and opportunities (Sáez, et al., 2019). In that context, the creation of every composite indicator in this study respects the following steps:

Data Normalization: To prevent composite indicators from comparing apples with oranges, it is suggested to normalize the data prior to any aggregation in order to avoid the usage of incomparable measurement units. Due to different type of data included in the final dataset, suitable techniques were applied, as explained below:

- Quantitative data: Most of the quantitative indicators are fractional with values varying between 0% up to 100%, usually following normal distribution. For the rest of the unbound indicators there is no limitation/threshold for the values so more actions are needed to normalize the data. The Min-Max technique was selected to set an identical range for indicators [0,1]. The maximum and minimum values were subtracted in order for the gap between the observed value of each city and the minimum value to be divided by their difference. Consideration was given to clearing the dataset from the presence of outliers before the subtraction of the max-min values to avoid widening the widening the range and distorting the transformed indicator

$$D_{new} = \frac{D_{observed} - MIN(\forall D)}{MAX(\forall D) - MIN(\forall D)} \quad (4)$$ Qualitative

- Scale 1: “None” received 0, “Under Development” received 0,5 and “Yes” received 1
- Scale 2: “Less often” received 0, “Annually” received 0,25, “Quarterly” received 0,5, “Monthly” received 0,75 and “More Often” received 1.

Weighting: The composite indicators are relying on weights in the indicators which are widely used. Weighing contributes to the balance of various factors even though it might reduce the importance of some indicators for some cities. On the contrary, there are methods that attempt to render better the special significance of some factors by allocating the weights to be used according to the contribution of specialists and citizens. Nonetheless, they cannot avoid subjectivity and they make the comparison more difficult (Akande, et al., 2019). Therefore, each composite indicator was estimated as the unweighted average of the cities’ scores after the explained above normalisation process took place.

4.2.7 Data analysis and Visualization

Sometimes due to arbitrary selection of indicators or lack of data the final systems of indicators are “indicator rich but information poor”. This leads to inconsistency and confusion which affect the decision makers because the interrelationship of the indicators is high. Therefore, an exploratory analysis of the above methodological choices is necessary in order to investigate the underlying structure compared with the theoretical framework. Moreover, it should be used to identify and study the groups of cities that have statistically similar performance providing an interpretation of the results. As a result, clustering analysis was used as a purely statistical method of exploring the impact of the proposed structure, serving as a diagnosis tool. Also, clustering analysis allows for a proper dissemination of the information without compromising the integrity of the individual indicators. Additionally, taking a decision on the assessment of the indicators’ robustness through correlation/sensitive analysis is equally essential. As explained by (Sáez, et al., 2019) most of the studies in the relevant literature haven’t adapted any similar analysis, weakening the methodological framework of the assessment. This correlation/sensitivity analysis provides useful insights regarding the impact of any uncertainties and regarding the influence of the various indicators.

The dissemination of the results (data, information and metadata) is a crucial stage and too often is neglected by statisticians even though it affects the assessment's credibility. This fact reflects on the ability of users to understand and analyze the information and on the consistency and coherence of the results. The visualization choices have huge influence on the final interpretation and interpretability impact. Thus, identifying the most coherent presentational tools according to the target audience is critical in order to communicate the most information. Decision makers, whether they are public or policy makers, do not generally study methodological reports or complex stats and charts, hence the recipients of this information in such cases are limited to the scientific community. But this is in direct contradiction to the general purpose of research on smart cites because the produced knowledge should be but is not shared with the decision makers.

The Power BI (PowerBI, 2021), which was used in this Thesis, is the leading software (Gartner, 2017) for supporting decision makers to create information and knowledge by assessing, gathering and analysing data with the use of various analytical methodologies and mathematical models, according to the definition of Business Intelligence (BI) given by Vercellis (Vercellis, 2011). Cities are acting in a lot of ways like business so the two concepts are closely linked transforming Power BI into a great tool for city intelligence. Microsoft developed Power BI as cloud-based tool for business users to analyze, interact and visualize data. It provides access to other Microsoft applications like Excel, Sharepoint, Forms etc maximizing the capabilities for data manipulation and dataset creation. The Power BI tool mainly consists of interactive reports and dashboards entirely accessible in terms of time and location. A dashboard in Power BI is a flexible and responsive collection of various visualization types designed to ensure alignment with the end user's requirements.

5 Results

This chapter contains the proposed by this Thesis system of indicators and discusses the results of the exploratory correlation analysis performed as part of this research. Moreover, the architecture of the resulting dashboard is explained along with practical implications.

5.1 System of Indicators

One of the most important factors of this research is the visibility and transparency of the methodological framework created in order to contribute towards a general framework for characterizing European cities. This section presented the system of 16 indicators identified as the outcome of applying the methodology explained in chapter 4. For each indicator, the corresponding category is presented along with the rationale for the selection justifying its suitability. Significantly, the structure for each indicator is clearly described because some of them are actually indexes. Moreover, information regarding the data source, availability and imputation techniques is provided in order to serve as an indication of both the quality of data and its openness. Finally, the indicators were grouped into sectors in order to increase the comparison capabilities of this study and the general understanding of the indicators' essence. An overview of the indicators and their corresponding categories and sectors is presented in Appendix B.

5.1.1 Environment

Category:	**Air Quality**
Indicators	PM_{10} and NO_2 Annual Mean Concentration μg/m3
Structure:	Unweighted average of the normalised scores.
Rationale:	NOx is the most important pollutant in the mobility sector, it is contributing to the creation of a mixture acidifying substances, nitric oxide (NO) and nitrogen dioxide (NO_2) which causes almost 80,000 premature deaths. Overall, the road transport in Europe is estimated that contributes approximately 70% for excessive concentration of nitrogen dioxide (NO_2) causing more than 7% of Europe's population to be exposed in levels higher than EU/WHO limit value (EEA, 2021a). Moreover, road transport is responsible for the 30% of the excessive concentration of Particulate Matter (PM) which in total affects more than 75 % of Europe's population. As a result of the above analysis the air quality is expressed as the mean value of the cities' normalized performance on PM_{10} and NO_2 because they are the most frequently measured air pollutants.

Data Source:	The emission sources from EEA Air Quality Statistics-Dashboard (EEA, 2021b; PUM, 2019). The Dashboard provides the open data in the form of CSV file. The various sources are classified in many ways so the following clarification are useful for the quality of the data. The mean value of the emissions from urban traffic stations located in continuously built-up urban areas in close distance to a single major road was assigned for each city.
Data Availability:	The data was available for 87% of the cities- Reference year 2019, at city level.
Imputation Technique:	The missing data was estimated by using the Air Quality report of European Environmental Agency in 2019 (EEA, 2020) and other sources (EKPAA, 2019) (Municipality of Guimaraes, 2017)

Category:	Noise
Indicators	Percentage of population exposed in L_{den} higher than 55 dB
Structure:	Single Indicator
Rationale:	The World Health Organization (WHO) has announced that traffic noise is the second most important factor for citizens' health in Western Europe. According to European Environment Agency in urban areas due to road traffic noise more than 41 million people are exposed above 55 dB L_{den}. In general road traffic noise is the most impactful environmental noise which together with the others is estimated to contribute to 12,000 premature deaths (EEA, 2021a). Also, it is calculated that 22 million people are suffering from chronic high annoyance and 6.5 million from chronic high sleep disturbance. L_{den} is descriptor of day-evening- night noise level which included a penalty for higher levels during the evening and night.
Data Source:	The noise population exposure data based on strategic traffic noise maps issued by EEA countries according with 2002/49/EC illustrated on EEA Noise Observation & Information Service for Europe (EEA, 2019b).
Data Availability:	Data was available for 86% of the cities. The data's reference year was 2017 and the reference area was greater-metro areas.
Imputation Technique:	In order to find some missing values for Greek cities except Trikala and Guimarāes different source was used (EKPAA, 2019), (Municipality of Guimaraes, 2017) with reference years 2012-2013. For all the other missing values a unique data mining technique was performed.

Category:	Land Consumption
Indicators	Green public areas per capita
Structure:	Single Indicator
Rationale:	One of the constant external challenges for urban mobility is the loss of public/green areas. Terms like 'pedestrianisation' and 'green streets' are becoming part of cities action plans describing the recreation that many cities are undergoing open up the public urban space activating the transformation of roads into green areas and pedestrian zones (ELTIS, 2020). Nowadays, public green areas are hosting social and micro mobility both important concepts for

	future mobility. For that reason, in 2009 in the Action Plan on Urban Mobility the seventh action is dedicated to the accessibility of green zones.
Data Source:	The available public green in European cities per inhabitant JRC dashboard (Maes, et al., 2019). This interactive tool is using data from the Copernicus maps in order to calculate the public green per capita for European Cities solving the problems with different definitions of public green and different measurement techniques. The data for each city was manually extracted by the dashboard.
Data Availability:	Data was available for 95% of the cities
Imputation Technique:	Data were missing for the Norwegian cities and only an indicative range was provided. For that reason, other sources were used to estimate the missing values (EGCA, 2018), (Statistics Norway, 2013).

Category:	**Renewable Energy**
Indicators	Percentage of renewable sources penetration on transport
Structure:	Single Indicator
Rationale:	The European Commission in 2018, set the target of 14% of the total energy used in transport to be originated from renewable energy sources (RES) by 2030, gradually decoupling energy from fossil fuels. The use of RES is contributing to the reduction of emission and the lowering of dependence on fuel imports expressing the quality of the energy used for transportation. Furthermore, electrification is seen as key element, the mean towards a sustainable mobility based on biofuels, bioliquids etc.
Data Source:	Eurostat is monitoring the share of energy from renewable sources (Eurostat, 2021b)
Data Availability:	Data was available for 100% of the cities. The data's reference year was 2019 and the data was only available and meaningful for the countries (at national level).

Category:	**Climate Change**
Indicators	Greenhouse gas (GHG) or CO2e emissions from transport (tn/year)
Structure:	Single Indicator
Rationale:	The Climate change after the Paris Agreement in 2015 is the dominant factor for assessing the progress of our society. Due to that, an overall reduction of GHG emissions was observed in the last years. According to EEA the transport sector hasn't followed this trend, increasing the magnitude of transport's contribution. EEA projection forecast that the transport emissions will increase 32% by 2030 with the current policies and if EU adopts more drastic measures this percentage will fall to 17%. The urban mobility is responsible for almost the 30% of overall emissions, for that reason cities around the world commit to a cross sectoral reduction by signing the covenant of mayors.

Data Source:	The online emissions inventory of the Covenant of Mayors provides information regarding the GHG emission of the cities which have official committed to reduce their climate impact (CoM , 2021). The emissions values produced by transport for each city was manually extracted by first searching for the latest updates for the cities that have documented their progress.
Data Availability:	Data were available for 82% of the cities. The reference year may vary as the cities have different strategies and action plans. But for the purpose of this assessment only reference years between 2012-2019 were considered acceptable for comparing the cities. That happened, mainly because it has been reported in 2017 that the overall improvement from the sector was only 7% (Kona, et al., 2017). The performance of cities with older reference year was treated as missing values. Moreover, the data was available for the city and municipality level.
Imputation Technique:	The missing data were estimated using ML based on a special database created for this purpose.

5.1.2 Strategy & Governance

Category:	Strategy
Indicators	The adaptation of SUMP, Data Sharing Strategy, MaaS Strategy and Covenant of Mayors
Structure:	Unweighted average of the normalised scores. - The scores were calculated using Scale 1 (Section 4.2.6)
Rationale:	Strategies represent the quantified vision of the cities and the foundation of their action plans. Moreover, they are capturing the cities level of grounded view and readiness to approach sustainable innovation and the capability to achieve it. This indicator combines a set of Strategies, which assess the following features: **Sustainable Urban Mobility Plan (SUMP):** It was issued in 2013 by the European Commission and has introduced a new approach for integrative planning aligned with the ever-changing mobility dialogue. SUMPs support the defining of a long-term vision on city policies and actions and a clear implementation plan of these by cities. **Data Sharing Strategy (or Digital):** It is a catalyst for the overall transparency and interoperability of the cities' operations. Data Strategies are defining the digital capabilities of cities for better harnessing and using of data to empower responsive mobility solutions. **EU Covenant of Mayors for Climate and Energy:** It was launched in 2008 aiming to engage local governments to tackle climate change. The cities are committing to implement a series of actions in transport, construction and other sectors in order to achieve specific targets. **MaaS Strategy:** It is paving the way for seamless multimodality which has become vital for urban and sub-urban areas. Mobility as a Service (MaaS)

	strategy crates the crucial pathway towards the integration of various transport systems in order to be accessible on demand.
Data Source:	Ertico ITS Europe City Moonshot and Covenant of Mayors (CoM , 2021)
Data Availability:	Data were available for 100% of cities at the city level.

Category:	**Governance**
Indicators	Score in European Quality of Government Index 2017
Structure:	Single Indicator
Rationale:	In the last decades, the quality of governance concerned urban planners and policy makers in the mobility sector. Effective governance is a prerequisite for the endurance of transport improvements and for the sustainability of transformative innovation. The Quality of Government Institute of Gothenburg University developed the regional European Quality of Government Index (EQI) to assess the multidimensional nature of institutional governance. EQI considers the corruption, the impartiality of public services by capturing the citizens experiences and believes at a sub-national level.
Data Source:	EC Regional Policy (Charron, et al., 2019). The EQI dataset was available in csv form.
Data Availability:	Data were available for 99% of the cities at the NUTS1 or NUTS2 level depending on the country.
Imputation Technique:	Data were missing for the Norwegian cities due to EQI's focus on EU countries, so data from OECD (OECD, 2019) regarding citizens' satisfaction with public services and institutions was used. According to the assessment of the OECD member countries Norway is a top performer in terms of quality of governance therefore the highest value was assigned to each Norwegian city.

5.1.3 Mobility Performance

Category:	**Accessibility**
Indicators	Percentage of Population with public transport stop within 500 m of walking distance
Structure:	Single Indicator
Rationale:	Modal change is one of the most important factors for reshaping urban mobility. The urban profiles of cities can be transformed by reducing the journeys of private cars over using public transport. One critical factor towards the necessary citizen's behavioral change is the accessibility of the given public transport services as many mobility intervention case studies have shown (PUM, 2019). More importantly the public transport accessibility reflects on the social inclusion of each city.
Data Source:	The European Commission published an assessment of the urban accessibility available in both pdf and xls form (Poelman, et al., 2020).

Data Availability:	Data were available for 86 % of cities at city level.
Imputation Technique:	The missing data were estimated using ML based on a special database created for this purpose.

Category:	**Congestion**
Indicators	TomTom Traffic Index
Structure:	Single Indicator
Rationale:	Congestion is a major mobility issue cities are faced with. It is one of the factors that hinder the advantages of living in high density agglomerations. Congestion has primarily an impact on private transport but it equally affects the public transport either because of the modal shift or as a result of the fact that both public and private transport modes are sharing the same roads. Congestion, increases the travel time of the transport users while it is reducing the timely access to destinations. The TomTom Index estimates the extra travel time worldwide and it is the most dominant metric of congestion in the literature.
Data Source:	The dataset of TomTom Traffic Index (TomTom, 2021) is available online and it can be downloaded in pdf form. Therefore, the data from TomTom Index 2019 was manually extracted.
Data Availability:	Data was available for 81% of cities at the city level and for 2019 as the reference year.
Imputation Technique:	The missing data was estimated using machine learning based on a special database created for this purpose.

Category:	**Safety**
Indicators	People killed in road accidents per 10,000 people
Structure:	Single Indicator
Rationale:	The annual road fatality rate captures a series of different factors such as improvement in vehicle technologies and safety standards. Moreover, reflects on the quality of road' infrastructure and the efficiency of traffic management systems. Therefore, it is affected by changes in legislation on speed limits, drinking and driving in general (EEA, 2021a).
Data Source:	The regions and cities statics of Eurostat (Eurostat, 2021a). The datasets for Cities and Regions are available in xlsx form.
Data Availability:	Data was available for 62% of cities at the city level and for 2019 as the reference year.
Imputation Technique:	Data was available for the cities in past years so together with the available corresponding data at NUTS2 level the missing values were imputed. Also for cities for United Kingdom the data from 2018 was used because it was the most updated available data. Also for the Greek cities in order to have up to date data a different source was used (ELSTAT, 2020).

5.1.4 Mobility Infrastructure

Category:	**Alternative fuels –Renewable Energy**
Indicators	Electricity charging stations per 1,000 people.
Structure:	Single Indicator
Rationale:	The electrification of mobility is the most promising concept in order to abandon traditional fossil fuel-based mobility. The indicator shows the level of readiness for the deployment of e-mobility in the cities and together the maturity of their policies.
Data Source:	European Alternative Fuels Observatory (EAFO, 2021). The data was manually extracted from the interactive map.
Data Availability:	Data was available for 100% of the cities at the city level and it was referring to 2021.

Category:	**Bicycle-Walkability**
Indicators	Length of bicycle network (dedicated cycle paths and lanes) –m per 1,000 capita
Structure:	Single Indicator
Rationale:	Cycling is one of the cleanest and most efficient forms of mobility, well suited for urban distances. Clearly, it is a climate friendly option which does not contribute to air pollution but instead, it improves the health of citizens and contributes to the reduction of both congestion and noise pollution. Also, it is one of the vital components towards sharing and MaaS. Cycle lanes are usually dedicated on-road spaces separated from the other road traffic by either marking or physical barriers.
Data Source:	The regions and cities statistics of Eurostat (Eurostat, 2021a) and Ertico ITS Europe Database. The datasets for Cities and Regions are available in xlsx form.
Data Availability:	Data was available for 100% of cities and data from the final 5 years was considered valid.

Category:	**Mobility as a Service**
Indicators	Deployment of MaaS services
Structure:	Single Indicator The values were calculated using Scale 1 (Section 4.2.6)
Rationale:	The indicator investigates the integration of the transport system, at both a physical and digital level. The essence of the indicator is the provision of a unified service to the citizens reflecting on the quality of typical journey through an advanced stakeholder management on behalf of the city.
Data Source:	Ertico ITS Europe-City Moonshot
Data Availability:	Data was available for 100% of cities

5.1.5 Innovation

Category:	Readiness- Competitiveness
Indicators	EU Regional Innovation Scoreboard 2019
Structure:	Single Indicator
Rationale:	Innovation is the driving force for deploying new smart mobility solutions in the cities which can increase the overall urban sustainability (Lopez-CarreiroI. & Monzon, 2010). The urban capabilities in terms of citizens' education level and business development and the level of R&D investment of the cities are the core elements for innovation readiness and development. In that context, the EC published the Regional Innovation Scoreboard to enable a detailed comparative assessment for analyzing difference between regions (NUTS2) based on the 17 indicators dedicated to Education, Research, SMEs and R&D.
Data Source:	The database (xlsx) of Regional Innovation Scoreboard issued by European Commission (Hollanders, et al., 2019a)
Data Availability:	Data was available for 95% of cities at the Regional –NUTS2 level.
Imputation Technique:	Small counties such as Latvia and Estonia are not included in the Regional Innovation Scoreboard as in those the NUTS1-NUTS2 is considered identical. For that reason, the countries' score in the European Innovation Scoreboard (Hollanders, et al., 2019b) was used by first reverse engineering the correction factors mentioned in the RIS methodology.

Category:	Investment
Indicators:	Horizon 2020 funding at NUTS3 level and Horizon 2020 transport funding at NUTS2 level (€ per capita)
Structure:	Unweighted average of the normalised scores.
Rationale:	Horizon 2020 has been the most impactful EU Research and Innovation programme until now with nearly €80 billions of funding available over 7 years (2014 to 2020) (European Commission , 2021c) excluding the private investment for the projects which was additional. It promises more state-of-the-art technology implementations, innovations and world-firsts by taking great ideas from research to business world. The focus of Horizon 2020 was very broad covering multiple areas, and transport was one of them. In general, the performance of cities in finding funding from the Horizon 2020 is an extremely useful indicator because it captures as well the level of competitiveness and readiness of the cities to adapt and develop new technologies. Moreover, the compliance of the cities to EU targets for urban innovation is also implied (Garau, et al., 2016).
Data Source:	The EC Horizon 2020 Dashboard (European Commission, 2021d) and Dashboard of the Transport Research and Innovation Monitoring and Information System (TRIMIS) (European Commission, 2021b). The data from both dashboards was extracted manually and it was further analysed using the corresponding numbers of populations in order to create the € per capita values.

Data Availability:	Data was available for 100% of cities at NUTS3 and NUTS2 level for 2021 and 2018 respectively.

5.1.6 Engagement

Category:	Citizens participation in decision making
Indicators	City Moonshot Engagement Index
Structure:	Unweighted average of the normalised scores – The scores were calculated using Scale 2 (Section 4.2.6)
Rationale:	Participatory governance is a vital socio- technological element towards a smart and resilient city. Nowadays, cities services and policies should be to a large extent citizen - responsive in order to maximize the sustainability and the impact of the planned mobility interventions. The last years the decision-making instruments managing smart mobility are focusing on establishing consensus between stakeholders and improving the efficiency of solutions by understanding the needs and priorities of the end-users. In the past, it was viewed as a costly, time consuming and complex procedure and there was lack of skills and capacity but lately the ICT technologies promoted the use of online tools, e-workshops and surveys to involve citizens. The indicator assesses the frequency of the usage of such tools for Public surveys, Public consultation, Complaints handling and Mass media campaigns.
Data Source:	Ertico ITS Europe-City Moonshot
Data Availability:	Data was available for 100% of cities.

5.2 New datasets & Machine learning predictions

This system of indicators along with the cities' corresponding NUTS2, NUTS3, Country, GDP per capita and weighted density resulted in the addition of more than 1,100 cells to the City Moonshot Database. For that purpose, raw data from almost 17 different sources in various formats were aggregated to create the extended database. Furthermore, four new supplementary databases were created in order to perform data mining calculations for estimating missing data according to the methodology explained in the previous section 4.2.4. In more detail, the four indicators for which for them none of other imputation technique was valid for estimating their missing values due to their nature were the base for creating the following datasets. In total, the datasets have more than 25,000 cells.

- **GHG (or CO2e) emissions-Climate change:** A dataset of 389 rows of cities for various reference years in the range of 2008-2017. The Joint Research Centre (JRC) of the EC published a dataset of the CHG emission of 6.200 European cities/ municipalities (rural-urban) but just at their adhesion year (JRC, 2020). Therefore, most of the emissions referenced year in the dataset is ranging between 1990-2010. For that reason, the dataset was carefully prepossessed and manually enriched with progress of cities available on the online emissions inventory of the Covenant of Mayors provides (CoM , 2021). Furthermore, some additional indicators included in the original dataset were used, such as Mitigation target and adhesion year. In Figure 12c a sample of the dataset is presented.

- **TomTom Traffic Index –Congestion:** A dataset of 611 rows, a collection of almost 200 cities for the years 2017, 2018 and 2019. TomTom provides in pdf form the raw data for the extra travel time in European cities (TomTom, 2021). In Figure 12a a sample of the dataset is highlighted.
- **Percentage of population exposed in L_{den} higher than 55 dB- Noise:** A dataset of 436 rows including data for some 270 different cities for the years 2012 and 2017. The raw data are provided by the European Environment Agency available in xls form (EEA, 2019b). In Figure 12b a sample of the dataset is presented.
- **Percentage of Population with public transport stop within 500 m walking distance-Accessibility:** A dataset of 694 cities. The raw data were available in xls form (Poelman, et al., 2020) and was including more cites but the available raw data for the other indicators limited the final observations. In Figure 12d a sample of the dataset is shown.

ID	Country	NUTS2	City	Congestion level	Date	GDP	Populaion	Density	Ownerhip	Coastal	Climate Zone
1	Belgium	Région de Bru	Bruxelles / Bru	0,37	2017	68500	2651879	12008	0,52	0	zone 4
2	Belgium	Prov. Vlaams-I	Leuven	0,21	2017	36500	223463	4610	0,64	0	zone 4
3	Belgium	Prov. Liège	Liège	0,20	2017	30600	750318	4585	0,47	0	zone 4
4	Belgium	Prov. Namur	Namur	0,19	2017	30000	222180	3745	0,49	0	zone 4
5	Belgium	Prov. Oost-Vla	Gent	0,19	2017	47500	603140	5713	0,46	2	zone 4
6	Belgium	Prov. West-Vl	Brugge	0,16	2017	39000	227835	3443	0,5	1	zone 4
7	Belgium	Prov. Hainaut	Charleroi	0,14	2017	26700	488749	3245	0,47	0	zone 4
8	Belgium	Région de Bru	Bruxelles / Bru	0,37	2018	69200	2667824	12008	0,53	0	zone 4
9	Belgium	Prov. Antwerp	Antwerpen	0,31	2018	49400	1111919	8319	0,5	1	zone 4
10	Belgium	Prov. Namur	Namur	0,29	2018	30900	223018	3745	0,49	0	zone 4

(a)

ID	Country	City	Population	WDendity	Year	Lden	coastal	Ownership	GDP	Intensity
1	Austria	Graz	250653,00	5960,00	2012,00	0,50	0,00	0,60	44800,00	87,71
2	Austria	Innsbruck	124951,00	7219,00	2012,00	0,30	0,00	0,57	39500,00	87,71
3	Austria	Linz	212973,00	5052,00	2012,00	0,50	0,00	0,57	38100,00	87,71
4	Austria	Salzburg	149201,00	4903,00	2012,00	0,30	0,00	0,57	45000,00	87,71
5	Austria	Wien	1741333,00	13030,00	2012,00	0,50	0,00	0,50	47500,00	87,71
6	Belgium	Antwerpen	483353,00	8319,00	2012,00	0,90	1,00	0,47	43000,00	82,29
7	Belgium	Brugge	117313,00	3443,00	2012,00	0,80	1,00	0,48	35100,00	82,29
8	Belgium	Gent	237250,00	5713,00	2012,00	0,80	2,00	0,46	40200,00	82,29
9	Belgium	Charleroi	203464,00	3245,00	2012,00	0,80	0,00	0,45	25000,00	82,29
10	Belgium	Liège	194715,00	4585,00	2012,00	0,80	0,00	0,46	27400,00	82,29

(b)

ID	City	Country	Mitigation target	Adhesion Year	Year	CO2e capita	Population	GDP	Intensity	Coastal	Urban	Private	Climate	W.Density
1	Daugavpils	Latvia	0	2016	2010	0,5	95467	4600	59	0	2	0,31	zone 5	5423
2	Daugavpils	Latvia	0	2016	2016	0,6	90000	9600	48	0	2	0,34	zone 5	5423
3	Gent	Belgium	20	2009	2015	2,1	261344	26500	77	2	2	0,48	zone 4	5713
4	Burgas	Bulgaria	25	2008	2015	0,3	209202	5300	47	1	2	0,38	zone 3	12971
5	Dobrich	Bulgaria	25	2008	2014	0,3	90000	6200	48	1	2	0,46	zone 3	8425
6	Dobrich	Bulgaria	25	2008	2016	0,2	107486	6700	47	1	2	0,44	zone 3	8425
7	Strovolos	Cyprus	20	2009	2014	1,7	67904	20400	74	1	2	0,56	zone 1	3731
8	Bremen	Germany	40	2008	2012	2,3	550000	43600	68	1	1	0,41	zone 4	5778
9	Hamburg	Germany	40	2008	2015	2,5	1787000	32800	63	1	1	0,42	zone 4	6598
10	München	Germany	47	2009	2012	0,7	1356000	108700	68	0	1	0,54	zone 3	7929

(c)

ID	Country	City	Population	W.Density	Population with public transport stop within 500 m	GDP	Private Car Ownerhip	Coastal	Climate Zone
1	Belgium	uxelles / Bruss	1233487	12008	97,0	70800,0	0,5	0,0	zone 4
2	Belgium	Antwerpen	553789	8319	95,0	50500,0	0,5	1,0	zone 4
3	Belgium	Brugge	92630	3443	96,0	41400,0	0,5	1,0	zone 4
4	Belgium	Charleroi	204143	3245	94,0	28000,0	0,5	0,0	zone 4
5	Belgium	Gent	216134	5713	96,0	49300,0	0,5	2,0	zone 4
6	Belgium	apellen / Ekere	52845	2929	77,0	50500,0	0,51	1,0	zone 4
7	Belgium	La Louvière	55562	3121	94,0	21300,0	0,49	0,0	zone 4
8	Belgium	Leuven	70414	4610	96,0	38400,0	0,7	0,0	zone 4
9	Belgium	Liège	361543	4585	94,0	32600,0	0,5	0,0	zone 4
10	Belgium	Mechelen	65993	4648	93,0	46500,0	0,51	2,0	zone 4

(d)

Figure 12: Datasets for Mining (a) Congestion, (b) Noise, (c) CHG emissions, (d) Accessibility

The performance of the auto ML calculations based on the above datasets are illustrated in Table 6. The best techniques for each case as a result of the AutoML process are included in the Table. Interestingly, even though the datasets are similar the best technique varies implying the different correlation between the indicators. Also, the accuracy is presented as it is the selection criterion which measures the percentage of correct predictions. Clearly the accuracy is low in all cases, but this is reflecting on both the quality of data but more importantly on the size of the datasets.

ML requires millions of observations for high learning optimization. Nonetheless, the assessment of the performance of these techniques was further enhanced by using an automated feature of RapidMiner (Rapidminer, 2021). The reason was that the KNIME does not provide information regarding the relative error of the results which is useful for further evaluating the quality of the predictions since the aim is not to perfectly predict the missing values but rather to have a close approximation. Hence, the relative error, meaning the ratio of absolute error to actual measurement is higher in the case of climate change and the noise (24%). Apart of the aforementioned reasons was explained before this depends on the data availability of other indicators which can support better predictions. Overall, the quality of the results is considered satisfactory in the scope of clustering the cities. Mainly because the aim is to group the cities and not to rank them or rate them so even the values with relative error of 24% are providing sufficient information for the clustering. Although, there is clearly room for improvement and some solutions are presented in the chapter 6.

Table 6: Quality of prediction - Data Mining Performance

Missing Data	Best Technique	Accuracy	Relative Error
Accessibility	Decision Tree	23%	4%
Climate change	XGBoost Trees	16%	24%
Noise	Generalized Linear Model (H2O)	32%	24%
Congestion	XGBoost Trees	23%	9%

5.3 Dashboard

For dynamically analyzing the data and offering the ability to each user- decision maker to create custom-made analysis, an analytical exploratory dashboard was created using PowerBI. The Dashboard (decision making tool) aims to provide a sense of data at a glance and consists of three visible thematic pages and two supportive hidden pages. The three thematic pages are:

- **Environment Page:** includes all the indicators (including the sub-indicators of Air Quality) of the sector of Environment.
- **Mobility Page:** includes all the indicators of the sectors of Performance and Infrastructure.
- **Innovation Page:** includes all the indicators of the sectors of Innovation, Governance-Strategy together with the CHG emissions, Electric stations per capita, congestion and Engagement to enable a cross-sectoral analysis.

The visible pages have the same layout, and use the same widgets to avoid unnecessary complexity and increase the usability. Undoubtedly, the display of the graphs is crucial and so the structure is based on typical flow of reading. Moreover, the layout was divided in two main parts data visualization and data manipulation using different background colors.

Data manipulation: This section is the core of the dashboard because it allows the users to assess what matters to them by customizing the layout. Hence, the left part of the layout which is expected to receive more attention is dedicated to data filtering. Several slicers are used for muti-criteria configuration. The slicers narrow the portion of the data affecting all the other visualizations accordingly. Thus, the basic slicers are the Clusters of Cities, Climate zone, the Countries, the GDP per capita and Weighted Density. The reason for adding the GDP and Weighted Density is that their values can be easily estimated for each city, they are insightful explanatory factors and as it will be explained below, they have high correlation with many indicators. Therefore, they are useful for the user in order to configure the page according to those parameters. In addition, depending on the context of each thematic page, slicers of indicators relevant to the City Moonshot framework were included to support in depth customization.

Data visualization: The presentation of the data is performed with many graphs located at the right part of the layout (please see Figure 13). The top left part is dedicated to present a map of the cities and their population. It is worth mentioning that the presented population refers to the public authority that was interviewed, either city or municipality during the City Moonshot survey. Next, a correlation matrix was selected to help the user quickly evaluate the relationships in the data. The second row contains a word cloud for illustrating qualitative answers of cities in the most interesting City Moonshot questions. Also, at the bottom right corner is located a scatter chart highlighting data the distribution analysis of the indicators with the higher relationship in each thematic page.

Figure 13: Dashboard Layout

5.3.1 Correlation analysis

The correlation matrix in each page has been extremely insightful for identifying the mobility behaviour of the different clusters. To elaborate, the matrix shows the Pearson correlation value between attributes ranging within the spectrum of [-1,1] between two attributes. If the value is positive that implies that the two indicators are increasing or decreasing together. On the other hand, if the value is negative when one variable increases the other one decreases and vice versa. In the case that the correlation value between indicators is either -1,1 that means a strong relationship and that they actually are correlated. Moreover, the correlation matric provides a colour coding scale in order to support more efficient visualization effects. The dark blue displays strong positive correlation (1), dark red strong negative correlation (-1), no correlation is white (0) and all the other values have lighter shades of the above colours depending on the correlation value.

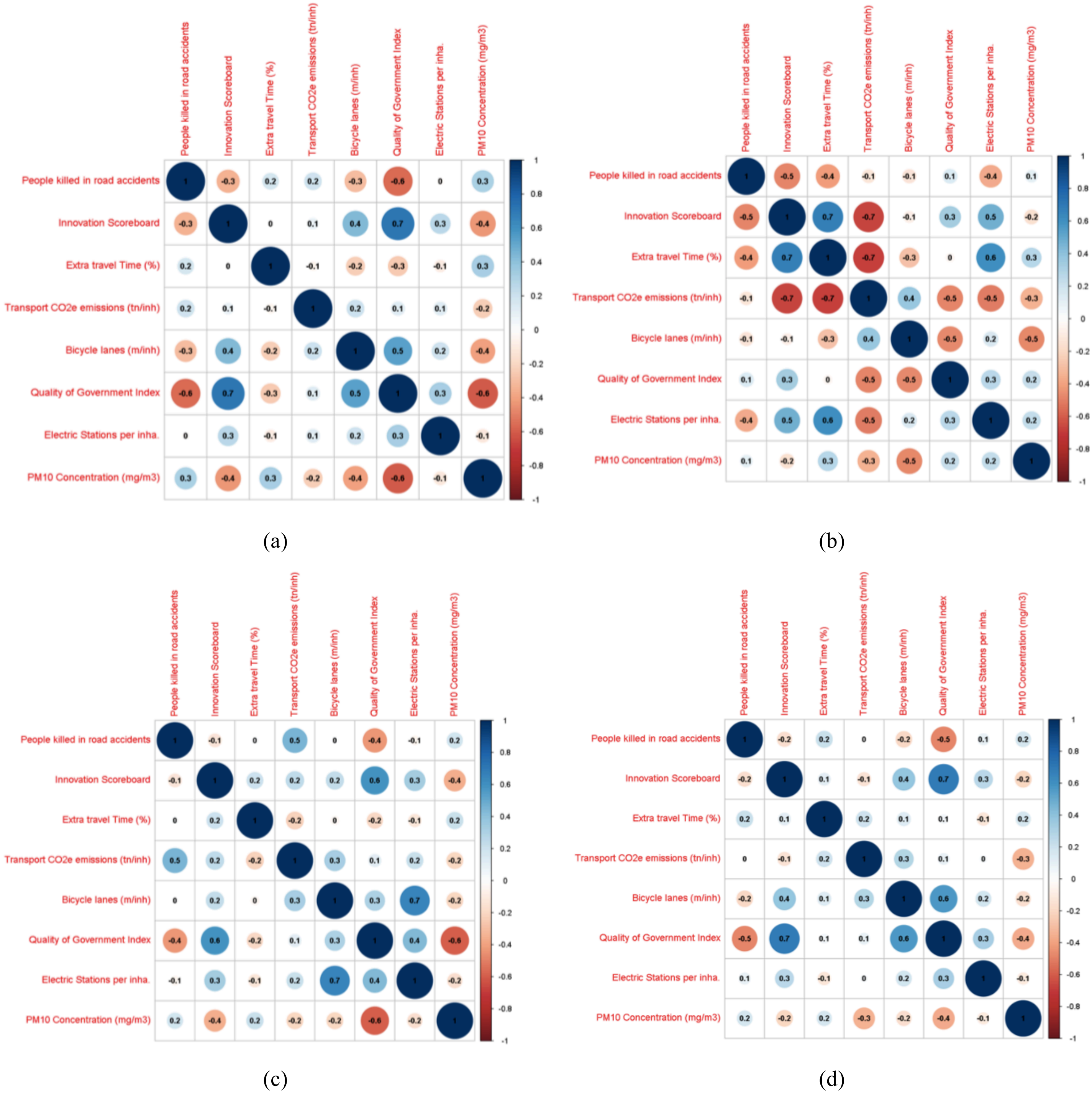


Figure 14: Correlation analysis (a) General, (b) Cluster 1, (c) Cluster 2, (d) Cluster 3

Figure 14 illustrates the corresponding correlation matrixes for each cluster and the general correlation analysis for all cities. Almost every single relationship between indicators is different among the clusters revealing the importance of assessing each cluster separately. Nevertheless, due to the small sample of cities, it is not yet valid to deduce conclusions with great certainty at the cluster level. In some cases, some paradoxical correlations are observed, such as the positive (0.4) correlation between CO2e emissions per capita and bicycle lanes in cluster 1. The reason for this phenomenon is purely statistical as the values for these cities are actually very close and to some extent, it might reflect on the data quality itself. In contrast, there are also some very meaningful observations in the cluster correlation analysis.

One of the most interesting is the strong positive relationship (+0.7) between bicycle lanes (m/inh) and electric stations per capita in cluster 2. This correlation implies that the interdependency of those urban solutions for these cities towards their decarbonization. Also, it is an indication that they share common policies and that they are rapidly developing their infrastructure because for the other clusters the relationship is just slightly positive (+0.2). Also, the strong negative (-0.7) correlation between CO2e emissions per capita and Regional Innovation Scoreboard in cluster 1 illustrates the impact of innovation towards changing the climate.

In general, the correlation review for all cities is more suitable for extracting conclusions based on strong correlations to identify some common patterns. However, it is important to remember the different reference areas for each indicator and that the sample is still relatively small. Nevertheless, these findings are a solid base for the users to explore the data and try to find which cities are not following that trend within the same cluster identifying possible "overachievers" and cluster leaders. Hence, the following points are presenting the most insightful relationships in the proposed system of indicators adding the GDP per capita, weighted density:

- Regional Innovation Scoreboard and Quality of Governance have a strong positive correlation +0.7. This correlation is the highest in this study and proves the complementary nature of this concepts but both indicators are referring to the NUTS2 level so further analysis in needed.
- Green public areas and PM_{10} concentrations have negative correlation -0.5. One of the reasons why public green is a vital element in the urban fabric is its ability to mitigate air pollution thus this correlation is reasonable. However, if the acceptable limit of 10 m2 per capita issued by World Health Organization (WHO) is included in then the analysis the correlation might be even higher.
- Noise has a strong positive correlation +0.6 with both PM_{10} concentration and weighted density population. This comes as no surprise as they both have traffic as their source but it is striking that the NO_2 concertation has not equally high correlation implying that PM_{10} concentration and noise are affected by other sources too.
- Quality of Governance has a strong negative correlation – 0.6 with both PM_{10} concentration and people killed in road accidents. The quality of governance has a strong impact on the most aspects of the urban metabolism as it is proven by this correlation underlying the importance of smart cities.

- GDP per capita has strong positive relationship with Quality of Governance and Regional Innovation Scoreboard with +0.6 and +0.5 respectively. The most interesting fact from these relationships is that the correlation between those indicators is realized even in the NUTS3 level
- Weighted Density has strong positive correlation with the accessibility (+0.5) and negative correlation with Public Green per capita, GDP per capita, Quality of Governance and Bicycle lanes. In principle, high density is core concept for cities. High densities enable cities to operate more efficiently mainly due to economy of scale. Densely populated cities traditionally reduce the cost of moving goods, and people. They also provide proximity to large public goods and specialized services. However, if the city does not follow a viable urban plan, very high density becomes a draw back. It is worth mentioning though that especially in the south Europe this phenomenon occurs mainly due to historical reasons that is why the correlation of governance is not very high.
- Population it has minor correlations with other indicators. This result really complements the purpose of this study as it emphasizes that the behavior of the modern cities is not directly dependent on the population.

5.3.2 Clustering

This section provides information regarding the characterization of the cities according to the clusters formed. Clustering was performed using the k-means algorithm provided by the Tables widget in Power BI. The "means" stands for the centroid of the cluster which is eventually, after some iterations, the mean of all the points in the cluster. On the other hand, "k" represents the number of arbitrary points used to initiate the clustering, and the number of clusters to be formed. Importantly, it does not allow any uncertainty in terms of cluster membership, thus it assigns each data point to a unique cluster. Also, k-means is suitable for generating clusters of continuous numerical attributes. Therefore, Power BI Tables automatically identified the number of clusters, and it assigned cities, accordingly, based on the 16 indicators.

It is worth mentioning that due to the high correlation of Quality of Governance and Regional Innovation, the unweighted average of Governance and Strategy was used instead. The distribution of the cities to each cluster in shown in Table 7, and the main characteristics of each cluster are explained below, based on the average score of each cluster in the different sectors as shown in Table 8. Although the limitations of composite indicators were thoroughly explained, the use of composite indicators of sectors is extremely beneficial for the comparison of the clusters as it simplifies the multicriteria analysis.

Cluster 1: The cities of this group are all cities from the Baltic Countries. All of them have high performance in Innovation and Environment and excellent score in the other sectors. Oslo and

Stockholm are widely considered as the two of the most sustainable cities in the world and both have been awarded as European Green Capitals (European Commission, 2021a). Also, Gothenburg was the most sustainable destination in the world for four consecutive years (Goteborg , 2021). Those, distinctions are highly related to the cities' performance on mobility and innovation.

Table 7: City Clusters

Cluster	Cities
Cluster 1	Bergen, Gothenburg, Malmo, Oslo, Stavanger, Stockholm, Tampere, Umea
Cluster 2	Barcelona, Bilbao, Brno, Brussels, Coventry, Glasgow, Grenoble, Guimaraes, Helmond, Kingston upon Hull, Larissa, Las Palmas, Logrono, London, Pamplona, Paris, Praha, Riga, Roma, Rotterdam, Thessaloniki, Torino, Trieste, Verona, Warsaw, Zagreb
Cluster 3	Amsterdam, Antwerp, Athens, Birmingham, Bremen Cambridge, Essen, Graz, Hamburg, Copenhagen, Leuven, Limerick, Lisbon, Madrid, Manchester, Milan, Milton Keynes, Munich, Oxford, Strasbourg, Tallinn, Toulouse, Trikala

Cluster 2: First of all, it is clear that there is no geographical correlation. This cluster is characterized by slightly lower performance in most sectors compared with Cluster 3. Specifically, the cities in Cluster 2 have by far the lowest score in terms of Innovation and Infrastructure on average. This may be explained by the lowest GDP per capita, which is a highly correlated factor. By contrast, some of the cities in this Cluster, such as Barcelona and Paris are undoubtedly leaders in Innovation, as they both have been awarded as European Capital of Innovation (European Innovation Council, 2021). In addition, London is one of the world's leading economic powers and it was in 2018 announced by Arcadis to be the most Sustainable city in the world (Arcadis, 2018). Overall, it is important to notice that the Score in Performance is almost the same with the other Clusters, implying that those cities probably depend more on other forms of mobility rather that those expressed in the Infrastructure sector.

Table 8: Average city profile of cities per cluster and sector

Cluster	Green	Performance	Infrastructure	Innovation	Engagement
C1	69	69	32	42	38
C2	39	63	8	29	36
C3	42	64	41	39	41
Cluster	GDP	Weighed Density	Population	Strategy- Governance	
C1	55055	5297	403,393	48	
C2	48120	5917	555,338	31	
C3	51232	7023	1,734,694	53	

Cluster 3: According to the results presented in Table 8 these cities were generally the densest and most populated ones. It includes cities that were European Green Capitals (Lisbon, Copenhagen, Hamburg, Essen) and European Innovation Capitals (Athens, Amsterdam, Leuven). Moreover, these cities have the higher average performance in Infrastructure, Strategy - Governance and Engagement, factors connected with the population and density. Likewise, cities in cluster 2 are not correlated geographically and belong to different climate zones. It is interesting to observe that cities in the same countries are split across the three clusters.

5.3.3 Scatter charts

The responsive scatter charts were selected to support users' in-depth analysis of the most insightful indicators. In each page, the indicators with the higher correlation to each other were presented in a 2D graph, with the size of each data point representing an additional third indicator: dimension. In addition, a fourth dimension: the cluster of each city is encoded by a distinct color to show its membership. Also, the form of the scatter points is an indication of the correlation. Hence, by applying different configurations the users can identify other cities with similar characteristics (in the same cluster) which have better performance values, in order to see how they can improve their activities. On a general note, the clusters in all three graphs follow the same trend. Most of the cities in cluster 1 (light blue) have better performance than cities in cluster 2 (orange) being really close to them (see Figures 15, 16, 17).

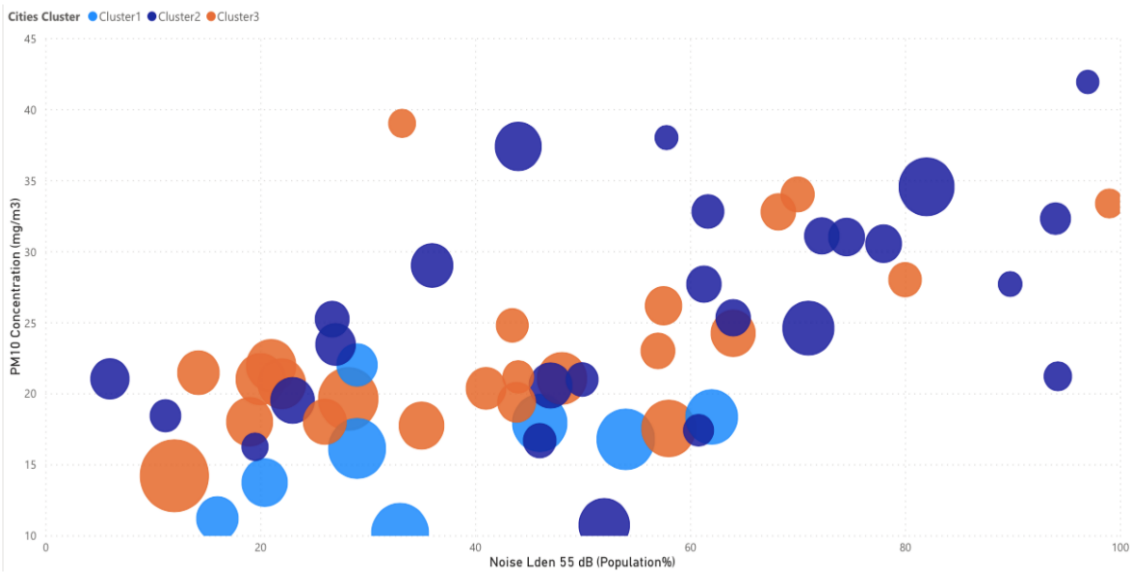


Figure 15: Environmental Factors: PM_{10} concertation and Noise (Bubble Size: Green per Cap.)

Overall, the relationships discussed in the chapter 4 are confirmed graphically. To elaborate, the correlation between Quality of Governance and Regional Innovation Scoreboard is the highest, thus the form of the data points is dense, following a positive linear pattern (see Figure 17). In the same manner, the first graph (see Figure 15) including Noise and PM_{10} concentration (+0.6 correlation) is more consistent than the second graph (see Figure 16) where the indicators have low correlation (-0.3) and thus the points are more scattered. Lastly, even though the correlations of the indicators used for the size of the data points with other indicators is relatively low, and so is hard to draw conclusion, in Figure 16 the difference between cluster 2 and cluster 3 is visible. Furthermore, since the indicator is MaaS projects which is included in the Infrastructure sector, the low performance of cluster 3 that was previously discussed, is efficiently presented in the following graph.

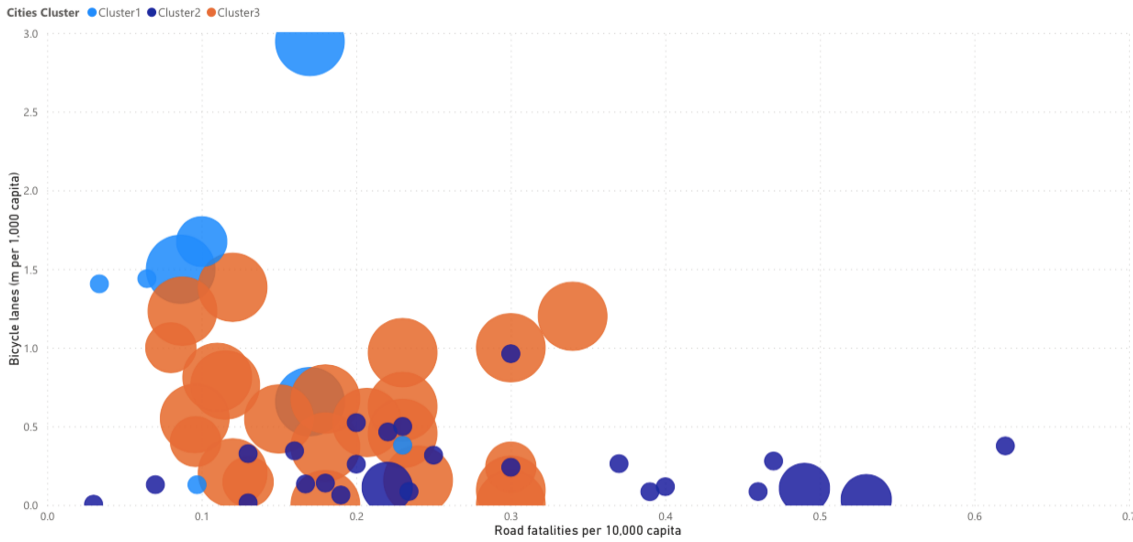


Figure 16: Mobility Factors: Bicycle lanes and Road fatalities (Bubble size: MaaS)

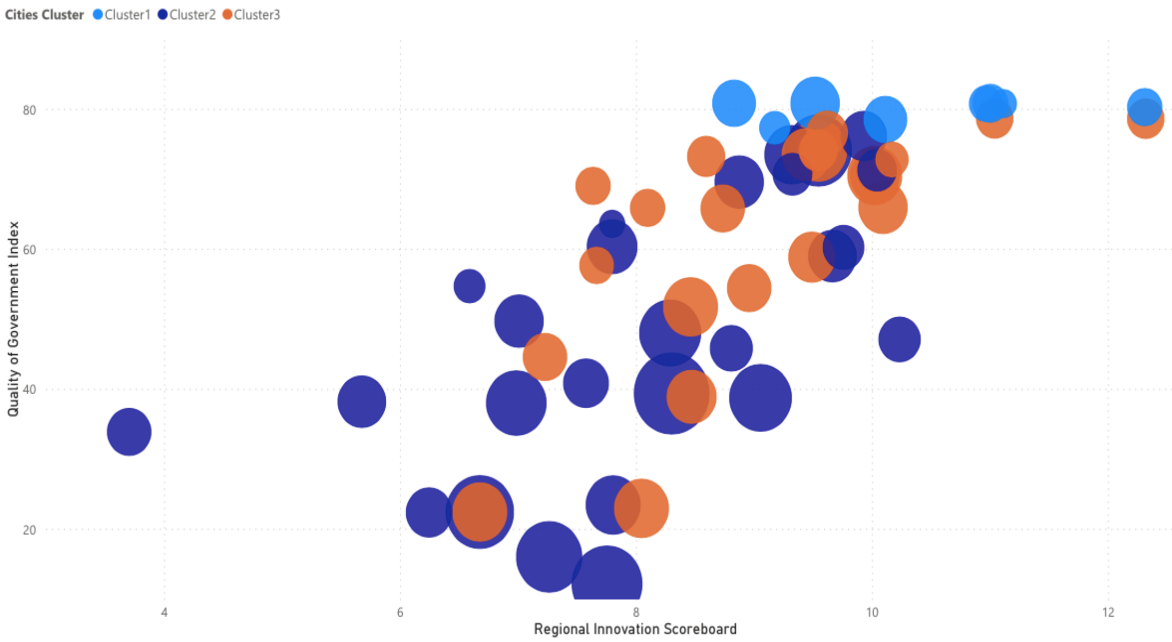
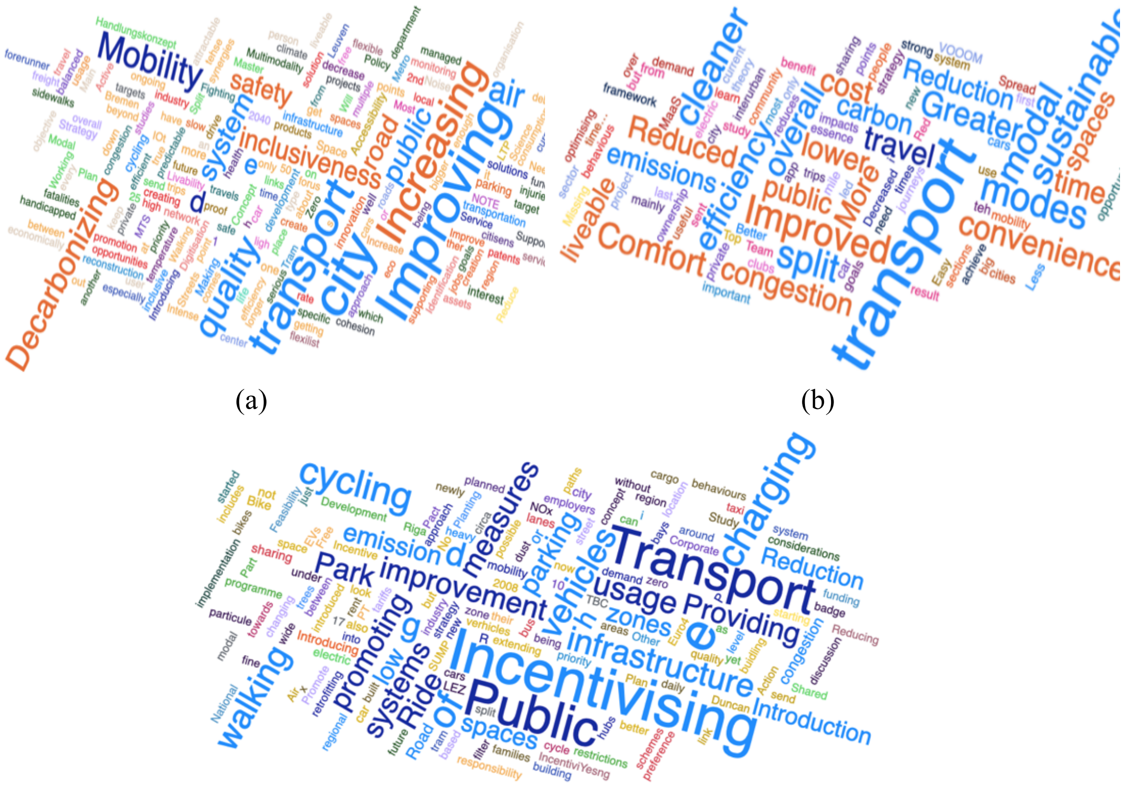


Figure 17: Innovation Factors: Quality of Governance and Regional Innovation Scoreboard (Bubble size: Road Fatalities)

5.3.4 Word clouds



(a) (b)

(c)

Figure 18: Word Clouds (a) Mobility objectives, (b) MaaS benefits, (c) Behavioural change objectives

Word clouds are attractive visualizations of data trends. The size of words is proportional to their frequency of occurrence. The word cloud widget was used for presenting the information of three important City Moonshot questions for two reasons. Firstly, the data of City Moonshot are not supposed to be published earlier than October 2021, when the results of the survey will be presented in the World ITS Conference. Therefore, word clouds allow the revelation of basic trends without further comparisons. Secondly, they are simple to understand, and they are transforming the qualitative data into a structured brainstorming session for each context, as they are responsive too. Nevertheless, this ability to recompose the source data needs the removal of common joiners like “and”, “or” etc. In that sense, meaningful context was extracted from the word clouds referring to all cities (Figure 18):

- Environment page: The words come from answers to the question: “Does your city encourage behavioral change to help achieving environmental improvements?”. Incentivizing the citizens to adopt more cycling and walking by also providing charging infrastructure efficient measures is the main strategy of the cities.

- Mobility page: The answers respond to the question: “What are the main objectives for your city when it comes to transport and mobility?”. Decarbonizing public transport and city mobility in general seems the main trend along with inclusiveness and safety.
- Innovation page: The source question is: “What in your view are the biggest benefits of MaaS for your city?”. Increasing the comfort and livability through reducing congestion and improving the modal split towards a cleaner environment is the cities’ expectation from MaaS.

6 Conclusions and future work

The final chapter is dedicated to discussing the challenges this research faced as well as the results this research produced. Also, recommendations and considerations for future steps towards further development of this work are proposed.

6.1 Conclusions

This thesis contributed to the establishment of the conceptual framework of smart cities highlighting one critical conceptual interdependency of smartness and sustainability and attempted to reveal that participatory governance is a core element in this. These realizations were necessary for establishing a meaningful system of indicators and concluding that gradually building a general framework for assessing European cities is vital. Furthermore, this study concluded that the concept of Smart Mobility seeks to narrow the gap between technological innovation and its environmental, economic and governance-social dimensions, something that is clearly illustrated in the preliminary results of the City Moonshot Questionnaire, which was used as a basis for this research.

From the results produced in this Thesis, the conclusion can be made that ideally, the governance of cities which is the critical point for achieving smartness, should be based on extensive evaluation and comparisons of cities' performance to set an empirical basis for the cities' targets. In this research, a new set of indicators was introduced in section 5.1 (RT2) focusing on the Innovation of Mobility, allowing a deeper understanding of the cities themselves by adopting a data-driven approach. Significantly, the final group of indicators used for the cities' analysis is multidimensional covering aspects of most of "smart cities" dimensions (Giffinger, et al., 2007) allowing a wider usage. More specifically, all indicators originate from and connect to indicators that resulted from similar studies that were used in this research, thus the proposed system is contributing to the ongoing science-policy dialogue for finding the set of the indicators with the highest consensus in the literature. In that way, initiatives such as the City Moonshot are complementary to other assessment initiatives maximizing the total impact of information received in terms of comparisons and data-driven narratives. For that purpose, it is equally important for future studies to have a clear methodological framework following the recommendations of the EC and OECD (European Commission & OECD, 2008) that provide guide to indicator analysis.

Moreover, from the more technical perspective, the importance of ICT capabilities for cities as enablers for extracting and utilizing information was attempted to be thoroughly presented in this study. The data analysis software used in this research was selected to be PowerBI and KNIME so that they can be easily accessible and usable by city stakeholders and decision-makers offering a variety of seamless and transparent applications in the urban context. In addition, the tool created as part of this research (RT5), i.e. the dashboard described in section 5.3, provides decision makers (public authorities or citizens) with the ability to interact with urban data in a comprehensive way, being aided in acquiring a "societal self-awareness" and an evidence based decision making tool (the dashboard). The fact that decision makers can perform personalized exploratory analysis through this dashboard is considered a foundational element for any decision-making process (Matheus, et al., 2018). The tool allows cities to understand where they need to invest more and which factors are the drivers for growth. Combining ICT with the open data policy of the EC, this thesis succeeds in further supporting the creation of city intelligence.

Methodology-wise, the research faced challenges due to the inconsistent publicly available datasets. In addition, there was some incomparability of data due to the variety of city definitions in the literature and primary (official) documents, as well as contradictions among various studies in the field. Nonetheless, this study in section 5.1 (RT1) demonstrated that the datasets of the EC are rich with potential for further aggregation and improvement using imputation techniques (4.2.4 &4.2.5). In that way, they allow the assessment of hundreds of European cities, including the capitals. Also, the implementation of ML technique described in section 4.2.5 is deemed to be promising for lifting the barriers of data availability and solving the problem of generalization, often found in the literature according to our literature review (RT3). Nevertheless, holistically these challenges are considered to be solvable trough juxtaposition with collective intelligence.

On a more specific note, clustering and correlation analysis results of the cities studied, are insightful and encouraging, paving the way for identifying certain city profiles and behavior (RT4). Consequently, any unexpected correlations between the cities could actually be the starting point for further research. This research could analyze and reveal more insights on how to transform the information into knowledge and wisdom for the city authorities.

This information will not only reflect on city performance, but it will also reflect on the quality of data and the integrity of the methodology in collecting and analyzing them. This certainly highlights some intrinsic weaknesses and opportunities, which should be guiding us towards more efficient collective intelligence. On the other hand, reasonable and valid connections already indicated in this study, such as that Quality of Governance and Innovation can be the solid ground for cities on which to identify cities to learn from and to evaluate the reasons for any possible inconsistency.

Overall, the findings of this study and the challenges regarding the openness, quality and the availability of data highlight the necessity for participatory governance and co-creation to be both core concepts of smart cities. Additionally, this study is contributing to knowledge democratization by offering an alternative way for decision makers to assess and analyse essential information in order to better understanding city needs. Thus, the tool highly supports and promotes better governance toward smart cities. In the researcher's opinion, this thesis is highlighting that smartness does not refer to the city's bricks and mortar, but rather the process of empowering cities and citizens to adopt efficient and sustainable solutions by utilizing ICT.

6.2 Recommendations for future work

The methodological approach adopted in this thesis could be replicated and expanded so that more cities and other factors determining a more elaborate or more restricted definition of smartness in cities can be assessed by future research. Nevertheless, the lack of city data, data fragmentation and quality of data publicly available will hinder efficient analysis, but as the United Nations explained in a similar context: "No index can be better than the data it uses but this is an argument for improving the data not abandoning the index" (United Nations , 1992). Hence, there are several recommendations and ideas that could be beneficial for further development in order to increase the impact of this research:

With regards to the data

- The addition of more data from other cities will enhance the results in terms of clustering and correlation analysis. Notably, as the dataset is currently synchronized with the dashboard, data form additional cities just need to be added so that the dashboard can be automatically updated.
- It would be useful if all datasets in the repository of the EC, JTC and H2020 projects had included the harmonised ID identifier for NUTS2, NUTS3 and cities. With regards to greater cities and municipalities, their harmonised IDs should be structured in a way so that it is easy for a researcher to find the corresponding higher NUTS levels.

With regards to indicators for smartness in cities and data estimation

- The use of Modal Split at city level and the Köppen classification at NUTS3 would be much recommended as a future addition in methodology for characterizing cities and increasing the accuracy of the ML predictions respectively. For example, in the case of Noise missing data prediction, the addition of the PM_{10} concentration in the learning data is considered a solution for further improvement based on the correlation analysis results in 5.3.1.
- The KNIME is an extremely powerful open source software available for modifications in contrast to other commercial software in the market. Thus, the absence of relative error in the AutoML criteria should be a field for further improvement.

With regards to the dashboard as a decision-making tool

- The operation of dashboard could be enriched in the future with more detailed graphs for the City Moonshot answers. On a similar note, if the datasets are systematically monitored, progress through time can be illustrated in the dashboard maximizing its impact.
- The Dashboard could be connected with Microsoft Forms to increase the automation of data collection and to provide instant participation for new cities.

In conclusion, further research can contribute to City intelligence – much sought after by European level decision-bodies such as the EC. For that purpose, the datasets resulted by this research along with the decision-making tool (Dashboard) will be open and available for free use after the ITS World Congress in October 2021.Also, the results, and conclusions of this study will be presented in a Scientific Conference or Journal to increase their visibility towards better dissemination and quality, new knowledge awareness and further research advancement in the field.

Bibliography

Akande, A., Cabral, P., Gomes, P. & Casteleyn, S., 2019. The Lisbon ranking for smart sustainable cities in Europe. *Sustainable Cities and Society,* 44(475-487).

Albalate, D. & Fageda, X., 2019. Congestion, Road Safety, and the Effectiveness of Public Policies in Urban Areas. *Sustainability.*

Albino, V., Berardi, U. & Dangelico, R., 2015. Smart Cities: Definitions, Dimensions, Performance, and Initiatives. *Journal of Urban Technology.*

Al-Khouri, A., 2015. *"Smart Government – Circle of Attention".* USA: Xlibris, s.n.

Allio, M. K., 2012. Strategic dashboards: Designing and deploying them to improve implementation.. *Strategy & Leadership,* Volume 40(5), p. 24–31.

Alonso Raposo, M. (. et al., 2019. *The future of road transport - Implications of automated, connected, low-carbon and shared mobility,* Luxembourg: Publications Office of the European Union.

Alonso, A., Monzón, A. & Cascajo, R., 2015. Comparative analysis of passenger transport sustainability in European cities. *Ecological Indicators,,* Volume 45, p. 578–592.

Arcadis, 2018. [Online]
Available at: https://www.arcadis.com/en/news/global/2018/10/european-and-asian-cities-lead-in-arcadis-sustainable-cities-index
[Accessed 22 05 2021].

Avramidou, A. & Tjortjis, C., 2021. *Predicting CO2 Emissions for Buildings Using Regression and Classification.* s.l., Proc. 17thIFIP Int'l Conf. on Artificial Intelligence Applications and Innovations (AIAI 21)..

Böckera, L., Uteng T., P., Liu, C. & Dijstde, M., 2019. Weather and daily mobility in international perspective: A cross-comparison of Dutch, Norwegian and Swedish city regions. *Transportation Research Part D: Transport and Environment,* Volume 77, pp. 491-505.

Bannister, F., 2005. E-government and administrative power. Electronic Government. Volume 2(2), pp. 160- 176..

Barrionuevo, J., Berrone, P. & Ricart, J., 2012. Smart Cities, Sustainable Progress. Opportunities for urban development. *IESE Insight,* Volume 14, pp. 50-57.

Battara, R., Gargiulo, C., Tremiterra, M. & Zucaro, F., 2018. Smart mobility in Italian metropolitan cities: A comparative analysis through indicators and actions. *Sustainable Cities and Society,* Volume 41, pp. 556-567.

Bauer, T. et al., 2006. Selection in the Information Age: The Impact of Privacy Concerns and Computer Experience on Applicant Reactions.. *Journal of Management,* Volume 32(5), pp. 601-621.

Baumgärtner , S. & Quaas , M., 2010. What is sustainability economics?. *Ecol Econ 2010,* 69(445–50).

Beatley, T. & Manning, K., 1997. The Ecology of Place: Planning for Environment, Economy, and Community. *D.C. Island Press.*

Belissent, J., 2010. Getting clever about smart cities: New opportunities require new business models..

Bernardo, M., 2017. *Smart City Governance: From E-Government to Smart Governance.* lSBON: Universidade Aberta, Portugal & CAPP-ISCSP Universidade de Lisboa.

Bessis, N. & Dobre, C., 2014. Big Data and Internet of Things: 169 A Roadmap for Smart Environments. *Studies in Computational Intelligence,* Volume 546.

Bohringer , C. & Jochem, P., 2017. Measuring the immeasurable—A survey of sustain- ability indices.. *Ecological Economics,* Volume 63, pp. 1-8.

Bollier, D., 1998. *How Smart Growth Can Stop Sprawl: A Fledgling Citizen Movement Expands.* Washington: D.C. : Essential Books.

Boob, T., 2015. Transformation of Urban Development in to Smart Cities : The Challenges. *J. Mech.Civ. Eng,* Volume 12, pp. 24-30.

Brown, M., 2007. Understanding e-Government Benefits. An Examination of Leading-Edge Local Governments. The American Review of Public Administration.. Volume 37(2), pp. 178-197.

Caragliu, A. D. B. C. &. N. P., 2011. Smart cities in Europe.. *Journal of Urban Technology,* Volume 18, p. 65–82.

Castelnovo, W., Misuraca , G. & Savoldelli, A., 2016. Smart Cities Governance: The Need for a Holistic Approach to Assessing Urban Participatory Policy Maki. *Social Science Computer Review,* Volume 34(6), pp. 724-739.

Cavanillas, J., Curry, E. & Wahlste, W., 2016. *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe.* s.l.:Springer.

Cebr, 2017. *Urban Mobility Index,* s.l.: Centre for Economics & Business Research Exploration - Qualcomm.

Charron, N., Lapuente, V. & Annoni, P., 2019. Measuring Quality of Government in EU Regions Across Space and Time.. *Papers in Regional Science.*

Chen, F. et al., 2015. Data Mining for the Internet of Things: Literature Review and Challenges. *International Journal of Distributed Sensor Networks,* p. 14.

Christantonis, K. et al., 2020. 'Smart Cities Data Classifi-cation for Electricity Consumption & Traffic Prediction. *Automatics & Software Enginery,* Volume 31(1).

Collins , K. & Ison, R., 2006. *Dare we jump off Arnstein's ladder? Social learning as a new policy paradigm.* s.l., s.n.

CoM , 2021. *Covenant of Mayors for Climate & Energy Europe.* [Online]
Available at: https://www.covenantofmayors.eu
[Accessed 01 05 2021].

Datta, A., 2015. New urban utopias of postcolonial India: 'Entrepreneurial urbanization' in Dholera smart city. *Dialogues in Human Geography,* Volume 5, pp. 3-22.

De Oña, J., Estevez, E. & De Oña, R., 2020. Perception of Public Transport Quality of Service among Regular Private Vehicle Users in Madrid, Spain. *Transportation Research Board,* Volume 2674(2), pp. 213-224.

Deloitte Insights , 2020. *The 2020 Deloitte City Mobility Index,* s.l.: Deloitte Development LLC.

Doe, J., 2020. A sample reference. *IEEE Transactions on Smart Cities and Communities,* 2(3), pp. 2-12.

EAFO, 2021. *European Alternative Fuels Observatory.* [Online]
Available at: https://www.eafo.eu/fuel-map#
[Accessed 23 04 2021].

Edenhofer, O. et al., 2014. *Technical Summary. In: Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* s.l.: Cambridge University Press, Cam- bridge, United Kingdom and New York, NY, USA..

EEA, 2019a. *Energy intensity in Europe.* [Online]
Available at: https://www.eea.europa.eu/data-and-maps/indicators/total-primary-energy-intensity-4/assessment-1
[Accessed 20 04 2021].

EEA, 2019b. *Reported data on noise exposure covered by Directive 2002/49/EC.* [Online]
Available at: https://www.eea.europa.eu/data-and-maps/data/data-on-noise-exposure-7
[Accessed 25 04 2021].

EEA, 2020. *Air quality in Europe — 2019 report,* Copenhagen : EEA Report No 10/2019.

EEA, 2021a. *European Environmental Agency.* [Online]
Available at: https://www.eea.europa.eu
[Accessed 21 04 2021].

EEA, 2021b. *Air quality statistics.* [Online]
Available at: https://www.eea.europa.eu/data-and-maps/dashboards/air-quality-statistics
[Accessed 22 04 2021].

EGCA, 2018. *Green Application Form for the European Green Capital Award 2019.* [Online]
Available at: https://ec.europa.eu/environment/europeangreencapital/wp-content/uploads/2017/06/Indicator_3_Green_Urban_Areas_incorporating_Sustainable_Land.pdf
[Accessed 21 05 2021].

Eger, J., 2009. Smart growth, smart cities, and the crisis at the pump a worldwide phenomenon.. *J. E-Government Policy and Regulation,* Volume 32, pp. 47-53.

EKPAA, 2019. *Greece Environmental Accessement 2018,* Athens: EKPAA.

Elkington, J., 1997. Cannibals with forks: the triple bottom line of the 21st century business.

ELSTAT, 2020. [Online]
Available at: https://www.statistics.gr/en/home/
[Accessed 21 05 2021].

ELTIS, 2020. *ELTIS The Urban Mobility Observatory.* [Online]
Available at: https://www.eltis.org/
[Accessed 20 04 2021].

Ertico ITS Europe, 2018. *Intelligent Transport Systems.* [Online]
Available at: https://ertico.com
[Accessed 21 05 2021].

Escolar, S. V. F. J. et al., 2018. A Multiple-Attribute Decision Making-based approach for smart city rankings design.. *Technological Forecasting and Social Change,* Volume 142, p. 42–55..

European Commission & OECD, 2008. *Handbook on Constructing Composite Indicators: METHODOLOGY AND USER GUIDE,* s.l.: OECD.

European Commission , 2021c. *What is Horizon 2020?.* [Online]
Available at: https://ec.europa.eu/programmes/horizon2020//en/what-horizon-2020
[Accessed 10 04 2021].

European Commission, 2015. *Creating Value through Open Data: Study on the Impact of Re-use of Public Data Resource,* s.l.: European Union.

European Commission, 2016. *Analytical underpinning for a New Skills Agenda for Europe,* Brussels: COMMISSION STAFF WORKING DOCUMENT.

European Commission, 2021a. *European Green Capital.* [Online]
Available at: https://ec.europa.eu/environment/europeangreencapital/
[Accessed 28 05 2021].

European Commission, 2021b. *TRIMIS Dashboard.* [Online]
Available at: https://trimis.ec.europa.eu/dashboard/
[Accessed 25 04 2021].

European Commission, 2021d. [Online]
Available at: https://webgate.ec.europa.eu/dashboard/sense/app/a976d168-2023-41d8-acec-e77640154726/sheet/0c8af38b-b73c-4da2-ba41-73ea34ab7ac4/state/analysis
[Accessed 25 04 2021].

European Innovation Council, 2021. [Online]
Available at: https://eic.ec.europa.eu/eic-funding-opportunities/eic-prizes/european-capital-innovation-awards_en
[Accessed 21 05 2021].

Eurostat, UN, OECD, The World Bank, 2021. *Applying the Degree of Urbanisation: A METHODOLOGICAL MANUAL TO DEFINE CITIES, TOWNS AND RURAL AREAS FOR INTERNATIONAL COMPARISONS,* Luxembourg: Publications Office of the European Union.

Eurostat, 2020. *Statistics on European cities.* [Online]
Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Statistics_on_European_cities
[Accessed 20 04 2021].

Eurostat, 2021a. *Regions and cities.* [Online]
Available at: https://ec.europa.eu/eurostat/web/regions-and-cities
[Accessed 23 04 2021].

Eurostat, 2021b. *Share of energy from renewable sources.* [Online]
Available at: https://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK_DS-1032380_QID_-1960B966_UID_-3F171EB0&layout=TIME,C,X,0;GEO,L,Y,0;NRG_BAL,L,Z,0;UNIT,L,Z,1;INDICATORS,C,Z,2;&zSelection=DS-1032380UNIT,PC;DS-1032380NRG_BAL,REN_TRA;DS-1032380INDICATORS,OBS_FLAG;&rankName1=UNIT_1_2_-1_2&rankName2=NRG-BAL_1_2_-1_2&rankName3=INDICATORS_1_2_-1_2&rankName4=TIME_1_0_0_0&rankName5=GEO_1_2_0_1&sortC=ASC_-1_FIRST&rStp=&cStp=&rDCh=&cDCh=&rDM=true&cDM=true&footnes=false&empty=false&wai=false&time_mode=ROLLING&time_most_recent=true&lang=EN&cfo=%23%23%23%2C%23%23%23.%23%23%23
[Accessed 21 04 2021].

Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P., 1996. From Data Mining to Knowledge Discovery in Databases. *AI Magazine,* Volume 17(3).

Fujitsu, 2014. *Making Secure, Prosperous Society a Reality,* s.l.: Fujitsu.

Gabrys, J., 2014. Programming environments: environmentality and citizen sensing in the smart city.. *Environment and Planning D: Society and Space,* Volume 32, pp. 30-48.

Garau, C., Masala, F. & & Pinna, F., 2016. Cagliari and smart urban mobility: Analysis and comparison.. *Cities,* Volume 56, p. 35–46.

Gartner, 2017. *Gartner Magic Quadrant for Data Science Platforms.* [Online]
Available at: https://www.gartner.com/en/documents/3606026/magic-quadrant-for-data-science-platforms
[Accessed 20 04 2021].

Gartner, 2019. *Gartner Hype Cycle for Emerging Technologies.* [Online]
Available at: https://www.gartner.com/smarterwithgartner/5-trends-appear-on-the-gartner-hype-cycle-for-emerging-technologies-2019/
[Accessed 01 05 2021].

German Federal Institute for Research on Building, Urban Affairs and Spatial Development,, 2017. *Smart City Charta. Sustainable Development of Digital Transformation for Municipalities,,* s.l.: BBSR.

Giffinger, E., Fertner, R., Milanovic, C. & Meijers, N., 2007. Smart cities Ranking of European medium-sized cities..

Global Footprint Network, 2020. [Online]
Available at: https://www.overshootday.org/about/
[Accessed 23 05 2021].

Golden, T. & Veiga, J., 2005. The Impact of Extent of Telecommuting on Job Satisfaction: Resolving Inconsistent Findings.. *Journal of Management,* Volume 31(2), pp. 301-318.

Goteborg , 2021. *The world's most sustainable destination.* [Online]
Available at: https://www.goteborg.com/en/guides/the-worlds-most-sustainable-destination
[Accessed 21 05 2021].

Gudmundsson, H., 2003. Making concepts matter: Sustainable mobility and indicator systems in transport policy. *International Social Science Journal,* Volume 55(2), p. 173.

Höjer, M. & Wangel, J., 2014. Smart Sustainable Cities: Definition and Challenges. *ICT Innovations for Sustainability,* pp. 333-349.

Haghshenas, H. & Vaziri, M., 2012. Urban sustainable transportation indicators for global comparison. *Ecological Indicators ,* Volume 15, pp. 115-121.

Han, J. & Kamber, M., 2000. *Data Mining Concepts and Techniques.* s.l.:Simon Fraser University.

Heinrich, C., 2007. Evidence based Policy and Performance Management.. *The American Review of Public Administration,* Volume 37(3), pp. 255-277.

Hollanders, H., Es-Sadki, N. & Merkelbach, I., 2019a. *Regional Innovation Scoreboard 2019,* s.l.: European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs..

Hollanders, H., Es-Sadki, N. & Merkelbach, I., 2019b. *European innovation scoreboard 2019,* s.l.: European Commission, Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs..

Horbaty , R., 2014. Smart cities e intelligente vernetzung kommunaler infrastruktur.. p. 31.

Janssen , M., Matheus, , R. & Zui, A., 2015. *Big and Open Linked Data (BOLD) to Create Smart Cities and Citizens: Insights from Smart Energy and Mobility Cases.* s.l., Electronic Government.

Joumard, R. & Gudmundsson, H., 2010. *Indicators of environmental sustainability in transport..* France: l'INRETSed Broncedex.

JRC, 2020. *A dataset of GHG emissions for 6,200 cities in Europe and the Southern Mediterranean countries.* [Online]
Available at: https://data.jrc.ec.europa.eu/dataset/57a615eb-cfbc-435a-a8c5-553bd40f76c9
[Accessed 10 05 2021].

Kelleher, D. J., 2020. *Fundamentals of Machine Learning for Predictive Data Analytics.* United States: MIT Press Ltd.

Khan, Z., Anjum, A., Soomro, K. & Atif Tahir, M., 2015. Towards cloud based big data analytics for smart future cities. *Journal of Cloud Computing.*

KNIME, 2021. *Open for Innovation.* [Online]
Available at: https://www.knime.com
[Accessed 21 05 2021].

Komninos, N. & Mora, L., 2018. Exploring the big picture of smart city research.. *Italian Journal of Regional Science,* Volume 1, pp. 15-38.

Komninos, N., 2011. Intelligent cities: variable geometries of spatial intelligence.. *Intell Build Int ,* Volume 3, pp. 172-88.

Kona, A. et al., 2017. *Covenant of Mayors in figures: 8-year assessment,* s.l.: European Union.

Koutra, S., Becue, V. & Ioakimidis, C., 2019. Searching for the 'smart' definition through its spatial approach. *Energy,* Volume 169, pp. 924-936.

Lazaroiu, G. & Roscia, m., 2012. Definition methodology for the smart cities model.. *Energy,* Volume 47, p. 326–332.

Lee, J. H. H. M. G. &. H. M. C., 2014. Towards an effective framework for building smart cities: Lessons from Seoul and San Francisco.. *Technological Forecasting and Social Change,* Volume 89, pp. 80-99.

Leskovec, J., Rajaraman, A. & Ullman, J., 2012. *Mining of Massive Datasets.* s.l.:Cambridge University Press.

Lluïsa, M. & Llacuna, M., 2020. The people's smart city dashboard (PSCD): Delivering on community-led T governance with blockchain. *Technological Forecasting & Social Change,* Volume 158.

Lopez-CarreiroI., I. & Monzon, A., 2010. Evaluating sustainability and innovation of mobility patterns in Spanish cities. Analysis by size and urban typology. *Sustainable Cities and Society,* Volume 38, pp. 684-696.

Maes, J. et al., 2019. *Enhancing Resilience Of Urban Ecosystems through Green Infrastructure (EnRoute),* Luxembourg: Publications Office of the European Union.

Makridis, M. et al., 2018. *Estimating reaction time in Adaptive Cruise Control System.* s.l., s.n., p. 1312–1317.

Matheus, R., Janssen, M. & Maheshwari, D., 2018. Data science empowering the public: Data-driven dashboards for transparent and accountable decision-making in smart cities. *Government Information Quarterly.*

Mehmood, Y. et al., 2017. Internet-of-Things-Based Smart Cities: Recent Advances and Challenges. *IEEE Communications Magazine.*

Meijer, A. &. B. M. P., 2016. Governing the smart city: A review of the literature on smart urban governance.. *International Review of Administrative Sciences,* Volume 82, p. 392–408..

Misuraca, G. et al., 2010. Envisioning Digital Europe 2030: Scenarios for ICT in Future Provenance and Policy Modelling. *European Commission.*

Mori, K. & Christodoulou, A., 2012. Review of sustainability indices and indicators: Towards a new City Sustainability Index (CSI). *Environmental Impact Assessment Review,* Volume 32, p. 94–106.

Mukhopadhyay, A., Maulik, U., Bandyopadhyay, S. & Coello, C. A. C., 2014. A survey of multiobjective evolutionary algorithms for data mining: part IIEEE Transactions on Evolutionary Computation.

Municipality of Guimaraes, 2017. *Application form European Green Capital Award 2020,* s.l.: Municipality of Guimaraes.

Mystakidis, A. & Tjortjis, C., 2020. *Big Data Mining for Smart Cities: Predicting Traffic Congestion using Classification.* s.l., Intelligence, Systems and Applications (IISA 20).

OECD, 2019. *Government at a Glance 2019,* Paris: OECD Publishing.

OECD, 2020. *Cities in the World : A New Perspective on Urbanisation.* [Online]
Available at: https://www.oecd-ilibrary.org/sites/d0efcbda-en/index.html?itemId=/content/publication/d0efcbda-en
[Accessed 23 05 2021].

Oxford University, 1998. *Oxford Dictionart of English.* s.l.:Oxford University Press.

Pasimeni, F., Fiorini, A. & Georgakaki, A., 2018. *Patent-based Estimation Procedure of Private R&D: The Case of Climate Change and Mitigation Technologies in Europe.* s.l., s.n., pp. 1-22.

PBL Netherlands Environmental Assessment Afency, 2016. *Cities in Europe: Facts and figures on cities and urban areas,* Hague: s.n.

Peters, J., 2018. *Urban Mobility:Preparing for the Future, Learning from the Past,* s.l.: Eurocities.

Poelman, H., Dijkstra, L. & Ackermans, L., 2020. *HOW MANY PEOPLE CAN YOU REACH BY PUBLIC TRANSPORT, BICYCLE OR ON FOOT IN EUROPEAN CITIES?Measuring urban accessibility for low-carbon modes,* Luxembourg: Publications Office of the European Union.

PowerBI, 2021. *Create a data-driven culture with business intelligence for all.* [Online]
Available at: https://powerbi.microsoft.com/en-us/
[Accessed 21 05 2021].

PUM, 2019. *Promoting Mobility Behaviour Change,* s.l.: Walk21 Foundation.

PVsites, 2016. *D2.2 European climate zones and bio- climatic design requirements,* s.l.: H2020.

Rapidminer, 2021. *Depth for Data Scientists, Simplified for Everyone Else.* [Online]
Available at: https://rapidminer.com
[Accessed 21 05 2021].

Rudloff, C. et al., 2015. Influ- ence of weather on transport demand: A case study from the Vienna region. *Transportation Research Record: Journal of the Transportation Research Board,* Volume 1, p. 110–116..

Sáez, L., Heras-Saizarbitoria, I. & Rodríguez-Núñez, E., 2019. Sustainable city rankings, benchmarking and indexes: Looking into the black box. *Sustainable Cities and Society ,* Volume 53.

SAE International, 2016. *J3016 Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles ,* s.l.: s.n.

Sang, Y., Park, C. N., Lee, Y. & Lee, Y., 2018. *Scaling Relationship between Traffic Congestion versus Population Size of 164 Global Cities ,* s.l.: s.n.

Sankowska, P., 2018. Smart Government: An European Approach toward Building Sustainable and Secure Cities of Tomorrow. *International Journal of Technology,* Volume 9, p. 1355.

Schaffers, . H., 2012. *Empowering citizens to realizing smart cities: results from FIREBALL Smart city case studies..* Aalborg, s.n.

Shaheen, S. & Chan, N., 2016. Mobility and the Sharing Economy: Potential to Overcome First- and Last-Mile Public Transit Connections. *Built Environment.*

Smith, J., 2021. Another sample article. *IEEE Transactions on Smart Cities and Communities,* 1(2), pp. 55-58.

Statistics Norway, 2013. *Godt urbant miljø i «framtidens byer»?.* [Online]
Available at: https://www.ssb.no/natur-og-miljo/artikler-og-publikasjoner/godt-urbant-miljo-i-framtidens-byer
[Accessed 21 05 2021].

Stenvall, J., Syväjärvi, A. & Harisalo, R., 2007. The Information Steering in Government – Information Resource steered by Administration in Welfare Sector.. *Managing Worldwide Operations & Communications with Information Technology,* pp. 1395-1399.

Stevenson, M. et al., 2016. Land use, transport, and population health: estimating the health benefits of compact cities. *The lancet,* Volume 388(10062), pp. 2925-2935.

Sumathi, S. & Sivanandam, N., 2006. *Introduction to Data Mining and its Applications.* s.l.:Springer.

Switzerland's Federal Statistical Office, 2021. *https://www.bfs.admin.ch/bfs/en/home.html.* [Online]
Available at:
https://www.bfs.admin.ch/bfs/de/home/statistiken/volkswirtschaft/volkswirtschaftliche-gesamtrechnung/bruttoinlandprodukt-kanton.assetdetail.10647599.html
[Accessed 23 03 2021].

Taborda, S., Yiangou, G. & Georgouli, C., 2017. *Urban Mobility Innovation Indec,* s.l.: Future Cities Catapult.

Tafidis, P., Sdoukopoulos, A. & Pitsiava-Latinopoulou, M., 2017. Transportation Research Procedia. *Sustainable urban mobility indicators: Policy versus practice in the case of Greek cities. ,* Volume 24, p. 304–312.

Tanguay, G., Rajaonson, J., Lefebvre, J. & Lanoie, P., 2010. Measuring the sustaina- bility of cities: an analysis of the use of local indicators.. *Ecological Indicators 10,* p. 407–418.

Tan, P., Steinbach, M. & Kumar, V., 2014. *Introduction to Data Mining.* Essex: Pearson Education Limited.

TomTom, 2021. *Traffic Index.* [Online]
Available at: https://www.tomtom.com/en_gb/traffic-index/ranking/
[Accessed 20 04 2021].

Toppeta, D., 2010. The smart city vision: How innovation and ICT can build smart,livable, sustainable cities,. *The Innovation Knowledge Foundation.*

UK National Statistics, 2021. *UK National Statistics.* [Online]
Available at: www.ons.gov.uk
[Accessed 22 03 2021].

UNECE, 2011. *Climate Neutral Clities.How to make Cities Less Energy and Carbon Intensive and More Resilient to Climatic Challenges.,* New York and Geneva: United Nations,.

United Nations , 1992. *Human Development Report. .,* s.l.: United Kingdom: Oxford University Press.

United Nations, 2019. *World Urbanixation Prospects,* New York: United Nations.

Vandecasteele, I. et al., 2019. *The Future of Cities – Opportunities, challenges and the way forward,* Luxembourg: Publications Office.

Vartak, M. et al., 2016. Towards Visualization Recommendation Systems. *SIGMOD Record,* Volume 45, p. 4.

Veenstra, A. & Kotterink, B., 2017. *Data-Driven Policy Making: The Policy Lab Approach.* s.l., s.n.

Vercellis, C., 2011. *Business Intelligence: Data Mining and Optimization for Decision Making..* s.l.:Wiley.

WCED, 1987. *Our Common Future,* s.l.: New York: Oxford University Press.

Wikipedia , 2021. *Wikipedia: The free encyclopedia.* [Online]
Available at: https://www.wikipedia.org
[Accessed 18 05 2021].

WorldBankGroup, 2015. *Open Data for Suistanable Development | Transport and ICT,* s.l.: s.n.

Zhou, Z., 2003. Three perspectives of data mining. *Artificial Intelligence,* Volume 143, p. 139–146.

Zuccalà, M. & SergioVerga, E., 2017. Enabling Energy Smart Cities through Urban Sharing Ecosystems. *Energy Procedia,* Volume 111, pp. 826-835.

Appendix A: Sustainable Transport

Categories of Sustainable Transport (Haghshenas & Vaziri, 2012).

No	**Categories**	**Number of Indicators**	**Frequency of Use**
TEII	Categories of Transportation Environmental	33	90
1	Air pollution	5	30
2	Energy consumption	3	11
3	Renewable energy type	4	8
4	Efficient vehicle	6	7
5	Noise pollution	4	13
6	Land consumption	1	9
7	Environment management	2	2
8	Transport facility environment impact	2	2
9	Wild life	2	3
10	Other resource	4	5
TCII	Categories of Transportation Economical	25	48
1	Consumer direct cost and benefit	6	16
2	Consumer indirect cost and benefit	6	12
3	Transport price	2	2
4	Commercial transport	2	2
TSII	Categories of Transportation Social Impact	27	59
1	Safety	4	17
2	Access	6	16
3	Transport for disabled	1	4
4	Equity	6	8
5	Citizen participation in transport decision	6	7
6	Security	1	1
	Sum	85	197

Appendix B: System of indicators

Summarization of the final system of indicators

No	Sector	Category	Indicators
1	Environment	Air Quality	PM_{10} and NO_2 Annual Mean Concentration μg/m3
2		Noise	Percentage of population exposed in L_{den} higher than 55 dB
3		Land Consumption	Green public areas per capita
4		Renewable Energy	Percentage of renewable sources penetration on transport
5		Climate Change	Greenhouse gas (GHG) emissions from transport (tn/year)
6	Strategy & Governance	Strategy	The adaptation of SUMP, Data Sharing Strategy, MaaS Strategy and Covenant of Mayors
7		Governance	Score in European Quality of Government Index 2017
8	Mobility Performance	Accessibility	Percentage of Population with public transport stop within 500 m walking
9		Congestion	TomTom Traffic Index
10		Safety	People killed in road accidents per 10,000 people
11	Mobility Infrastructure	Alternative fuels – Renewable Energy	Electricity charging stations per 1,000 people.
12		Bicycle-Walkability	Length of bicycle network (dedicated cycle paths and lanes) –m per 1,000 capita
13		Mobility as a Service	Deployment of MaaS services
14	Innovation	Readiness- Competitiveness	EU Regional Innovation Score 2019
15		Investment	Horizon 2020 funding at NUTS3 level and H2020 transport funding at NUTS2 level (€ per capita)
16	Engagement	Citizens participation in decision making	City Moonshot Engagement Index